

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(12) Rettet oversættelse af
europæisk patentskrift

(10) **DK/EP 3650459 T5**

-
- (51) Int.Cl.: **C 07 K 7/00 (2006.01)** **B 82 Y 5/00 (2011.01)** **C 07 H 21/00 (2006.01)**
C 07 K 14/00 (2006.01) **C 07 K 16/00 (2006.01)** **C 07 K 17/00 (2006.01)**
C 40 B 70/00 (2006.01) **G 01 N 33/543 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2024-08-26**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2023-09-06**
- (86) Europæisk ansøgning nr.: **19196019.4**
- (86) Europæisk indleveringsdag: **2016-02-24**
- (87) Den europæiske ansøgnings publiceringsdag: **2020-05-13**
- (30) Prioritet: **2015-02-24 US 201562120262 P**
- (62) Stamansøgningsnr: **16756316.2**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **City of Hope, 1500 East Duarte Road, Duarte, California 91010, USA**
- (72) Opfinder: **COPELAND, Gregory, 785 W. 9th St., Claremont, California 91711, USA**
ELISON, Kathleen, 2784 N. Sierra Way, San Bernardino, California 92405, USA
MURADYAN, Hurik, 1506 Glenwood Road, Glendale, California 91201, USA
BERLIN, Jacob, 1074 Norumbega Dr., Monrovia, California 91016, USA
- (74) Fuldmægtig i Danmark: **NORDIC PATENT SERVICE A/S, Bredgade 30, 1260 København K, Danmark**
- (54) Benævnelse: **KEMISK KODEDE, RUMLIGT ADRESSEREDE BIBLIOTEKSCREENINGSPLATFORME**
- (56) Fremdragne publikationer:
EP-A2- 1 239 286
WO-A1-95/12608
WO-A1-98/20019
WO-A1-2011/143583
WO-A1-2014/026283
WO-A2-00/71995
WO-A2-99/67641
WO-A2-2005/123959
WO-A2-2007/084192
WO-A2-2010/093705
US-A- 4 672 040
US-A1- 2003 027 221
US-A1- 2006 040 286
US-B2- 8 039 271
JAVIER RAMON-AZCON ET AL: "Sensitive and Spatially Multiplexed Detection System Based on Dielectrophoretic Manipulation of DNA-Encoded Particles Used as Immunoreactions Platform", ANALYTICAL

Fortsættes ...

CHEMISTRY, vol. 83, no. 3, 1 February 2011 (2011-02-01), pages 1053-1060, XP055503799, US ISSN: 0003-2700, DOI: 10.1021/ac102854z
JAVIER RAMON-AZCON ET AL: "Sensitive and Spatially Multiplexed Detection System Based on Dielectrophoretic Manipulation of DNA-Encoded Particles Used as Immunoreactions Platform", ANALYTICAL CHEMISTRY, vol. 83, no. 3, 1 February 2011 (2011-02-01), US, pages 1053 - 1060, XP055503799, ISSN: 0003-2700, DOI: 10.1021/ac102854z

DESCRIPTION

BACKGROUND OF THE INVENTION

[0001] The number of molecules displayed by arrays used for screening and detection methods is restricted by the synthetic method of the array. In theory, split pool synthesis can generate enormous libraries -limited only by the number of chemical steps and number of unique building blocks utilized per step (i.e. a 5 step library utilizing 100 unique building blocks per step would in theory yield a 100^5 or 10 billion member chemical library). However, in practice, encoded split pool strategies face numerous practical constraints. Libraries that are decodable but not screenable or vice versa are not useful. The encoding strategy may be practically limited in the number and type of chemical steps or building blocks used. An encoded split pool library platform which requires large particles for decoding (e.g., by radio frequency tags or mass spectrometry) will normally need to contain fewer library members than a similar library that can be created on smaller particles. If assays are to be performed on a particle, the ligand density on each particle and the surface chemistry environment around each ligand should not interfere with the assay. The serial nature of reported decoding strategies also limits the number of "hits" which can be identified in a cost effective manner in a given screen, and therefore can limit the size of a library that is screened. The present invention addresses these and other problems in the art.

WO 95/12608 relates to a device and method for efficiently synthesizing diverse molecular products on substrates.

WO 2011/143583 discloses compositions and methods for assaying the presence of a target analyte in a sample using a solid support.

US 8,039,271 discloses a method and apparatus for the manipulation of colloidal particulates and biomolecules at the interface between an insulating electrode such as silicon oxide and an electrolyte solution.

WO 98/20019 discloses compositions containing at least one bead conjugated to a solid support and further conjugated to at least one macromolecule, such as a nucleic acid and methods for making the compositions.

US 2006/040286 discloses screening methods, compositions, and kits for detecting the presence or absence of one or more target analytes, e.g. biomolecules, in a sample.

WO 2005/123959 discloses high density multiplexing assays for proteins, nucleic acids, and other molecules based upon using mass spectrometry detection.

WO 2014/026283 relates to dendrimers conjugated to multiple targeting peptides and one or more therapeutic, diagnostic, or imaging agents for delivery of such agents across the blood-brain barrier and into certain cell types including, cells expressing the LRP-1 receptor.

US 2003/027221 discloses candidates for biological activity screened in capsules, each capsule containing a candidate, a target (which may be a molecular species or a biological cell or multiple copies of such a species or cell), and an intelligent substance.

WO 2007/084192 discloses a method for detecting an analyte of interest via a bio-barcode assay and a colorimetric bio-barcode method that is capable of detecting minute

concentrations of an analyte by relying on porous particles.

Ramon-Azcon et al. (Analytical Chemistry, 83, 3, 2011, pages 1053-1060) discloses immunodevices that combine competitive immunoreactions on microparticles, accumulation of the microparticles by negative dielectrophoresis (n-DEP), and their subsequent capture through hybridization among single-stranded DNAs (ssDNAs).

EP1239286 discloses a biochemical sensor with probes uniformly fixed in sections. The probes used for detecting a substance of interest are caught in advance on particles, and the particles are fixed in each of the sections arranged in the form of a lattice using a chemical patterning method on the surface of a baseplate.

WO 00/71995 discloses assays and components for encoding and decoding microspheres.

WO 99/67641 discloses compositions and methods for decoding microsphere array sensors.

WO 2010/093705 relates to the therapeutic use of novel stabilized oligoribonucleotides as immune modulatory agents for immune therapy applications.

US 4,672,040 discloses methods for the use of magnetically responsive particles in systems in which the separation of certain molecules, macromolecules and cells from the surrounding medium is desirable.

BRIEF SUMMARY OF THE INVENTION

[0002] In one aspect, there is provided a method comprising the steps of:

1. (i) providing a solid support comprising a plurality of wells;
2. (ii) attaching a plurality of microparticles to said plurality of wells to form an ordered array of a plurality of immobilized microparticles, wherein each of said plurality of immobilized microparticles is further covalently attached to: (a) a ligand domain through a first linker; and (b) an encoding tag through a second linker; wherein said ligand domain is a peptide, a protein, a macrocyclic peptide, a peptidomimetic, a polyamide, or a macrolactam; wherein said encoding tag is a nucleic acid sequence identifying said ligand domain; wherein said second linker is cleavable and said first linker is not cleavable under a condition that said second linker is cleavable; wherein said first linker comprises -C(O)NH- or -NHC(O)-, and said second linker comprises -C(O)O- or -OC(O)-; and
3. (iii) performing a decoding procedure on the encoding tag, thereby identifying the composition of the ligand domain and its location on said solid support; and
4. (iv) cleaving the second linkers of said plurality of immobilized microparticles, thereby forming a cleaved ordered array.

BRIEF DESCRIPTION OF THE DRAWINGS

[0003]

FIG. 1: Schematic illustrating general structure of microparticles used to create library. Variation in ratios of A, B, and C are achieved through one or more synthetic transformations applied to the starting material ProMag 1 COOH series particles. Initial target ratios are set by relative ratios of mixtures of building blocks used to assemble the particle surface. The final ratios achieved are measured through a combination of TGA, LC/MS or gel analysis of cleaved products, and colorimetric or fluorescence based solid phase assays. Preferred ranges: X has a concentration of 0.1 to 100 nanomoles/mg; $1\%X \leq A \leq 20\%X$; $40\%X \leq B \leq 99\%X$; $0\%X \leq C \leq 50\%X$; wherein X is A (library molecule attachment point), B (encoding tag attachment point) or C (core immobilization point).

FIG. 2: General schematic for how encoded, split pool synthesis can be performed; emphasizing the exponential increase in chemical diversity observed with a linear increase in the number of chemical steps and building blocks. For a general review, see: Czamik, A. W. "Encoding methods for combinatorial chemistry," *Curr. Opin. Chem. Biol.* 1997, 1, 60-66. Example: 100 building blocks, 6 rounds of split-pool, 1×10^{12} unique members, 600 chemical reactions. n is the number of splitpool rounds, x is the number of building blocks, unique library members = X^n , chemical reactions = n^x

FIG. 3: An illustration of the chemical bonds being formed and broken during one round of an encoded synthesis step using peptide synthesis as the chemistry which is being encoded. In step 1 an Fmoc protected amino acid is coupled to the free amine on the microparticle through an amide coupling. In step 2 a nucleic acid tag bearing an alkyne and free primary amine is coupled to the azide group on the microparticle through a copper catalyzed Huisgen cycloaddition. In step 3, an azide group is coupled to the primary amine on the nucleic acid through an amide coupling. In step 4 the Fmoc group protecting the amine group of the incorporated amino acid is removed. The product following step 4 displays the same reactive moieties as the starting material (prior to step 1), but contains an additional nucleic acid tag and an additional amino acid. See materials and methods for reaction details. See materials and methods for a detailed example of how initial microparticles are prepared for the orthogonal synthesis.

FIG. 4: Schematic illustrating the connectivity of the nucleic acid tags in one of the potential tagging approaches. In this variant, during each tagging step, all of the tag attachment points on the microparticles are consumed, and an equivalent amount of new tag attachment points are then created on the ends of the newly attached nucleic acid tags. Two tagging steps in this example would lead to the generation of predominantly a single population of nucleic acid containing oligomers on each bead consisting of tag 1 linked to tag 2. Note that by design, each tag population is independently decodable, such that decodability is only dependent on the presence of each individual tag-i.e. decodability of tag 2 is not dependent on being directly attached to tag 1 and vice versa-such that subsets of beads which may have not gone to complete conversion in either tagging step are still decodable, as long as sufficient amounts of each tag are present for decoding.

FIG. 5: Schematic illustrating the connectivity of the nucleic acid tags in one of the potential tagging approaches. In this variant, during each tagging step, only a fraction of the tag

attachment points on the microparticle for the tags are consumed. As shown, two tagging steps in this example would lead to the generation two populations of nucleic acid containing oligomers on each bead-one population consisting of tag 1 and one population consisting of tag 2. For four tagging steps, four unique populations per bead would be generated.

FIG. 6: Schematic illustrating the connectivity of the nucleic acid tags in one of the potential tagging approaches. In this variant, during each tagging step, a fraction of the tag attachment points on the microparticles are consumed, and an equivalent amount of new tag attachment points are then created on the ends of the newly attached nucleic acid tags. Two tagging steps in this example would lead to the generation of three populations of nucleic acid containing oligomers on each bead, tag 1, tag 2, and an oligomer consisting of tag 1 linked to tag 2. Note that by design, each tag population is independently decodable, such that decodability of each tag is only dependent on the presence of each individual tag-ie decodability of tag 2 is not dependent on being directly attached to tag 1 and vice versa.

FIG. 7: LC traces of the different tag intermediates during one round of tagging and the addition of a new tag linking element. As demonstrated by changes in retention time and mass spectra, the solid phase cycloaddition chemistry yields predominantly the desired product. Following the solid phase amide coupling of a new tag linking element, LC/MS analysis indicates conversion to the desired product. See materials and methods for cycloaddition conditions, amide coupling conditions, and cleavage conditions.

FIG. 8: A portion of the PEG modified microparticles was functionalized with an acid labile linker and submitted to a seven step solid phase peptide synthesis. The product was cleaved from the microparticles using TFA. The TFA was removed and residue was analyzed by mass spectrometry, demonstrating successful synthesis of the desired peptide.

FIG. 9A-9B: Demonstration of two different modes of library immobilization. In FIG. 9A, a brightfield image of a portion of Applicants' microparticles covalently immobilized to an activated carboxymethyldextran coated slide through amide bond formation. In FIG. 9B on right, an SEM image of a portion of Applicants' microparticles noncovalently immobilized within a custom microfabricated silicon wafer.

FIG. 10: Microparticles undergoing four rounds of encoding/tagging steps (as outlined in Encoding Scheme of FIG. 6 can be successfully decoded. Two different portions of microparticles were submitted in parallel to four rounds of tagging as illustrated in FIG. 6. Each portion of microparticles was tagged with a unique 18mer at each tagging step. Following the four rounds of encoding, each portion was exposed to a series of hybridization solutions containing a 1:1 mixture of two different fluorescently labeled oligonucleotide probes. Shown are eight different two-channel fluorescent images of aliquots of the tagged microparticles following hybridization. In column 1, the two microparticle samples were hybridized with a 1:1 mixture of Cy3 labeled oligonucleotide complementary to A0, and a FITC labeled oligonucleotide complementary to A1. As expected, those particles encoded with tag A0 fluoresced in the red (Cy3) channel (top image, column 1) and those particles encoded with tag A1 fluoresced in the green (FITC) channel (bottom image, column 2). The trend holds true

for all of the other hybridization conditions, in which particles selectively fluoresce in the wavelength corresponding to the fluorescent label attached to the complementary DNA sequence tag, rather than a mismatched tag.

FIG. 11: Microparticles undergoing four rounds of encoding/tagging steps (as outlined in FIG. 5) can be successfully decoded. Two different portions of microparticles were submitted in parallel to four rounds of tagging as illustrated in FIG. 5. Each portion of microparticles was tagged with a unique 18mer at each tagging step. Following the four rounds of encoding, each portion was exposed to a series of hybridization solutions containing a 1:1 mixture of two different fluorescently labeled oligonucleotide probes. Shown are eight different two-channel fluorescent images of aliquots of the tagged microparticles following hybridization. In column 1, the two microparticle samples were hybridized with a 1:1 mixture of Cy3 labeled oligonucleotide complementary to A0, and a FITC labeled oligonucleotide complementary to A1. As expected, those particles encoded with tag A0 fluoresced in the red (Cy3) channel (top image, column 1) and those particles encoded with tag A1 fluoresced in the green (FITC) channel (bottom image, column 2). The trend holds true for all of the other hybridization conditions, in which particles selectively fluoresce in the wavelength corresponding to the fluorescent label attached to the complementary DNA sequence tag, rather than a mismatched tag.

FIG. 12: To demonstrate peptide synthesis is compatible with Applicants' encoding chemistry, two peptides (HA and Myc--two common antibody epitopes) were synthesized in parallel on the microparticles in which four of the synthetic steps were encoded with DNA tags. The particles were incubated with a Cy5 labeled antibody specific to the HA epitope at various steps in the synthesis. Microparticles that displayed DNA tags and fully protected HA peptide (in which the side chain protecting moieties have not been removed) did not bind the labeled antibody (top row images). Following cleavage of the DNA tags, the microparticles that display side chain protected HA peptide do not bind the labeled antibody (second row images). Following deprotection of the side chain protecting moieties (TFA treatment), the HA displaying microparticles now bind the labeled antibody (third row images). Microparticles displaying fully deprotected Myc peptide does not bind the labeled antibody (fourth row images)

FIG. 13: To demonstrate peptide synthesis is compatible with Applicants' encoding chemistry, two peptides (HA and Myc--two common antibody epitopes) were synthesized in parallel on microparticles in which four of the synthetic steps were encoded with DNA tags. The particles were incubated with an Alexafluor 488 labeled antibody specific to the Myc epitope at various steps in the synthesis. Microparticles that displayed DNA tags and fully protected Myc peptide (in which the side chain protecting moieties have not been removed) did not bind the labeled antibody (top row images). Following cleavage of the DNA tags, the microparticles that display side chain protected Myc peptide do not bind the labeled antibody (second row images). Following deprotection of the side chain protecting moieties (TFA treatment), the Myc displaying microparticles now bind the labeled antibody (third row images). Microparticles displaying fully deprotected HA peptide does not bind the labeled antibody (fourth row images)

FIG. 14: Mixing the beads and stains demonstrates ability to differentiate binding. Tagless beads bearing fully deprotected HA peptide or Myc peptide were mixed and stained with a

mixture of Anti-HA-Cy5 and Anti-Myc-Alexa 488. Fluorescence imaging indicates two distinct bead populations.

FIG. 15A- FIG. 15H: FIG. 15A shows SEM of the 0.88 μm Promag particles. FIG. 15B shows a Silicon chip with 1.3 μm center to center spacing of wells in a hexagonal pattern. FIG. 15C shows a Silicon chip with 1.3 μm center to center spacing of wells in a square pattern. FIG. 15D shows a Silicon chip with 1.3 μm center to center spacing of wells in a square pattern- partially filled with microspheres. FIG. 15E shows a Silicon chip with 1.3 μm center to center spacing of wells in a hexagonal pattern, filled with microspheres. FIG. 15F shows a Silicon chip with 1.3 μm center to center spacing of wells in a hexagonal pattern, filled with microspheres- tilted. FIG. 15G shows a fluorescent Image of Alexafluor 488-labeled microspheres immobilized in partially filled silicon chip with 1.3 μm center to center spacing of wells in a hexagonal pattern. FIG. 15H shows a bright field image of microspheres immobilized in a partially filled quartz chip with 1.3 μm center to center spacing of wells in a hexagonal pattern.

FIG. 16A- FIG. 16D: FIG. 16A shows autofluorescence of microspheres immobilized in a partially filled quartz chip with 1.3 μm center to center spacing of wells in a hexagonal pattern. FIG. 16B shows a Silicon chip with 2.4 μm center to center spacing of wells in a hexagonal pattern. FIG. 16C shows a Silicon chip with 2.4 μm center to center spacing of wells in a hexagonal pattern , partially filled with microspheres. FIG. 16D shows a fluorescent image of DNA tagged-microspheres hybridized with fluorescently labeled DNA complements immobilized in a silicon chip. The chip is partially filled and has 2.4 μm center to center spacing of wells in a hexagonal pattern.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS

[0004] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0005] The abbreviations used herein have their conventional meaning within the chemical and biological arts. The chemical structures and formulae set forth herein are constructed according to the standard rules of chemical valency known in the chemical arts.

[0006] Where substituent groups are specified by their conventional chemical formulae, written from left to right, they equally encompass the chemically identical substituents that would result from writing the structure from right to left, e.g., $-\text{CH}_2\text{O}-$ is equivalent to $-\text{OCH}_2-$.

[0007] The term "alkyl," by itself or as part of another substituent, means, unless otherwise

stated, a straight (i.e., unbranched) or branched non-cyclic carbon chain (or carbon), or combination thereof, which may be fully saturated, mono- or polyunsaturated and can include di- and multivalent radicals, having the number of carbon atoms designated (i.e., C₁-C₁₀ means one to ten carbons). Examples of saturated hydrocarbon radicals include, but are not limited to, groups such as methyl, ethyl, n-propyl, isopropyl, n-butyl, t-butyl, isobutyl, sec-butyl, (cyclohexyl)methyl, homologs and isomers of, for example, n-pentyl, n-hexyl, n-heptyl, n-octyl, and the like. An unsaturated alkyl group is one having one or more double bonds or triple bonds. Examples of unsaturated alkyl groups include, but are not limited to, vinyl, 2-propenyl, crotyl, 2-isopentenyl, 2-(butadienyl), 2,4-pentadienyl, 3-(1,4-pentadienyl), ethynyl, 1- and 3-propynyl, 3-butylnyl, and the higher homologs and isomers. An alkoxy is an alkyl attached to the remainder of the molecule via an oxygen linker (-O-). An alkyl moiety may be an alkenyl moiety. An alkyl moiety may be an alkynyl moiety. An alkyl moiety may be fully saturated.

[0008] The term "alkylene, " by itself or as part of another substituent, means, unless otherwise stated, a divalent radical derived from an alkyl, as exemplified, but not limited by, -CH₂CH₂CH₂CH₂-. Typically, an alkyl (or alkylene) group will have from 1 to 24 carbon atoms, with those groups having 10 or fewer carbon atoms being preferred in the present invention. A "lower alkyl" or "lower alkylene" is a shorter chain alkyl or alkylene group, generally having eight or fewer carbon atoms. The term "alkenylene," by itself or as part of another substituent, means, unless otherwise stated, a divalent radical derived from an alkene.

[0009] The term "heteroalkyl," by itself or in combination with another term, means, unless otherwise stated, a stable non-cyclic straight or branched chain, or combinations thereof, including at least one carbon atom and at least one heteroatom selected from the group consisting of O, N, P, Si, and S, and wherein the nitrogen and sulfur atoms may optionally be oxidized, and the nitrogen heteroatom may optionally be quaternized. The heteroatom(s) O, N, P, S, and Si may be placed at any interior position of the heteroalkyl group or at the position at which the alkyl group is attached to the remainder of the molecule. Examples include, but are not limited to: -CH₂-CH₂-O-CH₃, -CH₂-CH₂-NH-CH₃, -CH₂-CH₂-N(CH₃)-CH₃, -CH₂-S-CH₂-CH₃, -CH₂-C H₂, -S(O)-CH₃, -CH₂-CH₂-S(O)₂-CH₃, -CH=CH-O-CH₃, -Si(CH₃)₃, -CH₂-CH=N-OCH₃, -CH=CH-N(CH₃)-CH₃, -O-CH₃, -O-CH₂-CH₃, and -CN. Up to two or three heteroatoms may be consecutive, such as, for example, -CH₂-NH-OCH₃ and -CH₂-O-Si(CH₃)₃. A heteroalkyl moiety may include one heteroatom (e.g., O, N, S, Si, or P). A heteroalkyl moiety may include two optionally different heteroatoms (e.g., O, N, S, Si, or P). A heteroalkyl moiety may include three optionally different heteroatoms (e.g., O, N, S, Si, or P). A heteroalkyl moiety may include four optionally different heteroatoms (e.g., O, N, S, Si, or P). A heteroalkyl moiety may include five optionally different heteroatoms (e.g., O, N, S, Si, or P). A heteroalkyl moiety may include up to 8 optionally different heteroatoms (e.g., O, N, S, Si, or P).

[0010] Similarly, the term "heteroalkylene," by itself or as part of another substituent, means, unless otherwise stated, a divalent radical derived from heteroalkyl, as exemplified, but not limited by, -CH₂-CH₂-S-CH₂-CH₂- and -CH₂-S-CH₂-CH₂-NH-CH₂-. For heteroalkylene groups, heteroatoms can also occupy either or both of the chain termini (e.g., alkyleneoxy,

alkylenedioxy, alkyleneamino, alkylenediamino, and the like). Still further, for alkylene and heteroalkylene linking groups, no orientation of the linking group is implied by the direction in which the formula of the linking group is written. For example, the formula $-C(O)_2R'$ represents both $-C(O)_2R'$ and $-R'C(O)_2$. As described above, heteroalkyl groups, as used herein, include those groups that are attached to the remainder of the molecule through a heteroatom, such as $-C(O)R'$, $-C(O)NR'$, $-NR'R''$, $-OR'$, $-SR'$, and/or $-SO_2R'$. Where "heteroalkyl" is recited, followed by recitations of specific heteroalkyl groups, such as $-NR'R''$ or the like, it will be understood that the terms heteroalkyl and $-NR'R''$ are not redundant or mutually exclusive. Rather, the specific heteroalkyl groups are recited to add clarity. Thus, the term "heteroalkyl" should not be interpreted herein as excluding specific heteroalkyl groups, such as $-NR'R''$ or the like.

[0011] The terms "cycloalkyl" and "heterocycloalkyl," by themselves or in combination with other terms, mean, unless otherwise stated, non-aromatic cyclic versions of "alkyl" and "heteroalkyl," respectively, wherein the carbons making up the ring or rings do not necessarily need to be bonded to a hydrogen due to all carbon valencies participating in bonds with non-hydrogen atoms. Additionally, for heterocycloalkyl, a heteroatom can occupy the position at which the heterocycle is attached to the remainder of the molecule. Examples of cycloalkyl include, but are not limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, 1-cyclohexenyl, 3-cyclohexenyl, cycloheptyl, 3-hydroxy-cyclobut-3-enyl-1,2, dione, 1H-1,2,4-triazolyl-5(4H)-one, 4H-1,2,4-triazolyl, and the like. Examples of heterocycloalkyl include, but are not limited to, 1-(1,2,5,6-tetrahydropyridyl), 1-piperidiny, 2-piperidiny, 3-piperidiny, 4-morpholinyl, 3-morpholinyl, tetrahydrofuran-2-yl, tetrahydrofuran-3-yl, tetrahydrothien-2-yl, tetrahydrothien-3-yl, 1-piperazinyl, 2-piperazinyl, and the like. A "cycloalkylene" and a "heterocycloalkylene," alone or as part of another substituent, means a divalent radical derived from a cycloalkyl and heterocycloalkyl, respectively. A heterocycloalkyl moiety may include one ring heteroatom (e.g., O, N, S, Si, or P). A heterocycloalkyl moiety may include two optionally different ring heteroatoms (e.g., O, N, S, Si, or P). A heterocycloalkyl moiety may include three optionally different ring heteroatoms (e.g., O, N, S, Si, or P). A heterocycloalkyl moiety may include four optionally different ring heteroatoms (e.g., O, N, S, Si, or P). A heterocycloalkyl moiety may include five optionally different ring heteroatoms (e.g., O, N, S, Si, or P). A heterocycloalkyl moiety may include up to 8 optionally different ring heteroatoms (e.g., O, N, S, Si, or P).

[0012] The terms "halo" or "halogen," by themselves or as part of another substituent, mean, unless otherwise stated, a fluorine, chlorine, bromine, or iodine atom. Additionally, terms such as "haloalkyl" are meant to include monohaloalkyl and polyhaloalkyl. For example, the term "halo(C₁-C₄)alkyl" includes, but is not limited to, fluoromethyl, difluoromethyl, trifluoromethyl, 2,2,2-trifluoroethyl, 4-chlorobutyl, 3-bromopropyl, and the like.

[0013] The term "acyl" means, unless otherwise stated, $-C(O)R$ where R is a substituted or unsubstituted alkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl.

[0014] The term "aryl" means, unless otherwise stated, a polyunsaturated, aromatic, hydrocarbon substituent, which can be a single ring or multiple rings (preferably from 1 to 3 rings) that are fused together (i.e., a fused ring aryl) or linked covalently. A fused ring aryl refers to multiple rings fused together wherein at least one of the fused rings is an aryl ring. The term "heteroaryl" refers to aryl groups (or rings) that contain at least one heteroatom such as N, O, or S, wherein the nitrogen and sulfur atoms are optionally oxidized, and the nitrogen atom(s) are optionally quaternized. Thus, the term "heteroaryl" includes fused ring heteroaryl groups (i.e., multiple rings fused together wherein at least one of the fused rings is a heteroaromatic ring). A 5,6-fused ring heteroarylene refers to two rings fused together, wherein one ring has 5 members and the other ring has 6 members, and wherein at least one ring is a heteroaryl ring. Likewise, a 6,6-fused ring heteroarylene refers to two rings fused together, wherein one ring has 6 members and the other ring has 6 members, and wherein at least one ring is a heteroaryl ring. And a 6,5-fused ring heteroarylene refers to two rings fused together, wherein one ring has 6 members and the other ring has 5 members, and wherein at least one ring is a heteroaryl ring. A heteroaryl group can be attached to the remainder of the molecule through a carbon or heteroatom. Non-limiting examples of aryl and heteroaryl groups include phenyl, 1-naphthyl, 2-naphthyl, 4-biphenyl, 1-pyrrolyl, 2-pyrrolyl, 3-pyrrolyl, 3-pyrazolyl, 2-imidazolyl, 4-imidazolyl, pyrazinyl, 2-oxazolyl, 4-oxazolyl, 2-phenyl-4-oxazolyl, 5-oxazolyl, 3-isoxazolyl, 4-isoxazolyl, 5-isoxazolyl, 2-thiazolyl, 4-thiazolyl, 5-thiazolyl, 2-furyl, 3-furyl, 2-thienyl, 3-thienyl, 2-pyridyl, 3-pyridyl, 4-pyridyl, 2-pyrimidyl, 4-pyrimidyl, 5-benzothiazolyl, purinyl, 2-benzimidazolyl, 5-indolyl, 1-isoquinolyl, 5-isoquinolyl, 2-quinoxaliny, 5-quinoxaliny, 3-quinolyl, and 6-quinolyl. Substituents for each of the above noted aryl and heteroaryl ring systems are selected from the group of acceptable substituents described below. An "arylene" and a "heteroarylene," alone or as part of another substituent, mean a divalent radical derived from an aryl and heteroaryl, respectively. Non-limiting examples of aryl and heteroaryl groups include pyridinyl, pyrimidinyl, thiophenyl, thienyl, furanyl, indolyl, benzoxadiazolyl, benzodioxolyl, benzodioxanyl, thianaphthanyl, pyrrolopyridinyl, indazolyl, quinolinyl, quinoxaliny, pyridopyrazinyl, quinazolinonyl, benzoisoxazolyl, imidazopyridinyl, benzofuranyl, benzothiophenyl, phenyl, naphthyl, biphenyl, pyrrolyl, pyrazolyl, imidazolyl, pyrazinyl, oxazolyl, isoxazolyl, thiazolyl, furylthienyl, pyridyl, pyrimidyl, benzothiazolyl, purinyl, benzimidazolyl, isoquinolyl, thiadiazolyl, oxadiazolyl, pyrrolyl, diazolyl, triazolyl, tetrazolyl, benzothiadiazolyl, isothiazolyl, pyrazolopyrimidinyl, pyrrolopyrimidinyl, benzotriazolyl, benzoxazolyl, or quinolyl. The examples above may be substituted or unsubstituted and divalent radicals of each heteroaryl example above are non-limiting examples of heteroarylene. A heteroaryl moiety may include one ring heteroatom (e.g., O, N, or S). A heteroaryl moiety may include two optionally different ring heteroatoms (e.g., O, N, or S). A heteroaryl moiety may include three optionally different ring heteroatoms (e.g., O, N, or S). A heteroaryl moiety may include four optionally different ring heteroatoms (e.g., O, N, or S). A heteroaryl moiety may include five optionally different ring heteroatoms (e.g., O, N, or S). An aryl moiety may have a single ring. An aryl moiety may have two optionally different rings. An aryl moiety may have three optionally different rings. An aryl moiety may have four optionally different rings. A heteroaryl moiety may have one ring. A heteroaryl moiety may have two optionally different rings. A heteroaryl moiety may have three optionally different rings. A heteroaryl moiety may have four optionally different rings. A heteroaryl moiety may have five optionally different rings.

[0015] A fused ring heterocycloalkyl-aryl is an aryl fused to a heterocycloalkyl. A fused ring heterocycloalkyl-heteroaryl is a heteroaryl fused to a heterocycloalkyl. A fused ring heterocycloalkyl-cycloalkyl is a heterocycloalkyl fused to a cycloalkyl. A fused ring heterocycloalkyl-heterocycloalkyl is a heterocycloalkyl fused to another heterocycloalkyl. Fused ring heterocycloalkyl-aryl, fused ring heterocycloalkyl-heteroaryl, fused ring heterocycloalkyl-cycloalkyl, or fused ring heterocycloalkyl-heterocycloalkyl may each independently be unsubstituted or substituted with one or more of the substituents described herein.

[0016] The term "oxo," as used herein, means an oxygen that is double bonded to a carbon atom.

[0017] The term "alkylsulfonyl," as used herein, means a moiety having the formula $-S(O_2)-R'$, where R' is a substituted or unsubstituted alkyl group as defined above. R' may have a specified number of carbons (e.g., "C₁-C₄ alkylsulfonyl").

[0018] Each of the above terms (e.g., "alkyl," "heteroalkyl," "cycloalkyl," "heterocycloalkyl," "aryl," and "heteroaryl") includes both substituted and unsubstituted forms of the indicated radical. Preferred substituents for each type of radical are provided below.

[0019] Substituents for the alkyl and heteroalkyl radicals (including those groups often referred to as alkylene, alkenyl, heteroalkylene, heteroalkenyl, alkynyl, cycloalkyl, heterocycloalkyl, cycloalkenyl, and heterocycloalkenyl) can be one or more of a variety of groups selected from, but not limited to, $-OR'$, $=O$, $=NR'$, $=N-OR'$, $-NR'R''$, $-SR'$, $-halogen$, $-SiR'R''R'''$, $-OC(O)R'$, $-C(O)R'$, $-CO_2R'$, $-CONR'R''$, $-OC(O)NR'R''$, $-NR''C(O)R'$, $-NR'-C(O)NR''R'''$, $-NR''C(O)_2R'$, $-NR-C(NR'R'')=NR'''$, $-S(O)R'$, $-S(O)_2R'$, $-S(O)_2N(R)('R''-NRSO_2R')$, $-CN$, and $-NO_2$ in a number ranging from zero to $(2m'+1)$, where m' is the total number of carbon atoms in such radical. R' , R'' , R''' , and R'''' each preferably independently refer to hydrogen, substituted or unsubstituted heteroalkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl (e.g., aryl substituted with 1-3 halogens), substituted or unsubstituted alkyl, alkoxy, or thioalkoxy groups, or arylalkyl groups. When a compound used in the invention includes more than one R group, for example, each of the R groups is independently selected as are each R' , R'' , R''' , and R'''' group when more than one of these groups is present. When R' and R'' are attached to the same nitrogen atom, they can be combined with the nitrogen atom to form a 4-, 5-, 6-, or 7-membered ring. For example, $-NR'R''$ includes, but is not limited to, 1-pyrrolidinyl and 4-morpholinyl. From the above discussion of substituents, one of skill in the art will understand that the term "alkyl" is meant to include groups including carbon atoms bound to groups other than hydrogen groups, such as haloalkyl (e.g., $-CF_3$ and $-CH_2CF_3$) and acyl (e.g., $-C(O)CH_3$, $-C(O)CF_3$, $-C(O)CH_2OCH_3$, and the like).

[0020] Similar to the substituents described for the alkyl radical, substituents for the aryl and heteroaryl groups are varied and are selected from, for example: $-OR'$, $-NR'R''$, $-SR'$, $-halogen$, $-SiR'R''R'''$, $-OC(O)R'$, $-C(O)R'$, $-CO_2R'$, $-CONR'R''$, $-OC(O)NR'R''$, $-NR''C(O)R'$, $-NR'-C(O)NR''R'''$

"', $\text{NR}''\text{C}(\text{O})_2\text{R}'$, $\text{NRC}(\text{NR}'\text{R}'')=\text{NR}'''$, $\text{S}(\text{O})\text{R}'$, $-\text{S}(\text{O})_2\text{R}'$, $-\text{S}(\text{O})_2\text{N}(\text{R}')(\text{R}'')$, $-\text{NRSO}_2\text{R}'$), $-\text{CN}$, $-\text{NO}_2$, $-\text{R}'$, $-\text{N}_3$, $-\text{CH}(\text{Ph})_2$, fluoro($\text{C}_1\text{-C}_4$)alkoxy, and fluoro($\text{C}_1\text{-C}_4$)alkyl, in a number ranging from zero to the total number of open valences on the aromatic ring system; and where R' , R'' , R''' , and R'''' are preferably independently selected from hydrogen, substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl, and substituted or unsubstituted heteroaryl. When a compound used in the invention includes more than one R group, for example, each of the R groups is independently selected as are each R' , R'' , R''' , and R'''' groups when more than one of these groups is present.

[0021] Where a moiety is substituted with an R substituent, the group may be referred to as "R-substituted." Where a moiety is R-substituted, the moiety is substituted with at least one R substituent and each R substituent is optionally different. For example, where a moiety herein is $\text{R}^{1\text{A}}$ -substituted or unsubstituted alkyl, a plurality of $\text{R}^{1\text{A}}$ substituents may be attached to the alkyl moiety wherein each $\text{R}^{1\text{A}}$ substituent is optionally different. Where an R-substituted moiety is substituted with a plurality R substituents, each of the R-substituents may be differentiated herein using a prime symbol (') such as R' , R'' , etc. For example, where a moiety is $\text{R}^{1\text{A}}$ -substituted or unsubstituted alkyl, and the moiety is substituted with a plurality of $\text{R}^{1\text{A}}$ substituents, the plurality of $\text{R}^{1\text{A}}$ substituents may be differentiated as $\text{R}^{1\text{A}'}$, $\text{R}^{1\text{A}''}$, $\text{R}^{1\text{A}'''}$, etc. In some embodiments, the plurality of R substituents is 3. In some embodiments, the plurality of R substituents is 2.

[0022] In embodiments, a compound as described herein may include multiple instances of R^1 , R^2 , R^3 , R^4 , R^5 , R^6 , $\text{R}^{6\text{A}}$, R^7 , R^8 , R^9 , R^{10} , R^{11} , R^{12} , R^{13} , R^{14} , R^{15} , R^{16} , R^{17} , R^{18} , R^{19} , R^{20} , R^{21} , R^{22} , R^{23} , and/or other substituents and variables. In such embodiments, each variable may optionally be different and be appropriately labeled to distinguish each group for greater clarity. For example, where each $\text{R}^{6\text{A}}$ is different, they may be referred to, for example, as $\text{R}^{6\text{A}.1}$, $\text{R}^{6\text{A}.2}$, $\text{R}^{6\text{A}.3}$, or $\text{R}^{6\text{A}.4}$, respectively, wherein the definition of $\text{R}^{6\text{A}}$ is assumed by $\text{R}^{6\text{A}.1}$, $\text{R}^{6\text{A}.2}$, $\text{R}^{6\text{A}.3}$, and/or $\text{R}^{6\text{A}.4}$. The variables used within a definition of R^1 , R^2 , R^3 , R^4 , R^5 , R^6 , $\text{R}^{6\text{A}}$, R^7 , R^8 , R^9 , R^{10} , R^{11} , R^{12} , R^{13} , R^{14} , R^{15} , R^{16} , R^{17} , R^{18} , R^{19} , R^{20} , R^{21} , R^{22} , R^{23} , and/or other variables that appear at multiple instances and are different may similarly be appropriately labeled to distinguish each group for greater clarity.

[0023] Two or more substituents may optionally be joined to form aryl, heteroaryl, cycloalkyl, or heterocycloalkyl groups. Such so-called ring-forming substituents are typically, though not necessarily, found attached to a cyclic base structure. In one embodiment, the ring-forming substituents are attached to adjacent members of the base structure. For example, two ring-forming substituents attached to adjacent members of a cyclic base structure create a fused ring structure. In another embodiment, the ring-forming substituents are attached to a single member of the base structure. For example, two ring-forming substituents attached to a single member of a cyclic base structure create a spirocyclic structure. In yet another embodiment,

the ring-forming substituents are attached to non-adjacent members of the base structure.

[0024] Two of the substituents on adjacent atoms of the aryl or heteroaryl ring may optionally form a ring of the formula $-T-C(O)-(CRR')_q-U-$, wherein T and U are independently $-NR-$, $-O-$, $-CRR'-$, or a single bond, and q is an integer of from 0 to 3. Alternatively, two of the substituents on adjacent atoms of the aryl or heteroaryl ring may optionally be replaced with a substituent of the formula $-A-(CH_2)_r-B-$, wherein A and B are independently $-CRR'-$, $-O-$, $-NR-$, $-S-$, $-S(O)-$, $-S(O)_2-$, $-S(O)_2NR'-$, or a single bond, and r is an integer of from 1 to 4. One of the single bonds of the new ring so formed may optionally be replaced with a double bond. Alternatively, two of the substituents on adjacent atoms of the aryl or heteroaryl ring may optionally be replaced with a substituent of the formula $-(CRR')_s-X'-(C''R''R''')_d-$, where variables s and d are independently integers of from 0 to 3, and X' is $-O-$, $-NR'-$, $-S-$, $-S(O)-$, $-S(O)_2-$, or $-S(O)_2NR'-$. The substituents R, R', R'', and R''' are preferably independently selected from hydrogen, substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl, and substituted or unsubstituted heteroaryl.

[0025] As used herein, the terms "heteroatom" or "ring heteroatom" are meant to include, oxygen (O), nitrogen (N), sulfur (S), phosphorus (P), and silicon (Si).

[0026] A "substituent group," as used herein, means a group selected from the following moieties:

1. (A) oxo, halogen, $-CF_3$, $-CN$, $-OH$, $-NH_2$, $-COOH$, $-CONH_2$, $-NO_2$, $-SH$, $-SO_2Cl$, $-SO_3H$, $-SO_4H$, $-SO_2NH_2$, $-NHNH_2$, $-ONH_2$, $-NHC(=O)NHNH_2$, $-NHC(=O)NH_2$, $-NHSO_2H$, $-NHC(=O)H$, $-NHC(O)-OH$, $-NHOH$, $-OCF_3$, $-OCHF_2$, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl, and
2. (B) alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, heteroaryl, substituted with at least one substituent selected from:
 1. (i) oxo, halogen, $-CF_3$, $-CN$, $-OH$, $-NH_2$, $-COOH$, $-CONH_2$, $-NO_2$, $-SH$, $-SO_2Cl$, $-SO_3H$, $-SO_4H$, $-SO_2NH_2$, $-NHNH_2$, $-ONH_2$, $-NHC(=O)NHNH_2$, $-NHC(=O)NH_2$, $-NHSO_2H$, $-NHC(=O)H$, $-NHC(O)-OH$, $-NHOH$, $-OCF_3$, $-OCHF_2$, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl, and
 2. (ii) alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, heteroaryl, substituted with at least one substituent selected from:
 1. (a) oxo, halogen, $-CF_3$, $-CN$, $-OH$, $-NH_2$, $-COOH$, $-CONH_2$, $-NO_2$, $-SH$, $-SO_2Cl$, $-SO_3H$, $-SO_4H$, $-SO_2NH_2$, $-NHNH_2$, $-ONH_2$, $-NHC(=O)NHNH_2$, $-NHC(=O)NH_2$, $-NHSO_2H$, $-NHC(=O)H$, $-NHC(O)-OH$, $-NHOH$, $-OCF_3$, $-OCHF_2$, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl,

and

2. (b) alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, heteroaryl, substituted with at least one substituent selected from: oxo, halogen, -CF₃, -CN, -OH, -NH₂, -COOH, -CONH₂, -NO₂, -SH, -SO₂Cl, -SO₃H, -SO₄H, -SO₂NH₂, -NHNH₂, -ONH₂, -NHC(=O)NHNH₂, -NHC(=O)NH₂, -NHSO₂H, -NHC(=O)H, -NHC(O)-OH, -NHOH, -OCF₃, -OCHF₂, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl.

[0027] A "size-limited substituent" or "size-limited substituent group," as used herein, means a group selected from all of the substituents described above for a "substituent group," wherein each substituted or unsubstituted alkyl is a substituted or unsubstituted C₁-C₂₀ alkyl, each substituted or unsubstituted heteroalkyl is a substituted or unsubstituted 2 to 20 membered heteroalkyl, each substituted or unsubstituted cycloalkyl is a substituted or unsubstituted C₃-C₈ cycloalkyl, each substituted or unsubstituted heterocycloalkyl is a substituted or unsubstituted 3 to 8 membered heterocycloalkyl, each substituted or unsubstituted aryl is a substituted or unsubstituted C₆-C₁₀ aryl, and each substituted or unsubstituted heteroaryl is a substituted or unsubstituted 5 to 10 membered heteroaryl.

[0028] A "lower substituent" or "lower substituent group," as used herein, means a group selected from all of the substituents described above for a "substituent group," wherein each substituted or unsubstituted alkyl is a substituted or unsubstituted C_i-C_s alkyl, each substituted or unsubstituted heteroalkyl is a substituted or unsubstituted 2 to 8 membered heteroalkyl, each substituted or unsubstituted cycloalkyl is a substituted or unsubstituted C₃-C₇ cycloalkyl, each substituted or unsubstituted heterocycloalkyl is a substituted or unsubstituted 3 to 7 membered heterocycloalkyl, each substituted or unsubstituted aryl is a substituted or unsubstituted C₆-C₁₀ aryl, and each substituted or unsubstituted heteroaryl is a substituted or unsubstituted 5 to 9 membered heteroaryl.

[0029] In some embodiments, each substituted group described in the compounds herein is substituted with at least one substituent group. More specifically, in some embodiments, each substituted alkyl, substituted heteroalkyl, substituted cycloalkyl, substituted heterocycloalkyl, substituted aryl, substituted heteroaryl, substituted alkylene, substituted heteroalkylene, substituted cycloalkylene, substituted heterocycloalkylene, substituted arylene, and/or substituted heteroarylene described in the compounds herein are substituted with at least one substituent group. In other embodiments, at least one or all of these groups are substituted with at least one size-limited substituent group. In other embodiments, at least one or all of these groups are substituted with at least one lower substituent group.

[0030] In other embodiments of the compounds herein, each substituted or unsubstituted alkyl may be a substituted or unsubstituted C₁-C₂₀ alkyl, each substituted or unsubstituted

heteroalkyl is a substituted or unsubstituted 2 to 20 membered heteroalkyl, each substituted or unsubstituted cycloalkyl is a substituted or unsubstituted C₃-C₈ cycloalkyl, each substituted or unsubstituted heterocycloalkyl is a substituted or unsubstituted 3 to 8 membered heterocycloalkyl, each substituted or unsubstituted aryl is a substituted or unsubstituted C₆-C₁₀ aryl, and/or each substituted or unsubstituted heteroaryl is a substituted or unsubstituted 5 to 10 membered heteroaryl. In some embodiments of the compounds herein, each substituted or unsubstituted alkylene is a substituted or unsubstituted C₁-C₂₀ alkylene, each substituted or unsubstituted heteroalkylene is a substituted or unsubstituted 2 to 20 membered heteroalkylene, each substituted or unsubstituted cycloalkylene is a substituted or unsubstituted C₃-C₈ cycloalkylene, each substituted or unsubstituted heterocycloalkylene is a substituted or unsubstituted 3 to 8 membered heterocycloalkylene, each substituted or unsubstituted arylene is a substituted or unsubstituted C₆-C₁₀ arylene, and/or each substituted or unsubstituted heteroarylene is a substituted or unsubstituted 5 to 10 membered heteroarylene.

[0031] In some embodiments, each substituted or unsubstituted alkyl is a substituted or unsubstituted C₁-C₈ alkyl, each substituted or unsubstituted heteroalkyl is a substituted or unsubstituted 2 to 8 membered heteroalkyl, each substituted or unsubstituted cycloalkyl is a substituted or unsubstituted C₃-C₇ cycloalkyl, each substituted or unsubstituted heterocycloalkyl is a substituted or unsubstituted 3 to 7 membered heterocycloalkyl, each substituted or unsubstituted aryl is a substituted or unsubstituted C₆-C₁₀ aryl, and/or each substituted or unsubstituted heteroaryl is a substituted or unsubstituted 5 to 9 membered heteroaryl. In some embodiments, each substituted or unsubstituted alkylene is a substituted or unsubstituted C₁-C₈ alkylene, each substituted or unsubstituted heteroalkylene is a substituted or unsubstituted 2 to 8 membered heteroalkylene, each substituted or unsubstituted cycloalkylene is a substituted or unsubstituted C₃-C₇ cycloalkylene, each substituted or unsubstituted heterocycloalkylene is a substituted or unsubstituted 3 to 7 membered heterocycloalkylene, each substituted or unsubstituted arylene is a substituted or unsubstituted C₆-C₁₀ arylene, and/or each substituted or unsubstituted heteroarylene is a substituted or unsubstituted 5 to 9 membered heteroarylene. In some embodiments, the compound is a chemical species set forth in the Examples section, figures, or tables below.

[0032] The terms "a" or "an," as used in herein means one or more. In addition, the phrase "substituted with a[n]," as used herein, means the specified group may be substituted with one or more of any or all of the named substituents. For example, where a group, such as an alkyl or heteroaryl group, is "substituted with an unsubstituted C₁-C₂₀ alkyl, or unsubstituted 2 to 20 membered heteroalkyl," the group may contain one or more unsubstituted C₁-C₂₀ alkyls, and/or one or more unsubstituted 2 to 20 membered heteroalkyls. Moreover, where a moiety is substituted with an R substituent, the group may be referred to as "R-substituted." Where a moiety is R-substituted, the moiety is substituted with at least one R substituent and each R substituent is optionally different.

[0033] The symbol

""

denotes the point of attachment of a chemical moiety to the remainder of a molecule or chemical formula.

[0034] Descriptions of compounds used in the present invention are limited by principles of chemical bonding known to those skilled in the art. Accordingly, where a group may be substituted by one or more of a number of substituents, such substitutions are selected so as to comply with principles of chemical bonding and to give compounds which are not inherently unstable and/or would be known to one of ordinary skill in the art as likely to be unstable under ambient conditions, such as aqueous, neutral, and several known physiological conditions. For example, a heterocycloalkyl or heteroaryl is attached to the remainder of the molecule via a ring heteroatom in compliance with principles of chemical bonding known to those skilled in the art thereby avoiding inherently unstable compounds.

[0035] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by a person of ordinary skill in the art. See, e.g., Singleton et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY 2nd ed., J. Wiley & Sons (New York, NY 1994); Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL, Cold Springs Harbor Press (Cold Springs Harbor, NY 1989). The following definitions are provided to facilitate understanding of certain terms used frequently herein and are not meant to limit the scope of the present disclosure.

[0036] "Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single-, double- or multiple-stranded form, or complements thereof. The term "polynucleotide" refers to a linear sequence of nucleotides. The term "nucleotide" typically refers to a single unit of a polynucleotide, i.e., a monomer. Nucleotides can be ribonucleotides, deoxyribonucleotides, or modified versions thereof. Examples of polynucleotides contemplated herein include single and double stranded DNA, single and double stranded RNA (including siRNA), and hybrid molecules having mixtures of single and double stranded DNA and RNA. Nucleic acids can be linear or branched. For example, nucleic acids can be a linear chain of nucleotides or the nucleic acids can be branched, e.g., such that the nucleic acids comprise one or more arms or branches of nucleotides. Optionally, the branched nucleic acids are repetitively branched to form higher ordered structures such as dendrimers and the like.

[0037] Nucleic acids, including nucleic acids with a phosphothioate backbone can include one or more reactive moieties. As used herein, the term reactive moiety includes any group capable of reacting with another molecule, e.g., a nucleic acid or polypeptide through covalent, non-covalent or other interactions. By way of example, the nucleic acid can include an amino acid reactive moiety that reacts with an amino acid on a protein or polypeptide through a covalent, non-covalent or other interaction.

[0038] The terms also encompass nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-

naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphodiester derivatives including, e.g., phosphoramidate, phosphorodiamidate, phosphorothioate (also known as phosphothioate), phosphorodithioate, phosphonocarboxylic acids, phosphonocarboxylates, phosphonoacetic acid, phosphonoformic acid, methyl phosphonate, boron phosphonate, or O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, modified sugars, and non-ribose backbones (e.g. phosphorodiamidate morpholino oligos or locked nucleic acids (LNA)), including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, *Carbohydrate Modifications in Antisense Research*, Sanghui & Cook, eds. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g., to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. In embodiments, the internucleotide linkages in DNA are phosphodiester, phosphodiester derivatives, or a combination of both.

[0039] Nucleic acids can include nonspecific sequences. As used herein, the term "nonspecific sequence" refers to a nucleic acid sequence that contains a series of residues that are not designed to be complementary to or are only partially complementary to any other nucleic acid sequence. By way of example, a nonspecific nucleic acid sequence is a sequence of nucleic acid residues that does not function as an inhibitory nucleic acid when contacted with a cell or organism. An "inhibitory nucleic acid" is a nucleic acid (e.g. DNA, RNA, polymer of nucleotide analogs) that is capable of binding to a target nucleic acid (e.g. an mRNA translatable into a protein) and reducing transcription of the target nucleic acid (e.g. mRNA from DNA) or reducing the translation of the target nucleic acid (e.g.mRNA) or altering transcript splicing (e.g. single stranded morpholino oligo).

[0040] A "labeled nucleic acid or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the nucleic acid may be detected by detecting the presence of the detectable label bound to the nucleic acid. Alternatively, a method using high affinity interactions may achieve the same results where one of a pair of binding partners binds to the other, e.g., biotin, streptavidin. In embodiments, the nucleic acid domain includes a detectable label, as disclosed herein and generally known in the art.

[0041] The term "probe" or "primer", as used herein, is defined to be one or more nucleic acid fragments whose specific hybridization to a sample can be detected. A probe or primer can be of any length depending on the particular technique it will be used for. For example, PCR

primers are generally between 10 and 40 nucleotides in length, while nucleic acid probes for, e.g., a Southern blot, can be more than a hundred nucleotides in length. The probe may be unlabeled or labeled as described below so that its binding to the target or sample can be detected. The probe can be produced from a source of nucleic acids from one or more particular (preselected) portions of a chromosome, e.g., one or more clones, an isolated whole chromosome or chromosome fragment, or a collection of polymerase chain reaction (PCR) amplification products. The length and complexity of the nucleic acid fixed onto the target element is not critical to the invention. One of skill can adjust these factors to provide optimum hybridization and signal production for a given hybridization procedure, and to provide the required resolution among different genes or genomic locations.

[0042] The probe may also be isolated nucleic acids immobilized on a solid surface (e.g., nitrocellulose, glass, quartz, fused silica slides), as in an array. In some embodiments, the probe may be a member of an array of nucleic acids as described, for instance, in WO 96/17958. Techniques capable of producing high density arrays can also be used for this purpose (see, e.g., Fodor (1991) *Science* 767-773; Johnston (1998) *Curr. Biol.* 8: R171-R174; Schummer (1997) *Biotechniques* 23: 1087-1092; Kern (1997) *Biotechniques* 23: 120-124; U.S. Patent No. 5,143,854).

[0043] The words "complementary" or "complementarity" refer to the ability of a nucleic acid in a polynucleotide to form a base pair with another nucleic acid in a second polynucleotide. For example, the sequence A-G-T is complementary to the sequence T-C-A. Complementarity may be partial, in which only some of the nucleic acids match according to base pairing, or complete, where all the nucleic acids match according to base pairing.

[0044] Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a pre-sequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a pre-protein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are near each other, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

[0045] The term "gene" means the segment of DNA involved in producing a protein; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons). The leader, the trailer as well as the introns include regulatory elements that are necessary during the transcription and the translation of a gene. Further, a "protein gene product" is a protein expressed from a particular gene.

[0046] The word "expression" or "expressed" as used herein in reference to a gene means the transcriptional and/or translational product of that gene. The level of expression of a DNA molecule in a cell may be determined on the basis of either the amount of corresponding mRNA that is present within the cell or the amount of protein encoded by that DNA produced by the cell. The level of expression of non-coding nucleic acid molecules (e.g., siRNA) may be detected by standard PCR or Northern blot methods well known in the art. See, Sambrook et al., 1989 Molecular Cloning: A Laboratory Manual, 18.1-18.88.

[0047] The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all. Transgenic cells and plants are those that express a heterologous gene or coding sequence, typically as a result of recombinant methods.

[0048] The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion

The term "exogenous" refers to a molecule or substance (e.g., a compound, nucleic acid or protein) that originates from outside a given cell or organism. For example, an "exogenous promoter" as referred to herein is a promoter that does not originate from the plant it is expressed by. Conversely, the term "endogenous" or "endogenous promoter" refers to a molecule or substance that is native to, or originates within, a given cell or organism.

[0049] The term "isolated", when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It can be, for example, in a homogeneous state and may be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified.

[0050] The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. In some embodiments, the nucleic acid or protein is at least 50% pure, optionally at least 65% pure, optionally at least 75% pure, optionally at least 85% pure, optionally at least 95% pure, and optionally at least 99% pure.

[0051] The term "isolated" may also refer to a cell or sample cells. An isolated cell or sample cells are a single cell type that is substantially free of many of the components which normally accompany the cells when they are in their native state or when they are initially removed from their native state. In certain embodiments, an isolated cell sample retains those components from its natural state that are required to maintain the cell in a desired state. In some embodiments, an isolated (e.g. purified, separated) cell or isolated cells, are cells that are substantially the only cell type in a sample. A purified cell sample may contain at least 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of one type of cell. An isolated cell sample may be obtained through the use of a cell marker or a combination of cell markers, either of which is unique to one cell type in an unpurified cell sample. In some embodiments, the cells are isolated through the use of a cell sorter. In some embodiments, antibodies against cell proteins are used to isolate cells.

[0052] As used herein, the term "conjugate" refers to the association between atoms or molecules. The association can be direct or indirect. For example, a conjugate between a nucleic acid and a protein can be direct, e.g., by covalent bond, or indirect, e.g., by non-covalent bond (e.g. electrostatic interactions (e.g. ionic bond, hydrogen bond, halogen bond), van der Waals interactions (e.g. dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like). In embodiments, conjugates are formed using conjugate chemistry including, but are not limited to nucleophilic substitutions (e.g., reactions of amines and alcohols with acyl halides, active esters), electrophilic substitutions (e.g., enamine reactions) and additions to carbon-carbon and carbon-heteroatom multiple bonds (e.g., Michael reaction, Diels-Alder addition). These and other useful reactions are discussed in, for example, March, ADVANCED ORGANIC CHEMISTRY, 3rd Ed., John Wiley & Sons, New York, 1985; Hermanson, BIOCONJUGATE TECHNIQUES, Academic Press, San Diego, 1996; and Feeney et al., MODIFICATION OF PROTEINS; Advances in Chemistry Series, Vol. 198, American Chemical Society, Washington, D.C., 1982. In embodiments, the microparticle is non-covalently attached to a solid support through a non-covalent chemical reaction between a component of the microparticle and a component of the solid support. In other embodiments, the microparticle includes one or more reactive moieties, e.g., a covalent reactive moiety, as described herein (e.g., an amine reactive moiety). In other embodiments, the microparticle includes a linker with one or more reactive moieties, e.g., a covalent reactive moiety, as described herein (e.g., an amine reactive moiety).

[0053] Useful reactive moieties or reactive functional groups used for conjugate chemistries herein include, for example:

1. (a) carboxyl groups and various derivatives thereof including, but not limited to, N-hydroxysuccinimide esters, N-hydroxybenztriazole esters, acid halides, acyl imidazoles, thioesters, p-nitrophenyl esters, alkyl, alkenyl, alkynyl and aromatic esters;
2. (b) hydroxyl groups which can be converted to esters, ethers, aldehydes, etc.
3. (c) haloalkyl groups wherein the halide can be later displaced with a nucleophilic group such as, for example, an amine, a carboxylate anion, thiol anion, carbanion, or an alkoxide ion, thereby resulting in the covalent attachment of a new group at the site of

the halogen atom;

4. (d) dienophile groups which are capable of participating in Diels-Alder reactions such as, for example, maleimido groups;
5. (e) aldehyde or ketone groups such that subsequent derivatization is possible via formation of carbonyl derivatives such as, for example, imines, hydrazones, semicarbazones or oximes, or via such mechanisms as Grignard addition or alkyllithium addition;
6. (f) sulfonyl halide groups for subsequent reaction with amines, for example, to form sulfonamides;
7. (g) thiol groups, which can be converted to disulfides, reacted with acyl halides, or bonded to metals such as gold;
8. (h) amine or sulfhydryl groups, which can be, for example, acylated, alkylated or oxidized;
9. (i) alkenes, which can undergo, for example, cycloadditions, acylation, Michael addition, etc;
10. (j) epoxides, which can react with, for example, amines and hydroxyl compounds;
11. (k) phosphoramidites and other standard reactive moieties useful in nucleic acid synthesis;
12. (l) metal silicon oxide bonding;
13. (m) metal bonding to reactive phosphorus groups (e.g. phosphines) to form, for example, phosphate diester bonds; and
14. (n) sulfones, for example, vinyl sulfone.

[0054] The reactive moieties can be chosen such that they do not participate in, or interfere with, the chemical stability of the proteins or nucleic acids described herein. By way of example, the nucleic acids can include a vinyl sulfone or other reactive moiety (e.g., maleimide). Optionally, the nucleic acids can include a reactive moiety having the formula S-S-R. R can be, for example, a protecting moiety. Optionally, R is hexanol. As used herein, the term hexanol includes compounds with the formula $C_6H_{13}OH$ and includes, 1-hexanol, 2-hexanol, 3-hexanol, 2-methyl-1-pentanol, 3-methyl-1-pentanol, 4-methyl-1-pentanol, 2-methyl-2-pentanol, 3-methyl-2-pentanol, 4-methyl-2-pentanol, 2-methyl-3-pentanol, 3-methyl-3-pentanol, 2,2-dimethyl-1-butanol, 2,3-dimethyl-1-butanol, 3,3-dimethyl-1-butanol, 2,3-dimethyl-2-butanol, 3,3-dimethyl-2-butanol, and 2-ethyl-1-butanol. Optionally, R is 1-hexanol.

[0055] As used herein, the term "about" means a range of values including the specified value, which a person of ordinary skill in the art would consider reasonably similar to the specified value. In embodiments, the term "about" means within a standard deviation using measurements generally acceptable in the art. In embodiments, about means a range extending to +/- 10% of the specified value. In embodiments, about means the specified value.

[0056] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues, wherein the polymer may be conjugated to a moiety

that does not consist of amino acids. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. The terms apply to macrocyclic peptides, peptides that have been modified with non-peptide functionality, peptidomimetics, polyamides, and macrolactams. A "fusion protein" refers to a chimeric protein encoding two or more separate protein sequences that are recombinantly expressed as a single moiety.

[0057] The term "peptidyl" and "peptidyl moiety" means a monovalent peptide.

[0058] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid. The terms "non-naturally occurring amino acid" and "unnatural amino acid" refer to amino acid analogs, synthetic amino acids, and amino acid mimetics which are not found in nature.

[0059] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0060] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each

silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual probe sequences.

[0061] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

[0062] The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, *Proteins* (1984)).

[0063] "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0064] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity over a specified region, e.g., of the entire polypeptide sequences as disclosed or individual domains of the polypeptides as disclosed), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the complement of a test sequence. Optionally, the identity exists over a region that is at least about 50 nucleotides in length, or more preferably over a region that is 100 to 500 or 1000 or more nucleotides in length.

[0065] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and

reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0066] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of, e.g., a full length sequence or from 20 to 600, about 50 to about 200, or about 100 to about 150 amino acids or nucleotides in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1970) *Adv. Appl. Math.* 2:482c, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology* (1995 supplement)).

[0067] An example of an algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nuc. Acids Res.* 25:3389-3402, and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad.*

Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[0068] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0069] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross-reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

[0070] "Contacting" is used in accordance with its plain ordinary meaning and refers to the process of allowing at least two distinct species (e.g. chemical compounds including biomolecules or cells) to become sufficiently proximal to react, interact or physically touch. It should be appreciated; however, the resulting reaction product can be produced directly from a reaction between the added reagents or from an intermediate from one or more of the added reagents which can be produced in the reaction mixture.

[0071] The term "contacting" may include allowing two species to react, interact, or physically touch, wherein the two species may be, for example, a ligand domain as described herein and a ligand binder. In embodiments contacting includes, for example, allowing a ligand domain as described herein to interact with a ligand binder.

[0072] A "control" sample or value refers to a sample that serves as a reference, usually a known reference, for comparison to a test sample. For example, a test sample can be taken from a test condition, e.g., in the presence of a test compound, and compared to samples from known conditions, e.g., in the absence of the test compound (negative control), or in the presence of a known compound (positive control). A control can also represent an average value gathered from a number of tests or results. One of skill in the art will recognize that controls can be designed for assessment of any number of parameters. For example, a control can be devised to compare therapeutic benefit based on pharmacological data (e.g., half-life) or therapeutic measures (e.g., comparison of side effects). One of skill in the art will understand which standard controls are most appropriate in a given situation and be able to

analyze data based on comparisons to standard control values. Standard controls are also valuable for determining the significance (e.g. statistical significance) of data. For example, if values for a given parameter are widely variant in standard controls, variation in test samples will not be considered as significant.

[0073] A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, useful labels include ^{32}P , fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins or other entities which can be made detectable, e.g., by incorporating a radiolabel into a peptide or antibody specifically reactive with a target peptide. Any appropriate method known in the art for conjugating an antibody to the label may be employed, e.g., using methods described in Hermanson, *Bioconjugate Techniques* 1996, Academic Press, Inc., San Diego.

[0074] A "labeled protein or polypeptide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the labeled protein or polypeptide may be detected by detecting the presence of the label bound to the labeled protein or polypeptide. Alternatively, methods using high affinity interactions may achieve the same results where one of a pair of binding partners binds to the other, e.g., biotin, streptavidin.

[0075] "Biological sample" or "sample" refer to materials obtained from or derived from a subject or patient. A biological sample includes sections of tissues such as biopsy and autopsy samples, and frozen sections taken for histological purposes. Such samples include bodily fluids such as blood and blood fractions or products (e.g., serum, plasma, platelets, red blood cells, and the like), sputum, tissue, cultured cells (e.g., primary cultures, explants, and transformed cells) stool, urine, synovial fluid, joint tissue, synovial tissue, synoviocytes, fibroblast-like synoviocytes, macrophage-like synoviocytes, immune cells, hematopoietic cells, fibroblasts, macrophages, T cells, etc. A biological sample is typically obtained from a eukaryotic organism, such as a mammal such as a primate e.g., chimpanzee or human; cow; dog; cat; a rodent, e.g., guinea pig, rat, mouse; rabbit; or a bird; reptile; or fish.

[0076] A "cell" as used herein, refers to a cell carrying out metabolic or other function sufficient to preserve or replicate its genomic DNA. A cell can be identified by well-known methods in the art including, for example, presence of an intact membrane, staining by a particular dye, ability to produce progeny or, in the case of a gamete, ability to combine with a second gamete to produce a viable offspring. Cells may include prokaryotic and eukaryotic cells. Prokaryotic cells include but are not limited to bacteria. Eukaryotic cells include but are not limited to yeast cells and cells derived from plants and animals, for example mammalian, insect (e.g., spodoptera) and human cells.

[0077] The term "antibody" is used according to its commonly known meaning in the art. Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests

an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_{H1} by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (see Fundamental Immunology (Paul ed., 3d ed. 1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv) or those identified using phage display libraries (see, e.g., McCafferty et al., Nature 348:552-554 (1990)).

[0078] For preparation of monoclonal or polyclonal antibodies, any technique known in the art can be used (see, e.g., Kohler & Milstein, Nature 256:495-497 (1975); Kozbor et al., Immunology Today 4:72 (1983); Cole et al., pp. 77-96 in Monoclonal Antibodies and Cancer Therapy (1985)). "Monoclonal" antibodies (mAb) refer to antibodies derived from a single clone. Techniques for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce antibodies to polypeptides used in this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies. Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (see, e.g., McCafferty et al., Nature 348:552-554 (1990); Marks et al., Biotechnology 10:779-783 (1992)).

[0079] A "solid support" as provided herein refers to any material that can be modified to contain discrete individual sites appropriate for the attachment or association of a microparticle as provided herein including embodiments thereof and is amenable to the methods provided herein including embodiments thereof. Examples of solid supports include without limitation, glass and modified or functionalized glass (e.g., carboxymethyl dextran functionalized glass), plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon.TM., etc.), polysaccharides, nylon or nitrocellulose, composite materials, ceramics, and plastic resins, silica or silica-based materials including silicon and modified silicon (e.g., patterned silicon), carbon, metals, quartz (e.g., patterned quartz), inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In general, the substrates allow optical detection and do not appreciably fluoresce.

[0080] The solid support provided herein including embodiments thereof may form part of an ion-sensitive field-effect transistor (ISFET) microarray. The solid support may be planar (e.g., flat planar substrates such as glass, polystyrene and other plastics and acrylics). Although it will be appreciated by a person of ordinary skill in the art that other configurations of solid supports may be used as well, for example, three dimensional configurations can be used. The solid support may be modified to contain discrete, individual sites (also referred to herein as "wells") for microparticle binding. These sites generally include physically altered sites, i.e. physical configurations such as wells or small depressions in the substrate that can retain the microparticles. The wells may be formed using a variety of techniques well known in the art,

including, but not limited to, photolithography, stamping techniques, molding techniques and microetching techniques. It will be appreciated by a person of ordinary skill in the art that the technique used will depend on the composition and shape of the solid support. In embodiments, physical alterations are made in a surface of the solid support to produce wells. The required depth of the wells will depend on the size of the microparticle to be added to the well.

[0081] A "microparticle" as used herein refers to a non-planar (e.g. spherical) particle having a size sufficient to attach molecules (e.g., a first, a second or a third linker provided, a ligand domain and a nucleic acid domain), directly or indirectly, through either covalent or non-covalent bonds. The microparticle may include any material that is capable of providing physical support for the molecules (e.g., a first, a second or a third linker provided, a ligand domain and a nucleic acid domain) that are attached to the surface. The material is generally capable of enduring conditions related to the attachment of the molecules (e.g., a first, a second or a third linker provided, a ligand domain and a nucleic acid domain) to the surface and any subsequent treatment, handling, or processing encountered during the performance of an assay. The materials may be naturally occurring, synthetic, or a modification of a naturally occurring material. Suitable microparticle materials may include silicon, ceramics, plastics (including polymers such as, e.g., poly(vinyl chloride), cyclo-olefin copolymers, agarose, polyacrylamide, polyacrylate, polyethylene, polypropylene, poly(4-methylbutene), polystyrene, polymethacrylate, poly(ethylene terephthalate), polytetrafluoroethylene (PTFE or Teflon.RTM.), nylon, poly(vinyl butyrate)), germanium, gallium arsenide, gold or silver, copper or aluminum surfaces, magnetic surfaces, e.g. Fe, Mn, Ni, Co, and their oxides, quantum dots, e.g., III-V (GaN, GaP, GaAs, InP, or InAs) or II-VI (ZnO, ZnS, CdS, CdSe, or CdTe) semiconductors, or Ln-doped fluoride nanocrystals, rare earth-doped oxidic nanomaterials either used by themselves or in conjunction with other materials. Additional rigid materials may be considered, such as glass, which includes silica and further includes, for example, glass that is available as Bioglass. Other materials that may be employed include porous materials, such as, for example, controlled pore glass beads, crosslinked beaded Sepharose.RTM. or agarose resins, or copolymers of crosslinked bis-acrylamide and azalactone. Other beads include polymer beads, solid core beads, paramagnetic beads, or microbeads. Any other materials known in the art that are capable of having one or more moieties, such as any of an amino, carboxyl, thiol, or hydroxyl reactive moiety, for example, incorporated on its surface, are also contemplated. In embodiments, the microparticle is a magnetic polymer-based sphere. In embodiments, the microparticle is a ProMag™ microsphere. In embodiments, the longest dimension of the microparticle is less than 1000 μm.

COMPOSITIONS

[0082] The compositions provided herein are, *inter alia*, useful for the assembly of highly dense arrays suitable for a variety of high throughput screening methods. The microparticles provided herein include a ligand domain attached through a first linker and a nucleic acid

domain attached through a second linker. By binding to a solid support the microparticles provided herein including embodiments thereof form part of an array. The ligand domain and the nucleic acid domain are synthesized on the microparticle using methods of encoded split pool chemistry. Encoded split pool chemistry is a method well known in the art and described, *inter alia*, by the following references: Furka, A.; et al. Int. J. Pept. Protein Res. 1991, 37, 487-493; Kit Lam et al., Nature, 1991; 354: 82-84; US Patent 6,060,596; 5,770,358; 6,368,874; 5,565,324; 6,936,477 and 5,573,905. Each step of the ligand domain synthesis (e.g., peptide or chemical compound synthesis) is encoded in the nucleic acid domain by a short nucleic acid sequence serving as an identification bar code. Therefore, each microparticle includes a unique ligand domain and a corresponding nucleic acid domain encoding specific nucleic acid sequences. The specific nucleic acid sequences correspond to the building blocks of the ligand domain and the order in which they were incorporated in the ligand domain. Upon hybridization of a complementary nucleic acid to said nucleic acid domain, the composition of the ligand domain and its location on the array can be determined (decoded). After the identity of the ligand domain and its location on the array have been determined, the nucleic acid domain is removed, the ligand domain may be further modified and contacted with a ligand binder (e.g., biomolecule).

[0083] In methods of the invention, a plurality of microparticles are provided. Each microparticle is covalently attached to a ligand domain through a first linker; and a nucleic acid domain through a second linker, wherein the second linker is cleavable and the first linker is not cleavable under a condition that the second linker is cleavable.

[0084] In embodiments, the microparticle is a microbead. A "microbead" as referred to herein is a polymer-based microparticle of roughly spherical shape with a diameter of about 0.5 μm to about 500 μm . The term "polymer-based" or "polymeric" as provided herein refers to a microparticle or microbead including at least one polymer compound (e.g., polyethylene glycols, polyethylene imides, polysaccharides, polypeptides, or polynucleotides). In embodiments, the microbead is a ProMag™ microsphere. In embodiments, the microbead is a polymer-based magnetic microbead.

[0085] The microparticle provided herein including embodiments thereof may be less than 200 μm . Where the microparticle is less than 200 μm a person of ordinary skill in the art will immediately recognize that the longest dimension (e.g. diameter or length) of a microparticle is smaller than 200 μm . In other embodiments, the microparticle is about 20 nm. In some embodiments, the microparticle is from about 0.01 μm to about 200 μm , from about 0.02 μm to about 200 μm , from about 0.05 μm to about 200 μm , from about 0.1 μm to about 200 μm , from about 0.5 μm to about 200 μm , from about 1 μm to about 200 μm , from about 2 μm to about 200 μm , from about 5 μm to about 200 μm , from about 10 μm to about 200 μm , from about 15 μm to about 200 μm , from about 20 μm to about 200 μm , from about 25 μm to about 200 μm , from about 30 μm to about 200 μm , from about 35 μm to about 200 μm , from about 40 μm to about 200 μm , from about 45 μm to about 200 μm , from about 50 μm to about 200 μm , from about 55 μm to about 200 μm , from about 60 μm to about 200 μm , from about 65 μm to about 200 μm , from about 70 μm to about 200 μm , from about 75 μm to about 200 μm , from about

about 5 μm to about 10 μm .

[0090] In embodiments, the microparticle is about 0.9 μm . In embodiments, the microparticle has a diameter of about 0.9 μm . In embodiments, the microparticle is about 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 μm . In other embodiments, the microparticle has a diameter of about 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 μm . The numerical values above represent the size of the microparticle in μm .

[0091] In embodiments, the microparticle is a functionalized microbead. Where the microparticle is a functionalized microbead, the microparticle may include any reactive moiety suitable for the conjugate chemistries described herein. The term "functionalized" as provided herein refers to a compound or domain (e.g., microparticle, linker, ligand domain, nucleic acid domain, nucleic acid sequence) including a reactive moiety or reactive functional groups used for conjugate chemistries as described herein. For example, a functionalized microbead may include one or more reactive moieties, such as any of an amino, carboxyl, thiol, or hydroxyl reactive moiety, incorporated on its surface. In embodiments, a first functionalized group allows for attachment of the ligand domain through a first linker. In embodiments, a second functionalized group allows for attachment of the nucleic acid domain through a second linker. In embodiments, the first and the second functionalized group are independently different. Therefore, the ligand domain may be attached to the microparticle through a first linker by conjugation to a different functionalized group than the nucleic acid domain. In embodiments, a third functionalized group connects the microparticle to a solid support. Therefore, in embodiments the microparticle is covalently attached to a solid support.

[0092] The microparticle provided herein may include a polymer. In such a case the polymers will carry the reactive moieties to be activated. The polymer may be selected from any suitable class of compounds, for example, polyethylene glycols, polyethylene imides, polysaccharides, polypeptides, or polynucleotides. In embodiments, the microparticle includes bis-amino polyethyleneglycol 3000 and hydroxyl-polyethylene glycol 3000. In embodiments, the microparticle includes a polymer layer. Attachment of the polymers to the microparticle may be effected by a variety of methods which are readily apparent to a person skilled in the art. For example, polymers bearing trichlorosilyl or trisalkoxy groups may be reacted with hydroxyl groups on the microparticle to form siloxane bonds. Attachment to a gold or silver microparticle may take place via thiol groups on the polymer. Alternatively, the polymer may be attached via an intermediate species, such as a self-assembled monolayer of alkanethiols. The type of polymers selected, and the method selected for attaching the polymers to the microparticle, will thus depend on the polymer having suitable reactivity for being attached to the microparticle surface, and on the properties of the polymers regarding non-specific adsorption to, especially, DNA or peptides. The reactive moieties may be present on the polymer or may be added to

the polymer by the addition of single or multiple reactive moieties. Optionally, a spacer arm (e.g., linker) can be used to provide flexibility to the binding nucleic acid domain or ligand domain allowing it to interact with its environment in a way which minimizes steric hindrance with the microparticle.

[0093] In embodiments, the functionalized microbead is a magnetic polymer-based (polymeric) microbead. In embodiments, the microbead is a ProMag™ microspheres. In embodiments, the microbead includes more than one polymer. In embodiments, the microbead includes a first polymer and a second polymer, wherein the first polymer and the second polymer are chemically different. In embodiments, the first polymer is bis-amino polyethyleneglycol 3000 and the second polymer is hydroxyl-polyethylene glycol 3000. In embodiments, the first polymer includes a first reactive moiety and the second polymer includes a second reactive moiety. A reactive moiety as referred to herein includes any of the functional moieties useful for conjugate chemistry as described herein. In embodiments, the first reactive moiety is an amino functional group and the second reactive moiety is a hydroxyl functional group. In embodiments, the hydroxyl functional group is reacted to form an azidoacetate moiety. In embodiments, the first reactive moiety (e.g., amino functional group) is reacted with a reactive moiety (e.g., a carboxyl functional group) of the first linker. In embodiments, the azidoacetate moiety is reacted with a reactive moiety (e.g., an alkynyl functional group) of the second linker.

[0094] In embodiments, the microparticle is a polymeric microbead. In embodiments, the microparticle is a dendrimer. A "dendrimer" as referred to herein is a spherical polymeric molecule made from two monomers (e.g., acrylic acid and a diamine). Dendrimers are precisely defined chemical structures that consist of a series of chemical shells built on a small core molecule. Each shell consists of two chemicals, always in the same order. In embodiments, the microparticle is a branched polymer. In embodiments, the microparticle is a magnetic polymeric microbead. In embodiments, the microparticle is a carboxymethyl dextran functionalized microbead. In embodiments, the microparticle is a polyethylene glycol functionalized microbead. In further embodiments, the polyethylene glycol functionalized microbead includes orthogonally protected amines. In embodiments, the microparticle is a magnetic microbead. In embodiments, the microparticle is a metallic microbead. In embodiments, the microparticle is a silica microbead.

[0095] As depicted in FIG. 1 the microparticles provided herein including embodiments thereof may include a plurality of attachment points for the attachment of a plurality of a first, second and third linker. The microparticle may include a plurality of first attachment points for the first linker attaching the ligand domain, a plurality of second attachment points for the second linker attaching the nucleic acid domain and a plurality of third attachment points for the third linker attaching the microparticle to a solid support. The total number of attachment points per microparticle may be about 25-50 attomoles. Where the total number of attachment points corresponds to 100%, the number of first attachment points may be more than about 1% and less than about 20%. Where the total number of attachment points corresponds to 100%, the number of second attachment points may be more than about 40% and less than about 90%. Where the total number of attachment points corresponds to 100%, the number of third

attachment points may be more than about 0% and less than about 50%.

[0096] A "ligand domain" as provided herein is a domain capable of binding a ligand binder (e.g., analyte, biomolecule). In embodiments, the ligand domain is a peptide. In embodiments, the ligand domain is a polypeptide. In embodiments, the ligand domain includes a surface glycoprotein or fragments thereof. In embodiments, the ligand domain has a protein sequence corresponding to amino acid position 98-106 of Human influenza hemagglutinin (HA) protein. In embodiments, the ligand domain includes the sequence of SEQ ID NO: 17 or SEQ ID NO: 18.

[0097] In embodiments, the ligand domain includes a protecting moiety attached to a reactive moiety (e.g., a carboxyl functional group) of the ligand domain. As used herein, a protecting moiety is a chemical moiety covalently attached to a ligand domain that prevents the ligand domain from binding a ligand binder, wherein the protecting moiety may be removed, for example, by chemical means when desired. In embodiments, the protecting moiety is fluorenylmethyloxycarbonyl. In embodiments, the protecting moiety is tert-butyl or carboxybenzyl. Where the ligand domain includes a protecting moiety the ligand domain may be a side chain protected polyamide. Where the ligand domain includes a protecting moiety it may also be referred to herein as "synthetic intermediate" or "synthetic precursor." In embodiments, the protecting moiety includes an amino acid side chain. In embodiments, the protecting moiety includes an amino terminus (e.g. a terminal $-NH_2$ group) or a carboxy terminus (e.g. a terminal $-COOH$ group). In embodiments, the protecting moiety is attached to an amino acid side chain. In embodiments, the protecting moiety is attached to an amino terminus (e.g. a terminal $-NH_2$ group) or a carboxy terminus (e.g. a terminal $-COOH$ group). In the presence of the protecting moiety the ligand domain is not capable of binding a ligand binder. Thus, in embodiments the ligand domain includes a protecting moiety and is not bound to a ligand binder. Upon removal of the protecting moiety and reacting the reactive moiety with a ligand domain capable of binding a ligand binder is formed.

[0098] The ligand domain provided herein may be formed using any multi-step support-bound synthesis compatible with the composition of the microparticle and with the synthesis chemistry of the nucleic acid domain provided herein. The ligand domain and the nucleic acid domain may be simultaneously synthesized on the microparticle. Alternatively, the nucleic acid domain is synthesized on the microparticle after the synthesis of the ligand domain. In embodiments, the attachment of the ligand domain through the first linker is performed prior to the attachment of the nucleic domain through the second linker. In embodiments, the attachment of the ligand domain through the first linker is performed simultaneously with the attachment of the nucleic domain through the second linker. In embodiments, the ligand domain is a peptide. In embodiments, the ligand domain is a small molecule. In embodiments, the ligand domain is a protein. In embodiments, the ligand domain binds to a ligand binder. The ligand domain may be attached to a detectable moiety (e.g., a fluorescent moiety, luminescent moiety, colorimetric moiety, phosphorescent moiety, radioactive moiety or electroactive moiety).

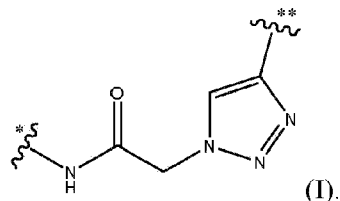
[0099] A "ligand binder" as used herein refers to an agent (e.g., atom, molecule, ion, molecular

ion, compound or particle) capable of binding a ligand domain provided herein including embodiments thereof. Ligand binders include without limitation, biomolecules (e.g., hormones, cytokines, proteins, nucleic acids, lipids, carbohydrates, cellular membrane antigens and receptors (neural, hormonal, nutrient, and cell surface receptors or their ligands); whole cells or lysates thereof (e.g., prokaryotic (e.g., pathogenic bacteria), eukaryotic cells (e.g., mammalian tumor cells); viruses (e.g., retroviruses, herpesviruses, adenoviruses, lentiviruses); and spores); chemicals (e.g., solvents, polymers, organic materials); therapeutic molecules (e.g., therapeutic drugs, abused drugs, antibiotics); or environmental pollutants (e.g., pesticides, insecticides, toxins). The ligand binder may be a protein, a mixture of proteins, a nucleic acid, a mixture of nucleic acids, a small molecule, a mixture of small molecules, an element, a mixture of elements, a synthetic polymer, a mixture of synthetic polymers, cell lysate. In embodiments, the ligand binder is a biomolecule. In embodiments, the biomolecule is a nucleic acid. In embodiments, the biomolecule is a protein (e.g. antibody). In embodiments, the ligand binder is an antibody. In embodiments, the ligand binder is an anti-HA antibody. In embodiments, the ligand binder is an anti-Myc antibody. In embodiments, the ligand domain binds a peptide of SEQ ID NO:17. In embodiments, the ligand domain binds a peptide of SEQ ID NO:18. In embodiments, the ligand binder is attached to a detectable moiety. In embodiments, the detectable moiety is a fluorescent moiety. In embodiments, the ligand binder is a small molecule. In embodiments, the ligand domain is not bound to a ligand binder. Where the ligand domain is not bound to a ligand binder, the ligand domain may include a protecting moiety or any other applicable modification rendering the ligand domain inert. The ligand binder may be attached to a detectable moiety (e.g., a fluorescent moiety, luminescent moiety, colorimetric moiety, phosphorescent moiety, radioactive moiety or electroactive moiety).

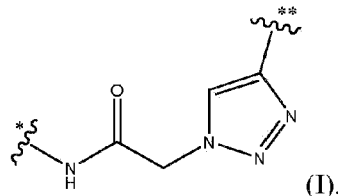
[0100] As described above a "nucleic acid domain" as provided herein includes a nucleic acid sequence corresponding to the individual building blocks of the ligand domain and the order in which these building blocks are incorporated in said ligand domain. Therefore, each microparticle includes a unique ligand domain and a corresponding nucleic acid domain encoding specific nucleic acid sequences corresponding to the building blocks of the ligand domain and the order in which they were incorporated in the ligand domain. The nucleic acid domains provided herein are also referred to as "tag" or "encoding tag." In embodiments, the nucleic acid domain includes a nucleic acid sequence. In embodiments, the nucleic acid sequence is about 18 base pairs in length. In embodiments, the nucleic acid sequence is about 20 base pairs in length. In embodiments, the nucleic acid sequence is 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or 80 base pairs in length. In embodiments, the nucleic acid sequence does not include cytosine.

[0101] In embodiments, the nucleic acid sequence includes a covalent linker. In embodiments, the covalent linker connects two nucleic acid sequences within a nucleic acid domain. In embodiments, the nucleic acid domain includes at least two nucleic acid sequences connected through a covalent linker. In embodiments, the nucleic acid domain includes at least four nucleic acid sequences connected through covalent linkers. Thus, in embodiments, the nucleic

acid domain includes a first nucleic acid sequence, a second nucleic acid sequence, a third nucleic acid sequence and a forth nucleic acid sequence, wherein the first nucleic acid sequence is connected to the second nucleic acid sequence through a first covalent linker, the second nucleic acid sequence is connected to the third nucleic acid sequence through a second covalent linker and the third nucleic acid sequence is connected to the forth nucleic acid sequence through a third covalent linker. In embodiments, the covalent linker (e.g., first, second, third covalent linker) is a bond, $-S(O)-$, $-S(O)_2NH-$, $-NHS(O)_2-$, $-C(O)O-$, $-OC(O)-$, $-C(O)-$, $-C(O)NH-$, $-NH-$, $-NHC(O)-$, $-O-$, $-S-$, substituted or unsubstituted alkylene, substituted or unsubstituted heteroalkylene, substituted or unsubstituted cycloalkylene, substituted or unsubstituted heterocycloalkylene, substituted or unsubstituted arylene, or substituted or unsubstituted heteroarylene. In embodiments, the covalent linker is a 1,3 triazolene linker. In embodiments, the covalent linker has the structure:



In embodiments, the covalent linker includes the structure:



In formula (I), the point of attachment marked by * indicates the attachment of the covalent linker to a first nucleic acid sequence and the attachment marked by ** indicates attachment point of the covalent linker to a second nucleic acid sequence.

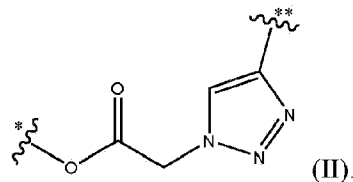
[0102] In embodiments, the nucleic acid domain includes a functionalized nucleic acid sequence. A functionalized nucleic acid as provided herein includes reactive functional groups used for conjugate chemistries as described herein. In embodiments, the nucleic acid domain includes a plurality of functionalized nucleic acid sequences. Where the nucleic acid domain includes a plurality of functionalized nucleic acid sequences, the functionalized nucleic acid sequences are connected through a plurality of covalent linkers. In embodiments, each of the plurality of covalent linkers is chemically different. Schematic illustrations of the synthesis of the nucleic domain and ligand domain on a microparticle are depicted in Figures 4, 5, and 6. As depicted in Figures 5 and 6 the nucleic acid domains attached to a microparticle may be independently different depending on the synthesis used and may include two or more nucleic acid sequences connected through a covalent linker.

[0103] The nucleic acid domains provided herein including embodiments thereof are compatible with (i) the multi-step support-bound synthesis methods applied to form a ligand domain as provided herein, (ii) the composition of the microparticle and (iii) the decoding procedures provided herein (i.e., identifying the composition of the ligand domain and its location on an array). Useful decoding procedures include without limitation sequencing by

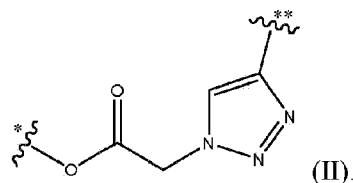
hybridization or enzymatic-based sequencing procedures (e.g., sequencing by synthesis, sequencing by ligation). Thus, in embodiments, the nucleic acid sequence is bound to a complementary nucleic acid sequence. In embodiments, the complementary nucleic acid sequence includes a detectable moiety. In embodiments, the detectable moiety is a fluorescent moiety. Upon hybridization of a complementary nucleic acid to said nucleic acid domain, the composition of the ligand domain and its location on the array can be determined. After the identity of the ligand domain and its location on the array have been determined the nucleic acid domain may be removed (e.g., through cleavage of the second linker), the ligand domain may be further modified (e.g., through reacting a reactive moiety of the ligand domain) and contacted with a ligand binder (e.g., biomolecule).

[0104] The linkers provided herein chemically link the microparticle and the ligand domain (first linker), the microparticle and the nucleic acid domain (second linker) or the microparticle and the solid support (third linker). As described above the nucleic acid domain provided herein including embodiments thereof may include two or more nucleic acid sequences connected through covalent linkers (e.g., a 1,3 triazoline linker). Thus, in embodiments, the nucleic acid domain includes two or more 1,3 triazoline linkers. The linkers provided herein (e.g., first linker, second linker, third linker) may be covalently attached to the microparticle applying methods well known in the art and compatible with the composition of the linker and the microparticle. The linkers provided herein may include the conjugated product of reactive moieties at the point of attachment to the microparticle, at the point of attachment to the ligand domain, at the point of attachment to the nucleic acid domain, or at the point of attachment to the solid support. Thus, the linkers provided herein may be polyvalent and may be formed by conjugate chemistry techniques.

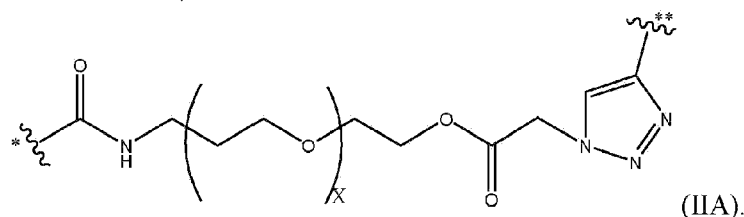
[0105] The first linker comprises $-C(O)NH-$ or $-NHC(O)-$. The second linker comprises $-C(O)O-$ or $-OC(O)-$. In embodiments, the first linker includes the structure $-N(H)-C(O)-$. Where the first linker has the structure $-N(H)-C(O)-$, the nitrogen is attached to the functionalized solid support (e.g., functionalize with bis-amino PEG 3000) and the carbon is attached to the ligand domain. As described above, after the identity of the ligand domain and its location on the array have been determined, the nucleic acid domain may be removed through cleavage of the second linker. In embodiments, the second linker is a photocleavable linker. In embodiments, the second linker is an acid labile linker. In embodiments, the second linker is an alkali labile linker. In embodiments, the second linker has the structure:



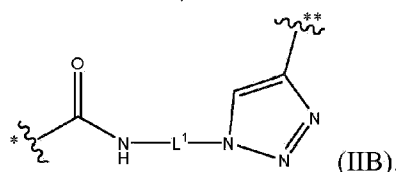
In embodiments, the second linker includes the structure:



In formula (II), the point of attachment marked by * indicates the point of attachment to the functionalized solid support (e.g., functionalized with hydroxyl-amin PEG 3000) and the point of attachment marked by ** indicates the point of attachment to the nucleic acid domain. In embodiments, the second linker has the structure:



In formula (IIA), X is an integer from 20-300. In embodiments, x is 68. In formula (IIA), the point of attachment marked by * indicates the point of attachment to the solid support and the point of attachment marked by ** indicates the point of attachment to the nucleic acid domain. In embodiments, the second linker has the structure



In formula (IIB), L^1 is $-C(O)O-$ or $-OC(O)-$. In formula (IIB), the point of attachment marked by * indicates the point of attachment to the solid support and the point of attachment marked by ** indicates the point of attachment to the nucleic acid domain.

[0106] According to the embodiments provided herein the microparticles provided herein may include a plurality of ligand domains and a plurality of nucleic acid domains attached through a plurality of first linkers and a plurality of second linkers, respectively (see FIG. 3, 4, or 5). Thus, in embodiments, the ligand domain is a plurality of ligand domains attached through a plurality of first linkers. In embodiments, the nucleic acid domain is a plurality of nucleic acid domains attached through a plurality of second linkers. In embodiments, the plurality of nucleic acid domains attached to a single microparticle may be the same or independently different (see FIG. 5 or 6).

[0107] The microparticles provided herein including embodiments thereof are attached to a solid support. In embodiments, the solid support is a planar support. In embodiments, the microparticle is connected through a third linker to the solid support. In embodiments, the microparticle is non-covalently attached to the solid support (e.g. through electrostatic interactions (e.g. ionic bond, hydrogen bond, halogen bond), van der Waals interactions (e.g. dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like). In embodiments, the microparticle is mechanically attached to the solid support. Where a microparticle is mechanically attached to the solid support it is physically held in place on the support through mechanical means (e.g., a well). In embodiments, a plurality of microparticles are covalently attached to the solid support. In embodiments, the microparticle is attached to the solid support through an amide linker. Thus, in embodiments, the third linker has the structure $-N(H)-C(O)-$. In embodiments, the solid support includes carboxymethyl dextran. In embodiments, the solid support includes carboxymethyl dextran functionalized glass. In embodiments, the solid support is a silicon

wafer.

[0108] In an example (not of the present invention), the plurality of microparticles form a disordered array. A "disordered array" as referred to herein is an array of microparticles, wherein the microparticles are randomly assembled on or attached to a solid support and do not form an ordered two- or three-dimensional structure. In the method of the present invention, the plurality of microparticles form an ordered array. In an ordered array, the microparticles are assembled on or attached to a solid support according to an two- or three-dimensional order. For example, a hexagonal array consists of a plurality of microparticles assembled on or attached to a solid support such that each microparticle forms part of a hexagon, wherein each microparticle occupies one angle of the hexagon, and wherein the center of the hexagon is occupied by a seventh microparticle. In embodiments, the plurality of microparticles form an hexagonal array. In embodiments, the plurality of microparticles form a square packed array. A square packed array consists of a plurality of microparticles assembled on or attached to a solid support such that each microparticle forms part of a square or rectangle consisting of at least four microparticles. The formation of arrays is a method well known and used in the art and is described, inter alia, in US Patents 6,110,426; 7,615,368; 7,932,213; 6,824,987; 5,143,854; 8,795,967 and Hughes TR et al. (2001) Nat. Biotech. 4, 342-347. In embodiments, at least about 10^6 of the microparticles are attached to the solid support. In embodiments, each of the microparticles is different. In embodiments, about 10^6 to 10^9 of the microparticles are attached to the solid support. In embodiments, about 10^9 of the microparticles are attached to the solid support. In embodiments, about 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , 10^{10} or 10^{11} of the microparticles are attached to the solid support. In embodiments, the array includes 10^6 microparticles per square millimeter.

[0109] In embodiments, the array includes at least about 10,000 microparticles per square millimeter. In embodiments, the array includes at least about 20,000 microparticles per square millimeter. In embodiments, the array includes at least about 30,000 microparticles per square millimeter. In embodiments, the array includes at least about 40,000 microparticles per square millimeter. In embodiments, the array includes at least about 50,000 microparticles per square millimeter. In embodiments, the array includes at least about 60,000 microparticles per square millimeter. In embodiments, the array includes at least about 70,000 microparticles per square millimeter. In embodiments, the array includes at least about 80,000 microparticles per square millimeter. In embodiments, the array includes at least about 90,000 microparticles per square millimeter. In embodiments, the array includes at least about 100,000 microparticles per square millimeter. In embodiments, the array includes at least about 200,000 microparticles per square millimeter. In embodiments, the array includes at least about 300,000 microparticles per square millimeter. In embodiments, the array includes at least about 400,000 microparticles per square millimeter. In embodiments, the array includes at least about 500,000 microparticles per square millimeter. In embodiments, the array includes at least about 600,000 microparticles per square millimeter. In embodiments, the array includes at least about 700,000 microparticles per square millimeter. In embodiments, the array includes at least about 800,000 microparticles per square millimeter. In embodiments, the

array includes at least about 900,000 microparticles per square millimeter.

[0110] In embodiments, the array includes about 200,000 microparticles per square millimeter. In embodiments, the array includes about 789,000 microparticles per square millimeter. In embodiments, the array includes about 591,715 microparticles per square millimeter. In embodiments, the array includes from about 200,000 to about 800,000 microparticles per square millimeter.

[0111] In embodiments, the array includes about one microparticle per 4.99 square microns. In embodiments, the array includes about one microparticle per 1.46 square microns. In embodiments, the array includes about one microparticle per 1.69 square microns. In embodiments, the array includes about one microparticle per square micron.

[0112] In embodiments, the solid support includes a plurality of wells each of the wells capturing one of the microparticles. In embodiments, the nucleic acid domain is a nucleic acid sequence as described herein. In embodiments, the nucleic acid sequence is bound to a complementary nucleic acid sequence. In embodiments, the complementary nucleic acid sequence includes a detectable moiety. In embodiments, the detectable moiety is a fluorescent moiety.

METHODS

[0113] The present invention provides a method of forming a cleaved ordered array of microparticles. The method includes (i) providing a solid support comprising a plurality of wells. (ii) Attaching a plurality of microparticles as provided herein including embodiments thereof to said plurality of wells, thereby forming an ordered array of immobilized microparticles. (iii) Performing a decoding procedure on the encoding tag, thereby identifying the composition of the ligand domain and its location on said solid support. (iv) The second linkers of the plurality of immobilized microparticles are cleaved, thereby forming a cleaved ordered array. The immobilized microparticle in step (i) is further covalently attached to (a) a ligand domain through a first linker; and (b) an encoding tag through a second linker. The ligand domain is a peptide, a protein, a macrocyclic peptide, a peptidomimetic, a polyamide, or a macrolactam. The encoding tag is a nucleic acid sequence identifying said ligand domain. The second linker is cleavable and said first linker is not cleavable under a condition that said second linker is cleavable. In embodiments, step (ii) includes binding a complementary nucleic acid sequence to the nucleic acid domain. In embodiments, the cleaving includes contacting the immobilized microparticle with a cleaving agent. In embodiments, the cleaving agent is an acid. In embodiments, the cleaving agent is trifluoroacetic acid. In embodiments, the cleaving agent is an alkali agent. In embodiments, the cleaving agent is ammonium hydroxide. In embodiments, the cleaving agent is ammonia. In embodiments, the cleaving agent is methylamine. In embodiments, the cleaving agent is a mixture of ammonium hydroxide and methylamine. In embodiments, the cleaving is performed at room temperature. In embodiments, the cleaving

agent is UV irradiation. In embodiments, the cleaving agent is light irradiation. In embodiments, the cleaving does not include cleaving the first linker.

[0114] In embodiments, the method includes after the cleaving of step (iii), a step (iv) of reacting a reactive moiety of the ligand domain, thereby forming a reactive ligand domain and (v) binding a ligand binder to the reactive ligand domain. In embodiments, the method includes after the cleaving of step (iii), a step (iv) of binding a ligand binder to the ligand domain. In embodiments, the step (iii) of cleaving includes binding a ligand binder to the ligand domain. Thus, the cleaving of the second linker may occur simultaneously with the binding of a ligand binder to the ligand domain. Alternatively, the binding of a ligand binder to the ligand domain may occur after the cleaving of the second linker. In embodiments, the binding of a ligand binder to the ligand domain includes reacting a reactive moiety of the ligand domain.

EXAMPLES

[0115] Using split pool library synthesis Applicants were able to increase the number of compounds displayed by at least 1,000x over current methods. The compositions provided herein are a highly diverse collection of molecules immobilized in an extremely dense array on solid support. To prepare Applicants' system, the assembled precursor library is immobilized into a planar array. The entire immobilized array is decoded, as each library member now occupies a permanent and discrete space in a planar array, the decoding converts what was a chemically encoded library into a spatially addressed library. The chemical encoding units are removed and subsequent synthetic transformations are performed across the immobilized library, completing the library synthesis. The library can then be screened to identify molecules demonstrating useful function.

[0116] When encoded split pool synthesis is combined with established methods of high density bead immobilization and oligonucleotide sequencing, the system allows for screens of fully decoded split pool libraries on a scale that was not previously possible. By analogy to next generation sequencing technology, fully decoded screening sets of up to 108 to 1010 in a microarray format should be achievable. Provided herein are "tagless" libraries for further chemistry. By decoding prior to screening, oligonucleotide tags can be removed which accomplishes two things: (i) additional chemistry that would be incompatible with the oligonucleotide tag can be done on the immobilized library (chemical incompatibility is one of the stated challenges of encoded combinatorial chemistry); (ii) removal of the tags eliminates the potential of the tags interfering with the assays of interest. The encoding chemistry enables Applicants to harness the power of split-pool synthesis for chemical library generation, yet allows for libraries to be evaluated in relatively information rich screens, rather than selections. The relative performance of all library members in a given assay are evaluated, not just selected 'hits'. Compounds that prove problematic in one or more assays can be easily identified and flagged, reducing the number of potential false positives in subsequent screens. This facilitates the development of structure activity relationships.

[0117] The final bioactive compounds that are screened for activity are formed after decoding and removal of the tags. What is immobilized and decoded are protected, synthetic intermediates. For the invention provided herein, at least one additional chemical step is performed after decoding and removal of the nucleic acid domain to finish preparing the "bioactive agents" (this could include: deprotection, macrocyclization etc).

[0118] The potential benefits of encoding a split-pool library synthesis using nucleic acids are well known as are the synthetic challenges. Applicants' specific encoding strategy allows for the creation of screenable DNA encoded libraries in fewer linear chemical transformations per step than previously demonstrated, with each encoding step being independently decodable.

[0119] The microparticle also referred to herein as "core" could be a dendrimer, a hyperbranched polymer, a functionalized silica particle, a functionalized polymer particle. The core could be magnetic. The core could range in size from 20 nm to 200 microns in diameter. The core is functionalized with a reactive moiety that allows attachment of building blocks to make the library molecules, a different reactive moiety that allows attachment of DNA tags for encoding, and, in some cases, a third different reactive moiety which aids in covalent immobilization of the core to a surface.

[0120] The multidentate core can be a 0.9 micron magnetic polymeric bead functionalized with a surface of polyethylene glycol terminated with orthogonally protected amines .

[0121] The synthetic precursor (ligand domain including a protecting moiety) could be any product of multi-step support-bound synthesis, with the restriction that any synthetic transformations used to create the synthetic precursor, once attached to the core particle must be compatible with the core chemical structure and are either orthogonal/compatible with the described encoding chemistry or performed prior to incorporation of the first encoding tags. The synthetic precursors may be side chain protected polyamides.

[0122] The Encoding tags (nucleic acid domains) consist of unique, pre-synthesized, functionalized nucleic acid derivatives covalently linked directly to the core particle through a cleavable linker or indirectly through other encoding tags. Tag sequences are of sufficient length and composition and encode entire library. To enable decoding of specific encoded synthetic steps, independent of other encoded synthetic steps. For example: decoding the tags used to encode for the second step in a synthesis should be completely independent of the ability to decode tags used to encode the first or third steps in a synthesis. Unsuccessful incorporation of tags at step one adversely impact incorporation of tags at step two. An inability to "decode" step one will affect the ability to "decode" step two for a given encoded library member. Applicants' approach may require longer nucleic acid sequences but provides a more robust encoding/decoding process. Tags are compatible with the chosen method of decoding or sequencing. For example; if sequencing of tags is to be performed by a process of sequential hybridizations, tags should be relatively isothermal to one another and contain sufficient sequence differences such that undesired cross hybridization is minimized. If sequencing of tags is to be done through any of the enzymatic based sequencing by synthesis

or sequencing by ligation approaches, tags may require common primer binding sites

[0123] Tag chemical structures: (i) stable and unreactive to the synthetic transformations used to construct the precursor library; (ii) removable from the core particle after it is sequenced; (iii) compatible with method of decoding/sequencing. For example, sequencing by hybridization could be done with DNA, PNA, LNA, RNA, modified RNA, modified DNA analogues or some combinations thereof. For any enzymatic based sequencing approaches, DNA is preferred. The protected oligonucleotide structures could be used to improve chemical orthogonality, wherein protecting moieties on the tag are removed prior to decoding. (for example, the exocyclic amines of the nucleobases may be protected prior to or immediately following tag incorporation)

[0124] Tag cleavage site structures: (i) tags may be directly attached to the core particle or indirectly to the core through other tags, all core particle attachment points for tags ultimately contain a site of cleavability. (ii) preferred cleavable structures, following cleavage, leave the core particle surface free of reactive moieties which may interfere with subsequent chemistry or assays (e.g., ester cleavage site which leave a free alcohol; trityl ether cleavage site which leave a free alcohol).

[0125] Tag linking structures: the encoding tags may be 20-mer DNA based oligonucleotides with zero C content and 25% G content. The tags differ from one another by at least a six base pair mismatch. The tags are linked to the support and or one another through 1,3 triazole linkages all ultimately connected to the core particle through a cleavable ester bond. Any method of attachment and any surface used must be compatible with the decoding conditions and microarray assay conditions, as well as tag removal conditions, and any synthetic reaction conditions applied to the array. Applicants have demonstrated immobilization on patterned quartz, patterned silicon, and carboxymethyl-dextran functionalized glass.

[0126] DNA tag removal: Conditions are dependent on structure at tag cleavage site. With ester linkage, Applicants have used ammonium hydroxide and methylamine. Ligand deprotection and further synthetic modifications. Applicants have removed all side chain protecting moieties from polyamide structures. Selective side chain de-protection followed by macrolactamization will be performed. Incorporation of fluorophore-quencher pairs or solvatochromic moieties will be performed. The compositions provided herein may be useful, inter alia, for microarray screens, fluorescence, colorimetric, or electrochemical readouts binding assays in which fluorescently labeled proteins of interest are incubated over the array to identify binders has been demonstrated.

[0127] In embodiments, the prototype chips are at 2.4 micron C-C (center to center) spacing in a hexagonal array. That results in 4.99 square microns per bead (less than the 5.76 square microns per bead expected because the beads are hexagonally packed), or 20 million beads per square centimeter, or 200,000 beads per square millimeter, or ~376 million particles within the area equivalent of a standard 25mm by 75mm microscope slide.

[0128] In embodiments, particles have been immobilized into a 1.3 micron C-C spacing hexagonal array. That results in 1.46 square microns per bead, or 78.9 million beads per square centimeter, or 789,000 beads per square millimeter, or 1.28 billion beads within the area equivalent to a 25mm by 75mm microscope slide.

[0129] In embodiments, particles have been immobilized into a 1.3 micron C-C spacing square packed array. That results in 1.69 square microns per bead, or 59 million beads per square centimeter, or 591,715 beads per square millimeter, or 1.11 billion beads within the area equivalent to a 25mm by 75mm microscope slide.

[0130] In embodiments, an array is hexagonally packed with a density of 1 bead per square micron (C-C spacing of ~1.075 microns).

MATERIALS AND METHODS

Reagents

[0131] Triethylamine (TEA), Diisopropylethylamine (DIPEA), Diisopropylcarbodiimide (DIC), dimethylaminopyridine (DMAP), dimethylformamide (DMF), dimethylsulfoxide (DMSO), Triton X-100 (TX100), azidoacetic acid (Aza), 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), (7-Azabenzotriazol-1-yl)oxy)tripyrrolidinophosphonium hexafluorophosphate (PYAOP), Tris(3-hydroxypropyl)triazolylmethylamine (THPTA), Copper(I) bromide dimethyl sulfide complex (CuBrDMS), Boc-Glycine, and Fmoc-Glycine were purchased from Sigma-Aldrich and used as received. All other Fmoc protected amino acids were purchased from Novabiochem or Advanced Chemtech. Water used in all wash and reaction buffers was obtained from a Millipore MilliQ purification system. Peg reagents used for the initial microparticle functionalization were purchased from Rapp Polymere or Quanta Biodesign. Promag Beads are provided by Bangs Laboratories as a 2.6% w/v solution in water and have the following characteristics: 880 nm average diameter. Composed of iron oxide embedded within a highly crosslinked polymer matrix. Surface of the particles displays free carboxylic acid reactive moieties at 440 nmole equivalents per milligram. Approximately 2 billion microparticles are contained within one milligram of the stock particles. All nucleic acid tags and fluorescently labeled complement sequences were purchased from IDT.

Table 1.

Oligo sequence name	IDT Sequence code	SEQ ID NO:
Tag A ₀	/5Hexynyl/AAC CAC ACA CAC ACA ACC /3AmMO/	1
Tag A ₁	/5Hexynyl/AAC CAC ACA CAC CAA ACC /3AmMO/	2

Oligo sequence name	IDT Sequence code	SEQ ID NO:
Tag B ₂	/5Hexynyl/ACG AAC ACA CAC GTA CGA /3AmMO/	3
Tag C ₂	/5Hexynyl/CAA GAC ACA CAC GTC AAG /3AmMO/	4
Tag D ₂	/5Hexynyl/CCT TAC ACA CAC GTC CTT /3AmMO/	5
Tag B ₃	/5Hexynyl/ACG AAC ACA CAC TGA CGA /3AmMO/	6
Tag C ₃	/5Hexynyl/CAA GAC ACA CAC TGC AAG /3AmMO/	7
Tag D ₃	/5Hexynyl/CCT TAC ACA CAC TGC CTT /3AmMO/	8
Anti-B ₂ -Cy3	/5Cy3/TCG TAC GTG TGT GTT CGT	9
Anti-C ₂ -Cy3	/5Cy3/CTT GAC GTG TGT GTC TTG	10
Anti-D ₂ -Cy3	/5Cy3/AAG GAC GTG TGT GTA AGG	11
Anti-B ₃ -AF647	/5Alex647N/TCG TCA GTG TGT GTT CGT	12
Anti-C ₃ -FITC	/56-FAM/CTT GCA GTG TGT GTC TTG	13
Anti-D ₃ -FITC	/56-FAM/AAG GCA GTG TGT GTA AGG	14
Anti-A ₀ -Cy3	/5Cy3/GGT TGT GTG TGT GTG GTT	15
Anti-A ₁ -FITC	/56-FAM/GGT TTG GTG TGT GTG GTT	16

Ligand domain sequences

[0132]

SEQ ID NO:17: YPYDVPDYA. (HA-tag)

SEQ ID NO:18 EQKLISEEDL (Myc-tag)

Additional Reagent Abbreviations

[0133]

DITx	MilliQ water containing 1% Triton X-100
------	---

DMSOTx	DMSO containing 1% Triton X-100
PBT	100 mM phosphate buffer, pH 7.0 containing 1% Triton X-100
AMA	1: 1 mix of ammonium hydroxide and 30% methylamine in ethanol

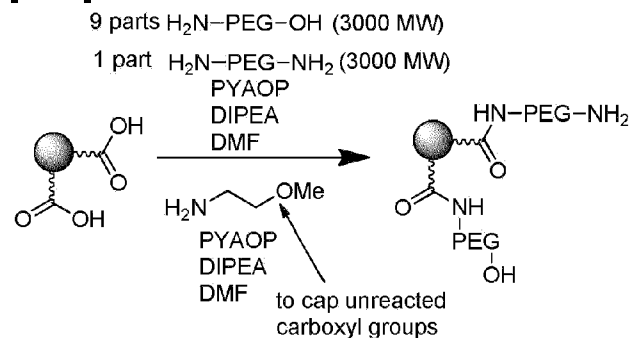
General microparticle handling procedures

[0134] Similar to most solid phase synthetic procedures, a typical reaction involves 1) dispersing the microparticles in a reaction solution 2) adding additional solvents or reagents as needed for the reaction 3) providing occasional or constant agitation at some specified reaction temperature for some period of time and 4) following the reaction, the reagents and soluble byproducts of the reaction are separated from the microparticles through a series of washes. The separation and washing steps for Applicants' microparticles consists of multiple rounds of 1) magnetically assisted pelleting of the microparticles and aspiration of the supernatant followed by 2) resuspension of the microparticle pellet in a suitable wash solution.

ProMag microparticle functionalization

Initial PEGylation:

[0135]

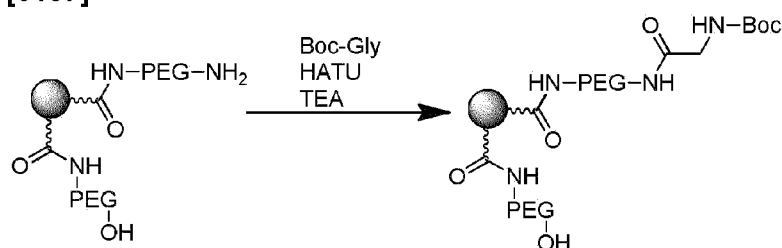


[0136] Large-scale functionalization to create pegylated beads (with a 4.5:1 ratio of hydroxyl groups to amine groups) are done in order to eliminate any batch-to-batch variations. 250uL of stock ProMag beads are washed with DMFTx (1mL, x3) and suspended in 150 uL DMFTx. Amino hydroxy PEG 3000 (270mg) and bis amine PEG 3000 (30mg) are weighed into a 1.5 mL conical tube and melted in an oil bath at 65C. The beads are added to the melted PEG and mixed thoroughly/heated. Fifty five microliters of DIPEA is added to the reaction mix and vortexed followed by the addition of solid PyAOP (160 mg). The reaction ran at 65C for 45 minutes while heating with an oil bath. The final concentrations in the reaction end up being

(200 mM PEG, 600mM PyAOP, 600 mM DIPEA). After 45 minutes the remaining carboxylic acids on the surface are capped with 2-methoxyethyl amine (250uL), let incubate at the same temperature for 10 minutes then wash the beads with DMFTx (1mL, x3), then resuspend with 250 uL DMFTx. A 50 uL aliquot taken out and washed with MilliQ water (400 uL, x3) and loaded on the burned off clean/ tared pan, put in the oven at 95C and the following method was ran. Jump to 95C, isotherm for 10 min, ramp 20C/min until 500C, isotherm for 30 minutes. [Starting mass: 1.5940mg. At 100C 98.86%. At 500C 31.30%] This corresponds to 25.4% added functionality, comparison to the last large batch which was 24.3% added functionalization (1.704 mg, 99.81% at 100C and 32.08% at 500C) as compared to the first two Large batches prepared by the same protocol which had an added funct of 25.6%.

BOC Protection (x2)

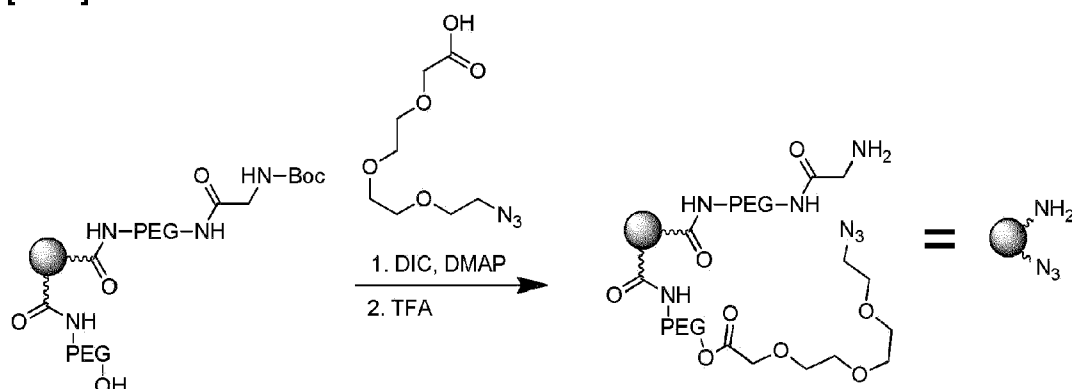
[0137]



[0138] The rest of the beads are Boc protected by washing into DMSOTx (1 mL, x3), suspending in 170uL DMFTx, adding 14 mg Boc-Gly OH, followed by 22.3 uL TEA, and 30.4 mg HATU, in that order. The reaction ran at 65C for 30 min in an oil bath, after which the beads were washed with DMFTx (500mL, x3) and subject to the same conditions one more time. The beads were then subject to AMA (200 uL, 65C, 5 min) then washed into DMFTx at stock concentration.

DIC/DMAP with Long Chain Azide (x3)

[0139]



[0140] Beads are resuspended in 400 uL DMFTx, to this is added 80 uL of 500 mM long chain azido acid (11.66mg/100uL), followed by 6.7 uL DIC, then 10 uL of DMAP (80mg/ml stock in DMF). These coupling conditions were repeated a total of three times--washing with DMFTx (200 uL, x3) in between. The Boc group is then removed by treatment with TFA for 15 minutes. This large batch is then washed into PBT (400uL, x3) and suspended in 400 uL (x4 dilution from stock). Protocol for attaching azido acetic acid in a alkali cleavable form (connected to the microparticle through an ester bond) is performed in a similar fashion.

Standard HATU Amino Acid Coupling Conditions

[0141] 400mM HATU, 400mM FMOC'ed amino acid, 800mM TEA in 10% DITx, 90% DMSOTx, 65C, 30 minutes.

[0142] Beads are washed into DITx (100uL, x3), and suspended in 10uL DITx. To this is added the FMOC'ed amino acid in 80uL DMSOTx (400mM for a 200uL rxn volume), followed by TEA (10uL, 101.19g/mole, 0.726g/mL), and lastly HATU (15mg, 380.23g/mole) is added last as a solid and the tube is set on the red heat block set at 65C for 30 minutes. Double couplings are done (without washing in between) with all amino acids to ensure 100% conversion.

Standard HATU coupling of azide to the amine reactive moiety on the nucleic acid tags

[0143] 400mM HATU, 400mM 2-azidoacetic acid, 800mM TEA in 10% DITx, 90% DMSOTx, 65C, 30 minutes

[0144] Beads are washed into PBT (100uL, x3), and suspended in 10uL DITx. To this is added DMSOTx (80uL), followed by 2-azidoacetic acid (3uL, 101.06g/mole, 1.35g/mL), followed by TEA (22.3uL, 101.19g/mole, 0.726g/mL), and lastly HATU (15mg, 380.23g/mole) is added last as a solid and the tube is set on the red heat block set at 65C for 30 minutes. Double couplings are done (without washing in between) with all acids to ensure 100% conversion.

Copper catalyzed azide alkyne (Huisgen) cycloaddition conditions (Encoding step)

[0145] Catalyst stock solution preparation: Cu(I)Br DMS (205.58g/mole) had previously been weighed out in the glovebox in 20mL scintillation vials. DMSO (that is kept in the glovebox) is added to the vial such that a concentration of 4.5mg/mL (22mM) is achieved. THPTA (tris(3-hydroxypropyltriazolylmethyl)amine) (434.5g/mole) is weighed out outside of the hood and dissolved to a concentration of 2.8mg/125uL (52mM). The Cu (I) and the ligand solutions are combined in a 1:2 v:v ration of Cu(I) to ligand giving the following concentrations: 35mM THPTA and 7.3mM Cu(I)Br DMS in 50% DMSO and 50% MilliQ water.

[0146] The beads are washed into PBT (100uL, x3) and suspended in 13uL PBT. To this is added 10uL oligo tag (from 500uM stock solution). This is transferred into the glovebox with a minimum of 4 pump/purge cycles. Two microliters of the catalyst stock solution is added and the PCR tube is set on the PCR block set to 60C for 30 minutes, after which the reaction is quenched with 500mM EDTA (outside of the glove box) by added 100uL EDTA and incubating for 3 minutes before washing with PBT (200uL, x5).

Standard Fmoc Deprotection Conditions

[0147] 100uL 20% piperidine in DMFTx for 10 minutes at rt (x3).

Competitive Hybridization Conditions, example using for distinguishing between tags A0 and A1

[0148] Anti-A0-Cy3, and Anti-A1-FITC are purchased from IDT and diluted to stocks of 500μM in MilliQ water. A hybridization solution that is to be applied to bead samples is prepared as follows. 40μL of formamide, 20 μL of 20X SSPE buffer (Sigma), 10 μL of DITx, 20 μL of DI, 5 μL of Anti-A0-Cy3 stock , and 5 μL of Anti-A1-FITC stock are mixed in a single PCR tube, and stored in the dark till use.

[0149] Samples of beads (1.6μg) displaying A0 or A1 on their respective surfaces were placed in separate PCR tube, pelleted, and the supernatant was removed by vacuum aspiration. Each bead pellet was immediately suspended in 25 μL of hybridization solution. The bead slurry mixtures were allowed to hybridize over the next 15 minutes at room temperature, away from light. After the indicated time period, the bead samples were pelleted, the hybridization solution was removed by vacuum aspiration. Beads were then washed 3X with 25 μL of PBT, and finally suspended in 25 μL pf PBT. 5 μL of this bead sample is removed, placed in a 1356 well plate and imaged at the microscopy core using a Zeiss Observer, 63X water objective, 1.6 optovar in the brightfeild, DsRed and EGFP channels. An extra 5 μL sample of each bead type was loaded in a separate well of the plate and used to set the exposure times for the DsRed and EGFP by the 2014 Zeiss ZenBlue software. The additional A0 sample was used to set the DsRed and the additional A1 sample was used to set the EGFP. After setting the exposure time the unexposed A0 and A1 samples were imaged using the same, fixed exposure times.

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [WO9512608A \[0001\]](#)
- [WO2011143583A \[0001\]](#)
- [US8039271B \[0001\]](#)
- [WO9820019A \[0001\]](#)
- [US2006040286A \[0001\]](#)
- [WO2005123959A \[0001\]](#)
- [WO2014026283A \[0001\]](#)
- [US2003027221A \[0001\]](#)
- [WO2007084192A \[0001\]](#)
- [EP1239286A \[0001\]](#)
- [WO0071995A \[0001\]](#)
- [WO9967641A \[0001\]](#)
- [WO2010093705A \[0001\]](#)
- [US4672040A \[0001\]](#)
- [US5235033A \[0038\]](#)
- [US5034506A \[0038\]](#)
- [WO9617958A \[0042\]](#)
- [US5143854A \[0042\] \[0108\]](#)
- [US4946778A \[0078\]](#)
- [US6060596A \[0082\]](#)
- [US5770358A \[0082\]](#)
- [US6368874B \[0082\]](#)
- [US5565324A \[0082\]](#)
- [US6936477B \[0082\]](#)
- [US5573905A \[0082\]](#)
- [US6110426A \[0108\]](#)
- [US7615368B \[0108\]](#)
- [US7932213B \[0108\]](#)
- [US6824987B \[0108\]](#)
- [US8795967B \[0108\]](#)

Non-patent literature cited in the description

- **RAMON-AZCON et al.** Analytical Chemistry, 2011, vol. 83, 31053-1060 [0001]
- **CZAMIK, A. W.** Encoding methods for combinatorial chemistry Curr. Opin. Chem. Biol., 1997, vol. 1, 60-66 [0003]
- **SINGLETON et al.** DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY J. Wiley & Sons 1994 0000 [0035]
- **SAMBROOK et al.** MOLECULAR CLONING, A LABORATORY MANUAL Cold Springs Harbor Press 1989 0000 [0035]
- **ECKSTEIN** Oligonucleotides and Analogues: A Practical Approach Oxford University Press [0038]
- Carbohydrate Modifications in Antisense Research ASC Symposium Series 580 [0038]
- **FODOR** Science, 1991, 767-773 [0042]
- **JOHNSTON** Curr. Biol., 1998, vol. 8, R171-R174 [0042]
- **SCHUMMER** Biotechniques, 1997, vol. 23, 1087-1092 [0042]
- **KERN** Biotechniques, 1997, vol. 23, 120-124 [0042]
- **SAMBROOK et al.** Molecular Cloning: A Laboratory Manual 1989 0000 1-88 [0046]
- **MARCH** ADVANCED ORGANIC CHEMISTRY John Wiley & Sons 1985 0000 [0052]
- **HERMANSON** BIOCONJUGATE TECHNIQUES Academic Press 1996 0000 [0052]
- MODIFICATION OF PROTEINS **FEENEY et al.** Advances in Chemistry Series American Chemical Society 1982 0000 vol. 198, [0052]
- **CREIGHTON** Proteins, 1984, [0062]
- **SMITH** WATERMAN Adv. Appl. Math., 1970, vol. 2, 482c- [0066]
- **NEEDLEMAN** WUNSCH J. Mol. Biol., 1970, vol. 48, 443- [0066]
- **PEARSON** LIPMAN Proc. Nat'l. Acad. Sci. USA, 1988, vol. 85, 2444- [0066]
- **AUSUBEL et al.** Current Protocols in Molecular Biology, 1995, [0066]
- **ALTSCHUL et al.** Nuc. Acids Res., 1977, vol. 25, 3389-3402 [0067]
- **ALTSCHUL et al.** J. Mol. Biol., 1990, vol. 215, 403-410 [0067]
- **HENIKOFF** HENIKOFF Proc. Natl. Acad. Sci. USA, 1989, vol. 89, 10915- [0067]
- **KARLIN** ALTSCHUL Proc. Natl. Acad. Sci. USA, 1993, vol. 90, 5873-5877 [0068]
- **HERMANSON** Bioconjugate Techniques Academic Press, Inc. 1996 0000 [0073]
- Fundamental Immunology 1993 0000 [0077]
- **MCCAFFERTY et al.** Nature, 1990, vol. 348, 552-554 [0077] [0078]
- **KOHLER** MILSTEIN Nature, 1975, vol. 256, 495-497 [0078]
- **KOZBOR et al.** Immunology Today, 1983, vol. 4, 72- [0078]
- **COLE et al.** Monoclonal Antibodies and Cancer Therapy, 1985, 77-96 [0078]
- **MARKS et al.** Biotechnology, 1992, vol. 10, 779-783 [0078]
- **FURKA, A. et al.** Int. J. Pept. Protein Res., 1991, vol. 37, 487-493 [0082]
- **KIT LAM et al.** Nature, 1991, vol. 354, 82-84 [0082]
- **HUGHES TR et al.** Nat. Biotech., 2001, vol. 4, 342-347 [0108]

Patentkrav

1. Fremgangsmåde, der omfatter følgende trin:
 - (i) tilvejebringelse af en fast bærer omfattende en flerhed af brønde;
 - 5 (ii) fastgørelse af en flerhed af mikropartikler til flerheden af brønde for at danne en ordnet række af en flerhed af immobiliserede mikropartikler, hvor hver af flerheden af immobiliserede mikropartikler endvidere er kovalent bundet til:
 - (a) et liganddomæne via en første linker; og
 - (b) et kodningstag via en anden linker;
 - 10 hvor liganddomænet er et peptid eller et protein, et makrocyclisk peptid, et peptidomimetikum, et polyamid eller et makrolactam;
hvor kodningstagget er en nukleinsyresekvens, der identificerer liganddomænet;
hvor den anden linker kan spaltes, og den første linker ikke kan spaltes under et forhold, hvor den anden linker kan spaltes;
 - 15 hvor den første linker omfatter -C(O)NH- eller -NHC(O)-, og den anden linker omfatter -C(O)O- eller -OC(O)-;
 - (iii) udførelse af en afkodningsprocedure på kodningstagget, hvorved liganddomænets sammensætning og dens placering på den faste bærer identificeres; og
 - (iv) spaltning af anden linker af flerheden af immobiliserede mikropartikler, hvorved
20 der dannes en spaltet, ordnet række.
2. Fremgangsmåde ifølge krav 1, hvor afkodningsproceduren omfatter sekvensering af kodningstagget ved hybridisering eller enzymbaserede sekvenseringsprocedurer.
- 25 3. Fremgangsmåde ifølge krav 1 eller krav 2, hvor afkodningsproceduren omfatter binding af en komplementær nukleinsyresekvens til kodningstagget, og hvor den komplementære nukleinsyresekvens indbefatter en detekterbar del.
4. Fremgangsmåde ifølge et hvilket som helst af krav 1-3, hvor spaltningen omfatter at
30 bringe hver af flerheden af immobiliserede mikropartikler fra den ordnede række i kontakt med et spaltningsmiddel, hvor spaltningsmidlet er en syre eller et alkalimiddel.
5. Fremgangsmåde ifølge krav 4, hvor syren er trifluoreddikesyre.

6. Fremgangsmåde ifølge krav 4, hvor alkalimidlet vælges fra gruppen bestående af ammoniumhydroxid, ammoniak, methylamin og en blanding af ammoniumhydroxid og methylamin.
- 5 7. Fremgangsmåde ifølge et hvilket som helst af krav 1-6, og som endvidere omfatter binding af ligandbindere til liganddomænerne af flerheden af immobiliserede mikropartikler fra den spaltede, ordnede række, hvorved flerheden af bundne ligandbindere dannes.
- 10 8. Fremgangsmåde ifølge krav 7, og som endvidere omfatter identificering af placeringerne for flerheden af bundne ligandbindere på den fast bærer, hvorved flerheden af ligandbindere detekteres.
9. Fremgangsmåde ifølge et hvilket som helst af krav 4-8, hvor hver af ligandbinderne omfatter en detekterbar del.
- 15 10. Fremgangsmåde ifølge et hvilket som helst af krav 4-9, hvor hver af ligandbinderne er et protein eller en nukleinsyre.
- 20 11. Fremgangsmåde ifølge et hvilket som helst af krav 1-10, hvor den ordnede række er en heksagonal række eller en firkantet, pakket række.
12. Fremgangsmåde ifølge et hvilket som helst af krav 1-11, hvor den ordnede række indbefatter mindst ca. 10.000 mikropartikler pr. kvadratmillimeter.
- 25 13. Fremgangsmåde ifølge et hvilket som helst af krav 1-12, hvor peptidet endvidere omfatter non-peptidfunktionalitet.
- 30 14. Fremgangsmåde ifølge et hvilket som helst af krav 1-13, hvor liganddomænet omfatter en beskyttelsesdel, der er kovalent bundet til en reaktiv del, hvor beskyttelsesdelen forhindrer liganddomænet i at binde en ligandbinder, hvor beskyttelsesdelen kan fjernes; eventuelt hvor den reaktive del er en funktionel carboxylgruppe.

DRAWINGS

FIG. 1

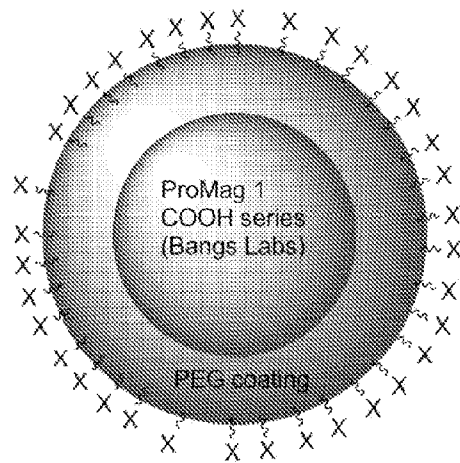
Preferred Ranges:

$X = 0.1$ to 100
nanomoles/mg

$1\% X \leq A \leq 20\% X$

$40\% X \leq B \leq 99\% X$

$0\% X \leq C \leq 50\% X$



A) library molecule attachment point
OR
X = B) encoding tag attachment point
OR
C) core immobilization point

FIG. 2

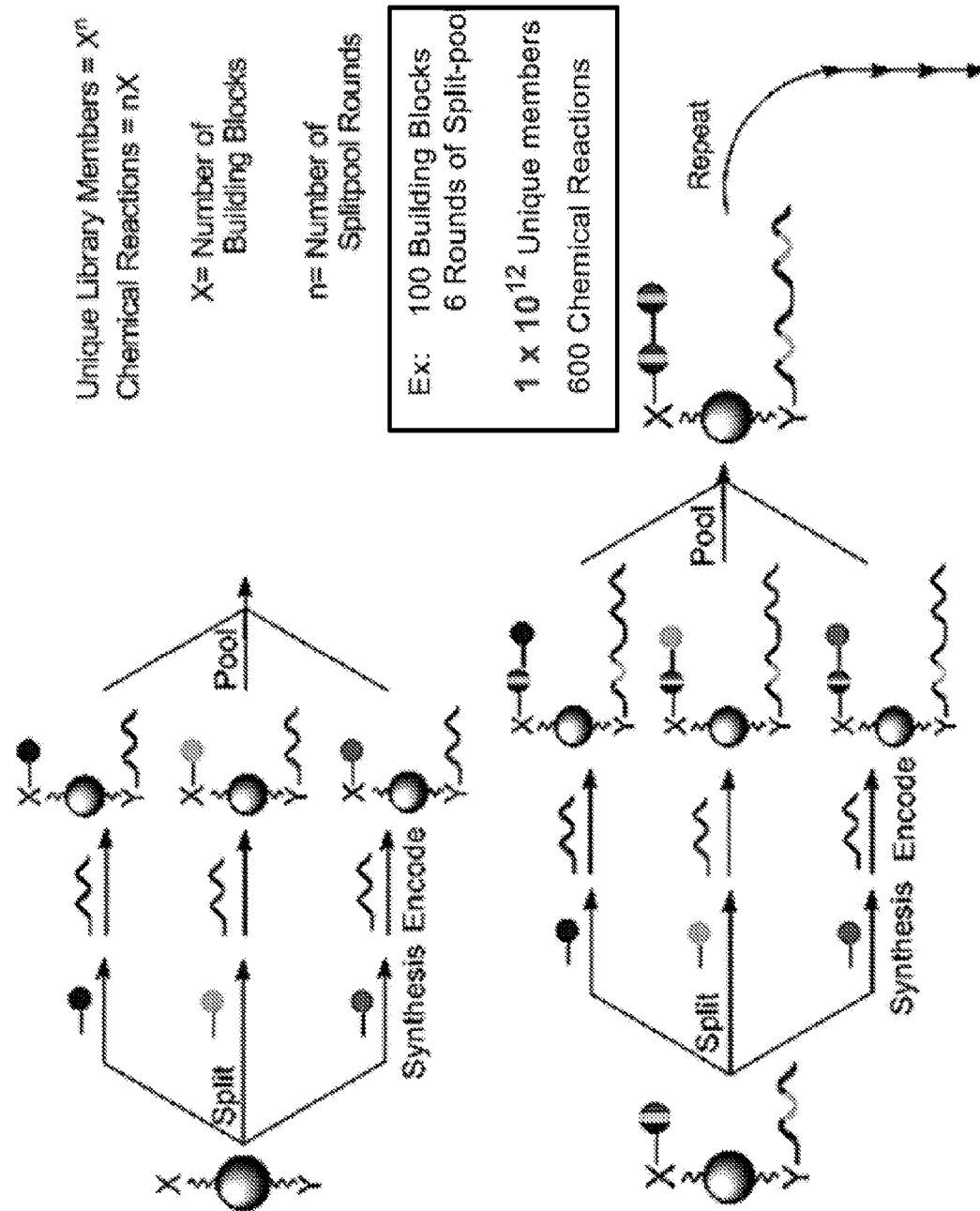


FIG. 3

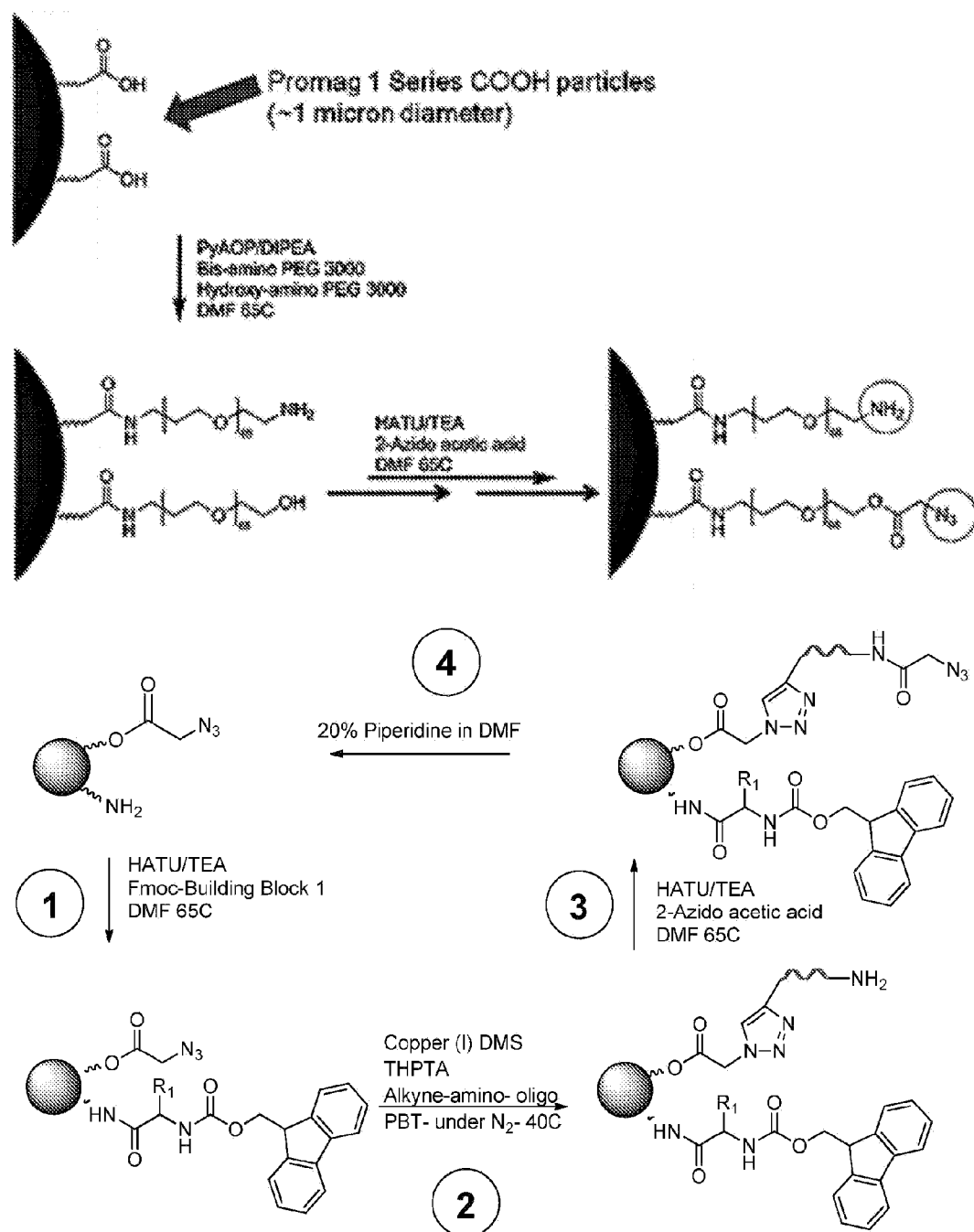


FIG. 4

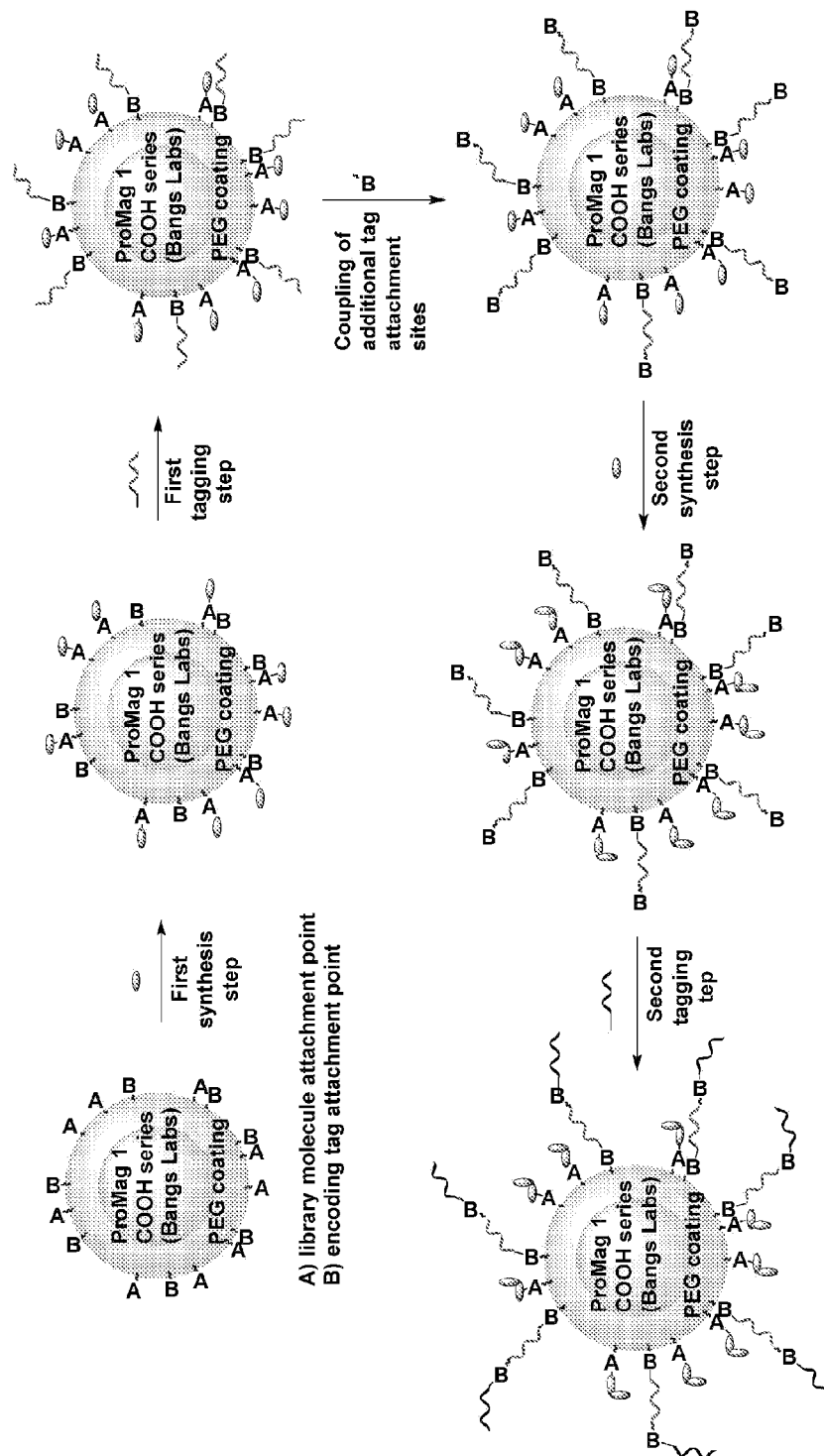


FIG. 5

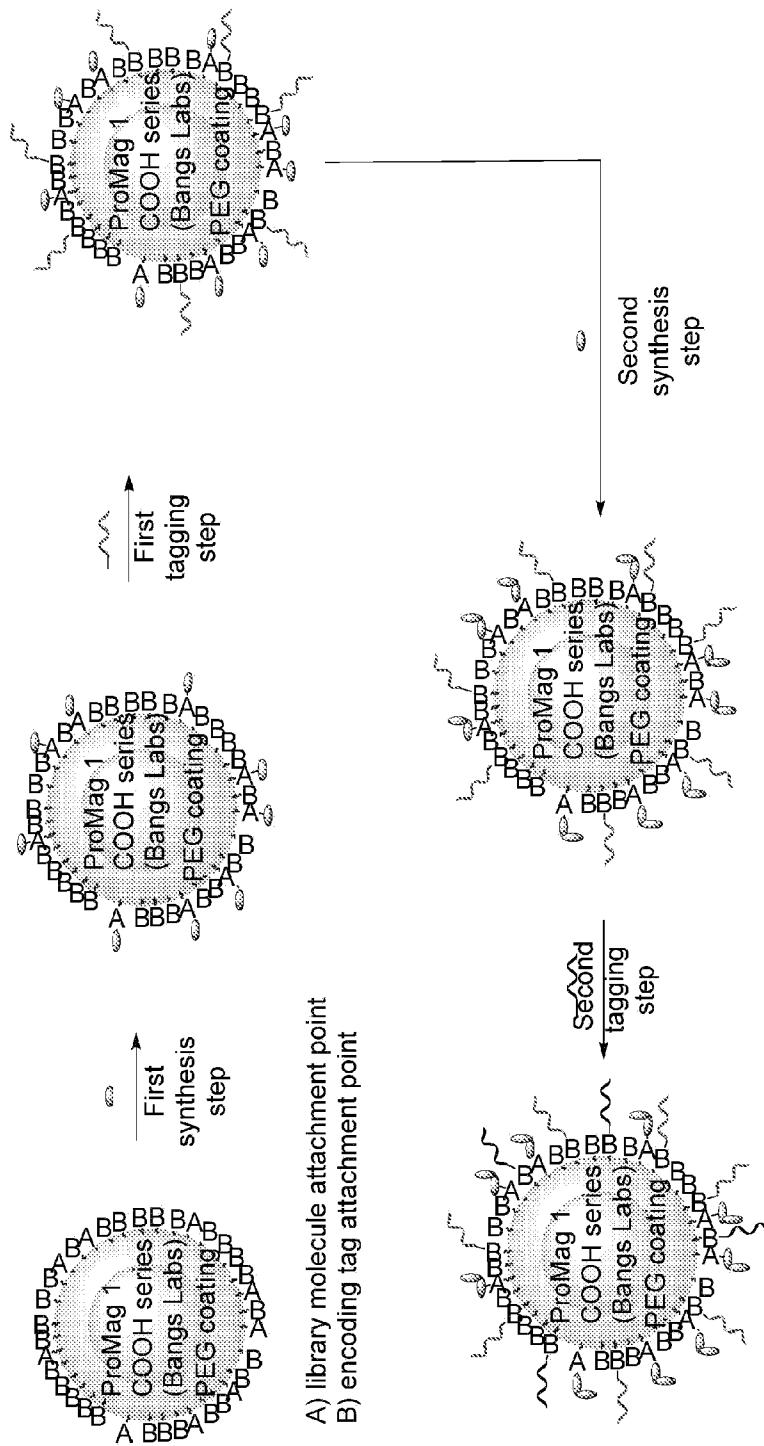


FIG. 6

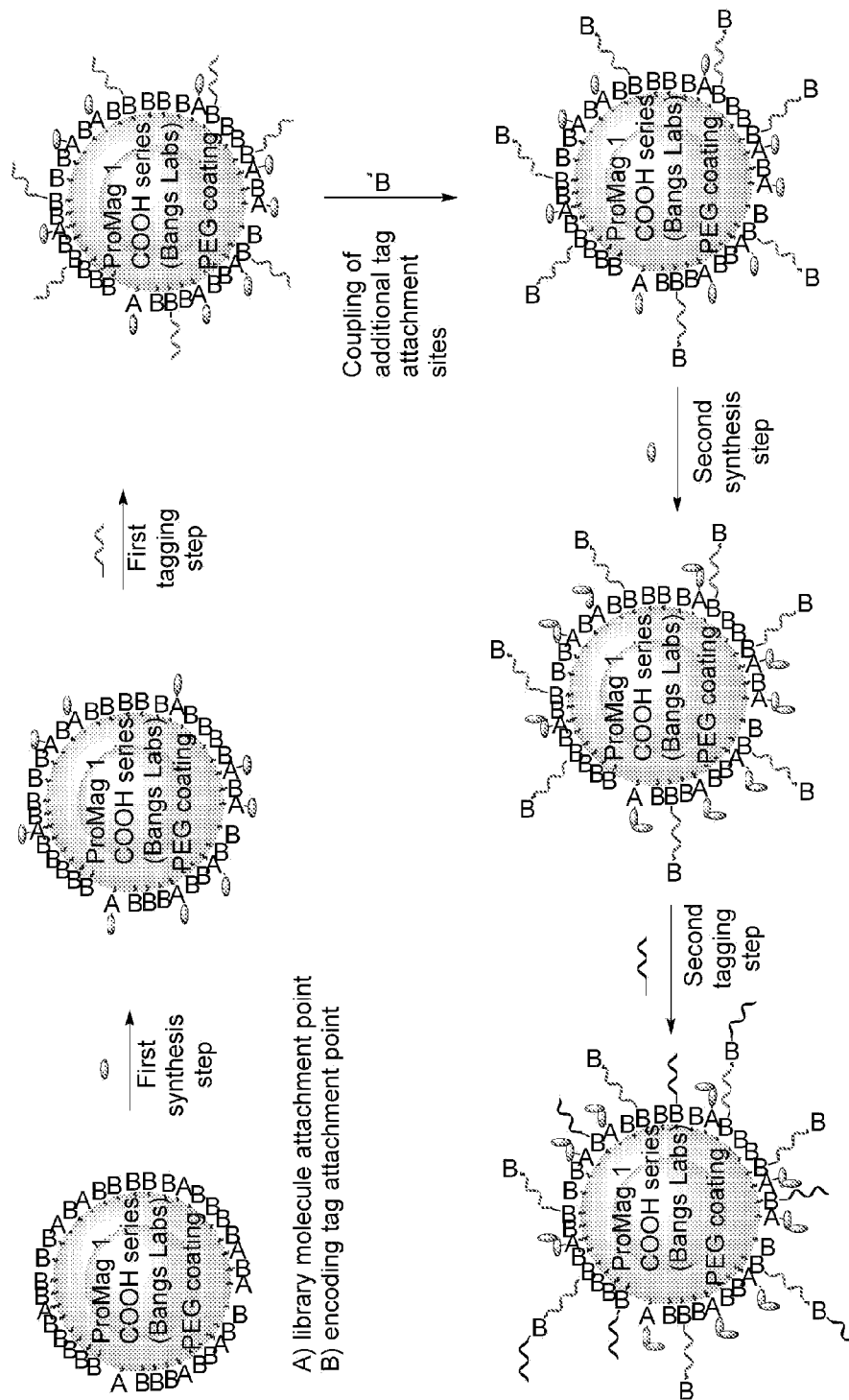
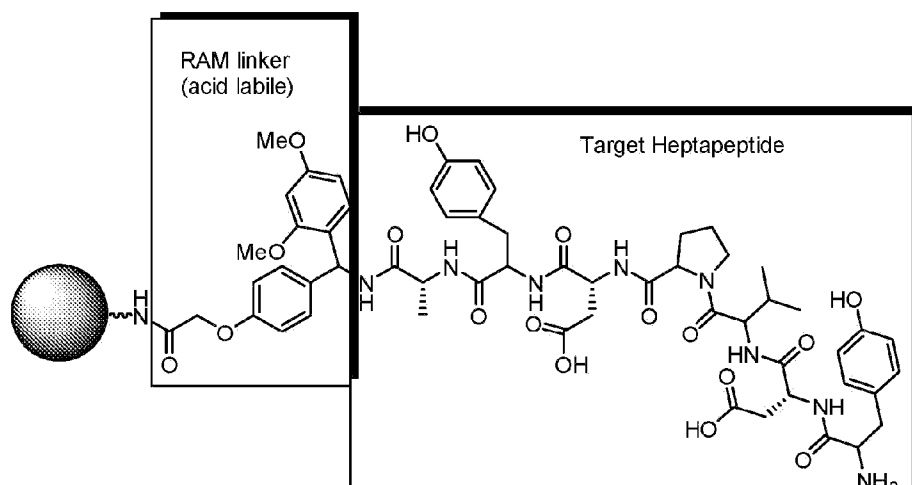


FIG. 8



rm13022D1101#1-126 RT: 0.01-1.91 AV: 125 NL: 3.92E6
T: FTMS + p NSI Full ms [150.00-2000.00]

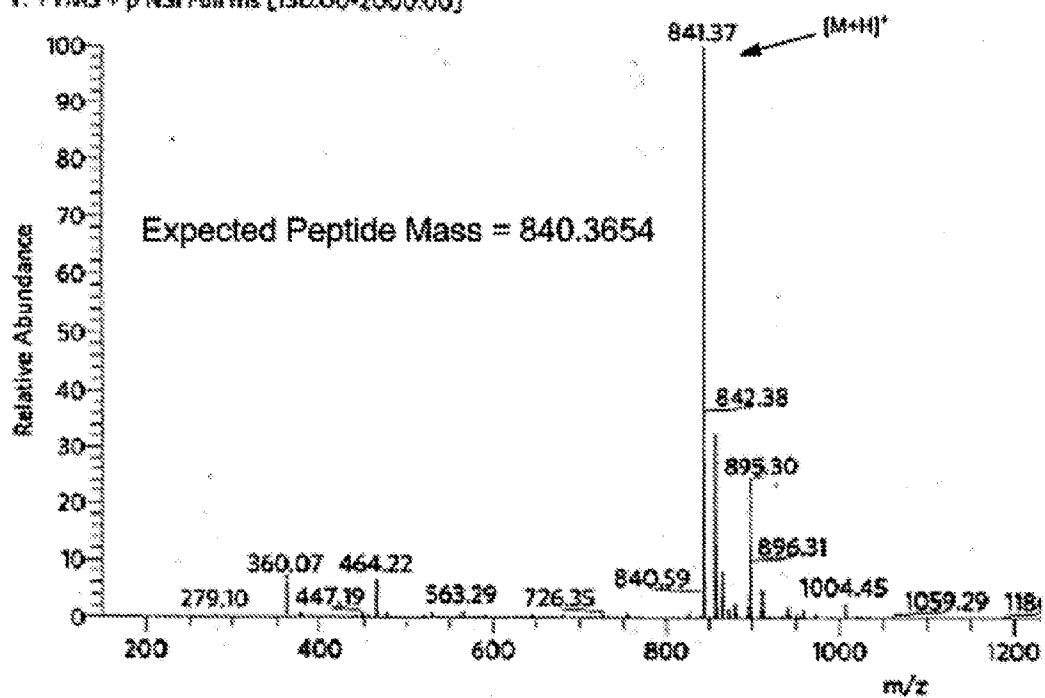


FIG. 10

Exposed to mixture of:

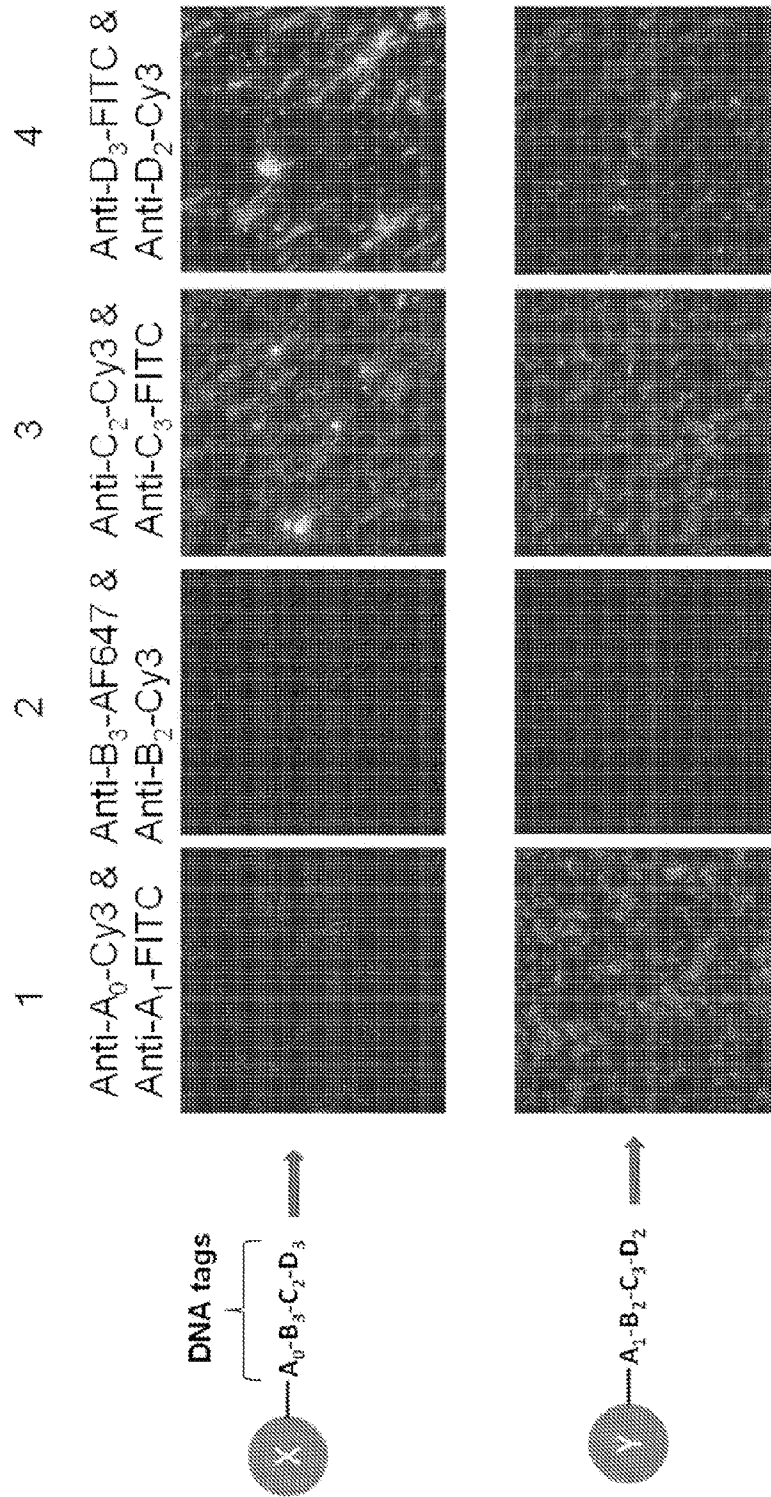


FIG. 11

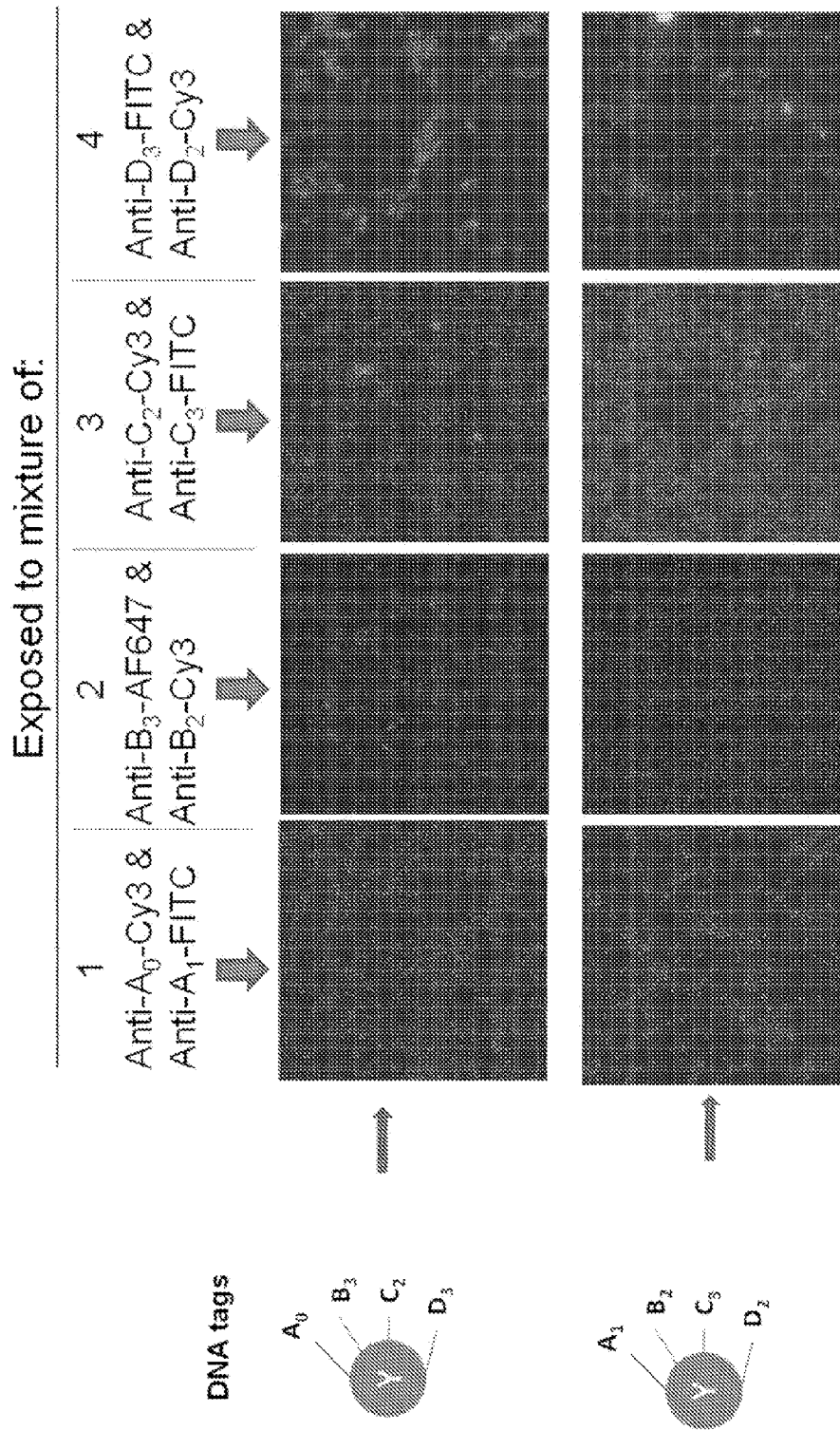


FIG. 12

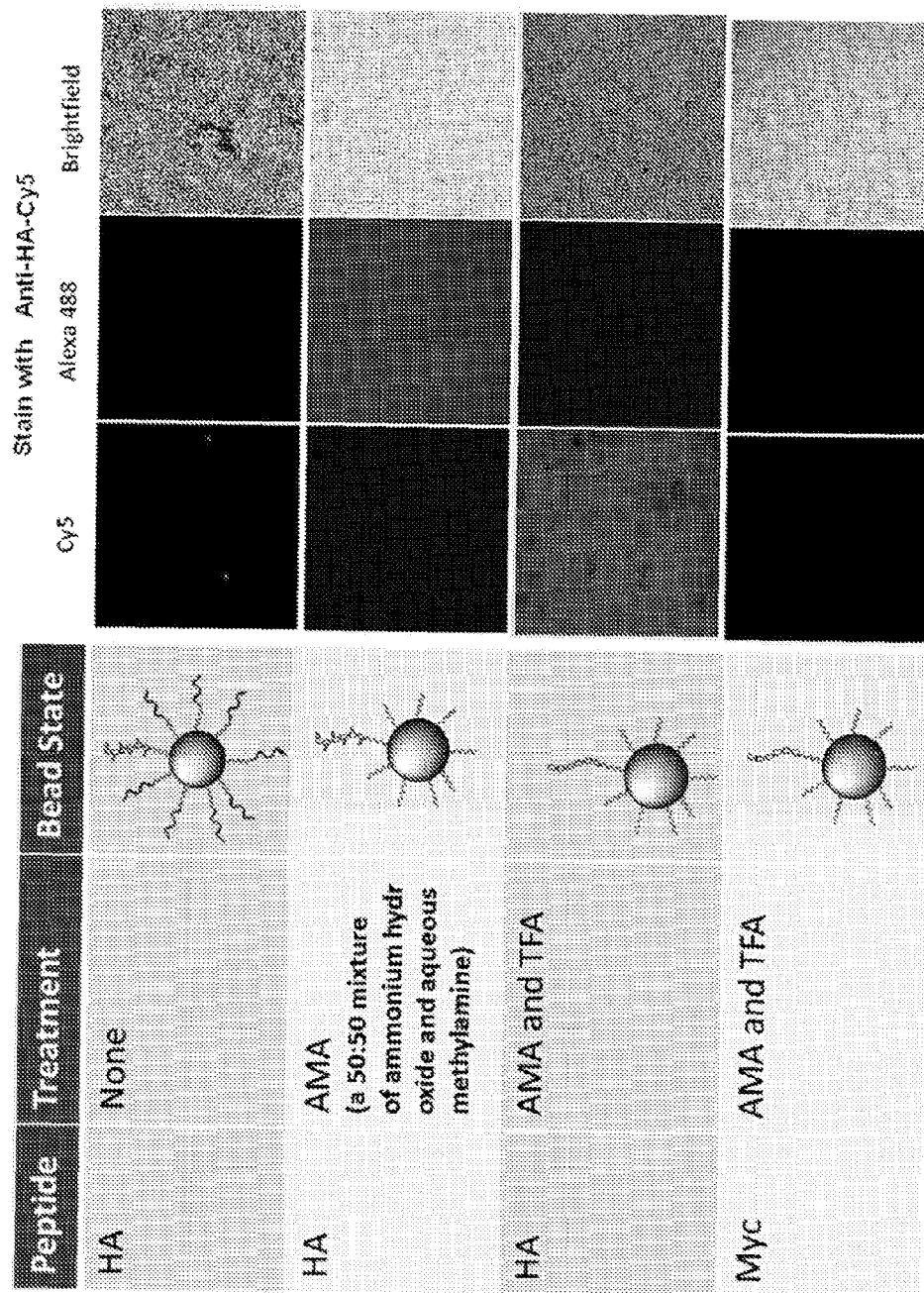


FIG. 13

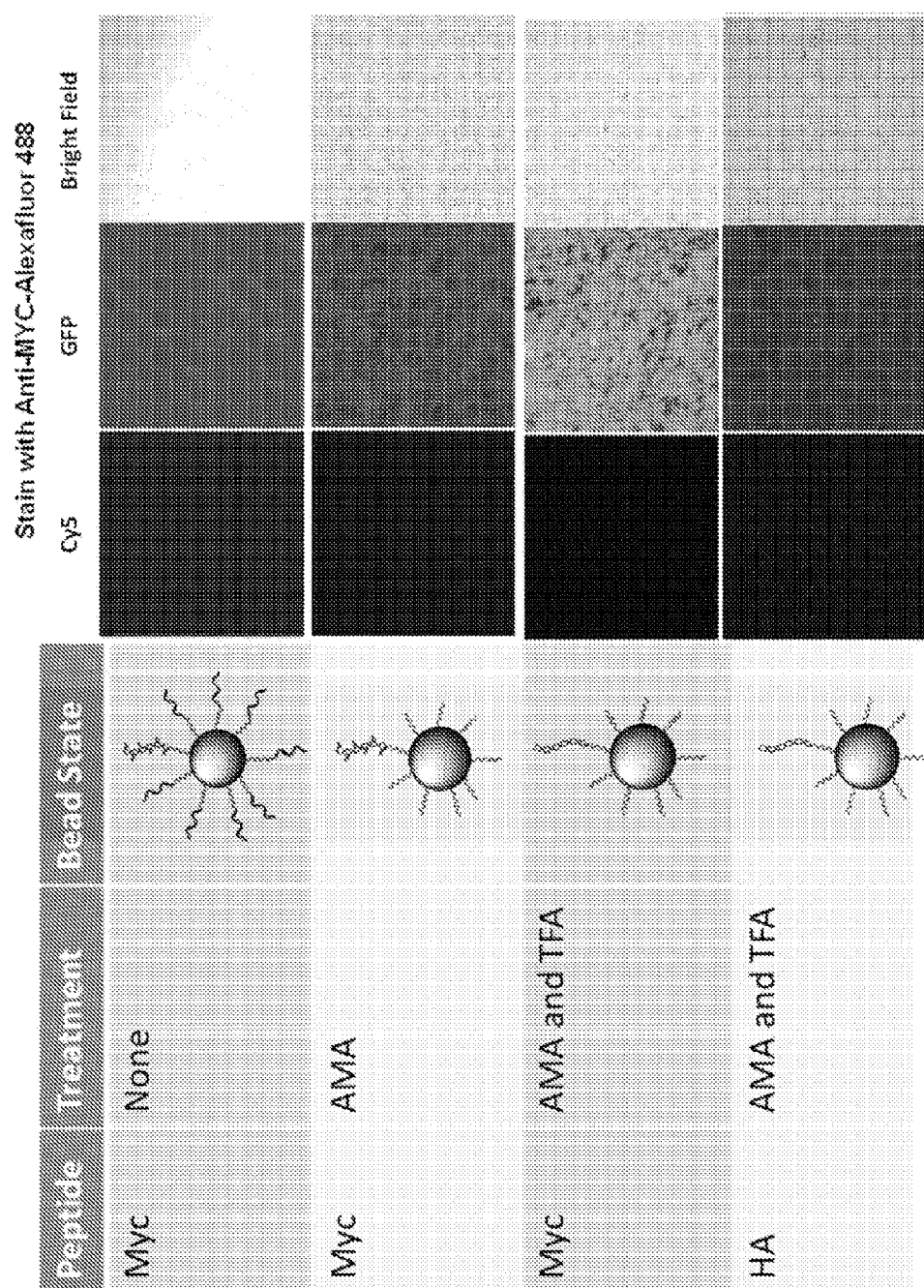


FIG. 14

Cy5 Channel
(Showing Anti-HA binding)

GFP Channel
(Showing Anti-Myc binding)

Cy5/GFP
Overlay

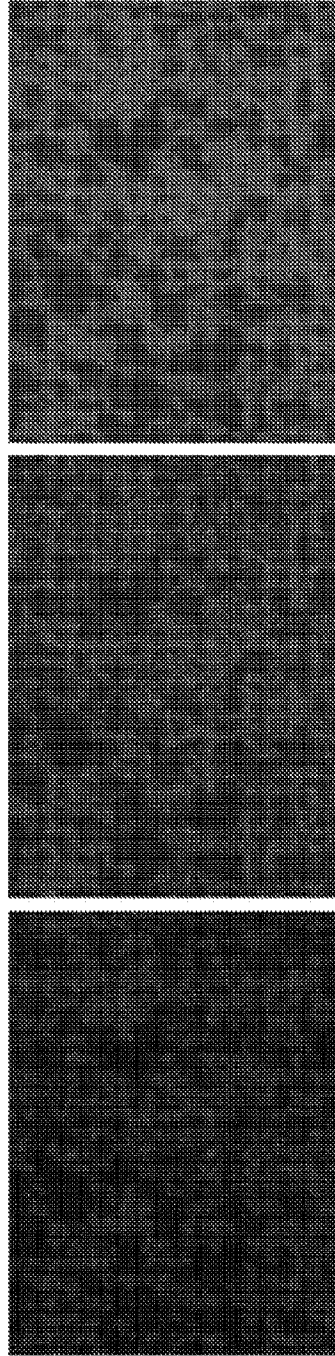


FIG. 16A

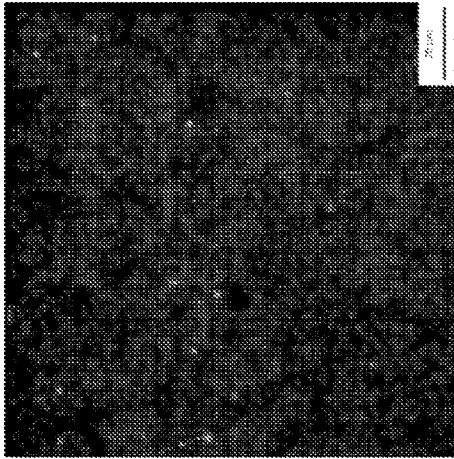


FIG. 16B

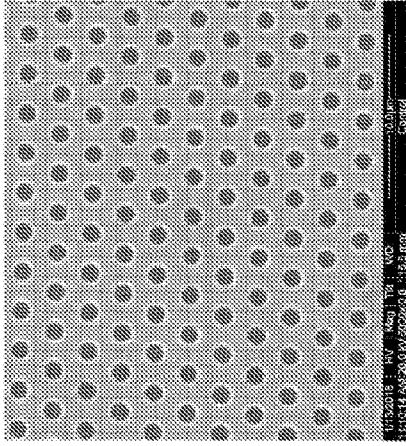


FIG. 16AC

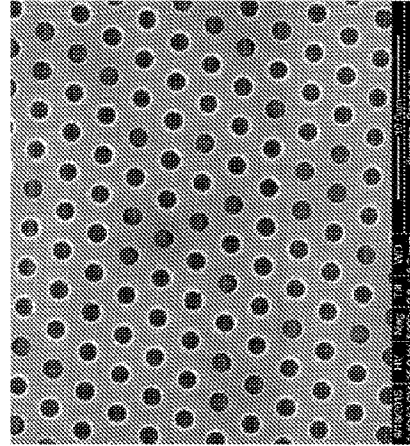
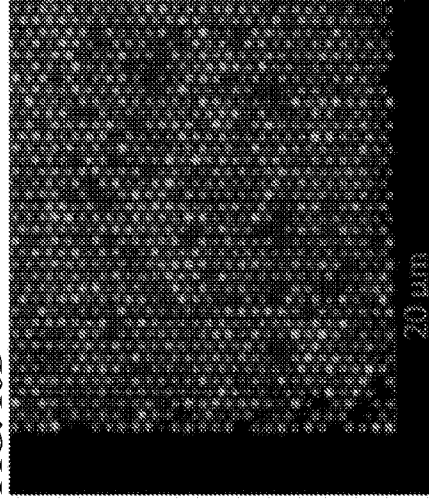


FIG. 16D



SEKVENSLISTE

Sekvenslisten er udeladt af skriftet og kan hentes fra det Europæiske Patent Register.

The Sequence Listing was omitted from the document and can be downloaded from the European Patent Register.

