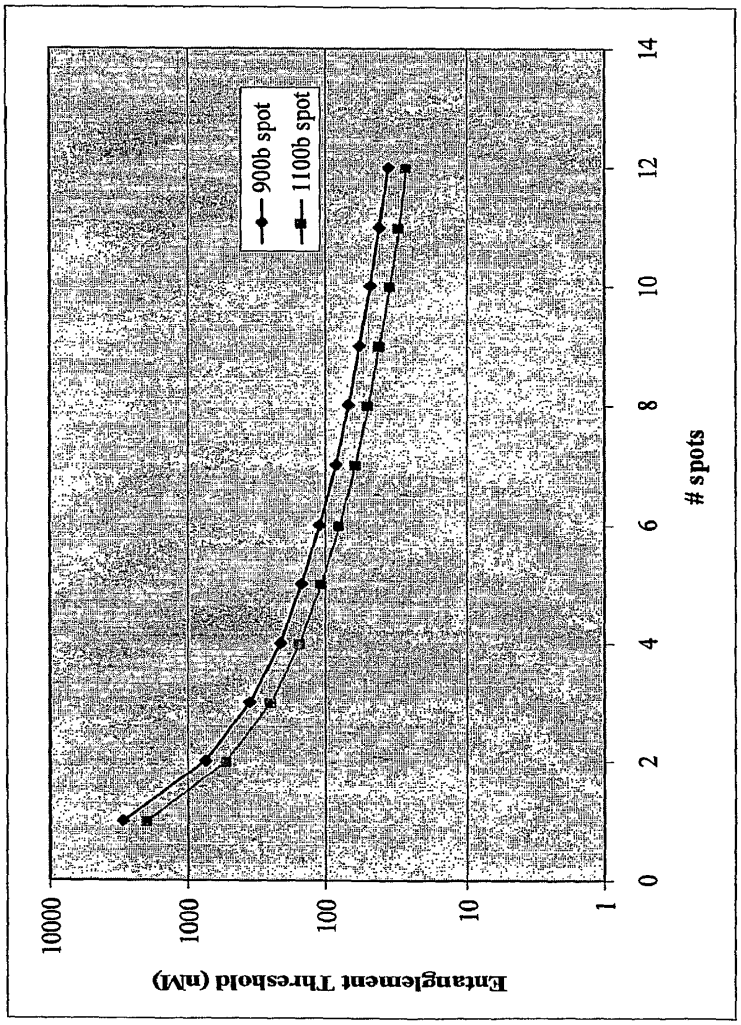
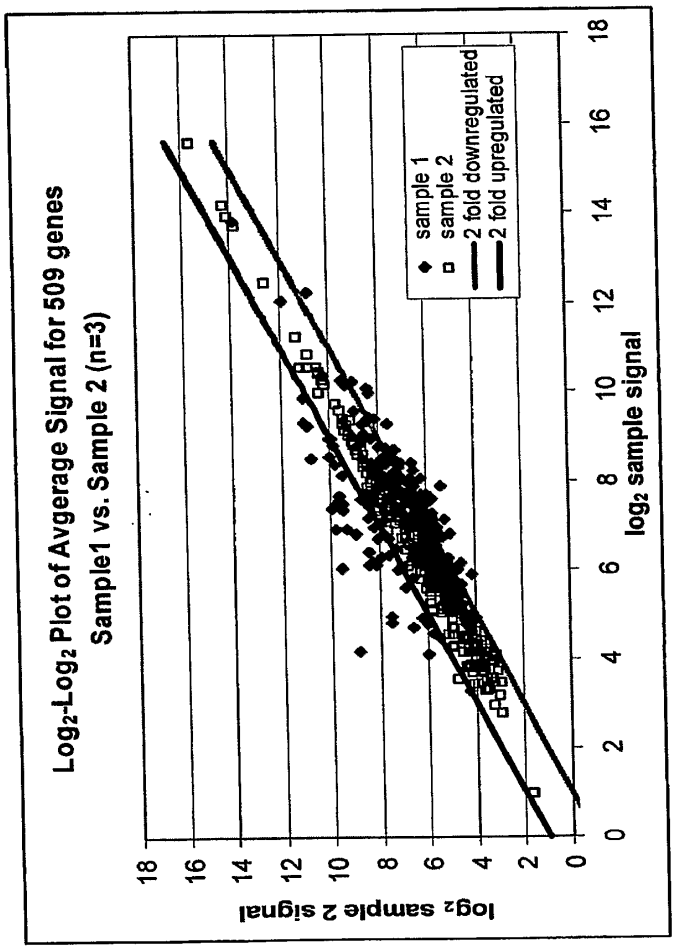

Entanglement Threshold (mM)
10000
1000
100
10
1
0
2
4
6
8
10
12
14
spots
900b spot
1100b spot

FIGURE 17

FIGURE 18


Log2-Log2 Plot of Avgerage Signal for 509 genes
Sample1 vs. Sample 2 (n=3)
log2 sample 2 signal
log2 sample signal
sample 1
sample 2
2 fold downregulated
2 fold upregulated
0
2
4
6
8
10
12
14
16
18

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 July 2007 (05.07.2007)

PCT

(10) International Publication Number
WO 2007/076129 A3

(51) International Patent Classification:
C07H 21/00 (2006.01)

(21) International Application Number:
PCT/US2006/049275

(22) International Filing Date:
22 December 2006 (22.12.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/753,758 23 December 2005 (23.12.2005) US
60/843,528 8 September 2006 (08.09.2006) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 201 Elliot Avenue West, Suite 300, Seattle, WA 98119 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GEISS, Gary, K.** [US/US]; 647 NW 78th Street, Seattle, WA 98117 (US). **FERREE, Sean, M.** [US/US]; 7018 3rd Avenue NW, Seattle, WA 98117 (US). **WEBSTER, Philippa, J.** [US/US]; 4317 Densmore Avenue N, Seattle, WA 98103 (US).

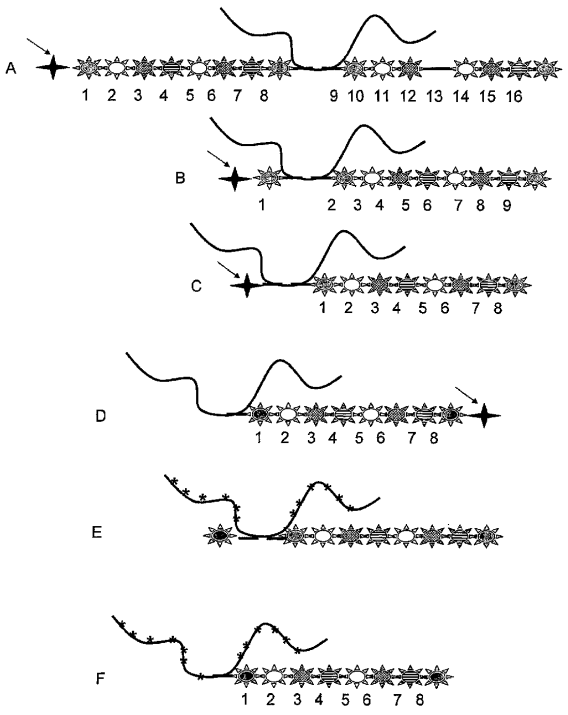
(74) Agents: **ANTLER, Andriane, M.** et al.; Jones Day, 222 East 41st Street, New York, NY 10017-6702 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

[Continued on next page]

(54) Title: NANOREPORTERS AND METHODS OF MANUFACTURING AND USE THEREOF



(57) Abstract: The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to coded, labeled probes that are capable of binding to and identifying target molecules based on the probes' label codes. Methods of making and using such probes are also provided. The probes can be used in diagnostic, prognostic, quality control and screening applications.

WO 2007/076129 A3

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 662

FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report: 6 March 2008

Published:

— *with international search report*

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 663

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 06/49275

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C07H 21/00 (2007.10)
USPC - 536/23.1, 23.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (PGPB, USPT, USOC, EPAB, JPAB); DialogPRO (Engineering); Google Scholar and PubMed.
Search Terms: nanoreporter, molecule, probe, label, monomer, nano, dual probe, probe pair, molecular probe, first label attachment, second label attachment, repeated sequence, nucleotide, target, region, affinity tag, first label,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,582,978 A (SHAH) 10 December 1996 (10.12.1996) col 14, ln 39 - 51.	1-7
Y	US 2005/0064435 A1 (SU et al.) 24 March 2005 (24.03.2005) para [0007], [0025], [0030], [0032], [0037] - [0039], [0073], [0145].	1-7, 68-71 and 73-74
Y	US 2003/0059955 (BAMDAD) 27 March 2003 (27.03.2003) para [0025], [0028], [0069].	1-7, 72
Y	US 2004/0002095 A1 (LIU et al.) 01 January 2004 (01.01.2004) fig 5; para [0007], [0008], [0036] - [0039], [0045], [0050], [0081], [0088], [0111], [0115].	29-75
Y	TUNTIWECHAPIKUL, W. Mechanism of in Vitro Expansion of Long DNA Repeats: Effect of Temperature, Repeat Length, Repeat Sequence, and DNA Polymerases. Biochemistry. January 2002, Vol. 41, No.3, pages 854 - 860.	29-75
Y	US 2005/0004078 A1 (BERGMANN et al.) 06 January 2005 (06.01.2005) para [0134], [0174].	36 and 73-74
Y	US 2003/0224357 A1 (SANTALUCIA et al.) 04 December 2003 (04.12.2003) para [0088].	62

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 November 2007 (12.11.2007)	17 DEC 2007
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 06/49275

Continuation of Box III:

Group I: Claims 1-7
Group II: Claims 17-28
Group III: Claims 29-75

The Search Authority finds that the invention contains three inventions described by the groups above, as the groups are not unified by a common innovative element. Specifically, Group I relates to a pair of probes, the first of which has two regions defined by light emitting monomers that also comprises a target sequence bound thereto; Group II relates to a probe bound to a target sequence, further comprising a synthetic nucleic acid hybridized in two regions to RNA probes, each of which comprises light-emitting monomers; Group III relates to a nucleic acid having one or more non-overlapping regions comprising a regularly repeated base. Although it is possible to utilize probes of the first and second groups in a common manner to achieve a means of detecting a target sequence, the means of achieving the embodiment of the detection contains only well-known elements of a target nucleic acid and light-emitting monomers as common elements. Further, no common elements can be found between the third group and either of the first two groups. Therefore, the three inventions in the application are considered separate.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 665

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 July 2007 (05.07.2007)

PCT

(10) International Publication Number
WO 2007/076132 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2006/049279

(22) International Filing Date:
22 December 2006 (22.12.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/753,816 23 December 2005 (23.12.2005) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 201 Elliot Avenue West, Suite 300, Seattle, WA 98119 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **FERREE, Sean, M.** [US/US]; 7018 3rd Avenue, Nw, Seattle, WA 98117 (US). **DUNAWAY, Dwayne, L.** [US/US]; 2751 NE 94th Street, Seattle, WA 98115 (US).

(74) Agents: **ANTLER, Adriane, M.** et al.; JONES DAY, 222 East 41st Street, New York, NY 10017-6702 (US).

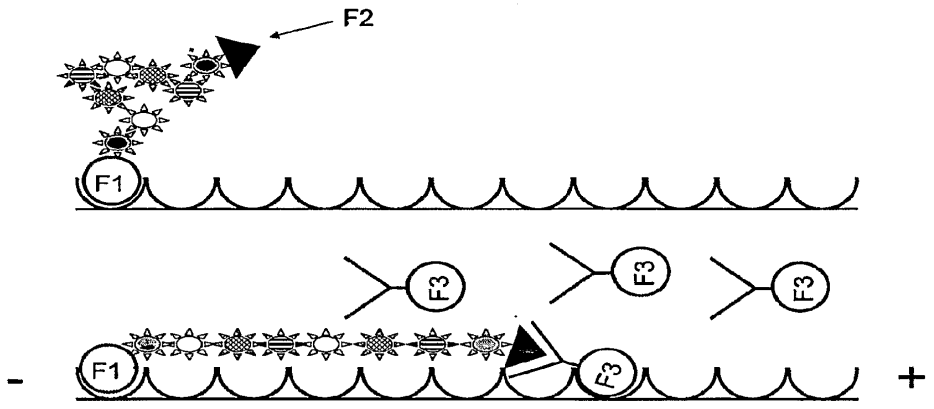
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** COMPOSITIONS COMPRISING ORIENTED, IMMOBILIZED MACROMOLECULES AND METHODS FOR THEIR PREPARATION



(57) **Abstract:** The present invention provides methods and compositions that facilitate the identification of structures of a variety of macromolecules. For instance, the present invention provides methods for the selective immobilization of macromolecules in an extended or oriented state. The present invention also provides compositions comprising macromolecules selectively immobilized in extended or oriented states.

WO 2007/076132 A2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 666

COMPOSITIONS COMPRISING ORIENTED, IMMOBILIZED MACROMOLECULES AND METHODS FOR THEIR PREPARATION

[0001] This application claims benefit of U.S. Provisional Application No. 60/753,816, filed December 23, 2005, which is incorporated by reference herein in its entirety.

[0002]

1. FIELD OF THE INVENTION

[0003] The present invention relates to the fields of macromolecular recognition, macromolecular labeling, macromolecular stretching, macromolecular orientation, macromolecular immobilization and nanotechnology.

2. BACKGROUND OF THE INVENTION

[0004] Macromolecules can comprise unique sequences of monomers or other moieties that can be used to readily distinguish one macromolecule from another. When a macromolecule is bound to or associated with another entity, that entity can be identified by features of the macromolecule. The entity can be any entity apparent to one to those of skill in the art such as a substrate, a surface, another molecule, a position in an array or any other entity with which a macromolecule can be associated. One can readily identify the entity by identifying the macromolecule.

[0005] Of the many features of macromolecules, their structure provides much of their diversity. Generally, their primary structure, *i.e.* the sequence of their monomers or other moieties, distinguishes one macromolecule from another macromolecule. However, many, if not most, macromolecules comprise further levels of structure, such as secondary, tertiary or even quaternary structure, that can actually hinder the ability of one of skill to identify the primary structure of the macromolecule. To illustrate, a simple macromolecule comprising a primary sequence of features A, B, C, D, E and F can be readily identified if each feature can be recognized one after the other along the primary sequence. However, if that macromolecule were twisted, kinked or folded into a three dimensional structure, as is often the case with naturally occurring macromolecules in solution, the features might not be readily identifiable in their proper sequence. One of skill in the art might not be able to distinguish a macromolecule having the primary structure A, B, F, C, E and D from the primary structure A, B, C, D, E and F due to secondary or tertiary or quaternary structure.

[0006] Persons of skill have developed techniques to tease primary structure out of a three dimensional macromolecule. For certain macromolecules, *e.g.* polynucleotides and polypeptides, chemical techniques for identifying the primary sequence of their features will be familiar to those of skill. In addition, a few techniques for extending or stretching or combing macromolecules have been developed to reduce the complexity of their structures and thereby facilitate the elucidation of primary structure. These techniques have generally involved the application of some force capable of extending the macromolecule. Some techniques have further involved the nonselective fixing of the macromolecule in an extended state, for instance by drying the macromolecule on a surface.

[0007] Methods and compositions that facilitate the identification of the primary structure of a macromolecule will further enhance their utility in the fields of macromolecular recognition and macromolecular labeling and other fields.

3. SUMMARY OF THE INVENTION

[0008] The present invention provides methods and compositions that facilitate the identification of primary structures of a variety of macromolecules. In certain aspects, the present invention provides methods for the selective immobilization of macromolecules in an extended state. Remarkably, according to the invention, a macromolecule can be selectively immobilized while fully extended under whatever force is used for the extension. In addition, the methods of the invention facilitate the selective immobilization of extended macromolecules that are oriented with respect to each other. In other words, according to the methods of the invention, a plurality of macromolecules can readily be immobilized in the same orientation with respect to each other.

[0009] In one aspect, the present invention provides methods for selectively immobilizing a macromolecule in an extended state. The macromolecule can be any macromolecule known to those of skill in the art such as a polymer, a polysaccharide, a polynucleotide or a polypeptide. For the methods of this aspect of the invention, generally, a first portion of the macromolecule is immobilized by any technique known to those of skill in the art. Indeed, the technique for immobilizing the first portion of the macromolecule is not critical to many embodiments of the invention. In certain embodiments, the first portion of the macromolecule can be immobilized selectively or non-selectively. In certain embodiments the first portion is immobilized by one or more covalent bonds. In certain embodiments, the first portion is immobilized by one or more non-covalent bonds. Exemplary immobilized first portions are described in the sections below.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 668

[0010] With an immobilized first portion, the macromolecule can be extended by any technique for extending a macromolecule apparent to those of skill in the art. In certain embodiments, the technique for extending the macromolecule is not critical for the methods of the invention. In certain embodiments, the technique for extending the macromolecule appropriate for the class of macromolecule according to the judgment of one of skill in the art. In certain embodiments, the macromolecule is extended by application of a force capable of extending the macromolecule. The force can be any force apparent to one of skill in the art for extending the macromolecule. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the macromolecule are described in the sections below.

[0011] The macromolecule is in an extended state if it would be recognized as extended by one of skill in the art. In certain embodiments, the macromolecule is in an extended state when it is in the field of a force capable of extending the macromolecule. In certain embodiments, the macromolecule is in an extended state when its average hydrodynamic radius is more than double the average hydrodynamic radius of the macromolecule in its native state as recognized by those of skill in the art.

[0012] In this aspect of the invention, the methods generally comprise the step of selectively immobilizing a second portion of the macromolecule while it is in an extended state. This can result in an immobilized macromolecule that is extended between the first and the second portion. Remarkably, since the macromolecule is selectively immobilized while extended, that extension can be preserved in the immobilized macromolecule. Generally, the first portion and the second portion of the macromolecule are not the same.

[0013] The selective immobilization can be according to any technique for selective immobilization of a portion of a macromolecule apparent to those of skill in the art. The selective immobilization can be through, for example, the formation of one or more covalent bonds or one or more non-covalent bonds, or both. Particular examples of selective immobilization techniques are described in the sections below. In particular embodiments, one or more binding pairs are used to immobilize the second portion of the macromolecule.

[0014] The second portion can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any substrate judged to be useful for immobilization known to those of skill in the art. In certain embodiments, the second portion can be immobilized to another molecule. Further useful substrates include surfaces, membranes, beads, porous materials, electrodes, arrays and any other substrate apparent to those of skill in the art.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 669

[0015] In another aspect, the present invention provides compositions comprising a selectively immobilized, extended macromolecule. The compositions generally comprise a substrate and an extended macromolecule selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the macromolecule are immobilized onto the substrate, and the macromolecule is in an extended state between the two portions. In certain embodiments, at least one portion of the macromolecule is selectively immobilized onto the substrate. In certain embodiments, two or more portions of the macromolecule are selectively immobilized onto the substrate. The macromolecule can be extended and/or immobilized by any technique apparent to those of skill, including particularly the methods of the present invention.

[0016] In another aspect, the present invention provides methods for selectively immobilizing a macromolecule in an oriented state. The macromolecule can be any macromolecule described above. In certain embodiments, the macromolecule can be flexible, or in certain embodiments the macromolecule can be rigid or semi-rigid. For the methods of this aspect of the invention, generally, a first portion of the macromolecule is immobilized as described above. With an immobilized first portion, the macromolecule can be oriented by any technique for extending a macromolecule apparent to those of skill in the art. In certain embodiments, the technique for orienting the macromolecule is not critical for the methods of the invention. In certain embodiments, the technique for orienting the macromolecule appropriate for the class of macromolecule according to the judgment of one of skill in the art. In certain embodiments, the macromolecule is oriented by application of a force capable of orienting the macromolecule. The force can be any force apparent to one of skill in the art for orienting the macromolecule. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the macromolecule are described in the sections below.

[0017] The macromolecule is in an oriented state if it would be recognized as oriented by one of skill in the art. In certain embodiments, the macromolecule is in an oriented state when it is in the field of a force capable of orienting the macromolecule. In certain embodiments, the macromolecule is in an oriented state when its termini are arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the macromolecule. In certain embodiments, a plurality of macromolecules is in an oriented state when the termini of the macromolecules are arranged in parallel, as recognized by those of skill in the art.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 670

[0018] In this aspect of the invention, the methods generally comprise the step of selectively immobilizing a second portion of the macromolecule while it is in an oriented state. This can result in an immobilized macromolecule that is oriented between the first and the second portion. Remarkably, since the macromolecule is selectively immobilized while extended, that orientation can be preserved in the immobilized macromolecule. The selective immobilization can according to the methods described above.

[0019] In another aspect, the present invention provides a compositions comprising a selectively immobilized, oriented macromolecule. The compositions generally comprise a substrate and an oriented macromolecule selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the macromolecule are immobilized onto the substrate, and the macromolecule is in an oriented state between the two portions. In certain embodiments, at least one portion of the macromolecule is selectively immobilized onto the substrate. In certain embodiments, both portions of the macromolecule are selectively immobilized onto the substrate. The macromolecule can be oriented and/or immobilized by any technique apparent to those of skill, including particularly the methods of the present invention.

[0020] The methods and compositions of the present invention can be used for any purpose apparent to those of skill in the art. For instance, the immobilized and extended and/or oriented macromolecule can be used as a label for a substrate on which the macromolecule is immobilized. The primary sequence of the immobilized and extended and/or oriented macromolecule can be identified by any technique apparent to those of skill. Advantageously, immobilization of the extended and/or oriented macromolecule can facilitate such techniques. In certain embodiments, the immobilized and extended and/or oriented macromolecule can be used to guide the manufacture of nanopaths, for example to create nanowires or nanocircuits. Further uses for the immobilized and extended and/or oriented macromolecules are described in the sections below.

4. BRIEF DESCRIPTION OF THE FIGURES

[0021] FIG. 1A provides an illustration of a macromolecule comprising an immobilized first portion F1;

[0022] FIG. 1B provides an illustration of a macromolecule extended in an electrical field and comprising immobilized first portion F1 and immobilized second portion F2, wherein F2 is immobilized via a complex with molecule F3;

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 671

[0023] FIG. 2A provides an illustration of a three-member complex for immobilization of an extended macromolecule;

[0024] FIG. 2B provides an illustration of a two-member complex for immobilization of an extended macromolecule;

[0025] FIG. 2C provides an illustration of an incomplete complex for immobilization of an extended macromolecule;

[0026] FIG. 3A provides an illustration of a macromolecule comprising an immobilized first portion F1;

[0027] FIG. 3B provides an illustration of an extended macromolecule immobilized at first portion F1 and at a second portion via complexes with F2;

[0028] FIG. 3C provides an illustration of a macromolecule comprising a first portion immobilized to an avidin surface via biotin;

[0029] FIG. 3D provides an illustration of an extended macromolecule immobilized at a first portion and at a second portion via selective binding of biotin to an avidin surface;

[0030] FIG. 4A illustrates immobilization of one terminus of a DNA molecule in a microfluidic device;

[0031] FIG. 4B illustrates extension of the DNA in an electric field;

[0032] FIG. 4C illustrates selective immobilization of a second terminus of the extended DNA molecule;

[0033] FIG. 5 provides an image of extended macromolecules selectively immobilized by the methods of the present invention;

[0034] FIG. 6A provides a nanoreporter in which alternate spots are labeled;

[0035] FIG. 6B depicts a nanoreporter in which every spot is labeled;

[0036] FIG. 7A illustrates a dual nanoreporter with a 16-position nanoreporter code, using two 8-position nanoreporter components;

[0037] FIG. 7B illustrates a dual nanoreporter with a 9-position nanoreporter code;

[0038] FIG. 7C illustrates a dual nanoreporter with an 8-position nanoreporter code, using one ghost probe and one 8-position nanoreporter component;

[0039] FIG. 8A provides a schematic illustration of the experiment shown in Figures 6B and 6C where the star represents biotin that was used to attach the complex by one end to the surface prior to stretching;

[0040] FIGS. 8B and 8C provide images from experiments in which S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA (FIG. 8B) or S2 target RNA (FIG. 8C) were hybridized;

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 672

[0041] FIG. 8D provides an image of a negative control experiment, in which S2-A ghost probe, S2-B labeled nanoreporter and no S2 target RNA were hybridized; and

[0042] FIG. 8E provides a close-up of a nanoreporter complexes from FIG. 8B.

5. DETAILED DESCRIPTION OF THE INVENTION

5.1 Definitions

[0043] All terms used herein have their ordinary meanings to those of skill in the art unless indicated otherwise. The following terms shall have the following meanings.

[0044] As used herein, the term "binding pair" refers to first and second molecules or moieties that are capable of selectively binding to each other, *i.e.* binding to each other with greater affinity than to other components in a composition. The binding between the members of the binding pair can be covalent or non-covalent. In certain embodiments, the binding is noncovalent. Exemplary binding pairs include immunological binding pairs (*e.g.* any haptenic or antigenic compound in combination with a corresponding antibody or binding portion or fragment thereof, for example digoxigenin and anti-digoxigenin, fluorescein and anti-fluorescein, dinitrophenol and anti-dinitrophenol, bromodeoxyuridine and anti-bromodeoxyuridine, mouse immunoglobulin and goat anti-mouse immunoglobulin) and nonimmunological binding pairs (*e.g.*, biotin-avidin, biotin-streptavidin, hormone-hormone binding protein, receptor-receptor ligand (*e.g.*, acetylcholine receptor-acetylcholine or an analog thereof), IgG-protein A, lectin-carbohydrate, enzyme-enzyme cofactor, enzyme-enzyme inhibitor, complementary polynucleotide pairs capable of forming nucleic acid duplexes, and the like). For instance, immunoreactive binding members may include antigens, haptens, aptamers, antibodies (primary or secondary), and complexes thereof, including those formed by recombinant DNA methods or peptide synthesis. An antibody may be a monoclonal or polyclonal antibody, a recombinant protein or a mixture(s) or fragment(s) thereof, as well as a mixture of an antibody and other binding members. Other common binding pairs include but are not limited to, biotin and avidin (or derivatives thereof), biotin and streptavidin, carbohydrates and lectins, complementary nucleotide sequences (including probe and capture nucleic acid sequences), complementary peptide sequences including those formed by recombinant methods, effector and receptor molecules, hormone and hormone binding protein, enzyme cofactors and enzymes, enzyme inhibitors and enzymes, and so forth.

[0045] "Selective binding" refers to the any preferential binding of a pair of molecules or moieties for each other with respect to other molecules or moieties in a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 673

composition that would be recognized by one of skill in the art. In certain embodiments, a pair of molecules or moieties selectively binds when they preferentially bind each other compared to other molecules or moieties. Selective binding can include affinity or avidity, or both, of one molecule or moiety for another molecule or moiety. In particular embodiments, selective binding requires a dissociation constant (K_D) of less than about $1x10^{-5}$ M or less than about $1x10^{-6}$ M, $1x10^{-7}$ M, $1x10^{-8}$ M, $1x10^{-9}$ M, or $1x10^{-10}$ M. In contrast, in certain embodiments, non-selective binding has significantly less affinity, for example, a K_D greater than $1x10^{-3}$ M.

[0046] "Extended state" refers to a macromolecule in a state that would be recognized as extended by one of skill in the art. In certain embodiments, a macromolecule is in an extended state when it is extended relative to its native conformation in solution. In certain embodiments, a macromolecule is in an extended state when it is in the field of a force capable of extending the macromolecule. In certain embodiments, an extended state of a macromolecule can be determined quantitatively. In such embodiments, those of skill in the art will recognize R as the end-to-end vector of the macromolecule, *i.e.* the distance between two termini of the macromolecule, and <R> as the average end-to-end vector such that 95% of R will be within 2<R> in a solution deemed appropriate to one of skill in the art. Exemplary solutions include, for example, a dilute solution of the macromolecule in water or in a pH buffer. In particular embodiments, a macromolecule is in an extended state when R is greater than 2.0<R>.

[0047] "Oriented state" refers to a macromolecule in a state that would be recognized as oriented by one of skill in the art. In certain embodiments, a macromolecule is in an oriented state when it is oriented relative to its native conformation in solution. In certain embodiments, the macromolecule is oriented when it is arranged in parallel with the field of a force capable of orienting the macromolecule. In certain embodiments, the macromolecule is oriented when it is one of a plurality of macromolecules that are arranged in parallel, as recognized by those of skill in the art.

5.2 Methods of Selective Immobilization

[0048] As described in the summary, the present invention provides methods for the selective immobilization of a macromolecule in an extended state. The macromolecule, once selectively immobilized, can be used for any purpose apparent to those of skill in the art.

5.2.1 Macromolecules

[0049] In the methods, the macromolecule can be any macromolecule known to those of skill in the art without limitation. In certain embodiments, the macromolecule is a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 674

macromolecule that is capable of being extended in the methods of the invention. In certain embodiments, the macromolecule is capable of being immobilized in one or two portions as described in the sections below.

[0050] In certain embodiments, the macromolecule is any polymer known to those of skill in the art. For instance, the macromolecule can be a polysaccharide, a polypeptide or a polynucleotide. Useful polynucleotides include ribonucleic acids, deoxyribonucleic acids and other polynucleotides known to those of skill in the art.

[0051] The macromolecule can be of any size that is sufficient to allow extension and immobilization of the macromolecule according to the methods of the invention. In certain embodiments when the macromolecule is a polynucleotide, the macromolecule can have a length of greater than 500 bp, greater than 750 bp, greater than 1 kb, greater than 1.5 kb, greater than 2.0 kb, greater than 2.5 kb, greater than 3.0 kb, greater than 4.0 kb or greater than 5.0 kb. In certain embodiments, when the macromolecule is a polypeptide, the macromolecule can have a size of greater than 50 amino acids, greater than 100 amino acids, greater than 200 amino acids, greater than 300 amino acids, greater than 400 amino acids, greater than 500 amino acids, greater than 750 amino acids, greater than 1000 amino acids, greater than 1500 amino acids, greater than 2000 amino acids, greater than 2500 amino acids, greater than 3000 amino acids, greater than 4000 amino acids or greater than 5000 amino acids. In certain embodiments, when the macromolecule is a polysaccharide, the macromolecule can have a size of greater than 50 saccharides, greater than 100 saccharides, greater than 200 saccharides, greater than 300 saccharides, greater than 400 saccharides, greater than 500 saccharides, greater than 750 saccharides, greater than 1000 saccharides, greater than 1500 saccharides, greater than 2000 saccharides, greater than 2500 saccharides, greater than 3000 saccharides, greater than 4000 saccharides or greater than 5000 saccharides.

[0052] The macromolecule can be a native macromolecule as understood by those of skill in the art, or the macromolecule can be a non-native macromolecule. In certain embodiments, when the macromolecule is a polypeptide, the macromolecule can comprise only naturally occurring amino acids, or the macromolecule can comprise naturally occurring amino acids and non-naturally occurring amino acids. The other amino acids can be any amino acids, or derivatives or analogs thereof, known to those of skill in the art. In certain embodiments, when the macromolecule is a polynucleotide, the polynucleotide can comprise only naturally occurring nucleotides, or the polynucleotide can comprise naturally occurring nucleotides and non-naturally occurring nucleotides. In certain embodiments, when the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 675

macromolecule is a polysaccharide, the polysaccharide can comprise only naturally occurring saccharides, or the polysaccharide can comprise naturally occurring saccharides and non-naturally occurring saccharides. In certain embodiments, the polymers can comprise only non-natural monomers. In further embodiments, the macromolecule can comprise a plurality of classes of monomers, such as amino acids, nucleotides and/or saccharides.

[0053] In certain embodiments, the macromolecule comprises only one primary, covalently linked chain of monomers. For instance, when the macromolecule is a polypeptide, in certain embodiments, the macromolecule comprises only one primary amino acid chain. When the macromolecule is a polynucleotide, in certain embodiments, the macromolecule is single stranded. In further embodiments, the macromolecule comprises two primary, covalently linked chains of monomers. For instance, when the macromolecule is a polypeptide, in certain embodiments, the macromolecule comprises two primary amino acid chains. When the macromolecule is a polynucleotide, in certain embodiments, the macromolecule comprises two polynucleotide strands; in certain embodiments, the macromolecule can be double stranded, in part or in whole. In further embodiments, the macromolecule comprises three or more primary, covalently linked chains of monomers. For instance, when the macromolecule is a polypeptide, in certain embodiments, the macromolecule comprises three primary amino acid chains. When the macromolecule is a polynucleotide, in certain embodiments, the macromolecule comprises three polynucleotide strands. For instance, the macromolecule can comprise three strands F1, X and F2 where a portion of strand X is complementary to strand F1 and a portion of strand X is complementary to strand F2. An example is illustrated in FIG. 2A. In certain embodiments, the macromolecule comprises more than three primary, covalently linked chains of monomers.

[0054] Advantageously, a macromolecule of the invention can comprise one or more labels that facilitate the detection, imaging or identification of the macromolecule by techniques known to those of skill in the art. The label can be any detectable moiety known to those of skill in the art. Exemplary labels for macromolecules include detectable isotopes, radioisotopes, fluors, dyes, enzymes, ligands, receptors, antigens, antibodies, lectins, carbohydrates, nucleotide sequences, and any other detectable label apparent to those of skill in the art. Useful labels, macromolecules comprising labels, and methods of their preparation are described in U.S. provisional application no. 60/753,758, filed December 23, 2005, entitled "Nanoreporters And Methods Of Manufacturing And Use Thereof," the contents of which are hereby incorporated by reference in their entirety.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 676

[0055] In certain embodiments, a polynucleotide is a polymer of natural (*e.g.* A, G, C, T, U) or synthetic nucleobases, or a combination of both. The backbone of the polynucleotide can be composed entirely of "native" phosphodiester linkages, or it may contain one or more modified linkages, such as one or more phosphorothioate, phosphorodithioate, phosphoramidate or other modified linkages. As a specific example, a polynucleotide may be a peptide nucleic acid (PNA), which contains amide interlinkages. Additional examples of synthetic bases and backbones that can be used in conjunction with the invention, as well as methods for their synthesis can be found, for example, in U.S. Patent No. 6,001,983; Uhlman & Peyman, 1990, *Chemical Review* 90(4):544 584; Goodchild, 1990, *Bioconjugate Chem.* 1(3):165 186; Egholm *et al.*, 1992, *J. Am. Chem. Soc.* 114:1895 1897; Gryaznov *et al.*, *J. Am. Chem. Soc.* 116:3143 3144, as well as the references cited in all of the above. Common synthetic nucleobases of which polynucleotides may be composed include 3-methlyuracil, 5,6-dihydrouracil, 4 thiouracil, 5 bromouracil, 5-thorouracil, 5-iodouracil, 6-dimethyl aminopurine, 6-methyl aminopurine, 2-aminopurine, 2,6-diamino purine, 6-amino-8-bromopurine, inosine, 5-methylcytosine, 7-deazaadenine, and 7-deazaguanosine. Additional non-limiting examples of synthetic nucleobases of which the target nucleic acid may be composed can be found in Fasman, *CRC Practical Handbook of Biochemistry and Molecular Biology*, 1985, pp. 385-392; *Beilstein's Handbuch der Organischen Chemie*, Springer Verlag, Berlin and Chemical Abstracts, all of which provide references to publications describing the structures, properties and preparation of such nucleobases.

[0056] The macromolecule can be prepared according to any technique apparent to those of skill in the art. Advantageously, macromolecules according to the invention can comprise labels and/or members of binding pairs, as described in the sections below, that can be used to facilitate preparation and/or purification of the macromolecule. In addition, certain macromolecules of the invention are capable of forming complexes with molecules that comprise members of binding pairs, as described below. These complexes can be used to facilitate preparation and/or purification of the macromolecule or complex.

5.2.2 Immobilization of First Portion

[0057] In the methods of the invention, a first portion of the macromolecule is immobilized. Generally, the first portion is immobilized if it would be recognized as immobilized by one of skill in the art. The first portion can be immobilized by any technique apparent to those of skill in the art. In certain embodiments, the technique for immobilization of the first portion of the macromolecule is not critical for the methods of the invention.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 677

[0058] The first portion of the macromolecule can be at any location in the macromolecule. In certain embodiments, the first portion is at a terminus of the macromolecule. For the purposes of the invention, a portion of a macromolecule can be "at a terminus" when it is less than five, four, three, two, one or zero monomers from a terminus of the macromolecule. Of course, although many macromolecules have two termini, the methods of the invention are applicable to macromolecules have more than two termini and to macromolecules having one or zero termini, *e.g.* circular macromolecules. In certain embodiments, the first portion is not at a terminus of the macromolecule.

[0059] The macromolecule can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any moiety to which the macromolecule can be immobilized without limitation. In certain embodiments, the substrate is a surface, membrane, bead, porous material, electrode or array.

[0060] In certain embodiments, the first portion of the macromolecule can be immobilized non-selectively. In further embodiments, the first portion of the macromolecule can be immobilized selectively. In advantageous embodiments, after the first portion of the macromolecule is immobilized, some portion of the macromolecule should be free to move sufficiently so that the macromolecule can be extended and/or oriented in following steps of the method. In particular, in certain embodiments, when the first portion of the macromolecule is immobilized non-selectively, it is important that the entire macromolecule not be immobilized non-selectively to an extent that prevents extension of any portion of the macromolecule.

[0061] The immobilization can be by any interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain embodiments, the immobilization can be via electrostatic interaction with an electrode. In further embodiments, the immobilization is via electrostatic interaction with a substrate other than the electrode.

[0062] In certain embodiments, the first portion of the macromolecule comprises a first member of a binding pair. The first member of the binding pair can be covalently bound to the first portion of the macromolecule, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. In useful embodiments, the substrate onto which the first portion of the macromolecule is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound. FIG. 1A illustrates a macromolecule

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 678

that comprises a moiety F1 that is capable of selectively binding a moiety of the substrate. Moiety F1 can be, for example, biotin, capable of binding, for example, a substrate coated with avidin.

[0063] In certain embodiments, the first portion of the macromolecule can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins, and antibodies. The first portion of the macromolecule would comprise a binding moiety capable of binding with the binding moiety of the substrate.

[0064] In advantageous embodiments, the first portion of the macromolecule can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the macromolecule can comprise a biotin moiety in its first portion. For instance, a polynucleotide macromolecule can comprise a biotinylated nucleotide residue. Similarly, a polypeptide macromolecule can comprise a biotinylated amino acid residue. The substrate comprising avidin can be any substrate comprising avidin known to those of skill in the art. Useful substrates comprising avidin are commercially available including TB0200 (Accelr8), SAD6, SAD20, SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), streptavidin slide (catalog #MPC 000, Xenopore) and STREPTAVIDINnslide (catalog #439003, Greiner Bio-one).

[0065] In certain embodiments, the first portion of the macromolecule can comprise a nucleotide sequence that is capable of selectively binding a nucleotide sequence on the substrate.

[0066] In further embodiments, the first portion of the macromolecule can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are commercially available including Optiarray-biotin (Accler8), BD6, BD20, BD100, BD500 and BD2000 (Xantec).

[0067] In further embodiments, the first portion of the macromolecule is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, a first portion of the macromolecule can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. FIG. 2A illustrates a macromolecule that is capable of selectively binding a second molecule X that is capable of selectively binding a third molecule that

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 679

comprises F1. F1 is capable of selectively binding a moiety on a substrate. FIG. 2B illustrates a macromolecule that is capable of selectively binding a second molecule that comprises F1, and F1 is capable of selectively binding a moiety on a substrate.

[0068] In further embodiments, the first portion of the macromolecule can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Useful substrates comprising reactive groups include those that comprise a reactive moiety selected from the group consisting of succinamides, amines, aldehydes, epoxies and thiols. Exemplary useful substrates comprising reactive moieties include, but are not limited to, surfaces comprising epoxy, aldehyde, gold, hydrazide, sulfhydryl, NHS-ester, amine, thiol, carboxylate, maleimide, hydroxymethyl phosphine, imidoester, isocyanate, hydroxyl, pentafluorophenyl-ester, psoralen, pyridyl disulfide or vinyl sulfone, or mixtures thereof. Such surfaces can be obtained from commercial sources or prepared according to standard techniques. The first portion of the macromolecule would comprise a reactive moiety capable of reacting with the reactive moiety of the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, OptArray-DNA NHS group (Accler8), Nexterion Slide AL (Schott) and Nexterion Slide E (Schott).

[0069] In certain embodiments, the first portion of the macromolecule can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the first portion of the macromolecule could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

[0070] In further embodiments, the first portion of the macromolecule can be immobilized to the substrate via other binding pairs apparent to those of skill in the art.

5.2.3 Extension of the Macromolecule

[0071] In certain methods of the invention, the macromolecule is in an extended state. Generally, any macromolecule is in an extended state if it would be recognized as such by one of skill in the art.

[0072] In certain embodiments, the macromolecule is in an extended state when it is in the field of a force capable of extending the macromolecule under conditions suitable for extending the macromolecule. Such forces and conditions should be apparent to those of skill in the art. For instance, many macromolecules can be extended by hydrodynamic force

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 680

or by gravity, and many charged macromolecules can be extended by electromagnetic force. In certain embodiments, the force can be applied to the macromolecule indirectly. For instance, the macromolecule can comprise or can be linked, covalently or noncovalently, to a moiety capable of being moved by a force. In certain embodiments, the macromolecule can be linked to a moiety capable of being moved by electromagnetic, hydrodynamic or optical force.

[0073] In certain embodiments, the force is an electromagnetic force. For instance, when the macromolecule is charged, such as a polynucleotide, the macromolecule can be extended in an electric or magnetic field. The field should be strong enough to extend the macromolecule according to the judgment of one of skill in the art. Exemplary techniques for extending a macromolecule in an electric or magnetic field are described in Matsuura *et al.*, 2002, *J Biomol Struct Dyn.* 20(3):429-36; Ferree & Blanch, 2003, *Biophys J.* 85(4):2539-46; Stigter & Bustamante, 1998, *Biophys J.* 1998 75(3):1197-210; Matsuura *et al.*, 2001, *Nucleic Acids Res.* 29(16); Ferree & Blanch, 2004, *Biophys J.* 87(1):468-75; the contents of which are hereby incorporated by reference in their entirety.

[0074] In certain embodiments, the force is a hydrodynamic force. For instance, many macromolecules, including polysaccharides, polypeptides, and polynucleotides, can be extended in the field of a moving fluid. The hydrodynamic force should be strong enough to extend the macromolecule according to the judgment of one of skill in the art. Exemplary techniques for extending a macromolecule in hydrodynamic field are described in Bensimon *et al.*, 1994, *Science* 265:2096-2098; Henegariu *et al.*, 2001, *BioTechniques* 31: 246-250; Kraus *et al.*, 1997, *Human Genetics* 99:374 - 380; Michalet *et al.*, 1997, *Science* 277:1518-1523; Yokota *et al.*, 1997, *Nucleic Acids Res.* 25(5):1064-70; Otobe *et al.*, 2001, *Nucleic Acids Research* 29:109; Zimmerman & Cox, 1994, *Nucleic Acids Res.* 22(3):492-7, and U.S. Patent Nos. 6,548,255, 6,344,319, 6,303,296, 6,265,153, 6,225,055, 6,054,327, 5,840,862, the contents of which are hereby incorporated by reference in their entirety.

[0075] In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force to extend the macromolecule. In certain embodiments, The force should be strong enough to extend the macromolecule according to the judgment of one of skill in the art. Exemplary techniques for extending a macromolecule with gravity are described in Michalet *et al.*, 1997, *Science* 277:1518-1523; Yokota *et al.*, 1997, *Nucleic Acids Res.* 25(5):1064-70; Kraus *et al.*, 1997, *Human Genetics* 99:374 - 380, the contents of which are hereby incorporated by reference in their entirety.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 681

[0076] In particular embodiments, the force is applied through a moving meniscus. Those of skill in the art will recognize that a moving meniscus can apply various forces to a macromolecule including hydrodynamic force, surface tension and/or any other force recognized by those of skill in the art. The meniscus can be moved by any technique apparent to those of skill in the art including evaporation and gravity. Exemplary techniques for extending a macromolecule with a moving meniscus are described in, for example, U.S. Patent Nos. 6,548,255, 6,344,319, 6,303,296, 6,265,153, 6,225,055, 6,054,327, 5,840,862, the contents of which are hereby incorporated by reference in their entireties.

[0077] In particular embodiments, the macromolecule can be extended by an optical trap or optical tweezers. For instance, the macromolecule can comprise or can be linked, covalently or noncovalently, to a particle capable of being trapped or moved by an appropriate source of optical force. Useful techniques for moving particles with optical traps or optical tweezers are described in Ashkin *et al.*, 1986, *Optics Letters* 11:288-290; Ashkin *et al.*, 1987, *Science* 235:1517–1520; Ashkin *et al.*, Nature 330:769–771; Perkins *et al.*, 1994, *Science* 264:822-826; Simmons *et al.*, 1996, *Biophysical Journal* 70:1813-1822; Block *et al.*, 1990, *Nature* 348:348 - 352; and Grier, 2003, *Nature* 424: 810-816; the contents of which are hereby incorporated by reference in their entireties.

[0078] In certain embodiments, the macromolecule can be extended by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain macromolecules are extended by a combination of an electric field and hydrodynamic force.

[0079] The macromolecule is extended when it would be recognized as extended by one of skill in the art according to standard criteria for extension of a macromolecule. In certain embodiments, the macromolecule is extended when it loses most of its tertiary structural features as recognized by those of skill in the art. In certain embodiments, the macromolecule is extended when it loses most of its secondary structural features as recognized by those of skill in the art. In certain embodiments, the macromolecule is extended when its primary structural features are detectable in sequence when imaged according to standard techniques. Exemplary imaging techniques are described in the examples below.

[0080] In certain embodiments, an extended state of a macromolecule can be recognized by comparing its hydrodynamic radius to its average hydrodynamic radius when free in dilute solution. For instance, in certain embodiments, a macromolecule, or portion thereof, is extended when its hydrodynamic radius is more than about double its average

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 682

hydrodynamic radius in dilute solution. More quantitatively, R represents the hydrodynamic radius of the macromolecule, or portion thereof, and <R> represents the average hydrodynamic radius of the macromolecule, or portion thereof, in dilute solution. The average <R> should be calculated such that R for the macromolecule, or portion thereof, when unbound in dilute solution is less than 2<R> 95% of the time. In certain embodiments, a macromolecule, or portion thereof, is in an extended state when R is greater than 1.5<R>, greater than 1.6<R>, greater than 1.7<R>, greater than 1.8<R>, greater than 1.9<R>, greater than 2.0<R>, greater than 2.1<R>, greater than 2.2<R>, greater than 2.3<R>, greater than 2.4<R>, greater than 2.5<R> or greater than 3.0<R>. In particular embodiments, a macromolecule, or portion thereof, is in an extended state when R is greater than 2.0<R>.

5.2.4 Orientation of the Macromolecule

[0081] In certain methods of the invention, the macromolecule is in an oriented state. Generally, any macromolecule is in an oriented state if it would be recognized as such by one of skill in the art.

[0082] In certain embodiments, the macromolecule is in an oriented state when it is in the field of a force capable of orienting the macromolecule under conditions suitable for orienting the macromolecule. Such forces and conditions should be apparent to those of skill in the art.

[0083] In certain embodiments, the force is an electromagnetic force. For instance, when the macromolecule is charged, such as a polynucleotide, the macromolecule can be oriented in an electric or magnetic field. The field should be strong enough to orient the macromolecule according to the judgment of one of skill in the art. Exemplary techniques for orienting a macromolecule in an electric or magnetic field are described above.

[0084] In certain embodiments, the force is a hydrodynamic force. For instance, many macromolecules, including polysaccharides, polypeptides, and polynucleotides, can be oriented in the field of a moving fluid. The hydrodynamic force should be strong enough to orient the macromolecule according to the judgment of one of skill in the art. Exemplary techniques for orienting a macromolecule in hydrodynamic field are described above.

[0085] In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force or surface tension to orient the macromolecule. In certain embodiments, The force should be strong enough to orient the macromolecule according to the judgment of one of skill in the art. Exemplary techniques for orienting a macromolecule with gravity are described above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 683

[0086] In certain embodiments, the force in an optical force. For instance, the macromolecule can comprise or can be linked, covalently or noncovalently, to a particle capable of being trapped or moved by an appropriate source of optical force as described above.

[0087] In certain embodiments, the macromolecule can be oriented by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain macromolecules are oriented by a combination of an electric field and hydrodynamic force.

[0088] The macromolecule is oriented when it would be recognized as oriented by one of skill in the art according to standard criteria for orientation of a macromolecule. In certain embodiments, the macromolecule is oriented when it is arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the macromolecule. In certain embodiments, the macromolecule is oriented when it is one of a plurality of macromolecules that are arranged in parallel, as recognized by those of skill in the art. For instance, a plurality of macromolecules can be oriented when the vector from a first terminus to a second terminus of a macromolecule is parallel, as recognized by those of skill in the art, to the vectors between corresponding termini of other macromolecules in the plurality.

5.2.5 Selective Immobilization of Second Portion of Macromolecule

[0089] As discussed above, in the methods of the invention, a second portion of the macromolecule is selectively immobilized. The second portion of the macromolecule can be any portion of the macromolecule that is not identical to the first portion of the macromolecule. In some embodiments, the second portion of the macromolecule does not overlap any part of the first portion of the macromolecule.

[0090] In certain embodiments, the present invention provides methods that comprise the single step of selectively immobilizing a second portion of a macromolecule while the macromolecule is in an extended or oriented state, and while a first portion of the macromolecule is immobilized. Exemplary methods for immobilization of the first portion of the macromolecule, and for extension or orientation of the macromolecule are described in detail in the sections above.

[0091] In certain embodiments, the present invention provides methods that comprise the step of extending a macromolecule, while a first portion of the macromolecule is immobilized, and the step of selectively immobilizing a second portion of a macromolecule while the macromolecule is in an extended state. Exemplary methods for immobilization of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 684

the first portion of the macromolecule, and for extension of the macromolecule are described in detail in the sections above.

[0092] In certain embodiments, the present invention provides methods that comprise the step of immobilizing a first portion of a macromolecule, the step of extending the macromolecule while the first portion is immobilized and the step of selectively immobilizing a second portion of a macromolecule while the macromolecule is in an extended state. Exemplary methods for immobilization of the first portion of the macromolecule, and for extension of the macromolecule are described in detail above.

[0093] In certain embodiments, the present invention provides methods that comprise the step of orienting a macromolecule, while a first portion of the macromolecule is immobilized, and the step of selectively immobilizing a second portion of a macromolecule while the macromolecule is in an oriented state. Exemplary methods for immobilization of the first portion of the macromolecule, and for orienting the macromolecule are described in detail in the sections above.

[0094] In certain embodiments, the present invention provides methods that comprise the step of immobilizing a first portion of a macromolecule, the step of orienting the macromolecule while the first portion is immobilized and the step of selectively immobilizing a second portion of a macromolecule while the macromolecule is in an oriented state. Exemplary methods for immobilization of the first portion of the macromolecule, and for orienting the macromolecule are described in detail above.

[0095] The selective immobilization of the second portion of the macromolecule can follow any technique for selective immobilization of a macromolecule apparent to those of skill in the art. Significantly, in advantageous embodiments of the invention, the second portion of the macromolecule is not immobilized non-selectively. Selective immobilization can allow the macromolecule to be immobilized while in a fully extended state or nearly fully extended state. Selective immobilization can also allow the macromolecule to be immobilized in an oriented manner. In other words, the first portion and second portion of the macromolecule can be immobilized along the direction of the field or fields used to extend the macromolecule, with the first portion preceding the second portion in the field. When a plurality of macromolecules are immobilized, the can be uniformly oriented along the field.

[0096] The second portion of the macromolecule can be at any location in the macromolecule. In certain embodiments, the second portion is at a terminus of the macromolecule. In certain embodiments, the second portion is not at a terminus of the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 685

macromolecule. In certain embodiments, the first portion, described in the sections above, is at one terminus of the macromolecule, and the second portion is at another terminus of the macromolecule.

[0097] As discussed above, the second portion of the macromolecule is immobilized selectively. The immobilization can be by any selective interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain embodiments, the immobilization can be via electrostatic interaction with an electrode. In further embodiments, the immobilization is via electrostatic interaction with a substrate other than the electrode.

[0098] If the first portion and the second portion of the macromolecule are selectively immobilized to the same substrate, the techniques of selective immobilization should of course be compatible with the substrate. In particular embodiments, the techniques of immobilization are the same. For instance, on a substrate coated with avidin, both the first and second portion of the macromolecule can be immobilized selectively via biotin - avidin interactions. However, as will be apparent to those of skill in the art, the same interaction need not be used at both the first and second portions for immobilization on the same substrate. For instance, the substrate can comprise multiple moieties capable of selective binding, or the first portion can be immobilized non-selectively, or other techniques apparent to those of skill in the art.

[0099] In certain embodiments, the second portion of the macromolecule comprises a first member of a binding pair. The second member of the binding pair can be covalently bound to the second portion of the macromolecule, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. In useful embodiments, the substrate onto which the second portion of the macromolecule is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound.

[00100] In certain embodiments, the second portion of the macromolecule can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins, and antibodies such as those described in the sections above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 686

[00101] In advantageous embodiments, the second portion of the macromolecule can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the macromolecule can comprise a biotin moiety in its first portion. For instance, a polynucleotide macromolecule can comprise a biotinylated nucleotide residue. Similarly, a polypeptide macromolecule can comprise a biotinylated amino acid residue. Useful substrates comprising avidin are described in the sections above.

[00102] In further embodiments, the second portion of the macromolecule can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are described in the sections above.

[00103] In further embodiments, the second portion of the macromolecule can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Exemplary useful substrates comprising reactive groups are described in the sections above.

[00104] In certain embodiments, the second portion of the macromolecule can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the second portion of the macromolecule could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

[00105] In further embodiments, the second portion of the macromolecule can be immobilized to the substrate via other binding pairs described in the sections above.

[00106] In further embodiments, the second portion of the macromolecule is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, the second portion of the macromolecule can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. FIG. 1B illustrates a macromolecule of selectively binding a second molecule that comprises F3 that is, in turn, capable of selectively binding a moiety on a substrate. The interaction between the second portion of the macromolecule and the molecule that comprises F3 can be mediated, for example, by an antigen-antibody interaction.

[00107] FIGS. 3A and 3B illustrate the selective immobilization of a macromolecule according to methods of the present invention. In FIG. 3A, a first portion of the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 687

macromolecule comprises binding moiety F1 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F1 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. The macromolecule of FIG. 3A is extended by a force as described in the sections above. In FIG. 3B, the force is an electrical potential. While extended, the macromolecule is contacted with molecules comprising binding moiety F2 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F2 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. Significantly, up to three molecules comprising F2 are capable of selectively binding a second portion of the macromolecule to selectively immobilize it in its extended state. As illustrated, the molecules comprise a second binding moiety that selectively binds a repeated binding moiety of the macromolecule. The binding moieties can be, for instance, complementary nucleic acid sequences, as illustrated in FIG. 3B. The resulting macromolecule is selectively immobilized in an extended state and should remain extended even when the force is removed. The selectively immobilized, extended macromolecule can be used for any purpose apparent to those of skill in the art.

5.2.6 Immobilization of Two Portions of an Extended or Oriented Macromolecule

[00108] In certain embodiments, the present invention provides methods for selective immobilization of a first portion and a second portion of a macromolecule that is in an extended or oriented state. Significantly, according to these methods of the invention, the macromolecule need not be immobilized prior to application of a force capable of extending or orienting the macromolecule.

[00109] In these methods, the macromolecule is extended or oriented, or both, by a force capable of extending or orienting the macromolecule. Such forces are described in detail in the sections above. In particular embodiments, the force is a force capable of extending or orienting the macromolecule while maintaining the macromolecule in one location, *i.e.* a force capable of extending or orienting without substantially moving the macromolecule. Exemplary forces include oscillating electromagnetic fields and oscillating hydrodynamic fields. In a particular embodiment, the force is an oscillating electrical field. Exemplary techniques for extending or orienting a macromolecule in an oscillating electric field are described in Asbury *et al.*, 2002, *Electrophoresis* 23(16):2658-66; Kabata *et al.*, 1993, *Science* 262(5139):1561-3; and Asbury and van den Engh, 1998, *Biophys J.* 74:1024-30, the contents of which are hereby incorporated by reference in their entirety.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 688

[00110] In the methods, the macromolecule is immobilized at a first portion and at a second portion while extended or oriented. Both the first portion and the second portion can be immobilized non-selectively, both can be immobilized selectively, or one can be immobilized selectively and the other non-selectively. Techniques for immobilization of the first portion and second portion are described in detail in the sections above.

5.2.7 Substrate for Immobilization

[00111] In the methods of the invention, the substrate for immobilization can be any substrate capable of selectively binding the macromolecule apparent to those of skill in the art. Further, in certain aspects, the present invention provides compositions comprising a selectively immobilized macromolecule in an extended state. The compositions comprise a substrate, as described herein, having immobilized thereto a macromolecule in an extended state. The macromolecule can be, of course, immobilized according to a method of the invention.

[00112] The only requirement of the substrate is that it be capable of selectively binding the second portion of the macromolecule as described above. Thus, the substrate can be a filter or a membrane, such as a nitrocellulose or nylon, glass, a polymer such as polyacrylamide, a gel such as agarose, dextran, cellulose, polystyrene, latex, or any other material known to those of skill in the art to which capture compounds can be immobilized. The substrate can be composed of a porous material such as acrylic, styrene methyl methacrylate copolymer and ethylene/acrylic acid.

[00113] The substrate can take on any form so long as the form does not prevent selective immobilization of the second portion of the macromolecule. For instance, the substrate can have the form of a disk, slab, strip, bead, submicron particle, coated magnetic bead, gel pad, microtiter well, slide, membrane, frit or other form known to those of skill in the art. The substrate is optionally disposed within a housing, such as a chromatography column, spin column, syringe barrel, pipette, pipette tip, 96 or 384 well plate, microchannel, capillary, etc., that aids the flow of liquid over or through the substrate.

[00114] The macromolecule can be immobilized on a single substrate or on a plurality of substrates. For instance, in certain embodiments, the first and second portions of macromolecule are immobilized on the same substrate, as recognized by those of skill in the art. In certain embodiments, the first portion of the macromolecule can be immobilized on a first substrate while the second portion of the macromolecule can be immobilized on a second substrate, distinct from the first.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 689

[00115] The substrate can be prepared according to any method apparent to those of skill in the art. For a review of the myriad techniques that can be used to activate exemplary substrates of the invention with a sufficient density of reactive groups, see, the *Wiley Encyclopedia of Packaging Technology*, 2d Ed., Brody & Marsh, Ed., "Surface Treatment," pp. 867 874, John Wiley & Sons (1997), and the references cited therein. Chemical methods suitable for generating amino groups on silicon oxide substrates are described in Atkinson & Smith, "Solid Phase Synthesis of Oligodeoxyribonucleotides by the Phosphite Triester Method," In: *Oligonucleotide Synthesis: A Practical Approach*, M J Gait, Ed., 1984, IRL Press, Oxford, particularly at pp. 45 49 (and the references cited therein); chemical methods suitable for generating hydroxyl groups on silicon oxide substrates are described in Pease *et al.*, 1994, *Proc. Natl. Acad. Sci. USA 91*:5022 5026 (and the references cited therein); chemical methods for generating functional groups on polymers such as polystyrene, polyamides and grafted polystyrenes are described in Lloyd Williams *et al.*, 1997, Chemical Approaches to the Synthesis of Peptides and Proteins, Chapter 2, CRC Press, Boca Raton, FL (and the references cited therein).

[00116] Exemplary useful substrates include surfaces coated with streptavidin, *e.g.* Accelr8 TB0200. Further useful substrates include surfaces coated with N-hydroxysuccinamide that are capable of reacting with a portion of a macromolecule that comprises an amine. One such surface is OptArray-DNA (Accelr8). Additional useful surfaces are coated with aldehyde (*e.g.* Nexterion Slide AL, Schott) and surfaces coated with epoxy (*e.g.* Nexterion Slide E, Schott). Another useful surface is a biotinylated BSA coated surface useful for selective immobilization of a portion of a macromolecule that comprises avidin or streptavidin.

5.3 Methods of Using Selectively Immobilized, Extended or Oriented Macromolecules

[00117] The selectively immobilized, extended and/or oriented macromolecules can be used for any purpose apparent to those of skill in the art. For instance, the selectively immobilized, extended and/or oriented macromolecules are useful for mapping, nanoassembly and surface plasmon resonance.

[00118] In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for macromolecular with a variety of techniques, *e.g.*, atomic force microscopy or electron microscopy.

[00119] In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for macromolecular mapping. For instance, they can be used to

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 690

determine the location of binding or hybridization along a macromolecule by, for example, fluorescent molecules or DNA binding proteins.

[00120] In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for nanoassembly. For instance, they can be used to faciliate crystal growth on extended and/or oriented macromolecules, or crystal growth on polypeptides linked or bound to extended and/or oriented macromolecules. In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for the construction of nanopaths. In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for directed transport using molecular motors, such as kinesin or myosin. In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for molecular computing or for the assembly of circuits comprising macromolecules, *i.e.* DNA computing. In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used to manipulate carbon nanotubes.

[00121] In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for the study of polynucleotide binding proteins. They can be used, for instance, to determine the presence or location of protein bound to a polynucleotide. Useful techniques include surface plasmon resonance.

[00122] In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for the study of protein fibers, such as amyloid, titin, and fibronectin.

[00123] In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used to create macromolecular barcodes for the purposes of separation and sequential detection of labels. These labels spaced along the molecule provide a unique code that can be read when the macromolecule is immobilized and extended and/or oriented. Extension and/or orientation with selective immobilization can facilitate the decoding of the macromolecular barcode.

[00124] The selectively immobilized, extended and/or oriented macromolecules can further be used for can be used in any context where detection or imaging of a macromolecule might be useful. They can be used for diagnostic, prognostic therapeutic and screening purposes. For instance, they can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease. They can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 691

and/or quantity of markers of bacterium or virus, respectively, in the sample. The compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state. In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient.

5.4 Kits Comprising Selectively Immobilized, Extended or Oriented Macromolecules

[00125] The invention further provides kits comprising one or more components of the invention. The kits can comprise, for example, a substrate according to the invention and one or more extended and/or oriented, or both, macromolecules selectively immobilized on the substrate. The kits can be used for any purpose apparent to those of skill in the art, including, those described above.

[00126] In certain embodiments, the present invention also provides kits useful for the extension and/or orientation and selective immobilization of macromolecules. The kits can comprise a substrate for immobilization and one or more binding partners to facilitate extension and/or orientation or immobilization of a macromolecule. The binding partners could, in certain embodiments, comprise a moiety useful for extension and/or orientation of the macromolecule in an appropriate force. In certain embodiments, the binding partners could facilitate immobilization or selective immobilization of the macromolecule to the surface. In further embodiments, the kit could comprise a macromolecule for extension and/or orientation and immobilization. In further embodiments, the kit could comprise a device capable of extending the macromolecule.

[00127] In certain embodiments, the present invention provides kits comprising a container and one or more components of the kits described above.

[00128] The following examples are offered to illustrate this invention and are not to be construed in any way as limiting the scope of this invention.

6. Examples

6.1 Example 1: Selective Immobilization of Extended DNA

[00129] A double stranded RNA-DNA hybrid 7.2 Kb in length is functionalized at one terminus with biotin. At the other terminus, the DNA comprises a single stranded sequence of 15 bases repeated 4 times (5'-GTC TAT CAT CAC AGC GTC TAT CAT CAC AGC

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 692

GTC TAT CAT CAC AGC GTC TAT CAT CAC AGC-3'; SEQ ID NO:1). Thus, the DNA comprises four binding sites at one terminus for selective immobilization. The hybrid also has 4 regions with Cy3 fluorophores incorporated into the RNA.

[00130] A small sample of the DNA (3 μL, 0.01 fmol/μL in 1X TAE, or 40 mM Tris acetate, 1 mM EDTA, pH 8.0) is transferred into a microfluidic device comprising a channel molded into polydimethylsiloxane that is passively adhered to a streptavadin coated coverslip (Accelr8, TB0200). The channel dimensions are 50 μm x 1 mm x 10 mm. *See* FIG. 4A. The sample is contacted with the coverslip at room temperature for 15 minutes allowing the DNA to selectively bind the streptavadin surface via the biotin at the terminus of the DNA. Unbound DNA is washed away by fluid flow. The 1X TAE buffer in the wells are exchanged for fresh buffer and fluid levels are evened out at 30 μL each well. *See* FIG. 4A.

[00131] An electric field of 200 V/cm is applied to extend the long negatively charged DNA (*see* FIG. 4B) toward the positive electrode.

[00132] An immobilization agent, a biotinylated oligonucleotide (5'-\5Biotin\GCTGTGATGATAGAC-3' (SEQ ID NO:2), 50 μL @ 100 nM, 1X TAE) complementary to the second terminus of the DNA, is added to the negative well. The additional volume raises the fluid level in the well and induces hydrostatic flow to introduce the immobilization reagent into the channel (*see* FIG. 4C). The flow also acts to further stretch the DNA in addition to the electric field.

[00133] The biotinylated oligonucleotide hybridizes with the second terminus of the DNA while it is extended and selectively binds the streptavidin of the coverslip. The sample can be effectively immobilized in an extended state in less than 5 minutes.

6.2 Example 2: Imaging of Selectively Immobilized, Extended Macromolecules

[00134] A macromolecule comprising fluorophore labels and biotin affinity tags is prepared and purified according to Example 3. The macromolecule is bound to a coverslip surface comprising biotin and stretched with an electric field according to Example 3. Finally macromolecule is illuminated with an Arc lamp and imaged with a camera. An exemplary image is provided in FIG. 5. Individual dyes and, significantly, the order of those dyes on individual macromolecules can be detected in the image.

6.3 Example 3: Preparation and Imaging of Selectively Immobilized, Extended Macromolecules

[00135] Herein is a step-by-step example of the construction of a nanoreporter from various components. It can be appreciated that various components can be constructed or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 693

added either at the same time, before or after other components. For example, annealing patch units or flaps to a scaffold can be done simultaneously or one after the other.

6.3.1 Scaffold Production

[00136] Single-stranded circular M13mp18 DNA (USB) is annealed to a 5-fold molar excess of an oligonucleotide complementary to the Bam HI recognition site (Bam Cutter oligo) and cut with Bam HI restriction enzyme to yield a linear single-stranded DNA backbone. An oligonucleotide complementary to the Bam Cutter oligonucleotide (anti-Bam oligonucleotide) is subsequently added in 50-fold excess to sequester free Bam Cutter oligonucleotide and thus prevent recircularization of the M13 during later steps.

[00137] The linear M13 molecule serves as a scaffold onto which RNA patches, or RNA segments, with incorporated fluorophores can be annealed.

6.3.2 PCR to form double-stranded positions on the M13 scaffold

[00138] Ten sets of oligonucleotide primer pairs were designed to create 10 different regions along the M13 scaffold. Each pair contains one primer which has a T7 RNA polymerase promoter at the 5' end. Regions 2-7 are designed to be 900 bases (approximately 300 nm) long, as this is the approximate size of a diffraction-limited spot (the smallest spot that can be achieved with standard optics). Regions 1 and 8 have both long and short versions: the long versions cover the whole 900-base region, while the short versions cover only a portion of the 900-base region to allow a target-specific sequence to be ligated. Thus a target-specific sequence can be attached to either end. The ends can also be used for attachment of anchors or tags.

[00139] PCR is performed using Taq polymerase and 0.5ng of double-stranded M13mp18 (USB) as a template. Reactions are cleaned up using a Qiaquick purification kit from Qiagen. Each PCR reaction yields a double-stranded fragment corresponding to one specific segment as illustrated below. These fragments are used as templates for the *in vitro* transcription of the RNA segments.

6.3.3 *In vitro* Transcription to Produce Dark RNA Segments

[00140] Using the PCR products described above as double-stranded templates, RNA segments are generated using an *in vitro* transcription kit from Ambion (Megascript T7 kit). The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 694

6.3.4 *In vitro* Transcription to Produce RNA Segments Modified With Aminoallyl Groups

[00141] Using the PCR products described above as double-stranded templates, RNA segments for later dye-coupling are generated using an *in vitro* transcription kit from Ambion (MessageAmp aRNA kit). Aminoallyl-modified UTP nucleotides are incorporated into the RNA segments during transcription. The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

6.3.5 Dye Coupling of Aminoallyl RNA Segments To Produce Colored Rna Segments

[00142] 20-100μg of aminoallyl-modified RNA segment is coupled with NHS-ester dyes using Ambion Aminoallyl Labeling Kit. Dyes used include Alexa 488, Alexa 594 and Alexa 647 (Invitrogen/Molecular Probes) as well as Cy3 (Amersham).

[00143] Each segment is made separately in 4 colors so that each position on the scaffold can be filled with a segment in any of the four colors; thus different colors can be added at different positions to create many unique color combinations.

[00144] In this particular embodiment, adjacent segments must be of different colors or there may be dark segments interspersed so that each segment is detected as an individual 'spot'. Dark segments may be used as part of the nanoreporter code.

6.3.6 Assembly of the Label Molecule

[00145] Segments for each position are annealed in a 2:1 ratio of segment to M13 scaffold in 1X SSPE buffer at 70°C for 2 hours.

[00146] An assembled nanoreporter with labeled RNA segments is depicted in FIG. 6A-6B. FIG. 6A depicts a nanoreporter in which only alternate "spots" (1, 3, 5 and 7) are labeled, and FIG. 6B depicts a nanoreporter in which every spot is labeled.

6.3.7 Synthesis Of Probe And Target Oligonucleotides

[00147] S2 DNA target oligonucleotide was synthesized and purified by polyacrylamide gel electrophoresis (Integrated DNA Technologies). S2 RNA target molecules were generated by *in vitro* transcription of PCR products corresponding to region of cloned SARS coronavirus gene (Invitrogen) using an Ambion Megascript™ kit per manufacturer's instructions. The S2 ghost probe (FIG. 8A (i) was complementary to a specific 50-base region of the S2 target sequence (S2-*a*) and was synthesized with a biotin-TEG monomer at the 5' end and purified by high performance liquid chromatograpy (Integrated DNA Technologies).A second oligonucleotide with 50bps complementary to the S2 target (S2-*b*) plus 9 bp of a additional sequence used for ligation to the M13 scaffold

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 695

(59bp total) was synthesized and purified by HPLC (Integrated DNA Technologies). Note that S2-*a* and S2-*b* target regions were not overlapping.

6.3.8 Nanoreporter Synthesis

[00148] Oligonucleotide S2-*b* was ligated to the 5' end of linearized M13 [FIG. 8A (iii)], and the resulting product was purified away from residual unligated oligonucleotide by size-exclusion filtration through a YM100 filter (Millipore) per manufacturer's instructions. Amino-allyl-modified RNA segments complementary to M13 is positions 2, 4, 6, and 8 (FIG. 7A) were generated from *in vitro*-transcription of DNA templates (PCR products) via the Ambion Megascript™ kit per manufacturer's instructions. The segments were then coupled to NHS-ester-modified Alexa 647 dye (Invitrogen) per Ambion's instructions (amino allyl MessageAmp™ II aRNA kit). RNA segments corresponding to positions 1, 3, 5, and 7 of the M13 scaffold (FIG. 7C) were generated as unmodified *in vitro*-transcribed RNAs from DNA templates as described above. Assembly of the nanoreporter was carried out by annealing 10 fmol/µl of each of the eight segments to 5 fmol/µl of the M13-S1-*b* scaffold for 2 hours at 70°C in 1X SSPE buffer (150 mM sodium chloride, 10 mM sodium phosphate, 1mM EDTA). The final product was a nanoreporter with 4 segments labeled with A647 (red) interspersed with dark segments.

6.3.9 Hybridization Conditions

[00149] Hybridization of nanoreporters and ghost probes to target were carried out under the following conditions: 5X SSPE (750 mM sodium chloride, 50 mM sodium phosphate, 5 mM disodium EDTA), 40pM ghost probe (attachment oligonucleotide S2-*a*), 40pM Nanoreporter S2-*b*, 100 ng/µl sheared salmon sperm DNA, 5X Denhardt's solution and 0.1% Tween. Final target concentrations were 20pM S2 DNA target (FIG. 8B) and 1pM S2 RNA target (FIG. 8C). No target was added to the negative control (FIG. 8D). The hybridization reaction was incubated at 65°C for at least 16h.

[00150] Hybridization reactions were diluted 1:2 with 100mM Borate buffer solution (pH 9.8) and introduced into a flow cell channel and bound to a streptavidin-coated coverslip forming the bottom of the channel (Streptavidin-OptiChem coverslips from Accelr8). Attachment to the slide by one end of the nanorerporter/target/ghost probe complex was achieved via interaction of the biotinylated ghost probe with the streptavidin surface. After rinsing the channel with additional borate buffer to remove excess reporters not bound to the surface, the buffer was exchanged with 1X TAE (40 mM Tris-acetate, 1 mM EDTA) and a current of 200V was applied to stretch out the nanoreporter/target complexes during image capture.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 696

6.3.10 Surface Attachment

[00151] Once the nanoreporters are attached to both target molecule and corresponding labeled nucleic acids, *i.e.*, nucleic acids attached to label monomers, they are attached to a surface and stretched in resolve the order of signals emitted by the label monomers and thus identify the target molecule. In this example, the nanoreporters are stretched to spatially resolve their fluorescent dye codes which correspond to a particular target molecule. The nanoreporters are stretched by attaching one end to a surface (in this example – a coverslip, see preparations below). Two methods for surface attachment may be used: A) streptavidin coated slides from Accelr8 Corporation with the nanoreporters being biotinylated and B) biotin coated slides with the nanoreporters having streptavidin. In buffer, the nanoreporters are brought into contact with the active surface and allowed to incubate for a period of time. The reaction is performed in flow cells which were made from PDMS molded in etched silicon wafers to make the channels. Metal tubing is used to core wells at the ends of the channels for buffer and sample insertion. Channel dimensions are 0.5 mm or 1 mm wide and 54 μm high. Once the sample has been loaded into the flow cell lane and incubated, the nanoreporters should be attached. Nanoreporters can be stretched either by applying a voltage or by removing the liquid with a receding meniscus leaving the strings stretched and dry.

6.3.11 Preparation of surface and assembly of device

[00152] The binding surfaces (Accelr8 brand Streptavidin-OptiChem, coated coverslips) are shipped in units of 5 surfaces per slide container, and each container is enclosed with a package of silica dessicant in a foil pouch. The pouches are stored at -20 °C until use.

[00153] To prepare the surface for binding, a pouch is first pulled from the freezer and allowed to come to room temperature over several minutes. If previously unopened, the pouch is then sliced along one edge to form a slit, and the container of surfaces is removed. Upon removal of the required surface, the container is replaced in the pouch with its dessicant, the slit is sealed closed with a strip of packaging tape, and the pouch is replaced in the freezer.

[00154] The surface is then lightly rinsed with a stream of Nanopure water (Barnstead Nanopure Diamond) and soaked for 10 minutes in 0.2μm-filtered 1X PBS in a clean, slotted Coplin jar. After soaking, the surface is dipped in Nanopure water and dried by blowing filtered nitrogen across the surface edge.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 697

[00155] The PDMS device used to mate with the surface and provide localization of the sample is cleaned just before use by applying cellophane tape to the PDMS surface and then peeling away dust or other particles which may have become attached during storage. The binding side of the Accelr8 surface is laid face-up, and the clean PDMS structure is centered, channel side down, on the surface. PDMS adheres readily to coated glass, and no further attachment mechanism is necessary.

6.3.12 Sample Binding and Washing

[00156] The sample is bound to the surface by first applying a 5 μL drop of the sample (currently diluted in 100mM sodium borate buffer, pH 9.8) in one well of the chosen lane. The drop should just touch the point at which the channel joins the well (some sample may wick into the channel at this point). The channel is filled, and binding is equalized throughout the channel, by pulling the droplet through the channel to the opposite well using a very weak vacuum (< 2 kPa). The process is repeated for the other samples in their respective lanes. Excess fluid is then removed from the wells, the wells are taped to reduce evaporation, and the device is incubated at room temperature in the dark for 20 minutes.

[00157] After binding, the tape is removed, and the top well of each lane is filled with 100 μL of the borate buffer described above. About 20 μL of that buffer is pulled through the channels to the other wells using the vacuum, and the process is repeated once. All borate buffer is then removed from all wells, and the top well is filled with 1X TAE, pH 8.3. About 50μL TAE is pulled through the channel, then all TAE is removed and the well is refilled. The process is repeated three times, for a total of about 150 μL of TAE rinse. Finally, all wells are filled with 100 μL 1X TAE.

6.3.13 Electrostretching

[00158] The bottom of the coverslip/PDMS device is spotted with immersion oil and placed on the microscope. Electrodes are inserted into the wells on opposite ends of the first PDMS channel (negative electrode in top well, positive in bottom). The first image of the channel will be taken close to the bottom well; the microscope stage is adjusted so that the area of interest is in focus.

[00159] Voltage (200 V) is then applied across the channel. Voltage is supplied by a DC power supply (Agilent E3630A) and amplified 100X through a amplified by a high voltage amplifier (Matsusada Precision Inc.). After the current is applied, focus is readjusted, and the imaging process begins.

[00160] The electrostretching and imaging process is then repeated with the remaining channels. Image the nanoreporters.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 698

6.3.14 Light source for the fluorescent dyes on the nanoreporter

[00161] In using an arc lamp as a light source, the best fluorophore selection is the brightest types without leading to fluorescent overlap such as Alexa 488, Cy3, and Alexa 594. Weaker fluorescent dyes such as Alexa 647 and Cy5.5 may also be used.

6.3.15 Filters to image the fluorescent dyes on the nanoreporter

[00162] For the selected fluorophores Alexa 488, Cy3, Alexa 594 and Alexa 647 there maybe an overlap between the Cy3 and Alexa 594. However, custom ordering an emission filter with a bandwidth of 572-600 nm minimizes the overlap.

6.3.16 Microscope and objective lens to image the nanoreporters

[00163] The microscope model used is the Nikon Eclipse TE2000E from Nikon Incorporation using the inverted fluorescence imaging station which has 6 filter cassettes that allow the selection of fluorescent emission from multiple fluorescent dye candidates. For the selected dyes, the optical resolution required is about 400 nm for all the wavelengths (500-700 nm). The selected objective lens is the Nikon Plan Apo TIRF lens which has a NA of 1.45 and magnification of 60. The optical resolution is ~210-300 nm for different wavelengths.

[00164] Five minutes before using the microscope (Nikon Eclipse TE2000E), turn on the light source (X-cite 120, Exfo Corporation) and make sure the intensity is the maximum. Turn on the CCD camera driver (Hamamatsu, Orca Ag) and the shutter controller. Use the oil objective of 60 x 1.45NA (Plan Apo TIRF, Nikon) to evaluate the nanoreporters. For all the nanoreporter evaluations the optivar is set at 1x. Open the Metamorph software (Universal Imaging Corporation). Acquire the images using the corresponding filter sets such as cy3, A647 (Chroma Technologies).

[00165] All publications, patents and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 699

What is claimed is:

1. A method for selectively immobilizing a macromolecule in an extended state, wherein a first portion of said extended macromolecule is immobilized, comprising the step of:

selectively immobilizing a second portion of said extended macromolecule such that said macromolecule is immobilized in an extended state.

2. The method of Claim 1 wherein said immobilized macromolecule is extended between the first portion and the second portion.

3. The method of Claim 1 wherein the first portion of said macromolecule is immobilized selectively.

4. The method of Claim 1 further comprising the step of selectively immobilizing the first portion of said extended macromolecule prior to the step of selectively immobilizing the second portion.

5. The method of Claim 1 wherein the first portion and the second portion of said immobilized macromolecule are selectively immobilized on a single surface.

6. The method of Claim 1 wherein said second portion of the macromolecule comprises a first member of a binding pair that is capable of selectively binding a second member of the binding pair to selectively immobilize said second portion of the macromolecule.

7. The method of Claim 6 wherein the first member is non-covalently linked to said macromolecule.

8. The method of Claim 6 wherein the first member is covalently linked to said macromolecule.

9. The method of Claim 6 wherein a surface comprises said second member of the binding pair for selective immobilization of the macromolecule.

10. The method of Claim 6 wherein said first member of the binding pair is selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins and antibodies.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 700

11. The method of Claim 6 wherein said first member of the binding pair is selected from the group consisting of biotin, digoxigenin, FITC, avidin, streptavidin, anti-digoxigenin and anti-FITC.

12. The method of Claim 6 wherein said second member of the binding pair is selected from the group consisting of ligands, antigens, nucleic acids, receptors and antibodies.

13. The method of Claim 6 wherein said second member of the binding pair is selected from the group consisting of biotin, digoxigenin, FITC, avidin, streptavidin, anti-digoxigenin and anti-FITC.

14. The method of Claim 1 wherein said first portion of the macromolecule comprises a first member of a binding pair that is capable of selectively binding a second member of the binding pair to selectively immobilize said first portion of the macromolecule.

15. The method of Claim 14 wherein the first member is non-covalently linked to said macromolecule.

16. The method of Claim 14 wherein the first member is covalently linked to said macromolecule.

17. The method of Claim 14 wherein a surface comprises said second member of the binding pair for selective immobilization of the macromolecule.

18. The method of Claim 14 wherein said first member of the binding pair is selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins and antibodies.

19. The method of Claim 14 wherein said first member of the binding pair is selected from the group consisting of biotin, digoxigenin, FITC, avidin, streptavidin, anti-digoxigenin and anti-FITC.

20. The method of Claim 14 wherein said second member of the binding pair is selected from the group consisting of ligands, antigens, nucleic acids, receptors and antibodies.

21. The method of Claim 14 wherein said second member of the binding pair is selected from the group consisting of biotin, digoxigenin, FITC, avidin, streptavidin, anti-digoxigenin and anti-FITC.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 701

22. The method of Claim 1 wherein the selective immobilization is via one or more non-covalent bonds.

23. The method of Claim 1 wherein the selective immobilization is via one or more covalent bonds.

24. The method of Claim 1 wherein said second portion of the macromolecule comprises a first moiety that is capable of selectively reacting with a second moiety to selectively immobilize said second portion of the macromolecule.

25. The method of Claim 24 wherein a surface comprises said second member of the binding pair for selective immobilization of the macromolecule.

26. The method of Claim 24 wherein said first moiety is selected from the group consisting of succinamides, amines, aldehydes, epoxies and thiols.

27. The method of Claim 24 wherein said second moiety is selected from the group consisting of succinamides, amines, aldehydes, epoxies and thiols.

28. The method of Claim 1 wherein said macromolecule is extended by a force.

29. The method of Claim 28 wherein said force is selected from the group consisting of gravity, electromagnetic force and hydrodynamic force.

30. The method of Claim 28 wherein said force is an electrical potential.

31. The method of Claim 28 wherein said force is gravity.

32. The method of Claim 28 wherein said force is hydrodynamic force.

33. The method of Claim 1 wherein said macromolecule is a polymer.

34. The method of Claim 30 wherein said polymer is a polynucleotide.

35. The method of Claim 34 wherein said polynucleotide is single stranded.

36. The method of Claim 34 wherein said polynucleotide comprises multiple complementary strands.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 702

37. The method of Claim 34 wherein said polynucleotide comprises two complementary strands.

38. The method of Claim 34 wherein said polynucleotide comprises three complementary strands.

39. The method of Claim 34 wherein said polynucleotide is deoxyribonucleic acid or ribonucleic acid.

40. The method of Claim 1 wherein the second portion is a terminus of said macromolecule.

41. The method of Claim 1 wherein the first portion is a terminus of said macromolecule.

42. The method of Claim 1 wherein the first portion and the second portion are termini of said macromolecule.

43. The method of Claim 1 wherein said second portion is not a terminus of said macromolecule.

44. The method of Claim 1 wherein said first portion is not a terminus of said macromolecule.

45. The method of Claim 1 wherein a plurality of portions, in addition to the first portion, of said extended macromolecule are selectively immobilized.

46. The method of Claim 1 further comprising selectively immobilizing a third portion of said extended macromolecule.

47. A method for extending and selectively immobilizing a macromolecule comprising the steps of:

a) applying to the macromolecule, wherein at least a first portion of said macromolecule is immobilized, a force sufficient to extend said macromolecule; and

b) selectively immobilizing a second portion of the extended macromolecule.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 703

48. A method for extending and selectively immobilizing a macromolecule comprising the steps of:

a) immobilizing a first portion of said macromolecule on a surface;

b) applying to said macromolecule a force capable of extending said macromolecule;

c) contacting a second portion of said macromolecule with a molecule, wherein said molecule comprises a first moiety capable of binding the second portion of said macromolecule and a second moiety capable of selectively binding said surface, under conditions wherein said molecule selectively binds the second portion of said macromolecule and selectively binds said surface thereby selectively immobilizing said macromolecule in an extended state.

49. A surface comprising a selectively immobilized, extended macromolecule.

50. The surface of Claim 49 wherein at least two portions of said immobilized, extended macromolecule are selectively immobilized on the surface.

51. The surface of Claim 50 wherein said immobilized, extended macromolecule is substantially extended at least between the two immobilized portions.

52. A surface comprising a selectively immobilized, extended macromolecule prepared according to Claim 1, 47 or 48.

53. A method for selectively immobilizing a macromolecule in an oriented state, wherein a first portion of said extended macromolecule is immobilized, comprising the step of:

selectively immobilizing a second portion of said extended macromolecule such that said macromolecule is immobilized in an oriented state.

54. A method for orienting and selectively immobilizing a macromolecule comprising the steps of:

a) applying to the macromolecule, wherein at least a first portion of said macromolecule is immobilized, a force sufficient to orient said macromolecule; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 704

b) selectively immobilizing a second portion of the oriented macromolecule.

55. A method for orienting and selectively immobilizing a macromolecule comprising the steps of:

a) immobilizing a first portion of said macromolecule on a surface;

b) applying to said macromolecule a force capable of orienting said macromolecule;

c) contacting a second portion of said macromolecule with a molecule, wherein said molecule comprises a first moiety capable of binding the second portion of said macromolecule and a second moiety capable of selectively binding said surface, under conditions wherein said molecule selectively binds the second portion of said macromolecule and selectively binds said surface thereby selectively immobilizing said macromolecule in an oriented state.

56. A surface comprising a selectively immobilized, oriented macromolecule.

57. The surface of Claim 56 wherein at least two portions of said immobilized, oriented macromolecule are selectively immobilized on the surface.

58. The surface of Claim 57 wherein said immobilized, oriented macromolecule is substantially extended at least between the two immobilized portions.

59. A surface comprising a selectively immobilized, extended macromolecule prepared according to Claim 53, 54 or 55.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 705

FIG. 1A

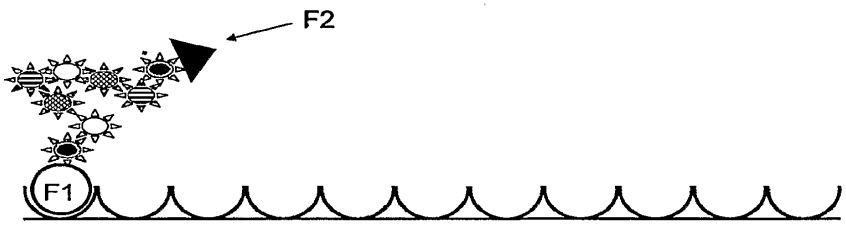
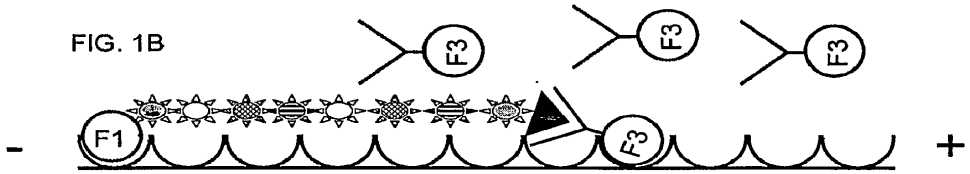


FIG. 1B



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 706

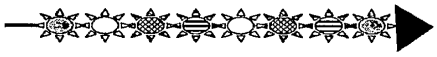
FIG. 2A



FIG. 2B



FIG. 2C



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 707

FIG. 3A

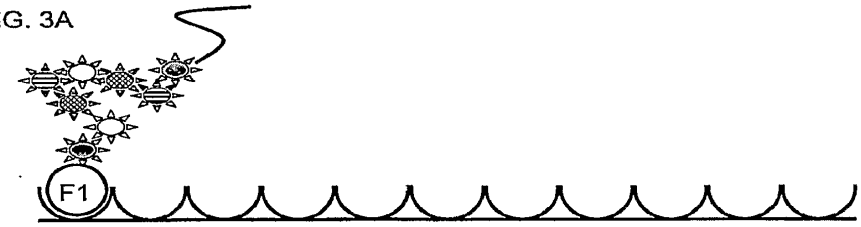


FIG. 3B

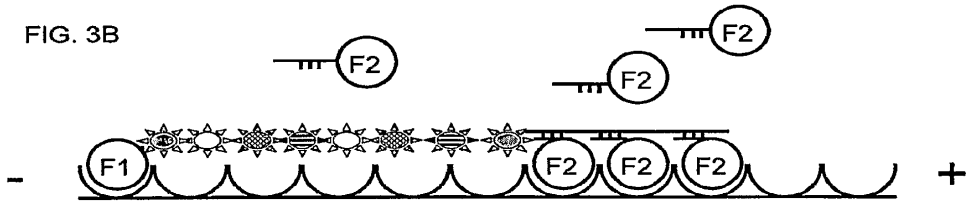


FIG. 3C

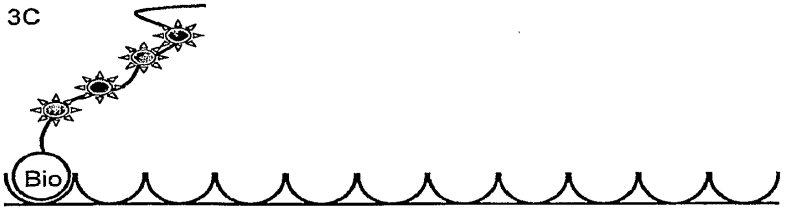
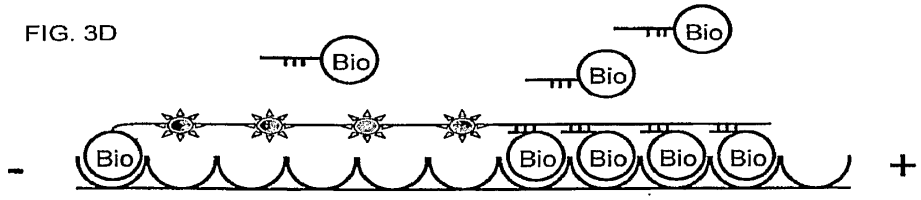


FIG. 3D



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 708

FIG. 4A

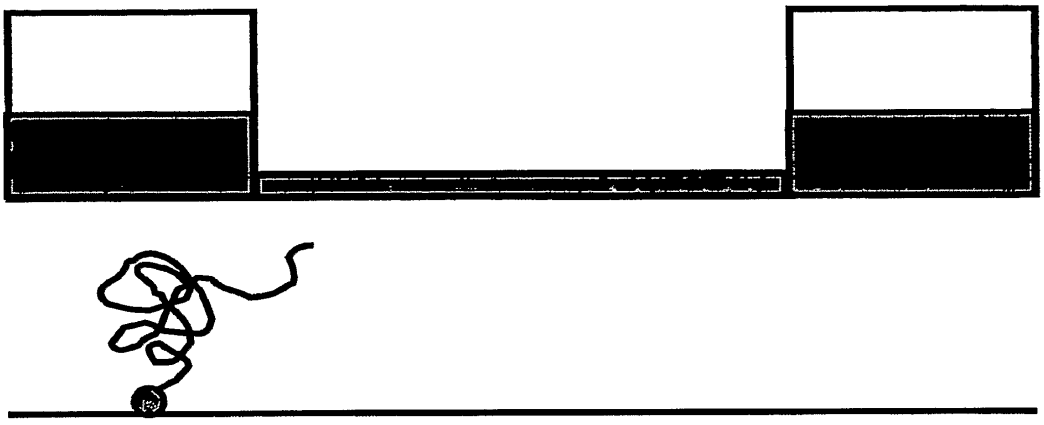
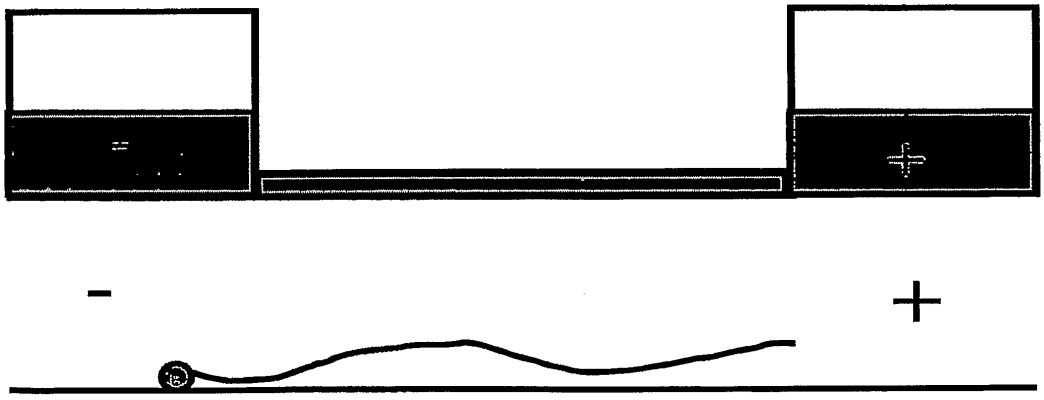
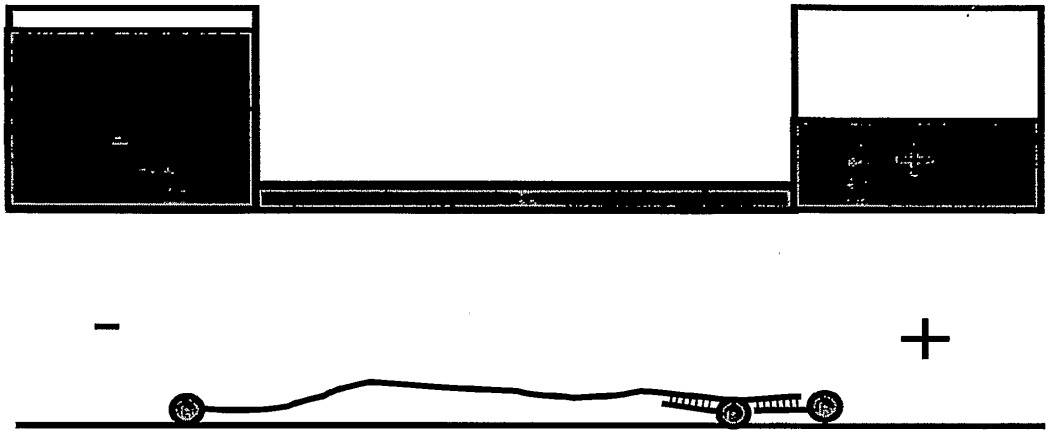


FIG. 4B



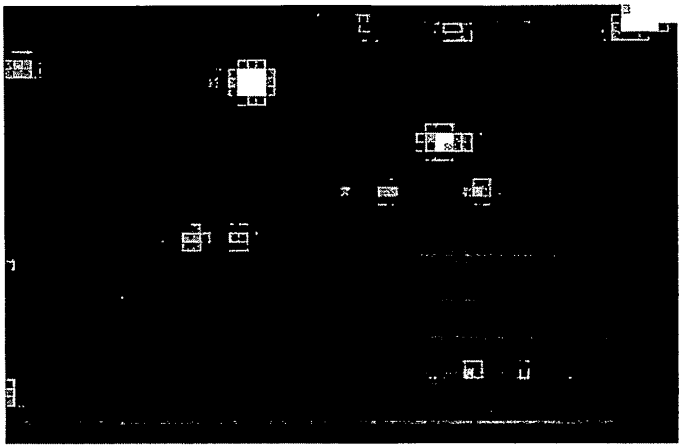
IPR2024-01177
California Institute of Technology - Ex. 2003 Page 709

FIG. 4C



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 710

FIG. 5



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 711

FIG. 6A

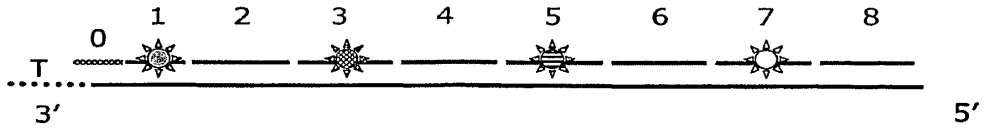


FIG. 6B

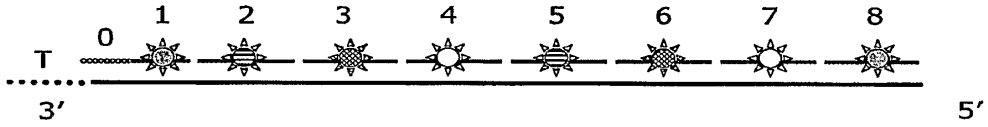


FIG. 7A

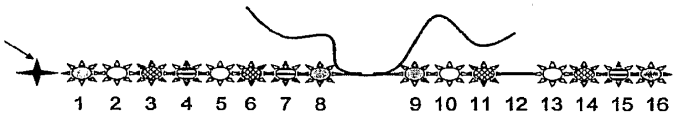


FIG. 7B

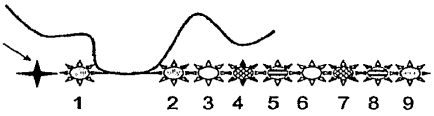
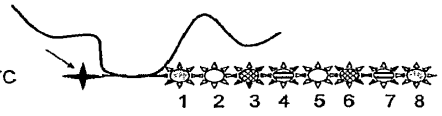


FIG. 7C



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 712

FIG. 8A

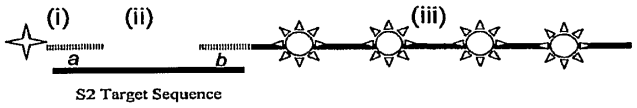


FIG. 8E



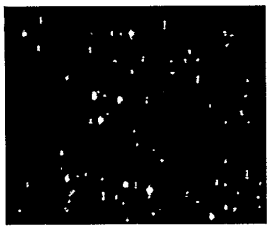


FIG. 8B

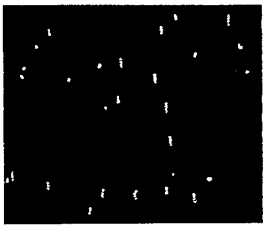


FIG. 8C



FIG. 8D

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 713

SEQUENCE LISTING

<110> NANOSTRING TECHNOLOGIES, INC.

<120> COMPOSITIONS COMPRISING ORIENTED, IMMOBILIZED MACROMOLECULES AND METHODS FOR THEIR PREPARATION

<130> 11612-014-228

<140> to be assigned
<141> on even date herewith

<150> 60/753,816
<151> 2005-12-23

<160> 2

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> a 15-bases repeats 4X functioning as binding sites for selective immobilization

<400> 1
gtctatcatc acagcgtcta tcatcacagc gtctatcatc acagcgtcta tcatcacagc 60

<210> 2
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> biotinylated oligonucleotide as an immobilization agent

<400> 2
gctgtgatga tagac 15

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 714

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 July 2007 (05.07.2007)

PCT

(10) International Publication Number
WO 2007/076132 A3

(51) International Patent Classification:
C12N 11/00 (2006.01) *C12N 11/16* (2006.01)
C12N 11/04 (2006.01)

(21) International Application Number:
PCT/US2006/049279

(22) International Filing Date:
22 December 2006 (22.12.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/753,816 23 December 2005 (23.12.2005) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 201 Elliot Avenue West, Suite 300, Seattle, WA 98119 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **FERREE, Sean, M.** [US/US]; 7018 3rd Avenue, Nw, Seattle, WA 98117 (US). **DUNAWAY, Dwayne, L.** [US/US]; 2751 NE 94th Street, Seattle, WA 98115 (US).

(74) Agents: **ANTLER, Adriane, M.** et al.; JONES DAY, 222 East 41st Street, New York, NY 10017-6702 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

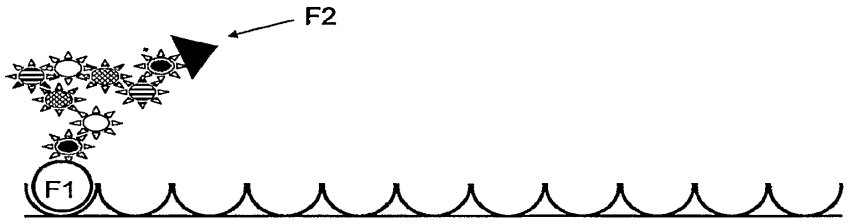
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
7 September 2007

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS COMPRISING ORIENTED, IMMOBILIZED MACROMOLECULES AND METHODS FOR THEIR PREPARATION



(57) Abstract: The present invention provides methods and compositions that facilitate the identification of structures of a variety of macromolecules. For instance, the present invention provides methods for the selective immobilization of macromolecules in an extended or oriented state. The present invention also provides compositions comprising macromolecules selectively immobilized in extended or oriented states.

WO 2007/076132 A3

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 715

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 06/49279

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) -C12N 11/00, 11/04, 11/16 (2007.01)
USPC -435/174
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
USPC -435/174

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC -435/182, 329; 436/43, 518

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(USPT,PGPB,EPAB,JPAB); PubMed; Google Scholar
Search Terms Used: Extend, macromolecule, immobilize, covalent

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2004/0077090 A1 (SHORT) 22April 2004 (22.04.2004) entire document, especially para [0316], [0855]-[0856], [0866]	1-6, 8-14, 16-21, 23-25, 28-59 --- 7, 15, 22, 26-27
Y	US 2004/0023205 A1 (LADNER et al.) 5 February 2004 (05.02.2004) entire document, especially para [0840]-[0841]	7, 15, 22, 26-27

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 May 2007 (24.05.2007)	18 JUL 2007
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2005)

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 December 2007 (06.12.2007)

PCT

(10) International Publication Number
WO 2007/139766 A2

(51) International Patent Classification:
G06F 19/00 (2006.01)

(21) International Application Number:
PCT/US2007/012130

(22) International Filing Date: 21 May 2007 (21.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/802,862 22 May 2006 (22.05.2006) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 210 Elliot Avenue West, Suite 300, Seattle, WA 98119 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HWANG, Jenq-Neng** [US/US]; 15125 S.E. 66th Street, Bellevue, WA 98006 (US). **MITTON, Jeffrey, D.** [US/US]; 4232 134th Avenue SE, Bellevue, WA 98006 (US).

(74) Agents: **ANTLER, Adriane, M.** et al.; Jones Day, 222 East 41st Street, New York, NY 10017-6702 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

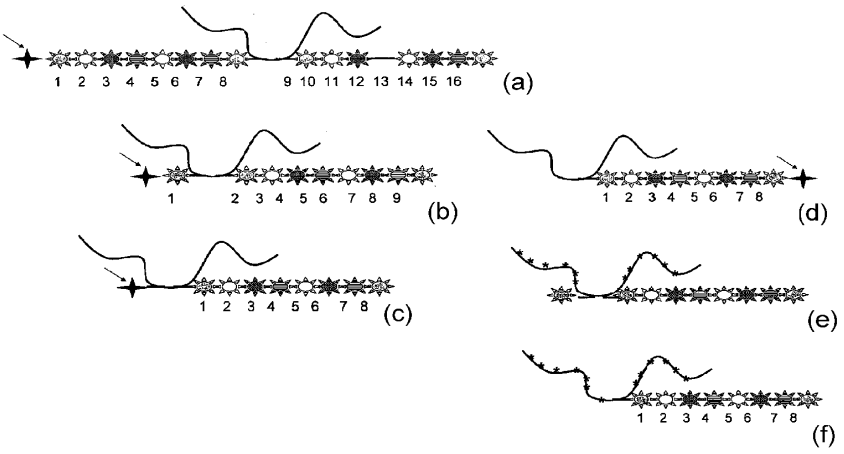
Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** SYSTEMS AND METHODS FOR ANALYZING NANOREPORTERS



(57) **Abstract:** Methods, computers, and computer program products for detecting the presence of a probe within a sample overlayed on a substrate are provided. The probe comprises a plurality of spatially arranged labels. A data storage module stores a plurality of light images, where each light image has light from the sample at a corresponding wavelength range in a plurality of different wavelength ranges. A label identification module identifies a plurality of labels in the plurality of light images that are proximate to each other on the substrate. A spatial order of the plurality of labels determines a string sequence of the plurality of labels. A probe identification module determines whether the string sequence of the plurality of labels comprises a valid reporter sequence.

WO 2007/139766 A2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 717

SYSTEMS AND METHODS FOR ANALYZING NANOREPORTERS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/802,862 filed May 22, 2006, which is incorporated herein by reference in its entirety.

1. FIELD OF THE INVENTION

[0002] The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to coded, labeled reporter molecules, referred to herein as labeled "nanoreporters," that are capable of binding individual target molecules. Through the nanoreporters' label codes, the binding of the nanoreporters to target molecules results in the identification of the target molecules. Methods of making and using such nanoreporters are also provided. The nanoreporters can be used in diagnostic, prognostic, quality control and screening applications.

2. BACKGROUND OF THE INVENTION

[0003] Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression are at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in normal cells compared to diseased cells can aid the discovery of new drug and diagnostic targets.

[0004] Nucleic acids can be detected and quantified based on their specific polynucleotide sequences. The basic principle underlying existing methods of detection and quantification is the hybridization of a labeled complementary probe sequence to a target sequence of interest in a sample. The formation of a duplex indicates the presence of the target sequence in the sample and the degree of duplex formation, as measured by the amount of label incorporated in it, is proportional to the amount of the target sequence.

[0005] This technique, called molecular hybridization, has been a useful tool for identifying and analyzing specific nucleic acid sequences in complex mixtures. This technique has been used in diagnostics, for example, to detect nucleic acid sequences of various microbes in biological samples. In addition, hybridization techniques have been used to map genetic differences or polymorphisms between individuals. Furthermore, these techniques have been

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 718

used to monitor changes in gene expression in different populations of cells or in cells treated with different agents.

[0006] In the past, only a few genes could be detected in a complex sample at one time. Within the past decade, several technologies have made it possible to monitor the expression level of a large number of transcripts within a cell at any one time (*see, e.g.*, Schena *et al.*, 1995, Science 270: 467-470; Lockhart *et al.*, 1996, Nature Biotechnology 14: 1675-1680; Blanchard *et al.*, 1996, Nature Biotechnology 14:1649). In organisms for which most or all of the genome is known, it is possible to analyze the transcripts of large numbers of the genes within the cell. Most of these technologies employ DNA microarrays, devices that consist of thousands of immobilized DNA sequences present on a miniaturized surface that have made this process more efficient. Using a microarray, it is possible in a single experiment to detect the presence or absence of thousands of genes in a biological sample. This allows researchers to simultaneously perform several diagnostic tests on one sample, or to observe expression level changes in thousands of genes in one experiment. Generally, microarrays are prepared by binding DNA sequences to a surface such as a nylon membrane or glass slide at precisely defined locations on a grid. Then nucleic acids in a biological sample are labeled and hybridized to the array. The labeled sample DNA marks the exact position on the array where hybridization occurs, allowing automatic detection.

[0007] Unfortunately, despite the miniaturization of array formats, this method still requires significant amounts of the biological sample. However, in several cases, such as biopsies of diseased tissues or samples of a discrete cell type, the biological sample is in limited supply. In addition, the kinetics of hybridization on the surface of a microarray is less efficient than hybridization in small amounts of aqueous solution. Moreover, while methods exists to estimate the amount of nucleic acid present in a sample based on microarray hybridization result, microarray technology thus far does not allow for detection of target molecules on an individual level, nor are there microarray-based methods for directly quantifying the amount of target molecule in a given sample.

[0008] Thus, there exists a need for accurate and sensitive detection, identification and quantification of target molecules in complex mixtures.

[0009] Discussion or citation of a reference herein shall not be construed as an admission that such reference is prior art to the present invention.

3. SUMMARY OF THE INVENTION

[0010] One aspect of the present invention provides a computer program product comprising a computer readable storage medium and a computer program mechanism

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 719

embedded therein. The computer program mechanism is for detecting the presence of a probe within a sample overlayed on a substrate. The probe comprises a plurality of spatially arranged labels. The computer program mechanism comprises a data storage module, a label identification module, and a probe identification module. The data storage module comprises instructions for storing a plurality of light images. Each light image in the plurality of light images is for light received from the sample at a wavelength range in a plurality of different wavelength ranges. The label identification module comprises instructions for identifying a plurality of labels, in the plurality of light images, that are proximate to each other on the substrate. A spatial order of the plurality of labels determines a string sequence of the plurality of labels. The probe identification module comprises instructions for determining whether the string sequence of the plurality of labels comprises a valid reporter sequence. When the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe. When the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

[0011] In some embodiments, a first label in the plurality of labels is associated with a first position on the substrate that emits light in a first wavelength range in the plurality of different wavelength ranges. Further, a second label in the plurality of labels is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges. In some embodiments, a portion of the first wavelength range overlaps with a portion of the second wavelength range. In some embodiments, no part of the first wavelength range overlaps with any part of the second wavelength range. In some embodiments, each label in the plurality of labels is associated with a position on the substrate that emits more than a threshold amount of light in any one light image in the plurality of light images.

[0012] In some embodiments, the computer program mechanism further comprises a lookup table comprising a plurality of valid reporter sequences. In such embodiments, the probe identification module further comprises instructions for comparing the string sequence of the plurality of labels to valid reporter sequences in the lookup table. In some embodiments, the lookup table is dimensioned and configured to hold up to 4^4, up to 7^4, or up to 20^{20} different valid reporter sequences.

[0013] In some embodiments, the probe identification module further comprises instructions for storing the string sequence of the plurality of labels that is not confirmed as a valid reporter sequence. In some embodiments, the data storage module further comprises instructions for aligning a first light image to a second light image in the plurality of light images using a plurality of fiducials present on the substrate. In some embodiments, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 720

position of the probe on the substrate is random. In some embodiments, the probe consists of a single molecule. In some embodiments, the probe comprises a molecular scaffold and each label in the plurality of labels represents a different position on the molecular scaffold. In some embodiments, each position on the molecular scaffold that is represented by a label is separated from a neighboring position on the scaffold by a spacer. In some embodiments, the probe comprises a single stranded deoxynucleic acid or ribonucleic acid scaffold and each label in the plurality of labels is represented by a dye laden single stranded deoxynucleic acid or ribonucleic acid sequence that hybridizes to a different position on the scaffold.

[0014] In some embodiments, the probe comprises a molecular scaffold having a first end and a second end. A target specific sequence is covalently attached to the first end. A binder sequence is covalently attached to the second end. In such embodiments, the probe is linearly arranged on the substrate through (i) the binding of the target specific sequence to a first molecular entity that is bound to a first position of the substrate and (ii) the binding of the binder sequence to a second molecular entity that is bound to a second position on the substrate. In some embodiments, the first molecular entity is a target (single stranded deoxynucleic acid or ribonucleic acid)-biotin complex and the second molecular entity is a predetermined (single stranded deoxynucleic acid or ribonucleic acid)-biotin complex.

[0015] In some embodiments, the label identification module further comprises instructions for identifying a plurality of candidate labels in the plurality of light images. This plurality of labels is a subset of the plurality of candidate labels that have been validated by the label identification module. In some embodiments, each candidate label in the plurality of candidate labels is associated with a position on the substrate that emits more than a threshold amount of light in any one light image in the plurality of light images. In some embodiments, the plurality of labels comprises a first candidate label that is associated with a first position on the substrate that emits light in a first wavelength range in the plurality of different wavelength ranges and a second candidate label that is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges. In some embodiments, a portion of the first wavelength range overlaps with a portion of the second wavelength range. In some embodiments, the first wavelength range does not overlap with the second wavelength range. In some embodiments, the instructions for identifying the plurality of labels applies a first distance criterion between a centroid of a first candidate label and a centroid of a second candidate label in the plurality of candidate labels. The first distance criterion can be determined by a calculated distance between a first label and a second label in the probe. In some embodiments, the instructions for identifying the plurality of labels applies a second distance criterion between a centroid of the second candidate label and a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 721

centroid of a third candidate label in the plurality of candidate labels. The second distance criterion can be determined by a calculated distance between a second label and a third label in the probe. In some instances, the first distance criterion is the same as the second distance criterion. In other instances, the first distance criterion is different from the second distance criterion. In some embodiments, a value of the first distance criterion and a value of the second distance criterion contribute to determining whether the plurality of labels is the probe.

[0016] In some embodiments, the computer program mechanism further comprises a lookup table comprising a plurality of valid reporter sequences. In some embodiments, each valid reporter sequence in the plurality of valid reporter sequences comprises a first distance between a first pair of labels and a second distance between a second pair of labels. Further, the probe identification module comprises instructions for comparing the string sequence of the plurality of labels, the first distance criterion, and the second distance criterion, to valid reporter sequences in the lookup table.

[0017] In some embodiments, the instructions for identifying the plurality of labels applies an angle criterion to triplets of candidate labels in the plurality of candidate labels. In some embodiments, the instructions for identifying the plurality of labels comprises instructions for applying a model (*e.g.*, linear regression) to select candidate labels in the plurality of candidate labels. In some embodiments, the label identification module further comprises instructions for verifying that a candidate label in the plurality of candidate labels satisfies a spot shape criterion. An example of a spot shape criterion is a match between an observed spot shape of the candidate label and the theoretical point spread of the diffraction limited point source light determined by a magnification of the candidate label. In some embodiments, the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for performing point spread function modeling on the candidate label. In some embodiments, the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for applying a spot segmentation algorithm to the candidate label. In some embodiments, the spot segmentation algorithm comprises a watershed transformation.

[0018] In some embodiments, the instructions for identifying the plurality of labels applies an absolute distance criterion between a centroid of a first terminal candidate label and a centroid of a second terminal candidate label in the plurality of candidate labels. In some embodiments, the instructions for identifying the plurality of labels comprises instructions for identifying a buffer zone around a portion of the substrate that has select candidate labels in the plurality of candidate labels, where there are no candidate labels in the buffer zone. In some embodiments, the plurality of labels are linearly arranged on the substrate (*e.g.*, in the same

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 722

linear orientation). In some embodiments, a linear orientation of each label in the plurality of labels is predetermined (*e.g.*, by an application of an electrical current across the substrate or by an application of a fluid across the substrate, *etc.*). In some embodiments, each respective label in the plurality of labels occupies between 4 and 20 pixels, between 1 and 30 pixels, or between 4 and 100 pixels in a pixilated representation of the respective label in a light image in the plurality of light images.

[0019] In some embodiments, the label identification module further comprises instructions for identifying a first candidate label in the plurality of light images and instructions for identifying a second candidate label in the plurality of light images that is within a predetermined distance away from the first candidate label. In such embodiments, the plurality of labels comprises the first candidate label and the second candidate label.

[0020] In some embodiments, the plurality of different wavelength ranges consists of between two different wavelength ranges and six different wavelength ranges, or between two different wavelength ranges and twenty different wavelength ranges. In some embodiments, the plurality of labels comprises four or five labels. In some embodiments, the plurality of labels consists of between two labels and twenty labels.

[0021] In some embodiments, a first subset of the labels in the string sequence error check an identity of the labels in a second subset of the labels in the string sequence. In some embodiments, a first subset of labels in the string sequence are a checksum for a second subset of labels in the string sequence.

[0022] In some embodiments, the label identification module comprises instructions for repeating the instructions for identifying a plurality of labels a plurality of times. Each time the instructions for identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images, that are proximate to each other on the substrate. The probe identification module determines whether each of the different plurality of labels identified by the label identification module comprises a valid reporter sequence. For each different plurality of labels identified, the probe identification module deems the different plurality of labels to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence and deems the different plurality of labels to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence. In some embodiments, a plurality of probes is identified. In some embodiments, the plurality of probes consists of three or more probes. In some embodiments, the plurality of probes consists of ten or more probes. In some embodiments, the plurality of probes consists of less than 50 probes. In some embodiments, the probe identification module stores each type of probe identified. In some embodiments, the probe identification model

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 723

stores each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence. In some embodiments, the probe identification model stores each string sequence of each different plurality of labels that is confirmed as a valid reporter sequence.

[0023] Another aspect of the invention provides a computer system for detecting the presence of a probe within a sample overlayed on a substrate. The probe comprises a plurality of spatially arranged labels. The computer system comprises a central processing unit and a memory, coupled to the central processing unit. The memory stores a data storage module, a label identification module, and a probe identification module. The data storage module comprises instructions for storing a plurality of light images. Each light image in the plurality of light images is for light received from the sample at a wavelength range in a plurality of different wavelength ranges. The label identification module comprises instructions for identifying a plurality of labels, in the plurality of light images, that are proximate to each other on the substrate. A spatial order of the plurality of labels determines a string sequence of the plurality of labels. The probe identification module comprises instructions for determining whether the string sequence of the plurality of labels comprises a valid reporter sequence. When the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe. When the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

[0024] Still another aspect of the invention comprises a system for detecting the presence of a probe within a sample overlayed on a substrate. The system comprises a light measuring mechanism, a data storage module, a label identification mechanism, and a probe identification mechanism. The light measuring mechanism measures a plurality of light images. Each light image in the plurality of light images is for light received from the sample at a wavelength range in a plurality of different wavelength ranges. The data storage module comprises instructions for storing the plurality of light images. The label identification mechanism identifies a plurality of labels in the plurality of light images that are proximate to each other on the substrate. A spatial order of the plurality of labels determines a string sequence of the plurality of labels. The probe identification mechanism determines whether the string sequence of the plurality of labels comprises a valid reporter sequence. When the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe. When the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe. In some embodiments, the system further comprises an illumination mechanism that illuminates the substrate. In some embodiments, the illumination mechanism comprises an excitation light

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 724

source and a plurality of excitation filters. Each excitation filter in the plurality of excitation filters is used in a corresponding light image in the plurality of light images to confine the light source to a corresponding different spectral range when the corresponding light image is measured. In some embodiments, the light measuring mechanism comprises a plurality of measurement wavelength filters, where each measurement wavelength filter in the plurality of measurement wavelength filters is used in a corresponding light image in the plurality of light images to reject light not within a corresponding spectral range. In some embodiments, the light measuring mechanism comprises a photodetector that forms a detection signal in response to light emitted from the sample.

[0025] In some embodiments, the light measuring mechanism comprises a detector circuit addressed by the detection signal that measures light emitted from the sample overlayed on the substrate. In such embodiments, the light measuring mechanism further comprises an electronic memory for storing a plurality of label positions, where each label position in the plurality of label positions represents a label and each label position in the plurality of label positions originates more than a threshold amount of light. In some embodiments, the label identification mechanism identifies the plurality of labels that are proximate to each other from among the plurality of label positions stored in electronic memory. In some embodiments, the label identification mechanism comprises instructions for identifying a plurality of candidate labels in the plurality of light images and the plurality of labels is a subset of the plurality of candidate labels.

[0026] Yet another aspect of the invention provides methods for detecting the presence of a probe within a sample overlayed on a substrate. In this aspect of the invention, the probe comprises a plurality of spatially arranged labels. In one such method, a plurality of labels, in a plurality of light images, that are proximate to each other on the substrate is identified. The spatial order of the plurality of labels determines a string sequence of the plurality of labels. Each light image in the plurality of light images is for light received from the sample at a wavelength range in a plurality of different wavelength ranges. In the method, a determination is made as to whether the string sequence of the plurality of labels comprises a valid reporter sequence. When the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe. When the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

[0027] In some embodiments, the determining step comprises comparing the string sequence of the plurality of labels to valid reporter sequences in a lookup table. In some embodiments, the method further comprises storing the string sequence of the plurality of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 725

labels that is not confirmed as a valid reporter sequence. For instance, the string sequence can be stored in an electronic memory. In some embodiments, the method further comprises aligning a first light image to a second light image in the plurality of light images using a plurality of fiducials present on the substrate.

[0028] In some embodiments, the step of identifying a plurality of labels is repeated a plurality of times. Each time the step of identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images, that are proximate to each other on the substrate. In some embodiments, the method further comprises determining whether each of the different plurality of labels comprises a valid reporter sequence. Each different plurality of labels is deemed to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence. Furthermore, each different plurality of labels is deemed to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence. In some instances according to this embodiment of the present invention, a plurality of probes is identified. For instance, in some embodiments, two or more probes, three or more probes, ten or more probes, at least 5, 10, 15, 20, 50, 75, 100, 150, 200, 300, or 400 probes or more are identified.

[0029] In some embodiments where a plurality of probes are identified, each type of probe identified is recorded. A probe "type" is identified by the string sequence of the probe. Each unique valid string sequence represents a different probe type. In some embodiments, each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence is stored. In this way, it is possible to determine common conditions that arise on the substrate. One type of condition that can be identified by tracking pluralities of labels that do not form valid string sequences is the condition where there are too many probes on the substrate. When there are too many probes on the substrate, the labels of neighboring probes become proximate to each other, making it difficult to determine which probe each label belongs to. Another type of condition that can be identified is the condition in which an excessive number of probes are not full length. In some embodiments, all species of labels, strings, invalid reporter sequences, valid reporter sequences, probe types is tracked in the methods of the present invention.

4. BRIEF DESCRIPTION OF THE FIGURES

[0030] Figures 1A-1F. Figure 1A illustrates a dual nanoreporter with a 16-position nanoreporter code, using two 8-position nanoreporter components. Figure 1B illustrates a dual nanoreporter with a 9-position nanoreporter code, using one 8-position nanoreporter component and one single-position nanoreporter component. Figure 1C illustrates a dual

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 726

nanoreporter with an 8-position nanoreporter code, using one ghost probe and one 8-position nanoreporter component. Figure 1D illustrates a single nanoreporter with an 8-position nanoreporter code. In Figures 1A-1D, the star shape (depicted with an arrow) is illustrative of an affinity tag, which can be used to purify the nanoreporter or immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging. The numbered region in Figures 1A-1D refer to separate label attachment regions. All except for position 12 of Figure 1A are labeled with one of four types of label monomers, depicted as grey, white, hatched or stripe "sun" diagrams. Position 12 of Figure 1A is an unlabeled "dark spot." Figures 1E and 1F represent variations on the nanoreporters of Figures 1B and 1D, respectively, in which the target molecule to which the nanoreporters are bound comprises biotin moieties (shown as small asterisks), for example biotin-modified nucleotides randomly incorporated into a target nucleic acid. The nanoreporters themselves further optionally comprise an affinity tag (not shown).

[0031] Figures 2A-2C. Figure 2A shows an illustration of a label unit of a nanoreporter, containing a scaffold with patch units and corresponding split flaps disposed along its length. Figure 2B illustrates the components of a single patch pair and its corresponding flap, containing: 1: a portion of a nanoreporter scaffold (e.g., M13 single-stranded DNA); 2: A patch pair; 3: a split flap pair; and 4: labeled oligonucleotides, each with a label monomer incorporated, hybridized to the split flap. Figure 2C shows a nanoreporter with 4 "spots," each spot designed to contain 9 patch pairs of 60-65 nucleotides, each attached to a split flap pair of 95-100 nucleotides. Each split flap pair had binding site for 12 oligonucleotides each attached to a single label monomer. Each spot therefore had binding sites for 108 label monomers.

[0032] Figure 3. A nanoreporter in which the patches that are RNA segments can be used with (Figure 3A) and without registers (Figure 3B). Both Figure 3A and 3B depict a (1) nanoreporter scaffold (heavy black line) to which are attached (2) eight RNA segments (lines 1-8), (3) a target-specific sequence (dotted line "T") and (4) an oligonucleotide (checkered line "0") that is partly complementary to the nanoreporter scaffold and partly complementary to the target-specific sequence. This oligonucleotide is referred to as a "ligator" oligonucleotide. In Figure 3A, only one register, *i.e.*, every alternate RNA segment is labeled. The second register positions serve as "spacers," making it possible to generate a nanoreporter code in which consecutive positions in the code are the same "color," or spectrally indistinguishable. In Figure 3B, both registers, *i.e.*, adjacent RNA segments with no intervening spacers, are labeled, with no nearest neighbor of the same "color."

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 727

[0033] Figure 4. is an image of a dual nanoreporter hybridized to a target molecule. Here, both registers are labeled. The nanoreporters are labeled with three different colors, Alexa 488, Cy3 and Alexa 647 (labeled 1, 2 and 3, respectively). The left brackets show one probe of the dual nanoreporter and the right brackets show the other probe of the dual nanoreporter. Colors 1, 2 and 3 were each acquired in different channels and the first and second registers, seen as rows of spots, were shifted up by several pixels to be able to show each register individually.

[0034] Figures 5A-5D. This figure illustrates the various components of the dual nanoreporters shown in Figure 4. Figure 5A illustrates one color (here, Alexa 488, depicted in the left column as open circles), which is spectrally distinguishable from Cy3 (shown in Figure 5B, depicted in the left column as vertically striped circles) and Alexa 647 (shown in Figure 5C as diagonally striped circles). The images obtained from each were superimposed to generate Figure 5D.

[0035] Figures 6A-6E. Figure 6A is a schematic illustration of the experiment shown in Figures 6B and 6C. In this case, the star represents biotin that was used to attach the complex by one end to the surface prior to stretching. Figures 6B and 6C show images from experiments in which S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA (Figure 6B) or S2 target RNA (Figure 6C) were hybridized. Figure 6E shows a close-up of a nanoreporter complexes from Figure 6B, each containing S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA. Figure 6D shows an image of a negative control experiment, in which S2-A ghost probe, S2-B labeled nanoreporter and no S2 target RNA were hybridized.

[0036] Figures 7A-G. Figures 7A, 7B, 7C and 7D depict different permutations of patches on a nanoreporter scaffold, Figures 7E and 7F depict different permutations of split flaps on a nanoreporter scaffold, optionally hybridized to one or more oligonucleotides, as in Figure 7G. In Figure 7A-G, α refers to a 5' or 3' molecule or end of a molecule, and β refers to a corresponding 3' or 5' molecule or end of a molecule.

[0037] Figure 8. Figure 8 depicts a scheme in which single-stranded M13 phage is linearized for use as a nanoreporter scaffold. The circular M13 phage is annealed to a five-fold excess of BamH1 cutter oligonucleotide (hatched lines) (1), and the resulting partially double-stranded M13 digested with the restriction endonuclease BamH1 (2), resulting in a linearized M13 in which BamH1 cutter oligonucleotide is still attached (3). This M13-oligonucleotide complex is heated in the presence of an excess oligonucleotide complementary to the BamH1 cutter oligonucleotide (an "anti-BamH1 oligonucleotide") (grey lines) (4). The BamH1 cutter oligonucleotide anneals to the excess of anti-BamH1

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 728

oligonucleotide, and the M13 molecule is purified from the oligonucleotide, for example by using size exclusion columns, to yield M13 scaffold.

[0038] Figure 9. Figure 9 illustrates a computer system in accordance with an embodiment of the present invention.

[0039] Figures 10A-10C. Figure 10A shows a labeled nanoreporter containing a single affinity tag, A1. Another affinity tag, A2, can be attached to the nanoreporter by direct binding of the nanoreporter to a molecule containing A2 (*e.g.*, if the nanoreporter is or comprises a nucleic acid, it can hybridize directly with another nucleic acid to which A2 is attached), as depicted in Figure 10B. Alternatively, the second affinity tag, A2, can be attached to the labeled nanoreporter via a bridging molecule, such as the bridging nucleic acid ("X") depicted in Figure 10C.

[0040] Figures 11A-11B depict a labeled (nucleic acid-based) nanoreporter with an affinity tag, A1, at one end. In Figure 11, the labeled nanoreporter is immobilized through the binding of A1 to an immobilized affinity partner. The other end of the nanoreporter is in solution (Figure 11A), but can be immobilized by hybridization to a complementary oligonucleotide which contains another affinity tag (A2) used to immobilize the nanoreporter (Figure 11B). A1 and A2 can be the same, for example biotin, for immobilization on an avidin- or streptavidin-coated surface. Upon immobilization of A1, the nanoreporter can be stretched, or "elongated" as depicted in Figure 11, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code. Optionally, while the nanoreporter is in an elongated state, A2 is introduced and binds the end of the nanoreporter that is complementary to A2 down to the surface.

[0041] Figures 12A-12B. Figure 12A provides an illustration of a nanoreporter comprising an immobilized first portion F1; and Figure 12B provides an illustration of a nanoreporter extended in an electrical field and comprising immobilized first portion F1 and immobilized second portion F2, wherein F2 is immobilized via a complex with molecule F3.

[0042] Figures 13A-13C. Figure 13A provides an illustration of a three-member complex for immobilization of an extended nanoreporter; Figure 13B provides an illustration of a two-member complex for immobilization of an extended nanoreporter; and Figure 13C provides an illustration of an incomplete complex for immobilization of an extended nanoreporter.

[0043] Figures 14A-14D. Figure 14A provides an illustration of a nanoreporter comprising an immobilized first portion F1; Figure 14B provides an illustration of an extended nanoreporter immobilized at first portion F1 and at a second portion via complexes with F2;

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 729

Figure 14C provides an illustration of a nanoreporter comprising a first portion immobilized to an avidin surface via biotin; and Figure 14D provides an illustration of an extended nanoreporter immobilized at a first portion and at a second portion via selective binding of biotin to an avidin surface.

[0044] Figures 15A-15C. Figure 15A illustrates immobilization of one terminus of a DNA molecule in a microfluidic device; Figure 15B illustrates extension of the DNA in an electric field; and Figure 15C illustrates selective immobilization of a second terminus of the extended DNA molecule.

[0045] Figure 16 provides an image of extended nanoreporters selectively immobilized by the methods of the present invention.

5. DETAILED DESCRIPTION OF THE INVENTION

[0046] The present invention pertains to nanoreporters, and their manufacture and use. A fully assembled and labeled nanoreporter comprises two main portions, a target-specific sequence that is capable of binding to a target molecule, and a labeled region which emits a "code" of signals (the "nanoreporter code") associated with the target-specific sequence. Upon binding of the nanoreporter to the target molecule, the nanoreporter code identifies the target molecule to which the nanoreporter is bound.

[0047] Nanoreporters are modular structures. Generally, a nanoreporter is a molecular entity containing three basic elements: a scaffold containing two or more label attachment regions, one or more patches attached to the scaffold, and a target-specific sequence, also attached to the scaffold. The elements of a nanoreporter can be found in a single molecular entity (a "singular" nanoreporter), or two distinct molecular entities (a "dual" nanoreporter). Each molecular entity may be composed of one molecule or more than one molecule attached to one another by covalent or non-covalent means. Generally, each component of a dual nanoreporter has a target-specific sequence that binds to a different site on the same target molecule. This allows for smaller nanoreporter components with more efficient kinetics of binding of the nanoreporter to the target molecule and better signal:noise ratios resulting from the greater binding specificity.

[0048] The patches attached to a nanoreporter scaffold serve to attach label monomers to a nanoreporter scaffold. Patches may be directly labeled, for example by covalent incorporation of one or more label monomers into nucleic acid patches. Alternatively, patches may be attached to flaps, which maybe labeled directly, for example by covalent incorporation of one or more label monomers into a nucleic acid flap, or indirectly, for example by hybridization of a nucleic acid flap to an oligonucleotide which is covalently attached to one or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 730

more label monomers. Where the label monomers attached to a label attachment region are not directly incorporated into a patch or flap, the patch or flap serves as a "bridge" between the label monomer and the label attachment region, and may be referred to as a "bridging molecule," *e.g.*, a bridging nucleic acid.

[0049] Additionally, nanoreporters may have affinity tags for purification and/or for immobilization (for example to a solid surface). Nanoreporters, or nanoreporter-target molecule complexes, are preferably purified in two or more affinity selection steps. For example, in a dual nanoreporter, one probe can comprise a first affinity tag and the other probe can comprise a second (different) affinity tag. The probes are mixed with target molecules, and complexes comprising the two probes of the dual nanoreporter are separated from unbound materials (*e.g.*, the target or the individual probes of the nanoreporter) by affinity purification against one or both individual affinity tags. In the first step, the mixture can be bound to an affinity reagent for the first affinity tag, so that only probes comprising the first affinity tag and the desired complexes are purified. The bound materials are released from the first affinity reagent and optionally bound to an affinity reagent for the second affinity tag, allowing the separation of complexes from probes comprising the first affinity tag. At this point only full complexes would be bound. The complexes are finally released from the affinity reagent for the second affinity tag and then preferably stretched and imaged. The affinity reagent can be any solid surface coated with a binding partner for the affinity tag, such as a column, bead (*e.g.*, latex or magnetic bead) or slide coated with the binding partner. Immobilizing and stretching nanoreporters using affinity reagents is fully described in U.S. provisional application no. 60/753,816 by Sean M. Ferree and Dwayne L. Dunaway, entitled "Compositions Comprising Oriented, Immobilized Macromolecules and Methods for Their Preparation," attorney docket no. 11616-014-888 filed on December 23, 2005, which is incorporated by reference herein in its entirety.

[0050] Nanoreporter and nanoreporter-target complexes which are or comprise nucleic acids may be affinity-purified or immobilized using a nucleic acid, such as an oligonucleotide, that is complementary to at least part of the nanoreporter or target. In a specific application where the target includes a poly A or poly dA stretch, the nanoreporter-target complex can be purified or immobilized by an affinity reagent coated with a poly dT oligonucleotide.

[0051] The sequence of signals emitted by the label monomers associated with the various label attachment regions of the scaffold of a given nanoreporter allows for the unique identification of the nanoreporter. A nanoreporter having a unique identity or unique spectral signature is associated with a target-specific sequence that recognizes a specific target molecule or a portion thereof. When a nanoreporter is exposed to a mixture containing the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 731

target molecule under conditions that permit binding of the target-specific sequence(s) of the nanoreporter to the target molecule, the target-specific sequence(s) preferentially bind(s) to the target molecule. Detection of the spectral code associated with the nanoreporter allows detection of the presence of the target molecule in the mixture (qualitative analysis). Counting all the label monomers associated with a given spectral code or signature allows the counting of all the molecules in the mixture associated with the target-specific sequence coupled to the nanoreporter (quantitative analysis). Nanoreporters are thus useful for the diagnosis or prognosis of different biological states (*e.g.*, disease vs. healthy) by quantitative analysis of known biological markers. Moreover, the exquisite sensitivity of single molecule detection and quantification provided by the nanoreporters of the invention allows for the identification of new diagnostic and prognostic markers, including those whose fluctuations among the different biological states is too slight detect a correlation with a particular biological state using traditional molecular methods. The sensitivity of nanoreporter-based molecular detection permits detailed pharmacokinetic analysis of therapeutic and diagnostic agents in small biological samples.

[0052] Many nanoreporters, referred to as singular nanoreporters, are composed of one molecular entity, as depicted in Figure 1D. However, to increase the specificity of a nanoreporter and/or to improve the kinetics of its binding to a target molecule, a preferred nanoreporter is a dual nanoreporter composed of two molecular entities, each containing a different target-specific sequence that binds to a different region of the same target molecule. Various embodiments of dual nanoreporters are depicted in Figures 1A-1C. In a dual nanoreporter, at least one of the two molecular entities is labeled. The other molecular entity is not necessarily labeled. Such unlabeled components of dual nanoreporters are referred to herein as "ghost probes" (see Figure 1C) and often have affinity tags attached, which are useful to immobilize and/or stretch the complex containing the dual nanoreporter and the target molecule to allow visualization and/or imaging of the complex.

[0053] Because of their modular structures, nanoreporters may be assembled and labeled in a variety of different ways. For example, a nanoreporter scaffold can be attached to a target-specific sequence (for example by hybridization and, optionally, ligation), and the structure comprising the scaffold and target-specific sequence attached to one or more patches and, where desired, flaps. Alternatively, the nanoreporter scaffold can first be attached to one or more patches (and, optionally, flaps), and the scaffold/patch structure then attached to a target specific sequence. Thus, unless stated otherwise, a discussion or listing of steps in nanoreporter assembly does not imply that a specific route of assembly must be followed.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 732

[0054] Nanoreporter assembly and use is exemplified herein largely by way of description of a variety of nucleic acid-based nanoreporters; however, one of skill in the art would recognize that the methods described herein are applicable to an amino acid-based (or hybrid nucleic acid-/amino acid-based) nanoreporter. Illustrative embodiments of partially and fully assembled nanoreporters are listed below.

[0055] At its simplest, a nanoreporter comprises a scaffold having at least two label attachment regions capable of being labeled and resolved. The scaffold can be any molecular entity that allows the formation of label attachment regions on the scaffold that can be separately labeled and resolved. The number of label attachment regions to be formed on a scaffold is based on the length and nature of the scaffold, the means of labeling the nanoreporter, as well as the type of label monomers emitting a signal to be attached to the label attachment regions of the scaffold. A nanoreporter according to the invention may have a scaffold including two or more label attachment regions. Suitable scaffold structures include DNA-based scaffolds.

[0056] The invention also provides labeled nanoreporters in which one or more label attachment regions are attached to corresponding label monomers, each label monomer emitting a signal. For example a labeled nanoreporter according to the invention is obtained when at least two label monomers are attached to two corresponding label attachment regions of the scaffold such that these labeled label attachment regions, or "spots," are distinguishable. Label monomers emitting a signal associated with different label attachment regions of the scaffold can emit signals that are spectrally indistinguishable under the detection conditions ("like" signals), or can emit signals that are spectrally distinguishable under the detection conditions.

[0057] The invention also provides a nanoreporter in which two or more label monomers are attached to a label attachment region. The signal emitted by the label monomers associated with the label attachment region produces an aggregate signal that is detected. The aggregate signal produced may be made up of like signals or made up of at least two spectrally distinguishable signals.

[0058] In one embodiment, the invention provides a nanoreporter in which at least two label monomers emitting like signals are attached to two corresponding label attachment regions of the scaffold and the two label monomers are spatially distinguishable. In another embodiment, the invention provides a nanoreporter in which at least two label monomers emitting two distinguishable signals are attached to two neighboring label attachment regions, for example two adjacent label attachment regions so that at least two label monomers are spectrally distinguishable.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 733

[0059] The invention provides a nanoreporter in which two spots emitting like signals are separated by a spacer region. Such a spacer region allows resolution or better resolution of the like signals emitted by label monomers attached to the two spots. In one embodiment, the spacer regions have a length determined by the resolution of an instrument employed in detecting the nanoreporter.

[0060] The invention provides a nanoreporter with one or more "double spots." Each double spot contains two or more (*e.g.*, three, four or five) adjacent spots that emit like signals without being separated by a spacer region. Double spots can be identified by their sizes. A label monomer emitting a signal according to the invention may be attached covalently or non-covalently (*e.g.*, via hybrization) to a patch that is attached to the label attachment region. The label monomers may also be attached covalently or non-covalently (*e.g.*, via hybrization) to a flap attached to a patch that is in turn attached to the scaffold. The flap can be formed by one molecule or two or more molecules ("flap pieces") that form a split flap.

[0061] The invention also provides a nanoreporter associated with a spectral code determined by the sequence of signals emitted by the label monomers attached (*e.g.*, indirectly via a patch) to label attachment regions on the scaffold of the nanoreporter, whereby detection of the spectral code allows identification of the nanoreporter.

[0062] In one embodiment, the invention provides a nanoreporter further comprising an affinity tag attached to the nanoreporter scaffold, such that attachment of the affinity tag to a support allows scaffold stretching and resolution of signals emitted by label monomers corresponding to different label attachment regions on the scaffold. Nanoreporter stretching may involve any stretching means known in the art including but not limited to, means involving physical, hydrodynamic or electrical means.

[0063] In yet another embodiment, the invention provides a nanoreporter further comprising flaps attached to label attachment regions of the scaffold, wherein a flap attached to a label attachment region of the scaffold attaches the label monomer corresponding to the label attachment region, thereby indirectly attaching label monomers to corresponding label attachment regions on the scaffold. In a further embodiment, each label monomer comprises a signal emitting portion and an oligonucleotide portion of a predetermined sequence, and the flaps comprise repeats of a flap sequence complementary to the oligonucleotide portion of a corresponding label, whereby one or more label monomers attach to a corresponding label attachment region through hybridization of the oligonucleotide portions of the label monomers to the repeats of the flap sequence thereby producing a labeled nanoreporter.

[0064] A nanoreporter according to the invention can further include a target-specific sequence coupled to the scaffold. The target-specific sequence is selected to allow the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 734

nanoreporter to recognize, bind or attach to a target molecule. The nanoreporters of the invention are suitable for identification of target molecules of all types. For example, appropriate target-specific sequences can be coupled to the scaffold of the nanoreporter to allow detection of a target molecule. Preferably the target molecule is DNA (including cDNA), RNA (including mRNA and cRNA), a peptide, a polypeptide, or a protein.

[0065] One embodiment of the invention provides increased flexibility in target molecule detection with label monomers according to the invention. In this embodiment, a dual nanoreporter comprising two different molecular entities, each with a separate target-specific region, at least one of which is labeled, bind to the same target molecule. Thus, the target-specific sequences of the two components of the dual nanoreporter bind to different portions of a selected target molecule, whereby detection of the spectral code associated with the dual nanoreporter provides detection of the selected target molecule in a biomolecular sample contacted with the dual nanoreporter.

[0066] The invention also provides a method of detecting the presence of a specific target molecule in a biomolecular sample comprising: (i) contacting the sample with a dual nanoreporter under conditions that allow binding of the target-specific sequences in the dual nanoreporter to the target molecule and (ii) detecting the spectral code associated with the dual nanoreporter. Depending on the nanoreporter architecture, the dual nanoreporter may be labeled before or after binding to the target molecule.

[0067] Structural stability of a nanoreporter can be increased through ligation of the patches and, optionally, ligation of the split flaps and/or the labeled oligonucleotides hybridized to the split flaps.

[0068] In addition to the qualitative analytical capabilities provided by the nanoreporters of the invention and the analytical techniques based thereon, the nanoreporters of the invention are uniquely suitable for conducting quantitative analyses. By providing a one to one binding between the nanoreporters (whether singular or dual nanoreporters) of the invention and their target molecules in a biomolecular sample, all or a representative portion of the target molecules present in the sample can be identified and counted. This individual counting of the various molecular species provides an accurate and direct method for determining the absolute or relative concentration of the target molecule in the biomolecular sample. Moreover, the ability to address each molecule in a mixture individually leverages benefits of miniaturization including high sensitivity, minimal sample quantity requirements, high reaction rates which are afforded by solution phase kinetics in a small volume, and ultimately very low reagent costs.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 735

[0069] As will be appreciated from the description and examples provided below, the present invention provides numerous advantages. For example, the complex modularity in forming nanoreporters according to the invention allows for systematic creation of libraries of unique nanoreporters having a very high degree of diversity (*e.g.*, millions of uniquely recognizable nanoreporters). This modularity allows flexibility in customizing nanoreporter populations to specific applications which in turn provides significant manufacturing efficiencies. Another advantage that will be appreciated through the following description stems from the flexibility in assembling the nanoreporters of the invention. That is, due to their modular structure, the nanoreporters of the invention can be assembled prior to shipment to a point of use or assembled at the point of use.

5.1 NANOREPORTER NOMENCLATURE

[0070] All terms used herein have their ordinary meanings to those of skill in the art unless indicated otherwise. The following terms shall have the following meanings.

[0071] BINDING PAIR. The term "binding pair" refers to first and second molecules or moieties that are capable of selectively binding to each other, *i.e.* binding to each other with greater affinity than to other components in a composition. The binding between the members of the binding pair can be covalent or non-covalent. In certain embodiments, the binding is noncovalent. Exemplary binding pairs include immunological binding pairs (*e.g.*, any haptenic or antigenic compound in combination with a corresponding antibody or binding portion or fragment thereof, for example digoxigenin and anti-digoxigenin, fluorescein and anti-fluorescein, dinitrophenol and anti-dinitrophenol, bromodeoxyuridine and anti-bromodeoxyuridine, mouse immunoglobulin and goat anti-mouse immunoglobulin) and nonimmunological binding pairs (*e.g.*, biotin-avidin, biotin-streptavidin, hormone-hormone binding protein, receptor-receptor ligand (*e.g.*, acetylcholine receptor-acetylcholine or an analog thereof), IgG-protein A, lectin-carbohydrate, enzyme-enzyme cofactor, enzyme-enzyme inhibitor, complementary polynucleotide pairs capable of forming nucleic acid duplexes, and the like). For instance, immunoreactive binding members may include antigens, haptens, aptamers, antibodies (primary or secondary), and complexes thereof, including those formed by recombinant DNA methods or peptide synthesis. An antibody may be a monoclonal or polyclonal antibody, a recombinant protein or a mixture(s) or fragment(s) thereof, as well as a mixture of an antibody and other binding members. Other common binding pairs include but are not limited to, biotin and avidin (or derivatives thereof), biotin and streptavidin, carbohydrates and lectins, complementary nucleotide sequences (including probe and capture nucleic acid sequences), complementary peptide sequences including those formed by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 736

recombinant methods, effector and receptor molecules, hormone and hormone binding protein, enzyme cofactors and enzymes, enzyme inhibitors and enzymes, and so forth.

[0072] DARK SPOT. The term "dark spot" refers to a lack of signal, or "spot," from a label attachment site on a nanoreporter. Dark spots can be incorporated into the nanoreporter code to add more coding permutations and generate greater nanoreporter diversity in a nanoreporter population.

[0073] EXTENDED STATE. The term "Extended state" refers to a nanoreporter in a state that would be recognized as extended by one of skill in the art. In certain embodiments, a nanoreporter is in an extended state when it is extended relative to its native conformation in solution. In certain embodiments, a nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter. In certain embodiments, an extended state of a nanoreporter can be determined quantitatively. In such embodiments, those of skill in the art will recognize R as the end-to-end vector of the nanoreporter, *i.e.* the distance between two termini of the nanoreporter, and <R> as the average end-to-end vector such that 95% of R will be within 2<R> in a solution deemed appropriate to one of skill in the art. Exemplary solutions include, for example, a dilute solution of the nanoreporter in water or in a pH buffer. In particular embodiments, a nanoreporter is in an extended state when R is greater than 2.0<R>.

[0074] FLAP. The term "flap" as used herein refers to a molecular entity attached to a patch or patch pair attached to a label attachment region. The flap is one or more molecules containing label monomers or capable of binding one or more molecules containing label monomers. By providing indirect labeling of the regions, the flaps provide more flexibility in controlling the number of signal emitting monomers associated with a region as well as the nature of those monomers. Flaps may be formed by a single molecular piece or several molecular pieces (*e.g.*, two pieces) forming a "split flap" (see, *e.g.*, Figure 7).

[0075] GHOST PROBE. A molecule comprising a target-specific sequence, but which is not labeled with a label monomer that emits a signal that contributes to the nanoreporter code.

[0076] LABELED NANOREPORTER. A labeled nanoreporter is a nanoreporter in which at least one patch of the nanoreporter is attached to one or more label monomers that generate(s) a signal that forms at least part of the nanoreporter code.

[0077] LABEL UNIT. The term "label unit" refers to the non-target-specific portions of a labeled nanoreporter.

[0078] NANOREPORTER. The term "nanoreporter" refers to a molecular entity that has (i) a molecule ("scaffold") containing at least two label attachment regions; (ii) at least one

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 737

patch attached to at least one label attachment region; and (iii) a target-specific sequence. As described in detail below, nanoreporters can be singular nanoreporters (all components being in a single molecular entity) or dual nanoreporters (all the components being in two separate molecular entities). Nanoreporters are preferably synthetic, *i.e.*, non-naturally-occurring molecules, for example are chimeric molecules made by joining two or more manmade and/or naturally occurring sequences that normally exist on more than one molecule (*e.g.*, plasmid, chromosome, viral genome, protein, etc.).

[0079] NANOREPORTER CODE. The order and nature (*e.g.*, primary emission wavelength(s), optionally also length) of spots from a nanoreporter serve as a nanoreporter code that identifies the target molecule capable of being bound by the nanoreporter through the nanoreporter's target specific sequence(s). When the nanoreporter is bound to a target molecule, the nanoreporter code also identifies the target molecule. Optionally, the length of a spot can be a component of the nanoreporter code.

[0080] ORIENTED STATE. The term "oriented state" refers to a nanoreporter in a state that would be recognized as oriented by one of skill in the art. In certain embodiments, a nanoreporter is in an oriented state when it is oriented relative to its native conformation in solution. In certain embodiments, the nanoreporter is oriented when it is arranged in parallel with the field of a force capable of orienting the nanoreporter. In certain embodiments, the nanoreporter is oriented when it is one of a plurality of nanoreporters that are arranged in parallel, as recognized by those of skill in the art.

[0081] PATCH. The term "patch" refers to a molecular entity attached to the label attachment region of the nanoreporter scaffold, generally for the purpose of labeling the nanoreporter. The patch can have one or more label monomers either directly (covalently or noncovalently) or indirectly attached to it, either prior to or after its attachment to the scaffold.

[0082] PROBE. This refers to a molecule that has a target-specific sequence. In the context of a singular nanoreporter, the term "probe" refers to the nanoreporter itself; in the context of a dual nanoreporter, the term "probe" refers to one or both of the two components of the nanoreporter.

[0083] PROBE PAIR. The term "probe pair" refers to a dual nanoreporter.

[0084] REGISTER. The term "register" refers to a set of alternating label attachment regions.

[0085] SELECTIVE BINDING. The term "selective binding" refers to the any preferential binding of a pair of molecules or moieties for each other with respect to other molecules or moieties in a composition that would be recognized by one of skill in the art. In certain embodiments, a pair of molecules or moieties selectively binds when they preferentially

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 738

bind each other compared to other molecules or moieties. Selective binding can include affinity or avidity, or both, of one molecule or moiety for another molecule or moiety. In particular embodiments, selective binding requires a dissociation constant (K_D) of less than about $1x10^{-5}$ M or less than about $1x10^{-6}$ M, $1x10^{-7}$ M, $1x10^{-8}$ M, $1x10^{-9}$ M, or $1x10^{-10}$ M. In contrast, in certain embodiments, non-selective binding has significantly less affinity, for example, a K_D greater than $1x10^{-3}$ M.

[0086] SPOT. A spot, in the context of nanoreporter detection, is the aggregate signal detected from the label monomers attached to a single label attachment site on a nanoreporter, and which, depending on the size of the label attachment region and the nature (*e.g.*, primary emission wavelength) of the label monomer, may appear as a single point source of light when visualized under a microscope. Spots from a nanoreporter may be overlapping or non-overlapping. The nanoreporter code that identifies that target molecule can comprise any permutation of the length of a spot, its position relative to other spots, and/or the nature (e.g., primary emission wavelength(s)) of its signal.

[0087] TARGET-SPECIFIC SEQUENCE. The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter scaffold. The target molecule is preferably (but not necessarily) a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of the cDNA.

5.2 THE NANOREPORTER SCAFFOLD

[0088] The nanoreporter scaffold can be any molecular entity, more preferably a nucleic acid molecule, containing label attachment regions to which label monomers can be directly or indirectly attached. In one embodiment, the nanoreporter scaffold is a protein scaffold; in a preferred embodiment, the nanoreporter scaffold is a nucleic acid scaffold in which the label attachment regions are single-stranded regions to which other nucleic acids, such as oligonucleotide patches, RNA patches, or DNA patches, can attach by hybridization. In specific embodiments, the nanoreporter scaffold is a nucleic acid molecule.

[0089] There are no particular limitations on the types of scaffolds that are suitable for forming the nanoreporters of the invention. A scaffold according to the invention can essentially have any structure including, for example, single stranded linear scaffold, double stranded linear scaffold, single stranded circular scaffold or double stranded circular scaffold. Examples of scaffold structures include, for example, a scaffold made of one molecular entity such as polypeptides, nucleic acids or carbohydrates. A scaffold may also include a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 739

combination of structures, for example, a scaffold may be made of one or more polypeptide stretches coupled to one or more carbohydrate stretches.

[0090] Suitable molecular entities for scaffolds according to the invention include polymeric structures particularly nucleic acid based polymeric structures such as DNA. DNA based structures offer numerous advantages in the context of the present invention due at least in part to the vast universe of existing techniques and methodologies that allow manipulation of DNA constructs.

[0091] As indicated above, the scaffold may be single stranded or double stranded. Double stranded scaffold can be either conventional double stranded DNA or a double strand that is composed of a linear single stranded stretch of nucleic acid with patch units or flat-patches attached. A scheme for forming a linearized scaffold is depicted in Figure 8.

[0092] A scaffold can have 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21-100 label attachment regions or more.

[0093] The label attachment regions of a nanoreporter scaffold will vary in size depending on the method of labeling. In various embodiments, a label attachment region has a length anywhere from 10 nm to 10,000 nm, but is more preferably from 50nm to 5,000 nm, and is more preferably from 100 nm to 1,000 nm. In various embodiments, the label attachment region is from about 100 nm to about 500 nm, from about 150 nm to about 450 nm, from about 200 nm to about 400 nm, or from 250 to about 350 nm. In a preferred embodiment, the label attachment region corresponds closely to the size of a diffraction-limited spot, *e.g.*, the smallest spot that can be detected with standard optics, which is about 300nm.

[0094] Where the scaffold is a nucleic acid, 1 nm corresponds to approximately 3 nucleotides; thus, an approximately 300nm-label attachment region corresponds to approximately 900 bases. In other preferred embodiments, the label attachment region is from about 300 nucleotides to about 1.5kb, from about 450 nucleotides to about 1.35 kb, from about 0.6 kb to about 1.2 kb, or from 0.75 kb to about 1.05 kb.

[0095] An illustrative example of a molecular entity for a nanoreporter scaffold according to the invention is M13 DNA, which is single-stranded. In one embodiment, the nanoreporter scaffold is circular at least partially single stranded DNA, such as circular M13. In a more preferred embodiment, the nanoreporter scaffold is linear at least partially single stranded DNA, such as linear M13. In a specific embodiment, the M13 single-stranded DNA obtained by operating a cut at the BamH1 site of circular M13 DNA.

[0096] It should be noted that within the context of the present invention, linear DNA provides additional advantages compared to circular DNA. One advantage of using linear DNA in forming a scaffold according to the invention relates to the significantly reduced

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 740

torsional stress associated with linear DNA. The added torsional stress associated with circular DNA may interfere with the structural integrity of the scaffold upon the addition to the scaffold of other components of the nanoreporter, such as patch units. Severe torsional stress may lead to the breaking of the structure of the scaffold. It should be noted however that the nanoreporters where only a few, short label attachment sites are labeled, circular DNA may be suitable.

5.2.1 NOVEL SYNTHETIC NANOREPORTER SCAFFOLD SEQUENCES

[0097] The present invention provides nanoreporter scaffold that are artificial nucleic acid molecules (DNA, RNA, or DNA/RNA hybrids) designed to have features that optimize labeling and detection of the nanoreporter. In these aspects of the invention, a nanoreporter scaffold is an artificial nucleic acid comprising one or more synthetic sequences from 50 to 50,000 bases long. Accordingly, the nanoreporter scaffold, which is preferably a DNA, is designed to have one or more regions, useful as label attachment regions, comprising a regular pattern of a particular base (the "regularly-repeated base"). In such regions, the regularly-repeated base occurs with a periodicity of every nth residue, where n is any number, and preferably from 4 to 25.

[0098] Preferably, not more than 25% of the regularly-repeated base in a Region appears at other than the regular intervals. For example, if in a Region of 100 nucleotides there are 12 thymidine bases, and thymidine is the regularly-repeated base, in this aspect of the invention not more than 25% of these, *i.e.*, 3 thymidine bases, appear outside the regular pattern of thymidines. In specific embodiments, not more than 20%, not more than 15%, not more than 10%, not more than 9%, not more than 8%, not more than 7%, not more than 6%, not more than 5%, not more than 4%, not more than 3%, not more than 2% or not more than 1% of the base appears at other than the regular intervals in the region.

[0099] The regularly-repeated base in the Regions in a nanoreporter scaffold, or its complementary regularly-repeated base in an annealed patch (or segment) can be used to attach label monomers, preferably light emitting label monomers, to the nanoreporter in a regular, evenly spaced pattern for better distribution of the nanoreporter signal. Preferably, where a Region is labeled, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or at least 98% of occurrences of the regularly-repeated base is attached to at least one light-emitting label monomer, either by covalent attachment of a label monomer to a base, or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 741

by hybridization to a nucleic acid in which the complements of the regularly-repeated base are so-labeled.

[00100] This percentage of occurrences can be measured by any means known in the art. In one method, the amount of nucleic acid produced in a labeling reaction is purified (for example, RNA can be purified using a Qiagen RNeasy kit) and subjected to UV spectrophotometry. The absorbance ("A") at the appropriate wavelengths is measured for each of the nucleic acid (260 nm) and the label monomer whose occurrence is to be measured (*e.g.*, 495 nm for Alexa Fluor 488; 590 nm for Alexa Fluor 594; 650 for Alexa Fluor 647; and 550 nm for Cy3). The absorbance of the nucleic acid is corrected by adjusting the value of the absorbance at 260 nm ("A260") to remove the "noise" contribution from the label monomer by subtracting the absorbance at the peak wavelength for the label monomer (A_{LM}) minus the correction factor for that label monomer. Where the nucleic acid is RNA, the number of label monomers per one thousand nucleotides can be calculated according to the formula:

$$\frac{\text{no. of label monomers}}{\text{1000 nucleotides}} = \frac{A260}{A_{LM}} \times \frac{9010}{EC_{LM}} \times 1000$$

where EC_{LM} is the extinction coefficient for the label monomer. From this formula, the percentage of occurrences of the regularly-repeated base that are attached to a light-emitting label monomer can be calculated.

[00101] Generally, the preferred regularly-repeating base in a label attachment region is thymidine, so that the region can be labeled by hybridization to one or more complementary patches (*e.g.*, RNA segments) in which the regularly-repeated base is uridine. This permits the use of amino-allyl-modified UTPs, which are readily commercially available, as label monomer attachment sites, in an otherwise random sequence. Preferably, in addition to the regular periodicity of the Regions, the regions (and the nucleic acid comprising them) contain minimal secondary structure. The overall GC-content is preferably maintained close to 50%, and is preferably consistent over relatively short stretches to make local Tm's similar.

[00102] The artificial nucleic acids of the invention, or at least the Regions therein, preferably do not have direct or inverted repeats that are greater than 12 bases in length. In other embodiments, the artificial nucleic acids and/or Regions do not have direct or inverted repeats that are greater than 11, 10 or 9 bases in length.

[00103] In an exemplary Region in which the regularly-repeated nucleotide is a thymidine and a GC content of approximately 50%, excess adenines would make up the loss in abundance of T's. To generate the selected sequence, random sequences with fixed patterns of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 742

T's ranging from every 4th base to every 25th base are created and screened to minimize the presence of inverted and direct repeats.

[00104] Sequences are also screened preferably to avoid common six-base-cutter restriction enzyme recognition sites. Selected sequences are additionally subjected to predicted secondary structure analysis, and those with the least secondary structure are chosen for further evaluation. Any program known in the art can be used to predict secondary structure, such as the MFOLD program (Zuker, 2003, Nucleic Acids Res. 31 (13):3406-15; Mathews *et al.*, 1999, J. Mol. Biol. 288:911-940).

[00105] An appropriate sequence is divided into label attachment regions ranging from 50 bases to 2 kilobases long (could be longer). Each label attachment region is a unique sequence, but contains a consistent number and spacing of T's in relation to the other label attachment regions in a given reporter sequence. These label attachment regions can interspersed with other regions whose sequence does not matter. The synthetic label attachment regions in a nanoreporter scaffold can be of different lengths and/or have different regularly-repeated bases. An optimized start sequence for transcription by RNA polymerase T7, T3, or SP6 (beginning at position +1 of the transcript) can be added to the 5' end of each label attachment region. Restriction sites are optionally added at the boundaries of each label attachment region to allow specific addition or deletion of individual label attachment regions to the sequence using conventional cloning techniques. The number of synthetic label attachment regions in a nanoreporter preferably ranges from 1 to 50. In yet other embodiments, the number of synthetic label attachment regions in a nanoreporter ranges from 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 synthetic label attachment regions to 15, 20, 30, 40, or 50 synthetic label attachment regions, or any range in between.

[00106] An example of such a novel synthetic label attachment region is given below. In this sequence, shown 5' to 3', the T's are placed in every 8th position and the region is bounded by a 5' Sac I restriction site and a 3' Kpn I restriction site. An optimized transcript start site for T7 polymerase (GGGAGA) is included at the 5' end of the region, downstream of the 5' restriction site. The complement of this sequence, when generated as a single-stranded molecule, forms the scaffold for the RNA molecule transcribed from this label attachment region.

GAGCTCGGGAGATGGCGAGCTGGAAGCATCAGAAAGTAGGAAGATGACAA
AATAGGGCCATAGAAGCATGAAGAACTGAACGCATGAGACAATAGGAAGC
TACGCCACTAGGGACCTGAGAAGCTGAGCGGCTCAGCGGGTCCGAGCGTC
AAAAAATAAAAGAGTGAAACAATAGACGAATGACGCGGTAAAACCATCCA
GAAGTAAACGGGTACAAACATACAGAGATAGCCACCTGGACCAATAGGCA

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 743

CGTACAAACGTACAAGCCTGGCGCGATGAGGCAATCCACACGTGCAGAGC
TGGAACAATGGAAAGATGCAAGAATAAACCGATACCGGGATCGAGGGCTC
AGCGAATAAAGCAGTCAACAACTGGAAAGATCCACACATACCGGCGTAAC
CGAGTCCAAACATACAGACCTGCAAGACTCGCGACATGGGACGGTAAAAC
CATCCGACCGTAAACCGGTAACCAGGTAGCCGGGTAAAAACATAGCAGGG
TGGAGACCTCAGAACGTAAAGACGTCCAAGGGTCGCCGGATAGCGAACTA
CGCGCATCGCCCAATGGGCCAATCAACAGATAAACGAGTAGAAAAGTCAG
AAAATAAGAAACTAACGAAATACGAGGGTCCAAGGATGCAAGACTGAGGC
CCTAAGGAGATAAGGAAATAGGCCGATGCAGACCTGAAACGATGCACCGA
TCCGACGGTAAAAGACTAGACACGTAGCCGGATCAGGGCCTGGGAGGCTG
GAACCGTGAGCACATAGCAAAGTCGCAGCGTCGGCAGATGCGCCGGTAAA
AAAGTAGAGGCATGACCGGATGGGCAAATAGCGACGTACAGCAGTGAAGC
ACTAAAAGCATCCAAGGGTAGGAGACTAGGCGCCTCGACGGGTAGGTACC
(SEQ ID NO: 1)

[00107] The synthetic nucleic acids of the present invention can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the label attachment region and the annealed patches or segments, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the synthetic nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl)uracil, (acp3)w, and 2,6-diaminopurine.

[00108] Alternatively, the synthetic nucleic acid can be produced biologically using a vector into which a nucleic acid has been subcloned.

[00109] In various embodiments, the synthetic nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 744

stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.*, 1996, Bioorganic & Medicinal Chemistry 4(1):5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.*, 1996, Bioorganic & Medicinal Chemistry 4(1): 5-23; Perry-O'Keefe *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93: 14670-675.

[00110] In an exemplary embodiment, the selected novel synthetic sequence can be constructed synthetically as double-stranded DNA by a commercial gene synthesis company and cloned in an oriented fashion into a "phagemid", a plasmid vector containing an M13 or f1 phage intergenic (IG) region which contains the *cis*-acting sequences necessary for DNA replication and phage encapsidation, such as pUC119. The appropriate orientation of the cloned insert relative to the phage origin of replication allows for the generation of a single-stranded DNA scaffold which is the reverse complement of the RNA molecules generated by *in vitro* transcription for each label attachment region.

[00111] By way of example, in order to generate the single-stranded DNA scaffold of the novel reporter, the phagemid is transformed into an *E. coli* strain containing an F' episome. Subsequent infection of the transformed bacteria with a helper phage such as the M13 mutant K07 results in the secretion of the phagemid carrying the novel reporter sequence as a single-stranded, packaged phage from which the circular, single-stranded DNA is prepared using a standard protocol. This DNA is linearized and the vector portion is excised by annealing short, complementary oligonucleotides to either end of the novel reporter sequence to generate double-stranded restriction sites, followed by treatment with the appropriate restriction enzymes.

[00112] By way of example, to make the RNA molecules (patches or "segments") for each label attachment region, PCR primers are designed to generate a double-stranded template beginning with an RNA polymerase promoter (T7, T3, or SP6) directly upstream (5') of the transcription start site and ending following the 3' restriction enzyme site. Using this template, *in vitro* transcription of RNA molecules is performed in the presence of amino-allyl modified regularly-repeated base in the RNA (*e.g.*, UTP) and unmodified other bases (*e.g.*, ATP, CTP and GTP). This leads to an RNA product in which every regularly-repeated base (*e.g.*, U) is modified to allow covalent coupling of a label monomer at that position in the RNA molecule.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 745

[00113] Coupling of light-emitting label monomers to the RNA molecules and annealing of the labeled RNA molecules to the scaffold are carried out as described below.

[00114] Some design considerations for the de novo sequence are listed in Table 1 below.

Table 1

Feature Of Synthetic Scaffold	Advantages
Novel synthetic sequence	Can be of any length and be designed to incorporate any desired sequence feature including but not limited to those listed in this table.
Minimal secondary structure (select against inverted repeats)	Allows for consistent transcription of full-length RNA molecules. Allows for consistent annealing of RNA molecules to scaffold at predictable temperatures. Minimizes self-annealing and/or cross-annealing between RNA molecules or scaffolds.
Minimal repeated sequences	Avoids mis-annealing between RNA molecules and inappropriate regions of the scaffold.
Unique restriction sites at borders of label attachment regions	Allows addition and deletion of individual label attachment regions using standard molecular cloning techniques.
Defined, even spacing of T's and transcription with amino-allyl-modified UTP (no unmodified UTP)	Controls number of coupling sites for monomers in each label attachment region, allowing for consistent brightness of individual labeled RNA molecules. Controls distance between monomers: spacing can be optimized to avoid stearic hindrance and fluorescence quenching.
Optimized start sequence for transcription by RNA polymerase T7, T3, or SP6	Promotes efficient in vitro transcription of each label attachment region.

5.3 PATCHES

[00115] Label monomers that emit signals that constitute all or part of the nanoreporter code are attached to label attachment region(s) of the nanoreporter scaffold through a structure referred to herein as a "patch." The label monomers can be directly (*e.g.*, covalently or noncovalently) attached to a patch, or indirectly attached to a patch (*e.g.*, through hybridization).

[00116] Nucleic acid patches can by anywhere from 25 nucleotides to several kilobases (*e.g.*, 5kb) in length, and are preferably 50 nucleotides to 2kb in length. In specific embodiments, nucleic acid patches are approximately 25-250, 50-200, 50-150, or 50-100 nucleotides in length. In other embodiments, nucleic acid patches are approximately 500-

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 746

2,000, 500-1,500, 500-1,000, 750-1,250, or 750-1,000 nucleotides in length. Nucleic acid patches can be RNA patches or DNA patches.

[00117] A label monomer can be covalently attached to a patch before or after the patch is attached to the label attachment region of a nanoreporter scaffold. For example, where the patch is a nucleic acid molecule, the label can be covalently attached by incorporation of a nucleotide containing a label monomer into the nucleic acid during its synthesis but before it is attached, *e.g.*, via hybridization, to the label attachment region of the scaffold. Alternatively, during the synthesis of a nucleic acid patch, a nucleotide containing a label monomer acceptor group be included, and the label monomer added to the nucleic acid patch after its synthesis, either before or after it is attached to the label attachment region of the scaffold. Alternatively, the label monomer can be indirectly attached to the patch, for example by hybridization of the patch to a "flap" that serves as a basis for attachment of the label monomer to the nanoreporter.

[00118] Thus, where a patch is a nucleic acid, it can range anywhere from 20 nucleotides to more than 5 kb in length, depending on the method of assembly of the nanoreporter.

[00119] For example, where a patch has covalently incorporated into it one or more label monomers that emit signals that art part of the nanoreporter code in the context of the labeled nanoreporter, it is preferably about 500-1500 nucleotides in length, and is generally referred to herein as a "segment," a "dark" segment being the patch prior to the incorporation of the label monomer (but, in a preferred embodiment, containing label monomer acceptor sites, such as amino allyl nucleotides), and a "colored" segment being one containing the desired label monomer or label monomers.

[00120] Where a patch merely serves as a template for flap attachment to the nanoreporter, then it is preferably smaller in size, for example about 25-250 nucleotides in length, and is most preferably about 50-100 nucleotides in length. Such patches are referred to herein as "oligonucleotide patches." As detailed in Section below, an oligonucleotide is preferably partially complimentary in sequence to a scaffold, such that when it is annealed to the scaffold, an overhang is generated that is complementary to all or portion of a flap.

[00121] The terms "segment" and "oligonucleotide patch" are used herein merely for convenience of description; however, there is no size cutoff to distinguish a "segment" from an "oligonucleotide patch." The purpose of both types of structures is to maximize the labeling -- and thus signal intensity-- from the nanoreporter, thereby allowing for single target molecule detection by a nanoreporter.

[00122] In certain aspects, the present invention provides a synthetic molecule, whose configuration is illustrated by reference to Figure 7A, comprising a strand of a nucleic acid

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 747

(scaffold) and a plurality of patch pairs hybridized to the strand, where each patch pair comprises an "A" patch and a "B" patch, and , for each patch pair, (a) each "A" patch is an oligonucleotide comprising a first region (1P) and a second region (2P), the first region being (i) at the alpha end of the "A" patch, and (ii) hybridized to a first portion of the strand, the second region being (ii) at the beta end of the "A" patch; (b) each "B" patch is an oligonucleotide comprising a third region (3P) and a fourth region (4P), the third region being (i) at the alpha end of the "B" patch, and (ii) hybridized to the second region of the "A" patch, the fourth region being (i) at the beta end of the "B" patch and (ii) hybridized to a second portion of the strand, the second portion of the strand being to the beta end of the first portion of the strand, where the second region or the third region further comprises at its beta end or alpha end, respectively, a hybridizable region that is not hybridized to the "B" patch or "A" patch, respectively.

[00123] In the synthetic molecule of Figure 7A, the second region may further comprise at its beta end a hybridizable region that is not hybridized to the "B" patch, as depicted in Figure 7B, or the third region further comprises at its alpha end a hybridizable region that is not hybridized to the "A" patch, as depicted in Figure 7C.

[00124] The present invention further provides a synthetic molecule, whose configuration is illustrated by reference to Figure 7D, comprising a strand of a nucleic acid (scaffold) and a plurality of patch pairs hybridized to the strand, where each patch pair comprises an "A" patch and a "B" patch, where, for each patch pair, (a) each "A" patch is an oligonucleotide comprising a first region (1P) and a second region (2P), the first region being (i) at the alpha end of the "A" patch, and (ii) hybridized to a first portion of the strand, the second region being (ii) at the beta end of the "A" patch; (b) each "B" patch is an oligonucleotide comprising a third region (3P) and a fourth region (4P), the third region being (i) at the alpha end of the "B" patch, and (ii) hybridized to the second region of the "A" patch, the fourth region being (i) at the beta end of the "B" patch and (ii) hybridized to a second portion of the strand, the second portion of the strand being to the first of the first portion of the strand, where the second region further comprises at its beta end a first hybridizable region that is not hybridized to the "B" patch, and where the third region further comprises at its alpha end a second hybridizable region that is not hybridized to the "A" patch.

[00125] In the synthetic molecule of Figure 7B, each patch pair can be attached to a flap pair, as depicted in Figure 7F, where each flap pair comprises an "A" flap and a "B" flap, where , for each flap pair, (a) each "A" flap is an oligonucleotide comprising a first flap region (1F) and a second flap region (2F); the first flap region being at the alpha end of the "A" flap; the second flap region (i) being at the beta end of the "A" flap and (ii) comprising at its beta

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 748

end a hybridizable region that is not hybridized to the "A" patch, "B" patch or "B" flap; and (b) each "B" flap is an oligonucleotide comprising a third flap region (3F), a fourth flap region (4F), and a fifth flap region (5F); the third flap region being (i) being at the alpha end of the "B" flap and (ii) comprising at its alpha end a hybridizable region that is not hybridized to the "A" patch, "B" patch or "A" flap; the fourth flap region (i) being between the third flap region and the fifth flap region and (ii) hybridized to the first flap region of the "A" flap; the fifth flap region being (i) at the beta end of the "B" flap, and (ii) hybridized to the hybridizable region of the second region of the "A" patch.

[00126] In the synthetic molecule of Figure 7C, each patch pair can be attached to a flap pair, as depicted in Figure 7E, where each flap pair comprises an "A" flap and a "B" flap, where, for each flap pair, (a) each "A" flap is an oligonucleotide comprising an first flap region (1F), a second flap region (2F), and a third flap region (3F); the "A" flap region being (i) at the alpha end of the "A" flap and (ii) hybridized to the hybridizable region of the third region of the "B" patch; the second flap region being between the first flap region and the third flap region; the third flap region (i) being at the beta end of the "A" flap and (ii) comprising at its beta end a hybridizable region that is not hybridized to the "A" patch, "B" patch or "B" flap, and (b) each "B" flap is an oligonucleotide comprising a fourth flap region (4F) and a fifth flap region (5F); the fourth flap region being (i) being at the alpha end of the "B" flap and (ii) comprising at its alpha end a hybridizable region that is not hybridized to the "A" patch, "B" patch or "A" flap; the fifth flap region being (i) at the beta end of the "B" flap, and (ii) hybridized to the second flap region of the "A" flap.

[00127] In the synthetic molecule of Figures 7D and 7E, the split flaps can be attached one (*e.g.*, (1O)), or more (*e.g.*, (2O) and (3O)) oligonucleotides, as depicted in Figure 7G. Thus, the one or more oligonucleotides can be attached to the all or a portion of the "A" flap individually (*e.g.*, (1O)), the "B" flap individually (*e.g.*, (3O)), or span all or a portion of each of the "A" flap and "B" flap (*e.g.*, (2O)). Such oligonucleotides are preferably covalently bound to one or more label monomers.

[00128] The hybridizable regions of the synthetic molecules may be hybridized to a plurality of oligonucleotides, each bound, preferably covalently bound, to at least one label monomer, more preferably to at least five label monomers. In certain embodiments, all the oligonucleotides attached to a single patch pair comprise the same label monomers, *e.g.*, comprise label monomers that emit light at the same wavelength(s); in specific embodiments, all the oligonucleotides attached to at least two, or at least four, adjacent patch pairs preferably comprise the same label monomers. One or more of the oligonucleotides may be bound to at least one affinity tag.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 749

[00129] In certain preferred embodiments, the label monomers are fluorophores or quantum dots.

[00130] In the synthetic molecule described above, alpha can refers to either 5'or 3', and the corresponding beta to either 3' or 5', respectively.

[00131] The region of complementary in each patch pair, or between a given patch and corresponding flap, is preferably 20-5,000 nucleotides. In certain embodiments, the region of complementary is 20-100 nucleotides, or 5-50 nucleotides.

[00132] In the synthetic molecules described above, each flap is preferably 50-5,000 nucleotides in length. In certain embodiments, each flap is 50-150 nucleotides.

[00133] The synthetic molecules described above may further comprise a target-specific region which binds to a target molecule. The target-specific region can be attached to the beta or alpha end of the strand.

[00134] In certain embodiments, the synthetic molecule described above comprise at least ten patch pairs, or at least fifty patch pairs.

[00135] In the synthetic molecules described above, the strand, or scaffold, can be a linearized vector, such as linearized M13.

[00136] The synthetic molecule described above may further comprise (a) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal; (b) a second label attachment region, which is non-overlapping with the first label attachment region, to which is attached one or more label monomers that emit light constituting a second signal; (c) a third label attachment region, which is non-overlapping with the first and second label attachment regions, to which is attached one or more label monomers that emit light constituting a third signal; where each attachment region comprises a plurality of patch pairs; wherein the first and second signals are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable; wherein the first and second signals are not spatially resolvable under conditions that can be used to detect the first, second and third signals; wherein the second and third signals are not spatially resolvable under conditions that can be used to detect the first, second and third signals; wherein the first and third signals are spatially resolvable under conditions that can be used to detect the first, second and third signals; and wherein the identities of the first, second and third signals and the locations of the first and third signal relative to each other constitute at least part of a code that identifies the target molecule.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 750

5.4 LABEL MONOMERS

[00137] The nanoreporters of the present invention can be labeled with any of a variety of label monomers, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other monomer known in the art that can be detected directly (*e.g.*, by light emission) or indirectly (*e.g.*, by binding of a fluorescently-labeled antibody). Generally, one or more of the label attachments regions in the nanoreporter is labeled with one or more label monomers, and the signals emitted by the label monomers attached to the label attachment regions of a nanoreporter constitute a code that identifies the target to which the target-specific region of the nanoreporter binds. In certain embodiments, the lack of a given signal from the label attachment region (*i.e.*, a "dark" spot) can also constitute part of the nanoreporter code. An example of a dark spot is depicted at position 12 of the nanoreporter in Figure 1A.

[00138] Radioisotopes are an example of label monomers that can be utilized by the invention. Several radioisotopes can be used as label monomers for labeling nucleotides or proteins, including, for example, ^{32}P, ^{33}P, ^{35}S, ^{3}H, and ^{125}I. These radioisotopes have different half-lives, types of decay, and levels of energy which can be tailored to match the needs of a particular experiment. For example, ^{3}H is a low energy emitter which results in low background levels, however this low energy also results in long time periods for autoradiography. Radioactively labeled ribonucleotides, deoxyribonucleotides and amino acids are commercially available. Nucleotides are available that are radioactively labeled at the first, or α, phosphate group, or the third, or γ, phosphate group. For example, both [α -^{32}P] dATP and [γ -^{32}P] dATP are commercially available. In addition, different specific activities for radioactively labeled nucleotides are also available commercially and can be tailored for different experiments.

[00139] Another example of label monomers that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, and Texas Red. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum. Also, different formulations of the same fluorophore have been produced for different applications. For example, fluorescein, can be used in its isothiocynanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or as isomeric dichlorotriazine forms of fluorescein (DTAF). These monomers are chemically distinct, but all emit light with a peak between 515-520 nm, thereby generating a similar signal. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 751

example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein. In addition, different formulations or related dyes have been developed around other fluorophores that emit light in other parts of the spectrum.

[00140] Non-radioactive and non-fluorescent label monomers are also available. For example, biotin can be attached directly to nucleotides and detected by specific and high affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin labeled nucleotides can also similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

[00141] Very small particles, termed nanoparticles, also can be used as label monomers to label nucleic acids. These particles range from 1-1000nm in size and include diverse chemical structures such as gold and silver particles and quantum dots.

[00142] When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle. Four to five different particles in close proximity will each scatter monochromatic light which when superimposed will give a specific, unique color. The particles are being manufactured by companies such as Genicon Sciences. Derivatized silver or gold particles can be attached to a broad array of molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivitized to allow attachment to a nucleotide.

[00143] Another type of nanoparticle that can be used as a label monomer are quantum dots. Quantum dots are fluorescing crystals 1-5nm in diameter that are excitable by a large range of wavelengths of light. These crystals emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties.

[00144] Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to desired size and/or emission wavelengths. Use of quantum dots for labeling particles, in the context of the present invention, is new, but is old in the art of semiconductors. Two examples of earlier references in which quantum dots are embedded

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 752

within intrinsic silicon epitaxial layers of semiconductor light emitting/detecting devices are United States Patent Nos. 5,293, 050 and 5,354,707 to Chapple Sokol, *et al.*, which are hereby incorporated by reference herein in their entireties.

[00145] In specific embodiments, one or more of the label attachments regions in the nanoreporter is labeled with one or more light-emitting dyes, each label attachment region containing, directly or indirectly, one or more label monomers. The light emitted by the dyes can be visible light or invisible light, such as ultraviolet or infra red light. In exemplary embodiments, the dye is a fluorescence resonance energy transfer (FRET) dye; a xanthene dye, such as fluorescein and rhodamine; a dye that has an amino group in the alpha or beta position (such as a naphthylamine dye, 1-dimethylaminonaphthyl-5-sulfonate, 1-anilino-8-naphthalende sulfonate and 2-p-touidinyl-6-naphthalene sulfonate); a dye that has 3-phenyl-7-isocyanatocoumarin; an acridine, such as 9-isothiocyanatoacridine and acridine orange; a pyrene, a bensoxadiazole and a stilbene; a dye that has 3-(ε-carboxypentyl)-3'-ethyl-5,5'-dimethyloxacarbocyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); 6-carboxy-X-rhodamine (ROX); 6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein (JOE); ALEXA Fluor™; Cy2; Texas Red and Rhodamine Red; 6-carboxy-2',4,7,7'-tetrachlorofluorescein (TET); 6-carboxy-2',4,4',5',7,7'-hexachlorofluorescein (HEX); 5-carboxy-2',4',5',7'-tetrachlorofluorescein (ZOE); NAN; NED; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; and Cy7.5; Alexa Fluor 350; Alexa Fluor 488; Alexa Fluor 532; Alexa Fluor 546; Alexa Fluor 568; Alexa Fluor 594; or Alexa Fluor 647.

[00146] The label monomers can be incorporated into a nanoreporter at different stages of its assembly, or into a component (*e.g.*, a "flap" of the nanoreporter prior to its assembly into the nanoreporter).

[00147] A label monomer can be directly attached to a nucleotide using methods well known in the art. Nucleotides can also be chemically modified or derivitized in order to attach a label monomer. For example, a fluorescent monomer such as a fluorescein molecule can be attached to dUTP (deoxyuridine-triphosphate) using a four-atom aminoalkynyl group. Each label monomer is attached to a nucleotide making a label monomer: nucleotide complex.

[00148] This label monomer: nucleotide complex can be incorporated into nucleic acids (for example, a DNA patch or a detection oligonucleotide) in a variety of ways. For example, a label monomer: nucleotide complex can be incorporated at only one location within a nucleic acid or at two or more locations within a nucleic acid.

[00149] Amine-reactive and thiol-reactive fluorophores are available and used for labeling nucleotides and biomolecules. Generally, nucleotides are fluorescently labeled during

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 753

chemical synthesis, for example, incorporation of amines or thiols during nucleotide synthesis permit addition of fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum.

[00150] A nucleotide can be attached to a label monomer first and then be incorporated into a nucleic acid. Alternatively, an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid. For example aminoallyl- ("AA-") modified UTP nucleotides can be incorporated into the RNA product during transcription. In various embodiments, 20% or more of UTP nucleotides in a transcription reaction to generate RNA patches are AA modified. In various embodiments, about 20%-100%, 20%-80%, 30-80%, 40-60% or 50%-75% of UTPs in a transcription reaction are AA-modified, in a preferred embodiment, approximately 50% of UTPs in a transcription reaction are AA-modified.

[00151] In addition, for example, different types of label monomer: nucleotide complexes can be incorporated into a single acid nucleic acid, where one component of the nanoreporter code comprises more than one type of signal.

[00152] Fluorescent dyes that can be bound directly to nucleotides can also be utilized as label monomers. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These fluorecently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-ddGTP and ROX-ddUTP, are commercially available.

[00153] Other types of label monomers that may be used to label a nanoreporter are quantum dots. Due to their very small size the quantum dots can be coupled into oligonucleotides directly without affecting the solubility or use of the oligonucleotide. In a preferred embodiment, only one oligonucleotide molecule is coupled to each nanoparticle. To synthesize an oligonucleotide-nanoparticle complex in a 1:1 ratio by conventional batch chemistry, both the oligonucleotide and the nanoparticle require a single reactive group of different kinds that can be reacted with each other. For example, if an oligonucleotide has an amino group and a nanoparticle has an aldehyde group, these groups can react to form a Schiff base. An oligonucleotide can be derivitized to attach a single amino or other functional group using chemistry well known in the art. However, when a nanoparticle is derivatized, it is covered with a chemical reagent which results in coating the entire surface of the nanoparticle with several functional groups.

[00154] The invention provides a method of coupling one oligonucleotide to one nanoparticle by chemically coupling the oligonucleotide on a solid surface such as the glass support used for the oligonucleotide synthesis.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 754

[00155] For example, commercially available resins for oligonucleotide synthesis such as long chain alkylamino controlled pore glass (lcaa CPG) can be used.

[00156] Alternatively, a flat surface such as a derivitized microscope slide can be used. The surface density of the nascent oligonucleotide chains should be lower than the diameter of the nanoparticle. This can be achieved by either choosing a glass support with low surface density of the reactive groups, or by using diluted reagent for the first step of the oligonucleotide synthesis so that the surface is not saturated. Another point of consideration when using the standard glass matrices for oligonucleotide synthesis is to use a pore diameter higher than the nanoparticle diameter to ensure the flow of the reagents. For example, an oligonucleotide can be synthesized on a diluted basis relative to the solid support, for example one tenth of a normal synthesis, to ensure good spacing of the oligonucleotides on the glass support. After the oligonucleotide is synthesized with a reactive functional group, for example, an amino group, derivitized nanoparticles are passed over the glass support to react with the oligonucleotides. A sufficiently large pore size of the glass support can be chosen to prevent clogging with nanoparticles. For example, a pore size of about 200 nm can be used. After the reaction is complete, un-reacted groups on the nanoparticle can be blocked and the complexes can be uncoupled from the glass support.

5.5 THE NANOREPORTER CODE

5.5.1 DUAL NANOREPORTERS

[00157] A nanoreporter whose components exist in two molecular entities is referred to as a dual nanoreporter. In a dual nanoreporter, generally each component contains a target-specific sequence, which improves the specificity of and binding kinetics of the nanoreporter to its target. The two different target-specific sequences are designed or selected such that each recognizes a different portion of a target molecule.

[00158] Figures 1A-1C illustrates embodiments of the invention involving dual nanoreporters. In Figures 1A and 1B, each of the two component of the nanoreporter is labeled, such that the nanoreporter's spectral code is formed only when the two components of the nanoreporter come together upon binding of the dual nanoreporter to its target molecule. However, in a dual nanoreporter, it is not necessary that both components are labeled. For example, as depicted in Figure 1C, one component of a dual nanoreporter is labeled with the nanoreporter code, and the other component attached to an affinity tag (arrow) that is useful to immobilize the nanoreporter for stretching and visualization.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 755

5.5.2 REGISTERS

[00159] The term "register" refers to a set of alternating (every other) label attachment regions. Registers are useful where it is desirable to label adjacent label attachment regions without a spacer region, and where the signal emanating from adjacent label attachment regions cannot be spatially resolved using the desired method of detection. Thus, the signals detected with use of a register is that form by the alternating, rather than adjacent, label attachment regions. Signals detected from a plurality of registers (*e.g.*, that together are all the label attachment regions) can be combined to form a nanoregister code. Generally when using registers, adjacent label attachment regions are labeled with spectrally distinguishable label monomers.

[00160] Examples of registers are depicted in Figures 3 and 5. For example, in Figures 3A-3B, there are eight label attachment regions 1-8. Alternating label attachment regions 1, 3, 5 and 7 form one register, and label attachment regions 2, 4, 6 and 8 form another register. In Figure 3A, only one of the registers (1, 3, 5 and 7) is labeled; in Figure 3B, both registers are labeled.

5.6 AFFINITY TAGS

[00161] A variety of affinity tags known in the art may be used to purify and/or immobilize nanoreporters. Where an affinity tag is used to immobilize a nanoreporter for the purpose of detection or imaging, it may be referred to herein as an "anchor." In a preferred embodiment, a biotin anchor is attached to the nanoreporter, allowing immobilization of the nanoreporter on a streptavidin coated slide.

[00162] Non-limiting examples of suitable affinity tags are provided below. It should be understood that most affinity tags could serve dual purposes: both as anchors for immobilization of the nanoreporters and tags for purification of the nanoreporters (whether fully or only partially assembled) or their components.

[00163] In certain embodiments, the affinity tag is a protein monomer. Examples of protein monomers include, but are not limited to, the immunoglobulin constant regions (see Petty, 1996, Metal-chelate affinity chromatography, in Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience), glutathione S-transferase (GST; Smith, 1993, Methods Mol. Cell Bio. 4:220-229), the *E. coli* maltose binding protein (Guan *et al.*, 1987, Gene 67:21-30), and various cellulose binding domains (U.S. Pat. Nos. 5,496,934; 5,202,247; 5,137,819; Tomme *et al.*, 1994, Protein Eng. 7:117-123), *etc.* Other affinity tags are recognized by specific binding partners and thus facilitate isolation and immobilization by affinity binding to the binding partner, which can be immobilized onto

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 756

a solid support. For example, the affinity tag can be an epitope, and the binding partner an antibody. Examples of such epitopes include, but are not limited to, the FLAG epitope, the myc epitope at amino acids 408-439, the influenza virus hemagglutinin (HA) epitope, or digoxigenin ("DIG"). In other embodiments, the affinity tag is a protein or amino acid sequence that is recognized by another protein or amino acid, for example the avidin/streptavidin and biotin.

[00164] In certain instances, the affinity tag can be used for labeling a nanoreporter in addition to purifying or immobilizing the nanoreporter. As will be appreciated by those skilled in the art, many methods can be used to obtain the coding region of the affinity tags, including but not limited to, DNA cloning, DNA amplification, and synthetic methods. Some of the affinity tags and reagents for their detection and isolation are available commercially.

5.7 TARGET-SPECIFIC SEQUENCES

[00165] The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter scaffold. The target specific sequence is generally an amino acid sequence (*i.e.*, a polypeptide or peptide sequence) or a nucleic acid sequence. In specific embodiments, where the target-specific sequence is an amino acid sequence, the target-specific sequence is an antibody fragment, such as an antibody Fab' fragment, a single chain Fv antibody.

[00166] The target-specific sequence is preferably a nucleic acid sequence, and is most preferably within an oligonucleotide that is either covalently attached (*e.g.*, by ligation) or noncovalently attached (*e.g.*, by hybridization) to the nanoreporter scaffold. A target-specific nucleic acid sequence is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific embodiments, the target-specific sequence is approximately 10-500, 20-400, 30-300, 40-200, or 50-100 nucleotides in length. In other embodiments, the target-specific sequence is approximately 30-70, 40-80, 50-90, or 60-100, 30-120, 40-140, or 50-150 nucleotides in length.

5.8 TARGET MOLECULES

[00167] The term "target molecule" refers to a molecule that is detected or measured by binding of a labeled nanoreporter whose target-specific sequence(s) recognize (are specific binding partners thereto). A target molecule can be, but is not limited to, any of the following: DNA, cDNA, RNA, mRNA, peptide, a polypeptide/protein (*e.g.*, a bacterial or viral protein or an antibody), a lipid, a carbohydrate, a glycoprotein, a glycolipid, a small molecule, an organic

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 757

monomer, or a drug. Generally, a target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.

[00168] A target molecule can be part of a biomolecular sample that contains other components or can be the sole or major component of the sample. A target molecule can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The target molecule can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also, the target molecule can have either a known or unknown structure or sequence.

[00169] In certain specific embodiments, that target molecule is not a chromosome. In other specific embodiments, the target molecule is no greater than 1,000 kb (or 1 mb) in size, no greater than 500 kb in size, no greater than 250 kb in size, no greater than 175 kb in size, no greater than 100 kb in size, no greater than 50 kb in size, no greater than 20 kb in size, or no greater than 10 kb in size. In yet other specific embodiments, the target molecule is isolated from its cellular milieu.

[00170] In specific, non-limiting embodiments, the target molecule is an antigen such as alpha fetoprotein, alpha-1 antitrypsin, α-2 macroglobulin, adiponectin, apoliprotein-A-1, apoliprotein-CIII, apoliprotein-H, BDNF, β-2 microglobulin, C reactive protein, calcitonin, cancer antigen 19-9, cancer antigen 125, CEA, CD 40, CD 40 ligand, complement 3, CK-MB, EGF, ENA-78, endothelin-1, enrage, eotaxin, erythropoietin, Factor VII, FABP, ferritin, FGF-basic, fibrinogen, G-CSF, GST, GM-CSF, growth hormone, haptoglobin, ICAM-1, IFN-gamma, IgA, IgE, IGF-1, IgM, IL-1α, IL-1β, IL-1ra, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12 p40, IL-12 p70, IL-13, IL-15, IL-16, insulin, leptin, lipoprotein (a), lymphotactin, MCP-1, MDC, MIP-1α, MIP-1β, MMP-2, MMP-3, MMP-9, myeloperoxidase, myoglobin, PAI-1, PAP, PAPP-A, SGOT, SHBG, PSA (free), RANTES, serum amyloid P, stem cell factor, TBG, thrombopoietin, TIMP-1, tissue factor, TNF-α, TNF-β, TNF RII, TSH, VCAM-1, VEGF, or vWF.

[00171] In some embodiments, the target molecule is an autoimmune related molecule such as ASCA, β-2 glycoprotein, C1q, centromere Prot. B, collagen type 1, collagen type 2, collagen type 4, collagen type 6, Cyto P450, ds DNA, histone, histone H1, histone H2A, histone H2B, histone H3, histone H4, HSC-70, HSP-32, HSP-65, HSP-71, HSP-90α, HSP-90β, insulin, JO-1, mitochondrial, myeloperoxidase, pancreatic islet cells, PCNA, PM-1, PR3, ribosomal P, RNP-A, RNP-C, RNP, Sel-70, Smith, SSA, SSB, T3, T4, thyroglobulin, tTG, (celiac disease), or thyroid microsomal.

[00172] In some embodiments, the target molecule is a component isolated from an infectious disease, such as Cholera Toxin, Cholera Toxin β, *Campylobacter jejuni*,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 758

cytomegalovirus, Diptheria toxin, Epstein-Barr NA, Epstein-Barr EA, Epstein-Barr VCA, *Heliobacter pylori*, HBV core, HBV envelope, HBV surface (Ad), HBV surface (Ay), HCV core, HCV NS3, HCV NS4, HCV NS5, hepatitis A, hepatitis D, HEV orf2 3KD, HEV orf2 6 KD, HEV orf 3KD, HIV-1 p24, HIV-1 gp41, HIV-1 gp120, HPV, HSV-1/2, HSV-1 gD, HSV-2 gD, HTLV-1/2, influenza A, influenza A H3N2, influenza B, *Leishmania donorani*, Lyme disease, mumps, *M. pneumonia*, *M. tuberculosis*, parainfluenza 1, parainfluenza 2, parainfluenza 3, polio virus, RSV, Rubella, Rubeola, Streptolysin O, Tetanus Toxin, *T. pallidum* 15kD, *T. pallidum* p47, *T. cruzi*, Toxoplasma, Varicella zoster.

5.9 NANOREPORTER POPULATIONS

[00173] The present invention provides nanoreporter or nanoreporter label unit populations, for example nanoreporter or nanoreporter label unit libraries, that contain at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, or at least 1,000 unique nanoreporters or nanoreporter label units, respectively. As used herein, "unique" when used in reference to a nanoreporter or nanoreporter label units within a population is intended to mean a nanoreporter or label unit that has a code that distinguishes it from other nanoreporters or label units in the same population.

[00174] In specific embodiments, the present invention provides nanoreporter populations with at least 5,000, at least 10,000, at least 20,000 or at least 50,000 unique nanoreporters or nanoreporter label units. The nanoreporters in a population of nanoreporters can be singular nanoreporters, dual nanoreporters, or a combination thereof. The nanoreporters can be labeled or unlabeled.

[00175] The size of a nanoreporter population and the nature of the target-specific sequences of the nanoreporters within it will depend on the intended use of the nanoreporter. Nanoreporter populations can be made in which the target-specific sequences correspond to markers of a given cell type, including a diseased cell type. In certain embodiments, nanoreporters populations are generated in which the target-specific sequences represent at least 0.1%, at least 0.25%, at least 0.5%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the different type of transcripts in a cell. In certain embodiments, nanoreporters populations are generated in which the target-specific sequences represent at least 0.1%, at least 0.25%, at least 0.5%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the different genes in a cell. In yet

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 759

other embodiments, nanoreporter populations are generated in which at least some of the target-specific sequences represent rare transcripts in a cell or tissue. Such nanoreporter populations preferably represent at least 5 rare transcripts. In specific embodiments, such nanoreporter populations represent at least 10, at least 20, at least 30, at least 40 or at least 50 rare transcripts. In a specific embodiment, the cell or tissue is a mammalian cell or tissue, and more preferably is a human cell or tissue.

[00176] In certain embodiments, the nanoreporter population is a diagnostic or prognostic nanoreporter population. For example, a diagnostic nanoreporter population can be generated that is useful for screening blood products, in which the target-specific sequences bind to the nucleic acids of contaminating viruses such the hepatitis B, hepatitis C, and the human immunodeficiency virus. Alternatively, the diagnostic nanoreporter population may contain target-specific sequences corresponding to cellular disease markers, such as tumor antigens. Prognostic nanoreporter populations generally include target-specific markers that represent different stages of a given disease such as cancer. By selecting appropriate target-specific sequences, a nanoreporter population can be used both to diagnose and prognose disease.

5.10 BIOMOLECULAR SAMPLES

[00177] The nanoreporter systems of the invention can be used to detect target molecules in any biomolecular sample. As will be appreciated by those skilled in the art, the sample may comprise any number of things, including, but not limited to, cells (including both primary cells and cultured cell lines), tissues and bodily fluids (including, but not limited to, blood, urine, serum, lymph, bile, cerebrospinal fluid, interstitial fluid, aqueous or vitreous humor, colostrum, sputum, amniotic fluid, saliva, anal and vaginal secretions, perspiration and semen), a transudate, an exudate (*e.g.*, fluid obtained from an abscess or any other site of infection or inflammation) or fluid obtained from a joint (*e.g.*, a normal joint or a joint affected by disease such as rheumatoid arthritis, osteoarthritis, gout or septic arthritis) of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred; environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples including extracellular fluids, extracellular supernatants from cell cultures, inclusion bodies in bacteria, cellular compartments, cellular periplasm, mitochondria compartment, etc.

[00178] The biomolecular samples can be indirectly derived from biological specimens. For example, where the target molecule of interest is a cellular transcript, *e.g.*, a messenger RNA, the biomolecular sample of the invention can be a sample containing cDNA produced by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 760

a reverse transcription of messenger RNA. In another example, the biomolecular sample of the invention is generated by subjecting a biological specimen to fractionation, *e.g.*, size fractionation or membrane fractionation.

[00179] The biomolecular samples of the invention may be either "native," *i.e.*, not subject to manipulation or treatment, or "treated," which can include any number of treatments, including exposure to candidate agents including drugs, genetic engineering (*e.g.*, the addition or deletion of a gene), *etc.*

5.11 SEPARATION OF LABEL MONOMERS

[00180] In addition to detecting an overall signal generated from a labeled nanoreporter, the invention provides for the determination of the spatial location of signals emanating from the label monomers (*i.e.*, spots) on a nanoreporter, each spot representing the aggregate signal from label monomers attached to a given label attachment region. A spot may contain signals of the same wavelength or of different wavelengths. Thus, the nature of the spots on a nanoreporter and their location constitutes the nanoreporter code.

[00181] Any of a variety of means can be used to "stretch" the nanoreporter to separate the individual spots. For example, a nanoreporter can be stretched using a flowstretch technique (Henegariu *et al.*, 2001, Biotechniques 31:246-250), a receding meniscus technique (Yokota *et al.*, 1997, Nuc. Acids Res. 25:1064-1070) or an electrostretching technique (Matsuura *et al.*, 2001, Nuc. Acids Res. 29: E79).

[00182] The use of flow-stretching, receding meniscus, or electro-stretching techniques allows for the separation of the label attachment regions within a nanoreporter so that one can determine spatially where a particular signal is positioned in the nanoreporter. Therefore, unique nanoreporters that have the same combination of label monomers and the same overall signal can be differentiated from one another based on the location of those label monomers within the nanoreporter.

[00183] This ability to locate the position of a label attachment region or spot within a nanoreporter allows for the position of the signal(s) emitted by the label monomers in each label attachment region to be used as a distinguishing characteristic when generating a set of unique nanoreporters. Hence, a complex set of nanoreporters can be generated using the same combination of starting label monomers by varying the positions of the label monomers within a nanoreporter.

[00184] Prior to stretching a nanoreporter, it is preferable to immobilize the nanoreporter to a solid surface using an affinity tag, as described in Section 5.6 above. In certain aspects of the invention, one end of a nanoreporter is immobilized, either through specific or non-specific

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 761

binding to a solid surface, the nanoreporter is stretched, and then the other end of the reporter is immobilized, also either through specific or non-specific binding to a solid surface. Accordingly, the nanoreporter is "frozen" in its stretched, or extended, state, to facilitate resolution of the nanoreporters code by detecting and/or imaging the signals emitted by the label monomers attached to a nanoreporter and their locations relative to one another. These aspects of the invention are described below.

5.12 NANOREPORTERS

[00185] In the methods, the nanoreporters are certain types of macromolecules. In certain embodiments, the macromolecule is a macromolecule that is capable of being extended in the methods of the invention. In certain embodiments, the macromolecule is capable of being immobilized in one or two portions as described in the sections below.

[00186] In certain embodiments, the nanoreporter is a polysaccharide, a polypeptide or a polynucleotide. Useful polynucleotides include ribonucleic acids, deoxyribonucleic acids and other polynucleotides known to those of skill in the art.

[00187] The nanoreporter can be of any size that is sufficient to allow extension and immobilization of the nanoreporter according to the methods of the invention. In certain embodiments when the nanoreporter is a polynucleotide, the nanoreporter can have a length of greater than 500 bp, greater than 750 bp, greater than 1 kb, greater than 1.5 kb, greater than 2.0 kb, greater than 2.5 kb, greater than 3.0 kb, greater than 4.0 kb or greater than 5.0 kb. In certain embodiments, when the nanoreporter is a polypeptide, the nanoreporter can have a size of greater than 50 amino acids, greater than 100 amino acids, greater than 200 amino acids, greater than 300 amino acids, greater than 400 amino acids, greater than 500 amino acids, greater than 750 amino acids, greater than 1000 amino acids, greater than 1500 amino acids, greater than 2000 amino acids, greater than 2500 amino acids, greater than 3000 amino acids, greater than 4000 amino acids or greater than 5000 amino acids. In certain embodiments, when the nanoreporter is a polysaccharide, the nanoreporter can have a size of greater than 50 saccharides, greater than 100 saccharides, greater than 200 saccharides, greater than 300 saccharides, greater than 400 saccharides, greater than 500 saccharides, greater than 750 saccharides, greater than 1000 saccharides, greater than 1500 saccharides, greater than 2000 saccharides, greater than 2500 saccharides, greater than 3000 saccharides, greater than 4000 saccharides or greater than 5000 saccharides.

[00188] The nanoreporter can be a native nanoreporter as understood by those of skill in the art, or the nanoreporter can be a non-native nanoreporter. In certain embodiments, when the nanoreporter is a polypeptide, the nanoreporter can comprise only naturally occurring

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 762

amino acids, or the nanoreporter can comprise naturally occurring amino acids and non-naturally occurring amino acids. The other amino acids can be any amino acids, or derivatives or analogs thereof, known to those of skill in the art. In certain embodiments, when the nanoreporter is a polynucleotide, the polynucleotide can comprise only naturally occurring nucleotides, or the polynucleotide can comprise naturally occurring nucleotides and non-naturally occurring nucleotides. In certain embodiments, when the nanoreporter is a polysaccharide, the polysaccharide can comprise only naturally occurring saccharides, or the polysaccharide can comprise naturally occurring saccharides and non-naturally occurring saccharides. In certain embodiments, the polymers can comprise only non-natural monomers. In further embodiments, the nanoreporter can comprise a plurality of classes of monomers, such as amino acids, nucleotides and/or saccharides.

[00189] In certain embodiments, the nanoreporter comprises only one primary, covalently linked chain of monomers. For instance, when the nanoreporter is a polypeptide, in certain embodiments, the nanoreporter comprises only one primary amino acid chain. When the nanoreporter is a polynucleotide, in certain embodiments, the nanoreporter is single stranded. In further embodiments, the nanoreporter comprises two primary, covalently linked chains of monomers. For instance, when the nanoreporter is a polypeptide, in certain embodiments, the nanoreporter comprises two primary amino acid chains. When the nanoreporter is a polynucleotide, in certain embodiments, the nanoreporter comprises two polynucleotide strands; in certain embodiments, the nanoreporter can be double stranded, in part or in whole. In further embodiments, the nanoreporter comprises three or more primary, covalently linked chains of monomers. For instance, when the nanoreporter is a polypeptide, in certain embodiments, the nanoreporter comprises three primary amino acid chains. When the nanoreporter is a polynucleotide, in certain embodiments, the nanoreporter comprises three polynucleotide strands. For instance, the nanoreporter can comprise three strands F1, X and F2 where a portion of strand X is complementary to strand F1 and a portion of strand X is complementary to strand F2. An example is illustrated in Figure 13A. In certain embodiments, the nanoreporter comprises more than three primary, covalently linked chains of monomers.

[00190] Advantageously, a nanoreporter of the invention can comprise one or more labels that facilitate the detection, imaging or identification of the nanoreporter by techniques known to those of skill in the art. The label can be any detectable moiety known to those of skill in the art. Exemplary labels for nanoreporters include detectable isotopes, radioisotopes, fluors, dyes, enzymes, ligands, receptors, antigens, antibodies, lectins, carbohydrates, nucleotide sequences, and any other detectable label apparent to those of skill in the art.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 763

Useful labels, macromolecules comprising labels, and methods of their preparation are described in U.S. provisional patent application no. 60/753,758, entitled "Nanoreporters and Methods of Manufacturing and Use Thereof," filed December 23, 2005, the contents of which are hereby incorporated by reference herein in their entirety.

[00191] In certain embodiments, a polynucleotide is a polymer of natural (*e.g.*, A, G, C, T, U) or synthetic nucleobases, or a combination of both. The backbone of the polynucleotide can be composed entirely of "native" phosphodiester linkages, or it may contain one or more modified linkages, such as one or more phosphorothioate, phosphorodithioate, phosphoramidate or other modified linkages. As a specific example, a polynucleotide may be a peptide nucleic acid (PNA), which contains amide interlinkages. Additional examples of synthetic bases and backbones that can be used in conjunction with the invention, as well as methods for their synthesis can be found, for example, in U.S. Patent No. 6,001,983; Uhlman & Peyman, 1990, *Chemical Review* 90(4):544 584; Goodchild, 1990, *Bioconjugate Chem.* 1(3):165 186; Egholm *et al.*, 1992, *J. Am. Chem. Soc.* 114:1895 1897; Gryaznov *et al.*, *J. Am. Chem. Soc.* 116:3143 3144, as well as the references cited in all of the above. Common synthetic nucleobases of which polynucleotides may be composed include 3-methlyuracil, 5,6-dihydrouracil, 4 thiouracil, 5 bromouracil, 5-thorouracil, 5-iodouracil, 6-dimethyl aminopurine, 6-methyl aminopurine, 2-aminopurine, 2,6-diamino purine, 6-amino-8-bromopurine, inosine, 5-methylcytosine, 7-deazaadenine, and 7-deazaguanosine. Additional non-limiting examples of synthetic nucleobases of which the target nucleic acid may be composed can be found in Fasman, *CRC Practical Handbook of Biochemistry and Molecular Biology*, 1985, pp. 385-392; *Beilstein's Handbuch der Organischen Chemie*, Springer Verlag, Berlin and Chemical Abstracts, all of which provide references to publications describing the structures, properties and preparation of such nucleobases.

[00192] The nanoreporter can be prepared according to any technique apparent to those of skill in the art. Advantageously, nanoreporterss according to the invention can comprise labels and/or members of binding pairs, as described in the sections below, that can be used to facilitate preparation and/or purification of the nanoreporter. In addition, certain nanoreporters of the invention are capable of forming complexes with molecules that comprise members of binding pairs, as described below. These complexes can be used to facilitate preparation and/or purification of the nanoreporter or complex.

5.13 IMMOBILIZATION OF STRETCHED NANOREPORTERS

[00193] A macromolecule can be selectively immobilized while fully extended under whatever force is used for the extension. In addition, the methods of the invention facilitate the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 764

selective immobilization of extended nanoreporters that are oriented with respect to each other. In other words, according to the methods of the invention, a plurality of nanoreporters can readily be immobilized in the same orientation with respect to each other.

[00194] In one aspect, the present invention provides methods for selectively immobilizing a nanoreporter in an extended state. The macromolecule can be any macromolecule known to those of skill in the art such as a polymer, a polysaccharide, a polynucleotide or a polypeptide. For the methods of this aspect of the invention, generally, a first portion of the nanoreporter is immobilized by any technique known to those of skill in the art. Indeed, the technique for immobilizing the first portion of the nanoreporter is not critical to many embodiments of the invention. In certain embodiments, the first portion of the nanoreporter can be immobilized selectively or non-selectively. In certain embodiments the first portion is immobilized by one or more covalent bonds. In certain embodiments, the first portion is immobilized by one or more non-covalent bonds.

[00195] With an immobilized first portion, the nanoreporter can be extended by any technique for extending a nanoreporter apparent to those of skill in the art. In certain embodiments, the technique for extending the nanoreporter is not critical for the methods of the invention. In certain embodiments, the technique for extending the nanoreporter appropriate for the class of nanoreporter according to the judgment of one of skill in the art. In certain embodiments, the nanoreporter is extended by application of a force capable of extending the nanoreporter. The force can be any force apparent to one of skill in the art for extending the nanoreporter. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the nanoreporter are described in the sections below.

[00196] The nanoreporter is in an extended state if it would be recognized as extended by one of skill in the art. In certain embodiments, the nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter. In certain embodiments, the nanoreporter is in an extended state when its average hydrodynamic radius is more than double the average hydrodynamic radius of the nanoreporter in its native state as recognized by those of skill in the art.

[00197] In this aspect of the invention, the methods generally comprise the step of selectively immobilizing a second portion of the nanoreporter while it is in an extended state. This can result in an immobilized nanoreporter that is extended between the first and the second portion. Remarkably, since the nanoreporter is selectively immobilized while extended, that extension can be preserved in the immobilized nanoreporter. Generally, the first portion and the second portion of the nanoreporter are not the same.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 765

[00198] The selective immobilization can be according to any technique for selective immobilization of a portion of a nanoreporter apparent to those of skill in the art. The selective immobilization can be through, for example, the formation of one or more covalent bonds or one or more non-covalent bonds, or both. Particular examples of selective immobilization techniques are described in the sections below. In particular embodiments, one or more binding pairs are used to immobilize the second portion of the nanoreporter.

[00199] The second portion can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any substrate judged to be useful for immobilization known to those of skill in the art. In certain embodiments, the second portion can be immobilized to another molecule. Further useful substrates include surfaces, membranes, beads, porous materials, electrodes, arrays and any other substrate apparent to those of skill in the art.

[00200] In another aspect, the present invention provides compositions comprising a selectively immobilized, extended nanoreporter. The compositions generally comprise a substrate and an extended nanoreporter selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the nanoreporter are immobilized onto the substrate, and the nanoreporter is in an extended state between the two portions. In certain embodiments, at least one portion of the nanoreporter is selectively immobilized onto the substrate. In certain embodiments, two or more portions of the nanoreporter are selectively immobilized onto the substrate. The nanoreporter can be extended and/or immobilized by any technique apparent to those of skill, including particularly the methods of the present invention.

[00201] In another aspect, the present invention provides methods for selectively immobilizing a nanoreporter in an oriented state. The nanoreporter can be any nanoreporter described above. In certain embodiments, the nanoreporter can be flexible, or in certain embodiments the nanoreporter can be rigid or semi-rigid. For the methods of this aspect of the invention, generally, a first portion of the nanoreporter is immobilized as described above. With an immobilized first portion, the nanoreporter can be oriented by any technique for extending a nanoreporter apparent to those of skill in the art. In certain embodiments, the technique for orienting the nanoreporter is not critical for the methods of the invention. In certain embodiments, the technique for orienting the nanoreporter appropriate for the class of nanoreporter according to the judgment of one of skill in the art. In certain embodiments, the nanoreporter is oriented by application of a force capable of orienting the nanoreporter. The force can be any force apparent to one of skill in the art for orienting the nanoreporter.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 766

Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof.

[00202] The nanoreporter is in an oriented state if it would be recognized as oriented by one of skill in the art. In certain embodiments, the nanoreporter is in an oriented state when it is in the field of a force capable of orienting the nanoreporter. In certain embodiments, the nanoreporter is in an oriented state when its termini are arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the nanoreporter. In certain embodiments, a plurality of nanoreporterss is in an oriented state when the termini of the nanoreporterss are arranged in parallel, as recognized by those of skill in the art.

[00203] In this aspect of the invention, the methods generally comprise the step of selectively immobilizing a second portion of the nanoreporter while it is in an oriented state. This can result in an immobilized nanoreporter that is oriented between the first and the second portion. Remarkably, since the nanoreporter is selectively immobilized while extended, that orientation can be preserved in the immobilized nanoreporter. The selective immobilization can be according to the methods described above.

[00204] In another aspect, the present invention provides compositions comprising a selectively immobilized, oriented nanoreporter. The compositions generally comprise a substrate and an oriented nanoreporter selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the nanoreporter are immobilized onto the substrate, and the nanoreporter is in an oriented state between the two portions. In certain embodiments, at least one portion of the nanoreporter is selectively immobilized onto the substrate. In certain embodiments, both portions of the nanoreporter are selectively immobilized onto the substrate. The nanoreporter can be oriented and/or immobilized by any technique apparent to those of skill, including particularly the methods of the present invention.

[00205] The methods and compositions of the present invention can be used for any purpose apparent to those of skill in the art. For instance, the immobilized and extended and/or oriented nanoreporter can be used as a label for a substrate on which the nanoreporter is immobilized. The primary sequence of the immobilized and extended and/or oriented nanoreporter can be identified by any technique apparent to those of skill in the art. Advantageously, immobilization of the extended and/or oriented nanoreporter can facilitate such techniques. In certain embodiments, the immobilized and extended and/or oriented nanoreporter can be used to guide the manufacture of nanopaths, for example to create

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 767

nanowires or nanocircuits. Further uses for the immobilized and extended and/or oriented nanoreporterss are described in the sections below.

5.13.1 METHODS OF SELECTIVE IMMOBILIZATION

[00206] As described above, the present invention provides methods for the selective immobilization of a nanonreporter in an extended state. The nanoreporter, once selectively immobilized, can be used for any purpose apparent to those of skill in the art.

5.13.2 IMMOBILIZATION OF FIRST PORTION

[00207] In the methods of the invention, a first portion of the nanoreporter is immobilized. Generally, the first portion is immobilized if it would be recognized as immobilized by one of skill in the art. The first portion can be immobilized by any technique apparent to those of skill in the art. In certain embodiments, the technique for immobilization of the first portion of the nanoreporter is not critical for the methods of the invention.

[00208] The first portion of the nanoreporter can be at any location in the nanoreporter. In certain embodiments, the first portion is at a terminus of the nanoreporter. For the purposes of the invention, a portion of a nanoreporter can be "at a terminus" when it is less than five, four, three, two, one or zero monomers from a terminus of the nanoreporter. Of course, although many nanoreporters have two termini, the methods of the invention are applicable to nanoreporters that have more than two termini and to nanoreporters having one or zero termini, *e.g.*, circular nanoreporters. In certain embodiments, the first portion is not at a terminus of the nanoreporter.

[00209] The nanoreporter can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any moiety to which the nanoreporter can be immobilized without limitation. In certain embodiments, the substrate is a surface, membrane, bead, porous material, electrode or array.

[00210] In certain embodiments, the first portion of the nanoreporter can be immobilized non-selectively. In further embodiments, the first portion of the nanoreporter can be immobilized selectively. In advantageous embodiments, after the first portion of the nanoreporter is immobilized, some portion of the nanoreporter should be free to move sufficiently so that the nanoreporter can be extended and/or oriented in the following steps of the method. In particular, in certain embodiments, when the first portion of the nanoreporter is immobilized non-selectively, it is important that the entire nanoreporter not be immobilized non-selectively to an extent that prevents extension of any portion of the nanoreporter.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 768

[00211] The immobilization can be by any interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain embodiments, the immobilization can be via electrostatic interaction with an electrode. In further embodiments, the immobilization is via electrostatic interaction with a substrate other than the electrode.

[00212] In certain embodiments, the first portion of the nanoreporter comprises a first member of a binding pair. The first member of the binding pair can be covalently bound to the first portion of the nanoreporter, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. In useful embodiments, the substrate onto which the first portion of the nanoreporter is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound. Figure 12 illustrates a nanoreporter that comprises a moiety F1 that is capable of selectively binding a moiety of the substrate. Moiety F1 can be, for example, biotin, capable of binding, for example, a substrate coated with avidin.

[00213] In certain embodiments, the first portion of the nanoreporter can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins, and antibodies. The first portion of the nanoreporter would comprise a binding moiety capable of binding with the binding moiety of the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, surfaces comprising epoxy, aldehyde, gold, hydrazide, sulfhydryl, NHS-ester, amine, thiol, carboxylate, maleimide, hydroxymethyl phosphine, imidoester, isocyanate, hydroxyl, pentafluorophenyl-ester, psoralen, pyridyl disulfide or vinyl sulfone, or mixtures thereof. Such surfaces can be obtained from commercial sources or prepared according to standard techniques.

[00214] In advantageous embodiments, the first portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. Similarly, a polypeptide nanoreporter can comprise a biotinylated amino acid residue. The substrate comprising avidin can be any substrate comprising avidin known to those of skill in the art. Useful substrates comprising avidin are commercially available including TB0200 (Accelr8), SAD6, SAD20,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 769

SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), streptavidin slide (catalog #MPC 000, Xenopore) and STREPTAVIDINnslide (catalog #439003, Greiner Bio-one).

[00215] In certain embodiments, the first portion of the nanoreporter can comprise a nucleotide sequence that is capable of selectively binding a nucleotide sequence on the substrate. In further embodiments, the first portion of the nanoreporter can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin that are commercially available include, but are not limited to, Optiarray-biotin (Accler8), BD6, BD20, BD100, BD500 and BD2000 (Xantec).

[00216] In further embodiments, the first portion of the nanoreporter is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, a first portion of the nanoreporter can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. Figure 13A illustrates a nanoreporter that is capable of selectively binding a second molecule X that is capable of selectively binding a third molecule that comprises F1. F1 is capable of selectively binding a moiety on a substrate. Figure 13B illustrates a nanoreporter that is capable of selectively binding a second molecule that comprises F1, and F1 is capable of selectively binding a moiety on a substrate.

[00217] In further embodiments, the first portion of the nanoreporter can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Usesful substrates comprising reactive groups include those that comprise a reactive moiety selected from the group consisting of succinamides, amines, aldehydes, epoxies and thiols. Exemplary useful substrates comprising reactive moieties include, but are not limited to, surfaces comprising epoxy, aldehyde, gold, hydrazide, sulfhydryl, NHS-ester, amine, thiol, carboxylate, maleimide, hydroxymethyl phosphine, imidoester, isocyanate, hydroxyl, pentafluorophenyl-ester, psoralen, pyridyl disulfide or vinyl sulfone, or mixtures thereof. Such surfaces can be obtained from commercial sources or prepared according to standard techniques. The first portion of the nanoreporter would comprise a reactive moiety capable of reacting with the reactive moiety of the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, OptArray-DNA NHS group (Accler8), Nexterion Slide AL (Schott) and Nexterion Slide E (Schott).

[00218] In certain embodiments, the first portion of the nanoreporter can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the first portion of the nanoreporter

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 770

could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

[00219] In further embodiments, the first portion of the nanoreporter can be immobilized to the substrate via other binding pairs apparent to those of skill in the art.

5.13.3 EXTENSION OF THE NANOREPORTER

[00220] In certain methods of the invention, the nanoreporter is in an extended state. Generally, any nanoreporter is in an extended state if it would be recognized as such by one of skill in the art.

[00221] In certain embodiments, the nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter under conditions suitable for extending the nanoreporter. Such forces and conditions should be apparent to those of skill in the art. For instance, many nanoreporter can be extended by hydrodynamic force or by gravity, and many charged nanoreporter can be extended by electromagnetic force. In certain embodiments, the force can be applied to the nanoreporter indirectly. For instance, the nanoreporter can comprise or can be linked, covalently or noncovalently, to a moiety capable of being moved by a force. In certain embodiments, the nanoreporter can be linked to a moiety capable of being moved by electromagnetic, hydrodynamic or optical force.

[00222] In certain embodiments, the force is an electromagnetic force. For instance, when the nanoreporter is charged, such as a polynucleotide, the nanoreporter can be extended in an electric or magnetic field. The field should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter in an electric or magnetic field are described in Matsuura *et al.*, 2002, *J Biomol Struct Dyn.* 20(3):429-36; Ferree & Blanch, 2003, *Biophys J.* 85(4):2539-46; Stigter & Bustamante, 1998, *Biophys J.* 1998 75(3):1197-210; Matsuura *et al.*, 2001, *Nucleic Acids Res.* 29(16); Ferree & Blanch, 2004, *Biophys J.* 87(1):468-75; the contents of which are hereby incorporated by reference in their entirety.

[00223] In certain embodiments, the force is a hydrodynamic force. For instance, many nanoreporters, including polysaccharides, polypeptides, and polynucleotides, can be extended in the field of a moving fluid. The hydrodynamic force should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter in hydrodynamic field are described in Bensimon *et al.*, 1994, *Science* 265:2096-2098; Henegariu *et al.*, 2001, *BioTechniques* 31: 246-250; Kraus *et al.*,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 771

1997, *Human Genetics* 99:374 - 380; Michalet *et al.*, 1997, *Science* 277:1518-1523; Yokota *et al.*, 1997, *Nucleic Acids Res.* 25(5):1064-70; Otobe *et al.*, 2001, *Nucleic Acids Research* 29:109; Zimmerman & Cox, 1994, *Nucleic Acids Res.* 22(3):492-7, and U.S. Patent Nos. 6,548,255, 6,344,319, 6,303,296, 6,265,153, 6,225,055, 6,054,327, 5,840,862, the contents of which are hereby incorporated by reference in their entirety.

[00224] In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force to extend the nanoreporter. In certain embodiments, The force should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter with gravity are described in Michalet *et al.*, 1997, *Science* 277:1518-1523; Yokota *et al.*, 1997, *Nucleic Acids Res.* 25(5):1064-70; Kraus *et al.*, 1997, *Human Genetics* 99:374 - 380, the contents of which are hereby incorporated by reference in their entirety.

[00225] In particular embodiments, the force is applied through a moving meniscus. Those of skill in the art will recognize that a moving meniscus can apply various forces to a nanoreporter including hydrodynamic force, surface tension and/or any other force recognized by those of skill in the art. The meniscus can be moved by any technique apparent to those of skill in the art including evaporation and gravity. Exemplary techniques for extending a nanoreporter with a moving meniscus are described in, for example, U.S. Patent Nos. 6,548,255, 6,344,319, 6,303,296, 6,265,153, 6,225,055, 6,054,327, 5,840,862, the contents of which are hereby incorporated by reference in their entireties.

[00226] In particular embodiments, the nanoreporter can be extended by an optical trap or optical tweezers. For instance, the nanoreporter can comprise or can be linked, covalently or noncovalently, to a particle capable of being trapped or moved by an appropriate source of optical force. Useful techniques for moving particles with optical traps or optical tweezers are described in Ashkin *et al.*, 1986, *Optics Letters* 11:288-290; Ashkin *et al.*, 1987, *Science* 235:1517–1520; Ashkin *et al.*, Nature 330:769–771; Perkins *et al.*, 1994, *Science* 264:822-826; Simmons *et al.*, 1996, *Biophysical Journal* 70:1813-1822; Block *et al.*, 1990, *Nature* 348:348 - 352; and Grier, 2003, *Nature* 424: 810-816; the contents of which are hereby incorporated by reference in their entireties.

[00227] In certain embodiments, the nanoreporter can be extended by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain nanoreporters are extended by a combination of an electric field and hydrodynamic force.

[00228] The nanoreporter is extended when it would be recognized as extended by one of skill in the art according to standard criteria for extension of a nanoreporter. In certain

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 772

embodiments, the nanoreporter is extended when it loses most of its tertiary structural features as recognized by those of skill in the art. In certain embodiments, the nanoreporter is extended when it loses most of its secondary structural features as recognized by those of skill in the art. In certain embodiments, the nanoreporter is extended when its primary structural features are detectable in sequence when imaged according to standard techniques. Exemplary imaging techniques are described in the examples below.

[00229] In certain embodiments, an extended state of a nanoreporter can be recognized by comparing its hydrodynamic radius to its average hydrodynamic radius when free in dilute solution. For instance, in certain embodiments, a nanoreporter, or portion thereof, is extended when its hydrodynamic radius is more than about double its average hydrodynamic radius in dilute solution. More quantitatively, R represents the hydrodynamic radius of the nanoreporter, or portion thereof, and <R> represents the average hydrodynamic radius of the nanoreporter, or portion thereof, in dilute solution. The average <R> should be calculated such that R for the nanoreporter, or portion thereof, when unbound in dilute solution is less than 2<R> 95% of the time. In certain embodiments, a nanoreporter, or portion thereof, is in an extended state when R is greater than 1.5<R>, greater than 1.6<R>, greater than 1.7<R>, greater than 1.8<R>, greater than 1.9<R>, greater than 2.0<R>, greater than 2.1<R>, greater than 2.2<R>, greater than 2.3<R>, greater than 2.4<R>, greater than 2.5<R> or greater than 3.0<R>. In particular embodiments, a nanoreporter, or portion thereof, is in an extended state when R is greater than 2.0<R>.

5.13.4 ORIENTATION OF THE NANOREPORTER

[00230] In certain methods of the invention, the nanoreporter is in an oriented state. Generally, any nanoreporter is in an oriented state if it would be recognized as such by one of skill in the art.

[00231] In certain embodiments, the nanoreporter is in an oriented state when it is in the field of a force capable of orienting the nanoreporter under conditions suitable for orienting the nanoreporter. Such forces and conditions should be apparent to those of skill in the art.

[00232] In certain embodiments, the force is an electromagnetic force. For instance, when the nanoreporter is charged, such as a polynucleotide, the nanoreporter can be oriented in an electric or magnetic field. The field should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter in an electric or magnetic field are described above.

[00233] In certain embodiments, the force is a hydrodynamic force. For instance, many nanoreporters, including polysaccharides, polypeptides, and polynucleotides, can be oriented in

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 773

the field of a moving fluid. The hydrodynamic force should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter in hydrodynamic field are described above.

[00234] In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force or surface tension to orient the nanoreporter. In certain embodiments, The force should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter with gravity are described above.

[00235] In certain embodiments, the force in an optical force. For instance, the macromolecule can comprise or can be linked, covalently or noncovalently, to a particle capable of being trapped or moved by an appropriate source of optical force as described above.

[00236] In certain embodiments, the nanoreporter can be oriented by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain nanoreporters are oriented by a combination of an electric field and hydrodynamic force.

[00237] The nanoreporter is oriented when it would be recognized as oriented by one of skill in the art according to standard criteria for orientation of a nanoreporter. In certain embodiments, the nanoreporter is oriented when it is arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the nanoreporter. In certain embodiments, the nanoreporter is oriented when it is one of a plurality of nanoreporter that are arranged in parallel, as recognized by those of skill in the art. For instance, a plurality of nanoreporters can be oriented when the vector from a first terminus to a second terminus of a nanoreporter is parallel, as recognized by those of skill in the art, to the vectors between corresponding termini of other nanoreporters in the plurality.

5.13.5 SELECTIVE IMMOBILIZATION OF SECOND PORTION OF NANOREPORTER

[00238] As discussed above, in the methods of the invention, a second portion of the nanoreporter is selectively immobilized. The second portion of the nanoreporter can be any portion of the nanoreporter that is not identical to the first portion of the nanoreporter. In some embodiments, the second portion of the nanoreporter does not overlap any part of the first portion of the nanoreporter.

[00239] In certain embodiments, the present invention provides methods that comprise the single step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended or oriented state, and while a first portion of the nanoreporter is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 774

immobilized. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension or orientation of the nanoreporter are described in detail in the sections above.

[00240] In certain embodiments, the present invention provides methods that comprise the step of extending a nanoreporter, while a first portion of the nanoreporter is immobilized, and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension of the nanoreporter are described in detail in the sections above.

[00241] In certain embodiments, the present invention provides methods that comprise the step of immobilizing a first portion of a nanoreporter, the step of extending the nanoreporter while the first portion is immobilized and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension of the nanoreporter are described in detail above.

[00242] In certain embodiments, the present invention provides methods that comprise the step of orienting a nanoreporter, while a first portion of the nanoreporter is immobilized, and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an oriented state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for orienting the nanoreporter are described in detail in the sections above.

[00243] In certain embodiments, the present invention provides methods that comprise the step of immobilizing a first portion of a nanoreporter, the step of orienting the nanoreporter while the first portion is immobilized and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an oriented state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for orienting the nanoreporter are described in detail above.

[00244] The selective immobilization of the second portion of the nanoreporter can follow any technique for selective immobilization of a nanoreporter apparent to those of skill in the art. Significantly, in advantageous embodiments of the invention, the second portion of the nanoreporter is not immobilized non-selectively. Selective immobilization can allow the nanoreporter to be immobilized while in a fully extended state or nearly fully extended state. Selective immobilization can also allow the nanoreporter to be immobilized in an oriented manner. In other words, the first portion and second portion of the ~~a~~ nanoreporter can be immobilized along the direction of the field or fields used to extend the nanoreporter, with the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 775

first portion preceding the second portion in the field. When a plurality of nanoreporters are immobilized, the can be uniformly oriented along the field.

[00245] The second portion of the nanoreporter can be at any location in the nanoreporter. In certain embodiments, the second portion is at a terminus of the nanoreporter. In certain embodiments, the second portion is not at a terminus of the nanoreporter. In certain embodiments, the first portion, described in the sections above, is at one terminus of the nanoreporter, and the second portion is at another terminus of the nanoreporter.

[00246] As discussed above, the second portion of the nanoreporter is immobilized selectively. The immobilization can be by any selective interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain embodiments, the immobilization can be via electrostatic interaction with an electrode. In further embodiments, the immobilization is via electrostatic interaction with a substrate other than the electrode.

[00247] If the first portion and the second portion of the nanoreporter are selectively immobilized to the same substrate, the techniques of selective immobilization should of course be compatible with the substrate. In particular embodiments, the techniques of immobilization are the same. For instance, on a substrate coated with avidin, both the first and second portion of the nanoreporter can be immobilized selectively via biotin - avidin interactions. However, as will be apparent to those of skill in the art, the same interaction need not be used at both the first and second portions for immobilization on the same substrate. For instance, the substrate can comprise multiple moieties capable of selective binding, or the first portion can be immobilized non-selectively, or other techniques apparent to those of skill in the art.

[00248] In certain embodiments, the second portion of the nanoreporter comprises a first member of a binding pair. The second member of the binding pair can be covalently bound to the second portion of the nanoreporter, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. In useful embodiments, the substrate onto which the second portion of the nanoreporter is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound.

[00249] In certain embodiments, the second portion of the nanoreporter can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 776

carbohydrates, nucleic acids, receptors, lectins, and antibodies such as those described in the sections above.

[00250] In advantageous embodiments, the second portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. Similarly, a polypeptide nanoreporter can comprise a biotinylated amino acid residue. Useful substrates comprising avidin are described in the sections above.

[00251] In further embodiments, the second portion of the nanoreporter can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are described in the sections above.

[00252] In further embodiments, the second portion of the nanoreporter can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Exemplary useful substrates comprising reactive groups are described in the sections above.

[00253] In certain embodiments, the second portion of the nanoreporter can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the second portion of the nanoreporter could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-bromo-deoxyuridine.

[00254] In further embodiments, the second portion of the nanoreporter can be immobilized to the substrate via other binding pairs described in the sections above.

[00255] In further embodiments, the second portion of the nanoreporter is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, the second portion of the nanoreporter can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. Figure 12B llustrates a method of selectively binding a second molecule that comprises F3 that is, in turn, capable of selectively binding a moiety on a substrate. The interaction between the second portion of the nanoreporter and the molecule that comprises F3 can be mediated, for example, by an antigen-antibody interaction.

[00256] FIGS. 14A and 14B illustrate the selective immobilization of a nanoreporter according to methods of the present invention. In Figure 14A, a first portion of the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 777

nanoreporter comprises binding moiety F1 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F1 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. The nanoreporter of Figure 14 is extended by a force as described in the sections above. In Figure 14B, the force is an electrical potential. While extended, the nanoreporter is contacted with molecules comprising binding moiety F2 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F2 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. Significantly, up to three molecules comprising F2 are capable of selectively binding to a second portion of the nanoreporter to selectively immobilize it in its extended state. As illustrated, the molecules comprise a second binding moiety that selectively binds a repeated binding moiety of the nanoreporter. The binding moieties can be, for instance, complementary nucleic acid sequences, as illustrated in Figure 14B. The resulting nanoreporter is selectively immobilized in an extended state and should remain extended even when the force is removed. The selectively immobilized, extended nanoreporter can be used for any purpose apparent to those of skill in the art.

5.13.6 IMMOBILIZATION OF TWO PORTIONS OF AN EXTENDED OR ORIENTED NANOREPORTER

[00257] In certain embodiments, the present invention provides methods for selective immobilization of a first portion and a second portion of a nanoreporter that is in an extended or oriented state. Significantly, according to these methods of the invention, the nanoreporter need not be immobilized prior to application of a force capable of extending or orienting the nanoreporter.

[00258] In these methods, the nanoreporter is extended or oriented, or both, by a force capable of extending or orienting the nanoreporter. Such forces are described in detail in the sections above. In particular embodiments, the force is a force capable of extending or orienting the nanoreporter while maintaining the nanoreporter in one location, *i.e.* a force capable of extending or orienting without substantially moving the nanoreporter. Exemplary forces include oscillating electromagnetic fields and oscillating hydrodynamic fields. In a particular embodiment, the force is an oscillating electrical field. Exemplary techniques for extending or orienting a nanoreporter in an oscillating electric field are described in Asbury *et al.*, 2002, *Electrophoresis* 23(16):2658-66; Kabata *et al.*, 1993, *Science* 262(5139):1561-3; and Asbury and van den Engh, 1998, *Biophys J.* 74:1024-30, the contents of which are hereby incorporated by reference in their entirety.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 778

[00259] In the methods, the nanoreporter is immobilized at a first portion and at a second portion while extended or oriented. Both the first portion and the second portion can be immobilized non-selectively, both can be immobilized selectively, or one can be immobilized selectively and the other non-selectively. Techniques for immobilization of the first portion and second portion are described in detail in the sections above.

5.13.7 SUBSTRATE FOR IMMOBILIZATION

[00260] In the methods of the invention, the substrate for immobilization can be any substrate capable of selectively binding the nanoreporter apparent to those of skill in the art. Further, in certain aspects, the present invention provides compositions comprising a selectively immobilized nanoreporter in an extended state. The compositions comprise a substrate, as described herein, having immobilized thereto a nanoreporter in an extended state. The nanoreporter can be, of course, immobilized according to a method of the invention.

[00261] The only requirement of the substrate is that it be capable of selectively binding the second portion of the nanoreporter as described above. Thus, the substrate can be a filter or a membrane, such as a nitrocellulose or nylon, glass, a polymer such as polyacrylamide, a gel such as agarose, dextran, cellulose, polystyrene, latex, or any other material known to those of skill in the art to which capture compounds can be immobilized. The substrate can be composed of a porous material such as acrylic, styrene methyl methacrylate copolymer and ethylene/acrylic acid.

[00262] The substrate can take on any form so long as the form does not prevent selective immobilization of the second portion of the nanoreporter. For instance, the substrate can have the form of a disk, slab, strip, bead, submicron particle, coated magnetic bead, gel pad, microtiter well, slide, membrane, frit or other form known to those of skill in the art. The substrate is optionally disposed within a housing, such as a chromatography column, spin column, syringe barrel, pipette, pipette tip, 96 or 384 well plate, microchannel, capillary, etc., that aids the flow of liquid over or through the substrate.

[00263] The nanoreporter can be immobilized on a single substrate or on a plurality of substrates. For instance, in certain embodiments, the first and second portions are immobilized on the same substrate, as recognized by those of skill in the art. In certain embodiments, the first portion of the nanoreporter can be immobilized on a first substrate while the second portion of the nanoreporter can be immobilized on a second substrate, distinct from the first.

[00264] The substrate can be prepared according to any method apparent to those of skill in the art. For a review of the myriad techniques that can be used to activate exemplary substrates of the invention with a sufficient density of reactive groups, see, the *Wiley*

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 779

Encyclopedia of Packaging Technology, 2d Ed., Brody & Marsh, Ed., "Surface Treatment," pp. 867 874, John Wiley & Sons (1997), and the references cited therein. Chemical methods suitable for generating amino groups on silicon oxide substrates are described in Atkinson & Smith, "Solid Phase Synthesis of Oligodeoxyribonucleotides by the Phosphite Triester Method," In: *Oligonucleotide Synthesis: A Practical Approach*, M J Gait, Ed., 1984, IRL Press, Oxford, particularly at pp. 45 49 (and the references cited therein); chemical methods suitable for generating hydroxyl groups on silicon oxide substrates are described in Pease *et al.*, 1994, *Proc. Natl. Acad. Sci. USA 91*:5022 5026 (and the references cited therein); chemical methods for generating functional groups on polymers such as polystyrene, polyamides and grafted polystyrenes are described in Lloyd Williams *et al.*, 1997, Chemical Approaches to the Synthesis of Peptides and Proteins, Chapter 2, CRC Press, Boca Raton, FL (and the references cited therein).

[00265] Exemplary useful substrates include surfaces coated with streptavidin, *e.g.*, Accelr8 TB0200. Further useful substrates include surfaces coated with N-hydroxysuccinamide that are capable of reacting with a portion of a nanoreporter that comprises an amine. One such surface is OptArray-DNA (Accelr8). Additional useful surfaces are coated with aldehyde (*e.g.*, Nexterion Slide AL, Schott) and surfaces coated with epoxy (*e.g.*, Nexterion Slide E, Schott). Another useful surface is a biotinylated BSA coated surface useful for selective immobilization of a portion of a nanoreporter that comprises avidin or streptavidin.

5.13.8 METHODS OF USING SELECTIVELY IMMOBILIZED, EXTENDED OR ORIENTED NANOREPORTERS

[00266] The selectively immobilized, extended and/or oriented nanoreporters can be used for any purpose apparent to those of skill in the art. For instance, the selectively immobilized, extended and/or oriented nanoreporters are useful for mapping, nanoassembly and surface plasmon resonance. In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for macromoleculars with a variety of techniques, *e.g.*, atomic force microscopy or electron microscopy.

[00267] In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for macromolecular mapping. For instance, they can be used to determine the location of binding or hybridization along a macromolecule by, for example, fluorescent molecules or DNA binding proteins.

[00268] In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for nanoassembly. For instance, they can be used to facilitate

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 780

crystal growth on extended and/or oriented nanoreporters, or crystal growth on polypeptides linked or bound to extended and/or oriented macromolecules. In certain embodiments, the selectively immobilized, extended and/or oriented macromolecules can be used for the construction of nanopaths. In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for directed transport using molecular motors, such as kinesin or myosin. In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for molecular computing or for the assembly of circuits comprising macromolecules, *e.g.*, DNA computing. In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used to manipulate carbon nanotubes.

[00269] In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for the study of polynucleotide binding proteins. They can be used, for instance, to determine the presence or location of protein bound to a polynucleotide. Useful techniques include surface plasmon resonance. In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used for the study of protein fibers, such as amyloid, titin, and fibronectin.

[00270] In certain embodiments, the selectively immobilized, extended and/or oriented nanoreporters can be used to create macromolecular barcodes for the purposes of separation and sequential detection of labels. These labels spaced along the molecule provide a unique code that can be read when the nanoreporter is immobilized and extended and /or oriented. Extension and/or orientation with selective immobilization can facilitate the decoding of the macromolecular barcode.

[00271] The selectively immobilized, extended and/or oriented nanoreporters can further be used for can be used in any context where detection or imaging of a nanoreporters might be useful. They can be used for diagnostic, prognostic therapeutic and screening purposes. For instance, they can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease. They can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample. The compositions and methods of the invention can be used to quantify target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state. In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 781

5.13.9 KITS COMPRISING SELECTIVELY IMMOBILIZED, EXTENDED OR ORIENTED NANOREPORTERS

[00272] The invention further provides kits comprising one or more components of the invention. The kits can comprise, for example, a substrate according to the invention and one or more extended and/or oriented, or both, nanoreporters selectively immobilized on the substrate. The kits can be used for any purpose apparent to those of skill in the art, including, those described above.

[00273] In certain embodiments, the present invention also provides kits useful for the extension and/or orientation and selective immobilization of nanoreporters. The kits can comprise a substrate for immobilization and one or more binding partners to facilitate extension and/or orientation or immobilization of a nanoreporter. The binding partners could, in certain embodiments, comprise a moiety useful for extension and/or orientation of the nanoreporter in an appropriate force. In certain embodiments, the binding partners could facilitate immobilization or selective immobilization of the nanoreporter to the surface. In further embodiments, the kit could comprise a nanoreporter for extension and/or orientation and immobilization. In further embodiments, the kit could comprise a device capable of extending the nanoreporter.

5.14 DETECTION OF NANOREPORTERS

[00274] Nanoreporters are detected by any means available in the art that are capable of detecting the specific signals on a given nanoreporter. Where the nanoreporter is fluorescently labeled, suitable consideration of appropriate excitation sources may be investigated. Possible sources may include but are not limited to arc lamp, xenon lamp, lasers, light emitting diodes or some combination thereof. The appropriate excitation source is used in conjunction with an appropriate optical detection system, for example an inverted fluorescent microscope, an epi-fluorescent microscope or a confocal microscope. Preferably, a microscope is used that can allow for detection with enough spatial resolution to determine the sequence of the spots on the nanoreporter.

5.14.1 MICROSCOPE AND OBJECTIVE LENS SELECTION

[00275] The major consideration regarding the microscope objective lens is with the optical resolution, which is determined by its numerical aperture (NA). Generally, the larger the NA, the better the optical resolution. The required NA is preferably at least 1.07 based on the relationship of $\delta = 0.61\lambda/NA$ (δ = optical resolution and λ = wavelength). The amount of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 782

light that is collected by an objective is determined by NA^4/Mag^2 (Mag = magnification of the objective). Therefore, in order to collect as much light as possible, objectives with high NA and low magnifications should be selected.

5.14.2 CCD CAMERA SELECTION AND IMAGE CAPTURE TECHNIQUES

[00276] When selecting a CCD camera, the first consideration is the pixel size, which partially determines the final resolution of the imaging system. Optimally the optical resolution should not be compromised by the CCD camera. For example, if the optical resolution is 210-300nm, which corresponds to 12.6-18μm on a CCD chip after a 60x magnification, in order to resolve and maintain the optical resolution there should be at least two pixels to sample each spot. Or the pixel size of the CCD chip should be at most 6.3- 9 μm.

[00277] The second consideration is detection sensitivity which can be determined by many factors that include but are not limited to pixel size, quantum efficiency, readout noise and dark noise. To achieve high sensitivity, select a qualitative camera with big pixel size (which can give big collection area), high quantum efficiency and low noise. An exemplary camera with these criteria is the Orca-Ag camera from Hamamatsu Inc.

5.15 COMPUTER SYSTEMS

[00278] The invention provides computer systems that may be used to automate nanoreporter image collection, perform nanoreporter identification and/or decode nanoreporter codes. Specifically, the invention provides various computer systems which run one or more programs described below (*e.g.*, data storage module 44, label identification module 50, probe identification module 54). The computer systems can control a camera that takes light images of nanoreporters bound on substrates. These light images are then used by the computers of the present invention to identify and decode nanoreporters.

[00279] Figure 9 details an exemplary system that supports the functionality described herein. The system is preferably a computer system 10 having:

- a central processing unit 22;
- a main non-volatile storage unit 14, for example, a hard disk drive, for storing software and data, the storage unit 14 controlled by controller 12;
- a system memory 36, preferably high speed random-access memory (RAM), for storing system control programs, data, and application programs, comprising programs and data loaded from non-volatile storage unit 14; system memory 36 may also include read-only memory (ROM);

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 783

- a user interface 32, comprising one or more input devices (*e.g.*, keyboard 28) and a display 26 or other output device;
- a network interface card 20 or other communication circuitry for connecting to detector 72 and, optionally, any wired or wireless communication network 34 (*e.g.*, the Internet or any other wide area network);
- an internal bus 30 for interconnecting the aforementioned elements of the system; and
- a power source 24 to power the aforementioned elements.

[00280] Operation of computer system 10 is controlled primarily by operating system 40, which is executed by central processing unit 22. Operating system 40 can be stored in system memory 36. In addition to operating system 40, in a typical implementation, system memory 36 can include one or more of the following:

- file system 42 for controlling access to the various files and data structures used by the present invention;
- a data storage module 44 comprising instructions for storing a plurality of light images 46;
- a label identification module 50 for identifying a plurality of labels 52 in the plurality of light images that are proximate to each other on a substrate, where a spatial order of the plurality of labels determines a string sequence of the plurality of labels;
- a probe identification module 54 for determining for determining whether the string sequence of a plurality of labels comprises a valid reporter sequence; and
- a lookup table 56 that includes a plurality of valid reporter sequences 58.

[00281] As illustrated in Fig. 9, computer system 10 comprises software program modules and data structures. The data structures stored in computer system 10 include, for example, images 46, labels 52, and lookup table 56. Each of these data structures can comprise any form of data storage including, but not limited to, a flat ASCII or binary file, an Excel spreadsheet, a relational database (SQL), an on-line analytical processing (OLAP) database (MDX and/or variants thereof), or a comma separated value file. In some embodiments, the data structures and software modules depicted in Figure 9 are not housed on computer system 10, but rather are housed on a computer or other type of storage device that is in electrical communication with computer system 10 across network 34.

[00282] One aspect of the present invention provides a computer program product comprising a computer readable storage medium (*e.g.*, memory 36, storage unit 14, and/or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 784

other computer readable storage media) and a computer program mechanism embedded therein. The computer program mechanism is for detecting the presence of a probe within a sample overlayed on a substrate. As described above, the probe comprises a plurality of spatially arranged label monomers in one or more label attachment regions of a given nanoreporter. The sequence of signals emitted by the label monomers associated with the various label attachment regions of the scaffold of a given nanoreporter allows for the unique identification of the nanoreporter. The computer program mechanism comprises data storage module 44, label identification module 50, and probe identification module 54.

[00283] *Data storage module 44.* Data storage module 44 comprises instructions for storing a plurality of light images 46. Each light image in the plurality of light images is from light emitted from a sample overlayed on a substrate. Representative biological samples are described in Section 5.10, above. Representative substrates are described in Section 5.13.7. In typical embodiments, each light image in the plurality of light images is a scan, taken by detector 72, of a portion of a substrate. Exemplary detectors 72 are described in Section 5.14.2. A nonlimiting exemplary detector 72 is the Orca-Ag camera from Hamamatsu Inc. Each light image is a picture of the substrate onto which the sample containing a nanoreporter is overlayed that is taken by detector 72. Each light image records light from the sample at a corresponding wavelength range in a plurality of different wavelength ranges. In other words, each light image measures light in a select wavelength range. In some embodiments, the wavelength range of a light image is determined by the specifications of a light filter that filters light received by detector 72. The filter effectively blocks out light that is not within a specified wavelength range. In practice, a plurality of light images is taken, and each light image in the plurality of light images is taken using a filter selected from a plurality of different filters. In some embodiments, 2 or more light images, 3 or more light images, 4 or more light images, 5 or more light images, 6 or more light images, 7 or more light images, 8 or more light images, 9 or more light images, or 10 or more light images, between 2 and 20 light images, between 2 and 100 light images, or less than 30 light images are taken. Filters for limiting light to predetermined wavelength ranges are commercially available. Nonlimiting examples of commercially available filters are HQ480/40, Q505LP, HQ535/50, HQ545/30x, 570LP, HQ587/30m, HQ630/20, Q649LP, and HQ655LP by Chroma Inc. Exemplary wavelength ranges in the plurality of different wavelength ranges could be, for example, the wavelength range between 485nm and 585nm, the wavelength range between 557nm and 617nm, the wavelength range between 637nm and 697nm, the wavelength range between 510nm and 650nm, the wavelength range between 586nm and 658nm, the wavelength range between 515nm and 575nm, and the wavelength range between 440nm and 520nm. Many

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 785

other wavelength ranges can be used in addition to or instead of some or all of the aforementioned wavelength ranges.

[00284] In some embodiments, the plurality of different wavelength ranges consists of between two and six different wavelength ranges, between two and twenty different wavelength ranges, more than ten different wavelength ranges, more than twenty different wavelength ranges, more than thirty different wavelength ranges, more than forty different wavelength ranges, more than 100 different wavelength ranges, or less than fifty different wavelength ranges. In some embodiments each wavelength range represents a different color. In some embodiments there is some degree of overlap between different wavelength ranges. In some embodiments there is no overlap between different wavelength ranges.

[00285] An advantage of the present invention is that there is no requirement that the nanoreporters of the present invention be overlayed at a predetermined position on the substrate as is the case in the microarray arts. While specialized substrates can be prepared such that the nanoreporters of the present invention will adhere to predetermined positions on a substrate, in preferred embodiments, the location where a nanostring adheres to a substrate is totally random. Furthermore, in preferred embodiments, all that needs to be determined is the identity of a nanoreporter on the substrate. Thus, signals are measured from the substrate for the purpose of nanoreporter identification. As such, signals are processed in a binary fashion (*e.g.*, absence or presence of at a given wavelength) in preferred embodiments using the software modules described below. In some embodiments, not only is the position of nanoreporters on the substrate random, so is the orientation. Even in instances where the orientation of the nanoreporter on the substrate is random, the software modules described below can identify the nanoreporters. Another advantage of the present invention is that there is no requirement that multiple regions of the substrate be stitched together into a composite light image. Multiple light images of the substrate are taken, but each of these light images is from the same region of the substrate, just at different wavelengths.

[00286] In some embodiments, a label will fluoresce at multiple wavelengths (e.g., red and blue). In such embodiments the label identification module described below detects such a label by measuring a spot in the same position on the substrate in multiple light images. Intensity criteria can be used in such instances. For example, a requirement could be imposed that (i) a minimum blue intensity be observed in a light image that stores blue light emitted by the sample and (ii) that a minimum red intensity be observed in the same spot in the light image that stores red light emitted by the sample.

[00287] *Label identification module 50.* Label identification module 50 comprises instructions for identifying a plurality of labels (label monomers), in the plurality of light

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 786

images 46, that are proximate to each other on the substrate. In typical embodiments, each label in a nanoreporter will emit light that causes a corresponding bright spot to appear in one or more light images 46. Each light image 46 may have dozens, hundreds, or even thousands of spots, where each spot potentially represents a label in a nanoreporter. Label identification module 50 deems these spots in the light images to be labels when specific criteria, described in detail below, are satisfied.

[00288] Label identification module 50 is capable of identifying labels that are proximate to each other on the substrate even when these labels appear in different light images. For example, consider the case in which label 1 (light monomer 1) and label 2 (light monomer 2) are in the same nanoreporter and are spatially proximate on the substrate. Further consider that label 1 and label 2 emit light within different wavelength ranges. Thus, in this example, light emitted by label 1 and by label 2 are recorded in different light images, say light image A and light image B, respectively. In some embodiments, the light emission of label 1 is recorded as a bright intensity spot in light image A whereas the light emission of label 2 is recorded as a bright intensity spot in light image B. In this example, label identification module 50 superimposes light image A and light image B. Light image A and light image B both cover the same region of the substrate. This region of the substrate may encompass several nanoreporters. In fact, this the region of the substrate covered by light image A and light image B may encompass hundreds or even thousands of nanoreporters. Therefore, light image A and light image B can each potentially record dozens, hundreds, or even thousands of labels, where each such label appears, for example, as a bright spot within light image A, light image B, or some other light image that is taken by detector 72. Light image A and light image B only differ by the wavelength range recorded. Because light image A and light image B are spatially superimposed on each other, label identification module 50 can identify the spatial proximity of label 1 and label 2 even though label 1 and label 2 originate in different light images, based on the spatial proximity of label 1 and label 2 in the superimposed light images.

[00289] Each wavelength range represented by a light image 46 is intended to measure a particular color emitted by one or more label monomers in a nanostring. In some embodiments, a wavelength range encompasses a contiguous set of wavelengths having a spectral width of up to 10nm, up to 20nm, up to 30nm, up to 40nm, up to 50nm, up to 60nm, up to 70nm, up to 80nm, up to 90nm, or up to 100nm.

[00290] In some embodiments, label identification module 50 aligns the light images using fiducials. A nonlimiting example of a fiducial is a latex bead that is impregnated with many different types of dye. One example of a fiducial is the MultiSpeck™ Multispectral

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 787

Fluorescence Microscopy Kit (Molecular Probes, Inc., Eugene, Oregon). The MultiSpeck™ Multispectral Fluorescence Microscopy Standards Kit is marketed as an external reference for comparing images collected with different optics, on different instruments and in different laboratories, as well as for monitoring routine day-to-day variations in instrumental performance. The kit includes two suspensions of submicron-diameter microspheres. The first suspension, the MultiSpeck suspension, is comprised of multispectral fluorescent microspheres that exhibit three relatively distinct excitation and emission bands, red, green and blue, all in the same particle. When excited with ultraviolet light, each sphere emits blue fluorescence, whereas when fluorescein or rhodamine/Texas Red® excitation filters are used, the spheres fluoresce at wavelengths similar to fluorescein or rhodamine/Texas Red emissions, respectively. Furthermore, because a single multispectral microsphere will appear different colors depending on the filters used for observation, these microspheres can be used for image registration across multiple wavelengths, thus allowing for the accurate determination of the spatial relationships of multiple labels in a multiparameter experiment. The second suspension, the RGB Mix suspension, is comprised of a mixture of "single-band" microspheres that exhibit the same three excitation/emission bands, red, green and blue, (RGB) as the multispectral microspheres, but in separate particles. Thus, a fiducial will emit light across several different wavelengths. Fiducials are randomly arranged on the substrate along with the sample. Because the fiducials emit light across a wide range of spectral frequencies, they are present in several if not all of the plurality of light images. Thus, label identification module 50 can align most if not all of the light images in a plurality of images to each other using the fiducial spots present in the light images provided that the light images are of the same region of the substrate (*e.g.*, the same field of view).

[00291] A spatial order of the plurality of labels determines a string sequence of the plurality of labels. For example, consider the case in which the following string of label monomers are detected on the substrate: red - red - green - blue. The spatial order of these label monomers on the substrate dictates the string sequence of a possible probe. Thus, if serially arranged label monomers are detected on the substrate with the serial arrangement: red - red - green - blue, the serially arranged label monomers form the string sequence: RRGB, where R stands for red, G stands for green, and B stands for blue. The string sequence RRGB is potentially a string sequence of a valid probe.

[00292] The present invention encompasses a broad range of different labels that may be used in a probe. In some embodiments, each label in a plurality of spatially arranged labels can emit light in one of four different wavelength ranges: a red wavelength range, a green wavelength range, a blue wavelength range, or blank (no emission). In some embodiments,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 788

each label in a plurality of spatially arranged labels can emit light in one of five different wavelength ranges, six different wavelength ranges, seven different wavelength ranges, eight different wavelength ranges, nine different wavelength ranges, ten different wavelength ranges, between five and fifteen different wavelength ranges, between four and twenty different wavelength ranges, between three and forty different wavelength ranges, more than 30 different wavelength ranges, or less than 100 different wavelength ranges. Furthermore, the present invention imposes no limitation on the number of labels (label monomers) that are present in a given probe. In some embodiments, there are between two and 100 labels in a probe, between two and 1000 labels in a probe, between two and 20 labels in a probe, between three and 40 labels in a probe, more than five labels in a probe, more than six labels in a probe, more than seven labels in a probe, more than ten labels in a probe, or less than fifty labels in a probe.

[00293] Label identification module 50 determines the spatial order of the labels on a substrate. To accomplish this, a plurality of labels in the images is identified. In typical embodiments, there is more than one probe on the portion of the substrate that is imaged, and each such probe has a plurality of spatially arranged labels that generate corresponding spots in the light images. In some embodiments, there are between two and 10,000 probes in the portion of the substrate being imaged. Thus, label identification module 50 has to determine which spots in images are labels as well as identify sets of labels that each belong to a single probe. This task is made more complex in preferred embodiments where probes randomly bind to the substrate.

[00294] In some embodiments, label identification module 50 first validates labels in light images. Then, label identification module 50 proceeds to determine which of the labels in the plurality of labels belong to the same probe using one or more rules. In some embodiments label identification module 50 first identifies a plurality of candidate labels giving rise to spots in the plurality of light images. Each candidate label in the plurality of candidate labels comprises a position on the substrate that emits more than a threshold amount of light in one or more light images in the plurality of light images. In some embodiments, the plurality of labels comprises a first candidate label that emits light in a first wavelength in the plurality of different wavelengths and a second candidate label in the plurality of labels that emits light in a second wavelength in the plurality of different wavelengths. In some embodiments, a candidate label is deemed to be a valid label when it emits more than a threshold amount of light in any one light image in the plurality of light images. In some embodiments label identification module 50 verifies that a candidate label is a valid label when the label satisfies a spot shape criterion (*e.g.*, a match between an observed spot shape of the candidate label and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 789

the theoretical point spread of the diffraction limited point source light determined by a magnification of the candidate label). In some embodiments, label identification module 50 verifies that a candidate label is a valid label using point spread function modeling. In some embodiments, label identification module 50 validates a candidate label using a spot segmentation algorithm (*e.g.*, a watershed transformation). Watershed transformations are described in Vincent and Soille, 1991, *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*, pp. 583-598, which is hereby incorporated by reference herein, in its entirety.

[00295] In some embodiments, label identification module 50 verifies that a candidate label is a valid label when the label satisfies a spot shape criterion (*e.g.*, a match between an observed spot shape of the label and the theoretical point spread of the diffraction limited point source light determined by a magnification of the label in an image in which the label was observed). In some embodiments, the spot shape criterion is evaluated using point spread function modeling. In some embodiments, the spot shape criterion is evaluated using a spot segmentation algorithm (*e.g.*, a watershed transformation). An exemplary application of a watershed transformation is described in Park *et al.*, 2004, "Automatic Microarray Image Segmentation Based on Watershed Transformation," Proceedings of the 17th International Conference on Pattern Recognition, volume 3, pages 786-789, which is hereby incorporated by reference herein in its entirety.

[00296] In some embodiments, an image segmentation step is performed by label identification module 50 on a candidate label in a light image in the plurality of light images to determine which pixels form the spot in the light image generated by the label, which pixels form the background, and which pixels are just noise or artifacts and should be eliminated. In some embodiments, a pure spatial-based signal segmentation approach is used by label identification module 50, in which a circle is placed over the spot generated by the label. In this approach, all pixels within this circle are counted as part of the spot and all pixels outside the circle are used to calculate background. The pixels between the two circles correspond to the transition area between the spot and its background and are discarded in order to improve the quality of the data. In this approach, all the pixels outside the circle within the boundary of a predetermined square are considered as background.

[00297] In some embodiments, intensity based segmentation is used by label identification module 50 to validate candidate labels. Methods in this category use exclusively intensity information to segment out signal pixels from background. Such approaches assume that the signal pixels are brighter on average than the background pixels. As an example, suppose that the target region around the spot taken from the light image consists of 40 x 40 pixels. The spot is about twenty pixels in diameter. Thus, from the total of 1600 (40 x 40)

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 790

pixels in the region, about 314 (π x 10^2, since the area of a circle is πr^2, where r stand for radius and is half the diameter) pixels, or 20%, are signal pixels and they are expected to have their intensity values higher than that of the background pixels. To identify these signal pixels, all the pixels from the target region are ordered in a one dimensional array from the lowest intensity pixel to the highest one $\{p_1, p_2, p_3, \ldots, p_{2500}\}$, in which p_i is the intensity value of the pixel of the i^{th} lowest intensity among all the pixels. If there is no contamination in the target region, the top 20% pixels in the intensity rank may be classified as the signal pixels. In some embodiments, a spot is only considered valid if the intensity measured by the intensity based segmentation approach exceeds a threshold value. However, in the present invention, spot size is typically substantially less than twenty pixels in diameter but the described method is still applicable.

[00298] In some embodiments, Mann-Whitney segmentation is used by label identification module 50 in order to validate candidate labels. Mann-Whitney segmentation is described in Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC, New York, p. 47, which is hereby incorporated by reference herein in its entirety. The Mann-Whitney segmentation approach combines the use of spatial information with some intensity based analysis. Based on the result of the spot finding operation performed by label identification module 50, a circle is placed in the target region to include the region in which the spot is expected to be found. Since the pixels outside of the circle are assumed to be the background, the statistical properties of these background pixels can be used to determine which pixels inside the circles are signal pixels. A Mann-Whitney test is used to obtain a threshold intensity level that will separate the signal pixels from other pixels (*e.g.*, background) even if they are inside the expected area of the spot. The Mann-Whitney test is described in Smith, 1991, *Statistical Reasoning*, Third Edition, Allyn and Bacon, Boston, pp. 724-730, which is hereby incorporated by reference herein in its entirety. Pixels inside of the circle having a higher intensity than the threshold intensity are treated as signal. In some embodiments, a spot is only considered valid if the intensity measured by the Mann-Whitney segmentation approach exceeds a threshold value.

[00299] In some embodiments, a combined intensity-spatial segmentation (trimmed measurements approach) is used by label identification module 50 in order to validate candidate labels. This approach combines both spatial and intensity information in segmenting the signal pixels from the background in a manner similar to the Mann-Whitney approach. In this approach, the assumption is that once the spot for a label is localized by label identification module 50 and a target circle placed in the target region, most of the pixels inside of the circle will be signal pixels and most of the pixels outside of the circle will be background. However,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 791

due to spot shape irregularity, some signal pixels may leak out of the circle and some background pixels may get into the circle. Background pixels within the circle may be considered as outliers in the intensity distribution of the signal pixels. Similarly, signal pixels that fall outside the circle will also appear as outliers with respect to the intensity distribution of the background pixels. Contamination pixels anywhere will appear as outliers in the intensity domain for both signal and background. These outliers would severely change the measurement of the mean and total signal intensity if they are not eliminated. To remove the effect of outliers on these measurements, some embodiments of label identification module 50 can simply "trim-off" a fixed percentage of pixels from the intensity distribution of the pixels for both signal and background regions. The Mann-Whitney approach performs a statistical analysis on the pixels outside the presumed spot area and then uses the threshold calculated there to segment the pixels inside the target area. The trimmed measurements approach performs a statistical analysis of both distributions (outside as well as inside the presumed spot) and eliminates the outliers from each such distribution independently without making the leap of faith that the characteristics of the distribution outside will also reflect the properties of the distribution inside. Eliminating approximately 5-10% of each distribution allows this approach to cope with artifacts such as dust particles and other impurities. In preferred embodiments, label identification module 50 does not have to quantify the intensity of the spots corresponding to labels in the images. However, in some embodiments, label intensity is quantified using any quantification technique known in the art including, but not limited to, total signal intensity, mean signal intensity, and median signal intensity. Exemplary quantification techniques are described in Draghici, 2003, *Data Analysis Tools for DNA Microarrays*, Chapman & Hall/CRC, New York, Section 3.4.3, which is hereby incorporated by reference herein in its entirety.

[00300] The above-identified techniques identifies a plurality of candidate labels and validates all or some of these candidate labels. These are referred to as spot level rules. In some embodiments, to be valid, the spot for a label needs to match the theoretical point spread function of the diffraction limited point source light. Spots can be identified, for example, by a watershed transformation. In some embodiments, spots that exceed the diffraction limit are recorded. Next, because there can be more than one probe in the light images, label identification module 50 applies *a priori* knowledge about probe architecture to determine which validated labels belong to the same probe. In essence, this *a priori* knowledge is used to form a model of what a probe should look like in the light images. This model can then be reduced to a set of reporter level rules that can be used to verify that a given set of labels belongs to a given probe. Examples of such rules include applying a distance criterion

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 792

between the centroid of one spot to another and requiring that spots considered to belong to the same probe do not form angles greater than a threshold amount. In essence, the expected model is fitted to observed validated labels. In some embodiments, a set of labels that gives the least amount of error when fitted to an expected model for a probe is deemed to belong to the same probe.

[00301] In some embodiments *a priori* knowledge used to formulate a model is the expectation that the labels belonging to the same probe will be linearly arranged. In some embodiments, a set of labels is subjected to linear regression in order to determine if the set of labels is linearly arranged. If the linear regression is satisfied, other rules may be applied to further validate that the set of labels belongs to the same probe. In some embodiments, a set of labels is considered linear if they fit a linear regression model criterion. In some embodiments, the set of labels is considered to fit a linear regression model criterion when an R-value for the regression model is 0.9 or greater. In some embodiments, as described herein, experimental conditions are imposed (*e.g.*, an electric field) to guarantee that the probes are linearly arranged. In some embodiments, probes are not linearly arranged and more complex models that incorporate expected probe curvature are invoked. In some embodiments, the linear direction of probes on the substrate is not known (*e.g.*, is not predetermined). In some embodiments, the linear direction of probes on the substrates is known (*e.g.*, is predetermined). In each such embodiment, this *a priori* knowledge is used to form the model. In addition to rules concerning the shape of the probe (*e.g.*, linear, curved, linear in a predetermined orientation), there are rules for label spacing. A spacing rule imposes constraints on the distance between labels in the set of labels. Additional types of rules concern the shape of spots that constitute the reporter. Spot shape has been discussed above in conjunction with algorithms that are designed to localize spots and measure spot intensity (*e.g.*, a match between an observed spot shape and the theoretical point spread of the diffraction limited point source light determined by a magnification of the label, point spread function modeling, a spot segmentation algorithm such as a watershed transformation).

[00302] In some embodiments the model used to identify labels belonging to the same probe requires that the spacing between each label in a probe be equal or within a defined range. In one example, the model comprises the rules that (i) the labels for the probe are linearly arranged and (ii) the labels are spaced between 450 nm and 550 nm apart from each other. In some embodiments the model used to identify labels belonging to the same probe imposes no requirement that the labels be equally spaced. In fact, in some embodiments, information about a probe identity is encoded in the form of unequal spacing between probe spots. Further, in some embodiments, the string sequence encoded by a probe includes spacer

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 793

label monomers that do not emit light in any of the wavelengths measured in the plurality of light images. The present invention imposes no constraints on the spacing between label monomers. However, once the nanoreporters are synthesized, the *a priori* knowledge about the nanoreporters can be used to construct the model. Accordingly, in some embodiments, a model imposes the constraint that labels belonging to the same probe are spaced between 100 nm and 150 nm of each other, between 150 nm and 200 nm of each other, between 200 nm and 250 nm of each other, between 250 nm and 300 nm of each other, between 350 nm and 400 nm of each other, between 400 nm and 450 nm of each other, between 450 nm and 500 nm of each other, between 500 nm and 550 nm of each other, or between 550 nm and 600 nm of each other. In some embodiments, the spacing between label monomers is unequal but is known. For example, in one four-label nanostring, the spacing between the first label and the second label is 400 nm, the spacing between the second label and the third label is 750 nm, and the spacing between the third label and the fourth label is 625 nm. This spacing information would be taken into account in constructing a model for this class of nanostring.

[00303] In models in which labels are equally spaced, the model is fit to observed labels and the model is satisfied if the observed labels fall within predetermined tolerances. In some embodiments, the models used to identify labels belonging to the same probe include angle rules. For example, consider the case of a four label nanostring. In some embodiments, a model for this nanostring will impose a largest tolerable angle constraint, such that no three labels in the nanostring form an angle larger than the largest tolerable angle constraint.

[00304] In some embodiments, a label identification module applies a first distance criterion between a centroid of a first label and a centroid of a second label in a plurality of labels. In some embodiments, the first distance criterion is determined by a calculated distance between a first label and a second label in the probe. In some embodiments, the label identification module applies a second distance criterion between a centroid of the second label and a centroid of a third label in the plurality of labels. The second distance criterion can be determined by, for example, a calculated distance between a second label and a third label in the probe. In some embodiments, the first distance criterion is the same as the second distance criterion. In some embodiments, the first distance criterion is different from the second distance criterion. In some embodiments, a value of the first distance criterion and a value of the second distance criterion contribute to determining whether the plurality of labels is the probe.

[00305] In some embodiments, the instructions for identifying the plurality of labels comprises instructions for identifying a buffer zone around a portion of the substrate that has select labels. A buffer zone is a region of the substrate that contains no labels. A buffer zone

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 794

surrounds a portion of the substrate that has a set of labels. Identification of a buffer zone around a set of labels ensures that the set of labels surrounded by the buffer zone is in fact a set of labels that potentially corresponds to a probe. If a buffer zone around a given set of labels cannot be identified, the given set of labels could potentially be from two or more probes that are proximate to each other on the substrate. This is not a desirable result because it does not lead to proper probe identification. Therefore, the use of a buffer zone criterion can help to ensure that a given set of labels belongs to a single probe. In fact, in some embodiments, a set of labels identified on the substrate are not considered to be validated labels unless there is a buffer zone around the set of labels. Thus, in some embodiments, a set of labels may still be considered "candidate" labels unless and until the buffer zone criterion is satisfied. In some embodiments, the buffer zone is elongated in shape so that it will fit around a linear array of spots that represent a possible probe.

[00306] *Probe identification module 54.* Once label identification module 50 identifies a candidate probe on the substrate, probe identification module 54 determines whether the string sequence defined by the labels of the identified candidate probe constitutes a valid reporter sequence. When the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe. When the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe. In such embodiments, probe identification module 54 compares the string sequence of the plurality of labels to valid reporter sequences in a lookup table. In preferred embodiments, the lookup table includes a list of all possible valid string sequences. The premise behind the lookup table is that only some of the possible set of string sequences are actually used. For example, consider the case where nanostrings (probes) are constructed with four label monomers and each label monomer can adopt one of four different colors (*e.g.*, red, green, blue, and blank). In this case, there are $4^4 = 64$ possible string sequences. Say that only twenty of these string sequences are actually used in the sample exposed to the substrate. In such embodiments, the lookup table would be populated with these twenty string sequences. If label identification module 50 identifies a probe on the substrate that has a string sequence that matches one of these twenty string sequences, the probe would be validated. If label identification module 50 identifies a probe on the substrate that has a string sequence that does not match one of these twenty string sequences, the probe would be discarded as an artifact. In many embodiments, the number of string sequences in the lookup table is a small fraction of the possible number of string sequences. This condition helps to ensure that only valid probes are identified on the substrate. In some embodiments, the probe has seven labels, each label adopting one of four different colors for a total of 7^4 different string sequence, a small fraction

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 795

of which are used to construct actual nanoreporters. In some embodiments, a probe has twenty labels, each label adopting one of twenty different colors for a total of 20^{20} different valid reporter sequences, a small fraction of which are used to construct actual nanoreporters. These exemplary embodiments merely serve to illustrate the dimensions of some embodiments of the present invention. As indicated above, a larger range of labels can be used in a given nanoreporter and each such label can adopt many different colors. In some embodiments, there are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more labels in a probe and each label can adopt any one of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more different colors. In some embodiments, there are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more labels in a probe and each label can adopt any two of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more different colors. In some embodiments, there are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more labels in a probe and each label can adopt any three of 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more different colors. In some embodiments, there are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more labels in a probe and each label can adopt multiple colors (*e.g.*, 2 colors, 3 colors, 4 colors) in a plurality of colors. In some embodiments, this plurality of colors consists of five or more colors, six or more colors, seven or more colors, between seven and one hundred colors, or more than twenty colors. In some embodiments, a color is a discrete wavelength range (*e.g.*, 340 nm to 350 nm, 355 nm to 365 nm, *etc.*).

[00307] In some embodiments, a first subset of the labels in the string sequence defined by a probe serves to error check the identities of the labels in a second subset of the labels in the string sequence. For example, consider the case in which there are eight labels in a probe. The last two labels may serve to error check the identity of the first six labels. For instance, the last two labels in the string sequence may serve as a checksum for the first six labels in the string sequence. In some embodiments, the checksum values can actually be used to error correct a string sequence that has been misread by label identification module 50. Thus, in some embodiments, it is possible to validate a string sequence that is not present in the look-up table by applying error correction techniques to the string sequence using the checksum or other form of error correcting code present in the string sequence.

[00308] In some embodiments, where distance information encodes string identity in part, the lookup table comprises valid reporter sequences, where each such valid reporter sequence comprises distance information between labels in addition to a string sequence. In such embodiments, there must be a match between both the distance information and the string sequence in order to identify a valid probe. For example, consider the case in which the string

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 796

sequence identified by label identification module 50 consist of three labels, where the distance between the first and second labels is d_1 and the distance between the second and third labels is d_2. In this example, probe identification module must find a string sequence in the lookup table 56 that encodes the same string sequence and has matching d_1 and d_2 distances between the first and second labels and the second and third labels, respectively.

[00309] In some embodiments, label identification module 50 comprises instructions for repeating the instructions for identifying a plurality of labels a plurality of times. Each time the instructions for identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images 46, that are proximate to each other on the substrate. In such embodiments, probe identification module 54 determines whether each such different plurality of labels identified by label identification module 50 comprises a valid reporter sequence. For each such different plurality of labels, probe identification module 54 deems the different plurality of labels to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence. Further, probe identification module 54 deems the different plurality of labels to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence. In some embodiments, a plurality of probes is identified. In some embodiments, the plurality of probes consists of three or more probes. two or more probes, three or more probes, ten or more probes, at least 5, 10, 15, 20, 50, 75, 100, 150, 200, 300, or 400 probes or more. In some embodiments, probe identification module 54 stores each type of probe identified. A probe "type" is identified by the string sequence of the probe. In some embodiments, probe identification model 54 stores each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence. In some embodiments, probe identification model 54 stores each string sequence of each different plurality of labels that is confirmed as a valid reporter sequence.

5.16 APPLICATIONS OF NANOREPORTER TECHNOLOGY

[00310] The compositions and methods of the invention can be used for diagnostic, prognostic therapeutic and screening purposes. The present invention provides the advantage that many different target molecules can be analyzed at one time from a single biomolecular sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 797

5.16.1 DIAGNOSTIC/PROGNOSTIC METHODS

[00311] The present methods can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease. For example, a blood sample can be assayed according to any of the methods described herein to determine the presence and/or quantity of markers of a cancerous cell type in the sample, thereby diagnosing or staging the cancer. Alternatively, the methods described herein can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample. Thus, the target molecules detected using the compositions and methods of the invention can be either patient markers (such as a cancer marker) or markers of infection with a foreign agent, such as bacterial or viral markers. Because of the quantitative nature of nanoreporters, the compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state.

[00312] In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient. For example, the amount of a particular marker for a tumor can be accurately quantified from even a small sample from a patient. For certain diseases like breast cancer, overexpression of certain genes, such as Her2-neu, indicate a more aggressive course of treatment will be needed.

5.16.2 SCREENING METHODS

[00313] The methods of the present invention can be used, *inter alia*, for determining the effect of a perturbation, including chemical compounds, mutations, temperature changes, growth hormones, growth factors, disease, or a change in culture conditions, on various target molecules, thereby identifying target molecules whose presence, absence or levels are indicative of a particular biological states. In a preferred embodiment, the present invention is used to elucidate and discover components and pathways of disease states. For example, the comparison of quantities of target molecules present in a disease tissue with "normal" tissue allows the elucidation of important target molecules involved in the disease, thereby identifying targets for the discovery/screening of new drug candidates that can be used to treat disease.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 798

5.17 METHODS FOR IDENTIFYING PROBES

[00314] One aspect of the invention provides methods for detecting the presence of a probe within a sample overlayed on a substrate. In this aspect of the invention, the probe comprises a plurality of spatially arranged labels. In one such method, a plurality of labels, in a plurality of light images, that are proximate to each other on the substrate is identified. The spatial order of the plurality of labels determines a string sequence of the plurality of labels. Each light image in the plurality of light images is for light received from the sample at a wavelength range in a plurality of different wavelength ranges. In the method, a determination is made as to whether the string sequence of the plurality of labels comprises a valid reporter sequence. When the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe. When the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

[00315] In some embodiments, the determining step comprises comparing the string sequence of the plurality of labels to valid reporter sequences in a lookup table. In some embodiments, the method further comprises storing the string sequence of the plurality of labels that is not confirmed as a valid reporter sequence. For instance, the string sequence can be stored in an electronic memory. In some embodiments, the method further comprises aligning a first light image to a second light image in the plurality of light images using a plurality of fiducials present on the substrate.

[00316] In some embodiments, the step of identifying a plurality of labels is repeated a plurality of times. Each time the step of identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images, that are proximate to each other on the substrate. In some embodiments, the method further comprises determining whether each of the different plurality of labels comprises a valid reporter sequence. Each different plurality of labels is deemed to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence. Furthermore, each different plurality of labels is deemed to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence. In some instances according to this embodiment of the present invention, a plurality of probes is identified. For instance, in some embodiments, two or more probes, three or more probes, ten or more probes, at least 5, 10, 15, 20, 50, 75, 100, 150, 200, 300, or 400 probes or more are identified.

[00317] In some embodiments where a plurality of probes are identified, each type of probe identified is recorded. A probe "type" is identified by the string sequence of the probe. Each unique valid string sequence represents a different probe type. In some embodiments,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 799

each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence is stored. In this way, it is possible to determine common errors that arising in the probes on the substrate. One type of error that can be identified by tracking pluralities of labels that do not form valid string sequences is the condition where there are too many probes on the substrate. When there are too many probes on the substrate, the labels of neighboring probes become proximate to each other, making it difficult to determine which probe each label belongs to. Another type of error that can be identified is the condition in which an excessive number of probes get ripped apart, leaving truncated probes on the substrate. In some embodiments, all species of labels, strings, invalid reporter sequences, valid reporter sequences, probe types is tracked in the methods of the present invention.

6. EXAMPLES

[00318] The following examples are offered to illustrate this invention and are not to be construed in any way as limiting the scope of this invention.

6.1 EXAMPLE 1: SELECTIVE IMMOBILIZATION OF EXTENDED DNA

[00319] A double stranded RNA-DNA hybrid 7.2 Kb in length is functionalized at one terminus with biotin. At the other terminus, the DNA comprises a single stranded sequence of 15 bases repeated 4 times (5'-GTC TAT CAT CAC AGC GTC TAT CAT CAC AGC GTC TAT CAT CAC AGC GTC TAT CAT CAC AGC-3'; SEQ ID NO:2). Thus, the DNA comprises four binding sites at one terminus for selective immobilization. The hybrid also has four regions with Cy3 fluorophores incorporated into the RNA.

[00320] A small sample of the DNA (3 μL, 0.01 fmol/μL in 1X TAE, or 40 mM Tris acetate, 1 mM EDTA, pH 8.0) is transferred into a microfluidic device comprising a channel molded into polydimethylsiloxane that is passively adhered to a streptavadin coated coverslip (Accelr8, TB0200). The channel dimensions are 50 μm x 1 mm x 10 mm. *See* Figure 15A. The sample is contacted with the coverslip at room temperature for 15 minutes allowing the DNA to selectively bind the streptavadin surface via the biotin at the terminus of the DNA. Unbound DNA is washed away by fluid flow. The 1X TAE buffer in the wells are exchanged for fresh buffer and fluid levels are evened out at 30 μL each well. *See* Figure 4A.

[00321] An electric field of 200 V/cm is applied to extend the long negatively charged DNA (*see* Figure 15B) toward the positive electrode.

[00322] An immobilization agent, a biotinylated oligonucleotide (5'-Biotin GCTGTGATGATAGAC-3', SEQ ID NO:3, 50 μL @ 100 nM, 1X TAE) complementary to the second terminus of the DNA, is added to the negative well. The additional volume raises

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 800

the fluid level in the well and induces hydrostatic flow to introduce the immobilization reagent into the channel (*see* Figure 15C). The flow also acts to further stretch the DNA in addition to the electric field.

[00323] The biotinylated oligonucleotide hybridizes with the second terminus of the DNA while it is extended and selectively binds the streptavidin of the coverslip. The sample can be effectively immobilized in an extended state in less than 5 minutes.

[00324] *Imaging of selectively immobilized, extended macromolecules.* A macromolecule comprising fluorophore labels and biotin affinity tags is prepared and purified according to examples provided below. The macromolecule is bound to a coverslip surface comprising biotin and stretched with an electric field. Finally macromolecule is illuminated with an Arc lamp and imaged with a camera. An exemplary image is provided in Figure 16. Individual dyes and, significantly, the order of those dyes on individual macromolecules can be detected in the image.

[00325] *Preparation and Imaging of Selectively Immobilized, Extended Macromolecules.* Herein is a step-by-step example of the construction of a nanoreporter from various components. It can be appreciated that various components can be constructed or added either at the same time, before or after other components. For example, annealing patch units or flaps to a scaffold can be done simultaneously or one after the other.

[00326] In this example the starting material is a circular M13mp18 viral vector. Using a single linear strand M13mp18, patch units are annealed to it to form a double stranded scaffold. Next, flaps are added and then a target-specific sequence is ligated. Meanwhile purification steps aid to filter out excess, unattached patch units and flaps. Constructions of labeled nucleic acids (patches and/or flaps and/or other labeled oligonucleotides) that bind the nanoreporter are also described.

[00327] Upon attachment (*e.g.*, via hybridization) of a target molecule, the nanoreporter is attached to a surface and stretched. Finally the nanoreporters are imaged by a camera. Nanoreporters were generated and successfully employed to detect target molecules using methods substantially as described in this example.

[00328] *Scaffold construction.* The oligonucleotide scaffold sequence selected was analyzed using Vector NTI® software. First, a single stranded nucleic acid was made from linearizing a circular M13mp18 single stranded DNA, which was commercially purchased from New England Biolabs. The circular M13mp18 was digested with BamH1 enzyme to linearize it. Materials used consisted of M13mp18 vector (250ng/μl), Patch_1L_BamH1.02 (10μM dilution of a 100μM stock), 10X BamH1 Buffer, BamH1 enzyme. The protocol for making 0.8pmol total of linear M13mp18 involves the following steps. 1) preheat heating

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 801

block to 37 °C; 2) in a 0.65ml ependorff tube combine 40µl of 250ng/µl M13mp18 vector, 2µl of 10µM Patch_1L_BamH1.02, and 5µl of 10X BamH1 Buffer; 3) place the ependorff tube in the 37°C heating block with foil over the top. Incubate the tube at 37°C for 15 minutes to allow the patch unit to hybridize to the M13mp18 scaffold; 4) after 15 minutes add 2 µl of BamH1 enzyme and let the reaction digest at 37 °C for 30 minutes, after which add an additional 2 µl of BamH1 enzyme and let the reaction continue to digest for another 30 minutes at 37 °C (final volume of BamH1 enzyme is 8%); and 5) aliquot 10 µl into 0.65ml ependorff tubes and store in freezer (final concentration of linear M13mp18 is 200ng/µl).

[00329] *Patch unit preparation of the base patch pools (BPP).* Second, patch units are prepared in pools. Patch oligonucleotide sequences were selected for optimal length and desired homology/non-homology to M13mp18 strand and the human genomic sequence. Patches were commercially manufactured oligonucleotides (purchased from Integrated DNA technologies) either 60 or 65 nucleotide bases in length. Fifty nucleotide bases of each patch oligonucleotide are complementary to the M13mp18 single stranded DNA, 10 nucleotide bases are complementary to an adjacent patch, and 5 nucleotides base pairs are complementary to a corresponding flap. The 10 nucleotide base match between patches forms a stem structure which stabilizes the structure and helps lift the flaps off the covered scaffold so they are more available to bind labeled oligonucleotides. Synthetic binding sites, the five nucleotide bases, on the patches for binding to the flaps make leveraging the power of a modular system possible.

[00330] The base patch pools contain nine patch units all corresponding to a specific letter grouping and position on the nanoreporter. For this example, there are four different fluorescent dyes (color) labeled A, B, C, and D and 8 different positions or regions where labeled nucleic acids can bind on a nanoreporter. For example, BPP A3 corresponds to all of the A patch units at position 3 (patch units 19-27) on the nanoreporter.

[00331] The nanoreporter positions are as follows:

Position 1: Patch units 1-9 (A or C)
Position 2: Patch units 10-18 (B or D)
Position 3: Patch units 19-27 (A or C)
Position 4: Patch units 28-36 (B or D)
Position 5: Patch units 37-45 (A or C)
Position 6: Patch units 46-54 (B or D)
Position 7: Patch units 55-63 (A or C)
Position 8: Patch units 64-72 (B or D)

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 802

[00332] Materials: right and left patches, pre-annealed to each other (each oligonucleotide is at a concentration of 10 μM). Materials for making 100pmol of BPP 1: (In position 1, patch coordinate 1L is used for the BamH1 digest -this patch is not included in BPP 1): 10 μl each pre-annealed (10 μM/each) patch unit (coordinates 2-9), 5 μl [20 μM] Patch_1R (A or C). Final concentration of each patch is 1.18 pmol/ μl. Materials for making 100pmol of BPP 2-8: 10 μl each pre-annealed (10 μM/each) appropriate patch unit. There are 9 patch units added to each, or 90 μl total. Final concentration of each patch is 1.11pmol/ μl.

[00333] Below is a table of all the patch unit pools made for this example, with eight positions or regions for dye-labeled nucleic acids to bind on the nanoreporter. Positions 1, 3, 5, and 7 can bind to nucleic acid labeled with dye A or dye C, and a positions 2, 4, 6, and 8 can bind to nucleic acid labeled with dye B or dye D.

Table of resulting Basic Patch Pools (correspond to labels on tubes)

BPP-A1 [Pre-Paired, Color=A, Coordinates 1-9] Patch_(1-9)R.A Patch_(2-9)L
BPP-B2 [Pre-Paired, Color=B, Coordinates 10-18] Patch_(10-18)R.B Patch_(10-18)L
BPP-A3 [Pre-Paired, Color=A, Coordinates 19-27] Patch_(19-27)R.A Patch_(19-27)L
BPP-B4 [Pre-Paired, Color=B, Coordinates 28-36] Patch_(28-36)R.B Patch_(28-36)L
BPP-A5 [Pre-Paired, Color=A, Coordinates 37-45] Patch_(37-45)R.A Patch_(37-45)L
BPP-B6 [Pre-Paired, Color=B, Coordinates 46-54] Patch_(46-54)R.B Patch_(46-54)L
BPP-A7 [Pre-Paired, Color=A, Coordinates 55-63] Patch_(55-63)R.A Patch_(55-63)L
BPP-B8 [Pre-Paired, Color=B, Coordinates 64-72] Patch_(64-72)R.B Patch_(64-72)L

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 803

BPP-C1 [Pre-Paired, Color=C, Coordinates 1-9] Patch_(1-9)R.C Patch_(2-9)L
BPP-D2 [Pre-Paired, Color=D, Coordinates 10-18] Patch_(10-18)R.D Patch_(10-18)L
BPP-C3 [Pre-Paired, Color=C, Coordinates 19-27] Patch_(19-27)R.C Patch_(19-27)L
BPP-D4 [Pre-Paired, Color=D, Coordinates 28-36] Patch_(28-36)R.D Patch_(28-36)L
BPP-C5 [Pre-Paired, Color=C, Coordinates 37-45] Patch_(37-45)R.C Patch_(37-45)L
BPP-D6 [Pre-Paired, Color=D, Coordinates 46-54] Patch_(46-54)R.D Patch_(46-54)L
BPP-C7 [Pre-Paired, Color=C, Coordinates 55-63] Patch_(55-63)R.C Patch_(55-63)L
BPP-D8 [Pre-Paired, Color=D, Coordinates 64-72] Patch_(64-72)R.D Patch_(64-72)L

[00334] *Materials and preparation for annealing the single stranded oligonucleotide with patch units for a double stranded scaffold.* Third, patch units are prepared to be annealed to the single stranded linear M13mp18, covering the strand in order to make a double stranded oligonucleotide scaffold. Conditions for annealing 60 and 65 nucleotide base patches to the M13mp18 need to occur at high salt concentrations so that binding will be very specific and patches will not anneal to an incorrect coordinate on the M13mp18 strand. For the annealing step, each patch unit is added at a 2:1 to 4:1 ratio with the single stranded M13mp18 sequence at 0.5pmol total volume. Excess patches are removed before annealing flaps.

[00335] Materials used consisted of 20x SSC, linear M13mp18 (BamH1 digested at 0.08pmol/μl or 200 ng/μl), appropriate base patch pools (BPP) (need 8 total at 1.11pmol/μl – see above) and digital heat block set at 45°C. Annealing reaction make up is as follows. General guidelines: 2X each patch unit per M13mp18 molecule, pre-ligated flaps/patches (in position 1 or 8) added for purification later, and 5X SSC. Example (0.5pmol of scaffold with

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 804

F8 hook flaps) reaction consists of: 7.1 µl BamH1 Digested M13mp18 strand at 0.071 µM, 0.9 µl each new Base Patch Pools at 1.11 µM for first 7 positions: A1, B2, A3, B4, C5, B6 and A7:

1.7 µl A1 BPP (Pre-Annealed, 12/15; at 1.18 µM/each patch)

1.8 µl B2 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl A3 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl B4 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl C5 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl B6 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl A7 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch),

2.4 µl BPP-D8 (pool of the first seven patch units – coordinates 64, 65, 66, 67, 68, 69, and 70 at position 8 – "D" specificity) with purification tags– F8 (FHF, which anneal to patch coordinates 71L, 71R, 72L, 72R, 73L making full split-flap/patch units that have "F" specificity for use as biotin linkers, at position F8) at 0.83 µM, and 7.3 µl 20X SSC. The final reaction volume will be 29.3 µl at 0.027 pmol/ µl.

[00336] Anti-Bam oligonucleotide is also added to anneal to region in M13 that is complementary to the (missing) 1L patch unit and to prevent recircularization of the M13 scaffold during ligation.

[00337] *Annealing patch units to single stranded m13mp18 to form a double stranded scaffold.* The fourth step involves the protocol to anneal the patch units to the single stranded linear M13mp18, covering the strand in order to make a double stranded oligonucleotide scaffold, is performed in the following steps: 1) preheat heating block to 42°C, heat above reaction solution to 45°C in small PCR (or strip) tube(s) with foil over top for 15 minutes, turn heat block to 65°C and incubate for an additional 1 hour and 45 minutes and remove tubes, place on ice or freeze.

[00338] *Purification of nanoreporter scaffold using biotin and magnetic beads with streptavidin.* The fifth step occurs before attaching the flaps, where excess patch units that have not annealed to the M13mp18 strand are separated from the double stranded oligonucleotide scaffold. A purification tag with a five nucleotide base homologous region to some of the patch units' complementary five nucleotide base overhang is annealed to 'hook' the scaffold. Biotinylated oligonucleotides are annealed to the 'purification tag' and magnetic beads with streptavidin are used to capture the scaffold using the biotinylated oligonucleotides. Excess patch units are removed with the supernatant. The scaffold melts off of the magnetic beads into solution for recovery.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 805

[00339] *Anneal the d-biotin catchers to the purification tags.* Anneal the D-Biotin catchers to the purification tags on the nanoreporter (making 2X to amount of D8-flap positions available in solution, which is 2X to M13, or 4X final): 0.5pmol X 25 hook oligonucleotide positions (5 multiplied by 5), 4X makes 50 pmols translates to 0.50 µl of 100 pmol/ µl D-biotin, add 0.5 µl (D, E, F) – Biotin (at 100 µM) to sample, mix and incubate at room temperature for 30 minutes.

[00340] *Purification protocol to wash off unattached patch units from double stranded scaffold.* Anneal F-hook oligonucleotides in a 25 fold excess to nanoreporters in 5X SSC for 30 min at room temperature. Pipet 200 µl DynaBead MyOne Streptavidin™ bead solution into 1.5 ml tubes, place on magnet and remove supernatant. Wash twice with 5X SSC by resuspending and clearing with magnet as in step above. Add 80 µl of sample in 5X SSC (80 fmoles of sample in this example). Resuspend well, by placing on vortex for 15 minutes. Clear solution with magnet and transfer supernatant to fresh tubes for later gel analysis. While on magnet, wash pellets (do not resuspend) with 80 µl TE by pipeting over pellet three times with the same 80 µl volume originally added. Remove wash, place in freshy "wash" tubes for analysis. Heat up TE buffer to 45°C, add 80 µl to each pellet and resuspend. Place tubes on 45°C heat block for 15 minutes, pipetting up/down once to insure beads remain suspended. Immediately clear product with magnet while warm and save. The majority of purified nanoreporters should be present in this product eluted at 45°C.

[00341] *Annealing and ligation of flaps to scaffold.* The sixth step involves split flap oligonucleotides which are annealed to the scaffold to make a 'covered scaffold.' Purification with magnetic beads is performed afterwards to remove excess split flaps. Ligation of the covered scaffold is done using T4 ligase to increase the stability of the structure. Only one type of flap is needed per fluorescent dye. Flaps are either 95 or 100 bases in length and have regions complementary to the patches, to labeled oligonucleotides and to each other. Each flap has 15 base repeating sequences for binding to labeled oligonucleotides. The repeat sequences are based on Lambda sequences that have been analyzed to remove any palindromes and hairpin structures.

[00342] Conditions for annealing the flaps are as follows. The sequence on the flaps that corresponds to the patch is 5 nucleotide base pairs long, and therefore the flaps anneal specifically to the patches even at high salt concentrations. The ratio of flaps to patches is 2:1. In order to increase stability at high temperatures, ligation of patches to each other and the flap to the patches may be carried out in the same reaction.

[00343] 1) Quantify the purified scaffold using a spectrometer at A260nm. Calculate the volume needed for appropriate amount of nanoreporter to prepare. For this example we

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 806

used 110 ng or 0.023 pmol, reading at A260nm shows 7.7ng/ µl, or 14.3 µl for 110ng. 2) Setup ligation reaction as follows (volume will vary, depending on the purification and scale). Currently using 1.5X flaps to patches, calculate accordingly. For this example, there are four different fluorescent dyes (color) labeled A, B, C, and D and 8 different positions or regions where dye-labeled nucleic acids can bind on a nanoreporter. The number of positions for each color (in this case 1-4) multiply by 9 multiply by 1.5 moles of scaffold = moles of flaps to use.

[00344] For the nanoreporter with fluorescent dye in the sequence/positions [ABABCBAD] :

ABABCBAD =

A:40.5 x .023 = .93pmol; vol: .93ul of SF (split flap)-AL @1µM

0.93µl of SF-AR at 1µM

B:40.5 x .023 = .93pmol; vol: .93ul of SF-BL @1 µM

0.93µl of SF-BR at 1µM

C:13.5 x .023 = .31pmol; vol: .31ul of SF-CL @1µM

0.31µl of SF-CR at 1µM

D:13.5 x .023 = .31pmol; vol: .31ul of SF-DL @1µM

0.31µl of SF-DR at 1µM

[00345] Ligation reaction (25 µl total) consists of : Split Flaps (see above; 4.96µl, or ~5µl total), 14.3µl of MODB-Scaffold at 0.0016 pmol/µl, 2.5µl 10X T4 ligation Buffer, 2.2µl NanoPure H2O and 1µl T4 ligase. Incubate tubes 5 minutes at 45°C. Move to 37°C water bath, inc. for 5min. Add 1µl T4 ligase to samples. Incubate for additional 1hour at 37°C. Freeze immediately, or heat at 75°C for 5 minutes to kill T4 ligase.

[00346] *Ligation of target-specific sequences to nanoreporters.* The seventh step involves ligation of a target-specific sequence to the nanoreporter. A DNA target-specific sequence is designed to be complementary to the target molecule, which can be RNA (*e.g.*, mRNA) or DNA (*e.g.*, cDNA or genomic DNA). The target-specific sequence can be from 35, 60 or 70 nucleotide bases in length. The target-specific sequence can be ligated to the scaffold using a single stranded overhanging region on the covered scaffold. The scaffold with a single type of target-specific sequence can be manufactured separately and then mixed to form libraries.

[00347] *Nanoreporter construction.* Addition of oligonucleotides to a nanoreporter can be done at any point during the construction of a nanoreporter. In certain aspects of the present invention, a labeled oligonucleotide is 15 nucleotide bases long. On the 5' end, a single fluorophore dye is attached. Oligonucleotides with a particular fluorophore dye will generally have the same sequence. These labeled oligonucleotides bind to the repeat sequences of the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 807

split flaps. Fluorophores best suited for this example include but are not limited to Alexa 488, cy3, Alexa 594, and Alexa 647. The 15 nucleotide base length holds the fluorophores far enough apart so that they cannot quench each other and ensure that the labeled nucleic acids will be stable (will not melt off complementary strand) at conditions in the visualization process. Labeled oligonucleotides are stable at 40 °C. This short length also allows for packing a large number of fluorescent dyes onto the flaps. In certain aspects of the invention, labeled oligonucleotides are introduced during the target sample processing.

[00348] *Attachment of nanoreporters to target molecules.* Nanoreporters can be attached to target molecules using any means known to one of skill in the art. In an exemplary embodiment, dual nanoreporters are hybridized to target molecules by mixing 250pmols each of both the first probe and the second probe with 125 pmols of target. The total volume is adjusted to 4µl and a final concentration of buffer of 5X SSC. This mixture is incubated in a covered PCR tube overnight at 42 degrees to allow hybridization to occur.

[00349] *Surface attachment.* Once the nanoreporters are attached to both target molecule and corresponding labeled nucleic acids, *i.e.*, nucleic acids attached to label monomers, they are attached to a surface and stretched in resolve the order of signals emitted by the label monomers and thus identify the target molecule. In this example, the nanoreporters are stretched to spatially resolve their fluorescent dye codes which correspond to a particular target molecule. The nanoreporters are stretched by attaching one end to a surface (in this example – a coverslip, see preparations below). Two methods for surface attachment may be used: A) streptavidin coated slides from Accelr8 Corporation with the nanoreporters being biotinylated and B) biotin coated slides with the nanoreporters having streptavidin. In buffer, the nanoreporters are brought into contact with the active surface and allowed to incubate for a period of time. The reaction is performed in flow cells which were made from PDMS molded in etched silicon wafers to make the channels. Metal tubing is used to core wells at the ends of the channels for buffer and sample insertion. Channel dimensions are 0.5 mm or 1 mm wide and 54 µm high. Once the sample has been loaded into the flow cell lane and incubated, the nanoreporters should be attached. Nanoreporters can be stretched either by applying a voltage or by removing the liquid with a receding meniscus leaving the strings stretched and dry.

[00350] *Preparation of surface and assembly of device.* The binding surfaces (Accelr8 brand Streptavidin-OptiChem, coated coverslips) are shipped in units of 5 surfaces per slide container, and each container is enclosed with a package of silica dessicant in a foil pouch. The pouches are stored at -20 °C until use. To prepare the surface for binding, a pouch is first pulled from the freezer and allowed to come to room temperature over several minutes. If

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 808

previously unopened, the pouch is then sliced along one edge to form a slit, and the container of surfaces is removed. Upon removal of the required surface, the container is replaced in the pouch with its dessicant, the slit is sealed closed with a strip of packaging tape, and the pouch is replaced in the freezer.

[00351] The surface is then lightly rinsed with a stream of Nanopure water (Barnstead Nanopure Diamond) and soaked for 10 minutes in 0.2μm-filtered 1X PBS in a clean, slotted Coplin jar. After soaking, the surface is dipped in Nanopure water and dried by blowing filtered nitrogen across the surface edge.

[00352] The PDMS device used to mate with the surface and provide localization of the sample is cleaned just before use by applying cellophane tape to the PDMS surface and then peeling away dust or other particles which may have become attached during storage. The binding side of the Accelr8 surface is laid face-up, and the clean PDMS structure is centered, channel side down, on the surface. PDMS adheres readily to coated glass, and no further attachment mechanism is necessary.

[00353] *Sample binding and washing.* The sample is bound to the surface by first applying a five μL drop of the sample (currently diluted in 100mM sodium borate buffer, pH 9.8) in one well of the chosen lane. The drop should just touch the point at which the channel joins the well (some sample may wick into the channel at this point). The channel is filled, and binding is equalized throughout the channel, by pulling the droplet through the channel to the opposite well using a very weak vacuum (< 2 kPa). The process is repeated for the other samples in their respective lanes. Excess fluid is then removed from the wells, the wells are taped to reduce evaporation, and the device is incubated at room temperature in the dark for 20 minutes.

[00354] After binding, the tape is removed, and the top well of each lane is filled with 100 μL of the borate buffer described above. About 20 μL of that buffer is pulled through the channels to the other wells using the vacuum, and the process is repeated once. All borate buffer is then removed from all wells, and the top well is filled with 1X TAE, pH 8.3. About 50μL TAE is pulled through the channel, then all TAE is removed and the well is refilled. The process is repeated three times, for a total of about 150 μL of TAE rinse. Finally, all wells are filled with 100 μL 1X TAE.

[00355] *Electrostretching.* The bottom of the coverslip/PDMS device is spotted with immersion oil and placed on the microscope. Electrodes are inserted into the wells on opposite ends of the first PDMS channel (negative electrode in top well, positive in bottom). The first image of the channel will be taken close to the bottom well; the microscope stage is adjusted so that the area of interest is in focus. Voltage (200 V) is then applied across the channel. Voltage

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 809

is supplied by a DC power supply (Agilent E3630A) and amplified 100X through a home-built amplifier. After the current is applied, focus is readjusted, and the imaging process begins. The electrostretching and imaging process is then repeated with the remaining channels. Image the bindings.

[00356] *Light source for the fluorescent dyes on the nanoreporter.* In using an arc lamp as a light source, the best fluorophore selection is the brightest types without leading to fluorescent overlap such as Alexa 488, Cy3, and Alexa 594. Weaker fluorescent dyes such as Alexa 647 and Cy5.5 may also be used.

[00357] *Filters to image the fluorescent dyes on the nanoreporter.* For the selected fluorophores Alexa 488, Cy3, Alexa 594 and Alexa 647 there maybe an overlap between the Cy3 and Alexa 594. However, custom ordering an emission filter with a bandwidth of 572-600 nm minimizes the overlap.

[00358] *Microscope and objective lens to image the nanoreporters.* The microscope model used was the Nikon Eclipse TE2000E from Nikon Incorporation using the inverted fluorescence imaging station which has 6 filter cassettes that allow the selection of fluorescent emission from multiple fluorescent dye candidates. For the selected dyes, the optical resolution required is about 400 nm for all the wavelengths (500-700 nm). The selected objective lens is the Nikon Plan Apo TIRF lens which has a NA of 1.45 and magnification of 60. The optical resolution is ~210-300 nm for different wavelengths.

6.2 EXAMPLE 2: PATCH/FLAP NANOREPORTER MANUFACTURING PROTOCOL

[00359] This example demonstrates another way of making a nanoreporter which consists of a single stranded linear M13mp18 viral DNA, oligonucleotide patch units and long flaps. Nanoreporter label units were successfully generated using methods substantially as described in this example. Pre-phosphorylated patch units and flaps are added together with the M13mp18 DNA vector and ligated together. After the ligation of the flaps to the patch units which are ligated to the M13mp18 DNA, the BamH1 enzyme is introduced to linearize the vector. Prepare a batch of nanoreporters starting with 5 μg of M13mp18 as a scaffold. The hybridization may be scaled up accordingly to the desired amount. This process will take about 1-2 days to complete.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 810

[00360] Materials:

Qty	Item	Vendor
20	250 ug/ul M13mp18 viral ssDNA	New England Biolabs
27 ul	0.74 pmol/ul Oligonucleotide Patch Unit Mix	IDT
8 ul	Long Flap Oligonucleotide A 100 pmol/ul	IDT
8 ul	Long Flap Oligonucleotide B 100 pmol/ul	IDT
0.5 ul	Flap patch Oligos @ 100 pmol/ul from plates #529916 and #610591	IDT
31 ul	T4 Ligase 10x buffer	Fermentas
19 ul	T4 Ligase	Fermentas
15 ul	Optikinase 10x buffer	USB
4.2 ul	100 mM ATP	ANY
5 ul	Optikinase Enzyme 10 units/ul	USB
1 ul	BamH1 oligonucleotide 10pmol/ul	IDT
20 ul	BamH1 10x buffer	Fermentas
3 ul	BamH1 Enzyme 10 units/ul	Fermentas

[00361] Preheat water bath to 37°C and 55°C before beginning protocol. Make sure buffers are all well mixed and thawed before using. A work plate should be available and labeled with the ordered oligos from IDT in plates #529916 and #610591. Take these two plates out and thaw at room temperature for 0.5-1 hours and spin down contents before removing the tape that covers the wells. Four separate reactions will be set up in 1.5 ml eppendorf tubes using specific oligonucleotides from these plates. To begin label these four separate tubes with roman numerals on their caps. Columns 5 and 6 A through H are for reaction I, Columns 7 and 8 A through H are for reaction ii are all found in plate #529916. Columns 1 and 2 are for reaction iv, and Columns 3 and 4 are for reaction iii.

[00362] *Flap Ligations (Step A)*: Label four separate 1.5 ml tubes with roman numerals i through iv (mentioned above). Add the reagents below accordingly to each 50 µl reaction containing: 5 µl 10x ligase buffer, 0.5 µl/oligonucleotide from designated wells from plates #529916 and #610591, 4 µl Long Flap Oligo/reaction (A or B) for reactions I, ii and iv. 3ul of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 811

LF for area iii, 29 H20 for reactions I,ii and iv. 32 µl H20 for reaction iii, and 4 µl T4 ligase. Preanneal oligos in this mix without the ligase at 37 °C for half an hour. Add ligase as last reagent and allow to ligate at room temperature for at least four hours. Product concentration is 1 pmol/flap/µl.

[00363] *Flap Ligation Phosphorylation (Step B)*: Label four separate 1.5 ml tubes with roman numerals again, one through four with a P inside a circle to designate that the products are phosphorylated. Add the following reagents to the corresponding tube: 10 µl/Flap ligation. reaction (take 10 µl/flap ligation reaction above), 2.5 µl Optikinase buffer, 0.5 µl 100 mM ATP, 11.5 µl H20, and 0.5 µl Optikinase enzyme. Incubate at 37 °C for 1 hour. Product concentration 0.4 pmol/flap/µl.

[00364] *Oligonucleotide Patch Unit Phosphorylation (Step C)*: 27 µl Oligonucleotide Patch Unit mix 0.74 pmol/µl, 5 µl 10x buffer, 1 µl 100 mM ATP, 3 µl Optikinase enzyme, and14 µl H20. Once reagents are all together gently mix the solution by flicking the tube a few times and spin down. Incubate at 37 °C for 1 hour.

[00365] *Hybridization to M13mp18 scaffold (Step D)*: In a new 1.5 ml tube add the following reagents: 20 µl M13mp18 at 250 ng/µl, 27 µl Phosphorylated Oligonucleotide Patch Units 0.4 pmol/µl (Step C), 12.5 µl/ Phosph. Flap Ligation (Step B) preheat at 55 °C for 5 minutes and put on ice, 11 µl 10x ligase buffer and heat entire mixture at 55 °C for 1 minute.. Hybridize mixture at 37 °C for at least 4 hours.

[00366] *Ligation (Step E)*: Spin down eppendorf contents. Add 1.2 µl 100 mM ATP and 3 µl T4 ligase. Gently mix contents by flicking the tube, then spin down.

[00367] *BamH1 Digest (Step F)*: 1 µl of 10 pmol BamH1 oligo, 20 µl 10x BamH1 buffer and hybride at 37 ~ 1 hour. Adjust volume to 200 µl. Add 3 µl BamH1 enzyme. Incubate at 37 °C for 1 hour.

[00368] First step: start by adding 20 µl of M13mp18 (NEB 250 µg/ml) to a clean 1.7 ml eppendorf tube. Take 5 µl of Phosphorylated Flap ligation reaction and preheat it at 70 for 2 minutes and immediately put on ice. Add the 5 µl of each Phosphorylated Flap Ligation reaction (1 pmol/flap/µl) to the tube and gently mix by pipetting a few times. Incubate the eppendorf tube at 37 °C for 1 hour.

[00369] Second step: put 13.5 µl Oligonucleotide Patch Unit Mix (0.74 pmol/µl) and 1 µl of Acrydite Mix (10 pmol/µl)in a new eppendorf 1.7 ml eppendorf tube. Add 5 µl 10x Optikinase buffer, 1 µl 100 mM ATP and 27.5 µl H20. Mix gently by pipetting the solution. Add 2 µl Optikinase enzyme, gently mix by pipetting and incubate at 37 °C for 1 hr.

[00370] Third step: take the phosporylated oligos rxn and add it entirely to the contents of the M13mp18+Flaps Hybridization. The reaction is mixed gently by pipetting and it is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 812

allowed to incubate at 30 °C for 1 hour. After the hybridization is complete adjust the ATP by adding 1 µl (100 ATP) to the reaction.

[00371] Fourth step: spin down contents in eppendorf tube and add 4 µl T4 Ligase enzyme (5 units/µl), mix gently by pipetting. Incubate at room temperature for at least four hours. Add 1 µl BamH1 oligonucleotide (10 pmol/µl) to hybridize at room temperature while ligation is taking place.

[00372] Fifth step: digest ligation reaction by adding 4 µl BamH1 enzyme (5 units/µl), mix gently by pipetting and incubate at 37 °C for 1 hour. Once the incubation period is over. Take an aliquot of 500 ng for QC.

[00373] Sixth step: treat with Psoralen, UV or DMPA light for 15 minutes.

[00374] Calculations include:

5 µg of M13 = 20 µl stock from New England Biolabs = 2 pmols

Oligonucleotide mix: 180-34 flap areas – 10 Acrydite modified Oligos = 0.74 pmol/oligo

10 pmols/oligonucleotide = 13.5 µl = 1350 pmols

Optikinase 1unit converts 1 nmol of phosphate to ends - use excess. 4 µl of Optikinase was used.

6.3 EXAMPLE 3: PROTOCOL FOR PRODUCTION OF RNA NANOREPORTERS

[00375] Nanoreporters were generated and successfully employed to detect target molecules using methods substantially as described in this example. An example of target detection using such this method is shown in Figure 6.

[00376] *Scaffold Production.* Single-stranded circular M13mp18 DNA (USB) was annealed to a 5-fold molar excess of an oligonucleotide complementary to the Bam HI recognition site (Bam Cutter oligo) and cut with Bam HI restriction enzyme to yield a linear single-stranded DNA backbone. An oligonucleotide complementary to the Bam Cutter oligonucleotide (anti-Bam oligonucleotide) was subsequently added in 50-fold excess to sequester free Bam Cutter oligonucleotide and thus prevent recircularization of the M13 during later steps. The linear M13 molecule serves as a scaffold onto which RNA patches, or RNA segments, with incorporated fluorophores can be annealed.

[00377] *PCR to form double-stranded positions on the M13 scaffold.* Ten sets of oligonucleotide primer pairs were designed to create 10 different regions along the M13 scaffold. Each pair contains one primer which has a T7 RNA polymerase promoter at the 5' end. Regions 2-7 are designed to be 900 bases (approximately 300 nm) long, as this is the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 813

approximate size of a diffraction-limited spot (the smallest spot that can be achieved with standard optics). Regions 1 and 8 have both long and short versions: the long versions cover the whole 900-base region, while the short versions cover only a portion of the 900-base region to allow a target-specific sequence to be ligated. Thus a target-specific sequence can be attached to either end. The ends can also be used for attachment of anchors or tags.

[00378] PCR is performed using Taq polymerase and 0.5ng of double-stranded M13mp18 (USB) as a template. Reactions are cleaned up using a Qiaquick purification kit from Qiagen. Each PCR reaction yields a double-stranded fragment corresponding to one specific segment as illustrated below. These fragments are used as templates for the *in vitro* transcription of the RNA segments.

[00379] *In vitro Transcription to Produce Dark RNA Segments.* Using the PCR products described above as double-stranded templates, RNA segments are generated using an *in vitro* transcription kit from Ambion (Megascript T7 kit). The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

[00380] *In vitro Transcription to Produce RNA Segments Modified With Aminoallyl Groups.* Using the PCR products described above as double-stranded templates, RNA segments for later dye-coupling are generated using an *in vitro* transcription kit from Ambion (MessageAmp aRNA kit). Aminoallyl-modified UTP nucleotides are incorporated into the RNA segments during transcription. The products of the transcription reactions are purified (including treatment with DNAse I to remove template) using a RNeasy Kit from Qiagen.

[00381] *Dye Coupling of Aminoallyl RNA Segments to Produce Colored RNA Segments.* 20-100μg of aminoallyl-modified RNA segment is coupled with NHS-ester dyes using Ambion Aminoallyl Labeling Kit. Dyes used include Alexa 488, Alexa 594 and Alexa 647 (Invitrogen/Molecular Probes) as well as Cy3 (Amersham).

[00382] Each segment is made separately in four colors so that each position on the scaffold can be filled with a segment in any of the four colors; thus different colors can be added at different positions to create many unique color combinations.

[00383] In this particular embodiment, adjacent segments must be of different colors or there may be dark segments interspersed so that each segment is detected as an individual 'spot'. Dark segments may be used as part of the nanoreporter code.

[00384] *Assembly of the Label Molecule.* Segments for each position are annealed in a 2:1 ratio of segment to M13 scaffold in 1X SSPE buffer at 70°C for 2 hours. An assembled nanoreporter with labeled RNA segments is depicted in Figures 3A and 3B. Figure 3A depicts

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 814

a nanoreporter in which only alternate "spots" (1, 3, 5 and 7) are labeled, and Figure 3B depicts a nanoreporter in which every spot is labeled.

6.4 EXAMPLE 4: DETECTION OF TARGET (S2) RNA AND DNA MOLECULES USING AN RNA NANOREPORTER/GHOST PROBE COMBINATION

[00385] *Synthesis of Probe and Target Oligonucleotides.* S2 DNA target oligonucleotide was synthesized and purified by polyacrylamide gel electrophoresis (Integrated DNA Technologies). S2 RNA target molecules were generated by *in vitro* transcription of PCR products corresponding to region of cloned SARS coronavirus gene (Invitrogen) using an Ambion Megascript™ kit per manufacturer's instructions. The S2 ghost probe (Figure 6A (i) was complementary to a specific 50-base region of the S2 target sequence (S2-*a*) and was synthesized with a biotin-TEG monomer at the 5' end and purified by high performance liquid chromatograpy (Integrated DNA Technologies).A second oligonucleotide with 50bps complementary to the S2 target (S2-*b*) plus 9 bp of a additional sequence used for ligation to the M13 scaffold (59bp total) was synthesized and purified by HPLC (Integrated DNA Technologies). Note that S2-*a* and S2-*b* target regions were not overlapping.

[00386] *Nanoreporter synthesis.* Oligonucleotide S2-*b* was ligated to the 5' end of linearized M13 [Figure 6A (iii)], and the resulting product was purified away from residual unligated oligonucleotide by size-exclusion filtration through a YM100 filter (Millipore) per manufacturer's instructions. Amino-allyl-modified RNA segments complementary to M13 is positions 2, 4, 6, and 8 (Figure 1C) were generated from *in vitro*-transcription of DNA templates (PCR products) via the Ambion Megascript™ kit per manufacturer's instructions. The segments were then coupled to NHS-ester-modified Alexa 647 dye (Invitrogen) per Ambion's instructions (amino allyl MessageAmp™ II aRNA kit). RNA segments corresponding to positions 1, 3, 5, and 7 of the M13 scaffold (Figure 1C) were generated as unmodified *in vitro*-transcribed RNAs from DNA templates as described above. Assembly of the nanoreporter was carried out by annealing 10 fmol/μl of each of the eight segments to 5 fmol/μl of the M13-S1-*b* scaffold for 2 hours at 70°C in 1X SSPE buffer (150 mM sodium chloride, 10 mM sodium phosphate, 1mM EDTA). The final product was a nanoreporter with 4 segments labeled with A647 (red) interspersed with dark segments.

[00387] *Hybridization conditions.* Hybridization of nanoreporters and ghost probes to target were carried out under the following conditions: 5X SSPE (750 mM sodium chloride, 50 mM sodium phosphate, 5 mM disodium EDTA), 40pM ghost probe (attachment oligonucleotide S2-*a*), 40pM Nanoreporter S2-*b*, 100 ng/μl sheared salmon sperm DNA, 5X

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 815

Denhardt's solution and 0.1% Tween. Final target concentrations were 20pM S2 DNA target (Figure 6B) and 1pM S2 RNA target (Figure 6C). No target was added to the negative control (Figure 6C). The hybridization reaction was incubated at 65°C for at least 16h.

[00388] Hybridization reactions were diluted 1:2 with 100mM Borate buffer solution (pH 9.8) and introduced into a flow cell channel and bound to a streptavidin-coated coverslip forming the bottom of the channel (Streptavidin-OptiChem coverslips from Accelr8). Attachment to the slide by one end of the nanoreporter/target/ghost probe complex was achieved via interaction of the biotinylated ghost probe with the streptavidin surface. After rinsing the channel with additional borate buffer to remove excess reporters not bound to the surface, the buffer was exchanged with 1X TAE (40 mM Tris-acetate, 1 mM EDTA) and a current of 200V was applied to stretch out the nanoreporter/target complexes during image capture.

[00389] Images were obtained using a Leica DMI 6000B microscope with a 63X oil immersion objective (1.4 NA), Xcite-120 light source (Exfo), customized filter sets (Chroma Technologies), an Orca-ER CCD camera (Hamamatsu) and Metamorph data acquisition software (Molecular Devices). As predicted, when the correct target molecule S2 hybridizes [Figure 6A (ii)] to both ghost probe [Figure 6A (i), S2-*a*] and S2-*b* target-specific nanoreporter [Figure 6A (iii)], the ghost probe/target/nanoreporter complex forms a single species that attaches to the slide and was visualized as 4 spots when exposed to 647nm wavelength light (Figure 6B, 6C, and 6E). The amount of binding was dependent on the target concentration. There was no significant binding in absence of S2 target sequence (Figure 6D).

6.5 EXAMPLE 5: NANOREPORTER COMPRISING A MONOVALENT ANTIBODY FRAGMENT

[00390] Where a target molecule is a protein or polypeptide, a nanoreporter can be generated in which the nanoreporter scaffold is a nucleic acid and the target-specific sequence is a monovalent antibody fragment. Using routine methods, an antibody that recognizes a target molecule of interest is optionally digested with pepsin to generate F(ab')2 fragments. The two parts of the antibody or the two F(ab')2 fragments generated by the pepsin digestion are separated by mild reduction, for example with 2-mercaptoethylamine. This reduction separates either the antibody or the two F(ab')2 fragments into two monovalent fragments with two sulfhydryl groups that can be functionalized.

[00391] A heterobifunctional crosslinking reagent (*e.g.*, m-Maleimidobenzoyl-N-hydroxysuccinimide ester from Pierce Biotechnology Inc.) is used to attach a maleimide to an oligonucleotide with an amine modification (which can be ordered from many sources, such as

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 816

Integrated DNA Technologies). The NHS on the cross-linking reagent is reacted with the amine on the oligonucleotides to produce a maleimide-conjugated oligonucleotide.

[00392] This maleimide conjugated oligonucleotide is then reacted with one of the sulfhydryl groups on the antibody fragment. Due to steric limitations, it is preferalbe that only one oligonucleotide is be attached to each fragment. This monovalent antibody fragment attached to an oligonucleotide can then be hybridized to a complementary sequence on a nanoreporter scaffold, to generate a nanoreporter in which the target-specific sequence is an antibody sequence.

[00393] *Surface attachment.* Once the nanoreporters are attached to both target molecule and corresponding labeled nucleic acids, *i.e.*, nucleic acids attached to label monomers, they are attached to a surface and stretched in resolve the order of signals emitted by the label monomers and thus identify the target molecule. In this example, the nanoreporters are stretched to spatially resolve their fluorescent dye codes which correspond to a particular target molecule. The nanoreporters are stretched by attaching one end to a surface (in this example – a coverslip, see preparations below). Two methods for surface attachment may be used: A) streptavidin coated slides from Accelr8 Corporation with the nanoreporters being biotinylated and B) biotin coated slides with the nanoreporters having streptavidin. In buffer, the nanoreporters are brought into contact with the active surface and allowed to incubate for a period of time. The reaction is performed in flow cells which were made from PDMS molded in etched silicon wafers to make the channels. Metal tubing is used to core wells at the ends of the channels for buffer and sample insertion. Channel dimensions are 0.5 mm or 1 mm wide and 54 μm high. Once the sample has been loaded into the flow cell lane and incubated, the nanoreporters should be attached. Nanoreporters can be stretched either by applying a voltage or by removing the liquid with a receding meniscus leaving the strings stretched and dry.

[00394] *Preparation of surface and assembly of device.* The binding surfaces (Accelr8 brand Streptavidin-OptiChem, coated coverslips) are shipped in units of 5 surfaces per slide container, and each container is enclosed with a package of silica dessicant in a foil pouch. The pouches are stored at -20 °C until use.

[00395] To prepare the surface for binding, a pouch is first pulled from the freezer and allowed to come to room temperature over several minutes. If previously unopened, the pouch is then sliced along one edge to form a slit, and the container of surfaces is removed. Upon removal of the required surface, the container is replaced in the pouch with its dessicant, the slit is sealed closed with a strip of packaging tape, and the pouch is replaced in the freezer.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 817

[00396] The surface is then lightly rinsed with a stream of Nanopure water (Barnstead Nanopure Diamond) and soaked for 10 minutes in 0.2μm-filtered 1X PBS in a clean, slotted Coplin jar. After soaking, the surface is dipped in Nanopure water and dried by blowing filtered nitrogen across the surface edge.

[00397] The PDMS device used to mate with the surface and provide localization of the sample is cleaned just before use by applying cellophane tape to the PDMS surface and then peeling away dust or other particles which may have become attached during storage. The binding side of the Accelr8 surface is laid face-up, and the clean PDMS structure is centered, channel side down, on the surface. PDMS adheres readily to coated glass, and no further attachment mechanism is necessary.

[00398] *Sample Binding and Washing.* The sample is bound to the surface by first applying a 5 μL drop of the sample (currently diluted in 100mM sodium borate buffer, pH 9.8) in one well of the chosen lane. The drop should just touch the point at which the channel joins the well (some sample may wick into the channel at this point). The channel is filled, and binding is equalized throughout the channel, by pulling the droplet through the channel to the opposite well using a very weak vacuum (< 2 kPa). The process is repeated for the other samples in their respective lanes. Excess fluid is then removed from the wells, the wells are taped to reduce evaporation, and the device is incubated at room temperature in the dark for 20 minutes.

[00399] After binding, the tape is removed, and the top well of each lane is filled with 100 μL of the borate buffer described above. About 20 μL of that buffer is pulled through the channels to the other wells using the vacuum, and the process is repeated once. All borate buffer is then removed from all wells, and the top well is filled with 1X TAE, pH 8.3. About 50μL TAE is pulled through the channel, then all TAE is removed and the well is refilled. The process is repeated three times, for a total of about 150 μL of TAE rinse. Finally, all wells are filled with 100 μL 1X TAE.

[00400] *Electrostretching.* The bottom of the coverslip/PDMS device is spotted with immersion oil and placed on the microscope. Electrodes are inserted into the wells on opposite ends of the first PDMS channel (negative electrode in top well, positive in bottom). The first image of the channel will be taken close to the bottom well; the microscope stage is adjusted so that the area of interest is in focus.

[00401] Voltage (200 V) is then applied across the channel. Voltage is supplied by a DC power supply (Agilent E3630A) and amplified 100X through a amplified by a high voltage amplifier (Matsusada Precision Inc.). After the current is applied, focus is readjusted, and the imaging process begins.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 818

[00402] The electrostretching and imaging process is then repeated with the remaining channels. Image the nanoreporters.

[00403] *Light source for the fluorescent dyes on the nanoreporter.* In using an arc lamp as a light source, the best fluorophore selection is the brightest types without leading to fluorescent overlap such as Alexa 488, Cy3, and Alexa 594. Weaker fluorescent dyes such as Alexa 647 and Cy5.5 may also be used.

[00404] *Filters to image the fluorescent dyes on the nanoreporter.* For the selected fluorophores Alexa 488, Cy3, Alexa 594 and Alexa 647 there maybe an overlap between the Cy3 and Alexa 594. However, custom ordering an emission filter with a bandwidth of 572-600 nm minimizes the overlap.

[00405] *Microscope and objective lens to image the nanoreporters.* The microscope model used is the Nikon Eclipse TE2000E from Nikon Incorporation using the inverted fluorescence imaging station which has 6 filter cassettes that allow the selection of fluorescent emission from multiple fluorescent dye candidates. For the selected dyes, the optical resolution required is about 400 nm for all the wavelengths (500-700 nm). The selected objective lens is the Nikon Plan Apo TIRF lens which has a NA of 1.45 and magnification of 60. The optical resolution is ~210-300 nm for different wavelengths.

[00406] Five minutes before using the microscope (Nikon Eclipse TE2000E), turn on the light source (X-cite 120, Exfo Corporation) and make sure the intensity is the maximum. Turn on the CCD camera driver (Hamamatsu, Orca Ag) and the shutter controller. Use the oil objective of 60 x 1.45NA (Plan Apo TIRF, Nikon) to evaluate the nanoreporters. For all the nanoreporter evaluations the optivar is set at 1x. Open the Metamorph software (Universal Imaging Corporation). Acquire the images using the corresponding filter sets such as cy3, A647 (Chroma Technologies).

7. REFERENCES CITED

[00407] All publications, patents and patent applications cited in this specification are herein incorporated by reference to the same extent as if each individual publication or patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety herein for all purposes.

[00408] The present invention can be implemented as a computer program product that comprises a computer program mechanism embedded in a computer readable storage medium. For instance, the computer program product could contain the program modules shown in Fig. 9. These program modules can be stored on a CD-ROM, DVD, magnetic disk storage product, or any other computer readable data or program storage product. The program

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 819

modules can also be embedded in permanent storage, such as ROM, one or more programmable chips, or one or more application specific integrated circuits (ASICs). Such permanent storage can be localized in a server, 802.11 access point, 802.11 wireless bridge/station, repeater, router, mobile phone, or other electronic devices. The program modules in the computer program product can also be distributed electronically, via the Internet or otherwise, by transmission of a computer data signal (in which the software modules are embedded) either digitally or on a carrier wave.

[00409] Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. For instance, data storage module 44, label identification module 50, and probe identification module 54 can be combined into a single program, can each be a separate program, or could, in fact, be dispersed in multiple (*e.g.*, three or more) programs. The specific embodiments described herein are offered by way of example only, and the invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 820

What is claimed is:

1. A computer program product, wherein the computer program product comprises a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism for detecting the presence of a probe within a sample overlayed on a substrate, wherein the probe comprises a plurality of spatially arranged labels, the computer program mechanism comprising:

a data storage module comprising instructions for storing a plurality of light images, each light image in the plurality of light images being for light received from the sample at a wavelength range in a plurality of different wavelength ranges;

a label identification module comprising instructions for identifying a plurality of labels, in the plurality of light images, that are proximate to each other on the substrate, wherein a spatial order of the plurality of labels determines a string sequence of the plurality of labels; and

a probe identification module comprising instructions for determining whether the string sequence of the plurality of labels comprises a valid reporter sequence, wherein

when the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe; and

when the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

2. The computer program product of claim 1, wherein a first label in the plurality of labels is associated with a first position on the substrate that emits light in a first wavelength range in the plurality of different wavelength ranges and a second label in the plurality of labels is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges.

3. The computer program product of claim 2, wherein a portion of the first wavelength range overlaps with a portion of the second wavelength range.

4. The computer program product of claim 2, wherein the first wavelength range does not overlap with the second wavelength range.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 821

5. The computer program product of claim 1, wherein each label in the plurality of labels is associated with a position on the substrate that emits more than a threshold amount of light in at least one light image in the plurality of light images.

6. The computer program product of claim 1, wherein the computer program mechanism further comprises a lookup table comprising a plurality of valid reporter sequences, and wherein the probe identification module further comprises instructions for comparing the string sequence of the plurality of labels to valid reporter sequences in the lookup table.

7. The computer program product of claim 6, wherein the lookup table is dimensioned and configured to hold up to 4^4 different valid reporter sequences.

8. The computer program product of claim 6, wherein the lookup table is dimensioned and configured to hold up to 7^4 different valid reporter sequences.

9. The computer program product of claim 6, wherein the lookup table is dimensioned and configured to hold more than 8 different valid reporter sequences.

10. The computer program product of claim 1, the probe identification module further comprising instructions for storing the string sequence of the plurality of labels that is not confirmed as a valid reporter sequence.

11. The computer program product of claim 1, wherein the data storage module further comprises instructions for aligning a first light image to a second light image in the plurality of light images using a plurality of fiducials present on the substrate.

12. The computer program product of claim 1, wherein the position of the probe on the substrate is random.

13. The computer program product of claim 1, wherein the probe consists of a single molecule.

14. The computer program product of claim 1, wherein the probe comprises a molecular scaffold and wherein each label in the plurality of labels represents a different position on said molecular scaffold.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 822

15. The computer program product of claim 14, wherein each position on said molecular scaffold that is represented by a label is separated from a neighboring position on the scaffold by a spacer.

16. The computer program product of claim 1, wherein the probe comprises a single stranded deoxynucleic acid or ribonucleic acid scaffold and wherein each label in the plurality of labels is represented by a dye laden single stranded deoxynucleic acid or ribonucleic acid sequence that hybridizes to a different position on the scaffold.

17. The computer program product of claim 1, wherein

the probe comprises a molecular scaffold having a first end and a second end;

a target specific sequence is covalently attached to the first end;

a binder sequence is covalently attached to the second end; and

the probe is linearly arranged on the substrate through (i) the binding of the target specific sequence to a first molecular entity that is bound to a first position of the substrate and (ii) the binding of the binder sequence to a second molecular entity that is bound to a second position on the substrate.

18. The computer program product of claim 17, wherein

the first molecular entity is a target (single stranded deoxynucleic acid or ribonucleic acid)-biotin complex, and

the second molecular entity is a predetermined (single stranded deoxynucleic acid or ribonucleic acid)-biotin complex.

19. The computer program product of claim 1, wherein the label identification module further comprises instructions for identifying a plurality of candidate labels in the plurality of light images, and wherein said plurality of labels is a subset of the plurality of candidate labels that have been validated by the label identification module.

20. The computer program product of claim 19, wherein each candidate label in the plurality of candidate labels is associated with a position on the substrate that emits more than a threshold amount of light in any one light image in the plurality of light images.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 823

21. The computer program product of claim 19, wherein the plurality of labels comprises a first candidate label that is associated with a first position on the substrate that emits light in a first wavelength range in the plurality of different wavelength ranges and a second candidate label that is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges.

22. The computer program product of claim 21, wherein a portion of the first wavelength range overlaps with a portion of the second wavelength range.

23. The computer program product of claim 21, wherein the first wavelength range does not overlap with the second wavelength range.

24. The computer program product of claim 19, wherein the instructions for identifying the plurality of labels applies a first distance criterion between a centroid of a first candidate label and a centroid of a second candidate label in the plurality of candidate labels.

25. The computer program product of claim 24, wherein the first distance criterion is determined by a calculated distance between a first label and a second label in the probe.

26. The computer program product of claim 24, wherein the instructions for identifying the plurality of labels applies a second distance criterion between a centroid of the second candidate label and a centroid of a third candidate label in the plurality of candidate labels.

27. The computer program product of claim 26, wherein the second distance criterion is determined by a calculated distance between a second label and a third label in the probe.

28. The computer program product of claim 26, wherein the first distance criterion is the same as the second distance criterion.

29. The computer program product of claim 26, wherein the first distance criterion is different from the second distance criterion.

30. The computer program product of claim 26, wherein a value of the first distance criterion and a value of the second distance criterion contribute to determining whether the plurality of labels is said probe.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 824

31. The computer program product of claim 26, wherein the computer program mechanism further comprises a lookup table comprising a plurality of valid reporter sequences, wherein each valid reporter sequence in the plurality of valid reporter sequences comprises a first distance between a first pair of labels and a second distance between a second pair of labels, and wherein the probe identification module further comprises instructions for comparing the string sequence of the plurality of labels, the first distance criterion, and the second distance criterion, to valid reporter sequences in the lookup table.

32. The computer program product of claim 19, wherein the instructions for identifying the plurality of labels applies an angle criterion to triplets of candidate labels in the plurality of candidate labels.

33. The computer program product of claim 19, wherein the instructions for identifying the plurality of labels comprises instructions for applying a model to select candidate labels in the plurality of candidate labels.

34. The computer program product of claim 33, wherein the instructions for applying a model to select candidate labels comprises instructions for applying linear regression to the select candidate labels.

35. The computer program product of claim 19, wherein the label identification module further comprises instructions for verifying that a candidate label in the plurality of candidate labels satisfies a spot shape criterion.

36. The computer program product of claim 35, wherein the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for performing point spread function modeling on the candidate label.

37. The computer program product of claim 35, wherein the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for applying a spot segmentation algorithm to the candidate label.

38. The computer program product of claim 37, wherein the spot segmentation algorithm comprises a watershed transformation.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 825

39. The computer program product of claim 19, wherein the instructions for identifying the plurality of labels applies an absolute distance criterion between a centroid of a first terminal candidate label and a centroid of a second terminal candidate label in the plurality of candidate labels.

40. The computer program product of claim 19, wherein the instructions for identifying the plurality of labels comprises instructions for identifying a buffer zone around a portion of the substrate that has select candidate labels in the plurality of candidate labels, wherein there are no candidate labels in the buffer zone.

41. The computer program product of claim 1, wherein the plurality of labels are linearly arranged on the substrate.

42. The computer program product of claim 1, wherein the plurality of labels are linearly arranged on the substrate in the same linear orientation.

43. The computer program product of claim 42, wherein a linear orientation of each label in the plurality of labels is predetermined.

44. The computer program product of claim 43, wherein a linear orientation of each label in the plurality of labels is determined by an application of an electrical current across said substrate.

45. The computer program product of claim 43, wherein a linear orientation of each label in the plurality of labels is determined by an application of a fluid across said substrate.

46. The computer program product of claim 1, wherein each respective label in the plurality of labels occupies between 4 and 20 pixels in a pixilated representation of the respective label in a light image in the plurality of light images.

47. The computer program product of claim 1, wherein each respective label in the plurality of labels occupies between 6 and 30 pixels in a pixilated representation of the respective label in a light image in the plurality of light images.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 826

48. The computer program product of claim 1, wherein each respective label in the plurality of labels occupies between 1 and 30 pixels in a pixilated representation of the respective label in a light image in the plurality of light images.

49. The computer program product of claim 1, wherein each respective label in the plurality of labels occupies between 4 and 100 pixels in a pixilated representation of the respective label in a light image in the plurality of light images.

50. The computer program product of claim 1, wherein the label identification module further comprises:

instructions for identifying a first candidate label in the plurality of light images; and

instructions for identifying a second candidate label in the plurality of light images that is within a predetermined distance away from the first candidate label;

wherein the plurality of labels comprises the first candidate label and the second candidate label.

51. The computer program product of claim 1, wherein the plurality of different wavelength ranges consists of between two different wavelength ranges and six different wavelength ranges.

52. The computer program product of claim 1, wherein the plurality of different wavelength ranges consists of between two different wavelength ranges and twenty different wavelength ranges.

53. The computer program product of claim 1, wherein the plurality of labels comprises four labels.

54. The computer program product of claim 1, wherein the plurality of labels comprises five labels.

55. The computer program product of claim 1, wherein the plurality of labels consists of between two labels and twenty labels.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 827

56. The computer program product of claim 1, wherein a first subset of the labels in the string sequence error check an identity of the labels in a second subset of the labels in the string sequence.

57. The computer program product of claim 1, wherein a first subset of labels in the string sequence are a checksum for a second subset of labels in the string sequence.

58. The computer program product of claim 1, wherein

the label identification module comprises instructions for repeating said instructions for identifying a plurality of labels a plurality of times, wherein each time said instructions for identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images, that are proximate to each other on the substrate; and

wherein the probe identification module determines whether each said different plurality of labels identified by the label identification module comprises a valid reporter sequence wherein, for each said different plurality of labels, said probe identification module:

deems the different plurality of labels to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence; and

deems the different plurality of labels to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence.

59. The computer program product of claim 58, wherein a plurality of probes is identified.

60. The computer program product of claim 59, wherein said plurality of probes consists of three or more probes.

61. The computer program product of claim 59, wherein said plurality of probes consists of ten or more probes.

62. The computer program product of claim 59, wherein said plurality of probes consists of less than 50 probes.

63. The computer program product of claim 58, wherein said probe identification module stores each type of probe identified.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 828

64. The computer program product of claim 58, wherein said probe identification model stores each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence.

65. The computer program product of claim 58, wherein said probe identification model stores each string sequence of each different plurality of labels that is confirmed as a valid reporter sequence.

66. A computer system for detecting the presence of a probe within a sample overlayed on a substrate, wherein the probe comprises a plurality of spatially arranged labels, the computer system comprising:

a central processing unit; and

a memory, coupled to the central processing unit, the memory storing:

a data storage module comprising instructions for storing a plurality of light images, each light image in the plurality of light images being for light received from the sample at a wavelength range in a plurality of different wavelength ranges;

a label identification module comprising instructions for identifying a plurality of labels, in the plurality of light images, that are proximate to each other on the substrate, wherein a spatial order of the plurality of labels determines a string sequence of the plurality of labels; and

a probe identification module comprising instructions for determining whether the string sequence of the plurality of labels comprises a valid reporter sequence, wherein

when the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe; and

when the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

67. The computer system of claim 66, wherein a first label in the plurality of labels is associated with a first position on the substrate that emits light in a first wavelength range in the plurality of different wavelength ranges and a second label in the plurality of labels is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges.

68. The computer system of claim 67, wherein a portion of the first wavelength range overlaps with a portion of the second wavelength range.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 829

69. The computer system of claim 67, wherein the first wavelength range does not overlap with the second wavelength range.

70. The computer system of claim 66, wherein each label in the plurality of labels is associated with a position on the substrate that emits more than a threshold amount of light in any one light image in the plurality of light images.

71. The computer system of claim 66, wherein the memory further stores a lookup table comprising a plurality of valid reporter sequences, and wherein the probe identification module further comprises instructions for comparing the string sequence of the plurality of labels to valid reporter sequences in the lookup table.

72. The computer system of claim 71, wherein the lookup table is dimensioned and configured to hold up to 4^4 different valid reporter sequences.

73. The computer system of claim 71, wherein the lookup table is dimensioned and configured to hold up to 20^{20} different valid reporter sequences.

74. The computer system of claim 66, the probe identification module further comprising instructions for storing the string sequence of the plurality of labels that is not confirmed as a valid reporter sequence.

75. The computer system of claim 66, wherein the data storage module further comprises instructions for aligning a first light image to a second light image in the plurality of light images using a plurality of fiducials present on the substrate.

76. The computer system of claim 66, wherein the position of the probe on the substrate is random.

77. The computer system of claim 66, wherein the probe consists of a single molecule.

78. The computer system of claim 66, wherein the probe comprises a molecular scaffold and wherein each label in the plurality of labels represents a different position on said molecular scaffold.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 830

79. The computer system of claim 78, wherein each position on said molecular scaffold that is represented by a label is separated from a neighboring position on the scaffold by a spacer.

80. The computer system of claim 66, wherein the probe comprises a single stranded deoxynucleic acid or ribonucleic acid scaffold and wherein each label in the plurality of labels is represented by a dye laden single stranded deoxynucleic acid or ribonucleic acid sequence that hybridizes to a different position on the scaffold.

81. The computer system of claim 66, wherein

the probe comprises a molecular scaffold having a first end and a second end;

a target specific sequence is covalently attached to the first end;

a binder sequence is covalently attached to the second end; and

the probe is linearly arranged on the substrate through (i) the binding of the target specific sequence to a first molecular entity that is bound to a first position of the substrate and (ii) the binding of the binder sequence to a second molecular entity that is bound to a second position on the substrate.

82. The computer system of claim 81, wherein

the first molecular entity is a target (single stranded nucleic acid or ribonucleic acid)-biotin complex, and

the second molecular entity is a predetermined (single stranded deoxynucleic acid or ribonucleic acid)-biotin complex.

83. The computer system of claim 66, wherein the label identification module further comprises instructions for identifying a plurality of candidate labels in the plurality of light images, and wherein said plurality of labels is a subset of the plurality of candidate labels that have been validated by the label identification module.

84. The computer system of claim 83, wherein each candidate label in the plurality of candidate labels is associated with a position on the substrate that emits more than a threshold amount of light in any one light image in the plurality of light images.

85. The computer system of claim 83, wherein the plurality of labels comprises a first candidate label is associated with a first position on the substrate that emits light in a first

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 831

wavelength range in the plurality of different wavelength ranges and a second candidate label is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges.

86. The computer system of claim 85, wherein a portion of the first wavelength range overlaps with a portion of the second wavelength range.

87. The computer system of claim 85, wherein the first wavelength range does not overlap with the second wavelength range.

88. The computer system of claim 83, wherein the instructions for identifying the plurality of labels applies a first distance criterion between a centroid of a first candidate label and a centroid of a second candidate label in the plurality of candidate labels.

89. The computer system of claim 88, wherein the first distance criterion is determined by a calculated distance between a first label and a second label in the probe.

90. The computer system of claim 88, wherein the instructions for identifying the plurality of labels applies a second distance criterion between a centroid of the second candidate label and a centroid of a third candidate label in the plurality of candidate labels.

91. The computer system of claim 90, wherein the second distance criterion is determined by a calculated distance between a second label and a third label in the probe.

92. The computer system of claim 90, wherein a value of the first distance criterion and a value of the second distance criterion contribute to determining whether the plurality of labels is said probe.

93. The computer system of claim 90, wherein the computer program mechanism further comprises a lookup table comprising a plurality of valid reporter sequences, wherein each valid reporter sequence in the plurality of valid reporter sequences comprises a first distance between a first pair of labels and a second distance between a second pair of labels, and wherein the probe identification module further comprises instructions for comparing the string sequence of the plurality of labels, the first distance criterion, and the second distance criterion, to valid reporter sequences in the lookup table.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 832

94. The computer system of claim 83, wherein the instructions for identifying the plurality of labels applies an angle criterion to triplets of candidate labels in the plurality of candidate labels.

95. The computer system of claim 83, wherein the instructions for identifying the plurality of labels comprises instructions for applying a model to select candidate labels in the plurality of candidate labels.

96. The computer system of claim 95, wherein the instructions for applying a model to select candidate labels comprises instructions for applying linear regression to the select candidate labels.

97. The computer system of claim 95, wherein the label identification module further comprises instructions for verifying that a candidate label in the plurality of candidate labels satisfies a spot shape criterion.

98. The computer system of claim 97, wherein said spot shape criterion is a match between an observed spot shape of the candidate label and the theoretical point spread of the diffraction limited point source light determined by a magnification of the candidate label.

99. The computer system of claim 97, wherein the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for performing point spread function modeling on the candidate label.

100. The computer system of claim 97, wherein the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for applying a spot segmentation algorithm to the candidate label.

101. The computer system of claim 100, wherein the spot segmentation algorithm comprises a watershed transformation.

102. The computer system of claim 83, wherein the instructions for identifying the plurality of labels applies an absolute distance criterion between a centroid of a first terminal candidate label and a centroid of a second terminal candidate label in the plurality of candidate labels.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 833

103. The computer system of claim 83, wherein the instructions for identifying the plurality of labels comprises instructions for identifying a buffer zone around a portion of the substrate that has select candidate labels in the plurality of candidate labels, wherein there are no candidate labels in the buffer zone.

104. The computer system of claim 66, wherein the plurality of labels are linearly arranged on the substrate.

105. The computer system of claim 66, wherein the plurality of labels are linearly arranged on the substrate in the same linear orientation.

106. The computer system of claim 66, wherein each respective label in the plurality of labels occupies between 4 and 20 pixels in a pixilated representation of the respective label in a light image in the plurality of light images.

107. The computer system of claim 66, wherein the label identification module further comprises:

instructions for identifying a first candidate label in the plurality of light images; and

instructions for identifying a second candidate label in the plurality of light images that is within a predetermined distance away from the first candidate label;

wherein the plurality of labels comprises the first candidate label and the second candidate label.

108. The computer system of claim 66, wherein the plurality of different wavelength ranges consists of between two different wavelength ranges and twenty different wavelength ranges.

109. The computer system of claim 66, wherein the plurality of labels consists of between two labels and twenty labels.

110. The computer system of claim 66, wherein a first subset of the labels in the string sequence error check an identity of the labels in a second subset of the labels in the string sequence.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 834

111. The computer system of claim 66, wherein a first subset of labels in the string sequence are a checksum for a second subset of labels in the string sequence.

112. The computer system of claim 66, wherein

the label identification module comprises instructions for repeating said instructions for identifying a plurality of labels a plurality of times, wherein each time said instructions for identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images, that are proximate to each other on the substrate; and

wherein the probe identification module determines whether each said different plurality of labels identified by the label identification module comprises a valid reporter sequence, wherein for each said different plurality of labels, said probe identification module:

deems the different plurality of labels to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence; and

deems the different plurality of labels to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence.

113. The computer system of claim 112, wherein a plurality of probes is identified.

114. The computer system of claim 113 wherein said plurality of probes consists of three or more probes.

115. The computer system of claim 112, wherein said probe identification model stores each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence.

116. The computer system of claim 112, wherein said probe identification model stores each string sequence of each different plurality of labels that is confirmed as a valid reporter sequence.

117. A system for detecting the presence of a probe within a sample overlayed on a substrate, the system comprising:

a light measuring mechanism that measures a plurality of light images, each light image in the plurality of light images being for light received from the sample at a wavelength range in a plurality of different wavelength ranges;

a data storage module comprising instructions for storing said plurality of light images;

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 835

a label identification mechanism that identifies a plurality of labels in the plurality of light images that are proximate to each other on the substrate, wherein a spatial order of the plurality of labels determines a string sequence of the plurality of labels; and

a probe identification mechanism that determines whether the string sequence of the plurality of labels comprises a valid reporter sequence, wherein

when the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe; and

when the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

118. The system of claim 117, wherein a first label in the plurality of labels is associated with a first position on the substrate that emits light in a first wavelength range in the plurality of different wavelength ranges and a second label in the plurality of labels is associated with a second position on the substrate that emits light in a second wavelength range in the plurality of different wavelength ranges.

119. The system of claim 118, wherein a portion of the first wavelength range overlaps with a portion of the second wavelength range.

120. The system of claim 117, wherein the system further comprises an illumination mechanism that illuminates the substrate.

121. The system of claim 120, wherein the illumination mechanism comprises an excitation light source and a plurality of excitation filters, where each excitation filter in the plurality of excitation filters is used in a corresponding light image in the plurality of light images to confine the light source to a corresponding different spectral range when the corresponding light image is measured.

122. The system of claim 117, wherein the light measuring mechanism comprises a plurality of measurement wavelength filters, wherein each measurement wavelength filter in the plurality of measurement wavelength filters is used in a corresponding light image in the plurality of light images to reject light not within a corresponding spectral range.

123. The system of claim 117, wherein the light measuring mechanism comprises a photodetector that forms a detection signal in response to light emitted from the sample.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 836

124. The system of claim 117, wherein the light measuring mechanism comprises a detector circuit addressed by the detection signal that measures light emitted from the sample overlayed on the substrate, the light measuring mechanism further comprising an electronic memory for storing a plurality of label positions, wherein each label position in the plurality of label positions represents a label and each label position in the plurality of label positions originates more than a threshold amount of light.

125. The system of claim 117, wherein the label identification mechanism identifies the plurality of labels that are proximate to each other from among the plurality of label positions stored in electronic memory.

126. The system of claim 117, wherein the label identification mechanism comprises instructions for identifying a plurality of candidate labels in the plurality of light images, and wherein the plurality of labels is a subset of the plurality of candidate labels.

127. The system of claim 126, wherein each candidate label in the plurality of candidate labels comprises a position on the substrate that emits more than a threshold amount of light in any one light image in the plurality of light images.

128. The system of claim 126, wherein the plurality of labels comprises a first candidate label that emits light in a first wavelength range in the plurality of different wavelength ranges and a second candidate label in the plurality of labels that emits light in a second wavelength range in the plurality of different wavelength ranges.

129. The system of claim 126, wherein the instructions for identifying the plurality of labels applies a first distance criterion between a centroid of a first candidate label and a centroid of a second candidate label in the plurality of candidate labels.

130. The system of claim 129, wherein the first distance criterion is determined by a calculated distance between a first label and a second label in the probe.

131. The system of claim 130, wherein the instructions for identifying the plurality of labels applies a second distance criterion between a centroid of the second candidate label and a centroid of a third candidate label in the plurality of candidate labels.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 837

132. The system of claim 131, wherein the second distance criterion is determined by a calculated distance between a second label and a third label in the probe.

133. The system of claim 131, wherein the first distance criterion is the same as the second distance criterion.

134. The system of claim 131, wherein the first distance criterion is different from the second distance criterion.

135. The system of claim 131, wherein a value of the first distance criterion and a value of the second distance criterion contribute to determining whether the plurality of labels is said probe.

136. The system of claim 126, wherein the instructions for identifying the plurality of labels applies an angle criterion to triplets of candidate labels in the plurality of candidate labels.

137. The system of claim 126, wherein the instructions for identifying the plurality of labels comprises instructions for applying a model to select candidate labels in the plurality of candidate labels.

138. The system of claim 137, wherein the instructions for applying a model to select candidate labels comprises instructions for applying linear regression to the select candidate labels.

139. The system of claim 126, wherein the label identification module further comprises instructions for verifying that a candidate label in the plurality of candidate labels satisfies a spot shape criterion.

140. The system of claim 139, wherein said spot shape criterion is a match between an observed spot shape of the candidate label and the theoretical point spread of the diffraction limited point source light determined by a magnification of the candidate label.

141. The system of claim 139, wherein the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for performing point spread function modeling on the candidate label.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 838

142. The system of claim 139, wherein the instructions for verifying that a candidate label in the plurality of candidate labels satisfies the spot shape criterion comprises instructions for applying a spot segmentation algorithm to the candidate label.

143. The system of claim 142, wherein the spot segmentation algorithm comprises a watershed transformation.

144. The system of claim 117, wherein the plurality of labels are linearly arranged on the substrate.

145. The system of claim 144, wherein a linear orientation of the plurality of labels is predetermined.

146. The system of claim 144, wherein a linear orientation of the plurality of labels is determined by an application of an electrical current across said substrate.

147. The system of claim 144, wherein the probe within the sample is overlayed on the substrate at a random position on the substrate.

148. A method for detecting the presence of a probe within a sample overlayed on a substrate, wherein the probe comprises a plurality of spatially arranged labels, the method comprising:

identifying a plurality of labels, in a plurality of light images, that are proximate to each other on the substrate, wherein a spatial order of the plurality of labels determines a string sequence of the plurality of labels, wherein each light image in the plurality of light images is for light received from the sample at a wavelength range in a plurality of different wavelength ranges; and

determining whether the string sequence of the plurality of labels comprises a valid reporter sequence, wherein

when the string sequence of the plurality of labels is confirmed as a valid reporter sequence, the plurality of labels is deemed to be the probe; and

when the string sequence of the plurality of labels is not confirmed as a valid reporter sequence, the plurality of labels is deemed to not be the probe.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 839

149. The method of claim 148, where the determining step comprises comparing the string sequence of the plurality of labels to valid reporter sequences in a lookup table.

150. The method of claim 148, the method further comprising storing the string sequence of the plurality of labels that is not confirmed as a valid reporter sequence.

151. The method of claim 148, the method further comprising aligning a first light image to a second light image in the plurality of light images using a plurality of fiducials present on the substrate.

152. The method of claim 148, wherein

the step of identifying a plurality of labels is repeated a plurality of times, wherein each time said step of identifying a plurality of labels is repeated, a different plurality of labels is identified, in the plurality of light images, that are proximate to each other on the substrate; the method further comprising

determining whether each said different plurality of labels comprises a valid reporter sequence, wherein for each said different plurality of labels

the different plurality of labels is deemed to be a probe when the string sequence of the different plurality of labels is confirmed as a valid reporter sequence; and

the different plurality of labels is deemed to not be a probe when the string sequence of the different plurality of labels is not confirmed as a valid reporter sequence.

153. The method of claim 152, wherein a plurality of probes is identified.

154. The method of claim 152, wherein said plurality of probes consists of three or more probes.

155. The method of claim 152, wherein said plurality of probes consists of ten or more probes.

156. The method of claim 152, the method further comprising storing each type of probe identified.

157. The method of claim 152, wherein each string sequence of each different plurality of labels that is not confirmed as a valid reporter sequence is stored.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 840

158. The method of claim 152, wherein each string sequence of each different plurality of labels that is confirmed as a valid reporter sequence is stored.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 841

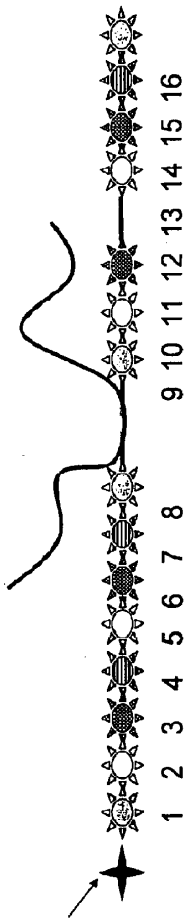


Figure 1A

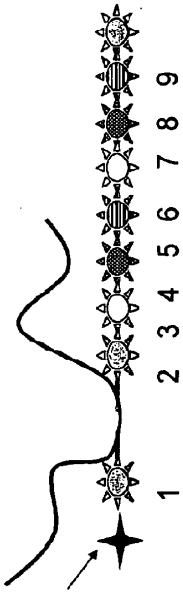


Figure 1B

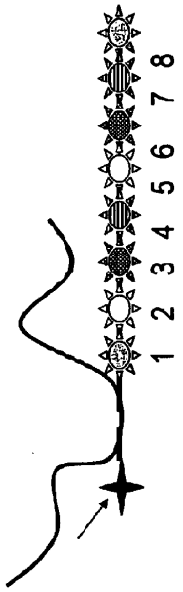


Figure 1C



Figure 1D



Figure 1E

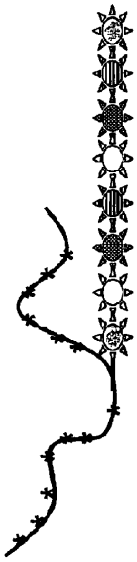


Figure 1F

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 842

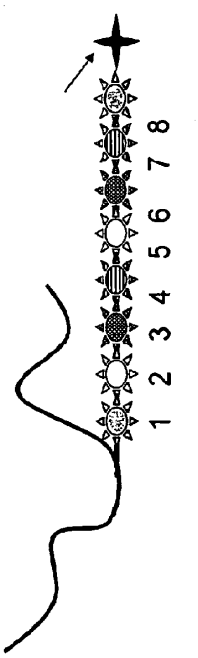


Figure 1D

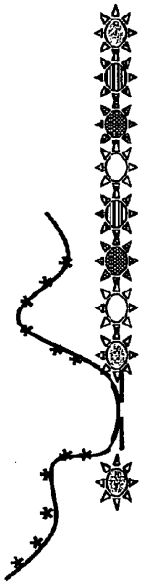


Figure 1E

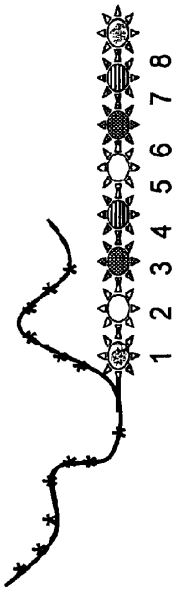
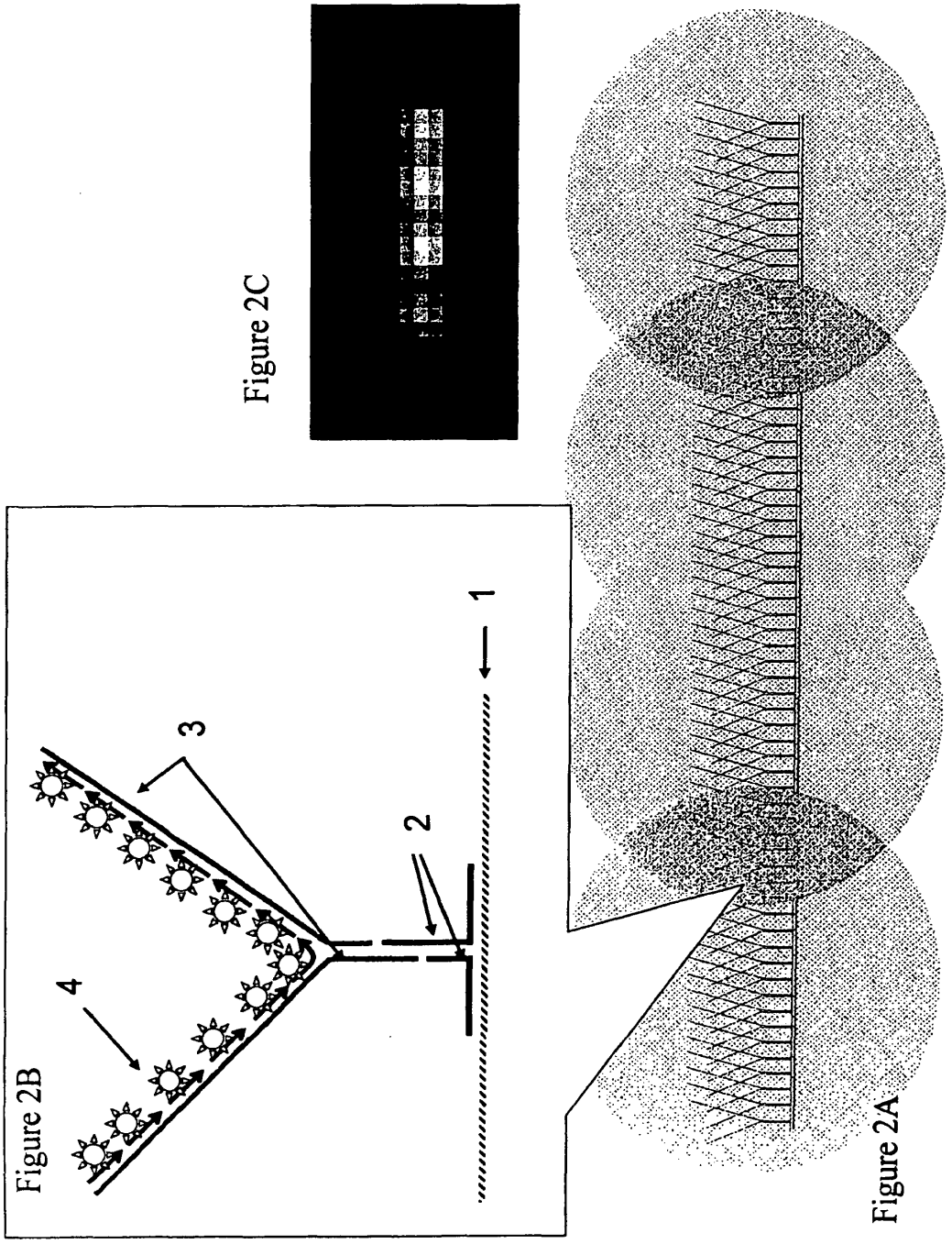


Figure 1F

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 843



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 844

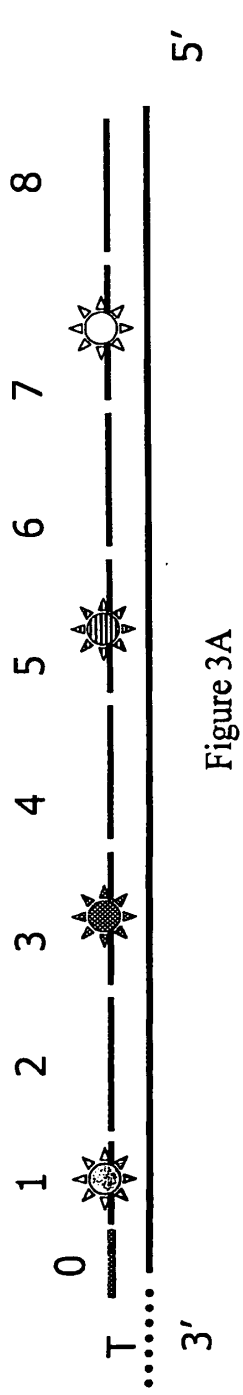


Figure 3A

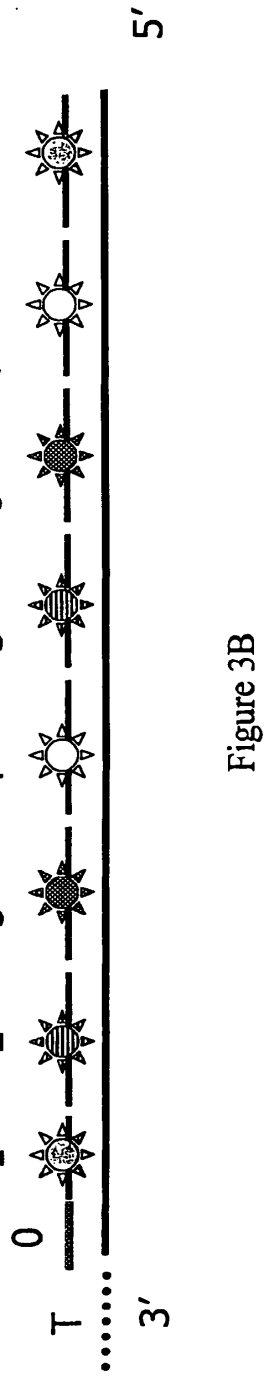


Figure 3B

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 845

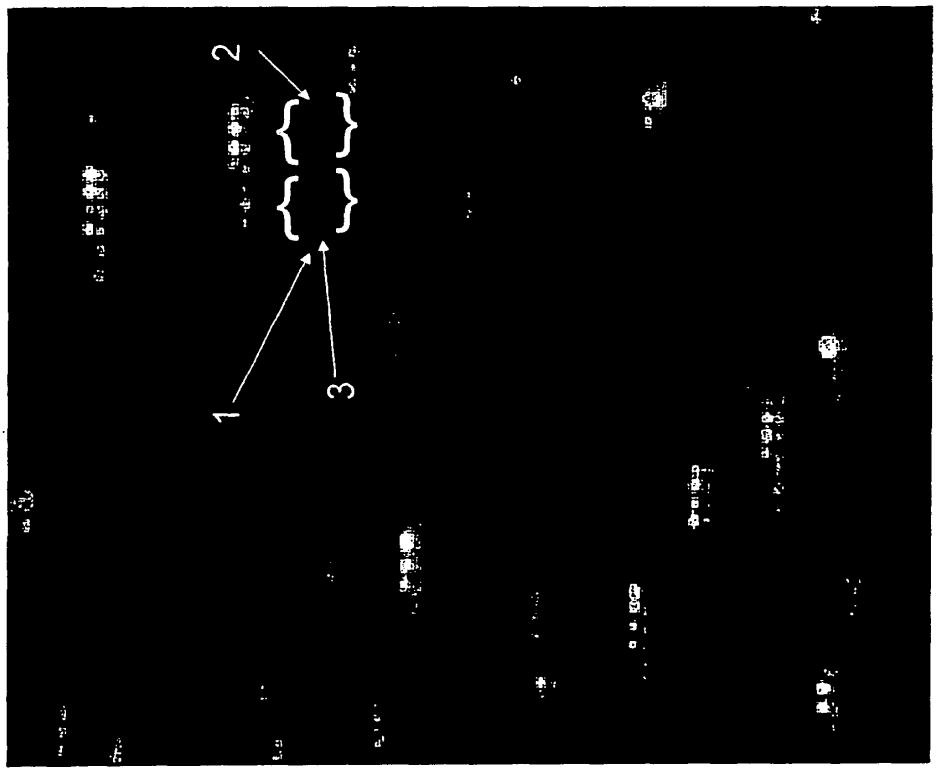
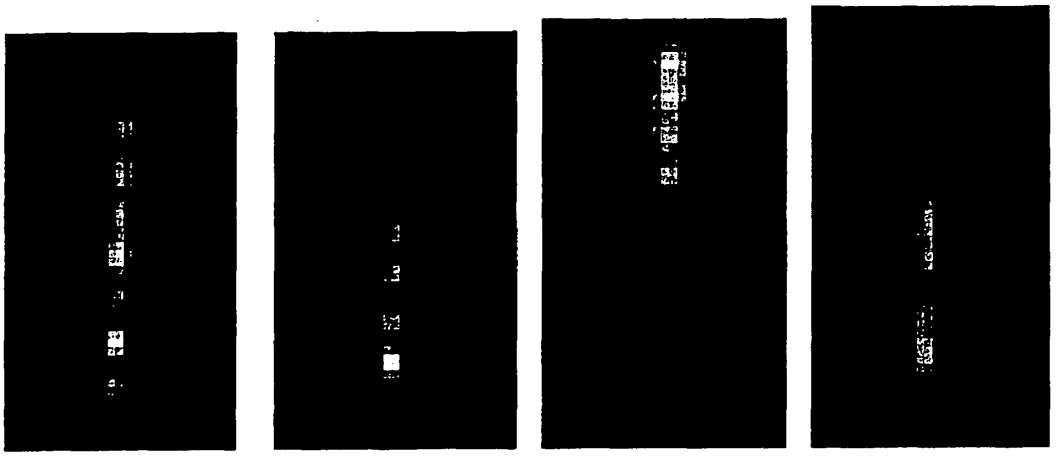


Figure 4

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 846



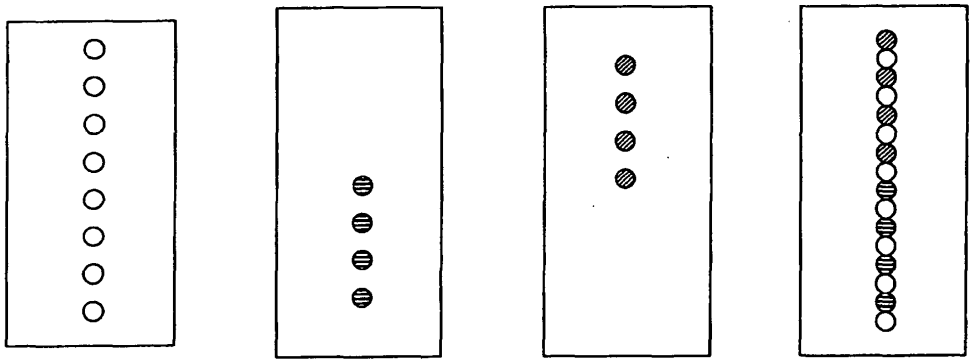


Figure 5A

Figure 5B

Figure 5C

Figure 5D

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 847

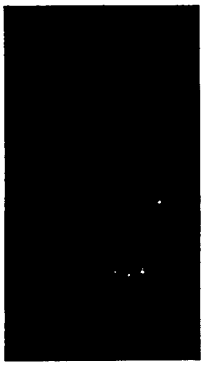


Figure 6E

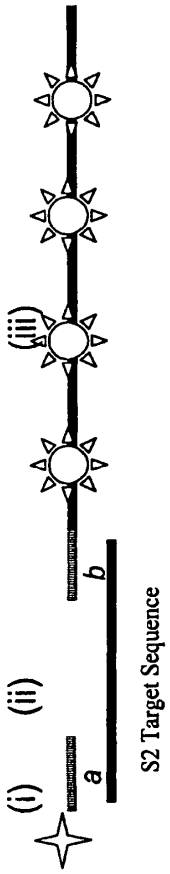


Figure 6A

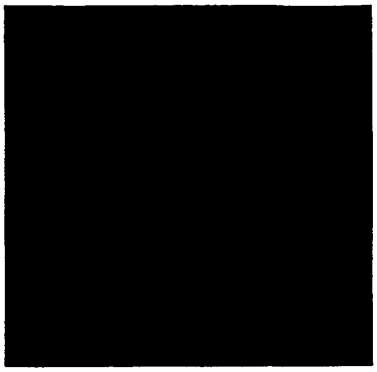


Figure 6D

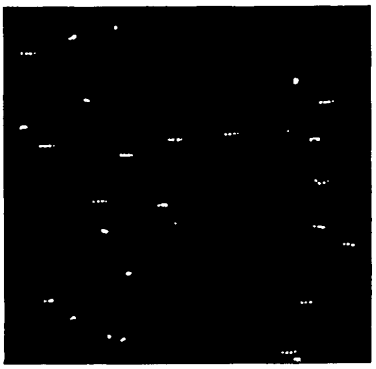


Figure 6C

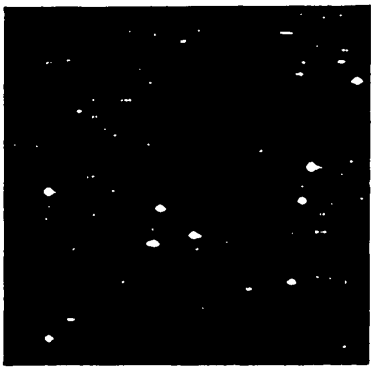
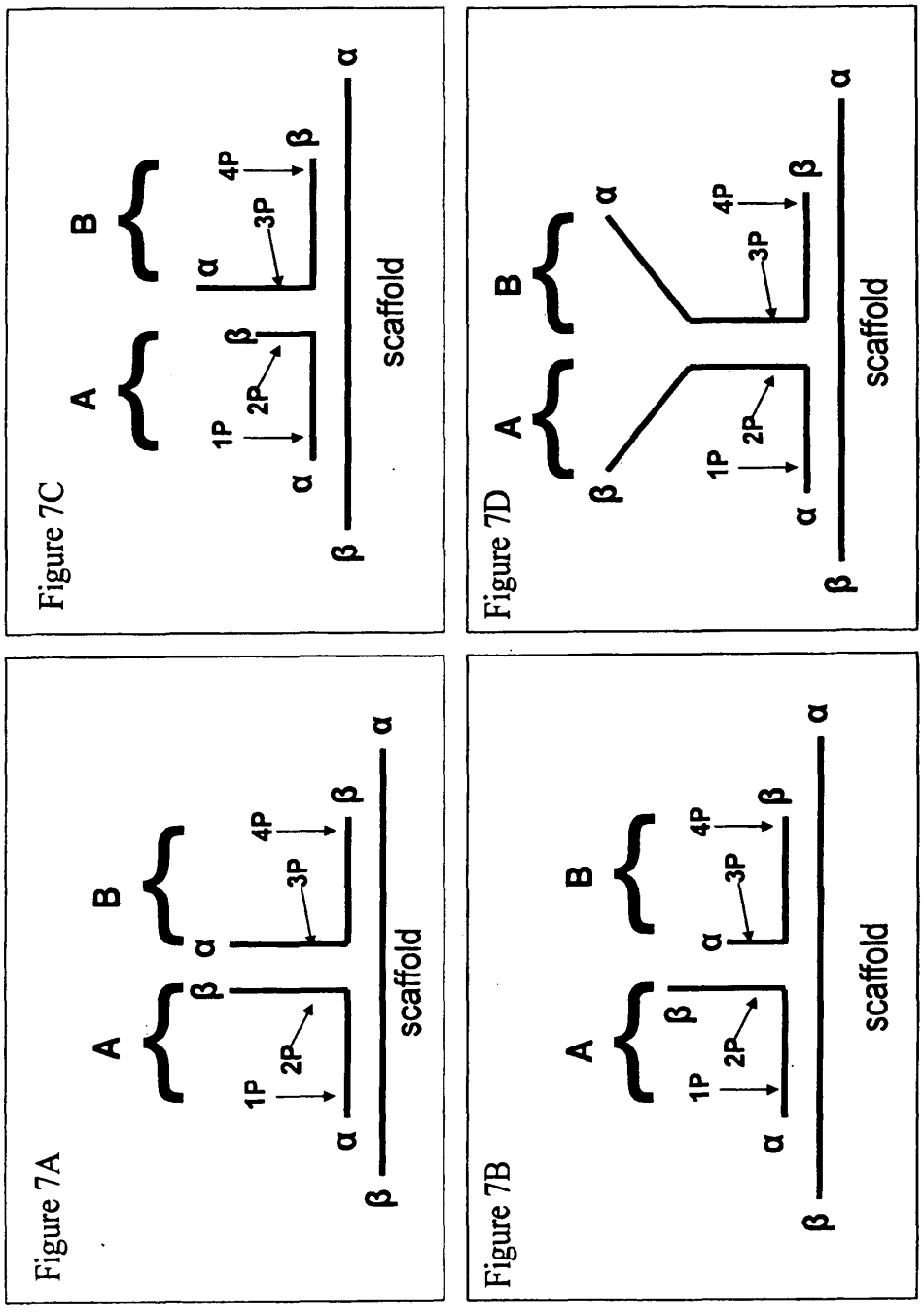
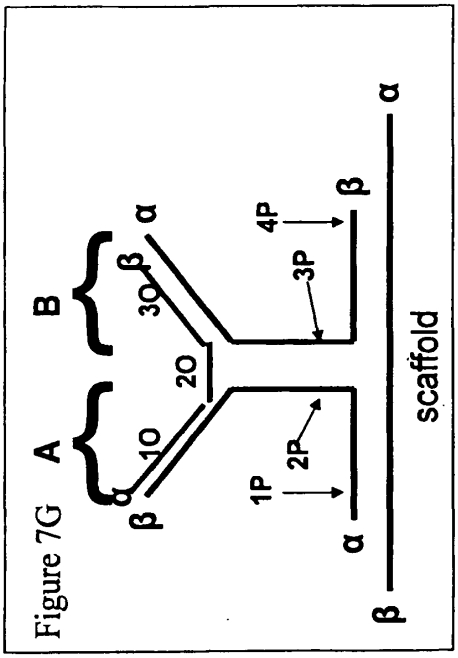


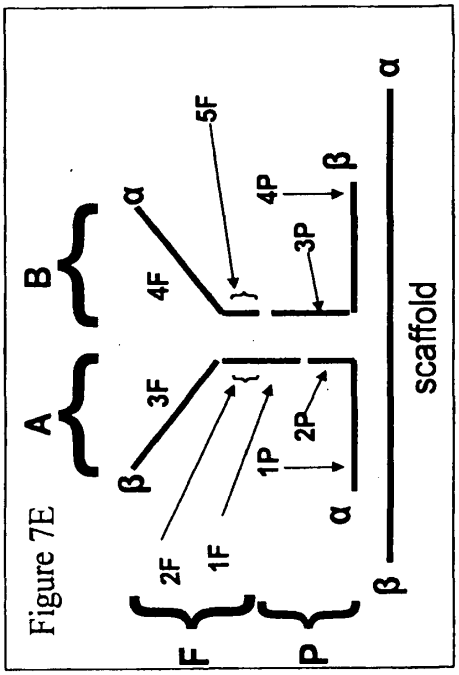
Figure 6B

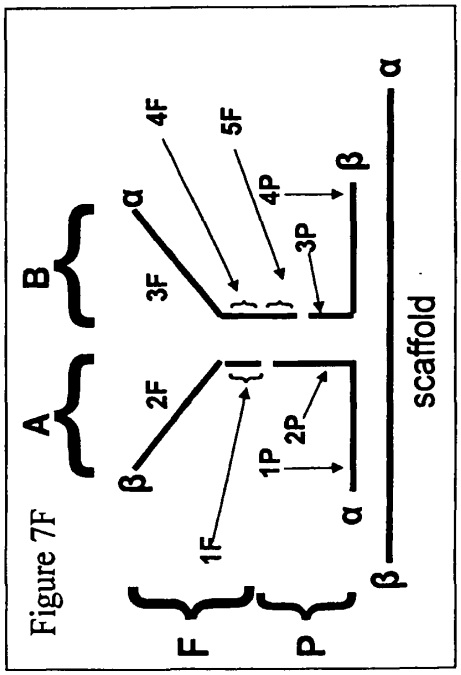
IPR2024-01177
California Institute of Technology - Ex. 2003 Page 848



IPR2024-01177
California Institute of Technology - Ex. 2003 Page 849







IPR2024-01177
California Institute of Technology - Ex. 2003 Page 850

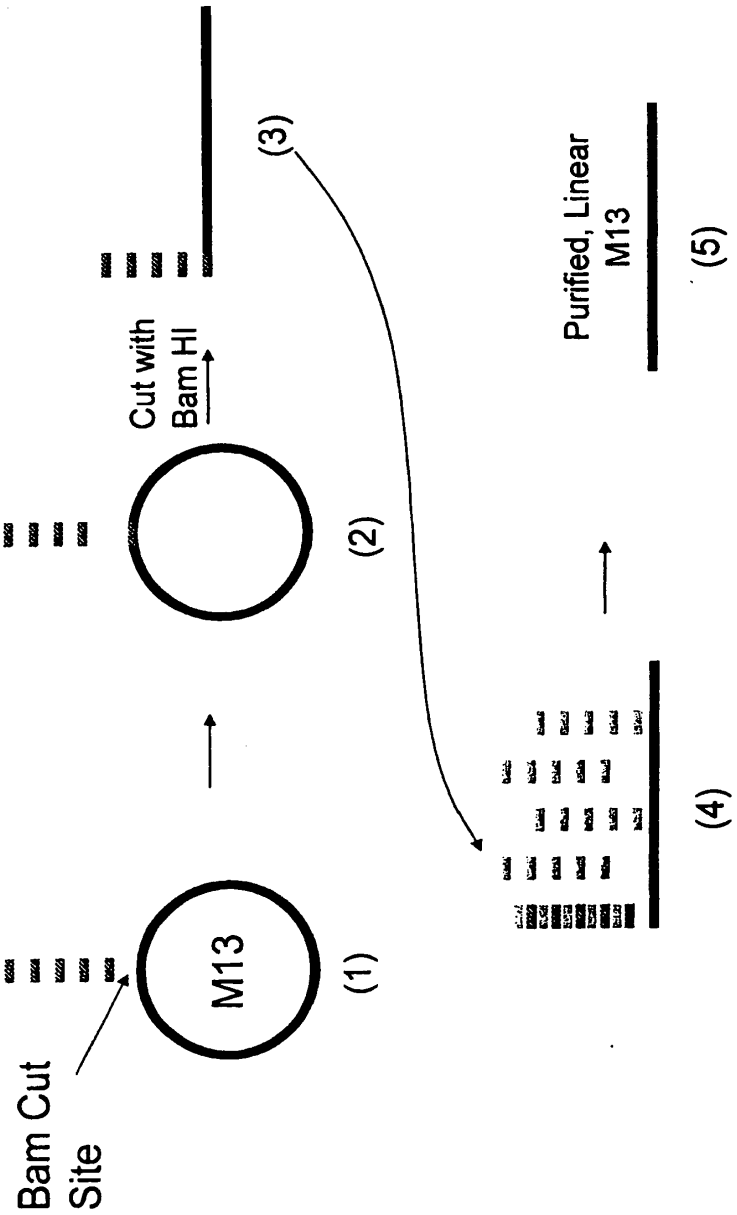


Figure 8

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 851

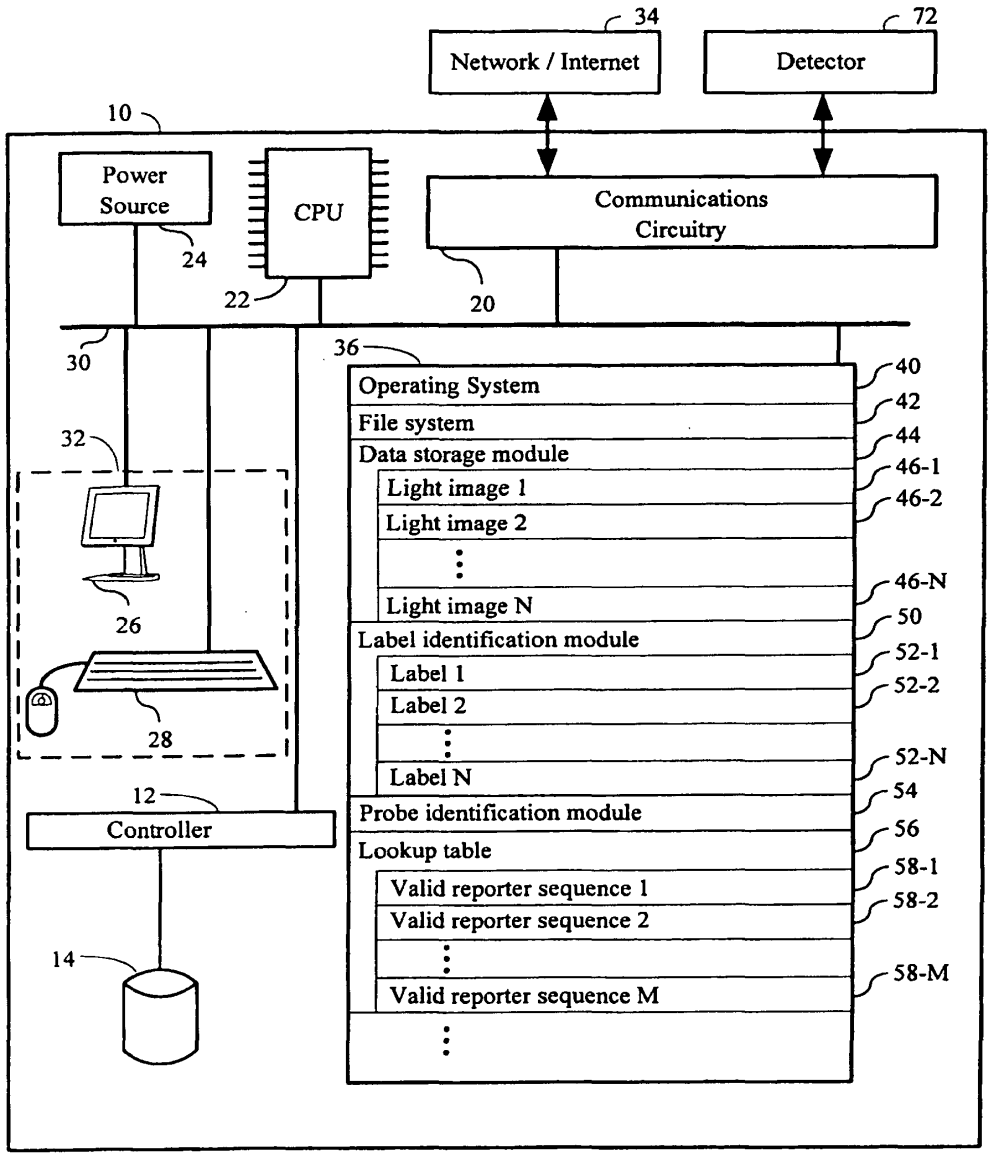


Figure 9

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 852

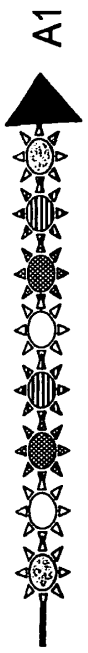


Figure 10A



Figure 10B

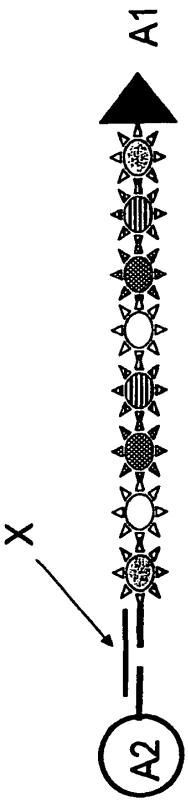


Figure 10C

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 853

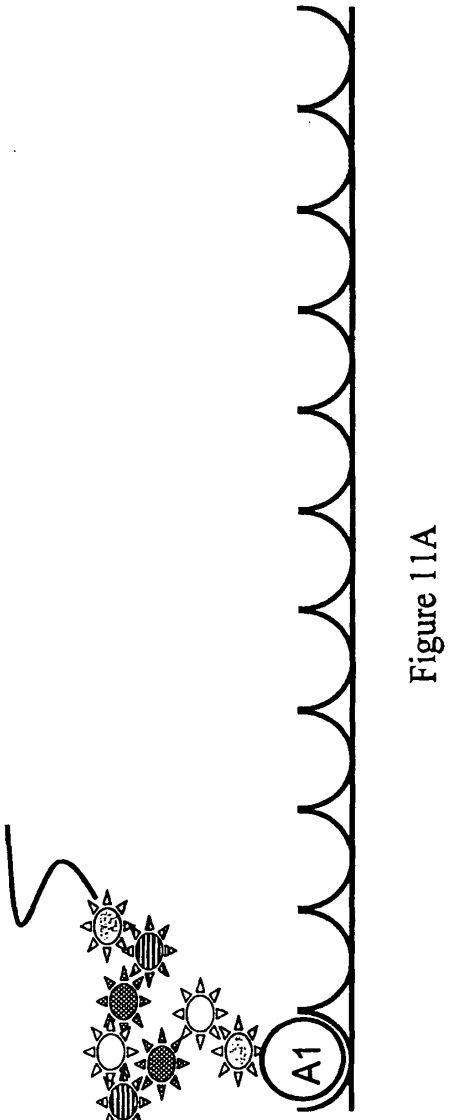


Figure 11A

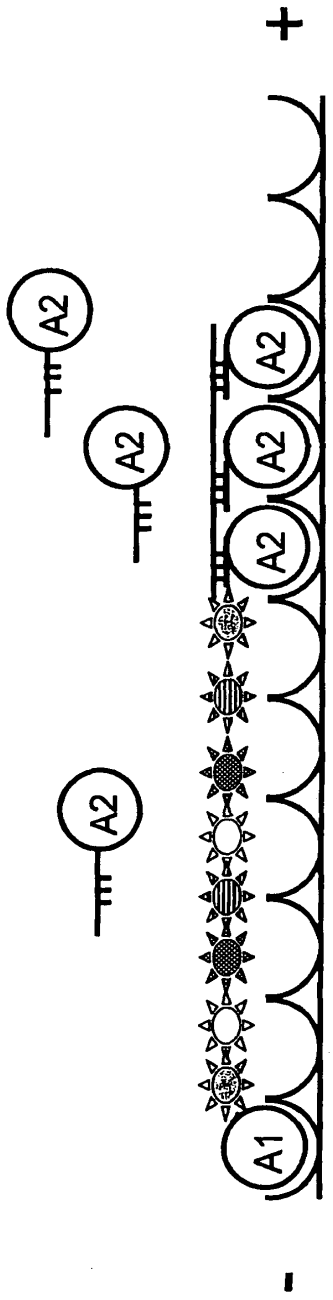


Figure 11B

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 854

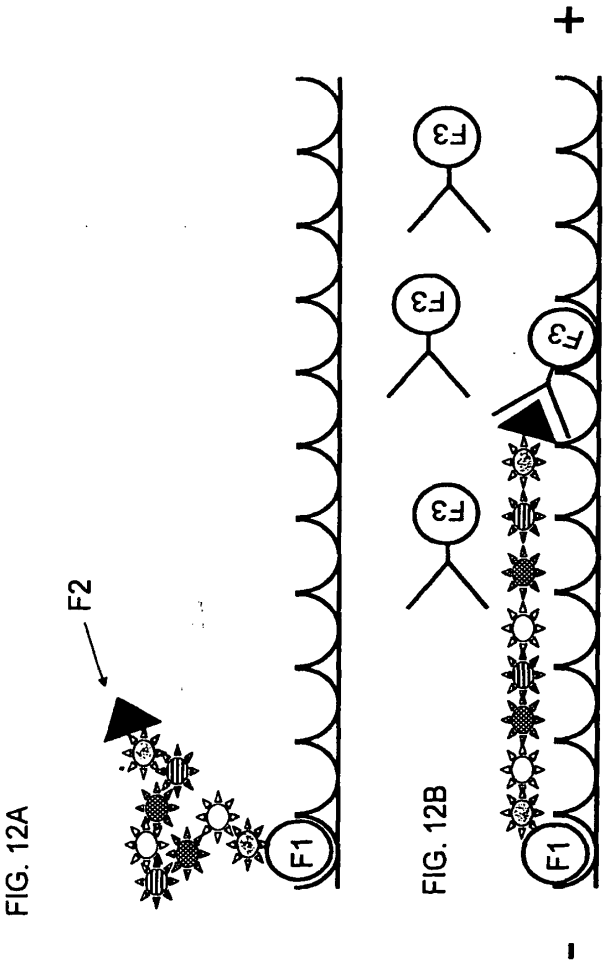


FIG. 12A

FIG. 12B

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 855

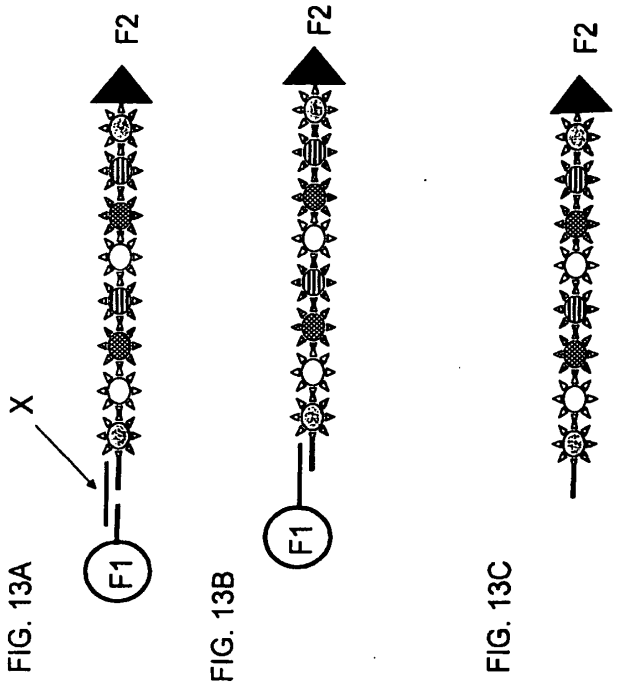


FIG. 13A FIG. 13B FIG. 13C

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 856

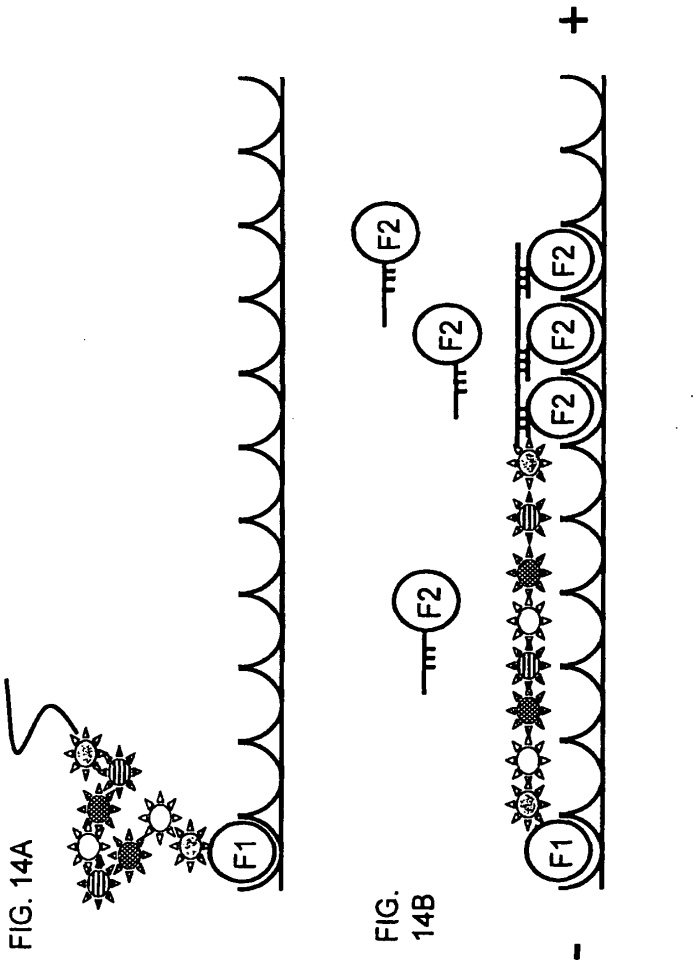


FIG. 14A

FIG. 14B

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 857

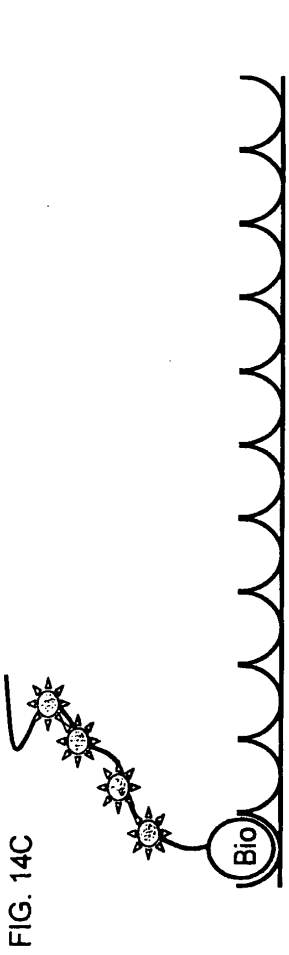


FIG. 14C

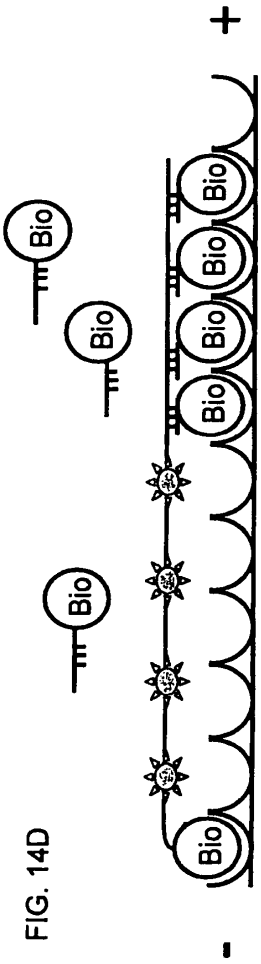


FIG. 14D

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 858

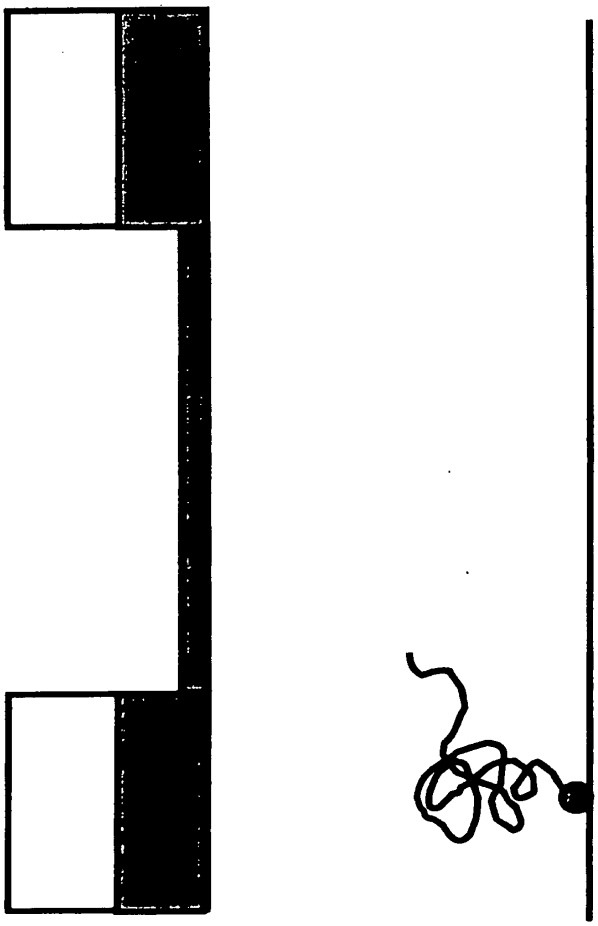


Figure 15A

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 859

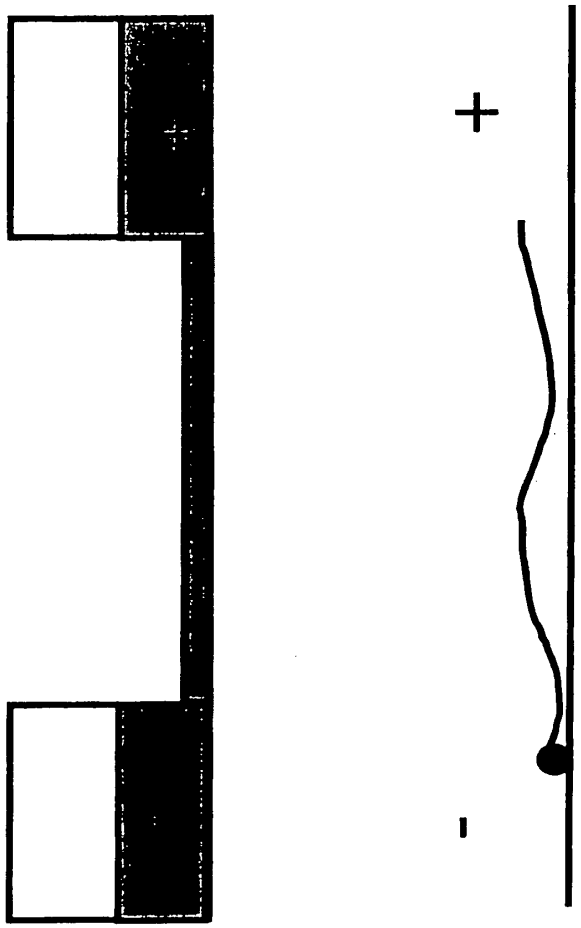


Figure 15B

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 860

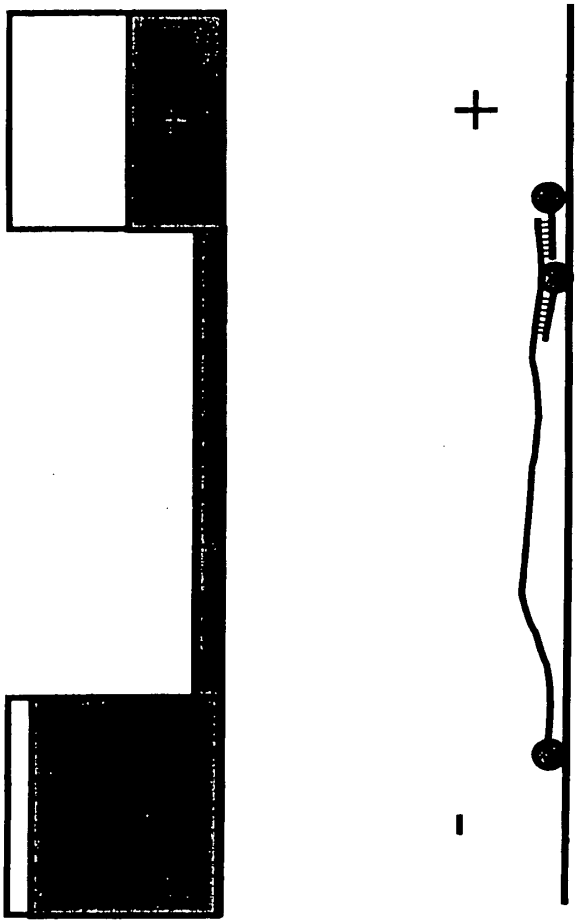


Figure 15C

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 861

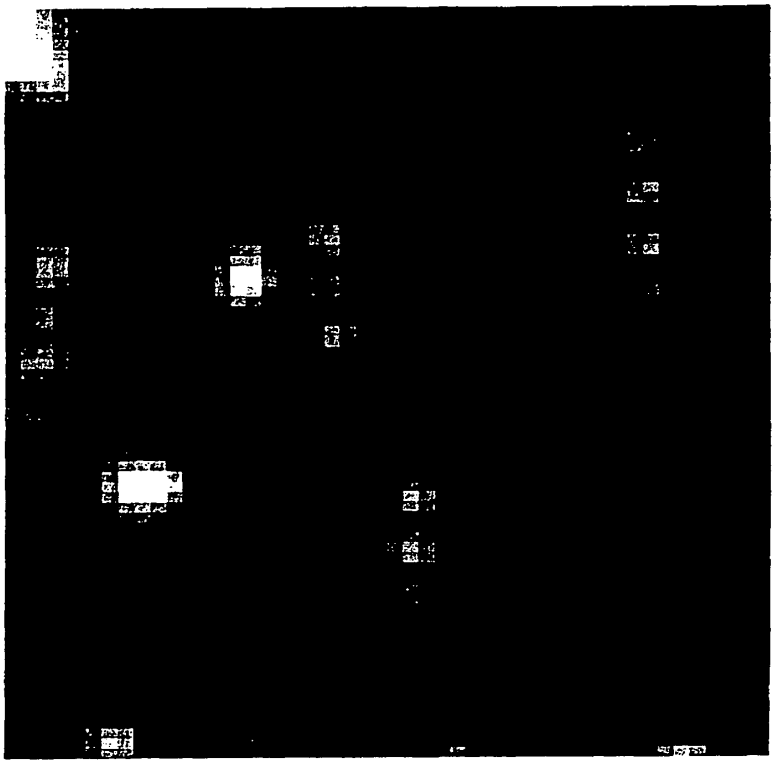


Figure 16

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 862

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 December 2007 (06.12.2007)

PCT

(10) International Publication Number
WO 2007/139766 A3

(51) International Patent Classification:
G01N 33/48 (2006.01)

(21) International Application Number:
PCT/US2007/012130

(22) International Filing Date: 21 May 2007 (21.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/802,862 22 May 2006 (22.05.2006) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 210 Elliot Avenue West, Suite 300, Seattle, WA 98119 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HWANG, Jenq-Neng** [US/US]; 15125 S.E. 66th Street, Bellevue, WA 98006 (US). **MITTON, Jeffrey, D.** [US/US]; 4232 134th Avenue SE, Bellevue, WA 98006 (US).

(74) Agents: **ANTLER, Adriane, M.** et al.; Jones Day, 222 East 41st Street, New York, NY 10017-6702 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
11 December 2008

(54) Title: SYSTEMS AND METHODS FOR ANALYZING NANOREPORTERS

(57) Abstract: Methods, computers, and computer program products for detecting the presence of a probe within a sample overlayed on a substrate are provided. The probe comprises a plurality of spatially arranged labels. A data storage module stores a plurality of light images, where each light image has light from the sample at a corresponding wavelength range in a plurality of different wavelength ranges. A label identification module identifies a plurality of labels in the plurality of light images that are proximate to each other on the substrate. A spatial order of the plurality of labels determines a string sequence of the plurality of labels. A probe identification module determines whether the string sequence of the plurality of labels comprises a valid reporter sequence.

WO 2007/139766 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US07/12130

A. CLASSIFICATION OF SUBJECT MATTER
IPC: **G01N 33/48**(2006.01)

USPC: 702/19
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 702/19

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2003/0157504 (CHEE et al.) 2003 August 21 (21.08.2003), page 3, paragraph 27; page 4, paragraphh 0042; page 7, paragraph 0072; page 13, paragraph 0131-0134; page 15, paragraph 0162-page 16, paragraph 0166; page 19, paragraph 199-page 22, paragraph 233; page 23, paragraph 0237-page 24, paragraph 0253.	66-77, 83-93, 95, 104-109, 112-135, 137, 144, 145, 147-158 --- 97-101, 139-143
Y	US 2005/0014217 A1 (MATTEAKIS et al.) 20 January 2005 (20.01.2005), page 7, paragraph 094-page 8, paragraph 0107).	97-101 and 139-143
A	GEISS et al., "Direct Multiplexed Measurements of Gene Expression with Color-Coded Probe Pairs," Nature Biotechnology (2008) Volume 26, number 3, pages 317-325.	66-158
A	FORTINA et al., "Digital mRNA Profiling," Nature Biotechnology (2008) Volume 25, Number 3, pages 293 and 294	66-158

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

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

"E" earlier application or patent published on or after the international filing date

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 September 2008 (04.09.2008)	17 OCT 2008
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201	Authorized officer /JERRY LIN/ Telephone No. (571) 272 1600

Form PCT/ISA/210 (second sheet) (April 2007)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US07/12130

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

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

1. ☒ Claims Nos.: 1-65
because they relate to subject matter not required to be searched by this Authority, namely:
Instant claims 1-65 are drawn to a computer program. However, a computer program is non-statutory under 35 U.S.C. 101. Thus, the instant claims will not be examined under Rule 67.1.

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

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 865

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US07/12130

Continuation of B. FIELDS SEARCHED Item 3:
WEST, embase, biosis medline, compusci
terms: microarray, image, porbe, colar, code tag, nanoreporter, nanostring

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 866

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 October 2008 (16.10.2008)

PCT

(10) International Publication Number
WO 2008/124847 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2008/059959

(22) International Filing Date: 10 April 2008 (10.04.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/922,817 10 April 2007 (10.04.2007) US
61/029,220 15 February 2008 (15.02.2008) US

(71) Applicant *(for all designated States except US)*: **NANOSTRING TECHNOLOGIES, INC.** [US/US]; 201 Elliott Avenue West, Suite 300, Seattle, Washington 98119 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **GEISS, Gary, K.** [US/US]; 201 Elliott Avenue West, Suite 300, Seattle, Washington 98119 (US). **DAHL, Timothy** [US/US]; 201 Elliott Avenue, Suite 300, Seattle, Washington 98119 (US). **DAVIDSON, Eric, H.** [US/US]; 201 Elliott Avenue, Suite 300, Seattle, Washington 98119 (US).

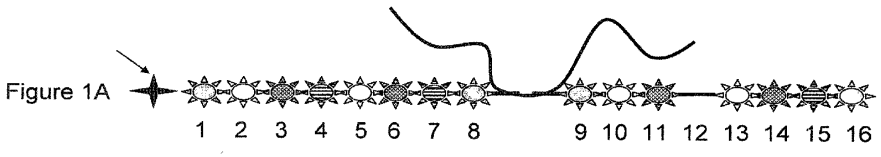
(74) Agents: **LAHERTY, Carol, D.** et al.; Seed Intellectual Property Law Group PLLC, Suite 5400, 701 Fifth Avenue, Seattle, Washington 98104-7064 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *without international search report and to be republished upon receipt of that report*

(54) **Title:** METHODS AND COMPUTER SYSTEMS FOR IDENTIFYING TARGET-SPECIFIC SEQUENCES FOR USE IN NANOREPORTERS



(57) **Abstract:** The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to coded, labeled probes that are capable of binding to and identifying target molecules based on the probes' label codes. Methods, computers, and computer program products for identifying target-specific sequences for inclusion in the probes are also provided, as are methods of making and using such probes. The probes can be used in diagnostic, prognostic, quality control and screening applications.

WO 2008/124847 A2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 867

METHODS AND COMPUTER SYSTEMS FOR IDENTIFYING TARGET-SPECIFIC SEQUENCES FOR USE IN NANOREPORTERS

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Patent Application No. 61/029,220, filed February 15, 2008 and U.S. Provisional Patent Application No. 60/922,817, filed April 10, 2007; where these provisional applications are incorporated herein by reference in their entireties.

FIELD OF THE INVENTION

The present invention relates to compositions and methods for detection and quantification of individual target molecules in biomolecular samples. In particular, the invention relates to coded, labeled reporter molecules, referred to herein as labeled "nanoreporters," that are capable of binding individual target molecules. Through the nanoreporters' label codes, the binding of the nanoreporters to target molecules results in the identification of the target molecules. Methods of making and using such nanoreporters are also provided. The nanoreporters can be used in diagnostic, prognostic, quality control and screening applications.

BACKGROUND OF THE INVENTION

This invention relates generally to the field of detection, identification, and quantification of target molecules in mixtures.

Although all cells in the human body contain the same genetic material, the same genes are not active in all of those cells. Alterations in gene expression patterns can have profound effects on biological functions. These variations in gene expression are at the core of altered physiologic and pathologic processes. Therefore, identifying and quantifying the expression of genes in normal cells compared to diseased cells can aid the discovery of new drug and diagnostic targets.

Nucleic acids can be detected and quantified based on their specific polynucleotide sequences. The basic principle underlying existing methods of detection and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 868

quantification is the hybridization of a labeled complementary probe sequence to a target sequence of interest in a sample. The formation of a duplex indicates the presence of the target sequence in the sample and the degree of duplex formation, as measured by the amount of label incorporated in it, is proportional to the amount of the target sequence.

This technique, called molecular hybridization, has been a useful tool for identifying and analyzing specific nucleic acid sequences in complex mixtures. This technique has been used in diagnostics, for example, to detect nucleic acid sequences of various microbes in biological samples. In addition, hybridization techniques have been used to map genetic differences or polymorphisms between individuals. Furthermore, these techniques have been used to monitor changes in gene expression in different populations of cells or in cells treated with different agents.

In the past, only a few genes could be detected in a complex sample at one time. Within the past decade, several technologies have made it possible to monitor the expression level of a large number of transcripts within a cell at any one time (see, *e.g.*, Schena *et al.*, 1995, Science 270: 467-470; Lockhart *et al.*, 1996, Nature Biotechnology 14: 1675-1680; Blanchard *et al.*, 1996, Nature Biotechnology 14:1649). In organisms for which most or all of the genome is known, it is possible to analyze the transcripts of large numbers of the genes within the cell. Most of these technologies employ, DNA microarrays, devices that consist of thousands of immobilized DNA sequences present on a miniaturized surface that have made this process more efficient. Using a microarray, it is possible in a single experiment to detect the presence or absence of thousands of genes in a biological sample. This allows researchers to simultaneously perform several diagnostic tests on one sample, or to observe expression level changes in thousands of genes in one experiment. Generally, microarrays are prepared by binding DNA sequences to a surface such as a nylon membrane or glass slide at precisely defined locations on a grid. Then nucleic acids in a biological sample are labeled and hybridized to the array. The labeled sample DNA marks the exact position on the array where hybridization occurs, allowing automatic detection.

Unfortunately, despite the miniaturization of array formats, this method still requires significant amounts of the biological sample. However, in several cases, such as biopsies of diseased tissues or samples of a discrete cell type, the biological sample is in

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 869

limited supply. In addition, the kinetics of hybridization on the surface of a microarray is less efficient than hybridization in small amounts of aqueous solution. Moreover, while methods exist to estimate the amount of nucleic acid present in a sample based on microarray hybridization result, microarray technology thus far does not allow for detection of target molecules on an individual level, nor are there microarray-based methods for directly quantifying the amount of target molecule in a given sample.

Thus, there exists a need for accurate and sensitive detection, identification and quantification of target molecules in complex mixtures.

Discussion or citation of a reference herein shall not be construed as an admission that such reference is prior art to the present invention.

SUMMARY OF THE INVENTION

One aspect of the present invention provides a computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein. The computer program mechanism is for identifying and selecting target-specific sequences useful in the probes of the invention. The computer program mechanism comprises a data storage module and a sequence selection module. The data storage module comprises one or more sequence databases. The sequence selection module comprises instructions for assessing the suitability of a sequence for use as a target-specific sequence in the probes of the invention and/or for selecting target-specific sequences for use in the probes of the invention. The sequence selection module can be a single-tiered or multi-tiered program that identifies useful target-specific sequences.

Each of the methods, computer program products, and computers disclosed herein optionally further comprise a step of, or instructions for, outputting a result (for example, to a monitor, to a user, to computer readable media, *e.g.*, storage media or to a remote computer). Here the result is any result obtained by the methods, computer program products, and computers disclosed herein.

In certain aspects, the present invention provides a method (reflected in Figure 20A-C) for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, comprising the steps of: (a) generating a first pool of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 870

candidate nucleotide sequences of a first predetermined length or lengths that are reverse complements of a target mRNA sequence, wherein each candidate nucleotide sequence can be divided into two adjacent nucleotide sequences of equal length consisting of a 5′ candidate sequence and a 3′ candidate sequence; (b) deleting from said first pool one or more candidate nucleotide sequences that meet at least two of the following criteria: (i) contain inverted repeats of greater than a predetermined length of consecutive nucleotides; (ii) contain direct repeats of greater than a predetermined length of consecutive nucleotides; (iii) whose 5′ candidate sequence and/or 3′ candidate sequence have a GC content outside a predetermined range; (iv) whose 5′ candidate sequence and/or 3′ candidate sequence contain contiguous stretches of C residues of greater than a predetermined length; and (v) whose 5′ candidate sequence and/or 3′ candidate sequence have melting temperatures that are outside a first predetermined melting temperature range; thereby generating a second pool of candidate nucleotide sequences; (c) deleting from said second pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a cross-hybridization potential to non-specific sequences that is higher than a predetermined threshold, thereby generating a third pool of candidate nucleotide sequences; (d) deleting from said third pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature outside a second predetermined temperature range, wherein the second predetermined melting temperature range is within the first predetermined melting temperature range; (e) determining the melting temperature for a modified 5′ candidate sequence or a modified 3′ candidate sequence, wherein the modified 5′ candidate sequence or a modified 3′ candidate sequence is a modified form of a 5′ candidate sequence or a 3′ candidate sequence, respectively, of a candidate nucleotide sequence deleted in step (d) because its 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature above the second predetermined range, wherein the modified 5′ candidate sequence has been modified by trimming at least one nucleotide from the 5′ end of the corresponding 5′ candidate sequence, and wherein the modified 3′ candidate sequence has been modified by trimming at least one nucleotide from the 3′ end of the corresponding 3′ candidate sequence; (f) in the event that: (A) the modified 5′ or modified 3′ candidate sequence, and (B) a 3′ or 5′, respectively, candidate sequence or

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 871

the modified form thereof; each have a melting temperature within the second predetermined melting temperature range and both are derived from the same candidate nucleotide sequence; adding to the third pool a modified candidate nucleotide sequence composed of (A) and (B), thereby generating a fourth pool of candidate nucleotide sequences; (g) in the event that the length of the modified 5′ or modified 3′ candidate sequence is greater than a second predetermined length, repeating step (e) one or more times wherein the modified 5′ candidate sequence or modified 3′ candidate sequence, respectively, has been trimmed by a greater number of nucleotides than in step (e) each time, until the length of the modified 5′ or modified 3′ candidate sequence is the earlier of (i) equal to, or (ii) lower than, the second predetermined length; (h) for each modified 5′ or modified 3′ candidate sequence of step (g) wherein: (C) said modified 5′ or modified 3′ candidate sequence, and (D) a 3′ or 5′, respectively, candidate sequence or the modified form thereof; each have a melting temperature within the second predetermined melting temperature range and both are derived from the same candidate nucleotide sequence; adding to the third pool a modified candidate sequence composed of (C) and (D), thereby generating a fifth pool of candidate nucleotide sequences; and (i) optionally repeating steps (e)-(h) for one or more different candidate nucleotide sequences deleted in step (d), thereby generating a sixth pool of candidate nucleotide sequences, whereby the fourth, fifth and sixth pools consist of candidate nucleotide sequences composed of pairs of adjacent target-specific sequences for use in a probe pair hybridizable to the target mRNA. Optionally, the method further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences or modified 5′ candidate sequences and/or 3′ candidate sequences or modified 3′ candidate sequences contained therein. Moreover, the candidate nucleotide sequences and/or modified candidate nucleotide sequences are optionally outputted as pairs of adjacent target-specific nucleotide sequences derived from said candidate nucleotide sequences and/or modified candidate nucleotide sequences, respectively.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 872

In certain aspects of the foregoing methods for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, the method comprises in step (b) deleting one or more candidate nucleotide sequences that at least 3 (*e.g.*, 3, 4 or all 5) criteria of step (b) from the first pool.

According to the above methods, in step (c) a 5′ candidate sequence or 3′ candidate sequence can, in certain embodiments, be deemed to have a cross-hybridization potential to non-specific sequences that is higher than said predetermined threshold if said 5′ candidate sequence or 3′ candidate sequence has (i) a sequence percentage identity with a first sequence (hereinafter "first non-target sequence") or its complement that is equal to or greater than a first predetermined cutoff, said first non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said first non-target sequence being present in a database comprising cellular mRNA sequences or cDNA sequences derived therefrom; and (ii) a contiguous block of sequence identity with a second sequence (hereinafter "second non-target sequence") or its complement that is equal to or greater than a second predetermined cutoff, said second non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said second non-target sequence being present in the database. The first non-target sequence and the second non-target sequence can be the same or they can be different.

According to the foregoing methods for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences contained therein are, in certain embodiments, outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences. In a specific embodiment, the weighted score is calculated according to the formula:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 873

(Tm score * WFa) + (MCB score * WFb) + (PID score * WFc)

where:

Tm score is a melting temperature score calculated according to the formula:

(differential score + general score)/3

where the differential score is calculated according to the following formula:

1-|(TmA - TmB)| / (TmHco - TmLco)

where the general score is calculated according to the following formula:

(((TmI - |(TmA - TmI)|) /TmI) + (((TmI - |(TmB - TmI)|) / TmI)))

where TmA is the melting temperature of the 5′ candidate sequence or modified 5′ candidate sequence in a pair of adjacent target-specific sequences, TmB is the melting temperature of the 3′ candidate sequence or modified 3′ candidate sequence in said pair of adjacent target-specific sequences, TmHco is the upper limit of the second predetermined temperature range; TmLco is the lower limit of the second predetermined temperature range; and TmI is a predetermined ideal melting temperature;

where:

MCB score is a maximum contiguous block score calculated according to the formula:

1 - (MCB / MCBco);

where MCB is the greater of (i) and (ii) below, where (i) and (ii) are respectively:

(i) the maximum contiguous block of identity between (A) and (B) below:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the maximum contiguous block of identity between (A) and (B) below:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 874

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA,

and wherein MCBco is the first predetermined cutoff;

where:

PID score is a percent identity score calculated according to the formula:

$$1 - (PID / PIDco));$$

where PID is the greater of (i) and (ii) below, where (i) and (ii) are respectively:

(i) the greatest percentage sequence identity between (A) and (B) below:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the greatest percentage sequence identity between (A) and (B) below:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA,

and wherein PIDco is the second predetermined cutoff,

and where WFa, WFb, and WFc are each independently a weighting factor, each of which is a real number.

In certain specific embodiments, the present invention provides a method for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA, comprising the steps of: (a) generating a first pool of candidate nucleotide

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 875

sequences of 100 nucleotides that are reverse complements of a target mRNA sequence, wherein each candidate nucleotide sequence can be divided into two adjacent nucleotide sequences of 50 nucleotides each, said adjacent nucleotide sequences consisting of a 5′ candidate sequence and a 3′ candidate sequence; (b) deleting from said first pool one or more candidate nucleotide sequences that meet the following criteria: (i) contain inverted repeats that are 6 consecutive nucleotides in length or greater; (ii) contain direct repeats that are 9 consecutive nucleotides in length or greater; (iii) whose 5′ candidate sequence and/or 3′ candidate sequence have a GC content outside 40-70%; (iv) whose 5′ candidate sequence and/or 3′ candidate sequence contain contiguous stretches of 3 C residues or greater; and (v) whose 5′ candidate sequence and/or 3′ candidate sequence have melting temperatures that are outside a range of (A) 60-90°C or (B) 65-85°C; thereby generating a second pool of candidate nucleotide sequences; (c) deleting from said second pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has (i) a sequence percentage identity of 85% or greater with a first sequence (hereinafter "first non-target sequence") or its complement, said first non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said first non-target sequence being present in a database comprising cellular mRNA sequences or cDNA sequences derived therefrom; and (ii) a contiguous block of sequence identity of 15 nucleotides or greater with a second sequence (hereinafter "second non-target sequence") or its complement, said second non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said second non-target sequence being present in the database; thereby generating a third pool of candidate nucleotide sequences; (d) deleting from said third pool one or more candidate nucleotide sequences whose 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature outside the range of 78-83°C; (e) determining the melting temperature for a modified 5′ candidate sequence or a modified 3′ candidate sequence, wherein the modified 5′ candidate sequence or a modified 3′ candidate sequence is a modified form of a 5′ candidate sequence or a 3′ candidate sequence, respectively, of a candidate nucleotide sequence deleted

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 876

in step (d) because its 5′ candidate sequence and/or 3′ candidate sequence has a melting temperature above 83°C, wherein the modified 5′ candidate sequence has been modified by trimming at least one nucleotide from the 5′ end of the corresponding 5′ candidate sequence, and wherein the modified 3′ candidate sequence has been modified by trimming at least one nucleotide from the 3′ end of the corresponding 3′ candidate sequence; (f) in the event that: (A) the modified 5′ or modified 3′ candidate sequence, and (B)a 3′ or 5′, respectively, candidate sequence or the modified form thereof, each have a melting temperature within the range of 78-83°C and both are derived from the same candidate nucleotide sequence, adding to the third pool a modified candidate nucleotide sequence composed of (A) and (B); thereby generating a fourth pool of candidate nucleotide sequences;(g) in the event that the length of the modified 5′ or modified 3′ candidate sequence is greater than 35 nucleotides, repeating step (e) one or more times wherein the modified 5′ candidate sequence or modified 3′ candidate sequence, respectively, has been trimmed by a greater number of nucleotides than in step (e) each time, until the length of the modified 5′ or modified 3′ candidate sequence is the earlier of (i) equal to, or (ii) lower than, 35 nucleotides; (h) for each modified 5′ or modified 3′ candidate sequence of step (g) wherein: (C) the modified 5′ or modified 3′ candidate sequence, and (D) a 3′ or 5′, respectively, candidate sequence or modified candidate sequence; each have a melting temperature in the range of 78-83°C and both are derived from the same candidate nucleotide sequence, adding to the third pool a modified candidate sequence composed of (C) and (D); thereby generating a fifth pool of candidate nucleotide sequences; and (i) optionally repeating steps (e)-(h) for one or more different candidate nucleotide sequences deleted in step (d), thereby generating a sixth pool of candidate nucleotide sequences, whereby the fourth, fifth and sixth pools consist of candidate nucleotide sequences composed of pairs of adjacent target-specific sequences for use in a probe pair hybridizable to the target mRNA. The first non-target sequence and the second non-target sequence can be the same or they can be different.

Optionally, the method further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 877

or modified 5′ candidate sequences and/or 3′ candidate sequences or modified 3′ candidate sequences contained therein. Moreover, the candidate nucleotide sequences and/or modified candidate nucleotide sequences are optionally outputted as a pair of adjacent target-specific nucleotide sequences derived from said candidate nucleotide sequences and/or modified candidate nucleotide sequences, respectively.

In certain embodiments of the foregoing method, a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools and/or the 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences contained therein are outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said 5′ candidate sequences or modified 5′ candidate sequences and 3′ candidate sequences or modified 3′ candidate sequences. In a specific embodiment, the weighted score is calculated according to the formula:

(Tm score * WFa) + (MCB score * WFb) + (PID score * WFc)

where:

Tm score is a melting temperature score calculated according to the formula:

(differential score + general score)/3

where the differential score is calculated according to the following formula:

1-|(TmA - TmB)| / 5

where the general score is calculated according to the following formula:

(((80.5 - |(TmA - 80.5)|) / 80.5) + (((80.5 - |(TmB - 80.5)|) / 80.5)))

where TmA is the melting temperature of the 5′ candidate sequence or modified 5′ candidate sequence of a pair of adjacent target-specific sequences and TmB is the melting temperature of the 3′ candidate sequence or modified 3′ candidate sequence of said pair of adjacent target-specific sequences;

where:

MCB score is a maximum contiguous block score calculated according to the formula:

1 - (MCB / 15);

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 878

where MCB is the greater of (i) and (ii) below, where (i) and (ii) are respectively:

(i) the maximum contiguous block of identity between (A) and (B) below:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the maximum contiguous block of identity between (A) and (B) below:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

where:

PID score is a percent identity score calculated according to the formula:

$$1 - (\text{PID} / 85\%);$$

where PID is the greater of (i) and (ii) below, wherein (i) and (ii) are respectively:

(i) the greatest percentage sequence identity between (A) and (B) below:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA;

and

(ii) the greatest percentage sequence identity between (A) and (B) below:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 879

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA,

and wherein PIDco is the second predetermined cutoff,

and where WFa, WFb, and WFc are each independently a weighting factor, each of which is a real number.

The foregoing methods for identifying a pair of adjacent target-specific sequences for use in a probe pair hybridizable to a target mRNA can be utilized for identifying a plurality of pairs of adjacent target-specific sequences for use in a respective plurality of probe pairs, each probe pair being hybridizable to a different target mRNA, comprising, for each target mRNA: identifying a pair of adjacent target-specific sequences according to any embodiment of the foregoing methods.

The present invention yet further provides a method (reflected in Figure 21A-C) for identifying a target-specific nucleotide sequence for use in a probe hybridizable to a target mRNA, comprising the steps of: (a) generating a first pool of candidate nucleotide sequences of a first predetermined length or lengths that are reverse complements of a target mRNA sequence; (b) deleting from said first pool one or more candidate nucleotide sequences that meet at least two of the following criteria: (i) contains inverted repeats of greater than a predetermined length of consecutive nucleotides; (ii) contains direct repeats of greater than a predetermined length of consecutive nucleotides; (iii) has a GC content outside a predetermined range; (iv) contains a contiguous stretch of C residues of greater than a predetermined length; and (v) has a melting temperature that is outside a first predetermined melting temperature range; thereby generating a second pool of candidate nucleotide sequences; (c) deleting from said second pool one or more candidate nucleotide sequences that have a cross-hybridization potential to non-specific sequences that is higher than a predetermined threshold, thereby generating a third pool of candidate nucleotide sequences; (d) deleting from said third pool one or more candidate nucleotide sequences that have a melting temperature outside a second predetermined temperature range, wherein the second predetermined melting temperature range is within the first predetermined melting temperature range; (e) determining the melting temperature for a modified candidate

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 880

nucleotide sequence, wherein the modified candidate nucleotide sequence is a modified form of a candidate nucleotide sequence deleted in step (d) because it has a melting temperature above the second predetermined range, wherein the modified candidate nucleotide sequence has been modified by trimming at least one nucleotide from the 5′ end or the 3′ end of said candidate nucleotide sequence; (f) in the event that the modified candidate nucleotide sequence has a melting temperature within the second predetermined melting temperature range, adding to the third pool the modified candidate nucleotide sequence, thereby generating a fourth pool of candidate nucleotide sequences; (g) in the event that the length of the modified candidate nucleotide sequence is greater than a second predetermined length, repeating step (e) one or more times wherein the modified candidate nucleotide sequence has been trimmed by a greater number of nucleotides than in step (e) each time, until the length of the modified candidate nucleotide sequence is the earlier of (i) equal to, or (ii) lower than, the second predetermined length; (h) adding to the third pool each modified candidate nucleotide sequence of step (g) which has a melting temperature within the second predetermined melting temperature range; thereby generating a fourth pool of candidate nucleotide sequences; and (i) optionally repeating steps (e)-(h) for one or more different candidate nucleotide sequences deleted in step (d), thereby generating a sixth pool of candidate nucleotide sequences, whereby the fourth, fifth and sixth pools consist of target-specific nucleotide sequences for use in a probe hybridizable to a target mRNA. The method optionally further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools.

According to the foregoing method for identifying a target-specific nucleotide sequence for use in a probe hybridizable to a target mRNA, in step (c) a candidate target-specific sequence can, in certain embodiments, be deemed to have a cross-hybridization potential to non-specific sequences that is higher than said predetermined threshold if said candidate target-specific sequence has (i) a sequence percentage identity with a first sequence (hereinafter "first non-target sequence") or its complement that is equal to or greater than a first predetermined cutoff, said first non-target sequence being other than the complement of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 881

the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said first non-target sequence being present in a database comprising cellular mRNA sequences or cDNA sequences derived therefrom; and (ii) a contiguous block of sequence identity with a second sequence (hereinafter "second non-target sequence") or its complement that is equal to or greater than a second predetermined cutoff, said second non-target sequence being other than the complement of the target mRNA and, optionally, other than the complements of one or more alternatively spliced mRNAs corresponding to the same gene as the target mRNA, and said second non-target sequence being present in the database. The first non-target sequence and the second non-target sequence can be the same or they can be different.

In the foregoing method for identifying a target-specific nucleotide sequence for use in a probe hybridizable to a target mRNA, a plurality of candidate nucleotide sequences and/or modified candidate nucleotide sequences in the fourth, fifth and/or sixth pools are optionally outputted or displayed in a ranked order based on a weighted score of the cross-hybridization potentials and the melting temperatures of said candidate nucleotide sequences and/or modified candidate nucleotide sequences.

The foregoing methods for identifying a target-specific nucleotide sequence for use in a probe hybridizable to a target mRNA can be utilized for identifying a plurality of target-specific sequences for use in a respective plurality of probes, each probe being hybridizable to a different target mRNA, comprising, for each target mRNA: identifying a target-specific sequence according to any embodiment of the foregoing method.

In certain aspects of the foregoing methods for identifying a target-specific nucleotide sequence for use in a probe hybridizable to a target mRNA, one or more candidate nucleotide sequences that meet 3 or more, *e.g.*, 3, 4 or all 5, criteria of step (b) are deleted from the first pool in step (b).

In any of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, if the fourth, fifth and/or sixth pools contains no candidate nucleotide sequences, the method may further comprise repeating steps (b) to (i), wherein step (b) is performed under more relaxed criteria (*e.g.*, with an increased predetermined length of direct and/or inverted repeats and/or a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 882

broader range of GC content and/or a broader range of melting temperatures and/or a greater predetermined length of contiguous C residues).

In specific embodiments of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, each repetition of the trimming step (e) can be performed in increments of 1 nucleotide, 2 nucleotides, 3 nucleotides, 4 nucleotides, or more.

In specific embodiments of the foregoing methods for identifying a pair of adjacent target-specific sequences, wherein the first predetermined length can be selected from the range of 70 to 120 nucleotides, and/or the second predetermined length can be selected from the range of 30 to 45 nucleotides. In a specific embodiment, the second predetermined length is selected from the range of 35-40 nucleotides.

In specific embodiments of the foregoing methods for identifying a target-specific nucleotide sequence, the first predetermined length can be selected from the range of 35 to 60 nucleotides and/or the second predetermined length can selected from the range of 30 to 45 nucleotides. In a specific embodiment, the second predetermined length is selected from the range of 35-40 nucleotides.

In certain embodiments of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, the predetermined length of the inverted repeats of step (b)(i) can be selected from the range of 5 to 7 consecutive nucleotides and/or the predetermined length of the direct repeats of step (b)(ii) can be selected from the range of 7 to 9 consecutive nucleotides and/or the predetermined range of GC content of step (b)(iii) can be from 35-45% at the lower limit to 65-80% at the upper limit. In a specific embodiment, the predetermined range of GC content of step (b)(iii) is 40-70%.

In certain embodiments of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, the predetermined length in step (b)(iv) is preferably 3.

In certain embodiments of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, the highest and lowest temperatures of the first predetermined melting temperature range

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 883

preferably differ by 15°C to 30°C, and most preferably by 20°C to 25°C. In specific embodiments, the first predetermined melting temperature range is from 60°C to 90°C, from 65°C to 85°C, or from 65°C to 90°C.

In aspects of the present methods that entail the use of a first predetermined cutoff for cross-hybridization potential determination, the first predetermined cutoff is preferably selected from the range of 70-95% sequence identity, more preferably selected from the range of 80-90% sequence identity. In a specific embodiment, the first predetermined cutoff is 85% sequence identity.

In aspects of the present methods that entail the use of a second predetermined cutoff for cross-hybridization potential determination, wherein the second predetermined cutoff is preferably selected from the range of 10-18 contiguous nucleotides, and more preferably from the range of 14-16 contiguous nucleotides.

In certain embodiments of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, the highest and lowest temperatures of the second predetermined melting temperature range preferably differ by 4°C to 8°C. In specific embodiments, the second predetermined melting temperature range is from 78°C to 83°C.

In certain embodiments, the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences further comprises the step of deleting from the fourth, fifth and/or sixth pools candidate nucleotide sequences that have a cross-hybridization potential to sequences present in other components of the probe or in a preparation step for the probe.

In certain embodiments of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences, the target mRNA is an alternatively spliced mRNA. In such embodiment, the methods may further comprise the step of determining whether one or more candidate nucleotide sequences are unique to one splice form or common to more than one splice form of the target mRNA. Alternatively, the first pool of candidate nucleotide sequences is designed to contain only candidate nucleotide sequences unique to one splice form or only candidate nucleotide sequences common to multiple splice forms.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 884

Any of the foregoing methods for identifying a target-specific nucleotide sequence and/or identifying a pair of adjacent target-specific sequences can be-a computer .implemented method

The present invention further provides a computer system for identifying a pair of adjacent target-specific sequences, for use in a probe pair hybridizable to a target mRNA, comprising a processor and a memory coupled with the processor comprising a plurality of machine instructions that cause the processor to perform the method of any one of the foregoing methods for identifying a pair of adjacent target-specific sequences. The present invention yet further provides computer system for identifying a target-specific sequence for use in a probe hybridizable to a target mRNA, comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform any one of the foregoing methods for identifying a target-specific sequence.

The present invention yet further provides a computer system for identifying a plurality of pairs of adjacent target-specific sequences for use in a respective plurality of probe pairs, each probe pair being hybridizable to a different target mRNA, comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform any of the foregoing methods for identifying a plurality of pairs of adjacent target-specific sequences.

The present invention yet further provides a computer system for identifying a plurality of target-specific sequences for use in a respective plurality of probes, each probe being hybridizable to a different target mRNA, each probe being specific sequences for use in a respective plurality of probes, comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform any one of the foregoing methods for identifying a target-specific sequence.

The present invention further provides computer system comprising: a processor and a memory coupled with the processor, the memory storing a plurality of machine instructions that cause the processor to perform anone of the foregoing method for identifying a target-specific sequence.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 885

The present invention yet further provides a computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising instructions for performing any of the foregoing methods for identifying a target-specific nucleotide sequence (or a plurality thereof) and/or identifying a pair of adjacent target-specific sequences (or a plurality thereof).

BRIEF DESCRIPTION OF THE FIGURES

FIGURE 1A-1F: Figure 1A illustrates a dual nanoreporter with a 16-position nanoreporter code, using two 8-position nanoreporter components. Figure 1B illustrates a dual nanoreporter with a 9-position nanoreporter code, using one 8-position nanoreporter component and one single-position nanoreporter component. Figure 1C illustrates a dual nanoreporter with an 8-position nanoreporter code, using one ghost probe and one 8-position nanoreporter component. Figure 1D illustrates a single nanoreporter with an 8-position nanoreporter code. In Figures 1A-1D, the star shape (depicted with an arrow) is illustrative of an affinity tag, which can be used to purify the nanoreporter or immobilize the nanoreporter (or nanoreporter-target molecule complex) for the purpose of imaging. The numbered regions in Figure 1A-1D refer to separate label attachment regions. All except for position 12 of Figure 1A are labeled with one of four types of label monomers, depicted as grey, white, hatched or stripe "sun" diagrams. Position 12 of Figure 1A is an unlabeled "dark spot." Figures 1E and 1F represent variations on the nanoreporters of Figures 1B and 1D, respectively, in which the target molecule to which the nanoreporters are bound comprises biotin moieties (shown as small asterisks), for example biotin-modified nucleotides randomly incorporated into a target nucleic acid. The nanoreporters themselves further optionally comprise an affinity tag (not shown).

FIGURE 2A-2C: Figure 2A shows an illustration of a label unit of a nanoreporter, containing a scaffold with patch units and corresponding split flaps disposed along its length. Figure 2B illustrates the components of a single patch pair and its corresponding flap, containing: 1: a portion of a nanoreporter scaffold (*e.g.*, M13 single-stranded DNA); 2: A patch pair; 3: a split flap pair; and 4: labeled oligonucleotides, each

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 886

with a label monomer incorporated, hybridized to the split flap. Figure 2C shows a nanoreporter with 4 "spots," each spot designed to contain 9 patch pairs of 60-65 nucleotides, each attached to a split flap pair of 95-100 nucleotides. Each split flap pair had binding site for 12 oligonucleotides each attached to a single label monomer. Each spot therefore had binding sites for 108 label monomers.

FIGURE 3: A nanoreporter in which the patches are RNA segments can be used with (Figure 3A) and without registers (Figure 3B). Both Figure 3A and 3B depict a (1) nanoreporter scaffold (heavy black line) to which are attached (2) 8 RNA segments (heavy grey lines 1-8), (3) a target-specific sequence (dotted line "T") and (4) an oligonucleotide (checkered line "0") that is partly complementary to the nanoreporter scaffold and partly complementary to the target-specific sequence. This oligonucleotide is referred to as a "ligator" oligonucleotide. In Figure 3A, only one register, *i.e.*, every alternate RNA segment is labeled. The second register positions serve as "spacers," making it possible to generate a nanoreporter code in which consecutive positions in the code are the same "color," or spectrally indistinguishable. In Figure 3B, both registers, *i.e.*, adjacent RNA segments with no intervening spacers, are labeled, with no nearest neighbor of the same "color."

FIGURE 4: Is an image of a dual nanoreporter hybridized to a target molecule. Here, both registers are labeled. The nanoreporters are labeled with three different colors, Alexa 488, Cy3 and Alexa 647 (labeled 1, 2 and 3, respectively). The left brackets show one probe of the dual nanoreporter and the right brackets show the other probe of the dual nanoreporter. Colors 1, 2 and 3 were each acquired in different channels and the first and second registers, seen as rows of spots, were shifted up by several pixels to be able to show each register individually.

FIGURE 5A-5D: This figure illustrates the various components of the dual nanoreporters shown in Figure 4. Figure 5A illustrates one color (here, Alexa 488, depicted in the left column as open circles), which is spectrally distinguishable from Cy3 (shown in Figure 5B, depicted in the left column as vertically striped circles) and Alexa 647 (shown in Figure 5C as diagonally striped circles). The images obtained from each were superimposed to generate Figure 5D.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 887

FIGURE 6A-6E: Figure 6A is a schematic illustration of the experiment shown in Figures 6B and 6C. In this case, the star represents biotin that was used to attach the complex by one end to the surface prior to stretching. Figures 6B and 6C show images from experiments in which S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA (Figure 6B) or S2 target RNA (Figure 6C) were hybridized. Figure 6E shows a close-up of a nanoreporter complexes from Figure 6B, each containing S2-A ghost probe, S2-B labeled nanoreporter and S2 target DNA. Figure 6D shows an image of a negative control experiment, in which S2-A ghost probe, S2-B labeled nanoreporter and no S2 target RNA were hybridized.

FIGURE 7A-G. Figures 7A, 7B, 7C and 7D depict different permutations of patches on a nanoreporter scaffold, Figures 7E and 7F depict different permutations of split flaps on a nanoreporter scaffold, optionally hybridized to one or more oligonucleotides, as in Figure 7G. In Figure 7A-G, α refers to a 5′ or 3′ molecule or end of a molecule, and β refers to a corresponding 3′ or 5′ molecule or end of a molecule.

FIGURE 8: Figure 8 depicts a scheme in which single-stranded M13 phage is linearized for use as a nanoreporter scaffold. The circular M13 phage is annealed to a five-fold excess of BamH1 cutter oligonucleotide (hatched lines) (1), and the resulting partially double-stranded M13 digested with the restriction endonuclease BamH1 (2), resulting in a linearized M13 in which BamH1 cutter oligonucleotide is still attached (3). This M13-oligonucleotide complex is heated in the presence of an excess oligonucleotide complementary to the BamH1 cutter oligonucleotide (an "anti-BamH1 oligonucleotide") (grey lines) (4). The BamH1 cutter oligonucleotide anneals to the excess of anti-BamH1 oligonucleotide, and the M13 molecule is purified from the oligonucleotide, for example by using size exclusion columns, to yield M13 scaffold.

FIGURE 9A-9B: Shows a labeled nanoreporter with an affinity tag at each end, Al and A2. In Figure 9, the labeled nanoreporter is immobilized through the binding of A1 to an immobilized affinity partner. In the absence of an affinity binding partner for A2, the A2 end of the nanoreporter remains in solution (Figure 9A), but in the presence of an affinity binding partner (A2'), the A2 end of the nanoreporter is also immobilized (Figure 9B). Upon immobilization, the nanoreporter can be stretched, or "elongated" as depicted in

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 888

Figure 9B, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code.

FIGURE 10A-10C: Figure 10A shows a labeled nanoreporter containing a single affinity tag, A1. Another affinity tag, A2, can be attached to the nanoreporter by direct binding of the nanoreporter to a molecule containing A2 (*e.g.*, if the nanoreporter is or comprises a nucleic acid, it can hybridize directly with another nucleic acid to which A2 is attached), as depicted in Figure 10B. Alternatively, the second affinity tag, A2, can be attached to the labeled nanoreporter via a bridging molecule, such as the bridging nucleic acid ("X") depicted in Figure 10C.

FIGURE 11A-11B: Shows a labeled (nucleic acid-based) nanoreporter with an affinity tag, Al, at one end. In Figure 11, the labeled nanoreporter is immobilized through the binding of Al to an immobilized affinity partner. The other end of the nanoreporter is in solution (Figure 11A), but can be immobilized by hybridization to a complementary oligonucleotide which contains another affinity tag (A2) used to immobilize the nanoreporter (Figure 11B). A1 and A2 can be the same, for example biotin, for immobilization on an avidin- or streptavidin-coated surface. Upon immobilization of Al, the nanoreporter can be stretched, or "elongated" as depicted in Figure 11, for example by electrostretching, for separation of the label attachment regions in a manner that permits detection of the nanoreporter code. Optionally, while the nanoreporter is in an elongated state, A2 is introduced and binds the end of the nanoreporter that is complementary to A2 down to the surface.

FIGURE 12A-12B. Figure 12A provides an illustration of a nanoreporter comprising an immobilized first portion F1; and Figure 12B provides an illustration of a nanoreporter extended in an electrical field and comprising immobilized first portion F1 and immobilized second portion F2, wherein F2 is immobilized via a complex with molecule F3.

FIGURE 13A-13C. Figure 13A provides an illustration of a three-member complex for immobilization of an extended nanoreporter; Figure 13B provides an illustration of a two-member complex for immobilization of an extended nanoreporter; and Figure 13C provides an illustration of an incomplete complex for immobilization of an extended nanoreporter.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 889

FIGURE 14A-14D. Figure 14A provides an illustration of a nanoreporter comprising an immobilized first portion F1; Figure 14B provides an illustration of an extended nanoreporter immobilized at first portion F1 and at a second portion via complexes with F2; Figure 14C provides an illustration of a nanoreporter comprising a first portion immobilized to an avidin surface via biotin; and Figure 14D provides an illustration of an extended nanoreporter immobilized at a first portion and at a second portion via selective binding of biotin to an avidin surface.

FIGURE 15A-15C. Figure 15A illustrates immobilization of one terminus of a DNA molecule in a microfluidic device; Figure 15B illustrates extension of the DNA in an electric field; and Figure 15C illustrates selective immobilization of a second terminus of the extended DNA molecule.

FIGURE 16 provides an image of extended nanoreporters selectively immobilized by the methods of the present invention.

FIGURE 17 depicts the relationship between the number of label attachment regions to the calculated entanglement threshold for nanoreporters for label attachment region sizes of 900 bp and 1100 bp

FIGURE 18 is a scatter plot showing normalized and average $\log_2$ signal values from each positive sample (n=3) for all 509 genes whose expression was measured in a nanoreporter multiplex assay as described in Example 9 (Section 14) below.

FIGURE 19 illustrates a computer system in accordance with an embodiment of the present invention.

FIGURE 20A-20C illustrates the steps of an exemplary method for the identification of a pair of adjacent target-specific sequences, which can be used in a probe pair hybridizable to a target mRNA.

FIGURE 21A-21C illustrates the steps of an exemplary method for the identification of a target-specific sequence, which can be used in a probe hybridizable to a target mRNA.

FIGURE 22A-22C provides a schematic representation of the hybridized complex (not to scale). Figure 22a shows the capture probe and reporter probe hybridized to a complementary target mRNA in solution via the gene-specific sequences. After

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 890

hybridization, the tripartite molecule is affinity-purified first by the 3'-repeat sequence and then by the 5'-repeat sequence to remove excess reporter and capture probes, respectively. Figure 22b provides a schematic representation of binding, electrophoresis, and immobilization. (*i*) The purified complexes are attached to a streptavidin-coated slide via biotinylated capture probes. (*ii*) Voltage is applied to elongate and align the molecules. Biotinylated anti-5' oligonucleotides that hybridize to the 5'-repeat sequence are added. (*iii*) The stretched reporters are immobilized by the binding of the anti-5' oligonucleotides to the slide surface via the biotin. Voltage is turned off and the immobilized reporters are prepared for imaging and counting. Figure 22c shows false-color image of immobilized reporter probes.

FIGURE 23A-23B demonstrates the linearity and reproducibility of the NanoString spike-in controls. Non-human DNA oligonucleotide targets were spiked into each sample at concentrations of 0.1, 0.5, 1, 5, 10 and 50 fM. No target was added for the two negative control probe pairs. Figure 23a shows signal (counts) on a log scale *vs.* concentration of the spike on a log scale. Each of three replicate measurements for each spike in Mock- and PV-infected RNA is shown. At this scale, the replicate measurements lie essentially on top of each other except at the lowest spike-in concentration. Figure 23b provides average signal *vs.* concentration on a linear scale for spikes in both mock- and PV-infected samples. The correlation coefficients (R^2 values) of a linear fit to the average signal are 0.9988 and 0.9992 for mock and PV-infected samples respectively. The normalized counts used to construct both graphs are available in Table 6.

FIGURE 24A-24B depicts the reproducibility and differential gene expression plots for 509 genes on the NanoString nCounter platform. Figure 24a is a scatter plot of normalized signal for all 509 genes assayed shown in log scale for technical replicates. Genes were not filtered based on detection. The R^2 value of a linear fit to this data is 0.9999 +/- 0.0002. The R^2 value for all pairwise comparisons of technical replicates for both NanoString and Affymetrix are shown in Table 6. Figure 24b is a scatter plot of mock-infected *vs.* PV-infected counts for 509 genes. The normalized average counts for the triplicate assays are shown. The top and bottom lines represent 2-fold increase and decrease in expression levels, respectively. All 509 data points are shown without filtering.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 891

FIGURE 25 shows a comparison of detected/undetected calls for the NanoString and Affymetrix assays. A set of 449 RefSeq mRNAs that had corresponding Affymetrix probe sets was used in this analysis. Figure 25a depicts mock-infected and Figure 25b depicts PV-infected samples. For the NanoString assay a gene was considered detected if the average normalized signal for the three replicates was significantly above that of the negative controls ($P \leq 0.05$). For Affymetrix assay, a gene was considered detected if any one of the three replicates was called "Present" or "Marginal" based on MAS 5.0 analysis.

FIGURE 26A-26C provides comparison plots of NanoString nCounter to Affymetrix GeneChip and Applied BioSystemsTaqMan platforms. Figure 26a provides $\log_2$ (PV-infected/mock-infected) ratios as measured by NanoString assay (x-axis) and Affymetrix arrays (y-axis). Genes were considered differentially regulated if the P-value in a Student's T-test performed on replicate data was ≤ 0.05 (n=3). Affymetrix ratios were based on RMA normalized data. A linear fit to the ratios that are deemed statistically significant in both assays (♦) yields a correlation coefficient of 0.79. Genes were not filtered based on the magnitude of fold-change or the detected/undetected calls for this analysis. A set of 14 genes whose expression levels were discordant between the two platforms and were selected for real-time PCR analysis are also shown (♦). Genes were selected based on criteria outlined in the Examples. Figure 26b demonstrates that the discordant 14 genes shown in Figure 26a as analyzed by TaqMan® real-time PCR performed in triplicate on 100ng of the same mock and PV-infected samples. The bar graph shows $\log_2$ ratios (PV-infected/mock-infected) for the NanoString (■), TaqMan (■) and Affymetrix (■) platforms in triplicate. The root mean square deviation of $\log_2$ ratios between NanoString to TaqMan was 0.34, DNA microarray to TaqMan was 1.20. Figure 26c shows the results obtained when a library of probe pairs to 35 RefSeq mRNAs that overlapped with the published MAQC consortium study was hybridized to commercially-available reference RNAs. Data was filtered to remove genes that were not detected in all samples (see Methods). The Affymetrix data shown here was downloaded from the MAQC study and represents data from a single site (site 1, Affymetrix). TaqMan real-time PCR data was performed at Applied Biosystems Inc. The R^2 values for 27 NanoString genes (♦) and 18 Affymetrix genes (■) that met the selection criteria were 0.95

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 892

and 0.83, respectively. The overall correlation of Affymetrix data for 469 genes (site 1) in the original study was 0.92.

FIGURE 27 is a graph depicting a comparison of the fold change results for all 509 genes examined. This graph is a scatter plot of log2 fold change for 317 genes that were measured by both NanoString and Affymetrix platforms. Genes are coded based upon the significance of their fold change values ($P<0.05$) in either both platforms (♦), NanoString platform only (■), Affymetrix platform only (▲), or neither platform (X). The R2 value shown represents the correlation of fold changes of genes that were found to be significant in both NanoString and microarray platforms.

FIGURE 28 shows the correlation between nCounter and real-time PCR. Individual line plots for 21 genes across 7 time points are shown. The normalized counts obtained from the NanoString system are shown ( ) on the left-hand y-axis scale. Quantitative real-time PCR results in copies/embryo are shown ( ) on the right-hand y-axis. The 7 time points (x-axis) were 0h (egg), 9.3h, 18h, 24h, 33h, 48h, and 70h. All data has been normalized to the expression levels of the polyubiquitin gene. Real-time PCR data is shown in copies/embryo and the NanoString data is shown in normalized counts. A quantitative comparison of the nCounter system and real-time PCR (not shown) revealed that estimates of the transcript number for some genes are similar in the two systems, whereas others disagree. The discrepancies are likely to reflect differences in the two platforms. The nCounter system is based on solution-hybridization kinetics, directly measures mRNA transcripts, and uses a standard curve in each reaction to estimate transcript number. In contrast, real-time PCR involves a reverse transcription step followed by amplification of a portion of the cDNA with specific primers, and transcript copy number is calculated relative to polyubiquitin expression levels.

DETAILED DESCRIPTION OF THE INVENTION

The present invention pertains to nanoreporters, and their manufacture and use. A fully assembled and labeled nanoreporter comprises two main portions, a target-specific sequence that is capable of binding to a target molecule, and a labeled region which emits a "code" of signals (the "nanoreporter code") associated with the target-specific

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 893

sequence. Upon binding of the nanoreporter to the target molecule, the nanoreporter code identifies the target molecule to which the nanoreporter is bound.

Nanoreporters are modular structures. Generally, a nanoreporter is a molecular entity containing three basic elements: a scaffold containing two or more label attachment regions, one or more patches attached to the scaffold, and a target-specific sequence, also attached to the scaffold. The elements of a nanoreporter can be found in a single molecular entity (a "singular" nanoreporter), or two distinct molecular entities (a "dual" nanoreporter). Each molecular entity may be composed of one molecule or more than one molecule attached to one another by covalent or non-covalent means. Generally, each component of a dual nanoreporter has a target-specific sequence that binds to a different site on the same target molecule. This allows for smaller nanoreporter components with more efficient kinetics of binding of the nanoreporter to the target molecule and better signal:noise ratios resulting from the greater binding specificity.

The patches attached to a nanoreporter scaffold serve to attach label monomers to a nanoreporter scaffold. Patches may be directly labeled, for example by covalent incorporation of one or more label monomers into nucleic acid patches. Alternatively, patches may be attached to flaps, which maybe labeled directly, for example by covalent incorporation of one or more label monomers into a nucleic acid flap, or indirectly, for example by hybridization of a nucleic acid flap to an oligonucleotide which is covalently attached to one or more label monomers. Where the label monomers attached to a label attachment region are not directly incorporated into a patch or flap, the patch or flap serves as a "bridge" between the label monomer and the label attachment region, and may be referred to as a "bridging molecule," *e.g.*, a bridging nucleic acid.

Additionally, nanoreporters may have affinity tags for purification and/or for immobilization (for example to a solid surface). Nanoreporters, or nanoreporter-target molecule complexes, are preferably purified in two or more affinity selection steps. For example, in a dual nanoreporter, one probe can comprise a first affinity tag and the other probe can comprise a second (different) affinity tag. The probes are mixed with target molecules, and complexes comprising the two probes of the dual nanoreporter are separated from unbound materials (*e.g.*, the target or the individual probes of the nanoreporter) by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 894

affinity purification against one or both individual affinity tags. In the first step, the mixture can be bound to an affinity reagent for the first affinity tag, so that only probes comprising the first affinity tag and the desired complexes are purified. The bound materials are released from the first affinity reagent and optionally bound to an affinity reagent for the second affinity tag, allowing the separation of complexes from probes comprising the first affinity tag. At this point only full complexes would be bound. The complexes are finally released from the affinity reagent for the second affinity tag and then preferably stretched and imaged. The affinity reagent can be any solid surface coated with a binding partner for the affinity tag, such as a column, bead (*e.g.*, latex or magnetic bead) or slide coated with the binding partner. Immobilizing and stretching nanoreporters using affinity reagents is fully described in U.S. provisional application no. 60/753,816 by Sean M. Ferree and Dwayne L. Dunaway, entitled "Compositions Comprising Oriented, Immobilized Macromolecules and Methods for Their Preparation," attorney docket no. 11616-014-888 filed on December 23, 2005, which is incorporated by reference herein in its entirety.

Nanoreporter and nanoreporter-target complexes which are or comprise nucleic acids may be affinity-purified or immobilized using a nucleic acid, such as an oligonucleotide, that is complementary to at least part of the nanoreporter or target. In a specific application where the target includes a poly A or poly dA stretch, the nanoreporter-target complex can be purified or immobilized by an affinity reagent coated with a poly dT oligonucleotide.

The sequence of signals emitted by the label monomers associated with the various label attachment regions of the scaffold of a given nanoreporter allows for the unique identification of the nanoreporter. A nanoreporter having a unique identity or unique spectral signature is associated with a target-specific sequence that recognizes a specific target molecule or a portion thereof. When a nanoreporter is exposed to a mixture containing the target molecule under conditions that permit binding of the target-specific sequence(s) of the nanoreporter to the target molecule, the target-specific sequence(s) preferentially bind(s) to the target molecule. Detection of the spectral code associated with the nanoreporter allows detection of the presence of the target molecule in the mixture (qualitative analysis). Counting all the label monomers associated with a given spectral code or signature allows the

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 895

counting of all the molecules in the mixture associated with the target-specific sequence coupled to the nanoreporter (quantitative analysis). Nanoreporters are thus useful for the diagnosis or prognosis of different biological states (*e.g.*, disease vs. healthy) by quantitative analysis of known biological markers. Moreover, the exquisite sensitivity of single molecule detection and quantification provided by the nanoreporters of the invention allows for the identification of new diagnostic and prognostic markers, including those whose fluctuations among the different biological states is too slight detect a correlation with a particular biological state using traditional molecular methods. The sensitivity of nanoreporter-based molecular detection permits detailed pharmacokinetic analysis of therapeutic and diagnostic agents in small biological samples.

Many nanoreporters, referred to as singular nanoreporters, are composed of one molecular entity, as depicted in Figure 1D. However, to increase the specificity of a nanoreporter and/or to improve the kinetics of its binding to a target molecule, a preferred nanoreporter is a dual nanoreporter composed of two molecular entities, each containing a different target-specific sequence that binds to a different region of the same target molecule. Various embodiments of dual nanoreporters are depicted in Figures 1A-1C. In a dual nanoreporter, at least one of the two molecular entities is labeled. The other molecular entity is not necessarily labeled. Such unlabeled components of dual nanoreporters are referred to herein as "ghost probes" (see Figure 1C) and often have affinity tags attached, which are useful to immobilize and/or stretch the complex containing the dual nanoreporter and the target molecule to allow visualization and/or imaging of the complex.

Because of their modular structures, nanoreporters may be assembled and labeled in a variety of different ways. For example, a nanoreporter scaffold can be attached to a target-specific sequence (for example by hybridization and, optionally, ligation), and the structure comprising the scaffold and target-specific sequence attached to one or more patches and, where desired, flaps. Alternatively, the nanoreporter scaffold can first be attached to one or more patches (and, optionally, flaps), and the scaffold/patch structure then attached to a target specific sequence. Thus, unless stated otherwise, a discussion or listing of steps in nanoreporter assembly does not imply that a specific route of assembly must be followed.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 896

Nanoreporter assembly and use is exemplified herein largely by way of description of a variety of nucleic acid-based nanoreporters; however, one of skill in the art would recognize that the methods described herein are applicable to an amino acid-based (or hybrid nucleic acid-/amino acid-based) nanoreporter. Illustrative embodiments of partially and fully assembled nanoreporters are listed below.

At its simplest, the invention provides a scaffold having at least two label attachment regions capable of being labeled and resolved. The scaffold can be any molecular entity that allows the formation of label attachment regions on the scaffold that can be separately labeled and resolved. The number of label attachment regions to be formed on a scaffold is based on the length and nature of the scaffold, the means of labeling the nanoreporter, as well as the type of label monomers emitting a signal to be attached to the label attachment regions of the scaffold. A nanoreporter according to the invention may have a scaffold including two or more label attachment regions. Suitable scaffold structures include DNA-based scaffolds.

The invention also provides labeled nanoreporters wherein one or more label attachment regions are attached to corresponding label monomers, each label monomer emitting a signal. For example a labeled nanoreporter according to the invention is obtained when at least two label monomers are attached to two corresponding label attachment regions of the scaffold such that these labeled label attachment regions, or "spots," are distinguishable. Label monomers emitting a signal associated with different label attachment regions of the scaffold can emit signals that are spectrally indistinguishable under the detections conditions ("like" signals), or can emit signals that are spectrally distinguishable, at least under the detection conditions (*e.g.*, when the nanoreporter is immobilized, stretched and observed under a microscope).

The invention also provides a nanoreporter wherein two or more label monomers are attached to a label attachment region. The signal emitted by the label monomers associated with said label attachment region produces an aggregate signal that is detected. The aggregate signal produced may be made up of like signals or made up of at least two spectrally distinguishable signals.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 897

In one embodiment, the invention provides a nanoreporter wherein at least two label monomers emitting like signals are attached to two corresponding label attachment regions of the scaffold and said two label monomers are spatially distinguishable. In another embodiment, the invention provides a nanoreporter wherein at least two label monomers emitting two distinguishable signals are attached to two neighboring label attachment regions, for example two adjacent label attachment regions, whereby said at least two label monomers are spectrally distinguishable.

The invention provides a nanoreporter wherein two spots emitting like signals are separated by a spacer region, whereby interposing the spacer region allows resolution or better resolution of said like signals emitted by label monomers attached to said two spots. In one embodiment, the spacer regions have a length determined by the resolution of an instrument employed in detecting the nanoreporter.

The invention provides a nanoreporter with one or more "double spots." Each double spot contains two or more (*e.g.*, three, four or five) adjacent spots that emit like signals without being separated by a spacer region. Double spots can be identified by their sizes.

A label monomer emitting a signal according to the invention may be attached covalently or non-covalently (*e.g.*, via hybridization) to a patch that is attached to the label attachment region. The label monomers may also be attached covalently or non-covalently (*e.g.*, via hybridization) to a flap attached to a patch that is in turn attached to the scaffold. The flap can be formed by one molecule or two or more molecules ("flap pieces") that form a split flap.

The invention also provides a nanoreporter associated with a spectral code determined by the sequence of signals emitted by the label monomers attached (*e.g.*, indirectly via a patch) to label attachment regions on the scaffold of the nanoreporter, whereby detection of the spectral code allows identification of the nanoreporter.

In one embodiment, the invention provides a nanoreporter further comprising an affinity tag attached to the nanoreporter scaffold, such that attachment of the affinity tag to a support allows scaffold stretching and resolution of signals emitted by label monomers corresponding to different label attachment regions on the scaffold. Nanoreporter stretching

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 898

may involve any stretching means known in the art including but not limited to, means involving physical, hydrodynamic or electrical means.

In yet another embodiment, the invention provides a nanoreporter further comprising flaps attached to label attachment regions of the scaffold, wherein a flap attached to a label attachment region of the scaffold attaches the label monomer corresponding to said label attachment region, thereby indirectly attaching label monomers to corresponding label attachment regions on said scaffold. In a further embodiment, each label monomer comprises a signal emitting portion and an oligonucleotide portion of a predetermined sequence, and the flaps comprise repeats of a flap sequence complementary to the oligonucleotide portion of a corresponding label, whereby one or more label monomers attach to a corresponding label attachment region through hybridization of said oligonucleotide portions of said label monomers to said repeats of said flap sequence thereby producing a labeled nanoreporter.

A nanoreporter according to the invention can further include a target-specific sequence coupled to the scaffold. The target-specific sequence is selected to allow the nanoreporter to recognize, bind or attach to a target molecule. The nanoreporters of the invention are suitable for identification of target molecules of all types. For example, appropriate target-specific sequences can be coupled to the scaffold of the nanoreporter to allow detection of a target molecule. Preferably the target molecule is DNA (including cDNA), RNA (including mRNA and cRNA), a peptide, a polypeptide, or a protein.

One embodiment of the invention provides increased flexibility in target molecule detection with label monomers according to the invention. In this embodiment, a dual nanoreporter comprising two different molecular entities, each with a separate target-specific region, at least one of which is labeled, bind to the same target molecule. Thus, the target-specific sequences of the two components of the dual nanoreporter bind to different portions of a selected target molecule, whereby detection of the spectral code associated with the dual nanoreporter provides detection of the selected target molecule in a biomolecular sample contacted with said dual nanoreporter.

The invention also provides a method of detecting the presence of a specific target molecule in a biomolecular sample comprising: (i) contacting said sample with a dual nanoreporter under conditions that allow binding of the target-specific sequences in the dual

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 899

nanoreporter to the target molecule and (ii) detecting the spectral code associated with the dual nanoreporter. Depending on the nanoreporter architecture, the dual nanoreporter may be labeled before or after binding to the target molecule.

In certain embodiments, the methods of detection are performed in multiplex assays, whereby a plurality of target molecules are detected in the same assay (a single reaction mixture). In a preferred embodiment, the assay is a hybridization assay in which the plurality of target molecules are detected simultaneously. In certain embodiments, the plurality of target molecules detected in the same assay is at least 5 different target molecules, at least 10 different target molecules, at least 20 different target molecules, at least 50 different target molecules, at least 75 different target molecules, at least 100 different target molecules, at least 200 different target molecules, at least 500 different target molecules, or at least 750 different target molecules, or at least 1000 different target molecules. In other embodiments, the plurality of target molecules detected in the same assay is up to 50 different target molecules, up to 100 different target molecules, up to 150 different target molecules, up to 200 different target molecules, up to 300 different target molecules, up to 500 different target molecules, up to 750 different target molecules, up to 1000 different target molecules, up to 2000 different target molecules, or up to 5000 different target molecules. In yet other embodiments, the plurality of target molecules detected is any range in between the foregoing numbers of different target molecules, such as, but not limited to, from 20 to 50 different target molecules, from 50 to 200 different target molecules, from 100 to 1000 different target molecules, from 500 to 5000 different target molecules, and so on and so forth.

In certain embodiments, the invention is directed to detecting different splice forms of the same RNA. The different splice forms can be detected using a plurality of nanoreporter probes, each with a different target-specific sequence complementary to a different exon of the same gene.

Structural stability of a nanoreporter can be increased through ligation of the patches and, optionally, ligation of the split flaps and/or the labeled oligonucleotides hybridized to the split flaps.

In addition to the qualitative analytical capabilities provided by the nanoreporters of the invention and the analytical techniques based thereon, the nanoreporters

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 900

of the invention are uniquely suitable for conducting quantitative analyses. By providing a one to one binding between the nanoreporters (whether singular or dual nanoreporters) of the invention and their target molecules in a biomolecular sample, all or a representative portion of the target molecules present in the sample can be identified and counted. This individual counting of the various molecular species provides an accurate and direct method for determining the absolute or relative concentration of the target molecule in the biomolecular sample. Moreover, the ability to address each molecule in a mixture individually leverages benefits of miniaturization including high sensitivity, minimal sample quantity requirements, high reaction rates which are afforded by solution phase kinetics in a small volume, and ultimately very low reagent costs.

As will be appreciated from the description and examples provided below, the present invention provides numerous advantages. For example, the complex modularity in forming nanoreporters according to the invention allows for systematic creation of libraries of unique nanoreporters having a very high degree of diversity (*e.g.*, millions of uniquely recognizable nanoreporters). This modularity allows flexibility in customizing nanoreporter populations to specific applications which in turn provides significant manufacturing efficiencies. Another advantage that will be appreciated through the following description stems from the flexibility in assembling the nanoreporters of the invention. That is, due to their modular structure, the nanoreporters of the invention can be assembled prior to shipment to a point of use or assembled at the point of use.

NANOREPORTER NOMENCLATURE

NANOREPORTER: The term "nanoreporter" refers to a molecular entity that has (i) a molecule ("scaffold") containing at least two label attachment regions; (ii) at least one patch attached to at least one label attachment region; and (iii) a target-specific sequence. As described in detail below, nanoreporters can be singular nanoreporters (all components being in a single molecular entity) or dual nanoreporters (all the components being in two separate molecular entities). Nanoreporters are preferably synthetic, *i.e.*, non-naturally-occurring molecules, for example are chimeric molecules made by joining two or more

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 901

manmade and/or naturally occurring sequences that normally exist on more than one molecule (*e.g.*, plasmid, chromosome, viral genome, protein, etc.).

LABELED NANOREPORTER: A labeled nanoreporter is a nanoreporter in which at least one patch of the nanoreporter is attached to one or more label monomers that generate(s) a signal that forms at least part of the nanoreporter code.

LABEL UNIT: The term "label unit" refers to the non-target-specific portions of a labeled nanoreporter.

PROBE: This refers to a molecule that has a target-specific sequence. In the context of a singular nanoreporter, the term "probe" refers to the nanoreporter itself; in the context of a dual nanoreporter, the term "probe" refers to one or both of the two components of the nanoreporter.

PROBE PAIR: This refers to a dual nanoreporter.

PATCH: The term "patch" refers to a molecular entity attached to the label attachment region of the nanoreporter scaffold, generally for the purpose of labeling the nanoreporter. The patch can have one or more label monomers either directly (covalently or noncovalently) or indirectly attached to it, either prior to or after its attachment to the scaffold.

FLAP: The term "flap" as used herein refers to a molecular entity attached to a patch or patch pair attached to a label attachment region. The flap is one or more molecule containing label monomers or capable of binding one or more molecules containing label monomers. By providing indirect labeling of the regions, the flaps provide more flexibility in controlling the number of signal emitting monomers associated with a region as well as the nature of those monomers. Flaps may be formed by a single molecular piece or several molecular pieces (*e.g.*, two pieces) forming a "split flap" (see, *e.g.*, Figure 7)

TARGET-SPECIFIC SEQUENCE: The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter scaffold. The target molecule is preferably (but not necessarily) a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 902

GHOST PROBE: A molecule comprising a target-specific sequence, but which is not labeled with a label monomer that emits a signal that contributes to the nanoreporter code.

REPORTER PROBE: A molecule comprising a target-specific sequence that is labeled with at least one label monomer that emits a signal that contributes to the nanoreporter code. A singular nanoreporter is a reporter probe, as is a labeled component of a dual nanoreporter.

F-HOOK and G-HOOK: In the context of a dual nanoreporter, F- and G-hooks are each an affinity tag that is capable of being selectively bound to one of the probes. In preferred embodiments, the F-hook and G-hook are biotinylated oligonucleotides that are hybridizable to respective complementary sequences present in (*e.g.*, via ligation) or attached to (*e.g.*, via hybridization) the respective nanoreporter probes in a dual nanoreporter. Thus, the F-hooks and G-hooks can be used for purification, immobilization and stretching of the nanoreporter. Generally, where a dual nanoreporter contains one reporter probe and one ghost probe, the G-hook becomes attached to the reporter probe and the F-hook becomes attached to the ghost probe. F-hooks and G-hooks can be biotinylated on either end or internally. They can also be amine-modified to allow for attachment to a solid substrate for affinity purification.

F-TAG and G-TAG: Tandemly-repeated sequences of about 10 to about 25 nucleotides that are complementary to the F-hook and G-hook, respectively. G-tags and F-tags are attached to the nanoreporter probes. Generally, an F-tag is present in or attached to a ghost probe via a ligator sequence and a G-tag is present in or attached the reporter probe scaffold via a ligator sequence.

SPOT: A spot, in the context of nanoreporter detection, is the aggregate signal detected from the label monomers attached to a single label attachment site on a nanoreporter, and which, depending on the size of the label attachment region and the nature (*e.g.*, primary emission wavelength) of the label monomer, may appear as a single point source of light when visualized under a microscope. Spots from a nanoreporter may be overlapping or non-overlapping. The nanoreporter code that identifies that target molecule can comprise any permutation of the length of a spot, its position relative to other spots,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 903

and/or the nature (*e.g.*, primary emission wavelength(s)) of its signal. Generally, for each probe or probe pair of the invention, adjacent label attachment regions are non-overlapping, and/or the spots from adjacent label attachment regions are spatially and/or spectrally distinguishable, at least under the detection conditions (*e.g.*, when the nanoreporter is immobilized, stretched and observed under a microscope, as described herein).

Occasionally, reference is made to a spot "size" as a certain number of bases or nucleotides. As would be readily understood by one of skill in the art, this refers to the number of bases or nucleotides in the corresponding label attachment region.

NANOREPORTER CODE: The order and nature (*e.g.*, primary emission wavelength(s), optionally also length) of spots from a nanoreporter serve as a nanoreporter code that identifies the target molecule capable of being bound by the nanoreporter through the nanoreporter's target specific sequence(s). When the nanoreporter is bound to a target molecule, the nanoreporter code also identifies the target molecule. Optionally, the length of a spot can be a component of the nanoreporter code.

DARK SPOT: The term "dark spot" refers to a lack of signal, or "spot," from a label attachment site on a nanoreporter. Dark spots can be incorporated into the nanoreporter code to add more coding permutations and generate greater nanoreporter diversity in a nanoreporter population.

REGISTER: The term "register" refers to a set of alternating label attachment regions.

THE NANOREPORTER SCAFFOLD

The nanoreporter scaffold can be any molecular entity, more preferably a nucleic acid molecule, containing label attachment regions to which label monomers can be directly or indirectly attached. In one embodiment, the nanoreporter scaffold is a protein scaffold; in a preferred embodiment, the nanoreporter scaffold is a nucleic acid scaffold in which the label attachment regions are single-stranded regions to which other nucleic acids, such as oligonucleotide patches, RNA patches, or DNA patches, can attach by hybridization. In specific embodiments, the nanoreporter scaffold is a nucleic acid molecule.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 904

There are no particular limitations on the types of scaffolds that are suitable for forming the nanoreporters of the invention. A scaffold according to the invention can essentially have any structure including, for example, single stranded linear scaffold, double stranded linear scaffold, single stranded circular scaffold or double stranded circular scaffold. Examples of scaffold structures include, for example, a scaffold made of one molecular entity such as polypeptides, nucleic acids or carbohydrates. A scaffold may also include a combination of structures, for example, a scaffold may be made of one or more polypeptide stretches coupled to one or more carbohydrate stretches.

Suitable molecular entities for scaffolds according to the invention include polymeric structures particularly nucleic acid based polymeric structures such as DNA. DNA based structures offer numerous advantages in the context of the present invention due at least in part to the vast universe of existing techniques and methodologies that allow manipulation of DNA constructs.

As indicated above, the scaffold may be single stranded or double stranded. Double stranded scaffold can be either conventional double stranded DNA or a double strand that is composed of a linear single stranded stretch of nucleic acid with patch units or flat-patches attached.

A scaffold can have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21-100 label attachment regions or more.

The label attachment regions of a nanoreporter scaffold will vary in size depending on the method of labeling. In various embodiments, a label attachment region can have a length anywhere from 10 nm to 10,000 nm, but is more preferably from 50 nm to 5,000 nm, and is more preferably from 100 nm to 1,000 nm. In various embodiments, the label attachment region is from about 100 nm to about 500 nm, from about 150 nm to about 450 nm, from about 200 nm to about 400 nm, or from 250 to about 350 nm. In a preferred embodiment, the label attachment region corresponds closely to the size of a diffraction-limited spot, *i.e.*, the smallest spot that can be detected with standard optics, which is about 300 nm.

Where the scaffold is a nucleic acid, 1 nm corresponds to approximately 3 nucleotides; thus, an approximately 300 nm-label attachment region corresponds to

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 905

approximately 900 bases. In other preferred embodiments, the label attachment region is from about 300 nucleotides to about 1.5kb, from about 450 nucleotides to about 1.35 kb, from about 0.6 kb to about 1.2 kb, or from 0.75 kb to about 1.05 kb.

An illustrative example of a molecular entity for a nanoreporter scaffold according to the invention is M13 DNA, which is single-stranded. In one embodiment, the nanoreporter scaffold is circular at least partially single stranded DNA, such as circular M13. In a more preferred embodiment, the nanoreporter scaffold is linear at least partially single stranded DNA, such as linear M13. In a specific embodiment, the M13 single-stranded DNA obtained by operating a cut at the BamH1 site of circular M13 DNA.

It should be noted that within the context of the present invention, linear DNA provides additional advantages compared to circular DNA. One advantage of using linear DNA in forming a scaffold according to the invention relates to the significantly reduced torsional stress associated with linear DNA. The added torsional stress associated with circular DNA may interfere with the structural integrity of the scaffold upon the addition to the scaffold of other components of the nanoreporter, such as patch units. Severe torsional stress may lead to the breaking of the structure of the scaffold. It should be noted however that the nanoreporters where only a few, short label attachment sites are labeled, circular DNA may be suitable.

NOVEL SYNTHETIC NANOREPORTER SCAFFOLD SEQUENCES

The present invention provides nanoreporter scaffold that are artificial nucleic acid molecules (DNA, RNA, or DNA/RNA hybrids) designed to have features that optimize labeling and detection of the nanoreporter. In these aspects of the invention, a nanoreporter scaffold is an artificial nucleic acids comprising one or more synthetic sequences from 50 to 50,000 bases long. Accordingly, the nanoreporter scaffold, which is preferably a DNA, is designed to have one or more Regions, useful as label attachment regions, comprising a regular pattern of a particular base (the "regularly-repeated base"). In such regions, the regularly-repeated base occurs with a periodicity of every nth residue, where n is any number, and preferably from 4 to 25.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 906

Preferably, not more than 25% of the regularly-repeated base in a Region appears at other than said regular intervals. For example, if in a Region of 100 nucleotides there are 12 thymidine bases, and thymidine is the regularly-repeated base, in this aspect of the invention not more than 25% of these, *i.e.*, 3 thymidine bases, appear outside the regular pattern of thymidines. In specific embodiments, not more than 20%, not more than 15%, not more than 10%, not more than 9%, not more than 8%, not more than 7%, not more than 6%, not more than 5%, not more than 4%, not more than 3%, not more than 2% or not more than 1% of said base appears at other than said regular intervals in said region.

The regularly-repeated base in the Regions in a nanoreporter scaffold, or its complementary regularly-repeated base in an annealed patch (or segment) can be used to attach label monomers, preferably light emitting label monomers, to the nanoreporter in a regular, evenly spaced pattern for better distribution of the nanoreporter signal. Preferably, where a Region is labeled, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95% or at least 98% of occurrences of the regularly-repeated base is attached to at least one light-emitting label monomer, either by covalent attachment of a label monomer to a base, or by hybridization to a nucleic acid in which the complements of the regularly-repeated base are so-labeled.

This percentage of occurrences can be measured by any means known in the art. In one method, the amount of nucleic acid produced in a labeling reaction is purified (for example, RNA can be purified using a Qiagen RNeasy kit) and subjected to UV spectrophotometry. The absorbance ("A") at the appropriate wavelengths is measured for each of the nucleic acid (260 nm) and the label monomer whose occurrence is to be measured (*e.g.*, 495 nm for Alexa Fluor 488; 590 nm for Alexa Fluor 594; 650 for Alexa Fluor 647; and 550 nm for Cy3). The absorbance of the nucleic acid is corrected by adjusting the value of the absorbance at 260 nm ("A260") to remove the "noise" contribution from the label monomer by subtracting the absorbance at the peak wavelength for the label monomer (A_{LM}) minus the correction factor for that label monomer. Where the nucleic acid is RNA, the number of label monomers per one thousand nucleotides is calculated according to the formula:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 907

$$\frac{\text{no. of label monomers}}{1000\,\text{nucleotides}} = \frac{A260}{A_{LM}} \times \frac{9010}{EC_{LM}} \times 1000$$

where EC_{LM} is the extinction coefficient for the label monomer. From this formula, the percentage of occurrences of the regularly-repeated base that are attached to a light-emitting label monomer can be calculated.

Generally, the preferred regularly-repeating base in a label attachment region is thymidine, so that the region can be labeled by hybridization to one or more complementary patches (*e.g.*, RNA segments) in which the regularly-repeated base is uridine. This permits the use of amino-allyl-modified UTPs, which are readily commercially available, as label monomer attachment sites, in an otherwise random sequence. Preferably, in addition to the regular periodicity of the Regions, the regions (and the nucleic acid comprising them) contain minimal secondary structure. The overall GC-content is preferably maintained close to 50%, and is preferably consistent over relatively short stretches to make local Tm's similar.

The artificial nucleic acids of the invention, or at least the Regions therein, preferably do not have direct or inverted repeats that are greater than 12 bases in length. In other embodiments, the artificial nucleic acids and/or Regions do not have direct or inverted repeats that are greater than about 11, about 10 or about 9 bases in length.

In an exemplary Region in which the regularly-repeated nucleotide is a thymidine and a GC content of approximately 50%, excess adenines would make up the loss in abundance of T's. To generate the selected sequence, random sequences with fixed patterns of T's ranging from every 4th base to every 25th base are created and screened to minimize the presence of inverted and direct repeats.

Sequences are also screened preferably to avoid common six-base-cutter restriction enzyme recognition sites. Selected sequences are additionally subjected to predicted secondary structure analysis, and those with the least secondary structure are chosen for further evaluation. Any program known in the art can be used to predict secondary structure, such as the MFOLD program (Zuker, 2003, Nucleic Acids Res. 31 (13):3406-15; Mathews *et al.*, 1999, J. Mol. Biol. 288:911-940).

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 908

An appropriate sequence is divided into label attachment regions ranging from 50 bases to 2 kilobases long (could be longer). Each label attachment region is a unique sequence, but contains a consistent number and spacing of T's in relation to the other label attachment regions in a given reporter sequence. These label attachment regions can interspersed with other regions whose sequence does not matter. The synthetic label attachment regions in a nanoreporter scaffold can be of different lengths and/or have different regularly-repeated bases. An optimized start sequence for transcription by RNA polymerase T7, T3, or SP6 (beginning at position +1 of the transcript) can be added to the 5′ end of each label attachment region. Restriction sites are optionally added at the boundaries of each label attachment region to allow specific addition or deletion of individual label attachment regions to the sequence using conventional cloning techniques. The number of synthetic label attachment regions in a nanoreporter preferably ranges from 1 to 50. In yet other embodiments, the number of synthetic label attachment regions in a nanoreporter ranges from 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 synthetic label attachment regions to 15, 20, 30, 40, or 50 synthetic label attachment regions, or any range in between.

An example of such a novel synthetic label attachment region is given below. In this sequence, shown 5′ to 3′, the T's are placed in every 8th position and the region is bounded by a 5′ Sac I restriction site and a 3′ Kpn I restriction site. An optimized transcript start site for T7 polymerase (GGGAGA) is included at the 5′ end of the region, downstream of the 5′ restriction site. The complement of this sequence, when generated as a single-stranded molecule, forms the scaffold for the RNA molecule transcribed from this label attachment region.

GAGCTCGGGAGATGGCGAGCTGGAAGCATCAGAAAGTAGGAAGATGACAAA

ATAGGGCCATAGAAGCATGAAGAACTGAACGCATGAGACAATAGGAAGCTA

CGCCACTAGGGACCTGAGAAGCTGAGCGGCTCAGCGGGTCCGAGCGTCAAAA

AATAAAAGAGTGAAACAATAGACGAATGACGCGGTAAAACCATCCAGAAGT

AAACGGGTACAAACATACAGAGATAGCCACCTGGACCAATAGGCACGTACA

AACGTACAAGCCTGGCGCGATGAGGCAATCCACACGTGCAGAGCTGGAACA

ATGGAAAGATGCAAGAATAAACCGATACCGGGATCGAGGGCTCAGCGAATA

AAGCAGTCAACAACTGGAAAGATCCACACATACCGGCGTAACCGAGTCCAAA

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 909

CATACAGACCTGCAAGACTCGCGACATGGGACGGTAAAACCATCCGACCGTA
AACCGGTAACCAGGTAGCCGGGTAAAAACATAGCAGGGTGGAGACCTCAGA
ACGTAAAGACGTCCAAGGGTCGCCGGATAGCGAACTACGCGCATCGCCCAAT
GGGCCAATCAACAGATAAACGAGTAGAAAAGTCAGAAAATAAGAAACTAAC
GAAATACGAGGGTCCAAGGATGCAAGACTGAGGCCCTAAGGAGATAAGGAA
ATAGGCCGATGCAGACCTGAAACGATGCACCGATCCGACGGTAAAAGACTAG
ACACGTAGCCGGATCAGGGCCTGGGAGGCTGGAACCGTGAGCACATAGCAA
AGTCGCAGCGTCGGCAGATGCGCCGGTAAAAAAGTAGAGGCATGACCGGAT
GGGCAAATAGCGACGTACAGCAGTGAAGCACTAAAAGCATCCAAGGGTAGG
AGACTAGGCGCCTCGACGGGTAGGTACC

The synthetic nucleic acids of the present invention can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the label attachment region and the annealed patches or segments, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the synthetic nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5′-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl)uracil, (acp3)w, and 2,6-diaminopurine.

Alternatively, the synthetic nucleic acid can be produced biologically using a vector into which a nucleic acid has been subcloned.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 910

In various embodiments, the synthetic nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.*, 1996, Bioorganic & Medicinal Chemistry 4(l):5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.*, 1996, Bioorganic & Medicinal Chemistry 4(1): 5-23; Perry-O'Keefe *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93: 14670-675.

In an exemplary embodiment, the selected novel synthetic sequence can be constructed synthetically as double-stranded DNA by a commercial gene synthesis company and cloned in an oriented fashion into a "phagemid", a plasmid vector containing an M13 or f1 phage intergenic (IG) region which contains the *cis*-acting sequences necessary for DNA replication and phage encapsidation, such as pUC119. The appropriate orientation of the cloned insert relative to the phage origin of replication allows for the generation of a single-stranded DNA scaffold which is the reverse complement of the RNA molecules generated by *in vitro* transcription for each label attachment region.

In order to generate the single-stranded DNA scaffold of the novel reporter, the phagemid is transformed into an *E. coli* strain containing an F' episome. Subsequent infection of the transformed bacteria with a helper phage such as the M13 mutant K07 results in the secretion of the phagemid carrying the novel reporter sequence as a single-stranded, packaged phage from which the circular, single-stranded DNA is prepared using a standard protocol. This DNA is linearized and the vector portion is excised by annealing short, complementary oligonucleotides to either end of the novel reporter sequence to generate double-stranded restriction sites, followed by treatment with the appropriate restriction enzymes.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 911

To make the RNA molecules (patches or "segments") for each label attachment region, polymerase chain reaction ("PCR") primers are designed to generate a double-stranded template beginning with an RNA polymerase promoter (T7, T3, or SP6) directly upstream (5′) of the transcription start site and ending following the 3′ restriction enzyme site. Using this template, *in vitro* transcription of RNA molecules is performed in the presence of amino-allyl modified regularly-repeated base in the RNA (*e.g.*, UTP) and unmodified other bases (*e.g.*, ATP, CTP and GTP). This leads to an RNA product in which every regularly-repeated base (*e.g.*, U) is modified to allow covalent coupling of a label monomer at that position in the RNA molecule.

Coupling of light-emitting label monomers to the RNA molecules and annealing of the labeled RNA molecules to the scaffold are carried out as described below.

Some design considerations for the de novo sequence are listed in Table 1 below.

Feature Of Synthetic Scaffold	**Advantages**
Novel synthetic sequence	Can be of any length and be designed to incorporate any desired sequence feature including but not limited to those listed in this table.
Minimal secondary structure (select against inverted repeats)	Allows for consistent transcription of full-length RNA molecules. Allows for consistent annealing of RNA molecules to scaffold at predictable temperatures. Minimizes self-annealing and/or cross-annealing between RNA molecules or scaffolds.
Minimal repeated sequences	Avoids mis-annealing between RNA molecules and inappropriate regions of the scaffold.
Unique restriction sites at borders of label attachment regions	Allows addition and deletion of individual label attachment regions using standard molecular cloning techniques.
Defined, even spacing of T's and transcription with amino-allyl-modified UTP (no unmodified UTP)	Controls number of coupling sites for monomers in each label attachment region, allowing for consistent brightness of individual labeled RNA molecules. Controls distance between monomers: spacing can be optimized to avoid stearic hindrance and fluorescence quenching.
Optimized start sequence for transcription by RNA polymerase T7, T3, or SP6	Promotes efficient *in vitro* transcription of each label attachment region.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 912

PATCHES

Label monomers that emit signals which constitute all or part of the nanoreporter code are attached to label attachment region(s) of the nanoreporter scaffold through a structure referred to herein as a "patch." The label monomers can be directly (*e.g.*, covalently or noncovalently) attached to a patch, or indirectly attached to a patch (*e.g.*, through hybridization).

Nucleic acid patches can by anywhere from 25 nucleotides to several kilobases (*e.g.*, 5kb) in length, and are preferably 50 nucleotides to 2kb in length. In specific embodiments, nucleic acid patches are approximately 25 to 250, 50 to 200, 50 to 150, or 50 to 100 nucleotides in length. In other embodiments, nucleic acid patches are approximately 500 to 2,000, 500 to 1,500, 500 to 1,000, 750 to 1,250, or 750 to 1,000 nucleotides in length. Nucleic acid patches can be RNA patches or DNA patches.

A label monomer can be covalently attached to a patch before or after the patch is attached to the label attachment region of a nanoreporter scaffold. For example, where the patch is a nucleic acid molecule, the label can be covalently attached by incorporation of a nucleotide containing a label monomer into the nucleic acid during its synthesis but before it is attached, *e.g.*, via hybridization, to the label attachment region of the scaffold. Alternatively, during the synthesis of a nucleic acid patch, a nucleotide containing a label monomer acceptor group can be included, and the label monomer added to the nucleic acid patch after its synthesis, either before or after it is attached to the label attachment region of the scaffold. Alternatively, the label monomer can be indirectly attached to the patch, for example by hybridization of the patch to a "flap" that serves as a basis for attachment of the label monomer to the nanoreporter.

Thus, where a patch is a nucleic acid, it can range anywhere from 20 nucleotides to more than 5 kb in length, depending on the method of assembly of the nanoreporter.

For example, where a patch has covalently incorporated into it one or more label monomers that emit signals that are part of the nanoreporter code in the context of the labeled nanoreporter, the patch is preferably about 100 to about 10,000 bases, more preferably 200 to about 2000 bases, and yet more preferably 700 to about 1200 nucleotides in

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 913

length, and is generally referred to herein as a "segment," a "dark segment" being the patch prior to the incorporation of the label monomer (but, in a preferred embodiment, containing label monomer acceptor sites, such as amino allyl nucleotides), and a "colored" segment being one containing the desired label monomer or label monomers. The Tm of a segment when hybridized to its label attachment region preferably is >80°C, more preferably >90°C, in 825 mM Na^+ (5X SSC).

Where a patch merely serves as a template for flap attachment to the nanoreporter, then it is preferably smaller in size, for example about 25-250 nucleotides in length, and is most preferably about 50-100 nucleotides in length. Such patches are referred to herein as "oligonucleotide patches." As detailed in Section below, an oligonucleotide is preferably partially complimentary in sequence to a scaffold, such that when it is annealed to the scaffold, an overhang is generated that is complementary to all or a portion of a flap.

The terms "segment" and "oligonucleotide patch" are used herein merely for convenience of description; however, there is no size cutoff to distinguish a "segment" from an "oligonucleotide patch." The purpose of both types of structures is to maximize the labeling --and thus signal intensity-- from the nanoreporter, thereby allowing for single target molecule detection by a nanoreporter.

In certain aspects, the present invention provides a synthetic molecule, whose configuration is illustrated by reference to **Figure 7A**, comprising a strand of a nucleic acid (**scaffold**) and a plurality of patch pairs hybridized to said strand, wherein each patch pair comprises an "A" patch and a "B" patch, wherein, for each patch pair, (a) each "A" patch is an oligonucleotide comprising a first region (**1P**) and a second region (**2P**), said first region being (i) at the alpha end of said "A" patch, and (ii) hybridized to a first portion of said strand, said second region being (ii) at the beta end of said "A" patch; (b) each "B" patch is an oligonucleotide comprising a third region (**3P**) and a fourth region (**4P**), said third region being (i) at the alpha end of said "B" patch, and (ii) hybridized to said second region of said "A" patch, said fourth region being (i) at the beta end of said "B" patch and (ii) hybridized to a second portion of said strand, said second portion of said strand being to the beta end of said first portion of said strand, wherein said second region or said third region further

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 914

comprises at its beta end or alpha end, respectively, a hybridizable region that is not hybridized to said "B" patch or "A" patch, respectively.

In the synthetic molecule of **Figure 7A**, the second region may further comprise at its beta end a hybridizable region that is not hybridized to said "B" patch, as depicted in **Figure 7B**, or the third region further comprises at its alpha end a hybridizable region that is not hybridized to said "A" patch, as depicted in **Figure 7C**.

The present invention further provides a synthetic molecule, whose configuration is illustrated by reference to **Figure 7D**, comprising a strand of a nucleic acid (**scaffold**) and a plurality of patch pairs hybridized to said strand, wherein each patch pair comprises an "A" patch and a "B" patch, wherein, for each patch pair, (a) each "A" patch is an oligonucleotide comprising a first region (**1P**) and a second region (**2P**), said first region being (i) at the alpha end of said "A" patch, and (ii) hybridized to a first portion of said strand, said second region being (ii) at the beta end of said "A" patch; (b) each "B" patch is an oligonucleotide comprising a third region (**3P**) and a fourth region (**4P**), said third region being (i) at the alpha end of said "B" patch, and (ii) hybridized to said second region of said "A" patch, said fourth region being (i) at the beta end of said "B" patch and (ii) hybridized to a second portion of said strand, said second portion of said strand being to the first of said first portion of said strand, wherein said second region further comprises at its beta end a first hybridizable region that is not hybridized to said "B" patch, and wherein said third region further comprises at its alpha end a second hybridizable region that is not hybridized to said "A" patch.

In the synthetic molecule of **Figure 7B**, each patch pair can be attached to a flap pair, as depicted in **Figure 7F**, wherein each flap pair comprises an "A" flap and a "B" flap, wherein, for each flap pair, (a) each "A" flap is an oligonucleotide comprising a first flap region (**1F**) and a second flap region (**2F**); said first flap region being at the alpha end of said "A" flap; said second flap region (i) being at the beta end of said "A" flap and (ii) comprising at its beta end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "B" flap; and (b) each "B" flap is an oligonucleotide comprising a third flap region (**3F**), a fourth flap region (**4F**), and a fifth flap region (**5F**); said third flap region (i) being at the alpha end of said "B" flap and (ii) comprising at its alpha end a hybridizable

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 915

region that is not hybridized to said "A" patch, "B" patch or "A" flap; said fourth flap region (i) being between the third flap region and the fifth flap region and (ii) hybridized to said first flap region of said "A" flap; said fifth flap region being (i) at the beta end of said "B" flap, and (ii) hybridized to said hybridizable region of said second region of said "A" patch.

In the synthetic molecule of **Figure 7C**, each patch pair can be attached to a flap pair, as depicted in **Figure 7E**, wherein each flap pair comprises an "A" flap and a "B" flap, wherein, for each flap pair, (a) each "A" flap is an oligonucleotide comprising a first flap region (**1F**), a second flap region (**2F**), and a third flap region (**3F**); said "A" flap region being (i) at the alpha end of said "A" flap and (ii) hybridized to said hybridizable region of said third region of said "B" patch; said second flap region being between the first flap region and the third flap region; said third flap region (i) being at the beta end of said "A" flap and (ii) comprising at its beta end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "B" flap, and (b) each "B" flap is an oligonucleotide comprising a fourth flap region (**4F**) and a fifth flap region (**5F**); said fourth flap region being (i) being at the alpha end of said "B" flap and (ii) comprising at its alpha end a hybridizable region that is not hybridized to said "A" patch, "B" patch or "A" flap; said fifth flap region being (i) at the beta end of said "B" flap, and (ii) hybridized to said second flap region of said "A" flap.

In the synthetic molecule of **Figures 7D** and **7E**, the split flaps can be attached one (*e.g.*, (**1O**)), or more (*e.g.*, (**2O**) and (**3O**)) oligonucleotides, as depicted in **Figure 7G**. Thus, the one or more oligonucleotides can be attached to the all or a portion of the "A" flap individually (*e.g.*, (**1O**)), the "B" flap individually (*e.g.*, (**3O**)), or span all or a portion of each of the "A" flap and "B" flap (*e.g.*, (**2O**)). Such oligonucleotides are preferably covalently bound to one or more label monomers.

The hybridizable regions of said synthetic molecules may be hybridized to a plurality of oligonucleotides, each bound, preferably covalently bound, to at least one label monomer, more preferably to at least five label monomers. In certain embodiments, all the oligonucleotides attached to a single patch pair comprise the same label monomers, *e.g.*, comprise label monomers that emit light at the same wavelength(s); in specific embodiments, all the oligonucleotides attached to at least two, or at least four, adjacent patch pairs

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 916

preferably comprise the same label monomers. One or more of the oligonucleotides may be bound to at least one affinity tag.

In certain preferred embodiments, the label monomers are fluorophores or quantum dots.

In the synthetic molecule described above, alpha can refers to either 5′ or 3′, and the corresponding beta to either 3′ or 5′, respectively.

The region of complementary in each patch pair, or between a given patch and corresponding flap, is preferably about 20 to 5,000 nucleotides. In certain embodiments, the region of complementary is about 20 to 100 nucleotides, or about 5 to 50 nucleotides.

In the synthetic molecules described above, each flap is preferably about 50 to 5,000 nucleotides in length. In certain embodiments, each flap is about 50 to 150 nucleotides.

The synthetic molecules described above may further comprise a target-specific region which binds to a target molecule. The target-specific region can be attached to the beta or alpha end of said strand.

In certain embodiments, the synthetic molecule described above may comprise at least ten patch pairs, or at least fifty patch pairs.

In the synthetic molecules described above, the strand, or scaffold, can be a linearized vector, such as linearized M13.

The synthetic molecule described above may further comprise (a) a first label attachment region to which are attached (directly or indirectly) one or more label monomers that emit light constituting a first signal; (b) a second label attachment region, which is non-overlapping with the first label attachment region, to which is attached one or more label monomers that emit light constituting a second signal; (c) a third label attachment region, which is non-overlapping with the first and second label attachment regions, to which is attached one or more label monomers that emit light constituting a third signal; wherein each attachment region comprises a plurality of patch pairs; wherein the first and second signals are spectrally distinguishable; wherein the second and third signals are spectrally distinguishable; wherein the first and second signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the second

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 917

and third signals are not spatially resolvable under conditions that can be used to detect said first, second and third signals; wherein the first and third signals are spatially resolvable under conditions that can be used to detect said first, second and third signals; and wherein the identities of the first, second and third signals and the locations of the first and third signal relative to each other constitute at least part of a code that identifies the target molecule.

LABEL MONOMERS

The nanoreporters of the present invention can be labeled with any of a variety of label monomers, such as a radioisotope, fluorochrome, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other monomer known in the art that can be detected directly (*e.g.*, by light emission) or indirectly (*e.g.*, by binding of a fluorescently-labeled antibody). Generally, one or more of the label attachment regions in the nanoreporter is labeled with one or more label monomers, and the signals emitted by the label monomers attached to the label attachment regions of a nanoreporter constitute a code that identifies the target to which the target-specific region of the nanoreporter binds. In certain embodiments, the lack of a given signal from the label attachment region (*i.e.*, a "dark" spot) can also constitute part of the nanoreporter code. An example of a dark spot is depicted at position 12 of the nanoreporter in Figure 1A.

Radioisotopes are an example of label monomers that can be utilized by the invention. Several radioisotopes can be used as label monomers for labeling nucleotides or proteins, including, for example, ^{32}P, ^{33}P, ^{35}S, ^{3}H, and ^{125}I. These radioisotopes have different half-lives, types of decay, and levels of energy which can be tailored to match the needs of a particular experiment. For example, ^{3}H is a low energy emitter which results in low background levels, however this low energy also results in long time periods for autoradiography. Radioactively labeled ribonucleotides, deoxyribonucleotides and amino acids are commercially available. Nucleotides are available that are radioactively labeled at the first, or α, phosphate group, or the third, or γ, phosphate group. For example, both [α - ^{32}P] dATP and [γ - ^{32}P] dATP are commercially available. In addition, different specific activities for radioactively labeled nucleotides are also available commercially and can be tailored for different experiments.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 918

Another example of label monomers that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, for example, fluorescein, tetramethylrhodamine, and Texas Red. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum. Also, different formulations of the same fluorophore have been produced for different applications. For example, fluorescein, can be used in its isothiocynanate form (FITC), as mixed isomer or single isomer forms of carboxyfluorescein succinimidyl ester (FAM), or as isomeric dichlorotriazine forms of fluorescein (DTAF). These monomers are chemically distinct, but all emit light with a peak between 515-520 nm, thereby generating a similar signal. In addition to the chemical modifications of fluorescein, completely different fluorophores have been synthesized that have the same or very similar emission peaks as fluorescein. For example, the Oregon Green dye has virtually superimposable excitation and emission spectra compared to fluorescein. Other fluorophores such as Rhodol Green and Rhodamine Green are only slightly shifted in their emission peaks and so also serve functionally as substitutes for fluorescein. In addition, different formulations or related dyes have been developed around other fluorophores that emit light in other parts of the spectrum.

Non-radioactive and non-fluorescent label monomers are also available. For example, biotin can be attached directly to nucleotides and detected by specific and high affinity binding to avidin or streptavidin which has been chemically coupled to an enzyme catalyzing a colorimetric reaction (such as phosphatase, luciferase, or peroxidase). Digoxigenin labeled nucleotides can also similarly be used for non-isotopic detection of nucleic acids. Biotinylated and digoxigenin-labeled nucleotides are commercially available.

Very small particles, termed nanoparticles, also can be used as label monomers to label nucleic acids. These particles range from 1-1000 nm in size and include diverse chemical structures such as gold and silver particles and quantum dots.

When irradiated with angled incident white light, silver or gold nanoparticles ranging from 40-120 nm will scatter monochromatic light with high intensity. The wavelength of the scattered light is dependent on the size of the particle. Four to five different particles in close proximity will each scatter monochromatic light which when superimposed will give a specific, unique color. The particles are being manufactured by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 919

companies such as Genicon Sciences. Derivatized silver or gold particles can be attached to a broad array of molecules including, proteins, antibodies, small molecules, receptor ligands, and nucleic acids. For example, the surface of the particle can be chemically derivatized to allow attachment to a nucleotide.

Another type of nanoparticle that can be used as a label monomer are quantum dots. Quantum dots are fluorescing crystals 1-5 nm in diameter that are excitable by a large range of wavelengths of light. These crystals emit light, such as monochromatic light, with a wavelength dependent on their chemical composition and size. Quantum dots such as CdSe, ZnSe, InP, or InAs possess unique optical properties.

Many dozens of classes of particles can be created according to the number of size classes of the quantum dot crystals. The size classes of the crystals are created either 1) by tight control of crystal formation parameters to create each desired size class of particle, or 2) by creation of batches of crystals under loosely controlled crystal formation parameters, followed by sorting according to desired size and/or emission wavelengths. Use of quantum dots for labeling particles, in the context of the present invention, is new, but is old in the art of semiconductors. Two examples of earlier references in which quantum dots are embedded within intrinsic silicon epitaxial layers of semiconductor light emitting/detecting devices are United States Patent Nos. 5,293,050 and 5,354,707 to Chapple Sokol, *et al.*

In specific embodiments, one or more of the label attachments regions in the nanoreporter is labeled with one or more light-emitting dyes, each label attachment region containing, directly or indirectly, one or more label monomers. The light emitted by the dyes can be visible light or invisible light, such as ultraviolet or infrared light. In exemplary embodiments, the dye is a fluorescence resonance energy transfer (FRET) dye; a xanthene dye, such as fluorescein and rhodamine; a dye that has an amino group in the alpha or beta position (such as a naphthylamine dye, 1-dimethylaminonaphthyl-5-sulfonate, 1-anilino-8-naphthalende sulfonate and 2-p-touidinyl-6-naphthalene sulfonate); a dye that has 3-phenyl-7-isocyanatocoumarin; an acridine, such as 9-isothiocyanatoacridine and acridine orange; a pyrene, a bensoxadiazole and a stilbene; a dye that has 3-(ε-carboxypentyl)-3′-ethyl-5,5′-dimethyloxacarbocyanine (CYA); 6-carboxy fluorescein (FAM); 5&6-carboxyrhodamine-110 (R110); 6-carboxyrhodamine-6G (R6G); N,N,N′,N′-tetramethyl-6-carboxyrhodamine

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 920

(TAMRA); 6-carboxy-X-rhodamine (ROX); 6-carboxy-4′,5′-dichloro-2′,7′-dimethoxyfluorescein (JOE); ALEXA Fluor™; Cy2; Texas Red and Rhodamine Red; 6-carboxy-2′,4,7,7′-tetrachlorofluorescein (TET); 6-carboxy-2′,4,4′,5′,7,7′-hexachlorofluorescein (HEX); 5-carboxy-2′,4′,5′,7′-tetrachlorofluorescein (ZOE); NAN; NED; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; and Cy7.5; Alexa Fluor 350; Alexa Fluor 488; Alexa Fluor 532; Alexa Fluor 546; Alexa Fluor 568; Alexa Fluor 594; or Alexa Fluor 647.

The label monomers can be incorporated into a nanoreporter at different stages of its assembly, or into a component (*e.g.*, a "flap" or of the nanoreporter prior to its assembly into the nanoreporter).

A label monomer can be directly attached to a nucleotide using methods well known in the art. Nucleotides can also be chemically modified or derivatized in order to attach a label monomer. For example, a fluorescent monomer such as a fluorescein molecule can be attached to dUTP (deoxyuridine-triphosphate) using a four-atom aminoalkynyl group. Each label monomer is attached to a nucleotide making a label monomer: nucleotide complex.

This label monomer: nucleotide complex can be incorporated into nucleic acids (for example, a DNA patch or a detection oligonucleotide) in a variety of ways. For example, a label monomer: nucleotide complex can be incorporated at only one location within a nucleic acid or at two or more locations within a nucleic acid.

Amine-reactive and thiol-reactive fluorophores are available and used for labeling nucleotides and biomolecules. Generally, nucleotides are fluorescently labeled during chemical synthesis, for example, incorporation of amines or thiols during nucleotide synthesis permit addition of fluorophores. Fluorescently labeled nucleotides are commercially available. For example, uridine and deoxyuridine triphosphates are available that are conjugated to ten different fluorophores that cover the spectrum.

A nucleotide can be attached to a label monomer first and then be incorporated into a nucleic acid. Alternatively, an existing nucleic acid can be labeled by attaching a label monomer to a nucleotide within the nucleic acid. For example aminoallyl- ("AA-") modified UTP nucleotides can be incorporated into the RNA product during transcription. In various embodiments, 20% or more of UTP nucleotides in a transcription reaction to generate RNA

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 921

patches are AA modified. In various embodiments, about 20% to 100%, 20% to 80%, 30% to 80%, 40% to 60% or 50% to 75% of UTPs in a transcription reaction are AA-modified, in a preferred embodiment, approximately 50% of UTPs in a transcription reaction are AA-modified.

In addition, for example, different types of label monomer: nucleotide complexes can be incorporated into a single acid nucleic acid, where one component of the nanoreporter code comprises more than one type of signal.

Fluorescent dyes that can be bound directly to nucleotides can also be utilized as label monomers. For example, FAM, JOE, TAMRA, and ROX are amine reactive fluorescent dyes that have been attached to nucleotides and are used in automated DNA sequencing. These fluorescently labeled nucleotides, for example, ROX-ddATP, ROX-ddCTP, ROX-ddGTP and ROX-ddUTP, are commercially available.

Other types of label monomers that may be used to label a nanoreporter are quantum dots. Due to their very small size the quantum dots can be coupled into oligonucleotides directly without affecting the solubility or use of the oligonucleotide. In a preferred embodiment, only one oligonucleotide molecule is coupled to each nanoparticle. To synthesize an oligonucleotide-nanoparticle complex in a 1:1 ratio by conventional batch chemistry, both the oligonucleotide and the nanoparticle require a single reactive group of different kinds that can be reacted with each other. For example, if an oligonucleotide has an amino group and a nanoparticle has an aldehyde group, these groups can react to form a Schiff base. An oligonucleotide can be derivatized to attach a single amino or other functional group using chemistry well known in the art. However, when a nanoparticle is derivatized, it is covered with a chemical reagent which results in coating the entire surface of the nanoparticle with several functional groups.

The invention provides a method of coupling one oligonucleotide to one nanoparticle by chemically coupling the oligonucleotide on a solid surface such as the glass support used for the oligonucleotide synthesis.

For example, commercially available resins for oligonucleotide synthesis such as long chain alkylamino controlled pore glass (lcaa CPG) can be used.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 922

Alternatively, a flat surface such as a derivatized microscope slide can be used. The surface density of the nascent oligonucleotide chains should be lower than the diameter of the nanoparticle. This can be achieved by either choosing a glass support with low surface density of the reactive groups, or by using diluted reagent for the first step of the oligonucleotide synthesis so that the surface is not saturated. Another point of consideration when using the standard glass matrices for oligonucleotide synthesis is to use a pore diameter higher than the nanoparticle diameter to ensure the flow of the reagents. For example, an oligonucleotide can be synthesized on a diluted basis relative to the solid support, for example one tenth of a normal synthesis, to ensure good spacing of the oligonucleotides on the glass support. After the oligonucleotide is synthesized with a reactive functional group, for example, an amino group, derivatized nanoparticles are passed over the glass support to react with the oligonucleotides. A sufficiently large pore size of the glass support can be chosen to prevent clogging with nanoparticles. For example, a pore size of about 200 nm can be used. After the reaction is complete, un-reacted groups on the nanoparticle can be blocked and the complexes can be uncoupled from the glass support.

THE NANOREPORTER CODE

DUAL NANOREPORTERS

A nanoreporter whose components exist in two molecular entities is referred to as a dual nanoreporter. In a dual nanoreporter, generally each component contains a target-specific sequence, which improves the specificity of and binding kinetics of the nanoreporter to its target. The two different target-specific sequences are designed or selected such that each recognizes a different portion of a target molecule.

Figures 1A-1C illustrate embodiments of the invention involving dual nanoreporters. In Figures 1A and 1B, each of the two components of the nanoreporter is labeled, such that the nanoreporter's spectral code is formed only when the two components of the nanoreporter come together upon binding of the dual nanoreporter to its target molecule. However, in a dual nanoreporter, it is not necessary that both components are labeled. For example, as depicted in Figure 1C, one component of a dual nanoreporter is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 923

labeled with the nanoreporter code, and the other component attached to an affinity tag (arrow) that is useful to immobilize the nanoreporter for stretching a visualization.

REGISTERS

The term "register" refers to a set of alternating (every other) label attachment regions. Registers are useful where it is desirable to label adjacent label attachment regions without a spacer region, and where the signal emanating from adjacent label attachment regions cannot be spatially resolved using the desired method of detection. Thus, the signals detected with use of a register is that form by the alternating, rather than adjacent, label attachment regions. Signals detected from a plurality of registers (*e.g.*, that together are all the label attachment regions) can be combined to form a nanoregister code. Generally when using registers, adjacent label attachment regions are labeled with spectrally distinguishable label monomers.

Examples of registers are depicted in Figures 3 and 5. For example, in Figures 3A-3B, there are 8 label attachment regions 1-8. Alternating label attachment regions 1, 3, 5 and 7 form one register, and label attachment regions 2, 4, 6 and 8 form another register. In Figure 3A, only one of the registers (1, 3, 5 and 7) is labeled; in Figure 3B, both registers are labeled.

AFFINITY TAGS

A variety of affinity tags known in the art may be used to purify and/or immobilize nanoreporters.

Where an affinity tag is used to immobilize a nanoreporter for the purpose of detection or imaging, it may be referred to herein as an "anchor." In a preferred embodiment, a biotin anchor is attached to the nanoreporter, allowing immobilization of the nanoreporter on a streptavidin coated slide.

An affinity tag that can be used for attachment to beads or other matrixes for a variety of useful applications including but not limited to purification.

Non-limiting examples of suitable affinity tags are provided below. It should be understood that most affinity tags could serve dual purposes: both as anchors for

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 924

immobilization of the nanoreporters and tags for purification of the nanoreporters (whether fully or only partially assembled) or their components.

In certain embodiments, the affinity tag is a protein monomer. Examples of protein monomers include, but are not limited to, the immunoglobulin constant regions (see Petty, 1996, Metal-chelate affinity chromatography, in Current Protocols in Molecular Biology, Vol. 2, Ed. Ausubel *et al.*, Greene Publish. Assoc. & Wiley Interscience), glutathione S-transferase (GST; Smith, 1993, Methods Mol. Cell Bio. 4:220-229), the *E. coli* maltose binding protein (Guan *et al.*, 1987, Gene 67:21-30), and various cellulose binding domains (U.S. Pat. Nos. 5,496,934; 5,202,247; 5,137,819; Tomme *et al.*, 1994, Protein Eng. 7:117-123), etc. Other affinity tags are recognized by specific binding partners and thus facilitate isolation and immobilization by affinity binding to the binding partner, which can be immobilized onto a solid support. For example, the affinity tag can be an epitope, and the binding partner an antibody. Examples of such epitopes include, but are not limited to, the FLAG epitope, the myc epitope at amino acids 408-439, the influenza virus hemagglutinin (HA) epitope, or digoxigenin ("DIG"). In other embodiments, the affinity tag is a protein or amino acid sequence that is recognized by another protein or amino acid, for example the avidin/streptavidin and biotin.

In certain aspects of the invention, the affinity tag is a nucleotide sequence. A large variety of sequences of about 8 to about 30 bases, more preferably of about 10 to about 20 bases, can be used for purification and immobilization of nanoreporters, and the sequence can be tandemly repeated (*e.g.*, from 1 to 10 tandem repeats). Such a sequence is preferably not widely represented (that is, present in fewer than 5% of the genes, more preferably, present in fewer than 3% of the genes, and, most preferably, present in fewer than 1% of the genes) in the sample being assayed (for example, where the nanoreporter is used for detection of human cellular RNA, the sequence is preferably not widely represented in the human genome); have little or no secondary structure or self-complementarity either internally or with copies of itself when multimerized (that is, all secondary structures of the multimerized tag preferably have a Tm less than 25°C at 1 M NaCl); have no significant identity or complementarity with scaffold or segment sequences (that is, the Tm of complementary

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 925

sequences is preferably less than 25°C at 0.2 M NaCl); and have a Tm of about 35-65°C, more preferably about 40-50°C, in 50 mM Na^+.

In certain embodiments, different sequences are used as purification and immobilization tags. In this case, for example, the purification tag can be as described above, but the immobilization tag can be in the range of 10 to 100 bases, with a Tm up to 95°C in 50 mM Na^+. An alternative embodiment would be to have the purification tag nested within the immobilization tag (*e.g.*, the affinity tag would comprise a 25-base sequence of which 15 bases are used as a purification tag and the entire 25 bases are used as the immobilization tag).

In certain instances, the affinity tag can be used for labeling a nanoreporter in addition to purifying or immobilizing the nanoreporter.

As will be appreciated by those skilled in the art, many methods can be used to obtain the coding region of the affinity tags, including but not limited to, DNA cloning, DNA amplification, and synthetic methods. Some of the affinity tags and reagents for their detection and isolation are available commercially.

TARGET-SPECIFIC SEQUENCES

The term "target-specific sequence" refers to a molecular entity that is capable of binding a target molecule. In the context of a nanoreporter, the target-specific sequence is attached to the nanoreporter scaffold.

The target specific sequence is generally an amino acid sequence (*i.e.*, a polypeptide or peptide sequence) or a nucleic acid sequence.

In specific embodiments, where the target-specific sequence is an amino acid sequence, the target-specific sequence is an antibody fragment, such as an antibody Fab' fragment, a single chain Fv antibody.

The target-specific sequence is preferably a nucleic acid sequence, and is most preferably within an oligonucleotide that is either covalently attached (*e.g.*, by ligation) or noncovalently attached (*e.g.*, by hybridization) to the nanoreporter scaffold. A target-specific nucleic acid sequence is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific embodiments, the target-specific sequence is

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 926

approximately 10 to 500, 20 to 400, 30 to 300, 40 to 200, or 50 to 100 nucleotides in length. In other embodiments, the target-specific sequence is approximately 30 to 70, 40 to 80, 50 to 90, or 60 to 100, 30 to 120, 40 to 140, or 50 to 150 nucleotides in length.

A target-specific nucleotide sequence preferably has a Tm of about 65-90°C for each probe in 825 mM Na^+ (5X SSC), most preferably about 78-83°C.

In certain preferred embodiments, the target specific sequence of each probe of a dual nanoreporter is about 35 to 100 nucleotides (for a total target sequence of about 70 to 200 nucleotides, covered by 2 probes), most preferably about 40 to 50 nucleotides for each probe (for a total of about 80 to 100 nucleotides).

COMPUTER PROGRAMS FOR SELECTION OF TARGET-SPECIFIC SEQUENCES

The invention provides methods, and computer systems and computer program products that may be used to automate the methods of the invention, for selecting target-specific sequences for use in nanoreporters. The invention provides methods, and various computer systems which run one or more programs described below (*e.g.*, target-specific sequence selection module 50), as well as computer program products that comprise computer-readable media and computer-program mechanisms embedded therein which comprise instructions for carrying out the methods of the invention, *i.e.*, running one or more programs described below.

Figure 19 details an exemplary system that supports the functionality described herein. The system is preferably a computer system 10 having:

- a central processing unit 22;
- a main non-volatile storage unit 14, for example, a hard disk drive, for storing software and data, the storage unit 14 controlled by controller 12;
- a system memory 36, preferably high speed random-access memory (RAM), for storing system control programs, data, and application programs, comprising programs and data loaded from non-volatile storage unit 14; system memory 36 may also include read-only memory (ROM);
- a user interface 32, comprising one or more input devices (*e.g.*,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 927

keyboard 28) and a display 26 or other output device;

- a network interface card 20 or other communication circuitry for connecting to detector 72 and, optionally, any wired or wireless communication network 34 (*e.g.*, the Internet or any other wide area network);
- an internal bus 30 for interconnecting the aforementioned elements of the system; and
- a power source 24 to power the aforementioned elements.

Operation of computer system 10 is controlled primarily by operating system 40, which is executed by central processing unit 22. Operating system 40 can be stored in system memory 36. In addition to operating system 40, in a typical implementation, system memory 36 can include one or more of the following:

- file system 42 for controlling access to the various files and data structures used by the present invention;
- a data storage module 44 comprising instructions for storing a plurality of sequences; and
- a target-specific sequence selection module 50 for identifying a plurality of target-specific sequences.

As illustrated in Figure 19, computer system 10 comprises software program modules and data structures. The data structures stored in computer system 10 include, for example, sequence databases of interest and sequences present in the nanoreporter structure (these are protocol- and fabrication-specific sequences). Each of these data structures can comprise any form of data storage including, but not limited to, a flat ASCII or binary file, an Excel spreadsheet, a relational database (SQL), an on-line analytical processing (OLAP) database (MDX and/or variants thereof), or a comma separated value file. In some embodiments, the data structures and software modules depicted in Figure 19 are not housed on computer system 10, but rather are housed on a computer or other type of storage device that is in electrical communication with computer system 10 across network 34.

One aspect of the present invention provides a computer program product comprising a computer readable storage medium (*e.g.*, memory 36, storage unit 14, and/or other computer readable storage media) and a computer program mechanism embedded

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 928

therein. The computer program mechanism is for identifying suitable target-specific sequences for use in nanoreporters. The computer program mechanism comprises data storage module 44 and the target-specific sequence identification module 50.

Data storage module 44. Data storage module 44 comprises sequence databases, for example, for use as reference sequences. For example, human reference sequences can be acquired from refseq database for mRNA sequences (Pruitt *et al.*, 2005, Nucleic Acids Res. 33(1):D501-D504).

In addition, the data storage module can comprise sequences of relevance to the user of the program, for example sequences used in nanoreporter assemblies that can be used as reference sequences in the Higher-Resolution Context Sensitive Structural Filter of the third selection tier of the target-specific sequence selection program (described below).

Target-specific sequence identification module 50.

This module is illustrated in Figures 21 and 22 for dual nanoreporter and single nanoreporter target-specific sequence selection, respectively. However, the methods described herein are useful for identifying target-specific sequences (or pairs thereof) for use in any other probe system, for example for use in gene expression analysis by RT-PCR or microarrays.

First Selection Tier

In a single-tiered program or in a first tier of a multi-tiered program, the program generates candidate target-specific sequences of a given size (*e.g.*, 100 bases) from each target mRNA. In Figures 20 and 21, this step is illustrated as step 2002 and 2202, respectively.

In various embodiments, target-specific sequences are selected for any where from 1 to 10,000 target mRNAs, for example from 1 to 20 target mRNAs, from 5 to 100 target mRNAs, from 20 to 250 target mRNAs, from 100 to 500 target mRNAs, from 200 to 1,000 target mRNAs, from 500 to 2,000 target mRNAs, from 1,000 to 10,000 mRNAs, or any range in between (*e.g.*, from 5 to 250 target mRNAs). In specific embodiments, target-specific sequences are selected for at least 10 target mRNAs, at least 25 target mRNAs, at

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 929

least 50 target mRNAs, at least 100 mRNAs, at least 200 target mRNAs, or at least 500 target mRNAs.

Candidate target-specific sequences are preferably 30-160 bases long. Candidate target-specific sequences for use in a single nanoreporters probe are preferably 30-80, more preferably 35-70, and most preferably 40-55 bases in length. Candidate target-specific sequences for use in two nanoreporter probes, the candidate target-specific sequences are preferably 60-160, more preferably 70-150, and most preferably 80-110 bases in length.

For each target molecule, the pool of candidate target-specific sequences may be all possible target-specific sequences of a selected size against the target molecule. The pool of target-specific sequences can be generated using a sliding window such that each candidate target-specific sequence will be adjacent to or overlap with the adjacent candidate target-specific sequence. In embodiments where the sliding window covers overlapping candidate target-specific sequences, the overlap can be of a step size of 1 to up to 1 base less than the length of the candidate target-specific sequence (*e.g.*, for a 100-base target-specific sequence, the step size can be anywhere from 1 to 99 bases, wherein a 1-base step size results in a 99-base overlap between adjacent candidate target-specific sequences and a 99-base step size results in a 1-base overlap between adjacent candidate target-specific sequences). In a preferred embodiment, the step size is not a multiple of 3. In other preferred embodiments, the step size is 2-20 bases less than the window size, most preferably 4-10 bases less than the window size. Where the candidate target-specific sequence will be divided into two target-specific sequences for use in two nanoreporter probes (*e.g.*, the two components of a dual nanoreporter), the step size is preferably less than half the window size (for example, for a 100-base target-specific sequence which will form the basis of two 50-base target-specific regions of a dual nanoreporter, the window size is preferably less than 50).

In some embodiments, each candidate target-specific sequence is assessed on any combination of two or more, preferably three or more, and more preferably four or all four of the following criteria (this step is reflected as step 2004 of Figure 20 and step 2104 of Figure 21, respectively):

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 930

(1) the candidate target-specific sequence has no inverted repeats of a predetermined length or greater, *e.g.*, five or more, preferably six or more, consecutive bases (this criterion prevents inter-probe interactions);

(2) the candidate target-specific sequence has no direct repeats of a predetermined length or greater, *e.g.*, five or more consecutive bases, more preferably six, seven or eight or more consecutive bases, and most preferably nine or more consecutive bases (this criterion prevents inter-probe interactions);

(3) each target-specific sequence (or each of the 5′ half and the 3′ half of the candidate target-specific sequence where the target-specific sequence will be the basis of the two target-specific sequences of a dual nanoreporter) has a GC content in a preferred range, *e.g.*, of 25-85%, more preferably 30-80%, yet more preferably 35-75% GC, and yet most preferably 40-70%, or any range in between (*e.g.*, 32%-76% or 38%-68%) (this criterion is used for identifying/selecting target-specific regions of dual nanoreporters and avoids skew in the hybridization properties of the two components of the dual nanoreporters); and

(4) the candidate target-specific sequence has no contiguous stretches of Cs longer than a predetermined length, *e.g.*, longer than three, longer than four, or longer than five (this criterion avoids complications in probe synthesis); and

(5) the candidate target-specific sequence has a melting temperature in a predetermined range, preferably from 60-75°C at its lower end to 80-90°C at its upper end.

In specific embodiments, mFOLD or Oligowalk (available on the Internet) may be used to predict probe folding. If for a given target molecule one or more candidate target-specific sequences that meet the predetermined combination of the foregoing criteria, the target-specific sequences can be selected for use in a nanoreporter probe of the invention (in a single-tier selection), or the candidate target-specific sequence can be subject to additional selection criteria, as described below. If, on the other hand, there are no candidate target-specific sequences against a particular target molecule that meet the predetermined combination of the foregoing four criteria (*e.g.*, all four criteria or some predetermined subset of the four criteria), one or more of the criteria used in this selection step are relaxed and candidate target-specific sequences are selected on the basis of the less stringent criteria.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 931

The melting temperatures of each candidate target-specific sequence are either actual melting temperatures (for example melting temperatures measured under conditions of interest) or calculated using standard algorithm and thermodynamic parameters. As used herein, a reference to a melting temperature, or Tm, refers to the melting temperature of a duplex consisting of the sequence in question (*e.g.*, the candidate target-specific sequence (usually DNA) and the reverse complement (usually mRNA). For RNA/DNA hybrids, for example, the Dan program of the EMBOSS freeware program suite (available on the Internet) calculates the melting temperature (Tm) and the percent G+C of a nucleic acid sequence. For the melting temperature profile, free energy values calculated from nearest neighbor thermodynamics are used (Breslauer *et al.*, Proc. Natl. Acad. Sci. USA 83:3746-3750 and Baldino *et al.*, Methods in Enzymol. 168:761-777). The Tm information can be used to discard candidate target-specific sequences of unsuitable melting temperatures (*e.g.*, outside the range of 65°C-90°C), and is used in a subsequent selection round of a multi-tiered program for further refinement of probe selection.

Many genes produce different RNAs, for example as a result of alternative splicing. The first selection tier can be used to identify specific products or all products of a particular gene, by running the first selection tier in "specific" mode or "common" mode. In "specific" mode, the sliding window only covers regions that are specific to one RNA, for example regions that are at splice junctions specific to that RNA. In "common" mode, the sliding window covers regions that are common to all products of interest of a given gene.

Either following or during the first selection round, an alignment such as a BLAST or FASTA alignment is performed on target-specific sequence (using algorithms such as NCBI BLAST, selecting dual strand BLAST with the following parameters: 'w11 q-1'). The alignment output is used in a subsequent selection round of a multi-tiered program. The alignment can be performed locally or remotely. Local alignments require that the local computer carry the alignment program (*e.g.*, BLAST) and the sequence database against which the candidate target-specific sequences are going to be compared; for example, where the target genes are human genes whose expression will be monitored, the sequence database can be a database of expressed human genes. Optionally, the sequence database contains

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 932

only sequences that are expressed in a target tissue of interest. Remote alignments require a connection to a remote site that can perform alignments, such as the NCBI web site.

Second Selection Tier

As reflected in step 2006 of Figure 20 and step 2106 of Figure 21, candidate sequences are eliminated from contention if they have the potential to cross-hybridize to non-specific sequences present in a biological sample of interest.

In one embodiment, the cross-hybridization potential of target-specific sequence is determined as follows. The sequence selection program performs an additional, second step of alignment output interpretation and scoring. In this step, for example, the BLAST (preferably in dual strand mode) or other alignment program results are used to calculate some basic metrics for every hit. In one embodiment, the BLAST hit coordinates (which, for 100-base candidate target-specific sequences, will range anywhere from ~12 to 100 bases when the 'w11 q-1' BLAST parameter set is used) are extended to line up with the candidate target-specific sequence, and the following is calculated:

(i) Percent identity calculated between each hit and the candidate target-specific sequence; and

(ii) Maximum contiguous block of identity (stretch of contiguous bases that align perfectly) between each hit and candidate target-specific sequence (or each of the 5′ half and the 3′ half of the candidate target-specific sequence where the target-specific sequence will be the basis of two halves of a dual nanoreporter).

Sequences are eliminated from contention when:

(1) the percentage identity between non-specific hits (*i.e.*, those sequence hits identified by the alignment program (*e.g.*, BLAST) that do not correspond to the gene to which the target-specific sequence corresponds) and the candidate target-specific sequence is greater than a predetermined amount; and/or

(2) the longest contiguous block of sequence identity between the candidate target-specific sequence and non-specific hits.

In certain embodiments, the cutoffs above are (i) a sequence identity with a non-specific hit of 95% or greater, 90% or greater, 85% or greater, or 80% or greater and (ii)

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 933

a contiguous block of sequence identity with a non-specific hit of 20 bases or greater, of 19 bases or greater, of 18 bases or greater, of 17 bases or greater, of 16 bases or greater, of 15 bases or greater, of 14 bases or greater, of 13 bases or greater, of 12 bases or greater, of 11 bases or greater or of 10 bases or greater.

Candidate target-specific sequences that meet criteria (i) and/or (ii) of the second selection tier are eliminated. This step allows the elimination of target-specific sequences that will cross-hybridize to transcripts other than the target transcript in a nanoreporter assay. In addition to criteria (i) and/or (ii) above, other criteria selected by the user may be used to score candidate eliminate target-specific sequences on the basis of their ability to cross-hybridize with non-target sequences.

The scored candidate target sequences of the second selection tier can be subject to further optional steps in a third selection tier, described below.

Third Selection Tier

This third selection tier consists of a series of various optional steps to optimize the target-specific sequence selection.

(a) *Higher-Resolution Context Sensitive Structural Filter*

A "Higher-Resolution Context Sensitive Structural Filter" or HRCSSF scans various parts of nanoreporters, such as the nanoreporter backbone (*e.g.*, M13), affinity tags (*e.g.*, G-hooks, F-hooks), and checks for inter- and intra-reporter interactions based on the context of when certain exposed sequences have potential to interact.

In certain embodiment, the HRCSSF contains two or three main features, described below:

(1) A structural check on the target-specific sequence (or pair of target-specific sequences). This is almost identical to the first two criteria of the first tier (*e.g.*, as reflected in (i) and (ii) of step 2004 of Figure 20 and step 2104 of Figure 21), but allows the addition of non-target-specific sequences present in the nanoreporters or mRNA sequence adjacent to the target-specific sequences. The two primary cutoffs are Direct Repeats (DR) and Inverted repeats (IR). Preferably those target-specific sequences with DRs of a

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 934

predetermined length of, *e.g.*, 6-10 bases or longer, are eliminated. For example, target-specific sequences with DRs 10 bases or longer, at least 8 bases or longer, 6 bases or longer, are eliminated). Preferably those target-specific sequences with IRs of a predetermined length of, *e.g.*, 4-8 bases or longer, are eliminated. For example, target-specific sequences with IRs of 4 bases or longer, 6 bases or longer, or 8 bases or longer, are eliminated.

(2) An intra-molecular check of each nanoreporter (or each component of a dual nanoreporter). Again, preferably, the two primary cutoffs are Direct Repeats (DR) and Inverted repeats (IR). Preferably, the cutoff size for each DR and IR is 8-12, such that target-specific sequences with DRs or IRs of 8 bases or longer, 10 bases or longer, or 12 bases or longer, are eliminated.

(3) Optionally, for dual nanoreporters, an inter-molecular check between the different components of the nanoreporters (for example between a ghost probe and a reporter probe). Again, the two primary cutoffs are Direct Repeats (DR) (preferably those target-specific sequences with direct repeats of 12-18 bases or longer, *e.g.*, 17 bases or longer, 16 bases or longer, or 15 bases or longer, are eliminated) and Inverted repeats (IR) (preferably those target-specific sequences with inverted repeats of 12-18 bases or longer, *e.g.*, 13 bases or longer, at least 15 bases or longer, or at least 17 bases or longer, are eliminated).

The algorithm contains no scoring, if a feature is found above cutoff, that a target-specific sequence (or pair of target-specific sequences) is discarded completely.

(b) *Dynamic Tm Filter*

To optimize the signal to noise ratio in multiplex nanoreporter detection assays (involving the detection of multiple target molecules in one experiment), it is preferable that the target-specific sequences of all reporter probes fall into a small melting temperature range, *e.g.*, three, four, five, six, or seven degrees Celsius between 72°C and 86°C (*e.g.*, from 78°C to 83°C or from 75°C to 82°C). The dynamic Tm filter takes candidate target-specific sequences that are above Tm Range and "trims" the target-specific sequences until they either fall into the range, or reach a minimum size. Preferably, the candidate target-specific sequences for dual nanoreporters are trimmed from their outside

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 935

ends (*i.e.*, the 5′ end for the 5′ candidate sequence and the 3′ end for the 3′ candidate sequence) or from either end for individual target-specific sequences. This embodiment is illustrated steps 2008, 2010, 2012, 2014, 2016, and 2018 of Figure 20, and steps 2108, 2110, 2112, 2114, 2116, and 2118 of Figure 21. For dual nanoreporters whose target-specific sequences are not adjacent, the opposite end can be trimmed also; however, it is preferable that each pair of target-specific sequences correspond to sequences no more than 5 nucleotides apart on the target mRNA, and more preferably no more than 3 nucleotides apart or even no more than 1 nucleotide apart on the target mRNA.

The dynamic Tm filter can be designed to also extend sequences that have too low a Tm (outside the preselected Tm range) until they fall into the range, or reach a maximum size. Preferably, the candidate target-specific sequences for dual nanoreporters are extended from their outside ends (*i.e.*, the 5′ end for the 5′ candidate sequence and from the 3′ end for the 3′ candidate sequence) or from either end for individual target-specific sequences.

Thus, in this dynamic Tm filter step, a candidate target-specific sequence may be modified to be longer or shorter than the initial window size of the first selection tier.

(c) *Transcript Specificity Check*

Many genes produce different RNAs, for example as a result of alternative splicing. In certain embodiments of the invention, a transcript specificity check is performed.

For the target-specific sequence to hybridize with multiple variants of the mRNA it must have a perfect or near-perfect alignment to the variants in common. Accordingly, after identifying one or more target-specific sequences, it is possible to check whether they hybridize to multiple splice forms (or other variations such as allelic variations) of the same RNA. In practice, it is possible to check whether in the second selection tier sequences are not eliminated if they only cross-hybridize to variants of the same RNA.

As an alternative to the transcript specificity check, it is possible to include in the pool for the first selection tier only candidate sequences present in only a specific variant of an RNA of interest (so as to identify target-specific sequences hybridizable only to that

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 936

variant), or candidate sequences present in multiple variants (so as to identify target-specific sequences hybridizable to the multiple variants).

Scoring

A scoring software module that calculates quality scores (the term score refers to any qualitative and quantitative values with regard to desired properties of a target-specific sequence) for candidate target-specific sequences. Scores based on the Tm values and non-specific hybridization potential of each candidate target-specific are inserted into the score sheet which is used to select "top scoring" target-specific sequence for each target molecule. Target-specific sequences (or, for dual nanoreporters, target-specific sequence pairs) passing all minimum requirements are given a score to choose the pair most likely to perform well. In an exemplary embodiment, this score is based on a weighted score of the cross-hybridization potentials and the melting temperatures of the adjacent target-specific sequences (whether unmodified or as modified by the dynamic Tm filter). In a specific embodiment, the weighted score is calculated according to the formula:

(Tm score * WFa) + (MCB score * WFb) + (PID score * WFc)

where:

Tm score is a melting temperature score calculated according to the formula:

(differential score + general score)/3

where the differential score is calculated according to the following formula:

$1-|(TmA - TmB)| / (TmHco - TmLco)$

where the general score is calculated according to the following formula:

$(((TmI - |(TmA - TmI)|) / TmI) + (((TmI - |(TmB - TmI)|) / TmI)))$

where TmA and TmB are the respective melting temperatures of the adjacent target-specific sequences (one or both of which is optionally modified by the dynamic Tm filter), TmHco is the upper limit of the second predetermined temperature range; TmLco is the lower limit of the second predetermined temperature range; and TmI is a predetermined ideal melting temperature;

where:

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 937

MCB score is a maximum contiguous block score calculated according to the formula:

1 - (MCB / MCBco);

where MCB is the greater of (i) and (ii) below, where (i) and (ii) are respectively:

(i) the maximum contiguous block of identity between (A) and (B) below:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more variants, such as alternatively spliced mRNAs, corresponding to the same gene as the target mRNA;

and

(ii) the maximum contiguous block of identity between (A) and (B) below:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more variants, such as alternatively spliced mRNAs, corresponding to the same gene as the target mRNA,

and wherein MCBco is the first predetermined cutoff;

where:

PID score is a percent identity score calculated according to the formula:

1 - (PID / PIDco));

where PID is the greater of (i) and (ii) below, where (i) and (ii) are respectively:

(i) the greatest percentage sequence identity between (A) and (B) below:

(A) a first target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 938

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more variants, such as alternatively spliced mRNAs, corresponding to the same gene as the target mRNA;

and

(ii) the greatest percentage sequence identity between (A) and (B) below:

(A) a second target-specific nucleotide sequence in said pair of adjacent target-specific sequences; and

(B) a sequence in the database other than the complement of the target mRNA and, optionally, other than the complements of one or more variants, such as alternatively spliced mRNAs, corresponding to the same gene as the target mRNA,

and wherein PIDco is the second predetermined cutoff,

and where WFa, WFb, and WFc are each independently a weighting factor, each of which is a real number.

For dual nanoreporters, the top scoring pair of target-specific sequences are selected, which are preferably complementary to portions of the target molecule no more than 10 bases apart, more preferably complementary to portions of the target molecule no more than 5, 4, 3, 2 or 1 base(s) apart, and most preferably complementary to immediately adjacent portions of the target molecule).

In a variation of the computer program of the present invention, instead of using the five criteria of the first selection tier as cutoff points, such criteria may be factored in to the scores of the candidate target-specific sequences.

Iterative rounds of selection according to the first selection tier, with progressively more relaxed parameters (*e.g.*, broader melting temperature range, broader %GC content range, higher cutoff for inverted and/or direct repeats), can be used to identify target-specific sequences of genes for which suitable target-specific sequences are identified under the more stringent criteria.

As will be appreciated by one of skill in the art, the present invention may be embodied as a method, computer system or program products. Accordingly, the present invention may take the form of data analysis systems, methods, analysis software, etc. Software written according to the present invention can be stored in some form of computer

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 939

readable medium, such as memory, or CD-ROM, or transmitted over a network, and executed by a processor. For a description of basic computer systems and computer networks, see, *e.g.*, Introduction to Computing Systems: From Bits and Gates to C and Beyond, by Yale N. Patt, Sanjay J. Patel, 1st edition (Jan. 15, 2000) McGraw Hill Text; ISBN: 0072376902; and Introduction to Client/Server Systems: A Practical Guide for Systems Professionals, by Paul E. Renaud, 2nd edition (June 1996), John Wiley & Sons; ISBN: 0471133337.

Each of the methods, computer program products, and computer systems disclosed herein optionally further comprise a step of, or instructions for, outputting or displaying a result (for example, to a monitor, to a user, to computer readable media, *e.g.*, storage media or to a remote computer). Here the result is any result obtained by the methods, computer program products, and computer systems disclosed herein. Optionally, the method further comprises the step of outputting to a user interface device, a computer readable storage medium, or a local or remote computer system, or displaying, one or a plurality of candidate target-specific sequences (optionally, modified by the dynamic Tm filter). Moreover, in certain embodiments, the candidate target-specific sequences (optionally, modified by the dynamic Tm filter) may be outputted as pairs of adjacent target-specific nucleotide sequences, *e.g.*, for use in dual nanoreporters. The candidate target-specific sequences outputted in this manner can be target-specific sequences that have undergone only the first selection tier; the first and second selection tiers; or the first selection tier, the second selection tier, and one or more embodiments of the third selection tier (such as the dynamic Tm filter and/or the HRCSSF and/or transcript specificity check). In certain specific embodiments, the candidate target-specific sequences are outputted or displayed in a ranked order based on a weighted score, for example a weighted score of the cross-hybridization potentials and the melting temperatures of the sequences (or one or both the adjacent target-specific nucleotide sequences contained therein). An example of a scoring algorithm is described above.

Computer software products may be written in any of various suitable programming languages, such as C, C++, Fortran and Java (Sun Microsystems). Preferably, the software products are written in Perl, a dynamic programming language that derives broadly from C. The computer software product may be an independent application with

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 940

data input and data display modules. Alternatively, the computer software products may be classes that may be instantiated as distributed objects. The computer software products may also be component software such as Java Beans (Sun Microsystems), Enterprise Java Beans (EJB), Microsoft™ COM/DCOM, etc.

TARGET MOLECULES

The term "target molecule" is the molecule detected or measured by binding of a labeled nanoreporter whose target-specific sequence(s) recognize (are specific binding partners thereto). Preferably, a target molecule can be, but is not limited to, any of the following: DNA, cDNA, RNA, mRNA, peptide, a polypeptide/protein (*e.g.*, a bacterial or viral protein or an antibody), a lipid, a carbohydrate, a glycoprotein, a glycolipid, a small molecule, an organic monomer, or a drug. Generally, a target molecule is a naturally occurring molecule or a cDNA of a naturally occurring molecule or the complement of said cDNA.

A target molecule can be part of a biomolecular sample that contains other components or can be the sole or major component of the sample. A target molecule can be a component of a whole cell or tissue, a cell or tissue extract, a fractionated lysate thereof or a substantially purified molecule. The target molecule can be attached in solution or solid-phase, including, for example, to a solid surface such as a chip, microarray or bead. Also the target molecule can have either a known or unknown structure or sequence.

In certain specific embodiments, that target molecule is not a chromosome. In other specific embodiments, the target molecule is no greater than 1,000 kb (or 1 mb) in size, no greater than 500 kb in size, no greater than 250 kb in size, no greater than 175 kb in size, no greater than 100 kb in size, no greater than 50 kb in size, no greater than 20 kb in size, or no greater than 10 kb in size. In yet other specific embodiments, the target molecule is isolated from its cellular milieu.

In specific, non-limiting embodiments, the target molecule is one of the following antibodies or an antigen recognized by one of the following antibodies: anti-estrogen receptor antibody, an anti-progesterone receptor antibody, an anti-p53 antibody, an anti-Her-2/neu antibody, an anti-EGFR antibody, an anti-cathepsin D antibody, an anti-Bcl-2

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 941

antibody, an anti-E-cadherin antibody, an anti-CA125 antibody, an anti-CA15-3 antibody, an anti-CA19-9 antibody, an anti-c-erbB-2 antibody, an anti-P-glycoprotein antibody, an anti-CEA antibody, an anti-retinoblastoma protein antibody, an anti-ras oncoprotein antibody, an anti-Lewis X antibody, an anti-Ki-67 antibody, an anti-PCNA antibody, an anti-CD3 antibody, an anti-CD4 antibody, an anti-CD5 antibody, an anti-CD7 antibody, an anti-CD8 antibody, an anti-CD9/p24 antibody, an anti-CD10 antibody, an anti-CD11c antibody, an anti-CD13 antibody, an anti-CD14 antibody, an anti-CD15 antibody, an anti-CD19 antibody, an anti-CD20 antibody, an anti-CD22 antibody, an anti-CD23 antibody, an anti-CD30 antibody, an anti-CD31 antibody, an anti-CD33 antibody, an anti-CD34 antibody, an anti-CD35 antibody, an anti-CD38 antibody, an anti-CD41 antibody, an anti-LCA/CD45 antibody, an anti-CD45RO antibody, an anti-CD45RA antibody, an anti-CD39 antibody, an anti-CD100 antibody, an anti-CD95/Fas antibody, an anti-CD99 antibody, an anti-CD106 antibody, an anti-ubiquitin antibody, an anti-CD71 antibody, an anti-c-myc antibody, an anti-cytokeratins antibody, an anti-vimentins antibody, an anti-HPV proteins antibody, an anti-kappa light chains antibody, an anti-lambda light chain antibody, an anti-melanosome antibody, an anti-prostate specific antigen antibody, an anti-S-100 antibody, an anti-tau antigen antibody, an anti-fibrin antibody, an anti-keratins antibody, an anti-Tn-antigen antibody receptor protein, a lymphokine, an enzyme, a hormone, a growth factor, or a nucleic acid binding protein, a ligand for a cell adhesion receptor; a ligand for a signal transduction receptor; a hormone; a molecule that binds to a death domain family molecule; an antigen; a viral particle, a viral coating protein or fragment thereof, a toxic polypeptide selected from the group consisting of: (a) ricin, (b) Pseudomonas exotoxin (PE); (c) bryodin; (d) gelonin; (e) α-sarcin; (f) aspergillin; (g) restrictocin; (h) angiogenin; (i) saporin; (j) abrin; (k) pokeweed antiviral protein (PAP); and (l) a functional fragment of any of (a)-(k); a cytokine, or a soluble cytokine selected from the group consisting of erythropoietin, interleukins, interferons, fibroblast growth factors, transforming growth factors, tumor necrosis factors, colony stimulating factors and epidermal growth factor, Class I MHC antigens, class II MHC antigens, internalizing cell-surface receptors and/or viral receptors.

In specific, non-limiting embodiments, the target molecule is an antigen such as alpha fetoprotein, alpha-1 antitrypsin, α-2 macroglobulin, adiponectin, apoliprotein-A-1,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 942

apoliprotein-CIII, apoliprotein-H, BDNF, β-2 microglobulin, C reactive protein, calcitonin, cancer antigen 19-9, cancer antigen 125, CEA, CD 40, CD 40 ligand, complement 3, CK-MB, EGF, ENA-78, endothelin-1, enrage, eotaxin, erythropoietin, Factor VII, FABP, ferritin, FGF-basic, fibrinogen, G-CSF, GST, GM-CSF, growth hormone, haptoglobin, ICAM-1, IFN-gamma, IgA, IgE, IGF-1, IgM, IL-Iα, IL-1β, IL-1ra, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12 p40, IL-12 p70, IL-13, IL-15, IL-16, insulin, leptin, lipoprotein (a), lymphotactin, MCP-1, MDC, MIP-1α, MIP-1β, MMP-2, MMP-3, MMP-9, myeloperoxidase, myoglobin, PAI-1, PAP, PAPP-A, SGOT, SHBG, PSA (free), RANTES, serum amyloid P, stem cell factor, TBG, thrombopoietin, TIMP-1, tissue factor, TNF-α, TNF-β, TNF RII, TSH, VCAM-1, VEGF, or vWF.

In some embodiments, the target molecule is an autoimmune related molecule such as ASCA, β-2 glycoprotein, C1q, centromere Prot. B, collagen type 1, collagen type 2, collagen type 4, collagen type 6, Cyto P450, ds DNA, histone, histone H1, histone H2A, histone H2B, histone H3, histone H4, HSC-70, HSP-32, HSP-65, HSP-71, HSP-90α, HSP-90β, insulin, JO-1, mitochondrial, myeloperoxidase, pancreatic islet cells, PCNA, PM-1, PR3, ribosomal P, RNP-A, RNP-C, RNP, Sel-70, Smith, SSA, SSB, T3, T4, thyroglobulin, tTG, (celiac disease), or thyroid microsomal.

In some embodiments, the target molecule is a component isolated from an infectious agent, such as Cholera Toxin, Cholera Toxin β, *Campylobacter jejuni*, cytomegalovirus, Diptheria toxin, Epstein-Barr NA, Epstein-Barr EA, Epstein-Barr VCA, *Heliobacter pylori*, HBV core, HBV envelope, HBV surface (Ad), HBV surface (Ay), HCV core, HCV NS3, HCV NS4, HCV NS5, hepatitis A, hepatitis D, HEV orf2 3KD, HEV orf2 6 KD, HEV orf 3KD, HIV-1 p24, HIV-1 gp4l, HIV-1 gp120, HPV, HSV-1/2, HSV-1 gD, HSV-2 gD, HTLV-1/2, influenza A, influenza A H3N2, influenza B, *Leishmania donorani*, Lyme disease, mumps, *M. pneumonia*, *M. tuberculosis*, parainfluenza 1, parainfluenza 2, parainfluenza 3, polio virus, RSV, Rubella, Rubeola, Streptolysin O, Tetanus Toxin, *T. pallidum* 15kD, *T. pallidum* p47, *T. cruzi*, Toxoplasma, Varicella zoster.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 943

NANOREPORTER POPULATIONS

The present invention provides nanoreporter or nanoreporter label unit populations, for example nanoreporter or nanoreporter label unit libraries, that contain at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 75, at least 100, at least 200, at least 300, at least 400, at least 500, at least 750, or at least 1,000 unique nanoreporters or nanoreporter label units, respectively. As used herein, "unique" when used in reference to a nanoreporter or nanoreporter label units within a population is intended to mean a nanoreporter or label unit that has a code that distinguishes it from other nanoreporters or label units in the same population.

In specific embodiments, the present invention provides nanoreporter populations with at least 5,000, at least 10,000, at least 20,000 or at least 50,000 unique nanoreporters or nanoreporter label units.

The nanoreporters in a population of nanoreporters can be singular nanoreporters, dual nanoreporters, or a combination thereof. The nanoreporters can be labeled or unlabeled.

The size of a nanoreporter population and the nature of the target-specific sequences of the nanoreporters within it will depend on the intended use of the nanoreporter. Nanoreporter populations can be made in which the target-specific sequences correspond to markers of a given cell type, including a diseased cell type. In certain embodiments, nanoreporters populations are generated in which the target-specific sequences represent at least 0.1%, at least 0.25%, at least 0.5%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the different type of transcripts in a cell. In certain embodiments, nanoreporters populations are generated in which the target-specific sequences represent at least 0.1%, at least 0.25%, at least 0.5%, at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the different genes in a cell. In yet other embodiments, nanoreporter populations are generated in which at least some of the target-specific sequences represent rare transcripts in a cell or tissue. Such nanoreporter populations preferably represent at least 5 rare transcripts. In specific embodiments, such

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 944

nanoreporter populations represent at least 10, at least 20, at least 30, at least 40 or at least 50 rare transcripts.

In a specific embodiment, the cell or tissue is a mammalian cell or tissue, and more preferably is a human cell or tissue.

In certain embodiments, the nanoreporter population is a diagnostic or prognostic nanoreporter populations. For example, a diagnostic nanoreporter population can be generated that is useful for screening blood products, in which the target-specific sequences bind to the nucleic acids of contaminating viruses such as hepatitis B, hepatitis C, and the human immunodeficiency virus. Alternatively, the diagnostic nanoreporter population may contain target-specific sequences corresponding to cellular disease markers, such as tumor antigens. Prognostic nanoreporter populations generally include target-specific markers that represent different stages of a given disease such as cancer. By selecting appropriate target-specific sequences, a nanoreporter population can be used both to diagnose and prognose disease.

BIOMOLECULAR SAMPLES

The nanoreporter systems of the invention can be used to detect target molecule in any biomolecular sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: cells (including both primary cells and cultured cell lines), cell lysates or extracts (including but not limited to RNA extracts; purified mRNA), tissues and tissue extracts (including but not limited to RNA extracts; purified mRNA); bodily fluids (including, but not limited to, blood, urine, serum, lymph, bile, cerebrospinal fluid, interstitial fluid, aqueous or vitreous humor, colostrum, sputum, amniotic fluid, saliva, anal and vaginal secretions, perspiration and semen, a transudate, an exudate (*e.g.*, fluid obtained from an abscess or any other site of infection or inflammation) or fluid obtained from a joint (*e.g.*, a normal joint or a joint affected by disease such as rheumatoid arthritis, osteoarthritis, gout or septic arthritis) of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred; environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples including extracellular fluids,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 945

extracellular supernatants from cell cultures, inclusion bodies in bacteria, cellular compartments, cellular periplasm, mitochondria compartment, etc.

The biomolecular samples can be indirectly derived from biological specimens. For example, where the target molecule of interest is a cellular transcript, *e.g.*, a messenger RNA, the biomolecular sample of the invention can be a sample containing cDNA produced by a reverse transcription of messenger RNA. In another example, the biomolecular sample of the invention is generated by subjecting a biological specimen to fractionation, *e.g.*, size fractionation or membrane fractionation.

The biomolecular samples of the invention may be either "native," *i.e.*, not subject to manipulation or treatment, or "treated," which can include any number of treatments, including exposure to candidate agents including drugs, genetic engineering (*e.g.*, the addition or deletion of a gene), etc.

SEPARATION OF LABEL MONOMERS

In addition to detecting an overall signal generated from a labeled nanoreporter, the invention provides for the determination of the spatial location of signals emanating from the label monomers (*i.e.*, spots) on a nanoreporter, each spot representing the aggregate signal from label monomers attached to a given label attachment region. A spot may contain signals of the same wavelength or of different wavelengths. Thus, the nature of the spots on a nanoreporter and their location constitutes the nanoreporter code.

Any of a variety of means can be used to "stretch" the nanoreporter to separate the individual spots. For example, a nanoreporter can be stretched using a flowstretch technique (Henegariu *et al.*, 2001, Biotechniques 31:246-250), a receding meniscus technique (Yokota *et al.*, 1997, Nuc. Acids Res. 25:1064-1070) or an electrostretching technique (Matsuura *et al.*, 2001, Nuc. Acids Res. 29: E79).

The use of flow-stretching, receding meniscus, or electro-stretching techniques allows for the separation of the label attachment regions within a nanoreporter so that one can determine spatially where a particular signal is positioned in the nanoreporter. Therefore, unique nanoreporters that have the same combination of label monomers and the same

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 946

overall signal can be differentiated from one another based on the location of those label monomers within the nanoreporter.

This ability to locate the position of a label attachment region or spot within a nanoreporter allows for the position of the signal(s) emitted by the label monomers in each label attachment region to be used as a distinguishing characteristic when generating a set of unique nanoreporters. Hence, a complex set of nanoreporters can be generated using the same combination of starting label monomers by varying the positions of the label monomers within a nanoreporter.

Prior to stretching a nanoreporter, it is preferable to immobilize the nanoreporter to a solid surface using an affinity tag, as described in Section 5.6 above.

In certain aspects of the invention, one end of a nanoreporter is immobilized, either through specific or non-specific binding to a solid surface, the nanoreporter is stretched, and then the other end of the reporter is immobilized, also either through specific or non-specific binding to a solid surface. Accordingly, the nanoreporter is "frozen" in its stretched, or extended, state, to facilitate resolution of the nanoreporters code by detecting and/or imaging the signals emitted by the label monomers attached to a nanoreporter and their locations relative to one another. These aspects of the invention are described below in Section 5.13.

IMMOBILIZATION OF STRETCHED NANOREPORTERS

The present invention provides methods and compositions that facilitate the identification of primary structures of a variety of nanoreporters. In certain aspects, the present invention provides methods for the selective immobilization of nanoreporters in an extended state. According to the invention, a nanoreporter can be selectively immobilized while fully extended under whatever force is used for the extension. In addition, the methods of the invention facilitate the selective immobilization of extended nanoreporters that are oriented with respect to each other. In other words, according to the methods of the invention, a plurality of nanoreporters can readily be immobilized in the same orientation with respect to each other.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 947

In one aspect, the present invention provides methods for selectively immobilizing a nanoreporter in an extended state. For the methods of this aspect of the invention, generally, a first portion of the nanoreporter is immobilized by any technique known to those of skill in the art. Indeed, the technique for immobilizing the first portion of the nanoreporter is not critical to many embodiments of the invention. In certain embodiments, the first portion of the nanoreporter can be immobilized selectively or non-selectively. In certain embodiments the first portion is immobilized by one or more covalent bonds. In certain embodiments, the first portion is immobilized by one or more non-covalent bonds. Exemplary immobilized first portions are described in the sections below.

With an immobilized first portion, the nanoreporter can be extended by any technique for extending a nanoreporter apparent to those of skill in the art. In certain embodiments, the technique for extending the nanoreporter is not critical for the methods of the invention. In certain embodiments, the technique for extending the nanoreporter appropriate for the class of nanoreporter according to the judgment of one of skill in the art. In certain embodiments, the nanoreporter is extended by application of a force capable of extending the nanoreporter. The force can be any force apparent to one of skill in the art for extending the nanoreporter. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the nanoreporter are described in the sections below.

The nanoreporter is in an extended state if it would be recognized as extended by one of skill in the art. In certain embodiments, the nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter. In certain embodiments, the nanoreporter is in an extended state when its average hydrodynamic radius is more than double the average hydrodynamic radius of the nanoreporter in its native state as recognized by those of skill in the art.

In this aspect of the invention, the methods generally comprise the step of selectively immobilizing a second portion of the nanoreporter while it is in an extended state. This can result in an immobilized nanoreporter that is extended between the first and the second portion. Remarkably, since the nanoreporter is selectively immobilized while

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 948

extended, that extension can be preserved in the immobilized nanoreporter. Generally, the first portion and the second portion of the nanoreporter are not the same.

The selective immobilization can be according to any technique for selective immobilization of a portion of a nanoreporter apparent to those of skill in the art. The selective immobilization can be through, for example, the formation of one or more covalent bonds or one or more non-covalent bonds, or both. Particular examples of selective immobilization techniques are described in the sections below. In particular embodiments, one or more binding pairs are used to immobilize the second portion of the nanoreporter.

The second portion can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any substrate judged to be useful for immobilization known to those of skill in the art. In certain embodiments, the second portion can be immobilized to another molecule. Further useful substrates include surfaces, membranes, beads, porous materials, electrodes, arrays and any other substrate apparent to those of skill in the art.

In another aspect, the present invention provides a composition comprising a selectively immobilized, extended nanoreporter. The compositions generally comprise a substrate and an extended nanoreporter selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the nanoreporter are immobilized onto the substrate, and the nanoreporter is in an extended state between the two portions. In certain embodiments, at least one portion of the nanoreporter is selectively immobilized onto the substrate. In certain embodiments, two or more portions of the nanoreporter are selectively immobilized onto the substrate. The nanoreporter can be extended and/or immobilized by any technique apparent to those of skill, including particularly the methods of the present invention.

In another aspect, the present invention provides methods for selectively immobilizing a nanoreporter in an oriented state. The nanoreporter can be any nanoreporter described above. In certain embodiments, the nanoreporter can be flexible, or in certain embodiments the nanoreporter can be rigid or semi-rigid. For the methods of this aspect of the invention, generally, a first portion of the nanoreporter is immobilized as described above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 949

With an immobilized first portion, the nanoreporter can be oriented by any technique for extending a nanoreporter apparent to those of skill in the art. In certain embodiments, the technique for orienting the nanoreporter is not critical for the methods of the invention. In certain embodiments, the technique for orienting the nanoreporter appropriate for the class of nanoreporter is determined according to the judgment of one of skill in the art. In certain embodiments, the nanoreporter is oriented by application of a force capable of orienting the nanoreporter. The force can be any force apparent to one of skill in the art for orienting the nanoreporter. Exemplary forces include gravity, hydrodynamic force, electromagnetic force and combinations thereof. Specific techniques for extending the nanoreporter are described in the subsections below.

The nanoreporter is in an oriented state if it would be recognized as oriented by one of skill in the art. In certain embodiments, the nanoreporter is in an oriented state when it is in the field of a force capable of orienting the nanoreporter. In certain embodiments, the nanoreporter is in an oriented state when its termini are arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the nanoreporter. In certain embodiments, a plurality of nanoreporters is in an oriented state when the termini of the nanoreporters are arranged in parallel, as recognized by those of skill in the art.

In this aspect of the invention, the methods generally comprise the step of selectively immobilizing a second portion of the nanoreporter while it is in an oriented state. This can result in an immobilized nanoreporter that is oriented between the first and the second portion. Remarkably, since the nanoreporter is selectively immobilized while extended, that orientation can be preserved in the immobilized nanoreporter. The selective immobilization can according to the methods described above.

In another aspect, the present invention provides a composition comprising a selectively immobilized, oriented nanoreporter. The compositions generally comprise a substrate and an oriented nanoreporter selectively immobilized onto the substrate. The substrate can be any substrate known to those of skill in the art. Exemplary substrates include those described in the sections below. At least two portions of the nanoreporter are immobilized onto the substrate, and the nanoreporter is in an oriented state between the two

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 950

portions. In certain embodiments, at least one portion of the nanoreporter is selectively immobilized onto the substrate. In certain embodiments, both portions of the nanoreporter are selectively immobilized onto the substrate. The nanoreporter can be oriented and/or immobilized by any technique apparent to those of skill, including particularly the methods of the present invention.

The methods and compositions of the present invention can be used for any purpose apparent to those of skill in the art. For instance, the immobilized and extended and/or oriented nanoreporter can be used as a label for a substrate on which the nanoreporter is immobilized. The primary sequence of the immobilized and extended and/or oriented nanoreporter can be identified by any technique apparent to those of skill. Advantageously, immobilization of the extended and/or oriented nanoreporter can facilitate such techniques. In certain embodiments, the immobilized and extended and/or oriented nanoreporter can be used to guide the manufacture of nanopaths, for example to create nanowires or nanocircuits. Further uses for the immobilized and extended and/or oriented nanoreporters are described in the sections below.

All terms used herein have their ordinary meanings to those of skill in the art unless indicated otherwise. The following terms shall have the following meanings.

As used herein, the term "binding pair" refers to first and second molecules or moieties that are capable of selectively binding to each other, *i.e.*, binding to each other with greater affinity than to other components in a composition. The binding between the members of the binding pair can be covalent or non-covalent. In certain embodiments, the binding is noncovalent. Exemplary binding pairs include immunological binding pairs (*e.g.*, any haptenic or antigenic compound in combination with a corresponding antibody or binding portion or fragment thereof, for example digoxigenin and anti-digoxigenin, fluorescein and anti-fluorescein, dinitrophenol and anti-dinitrophenol, bromodeoxyuridine and anti-bromodeoxyuridine, mouse immunoglobulin and goat anti-mouse immunoglobulin) and nonimmunological binding pairs (*e.g.*, biotin-avidin, biotin-streptavidin, hormone-hormone binding protein, receptor-receptor ligand (*e.g.*, acetylcholine receptor-acetylcholine or an analog thereof), IgG-protein A, lectin-carbohydrate, enzyme-enzyme cofactor, enzyme-enzyme inhibitor, complementary polynucleotide pairs capable of forming nucleic acid

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 951

duplexes, and the like). For instance, immunoreactive binding members may include antigens, haptens, aptamers, antibodies (primary or secondary), and complexes thereof, including those formed by recombinant DNA methods or peptide synthesis. An antibody may be a monoclonal or polyclonal antibody, a recombinant protein or a mixture(s) or fragment(s) thereof, as well as a mixture of an antibody and other binding members. Other common binding pairs include but are not limited to, biotin and avidin (or derivatives thereof), biotin and streptavidin, carbohydrates and lectins, complementary nucleotide sequences (including probe and capture nucleic acid sequences), complementary peptide sequences including those formed by recombinant methods, effector and receptor molecules, hormone and hormone binding protein, enzyme cofactors and enzymes, enzyme inhibitors and enzymes, and so forth.

"Selective binding" refers to the preferential binding of a pair of molecules or moieties for each other with respect to other molecules or moieties in a composition that would be recognized by one of skill in the art. In certain embodiments, a pair of molecules or moieties selectively binds when they preferentially bind each other compared to other molecules or moieties. Selective binding can include affinity or avidity, or both, of one molecule or moiety for another molecule or moiety. In particular embodiments, selective binding requires a dissociation constant (K_D) of less than about $1x10^{-5}$ M or less than about $1x10^{-6}$ M, $1x10^{-7}$ M, $1x10^{-8}$ M, $1x10^{-9}$ M, or $1x10^{-10}$ M. In contrast, in certain embodiments, non-selective binding has significantly less affinity, for example, a K_D greater than $1x10^{-3}$ M.

"Extended state" refers to a nanoreporter in a state that would be recognized as extended by one of skill in the art. In certain embodiments, a nanoreporter is in an extended state when it is extended relative to its native conformation in solution. In certain embodiments, a nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter. In certain embodiments, an extended state of a nanoreporter can be determined quantitatively. In such embodiments, those of skill in the art will recognize R as the end-to-end vector of the nanoreporter, *i.e.*, the distance between two termini of the nanoreporter, and <R> as the average end-to-end vector such that 95% of R will be within 2<R> in a solution deemed appropriate to one of skill in the art. Exemplary solutions include, for example, a dilute solution of the nanoreporter in water or in a pH

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 952

buffer. In particular embodiments, a nanoreporter is in an extended state when R is greater than 2.0<R>.

"Oriented state" refers to a nanoreporter in a state that would be recognized as oriented by one of skill in the art. In certain embodiments, a nanoreporter is in an oriented state when it is oriented relative to its native conformation in solution. In certain embodiments, the nanoreporter is oriented when it is arranged in parallel with the field of a force capable of orienting the nanoreporter. In certain embodiments, the nanoreporter is oriented when it is one of a plurality of nanoreporters that are arranged in parallel, as recognized by those of skill in the art.

METHODS OF SELECTIVE IMMOBILIZATION

As described above, the present invention provides methods for the selective immobilization of a nanoreporter in an extended state. The nanoreporter, once selectively immobilized, can be used for any purpose apparent to those of skill in the art.

In certain embodiments, the nanoreporter is any polymer known to those of skill in the art. For instance, the nanoreporter can be a polysaccharide, a polypeptide or a polynucleotide. Useful polynucleotides include ribonucleic acids, deoxyribonucleic acids and other polynucleotides known to those of skill in the art.

The nanoreporter can be of any size that is sufficient to allow extension and immobilization of the nanoreporter according to the methods of the invention. In certain embodiments when the nanoreporter is a polynucleotide, the nanoreporter can have a length of greater than 500 bp, greater than 750 bp, greater than 1 kb, greater than 1.5 kb, greater than 2.0 kb, greater than 2.5 kb, greater than 3.0 kb, greater than 4.0 kb or greater than 5.0 kb. In certain embodiments, when the nanoreporter is a polypeptide, the nanoreporter can have a size of greater than 50 amino acids, greater than 100 amino acids, greater than 200 amino acids, greater than 300 amino acids, greater than 400 amino acids, greater than 500 amino acids, greater than 750 amino acids, greater than 1000 amino acids, greater than 1500 amino acids, greater than 2000 amino acids, greater than 2500 amino acids, greater than 3000 amino acids, greater than 4000 amino acids or greater than 5000 amino acids. In certain embodiments, when the nanoreporter is a polysaccharide, the nanoreporter can have a size of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 953

greater than 50 saccharides, greater than 100 saccharides, greater than 200 saccharides, greater than 300 saccharides, greater than 400 saccharides, greater than 500 saccharides, greater than 750 saccharides, greater than 1000 saccharides, greater than 1500 saccharides, greater than 2000 saccharides, greater than 2500 saccharides, greater than 3000 saccharides, greater than 4000 saccharides or greater than 5000 saccharides.

The nanoreporter can be a native nanoreporter as understood by those of skill in the art, or the nanoreporter can be a non-native nanoreporter. In certain embodiments, when the nanoreporter is a polypeptide, the nanoreporter can comprise only naturally occurring amino acids, or the nanoreporter can comprise naturally occurring amino acids and non-naturally occurring amino acids. The other amino acids can be any amino acids, or derivatives or analogs thereof, known to those of skill in the art. In certain embodiments, when the nanoreporter is a polynucleotide, the polynucleotide can comprise only naturally occurring nucleotides, or the polynucleotide can comprise naturally occurring nucleotides and non-naturally occurring nucleotides. In certain embodiments, when the nanoreporter is a polysaccharide, the polysaccharide can comprise only naturally occurring saccharides, or the polysaccharide can comprise naturally occurring saccharides and non-naturally occurring saccharides. In certain embodiments, the polymers can comprise only non-natural monomers. In further embodiments, the nanoreporter can comprise a plurality of classes of monomers, such as amino acids, nucleotides and/or saccharides.

In certain embodiments, the nanoreporter comprises only one primary, covalently linked chain of monomers. For instance, when the nanoreporter is a polypeptide, in certain embodiments, the nanoreporter comprises only one primary amino acid chain. When the nanoreporter is a polynucleotide, in certain embodiments, the nanoreporter is single stranded. In further embodiments, the nanoreporter comprises two primary, covalently linked chains of monomers. For instance, when the nanoreporter is a polypeptide, in certain embodiments, the nanoreporter comprises two primary amino acid chains. When the nanoreporter is a polynucleotide, in certain embodiments, the nanoreporter comprises two polynucleotide strands; in certain embodiments, the nanoreporter can be double stranded, in part or in whole. In further embodiments, the nanoreporter comprises three or more primary, covalently linked chains of monomers. For instance, when the nanoreporter is a polypeptide,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 954

in certain embodiments, the nanoreporter comprises three primary amino acid chains. When the nanoreporter is a polynucleotide, in certain embodiments, the nanoreporter comprises three polynucleotide strands. For instance, the nanoreporter can comprise three strands F1, X and F2 where a portion of strand X is complementary to strand F1 and a portion of strand X is complementary to strand F2. An example is illustrated in Figure 13A. In certain embodiments, the nanoreporter comprises more than three primary, covalently linked chains of monomers.

Advantageously, a nanoreporter of the invention can comprise one or more labels that facilitate the detection, imaging or identification of the nanoreporter by techniques known to those of skill in the art. The label can be any detectable moiety known to those of skill in the art. Exemplary labels for nanoreporters include detectable isotopes, radioisotopes, fluors, dyes, enzymes, ligands, receptors, antigens, antibodies, lectins, carbohydrates, nucleotide sequences, and any other detectable label apparent to those of skill in the art.

In certain embodiments, a polynucleotide is a polymer of natural (*e.g.*, A, G, C, T, U) or synthetic nucleobases, or a combination of both. The backbone of the polynucleotide can be composed entirely of "native" phosphodiester linkages, or it may contain one or more modified linkages, such as one or more phosphorothioate, phosphorodithioate, phosphoramidate or other modified linkages. As a specific example, a polynucleotide may be a peptide nucleic acid (PNA), which contains amide interlinkages. Additional examples of synthetic bases and backbones that can be used in conjunction with the invention, as well as methods for their synthesis can be found, for example, in U.S. Patent No. 6,001,983; Uhlman & Peyman, 1990, *Chemical Review* 90(4):544-584; Goodchild, 1990, *Bioconjugate Chem.* 1(3):165-186; Egholm *et al.*, 1992, *J. Am. Chem.* Soc. 114:1895-1897; Gryaznov *et al.*, *J. Am. Chem. Soc.* 116:3143-3144, as well as the references cited in all of the above. Common synthetic nucleobases of which polynucleotides may be composed include 3-methlyuracil, 5,6-dihydrouracil, 4 thiouracil, 5 bromouracil, 5-thorouracil, 5-iodouracil, 6-dimethyl aminopurine, 6-methyl aminopurine, 2-aminopurine, 2,6-diamino purine, 6-amino-8-bromopurine, inosine, 5-methylcytosine, 7-deazaadenine, and 7-deazaguanosine. Additional non-limiting examples of synthetic nucleobases of which the target nucleic acid may be composed can be found in Fasman, *CRC Practical Handbook of Biochemistry and*

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 955

Molecular Biology, 1985, pp. 385-392; *Beilstein's Handbuch der Organischen Chemie*, Springer Verlag, Berlin and Chemical Abstracts, all of which provide references to publications describing the structures, properties and preparation of such nucleobases.

The nanoreporter can be prepared according to any technique apparent to those of skill in the art. Advantageously, nanoreporters according to the invention can comprise labels and/or members of binding pairs, as described in the sections below, that can be used to facilitate preparation and/or purification of the nanoreporter. In addition, certain nanoreporters of the invention are capable of forming complexes with molecules that comprise members of binding pairs, as described below. These complexes can be used to facilitate preparation and/or purification of the nanoreporter or complex.

IMMOBILIZATION OF FIRST PORTION

In the methods of the invention, a first portion of the nanoreporter is immobilized.

Generally, the first portion is immobilized if it would be recognized as immobilized by one of skill in the art. The first portion can be immobilized by any technique apparent to those of skill in the art. In certain embodiments, the technique for immobilization of the first portion of the nanoreporter is not critical for the methods of the invention.

The first portion of the nanoreporter can be at any location in the nanoreporter. In certain embodiments, the first portion is at a terminus of the nanoreporter. For the purposes of the invention, a portion of a nanoreporter can be "at a terminus" when it is less than five, four, three, two, one or zero monomers from a terminus of the nanoreporter. Of course, although many nanoreporters have two termini, the methods of the invention are applicable to nanoreporters having more than two termini and to nanoreporters having one or zero termini, *e.g.*, circular nanoreporters. In certain embodiments, the first portion is not at a terminus of the nanoreporter.

The nanoreporter can be immobilized onto any substrate apparent to those of skill in the art. The substrate can be any moiety to which the nanoreporter can be immobilized without limitation. In certain embodiments, the substrate is a surface, membrane, bead, porous material, electrode or array.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 956

In certain embodiments, the first portion of the nanoreporter can be immobilized non-selectively. In further embodiments, the first portion of the nanoreporter can be immobilized selectively. In advantageous embodiments, after the first portion of the nanoreporter is immobilized, some portion of the nanoreporter should be free to move sufficiently so that the nanoreporter can be extended in the following steps of the method. In particular, in certain embodiments, when the first portion of the nanoreporter is immobilized non-selectively, it is important that the entire nanoreporter not be immobilized non-selectively to an extent that prevents extension of any portion of the nanoreporter.

The immobilization can be by any interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain embodiments, the immobilization can be via electrostatic interaction with an electrode. In further embodiments, the immobilization is via electrostatic interaction with a substrate other than the electrode.

In certain embodiments, the first portion of the nanoreporter comprises a first member of a binding pair. The first member of the binding pair can be covalently bound to the first portion of the nanoreporter, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. In useful embodiments, the substrate onto which the first portion of the nanoreporter is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound. Figure 12A illustrates a nanoreporter that comprises a moiety F1 that is capable of selectively binding a moiety of the substrate. Moiety F1 can be, for example, biotin, capable of binding, for example, a substrate coated with avidin.

In certain embodiments, the first portion of the nanoreporter can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins, and antibodies. The first portion of the nanoreporter would comprise a binding moiety capable of binding with the binding moiety of

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 957

the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, surfaces comprising epoxy, aldehyde, gold, hydrazide, sulfhydryl, NHS-ester, amine, thiol, carboxylate, maleimide, hydroxymethyl phosphine, imidoester, isocyanate, hydroxyl, pentafluorophenyl-ester, psoralen, pyridyl disulfide or vinyl sulfone, or mixtures thereof. Such surfaces can be obtained from commercial sources or prepared according to standard techniques.

In advantageous embodiments, the first portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. Similarly, a polypeptide nanoreporter can comprise a biotinylated amino acid residue. The substrate comprising avidin can be any substrate comprising avidin known to those of skill in the art. Useful substrates comprising avidin are commercially available including TB0200 (Accelr8), SAD6, SAD20, SAD100, SAD500, SAD2000 (Xantec), SuperAvidin (Array-It), streptavidin slide (catalog #MPC 000, Xenopore) and STREPTAVIDINnslide (catalog #439003, Greiner Bio-one).

In certain embodiments, the first portion of the nanoreporter can comprise a nucleotide sequence that is capable of selectively binding a nucleotide sequence on the substrate.

In further embodiments, the first portion of the nanoreporter can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are commercially available including Optiarray-biotin (Accelr8), BD6, BD20, BD100, BD500 and BD2000 (Xantec).

In further embodiments, the first portion of the nanoreporter is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, a first portion of the nanoreporter can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. Figure 13A illustrates a nanoreporter that is capable of selectively binding a second molecule X that is capable of selectively binding a third molecule that comprises F1.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 958

F1 is capable of selectively binding a moiety on a substrate. Figure 13B illustrates a nanoreporter that is capable of selectively binding a second molecule that comprises F1, and F1 is capable of selectively binding a moiety on a substrate.

In further embodiments, the first portion of the nanoreporter can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Exemplary useful substrates comprising reactive groups include those that comprise a reactive moiety selected from the group consisting of succinamides, amines, aldehydes, epoxies and thiols. The first portion of the nanoreporter would comprise a reactive moiety capable of reacting with the reactive moiety of the substrate. Exemplary useful substrates comprising reactive moieties include, but are not limited to, OptArray-DNA NHS group (Accelr8), Nexterion Slide AL (Schott) and Nexterion Slide E (Schott).

In certain embodiments, the first portion of the nanoreporter can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the first portion of the nanoreporter could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

In further embodiments, the first portion of the nanoreporter can be immobilized to the substrate via other binding pairs apparent to those of skill in the art.

EXTENSION OF THE NANOREPORTER

In certain methods of the invention, the nanoreporter is in an extended state. Generally, any nanoreporter is in an extended state if it would be recognized as such by one of skill in the art.

In certain embodiments, the nanoreporter is in an extended state when it is in the field of a force capable of extending the nanoreporter under conditions suitable for extending the nanoreporter. Such forces and conditions should be apparent to those of skill in the art. For instance, many nanoreporters can be extended by hydrodynamic force or by

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 959

gravity, and many charged nanoreporters can be extended by electromagnetic force. In certain embodiments, the force can be applied to the nanoreporter indirectly. For instance, the nanoreporter can comprise or can be linked, covalently or noncovalently, to a moiety capable of being moved by a force. In certain embodiments, the nanoreporter can be linked to a moiety.

In certain embodiments, the force is an electromagnetic force. For instance, when the nanoreporter is charged, such as a polynucleotide, the nanoreporter can be extended in an electric or magnetic field. The field should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter in an electric or magnetic field are described in Matsuura *et al.*, 2002, *J Biomol Struct Dyn.* 20(3):429-36; Ferree & Blanch, 2003, *Biophys J.* 85(4):2539-46; Stigter & Bustamante, 1998, *Biophys J.* 1998 75(3):1197-210; Matsuura *et al.*, 2001, *Nucleic Acids Res.* 29(16); Ferree & Blanch, 2004, *Biophys J.* 87(1):468-75; the contents of which are hereby incorporated by reference in their entirety.

In certain embodiments, the force is a hydrodynamic force. For instance, many nanoreporters, including polysaccharides, polypeptides, and polynucleotides, can be extended in the field of a moving fluid. The hydrodynamic force should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter in a hydrodynamic field are described in Bensimon *et al.*, 1994, *Science* 265:2096-2098; Henegariu *et al.*, 2001, *BioTechniques* 31:246-250; Kraus *et al.*, 1997, *Human Genetics* 99:374-380; Michalet *et al.*, 1997, *Science* 277:1518-1523; Yokota *et al.*, 1997, *Nucleic Acids Res.* 25(5):1064-70; Otobe *et al.*, 2001, *Nucleic Acids Research* 29:109; Zimmerman & Cox, 1994, *Nucleic Acids Res.* 22(3):492-7, and U.S. Patent Nos. 6,548,255; 6,344,319; 6,303,296; 6,265,153; 6,225,055; 6,054,327; and 5,840,862, the contents of which are hereby incorporated by reference in their entirety.

In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force to extend the nanoreporter. In certain embodiments, the force should be strong enough to extend the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for extending a nanoreporter with gravity are described in Michalet *et al.*, 1997, *Science*

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 960

277:1518-1523; Yokota *et al.*, 1997, *Nucleic Acids Res.* 25(5):1064-70; Kraus *et al.*, 1997, *Human Genetics* 99:374-380, the contents of which are hereby incorporated by reference in their entirety.

In particular embodiments, the force is applied through a moving meniscus. Those of skill in the art will recognize that a moving meniscus can apply various forces to a nanoreporter including hydrodynamic force, surface tension and any other force recognized by those of skill in the art. The meniscus can be moved by any technique apparent to those of skill in the art including evaporation and gravity. Exemplary techniques for extending a nanoreporter with a moving meniscus are described in, for example, U.S. Patent Nos. 6,548,255; 6,344,319; 6,303,296; 6,265,153; 6,225,055; 6,054,327; and 5,840,862, the contents of which are hereby incorporated by reference in their entireties.

In particular embodiments, the nanoreporter can be extended by an optical trap or optical tweezers. For instance, the nanoreporter can comprise or can be linked, covalently or noncovalently, to a particle capable of being trapped or moved by an appropriate source of optical force. Useful techniques for moving particles with optical traps or optical tweezers are described in Ashkin *et al.*, 1986, *Optics Letters* 11:288-290; Ashkin *et al.*, 1987, *Science* 235:1517-1520; Ashkin *et al.*, *Nature* 330:769-771; Perkins *et al.*, 1994, *Science* 264:822-826; Simmons *et al.*, 1996, *Biophysical Journal* 70:1813-1822; Block *et al.*, 1990, *Nature* 348:348-352; and Grier, 2003, *Nature* 424:810-816; the contents of which are hereby incorporated by reference in their entireties.

In certain embodiments, the nanoreporter can be extended by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain nanoreporters are extended by a combination of an electric field and hydrodynamic force.

The nanoreporter is extended when it would be recognized as extended by one of skill in the art according to standard criteria for extension of a nanoreporter. In certain embodiments, the nanoreporter is extended when it loses most of its tertiary structural features as recognized by those of skill in the art. In certain embodiments, the nanoreporter is extended when it loses most of its secondary structural features as recognized by those of skill in the art. In certain embodiments, the nanoreporter is extended when its primary

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 961

structural features are detectable in sequence when imaged according to standard techniques. Exemplary imaging techniques are described in the examples below.

In certain embodiments, an extended state of a nanoreporter can be recognized by comparing its hydrodynamic radius to its average hydrodynamic radius when free in dilute solution. For instance, in certain embodiments, a nanoreporter, or portion thereof, is extended when its hydrodynamic radius is more than about double its average hydrodynamic radius in dilute solution. More quantitatively, R represents the hydrodynamic radius of the nanoreporter, or portion thereof, and <R> represents the average hydrodynamic radius of the nanoreporter, or portion thereof, in dilute solution. The average <R> should be calculated such that R for the nanoreporter, or portion thereof, when unbound in dilute solution is less than 2<R> 95% of the time. In certain embodiments, a nanoreporter, or portion thereof, is in an extended state when R is greater than 1.5<R>, greater than 1.6<R>, greater than 1.7<R>, greater than 1.8<R>, greater than 1.9<R>, greater than 2.0<R>, greater than 2.1<R>, greater than 2.2<R>, greater than 2.3<R>, greater than 2.4<R>, greater than 2.5<R> or greater than 3.0<R>. In particular embodiments, a nanoreporter, or portion thereof, is in an extended state when R is greater than 2.0<R>.

ORIENTATION OF THE NANOREPORTER

In certain methods of the invention, the nanoreporter is in an oriented state. Generally, any nanoreporter is in an oriented state if it would be recognized as such by one of skill in the art.

In certain embodiments, the nanoreporter is in an oriented state when it is in the field of a force capable of orienting the nanoreporter under conditions suitable for orienting the nanoreporter. Such forces and conditions should be apparent to those of skill in the art.

In certain embodiments, the force is an electromagnetic force. For instance, when the nanoreporter is charged, such as a polynucleotide, the nanoreporter can be oriented in an electric or magnetic field. The field should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter in an electric or magnetic field are described above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 962

In certain embodiments, the force is a hydrodynamic force. For instance, many nanoreporters, including polysaccharides, polypeptides, and polynucleotides, can be oriented in the field of a moving fluid. The hydrodynamic force should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter in hydrodynamic field are described above.

In certain embodiments, the force is gravity. In advantageous embodiments, the force of gravity can be combined with, for example, hydrodynamic force to orient the nanoreporter. In certain embodiments, the force should be strong enough to orient the nanoreporter according to the judgment of one of skill in the art. Exemplary techniques for orienting a nanoreporter with gravity are described above.

In certain embodiments, the nanoreporter can be oriented by combinations of the above forces that are apparent to those of skill in the art. In the examples, below, certain nanoreporters are oriented by a combination of an electric field and hydrodynamic force.

The nanoreporter is oriented when it would be recognized as oriented by one of skill in the art according to standard criteria for orientation of a nanoreporter. In certain embodiments, the nanoreporter is oriented when it is arranged in parallel, as recognized by those of skill in the art, with the field of a force capable of orienting the nanoreporter. In certain embodiments, the nanoreporter is oriented when it is one of a plurality of nanoreporters that are arranged in parallel, as recognized by those of skill in the art. For instance, a plurality of nanoreporters can be oriented when the vector from a first terminus to a second terminus of a nanoreporter is parallel, as recognized by those of skill in the art, to the vectors between corresponding termini of other nanoreporters in the plurality.

SELECTIVE IMMOBILIZATION OF SECOND PORTION OF NANOREPORTER

As discussed above, in the methods of the invention, a second portion of the nanoreporter is selectively immobilized. The second portion of the nanoreporter can be any portion of the nanoreporter that is not identical to the first portion of the nanoreporter.

In some embodiments, the second portion of the nanoreporter does not overlap any part of the first portion of the nanoreporter.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 963

In certain embodiments, the present invention provides methods that comprise the single step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended or oriented state, and while a first portion of the nanoreporter is immobilized. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension or orientation of the nanoreporter are described in detail in the sections above.

In certain embodiments, the present invention provides methods that comprise the step of extending a nanoreporter, while a first portion of the nanoreporter is immobilized, and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension of the nanoreporter are described in detail in the sections above.

In certain embodiments, the present invention provides methods that comprise the step of immobilizing a first portion of a nanoreporter, the step of extending the nanoreporter while the first portion is immobilized and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an extended state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for extension of the nanoreporter are described in detail above.

In certain embodiments, the present invention provides methods that comprise the step of orienting a nanoreporter, while a first portion of the nanoreporter is immobilized, and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an oriented state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for orienting the nanoreporter are described in detail in the sections above.

In certain embodiments, the present invention provides methods that comprise the step of immobilizing a first portion of a nanoreporter, the step of orienting the nanoreporter while the first portion is immobilized and the step of selectively immobilizing a second portion of a nanoreporter while the nanoreporter is in an oriented state. Exemplary methods for immobilization of the first portion of the nanoreporter, and for orienting the nanoreporter are described in detail above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 964

The selective immobilization of the second portion of the nanoreporter can follow any technique for selective immobilization of a nanoreporter apparent to those of skill in the art. Significantly, in advantageous embodiments of the invention, the second portion of the nanoreporter is not immobilized non-selectively. Selective immobilization can allow the nanoreporter to be immobilized while in a fully extended state or nearly fully extended state. Selective immobilization can also allow the nanoreporter to be immobilized in an oriented manner. In other words, the first portion and second portion of the nanoreporter can be immobilized along the direction of the field or fields used to extend the nanoreporter, with the first portion preceding the second portion in the field. When a plurality of nanoreporters are immobilized, the plurality can be uniformly oriented along the field.

The second portion of the nanoreporter can be at any location in the nanoreporter. In certain embodiments, the second portion is at a terminus of the nanoreporter. In certain embodiments, the second portion is not at a terminus of the nanoreporter. In certain embodiments, the first portion, described in the sections above, is at one terminus of the nanoreporter, and the second portion is at another terminus of the nanoreporter.

As discussed above, the second portion of the nanoreporter is immobilized selectively. The immobilization can be by any selective interaction with the substrate apparent to those of skill in the art. The immobilization can be via electrostatic or ionic interaction, via one or more covalent bonds, via one or more non-covalent bonds or combinations thereof. In certain embodiments, the immobilization can be via electrostatic interaction with an electrode. In further embodiments, the immobilization is via electrostatic interaction with a substrate other than the electrode.

If the first portion and the second portion of the nanoreporter are selectively immobilized to the same substrate, the techniques of selective immobilization should of course be compatible with the substrate. In particular embodiments, the techniques of immobilization are the same. For instance, on a substrate coated with avidin, both the first and second portion of the nanoreporter can be immobilized selectively via biotin-avidin interactions. However, as will be apparent to those of skill in the art, the same interaction need not be used at both the first and second portions for immobilization on the same

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 965

substrate. For instance, the substrate can comprise multiple moieties capable of selective binding, or the first portion can be immobilized non-selectively, or other techniques apparent to those of skill in the art.

In certain embodiments, the second portion of the nanoreporter comprises a first member of a binding pair. The second member of the binding pair can be covalently bound to the second portion of the nanoreporter, or they can be non-covalently bound. Useful covalent bonds and non-covalent bonds will be apparent to those of skill in the art. In useful embodiments, the substrate onto which the second portion of the nanoreporter is bound will comprise a second member of the binding pair. The substrate can be covalently bound to the second member, or they can be non-covalently bound.

In certain embodiments, the second portion of the nanoreporter can comprise a member of a binding pair that is capable of binding with a member of a binding pair on the substrate to form one or more non-covalent bonds. Exemplary useful substrates include those that comprise a binding moiety selected from the group consisting of ligands, antigens, carbohydrates, nucleic acids, receptors, lectins, and antibodies such as those described in the sections above.

In advantageous embodiments, the second portion of the nanoreporter can be immobilized to the substrate via an avidin-biotin binding pair. In certain embodiments, the nanoreporter can comprise a biotin moiety in its first portion. For instance, a polynucleotide nanoreporter can comprise a biotinylated nucleotide residue. Similarly, a polypeptide nanoreporter can comprise a biotinylated amino acid residue. Useful substrates comprising avidin are described in the sections above.

In further embodiments, the second portion of the nanoreporter can comprise avidin, and the substrate can comprise biotin. Useful substrates comprising biotin are described in the sections above.

In further embodiments, the second portion of the nanoreporter can comprise a member of a binding pair that is capable of reacting with a member of a binding pair on the substrate to form one or more covalent bonds. Exemplary useful substrates comprising reactive groups are described in the sections above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 966

In certain embodiments, the second portion of the nanoreporter can comprise a reactive moiety that is capable of being bound to the substrate by photoactivation. The substrate could comprise the photoreactive moiety, or the second portion of the nanoreporter could comprise the photoreactive moiety. Some examples of photoreactive moieties include aryl azides, such as *N*-((2-pyridyldithio)ethyl)-4-azidosalicylamide; fluorinated aryl azides, such as 4-azido-2,3,5,6-tetrafluorobenzoic acid; benzophenone-based reagents, such as the succinimidyl ester of 4-benzoylbenzoic acid; and 5-Bromo-deoxyuridine.

In further embodiments, the second portion of the nanoreporter can be immobilized to the substrate via other binding pairs described in the sections above.

In further embodiments, the second portion of the nanoreporter is capable of forming a complex with one or more other molecules that, in turn, are capable of binding, covalently or non-covalently, a binding moiety of the substrate. For instance, the second portion of the nanoreporter can be capable of selectively binding another molecule that comprises, for instance, a biotin moiety that is capable of selectively binding, for instance, an avidin moiety of the substrate. Figure 12B illustrates a nanoreporter of selectively binding a second molecule that comprises F3 that is, in turn, capable of selectively binding a moiety on a substrate. The interaction between the second portion of the nanoreporter and the molecule that comprises F3 can be mediated, for example, by an antigen-antibody interaction.

Figures 14A and 14B illustrate the selective immobilization of a nanoreporter according to methods of the present invention. In Figure 14A, a first portion of the nanoreporter comprises binding moiety F1 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F1 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. The nanoreporter of Figure 14A is extended by a force as described in the sections above. In Figure 14B, the force is an electrical potential. While extended, the nanoreporter is contacted with molecules comprising binding moiety F2 that is capable of selectively binding a moiety on the illustrated substrate S. Binding moiety F2 can be, for instance, biotin, and substrate S can be coated with, for instance, avidin. Significantly, up to three molecules comprising F2 are capable of selectively binding to a second portion of the nanoreporter to selectively immobilize it in its extended state. As illustrated, the molecules comprise a second binding moiety that selectively binds a repeated binding moiety

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 967

of the nanoreporter. The binding moieties can be, for instance, complementary nucleic acid sequences, as illustrated in Figure 14B. The resulting nanoreporter is selectively immobilized in an extended state and should remain extended even when the force is removed. The selectively immobilized, extended nanoreporter can be used for any purpose apparent to those of skill in the art.

IMMOBILIZATION OF TWO PORTIONS OF AN EXTENDED OR ORIENTED NANOREPORTER

In certain embodiments, the present invention provides methods for selective immobilization of a first portion and a second portion of a nanoreporter that is in an extended or oriented state. Significantly, according to these methods of the invention, the nanoreporter need not be immobilized prior to application of a force capable of extending or orienting the nanoreporter.

In these methods, the nanoreporter is extended or oriented, or both, by a force capable of extending or orienting the nanoreporter. Such forces are described in detail in the sections above. In particular embodiments, the force is a force capable of extending or orienting the nanoreporter while maintaining the nanoreporter in one location, *i.e.*, a force capable of extending or orienting without substantially moving the nanoreporter. Exemplary forces include oscillating electromagnetic fields and oscillating hydrodynamic fields. In a particular embodiment, the force is an oscillating electrical field. Exemplary techniques for extending or orienting a nanoreporter in an oscillating electric field are described in Asbury *et al.*, 2002, *Electrophoresis* 23(16):2658-66; Kabata *et al.*, 1993, *Science* 262(5139):1561-3; and Asbury and van den Engh, 1998, *Biophys J.* 74:1024-30, the contents of which are hereby incorporated by reference in their entirety.

In the methods, the nanoreporter is immobilized at a first portion and at a second portion while extended or oriented. Both the first portion and the second portion can be immobilized non-selectively, both can be immobilized selectively, or one can be immobilized selectively and the other non-selectively. Techniques for immobilization of the first portion and second portion are described in detail in the sections above.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 968

SUBSTRATE FOR IMMOBILIZATION

In the methods of the invention, the substrate for immobilization can be any substrate capable of selectively binding the nanoreporter apparent to those of skill in the art. Further, in certain aspects, the present invention provides compositions comprising a selectively immobilized nanoreporter in an extended state. The compositions comprise a substrate, as described herein, having immobilized thereto a nanoreporter in an extended state. The nanoreporter can be, of course, immobilized according to a method of the invention.

The only requirement of the substrate is that it be capable of selectively binding the second portion of the nanoreporter as described above. Thus, the substrate can be a filter or a membrane, such as a nitrocellulose or nylon, glass, a polymer such as polyacrylamide, a gel such as agarose, dextran, cellulose, polystyrene, latex, or any other material known to those of skill in the art to which capture compounds can be immobilized. The substrate can be composed of a porous material such as acrylic, styrene methyl methacrylate copolymer and ethylene/acrylic acid.

The substrate can take on any form so long as the form does not prevent selective immobilization of the second portion of the nanoreporter. For instance, the substrate can have the form of a disk, slab, strip, bead, submicron particle, coated magnetic bead, gel pad, microtiter well, slide, membrane, frit or other form known to those of skill in the art. The substrate is optionally disposed within a housing, such as a chromatography column, spin column, syringe barrel, pipette, pipette tip, 96 or 384 well plate, microchannel, capillary, etc., that aids the flow of liquid over or through the substrate.

The nanoreporter can be immobilized on a single substrate or on a plurality of substrates. For instance, in certain embodiments, the first and second portions of nanoreporter are immobilized on the same substrate, as recognized by those of skill in the art. In certain embodiments, the first portion of the nanoreporter can be immobilized on a first substrate while the second portion of the nanoreporter can be immobilized on a second substrate, distinct from the first.

The substrate can be prepared according to any method apparent to those of skill in the art. For a review of the myriad techniques that can be used to activate exemplary

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 969

substrates of the invention with a sufficient density of reactive groups, see *Wiley Encyclopedia of Packaging Technology*, 2d Ed., Brody & Marsh, Ed., "Surface Treatment," pp. 867-874, John Wiley & Sons (1997), and the references cited therein. Chemical methods suitable for generating amino groups on silicon oxide substrates are described in Atkinson & Smith, "Solid Phase Synthesis of Oligodeoxyribonucleotides by the Phosphite Triester Method," In: *Oligonucleotide Synthesis: A Practical Approach*, M. J. Gait, Ed., 1984, IRL Press, Oxford, particularly at pp. 45-49 (and the references cited therein); chemical methods suitable for generating hydroxyl groups on silicon oxide substrates are described in Pease *et al.*, 1994, *Proc. Natl. Acad. Sci. USA 91*:5022-5026 (and the references cited therein); chemical methods for generating functional groups on polymers such as polystyrene, polyamides and grafted polystyrenes are described in Lloyd Williams *et al.*, 1997, Chemical Approaches to the Synthesis of Peptides and Proteins, Chapter 2, CRC Press, Boca Raton, FL (and the references cited therein).

Exemplary useful substrates include surfaces coated with streptavidin, *e.g.*, Accelr8 TB0200. Further useful substrates include surfaces coated with N-hydroxysuccinamide that are capable of reacting with a portion of a nanoreporter that comprises an amine. One such surface is OptArray-DNA (Accelr8). Additional useful surfaces are coated with aldehyde (*e.g.*, Nexterion Slide AL, Schott) and surfaces coated with epoxy (*e.g.*, Nexterion Slide E, Schott). Another useful surface is a biotinylated BSA coated surface useful for selective immobilization of a portion of a nanoreporter that comprises avidin or streptavidin.

METHODS OF USING SELECTIVELY IMMOBILIZED, EXTENDED OR ORIENTED NANOREPORTERS

In certain embodiments, the selectively immobilized, elongated nanoreporters can be used to create macromolecular barcodes for the purposes of separation and sequential detection of labels. These labels spaced along the molecule provide a unique code that can be read when the nanoreporter is extended and immobilized. Extension and selective immobilization can facilitate the decoding of the macromolecular barcode.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 970

The selectively immobilized, elongated nanoreporters can be used in any context where detection or imaging of a nanoreporter might be useful. They can be used for diagnostic, prognostic therapeutic and screening purposes. For instance, they can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease. They can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample. The compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state. In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient.

KITS COMPRISING SELECTIVELY IMMOBILIZED EXTENDED OR ORIENTED NANOREPORTERS

The invention further provides kits comprising one or more components of the invention. The kits can comprise, for example, a substrate according to the invention and one or more extended or oriented, or both, nanoreporters selectively immobilized on the substrate. The kits can be used for any purpose apparent to those of skill in the art, including those described above.

In certain embodiments, the present invention also provides kits useful for the extension and selective immobilization of nanoreporters. The kits can comprise a substrate for immobilization and one or more binding partners to facilitate extension or immobilization of a nanoreporter. The binding partners could, in certain embodiments, comprise a moiety useful for extension of the nanoreporter in an appropriate force. In certain embodiments, the binding partners could facilitate immobilization or selective immobilization of the nanoreporter to the surface. In further embodiments, the kit could comprise a nanoreporter for extension and immobilization. In further embodiments, the kit could comprise a device capable of extending the nanoreporter.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 971

DETECTION OF NANOREPORTERS

Nanoreporters are detected by any means available in the art that is capable of detecting the specific signals on a given nanoreporter. Where the nanoreporter is fluorescently labeled, suitable consideration of appropriate excitation sources may be investigated. Possible sources may include but are not limited to arc lamp, xenon lamp, lasers, light emitting diodes or some combination thereof. The appropriate excitation source is used in conjunction with an appropriate optical detection system, for example an inverted fluorescent microscope, an epi-fluorescent microscope or a confocal microscope. Preferably, a microscope is used that can allow for detection with enough spatial resolution to determine the sequence of the spots on the nanoreporter.

MICROSCOPE AND OBJECTIVE LENS SELECTION.

The major consideration regarding the microscope objective lens is with the optical resolution, which is determined by its numerical aperture (NA). Generally, the larger the NA, the better the optical resolution. The required NA is preferably at least 1.07 based on the relationship of $\delta = 0.61\lambda/NA$ (δ = optical resolution and λ = wavelength). The amount of light that is collected by an objective is determined by NA^4/Mag^2 (Mag = magnification of the objective). Therefore, in order to collect as much light as possible, objectives with high NA and low magnifications should be selected.

CCD CAMERA SELECTION AND IMAGE CAPTURE TECHNIQUES.

When selecting a CCD camera, the first consideration is the pixel size, which partially determines the final resolution of the imaging system. Optimally the optical resolution should not be compromised by the CCD camera. For example, if the optical resolution is 210-300 nm, which corresponds to 12.6-18 μm on a CCD chip after a 60 x magnification, in order to resolve and maintain the optical resolution there should be at least two pixels to sample each spot. Or the pixel size of the CCD chip should be at most 6.3-9 μm.

The second consideration is detection sensitivity which can be determined by many factors that include but are not limited to pixel size, quantum efficiency, readout noise

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 972

and dark noise. To achieve high sensitivity, select a qualitative camera with big pixel size (which can give big collection area), high quantum efficiency and low noise. An exemplary camera with these criteria is the Orca-Ag camera from Hamamatsu Inc. The chip size is 1344 x 1024 pixels; when using the 60x objective, the field of view is 144 x 110 μm^2.

COMPUTER SYSTEMS

The invention provides computer systems that may be used to computerize nanoreporter image collection, nanoreporter identification and/or decoding of the nanoreporter code. Specifically, the invention provides various computer systems comprising a processor and a memory coupled to the processor and encoding one or more programs. The computer systems can be connected to the microscopes employed in imaging the nanoreporter, allowing imaging, identification and decoding the nanoreporter, as well as storing the nanoreporter image and associated information, by a single apparatus. The one or more programs encoded by the memory cause the processor to perform the methods of the invention.

In still other embodiments, the invention provides computer program products for use in conjunction with a computer system (*e.g.*, one of the above-described computer systems of the invention) having a processor and a memory connected to the processor. The computer program products of the invention comprise a computer readable storage medium having a computer program mechanism encoded or embedded thereon. The computer program mechanism can be loaded into the memory of the computer and cause the processor to execute the steps of the methods of the invention.

The methods described in the previous subsections can preferably be implemented by use of the following computer systems, and according to the following methods. An exemplary computer system suitable for implementation of the methods of this invention comprises internal components and being linked to external components. The internal components of this computer system include a processor element interconnected with main memory. For example, the computer system can be an Intel Pentium-based processor of 200 MHz or greater clock rate and with 32 MB or more of main memory.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 973

The external components include mass storage. This mass storage can be one or more hard disks which are typically packaged together with the processor and memory. Such hard disks are typically of 1 GB or greater storage capacity. Other external components include user interface device, which can be a monitor and a keyboard, together with pointing device, which can be a "mouse", or other graphical input devices (not illustrated). Typically, the computer system is also linked to a network link, which can be part of an Ethernet link to other local computer systems, remote computer systems, or wide area communication networks, such as the Internet. This network link allows the computer system to share data and processing tasks with other computer systems.

Loaded into memory during operation of this system are several software components, which are both standard in the art and special to the instant invention. These software components collectively cause the computer system to function according to the methods of the invention. The software components are typically stored on mass storage. A first software component is an operating system, which is responsible for managing the computer system and its network interconnections. This operating system can be, for example, of the Microsoft Windows® family, such as Windows 95, Windows 2000, or Windows XP, or, alternatively, a Macintosh operating system, a Linux operating system or a Unix operating system. A second software component may include common languages and functions conveniently present in the system to assist programs implementing the methods specific to this invention. Languages that can be used to program the analytic methods of the invention include, for example, C, C++, JAVA, and, less preferably, FORTRAN, PASCAL, and BASIC. Another software component of the present invention comprises the analytic methods of this invention as programmed in a procedural language or symbolic package.

In an exemplary implementation, to practice the methods of the present invention, a nanoreporter code (*i.e.*, a correlation between the order and nature of spots on a nanoreporter and the identity of a target molecule to which such a nanoreporter binds) is first loaded in the computer system. Next the user causes execution of analysis software which performs the steps of determining the presence and, optionally, quantity of nanoreporters with a given nanoreporter code.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 974

The analytical systems of the invention also include computer program products that contain one or more of the above-described software components such that the software components may be loaded into the memory of a computer system. Specifically, a computer program product of the invention includes a computer readable storage medium having one or more computer program mechanisms embedded or encoded thereon in a computer readable format. The computer program mechanisms encoded, *e.g.*, one or more of the analytical software components described above which can be loaded into the memory of a computer system and cause the processor of the computer system to execute the analytical methods of the present invention.

The computer program mechanisms or mechanisms are preferably stored or encoded on a computer readable storage medium. Exemplary computer readable storage media are discussed above and include, but are not limited to: a hard drive, which may be, *e.g.*, an external or an internal hard drive of a computer system of the invention, or a removable hard drive; a floppy disk; a CD-ROM; or a tape such as a DAT tape. Other computer readable storage media will also be apparent to those skilled in the art that can be used in the computer program mechanisms of the present invention.

The present invention also provides databases useful for practicing the methods of the present invention. The databases may include reference nanoreporter codes for a large variety of target molecules. Preferably, such a database will be in an electronic form that can be loaded into a computer system. Such electronic forms include databases loaded into the main memory of a computer system used to implement the methods of this invention, or in the main memory of other computers linked by network connection, or embedded or encoded on mass storage media, or on removable storage media such as a CD-ROM or floppy disk.

Alternative systems and methods for implementing the methods of this invention are intended to be comprehended within the accompanying claims. In particular, the accompanying claims are intended to include the alternative program structures for implementing the methods of this invention that will be readily apparent to one of skill in the art.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 975

APPLICATIONS OF NANOREPORTER TECHNOLOGY

The compositions and methods of the invention can be used for diagnostic, prognostic therapeutic and screening purposes. The present invention provides the advantage that many different target molecules can be analyzed at one time from a single biomolecular sample using the methods of the invention. This allows, for example, for several diagnostic tests to be performed on one sample.

DIAGNOSTIC/PROGNOSTIC METHODS

The present methods can be applied to the analysis of biomolecular samples obtained or derived from a patient so as to determine whether a diseased cell type is present in the sample and/or to stage the disease.

For example, a blood sample can be assayed according to any of the methods described herein to determine the presence and/or quantity of markers of a cancerous cell type in the sample, thereby diagnosing or staging the cancer.

Alternatively, the methods described herein can be used to diagnose pathogen infections, for example infections by intracellular bacteria and viruses, by determining the presence and/or quantity of markers of bacterium or virus, respectively, in the sample.

Thus, the target molecules detected using the compositions and methods of the invention can be either patient markers (such as a cancer marker) or markers of infection with a foreign agent, such as bacterial or viral markers.

Because of the quantitative nature of nanoreporters, the compositions and methods of the invention can be used to quantitate target molecules whose abundance is indicative of a biological state or disease condition, for example, blood markers that are upregulated or downregulated as a result of a disease state.

In addition, the compositions and methods of the invention can be used to provide prognostic information that assists in determining a course of treatment for a patient. For example, the amount of a particular marker for a tumor can be accurately quantified from even a small sample from a patient. For certain diseases like breast cancer, overexpression of certain genes, such as Her2-neu, indicate a more aggressive course of treatment will be needed.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 976

ANALYSIS OF PATHOLOGY SAMPLES

RNA extracted from formaldehyde- or paraformaldehyde-fixed paraffin-embedded tissue samples is typically poor in quality (fragmented) and low in yield. This makes gene expression analysis of low-expressing genes in histology samples or archival pathology tissues extremely difficult and often completely infeasible. The nanoreporter technology can fill this unmet need by allowing the analysis of very small quantities of low-quality total RNA.

To use nanoreporter technology in such an application, total RNA can be extracted from formaldehyde- or paraformaldehyde-fixed paraffin-embedded tissue samples (or similar) using commercially available kits such as RecoverAll Total Nucleic Acid Isolation Kit (Ambion) following manufacturer's protocols. RNA in such samples is frequently degraded to small fragments (200 to 500 nucleotides in length), and many paraffin-embedded histology samples only yield tens of nanograms of total RNA. Small amounts (5 to 100 ng) of this fragmented total RNA can be used directly as target material in a nanoreporter hybridization following the assay conditions described herein. As described in Example 6 in Section 11 below, nanoreporter analysis of approximately 3.3 ng cellular RNA permitted detection of transcripts present at approximately 0.5 copy/cell.

SCREENING METHODS

The methods of the present invention can be used, *inter alia*, for determining the effect of a perturbation, including chemical compounds, mutations, temperature changes, growth hormones, growth factors, disease, or a change in culture conditions, on various target molecules, thereby identifying target molecules whose presence, absence or levels are indicative of particular biological states. In a preferred embodiment, the present invention is used to elucidate and discover components and pathways of disease states. For example, the comparison of quantities of target molecules present in a disease tissue with "normal" tissue allows the elucidation of important target molecules involved in the disease, thereby identifying targets for the discovery/screening of new drug candidates that can be used to treat disease.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 977

5.17 KITS

The invention further provides kits comprising one or more components of the invention. The kits can contained pre-labeled nanoreporters, or unlabeled nanoreporters with one or more components for labeling the nanoreporters. Moreover, the nanoreporters provided in a kit may or may not have target-specific sequences pre-attached. In one embodiment, the target sequences are provided in the kit unattached to the nanoreporter scaffold.

The kit can include other reagents as well, for example, buffers for performing hybridization reactions, linkers, restriction endonucleases, and DNA ligases.

The kit also will include instructions for using the components of the kit, and/or for making and/or using the labeled nanoreporters.

EXAMPLE 1: NANOREPORTER MANUFACTURING AND PROTOCOL

Herein is a step-by-step example of a method construction of a nanoreporter from various components.

It can be appreciated that various components can be constructed or added either at the same time, before or after other components. For example, annealing patch units or flaps to a scaffold can be done simultaneously or one after the other.

In this example the starting material is a circular M13mp18 viral vector. Using a single linear strand M13mp18, patch units are annealed to it to form a double stranded scaffold. Next, flaps are added, then a target-specific sequence is ligated. Meanwhile purification steps aid to filter out excess, unattached patch units and flaps. Construction of labeled nucleic acids (patches and/or flaps and/or other labeled oligonucleotides) that bind the nanoreporter are also described.

Upon attachment (*e.g.*, via hybridization) of a target molecule, the nanoreporter is attached to a surface and stretched. Finally the nanoreporters are imaged by a camera.

Nanoreporters were generated and successfully employed to detect target molecules using methods substantially as described in this example. An example of target detection using this method is shown in Figure 4.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 978

SCAFFOLD CONSTRUCTION

The oligonucleotide scaffold sequence selected was analyzed using Vector NTI® software. First, a single stranded nucleic acid was made from linearizing a circular M13mp18 single stranded DNA, which was commercially purchased from New England Biolabs. The circular M13mp18 was digested with BamH1 enzyme to linearize it. Materials used consisted of M13mp18 vector (250 ng/μl), Patch_1L_BamH1.02 (10 μM dilution of a 100 μM stock), 10X BamH1 Buffer, BamH1 enzyme. Protocol for making 0.8 pmol total of linear M13mp18 involve the following steps. 1) preheat heating block to 37°C; 2) in a 0.65ml ependorff tube combine 40 μl of 250 ng/μl M13mp18 vector, 2 μl of 10 μM Patch_1L_BamH1.02, and 5 μl of 10X BamH1 Buffer; 3) place the ependorff tube in the 37°C heating block with foil over the top. Incubate the tube at 37°C for 15 minutes to allow the patch unit to hybridize to the M13mp18 scaffold; 4) after 15 minutes add 2 μl of BamH1 enzyme and let the reaction digest at 37°C for 30 minutes, after which add an additional 2 μl of BamH1 enzyme and let the reaction continue to digest for another 30 minutes at 37°C (final volume of BamH1 enzyme is 8%); and 5) aliquot 10 μl into 0.65ml ependorff tubes and store in freezer (final concentration of linear M13mp18 is 200 ng/μl).

PATCH UNIT PREPARATION OF THE BASE PATCH POOLS (BPP).

Second, patch units are prepared in pools. Patch oligonucleotide sequences were selected for optimal length and desired homology/non-homology to M13mp18 strand and the human genomic sequence. Patches were commercially manufactured oligonucleotides (purchased from Integrated DNA technologies) either 60 or 65 nucleotide bases in length. 50 nucleotide bases of each patch oligonucleotide are complementary to the M13mp18 single stranded DNA, 10 nucleotide bases are complementary to an adjacent patch, and 5 nucleotide base pairs are complementary to a corresponding flap. The 10 nucleotide base match between patches forms a stem structure which stabilizes the structure and helps lift the flaps off the covered scaffold so they are more available to bind labeled oligonucleotides. Synthetic binding sites, the 5 nucleotide bases, on the patches for binding to the flaps make leveraging the power of a modular system possible.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 979

The base patch pools contain nine patch units all corresponding to a specific letter grouping and position on the nanoreporter. For this example, there are four different fluorescent dyes (color) labeled A, B, C, and D and 8 different positions or regions where labeled nucleic acids can bind on a nanoreporter. For example, BPP A3 corresponds to all of the A patch units at position 3 (patch units 19-27) on the nanoreporter.

The nanoreporter positions are as follows:

Position 1: Patch units 1-9 (A or C)

Position 2: Patch units 10-18 (B or D)

Position 3: Patch units 19-27 (A or C)

Position 4: Patch units 28-36 (B or D)

Position 5: Patch units 37-45 (A or C)

Position 6: Patch units 46-54 (B or D)

Position 7: Patch units 55-63 (A or C)

Position 8: Patch units 64-72 (B or D)

Materials: right and left patches, pre-annealed to each other (each oligonucleotide is at a concentration of 10 μM). Materials for making 100 pmol of BPP 1: (In position 1, patch coordinate 1L is used for the BamH1 digest - this patch is not included in BPP 1): 10 μl each pre-annealed (10 μM/each) patch unit (coordinates 2-9), 5 μl [20 μM] Patch_1R (A or C). Final concentration of each patch is 1.18 pmol/μl. Materials for making 100 pmol of BPP 2-8: 10 μl each pre-annealed (10 μM/each) appropriate patch unit. There are 9 patch units added to each, or 90 μl total. Final concentration of each patch is 1.11 pmol/μl.

Below is a table of all the patch unit pools made for this example, with 8 positions or regions for dye-labeled nucleic acids to bind on the nanoreporter. Positions 1, 3, 5, and 7 can bind to nucleic acid labeled with dye A or dye C, and a positions 2, 4, 6, and 8 can bind to nucleic acid labeled with dye B or dye D.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 980

Table 2 of resulting Basic Patch Pools (correspond to labels on tubes)

BPP-A1 [Pre-Paired, Color=A, Coordinates 1-9] Patch_(1-9)R.A Patch_(2-9)L
BPP-B2 [Pre-Paired, Color=B, Coordinates 10-18] Patch_(10-18)R.B Patch_(10-18)L
BPP-A3 [Pre-Paired, Color=A, Coordinates 19-27] Patch_(19-27)R.A Patch_(19-27)L
BPP-B4 [Pre-Paired, Color=B, Coordinates 28-36] Patch_(28-36)R.B Patch_(28-36)L
BPP-A5 [Pre-Paired, Color=A, Coordinates 37-45] Patch_(37-45)R.A Patch_(37-45)L
BPP-B6 [Pre-Paired, Color=B, Coordinates 46-54] Patch_(46-54)R.B Patch_(46-54)L
BPP-A7 [Pre-Paired, Color=A, Coordinates 55-63] Patch_(55-63)R.A Patch_(55-63)L
BPP-B8 [Pre-Paired, Color=B, Coordinates 64-72] Patch_(64-72)R.B Patch_(64-72)L
BPP-C1 [Pre-Paired, Color=C, Coordinates 1-9] Patch_(1-9)R.C Patch_(2-9)L
BPP-D2 [Pre-Paired, Color=D, Coordinates 10-18] Patch_(10-18)R.D Patch_(10-18)L
BPP-C3 [Pre-Paired, Color=C, Coordinates 19-27] Patch_(19-27)R.C Patch_(19-27)L

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 981

BPP-D4 [Pre-Paired, Color=D, Coordinates 28-36] Patch_(28-36)R.D Patch_ (28-36)L
BPP-C5 [Pre-Paired, Color=C, Coordinates 37-45] Patch_(37-45)R.C Patch_(37-45)L
BPP-D6 [Pre-Paired, Color=D, Coordinates 46-54] Patch_(46-54)R.D Patch_(46-54)L
BPP-C7 [Pre-Paired, Color=C, Coordinates 55-63] Patch_(55-63)R.C Patch_(55-63)L
BPP-D8 [Pre-Paired, Color=D, Coordinates 64-72] Patch_(64-72)R.D Patch_(64-72)L

MATERIALS AND PREPARATION FOR ANNEALING THE SINGLE STRANDED OLIGONUCLEOTIDE WITH PATCH UNITS FOR A DOUBLE STRANDED SCAFFOLD.

Third, patch units are prepared to be annealed to the single stranded linear M13mp18, covering the strand in order to make a double stranded oligonucleotide scaffold. Conditions for annealing 60 and 65 nucleotide base patches to the M13mp18 need to occur at high salt concentrations so that binding will be very specific and patches will not anneal to an incorrect coordinate on the M13mp18 strand. For the annealing step, each patch unit is added at a 2:1 to 4:1 ratio with the single stranded M13mp18 sequence at 0.5 pmol total volume. Excess patches are removed before annealing flaps.

Materials used consisted of 20x SSC, linear M13mp18 (BamH1 digested at 0.08 pmol/μl or 200 ng/μl), appropriate base patch pools (BPP) (need 8 total at 1.11 pmol/μl - see above) and digital heat block set at 45°C. Annealing reaction make up is as follows. General guidelines: 2X each patch unit per M13mp18 molecule, pre-ligated flaps/patches (in position 1 or 8) added for purification later, and 5X SSC. Example (0.5 pmol of scaffold with F8 hook flaps) reaction consists of: 7.1 μl BamH1 Digested M13mp18 strand at 0.071 μM,

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 982

0.9 µl each new Base Patch Pools at 1.11 µM for first 7 positions: A1, B2, A3, B4, C5, B6 and A7:

1.7 µl A1 BPP (Pre-Annealed, 12/15; at 1.18 µM/each patch)

1.8 µl B2 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl A3 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl B4 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl C5 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl B6 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch)

1.8 µl A7 BPP (Pre-Annealed, 12/15; at 1.11 µM/each patch),

2.4 µl BPP-D8 (pool of the first seven patch units - coordinates 64, 65, 66, 67, 68, 69, and 70 at position 8 - "D" specificity) with purification tags- F8 (FHF, which anneal to patch coordinates 71L, 71R, 72L, 72R, 73L making full split-flap/patch units that have "F" specificity for use as biotin linkers, at position F8) at 0.83 µM, and 7.3 µl 20X SSC. The final reaction volume will be 29.3 µl at 0.027 pmol/µl.

Anti-Bam oligonucleotide is also added to anneal to region in M13 that is complementary to the (missing) 1L patch unit and to prevent recircularization of the M13 scaffold during ligation.

ANNEALING PATCH UNITS TO SINGLE STRANDED M13MP18 TO FORM A DOUBLE STRANDED SCAFFOLD.

The fourth step involves the protocol to anneal the patch units to the single stranded linear M13mp18, covering the strand in order to make a double stranded oligonucleotide scaffold, is performed in the following steps: 1) preheat heating block to 42°C, heat above reaction solution to 45°C in small PCR (or strip) tube(s) with foil over top for 15 minutes, turn heat block to 65°C and incubate for an additional 1 hour and 45 minutes and remove tubes, place on ice or freeze.

IPR2024-01177
California Institute of Technology - Ex. 2003 Page 983