



- (51) **International Patent Classification:**  
*C09B 57/00* (2006.01)      *C07D 413/14* (2006.01)  
*C09B 11/24* (2006.01)      *G01N 33/58* (2006.01)  
*C07H 21/00* (2006.01)
- (21) **International Application Number:**  
PCT/US2019/041717
- (22) **International Filing Date:**  
12 July 2019 (12.07.2019)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
62/698,006      13 July 2018 (13.07.2018)      US
- (71) **Applicant: QUANTUM-SI INCORPORATED** [US/US];  
530 Old Whitfield Street, Guilford, CT 06437 (US).
- (72) **Inventors: ROTHBERG, Jonathan, M.**; 215 Uncas Point  
Road, Guilford, CT 06437 (US). **LACKEY, Jeremy**;  
636 Nortontown Road, Guilford, CT 06437 (US). **NANI,**  
**Roger**; 17 Yankee Glen Road, Madison, CT 06443 (US).  
**DODD, David**; 43 High Street, Guilford, CT 06437 (US).
- (74) **Agent: WALLER, Patrick, R.H.** et al.; Wolf, Green-  
field & Sacks, P.C., 600 Atlantic Avenue, Boston, MA  
02210-2206 (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, BN, BR, BW, BY, BZ,  
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,  
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,  
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,  
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,  
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,  
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,  
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,  
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,  
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,  
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,  
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,  
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,  
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** BICONJUGATABLE LABELS AND METHODS OF USE

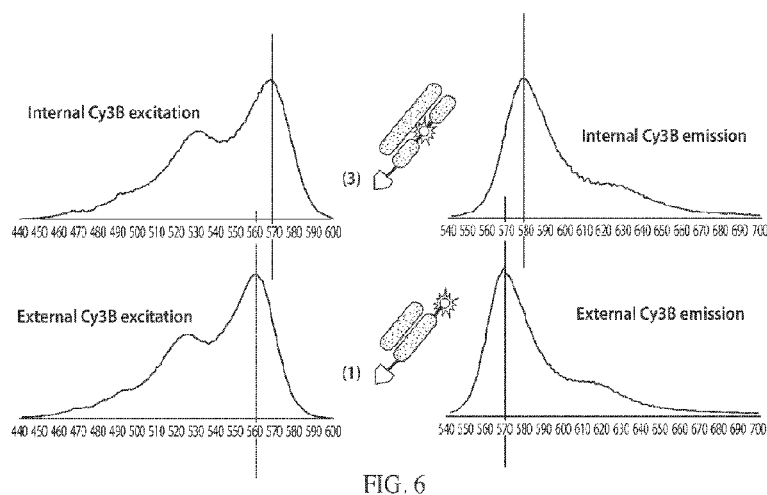
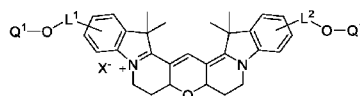


FIG. 6



(57) **Abstract:** Provided herein is a biomolecule comprising a fluorescent marker. The label is a Cy3B having formula (I) wherein, Q<sup>1</sup> and Q<sup>2</sup> are independently monomeric or oligomeric biomolecules.



**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

## BICONJUGATABLE LABELS AND METHODS OF USE

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. Provisional Patent Application No. 62/698,006, filed July 13, 2018, which is hereby incorporated by reference in its entirety.

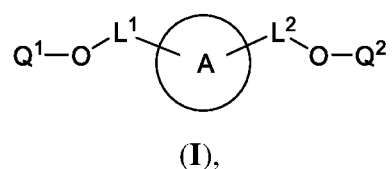
### BACKGROUND

[0002] Labeled probes are widely used in methods for detecting biological analytes and analyzing biological processes. Some of these techniques involve monitoring a biological reaction in real-time using luminescently labeled reaction components. The labels are illuminated with a light source to cause luminescence, and the luminescent light is detected with a photodetector. These events can be recorded and analyzed to identify individual reaction components based on corresponding luminescent properties. In identifying a specific type of labeled molecule among a plurality of types, it is critical that each type exhibits unique and readily differentiable luminescent properties. However, the inherent sensitivity of complex biological processes requires careful consideration when designing labeled probes for use in these systems.

### SUMMARY

[0003] Aspects of the disclosure relate to labeled biomolecules comprising internally-conjugated luminescent labels. In some aspects, the disclosure provides labeled biomolecules comprising a substrate configured for use in a reaction. In some aspects, the disclosure provides labeled biomolecules comprising a nucleotide configured for use in a polymerization reaction. In some aspects, the disclosure provides methods of sequencing using labeled nucleotides described herein. In some aspects, the disclosure provides biconjugatable luminescent labels and methods of making the same.

[0004] In some aspects, the application provides labeled biomolecules of Formula (I):



wherein: Q<sup>1</sup> and Q<sup>2</sup> are independently monomeric or oligomeric biomolecules; A is a polycyclic fluorophore; and L<sup>1</sup> and L<sup>2</sup> are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof.

**[0005]** In some aspects, provided herein is a labeled nucleotide comprising one or more nucleotides associated with a labeled biomolecule according to the present application. In some embodiments, the one or more nucleotides comprise one type of nucleotide selected from guanine, cytosine, adenine, and thymine or uracil. In some embodiments, the one or more nucleotides are cleaved from the labeled biomolecule by a polymerase when subjected to polymerization reaction conditions. In some embodiments, the one or more nucleotides comprise nucleoside polyphosphates. In some embodiments, the one or more nucleotides comprise nucleoside triphosphates. In some embodiments, the one or more nucleotides comprise nucleoside hexaphosphates. In some embodiments, the one or more nucleotides (e.g., nucleoside polyphosphates) are attached through a terminal phosphate to the labeled biomolecule. In some aspects, provided herein are compositions comprising a labeled nucleotide according to the present application.

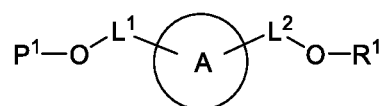
**[0006]** In some aspects, provided herein is a nucleic acid sequencing reaction composition comprising two or more different types of labeled nucleotides in a reaction mixture. In some embodiments, at least one type of labeled nucleotide of the nucleic acid sequencing reaction composition is a labeled nucleotide according to the present application. In some embodiments, the nucleic acid sequencing reaction composition comprises four different types of labeled nucleotides. In some embodiments, the nucleic acid sequencing reaction composition comprises a first labeled nucleotide comprising guanine, a second labeled nucleotide comprising cytosine, a third labeled nucleotide comprising adenine, and a fourth labeled nucleotide comprising thymine or uracil.

**[0007]** In some aspects, provided herein are methods of determining a sequence of a template nucleic acid. In some embodiments, the methods comprise exposing a complex in a target volume, the complex comprising the template nucleic acid, a primer, and a polymerizing enzyme, to a nucleic acid sequencing reaction composition according to the present application. In some embodiments, the methods further comprise directing a series of pulses of one or more excitation energies towards a vicinity of the target volume. In some

embodiments, the methods further comprise detecting a plurality of emitted photons from luminescently labeled nucleotides during sequential incorporation into a nucleic acid comprising the primer. In some embodiments, the methods further comprise identifying the sequence of incorporated nucleotides by determining timing and optionally luminescence intensity of the emitted photons.

**[0008]** In some aspects, provided herein is a kit for sequencing a template nucleic acid. In some embodiments, the kit comprises two or more different types of labeled nucleotides. In some embodiments, at least one of the two or more different types of labeled nucleotides comprises a labeled nucleotide according to the present application.

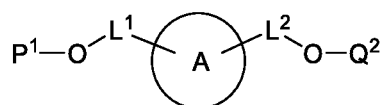
**[0009]** In some aspects, the application provides compounds Formula (II):



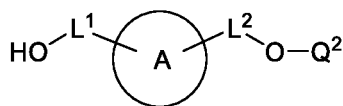
(II),

or a salt thereof, wherein: A is a polycyclic fluorophore; L<sup>1</sup> and L<sup>2</sup> are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof; P<sup>1</sup> is an oxygen protecting group; and R<sup>1</sup> is a reactive moiety. In some aspects, provided herein are compositions comprising a compound of Formula (II).

**[0010]** In some aspects, provided herein is a method for preparing a labeled biomolecule of the present application. In some embodiments, the method comprises (i) contacting a monomeric or oligomeric biomolecule of formula Q<sup>2</sup>-OH, or a salt thereof, with a compound of Formula (II), or a salt thereof, under conditions sufficient to promote conjugation to yield a conjugate of the formula:



**[0011]** In some embodiments, the method further comprises (ii) deprotecting the conjugate formed in step (i) under conditions sufficient to cleave the P<sup>1</sup> protecting group and yield a conjugate of the formula:



[0012] In some embodiments, the method further comprises (iii) contacting the conjugate formed in step (ii) with a monomeric or oligomeric biomolecule of formula  $Q^1-O-R^1$ , or a salt thereof, under conditions sufficient to promote conjugation to yield a labeled biomolecule of Formula (I).

[0013] The details of certain embodiments of the invention are set forth in the Detailed Description of Certain Embodiments, as described below. Other features, objects, and advantages of the invention will be apparent from the Definitions, Examples, Figures, and Claims.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The accompanying drawings, which constitute a part of this specification, illustrate several embodiments of the invention and together with the description, serve to explain the principles of the invention.

[0015] **FIGs. 1A-1G** show various examples of labeled biomolecules in accordance with the application. FIG. 1A shows a labeled biomolecule having one internal label. FIG. 1B shows a labeled biomolecule having two internal labels. FIG. 1C shows a circularized labeled biomolecule. FIG. 1D shows a labeled oligonucleotide that comprises an internally-labeled strand hybridized to an unlabeled strand. FIG. 1E shows a labeled oligonucleotide that comprises one internally-labeled strand hybridized to another internally-labeled strand. FIG. 1F shows a labeled oligonucleotide that comprises an unlabeled strand hybridized to a labeled strand having two internal labels. FIG. 1G shows a labeled oligonucleotide strand that is self-hybridized to form a stem-loop motif.

[0016] **FIGs. 2A-2C** depict examples of uses of labeled biomolecules in accordance with the application. FIG. 2A depicts an internally-labeled oligonucleotide hybridized to a target nucleic acid sequence. FIG. 2B depicts an internally-labeled biomolecule bound by a target protein. FIG. 2C depicts an internally-labeled antibody bound to a target protein.

[0017] **FIGs. 3A-3B** illustrate examples of labeled biomolecules modified with functional moieties in accordance with the application. FIG. 3A depicts an internally-labeled oligonucleotide modified with a quenching moiety that prevents detection of an internal label unless cleaved from the biomolecule. FIG. 3B depicts an internally-labeled biomolecule

modified with a ligand that is bound by a target protein, where binding of the ligand to the target protein permits detection of an internal label.

**[0018] FIGs. 4A-4C** depict examples of labeled nucleotides bound by a polymerizing enzyme in accordance with the application. FIG. 4A depicts a polymerizing enzyme bound to a nucleotide that comprises an external label. FIG. 4B depicts a polymerizing enzyme bound to a nucleotide that comprises an internally-labeled biomolecule. FIG. 4C depicts a polymerizing enzyme bound to a nucleotide that comprises an internally-labeled oligonucleotide.

**[0019] FIGs. 5A-5B** show a comparative sequencing analysis of externally-labeled biomolecules and an internally-labeled biomolecule in accordance with the application. FIG. 5A generically depicts a set of externally-labeled and internally-labeled nucleotides that were prepared and subjected to further analysis. FIG. 5B shows the results of a single-molecule sequencing reaction that was performed using externally-labeled and internally-labeled nucleotides.

**[0020] FIG. 6** shows an alignment of excitation and emission spectra for an externally-labeled and an internally-labeled nucleotide.

#### **DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS**

**[0021]** Among other aspects, the disclosure provides labeled biomolecules comprising internally-conjugated luminescent labels (e.g., internal labels). In some embodiments, an internal label is configured with enhanced conformational restraint to restrict rotation and block deactivation pathways that shorten luminescence lifetime. In some embodiments, a labeled biomolecule is configured to provide a rigid molecular scaffold for the internal label that avoids label-label interactions and other quenching effects which could reduce luminescence intensity or other emission characteristics.

**[0022]** Without wishing to be bound by any particular theory, labeled biomolecules provided herein offer a number of distinct advantages, such as improved quantum yields and extended luminescence lifetimes, increased luminescence intensity and/or brightness, and decreased exposure to bulk solvent molecules to limit formation of reactive species. Accordingly, in some embodiments, provided herein are labeled biomolecules and methods of use. In some embodiments, the disclosure provides compositions and methods related to the preparation of labeled biomolecules. In some embodiments, the disclosure provides compositions and methods related to the preparation of biconjugatable labels.

**Internally-Labeled Biomolecules**

**[0023]** One aspect of the present invention relates to internally-labeled biomolecules (“labeled biomolecules”). As described herein, the labeled biomolecules (e.g., oligonucleotides, nucleic acids, polypeptides, proteins, polysaccharides) comprise internally-conjugated luminescent compounds (“luminescent labels” or “labels”, e.g., polycyclic fluorophores). “Internally-labeled” or “internally-conjugated” are used interchangeably herein and refer to when one portion of the biomolecule is conjugated to a first site on the luminescent compound, and another portion of the biomolecule is conjugated to a second site on the luminescent compound. Internal conjugation of luminescent compounds can have several advantages. For example, internal conjugation of a label into a biomolecule can alter the lifetime and other photophysical properties of the label (e.g., via restricted rotation of the label). As another example, internal conjugation of a label can immobilize the label, separating it from other labels so as to mitigate self-quenching. Further, internal conjugation of a label can limit the label’s access to bulk solvents in solution so as to mitigate radical formation which may be damaging to other components in the solution. Internal incorporation of labels have other advantages as described herein. A non-limiting set of examples of labeled biomolecules is shown in FIGs. 1A-1G.

**[0024]** FIGs. 1A-1G generically depict various configurations of labeled biomolecules in accordance with the present application. Each example is shown as having an internal label and a biomolecule (shown as stippled shapes). FIG. 1A is a labeled biomolecule comprising one internal label that conjugates one portion of a biomolecule to another portion of the biomolecule. In some embodiments, a labeled biomolecule comprises two or more internal labels.

**[0025]** FIG. 1B is a labeled biomolecule comprising two internal labels. As shown, one label conjugates a first portion of a biomolecule to a second portion of the biomolecule, and another label conjugates the second portion of the biomolecule to a third portion of the biomolecule. In some embodiments, labeled biomolecules having two or more copies of the same internal label exhibit increased luminescence intensity and/or brightness relative to a labeled biomolecule having one copy of the internal label. In some embodiments, labeled biomolecules having two or more different types of internal labels provide two or more unique detectable signals. Examples of various configurations and uses of multiply-labeled biomolecules are described elsewhere herein.

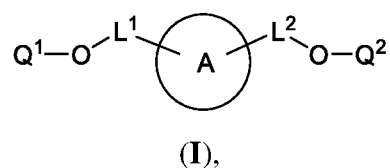
[0026] In accordance with the application, increased rigidity of internally-conjugated labels has been shown to enhance one or more luminescent properties of labeled biomolecules, e.g., via conformational restraint provided by a polycyclic fluorophore of the internal label. Advantageously, a biomolecular scaffold upon which an internal label is conjugated can be engineered to further promote rigidity and further enhance the one or more luminescent properties in these systems.

[0027] FIG. 1C is a labeled biomolecule comprising an internal label conjugated to a circularized biomolecule. As depicted by this example, in some embodiments, an internal label conjugates one end of a biomolecule to another end of the biomolecule such that the labeled biomolecule is circularized. Without wishing to be bound by any particular theory, it is thought that circularization promotes structural rigidity throughout the backbone of a biomolecule, which enhances the favorable luminescent properties of internal labels provided herein. Examples of circular biomolecules include, without limitation, cyclic peptides, cyclic proteins, and circular nucleic acids (e.g., circular RNA, DNA plasmids).

[0028] The inventors have further recognized and appreciated that oligonucleotides (e.g., polynucleotides, nucleic acids) provide rigid, highly tunable biomolecular scaffolds for internal labels of the application. Various examples of internally-labeled oligonucleotides are shown in FIGs. 1D-1G. These example constructs and additional embodiments related to internally-labeled oligonucleotides are described in detail elsewhere herein.

[0029] As generally illustrated in the example structures shown in FIGs. 1A-1G, in some embodiments, internal label **100** corresponds to A of Formula (I), as depicted herein. In some embodiments, a biomolecule (shown as stippled shapes) of the example structures corresponds to Q<sup>1</sup> and/or Q<sup>2</sup> of Formula (I), as depicted herein. In some embodiments, internal label **100** corresponds to A and a biomolecule (shown as stippled shapes) of the example structures corresponds to Q<sup>1</sup> and Q<sup>2</sup> of Formula (I), as depicted herein.


[0030] In one aspect, provided herein are labeled biomolecules of Formula (I):



wherein:

- Q<sup>1</sup> and Q<sup>2</sup> are independently monomeric or oligomeric biomolecules;
- A is a polycyclic fluorophore; and

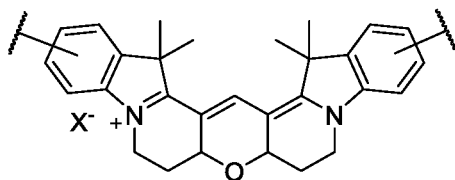
$L^1$  and  $L^2$  are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof.

**[0031]** As described herein,  (also depicted herein as “A”) is a luminescent compound or dye. As represented by Formula (I), A is conjugated to one portion of a biomolecule (group  $Q^1$ ) through a linker represented by  $L^1$ , and is conjugated to another portion of the biomolecule (group  $Q^2$ ) through linker  $L^2$ .  $Q^1$  and  $Q^2$  together form a biomolecule, which is interrupted by a structure:  $-O-L^1-A-L^2-O-$ , and is thereby internally labeled.

**[0032]** In certain embodiments, A is a polycyclic fluorophore. In certain embodiments, incorporating a polycyclic fluorophore is advantageous insofar as the polycyclic structure can impart greater rigidity to the system as compared with linear or non-polycyclic fluorophores. In certain embodiments,  $L^1$  and  $L^2$  are directly linked (e.g., through a covalent bonds) to one or more rings (e.g., benzenoid or heteroaromatic rings) of the polycyclic structure. This direct linkage can also impart greater rigidity to the system (e.g., via immobilization/restricted rotation of the label). In certain embodiments,  $L^1$  and  $L^2$  are directly linked to different rings on A, a design feature which can also impart greater rigidity to the system and/or help immobilize the dye.

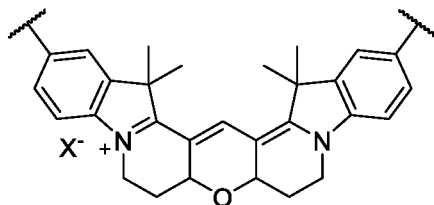
**[0033]** In certain embodiments, A is a polycyclic cyanine, fluorone, acridine, phenoxazine, coumarin, or boron-dipyrromethene (BODIPY) fluorophore. In certain embodiments, A is a porphyrin, phthalocyanine, or naphthalimide. These are non-limiting examples. In certain embodiments, A is a polycyclic cyanine fluorophore. In certain embodiments, A is a polycyclic fluorone fluorophore. In certain embodiments, A is a polycyclic acridine fluorophore. In certain embodiments, A is a polycyclic phenoxazine fluorophore. In certain embodiments, A is a polycyclic coumarin fluorophore. In certain embodiments, A is a polycyclic BODIPY fluorophore. In certain embodiments, A is a porphyrin. In certain embodiments, A is a phthalocyanine. In certain embodiments, A is a naphthalimide. Other embodiments of Ring A are described below and herein.

**[0034]** In certain embodiments, A is a polycyclic cyanine fluorophore. In certain embodiments, A is an optionally substituted Cy3B dye. In certain embodiments, A is Cy3B. In certain embodiments, A is of the following formula:



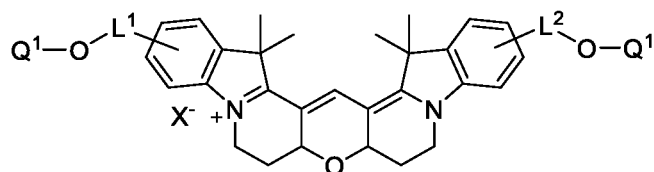
wherein  $X^-$  is a counterion or is absent; and wherein the structure is optionally substituted at any position. In certain embodiments, the structure is unsubstituted.

**[0035]** In certain embodiments, A is of the following formula:



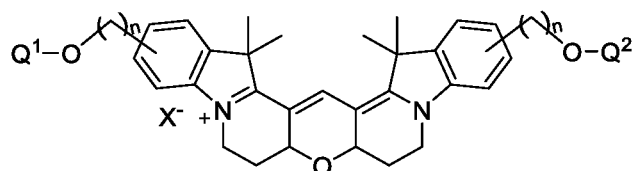
wherein  $X^-$  is a counterion or is absent; and wherein the structure is optionally substituted at any position. In certain embodiments, the structure is unsubstituted.

**[0036]** Accordingly, in certain embodiments, the labeled biomolecule is of the formula:



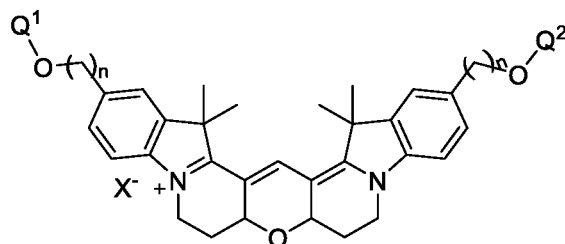
wherein  $X^-$  is a counterion or is absent.

**[0037]** In certain embodiments, the labeled biomolecule is of the formula:

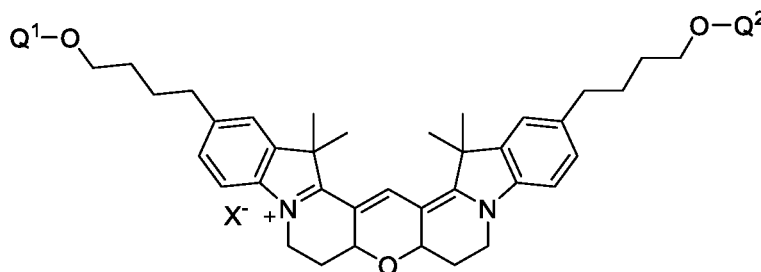


wherein  $n$  is independently an integer from 1-20, inclusive.

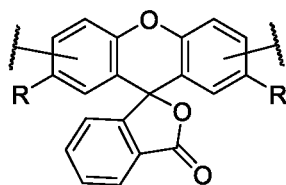
**[0038]** In certain embodiments, the labeled biomolecule is of the formula:



**[0039]** In certain embodiments, the labeled biomolecule is of the formula:



**[0040]** As described herein, in certain embodiments, Ring A is a polycyclic fluorone fluorophore (e.g., fluorescein or rhodamine). In certain embodiments, A is a fluorone dye. In certain embodiments, A is a rhodamine dye. In certain embodiments, A is of the following formula:

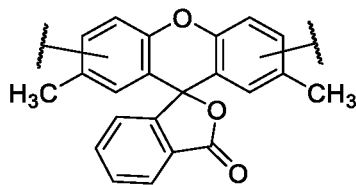


wherein each R is independently hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $-N(R^N)_2$ . In certain embodiments, the structure is optionally substituted at any position. In certain embodiments, the structure is unsubstituted.

**[0041]** As defined herein, each R is independently hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $-N(R^N)_2$ . In certain embodiments, R is hydrogen. In certain embodiments, R is halogen. In certain embodiments, R is  $-N_3$ . In certain embodiments, R is  $-CN$ . In certain embodiments, R is  $-NO_2$ . In certain embodiments, R is optionally substituted alkyl. In certain embodiments, R is optionally substituted alkenyl. In certain embodiments, R is optionally substituted alkynyl. In certain embodiments, R is optionally substituted carbocyclyl. In certain embodiments, R is optionally substituted aryl. In certain embodiments, R is optionally substituted heterocyclyl. In certain embodiments, R is optionally substituted heteroaryl. In certain embodiments, R is  $-OR^O$ . In certain embodiments, R is  $-SR^S$ . In certain embodiments, R is  $-N(R^N)_2$ . In certain embodiments, R is optionally substituted  $C_{1-6}$  alkyl. In certain embodiments, R is unsubstituted  $C_{1-6}$  alkyl. In certain embodiments, R is optionally substituted  $C_{1-3}$  alkyl. In certain embodiments, R is unsubstituted  $C_{1-3}$  alkyl. In certain embodiments, R is selected

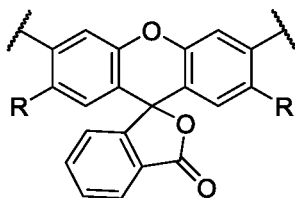
from the group consisting of methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl, sec-butyl, and tert-butyl. In certain embodiments, R is methyl. In certain embodiments, each R is methyl.

[0042] In certain embodiments, A is of the formula:



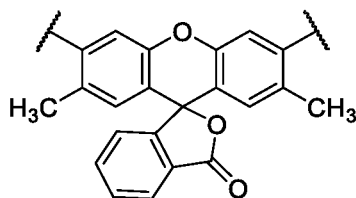
wherein the structure is optionally substituted at any position. In certain embodiments, the structure is unsubstituted.

[0043] In certain embodiments, A is of the formula:



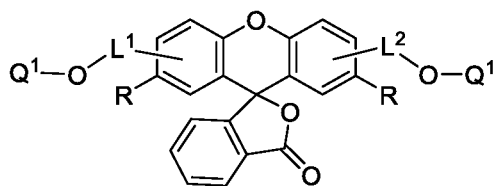
wherein the structure is optionally substituted at any position. In certain embodiments, the structure is unsubstituted.

[0044] In certain embodiments, A is of the formula:

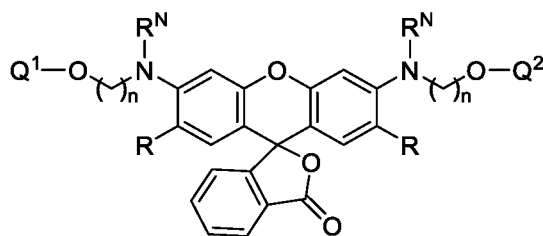


wherein the structure is optionally substituted at any position. In certain embodiments, the structure is unsubstituted.

[0045] Accordingly, in certain embodiments, the labeled biomolecule is of the formula:



[0046] In certain embodiments, the labeled biomolecule is of the formula:

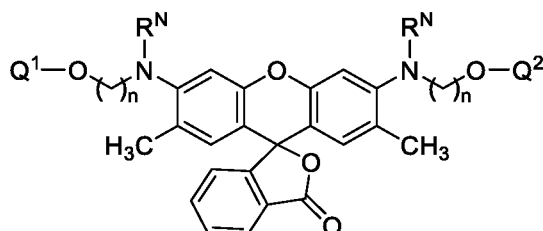


wherein:

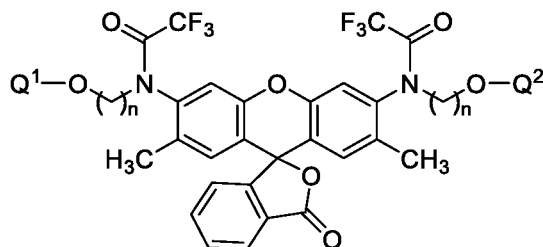
$n$  is independently an integer from 1-20, inclusive; and

each instance of  $R^N$  is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group.

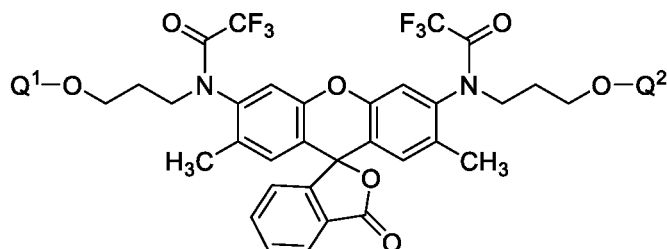
[0047] In certain embodiments, the labeled biomolecule is of the formula:



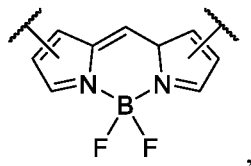
[0048] In certain embodiments, the labeled biomolecule is of the formula:



[0049] In certain embodiments, the labeled biomolecule is of the formula:

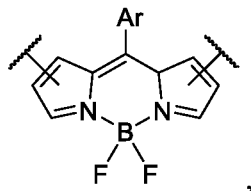


[0050] As described herein, in certain embodiments, A is a boron-dipyrromethene (BODIPY) fluorophore. In certain embodiments, A is of the following formula:



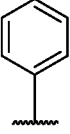
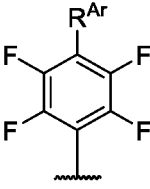
wherein the structure is optionally substituted at any position.

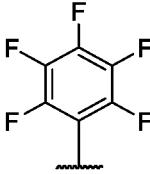
[0051] In certain embodiments, A is of the formula:

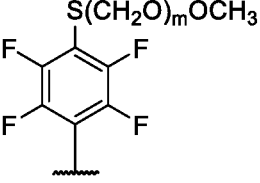


wherein Ar is optionally substituted aryl or optionally substituted heteroaryl; and wherein the structure is optionally substituted at any position.

[0052] As defined herein, Ar is optionally substituted aryl or optionally substituted heteroaryl. In certain embodiments, Ar is optionally substituted aryl. In certain embodiments, Ar is optionally substituted heteroaryl. In certain embodiments, Ar is optionally substituted phenyl. In certain embodiments, Ar is polyfluorophenyl. In certain embodiments, Ar is of the

formula: . In certain embodiments, Ar is of the formula: . In certain

embodiments, Ar is of the formula: . In certain embodiments, Ar is of the

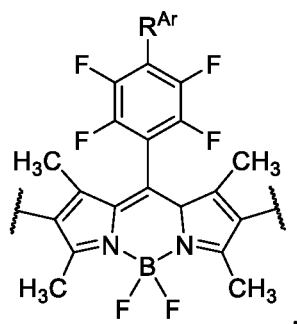
formula: , wherein m is as defined herein.

[0053] As defined herein, R<sup>Ar</sup> is hydrogen, halogen, -N<sub>3</sub>, -CN, -NO<sub>2</sub>, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl, -OR<sup>O</sup>, -SR<sup>S</sup>, or -N(R<sup>N</sup>)<sub>2</sub>. In certain embodiments, R<sup>Ar</sup> is hydrogen. In certain embodiments, R<sup>Ar</sup> is halogen (-Cl, -I, -Br, -F). In certain embodiments, R<sup>Ar</sup> is -N<sub>3</sub>.

In certain embodiments,  $R^{Ar}$  is  $-CN$ . In certain embodiments,  $R^{Ar}$  is  $-NO_2$ . In certain embodiments,  $R^{Ar}$  is optionally substituted alkyl. In certain embodiments,  $R^{Ar}$  is optionally substituted alkenyl. In certain embodiments,  $R^{Ar}$  is optionally substituted alkynyl. In certain embodiments,  $R^{Ar}$  is optionally substituted carbocyclyl. In certain embodiments,  $R^{Ar}$  is optionally substituted aryl. In certain embodiments,  $R^{Ar}$  is optionally substituted heterocyclyl. In certain embodiments,  $R^{Ar}$  is optionally substituted heteroaryl. In certain embodiments,  $R^{Ar}$  is  $-OR^O$ . In certain embodiments,  $R^{Ar}$  is  $-SR^S$ . In certain embodiments,  $R^{Ar}$  is or  $-N(R^N)_2$ . In certain embodiments,  $R^{Ar}$  is  $-F$ . In certain embodiments,  $R^{Ar}$  is  $-S(CH_2CH_2O)_mOCH_3$ , wherein  $m$  is as defined herein.

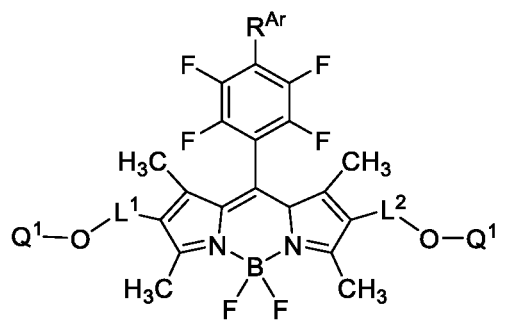
**[0054]** As defined herein,  $m$  is an integer from 1-6, inclusive.

**[0055]** In certain embodiments,  $A$  is of the formula:



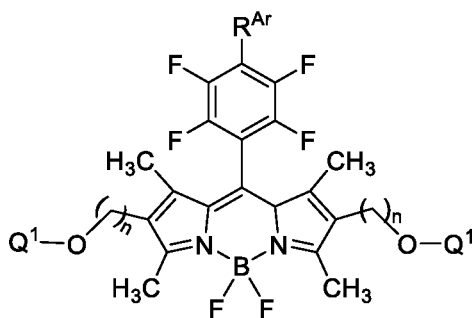
wherein  $R^{Ar}$  is hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $-N(R^N)_2$ . In certain embodiments, the structure is substituted at any position. In certain embodiments, the structure is unsubstituted.

**[0056]** Accordingly, in certain embodiments, the labeled biomolecule is of the formula:

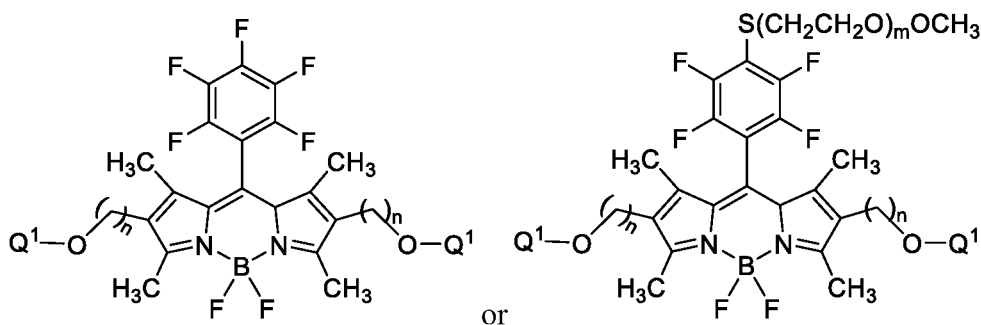


wherein  $R^{Ar}$  is hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ .

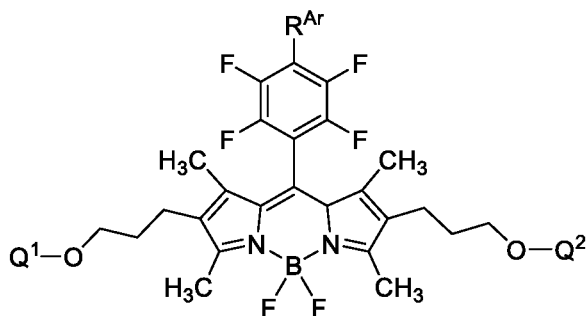
[0057] In certain embodiments, the labeled biomolecule is of the formula:



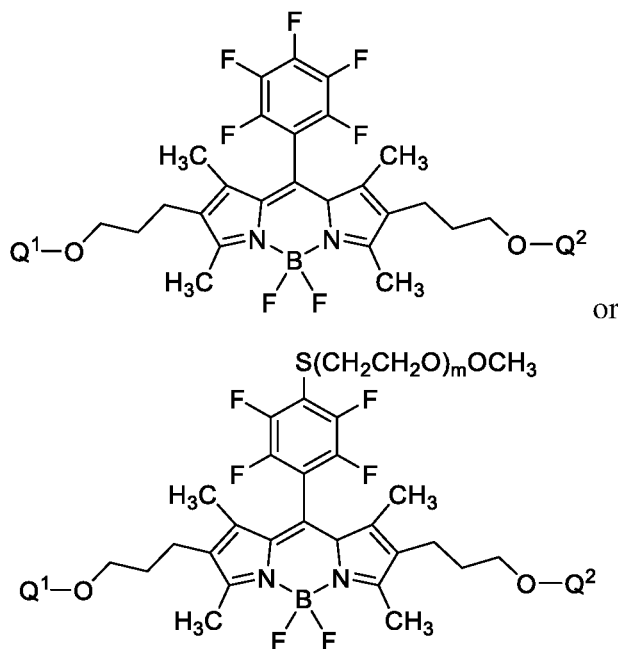
[0058] In certain embodiments, the labeled biomolecule is of one of the following formulae:



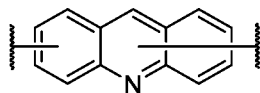
[0059] In certain embodiments, the labeled biomolecule is of the formula:



[0060] In certain embodiments, the labeled biomolecule is of one of the following formulae:

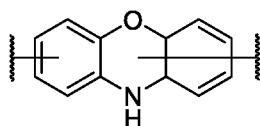


**[0061]** As described herein, in certain embodiments A is an acridine fluorophore. In certain embodiments, A is an acridine fluorophore of the following formula:



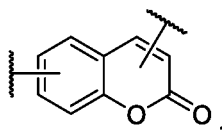
wherein the structure is optionally substituted.

**[0062]** As described herein, in certain embodiments A is a phenoxazine fluorophore. In certain embodiments, A is a phenoxazine fluorophore of the following formula:

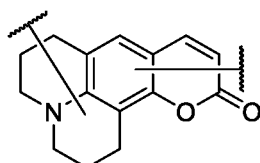


wherein the structure is optionally substituted.

**[0063]** As described herein, in certain embodiments A is a coumarin fluorophore. In certain embodiments, A is a coumarin fluorophore of the following formula:

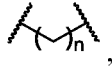


wherein the structure is optionally substituted. In certain embodiments, A is a coumarin fluorophore of the following formula:

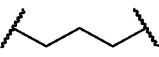


wherein the structure is optionally substituted.

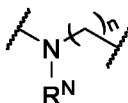
[0064] As defined herein,  $L^1$  and  $L^2$  are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof. In certain embodiments,  $L^1$  comprises optionally substituted alkylene. In certain embodiments,  $L^1$  comprises optionally substituted alkenylene. In certain embodiments,  $L^1$  comprises optionally substituted alkynylene. In certain embodiments,  $L^1$  comprises optionally substituted heteroalkylene. In certain embodiments,  $L^1$  comprises optionally substituted heteroalkenylene. In certain embodiments,  $L^1$  comprises optionally substituted heteroalkynylene. In certain embodiments,  $L^1$  comprises optionally substituted carbocyclylene. In certain embodiments,  $L^1$  comprises optionally substituted heterocyclylene. In certain embodiments,  $L^1$  comprises optionally substituted arylene. In certain embodiments,  $L^1$  comprises optionally substituted heteroarylene. In certain embodiments,  $L^1$  is optionally substituted  $C_{1-20}$  alkylene. In certain embodiments,  $L^1$  is optionally substituted  $C_{1-10}$  alkylene. In certain embodiments,  $L^1$  is

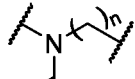
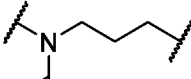
optionally substituted  $C_{1-6}$  alkylene. In certain embodiments,  $L^1$  is of the formula: ,

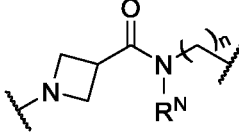
wherein n is as defined herein. In certain embodiments,  $L^1$  is of the formula: .

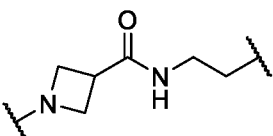
In certain embodiments,  $L^1$  is of the formula: . In certain embodiments,  $L^1$  is of

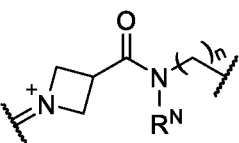
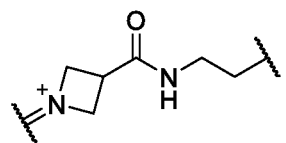
the formula: . In certain embodiments,  $L^1$  is of the formula: . In

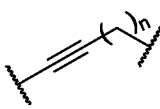
certain embodiments,  $L^1$  is of the formula: . In certain embodiments,  $L^1$  is of the

formula: . In certain embodiments,  $L^1$  is of the formula: . In certain

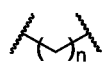
embodiments,  $L^1$  is of the formula: . In certain embodiments,  $L^1$  is of the

formula:  . In certain embodiments, L<sup>1</sup> is of the formula:

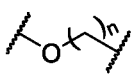
 . In certain embodiments, L<sup>1</sup> is of the formula:  . In

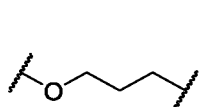
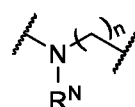
certain embodiments, L<sup>1</sup> is of the formula:  .

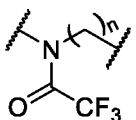
**[0065]** In certain embodiments, L<sup>2</sup> comprises optionally substituted alkylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted alkenylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted alkynylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted heteroalkylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted heteroalkenylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted heteroalkynylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted carbocyclylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted heterocyclylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted arylene. In certain embodiments, L<sup>2</sup> comprises optionally substituted heteroarylene. In certain embodiments, L<sup>2</sup> is optionally substituted C<sub>1-20</sub> alkylene. In certain embodiments, L<sup>2</sup> is optionally substituted C<sub>1-10</sub> alkylene. In certain embodiments, L<sup>2</sup> is optionally substituted C<sub>1-6</sub> alkylene. In certain embodiments, L<sup>2</sup>

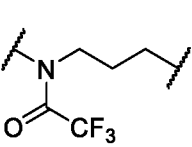
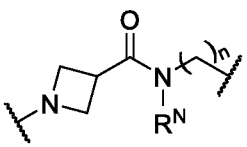
is of the formula:  , wherein n is as defined herein. In certain embodiments, L<sup>2</sup> is of

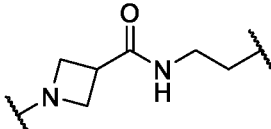
the formula:  . In certain embodiments, L<sup>2</sup> is of the formula:  . In

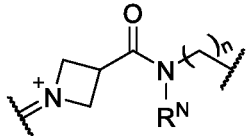
certain embodiments, L<sup>2</sup> is of the formula:  . In certain embodiments, L<sup>2</sup> is of the

formula:  . In certain embodiments, L<sup>2</sup> is of the formula:  . In certain

embodiments, L<sup>2</sup> is of the formula:  . In certain embodiments, L<sup>2</sup> is of the formula:

 . In certain embodiments, L<sup>2</sup> is of the formula:  . In certain

embodiments,  $L^2$  is of the formula:  . In certain embodiments,  $L^2$  is of

the formula:  . In certain embodiments,  $L^2$  is of the formula:

 . In certain embodiments,  $L^2$  is of the formula:  .

**[0066]** As defined herein,  $n$  is independently an integer from 1-20, inclusive. In certain embodiments,  $n$  is 1. In certain embodiments,  $n$  is 2. In certain embodiments,  $n$  is 3. In certain embodiments,  $n$  is 4. In certain embodiments,  $n$  is 5. In certain embodiments,  $n$  is 6. In certain embodiments,  $n$  is 7. In certain embodiments,  $n$  is 8. In certain embodiments,  $n$  is 9. In certain embodiments,  $n$  is 10. In certain embodiments,  $n$  is 11. In certain embodiments,  $n$  is 12. In certain embodiments,  $n$  is 13. In certain embodiments,  $n$  is 14. In certain embodiments,  $n$  is 15. In certain embodiments,  $n$  is 16. In certain embodiments,  $n$  is 17. In certain embodiments,  $n$  is 18. In certain embodiments,  $n$  is 19. In certain embodiments,  $n$  is 20.

**[0067]** As described herein,  $Q^1$  and  $Q^2$  are independently monomeric or oligomeric biomolecules. As described herein,  $Q^1$  and  $Q^2$  together form an oligomeric or polymeric biomolecule, which is interrupted by  $-O-L^1-A-L^2-O-$  and is thereby internally labeled (e.g., forming a labeled biomolecule).

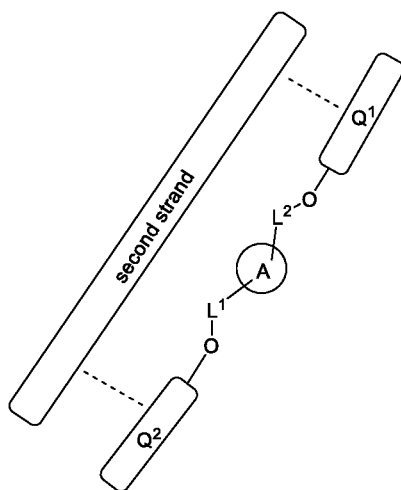
**[0068]** In some embodiments, the labeled biomolecule is an oligomeric or polymeric biomolecule comprising at least 5 monomeric biomolecules (e.g., at least 5 nucleotides, at least 5 amino acids, at least 5 monosaccharides). In some embodiments, an oligomeric or polymeric biomolecule comprises at least 10 monomeric biomolecules. In some embodiments, an oligomeric or polymeric biomolecule comprises at least 10 and fewer than 200 monomeric biomolecules. For example, in some embodiments, an oligomeric or polymeric biomolecule comprises at least 10 and fewer than 150 monomeric biomolecules, at least 10 and fewer than 100 monomeric biomolecules, at least 10 and fewer than 50 monomeric biomolecules, at least 10 and fewer than 40 monomeric biomolecules, at least 10 and fewer than 30 monomeric biomolecules, or at least 10 and fewer than 20 monomeric biomolecules.

**[0069]** In certain embodiments, the labeled biomolecule is an oligonucleotide or nucleic acid.

[0070] In certain embodiments,  $Q^1$  and  $Q^2$  are independently nucleosides, nucleotides, oligonucleotides, nucleic acids, or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently nucleosides or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently nucleotides derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently oligonucleotides or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently nucleic acids or derivatives or fragments thereof.

[0071] In certain embodiments,  $Q^1$  and  $Q^2$  are independently deoxyribonucleic acids, ribonucleic acids, peptide nucleic acids, locked nucleic acids, or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently deoxyribonucleic acids or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently ribonucleic acids or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently peptide nucleic acids or derivatives or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are locked nucleic acids, or derivatives or fragments thereof.

[0072] As described herein,  $Q^1$  and  $Q^2$  together form an oligomeric or polymeric biomolecule, which is interrupted by  $-O-L^1-A-L^2-O-$  and is thereby internally labeled (e.g., forming a labeled biomolecule). In certain embodiments, the labeled biomolecule is a single-stranded nucleic acid. In this instance, the single-stranded nucleic acid comprises a first oligonucleotide strand ( $Q^1$  and/or  $Q^2$ ). In certain embodiments, the labeled biomolecule comprises a second oligonucleotide strand hybridized to the first oligonucleotide strand. For example, in certain embodiments, the second oligonucleotide strand is hybridized to  $Q^1$  and/or  $Q^2$ . A visual representation of this internally-labeled system is represented below (wherein - - - represents hybridization interactions (e.g., one or more Watson-Crick base interactions)):



A visual representation of this system can also be found in FIG. 1D.

**[0073]** Additional examples of internally-labeled oligonucleotides are shown in FIGs. 1E-1G. As illustrated by these and other examples described herein, various oligonucleotide strand hybridization strategies are provided to promote rigidity and/or shield an internal label from bulk solvent. Accordingly, oligonucleotide strand hybridization can be used as a general design strategy in preparing labeled oligonucleotides of the application. In some embodiments, oligonucleotide strand hybridization involves self-strand hybridization (e.g., self-hybridizing within a single strand). In some embodiments, oligonucleotide strand hybridization involves hybridization of different oligonucleotide strands.

**[0074]** FIG. 1D is a labeled oligonucleotide comprising an internally-labeled oligonucleotide strand hybridized to an unlabeled oligonucleotide strand. In some embodiments, an unlabeled oligonucleotide strand is used to increase rigidity in specific regions of a labeled oligonucleotide strand (e.g., in a region comprising an internal label). In some embodiments, an unlabeled oligonucleotide strand is hybridized to an internally-labeled oligonucleotide strand that comprises two or more internal labels provided herein.

**[0075]** FIG. 1E is a labeled oligonucleotide comprising one internally-labeled oligonucleotide strand hybridized to another internally-labeled oligonucleotide strand. In some embodiments, an internal label of one oligonucleotide strand comprises the same fluorophore as an internal label of the other oligonucleotide strand. In some embodiments, an internal label of one oligonucleotide strand comprises a different fluorophore from an internal label of the other oligonucleotide strand. In some embodiments, one of the internally-labeled oligonucleotide strands comprises two or more internal labels of the present application. In some embodiments, both of the internally-labeled oligonucleotide strands comprise two or more internal labels of the present application.

**[0076]** In some embodiments, oligonucleotide strand hybridization promotes formation of one or more structural motifs, such as stem-loops, junctions, pseudoknots, and double helices. Structural motifs, in accordance with the present application, are useful for enhancing rigidity of a labeled oligonucleotide and/or limiting the extent to which an internal label is exposed to bulk solvent. FIGs. 1F-1G depict examples of labeled oligonucleotides having higher order structural motifs formed through strand hybridization.

**[0077]** FIG. 1F is a labeled oligonucleotide comprising an unlabeled oligonucleotide strand hybridized to an oligonucleotide strand that comprises two internal labels. As generically shown in this example, formation of a double helix can promote separation of two internal

labels of the same oligonucleotide strand. In some embodiments, hybridized oligonucleotide strands form a double helix having approximately 10 to 12 base pairs per turn.

**[0078]** Accordingly, in some embodiments, where two internal labels of the same oligonucleotide strand occupy approximately the same amount of space as one nucleotide within the strand, the internal labels are minimally separated by 5 to 6 nucleotides along the strand such that the labels are on approximately opposite sides of a double helix. In some embodiments, internal labels are separated by between 4 to 8 (e.g., 4, 5, 6, 7, or 8) nucleotides along the oligonucleotide strand. Without wishing to be bound by theory, such design strategies may be utilized to limit the extent of label-label interaction (e.g., quenching effects) due to the intervening helical structure absorbing any radiative and/or non-radiative decay.

**[0079]** Internal labels of the disclosure, in some embodiments, are integrated into the oligonucleotide backbone to minimize the extent of label-label interaction. Accordingly, in some embodiments, internal labels are separated by between 1 to 3 (e.g., 1, 2, or 3) or 9 to 13 (e.g., 9, 10, 11, 12, or 13) nucleotides along the same oligonucleotide strand. As illustrated by this example, labeled oligonucleotides may be designed by consideration of a predicted or known helical structure such that the relative location of one internal label to another through space can be manipulated to the desired application. Additional examples of structural motifs useful in the design of labeled oligonucleotides are known in the art and described herein.

**[0080]** FIG. 1G is an internally-labeled oligonucleotide strand that is self-hybridized to form a stem-loop motif. A stem-loop, or hairpin loop, is an unpaired loop of nucleotides on an oligonucleotide strand that is formed when the oligonucleotide strand folds and forms base pairs with another section of the same strand. In some embodiments, the unpaired loop of a stem-loop comprises three to ten nucleotides. Accordingly, a stem-loop can be formed by two regions of an oligonucleotide strand having reverse complementary sequences that hybridize to form a stem, where the two regions are separated by the three to ten nucleotides that form the unpaired loop. In some embodiments, the stem can be designed to have one or more G/C nucleotides, which can provide added stability with the additional hydrogen bonding interactions that form compared to A/T/U nucleotides. In some embodiments, the stem comprises G/C nucleotides immediately adjacent to an unpaired loop sequence. In some embodiments, the stem comprises G/C nucleotides within the first 2, 3, 4, or 5 nucleotides adjacent to an unpaired loop sequence.

**[0081]** As described herein, in some embodiments, an internal label conjugates one portion of an oligonucleotide strand to another portion of the oligonucleotide strand. As generally

depicted in the example structures of FIGs. 1D-1G, in some embodiments, both portions of an oligonucleotide strand conjugated by an internal label are hybridized to the same oligonucleotide strand. Accordingly, an internal label can be adjacent to one or more (e.g., 1, 2, 3, 4, 5, or more) unpaired bases of a hybridized oligonucleotide strand.

**[0082]** In some embodiments, an internal label can interact with guanine nucleobases via radiative and/or non-radiative decay to effect diminished luminescence lifetime. In some embodiments, the one or more unpaired bases adjacent to an internal label are designed to exclude or minimize guanine. In some embodiments, regions surrounding an internally conjugated label are designed to exclude or minimize G/C content. In some embodiments, an internally-conjugated label is at least 2 nucleotides separated from a G or C nucleotide on the oligonucleotide strand (e.g., 2, 3, 4, 5, 6, 7, 8, 9, or more than 10 nucleotides separated from a G or C nucleotide). Thus, in some embodiments, each internal label is flanked on either side by at least 2 consecutive nucleotides selected from A or T/U.

**[0083]** The labels provided herein have applications in systems other than oligonucleotides and nucleic acids. In certain embodiments, the labeled biomolecule is a polypeptide or protein. For instance, in certain embodiments,  $Q^1$  and  $Q^2$  are independently amino acids, oligopeptides, polypeptides, proteins, or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently amino acids. In certain embodiments,  $Q^1$  and  $Q^2$  are independently oligopeptides or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently polypeptides or proteins, or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are joined together to form a cyclic peptide or cyclic protein.

**[0084]** Non-limiting examples of oligopeptides and polypeptides peptides suitable for use in labeled biomolecules of the application include, without limitation, oligopeptides, cyclic peptides, and small proteins (e.g., avian pancreatic peptide-based miniature proteins, such as described in Hodges, A.M. and Schepartz, A. (2007) *J. Am. Chem. Soc.* 129:11024-11025). Methods of engineering structural constraints into polypeptides are well known in the art and are envisioned to be particularly useful, e.g., to impart rigidity and enhance one or more luminescent properties discussed herein. For example, proline content of a peptide amino acid sequence can be modified to control peptide shape and impart rigidity (see, e.g., Kritzer, J.A., et al. (2006) *ChemBioChem* 7:29-31). Additional non-limiting examples of useful peptide engineering techniques include peptide cyclization (see, e.g., Maltsev, O.V., et al. (2016) *Angewandte Chemie* 55(4):1535-1539),  $\alpha$ -helical peptide constraint via stapling and/or H-bond surrogates (see, e.g., Douse, C.H., et al. (2014) *ACS Chem. Biol.* 9:2204-2209), peptide

constraint via cyclic  $\beta$ -sheet and  $\beta$ -hairpin mimics (see, e.g., Gibbs, A.C., et al. (1998) Nat. Struc. Biol. 5:284-288).

**[0085]** In certain embodiments, the labeled biomolecule is an oligosaccharide or a polysaccharide. For instance, in certain embodiments,  $Q^1$  and  $Q^2$  are independently monosaccharides, oligosaccharides, polysaccharides, or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently monosaccharides. In certain embodiments,  $Q^1$  and  $Q^2$  are independently oligosaccharides or fragments thereof. In certain embodiments,  $Q^1$  and  $Q^2$  are independently polysaccharides or fragments thereof. Examples of oligosaccharides and polysaccharides suitable for use in labeled biomolecules of the application are known in the art (e.g., as described in Solid Support Oligosaccharide Synthesis and Combinatorial Carbohydrate Libraries, Wiley 2001).

**[0086]** In accordance with the application, labeled biomolecules provided herein comprise a biomolecule that functions as a rigid scaffold upon which an internal label of the application is incorporated. In some embodiments, the biomolecule comprises one or more features which confer additional functions useful in various methods of detection, quantitative analysis, and imaging. For example, FIGs. 2A-2C illustrate non-limiting examples of internally-labeled biomolecules comprising a biomolecule that interacts with a desired target molecule.

**[0087]** In some embodiments, a labeled biomolecule comprising an oligonucleotide can function as a hybridization probe. A hybridization probe is a labeled fragment of DNA or RNA (e.g., an oligonucleotide) that can be added to a sample of known or unknown content to detect the presence of a desired target nucleic acid that is complementary to the hybridization probe. For example, FIG. 2A depicts an internally-labeled hybridization probe **200** hybridized with a target nucleic acid **210**. As generally illustrated by this example, internally-labeled hybridization probe **200** forms base pairing interactions with target nucleic acid **210** which results in a detectable increase in luminescence from the internal label. In some embodiments, however, hybridization of internally-labeled hybridization probe **200** with target nucleic acid **210** results in a detectable decrease in luminescence.

**[0088]** In some embodiments, an internally-labeled hybridization probe of the application comprises a sequence that is substantially complementary to a target nucleic acid sequence such that the probe and target form base pairing interactions under hybridization conditions. As described herein, an internal label conjugates one portion of a biomolecule to another portion of a biomolecule. Accordingly, one or both portions of the biomolecule (e.g., the oligonucleotide) conjugated by an internal label can be designed to hybridize with a target

nucleic acid sequence. In some embodiments, a target nucleic acid comprises RNA (e.g., mRNA). In some embodiments, a target nucleic acid comprises DNA (e.g., cDNA, genomic DNA or fragments thereof).

**[0089]** Internally-labeled hybridization probes of the application can be utilized in any methodology known in the art which utilizes hybridization probes. Examples of such techniques include, without limitation, fluorescent in situ hybridization (FISH), Northern blotting, Southern blotting, and general techniques involving SNP detection, real-time nucleic acid detection, real-time PCR quantification, allelic discrimination and identification, multiplex PCR assays, and diagnostic clinical assays.

**[0090]** Labeled biomolecules of the application, in some embodiments, comprise a biomolecule that functions as a protein ligand. For example, FIG. 2B depicts a labeled biomolecule **201** bound by a target protein **211**. As generally illustrated by this example, target protein **211** associates with (e.g., binds) at least a portion of the biomolecule (shown as dashed lines) of labeled biomolecule **201** which results in a detectable increase in luminescence from the internal label. In some embodiments, however, binding of target protein **211** to labeled biomolecule **201** results in a detectable decrease in luminescence.

**[0091]** In some embodiments, labeled biomolecule **201** and target protein **211** comprise a known binding pair. In some embodiments, labeled biomolecule **201** can be added to a sample of known or unknown content to detect presence of target protein **211**. In some embodiments, target protein **211** is a receptor and labeled biomolecule **201** comprises a receptor ligand. In some embodiments, target protein **211** is an antibody specific for at least a portion of labeled biomolecule **201**. In some embodiments, target protein **211** is an antibody and labeled biomolecule **201** comprises an antigen. In some embodiments, target protein **211** is a nucleic acid-binding protein (e.g., a DNA-binding protein) and labeled biomolecule **201** comprises a nucleic acid. Such labeled biomolecules are contemplated to be useful in methodologies that employ labeled protein ligands to detect presence of a target protein or to evaluate a protein-ligand binding interaction (e.g., fluorescence polarization and other techniques known in the art or described herein).

**[0092]** In some embodiments, a labeled biomolecule comprising a polypeptide can function as an antibody. FIG. 2C depicts an internally-labeled antibody **202** bound to a target protein **212**. As generally illustrated by this example, internally-labeled antibody **202** comprises a Fab region configured to specifically bind target protein **212** which results in a detectable increase in luminescence from the internal label. In some embodiments, however, binding of

internally-labeled antibody **202** with target protein **212** results in a detectable decrease in luminescence.

**[0093]** Internally-labeled antibodies of the application can be utilized in methods known in the art which utilizes luminescently labeled antibodies. Examples of such techniques include, without limitation, fluorescent in situ hybridization (FISH), Western blotting, general methodologies involving immunolabeling, such as immunocytochemistry and immunohistochemistry techniques, and other techniques known in the art or described herein.

**[0094]** As described above, an internal label of the application can be conjugated to a biomolecule that interacts with a desired target molecule. In some embodiments, an internal label is conjugated to a protein that associates with (e.g., binds to) a target ligand. In some embodiments, the protein is an antibody or an antigen-binding portion of an antibody. In some embodiments, the protein is an enzyme, such as a peptidase (e.g., an exopeptidase or an endopeptidase), a ribozyme, an aptazyme, a ligase, a transferase, or a tRNA synthetase. In some embodiments, an internal label is conjugated to a nucleic acid that associates with (e.g., binds to) a target ligand. In some embodiments, the nucleic acid is a nucleic acid aptamer (e.g., a DNA aptamer, an RNA aptamer, or a derivative or analog thereof).

**[0095]** In some embodiments, a labeled biomolecule comprises a biomolecule that is modified with one or more functional moieties. For example, FIGs. 3A-3B illustrate non-limiting examples of labeled biomolecules comprising moieties that interact with a target molecule and/or an internal label.

**[0096]** In accordance with the application, biomolecular scaffolds provide rigid labeling scaffolds which can be of particular benefit with techniques in which a strongly defined position of an internal label is desired. For example, Förster resonance energy transfer (FRET) and fluorescence correlation spectroscopy (FCS) have become important tools for the in vitro and in vivo investigation of conformational dynamics in biomolecules. These methods rely on the distance-dependent quenching of the fluorescence signal of a donor fluorophore either by a fluorescent acceptor fluorophore (FRET) or a non-fluorescent quencher, as used in FCS with photoinduced electron transfer (PET).

**[0097]** In some embodiments, a labeled biomolecule comprises one or more quenching moieties (e.g., fluorescent and/or non-fluorescent quenching moieties) that interact with an internal label of the labeled biomolecule. In some embodiments, such moieties can be useful in real-time PCR where an internal label's position is well defined in relation to a quencher that is cleaved by exonuclease activity. An example of this process is illustrated in FIG. 3A.

[0098] As shown in panel I, an internally-labeled hybridization probe comprising a quenching moiety 300 is hybridized with a target nucleic acid. In some embodiments, quenching moiety 300 is a non-fluorescent quencher that absorbs emission from an internal label. In some embodiments, quenching moiety 300 is a fluorescent quencher that absorbs emission of one wavelength from an internal label and emits at another wavelength.

[0099] As shown in panel II, quenching moiety 300 has been cleaved from the internally-labeled hybridization probe (e.g., by an exonuclease). This separation of quenching moiety 300 from the internal label eliminates the distance-dependent quenching effects to permit detection of a luminescence from the internal label. It should be appreciated that in some embodiments, an internally-labeled hybridization probe comprises an internal label that functions as a quencher of another label of the hybridization probe.

[00100] Quenching moiety-modified hybridization probes are known in the art and are contemplated to be useful with internal labels of the application. Examples of such hybridization probes include, without limitation, molecular beacons, TaqMan probes, Exciton-controlled hybridization-sensitive fluorescent oligonucleotide (ECHO) probes, and cycling probe technology (CPT) probes.

[00101] Accordingly, in some embodiments, multiple internal labels (e.g., two, three, four, five, or more internal labels) can be incorporated into a biomolecule according to the desired luminescent properties of a labeled biomolecule provided herein. For example, in some embodiments, a labeled biomolecule having two or more internal labels exhibits increased luminescence intensity and/or brightness relative to the biomolecule having one internal label. In some embodiments, the two or more internal labels are configured to provide independent reporter signals. In some embodiments, the two or more internal labels are configured to provide dependent reporter signals (e.g., a donor label and an acceptor label of a FRET pair).

[00102] In some aspects, the application provides internal labels configured for use in conventional solid-phase synthesis techniques, e.g., phosphoramidite analogs useful in oligonucleotide synthesis. Accordingly, in some embodiments, labels provided herein can be readily incorporated into a biomolecule to generate a labeled biomolecule having a number of internal labels that may be limited only by the desired size of the biomolecule. For example, in some embodiments, a labeled biomolecule comprises two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) internal labels. In some embodiments, a labeled biomolecule comprises between 2 and 5, between 2 and 10, between 5 and 10, between 5 and 15, between 10 and 15, between 15 and 20, or more internal labels.

**[00103]** In some embodiments, a labeled biomolecule provided herein further comprises one or more luminescent labels other than the polycyclic fluorophore of Formula (I). For example, in some embodiments, a labeled biomolecule comprises at least one internal label according to Formula (I), and one or more internal labels comprising a linear or non-polycyclic fluorophore. In some embodiments, a labeled biomolecule comprises at least one internal label according to Formula (I), and one or more external labels.

**[00104]** In some embodiments, an external label refers to a label (e.g., a fluorophore) conjugated to a single site of a labeled biomolecule provided herein. In some embodiments, an external label is conjugated to a labeled biomolecule at a terminal end. For example, in some embodiments, an external label is conjugated to a 5' or 3' end of a labeled oligonucleotide. In some embodiments, an external label is conjugated to an N- or C-terminus of a labeled polypeptide. In some embodiments, an external label is conjugated to a labeled biomolecule at a terminal monomer of the biomolecule (e.g., conjugated to a base of a terminal nucleotide in an oligonucleotide strand, conjugated to a side chain of a terminal amino acid in a polypeptide strand).

**[00105]** In some embodiments, an external label is conjugated to a labeled biomolecule at a non-terminal site of the biomolecule. In some embodiments, an external label is conjugated to a labeled biomolecule at a site that is between monomers of the biomolecule. In some embodiments, an external label is conjugated to an abasic site of a labeled oligonucleotide. In some embodiments, an external label is conjugated to a labeled biomolecule at a non-terminal monomer of the biomolecule (e.g., conjugated to a base of a non-terminal nucleotide in an oligonucleotide strand, conjugated to a side chain of a non-terminal amino acid in a polypeptide strand).

**[00106]** Labeled biomolecules of the application, in some embodiments, comprise a biomolecule modified with one or more moieties that function as protein ligands. For example, FIG. 3B illustrates a process whereby a labeled biomolecule comprising a ligand moiety is detectably bound by a target protein. As shown in panel I, a target protein is exposed to a labeled biomolecule comprising a ligand moiety **301** configured to bind the target protein. In the absence of binding between the target protein and ligand moiety **301**, the internal label of the labeled biomolecule does not emit a detectable signal.

**[00107]** As shown in panel II, the target protein associates with (e.g., binds) ligand moiety **301** which results in a detectable increase in luminescence from the internal label. In some embodiments, however, binding of the target protein to ligand moiety **301** results in a detectable decrease in luminescence. In some embodiments, the change in detectable

luminescence upon binding occurs as a result of confinement of the internal label biomolecule to an observation region (e.g., immobilization to a surface for a period of time sufficient to permit detection). In some embodiments, the change in luminescence upon binding occurs as a result of FRET interactions (e.g., with the target protein or a surface-conjugated acceptor/donor).

**[00108]** In some embodiments, labeled biomolecules comprising one or more ligand moieties can be used, e.g., for purposes of immobilizing a labeled biomolecule to a surface or a material, for detection of target protein in a known or unknown sample, for detection (e.g., quantitation) of a binding interaction between the target protein and a ligand moiety that is known or not known to bind the target protein.

**[00109]** In certain embodiments, the labeled biomolecule is associated with a reactant configured for use as a substrate in a reaction. For example, in certain embodiments,  $Q^1$  and  $Q^2$  of Formula (I) are independently optionally associated with a reactant configured for use as a substrate in a reaction. In the case of an oligonucleotide or nucleic acid system, for example, the first and second oligonucleotide strands are independently optionally associated with a reactant configured for use as a substrate in a reaction. In some embodiments, the reactant is configured for use as a substrate in a polymerization reaction. In some embodiments, the reactant is cleaved from the labeled biomolecule by a polymerase when subjected to polymerization reaction conditions. For example, in some embodiments, the reactant is a nucleotide (e.g., for use in a method of sequencing a nucleic acid).

#### Labeled Nucleotides

**[00110]** Also provided herein are labeled nucleotides comprising one or more nucleotides associated with a labeled biomolecule described herein. In some embodiments, the one or more nucleotides comprise one type of nucleotide selected from guanine, cytosine, adenine, and thymine or uracil. In some embodiments, the one or more nucleotides are cleaved from the labeled biomolecule by a polymerase when subjected to polymerization reaction conditions.

**[00111]** Without wishing to be bound by any particular theory, labeled nucleotides provided herein offer a number of distinct advantages over those currently used in sequencing reactions, such as increased readlength and increased accuracy, in addition to the advantages described elsewhere herein. FIGs. 4A-4C highlight several features of labeled nucleotides of the disclosure.

**[00112]** Each of FIGs. 4A-4C depicts a nucleotide **400** bound by a polymerase **410**. Nucleotide **400** of FIG. 4A is externally-conjugated to a luminescent label, whereas nucleotides of FIGs. 4B and 4C are associated with different labeled biomolecules described herein. As illustrated in FIG. 4A, the externally-conjugated label is relatively proximal to the polymerase and has a relatively high degree of access to bulk solvent molecules. In some embodiments, such characteristics are adverse to a polymerization reaction. For example, in some embodiments, a shorter distance between a label and a polymerase can result in label-induced damage to the polymerase via radiative and/or non-radiative decay (shown as path (i)). In some embodiments, a higher degree of a label's access to bulk solvent molecules can result in a higher incidence of reactive oxygen species (ROS) formation. Once formed, ROS can damage the polymerase to adversely affect enzymatic activity (shown as path (ii)).

**[00113]** FIG. 4B depicts a nucleotide associated with a labeled biomolecule, such as a labeled oligonucleotide strand. As shown relative to the externally-conjugated label, the biomolecule increases separation between the label and the polymerase. Additionally, the portion of the biomolecule between the nucleotide and the internal label provides a protective barrier between the label and polymerase. Accordingly, the occurrence of label-induced damage to the polymerase can be decreased, due to the label-polymerase separation and/or due to the biomolecule absorbing any decay emitted from the label (shown as path (iii)).

**[00114]** Also as shown relative to FIG. 4A, integration of the internal dye into the biomolecule shown in FIG. 4B decreases the extent to which the label is exposed to bulk solvent molecules. Accordingly, ROS-induced damage can be decreased due to lowered incidence of ROS formation as a result of decreased access of label to bulk solvent, and/or due to the biomolecule absorbing any ROS-induced damage, and/or due to free radical decay over the label-polymerase separation distance (shown as path (iv)).

**[00115]** FIG. 4C depicts a nucleotide associated with a labeled oligonucleotide. As shown, the labeled oligonucleotide is hybridized with an unlabeled oligonucleotide strand. In accordance with the application, such constructs impart a high degree of rigidity that enhance each of the above advantages described for the labeled nucleotide of FIG. 4B.

**[00116]** For example, the hybridized strand provides increased rigidity, meaning less overall flexibility which further promotes label-polymerase separation. The hybridized strand also provides another barrier between the label and the polymerase that can absorb any label-induced decay (shown as path (v)). Additionally, ROS-induced damage is further decreased due to the hybridized strand further restricting access of the label to bulk solvent, and/or due

to the hybridized strand absorbing any ROS-induced damage, and/or due to free radical decay over the increased label-polymerase separation distance.

**[00117]** Accordingly, in each of the above examples, the advantages provided by the labeled biomolecules of the application provide increased readlength in sequencing reactions by limiting the extent of photo-induced damage to the polymerase. Additionally, labeled nucleotides of the application provide increased accuracy, e.g., as a result of the enhancement of one or more emission characteristics.

**[00118]** It should be understood that, in the context of a labeled biomolecule, a “nucleotide” or “nucleoside polyphosphate” attached thereto refers to the one or more nucleotides (e.g., nucleoside polyphosphates) that are configured to be incorporated into a growing nucleic acid strand (e.g., during a sequencing reaction). In some embodiments, the one or more nucleotides comprise one or more nucleoside monophosphates or nucleoside polyphosphates. Examples of nucleoside polyphosphates include, in some embodiments, nucleoside di- or triphosphates, or nucleosides with more than three 5' phosphates, such as nucleoside hexaphosphates. In some embodiments of any of the compositions or methods described in this application, a phosphate portion (e.g., a polyphosphate portion) of a nucleotide includes one or more phosphates (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more phosphate groups) or variants thereof. For example, in some embodiments, a phosphate portion (e.g., a polyphosphate portion) of a nucleotide can include a phosphate ester, a thioester, a phosphoramidate, an alkyl phosphonate linkage, other suitable linkage, or more than one such modifications, or a combination of two or more thereof.

**[00119]** A labeled nucleotide can be a terminal phosphate labeled nucleotide, such that a labeled biomolecule of the application is attached to a terminal phosphate of the nucleotide. For example, in some embodiments, one or more nucleotides may be attached through a terminal phosphate to a biomolecule (e.g., Q<sup>1</sup> and/or Q<sup>2</sup> of Formula (I)) that forms part of a labeled biomolecule as described in this application. Accordingly, in some embodiments, a “labeled nucleotide” of the application refers to a nucleotide attached to a labeled biomolecule of Formula (I). In some embodiments, the one or more nucleotides may be attached through a terminal phosphate to an oligonucleotide (e.g., an unlabeled oligonucleotide strand) that forms part of a labeled biomolecule as described in this application.

**[00120]** A labeled biomolecule can be attached to a terminal phosphate of a nucleotide through a linker. The linker can include, for example, at least one or a plurality of hydroxyl groups, sulfhydryl groups, amino groups or haloalkyl groups, which may be suitable for

forming, for example, a phosphate ester, a thioester, a phosphoramidate or an alkyl phosphonate linkage at the terminal phosphate of a natural or modified nucleotide. A linker can be cleavable so as to separate a label from the terminal phosphate, such as with the aid of a polymerization enzyme. Examples of nucleotides and linkers are provided in U.S. Patent No. 7,041,812, which is entirely incorporated herein by reference. In some embodiments, the linker comprises optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof. Additional examples of linkers useful for attaching a label to a nucleotide can be found in co-pending U.S. Patent App. No. 15/600,979, the relevant portions of which are incorporated herein by reference in entirety.

**[00121]** A nucleotide (e.g., a nucleoside polyphosphate) can comprise any of an adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U), or variants thereof. A nucleotide (e.g., a nucleoside polyphosphate) can comprise a methylated nucleobase. For example, a methylated nucleotide can be a nucleotide that comprises one or more methyl groups attached to the nucleobase (e.g., attached directly to a ring of the nucleobase, attached to a substituent of a ring of the nucleobase). Exemplary methylated nucleobases include 1-methylthymine, 1-methyluracil, 3-methyluracil, 3-methylcytosine, 5-methylcytosine, 1-methyladenine, 2-methyladenine, 7-methyladenine, N6-methyladenine, N6,N6-dimethyladenine, 1-methylguanine, 7-methylguanine, N2-methylguanine, and N2,N2-dimethylguanine.

**[00122]** The term “nucleic acid,” as used herein, generally refers to a molecule comprising one or more nucleic acid subunits. A nucleic acid may include one or more subunits selected from adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U), or variants thereof. In some examples, a nucleic acid is deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), or derivatives thereof. In some embodiments, the nucleic acid is a modified nucleic acid, including, without limitation, a locked nucleic acid (LNA), a peptide nucleic acid (PNA), a triazole-linked nucleic acid, a 2'-F-modified nucleic acid, and derivatives and analogs thereof. A nucleic acid may be single-stranded or double stranded. In some embodiments, a nucleic acid generally refers to any polymer of nucleotides.

Emission Characteristics

**[00123]** As described herein, internal conjugation of a label into a biomolecule can alter the photophysical properties of the label (e.g., via restricted rotation or immobilization of the label within the biomolecule). Therefore, in certain embodiments, one or more emission characteristics of the labeled biomolecule are altered (e.g., increased) relative to an unconjugated molecule comprising the label. The “unconjugated molecule” comprising the label, as described herein, does not comprise one or both of Q<sup>1</sup> and Q<sup>2</sup>. The one or more emission characteristics that are altered by internal conjugation can include, but are not limited to, luminescence lifetime, luminescence intensity, brightness, emission maximum, luminescence quantum yield, and photostability.

**[00124]** In certain embodiments, luminescence lifetime of a labeled biomolecule is increased relative to the unconjugated molecule. In certain embodiments, the luminescence lifetime of the labeled biomolecule is increased by at least 10% relative to the unconjugated molecule. In certain embodiments, the luminescence lifetime of the labeled biomolecule is increased by between approximately 10% and 50% (e.g., between about 10% and 25%, between about 10% and 15%, between about 25% and 50%, between about 40% and 50%) relative to the unconjugated molecule. An example of increased luminescence lifetime is shown in FIGs. 5A-5B.

**[00125]** A set of labeled biomolecules comprising nucleotides were prepared in accordance with the constructs generically depicted in FIG. 5A. Labeled nucleotides (1) and (2) represent conjugates wherein a Cy3B dye was externally conjugated to a DNA linker at the terminus or a branch point, respectively, of the oligonucleotide. By contrast, labeled nucleotide (3) exchanges a base within the oligonucleotide strand for the Cy3B dye itself to become internally incorporated into the DNA backbone. These labeled nucleotides were used in sequencing experiments, and lifetime measurements for each conjugate were obtained. The externally conjugated Cy3B constructs (1) and (2) produced lifetimes of approximately 2.2 nanoseconds, whereas the measured lifetime for the internally conjugated Cy3B construct (3) was 2.6 nanoseconds—an increase in lifetime of approximately 15-20% for the internally conjugated dye.

**[00126]** In certain embodiments, luminescence intensity of a labeled biomolecule is increased relative to the unconjugated molecule. In certain embodiments, luminescence intensity of the labeled biomolecule is increased by between approximately 5% and 25% (e.g., between about 5% and 20%, between about 5% and 15%, between about 5% and 10%, between about 10% and 25%, between about 15% and 25%, between about 20% and 25%) relative to the unconjugated molecule. In some embodiments, luminescence intensity of the

labeled biomolecule is increased by approximately 5%, approximately 10%, approximately 15%, approximately 20%, approximately 25%, or more relative to the unconjugated molecule.

**[00127]** In certain embodiments, brightness of a labeled biomolecule is increased relative to the unconjugated molecule. In certain embodiments, brightness of the labeled biomolecule is increased by between approximately 5% and 10% relative to the unconjugated molecule. In certain embodiments, brightness of the labeled biomolecule is increased by between approximately 5% and 25% (e.g., between about 5% and 20%, between about 5% and 15%, between about 5% and 10%, between about 10% and 25%, between about 15% and 25%, between about 20% and 25%) relative to the unconjugated molecule. In some embodiments, brightness of the labeled biomolecule is increased by approximately 5%, approximately 10%, approximately 15%, approximately 20%, approximately 25%, or more relative to the unconjugated molecule.

**[00128]** In certain embodiments, emission maximum of a labeled biomolecule is increased by at least 1% relative to the unconjugated molecule. In certain embodiments, emission maximum of the labeled biomolecule is increased by between approximately 1% and 10% (e.g., between about 1% and 5%, between about 5% and 10%) relative to the unconjugated molecule. An example of increased emission maximum is shown in FIG. 6. In some embodiments, emission maximum of the labeled biomolecule is increased by approximately 1%, approximately 2%, approximately 5%, approximately 10%, or more relative to the unconjugated molecule.

**[00129]** Bulk fluorescence data obtained for an internally-conjugated Cy3B and an externally-conjugated Cy3B are depicted in FIG. 6. As shown, the excitation spectrum of the internal dye is red-shifted by 8 nm relative to the external dye, which places the vibronic shoulder (the hump on the left of the trace) closer to 532 nm. Also as shown, the emission spectrum of the internal dye is red-shifted, which was found to advantageously increase detectable signals as more light was permitted past the filter.

**[00130]** In certain embodiments, luminescence quantum yield of a labeled biomolecule is increased relative to the unconjugated molecule. In certain embodiments, luminescence quantum yield of the labeled biomolecule is increased by between approximately 5% and 25% (e.g., between about 5% and 20%, between about 5% and 15%, between about 5% and 10%, between about 10% and 25%, between about 15% and 25%, between about 20% and 25%) relative to the unconjugated molecule. In some embodiments, luminescence quantum yield of the labeled biomolecule is increased by approximately 5%, approximately 10%,

approximately 15%, approximately 20%, approximately 25%, or more relative to the unconjugated molecule.

**[00131]** In certain embodiments, photostability of a labeled biomolecule is increased relative to the unconjugated molecule. As used herein, in some embodiments, photostability refers to the ability of a luminescent molecule to continue to fluoresce over time. In some embodiments, photostability can be evaluated by measuring the rate of photobleaching. For example, in some embodiments, the rate of photobleaching can be measured for a labeled biomolecule (e.g., a biomolecule having an internally-conjugated label) and compared to the rate of photobleaching measured for the unconjugated molecule. A measured decrease in the rate of photobleaching would be indicative of increased photostability. Methods of measuring photobleaching rates are known in the art, e.g., as described in Wüstner, D., et al. (2014) *Molecules*, 9:11096-11130; Brakenhoff, G.J., et al. (1994) *Journal of Microscopy*, 175(2):154-161; and Song, L., et al. (1995) *Biophys J.* 68(6):2588-2600. In some embodiments, a labeled biomolecule has decreased photobleaching relative to the unconjugated molecule, as measured by fluorescence recovery after photobleaching (FRAP), e.g., as described in Meyvis, T., et al. (1999) *Pharmaceutical Research*, 16(8):1153-1162. In some embodiments, a labeled biomolecule has decreased photobleaching relative to the unconjugated molecule, as measured by fluorescence loss in photobleaching (FLIP), e.g., as described in Wüstner, D., et al. (2012) *BMC Bioinformatics*, 13:296.

**[00132]** In some embodiments, the disclosure provides new compositions for identifying single molecules based on one or more luminescent properties of those molecules. In some embodiments, a molecule (e.g., a luminescently labeled nucleotide) is identified based on its brightness, luminescence lifetime, absorption spectra, emission spectra, luminescence quantum yield, luminescence intensity, or a combination of two or more thereof. Identifying may mean assigning the exact molecular identity of a molecule, or may mean distinguishing or differentiating the particular molecule from a set of possible molecules. In some embodiments, a plurality of single molecules can be distinguished from each other based on different brightnesses, luminescence lifetimes, absorption spectra, emission spectra, luminescence quantum yields, luminescence intensities, or combinations of two or more thereof. In some embodiments, a single molecule is identified (e.g., distinguished from other molecules) by exposing the molecule to a series of separate light pulses and evaluating the timing or other properties of each photon that is emitted from the molecule. In some embodiments, information for a plurality of photons emitted sequentially from a single molecule is aggregated and evaluated to identify the molecule. In some

embodiments, a luminescence lifetime of a molecule is determined from a plurality of photons that are emitted sequentially from the molecule, and the luminescence lifetime can be used to identify the molecule. In some embodiments, a luminescence intensity of a molecule is determined from a plurality of photons that are emitted sequentially from the molecule, and the luminescence intensity can be used to identify the molecule. In some embodiments, a luminescence lifetime and luminescence intensity of a molecule is determined from a plurality of photons that are emitted sequentially from the molecule, and the luminescence lifetime and luminescence intensity can be used to identify the molecule.

**[00133]** Accordingly, in some aspects of the application, a reaction sample is exposed to a plurality of separate light pulses and a series of emitted photons are detected and analyzed. In some embodiments, the series of emitted photons provides information about a single molecule that is present and that does not change in the reaction sample over the time of the experiment. However, in some embodiments, the series of emitted photons provides information about a series of different molecules that are present at different times in the reaction sample (e.g., as a reaction or process progresses).

**[00134]** Determination of a luminescence lifetime of a molecule can be performed using any suitable method (e.g., by measuring the lifetime using a suitable technique or by determining time-dependent characteristics of emission). In some embodiments, determining the luminescence lifetime of a molecule comprises determining the lifetime relative to one or more molecules (e.g., different luminescently labeled nucleotides in a sequencing reaction). In some embodiments, determining the luminescence lifetime of a molecule comprises determining the lifetime relative to a reference. In some embodiments, determining the luminescence lifetime of a molecule comprises measuring the lifetime (e.g., fluorescence lifetime). In some embodiments, determining the luminescence lifetime of a molecule comprises determining one or more temporal characteristics that are indicative of lifetime. In some embodiments, the luminescence lifetime of a molecule can be determined based on a distribution of a plurality of emission events (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, or more emission events) occurring across one or more time-gated windows relative to an excitation pulse. For example, a luminescence lifetime of a single molecule can be distinguished from a plurality of molecules having different luminescence lifetimes based on the distribution of photon arrival times measured with respect to an excitation pulse.

**[00135]** It should be appreciated that a luminescence lifetime of a single molecule is indicative of the timing of photons emitted after the single molecule reaches an excited state

and the single molecule can be distinguished by information indicative of the timing of the photons. Some embodiments may include distinguishing a molecule from a plurality of molecules based on the molecule's luminescence lifetime by measuring times associated with photons emitted by the molecule. The distribution of times may provide an indication of the luminescence lifetime which may be determined from the distribution. In some embodiments, the single molecule is distinguishable from the plurality of molecules based on the distribution of times, such as by comparing the distribution of times to a reference distribution corresponding to a known molecule. In some embodiments, a value for the luminescence lifetime is determined from the distribution of times.

**[00136]** As used herein for single molecules, luminescence intensity refers to the number of emitted photons per unit time that are emitted by a molecule which is being excited by delivery of a pulsed excitation energy. In some embodiments, the luminescence intensity refers to the detected number of emitted photons per unit time that are emitted by a molecule which is being excited by delivery of a pulsed excitation energy, and are detected by a particular sensor or set of sensors.

**[00137]** In some aspects, the disclosure provides methods and compositions related to labeled biomolecules having enhanced emission brightness. As used herein, in some embodiments, "brightness" (and variations thereof, e.g., "bright," "brightly," etc.) refers to a parameter that reports on the average emission intensity per labeled reactant molecule. Thus, in some embodiments, "emission intensity" may be used to generally refer to brightness of a composition comprising brightly labeled reactants. In some embodiments, brightness of a labeled reactant is equal to the product of its quantum yield and extinction coefficient. In some embodiments, the labeled biomolecules of the disclosure are engineered to maximize quantum yield to promote increased brightness.

**[00138]** Luminescence quantum yield refers to the fraction of excitation events at a given wavelength or within a given spectral range that lead to an emission event, and is typically less than 1. In some embodiments, the luminescence quantum yield of a molecule described herein is between 0 and about 0.001, between about 0.001 and about 0.01, between about 0.01 and about 0.1, between about 0.1 and about 0.5, between about 0.5 and 0.9, or between about 0.9 and 1. In some embodiments, a molecule is identified by determining or estimating the luminescence quantum yield.

**[00139]** In some embodiments, internal labels described herein allow for the addition of successive luminescent labels to a labeled biomolecule for increasing brightness and/or luminescence intensity. In some embodiments, internally-labeled biomolecules comprising

two or more luminescent labels exhibit brightness and/or luminescent intensity according to the formula  $L_n(x)$ , where  $L_n$  is equal to the total number of luminescent labels on a labeled reactant and  $x$  is equal to the measured brightness or fluorescent intensity of the corresponding singly-labeled reactant. Accordingly, in some embodiments, a two-dye labeled reaction component possesses brightness and/or luminescent intensity that is doubled compared to the one-dye labeled analog. In some embodiments, a three- or four-dye labeled reaction component possesses brightness and/or luminescent intensity that is tripled or quadrupled, respectively, compared to the one-dye labeled analog. In some embodiments, the brightly labeled reactants described herein exhibit brightness and/or luminescent intensity that is at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, or at least 99% of the value predicted by  $L_n(x)$ .

#### Nucleic Acid Sequencing Reaction Compositions

**[00140]** Also provided herein are nucleic acid sequencing reaction compositions comprising two or more different types of labeled nucleotides in a reaction mixture, wherein at least one type of labeled nucleotide is a labeled nucleotide comprising a labeled biomolecule described herein.

**[00141]** In some embodiments, a nucleic acid sequencing reaction composition comprises two or more (e.g., two, three, four, five, or more) different types of labeled nucleotides. In some embodiments, a nucleic acid sequencing reaction composition comprises four different types of labeled nucleotides. In some embodiments, the four different types of labeled nucleotides comprising a first labeled nucleotide comprising guanine, a second labeled nucleotide comprising cytosine, a third labeled nucleotide comprising adenine, and a fourth labeled nucleotide comprising thymine or uracil.

**[00142]** In some embodiments, each type of labeled nucleotide in a nucleic acid sequencing reaction composition is present at a concentration of between about 100 and 1000 nM (e.g., between about 100 and 800 nM, between about 150 and 700 nM, between about 200 and 600 nM, or between about 250 and 500 nM).

**[00143]** In some embodiments, a nucleic acid sequencing reaction composition comprises a sequencing template comprising a polymerase in a complex with a target nucleic acid. In some embodiments, the complex further comprises a primer oligonucleotide having a sequence complementary to a portion of the target nucleic acid. In some embodiments, the complex is present at a concentration of between about 10 pM and 10 nM (e.g., between about 25 pM and 5 nM, between about 50 pM and 2 nM, between about 50 pM and 1 nM,

between about 50 pM and 500 pM, between about 50 pM and 100 pM, between about 250 pM and 5 nM, between about 250 pM and 2 nM, between about 250 pM and 1 nM, or between about 250 pM and 500 pM.

**[00144]** In some embodiments, a nucleic acid sequencing reaction composition comprises one or more buffering agents (e.g., MES, MOPS, MOPSO, HEPES, Tris, TAPS, and other such suitable buffering agents known in the art). In some embodiments, the one or more buffering agents comprises MOPS. In some embodiments, a buffering agent is present at a concentration of between about 25 and 100 mM (e.g., between about 25 and 75 mM, between about 50 and 75 mM, or approximately 65 mM).

**[00145]** In some embodiments, a nucleic acid sequencing reaction composition comprises a divalent cation (e.g., magnesium ion, calcium ion). In some embodiments, the divalent cation comprises a magnesium or calcium salt (e.g., a salt comprising magnesium or calcium and acetate, chloride, phosphate, sulfate). In some embodiments, the salt is magnesium acetate. In some embodiments, a divalent cation is present at a concentration of between about 5 and 50 mM (e.g., between about 10 and 40 mM, between about 15 and 35 mM, between about 20 and 30 mM, or approximately 25 mM).

**[00146]** In some embodiments, a nucleic acid sequencing reaction composition comprises one or more monovalent salts (e.g., a sodium or potassium salt, such as sodium chloride, sodium acetate, potassium chloride, or potassium acetate). In some embodiments, the monovalent salt is present at a concentration of between about 10 and 200 mM (e.g., between about 25 and 150 mM, between about 25 and 40 mM, between about 50 and 150 mM, between about 100 and 150 mM). In some embodiments, a nucleic acid sequencing reaction composition comprises approximately 40 mM monovalent salt, such as 40 mM sodium chloride. In some embodiments, a nucleic acid sequencing reaction composition comprises approximately 120 mM monovalent salt, such as 120 mM potassium acetate.

**[00147]** In some embodiments, a nucleic acid sequencing reaction composition comprises one or more photostabilizers (e.g., one or more photoprotective additives, such as antioxidants, oxygen scavengers, triplet state quenchers, and similar energy-absorbing additives known in the art). In some embodiments, a photostabilizer comprises protocatechuic acid (PCA). In some embodiments, a photostabilizer comprises 4-nitrobenzyl alcohol (NBA). In some embodiments, a photostabilizer comprises trolox, or a derivative thereof. In some embodiments, a photostabilizer is present in a concentration of between about 0.1 mM and about 20 mM. In some embodiments, the concentration of trolox is about 5 mM. In some embodiments, the concentration of PCA is about 3 mM. In some embodiments,

the concentration of PCA is about 8 mM. In some embodiments, the concentration of NBA is about 3 mM. A mixture with a photostabilizer (e.g., PCA) may also comprise an enzyme to regenerate the photostabilizer (e.g., protocatechuic acid dioxygenase (PCD)). In some embodiments, the concentration of PCD is about 0.3 mM. In some embodiments, the concentration of PCD is about 0.5 mg/mL.

**[00148]** In some embodiments, a nucleic acid sequencing reaction composition comprises one or more reducing agents. For example, in some embodiments, a nucleic acid sequencing reaction composition comprises between about 10 and 100 mM DTT (e.g., approximately 40 mM DTT).

### Sequencing

**[00149]** Some aspects of the application are useful for sequencing biological polymers, such as nucleic acids and proteins. In some aspects, compositions and techniques described in the application can be used to identify a series of nucleotide or amino acid monomers that are incorporated into a nucleic acid or protein (e.g., by detecting a time-course of incorporation of a series of labeled nucleotide or amino acid monomers). In some embodiments, compositions and techniques described in the application can be used to identify a series of nucleotides that are incorporated into a template-dependent nucleic acid sequencing reaction product synthesized by a polymerase enzyme.

**[00150]** Accordingly, also provided herein are methods of determining the sequence of a template nucleic acid using the nucleic acid sequencing reaction compositions of the application. In some embodiments, methods of sequencing comprise steps of: (i) exposing a complex in a target volume, the complex comprising the template nucleic acid, a primer, and a polymerizing enzyme, to a nucleic acid sequencing reaction composition according to the disclosure (e.g., at least one labeled nucleotide comprising a labeled biomolecule described herein); (ii) directing a series of pulses of one or more excitation energies towards a vicinity of the target volume; (iii) detecting a plurality of emitted photons from luminescently labeled nucleotides during sequential incorporation into a nucleic acid comprising the primer; and (iv) identifying the sequence of incorporated nucleotides by determining timing and optionally luminescence intensity of the emitted photons.

**[00151]** In some embodiments, as used herein, an excitation energy is a pulse of light from a light source. In some embodiments, an excitation energy is in the visible spectrum. In some embodiments, an excitation energy is in the ultraviolet spectrum. In some embodiments, an excitation energy is in the infrared spectrum. In some embodiments, an

excitation energy is at or near the absorption maximum of a luminescently labeled molecule from which a plurality of emitted photons are to be detected. In certain embodiments, the excitation energy is between about 500 nm and about 700 nm (e.g., between about 500 nm and about 600 nm, between about 600 nm and about 700 nm, between about 500 nm and about 550 nm, between about 550 nm and about 600 nm, between about 600 nm and about 650 nm, or between about 650 nm and about 700 nm). In certain embodiments, an excitation energy may be monochromatic or confined to a spectral range. In some embodiments, a spectral range has a range of between about 0.1 nm and about 1 nm, between about 1 nm and about 2 nm, or between about 2 nm and about 5 nm. In some embodiments a spectral range has a range of between about 5 nm and about 10 nm, between about 10 nm and about 50 nm, or between about 50 nm and about 100 nm.

**[00152]** Upon base pairing between a nucleobase of a target nucleic acid and the complementary nucleoside polyphosphate (e.g., dNTP), the polymerase incorporates the dNTP into the newly synthesized nucleic acid strand by forming a phosphodiester bond between the 3' hydroxyl end of the newly synthesized strand and the alpha phosphate of the dNTP. In examples in which a luminescent molecule (e.g., a labeled biomolecule as described herein) conjugated to the dNTP comprises a fluorophore, its presence is signaled by excitation and a pulse of emission is detected during and/or after the step of incorporation. For luminescent molecules (e.g., labeled biomolecules) that are conjugated to the terminal (gamma) phosphate of the dNTP, incorporation of the dNTP into the newly synthesized strand results in release of the beta and gamma phosphates and the luminescent molecule, which is free to diffuse in the sample well, resulting in a decrease in emission detected from the fluorophore.

**[00153]** In certain embodiments, the template-dependent nucleic acid sequencing product is carried out by naturally occurring nucleic acid polymerases. In some embodiments, the polymerase is a mutant or modified variant of a naturally occurring polymerase. In some embodiments, the template-dependent nucleic acid sequence product will comprise one or more nucleotide segments complementary to the template nucleic acid strand. In one aspect, the application provides a method of determining the sequence of a template (or target) nucleic acid strand by determining the sequence of its complementary nucleic acid strand.

**[00154]** The term "polymerase," as used herein, generally refers to any enzyme (or polymerizing enzyme) capable of catalyzing a polymerization reaction. Examples of polymerases include, without limitation, a nucleic acid polymerase, a transcriptase or a ligase. A polymerase can be a polymerization enzyme. Embodiments directed towards single

molecule nucleic acid extension (e.g., for nucleic acid sequencing) may use any polymerase that is capable of synthesizing a nucleic acid complementary to a target nucleic acid molecule. In some embodiments, a polymerase may be a DNA polymerase, an RNA polymerase, a reverse transcriptase, and/or a mutant or altered form of one or more thereof.

**[00155]** Examples of polymerases include, but are not limited to, a DNA polymerase, an RNA polymerase, a thermostable polymerase, a wild-type polymerase, a modified polymerase, E. coli DNA polymerase I, T7 DNA polymerase, bacteriophage T4 DNA polymerase  $\phi 29$  (psi29) DNA polymerase, Taq polymerase, Tth polymerase, Tli polymerase, Pfu polymerase, Pwo polymerase, VENT polymerase, DEEPVENT polymerase, EX-Taq polymerase, LA-Taq polymerase, Sso polymerase, Poc polymerase, Pab polymerase, Mth polymerase, ES4 polymerase, Tru polymerase, Tac polymerase, Tne polymerase, Tma polymerase, Tca polymerase, Tih polymerase, Tfi polymerase, Platinum Taq polymerases, Tbr polymerase, Tfl polymerase, Tth polymerase, Pfutubo polymerase, Pyrobest polymerase, Pwo polymerase, KOD polymerase, Bst polymerase, Sac polymerase, Klenow fragment, polymerase with 3' to 5' exonuclease activity, and variants, modified products and derivatives thereof. In some embodiments, the polymerase is a single subunit polymerase. Non-limiting examples of DNA polymerases and their properties are described in detail in, among other places, DNA Replication 2nd edition, Kornberg and Baker, W. H. Freeman, New York, N.Y. (1991).

**[00156]** In another aspect, the application provides methods of sequencing target nucleic acids by sequencing a plurality of nucleic acid fragments, wherein the target nucleic acid comprises the fragments. In certain embodiments, the method comprises combining a plurality of fragment sequences to provide a sequence or partial sequence for the parent target nucleic acid. In some embodiments, the step of combining is performed by computer hardware and software. The methods described herein may allow for a set of related target nucleic acids, such as an entire chromosome or genome to be sequenced.

**[00157]** During sequencing, a polymerizing enzyme may couple (e.g., attach) to a priming location of a target nucleic acid molecule. The priming location can be a primer that is complementary to a portion of the target nucleic acid molecule. As an alternative the priming location is a gap or nick that is provided within a double stranded segment of the target nucleic acid molecule. A gap or nick can be from 0 to at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, or 40 nucleotides in length. A nick can provide a break in one strand of a double stranded sequence, which can provide a priming location for a polymerizing enzyme, such as, for example, a strand displacing polymerase enzyme.

**[00158]** In some cases, a sequencing primer can be annealed to a target nucleic acid molecule that may or may not be immobilized to a solid support. A solid support can comprise, for example, a sample well (e.g., a nanoaperture, a reaction chamber) on a chip used for nucleic acid sequencing. In some embodiments, a sequencing primer may be immobilized to a solid support and hybridization of the target nucleic acid molecule also immobilizes the target nucleic acid molecule to the solid support. In some embodiments, a polymerase is immobilized to a solid support and soluble primer and target nucleic acid are contacted to the polymerase. However, in some embodiments a complex comprising a polymerase, a target nucleic acid and a primer is formed in solution and the complex is immobilized to a solid support (e.g., via immobilization of the polymerase, primer, and/or target nucleic acid). In some embodiments, none of the components in a sample well (e.g., a nanoaperture, a reaction chamber) are immobilized to a solid support. For example, in some embodiments, a complex comprising a polymerase, a target nucleic acid, and a primer is formed in solution and the complex is not immobilized to a solid support.

**[00159]** Under appropriate conditions, a polymerase enzyme that is contacted to an annealed primer/target nucleic acid can add or incorporate one or more nucleotides onto the primer, and nucleotides can be added to the primer in a 5' to 3', template-dependent fashion. Such incorporation of nucleotides onto a primer (e.g., via the action of a polymerase) can generally be referred to as a primer extension reaction. Each nucleotide can be associated with a detectable label that can be detected and identified (e.g., based on its luminescent lifetime and/or other characteristics) during the nucleic acid extension reaction and used to determine each nucleotide incorporated into the extended primer and, thus, a sequence of the newly synthesized nucleic acid molecule. Via sequence complementarity of the newly synthesized nucleic acid molecule, the sequence of the target nucleic acid molecule can also be determined. In some cases, annealing of a sequencing primer to a target nucleic acid molecule and incorporation of nucleotides to the sequencing primer can occur at similar reaction conditions (e.g., the same or similar reaction temperature) or at differing reaction conditions (e.g., different reaction temperatures). In some embodiments, sequencing by synthesis methods can include the presence of a population of target nucleic acid molecules (e.g., copies of a target nucleic acid) and/or a step of amplification of the target nucleic acid to achieve a population of target nucleic acids. However, in some embodiments sequencing by synthesis is used to determine the sequence of a single molecule in each reaction that is being evaluated (and nucleic acid amplification is not required to prepare the target template for sequencing). In some embodiments, a plurality of single molecule sequencing reactions

are performed in parallel (e.g., on a single chip) according to aspects of the present application. For example, in some embodiments, a plurality of single molecule sequencing reactions are each performed in separate reaction chambers (e.g., nanoapertures, sample wells) on a single chip.

**[00160]** Embodiments are capable of sequencing single nucleic acid molecules with high accuracy and long read lengths, such as an accuracy of at least about 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, 99.99%, 99.999%, or 99.9999%, and/or read lengths greater than or equal to about 10 base pairs (bp), 50 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 1000 bp, 10,000 bp, 20,000 bp, 30,000 bp, 40,000 bp, 50,000 bp, or 100,000 bp. In some embodiments, the target nucleic acid molecule used in single molecule sequencing is a single stranded target nucleic acid (e.g., deoxyribonucleic acid (DNA), DNA derivatives, ribonucleic acid (RNA), RNA derivatives) template that is added or immobilized to a sample well (e.g., nanoaperture) containing at least one additional component of a sequencing reaction (e.g., a polymerase such as, a DNA polymerase, a sequencing primer) immobilized or attached to a solid support such as the bottom or side walls of the sample well. The target nucleic acid molecule or the polymerase can be attached to a sample wall, such as at the bottom or side walls of the sample well directly or through a linker. The sample well (e.g., nanoaperture) also can contain any other reagents needed for nucleic acid synthesis via a primer extension reaction, such as, for example suitable buffers, co-factors, enzymes (e.g., a polymerase) and deoxyribonucleoside polyphosphates, such as, e.g., deoxyribonucleoside triphosphates, including deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), deoxyuridine triphosphate (dUTP) and deoxythymidine triphosphate (dTTP) dNTPs, that include luminescent labels, such as luminescent labels of a labeled biomolecule provided herein.

**[00161]** In some embodiments, each class of dNTPs (e.g., adenine-containing dNTPs (e.g., dATP), cytosine-containing dNTPs (e.g., dCTP), guanine-containing dNTPs (e.g., dGTP), uracil-containing dNTPs (e.g., dUTPs) and thymine-containing dNTPs (e.g., dTTP)) is conjugated to a luminescent molecule that comprises distinct luminescent properties such that detection of light emitted from the luminescent molecule indicates the identity of the dNTP that was incorporated into the newly synthesized nucleic acid. Emitted light from the luminescent molecule (e.g., emitted light from a labeled biomolecule comprising at least one luminescent label) can be detected and attributed to its appropriate luminescent molecule (and, thus, associated dNTP) via any suitable device and/or method. The luminescent molecule may be conjugated to the dNTP at any position such that the presence of the

luminescent molecule (e.g., a labeled biomolecule of the application) does not inhibit the incorporation of the dNTP into the newly synthesized nucleic acid strand or the activity of the polymerase. In some embodiments, the luminescent molecule is conjugated to the terminal phosphate (e.g., the gamma phosphate) of the dNTP.

**[00162]** In some embodiments, the single-stranded target nucleic acid template can be contacted with a sequencing primer, dNTPs, polymerase and other reagents necessary for nucleic acid synthesis. In some embodiments, all appropriate dNTPs can be contacted with the single-stranded target nucleic acid template simultaneously (e.g., all dNTPs are simultaneously present) such that incorporation of dNTPs can occur continuously. In other embodiments, the dNTPs can be contacted with the single-stranded target nucleic acid template sequentially, where the single-stranded target nucleic acid template is contacted with each appropriate dNTP separately, with washing steps in between contact of the single-stranded target nucleic acid template with differing dNTPs. Such a cycle of contacting the single-stranded target nucleic acid template with each dNTP separately followed by washing can be repeated for each successive base position of the single-stranded target nucleic acid template to be identified.

**[00163]** In some embodiments, the sequencing primer anneals to the single-stranded target nucleic acid template and the polymerase consecutively incorporates the dNTPs (or other nucleoside polyphosphate) to the primer based on the single-stranded target nucleic acid template. The unique luminescent molecule, such as a labeled biomolecule described herein, associated with each incorporated dNTP can be excited with the appropriate excitation light during or after incorporation of the dNTP to the primer and its emission can be subsequently detected, using, any suitable device(s) and/or method(s). Detection of a particular emission of light (e.g., having a particular emission lifetime, intensity, spectrum and/or combination thereof) can be attributed to a particular dNTP incorporated. The sequence obtained from the collection of detected luminescent molecules can then be used to determine the sequence of the single-stranded target nucleic acid template via sequence complementarity.

**[00164]** In some embodiments, the present disclosure provides methods and compositions that may be advantageously utilized in the technologies described in co-pending U.S. Patent App. Nos.: 14/543,865, 14/543,867, 14/543,888, 14/821,656, 14/821,686, 14/821,688, 15/161,067, 15/161,088, 15/161,125, 15/255,245, 15/255,303, 15/255,624, 15/261,697, 15/261,724, 15/600,979, 15/846,967, 15/847,001, 62/289,019, 62/296,546, 62/310,398, 62/339,790, 62/343,997, 62/344,123, 62/426,144, and 62/505,525 the contents of each of which are incorporated herein by reference.

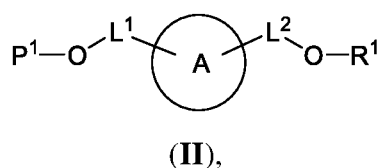
**Kits**

**[00165]** Also provided herein are kits for sequencing a template nucleic acid, the kit comprising two or more different types of labeled nucleotides, wherein at least one type of labeled nucleotide is a labeled nucleotide comprising a labeled biomolecule as described herein. In some embodiments, a kit comprises two or more (e.g., two, three, four, five, or more) different types of labeled nucleotides. In some embodiments, a kit comprises four different types of labeled nucleotides. In some embodiments, a kit comprises a polymerizing enzyme. In some embodiments, a kit comprises a primer complementary to the template nucleic acid.

**[00166]** In some embodiments, a kit comprises a plurality of types of labeled nucleotides comprising a labeled biomolecule as described herein. In some embodiments, at least one type (e.g., two, three, four, five, or more types) of labeled nucleotide comprises a labeled biomolecule having two or more internally-conjugated labels according to the application. In some embodiments, the plurality of nucleotides is selected from the labeled nucleotides depicted in FIGs. 1A-1G, 4A-4C, 5A, and 5C. In some embodiments, the kit further comprises a polymerizing enzyme (e.g., a DNA polymerase, as described elsewhere herein). In some embodiments, the kit further comprises a primer complementary to the template nucleic acid being sequenced.

**Biconjugatable Labels**

**[00167]** In another aspect, the present invention provides compounds. The compounds provided herein can be used as labels – that is, used in conjugation reactions to form the labeled biomolecules described herein. For example, provided herein are compounds of Formula (II):



and salts thereof, wherein:

A is a polycyclic fluorophore;

L<sup>1</sup> and L<sup>2</sup> are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally

substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof;

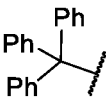
P<sup>1</sup> is an oxygen protecting group; and

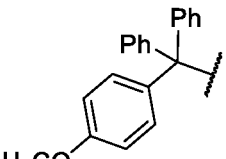
R<sup>1</sup> is a reactive moiety.

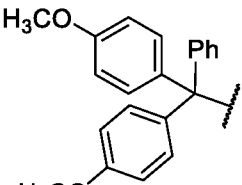
**[00168]** The compounds of Formula (II), and salts thereof, as described herein, are bifunctional (e.g., “asymmetrical”). R<sup>1</sup> is a reactive moiety that can be used as a reactive handle in a conjugation reaction. P<sup>1</sup> is an oxygen protecting group that can be cleaved or removed after the conjugation reaction involving R<sup>1</sup>, thereby revealing a free –OH group, which can then be used as a reactive moiety in a subsequent conjugation reaction. In certain embodiments, R<sup>1</sup> is a reactive moiety reactive in nucleoside coupling reactions (e.g., a phosphoramidite). Various oxygen protecting groups can be used at the position corresponding to P<sup>1</sup>, and examples are provided herein.

**[00169]** L<sup>1</sup> and L<sup>2</sup> are defined herein, and exemplarily embodiments are provided. Furthermore, exemplary embodiments of A are provided herein. All definitions and embodiments provided herein, including but not limited to those provided in the section INTERNALLY-LABELED BIOMOLECULES, are applicable to the compounds provided herein.

**[00170]** As defined herein, P<sup>1</sup> is an oxygen protecting group. Several examples of oxygen protecting groups are provided herein. In certain embodiments, P<sup>1</sup> is an optionally substituted triphenyl protecting group (e.g., trityl). In certain embodiments, P<sup>1</sup> is trityl, of the

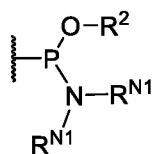
formula: . In certain embodiments, P<sup>1</sup> is 4-monomethoxytrityl (MMT), of the

formula: . In certain embodiments, P<sup>1</sup> is 4,4-dimethoxytrityl (DMT), of the

formula: .

**[00171]** In certain embodiments, R<sup>1</sup> is a reactive moiety. For example, R<sup>1</sup> is a reactive handle useful in polynucleotide synthesis, polypeptide synthesis, polysaccharide synthesis,

etc. A reactive moiety can be any group capable of reacting with the second reactive moiety (e.g., an  $-OH$  or  $-NH_2$  group) to form a covalent bond. A person of skill in the art would know what reactive moieties can be used to form the bonds in polynucleotide synthesis, polypeptide synthesis, polysaccharide synthesis, etc. In certain embodiments,  $R^1$  is a moiety useful in polynucleotide synthesis (e.g., a phosphoramidite). In certain embodiments,  $R^1$  is a phosphoramidite. In certain embodiments,  $R^1$  is a phosphoramidite of the formula:



, wherein  $R^{N1}$  and  $R^2$  are as defined herein. In certain embodiments,  $R^1$  is a

phosphoramidite of the formula: . In certain embodiments,  $R^1$  is a

phosphoramidite of the formula: .

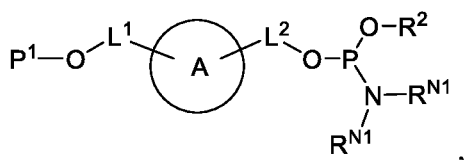
**[00172]** As defined herein,  $R^2$  is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or an oxygen protecting group. In certain embodiments,  $R^2$  is optionally substituted alkyl. In certain embodiments,  $R^2$  is optionally substituted  $C_{1-6}$  alkyl. In certain embodiments,  $R^2$  is optionally substituted  $C_{1-3}$  alkyl. In

certain embodiments,  $R^2$  is of the formula: .

**[00173]** As defined herein, each instance of  $R^{N1}$  is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group; optionally wherein two  $R^{N1}$  bonded to the same nitrogen are joined together with the intervening atoms to form optionally substituted heterocyclyl or optionally substituted heteroaryl. In certain embodiments,  $R^{N1}$  is optionally substituted alkyl. In certain embodiments,  $R^{N1}$  is optionally substituted  $C_{1-6}$  alkyl. In certain embodiments,  $R^{N1}$  is

unsubstituted C<sub>1-6</sub> alkyl. In certain embodiments, R<sup>N1</sup> is optionally substituted C<sub>1-3</sub> alkyl. In certain embodiments, R<sup>N1</sup> is unsubstituted C<sub>1-3</sub> alkyl. In certain embodiments, R<sup>N1</sup> is selected from the group consisting of methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl, sec-butyl, and tert-butyl. In certain embodiments, R<sup>N1</sup> is iso-propyl. In certain embodiments, both instances of R<sup>N1</sup> are iso-propyl.

**[00174]** In certain embodiments, R<sup>1</sup> is a moiety reactive in nucleoside coupling reactions, such as a phosphoramidite. In certain embodiments, the compound of Formula (II) is of the formula:

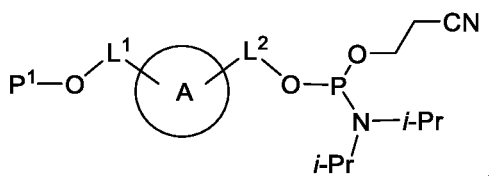


or a salt thereof, wherein:

R<sup>2</sup> is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or an oxygen protecting group;

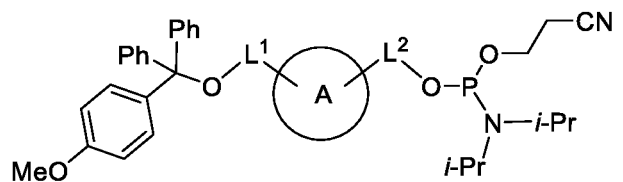
each instance of R<sup>N1</sup> is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group; optionally wherein two R<sup>N1</sup> bonded to the same nitrogen are joined together with the intervening atoms to form optionally substituted heterocyclyl or optionally substituted heteroaryl.

**[00175]** In certain embodiments, the compound of Formula (II) is of the formula:



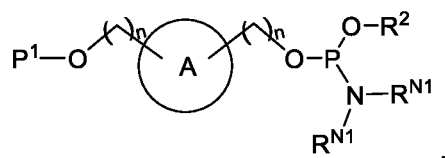
or a salt thereof.

**[00176]** In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof.

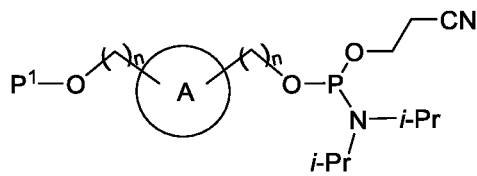
[00177] In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof, wherein:

each instance of n is an integer from 1-20, inclusive.

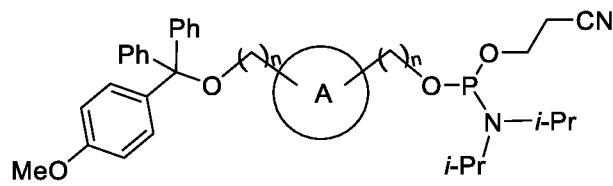
[00178] In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof, wherein:

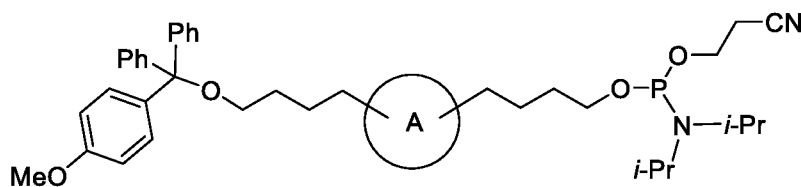
each instance of n is an integer from 1-20, inclusive.

[00179] In certain embodiments, the compound of Formula (II) is of the formula:



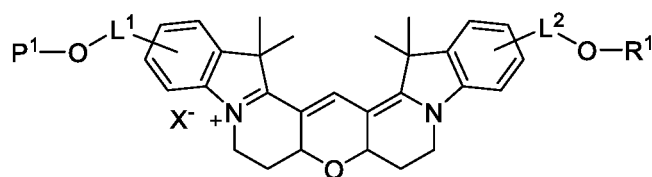
or a salt thereof.

[00180] In certain embodiments, the compound of Formula (II) is of the formula:



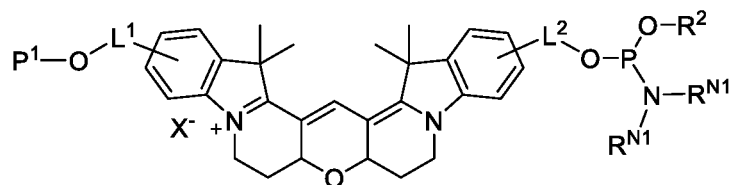
or a salt thereof.

[00181] As described herein, Ring A can be a polycyclic cyanine, such as Cy3B. In certain embodiments, the compound of Formula (II) is of the formula:



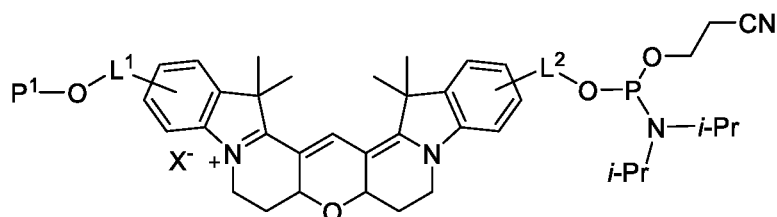
or a salt thereof, wherein X<sup>-</sup> is a counterion or is absent.

[00182] In certain embodiments, the compound of Formula (II) is of the formula:



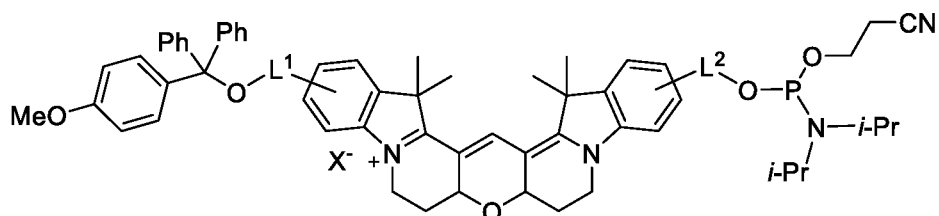
or a salt thereof.

**[00183]** In certain embodiments, the compound of Formula (II) is of the formula:



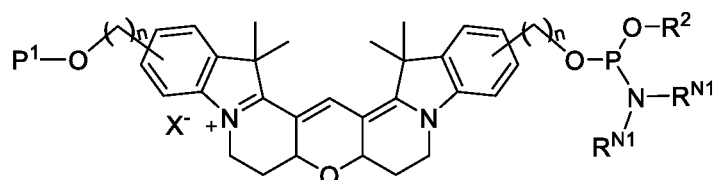
or a salt thereof.

**[00184]** In certain embodiments, the compound of Formula (II) is of the formula:



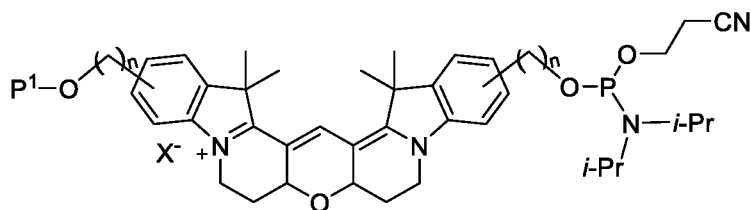
or a salt thereof.

**[00185]** In certain embodiments, the compound of Formula (II) is of the formula:



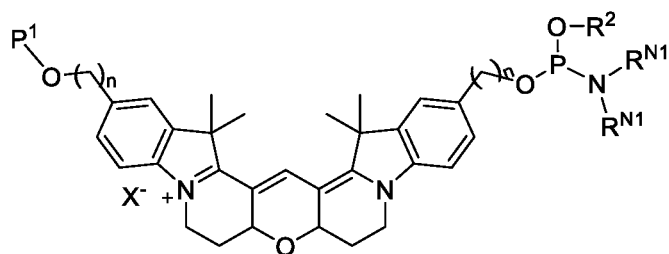
or a salt thereof.

**[00186]** In certain embodiments, the compound of Formula (II) is of the formula:



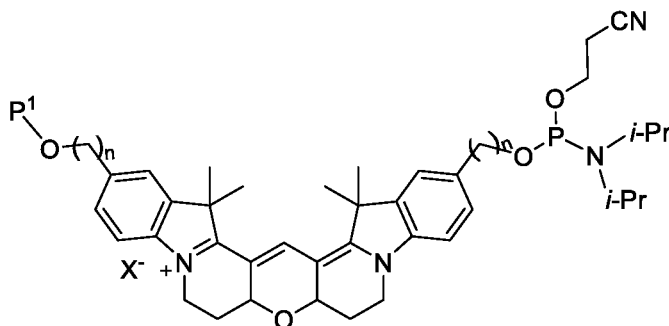
or a salt thereof.

[00187] In certain embodiments, the compound of Formula (II) is of the formula:



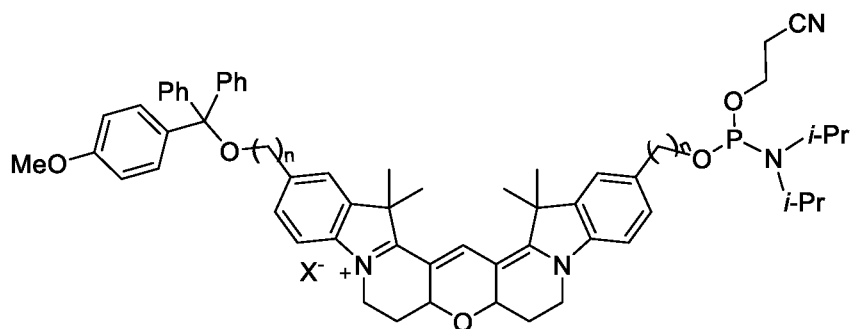
or a salt thereof.

[00188] In certain embodiments, the compound of Formula (II) is of the formula:



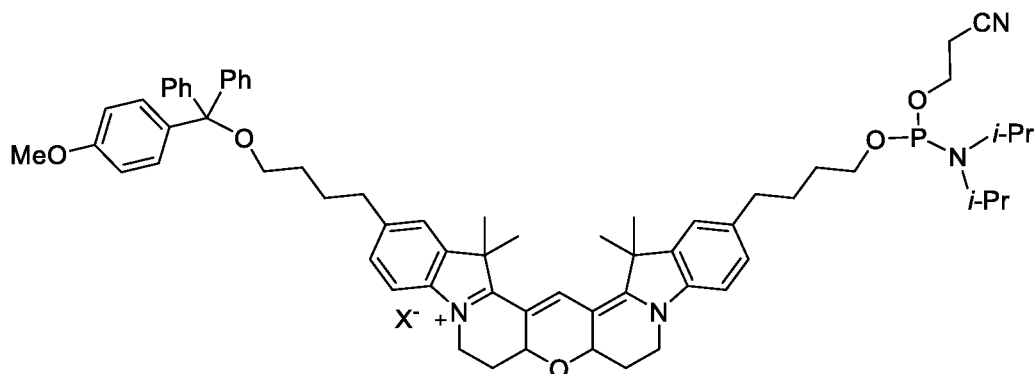
or a salt thereof.

[00189] In certain embodiments, the compound of Formula (II) is of the formula:



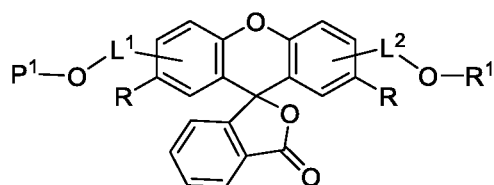
or a salt thereof.

[00190] In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof.

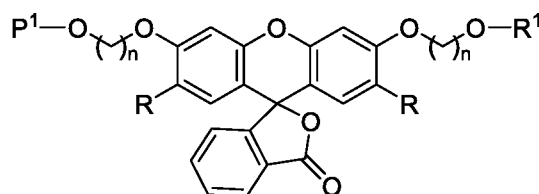
**[00191]** In certain embodiments, A can be a fluorone fluorophore, such as fluorescein or rhodamine, for example. In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof, wherein:

each R is independently hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ .

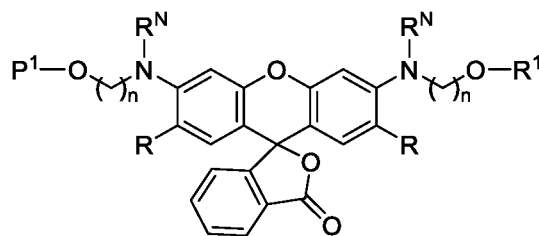
**[00192]** In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof, wherein:

n is independently an integer from 1-20, inclusive.

**[00193]** In certain embodiments, the compound of Formula (II) is of the formula:

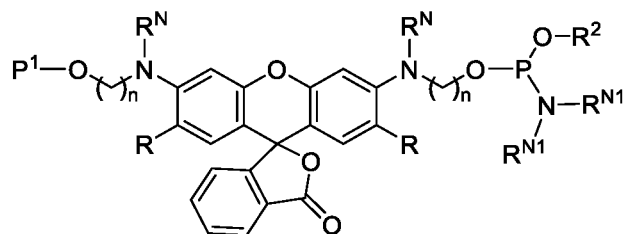


or a salt thereof, wherein:

n is independently an integer from 1-20, inclusive; and

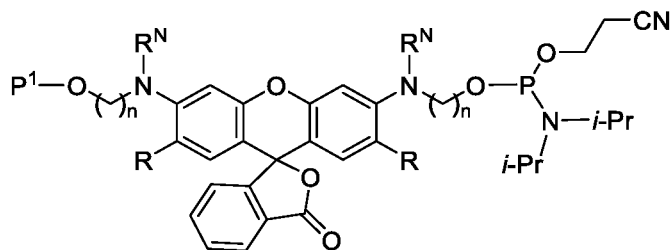
each instance of  $R^N$  is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group.

[00194] In certain embodiments, the compound of Formula (II) is of the formula:



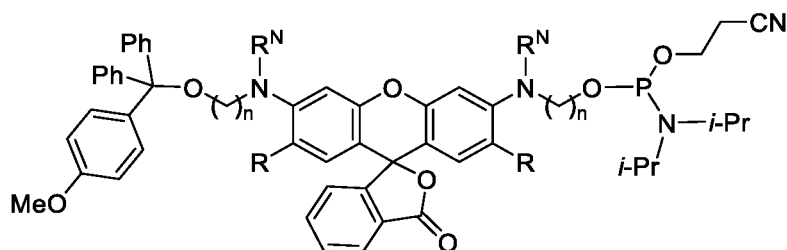
or a salt thereof.

[00195] In certain embodiments, the compound of Formula (II) is of the formula:



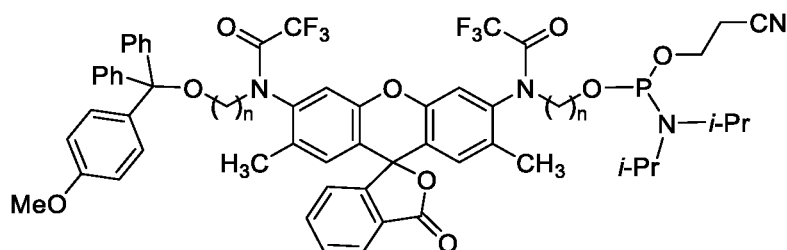
or a salt thereof.

[00196] In certain embodiments, the compound of Formula (II) is of the formula:



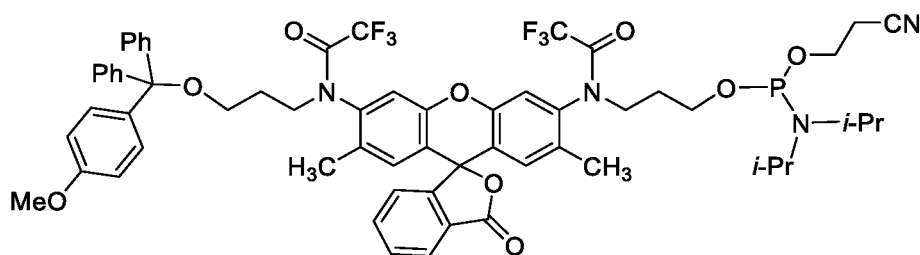
or a salt thereof.

[00197] In certain embodiments, the compound of Formula (II) is of the formula:



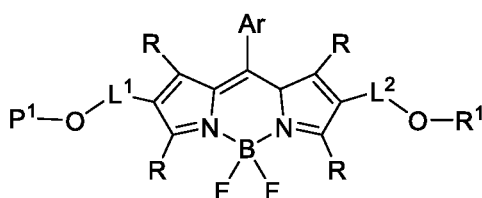
or a salt thereof.

[00198] In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof.

[00199] As described herein, in certain embodiments, A is a BODIPY fluorophore. Therefore, in certain embodiments, the compound of Formula (II) is of the following formula:

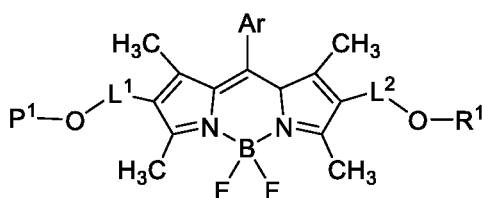


or a salt thereof, wherein:

each R is independently hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ ; and

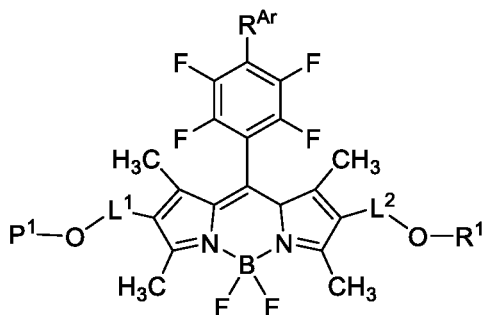
Ar is optionally substituted aryl or optionally substituted heteroaryl.

[00200] In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof.

[00201] In certain embodiments, the compound of Formula (II) is of the formula:



or a salt thereof, wherein:

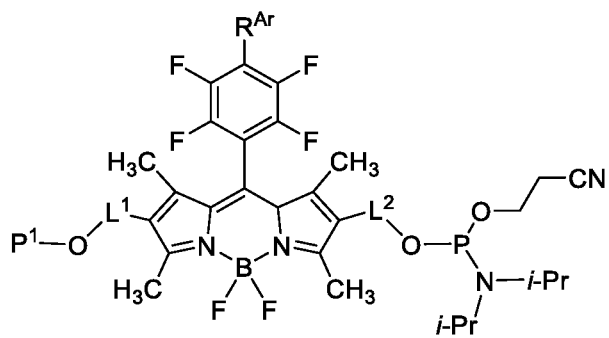
$R^{Ar}$  is hydrogen, halogen,  $-N_3$ ,  $-CN$ ,  $-NO_2$ , optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted aryl, optionally substituted heterocyclyl, optionally substituted heteroaryl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ .

**[00202]** As generally defined herein,  $R^O$  is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or an oxygen protecting group. In certain embodiments,  $R^O$  is hydrogen. In certain embodiments,  $R^O$  is optionally substituted alkyl. In certain embodiments,  $R^O$  is optionally substituted  $C_{1-6}$  alkyl. In certain embodiments,  $R^O$  is unsubstituted  $C_{1-6}$  alkyl. In certain embodiments,  $R^O$  is selected from methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl, sec-butyl, and tert-butyl. In certain embodiments,  $R^O$  is optionally substituted alkenyl. In certain embodiments,  $R^O$  is optionally substituted alkynyl. In certain embodiments,  $R^O$  is optionally substituted carbocyclyl. In certain embodiments,  $R^O$  is optionally substituted heterocyclyl. In certain embodiments,  $R^O$  is optionally substituted aryl. In certain embodiments,  $R^O$  is optionally substituted heteroaryl. In certain embodiments,  $R^O$  is optionally substituted acyl. In certain embodiments,  $R^O$  is an oxygen protecting group. In certain embodiments,  $R^O$  is  $-C(=O)Ph$ .

**[00203]** As generally defined herein, each instance of  $R^N$  is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group; or optionally two  $R^N$  are joined together with the intervening atoms to form optionally substituted heterocyclyl or optionally substituted heteroaryl. In certain embodiments,  $R^N$  is hydrogen. In certain embodiments,  $R^N$  is optionally substituted alkyl. In certain embodiments,  $R^N$  is optionally substituted  $C_{1-6}$  alkyl. In certain embodiments,  $R^N$  is unsubstituted  $C_{1-6}$  alkyl. In certain embodiments,  $R^N$  is selected from methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl, sec-butyl, and tert-butyl. In certain embodiments,  $R^N$  is optionally substituted alkenyl. In certain embodiments,  $R^N$  is optionally substituted alkynyl. In certain embodiments,  $R^N$  is optionally substituted carbocyclyl. In certain embodiments,  $R^N$  is optionally substituted heterocyclyl. In certain embodiments,  $R^N$  is optionally substituted aryl. In certain embodiments,  $R^N$  is optionally substituted heteroaryl. In certain embodiments,  $R^N$  is optionally substituted acyl. In certain embodiments,  $R^N$  is a nitrogen protecting group. In

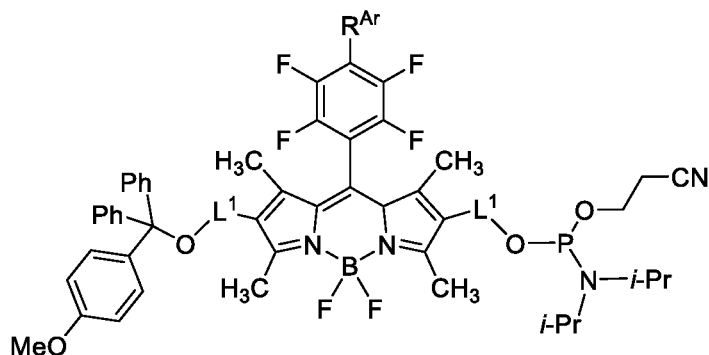


[00206] In certain embodiments, the compound of Formula (II) is of the formula:



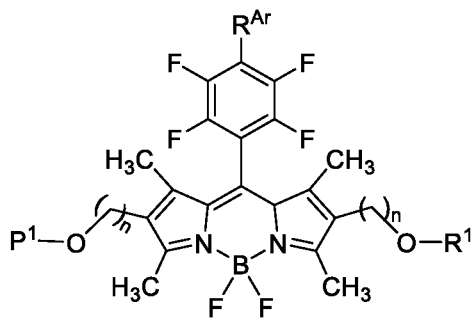
or a salt thereof.

[00207] In certain embodiments, the compound of Formula (II) is of the formula:



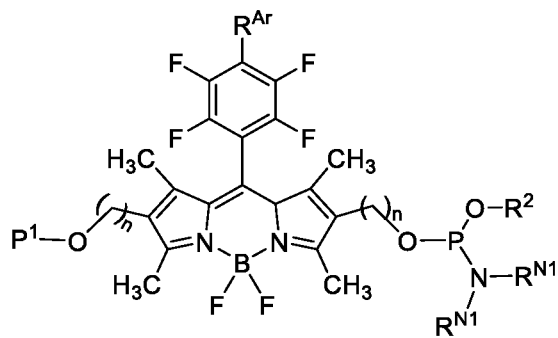
or a salt thereof.

[00208] In certain embodiments, the compound of Formula (II) is of the formula:



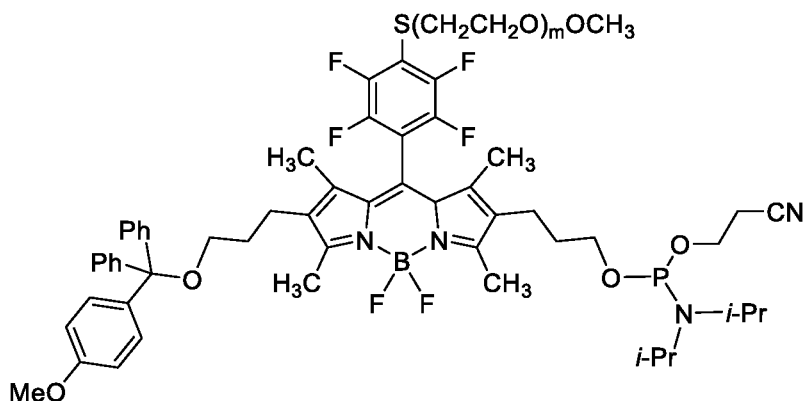
or a salt thereof, wherein n is independently an integer from 1-20, inclusive.

[00209] In certain embodiments, the compound of Formula (II) is of the formula:





[00213] In certain embodiments, the compound of Formula (II) is of the formula:



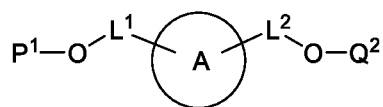
or a salt thereof.

### Methods for Preparing Internally-Labeled Biomolecules

[00214] Also provided herein are methods for preparing the internally-labeled biomolecules (e.g., biomolecules of Formula (I)) described herein. In general, the methods comprise two subsequent conjugation steps involving the bifunctional compounds (e.g., compounds of Formula (II)) provided herein.

[00215] Thus, provided herein are methods for preparing labeled biomolecules, the methods comprising:

(i) contacting a monomeric or oligomeric biomolecule of formula Q<sup>2</sup>-OH, or a salt thereof, with a compound of Formula (II), or a salt thereof, under conditions sufficient to promote conjugation to yield a conjugate of the formula:



or a salt thereof;

(ii) deprotecting the conjugate formed in step (i) under conditions sufficient to cleave the P<sup>1</sup> protecting group and yield a conjugate of the formula:



or a salt thereof;

(iii) contacting the conjugate formed in step (ii) with a monomeric or oligomeric biomolecule of formula Q<sup>1</sup>-O-R<sup>1</sup>, or a salt thereof, under conditions sufficient to promote conjugation to yield a labeled biomolecule of Formula (I):



to 21 °C). In certain embodiments, a reaction is carried out at or around -78 °C. In certain embodiments, a reaction is carried out at or around -10 °C. In certain embodiments, a reaction is carried out at around 0 °C. In certain embodiments, a reaction is carried out at above room temperature. In certain embodiment, a reaction is carried out at 30, 40, 50, 60, 70, 80, 110, 120, 130, 140, or 150 °C. In certain embodiments, a reaction is carried out at above 150 °C.

**[00221]** A reaction described herein may be carried out in a solvent, or a mixture of solvents (e.g., cosolvents). Solvents can be polar or non-polar, protic or aprotic. Any solvent may be used in the reactions described herein, and the reactions are not limited to particular solvents or combinations of solvents. Common organic solvents useful in the methods described herein include, but are not limited to, acetone, acetonitrile, benzene, benzonitrile, 1-butanol, 2-butanone, butyl acetate, tert-butyl methyl ether, carbon disulfide carbon tetrachloride, chlorobenzene, 1-chlorobutane, chloroform, cyclohexane, cyclopentane, 1,2-dichlorobenzene, 1,2-dichloroethane, dichloromethane (DCM), N,N-dimethylacetamide N,N-dimethylformamide (DMF), 1,3-dimethyl-3,4,5,6-tetrahydro-2-pyrimidinone (DMPU), 1,4-dioxane, 1,3-dioxane, diethylether, 2-ethoxyethyl ether, ethyl acetate, ethyl alcohol, ethylene glycol, dimethyl ether, heptane, n-hexane, hexanes, hexamethylphosphoramide (HMPA), 2-methoxyethanol, 2-methoxyethyl acetate, methyl alcohol, 2-methylbutane, 4-methyl-2-pentanone, 2-methyl-1-propanol, 2-methyl-2-propanol, 1-methyl-2-pyrrolidinone, dimethylsulfoxide (DMSO), nitromethane, 1-octanol, pentane, 3-pentanone, 1-propanol, 2-propanol, pyridine, tetrachloroethylene, tetrahyrdofuran (THF), 2-methyltetrahydrofuran, toluene, trichlorobenzene, 1,1,2-trichlorotrifluoroethane, 2,2,4-trimethylpentane, trimethylamine, triethylamine, N,N-diisopropylethylamine, diisopropylamine, water, o-xylene, p-xylene.

**[00222]** A reaction described herein may be carried out over any amount of time. In certain embodiments, a reaction is allowed to run for seconds, minutes, hours, or days.

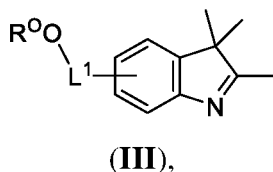
**[00223]** Methods described herein can be used to prepare compounds in any chemical yield. In certain embodiments, a compound is produced in from 1-10%, 10-20% 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90%, or 90-100% yield. In certain embodiments, the yield is the percent yield after one synthetic step. In certain embodiments, the yield is the percent yield after more than one synthetic step (e.g., 2, 3, 4, or 5 synthetic steps).

**[00224]** Methods described herein may further comprise one or more purification steps. For example, in certain embodiments, a compound produced by a method described herein may be purified by chromatography, extraction, filtration, precipitation, crystallization,

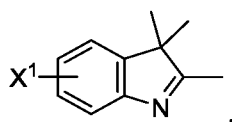
or any other method known in the art. In certain embodiments, a compound or mixture is carried forward to the next synthetic step without purification (e.g., crude).

#### Methods for Preparing Cy3B-Based Labels

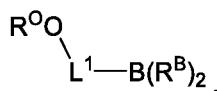
**[00225]** Provided herein is a method of preparing a compound of Formula **(III)**:



or a salt thereof, the method comprising coupling a compound of the formula:



or a salt thereof, with a compound of the formula:



or a salt thereof, in the presence of palladium to yield a compound of Formula **(III)**, or a salt thereof, wherein:

$X^1$  is halogen or a leaving group;

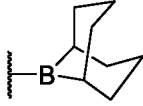
$B(R^B)_2$  is a borane, boronic acid, or boronic ester;

$L^1$  is a linker selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof; and

$R^O$  is an oxygen protecting group.

**[00226]** In certain embodiments, the reaction to yield a compound of Formula **(III)** is carried out in the presence of a metal other than palladium. For example, the reaction can be palladium-catalyzed, or catalyzed by a different metal. In certain embodiments, the metal is a transition metal.

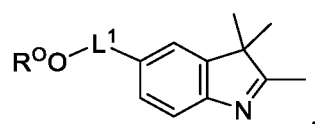
**[00227]** As defined herein, the group  $-B(R^B)_2$  is a borane, a boronic acid, or a boronic ester. In certain embodiments,  $-B(R^B)_2$  is a borane. In certain embodiments,  $-B(R^B)_2$  is a boronic acid. In certain embodiments,  $-B(R^B)_2$  is a boronic ester. In certain embodiments, —

$B(R^B)_2$  is a borane of the formula: . As defined herein, each instance of  $R^B$  is independently optionally substituted alkyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, -OH, or -OR<sup>O</sup>. Optionally, two  $R^B$  are joined together with the intervening atoms to form optionally substituted carbocyclyl or optionally substituted heterocyclyl. In general,  $-B(R^B)_2$  is any suitable borane, boronic acid, or boronic ester useful in a metal-promoted or metal-catalyzed cross-coupling reaction.

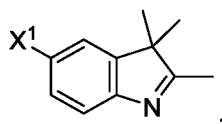
**[00228]** The linker  $L^1$  is as defined herein.

**[00229]** As defined herein,  $X^1$  is a halogen or a leaving group. In certain embodiments,  $X^1$  is a halogen. In certain embodiments,  $X^1$  is a leaving group. In certain embodiments,  $X^1$  is -Cl, -Br, or -I. In certain embodiments,  $X^1$  is -I

**[00230]** In certain embodiments of the coupling reaction, the compound of Formula (III) is of the formula:

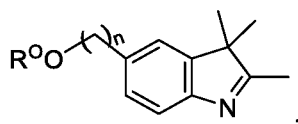


or a salt thereof; and the starting material is therefore of the formula:

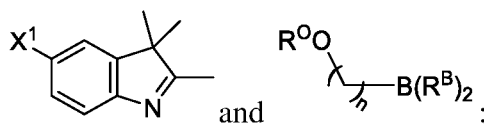


or a salt thereof.

**[00231]** In certain embodiments, the compound of Formula (III) is of the formula:



or a salt thereof; wherein  $n$  is an integer from 1-20, inclusive; and therefore the starting materials are of the formulae:



or salts thereof.

**[00232]** As described above, the coupling reaction is carried out in the presence of palladium. In certain embodiments, the palladium is a palladium complex. In certain embodiments, the palladium complex is a palladium(II) complex. In certain embodiments, the

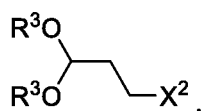
palladium complex is PdCl<sub>2</sub>(dppf). In certain embodiments, the palladium is present in a catalytic amount. In other embodiments, the palladium is present in a stoichiometric or excess amount.

**[00233]** In certain embodiments, the coupling reaction is carried out in the presence of a base. In certain embodiments, the base is a carbonate base. In certain embodiments, the base is Cs<sub>2</sub>CO<sub>3</sub>.

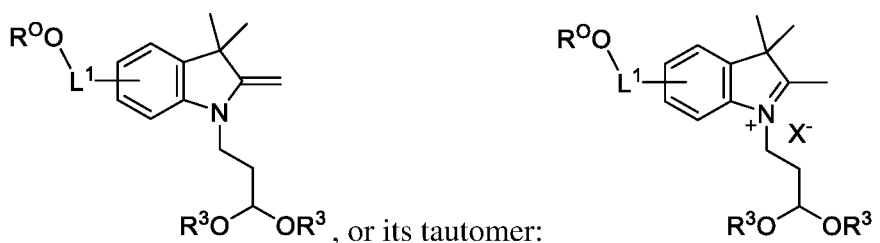
**[00234]** In certain embodiments, the coupling reaction is carried out in a solvent. In certain embodiments, the solvent is THF, DMF, or a mixture thereof.

**[00235]** The coupling reaction can be carried out at any temperature. In certain embodiments, the reaction is carried out at room temperature. In certain embodiments, the reaction is carried out at above room temperature (i.e., elevated temperature). In certain embodiments, the reaction is carried out at between room temperature and 100 °C. In certain embodiments, the reaction is carried out at between 50° and 100° C. In certain embodiments, the reaction is carried out at around 70° C.

**[00236]** In certain embodiments, the method further comprises a step of alkylating a compound of Formula (III) with a compound of the formula:



or a salt thereof, to yield a compound of the formula:

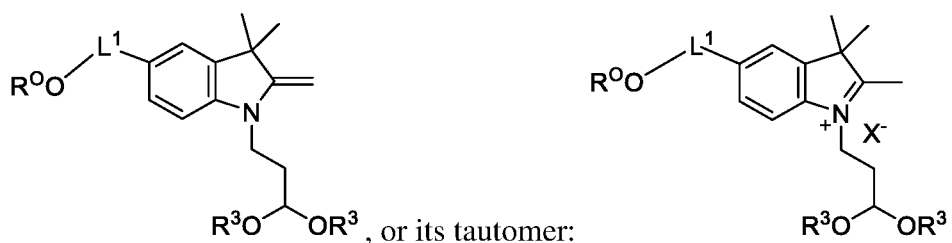


or a salt thereof, wherein:

X<sup>2</sup> is halogen or a leaving group; and

each R<sup>3</sup> is independently optionally substituted alkyl, optionally substituted acyl, or an oxygen protecting group; or optionally two R<sup>3</sup> are joined together with the intervening atoms to form optionally substituted heterocyclyl.

**[00237]** In certain embodiments, the product is a compound of the formula:

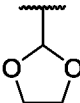


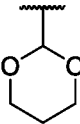
or a salt thereof.

**[00238]** In certain embodiments, the step of alkylating is carried out in the presence of a base. In certain embodiments, the base is a halide salt. In certain embodiments, the base is an iodide salt. In certain embodiments, the base is KI. In certain embodiments, the step of alkylating is carried out in a solvent. In certain embodiments, the solvent is acetonitrile (MeCN). In certain embodiments, the reaction is carried out at room temperature. In certain embodiments, the reaction is carried out at elevated temperature. In certain embodiments, the reaction is carried out at around 100 °C.

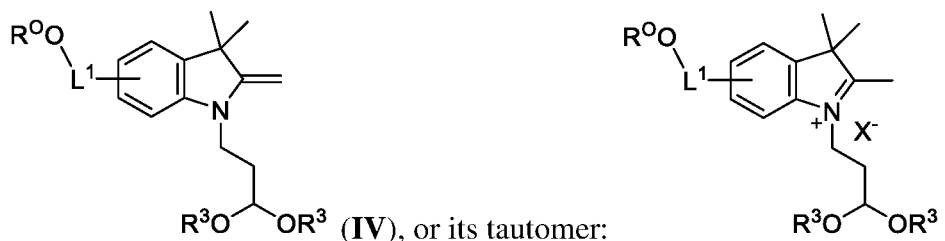
**[00239]** As defined herein,  $X^2$  is halogen or a leaving group. In certain embodiments,  $X^2$  is a halogen. In certain embodiments,  $X^2$  is  $-Cl$ ,  $-Br$ , or  $-I$ . In certain embodiments,  $X^2$  is  $-Br$ . In certain embodiments,  $X^2$  is a leaving group.

**[00240]** As defined herein, each  $R^3$  is independently optionally substituted alkyl, optionally substituted acyl, or an oxygen protecting group; or optionally two  $R^3$  are joined together with the intervening atoms to form optionally substituted heterocyclyl. In certain embodiments,  $R^3$  is optionally substituted alkyl. In certain embodiments,  $R^3$  is optionally substituted acyl. In certain embodiments,  $R^3$  is an oxygen protecting group. In certain embodiments, two  $R^3$  are joined together with the intervening atoms to form optionally substituted heterocyclyl. In certain embodiments, two  $R^3$  are joined together with the

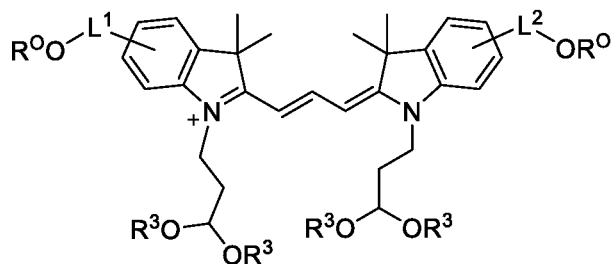
intervening atoms to form: . In certain embodiments, two  $R^3$  are joined together with

the intervening atoms to form: .

**[00241]** In certain embodiments, the method further comprises reacting the compound of formula:



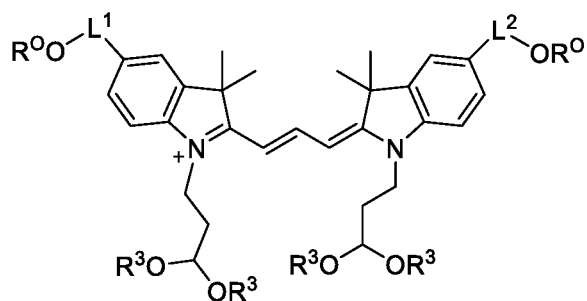
or a salt thereof, in the presence of a formamidine, to yield a compound of formula:



or a salt thereof. In certain embodiments, the reaction entails (i) reacting the compound of Formula (IV), or a salt thereof, in the presence of a formamidine to form an intermediate; and (ii) reacting the intermediate formed in step (i) with another compound of Formula (IV), or a salt thereof, to yield the product shown above.

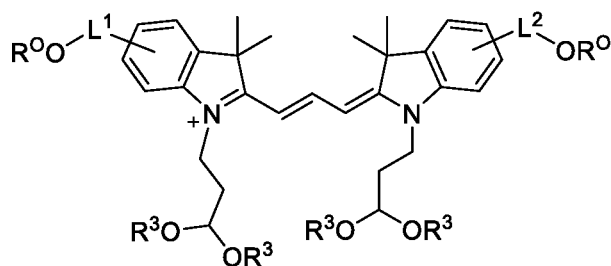
**[00242]** In certain embodiments, the formamidine in step (i) is diphenylformamidine. In certain embodiments, the reaction in step (i) is carried out in the presence of a base. In certain embodiments, the base is a pyridine base. In certain embodiments, the base is DMAP. In certain embodiments, the reaction in step (i) is carried out in the presence of an anhydride. In certain embodiments, the anhydride is acetic anhydride ( $\text{Ac}_2\text{O}$ ). In certain embodiments, the reaction in step (i) is carried out at elevated temperature (e.g., around  $125^\circ\text{C}$ ). In certain embodiments, the reaction is carried out in a solvent. In certain embodiments, the reaction in step (ii) is carried out in the presence of a base. In certain embodiments, the base is an amine base (e.g., a trialkylamine base). In certain embodiments, the base is  $\text{Et}_3\text{N}$ . wherein the reaction is carried out in the presence of a base. In certain embodiments, the reaction is carried out in a solvent. In certain embodiments, the solvent is EtOH. In certain embodiments, the reaction is carried out at elevated temperature (e.g., around  $80^\circ\text{C}$ ).

**[00243]** In certain embodiments, the product is of the formula:

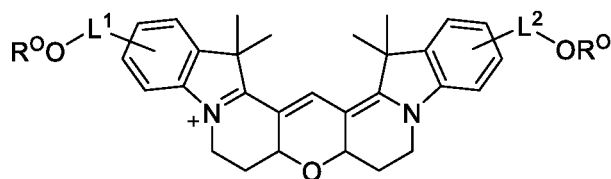


or a salt thereof.

**[00244]** In certain embodiments, the method further comprises cyclizing a compound of formula:

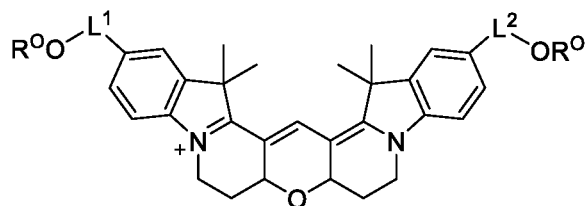


or a salt thereof, in the presence of an acid, to yield a compound of the formula:



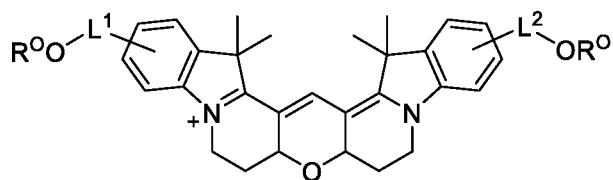
or a salt thereof. In certain embodiments, the acid is a sulfonic acid. In certain embodiments, the acid is sulfuric acid. In certain embodiments, the reaction is carried out in a solvent. In certain embodiments, the solvent is  $\text{CH}_2\text{Cl}_2$ . In certain embodiments, the reaction is carried out at elevated temperature (e.g., around  $60\text{ }^\circ\text{C}$ ).

**[00245]** In certain embodiments, the product is of the formula:

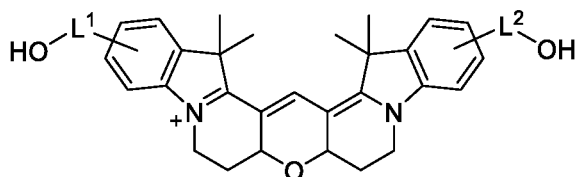


or a salt thereof.

**[00246]** In certain embodiments, the method further comprises a step of deprotecting a compound of the formula:

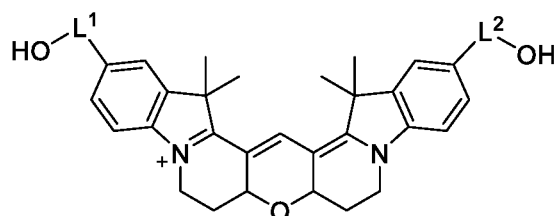


or a salt thereof, to yield a compound of the formula:



or a salt thereof.

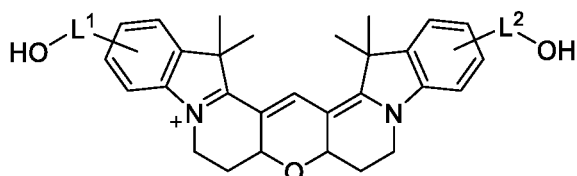
**[00247]** In certain embodiments, the compound is of the formula:



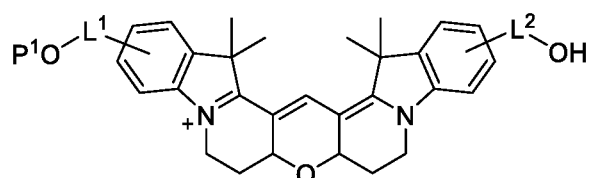
or a salt thereof.

**[00248]** In certain embodiments, the method further comprises the steps of:

- (i) protecting a compound of the formula:

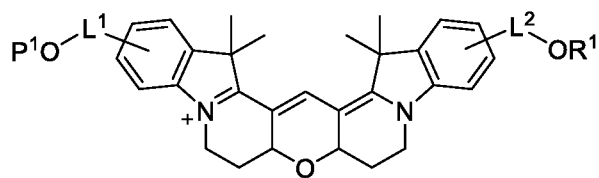


or a salt thereof, to yield a compound of the formula:



or a salt thereof; and

(ii) reacting the compound produced in step (i) under conditions sufficient for a compound of the formula:



or a salt thereof.

## DEFINITIONS

### Chemical Definitions

**[00249]** Definitions of specific functional groups and chemical terms are described in more detail below. The chemical elements are identified in accordance with the Periodic Table of the Elements, CAS version, Handbook of Chemistry and Physics, 75<sup>th</sup> Ed., inside cover, and specific functional groups are generally defined as described therein. Additionally, general principles of organic chemistry, as well as specific functional moieties and reactivity, are described in Organic Chemistry, Thomas Sorrell, University Science Books, Sausalito, 1999; Smith and March, *March's Advanced Organic Chemistry*, 5<sup>th</sup> Edition, John Wiley & Sons, Inc., New York, 2001; Larock, *Comprehensive Organic Transformations*, VCH Publishers, Inc., New York, 1989; and Carruthers, *Some Modern Methods of Organic Synthesis*, 3<sup>rd</sup> Edition, Cambridge University Press, Cambridge, 1987.

**[00250]** Compounds described herein can comprise one or more asymmetric centers, and thus can exist in various stereoisomeric forms, e.g., enantiomers and/or diastereomers. For example, the compounds described herein can be in the form of an individual enantiomer, diastereomer or geometric isomer, or can be in the form of a mixture of stereoisomers, including racemic mixtures and mixtures enriched in one or more stereoisomer. Isomers can be isolated from mixtures by methods known to those skilled in the art, including chiral high pressure liquid chromatography (HPLC) and the formation and crystallization of chiral salts; or preferred isomers can be prepared by asymmetric syntheses. See, for example, Jacques et al., *Enantiomers, Racemates and Resolutions* (Wiley Interscience, New York, 1981); Wilen et al., *Tetrahedron* 33:2725 (1977); Eliel, E.L. *Stereochemistry of Carbon Compounds* (McGraw-Hill, NY, 1962); and Wilen, S.H., *Tables of Resolving Agents and Optical Resolutions* p. 268 (E.L. Eliel, Ed., Univ. of Notre Dame Press, Notre Dame, IN 1972). The invention additionally encompasses compounds as individual isomers substantially free of other isomers, and alternatively, as mixtures of various isomers.

**[00251]** Unless otherwise stated, structures depicted herein are also meant to include compounds that differ only in the presence of one or more isotopically enriched atoms. For example, compounds having the present structures except for the replacement of hydrogen by deuterium or tritium, replacement of <sup>19</sup>F with <sup>18</sup>F, or the replacement of <sup>12</sup>C with <sup>13</sup>C or <sup>14</sup>C

are within the scope of the disclosure. Such compounds are useful, for example, as analytical tools or probes in biological assays.

**[00252]** When a range of values is listed, it is intended to encompass each value and sub-range within the range. For example “C<sub>1-6</sub> alkyl” is intended to encompass, C<sub>1</sub>, C<sub>2</sub>, C<sub>3</sub>, C<sub>4</sub>, C<sub>5</sub>, C<sub>6</sub>, C<sub>1-6</sub>, C<sub>1-5</sub>, C<sub>1-4</sub>, C<sub>1-3</sub>, C<sub>1-2</sub>, C<sub>2-6</sub>, C<sub>2-5</sub>, C<sub>2-4</sub>, C<sub>2-3</sub>, C<sub>3-6</sub>, C<sub>3-5</sub>, C<sub>3-4</sub>, C<sub>4-6</sub>, C<sub>4-5</sub>, and C<sub>5-6</sub> alkyl.

**[00253]** The term “aliphatic” refers to alkyl, alkenyl, alkynyl, and carbocyclic groups. Likewise, the term “heteroaliphatic” refers to heteroalkyl, heteroalkenyl, heteroalkynyl, and heterocyclic groups.

**[00254]** The term “alkyl” refers to a radical of a straight-chain or branched saturated hydrocarbon group having from 1 to 10 carbon atoms (“C<sub>1-10</sub> alkyl”). In some embodiments, an alkyl group has 1 to 9 carbon atoms (“C<sub>1-9</sub> alkyl”). In some embodiments, an alkyl group has 1 to 8 carbon atoms (“C<sub>1-8</sub> alkyl”). In some embodiments, an alkyl group has 1 to 7 carbon atoms (“C<sub>1-7</sub> alkyl”). In some embodiments, an alkyl group has 1 to 6 carbon atoms (“C<sub>1-6</sub> alkyl”). In some embodiments, an alkyl group has 1 to 5 carbon atoms (“C<sub>1-5</sub> alkyl”). In some embodiments, an alkyl group has 1 to 4 carbon atoms (“C<sub>1-4</sub> alkyl”). In some embodiments, an alkyl group has 1 to 3 carbon atoms (“C<sub>1-3</sub> alkyl”). In some embodiments, an alkyl group has 1 to 2 carbon atoms (“C<sub>1-2</sub> alkyl”). In some embodiments, an alkyl group has 1 carbon atom (“C<sub>1</sub> alkyl”). In some embodiments, an alkyl group has 2 to 6 carbon atoms (“C<sub>2-6</sub> alkyl”). Examples of C<sub>1-6</sub> alkyl groups include methyl (C<sub>1</sub>), ethyl (C<sub>2</sub>), propyl (C<sub>3</sub>) (e.g., n-propyl, isopropyl), butyl (C<sub>4</sub>) (e.g., n-butyl, tert-butyl, sec-butyl, iso-butyl), pentyl (C<sub>5</sub>) (e.g., n-pentyl, 3-pentanyl, amyl, neopentyl, 3-methyl-2-butanyl, tertiary amyl), and hexyl (C<sub>6</sub>) (e.g., n-hexyl). Additional examples of alkyl groups include n-heptyl (C<sub>7</sub>), n-octyl (C<sub>8</sub>), and the like. Unless otherwise specified, each instance of an alkyl group is independently unsubstituted (an “unsubstituted alkyl”) or substituted (a “substituted alkyl”) with one or more substituents (e.g., halogen, such as F). In certain embodiments, the alkyl group is an unsubstituted C<sub>1-10</sub> alkyl (such as unsubstituted C<sub>1-6</sub> alkyl, e.g., -CH<sub>3</sub> (Me), unsubstituted ethyl (Et), unsubstituted propyl (Pr, e.g., unsubstituted n-propyl (n-Pr), unsubstituted isopropyl (i-Pr)), unsubstituted butyl (Bu, e.g., unsubstituted n-butyl (n-Bu), unsubstituted tert-butyl (tert-Bu or t-Bu), unsubstituted sec-butyl (sec-Bu), unsubstituted isobutyl (i-Bu)). In certain embodiments, the alkyl group is a substituted C<sub>1-10</sub> alkyl (such as substituted C<sub>1-6</sub> alkyl, e.g., -CF<sub>3</sub>, Bn).

**[00255]** The term “haloalkyl” is a substituted alkyl group, wherein one or more of the hydrogen atoms are independently replaced by a halogen, e.g., fluoro, bromo, chloro, or iodo.

In some embodiments, the haloalkyl moiety has 1 to 8 carbon atoms (“C<sub>1-8</sub> haloalkyl”). In some embodiments, the haloalkyl moiety has 1 to 6 carbon atoms (“C<sub>1-6</sub> haloalkyl”). In some embodiments, the haloalkyl moiety has 1 to 4 carbon atoms (“C<sub>1-4</sub> haloalkyl”). In some embodiments, the haloalkyl moiety has 1 to 3 carbon atoms (“C<sub>1-3</sub> haloalkyl”). In some embodiments, the haloalkyl moiety has 1 to 2 carbon atoms (“C<sub>1-2</sub> haloalkyl”). Examples of haloalkyl groups include –CHF<sub>2</sub>, –CH<sub>2</sub>F, –CF<sub>3</sub>, –CH<sub>2</sub>CF<sub>3</sub>, –CF<sub>2</sub>CF<sub>3</sub>, –CF<sub>2</sub>CF<sub>2</sub>CF<sub>3</sub>, –CCl<sub>3</sub>, –CFCl<sub>2</sub>, –CF<sub>2</sub>Cl, and the like.

**[00256]** The term “heteroalkyl” refers to an alkyl group, which further includes at least one heteroatom (e.g., 1, 2, 3, or 4 heteroatoms) selected from oxygen, nitrogen, or sulfur within (e.g., inserted between adjacent carbon atoms of) and/or placed at one or more terminal position(s) of the parent chain. In certain embodiments, a heteroalkyl group refers to a saturated group having from 1 to 10 carbon atoms and 1 or more heteroatoms within the parent chain (“heteroC<sub>1-10</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 9 carbon atoms and 1 or more heteroatoms within the parent chain (“heteroC<sub>1-9</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 8 carbon atoms and 1 or more heteroatoms within the parent chain (“heteroC<sub>1-8</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 7 carbon atoms and 1 or more heteroatoms within the parent chain (“heteroC<sub>1-7</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 6 carbon atoms and 1 or more heteroatoms within the parent chain (“heteroC<sub>1-6</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 5 carbon atoms and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>1-5</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 4 carbon atoms and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>1-4</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 3 carbon atoms and 1 heteroatom within the parent chain (“heteroC<sub>1-3</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 to 2 carbon atoms and 1 heteroatom within the parent chain (“heteroC<sub>1-2</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 1 carbon atom and 1 heteroatom (“heteroC<sub>1</sub> alkyl”). In some embodiments, a heteroalkyl group is a saturated group having 2 to 6 carbon atoms and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-6</sub> alkyl”). Unless otherwise specified, each instance of a heteroalkyl group is independently unsubstituted (an “unsubstituted heteroalkyl”) or substituted (a “substituted heteroalkyl”) with one or more substituents. In certain embodiments, the heteroalkyl group is an unsubstituted heteroC<sub>1-10</sub> alkyl. In certain embodiments, the heteroalkyl group is a substituted heteroC<sub>1-10</sub> alkyl.

**[00257]** The term “alkenyl” refers to a radical of a straight-chain or branched hydrocarbon group having from 2 to 10 carbon atoms and one or more carbon-carbon double bonds (e.g., 1, 2, 3, or 4 double bonds). In some embodiments, an alkenyl group has 2 to 9 carbon atoms (“C<sub>2-9</sub> alkenyl”). In some embodiments, an alkenyl group has 2 to 8 carbon atoms (“C<sub>2-8</sub> alkenyl”). In some embodiments, an alkenyl group has 2 to 7 carbon atoms (“C<sub>2-7</sub> alkenyl”). In some embodiments, an alkenyl group has 2 to 6 carbon atoms (“C<sub>2-6</sub> alkenyl”). In some embodiments, an alkenyl group has 2 to 5 carbon atoms (“C<sub>2-5</sub> alkenyl”). In some embodiments, an alkenyl group has 2 to 4 carbon atoms (“C<sub>2-4</sub> alkenyl”). In some embodiments, an alkenyl group has 2 to 3 carbon atoms (“C<sub>2-3</sub> alkenyl”). In some embodiments, an alkenyl group has 2 carbon atoms (“C<sub>2</sub> alkenyl”). The one or more carbon-carbon double bonds can be internal (such as in 2-butenyl) or terminal (such as in 1-butenyl). Examples of C<sub>2-4</sub> alkenyl groups include ethenyl (C<sub>2</sub>), 1-propenyl (C<sub>3</sub>), 2-propenyl (C<sub>3</sub>), 1-butenyl (C<sub>4</sub>), 2-butenyl (C<sub>4</sub>), butadienyl (C<sub>4</sub>), and the like. Examples of C<sub>2-6</sub> alkenyl groups include the aforementioned C<sub>2-4</sub> alkenyl groups as well as pentenyl (C<sub>5</sub>), pentadienyl (C<sub>5</sub>), hexenyl (C<sub>6</sub>), and the like. Additional examples of alkenyl include heptenyl (C<sub>7</sub>), octenyl (C<sub>8</sub>), octatrienyl (C<sub>8</sub>), and the like. Unless otherwise specified, each instance of an alkenyl group is independently unsubstituted (an “unsubstituted alkenyl”) or substituted (a “substituted alkenyl”) with one or more substituents. In certain embodiments, the alkenyl group is an unsubstituted C<sub>2-10</sub> alkenyl. In certain embodiments, the alkenyl group is a substituted C<sub>2-10</sub> alkenyl. In an alkenyl group, a C=C double bond for which the stereochemistry is not specified (e.g., -CH=CHCH<sub>3</sub> or ) may be an (E)- or (Z)-double bond.

**[00258]** The term “heteroalkenyl” refers to an alkenyl group, which further includes at least one heteroatom (e.g., 1, 2, 3, or 4 heteroatoms) selected from oxygen, nitrogen, or sulfur within (e.g., inserted between adjacent carbon atoms of) and/or placed at one or more terminal position(s) of the parent chain. In certain embodiments, a heteroalkenyl group refers to a group having from 2 to 10 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-10</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 9 carbon atoms at least one double bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-9</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 8 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-8</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 7 carbon atoms, at least one double bond, and 1 or more

heteroatoms within the parent chain (“heteroC<sub>2-7</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 6 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-6</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 5 carbon atoms, at least one double bond, and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-5</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 4 carbon atoms, at least one double bond, and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-4</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 3 carbon atoms, at least one double bond, and 1 heteroatom within the parent chain (“heteroC<sub>2-3</sub> alkenyl”). In some embodiments, a heteroalkenyl group has 2 to 6 carbon atoms, at least one double bond, and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-6</sub> alkenyl”). Unless otherwise specified, each instance of a heteroalkenyl group is independently unsubstituted (an “unsubstituted heteroalkenyl”) or substituted (a “substituted heteroalkenyl”) with one or more substituents. In certain embodiments, the heteroalkenyl group is an unsubstituted heteroC<sub>2-10</sub> alkenyl. In certain embodiments, the heteroalkenyl group is a substituted heteroC<sub>2-10</sub> alkenyl.

**[00259]** The term “alkynyl” refers to a radical of a straight-chain or branched hydrocarbon group having from 2 to 10 carbon atoms and one or more carbon-carbon triple bonds (e.g., 1, 2, 3, or 4 triple bonds) (“C<sub>2-10</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 9 carbon atoms (“C<sub>2-9</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 8 carbon atoms (“C<sub>2-8</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 7 carbon atoms (“C<sub>2-7</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 6 carbon atoms (“C<sub>2-6</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 5 carbon atoms (“C<sub>2-5</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 4 carbon atoms (“C<sub>2-4</sub> alkynyl”). In some embodiments, an alkynyl group has 2 to 3 carbon atoms (“C<sub>2-3</sub> alkynyl”). In some embodiments, an alkynyl group has 2 carbon atoms (“C<sub>2</sub> alkynyl”). The one or more carbon-carbon triple bonds can be internal (such as in 2-butyne) or terminal (such as in 1-butyne). Examples of C<sub>2-4</sub> alkynyl groups include, without limitation, ethynyl (C<sub>2</sub>), 1-propynyl (C<sub>3</sub>), 2-propynyl (C<sub>3</sub>), 1-butyne (C<sub>4</sub>), 2-butyne (C<sub>4</sub>), and the like. Examples of C<sub>2-6</sub> alkenyl groups include the aforementioned C<sub>2-4</sub> alkynyl groups as well as pentynyl (C<sub>5</sub>), hexynyl (C<sub>6</sub>), and the like. Additional examples of alkynyl include heptyne (C<sub>7</sub>), octyne (C<sub>8</sub>), and the like. Unless otherwise specified, each instance of an alkynyl group is independently unsubstituted (an “unsubstituted alkynyl”) or substituted (a “substituted alkynyl”) with one or more substituents. In certain embodiments, the alkynyl group is an unsubstituted C<sub>2-10</sub> alkynyl. In certain embodiments, the alkynyl group is a substituted C<sub>2-10</sub> alkynyl.

**[00260]** The term “heteroalkynyl” refers to an alkynyl group, which further includes at least one heteroatom (e.g., 1, 2, 3, or 4 heteroatoms) selected from oxygen, nitrogen, or sulfur within (e.g., inserted between adjacent carbon atoms of) and/or placed at one or more terminal position(s) of the parent chain. In certain embodiments, a heteroalkynyl group refers to a group having from 2 to 10 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-10</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 9 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-9</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 8 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-8</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 7 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-7</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 6 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain (“heteroC<sub>2-6</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 5 carbon atoms, at least one triple bond, and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-5</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 4 carbon atoms, at least one triple bond, and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-4</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 3 carbon atoms, at least one triple bond, and 1 heteroatom within the parent chain (“heteroC<sub>2-3</sub> alkynyl”). In some embodiments, a heteroalkynyl group has 2 to 6 carbon atoms, at least one triple bond, and 1 or 2 heteroatoms within the parent chain (“heteroC<sub>2-6</sub> alkynyl”). Unless otherwise specified, each instance of a heteroalkynyl group is independently unsubstituted (an “unsubstituted heteroalkynyl”) or substituted (a “substituted heteroalkynyl”) with one or more substituents. In certain embodiments, the heteroalkynyl group is an unsubstituted heteroC<sub>2-10</sub> alkynyl. In certain embodiments, the heteroalkynyl group is a substituted heteroC<sub>2-10</sub> alkynyl.

**[00261]** The term “carbocyclyl” or “carbocyclic” refers to a radical of a non-aromatic cyclic hydrocarbon group having from 3 to 14 ring carbon atoms (“C<sub>3-14</sub> carbocyclyl”) and zero heteroatoms in the non-aromatic ring system. In some embodiments, a carbocyclyl group has 3 to 10 ring carbon atoms (“C<sub>3-10</sub> carbocyclyl”). In some embodiments, a carbocyclyl group has 3 to 8 ring carbon atoms (“C<sub>3-8</sub> carbocyclyl”). In some embodiments, a carbocyclyl group has 3 to 7 ring carbon atoms (“C<sub>3-7</sub> carbocyclyl”). In some embodiments, a carbocyclyl group has 3 to 6 ring carbon atoms (“C<sub>3-6</sub> carbocyclyl”). In some embodiments, a carbocyclyl group has 4 to 6 ring carbon atoms (“C<sub>4-6</sub> carbocyclyl”). In some embodiments, a carbocyclyl

group has 5 to 6 ring carbon atoms (“C<sub>5-6</sub> carbocyclyl”). In some embodiments, a carbocyclyl group has 5 to 10 ring carbon atoms (“C<sub>5-10</sub> carbocyclyl”). Exemplary C<sub>3-6</sub> carbocyclyl groups include, without limitation, cyclopropyl (C<sub>3</sub>), cyclopropenyl (C<sub>3</sub>), cyclobutyl (C<sub>4</sub>), cyclobutenyl (C<sub>4</sub>), cyclopentyl (C<sub>5</sub>), cyclopentenyl (C<sub>5</sub>), cyclohexyl (C<sub>6</sub>), cyclohexenyl (C<sub>6</sub>), cyclohexadienyl (C<sub>6</sub>), and the like. Exemplary C<sub>3-8</sub> carbocyclyl groups include, without limitation, the aforementioned C<sub>3-6</sub> carbocyclyl groups as well as cycloheptyl (C<sub>7</sub>), cycloheptenyl (C<sub>7</sub>), cycloheptadienyl (C<sub>7</sub>), cycloheptatrienyl (C<sub>7</sub>), cyclooctyl (C<sub>8</sub>), cyclooctenyl (C<sub>8</sub>), bicyclo[2.2.1]heptanyl (C<sub>7</sub>), bicyclo[2.2.2]octanyl (C<sub>8</sub>), and the like. Exemplary C<sub>3-10</sub> carbocyclyl groups include, without limitation, the aforementioned C<sub>3-8</sub> carbocyclyl groups as well as cyclononyl (C<sub>9</sub>), cyclononenyl (C<sub>9</sub>), cyclodecyl (C<sub>10</sub>), cyclodecenyl (C<sub>10</sub>), octahydro-1H-indenyl (C<sub>9</sub>), decahydronaphthalenyl (C<sub>10</sub>), spiro[4.5]decanyl (C<sub>10</sub>), and the like. As the foregoing examples illustrate, in certain embodiments, the carbocyclyl group is either monocyclic (“monocyclic carbocyclyl”) or polycyclic (e.g., containing a fused, bridged or spiro ring system such as a bicyclic system (“bicyclic carbocyclyl”) or tricyclic system (“tricyclic carbocyclyl”)) and can be saturated or can contain one or more carbon-carbon double or triple bonds. “Carbocyclyl” also includes ring systems wherein the carbocyclyl ring, as defined above, is fused with one or more aryl or heteroaryl groups wherein the point of attachment is on the carbocyclyl ring, and in such instances, the number of carbons continue to designate the number of carbons in the carbocyclic ring system. Unless otherwise specified, each instance of a carbocyclyl group is independently unsubstituted (an “unsubstituted carbocyclyl”) or substituted (a “substituted carbocyclyl”) with one or more substituents. In certain embodiments, the carbocyclyl group is an unsubstituted C<sub>3-14</sub> carbocyclyl. In certain embodiments, the carbocyclyl group is a substituted C<sub>3-14</sub> carbocyclyl.

**[00262]** In some embodiments, “carbocyclyl” is a monocyclic, saturated carbocyclyl group having from 3 to 14 ring carbon atoms (“C<sub>3-14</sub> cycloalkyl”). In some embodiments, a cycloalkyl group has 3 to 10 ring carbon atoms (“C<sub>3-10</sub> cycloalkyl”). In some embodiments, a cycloalkyl group has 3 to 8 ring carbon atoms (“C<sub>3-8</sub> cycloalkyl”). In some embodiments, a cycloalkyl group has 3 to 6 ring carbon atoms (“C<sub>3-6</sub> cycloalkyl”). In some embodiments, a cycloalkyl group has 4 to 6 ring carbon atoms (“C<sub>4-6</sub> cycloalkyl”). In some embodiments, a cycloalkyl group has 5 to 6 ring carbon atoms (“C<sub>5-6</sub> cycloalkyl”). In some embodiments, a cycloalkyl group has 5 to 10 ring carbon atoms (“C<sub>5-10</sub> cycloalkyl”). Examples of C<sub>5-6</sub> cycloalkyl groups include cyclopentyl (C<sub>5</sub>) and cyclohexyl (C<sub>5</sub>). Examples of C<sub>3-6</sub> cycloalkyl groups include the aforementioned C<sub>5-6</sub> cycloalkyl groups as well as cyclopropyl (C<sub>3</sub>) and

cyclobutyl (C<sub>4</sub>). Examples of C<sub>3-8</sub> cycloalkyl groups include the aforementioned C<sub>3-6</sub> cycloalkyl groups as well as cycloheptyl (C<sub>7</sub>) and cyclooctyl (C<sub>8</sub>). Unless otherwise specified, each instance of a cycloalkyl group is independently unsubstituted (an “unsubstituted cycloalkyl”) or substituted (a “substituted cycloalkyl”) with one or more substituents. In certain embodiments, the cycloalkyl group is an unsubstituted C<sub>3-14</sub> cycloalkyl. In certain embodiments, the cycloalkyl group is a substituted C<sub>3-14</sub> cycloalkyl.

**[00263]** The term “heterocyclyl” or “heterocyclic” refers to a radical of a 3- to 14-membered non-aromatic ring system having ring carbon atoms and 1 to 4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“3-14 membered heterocyclyl”). In heterocyclyl groups that contain one or more nitrogen atoms, the point of attachment can be a carbon or nitrogen atom, as valency permits. A heterocyclyl group can either be monocyclic (“monocyclic heterocyclyl”) or polycyclic (e.g., a fused, bridged or spiro ring system such as a bicyclic system (“bicyclic heterocyclyl”) or tricyclic system (“tricyclic heterocyclyl”)), and can be saturated or can contain one or more carbon-carbon double or triple bonds. Heterocyclyl polycyclic ring systems can include one or more heteroatoms in one or both rings. “Heterocyclyl” also includes ring systems wherein the heterocyclyl ring, as defined above, is fused with one or more carbocyclyl groups wherein the point of attachment is either on the carbocyclyl or heterocyclyl ring, or ring systems wherein the heterocyclyl ring, as defined above, is fused with one or more aryl or heteroaryl groups, wherein the point of attachment is on the heterocyclyl ring, and in such instances, the number of ring members continue to designate the number of ring members in the heterocyclyl ring system. Unless otherwise specified, each instance of heterocyclyl is independently unsubstituted (an “unsubstituted heterocyclyl”) or substituted (a “substituted heterocyclyl”) with one or more substituents. In certain embodiments, the heterocyclyl group is an unsubstituted 3-14 membered heterocyclyl. In certain embodiments, the heterocyclyl group is a substituted 3-14 membered heterocyclyl.

**[00264]** In some embodiments, a heterocyclyl group is a 5-10 membered non-aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“5-10 membered heterocyclyl”). In some embodiments, a heterocyclyl group is a 5-8 membered non-aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“5-8 membered heterocyclyl”). In some embodiments, a heterocyclyl group is a 5-6 membered non-aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms, wherein each heteroatom is independently selected

from nitrogen, oxygen, and sulfur (“5-6 membered heterocyclyl”). In some embodiments, the 5-6 membered heterocyclyl has 1-3 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heterocyclyl has 1-2 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heterocyclyl has 1 ring heteroatom selected from nitrogen, oxygen, and sulfur.

**[00265]** Exemplary 3-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azirdinyl, oxiranyl, and thiiranyl. Exemplary 4-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azetidiny, oxetanyl, and thietanyl. Exemplary 5-membered heterocyclyl groups containing 1 heteroatom include, without limitation, tetrahydrofuranyl, dihydrofuranyl, tetrahydrothiophenyl, dihydrothiophenyl, pyrrolidinyl, dihydropyrrolyl, and pyrrolyl-2,5-dione. Exemplary 5-membered heterocyclyl groups containing 2 heteroatoms include, without limitation, dioxolanyl, oxathiolanyl and dithiolanyl. Exemplary 5-membered heterocyclyl groups containing 3 heteroatoms include, without limitation, triazoliny, oxadiazoliny, and thiadiazoliny. Exemplary 6-membered heterocyclyl groups containing 1 heteroatom include, without limitation, piperidinyl, tetrahydropyranyl, dihydropyridiny, and thianyl. Exemplary 6-membered heterocyclyl groups containing 2 heteroatoms include, without limitation, piperazinyl, morpholinyl, dithianyl, and dioxanyl. Exemplary 6-membered heterocyclyl groups containing 3 heteroatoms include, without limitation, triazinyl. Exemplary 7-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azepanyl, oxepanyl and thiepanyl. Exemplary 8-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azocanyl, oxecanyl and thiocanyl. Exemplary bicyclic heterocyclyl groups include, without limitation, indolinyl, isoindolinyl, dihydrobenzofuranyl, dihydrobenzothieryl, tetrahydrobenzothieryl, tetrahydrobenzofuranyl, tetrahydroindolyl, tetrahydroquinoliny, tetrahydroisoquinoliny, decahydroquinoliny, decahydroisoquinoliny, octahydrochromenyl, octahydroisochromenyl, decahydronaphthyridiny, decahydro-1,8-naphthyridiny, octahydropyrrolo[3,2-b]pyrrole, indolinyl, phthalimidyl, naphthalimidyl, chromanyl, chromenyl, 1H-benzo[e][1,4]diazepiny, 1,4,5,7-tetrahydropyrano[3,4-b]pyrrolyl, 5,6-dihydro-4H-furo[3,2-b]pyrrolyl, 6,7-dihydro-5H-furo[3,2-b]pyranyl, 5,7-dihydro-4H-thieno[2,3-c]pyranyl, 2,3-dihydro-1H-pyrrolo[2,3-b]pyridiny, 2,3-dihydrofuro[2,3-b]pyridiny, 4,5,6,7-tetrahydro-1H-pyrrolo[2,3-b]pyridiny, 4,5,6,7-tetrahydrofuro[3,2-c]pyridiny, 4,5,6,7-tetrahydrothieno[3,2-b]pyridiny, 1,2,3,4-tetrahydro-1,6-naphthyridiny, and the like.

**[00266]** The term “aryl” refers to a radical of a monocyclic or polycyclic (e.g., bicyclic or tricyclic)  $4n+2$  aromatic ring system (e.g., having 6, 10, or 14  $\pi$  electrons shared in a cyclic array) having 6-14 ring carbon atoms and zero heteroatoms provided in the aromatic ring system (“C<sub>6-14</sub> aryl”). In some embodiments, an aryl group has 6 ring carbon atoms (“C<sub>6</sub> aryl”; e.g., phenyl). In some embodiments, an aryl group has 10 ring carbon atoms (“C<sub>10</sub> aryl”; e.g., naphthyl such as 1-naphthyl and 2-naphthyl). In some embodiments, an aryl group has 14 ring carbon atoms (“C<sub>14</sub> aryl”; e.g., anthracyl). “Aryl” also includes ring systems wherein the aryl ring, as defined above, is fused with one or more carbocyclyl or heterocyclyl groups wherein the radical or point of attachment is on the aryl ring, and in such instances, the number of carbon atoms continue to designate the number of carbon atoms in the aryl ring system. Unless otherwise specified, each instance of an aryl group is independently unsubstituted (an “unsubstituted aryl”) or substituted (a “substituted aryl”) with one or more substituents. In certain embodiments, the aryl group is an unsubstituted C<sub>6-14</sub> aryl. In certain embodiments, the aryl group is a substituted C<sub>6-14</sub> aryl.

**[00267]** The term “heteroaryl” refers to a radical of a 5-14 membered monocyclic or polycyclic (e.g., bicyclic, tricyclic)  $4n+2$  aromatic ring system (e.g., having 6, 10, or 14  $\pi$  electrons shared in a cyclic array) having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“5-14 membered heteroaryl”). In heteroaryl groups that contain one or more nitrogen atoms, the point of attachment can be a carbon or nitrogen atom, as valency permits. Heteroaryl polycyclic ring systems can include one or more heteroatoms in one or both rings. “Heteroaryl” includes ring systems wherein the heteroaryl ring, as defined above, is fused with one or more carbocyclyl or heterocyclyl groups wherein the point of attachment is on the heteroaryl ring, and in such instances, the number of ring members continue to designate the number of ring members in the heteroaryl ring system. “Heteroaryl” also includes ring systems wherein the heteroaryl ring, as defined above, is fused with one or more aryl groups wherein the point of attachment is either on the aryl or heteroaryl ring, and in such instances, the number of ring members designates the number of ring members in the fused polycyclic (aryl/heteroaryl) ring system. Polycyclic heteroaryl groups wherein one ring does not contain a heteroatom (e.g., indolyl, quinolinyl, carbazolyl, and the like) the point of attachment can be on either ring, e.g., either the ring bearing a heteroatom (e.g., 2-indolyl) or the ring that does not contain a heteroatom (e.g., 5-indolyl).

**[00268]** In some embodiments, a heteroaryl group is a 5-10 membered aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“5-10 membered heteroaryl”). In some embodiments, a heteroaryl group is a 5-8 membered aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“5-8 membered heteroaryl”). In some embodiments, a heteroaryl group is a 5-6 membered aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur (“5-6 membered heteroaryl”). In some embodiments, the 5-6 membered heteroaryl has 1-3 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heteroaryl has 1-2 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heteroaryl has 1 ring heteroatom selected from nitrogen, oxygen, and sulfur. Unless otherwise specified, each instance of a heteroaryl group is independently unsubstituted (an “unsubstituted heteroaryl”) or substituted (a “substituted heteroaryl”) with one or more substituents. In certain embodiments, the heteroaryl group is an unsubstituted 5-14 membered heteroaryl. In certain embodiments, the heteroaryl group is a substituted 5-14 membered heteroaryl.

**[00269]** Exemplary 5-membered heteroaryl groups containing 1 heteroatom include, without limitation, pyrrolyl, furanyl, and thiophenyl. Exemplary 5-membered heteroaryl groups containing 2 heteroatoms include, without limitation, imidazolyl, pyrazolyl, oxazolyl, isoxazolyl, thiazolyl, and isothiazolyl. Exemplary 5-membered heteroaryl groups containing 3 heteroatoms include, without limitation, triazolyl, oxadiazolyl, and thiadiazolyl. Exemplary 5-membered heteroaryl groups containing 4 heteroatoms include, without limitation, tetrazolyl. Exemplary 6-membered heteroaryl groups containing 1 heteroatom include, without limitation, pyridinyl. Exemplary 6-membered heteroaryl groups containing 2 heteroatoms include, without limitation, pyridazinyl, pyrimidinyl, and pyrazinyl. Exemplary 6-membered heteroaryl groups containing 3 or 4 heteroatoms include, without limitation, triazinyl and tetrazinyl, respectively. Exemplary 7-membered heteroaryl groups containing 1 heteroatom include, without limitation, azepinyl, oxepinyl, and thiepinyl. Exemplary 5,6-bicyclic heteroaryl groups include, without limitation, indolyl, isoindolyl, indazolyl, benzotriazolyl, benzothiophenyl, isobenzothiophenyl, benzofuranyl, benzoisofuranyl, benzimidazolyl, benzoxazolyl, benzisoxazolyl, benzoxadiazolyl, benzthiazolyl, benzisothiazolyl, benzthiadiazolyl, indoliziny, and purinyl. Exemplary 6,6-bicyclic

heteroaryl groups include, without limitation, naphthyridinyl, pteridinyl, quinolinyl, isoquinolinyl, cinnolinyl, quinoxalinyl, phthalazinyl, and quinazolinyl. Exemplary tricyclic heteroaryl groups include, without limitation, phenanthridinyl, dibenzofuranyl, carbazolyl, acridinyl, phenothiazinyl, phenoxazinyl, and phenazinyl.

**[00270]** The term “unsaturated bond” refers to a double or triple bond.

**[00271]** The term “unsaturated” or “partially unsaturated” refers to a moiety that includes at least one double or triple bond.

**[00272]** The term “saturated” refers to a moiety that does not contain a double or triple bond, i.e., the moiety only contains single bonds.

**[00273]** Affixing the suffix “-ene” to a group indicates the group is a divalent moiety, e.g., alkylene is the divalent moiety of alkyl, alkenylene is the divalent moiety of alkenyl, alkynylene is the divalent moiety of alkynyl, heteroalkylene is the divalent moiety of heteroalkyl, heteroalkenylene is the divalent moiety of heteroalkenyl, heteroalkynylene is the divalent moiety of heteroalkynyl, carbocyclylene is the divalent moiety of carbocyclyl, heterocyclylene is the divalent moiety of heterocyclyl, arylene is the divalent moiety of aryl, and heteroarylene is the divalent moiety of heteroaryl.

**[00274]** A group is optionally substituted unless expressly provided otherwise. The term “optionally substituted” refers to being substituted or unsubstituted. In certain embodiments, alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl groups are optionally substituted. “Optionally substituted” refers to a group which may be substituted or unsubstituted (e.g., “substituted” or “unsubstituted” alkyl, “substituted” or “unsubstituted” alkenyl, “substituted” or “unsubstituted” alkynyl, “substituted” or “unsubstituted” heteroalkyl, “substituted” or “unsubstituted” heteroalkenyl, “substituted” or “unsubstituted” heteroalkynyl, “substituted” or “unsubstituted” carbocyclyl, “substituted” or “unsubstituted” heterocyclyl, “substituted” or “unsubstituted” aryl or “substituted” or “unsubstituted” heteroaryl group). In general, the term “substituted” means that at least one hydrogen present on a group is replaced with a permissible substituent, e.g., a substituent which upon substitution results in a stable compound, e.g., a compound which does not spontaneously undergo transformation such as by rearrangement, cyclization, elimination, or other reaction. Unless otherwise indicated, a “substituted” group has a substituent at one or more substitutable positions of the group, and when more than one position in any given structure is substituted, the substituent is either the same or different at each position. The term “substituted” is contemplated to include substitution with all permissible substituents of organic compounds, and includes any of the

substituents described herein that results in the formation of a stable compound. The present invention contemplates any and all such combinations in order to arrive at a stable compound. For purposes of this invention, heteroatoms such as nitrogen may have hydrogen substituents and/or any suitable substituent as described herein which satisfy the valencies of the heteroatoms and results in the formation of a stable moiety. The invention is not intended to be limited in any manner by the exemplary substituents described herein.

**[00275]** Exemplary carbon atom substituents include, but are not limited to, halogen,  $-\text{CN}$ ,  $-\text{NO}_2$ ,  $-\text{N}_3$ ,  $-\text{SO}_2\text{H}$ ,  $-\text{SO}_3\text{H}$ ,  $-\text{OH}$ ,  $-\text{OR}^{\text{aa}}$ ,  $-\text{ON}(\text{R}^{\text{bb}})_2$ ,  $-\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{N}(\text{R}^{\text{bb}})_3^+\text{X}^-$ ,  $-\text{N}(\text{OR}^{\text{cc}})\text{R}^{\text{bb}}$ ,  $-\text{SH}$ ,  $-\text{SR}^{\text{aa}}$ ,  $-\text{SSR}^{\text{cc}}$ ,  $-\text{C}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{CO}_2\text{H}$ ,  $-\text{CHO}$ ,  $-\text{C}(\text{OR}^{\text{cc}})_3$ ,  $-\text{CO}_2\text{R}^{\text{aa}}$ ,  $-\text{OC}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{OCO}_2\text{R}^{\text{aa}}$ ,  $-\text{C}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{OC}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{NR}^{\text{bb}}\text{C}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{NR}^{\text{bb}}\text{CO}_2\text{R}^{\text{aa}}$ ,  $-\text{NR}^{\text{bb}}\text{C}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{C}(=\text{NR}^{\text{bb}})\text{R}^{\text{aa}}$ ,  $-\text{C}(=\text{NR}^{\text{bb}})\text{OR}^{\text{aa}}$ ,  $-\text{OC}(=\text{NR}^{\text{bb}})\text{R}^{\text{aa}}$ ,  $-\text{OC}(=\text{NR}^{\text{bb}})\text{OR}^{\text{aa}}$ ,  $-\text{C}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{OC}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{NR}^{\text{bb}}\text{C}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{C}(=\text{O})\text{NR}^{\text{bb}}\text{SO}_2\text{R}^{\text{aa}}$ ,  $-\text{NR}^{\text{bb}}\text{SO}_2\text{R}^{\text{aa}}$ ,  $-\text{SO}_2\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{SO}_2\text{R}^{\text{aa}}$ ,  $-\text{SO}_2\text{OR}^{\text{aa}}$ ,  $-\text{OSO}_2\text{R}^{\text{aa}}$ ,  $-\text{S}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{OS}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{Si}(\text{R}^{\text{aa}})_3$ ,  $-\text{OSi}(\text{R}^{\text{aa}})_3$ ,  $-\text{C}(=\text{S})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{C}(=\text{O})\text{SR}^{\text{aa}}$ ,  $-\text{C}(=\text{S})\text{SR}^{\text{aa}}$ ,  $-\text{SC}(=\text{S})\text{SR}^{\text{aa}}$ ,  $-\text{SC}(=\text{O})\text{SR}^{\text{aa}}$ ,  $-\text{OC}(=\text{O})\text{SR}^{\text{aa}}$ ,  $-\text{SC}(=\text{O})\text{OR}^{\text{aa}}$ ,  $-\text{SC}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{P}(=\text{O})(\text{R}^{\text{aa}})_2$ ,  $-\text{P}(=\text{O})(\text{OR}^{\text{cc}})_2$ ,  $-\text{OP}(=\text{O})(\text{R}^{\text{aa}})_2$ ,  $-\text{OP}(=\text{O})(\text{OR}^{\text{cc}})_2$ ,  $-\text{P}(=\text{O})(\text{N}(\text{R}^{\text{bb}})_2)_2$ ,  $-\text{OP}(=\text{O})(\text{N}(\text{R}^{\text{bb}})_2)_2$ ,  $-\text{NR}^{\text{bb}}\text{P}(=\text{O})(\text{R}^{\text{aa}})_2$ ,  $-\text{NR}^{\text{bb}}\text{P}(=\text{O})(\text{OR}^{\text{cc}})_2$ ,  $-\text{NR}^{\text{bb}}\text{P}(=\text{O})(\text{N}(\text{R}^{\text{bb}})_2)_2$ ,  $-\text{P}(\text{R}^{\text{cc}})_2$ ,  $-\text{P}(\text{OR}^{\text{cc}})_2$ ,  $-\text{P}(\text{R}^{\text{cc}})_3^+\text{X}^-$ ,  $-\text{P}(\text{OR}^{\text{cc}})_3^+\text{X}^-$ ,  $-\text{P}(\text{R}^{\text{cc}})_4$ ,  $-\text{P}(\text{OR}^{\text{cc}})_4$ ,  $-\text{OP}(\text{R}^{\text{cc}})_2$ ,  $-\text{OP}(\text{R}^{\text{cc}})_3^+\text{X}^-$ ,  $-\text{OP}(\text{OR}^{\text{cc}})_2$ ,  $-\text{OP}(\text{OR}^{\text{cc}})_3^+\text{X}^-$ ,  $-\text{OP}(\text{R}^{\text{cc}})_4$ ,  $-\text{OP}(\text{OR}^{\text{cc}})_4$ ,  $-\text{B}(\text{R}^{\text{aa}})_2$ ,  $-\text{B}(\text{OR}^{\text{cc}})_2$ ,  $-\text{BR}^{\text{aa}}(\text{OR}^{\text{cc}})$ ,  $\text{C}_{1-10}$  alkyl,  $\text{C}_{1-10}$  perhaloalkyl,  $\text{C}_{2-10}$  alkenyl,  $\text{C}_{2-10}$  alkynyl, hetero $\text{C}_{1-10}$  alkyl, hetero $\text{C}_{2-10}$  alkenyl, hetero $\text{C}_{2-10}$  alkynyl,  $\text{C}_{3-10}$  carbocyclyl, 3-14 membered heterocyclyl,  $\text{C}_{6-14}$  aryl, and 5-14 membered heteroaryl, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5  $\text{R}^{\text{dd}}$  groups; wherein  $\text{X}^-$  is a counterion;

or two geminal hydrogens on a carbon atom are replaced with the group  $=\text{O}$ ,  $=\text{S}$ ,  $=\text{NN}(\text{R}^{\text{bb}})_2$ ,  $=\text{NNR}^{\text{bb}}\text{C}(=\text{O})\text{R}^{\text{aa}}$ ,  $=\text{NNR}^{\text{bb}}\text{C}(=\text{O})\text{OR}^{\text{aa}}$ ,  $=\text{NNR}^{\text{bb}}\text{S}(=\text{O})_2\text{R}^{\text{aa}}$ ,  $=\text{NR}^{\text{bb}}$ , or  $=\text{NOR}^{\text{cc}}$ ;

each instance of  $\text{R}^{\text{aa}}$  is, independently, selected from  $\text{C}_{1-10}$  alkyl,  $\text{C}_{1-10}$  perhaloalkyl,  $\text{C}_{2-10}$  alkenyl,  $\text{C}_{2-10}$  alkynyl, hetero $\text{C}_{1-10}$  alkyl, hetero $\text{C}_{2-10}$  alkenyl, hetero $\text{C}_{2-10}$  alkynyl,  $\text{C}_{3-10}$  carbocyclyl, 3-14 membered heterocyclyl,  $\text{C}_{6-14}$  aryl, and 5-14 membered heteroaryl, or two  $\text{R}^{\text{aa}}$  groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5  $\text{R}^{\text{dd}}$  groups;

each instance of R<sup>bb</sup> is, independently, selected from hydrogen, -OH, -OR<sup>aa</sup>, -N(R<sup>cc</sup>)<sub>2</sub>, -CN, -C(=O)R<sup>aa</sup>, -C(=O)N(R<sup>cc</sup>)<sub>2</sub>, -CO<sub>2</sub>R<sup>aa</sup>, -SO<sub>2</sub>R<sup>aa</sup>, -C(=NR<sup>cc</sup>)OR<sup>aa</sup>, -C(=NR<sup>cc</sup>)N(R<sup>cc</sup>)<sub>2</sub>, -SO<sub>2</sub>N(R<sup>cc</sup>)<sub>2</sub>, -SO<sub>2</sub>R<sup>cc</sup>, -SO<sub>2</sub>OR<sup>cc</sup>, -SOR<sup>aa</sup>, -C(=S)N(R<sup>cc</sup>)<sub>2</sub>, -C(=O)SR<sup>cc</sup>, -C(=S)SR<sup>cc</sup>, -P(=O)(R<sup>aa</sup>)<sub>2</sub>, -P(=O)(OR<sup>cc</sup>)<sub>2</sub>, -P(=O)(N(R<sup>cc</sup>)<sub>2</sub>)<sub>2</sub>, C<sub>1-10</sub> alkyl, C<sub>1-10</sub> perhaloalkyl, C<sub>2-10</sub> alkenyl, C<sub>2-10</sub> alkynyl, heteroC<sub>1-10</sub> alkyl, heteroC<sub>2-10</sub> alkenyl, heteroC<sub>2-10</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, 3-14 membered heterocyclyl, C<sub>6-14</sub> aryl, and 5-14 membered heteroaryl, or two R<sup>bb</sup> groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R<sup>dd</sup> groups; wherein X<sup>-</sup> is a counterion;

each instance of R<sup>cc</sup> is, independently, selected from hydrogen, C<sub>1-10</sub> alkyl, C<sub>1-10</sub> perhaloalkyl, C<sub>2-10</sub> alkenyl, C<sub>2-10</sub> alkynyl, heteroC<sub>1-10</sub> alkyl, heteroC<sub>2-10</sub> alkenyl, heteroC<sub>2-10</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, 3-14 membered heterocyclyl, C<sub>6-14</sub> aryl, and 5-14 membered heteroaryl, or two R<sup>cc</sup> groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R<sup>dd</sup> groups;

each instance of R<sup>dd</sup> is, independently, selected from halogen, -CN, -NO<sub>2</sub>, -N<sub>3</sub>, -SO<sub>2</sub>H, -SO<sub>3</sub>H, -OH, -OR<sup>ee</sup>, -ON(R<sup>ff</sup>)<sub>2</sub>, -N(R<sup>ff</sup>)<sub>2</sub>, -N(R<sup>ff</sup>)<sub>3</sub><sup>+</sup>X<sup>-</sup>, -N(OR<sup>ee</sup>)R<sup>ff</sup>, -SH, -SR<sup>ee</sup>, -SSR<sup>ee</sup>, -C(=O)R<sup>ee</sup>, -CO<sub>2</sub>H, -CO<sub>2</sub>R<sup>ee</sup>, -OC(=O)R<sup>ee</sup>, -OCO<sub>2</sub>R<sup>ee</sup>, -C(=O)N(R<sup>ff</sup>)<sub>2</sub>, -OC(=O)N(R<sup>ff</sup>)<sub>2</sub>, -NR<sup>ff</sup>C(=O)R<sup>ee</sup>, -NR<sup>ff</sup>CO<sub>2</sub>R<sup>ee</sup>, -NR<sup>ff</sup>C(=O)N(R<sup>ff</sup>)<sub>2</sub>, -C(=NR<sup>ff</sup>)OR<sup>ee</sup>, -OC(=NR<sup>ff</sup>)R<sup>ee</sup>, -OC(=NR<sup>ff</sup>)OR<sup>ee</sup>, -C(=NR<sup>ff</sup>)N(R<sup>ff</sup>)<sub>2</sub>, -OC(=NR<sup>ff</sup>)N(R<sup>ff</sup>)<sub>2</sub>, -NR<sup>ff</sup>C(=NR<sup>ff</sup>)N(R<sup>ff</sup>)<sub>2</sub>, -NR<sup>ff</sup>SO<sub>2</sub>R<sup>ee</sup>, -SO<sub>2</sub>N(R<sup>ff</sup>)<sub>2</sub>, -SO<sub>2</sub>R<sup>ee</sup>, -SO<sub>2</sub>OR<sup>ee</sup>, -OSO<sub>2</sub>R<sup>ee</sup>, -S(=O)R<sup>ee</sup>, -Si(R<sup>ee</sup>)<sub>3</sub>, -OSi(R<sup>ee</sup>)<sub>3</sub>, -C(=S)N(R<sup>ff</sup>)<sub>2</sub>, -C(=O)SR<sup>ee</sup>, -C(=S)SR<sup>ee</sup>, -SC(=S)SR<sup>ee</sup>, -P(=O)(OR<sup>ee</sup>)<sub>2</sub>, -P(=O)(R<sup>ee</sup>)<sub>2</sub>, -OP(=O)(R<sup>ee</sup>)<sub>2</sub>, -OP(=O)(OR<sup>ee</sup>)<sub>2</sub>, C<sub>1-6</sub> alkyl, C<sub>1-6</sub> perhaloalkyl, C<sub>2-6</sub> alkenyl, C<sub>2-6</sub> alkynyl, heteroC<sub>1-6</sub> alkyl, heteroC<sub>2-6</sub> alkenyl, heteroC<sub>2-6</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, 3-10 membered heterocyclyl, C<sub>6-10</sub> aryl, 5-10 membered heteroaryl, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R<sup>gg</sup> groups, or two geminal R<sup>dd</sup> substituents can be joined to form =O or =S; wherein X<sup>-</sup> is a counterion;

each instance of R<sup>ee</sup> is, independently, selected from C<sub>1-6</sub> alkyl, C<sub>1-6</sub> perhaloalkyl, C<sub>2-6</sub> alkenyl, C<sub>2-6</sub> alkynyl, heteroC<sub>1-6</sub> alkyl, heteroC<sub>2-6</sub> alkenyl, heteroC<sub>2-6</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, C<sub>6-10</sub> aryl, 3-10 membered heterocyclyl, and 3-10 membered heteroaryl, wherein

each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R<sup>gg</sup> groups;

each instance of R<sup>ff</sup> is, independently, selected from hydrogen, C<sub>1-6</sub> alkyl, C<sub>1-6</sub> perhaloalkyl, C<sub>2-6</sub> alkenyl, C<sub>2-6</sub> alkynyl, heteroC<sub>1-6</sub> alkyl, heteroC<sub>2-6</sub> alkenyl, heteroC<sub>2-6</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, 3-10 membered heterocyclyl, C<sub>6-10</sub> aryl and 5-10 membered heteroaryl, or two R<sup>ff</sup> groups are joined to form a 3-10 membered heterocyclyl or 5-10 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R<sup>gg</sup> groups; and

each instance of R<sup>gg</sup> is, independently, halogen, -CN, -NO<sub>2</sub>, -N<sub>3</sub>, -SO<sub>2</sub>H, -SO<sub>3</sub>H, -OH, -OC<sub>1-6</sub> alkyl, -ON(C<sub>1-6</sub> alkyl)<sub>2</sub>, -N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -N(C<sub>1-6</sub> alkyl)<sub>3</sub><sup>+</sup>X<sup>-</sup>, -NH(C<sub>1-6</sub> alkyl)<sub>2</sub><sup>+</sup>X<sup>-</sup>, -NH<sub>2</sub>(C<sub>1-6</sub> alkyl)<sup>+</sup>X<sup>-</sup>, -NH<sub>3</sub><sup>+</sup>X<sup>-</sup>, -N(OC<sub>1-6</sub> alkyl)(C<sub>1-6</sub> alkyl), -N(OH)(C<sub>1-6</sub> alkyl), -NH(OH), -SH, -SC<sub>1-6</sub> alkyl, -SS(C<sub>1-6</sub> alkyl), -C(=O)(C<sub>1-6</sub> alkyl), -CO<sub>2</sub>H, -CO<sub>2</sub>(C<sub>1-6</sub> alkyl), -OC(=O)(C<sub>1-6</sub> alkyl), -OCO<sub>2</sub>(C<sub>1-6</sub> alkyl), -C(=O)NH<sub>2</sub>, -C(=O)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OC(=O)NH(C<sub>1-6</sub> alkyl), -NHC(=O)(C<sub>1-6</sub> alkyl), -N(C<sub>1-6</sub> alkyl)C(=O)(C<sub>1-6</sub> alkyl), -NHCO<sub>2</sub>(C<sub>1-6</sub> alkyl), -NHC(=O)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -NHC(=O)NH(C<sub>1-6</sub> alkyl), -NHC(=O)NH<sub>2</sub>, -C(=NH)O(C<sub>1-6</sub> alkyl), -OC(=NH)(C<sub>1-6</sub> alkyl), -OC(=NH)OC<sub>1-6</sub> alkyl, -C(=NH)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -C(=NH)NH(C<sub>1-6</sub> alkyl), -C(=NH)NH<sub>2</sub>, -OC(=NH)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OC(=NH)NH(C<sub>1-6</sub> alkyl), -OC(=NH)NH<sub>2</sub>, -NHC(=NH)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -NHC(=NH)NH<sub>2</sub>, -NHCO<sub>2</sub>(C<sub>1-6</sub> alkyl), -SO<sub>2</sub>N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -SO<sub>2</sub>NH(C<sub>1-6</sub> alkyl), -SO<sub>2</sub>NH<sub>2</sub>, -SO<sub>2</sub>(C<sub>1-6</sub> alkyl), -SO<sub>2</sub>O(C<sub>1-6</sub> alkyl), -OSO<sub>2</sub>(C<sub>1-6</sub> alkyl), -SO(C<sub>1-6</sub> alkyl), -Si(C<sub>1-6</sub> alkyl)<sub>3</sub>, -OSi(C<sub>1-6</sub> alkyl)<sub>3</sub>, -C(=S)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, C(=S)NH(C<sub>1-6</sub> alkyl), C(=S)NH<sub>2</sub>, -C(=O)S(C<sub>1-6</sub> alkyl), -C(=S)SC<sub>1-6</sub> alkyl, -SC(=S)SC<sub>1-6</sub> alkyl, -P(=O)(OC<sub>1-6</sub> alkyl)<sub>2</sub>, -P(=O)(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OP(=O)(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OP(=O)(OC<sub>1-6</sub> alkyl)<sub>2</sub>, C<sub>1-6</sub> alkyl, C<sub>1-6</sub> perhaloalkyl, C<sub>2-6</sub> alkenyl, C<sub>2-6</sub> alkynyl, heteroC<sub>1-6</sub> alkyl, heteroC<sub>2-6</sub> alkenyl, heteroC<sub>2-6</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, C<sub>6-10</sub> aryl, 3-10 membered heterocyclyl, 5-10 membered heteroaryl; or two geminal R<sup>gg</sup> substituents can be joined to form =O or =S; wherein X<sup>-</sup> is a counterion.

**[00276]** In certain embodiments, carbon atom substituents include: halogen, -CN, -NO<sub>2</sub>, -N<sub>3</sub>, -SO<sub>2</sub>H, -SO<sub>3</sub>H, -OH, -OC<sub>1-6</sub> alkyl, -ON(C<sub>1-6</sub> alkyl)<sub>2</sub>, -N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -N(C<sub>1-6</sub> alkyl)<sub>3</sub><sup>+</sup>X<sup>-</sup>, -NH(C<sub>1-6</sub> alkyl)<sub>2</sub><sup>+</sup>X<sup>-</sup>, -NH<sub>2</sub>(C<sub>1-6</sub> alkyl)<sup>+</sup>X<sup>-</sup>, -NH<sub>3</sub><sup>+</sup>X<sup>-</sup>, -N(OC<sub>1-6</sub> alkyl)(C<sub>1-6</sub> alkyl), -N(OH)(C<sub>1-6</sub> alkyl), -NH(OH), -SH, -SC<sub>1-6</sub> alkyl, -SS(C<sub>1-6</sub> alkyl), -C(=O)(C<sub>1-6</sub> alkyl), -CO<sub>2</sub>H, -CO<sub>2</sub>(C<sub>1-6</sub> alkyl), -OC(=O)(C<sub>1-6</sub> alkyl), -OCO<sub>2</sub>(C<sub>1-6</sub> alkyl), -C(=O)NH<sub>2</sub>, -C(=O)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OC(=O)NH(C<sub>1-6</sub> alkyl), -NHC(=O)(C<sub>1-6</sub> alkyl), -N(C<sub>1-6</sub>

alkyl)C(=O)(C<sub>1-6</sub> alkyl), -NHCO<sub>2</sub>(C<sub>1-6</sub> alkyl), -NHC(=O)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -NHC(=O)NH(C<sub>1-6</sub> alkyl), -NHC(=O)NH<sub>2</sub>, -C(=NH)O(C<sub>1-6</sub> alkyl), -OC(=NH)(C<sub>1-6</sub> alkyl), -OC(=NH)OC<sub>1-6</sub> alkyl, -C(=NH)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -C(=NH)NH(C<sub>1-6</sub> alkyl), -C(=NH)NH<sub>2</sub>, -OC(=NH)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OC(=NH)NH(C<sub>1-6</sub> alkyl), -OC(=NH)NH<sub>2</sub>, -NHC(=NH)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -NHC(=NH)NH<sub>2</sub>, -NHSO<sub>2</sub>(C<sub>1-6</sub> alkyl), -SO<sub>2</sub>N(C<sub>1-6</sub> alkyl)<sub>2</sub>, -SO<sub>2</sub>NH(C<sub>1-6</sub> alkyl), -SO<sub>2</sub>NH<sub>2</sub>, -SO<sub>2</sub>(C<sub>1-6</sub> alkyl), -SO<sub>2</sub>O(C<sub>1-6</sub> alkyl), -OSO<sub>2</sub>(C<sub>1-6</sub> alkyl), -SO(C<sub>1-6</sub> alkyl), -Si(C<sub>1-6</sub> alkyl)<sub>3</sub>, -OSi(C<sub>1-6</sub> alkyl)<sub>3</sub> -C(=S)N(C<sub>1-6</sub> alkyl)<sub>2</sub>, C(=S)NH(C<sub>1-6</sub> alkyl), C(=S)NH<sub>2</sub>, -C(=O)S(C<sub>1-6</sub> alkyl), -C(=S)SC<sub>1-6</sub> alkyl, -SC(=S)SC<sub>1-6</sub> alkyl, -P(=O)(OC<sub>1-6</sub> alkyl)<sub>2</sub>, -P(=O)(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OP(=O)(C<sub>1-6</sub> alkyl)<sub>2</sub>, -OP(=O)(OC<sub>1-6</sub> alkyl)<sub>2</sub>, C<sub>1-6</sub> alkyl, C<sub>1-6</sub> perhaloalkyl, C<sub>2-6</sub> alkenyl, C<sub>2-6</sub> alkynyl, heteroC<sub>1-6</sub> alkyl, heteroC<sub>2-6</sub> alkenyl, heteroC<sub>2-6</sub> alkynyl, C<sub>3-10</sub> carbocyclyl, C<sub>6-10</sub> aryl, 3-10 membered heterocyclyl, 5-10 membered heteroaryl; or two geminal R<sup>gg</sup> substituents can be joined to form =O or =S; wherein X<sup>-</sup> is a counterion.

**[00277]** The term “halo” or “halogen” refers to fluorine (fluoro, -F), chlorine (chloro, -Cl), bromine (bromo, -Br), or iodine (iodo, -I).

**[00278]** The term “hydroxyl” or “hydroxy” refers to the group -OH. The term “substituted hydroxyl” or “substituted hydroxyl,” by extension, refers to a hydroxyl group wherein the oxygen atom directly attached to the parent molecule is substituted with a group other than hydrogen, and includes groups selected from -OR<sup>aa</sup>, -ON(R<sup>bb</sup>)<sub>2</sub>, -OC(=O)SR<sup>aa</sup>, -OC(=O)R<sup>aa</sup>, -OCO<sub>2</sub>R<sup>aa</sup>, -OC(=O)N(R<sup>bb</sup>)<sub>2</sub>, -OC(=NR<sup>bb</sup>)R<sup>aa</sup>, -OC(=NR<sup>bb</sup>)OR<sup>aa</sup>, -OC(=NR<sup>bb</sup>)N(R<sup>bb</sup>)<sub>2</sub>, -OS(=O)R<sup>aa</sup>, -OSO<sub>2</sub>R<sup>aa</sup>, -OSi(R<sup>aa</sup>)<sub>3</sub>, -OP(R<sup>cc</sup>)<sub>2</sub>, -OP(R<sup>cc</sup>)<sub>3</sub><sup>+</sup>X<sup>-</sup>, -OP(OR<sup>cc</sup>)<sub>2</sub>, -OP(OR<sup>cc</sup>)<sub>3</sub><sup>+</sup>X<sup>-</sup>, -OP(=O)(R<sup>aa</sup>)<sub>2</sub>, -OP(=O)(OR<sup>cc</sup>)<sub>2</sub>, and -OP(=O)(N(R<sup>bb</sup>)<sub>2</sub>)<sub>2</sub>, wherein X<sup>-</sup>, R<sup>aa</sup>, R<sup>bb</sup>, and R<sup>cc</sup> are as defined herein.

**[00279]** The term “amino” refers to the group -NH<sub>2</sub>. The term “substituted amino,” by extension, refers to a monosubstituted amino, a disubstituted amino, or a trisubstituted amino. In certain embodiments, the “substituted amino” is a monosubstituted amino or a disubstituted amino group.

**[00280]** The term “monosubstituted amino” refers to an amino group wherein the nitrogen atom directly attached to the parent molecule is substituted with one hydrogen and one group other than hydrogen, and includes groups selected from -NH(R<sup>bb</sup>), -NHC(=O)R<sup>aa</sup>, -NHCO<sub>2</sub>R<sup>aa</sup>, -NHC(=O)N(R<sup>bb</sup>)<sub>2</sub>, -NHC(=NR<sup>bb</sup>)N(R<sup>bb</sup>)<sub>2</sub>, -NHSO<sub>2</sub>R<sup>aa</sup>, -NHP(=O)(OR<sup>cc</sup>)<sub>2</sub>, and -NHP(=O)(N(R<sup>bb</sup>)<sub>2</sub>)<sub>2</sub>, wherein R<sup>aa</sup>, R<sup>bb</sup> and R<sup>cc</sup> are as defined herein, and wherein R<sup>bb</sup> of the group -NH(R<sup>bb</sup>) is not hydrogen.

**[00281]** The term “disubstituted amino” refers to an amino group wherein the nitrogen atom directly attached to the parent molecule is substituted with two groups other than

hydrogen, and includes groups selected from  $-N(R^{bb})_2$ ,  $-NR^{bb}C(=O)R^{aa}$ ,  $-NR^{bb}CO_2R^{aa}$ ,  $-NR^{bb}C(=O)N(R^{bb})_2$ ,  $-NR^{bb}C(=NR^{bb})N(R^{bb})_2$ ,  $-NR^{bb}SO_2R^{aa}$ ,  $-NR^{bb}P(=O)(OR^{cc})_2$ , and  $-NR^{bb}P(=O)(N(R^{bb})_2)_2$ , wherein  $R^{aa}$ ,  $R^{bb}$ , and  $R^{cc}$  are as defined herein, with the proviso that the nitrogen atom directly attached to the parent molecule is not substituted with hydrogen.

**[00282]** The term “trisubstituted amino” refers to an amino group wherein the nitrogen atom directly attached to the parent molecule is substituted with three groups, and includes groups selected from  $-N(R^{bb})_3$  and  $-N(R^{bb})_3^+X^-$ , wherein  $R^{bb}$  and  $X^-$  are as defined herein.

**[00283]** The term “sulfonyl” refers to a group selected from  $-SO_2N(R^{bb})_2$ ,  $-SO_2R^{aa}$ , and  $-SO_2OR^{aa}$ , wherein  $R^{aa}$  and  $R^{bb}$  are as defined herein.

**[00284]** The term “sulfinyl” refers to the group  $-S(=O)R^{aa}$ , wherein  $R^{aa}$  is as defined herein.

**[00285]** The term “acyl” refers to a group having the general formula  $-C(=O)R^{X1}$ ,  $-C(=O)OR^{X1}$ ,  $-C(=O)-O-C(=O)R^{X1}$ ,  $-C(=O)SR^{X1}$ ,  $-C(=O)N(R^{X1})_2$ ,  $-C(=S)R^{X1}$ ,  $-C(=S)N(R^{X1})_2$ ,  $-C(=S)O(R^{X1})$ ,  $-C(=S)S(R^{X1})$ ,  $-C(=NR^{X1})R^{X1}$ ,  $-C(=NR^{X1})OR^{X1}$ ,  $-C(=NR^{X1})SR^{X1}$ , and  $-C(=NR^{X1})N(R^{X1})_2$ , wherein  $R^{X1}$  is hydrogen; halogen; substituted or unsubstituted hydroxyl; substituted or unsubstituted thiol; substituted or unsubstituted amino; substituted or unsubstituted acyl, cyclic or acyclic, substituted or unsubstituted, branched or unbranched aliphatic; cyclic or acyclic, substituted or unsubstituted, branched or unbranched heteroaliphatic; cyclic or acyclic, substituted or unsubstituted, branched or unbranched alkyl; cyclic or acyclic, substituted or unsubstituted, branched or unbranched alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy, heteroalkylthioxy, arylthioxy, heteroarylthioxy, mono- or di- aliphaticamino, mono- or di- heteroaliphaticamino, mono- or di- alkylamino, mono- or di- heteroalkylamino, mono- or di-arylamino, or mono- or di-heteroarylamino; or two  $R^{X1}$  groups taken together form a 5- to 6-membered heterocyclic ring. Exemplary acyl groups include aldehydes ( $-CHO$ ), carboxylic acids ( $-CO_2H$ ), ketones, acyl halides, esters, amides, imines, carbonates, carbamates, and ureas. Acyl substituents include, but are not limited to, any of the substituents described herein, that result in the formation of a stable moiety (e.g., aliphatic, alkyl, alkenyl, alkynyl, heteroaliphatic, heterocyclic, aryl, heteroaryl, acyl, oxo, imino, thiooxo, cyano, isocyano, amino, azido, nitro, hydroxyl, thiol, halo, aliphaticamino, heteroaliphaticamino, alkylamino, heteroalkylamino, arylamino, heteroarylamino, alkylaryl, arylalkyl, aliphaticoxy, heteroaliphaticoxy, alkyloxy, heteroalkyloxy, aryloxy, heteroaryloxy, aliphaticthioxy, heteroaliphaticthioxy, alkylthioxy,

heteroalkylthioxy, arylthioxy, heteroarylthioxy, acyloxy, and the like, each of which may or may not be further substituted).

**[00286]** The term “carbonyl” refers a group wherein the carbon directly attached to the parent molecule is  $sp^2$  hybridized, and is substituted with an oxygen, nitrogen or sulfur atom, e.g., a group selected from ketones (e.g.,  $-C(=O)R^{aa}$ ), carboxylic acids (e.g.,  $-CO_2H$ ), aldehydes ( $-CHO$ ), esters (e.g.,  $-CO_2R^{aa}$ ,  $-C(=O)SR^{aa}$ ,  $-C(=S)SR^{aa}$ ), amides (e.g.,  $-C(=O)N(R^{bb})_2$ ,  $-C(=O)NR^{bb}SO_2R^{aa}$ ,  $-C(=S)N(R^{bb})_2$ ), and imines (e.g.,  $-C(=NR^{bb})R^{aa}$ ,  $-C(=NR^{bb})OR^{aa}$ ,  $-C(=NR^{bb})N(R^{bb})_2$ ), wherein  $R^{aa}$  and  $R^{bb}$  are as defined herein.

**[00287]** The term “silyl” refers to the group  $-Si(R^{aa})_3$ , wherein  $R^{aa}$  is as defined herein.

**[00288]** The term “oxo” refers to the group  $=O$ , and the term “thiooxo” refers to the group  $=S$ .

**[00289]** Nitrogen atoms can be substituted or unsubstituted as valency permits, and include primary, secondary, tertiary, and quaternary nitrogen atoms. Exemplary nitrogen atom substituents include, but are not limited to, hydrogen,  $-OH$ ,  $-OR^{aa}$ ,  $-N(R^{cc})_2$ ,  $-CN$ ,  $-C(=O)R^{aa}$ ,  $-C(=O)N(R^{cc})_2$ ,  $-CO_2R^{aa}$ ,  $-SO_2R^{aa}$ ,  $-C(=NR^{bb})R^{aa}$ ,  $-C(=NR^{cc})OR^{aa}$ ,  $-C(=NR^{cc})N(R^{cc})_2$ ,  $-SO_2N(R^{cc})_2$ ,  $-SO_2R^{cc}$ ,  $-SO_2OR^{cc}$ ,  $-SOR^{aa}$ ,  $-C(=S)N(R^{cc})_2$ ,  $-C(=O)SR^{cc}$ ,  $-C(=S)SR^{cc}$ ,  $-P(=O)(OR^{cc})_2$ ,  $-P(=O)(R^{aa})_2$ ,  $-P(=O)(N(R^{cc})_2)_2$ ,  $C_{1-10}$  alkyl,  $C_{1-10}$  perhaloalkyl,  $C_{2-10}$  alkenyl,  $C_{2-10}$  alkynyl, hetero $C_{1-10}$ alkyl, hetero $C_{2-10}$ alkenyl, hetero $C_{2-10}$ alkynyl,  $C_{3-10}$  carbocyclyl, 3-14 membered heterocyclyl,  $C_{6-14}$  aryl, and 5-14 membered heteroaryl, or two  $R^{cc}$  groups attached to an N atom are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5  $R^{dd}$  groups, and wherein  $R^{aa}$ ,  $R^{bb}$ ,  $R^{cc}$  and  $R^{dd}$  are as defined above.

**[00290]** In certain embodiments, the substituent present on the nitrogen atom is an nitrogen protecting group (also referred to herein as an “amino protecting group”). Nitrogen protecting groups include, but are not limited to,  $-OH$ ,  $-OR^{aa}$ ,  $-N(R^{cc})_2$ ,  $-C(=O)R^{aa}$ ,  $-C(=O)N(R^{cc})_2$ ,  $-CO_2R^{aa}$ ,  $-SO_2R^{aa}$ ,  $-C(=NR^{cc})R^{aa}$ ,  $-C(=NR^{cc})OR^{aa}$ ,  $-C(=NR^{cc})N(R^{cc})_2$ ,  $-SO_2N(R^{cc})_2$ ,  $-SO_2R^{cc}$ ,  $-SO_2OR^{cc}$ ,  $-SOR^{aa}$ ,  $-C(=S)N(R^{cc})_2$ ,  $-C(=O)SR^{cc}$ ,  $-C(=S)SR^{cc}$ ,  $C_{1-10}$  alkyl (e.g., aralkyl, heteroaralkyl),  $C_{2-10}$  alkenyl,  $C_{2-10}$  alkynyl, hetero $C_{1-10}$  alkyl, hetero $C_{2-10}$  alkenyl, hetero $C_{2-10}$  alkynyl,  $C_{3-10}$  carbocyclyl, 3-14 membered heterocyclyl,  $C_{6-14}$  aryl, and 5-14 membered heteroaryl groups, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aralkyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5  $R^{dd}$  groups, and wherein  $R^{aa}$ ,  $R^{bb}$ ,  $R^{cc}$  and  $R^{dd}$  are as defined herein. Nitrogen protecting groups are well known in the art and include those

described in detail in *Protecting Groups in Organic Synthesis*, T. W. Greene and P. G. M. Wuts, 3<sup>rd</sup> edition, John Wiley & Sons, 1999, incorporated herein by reference.

**[00291]** For example, nitrogen protecting groups such as amide groups (e.g.,  $-C(=O)R^{aa}$ ) include, but are not limited to, formamide, acetamide, chloroacetamide, trichloroacetamide, trifluoroacetamide, phenylacetamide, 3-phenylpropanamide, picolinamide, 3-pyridylcarboxamide, N-benzoylphenylalanyl derivative, benzamide, p-phenylbenzamide, o-nitrophenylacetamide, o-nitrophenoxyacetamide, acetoacetamide, (N'-dithiobenzoyloxyacylamino)acetamide, 3-(p-hydroxyphenyl)propanamide, 3-(o-nitrophenyl)propanamide, 2-methyl-2-(o-nitrophenoxy)propanamide, 2-methyl-2-(o-phenylazophenoxy)propanamide, 4-chlorobutanamide, 3-methyl-3-nitrobutanamide, o-nitrocinnamide, N-acetylmethionine derivative, o-nitrobenzamide and o-(benzoyloxymethyl)benzamide.

**[00292]** Nitrogen protecting groups such as carbamate groups (e.g.,  $-C(=O)OR^{aa}$ ) include, but are not limited to, methyl carbamate, ethyl carbamate, 9-fluorenylmethyl carbamate (Fmoc), 9-(2-sulfo)fluorenylmethyl carbamate, 9-(2,7-dibromo)fluorenylmethyl carbamate, 2,7-di-t-butyl-[9-(10,10-dioxo-10,10,10-tetrahydrothioxanthyl)]methyl carbamate (DBD-Tmoc), 4-methoxyphenacyl carbamate (Phenoc), 2,2,2-trichloroethyl carbamate (Troc), 2-trimethylsilylethyl carbamate (Teoc), 2-phenylethyl carbamate (hZ), 1-(1-adamantyl)-1-methylethyl carbamate (Adpoc), 1,1-dimethyl-2-haloethyl carbamate, 1,1-dimethyl-2,2-dibromoethyl carbamate (DB-t-BOC), 1,1-dimethyl-2,2,2-trichloroethyl carbamate (TCBOC), 1-methyl-1-(4-biphenyl)ethyl carbamate (Bpoc), 1-(3,5-di-t-butylphenyl)-1-methylethyl carbamate (t-Bumeoc), 2-(2'- and 4'-pyridyl)ethyl carbamate (Pyoc), 2-(N,N-dicyclohexylcarboxamido)ethyl carbamate, t-butyl carbamate (BOC or Boc), 1-adamantyl carbamate (Adoc), vinyl carbamate (Voc), allyl carbamate (Alloc), 1-isopropylallyl carbamate (Ipaoc), cinnamyl carbamate (Coc), 4-nitrocinnamyl carbamate (Noc), 8-quinolyl carbamate, N-hydroxypiperidinyl carbamate, alkylthio carbamate, benzyl carbamate (Cbz), p-methoxybenzyl carbamate (Moz), p-nitobenzyl carbamate, p-bromobenzyl carbamate, p-chlorobenzyl carbamate, 2,4-dichlorobenzyl carbamate, 4-methylsulfinylbenzyl carbamate (Msz), 9-anthrylmethyl carbamate, diphenylmethyl carbamate, 2-methylthioethyl carbamate, 2-methylsulfonylethyl carbamate, 2-(p-toluenesulfonyl)ethyl carbamate, [2-(1,3-dithianyl)]methyl carbamate (Dmoc), 4-methylthiophenyl carbamate (Mtpc), 2,4-dimethylthiophenyl carbamate (Bmpc), 2-phosphonioethyl carbamate (Peoc), 2-triphenylphosphonioisopropyl carbamate (Ppoc), 1,1-

dimethyl-2-cyanoethyl carbamate, m-chloro-p-acyloxybenzyl carbamate, p-(dihydroxyboryl)benzyl carbamate, 5-benzisoxazolylmethyl carbamate, 2-(trifluoromethyl)-6-chromonylmethyl carbamate (Tcroc), m-nitrophenyl carbamate, 3,5-dimethoxybenzyl carbamate, o-nitrobenzyl carbamate, 3,4-dimethoxy-6-nitrobenzyl carbamate, phenyl(o-nitrophenyl)methyl carbamate, t-amyl carbamate, S-benzyl thiocarbamate, p-cyanobenzyl carbamate, cyclobutyl carbamate, cyclohexyl carbamate, cyclopentyl carbamate, cyclopropylmethyl carbamate, p-decyloxybenzyl carbamate, 2,2-dimethoxyacylvinyl carbamate, o-(N,N-dimethylcarboxamido)benzyl carbamate, 1,1-dimethyl-3-(N,N-dimethylcarboxamido)propyl carbamate, 1,1-dimethylpropynyl carbamate, di(2-pyridyl)methyl carbamate, 2-furanylmethyl carbamate, 2-iodoethyl carbamate, isoborynl carbamate, isobutyl carbamate, isonicotinyl carbamate, p-(p'-methoxyphenylazo)benzyl carbamate, 1-methylcyclobutyl carbamate, 1-methylcyclohexyl carbamate, 1-methyl-1-cyclopropylmethyl carbamate, 1-methyl-1-(3,5-dimethoxyphenyl)ethyl carbamate, 1-methyl-1-(p-phenylazophenyl)ethyl carbamate, 1-methyl-1-phenylethyl carbamate, 1-methyl-1-(4-pyridyl)ethyl carbamate, phenyl carbamate, p-(phenylazo)benzyl carbamate, 2,4,6-tri-t-butylphenyl carbamate, 4-(trimethylammonium)benzyl carbamate, and 2,4,6-trimethylbenzyl carbamate.

**[00293]** Nitrogen protecting groups such as sulfonamide groups (e.g.,  $-S(=O)_2R^{aa}$ ) include, but are not limited to, p-toluenesulfonamide (Ts), benzenesulfonamide, 2,3,6-trimethyl-4-methoxybenzenesulfonamide (Mtr), 2,4,6-trimethoxybenzenesulfonamide (Mtb), 2,6-dimethyl-4-methoxybenzenesulfonamide (Pme), 2,3,5,6-tetramethyl-4-methoxybenzenesulfonamide (Mte), 4-methoxybenzenesulfonamide (Mbs), 2,4,6-trimethylbenzenesulfonamide (Mts), 2,6-dimethoxy-4-methylbenzenesulfonamide (iMDs), 2,2,5,7,8-pentamethylchroman-6-sulfonamide (Pmc), methanesulfonamide (Ms),  $\beta$ -trimethylsilylethanesulfonamide (SES), 9-anthracenesulfonamide, 4-(4',8'-dimethoxynaphthylmethyl)benzenesulfonamide (DNMBS), benzylsulfonamide, trifluoromethylsulfonamide, and phenacylsulfonamide.

**[00294]** Other nitrogen protecting groups include, but are not limited to, phenothiazinyl-(10)-acyl derivative, N'-p-toluenesulfonylaminoacyl derivative, N'-phenylaminothioacyl derivative, N-benzoylphenylalanyl derivative, N-acetylmethionine derivative, 4,5-diphenyl-3-oxazolin-2-one, N-phthalimide, N-dithiasuccinimide (Dts), N-2,3-diphenylmaleimide, N-2,5-dimethylpyrrole, N-1,1,4,4-tetramethyldisilylazacyclopentane adduct (STABASE), 5-substituted 1,3-dimethyl-1,3,5-triazacyclohexan-2-one, 5-substituted

1,3-dibenzyl-1,3,5-triazacyclohexan-2-one, 1-substituted 3,5-dinitro-4-pyridone, N-methylamine, N-allylamine, N-[2-(trimethylsilyl)ethoxy]methylamine (SEM), N-3-acetoxypyrrolamine, N-(1-isopropyl-4-nitro-2-oxo-3-pyrroline-3-yl)amine, quaternary ammonium salts, N-benzylamine, N-di(4-methoxyphenyl)methylamine, N-5-dibenzosuberylamine, N-triphenylmethylamine (Tr), N-[(4-methoxyphenyl)diphenylmethyl]amine (MMTr), N-9-phenylfluorenylamine (PhF), N-2,7-dichloro-9-fluorenylmethyleneamine, N-ferrocenylmethylamino (Fcm), N-2-picolylamino N<sup>2</sup>-oxide, N-1,1-dimethylthiomethyleneamine, N-benzylideneamine, N-p-methoxybenzylideneamine, N-diphenylmethyleneamine, N-[(2-pyridyl)mesityl]methyleneamine, N-(N',N'-dimethylaminomethylene)amine, N,N'-isopropylidenediamine, N-p-nitrobenzylideneamine, N-salicylideneamine, N-5-chlorosalicylideneamine, N-(5-chloro-2-hydroxyphenyl)phenylmethyleneamine, N-cyclohexylideneamine, N-(5,5-dimethyl-3-oxo-1-cyclohexenyl)amine, N-borane derivative, N-diphenylborinic acid derivative, N-[phenyl(pentaacylchromium- or tungsten)acyl]amine, N-copper chelate, N-zinc chelate, N-nitroamine, N-nitrosoamine, amine N-oxide, diphenylphosphinamide (Dpp), dimethylthiophosphinamide (Mpt), diphenylthiophosphinamide (Ppt), dialkyl phosphoramidates, dibenzyl phosphoramidate, diphenyl phosphoramidate, benzenesulfenamide, o-nitrobenzenesulfenamide (Nps), 2,4-dinitrobenzenesulfenamide, pentachlorobenzenesulfenamide, 2-nitro-4-methoxybenzenesulfenamide, triphenylmethylsulfenamide, and 3-nitropyridinesulfenamide (Npys). In certain embodiments, a nitrogen protecting group is benzyl (Bn), tert-butylloxycarbonyl (BOC), carbobenzyloxy (Cbz), 9-fluorenylmethyloxycarbonyl (Fmoc), trifluoroacetyl, triphenylmethyl, acetyl (Ac), benzoyl (Bz), p-methoxybenzyl (PMB), 3,4-dimethoxybenzyl (DMPM), p-methoxyphenyl (PMP), 2,2,2-trichloroethyloxycarbonyl (Troc), triphenylmethyl (Tr), tosyl (Ts), brosyl (Bs), nosyl (Ns), mesyl (Ms), triflyl (Tf), or dansyl (Ds).

**[00295]** In certain embodiments, the substituent present on an oxygen atom is an oxygen protecting group (also referred to herein as a “hydroxyl protecting group”). Oxygen protecting groups include, but are not limited to,  $-R^{aa}$ ,  $-N(R^{bb})_2$ ,  $-C(=O)SR^{aa}$ ,  $-C(=O)R^{aa}$ ,  $-CO_2R^{aa}$ ,  $-C(=O)N(R^{bb})_2$ ,  $-C(=NR^{bb})R^{aa}$ ,  $-C(=NR^{bb})OR^{aa}$ ,  $-C(=NR^{bb})N(R^{bb})_2$ ,  $-S(=O)R^{aa}$ ,  $-SO_2R^{aa}$ ,  $-Si(R^{aa})_3$ ,  $-P(R^{cc})_2$ ,  $-P(R^{cc})_3^+X^-$ ,  $-P(OR^{cc})_2$ ,  $-P(OR^{cc})_3^+X^-$ ,  $-P(=O)(R^{aa})_2$ ,  $-P(=O)(OR^{cc})_2$ , and  $-P(=O)(N(R^{bb})_2)_2$ , wherein  $X^-$ ,  $R^{aa}$ ,  $R^{bb}$ , and  $R^{cc}$  are as defined herein. Oxygen protecting groups are well known in the art and include those described in detail in

Protecting Groups in Organic Synthesis, T. W. Greene and P. G. M. Wuts, 3<sup>rd</sup> edition, John Wiley & Sons, 1999, incorporated herein by reference.

**[00296]** Exemplary oxygen protecting groups include, but are not limited to, methyl, methoxymethyl (MOM), methylthiomethyl (MTM), t-butylthiomethyl, (phenyldimethylsilyl)methoxymethyl (SMOM), benzyloxymethyl (BOM), p-methoxybenzyloxymethyl (PMBM), (4-methoxyphenoxy)methyl (p-AOM), guaiacolmethyl (GUM), t-butoxymethyl, 4-pentenylloxymethyl (POM), siloxymethyl, 2-methoxyethoxymethyl (MEM), 2,2,2-trichloroethoxymethyl, bis(2-chloroethoxy)methyl, 2-(trimethylsilyl)ethoxymethyl (SEMOR), tetrahydropyranyl (THP), 3-bromotetrahydropyranyl, tetrahydrothiopyranyl, 1-methoxycyclohexyl, 4-methoxytetrahydropyranyl (MTHP), 4-methoxytetrahydrothiopyranyl, 4-methoxytetrahydrothiopyranyl S,S-dioxide, 1-[(2-chloro-4-methyl)phenyl]-4-methoxypiperidin-4-yl (CTMP), 1,4-dioxan-2-yl, tetrahydrofuran-2-yl, tetrahydrothiofuran-2-yl, 2,3,3a,4,5,6,7,7a-octahydro-7,8,8-trimethyl-4,7-methanobenzofuran-2-yl, 1-ethoxyethyl, 1-(2-chloroethoxy)ethyl, 1-methyl-1-methoxyethyl, 1-methyl-1-benzyloxyethyl, 1-methyl-1-benzyloxy-2-fluoroethyl, 2,2,2-trichloroethyl, 2-trimethylsilylethyl, 2-(phenylselenyl)ethyl, t-butyl, allyl, p-chlorophenyl, p-methoxyphenyl, 2,4-dinitrophenyl, benzyl (Bn), p-methoxybenzyl, 3,4-dimethoxybenzyl, o-nitrobenzyl, p-nitrobenzyl, p-halobenzyl, 2,6-dichlorobenzyl, p-cyanobenzyl, p-phenylbenzyl, 2-picoyl, 4-picoyl, 3-methyl-2-picoyl N-oxido, diphenylmethyl, p,p'-dinitrobenzhydryl, 5-dibenzosuberyl, triphenylmethyl,  $\alpha$ -naphthyldiphenylmethyl, p-methoxyphenyldiphenylmethyl, di(p-methoxyphenyl)phenylmethyl, tri(p-methoxyphenyl)methyl, 4-(4'-bromophenacyloxyphenyl)diphenylmethyl, 4,4',4''-tris(4,5-dichlorophthalimidophenyl)methyl, 4,4',4''-tris(levulinoyloxyphenyl)methyl, 4,4',4''-tris(benzoyloxyphenyl)methyl, 3-(imidazol-1-yl)bis(4',4''-dimethoxyphenyl)methyl, 1,1-bis(4-methoxyphenyl)-1'-pyrenylmethyl, 9-anthryl, 9-(9-phenyl)xanthenyl, 9-(9-phenyl-10-oxo)anthryl, 1,3-benzodithiolan-2-yl, benzisothiazolyl S,S-dioxido, trimethylsilyl (TMS), triethylsilyl (TES), triisopropylsilyl (TIPS), dimethylisopropylsilyl (IPDMS), diethylisopropylsilyl (DEIPS), dimethylhexylsilyl, t-butyl dimethylsilyl (TBDMS), t-butyl diphenylsilyl (TBDPS), tribenzylsilyl, tri-p-xylylsilyl, triphenylsilyl, diphenylmethylsilyl (DPMS), t-butylmethoxyphenylsilyl (TBMPS), formate, benzoylformate, acetate, chloroacetate, dichloroacetate, trichloroacetate, trifluoroacetate, methoxyacetate, triphenylmethoxyacetate, phenoxyacetate, p-chlorophenoxyacetate, 3-phenylpropionate, 4-oxopentanoate (levulinate), 4,4-(ethylenedithio)pentanoate (levulinoyldithioacetal), pivaloate,

adamantoate, crotonate, 4-methoxycrotonate, benzoate, p-phenylbenzoate, 2,4,6-trimethylbenzoate (mesitoate), methyl carbonate, 9-fluorenylmethyl carbonate (Fmoc), ethyl carbonate, 2,2,2-trichloroethyl carbonate (Troc), 2-(trimethylsilyl)ethyl carbonate (TMSEC), 2-(phenylsulfonyl) ethyl carbonate (Psec), 2-(triphenylphosphonio) ethyl carbonate (Peoc), isobutyl carbonate, vinyl carbonate, allyl carbonate, t-butyl carbonate (BOC or Boc), p-nitrophenyl carbonate, benzyl carbonate, p-methoxybenzyl carbonate, 3,4-dimethoxybenzyl carbonate, o-nitrobenzyl carbonate, p-nitrobenzyl carbonate, S-benzyl thiocarbonate, 4-ethoxy-1-naphthyl carbonate, methyl dithiocarbonate, 2-iodobenzoate, 4-azidobutyrate, 4-nitro-4-methylpentanoate, o-(dibromomethyl)benzoate, 2-formylbenzenesulfonate, 2-(methylthiomethoxy)ethyl, 4-(methylthiomethoxy)butyrate, 2-(methylthiomethoxymethyl)benzoate, 2,6-dichloro-4-methylphenoxyacetate, 2,6-dichloro-4-(1,1,3,3-tetramethylbutyl)phenoxyacetate, 2,4-bis(1,1-dimethylpropyl)phenoxyacetate, chlorodiphenylacetate, isobutyrate, monosuccinoate, (E)-2-methyl-2-butenoate, o-(methoxyacyl)benzoate,  $\alpha$ -naphthoate, nitrate, alkyl N,N,N',N'-tetramethylphosphorodiamidate, alkyl N-phenylcarbamate, borate, dimethylphosphinothioyl, alkyl 2,4-dinitrophenylsulfenate, sulfate, methanesulfonate (mesylate), benzylsulfonate, and tosylate (Ts). In certain embodiments, an oxygen protecting group is silyl. In certain embodiments, an oxygen protecting group is t-butyl diphenylsilyl (TBDPS), t-butyl dimethylsilyl (TBDMS), triisopropylsilyl (TIPS), triphenylsilyl (TPS), triethylsilyl (TES), trimethylsilyl (TMS), triisopropylsilyloxymethyl (TOM), acetyl (Ac), benzoyl (Bz), allyl carbonate, 2,2,2-trichloroethyl carbonate (Troc), 2-trimethylsilylethyl carbonate, methoxymethyl (MOM), 1-ethoxyethyl (EE), 2-methoxy-2-propyl (MOP), 2,2,2-trichloroethoxyethyl, 2-methoxyethoxymethyl (MEM), 2-trimethylsilylethoxymethyl (SEM), methylthiomethyl (MTM), tetrahydropyranyl (THP), tetrahydrofuranyl (THF), p-methoxyphenyl (PMP), triphenylmethyl (Tr), methoxytrityl (MMT), dimethoxytrityl (DMT), allyl, p-methoxybenzyl (PMB), t-butyl, benzyl (Bn), allyl, or pivaloyl (Piv).

**[00297]** In certain embodiments, the substituent present on a sulfur atom is a sulfur protecting group (also referred to as a “thiol protecting group”). Sulfur protecting groups include, but are not limited to,  $-R^{aa}$ ,  $-N(R^{bb})_2$ ,  $-C(=O)SR^{aa}$ ,  $-C(=O)R^{aa}$ ,  $-CO_2R^{aa}$ ,  $-C(=O)N(R^{bb})_2$ ,  $-C(=NR^{bb})R^{aa}$ ,  $-C(=NR^{bb})OR^{aa}$ ,  $-C(=NR^{bb})N(R^{bb})_2$ ,  $-S(=O)R^{aa}$ ,  $-SO_2R^{aa}$ ,  $-Si(R^{aa})_3$ ,  $-P(R^{cc})_2$ ,  $-P(R^{cc})_3^+X^-$ ,  $-P(OR^{cc})_2$ ,  $-P(OR^{cc})_3^+X^-$ ,  $-P(=O)(R^{aa})_2$ ,  $-P(=O)(OR^{cc})_2$ , and  $-P(=O)(N(R^{bb})_2)_2$ , wherein  $R^{aa}$ ,  $R^{bb}$ , and  $R^{cc}$  are as defined herein. Sulfur protecting groups are well known in the art and include those described in detail in *Protecting Groups in Organic Synthesis*, T. W. Greene and P. G. M. Wuts, 3<sup>rd</sup> edition, John Wiley & Sons, 1999,

incorporated herein by reference. In certain embodiments, a sulfur protecting group is acetamidomethyl, t-butyl, 3-nitro-2-pyridine sulfenyl, 2-pyridine-sulfenyl, or triphenylmethyl.

**[00298]** A “counterion” or “anionic counterion” is a negatively charged group associated with a positively charged group in order to maintain electronic neutrality. An anionic counterion may be monovalent (*i.e.*, including one formal negative charge). An anionic counterion may also be multivalent (*i.e.*, including more than one formal negative charge), such as divalent or trivalent. Exemplary counterions include halide ions (*e.g.*, F<sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup>, I<sup>-</sup>), NO<sub>3</sub><sup>-</sup>, ClO<sub>4</sub><sup>-</sup>, OH<sup>-</sup>, H<sub>2</sub>PO<sub>4</sub><sup>-</sup>, HCO<sub>3</sub><sup>-</sup>, HSO<sub>4</sub><sup>-</sup>, sulfonate ions (*e.g.*, methanesulfonate, trifluoromethanesulfonate, *p*-toluenesulfonate, benzenesulfonate, 10-camphor sulfonate, naphthalene-2-sulfonate, naphthalene-1-sulfonic acid-5-sulfonate, ethan-1-sulfonic acid-2-sulfonate, and the like), carboxylate ions (*e.g.*, acetate, propanoate, benzoate, glycerate, lactate, tartrate, glycolate, gluconate, and the like), BF<sub>4</sub><sup>-</sup>, PF<sub>4</sub><sup>-</sup>, PF<sub>6</sub><sup>-</sup>, AsF<sub>6</sub><sup>-</sup>, SbF<sub>6</sub><sup>-</sup>, B[3,5-(CF<sub>3</sub>)<sub>2</sub>C<sub>6</sub>H<sub>3</sub>]<sub>4</sub><sup>-</sup>, B(C<sub>6</sub>F<sub>5</sub>)<sub>4</sub><sup>-</sup>, BPh<sub>4</sub><sup>-</sup>, Al(OC(CF<sub>3</sub>)<sub>3</sub>)<sub>4</sub><sup>-</sup>, and carborane anions (*e.g.*, CB<sub>11</sub>H<sub>12</sub><sup>-</sup> or (HCB<sub>11</sub>Me<sub>5</sub>Br<sub>6</sub>)<sup>-</sup>). Exemplary counterions which may be multivalent include CO<sub>3</sub><sup>2-</sup>, HPO<sub>4</sub><sup>2-</sup>, PO<sub>4</sub><sup>3-</sup>, B<sub>4</sub>O<sub>7</sub><sup>2-</sup>, SO<sub>4</sub><sup>2-</sup>, S<sub>2</sub>O<sub>3</sub><sup>2-</sup>, carboxylate anions (*e.g.*, tartrate, citrate, fumarate, maleate, malate, malonate, gluconate, succinate, glutarate, adipate, pimelate, suberate, azelate, sebacate, salicylate, phthalates, aspartate, glutamate, and the like), and carboranes.

**[00299]** The term “leaving group” is given its ordinary meaning in the art of synthetic organic chemistry and refers to an atom or a group capable of being displaced by a nucleophile. See, for example, Smith, March Advanced Organic Chemistry 6th ed. (501-502). Examples of suitable leaving groups include, but are not limited to, halogen (such as F, Cl, Br, or I (iodine)), alkoxycarbonyloxy, aryloxy carbonyloxy, alkanesulfonyloxy, arenesulfonyloxy, alkyl-carbonyloxy (*e.g.*, acetoxy), arylcarbonyloxy, aryloxy, methoxy, N,O-dimethylhydroxylamino, pixyl, and haloformates. In some cases, the leaving group is a sulfonic acid ester, such as toluenesulfonate (tosylate, -OTs), methanesulfonate (mesylate, -OMs), *p*-bromobenzenesulfonyloxy (brosylate, -OBs), -OS(=O)<sub>2</sub>(CF<sub>2</sub>)<sub>3</sub>CF<sub>3</sub> (nonaflate, -ONf), or trifluoromethanesulfonate (triflate, -OTf). In some cases, the leaving group is a brosylate, such as *p*-bromobenzenesulfonyloxy. In some cases, the leaving group is a nosylate, such as 2-nitrobenzenesulfonyloxy. The leaving group may also be a phosphineoxide (*e.g.*, formed during a Mitsunobu reaction) or an internal leaving group such as an epoxide or cyclic sulfate. Other non-limiting examples of leaving groups are water, ammonia, alcohols, ether moieties, thioether moieties, zinc halides, magnesium moieties, diazonium salts, and copper moieties. Further exemplary leaving groups include, but are not limited to, halo (*e.g.*, chloro,

bromo, iodo) and activated substituted hydroxyl groups (e.g.,  $-\text{OC}(=\text{O})\text{SR}^{\text{aa}}$ ,  $-\text{OC}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{OCO}_2\text{R}^{\text{aa}}$ ,  $-\text{OC}(=\text{O})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{OC}(=\text{NR}^{\text{bb}})\text{R}^{\text{aa}}$ ,  $-\text{OC}(=\text{NR}^{\text{bb}})\text{OR}^{\text{aa}}$ ,  $-\text{OC}(=\text{NR}^{\text{bb}})\text{N}(\text{R}^{\text{bb}})_2$ ,  $-\text{OS}(=\text{O})\text{R}^{\text{aa}}$ ,  $-\text{OSO}_2\text{R}^{\text{aa}}$ ,  $-\text{OP}(\text{R}^{\text{cc}})_2$ ,  $-\text{OP}(\text{R}^{\text{cc}})_3$ ,  $-\text{OP}(=\text{O})_2\text{R}^{\text{aa}}$ ,  $-\text{OP}(=\text{O})(\text{R}^{\text{aa}})_2$ ,  $-\text{OP}(=\text{O})(\text{OR}^{\text{cc}})_2$ ,  $-\text{OP}(=\text{O})_2\text{N}(\text{R}^{\text{bb}})_2$ , and  $-\text{OP}(=\text{O})(\text{NR}^{\text{bb}})_2$ , wherein  $\text{R}^{\text{aa}}$ ,  $\text{R}^{\text{bb}}$ , and  $\text{R}^{\text{cc}}$  are as defined herein).

**[00300]** As used herein, use of the phrase “at least one instance” refers to 1, 2, 3, 4, or more instances, but also encompasses a range, e.g., for example, from 1 to 4, from 1 to 3, from 1 to 2, from 2 to 4, from 2 to 3, or from 3 to 4 instances, inclusive.

**[00301]** A “non-hydrogen group” refers to any group that is defined for a particular variable that is not hydrogen.

**[00302]** The following definitions are more general terms used throughout the present application.

**[00303]** As used herein, the term “salt” refers to any and all salts, and encompasses pharmaceutically acceptable salts. The term “pharmaceutically acceptable salt” refers to those salts which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of humans and lower animals without undue toxicity, irritation, allergic response, and the like, and are commensurate with a reasonable benefit/risk ratio. Pharmaceutically acceptable salts are well known in the art. For example, Berge et al. describe pharmaceutically acceptable salts in detail in *J. Pharmaceutical Sciences*, 1977, 66, 1-19, incorporated herein by reference. Pharmaceutically acceptable salts of the compounds of this invention include those derived from suitable inorganic and organic acids and bases. Examples of pharmaceutically acceptable, nontoxic acid addition salts are salts of an amino group formed with inorganic acids, such as hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid, and perchloric acid or with organic acids, such as acetic acid, oxalic acid, maleic acid, tartaric acid, citric acid, succinic acid, or malonic acid or by using other methods known in the art such as ion exchange. Other pharmaceutically acceptable salts include adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate, glucoheptonate, glycerophosphate, gluconate, hemisulfate, heptanoate, hexanoate, hydroiodide, 2-hydroxy-ethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate, picrate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, p-toluenesulfonate, undecanoate, valerate salts, and the like.

Salts derived from appropriate bases include alkali metal, alkaline earth metal, ammonium, and  $N^+(C_{1-4} \text{ alkyl})_4$  salts. Representative alkali or alkaline earth metal salts include sodium, lithium, potassium, calcium, magnesium, and the like. Further pharmaceutically acceptable salts include, when appropriate, nontoxic ammonium, quaternary ammonium, and amine cations formed using counterions such as halide, hydroxide, carboxylate, sulfate, phosphate, nitrate, lower alkyl sulfonate, and aryl sulfonate.

**[00304]** It is also to be understood that compounds that have the same molecular formula but differ in the nature or sequence of bonding of their atoms or the arrangement of their atoms in space are termed “isomers”. Isomers that differ in the arrangement of their atoms in space are termed “stereoisomers”.

**[00305]** Stereoisomers that are not mirror images of one another are termed “diastereomers” and those that are non-superimposable mirror images of each other are termed “enantiomers”. When a compound has an asymmetric center, for example, it is bonded to four different groups, a pair of enantiomers is possible. An enantiomer can be characterized by the absolute configuration of its asymmetric center and is described by the R- and S-sequencing rules of Cahn and Prelog, or by the manner in which the molecule rotates the plane of polarized light and designated as dextrorotatory or levorotatory (i.e., as (+) or (-)-isomers respectively). A chiral compound can exist as either individual enantiomer or as a mixture thereof. A mixture containing equal proportions of the enantiomers is called a “racemic mixture”.

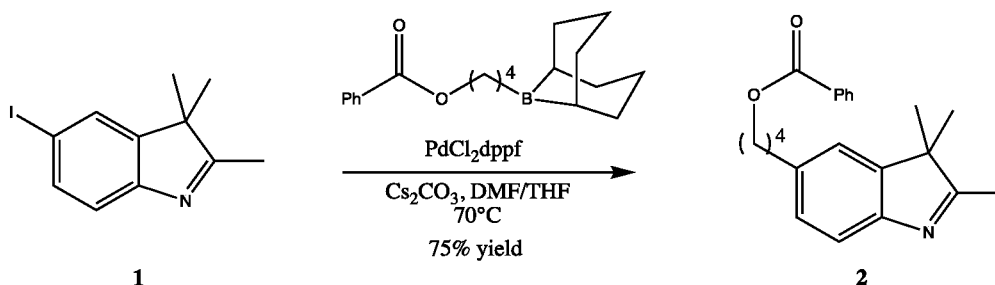
**[00306]** The term “catalysis,” “catalyze,” or “catalytic” refers to the increase in rate of a chemical reaction due to the participation of a substance called a “catalyst.” In certain embodiments, the amount and nature of a catalyst remains essentially unchanged during a reaction. In certain embodiments, a catalyst is regenerated, or the nature of a catalyst is essentially restored after a reaction. A catalyst may participate in multiple chemical transformations. The effect of a catalyst may vary due to the presence of other substances known as inhibitors or poisons (which reduce the catalytic activity) or promoters (which increase the activity). Catalyzed reactions have lower activation energy (rate-limiting free energy of activation) than the corresponding uncatalyzed reaction, resulting in a higher reaction rate at the same temperature. Catalysts may affect the reaction environment favorably, bind to the reagents to polarize bonds, form specific intermediates that are not typically produced by a uncatalyzed reaction, or cause dissociation of reagents to reactive forms.

**[00307]** The term “solvent” refers to a substance that dissolves one or more solutes, resulting in a solution. A solvent may serve as a medium for any reaction or transformation described herein. The solvent may dissolve one or more reactants or reagents in a reaction mixture. The solvent may facilitate the mixing of one or more reagents or reactants in a reaction mixture. The solvent may also serve to increase or decrease the rate of a reaction relative to the reaction in a different solvent. Solvents can be polar or non-polar, protic or aprotic. Common organic solvents useful in the methods are described herein.

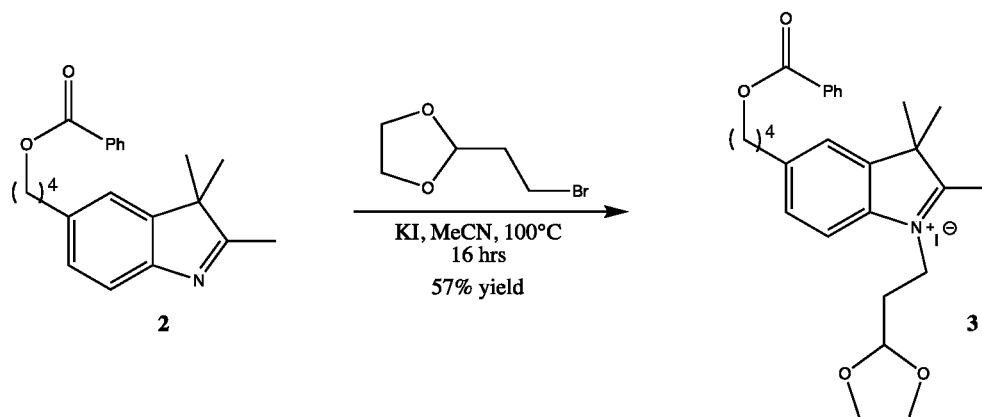
## EXAMPLES

### Synthesis of Bifunctional Labels

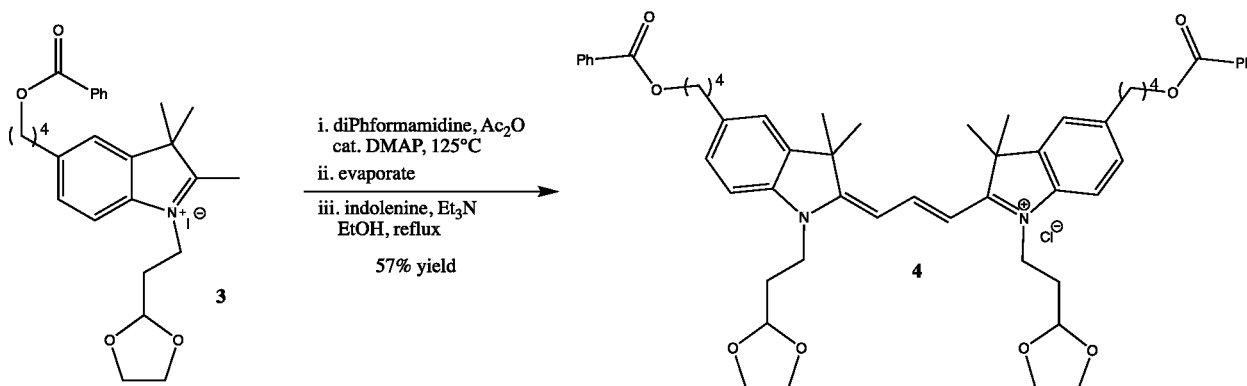
#### Indolenine benzoate **2**



**[00308]** To a flask containing 3-buten-1-ol benzoate (3.0 g, 17.1 mmol) was added a 0.5 M solution of 9-borabicyclo[3.3.1]nonane (34.1 mL, 17.1 mmol) in THF. The clear, colorless solution was stirred at room temperature for 3 hours. In a separate flask fitted with a condenser was charged iodo indolenine (**1**, 3.4 g, 11.9 mmol), [1,1'-Bis(diphenylphosphino)ferrocene]palladium(II) dichloride (0.7 g, 0.8 mmol), and cesium chloride (6.0 g, 18.5 mmol). DMF (20 mL) was charged to the flask and the dark suspension was sparged with argon for 10 minutes. After addition of the benzoate borane/THF solution the reaction was heated to 70°C for 12 hours. After this time HPLC indicated complete conversion of the starting indolenine **1**. The reaction was cooled to room temperature, diluted with EtOAc (50 mL) and hexanes (50 mL), and filtered through celite. The organic layer was washed water (3x), brine, and dried over magnesium sulfate. After filtration and evaporation the crude residue was purified by normal phase chromatography (0→50% EtOAc/hexanes, SiO<sub>2</sub>), affording indolenine **2** (3.0 g, 75% yield) as a viscous yellow oil. HRMS (ESI) calculated for C<sub>22</sub>H<sub>26</sub>NO<sub>2</sub> (M+H)<sup>+</sup> 336.1964, observed 336.1962.

Alkylated indolenine **3**

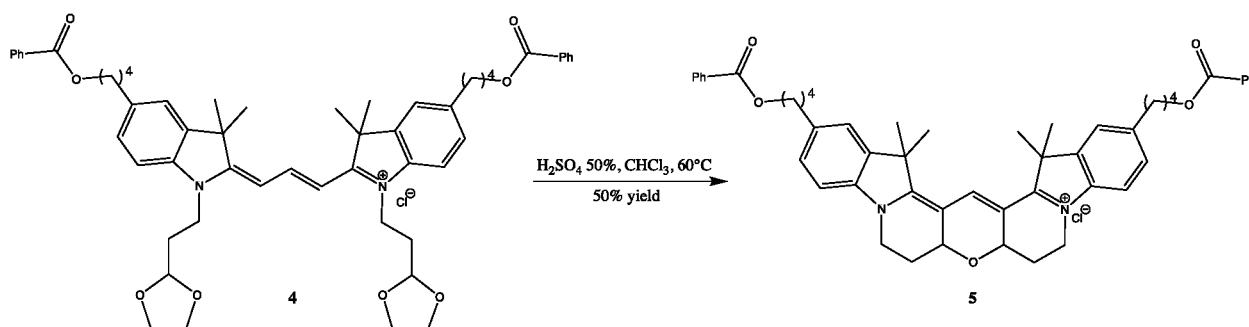
**[00309]** A schlenk flask was charged with potassium iodide (2.0 g, 12.0 mmol) and flushed with argon. Indolenine (**2**, 2.0 g, 6.0 mmol) was added as a solution in acetonitrile (8.0 mL). 2-(2-bromoethyl)-1,3-dioxolane (1.4 mL, 12.0 mmol) was added and the sealed vessel was heated to  $100^\circ\text{C}$  for 14 hours. The reaction was diluted with dichloromethane (20 mL) and the suspension was filtered through a fritted glass funnel. The filtrate was purified directly by normal phase chromatography (0→7% MeOH/DCM,  $\text{SiO}_2$ ), affording indolenine **3** (1.9 g, 57% yield) as a beige solid. HRMS (ESI) calculated for  $\text{C}_{27}\text{H}_{34}\text{NO}_4$  ( $\text{M}^+$ ) 436.2488, observed 436.2487.

Trimethine cyanine **4**

**[00310]** To an admixture of indolenine (**3**, 0.96 g, 1.7 mmol), diphenylformamidinium (0.40 g, 2.1 mmol) and DMAP (21 mg, 0.17 mmol) was added acetic anhydride (5 mL). The brown mixture was heated to  $120^\circ\text{C}$  for 1 hour. After cooling to room temperature the volatiles were concentrated in vacuo. To the crude intermediate was added an additional portion of indolenine (**3**, 1.6 g, 2.8 mmol), followed by ethanol (5 mL) and trimethylamine (1.2 mL, 8.6 mmol). The reaction was heated to reflux under argon for 1 hour. The reaction was diluted with aqueous sodium chloride and extracted with DCM. The organic layer was

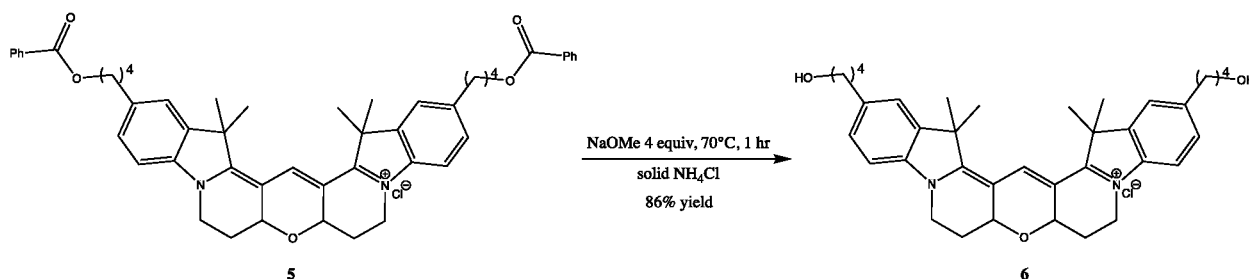
dried over magnesium sulfate, filtered and concentrated in vacuo. The crude residue was purified by normal phase chromatography (100% EtOAc, then 0→10% MeOH/DCM, SiO<sub>2</sub>), affording Cy3 **4** (0.9 g, 57% yield) as a dark purple solid. HRMS (ESI) calculated for C<sub>55</sub>H<sub>65</sub>N<sub>2</sub>O<sub>8</sub> (M)<sup>+</sup> 881.4735, observed 881.4717.

### Cy3B Analog **5**



**[00311]** To a flask containing Cy3 **4** (400 mg, 0.437 mmol) was added chloroform (6 mL) and sulfuric acid (4 mL, 50 % v/v water). The biphasic mixture was stirred vigorously at 60°C for 30 minutes, transitioning from a deep red to purple color. After cooling to room temperature the reaction was diluted with water (20 mL) and extracted with EtOAc (50 mL). The organic layer was washed with saturated aqueous sodium chloride, dried over magnesium sulfate, filtered, and concentrated in vacuo. The crude residue was purified by normal phase chromatography (100% EtOAc, then 0→15% MeOH/DCM, SiO<sub>2</sub>), affording Cy3B **5** (177 mg, 50% yield) as a dark purple solid. HRMS (ESI) calculated for C<sub>51</sub>H<sub>55</sub>N<sub>2</sub>O<sub>5</sub> (M)<sup>+</sup> 775.4105, observed 775.4091.

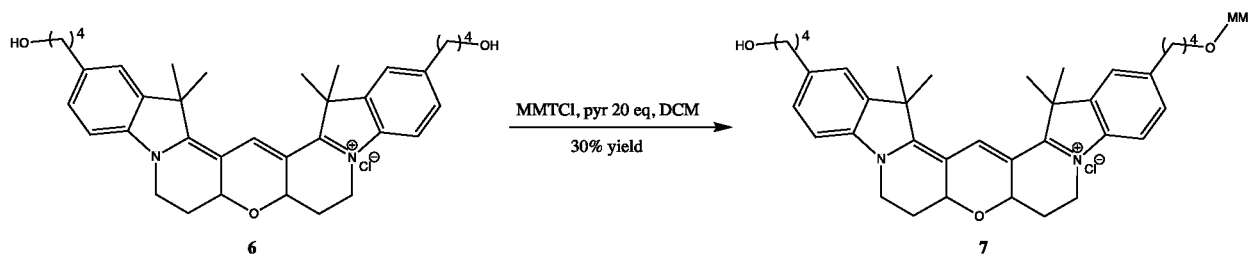
### Cy3B Diol **6**



**[00312]** To a solution of Cy3B dibenzoate **5** (150 mg, 0.185 mmol) in methanol (4 mL), was added sodium methoxide (0.37 mL, 0.74 mmol, 2.0 M in MeOH). The reaction was heated to 70°C for 1 hour. The reaction was quenched by addition of solid ammonium chloride (66 mg) and stirred for 30 minutes at room temperature. The volatiles were concentrated in vacuo and the crude re-dissolved in DCM (10 mL). The suspension was

filtered and concentrated in vacuo. The crude residue was purified by normal phase chromatography (100% EtOAc, then 0→25% MeOH/DCM, SiO<sub>2</sub>), affording Cy3B **6** (99 mg, 86% yield) as a dark purple solid. HRMS (ESI) calculated for C<sub>37</sub>H<sub>47</sub>N<sub>2</sub>O<sub>3</sub> (M)<sup>+</sup> 567.3581, observed 567.3568

### Cy3B mono-MMT **7**



To a solution of Cy3B diol **6** (100 mg, 0.166 mmol) and monomethoxytrityl chloride (62 mg, 0.20 mmol) in dichloromethane was added pyridine (0.26 mL, 3.3 mmol). The dark purple reaction was stirred at room temperature for 3

normal phase chromatography (0→2% MeOH/DCM with 1% Et<sub>3</sub>N, basic alumina). The product was redissolved in DCM (1 mL) and precipitated into hexanes (25 mL). Drying under high vacuum afforded Cy3B **8** (20 mg, 51% yield) as a dark purple solid. HRMS (ESI) calculated for C<sub>66</sub>H<sub>80</sub>N<sub>4</sub>O<sub>5</sub>P (M)<sup>+</sup> 1039.5861, observed 1039.5835.

#### EQUIVALENTS AND SCOPE

**[00314]** In the claims articles such as “a,” “an,” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

**[00315]** Furthermore, the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Where elements are presented as lists, e.g., in Markush group format, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements and/or features, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements and/or features. For purposes of simplicity, those embodiments have not been specifically set forth in haec verba herein.

**[00316]** The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements

specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

**[00317]** As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

**[00318]** As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

**[00319]** It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or

acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

**[00320]** In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03. It should be appreciated that embodiments described in this document using an open-ended transitional phrase (e.g., “comprising”) are also contemplated, in alternative embodiments, as “consisting of” and “consisting essentially of” the feature described by the open-ended transitional phrase. For example, if the disclosure describes “a composition comprising A and B,” the disclosure also contemplates the alternative embodiments “a composition consisting of A and B” and “a composition consisting essentially of A and B.”

**[00321]** Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

**[00322]** This application refers to various issued patents, published patent applications, journal articles, and other publications, all of which are incorporated herein by reference. If there is a conflict between any of the incorporated references and the instant specification, the specification shall control. In addition, any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the invention can be excluded from any claim, for any reason, whether or not related to the existence of prior art.

**[00323]** Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein. The scope of the present embodiments described herein is not intended to be limited to the above Description, but rather is as set forth in the appended claims. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made

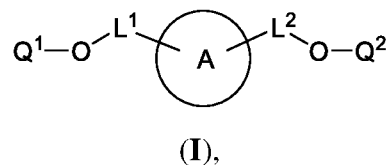
without departing from the spirit or scope of the present invention, as defined in the following claims.

**[00324]** The recitation of a listing of chemical groups in any definition of a variable herein includes definitions of that variable as any single group or combination of listed groups. The recitation of an embodiment for a variable herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

## CLAIMS

What is claimed is:

1. A labeled biomolecule of Formula (I):



wherein:

$Q^1$  and  $Q^2$  are independently monomeric or oligomeric biomolecules;

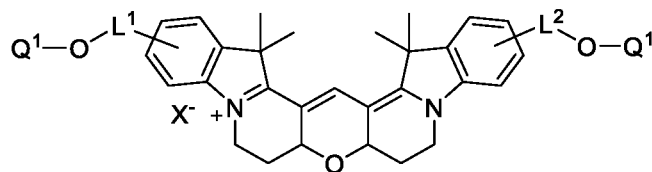
A is a polycyclic fluorophore; and

$L^1$  and  $L^2$  are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof.

2. The labeled biomolecule of claim 1, wherein A is a cyanine, fluorone, acridine, phenoxazine, coumarin, BODIPY, porphyrin, phthalocyanine, naphthalimide, or squaraine fluorophore.
3. The labeled biomolecule of claim 1 or 2, wherein  $L^1$  is optionally substituted alkylene.
4. The labeled biomolecule of claim 3, wherein  $L^1$  is unsubstituted  $C_{1-20}$  alkylene.
5. The labeled biomolecule of any one of claims 1-4, wherein  $L^2$  is optionally substituted alkylene.
6. The labeled biomolecule of claim 5, wherein  $L^2$  is unsubstituted  $C_{1-20}$  alkylene.
7. The labeled biomolecule of any one of claims 1-6, wherein A is a polycyclic cyanine.

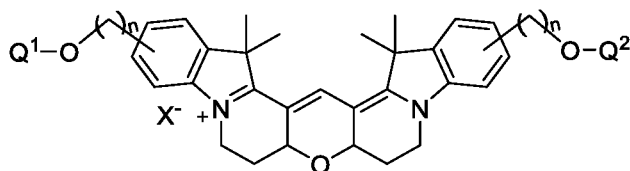
8. The labeled biomolecule of any one of claims 1-7, wherein A is Cy3B.

9. The labeled biomolecule of any one of claims 1-8, wherein the labeled biomolecule is of the formula:



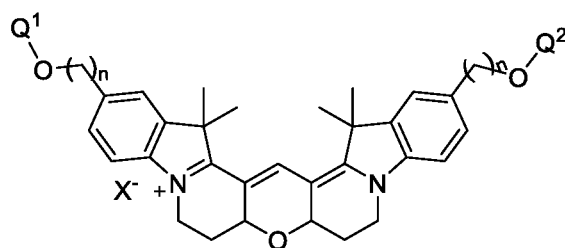
wherein  $X^-$  is a counterion or is absent.

10. The labeled biomolecule of any one of claims 1-9, wherein the labeled biomolecule is of the formula:



wherein  $n$  is independently an integer from 1-20, inclusive.

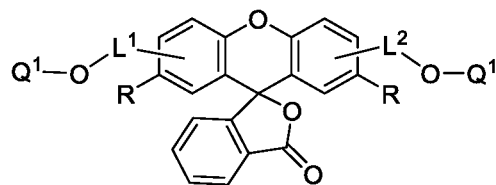
11. The labeled biomolecule of any one of claims 1-10, wherein the labeled biomolecule is of the formula:



12. The labeled biomolecule of any one of claims 1-6, wherein A is a fluorone.

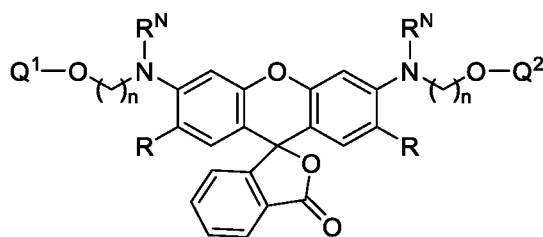
13. The labeled biomolecule of any one of claims 1-6, wherein A is rhodamine or fluorescein.

14. The labeled biomolecule of any one of claims 1-6, wherein the labeled biomolecule is of the formula:



wherein R is independently hydrogen or C<sub>1-6</sub> alkyl.

15. The labeled biomolecule of claim 14, wherein the labeled biomolecule is of the formula:

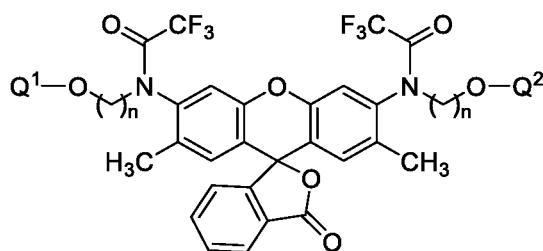


wherein:

n is independently an integer from 1-20, inclusive; and

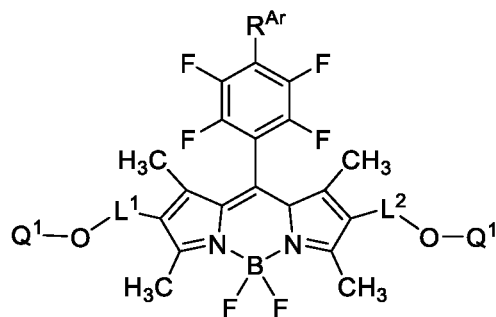
each instance of R<sup>N</sup> is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group.

16. The labeled biomolecule of claim 15, wherein the labeled biomolecule is of the formula:



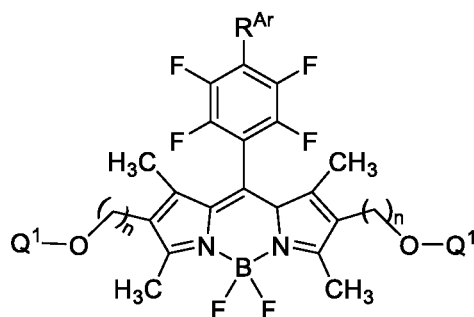
17. The labeled biomolecule of any one of claims 1-6, wherein A is a BODIPY.

18. The labeled biomolecule of claim 17, wherein the labeled biomolecule is of the formula:



wherein  $R^{Ar}$  is hydrogen, halogen, optionally substituted alkyl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ .

19. The labeled biomolecule of claim 18, wherein the labeled biomolecule is of the formula:



wherein  $R^{Ar}$  is hydrogen, halogen, optionally substituted alkyl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ .

20. The labeled biomolecule of any one of claims 1-19, wherein  $Q^1$  and  $Q^2$  are independently nucleosides, nucleotides, oligonucleotides, nucleic acids, or derivatives or fragments thereof.

21. The labeled biomolecule of claim 20, wherein  $Q^1$  and  $Q^2$  are independently deoxyribonucleic acids, ribonucleic acids, peptide nucleic acids, locked nucleic acids, or derivatives or fragments thereof.

22. The labeled biomolecule of claim 20 or 21, wherein the labeled biomolecule is a single-stranded nucleic acid comprising a first oligonucleotide strand.

23. The labeled biomolecule of claim 22, further comprising a second oligonucleotide strand hybridized to the first oligonucleotide strand.

24. The labeled biomolecule of claim 23, wherein the second oligonucleotide strand is hybridized to  $Q^1$  and  $Q^2$ .
25. The labeled biomolecule of claim 23, wherein the second oligonucleotide strand is hybridized to  $Q^1$  or  $Q^2$ .
26. The labeled biomolecule of any one of claims 1-19, wherein  $Q^1$  and  $Q^2$  are independently amino acids, oligopeptides, polypeptides, proteins, or fragments thereof.
27. The labeled biomolecule of claim 26, wherein  $Q^1$  and  $Q^2$  are joined together to form a cyclic peptide or cyclic protein.
28. The labeled biomolecule of any one of claims 1-19, wherein  $Q^1$  and  $Q^2$  are independently monosaccharides, oligosaccharides, polysaccharides, or fragments thereof.
29. The labeled biomolecule of any one of claims 1-28, wherein one or more emission characteristics of the labeled biomolecule are increased relative to an unconjugated molecule comprising the polycyclic fluorophore.
30. The labeled biomolecule of claim 29, wherein the one or more emission characteristics are selected from the group consisting of luminescence lifetime, luminescence intensity, luminescence quantum yield, brightness, and emission maximum.
31. The labeled biomolecule of claim 30, wherein luminescence lifetime of the labeled biomolecule is increased by at least 10% relative to the unconjugated molecule.
32. The labeled biomolecule of claim 31, wherein luminescence lifetime of the labeled biomolecule is increased by between approximately 10% and 25% relative to the unconjugated molecule.
33. The labeled biomolecule of claim 31, wherein luminescence lifetime of the labeled biomolecule is increased by between approximately 25% and 50% relative to the unconjugated molecule.

34. The labeled biomolecule of claim 30, wherein emission maximum of the labeled biomolecule is increased by at least 1% relative to the unconjugated molecule.
35. The labeled biomolecule of claim 34, wherein emission maximum of the labeled biomolecule is increased by between approximately 1% and 10% relative to the unconjugated molecule.
36. The labeled biomolecule of any one of claims 29-35, wherein the unconjugated molecule does not comprise one or either of  $Q^1$  and  $Q^2$ .
37. The labeled biomolecule of any one of claims 1-36, wherein  $Q^1$  and  $Q^2$  are independently optionally associated with one or more fluorophores other than the polycyclic fluorophore.
38. The labeled biomolecule of claim 37, wherein each fluorophore is at least 5 angstroms separated from any other fluorophore.
39. The labeled biomolecule of any one of claims 1-38, wherein  $Q^1$  and  $Q^2$  are independently optionally associated with a reactant configured for use as a substrate in a reaction.
40. The labeled biomolecule of any one of claims 23-25, wherein the first and second oligonucleotide strands are independently optionally associated with a reactant configured for use as a substrate in a reaction.
41. The labeled biomolecule of claim 39 or 40, wherein the reaction is a polymerization reaction.
42. The labeled biomolecule of any one claims 39-41, wherein the reactant is cleaved from the labeled biomolecule by a polymerase when subjected to polymerization reaction conditions.
43. A labeled nucleotide comprising one or more nucleotides associated with a labeled biomolecule of any one of claims 1-42.

44. The labeled nucleotide of claim 43, wherein the one or more nucleotides comprise one type of nucleotide selected from guanine, cytosine, adenine, and thymine or uracil.
45. The labeled nucleotide of claim 43 or 44, wherein the one or more nucleotides are cleaved from the labeled biomolecule by a polymerase when subjected to polymerization reaction conditions.
46. The labeled nucleotide of any one of claims 43-45, wherein the one or more nucleotides comprise nucleoside triphosphates.
47. The labeled nucleotide of any one of claims 43-46, wherein the one or more nucleotides comprise nucleoside hexaphosphates.
48. The labeled nucleotide of claim 46 or 47, wherein the one or more nucleotides are attached through a terminal phosphate to the labeled biomolecule.
49. A composition comprising a labeled nucleotide of any one of claims 43-48.
50. A nucleic acid sequencing reaction composition comprising two or more different types of labeled nucleotides in a reaction mixture, wherein at least one type of labeled nucleotide is a labeled nucleotide according to any one of claims 43-48.
51. The nucleic acid sequencing reaction composition of claim 50 comprising four different types of labeled nucleotides.
52. The nucleic acid sequencing reaction composition of claim 51, wherein the four different types of labeled nucleotides comprise a first labeled nucleotide comprising guanine, a second labeled nucleotide comprising cytosine, a third labeled nucleotide comprising adenine, and a fourth labeled nucleotide comprising thymine or uracil.
53. A method of determining a sequence of a template nucleic acid, the method comprising:

(i) exposing a complex in a target volume, the complex comprising the template nucleic acid, a primer, and a polymerizing enzyme, to a nucleic acid sequencing reaction composition according to any one of claims 50-52;

(ii) directing a series of pulses of one or more excitation energies towards a vicinity of the target volume;

(iii) detecting a plurality of emitted photons from luminescently labeled nucleotides during sequential incorporation into a nucleic acid comprising the primer; and

(iv) identifying the sequence of incorporated nucleotides by determining timing and optionally luminescence intensity of the emitted photons.

54. A kit for sequencing a template nucleic acid, the kit comprising:

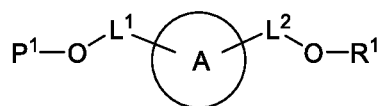
two or more different types of labeled nucleotides, wherein at least one type of labeled nucleotide comprises a labeled nucleotide according to any one of claims 43-48.

55. The kit of claim 54 comprising four different types of labeled nucleotides.

56. The kit of claim 54 or 55 further comprising a polymerizing enzyme.

57. The kit of any one of claims 54-56 further comprising a primer complementary to the template nucleic acid.

58. A compound of Formula (II):



(II),

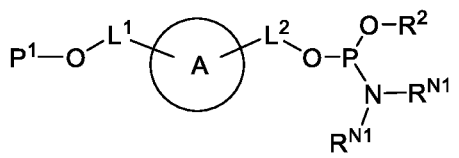
or a salt thereof, wherein:

A is a polycyclic fluorophore;

$L^1$  and  $L^2$  are independently linkers selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof;

$P^1$  is an oxygen protecting group; and  
 $R^1$  is a reactive moiety.

59. The compound of claim 58, wherein  $R^1$  is a phosphoramidite group.
60. The compound of claim 58 or 59, wherein  $L^1$  is optionally substituted alkylene.
61. The compound of claim 60, wherein  $L^1$  is unsubstituted  $C_{1-20}$  alkylene.
62. The compound of any one of claims 58-61, wherein  $L^2$  is optionally substituted alkylene.
63. The compound of claim 62, wherein  $L^2$  is unsubstituted  $C_{1-20}$  alkylene.
64. The compound of any one of claims 58-63, wherein  $P^1$  is an optionally substituted triphenylmethyl (trityl) protecting group.
65. The compound of claim 64, wherein  $P^1$  is 4-methoxytriphenylmethyl (MMT).
66. The compound of any one of claims 58-65, wherein the compound is of the formula:



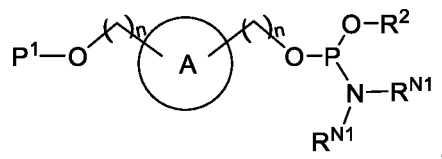
or a salt thereof, wherein:

$R^2$  is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or an oxygen protecting group;

each instance of  $R^{N1}$  is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group; optionally

wherein two R<sup>N1</sup> bonded to the same nitrogen are joined together with the intervening atoms to form optionally substituted heterocyclyl or optionally substituted heteroaryl.

67. The compound of any one of claims 58-66, wherein the compound is of the formula:



or a salt thereof, wherein:

each instance of n is an integer from 1-20, inclusive.

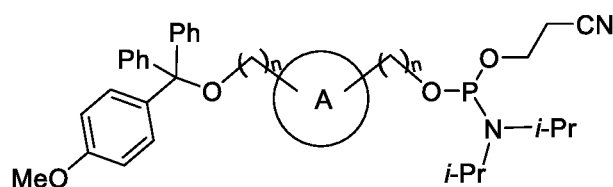
68. The compound of claim 66 or 67, wherein R<sup>2</sup> is optionally substituted C<sub>1</sub>-C<sub>6</sub> alkyl.

69. The compound of claim 68, wherein R<sup>2</sup> is:

70. The compound of any one of claims 66-69, wherein each R<sup>N</sup> is optionally substituted C<sub>1-6</sub> alkyl.

71. The compound of claim 70, wherein each R<sup>N</sup> is iso-propyl.

72. The compound of any one of claims 58-71, wherein the compound is of the formula:



or a salt thereof.

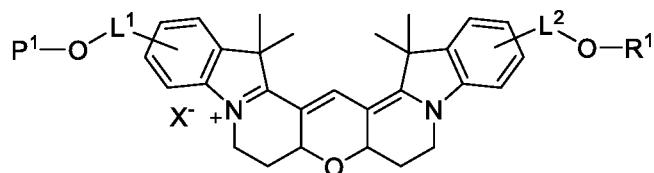
73. The compound of claim 72, wherein each n is 4.

74. The compound of any one of claims 58-73, wherein A is a cyanine, fluorone, acridine, phenoxazine, coumarin, BODIPY, or squaraine fluorophore.

75. The compound of any one of claims 58-74, wherein A is a polycyclic cyanine.

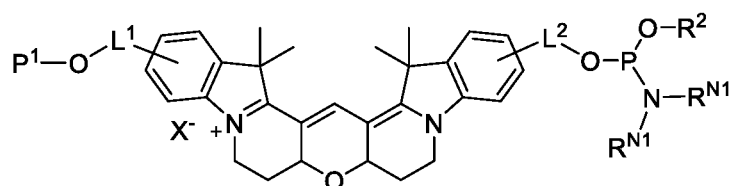
76. The compound of any one of claims 58-75, wherein A is Cy3B.

77. The compound of any one of claims 58-76, wherein the compound is of the formula:



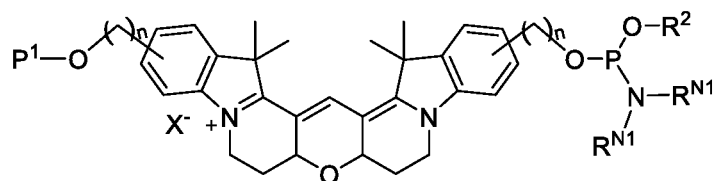
or a salt thereof, wherein  $X^-$  is a counterion or is absent.

78. The compound of any one of claims 58-77, wherein the compound is of the formula:



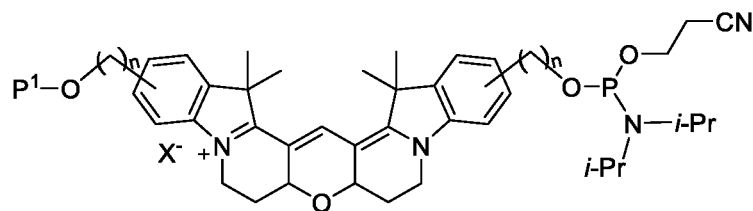
or a salt thereof.

79. The compound of any one of claims 58-78, wherein the compound is of the formula:



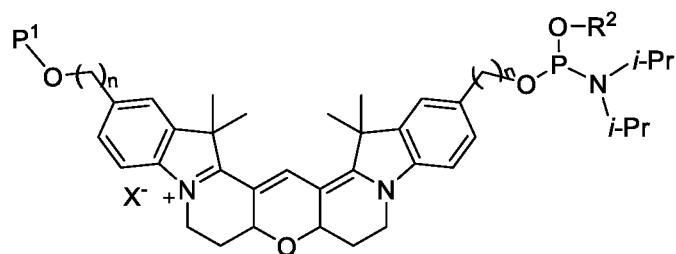
or a salt thereof.

80. The compound of any one of claims 58-79, wherein the compound is of the formula:



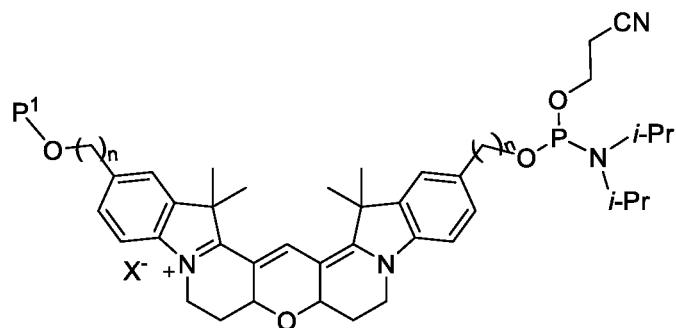
or a salt thereof.

81. The compound of any one of claims 58-80, wherein the compound is of the formula:



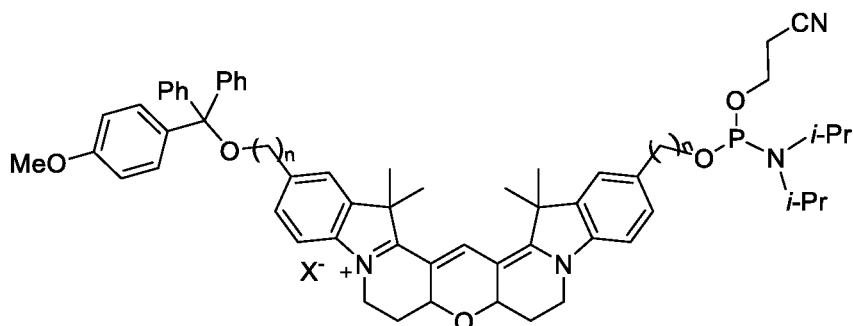
or a salt thereof.

82. The compound of any one of claims 58-81, wherein the compound is of the formula:



or a salt thereof.

83. The compound of any one of claims 58-82, wherein the compound is of the formula:

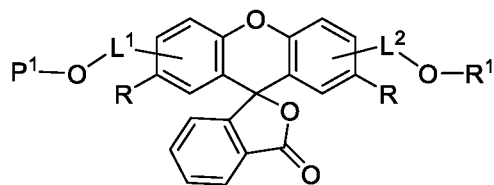


or a salt thereof.

84. The compound of any one of claims 58-74, wherein A is a fluorone.

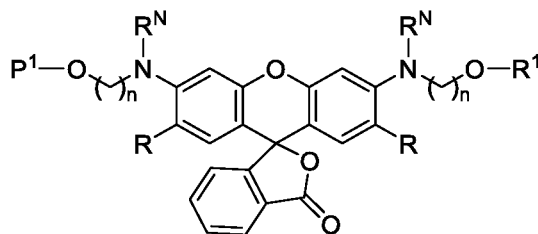
85. The compound of any one of claims 58-74, wherein A is rhodamine or fluorescein.

86. The compound of any one of claims 58-74, wherein the compound is of the formula:



wherein R is independently hydrogen or C<sub>1-6</sub> alkyl.

87. The compound of claim 86, wherein the compound is of the formula:

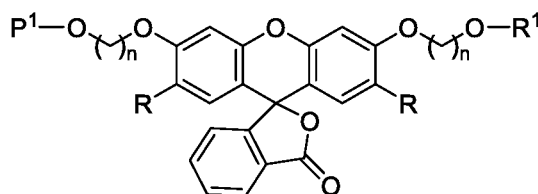


wherein:

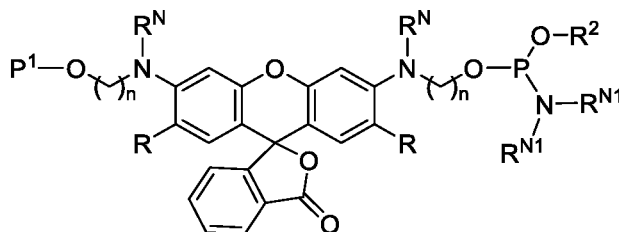
n is independently an integer from 1-20, inclusive; and

each instance of R<sup>N</sup> is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, optionally substituted acyl, or a nitrogen protecting group.

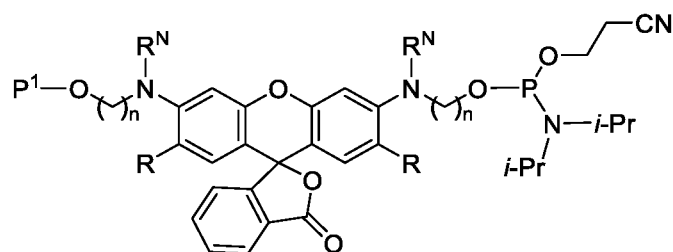
88. The compound of claim 87, wherein the compound is of the formula:



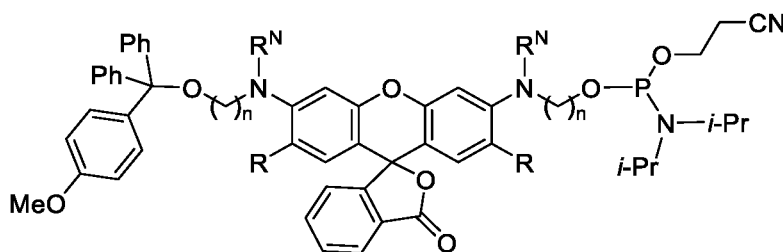
89. The compound of claim 87, wherein the compound is of the formula:



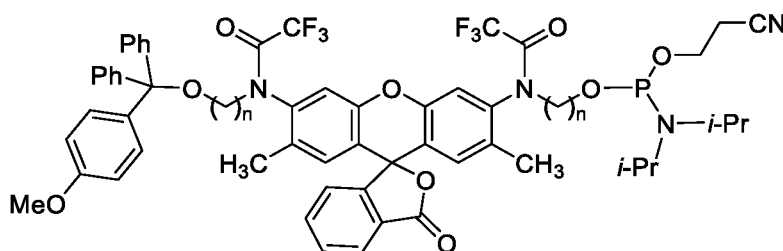
90. The compound of claim 89, wherein the compound is of the formula:



91. The compound of claim 90, wherein the compound is of the formula:

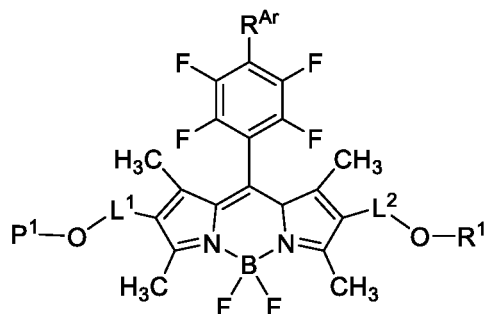


92. The compound of claim 91, wherein the compound is of the formula:



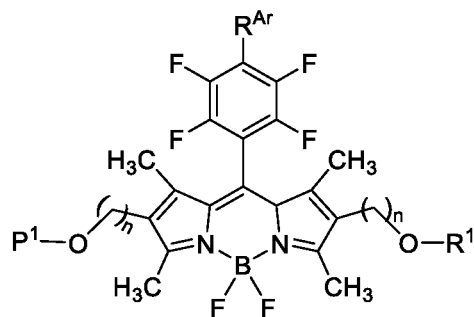
93. The compound of any one of claims 58-74, wherein A is BODIPY.

94. The compound of claim 93, wherein the compound is of the formula:



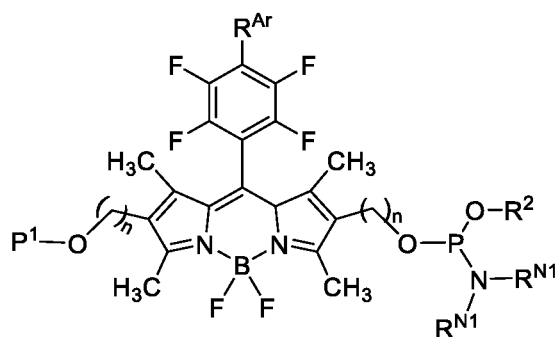
wherein  $R^{Ar}$  is hydrogen, halogen, optionally substituted alkyl,  $-OR^O$ ,  $-SR^S$ , or  $N(R^N)_2$ .

95. The compound of claim 94, wherein the compound is of the formula:

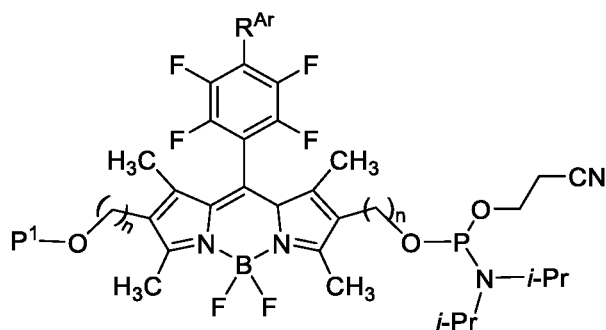


wherein n is independently an integer from 1-20, inclusive.

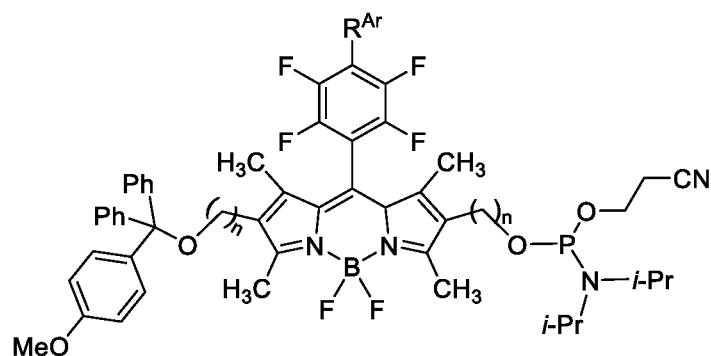
96. The compound of claim 95, wherein the compound is of the formula:



97. The compound of claim 96, wherein the compound is of the formula:



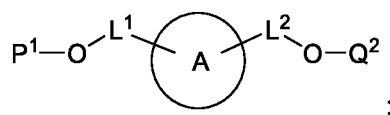
98. The compound of claim 97, wherein the compound is of the formula:



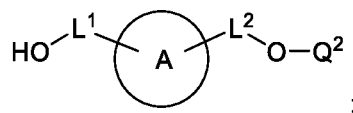
99. A composition comprising a compound of any one of claims 58-98.

100. A method for preparing a labeled biomolecule of any one of claims 1-42, the method comprising:

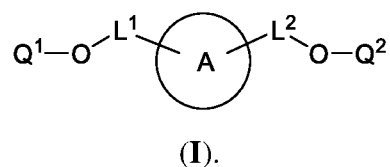
(i) contacting a monomeric or oligomeric biomolecule of formula  $Q^2-OH$ , or a salt thereof, with a compound of any one of claims 58-98, or a salt thereof, under conditions sufficient to promote conjugation to yield a conjugate of the formula:



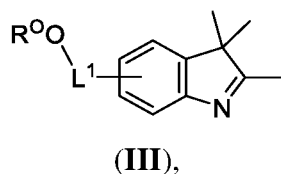
(ii) deprotecting the conjugate formed in step (i) under conditions sufficient to cleave the  $P^1$  protecting group and yield a conjugate of the formula:



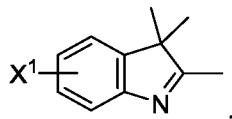
(iii) contacting the conjugate formed in step (ii) with a monomeric or oligomeric biomolecule of formula  $Q^1-O-R^1$ , or a salt thereof, under conditions sufficient to promote conjugation to yield a labeled biomolecule of Formula (I):



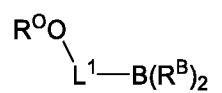
101. A method for preparing a compound of Formula (III):



or a salt thereof, the method comprising coupling a compound of the formula:



or a salt thereof, with a compound of the formula:



or a salt thereof, in the presence of palladium to yield a compound of Formula (III), or a salt thereof, wherein:

$X^1$  is halogen or a leaving group;

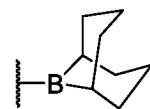
$B(R^B)_2$  is a borane, boronic acid, or boronic ester;

$L^1$  is a linker selected from the group consisting of optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, optionally substituted heteroarylene, and combinations thereof; and

$R^O$  is an oxygen protecting group.

102. The method of claim 101, wherein  $B(R^B)_2$  is borane.

103. The method of claim 102, wherein  $B(R^B)_2$  is of the formula:



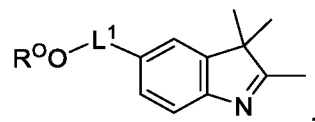
104. The method of any one of claims 101-103, wherein  $L^1$  is optionally substituted alkylene.

105. The method of claim 104, wherein  $L^1$  is unsubstituted  $C_{1-20}$  alkylene.

106. The method of any one of claims 101-105, wherein  $X^1$  is a halogen.

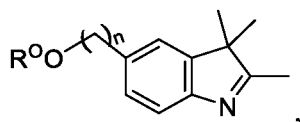
107. The method of claim 106, wherein  $X^1$  is  $-I$ .

108. The method of any one of claims 101-107, wherein the compound of Formula (III) is of the formula:



or a salt thereof.

109. The method of claim 108, wherein the compound of Formula (III) is of the formula:



or a salt thereof.

110. The method of any one of claims 101-109, wherein the reaction is carried out in the presence of a palladium(II) complex.

111. The method of claim 110, wherein the palladium complex is PdCl<sub>2</sub>(dppf).

112. The method of any one of claims 101-111, wherein the palladium is present in a catalytic amount.

113. The method of any one of claims 101-111, wherein the reaction is carried out in the presence of a base.

114. The method of claim 113, wherein the base is a carbonate base.

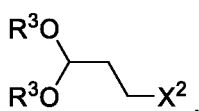
115. The method of claim 114, wherein the base is Cs<sub>2</sub>CO<sub>3</sub>.

116. The method of any one of claims 101-115, wherein the reaction is carried out in a solvent.

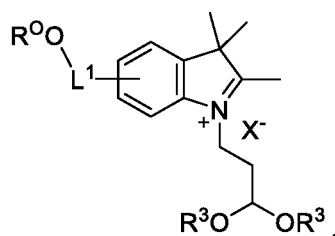
117. The method of claim 116, wherein the solvent is THF, DMF, or a mixture thereof.

118. The method of any one of claims 101-117, wherein the reaction is carried out at elevated temperature.

119. The method of any one of claims 101-118, further comprising a step of alkylating a compound of Formula (III) with a compound of the formula:



or a salt thereof, to yield a compound of the formula:



or a salt or tautomer thereof, wherein:

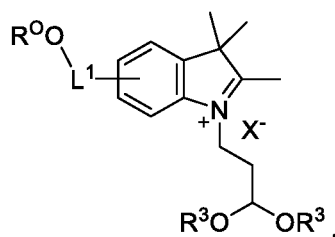
$X^2$  is halogen or a leaving group; and

each  $R^3$  is independently optionally substituted alkyl, optionally substituted acyl, or an oxygen protecting group; or optionally two  $R^3$  are joined together with the intervening atoms to form optionally substituted heterocyclyl.

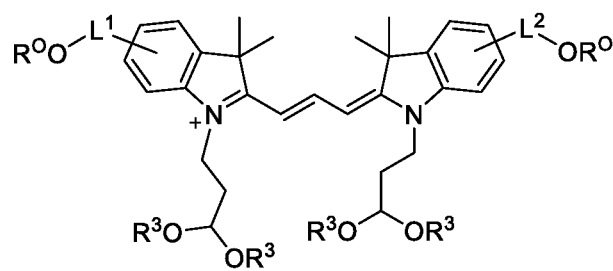
120. The method of claim 119, wherein the reaction is carried out in the presence of a base.

121. The method of claim 120, wherein the base is KI.

122. The method of any one of claims 118-120 further comprising reacting the compound of formula:



or a salt or tautomer thereof, in the presence of a formamidine, to yield a compound of formula:



or a salt or tautomer thereof.

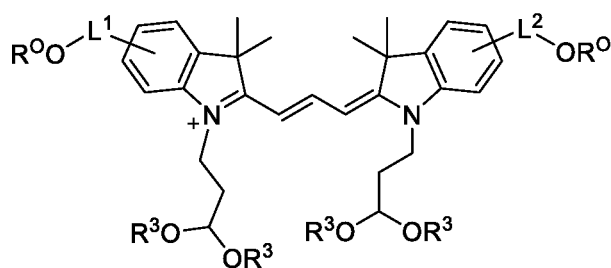
123. The method of claim 122, wherein the formamidine is diphenylformamidine.

124 The method of claim 122 or 123, wherein the reaction is carried out in the presence of a base.

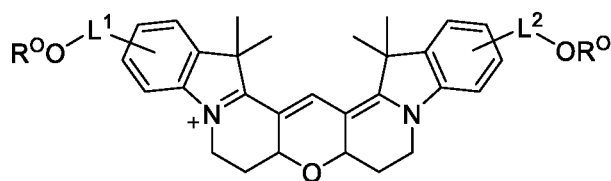
125. The method of claim 124, wherein the base is DMAP and/or Et<sub>3</sub>N.

126. The method of any one of claims 122-125, wherein the reaction is carried out in the presence of Ac<sub>2</sub>O.

127. The method of any one of claims 122-126 further comprising cyclizing a compound of formula:



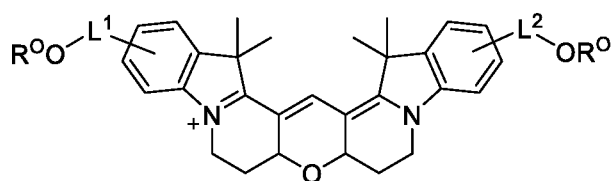
or a salt thereof, in the presence of an acid, to yield a compound of the formula:



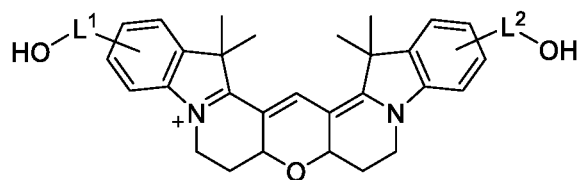
or a salt thereof.

128. The method of claim 127, wherein the acid is sulfuric acid.

129. The method of claim 127 or 128 further comprising a step of deprotecting a compound of the formula:



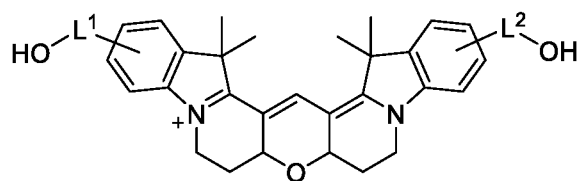
or a salt thereof, to yield a compound of the formula:



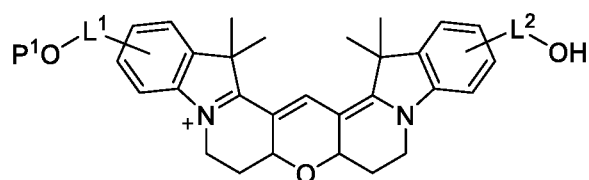
or a salt thereof.

130. The method of claim 129 further comprising the steps of:

(i) protecting a compound of the formula:

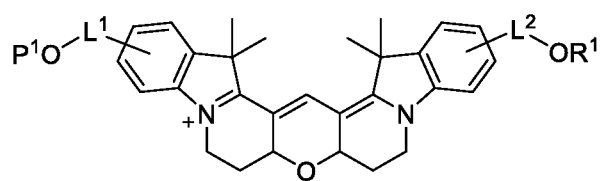


or a salt thereof, to yield a compound of the formula:



or a salt thereof; and

(ii) reacting the compound produced in step (i) under conditions sufficient to form a compound of the formula:



or a salt thereof.

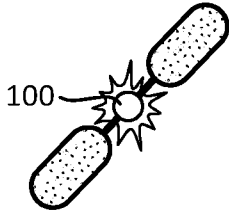


FIG. 1A

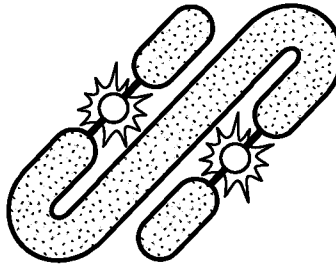


FIG. 1B

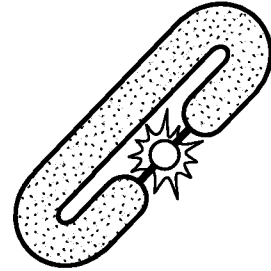


FIG. 1C

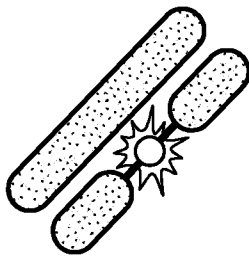


FIG. 1D

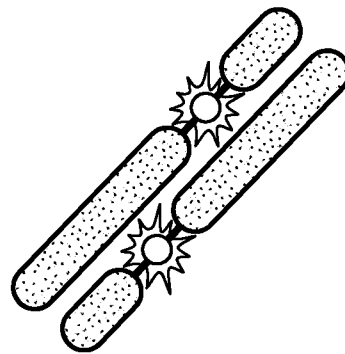


FIG. 1E

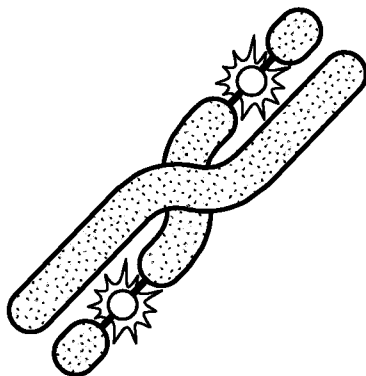


FIG. 1F

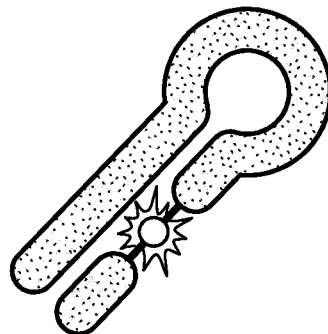


FIG. 1G

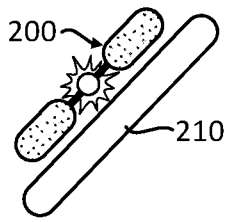


FIG. 2A

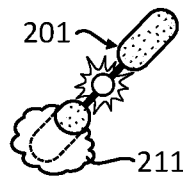


FIG. 2B

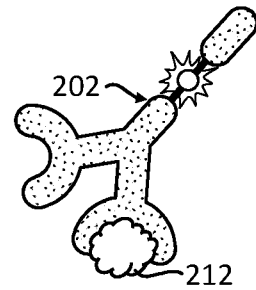
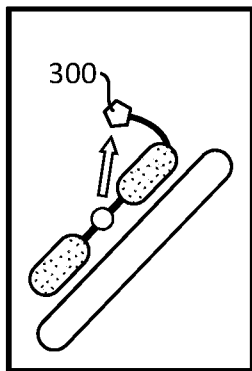
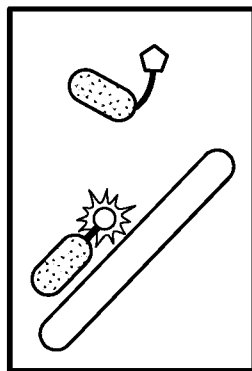


FIG. 2C

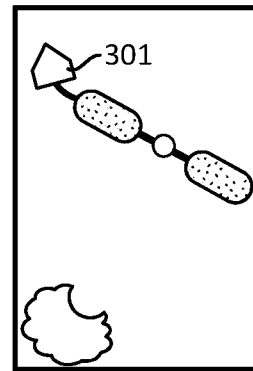


I

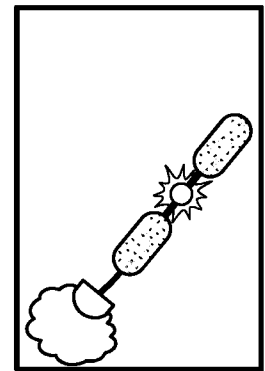


II

FIG. 3A



I



II

FIG. 3B

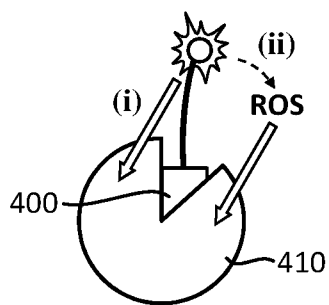


FIG. 4A

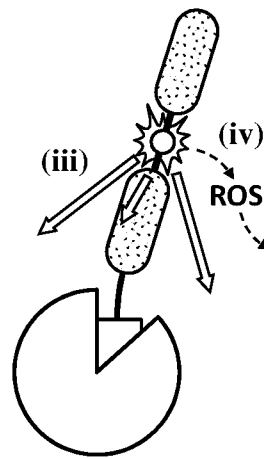


FIG. 4B

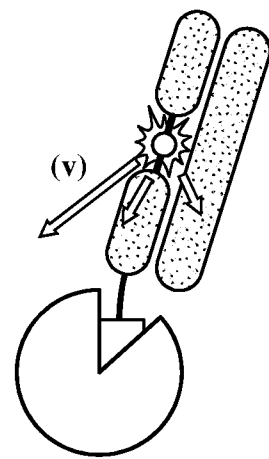


FIG. 4C

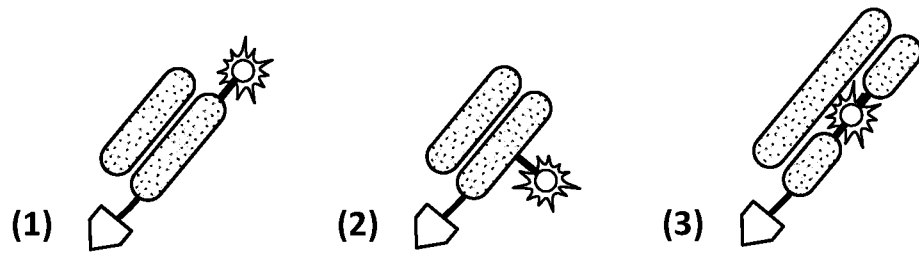


FIG. 5A

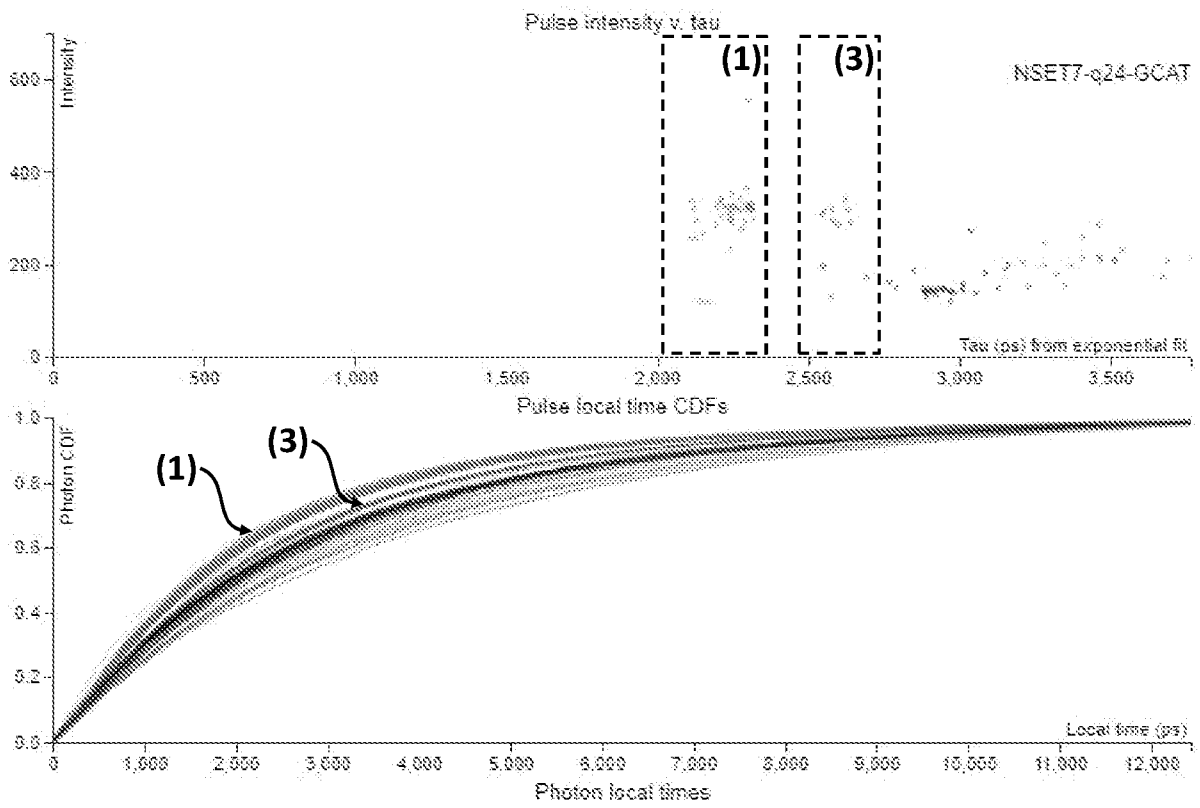


FIG. 5B

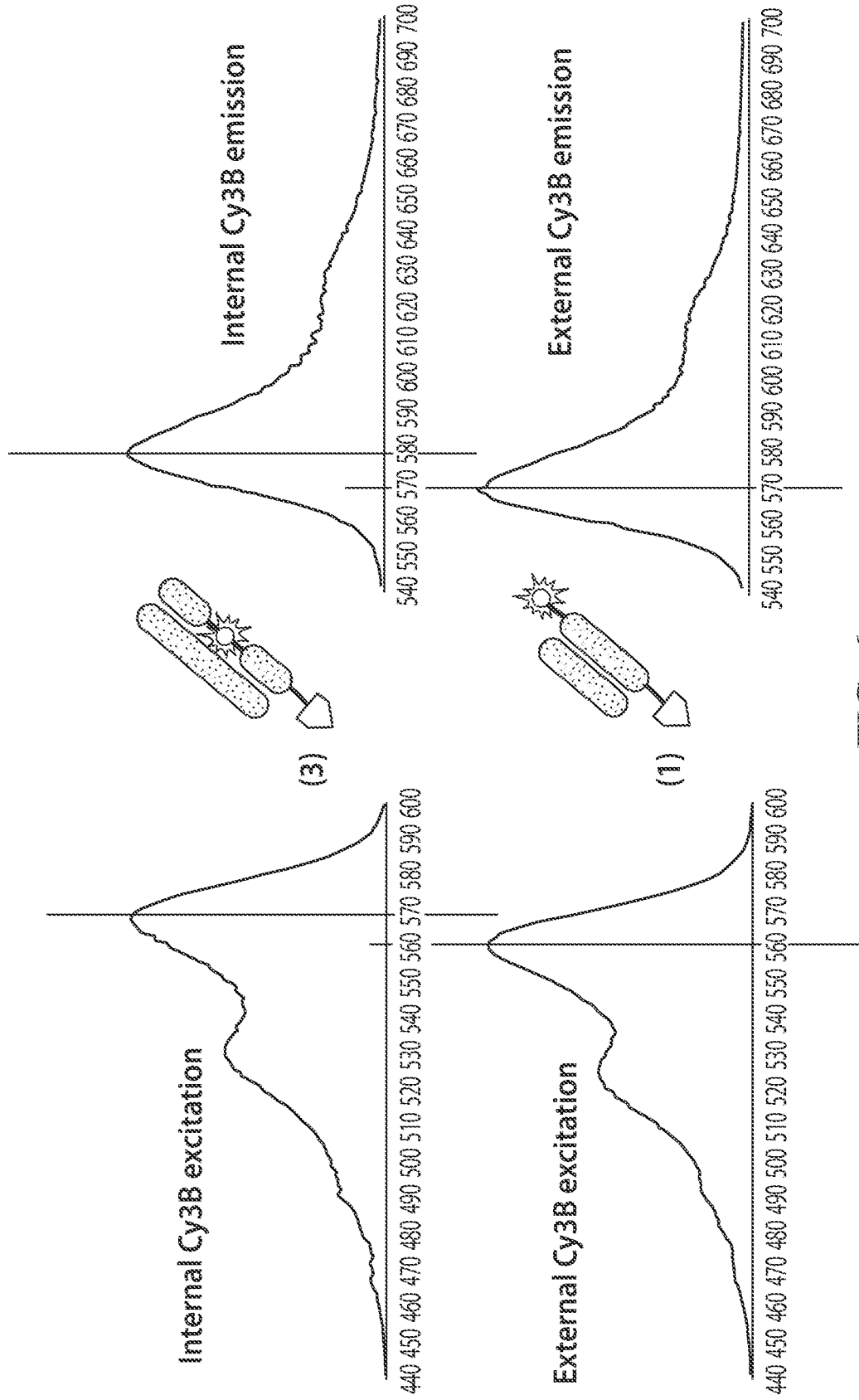


FIG. 6