Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3224572 A1 2020/08/27

(21) 3 224 572

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(22) **Date de dépôt/Filing Date:** 2020/02/18

(41) **Mise à la disp. pub./Open to Public Insp.:** 2020/08/27

(62) **Demande originale/Original Application:** 3 130 693

(30) **Priorité/Priority:** 2019/02/19 (US62/807,550)

(51) **Cl.Int./Int.Cl.** ***C12Q 1/6869*** (2018.01),
***C07H 19/04*** (2006.01), ***C07K 5/02*** (2006.01),
***C12Q 1/68*** (2018.01), ***C12Q 1/6844*** (2018.01)

(71) **Demandeur/Applicant:**
ULTIMA GENOMICS, INC., US

(72) **Inventeurs/Inventors:**
LEE, LINDA G., US;
ALMOGY, GILAD, US;
MENCHEN, STEVEN, US

(74) **Agent:** GOWLING WLG (CANADA) LLP

(54) **Titre : LIEURS ET PROCEDES DE DETECTION ET DE SEQUENCAGE OPTIQUES**
(54) **Title: LINKERS AND METHODS FOR OPTICAL DETECTION AND SEQUENCING**

(57) **Abrégé/Abstract:**
The present disclosure provides labeling reagents for labeling substrates such as nucleotides, proteins, antibodies, lipids, and cells. The labeling reagents provided herein may comprise fluorescent labels and semi-rigid linkers. Methods for nucleic acid sequencing using materials comprising such labeling reagents are also provided herein.

Abstract

The present disclosure provides labeling reagents for labeling substrates such as nucleotides, proteins, antibodies, lipids, and cells. The labeling reagents provided herein may comprise fluorescent labels and semi-rigid linkers. Methods for nucleic acid sequencing using materials comprising such labeling reagents are also provided herein.

Date Reçue/Date Received 2023-12-22

# LINKERS AND METHODS FOR OPTICAL DETECTION AND SEQUENCING

[0001]

## BACKGROUND

[0002] The detection, quantification and sequencing of cells and biological molecules may be important for molecular biology and medical applications, such as diagnostics. Genetic testing may be useful for a number of diagnostic methods. For example, disorders that are caused by rare genetic alterations (e.g., sequence variants) or changes in epigenetic markers, such as cancer and partial or complete aneuploidy, may be detected or more accurately characterized with deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) sequence information.

[0003] Nucleic acid sequencing is a process that can be used to provide sequence information for a nucleic acid sample. Such sequence information may be helpful in diagnosing and/or treating a subject with a condition. For example, the nucleic acid sequence of a subject may be used to identify, diagnose and potentially develop treatments for genetic diseases. As another example, research into pathogens may lead to treatment of contagious diseases.

[0004] Nucleic acid sequencing may comprise the use of fluorescently labeled moieties. Such moieties may be labeled with organic fluorescent dyes. The sensitivity of a detection scheme can be improved by using dyes with both a high extinction coefficient and quantum yield, where the product of these characteristics may be termed the dye's "brightness." Dye brightness may be attenuated by quenching phenomena, including quenching by biological materials, quenching by proximity to other dyes, and quenching by solvent. Other routes to brightness loss include photobleaching, reactivity to molecular oxygen, and chemical decomposition.

## SUMMARY

[0005] The present disclosure provides improved optical (e.g., fluorescent) labeling reagents and methods of nucleic acid processing comprising the use of optically (e.g., fluorescently) labeled moieties. The materials and methods provided herein may comprise the use of organic fluorescent dyes. The materials provided herein may allow for optimized molecular quenching to facilitate efficient nucleic acid processing and detection. Molecular quenching mechanisms can include photoinduced electron transfer, photoinduced hole transfer, Forster energy transfer, Dexter quenching, and the like. A general solution to many types of quenching requires physical separation of the dye from the quencher moiety, but existing solutions all have advantages and

disadvantages in terms of ease of use, cost, solvent-dependence and polydispersity. Accordingly, the present disclosure recognizes the need for materials and methods that address these limitations and provides materials comprising improved linker moieties.

**[0006]** In an aspect, the present disclosure provides a fluorescent labeling reagent comprising: (a) a fluorescent dye; and (b) a linker that is connected to the fluorescent dye and configured to couple to a substrate for fluorescently labelling the substrate, wherein the linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two atoms, and wherein the linker comprises a non-proteinogenic amino acid comprising a ring system of the two or more ring systems.

**[0007]** In some embodiments, the fluorescent labeling reagent coupled to the substrate is configured to emit a fluorescent signal.

**[0008]** In some embodiments, the linker is configured to establish a functional length of at least about 0.5 nanometers (nm) between the fluorescent dye and the substrate upon association of the linker and the substrate. In some embodiments, the functional length varies based on one or more members selected from the group consisting of temperature, solvent, pH, and salt concentration of a solution comprising the fluorescent labeling reagent. In some embodiments, the functional length is between about 0.5 nanometers (nm) and 50 nm.

**[0009]** In some embodiments, the linker is configured to form a bond to a plurality of fluorescent dyes or substrates.

**[0010]** In some embodiments, the linker comprises a plurality of amino acids. In some embodiments, the plurality of amino acids comprises a plurality of non-proteinogenic amino acids. In some embodiments, the plurality of amino acids comprises a plurality of hydroxyprolines. In some embodiments, the plurality of amino acids comprises three or more hydroxyprolines. In some embodiments, the plurality of amino acids comprises ten or more hydroxyprolines.

**[0011]** In some embodiments, the plurality of amino acids comprises a comprises a homopolymer. In some embodiments, the homopolymer comprises a repeating unit that is an amino acid. In some embodiments, the repeating unit is hydroxyproline. In some embodiments, the homopolymer of the linker comprises three or more hydroxyprolines. In some embodiments, the homopolymer of the linker comprises ten or more hydroxyprolines.

**[0012]** In some embodiments, the linker comprises a copolymer. In some embodiments, the copolymer comprises two or more repeating units, wherein at least one of the two or more



Date Reçue/Date Received 2023-12-22

repeating units is an amino acid. In some embodiments, the amino acid is a non-proteinogenic amino acid.

**[0013]** In some embodiments, the two or more ring systems comprise aromatic or aliphatic rings. In some embodiments, the two or more ring systems comprise rings having 5 or 6 members.

**[0014]** In some embodiments, at least two of the two or more ring systems are connected to each other by one or two $sp^3$ carbon atoms. In some embodiments, at least two of the two or more ring systems are connected to each other by an $sp^2$ carbon atoms. In some embodiments, the at least two of the two or more ring systems are connected to each other directly without an intervening carbon atom.

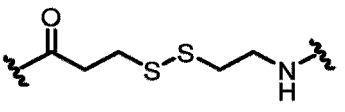
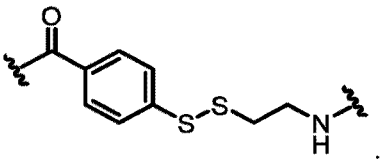
**[0015]** In some embodiments, at least two of the two or more ring systems comprises a water-soluble group of the one or more water soluble groups. In some embodiments, at least one water-soluble group of the one or more water-soluble groups is appended to a ring system of the two or more ring systems. In some embodiments, at least one water-soluble group of the one or more water-soluble groups is a constituent part of a ring system of the two or more ring systems. In some embodiments, at least one water-soluble group of the one or more water-soluble groups is positively charged. In some embodiments, the one or more water-soluble groups are selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester. In some embodiments, the one or more water-soluble groups decrease the logP of the fluorescent labeling reagent. In some embodiments, the fluorescent labeling reagent comprises more ring systems than water-soluble groups.

**[0016]** In some embodiments, the linker is configured to form a covalent bond with the substrate. In some embodiments, the linker is configured to form a non-covalent bond with the substrate.

**[0017]** In some embodiments, the fluorescent labeling reagent further comprises a cleavable group that is configured to be cleaved to separate the fluorescent labeling reagent or portion thereof from the substrate. In some embodiments, the cleavable group is configured to be cleaved to separate a first portion of the fluorescent labeling reagent comprising the fluorescent dye and a first portion of the linker and a second portion of the fluorescent labeling reagent comprising a second portion of the linker. In some embodiments, the cleavable group is selected from the group consisting of an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group. In some embodiments, the cleavable group is cleavable by application of one or more members of the group consisting of tris(2-carboxyethyl)phosphine



Date Reçue/Date Received 2023-12-22

(TCEP), dithiothreitol (DTT), tetrahydropyranyl (THP), ultraviolet (UV) light, and a combination thereof. In some embodiments, the linker comprises a moiety selected from the group consisting of  and .

**[0018]** In some embodiments, the fluorescent labeling reagent is configured to emit a signal between about 625 nanometers (nm) - 740 nm. In some embodiments, the fluorescent labeling reagent is configured to emit a signal between about 500 nanometers (nm) - 565 nm.

**[0019]** In some embodiments, the substrate is a protein, lipid, cell, or antibody. In some embodiments, the substrate is a nucleotide. In some embodiments, the linker is attached to the nucleotide via the nucleobase of the nucleotide. In some embodiments, the substrate is a fluorescence quencher, a fluorescence donor, or a fluorescence acceptor.

**[0020]** In another aspect, the present disclosure provides a composition comprising a solution comprising a fluorescently labeled nucleotide, wherein the fluorescently labeled nucleotide comprises a fluorescent dye that is connected to a nucleotide via a linker, wherein the linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two atoms, and wherein the linker comprises a non-proteinogenic amino acid comprising a ring system of the two or more ring systems.

**[0021]** In some embodiments, the fluorescently labeled nucleotide is configured to emit a fluorescent signal.

**[0022]** In some embodiments, the linker comprises a plurality of amino acids. In some embodiments, the plurality of amino acids comprises a plurality of non-proteinogenic amino acids. In some embodiments, the linker comprises a plurality of hydroxyprolines.

**[0023]** In some embodiments, the at least two ring systems of the two or more ring systems are connected to each other by an $sp^2$ carbon atom. In some embodiments, the at least two ring systems of the two or more ring systems are directly connected to each other without an intervening carbon atom.

**[0024]** In some embodiments, at least one water-soluble group of the one or more water-soluble groups is appended to a ring system of the two or more ring systems. In some embodiments, the one or more water soluble groups are selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester.



Date Reçue/Date Received 2023-12-22

**[0025]** In some embodiments, the linker further comprises a cleavable group that is configured to be cleaved to separate the fluorescent dye from the nucleotide. In some embodiments, the cleavable group is selected from the group consisting of an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group.

**[0026]** In some embodiments, the solution comprises a plurality of fluorescently labeled nucleotides, wherein each fluorescently labeled nucleotide of the plurality of the fluorescently labeled nucleotides comprises a fluorescent dye of a same type, a linker of a same type, and a nucleotide of a same type. In some embodiments, each the linker of each fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides has the same molecular weight. In some embodiments, the solution further comprises a plurality of unlabeled nucleotides, wherein each nucleotide of the plurality of unlabeled nucleotides is of a same type as each the nucleotide of the plurality of fluorescently labeled nucleotides. In some embodiments, the ratio of the plurality of fluorescently labeled nucleotides to the plurality of unlabeled nucleotides in the solution is at least about 1:4. In some embodiments, the ratio is at least about 1:1.

**[0027]** The present disclosure also provides a method comprising providing a composition described herein to a template nucleic acid molecule coupled to a nucleic acid strand.

**[0028]** In some embodiments, the method further comprises subjecting the template nucleic acid molecule and the composition to conditions sufficient to incorporate the fluorescently labeled nucleotide into the nucleic acid strand coupled to the template nucleic acid molecule. In some embodiments, the composition further comprises a polymerase enzyme, wherein the polymerase enzyme incorporates the fluorescently labeled nucleotide into the nucleic acid strand.

**[0029]** In some embodiments, the method further comprises detecting a signal from the fluorescently labeled nucleotide.

**[0030]** In some embodiments, the method further comprises contacting the fluorescently labeled nucleotide with a cleavage reagent configured to cleave the fluorescent dye from the nucleotide. In some embodiments, the cleavage reagent is configured to cleave the linker to provide the nucleotide attached to a portion of the linker. In some embodiments, the portion of the linker attached to the nucleotide comprises a thiol moiety, an aromatic moiety, or a combination thereof.

**[0031]** In some embodiments, the method further comprises, subsequent to the contacting the fluorescently labeled nucleotide with the cleavage reagent, subjecting the template nucleic acid molecule and the composition to conditions sufficient to incorporate an additional fluorescently labeled nucleotide into the nucleic acid strand coupled to the template nucleic acid molecule.

**[0032]** In some embodiments, the template nucleic acid molecule is immobilized to a support.



Date Reçue/Date Received 2023-12-22

**[0033]** In a further aspect, the present disclosure provides a method comprising providing a fluorescent labeling reagent, wherein the fluorescent labeling reagent comprises a fluorescent dye and a linker that is connected to the fluorescent dye, wherein the linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two atoms, and wherein the linker comprises a non-proteinogenic amino acid comprising a ring system of the two or more ring systems.

**[0034]** In some embodiments, the method further comprises contacting the fluorescent labeling reagent with a substrate to generate a fluorescently labeled substrate, wherein the linker connected to the fluorescent dye is coupled to the substrate. In some embodiments, the substrate is a nucleotide. In some embodiments, the substrate is a protein, lipid, cell, or antibody. In some embodiments, the fluorescently labeled substrate is configured to emit a fluorescent signal.

**[0035]** In some embodiments, the method further comprises contacting the fluorescently labeled substrate with a cleavage reagent, wherein the cleavage reagent is configured to cleave the fluorescent labeling reagent or a portion thereof from the fluorescently labeled substrate to generate a scarred substrate. In some embodiments, the cleavage reagent is configured to cleave a cleavable group of the linker, wherein the cleavable group is selected from the group consisting of an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group. In some embodiments, the scarred substrate comprises a thiol moiety, an aromatic moiety, or a combination thereof.

**[0036]** In some embodiments, the method further comprises, prior to generating the scarred substrate, subjecting the fluorescently labeled substrate and a nucleic acid molecule to conditions sufficient to incorporate the fluorescently labeled substrate into the nucleic acid molecule. In some embodiments, the fluorescently labeled substrate is incorporated into the nucleic acid molecule using a polymerase enzyme.

**[0037]** In some embodiments, the method further comprises, prior to generating the scarred substrate, subjecting an additional substrate and the nucleic acid molecule to conditions sufficient to incorporate the additional substrate into the nucleic acid molecule at a position adjacent to the fluorescently labeled substrate. In some embodiments, the additional substrate does not comprise a fluorescent labeling reagent. In some embodiments, the additional substrate comprises a fluorescent labeling reagent.

**[0038]** In some embodiments, the method further comprises, subsequent to generating the scarred substrate, subjecting an additional substrate and the nucleic acid molecule to conditions sufficient to incorporate the additional substrate into the nucleic acid molecule at a position



Date Reçue/Date Received 2023-12-22

adjacent to the scarred substrate. In some embodiments, the additional substrate does not comprise a fluorescent labeling reagent. In some embodiments, the additional substrate comprises a fluorescent labeling reagent.

**[0039]** In some embodiments, the nucleic acid molecule is immobilized to a support.

**[0040]** In some embodiments, the linker comprises a plurality of amino acids. In some embodiments, the plurality of amino acids comprises a plurality of non-proteinogenic amino acids. In some embodiments, the linker comprises a plurality of hydroxyprolines.

**[0041]** In some embodiments, the at least two ring systems of the two or more ring systems are connected to each other by an $sp^2$ carbon atom. In some embodiments, the at least two ring systems of the two or more ring systems are directly connected to each other without an intervening carbon atom.

**[0042]** In some embodiments, at least one water-soluble group of the one or more water-soluble groups is appended to a ring system of the two or more ring systems. In some embodiments, the one or more water soluble groups are selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester.

**[0043]** In another aspect, the present disclosure provides a kit comprising: a plurality of linkers, wherein a linker of the plurality of linkers comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms, and wherein the linker comprises a non-proteinogenic amino acid comprising a ring system of the two or more ring systems.

**[0044]** In some embodiments, the linker comprises a plurality of amino acids. In some embodiments, the plurality of amino acids comprises a plurality of non-proteinogenic amino acids. In some embodiments, the linker comprises a plurality of hydroxyprolines.

**[0045]** In some embodiments, the at least two ring systems of the two or more ring systems are connected to each other by an $sp^2$ carbon atom. In some embodiments, the at least two ring systems of the two or more ring systems are directly connected to each other without an intervening carbon atom.

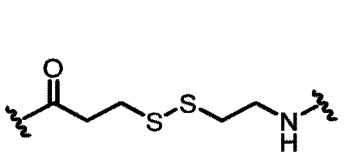
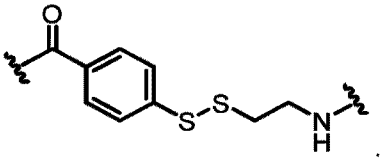
**[0046]** In some embodiments, at least one water-soluble group of the one or more water-soluble groups is appended to a ring system of the two or more ring systems. In some embodiments, the one or more water soluble groups are selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an



Date Reçue/Date Received 2023-12-22

imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester.

**[0047]** In some embodiments, the linker further comprises a cleavable group that is configured to be cleaved to separate a first portion of the linker from a second portion of the linker. In some embodiments, the cleavable group is selected from the group consisting of an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group. In some embodiments, the cleavable group is cleavable by application of one or more members of the group consisting of tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), tetrahydropyranyl (THP), ultraviolet (UV) light, and a combination thereof. In some embodiments, the linker comprises a moiety selected from the group consisting of


and

.

**[0048]** In some embodiments, the linker is connected to a fluorescent dye.

**[0049]** In some embodiments, the linker is associated with a substrate. In some embodiments, the substrate comprises a protein, lipid, cell, or antibody. In some embodiments, the substrate comprises a nucleotide.

**[0050]** In some embodiments, the plurality of linkers comprises a first linker associated with a first substrate and a second linker associated with a second substrate, wherein the first substrate and the second substrate are of different types. In some embodiments, the first linker and the second linker comprise the same chemical structure. In some embodiments, the first substrate and the second substrate are nucleotides comprising nucleobases of different types. In some embodiments, the kit further comprises a third linker associated with a third substrate and a fourth linker associated with a fourth substrate, wherein the first substrate, the second substrate, the third substrate, and the fourth substrate are of different types. In some embodiments, the first substrate, the second substrate, the third substrate, and the fourth substrate are nucleotides comprising nucleobases of different types. In some embodiments, the first linker and the third linker comprise different chemical structures. In some embodiments, the first linker and the third linker comprise a same chemical group. In some embodiments, the same chemical group comprises a disulfide bond.

**[0051]** In a further aspect, the present disclosure provides an oligonucleotide molecule comprising a fluorescent labeling reagent described herein, or a derivative thereof.

**[0052]** In some embodiments, the oligonucleotide molecule further comprises one or more additional fluorescent labeling reagents. In some embodiments, the fluorescent labeling reagent



Date Reçue/Date Received 2023-12-22

and the one or more additional fluorescent labeling reagents comprise linkers having the same chemical structure. In some embodiments, the fluorescent labeling reagent and the one or more additional fluorescent labeling reagents comprise fluorescent dyes having the same chemical structure. In some embodiments, the fluorescent labeling reagent and the one or more additional fluorescent labeling reagents are associated with substrates of a same type, wherein the substrates are nucleotides. In some embodiments, the fluorescent labeling reagent and the one or more additional fluorescent labeling reagents are connected to nucleobases of the nucleotides. In some embodiments, the fluorescent labeling reagent and the one or more additional fluorescent labeling reagents are connected to adjacent nucleotides of the oligonucleotide molecule. In some embodiments, the fluorescent labeling reagent and the one or more additional fluorescent labeling reagents are connected to nucleotides of the oligonucleotide molecule that are separated by one or more nucleotides that are not connected to fluorescent labeling reagents. In some embodiments, the linker of the fluorescent labeling reagent comprises a cleavable group that is configured to be cleaved to separate the fluorescent dye from a substrate with which it is associated. In some embodiments, the fluorescent labeling reagent is configured to emit a fluorescent signal.

**[0053]** In another aspect, the present disclosure provides a method, comprising: (a) contacting a nucleic acid molecule with a solution comprising a plurality of nucleotides under conditions sufficient to incorporate a first labeled nucleotide and a second labeled nucleotide of the plurality of nucleotides into a growing strand that is complementary to the nucleic acid molecule, wherein at least about 20% of the plurality of nucleotides are labeled nucleotides; (b) detecting one or more signals or signal changes from the first labeled nucleotide and the second labeled nucleotide, wherein the one or more signals or signal changes are indicative of incorporation of the first labeled nucleotide and the second labeled nucleotide; and (c) resolving the one or more signals or signal changes to determine a sequence of the nucleic acid molecule.

**[0054]** In some embodiments, the first labeled nucleotide and the second labeled nucleotide are a same canonical base type. In some embodiments, the first labeled nucleotide comprises a fluorescent dye. In some embodiments, the second labeled nucleotide comprises the fluorescent dye. In some embodiments, the fluorescent dye is cleavable. In some embodiments, the method further comprises (i) cleaving the fluorescent dye; (ii) contacting the nucleic acid molecule with a second solution comprising a second plurality of nucleotides under conditions sufficient to incorporate a third labeled nucleotide of the second plurality of nucleotides into the growing strand, wherein at least about 20% of the second plurality of nucleotides are labeled nucleotides; (iii) detecting one or more second signals or signal changes from the third labeled nucleotide;



Date Reçue/Date Received 2023-12-22

and (iv) resolving the one or more second signals or signal changes to determine a second sequence of the nucleic acid molecule. In some embodiments, the first labeled nucleotide and the third labeled nucleotide are different canonical base types. In some embodiments, the third labeled nucleotide comprises the fluorescent dye.

**[0055]** In some embodiments, the method further comprises (i) contacting the nucleic acid molecule with a second solution comprising a second plurality of nucleotides under conditions sufficient to incorporate a third labeled nucleotide of the second plurality of nucleotides into the growing strand, wherein at least about 20% of the second plurality of nucleotides are labeled nucleotides; (ii) detecting one or more second signals or signal changes from the third labeled nucleotide; and (iii) resolving the one or more second signals or signal changes to determine a second sequence of the nucleic acid molecule. In some embodiments, the first labeled nucleotide and the third labeled nucleotide are different canonical base types. In some embodiments, the third labeled nucleotide comprises the fluorescent dye. In some embodiments, the contacting in (i) is performed in absence of cleaving a fluorescent dye from the first labeled nucleotide or the second labeled nucleotide. In some embodiments, the method further comprises repeating (i)-(iii) at least 5 times, each with a different solution of nucleotides that comprises at least 20% labeled nucleotides, in absence of cleaving a fluorescent dye from the first labeled nucleotide or the second labeled nucleotide.

**[0056]** In some embodiments, at least about 50%, 70%, 80%, 90%, 95%, or 99% of the plurality of nucleotides are labeled nucleotides. In some embodiments, substantially all of the plurality of nucleotides are labeled nucleotides. In some embodiments, the resolving in (c) comprises determining a number of consecutive nucleotides from the solution that incorporated into the growing strand. In some embodiments, the number is selected from the group consisting of 2, 3, 4, 5, 6, 7, or 8 nucleotides. In some embodiments, the resolving in (c) comprises processing a tolerance of the solution.

**[0057]** In some embodiments, subsequent to (a) a third nucleotide of the plurality of nucleotides has incorporated into the growing strand. In some embodiments, the third nucleotide is unlabeled. In some embodiments, the third nucleotide is labeled. In some embodiments, the first labeled nucleotide and the third nucleotide are a same canonical base type. In some embodiments, the first labeled nucleotide and the third nucleotide are different canonical base types.

**[0058]** In another aspect, the present disclosure provides a method, comprising: (a) contacting a nucleic acid molecule with a solution comprising a plurality of non-terminated nucleotides under conditions sufficient to incorporate a first nucleotide and a second nucleotide of the plurality of



Date Reçue/Date Received 2023-12-22

non-terminated nucleotides into a growing strand that is complementary to the nucleic acid molecule, wherein the first nucleotide is labeled, and wherein at least about 20% of the plurality of nucleotides are labeled nucleotides; (b) detecting one or more signals or signal changes from the first nucleotide, wherein the one or more signals or signal changes are indicative of incorporation of the first nucleotide; and (c) resolving the one or more signals or signal changes to determine a sequence of the nucleic acid molecule.

**[0059]** In some embodiments, the plurality of non-terminated nucleotides comprises nucleotides of a same canonical base type. In some embodiments, the first nucleotide comprises a fluorescent dye. In some embodiments, the fluorescent dye is cleavable. In some embodiments, the method further comprises: (i) cleaving the fluorescent dye; (ii) contacting the nucleic acid molecule with a second solution comprising a second plurality of non-terminated nucleotides under conditions sufficient to incorporate a third nucleotide of the second plurality of non-terminated nucleotides into the growing strand, wherein at least about 20% of the second plurality of non-terminated nucleotides are labeled nucleotides, wherein the third nucleotide is a labeled nucleotide; (iii) detecting one or more second signals or signal changes from the third nucleotide; and (iv) resolving the one or more second signals or signal changes to determine a second sequence of the nucleic acid molecule. In some embodiments, the first nucleotide and the third nucleotide are different canonical base types. In some embodiments, the third nucleotide comprises the fluorescent dye.

**[0060]** In some embodiments, the method further comprises: (i) contacting the nucleic acid molecule with a second solution comprising a second plurality of non-terminated nucleotides under conditions sufficient to incorporate a third nucleotide of the second plurality of non-terminated nucleotides into the growing strand, wherein at least about 20% of the second plurality of nucleotides are labeled nucleotides, wherein the third nucleotide is a labeled nucleotide; (ii) detecting one or more second signals or signal changes from the third nucleotide; and (iii) resolving the one or more second signals or signal changes to determine a second sequence of the nucleic acid molecule. In some embodiments, the first nucleotide and the third nucleotide are different canonical base types. In some embodiments, the third nucleotide comprises the fluorescent dye. In some embodiments, the contacting in (i) is performed in absence of cleaving a fluorescent dye from the first nucleotide. In some embodiments, the method further comprises repeating (i)-(iii) at least 5 times, each with a different solution of non-terminated nucleotides that comprises at least 20% labeled nucleotides, in absence of cleaving a fluorescent dye from the first nucleotide.



Date Reçue/Date Received 2023-12-22

**[0061]** In some embodiments, at least about 50%, 70%, 80%, 90%, 95%, or 99% of the plurality of non-terminated nucleotides are labeled nucleotides. In some embodiments, substantially all of the plurality of non-terminated nucleotides are labeled nucleotides. In some embodiments, the resolving in (c) comprises determining a number of consecutive nucleotides from the solution that incorporated into the growing strand. In some embodiments, the number is selected from the group consisting of 2, 3, 4, 5, 6, 7, or 8 nucleotides. In some embodiments, the resolving in (c) comprises processing a tolerance of the solution.

**[0062]** In some embodiments, the second nucleotide is unlabeled. In some embodiments, the second nucleotide is labeled. In some embodiments, the first nucleotide and the second nucleotide are a same canonical base type. In some embodiments, the first nucleotide and the second nucleotide are different canonical base types.

**[0063]** In an aspect, the present disclosure provides a fluorescent labeling reagent comprising: (a) a fluorescent dye; and (b) a linker that is connected to the fluorescent dye and capable of associating with a substrate for fluorescently labelling the substrate, wherein the linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms. In some embodiments, the linker is configured to establish a functional length between the fluorescent dye and the substrate of at least about 0.5 nanometers (nm) upon association of the linker and the substrate.

**[0064]** In some embodiments, the functional length is as measured in a solution. In some embodiments, the fluorescent labeling reagent coupled to the substrate is capable of emitting a fluorescent signal in the solution. In some embodiments, the functional length varies based on the temperature, solvent, pH, or salt concentration of the solution.

**[0065]** In some embodiments, the functional length is between about 0.5 and 50 nm.

**[0066]** In some embodiments, the linker is capable of forming a bond with a plurality of fluorescent dyes and/or substrates.

**[0067]** In some embodiments, the linker has a defined molecular weight.

**[0068]** In some embodiments, the linker comprises a polymer having a regularly repeating unit. In some embodiments, the linker is a co-polymer without a regularly repeating unit.

**[0069]** In some embodiments, the two or more ring systems comprise aromatic or aliphatic rings. In some embodiments, the two or more ring systems comprise rings having 5 or 6 members. In some embodiments, at least one of the two or more ring systems comprises hydroxyproline.



Date Reçue/Date Received 2023-12-22

**[0070]** In some embodiments, the two or more ring systems are connected to each other by one or two $sp^3$ carbon atoms. In some embodiments, the two or more ring systems are connected to each other directly without an intervening carbon atom.

**[0071]** In some embodiments, each of the two or more ring systems comprises a water-soluble group. In some embodiments, the fluorescent labeling reagent comprises more ring systems than water-soluble groups. In some embodiments, at least one water-soluble group of the one or more water-soluble groups is appended to a ring system of the two or more ring systems. In some embodiments, at least one water-soluble group of the one or more water-soluble groups is a constituent part of a ring system of the two or more ring systems. In some embodiments, at least one water-soluble group of the one or more water-soluble groups is positively charged. In some embodiments, the one or more water-soluble groups are selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester. In some embodiments, the one or more water-soluble groups decrease the logP of the fluorescent labeling reagent.

**[0072]** In some embodiments, the substrate is capable of associating with one or more different moieties of the fluorescent labeling reagent.

**[0073]** In some embodiments, the linker is capable of forming a covalent bond with the substrate.

**[0074]** In some embodiments, the linker is capable of forming a non-covalent bond with the substrate. In some embodiments, the non-covalent bond is a biotin-streptavidin bond.

**[0075]** In some embodiments, the fluorescent labeling reagent coupled to the substrate is capable of emitting a fluorescent signal, which fluorescent signal is proportional to the number of fluorescent labeling reagents associated with the substrate.

**[0076]** In some embodiments, the fluorescent labeling reagent further comprises a cleavable group that is capable of being cleaved to separate the fluorescent labeling reagent or portion thereof from the substrate. In some embodiments, cleavage of the cleavable group leaves a scar group associated with substrate. In some embodiments, the cleavable group is an azidomethyl group capable of being cleaved by tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), or tetrahydropyranyl (THP) to leave a hydroxyl scar group. In some embodiments, the cleavable group is a disulfide bond capable of being cleaved by tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), or tetrahydropyranyl (THP) to leave a thiol scar group. In some embodiments, the cleavable group is a hydrocarbyldithiomethyl group capable of being cleaved by tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), or tetrahydropyranyl (THP) to



Date Reçue/Date Received 2023-12-22

leave a hydroxyl scar group. In some embodiments, the cleavable group is a 2-nitrobenzyloxy group capable of being cleaved by ultraviolet (UV) light to leave a hydroxyl scar group.

**[0077]** In some embodiments, the fluorescent dye is Atto633.

**[0078]** In some embodiments, the substrate to be labeled is a protein, lipid, cell, or antibody. In some embodiments, the substrate is a nucleotide. In some embodiments, the linker is attached to the nucleobase of the nucleotide. In some embodiments, the substrate is a fluorescence quencher, a fluorescence donor, or fluorescence acceptor.

**[0079]** In some embodiments, the linker is capable of being made by peptide synthesis chemistry.

**[0080]** In some embodiments, the linker comprises a plurality of amino acids. In some embodiments, the plurality of amino acids comprises a plurality of non-proteinogenic (e.g., non-natural) amino acids. In some embodiments, the linker comprises a polymerization product of two half-monomers. In some embodiments, the two half-monomers have water-solubilizing groups. In some embodiments, at least one of the two or more ring systems comprises hydroxyproline.

**[0081]** In another aspect, the present disclosure provides a method for sequencing a nucleic acid molecule, the method comprising: (a) contacting the nucleic acid molecule with a primer under conditions sufficient to hybridize the primer to the nucleic acid molecule, thereby generating a sequencing template; (b) contacting the sequencing template with a polymerase and a solution comprising a plurality of fluorescently labeled nucleotides, wherein each fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides is of a same type, and wherein a fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides is complementary to the nucleic acid molecule at a plurality of positions adjacent to the primer hybridized to the nucleic acid molecule, thereby incorporating two or more fluorescently labeled nucleotides of the plurality of fluorescently labeled nucleotides into the sequencing template; (c) washing the solution comprising the plurality of fluorescently labeled nucleotides away from the sequencing template; and (d) measuring a fluorescent signal emitted by the sequencing template, wherein the intensity of the measured fluorescent signal is greater than a fluorescent signal that may be measured if a single fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides had been incorporated into the sequencing template, wherein a fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides comprises a fluorescent dye and a linker that is connected to the fluorescent dye and a nucleotide, wherein the linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms; and



Date Reçue/Date Received 2023-12-22

wherein the linker establishes a functional length between the fluorescent dye and the nucleotide of at least about 0.5 nanometers.

**[0082]** In some embodiments, the fluorescently labeled nucleotide comprises any optical (e.g., fluorescent) labeling reagent described herein.

**[0083]** In some embodiments, the intensity of the measured fluorescent signal is proportional to the number of fluorescently labeled nucleotides incorporated into the sequencing template. In some embodiments, the intensity of the measured fluorescent signal is linearly proportional to the number of fluorescently labeled nucleotides incorporated into the sequencing template. In some embodiments, the intensity of the measured fluorescent signal is linearly proportional with a slope of approximately 1.0 when plotted against the number of fluorescently labeled nucleotides incorporated into the sequencing template.

**[0084]** In some embodiments, the solution comprising the plurality of fluorescently labeled nucleotides also contains un-labeled nucleotides. In some embodiments, at least about 20% of nucleotides in the solution are fluorescently labeled.

**[0085]** In some embodiments, three or more fluorescently labeled nucleotides of the plurality of fluorescently labeled nucleotides are incorporated into the sequencing template.

**[0086]** In some embodiments, a first fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides is incorporated within four positions of a second fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides.

**[0087]** In some embodiments, the method further comprises, subsequent to (d), cleaving fluorescent labels of the two or more fluorescently labeled nucleotides incorporated into the sequencing template.

**[0088]** In a further aspect, the present disclosure provides a method for sequencing a nucleic acid molecule, the method comprising: (a) contacting the nucleic acid molecule with a primer under conditions sufficient to hybridize the primer to the nucleic acid molecule, thereby generating a sequencing template; (b) contacting the sequencing template with a polymerase and a first solution comprising a plurality of first fluorescently labeled nucleotides, wherein each first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides is of a same type, and wherein a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides is complementary to the nucleic acid molecule at a position adjacent to the primer hybridized to the nucleic acid molecule, thereby incorporating a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides into the sequencing template to generate an extended primer; (c) washing the first solution comprising the plurality of first fluorescently labeled nucleotides away from the sequencing template; (d)



Date Reçue/Date Received 2023-12-22

measuring a first fluorescent signal emitted by the sequencing template; (e) contacting the sequencing template with a polymerase and a second solution comprising a plurality of second fluorescently labeled nucleotides, wherein each second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides is of a same type, and wherein a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides is complementary to the nucleic acid molecule at a position adjacent to the extended primer hybridized to the nucleic acid molecule, thereby incorporating a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides into the sequencing template to generate a further extended primer; (f) washing the second solution comprising the plurality of second fluorescently labeled nucleotides away from the sequencing template; and (g) measuring a second fluorescent signal emitted by the sequencing template, wherein the intensity of the second fluorescent signal is greater than the intensity of the first fluorescent signal, wherein a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides comprises a first fluorescent dye and a first linker that is connected to the first fluorescent dye and a first nucleotide, and a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides comprises a second fluorescent dye and a second linker that is connected to the second fluorescent dye and a second nucleotide; and wherein (I) the first linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms; and wherein the first linker establishes a functional length between the first fluorescent dye and the first nucleotide of at least about 0.5 nanometers; and/or (II) the second linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms; and wherein the second linker establishes a functional length between the second fluorescent dye and the second nucleotide of at least about 0.5 nanometers.

**[0089]** In some embodiments, the first fluorescently labeled nucleotide and/or the second fluorescently labeled nucleotide comprises any optical (e.g., fluorescent) labeling reagent described herein.

**[0090]** In some embodiments, the first linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms; and wherein the first linker establishes a functional length between the first fluorescent dye and the first nucleotide of at least about 0.5 nanometers.

**[0091]** In some embodiments, the second linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each



Date Reçue/Date Received 2023-12-22

other by no more than two $sp^3$ carbon atoms; and wherein the second linker establishes a functional length between the second fluorescent dye and the second nucleotide of at least about 0.5 nanometers.

**[0092]** In some embodiments, the first solution comprising the plurality of first fluorescently labeled nucleotides also contains first un-labeled nucleotides.

**[0093]** In some embodiments, the second solution comprising the plurality of second fluorescently labeled nucleotides also contains second un-labeled nucleotides.

**[0094]** In some embodiments, the plurality of first fluorescently labeled nucleotides is different than the plurality of second fluorescently labeled nucleotides. In some embodiments, the first fluorescent dye of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides and the second fluorescent dye of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides are the same, and the first nucleotide of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides and the second nucleotide of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides are plurality of second fluorescently labeled nucleotides are different. In some embodiments, the first fluorescent dye of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides and the second fluorescent dye of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides are different, and the first nucleotide of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides and the second nucleotide of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides are plurality of second fluorescently labeled nucleotides are the same. In some embodiments, the first fluorescent dye of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides and the second fluorescent dye of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides are different, and the first nucleotide of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides and the second nucleotide of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides are plurality of second fluorescently labeled nucleotides are different.

**[0095]** In some embodiments, two or more first fluorescently labeled nucleotides are incorporated into the sequencing template. In some embodiments, two or more second fluorescently labeled nucleotides are incorporated into the sequencing template.

**[0096]** In some embodiments, the method further comprises: (h) contacting the sequencing template with a polymerase and a third solution comprising a plurality of third fluorescently



Date Reçue/Date Received 2023-12-22

labeled nucleotides, wherein each third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides is of a same type, and wherein a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides is complementary to the nucleic acid molecule at a position adjacent to the further extended primer hybridized to the nucleic acid molecule, thereby incorporating a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides into the sequencing template; (i) washing the third solution comprising the plurality of third fluorescently labeled nucleotides away from the sequencing template; and (j) measuring a third fluorescent signal emitted by the sequencing template, wherein the intensity of the third fluorescent signal is greater than the intensity of the first fluorescent signal and the intensity of the second fluorescent signal, wherein a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides comprises a third fluorescent dye and a third linker that is connected to the third fluorescent dye and a third nucleotide.

**[0097]** In some embodiments, the third linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms; and wherein the third linker establishes a functional length between the third fluorescent dye and the third nucleotide of at least about 0.5 nanometers.

**[0098]** In some embodiments, the third fluorescent dye of a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides is different than the first fluorescent dye of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides. In some embodiments, the third fluorescent dye of a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides is different than the second fluorescent dye of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides. In some embodiments, the third nucleotide of a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides is different than the first nucleotide of a first fluorescently labeled nucleotide of the plurality of first fluorescently labeled nucleotides. In some embodiments, the third nucleotide of a third fluorescently labeled nucleotide of the plurality of third fluorescently labeled nucleotides is different than the second nucleotide of a second fluorescently labeled nucleotide of the plurality of second fluorescently labeled nucleotides.

**[0099]** In some embodiments, the method further comprises subsequent to (d), cleaving the first fluorescent dye of the first fluorescently labeled nucleotide incorporated into the sequencing template.



Date Reçue/Date Received 2023-12-22

**[00100]** In some embodiments, the method further comprises, subsequent to (g), cleaving the second fluorescent dye of the second fluorescently labeled nucleotide incorporated into the sequencing template.

**[00101]** In yet another aspect, the present disclosure provides a method for sequencing a nucleic acid molecule, the method comprising: (a) providing a solution comprising a plurality of fluorescently labeled nucleotides, wherein each fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides is of a same type, and wherein a given fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides comprises a fluorescent dye that is connected to a nucleotide via a semi-rigid water-soluble linker having a defined molecular weight and a length of at least about 0.5 nanometers (nm); (b) contacting the nucleic acid molecule with a primer under conditions sufficient to hybridize the primer to the nucleic acid molecule, thereby generating a sequencing template; (c) contacting the sequencing template with a polymerase and the solution comprising the plurality of fluorescently labeled nucleotides, wherein a fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides is complementary to the nucleic acid molecule at a position adjacent to the primer hybridized to the nucleic acid molecule, thereby incorporating one or more fluorescently labeled nucleotides of the plurality of fluorescently labeled nucleotides into the sequencing template; (d) washing the solution comprising the plurality of fluorescently labeled nucleotides away from the sequencing template; and (e) measuring a fluorescent signal emitted by the sequencing template.

**[00102]** In some embodiments, the nucleotide is guanine (G).

**[00103]** In some embodiments, the linker decreases quenching between the nucleotide and the fluorescent dye.

**[00104]** In some embodiments, a fluorescently labeled nucleotide of the one or more fluorescently labeled nucleotides is more efficiently incorporated into the sequencing template than another fluorescently labeled nucleotide that comprises the same nucleotide and fluorescent dye but does not include the linker.

**[00105]** In some embodiments, a fluorescently labeled nucleotide of the one or more fluorescently labeled nucleotides is incorporated into the sequencing template with higher fidelity than another fluorescently labeled nucleotide that comprises the same nucleotide and fluorescent dye but does not include the linker.

**[00106]** In some embodiments, the polymerase is a Family A polymerase selected from the group consisting of Taq polymerase, Klenow polymerase, and Bst polymerase.

**[00107]** In some embodiments, the polymerase is a Family B polymerase selected from the group consisting of Vent(exo-) polymerase and Therminator™ polymerase.



Date Reçue/Date Received 2023-12-22

**[00108]** In some embodiments, the linker comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms; and wherein the linker establishes a functional length between the fluorescent dye and the nucleotide of at least about 0.5 nanometers.

**[00109]** Another aspect of the present disclosure provides a non-transitory computer readable medium comprising machine executable code that, upon execution by one or more computer processors, implements any of the methods above or elsewhere herein.

**[00110]** Another aspect of the present disclosure provides a system comprising one or more computer processors and computer memory coupled thereto. The computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements any of the methods above or elsewhere herein.

**[00111]** Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

**[00112]**

## BRIEF DESCRIPTION OF THE DRAWINGS

**[00113]** The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also "Figure" and "FIG." herein), of which:

**[00114]** **FIG. 1A** shows examples of linkers of the present disclosure;

**[00115]** **FIG. 1B** shows an example of a linker of the present disclosure;

**[00116]** **FIG. 1C** shows an example of a linker of the present disclosure, where R is a water solubilizing group;

**[00117]** **FIG. 2A** shows an example of a method for synthesizing a linker of the present disclosure having an effective length of about 2 nanometers;

**[00118]** **FIG. 2B** shows an example of reagents that can be used in the method of **FIG. 2A** for synthesizing a linker of the present disclosure as well as some trifunctional reagents;

**[00119]** **FIG. 2C** shows an example of a method for synthesizing a linker of the present disclosure that is polymeric with defined molecular weight and linking groups; and

**[00120]** **FIG. 3** shows a computer system that is programmed or otherwise configured to implement methods provided herein;

**[00121]** **FIG. 4** shows an example of a method for constructing a labeled nucleotide comprising a propargyl-derivatized nucleotide, a linker, and a dye.

**[00122]** **FIGs. 5A** and **5B** show an example method for preparing a labeled nucleotide comprising a dGTP analog.

**[00123]** **FIG. 6** shows an example method for the preparing a labeled nucleotide comprising dCTP.

**[00124]** **FIG. 7** shows components used to construct dye-labeled nucleotides for excitation at about 530 nm.

**[00125]** **FIG. 8** shows an example method for preparing a labeled nucleotide comprising a guanine analog.

**[00126]** **FIG. 9** shows a schematic of a bead-based assay for evaluating labeled nucleotides.

**[00127]** **FIG. 10** shows results of a bead-based assay for different labeled dUTPs.

**[00128]** **FIG. 11** shows results of a bead-based assay for different labeled dATPs.

**[00129]** **FIG. 12** shows results of a bead-based assay for different labeled dGTPs.

**[00130]** **FIGs. 13A-13C** show an example method for preparing a labeled nucleotide comprising a guanine analog.

**[00131]** **FIGs. 14A** and **14B** show an example method for preparing a labeled nucleotide comprising repeating units of an amino acid.

**[00132]** **FIG. 15** shows a schematic of an assay for evaluating quenching.

**[00133]** **FIG. 16** shows quenching results for red dye linkers.

**[00134]** **FIG. 17** shows quenching results for green dye linkers.

**[00135]** **FIG. 18** shows an example sequencing procedure.

**[00136]** **FIG. 19** shows tolerances of different labeled nucleotides.



Date Reçue/Date Received 2023-12-22

**[00137]** **FIGs. 20A** and **20B** show examples of constructs including homopolymeric regions.

**[00138]** **FIG. 20C** shows signals detected from sequencing a template having a homopolymeric region using labeled nucleotides.

**[00139]** **FIG. 21A** shows example results of a sequencing analysis utilizing populations of nucleotides comprising 20% fluorophore labeled dNTPs.

**[00140]** **FIG. 21B** shows fluorescence signal intensity as a function of homopolymer length.

**[00141]** **FIG. 22** shows example results of a sequencing analysis utilizing populations of nucleotides comprising 100% fluorophore labeled dNTPs.

## DETAILED DESCRIPTION

**[00142]** While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

**[00143]** Where values are described as ranges, it will be understood that such disclosure includes the disclosure of all possible sub-ranges within such ranges, as well as specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated.

**[00144]** The terms "about" and "approximately" shall generally mean an acceptable degree of error or variation for a given value or range of values, such as, for example, a degree of error or variation that is within 20 percent (%), within 15%, within 10%, or within 5% of a given value or range of values.

**[00145]** The term "subject," as used herein, generally refers to an individual or entity from which a biological sample (e.g., a biological sample that is undergoing or will undergo processing or analysis) may be derived. A subject may be an animal (e.g., mammal or non-mammal) or plant. The subject may be a human, dog, cat, horse, pig, bird, non-human primate, simian, farm animal, companion animal, sport animal, or rodent. A subject may be a patient. The subject may have or be suspected of having a disease or disorder, such as cancer (e.g., breast cancer, colorectal cancer, brain cancer, leukemia, lung cancer, skin cancer, liver cancer, pancreatic cancer, lymphoma, esophageal cancer or cervical cancer) or an infectious disease. Alternatively or in addition to, a subject may be known to have previously had a disease or



Date Reçue/Date Received 2023-12-22

disorder. The subject may have or be suspected of having a genetic disorder such as achondroplasia, alpha-1 antitrypsin deficiency, antiphospholipid syndrome, autism, autosomal dominant polycystic kidney disease, Charcot-Marie-tooth, cri du chat, Crohn's disease, cystic fibrosis, Dercum disease, down syndrome, Duane syndrome, Duchenne muscular dystrophy, factor V Leiden thrombophilia, familial hypercholesterolemia, familial Mediterranean fever, fragile x syndrome, Gaucher disease, hemochromatosis, hemophilia, holoprosencephaly, Huntington's disease, Klinefelter syndrome, Marfan syndrome, myotonic dystrophy, neurofibromatosis, Noonan syndrome, osteogenesis imperfecta, Parkinson's disease, phenylketonuria, Poland anomaly, porphyria, progeria, retinitis pigmentosa, severe combined immunodeficiency, sickle cell disease, spinal muscular atrophy, Tay-Sachs, thalassemia, trimethylaminuria, Turner syndrome, velocardiofacial syndrome, WAGR syndrome, or Wilson disease. A subject may be undergoing treatment for a disease or disorder. A subject may be symptomatic or asymptomatic of a given disease or disorder. A subject may be healthy (e.g., not suspected of having disease or disorder). A subject may have one or more risk factors for a given disease. A subject may have a given weight, height, body mass index, or other physical characteristic. A subject may have a given ethnic or racial heritage, place of birth or residence, nationality, disease or remission state, family medical history, or other characteristic.

**[00146]** As used herein, the term "biological sample" generally refers to a sample obtained from a subject. The biological sample may be obtained directly or indirectly from the subject. A sample may be obtained from a subject via any suitable method, including, but not limited to, spitting, swabbing, blood draw, biopsy, obtaining excretions (e.g., urine, stool, sputum, vomit, or saliva), excision, scraping, and puncture. A sample may be obtained from a subject by, for example, intravenously or intraarterially accessing the circulatory system, collecting a secreted biological sample (e.g., stool, urine, saliva, sputum, etc.), breathing, or surgically extracting a tissue (e.g., biopsy). The sample may be obtained by non-invasive methods including but not limited to: scraping of the skin or cervix, swabbing of the cheek, or collection of saliva, urine, feces, menses, tears, or semen. Alternatively, the sample may be obtained by an invasive procedure such as biopsy, needle aspiration, or phlebotomy. A sample may comprise a bodily fluid such as, but not limited to, blood (e.g., whole blood, red blood cells, leukocytes or white blood cells, platelets), plasma, serum, sweat, tears, saliva, sputum, urine, semen, mucus, synovial fluid, breast milk, colostrum, amniotic fluid, bile, bone marrow, interstitial or extracellular fluid, or cerebrospinal fluid. For example, a sample may be obtained by a puncture method to obtain a bodily fluid comprising blood and/or plasma. Such a sample may comprise both cells and cell-free nucleic acid material. Alternatively, the sample may be obtained from any other source



Date Reçue/Date Received 2023-12-22

including but not limited to blood, sweat, hair follicle, buccal tissue, tears, menses, feces, or saliva. The biological sample may be a tissue sample, such as a tumor biopsy. The sample may be obtained from any of the tissues provided herein including, but not limited to, skin, heart, lung, kidney, breast, pancreas, liver, intestine, brain, prostate, esophagus, muscle, smooth muscle, bladder, gall bladder, colon, or thyroid. The methods of obtaining provided herein include methods of biopsy including fine needle aspiration, core needle biopsy, vacuum assisted biopsy, large core biopsy, incisional biopsy, excisional biopsy, punch biopsy, shave biopsy or skin biopsy. The biological sample may comprise one or more cells. A biological sample may comprise one or more nucleic acid molecules such as one or more deoxyribonucleic acid (DNA) and/or ribonucleic acid (RNA) molecules (e.g., included within cells or not included within cells). Nucleic acid molecules may be included within cells. Alternatively or in addition to, nucleic acid molecules may not be included within cells (e.g., cell-free nucleic acid molecules). The biological sample may be a cell-free sample.

**[00147]** The term "cell-free sample," as used herein, generally refers to a sample that is substantially free of cells (e.g., less than 10% cells on a volume basis). A cell-free sample may be derived from any source (e.g., as described herein). For example, a cell-free sample may be derived from blood, sweat, urine, or saliva. For example, a cell-free sample may be derived from a tissue or bodily fluid. A cell-free sample may be derived from a plurality of tissues or bodily fluids. For example, a sample from a first tissue or fluid may be combined with a sample from a second tissue or fluid (e.g., while the samples are obtained or after the samples are obtained). In an example, a first fluid and a second fluid may be collected from a subject (e.g., at the same or different times) and the first and second fluids may be combined to provide a sample. A cell-free sample may comprise one or more nucleic acid molecules such as one or more DNA or RNA molecules.

**[00148]** A sample that is not a cell-free sample (e.g., a sample comprising one or more cells) may be processed to provide a cell-free sample. For example, a sample that includes one or more cells as well as one or more nucleic acid molecules (e.g., DNA and/or RNA molecules) not included within cells (e.g., cell-free nucleic acid molecules) may be obtained from a subject. The sample may be subjected to processing (e.g., as described herein) to separate cells and other materials from the nucleic acid molecules not included within cells, thereby providing a cell-free sample (e.g., comprising nucleic acid molecules not included within cells). The cell-free sample may then be subjected to further analysis and processing (e.g., as provided herein). Nucleic acid molecules not included within cells (e.g., cell-free nucleic acid molecules) may be derived from cells and tissues. For example, cell-free nucleic acid molecules may derive from a tumor tissue



Date Reçue/Date Received 2023-12-22

or a degraded cell (e.g., of a tissue of a body). Cell-free nucleic acid molecules may comprise any type of nucleic acid molecules (e.g., as described herein). Cell-free nucleic acid molecules may be double-stranded, single-stranded, or a combination thereof. Cell-free nucleic acid molecules may be released into a bodily fluid through secretion or cell death processes, e.g., cellular necrosis, apoptosis, or the like. Cell-free nucleic acid molecules may be released into bodily fluids from cancer cells (e.g., circulating tumor DNA (ctDNA)). Cell free nucleic acid molecules may also be fetal DNA circulating freely in a maternal blood stream (e.g., cell-free fetal nucleic acid molecules such as cffDNA). Alternatively or in addition to, cell-free nucleic acid molecules may be released into bodily fluids from healthy cells.

**[00149]** A biological sample may be obtained directly from a subject and analyzed without any intervening processing, such as, for example, sample purification or extraction. For example, a blood sample may be obtained directly from a subject by accessing the subject's circulatory system, removing the blood from the subject (e.g., via a needle), and transferring the removed blood into a receptacle. The receptacle may comprise reagents (e.g., anti-coagulants) such that the blood sample is useful for further analysis. Such reagents may be used to process the sample or analytes derived from the sample in the receptacle or another receptacle prior to analysis. In another example, a swab may be used to access epithelial cells on an oropharyngeal surface of the subject. Following obtaining the biological sample from the subject, the swab containing the biological sample may be contacted with a fluid (e.g., a buffer) to collect the biological fluid from the swab.

**[00150]** Any suitable biological sample that comprises one or more nucleic acid molecules may be obtained from a subject. A sample (e.g., a biological sample or cell-free biological sample) suitable for use according to the methods provided herein may be any material comprising tissues, cells, degraded cells, nucleic acids, genes, gene fragments, expression products, gene expression products, and/or gene expression product fragments of an individual to be tested. A biological sample may be solid matter (e.g., biological tissue) or may be a fluid (e.g., a biological fluid). In general, a biological fluid may include any fluid associated with living organisms. Non-limiting examples of a biological sample include blood (or components of blood - e.g., white blood cells, red blood cells, platelets) obtained from any anatomical location (e.g., tissue, circulatory system, bone marrow) of a subject, cells obtained from any anatomical location of a subject, skin, heart, lung, kidney, breath, bone marrow, stool, semen, vaginal fluid, interstitial fluids derived from tumorous tissue, breast, pancreas, cerebral spinal fluid, tissue, throat swab, biopsy, placental fluid, amniotic fluid, liver, muscle, smooth muscle, bladder, gall bladder, colon, intestine, brain, cavity fluids, sputum, pus, microbiota, meconium, breast milk,



Date Reçue/Date Received 2023-12-22

prostate, esophagus, thyroid, serum, saliva, urine, gastric and digestive fluid, tears, ocular fluids, sweat, mucus, earwax, oil, glandular secretions, spinal fluid, hair, fingernails, skin cells, plasma, nasal swab or nasopharyngeal wash, spinal fluid, cord blood, emphatic fluids, and/or other excretions or body tissues. Methods for determining sample suitability and/or adequacy are provided. A sample may include, but is not limited to, blood, plasma, tissue, cells, degraded cells, cell-free nucleic acid molecules, and/or biological material from cells or derived from cells of an individual such as cell-free nucleic acid molecules. The sample may be a heterogeneous or homogeneous population of cells, tissues, or cell-free biological material. The biological sample may be obtained using any method that can provide a sample suitable for the analytical methods described herein.

**[00151]** A sample (e.g., a biological sample or cell-free biological sample) may undergo one or more processes in preparation for analysis, including, but not limited to, filtration, centrifugation, selective precipitation, permeabilization, isolation, agitation, heating, purification, and/or other processes. For example, a sample may be filtered to remove contaminants or other materials. In an example, a sample comprising cells may be processed to separate the cells from other material in the sample. Such a process may be used to prepare a sample comprising only cell-free nucleic acid molecules. Such a process may consist of a multi-step centrifugation process. Multiple samples, such as multiple samples from the same subject (e.g., obtained in the same or different manners from the same or different bodily locations, and/or obtained at the same or different times (e.g., seconds, minutes, hours, days, weeks, months, or years apart)) or multiple samples from different subjects may be obtained for analysis as described herein. In an example, the first sample is obtained from a subject before the subject undergoes a treatment regimen or procedure and the second sample is obtained from the subject after the subject undergoes the treatment regimen or procedure. Alternatively or in addition to, multiple samples may be obtained from the same subject at the same or approximately the same time. Different samples obtained from the same subject may be obtained in the same or different manner. For example, a first sample may be obtained via a biopsy and a second sample may be obtained via a blood draw. Samples obtained in different manners may be obtained by different medical professionals, using different techniques, at different times, and/or at different locations. Different samples obtained from the same subject may be obtained from different areas of a body. For example, a first sample may be obtained from a first area of a body (e.g., a first tissue) and a second sample may be obtained from a second area of the body (e.g., a second tissue).

**[00152]** A biological sample as used herein (e.g., a biological sample comprising one or more nucleic acid molecules) may not be purified when provided in a reaction vessel.



Date Reçue/Date Received 2023-12-22

Furthermore, for a biological sample comprising one or more nucleic acid molecules, the one or more nucleic acid molecules may not be extracted when the biological sample is provided to a reaction vessel. For example, ribonucleic acid (RNA) and/or deoxyribonucleic acid (DNA) molecules of a biological sample may not be extracted from the biological sample when providing the biological sample to a reaction vessel. Moreover, a target nucleic acid (e.g., a target RNA or target DNA molecules) present in a biological sample may not be concentrated when providing the biological sample to a reaction vessel. Alternatively, a biological sample may be purified and/or nucleic acid molecules may be isolated from other materials in the biological sample.

**[00153]** A biological sample as described herein may contain a target nucleic acid. As used herein, the terms "template nucleic acid", "target nucleic acid", "nucleic acid molecule," "nucleic acid sequence," "nucleic acid fragment," "oligonucleotide," "polynucleotide," and "nucleic acid" generally refer to polymeric forms of nucleotides of any length, such as deoxyribonucleotides (dNTPs) or ribonucleotides (rNTPs), or analogs thereof, and may be used interchangeably. Nucleic acids may have any three-dimensional structure, and may perform any function, known or unknown. A nucleic acid molecule may have a length of at least about 10 nucleic acid bases ("bases"), 20 bases, 30 bases, 40 bases, 50 bases, 100 bases, 200 bases, 300 bases, 400 bases, 500 bases, 1 kilobase (kb), 2 kb, 3, kb, 4 kb, 5 kb, 10 kb, 50 kb, or more. An oligonucleotide is typically composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); and thymine (T) (uracil (U) for thymine (T) when the polynucleotide is RNA). Oligonucleotides may include one or more nonstandard nucleotide(s), nucleotide analog(s) and/or modified nucleotides. Non-limiting examples of nucleic acids include DNA, RNA, genomic DNA (e.g., gDNA such as sheared gDNA), cell-free DNA (e.g., cfDNA), synthetic DNA/RNA, coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short- hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, complementary DNA (cDNA), recombinant nucleic acids, branched nucleic acids, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A nucleic acid may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be made before or following assembly of the nucleic acid. The sequence of nucleotides of a nucleic acid may be interrupted by non-nucleotide components. A nucleic acid may be further modified following polymerization, such as by conjugation or binding with a reporter agent.



Date Reçue/Date Received 2023-12-22

**[00154]** A target nucleic acid or sample nucleic acid as described herein may be amplified to generate an amplified product. A target nucleic acid may be a target RNA or a target DNA. When the target nucleic acid is a target RNA, the target RNA may be any type of RNA, including types of RNA described elsewhere herein. The target RNA may be viral RNA and/or tumor RNA. A viral RNA may be pathogenic to a subject. Non-limiting examples of pathogenic viral RNA include human immunodeficiency virus I (HIV I), human immunodeficiency virus n (HIV 11), orthomyxoviruses, Ebola virus. Dengue virus, influenza viruses (e.g., H1N1, H3N2, H7N9, or H5N1), hepesvirus, hepatitis A virus, hepatitis B virus, hepatitis C (e.g., armored RNA-HCV virus) virus, hepatitis D virus, hepatitis E virus, hepatitis G virus, Epstein-Barr virus, mononucleosis virus, cytomegalovirus, SARS virus, West Nile Fever virus, polio virus, and measles virus.

**[00155]** A biological sample may comprise a plurality of target nucleic acid molecules. For example, a biological sample may comprise a plurality of target nucleic acid molecules from a single subject. In another example, a biological sample may comprise a first target nucleic acid molecule from a first subject and a second target nucleic acid molecule from a second subject.

**[00156]** The term "nucleotide," as used herein, generally refers to a substance including a base (e.g., a nucleobase), sugar moiety, and phosphate moiety. A nucleotide may comprise a free base with attached phosphate groups. A substance including a base with three attached phosphate groups may be referred to as a nucleoside triphosphate. When a nucleotide is being added to a growing nucleic acid molecule strand, the formation of a phosphodiester bond between the proximal phosphate of the nucleotide to the growing chain may be accompanied by hydrolysis of a high-energy phosphate bond with release of the two distal phosphates as a pyrophosphate. The nucleotide may be naturally occurring or non-naturally occurring (e.g., a modified or engineered nucleotide).

**[00157]** The term "nucleotide analog," as used herein, may include, but is not limited to, a nucleotide that may or may not be a naturally occurring nucleotide. For example, a nucleotide analog may be derived from and/or include structural similarities to a canonical nucleotide such as adenine- (A), thymine- (T), cytosine- (C), uracil- (U), or guanine- (G) including nucleotide. A nucleotide analog may comprise one or more differences or modifications relative to a natural nucleotide. Examples of nucleotide analogs include inosine, diaminopurine, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, deazaxanthine, deazaguanine, isocytosine, isoguanine, 4- acetylcytosine, 5-(carboxyhydroxylmethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-



Date Reçue/Date Received 2023-12-22

dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-D46-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, 2,6-diaminopurine, ethynyl nucleotide bases, 1-propynyl nucleotide bases, azido nucleotide bases, phosphoroselenoate nucleic acids, and modified versions thereof (e.g., by oxidation, reduction, and/or addition of a substituent such as an alkyl, hydroxyalkyl, hydroxyl, or halogen moiety). Nucleic acid molecules (e.g., polynucleotides, double-stranded nucleic acid molecules, single-stranded nucleic acid molecules, primers, adapters, etc.) may be modified at the base moiety (e.g., at one or more atoms that typically are available to form a hydrogen bond with a complementary nucleotide and/or at one or more atoms that are not typically capable of forming a hydrogen bond with a complementary nucleotide), sugar moiety, or phosphate backbone. In some cases, a nucleotide may include a modification in its phosphate moiety, including a modification to a triphosphate moiety. Additional, non-limiting examples of modifications include phosphate chains of greater length (e.g., a phosphate chain having, 4, 5, 6, 7, 8, 9, 10 or more phosphate moieties), modifications with thiol moieties (e.g., alpha-thio triphosphate and beta-thiotriphosphates), and modifications with selenium moieties (e.g., phosphoroselenoate nucleic acids). A nucleotide or nucleotide analog may comprise a sugar selected from the group consisting of ribose, deoxyribose, and modified versions thereof (e.g., by oxidation, reduction, and/or addition of a substituent such as an alkyl, hydroxyalkyl, hydroxyl, or halogen moiety). A nucleotide analog may also comprise a modified linker moiety (e.g., in lieu of a phosphate moiety). Nucleotide analogs may also contain amine-modified groups, such as aminoallyl-dUTP (aa-dUTP) and aminohexylacrylamide-dCTP (aha-dCTP) to allow covalent attachment of amine reactive moieties, such as N-hydroxysuccinimide esters (NHS). Alternatives to standard DNA base pairs or RNA base pairs in the oligonucleotides of the present disclosure may provide, for example, higher density in bits per cubic mm, higher safety (resistant to accidental or purposeful synthesis of natural toxins), easier discrimination in photo-programmed polymerases, and/or lower secondary structure. Nucleotide analogs may be capable of reacting or bonding with detectable moieties for nucleotide detection.

**[00158]** The term "homopolymer," as used herein, generally refers to a polymer or a portion of a polymer comprising identical monomer units. A homopolymer may have a homopolymer sequence. A nucleic acid homopolymer may refer to a polynucleotide or an



Date Reçue/Date Received 2023-12-22

oligonucleotide comprising consecutive repetitions of a same nucleotide or any nucleotide variants thereof. For example, a homopolymer can be poly(dA), poly(dT), poly(dG), poly(dC), poly(rA), poly(U), poly(rG), or poly(rC). A homopolymer can be of any length. For example, the homopolymer can have a length of at least 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, or more nucleic acid bases. The homopolymer can have from 10 to 500, or 15 to 200, or 20 to 150 nucleic acid bases. The homopolymer can have a length of at most 500, 400, 300, 200, 100, 50, 40, 30, 20, 10, 5, 4, 3, or 2 nucleic acid bases. A molecule, such as a nucleic acid molecule, can include one or more homopolymer portions and one or more non-homopolymer portions. The molecule may be entirely formed of a homopolymer, multiple homopolymers, or a combination of homopolymers and non-homopolymers. In nucleic acid sequencing, multiple nucleotides can be incorporated into a homopolymeric region of a nucleic acid strand. Such nucleotides may be non-terminated to permit incorporation of consecutive nucleotides (e.g., during a single nucleotide flow).

**[00159]** The terms "amplifying," "amplification," and "nucleic acid amplification" are used interchangeably and generally refer to generating one or more copies of a nucleic acid or a template. For example, "amplification" of DNA generally refers to generating one or more copies of a DNA molecule. An amplicon may be a single-stranded or double-stranded nucleic acid molecule that is generated by an amplification procedure from a starting template nucleic acid molecule. Such an amplification procedure may include one or more cycles of an extension or ligation procedure. The amplicon may comprise a nucleic acid strand, of which at least a portion may be substantially identical or substantially complementary to at least a portion of the starting template. Where the starting template is a double-stranded nucleic acid molecule, an amplicon may comprise a nucleic acid strand that is substantially identical to at least a portion of one strand and is substantially complementary to at least a portion of either strand. The amplicon can be single-stranded or double-stranded irrespective of whether the initial template is single-stranded or double-stranded. Amplification of a nucleic acid may linear, exponential, or a combination thereof. Amplification may be emulsion based or may be non-emulsion based. Non-limiting examples of nucleic acid amplification methods include reverse transcription, primer extension, polymerase chain reaction (PCR), ligase chain reaction (LCR), helicase-dependent amplification, asymmetric amplification, rolling circle amplification, and multiple displacement amplification (MDA). Where PCR is used, any form of PCR may be used, with non-limiting examples that include real-time PCR, allele-specific PCR, assembly PCR, asymmetric PCR, digital PCR, emulsion PCR, dial-out PCR, helicase-dependent PCR, nested PCR, hot start PCR, inverse PCR, methylation-specific PCR, miniprimer PCR, multiplex PCR, nested PCR, overlap-



Date Reçue/Date Received 2023-12-22

extension PCR, thermal asymmetric interlaced PCR and touchdown PCR. Moreover, amplification can be conducted in a reaction mixture comprising various components (e.g., a primer(s), template, nucleotides, a polymerase, buffer components, co-factors, etc.) that participate or facilitate amplification. In some cases, the reaction mixture comprises a buffer that permits context independent incorporation of nucleotides. Non-limiting examples include magnesium-ion, manganese-ion and isocitrate buffers. Additional examples of such buffers are described in Tabor, S. et al. C.C. PNAS, 1989, 86, 4076-4080 and U.S. Patent Nos. 5,409,811 and 5,674,716 .

**[00160]** Amplification may be clonal amplification. The term "clonal," as used herein, generally refers to a population of nucleic acids for which a substantial portion (e.g., greater than about 50%, 60%, 70%, 80%, 90%, 95%, or 99%) of its members have sequences that are at least about 50%, 60%, 70%, 80%, 90%, 95%, or 99% identical to one another. Members of a clonal population of nucleic acid molecules may have sequence homology to one another. Such members may have sequence homology to a template nucleic acid molecule. The members of the clonal population may be double stranded or single stranded. Members of a population may not be 100% identical or complementary, e.g., "errors" may occur during the course of synthesis such that a minority of a given population may not have sequence homology with a majority of the population. For example, at least 50% of the members of a population may be substantially identical to each other or to a reference nucleic acid molecule (i.e., a molecule of defined sequence used as a basis for a sequence comparison). At least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or more of the members of a population may be substantially identical to the reference nucleic acid molecule. Two molecules may be considered substantially identical (or homologous) if the percent identity between the two molecules is at least 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 99.9% or greater. Two molecules may be considered substantially complementary if the percent complementarity between the two molecules is at least 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99%, 99.9% or greater. A low or insubstantial level of mixing of non-homologous nucleic acids may occur, and thus a clonal population may contain a minority of diverse nucleic acids (e.g., less than 30%, e.g., less than 10%).

**[00161]** Useful methods for clonal amplification from single molecules include rolling circle amplification (RCA) (Lizardi et al., Nat. Genet. 19:225-232 (1998)),

bridge PCR (Adams and Kron, Method for Performing Amplification of Nucleic Acid with Two Primers Bound to a Single Solid Support, Mosaic Technologies, Inc. (Winter Hill, Mass.); Whitehead Institute for Biomedical Research, Cambridge, Mass., (1997);

Adessi et al., Nucl. Acids Res. 28:E87 (2000); Pemov et al., Nucl. Acids Res. 33:e11(2005); or U.S. Pat. No. 5,641,658), polony generation (Mitra et al., Proc. Natl. Acad. Sci. USA 100:5926-5931 (2003); Mitra et al., Anal. Biochem. 320:55-65(2003)), and clonal amplification on beads using emulsions (Dressman et al., Proc. Natl. Acad. Sci. USA 100:8817-8822 (2003)), or ligation to bead-based adapter libraries (Brenner et al., Nat. Biotechnol. 18:630-634 (2000); Brenner et al., Proc. Natl. Acad. Sci. USA 97:1665-1670 (2000)); Reinartz, et al., Brief Funct. Genomic Proteomic 1:95-104 (2002)).

The enhanced signal-to-noise ratio provided by clonal amplification more than outweighs the disadvantages of the cyclic sequencing requirement.

**[00162]** The term "polymerizing enzyme" or "polymerase," as used herein, generally refers to any enzyme capable of catalyzing a polymerization reaction. A polymerizing enzyme may be used to extend a nucleic acid primer paired with a template strand by incorporation of nucleotides or nucleotide analogs. A polymerizing enzyme may add a new strand of DNA by extending the 3' end of an existing nucleotide chain, adding new nucleotides matched to the template strand one at a time via the creation of phosphodiester bonds. The polymerase used herein can have strand displacement activity or non-strand displacement activity. Examples of polymerases include, without limitation, a nucleic acid polymerase. An example polymerase is a Φ29 DNA polymerase or a derivative thereof. A polymerase can be a polymerization enzyme. In some cases, a transcriptase or a ligase is used (i.e., enzymes which catalyze the formation of a bond). Examples of polymerases include 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 Φ29 (phi29) 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, Tea polymerase, Tih polymerase, Tfi polymerase, Platinum Taq polymerases, Tbr polymerase, Tfl polymerase, Pfu-turbo 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 cases, the polymerase is a single subunit polymerase. The polymerase can have high processivity, namely the capability of the polymerase to consecutively incorporate nucleotides into a nucleic acid template without releasing the nucleic acid template. In some cases, a polymerase is a polymerase modified to accept

dideoxynucleotide triphosphates, such as for example, Taq polymerase having a 667Y mutation (see e.g., Tabor et al, PNAS, 1995, 92, 6339-6343).

In some cases, a polymerase is a polymerase having a modified nucleotide binding, which may be useful for nucleic acid sequencing, with non-limiting examples that include ThermoSequenas polymerase (GE Life Sciences), AmpliTaq FS (ThermoFisher) polymerase and Sequencing Pol polymerase (Jena Bioscience). In some cases, the polymerase is genetically engineered to have discrimination against dideoxynucleotides, such as for example, Sequenase DNA polymerase (ThermoFisher).

**[00163]** A polymerase may be Family A polymerase or a Family B DNA polymerase. Family A polymerases include, for example, Taq, Klenow, and Bst polymerases. Family B polymerases include, for example, Vent(exo-) and Therminator polymerases. Family B polymerases are known to accept more varied nucleotide substrates than Family A polymerases. Family A polymerases are used widely in sequencing by synthesis methods, likely due to their high processivity and fidelity.

**[00164]** The term "complementary sequence," as used herein, generally refers to a sequence that hybridizes to another sequence. Hybridization between two single-stranded nucleic acid molecules may involve the formation of a double-stranded structure that is stable under certain conditions. Two single-stranded polynucleotides may be considered to be hybridized if they are bonded to each other by two or more sequentially adjacent base pairings. A substantial proportion of nucleotides in one strand of a double-stranded structure may undergo Watson-Crick base-pairing with a nucleoside on the other strand. Hybridization may also include the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, and the like, that may be employed to reduce the degeneracy of probes, whether or not such pairing involves formation of hydrogen bonds.

**[00165]** The term "denaturation," as used herein, generally refers to separation of a double-stranded molecule (e.g., DNA) into single-stranded molecules. Denaturation may be complete or partial denaturation. In partial denaturation, a single-stranded region may form in a double-stranded molecule by denaturation of the two deoxyribonucleic acid (DNA) strands flanked by double-stranded regions in DNA.

**[00166]** The term "melting temperature" or "melting point," as used herein, generally refers to the temperature at which at least a portion of a strand of a nucleic acid molecule in a sample has separated from at least a portion of a complementary strand. The melting temperature may be the temperature at which a double-stranded nucleic acid molecule has partially or completely denatured. The melting temperature may refer to a temperature of a sequence among a plurality

of sequences of a given nucleic acid molecule, or a temperature of the plurality of sequences. Different regions of a double-stranded nucleic acid molecule may have different melting temperatures. For example, a double-stranded nucleic acid molecule may include a first region having a first melting point and a second region having a second melting point that is higher than the first melting point. Accordingly, different regions of a double-stranded nucleic acid molecule may melt (e.g., partially denature) at different temperatures. The melting point of a nucleic acid molecule or a region thereof (e.g., a nucleic acid sequence) may be determined experimentally (e.g., via a melt analysis or other procedure) or may be estimated based upon the sequence and length of the nucleic acid molecule. For example, a software program such as MELTING may be used to estimate a melting temperature for a nucleic acid sequence (Dumousseau M, Rodriguez N, Juty N, Le Novère N, MELTING, a flexible platform to predict the melting temperatures of nucleic acids. BMC Bioinformatics. 2012 May 16;13:101. doi: 10.1186/1471-2105-13-101). Accordingly, a melting point as described herein may be an estimated melting point. A true melting point of a nucleic acid sequence may vary based upon the sequences or lack thereof adjacent to the nucleic acid sequence of interest as well as other factors.

**[00167]** The term "sequencing," as used herein, generally refers to a process for generating or identifying a sequence of a biological molecule, such as a nucleic acid molecule or a polypeptide. Such sequence may be a nucleic acid sequence, which may include a sequence of nucleic acid bases (e.g., nucleobases). Sequencing may be, for example, single molecule sequencing, sequencing by synthesis, sequencing by hybridization, or sequencing by ligation. Sequencing may be performed using template nucleic acid molecules immobilized on a support, such as a flow cell or one or more beads. A sequencing assay may yield one or more sequencing reads corresponding to one or more template nucleic acid molecules.

**[00168]** The term "read," as used herein, generally refers to a nucleic acid sequence, such as a sequencing read. A sequencing read may be an inferred sequence of nucleic acid bases (e.g., nucleotides) or base pairs obtained via a nucleic acid sequencing assay. A sequencing read may be generated by a nucleic acid sequencer, such as a massively parallel array sequencer (e.g., Illumina or Pacific Biosciences of California). A sequencing read may correspond to a portion, or in some cases all, of a genome of a subject. A sequencing read may be part of a collection of sequencing reads, which may be combined through, for example, alignment (e.g., to a reference genome), to yield a sequence of a genome of a subject.

**[00169]** The term "detector," as used herein, generally refers to a device that is capable of detecting or measuring a signal, such as a signal indicative of the presence or absence of an



Date Reçue/Date Received 2023-12-22

incorporated nucleotide or nucleotide analog. A detector may include optical and/or electronic components that may detect and/or measure signals. Non-limiting examples of detection methods involving a detector include optical detection, spectroscopic detection, electrostatic detection, and electrochemical detection. Optical detection methods include, but are not limited to, fluorimetry and UV-vis light absorbance. Spectroscopic detection methods include, but are not limited to, mass spectrometry, nuclear magnetic resonance (NMR) spectroscopy, and infrared spectroscopy. Electrostatic detection methods include, but are not limited to, gel based techniques, such as, for example, gel electrophoresis. Electrochemical detection methods include, but are not limited to, electrochemical detection of amplified product after high-performance liquid chromatography separation of the amplified products.

**[00170]** The term "support", as used herein, generally refers to any solid or semi-solid article on which reagents such as nucleic acid molecules may be immobilized. Nucleic acid molecules may be synthesized, attached, ligated, or otherwise immobilized. Nucleic acid molecules may be immobilized on a support by any method including, but not limited to, physical adsorption, by ionic or covalent bond formation, or combinations thereof. A support may be 2-dimensional (e.g., a planar 2D support) or 3-dimensional. In some cases, a support may be a component of a flow cell and/or may be included within or adapted to be received by a sequencing instrument. A support may include a polymer, a glass, or a metallic material. Examples of supports include a membrane, a planar support, a microtiter plate, a bead (e.g., a magnetic bead), a filter, a test strip, a slide, a cover slip, and a test tube. A support may comprise organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide (e.g., polyacrylamide gel), as well as co-polymers and grafts thereof. A support may comprise latex or dextran. A support may also be inorganic, such as glass, silica, gold, controlled-pore-glass (CPG), or reverse-phase silica. The configuration of a support may be, for example, in the form of beads, spheres, particles, granules, a gel, a porous matrix, or a support. In some cases, a support may be a single solid or semi-solid article (e.g., a single particle), while in other cases a support may comprise a plurality of solid or semi-solid articles (e.g., a collection of particles). Supports may be planar, substantially planar, or non-planar. Supports may be porous or non-porous, and may have swelling or non-swelling characteristics. A support may be shaped to comprise one or more wells, depressions, or other containers, vessels, features, or locations. A plurality of supports may be configured in an array at various locations. A support may be addressable (e.g., for robotic delivery of reagents), or by detection approaches, such as scanning by laser illumination and confocal or deflective light gathering. For example, a support may be in optical and/or physical communication with a detector. Alternatively, a support may



Date Reçue/Date Received 2023-12-22

be physically separated from a detector by a distance. An amplification support (e.g., a bead) can be placed within or on another support (e.g., within a well of a second support).

**[00171]** The term "label," as used herein, generally refers to a moiety that is capable of coupling with a species, such as, for example a nucleotide analog. A label may include an affinity moiety. In some cases, a label may be a detectable label that emits a signal (or reduces an already emitted signal) that can be detected. In some cases, such a signal may be indicative of incorporation of one or more nucleotides or nucleotide analogs. In some cases, a label may be coupled to a nucleotide or nucleotide analog, which nucleotide or nucleotide analog may be used in a primer extension reaction. In some cases, the label may be coupled to a nucleotide analog after a primer extension reaction. The label, in some cases, may be reactive specifically with a nucleotide or nucleotide analog. Coupling may be covalent or non-covalent (e.g., via ionic interactions, Van der Waals forces, etc.). In some cases, coupling may be via a linker, which may be cleavable, such as photo-cleavable (e.g., cleavable under ultra-violet light), chemically-cleavable (e.g., via a reducing agent, such as dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP), tris(hydroxypropyl)phosphine (THP) or enzymatically cleavable (e.g., via an esterase, lipase, peptidase or protease). In some cases, the label may be luminescent; that is, fluorescent or phosphorescent. For example, the label may be or comprise a fluorescent moiety (e.g., a dye). Dyes and labels may be incorporated into nucleic acid sequences. Dyes and labels may also be incorporated into or attached to linkers, such as linkers for linking one or more beads to one another. For example, labels such as fluorescent moieties may be linked to nucleotides or nucleotide analogs via a linker (e.g., as described herein). Non-limiting examples of dyes include SYBR green, SYBR blue, DAPI, propidium iodine, Hoechst, SYBR gold, ethidium bromide, acridine, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, phenanthridines and acridines, propidium iodide, hexidium iodide, dihydroethidium, ethidium homodimer-1 and -2, ethidium monoazide, ACMA, Hoechst 33258, Hoechst 33342, Hoechst 34580, DAPI, acridine orange, 7-AAD, actinomycin D, LDS751, hydroxystilbamidine, SYTOX Blue, SYTOX Green, SYTOX Orange, POPO-1, POPO-3, YOYO-1, YOYO-3, TOTO-1, TOTO-3, JOJO-1, LOLO-1, BOBO-1, BOBO-3, PO-PRO-1, PO-PRO-3, BO-PRO-1, BO-PRO-3, TO-PRO-1, TO-PRO-3, TO-PRO-5, JO-PRO-1, LO-PRO-1, YO-PRO-1, YO-PRO-3, PicoGreen, OliGreen, RiboGreen, SYBR Gold, SYBR Green I, SYBR Green II, SYBR DX, SYTO labels (e.g., SYTO-40, -41, -42, -43, -44, and -45 (blue); SYTO-13, -16, -24, -21, -23, -12, -11, -20, -22, -15, -14, and -25 (green); SYTO-81, -80, -82, -83, -84, and-85 (orange); and SYTO-64, -17, -59, -61, -62, -60, and -63 (red)), fluorescein, fluorescein



Date Reçue/Date Received 2023-12-22

isothiocyanate (FITC), tetramethyl rhodamine isothiocyanate (TRITC), rhodamine, tetramethyl rhodamine, R-phycoerythrin, Cy-2, Cy-3, Cy-3.5, Cy-5, Cy5.5, Cy-7, Texas Red, Phar-Red, allophycocyanin (APC), Sybr Green I, Sybr Green II, Sybr Gold, CellTracker Green, 7-AAD, ethidium homodimer I, ethidium homodimer II, ethidium homodimer III, ethidium bromide, umbelliferone, eosin, green fluorescent protein, erythrosin, coumarin, methyl coumarin, pyrene, malachite green, stilbene, lucifer yellow, cascade blue, dichlorotriazinylamine fluorescein, dansyl chloride, fluorescent lanthanide complexes such as those including europium and terbium, carboxy tetrachloro fluorescein, 5 and/or 6-carboxy fluorescein (FAM), VIC, 5- (or 6-) iodoacetamidofluorescein, 5-{[2(and 3)-5-(Acetylmercapto)-succinyl]amino} fluorescein (SAMSA-fluorescein), lissamine rhodamine B sulfonyl chloride, 5 and/or 6 carboxy rhodamine (ROX), 7-amino-methyl-coumarin, 7-Amino-4-methylcoumarin-3-acetic acid (AMCA), BODIPY fluorophores, 8-methoxypyrene-1,3,6-trisulfonic acid trisodium salt, 3,6-Disulfonate-4-amino-naphthalimide, phycobiliproteins, AlexaFluor labels (e.g., AlexaFluor 350, 405, 430, 488, 532, 546, 555, 568, 594, 610, 633, 635, 647, 660, 680, 700, 750, and 790 dyes), DyLight labels (e.g., DyLight 350, 405, 488, 550, 594, 633, 650, 680, 755, and 800 dyes), Black Hole Quencher Dyes (Biosearch Technologies) (e.g., BH1-0, BHQ-1, BHQ-3, and BHQ-10), QSY Dye fluorescent quenchers (Molecular Probes/Invitrogen) (e.g., QSY7, QSY9, QSY21, and QSY35), Dabcyl, Dabsyl, Cy5Q, Cy7Q, Dark Cyanine dyes (GE Healthcare), Dy-Quenchers (Dyomics) (e.g., DYQ-660 and DYQ-661), ATTO fluorescent quenchers (ATTO-TEC GmbH) (e.g., ATTO 540Q, ATTO 580Q, ATTO 612Q, Atto532 [e.g., Atto 532 succinimidyl ester], and Atto633), and other fluorophores and/or quenchers. Additional examples are included in structures provided herein. Dyes included in structures provided herein are contemplated for use in combination with any linker and substrate described herein. A fluorescent dye may be excited by application of energy corresponding to the visible region of the electromagnetic spectrum (e.g., between about 430-770 nanometers (nm)). Excitation may be done using any useful apparatus, such as a laser and/or light emitting diode. Optical elements including, but not limited to, mirrors, waveplates, filters, monochromaters, gratings, beam splitters, and lenses may be used to direct light to or from a fluorescent dye. A fluorescent dye may emit light (e.g., fluoresce) in the visible region of the electromagnetic spectrum ((e.g., between about 430-770 nm). A fluorescent dye may be excited over a single wavelength or a range of wavelengths. A fluorescent dye may be excitable by light in the red region of the visible portion of the electromagnetic spectrum (about 625-740 nm) (e.g., have an excitation maximum in the red region of the visible portion of the electromagnetic spectrum). Alternatively or in addition to, fluorescent dye may be excitable by light in the green region of the visible portion of the electromagnetic spectrum (about 500-565



Date Reçue/Date Received 2023-12-22

nm) (e.g., have an excitation maximum in the green region of the visible portion of the electromagnetic spectrum). A fluorescent dye may emit signal in the red region of the visible portion of the electromagnetic spectrum (about 625-740 nm) (e.g., have an emission maximum in the red region of the visible portion of the electromagnetic spectrum). Alternatively or in addition to, fluorescent dye may emit signal in the green region of the visible portion of the electromagnetic spectrum (about 500-565 nm) (e.g., have an emission maximum in the green region of the visible portion of the electromagnetic spectrum).

**[00172]** Labels may be quencher molecules. The term "quencher," as used herein, generally refers to molecules that may be energy acceptors. A quencher may be a molecule that can reduce an emitted signal. For example, a template nucleic acid molecule may be designed to emit a detectable signal. Incorporation of a nucleotide or nucleotide analog comprising a quencher can reduce or eliminate the signal, which reduction or elimination is then detected. Luminescence from labels (e.g., fluorescent moieties, such as fluorescent moieties linked to nucleotides or nucleotide analogs) may also be quenched (e.g., by incorporation of other nucleotides that may or may not comprise labels). In some cases, as described elsewhere herein, labelling with a quencher can occur after nucleotide or nucleotide analog incorporation (e.g., after incorporation of a nucleotide or nucleotide analog comprising a fluorescent moiety). In some cases, the label may be a type that does not self-quench or exhibit proximity quenching. Non-limiting examples of a label type that does not self-quench or exhibit proximity quenching include Bimane derivatives such as Monobromobimane. The term "proximity quenching," as used herein, generally refers to a phenomenon where one or more dyes near each other may exhibit lower fluorescence as compared to the fluorescence they exhibit individually. In some cases, the dye may be subject to proximity quenching wherein the donor dye and acceptor dye are within 1 nm to 50 nm of each other. Examples of quenchers include, but are not limited to, Black Hole Quencher Dyes (Biosearch Technologies) (e.g., BH1-0, BHQ-1, BHQ-3, and BHQ-10), QSY Dye fluorescent quenchers (Molecular Probes/Invitrogen) (e.g., QSY7, QSY9, QSY21, and QSY35), Dabcyl, Dabsyl, Cy5Q, Cy7Q, Dark Cyanine dyes (GE Healthcare), Dy-Quenchers (Dyomics) (e.g., DYQ-660 and DYQ-661), and ATTO fluorescent quenchers (ATTO-TEC GmbH) (e.g., ATTO 540Q, ATTO 580Q, and ATTO 612Q). Fluorophore donor molecules may be used in conjunction with a quencher. Examples of fluorophore donor molecules that can be used in conjunction with quenchers include, but are not limited to, fluorophores such as Cy3B, Cy3, or Cy5; Dy-Quenchers (Dyomics) (e.g., DYQ-660 and DYQ-661); and ATTO fluorescent quenchers (ATTO-TEC GmbH) (e.g., ATTO 540Q, 580Q, and 612Q).



Date Reçue/Date Received 2023-12-22

**[00173]** The term "labeling fraction," as used herein, generally refers to the ratio of dye-labeled nucleotide or nucleotide analog to natural/unlabeled nucleotide or nucleotide analog of a single canonical type in a flow solution. The labeling fraction can be expressed as the concentration of the labeled nucleotide or nucleotide analog divided by the sum of the concentrations of labeled and unlabeled nucleotide or nucleotide analog. The labeling fraction may be expressed as a % of labeled nucleotides included in a solution (e.g., a nucleotide flow). The labeling fraction may be at least about 0.5%, 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or higher. For example, the labeling fraction may be at least about 20%. The labeling fraction may be about 100%. The labeling fraction may also be expressed as a ratio of labeled nucleotides to unlabeled nucleotides included in a solution. For example, the ratio of labeled nucleotides to unlabeled nucleotides may be at least about 1:5, 1:4, 1:3, 1:2, 1:1, 2:1, 3:1, 4:1, 5:1, or higher. For example, the ratio of labeled nucleotides to unlabeled nucleotides may be at least about 1:4. For example, the ratio of labeled nucleotides to unlabeled nucleotides may be at least about 1:1. For example, the ratio of labeled nucleotides to unlabeled nucleotides may be at least about 5:1.

**[00174]** The term "labeled fraction," as used herein, generally refers to the actual fraction of labeled nucleic acid (e.g., DNA) resulting after treatment of a primer-template with a mixture of the dye-labeled and natural nucleotide or nucleotide analog. The labeled fraction may be about the same as the labeling fraction. For example, if 20% of nucleotides in a nucleotide flow are labeled, about 20% of nucleotides incorporated into a growing nucleic acid strand (e.g., during nucleic acid sequencing) may be labeled. Alternatively, the labeled fraction may be greater than the labeled fraction. For example, if 20% of nucleotides in a nucleotide flow are labeled, greater than 20% of nucleotides incorporated into a growing nucleic acid strand (e.g., during nucleic acid sequencing) may be labeled. Alternatively, the labeled fraction may be less than the labeled fraction. For example, if 20% of nucleotides in a nucleotide flow are labeled, less than 20% of nucleotides incorporated into a growing nucleic acid strand (e.g., during nucleic acid sequencing) may be labeled.

**[00175]** When a solution including less than 100% labeled nucleotides or nucleotide analogs is used in an incorporation process such as a sequencing process (e.g., as described herein), both labeled ("bright") and unlabeled ("dark") nucleotides or nucleotide analogs may be incorporated into a growing nucleic acid strand. The term "tolerance," as used herein, generally refers to the ratio of the labeled fraction (e.g., "bright" incorporated fraction) to the labeling fraction (e.g., "bright" fraction in solution). For example, if a labeling fraction of 0.2 is used resulting in a labeled fraction of 0.4 the tolerance is 2. Similarly, if an incorporation process



Date Reçue/Date Received 2023-12-22

such as a sequencing process is performed using 2.5% labeled fraction in solution ($b_f$, bright solution fraction) and 5% is labeled ($b_i$, bright incorporated fraction), the tolerance may be 2 (e.g., tolerance). This model may be linear for low labeling fractions (e.g., 10% or lower labeling fraction). For higher labeling fractions, tolerance may take into account competing dark incorporation. Tolerance may refer to a comparison of the ratio of bright incorporated fraction to dark incorporated fraction ($b_i/d_i$) to the ratio of bright solution fraction to dark solution fraction ($b_f/d_f$):

**[00176]** $Tolerance = \frac{b_i/d_i}{b_f/d_f}$ where $d_i = 1 - b_i$ (e.g., dark incorporated fraction and bright incorporated fraction sum to 1 assuming 100% bright fraction is normalized to 1)

**[00177]** Though $d_i$ cannot easily be measured, $b_i$, the bright incorporated fraction, can be measured (e.g., as described herein) and used to determine tolerance by fitting a curve of bright solution fraction ($b_f$) vs. bright incorporated fraction ($b_i$):

**[00178]** $b_i = \frac{tol(b_f/d_f)}{1+tol(b_f/d_f)}$

**[00179]** A "positive" tolerance number (>1) indicates that at 50% labeling fraction, more than 50% is labeled. A "negative" tolerance number (<1) indicates that at 50% labeling fraction, less than 50% is labeled.

**[00180]** The term "context," as used herein, generally refers to the sequence of the neighboring nucleotides, or context, has been observed to affect the tolerance in an incorporation reaction. The nature of the enzyme, the pH and other factors may also affect the tolerance. Reducing context effects to a minimum greatly simplifies base determination.

**[00181]** The term "scar," as used herein, generally refers to a residue left on a previously labeled nucleotide or nucleotide analog after cleavage of an optical (e.g., fluorescent) dye and, optionally, all or a portion of a linker attaching the optical dye to the nucleotide or nucleotide analog. Examples of scars include, but are not limited to, hydroxyl moieties (e.g., resulting from cleavage of an azidomethyl group, hydrocarbyldithiomethyl linkage, or 2-nitrobenzyloxy linkage), thiol moieties (e.g., resulting from cleavage of a disulfide linkage), and benzyl moieties. For example, a scar may comprise an aromatic group such as a phenyl or benzyl group. The size and nature of a scar may affect subsequent incorporations.

**[00182]** The term "misincorporation," as used herein, generally refers to occurrences when the DNA polymerase incorporates a nucleotide, either labeled or unlabeled, that is not the correct Watson-Crick partner for the template base. Misincorporation can occur more frequently in methods that lack competition of all four bases in an incorporation event, and leads to strand loss, and thus limits the read length of a sequencing method.



Date Reçue/Date Received 2023-12-22

**[00183]** The term "mispair extension", as used herein, generally refers to occurrences when the DNA polymerase incorporates a nucleotide, either labeled or unlabeled, that is not the correct Watson-Crick partner for the template base, then subsequently incorporates the correct Watson-Crick partner for the following base. Mispair extension generally results in lead phasing and limits the read length of a sequencing method.

**[00184]** Regarding quenching, dye-dye quenching between two dye moieties linked to different nucleotides (e.g., adjacent nucleotides in a growing nucleic acid strand, or nucleotides in a nucleic acid strand that are separated by one or more other nucleotides) may be strongly dependent on the distance between the two dye moieties. The distance between two dye moieties may be at least partially dependent on the properties of linkers connecting the two dye moieties to respective nucleotides or nucleotide analogs, including the linker compositions and functional lengths. Features of the linkers, including composition and functional length, may be affected by temperature, solvent, pH and salt concentration (e.g., within a solution). Quenching may also vary based on the nature of the dyes used. Quenching may also take place between dye moieties and nucleobase moieties (e.g., between a fluorescent dye and a nucleobase of a nucleotide with which it is associated). Controlling quenching phenomena may be a key feature of the methods described herein.

**[00185]** Regarding flows, a nucleotide flow can consist of a mixture of labeled and unlabeled nucleotides or nucleotide analogs (e.g., nucleotides or nucleotide analogs of a single canonical type). For example, a solution comprising a plurality of optically (e.g., fluorescently) labeled nucleotides and a plurality of unlabeled nucleotides may be contacted with, e.g., a sequencing template (as described herein). The plurality of optically labeled nucleotides and a plurality of unlabeled nucleotides may each comprise the same canonical nucleotide or nucleotide analog. A flow may include only labeled nucleotides or nucleotide analogs. Alternatively, a flow may include only unlabeled nucleotides or nucleotide analogs. A flow may include a mixture of nucleotide or nucleotide analogs of different types (e.g., A and G).

**[00186]** A wash flow (e.g., a solution comprising a buffer) may be used to remove any nucleotides that are not incorporated into a nucleic acid complex (e.g., a sequencing template, as described herein). A cleavage flow (e.g., a solution comprising a cleavage reagent) may be used to remove dye moieties (e.g., fluorescent dye moieties) from optically (e.g., fluorescently) labeled nucleotides or nucleotide analogs. In some cases, different dyes (e.g., fluorescent dyes) may be removable using different cleavage reagents. In other cases, different dyes (e.g., fluorescent dyes) may be removable using the same cleavage reagents. Cleavage of dye moieties



Date Reçue/Date Received 2023-12-22

from optically labeled nucleotides or nucleotide analogs may comprise cleavage of all or a portion of a linker connecting a nucleotide or nucleotide analog to a dye moiety.

**[00187]** The term "cycle," as used herein, generally refers to a process in which a nucleotide flow, a wash flow, and a cleavage flow corresponding to each canonical nucleotide (e.g., dATP, dCTP, dGTP, and dTTP or dUTP, or modified versions thereof) are used (e.g., provided to a sequencing template, as described herein). Multiple cycles may be used to sequence and/or amplify a nucleic acid molecule. The order of nucleotide flows can be varied.

**[00188]** Phasing can be lead or lag phasing. Lead phasing generally refers to the phenomenon in which a population of strands show incorporation of a nucleotide a flow ahead of the expected cycle (*e.g.*, due to contamination in the system). Lag phasing refers to the phenomenon in which a population of strands shows incorporation of a nucleotide a flow behind the expected cycle (*e.g.*, due to incompletion of extension in an earlier cycle).

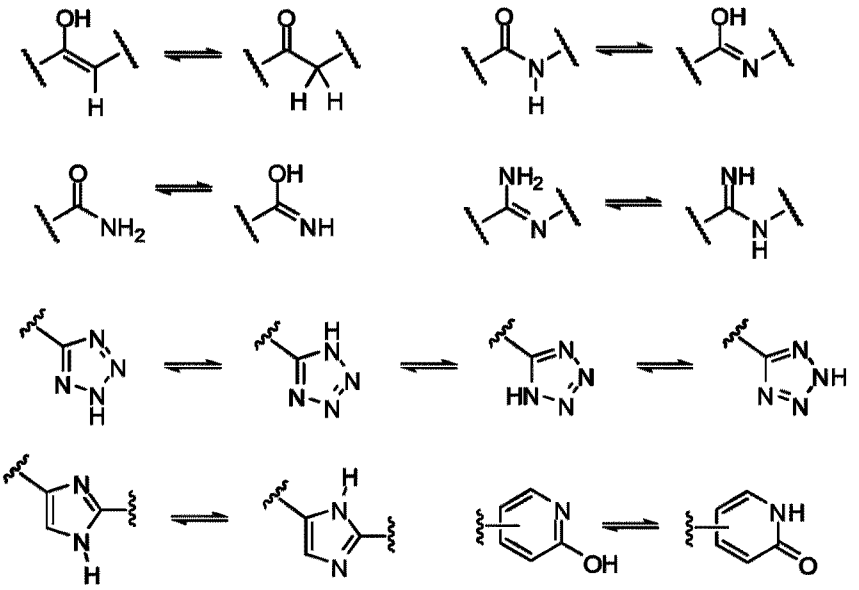
**[00189]** Compounds and chemical moieties described herein, including linkers, may contain one or more asymmetric centers and thus give rise to enantiomers, diastereomers, and other stereoisomeric forms that are defined, in terms of absolute stereochemistry, as (*R*)- or (*S*)-, and, in terms of relative stereochemistry, as (*D*)- or (*L*)-. The D/L system relates molecules to the chiral molecule glyceraldehyde and is commonly used to describe biological molecules including amino acids. Unless stated otherwise, it is intended that all stereoisomeric forms of the compounds disclosed herein are contemplated by this disclosure. When the compounds described herein contain alkene double bonds, and unless specified otherwise, it is intended that this disclosure includes both *E* and Z geometric isomers (*e.g., cis* or *trans*.) Likewise, all possible isomers, as well as their racemic and optically pure forms, and all tautomeric forms are also intended to be included. The term "geometric isomer" refers to *E* or Z geometric isomers (*e.g., cis* or *trans*) of an alkene double bond. The term "positional isomer" refers to structural isomers around a central ring, such as *ortho*-, *meta*-, and *para*- isomers around a phenyl ring. Separation of stereoisomers may be performed by chromatography or by forming diastereomers and separating by recrystallization, or chromatography, or any combination thereof. (Jean Jacques, Andre Collet, Samuel H. Wilen, "Enantiomers, Racemates and Resolutions", John Wiley and Sons, Inc., 1981, herein incorporated by reference for this disclosure). Stereoisomers may also be obtained by stereoselective synthesis.

**[00190]** Compounds and chemical moieties described herein, including linkers, may exist as tautomers. A "tautomer" refers to a molecule wherein a proton shift from one atom of a molecule to another atom of the same molecule is possible. In circumstances where tautomerization is possible, a chemical equilibrium of the tautomers will exist. Unless otherwise



Date Reçue/Date Received 2023-12-22

stated, chemical structures depicted herein are intended to include structures which are different tautomers of the structures depicted. For example, the chemical structure depicted with an enol moiety also includes the keto tautomer form of the enol moiety. The exact ratio of the tautomers depends on several factors, including physical state, temperature, solvent, and pH. Some examples of tautomeric equilibrium include:



**[00191]** Compounds and chemical moieties described herein, including linkers and dyes, may be provided in different enriched isotopic forms. For example, compounds may be enriched in the content of $^{2}H$, $^{3}H$, $^{11}C$, $^{13}C$ and/or $^{14}C$. For example, a linker, substrate (e.g., nucleotide or nucleotide analog), or dye may be deuterated in at least one position. In some examples, a linker, substrate (e.g., nucleotide or nucleotide analog), or dye may be fully deuterated. Such deuterated forms can be made by the procedure described in U.S. Patent Nos. 5,846,514 and 6,334,997. As described in U.S. Patent Nos. 5,846,514 and 6,334,997, deuteration can improve the metabolic stability and or efficacy, thus increasing the duration of action of drugs.

**[0001]** Unless otherwise stated, structures depicted and described herein are intended to include compounds which differ only in the presence of one or more isotopically enriched atoms. For example, compounds and chemical moieties having the present structures except for the replacement of a hydrogen by a deuterium or tritium, or the replacement of a carbon by $^{13}C$- or $^{14}C$-enriched carbon are within the scope of the present disclosure.

**[0002]** The compounds and chemical moieties of the present disclosure may contain unnatural proportions of atomic isotopes at one or more atoms that constitute such compounds. For example, a compound or chemical moiety such as a linker, substrate (e.g., nucleotide or



E

nucleotide analog), or dye, or a combination thereof, may be labeled with one or more isotopes, such as deuterium ($^{2}H$), tritium ($^{3}H$), iodine-125 ($^{125}I$) or carbon-14 ($^{14}C$). Isotopic substitution with $^{2}H$, $^{11}C$, $^{13}C$, $^{14}C$, $^{15}C$, $^{12}N$, $^{13}N$, $^{15}N$, $^{16}N$, $^{16}O$, $^{17}O$, $^{14}F$, $^{15}F$, $^{16}F$, $^{17}F$, $^{18}F$, $^{33}S$, $^{34}S$, $^{35}S$, $^{36}S$, $^{35}Cl$, $^{37}Cl$, $^{79}Br$, $^{81}Br$, and $^{125}I$ are all contemplated. All isotopic variations of the compounds and chemical moieties described herein, whether radioactive or not, are encompassed within the scope of the present disclosure.

**Linkers for optical detection**

**[00192]** The present disclosure provides an optical (e.g., fluorescent) labeling reagent comprising a dye (e.g., fluorescent dye) and a linker that is connected to the dye and capable of associating with a substrate to be optically (e.g., fluorescently) labeled. The substrate can be any suitable molecule, analyte, cell, tissue or surface that is to be optically labeled. Examples include cells, including eukaryotic cells, prokaryotic cells, healthy cells, and diseased cells; cellular receptors; antibodies; proteins; lipids; metabolites; probes; reagents; nucleotides and nucleotide analogs; and nucleic acid molecules. The association between the linker and the substrate can be any suitable association including a covalent or non-covalent bond, such as an association between a purine-containing nucleotide and a pyrimidine-containing nucleotide in a nucleic acid molecule. In some cases, such an association may be a biotin-avidin interaction. In other cases, the association between the linker and the substrate may be via a propargylamino moiety. In some cases, the association between the linker and the substrate may be via an amide bond (e.g., a peptide bond).

**[00193]** A linker can be semi-rigid. The semi-rigid nature of the linker can be most readily achieved by use of structure that comprises a series of ring systems (e.g., aliphatic and aromatic rings). As used herein, a ring (e.g., ring structure) is a cyclic moiety comprising any number of atoms connected in a closed, essentially circular fashion, as used in the field of organic chemistry. A ring may be defined by any number of atoms. For example, a ring may include between 3-12 atoms, such as between 3-12 carbon atoms. In certain examples, a ring may be a five-membered ring (*i.e.*, a pentagon) or a six-membered ring (*i.e.*, a hexagon). A ring can be aromatic or non-aromatic. A ring may be aliphatic. A ring may comprise one or more double bonds.

**[00194]** A ring (e.g., ring structure) may be a component of a ring system that may comprise one or more ring structures (e.g., a multi-cycle system). For example, a ring system may comprise a monocycle. In another example, a ring system may be a bicycle or bridged system. A ring structure may be a carbocycle or component thereof formed of carbon atoms. A



Date Reçue/Date Received 2023-12-22

carbocycle may be a saturated, unsaturated, or aromatic ring in which each atom of the ring is carbon. A carbocycle includes 3- to 10-membered monocyclic rings, 4- to 12-membered bicyclic rings (e.g., 6- to 12-membered bicyclic rings), and 5- to 12-membered bridged rings. Each ring of a bicyclic carbocycle may be selected from saturated, unsaturated, and aromatic rings. For example, a bicyclic carbocycle may include an aromatic ring (e.g., phenyl) fused to a saturated or unsaturated ring (e.g., cyclohexane, cyclopentane, or cyclohexene). A bicyclic carbocycle may include any combination of saturated, unsaturated, and aromatic bicyclic rings, as valence permits. A bicyclic carbocycle may include any combination of ring sizes such as 4-5 fused ring systems, 5-5 fused ring systems, 5-6 fused ring systems, and 6-6 fused ring systems. A carbocycle may be, for example, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cyclohexenyl, adamantyl, phenyl, indanyl, or naphthyl. A saturated carbocycle includes no multiple bonds (e.g., double or triple bonds). A saturated carbocycle may be, for example, cyclopropane, cyclobutane, cyclopentane, or cyclohexane. An unsaturated carbocycle includes at least one multiple bond (e.g., double or triple bond) but is not an aromatic carbocycle. An unsaturated carbocycle may be, for example, cyclohexadiene, cyclohexene, or cyclopentene. Other examples of carbocycles include, but are not limited to, cyclopropane, cyclobutane, cyclopentane, cyclopentadiene, cyclohexane, cycloheptane, cycloheptene, naphthalene, and adamantine. An aromatic carbocycle (e.g., aryl moiety) may be, for example, phenyl, naphthyl, or dihydronaphthyl.

**[00195]** In some cases, a ring may include one or more heteroatoms, such as one or more oxygen, nitrogen, silicon, phosphorous, boron, or sulfur atoms. A ring may be a heterocycle or component thereof including one or more heteroatoms. A heterocycle may be a saturated, unsaturated, or aromatic ring in which at least one atom is a heteroatom. A heteroatom includes 3- to 10-membered monocyclic rings, 6- to 12-membered bicyclic rings, and 6- to 12-membered bridged rings. A bicyclic heterocycle may include any combination of saturated, unsaturated and aromatic bicyclic rings, as valence permits. For example, a heteroaromatic ring (e.g., pyridyl) may be fused to a saturated or unsaturated ring (e.g., cyclohexane, cyclopentane, morpholine, piperidine or cyclohexene). A bicyclic heterocycle may include any combination of ring sizes such as 4-5 fused ring systems, 5-5 fused ring systems, 5-6 fused ring systems, and 6-6 fused ring systems. An unsaturated heterocycle includes at least one multiple bond (e.g., double or triple bond) but is not an aromatic heterocycle. An unsaturated heterocycle may be, for example, dihydropyrrole, dihydrofuran, oxazoline, pyrazoline, or dihydropyridine. Additional examples of heterocycles include, but are not limited to, indole, benzothiophene, benzthiazole, benzoxazole, benzimidazole, oxazolopyridine, imidazopyridine, thiazolopyridine, furan, oxazole,



Date Reçue/Date Received 2023-12-22

pyrrole, pyrazole, imidazole, thiophene, thiazole, isothiazole, and isoxazole. A heteroaryl moiety may be an aromatic single ring structure, such as a 5- to 7-membered ring, including at least one heteroatom, such as one to four heteroatoms. Alternatively, a heteroaryl moiety may be a polycyclic ring system having two or more cyclic rings in which two or more atoms are common to two adjoining rings wherein at least one of the rings is heteroaromatic. Heteroaryl groups include, for example, pyrrole, furan, thiophene, imidazole, oxazole, thiazole, pyrazole, pyridine, pyrazine, pyridazine, and pyrimidine, and the like.

**[00196]** A ring can be substituted or un-substituted. A substituent replaces a hydrogen atom on one or more atoms of a ring or a substitutable heteroatom of a ring (e.g., NH or $NH_2$). Substitution is in accordance with permitted valence of the various components of the ring system and provides a stable compound (e.g., a compound that does not undergo spontaneous transformation by, for example, rearrangement, elimination, or cyclization). A substituent may replace a single hydrogen atom or multiple hydrogen atoms (e.g., on the same ring atom or different ring atoms). A substituent on a ring may be, for example, halogen, hydroxy, oxo, thioxo, thiol, amido, amino, carboxy, nitrilo, cyano, nitro, imino, oximo, hydrazino, alkoxy, alkenyl, alkynyl, aryl, aralkyl, aralkenyl, aralkynyl, cycloalkyl, cycloalkylalkyl, alkylcycloalkyl, heterocycloalkyl, heterocycyl, alkylheterocycyl, or any other useful substituent. A substituent may be water-soluble. Examples of water-soluble substituents include, but are not limited to, a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester.

**[00197]** A linker can have any number of rings, including at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more rings. The rings can share an edge in some cases (e.g., be components of a bicyclic ring system). In general, the ring portion of the linker can provide a degree of physical rigidity to the linker and/or can serve to physically separate the dye (e.g., fluorescent dye) on one end of the linker from the substrate to be labeled and/or from a second dye (e.g., fluorescent dye) associated with the substrate and/or associated with the linker. A ring can be a component of an amino acid (e.g., a non-proteinogenic amino acid, as described herein).

**[00198]** In some cases, a linker may be "fully rigid" (e.g., substantially inflexible). For example, ring systems of the linker may not be separated by any $sp^2$ or $sp^3$ carbon atoms. In general, $sp^2$ and $sp^3$ carbon atoms (e.g., between ring systems) provide the linker with a degree of physical flexibility. $sp^3$ carbon atoms in particular can confer significant flexibility. Without limitation, flexibility can allow a polymerase to accept a substrate (e.g., a nucleotide or



Date Reçue/Date Received 2023-12-22

nucleotide analog) modified with the linker and the dye (e.g., fluorescent dye), or otherwise improve the performance of a labeled system. However, in a multiple dye system (e.g., a system comprising multiple fluorescent labeling reagents, such as a polynucleotide including two or more nucleotides coupled to two or more fluorescent labeling reagents), an overly flexible linker may defeat the feature of rigidity and allow two dyes (e.g., fluorescent dyes) to come into close association and be quenched. Accordingly, ring systems of a linker may be connected to each other by a limited number of $sp^3$ bonds, such as by no more than two $sp^3$ bonds (e.g., 0, 1, or 2 $sp^3$ bonds). For example, at least two ring systems of a linker may be connected to each other by no more than two $sp^3$ bonds (e.g., by 0, 1, or 2 $sp^3$ bonds). For example, at least two ring systems of a linker may be connected to each other by a no more than two $sp^2$ bonds, such as by no more than 1 $sp^2$ bond. Ring systems of a linker may be connected to each other by a limited number of atoms, such as by no more than 2 atoms. For example, at least two ring systems of a linker may be connected to each other by no more than 2 atoms, such as by only 1 atom or by no atoms (e.g., directly connected).

**[00199]** The series of ring systems of a linker may comprise aromatic and/or aliphatic rings. At least two ring systems of a linker may be connected to each other directly without an intervening carbon atom. A linker may comprise at least one amino acid that may comprise a ring system. For example, a linker may comprise at least one non-proteinogenic amino acid (e.g., as described herein), such as a hydroxyproline.

**[00200]** Many applications of optical (e.g., fluorescent) labeling reagents (e.g., nucleic acid sequencing reactions) can be performed in aqueous solutions. In some cases, a linker that has too high of a proportion of carbon and hydrogen atoms and/or a lack of charged chemical groups can be insufficiently water-soluble to be useful in an aqueous solution. Therefore, the linkers described herein can have a water-soluble group or groups.

**[00201]** A linker may include a water-soluble group at any useful position. For example, a linker may comprise a water-soluble group at or near a point of attachment to a label (e.g., dye, as described herein). Alternatively or in addition to, a linker may comprise a water-soluble group at or near a point of attachment to a substrate (e.g., a protein or a nucleotide or nucleotide analog). Alternatively or in addition to, a linker may comprise a water-soluble group between points of attachment to a label (e.g., dye, as described herein) and a substrate (e.g., a protein or a nucleotide or nucleotide analog). One or more rings of a linker may comprise a water-soluble group. For example, each of the rings may comprise a water-soluble group, two or more rings may comprise a water-soluble group, only one of the rings may comprise a water-soluble group, or anywhere there between. A given ring may comprise one or more water-soluble moieties. For



Date Reçue/Date Received 2023-12-22

example, a ring of a linker may comprise two water-soluble moieties. The water-soluble group(s) can be a constituent part of the backbone of a ring of a linker or can be appended to a ring of a linker (e.g., as a substituent). Each water-soluble moiety of a linker may be different. Alternatively, one or more water-soluble moieties of a linker may be the same. For example, each water-soluble moiety of a linker may be the same. In some cases, the water-soluble group is positively charged. Examples of suitable water-soluble groups include, but are not limited to, a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, and a boronic acid or boronic ester.

**[00202]** A water-soluble group can be any functional group that decreases (including making more negative) the logP of the optical (e.g., fluorescent) labeling reagent. LogP is the partition coefficient for a molecule between water and *n*-octanol. A greasy molecule is more likely to partition into octanol, giving a positive and large logP value. A formula for LogP can be represented as log $P_{octanol/water}$= log ($[solute]_{octanol}/[solute]_{water}$), where $[solute]_{octanol}$ is the concentration of the solute (*i.e.*, the labeling reagent) in octanol and $[solute]_{water}$ is the concentration of the solute in water. Therefore, the more a compound partitions into water compared to octanol, the more negative the logP. LogP can be measured experimentally or predicted using software algorithms. The water-soluble group can have any suitable LogP value. In some cases, the LogP is less than about 2, less than about 1.5, less than about 1, less than about 0.5, less than about 0, less than about -0.5, less than about -1, less than about -1.5, less than about -2, or lower. In some cases, the LogP is between about 2.0 and about -2.0.

**[00203]** A linker may include one or more asymmetric (e.g., chiral) centers (e.g., as described herein). All stereochemical isomers of linkers are contemplated, including racemates and enantiomerically pure linkers.

**[00204]** A linker, and/or a substrate (e.g., protein or nucleotide or nucleotide analog) or dye to which it may be attached, may include one or more isotopic (e.g., radio) labels (e.g., as described herein). All isotopic variations of linkers are contemplated.

**[00205]** The structural features of a linker, including the number of rings, the rigidity of the linker, and the like, can combine to establish a functional distance between a dye (e.g., fluorescent) dye and a substrate (e.g., protein or nucleotide or nucleotide analog) that are linked by the linker. In some cases, the distance corresponds to the length (and/or the functional length) of the linker. In some cases, the functional length varies based on the temperature, solvent, pH, and/or salt concentration of the solution in which the length is measured or estimated. The functional length can be measured in a solution in which an optical (e.g., fluorescent) signal from



Date Reçue/Date Received 2023-12-22

the substrate is measured. The functional length may an average or ensemble value of a distribution of functional lengths (e.g., over rotational, vibrational, and translational motions) and may differ based on, e.g., temperature, solvent, pH, and/or salt concentrations. The functional length may be estimated (e.g., based on bond lengths and steric considerations, such as by use of a chemical drawing or modeling program) and/or measured (e.g., using molecular imaging and/or crystallographic techniques).

**[00206]** A linker can establish any suitable functional length between a dye (e.g., fluorescent dye) and a substrate (e.g., protein or nucleotide or nucleotide analog). In some cases, the functional length is at most about 500 nanometers (nm), about 200 nm, about 100 nm, about 75 nm, about 50 nm, about 40 nm, about 30 nm, about 20 nm, about 10 nm, about 5 nm, about 2 nm, about 1.0 nm, about 0.5 nm, about 0.3 nm, about 0.2 nm, or less. In some instances, the functional length is at least about 0.2 nanometers (nm), at least about 0.3 nm, at least about 0.5 nm, at least about 1.0 nm, at least about 2 nm, at least about 5 nm, at least about 10 nm, at least about 20 nm, at least about 30 nm, at least about 40 nm, at least about 50 nm, at least about 75 nm, at least about 100 nm, at least about 200 nm, at least about 500 nm, or more. In some instances, the functional length is between about 0.5 nm and about 50 nm.

**[00207]** In some cases, the linker forms a straight and/or contiguous chain. In some instances, the linker is branched. The linker can be capable of forming a bond with a plurality of dyes (e.g., fluorescent dyes) and/or substrates (e.g., nucleotides and/or nucleotide analogs).

**[00208]** A linker may be a polymer having a regularly repeating unit. Alternatively, a linker may be a co-polymer without a regularly repeating unit. In some cases, the linker is not the result of a polymerization process. In general, a polymerization process can generate products having a variety of degrees of polymerization and molecular weights. In contrast, in some cases, the linkers described herein have a defined (*i.e.*, known) molecular weight.

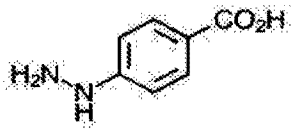
**[00209]** A linker may be constructed from one or more amino acids. For example, a linker may be constructed from two or more amino acids. An amino acid may be a natural amino acid or a non-natural amino acid. An amino acid may be a proteinogenic amino acid or a non-proteinogenic amino acid. A "proteinogenic amino acid," as used herein, generally refers to a genetically encoded amino acid that may be incorporated into a protein during translation. Proteinogenic amino acids include arginine, histidine, lysine, aspartic acid, glutamic acid, serine, threonine, asparagine, glutamine, cysteine, selenocysteine, glycine, proline, alanine, isoleucine, leucine, methionine, phenylalanine, tryptophan, tyrosine, valine, selenocysteine, and pyrrolysine. A "non-proteinogenic amino acid," as used herein, is an amino acid that is not a proteinogenic amino acid. A non-proteinogenic amino acid may be a naturally occurring amino acid or a non-



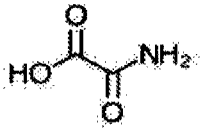
Date Reçue/Date Received 2023-12-22

naturally occurring amino acid. Non-proteinogenic amino acids include amino acids that are not found in proteins and/or are not naturally encoded or found in the genetic code of an organism. Examples of non-proteinogenic amino acids include, but are not limited to, hydroxyproline, selenomethionine, hypusine, 2-aminoisobutyric acid, αγ-aminobutyric acid, ornithine, citrulline, β-alanine (3-aminopropanoic acid), δ-aminolevulinic acid, 4-aminobenzoic acid, dehydroalanine, carboxyglutamic acid, pyroglutamic acid, norvaline, norleucine, alloisoleucine, t-leucine, pipecolic acid, allothreonine, homocysteine, homoserine, α-amino-n-heptanoic acid, α,β-diaminopropionic acid, α,γ-diaminobutyric acid, β-amino-n-butyric acid, β-aminoisobutyric acid, isovaline, sarcosine, N-ethyl glycine, N-propyl glycine, N-isopropyl glycine, N-methyl alanine, N-ethyl alanine, N-methyl β-alanine, N-ethyl β-alanine, isoserine, and α-hydroxy- γ-aminobutyric acid. Additional examples of non-proteinogenic amino acids include the non-natural amino acids described herein. A non-proteinogenic amino acid may comprise a ring structure. For example, a non-proteinogenic amino acid may be trans-4-aminomethylcyclohexane carboxylic acid or 4-hydrazinobenzoic acid. Such compounds may be FMOC-protected with FMOC (fluorenylmethyloxycarbohyl chloride) and utilized in solid-phase peptide synthesis. The structures of these compounds are shown below:




**[00210]** Where a linker comprises multiple amino acids, such as multiple non-proteinogenic amino acids, an amine moiety adjacent to a ring moiety (e.g., the amine moiety in the hydrazine moiety) can function as a water-solubilizing group. To synthesize a water-soluble peptide, a hybrid linker can be made that comprises alternating non-water-soluble amino acids and water-soluble amino acids (e.g., hydroxyproline). Other moieties can be used to increase water-solubility. For example, linking amino acids with oxamate moieties can provide water-solubility through the additional hydrogen bonding without adding any $sp^3$ linkages. The structure of the oxamate precursor 2-amino-2-oxoacetic acid is shown below:

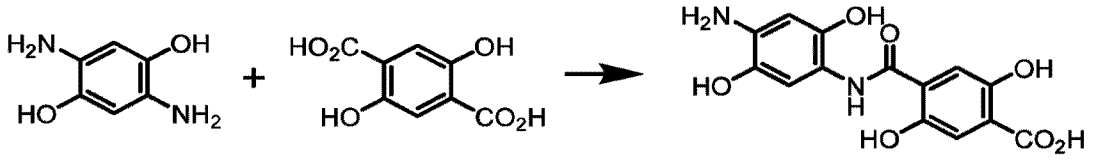



**[00211]** In some cases, a component (e.g., a monomer unit) of a linker may have an amino group, a carboxy group, and a water-solubilizing moiety. In some cases, a monomer may be deconstructed as two "half-monomers." That is, by using two different units, one that contains



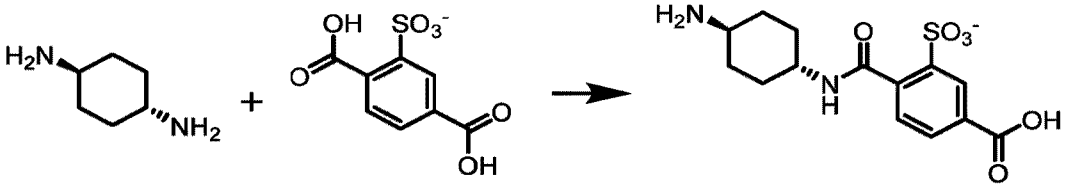
Date Reçue/Date Received 2023-12-22

two amino groups and another that contains two carboxy groups, an amino acid moiety can be constructed, which amino acid moiety may be a unit (e.g., a repeated unit) of a linker. One or both units may include one or more water solubilizing moieties. For example, at least one unit may include a water-soluble group (e.g., as described herein). For example, 2,5-diaminohydroquinone can be one half-monomer (A), and 2,5-dihydroxyterephthalic acid may be the other half-monomer (B). Such a scheme is shown below:




A B

**[00212]** As shown above, A is a diamine and B is a diacid. Accordingly, non-proteinogenic (e.g., non-natural) amino acids may be constructed from diamines and diacids. An additional example of such a construction is shown below:




Diamine Diacid Amino acid

**[00213]** A polymer based on two half-monomers (e.g., as shown above) can be constructed via solid phase synthesis. Because the half-monomers can be homobifunctional in the linking moiety, in some cases no FMOC protection is required. For example, the dicarboxylic acid can be appended to the solid support, then an excess of the diamine added with appropriate coupling reagent (HBTU / HOBT / collidine). After washing away excess reagent, an excess of the dicarboxylic acid can be added with the coupling reagent. Side-products consisting of one molecule of the fluid phase reagent reacting with two solid-phase attached reagent can result in truncation of the synthesis. These side products can be separated from a product after cleavage from the support and purification by HPLC.

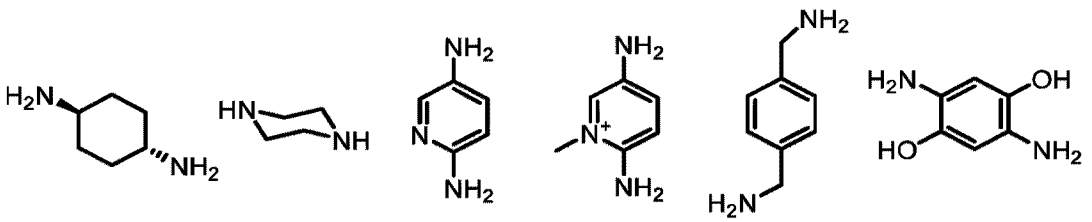
**[00214]** An advantage of the half-monomers approach can be increased flexibility in creating polymers. The diamine (A) can be replaced in a subsequent step by a different diamine (A') to change the properties of the polymer, in a repeating or non-repeating manner. Such a scheme may facilitate construction of a polymer such as ABA'BABA'B.

**[00215]** Additional examples of half-monomers for use according to the schemes described above include 2,5-diaminopyridine and 2,5-dicarboxypyridine, both of which are shown below, as well as the other moieties shown below:

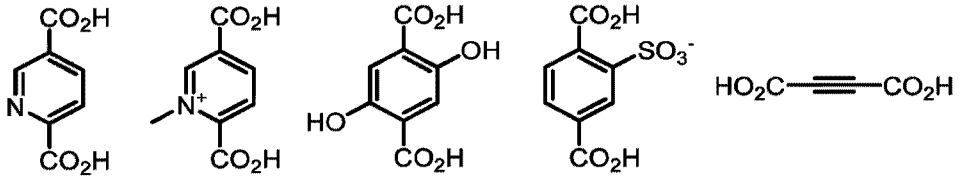


Date Reçue/Date Received 2023-12-22

Diamines:

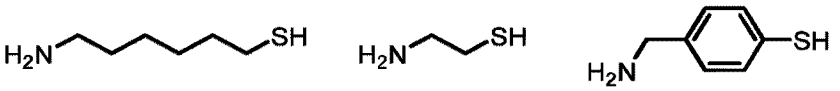


Dicarboxylic acids:

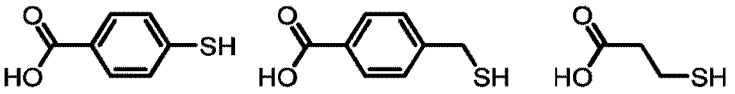


**[00216]** As described above, an amino acid (e.g., a non-proteinogenic amino acid that may be a non-natural amino acid) may be constructed from a diamine and a dicarboxylic acid. An amino acid (e.g., a non-proteinogenic amino acid that may be a non-natural amino acid) may also be constructed from an amino thiol and a thiol carboxylic acid. Examples of amino thiols and thiol carboxylic acids are shown below:

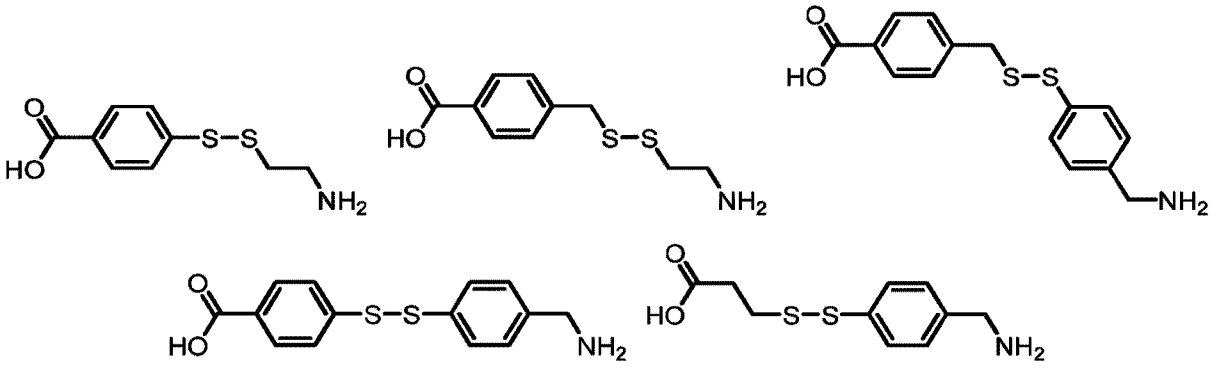
Amino thiols:



Thiol carboxylic acids:



**[00217]** Examples of amino acids (e.g., non-natural amino acids) constructed from an amino thiol and a thiol carboxylic acid are shown below:

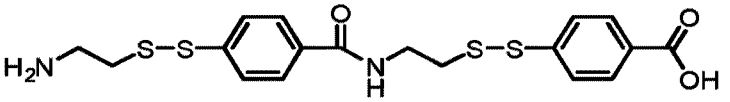


**[00218]** As shown above, amino acids constructed using an amino thiol and a thiol carboxylic acid may include a disulfide bond. As described elsewhere herein, a disulfide bond may be cleavable using a cleavage reagent (e.g., as described herein). Accordingly, an amino acid constructed from an amino thiol and a thiol carboxylic acid may serve as a cleavable portion of a linker. An amino acid constructed from an amino thiol and a carboxylic acid may be a component of a linker (e.g., as described herein) that may couple labeling moiety (e.g., a



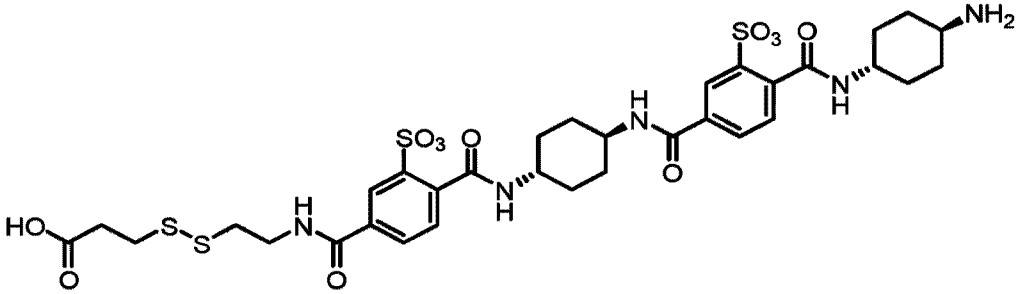
Date Reçue/Date Received 2023-12-22

fluorescent dye) to a substrate (e.g., a nucleotide or nucleotide analog). The various structures allow different hydrophobicities for incorporation and may provide different "scar" moieties subsequent to interaction with a cleavage reagent (e.g., as described herein). Two or more amino acids, such as two or more amino acids constructed from an amino thiol and a thiol carboxylic acid, may be included in a linker. For example, two or more amino acids may be included in a linker and separated by no more than 2 $sp^3$ carbon atoms, such as by no more than 2 $sp^2$ carbon atoms or by no more than 2 atoms. Where two or more amino acids formed of amino thiols and thiol carboxylic acids are connected to one another within a linker, cleavage may be more rapid as there will be multiple possible sites for cleavage. An example of a portion of a linker including such a component is shown below:

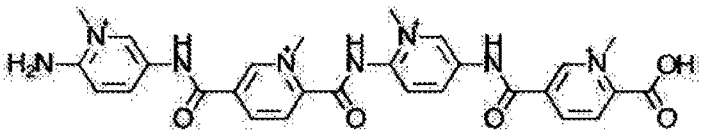


**[00219]** As described above, two half-monomers may combine to provide an amino acid (e.g., a non-proteinogenic amino acid, such as a non-natural amino acid). Accordingly, a non-natural amino acid may include any known non-natural amino acid, as well as any non-natural amino acid that may be constructed as described herein.

**[00220]** Half-monomers such as those described herein can be constructed into polypeptide polymers. An example of a nucleotide constructed with two repeating units of an amino acid is shown below:



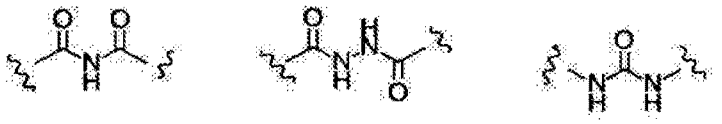
**[00221]** In some cases, before or after peptide coupling, the nitrogen in a nitrogen-containing ring can be quaternized to provide pyridinium moieties, thereby improving water-solubility of the final product. An example linker sequence generated in this manner is shown below:





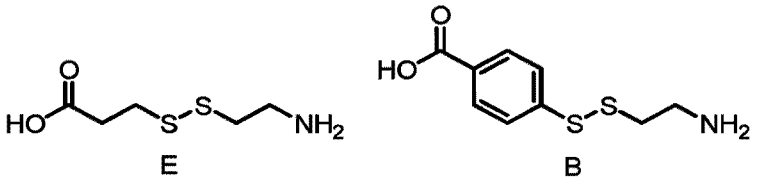
Date Reçue/Date Received 2023-12-22

**[00222]** Water-solubilizing linkages that can work with the half-monomer method include, for example, those that have symmetrical functional groups, such as secondary amides, bishydrazides, and ureas. Examples of such moieties are shown below:



**[00223]** Amino acid linker subunits may be assembled into polymers by peptide synthesis methods. For example, a solid support method known as SPPS (Solid Phase Peptide Synthesis) or by liquid-phase synthesis may be used to assemble amino acids into a linker. SPPS methods can use a solid phase bead where the initial step is attachment of the C-terminal amino acid via its carboxylic acid moiety, leaving its free amine ready for coupling. Peptide synthesis can be initiated by flowing FMOC amine-protected monomers with peptide coupling reagents such as HBTU and an organic base. Excess reagent can be washed away and the next monomer is introduced. After one or more amino acids have been appended the final peptide can be cleaved from the beads and purified by HPLC. Liquid phase synthesis can use the same reagents (except the beads) but purification occurs after each step. The advantage of either stepwise polymerization process is that the resultant linkers can have a defined molecular weight that may be confirmed by mass spectrometry.

**[00224]** A linker may include one or more components. For example, a linker may include a first component that includes a polymeric region (e.g., that includes a repeating unit) and a second unit that does not include a polymeric region. The second component may include a cleavable component (e.g., as described herein). Examples of cleavable linkers include, but are not limited to, the structures E and B shown below:

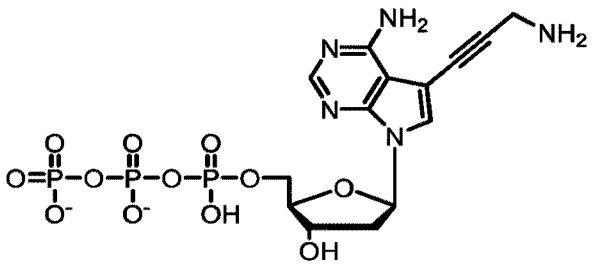



In the structures shown above, the disulfide moieties may be cleaved (e.g., as described herein) to provide thiol scars. The cleavable linkers may be attached to substrates upon reaction between a carboxyl moiety of the linker moiety and an amine moiety attached to a substrate (e.g., protein or nucleotide or nucleotide analog) to provide the substrate attached to the cleavable linker via an amide moiety. For example, the substrate may be a nucleotide or nucleotide analog including a propargylamino moiety, and a fluorescent labeling reagent comprising a dye and a

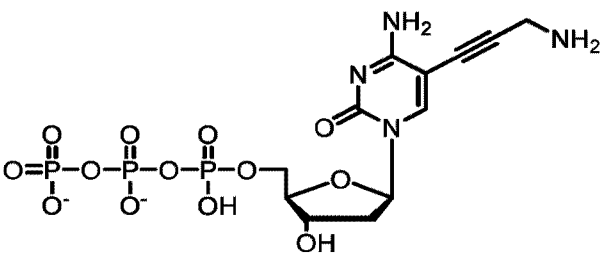


Date Reçue/Date Received 2023-12-22

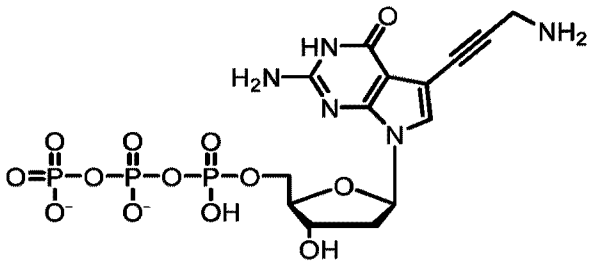
linker described herein may be configured to associate with the substrate via the propargylamino moiety. Examples of such substrates are shown below:



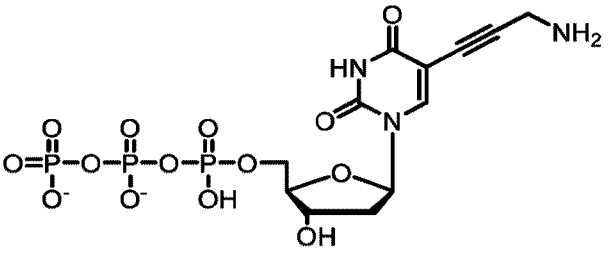
Modified dATP



Modified dCTP

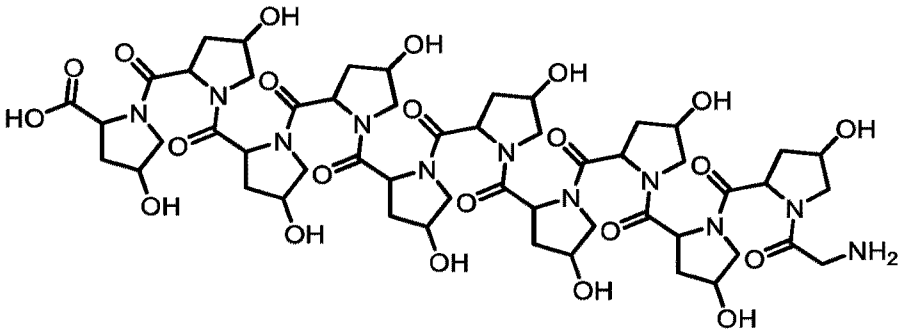


Modified dGTP



Modified dUTP

**[00225]** The first component of a linker including first and second components may include a repeating unit. For example, the linker may include a first component including one or more hydroxyproline moieties. An example of such a linker component is shown below:



hyp10

The linker shown above includes 10 hydroxyproline moieties and a glycine moiety and is referred to herein as "H" or "hyp10". An alternate version of the linker above includes 20 hydroxyproline moieties and a glycine moiety and is referred to herein as "hyp20". As described herein, all stereoisomers of hyp10 and hyp20, as well as combinations thereof, are contemplated. A linker component such as hyp10 can be linked to a cleavable linker via reaction between a free carboxyl moiety of the linker component and an amino moiety of a cleavable linker. A linker component such as hyp10 can be linked to a dye via the free amino moiety of the linker component. Examples of optical labeling reagent including a first linker component including a



Date Reçue/Date Received 2023-12-22

repeating unit (e.g., hyp10) and a second linker component including a cleavable linker are provided elsewhere herein.

**[00226]** Linkers may provide linkages between fluorescent moieties (e.g., dyes, as described herein) and substrates (e.g., proteins or nucleotides or nucleotide analogs). For example, an optical (e.g., fluorescent) labeling reagent may comprise an optical dye (e.g., fluorescent dye) attached to a linker (e.g., as described herein). Non-limiting examples of dyes (e.g., fluorescent dyes) include SYBR green, SYBR blue, DAPI, propidium iodine, Hoechst, SYBR gold, ethidium bromide, acridine, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, phenanthridines and acridines, propidium iodide, hexidium iodide, dihydroethidium, ethidium homodimer-1 and -2, ethidium monoazide, ACMA, Hoechst 33258, Hoechst 33342, Hoechst 34580, DAPI, acridine orange, 7-AAD, actinomycin D, LDS751, hydroxystilbamidine, SYTOX Blue, SYTOX Green, SYTOX Orange, POPO-1, POPO-3, YOYO-1, YOYO-3, TOTO-1, TOTO-3, JOJO-1, LOLO-1, BOBO-1, BOBO-3, PO-PRO-1, PO-PRO-3, BO-PRO-1, BO-PRO-3, TO-PRO-1, TO-PRO-3, TO-PRO-5, JO-PRO-1, LO-PRO-1, YO-PRO-1, YO-PRO-3, PicoGreen, OliGreen, RiboGreen, SYBR Gold, SYBR Green I, SYBR Green II, SYBR DX, SYTO dyes (e.g., SYTO-40, -41, -42, -43, -44, and -45 (blue); SYTO-13, -16, -24, -21, -23, -12, -11, -20, -22, -15, -14, and -25 (green); SYTO-81, -80, -82, -83, -84, and -85 (orange); SYTO-64, -17, -59, -61, -62, -60, and -63 (red)), fluorescein, fluorescein isothiocyanate (FITC), tetramethyl rhodamine isothiocyanate (TRITC), rhodamine, tetramethyl rhodamine, R-phycoerythrin, Cy-2, Cy-3, Cy-3.5, Cy-5, Cy5.5, Cy-7, Texas Red, Phar-Red, allophycocyanin (APC), Sybr Green I, Sybr Green II, Sybr Gold, CellTracker Green, 7-AAD, ethidium homodimer I, ethidium homodimer II, ethidium homodimer III, ethidium bromide, umbelliferone, eosin, green fluorescent protein, erythrosin, coumarin, methyl coumarin, pyrene, malachite green, stilbene, lucifer yellow, cascade blue, dichlorotriazinylamine fluorescein, dansyl chloride, fluorescent lanthanide complexes such as those including europium and terbium, carboxy tetrachloro fluorescein, 5 and/or 6-carboxy fluorescein (FAM), VIC, 5- (or 6-) iodoacetamidofluorescein, 5-{[2(and 3)-5-(Acetylmercapto)-succinyl]amino} fluorescein (SAMSA-fluorescein), lissamine rhodamine B sulfonyl chloride, 5 and/or 6 carboxy rhodamine (ROX), 7-amino-methyl-coumarin, 7-Amino-4-methylcoumarin-3-acetic acid (AMCA), BODIPY fluorophores, 8-methoxypyrene-1,3,6-trisulfonic acid trisodium salt, 3,6-Disulfonate-4-amino-naphthalimide, phycobiliproteins, AlexaFluor dyes (e.g., AlexaFluor 350, 405, 430, 488, 532, 546, 555, 568, 594, 610, 633, 635, 647, 660, 680, 700, 750, and 790 dyes), DyLight dyes (e.g., DyLight 350, 405, 488, 550, 594, 633, 650, 680, 755, and 800 dyes), Black Hole Quencher



Date Reçue/Date Received 2023-12-22

Dyes (Biosearch Technologies) (e.g.,BH1-0, BHQ-1, BHQ-3, and BHQ-10), QSY Dye fluorescent quenchers (from Molecular Probes/Invitrogen)(e.g., QSY7, QSY9, QSY21, and QSY35), Dabcyl, Dabsyl. Cy5Q, Cy7Q, Dark Cyanine dyes (GE Healthcare), Dy-Quenchers (Dyomics) (e.g., DYQ-660 and DYQ-661), ATTO fluorescent quenchers (ATTO-TEC GmbH) (e.g.,ATTO 540Q, 580Q, 612Q, 532, and 633), and other fluorophores and quenchers (e.g., as described herein). In some cases, the label may be a type that does not self-quench or exhibit proximity quenching. Non-limiting examples of a label type that does not self-quench or exhibit proximity quenching include Bimane derivatives such as Monobromobimane. Additional dyes included in structures provided herein may also be utilized in combination with any of the linkers provided herein, and with any substrate described herein, regardless of the context of their disclosure.

**[00227]** An optical (e.g., fluorescent) labeling reagent comprising an optical dye (e.g., fluorescent dye) and a linker can further comprise a cleavable group that is capable of being cleaved to separate the optical dye from a substrate with which the optical labeling reagent is associated. All or a portion of the linker may be part of the cleavable group. In some cases, cleaving a cleavable group may leave a scar group associated with substrate. The cleavable group can be, for example, an azidomethyl group capable of being cleaved by tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), or tetrahydropyranyl (THP) to leave a hydroxyl scar group. The cleavable group can be, for example, a disulfide bond capable of being cleaved by TCEP, DTT or THP to leave a thiol scar group. The cleavable group can be, for example, a hydrocarbyldithiomethyl group capable of being cleaved by TCEP, DTT or THP to leave a hydroxyl scar group. The cleavable group can be, for example, a 2-nitrobenzyloxy group capable of being cleaved by ultraviolet (UV) light to leave a hydroxyl scar group. A scar may also be, for example, an aromatic group such as a phenyl or benzyl moiety.

**[00228]** An optical (e.g., fluorescent) labeling reagent may be configured to associate with a substrate such as a nucleotide or nucleotide analog (e.g., as described herein). Alternatively or in addition to, an optical (e.g., fluorescent) labeling reagent may be configured to associate with a substrate such as a protein, cell, lipid, or antibody. For example, the optical labeling reagent may be configured to associate with a protein. A protein substrate may be any protein, and may include any useful modification, mutation, or label, including any isotopic label. For example, a protein may be an antibody such as a monoclonal antibody. A protein associated with one or more optical (e.g., fluorescent) labeling reagents (e.g., as described herein) may be, for example, an antibody (e.g., a monoclonal antibody) useful for labeling a cell, which labeled cell may be analyzed and sorted using flow cytometry.



Date Reçue/Date Received 2023-12-22

**[00229]** An optical (e.g., fluorescent) labeling reagent (e.g., as described herein) can decrease quenching (e.g., between dyes coupled to nucleotides or nucleotide analogs incorporated into a growing nucleic acid strand, such as during nucleic acid sequencing). For example, an optical (e.g., fluorescent) signal emitted by a substrate (e.g., a nucleotide or nucleotide analog that may be incorporated into a growing nucleic acid strand) can be proportional to the number of optical (e.g., fluorescent) labels associated with the substrate (e.g., to the number of optical labels incorporated adjacent or in proximity to the substrate). For example, multiple optical labeling reagents including substrates of the same or different types (e.g., nucleotides or nucleotide analogs of a same or different type) may be incorporated in proximity to one another in a growing nucleic acid strand (e.g., during nucleic acid sequencing). In such a system, signal emitted by the collective substrates may be approximately proportional (e.g., linearly proportional) to the number of dye-labeled substrates incorporated. In other words, quenching may not significantly impact the signal emitted. This may be observable in a system in which 100% labeling fractions are used. Where less than 100% of substrates are labeled (e.g., less than 100% of nucleotides in a nucleotide flow are labeled), an optical (e.g., fluorescent) signal emitted by substrates (e.g., nucleotides or nucleotide analogs) incorporated into a plurality of growing nucleic acid strands (e.g., a plurality of growing nucleic acid strands coupled to sequencing templates coupled to a support, as described herein) may be proportional to the length of a homopolymer region of the growing nucleic acid strands. Similarly, where less than 100% of substrates are labeled (e.g., less than 100% of nucleotides in each of successive nucleotide flows are labeled), an optical (e.g., fluorescent) signal emitted by substrates (e.g., nucleotides or nucleotide analogs) incorporated into a plurality of growing nucleic acid strands (e.g., a plurality of growing nucleic acid strands coupled to sequencing templates coupled to a support, as described herein) may be proportional to the length of a heteropolymeric and/or homopolymer region of the growing nucleic acid strands. In some such cases, the intensity of a measured optical (e.g., fluorescent) signal may be linearly proportional to the length of a heteropolymeric and/or homopolymeric region into which substrates have incorporated. For example, a measured optical (e.g., fluorescent) signal may be linearly proportional with a slope of approximately 1.0 when optical (e.g., fluorescent) signal is plotted against the length in substrates of a heteropolymeric and/or homopolymeric region into which substrates have incorporated.

**[00230]** An optical (e.g., fluorescent) labeling reagent (e.g., as described herein) can decrease quenching in a protein system. When labeling proteins, quenching may start to happen at a fluorophore to protein ratio (F/P) of around 3. Using optical labeling reagents provided



Date Reçue/Date Received 2023-12-22

herein, higher F/P ratios, and thus brighter reagents, may be obtained. This may be useful for analyzing proteins (e.g., using imaging) and/or for analyzing cells labeled with proteins (e.g., antibodies) associated with one or more optical (e.g., fluorescent) labeling reagents.

**[00231]** Examples of the linkers described herein are found, e.g., in **FIGs. 1A-1C, 2A, 4, 5A, 5B, 6, 7, 8, 13A-13C, 14A, 14B, 16,** and **17**. In some cases, the R group included in these linkers (e.g., as in **FIG. 1C**) confers sufficient water solubility on the labeling reagent. Additional examples are included elsewhere herein, including in the Examples below.

**[00232]** In as aspect, the present disclosure provides an oligonucleotide molecule comprising a fluorescent labeling reagent or derivative thereof (e.g., as described herein). The oligonucleotide molecule may comprise one or more additional fluorescent labeling reagents of a same type (e.g., comprising linkers having the same chemical structure, dyes comprising the same chemical structure, and/or associated with substrates (e.g., nucleotides) of a same type). The fluorescent labeling reagent and one or more additional fluorescent labeling reagents of the oligonucleotide molecule may be associated with nucleotides. For example, the fluorescent labeling reagents may be connected to nucleobases of nucleotides of the oligonucleotide molecule. A fluorescent labeling reagent and one or more additional fluorescent labeling reagent may be connected to adjacent nucleotides of the oligonucleotide molecule. Alternatively or in addition to, the fluorescent labeling reagent and the one or more additional fluorescent labeling reagents may be connected to nucleotides of the oligonucleotide molecule that are separated by one or more nucleotides that are not connected to fluorescent labeling reagents. The oligonucleotide molecule may be a single-stranded molecule. Alternatively, the oligonucleotide molecule may be a double-stranded or partially double-stranded molecule. A double-stranded or partially double-stranded molecule may comprise fluorescent labeling reagents associated with a single strand or both strands. The oligonucleotide molecule may be a deoxyribonucleic acid molecule. The oligonucleotide molecule may a ribonucleic acid molecule. The oligonucleotide molecule may be generated and/or modified via a nucleic acid sequencing process (e.g., as described herein).

**[00233]** The linker of the fluorescent labeling reagent may comprise a cleavable group that is configured to be cleaved to separate the fluorescent dye of the fluorescent labeling reagent from a substrate (e.g., nucleotide) with which it is associated. For example, the linker may comprise a cleavable group comprising an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, or a 2-nitrobenzyloxy group. The cleavable group may be configured to be cleaved by application of one or more members of the group consisting of tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), tetrahydropyranyl (THP),



Date Reçue/Date Received 2023-12-22

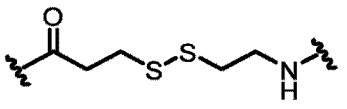
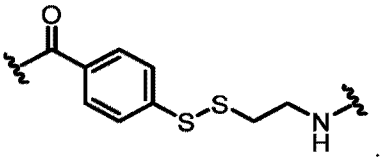
ultraviolet (UV) light, and a combination thereof. The oligonucleotide molecule comprising a fluorescent labeling reagent may be configured to emit a fluorescent signal (e.g., upon excitation at an appropriate range of energy, as described herein).

**[00234]** In another aspect, the present disclosure provides a kit comprising a plurality of linkers (e.g., as described herein). A linker of the plurality of linkers may comprise (i) one or more water soluble groups and (ii) two or more ring systems. At least two of the two or more ring systems may be connected to each other by no more than two $sp^3$ carbon atoms. For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. At least two of the two or more ring systems may be connected to each other by no more than two atoms. The linker may comprise a non-proteinogenic amino acid (e.g., as described herein) comprising a ring system of the two or more ring systems. For example, the linker may comprise a hydroxyproline or an amino acid constructed from, e.g., a diamine and a dicarboxylic acid or an amino thiol and a thiol carboxylic acid. The linker may be connected to a fluorescent dye (e.g., as described herein) and/or associated with a substrate. For example, the linker may be connected to a fluorescent dye and coupled to a substrate selected from a nucleotide, a protein, a lipid, a cell, and an antibody. For example, the linker may be connected to a fluorescent dye and a nucleotide.

**[00235]** The linker may comprise a plurality of amino acids, such as a plurality of non-proteinogenic (e.g., non-natural) amino acids. For example, the linker may comprise a plurality of hydroxyprolines (e.g., a hyp10 moiety). At least one water-soluble group of the one or more water-soluble groups may be appended to a ring structure of the two or more ring systems. The one or more water soluble groups may be selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester. The linker may comprise a cleavable group that is configured to be cleaved to separate a first portion of the linker from a second portion of the linker. The cleavable group may be selected from the group consisting of an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group. The cleavable group may be cleavable by application of one or more members of the group consisting of tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), tetrahydropyranyl (THP), ultraviolet (UV) light, and a combination thereof. The linker may comprise a moiety selected from the



Date Reçue/Date Received 2023-12-22

group consisting of  and . These moieties both comprise disulfide groups and so may be considered cleavable groups.

**[00236]** The plurality of linkers of the kit may comprise a first linker associated with a first substrate (e.g., a first nucleotide) and a second linker associated with a second substrate (e.g., a second nucleotide). The first substrate and the second substrate may be of different types (e.g., different canonical nucleotides). The first substrate and the second substrate may be nucleotides comprising nucleobases of different types (e.g., A, C, G, U, and T). The first linker and the second linker may comprise the same chemical structure. Similarly, the first linker may be connected to a first fluorescent dye and the second linker may be connected to a second fluorescent dye. The first fluorescent dye and the second fluorescent dye may be of different types. For example, the first and second fluorescent dyes may fluoresce at different wavelengths and/or have different maximum excitation wavelengths. The first and second fluorescent dyes may fluoresce at similar wavelengths and/or have similar maximum excitation wavelengths regardless of whether they share the same chemical structure.

**[00237]** The plurality of linkers of the kit may further comprise a third linker associated with a third substrate and a fourth linker associated with a fourth substrate. The first substrate, the second substrate, the third substrate, and the fourth substrate may be of different types. For example, the first substrate, the second substrate, the third substrate, and the fourth substrate may be nucleotides comprising nucleobases of different types (e.g., A, C, G, and U/T). The first linker and the third linker may comprise different chemical structures. The first and third linker may comprise a same chemical group, such as a same cleavable group (e.g., as described herein). For example, the first linker and the third linker may each comprise a moiety comprising a disulfide bond. Similarly, the first linker and the fourth linker may comprise different chemical structures. The first and fourth linker may comprise a same chemical group, such as a same cleavable group (e.g., as described herein). For example, the first linker and the fourth linker may each comprise a moiety comprising a disulfide bond.

**[00238]** In an example, the first linker comprises a hyp10 moiety and a first cleavable moiety, the second linker comprises a hyp10 moiety and a second cleavable moiety, the third linker comprises a third cleavable moiety and does not comprise a hyp10 moiety, and the fourth linker comprises a fourth cleavable moiety and does not comprise a hyp10 moiety. The second cleavable moiety may have a chemical structure that is different than the first cleavable moiety. Alternatively, the second cleavable moiety and the first cleavable moiety may have the same



Date Reçue/Date Received 2023-12-22

chemical structures. The third cleavable moiety and the fourth cleavable moiety may have the same chemical structure. Alternatively, the third cleavable moiety and the fourth cleavable moiety may have different chemical structures. In an example, the first linker and the second linker each have a first chemical structure and the third linker and the fourth linker each have a second chemical structure, which second structure is different than the first chemical structure. In another example, the first linker, the second linker, the third linker, and the fourth linker all have the same chemical structure. In another example, the first linker, the second linker, the third linker, and the fourth linker all have different chemical structures.

**Methods for using the optical labeling reagents**

**[00239]** There are several different types of quenching that can be reduced and different types of applications that can be performed using the optical (e.g., fluorescent) labeling reagents described herein.

**[00240]** The methods described herein can be used to reduce quenching, including G-quenching. Attachment of dyes (e.g., fluorescent dyes) to nucleotides (e.g., via a linker provided herein) can result in dye-quenching for many dyes, particularly when the dye is attached to a guanosine nucleotide. Dye quenching may take place between a dye and a nucleotide with which it is associated, as well as between dye moieties, such as between dye moieties coupled to different nucleotides (e.g., adjacent nucleotides or nucleotides separated by one or more other nucleotides). Use of the linkers provided herein can alleviate the quenching allowing more sensitive detection of sequences containing G. In addition, a dye-labeled nucleotide in proximity to a G-homopolymer region may show reduced fluorescence. Any nucleic acid sequencing method that requires attachment of a dye to dGTP may benefit from these linkers, including single molecule detection, sequencing using 3'-blocked nucleotides, and sequencing by hybridization.

**[00241]** The methods described herein can be used to reduce dye-dye quenching on adjacent or neighboring nucleotides (e.g., nucleotides separated by one, two, or more other nucleotides) on the same DNA strand. Methods that require dyes on adjacent or neighboring nucleotides can result in proximity quenching; that is, two dyes next to each other are less bright than twice the brightness of one dye, or often, less bright than even a single dye. Use of the linkers provided herein may alleviate the quenching, allowing quantitative detection of multiple dyes. For example, in sequencing methods such as mostly natural nucleotide flow sequencing, the fraction of labeled dye is typically less than 5%, since homopolymers are not linear in signal to homopolymer length at higher fractions due to the quenching problem. The reagents described



Date Reçue/Date Received 2023-12-22

herein can allow more (e.g., more than 5%, in some cases up to 100%) of the nucleotides to be labeled while facilitating sensitive and accurate detection of incorporated nucleotides.

**[00242]** The use of a dye-linker-nucleotide provided herein may result in more efficient incorporation into a growing nucleic acid strand (e.g., increased tolerance) by a polymerase (e.g., as described herein), compared to a dye-nucleotide lacking the linker (e.g., during nucleic acid sequencing). The result may be that a lower amount of the dye-labeled nucleotide is used to achieve the same signal.

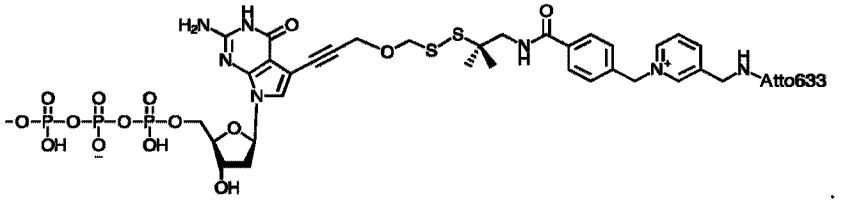
**[00243]** The use of a dye-linker-nucleotide provided herein may result in less misincorporation by a polymerase (e.g., as described herein) (e.g., during nucleic acid sequencing). The result may be less loss of template strands, and thus longer sequencing reads.

**[00244]** The use of a dye-linker-nucleotide provided herein may result in less mispair extension (e.g., during nucleic acid sequencing), and thus reduced lead phasing.

**[00245]** The methods described herein can be used to reduce dye-dye quenching in multi-dye applications. Hybridization assays can also benefit from linkers that prevent quenching. Quenching effects may result in non-linearity of target to signal.

**[00246]** The methods described herein can be used in combination with oligomers and dendrimers for signal amplification. Non-quenching linkers may allow the synthesis of very bright polymers for antibody labeling. These bright antibodies may be used for cell-surface labeling in flow cytometry or for antigen detection methods such as lateral flow tests and fluorescent immunoassays.

**[00247]** The optical (e.g., fluorescent) labeling reagent of the present disclosure may be used as a molecular ruler. The substrate can be a fluorescence quencher, a fluorescence donor, or a fluorescence acceptor. In some cases, the substrate is a nucleotide. The linker can be attached to the nucleotide on the nucleobase as shown below, where the dye is Atto633:




**[00248]** The structure shown above is an optical (e.g., fluorescent) labeling reagent comprising a cleavable (via the disulfide bond) moiety and a fluorescent dye attached via a pyridinium linker to a dGTP analog (dGTP-SS-py-Atto633). Additional examples of optical labeling reagents are provided throughout the disclosure.

**[00249]** The dye-labeled nucleotides described herein can be used in a sequencing by synthesis method using a mixture of dye-labeled and natural nucleotides in a flow-based scheme.



Date Reçue/Date Received 2023-12-22

Such methods often use a low percentage of labeled nucleotides compared to natural nucleotides. However, using a low percentage of labeled nucleotides compared to natural nucleotides in flow mixtures (*e.g.*, less than 20%) can have multiple drawbacks: (a) since a small fraction of the template provides sequence information, the method requires a high template copy number; (b) variability in DNA polymerase extension rates between labeled and unlabeled nucleotides can result in context-dependent labeling fractions, thus increasing the difficulty of distinguishing a single base incorporation from multiple base incorporations; and (c) the low fraction of labeling moieties can result in high binomial noise in the populations of labeled product. Methods for flow-based sequencing using mostly natural nucleotides are further described in U.S. Pat. No. 8,772,473, which is incorporated herein by reference in its entirety for all purposes.

**[00250]** The semi-rigid linkers provided herein may allow a labeled fraction of dye-labeled nucleotide to natural nucleotide in each flow to be sufficiently high (*e.g.*, 20-100% labeling) to avoid or reduce the effect of the aforementioned disadvantages of such schemes. This higher percentage labeling can result in greater optical (e.g., fluorescent) signal and thus a lower template requirement. If 100% labeling is used, the binomial noise and context variation may be essentially eliminated. The key technical barrier overcome by the solution described herein is that the dye-labeled nucleotides on adjacent or nearby nucleotides must show minimal quenching. The overall result of the combined advantages may be more accurate DNA sequencing.

**[00251]** The present disclosure provides a method for sequencing a nucleic acid molecule. The method can comprise contacting the nucleic acid molecule with a primer under conditions sufficient to hybridize the primer to the nucleic acid molecule, thereby generating a sequencing template. The sequencing template may then be contacted with a polymerase (e.g., as described herein) and a solution (e.g., a nucleotide flow) comprising a plurality of optically (e.g., fluorescently) labeled nucleotides (e.g., as described herein). Each optically (e.g., fluorescently) labeled nucleotide of the plurality of optically (e.g., fluorescently) labeled nucleotides may comprise the same chemical structure (e.g., each labeled nucleotide may comprise a dye of a same type, a linker of a same type, and a nucleotide or nucleotide analog of a same type). An optically labeled nucleotide of the plurality of optically labeled nucleotides may be complementary to the nucleic acid molecule at a plurality of positions adjacent to the primer hybridized to the nucleic acid molecule. Accordingly, one or more optically labeled nucleotides of the plurality of optically labeled nucleotides may be incorporated into the sequencing template. Where the nucleic acid molecule includes a homopolymeric region, multiple nucleotides (e.g., labeled and unlabeled nucleotides) may be incorporated. Incorporation of



Date Reçue/Date Received 2023-12-22

multiple nucleotides adjacent to one another may be facilitated by the use of non-terminated nucleotides. The solution comprising the plurality of optically labeled nucleotides may then be washed away from the sequencing template (e.g., using a wash flow, as described herein). An optical (e.g., fluorescent) signal from the sequencing template may be measured. Where two or more labeled nucleotides are incorporated into a homopolymeric region, the intensity of the measured optical (e.g., fluorescent) signal may be greater than an optical (e.g., fluorescent) signal that may be measured if a single optically (e.g., fluorescently) labeled nucleotide of the plurality of optically (e.g., fluorescently) labeled nucleotides had been incorporated into the sequencing template. Such a method may be particularly useful for sequencing of homopolymers or portions of nucleic acids that are homopolymeric (*i.e.,* have a plurality of the same base in a row). An optically labeled nucleotide of the plurality of optically labeled nucleotides may comprise a dye (e.g., fluorescent dye) and a linker connected to the dye and a nucleotide (e.g., as described herein). The linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms, such as by no more than two atoms. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. For example, the linker may comprise a hydroxyproline or an amino acid constructed from, e.g., a diamine and a dicarboxylic acid or an amino thiol and a thiol carboxylic acid. The linker may be configured to establish a functional length between the dye and the nucleotide of at least about 0.5 nanometers.

**[00252]** The intensity of the measured optical (e.g., fluorescent) signal may be proportional to the number of optically (e.g., fluorescently) labeled nucleotides incorporated into the sequencing template (e.g., where 100% labeling fraction is used). In other words, quenching may not significantly impact the signal emitted. For example, the intensity may be linearly proportional to the number of optically (e.g., fluorescently) labeled nucleotides incorporated into the sequencing template. The intensity of the measured optical (e.g., fluorescent) signal may be linearly proportional with a slope of approximately 1.0 when plotted against the number of optically (e.g., fluorescently) labeled nucleotides incorporated into the sequencing template. Where less than 100% of substrates are labeled (e.g., less than 100% of nucleotides in a nucleotide flow are labeled), an optical (e.g., fluorescent) signal emitted by substrates (e.g., nucleotides or nucleotide analogs) incorporated into a plurality of growing nucleic acid strands (e.g., a plurality of growing nucleic acid strands coupled to sequencing templates coupled to a support, as described herein) may be proportional to the length of a homopolymer region of the growing nucleic acid strands. Similarly, where less than 100% of substrates are labeled (e.g.,



Date Reçue/Date Received 2023-12-22

less than 100% of nucleotides in each of successive nucleotide flows are labeled), an optical (e.g., fluorescent) signal emitted by substrates (e.g., nucleotides or nucleotide analogs) incorporated into a plurality of growing nucleic acid strands (e.g., a plurality of growing nucleic acid strands coupled to sequencing templates coupled to a support, as described herein) may be proportional to the length of a heteropolymeric and/or homopolymer region of the growing nucleic acid strands. In some such cases, the intensity of a measured optical (e.g., fluorescent) signal may be linearly proportional to the length of a heteropolymeric and/or homopolymeric region into which substrates have incorporated. For example, a measured optical (e.g., fluorescent) signal may be linearly proportional with a slope of approximately 1.0 when optical (e.g., fluorescent) signal is plotted against the length in substrates of a heteropolymeric and/or homopolymeric region into which substrates have incorporated

**[00253]** The solution comprising the plurality of optically (e.g., fluorescently) labeled nucleotides may also contain un-labeled nucleotides (e.g., the labeling fraction may be less than 100%). For example, at least about 20% of nucleotides in the solution may be optically labeled, and at least about 80% of nucleotides in the solution may not be optically labeled. In some cases, the majority of the nucleotides in the solution may be optically labeled (e.g., between about 50-100%).

**[00254]** In some cases, two or more optically (e.g., fluorescently) labeled nucleotides of the plurality of optically (e.g., fluorescently) labeled nucleotides are incorporated into the sequencing template (e.g., into a homopolymeric region). In some cases, three or more optically (e.g., fluorescently) labeled nucleotides of the plurality of optically (e.g., fluorescently) labeled nucleotides are incorporated into the sequencing template. The number of optically labeled nucleotides incorporated into the sequencing template during a given nucleotide flow may depend on the homopolymeric nature of the nucleic acid molecule. In some cases, a first optically (e.g., fluorescently) labeled nucleotide of the plurality of optically (e.g., fluorescently) labeled nucleotides is incorporated within four positions of a second optically (e.g., fluorescently) labeled nucleotide of the plurality of optically (e.g., fluorescently) labeled nucleotides.

**[00255]** An optically (e.g., fluorescently) labeled nucleotide may comprise a cleavable group to facilitate cleavage of the optical (e.g., fluorescent) label (e.g., as described herein). In some cases, a method may further comprise, subsequent to incorporation of the one or more optically (e.g., fluorescently) labeled nucleotides and washing away of residual solution, cleaving optical (e.g., fluorescent) labels of the one or more optically (e.g., fluorescently) labeled



Date Reçue/Date Received 2023-12-22

nucleotides incorporated into the sequencing template (e.g., as described herein). The cleavage flow may be followed by an additional wash flow.

**[00256]** In some cases, a nucleotide flow and wash flow may be followed by a "chase" flow comprising unlabeled nucleotides and no labeled nucleotides. The chase flow may be used to complete the sequencing reaction for a given nucleotide position or positions of the sequencing template (e.g., across a plurality of such templates immobilized to a support). The chase flow may precede detection of an optical signal from a template. Alternatively, the chase flow may follow detection of an optical signal from a template. The chase flow may precede a cleavage flow. Alternatively, the chase flow may follow a cleavage flow. The chase flow may be followed by a wash flow.

**[00257]** The methods provided herein can also be used to sequence heteropolymers and/or heteropolymeric regions of a nucleic acid molecule (i.e., portions that are not homopolymeric). Accordingly, the methods described herein can be used to sequence a nucleic acid molecule having any degree of heteropolymeric or homopolymeric nature.

**[00258]** Regarding homopolymers, a nucleotide flow at a homopolymer region may incorporate several nucleotides in a row. Contacting a sequencing template comprising a nucleic acid molecule (e.g., a nucleic acid molecule hybridized to an unextended primer) comprising a homopolymer region with a solution comprising a plurality of nucleotides (e.g., labeled and unlabeled nucleotides), where each nucleotide of the plurality of nucleotides is of a same type, may result in multiple nucleotides of the plurality of nucleotides being incorporated into the sequencing template. In some cases, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 nucleotides are incorporated (*i.e.*, in a homopolymeric region of a nucleic acid molecule). The plurality of nucleotides incorporated into the sequencing template may comprise a plurality of labeled nucleotides (e.g., optically labeled, such as fluorescently labeled), as described herein. In such an instance, one or more of said nucleotides incorporated into a homopolymer region may be labeled, and may either occupy adjacent or non-adjacent positions to other labeled nucleotides incorporated into the homopolymeric region. The intensity of a signal obtained from a nucleic acid molecule may be proportional to the number of incorporated labeled nucleotides (e.g., where a labeling fraction of 100% is used). For example, the intensity of an optical signal (e.g., fluorescent signal) obtained from a nucleic acid molecule containing two labeled nucleotides may be of greater intensity than the optical signal obtained from a nucleic acid molecule containing one labeled nucleotide. Furthermore, the intensity of a signal obtained from a nucleic acid molecule may depend on the relative positioning of labeled nucleotides within a nucleic acid molecule. For example, a nucleic acid molecule containing two



Date Reçue/Date Received 2023-12-22

labeled nucleotides in non-adjacent positions may provide a different signal intensity than a nucleic acid molecule containing two labeled nucleotides in adjacent positions. Quenching in such systems may be optimized by careful selection of linkers and dyes (e.g., fluorescent dyes). In some cases, a plot of optical signal (e.g., fluorescence) vs. homopolymer length can be linear. For example, measured optical signal for an ensemble of growing nucleic acid strands including homopolymeric regions into which labeled nucleotides are incorporated may be approximately linearly proportional to the nucleotide length of the homopolymeric region.

**[00259]** In another aspect, the present disclosure provides a method for sequencing a nucleic acid molecule. The method can comprise contacting the nucleic acid molecule with a primer under conditions sufficient to hybridize the primer to the nucleic acid molecule, thereby generating a sequencing template. The may then be contacted with a polymerase and a first solution comprising a plurality of first optically (e.g., fluorescently) labeled nucleotides (and, optionally, a plurality of first unlabeled nucleotides). Each first optically (e.g., fluorescently) labeled nucleotide of the plurality of first optically (e.g., fluorescently) labeled nucleotides is of a same type. A first optically (e.g., fluorescently) labeled nucleotide of the plurality of first optically (e.g., fluorescently) labeled nucleotides may be complementary to the nucleic acid molecule to be sequenced at a position adjacent to the primer. A first optically (e.g., fluorescently) labeled nucleotide of the plurality of first optically (e.g., fluorescently) labeled nucleotides may thus be incorporated into the sequencing template to generate an extended primer. The first solution comprising the plurality of first optically (e.g., fluorescently) labeled nucleotides may then be washed away from the sequencing template (e.g., using a wash solution). A first optical (e.g., fluorescent) signal emitted by the sequencing template may then be measured (e.g., as described herein). The sequencing template may then be contacted with a polymerase and a second solution comprising a plurality of second optically (e.g., fluorescently) labeled nucleotides (and, optionally, a plurality of second unlabeled nucleotides). Each second optically (e.g., fluorescently) labeled nucleotide of the plurality of second optically (e.g., fluorescently) labeled nucleotides may be of a same type. A second optically (e.g., fluorescently) labeled nucleotide of the plurality of second optically (e.g., fluorescently) labeled nucleotides may be complementary to the nucleic acid molecule to be sequenced at a position adjacent to the extended primer. A second optically (e.g., fluorescently) labeled nucleotide of the plurality of second optically (e.g., fluorescently) labeled nucleotides may thus be incorporated into the sequencing template. The second solution comprising the plurality of second optically (e.g., fluorescently) labeled nucleotides may then be washed away from the sequencing template. A second optical (e.g., fluorescent) signal emitted by the sequencing



Date Reçue/Date Received 2023-12-22

template may then be measured. In some cases, the intensity of the second optical (e.g., fluorescent) signal may be greater than the intensity of the first optical (e.g., fluorescent) signal.

**[00260]** A first optically labeled nucleotide of the plurality of first optically labeled nucleotides may comprise a first dye (e.g., fluorescent dye) and a first linker connected to the first dye and a first nucleotide (e.g., as described herein). Similarly, a second optically labeled nucleotide of the plurality of second optically labeled nucleotides may comprise a second dye (e.g., fluorescent dye) and a second linker connected to the second dye and a second nucleotide (e.g., as described herein). The first linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms, such as by no more than two atoms. For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. For example, the first linker may comprise one or more hydroxyproline moieties (e.g., as described herein). The first linker may be configured to establish a functional length between the first dye and the first nucleotide of at least about 0.5 nanometers. Similarly, the second linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms, such as by no more than two atoms. For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. For example, the second linker may comprise one or more hydroxyproline moieties (e.g., as described herein). The second linker may be configured to establish a functional length between the second dye and the second nucleotide of at least about 0.5 nanometers. The first linker and the second linker may have the same structure. Alternatively, the first linker and the second linker may have different structures. The first linker and the second linker may comprise a shared structural motif, such as a shared cleavable component (e.g., as described herein). The first linker and/or the second linker may comprise a cleavable group configured to be cleaved with a cleavage reagent (e.g., as described herein).

**[00261]** The first solution comprising the plurality of first optically (e.g., fluorescently) labeled nucleotides may also contain first un-labeled nucleotides. For example, about 20% of the nucleotides of the first solution may be un-labeled. In some cases, at least 20% of the nucleotides of the first solution may be optically labeled, such as at least 50% or at least 80%. The un-labeled nucleotides may comprise the same nucleotide moiety (e.g., canonical nucleotide



Date Reçue/Date Received 2023-12-22

moiety) as the optically labeled nucleotides. Similarly, the second solution comprising the plurality of first optically labeled nucleotides may also contain second un-labeled nucleotides. For example, about 20% of the nucleotides of the second solution may be un-labeled. In some cases, at least 20% of the nucleotides of the second solution may be optically labeled, such as at least 50% or at least 80%. The un-labeled nucleotides may comprise the same nucleotide moiety (e.g., canonical nucleotide moiety) as the optically labeled nucleotides.

**[00262]** The plurality of first optically (e.g., fluorescently) labeled nucleotides may be different than the plurality of second optically (e.g., fluorescently) labeled nucleotides. For example, the plurality of first optically (e.g., fluorescently) labeled and the plurality of second optically (e.g., fluorescently) labeled nucleotides may comprise the same optical (e.g., fluorescent) label (e.g., the same dye) and different nucleotides. Alternatively, the plurality of first optically (e.g., fluorescently) labeled and the plurality of second optically (e.g., fluorescently) labeled nucleotides may comprise different optical (e.g., fluorescent) labels (e.g., different dyes) and the same nucleotides. In some cases, the plurality of first optically (e.g., fluorescently) labeled and the plurality of second optically (e.g., fluorescently) labeled nucleotides may comprise different optical (e.g., fluorescent) labels (e.g., different dyes) and different nucleotides. The first dye of the first plurality of optically labeled nucleotides and the second dye of the second plurality of optically labeled nucleotides may emit signal at approximately the same wavelength or range of wavelengths (e.g., whether the first and second dyes have the same or different chemical structures). For example, the first dye and the second dye may both emit signal in the green region of the visible portion of the electromagnetic spectrum.

**[00263]** In some cases, two or more first optically (e.g., fluorescently) labeled nucleotides may be incorporated into the sequencing template (e.g., in a homopolymeric region of the nucleic acid molecule). In some cases, two or more second optically (e.g., fluorescently) labeled nucleotides may be incorporated into the sequencing template.

**[00264]** Additional optically (e.g., fluorescently) labeled nucleotides may also be provided and incorporated into the sequencing template (e.g., in successive nucleotide flows, as described herein). For example, the method may further comprise contacting the sequencing template with a polymerase and a third solution comprising a plurality of third optically (e.g., fluorescently) labeled nucleotides, wherein each third optically (e.g., fluorescently) labeled nucleotide of the plurality of third optically (e.g., fluorescently) labeled nucleotides is of a same type, and wherein a third optically (e.g., fluorescently) labeled nucleotide of the plurality of third optically (e.g., fluorescently) labeled nucleotides is complementary to the nucleic acid molecule at a position



Date Reçue/Date Received 2023-12-22

adjacent to the further extended primer hybridized to the nucleic acid molecule, thereby incorporating a third optically (e.g., fluorescently) labeled nucleotide of the plurality of third optically (e.g., fluorescently) labeled nucleotides into the sequencing template; washing the third solution comprising the plurality of third optically (e.g., fluorescently) labeled nucleotides away from the sequencing template; and measuring a third optical (e.g., fluorescent) signal emitted by the sequencing template. In some cases, the intensity of the third optical signal may be greater than the intensity of the first optical (e.g., fluorescent) signal and the intensity of the second optical (e.g., fluorescent) signal. This process may be repeated with a fourth solution, etc. The third and fourth solutions may comprise optically (e.g., fluorescently) labeled nucleotides having different nucleotides than the first and second solutions, such that each canonical nucleotide (A, C, G, and U/T) may be provided in sequence to the sequencing template. A cycle in which each canonical nucleotide is provided to the sequencing template may be repeated one or more times to sequence and/or amplify the nucleic acid molecule.

**[00265]** A third optically labeled nucleotide of the plurality of third optically labeled nucleotides may comprise a third dye (e.g., fluorescent dye) and a third linker connected to the third dye and a third nucleotide (e.g., as described herein). The third linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms, such as by no more than two atoms. For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. For example, the third linker may comprise one or more hydroxyproline moieties (e.g., as described herein). The third linker may be configured to establish a functional length between the third dye and the third nucleotide of at least about 0.5 nanometers. The third linker and the first linker may have the same or different structures. Similarly, the third linker and the second linker may have the same or different structures. The third dye may have the same or a different structure as the first dye. Similarly, the third dye may have the same or a different structure as the second dye. The third dye and the first and/or second dye may emit at approximately the same wavelength or range of wavelengths (e.g., whether these dyes have the same or different chemical structures). Further, the third nucleotide may be of a same or different type as the first nucleotide, or the third nucleotide may be of a same or different type as the second nucleotide.

**[00266]** The method may further comprise, subsequent to washing a given solution (e.g., nucleotide flow) away (e.g., using a wash solution), cleaving the optical (e.g., fluorescent) label of its respective nucleotides. For example, after the first solution is washed away, the optical



Date Reçue/Date Received 2023-12-22

(e.g., fluorescent) label of the first optically (e.g., fluorescently) labeled nucleotide incorporated into the sequencing template may be cleaved (e.g., using a cleavage reagent to cleave a cleavable group of a linker of the first optically labeled nucleotide, as described herein). For example, the fluorescent dye(s) of the first optically labeled nucleotide(s) incorporated into the sequencing template may be cleaved prior to contacting the sequencing template with second optically labeled nucleotides (e.g., in a second nucleotide flow, as described herein). Accordingly, signal may be detected from one or more first optically labeled nucleotides prior to incorporation of one or more second optically labeled nucleotides into the sequencing template. Separation of the fluorescent dye (s) of the first optically labeled nucleotide(s) incorporated into the sequencing template may provide a scarred nucleotide(s) comprising a portion of the linker of the first optically labeled nucleotide, or a derivative thereof. Similarly, after the second solution (e.g., second nucleotide flow) is washed away, the optical (e.g., fluorescent) label of the second optically (e.g., fluorescently) labeled nucleotide incorporated into the sequencing template may be cleaved. All of a portion of the first and second linkers may be cleaved during the respective cleaving processes.

**[00267]** In another aspect, provided herein is a method for sequencing a nucleic acid molecule. The method can comprise providing a solution comprising a plurality of optically (e.g., fluorescently) labeled nucleotides, wherein each optically (e.g., fluorescently) labeled nucleotide of the plurality of optically (e.g., fluorescently) labeled nucleotides is of a same type. A given optically (e.g., fluorescently) labeled nucleotide of the plurality of fluorescently labeled nucleotides may comprise an optical (e.g., fluorescent) dye that is connected to a nucleotide via a semi-rigid water-soluble linker having a defined molecular weight. The linker connecting the dye and nucleotide may provide a functional length of at least about 0.5 nanometers (nm) between the dye and nucleotide. The nucleic acid molecule may then be contacted with a primer under conditions sufficient to hybridize the primer to a nucleic acid molecule to be sequenced to generate a sequencing template. The sequencing template may then be contacted with a polymerase and the solution containing the plurality of optically (e.g., fluorescently) labeled nucleotides, wherein an optically (e.g., fluorescently) labeled nucleotide of the plurality of optically (e.g., fluorescently) labeled nucleotides is complementary to the nucleic acid molecule to be sequenced at a position adjacent to the primer. One or more optically (e.g., fluorescently) labeled nucleotides of the plurality of optically (e.g., fluorescently) labeled nucleotides may thus be incorporated into the sequencing template. The solution comprising the plurality of optically (e.g., fluorescently) labeled nucleotides may be washed away from the sequencing template (e.g.,



Date Reçue/Date Received 2023-12-22

using a wash solution). An optical (e.g., fluorescent) signal emitted by the sequencing template may then be measured.

**[00268]** The linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems, wherein at least two of the two or more ring systems are connected to each other by no more than two $sp^3$ carbon atoms, such as by no more than two atoms (e.g., as described herein). For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. For example, the linker may comprise one or more hydroxyproline moieties (e.g., as described herein). The linker may establish a functional length between the fluorescent dye and the nucleotide of at least about 0.5 nanometers (e.g., as described herein).

**[00269]** The measured optical (e.g., fluorescent) signal may be proportional to the number of optically (e.g., fluorescently) labeled nucleotides that were incorporated into the sequencing template. For example, where 100% labeling fraction is used (e.g., all nucleotides in the solution are labeled), quenching may not diminish the emitted signal. In such a system, the measured optical (e.g., fluorescent) signal can be linearly proportional to the number of optically (e.g., fluorescently) labeled nucleotides that were incorporated into the sequencing template. The measured optical (e.g., fluorescent) signal may be linearly proportional with a slope of approximately 1.0 when plotted against the number of optically (e.g., fluorescently) labeled nucleotides that were incorporated into the sequencing template. Where less than 100% of nucleotides are labeled (e.g., less than 100% of nucleotides in the solution are labeled), an optical (e.g., fluorescent) signal emitted by nucleotides incorporated into a plurality of growing nucleic acid strands (e.g., a plurality of growing nucleic acid strands coupled to sequencing templates coupled to a support, as described herein) may be proportional to the length of a homopolymer region of the growing nucleic acid strands. Similarly, where less than 100% of nucleotides are labeled, an optical (e.g., fluorescent) signal emitted by nucleotides incorporated into a plurality of growing nucleic acid strands (e.g., a plurality of growing nucleic acid strands coupled to sequencing templates coupled to a support, as described herein) may be proportional to the length of a heteropolymeric and/or homopolymer region of the growing nucleic acid strands. In some such cases, the intensity of a measured optical (e.g., fluorescent) signal may be linearly proportional to the length of a heteropolymeric and/or homopolymeric region into which nucleotides have incorporated. For example, a measured optical (e.g., fluorescent) signal may be linearly proportional with a slope of approximately 1.0 when optical (e.g., fluorescent) signal is



Date Reçue/Date Received 2023-12-22

plotted against the length in nucleotides of a heteropolymeric and/or homopolymeric region into which nucleotides have incorporated

**[00270]** In some cases, the solution containing an optically (e.g., fluorescently) labeled nucleotide also contains un-labeled nucleotides. The un-labeled nucleotides may comprise the same nucleotide moiety (e.g., the same canonical nucleotide). In some embodiments, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, or about 100% of nucleotides in the solution are fluorescently labeled. In some cases, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or more of nucleotides in the solution are fluorescently labeled. In some cases, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or more of nucleotides in the solution are not fluorescently labeled.

**[00271]** A plurality of labeled nucleotides can be incorporated at locations along a nucleic acid molecule in proximity to each other. In some cases, a first optically (e.g., fluorescently) labeled nucleotide is incorporated within 4 positions, within 3 positions, within 2 positions, or next to a second optically (e.g., fluorescently) labeled nucleotide (e.g., a second optically labeled nucleotide of a same or different nucleotide type). In some cases, the method further comprises cleaving the optical (e.g., fluorescent) labels from the nucleotides after measuring the optical (e.g., fluorescent) signal (e.g., as described herein). Cleaving an optical (e.g., fluorescent) label may leave behind a scar (e.g., as described herein). A nucleic acid sequencing assay may be used to evaluate dye-labeled nucleotides. The assay may use a nucleic acid template having a known sequence, which sequence may include one or more homopolymeric regions. The template may be immobilized to a support (e.g., as described herein) via an adapter. A primer having a sequence at least partly complementary to the adapter or a portion thereof may hybridize to the adapter or portion thereof and provide a starting point for generation of a nucleic acid strand having a sequence complementary to that of the template via incorporation of labeled and unlabeled nucleotides (e.g., as described herein). The sequencing assay may use four distinct four nucleotide flows including different canonical nucleobases that may be repeated in cyclical fashion (e.g., cycle 1: A, G, C, U; cycle 2 A, G, C, U; etc.). Each nucleotide flow may include nucleotides including nucleobases of a single canonical type (or analogs thereof), some of which may be include optical labeling reagents provided herein. The labeling fraction (e.g., % of nucleotides included in the flow that are attached to an optical labeling reagent) may be varied between, e.g., 0.5% to 100%. Labeling fractions may be different for different nucleotide flows.



Date Reçue/Date Received 2023-12-22

Nucleotides may not be terminated to facilitate incorporation into homopolymeric regions. The template may be contacted with a nucleotide flow, followed by one or more wash flows (e.g., as described herein). The template may also be contacted with a cleavage flow (e.g., as described herein) including a cleavage reagent configured to cleave a portion of the optical labeling reagents attached to labeled nucleotides incorporated into the growing nucleic acid strand. A wash flow may be used to remove cleavage reagent and prepare the template for contact with a subsequent nucleotide flow. Emission may be detected from labeled nucleotides incorporated into the growing nucleic acid strand after each nucleotide flow.

**[00272]** An example sequencing procedure **1800** is provided in **FIG. 18**. In process **1802**, a template and primer configured for nucleotide incorporation are provided. A first sequencing cycle **1804** is subsequently performed. First sequencing cycle **1804** includes four flow processes **1804a**, **1804b**, **1804c**, and **1804d**, each of which multiple flows. Nucleotides 1, 2, 3, and 4 may each include nucleobases of different canonical types (e.g., A, G, C, and U). A given nucleotide flow may include both labeled nucleotides (e.g., nucleotides labeled with an optical labeling reagent provided herein) and unlabeled nucleotides. The labeling fraction of each nucleotide flow may be different. That is, A, B, C, and D in **FIG. 18** may be the same or different and may range from 0% to 100% (e.g., as described herein). Labels and linkers used to label nucleotides 1, 2, 3, and 4 may be of the same or different types. For example, nucleotide 1 may have a linker including a cleavable linker and a hyp10 linker and a first green dye, and nucleotide 2 may have a linker including a cleavable linker but not a hyp10 linker and a second green dye. The first green dye may be the same as or different than the first green dye. The cleavable linkers associated with the different nucleotides may be the same or different. Flow process **1804a** may include a nucleotide flow (e.g., a flow including a plurality of nucleotides of type Nucleotide 1, A% of which may be labeled). During this flow, labeled and unlabeled nucleotides may be incorporated into the growing strand (e.g., using a polymerase enzyme). A first wash flow ("wash flow 1") may be used to remove unincorporated nucleotides and associated reagents. A cleavage flow including a cleavage reagent may be provided to all or portions of the optical labeling reagents attached to incorporated nucleotides. For example, labeled nucleotides may include a cleavable linker portion that may by cleaved upon contact with the cleavage reagent to provide a scarred nucleotide. A second wash flow ("wash flow 2") may be used to remove the cleavage reagent and cleaved materials. Nucleotide flow process **1804a** may also include a "chase" process in which a nucleotide flow including only unlabeled nucleotides of type Nucleotide 1 may be flowed. Such a chase process may be followed by a wash flow. The chase process and its accompanying wash flow may take place after the initial nucleotide flow and



Date Reçue/Date Received 2023-12-22

wash flow 1, or after the cleavage flow and wash flow 2. The next nucleotide flow process **1804b** may then begin and proceed in similar fashion. Following completion of processes **1804b**, **1804c**, and **1804d**, the first flow cycle **1804** may be complete. A second flow cycle **1806** may begin. Cycle **1806** may include the same flow processes in the same or different order. Additional cycles may be performed until all or a portion of the template has been sequenced. Detection of incorporated nucleotides via emission detection may be performed after nucleotide flows and initial wash flows and before cleavage flows for each nucleotide flow process (e.g., flow process **1804a** may include a detection process between wash flow 1 and cleavage flow, etc.). A template interrogated by such a sequencing process may be immobilized to a support (e.g., as described herein). A plurality of such templates (e.g., at least about 100, 200, 500, 1000, 10000, 100,000, 500,000, 1,000,000, or more templates) may be interrogated contemporaneously in this fashion (e.g., in clonal fashion). In such a system, incorporation of nucleotides may be detected as an average over the plurality of templates, which may permit the use of labeling fractions of less than 100%.

**[00273]** In some cases, for any of the preceding methods, the nucleotide is guanine (G) and the linker decreases quenching between the nucleotide and the dye (e.g., fluorescent) dye.

**[00274]** In some cases, for any of the preceding methods, an optically (e.g., fluorescently) labeled nucleotide comprising a linker provided herein is more efficiently incorporated into a sequencing template than another optically (e.g., fluorescently) labeled nucleotide that comprises the same nucleotide and optical (e.g., fluorescent) dye but does not include the linker. In some cases, for any of the preceding methods, an optically (e.g., fluorescently) labeled nucleotide comprising a linker provided herein is incorporated into a sequencing template with higher fidelity than another optically (e.g., fluorescently) labeled nucleotide that comprises the same nucleotide and optical (e.g., fluorescent) dye but does not include the linker.

**[00275]** For any of the sequencing methods provided herein, the polymerase used may be a Family A polymerase such as Taq, Klenow, or Bst polymerase. Alternatively, for any of the sequencing methods provided herein, the polymerase may be a Family B polymerase such as Vent(exo-) or Therminator™ polymerase.

**[00276]** In an aspect, the present disclosure provides methods for sequencing a nucleic acid molecule using the optically (e.g., fluorescently) labeled nucleotides described herein. A method may comprise providing a plurality of nucleic acid molecules, which plurality of nucleic acid molecules may comprise or be part of a colony or a plurality of colonies. The plurality of nucleic acid molecules may have sequence homology to a template sequence. The method may comprise contacting the plurality of nucleic acid molecules with a solution comprising a plurality



Date Reçue/Date Received 2023-12-22

of nucleotides (e.g., a solution comprising a plurality of optically labeled nucleotides) under conditions sufficient to incorporate a subset of the plurality of nucleotides into a plurality of growing nucleic acid strands that is complementary to the plurality of nucleic acid molecules. In some instances, at least about 20% of the subset of the plurality of nucleotides are optically (e.g., fluorescently) labeled nucleotides (e.g., as described herein). The method may comprise detecting one or more signals or signal changes from the labeled nucleotides incorporated into the plurality of growing nucleic acid strands, wherein the one or more signals or signal changes are indicative of the labeled nucleotides having incorporated into the plurality of growing nucleic acid strands.

**[00277]** The optically (e.g., fluorescently) labeled nucleotides of the plurality of nucleotides may be non-terminated. In such cases, the growing strands may incorporate one or more consecutive nucleotides during (e.g., a complimentary base to the plurality of nucleotides in solution is not present at a plurality of positions adjacent to the primer hybridized to the nucleic acid molecule). The one or more signals or signal changes detected from the optically (e.g., fluorescently) labeled nucleotides may be indicative of consecutive nucleotides having incorporated into the plurality of growing nucleic acid strands. Methods for determining a number of fluorophores from the detected signals or signal changes are described elsewhere herein.

**[00278]** Alternatively, the optically (e.g., fluorescently) labeled nucleotides may be terminated. In such cases, each growing strand may incorporate no more than one nucleotide per flow cycle until synthesis is terminated. The one or more signals or signal changes detected from the optically (e.g., fluorescently) labeled nucleotides may be indicative of nucleotides having incorporated into the plurality of growing nucleic acid strands. Prior to, during, or subsequent to detection, a terminating group of the labeled nucleotides may be cleaved (e.g., to facilitate sequencing of homopolymers, and/or to reduce potential context and/or quenching issues).

**[00279]** Alternatively or in addition to, the optically (e.g., fluorescently) labeled nucleotides may include a mixture of terminated and non-terminated nucleotides. In such cases, the growing strands may incorporate one or more consecutive nucleotides generating an extended primer. The solution comprising the plurality of terminated and non-terminated nucleotides may then be washed away from the sequencing template. Un-labeled nucleotides of the plurality of nucleotides may comprise nucleotide moieties of the same type as labeled nucleotides of the plurality of nucleotides (e.g., the same canonical nucleotide).



Date Reçue/Date Received 2023-12-22

**[00280]** In an aspect, the present disclosure provides compositions comprising one or more fluorescently labeled nucleotides and methods of using the same. A composition may comprise a solution comprising a fluorescently labeled nucleotide (e.g., as described herein). The fluorescently labeled nucleotide may comprise a fluorescent dye that is connected to a nucleotide or nucleotide analog (e.g., as described herein) via a linker (e.g., as described herein). The linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems. At least two of the two or more ring systems may be connected to each other by no more than two $sp^3$ carbon atoms, such as by no $sp^3$ carbon atoms. For example, at least two of the two or more ring systems may be connected to each other by no more than two atoms. For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. The fluorescently labeled nucleotide may be configured to emit a fluorescent signal. The fluorescently labeled nucleotide may comprise a plurality of amino acids, such as a plurality of non-proteinogenic (e.g., non-natural) amino acids. For example, the linker may comprise a plurality of hydroxyprolines. At least one water-soluble group of the one or more water-soluble groups may be appended to a ring structure of the two or more ring systems. The one or more water soluble groups may be selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester. The linker may comprise a cleavable group (e.g., an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group) that is configured to be cleaved to separate the fluorescent dye from the nucleotide.

**[00281]** The solution (e.g., nucleotide flow) may comprise a plurality of fluorescently labeled nucleotides, each or which may comprise a fluorescent dye of a same type, a linker of a same type, and a nucleotide of a same type. Each linker of each fluorescently labeled nucleotide of the plurality of fluorescently labeled nucleotides may have the same molecular weight (e.g., they might not comprise polymers with a range of molecular weights). The solution may also comprise a plurality of unlabeled nucleotides, in which each nucleotide of the plurality of unlabeled nucleotides is of a same type as each nucleotide of the plurality of fluorescently labeled nucleotides. The ratio of the plurality of fluorescently labeled nucleotides to the plurality of unlabeled nucleotides in the solution may be at least about 1:4 (e.g., the labeling fraction may be at least 20%). For example, the ratio may be at least 1:1 (e.g., the labeling fraction may be at least 50%). Alternatively, the solution may not comprise any unlabeled nucleotides and the labeling fraction may be 100%.



Date Reçue/Date Received 2023-12-22

**[00282]** The solution (e.g., nucleotide flow) may be provided to a template nucleic acid molecule coupled to a nucleic acid strand. The template nucleic acid molecule may be immobilized to a support (e.g., as described herein). For example, the template nucleic acid molecule may be immobilized to a support via an adapter. For example, the template nucleic acid molecule may be immobilized to a support via a primer to which it is hybridized. The nucleic acid strand may be at least partially complementary to a portion of the template nucleic acid molecule. The template nucleic acid molecule and nucleic acid strand coupled thereto may be subjected to conditions sufficient to incorporate a fluorescently labeled nucleotide of the solution into the nucleic acid strand coupled to the template nucleic acid molecule. Incorporation of the fluorescently labeled nucleotide may be accomplished using a polymerase enzyme (e.g., as described herein). More than one fluorescently labeled nucleotide of the solution may be incorporated, such as into a homopolymeric region of the template nucleic acid molecule. Alternatively or in addition to, an unlabeled nucleotide may be incorporated (e.g., adjacent to the fluorescently labeled nucleotide), such as into a homopolymeric region of the template nucleic acid molecule. A signal (e.g., a fluorescent signal) may be detected from the fluorescently labeled nucleotide incorporated into the nucleic acid strand. Prior to detection of the signal, a wash solution may be used to used to remove fluorescently labeled nucleotides that are not incorporated into the nucleic acid strand. After detection of the signal, the fluorescently labeled nucleotide incorporated into the nucleic acid strand may be contacted with a cleavage reagent configured to cleave the fluorescent dye from the nucleotide. The cleavage reagent may be configured to cleave the linker to provide the nucleotide attached to a portion of the linker, which portion may comprise a thiol moiety, an aromatic moiety, or a combination thereof. The nucleic acid strand, such as a nucleic acid strand of a plurality of nucleic acid strands coupled to a plurality of template nucleic acid molecules, may be contacted with a chase flow comprising only unlabeled nucleotides of a same nucleotide type (e.g., before or after detection of a signal). The nucleic acid strand coupled to the template nucleic acid molecule may also be contacted with one or more additional wash flows. The nucleic acid strand coupled to the template nucleic acid molecule may be contacted with an additional solution comprising an additional fluorescently labeled nucleotide, such as an additional fluorescently labeled nucleotide including a nucleotide of a different type. The dye of the additional fluorescently labeled nucleotide may be of a same type as the dye of the fluorescently labeled nucleotide. Similarly, the linker of the additional fluorescently labeled nucleotide may be of a same type as the linker of the fluorescently labeled nucleotide.



Date Reçue/Date Received 2023-12-22

**[00283]** In another aspect, the present disclosure provides a method comprising providing a fluorescent labeling reagent (e.g., as described herein). The fluorescent labeling reagent may comprise a fluorescent dye and a linker that is connected to the fluorescent dye. The linker may comprise (i) one or more water soluble groups and (ii) two or more ring systems. At least two of the two or more ring systems may be connected to each other by no more than two $sp^3$ carbon atoms, such as by no more than two atoms. For example, at least two of the two or more ring structures may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. The fluorescent labeling reagent may be configured to emit a fluorescent signal. The fluorescent labeling reagent may comprise a plurality of amino acids, such as a plurality of non-proteinogenic (e.g., non-natural) amino acids. For example, the linker may comprise a plurality of hydroxyprolines. At least one water-soluble group of the one or more water-soluble groups may be appended to a ring structure of the two or more ring systems. The one or more water soluble groups may be selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester.

**[00284]** A substrate may be contacted with the fluorescent labeling reagent to generate a fluorescently labeled substrate, in which the linker connected to the fluorescent dye is associated with the substrate. The substrate may be a nucleotide or nucleotide analog (e.g., as described herein). Alternatively, the substrate may be a protein, lipid, cell, or antibody. The fluorescently labeled substrate may be configured to emit a fluorescent signal (e.g., upon excitation at an appropriate energy range), which signal may be detected (e.g., using imaging-based detection). The linker may comprise a cleavable group (e.g., an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group) that is configured to be cleaved to separate the fluorescent dye from the substrate. The fluorescently labeled substrate may be contacted with a cleavage reagent configured to cleave the fluorescent labeling reagent or a portion thereof from the fluorescently labeled substrate to generate a scarred substrate. The scarred substrate may comprise a thiol moiety, an aromatic moiety, or a combination thereof. Prior to generating the scarred substrate, the fluorescently labeled substrate and a nucleic acid molecule may be subjected to conditions sufficient to incorporate the fluorescently labeled substrate into the nucleic acid molecule. Incorporation may be accomplished using a polymerase enzyme (e.g., as described herein). More than one fluorescently labeled substrate may be incorporated, such as into a homopolymeric region of the nucleic acid molecule. For example, an additional fluorescently labeled substrate may be incorporated into a position adjacent to the



Date Reçue/Date Received 2023-12-22

position into which the fluorescently labeled substrate is incorporated. Alternatively or in addition to, an unlabeled substrate (e.g., a nucleotide of a same type as the nucleotide of a fluorescently labeled nucleotide) may also be incorporated into the nucleic acid molecule, such as into adjacent positions of the nucleic acid molecule. Incorporation of an additional fluorescently labeled substrate may be done before or after generation of the scarred substrate. Similarly, incorporation of an unlabeled substrate may be done before or after generation of the scarred substrate.

**[00285]** The nucleic acid molecule, such as a nucleic acid molecule of a plurality of nucleic acid molecules, may be contacted with a chase flow comprising only unlabeled substrates of a same type (e.g., before or after detection of a signal from the nucleic acid molecule). The nucleic acid molecule may also be contacted with one or more additional wash flows. The nucleic acid molecule may be contacted with an additional solution comprising an additional fluorescently labeled substrate, such as an additional fluorescently labeled substrate including a nucleotide of a different type. The dye of the additional fluorescently labeled substrate may be of a same type as the dye of the fluorescently labeled substrate. Similarly, the linker of the additional fluorescently labeled substrate may be of a same type as the linker of the fluorescently labeled substrate.

**[00286]** The nucleic acid molecule may be immobilized to a support (e.g., as described herein). For example, the nucleic acid molecule may be immobilized to a support via an adapter. For example, the nucleic acid molecule may be immobilized to a support via a primer to which it is hybridized. The nucleic acid molecule may comprise a first nucleic acid strand that is at least partially complementary to a portion of a second nucleic acid strand. The second nucleic acid strand may comprise a template nucleic acid sequence, or a complement thereof.

**[00287]** The labeled nucleotides of the present disclosure may be used during sequencing operations that involve a high fraction of labeled nucleotides. For example, the present disclosure provides a method comprising contacting a nucleic acid molecule (e.g., a template nucleic acid molecule) with a solution comprising a plurality of nucleotides under conditions sufficient to incorporate a first labeled nucleotide and a second labeled nucleotide of the plurality of nucleotides into a growing strand that is at least partially complementary to the nucleic acid molecule. The first labeled nucleotide and the second labeled nucleotide may be of a same canonical base type. The first nucleotide may comprise a fluorescent dye (e.g., as described herein), which fluorescent dye may be associated with the first nucleotide via a linker (e.g., as described herein). The second nucleotide may comprise the same fluorescent dye (e.g., associated with the second nucleotide via a linker having the same chemical structure of the



Date Reçue/Date Received 2023-12-22

linker associating the first nucleotide and the fluorescent dye). A fluorescent dye coupled to a nucleotide (e.g., the first and/or second nucleotide) may be cleavable (e.g., upon application of a cleavage reagent). At least about 20% of the plurality of nucleotides may be labeled nucleotides. For example, at least 20% of the plurality of nucleotides may be associated with a fluorescent labeling reagent (e.g., as described herein). For example, at least about 50%, 70%, 80%, 90%, 95%, or 99% of the plurality of nucleotides may be labeled nucleotides. For example, all of the nucleotides of the plurality of nucleotides may be labeled nucleotides (e.g., the labeling fraction may be 100%). One or more signals or signal changes may be detected from the first labeled nucleotide and the second labeled nucleotide (e.g., as described herein). The one or more signals or signal changes may comprise fluorescent signals or signal changes. The one or more signals or signal changes may be indicative of incorporation of the first labeled nucleotide and the second labeled nucleotide. The one or more signals or signal changes may be resolved to determine a sequence of the nucleic acid molecule, or a portion thereof. Resolving the one or more signals or signal changes may comprise determining a number of consecutive nucleotides from the solution that incorporated into the growing strand. The number of consecutive nucleotides may be selected from the group consisting of 2, 3, 4, 5, 6, 7, or 8 nucleotides. Resolving the one or more signals or signal changes may comprise processing a tolerance of the solution. A third nucleotide may also be incorporated into the growing strand (e.g., before or after detection of the one or more signals or signal changes). The third nucleotide may be a nucleotide of the plurality of nucleotides of the solution. Alternatively, the third nucleotide may be provided in a separate solution, such as in a "chase" flow (e.g., as described herein). The third nucleotide may be unlabeled. Alternatively, the third nucleotide may be labeled. The first labeled nucleotide and the third nucleotide may be of a same canonical base type. Alternatively, the first labeled nucleotide and the third nucleotide may be of different canonical base types.

**[00288]** The method may further comprise cleaving the fluorescent dye coupled to the first labeled nucleotide. The fluorescent dye may be cleaved by application of a cleavage reagent configured to cleave a linker associating the first labeled nucleotide and the fluorescent dye. The nucleic acid molecule may be contacted with a second solution comprising a second plurality of nucleotides under conditions sufficient to incorporate a third labeled nucleotide of the second plurality of nucleotides into the growing strand. At least about 20% of the second plurality of nucleotides may be labeled nucleotides (e.g., as described herein). One or more second signals or signal changes may be detected from the third labeled nucleotide (e.g., as described herein). The one or more second signals or signal changes may be resolved to determine a second sequence of the nucleic acid molecule, or a portion thereof. The first labeled nucleotide and the



Date Reçue/Date Received 2023-12-22

third labeled nucleotide may be different canonical base types (e.g., A, C, U/T, or G). The third labeled nucleotide may comprise the fluorescent dye. The fluorescent dye may be coupled to the third labeled nucleotide via a linker (e.g., as described herein), which linker may have the same chemical structure as the linker connecting the fluorescent dye to the first labeled nucleotide or a different chemical structure.

**[00289]** Alternatively, the method may comprise contacting the nucleic acid molecule with a second solution comprising a second plurality of nucleotides under conditions sufficient to incorporate a third labeled nucleotide of the second plurality of nucleotides into the growing strand. At least about 20% of the second plurality of nucleotides may be labeled nucleotides (e.g., as described herein). One or more second signals or signal changes may be detected from the third labeled nucleotide (e.g., as described herein). The one or more second signals or signal changes may be resolved to determine a second sequence of the nucleic acid molecule, or a portion thereof. The first labeled nucleotide and the third labeled nucleotide may be different canonical base types (e.g., A, C, U/T, or G). The third labeled nucleotide may comprise the fluorescent dye. The fluorescent dye may be coupled to the third labeled nucleotide via a linker (e.g., as described herein), which linker may have the same chemical structure as the linker connecting the fluorescent dye to the first labeled nucleotide or a different chemical structure. Contacting the nucleic acid molecule with the second solution may be performed in absence of cleaving a fluorescent dye from the first labeled nucleotide or the second labeled nucleotide. This process may be repeated one or more times, such as 1, 2, 3, 4, 5, or more times, each with a different solution of nucleotides, in absence of cleaving a fluorescent dye from the first labeled nucleotide or the second labeled nucleotide. One or more of these different solutions of nucleotides may comprise at least 20% labeled nucleotides.

**[00290]** The present disclosure also provides a method comprising contacting a nucleic acid molecule with a solution comprising a plurality of non-terminated nucleotides under conditions sufficient to incorporate a labeled nucleotide and a second nucleotide of the plurality of non-terminated nucleotides into a growing strand that is at least partly complementary to the nucleic acid molecule, or a portion thereof. The labeled nucleotide and the second nucleotide may be of a same canonical base type. Alternatively, the labeled nucleotide and the second nucleotide may be of different canonical base types. The labeled nucleotide may comprise a fluorescent dye (e.g., as described herein), which fluorescent dye may be associated with the labeled nucleotide via a linker (e.g., as described herein). The second nucleotide may be a labeled nucleotide. For example, the second nucleotide may comprise the same fluorescent dye (e.g., associated with the second nucleotide via a linker having the same chemical structure of



Date Reçue/Date Received 2023-12-22

the linker associating the first nucleotide and the fluorescent dye). Alternatively, the second nucleotide may not be coupled to a fluorescent dye (e.g., the second nucleotide may be unlabeled). A fluorescent dye coupled to a nucleotide (e.g., the first and/or second nucleotide) may be cleavable (e.g., upon application of a cleavage reagent). The plurality of non-terminated nucleotides may comprise nucleotides of a same canonical base type. At least about 20% of said plurality of nucleotides may be labeled nucleotides. For example, at least 20% of the plurality of nucleotides may be associated with a fluorescent labeling reagent (e.g., as described herein). For example, at least about 50%, 70%, 80%, 90%, 95%, or 99% of the plurality of non-terminated nucleotides may be labeled nucleotides. For example, substantially all of the plurality of non-terminated nucleotides may be labeled nucleotides. For example, all of the nucleotides of the plurality of non-terminated nucleotides may be labeled nucleotides (e.g., the labeling fraction may be 100%). One or more signals or signal changes may be detected from the labeled nucleotide (e.g., as described herein). The one or more signals or signal changes may comprise fluorescent signals or signal changes. The one or more signals or signal changes may be indicative of incorporation of the labeled nucleotide. The one or more signals or signal changes may be resolved to determine a sequence of the nucleic acid molecule, or a portion thereof. Resolving the one or more signals or signal changes may comprise determining a number of consecutive nucleotides from the solution that incorporated into the growing strand. The number of consecutive nucleotides may be selected from the group consisting of 2, 3, 4, 5, 6, 7, or 8 nucleotides. Resolving the one or more signals or signal changes may comprise processing a tolerance of the solution. A third nucleotide may also be incorporated into the growing strand (e.g., before or after detection of the one or more signals or signal changes). The third nucleotide may be a nucleotide of the plurality of non-terminated nucleotides of the solution. Alternatively, the third nucleotide may be provided in a separate solution, such as in a "chase" flow (e.g., as described herein). The third nucleotide may be unlabeled. Alternatively, the third nucleotide may be labeled. The labeled nucleotide and the third nucleotide may be of a same canonical base type. Alternatively, the labeled nucleotide and the third nucleotide may be of different canonical base types.

**[00291]** The method may further comprise cleaving the fluorescent dye coupled to the labeled nucleotide. The fluorescent dye may be cleaved by application of a cleavage reagent configured to cleave a linker associating the labeled nucleotide and the fluorescent dye. The nucleic acid molecule may be contacted with a second solution comprising a second plurality of non-terminated nucleotides under conditions sufficient to incorporate a third labeled nucleotide of the second plurality of non-terminated nucleotides into the growing strand. At least about



Date Reçue/Date Received 2023-12-22

20% of the second plurality of non-terminated nucleotides may be labeled nucleotides (e.g., as described herein). One or more second signals or signal changes may be detected from the third labeled nucleotide (e.g., as described herein). The one or more second signals or signal changes may be resolved to determine a second sequence of the nucleic acid molecule, or a portion thereof. The first labeled nucleotide and the third labeled nucleotide may be different canonical base types (e.g., A, C, U/T, or G). The third labeled nucleotide may comprise the fluorescent dye. The fluorescent dye may be coupled to the third labeled nucleotide via a linker (e.g., as described herein), which linker may have the same chemical structure as the linker connecting the fluorescent dye to the first labeled nucleotide or a different chemical structure.

**[00292]** Alternatively, the method may comprise contacting the nucleic acid molecule with a second solution comprising a second plurality of non-terminated nucleotides under conditions sufficient to incorporate a third labeled nucleotide of the second plurality of non-terminated nucleotides into the growing strand. At least about 20% of the second plurality of nucleotides may be labeled nucleotides (e.g., as described herein). One or more second signals or signal changes may be detected from the third labeled nucleotide (e.g., as described herein). The one or more second signals or signal changes may be resolved to determine a second sequence of the nucleic acid molecule, or a portion thereof. The first labeled nucleotide and the third labeled nucleotide may be different canonical base types (e.g., A, C, U/T, or G). The third labeled nucleotide may comprise the fluorescent dye. The fluorescent dye may be coupled to the third labeled nucleotide via a linker (e.g., as described herein), which linker may have the same chemical structure as the linker connecting the fluorescent dye to the first labeled nucleotide or a different chemical structure. Contacting the nucleic acid molecule with the second solution may be performed in absence of cleaving a fluorescent dye from the first labeled nucleotide or the second labeled nucleotide. This process may be repeated one or more times, such as 1, 2, 3, 4, 5, or more times, each with a different solution of nucleotides, in absence of cleaving a fluorescent dye from the first labeled nucleotide or the second labeled nucleotide. One or more of these different solutions of nucleotides may comprise at least 20% labeled nucleotides.

**Methods for synthesis of optical labeling reagents**

**[00293]** In some cases, the linkers provided herein may be prepared using peptide synthesis chemistry.

**[00294]** For example, a linker comprising a pyridinium moiety may be prepared using peptide synthesis chemistry. Such a method may use four bifunctional reagents to make the linker, namely: (a) $R^1A$, (b) BB, (c) AA, and (d) $AR^2$. Reagent A reacts with B to form a



Date Reçue/Date Received 2023-12-22

pyridinium group; $R^1$ and $R^2$ are hetero-bifunctional attachment groups. The synthesis begins with the group $R^1A$ (or $R^2A$). Excess BB is added to $R^1A$ to form $R^1A$-BB. The product is precipitated and washed in a less polar solvent (such as ethyl acetate or tetrahydrofuran) to remove excess BB. Excess AA is added with heat in N-methylpyrrolidone (NMP) to produce $R^1A$-BB-AA. The product is precipitated and washed in a less polar solvent. The synthesis proceeds until a linker of a particular length is formed. The group $AR^2$ is appended in the final step.

1) $R^1A + 10BB \rightarrow R^1A$-BB (wash away excess BB)

2) $R^1A$-BB + 10 AA $\rightarrow R^1A$-BB-AA (wash away excess AA)

3) $R^1A$-BB-AA + 10 BB $\rightarrow R^1A$-BB-AA-BB (wash away excess BB)

4) $R^1A$-BB-AA-BB + $AR^2 \rightarrow R^1A$-BB-AA-BB-$AR^2$ (use terminating reagent)

**[00295]** **FIG. 2A** shows an example of a method for synthesizing a linker of the present disclosure having an effective length of about 2 nanometers.

**[00296]** **FIG. 2B** shows examples of reagents that can be used in the method of **FIG. 2A** for synthesizing a linker of the present disclosure, as well as some trifunctional reagents.

**[00297]** **FIG. 2C** shows an example of a method for synthesizing a linker of the present disclosure that is polymeric with defined molecular weight and linking groups.

**[00298]** Additional synthetic methods for preparing optical labeling reagents (e.g., as described herein) are described elsewhere and in the Examples below.

**Methods for constructing labeled nucleotides**

**[00299]** In an aspect, the present disclosure provides methods for constructing labeled nucleotides (e.g., optically labeled nucleotides).

**[00300]** Labeled nucleotides can be constructed using modular chemical building blocks. A nucleotide or nucleotide analog can be derivatized with, e.g., a propargylamino moiety to provide a handle for attachment to a linker or detectable label (e.g., dye). One or more detectable labels, such as one or more dyes, can be attached to a nucleotide or nucleotide analog via a covalent bond. Alternatively or in addition to, one or more detectable labels can be attached to a nucleotide or nucleotide analog via a non-covalent bond. A detectable label may be attached to a nucleotide or nucleotide analog via a linker (e.g., as described herein). A linker may include one or more moieties. For example, a linker may include a first moiety including a disulfide bond within it to facilitate cleaving the linker and releasing the detectable label (e.g., during a sequencing process). Additional linker moieties can be added using sequential peptide bonds. Linker moieties can have various lengths and charges. A linker moiety may include one or more



Date Reçue/Date Received 2023-12-22

different components, such as one or more different ring systems, and/or a repeating unit (e.g., as described herein). Examples of linkers include, but are not limited to, aminoethyl-SS-propionic acid (epSS), aminoethyl-SS-benzoic acid, aminohexyl-SS-propionic acid, hyp10, and hyp20.

**[00301]** Examples of methods for constructing labeled nucleotides are shown in **FIGs. 4, 5A,** and **5B**. As shown in **FIG. 4**, a labeled nucleotide may be constructed from a nucleotide, a dye, and one or more linker moieties. The one or more linker moieties together comprise a linker as described herein. A nucleotide functionalized with a propargylamino moiety can be attached to a first linker moiety via a peptide bond. This first linker moiety may comprise a cleavable moiety, such as a disulfide moiety. The first linker moiety can also be attached to one or more additional linker moieties in linear or branching fashions. For example, a second linker moiety may include two or more ring systems, wherein at least two of the two or more ring systems are separated by no more than two sp3 carbon atoms, such as by no more than two atoms. For example, at least two of the two or more ring systems may be connected to each other by an $sp^2$ carbon atom. The linker may comprise a non-proteinogenic amino acid comprising a ring system of the two or more ring systems. For example, the second linker moiety may comprise a two or more hydroxyproline moieties. An amine handle on a linker moiety may be used to attach the linker and a dye, such as a dye that fluoresces in the red or green portions of the visible electromagnetic spectrum. The labeled nucleotide generated in **FIG. 4** comprises a modified deoxyadeninosine triphosphate moiety, a linker comprising a first linker moiety including a disulfide moiety and a second linker moiety including at least two ring systems, and a dye.

**[00302]** Construction of a labeled nucleotide can begin from either the nucleotide terminus or the dye terminus. Construction from the dye terminus permits the use of unlabeled, unactivated amino acid moieties, while construction from the nucleotide terminus may require amine-protected, carboxy-activated amino acid moieties.

**[00303]** **FIGs. 5A** and **5B** show an example synthesis of a labeled nucleotide including a propargylamino functionalized dGTP moiety, a first linker moiety including a disulfide group, a second linker moiety that is hyp10, and the dye moiety Atto633. Details of this synthesis are provided in Example 3 below.

**[00304]** A nucleotide or nucleotide analog of a labeled nucleotide may include one or more modifications, such as one or more modifications on the nucleobase. Alternatively, a nucleotide or nucleotide analog of a labeled nucleotide may include one or more modifications not on the nucleobase. Modifications can include, but are not limited to, covalent attachment of



Date Reçue/Date Received 2023-12-22

one or more linker or label moieties, alkylation, amination, amidation, esterification, hydroxylation, halogenation, sulfurylation, and/or phosphorylation.

**[00305]** A nucleotide or nucleotide analog of a labeled nucleotide may include one or more modifications that are configured prevent subsequent nucleotide additions to a position adjacent to the labeled nucleotide upon its incorporation into a growing nucleic acid strand. For example, the labeled nucleotide may include a terminating or blocking group (e.g., dimethoxytrityl, phosphoramidite, or nitrobenzyl molecules). In some instances, the terminating or blocking group may be cleavable.

**Computer systems**

**[00306]** The present disclosure provides computer systems that are programmed to implement methods of the disclosure. **FIG. 3** shows a computer system **301** that is programmed or otherwise configured to perform nucleic acid sequencing. The computer system **301** can determine sequence reads based at least in part on intensities of detected optical signals. The computer system **301** can regulate various aspects of the present disclosure, such as, for example, performing nucleic acid sequencing, sequence analysis, and regulating conditions of transient binding and non-transient binding (e.g., incorporation) of nucleotides. The computer system **301** can be an electronic device of a user or a computer system that is remotely located with respect to the electronic device. The electronic device can be a mobile electronic device.

**[00307]** The computer system **301** includes a central processing unit (CPU, also "processor" and "computer processor" herein) **305**, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system **301** also includes memory or memory location **310** (e.g., random-access memory, read-only memory, flash memory), electronic storage unit **315** (e.g., hard disk), communication interface **320** (e.g., network adapter) for communicating with one or more other systems, and peripheral devices **325**, such as cache, other memory, data storage and/or electronic display adapters. The memory **310**, storage unit **315**, interface **320** and peripheral devices **325** are in communication with the CPU 305 through a communication bus (solid lines), such as a motherboard. The storage unit **315** can be a data storage unit (or data repository) for storing data. The computer system **301** can be operatively coupled to a computer network ("network") **330** with the aid of the communication interface **320**. The network **330** can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network **330** in some cases is a telecommunication and/or data network. The network **330** can include one or more computer servers, which can enable distributed computing, such as cloud computing. The



Date Reçue/Date Received 2023-12-22

network **330**, in some cases with the aid of the computer system **301**, can implement a peer-to-peer network, which may enable devices coupled to the computer system **301** to behave as a client or a server.

**[00308]** The CPU **305** can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions may be stored in a memory location, such as the memory **310**. The instructions can be directed to the CPU **305**, which can subsequently program or otherwise configure the CPU **305** to implement methods of the present disclosure. Examples of operations performed by the CPU **305** can include fetch, decode, execute, and writeback.

**[00309]** The CPU **305** can be part of a circuit, such as an integrated circuit. One or more other components of the system **301** can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

**[00310]** The storage unit **315** can store files, such as drivers, libraries and saved programs. The storage unit **315** can store user data, e.g., user preferences and user programs. The computer system **301** in some cases can include one or more additional data storage units that are external to the computer system **301**, such as located on a remote server that is in communication with the computer system **301** through an intranet or the Internet.

**[00311]** The computer system **301** can communicate with one or more remote computer systems through the network **330**. For instance, the computer system **301** can communicate with a remote computer system of a user. Examples of remote computer systems include personal computers (e.g., portable PC), slate or tablet PC's (e.g., Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (e.g., Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system 301 via the network **330**.

**[00312]** Methods as described herein can be implemented by way of machine (e.g., computer processor) executable code stored on an electronic storage location of the computer system **301**, such as, for example, on the memory **310** or electronic storage unit **315**. The machine executable or machine readable code can be provided in the form of software. During use, the code can be executed by the processor **305**. In some cases, the code can be retrieved from the storage unit **315** and stored on the memory **310** for ready access by the processor **305**. In some situations, the electronic storage unit **315** can be precluded, and machine-executable instructions are stored on memory **310**.

**[00313]** The code can be pre-compiled and configured for use with a machine having a processer adapted to execute the code, or can be compiled during runtime. The code can be



Date Reçue/Date Received 2023-12-22

supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

**[00314]** Aspects of the systems and methods provided herein, such as the computer system **301**, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such as memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. "Storage" type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" refer to any medium that participates in providing instructions to a processor for execution.

**[00315]** Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch



Date Reçue/Date Received 2023-12-22

cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

**[00316]** The computer system **301** can include or be in communication with an electronic display **335** that comprises a user interface (UI) **340** for providing, for example, results of nucleic acid sequence and optical signal detection (e.g., sequence reads, intensity maps, etc.). Examples of UI's include, without limitation, a graphical user interface (GUI) and web-based user interface.

**[00317]** Methods and systems of the present disclosure can be implemented by way of one or more algorithms. An algorithm can be implemented by way of software upon execution by the central processing unit **305**. The algorithm can, for example, implement methods and systems of the present disclosure, such as determine sequence reads based at least in part on intensities of detected optical signals.

## EXAMPLES

**Example 1: General Synthetic Principles**

**[00318]** Certain examples of the following examples illustrate various methods of making linkers and labeled substrates described herein. It is understood that one skilled in the art may be able to make these compounds by similar methods or by combining other methods known to one skilled in the art. It is also understood that one skilled in the art would be able to make other compounds in a similar manner as described below by using the appropriate starting materials and modifying synthetic routes as needed. In general, starting materials and reagents can be obtained from commercial vendors or synthesized according to sources known to those skilled in the art or prepared as described herein.

**[00319]** Unless otherwise noted, reagents and solvents used in synthetic methods described herein are obtained from commercial suppliers. Anhydrous solvents and oven-dried glassware may be used for synthetic transformations sensitive to moisture and/or oxygen. Yields may not be optimized. Reaction times may be approximate and may not be optimized. Materials and instrumentation used in synthetic procedures may be substituted with appropriate alternatives. Column chromatography and thin layer chromatography (TLC) may be performed

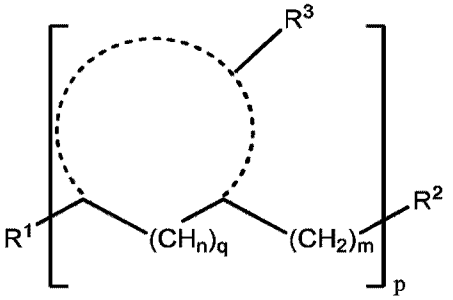


Date Reçue/Date Received 2023-12-22

on reverse-phase silica gel unless otherwise noted. Nuclear magnetic resonance (NMR) and mass spectra may be obtained to characterize reaction products and/or monitor reaction progress.

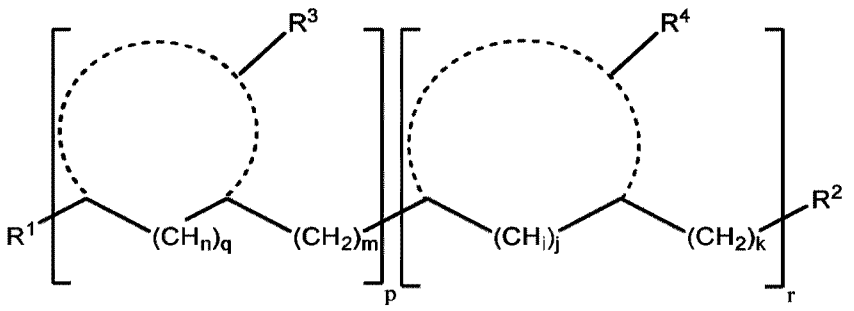
**Example 2: A structure of a labeling reagent**

**[00320]** Described herein is an example of a semi-rigid, water-soluble linker of a defined molecular weight that can efficiently accomplish a dye-dye or dye-quencher separation. A semi-rigid structure can be achieved through a series of linked, aromatic or non-aromatic ring systems connected by zero or one linkages with $sp^3$ bonding, and zero or more sp or $sp^2$ bonds. Water-solubility can be achieved with the inclusion (e.g., in each subunit) of at least one of the moieties selected from the group: hydroxyl, pyridinium, imidazolium, sulfonate, amino, thiol, carboxyl, and quaternary ammonium. A linker can be a hetero- or homobi-(or tri-)functional reagent that allows attachment of a dye (e.g., fluorescent dye) at one end and a biological ligand (e.g., a nucleotide) at the other end. An example of a general formula for such a linker is shown below:




in which *p* is a number of repeating units selected from 1-100; each $R^3$ is a water-soluble moiety independently selected from, for example, pyridinium and sulfonate; $R^1$ and $R^2$ are attachment groups such as amino and carboxy moieties; each *n* is independently 1 or 2; each *m* is independently selected from 1 and 2; and each *q* is independently selected from 4-8. In the structure above, *m* represents the number of $sp^3$ carbons linking ring moieties to one another. A ring moiety may be an aliphatic or an aromatic ring.

**[00321]** Multiple such subunits may be connected to one another. For example, a linker may be represented by the below formula:






Date Reçue/Date Received 2023-12-22

in which $p$ and $r$ are each a number of repeating units independently selected from 1-100; each $R^3$ and $R^4$ is a water-soluble moiety independently selected from, for example, pyridinium and sulfonate; $R^1$ and $R^2$ are attachment groups such as amino and carboxy moieties; each $n$ and $i$ is independently 1 or 2; each $m$ and $k$ is independently selected from 1 and 2; and each $q$ and $j$ is independently selected from 4-8. In the structure above, $m$ and $k$ represent the number of $sp^3$ carbons linking ring moieties to one another. A ring moiety may be an aliphatic or an aromatic ring. In some cases, ring moieties in the left portion of the structure are aliphatic and ring moieties in the right portion of the structure are aromatic, or vice versa.

**[00322]** Note that the above structures do not capture all embodiments of the disclosure. For example, the linker does not have to be a polymer of "P-repeating" units. Similarly, the water-soluble functional group can be a constituent component of the ring rather than attached to the ring.

**Example 3: Synthesis of dGTP-AP-SS-hyp10-Atto633**

**[00323]** Described herein is a method for constructing the labeled nucleotide dGTP-AP-SS-hyp10-Atto633. **FIG. 5A** illustrates an example method for the synthesis of a fluorescently labeled dGTP reagent. **FIG. 5B** illustrates the same synthesis with the full structures of the dye and linker. The method involves formation of a covalent linkage between Gly-Hyp10 and the fluorophore Atto633 (process (a)), esterification to couple Atto633-Gly-Hyp10 with pentafluorophenol (process (b)), substitution with the linker molecule epSS (process(c)), esterification to form Atto633-Gly-Hyp10-epSS-PFP (process (d)), and substitution with dGTP to provide the fluorescently labeled nucleotide(process (e)). Details of the synthesis are provided below.

**[00324]** *Preparation of Atto633-Gly-Hyp10.* (**FIG. 5** process (a)) A stock solution of Gly-Hyp10 (also referred to herein as "hyp10") in bicarbonate is prepared by dissolving 25 milligrams (mg) of the 11 amino acid peptide in 500 microliters (μL) of 0.2 molar (M) sodium bicarbonate in a 1.5 milliliter (mL) Eppendorf tube. 7 mg of Atto633-NHS is weighed into another Eppendorf tube and dissolved in 200 μL of dimethylformamide (DMF). A volume of 300μL of the peptide solution is added to the solution containing Atto633-NHS. The resulting solution is mixed and heated to 50°C for 20 minutes (min). The extent of the reaction is followed with reverse-phase thin layer chromatography (TLC). A 1 μL aliquot of the reaction solution is removed and dissolved in 40 μL water and spotted on reverse phase TLC. A co-spot with Atto633 acid is included, and Atto633 is also run alone. The plate is eluted with a 2:1 solution of acetonitrile 0.1M triethylammonium acetate (TEAA). Atto633 acid and Atto633-



Date Reçue/Date Received 2023-12-22

NHS both have an $R_f$ of zero, while Gly-Hyp10 has an $R_f$ of 0.4. The product is purified by injecting the solution onto a C18 reverse phase column using the gradient 20%→50% acetonitrile vs. 0.1M TEAA over 16 minutes at 2.5 mL/min. The desired product is the major product, Atto633-Gly-Hyp10, eluting at 15.2 minutes. The fractions containing the desired material are collected in Eppendorf tubes and dried, yielding a blue solid. A major peak was observed on ESI mass spec: m/z calculated for $C_{87}H_{115}N_{14}O_{24}^+$, $[M]^+ = 1739.8$; found: 1740.6.

**[00325]** *Preparation of Atto633-Gly-Hyp10-PFP.* (**FIG. 5** process (b)) Atto633-Gly-Hyp10 is suspended in 100 µL DMF in a 1.5 mL Eppendorf tube. Pyridine (20 µL) and pentafluorophenyl trifluoroacetate (PFP-TFA, 20 µL) are added to the tube. The reaction mixture is warmed to 50°C in a heat block for 20 min. The reaction is monitored by removing 1 µL aliquots and adding to 1 mL of dilute HCl (0.4%). When the reaction is complete the aqueous solution is colorless. After 10 min the dilute HCl solution is light blue. Additional PFP-TFA (30 µL) is added. After another 100 min at 50°C a retest of precipitation gives a colorless solution. The remaining reaction mixture is precipitated into 1 mL dilute HCl in 20 µL portions. 20 µL is added to 1 mL dilute HCl, the tube spun down, and aqueous solution discarded. The process is repeated until all of the product is precipitated. The residue is thoroughly dried. After drying, the solid is washed twice with 1 mL methyl tert-butyl ether (MTBE). The product is a dark blue powder. The product gives a major peak on electrospray ionization (ESI)-mass spectrometry (MS): m/z calculated for $C_{93}H_{115}F_5N_{14}O_{24}^{2+}$, $[M + H]^{2+} = 1906.8/2 = 953.4$; found: 953.4.

**[00326]** *Preparation of Atto633-Gly-Hyp10-epSS.* (**FIG. 5** process (c)) Atto633-Gly-Hyp10-PFP (1.6 micromoles (µmol)) is dissolved in 100 µL DMF in an Eppendorf tube. A solution of aminoethyl-SS-propionic acid (Broadpharm; 6 mg in 200 µL 0.1 M bicarbonate) is mixed with the Atto633-gly-hyp10-PFP and heated to 50°C in a heat block for 20 min. Atto633-Gly-Hyp10-epSS is purified from the resulting reaction mixture by reverse phase HPLC using a gradient of 20%→50% acetonitrile over 16 min. Atto633-Gly-Hyp10 elutes at 15 min and Atto633-Gly-Hyp10-epSS elutes at 15.6 min. The fractions containing the product, Atto633-Gly-Hyp10-epSS, are combined and dried. The product has a major peak on ESI-MS: m/z calculated for $C_{92}H_{124}N_{15}O_{25}S_2^+$, $[M]^+ = 1902.8$; Found: 1902.6.

**[00327]** *Preparation of Atto633-Gly-Hyp10-epSS-PFP.* (**FIG. 5** process (d)) Atto633-Gly-Hyp10-epSS is dissolved in 100 µL DMF in an Eppendorf tube. Pyridine (20 µL) and PFP-TFA (20 µL) are added and the mixture is heated to 50°C in a heat block for 20 min. A test aliquot (1 µL) in dilute HCl gives a colorless solution and a blue precipitate. The reaction is precipitated in 20 µL aliquots in 1 mL dilute HCl, the tube spun down, and the aqueous solution



Date Reçue/Date Received 2023-12-22

discarded. The process is repeated until all of the PFP ester is precipitated. The residue is thoroughly dried under vacuum and washed with MTBE.

**[00328]** *Preparation of dGTP-AP-SS-Atto633.* (**FIG. 5** process (e)) A solution of aminopropargyl dGTP (Trilink; 1 μmol in 100 μL of 0.2 M bicarbonate) is added to 50 μL of a DMF solution comprising Atto633-gly-hyp10-epSS-PFP. The mixture is heated to 50°C for 10 min. The product, dGTP-AP-epSS-Atto633, is purified by reverse-phase HPLC using a gradient of 20%→50% acetonitrile 16 min. The product elutes at 15.3 min. Preparative HPLC provides 0.65 μmol. The product gives a major peak on ESI-MS: m/z calculated for $C_{106}H_{139}N_{20}O_{37}P_3S_2^{2-}$, $[M-H]^{2-}$, 1220.4; found: 1220.6.

**[00329]** While synthesis of dGTP-Atto633-Gly-Hyp0-epSS-PFP is described, a skilled practitioner will recognize that other fluorescently labeled nucleotides can be produced in a similar manner using appropriate starting materials.

**Example 4: Synthesis of dCTP-epSS-Atto633**

**[00330]** dCTP-SS12-Atto633 can be prepared in manner similar to the method outlined in Example 3. Briefly, Atto633-epSS is prepared (**FIG. 6** process (a)) by mixing a 200 μL DMF solution comprising 11 mg Atto633-NHS with a 200 μL aqueous solution comprising 0.2M sodium bicarbonate and 24 mg epSS, heating the resulting mixture to 50°C for 15 min, purifying Atto633-epSS from the mixture by reverse phase HPLC using a gradient of 40%→60% acetonitrile vs. 0.1 M TEAA over 16 min at 4.5 mL/min, and confirming the product identity with ESI-MS. The product elutes at 7.3 min and the free dye elutes at 6.4 min. The yield is about 80%. The product gives a major peak on ESI-MS: calculated for $C_{40}H_{51}N_4O_4S_2^+$, [M]+ = 715.3; Found [M]+ = 715.3.

**[00331]** Atto633-epSS is then converted to Atto-epSS-PFP (**FIG. 6** process (b)) by mixing a solution of Atto633-epSS dissolved in 100 μL DMF, 20 μL pyridine, and 20 μL PFP-TFA; heating the solution at 50°C for 5 min before adding an additional 20-40 μL PFP-TFA; heating back to 50°C for 5 min; and precipitating the product in 1 mL of dilute HCl. The product is washed with an additional 1 mL of dilute HCl and the supernatant removed by pipette and evaporation, yielding a blue solid.

**[00332]** dCTP-epSS-Atto633 is formed by reacting Atto-epSS-PFP with aminopropargyl dCTP (AP-dCTP) (**FIG. 6** process (c)). AP-dCTP stock solution (Trilink; 1 μmol) is added to a 100 μL DMF solution comprising 0.2 M sodium bicarbonate and combined with a solution of Atto-epSS-PFP dissolved in 100 μL DMF. The mixture is left to sit overnight. dCTP-epSS-Atto633 is purified from the mixture on a C18 reverse phase column using the gradient



Date Reçue/Date Received 2023-12-22

20%→100% acetonitrile vs. 0.1 M TEAA over 16 minutes at 2.5 mL/min. The product elutes at 10.7 min. The fractions including the product are collected and dried. The product gives a major peak on ESI-MS: Calculated m/z for $C_{52}H_{66}N_8O_{16}P_3S_2^-$, $[M]^- = 1215.3$; found: 1215.5.

**Example 5: Preparation of dye-labeled nucleotides**

**[00333]** A set of dye-labeled nucleotides designed for excitation at about 530 nm is prepared. Excitation at 530 nm may be achieved using a green laser, which may be readily available, high-powered, and stable. There are many commercially available fluorescent dyes with excitation at or near 530 nm that are inexpensive and have a variety of properties (hydrophobic, hydrophilic, positively charged, negatively charged). Synthetic routes to such dyes may be shorter and cheaper than those for longer wavelength dyes. Moreover, certain green dyes may have significantly less self-quenching than red dyes, potentially allowing for the use of higher labeling fractions (e.g., as described herein).

**[00334]** A viable reagent set for use in, e.g., a sequencing application consists of each of four canonical nucleotides or analogs thereof with cleavable green dyes that perform well in sequencing. An optimal set may be prepared by varying each component of a labeled nucleotide structure to obtain an array of candidate labeled nucleotides with varying properties. The resultant nucleotides are evaluated (e.g., as described below), and certain labeled nucleotides are optimized for concentration and labeling fraction (the ratio of labeled to unlabeled nucleotide in a flow).

**[00335]** **FIG. 7** shows a variety of components used in the construction of candidate labeled nucleotides. Each of four propargylamino functionalized nucleotides (A, C, G, and U) can be modified with one of two cleavable linkers, E and B; a hydroxyproline linker (hyp10) or not; and one of three fluorescent dyes, *, #, and $. Using these components, there are 48 possible nucleotide variations. The labeled nucleotides may be prepared according to the synthetic route and principles described herein. An example synthesis of the G*-B-H labeled nucleotide is described in Example 6.

**Example 6: Synthesis of G*-B-H labeled nucleotide**

**[00336]** A synthetic method for preparing G*-B-H (see Example 5) is shown in **FIG. 8**. Similar methods may be used to prepare other labeled nucleotides described in Example 5 and elsewhere herein. As the components used include amino acids, there are multiple routes to the final product. Synthetic considerations include the tendency for hydrolysis of the triphosphate (to the diphosphate and monophosphates) under heat or acidic conditions, the tendency for



Date Reçue/Date Received 2023-12-22

disulfide to decompose in the presence of triethylamine and ammonia, preventing the use of acid-labile protecting groups, and preventing the use of trifluoroacetamide or FMOC protecting groups.

**[00337]** *Preparation of PN 40142.* A solution of Atto 532 succinimidyl ester (Atto-tec, PN 40183; 5 mg = 4.6 μmol) in 100 μL of DMF is mixed with gly-hyp-hyp-hyp-hyp-hyp-hyp-hyp-hyp-hyp-hyp (custom synthesis from Genscript, PN 40035; 8.5 mg = 7 μmol) in 170 μL 0.1 M bicarbonate in a 1.5 mL Eppendorf tube. The reaction is purified on a Phenomenex reverse phase C18 semi-prep column (Gemini 5 μM C18, 250 x 10 mm) using a gradient of 10%→40% acetonitrile vs. 0.1 M triethylammonium acetate over 16 minutes. The fractions containing product 40142 are combined and concentrated to dryness. The yield is determined by diluting a fraction and measuring the optical density (OD) at 633 nm and using an extinction coefficient for the dye of 130,000 $cm^{-1}M^{-1}$. The yield is 50%. The structure is confirmed by mass spectrometry in negative ion mode: m/z calculated for $C_{81}H_{103}N_{14}O_{31}S_2^-$, 1831.6; found: 1831.8.

**[00338]** *Preparation of PN 40143.* PN 40142 (4 μmol) is suspended in 100 μL DMF in a 1.5 mL eppendorf tube. Pyridine (20 μL) and pentafluorophenyl trifluoroacetate (20 μL) are added to the DMF solution and heated to 50 °C for five minutes. A portion (1 μL) of the reaction mixture is precipitated into 0.4% HCl; the aqueous solution remains colorless, indicating complete conversion to the active, pentafluorophenyl ester. The remainder of the reaction is precipitated into the dilute acidic solution and the aqueous solution pipetted off. The residue is washed with hexane and dried to a highly colored solid (PN 40143)

**[00339]** *Preparation of PN 40146.* PN 40143 is dissolved in 100 μL DMF and mixed with disulfide PN 40113 (5 mg, 20 μmol) in DMF. Diisopropylethylamine (5 μL) is added to the mixture. The mixture is purified on reverse phase HPLC using a gradient of 20%→50% acetonitrile vs. 0.1 M TEAA over 16 minutes. Two dye-colored fractions are obtained at 8.8 min and 9.5 min. The fraction at 9.5 min is identified by mass spectrometry to be the desired product: m/z calculated for $C_{90}H_{111}N_{15}O_{32}S_4^{2-}$, $[M-H]^{2-}$, 1020.84; found: 1021.1.

**[00340]** *Preparation of PN 40147.* PN 40146 is suspended in 100 μL DMF in a 1.5 mL eppendorf tube. Pyridine (20 μL) and pentafluorophenyl trifluoroacetate (20 μL) are added to the DMF solution and heated to 50 °C for five minutes. A portion (1 μL) of the reaction mixture is precipitated into 0.4% HCl; the aqueous solution remains colorless, indicating complete conversion to the active, pentafluorophenyl ester. The remainder of the reaction is precipitated into the dilute acidic solution and the aqueous solution pipetted off. The residue is washed with hexane and dried to a highly colored solid (PN 40147)



Date Reçue/Date Received 2023-12-22

**[00341]** *Preparation of PN 40150.* PN 40147 is dissolved in 50 μL DMF in a 1.5 mL eppendorf tube. A solution of 0.5 μmol 7-deaza-7-propargylamino-2'-deoxyguanosine-5'-triphosphate in 50 μL 1 M bicarbonate is prepared and added to the tube. After remaining overnight at 4 °C the product is purified on HPLC; the fraction at 12 min using a 20%→50% acetonitrile vs. 0.1 M TEAA gradient over 16 minutes contains the desired product: m/z calculated for $C_{104}H_{129}N_{20}O_{44}P_3S_4^{2-}$, $[M-H]^{2-}$, 1291.33; found: 1292.4.

**Example 7: Evaluation of dye-labeled nucleotides**

**[00342]** A bead-based assay is used to evaluate dye-labeled nucleotides of Example 5. A streptavidin bead is prepared with a 5'-biotinylated template strand annealed to a primer strand. The primer strand is designed so that the next cognate base incorporated by a DNA polymerase is a thymidine. A DNA polymerase is bound to the bead complex. Various mixtures containing different ratios of the dye-labeled nucleotide (dUTP*) and the natural base (TTP) is then presented to the beads. After washing away excess reagent, the fluorescence of the beads is read on a flow cytometer using the PE channel (excitation=488 nm, emission=580 nm). A schematic of this assay is shown in **FIG. 9**.

**[00343]** The results of the bead assay for different labeled dUTPs is shown in **FIG. 10**. The total concentration of the sum of the nucleotides is maintained at 2 μM; a labeling fraction of 10% means 0.2 μM of dUTP* and 1.8 μM of TTP. The behavior for the two nucleotides is noticeably different: U#-E has a "tolerance" of about one, meaning that there is no difference in incorporation of the dye-labeled vs the natural nucleotide over all the ratios tested; i.e., a 50% labeling fraction results in 50% of the beads getting labeled. U*-E, on the other hand, has a negative tolerance, meaning that at every ratio it falls below the line drawn between zero and the signal at 100% labeled. A negative tolerance suggests that the dye-label makes the nucleotide a worse substrate than the natural substrate. This result is consistent with the observation that negatively charged dyes such as Atto532 (the dye denoted by U*-E) inhibit incorporation by many polymerases while dyes such as 5-carboxyrhodamine-6G (the dye denoted by U#-E) are zwitterionic and are known to be good substrates.

**[00344]** Additional labeled nuclotides are evaluated using a similar assay. **FIG. 11** shows the result of the bead assay for labeled dATPs. **FIG. 12** shows the result of the bead assay for labeled dGTPs. For labeled dATPs, very low fluorescence is observed at 100% labeling for A*-B compared to A*-B-H and A*-E-H. This indicates that the hydroxyproline linker (H) relieves quenching of the dye by the nucleotide. A similar result is observed for labeled dGTPs. This result is expected for labeled dGTP, as G quenching via photoinduced electron transfer is well



Date Reçue/Date Received 2023-12-22

known. A quenching effect from the disulfide linker, B, may also contribute to the lower fluorescence observed for labeled dATPs and dGTPs.

**Example 8: Sequencing using dye-labeled nucleotides**

**[00345]** A nucleic acid sequencing assay may be used to evaluate dye-labeled nucleotides (e.g., as described herein). An example procedure is shown in **FIG. 18**.

**[00346]** Sequencing may be performed using an instrument outfitted with a light emitting device (LED) and/or a laser. Each nucleotide evaluated may include a dye that is configured for excitement and emission over similar wavelengths (e.g., all red or all green emission). One or more different nucleotide types may be coupled to different dyes. Sequencing performance may be evaluated based on base calling quality, phase lag, phase lead, and homopolymer completion.

**[00347]** Beads with amplified templates are primed, immobilized on a support, and incubated with a tight-binding DNA polymerase. Beads are then subjected to multiple cycles of sequencing. Each sequencing cycle may comprise incubation with U*/T (a fixed ratio of dye-labeled and natural TTP), a "chase" process (TTP alone), imaging, and a cleavage process (10 mM tris(hydroxypropyl)phosphine (THP)) to release the dye. Each process may have a wash process in between. This process may be repeated for A, C, and G-including nucleotides or nucleotide analogs. This sequencing procedure may effectively identify homopolymeric regions of at least 2, 3, 4, 5, 6, 7, 8, or more nucleotides.

**[00348]** Sequencing is also evaluated for an all hyp-linker set in which dye-labeled nucleotides including each canonical nucleotide include the hyp10 or hyp20 linker. This evaluation is performed to identify a set where higher fractions may be used with minimal quenching. Higher quenching may lead to higher scarring (e.g., as described herein), which may reduce incorporation efficiency by a polymerase enzyme. However, family B enzymes such as PolD may perform well with scars. Sequencing may be evaluated with 2.5% and 20% labeling fractions with a dye such as Atto633.

**[00349]** Sequencing may be used to evaluate the tolerance for various labeled nucleotides. **FIG. 19** shows normalized bead data for nucleotides labeled with a red-emitting dye. Bright solution fraction ($b_f$) is plotted against bright incorporation fraction ($b_i$). The curves are fitted to the following equation:

$$b_i = \frac{tol(b_f/d_f)}{1 + tol(b_f/d_f)}$$

in which $d_f$ is the dark solution fraction. In **FIG. 19**, the calculated tolerances are 10.6 for G*, 2.8 for A*, 2.0 for U*, and 1.2 for C*. The positive tolerance numbers indicate that at 50%

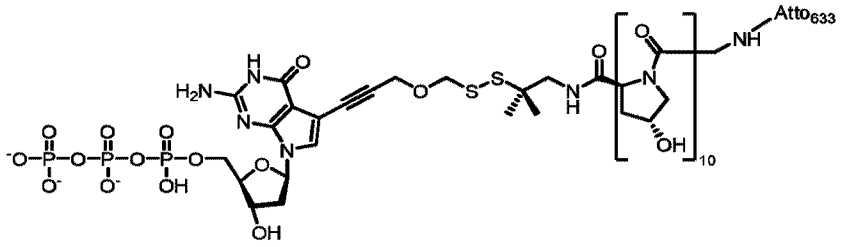


Date Reçue/Date Received 2023-12-22

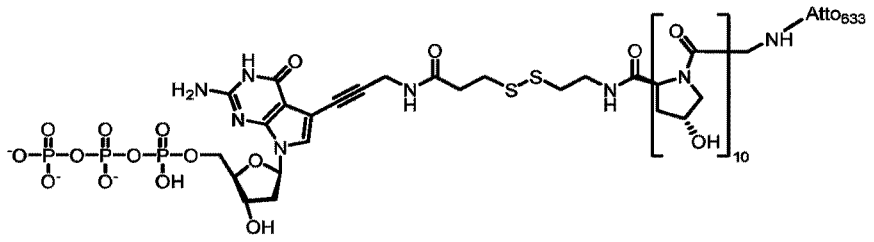
labeling fraction, more than 50% is labeled. Reagents with a tolerance of 1 may have the least "context" in sequencing. Reagents with a very negative tolerance (e.g., tolerance << 1) may have issues with uniform incorporation across a plurality of templates coupled to a support because they must be used at such low concentrations that they may fall below saturation and be consumed at an uneven rate.

**Example 9: Dye-labeled nucleotides including guanine or analogs thereof**

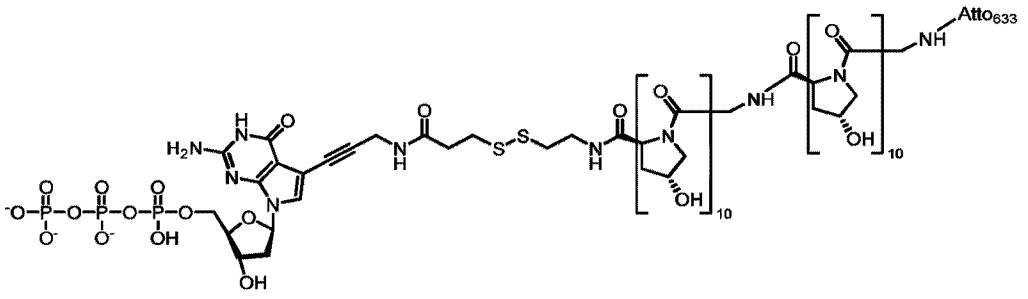
**[00350]** Nucleotides including guanine or analogs thereof may perform more poorly in sequencing applications (e.g., as described herein) in base-calling accuracy. This may be related to photoinduced electron transfer from the nucleobase to a dye linked to the nucleobase, which may quench signal emitted by the dye and thus less dynamic range of signal. Accordingly, various dye-labeled nucleotides including guanine or analogs thereof are prepared and evaluated as provided herein. Examples of such dye-labeled nucleotides include:




**G1**



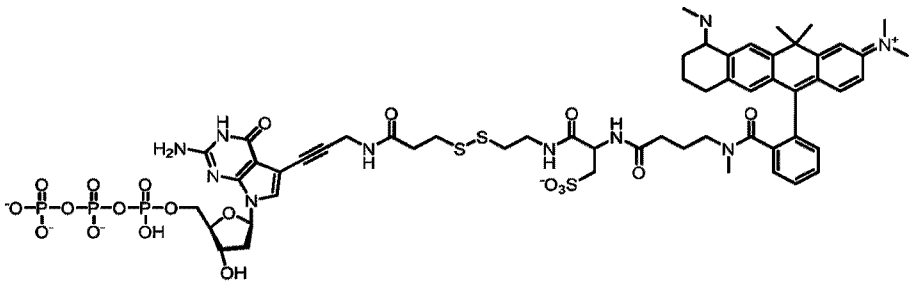

**G2**



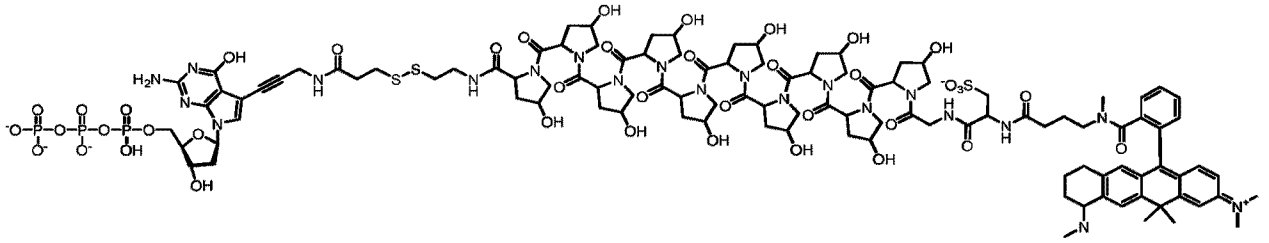

**G3**



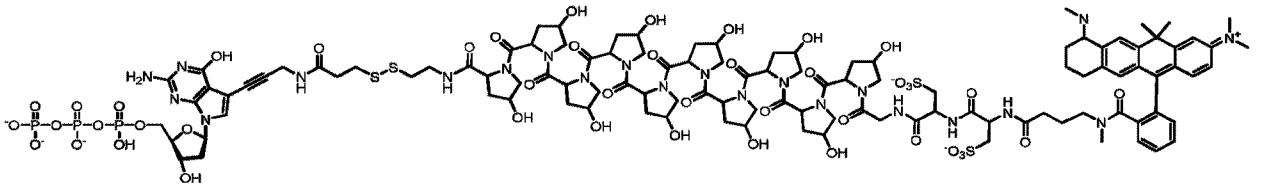
Date Reçue/Date Received 2023-12-22



**G4**



**G5 (Hyp10 linker, Cya dye)**



**G6 (Hyp10 linker, Cya2 dye)**

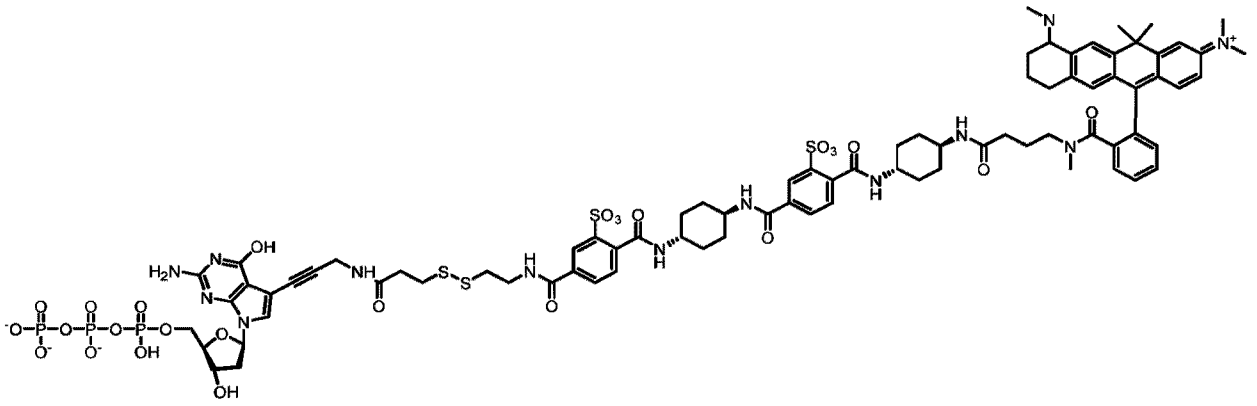
**[00351]** Several of the structures shown above include the hyp10 linker which includes the sequence Gly-Hyp-Hyp-Hyp-Hyp-Hyp-Hyp-Hyp-Hyp-Hyp-Hyp from the N-terminal end. G4, which lacked the hyp10 linker, is highly quenched. The remaining dye-labeled nucleotides are evaluated in a sequencing assay, as described herein. Of the structures shown, G6 provides the highest accuracy. A synthetic route for preparation of G6 is shown in **FIGs. 13A-13C**.

**Example 10: Preparation of dye-labeled nucleotides**

**[00352]** A dye-labeled nucleotide may include one or more amino acids. As described above, diamines and diacids may be used to construct amino acids. A dye-labeled nucleotide may include two or more of a given amino acid as a repeating unit. An example of a dye-labeled nucleotide including two repeating units of an amino acid is shown below:



Date Reçue/Date Received 2023-12-22



**[00353]** **FIGs. 14A** and **14B** show a synthetic route for preparation of the dye-labeled nucleotide shown above. The composition of each intermediate is confirmed by mass spectrometry. The dye-labeled nucleotide is evaluated in a bead assay, as described in Example 7. The linker provides a G* that is less bright than G*s with polyhydroxyproline linkers, but is more efficient in reducing quenching than a G* without a linker.

**Example 11: Evaluation of quenching**

**[00354]** The dye-labeled nucleotides provided herein may improve quenching between nucleobases and the dyes to which they are attached and/or between dyes in a nucleic acid molecule (e.g., a growing nucleic acid strand), such as in a homopolymeric region of a nucleic acid molecule. Quenching may be evaluated in an enzyme-independent manner.

**[00355]** **FIG. 15** shows a schematic for evaluating quenching. Synthetic oligos are constructed with one or two "linker arm nucleotides". Linker arm nucleotides are thymidine analogs with a linker arm containing a primary amine. The oligonucleotide containing the linker arm nucleotide can be labeled with linkers and dyes and HPLC purified. The advantage of using the bead-labeled assay is that exact quantitation of the reagents is not necessary; a large excess can be used in each step and the beads washed, ensuring that only stoichiometric amounts of oligonucleotides are bound to the template. Each dye-linker is put on both oligonucleotides. The beads are measured on the flow cytometer in the APC (red) channel. The percent quenching is determined by the formula: $\%\ quenching = 100 \times (1 - Fl_{bis}/(2 * Fl_{mono}))$.

**[00356]** **FIGs. 16** and **17** show quenching results for red dye linkers (**FIG. 16**) and green dye linkers (**FIG. 17**). The results show that the nature of the dye affects quenching. Negative charge (see Atto532 vs AttoRho6G) can improve quenching but if the dye is extremely large and flat (see Cy5, Alexa 647) quenching may not be improved. The hyp10 or hyp20 linkers improve quenching. As shown in **FIG. 16**, hyp10 improves quenching with Atto633, and cyanine dyes quench even with four sulfonic acid groups. As shown in **FIG. 17**, sulfonic acid groups on



Date Reçue/Date Received 2023-12-22

Atto532 improve quenching, and the combination of Atto532 and hyp10 also improves quenching.

**Example 12: Interrogation of homopolymers**

**[00357]** A nucleic acid template is provided that has various lengths of a homopolymer region comprising cytosines (1C, 2C, 3C, 4C, 5C). The template is contacted with guanosine-containing nucleotides labeled with Atto532 fluorophore (e.g., as described herein; denoted herein as G*). The labeled nucleotides may be provided in a solution as a nucleotide flow (e.g., as described herein). The nucleotide flow may include 100% labeled nucleotides (e.g., the nucleotide flow may include only labeled nucleotides and no unlabeled nucleotides) or may include both labeled and unlabeled nucleotides (e.g., as described herein). The labeled and, where present, unlabeled nucleotides may not be terminated so that multiple nucleotides can be incorporated into as many positions in succession as there appear cytosines in the template. An enzyme (e.g., a polymerase enzyme, such as Bst 3.0) may be used to incorporate labeled and/or unlabeled nucleotides into an extended primer using the nucleic acid having a polycytosine sequence as a template. A plurality of copies of the template may be immobilized to a bead or other support (e.g., as described herein). This procedure is schematically illustrated in **FIGs. 20A** and **20B**.

**[00358]** In some cases, the labeled nucleotide incorporates into as many positions in succession as there appear cytosine in the template. In other cases, less than all potential G* are incorporated. Where unlabeled nucleotides are included in the nucleotide flow, both unlabeled and labeled nucleotides may be incorporated. For example, for a template including a homopolymeric region including three cytosines, the incorporated nucleotides may have the sequence GGG, GG*G, GGG*, G*GG, G*G*G, G*GG*, GG*G*, or G*G*G*, where G* indicates a labeled nucleotide and G indicates an unlabeled nucleotide. The sequence of the incorporated nucleotides may vary based on, for example, the labeling fraction of the nucleotide flow (e.g., the ratio of labeled to unlabeled nucleotides in the flow) and the optical (e.g., fluorescent) labeling reagent used to label the nucleotides.

**[00359]** Labeled polynucleotide products are separated on a Biorad denaturing acrylamide gel and imaged using blue and green LEDs to detect incorporated labeled nucleotides. As shown in **FIG. 20C**, 1, 2, 3, 4, and 5 consecutive cytosines can be detected using this method.

**Example 13: Sequencing by synthesis using a high fraction of labeled nucleotides**



Date Reçue/Date Received 2023-12-22

**[00360]** A template nucleic acid having a length of at least 30 nucleotides is sequenced using the procedures and labeled nucleotides described herein. The template to be sequenced may be immobilized to a support (e.g., as described herein). The template is subjected to a sequencing by synthesis reaction, in which the template is sequentially contacted with solutions (e.g., nucleotide flows) comprising PolD polymerase (New England Biolabs) and a plurality of nucleotides of a single canonical type (e.g., T, A, C, or G). In each nucleotide flow, approximately 20% of the nucleotide population is labeled with Atto633 as described herein above to provide a labeling fraction of about 20%. The remaining nucleotides are unlabeled. Nucleotides included in nucleotide flows are not terminated to allow efficient sequencing of homopolymeric regions of the template. After contacting the template with a first nucleotide flow including nucleotides of a first canonical type, the template is contacted with a wash flow to remove unincorporated nucleotides. A fluorescent image is collected. The linker of the fluorescent labeling reagent associated with incorporated labeled nucleotides is contacted with a cleavage flow comprising a cleavage reagent configured to cleave a cleavable group of the linker to separate the fluorescent dye (e.g., Atto633) of the fluorescent labeling reagent from the incorporated nucleotide. An additional wash flow may be used to remove the cleavage flow. In some cases, a chase flow including unlabeled nucleotides of the first canonical type may follow the initial nucleotide flow and precede or follow the imaging process. The process is repeated for the second, third, and fourth nucleotide types in succession, and then the entire cycle is repeated.

**[00361]** **FIG. 21A** shows the results of application of this method to a sample template. A black circle indicates that a nucleotide was incorporated and a gray circle indicates that no nucleotide was incorporated in a particular flow cycle. As shown in the figure, the incorporation of one or more nucleotides in a flow cycle can be determined with a high degree of accuracy. Furthermore, as is shown in **FIG. 21B**, the relationship between signal intensity and labeled nucleotide homopolymer length may be substantially linear across a plurality of templates (e.g., as described herein). For example, the signal intensity may be proportional to the length of a homopolymeric region of the template. This proportionality indicates that quenching effects have been substantially overcome. In **FIG. 21B**, the slope for G is 0.96, for C is 0.80, for A is 079, and for T is 0.70. The dotted line indicates the actual signal, while the solid line indicates the signal after correction for phasing.



Date Reçue/Date Received 2023-12-22

**Example 14: Sequencing by synthesis using 100% labeled nucleotides**

**[00362]** A template nucleic acid having a length of at least 30 nucleotides is sequenced as described in Example 13, but with solutions in which 100% of the nucleotides are labeled. In **FIG 22**, black circles indicate that a base was incorporated in a given flow cycle, while gray circles indicate that a base was not incorporated in a given flow cycle. As can be seen from **FIG. 22**, the sequencing method can be used to detect base incorporation through 50 flow cycles.

**Example 15: Labeled proteins**

**[00363]** A protein is labeled with a plurality of optical (e.g., fluorescent) labeling reagents (e.g., as described herein). For example, the protein may be labeled with three or more optical labeling reagents. The optical labeling reagents associated with the protein may all comprise a fluorescent dye of the same type. The optical labeling reagents associated with the protein may all comprise a linker of the same type. The protein may be an antibody, such as a monoclonal antibody.

**[00364]** The protein is used to label a cell. The cell may be a component of sample, which sample may comprise a plurality of cells. The cells of the sample may be analyzed and sorted using flow cytometry. Flow cytometric analysis may identify the cell as being labeled with the protein associated with the plurality of optical labeling reagents. In some cases, a plurality of cells of a sample may be labeled with optical labeling reagents (e.g., as described herein). For example, cells comprising a particular cell surface feature (e.g., an antigen) configured to associate with a protein (e.g., a protein labeled with a plurality of optical labeling reagents, such as an antibody labeled with a plurality of optical labeling reagents) may be labeled with labeled proteins and analyzed and/or sorted using flow cytometry. Analyzed and/or sorted cells may be subjected to further downstream analysis and processing, including, for example, nucleic acid sequencing, staining, imaging, function assays, immunoassays, isolation/expansion, additional labeling, immunoprecipitation, etc.

**[00365]** While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it



Date Reçue/Date Received 2023-12-22

shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents.



Dat

## CLAIMS

WHAT IS CLAIMED IS:

1. A method, comprising:
   (a) contacting a nucleic acid molecule with a solution comprising a plurality of non-terminated nucleotides under conditions sufficient to incorporate, in succession, a first nucleotide and a second nucleotide of said plurality of non-terminated nucleotides into a growing strand that is complementary to said nucleic acid molecule, wherein said first nucleotide is labeled, and wherein at least about 20% of said plurality of non-terminated nucleotides are labeled nucleotides;
   (b) detecting one or more signals or signal changes from said first nucleotide, wherein said one or more signals or signal changes are indicative of incorporation of said first nucleotide; and
   (c) resolving said one or more signals or signal changes to determine a sequence of said nucleic acid molecule.
2. The method of claim 1, wherein said plurality of non-terminated nucleotides comprises nucleotides of a same canonical base type.
3. The method of claim 1, wherein said first nucleotide comprises a fluorescent dye.
4. The method of claim 3, wherein said fluorescent dye is cleavable.
5. The method of claim 4, further comprising:
   (i) cleaving said fluorescent dye;
   (ii) contacting said nucleic acid molecule with a second solution comprising a second plurality of non-terminated nucleotides under conditions sufficient to incorporate a third nucleotide of said second plurality of non-terminated nucleotides into said growing strand, wherein at least about 20% of said second plurality of non-terminated nucleotides are labeled nucleotides, wherein said third nucleotide is a labeled nucleotide;
   (iii) detecting one or more second signals or signal changes from said third nucleotide; and
   (iv) resolving said one or more second signals or signal changes to determine a second sequence of said nucleic acid molecule.



Date Reçue/Date Received 2023-12-22

6. The method of claim 5, wherein said plurality of non-terminated nucleotides and said second plurality of non-terminated nucleotides comprise nucleotides of different canonical base types.

7. The method of claim 5, wherein said third nucleotide comprises an additional fluorescent dye of a same type as said fluorescent dye.

8. The method of claim 1, further comprising:

(i) contacting said nucleic acid molecule with a second solution comprising a second plurality of non-terminated nucleotides under conditions sufficient to incorporate a third nucleotide of said second plurality of non-terminated nucleotides into said growing strand, wherein at least about 20% of said second plurality of non-terminated nucleotides are labeled nucleotides, wherein said third nucleotide is a labeled nucleotide;

(ii) detecting one or more second signals or signal changes from said third nucleotide; and

(iii) resolving said one or more second signals or signal changes to determine a second sequence of said nucleic acid molecule.

9. The method of claim 8, wherein said plurality of non-terminated nucleotides and said second plurality of non-terminated nucleotides comprise nucleotides of different canonical base types.

10. The method of claim 8, wherein said plurality of non-terminated nucleotides and said second plurality of non-terminated nucleotides comprise nucleotides of a same canonical base type.

11. The method of claim 8, wherein said third nucleotide comprises a fluorescent dye.

12. The method of claim 11, wherein said contacting in (i) is performed in absence of cleaving a fluorescent dye from said first nucleotide.

13. The method of claim 11, further comprising repeating (i)-(iii) at least 5 times, each with a different solution of non-terminated nucleotides that comprises at least 20% labeled nucleotides, in absence of cleaving a fluorescent dye from said first nucleotide.

14. The method of claim 1, wherein at least about 50%, 70%, 80%, 90%, 95%, or 99% of said plurality of non-terminated nucleotides are labeled nucleotides.



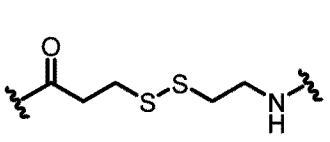
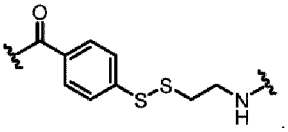
Date Reçue/Date Received 2023-12-22

15. The method of claim 1, wherein substantially all of said plurality of non-terminated nucleotides are labeled nucleotides.
16. The method of claim 1, wherein said resolving in (c) comprises determining a number of consecutive nucleotides from said solution incorporated into said growing strand.
17. The method of claim 16, wherein said number is selected from the group consisting of 2, 3, 4, 5, 6, 7, or 8 nucleotides.
18. The method of claim 16, wherein said resolving in (c) comprises processing a tolerance of said solution, wherein said tolerance comprises a comparison of a ratio of incorporated labeled nucleotides to incorporated unlabeled nucleotides in said solution.
19. The method of claim 1, wherein said second nucleotide is unlabeled.
20. The method of claim 1, wherein said second nucleotide is labeled.
21. The method of claim 1, wherein said first nucleotide and said second nucleotide are labeled.
22. The method of claim 21, wherein (b) comprises detecting one or more signals or signal changes from said first nucleotide and said second nucleotide.
23. The method of claim 22, wherein (c) comprises resolving said one or more signals or signal changes from said first nucleotide and said second nucleotide incorporated into said growing strand.
24. A kit comprising: a plurality of linkers, wherein a linker of said plurality of linkers comprises (i) one or more water soluble groups and (ii) two or more ring systems, wherein said linker is coupled to a nucleotide via a nucleobase of said nucleotide, wherein at least two ring systems of said two or more ring systems are connected to each other by no more than two atoms, and wherein said linker comprises a non-proteinogenic amino acid comprising a ring system of said two or more ring systems.
25. The kit of claim 24, wherein said linker is coupled to a plurality of substrates.
26. The kit of claim 25, wherein said plurality of substrates is selected from the group consisting of a nucleotide, a protein, a lipid, a cell, and an antibody.
27. The kit of claim 24, wherein said linker is connected to a fluorescent dye.
28. The kit of claim 24, wherein said linker comprises a plurality of non-proteinogenic amino acids.
29. The kit of claim 28, wherein said linker comprises a plurality of hydroxyprolines.



Date Reçue/Date Received 2023-12-22

30. The kit of claim 29, wherein said linker comprises at least three hydroxyprolines.
31. The kit of claim 30, wherein said linker comprises at least ten hydroxyprolines.
32. The kit of claim 24, wherein said at least two ring systems of said two or more ring systems are connected to each other by an $sp^2$ carbon atom.
33. The kit of claim 24, wherein said at least two ring systems of said two or more ring systems are connected to each other directly without an intervening atom.
34. The kit of claim 24, wherein said one or more water soluble groups are selected from the group consisting of a pyridinium, an imidazolium, a quaternary ammonium group, a sulfonate, a phosphate, an alcohol, an amine, an imine, a nitrile, an amide, a thiol, a carboxylic acid, a polyether, an aldehyde, a boronic acid, and a boronic ester.
35. The kit of claim 24, wherein said linker further comprises a cleavable group that is configured to be cleaved to separate a first portion of said linker from a second portion of said linker.
36. The kit of claim 35, wherein said cleavable group is selected from the group consisting of an azidomethyl group, a disulfide bond, a hydrocarbyldithiomethyl group, and a 2-nitrobenzyloxy group.
37. The kit of claim 24, wherein said linker comprises a moiety selected from the group consisting of


and

.

38. The kit of claim 24, wherein said plurality of linkers comprises a first linker associated with a first substrate and a second linker associated with a second substrate, wherein said first substrate and said second substrate are different types of substrates.
39. The kit of claim 38, wherein said first linker and said second linker comprise a same chemical structure.
40. The kit of claim 38, wherein said first substrate and said second substrate are nucleotides comprising nucleobases of different types.
41. The kit of claim 38, wherein said plurality of linkers further comprises a third linker associated with a third substrate and a fourth linker associated with a fourth substrate, and



Date Reçue/Date Received 2023-12-22

wherein said first substrate, said second substrate, said third substrate, and said fourth substrate are different types of substrates.

42. The kit of claim 41, wherein said first substrate, said second substrate, said third substrate, and said fourth substrate are nucleotides comprising nucleobases of different types.

43. The kit of claim 41, wherein said first linker and said second linker comprise a same chemical structure, and wherein said first linker and said third linker comprise different chemical structures.



Date Reçue/Date Received 2023-12-22

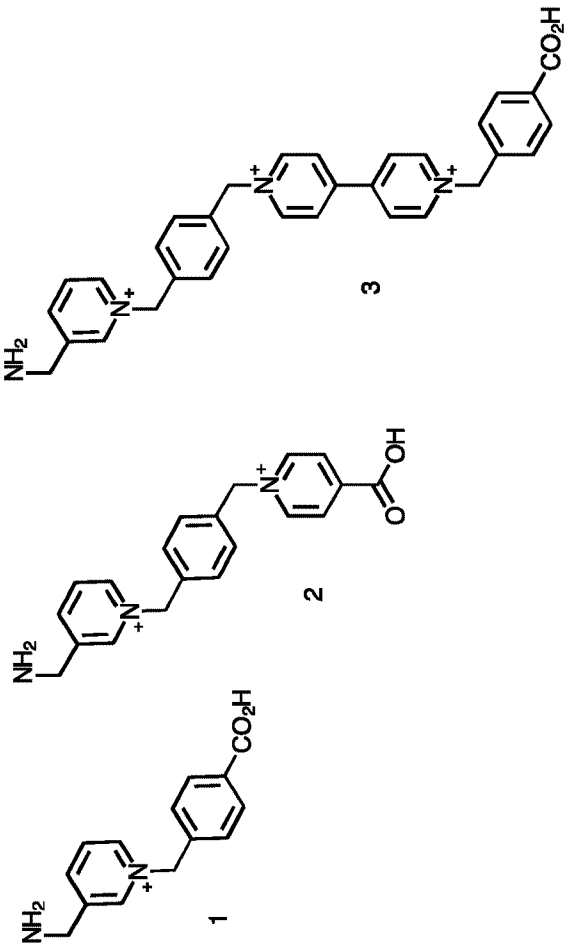



*FIG. 1A*



Date Reçue/Date Received 2023-12-22

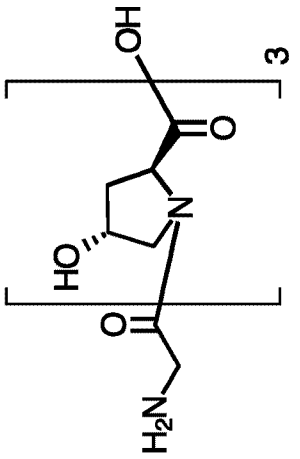



*FIG. 1B*



Date Reçue/Date Received 2023-12-22

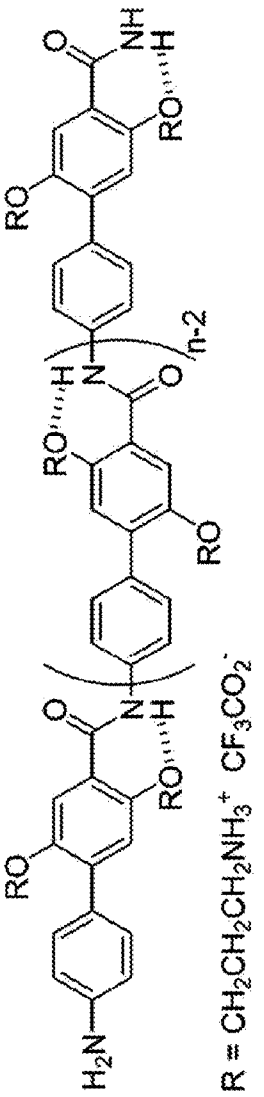



*FIG. 1C*



Date Reçue/Date Received 2023-12-22

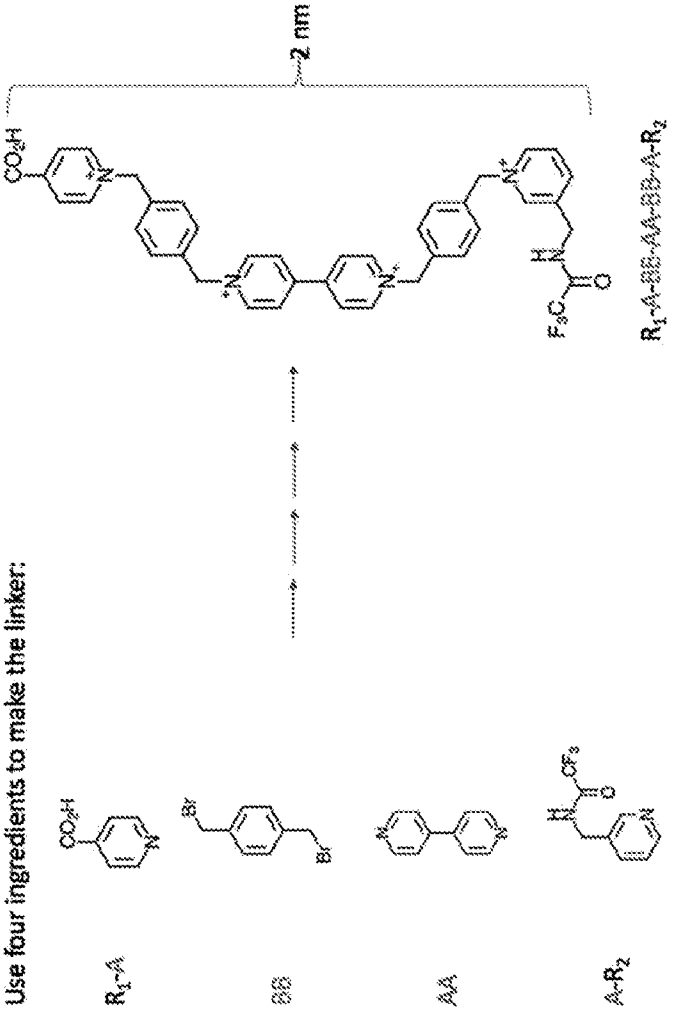



*FIG. 2A*



Date Reçue/Date Received 2023-12-22

*FIG. 2B*



Date Reçue/Date Received 2023-12-22

*FIG. 2C*



Date Reçue/Date Received 2023-12-22

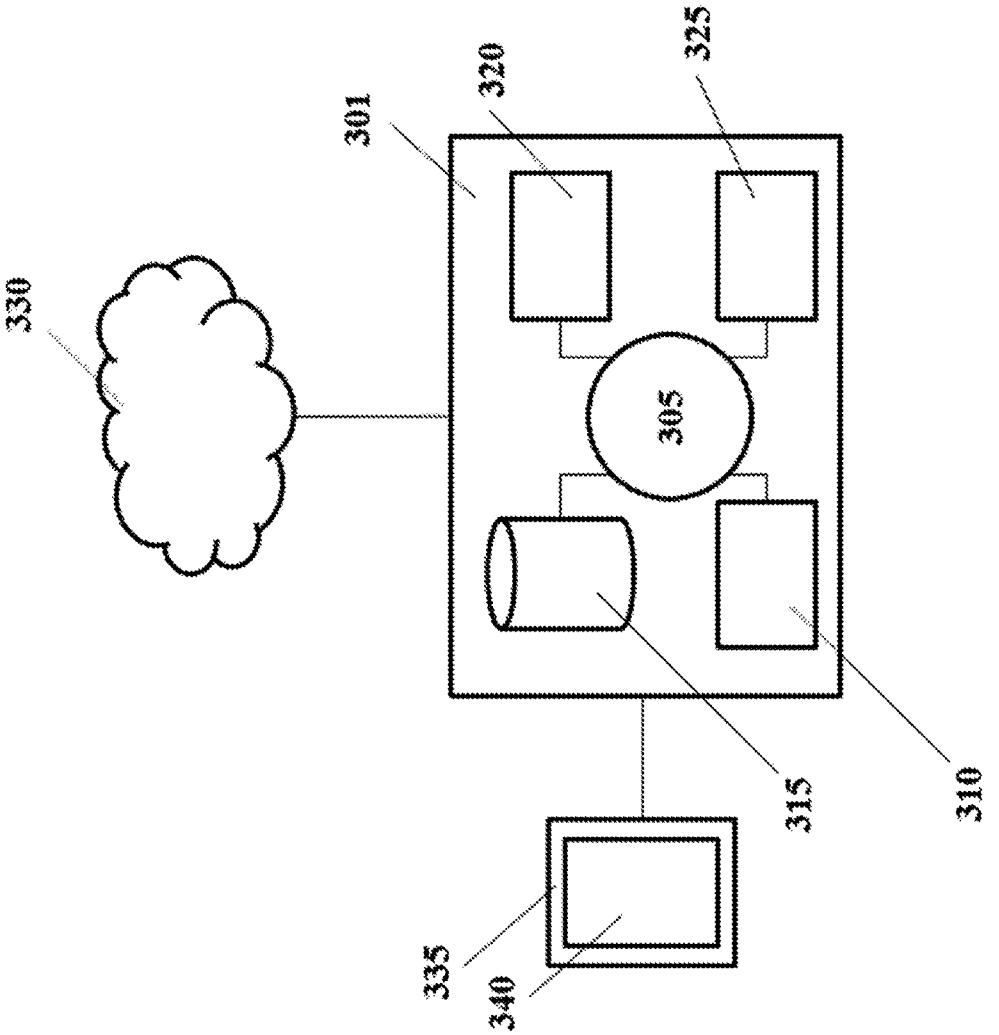



*FIG. 3*



Date Reçue/Date Received 2023-12-22

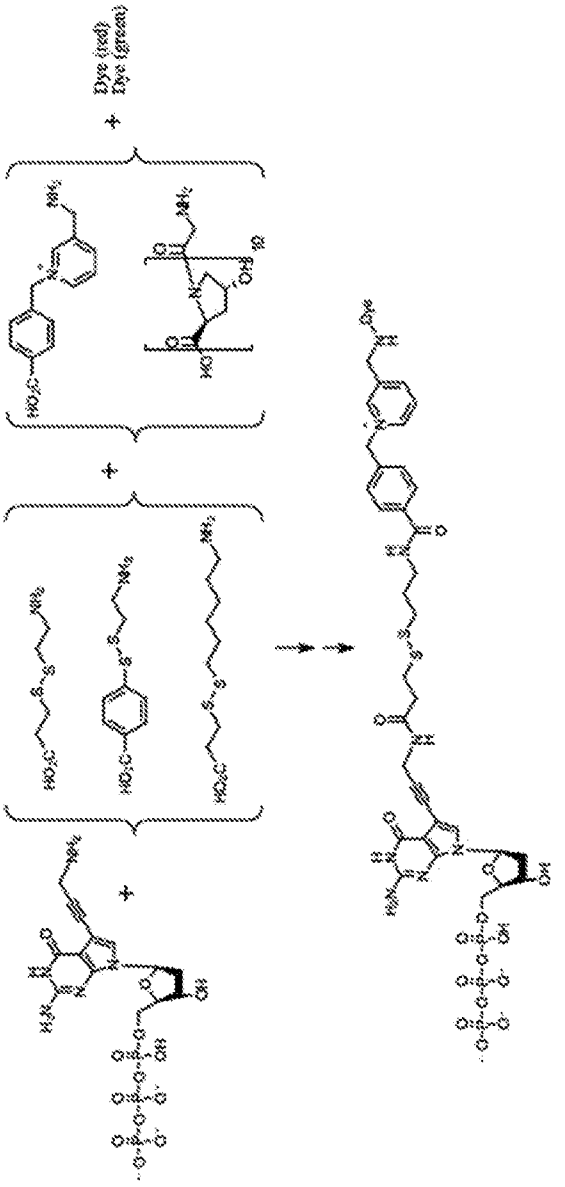



*FIG. 4*



Date Reçue/Date Received 2023-12-22

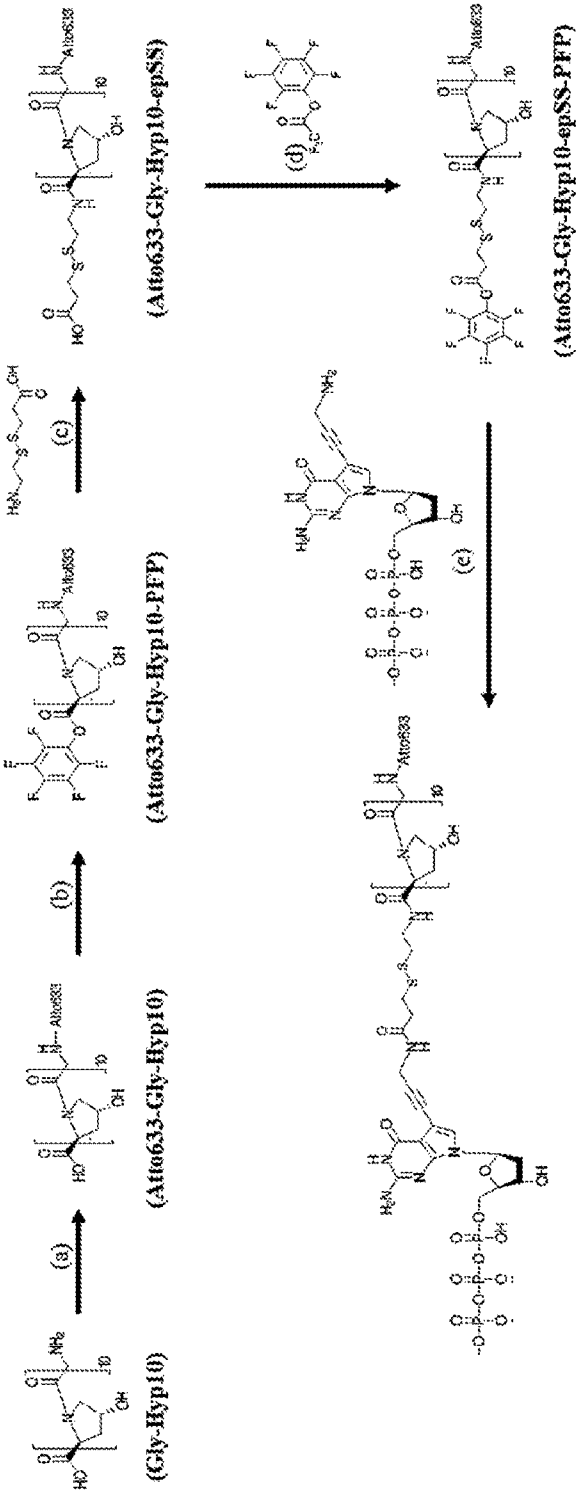



*FIG. 5A*



Date Reçue/Date Received 2023-12-22

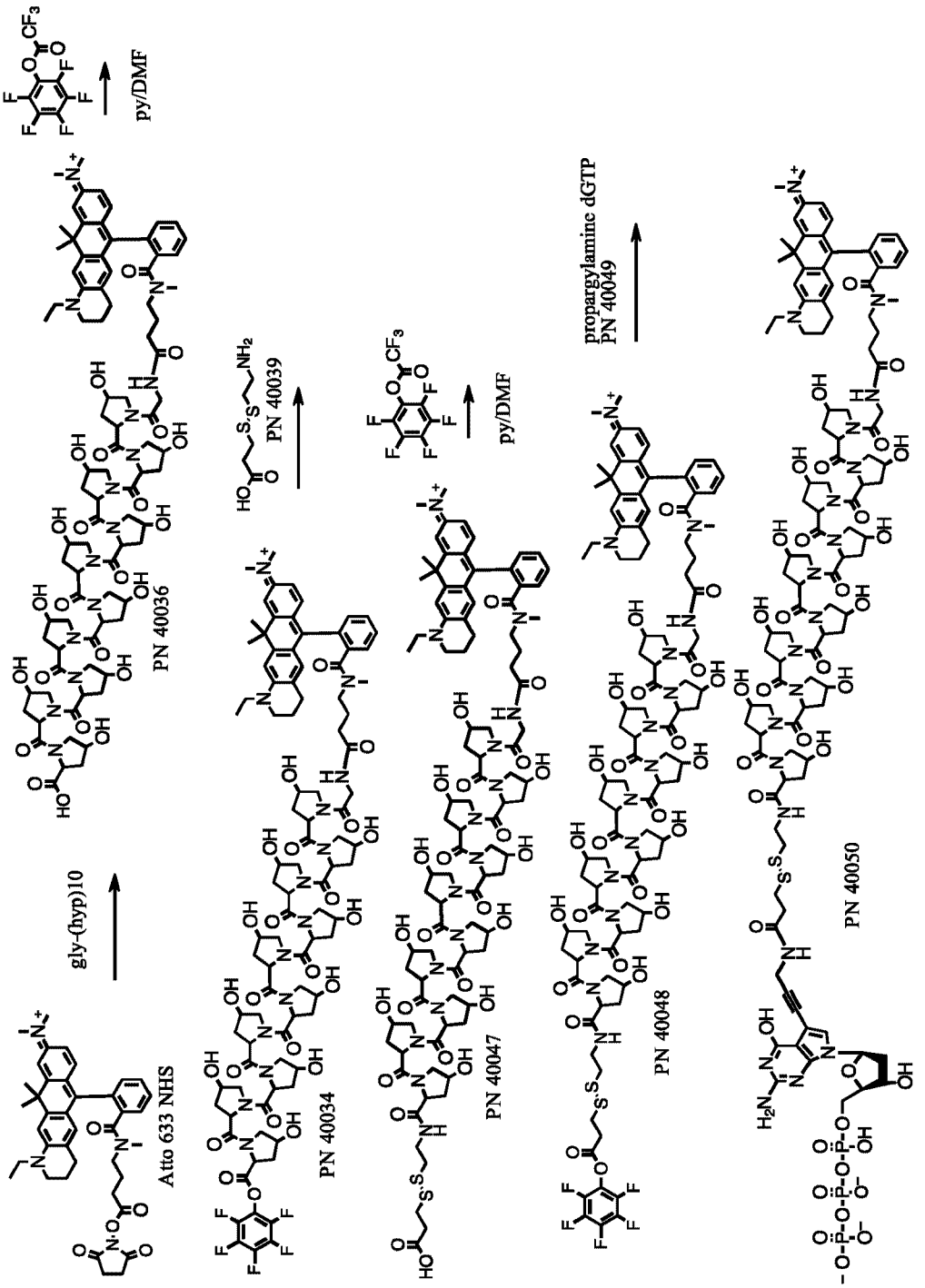



*FIG. 5B*



Date Reçue/Date Received 2023-12-22

*FIG. 6*



Date Reçue/Date Received 2023-12-22

*FIG. 7*



Date Reçue/Date Received 2023-12-22

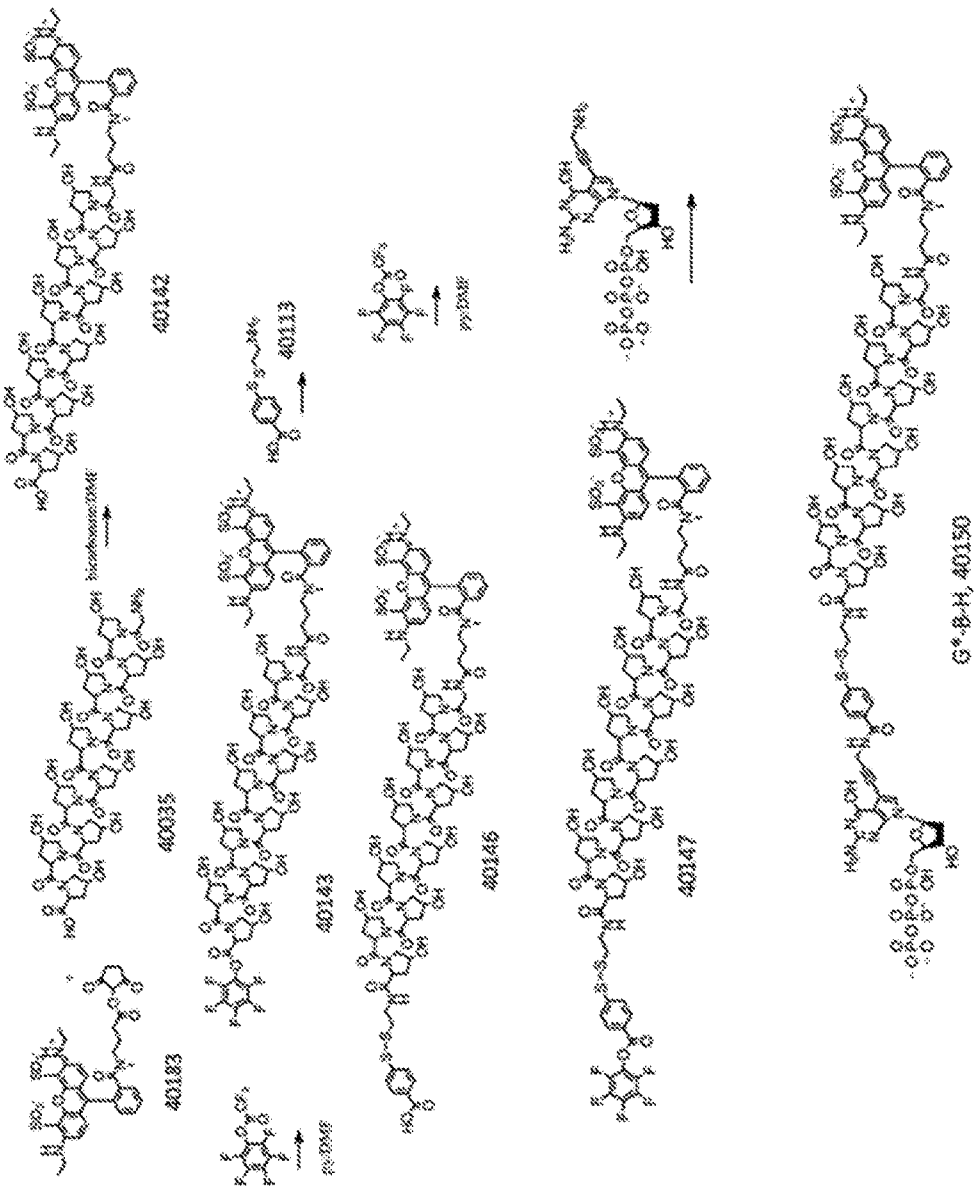



*FIG. 8*



Date Reçue/Date Received 2023-12-22

*FIG. 9*



Date Reçue/Date Received 2023-12-22

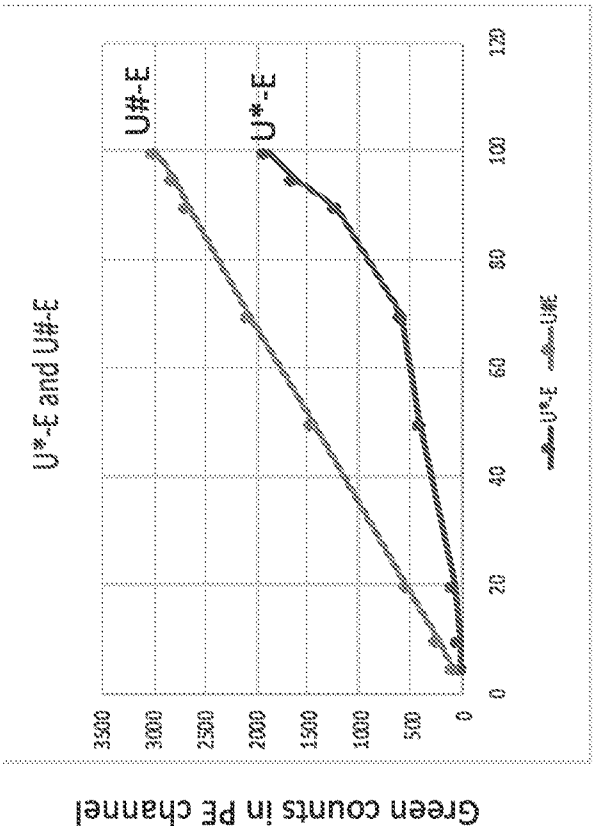



*FIG. 10*



Date Reçue/Date Received 2023-12-22

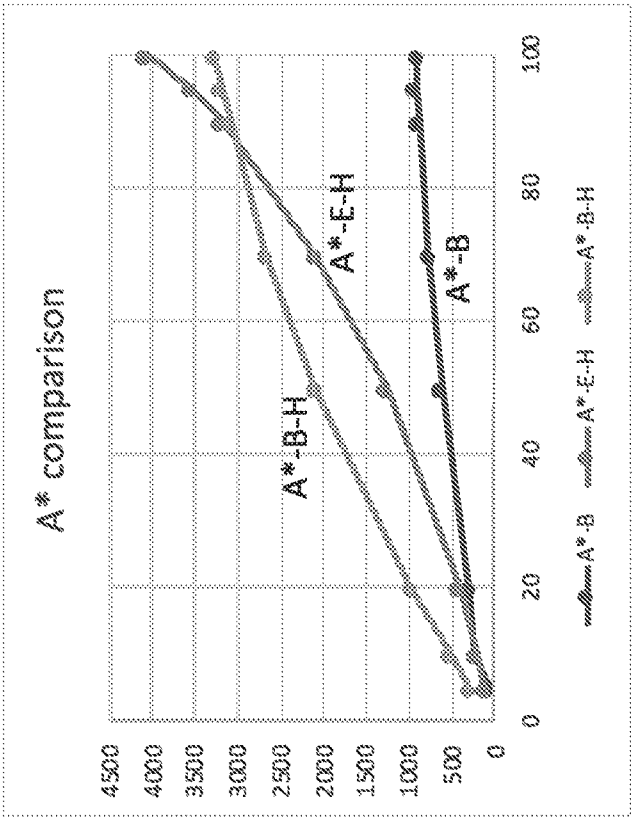



*FIG. 11*



Date Reçue/Date Received 2023-12-22

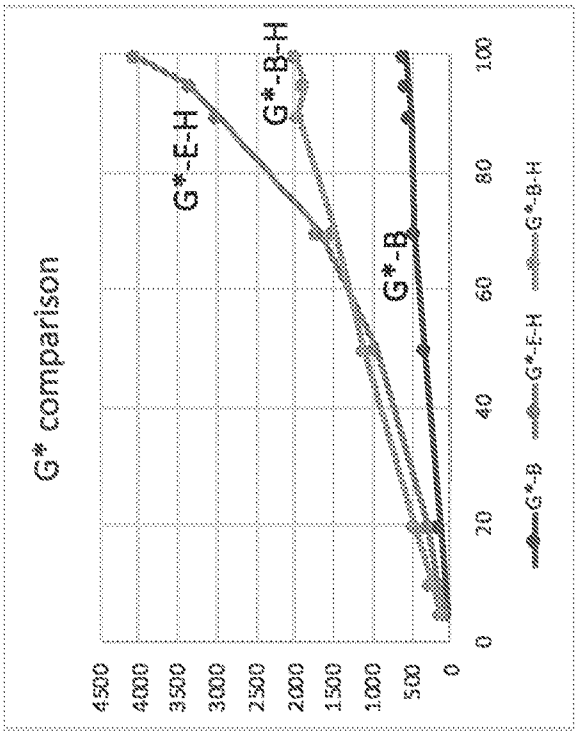



*FIG. 12*



Date Reçue/Date Received 2023-12-22

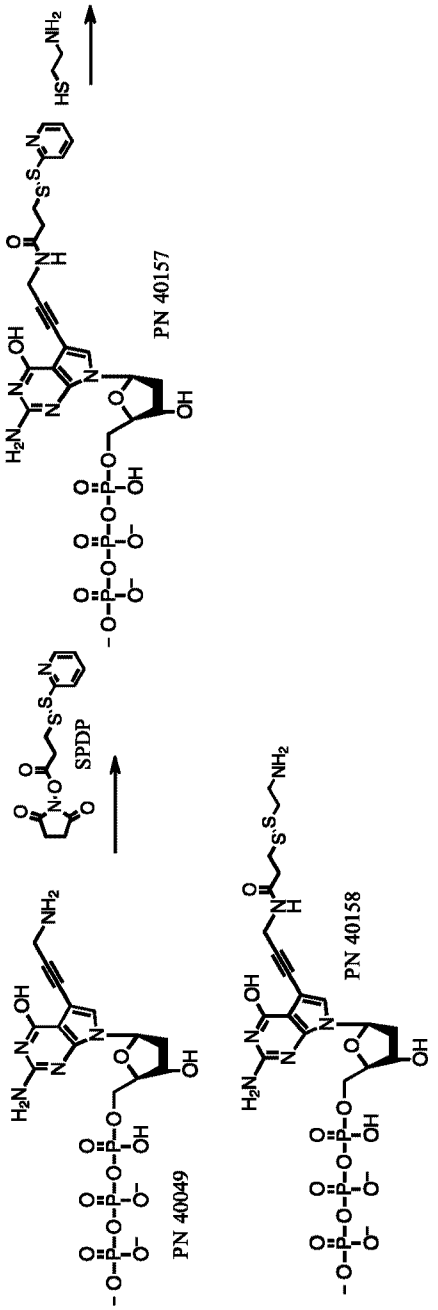



*FIG. 13A*



Date Reçue/Date Received 2023-12-22

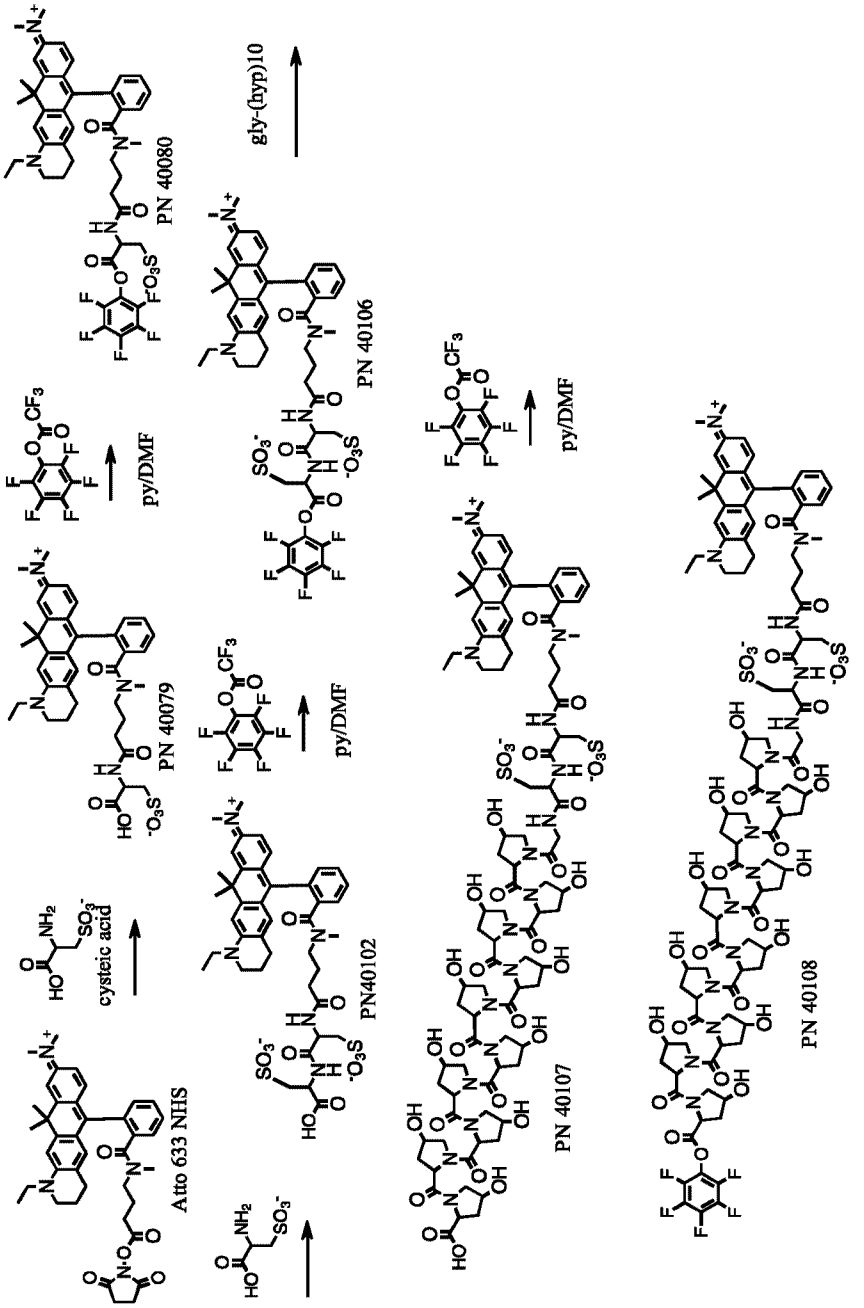



*FIG. 13B*



Date Reçue/Date Received 2023-12-22

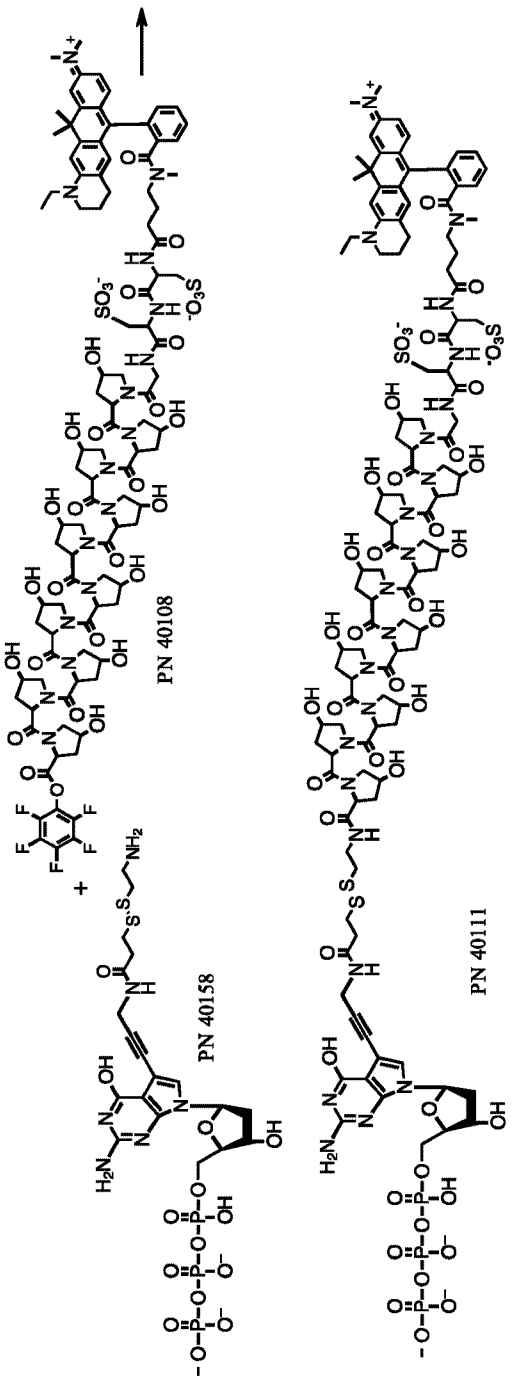



*FIG. 13C*



Date Reçue/Date Received 2023-12-22

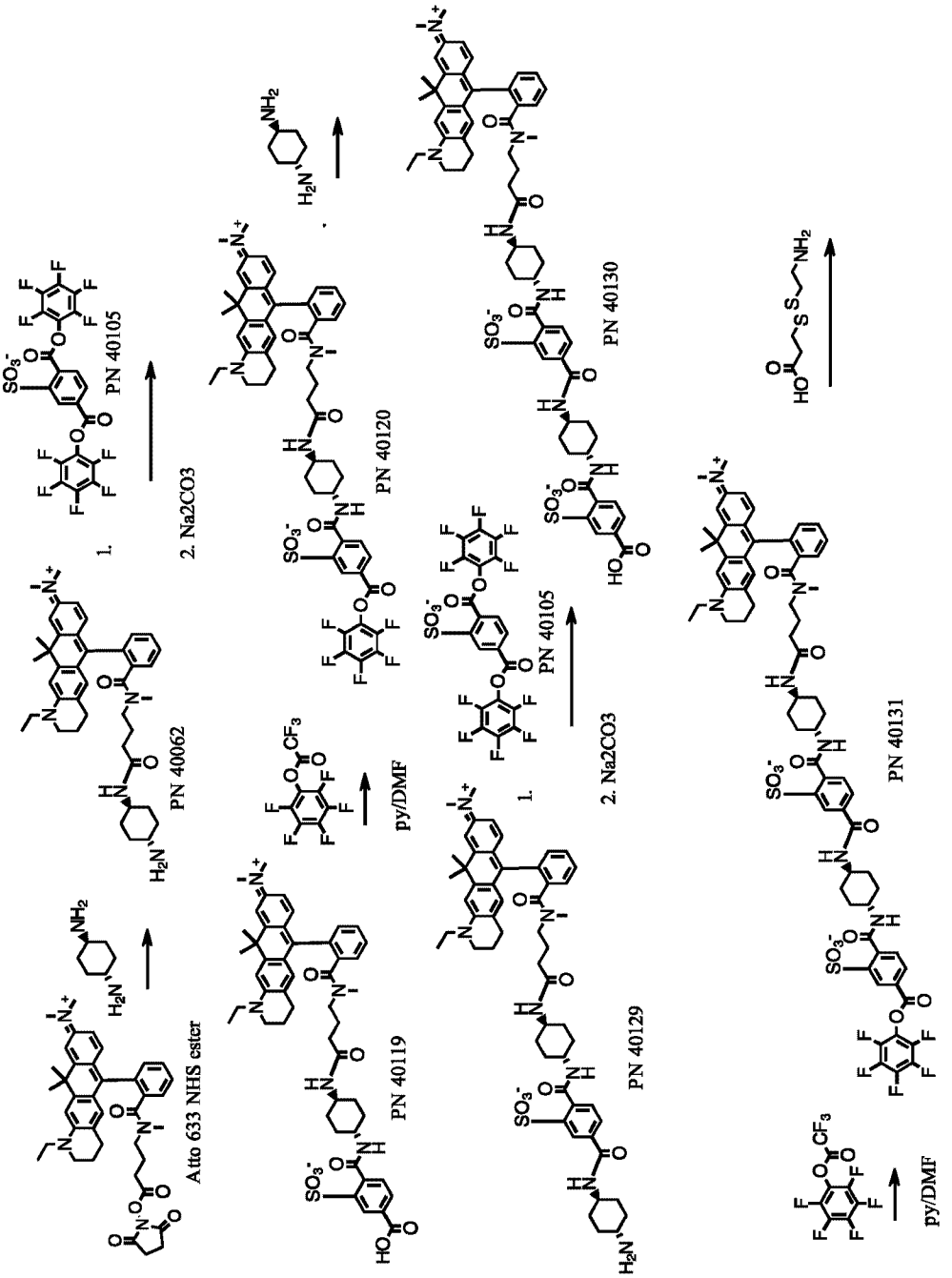



*FIG. 14A*



Date Reçue/Date Received 2023-12-22

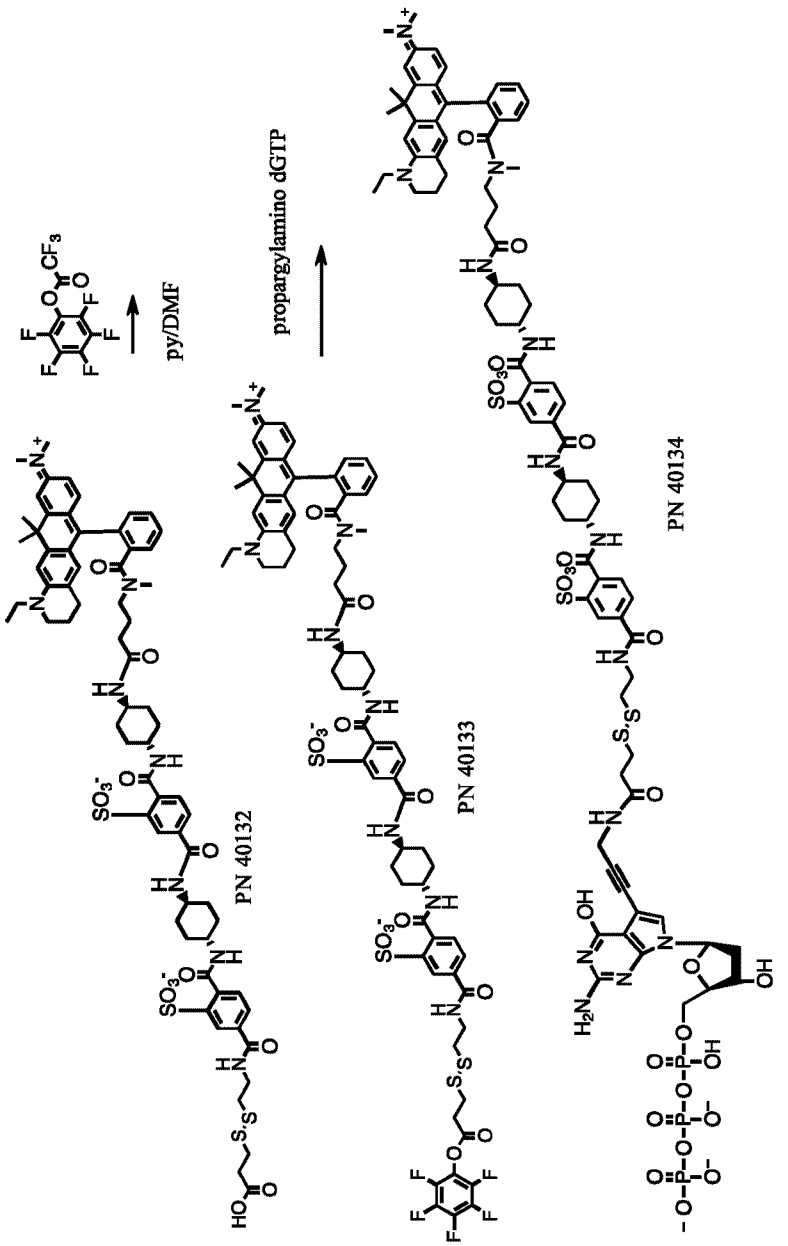



*FIG. 14B*



Date Reçue/Date Received 2023-12-22

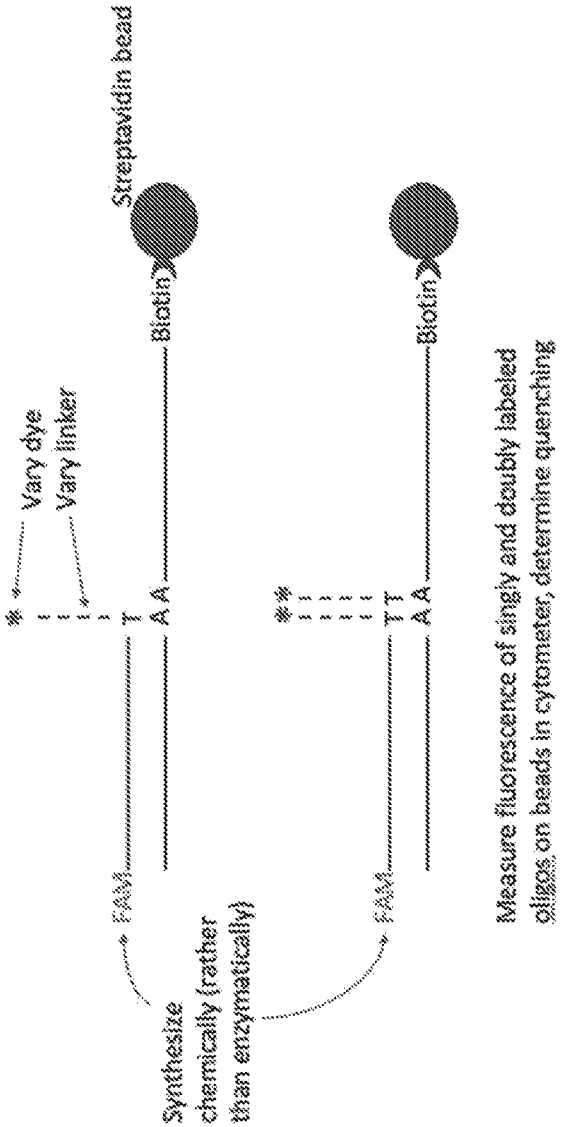



*FIG. 15*



Date Reçue/Date Received 2023-12-22

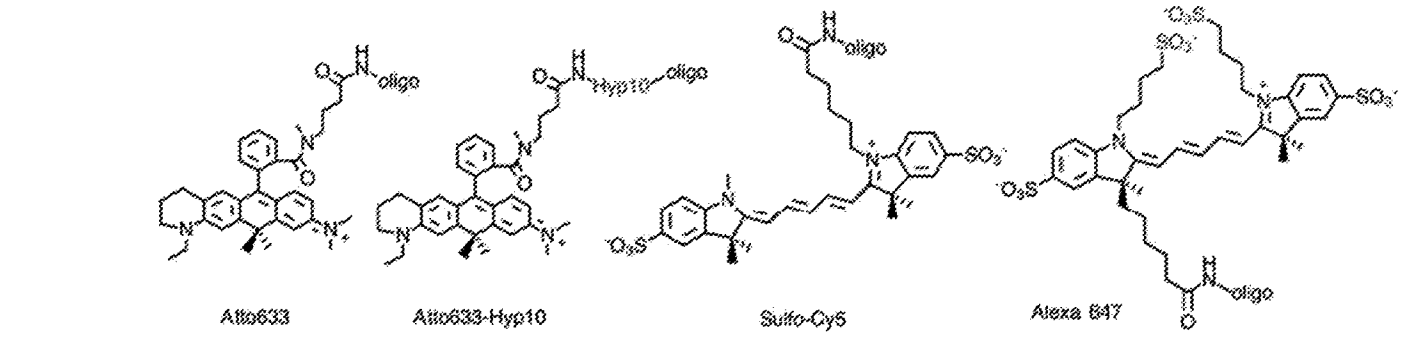

Atto633
Atto633-Hyp10
Sulfo-Cy5
Alexa 647
% Quenched:
80%
50%
90%
90%

*FIG. 16*

*FIG. 17*



Date Reçue/Date Received 2023-12-22

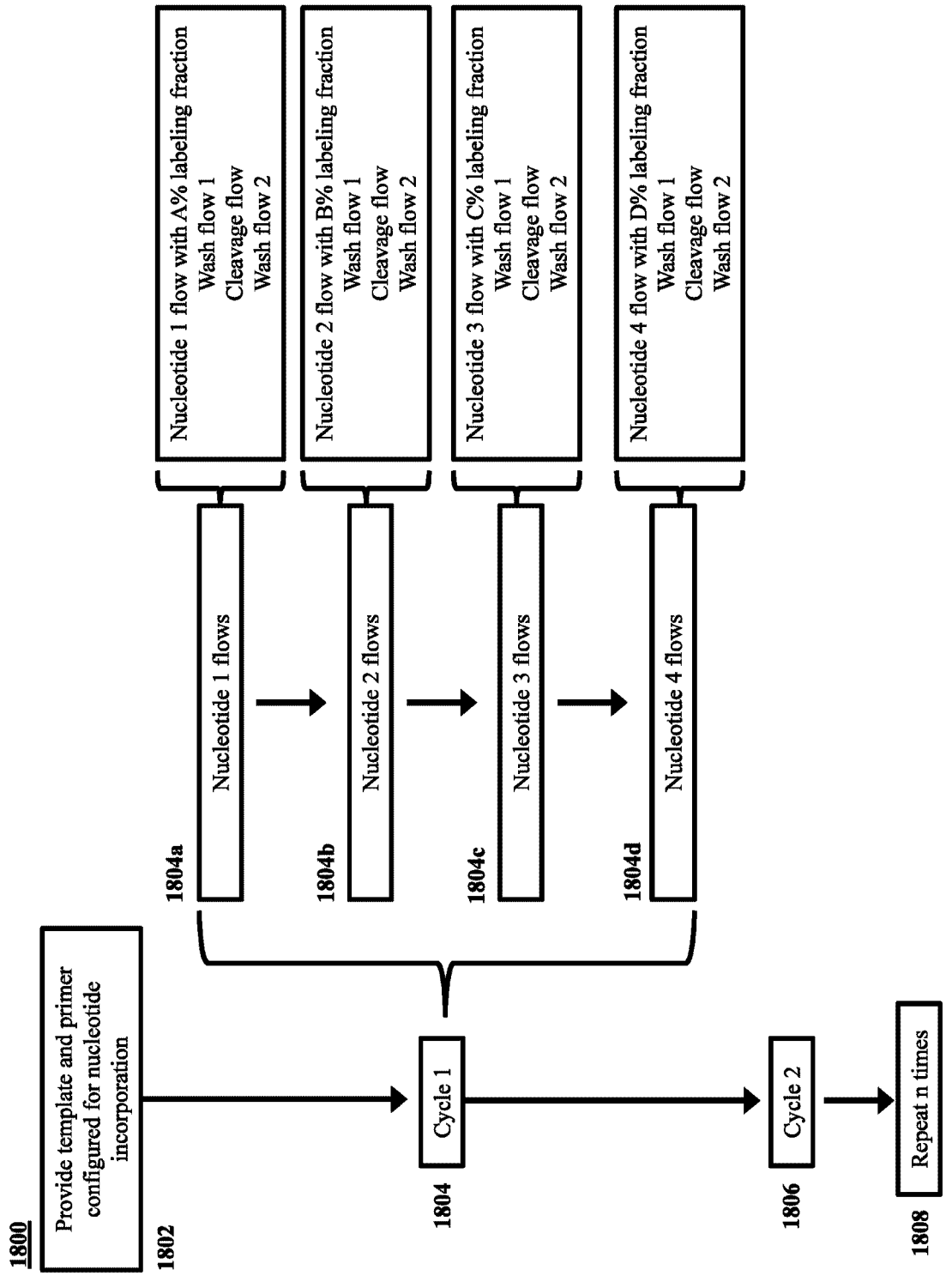



*FIG. 18*



Date Reçue/Date Received 2023-12-22

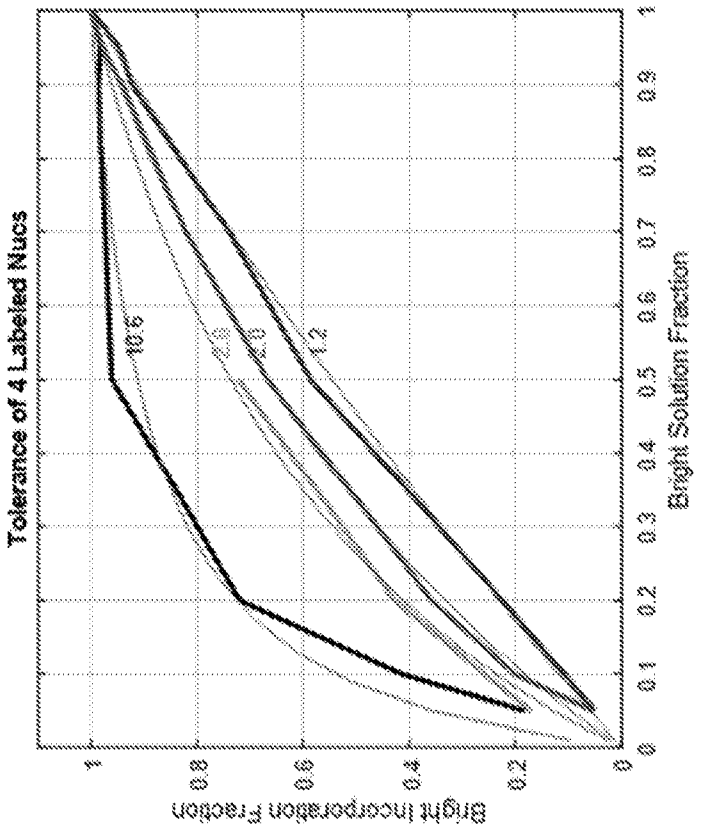


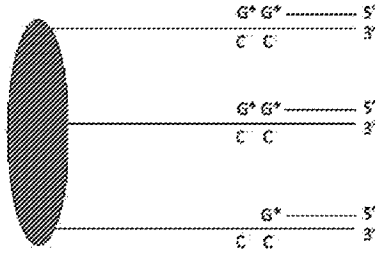

*FIG. 19*



Date Reçue/Date Received 2023-12-22

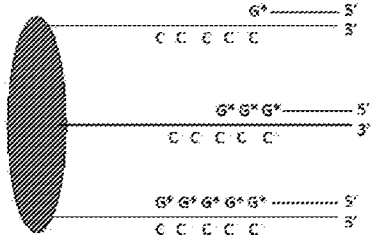
**20A**

Example of construct having 2 consecutive C's.

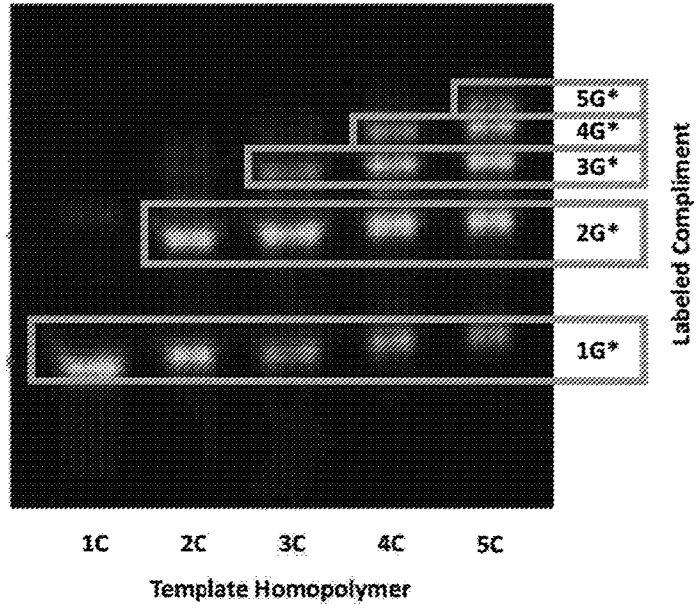



**20B**

Example of construct having 5 consecutive C's.




**20C**




*FIGs. 20A-20C*

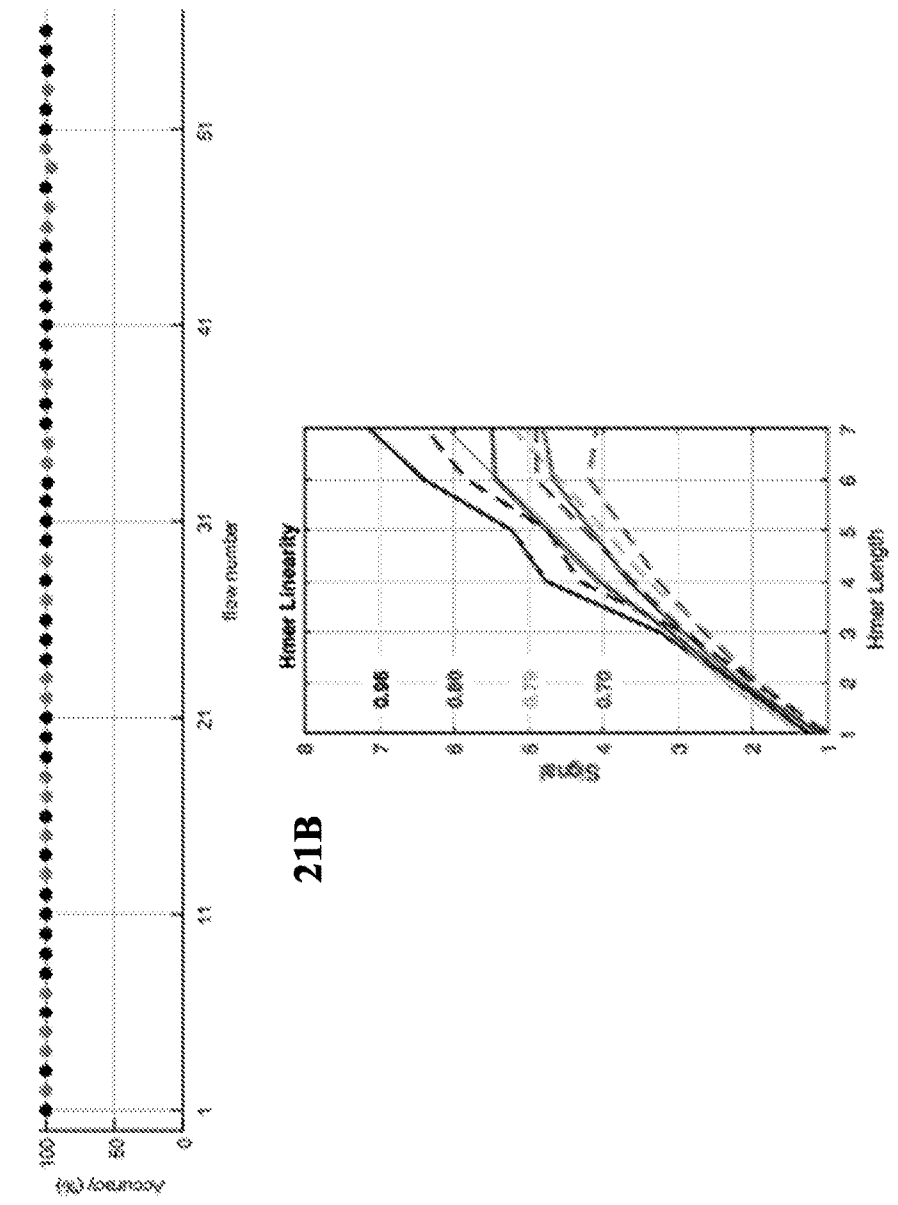



*FIG. 21A-21B*



Date Reçue/Date Received 2023-12-22

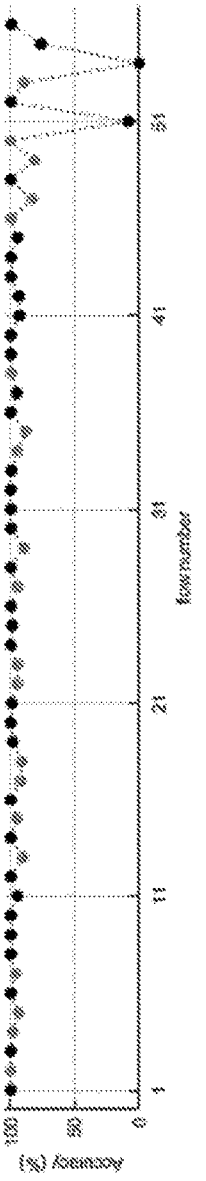



*FIG. 22*



Date Reçue/Date Received 2023-12-22