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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
03 October 2024 (03.10.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/206407 A2

(51) International Patent Classification:
C09B 57/08 (2006.01) *G01N 21/64* (2006.01)
G01N 21/00 (2006.01) *C12Q 1/6869* (2018.01)
C12Q 1/68 (2018.01)

(21) International Application Number:
PCT/US2024/021630

(22) International Filing Date:
27 March 2024 (27.03.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/492,896 29 March 2023 (29.03.2023) US

(71) Applicant: **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).

(72) Inventors: **PIEKOS, Justyna**; c/o Illumina, Inc., 5200 Illumina Way, San Diego, California 92122 (US). **VY-BORNYI, Mykhailo**; c/o Illumina, Inc., 5200 Illumina Way, San Diego, California 92122 (US). **CALLINGHAM, Michael**; c/o Illumina, Inc., 5200 Illumina Way, San Diego, California 92122 (US).

(74) Agent: **DAI, Jane Qian**; KNOBBE MARTENS OLSON & BEAR LLP, 2040 MAIN STREET, FOURTEENTH FLOOR, IRVINE, California 92614 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

(54) **Title:** NAPHTHALIMIDE DYES AND USES IN NUCLEIC ACID SEQUENCING

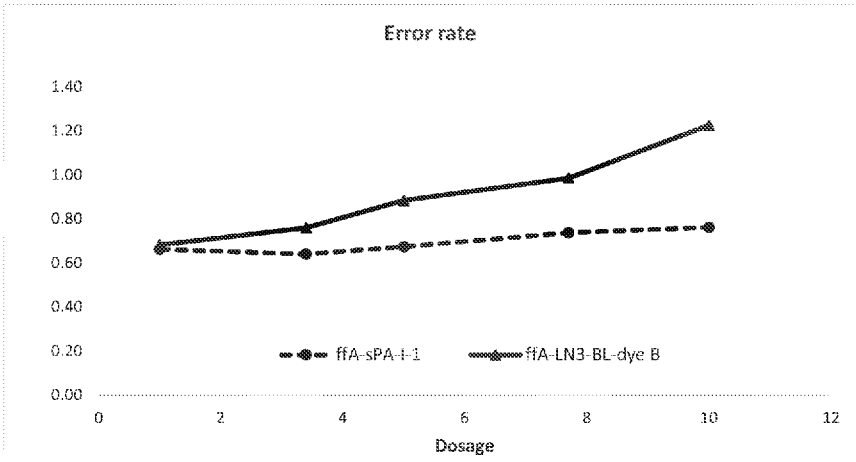


FIG. 2B

(57) **Abstract:** The present application relates to dyes containing a naphthalimide core and their uses as fluorescent labels. These dyes may be used as fluorescent labels for nucleotides in nucleic acid sequencing applications.

WO 2024/206407 A2

[Continued on next page]

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

NAPHTHALIMIDE DYES AND USES IN NUCLEIC ACID SEQUENCING

Field

[0001] The present application relates to dyes containing a 1,8-naphthalimide core and their uses as fluorescent labels in nucleic acid sequencing applications.

BACKGROUND

[0002] Non-radioactive detection of nucleic acids bearing fluorescent labels is an important technology in molecular biology. Many procedures employed in recombinant DNA technology previously relied on the use of nucleotides or polynucleotides radioactively labeled with, for example ^{32}P. Radioactive compounds permit sensitive detection of nucleic acids and other molecules of interest. However, there are serious limitations in the use of radioactive isotopes such as their expense, limited shelf life, insufficient sensitivity, and, more importantly, safety considerations. Eliminating the need for radioactive labels reduces both the safety risks and the environmental impact and costs associated with, for example, reagent disposal. Methods amenable to non-radioactive fluorescent detection include by way of non-limiting examples, automated DNA sequencing, hybridization methods, real-time detection of polymerase-chain-reaction products, and immunoassays.

[0003] For many applications, it is desirable to employ multiple spectrally distinguishable fluorescent labels to achieve independent detection of a plurality of spatially-overlapping analytes. In such multiplex methods, the number of reaction vessels may be reduced, simplifying experimental protocols and facilitating the production of application-specific reagent kits. In multi-color automated DNA sequencing systems for example, multiplex fluorescent detection allows for the analysis of multiple nucleotide bases in a single electrophoresis lane, thereby increasing throughput over single-color methods, and reducing uncertainties associated with inter-lane electrophoretic mobility variations.

[0004] However, multiplex fluorescent detection can be problematic and there are a number of important factors that constrain selection of appropriate fluorescent labels. First, it may be difficult to find dye compounds with substantially resolved absorption and emission spectra in a given application. In addition, when several fluorescent dyes are used together, generating fluorescence signals in distinguishable spectral regions by simultaneous excitation may be complicated because absorption bands of the dyes are usually widely separated, so it is difficult to achieve comparable fluorescence excitation efficiencies even for two dyes. Many excitation methods use high power light sources like lasers and therefore the dye must have sufficient photo-

stability to withstand such excitation. A final consideration of particular importance to molecular biology methods is the extent to which the fluorescent dyes must be compatible with reagent chemistries such as, for example, DNA synthesis solvents and reagents, buffers, polymerase enzymes, and ligase enzymes.

[0005] Fluorescent dye molecules with improved fluorescence properties such as suitable fluorescence intensity, shape, and wavelength maximum of fluorescence band can improve the speed and accuracy of nucleic acid sequencing. Strong fluorescence signals are especially important when measurements are made in water-based biological buffers and at higher temperatures as the fluorescence intensities of most organic dyes are significantly lower under such conditions. Moreover, the nature of the base to which a dye is attached also affects the fluorescence maximum, fluorescence intensity, and others spectral dye properties. The sequence-specific interactions between the nucleobases and the fluorescent dyes can be tailored by specific design of the fluorescent dyes. Optimization of the structure of the fluorescent dyes can improve the efficiency of nucleotide incorporation, reduce the level of sequencing errors, and decrease the usage of reagents in, and therefore the costs of, nucleic acid sequencing.

[0006] Some optical and technical developments have already led to greatly improved image quality but were ultimately limited by poor optical resolution. Generally, optical resolution of light microscopy is limited to objects spaced at approximately half of the wavelength of the light used. In practical terms, then, only objects that are laying quite far apart (at least 200 to 350 nm) could be resolved by light microscopy. One way to improve image resolution and increase the number of resolvable objects per unit of surface area is to use excitation light of a shorter wavelength. For example, if light wavelength is shortened by $\Delta\lambda$~100 nm with the same optics, resolution will be better (about Δ 50 nm / (about 15 %)), less-distorted images will be recorded, and the density of objects on the recognizable area will be increased about 35%.

[0007] Certain nucleic acid sequencing methods employ laser light to excite and detect dye-labeled nucleotides. These instruments use longer wavelength light, such as red lasers, along with appropriate dyes that are excitable at 660 nm. To detect more densely packed nucleic acid sequencing clusters while maintaining useful resolution, a shorter wavelength blue light source (450-460 nm) may be used. In this case, optical resolution will be limited not by the emission wavelength of the longer wavelength red fluorescent dyes but rather by the emission of dyes excitable by the next longest wavelength light source, for example, by a "green laser" at 532 nm. Thus, there is a need for blue dye labels for use in fluorescence detection in sequencing applications.

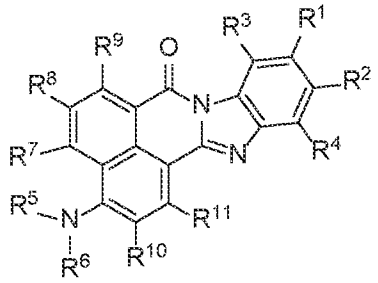
[0008] Although blue-dye chemistry and associated laser technologies have improved, appropriate commercially available blue dyes with strong fluorescence for nucleotide labeling are

still quite rare. However, certain dyes such as coumarin dyes are not stable in an aqueous environment for a prolonged period of time. For example, in basic conditions certain coumarin dyes may be easily attacked by nucleophiles, thus resulted in disturbance or deterioration of the dyes. Naphthalimide dyes have been reported in scientific literature, for example, *Dyes and Pigments* 2008, 79, 24–32; *Dyes and Pigments* 2018, 148, 16–24; *Chembiochem* 2016, 17 (18):1719–1724. However, designing new naphthalimide dyes with appropriate adsorption, good chemical stability, and tailored Stokes shifts as nucleic acid labels for sequencing application remains challenging.

SUMMARY

[0009] Described herein are dyes containing naphthalimide cores with improved chemical stability and strong fluorescence under blue light excitation (e.g., blue LED or laser at about 450 nm to about 460 nm). These dyes also have highly tunable absorption and emission properties that are suitable for nucleic acid labeling.

[0010] One aspect of the present disclosure relates to a compound of Formula (I):



(I), a salt or a mesomeric form thereof,

wherein each of R^1, R^2, R^3, R^4, R^7, R^8, R^9, R^{10}, and R^{11} is independently H, unsubstituted or substituted C_1-C_6 alkyl, C_1-C_6 alkoxy, unsubstituted or substituted C_2-C_6 alkenyl, unsubstituted or substituted C_2-C_6 alkynyl, C_1-C_6 haloalkyl, C_1-C_6 haloalkoxy, C_1-C_6 hydroxyalkyl, (C_1-C_6 alkoxy)C_1-C_6 alkyl, -O-(C_1-C_6 alkoxy)C_1-C_6 alkyl, unsubstituted or substituted amino, halo, cyano, carboxy, hydroxy, nitro, sulfonyl, sulfino, sulfo, sulfonate, S-sulfonamido, N-sulfonamido,–C(=O)OR^A, –C(=O)NR^BR^C or –SO_3R^D;

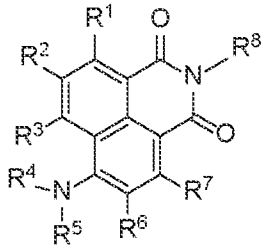
each of R^A, R^B, R^C, and R^D is independently H, or unsubstituted or substituted C_1-C_6 alkyl; and

each of R^5 and R^6 is independently H, unsubstituted or substituted C_1-C_6 alkyl; unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

alternatively, R^5 and R^6 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl;

provided that the compound of Formula (I) comprises one carboxy group.

[0011] Another aspect of the present disclosure relates to a compound of Formula (II):



(II), a salt or a mesomeric form thereof,

wherein each of R^1, R^2, R^3, R^6, and R^7 is independently H, unsubstituted or substituted C_1-C_6 alkyl, C_1-C_6 alkoxy, unsubstituted or substituted C_2-C_6 alkenyl, unsubstituted or substituted C_2-C_6 alkynyl, C_1-C_6 haloalkyl, C_1-C_6 haloalkoxy, C_1-C_6 hydroxyalkyl, (C_1-C_6 alkoxy)C_1-C_6 alkyl, -O-(C_1-C_6 alkoxy)C_1-C_6 alkyl, unsubstituted or substituted amino, halo, cyano, carboxy, hydroxy, nitro, sulfonyl, sulfino, sulfo, sulfonate, S-sulfonamido, N-sulfonamido, $-C(=O)OR^A$, $-C(=O)NR^BR^C$ or $-SO_3R^D$;

R^8 is unsubstituted or substituted C_1-C_6 alkyl, unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

each of R^A, R^B, R^C, and R^D is independently H, or unsubstituted or substituted C_1-C_6 alkyl; and

each of R^4 and R^5 is independently H, unsubstituted or substituted C_1-C_6 alkyl; unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

alternatively, R^4 and R^5 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl;

provided that the compound of Formula (II) comprises one carboxy group.

[0012] In some aspect, a compound of the present disclosure is labeled or conjugated with a substrate moiety such as, for example, a nucleoside, nucleotide, polynucleotide, polypeptide, carbohydrate, ligand, particle, cell, semi-solid surface (e.g., gel), or solid surface.

The labeling or conjugation may be carried out via a carboxyl group, which can be reacted using methods known in the art with an amino or hydroxy group on a moiety (such as a nucleotide) or a linker bound thereto, to form an amide or ester.

[0013] Another aspect of the present disclosure relates to dye compounds comprising linker groups to enable, for example, covalent attachment to a substrate moiety (such as a nucleotide). Linking may be carried out at any position of the dye. In some embodiments, linking may be carried out via one of R^1, R^2, R^3, R^4, R^5, R^6, R^7, R^8, R^9, R^{10}, and R^{11} of Formula (I) or R^1, R^2, R^3, R^4, R^5, R^6, R^7 and R^8 of Formula (II). In some embodiments, linking may be carried out via a carboxyl group of Formula (I) or Formula (II).

[0014] A further aspect of the present disclosure provides a labeled nucleoside or nucleotide compound defined by the formula:

N-L-Dye

wherein N is a nucleoside or nucleotide;

L is an optional linker moiety; and

Dye is a naphthalimide moiety of Formula (I) or Formula (II) according to the present disclosure, where a functional group (e.g., a carboxyl group) of the compound of Formula (I) or Formula (II) reacts with an amino or hydroxy group of the linker moiety or the nucleoside/nucleotide to form covalent bonding.

[0015] Some additional aspect of the present disclosure relates to an oligonucleotide or polynucleotide comprising a nucleotide labeled with a compound of Formula (I) or Formula (II) incorporated therein.

[0016] Some additional aspect of the present disclosure relates to a kit comprising a dye compound (free or in labeled form) that may be used in various immunological assays, oligonucleotide or nucleic acid labeling, or for DNA sequencing by synthesis. In yet another aspect, the disclosure provides kits comprising dye "sets" particularly suited to cycles of sequencing by synthesis on an automated instrument platform. In some aspect, are kits containing one or more nucleotides where at least one nucleotide is a labeled nucleotide described herein.

[0017] A further aspect of the disclosure relates to a method of determining the sequences of a plurality of target polynucleotides, comprising:

(a) contacting a solid support with a solution comprising sequencing primers under hybridization conditions, wherein the solid support comprises a plurality of different target polynucleotides immobilized thereon; and the sequencing primers are complementary to at least a portion of the target polynucleotides;

(b) contacting the solid support with an aqueous solution comprising DNA polymerase and one or more of four different types of nucleotides (A, C, T, G or U; dATP, dCTP, dTTP and

dGTP or dUTP) under conditions suitable for DNA polymerase-mediated primer extension, and incorporating one type of nucleotides into the sequencing primers to produce extended copy polynucleotides, wherein at least one type of nucleotide is a labeled nucleotide described herein, and wherein each of the one or more of four different type of nucleotides comprises a 3′ blocking group covalently attached to the 2′ deoxyribose of the nucleotide;

(c) imaging and performing one or more fluorescent measurements of the extended copy polynucleotides; and

(d) removing the 3′ blocking group from the nucleotides incorporated into the extended copy polynucleotides.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] **FIGs. 1A and 1B** are the scatterplots obtained at cycle 26 on an Illumina MiSeq® instrument with 2Ex-2Ch mode and 1Ex-2Ch mode, respectively, using an incorporation mix including fully functionalized A nucleotide (ffA) labeled with dye I-1.

[0019] **FIG. 2A** is a line chart showing percentage of signal decay as a function of blue laser dosage on an Illumina MiSeq® instrument with 2Ex-2Ch mode comparing an incorporation mix with ffA labeled with a known blue coumarin dye C, to an incorporation mix with ffA labeled with dye I-1. **FIG. 2B** is a line chart showing the percent error rate as a function of blue laser dosage for the same sequencing runs. **FIG. 2C** is a line chart showing the percent phasing as a function of blue laser dosage for the same sequencing runs.

[0020] **FIG. 3A** is a line chart showing percentage of signal decay as a function of blue laser dosage on an Illumina MiSeq® instrument with 1Ex-2Ch mode comparing an incorporation mix with ffA labeled with a known blue coumarin dye C, to an incorporation mix with ffA labeled with dye I-1. **FIG. 3B** is a line chart showing the percent error rate as a function of blue laser dosage for the same sequencing runs. **FIG. 3C** is a line chart showing the percent phasing as a function of blue laser dosage for the same sequencing runs.

DETAILED DESCRIPTION

[0021] Embodiments of the present disclosure relate to dyes containing naphthalimide cores with enhanced fluorescent intensity, tunable Stokes shift and improved chemical stability. In some embodiments, Stokes shift of the dyes described herein range from about 25 nm to 100 nm, or from 40 nm to about 80 nm. The naphthalimide dyes described herein may be used in Illumina's sequencing platforms, for example, MiSeq™ with two-channel detection and one blue light excitation or two-channel detection and blue/green excitations.

Definitions

[0022] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0023] It is noted that, as used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless expressly and unequivocally limited to one referent. It will be apparent to those skilled in the art that various modifications and variations can be made to various embodiments described herein without departing from the spirit or scope of the present teachings. Thus, it is intended that the various embodiments described herein cover other modifications and variations within the scope of the appended claims and their equivalents.

[0024] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art. The use of the term "including" as well as other forms, such as "include", "includes," and "included," is not limiting. The use of the term "having" as well as other forms, such as "have", "has," and "had," is not limiting. As used in this specification, whether in a transitional phrase or in the body of the claim, the terms "comprise(s)" and "comprising" are to be interpreted as having an open-ended meaning. That is, the above terms are to be interpreted synonymously with the phrases "having at least" or "including at least." For example, when used in the context of a process, the term "comprising" means that the process includes at least the recited steps but may include additional steps. When used in the context of a compound, composition, or device, the term "comprising" means that the compound, composition, or device includes at least the recited features or components, but may also include additional features or components.

[0025] As used herein, common organic abbreviations are defined as follows:

°C	Temperature in degrees Centigrade
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
ddNTP	Dideoxynucleotide triphosphate
ffA	Fully functionalized A nucleotide
ffC	Fully functionalized C nucleotide
ffG	Fully functionalized G nucleotide
ffN	Fully functionalized nucleotide
ffT	Fully functionalized T nucleotide
h	Hour(s)
RT	Room temperature
SBS	Sequencing by Synthesis
USM	Universal scan mix

[0026] As used herein, the term "array" refers to a population of different probe molecules that are attached to one or more substrates such that the different probe molecules can be differentiated from each other according to relative location. An array can include different probe molecules that are each located at a different addressable location on a substrate. Alternatively, or additionally, an array can include separate substrates each bearing a different probe molecule, wherein the different probe molecules can be identified according to the locations of the substrates on a surface to which the substrates are attached or according to the locations of the substrates in a liquid. Exemplary arrays in which separate substrates are located on a surface include, without limitation, those including beads in wells as described, for example, in U.S. Patent No. 6,355,431 B1, US 2002/0102578 and PCT Publication No. WO 00/63437. Exemplary formats that can be used in the invention to distinguish beads in a liquid array, for example, using a microfluidic device, such as a fluorescent activated cell sorter (FACS), are described, for example, in US Pat. No. 6,524,793. Further examples of arrays that can be used in the invention include, without limitation, those described in U.S. Pat Nos. 5,429,807; 5,436,327; 5,561,071; 5,583,211; 5,658,734; 5,837,858; 5,874,219; 5,919,523; 6,136,269; 6,287,768; 6,287,776; 6,288,220; 6,297,006; 6,291,193; 6,346,413; 6,416,949; 6,482,591; 6,514,751 and 6,610,482; and WO 93/17126; WO 95/11995; WO 95/35505; EP 742 287; and EP 799 897.

[0027] As used herein, the term "covalently attached" or "covalently bonded" refers to the forming of a chemical bonding that is characterized by the sharing of pairs of electrons between atoms. For example, a covalently attached polymer coating refers to a polymer coating that forms chemical bonds with a functionalized surface of a substrate, as compared to attachment to the surface via other means, for example, adhesion or electrostatic interaction. It will be appreciated that polymers that are attached covalently to a surface can also be bonded via other means in addition to covalent attachment.

[0028] The term "halogen" or "halo," as used herein, means any one of the radio-stable atoms of column 7 of the Periodic Table of the Elements, *e.g.*, fluorine, chlorine, bromine, or iodine, with fluorine and chlorine being preferred.

[0029] As used herein, "C_a to C_b" in which "a" and "b" are integers refer to the number of carbon atoms in an alkyl, alkenyl or alkynyl group, or the number of ring atoms of a cycloalkyl or aryl group. That is, the alkyl, the alkenyl, the alkynyl, the ring of the cycloalkyl, and ring of the aryl can contain from "a" to "b", inclusive, carbon atoms. For example, a "C_1 to C_4 alkyl" group refers to all alkyl groups having from 1 to 4 carbons, that is, CH_3-, CH_3CH_2-, $CH_3CH_2CH_2$-, $(CH_3)_2CH$-, $CH_3CH_2CH_2CH_2$-, $CH_3CH_2CH(CH_3)$- and $(CH_3)_3C$-; a C_3 to C_4 cycloalkyl group refers to all cycloalkyl groups having from 3 to 4 carbon atoms, that is, cyclopropyl and cyclobutyl. Similarly, a "4 to 6 membered heterocyclyl" group refers to all heterocyclyl groups

with 4 to 6 total ring atoms, for example, azetidine, oxetane, oxazoline, pyrrolidine, piperidine, piperazine, morpholine, and the like. If no "a" and "b" are designated with regard to an alkyl, alkenyl, alkynyl, cycloalkyl, or aryl group, the broadest range described in these definitions is to be assumed. As used herein, the term "C_1-C_6" includes C_1, C_2, C_3, C_4, C_5 and C_6, and a range defined by any of the two numbers. For example, C_1-C_6 alkyl includes C_1, C_2, C_3, C_4, C_5 and C_6 alkyl, C_2-C_6 alkyl, C_1-C_3 alkyl, etc. Similarly, C_2-C_6 alkenyl includes C_2, C_3, C_4, C_5 and C_6 alkenyl, C_2-C_5 alkenyl, C_3-C_4 alkenyl, etc.; and C_2-C_6 alkynyl includes C_2, C_3, C_4, C_5 and C_6 alkynyl, C_2-C_5 alkynyl, C_3-C_4 alkynyl, etc. C_3-C_8 cycloalkyl each includes hydrocarbon ring containing 3, 4, 5, 6, 7 and 8 carbon atoms, or a range defined by any of the two numbers, such as C_3-C_7 cycloalkyl or C_5-C_6 cycloalkyl.

[0030] As used herein, "alkyl" refers to a straight or branched hydrocarbon chain that is fully saturated (i.e., contains no double or triple bonds). The alkyl group may have 1 to 20 carbon atoms (whenever it appears herein, a numerical range such as "1 to 20" refers to each integer in the given range; *e.g.*, "1 to 20 carbon atoms" means that the alkyl group may consist of 1 carbon atom, 2 carbon atoms, 3 carbon atoms, *etc.*, up to and including 20 carbon atoms, although the present definition also covers the occurrence of the term "alkyl" where no numerical range is designated). The alkyl group may also be a medium size alkyl having 1 to 9 carbon atoms. The alkyl group could also be a lower alkyl having 1 to 6 carbon atoms. By way of example only, "C_1-$_6$ alkyl" or "C_1-C_6 alkyl" indicates that there are one to six carbon atoms in the alkyl chain, i.e., the alkyl chain is selected from the group consisting of methyl, ethyl, propyl, iso-propyl, n-butyl, iso-butyl, sec-butyl, and t-butyl. Typical alkyl groups include, but are in no way limited to, methyl, ethyl, propyl, isopropyl, butyl, isobutyl, tertiary butyl, pentyl, hexyl, and the like.

[0031] As used herein, "alkoxy" refers to the formula –OR wherein R is an alkyl as is defined above, such as ""C_1-$_9$ alkoxy" or "C_1-C_9 alkoxy", including but not limited to methoxy, ethoxy, n-propoxy, 1-methylethoxy (isopropoxy), n-butoxy, iso-butoxy, sec-butoxy, and tert-butoxy, and the like.

[0032] As used herein, "alkenyl" refers to a straight or branched hydrocarbon chain containing one or more double bonds. The alkenyl group may have 2 to 20 carbon atoms, although the present definition also covers the occurrence of the term "alkenyl" where no numerical range is designated. The alkenyl group may also be a medium size alkenyl having 2 to 9 carbon atoms. The alkenyl group could also be a lower alkenyl having 2 to 6 carbon atoms. By way of example only, "C_2-C_6 alkenyl" or "C_2-$_6$ alkenyl" indicates that there are two to six carbon atoms in the alkenyl chain, i.e., the alkenyl chain is selected from the group consisting of ethenyl, propen-1-yl, propen-2-yl, propen-3-yl, buten-1-yl, buten-2-yl, buten-3-yl, buten-4-yl, 1-methyl-propen-1-yl, 2-methyl-propen-1-yl, 1-ethyl-ethen-1-yl, 2-methyl-propen-3-yl, buta-1,3-dienyl, buta-1,2,-

dienyl, and buta-1,2-dien-4-yl. Typical alkenyl groups include, but are in no way limited to, ethenyl, propenyl, butenyl, pentenyl, and hexenyl, and the like.

[0033] As used herein, "alkynyl" refers to a straight or branched hydrocarbon chain containing one or more triple bonds. The alkynyl group may have 2 to 20 carbon atoms, although the present definition also covers the occurrence of the term "alkynyl" where no numerical range is designated. The alkynyl group may also be a medium size alkynyl having 2 to 9 carbon atoms. The alkynyl group could also be a lower alkynyl having 2 to 6 carbon atoms. By way of example only, "C_{2-6} alkynyl" or "C_2-C_6 alkenyl" indicates that there are two to six carbon atoms in the alkynyl chain, i.e., the alkynyl chain is selected from the group consisting of ethynyl, propyn-1-yl, propyn-2-yl, butyn-1-yl, butyn-3-yl, butyn-4-yl, and 2-butynyl. Typical alkynyl groups include, but are in no way limited to, ethynyl, propynyl, butynyl, pentynyl, and hexynyl, and the like.

[0034] The term "aromatic" refers to a ring or ring system having a conjugated pi electron system and includes both carbocyclic aromatic (e.g., phenyl) and heterocyclic aromatic groups (e.g., pyridine). The term includes monocyclic or fused-ring polycyclic (i.e., rings which share adjacent pairs of atoms) groups provided that the entire ring system is aromatic.

[0035] As used herein, "aryl" refers to an aromatic ring or ring system (i.e., two or more fused rings that share two adjacent carbon atoms) containing only carbon in the ring backbone. When the aryl is a ring system, every ring in the system is aromatic. The aryl group may have 6 to 18 carbon atoms, although the present definition also covers the occurrence of the term "aryl" where no numerical range is designated. In some embodiments, the aryl group has 6 to 10 carbon atoms. The aryl group may be designated as "C_6-C_{10} aryl," "C_6 or C_{10} aryl," or similar designations. Examples of aryl groups include, but are not limited to, phenyl, naphthyl, azulenyl, and anthracenyl.

[0036] An "aralkyl" or "arylalkyl" is an aryl group connected, as a substituent, via an alkylene group, such as "C_{7-14} aralkyl" and the like, including but not limited to benzyl, 2-phenylethyl, 3-phenylpropyl, and naphthylalkyl. In some cases, the alkylene group is a lower alkylene group (i.e., a C_{1-6} alkylene group).

[0037] As used herein, "heteroaryl" refers to an aromatic ring or ring system (i.e., two or more fused rings that share two adjacent atoms) that contain(s) one or more heteroatoms, that is, an element other than carbon, including but not limited to, nitrogen, oxygen and sulfur, in the ring backbone. When the heteroaryl is a ring system, every ring in the system is aromatic. The heteroaryl group may have 5-18 ring members (i.e., the number of atoms making up the ring backbone, including carbon atoms and heteroatoms), although the present definition also covers the occurrence of the term "heteroaryl" where no numerical range is designated. In some

embodiments, the heteroaryl group has 5 to 10 ring members or 5 to 7 ring members. The heteroaryl group may be designated as "5-7 membered heteroaryl," "5-10 membered heteroaryl," or similar designations. Examples of heteroaryl rings include, but are not limited to, furyl, thienyl, phthalazinyl, pyrrolyl, oxazolyl, thiazolyl, imidazolyl, pyrazolyl, isoxazolyl, isothiazolyl, triazolyl, thiadiazolyl, pyridinyl, pyridazinyl, pyrimidinyl, pyrazinyl, triazinyl, quinolinyl, isoquinolinyl, benzimidazolyl, benzoxazolyl, benzothiazolyl, indolyl, isoindolyl, and benzothienyl.

[0038] A "heteroaralkyl" or "heteroarylalkyl" is heteroaryl group connected, as a substituent, via an alkylene group. Examples include but are not limited to 2-thienylmethyl, 3-thienylmethyl, furylmethyl, thienylethyl, pyrrolylalkyl, pyridylalkyl, isoxazolylalkyl, and imidazolylalkyl. In some cases, the alkylene group is a lower alkylene group (i.e., a C_{1-6} alkylene group).

[0039] As used herein, "carbocyclyl" means a non-aromatic cyclic ring or ring system containing only carbon atoms in the ring system backbone. When the carbocyclyl is a ring system, two or more rings may be joined together in a fused, bridged or spiro-connected fashion. Carbocyclyls may have any degree of saturation provided that at least one ring in a ring system is not aromatic. Thus, carbocyclyls include cycloalkyls, cycloalkenyls, and cycloalkynyls. The carbocyclyl group may have 3 to 20 carbon atoms, although the present definition also covers the occurrence of the term "carbocyclyl" where no numerical range is designated. The carbocyclyl group may also be a medium size carbocyclyl having 3 to 10 carbon atoms. The carbocyclyl group could also be a carbocyclyl having 3 to 6 carbon atoms. The carbocyclyl group may be designated as "C_{3-6} carbocyclyl", "C_3-C_6 carbocyclyl" or similar designations. Examples of carbocyclyl rings include, but are not limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cyclohexenyl, 2,3-dihydro-indene, bicycle[2.2.2]octanyl, adamantyl, and spiro[4.4]nonanyl.

[0040] As used herein, "cycloalkyl" means a fully saturated carbocyclyl ring or ring system. Examples include cyclopropyl, cyclobutyl, cyclopentyl, and cyclohexyl.

[0041] As used herein, "heterocyclyl" means a non-aromatic cyclic ring or ring system containing at least one heteroatom in the ring backbone. Heterocyclyls may be joined together in a fused, bridged or spiro-connected fashion. Heterocyclyls may have any degree of saturation provided that at least one ring in the ring system is not aromatic. The heteroatom(s) may be present in either a non-aromatic or aromatic ring in the ring system. The heterocyclyl group may have 3 to 20 ring members (i.e., the number of atoms making up the ring backbone, including carbon atoms and heteroatoms), although the present definition also covers the occurrence of the term "heterocyclyl" where no numerical range is designated. The heterocyclyl group may also be a medium size heterocyclyl having 3 to 10 ring members. The heterocyclyl group could also be a

heterocyclyl having 3 to 6 ring members. The heterocyclyl group may be designated as "3-6 membered heterocyclyl" or similar designations. In preferred six membered monocyclic heterocyclyls, the heteroatom(s) are selected from one up to three of O, N or S, and in preferred five membered monocyclic heterocyclyls, the heteroatom(s) are selected from one or two heteroatoms selected from O, N, or S. Examples of heterocyclyl rings include, but are not limited to, aziridinyl, azetidinyl, azepanyl, azepinyl, acridinyl, carbazolyl, cinnolinyl, dioxolanyl, imidazolinyl, imidazolidinyl, morpholinyl, oxiranyl, oxepanyl, thiepanyl, piperidinyl, piperazinyl, dioxopiperazinyl, pyrrolidinyl, pyrrolidonyl, pyrrolidionyl, 4-piperidonyl, pyrazolinyl, pyrazolidinyl, 1,3-dioxinyl, 1,3-dioxanyl, 1,4-dioxinyl, 1,4-dioxanyl, 1,3-oxathianyl, 1,4-oxathiinyl, 1,4-oxathianyl, 2*H*-1,2-oxazinyl, trioxanyl, hexahydro-1,3,5-triazinyl, 1,3-dioxolyl, 1,3-dioxolanyl, 1,3-dithiolyl, 1,3-dithiolanyl, isoxazolinyl, isoxazolidinyl, oxazolinyl, oxazolidinyl, oxazolidinonyl, thiazolinyl, thiazolidinyl, 1,3-oxathiolanyl, indolinyl, isoindolinyl, tetrahydrofuranyl, tetrahydropyranyl, tetrahydrothiophenyl, tetrahydrothiopyranyl, tetrahydro-1,4-thiazinyl, thiamorpholinyl, dihydrobenzofuranyl, benzimidazolidinyl, and tetrahydroquinoline.

[0042] As used herein, (cycloalkyl)alkyl refers to a cycloalkyl group connected via an alkylene group, such as (C_3-C_7 cycloalkyl)C_1-C_6 alkyl. Non-limiting examples include cyclohexyl-$(CH_2)_{1-6}$-, cyclopentyl-$(CH_2)_{1-6}$-, or cyclopropyl-$(CH_2)_{1-6}$-.

[0043] As used herein, (aryl)alkyl refers to an aryl group connected via an alkylene group, such as (C_6-C_{10} aryl)C_1-C_6 alkyl. Non-limiting example includes phenyl-$(CH_2)_{1-6}$.

[0044] As used herein, (heteroaryl)alkyl refers to a heteroaryl group connected via an alkylene group, such as (5 to 10 membered heteroaryl)C_1-C_6 alkyl. Non-limiting examples include pyridyl-$(CH_2)_{1-6}$-, pyrimidinyl-$(CH_2)_{1-6}$-, or pyrrolyl-$(CH_2)_{1-6}$-.

[0045] As used herein, (heterocyclyl)alkyl refers to a heterocyclyl group connected via an alkylene group, such as (3 to 10 membered heterocyclyl)C_1-C_6 alkyl. Non-limiting examples include morpholinyl-$(CH_2)_{1-6}$-, piperidinyl-$(CH_2)_{1-6}$-, piperazinyl-$(CH_2)_{1-6}$-, pyrrolidinyl-$(CH_2)_{1-6}$-, or azetidinyl-$(CH_2)_{1-6}$-.

[0046] As used herein, "alkoxyalkyl" or "(alkoxy)alkyl" refers to an alkoxy group connected via an alkylene group, such as C_2-C_8 alkoxyalkyl, or (C_1-C_6 alkoxy)C_1-C_6 alkyl, for example, –$(CH_2)_{1-3}$-OCH_3.

[0047] An "aminoalkyl" group refers to an amino group connected via an alkylene group.

[0048] An "O-carboxy" group refers to a "-OC(=O)R" group in which R is selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein.

[0049] A "C-carboxy" group refers to a "-C(=O)OR" group in which R is selected from the group consisting of hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein. A non-limiting example includes carboxyl (i.e., -C(=O)OH).

[0050] A "sulfonyl" group refers to an "-SO_2R" group in which R is selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein.

[0051] A "sulfino" group refers to a "-S(=O)OH" group.

[0052] A "sulfo" group refers to a"-$S(=O)_2OH$" or "-SO_3H" group.

[0053] A "sulfonate" group refers to a "-SO_3^-" group.

[0054] A "sulfate" group refers to "-SO_4^-" group.

[0055] A "S-sulfonamido" group refers to a "-$SO_2NR_AR_B$" group in which R_A and R_B are each independently selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein.

[0056] An "N-sulfonamido" group refers to a "-$N(R_A)SO_2R_B$" group in which R_A and R_b are each independently selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein.

[0057] A "C-amido" group refers to a "-$C(=O)NR_AR_B$" group in which R_A and R_B are each independently selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein.

[0058] An "N-amido" group refers to a "-$N(R_A)C(=O)R_B$" group in which R_A and R_B are each independently selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein.

[0059] An "amino" group refers to a "-NR_AR_B" group in which R_A and R_B are each independently selected from hydrogen, C_{1-6} alkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{3-7} carbocyclyl, C_{6-10} aryl, 5-10 membered heteroaryl, and 3-10 membered heterocyclyl, as defined herein. A non-limiting example includes free amino (i.e., -NH_2).

[0060] As used herein, "–OAc" or "-O-acyl" refers to acetyloxy with the structure –O-$C(=O)CH_3$.

[0061] When a group is described as "optionally substituted" it may be either unsubstituted or substituted. Likewise, when a group is described as being "substituted", the substituent may be selected from one or more of the indicated substituents. As used herein, a

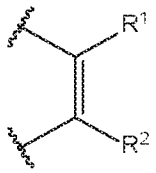
substituted group is derived from the unsubstituted parent group in which there has been an exchange of one or more hydrogen atoms for another atom or group. Unless otherwise indicated, when a group is deemed to be "substituted," it is meant that the group is substituted with one or more substituents independently selected from C_1-C_6 alkyl, C_1-C_6 alkenyl, C_1-C_6 alkynyl, C_3-C_7 carbocyclyl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), C_3-C_7-carbocyclyl-C_1-C_6-alkyl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), 3-10 membered heterocyclyl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), 3-10 membered heterocyclyl-C_1-C_6-alkyl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), aryl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), aryl(C_1-C_6)alkyl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), 5-10 membered heteroaryl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), 5-10 membered heteroaryl(C_1-C_6)alkyl (optionally substituted with halo, C_1-C_6 alkyl, C_1-C_6 alkoxy, C_1-C_6 haloalkyl, and C_1-C_6 haloalkoxy), halo, -CN, hydroxy, C_1-C_6 alkoxy, C_1-C_6 alkoxy(C_1-C_6)alkyl (i.e., ether), aryloxy, sulfhydryl (mercapto), halo(C_1-C_6)alkyl (e.g., $-CF_3$), halo(C_1-C_6)alkoxy (e.g., $-OCF_3$), C_1-C_6 alkylthio, arylthio, amino, amino(C_1-C_6)alkyl, nitro, O-carbamyl, N-carbamyl, O-thiocarbamyl, N-thiocarbamyl, C-amido, N-amido, S-sulfonamido, N-sulfonamido, C-carboxy, O-carboxy, acyl, cyanato, isocyanato, thiocyanato, isothiocyanato, sulfinyl, sulfonyl, $-SO_3H$, sulfonate, sulfate, sulfino, $-OSO_2C_1$-C_4alkyl, and oxo (=O). Wherever a group is described as "optionally substituted" that group can be substituted with the above substituents. In some embodiments, when an alkyl, alkenyl, alkynyl, aryl, heteroaryl, carbocyclyl or heterocyclyl group is substituted, each is independently substituted with one or more substituents selected from the group consisting of halo, -CN, $-SO_3^-$, $-OSO_3^-$, $-SO_3H$, $-SR^A$, $-OR^A$, $-NR^BR^C$, oxo, $-CONR^BR^C$, $-SO_2NR^BR^C$, -COOH, and $-COOR^B$, where R^A, R^B and R^C are each independently selected from H, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, aryl, and substituted aryl.

[0062] As understood by one of ordinary skill in the art, a compound described herein may exist in ionized form, e.g., $-CO_2^-$, $-SO_3^-$ or $-O-SO_3^-$. If a compound contains a positively or negatively charged substituent group, for example, $-SO_3^-$, it may also contain a negatively or positively charged counterion such that the compound as a whole is neutral. In other aspects, the compound may exist in a salt form, where the counterion is provided by a conjugate acid or base.

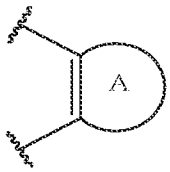
[0063] It is to be understood that certain radical naming conventions can include either a mono-radical or a di-radical, depending on the context. For example, where a substituent

requires two points of attachment to the rest of the molecule, it is understood that the substituent is a di-radical. For example, a substituent identified as alkyl that requires two points of attachment includes di-radicals such as –CH_2–, –CH_2CH_2–, –$CH_2CH(CH_3)CH_2$–, and the like. Other radical naming conventions clearly indicate that the radical is a di-radical such as "alkylene" or "alkenylene."

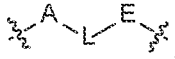
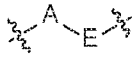
[0064] When two "adjacent" R groups are said to form a ring "together with the atom to which they are attached," it is meant that the collective unit of the atoms, intervening bonds, and the two R groups are the recited ring. For example, when the following substructure is present:



and R^1 and R^2 are defined as selected from the group consisting of hydrogen and alkyl, or R^1 and R^2 together with the atoms to which they are attached form an aryl or carbocyclyl, it is meant that R^1 and R^2 can be selected from hydrogen or alkyl, or alternatively, the substructure has structure:



where A is an aryl ring or a carbocyclyl containing the depicted double bond.

[0065] Wherever a substituent is depicted as a di-radical (*i.e.*, has two points of attachment to the rest of the molecule), it is to be understood that the substituent can be attached in any directional configuration unless otherwise indicated. Thus, for example, a substituent depicted as –AE– or  includes the substituent being oriented such that the **A** is attached at the leftmost attachment point of the molecule as well as the case in which **A** is attached at the rightmost attachment point of the molecule. In addition, if a group or substituent is depicted as , and L is defined an optionally present linker moiety; when L is not present (or absent), such group or substituent is equivalent to .

[0066] As used herein, a "nucleotide" includes a nitrogen containing heterocyclic base, a sugar, and one or more phosphate groups. They are monomeric units of a nucleic acid sequence. In RNA, the sugar is a ribose, and in DNA a deoxyribose, *i.e.* a sugar lacking a hydroxy group that is present in ribose. The nitrogen containing heterocyclic base can be purine, deazapurine, or pyrimidine base. Purine bases include adenine (A) and guanine (G), and modified derivatives or analogs thereof, such as 7-deaza adenine or 7-deaza guanine. Pyrimidine bases include cytosine

(C), thymine (T), and uracil (U), and modified derivatives or analogs thereof. The C-1 atom of deoxyribose is bonded to N-1 of a pyrimidine or N-9 of a purine.

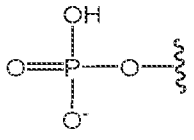
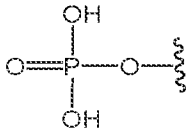
[0067] As used herein, a "nucleoside" is structurally similar to a nucleotide, but is missing the phosphate moieties. An example of a nucleoside analogue would be one in which the label is linked to the base and there is no phosphate group attached to the sugar molecule. The term "nucleoside" is used herein in its ordinary sense as understood by those skilled in the art. Examples include, but are not limited to, a ribonucleoside comprising a ribose moiety and a deoxyribonucleoside comprising a deoxyribose moiety. A modified pentose moiety is a pentose moiety in which an oxygen atom has been replaced with a carbon and/or a carbon has been replaced with a sulfur or an oxygen atom. A "nucleoside" is a monomer that can have a substituted base and/or sugar moiety. Additionally, a nucleoside can be incorporated into larger DNA and/or RNA polymers and oligomers.

[0068] The term "purine base" is used herein in its ordinary sense as understood by those skilled in the art, and includes its tautomers. Similarly, the term "pyrimidine base" is used herein in its ordinary sense as understood by those skilled in the art, and includes its tautomers. A non-limiting list of optionally substituted purine-bases includes purine, adenine, guanine, deazapurine, 7-deaza adenine, 7-deaza guanine. hypoxanthine, xanthine, alloxanthine, 7-alkylguanine (e.g., 7-methylguanine), theobromine, caffeine, uric acid and isoguanine. Examples of pyrimidine bases include, but are not limited to, cytosine, thymine, uracil, 5,6-dihydrouracil and 5-alkylcytosine (e.g., 5-methylcytosine).

[0069] As used herein, when an oligonucleotide or polynucleotide is described as "comprising" a nucleoside or nucleotide described herein, it means that the nucleoside or nucleotide described herein forms a covalent bond with the oligonucleotide or polynucleotide. Similarly, when a nucleoside or nucleotide is described as part of an oligonucleotide or polynucleotide, such as "incorporated into" an oligonucleotide or polynucleotide, it means that the nucleoside or nucleotide described herein forms a covalent bond with the oligonucleotide or polynucleotide. In some such embodiments, the covalent bond is formed between a 3′ hydroxy group of the oligonucleotide or polynucleotide with the 5′ phosphate group of a nucleotide described herein as a phosphodiester bond between the 3′ carbon atom of the oligonucleotide or polynucleotide and the 5′ carbon atom of the nucleotide.

[0070] As used herein, the term "cleavable linker" is not meant to imply that the whole linker is required to be removed. The cleavage site can be located at a position on the linker that ensures that part of the linker remains attached to the detectable label and/or nucleoside or nucleotide moiety after cleavage.

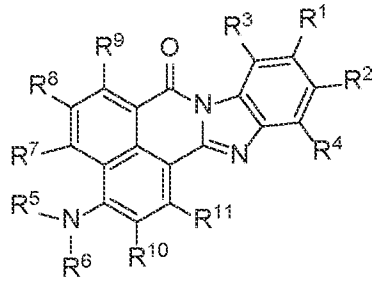
[0071] As used herein, "derivative" or "analog" means a synthetic nucleotide or nucleoside derivative having modified base moieties and/or modified sugar moieties. Such derivatives and analogs are discussed in, e.g., Scheit, *Nucleotide Analogs* (John Wiley & Son, 1980) and Uhlman *et al.*, *Chemical Reviews* 90:543-584, 1990. Nucleotide analogs can also comprise modified phosphodiester linkages, including phosphorothioate, phosphorodithioate, alkyl-phosphonate, phosphoranilidate and phosphoramidate linkages. "Derivative", "analog" and "modified" as used herein, may be used interchangeably, and are encompassed by the terms "nucleotide" and "nucleoside" defined herein.

[0072] As used herein, the term "phosphate" is used in its ordinary sense as understood by those skilled in the art, and includes its protonated forms (for example,  and ). As used herein, the terms "monophosphate," "diphosphate," and "triphosphate" are used in their ordinary sense as understood by those skilled in the art, and include protonated forms.

[0073] As used herein, the term "phasing" refers to a phenomenon in SBS that is caused by incomplete removal of the 3′ terminators and fluorophores, and/or failure to complete the incorporation of a portion of DNA strands within clusters by polymerases at a given sequencing cycle. Prephasing is caused by the incorporation of nucleotides without effective 3′ terminators, wherein the incorporation event goes 1 cycle ahead due to a termination failure. Phasing and prephasing cause the measured signal intensities for a specific cycle to consist of the signal from the current cycle as well as noise from the preceding and following cycles. As the number of cycles increases, the fraction of sequences per cluster affected by phasing and prephasing increases, hampering the identification of the correct base. Prephasing can be caused by the presence of a trace amount of unprotected or unblocked 3′-OH nucleotides during sequencing by synthesis (SBS). The unprotected 3′-OH nucleotides could be generated during the manufacturing processes or possibly during the storage and reagent handling processes. Accordingly, the discovery of nucleotide analogues which decrease the incidence of prephasing is surprising and provides a great advantage in SBS applications over existing nucleotide analogues. For example, the nucleotide analogues provided can result in faster SBS cycle time, lower phasing and prephasing values, and longer sequencing read lengths.

Naphthalimide Dyes of Formula (I)

[0074] Some aspects of the disclosure relate to naphthalimide dyes of Formula (I), and salts and mesomeric forms thereof:



(I),

wherein each of R^1, R^2, R^3, R^4, R^7, R^8, R^9, R^{10}, and R^{11} is independently H, unsubstituted or substituted C_1-C_6 alkyl, C_1-C_6 alkoxy, unsubstituted or substituted C_2-C_6 alkenyl, unsubstituted or substituted C_2-C_6 alkynyl, C_1-C_6 haloalkyl, C_1-C_6 haloalkoxy, C_1-C_6 hydroxyalkyl, (C_1-C_6 alkoxy)C_1-C_6 alkyl, -O-(C_1-C_6 alkoxy)C_1-C_6 alkyl, unsubstituted or substituted amino, halo, cyano, carboxy, hydroxy, nitro, sulfonyl, sulfino, sulfo, sulfonate, S-sulfonamido, N-sulfonamido,–C(=O)OR^A, –C(=O)NR^BR^C or –SO_3R^D;

each of R^A, R^B, R^C, and R^D is independently H, or unsubstituted or substituted C_1-C_6 alkyl; and

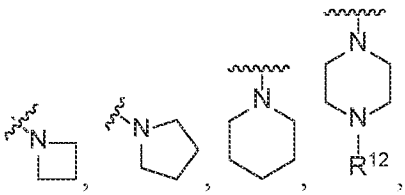
each of R^5 and R^6 is independently H, unsubstituted or substituted C_1-C_6 alkyl; unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

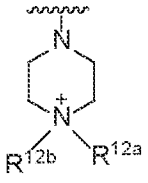
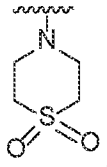
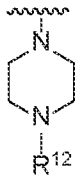
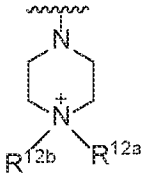
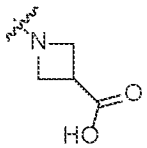
alternatively, R^5 and R^6 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl. In some embodiments, the compound of Formula (I) comprises one carboxy group. In some further embodiments, the compound of Formula (I) comprises at least one carboxy group.

[0075] In some embodiments of the compound of Formula (I), each of R^1 and R^2 is independently H, unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, t-butyl, n-pentyl, or n-hexyl), C_1-C_6 alkoxy, –C(=O)OR^A, or –SO_3R^D. In some embodiments, R^1 is H, methoxy, –C(=O)OH or –SO_3H, and R^2 is H or methoxy. In other embodiments, R^2 is H, methoxy, –C(=O)OH or –SO_3H, and R^1 is H or methoxy. In some embodiments, each of R^1 and R^2 is H. In some embodiments, each of R^1 and R^2 is C_1-C_6 alkoxy. In one embodiment, each of R^1 and R^2 is methoxy. In some embodiments, R^1 is –C(=O)OH and R^2 is H. In other embodiments, R^1 is –SO_3H and R^2 is H.

[0076] In some embodiments of the compound of Formula (I), each of R^3 and R^4 is H.

[0077] In some embodiments of the compound of Formula (I), each of R^5 and R^6 is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some embodiments, each of R^5 and R^6 is H or unsubstituted C_1-C_6 alkyl. In some embodiments, each of R^5 and R^6 is C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some other embodiments of the compound of Formula (I), R^5 is H or unsubstituted C_1-C_6 alkyl, and R^6 is C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some such embodiments, R^6 is C_2, C_3, C_4, or C_5 alkyl (e.g., ethyl, n-propyl, isopropyl, n-butyl, sec-butyl, t-butyl, or n-pentyl) substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some other embodiments of the compound of Formula (I), R^5 is H, and R^6 is C_2, C_3, C_4, or C_5 alkyl (e.g., ethyl, n-propyl, isopropyl, n-butyl, sec-butyl, t-butyl, or n-pentyl) substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In other embodiments, R^5 is H, and R^6 is unsubstituted C_3-C_7 cycloalkyl, unsubstituted phenyl, or C_3-C_7 cycloalkyl or phenyl each independently substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some further embodiments, R^5 is H, and R^6 is cyclobutyl, cyclopentyl, or cyclohexyl substituted with –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some such embodiments, each of R^A, R^B, R^C, and R^D is H. In other such embodiments, R^B is H and R^C is unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of $–C(=O)OH$ and $–SO_3H$. In some specific embodiments, R^6 is cyclohexyl substituted with -C(=O)OH.

[0078] In some other embodiments of the compound of Formula (I), R^5 and R^6 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl (for example, 4, 5, 6, or 7 membered heterocyclyl containing one, two, three or four heteroatoms selected from N, O and S). In some such embodiments, R^5 and R^6 together with the nitrogen atom to which they are attached form ,

, or , each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D, wherein each of R^{12}, R^{12a} and R^{12b} is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In one embodiment, R^5 and R^6 together with the nitrogen atom to which they are attached form , wherein R^{12} is unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, or t-butyl). In another embodiment, R^5 and R^6 together with the nitrogen atom to which they are attached form , wherein each of R^{12a} and R^{12b} is independently unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, or t-butyl). In one embodiment, both R^{12a} and R^{12b} are methyl. In other embodiments, R^{12a} is unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, or t-butyl) and R^{12b} is C_1-C_6 alkyl (such as C_2, C_3, C_4 or C_5 alkyl) substituted with -C(=O)OR^A (e.g., -C(O)OH), –C(=O)NR^BR^C or –SO_3R^D (e.g., -SO_3H). In another embodiment, R^5 and R^6 together with the nitrogen atom to which they are attached form  substituted with–C(=O)OH. In some specific embodiments, R^5 and R^6 together with the nitrogen atom to which they are attached form 

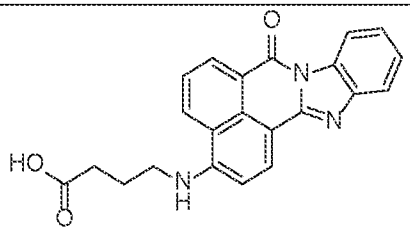
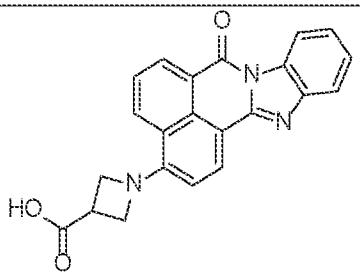
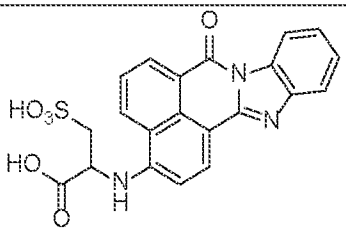
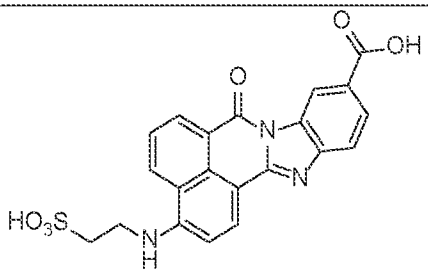
[0079] In some embodiments of the compound of Formula (I), each of R^A, R^B, R^C, and R^D is H. In other such embodiments, R^B is H and R^C is unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –C(=O)OH and –SO_3H.

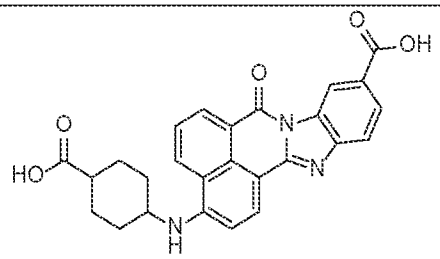
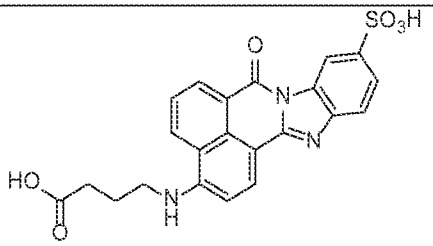
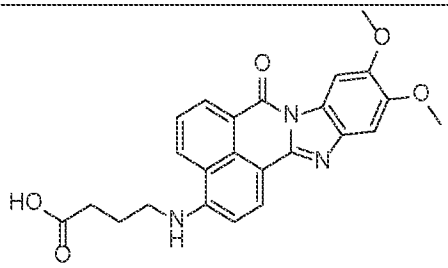
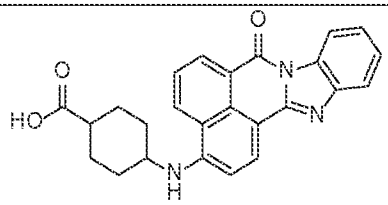
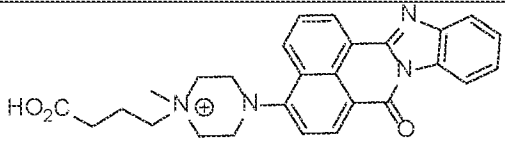
[0080] In some embodiments of the compound of Formula (I), R^7 is H. In some embodiments, R^8 is H. In some embodiments, R^9 is H. In some embodiments, R^{10} is H. In some embodiments, R^{11} is H. In some further embodiments, each of R^7, R^8, R^9, R^{10}, and R^{11} is H.

[0081] In any embodiments of the compound of Formula (I), when the compound of Formula (I) includes one or more $-SO_3H$ groups, the $-SO_3H$ may be in the ionized form (i.e., $-SO_3^-$), and the compound may be in a salt form containing one or more counterions (i.e., cations) such that the compound as a whole is neutral and not bear any charges.

[0082] Additional embodiments of the compound of Formula (I) include but not limited to those described in Table A below, and salts, esters, and mesomeric forms thereof. Non-limiting examples of esters include corresponding C_1-C_6 alkyl carboxylic esters (such as methyl esters, ethyl esters isopropyl esters, and t-butyl esters formed from the carboxylic group of the compounds).

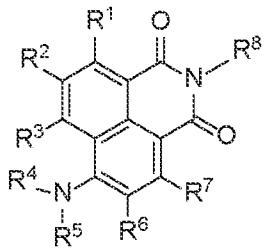
Table A.

Compd. #	Structure
I-1	
I-2	
I-3	
I-4	

I-5	
I-6	
I-7	
I-8	
I-9	

Naphthalimide Dyes of Formula (II)

[0083] Some aspects of the disclosure relate to naphthalimide dyes of Formula (II), and salts and mesomeric forms thereof:



(II),

wherein each of R^1, R^2, R^3, R^6, and R^7 is independently H, unsubstituted or substituted C_1-C_6 alkyl, C_1-C_6 alkoxy, unsubstituted or substituted C_2-C_6 alkenyl, unsubstituted or substituted C_2-C_6 alkynyl, C_1-C_6 haloalkyl, C_1-C_6 haloalkoxy, C_1-C_6 hydroxyalkyl, (C_1-C_6 alkoxy)C_1-C_6 alkyl, -O-(C_1-C_6 alkoxy)C_1-C_6 alkyl, unsubstituted or substituted amino, halo, cyano, carboxy, hydroxy, nitro, sulfonyl, sulfino, sulfo, sulfonate, S-sulfonamido, N-sulfonamido, –C(=O)OR^A, –C(=O)NR^BR^C or –SO_3R^D;

R^8 is unsubstituted or substituted C_1-C_6 alkyl, unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

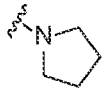
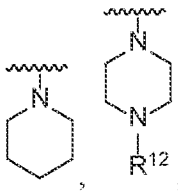
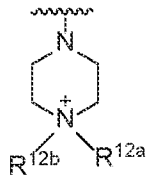
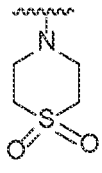
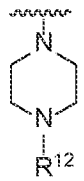
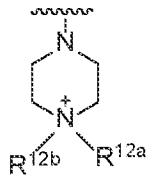
each of R^A, R^B, R^C, and R^D is independently H, or unsubstituted or substituted C_1-C_6 alkyl; and

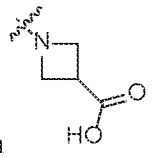
each of R^4 and R^5 is independently H, unsubstituted or substituted C_1-C_6 alkyl; unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

alternatively, R^4 and R^5 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl. In some embodiments, the compound of Formula (II) comprises one carboxy group. In some further embodiments, the compound of Formula (II) comprises at least one carboxy group.

[0084] In some embodiments of the compound of Formula (II), each of R^1, R^2, R^3, R^6, and R^7 is independently H, unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, t-butyl, n-pentyl, or n-hexyl), C_1-C_6 alkoxy, –C(=O)OR^A, or –SO_3R^D. In some such embodiments, each of R^1, R^2, R^3, R^6, and R^7 is H.

[0085] In some embodiments of the compound of Formula (II), each of R^4 and R^5 is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In some such embodiments, both R^4 and R^5 are H or unsubstituted C_1-C_6 alkyl. In some other embodiments, both R^4 and R^5 are C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In some other embodiments, R^4 is H, and R^5 is C_2, C_3, C_4, or C_5 alkyl (e.g., ethyl, n-propyl, isopropyl, n-butyl, sec-butyl, t-butyl, or n-pentyl) substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In other embodiments, R^4 is H, and R^5 is unsubstituted C_3-C_7 cycloalkyl, unsubstituted phenyl, or C_3-C_7 cycloalkyl or phenyl each independently substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano,

C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In some further embodiments, R^4 is H, and R^5 is cyclobutyl, cyclopentyl, or cyclohexyl substituted with –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In some such embodiments, each of R^A, R^B, R^C, and R^D is H. In other such embodiments, R^B is H and R^C is unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –C(=O)OH and –SO_3H. In some specific embodiments, R^5 is cyclohexyl substituted with –C(=O)OH. In some embodiments, R^4 and R^5 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl (for example, 4, 5, 6, or 7 membered heterocyclyl containing one, two, three or four heteroatoms selected from N, O and S). In some such embodiments, R^4 and R^5 together with the nitrogen atom to which they are attached form , , or , , or , each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, C_1-C_6 alkoxy, halo (e.g., fluoro or chloro), cyano, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D, wherein each of R^{12}, R^{12a} and R^{12b} is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D. In one embodiment, R^4 and R^5 together with the nitrogen atom to which they are attached form , wherein R^{12} is unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, or t-butyl). In another embodiment, R^5 and R^6 together with the nitrogen atom to which they are attached form , wherein each of R^{12a} and R^{12b} is independently unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, or t-butyl). In one embodiment, both R^{12a} and R^{12b} are methyl. In other embodiments, R^{12a} is unsubstituted C_1-C_6 alkyl (e.g., methyl, ethyl, n-propyl, isopropyl, n-butyl, or t-butyl) and R^{12b} is C_1-C_6 alkyl (such as C_2, C_3, C_4 or C_5 alkyl) substituted with -C(=O)OR^A (e.g., -C(O)OH), –C(=O)NR^BR^C or –SO_3R^D (e.g., -SO_3H). In another embodiment, R^4 and R^5 together with the nitrogen atom to which they are attached form 

substituted with –C(=O)OH. In some specific embodiments, R^5 and R^6 together with the nitrogen atom to which they are attached form .

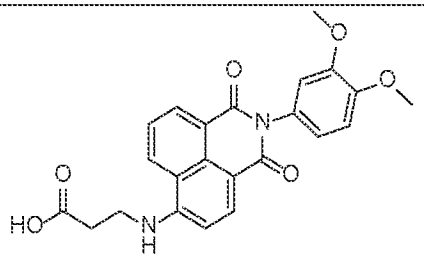
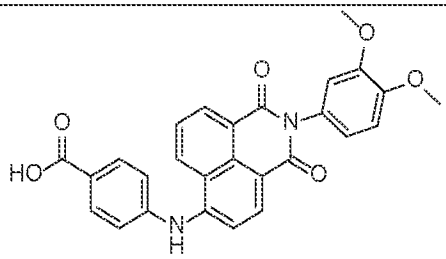
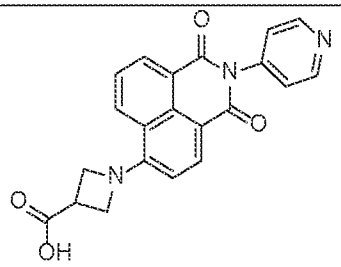
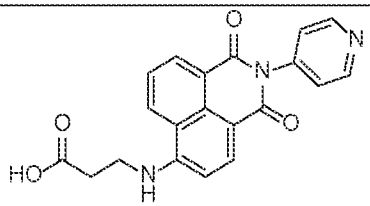
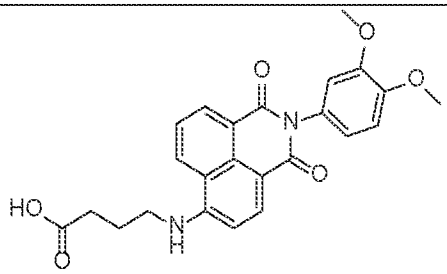
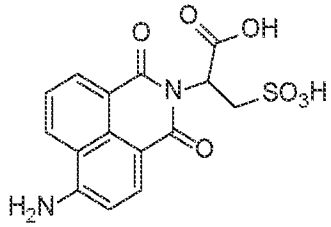
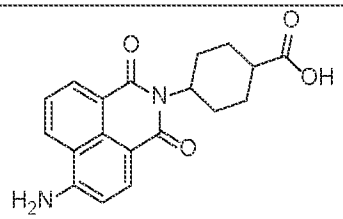
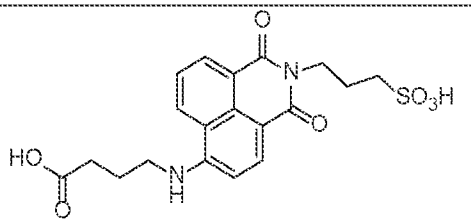
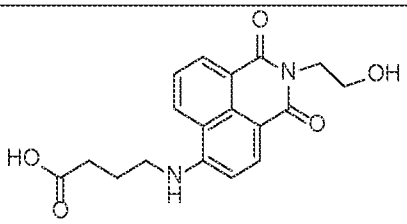
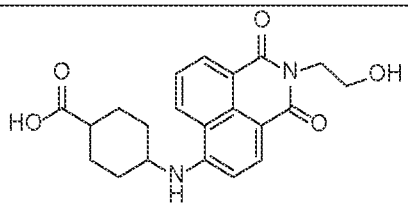
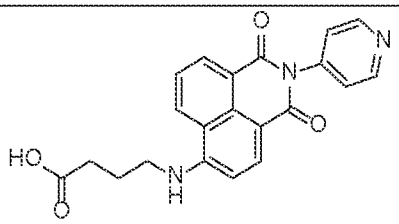
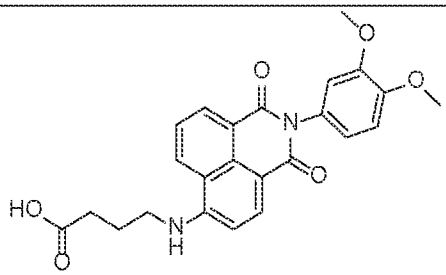
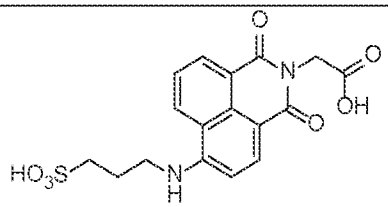
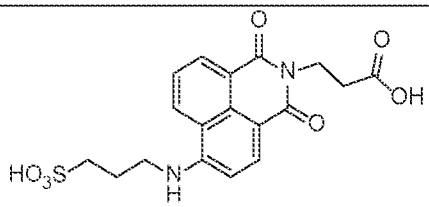
[0086] In some embodiments of the compound of Formula (II), R^8 is C_1-C_6 alkyl unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some such embodiments, R^8 is C_2, C_3, C_4, or C_5 alkyl (e.g., ethyl, n-propyl, isopropyl, n-butyl, sec-butyl, t-butyl, or n-pentyl) substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some embodiments, R^8 is $–CH_2CH_2OH$. In other embodiments, R^8 is ethyl, n-propyl, n-butyl, or n-pentyl substituted with one or more substituents selected from –C(=O)OH and $–SO_3H$. In some specific embodiments, R^8 is $-CH(COOH)CH_2SO_3H$. In some embodiments, R^8 is phenyl, pyridyl, cyclopentyl, or cyclohexyl, each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$. In some such embodiments, R^8 is cyclohexyl substituted with –C(=O)OH. In other such embodiments, R^8 is pyridyl. In yet other such embodiments, R^8 is dimethoxyphenyl. In some other embodiments, R^8 is (phenyl)C_1-C_6 alkyl, (pyridyl)C_1-C_6 alkyl, (cyclopentyl)C_1-C_6 alkyl, or (cyclohexyl)C_1-C_6 alkyl, each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo (e.g., fluoro or chloro), cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$.

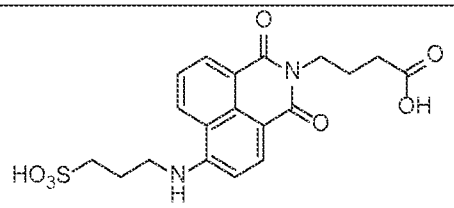
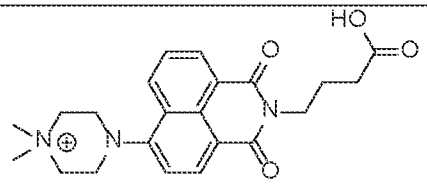
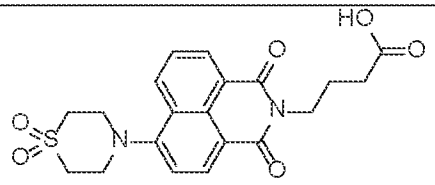
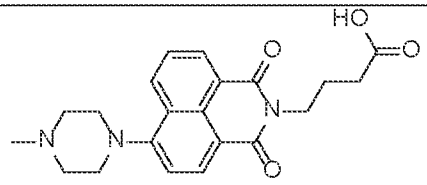
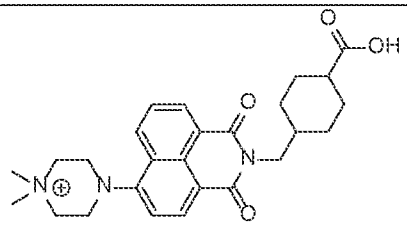
[0087] In some embodiments of the compound of Formula (II), each of R^A, R^B, R^C, and R^D is H. In other embodiments, R^B is H and R^C is unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –C(=O)OH and $–SO_3H$.

[0088] In any embodiments of the compound of Formula (II), when the compound of Formula (II) includes one or more $-SO_3H$ groups, the $-SO_3H$ may be in the ionized form (i.e., $-SO_3^-$), and the compound may be in a salt form containing one or more counterions (i.e., cations) such that the compound as a whole is neutral and not bear any charges.

[0089] Additional embodiments of the compound of Formula (II) include but not limited to those described in Table B below.

Table B.

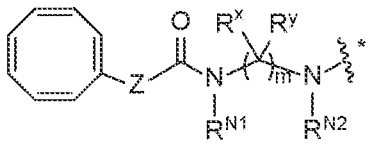
Compd. #	Structure	Compd. #	Structure
II-1		II-2	
II-3		II-4	
II-5		II-6	
II-7		II-8	
II-9		II-10	
II-11		II-12	
II-13		II-14	

II-15		II-16	
II-17		II-18	
II-19			

and salts, esters and mesomeric forms thereof. Non-limiting examples of the esters include corresponding C_1-C_6 alkyl carboxylic esters (such as methyl esters, ethyl esters isopropyl esters, and t-butyl esters formed from the carboxylic group of the compounds).

Photo-protecting Moieties

[0090] In any embodiments of the naphthalimide compounds described herein, the compound may be further modified to introduce a photo-protecting moiety covalently bonded thereto, for example, a photo-protecting cyclooctatetraene (COT) moiety as described in U.S. Pub. No. 2021/0155983 A1, which is incorporated by reference in its entirety. For example, the naphthalimide compound of Formula (I) or Formula (II) may be covalently attached to a COT moiety via an amide bond, wherein the photo-protecting COT moiety comprises the structure of Formula (III):



(III)

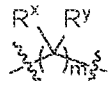
wherein Z is absent, optionally substituted C_{2-6} alkenylene, or optionally substituted C_{2-6} alkynylene;

each R^X and R^Y is independently H, carboxyl, carboxylate, amino, sulfo, sulfonate, $-C(O)OR^A$, or $-C(O)NR^BR^C$, or C_1-C_6 alkyl substituted with amino, carboxyl, carboxylate, sulfo, sulfonate, $-C(O)OR^A$, or $-C(O)NR^BR^C$,

each R^{N1} and R^{N2} is independently H or C_1-C_6 alkyl substituted with amino, carboxyl, carboxylate, sulfo, sulfonate, $-C(O)OR^A$, or $-C(O)NR^BR^C$;

R^A is optionally substituted C_1-C_6 alkyl, optionally substituted C_6-C_{10} aryl, optionally substituted 5 to 10 membered heteroaryl, or optionally substituted C_3-C_7 cycloalkyl;

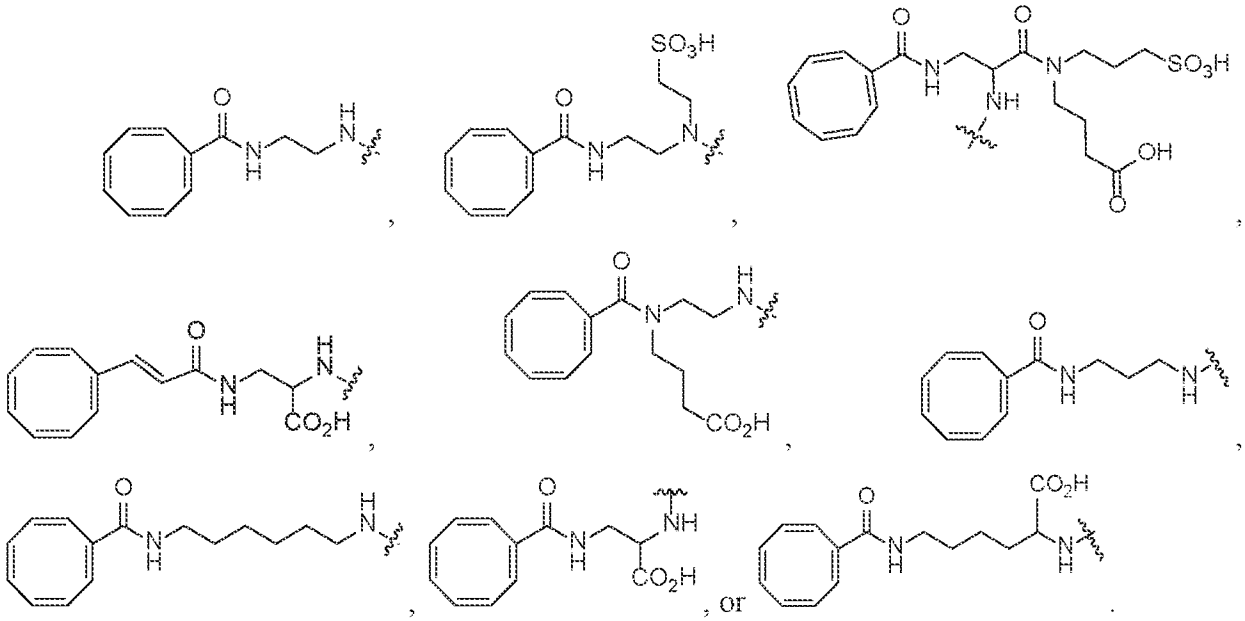
each of R^B and R^C is independently H, optionally substituted C_1-C_6 alkyl, optionally substituted C_6-C_{10} aryl, optionally substituted 5 to 10 membered heteroaryl, or optionally substituted C_3-C_7 cycloalkyl;

the carbon atom to which R^X and R^Y are attached in  is optionally replaced with O, S, or N, provided that when said carbon atom is replaced with O or S, then R^X and R^Y are both absent; when said carbon atom is replaced with N, then R^Y is absent; and

m is an integral number between 1 and 10;

wherein the asterisk * indicates the attachment point of the photo-protecting cyclooctatetraene moiety to the carbonyl group of amide bond formed by the reaction of an amino group of the photo-protecting cyclooctatetraene moiety with the carboxyl group of the compound of Formula (I) or Formula (II) as described herein.

[0091] In some further embodiments, the photo-protecting COT moiety may comprise the following structures:



[0092] The COT moiety may be covalently attached to the naphthalimide dye described herein by reaction between a functional group of the naphthalimide dye described herein (e.g., a carboxyl group) and an amino group of a COT derivative to form an amide bond (where the carbonyl group of the amide bond is not shown). Alternatively, the naphthalimide dye described herein may have an amino group (instead of a carboxyl group) that forms the amide bond with a carboxyl group of the COT derivative.

[0093] Additional photoprotective moieties for use as triplet state quencher (TSQ) or triplet-triplet annihilator (TTA) including 1,4-substituted COT moiety, vinyl-fluorene moieties, or stilbene moieties are described in U.S. Ser. No. 18/476911, which is also incorporated by reference in its entirety.

Labeled Nucleotides

[0094] According to an aspect of the disclosure, dye compounds described herein are suitable for attachment to substrate moieties, particularly comprising linker groups to enable attachment to substrate moieties. Substrate moieties can be virtually any molecule or substance to which the dyes of the disclosure can be conjugated, and, by way of non-limiting example, may include nucleosides, nucleotides, polynucleotides, carbohydrates, ligands, particles, solid surfaces, organic and inorganic polymers, chromosomes, nuclei, living cells, and combinations or assemblages thereof. The dyes can be conjugated by an optional linker by a variety of means including hydrophobic attraction, ionic attraction, and covalent attachment. In some aspect, the dyes are conjugated to the substrate by covalent attachment. More particularly, the covalent attachment is by means of a linker group. In some instances, such labeled nucleotides are also referred to as "modified nucleotides."

[0095] Some aspects of the present disclosure relate to a nucleotide labeled with a dye of Formula (I) or Formula (II), or a salt or mesomeric form thereof as described herein, or a derivative thereof containing a photo-protecting moiety (e.g., COT) as described herein. The labeled nucleotide or oligonucleotide may be attached to the dye compound disclosed herein via a carboxyl (-CO_2H) or an alkyl-carboxyl group to form an amide or alkyl-amide bond. In some further embodiments, the carboxyl group may be in the form of an activated form of carboxyl group, for example, an amide or ester, which may be used for attachment to an amino or hydroxy group of the nucleotide or oligonucleotide The term "activated ester" as used herein, refers to a carboxyl group derivative which is capable of reacting in mild conditions, for example, with a compound containing an amino group. Non-limiting examples of activated esters include but not limited to p-nitrophenyl, pentafluorophenyl and succinimidyl esters.

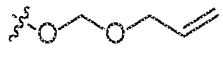
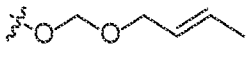
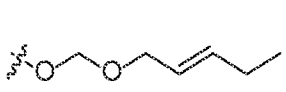
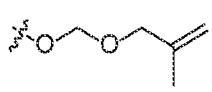
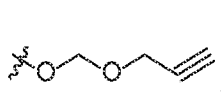
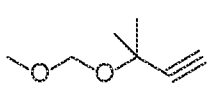
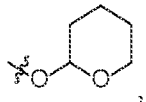
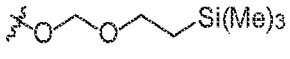
[0096] For example, the dye compound of Formula (I) or Formula (II) may be attached to the nucleotide via one of R^1, R^2, R^3, R^4, R^5, R^6, R^7, R^8, R^9, R^{10}, R^{11}, R^{12}, R^{12a} and R^{12b} of Formula (I) or R^1, R^2, R^3, R^4, R^5, R^6, R^7, R^8, R^{12}, R^{12a} and R^{12b} of Formula (II). In some such embodiments, one of R^1, R^2, R^3, R^4, R^5, R^6, R^7, R^8, R^9, R^{10}, R^{11}, R^{12}, R^{12a} and R^{12b} of Formula (I) or one of R^1, R^2, R^3, R^4, R^5, R^6, R^7, R^8, R^{12}, R^{12a} and R^{12b} of Formula (II) comprises a carboxyl group and the attachment forms an amide moiety between the carboxyl functional group of the compound of Formula (I) or Formula (II) and the amino functional group of a nucleotide or a nucleotide linker.

In some further embodiments, one of R^1, R^2, R^5, R^6, or R^{12}, R^{12a} and R^{12b} of Formula (I) comprises a carboxyl functional group. In some further embodiments, one of R^4, R^5, R^8, R^{12}, R^{12a} and R^{12b} of Formula (II) comprises a carboxyl functional group.

[0097] In some embodiments, the dye compound may be covalently attached to the nucleotide via the nucleotide base. In some such embodiments, the labeled nucleotide may have the dye attached to the C5 position of a pyrimidine base or the C7 position of a 7-deaza purine base, optionally through a linker moiety. For example, the nucleobase may be 7-deaza adenine and the dye is attached to the 7-deaza adenine at the C7 position, optionally through a linker. The nucleobase may be 7-deaza guanine and the dye is attached to the 7-deaza guanine at the C7 position, optionally through a linker. The nucleobase may be cytosine and the dye is attached to the cytosine at the C5 position, optionally through a linker. As another example, the nucleobase may be thymine or uracil and the dye is attached to the thymine or uracil at the C5 position, optionally through a linker.

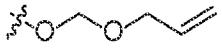
3′ Blocking Groups

[0098] The labeled nucleotide may also have a blocking group covalently attached to the ribose or deoxyribose sugar of the nucleotide. The blocking group may be attached at any position on the ribose or deoxyribose sugar. In particular embodiments, the blocking group is at the 3′ OH position of the ribose or deoxyribose sugar of the nucleotide. Various 3' OH blocking group are disclosed in WO2004/018497 and WO2014/139596, which are hereby incorporated by references. For example, the blocking group may be azidomethyl (-CH_2N_3) or substituted azidomethyl (e.g., -$CH(CHF_2)N_3$ or $CH(CH_2F)N_3$), or allyl connecting to the 3' oxygen atom of the ribose or deoxyribose moiety. In some embodiments, the 3' blocking group is azidomethyl, forming 3′-OCH_2N_3 with the 3′ carbon of the ribose or deoxyribose.

[0099] Additional 3′ blocking groups are disclosed in U.S. Publication No. 2020/0216891 A1, which is incorporated by reference in its entirety. Non-limiting examples of the 3′ blocking group include:  (AOM), , , , , , , and , each covalently attached to the 3′ carbon of the ribose or deoxyribose.

Deprotection of the 3′ Blocking Groups

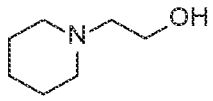
[0100] In some embodiments, the 3′ blocking group may be removed or deprotected by a chemical reagent to generate a free hydroxy group, for example, in the presence of a water

soluble phosphine reagent. Non-limiting examples include tris(hydroxymethyl)phosphine (THMP), tris(hydroxyethyl)phosphine (THEP) or tris(hydroxylpropyl)phosphine (THP or THPP). 3′-acetal blocking groups described herein may be removed or cleaved under various chemical conditions. For 3′ acetal blocking groups such as , non-limiting cleaving condition includes a Pd(II) complex, such as $Pd(OAc)_2$ or allylPd(II) chloride dimer, in the presence of a phosphine ligand, for example tris(hydroxymethyl)phosphine (THMP), or tris(hydroxylpropyl)phosphine (THP or THPP). For those blocking groups containing an alkynyl group (e.g., an ethynyl), they may also be removed by a Pd(II) complex (e.g., $Pd(OAc)_2$ or allyl Pd(II) chloride dimer) in the presence of a phosphine ligand (e.g., THP or THMP).

Palladium Cleavage Reagents

[0101] In some other embodiments, the 3' blocking group described herein such as allyl or AOM may be cleaved by a palladium catalyst. In some such embodiments, the Pd catalyst is water soluble. In some such embodiments, is a Pd(0) complex (e.g., Tris(3,3′,3″-phosphinidynetris(benzenesulfonato)palladium(0) nonasodium salt nonahydrate). In some instances, the Pd(0) complex may be generated *in situ* from reduction of a Pd(II) complex by reagents such as alkenes, alcohols, amines, phosphines, or metal hydrides. Suitable palladium sources include Na_2PdCl_4, Li_2PdCl_4, $Pd(CH_3CN)_2Cl_2$, $(PdCl(C_3H_5))_2$, $[Pd(C_3H_5)(THP)]Cl$, $[Pd(C_3H_5)(THP)_2]Cl$, $Pd(OAc)_2$, $Pd(Ph_3)_4$, $Pd(dba)_2$, $Pd(Acac)_2$, $PdCl_2(COD)$, $Pd(TFA)_2$, Na_2PdBr_4, K_2PdBr_4, $PdCl_2$, $PdBr_2$, and $Pd(NO_3)_2$. In one such embodiment, the Pd(0) complex is generated *in situ* from Na_2PdCl_4 or K_2PdCl_4. In another embodiment, the palladium source is allyl palladium(II) chloride dimer $[(PdCl(C_3H_5))_2]$. In some embodiments, the Pd(0) complex is generated in an aqueous solution by mixing a Pd(II) complex with a phosphine. Suitable phosphines include water soluble phosphines, such as THP, THMP, PTA, TCEP, bis(p-sulfonatophenyl)phenylphosphine dihydrate potassium salt, or triphenylphosphine-3,3',3''-trisulfonic acid trisodium salt.

[0102] In some embodiments, the palladium catalyst is prepared by mixing $[(Allyl)PdCl]_2$ with THP *in situ*. The molar ratio of $[(Allyl)PdCl]_2$ and the THP may be about 1:1, 1:1.5, 1:2, 1:2.5, 1:3, 1:3.5, 1:4, 1:4.5, 1:5, 1:5.5, 1:6, 1:6.5, 1:7, 1:7.5, 1:8, 1:8.5, 1:9, 1:9.5 or 1:10. In one embodiment, the molar ratio of $[(Allyl)PdCl]_2$ to THP is 1:10. In some other embodiment, the palladium catalyst is prepared by mixing a water soluble Pd reagent such as Na_2PdCl_4 or K_2PdCl_4 with THP *in situ*. The molar ratio of Na_2PdCl_4 or K_2PdCl_4 and THP may be about 1:1, 1:1.5, 1:2, 1:2.5, 1:3, 1:3.5, 1:4, 1:4.5, 1:5, 1:5.5, 1:6, 1:6.5, 1:7, 1:7.5, 1:8, 1:8.5, 1:9, 1:9.5 or 1:10. In one embodiment, the molar ratio of Na_2PdCl_4 or K_2PdCl_4 to THP is about 1:3. In another embodiment, the molar ratio of Na_2PdCl_4 or K_2PdCl_4 to THP is about 1:3.5. In yet another

embodiment, the molar ratio of Na_2PdCl_4 or K_2PdCl_4 to THP is about 1:2.5. In some further embodiments, one or more reducing agents may be added, such as ascorbic acid or a salt thereof (e.g., sodium ascorbate). In some embodiments, the cleavage mixture may contain additional buffer reagents, such as a primary amine, a secondary amine, a tertiary amine, a carbonate salt, a phosphate salt, or a borate salt, or combinations thereof. In some further embodiments, the buffer reagent comprises ethanolamine (EA), tris(hydroxymethyl)aminomethane (Tris), glycine, sodium carbonate, sodium phosphate, sodium borate, 2-dimethylethanolamine (DMEA), 2-diethylethanolamine (DEEA), N,N,N′,N′-tetramethylethylenediamine (TEMED), N,N,N′,N′-tetraethylethylenediamine (TEEDA), or 2-piperidine ethanol (also known as (2-hydroxyethyl)piperidine, having the structure ), or combinations thereof. In one embodiment, the buffer reagent comprises or is DEEA. In another embodiment, the buffer reagent comprises or is (2-hydroxyethyl)piperidine. In another embodiment, the buffer reagent contains one or more inorganic salts such as a carbonate salt, a phosphate salt, or a borate salt, or combinations thereof. In one embodiment, the inorganic salt is a sodium salt.

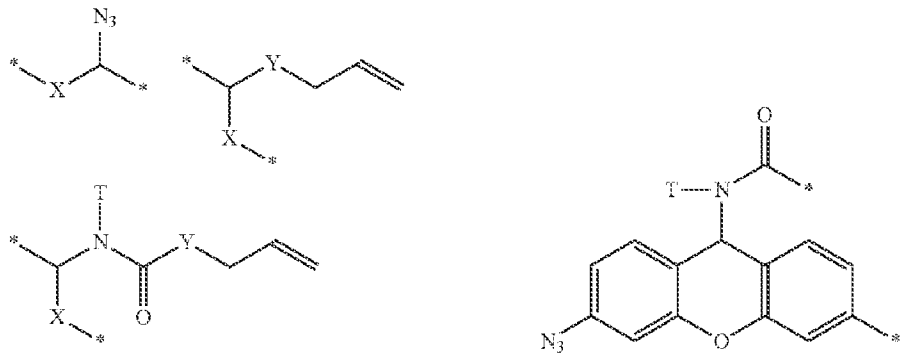
Cleavable Linkers

[0103] The dye compounds as disclosed herein may include a reactive linker group at one of the substituent positions for covalent attachment of the compound to a substrate or another molecule. Reactive linking groups are moieties capable of forming a bond (e.g., a covalent or non-covalent bond), in particular a covalent bond. In a particular embodiment the linker may be a cleavable linker. Use of the term "cleavable linker" is not meant to imply that the whole linker is required to be removed. The cleavage site can be located at a position on the linker that ensures that part of the linker remains attached to the dye and/or substrate moiety after cleavage. Cleavable linkers may be, by way of non-limiting example, electrophilically cleavable linkers, nucleophilically cleavable linkers, photocleavable linkers, cleavable under reductive conditions (for example disulfide or azide containing linkers), oxidative conditions, cleavable via use of safety-catch linkers and cleavable by elimination mechanisms. The use of a cleavable linker to attach the dye compound to a substrate moiety ensures that the label can, if required, be removed after detection, avoiding any interfering signal in downstream steps.

[0104] Useful linker groups may be found in PCT Publication No. WO 2004/018493 (herein incorporated by reference), examples of which include linkers that may be cleaved using water-soluble phosphines or water-soluble transition metal catalysts formed from a transition metal and at least partially water-soluble ligands. In aqueous solution the latter form at least

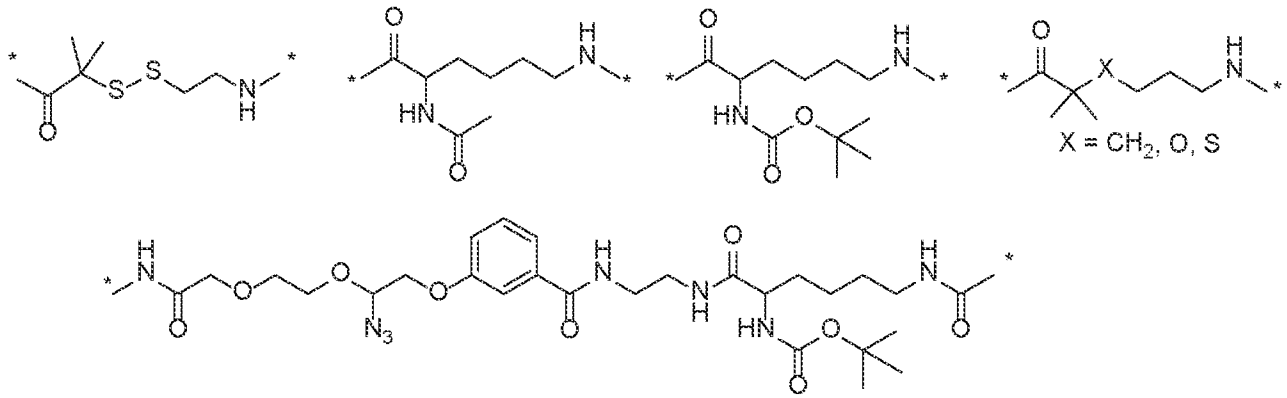
partially water-soluble transition metal complexes. Such cleavable linkers can be used to connect bases of nucleotides to labels such as the dyes set forth herein.

[0105] Particular linkers include those disclosed in PCT Publication No. WO 2004/018493 (herein incorporated by reference) such as those that include moieties of the formulae:



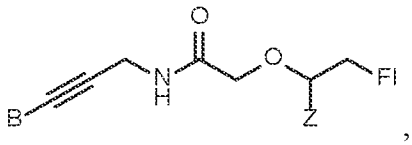
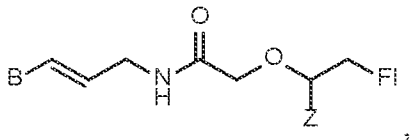
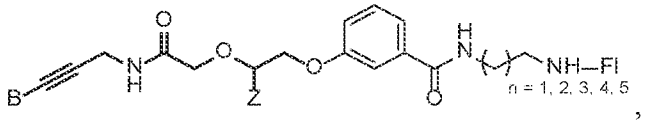
(wherein X is selected from the group comprising O, S, NH and NQ wherein Q is a C1-10 substituted or unsubstituted alkyl group, Y is selected from the group comprising O, S, NH and N(allyl), T is hydrogen or a C_1-C_{10} substituted or unsubstituted alkyl group and * indicates where the moiety is connected to the remainder of the nucleotide or nucleoside). In some aspect, the linkers connect the bases of nucleotides to labels such as, for example, the dye compounds described herein.

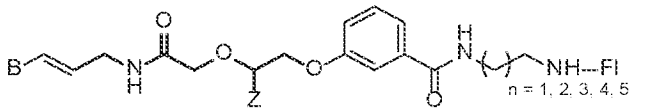
[0106] Additional examples of linkers include those disclosed in U.S. Publication No. 2016/0040225 (herein incorporated by reference), such as those include moieties of the formulae:

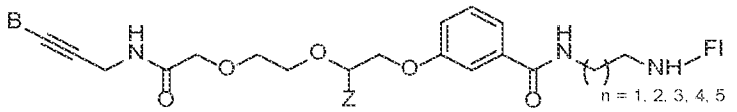


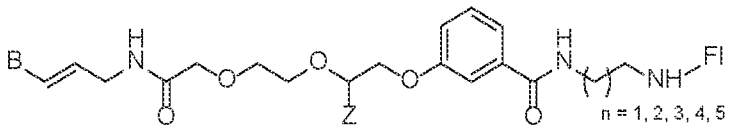
(wherein * indicates where the moiety is connected to the remainder of the nucleotide or nucleoside). The linker moieties illustrated herein may comprise the whole or partial linker structure between the nucleotides/nucleosides and the labels. The linker moieties illustrated herein may comprise the whole or partial linker structure between the nucleotides/nucleosides and the labels.

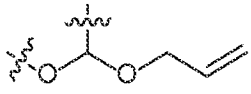
[0107] Additional examples of linkers include moieties of the formula:







or wherein B is a nucleobase; Z is $-N_3$ (azido), $-O-C_1-C_6$ alkyl, $-O-C_2-C_6$ alkenyl, or $-O-C_2-C_6$ alkynyl; and Fl comprises a dye moiety, which may contain additional linker structure. One of ordinary skill in the art understands that the dye compound described herein is covalently bounded to the linker by reacting a functional group of the dye compound (e.g., carboxyl) with a functional group of the linker (e.g., amino). In one embodiment, the cleavable linker comprises  ("AOL" linker moiety) where Z is –O-allyl.

[0108] In particular embodiments, the length of the linker between a fluorescent dye (fluorophore) and a guanine base can be altered, for example, by introducing a polyethylene glycol spacer group, thereby increasing the fluorescence intensity compared to the same fluorophore attached to the guanine base through other linkages known in the art. Exemplary linkers and their properties are set forth in PCT Publication No. WO 2007/020457 (herein incorporated by reference). The design of linkers, and especially their increased length, can allow improvements in the brightness of fluorophores attached to the guanine bases of guanosine nucleotides when incorporated into polynucleotides such as DNA. Thus, when the dye is for use in any method of analysis which requires detection of a fluorescent dye label attached to a guanine-containing nucleotide, it is advantageous if the linker comprises a spacer group of formula $-((CH_2)_2O)_n-$, wherein n is an integer between 2 and 50, as described in WO 2007/020457.

[0109] Nucleosides and nucleotides may be labeled at sites on the sugar or nucleobase. As known in the art, a "nucleotide" consists of a nitrogenous base, a sugar, and one or more phosphate groups. In RNA, the sugar is ribose and in DNA is a deoxyribose, i.e., a sugar lacking

a hydroxy group that is present in ribose. The nitrogenous base is a derivative of purine or pyrimidine. The purines are adenine (A) and guanine (G), and the pyrimidines are cytosine (C) and thymine (T) or in the context of RNA, uracil (U). The C-1 atom of deoxyribose is bonded to N-1 of a pyrimidine or N-9 of a purine. A nucleotide is also a phosphate ester of a nucleoside, with esterification occurring on the hydroxy group attached to the C-3 or C-5 of the sugar. Nucleotides are usually mono, di- or triphosphates.

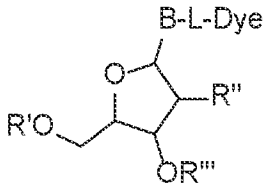
[0110] A "nucleoside" is structurally similar to a nucleotide but is missing the phosphate moieties. An example of a nucleoside analog would be one in which the label is linked to the base and there is no phosphate group attached to the sugar molecule.

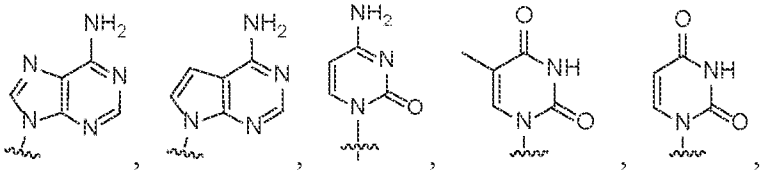
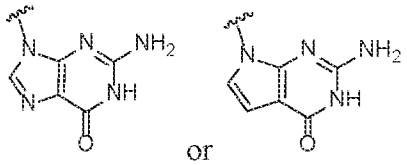
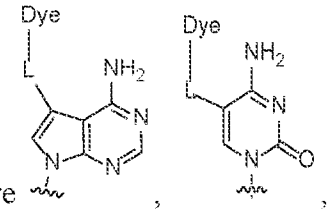
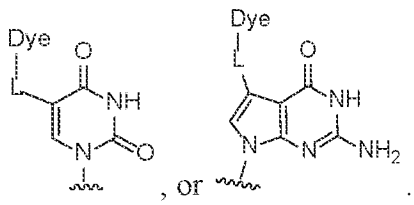
[0111] Although the base is usually referred to as a purine or pyrimidine, the skilled person will appreciate that derivatives and analogues are available which do not alter the capability of the nucleotide or nucleoside to undergo Watson-Crick base pairing. "Derivative" or "analogue" means a compound or molecule whose core structure is the same as, or closely resembles that of a parent compound but which has a chemical or physical modification, such as, for example, a different or additional side group, which allows the derivative nucleotide or nucleoside to be linked to another molecule. For example, the base may be a deazapurine. In particular embodiments, the derivatives should be capable of undergoing Watson-Crick pairing. "Derivative" and "analogue" also include, for example, a synthetic nucleotide or nucleoside derivative having modified base moieties and/or modified sugar moieties. Such derivatives and analogues are discussed in, for example, Scheit, *Nucleotide analogs* (John Wiley & Son, 1980) and Uhlman *et al.*, *Chemical Reviews* 90:543-584, 1990. Nucleotide analogues can also comprise modified phosphodiester linkages including phosphorothioate, phosphorodithioate, alkyl-phosphonate, phosphoranilidate, phosphoramidate linkages and the like.

[0112] A dye may be attached to any position on the nucleotide base, for example, through a linker. In particular embodiments, Watson-Crick base pairing can still be carried out for the resulting analog. Particular nucleobase labeling sites include the C5 position of a pyrimidine base or the C7 position of a 7-deaza purine base. As described above a linker group may be used to covalently attach a dye to the nucleoside or nucleotide.

[0113] In particular embodiments the labeled nucleotide or oligonucleotide may be enzymatically incorporable and enzymatically extendable. Accordingly, a linker moiety may be of sufficient length to connect the nucleotide to the compound such that the compound does not significantly interfere with the overall binding and recognition of the nucleotide by a nucleic acid replication enzyme. Thus, the linker can also comprise a spacer unit. The spacer distances, for example, the nucleotide base from a cleavage site or label.

[0114] Nucleosides or nucleotides labeled with the dyes described herein may have the formula:



[0115] where Dye is a naphthalimide dye compound (label) moiety described herein (after covalent bonding between a functional group of the dye and a functional group of the linker "L"); B is a nucleobase, such as, for example uracil, thymine, cytosine, adenine, 7-deaza adenine, guanine, 7-deaza guanine, and the like; L is an optional linker which may or may not be present; R' can be H, or -OR' is monophosphate, diphosphate, triphosphate, thiophosphate, a phosphate ester analog, –O– attached to a reactive phosphorous containing group, or –O– protected by a blocking group; R'' is H or OH; and R''' is H, a 3' blocking group described herein, or -OR''' forms a phosphoramidite. Where -OR''' is phosphoramidite, R' is an acid-cleavable hydroxyl protecting group which allows subsequent monomer coupling under automated synthesis conditions. In some further embodiments, B comprises ,  or , or optionally substituted derivatives and analogs thereof. In some further embodiments, the labeled nucleobase comprises the structure , , or .

[0116] In a particular embodiment, the blocking group is separate and independent of the dye compound, i.e., not attached to it. Alternatively, the dye may comprise all or part of the 3'-OH blocking group. Thus R''' can be a 3' OH blocking group which may or may not comprise the dye compound.

[0117] In yet another alternative embodiment, there is no blocking group on the 3' carbon of the pentose sugar and the dye (or dye and linker construct) attached to the base, for example, can be of a size or structure sufficient to act as a block to the incorporation of a further nucleotide. Thus, the block can be due to steric hindrance or can be due to a combination of size, charge and structure, whether or not the dye is attached to the 3' position of the sugar.

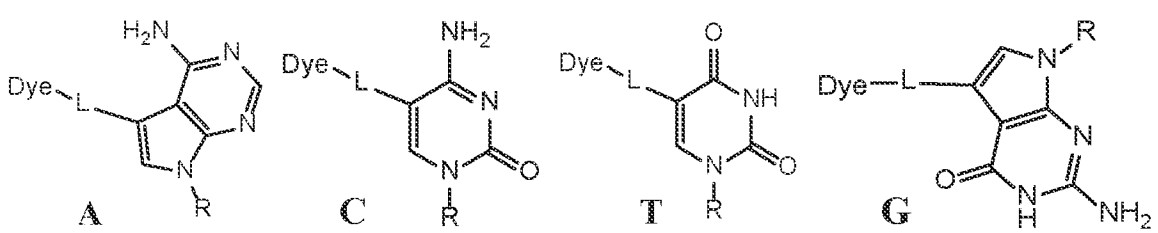
[0118] In still yet another alternative embodiment, the blocking group is present on the 2' or 4' carbon of the pentose sugar and can be of a size or structure sufficient to act as a block to the incorporation of a further nucleotide.

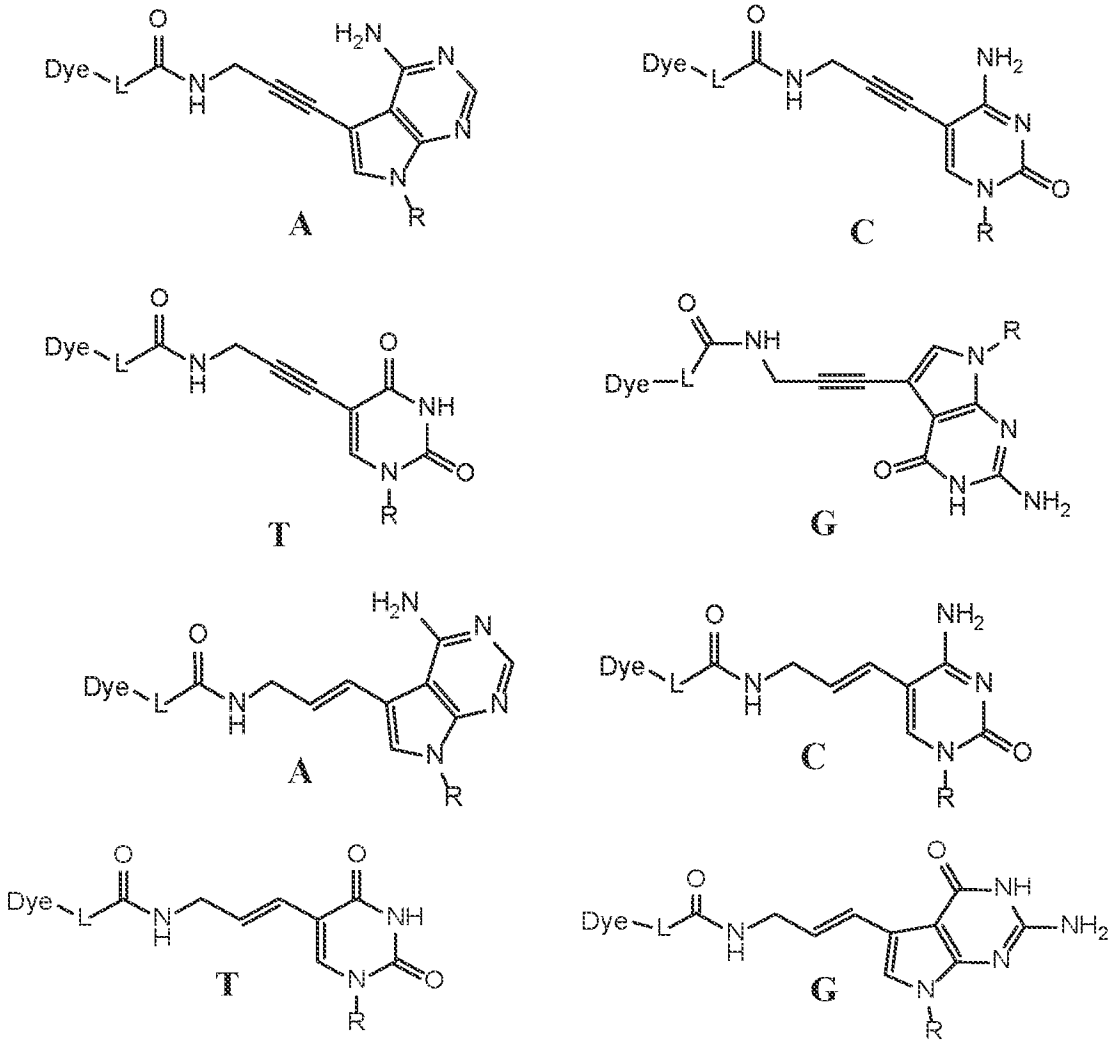
[0119] The use of a blocking group allows polymerization to be controlled, such as by stopping extension when a labeled nucleotide is incorporated. If the blocking effect is reversible, for example, by way of non-limiting example by changing chemical conditions or by removal of a chemical block, extension can be stopped at certain points and then allowed to continue.

[0120] In a particular embodiment, the linker (between dye and nucleotide) and blocking group are both present and are separate moieties. In particular embodiments, the linker and blocking group are both cleavable under the same or substantially similar conditions. Thus, deprotection and deblocking processes may be more efficient because only a single treatment will be required to remove both the dye compound and the blocking group. However, in some embodiments a linker and blocking group need not be cleavable under similar conditions, instead being individually cleavable under distinct conditions.

[0121] The disclosure also encompasses polynucleotides incorporating dye compounds. Such polynucleotides may be DNA or RNA comprised respectively of deoxyribonucleotides or ribonucleotides joined in phosphodiester linkage. Polynucleotides may comprise naturally occurring nucleotides, non-naturally occurring (or modified) nucleotides other than the labeled nucleotides described herein or any combination thereof, in combination with at least one modified nucleotide (e.g., labeled with a dye compound) as set forth herein. Polynucleotides according to the disclosure may also include non-natural backbone linkages and/or non-nucleotide chemical modifications. Chimeric structures comprised of mixtures of ribonucleotides and deoxyribonucleotides comprising at least one labeled nucleotide are also contemplated.

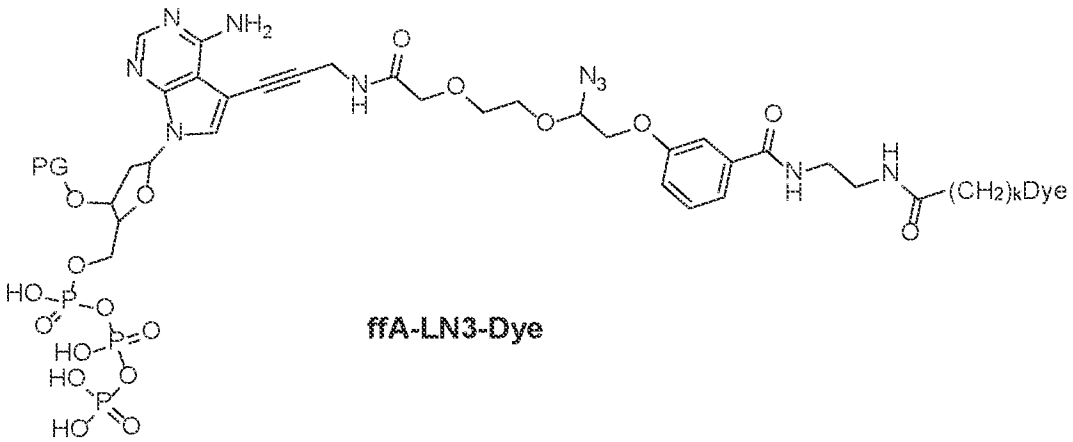
[0122] Non-limiting exemplary labeled nucleotides as described herein include:





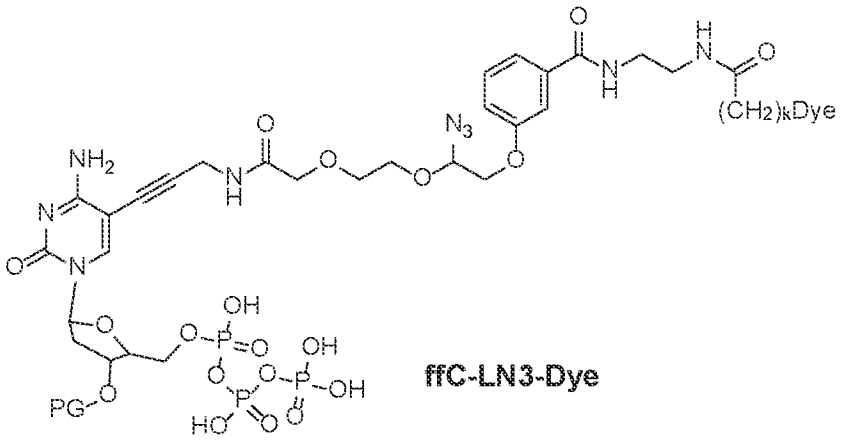
wherein L represents a linker and R represents a ribose or deoxyribose moiety as described above, or a ribose or deoxyribose moiety with the 5' position substituted with mono-, di- or tri-phosphates.

[0123] In some embodiments, non-limiting exemplary fluorescent dye conjugates are shown below:

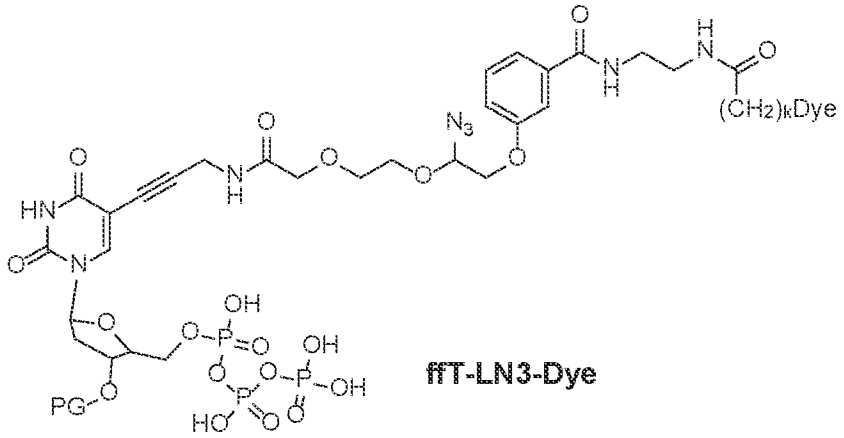


ffA-LN3-Dye

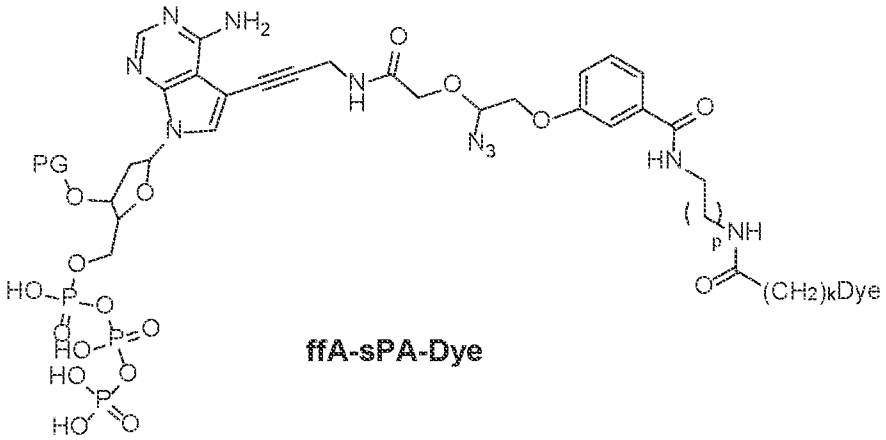
,



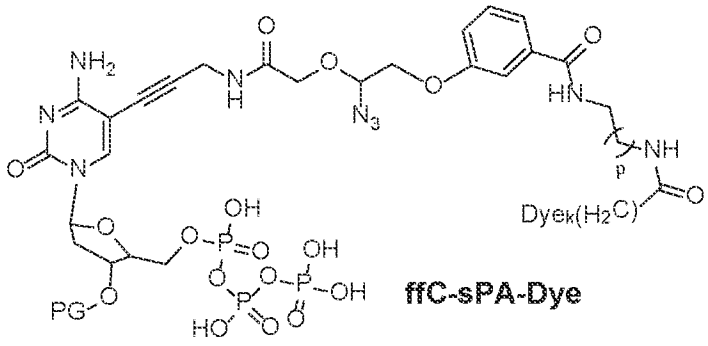
ffC-LN3-Dye



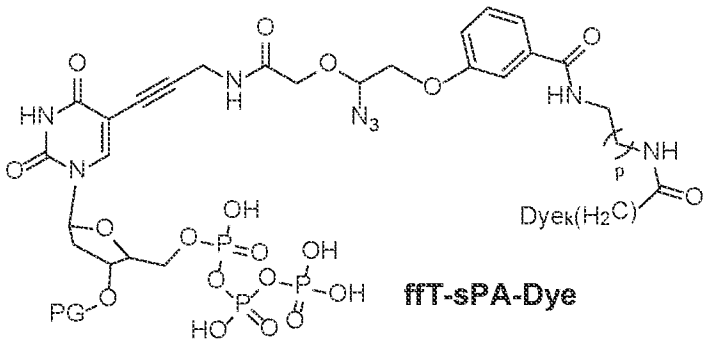
ffT-LN3-Dye



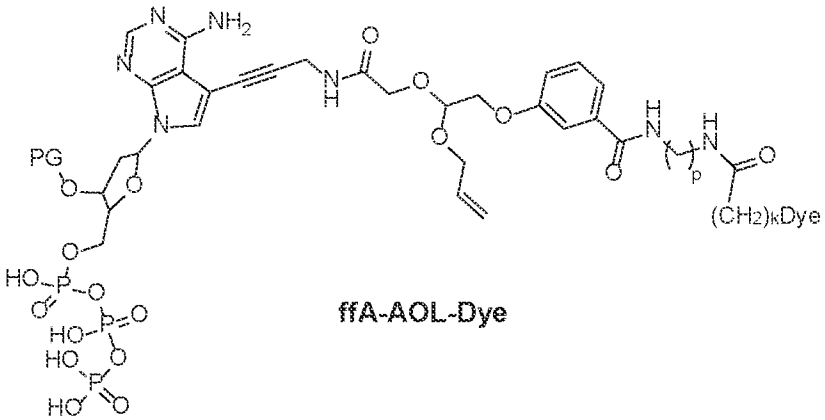
ffA-sPA-Dye



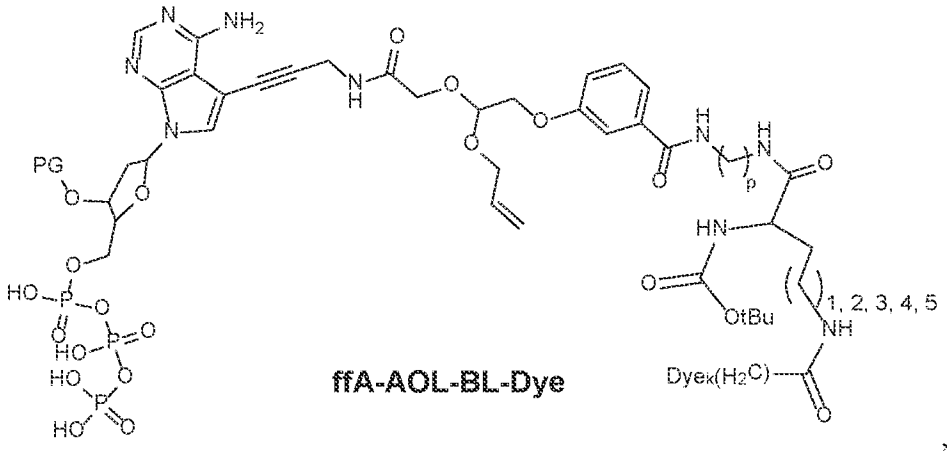
ffC-sPA-Dye



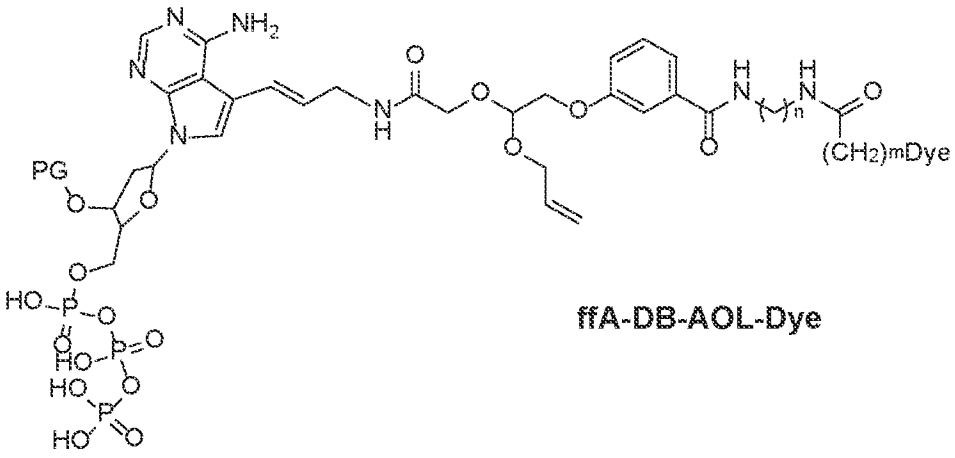
ffT-sPA-Dye



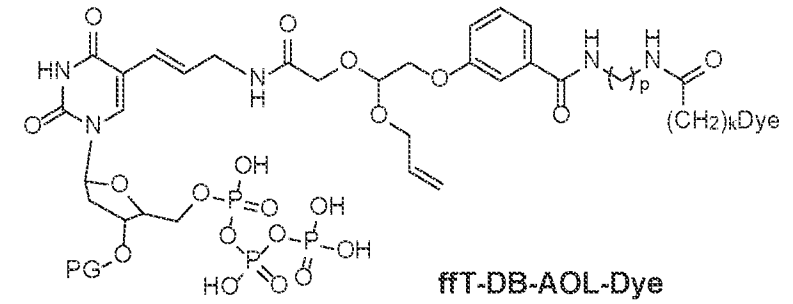
ffA-AOL-Dye



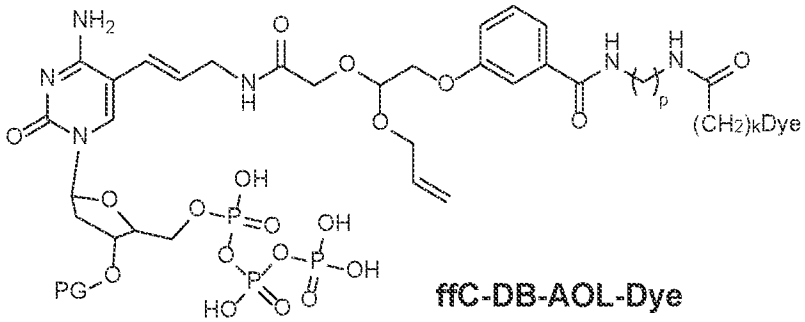
ffA-AOL-BL-Dye



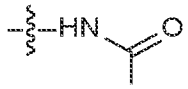
ffA-DB-AOL-Dye



ffT-DB-AOL-Dye



ffC-DB-AOL-Dye

wherein PG stands for the 3′ OH blocking groups described herein; p is an integer of 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10; and k is 0, 1, 2, 3, 4, or 5. In one embodiment, –O–PG is AOM. In another embodiment, –O–PG is –O–azidomethyl (AZM). In one embodiment, k is 5. In some further embodiments, p is 1, 2 or 3; and k is 5.  refers to the connection point of the Dye with the cleavable linker as a result of a reaction between an amino group of the linker moiety and the carboxyl group of the Dye. In any embodiments of the labeled nucleotide described herein, the nucleotide is a nucleotide triphosphate.

[0124] Additional aspects of the present disclosure relate to an oligonucleotide or polynucleotide comprising or incorporating a labeled nucleotide described herein. In some embodiments, the oligonucleotide or polynucleotide is hybridized to at least a portion of a target polynucleotide. In some embodiments, the target polynucleotide is immobilized on a solid support. In some further embodiments, the solid support comprises an array or a plurality of different immobilized target polynucleotides. In further embodiments, the solid support comprises a patterned flow cell. In further embodiments, the patterned flow cell comprises a plurality of nanowells. In further embodiments, the solid support comprises at least 5,000,000 spatially distinguishable sites/cm^2 that comprise multiple copies of target polynucleotides.

[0125] Additional aspects of the present disclosure relate to a protein tag or an antibody comprising one or more naphthalimide dyes described herein. In particular, the protein tag or an antibody may comprise multiple copies of the same dye for increased fluorescent intensity. The protein tag or antibody may be used as an affinity reagent that binds superficially to a particular type of unlabeled 3' blocked nucleotide.

Kits

[0126] Also provided herein are kits including a first type of nucleotide labeled with a naphthalimide compound of the present disclosure (i.e., a first label). In some embodiments, the kit also comprises a second type of labeled nucleotide, which is labeled with a second compound that is different than the naphthalimide dye in the first type of labeled nucleotide (i.e., a second label). In some embodiments, the first and second type labeled nucleotides are excitable using a single excitation source, which may be a first light source having a first excitation wavelength. For example, the excitation bands for the first and the second labels may be at least partially overlapping such that excitation in the overlap region of the spectrum causes both labels to emit fluorescence. In some other embodiments, the second type of labeled nucleotides is excitable using a second excitation source, which may be a second light source having a second excitation wavelength that is different from the first excitation wavelength. In some further embodiments, the kit may include a third type nucleotide, wherein the third nucleotide is labeled with a third compound that is different from the first and the second labels (i.e., a third label). Alternatively, the third type of nucleotide is labeled with both the first label and the second label. In some such embodiments, the third type labeled nucleotide is excitable using the first light source having the first excitation wavelength. That is, each of the first type, second type and the third type of nucleotide is excitable using the same light source with a single wavelength. In some other embodiments, the third type labeled nucleotide is excitable using both the first light source having the first excitation wavelength, or the second light source having the second excitation wavelength. In still other embodiments, the third type of nucleotide is excitable using a third light source with a third excitation wavelength. In some further embodiments, the kit may further comprise a fourth type of nucleotide. In some such embodiments, the fourth nucleotide is unlabeled (dark). In other embodiments, the fourth nucleotide is labeled with a different compound than the first, second and the third nucleotide, and each label has a distinct absorbance maximum that is distinguishable from the other labels. In still other embodiments, the fourth nucleotide is unlabeled. In some embodiments, the first light source has an excitation wavelength from about 400 nm to about 480 nm, from about 420nm to about 470 nm, or from 450 nm to about 460 nm (e.g., 450 nm). The second excitation light source has a wavelength from about 500 nm to about 550 nm, from about 510 to about 540 nm, or from about 520 to about 530 nm (e.g., 520 nm). The second light source has an excitation wavelength from about 400 nm to about 480 nm, from about 420nm to about 470 nm, or from 450 nm to about 460 nm (e.g., about 452 nm). In some embodiments, the emissions of the first type of labeled nucleotide, the second type of labeled nucleotide and the third type of labeled nucleotide are detectable in two detection channels with

different wavelengths (e.g., at blue region with a wavelength ranging from about 472 to about 520 nm, and at a green region with a wavelength ranging from about 540 nm to about 640nm). In other embodiments, each of the first type, the second type and the third type of nucleotide has an emission spectrum that can be collected in a single emission collection filter or channel.

[0127] In some embodiments, the kit may contain four types of labeled nucleotides (A, C, G and T or U), where the first of the four nucleotides is labeled with a compound as disclosed herein. In such a kit, each of the four nucleotides can be labeled with a compound that is the same or different from the label on the other three nucleotides. Alternatively, a first of the four nucleotides is a labeled nucleotide describe herein, a second of the four nucleotides carries a second label, a third nucleotide carries a third label, and a fourth nucleotide is unlabeled (dark). As another example, a first of the four nucleotides is a labeled nucleotide described herein, a second of the four nucleotides carries a second label, a third nucleotide carries a mixture of two labels, and a fourth nucleotide is unlabeled (dark). Thus, one or more of the label compounds can have a distinct absorbance maximum and/or emission maximum such that the compound(s) is(are) distinguishable from other compounds. For example, each compound can have a distinct absorbance maximum and/or emission maximum such that each of the compounds is spectrally distinguishable from the other three compounds (or two compounds if the fourth nucleotide is unlabeled). It will be understood that parts of the absorbance spectrum and/or emission spectrum other than the maxima can differ and these differences can be exploited to distinguish the compounds. The kit may be such that two or more of the compounds have a distinct absorbance maximum.

[0128] The compounds, nucleotides, or kits that are set forth herein may be used to detect, measure, or identify a biological system (including, for example, processes or components thereof). Exemplary techniques that can employ the compounds, nucleotides or kits include sequencing, expression analysis, hybridization analysis, genetic analysis, RNA analysis, cellular assay (e.g., cell binding or cell function analysis), or protein assay (e.g., protein binding assay or protein activity assay). The use may be on an automated instrument for carrying out a particular technique, such as an automated sequencing instrument. The sequencing instrument may contain two light sources operating at different wavelengths.

[0129] In a particular embodiment, the labeled nucleotide(s) described herein may be supplied in combination with unlabeled or native nucleotides, or any combination thereof. Combinations of nucleotides may be provided as separate individual components (e.g., one nucleotide type per vessel or tube) or as nucleotide mixtures (e.g., two or more nucleotides mixed in the same vessel or tube).

[0130] Where kits comprise a plurality, particularly two, or three, or more particularly four, nucleotides, the different nucleotides may be labeled with different dye compounds, or one may be dark, with no dye compounds. Where the different nucleotides are labeled with different dye compounds, it is a feature of the kits that the dye compounds are spectrally distinguishable fluorescent dyes. As used herein, the term "spectrally distinguishable fluorescent dyes" refers to fluorescent dyes that emit fluorescent energy at wavelengths that can be distinguished by fluorescent detection equipment (for example, a commercial capillary-based DNA sequencing platform) when two or more such dyes are present in one sample. When two nucleotides labeled with fluorescent dye compounds are supplied in kit form, it is a feature of some embodiments that the spectrally distinguishable fluorescent dyes can be excited at the same wavelength, such as, for example by the same light source. When four nucleotides labeled with fluorescent dye compounds are supplied in kit form, it is a feature of some embodiments that two of the spectrally distinguishable fluorescent dyes can both be excited at one wavelength and the other two spectrally distinguishable dyes can both be excited at another wavelength. Particular excitation wavelengths for the dyes are between 450-460 nm, 490-500 nm, or 520 nm or above (e.g., 532 nm).

[0131] In some embodiments, a kit includes a first type nucleotide labeled with a naphthalimide dye of the present disclosure. In addition, the kit may contain a second type of labeled having a long Stokes shift of about or above 70 nm, above 80 nm, about 90 nm, or above 100 nm (where "Stokes shift" is the distance between the peak absorption and peak emission wavelengths). In addition, the kit may contain a third type of labeled nucleotide having a short Stokes shift of about or less than 50 nm, 40 nm, or 30 nm. In some such embodiments, the first type of labeled nucleotide with the naphthalimide dye may have a Stokes shift range falling between the second and the third type of nucleotides, e.g., between about 30 nm to about 70 nm, between about 40 nm to about 60 nm, or between about 45 nm to about 55 nm. In further embodiments, each of the first type, second type and third type of nucleotides are excitable by a single light source (e.g., a blue light having a wavelength of between about 400 nm to about 460 nm, or about 450 nm to about 460 nm). In some embodiments, the emission spectra of the first type, second type and third type of nucleotides are collectable in two different detection channels or regions (e.g., at blue region with a wavelength ranging from about 472 to about 520 nm, and at a green region with a wavelength ranging from about 540 nm to about 640nm).

[0132] Although kits are exemplified herein in regard to configurations having different nucleotides that are labeled with different dye compounds, it will be understood that kits can include 2, 3, 4 or more different nucleotides that have the same dye compound.

[0133] In addition to the labeled nucleotides, the kit may comprise together at least one additional component. The further component(s) may be one or more of the components identified in a method set forth herein or in the Examples section below. Some non-limiting examples of components that can be combined into a kit of the present disclosure are set forth below. In some embodiments, the kit further comprises a DNA polymerase (such as a mutant DNA polymerase) and one or more buffer compositions. Non-limiting examples of DNA polymerase may be used in the present disclosure include those disclosed in WO 2005/024010, U.S. Publication Nos. 2020/0131484 A1, 2020/0181587 A1, and U.S. Ser. Nos. 63/412,241 and 63/433,971, each of which is incorporated by reference herein in its entirety. One buffer composition may comprise antioxidants such as ascorbic acid or sodium ascorbate, which can be used to protect the dye compounds from photo damage during detection. Additional buffer composition may comprise a reagent can may be used to cleave the 3′ blocking group and/or the cleavable linker. For example, a water-soluble phosphines or water-soluble transition metal catalysts formed from a transition metal and at least partially water-soluble ligands, such as a palladium complex. Various components of the kit may be provided in a concentrated form to be diluted prior to use. In such embodiments a suitable dilution buffer may also be included. Again, one or more of the components identified in a method set forth herein can be included in a kit of the present disclosure. In any embodiments of the nucleotide or labeled nucleotide described herein, the nucleotide contains a 3′ blocking group.

Methods of Sequencing

[0134] Nucleotides comprising a dye compound according to the present disclosure may be used in any method of analysis such as method that include detection of a fluorescent label attached to such nucleotide, whether on its own or incorporated into or associated with a larger molecular structure or conjugate. In this context the term "incorporated into a polynucleotide" can mean that the 5' phosphate is joined in phosphodiester linkage to the 3' hydroxyl group of a second nucleotide, which may itself form part of a longer polynucleotide chain. The 3' end of a nucleotide set forth herein may or may not be joined in phosphodiester linkage to the 5' phosphate of a further nucleotide. Thus, in one non-limiting embodiment, the disclosure provides a method of detecting a labeled nucleotide incorporated into a polynucleotide which comprises: (a) incorporating at least one labeled nucleotide of the disclosure into a polynucleotide and (b) determining the identity of the nucleotide(s) incorporated into the polynucleotide by detecting the fluorescent signal from the dye compound attached to said nucleotide(s). Particular embodiments of the method of sequencing utilizes a one-excitation, two-channel detection system (also known as 1Ex-2Ch). Detailed disclosures are provided in WO 2018/165099 and U.S. Publication No.

2022/0403450, each of which is incorporated by reference in its entirety. In particular, the 1Ex-2Ch may contain a blue excitation light source having a wavelength between about 450 nm to about 460 nm, and two separate collection channels at both the blue and green regions (e.g., at a blue region with a wavelength ranging from about 472 to about 520 nm, and at a green region with a wavelength ranging from about 540 nm to about 640nm).

[0135] This method can include: a synthetic step (a) in which one or more labeled nucleotides according to the disclosure are incorporated into a polynucleotide and a detection step (b) in which one or more labeled nucleotide(s) incorporated into the polynucleotide are detected by detecting or quantitatively measuring their fluorescence.

[0136] Some embodiments of the present application are directed to a method for determining the sequences of a plurality of different target polynucleotides, comprising:

(a) contacting a solid support with a solution comprising sequencing primers under hybridization conditions, wherein the solid support comprises a plurality of different target polynucleotides immobilized thereon; and the sequencing primers are complementary to at least a portion of the target polynucleotides;

(b) contacting the solid support with an aqueous solution comprising DNA polymerase and one more of four different types of nucleotides (e.g., dATP, dGTP, dCTP and dTTP or dUTP), under conditions suitable for DNA polymerase-mediated primer extension, and incorporating one type of nucleotides into the sequencing primers to produce extended copy polynucleotides, wherein at least one type of nucleotide is a naphthalimide dye labeled nucleotide described herein, and wherein each of the four types of nucleotides comprises a 3′ blocking group covalently attached to the deoxyribose sugar of the nucleotide;

(c) imaging the solid support and performing one or more fluorescent measurements of the extended copy polynucleotides; and

(d) removing the 3′ blocking group of the nucleotides incorporated into the extended copy polynucleotides. In some embodiments, step (d) also removes the labels of the incorporated nucleotides (if the incorporated nucleotides are labeled). In some such embodiments, the labels and the 3′ blocking groups of the incorporated nucleotides are removed in a single chemical reaction. In some further embodiments, the method may also comprises (e) washing the solid support with an aqueous wash solution (e.g., washing the removed label moiety and the 3′ blocking group away from the extended copy polynucleotides). In some embodiments, steps (b) through (e) are repeated at least 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, or 300 cycles to determine the target polynucleotide sequences. In some embodiments, the four types of nucleotides comprise dATP, dCTP, dGTP and dTTP or dUTP, or non-natural nucleotide analogs thereof. In some embodiments, the sequence determination is

conducted after the completion of repeated cycles of the sequencing steps described herein. In some embodiments, the naphthalimide dyes described herein may be used as any one of the first, the second or the third label described in the method.

[0137] In some further embodiments, the method is performed on an automated sequencing instrument, and wherein the automated sequencing instrument comprises a single light source operating with a blue laser at about 450 nm to about 460 nm. The incorporation of the first type of the nucleotide is determined by detection in the one of the blue or green channel/region (e.g., at a blue region with a wavelength ranging from about 472 to about 520 nm, or at a green region with a wavelength ranging from about 540 nm to about 640nm). The incorporation of the second type of nucleotide is determined by detection in the other one of the blue or green detection channel/region. The incorporation of the third type of nucleotide is determined by detection in both the blue and green channels/regions. The incorporation of the fourth type of nucleotide is determined by no detection in either the blue or the green channel/region.

[0138] In other embodiments, the automatic sequencing instrument may comprise two light sources operating at different wavelengths (e.g., at 450-460 nm and 520-530 nm). The incorporation of the first type of the nucleotide conjugates is determined by a signal state in the first imaging event and a dark state in the second imaging event. The incorporation of the second type of the nucleotide conjugates is determined by a dark state in the first imaging event and a signal state in the second imaging event. The incorporation of the third type of the nucleotide conjugates is determined by a signal state in both the first imaging event and the second imaging event. The incorporation of the fourth type of the nucleotide conjugates is determined by a dark state in both the first imaging event and the second imaging event.

[0139] In some embodiments, at least one nucleotide is incorporated into a polynucleotide (such as a single stranded primer polynucleotide described herein) in the synthetic step by the action of a polymerase enzyme. However, other methods of joining nucleotides to polynucleotides, such as, for example, chemical oligonucleotide synthesis or ligation of labeled oligonucleotides to unlabeled oligonucleotides, can be used. Therefore, the term "incorporating," when used in reference to a nucleotide and polynucleotide, can encompass polynucleotide synthesis by chemical methods as well as enzymatic methods.

[0140] In a specific embodiment, a synthetic step is carried out and may optionally comprise incubating a template or target polynucleotide strand with a reaction mixture comprising fluorescently labeled nucleotides of the disclosure. A polymerase can also be provided under conditions which permit formation of a phosphodiester linkage between a free 3' hydroxyl group on a polynucleotide strand annealed to the template or target polynucleotide strand and a 5' phosphate group on the labeled nucleotide. Thus, a synthetic step can include formation of a

polynucleotide strand as directed by complementary base pairing of nucleotides to a template/target strand.

[0141] In all embodiments of the methods, the detection step may be carried out while the polynucleotide strand into which the labeled nucleotides are incorporated is annealed to a template/target strand, or after a denaturation step in which the two strands are separated. Further steps, for example chemical or enzymatic reaction steps or purification steps, may be included between the synthetic step and the detection step. In particular, the polynucleotide strand incorporating the labeled nucleotide(s) may be isolated or purified and then processed further or used in a subsequent analysis. By way of example, polynucleotide strand incorporating the labeled nucleotide(s) as described herein in a synthetic step may be subsequently used as labeled probes or primers. In other embodiments, the product of the synthetic step set forth herein may be subject to further reaction steps and, if desired, the product of these subsequent steps purified or isolated.

[0142] Suitable conditions for the synthetic step will be well known to those familiar with standard molecular biology techniques. In one embodiment, a synthetic step may be analogous to a standard primer extension reaction using nucleotide precursors, including the labeled nucleotides as described herein, to form an extended polynucleotide strand (primer polynucleotide strand) complementary to the template/target strand in the presence of a suitable polymerase enzyme. In other embodiments, the synthetic step may itself form part of an amplification reaction producing a labeled double stranded amplification product comprised of annealed complementary strands derived from copying of the primer and template polynucleotide strands. Other exemplary synthetic steps include nick translation, strand displacement polymerization, random primed DNA labeling, etc. A particularly useful polymerase enzyme for a synthetic step is one that is capable of catalyzing the incorporation of the labeled nucleotides as set forth herein. A variety of naturally occurring or mutant/modified polymerases can be used. By way of example, a thermostable polymerase can be used for a synthetic reaction that is carried out using thermocycling conditions, whereas a thermostable polymerase may not be desired for isothermal primer extension reactions. Suitable thermostable polymerases which are capable of incorporating the labeled nucleotides according to the disclosure include those described in WO 2005/024010 or WO06120433, each of which is incorporated herein by reference. In synthetic reactions which are carried out at lower temperatures such as 37 °C, polymerase enzymes need not necessarily be thermostable polymerases, therefore the choice of polymerase will depend on a number of factors such as reaction temperature, pH, strand-displacing activity and the like.

[0143] In specific non-limiting embodiments, the disclosure encompasses methods of nucleic acid sequencing, re-sequencing, whole genome sequencing, single nucleotide

polymorphism scoring, any other application involving the detection of the modified nucleotide or nucleoside labeled with dyes set forth herein when incorporated into a polynucleotide.

[0144] A particular embodiment of the disclosure provides use of labeled nucleotides comprising dye moiety according to the disclosure in a polynucleotide sequencing-by-synthesis reaction. Sequencing-by-synthesis generally involves sequential addition of one or more nucleotides or oligonucleotides to a growing polynucleotide chain in the 5' to 3' direction using a polymerase or ligase in order to form an extended polynucleotide chain complementary to the template/target nucleic acid to be sequenced. The identity of the base present in one or more of the added nucleotide(s) can be determined in a detection or "imaging" step. The identity of the added base may be determined after each nucleotide incorporation step. The sequence of the template may then be inferred using conventional Watson-Crick base-pairing rules. The use of the nucleotides labeled with dyes set forth herein for determination of the identity of a single base may be useful, for example, in the scoring of single nucleotide polymorphisms, and such single base extension reactions are within the scope of this disclosure.

[0145] In an embodiment of the present disclosure, the sequence of a template/target polynucleotide is determined by detecting the incorporation of one or more nucleotides into a nascent strand complementary to the template polynucleotide to be sequenced through the detection of fluorescent label(s) attached to the incorporated nucleotide(s). Sequencing of the template polynucleotide can be primed with a suitable primer (or prepared as a hairpin construct which will contain the primer as part of the hairpin), and the nascent chain is extended in a stepwise manner by addition of nucleotides to the 3' end of the primer in a polymerase-catalyzed reaction.

[0146] In particular embodiments, each of the different nucleotide triphosphates (A, T, G and C) may be labeled with a unique fluorophore and also comprises a blocking group at the 3' position to prevent uncontrolled polymerization. Alternatively, one of the four nucleotides may be unlabeled (dark). The polymerase enzyme incorporates a nucleotide into the nascent chain complementary to the template/target polynucleotide, and the blocking group prevents further incorporation of nucleotides. Any unincorporated nucleotides can be washed away and the fluorescent signal from each incorporated nucleotide can be "read" optically by suitable means, such as a charge-coupled device using light source excitation and suitable emission filters. The 3' blocking group and fluorescent dye compounds can then be removed (deprotected) (simultaneously or sequentially) to expose the nascent chain for further nucleotide incorporation. Typically, the identity of the incorporated nucleotide will be determined after each incorporation step, but this is not strictly essential. Similarly, U.S. Pat. No. 5,302,509 (which is incorporated

herein by reference) discloses a method to sequence polynucleotides immobilized on a solid support.

[0147] The method, as exemplified above, utilizes the incorporation of fluorescently labeled, 3'-blocked nucleotides A, G, C, and T into a growing strand complementary to the immobilized polynucleotide, in the presence of DNA polymerase. The polymerase incorporates a base complementary to the target polynucleotide but is prevented from further addition by the 3'-blocking group. The label of the incorporated nucleotide can then be determined, and the blocking group removed by chemical cleavage to allow further polymerization to occur. The nucleic acid template to be sequenced in a sequencing-by-synthesis reaction may be any polynucleotide that it is desired to sequence. The nucleic acid template for a sequencing reaction will typically comprise a double stranded region having a free 3' hydroxyl group that serves as a primer or initiation point for the addition of further nucleotides in the sequencing reaction. The region of the template to be sequenced will overhang this free 3' hydroxyl group on the complementary strand. The overhanging region of the template to be sequenced may be single stranded but can be double-stranded, provided that a "nick is present" on the strand complementary to the template strand to be sequenced to provide a free 3' OH group for initiation of the sequencing reaction. In such embodiments, sequencing may proceed by strand displacement. In certain embodiments, a primer bearing the free 3' hydroxyl group may be added as a separate component (e.g., a short oligonucleotide) that hybridizes to a single-stranded region of the template to be sequenced. Alternatively, the primer and the template strand to be sequenced may each form part of a partially self-complementary nucleic acid strand capable of forming an intra-molecular duplex, such as for example a hairpin loop structure. Hairpin polynucleotides and methods by which they may be attached to solid supports are disclosed in PCT Publication Nos. WO0157248 and WO2005/047301, each of which is incorporated herein by reference. Nucleotides can be added successively to a growing primer, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the base which has been added may be determined, particularly but not necessarily after each nucleotide addition, thus providing sequence information for the nucleic acid template. Thus, a nucleotide is incorporated into a nucleic acid strand (or polynucleotide) by joining of the nucleotide to the free 3' hydroxyl group of the nucleic acid strand via formation of a phosphodiester linkage with the 5' phosphate group of the nucleotide.

[0148] The nucleic acid template to be sequenced may be DNA or RNA, or even a hybrid molecule comprised of deoxynucleotides and ribonucleotides. The nucleic acid template may comprise naturally occurring and/or non-naturally occurring nucleotides and natural or non-natural backbone linkages, provided that these do not prevent copying of the template in the sequencing reaction.

[0149] In certain embodiments, the nucleic acid template to be sequenced may be attached to a solid support via any suitable linkage method known in the art, for example via covalent attachment. In certain embodiments template polynucleotides may be attached directly to a solid support (e.g., a silica-based support). However, in other embodiments of the disclosure the surface of the solid support may be modified in some way so as to allow either direct covalent attachment of template polynucleotides, or to immobilize the template polynucleotides through a hydrogel or polyelectrolyte multilayer, which may itself be non-covalently attached to the solid support.

[0150] Arrays in which polynucleotides have been directly attached to a support (for example, silica-based supports such as those disclosed in WO00/06770 (incorporated herein by reference), wherein polynucleotides are immobilized on a glass support by reaction between a pendant epoxide group on the glass with an internal amino group on the polynucleotide. In addition, polynucleotides can be attached to a solid support by reaction of a sulfur-based nucleophile with the solid support, for example, as described in W02005/047301 (incorporated herein by reference). A still further example of solid-supported template polynucleotides is where the template polynucleotides are attached to hydrogel supported upon silica-based or other solid supports, for example, as described in WO00/31148, WO01/01143, WO02/12566, WO03/014392, U.S. Pat. No. 6,465,178 and WO00/53812, each of which is incorporated herein by reference.

[0151] A particular surface to which template polynucleotides may be immobilized is a polyacrylamide hydrogel. Polyacrylamide hydrogels are described in the references cited above and in WO2005/065814, which is incorporated herein by reference. Specific hydrogels that may be used include those described in WO2005/065814 and U.S. Pub. No. 2014/0079923. In one embodiment, the hydrogel is PAZAM (poly(N-(5-azidoacetamidylpentyl) acrylamide-co-acrylamide)).

[0152] DNA template molecules can be attached to beads or microparticles, for example, as described in U.S. Pat. No. 6,172,218 (which is incorporated herein by reference). Attachment to beads or microparticles can be useful for sequencing applications. Bead libraries can be prepared where each bead contains different DNA sequences. Exemplary libraries and methods for their creation are described in Nature, 437, 376-380 (2005); Science, 309, 5741, 1728-1732 (2005), each of which is incorporated herein by reference. Sequencing of arrays of such beads using nucleotides set forth herein is within the scope of the disclosure.

[0153] Template(s) that are to be sequenced may form part of an "array" on a solid support, in which case the array may take any convenient form. Thus, the method of the disclosure is applicable to all types of high-density arrays, including single-molecule arrays, clustered arrays,

and bead arrays. Nucleotides labeled with dye compounds of the present disclosure may be used for sequencing templates on essentially any type of array, including but not limited to those formed by immobilization of nucleic acid molecules on a solid support.

[0154] However, nucleotides labeled with dye compounds of the disclosure are particularly advantageous in the context of sequencing of clustered arrays. In clustered arrays, distinct regions on the array (often referred to as sites, or features) comprise multiple polynucleotide template molecules. Generally, the multiple polynucleotide molecules are not individually resolvable by optical means and are instead detected as an ensemble. Depending on how the array is formed, each site on the array may comprise multiple copies of one individual polynucleotide molecule (e.g., the site is homogenous for a particular single- or double-stranded nucleic acid species) or even multiple copies of a small number of different polynucleotide molecules (e.g., multiple copies of two different nucleic acid species). Clustered arrays of nucleic acid molecules may be produced using techniques generally known in the art. By way of example, WO 98/44151 and WO00/18957, each of which is incorporated herein, describe methods of amplification of nucleic acids wherein both the template and amplification products remain immobilized on a solid support in order to form arrays comprised of clusters or "colonies" of immobilized nucleic acid molecules. The nucleic acid molecules present on the clustered arrays prepared according to these methods are suitable templates for sequencing using nucleotides labeled with dye compounds of the disclosure.

[0155] Nucleotides labeled with dye compounds of the present disclosure are also useful in sequencing of templates on single molecule arrays. The term "single molecule array" or "SMA" as used herein refers to a population of polynucleotide molecules, distributed (or arrayed) over a solid support, wherein the spacing of any individual polynucleotide from all others of the population is such that it is possible to individually resolve the individual polynucleotide molecules. The target nucleic acid molecules immobilized onto the surface of the solid support can thus be capable of being resolved by optical means in some embodiments. This means that one or more distinct signals, each representing one polynucleotide, will occur within the resolvable area of the particular imaging device used.

[0156] Single molecule detection may be achieved wherein the spacing between adjacent polynucleotide molecules on an array is at least 100 nm, more particularly at least 250 nm, still more particularly at least 300 nm, even more particularly at least 350 nm. Thus, each molecule is individually resolvable and detectable as a single molecule fluorescent point, and fluorescence from said single molecule fluorescent point also exhibits single step photobleaching.

[0157] The terms "individually resolved" and "individual resolution" are used herein to specify that, when visualized, it is possible to distinguish one molecule on the array from its

neighboring molecules. Separation between individual molecules on the array will be determined, in part, by the particular technique used to resolve the individual molecules. The general features of single molecule arrays will be understood by reference to published applications WO00/06770 and WO 01/57248, each of which is incorporated herein by reference. Although one use of the labeled nucleotides of the disclosure is in sequencing-by-synthesis reactions, the utility of such nucleotides is not limited to such methods. In fact, the labeled nucleotides described herein may be used advantageously in any sequencing methodology which requires detection of fluorescent labels attached to nucleotides incorporated into a polynucleotide.

[0158] In particular, nucleotides labeled with dye compounds of the disclosure may be used in automated fluorescent sequencing protocols, particularly fluorescent dye-terminator cycle sequencing based on the chain termination sequencing method of Sanger and co-workers. Such methods generally use enzymes and cycle sequencing to incorporate fluorescently labeled dideoxynucleotides in a primer extension sequencing reaction. So-called Sanger sequencing methods, and related protocols (Sanger-type), utilize randomized chain termination with labeled dideoxynucleotides.

[0159] Thus, the present disclosure also encompasses nucleotides labeled with dye compounds which are dideoxynucleotides lacking hydroxyl groups at both of the 3' and 2' positions, such modified dideoxynucleotides being suitable for use in Sanger type sequencing methods and the like.

[0160] Nucleotides labeled with dye compounds of the present disclosure incorporating 3' blocking groups, it will be recognized, may also be of utility in Sanger methods and related protocols since the same effect achieved by using dideoxy nucleotides may be achieved by using nucleotides having 3′ OH blocking groups: both prevent incorporation of subsequent nucleotides. Where nucleotides according to the present disclosure, and having a 3' blocking group are to be used in Sanger-type sequencing methods it will be appreciated that the dye compounds or detectable labels attached to the nucleotides need not be connected via cleavable linkers, since in each instance where a labeled nucleotide of the disclosure is incorporated; no nucleotides need to be subsequently incorporated and thus the label need not be removed from the nucleotide.

[0161] Alternatively, the sequencing methods described herein may also be carried out using unlabeled nucleotides and affinity reagents containing a naphthalimide dye described herein. For example, one, two, three or each of the four different types of nucleotides (e.g., dATP, dCTP, dGTP and dTTP or dUTP) in the incorporation mixture of step (a) may be unlabeled. Each of the four types of nucleotides (e.g., dNTPs) has a 3′ blocking group to ensure that only a single base can be added by a polymerase to the 3′ end of the primer polynucleotide. After incorporation of

an unlabeled nucleotide in step (b), the remaining unincorporated nucleotides are washed away. An affinity/labeling reagent is then introduced that specifically recognizes and binds to the incorporated dNTP to provide a labeled extension product comprising the incorporated dNTP. Uses of unlabeled nucleotides and affinity/labeling reagents in sequencing by synthesis have been disclosed in WO 2018/129214 and WO 2020/097607, as well as U.S. Application No. 18/325610, which is incorporated by reference in its entirety. A modified sequencing method of the present disclosure using unlabeled nucleotides may include the following steps:

(a') contacting a solid support with a solution comprising sequencing primers under hybridization conditions, wherein the solid support comprises a plurality of different target polynucleotides immobilized thereon; and the sequencing primers are complementary to at least a portion of the target polynucleotides;

(b') contacting the solid support with an aqueous solution comprising DNA polymerase and one more of four different types of unlabeled nucleotides (e.g., dATP, dCTP, dGTP, and dTTP or dUTP) under conditions suitable for DNA polymerase-mediated primer extension, and incorporating one type of nucleotides into the sequencing primers to produce extended copy polynucleotides, and wherein each of the four types of nucleotides comprises a 3′ blocking group;

(c') contacting the extended copy polynucleotides with a set of affinity/labeling reagents under conditions wherein one affinity reagent binds specifically to the incorporated unlabeled nucleotides to provide labeled extended copy polynucleotides;

(d') imaging the solid support and performing one or more fluorescent measurements of the extended copy polynucleotides; and

(e') removing the 3′ blocking group of the incorporated nucleotides.

[0162] In some embodiments of the modified sequencing method described herein, the method further comprises removing the affinity reagents from the incorporated nucleotides. In still further embodiments, the 3′ blocking group and the affinity reagent are removed in the same reaction. In some embodiments, the method further comprises a step (f') washing the solid support with an aqueous wash solution. In further embodiments, steps (b') through (f') are repeated at least 50, 100, 150, 200, 250 or 300 cycles to determine the target polynucleotide sequences. In some embodiments, the set of affinity reagents may comprise a first affinity reagent that binds specifically to the first type of nucleotide, a second affinity reagent that binds specifically to the second type of nucleotide, and a third affinity reagent that binds specifically to the third type of nucleotide. In some further embodiments, each of the first, second and the third affinity reagents comprises a detectable labeled that is spectrally distinguishable. In some embodiments, the affinity reagents may include protein tags, antibodies (including but not limited to binding fragments of

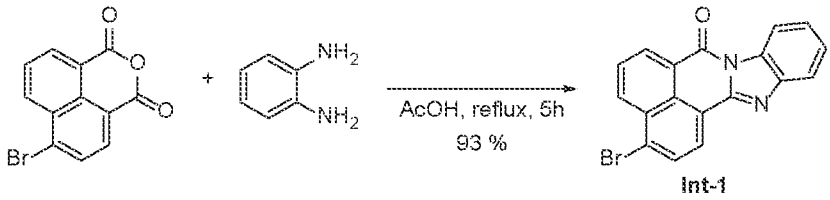
antibodies, single chain antibodies, bispecific antibodies, and the like), aptamers, knottins, affimers, or any other known agent that binds an incorporated nucleotide with a suitable specificity and affinity. In one embodiment, at least one affinity reagent is an antibody or a protein tag. In another embodiment, at least one of the first type, the second type, and the third type of affinity reagents is an antibody or a protein tag comprising one or more detectable labels (e.g., multiple copies of the same detectable label), wherein the detectable label is or comprises a naphthalimide dye moiety described herein.

EXAMPLES

[0163] Additional embodiments are disclosed in further detail in the following examples, which are not in any way intended to limit the scope of the claims.

Example 1. Synthesis of Naphthalimide Dyes

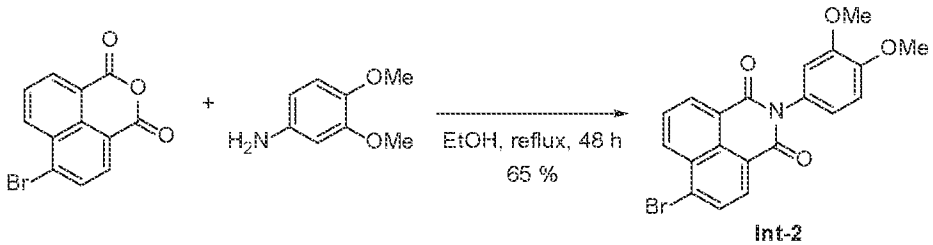
Synthesis of Intermediate 1 (Int-1)



Int-1

[0164] 4-bromo-1,8-naphtalic anhydride (1.0 equiv) and *o*-phenyl diamine (2.0 equiv) in glacial acetic acid were stirred at reflux for 5 h. The crude reaction mixture was cooled down and poured into cold water. The product was precipitated as a yellow solid. The precipitate was dissolved in DCM and dried over anhydrous Na_2SO_4. The solvent was removed under vacuum to obtain bromo-benzo[de]benzo[4,5]imidazole [2,1-a]isoquinoline-7-one as pale yellow solid (93%) as two inseparable isomers (cis and trans). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.91 (dd, J = 8.0 Hz, 1.03 Hz, 1H), 8.83 (dd, J = 8.0 Hz 1.23 Hz, 1H), 8.67 (dd, J = 8.0 Hz, 1.2 Hz, 2H), 8.58 (d, J = 8.0 Hz, 1H), 8.53 (m, 3H), 8.11 (d, J = 8.0 Hz, 1H), 8.08 (d, J = 8.0 Hz, 1H), 7.88 (m, 4H), 7.50 (m, 4H). Mass Spec: $[M+H]^+$ = 351.18.

Synthesis of Intermediate 2 (Int-2)

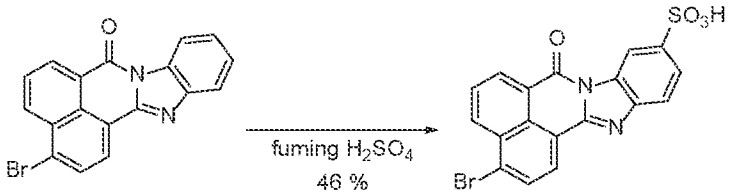


Int-2

[0165] 4-bromo-1,8-naphtalic anhydride (1.0 equiv.) and 3,4-dimethoxyaniline (1.2 equiv.) in ethanol were stirred at reflux for total of 48 h. The crude reaction mixture was cooled down and the solvent was removed under vacuum. The crude reaction mixture was poured into

cold water and precipitated as a yellow solid. The precipitate was dissolved in DCM and dried over anhydrous Na_2SO_4. The solvent was removed under vacuum to obtain 6-bromo-2-(3,4-dimethoxyphenyl)-1H-benzo[de]isochinoline-1,3-(2H)-dione (Int-2) as white solid (65 %). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.72 1 (dd, J = 8.0 Hz, 1.2 Hz, 1H), 8.65 (dd, J = 8.0 Hz 1.12 Hz, 1H), 8.47 (d, J = 8.0 Hz, 1H), 8.09 (d, J = 8.0 Hz, 1H), 7.91 (dd, J = 1.28 Hz, 8.0 Hz, 1H), 7.04 (d, J = 8.0 Hz, 1H), 6.90 (dd, J = 8.0 Hz, 2.36 Hz, 1H), 6.82 (d, J = 2.36 Hz, 1H), 3.95 (s, 3H), 3.88 (s, 3H). Mass Spec: $[M+H]^+$ = 414.15.

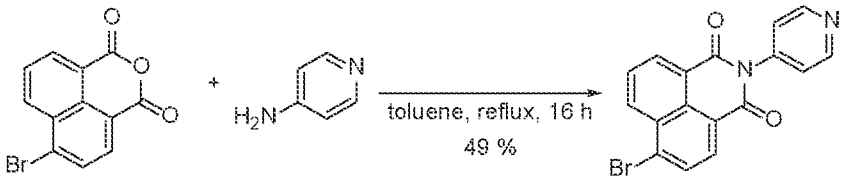
Synthesis of Intermediate 3 (Int-3)



Int-1 Int-3

[0166] 3-Bromo-7H-benzo[de]benzo[4,5]imidazole[2,1-a]isoquinolin-7-one (Int-1) (1.0 equiv) was treated with fuming H_2SO_4 at 0 °C. The cooling bath was removed, and the reaction mixture was stirred at 4 h. The crude was purified by RP (C18 column, H_2O:ACN, 6:4) to obtain the product (Int-3) as yellow solid (46 %). NMR shows two inseparable isomers cis and trans. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.69 (br t, 2H), 8.59 (td, J = 7.16 Hz, 1.2 Hz, 2H), 8.47 (dd, J = 8.57 Hz, 1.16 Hz, 1H), 8.36 (d, J = 8.0 Hz, 1H), 8.35 (d, J = 8.0 Hz, 1H), 8.31 (dd, J = 8.48 Hz, 1.2 Hz, 1H), 8.13 (d, J = 8.0 Hz, 1H), 8.08 (d, J = 8.0 Hz, 1H), 7.93 (d, J = 8.0 Hz, 1H), 7.91 (d, J = 8.0 Hz, 1H), 7.88 (d, J = 8.0 Hz, 1H), 7.86 (d, J = 8.0 Hz, 1H), 7.84 (d, J = 8.0 Hz, 1H), 7.81 (d, J = 8.0 Hz, 1H), 7.81 (m, 4H). Mass Spec: $[M-H]^-$ = 428.93.

Synthesis of Intermediate 4 (Int-4)

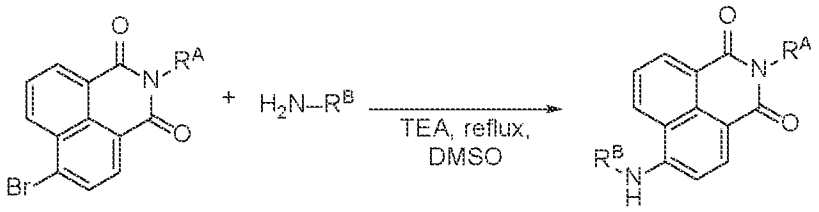


Int-4

[0167] Relevant 4-bromo-1,8-naphtalic anhydride (1.0 equiv) and 4-aminopyridine (1.2 equiv) in toluene were stirred at reflux for 16 h. The crude was cooled down, the solvent was removed under vacuum. The crude was precipitated from MeOH. The precipitate was dried under high vacuum to obtain Int-4 as light brown solid (49 %). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.84 (dd, J = 4.48 Hz, 1.64 Hz, 2H), 8.72 (dd, J = 8.0 Hz, 1.2 Hz, 1H), 8.69 (dd, J = 8.0 Hz, 1.2 Hz, 1H), 8.48 (d, J = 8.0 Hz, 1H), 8.12 (d, J = 8.0 Hz, 1H), 7.93 (dd, J = 8.52 Hz, 7.24 Hz, 1H), 7.32 (dd, J = 4.48 Hz, 1.64 Hz, 2H). Mass Spec: $[M+H]^+$ = 354.92.

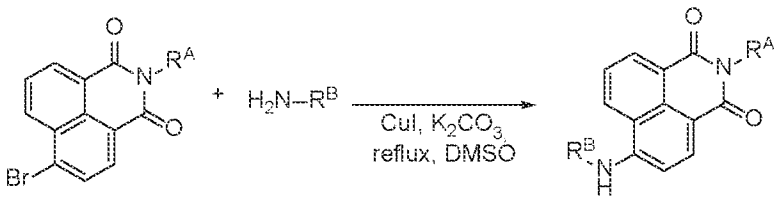
General Synthesis of naphthalimide dyes

General Procedure A



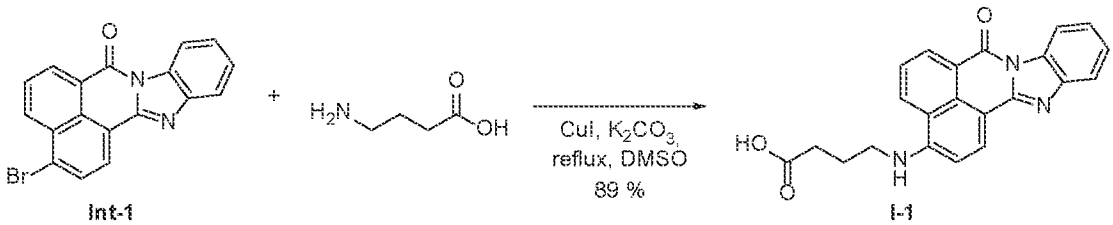
[0168] The relevant bromo-1,8-naphtalimide derivative (1.0 equiv) in DMSO was treated with relevant amine (10.0 equiv) and TEA (20.0 equiv). The reaction was stirred at reflux for 5 h. The crude was filtered and purified by prep HPLC.

General Procedure B



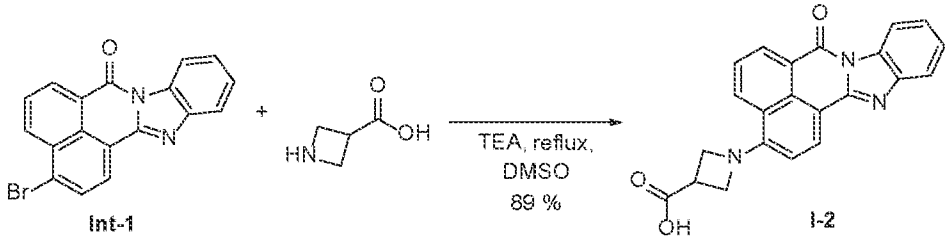
[0169] The relevant bromo-1,8-naphtalimide derivative (1.0 equiv) in DMSO was treated with CuI (0.3 equiv) then relevant amine (10.0 equiv) and K_2CO_3 (0.4 equiv) were added. The reaction was stirred at reflux for 5 h. The crude was extracted with DCM and the organic layer was washed with H_2O/brine, then dried over anhydrous Na_2SO_4. The solvent was removed under vacuum. The crude was purified by normal phase flash chromatography or filtered and purified by prep HPLC.

Synthesis of Compound I-1



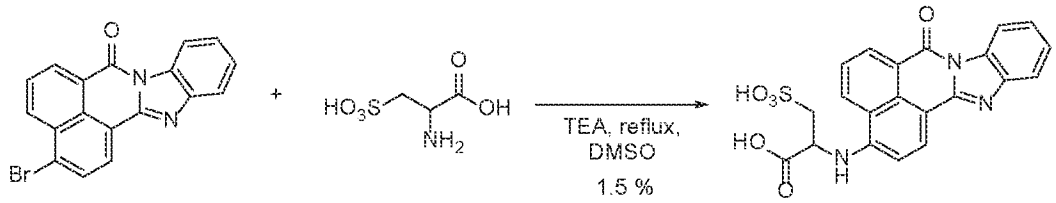
[0170] Based on General Procedure B, the crude was purified by prep HPLC to obtain I-1 as red solid (3.5 %). Mass Spec: $[M+H]^+$ = 372.27, $[M-H]^-$ = 370.28.

Synthesis of Compound I-2



[0171] Based on General Procedure A, the crude was purified by prep HPLC to obtain I-2 as red solid (89 %). Mass Spec: $[M+H]^+ = 370.28$, $[M-H]^- = 368.47$.

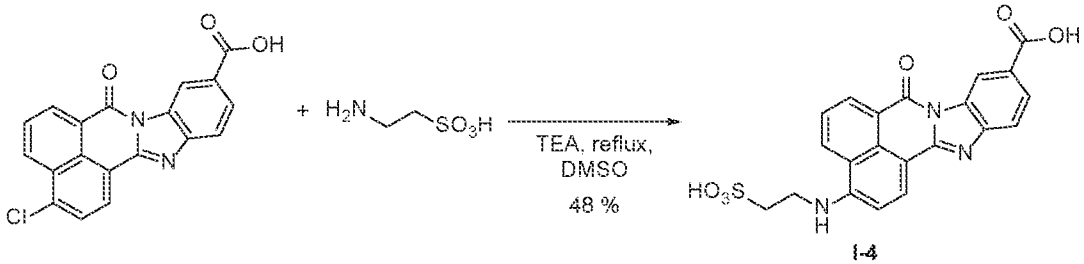
Synthesis of Compound I-3



Int-1 I-3

[0172] Based on General Procedure A, the crude was purified by prep HPLC to obtain I-3 as red solid (1.5 %). Mass Spec: $[M+H]^+ = 438.35$, $[M-H]^- = 436.37$.

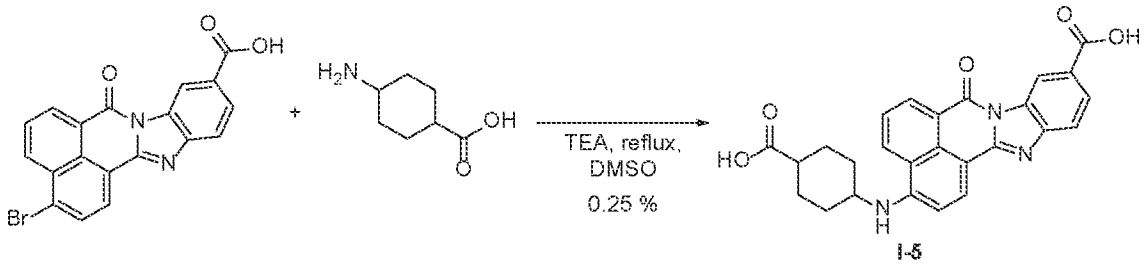
Synthesis of Compound I-4



I-4

[0173] Based on General Procedure A, the crude was purified by prep HPLC to obtain I-4 as red solid (48 %). Mass Spec: $[M+H]^+ = 438.52$, $[M-H]^- = 436.37$.

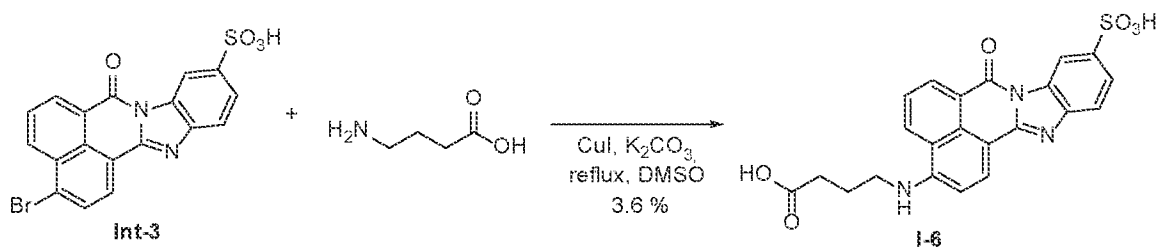
Synthesis of Compound I-5



I-5

[0174] Based on General Procedure A, the crude was purified by prep HPLC to obtain I-5 as red solid (48 %). Mass Spec: $[M+H]^+ = 438.52$, $[M-H]^- = 436.37$.

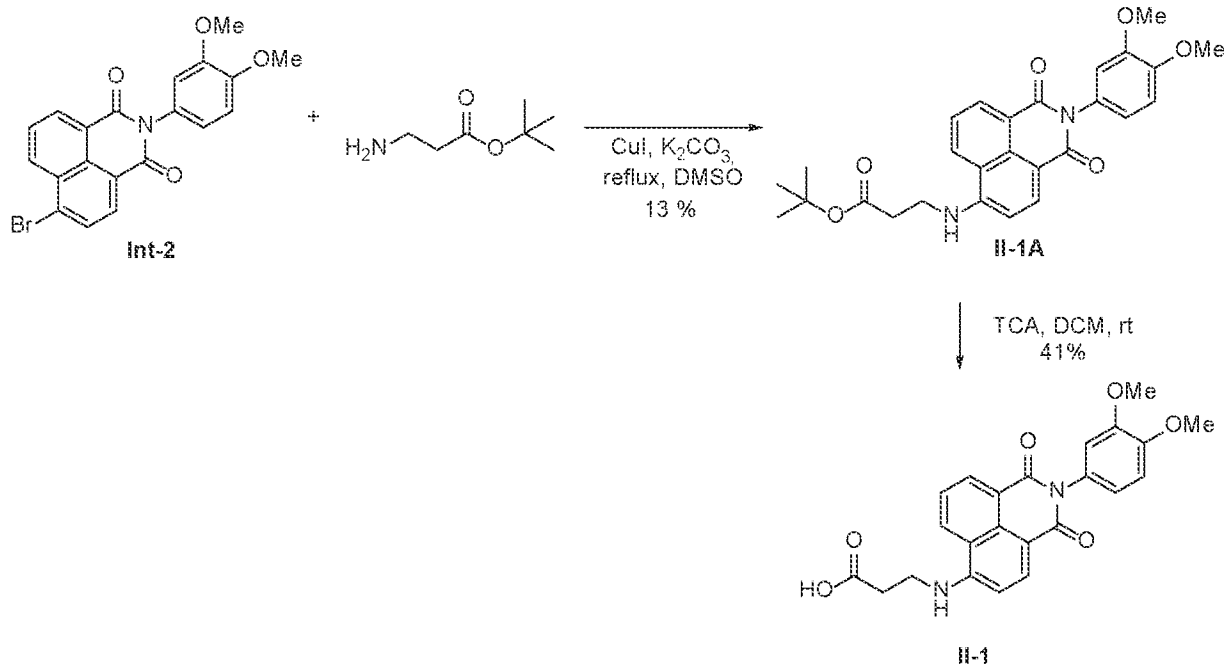
Synthesis of Compound I-6



Int-3 I-6

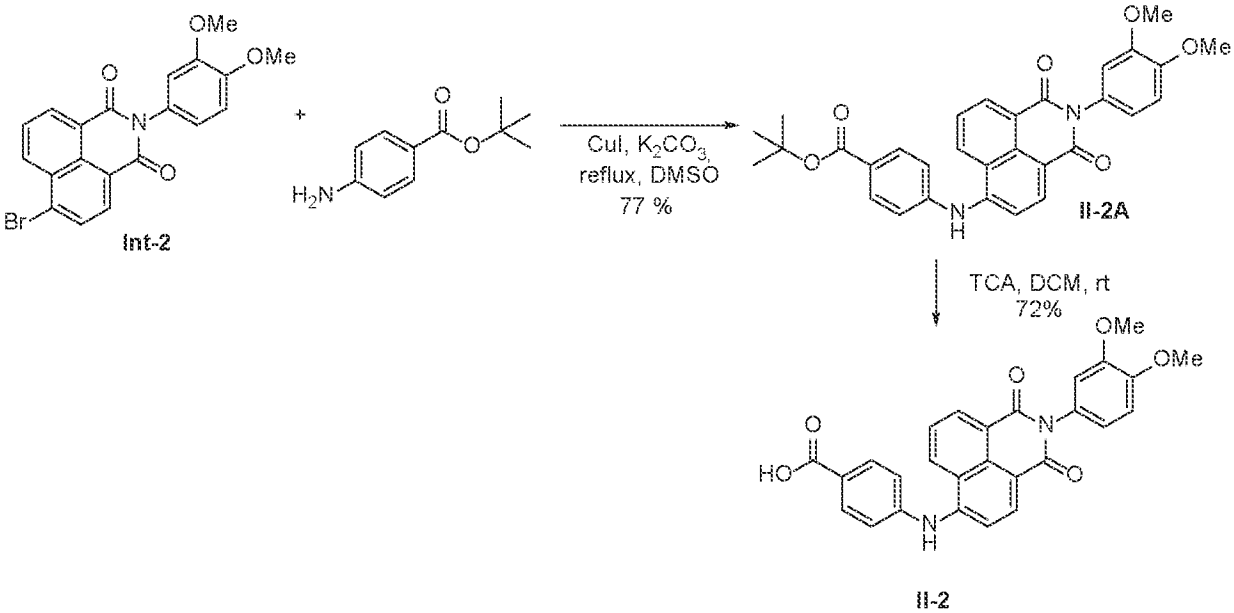
[0175] Based on General Procedure B, the crude was purified by prep HPLC to obtain I-6 as red solid (3.6 %). Mass Spec: $[M-H]^- = 450.16$.

Synthesis of Compound II-1



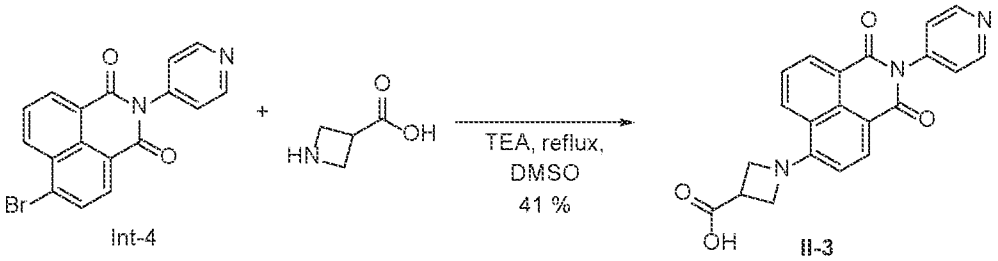
[0176] Based on General Procedure B, the crude reaction mixture was purified by NP (Silica column, DCM:EtOAc, 8:2) to obtain II-1A as yellow solid (13 %). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.61 1 (dd, J = 7.32 Hz, 1.08 Hz, 1H), 8.50 (d, J = 8.44 Hz, 1H), 8.16 (dd, J = 8.44 Hz, 1.12 Hz, 1H), 7.65 (dd, J = 8.60 Hz, 1.44 Hz, 1H), 7.00 (d, J = 8.52 Hz, 1H), 6.88 (dd, J = 8.44 Hz, 2.32 Hz, 1H), 6.86 (d, J = 8.0 Hz, 2.36 Hz, 1H), 6.82 (d, J = 2.28 Hz, 1H), 6.76 (d, J = 4.52 Hz, 1H), 6.18 (br t, J = 5.8 Hz, 1H), 3.92 (s, 3H), 3.85 (s, 3H), 3.71 (q, J = 5.88 Hz, 2H), 2.72 (q, J = 5.88 Hz, 2H), 1.48 (s, 9H). Mass Spec: $[M+H]^+$ = 377.37, $[M-H]^-$ = 375.26. II-1A was subsequently treated with trifluoroacetic acid in DCM at room temperature to obtain II-1 as yellow solid (41 %). Mass Spec: $[M+H]^+$ = 421.20, $[M-H]^-$ = 419.17.

Synthesis of Compound II-2



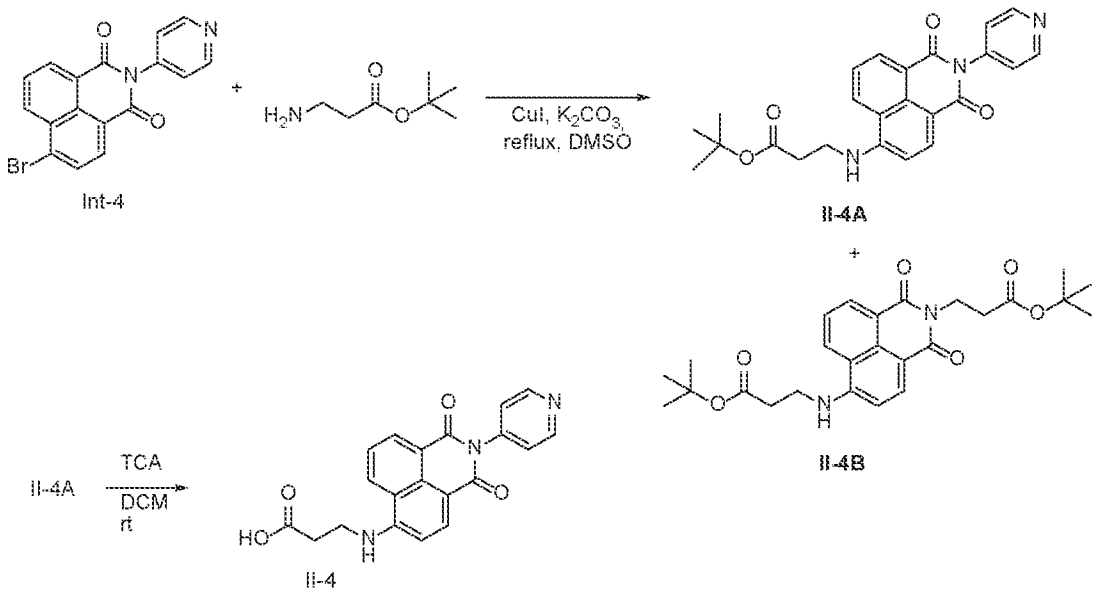
[0177] Based on General Procedure B, the crude was purified by NP (Silica column, DCM:EtOAc, 8:2) to obtain II-2A as yellow solid (77 %). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.64 (d, *J* = 6.52 Hz, 1H), 8.50 (d, *J* = 8.24 Hz, 1H), 8.40 (d, *J* = 8.44 Hz, 1H), 8.00 (d, *J* = 8.72 Hz, 2H), 7.68 (t, *J* = 8.52 Hz, 1H), 7.56 (d, *J* = 8.32 Hz, 1H), 7.40 (br s, 1H), 7.27 (d, *J* = 8.80 Hz, 2H), 6.94 (d, *J* = 4.52 Hz, 1H), 6.87 (dd, *J* = 8.44 Hz, 2.28 Hz, 1H), 6.80 (d, *J* = 2.32 Hz, 1H), 3.87 (s, 3H), 3.80 (s, 3H), 1.60 (s, 9H). Mass Spec: $[M+H]^+$ = 525.32, $[M-H]^-$ = 523.23. II-2A was subsequently treated with trifluoroacetic acid in DCM at room temperature to obtain II-2 as yellow solid (72 %). Mass Spec: $[M+H]^+$ = 469.14, $[M-H]^-$ = 467.33.

Synthesis of Compound II-3



[0178] Based on General Procedure A, II-3 was prepared from Int-4 and purified by prep HPLC as yellow solid (41%). Mass Spec: $[M+H]^+$ = 374.16, $[M-H]^-$ = 372.12.

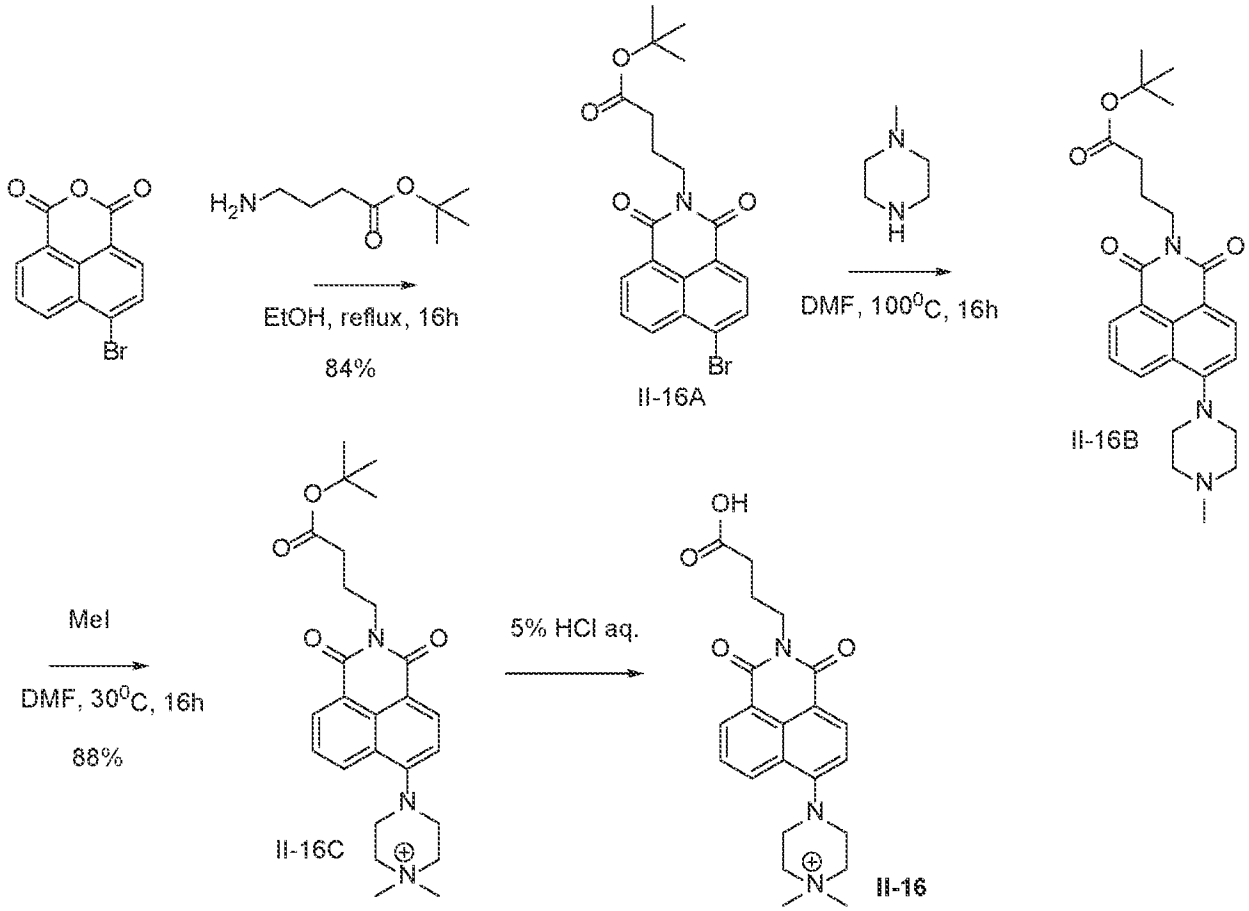
Synthesis of Compound II-4



[0179] Based on General Procedure B, the crude was purified by NP (Silica column, DCM:EtOAc, 8:2) two different products were obtained II-4A (45 %) and side product II-4B (41 %). ^{1}H NMR II-4A (400 MHz, $CDCl_3$): δ 8.77 (br s, 2H), 8.57 (d, *J* = 8.44 Hz, 1H), 8.47 (d, *J* = 8.44 Hz, 1H), 8.19 (d, *J* = 8.56 Hz, 1H), 7.64 (t, *J* = 7.64 Hz, 1H), 7.31 (br s, 2H), 6.75 (d, *J* = 8.52 Hz, 1H), 6.18 (br t, *J* = 5.84 Hz, 1H), 3.71 (q, *J* = 5.96 Hz, 2H), 2.72 (t, *J* = 6.32 Hz, 2H), 1.47 (s, 9H). Mass Spec: $[M+H]^+$ = 418.44, $[M-H]^-$ = 416.22. ^{1}H NMR II-4B (400 MHz, $CDCl_3$): δ 8.52 (dd, *J* = 7.32 Hz, 1.00 Hz, 1H), 8.41 (d, *J* = 8.44 Hz, 1H), 8.08 (dd, *J* = 8.56 Hz, 1.20 Hz,

1H), 7.58 (dd, J = 1.08 Hz, 7.32 Hz, 1H), 6.68 (d, J = 8.52 Hz, 1H), 6.10 (br t, J = 5.96 Hz, 1H), 4.43 (t, J = 7.60 Hz, 2H), 3.68 (q, J = 5.96 Hz, 2H), 2.70 (m, 4H), 1.46 (s, 9H), 1.41 (s, 9H). Mass Spec: $[M-H]^-$ = 467.28. Compound II-4 can be prepared from II-4A following similar procedure as described in the preparation of II-1 from II-1A.

Synthesis of Compound II-16



[0180] Triethylamine (1 mL) was added to the mixture of 4-bromo-1, 8-naphthalic anhydride (1 g, 3.6 mmol), tert-butyl 4-aminobutyrate hydrochloride and ethanol (780 mg, 3.97 mmol), then and the mixture was stirred and refluxed for 16 h. The solvent was removed under reduced pressure, 100 ml of water added to the crude material. The solid was filtered and vacuum dried to afford II-16A (yield 84%).

[0181] Compound II-16A (1 eq.) and 1-methylpiperazine (2.1 eq.) were dissolved in DMF and heated at 110°C for 16h.The mixture was poured in water and the solid was collected and dried to afford II-16B (yield 87%).

[0182] Compound II-16B (90 mg, 1 eq.) was dissolved in 2 mL of DMF. MeI (20 eq.) was added to the mixture. After 72h stirring at 35°C, the mixture was poured on cold Et_2O. The solid was filtered and dried to afford II-16C (yield 33%).

[0183] Compound II-16C (30mg) was dissolved in 5 mL of 5% aqueous HCl. After 72h, the solvent was evaporated and dried under high vacuum to afford compound II-16 with

quantitative yield. ^{1}H NMR (400 MHz, DMSO) δ 8.56 – 8.45 (m, 3H), 7.92 – 7.83 (m, 1H), 7.51 (d, J = 8.1 Hz, 1H), 4.09 (t, J = 7.0 Hz, 2H), 3.76 (s, 4H), 3.60 (s, 4H), 3.29 (s, 6H), 2.31 (t, J = 7.5 Hz, 2H), 1.90 (q, J = 7.3 Hz, 2H).

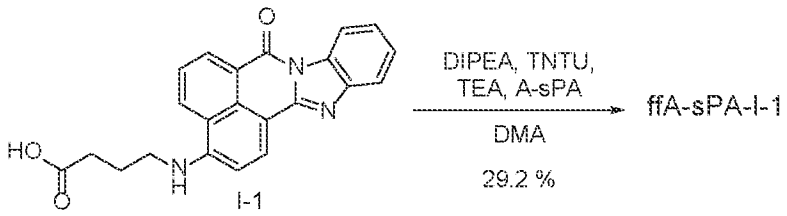
[0184] The fluorescent spectra of some exemplary dyes disclosed herein are summarized in Table 1 below.

Table 1. Absorption and Emission Spectra of Compounds of Formulas (I) and (II)

Dye	$\lambda_{max\ abs}$ [nm]	$\lambda_{max\ em}$ [nm]	Stokes shift SS [nm]	Solvent
I-1	461	508	47	MeOH
I-2	456	514	58	MeOH
I-3	440	489	49	ACN
I-4	457	512	55	MeOH
I-5	468	510	42	MeOH
	464	508	44	ACN
I-6	468	519	51	MeOH
	460	509	49	ACN
I-7	444	540	96	MeOH
II-1	451	---	---	MeOH
II-2	450	546	96	MeOH
II-16	387	524	137	EtOH

Example 2. Preparation of fully functionalized nucleotides coupled with naphthalimide dyes

General Procedure D (Shown for I-1)



[0185] A 25-mL round-bottomed flask was charged, under nitrogen, with a solution of the dye in the form of a free acid (10 μmol) in N,N-dimethylacetamide (DMA, 3 mL) and N,N-diisopropylamine (DIPEA) (10 eq.) was added along with 2-(5-Norborene-2,3-dicarboximido)-1,1,3,3-tetramethyluronium tetrafluoroborate (TNTU, 1.2 eq.). Reaction was left under stirring under nitrogen at room temperature and monitored by means of UPLC and TLC. Upon conversion to desired intermediate, sPA-A (1 eq) was added to the stirring solution with triethylamine (7.2 eq.). Reaction was left under stirring at room temperature under inert atmosphere for 8 hours. After complete consumption of intermediate and formation of desired material, reaction was quenched by the addition of 20 mL of 0.1M TEAB. Reaction was loaded onto a Sephadex column and processed with a gradient of 1M TEAB in 0.1M TEAB. Product containing fractions were then processed onto a 60g C18 cartridge with a gradient of 0.1M TEAB in ACN. Finally, product was purified by Preparative HPLC using a gradient of ACN in 0.1M TEAB. Nature and purity of

product were confirmed by UPLC and LC-MS. Yield was determined by OD of a dilution as summarized below:

[0186] **ffA-sPA-I-1** was prepared with 29.2% yield. Mass Spec: $[M-H]^-$ = 1297.84, $[M-2H]^{2-}$ =648.60, $[M+H]^+$ = 1301.60, $[M+2H]^{2+}$ = 650.70.

[0187] **ffA-sPA-I-2** was prepared with 15.2% yield. Mass Spec: $[M-H]^-$ = 1295.81, $[M-2H]^{2-}$ =647.99, $[M+H]^+$ = 1298.15, $[M+2H]^{2+}$ = 649.65.

[0188] **ffA-sPA-I-3** was prepared with 48.3% yield. Mass Spec: $[M-H]^-$ = 1364.57, $[M-2H]^{2-}$ =681.78, $[M+H]^+$ = 1366.55, $[M+2H]^{2+}$ = 683.76.

[0189] **ffA-sPA-I-6** was prepared with 22.8% yield. Mass Spec: $[M-H]^-$ = 1377.43, $[M-2H]^{2-}$ =688.43.

[0190] **ffC-sPA-II-1** was prepared with 57.9% yield. Mass Spec: $[M-H]^-$ = 1324.03, $[M-2H]^{2-}$ =661.50, $[M+H]^+$ = 1326.06, $[M+2H]^{2+}$ = 663.41.

[0191] The fluorescent spectra of exemplary ffNs disclosed herein are summarized in Table 2 below.

Table 2. Absorption and Emission Spectra of ffN conjugated with dyes of Formula (I) and (II)

ffN	$\lambda_{max\ abs}$ [nm]	$\lambda_{max\ em}$ [nm]	Stokes shift SS [nm]	Solvent
ffA-sPA-I-1	474	514	47	H_2O
ffA-sPA-I-2	465	522	47	H_2O
ffA-sPA-I-3	440	500	19	H_2O
ffA-sPA-I-6	478	525	43	H_2O
ffC-sPA-II-1	450	538	46	H_2O

Example 3. Sequencing experiments on Illumina MiSeq™ Platform

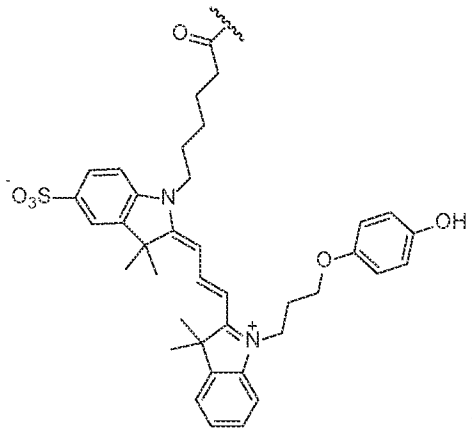
[0192] The ffA labeled with the naphthalimide dyes described herein were tested on an Illumina MiSeq™ instrument, which was set up to either take the first image with a blue excitation light (~ 450 nm) and the second image with a blue excitation light (~ 450 nm) or take the first image with a blue excitation light (~ 450 nm) and the second image with a green excitation light (~ 532 nm). Images were taken simultaneously through collection channels which are in blue (472-520nm) and green (540-640nm).

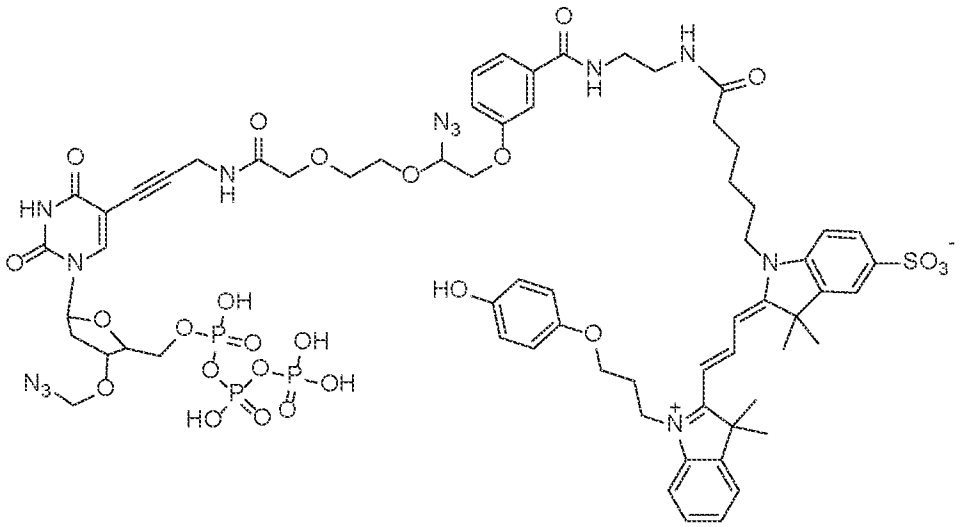
[0193] Sequencing Run A (2ex-2ch blue/green MiSeq™): For **FIG. 1A**, the incorporation mix used this experiment include the following five 3′-AZM blocked ffNs: 2 μM ffT labeled with a green cyanine dye A (ffT-LN3-dye A), 1 μM ffC labeled with a blue coumarin dye B (ffC-sPA-dye B), 1 μM ffC labeled with a known cyanine dye (ffC-sPA-NR550S0), 2 μM ffA labeled with a naphthalimide dye I-1 (ffA-sPA-I-1), and 2 μM unlabeled ffG (dark G) in 50

mM glycine buffer, pH 9.85, 50 mM NaCl, 1 mM EDTA, 0.2% CHAPS, 4 mM $MgSO_4$ and a DNA polymerase. The scatterplot was generated at cycle 26 at 22°C.

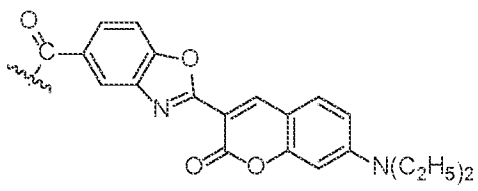
[0194] Sequencing Run B (1ex-2ch blue/blue MiSeq™): For **FIG. 1B**, the incorporation mix used this experiment include the following four3′-AZM blocked ffNs: 2 μM ffT labeled with a blue chromenoquinoline dye C (ffT-LN3-dye C), 2 μM ffC labeled with a blue coumarin dye D (ffC-sPA-dye D), 2 μM ffA labeled with a naphthalimide dye I-1 (ffA-sPA-I-1), and 2 μM unlabeled ffG (dark G) in 50 mM glycine buffer, pH 9.6, 50 mM NaCl, 1 mM EDTA, 0.2% CHAPS, 4 mM $MgSO_4$ and a DNA polymerase. The scatterplot was generated at cycle 26 at 22°C.

[0195] The standard incorporation mix used on the MiSeq™ platform was used as a control for comparing the primary sequencing metrics, including ffA labeled with a coumarin dye B (ffA-LN3-BL-dye B).

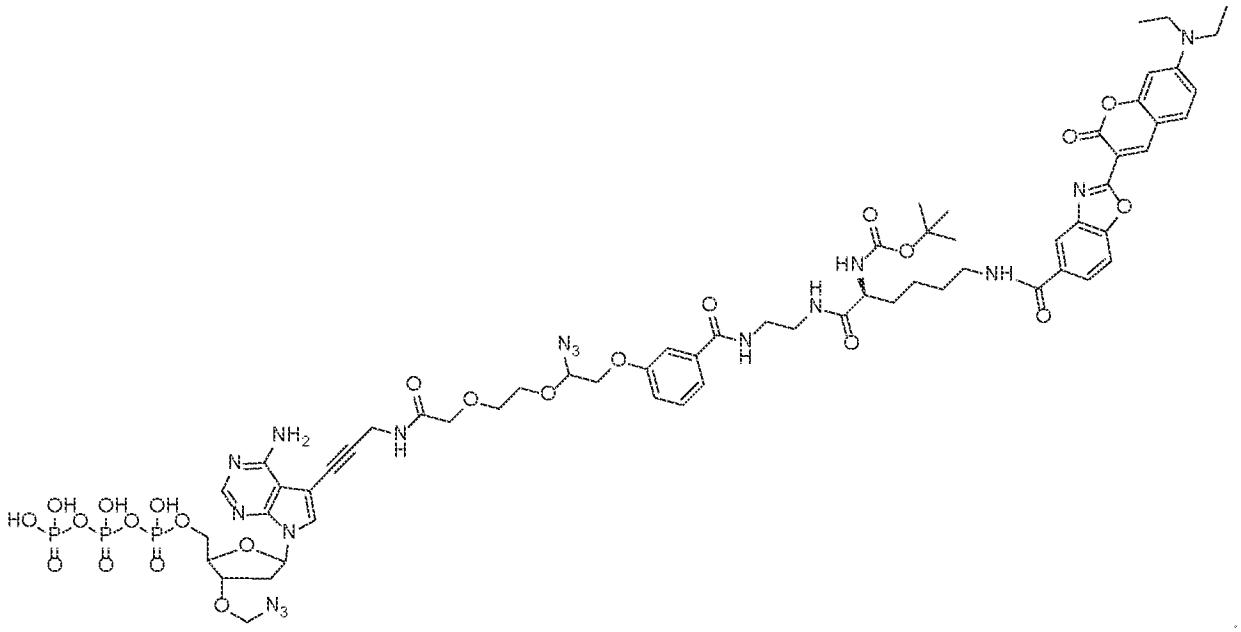
[0196] Cyanine dye A is disclosed in U.S. Publication No. 2018/0282791 A1, having the structure moiety  when conjugated with the ffT. The structure of ffT-LN3-cyanine dye A is:



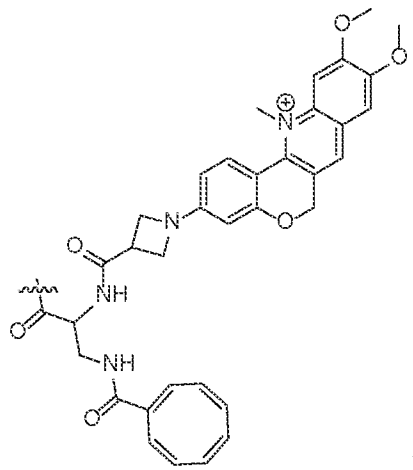
[0197] Coumarin dye B has strong fluorescence and great stability. This dye is disclosed in U.S. Publication No. 2020/0277670 A1, having the structure moiety



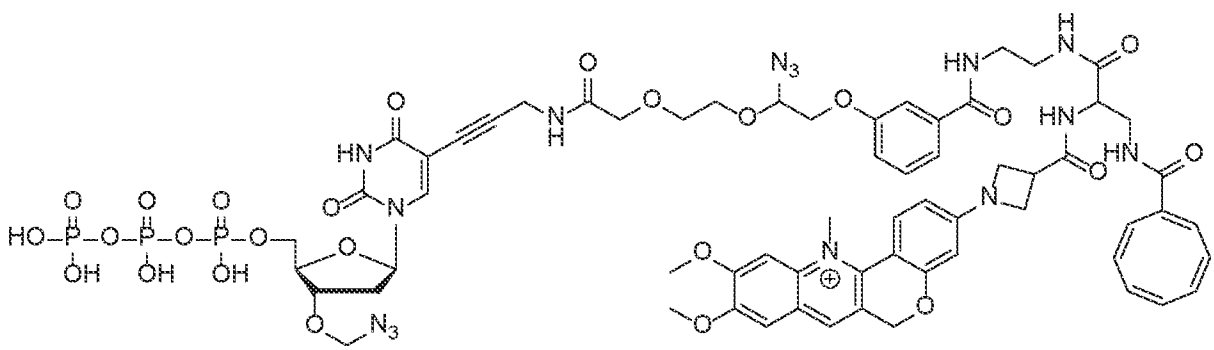
when conjugated with the ffA. The structure of ffA-LN3-BL-dye B is:



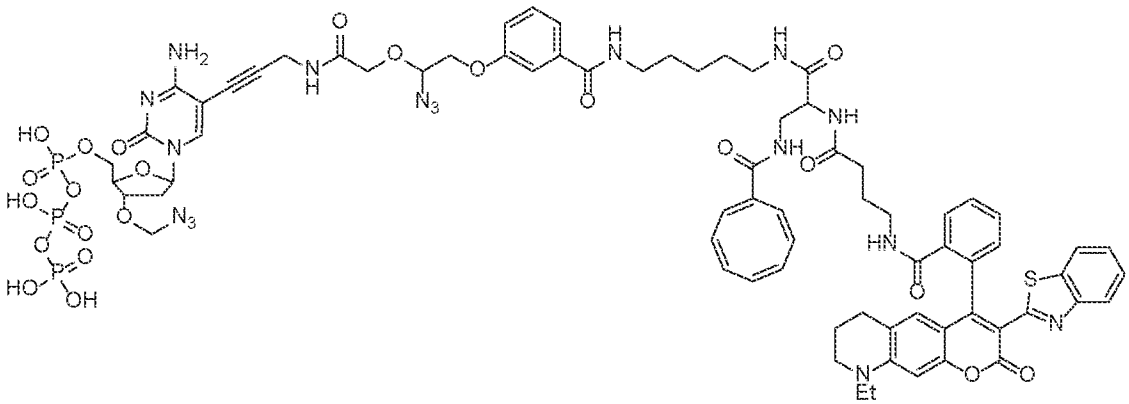
[0198] Chromenoquinoline dye C is disclosed in U.S. Pub. No. 2023/0313292, having the structure moiety:



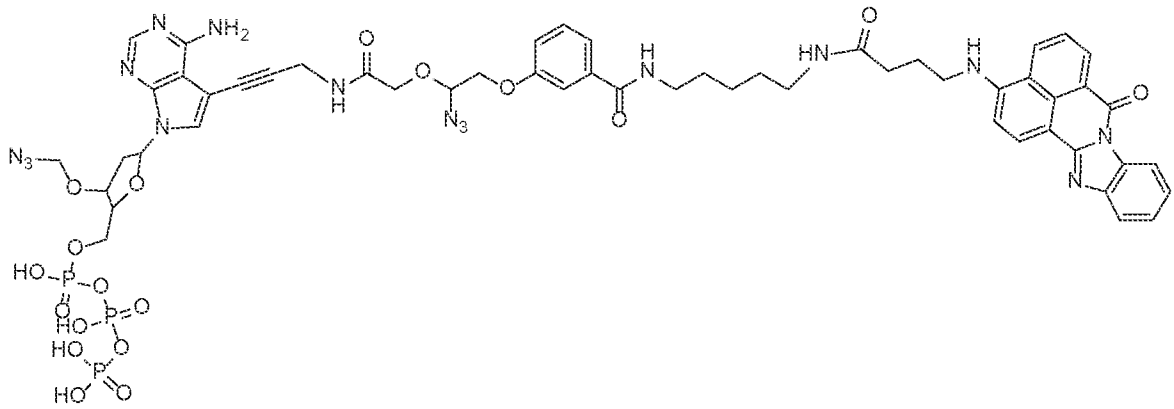
when conjugated with the ffT. The structure of ffT-LN3-dye C is:



[0199] Coumarin dye D is disclosed in U.S. Publication No. 2018/0094140, which is incorporated by reference. The structure of ffC-sPA- dye D is:



[0200] The structure of ffA-sPA-I-1 is:

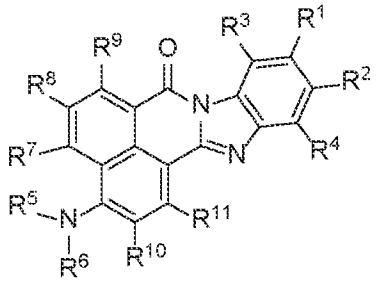


[0201] **FIGs. 2A**, **2B**, and **2C** demonstrate the percent remaining signal, error rate and percent phasing of the comparative sequencing run A over 151 cycles using the ffN sets with either (ffA-LN3-BL-dye B) or ffA-sPA-I-1 respectively, where different dosages of blue laser at 1x, 3x, 5x, 7.5x and 10x were assigned to different areas of the same flow cell. For percent signal decay, the observation was also made in absence of light and photodamage at 0x light dosage. The phasing values were from 0.104 to 0.147 for using ffN set with ffA-coumarin dye C and from 0.093 to 0.132for using ffN set with ffA-sPA-I-1. The average prephasing values, which was not light dependent, were 0.197 and 0.093 respectively.

[0202] **FIGs. 3A**, **3B**, and **3C** demonstrate the percent signal decay, error rate and percent phasing of the comparative sequencing run B over 151 cycles using the ffN sets with either ffA-LN3-BL-dye B or ffA-spA-I-1 respectively, where different dosages of blue laser at 1x, 3x, 5x, 7.5x and 10x were assigned to different areas of the same flow cell. For percent signal decay, the observation was also made in absence of light and photodamage at 0x light dosage. Again, it was observed that ffA-sPA-dye I-1 provided improved signal decay and phasing value as compared to ffA-LN3-BL-dye B.

WHAT IS CLAIMED IS:

1. A compound of Formula (I):



(I), a salt or a mesomeric form thereof,

wherein each of R^1, R^2, R^3, R^4, R^7, R^8, R^9, R^{10}, and R^{11} is independently H, unsubstituted or substituted C_1-C_6 alkyl, C_1-C_6 alkoxy, unsubstituted or substituted C_2-C_6 alkenyl, unsubstituted or substituted C_2-C_6 alkynyl, C_1-C_6 haloalkyl, C_1-C_6 haloalkoxy, C_1-C_6 hydroxyalkyl, (C_1-C_6 alkoxy)C_1-C_6 alkyl, -O-(C_1-C_6 alkoxy)C_1-C_6 alkyl, unsubstituted or substituted amino, halo, cyano, carboxy, hydroxy, nitro, sulfonyl, sulfino, sulfo, sulfonate, S-sulfonamido, N-sulfonamido,–C(=O)OR^A, –C(=O)NR^BR^C or –SO_3R^D;

each of R^A, R^B, R^C, and R^D is independently H, or unsubstituted or substituted C_1-C_6 alkyl; and

each of R^5 and R^6 is independently H, unsubstituted or substituted C_1-C_6 alkyl; unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

alternatively, R^5 and R^6 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl;

provided that the compound of Formula (I) comprises one carboxy group.

2. The compound of claim 1, wherein each of R^1 and R^2 is independently H, unsubstituted C_1-C_6 alkyl, C_1-C_6 alkoxy, –C(=O)OR^A, or –SO_3R^D.

3. The compound of claim 1 or 2, wherein R^1 is H, methoxy, –C(=O)OH or –SO_3H, and R^2 is H or methoxy.

4. The compound of claim 1 or 2 , wherein R^2 is H, methoxy, –C(=O)OH or –SO_3H, and R^1 is H or methoxy.

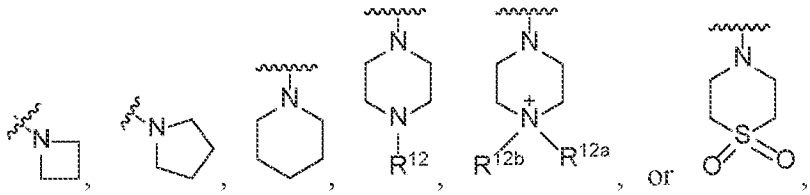
5. The compound of any one of claims 1 to 4, wherein each of R^3 and R^4 is H.

6. The compound of any one of claims 1 to 5, wherein each of R^5 and R^6 is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$.

7. The compound of claim 6, wherein R^5 is H or unsubstituted C_1-C_6 alkyl, and R^6 is C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$.

8. The compound of any one of claims 1 to 5, wherein R^5 is H and R^6 is unsubstituted C_3-C_7 cycloalkyl, unsubstituted phenyl, or C_3-C_7 cycloalkyl or phenyl each independently substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy,
$–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$.

9. The compound of any one of claims 1 to 5, wherein R^5 and R^6 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl.

10. The compound of claim 9, wherein R^5 and R^6 together with the nitrogen atom to which they are attached form , each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$, wherein each of R^{12}, R^{12a} and R^{12b} is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, $–C(=O)OR^A$, $–C(=O)NR^BR^C$ and $–SO_3R^D$.

11. The compound of claim 10, wherein each of R^{12}, R^{12a} and R^{12b} is independently unsubstituted C_1-C_6 alkyl.

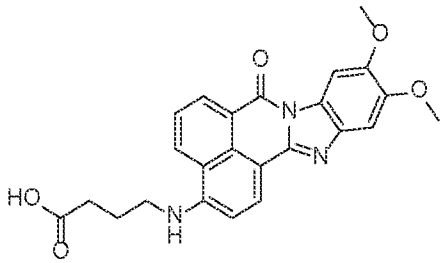
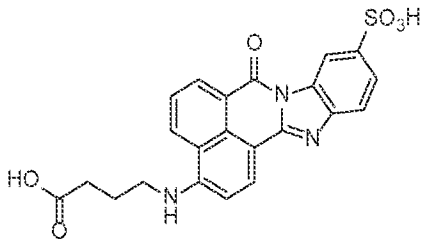
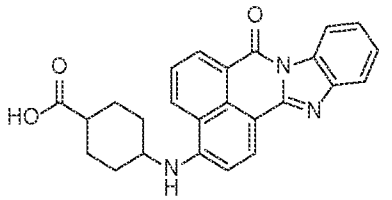
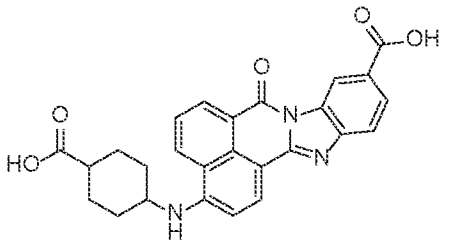
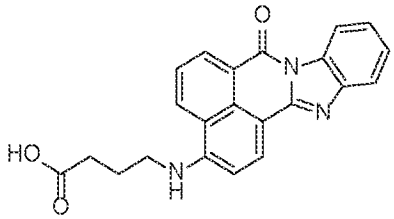
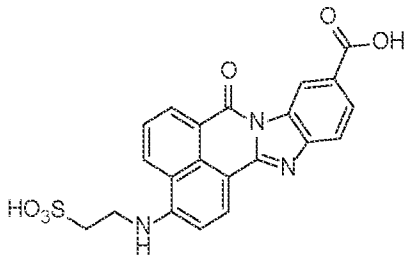
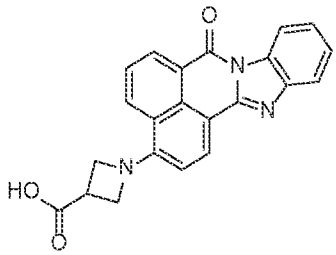
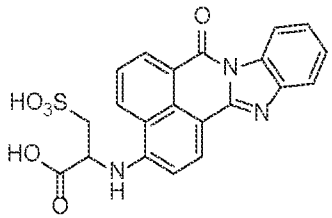
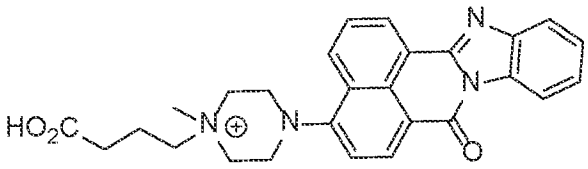
12. The compound of claim 10, wherein R^{12a} is unsubstituted C_1-C_6 alkyl and R^{12b} is C_1-C_6 alkyl substituted with $C(=O)OR^A$,$–C(=O)NR^BR^C$ or $–SO_3R^D$.

13. The compound of any one of claims 1 to 12, wherein each of R^7, R^8, R^9, R^{10}, and R^{11} is H.

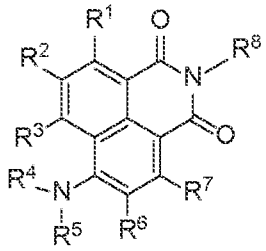
14. The compound of any one of claims 1 to 13, wherein each of R^A, R^B, R^C, and R^D is H.

15. The compound of any one of claims 1 to 13, wherein R^B is H and R^C is unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of $–C(=O)OH$ and$–SO_3H$.

16. The compound of claim 1, selected from the group consisting of:


,

,

,

,

,

,

,

,

, and salts and mesomeric forms thereof.

17. A compound of Formula (II):



(II), a salt or a mesomeric form thereof,

wherein each of R^1, R^2, R^3, R^6, and R^7 is independently H, unsubstituted or substituted C_1-C_6 alkyl, C_1-C_6 alkoxy, unsubstituted or substituted C_2-C_6 alkenyl, unsubstituted or substituted C_2-C_6 alkynyl, C_1-C_6 haloalkyl, C_1-C_6 haloalkoxy, C_1-C_6 hydroxyalkyl, (C_1-C_6 alkoxy)C_1-C_6 alkyl, -O-(C_1-C_6 alkoxy)C_1-C_6 alkyl, unsubstituted or substituted amino, halo, cyano, carboxy, hydroxy, nitro, sulfonyl, sulfino, sulfo, sulfonate, S-sulfonamido, N-sulfonamido, $-C(=O)OR^A$, $-C(=O)NR^BR^C$ or $-SO_3R^D$;

R^8 is unsubstituted or substituted C_1-C_6 alkyl, unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10

membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

each of R^A, R^B, R^C, and R^D is independently H, or unsubstituted or substituted C_1-C_6 alkyl; and

each of R^4 and R^5 is independently H, unsubstituted or substituted C_1-C_6 alkyl; unsubstituted or substituted C_3-C_7 cycloalkyl, unsubstituted or substituted C_6-C_{10} aryl, unsubstituted or substituted 5 to 10 membered heteroaryl, unsubstituted or substituted 3 to 10 membered heterocyclyl, unsubstituted or substituted (C_3-C_7 cycloalkyl)C_1-C_6 alkyl, unsubstituted or substituted (C_6-C_{10} aryl)C_1-C_6 alkyl, unsubstituted or substituted (5 to 10 membered heteroaryl)C_1-C_6 alkyl, unsubstituted or substituted (3 to 10 membered heterocyclyl)C_1-C_6 alkyl;

alternatively, R^4 and R^5 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl;

provided that the compound of Formula (II) comprises one carboxy group.

18. The compound of claim 17, wherein each of R^1, R^2, R^3, R^6, and R^7 is independently H, unsubstituted C_1-C_6 alkyl, C_1-C_6 alkoxy, –C(=O)OR^A, or –SO_3R^D.

19. The compound of claim 18, wherein each of R^1, R^2, R^3, R^6, and R^7 is H.

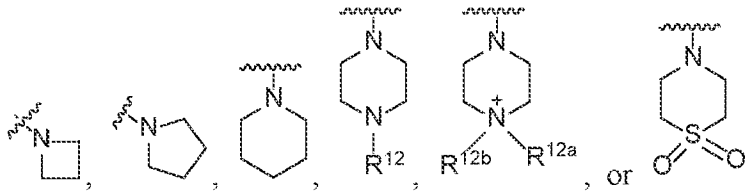
20. The compound of any one of claims 17 to 19, wherein each of R^4 and R^5 is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

21. The compound of claim 20, wherein both R^4 and R^5 are H or unsubstituted C_1-C_6 alkyl.

22. The compound of claim 20, wherein R^4 is H or unsubstituted C_1-C_6 alkyl, and R^5 is C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

23. The compound of any one of claims 17 to 22, wherein R^4 is H and R^5 is unsubstituted C_3-C_7 cycloalkyl, unsubstituted phenyl, or C_3-C_7 cycloalkyl or phenyl each independently substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

24. The compound of any one of claims 17 to 19, wherein R^4 and R^5 together with the nitrogen atom to which they are attached form an unsubstituted or substituted 3 to 10 membered heterocyclyl.

25. The compound of claim 24, wherein R^4 and R^5 together with the nitrogen atom to which they are attached form , each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D, wherein each of R^{12}, R^{12a} and R^{12b} is independently H, unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

26. The compound of claim 10, wherein each of R^{12}, R^{12a} and R^{12b} is independently unsubstituted C_1-C_6 alkyl.

27. The compound of claim 10, wherein R^{12a} is unsubstituted C_1-C_6 alkyl and R^{12b} is C_1-C_6 alkyl substituted with –C(=O)OR^A, –C(=O)NR^BR^C or –SO_3R^D.

28. The compound of any one of claims 17 to 27, wherein R^8 is C_1-C_6 alkyl unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

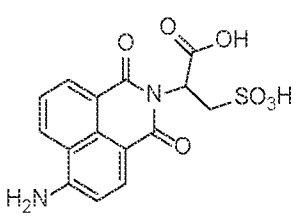
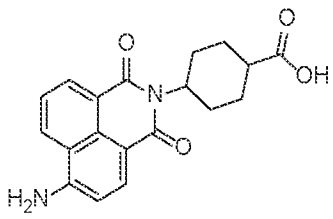
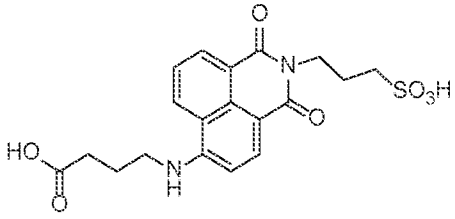
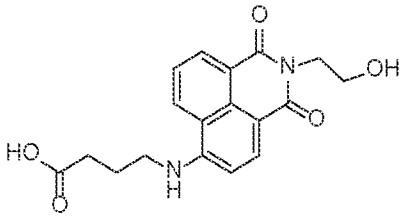
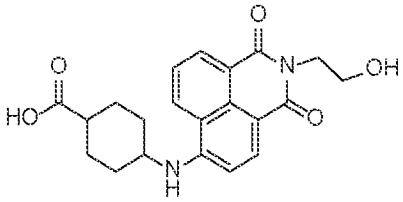
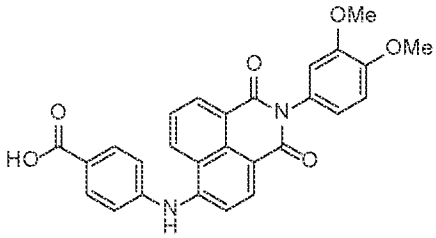
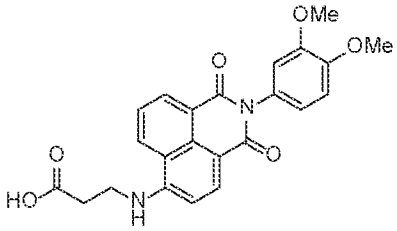
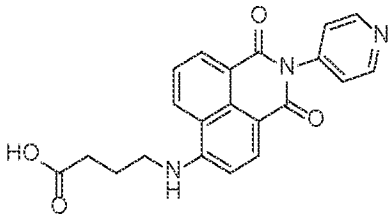
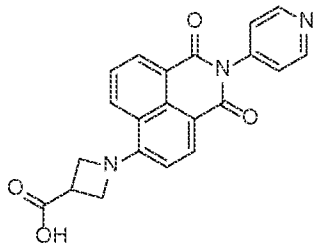
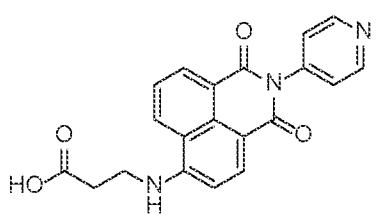
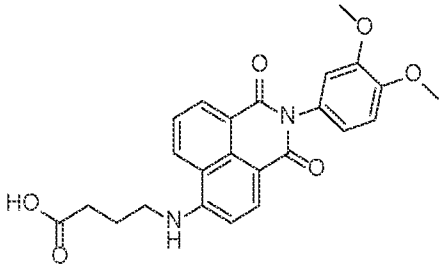
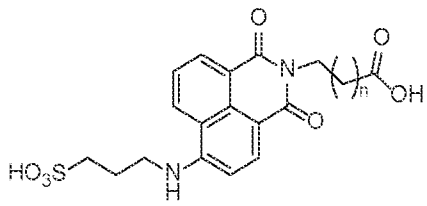
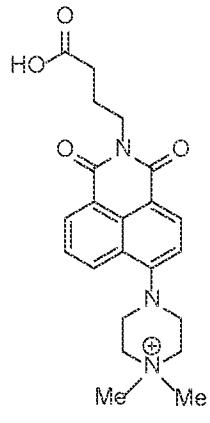
29. The compound of any one of claims 17 to 27, wherein R^8 is phenyl, pyridyl, cyclopentyl, or cyclohexyl, each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

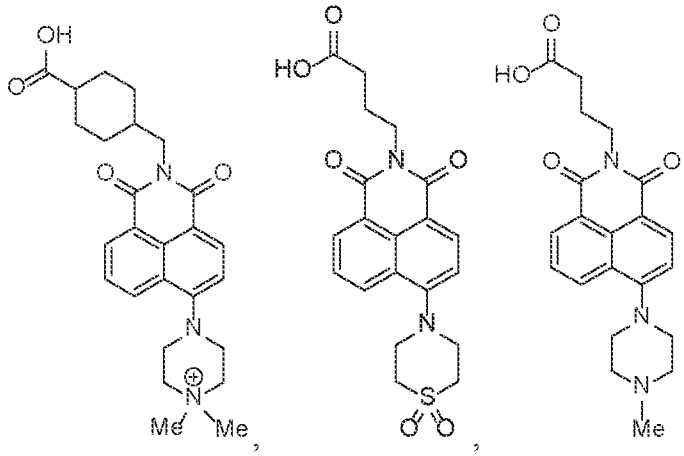
30. The compound of any one of claims 17 to 27, wherein R^8 is (phenyl)C_1-C_6 alkyl, (pyridyl)C_1-C_6 alkyl, (cyclopentyl)C_1-C_6 alkyl, or (cyclohexyl)C_1-C_6 alkyl, each independently unsubstituted or substituted with one or more substituents selected from the group consisting of –OH, halo, cyano, C_1-C_6 alkoxy, –C(=O)OR^A, –C(=O)NR^BR^C and –SO_3R^D.

31. The compound of any one of claims 17 to 30, wherein each of R^A, R^B, R^C, and R^D is H.

32. The compound of any one of claims 17 to 30, wherein R^B is H and R^C is unsubstituted C_1-C_6 alkyl, or C_1-C_6 alkyl substituted with one or more substituents selected from the group consisting of –C(=O)OH and–SO_3H.

33. The compound of claim 17, selected from the group consisting of:


,

,

,

,

,

,

,

,

,

,

,

,

,



, wherein n is 0, 1, 2, 3, 4, or 5; and salts and mesomeric forms thereof.

34. A nucleotide labeled with a compound according to any one of claims 1 to 33.

35. The labeled nucleotide of claim 34, wherein the compound is attached to the nucleotide via a carboxyl group of the compound of Formula (I) or the compound of Formula (II).

36. The labeled nucleotide of claim 34 or 35, wherein the compound is attached to the C5 position of a pyrimidine base or the C7 position of a 7-deaza purine base through a linker moiety.

37. The labeled nucleotide of any one of claims 34 to 36, further comprising a 3′ blocking group covalently attached to the ribose or 2′ deoxyribose of the nucleotide.

38. An oligonucleotide or polynucleotide comprising the labeled nucleotide according to any one of claims 34 to 37 incorporated thereto.

39. The oligonucleotide or polynucleotide of claim 38, wherein the oligonucleotide or polynucleotide is at least partially complementary and hybridized to a target polynucleotide immobilized on a surface of a solid support.

40. The oligonucleotide or polynucleotide of claim 39, wherein the solid support comprises an array of a plurality of target polynucleotides immobilized thereon.

41. A kit comprising a first type of labeled nucleotide according to any one of claims 34 to 37.

42. The kit of claim 41, comprising four types of nucleotides, wherein a second type of nucleotide carries a second label, a third type of nucleotide carries a third label, and a fourth type of nucleotide is unlabeled (dark), and wherein each of label has a distinct emission maximum that is distinguishable from the other labels.

43. The kit of claim 41, comprising four types of nucleotides, wherein a second type of nucleotide carries a second label, a third type of nucleotide comprises a mixture of third type of nucleotides carry two labels, and a fourth type of nucleotide is unlabeled (dark).

44. The kit according to any one of claims 41 to 43, further comprising a DNA polymerase and one or more buffer compositions.

45. A method of determining the sequences of a plurality of target polynucleotides, comprising:

(a) contacting a solid support with a solution comprising sequencing primers under hybridization conditions, wherein the solid support comprises a plurality of different target polynucleotides immobilized thereon; and the sequencing primers are complementary to at least a portion of the target polynucleotides;

(b) contacting the solid support with an aqueous solution comprising DNA polymerase and one or more of four different types of nucleotides under conditions suitable for DNA polymerase-mediated primer extension, and incorporating one type of nucleotides into the sequencing primers to produce extended copy polynucleotides, wherein at least one type of nucleotide is a labeled nucleotide of any one of claims 34 to 37, and wherein each of the one or more of four different type of nucleotides comprises a 3′ blocking group covalently attached to the 2′ deoxyribose of the nucleotide;

(c) imaging and performing one or more fluorescent measurements of the extended copy polynucleotides; and

(d) removing the 3′ blocking group from the nucleotides incorporated into the extended copy polynucleotides.

46. The method of claim 45, wherein the label and the 3′ blocking group from the nucleotides incorporated into the extended copy polynucleotides are removed in a single chemical reaction.

47. The method of claim 45 or 46, further comprising: (e) washing the solid support after said removing of the 3′ blocking group from the incorporated nucleotides.

48. The method of claim 47, further comprising repeating steps (b) to (e) until the sequences of at least a portion of the target polynucleotides are determined.

49. The method of claim 48, wherein steps (b) to (e) are repeated at least 50, 100, 150, 200, 250, or 300 cycles.

50. The method of any one of claims 45 to 49, wherein the method is performed on an automated sequencing instrument comprising a single light source.

51. The method of claim 50, wherein the single light source operates at a wavelength from about 450 nm to about 460 nm.

52. The method of any one of claims 45 to 49, wherein the method is performed on an automated sequencing instrument comprising two light sources operating at two different wavelengths.

53. The method of claim 52, wherein one light source operates at a wavelength from about 450 nm to about 460 nm, and the other light source operates at a wavelength from about 520 nm to about 540 nm.

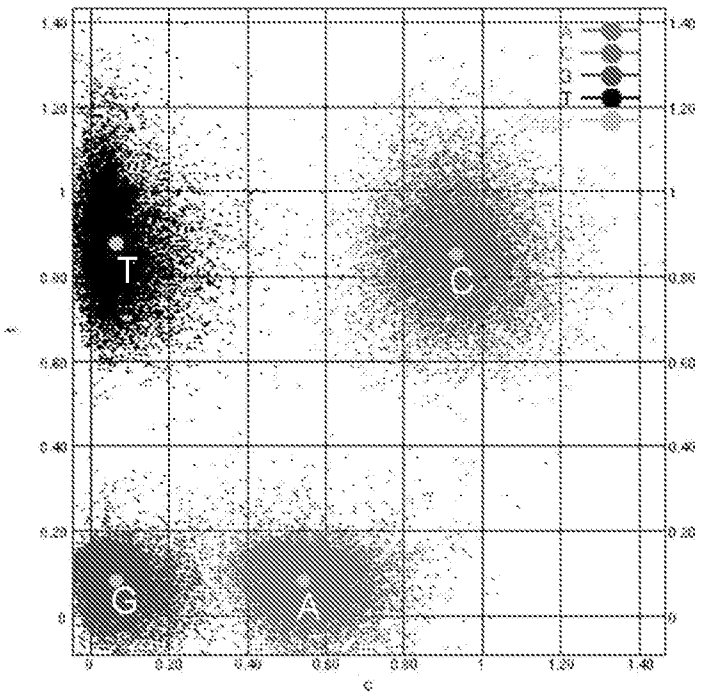


FIG. 1A

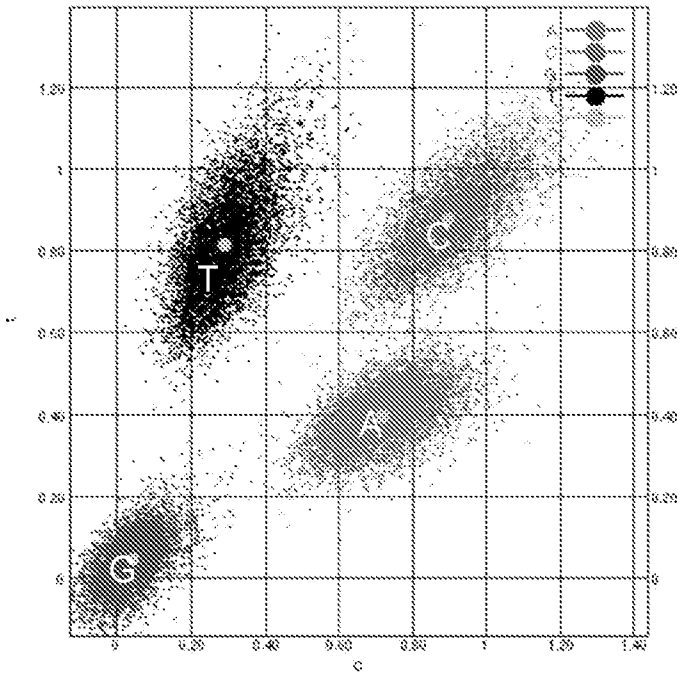


FIG. 1B

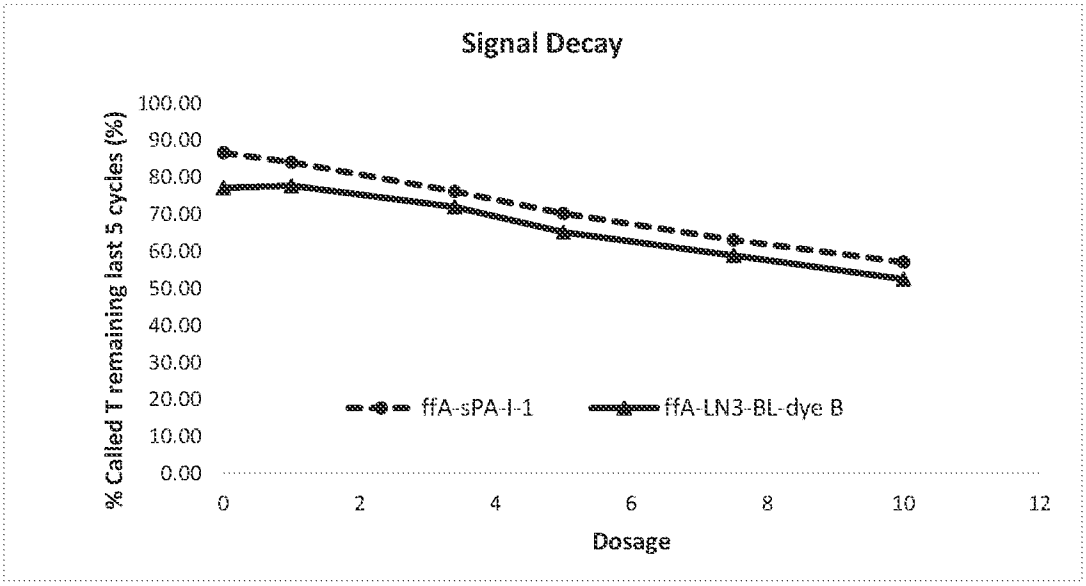

Signal Decay
% Called T remaining last 5 cycles (%)
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
0
2
4
6
8
10
12
Dosage
ffA-sPA-I-1
ffA-LN3-BL-dye B

FIG. 2A

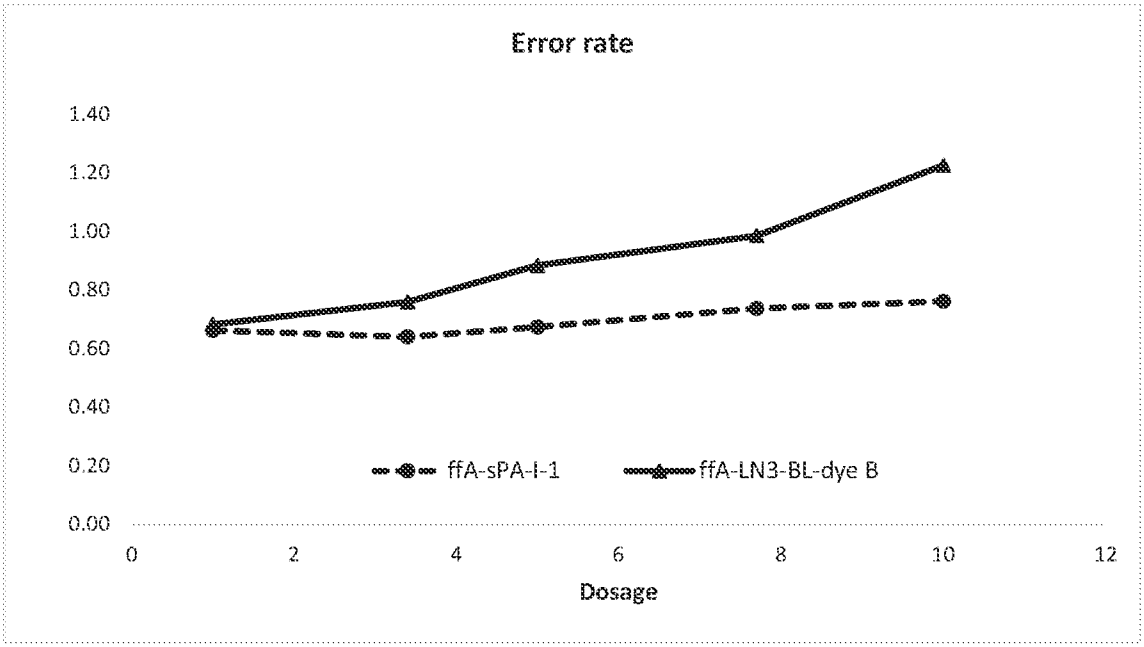

Error rate
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
0
2
4
6
8
10
12
Dosage
ffA-sPA-I-1
ffA-LN3-BL-dye B

FIG. 2B

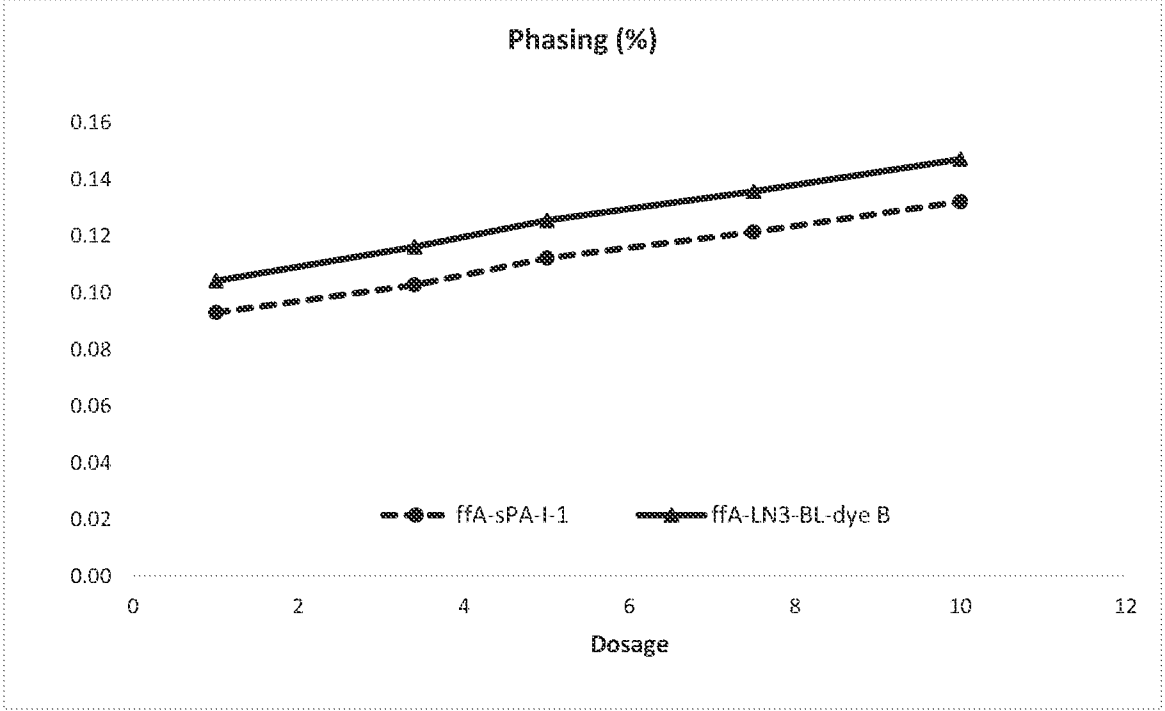

Phasing (%)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
10
12
Dosage
ffA-sPA-I-1
ffA-LN3-BL-dye B

FIG. 2C

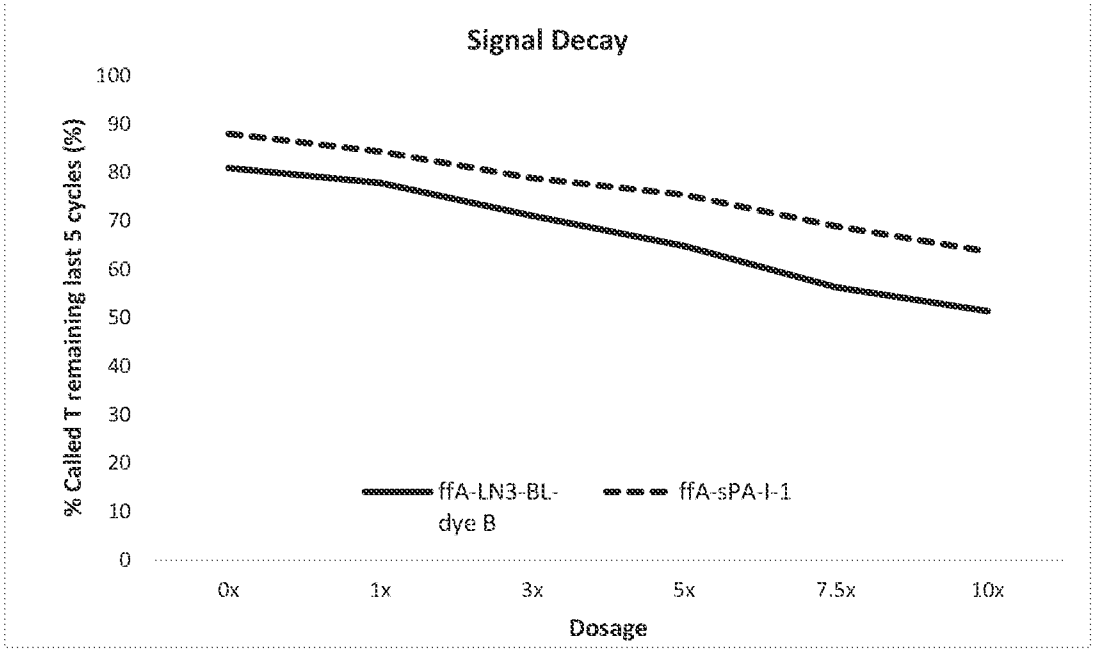

Signal Decay
% Called T remaining last 5 cycles (%)
100
90
80
70
60
50
40
30
20
10
0
ffA-LN3-BL-dye B
ffA-sPA-I-1
0x
1x
3x
5x
7.5x
10x
Dosage

FIG. 3A

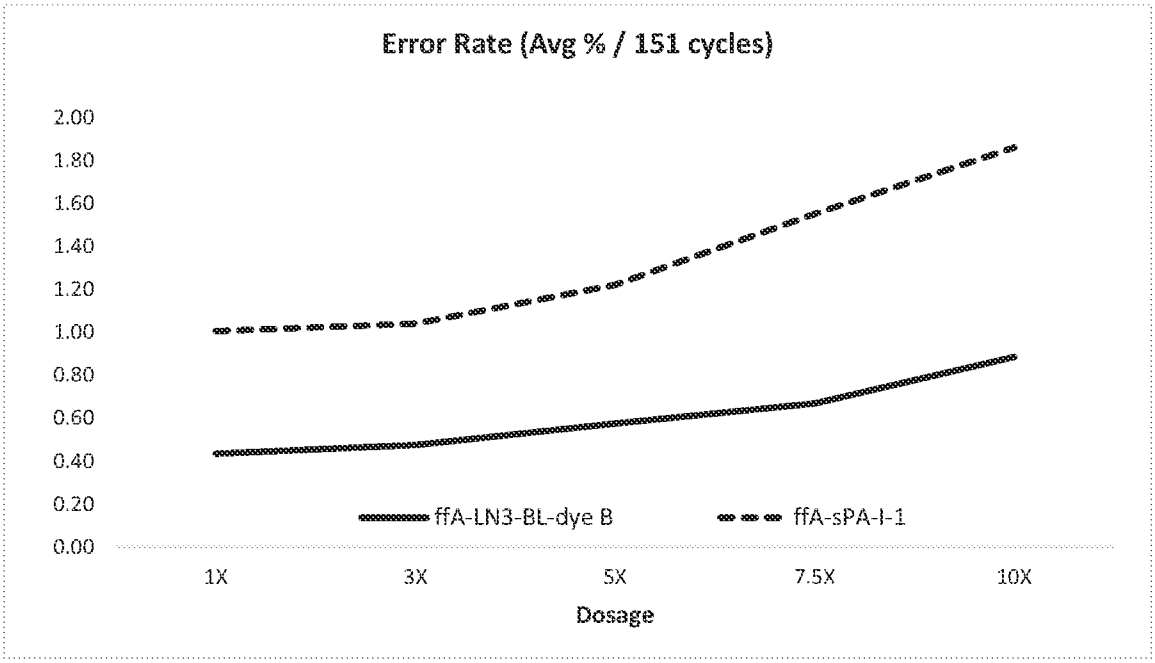

Error Rate (Avg % / 151 cycles)
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
ffA-LN3-BL-dye B
ffA-sPA-I-1
1X
3X
5X
7.5X
10X
Dosage

FIG. 3B

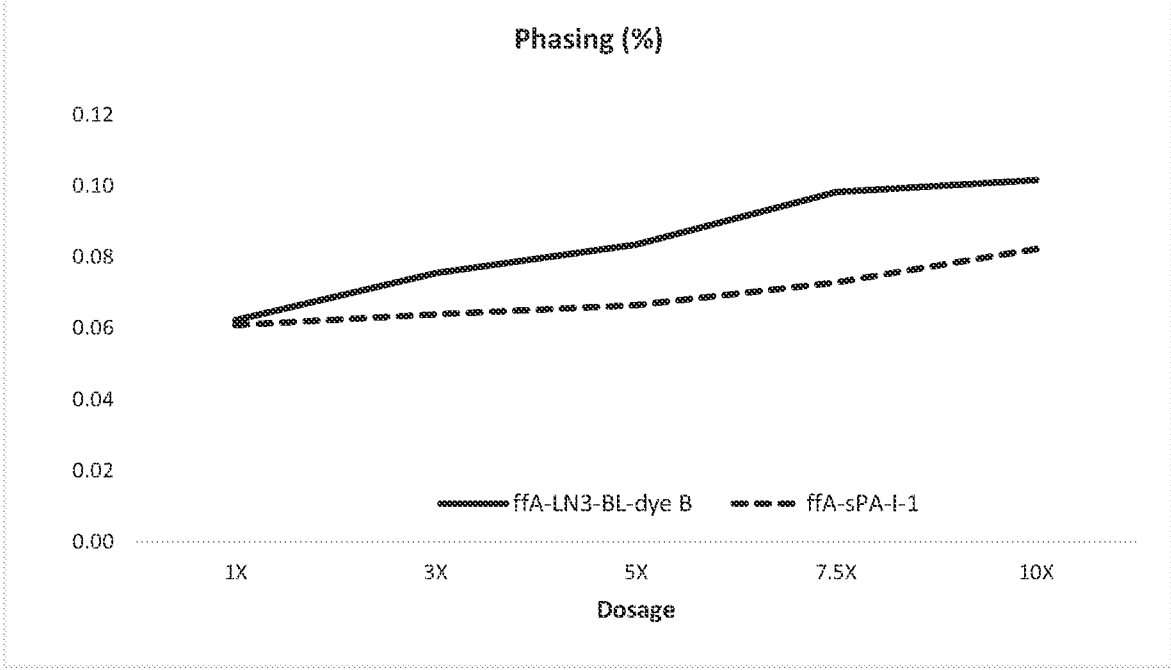

Phasing (%)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
ffA-LN3-BL-dye B
ffA-sPA-I-1
1X
3X
5X
7.5X
10X
Dosage

FIG. 3C