



(51) International Patent Classification:

C07H 19/073 (2006.01) C07H 19/173 (2006.01)

C07H 19/10 (2006.01) C07H 19/20 (2006.01)

C07H 19/14 (2006.01)

(21) International Application Number:

PCT/EP2013/055466

(22) International Filing Date:

15 March 2013 (15.03.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: ILLUMINA CAMBRIDGE LIMITED

[GB/GB]; Chesterford Research Park, Little Chesterford, Nr. Saffron Walden, Essex CB10 1XL (GB).

(72) Inventors: LIU, Xiaohai; c/o Illumina Cambridge Limited, Chesterford Research Park, Little Chesterford, Nr. Saffron Walden, Essex CB10 1XL (GB). WU, Xiaolin; c/o Illumina Cambridge Limited, Chesterford Research Park, Little Chesterford, Nr. Saffron Walden, Essex CB10 1XL (GB). SMITH, Geoff; c/o Illumina Cambridge Limited, Chesterford Research Park, Little Chesterford, Nr. Saffron Walden, Essex CB10 1XL (GB).

(74) Agents: BAILEY, Sam et al.; Mewburn Ellis LLP, 33 Gutter Lane, London EC2V 8AS (GB).

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

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

Published:

— with international search report (Art. 21(3))

(54) Title: MODIFIED NUCLEOSIDES OR NUCLEOTIDES

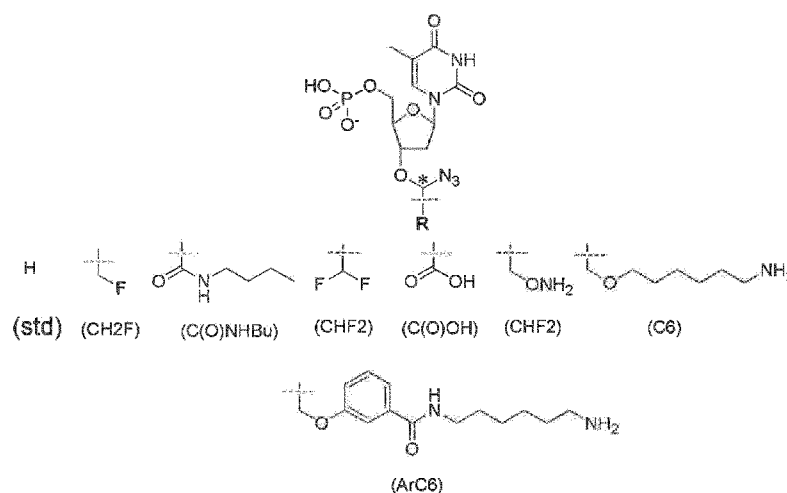


FIG. 1A

(57) Abstract: Some embodiments described herein relate to modified nucleotide and nucleoside molecules with novel 3'-hydroxy protecting groups. Said 3'-hydroxy protecting groups form a structure -O-C(R)₂N₃ covalently attached to the 3'-carbon atom wherein R is as defined in the claims. Also provided herein are methods to prepare such modified nucleotide and nucleoside molecules and sequencing by synthesis processes using such modified nucleotide and nucleoside molecules.

MODIFIED NUCLEOSIDES OR NUCLEOTIDES

BACKGROUND

Field of the Invention

[0001] Some embodiments described herein relate to modified nucleotides or nucleosides comprising 3'-hydroxy protecting groups and their use in polynucleotide sequencing methods. Some embodiments described herein relate to method of preparing the 3'-hydroxy protected nucleotides or nucleosides.

Description of the Related Art

[0002] Advances in the study of molecules have been led, in part, by improvement in technologies used to characterize the molecules or their biological reactions. In particular, the study of the nucleic acids DNA and RNA has benefited from developing technologies used for sequence analysis and the study of hybridization events.

[0003] An example of the technologies that have improved the study of nucleic acids is the development of fabricated arrays of immobilized nucleic acids. These arrays consist typically of a high-density matrix of polynucleotides immobilized onto a solid support material. *See, e.g., Fodor et al., Trends Biotech.* 12: 19-26, 1994, which describes ways of assembling the nucleic acids using a chemically sensitized glass surface protected by a mask, but exposed at defined areas to allow attachment of suitably modified nucleotide phosphoramidites. Fabricated arrays can also be manufactured by the technique of "spotting" known polynucleotides onto a solid support at predetermined positions (e.g., Stimpson et al., *Proc. Natl. Acad. Sci.* 92: 6379-6383, 1995).

[0004] One way of determining the nucleotide sequence of a nucleic acid bound to an array is called "sequencing by synthesis" or "SBS". This technique for determining the sequence of DNA ideally requires the controlled (i.e., one at a time) incorporation of the correct complementary nucleotide opposite the nucleic acid being sequenced. This allows for accurate sequencing by adding nucleotides in multiple cycles as each nucleotide residue is sequenced one at a time, thus preventing an uncontrolled series of incorporations occurring. The incorporated nucleotide is read using an appropriate label attached thereto before removal of the label moiety and the subsequent next round of sequencing.

[0005] In order to ensure only a single incorporation occurs, a structural modification ("protecting group") is added to each labeled nucleotide that is added to the growing chain to ensure that only one nucleotide is incorporated. After the nucleotide with the protecting group has been added, the protecting group is then removed, under reaction conditions which do not interfere with the integrity of the DNA being sequenced. The sequencing cycle can then continue with the incorporation of the next protected, labeled nucleotide.

[0006] To be useful in DNA sequencing, nucleotides, and more usually nucleotide triphosphates, generally require a 3'-hydroxy protecting group so as to prevent the polymerase used to incorporate it into a polynucleotide chain from continuing to replicate once the base on the nucleotide is added. There are many limitations on types of groups that can be added onto a nucleotide and still be suitable. The protecting group should prevent additional nucleotide molecules from being added to the polynucleotide chain whilst simultaneously being easily removable from the sugar moiety without causing damage to the polynucleotide chain. Furthermore, the modified nucleotide needs to be tolerated by the polymerase or other appropriate enzyme used to incorporate it into the polynucleotide chain. The ideal protecting group therefore exhibits long term stability, be efficiently incorporated by the polymerase enzyme, cause blocking of secondary or further nucleotide incorporation and have the ability to be removed under mild conditions that do not cause damage to the polynucleotide structure, preferably under aqueous conditions.

[0007] Reversible protecting groups have been described previously. For example, Metzker *et al.*, (*Nucleic Acids Research*, 22 (20): 4259-4267, 1994) discloses the synthesis and use of eight 3'-modified 2-deoxyribonucleoside 5'-triphosphates (3'-modified dNTPs) and testing in two DNA template assays for incorporation activity. WO 2002/029003 describes a sequencing method which may include the use of an allyl protecting group to cap the 3'-OH group on a growing strand of DNA in a polymerase reaction.

[0008] In addition, we previously reported the development of a number of reversible protecting groups and methods of deprotecting them under DNA compatible conditions in International Application Publication No. WO 2004/018497, which is hereby incorporated by reference in its entirety.

SUMMARY

[0009] Some embodiments described herein relate to a modified nucleotide or nucleoside molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar

moiety having a removable 3'-hydroxy protecting group forming a structure $-O-C(R)_2N_3$ covalently attached to the 3'-carbon atom, wherein

R is selected from the group consisting of hydrogen, $-C(R^1)_m(R^2)_n$, $-C(=O)OR^3$, $-C(=O)NR^4R^5$, $-C(R^6)_2O(CH_2)_pNR^7R^8$ and $-C(R^9)_2O-Ph-C(=O)NR^{10}R^{11}$;

each R^1 and R^2 is independently selected from hydrogen, optionally substituted alkyl or halogen;

R^3 is selected from hydrogen or optionally substituted alkyl;

each R^4 and R^5 is independently selected from hydrogen, optionally substituted alkyl, optionally substituted aryl, optionally substituted heteroaryl, or optionally substituted aralkyl;

each R^6 and R^9 is selected from hydrogen, optionally substituted alkyl or halogen;

each R^7 , R^8 , R^{10} and R^{11} is independently selected from hydrogen, optionally substituted alkyl, optionally substituted aryl, optionally substituted heteroaryl, or optionally substituted aralkyl;

m is an integer of 0 to 3; and

n is an integer of 0 to 3; provided that the total of m + n equals to 3; and

p is an integer of 0 to 6; provided that

R^1 and R^2 cannot both be halogen; and

at least one R is not hydrogen.

[0010] Some embodiments described herein relate to a method of preparing a growing polynucleotide complementary to a target single-stranded polynucleotide in a sequencing reaction, comprising incorporating a modified nucleotide molecule described herein into the growing complementary polynucleotide, wherein the incorporation of the modified nucleotide prevents the introduction of any subsequent nucleotide into the growing complementary polynucleotide.

[0011] Some embodiments described herein relate to a method for determining the sequence of a target single-stranded polynucleotide, comprising monitoring the sequential incorporation of complementary nucleotides, wherein at least one complementary nucleotide incorporated is a modified nucleotide molecule described herein; and detecting the identity of the modified nucleotide molecule. In some embodiments, the incorporation of the modified nucleotide molecule is accomplished by a terminal transferase, a terminal polymerase or a reverse transcriptase.

[0012] Some embodiments described herein relate to a kit comprising a plurality of modified nucleotide or nucleoside molecule described herein, and packaging materials therefor. In

some embodiments, the identity of the modified nucleotide is determined by detecting the detectable label linked to the base. In some such embodiments, the 3'-hydroxy protecting group and the detectable label are removed prior to introducing the next complementary nucleotide. In some such embodiments, the 3'-hydroxy protecting group and the detectable label are removed in a single step of chemical reaction.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0013] FIG. 1A illustrates a variety of 3'-OH protecting groups.
- [0014] FIG. 1B illustrated the thermal stability of various 3'-OH protecting groups.
- [0015] FIG. 2A illustrates the deprotection rate curve of three different 3'-OH protecting groups.
- [0016] FIG. 2B shows a chart of the deprotection half time of three different 3'-OH protecting groups.
- [0017] FIG. 3 shows the phasing and prephasing values of various modified nucleotide with a thermally stable 3'-OH protecting group in comparison and the standard protecting group.
- [0018] FIG. 4A shows the 2×400bp sequencing data of mono-F ffNs-A-isomer in incorporation mix (IMX).
- [0019] FIG. 4B shows the 2×400bp sequencing data of mono-F ffNs-B-isomer in incorporation mix (IMX).

DETAILED DESCRIPTION

[0020] One embodiment is a modified nucleotide or nucleoside comprising a 3'-OH protecting group. In one embodiment, the 3'-OH protecting group is a monofluoromethyl substituted azidomethyl protecting group. In another embodiment, the 3'-OH protecting group is a C-amido substituted azidomethyl protecting group. Still another embodiment relates to modified nucleotides having difluoromethyl substituted azidomethyl 3'-OH protecting groups.

Definitions

[0021] 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.

[0022] As used herein, common organic abbreviations are defined as follows:

Ac	Acetyl
Ac ₂ O	Acetic anhydride
aq.	Aqueous
Bn	Benzyl
Bz	Benzoyl
BOC or Boc	tert-Butoxycarbonyl
Bu	n-Butyl
cat.	Catalytic
Cbz	Carbobenzyloxy
°C	Temperature in degrees Centigrade
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
ddNTP(s)	Dideoxynucleotide(s)
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DCA	Dichloroacetic acid
DCE	1,2-Dichloroethane
DCM	Methylene chloride
DIEA	Diisopropylethylamine
DMA	Dimethylacetamide
DME	Dimethoxyethane
DMF	N,N'-Dimethylformamide

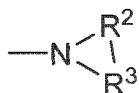
DMSO	Dimethylsulfoxide
DPPA	Diphenylphosphoryl azide
Et	Ethyl
EtOAc	Ethyl acetate
ffN	Fully functional nucleotide
g	Gram(s)
GPC	Gel permeation chromatography
h or hr	Hour(s)
iPr	Isopropyl
KPi	10 mM potassium phosphate buffer at pH 7.0
KPS	Potassium persulfate
IPA	Isopropyl Alcohol
IMX	Incorporation mix
LCMS	Liquid chromatography-mass spectrometry
LDA	Lithium diisopropylamide
m or min	Minute(s)
mCPBA	meta-Chloroperoxybenzoic Acid
MeOH	Methanol
MeCN	Acetonitrile
Mono-F	-CH ₂ F
Mono-F ffN	modified nucleotides with -CH ₂ F substituted on methylene position of azidomethyl 3'-OH protecting group
mL	Milliliter(s)
MTBE	Methyl tertiary-butyl ether
NaN ₃	Sodium Azide
NHS	N-hydroxysuccinimide
PG	Protecting group
Ph	Phenyl
ppt	Precipitate
rt	Room temperature
SBS	Sequencing by Synthesis
TEA	Triethylamine

TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TCDI	1,1'-Thiocarbonyl diimidazole
Tert, t	tertiary
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TEMED	Tetramethylethylenediamine
μL	Microliter(s)

[0023] 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.

[0024] 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 means in addition to covalent attachment.

[0025] As used herein, any “R” group(s) such as, without limitation, R², R³, R⁴, R⁵, R⁶, R⁷, and R⁸ represent substituents that can be attached to the indicated atom. An R group may be substituted or unsubstituted. If two “R” groups are described as being “taken together” the R groups and the atoms they are attached to can form a cycloalkyl, aryl, heteroaryl, or heterocycle. For example, without limitation, if R² and R³, or R², R³, or R⁴, and the atom to which it is attached, are indicated to be “taken together” or “joined together” it means that they are covalently bonded to one another to form a ring, an example of which is set forth below:



[0026] Whenever a group is described as being “optionally substituted” that group may be unsubstituted or substituted with one or more of the indicated substituents. Likewise, when a group is described as being “unsubstituted or substituted” if substituted, the substituent may be selected from one or more the indicated substituents. If no substituents are indicated, it is meant that the indicated “optionally substituted” or “substituted” group may be individually and independently substituted with one or more group(s) individually and independently selected from a group of functionalities including, but not limited to, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, aryl, heteroaryl, heteroalicycyl, aralkyl, heteroaralkyl, (heteroalicycyl)alkyl, hydroxy, protected hydroxyl, alkoxy, aryloxy, acyl, mercapto, alkylthio, arylthio, cyano, halogen, thiocarbonyl, O-carbamyl, N-carbamyl, O-thiocarbamyl, N-thiocarbamyl, C-amido, N-amido, S-sulfonamido, N-sulfonamido, C-carboxy, protected C-carboxy, O-carboxy, isocyanato, thiocyanato, isothiocyanato, nitro, silyl, sulfenyl, sulfinyl, sulfonyl, haloalkyl, haloalkoxy, trihalomethanesulfonyl, trihalomethanesulfonamido, amino, mono-substituted amino group, di-substituted amino group, and protected derivatives thereof.

[0027] As used herein, “alkyl” refers to a straight or branched hydrocarbon chain that comprises a fully saturated (no double or triple bonds) hydrocarbon group. In some embodiments, 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 inclusive of the endpoints; *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 about 7 to about 10 carbon atoms. The alkyl group can also be a lower alkyl having 1 to 6 carbon atoms. The alkyl group of the compounds may be designated as

“C₁-C₄ alkyl” or similar designations. By way of example only, “C₁-C₄ alkyl” indicates that there are one to four carbon atoms in the alkyl chain, *i.e.*, the alkyl chain is selected from 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, and hexyls. The alkyl group may be substituted or unsubstituted.

[0028] As used herein, “alkenyl” refers to an alkyl group that contains in the straight or branched hydrocarbon chain one or more double bonds. An alkenyl group may be unsubstituted or substituted.

[0029] As used herein, “alkynyl” refers to an alkyl group that contains in the straight or branched hydrocarbon chain one or more triple bonds. An alkynyl group may be unsubstituted or substituted.

[0030] As used herein, “cycloalkyl” refers to a completely saturated (no double or triple bonds) mono- or multi- cyclic hydrocarbon ring system. When composed of two or more rings, the rings may be joined together in a fused fashion. Cycloalkyl groups can contain 3 to 10 atoms in the ring(s). In some embodiments, cycloalkyl groups can contain 3 to 8 atoms in the ring(s). A cycloalkyl group may be unsubstituted or substituted. Typical cycloalkyl groups include, but are in no way limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl, and cyclooctyl.

[0031] As used herein, “aryl” refers to a carbocyclic (all carbon) monocyclic or polycyclic aromatic ring system (including, *e.g.*, fused, bridged, or spiro ring systems where two carbocyclic rings share a chemical bond, *e.g.*, one or more aryl rings with one or more aryl or non-aryl rings) that has a fully delocalized pi-electron system throughout at least one of the rings. The number of carbon atoms in an aryl group can vary. For example, in some embodiments, the aryl group can be a C₆-C₁₄ aryl group, a C₆-C₁₀ aryl group, or a C₆ aryl group. Examples of aryl groups include, but are not limited to, benzene, naphthalene, and azulene. An aryl group may be substituted or unsubstituted.

[0032] As used herein, “heterocyclyl” refers to ring systems including at least one heteroatom (*e.g.*, O, N, S). Such systems can be unsaturated, can include some unsaturation, or can contain some aromatic portion, or be all aromatic. A heterocyclyl group may be unsubstituted or substituted.

[0033] As used herein, “heteroaryl” refers to a monocyclic or polycyclic aromatic ring system (a ring system having at least one ring with a fully delocalized pi-electron system) that contain(s) one or more heteroatoms, that is, an element other than carbon, including but not limited

to, nitrogen, oxygen, and sulfur, and at least one aromatic ring. The number of atoms in the ring(s) of a heteroaryl group can vary. For example, in some embodiments, a heteroaryl group can contain 4 to 14 atoms in the ring(s), 5 to 10 atoms in the ring(s) or 5 to 6 atoms in the ring(s). Furthermore, the term “heteroaryl” includes fused ring systems where two rings, such as at least one aryl ring and at least one heteroaryl ring, or at least two heteroaryl rings, share at least one chemical bond. Examples of heteroaryl rings include, but are not limited to, furan, furazan, thiophene, benzothiophene, phthalazine, pyrrole, oxazole, benzoxazole, 1,2,3-oxadiazole, 1,2,4-oxadiazole, thiazole, 1,2,3-thiadiazole, 1,2,4-thiadiazole, benzothiazole, imidazole, benzimidazole, indole, indazole, pyrazole, benzopyrazole, isoxazole, benzoisoxazole, isothiazole, triazole, benzotriazole, thiadiazole, tetrazole, pyridine, pyridazine, pyrimidine, pyrazine, purine, pteridine, quinoline, isoquinoline, quinazoline, quinoxaline, cinnoline, and triazine. A heteroaryl group may be substituted or unsubstituted.

[0034] As used herein, “heteroalicyclic” or “heteroalicycyl” refers to three-, four-, five-, six-, seven-, eight-, nine-, ten-, up to 18-membered monocyclic, bicyclic, and tricyclic ring system wherein carbon atoms together with from 1 to 5 heteroatoms constitute said ring system. A heterocycle may optionally contain one or more unsaturated bonds situated in such a way, however, that a fully delocalized pi-electron system does not occur throughout all the rings. The heteroatoms are independently selected from oxygen, sulfur, and nitrogen. A heterocycle may further contain one or more carbonyl or thiocarbonyl functionalities, so as to make the definition include oxo-systems and thio-systems such as lactams, lactones, cyclic imides, cyclic thioimides, and cyclic carbamates. When composed of two or more rings, the rings may be joined together in a fused fashion. Additionally, any nitrogens in a heteroalicyclic may be quaternized. Heteroalicycyl or heteroalicyclic groups may be unsubstituted or substituted. Examples of such “heteroalicyclic” or “heteroalicycyl” groups include but are not limited to, 1,3-dioxin, 1,3-dioxane, 1,4-dioxane, 1,2-dioxolane, 1,3-dioxolane, 1,4-dioxolane, 1,3-oxathiane, 1,4-oxathiin, 1,3-oxathiolane, 1,3-dithiole, 1,3-dithiolane, 1,4-oxathiane, tetrahydro-1,4-thiazine, 2H-1,2-oxazine, maleimide, succinimide, barbituric acid, thiobarbituric acid, dioxopiperazine, hydantoin, dihydrouracil, trioxane, hexahydro-1,3,5-triazine, imidazoline, imidazolidine, isoxazoline, isoxazolidine, oxazoline, oxazolidine, oxazolidinone, thiazoline, thiazolidine, morpholine, oxirane, piperidine *N*-Oxide, piperidine, piperazine, pyrrolidine, pyrrolidone, pyrrolidione, 4-piperidone, pyrazoline, pyrazolidine, 2-oxopyrrolidine, tetrahydropyran, 4H-pyran, tetrahydrothiopyran, thiamorpholine, thiamorpholine

sulfoxide, thiamorpholine sulfone, and their benzo-fused analogs (*e.g.*, benzimidazolidinone, tetrahydroquinoline, 3,4-methylenedioxyphenyl).

[0035] As used herein, “aralkyl” and “aryl(alkyl)” refer to an aryl group connected, as a substituent, via a lower alkylene group. The lower alkylene and aryl group of an aralkyl may be substituted or unsubstituted. Examples include but are not limited to benzyl, 2-phenylalkyl, 3-phenylalkyl, and naphthylalkyl.

[0036] As used herein, “heteroaralkyl” and “heteroaryl(alkyl)” refer to a heteroaryl group connected, as a substituent, via a lower alkylene group. The lower alkylene and heteroaryl group of heteroaralkyl may be substituted or unsubstituted. Examples include but are not limited to 2-thienylalkyl, 3-thienylalkyl, furylalkyl, thienylalkyl, pyrrolylalkyl, pyridylalkyl, isoxazolylalkyl, and imidazolylalkyl, and their benzo-fused analogs.

[0037] As used herein, “alkoxy” refers to the formula –OR wherein R is an alkyl, an alkenyl, an alkynyl, a cycloalkyl, a cycloalkenyl or a cycloalkynyl is defined as above. A non-limiting list of alkoxys is methoxy, ethoxy, n-propoxy, 1-methylethoxy (isopropoxy), n-butoxy, isobutoxy, sec-butoxy, and tert-butoxy. An alkoxy may be substituted or unsubstituted.

[0038] As used herein, a “C-amido” group refers to a “–C(=O)N(R_aR_b)” group in which R_a and R_b can be independently hydrogen, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, aryl, heteroaryl, heteroalicycyl, aralkyl, or (heteroalicycyl)alkyl. A C-amido may be substituted or unsubstituted.

[0039] As used herein, an “N-amido” group refers to a “RC(=O)N(R_a)–” group in which R and R_a can be independently hydrogen, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, aryl, heteroaryl, heteroalicycyl, aralkyl, or (heteroalicycyl)alkyl. An N-amido may be substituted or unsubstituted.

[0040] The term “halogen atom”, “halogen” or “halo” as used herein, means any one of the radio-stable atoms of column 7 of the Periodic Table of the Elements, such as, fluorine, chlorine, bromine, and iodine.

[0041] The term “amine” as used herein refers to a –NH₂ group wherein one or more hydrogen can be optionally substituted by a R group. R can be independently hydrogen, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, aryl, heteroaryl, heteroalicycyl, aralkyl, or (heteroalicycyl)alkyl.

[0042] The term “aldehyde” as used herein refers to a –R_c-C(O)H group, wherein R_c can be absent or independently selected from alkylene, alkenylene, alkynylene, cycloalkylene,

cycloalkenylene, cycloalkynylene, arylene, heteroarylene, heteroalicycylene, aralkylene, or (heteroalicycyl)alkylene.

[0043] The term “amino” as used herein refers to a -NH_2 group.

[0044] The term “hydroxy” as used herein refers to a -OH group.

[0045] The term “cyano” group as used herein refers to a “ -CN ” group.

[0046] The term “azido” as used herein refers to a -N_3 group.

[0047] The term “thiol” as used herein refers to a -SH group.

[0048] The term “carboxylic acid” as used herein refers to -C(O)OH .

[0049] The term “thiocyanate” as used herein refers to $\text{-S-C}\equiv\text{N}$ group.

[0050] The term “oxo-amine” as used herein refers to -O-NH_2 group, wherein one or more hydrogen of the -NH_2 can be optionally substituted by a R group. R can be independently hydrogen, alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkenyl, cycloalkynyl, aryl, heteroaryl, heteroalicycyl, aralkyl, or (heteroalicycyl)alkyl. .

[0051] 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 hydroxyl group that is present in ribose. The nitrogen containing heterocyclic base can be purine or pyrimidine base. Purine bases include adenine (A) and guanine (G), and modified derivatives or analogs thereof. 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.

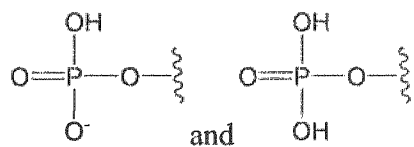
[0052] As used herein, a “nucleoside” is structurally similar to a nucleotide, but is missing the phosphate moieties. An example of a nucleoside analogue would be one in which the label is linked to the base and there is no phosphate group attached to the sugar molecule. The term “nucleoside” is used 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.

[0053] 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, 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).

[0054] As used herein, “derivative” or “analogue” 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.

[0055] 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.

[0056] The terms “protecting group” and “protecting groups” as used herein refer to any atom or group of atoms that is added to a molecule in order to prevent existing groups in the molecule from undergoing unwanted chemical reactions. Sometimes, “protecting group” and “blocking group” can be used interchangeably.

[0057] As used herein, the prefixes “photo” or “photo-” mean relating to light or electromagnetic radiation. The term can encompass all or part of the electromagnetic spectrum including, but not limited to, one or more of the ranges commonly known as the radio, microwave, infrared, visible, ultraviolet, X-ray or gamma ray parts of the spectrum. The part of the spectrum can be one that is blocked by a metal region of a surface such as those metals set forth herein. Alternatively or additionally, the part of the spectrum can be one that passes through an interstitial region of a surface such as a region made of glass, plastic, silica, or other material set forth herein. In particular embodiments, radiation can be used that is capable of passing through a metal.

Alternatively or additionally, radiation can be used that is masked by glass, plastic, silica, or other material set forth herein.

[0058] As used herein, the term “phasing” refers to phenomena in SBS that is caused by incomplete removal of the 3' terminators and fluorophores, and failure to complete the incorporation of a portion of DNA strands within clusters by polymerases at a given sequencing cycle. Pre-phasing is caused by the incorporation of nucleotides without effective 3' terminators and the incorporation event goes 1 cycle ahead. Phasing and pre-phasing cause the extracted intensities for a specific cycle to consist of the signal of 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 increases, hampering the identification of the correct base. Pre-phasing 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 pre-phasing 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 pre-phasing values, and longer sequencing read length.

3'-OH Protecting Groups $-C(R)_2N_3$

[0059] Some embodiments described herein relate to a modified nucleotide or nucleoside molecule having a removable 3'-hydroxy protecting group $-C(R)_2N_3$, wherein R is selected from the group consisting of hydrogen, $-C(R^1)_m(R^2)_n$, $-C(=O)OR^3$, $-C(=O)NR^4R^5$, $-C(R^6)_2O(CH_2)_pNR^7R^8$ and $-C(R^9)_2O-Ph-C(=O)NR^{10}R^{11}$, wherein R^1 , R^2 , R^3 , R^4 , R^5 , R^6 , R^7 , R^8 , R^9 , R^{10} , R^{11} , m, n and p are defined above.

[0060] In some embodiments, one of R is hydrogen and the other R is $-C(R^1)_m(R^2)_n$. In some such embodiments, $-C(R^1)_m(R^2)_n$ is selected from $-CHF_2$, $-CH_2F$, $-CHCl_2$ or $-CH_2Cl$. In one embodiment, $-C(R^1)_m(R^2)_n$ is $-CHF_2$. In another embodiment, $-C(R^1)_m(R^2)_n$ is $-CH_2F$.

[0061] In some embodiments, one of R is hydrogen and the other R is $-C(=O)OR^3$. In some such embodiment, R^3 is hydrogen.

[0062] In some embodiments, one of R is hydrogen and the other R is $-C(=O)NR^4R^5$. In some such embodiments, both R^4 and R^5 are hydrogen. In some other such embodiments, R^4 is

hydrogen and R^5 is C_{1-6} alkyl. In still some other embodiments, both R^4 and R^5 are C_{1-6} alkyl. In one embodiment, R^5 is n-butyl. In another embodiment, both R^4 and R^5 are methyl.

[0063] In some embodiments, one of R is hydrogen and the other R is $-C(R^6)_2O(CH_2)_pNR^7R^8$. In some such embodiments, both R^6 are hydrogen. In some such embodiments, both R^7 and R^8 are hydrogen. In some such embodiment, p is 0. In some other such embodiment, p is 6.

[0064] In some embodiments, one of R is hydrogen and the other R is $-C(R^9)_2O-Ph-C(=O)NR^{10}R^{11}$. In some such embodiments, both R^9 are hydrogen. In some such embodiments, both R^{10} and R^{11} are hydrogen. In some other such embodiments, R^{10} is hydrogen and R^{11} is a substituted alkyl. In one embodiment, R^{11} is an amino substituted alkyl.

Deprotection of the 3'-OH Protecting Groups

[0065] In some embodiments, the 3'-OH protecting group is removed in a deprotecting reaction with a phosphine. The azido group in $-C(R)_2N_3$ can be converted to an amino group by contacting the modified nucleotide or nucleoside molecules with a phosphine. Alternatively, the azido group in $-C(R)_2N_3$ may be converted to an amino group by contacting such molecules with the thiols, in particular water-soluble thiols such as dithiothreitol (DTT). In one embodiment, the phosphine is tris(hydroxymethyl)phosphine (THP). Unless indicated otherwise, the reference to nucleotides is also intended to be applicable to nucleosides.

Detectable Labels

[0066] Some embodiments described herein relate to the use of conventional detectable labels. Detection can be carried out by any suitable method, including fluorescence spectroscopy or by other optical means. The preferred label is a fluorophore, which, after absorption of energy, emits radiation at a defined wavelength. Many suitable fluorescent labels are known. For example, Welch *et al.* (*Chem. Eur. J.* 5(3):951-960, 1999) discloses dansyl-functionalised fluorescent moieties that can be used in the present invention. Zhu *et al.* (*Cytometry* 28:206-211, 1997) describes the use of the fluorescent labels Cy3 and Cy5, which can also be used in the present invention. Labels suitable for use are also disclosed in Prober *et al.* (*Science* 238:336-341, 1987); Connell *et al.* (*BioTechniques* 5(4):342-384, 1987), Ansorge *et al.* (*Nucl. Acids Res.* 15(11):4593-4602, 1987) and Smith *et al.* (*Nature* 321:674, 1986). Other commercially available fluorescent

labels include, but are not limited to, fluorescein, rhodamine (including TMR, texas red and Rox), alexa, bodipy, acridine, coumarin, pyrene, benzanthracene and the cyanins.

[0067] Multiple labels can also be used in the present application, for example, bi-fluorophore FRET cassettes (*Tet. Let.* 46:8867-8871, 2000). Multi-fluor dendrimeric systems (*J. Am. Chem. Soc.* 123:8101-8108, 2001) can also be used. Although fluorescent labels are preferred, other forms of detectable labels will be apparent as useful to those of ordinary skill in the art. For example, microparticles, including quantum dots (Empodocles *et al.*, *Nature* 399:126-130, 1999), gold nanoparticles (Reichert *et al.*, *Anal. Chem.* 72:6025-6029, 2000) and microbeads (Lacoste *et al.*, *Proc. Natl. Acad. Sci USA* 97(17):9461-9466, 2000) can all be used.

[0068] Multi-component labels can also be used in the present application. A multi-component label is one which is dependent on the interaction with a further compound for detection. The most common multi-component label used in biology is the biotin-streptavidin system. Biotin is used as the label attached to the nucleotide base. Streptavidin is then added separately to enable detection to occur. Other multi-component systems are available. For example, dinitrophenol has a commercially available fluorescent antibody that can be used for detection.

[0069] Unless indicated otherwise, the reference to nucleotides is also intended to be applicable to nucleosides. The present application will also be further described with reference to DNA, although the description will also be applicable to RNA, PNA, and other nucleic acids, unless otherwise indicated.

Linkers

[0070] In some embodiments described herein, the purine or pyrimidine base of the modified nucleotide or nucleoside molecules can be linked to a detectable label as described above. In some such embodiments, the linkers used are cleavable. The use of a cleavable linker ensures that the label can, if required, be removed after detection, avoiding any interfering signal with any labeled nucleotide or nucleoside incorporated subsequently.

[0071] In some other embodiments, the linkers used are non-cleavable. Since in each instance where a labeled nucleotide of the invention is incorporated, no nucleotides need to be subsequently incorporated and thus the label need not be removed from the nucleotide.

[0072] Those skilled in the art will be aware of the utility of dideoxynucleoside triphosphates in so-called Sanger sequencing methods, and related protocols (Sanger-type), which

rely upon randomized chain-termination at a particular type of nucleotide. An example of a Sanger-type sequencing protocol is the BASS method described by Metzker.

[0073] Sanger and Sanger-type methods generally operate by the conducting of an experiment in which eight types of nucleotides are provided, four of which contain a 3'-OH group; and four of which omit the OH group and which are labeled differently from each other. The nucleotides used which omit the 3'-OH group - dideoxy nucleotides (ddNTPs). As known by one skilled in the art, since the ddNTPs are labeled differently, by determining the positions of the terminal nucleotides incorporated, and combining this information, the sequence of the target oligonucleotide may be determined.

[0074] The nucleotides of the present application, it will be recognized, may be of utility in Sanger methods and related protocols since the same effect achieved by using ddNTPs may be achieved by using the 3'-OH protecting groups described herein: both prevent incorporation of subsequent nucleotides.

[0075] Moreover, it will be appreciated that monitoring of the incorporation of 3'-OH protected nucleotides may be determined by use of radioactive ^{32}P in the phosphate groups attached. These may be present in either the ddNTPs themselves or in the primers used for extension.

[0076] Cleavable linkers are known in the art, and conventional chemistry can be applied to attach a linker to a nucleotide base and a label. The linker can be cleaved by any suitable method, including exposure to acids, bases, nucleophiles, electrophiles, radicals, metals, reducing or oxidizing agents, light, temperature, enzymes etc. The linker as discussed herein may also be cleaved with the same catalyst used to cleave the 3'-O-protecting group bond. Suitable linkers can be adapted from standard chemical protecting groups, as disclosed in Greene & Wuts, *Protective Groups in Organic Synthesis*, John Wiley & Sons. Further suitable cleavable linkers used in solid-phase synthesis are disclosed in Guillier *et al.* (*Chem. Rev.* 100:2092-2157, 2000).

[0077] The use of the term "cleavable linker" is not meant to imply that the whole linker is required to be removed from, *e.g.*, the nucleotide base. Where the detectable label is attached to the base, the nucleoside cleavage site can be located at a position on the linker that ensures that part of the linker remains attached to the nucleotide base after cleavage.

[0078] Where the detectable label is attached to the base, the linker can be attached at any position on the nucleotide base provided that Watson-Crick base pairing can still be carried out. In the context of purine bases, it is preferred if the linker is attached via the 7-position of the purine or the preferred deazapurine analogue, via an 8-modified purine, via an N-6 modified adenosine or

an N-2 modified guanine. For pyrimidines, attachment is preferably via the 5-position on cytosine, thymidine or uracil and the N-4 position on cytosine.

A. Electrophilically cleaved linkers

[0079] Electrophilically cleaved linkers are typically cleaved by protons and include cleavages sensitive to acids. Suitable linkers include the modified benzylic systems such as trityl, p-alkoxybenzyl esters and p-alkoxybenzyl amides. Other suitable linkers include tert-butyloxycarbonyl (Boc) groups and the acetal system.

[0080] The use of thiophilic metals, such as nickel, silver or mercury, in the cleavage of thioacetal or other sulfur-containing protecting groups can also be considered for the preparation of suitable linker molecules.

B. Nucleophilically cleaved linkers

[0081] Nucleophilic cleavage is also a well recognised method in the preparation of linker molecules. Groups such as esters that are labile in water (i.e., can be cleaved simply at basic pH) and groups that are labile to non-aqueous nucleophiles, can be used. Fluoride ions can be used to cleave silicon-oxygen bonds in groups such as triisopropyl silane (TIPS) or t-butyldimethyl silane (TBDMS).

C. Photocleavable linkers

[0082] Photocleavable linkers have been used widely in carbohydrate chemistry. It is preferable that the light required to activate cleavage does not affect the other components of the modified nucleotides. For example, if a fluorophore is used as the label, it is preferable if this absorbs light of a different wavelength to that required to cleave the linker molecule. Suitable linkers include those based on O-nitrobenzyl compounds and nitroveratryl compounds. Linkers based on benzoin chemistry can also be used (Lee *et al.*, *J. Org. Chem.* 64:3454-3460, 1999).

D. Cleavage under reductive conditions

[0083] There are many linkers known that are susceptible to reductive cleavage. Catalytic hydrogenation using palladium-based catalysts has been used to cleave benzyl and benzyloxycarbonyl groups. Disulfide bond reduction is also known in the art.

E. Cleavage under oxidative conditions

[0084] Oxidation-based approaches are well known in the art. These include oxidation of p-alkoxybenzyl groups and the oxidation of sulfur and selenium linkers. The use of aqueous iodine to cleave disulfides and other sulfur or selenium-based linkers is also within the scope of the invention.

F. Safety-catch linkers

[0085] Safety-catch linkers are those that cleave in two steps. In a preferred system the first step is the generation of a reactive nucleophilic center followed by a second step involving an intra-molecular cyclization that results in cleavage. For example, levulinic ester linkages can be treated with hydrazine or photochemistry to release an active amine, which can then be cyclised to cleave an ester elsewhere in the molecule (Burgess *et al.*, J. Org. Chem. 62:5165-5168, 1997).

G. Cleavage by elimination mechanisms

[0086] Elimination reactions can also be used. For example, the base-catalysed elimination of groups such as Fmoc and cyanoethyl, and palladium-catalysed reductive elimination of allylic systems, can be used.

[0087] In some embodiments, the linker can comprise a spacer unit. The length of the linker is unimportant provided that the label is held a sufficient distance from the nucleotide so as not to interfere with any interaction between the nucleotide and an enzyme.

[0088] In some embodiments, the linker may consist of the similar functionality as the 3'-OH protecting group. This will make the deprotection and deprotecting process more efficient, as only a single treatment will be required to remove both the label and the protecting group. Particularly preferred linkers are phosphine-cleavable azide containing linkers.

Sequencing methods

[0089] The modified nucleosides or nucleotides described herein can be used in conjunction with a variety of sequencing techniques. In some embodiments, the process to determine the nucleotide sequence of a target nucleic acid can be an automated process.

[0090] The nucleotide analogues presented herein can be used in a sequencing procedure, such as a sequencing-by-synthesis (SBS) technique. Briefly, SBS can be initiated by contacting the target nucleic acids with one or more labeled nucleotides, DNA polymerase, etc. Those features where a primer is extended using the target nucleic acid as template will incorporate a labeled nucleotide that can be detected. Optionally, the labeled nucleotides can further include a reversible termination property that terminates further primer extension once a nucleotide has been added to a primer. For example, a nucleotide analog having a reversible terminator moiety can be added to a primer such that subsequent extension cannot occur until a deblocking agent is delivered to remove the moiety. Thus, for embodiments that use reversible termination, a deblocking reagent can be delivered to the flow cell (before or after detection occurs). Washes can be carried out

between the various delivery steps. The cycle can then be repeated n times to extend the primer by n nucleotides, thereby detecting a sequence of length n . Exemplary SBS procedures, fluidic systems and detection platforms that can be readily adapted for use with an array produced by the methods of the present disclosure are described, for example, in Bentley et al., *Nature* 456:53-59 (2008), WO 04/018497; WO 91/06678; WO 07/123744; US Pat. Nos. 7,057,026; 7,329,492; 7,211,414; 7,315,019 or 7,405,281, and US Pat. App. Pub. No. 2008/0108082 A1, each of which is incorporated herein by reference.

[0091] Other sequencing procedures that use cyclic reactions can be used, such as pyrosequencing. Pyrosequencing detects the release of inorganic pyrophosphate (PPi) as particular nucleotides are incorporated into a nascent nucleic acid strand (Ronaghi, et al., *Analytical Biochemistry* 242(1), 84-9 (1996); Ronaghi, *Genome Res.* 11(1), 3-11 (2001); Ronaghi et al. *Science* 281(5375), 363 (1998); US Pat. Nos. 6,210,891; 6,258,568 and 6,274,320, each of which is incorporated herein by reference). In pyrosequencing, released PPi can be detected by being converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the resulting ATP can be detected via luciferase-produced photons. Thus, the sequencing reaction can be monitored via a luminescence detection system. Excitation radiation sources used for fluorescence based detection systems are not necessary for pyrosequencing procedures. Useful fluidic systems, detectors and procedures that can be used for application of pyrosequencing to arrays of the present disclosure are described, for example, in WIPO Pat. App. Ser. No. PCT/US11/57111, US Pat. App. Pub. No. 2005/0191698 A1, US Pat. No. 7,595,883, and US Pat. No. 7,244,559, each of which is incorporated herein by reference.

[0092] Sequencing-by-ligation reactions are also useful including, for example, those described in Shendure et al. *Science* 309:1728-1732 (2005); US Pat. No. 5,599,675; and US Pat. No. 5,750,341, each of which is incorporated herein by reference. Some embodiments can include sequencing-by-hybridization procedures as described, for example, in Bains et al., *Journal of Theoretical Biology* 135(3), 303-7 (1988); Drmanac et al., *Nature Biotechnology* 16, 54-58 (1998); Fodor et al., *Science* 251(4995), 767-773 (1995); and WO 1989/10977, each of which is incorporated herein by reference. In both sequencing-by-ligation and sequencing-by-hybridization procedures, nucleic acids that are present in gel-containing wells (or other concave features) are subjected to repeated cycles of oligonucleotide delivery and detection. Fluidic systems for SBS methods as set forth herein, or in references cited herei, can be readily adapted for delivery of reagents for sequencing-by-ligation or sequencing-by-hybridization procedures. Typically, the

oligonucleotides are fluorescently labeled and can be detected using fluorescence detectors similar to those described with regard to SBS procedures herein or in references cited herein.

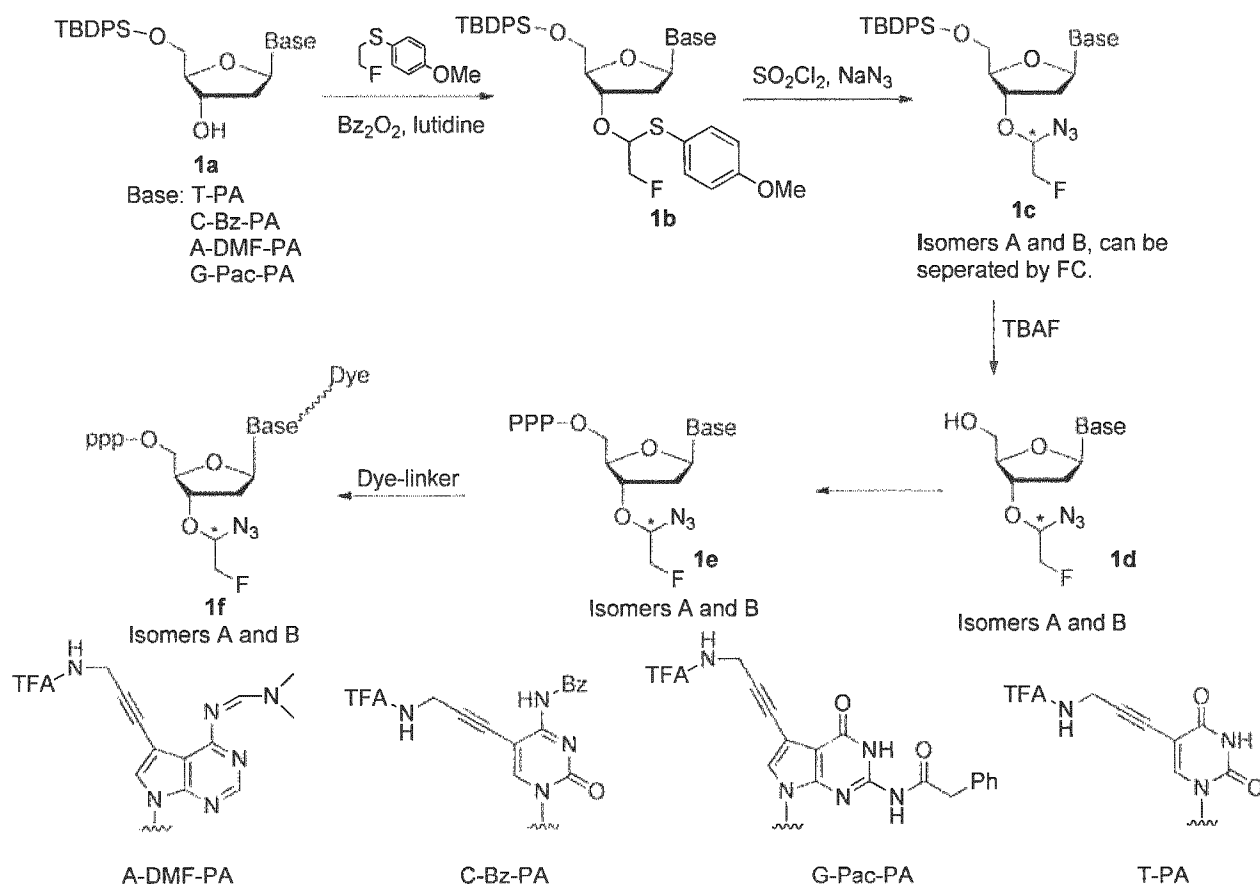
[0093] Some embodiments can utilize methods involving the real-time monitoring of DNA polymerase activity. For example, nucleotide incorporations can be detected through fluorescence resonance energy transfer (FRET) interactions between a fluorophore-bearing polymerase and γ -phosphate-labeled nucleotides, or with zeromode waveguides. Techniques and reagents for FRET-based sequencing are described, for example, in Levene et al. *Science* 299, 682–686 (2003); Lundquist et al. *Opt. Lett.* 33, 1026–1028 (2008); Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176–1181 (2008), the disclosures of which are incorporated herein by reference.

[0094] Some SBS embodiments include detection of a proton released upon incorporation of a nucleotide into an extension product. For example, sequencing based on detection of released protons can use an electrical detector and associated techniques that are commercially available from Ion Torrent (Guilford, CT, a Life Technologies subsidiary) or sequencing methods and systems described in US Pat. App. Pub. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; or 2010/0282617 A1, each of which is incorporated herein by reference.

EXAMPLES

[0095] 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. The synthesis of various modified nucleotide with protected 3'-hydroxy group are demonstrated in Examples 1-3.

Example 1
Synthesis of Nucleotides with 3'-OH protecting group



Scheme 1.

[0096] Scheme 1 illustrates a synthetic route for the preparation of the modified nucleotides with monofluoromethyl substituted azidomethyl as 3'-OH protecting groups. Compounds **1a-1f** employ a modified thymine (T-PA) as the base. Other non-limiting examples of the bases that can be used include Cbz-PA, ADMF-PA, and GPac-PA, the structures of which are shown above in Scheme 1.

Experimental Procedures

[0097] To a solution of the starting nucleoside **1a** (1.54 g, 2.5 mmol) in anhydrous CH₃CN (25 ml) was added 2,6-lutidine (0.87 mL, 7.5 mmol), (2-fluoroethyl)(4-methoxyphenyl)sulfane (MPSF) (3.26 g, 17.5 mmol) and then Bz₂O₂ (50% pure, 8.47 g, 17.5 mmol) at 4°C. The reaction mixture was allowed to warm up slowly to room temperature. The mixture was stirred for other 6 hours. TLC monitored (EtOAc:DCM=2:8 v/v) to see complete consumption of

the starting nucleoside. The reaction was then concentrated under reduced pressure to oily residue. To this mixture, petroleum ether (500 ml) was added and stirred vigorously for 10 min. The petroleum ether layer was decanted and the residue was repeated to treat with petroleum ether ($\times 2$). The oily residue was partitioned between DCM/ NaHCO_3 (1:1) (300 mL). The organic layer was separated and the aqueous was further extracted into DCM (2×150 mL). Combined organic layers were dried over MgSO_4 , filtered and the volatiles evaporated under reduced pressure. Crude product **1c** was purified by Biotag silica gel column (50g) using a gradient of petroleum ether to petroleum ether:EtOAc 1:1 (v/v) to afford 1.63g nucleoside **1b** as a pale yellow foam (diastereomers, 82% yield). ^1H NMR (d_6 DMSO, 400 MHz): δ , 0.95 (s, 9H, tBu), 2.16 – 2.28 (m, 2H, *H*-2'), 3.67 (s, OMe), 3.65 - 3.85 (m, 2H, *HH*-5'), 3.77 (dd, $J = 11.1, 4.5$ Hz, 1H, *HH*-5'), 3.95-3.98 (m, 1H, *H*-4'), 4.04 (m, 2H, CH_2F), 4.63-4.64 (m, 1H, *H*-3'), 5.01-5.32 (s, 1H, *CH*), 6.00 (m, 1H, *H*-1'), 6.72-6.87 (m, 3H, Ar), 7.35-7.44 (m, 7H, Ar), 7.55-7.60 (m, 4H, Ar), 7.88 (s, 1H, *H*-6), 9.95 (brt, 1H, NH), 11.70 (s, 1H, NH).

[0098] To a solution of the starting nucleoside **1b** (1.14 g, 1.4 mmol) in anhydrous CH_2Cl_2 (14 mL) with molecular sieve (4 Å) under N_2 was added cyclohexene (1.44 mL, 14 mmol). The mixture was cooled with a dry ice/acetone bath to -78°C . The solution of sulfonyl chloride (580 μL , 7.2 mmol) in DCM (14 ml) was slowly added over 90 minutes under N_2 . After 20 mins at that temperature TLC (EtOAc: petroleum ether=1:1 v/v) indicated the full consumption of the starting nucleoside. Volatiles were evaporated under reduced pressure (and room temperature of 25°C) and the oily residue was quickly subjected to high vacuum for a further 10 minutes until it foamed. The crude product was purged with N_2 and then dissolved in anhydrous DMF (5 mL) and NaN_3 (470 mg, 7 mmol) added at once. The resulting suspension was stirred at room temperature for 2 hours or until TLC indicated the completion of the reaction and formation of **1c** as two isomer (a and b) The reaction mixture was partitioned between EtOAc: NaHCO_3 (1:1) (200 mL). The organic layer was separated and the aqueous was further extracted into EtOAc (2×100 mL). Combined organic extracts were dried over MgSO_4 , filtered and the volatiles evaporated under reduced pressure. The two diastereoisomers of **1c** (A and B) were separated by Biotag silica gel column (25 g) using a gradient of petroleum ether to petroleum ether: EtOAc 1:1 (v/v) as pale yellow foam.

[0099] Isomer A (370 mg, yield: 38%). ^1H NMR (d_6 DMSO, 400 MHz): δ 1.02 (s, 9H, tBu), 2.35 – 2.43 (m, 2H, *H*-2'), 3.76-3.80 (m, 1H, *H*-5'), 3.88 - 3.92 (m, 1H, *H*-5'), 4.10 - 4.12 (m, 1H, *H*-4'), 4.14 (d, $J = 4.1$ Hz 2H, NHCH_2), 4.46-4.60 (m, 3H, *H*-3', CH_2F), 5.05-5.09 (m, 1H,

CHN₃), 6.11 (t, $J = 6.1$ Hz, 1H, H-1'), 7.47 - 7.51 (m, 6H, Ar), 7.64 - 7.68 (m, 4H, Ar), 7.97 (s, 1H, H-6), 10.03 (bt, 1H, $J = 10.0$ Hz, NH), 11.76 (s, 1H, NH). ¹⁹F NMR: -74.3 (CF₃), -230.2 (CH₂F).

[0100] Isomer B (253 mg, yield:26%). ¹H NMR (d₆ DMSO, 400 MHz): δ 1.01 (s, 9H, tBu), 2.38 - 2.42 (m, 2H, H-2'), 3.74-3.78 (m, 1H, H-5'), 3.86-3.90 (m, 1H, H-5'), 4.00-4.05 (m, 1H, H-4'), 4.12 (d, $J = 4.1$ Hz 2H, NHCH₂), 4.45-4.60 (m, 3H, H-3', CH₂F), 5.00-5.14 (m, 1H, CHN₃), 6.09 (t, $J = 6.1$ Hz, 1H, H-1'), 7.41 - 7.50 (m, 6H, Ar), 7.63-7.66 (m, 4H, Ar), 7.95 (s, 1H, H-6), 10.01 (bs, 1H, NH), 11.74 (s, 1H, NH). ¹⁹F NMR: -74.5 (CF₃), -230.4 (CH₂F).

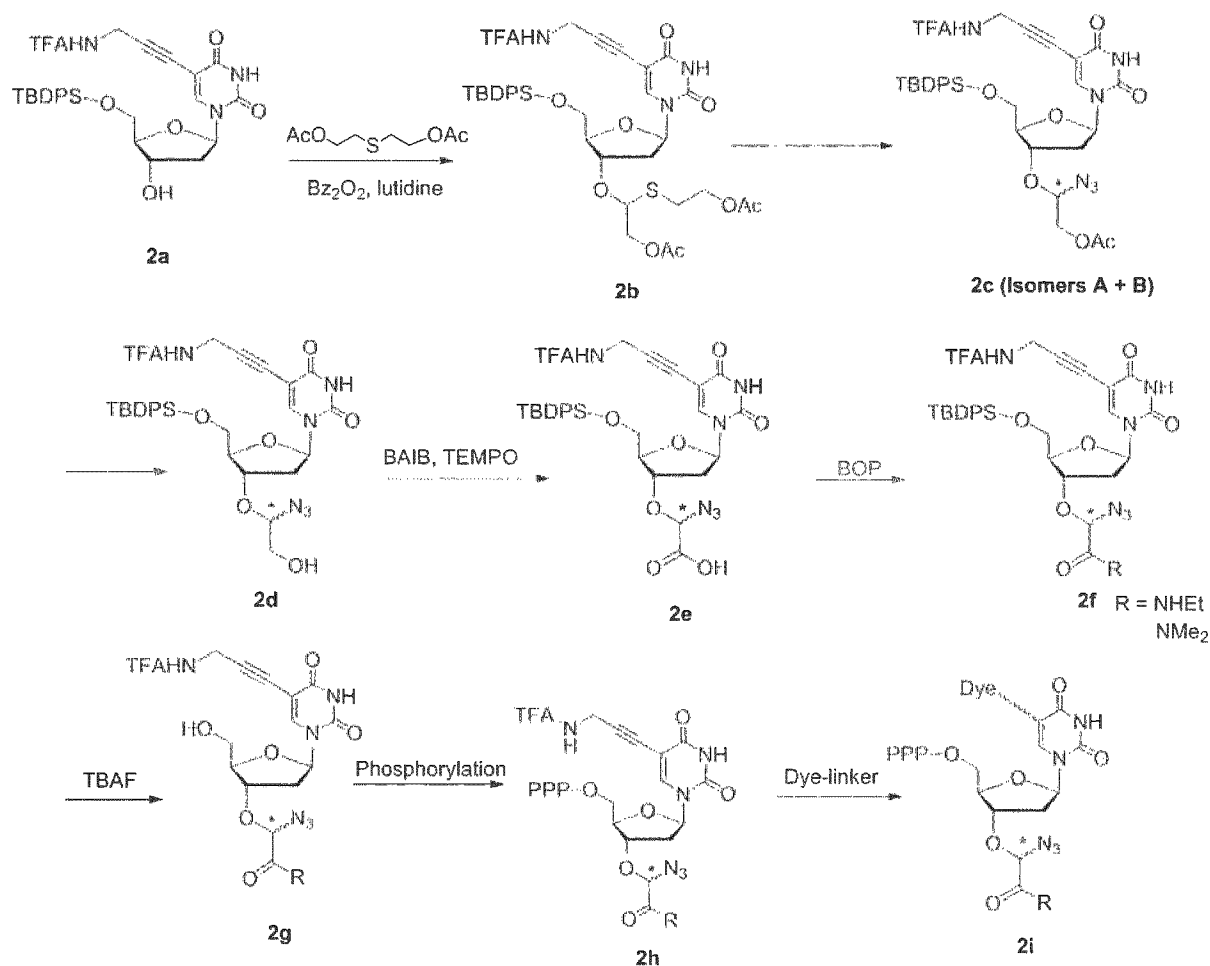
[0101] The starting material **1c** (isomer A) (500 mg, 0.71 mmol) was dissolved in THF (3 mL) and cooled to 4°C in ice-bath. Then TBAF (1.0 M in THF, 5 wt.% water, 1.07 mL, 1.07 mmol) was added slowly over a period of 5 mins. The reaction mixture was slowly warmed up to room temperature. Reaction progress was monitored by TLC (petroleum ether: EtOAc 3:7 (v/v)). The reaction was stopped after 1 hour when no more starting material was visible by TLC. The reaction solution was dissolved in EtOAc (50 mL) and added to NaHCO₃ (60 mL). The two layers were separated and the aqueous layer was extracted with additional DCM (50 mL×2). The organic extractions were combined, dried (MgSO₄), filtered, and evaporated to give a yellow oil. Crude product **1d** (isomer A) was purified by Biotag silica gel column (10 g) using a gradient of petroleum ether: EtOAc 8:2 (v/v) to EtOAc as a white solid (183 mg, yield:56%).

[0102] Isomer A: ¹H NMR (400 MHz, d₆-DMSO): δ 2.24-2.35 (m, 2H, H-2'), 3.56-3.66 (m, 2H, H-5'), 3.96-4.00 (m, 1H, H-4'), 4.23 (s, 2H, CH₂NH), 4.33-4.37 (m, 1H, H-3'), 4.43-4.51 (m, CH₂F), 5.12 (br.s, 1H, CHN₃), 5.23 (br.s, 1H, 5'-OH), 6.07 (t, $J=6.7$ Hz, 1H, H-1'), 8.26 (s, 1H, H-6), 10.11 (br s, 1H, NH), 11.72 (br s, 1H, NH). ¹⁹F NMR: -74.3 (CF₃), -230.5 (CH₂F)

[0103] The same reaction was performed for **1c** (isomer B) at 360 mg scale and afforded the corresponding product **1d** (Isomer B, 150 mg, 63%). ¹H NMR (400 MHz, d₆-DMSO): δ 2.24-2.37 (m, 2H, H-2'), 3.57-3.70 (m, 2H, H-5'), 3.97-4.01 (m, 1H, H-4'), 4.23 (br.s, 2H, CH₂NH), 4.33-4.37 (m, 1H, H-3'), 4.44-4.53 (m, CH₂F), 5.11-5.21 (br.s, 1H, CHN₃), 5.23 (br.s, 1H, 5'-OH), 6.07 (t, $J=6.6$ Hz, 1H, H-1'), 8.23 (s, 1H, H-6), 10.09 (br s, 1H, NH), 11.70 (br s, 1H, NH). ¹⁹F NMR: -74.1 (CF₃), -230.1 (CH₂F).

[0104] The preparation of the corresponding triphosphates **1e** and the further attachment of dye to the nucleobase to afford the fully functional nucleoside triphosphate (ffN) **1f** have been reported in WO 2004/018497 and are generally known by one skilled in the art.

Example 2
Synthesis of Nucleotides with 3'-OH protecting group



Scheme 2.

[0105] Scheme 2 illustrates a synthetic route for the preparation of the modified nucleotides with C-amido substituted azidomethyl as 3'-OH protecting groups. Compounds **2a-2i** employ a modified thymine (T-PA) as the base. Other non-limiting examples of the bases that can be used include Cbz-PA, ADMF-PA, and GPac-PA, the structures of which are shown above in Scheme 1. In the experimental procedure, compound **2f** with a N,N-dimethyl-C(=O)- substituted azidomethyl protecting group (R = NMe₂) and the subsequent reactions were reported. Compounds with other C-amido groups were also prepared, such as N-ethyl-C(=O)- (R = NHEt).

Experimental Procedures

[0106] To a solution of the starting nucleoside **2a** (4.27 g, 6.9 mmol) in anhydrous CH₃CN (50 ml) was added 2,6-lutidine (2.4 mL, 20.7 mmol), S(CH₂CH₂OAc)₂ (12.2 g, 69 mmol)

and then Bz_2O_2 (50% pure, 33.4 g, 69 mmol) at 4°C. The reaction mixture was allowed to warm up slowly to room temperature. The mixture was stirred for other 12 hours. TLC monitored (EtOAc:DCM = 4:6 v/v) to see complete consumption of the starting nucleoside. The reaction was then concentrated under reduced pressure to an oily residue. To this mixture, petroleum ether (800 ml) was added and stirred vigorously for 10 min. The petroleum ether layer was decanted and the residue was repeatedly treated with petroleum ether (x2). The oily residue was then partitioned between DCM/ NaHCO_3 (1:1) (1000 mL). The organic layer was separated and the aqueous layer was further extracted into DCM (2x500 mL). Combined organic layers were dried over MgSO_4 , filtered and the volatiles evaporated under reduced pressure. Crude product **2b** was purified by a Biotag silica gel column (100 g) using a gradient of petroleum ether to petroleum ether: EtOAc 2:8 (v/v) as a pale yellow foam (4.17g, yield: 74%, diastereoisomers).

[0107] To a solution of the starting nucleoside **2b** (4.54 g, 5.56 mmol) in anhydrous CH_2Cl_2 (56 mL) with molecular sieve (4 Å) under N_2 was added cyclohexene (5.62 mL, 56 mmol). The mixture was cooled with an ice bath to 4°C. The solution of sulfonyl chloride (1.13 mL, 13.9 mmol) in DCM (25 mL) was slowly added over 90 minutes under N_2 . After 30 min at that temperature TLC (EtOAc: DCM = 4:6 v/v) indicated 10% of the starting nucleoside **2b** was left. Additional sulfonyl chloride (0.1 mL) was added into reaction mixture. TLC indicated complete conversion of **2b**. Volatiles were evaporated under reduced pressure (and room temperature of 25°C) and the oily residue was quickly subjected to a high vacuum for a further 10 minutes until it foamed. The crude product was purged with N_2 and then dissolved in anhydrous DMF (5 mL) and NaN_3 (1.8 g, 27.8 mmol) added at once. The resulting suspension was stirred at room temperature for 2 hours or until TLC indicated the completion of the reaction and formation of **2c** as two isomers (A and B). The reaction mixture was partitioned between EtOAc: NaHCO_3 (1:1) (1000 mL). The organic layer was separated and the aqueous layer was further extracted into EtOAc (2x300 mL). Combined organic extracts were then dried over MgSO_4 , filtered and the volatiles evaporated under reduced pressure. The two diastereoisomers **2c** (isomer A and B) were separated by a Biotag silica gel column (100 g) using a gradient of petroleum ether to petroleum ether: EtOAc 1:1 (v/v) as pale yellow foam. Isomer A: 1.68 g, yield: 40.7%. Isomer B: 1.79 g, yield: 43.2%.

[0108] To a solution of the starting nucleoside **2c** (isomer A) (1.63 g, 2.2 mmol) in MeOH/THF (1:1) (20 mL) was slowly added NaOH (1M in water) (2.2 mL, 2.2 mmol) and stirred in 4°C. The reaction progress was monitored by TLC (EtOAc: DCM = 4:6 v/v). The reaction was stopped after 1 hour when no more starting material was visible by TLC. The reaction mixture was

partitioned between DCM: NaHCO_3 (1:1) (150 mL). The organic layer was separated and the aqueous layer was further extracted into DCM (2×70 mL). Combined organic extracts were dried over MgSO_4 , filtered and the volatiles evaporated under reduced pressure. The crude product **2d** was purified by a Biotag silica gel column (10 g) using a gradient of petroleum ether: EtOAc (8:2) (v/v) to EtOAc as a pale yellow foam (1.1g, yield:71%).

[0109] The same reaction was repeated for **2c** (isomer B, 1.57g) and afforded the corresponding product **2d** (isomer B, 1.01 g, 69% yield).

[0110] To a solution of the starting nucleoside **2d** (isomer A) (700 mg, 1 mmol) in CH_3CN (10 mL) was treated with TEMPO (63 mg, 0.4 mmol) and BAIB (644 mg, 2 mmol) at room temperature. The reaction progress was monitored by TLC (EtOAc:DCM = 7:3 v/v). The reaction was stopped after 2 hour when no more starting material was visible by TLC. The reaction mixture was partitioned between DCM: $\text{Na}_2\text{S}_2\text{O}_3$ (1:1) (100 mL). The organic layer was separated and the aqueous layer was further extracted into DCM (2×70 mL). Combined organic extracts were then washed with NaCl (sat.). The organic layer was evaporated under reduced pressure without drying over MgSO_4 in order to prevent the product from precipitating out. The crude product **2e** was purified by a Biotag silica gel column (10 g) using a gradient of petroleum ether: EtOAc (1:1) (v/v) to EtOAc to MeOH: EtOAc (1:9) as a pale yellow foam (isomer A, 482 mg, 68% yield).

[0111] The same reaction was performed for **2d** (isomer B, 700 mg) and afforded the corresponding product **2e** (isomer B, 488 mg, 69% yield).

[0112] To a solution of the starting nucleoside **2e** (isomer A) (233 mg, 0.33 mmol) in CH_3CN (10 mL) was added Hunig's base (173 μL , 1 mmol) and BOP (165 mg, 0.39 mmol) at room temperature. After stirring for 5 min, the solution was treated with Me_2NH (2 M in THF) (0.41 mL, 0.82 mmol). The reaction progress was monitored by TLC (MeOH: DCM = 1:9 v/v). The reaction was stopped after 2 hours when no more starting material was visible by TLC. The reaction mixture was partitioned between DCM: NaHCO_3 (1:1) (50 mL). The organic layer was separated and the aqueous layer was further extracted into DCM (2×30 mL). Combined organic extracts were dried over MgSO_4 , filtered and the volatiles evaporated under reduced pressure. The crude product **2f** (R = NMe_2) was purified by a Biotag silica gel column (10 g) using a gradient of DCM:EtOAc (8:2) (v/v) to EtOAc as a pale yellow foam (isomer A, 220 mg, 90% yield).

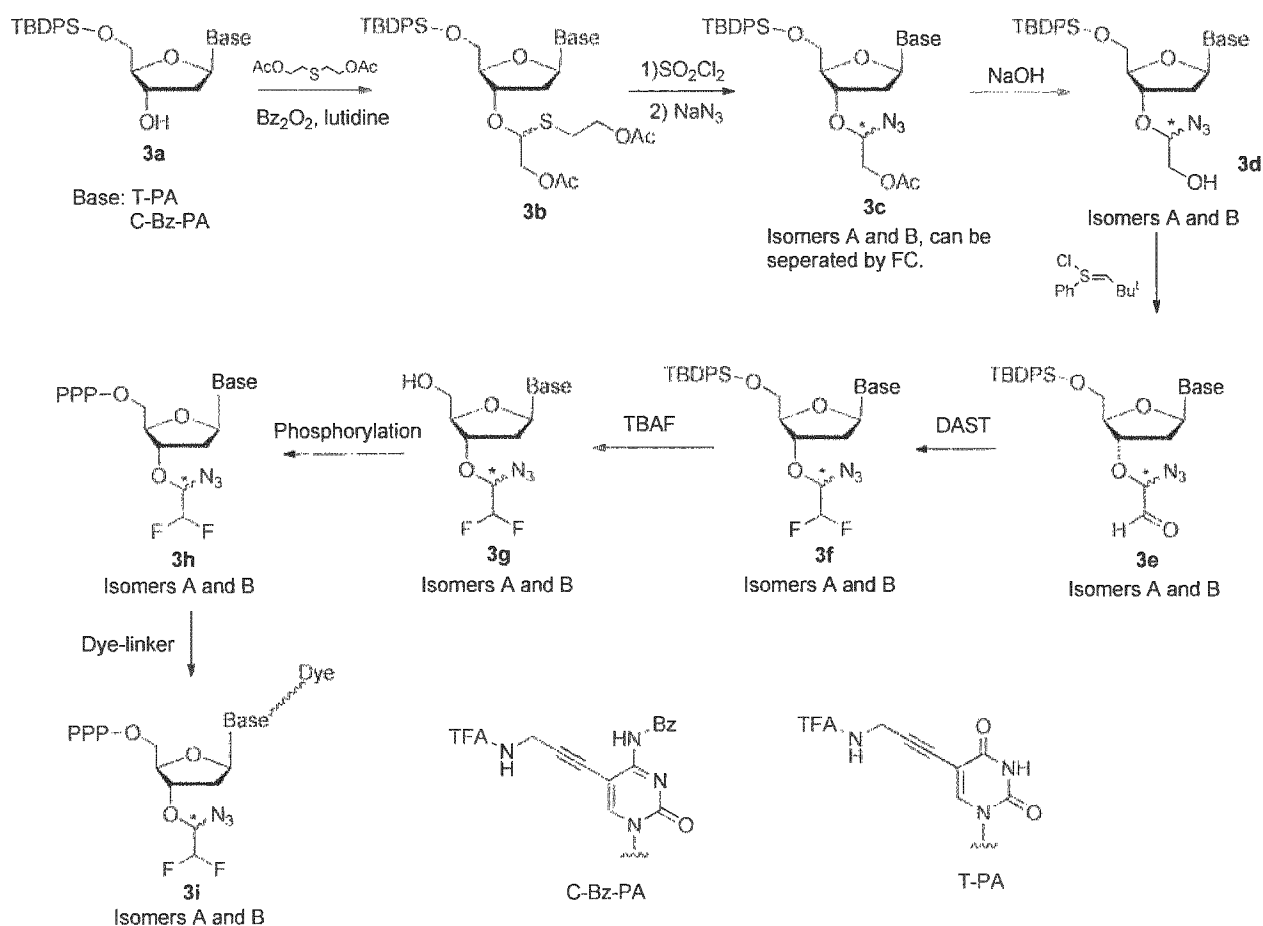
[0113] The same reaction was performed for **2e** (isomer B, 249 mg) and afforded the corresponding product **2f** (isomer B, 240 mg, 92% yield).

[0114] The starting material **2f** (mixture of isomer A and B) (455 mg, 0.61 mmol) was dissolved in THF (2 mL) and cooled to 4°C with ice-bath. Then, TBAF (1.0 M in THF, 5 wt.% water, 1.0 mL, 1.0 mmol) was added slowly over a period of 5 min. The reaction mixture was slowly warmed up to room temperature. The reaction progress was monitored by TLC (EtOAc). The reaction was stopped after 1 hour when no more starting material was visible by TLC. The reaction solution was dissolved in DCM (30 mL) and added to NaHCO₃ (30 mL). The two layers were separated and the aqueous layer was extracted with additional DCM (30 mL×2). The organic extractions were combined, dried (MgSO₄), filtered, and evaporated to give a yellow oil. Crude product **2g** was purified by a Biotag silica gel column (10 g) using a gradient of DCM: EtOAc 8:2 (v/v) to EtOAc to MeOH: EtOAc (2:8) as a white solid (52% yield, 160 mg).

[0115] The preparation of the corresponding triphosphates **2h** and the further attachment of dye to the nucleobase to afford the fully functional nucleoside triphosphate (ffN) **2i** have been reported in WO 2004/018497 and are generally known by one skilled in the art.

Example 3

Synthesis of Nucleotides with 3'-OH protecting group



Scheme 3.

[0116] Scheme 3 illustrates a synthetic route for the preparation of modified nucleotides with difluoromethyl substituted azidomethyl 3'-OH protecting groups. Compounds **3a-3i** employ a modified thymine (T-PA) as the base. Other non-limiting examples of the bases that can be used include Cbz-PA, ADMF-PA, and GPac-PA, the structures of which are shown above in Scheme 1. The procedure for the synthesis of **3b**, **3c** and **3d** were described in Example 2.

Experimental Procedures

[0117] To a solution of the starting nucleoside **3d (isomer A)** (490 mg, 0.7 mmol) and DBU (209 μ L, 1.4 mmol) in anhydrous DCM (5 mL) was added slowly a solution of N-tert-butyl benzene sulfinimidoyl chloride (181 mg, 0.84 mmol) in anhydrous DCM (2 mL) at -78°C. The reaction mixture was stirred for 2 h at -78°C. The reaction progress was monitored by TLC (EtOAc : DCM 4:6 v/v). The reaction was stopped after 2 hours when there was still 10% starting material

left by TLC, to prevent over-reacting. The reaction mixture was partitioned between DCM:NaHCO₃ (1:1) (50 mL). The aqueous layer was further extracted into DCM (2×30 mL). The organic extractions were combined, dried (MgSO₄), filtered, and evaporated to give a yellow oil. The crude product **3e** was purified by a Biotag silica gel column (10 g) using a gradient of petroleum ether: EtOAc (8:2) (v/v) to petroleum ether: EtOAc (2:8) (v/v) as a pale yellow foam (isomer A, 250 mg, 51% yield).

[0118] The same reaction was performed for **3d** (isomer B, 480 mg) and afforded the corresponding product **3e** (isomer B, 240 mg, 50% yield).

[0119] To a solution of the starting nucleoside **3e** (isomer A) (342 mg, 0.49 mmol), EtOH (15 µL, 0.25 mmol) in DCM (2.5 mL) was added slowly to the solution of DAST (181 mg, 0.84 mmol) in DCM (2.5 mL) at 4°C (ice bath). The reaction mixture was stirred for 1 h at 4°C. The reaction progress was monitored by TLC (EtOAc: petroleum ether = 3:7 v/v). The reaction was stopped after 1 hour. The reaction mixture was partitioned between DCM: NaHCO₃ (1:1) (50 mL). The aqueous layer was further extracted into DCM (2×30 mL). The organic extractions were combined, dried (MgSO₄), filtered, and evaporated to give a yellow oil. The crude product **3f** was purified by a Biotag silica gel column (10 g) using a gradient of petroleum ether: EtOAc (9:1) (v/v) to petroleum ether : EtOAc (2:8) (v/v) as a pale yellow foam (isomer A, 100 mg, 28%).

[0120] The same reaction was performed for **3e** (isomer B, 480 mg) and afforded the corresponding product **3f** (isomer B, 240 mg, 50% yield).

[0121] The starting material **3f** (isomer A) (124 mg, 0.17 mmol) was dissolved in THF (2 mL) and cooled to 4°C with an ice bath. Then, TBAF (1.0 M in THF, 5 wt.% water, 255 µL, 10.255 mmol) was added slowly over a period of 5 min. The reaction mixture was slowly warmed up to room temperature. The reaction progress was monitored by TLC (EtOAc). The reaction was stopped after 1 hour when no more starting material was visible by TLC. The reaction solution was dissolved in DCM (30 mL) and added to NaHCO₃ (30 mL). The two layers were separated and the aqueous layer was extracted with additional DCM (30 mL ×2). The organic extractions were combined, dried (MgSO₄), filtered, and evaporated to give a yellow oil. Crude product **3g** was purified by a Biotag silica gel column (4 g) using a gradient of DCM : EtOAc 8:2 (v/v) to EtOAc to MeOH: EtOAc (2:8) as a pale yellow foam (isomer A, 54% yield, 44 mg).

[0122] Isomer A: ¹H NMR (400 MHz, d₆-DMSO): δ 2.24-2.35 (m, 2H, H-2'), 3.56-3.66 (m, 2H, H-5'), 3.96-4.00 (m, 1H, H-4'), 4.23 (s, 2H, CH₂NH), 4.33-4.37 (m, 1H, H-3'), 4.85 (s, 2H,

OCH₂N₃), 5.23 (t, $J=5.1$ Hz, 1H, 5'-OH), 6.07 (t, $J=6.7$ Hz, 1H, H-1'), 8.19 (s, 1H, H-6), 10.09 (br s, 1H, NH), 11.70 (br s, 1H, NH). ¹⁹F NMR: -74.4 (CF₃), -131.6 (CH₂F).

[0123] The same reaction was performed for **3f** (isomer B, 133 mg) and afforded the corresponding product **3g** (isomer B, 48 mg, 54% yield). ¹H NMR (400 MHz, d₆-DMSO): δ 2.27-2.44 (m, 2H, H-2'), 3.58-3.67 (m, 2H, H-5'), 4.00-4.02 (m, 1H, H-4'), 4.24 (d, $J=4.1$ Hz, 2H, CH₂NH), 4.57-4.58 (m, 1H, H-3'), 5.24-5.29 (m, 2H, 5'-OH, OCHN₃), 6.07-6.34 (m, 2H, H-1', CHF₂), 8.19 (s, 1H, H-6), 10.09 (br s, 1H, NH), 11.70 (br s, 1H, NH). ¹⁹F NMR: -74.2 (CF₃), -131.4 (CH₂F).

[0124] The preparation of the corresponding triphosphates **3h** and the further attachment of dye to the nucleobase to afford the fully functional nucleotide (ffN) **3i** have been reported in WO 2004/018497 and are generally known by one skilled in the art.

Example 4 **Thermal Stability Testing of the 3'-OH protecting groups**

[0125] A variety of 3'-OH protecting groups were investigated in regard to their thermal stability (**FIG. 1A**). The thermal stability was evaluated by heating 0.1mM of each 3'-OH protected nucleotide in a pH = 9 buffer (tris-HCl 50mM, NaCl 50mM, tween 0.05%, Mg₂SO₄ 6mM) at 60°C. Various times points were taken and HPLC was used to analyze the formation of un-blocked materials. The stabilities of -CH₂F and -C(O)NHBu were found to be about 2-fold greater than the standard azidomethyl (-CH₂N₃) protecting group. The stability of -CF₂H group was found to be about 10-fold greater than the standard (**FIG. 1B**).

Example 5 **Deprotection of the 3'-OH protecting groups**

[0126] The deprotecting reaction rates of several 3'-OH protecting groups were also studied. The deprotection rate of the standard azidomethyl protecting group was compared with the -CH₂F substituted azidomethyl and -C(O)NHBu substituted azidomethyl. It was observed that both of the more thermally stable 3'-OH blocking groups were removed faster than the standard azidomethyl protecting group using phosphines (1mM THP) as the deprotecting agent. See **FIG. 2A**. For example, the half-life of -CH₂F and -C(O)NHBu was 8.9 minutes and 2.9 minutes respectively, compared to the 20.4 minutes half-life of azidomethyl (**FIG. 2B**).

Example 6 Sequencing Test

[0127] Modified nucleotides with $-\text{CH}_2\text{F}$ (mono-F) substituted azidomethyl 3'-OH protecting group were prepared and their sequencing performance was evaluated on Miseq platforms. It was envisaged that increased thermal stability of 3'-OH protecting groups would lead to a higher quality of nucleotides for sequencing chemistry with less contaminated 3'-unblocked nucleotides. The presence of 3'-unblocked nucleotides in the SBS-sequencing kits would therefore result in pre-phasing events, which were numerated as pre-phasing values.

[0128] Short 12-cycle sequencing experiments were first used to generate phasing and pre-phasing values. Mono-F substituted azidomethyl protected ffNs were used according to the following concentration: ffA-dye 1 (2uM); ffT-dye 2 (10uM), ffC-dye 3 (2uM) and ffG-dye 4 (5uM). Mono-F substituted azidomethyl group comprises both isomer A and B. Two dyes - dye 2 as in standard Miseq kits and dye 5 were used to label ffT. Table 1 shows various nucleotide combinations with A and B isomers of mono-F substituted azidomethyl that were evaluated in regard to phasing and pre-phasing impacts. In all cases, the pre-phasing values were substantially lower than the control that standard V2 Miseq kits nucleotides used (FIG. 3).

Table 1.

Sample	3'-OH Protecting Group	Phasing (%)	Pre-phasing (%)
1	Std Miseq V2 IMX control	0.119	0.177
2	Mono-F-A-isomer	0.11	0.085
3	Mono-F-A isomer (ffT-Dye 5)	0.076	0.032
4	Mono-F-A isomer (A, C and G) + Mono-F-B-ffT-Dye 5	0.095	0.083
5	Mono-F-A (G, ffT-Dye 5) and Mono-F-B (A,C)	0.104	0.05
6	Mono-F-A (G, T) and Mono-F-B (A,C)	0.098	0.095
7	Std Miseq V2 IMX control	0.145	0.167

Sequencing Quality Testing

[0129] 2×400bp sequencing was carried out on Miseq to evaluate the potential of these nucleotides for sequencing quality improvement. The sequencing run was performed according to manufacturer's instructions (Illumina Inc., San Diego, CA). The standard incorporation buffer was replaced with an incorporation buffer containing all mono-F blocked FFNs, each with a separate dye label: ffA-dye 1 (2uM), ffT-dye 2 (1uM), ffC-dye 3 (2uM) and ffG-dye 4 (5uM). The DNA library used was made following the standard TruSeq HT protocol from B cereus genomic DNA.

[0130] In both sequencing experiments (with mono-F block A and B isomer), very low pre-phasing values were observed. Coupled with low phasing values, application of these new nucleotides has generated superior 2×400bp sequencing data with >80% of bases above Q30 in both cases (see **FIG. 4A** for the Q score of isomer A and **FIG. 4B** for the Q score chart of isomer B). These results demonstrate a great improvement compared with Miseq v2 kits (2×250bp, 80% bases >Q30 in a typical R&D sequencing experiments, or 70% bases > Q30 as the stated specs). As shown below, Table 2 summarizes the sequencing data when using all mono-F ffNs-A-isomer in IMX. Table 3 summarizes the sequencing data using all mono-F ffNs-B-isomer in IMX.

Table 2.

Lane	Tiles	Density (K/mm ²)	Clusters PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% Mismatch Rate (PF)	% ≥ Q30	Yield Total (G)
R1	28	690 +/- 14	93.1 +/- 0.7	0.075 / 0.051	13.35	12.43	0.58 ± 0.11	89.7	5
R2	28	690 +/- 14	93.1 +/- 0.7	0.092 / 0.078	13.35	12.43	1.31 ± 0.25	81.9	5
Total								85.8	9.9

Table 3.

Lane	Tiles	Density (K/mm ²)	Clusters PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% Mismatch Rate (PF)	% ≥ Q30	Yield Total (G)
R1	28	816 +/- 9	92.7 +/- 0.6	0.073 / 0.033	15.79	14.64	0.44 ± 0.11	91.2	5.9
R2	28	816 +/- 9	92.7 +/- 0.6	0.078 / 0.059	15.79	14.64	1.03 ± 0.19	83.4	5.9
Total								87.3	11.7

WHAT IS CLAIMED IS:

1. A modified nucleotide or nucleoside molecule comprising a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-hydroxy protecting group forming a structure $-O-C(R)_2N_3$ covalently attached to the 3'-carbon atom, wherein

R is selected from the group consisting of hydrogen, $-C(R^1)_m(R^2)_n$, $-C(=O)OR^3$, $-C(=O)NR^4R^5$, $-C(R^6)_2O(CH_2)_pNR^7R^8$ and $-C(R^9)_2O-Ph-C(=O)NR^{10}R^{11}$;

each R^1 and R^2 is independently selected from hydrogen, optionally substituted alkyl or halogen;

R^3 is selected from hydrogen or optionally substituted alkyl;

each R^4 and R^5 is independently selected from hydrogen, optionally substituted alkyl, optionally substituted aryl, optionally substituted heteroaryl, or optionally substituted aralkyl;

each R^6 and R^9 is selected from hydrogen, optionally substituted alkyl or halogen;

each R^7 , R^8 , R^{10} and R^{11} is independently selected from hydrogen, optionally substituted alkyl, optionally substituted aryl, optionally substituted heteroaryl, or optionally substituted aralkyl;

m is an integer of 0 to 3;

n is an integer of 0 to 3; provided that the total of m + n equals to 3; and

p is an integer of 0 to 6; provided that

R^1 and R^2 cannot both be halogen; and

at least one R is not hydrogen.

2. The modified nucleotide or nucleoside molecule of Claim 1, wherein one of R is hydrogen and the other R is $-C(R^1)_m(R^2)_n$.

3. The modified nucleotide or nucleoside molecule of Claim 1 or 2, wherein $-C(R^1)_m(R^2)_n$ is selected from $-CHF_2$, $-CH_2F$, $-CHCl_2$ or $-CH_2Cl$.

4. The modified nucleotide or nucleoside molecule of any one of Claims 1 to 3, wherein $-C(R^1)_m(R^2)_n$ is $-CHF_2$.

5. The modified nucleotide or nucleoside molecule of any one of Claims 1 to 3, wherein $-C(R^1)_m(R^2)_n$ is $-CH_2F$.

6. The modified nucleotide or nucleoside molecule of Claim 1, wherein one of R is hydrogen and the other R is $-C(=O)OR^3$.

7. The modified nucleotide or nucleoside molecule of Claim 6, wherein R^3 is hydrogen.

8. The modified nucleotide or nucleoside molecule of Claim 1, wherein one of R is hydrogen and the other R is $-C(=O)NR^4R^5$.
9. The modified nucleotide or nucleoside molecule of Claim 1 or 8, wherein both R^4 and R^5 are hydrogen.
10. The modified nucleotide or nucleoside molecule of Claim 1 or 8, wherein R^4 is hydrogen and R^5 is C_{1-6} alkyl.
11. The modified nucleotide or nucleoside molecule of Claim 1 or 8, wherein both R^4 and R^5 are C_{1-6} alkyl.
12. The modified nucleotide or nucleoside molecule of Claim 1, wherein one of R is hydrogen and the other R is $-C(R^6)_2O(CH_2)_pNR^7R^8$.
13. The modified nucleotide or nucleoside molecule of Claim 1 or 12, wherein both R^6 are hydrogen.
14. The modified nucleotide or nucleoside molecule of any one of Claims 1, 12 and 13, wherein both R^7 and R^8 are hydrogen.
15. The modified nucleotide or nucleoside molecule of any one of Claims 1, and 12 to 14, wherein p is 0.
16. The modified nucleotide or nucleoside molecule of any one of Claims 1, and 12 to 14, wherein p is 6.
17. The modified nucleotide or nucleoside molecule of Claim 1, wherein one of R is hydrogen and the other R is $-C(R^9)_2O-Ph-C(=O)NR^{10}R^{11}$.
18. The modified nucleotide or nucleoside molecule of Claim 1 or 17, wherein both R^9 are hydrogen.
19. The modified nucleotide or nucleoside molecule of any one of Claim 1, 17 and 18, wherein both R^{10} and R^{11} are hydrogen.
20. The modified nucleotide or nucleoside molecule of any one of Claim 1, 17 and 18, wherein R^{10} is hydrogen and R^{11} is an amino substituted alkyl.
21. The modified nucleotide or nucleoside molecule of any one of Claims 1 to 14, wherein the 3'-hydroxy protecting group is removed in a deprotecting reaction with a phosphine.
22. The modified nucleotide or nucleoside molecule of Claim 21, wherein the phosphine is tris(hydroxymethyl)phosphine (THP).
23. The modified nucleotide or nucleoside molecule of any one of Claims 1 to 22, wherein said base is linked to a detectable label via a cleavable linker or a non-cleavable linker.

24. The modified nucleotide or nucleoside molecule of any one of Claims 1 to 22, wherein said 3'-hydroxy protecting group is linked to a detectable label via a cleavable linker or a non-cleavable linker.

25. The modified nucleotide or nucleoside molecule of Claim 23 or 24, wherein the linker is cleavable.

26. The modified nucleotide or nucleoside molecule of any one of Claims 23 to 25, wherein the detectable label is a fluorophore.

27. The modified nucleotide or nucleoside molecule of any one of Claims 23 to 26, wherein the linker is acid labile, photolabile or contains a disulfide linkage.

28. A method of preparing a growing polynucleotide complementary to a target single-stranded polynucleotide in a sequencing reaction, comprising incorporating a modified nucleotide molecule of any one of Claims 1 to 27 into the growing complementary polynucleotide, wherein the incorporation of the modified nucleotide prevents the introduction of any subsequent nucleotide into the growing complementary polynucleotide.

29. The method of Claim 28, wherein the incorporation of the modified nucleotide molecule is accomplished by a terminal transferase, a terminal polymerase or a reverse transcriptase.

30. A method for determining the sequence of a target single-stranded polynucleotide, comprising

monitoring the sequential incorporation of complementary nucleotides, wherein at least one complementary nucleotide incorporated is a modified nucleotide molecule of any one of Claims 23 to 27; and

detecting the identity of the modified nucleotide molecule.

31. The method of Claim 30, wherein the identity of the modified nucleotide is determined by detecting the detectable label linked to the base.

32. The method of Claim 30 or 31, wherein the 3'-hydroxy protecting group and the detectable label are removed prior to introducing the next complementary nucleotide.

33. The method of Claim 32, wherein the 3'-hydroxy protecting group and the detectable label are removed in a single step of chemical reaction.

34. A kit comprising a plurality of modified nucleotide or nucleoside molecule of any one of Claims 1 to 27, and packaging materials therefor.

35. The kit of Claim 34, further comprising an enzyme and buffers appropriate for the action of the enzyme.

1/4

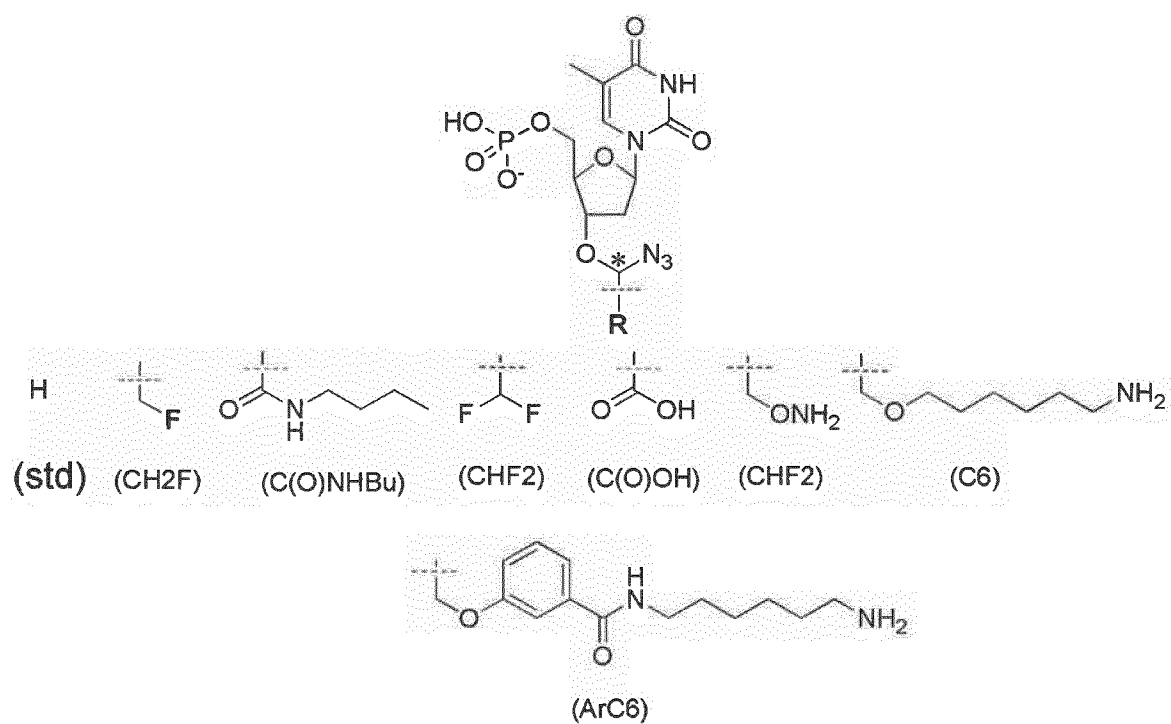


FIG. 1A

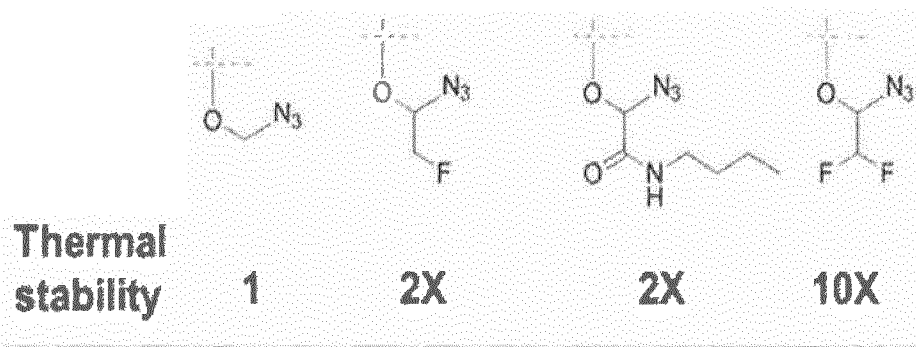


FIG. 1B

2/4

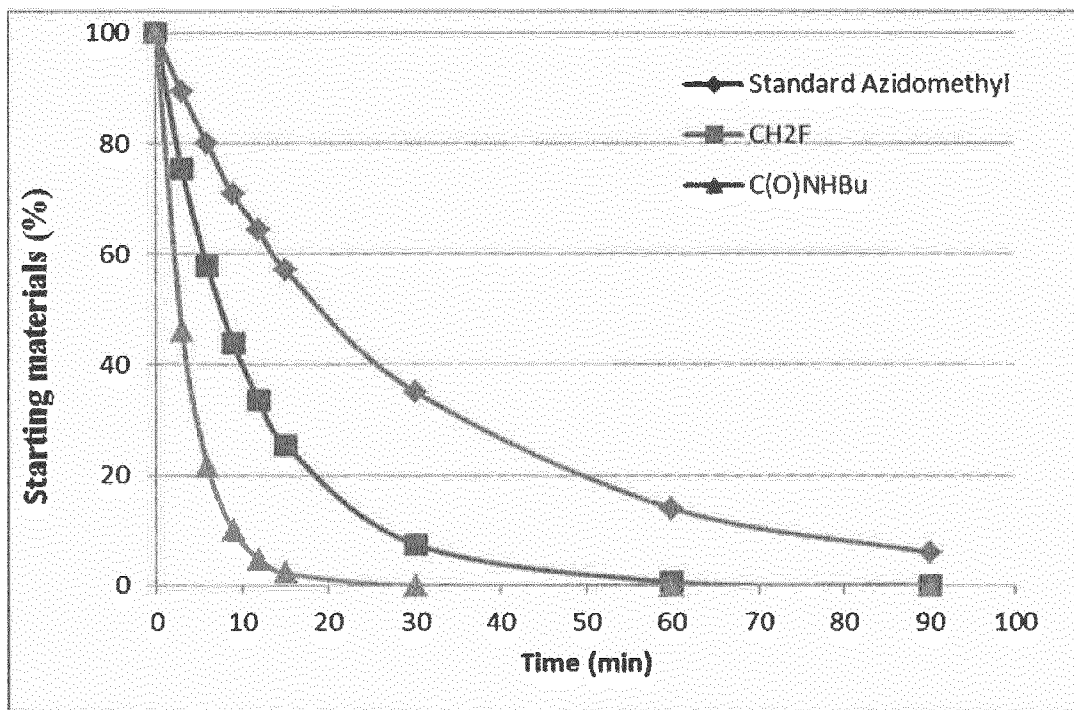


FIG. 2A

	Half-life (min)	Fold improvement
Std	20.4	1
-CH ₂ F	8.9	2.3
-C(O)NHBu	2.9	7

FIG. 2B

3/4

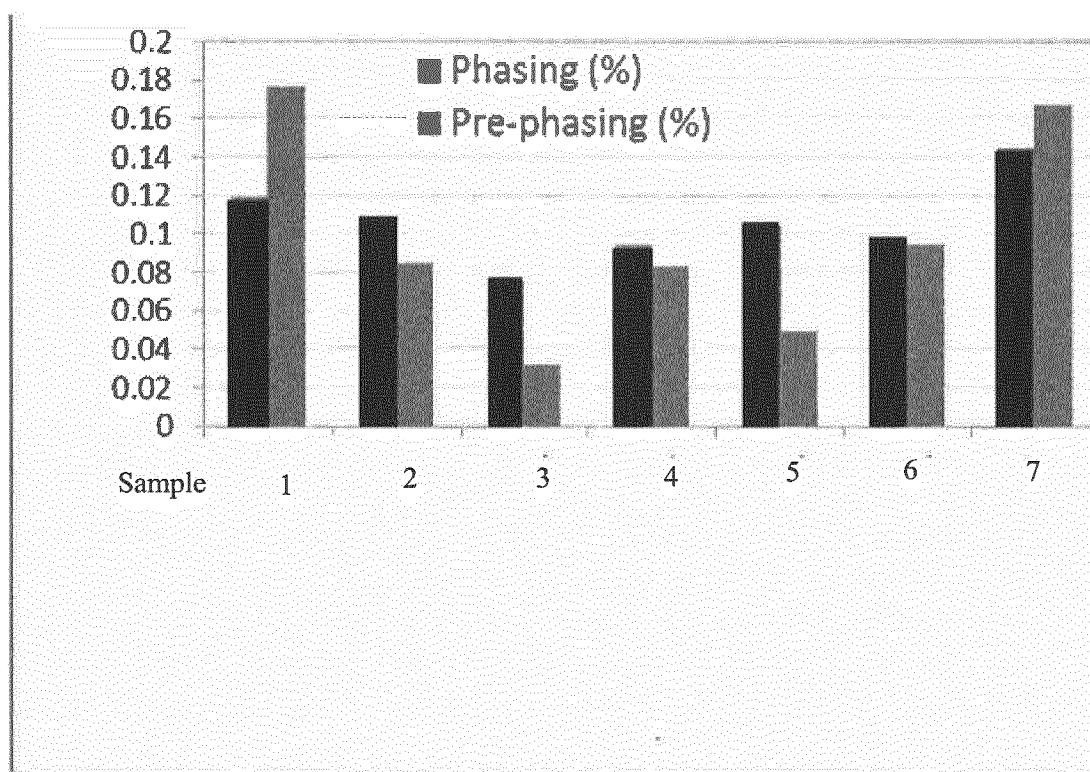


FIG. 3

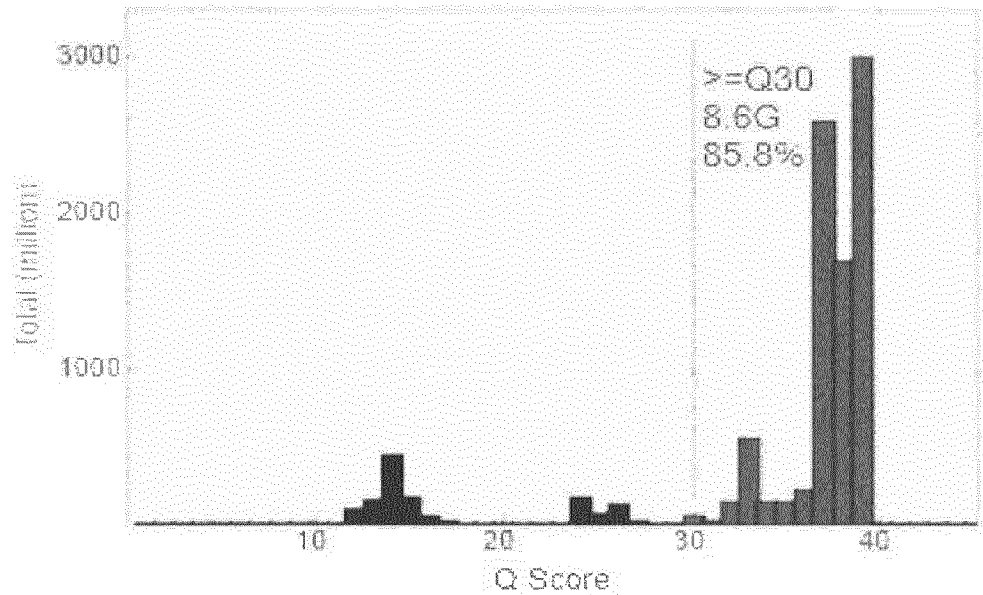


FIG. 4A

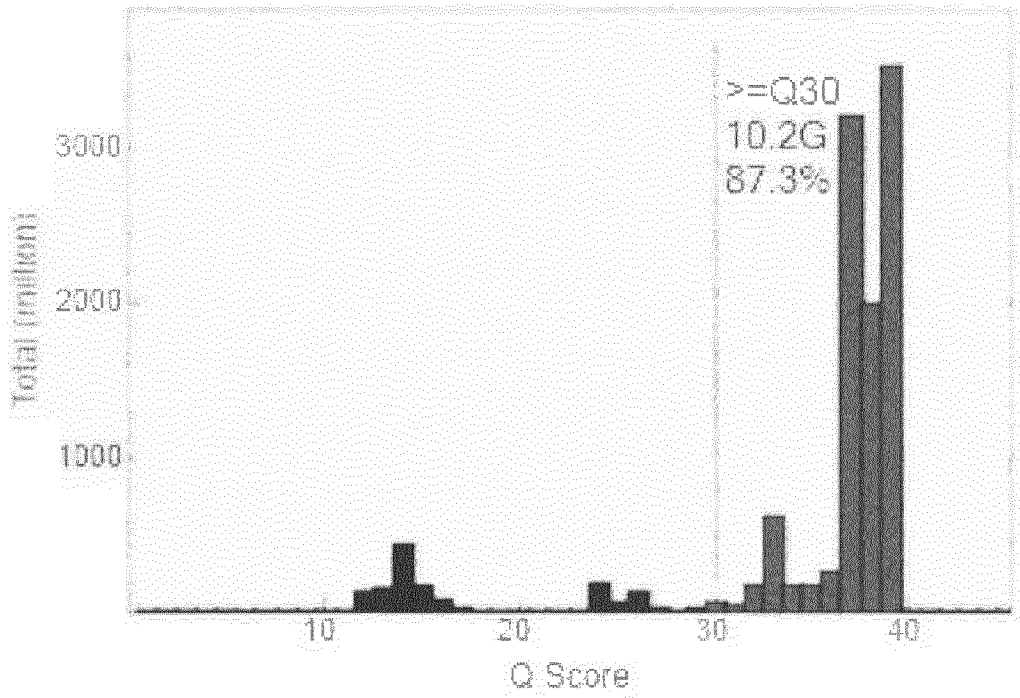


FIG. 4B

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/055466

A. CLASSIFICATION OF SUBJECT MATTER

INV. C07H19/073 C07H19/10 C07H19/14 C07H19/173 C07H19/20
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07H

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, CHEM ABS Data, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/162429 A2 (UNIV COLUMBIA [US]; JU JINGYUE [US]; WU JIAN [US]; LI ZENGMIN [US]) 29 November 2012 (2012-11-29)	1,2,21, 22, 28-30, 34,35
Y	Whole document, in particular claims 1, 4 and 9-19 and Figures 2, 18, 19 and 21	3-20, 23-27, 31-33
Y	----- WO 2004/018497 A2 (SOLEXA LTD [GB]; MILTON JOHN [GB]; WU XIAOLIN [GB]; SMITH MARK [GB]; B) 4 March 2004 (2004-03-04) cited in the application Whole document, in particular claims 1-4 and 29-48 and examples on pages 43-84 ----- -/-	1-35



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 October 2013

Date of mailing of the international search report

29/10/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Lécaillon, Jennifer

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2013/055466

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/083249 A2 (UNIV COLUMBIA [US]; JU JINGYUE [US]; KUMAR SHIV [US]; LI ZENGMIN [US];) 21 June 2012 (2012-06-21)	1,2,21, 22,24, 28-30, 32-35
Y	Claims and examples, in particular claims 3, 20 and 58 and Figure 10	23, 25-27,31
A		3-20
Y	----- US 2006/160081 A1 (MILTON JOHN [GB] ET AL) 20 July 2006 (2006-07-20)	1-23, 27-35
A	Claims and example (in particular page 17)	24
Y	----- WO 2009/054922 A1 (UNIV COLUMBIA [US]; JU JINGYUE [US]; KIM DAE H [US]; GUO JIA [US]; MEN) 30 April 2009 (2009-04-30)	1-23, 25-35
A	Claims and examples (in particular claims 2, 56-57 and Figure 1)	24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2013/055466

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012162429	A2	29-11-2012	NONE
WO 2004018497	A2	04-03-2004	AU 2003259350 A1 11-03-2004
		EP 1530578 A2 18-05-2005	
		EP 2607369 A1 26-06-2013	
		ES 2407681 T3 13-06-2013	
		GB 2395954 A 09-06-2004	
		JP 2006509040 A 16-03-2006	
		JP 2011088898 A 06-05-2011	
		WO 2004018497 A2 04-03-2004	
WO 2012083249	A2	21-06-2012	CN 103282518 A 04-09-2013
		EP 2652153 A2 23-10-2013	
		US 2013264207 A1 10-10-2013	
		WO 2012083249 A2 21-06-2012	
US 2006160081	A1	20-07-2006	US 2006160081 A1 20-07-2006
		US 2009062145 A1 05-03-2009	
		US 2010311178 A1 09-12-2010	
		US 2012156671 A1 21-06-2012	
WO 2009054922	A1	30-04-2009	EP 2209911 A1 28-07-2010
		US 2011039259 A1 17-02-2011	
		WO 2009054922 A1 30-04-2009	