

## (51) International Patent Classification:

C07H 19/00 (2006.01) C12N 15/09 (2006.01)

## (21) International Application Number:

PCT/US2014/050423

## (22) International Filing Date:

8 August 2014 (08.08.2014)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

61/863,649 8 August 2013 (08.08.2013) US

(71) Applicant: THE SCRIPPS RESEARCH INSTITUTE  
[US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

## (72) Inventors: ROMESBERG, Floyd, E.; c/o 10550 North Torrey Pines Road, La Jolla, CA 92037 (US). MALYSHEV, Denis, A.; c/o 10550 North Torrey Pines Road, La Jolla, CA 92037 (US). LI, Lingjun; c/o 10550 North Torrey Pines Road, La Jolla, CA 92037 (US). LAVERGNE, Thomas; c/o 10550 North Torrey Pines Road, La Jolla, CA 92037 (US). LI, Zhengtao; c/o 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

## (74) Agent: UMPIERRE, Alicia; Wilson Sonsini Goodrich &amp; Rosati, 650 Page Mill Road, Palo Alto, CA 92037 (US).

## (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, 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, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

## Declarations under Rule 4.17:

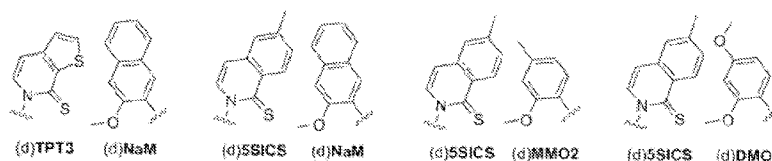
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

## Published:

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

(54) Title: A METHOD FOR THE SITE-SPECIFIC ENZYMATIC LABELLING OF NUCLEIC ACIDS *IN VITRO* BY INCORPORATION OF UNNATURAL NUCLEOTIDES

Figure 1



(57) Abstract: Provided herein are analogs of unnatural nucleotides bearing predominantly hydrophobic nucleobase analogs that form unnatural base pairs during DNA polymerase-mediated replication of DNA or RNA polymerase-mediated transcription of RNA. In this manner, the unnatural nucleobases can be introduced in a site-specific way into oligonucleotides (single or double stranded DNA or RNA), where they can provide for site-specific cleavage, or can provide a reactive linker than can undergo functionalization with a cargo -bearing reagent by means of reaction with a primary amino group or by means of click chemistry with an alkyne group of the unnatural nucleobase linker.

**A METHOD FOR THE SITE-SPECIFIC ENZYMATIC LABELLING OF NUCLEIC  
ACIDS *IN VITRO* BY INCORPORATION OF UNNATURAL NUCLEOTIDES**

**STATEMENT OF GOVERNMENT SUPPORT**

[0001] This invention was made with government support under grant number GM060005, awarded by the National Institutes of Health. The U.S. government has certain rights in the invention.

**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0002] This application claims the benefit of U.S. Provisional Application No 61/863,649 filed August 8, 2013; which is incorporated by reference in its entirety.

**BACKGROUND**

[0003] Oligonucleotides and their applications have revolutionized biotechnology. However, the oligonucleotides including both DNA and RNA each includes only the four natural nucleotides of adenosine (A), guanosine (G), cytosine (C), thymine (T) for DNA, and the four natural nucleotides of adenosine (A), guanosine (G), cytosine (C), and uridine (U) for RNA, and which significantly restricts the potential functions and applications of the oligonucleotides.

[0004] The ability to sequence-specifically synthesize/amplify oligonucleotides (DNA or RNA) with polymerases, for example by PCR or isothermal amplification systems (e.g., transcription with T7 RNA polymerase), has revolutionized biotechnology. In addition to all of the potential applications in nanotechnology, this has enabled a diverse range of new technologies such as the *in vitro* evolution via *SELEX* (Systematic Evolution of Ligands by Exponential Enrichment) of RNA and DNA aptamers and enzymes. See, for example, Oliphant AR, Brandl CJ & Struhl K (1989), Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 proteins, *Mol. Cell Biol.*, 9:2944-2949; Tuerk C & Gold L (1990), Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science*, 249:505-510; Ellington AD & Szostak JW (1990), In vitro selection of RNA molecules that bind specific ligands, *Nature*, 346:818-822.

[0005] Unfortunately, these applications are restricted by the limited chemical/physical diversity present in the natural genetic alphabet (the four natural nucleotides A, C, G, and T in DNA, and the four natural nucleotides A, C, G, and U in RNA). There is accordingly much interest in techniques that would enable the enzymatic synthesis/amplification of oligonucleotides site-specifically labeled with functional groups not present among the nucleotides of the natural genetic alphabet. Currently, the options available for site-specific nucleic acid derivatization include solid-support based chemical synthesis, combined

chemical/enzymatic synthesis, and end-labeling procedures. End-labeling procedures are limited to the oligonucleotide termini, and chemical synthesis is limited to short oligonucleotides (<200 nucleotides for DNA and <70 nucleotides for RNA). Enzymatic functionalization is dependent upon enzymatic recognition of the modification of interest and more problematically it is not site-specific.

### SUMMARY

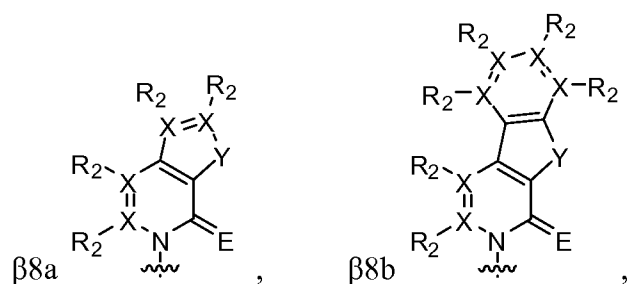
[0006] The compositions and methods described herein are based on the expansion of the genetic alphabet *in vitro*, and provide site-specific incorporation of unnatural nucleotides as described herein, bearing reactive linkers adapted to react with cargo reagents comprising groups of complementary reactivity, or linkers bearing cargos bonded thereto, into any position of any DNA or RNA sequence, for example, by using standard PCR or isothermal transcription methodologies.

[0007] In various embodiments, the linkers are attached to a cargo at the nucleotide triphosphate stage, thus allowing for the direct production of the desired site-specifically labeled oligonucleotide, e.g., by automated polynucleotide synthesis machines such as phosphoroamidite polynucleotide synthesis machines.

[0008] The linkers, in other embodiments, include a reactive center (e.g., primary amine, alkyne, thiol, aldehyde, or azide), providing a reactive linker, allowing for the site-specific modification of the DNA or RNA oligonucleotide after its synthesis. This can be accomplished using a cargo reagent comprising a cargo (e.g., molecule, liposome, nanoparticle, etc.) and a group of reactivity complementary to the reactive center of the reactive linker moiety. In some embodiments, the reactive center of the linker moiety is protected with a standard protecting group. Reaction of a nucleobase disclosed herein bearing a reactive linker (after deprotection, if required), and a cargo reagent incorporating a cargo and a group of complementary reactivity to the reactive linker, serves to provide a nucleobase linked to a cargo via a coupled linker moiety.

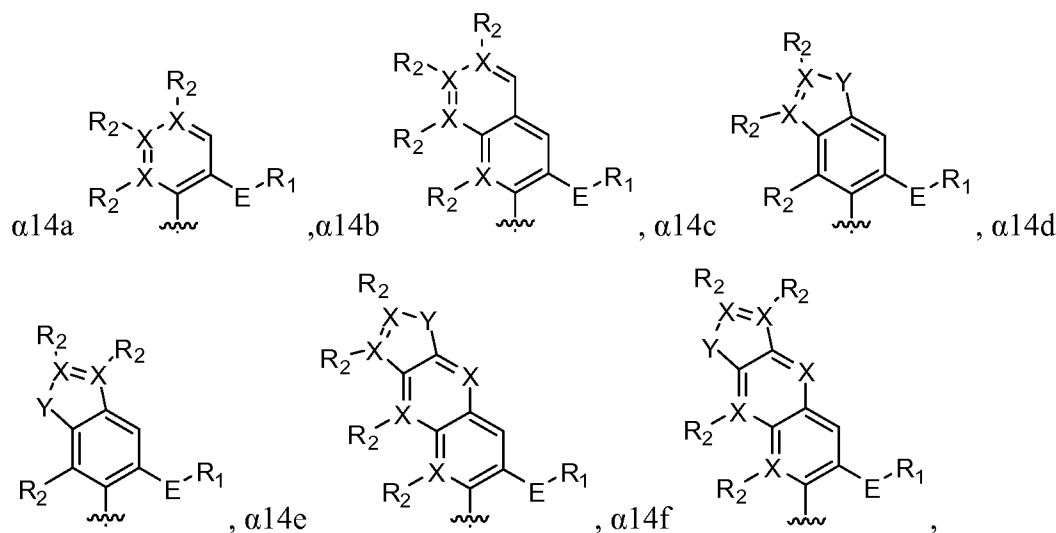
[0009] The compositions of this disclosure, in various embodiments, enable the expansion of the limited repertoire of functionality of the four natural DNA nucleotides and of the four natural RNA nucleotides to include virtually any functionality of interest, site-specifically incorporated into a DNA or RNA oligonucleotide or into a DNA analog or an RNA analog, such as a PNA or an LNA or the like. The cargo optionally includes functionality to enable altered molecular recognition (i.e. for the development of aptamers), altered reactivity (including for catalysis), and/or probes for visualization and/or characterization (i.e. for the development of diagnostics).

[0010] Provided herein, in various embodiments, is a compound comprising a nucleobase analog of any of the following formulas  $\beta 8a$  or  $\beta 8b$ :



wherein each X is independently carbon or nitrogen; wherein each  $R_2$  is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen; and wherein the nucleobase analog is not 4TFP or 7TFP or a linker-derivatization thereof.

**[0011]** Provided herein, in various embodiments, is a compound comprising a nucleobase analog of any of the following formulas:



wherein each X is independently carbon or nitrogen; wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each  $R_2$  is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen; and wherein the nucleobase analog is not FIMO, MIMO, FEMO, PrMO, EMO, MEMO, IMO,

MMO2, DMO, NMO, 5FM, 2OMe, TMO, FDMO, VMO, ZMO, CIMO, TfMO, CNMO, NaM, or QMO.

**[0012]** Here and throughout, the wavy line indicates a point of attachment to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety; or to an analog of a ribosyl, deoxyribosyl, or dideoxyribosyl moiety, such as a locked ribose analog, a peptide group, or the like. In some embodiments, the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, connected to a mono-phosphate, diphosphate, or triphosphate group; optionally comprising an  $\alpha$ -thiotriphosphate,  $\beta$ -thiophosphate, or  $\gamma$ -thiophosphate group; or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

**[0013]** In some embodiments, when referring to either a ribosyl moiety or deoxyribosyl moiety of an unnatural nucleobase provided herein X, dX or (d)X is used, for example, dTPT3 or (d)TPT3 refers to the TPT3 nucleobase bonded to any of the options at the position indicated by the wavy line. Thus, the general appellation of dX refers to compounds having ribose or deoxyribose analogs bonded thereto as indicated by the wavy line. When specifically referring to a ribosyl nucleoside, the prefix “d” is dropped, i.e., TPT3 refers to a ribosyl form. When incorporated into a triphosphate polymerase substrate, (i.e. TPT3TP, dTPT3TP), the nucleotide, or linker-derivatized variants, is considered to be incorporated into an RNA or DNA oligonucleotide using an RNA or DNA polymerase, respectively.

**[0014]** In some embodiments, a ribosyl, deoxyribosyl, or dideoxyribosyl analog of a nucleoside analog provided (e.g.,  $\beta$ 8a,  $\beta$ 8b,  $\alpha$ 14a,  $\alpha$ 14b,  $\alpha$ 14c,  $\alpha$ 14d,  $\alpha$ 14e,  $\alpha$ 14f) comprises a 2' functional group. Examples of functional groups include, without limitation, methoxy, halogen, -O-allyl, -O-methoxyethyl, primary amine, -O-dinitrophenol, -O-dinitrophenyl ether, alkyl, -O-alkyl, thiol, aminoethoxymethyl, aminopropoxymethyl, aminoethyl, cyanoethyl, and guanidinoethyl groups. In some embodiments, the ribosyl, deoxyribosyl, or dideoxyribosyl analog comprises a 4'-thio substitution (e.g., the oxygen of the sugar moiety is replaced with a sulfur).

**[0015]** In some embodiments, an alkyl group of a nucleobase analog includes, without limitation, a methyl, ethyl, propyl, and isopropyl group. In some embodiments, a halogen group of a nucleobase analog includes, without limitation, fluorine, chlorine, bromine, and iodine.

**[0016]** In some embodiments, a reactive linker of a nucleobase analog comprises a functional group including, but not limited to, alkyl, alkenyl, alkynyl, phenyl, benzyl, halo, hydroxyl, carbonyl, aldehyde, haloformyl, carbonate ester, carboxylate, carboxyl, ester, methoxy, hydroperoxy, peroxy, ether, hemiacetal, hemiketal, acetal, ketal, orthoester, methylenedioxy, orthocarbonate ester, carboxamide, primary amine, secondary amine, imide, azide, azo, cyanate, isocyanate, nitrate, nitrile, isonitrile, nitrosooxy, nitro, nitroso, pyridyl, sulfhydryl, sulfide, disulfide, sulfinyl, sulfo, thiocyanate, isothiocyanate, carbonothioyl, phoshino,

phosphono, phosphate, borono, boronate, borino, borinate, and a combination thereof. In some embodiments, a reactive linker of a nucleobase analog comprises an amino group, an acetylenic group, a thiol group, an aldehyde group, or an azide group.

**[0017]** In some embodiments, a ribosyl or deoxyribosyl moiety comprises a triphosphate or an  $\alpha$ -thiotriphosphate group bonded to a 5'-hydroxyl thereof. In some embodiments, a ribosyl or deoxyribosyl moiety is incorporated into a RNA or DNA oligonucleotide chain, respectively, or the ribosyl or deoxyribosyl moiety or analog thereof is incorporated into an RNA or a DNA analog. The RNA analog or DNA analog, in certain embodiments, is a peptide nucleic acid (PNA) or a locked nucleic acid (LNA). The RNA analog or DNA analog, in certain embodiments, is a bicyclic derivative. Bicyclic derivatives include, without limitation, 2'-O,4'-C-ethylene-bridged nucleic acid (ENA), carbocyclic locked nucleic acid (CLNA), cyclohexene nucleic acid (CENA), and 2'-deoxy-2'-N,4'-C-ethylene-locked nucleic acid (AENA). In certain embodiments, the RNA analog or the DNA analog is an acyclic derivative. In certain embodiments, the RNA analog or the DNA analog is an unlocked nucleic acid (UNA). In certain embodiments, the RNA analog or the DNA analog comprises a pyranose ring instead of a ribose. In certain embodiments, the RNA analog or the DNA analog is an arabino nucleic acid (ANA) or a hexitol nucleic acid (HNA).

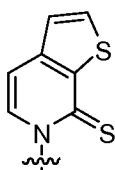
**[0018]** In some embodiments, a ribosyl or deoxyribosyl moiety or analog thereof is substituted with protecting and activating groups suitable for use in an automated chemical oligonucleotide synthesis machine. An example of an automated chemical oligonucleotide synthesis machine is a phosphoroamidite synthesis machine.

**[0019]** In some embodiments, at least one R2 of a nucleobase analog independently comprises a  $-C\equiv C-CH_2NHR_3$  group, wherein R3 is hydrogen or is an amino-protecting group. An example of an amino-protecting group is a dichloroacetyl group. In some embodiments, at least one R2 of a nucleobase analog independently comprises an acetylenic group suitable for use in a click reaction with a cargo reagent comprising a cargo and an acetylene-reactive group. In some embodiments, at least one R2 of a nucleobase analog independently comprises a thiol group suitable for use in a reaction with a cargo reagent comprising a cargo and a thiol-reactive group. In some embodiments, at least one R2 of a nucleobase analog independently comprises an aldehyde group suitable for use in a reaction with a cargo reagent comprising a cargo and an aldehyde-reactive group. In some embodiments, at least one R2 of a nucleobase analog independently comprises an azide group suitable for use in a reaction with a cargo reagent comprising a cargo and an azide-reactive group. In some embodiments, at least one R2 of a nucleobase analog independently comprises  $-C\equiv C-(CH_2)_n-C\equiv CH$ , wherein n is 1, 2, 3, 4, 5, or 6; or R2 is  $-C\equiv C-(CH_2)_{n1}-O-(CH_2)_{n2}-C\equiv CH$ , wherein n1 and n2 are each independently 1, 2,

or 3. In some embodiments, at least one R2 is independently a coupled linker bonded to a cargo by reaction of an amino group and an amino-reactive group. An example of an amino-reactive group is an acylating group or an alkylating group, or is an N-hydroxysuccinimide ester. In some embodiments, at least one R2 is independently a coupled linker bonded to a cargo by reaction of an acetylene group and an acetylene-reactive group. An example of an acetylene-reactive group is an azide group. In certain embodiments, the acetylene group and the azide group are coupled with a copper-catalyzed click reaction. In some embodiments, at least one R2 is independently a coupled linker bonded to a cargo by reaction of a thiol and a thiol-reactive group. In some embodiments, at least one R2 is independently a coupled linker bonded to a cargo by reaction of an aldehyde and an aldehyde-reactive group. In some embodiments, at least one R2 is independently a coupled linker bonded to a cargo by reaction of an azide and an azide-reactive group. An example of an azide-reactive group is a terminal alkyne or a strained cyclooctyne. In some embodiments, at least one R2 is independently hydrogen, the compound comprising an  $\alpha$ -thiotriphosphate group, wherein a cargo reagent comprising a  $\gamma$ -bromo- $\alpha,\beta$ -unsaturated carbonyl, iodo, bromoacetyl, or aziridinylsulfonamide group is coupled thereto.

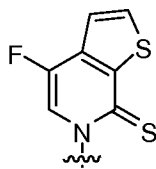
[0020] In some embodiments, a cargo of a nucleobase analog includes, without limitation, proteins, peptides, amino acids, oligonucleotides, small molecule drugs, aliphatic groups, compounds comprising photoreactive groups, compounds comprising chemically-reactive groups, compounds comprising catalytic groups, compounds comprising chloromethylketones, lipids, biotin, fluorescent compounds, fluorescence quenching compounds, liposomes, and nanoparticles.

[0021] Further provided herein, in various embodiments, is the nucleobase analog



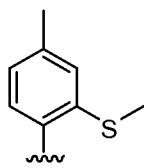
TPT3, and derivatives and analogs thereof.

[0022] Further provided herein, in various embodiments, is the nucleobase analog



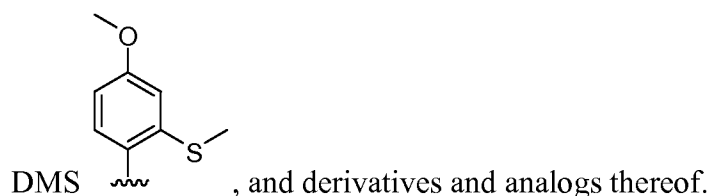
FTPT3, and derivatives and analogs thereof.

[0023] Further provided herein, in various embodiments, is the nucleobase analog

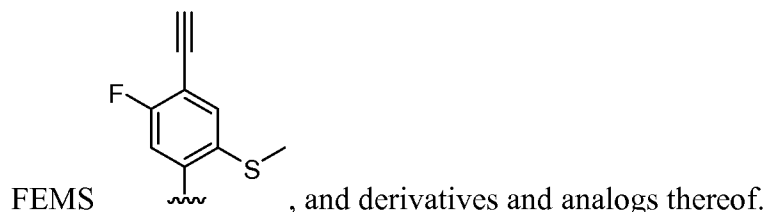


MMS, and derivatives and analogs thereof.

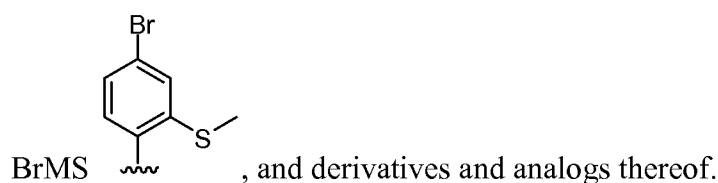
[0024] Further provided herein, in various embodiments, is the nucleobase analog



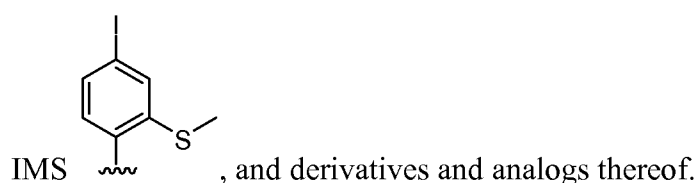
[0025] Further provided herein, in various embodiments, is the nucleobase analog



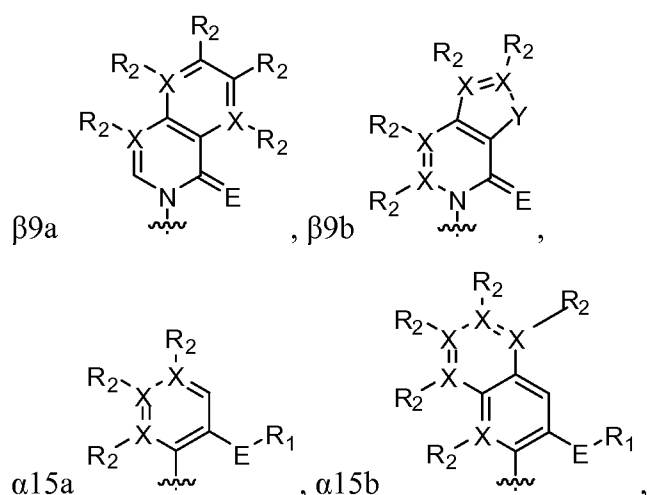
[0026] Further provided herein, in various embodiments, is the nucleobase analog



[0027] Further provided herein, in various embodiments, is the nucleobase analog



[0028] Provided herein, in some embodiments, are nucleobase pairs comprising a first nucleobase analog having any of the formulas  $\beta 9a$  or  $\beta 9b$ ; and a second nucleobase analog having any of the formulas  $\alpha 15a$  or  $\alpha 15b$ :

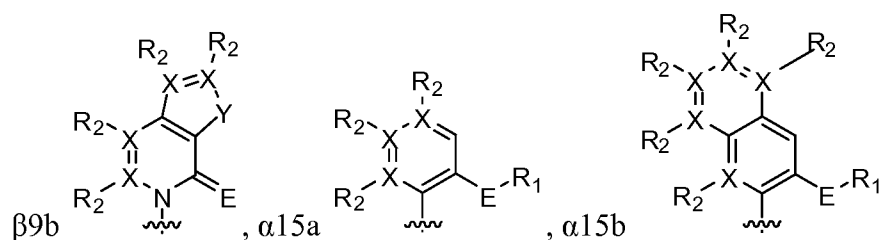


wherein each X is independently carbon or nitrogen; wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each  $R_2$  is optional and when present is



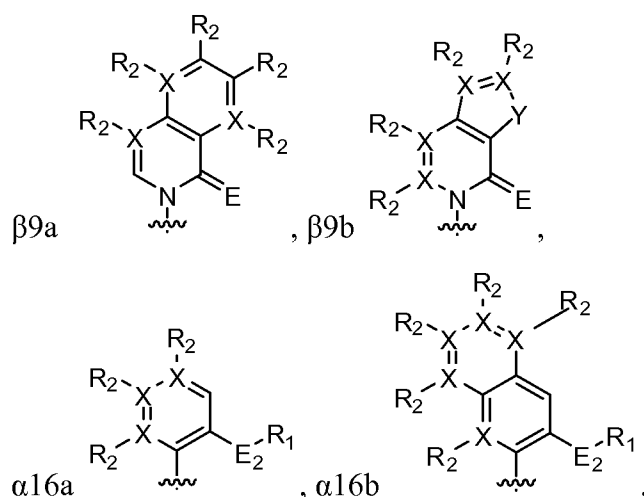
independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently oxygen, sulfur or selenium; and wherein the nucleobase pair is not dICS-dMMO2, dICS-2OMe, dSICS-dMMO2, dSICS-d2OMe, dSNICS-dMMO2, dSNICS-d2OMe, d4SICS-dMMO2, d4SICS-d2OMe, d5SICS-dFIMO, d5SICS-dMIMO, d5SICS-dFEMO, d5SICS-dPrMO, d5SICS-dEMO, d5SICS-dMEMO, d5SICS-dIMO, d5SICS-dMMO2, d5SICS-dDMO, d5SICS-dNMO, d5SICS-d5FM, d5SICS-d2OMe, d5SICS-dTMO, d5SICS-dFDMO, d5SICS-dVMO, d5SICS-dZMO, d5SICS-dCIMO, d5SICS-dTfMO, and d5SICS-dCNMO.

[0029] Provided herein, in some embodiments, is a nucleobase pair comprising a first nucleobase analog having the formula  $\beta 9b$ , and a second nucleobase analog having any of the formulas  $\alpha 15a$  or  $\alpha 15b$ :



wherein each X is independently carbon or nitrogen; wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each  $R_2$  is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; and wherein each E is independently oxygen, sulfur or selenium.

[0030] Provided herein, in some embodiments, is a nucleobase pair comprising a first nucleobase analog having any of the formulas  $\beta 9a$  or  $\beta 9b$ , and a second nucleobase analog having any of the formulas  $\alpha 16a$  or  $\alpha 16b$ :



wherein each X is independently carbon or nitrogen; wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each  $R_2$  is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently oxygen, sulfur or selenium; and wherein each  $E_2$  is independently sulfur or selenium.

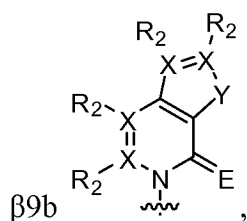
**[0031]** The wavy line indicates a point of attachment to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety; or to an analog of a ribosyl, deoxyribosyl, or dideoxyribosyl moiety, such as a locked ribose analog, a peptide group, or the like. In some embodiments, the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, connected to a monophosphate, diphosphate, or triphosphate group; optionally comprising an  $\alpha$ -thiotriphosphate,  $\beta$ -thiophosphate, or  $\gamma$ -thiophosphate group; or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

**[0032]** In some embodiments, an alkynyl group is an ethynyl or a propynyl group. In some embodiments, an alkyl group is a methyl, ethyl, propyl, or isopropyl group. In some embodiments, a halogen is fluorine, chlorine, bromine, or iodine.

**[0033]** In some embodiments, a ribosyl, deoxyribosyl, or dideoxyribosyl analog of the nucleobase pair comprises a 2' functional group. Exemplary functional groups include, without limitation, methoxy, halogen, -O-allyl, -O-methoxyethyl, primary amine, alkyl, -O-alkyl, thiol, -O-dinitrophenol, -O-dinitrophenyl ether, aminoethoxymethyl, aminopropoxymethyl, aminoethyl, cyanoethyl, and guanidinoethyl groups. In some embodiments, a ribosyl, deoxyribosyl, or dideoxyribosyl analog of the nucleobase pair comprises a 4'-thio substitution.

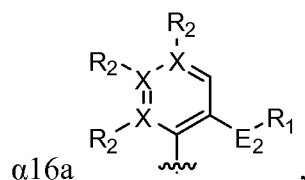
[0034] In some embodiments, a nucleobase pair comprises a nucleobase having the formula  $\alpha 15b$ . In some embodiments, a nucleobase pair comprises a nucleobase having the formula  $\alpha 15a$ . In some embodiments, a nucleobase pair comprises a nucleobase having the formula  $\beta 9b$ . In some embodiments, a nucleobase pair comprises a nucleobase having the formula  $\beta 9b$ , wherein each X is carbon, Y is sulfur, each R<sub>2</sub> is hydrogen, and E is sulfur. In some embodiments, a nucleobase pair comprises a nucleobase having the formula  $\alpha 15b$ , wherein each X is carbon, each R<sub>2</sub> is hydrogen, R<sub>1</sub> is a methyl group, and E is oxygen. In some embodiments, a nucleobase pair comprises a first nucleobase analog having the formula  $\beta 9a$  and a second nucleobase analog having the formula  $\alpha 16a$ . In some embodiments, a nucleobase pair comprises a first nucleobase analog having the formula  $\beta 9a$  and a second nucleobase analog having the formula  $\alpha 16b$ . In some embodiments, a nucleobase pair comprises a first nucleobase analog having the formula  $\beta 9b$  and a second nucleobase analog having the formula  $\alpha 16a$ . In some embodiments, a nucleobase pair comprises a first nucleobase analog having the formula  $\beta 9b$  and a second nucleobase analog having the formula  $\alpha 16b$ .

[0035] Provided herein, in certain embodiments, is a nucleobase pair comprising a first nucleobase analog having the formula  $\beta 9b$  and a second nucleobase analog having the formula  $\beta 9b$ :



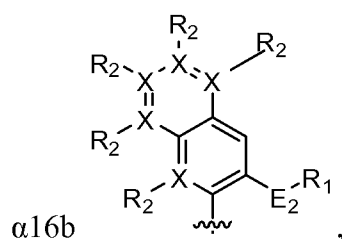
wherein each X is independently carbon or nitrogen; wherein each R<sub>2</sub> is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; and wherein each E is independently oxygen, sulfur or selenium. In some embodiments, the nucleobase pair is a homo-nucleobase pair.

[0036] Provided herein, in certain embodiments, is a nucleobase pair comprising a first nucleobase analog having the formula  $\alpha 16a$  and a second nucleobase analog having the formula  $\alpha 16a$ :



wherein each X is independently carbon or nitrogen; wherein each R<sub>1</sub> is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each R<sub>2</sub> is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; and wherein each E<sub>2</sub> is independently sulfur or selenium. In some embodiments, the nucleobase pair is a homo-nucleobase pair.

**[0037]** Provided herein, in certain embodiments, is a nucleobase pair comprising a first nucleobase analog having the formula  $\alpha 16b$  and a second nucleobase analog having the formula  $\alpha 16b$ :



wherein each X is independently carbon or nitrogen; wherein each R<sub>1</sub> is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each R<sub>2</sub> is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; and wherein each E<sub>2</sub> is independently sulfur or selenium. In some embodiments, the nucleobase pair is a homo-nucleobase pair.

**[0038]** Provided herein, in certain embodiments, is a nucleobase pair comprising a first nucleobase analog having any of the formulas  $\beta 9a$ ,  $\beta 9b$ ,  $\alpha 15a$ ,  $\alpha 15b$ ,  $\alpha 16a$ , or  $\alpha 16b$ ; and a second nucleobase selected from the group consisting of cytosine, guanine, adenine, thymine, uracil, 2-aminoadenin-9-yl, 2-aminoadenine, 2-F-adenine, 2-thiouracil, 2-thio-thymine, 2-thiocytosine, 2-propyl and alkyl derivatives of adenine and guanine, 2-amino-adenine, 2-amino-propyl-adenine, 2-aminopyridine, 2-pyridone, 2'-deoxyuridine, 2-amino-2'-deoxyadenosine 3-

deazaguanine, 3-deazaadenine, 4-thio-uracil, 4-thio-thymine, uracil-5-yl, hypoxanthin-9-yl (I), 5-methyl-cytosine, 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 5-bromo, and 5-trifluoromethyl uracils and cytosines; 5-halouracil, 5-halocytosine, 5-propynyl-uracil, 5-propynyl cytosine, 5-uracil, 5-substituted, 5-halo, 5-substituted pyrimidines, 5-hydroxycytosine, 5-bromocytosine, 5-bromouracil, 5-chlorocytosine, chlorinated cytosine, cyclocytosine, cytosine arabinoside, 5-fluorocytosine, fluoropyrimidine, fluorouracil, 5,6-dihydrocytosine, 5-iodocytosine, hydroxyurea, iodouracil, 5-nitrocytosine, 5-bromouracil, 5-chlorouracil, 5-fluorouracil, and 5-iodouracil, 6-alkyl derivatives of adenine and guanine, 6-azapyrimidines, 6-azo-uracil, 6-azo cytosine, azacytosine, 6-azo-thymine, 6-thio-guanine, 7-methylguanine, 7-methyladenine, 7-deazaguanine, 7-deazaguanosine, 7-deaza-adenine, 7-deaza-8-azaguanine, 8-azaguanine, 8-azaadenine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted adenines and guanines; N4-ethylcytosine, N-2 substituted purines, N-6 substituted purines, O-6 substituted purines, those that increase the stability of duplex formation, universal nucleic acids, hydrophobic nucleic acids, promiscuous nucleic acids, size-expanded nucleic acids, fluorinated nucleic acids, tricyclic pyrimidines, phenoxazine cytidine( [5,4-b][1,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H- pyrimido[5,4-b][1,4]benzothiazin-2(3H)-one), G-clamps, phenoxazine cytidine (9-(2-aminoethoxy)-H-pyrimido[5,4-b][1,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5- b]indol-2-one), pyridoindole cytidine (H-pyrido [3',2':4,5]pyrrolo [2,3-d]pyrimidin-2-one), 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxoacetic acid, wybutosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid methylester, uracil-5-oxoacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine and those in which the purine or pyrimidine base is replaced with a heterocycle.

**[0039]** A base pair comprising one or more unnatural nucleobases is exemplified by the dTPT3PA-dNaM unnatural base pair (i.e. the pair formed between dTPT3PA and dNaM; Figure 1A and 1B). In addition, the orthogonal reactivity of the different reactive centers/linkers developed (i.e. phosphorothioates, amines, and alkynes) allows for the selective arraying of different moieties to the same oligonucleotide (DNA or RNA). Another composition is further

exemplified by the dTPT3PA-dMMO2pCO unnatural base pair, wherein, in various embodiments, the alkynyl group of dMMO2pCO is used to attach one functional group via Copper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC) and, after deprotection of dTPT3PA, the free amine is used to attach a different functional group via N-hydroxysuccinimide (NHS) coupling.

**[0040]** While several other unnatural base pairs have been reported and derivatized with linkers, the linker-derivatized pairs described here, dTPT3PA-dNaM, d5SICSCO-dNaM, dTPT3PA-dMMO2pCO (Figure 2), are both more validated and are better replicated within DNA and better transcribed into RNA. In particular, dTPT3PA-dNaM, is well replicated and thus is suitable for use in practicing the methods disclosed and claimed herein.

**[0041]** Provided herein are unnatural base pairs comprising dTPT3, wherein dTPT3 in some instances, is linker-derivatized. Unnatural base pairs comprising dTPT3 or linker-derivatized dTPT3 (e.g., dTPT3PA) include, without limitation, dTPT3-MMS, dTPT3-DMS, dTPT3-FEMS, dTPT3-BrMS, dTPT3-IMS, dTPT3-dDMN, dTPT3-d4OMe, dTPT3-dIQ, dTPT3-d2MN, dTPT3-d3OMe, dTPT3-dQL, dTPT3-d2Np, dTPT3-dDM4, dTPT3-dDM, dTPT3-dBEN, dTPT3-d3FB, dTPT3-dMM1, dTPT3-dMMO1, dTPT3-dDM2, dTPT3-dDM5, dTPT3-d2Py, dTPT3-d5MPy, dTPT3-dEPy, dTPT3-d3MPy, dTPT3-d34DMPy, dTPT3-d45DMPy, dTPT3-d4MPy, dTPT3-d35DMPy, dTPT3-dBP, dTPT3-dBTp, dTPT3-dBF, dTPT3-dIN, dTPT3-dTp, dTPT3-dBTz, dTPT3-dMTp, dTPT3-dAM, dTPT3-dMAN, dTPT3-dMMAN, dTPT3-dADM, dTPT3-dMMAN, dTPT3-dTOK588, dTPT3-dTOK576, dTPT3-dTOK587, dTPT3-dTOK586, dTPT3-dTOK580, dTPT3-dPhMO, dTPT3-dPyMO1, dTPT3-PyMO2, dTPT3-dPMO1, dTPT3-dPMO2, dTPT3-dPMO3, dTPT3-dFuMO1, dTPT3-dFuMO2, dTPT3-TpMO1, dTPT3-dTpMO2, dTPT3-dFIMO, dTPT3-dIMO, dTPT3-dMIMO, dTPT3-dMEMO, dTPT3-dFEMO, dTPT3-dPrMO, dTPT3-dMMO2, dTPT3-d2OMe, dTPT3-dDMO, dTPT3-dTMO, dTPT3-dNMO, dTPT3-dNOPy, dTPT3-d5FM, dTPT3-dNAM, dTPT3-dAMO1, dTPT3-dAPy, dTPT3-dAMO2, dTPT3-dMAPy, dTPT3-dAMO3, dTPT3-dDMAPy, dTPT3-dFDMO, dTPT3-dVMO, dTPT3-dQMO, dTPT3-dZMO, dTPT3-dCIMO, dTPT3-dTfMO, and dTPT3-CNMO, wherein the dTPT3 complementary base is or is not linker-derivatized (e.g. dMMO2pCO). dTPT3 is illustrated in Figure 9 as a  $\beta$ 6 analog. An example of a linker-derivatized dTPT3 is illustrated in Figure 2, wherein in some instances, R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dTPT3, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11 and 15. In some embodiments, dTPT3 or a linker-derivatized dTPT3 is base paired with dTPT3 or a linker-derivatized dTPT3, to form a homo-nucleobase pair. In some embodiments, dTPT3 or a linker-derivatized dTPT3 is base

paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group, e.g., dTPT3PA. In some embodiments, a linker moiety is not protected with a protecting group, e.g. dTPT3A, wherein in some instances, a protecting group was removed.

**[0042]** Provided herein are unnatural base pairs comprising dMMS, wherein dMMS in some instances, is linker-derivatized. Unnatural base pairs comprising dMMS or linker-derivatized dMMS (e.g., dMMSPA) include, without limitation, d7AI-dMMS, dM7AI-dMMS, dImPy-dMMS, dP7AI-dMMS, dPPP-dMMS, d8Q-dMMS, dICS-dMMS, dPICS-dMMS, dMICS-dMMS, d4MICS-dMMS, d5MICS-dMMS, dNICS-dMMS, dONICS-dMMS, d7OFP-dMMS, d7OTP-dMMS, d4OTP-dMMS, dPYR-dMMS, d4MP-dMMS, d3MP-dMMS, dPPYR-dMMS, dMOP-dMMS, d4MOP-dMMS, dSICS-dMMS, dSNICS-dMMS, d5SICS-dMMS, d4SICS-dMMS, dTPT1-dMMS, dTPT2-dMMS, dFPT1-dMMS, and dFTPT3-dMMS, wherein the dMMS complementary base is or is not linker-derivatized (e.g. pFTPT3pA). dMMS is illustrated in Figure 11 as an  $\alpha$ 14a analog. In some embodiments, a linker-derivatized dMMS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dMMS, include, without limitation,  $\beta$  analogs or linker-derivatized  $\beta$  analogs illustrated in Figures 9, 12, and 13. In some embodiments, dMMS or a linker-derivatized dMMS is base paired with dMMS or a linker-derivatized dMMS, to form a homo-nucleobase pair. In some embodiments, dMMS or a linker-derivatized dMMS is base paired with an  $\alpha$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 8, 10, 11, and 15, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0043]** Provided herein are unnatural base pairs comprising dDMS, wherein dDMS in some instances, is linker-derivatized. Unnatural base pairs comprising dDMS or linker-derivatized dDMS (e.g., dDMSPA) include, without limitation, d7AI-dDMS, dM7AI-dDMS, dImPy-dDMS, dP7AI-dDMS, dPPP-dDMS, d8Q-dDMS, dICS-dDMS, dPICS-dDMS, dMICS-dDMS, d4MICS-dDMS, d5MICS-dDMS, dNICS-dDMS, dONICS-dDMS, d7OFP-dDMS, d7OTP-dDMS, d4OTP-dDMS, dPYR-dDMS, d4MP-dDMS, d3MP-dDMS, dPPYR-dDMS, dMOP-dDMS, d4MOP-dDMS, dSICS-dDMS, dSNICS-dDMS, d5SICS-dDMS, d4SICS-dDMS, dTPT1-dDMS, dTPT2-dDMS, dFPT1-dDMS, dFTPT3-dDMS, wherein the dDMS complementary base is or is not linker-derivatized (e.g. pFTPT3pA). dDMS is illustrated in Figure 11 as an  $\alpha$ 14a analog. In some embodiments, a linker-derivatized dDMS comprises a

functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dDMS, include, without limitation,  $\beta$  analogs or linker-derivatized  $\beta$  analogs illustrated in Figures 9, 12, and 13. In some embodiments, dDMS or a linker-derivatized dDMS is base paired with dDMS or a linker-derivatized dDMS, to form a homo-nucleobase pair. In some embodiments, dMMS or a linker-derivatized dMMS is base paired with an  $\alpha$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 8, 10, 11, and 15, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0044]** Provided herein are unnatural base pairs comprising dFEMS, wherein dFEMS in some instances, is linker-derivatized. Unnatural base pairs comprising dFEMS or linker-derivatized dFEMS (e.g., dFEMSPA) include, without limitation, d7AI-dFEMS, dM7AI-dFEMS, dImPy-dFEMS, dP7AI-dFEMS, dPPP-dFEMS, d8Q-dFEMS, dICS-dFEMS, dPICS-dFEMS, dMICS-dFEMS, d4MICS-dFEMS, d5MICS-dFEMS, dNICS-dFEMS, dONICS-dFEMS, d7OFP-dFEMS, d7OTP-dFEMS, d4OTP-dFEMS, dPYR-dFEMS, d4MP-dFEMS, d3MP-dFEMS, dPPYR-dFEMS, dMOP-dFEMS, d4MOP-dFEMS, dSICS-dFEMS, dSNICS-dFEMS, d5SICS-dFEMS, d4SICS-dFEMS, dTPT1-dFEMS, dTPT2-dFEMS, dFPT1-dFEMS, dFTPT3-dFEMS, wherein the dFEMS complementary base is or is not linker-derivatized (e.g. pFTPT3pA). dFEMS is illustrated in Figure 11 as an  $\alpha$ 14a analog. In some embodiments, a linker-derivatized dFEMS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dFEMS, include, without limitation,  $\beta$  analogs or linker-derivatized  $\beta$  analogs illustrated in Figures 9, 12, and 13. In some embodiments, dFEMS or a linker-derivatized dFEMS is base paired with dFEMS or a linker-derivatized dFEMS, to form a homo-nucleobase pair. In some embodiments, dFEMS or a linker-derivatized dFEMS is base paired with an  $\alpha$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 8, 10, 11, and 15, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0045]** Provided herein are unnatural base pairs comprising dBrMS, wherein dBrMS in some instances, is linker-derivatized. Unnatural base pairs comprising dBrMS or linker-derivatized dBrMS (e.g., dBrMSPA) include, without limitation, d7AI-dBrMS, dM7AI-dBrMS, dImPy-dBrMS, dP7AI-dBrMS, dPPP-dBrMS, d8Q-dBrMS, dICS-dBrMS, dPICS-dBrMS, dMICS-



dBrMS, d4MICS-dBrMS, d5MICS-dBrMS, dNICS-dBrMS, dONICS-dBrMS, d7OFP-dBrMS, d7OTP-dBrMS, d4OTP-dBrMS, dPYR-dBrMS, d4MP-dBrMS, d3MP-dBrMS, dPPYR-dBrMS, dMOP-dBrMS, d4MOP-dBrMS, dSICS-dBrMS, dSNICS-dBrMS, d5SICS-dBrMS, d4SICS-dBrMS, dTPT1-dBrMS, dTPT2-dBrMS, dFPT1-dBrMS, dFTPT3-dBrMS, wherein the dBrMS complementary base is or is not linker-derivatized (e.g. pFTPT3pA). dBrMS is illustrated in Figure 11 as an  $\alpha$ 14a analog. In some embodiments, a linker-derivatized dBrMS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dBrMS, include, without limitation,  $\beta$  analogs or linker-derivatized  $\beta$  analogs illustrated in Figures 9, 12, and 13. In some embodiments, dBrMS or a linker-derivatized dBrMS is base paired with dBrMS or a linker-derivatized dBrMS, to form a homo-nucleobase pair. In some embodiments, dBrMS or a linker-derivatized dBrMS is base paired with an  $\alpha$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 8, 10, 11, and 15, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0046]** Provided herein are unnatural base pairs comprising dIMS, wherein dIMS in some instances, is linker-derivatized. Unnatural base pairs comprising dIMS or linker-derivatized dIMS (e.g., dIMSPA) include, without limitation, d7AI-dIMS, dM7AI-dIMS, dImPy-dIMS, dP7AI-dIMS, dPPP-dIMS, d8Q-dIMS, dICS-dIMS, dPICS-dIMS, dMICS-dIMS, d4MICS-dIMS, d5MICS-dIMS, dNICS-dIMS, dONICS-dIMS, d7OFP-dIMS, d7OTP-dIMS, d4OTP-dIMS, dPYR-dIMS, d4MP-dIMS, d3MP-dIMS, dPPYR-dIMS, dMOP-dIMS, d4MOP-dIMS, dSICS-dIMS, dSNICS-dIMS, d5SICS-dIMS, d4SICS-dIMS, dTPT1-dIMS, dTPT2-dIMS, dFPT1-dIMS, dFTPT3-dIMS, wherein the dIMS complementary base is or is not linker-derivatized (e.g. pFTPT3pA). dIMS is illustrated in Figure 11 as an  $\alpha$ 14a analog. In some embodiments, a linker-derivatized dIMS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dIMS, include, without limitation,  $\beta$  analogs or linker-derivatized  $\beta$  analogs illustrated in Figures 9, 12, and 13. In some embodiments, dIMS or a linker-derivatized dIMS is base paired with dIMS or a linker-derivatized dIMS, to form a homo-nucleobase pair. In some embodiments, dIMS or a linker-derivatized dIMS is base paired with an  $\alpha$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 8, 10, 11, and 15, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a

linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0047]** Provided herein are unnatural base pairs comprising dICS, wherein dICS in some instances, is linker-derivatized. Unnatural base pairs comprising dICS or linker-derivatized dICS (e.g., dICSPA) include, without limitation, dICS-dFIMO, dICS-dMIMO, dICS-dFEMO, dICS-dPrMO, dICS-dEMO, dICS-dMEMO, dICS-dIMO, dICS-dDMO, dICS-dNMO, dICS-d5FM, dICS-dTMO, dICS-dFDMO, dICS-dVMO, dICS-dZMO, dICS-dCIMO, dICS-dTfMO, dICS-dCNMO, dICS-dNAM, dICS-dQMO, wherein the dICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). dICS is illustrated in Figure 9 as a  $\beta$ 2 analog. In some embodiments, a linker-derivatized dICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dICS or a linker-derivatized dICS is base paired with dICS or a linker-derivatized dICS, to form a homo-nucleobase pair. In some embodiments, dICS or a linker-derivatized dICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0048]** Provided herein are unnatural base pairs comprising dPICS, wherein dPICS in some instances, is linker-derivatized. Unnatural base pairs comprising dPICS or linker-derivatized dPICS (e.g., dPICSPA) include, without limitation, dPICS-dFIMO, dPICS-dMIMO, dPICS-dFEMO, dPICS-dPrMO, dPICS-dEMO, dPICS-dMEMO, dPICS-dIMO, dPICS-dMMO2, dPICS-dDMO, dPICS-dNMO, dPICS-d5FM, dPICS-d2OMe, dPICS-dTMO, dPICS-dFDMO, dPICS-dVMO, dPICS-dZMO, dPICS-dCIMO, dPICS-dTfMO, dPICS-dCNMO, dPICS-dNAM, dPICS-dQMO, wherein the dPICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). dPICS is illustrated in Figure 9 as a  $\beta$ 2 analog. In some embodiments, a linker-derivatized dPICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dPICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dPICS or a linker-derivatized dPICS is base paired with dPICS or a linker-derivatized dPICS, to form a homo-nucleobase pair. In some embodiments, dPICS or a linker-derivatized dPICS is base paired with a  $\beta$  nucleobase, including but not

limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0049]** Provided herein are unnatural base pairs comprising dMICS, wherein dMICS in some instances, is linker-derivatized. Unnatural base pairs comprising dMICS or linker-derivatized dMICS (e.g., dMICSPA) include, without limitation, dMICS-dFIMO, dMICS-dMIMO, dMICS-dFEMO, dMICS-dPrMO, dMICS-dEMO, dMICS-dMEMO, dMICS-dIMO, dMICS-dMMO2, dMICS-dDMO, dMICS-dNMO, dMICS-d5FM, dMICS-d2OMe, dMICS-dTMO, dMICS-dFDMO, dMICS-dVMO, dMICS-dZMO, dMICS-dCIMO, dMICS-dTfMO, dMICS-dCNMO, dMICS-dNAM, dMICS-dQMO, wherein the dMICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). dMICS is illustrated in Figure 9 as a  $\beta$ 2 analog. In some embodiments, a linker-derivatized dMICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dMICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dMICS or a linker-derivatized dMICS is base paired with dMICS or a linker-derivatized dMICS, to form a homo-nucleobase pair. In some embodiments, dMICS or a linker-derivatized dMICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0050]** Provided herein are unnatural base pairs comprising d4MICS, wherein d4MICS in some instances, is linker-derivatized. Unnatural base pairs comprising d4MICS or linker-derivatized d4MICS (e.g., d4MICSPA) include, without limitation, d4MICS-dFIMO, d4MICS-dMIMO, d4MICS-dFEMO, d4MICS-dPrMO, d4MICS-dEMO, d4MICS-dMEMO, d4MICS-dIMO, d4MICS-dMMO2, d4MICS-dDMO, d4MICS-dNMO, d4MICS-d5FM, d4MICS-d2OMe, d4MICS-dTMO, d4MICS-dFDMO, d4MICS-dVMO, d4MICS-dZMO, d4MICS-dCIMO, d4MICS-dTfMO, d4MICS-dCNMO, d4MICS-dNAM, d4MICS-dQMO, wherein the d4MICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). d4MICS is illustrated in Figure 9 as a  $\beta$ 2 analog. In some embodiments, a linker-derivatized d4MICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to d4MICS, include, without limitation,  $\alpha$  analogs

or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, d4MICS or a linker-derivatized d4MICS is base paired with d4MICS or a linker-derivatized d4MICS, to form a homo-nucleobase pair. In some embodiments, d4MICS or a linker-derivatized d4MICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0051]** Provided herein are unnatural base pairs comprising d5MICS, wherein d5MICS in some instances, is linker-derivatized. Unnatural base pairs comprising d5MICS or linker-derivatized d5MICS (e.g., d5MICSPA) include, without limitation, d5MICS-dFIMO, d5MICS-dMIMO, d5MICS-dFEMO, d5MICS-dPrMO, d5MICS-dEMO, d5MICS-dMEMO, d5MICS-dIMO, d5MICS-dMMO2, d5MICS-dDMO, d5MICS-dNMO, d5MICS-d5FM, d5MICS-d2OMe, d5MICS-dTMO, d5MICS-dFDMO, d5MICS-dVMO, d5MICS-dZMO, d5MICS-dCIMO, d5MICS-dTfMO, d5MICS-dCNMO, d5MICS-dNAM, d5MICS-dQMO, wherein the d5MICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). d5MICS is illustrated in Figure 9 as a  $\beta$ 2 analog. In some embodiments, a linker-derivatized d5MICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to d5MICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, d5MICS or a linker-derivatized d5MICS is base paired with d5MICS or a linker-derivatized d5MICS, to form a homo-nucleobase pair. In some embodiments, d5MICS or a linker-derivatized d5MICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0052]** Provided herein are unnatural base pairs comprising dNICS, wherein dNICS in some instances, is linker-derivatized. Unnatural base pairs comprising dNICS or linker-derivatized dNICS (e.g., dNICSPA) include, without limitation, dNICS-dFIMO, dNICS-dMIMO, dNICS-dFEMO, dNICS-dPrMO, dNICS-dEMO, dNICS-dMEMO, dNICS-dIMO, dNICS-dDMO, dNICS-dNMO, dNICS-d5FM, dNICS-dTMO, dNICS-dFDMO, dNICS-dVMO, dNICS-dZMO, dNICS-dCIMO, dNICS-MMO2, dNICS-2OMe, dNICS-dTfMO, dNICS-dCNMO, dNICS-dNAM, dNICS-dQMO, wherein the dNICS complementary base is or is not linker-derivatized

(e.g. dDMOpCO, dDMOpCC). dNICS is illustrated in Figure 9 as a  $\beta$ 3 analog. In some embodiments, a linker-derivatized dNICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dNICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dNICS or a linker-derivatized dNICS is base paired with dNICS or a linker-derivatized dNICS, to form a homo-nucleobase pair. In some embodiments, dNICS or a linker-derivatized dNICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0053]** Provided herein are unnatural base pairs comprising dONICS, wherein dONICS in some instances, is linker-derivatized. Unnatural base pairs comprising dONICS or linker-derivatized dONICS (e.g., dONICSPA) include, without limitation, dONICS-dFIMO, dONICS-dMIMO, dONICS-dFEMO, dONICS-dPrMO, dONICS-dEMO, dONICS-dMEMO, dONICS-dIMO, dONICS-dDMO, dONICS-dNMO, dONICS-d5FM, dONICS-dTMO, dONICS-dFDMO, dONICS-dVMO, dONICS-dZMO, dONICS-dCIMO, dONICS-MMO2, dONICS-2OMe, dONICS-dTfMO, dONICS-dCNMO, dONICS-dNAM, dONICS-dQMO, wherein the dONICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). dONICS is illustrated in Figure 9 as a  $\beta$ 3 analog. In some embodiments, a linker-derivatized dONICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dONICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dONICS or a linker-derivatized dONICS is base paired with dONICS or a linker-derivatized dONICS, to form a homo-nucleobase pair. In some embodiments, dONICS or a linker-derivatized dONICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0054]** Provided herein are unnatural base pairs comprising dSICS, wherein dSICS in some instances, is linker-derivatized. Unnatural base pairs comprising dSICS or linker-derivatized dSICS (e.g., dSICSPA) include, without limitation, dSICS-dFIMO, dSICS-dMIMO, dSICS-

dFEMO, dSICS-dPrMO, dSICS-dEMO, dSICS-dMEMO, dSICS-dIMO, dSICS-dDMO, dSICS-dNMO, dSICS-d5FM, dSICS-dTMO, dSICS-dFDMO, dSICS-dVMO, dSICS-dZMO, dSICS-dCIMO, dSICS-dTfMO, dSICS-dCNMO, dSICS-dNAM, dSICS-dQMO, wherein the dSICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). dSICS is illustrated in Figure 9 as a  $\beta$ 5 analog. In some embodiments, a linker-derivatized dSICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dSICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dSICS or a linker-derivatized dSICS is base paired with dSICS or a linker-derivatized dSICS, to form a homo-nucleobase pair. In some embodiments, dSICS or a linker-derivatized dSICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0055]** Provided herein are unnatural base pairs comprising dSNICS, wherein dSNICS in some instances, is linker-derivatized. Unnatural base pairs comprising dSNICS or linker-derivatized dSNICS (e.g., dSNICSPA) include, without limitation, dSNICS-dFIMO, dSNICS-dMIMO, dSNICS-dFEMO, dSNICS-dPrMO, dSNICS-dEMO, dSNICS-dMEMO, dSNICS-dIMO, dSNICS-dDMO, dSNICS-dNMO, dSNICS-d5FM, dSNICS-dTMO, dSNICS-dFDMO, dSNICS-dVMO, dSNICS-dZMO, dSNICS-dCIMO, dSNICS-dTfMO, dSNICS-dCNMO, dSNICS-dNAM, dSNICS-dQMO, wherein the dSNICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). dSNICS is illustrated in Figure 9 as a  $\beta$ 5 analog. In some embodiments, a linker-derivatized dSNICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to dSNICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, dSNICS or a linker-derivatized dSNICS is base paired with dSNICS or a linker-derivatized dSNICS, to form a homo-nucleobase pair. In some embodiments, dSNICS or a linker-derivatized dSNICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0056]** Provided herein are unnatural base pairs comprising d4SICS, wherein d4SICS in some instances, is linker-derivatized. Unnatural base pairs comprising d4SICS or linker-derivatized d4SICS (e.g., d4SICSPA) include, without limitation, d4SICS-dFIMO, d4SICS-dMIMO, d4SICS-dFEMO, d4SICS-dPrMO, d4SICS-dEMO, d4SICS-dMEMO, d4SICS-dIMO, d4SICS-dDMO, d4SICS-dNMO, d4SICS-d5FM, d4SICS-dTMO, d4SICS-dFDMO, d4SICS-dVMO, d4SICS-dZMO, d4SICS-dCIMO, d4SICS-dTfMO, d4SICS-dCNMO, d4SICS-dNAM, d4SICS-dQMO, wherein the d4SICS complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). d4SICS is illustrated in Figure 9 as a  $\beta 5$  analog. In some embodiments, a linker-derivatized d4SICS comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to d4SICS, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, d4SICS or a linker-derivatized d4SICS is base paired with d4SICS or a linker-derivatized d4SICS, to form a homo-nucleobase pair. In some embodiments, d4SICS or a linker-derivatized d4SICS is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0057]** Provided herein are unnatural base pairs comprising d7OFP, wherein d7OFP in some instances, is linker-derivatized. Unnatural base pairs comprising d7OFP or linker-derivatized d7OFP (e.g., d7OFPPA) include, without limitation, d7OFP-dFIMO, d7OFP-dMIMO, d7OFP-dFEMO, d7OFP-dPrMO, d7OFP-dEMO, d7OFP-dMEMO, d7OFP-dIMO, d7OFP-dMMO2, d7OFP-dDMO, d7OFP-dNMO, d7OFP-d5FM, d7OFP-d2OMe, d7OFP-dTMO, d7OFP-dFDMO, d7OFP-dVMO, d7OFP-dZMO, d7OFP-dCIMO, d7OFP-dTfMO, d7OFP-dCNMO, d7OFP-dNAM, d7OFP-dQMO, wherein the d7OFP complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). d7OFP is illustrated in Figure 9 as a  $\beta 5$  analog. In some embodiments, a linker-derivatized d7OFP comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to d7OFP, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, d7OFP or a linker-derivatized d7OFP is base paired with d7OFP or a linker-derivatized d7OFP, to form a homo-nucleobase pair. In some embodiments, d7OFP or a linker-derivatized d7OFP is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase

thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0058]** Provided herein are unnatural base pairs comprising d7OTP, wherein d7OTP in some instances, is linker-derivatized. Unnatural base pairs comprising d7OTP or linker-derivatized d7OTP (e.g., d7OTPPA) include, without limitation, d7OTP-dFIMO, d7OTP-dMIMO, d7OTP-dFEMO, d7OTP-dPrMO, d7OTP-dEMO, d7OTP-dMEMO, d7OTP-dIMO, d7OTP-dMMO2, d7OTP-dDMO, d7OTP-dNMO, d7OTP-d5FM, d7OTP-d2OMe, d7OTP-dTMO, d7OTP-dFDMO, d7OTP-dVMO, d7OTP-dZMO, d7OTP-dCIMO, d7OTP-dTfMO, d7OTP-dCNMO, d7OTP-dNAM, d7OTP-dQMO, wherein the d7OTP complementary base is or is not linker-derivatized (e.g. dDMOpCO, dDMOpCC). d7OTP is illustrated in Figure 9 as a  $\beta$ 5 analog. In some embodiments, a linker-derivatized d7OTP comprises a functional group R, wherein R is a reactive linker comprising a reactive center adapted to bond to a cargo reagent or R is a coupled linker to which a cargo is bonded. Nucleobase analogs which are complementary to d7OTP, include, without limitation,  $\alpha$  analogs or linker-derivatized  $\alpha$  analogs illustrated in Figures 8, 10, 11, and 15. In some embodiments, d7OTP or a linker-derivatized d7OTP is base paired with d7OTP or a linker-derivatized d7OTP, to form a homo-nucleobase pair. In some embodiments, d7OTP or a linker-derivatized d7OTP is base paired with a  $\beta$  nucleobase, including but not limited to, any nucleobase illustrated in Figures 9, 12, and 13, or a derivatized nucleobase thereof. In some embodiments, a linker moiety is protected with a protecting group. In some embodiments, a linker moiety is not protected with a protecting group, wherein in some instances, a protecting group was removed.

**[0059]** Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase TPT3 and a second nucleobase MMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase TPT3 and a second nucleobase DMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase TPT3 and a second nucleobase FEMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase TPT3 and a second nucleobase BrMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase TPT3 and a second nucleobase IMS.

**[0060]** Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase FTPT3 and a second nucleobase MMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase FTPT3 and a second nucleobase DMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase FTPT3 and a second nucleobase FEMS. Further provided herein,



in various embodiments, is a nucleobase pair comprising a first nucleobase FTPT3 and a second nucleobase BrMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase FTPT3 and a second nucleobase IMS.

[0061] Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase 5SICS and a second nucleobase MMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase 5SICS and a second nucleobase DMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase 5SICS and a second nucleobase FEMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase 5SICS and a second nucleobase BrMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase 5SICS and a second nucleobase IMS.

[0062] Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase SICS and a second nucleobase MMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase SICS and a second nucleobase DMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase SICS and a second nucleobase FEMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase SICS and a second nucleobase BrMS. Further provided herein, in various embodiments, is a nucleobase pair comprising a first nucleobase SICS and a second nucleobase IMS.

[0063] Further provided herein, in various embodiments, is a double stranded oligonucleotide duplex wherein a first oligonucleotide strand includes an unnatural nucleobase (e.g., nucleobase analog) disclosed herein, and a second complementary oligonucleotide strand comprising a complementary base-pairing nucleobase in a complementary base-pairing site thereof. For instance, for dTPT3, a second complementary oligonucleotide strand comprises dNaM, dDMO, or dMMO2, or a linker-derivatized analog thereof, in a complementary base-pairing site. In this way, the pairing interaction between the first oligonucleotide strand and the second oligonucleotide strand includes a specific nucleobase-pairing interaction between an unnatural nucleobase moiety provided herein and a complementary nucleobase, which can be a natural or an unnatural nucleobase.

[0064] Provided herein, in some embodiments, is a double stranded oligonucleotide duplex wherein a first oligonucleotide strand comprises a compound having the formula  $\beta$ 8a or  $\beta$ 8b, and a second complementary oligonucleotide strand comprises a complementary base-pairing nucleobase in a complementary base-pairing site thereof. In some embodiments, the complementary base-pairing nucleobase is a compound having the formula  $\alpha$ 14a,  $\alpha$ 14b,  $\alpha$ 14c,  $\alpha$ 14d,  $\alpha$ 14e, or  $\alpha$ 14f. In some embodiments, the complementary base-pairing nucleobase

includes, without limitation, cytosine, guanine, adenine, thymine, uracil, 2-aminoadenin-9-yl, 2-aminoadenine, 2-F-adenine, 2-thiouracil, 2-thio-thymine, 2-thiocytosine, 2-propyl and alkyl derivatives of adenine and guanine, 2-amino-adenine, 2-amino-propyl-adenine, 2-aminopyridine, 2-pyridone, 2'-deoxyuridine, 2-amino-2'-deoxyadenosine 3-deazaguanine, 3-deazaadenine, 4-thio-uracil, 4-thio-thymine, uracil-5-yl, hypoxanthin-9-yl (I), 5-methylcytosine, 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 5-bromo, and 5-trifluoromethyl uracils and cytosines; 5-halouracil, 5-halocytosine, 5-propynyl-uracil, 5-propynyl cytosine, 5-uracil, 5-substituted, 5-halo, 5-substituted pyrimidines, 5-hydroxycytosine, 5-bromocytosine, 5-bromouracil, 5-chlorocytosine, chlorinated cytosine, cyclocytosine, cytosine arabinoside, 5-fluorocytosine, fluoropyrimidine, fluorouracil, 5,6-dihydrocytosine, 5-iodocytosine, hydroxyurea, iodouracil, 5-nitrocytosine, 5-bromouracil, 5-chlorouracil, 5-fluorouracil, and 5-iodouracil, 6-alkyl derivatives of adenine and guanine, 6-azapyrimidines, 6-azo-uracil, 6-azocytosine, azacytosine, 6-azo-thymine, 6-thio-guanine, 7-methylguanine, 7-methyladenine, 7-deazaguanine, 7-deazaguanosine, 7-deaza-adenine, 7-deaza-8-azaguanine, 8-azaguanine, 8-azaadenine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted adenines and guanines; N4-ethylcytosine, N-2 substituted purines, N-6 substituted purines, O-6 substituted purines, those that increase the stability of duplex formation, universal nucleic acids, hydrophobic nucleic acids, promiscuous nucleic acids, size-expanded nucleic acids, fluorinated nucleic acids, tricyclic pyrimidines, phenoxazine cytidine( [5,4-b][1,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido[5,4-b][1,4]benzothiazin-2(3H)-one), G-clamps, phenoxazine cytidine (9-(2-aminoethoxy)-H-pyrimido[5,4-b][1,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5-b]indol-2-one), pyridoindole cytidine (H-pyrido[3',2':4,5]pyrrolo [2,3-d]pyrimidin-2-one), 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxoacetic acid, wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid methylester, uracil-5-oxoacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine and those in which the purine or pyrimidine base is replaced with a heterocycle.

[0065] Provided herein, in some embodiments, is a double stranded oligonucleotide duplex wherein a first oligonucleotide strand comprises a compound having the formula  $\alpha 14a$ ,  $\alpha 14b$ ,  $\alpha 14c$ ,  $\alpha 14d$ ,  $\alpha 14e$ , or  $\alpha 14f$ , and a second complementary oligonucleotide strand comprises a complementary base-pairing nucleobase in a complementary base-pairing site thereof. In some embodiments, the complementary base-pairing nucleobase is a compound having the formula  $\beta 8a$  or  $\beta 8b$ . In some embodiments, the complementary base-pairing nucleobase includes, without limitation, cytosine, guanine, adenine, thymine, uracil, 2-aminoadenin-9-yl, 2-aminoadenine, 2-F-adenine, 2-thiouracil, 2-thio-thymine, 2-thiocytosine, 2-propyl and alkyl derivatives of adenine and guanine, 2-amino-adenine, 2-amino-propyl-adenine, 2-aminopyridine, 2-pyridone, 2'-deoxyuridine, 2-amino-2'-deoxyadenosine 3-deazaguanine, 3-deazaadenine, 4-thio-uracil, 4-thio-thymine, uracil-5-yl, hypoxanthin-9-yl (I), 5-methylcytosine, 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 5-bromo, and 5-trifluoromethyl uracils and cytosines; 5-halouracil, 5-halocytosine, 5-propynyl-uracil, 5-propynyl cytosine, 5-uracil, 5-substituted, 5-halo, 5-substituted pyrimidines, 5-hydroxycytosine, 5-bromocytosine, 5-bromouracil, 5-chlorocytosine, chlorinated cytosine, cyclocytosine, cytosine arabinoside, 5-fluorocytosine, fluoropyrimidine, fluorouracil, 5,6-dihydrocytosine, 5-iodocytosine, hydroxyurea, iodouracil, 5-nitrocytosine, 5-bromouracil, 5-chlorouracil, 5-fluorouracil, and 5-iodouracil, 6-alkyl derivatives of adenine and guanine, 6-azapyrimidines, 6-azo-uracil, 6-azocytosine, azacytosine, 6-azo-thymine, 6-thio-guanine, 7-methylguanine, 7-methyladenine, 7-deazaguanine, 7-deazaguanosine, 7-deaza-adenine, 7-deaza-8-azaguanine, 8-azaguanine, 8-azaadenine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted adenines and guanines; N4-ethylcytosine, N-2 substituted purines, N-6 substituted purines, O-6 substituted purines, those that increase the stability of duplex formation, universal nucleic acids, hydrophobic nucleic acids, promiscuous nucleic acids, size-expanded nucleic acids, fluorinated nucleic acids, tricyclic pyrimidines, phenoxazine cytidine( [5,4-b][1,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido[5,4-b][1,4]benzothiazin-2(3H)-one), G-clamps, phenoxazine cytidine (9-(2-aminoethoxy)-H-pyrimido[5,4-b][1,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5-b]indol-2-one), pyridoindole cytidine (H-pyrido [3',2':4,5]pyrrolo [2,3-d]pyrimidin-2-one), 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-

methythio-N6-isopentenyladenine, uracil-5-oxoacetic acid, wybutosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid methylester, uracil-5-oxoacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)<sub>w</sub>, and 2,6-diaminopurine and those in which the purine or pyrimidine base is replaced with a heterocycle.

**[0066]** In some embodiments, at least one R2 of a nucleobase in a double stranded oligonucleotide duplex is a coupled linker bonded with a cargo. In some embodiments, the cargo is a reporter group, protein, or compound comprising catalytic functionality.

**[0067]** In some embodiments, a first oligonucleotide strand comprising a nucleobase analog disclosed herein is prepared by synthesis with a nucleobase comprising a reactive linker, followed by coupling of the cargo reagent with the first oligonucleotide strand, or wherein the first oligonucleotide strand is prepared by synthesis with a nucleobase comprising a coupled linker bonded to a cargo.

**[0068]** In some embodiments, a double stranded oligonucleotide duplex has a first strand comprising dTPT3 or a derivative thereof, and a second strand comprising dNaM, dDMO, or dMMO2 or a derivative thereof in a complementary base-pairing site thereof.

**[0069]** Further provided herein, in various embodiments, is a method of carrying out a site-specific functionalization of a double stranded oligonucleotide duplex, comprising: incorporating an unnatural nucleobase comprising a reactive linker comprising a reactive center, the nucleobase having any of the following formulas  $\alpha$ 14a,  $\alpha$ 14b,  $\alpha$ 14c,  $\alpha$ 14d,  $\alpha$ 14e,  $\alpha$ 14f,  $\beta$ 8a,  $\beta$ 8b,  $\beta$ 9a,  $\beta$ 9b,  $\alpha$ 15a,  $\alpha$ 15b,  $\alpha$ 16a, or  $\alpha$ 16b, into a first oligonucleotide strand; then, synthesizing a second strand complementary to the first strand, the second strand comprising a nucleobase complementary to the unnatural nucleobase at a site-specific complementary position therein, under conditions such that the first strand and the second strand form a double stranded oligonucleotide duplex; then, contacting the double stranded oligonucleotide duplex incorporating the unnatural nucleobase comprising the reactive linker moiety with a cargo reagent comprising a cargo and a group of complementary reactivity, under conditions suitable for reaction of the reactive linker and the group of complementary reactivity to occur to yield a coupled linker; to provide the functionalized double stranded oligonucleotide duplex with the cargo bonded thereto via a coupled linker.

**[0070]** Further provided herein, in various embodiments, is a method of carrying out a site-specific functionalization of a double stranded oligonucleotide duplex, comprising: incorporating an unnatural nucleobase comprising a reactive linker comprising a reactive center, the nucleobase being selected from the group consisting of d5SICSCO, d5SICSCC, dDMOCO, dDMOCC, dMMO2pCO, dMMO2pCC, dTPT3, dTPT3A, dTPT3PA, dTPT3CO, and

dTPT3CC, into a first oligonucleotide strand; then, synthesizing a second strand complementary to the first strand, the second strand comprising a nucleobase complementary to the unnatural nucleobase at a site-specific complementary position therein, under conditions such that the first strand and the second strand form a double stranded oligonucleotide duplex; then, contacting the double stranded oligonucleotide duplex incorporating the unnatural nucleobase comprising the reactive linker moiety with a cargo reagent comprising a cargo and a group of complementary reactivity, under conditions suitable for reaction of the reactive linker and the group of complementary reactivity to occur to yield a coupled linker; to provide the functionalized double stranded oligonucleotide duplex with the cargo bonded thereto via a coupled linker.

**[0071]** In an embodiment, the linker is bonded with a cargo after the corresponding 5' triphosphate is incorporated into a DNA or RNA oligonucleotide using a DNA or RNA polymerase (after deprotection, if required). In another embodiment, a second oligonucleotide is synthesized that is complementary to the first strand, the second strand containing an unnatural nucleotide at a position complementary to the unnatural nucleotide of the first strand; then, reacting the resulting double stranded oligonucleotide with a cargo-bearing reagent that reacts selectively with the reactive center of the reactive linker provides a functionalized double stranded oligonucleotide (e.g. DNA/DNA, DNA/RNA, or RNA/RNA) bearing a cargo.

**[0072]** Further provided herein, in various embodiments, are structures comprising the formula:  $N1 - Z_x - N2$ , wherein N1 is a nucleotide or analog thereof, or terminal phosphate group; wherein N2 is a nucleotide or analog thereof, or terminal hydroxyl group; wherein Z is a compound of having any of the formulas  $\alpha 14a$ ,  $\alpha 14b$ ,  $\alpha 14c$ ,  $\alpha 14d$ ,  $\alpha 14e$ ,  $\alpha 14f$ ,  $\beta 8a$ ,  $\beta 8b$ ,  $\beta 9a$ ,  $\beta 9b$ ,  $\alpha 15a$ ,  $\alpha 15b$ ,  $\alpha 16a$ , or  $\alpha 16b$ ; and wherein x is an integer from 1 to 20. In some embodiments, the structure is an oligonucleotide. In some embodiments, the oligonucleotide is a ribonucleic acid or a deoxyribonucleic acid. In some embodiments, the oligonucleotide is an aptamer or nucleic acid based sensor. In some embodiments, the oligonucleotide is a molecular beacon. In some embodiments, the oligonucleotide is an RNA analog or DNA analog.

**[0073]** Further provided herein, in various embodiments, is a method for identifying a nucleic acid aptamer comprising at least one compound provided herein (e.g.  $\alpha 14a$ ,  $\alpha 14b$ ,  $\alpha 14c$ ,  $\alpha 14d$ ,  $\alpha 14e$ ,  $\alpha 14f$ ,  $\beta 8a$ ,  $\beta 8b$ ,  $\beta 9a$ ,  $\beta 9b$ ,  $\alpha 15a$ ,  $\alpha 15b$ ,  $\alpha 16a$ ,  $\alpha 16b$ ), as having an enhanced desired property with respect to a target molecule, the method comprising: a) preparing a candidate mixture of single-stranded nucleic acid aptamers, wherein each nucleic acid aptamer of the candidate mixture of aptamers comprises at least one compound provided herein (e.g.  $\alpha 14a$ ,  $\alpha 14b$ ,  $\alpha 14c$ ,  $\alpha 14d$ ,  $\alpha 14e$ ,  $\alpha 14f$ ,  $\beta 8a$ ,  $\beta 8b$ ,  $\beta 9a$ ,  $\beta 9b$ ,  $\alpha 15a$ ,  $\alpha 15b$ ,  $\alpha 16a$ ,  $\alpha 16b$ ); then, b) contacting the candidate mixture with the target molecule under conditions suitable for binding to the target molecule to occur; then, c) partitioning the one or more nucleic acid aptamer having the desired

property with respect to the target molecule from among the aptamers of the candidate mixture; and then, d) amplifying the one or more nucleic acid aptamer with the desired property, in vitro, to yield the one or more nucleic acid aptamers, having an enhanced desired property with respect to the target molecule. In some embodiments, the method further comprises step e) repeating steps c) and d). In some embodiments, the single-stranded nucleic acids aptamers are selected from the group consisting of single-stranded DNA and single-stranded RNA. In some embodiments, the desired property is a binding affinity for a target. In some embodiments, the desired property is a target binding induced activity. In some embodiments, the desired property is a catalytic activity. In some embodiments, the desired property is an inhibition activity, an activation activity, or a modification of an inhibition activity or an activation activity. In some embodiments, the desired property is a structure switching activity or a modification of a structure switching activity. In some embodiments, the desired property is a cooperative activity. In some embodiments, the desired activity is an enhanced cellular uptake efficacy.

**[0074]** Further provided herein, in certain embodiments, is an aptamer comprising a compound having any of the following formulas:  $\alpha 14a$ ,  $\alpha 14b$ ,  $\alpha 14c$ ,  $\alpha 14d$ ,  $\alpha 14e$ ,  $\alpha 14f$ ,  $\beta 8a$ ,  $\beta 8b$ ,  $\beta 9a$ ,  $\beta 9b$ ,  $\alpha 15a$ ,  $\alpha 15b$ ,  $\alpha 16a$ ,  $\alpha 16b$ .

#### BRIEF DESCRIPTION OF THE FIGURES

**[0075]** **Figure 1** illustrates the pairing of dTPT3-dNaM, d5SICS-dNaM, d5SICS-dMMO2, d5SICS-dDMO in DNA or RNA.

**[0076]** **Figure 2** illustrates the linker-derivatized nucleotides dTPT3R, d5SICSR, dMMO2R, dMMO2pR, dDMOR, dNaMpR, dNaMpR, dFEMO and dEMO, where R = 3-aminopropyn-1-yl (denoted as A, e.g. dTPT3A); R = dichloroacetyl-3-aminopropyn-1-yl (denoted as PA); R = 4-oxahepta-1,6-diyn-1-yl (denoted as CO); R = hepta-1,6-diyn-1-yl (denoted as CC).

**[0077]** **Figure 3** shows an overview of the phosphorothioate-based post-synthesis site-specific labeling strategy.

**[0078]** **Figure 4** shows an overview of the amino-based post-synthesis site-specific labeling strategy. The linker-modified nucleotides can also be directly incorporated into the template DNA using standard solid phase synthesis of oligonucleotides and the corresponding phosphoroamidites.

**[0079]** **Figure 5** shows an overview of the click chemistry-based post-synthesis site-specific labeling strategy. The linker-modified nucleotides can also be directly incorporated into the template DNA using standard solid phase synthesis of oligonucleotides and the corresponding phosphoroamidites.

**[0080]** **Figure 6** shows representative data illustrating the post-amplification labeling of DNA analyzed via streptavidin (SA) gel shift. The faster migrating band corresponds to dsDNA, while

the slower migrating band corresponds to the 1:1 complex between dsDNA and streptavidin. (A) The labeling efficiency is 72% with d5SICSPA-dNaM and 80% with dTPT3PA-dNaM. (B) The labeling efficiency is 6%, 84%, and 56% with d5SICSPA-dNaM at the first (primer 1: unnatural base pair at position 1), ninth (primer 2: unnatural base pair at position eleven), and eleventh (primer 3: unnatural base pair at position nine) position. The corresponding labeling efficiencies with dTPT3PA-dNaM are 72%, 94%, and 81%.

[0081] **Figure 7** shows gel electrophoresis data confirming the full-length transcription of RNA containing linker-derivatized analogs of 5SICS or MMO2.

[0082] **Figure 8** illustrates 12 groupings of  $\alpha$  nucleobase analogs,  $\alpha 1 - \alpha 12$ .

[0083] **Figure 9** illustrates 6 groupings of  $\beta$  nucleobase analogs,  $\beta 1 - \beta 6$ .

[0084] **Figure 10** illustrates 2 groupings of  $\alpha$  nucleobase analogs,  $\alpha 13$  and  $\alpha 14$ ; wherein each X is independently carbon or nitrogen; wherein each R1 is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each R2 is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each R is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen.

[0085] **Figure 11** illustrates examples of  $\alpha 14a$  nucleobase analogs, including linker-derivatized nucleobase analogs, MMSPCO and MMSPA.

[0086] **Figure 12** illustrates 2 groupings of  $\beta$  nucleobase analogs,  $\beta 7$  and  $\beta 8$ ; wherein each X is independently carbon or nitrogen; wherein each R2 is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen; and wherein R is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker

comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded.

[0087] **Figure 13** illustrates examples of  $\beta$ 8 linker-derivatized nucleobase analogs.

[0088] **Figure 14** shows percentages of unnatural base pairs retained in DNA after amplification during 6 rounds of screenings.

[0089] **Figure 15** illustrates  $\alpha$  nucleobase analogs,  $\alpha$ 15a,  $\alpha$ 15b,  $\alpha$ 16a, and  $\alpha$ 16b; wherein each X is independently carbon or nitrogen; wherein each R1 is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each R2 is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently oxygen, sulfur or selenium; and wherein each E2 is independently sulfur or selenium.

[0090] **Figure 16** illustrates  $\beta$  nucleobase analogs,  $\beta$ 9a and  $\beta$ 9b; and wherein each X is independently carbon or nitrogen; wherein each R1 is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each R2 is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently oxygen, sulfur or selenium.

#### DETAILED DESCRIPTION

[0091] Phrases such as “under conditions suitable to provide” or “under conditions sufficient to yield” or the like, in the context of methods of synthesis, as used herein refers to reaction conditions, such as time, temperature, solvent, reactant concentrations, and the like, that are within ordinary skill for an experimenter to vary, that provide a useful quantity or yield of a reaction product. It is not necessary that the desired reaction product be the only reaction product or that the starting materials be entirely consumed, provided the desired reaction product can be isolated or otherwise further used.



[0092] By "chemically feasible" is meant a bonding arrangement or a compound where the generally understood rules of organic structure are not violated; for example a structure within a definition of a claim that would contain in certain situations a pentavalent carbon atom that would not exist in nature would be understood to not be within the claim. The structures disclosed herein, in all of their embodiments are intended to include only "chemically feasible" structures, and any recited structures that are not chemically feasible, for example in a structure shown with variable atoms or groups, are not intended to be disclosed or claimed herein.

[0093] An "analog" of a chemical structure, as the term is used herein, refers to a chemical structure that preserves substantial similarity with the parent structure, although it may not be readily derived synthetically from the parent structure. In some embodiments, a nucleotide analog is an unnatural nucleotide. In some embodiments, a nucleoside analog is an unnatural nucleoside. A related chemical structure that is readily derived synthetically from a parent chemical structure is referred to as a "derivative."

[0094] Accordingly, a "DNA analog" or an "RNA analog", as the terms are used herein, refer to DNA or RNA-like polymers such as peptide nucleic acids (PNA), locked nucleic acids (LNA), phosphorothioates, and the like, which are well-known in the art. DNA and RNA analogs, as well as DNA and RNA, can be synthesized in automated synthesizers, e.g., using phosphoroamidite chemistry or other chemical approaches adapted for synthesizer use.

[0095] DNA includes, but is not limited to, cDNA and genomic DNA. DNA may be attached, by covalent or non-covalent means, to another biomolecule, including, but not limited to, RNA and peptide. RNA includes coding RNA, e.g. messenger RNA (mRNA). In some embodiments, RNA is rRNA, RNAi, snoRNA, microRNA, siRNA, snRNA, exRNA, piRNA, long ncRNA, or any combination or hybrid thereof. In some instances, RNA is a component of a ribozyme. DNA and RNA can be in any form, including, but not limited to, linear, circular, supercoiled, single-stranded, and double-stranded.

[0096] The term "amino protecting group" or "amino-protected" as used herein refers to those groups intended to protect an amino group against undesirable reactions during synthetic procedures and which can later be removed to reveal the amine. Commonly used amino protecting groups are disclosed in Protective Groups in Organic Synthesis, Greene, T.W.; Wuts, P. G. M., John Wiley & Sons, New York, NY, (3rd Edition, 1999). Amino protecting groups include acyl groups such as formyl, acetyl, propionyl, pivaloyl, t-butylacetyl, 2-chloroacetyl, 2-bromoacetyl, trifluoroacetyl, trichloroacetyl, o-nitrophenoxyacetyl,  $\alpha$ -chlorobutyryl, benzoyl, 4-chlorobenzoyl, 4-bromobenzoyl, 4-nitrobenzoyl, and the like; sulfonyl groups such as benzenesulfonyl, p-toluenesulfonyl and the like; alkoxy- or aryloxy-carbonyl groups (which form urethanes with the protected amine) such as benzyloxycarbonyl (Cbz), p-

chlorobenzyloxycarbonyl, p-methoxybenzyloxycarbonyl, p-nitrobenzyloxycarbonyl, 2-nitrobenzyloxycarbonyl, p-bromobenzyloxycarbonyl, 3,4-dimethoxybenzyloxycarbonyl, 3,5-dimethoxybenzyloxycarbonyl, 2,4-dimethoxybenzyloxycarbonyl, 4-methoxybenzyloxycarbonyl, 2-nitro-4,5-dimethoxybenzyloxycarbonyl, 3,4,5-trimethoxybenzyloxycarbonyl, 1-(p-biphenyl)-1-methylethoxycarbonyl,  $\alpha,\alpha$ -dimethyl-3,5-dimethoxybenzyloxycarbonyl, benzhydryloxycarbonyl, t-butyloxycarbonyl (Boc), diisopropylmethoxycarbonyl, isopropylloxycarbonyl, ethoxycarbonyl, methoxycarbonyl, allyloxycarbonyl (Alloc), 2,2,2-trichloroethoxycarbonyl, 2-trimethylsilylethylloxycarbonyl (Teoc), phenoxycarbonyl, 4-nitrophenoxycarbonyl, fluorenyl-9-methoxycarbonyl (Fmoc), cyclopentylloxycarbonyl, adamantylloxycarbonyl, cyclohexylloxycarbonyl, phenylthiocarbonyl and the like; aralkyl groups such as benzyl, triphenylmethyl, benzyloxymethyl and the like; and silyl groups such as trimethylsilyl and the like. Amine protecting groups also include cyclic amino protecting groups such as phthaloyl and dithiosuccinimidyl, which incorporate the amino nitrogen into a heterocycle. Typically, amino protecting groups include formyl, acetyl, benzoyl, pivaloyl, t-butylacetyl, phenylsulfonyl, Alloc, Teoc, benzyl, Fmoc, Boc and Cbz. Protecting groups also include methyl carbamate, 9-fluorenylmethyl carbamate, 2,2,2-trichloroethyl carbamate, t-butyl carbamate, 2-(trimethylsilyl)ethyl carbamate, allyl carbamate, benzyl carbamate, m-nitrophenyl carbamate, trifluoroacetamide, benzylamine, allylamine, and tritylamine. Protecting groups also include, formamides, acetamides, trifluoroacetamides, p-toluenesulfonyl, trifluoromethanesulfonyl, trimethylsilylethanesulfonamide, and tert-butylsulfonyl. It is well within the skill of the ordinary artisan to select and use the appropriate amino protecting group for the synthetic task at hand.

**[0097]** DNA and RNA analogs include PNA (peptide nucleic acid) and LNA (locked nucleic acid) analogs.

**[0098]** A peptide nucleic acid (PNA) is a synthetic DNA/RNA analog wherein a peptide-like backbone replaces the sugar-phosphate backbone of DNA or RNA. PNA oligomers show higher binding strength and greater specificity in binding to complementary DNAs, with a PNA/DNA base mismatch being more destabilizing than a similar mismatch in a DNA/DNA duplex. This binding strength and specificity also applies to PNA/RNA duplexes. PNAs are not easily recognized by either nucleases or proteases, making them resistant to enzyme degradation. PNAs are also stable over a wide pH range. See also Nielsen PE, Egholm M, Berg RH, Buchardt O (December 1991). "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide", *Science* **254** (5037): 1497–500. doi:10.1126/science.1962210. PMID 1962210; and, Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Nordén B, and Nielsen PE (1993), "PNA

Hybridizes to Complementary Oligonucleotides Obeying the Watson-Crick Hydrogen Bonding Rules". *Nature* **365** (6446): 566–8. doi:10.1038/365566a0. PMID 7692304

[0099] A locked nucleic acid (LNA) is a modified RNA nucleotide, wherein the ribose moiety of an LNA nucleotide is modified with an extra bridge connecting the 2' oxygen and 4' carbon. The bridge "locks" the ribose in the 3'-*endo* (North) conformation, which is often found in the A-form duplexes. LNA nucleotides can be mixed with DNA or RNA residues in the oligonucleotide whenever desired. Such oligomers can be synthesized chemically and are commercially available. The locked ribose conformation enhances base stacking and backbone pre-organization. See, for example, Kaur, H; Arora, A; Wengel, J; Maiti, S (2006), "Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes", *Biochemistry* **45** (23): 7347–55. doi:10.1021/bi060307w. PMID 16752924; Owczarzy R.; You Y., Groth C.L., Tataurov A.V. (2011), "Stability and mismatch discrimination of locked nucleic acid-DNA duplexes.", *Biochem.* **50** (43): 9352–9367. doi:10.1021/bi200904e. PMC 3201676. PMID 21928795; Alexei A. Koshkin; Sanjay K. Singh, Poul Nielsen, Vivek K. Rajwanshi, Ravindra Kumar, Michael Meldgaard, Carl Erik Olsen, Jesper Wengel (1998), "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition", *Tetrahedron* **54** (14): 3607–30. doi:10.1016/S0040-4020(98)00094-5; and, Satoshi Obika; Daishu Nanbu, Yoshiyuki Hari, Ken-ichiro Morio, Yasuko In, Toshimasa Ishida, Takeshi Imanishi (1997), "Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3'-*endo* sugar puckering", *Tetrahedron Lett.* **38** (50): 8735–8. doi:10.1016/S0040-4039(97)10322-7.

[00100] A molecular beacon or molecular beacon probe is an oligonucleotide hybridization probe that can detect the presence of a specific nucleic acid sequence in a homogenous solution. Molecular beacons are hairpin shaped molecules with an internally quenched fluorophore whose fluorescence is restored when they bind to a target nucleic acid sequence. See, for example, Tyagi S, Kramer FR (1996), "Molecular beacons: probes that fluoresce upon hybridization", *Nat Biotechnol.* **14** (3): 303–8. PMID 9630890; Täpp I, Malmberg L, Rennel E, Wik M, Syvänen AC (2000 Apr), "Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5'-nuclease TaqMan assay and Molecular Beacon probes", *Biotechniques* **28** (4): 732–8. PMID 10769752; and, Akimitsu Okamoto (2011), "ECHO probes: a concept of fluorescence control for practical nucleic acid sensing", *Chem. Soc. Rev.* **40**: 5815–5828.

[00101] In some embodiments, a nucleobase is generally the heterocyclic base portion of a nucleoside. Nucleobases may be naturally occurring, may be modified, may bear no similarity to natural bases, and may be synthesized, e.g., by organic synthesis. In certain embodiments, a

nucleobase comprises any atom or group of atoms capable of interacting with a base of another nucleic acid with or without the use of hydrogen bonds. In certain embodiments, an unnatural nucleobase is not derived from a natural nucleobase. It should be noted that unnatural nucleobases do not necessarily possess basic properties, however, are referred to as nucleobases for simplicity. In some embodiments, when referring to a nucleobase, a “(d)” indicates that the nucleobase can be attached to a deoxyribose or a ribose.

**[00102]** In some embodiments, a nucleoside is a compound comprising a nucleobase moiety and a sugar moiety. Nucleosides include, but are not limited to, naturally occurring nucleosides (as found in DNA and RNA), abasic nucleosides, modified nucleosides, and nucleosides having mimetic bases and/or sugar groups. Nucleosides include nucleosides comprising any variety of substituents. A nucleoside can be a glycoside compound formed through glycosidic linking between a nucleic acid base and a reducing group of a sugar.

**[00103]** In some embodiments, a nucleotide is a compound in which the sugar moiety of a nucleoside forms an ester with phosphoric acid, more preferably a mono-, di- or tri-phosphate ester. The sugar moiety of such a nucleoside or nucleotide may be ribofuranosyl, 2'-deoxyribofuranosyl, or 2'-substituted ribofuranosyl having a substituent at the 2'-position. Likewise, the phosphoric acid moiety may be thiophosphoric acid. Namely, the sugar and phosphoric acid moieties may be in the same form as found in known nucleosides, nucleotides, or derivatives thereof. A ribonucleotide whose sugar moiety is ribofuranosyl can be used as a member constituting RNA. A deoxyribonucleotide whose sugar moiety is deoxyribofuranosyl can be used as a member constituting DNA. A nucleotide can be a nucleoside further comprising a phosphate linking group. Nucleotides may include nucleosides containing a phosphate moiety.

**[00104]** A class of unnatural base pairs, exemplified by d5SICS-dNaM and d5SICS-dMMO2 (Figure 1), has been developed and shown by us to be replicated (including via PCR) and transcribed by a wide range of natural polymerases with efficiencies and fidelities approaching those of a natural base pair (See Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E., PCR with an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* 2009, 131 (41), 14620-14621; Seo, Y. J.; Matsuda, S.; Romesberg, F. E., Transcription of an Expanded Genetic Alphabet. *J. Am. Chem. Soc.* 2009, 131 (14), 5046-5047; Lavergne T.; Degardin M.; Malyshev D.A.; Quach H.T.; Dhami K.; Ordoukhanian P.; Romesberg, F.E.; Expanding the scope of replicable unnatural DNA: Stepwise optimization of a predominantly hydrophobic base pair. *J. Am. Chem. Soc.* 2013, 135, 5408-5419; Seo, Y.J., Malyshev D.A., Lavergne T., Ordoukhanian P., and Romesberg, F.E., *J. Am. Chem. Soc.* 2011, 133, 19878.). These unnatural base pairs are formed between nucleotide analogs bearing unnatural, predominantly hydrophobic nucleobases. The base pairs are shown in Figure 1, with each unnatural nucleotide being incorporated into

oligonucleotides at complementary (*i.e.* paring) positions; the nucleobases are understood to be bonded at the position indicated by the wavy line to the 1'-position of a ribosyl or 2'-deoxyribosyl moiety, which is itself incorporated into RNA or DNA by phosphate or phosphorothioate groups bonding the 3' and 5' hydroxyl groups of the ribosyl or deoxyribosyl groups, respectively, as it is in the case with fully natural nucleic acids. The base pairing thus takes place as part of a complementary base paired structure as is well-known in the formation of oligonucleotide duplex structures. Provided herein, in various embodiments, is an unnatural nucleobase dTPT3 and its linker-derivatized variants, which are thought to pair with unnatural nucleobases dNaM, dMMO2, and dDMO (or their linker-derivatized variants) in a similar fashion.

[00105] We have demonstrated that the unnatural nucleotides dTPT3 and dTPT3<sup>PA</sup> are efficiently incorporated into DNA by DNA polymerases opposite dNaM (Figure 2). Both dTPT3 and dTPT3<sup>PA</sup> (PA = dichloroacetyl-3-aminopropyn-1-yl) as well as other linker-derivatized variants of dTPT3 (Figure 3) including, when R = 3-aminopropyn-1-yl (dTPT3<sup>A</sup>), R = 4-oxahept-1,6-diyn-1-yl (dTPT3<sup>CO</sup>), or R = hepta-1,6-diyn-1-yl (dTPT3<sup>CC</sup>) are also expected to pair with dNaM, dDMO or dMMO2, or linker-derivatized analogs thereof. Incorporation rates of dTPT3 and linker-derivatized variants thereof, opposite dNaM approach those of a natural base pair. Additional unnatural base pairs identified with efficient incorporation rates include dTPT3-dFEMO, dTPT3-dFIMO, dTPT3-dIMO, dFTPT3-dNaM, dFTPT3-dFEMO, dFTPT3-dFIMO, and dFTPT3-dIMO.

[00106] Further provided herein, in various embodiments, are unnatural nucleotides with nucleobase analogs including,  $\alpha$  analogs (e.g., any one of Figures 8, 10, 11, 15 and derivatives thereof);  $\beta$  analogs (e.g., any one of Figures 9, 12, 13, 16 and derivatives thereof); d5SICS<sup>CO</sup>, d5SICS<sup>CC</sup>, dDMO<sup>CO</sup>, dDMO<sup>CC</sup>, dMMO2<sup>pCO</sup>, dMMO2<sup>pCC</sup>, dTPT3, dTPT3<sup>PA</sup>, dTPT3<sup>A</sup>, dTPT3<sup>CO</sup>, dTPT3<sup>CC</sup>, and ribosyl forms thereof, and analogs thereof (See Figure 2); in the form of nucleosides, nucleoside 5' triphosphates, and analogs thereof (e.g., ribosyl and 2'-deoxyribosyl), nucleotides and analogs thereof (e.g., ribosyl and 2'-deoxyribosyl, phosphate and phosphorothioate), including nucleotide reagents derived there from for use in RNA/DNA synthesis (DMT-protected phosphoramidites) and for use in enzymatic incorporation into oligonucleotides as by PCR or T7 RNA polymerase-mediate transcription, and incorporated into nucleic acids (oligonucleotides) such as DNA and RNA. The compounds comprising the unnatural nucleobase analogs can also be incorporated into DNA analogs or into RNA analogs, such as PNA, LNA, and other like polynucleotide-analogous polymers. Exemplary nucleobase analogs provided herein include  $\beta$ 8 analogs comprising the formulas  $\beta$ 8a and  $\beta$ 8b, as shown in Figure 12, wherein each X is independently carbon or nitrogen; wherein each R<sub>2</sub> is optional and

when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen; and wherein R is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded. Examples of  $\beta$ 8 analogs include dTPT3 and linker-derivatized analogs thereof. Exemplary nucleobase analogs provided herein include  $\alpha$ 14 analogs comprising the formulas  $\alpha$ 14a-  $\alpha$ 14f, as shown in Figure 10, wherein each X is independently carbon or nitrogen; wherein each R<sub>1</sub> is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each R<sub>2</sub> is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each R is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen. Examples of  $\alpha$ 14 analogs include dMMS, dDMS, dFEMS, dBrMS, dIMS, and linker-derivatized analogs thereof.

[00107] Further provided herein, in various embodiments, are unnatural base pairs comprising any one  $\alpha$  analog or derivative thereof disclosed herein, and any one  $\beta$  analog or derivative thereof disclosed herein. Derivatives include, but are not limited to, atom substitutions and additions of linker moieties. Linker moieties may be attached to the analogs during synthesis or after nucleobase incorporation into a nucleic acid. Exemplary unnatural base pairs include, but are not limited to, dTPT3-dNaM, dTPT3-dFEMO, dTPT3-dFIMO, dTPT3-dIMO, dFTPT3-dNaM, dFTPT3-dFEMO, dFTPT3-dFIMO, dFTPT3-dIMO. Unnatural base pairs include, but are not limited to, dTPT3-MMS, dTPT3-DMS, dTPT3-FEMS, dTPT3-BrMS, dTPT3-IMS, dTPT3-dDMN, dTPT3-d4OMe, dTPT3-dIQ, dTPT3-d2MN, dTPT3-d3OMe, dTPT3-dQL,

dTPT3-d2Np, dTPT3-dDM4, dTPT3-dDM, dTPT3-dBEN, dTPT3-d3FB, dTPT3-dMM1, dTPT3-dMMO1, dTPT3-dDM2, dTPT3-dDM5, dTPT3-d2Py, dTPT3-d5MPy, dTPT3-dEPy, dTPT3-d3MPy, dTPT3-d34DMPy, dTPT3-d45DMPy, dTPT3-d4MPy, dTPT3-d35DMPy, dTPT3-dBP, dTPT3-dBTp, dTPT3-dBF, dTPT3-dIN, dTPT3-dTp, dTPT3-dBTz, dTPT3-dMTp, dTPT3-dAM, dTPT3-dMAN, dTPT3-dDMMAN, dTPT3-dADM, dTPT3-dMMAN, dTPT3-dTOK588, dTPT3-dTOK576, dTPT3-dTOK587, dTPT3-dTOK586, dTPT3-dTOK580, dTPT3-dPhMO, dTPT3-dPyMO1, dTPT3-PyMO2, dTPT3-dPMO1, dTPT3-dPMO2, dTPT3-dPMO3, dTPT3-dFuMO1, dTPT3-dFuMO2, dTPT3-TpMO1, dTPT3-dTpMO2, dTPT3-dFIMO, dTPT3-dIMO, dTPT3-dMIMO, dTPT3-dMEMO, dTPT3-dFEMO, dTPT3-dPrMO, dTPT3-dMMO2, dTPT3-d2OMe, dTPT3-dDMO, dTPT3-dTMO, dTPT3-dNMO, dTPT3-dNOPy, dTPT3-d5FM, dTPT3-dNAM, dTPT3-dAMO1, dTPT3-dAPy, dTPT3-dAMO2, dTPT3-dMAPy, dTPT3-dAMO3, dTPT3-dDMAPy, dTPT3-dFDMO, dTPT3-dVMO, dTPT3-dQMO, dTPT3-dZMO, dTPT3-dCIMO, dTPT3-dTfMO, dTPT3-CNMO, d7AI-dMMS, dM7AI-dMMS, dImPy-dMMS, dP7AI-dMMS, dPPP-dMMS, d8Q-dMMS, dICS-dMMS, dPICS-dMMS, dMICS-dMMS, d4MICS-dMMS, d5MICS-dMMS, dNICS-dMMS, dONICS-dMMS, d7OFP-dMMS, d7OTP-dMMS, d4OTP-dMMS, dPYR-dMMS, d4MP-dMMS, d3MP-dMMS, dPPYR-dMMS, dMOP-dMMS, d4MOP-dMMS, dSICS-dMMS, dSNICS-dMMS, d5SICS-dMMS, d4SICS-dMMS, dTPT1-dMMS, dTPT2-dMMS, dFPT1-dMMS, dFTPT3-dMMS, d7AI-dDMS, dM7AI-dDMS, dImPy-dDMS, dP7AI-dDMS, dPPP-dDMS, d8Q-dDMS, dICS-dDMS, dPICS-dDMS, dMICS-dDMS, d4MICS-dDMS, d5MICS-dDMS, dNICS-dDMS, dONICS-dDMS, d7OFP-dDMS, d7OTP-dDMS, d4OTP-dDMS, dPYR-dDMS, d4MP-dDMS, d3MP-dDMS, dPPYR-dDMS, dMOP-dDMS, d4MOP-dDMS, dSICS-dDMS, dSNICS-dDMS, d5SICS-dDMS, d4SICS-dDMS, dTPT1-dDMS, dTPT2-dDMS, dFPT1-dDMS, dFTPT3-dDMS, d7AI-dFEMS, dM7AI-dFEMS, dImPy-dFEMS, dP7AI-dFEMS, dPPP-dFEMS, d8Q-dFEMS, dICS-dFEMS, dPICS-dFEMS, dMICS-dFEMS, d4MICS-dFEMS, d5MICS-dFEMS, dNICS-dFEMS, dONICS-dFEMS, d7OFP-dFEMS, d7OTP-dFEMS, d4OTP-dFEMS, dPYR-dFEMS, d4MP-dFEMS, d3MP-dFEMS, dPPYR-dFEMS, dMOP-dFEMS, d4MOP-dFEMS, dSICS-dFEMS, dSNICS-dFEMS, d5SICS-dFEMS, d4SICS-dFEMS, dTPT1-dFEMS, dTPT2-dFEMS, dFPT1-dFEMS, dFTPT3-dFEMS, d7AI-dBrMS, dM7AI-dBrMS, dImPy-dBrMS, dP7AI-dBrMS, dPPP-dBrMS, d8Q-dBrMS, dICS-dBrMS, dPICS-dBrMS, dMICS-dBrMS, d4MICS-dBrMS, d5MICS-dBrMS, dNICS-dBrMS, dONICS-dBrMS, d7OFP-dBrMS, d7OTP-dBrMS, d4OTP-dBrMS, dPYR-dBrMS, d4MP-dBrMS, d3MP-dBrMS, dPPYR-dBrMS, dMOP-dBrMS, d4MOP-dBrMS, dSICS-dBrMS, dSNICS-dBrMS, d5SICS-dBrMS, d4SICS-dBrMS, dTPT1-dBrMS, dTPT2-dBrMS, dFPT1-dBrMS, dFTPT3-dBrMS, d7AI-dIMS, dM7AI-dIMS, dImPy-dIMS, dP7AI-dIMS, dPPP-dIMS, d8Q-dIMS, dICS-dIMS, dPICS-dIMS, dMICS-dIMS, d4MICS-dIMS,

d5MICS-dIMS, dNICS-dIMS, dONICS-dIMS, d7OFP-dIMS, d7OTP-dIMS, d4OTP-dIMS, dPYR-dIMS, d4MP-dIMS, d3MP-dIMS, dPPYR-dIMS, dMOP-dIMS, d4MOP-dIMS, dSICS-dIMS, dSNICS-dIMS, d5SICS-dIMS, d4SICS-dIMS, dTPT1-dIMS, dTPT2-dIMS, dFPT1-dIMS, and dFTPT3-dIMS; wherein one or two unnatural nucleobases of the unnatural base pair may be derivatized with a linker. Exemplary unnatural base pairs of this disclosure further include any pair described in Example 1. Exemplary  $\beta$  analogs include those which are presented in Figs. 9, 12, and 13. Exemplary  $\beta$  nucleobase analogs include  $\beta 8$  analogs comprising the formulas  $\beta 8a$  and  $\beta 8b$ , as shown in Figure 12, wherein each X is independently carbon or nitrogen; wherein each  $R_2$  is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen; and wherein R is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded. Examples of  $\beta$  analogs include dTPT3, d5SICS, dFTPT3 and derivatives or analogs thereof. Exemplary  $\alpha$  analogs include those which are presented in Figs. 8, 10, and 11. Exemplary  $\alpha$  analogs include  $\alpha 14$  analogs comprising the formulas  $\alpha 14a$ -  $\alpha 14f$ , as shown in Figure 10, wherein each X is independently carbon or nitrogen; wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded; wherein each  $R_2$  is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each R is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently sulfur, selenium or oxygen. Examples of  $\alpha$  analogs include dMMS, dDMS, dBrMS, dIMS, dFEMS, dNAM, dMMO2,



dDMO, dEMO, dFEMO, and derivatives or analogs thereof. In some embodiments, an unnatural base pair includes an  $\alpha$  analog and a natural base. In some embodiments, an unnatural base pair includes a  $\beta$  analog and a natural base. Further provided herein, in some aspects, are unnatural base pairs comprising the same two unnatural nucleoside analogs or derivatives thereof.

**[00108]** An unnatural base pair, in various aspects, comprises one unnatural nucleobase disclosed herein (e.g.  $\alpha$  analog or derivative thereof,  $\beta$  analog or derivative thereof) and another unnatural nucleobase including, but not limited to, 2-aminoadenin-9-yl, 2-aminoadenine, 2-F-adenine, 2-thiouracil, 2-thio-thymine, 2-thiocytosine, 2-propyl and alkyl derivatives of adenine and guanine, 2-amino-adenine, 2-amino-propyl-adenine, 2-aminopyridine, 2-pyridone, 2'-deoxyuridine, 2-amino-2'-deoxyadenosine 3-deazaguanine, 3-deazaadenine, 4-thio-uracil, 4-thio-thymine, uracil-5-yl, hypoxanthin-9-yl (I), 5-methyl-cytosine, 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 5-bromo, and 5-trifluoromethyl uracils and cytosines; 5-halouracil, 5-halocytosine, 5-propynyl-uracil, 5-propynyl cytosine, 5-uracil, 5-substituted, 5-halo, 5-substituted pyrimidines, 5-hydroxycytosine, 5-bromocytosine, 5-bromouracil, 5-chlorocytosine, chlorinated cytosine, cyclocytosine, cytosine arabinoside, 5-fluorocytosine, fluoropyrimidine, fluorouracil, 5,6-dihydrocytosine, 5-iodocytosine, hydroxyurea, iodouracil, 5-nitrocytosine, 5-bromouracil, 5-chlorouracil, 5-fluorouracil, and 5-iodouracil, 6-alkyl derivatives of adenine and guanine, 6-azapyrimidines, 6-azo-uracil, 6-azo cytosine, azacytosine, 6-azo-thymine, 6-thio-guanine, 7-methylguanine, 7-methyladenine, 7-deazaguanine, 7-deazaguanosine, 7-deaza-adenine, 7-deaza-8-azaguanine, 8-azaguanine, 8-azaadenine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted adenines and guanines; N4-ethylcytosine, N-2 substituted purines, N-6 substituted purines, O-6 substituted purines, those that increase the stability of duplex formation, universal nucleic acids, hydrophobic nucleic acids, promiscuous nucleic acids, size-expanded nucleic acids, fluorinated nucleic acids, tricyclic pyrimidines, phenoxazine cytidine( [5,4-b][1,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H- pyrimido[5,4-b][1,4]benzothiazin-2(3H)-one), G-clamps, phenoxazine cytidine (9-(2-aminoethoxy)-H-pyrimido[5,4-b][1,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5- b]indol-2-one), pyridoindole cytidine (H-pyrido [3',2':4,5]pyrrolo [2,3-d]pyrimidin-2-one), 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-

methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxoacetic acid, wybutosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid methylester, uracil-5-oxoacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)<sub>w</sub>, and 2,6-diaminopurine and those in which the purine or pyrimidine base is replaced with a heterocycle. The  $\alpha$  analogs of the unnatural base pair include, without limitation, dMMS, dDMS, dBrMS, dIMS, dFEMS, dNAM, dMMO<sub>2</sub>, dDMO, dEMO, dFEMO, and derivatives or analogs thereof. The  $\beta$  analogs of the unnatural base pair include, without limitation, dTPT3, d5SICS, and dFTPT3.

[00109] In some embodiments, the unnatural nucleobases and unnatural base pairs disclosed herein have efficient incorporation and extension with natural polymerases. In some embodiments, the unnatural nucleobases and unnatural base pairs disclosed herein have efficient incorporation and extension with modified polymerases. The effect of an unnatural nucleobase or unnatural nucleobase derivative on polymerase recognition is assessed, in exemplary embodiments, by determining the steady-state efficiency (e.g., second order rate constant  $k_{\text{cat}}/K_M$ ) with which the polymerase synthesizes an unnatural base pair, by insertion of the unnatural nucleotide opposite its complementary base in a template, and extends the resulting unnatural primer terminus, by insertion of the next correct natural nucleotide. Corresponding rates of synthesis and extension for mispairs with natural nucleotides may also be measured to determine fidelity. In some embodiments, polymerases do not need to be modified to improve incorporation or extension rates. The embodiments and examples disclosed herein may be performed with any known polymerase. Polymerases include naturally-occurring polymerases and any modified variations thereof, including, but not limited to, mutants, recombinants, fusions, genetic modifications, chemical modifications, synthetics, and analogs. Naturally-occurring polymerases and modified variations thereof are not limited to polymerases which retain the ability to catalyze a polymerization reaction. In some instances, the naturally-occurring and/or modified variations thereof retain the ability to catalyze a polymerization reaction. Mutant polymerases include polymerases wherein one or more amino acids are replaced with other amino acids (naturally or non-naturally occurring), and polymerases having one or more amino acid insertions or deletions. In some embodiments a polymerase refers to fusion proteins comprising at least two portions linked to each other, for example, where one portion comprises a peptide that can catalyze the polymerization of nucleotides into a nucleic acid strand is linked to another portion that comprises a second moiety, such as, a reporter enzyme or a processivity-modifying domain. One exemplary embodiment of such a polymerase is T7 DNA polymerase, which comprises a nucleic acid polymerizing domain and a thioredoxin

binding domain, wherein thioredoxin binding enhances the processivity of the polymerase. Absent the thioredoxin binding, T7 DNA polymerase is a distributive polymerase with processivity of only one to a few bases. DNA polymerases include, but are not limited to, bacterial DNA polymerases, eukaryotic DNA polymerases, archaeal DNA polymerases, viral DNA polymerases and phage DNA polymerases. Bacterial DNA polymerases include *E. coli* DNA polymerases I, II and III, IV and V, the Klenow fragment of *E. coli* DNA polymerase, *Clostridium stercoarium* (Cst) DNA polymerase, *Clostridium thermocellum* (Cth) DNA polymerase and *Sulfolobus solfataricus* (Sso) DNA polymerase. Eukaryotic DNA polymerases include DNA polymerases  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\eta$ ,  $\zeta$ ,  $\sigma$ ,  $\lambda$ ,  $\mu$ ,  $\iota$ , and  $\kappa$ , as well as the Rev1 polymerase (terminal deoxycytidyl transferase) and terminal deoxynucleotidyl transferase (TdT). Viral DNA polymerases include T4 DNA polymerase, phi-29 DNA polymerase, GA-1, phi-29-like DNA polymerases, PZA DNA polymerase, phi-15 DNA polymerase, Cpl DNA polymerase, Cp7 DNA polymerase, T7 DNA polymerase, and T4 polymerase. Archaeal DNA polymerases include thermostable and/or thermophilic DNA polymerases such as DNA polymerases isolated from *Thermus aquaticus* (Tag) DNA polymerase, *Thermus filiformis* (Tfi) DNA polymerase, *Thermococcus zilligi* (Tzi) DNA polymerase, *Thermus thermophilus* (Tth) DNA polymerase, *Thermus flavus* (Tfl) DNA polymerase, *Pyrococcus woesei* (Pwo) DNA polymerase, *Pyrococcus furiosus* (Pfu) DNA polymerase and Turbo Pfu DNA polymerase, *Thermococcus litoralis* (Tli) DNA polymerase, *Pyrococcus* sp. GB-D polymerase, *Thermotoga maritima* (Tma) DNA polymerase, *Bacillus stearothermophilus* (Bst) DNA polymerase, *Pyrococcus kodakaraensis* (KOD) DNA polymerase, Pfx DNA polymerase, *Thermococcus* sp. JDF-3 (JDF-3) DNA polymerase, *Thermococcus gorgonarius* (Tgo) DNA polymerase, *Thermococcus acidophilum* DNA polymerase; *Sulfolobus acidocaldarius* DNA polymerase; *Thermococcus* sp. 9° N-7 DNA polymerase; *Pyrodictium occultum* DNA polymerase; *Methanococcus voltae* DNA polymerase; *Methanococcus thermoautotrophicum* DNA polymerase; *Methanococcus jannaschii* DNA polymerase; *Desulfurococcus* strain TOK DNA polymerase (D. Tok Pol); *Pyrococcus abyssi* DNA polymerase; *Pyrococcus horikoshii* DNA polymerase; *Pyrococcus islandicum* DNA polymerase; *Thermococcus fumicolans* DNA polymerase; *Aeropyrum pernix* DNA polymerase; and the heterodimeric DNA polymerase DP1/DP2. RNA polymerases include, but are not limited to, viral RNA polymerases such as T7 RNA polymerase, T3 polymerase, SP6 polymerase, and K11 polymerase; Eukaryotic RNA polymerases such as RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, and RNA polymerase V; and Archaea RNA polymerase.

[00110] In some embodiments, a polymerase has a specificity for an unnatural nucleotide comprising an  $\alpha$  or  $\beta$  nucleobase analog that is at least about 10%, 20%, 30%, 40%, 50%, 60%,

70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% of the specificity of the polymerase toward a natural nucleotide. In some embodiments, a polymerase has a specificity for an unnatural nucleotide comprising an  $\alpha$  or  $\beta$  nucleobase analog and a modified sugar that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% of the specificity of the polymerase toward a natural nucleotide and/or the unnatural nucleotide without the modified sugar. In some embodiments, a polymerase has a specificity for an unnatural nucleotide comprising a linker-derivatized  $\alpha$  or  $\beta$  nucleobase analog that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the polymerase toward a natural nucleotide and/or the unnatural nucleotide without the linker. In some embodiments, the unnatural nucleobase is dTPT3. In some embodiments, the unnatural nucleobase is dMMS. In some embodiments, the unnatural nucleobase is dDMS. In some embodiments, the unnatural nucleobase is dBrMS. In some embodiments, the unnatural nucleobase is IMS. In some embodiments, the unnatural nucleobase is dFEMS. In some embodiments, the unnatural nucleobase is  $\text{MMS}^{\text{PCO}}$ . In some embodiments, the unnatural nucleobase is  $\text{dMMS}^{\text{PA}}$ . In some embodiments, the unnatural nucleobase is dFTPT3. In some embodiments, the unnatural nucleobase is  $\text{dTPT}^{\text{PA}}$ . In some embodiments, the unnatural nucleobase is  $\text{dTPT3}^{\text{CO}}$ . In some embodiments, the unnatural nucleobase comprises the formula  $\alpha 14a$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\alpha 14b$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\alpha 14c$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\alpha 14d$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\alpha 14e$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\alpha 14f$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\beta 8a$  or a derivative or analog thereof. In some embodiments, the unnatural nucleobase comprises the formula  $\beta 8b$  or a derivative or analog thereof.

**[00111]** Polymerases can be characterized according to their fidelity when used with a particular natural and/or unnatural nucleotide or collections of natural and/or unnatural nucleotides, wherein the unnatural nucleotide comprises an  $\alpha$  or  $\beta$  nucleobase analog disclosed herein. In various embodiments, fidelity generally refers to the accuracy with which a polymerase incorporates correct nucleotides into a growing oligonucleotide when making a copy of an oligonucleotide template. Polymerase fidelity can be measured as the ratio of correct to incorrect natural and unnatural nucleotide incorporations when the natural and unnatural nucleotides are present, e.g., at equal concentrations, to compete for strand synthesis at the same

site in the polymerase-strand-template nucleic acid binary complex. DNA polymerase fidelity can be calculated as the ratio of  $(k_{\text{cat}}/K_M)$  for the natural and unnatural nucleotide and  $(k_{\text{cat}}/K_M)$  for the incorrect natural and unnatural nucleotide; where  $k_{\text{cat}}$  and  $K_M$  are Michaelis-Menten parameters in steady state enzyme kinetics. In some embodiments, a polymerase has a fidelity value of at least about 100, 1000, 10,000, 100,000, or  $1 \times 10^6$ , with or without a proofreading activity. In some embodiments, a polymerase has a fidelity value of at least about 100, 1000, 10,000, 100,000, or  $1 \times 10^6$  for unnatural nucleotide incorporation. In some embodiments, the unnatural nucleotide is dTPT3TP or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide is dNAM or a derivative thereof. In some embodiments, the unnatural nucleotide is dNaMTP or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide is dTPT3 or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8a$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14a$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8a$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14b$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8a$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14c$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8a$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14d$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8a$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14e$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8a$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14f$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8b$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14a$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8b$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14b$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8b$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14c$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8b$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14d$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8e$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14e$  or a derivative thereof. In some embodiments, the unnatural nucleotide comprises  $\beta 8f$  or a derivative thereof, and its corresponding nucleobase on the template oligonucleotide comprises  $\alpha 14f$  or a derivative thereof.

[00112] The unnatural base pairs exemplified herein, in some embodiments, are synthesized/amplified with natural base pair-like efficiency and fidelity. Unnatural base pairs comprise, in various embodiments, any  $\alpha$  nucleobase analog or derivative thereof, and/or any  $\beta$  nucleobase analog or derivative thereof. Examples of  $\beta$  analogs include dTPT3, d5SICS, dFTPT3 and derivatives or analogs thereof. Examples of  $\alpha$  analogs include dMMS, dDMS, dBrMS, dIMS, dFEMS, dNAM, dMMO2, dDMO, dEMO, dFEMO, and derivatives or analogs thereof. In some embodiments, an unnatural base pair is efficiently amplified in a variety of different sequence contexts, including GC- and AT-rich sequences, randomized sequences, and sequences comprising multiple unnatural nucleobase pairs, with greater than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.85, 99.9 or higher fidelity per doubling. For example, an unnatural nucleobase pair comprising one or more unnatural nucleobases has a synthesis efficiency and/or fidelity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% similar to an amplification efficiency and/or fidelity of a natural base pair. As another example, an unnatural nucleobase pair comprising one or more unnatural nucleobases has a synthesis efficiency and/or fidelity that is at most 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or 0.5% less efficient and/or accurate than that of a natural base pair. In some embodiments, an unnatural nucleobase pair is transcribed with good efficiency and selectivity in both strand contexts (e.g., dX must template YTP insertion and dY must template XTP insertion). In some embodiments, relative to the rate at which a fully natural sequence is transcribed, the incorporation of an unnatural nucleotide does not reduce the rate of full-length transcription. In some embodiments, relative to the rate at which a fully natural sequence is transcribed, the incorporation of an unnatural nucleotide reduces the rate of full-length transcription by a factor less than 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, or 40. In some embodiments, the unnatural base pair comprises dTPT3 or a derivative or analog thereof, and dNaM or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises dTPT3 or a derivative or analog thereof, and dNaM or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises dTPT3 or a derivative or analog thereof, and dNaM or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises dTPT3 or a derivative or analog thereof, and dNaM or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises dTPT3 or a derivative or analog thereof, and dNaM or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8a or a derivative or analog thereof, and  $\alpha$ 14a or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8a or a derivative or analog thereof, and  $\alpha$ 14b or a derivative or analog thereof. In

some embodiments, the unnatural base pair comprises  $\beta$ 8a or a derivative or analog thereof, and  $\alpha$ 14c or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8a or a derivative or analog thereof, and  $\alpha$ 14d or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8a or a derivative or analog thereof, and  $\alpha$ 14e or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8a or a derivative or analog thereof, and  $\alpha$ 14f or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8b or a derivative or analog thereof, and  $\alpha$ 14a or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8b or a derivative or analog thereof, and  $\alpha$ 14b or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8b or a derivative or analog thereof, and  $\alpha$ 14c or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8b or a derivative or analog thereof, and  $\alpha$ 14d or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8b or a derivative or analog thereof, and  $\alpha$ 14e or a derivative or analog thereof. In some embodiments, the unnatural base pair comprises  $\beta$ 8b or a derivative or analog thereof, and  $\alpha$ 14f or a derivative or analog thereof.

[00113] Further provided herein, in various embodiments, are unnatural base pairs comprising one or more unnatural nucleobases (e.g.  $\alpha$  nucleobase,  $\beta$  nucleobase, or  $\alpha$  nucleobase and  $\beta$  nucleobase), wherein one or two nucleobases comprise a linker. A linker comprises a reactive center. Exemplary reactive centers include, but are not limited to, alkyl, alkenyl, alkynyl, phenyl, benzyl, halo, hydroxyl, carbonyl, aldehyde, haloformyl, carbonate ester, carboxylate, carboxyl, ester, methoxy, hydroperoxy, peroxy, ether, hemiacetal, hemiketal, acetal, ketal, orthoester, methylenedioxy, orthocarbonate ester, carboxamide, primary amine, secondary amine, imide, azide, azo, cyanate, isocyanate, nitrate, nitrile, isonitrile, nitrosooxy, nitro, nitroso, pyridyl, sulfhydryl, sulfide, disulfide, sulfinyl, sulfo, thiocyanate, isothiocyanate, carbonothioyl, phosphino, phosphono, phosphate, borono, boronate, borino, borinate, and a combination thereof. An example of a linker-derivatized nucleobase is TPT3<sup>R</sup> shown in Figure 2, wherein the superscript R indicates the linker. In some embodiments, a linker is modified with a protecting group, for example, TPT3<sup>PA</sup>, where the linker is a protected propargyl linker.

[00114] In some embodiments, a nucleobase analog provided herein comprises an amino-functional linker or a protected amino-functional linker (e.g., dX<sup>PA</sup>). In certain embodiments, the amino-functional linker is 3-aminopropyn-1-yl. In some embodiments, a nucleobase analog provided herein comprises an alkyne-azide ether linker for derivatization via click chemistry or a protected alkyne-azide ether linker for derivatization via click chemistry. In certain embodiments, the alkyne-azide ether linker is 4-oxahept-1,6-diyn-1-yl. In some embodiments, a nucleobase analog provided herein comprises an alkyne-azide trimethylene linker for

derivatization via click chemistry or a protected alkyne-azide trimethylene linker for derivatization via click chemistry. In certain embodiments, the alkyne-azide trimethylene linker is hepta-1,6-diyn-1-yl. In some embodiments, X is a  $\beta$  nucleoside analog having any of the formulas from Figures 9, 12, 13, and 16. In some embodiments, X is ICS, PICS, MICS, 4MICS, 5MICS, NICS, ONICS, SICS, SNICS, 5SICS, 4SICS, 7OFP, 7OTP, TPT2, TPT3, or FTPT3. In some embodiments, X is an  $\alpha$  nucleoside analog having any of the formulas from Figures 8, 10, 11, and 15. In some embodiments, X is FIMO, MIMO, FEMO, PrMO, EMO, MEMO, IMO, MMO2, DMO, NMO, 5FM, 2OMe, TMO, FDMO, VMO, ZMO, CIMO, TfMO, CNMO, MMS, DMS, BrMS, IMS, FEMS, NAM, or QMO.

[00115] In some embodiments, a linker is a propinyl linker, such as those used with natural nucleotides. These linkers comprise propargyl amines, with the amine serving as a reactive site to attach other functionalities.

[00116] In various embodiments, a linker-derivatized nucleobase comprises a spacer. An exemplary spacer is acetamidohexanamide. A spacer may be hydrophilic. A spacer may connect a linker to a functional group. Spacers include, but are not limited to, Spacer C3 (3-carbon spacer), Spacer C6 (6-carbon spacer), photo-cleavable spacer, hexanediol spacer, Spacer 9 (triethylene glycol spacer), Spacer C12 (12-carbon spacer), Spacer 18 (18-atom hexa-ethyleneglycol spacer), and 1',2'-Dideoxyribose spacer.

[00117] An unnatural nucleobase pair comprising one or two linker-derivatized nucleobases, in some instances, is amplified with an efficiency and fidelity that is similar to that of a natural base pair or a non-linker derivatized unnatural base pair. For example, an unnatural nucleobase pair comprising one or two linker-derivatized unnatural nucleobases has a synthesis efficiency and/or fidelity that has at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% similar to a synthesis efficiency and/or fidelity of a natural base pair or a non-linker derivatized unnatural base pair. As another example, an unnatural nucleobase pair comprising one or two linker-derivatized unnatural nucleobases has a synthesis efficiency and/or fidelity that is at most 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or 0.5% less efficient and/or accurate than that of a natural base pair or a non-linker derivatized unnatural base pair. In some embodiments, an unnatural nucleobase pair comprises dTPT3<sup>PA</sup>. In some embodiments, an unnatural nucleobase pair comprises dTPT3<sup>CO</sup>. In some embodiments, an unnatural nucleobase pair comprises dMMS<sup>CO</sup>. In some embodiments, an unnatural nucleobase pair comprises dMMS<sup>PA</sup>. In some embodiments, an unnatural nucleobase pair comprises dNaM<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dMMO2<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dDMO<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises



d5SICS<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dMMS<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dDMS<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dFEMS<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dBrMS<sup>R</sup>. In some embodiments, an unnatural nucleobase pair comprises dIMS<sup>R</sup>.

**[00118]** In some embodiments, a linker-derivatized unnatural nucleobase has an increased insertion efficiency during oligonucleotide synthesis, as compared to the same unnatural nucleobase which does not comprise a linker. In some embodiments, a linker-derivatized unnatural nucleobase has a decreased insertion efficiency during oligonucleotide synthesis, as compared to the same unnatural nucleobase which does not comprise a linker. In some instances, a linker-derivatized unnatural nucleobase has about the same insertion efficiency during oligonucleotide synthesis, as compared to the same unnatural nucleobase which does not comprise a linker. In some embodiments, a protected linker-derivatized unnatural nucleobase has an increased insertion efficiency during oligonucleotide synthesis, as compared to the same unnatural nucleobase which does not comprise a protected linker. In some embodiments, a protected linker-derivatized unnatural nucleobase has a decreased insertion efficiency during oligonucleotide synthesis, as compared to the same unnatural nucleobase which does not comprise a protected linker. In some instances, a protected linker-derivatized unnatural nucleobase has about the same efficiency during oligonucleotide synthesis, as compared to the same unnatural nucleobase which does not comprise a protected linker.

**[00119]** Exemplary methods for analyzing unnatural base pair synthesis efficiency (insertion of an unnatural nucleobase opposite its partner in a template) and extension (continued primer elongation) are provided herein. One or both of the nucleobases in an unnatural base pair, in various embodiments, may be a linker-derivatized unnatural nucleobase. One method uses a presteady-state assay. The assay is based on determining, under a fixed set of conditions, the amount of a primer (e.g. 23-mer) that is extended by addition of the unnatural nucleotide opposite its complementary nucleotide in a template (e.g., 45-mer) by a polymerase (e.g., the Klenow fragment of *E. coli* DNA polymerase I). In this assay, the efficiency of unnatural base pair synthesis is characterized by measuring the percent incorporation (%inc) at a given concentration of the unnatural and next correct triphosphate, for example using a ratio such as [24-mer + 25-mer]/[23-mer + 24-mer + 25-mer]. In this assay, the efficiency of extension is characterized by measuring the percent extension (%ext) at a given concentration of the next correct nucleotide and saturating concentrations of unnatural nucleotide, for example using a ratio [25-mer]/[24-mer + 25-mer]. Results from an exemplary presteady-state assay are shown in Table 1, wherein the unnatural triphosphate is 5SICS, FPT1, TPT1, TPT2, TPT3, FTPT3, TPT3<sup>PA</sup>, or 5SICS<sup>PA</sup>. In some embodiments, the percent incorporation of an unnatural

nucleobase is at least 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. In some embodiments, the percent extension of a next correct nucleotide following insertion of an unnatural nucleobase is at least 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or greater. In some embodiments, synthesis efficiency is increased by derivatizing the unnatural nucleobase. For example, by the addition of a linker, a protected linker, and/or a linker conjugated to a cargo molecule. As another example, derivatization includes atom substitutions, additions, or deletions. In some embodiments, percent extension is increased by derivatizing the unnatural nucleobase. Derivatization of an unnatural nucleobase, in some instances, increases by at least 1-2 orders of magnitude the efficiency of insertion of the nucleotide complementary to the base pair in the template. This increase in efficiency may be due to an increase  $k_{cat}$  and a decreased  $K_M$ .

Table 1. Presteady-state kinetics.

dXTP	% Incorporation <sup>a</sup>	% Extension <sup>b</sup>
5SICS	57.0 ± 0.2	15.1 ± 1.1
FPT1	7.2 ± 0.2	32.0 ± 1.5
TPT1	28.7 ± 0.5	8.8 ± 0.2
TPT2	65.7 ± 0.5	34.5 ± 0.5
TPT3	72.3 ± 0.5	49.8 ± 1.3
FTPT3	66.3 ± 0.5	33.8 ± 0.2
TPT3 <sup>PA</sup>	68.3 ± 0.4	31.5 ± 0.7
5SICS <sup>PA</sup>	7.0 ± 0.2	5.5 ± 0.1

<sup>a</sup>Incorporation assay conditions: 40 nM unnatural triphosphate, 2 μM dCTP, 10s. <sup>b</sup>Extension assay conditions: 10 μM unnatural triphosphate, 2 μM dCTP, 10s. dXTPs are paired with dNaM.

[00120] Further provided herein are replication evaluation methods. In one method, a template nucleic acid duplex comprising an unnatural base pair (e.g., dTPT3-dNaM or analogs thereof), is amplified by PCR. In one example, a set of PCR reactions employs 48 cycles with OneTaq polymerase. In another example, a set of PCR reactions employs 20 cycles of amplification with exonuclease-negative Taq. Efficiency is determined by monitoring the amplification level. Fidelity, generally defined as unnatural base pair extension per doubling, is determined from the percentage of the amplified DNA that retains the unnatural base pair. The percentage of amplified DNA that retains the unnatural base pair may be determined from the relative peak intensities of a sequencing chromatogram. In some embodiments, the fidelity of unnatural base pair replication is at least 98%, 98.1%, 98.2%, 98.3%, 98.4%, 98.5%, 98.6%, 98.7%, 98.8%,

98.9%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, 99.91%, 99.92%, 99.93%, 99.94%, 99.95%, 99.96%, 99.97%, 99.98% or 99.99%. Replication of an unnatural base pair may proceed with little or no sequence bias, wherein little sequence bias indicates that an unfavorable sequence decreases fidelity by less than 1%. Exemplary fidelities are described in Example 1 and shown in Tables 4, 5, and 6.

[00121] Further provided herein, in various embodiments, are oligonucleotides, including single-stranded and double-stranded (e.g., duplex) DNA and/or RNA, comprising one or more unnatural nucleobases described herein (e.g., any  $\alpha$  nucleobase or analog or derivative thereof and/or any  $\beta$  nucleobase or analog or derivative thereof). The nucleobase may be any  $\alpha$  nucleobase or  $\beta$  nucleobase described herein, including those in Figures 2, 8, 9, 10, 11, 12, 13, 15, and 16. A double-stranded oligonucleotide includes a DNA-DNA duplex, DNA-RNA hybrid duplex, and RNA-RNA duplex. In some embodiments, the oligonucleotide comprises a linker-derivatized nucleobase.

[00122] In some embodiments, an oligonucleotide comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50 or more unnatural nucleobases. In some embodiments, the percentage of unnatural nucleobases in an oligonucleotide is between about 0% and about 1%, between about 0% and about 2%, between about 0% and about 3%, between about 0% and about 4%, between about 0% and about 5%, between about 0% and about 10%, between about 1% and about 10%, between about 1% and about 15%, between about 1% and about 20%, between about 5% and about 10%, between about 5% and about 20%, between about 10% and about 30%, between about 1% and about 50%, or between about 1% and about 100%.

[00123] Examples of oligonucleotides comprising one or more unnatural nucleobases include, but are not limited to, DNA aptamers and RNA aptamers. DNA and RNA aptamers include, but are not limited to, primers and molecular beacons. A DNA aptamer may include a barcode.

[00124] In some embodiments, an oligonucleotide comprises dTPT3 or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises d5SICS or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises dNaM or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises dMMS or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises dDMS or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises dFEMS or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises dBrMS or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises dIMS or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\beta$ 8a or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\beta$ 8b or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\alpha$ 14a or a derivative or analog thereof.

thereof. In some embodiments, an oligonucleotide comprises  $\alpha$ 14b or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\alpha$ 14c or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\alpha$ 14d or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\alpha$ 14e or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises  $\alpha$ 14f or a derivative or analog thereof.

[00125] In some embodiments, an oligonucleotide comprises dTPT3 or a derivative or analog thereof, and dNaM or a derivative or analog thereof. In some embodiments, an oligonucleotide comprises a dTPT3-dNaM base pair. In some embodiments, an oligonucleotide comprises one or more base pairs selected from dTPT3-dFEMO, dTPT3-dFIMO, dTPT3-dIMO, dFTPT3-dNaM, dFTPT3-dFEMO, dFTPT3-dFIMO, and dFTPT3-dIMO. In some embodiments, an oligonucleotide comprises one or more base pairs selected from dTPT3-MMS, dTPT3-DMS, dTPT3-FEMS, dTPT3-BrMS, dTPT3-IMS, dTPT3-dDMN, dTPT3-d4OMe, dTPT3-dIQ, dTPT3-d2MN, dTPT3-d3OMe, dTPT3-dQL, dTPT3-d2Np, dTPT3-dDM4, dTPT3-dDM, dTPT3-dBEN, dTPT3-d3FB, dTPT3-dMM1, dTPT3-dMMO1, dTPT3-dDM2, dTPT3-dDM5, dTPT3-d2Py, dTPT3-d5MPy, dTPT3-dEPy, dTPT3-d3MPy, dTPT3-d34DMPy, dTPT3-d45DMPy, dTPT3-d4MPy, dTPT3-d35DMPy, dTPT3-dBP, dTPT3-dBTp, dTPT3-dBF, dTPT3-dIN, dTPT3-dTp, dTPT3-dBTz, dTPT3-dMTp, dTPT3-dAM, dTPT3-dMAN, dTPT3-dDMAN, dTPT3-dADM, dTPT3-dMMAN, dTPT3-dTOK588, dTPT3-dTOK576, dTPT3-dTOK587, dTPT3-dTOK586, dTPT3-dTOK580, dTPT3-dPhMO, dTPT3-dPyMO1, dTPT3-PyMO2, dTPT3-dPMO1, dTPT3-dPMO2, dTPT3-dPMO3, dTPT3-dFuMO1, dTPT3-dFuMO2, dTPT3-TpMO1, dTPT3-dTpMO2, dTPT3-dFIMO, dTPT3-dIMO, dTPT3-dMIMO, dTPT3-dMEMO, dTPT3-dFEMO, dTPT3-dPrMO, dTPT3-dMMO2, dTPT3-d2OMe, dTPT3-dDMO, dTPT3-dTMO, dTPT3-dNMO, dTPT3-dNOPy, dTPT3-d5FM, dTPT3-dNAM, dTPT3-dAMO1, dTPT3-dAPy, dTPT3-dAMO2, dTPT3-dMAPy, dTPT3-dAMO3, dTPT3-dDMAPy, dTPT3-dFDMO, dTPT3-dVMO, dTPT3-dQMO, dTPT3-dZMO, dTPT3-dCIMO, dTPT3-dTfMO, dTPT3-CNMO, d7AI-dMMS, dM7AI-dMMS, dImPy-dMMS, dP7AI-dMMS, dPPP-dMMS, d8Q-dMMS, dICS-dMMS, dPICS-dMMS, dMICS-dMMS, d4MICS-dMMS, d5MICS-dMMS, dNICS-dMMS, dONICS-dMMS, d7OFP-dMMS, d7OTP-dMMS, d4OTP-dMMS, dPYR-dMMS, d4MP-dMMS, d3MP-dMMS, dPPYR-dMMS, dMOP-dMMS, d4MOP-dMMS, dSICS-dMMS, dSNICS-dMMS, d5SICS-dMMS, d4SICS-dMMS, dTPT1-dMMS, dTPT2-dMMS, dFPT1-dMMS, dFTPT3-dMMS, d7AI-dDMS, dM7AI-dDMS, dImPy-dDMS, dP7AI-dDMS, dPPP-dDMS, d8Q-dDMS, dICS-dDMS, dPICS-dDMS, dMICS-dDMS, d4MICS-dDMS, d5MICS-dDMS, dNICS-dDMS, dONICS-dDMS, d7OFP-dDMS, d7OTP-dDMS, d4OTP-dDMS, dPYR-dDMS, d4MP-dDMS, d3MP-dDMS, dPPYR-dDMS, dMOP-dDMS, d4MOP-

dDMS, dSICS-dDMS, dSNICS-dDMS, d5SICS-dDMS, d4SICS-dDMS, dTPT1-dDMS, dTPT2-dDMS, dFPT1-dDMS, dFTPT3-dDMS, d7AI-dFEMS, dM7AI-dFEMS, dImPy-dFEMS, dP7AI-dFEMS, dPPP-dFEMS, d8Q-dFEMS, dICS-dFEMS, dPICS-dFEMS, dMICS-dFEMS, d4MICS-dFEMS, d5MICS-dFEMS, dNICS-dFEMS, dONICS-dFEMS, d7OFP-dFEMS, d7OTP-dFEMS, d4OTP-dFEMS, dPYR-dFEMS, d4MP-dFEMS, d3MP-dFEMS, dPPYR-dFEMS, dMOP-dFEMS, d4MOP-dFEMS, dSICS-dFEMS, dSNICS-dFEMS, d5SICS-dFEMS, d4SICS-dFEMS, dTPT1-dFEMS, dTPT2-dFEMS, dFPT1-dFEMS, dFTPT3-dFEMS, d7AI-dBrMS, dM7AI-dBrMS, dImPy-dBrMS, dP7AI-dBrMS, dPPP-dBrMS, d8Q-dBrMS, dICS-dBrMS, dPICS-dBrMS, dMICS-dBrMS, d4MICS-dBrMS, d5MICS-dBrMS, dNICS-dBrMS, dONICS-dBrMS, d7OFP-dBrMS, d7OTP-dBrMS, d4OTP-dBrMS, dPYR-dBrMS, d4MP-dBrMS, d3MP-dBrMS, dPPYR-dBrMS, dMOP-dBrMS, d4MOP-dBrMS, dSICS-dBrMS, dSNICS-dBrMS, d5SICS-dBrMS, d4SICS-dBrMS, dTPT1-dBrMS, dTPT2-dBrMS, dFPT1-dBrMS, dFTPT3-dBrMS, d7AI-dIMS, dM7AI-dIMS, dImPy-dIMS, dP7AI-dIMS, dPPP-dIMS, d8Q-dIMS, dICS-dIMS, dPICS-dIMS, dMICS-dIMS, d4MICS-dIMS, d5MICS-dIMS, dNICS-dIMS, dONICS-dIMS, d7OFP-dIMS, d7OTP-dIMS, d4OTP-dIMS, dPYR-dIMS, d4MP-dIMS, d3MP-dIMS, dPPYR-dIMS, dMOP-dIMS, d4MOP-dIMS, dSICS-dIMS, dSNICS-dIMS, d5SICS-dIMS, d4SICS-dIMS, dTPT1-dIMS, dTPT2-dIMS, dFPT1-dIMS, dFTPT3-dIMS; wherein one or two unnatural nucleobases of the unnatural base pair may be derivatized with a linker.

**[00126]** An oligonucleotide comprising an unnatural nucleobase disclosed herein may further comprise one or more additional unnatural bases, including, but not limited to, 2-aminoadenin-9-yl, 2-aminoadenine, 2-F-adenine, 2-thiouracil, 2-thio-thymine, 2-thiocytosine, 2-propyl and alkyl derivatives of adenine and guanine, 2-amino-adenine, 2-amino-propyl-adenine, 2-aminopyridine, 2-pyridone, 2'-deoxyuridine, 2-amino-2'-deoxyadenosine 3-deazaguanine, 3-deazaadenine, 4-thio-uracil, 4-thio-thymine, uracil-5-yl, hypoxanthin-9-yl (I), 5-methylcytosine, 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 5-bromo, and 5-trifluoromethyl uracils and cytosines; 5-halouracil, 5-halocytosine, 5-propynyl-uracil, 5-propynyl cytosine, 5-uracil, 5-substituted, 5-halo, 5-substituted pyrimidines, 5-hydroxycytosine, 5-bromocytosine, 5-bromouracil, 5-chlorocytosine, chlorinated cytosine, cyclocytosine, cytosine arabinoside, 5-fluorocytosine, fluoropyrimidine, fluorouracil, 5,6-dihydrocytosine, 5-iodocytosine, hydroxyurea, iodouracil, 5-nitrocytosine, 5-bromouracil, 5-chlorouracil, 5-fluorouracil, and 5-iodouracil, 6-alkyl derivatives of adenine and guanine, 6-azapyrimidines, 6-azo-uracil, 6-azocytosine, azacytosine, 6-azo-thymine, 6-thio-guanine, 7-methylguanine, 7-methyladenine, 7-deazaguanine, 7-deazaguanosine, 7-deaza-adenine, 7-deaza-8-azaguanine, 8-azaguanine, 8-azaadenine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted adenines and guanines; N4-ethylcytosine, N-2 substituted purines, N-6 substituted purines, O-6 substituted

purines, those that increase the stability of duplex formation, universal nucleic acids, hydrophobic nucleic acids, promiscuous nucleic acids, size-expanded nucleic acids, fluorinated nucleic acids, tricyclic pyrimidines, phenoxazine cytidine( [5,4-b][1,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H- pyrimido[5,4-b][1,4]benzothiazin-2(3H)-one), G-clamps, phenoxazine cytidine (9-(2-aminoethoxy)-H-pyrimido[5,4-b][1,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5- b]indol-2-one), pyridoindole cytidine (H-pyrido [3',2':4,5]pyrrolo [2,3-d]pyrimidin-2-one), 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylecytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxoacetic acid, wybutosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxoacetic acid methylester, uracil-5-oxoacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine and those in which the purine or pyrimidine base is replaced with a heterocycle.

**[00127]** An oligonucleotide comprising an unnatural nucleobase disclosed herein, may further comprise an unnatural sugar moiety, including, but not limited to, a modification at the 2' position: OH; substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH<sub>3</sub>, OCN, Cl, Br, CN, CF<sub>3</sub>, OCF<sub>3</sub>, SOCH<sub>3</sub>, SO<sub>2</sub> CH<sub>3</sub>, ONO<sub>2</sub>, NO<sub>2</sub>, N<sub>3</sub>, NH<sub>2</sub>F; O-alkyl, S-alkyl, N-alkyl; O-alkenyl, S- alkenyl, N-alkenyl; O-alkynyl, S-alkynyl, N-alkynyl; O-alkyl-O-alkyl, 2'-F, 2'-OCH<sub>3</sub>, 2'-O(CH<sub>2</sub>)<sub>2</sub>OCH<sub>3</sub> wherein the alkyl, alkenyl and alkynyl may be substituted or substituted C1-C10, alkyl, C2-C10 alkenyl, C2-C10 alkynyl, -O[(CH<sub>2</sub>)<sub>n</sub> O]<sub>m</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>OCH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub> NH<sub>2</sub>, -O(CH<sub>2</sub>)<sub>n</sub> CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>-ONH<sub>2</sub>, and -O(CH<sub>2</sub>)<sub>n</sub>ON[(CH<sub>2</sub>)<sub>n</sub> CH<sub>3</sub>)]<sub>2</sub>, where n and m are from 1 to about 10; and/or a modification at the 5' position: 5'-vinyl, 5'-methyl (R or S), a modification at the 4' position, 4'-S, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and any combination thereof.

**[00128]** In some embodiments, the oligonucleotide comprising an unnatural nucleobase disclosed herein, further comprises an unnatural backbone. An unnatural backbone includes, but is not limited to, phosphorothioate, chiral phosphorothioate, phosphorodithioate,

phosphotriester, aminoalkylphosphotriester, C1-C10 phosphonates, 3'-alkylene phosphonate, chiral phosphonates, phosphinates, phosphoramidates, 3'-amino phosphoramidate, aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates.

**[00129]** Methods for determining the stability of oligonucleotide duplexes comprising unnatural base pairs (with or without linkers) include thermodynamic analysis by circular dichroism (CD) measurements and UV melting experiments. In some embodiments, DNA duplex stability studies are employed to facilitate the selection of a suitable unnatural nucleotide base pair, unnatural nucleobase, or unnatural nucleobase derivatives or substitutions. Suitably selected unnatural base pairs include those which increase oligonucleotide hybridization fidelity at other positions within the duplex. Suitably selected unnatural base pairs include those which increase oligonucleotide duplex stability. Suitably selected nucleobases may be used to optimize oligonucleotides for biotechnological or therapeutic applications where high fidelity hybridization and discrimination is critical. In some instances, an unnatural base pair is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more, as stable as a natural base pair in an oligonucleotide duplex. In some instances, the  $T_m$  of a duplex comprising one or more unnatural base pairs is less than 10 °C, 9 °C, 8 °C, 7 °C, 6 °C, 5 °C, 4.5 °C, 4 °C, 3.5 °C, 3 °C, 2.9 °C, 2.8 °C, 2.7 °C, 2.6 °C, 2.5 °C, 2.4 °C, 2.3 °C, 2.2 °C, 2.1 °C, 2 °C, 1.9 °C, 1.8 °C, 1.7 °C, 1.6 °C, 1.5 °C, 1.4 °C, 1.3 °C, 1.2 °C, 1.1 °C, 1 °C, 0.9 °C, 0.8 °C, 0.7 °C, 0.6 °C, 0.5 °C, 0.4 °C, 0.3 °C, 0.2 °C, 0.1 °C below the  $T_m$  of the same duplex wherein the one or more unnatural nucleobases are replaced with one or more natural nucleobases. In some embodiments, the presence of an unnatural base pair in an oligonucleotide duplex does not significantly perturb duplex structure.

**[00130]** In some embodiments, an oligonucleotide comprising a linker-derivatized nucleobase allows for the site-specific modification of that DNA or RNA during or after enzymatic synthesis. An unnatural nucleotide disclosed herein (e.g. a nucleotide comprising an unnatural  $\alpha$  or  $\beta$  nucleobase analog), in some instances, is modified with a linker that enables the attachment of different functional groups (e.g., cargo) without ablating polymerase recognition. Site specific functionalities include, but are not limited to fluorophores, NMR handles for characterization (e.g., F19), IR probes (e.g., azido and cyano groups), biotin (e.g. to facilitate identification and/or purification), affinity tags, liposomes, and nanoparticles. In one embodiment, a linker provides bioconjugation via cross-coupling (e.g., iodo group). In one embodiment, a linker provides a handle for bioconjugation via click chemistry (e.g., azido and alkyne substituents). In one embodiment, an oligonucleotide comprising a linker-derivatized nucleobase is useful as a primer and/or molecular beacon.

[00131] Further provided herein, in various embodiments, is the use of any nucleoside analogs disclosed herein ( $\alpha$  or  $\beta$ ), or analogs or derivatives thereof, in site-specific cleavage or functionalization of an oligonucleotide. In some embodiments, a nucleoside analog comprises one or more linkers configured for site-specific modification. Examples of nucleotide analogs comprising a linker moiety include, but are not limited to, d5SICSCO, d5SICSCC, dDMOCO, dDMOCC, dMMO2pCO, dMMO2pCC, dTPT3, dTPT3A, dTPT3PA, dTPT3CO, dMMSpCO, dMMSPA, and dTPT3CC, or ribosyl forms thereof, or analogs thereof. Provided herein, in various embodiments, are compositions of matter per se of the functionalized oligonucleotides, methods of preparation of the functionalized oligonucleotides, and methods of use of the functionalized oligonucleotides. Various embodiments provide dTPT3, dTPT3PA, dTPT3A, dTPT3CO, and dTPT3CC, or other linker-derivatized analogs of dTPT3, incorporated into oligonucleotides and the further reaction or derivatization of these unnatural nucleobase analogs incorporated in a oligonucleotide with various reagents for selective reaction with the unnatural nucleobase analogs in a oligonucleotide wherein the naturally occurring nucleobases (A, T, G, C, U) do not react with these reagents to any appreciable extent. The dTPT3-based family of linker-bearing unnatural nucleotides is especially central, as we have found that they are more efficiently replicated by DNA polymerases than are base pairs that include d5SICS or its linker-derivatized variants, which will significantly facilitate many of the potential applications. In some embodiments, the percent incorporation of an unnatural nucleotide comprising a linker into an oligonucleotide is at least 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%. In some embodiments, the percent extension of a next correct nucleotide into an oligonucleotide, wherein the next correct nucleotide follows incorporation of an unnatural nucleotide comprising a linker, is at least 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or greater. In some embodiments, the addition of a site-specific functionality decreases the percent incorporation of an unnatural nucleotide into an oligonucleotide by at most about 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1%. In some embodiments, the fidelity of a linker-derivatized unnatural nucleotide is at least 98%, 98.1%, 98.2%, 98.3%, 98.4%, 98.5%, 98.6%, 98.7%, 98.8%, 98.9%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%, 99.91%, 99.92%, 99.93%, 99.94%, 99.95%, 99.96%, 99.97%, 99.98% or 99.99%. Accordingly, in various embodiments, provided herein are methods for using the linker-derivatized unnatural nucleotides to produce DNA or RNA that is site-specifically modified with another molecule of interest. In some



embodiments, site-specific inclusion of different functionalities occurs either pre- or post-amplification. In some embodiments, site-specific functionalization is employed for SELEX applications.

[00132] An exemplary strategy to produce DNA or RNA that is site-specifically modified with another molecule of interest is referred to as the phosphorothioate strategy (Figure 3), which relies on the site-specific incorporation of a phosphorothioate group into an DNA or RNA via a ribo or deoxyribo  $\alpha$ -thiotriphosphate of one of the unnatural nucleosides from Figure 1 or 2. After incorporation into DNA or RNA, the phosphorothioate may be used to couple reagents that bear  $\gamma$ -bromo- $\alpha,\beta$ -unsaturated carbonyl-, iodo (or bromo)acetyl-, or aziridinylsulfonamide moieties to produce site-specifically functionalized DNA or RNA. Alternatively, after incorporation into DNA or RNA, the phosphorothioate may be used to site-specifically cleave the DNA or RNA using iodine in an alkaline solution or iodoethanol, respectively. Thus, the phosphorothioate strategy provides site-specific modification of the nucleic acid backbone, and provides for a method of site-specific cleavage of the oligonucleotide chain.

[00133] Another strategy to produce DNA or RNA that is site-specifically modified with another molecule of interest, referred to as the *linker strategy* (Figures 4 and 5), makes use of the derivatization of an unnatural nucleobases with a linker (Figure 2) that may be used to attach functional groups of interest, either before polymerization (via PCR or T7 RNA polymerase-mediated transcription using an appropriate functionalized nucleobase triphosphate reagent that is incorporated into the DNA or RNA chain being synthesized), or by reaction of the linker of the unnatural nucleobase after incorporation into the oligonucleotide chain with an appropriate functionalization reagent, e.g., an NHS containing reagent also comprising the desired functional group, wherein the NHS reacts with the free amino group of an amino-functionalized unnatural nucleobase such as d5SICS<sup>A</sup>, dMMO2<sup>A</sup>, or dTPT3<sup>A</sup>. Figure 4 also shows the amino functionalization linker strategy using d5SICS<sup>A</sup> and dMMO2<sup>PA</sup> that allows site-specific double labeling of duplex DNA.

[00134] For example, functionalization can be accomplished after incorporation into the oligonucleotide of the unnatural nucleobase with the linker bearing a primary amino group (e.g., dTPT3<sup>A</sup>). More specifically, the functionalization can be carried out via reaction of the primary amino (e.g., propargylamino group) and a cargo-bearing reagent including an N-hydroxysuccinimide (NHS) ester (Figure 4). The analogs developed for this application include d5SICS<sup>A</sup>, d5SICS<sup>PA</sup>, dMMO2<sup>A</sup>, dMMO2<sup>PA</sup>, dTPT3<sup>PA</sup>, and dTPT3<sup>A</sup> (recall that “A” refers to the nucleotide with a propargyl amine and “PA” refers to the same linker with a protecting group, see Figure 2 and its caption). The use of dTPT3<sup>PA</sup>, bearing a protected primary propargylamino

group, and dTPT3<sup>A</sup>, bearing the primary propargylamino group, for sequence-specific functionalization of oligonucleotides, is disclosed and claimed herein.

[00135] The site-specific functionalization of a oligonucleotide can also be accomplished using the Copper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC) (*i.e.* “Click chemistry” linker strategy; Fig. 5), and for these applications d5SICS<sup>CO</sup>, d5SICS<sup>CC</sup>, dDMO<sup>CO</sup>, dDMO<sup>CC</sup>, dMMO2<sup>pCO</sup>, dMMO2<sup>pCC</sup>, dTPT3<sup>CO</sup> and dTPT3<sup>CC</sup> (Figure 2), may be used. In each case, the ribosyltriphosphates of the unnatural nucleobases can be employed for transcription to produce site-specifically labeled RNA, and the deoxyribosyltriphosphates of the unnatural nucleobases can be used, e.g., in PCR, to produce site-specifically labeled DNA. The unnatural nucleobases comprising an acetylenic (alkynyl) linker group suitable for use in CuAAC conjugation, d5SICS<sup>CO</sup>, d5SICS<sup>CC</sup>, dDMO<sup>CO</sup>, dDMO<sup>CC</sup>, dMMO2<sup>pCO</sup>, dMMO2<sup>pCC</sup>, dTPT3<sup>CO</sup> and dTPT3<sup>CC</sup>, methods of their preparation, and methods of their use in preparing such site-specifically labeled oligonucleotides, are disclosed and claimed herein.

[00136] *Demonstration of general phosphorothioate strategy* (Figure 3). To demonstrate the feasibility of our system, we have prepared the  $\alpha$ -thiotriphosphate of the unnatural nucleotide, d5SICS (d5SICS- $\alpha$ S), and incorporated it into DNA opposite its cognate unnatural nucleotide dNaM, using standard PCR. The amplification efficiency and fidelity of incorporation of d(5SICS- $\alpha$ S)TP is greater than 99% and virtually identical to results obtained with d5SICS. To functionalize this unnatural base pair, we reacted the site-specifically incorporated phosphorothioate bond with iodoacetyl-PEG<sub>2</sub>-biotin to label the DNA duplex with the biotin functionality.<sup>8</sup> To characterize this site-specific adduct, we incubated it in the presence of streptavidin and then quantified the functionalization by gel shift assay. We were able to convert 60–70% of the phosphorothioate bond to the functionalized derivative, which is a standard efficiency (70%) for labeling protocols previously reported in literature (See Fidanza, J. A.; Ozaki, H.; McLaughlin, L. W., Site-specific labeling of DNA sequences containing phosphorothioate diesters. *J. Am. Chem. Soc.* **2002**, *114* (14), 5509-5517.). These conjugated derivatives show high stability under conditions typical for heat denaturation of DNA duplexes, *i.e.* at 50 °C overnight within the range of pH 6.0–8.3 (<10% decomposition), as well as for at 95 °C for 3 minutes at pH 8.3 (<5 % decomposition). We envision that the phosphorothioate strategy can be equally well employed with other unnatural base pairs, including d5SICS-dMMO2 and d5SICS-dNaM.

[00137] Provided herein, in various embodiments, is a phosphorothioate strategy using an unnatural base pair dTPT3-dNaM, dTPT3-dMMO2, or dTPT3-dDMO, and linker-derivatized variants thereof.

[00138] The phosphorothioate and linker-based strategies are not mutually exclusive and when combined should allow for a given site to be simultaneously modified with up to three different functional groups, one attached to a first nucleobase of a nucleobase pair, a second attached to a second nucleobase of a nucleobase pair, and a third attached to the backbone immediately 5' to an unnatural nucleotide.

[00139] *Demonstration of linker strategy with primary amine* (Figure 4). To further demonstrate the feasibility of our system, we have synthesized and characterized the amino- and protected amino-linker derivatized variants of d5SICS and dMMO2 (Figure 2). When paired in DNA opposite their cognate unnatural partners, we showed that each was well amplified by PCR and transcribed into RNA. Coupling of DNA containing dMMO2<sup>A</sup> or d5SICS<sup>A</sup> prepared by PCR amplification with NHS-ester biotin proceeds with 55% and 70% efficiency, respectively.

[00140] We have shown that the ribonucleotide triphosphates of 5SICS<sup>PA</sup>, 5SICS<sup>A</sup>, MMO2<sup>PA</sup> or MMO2<sup>A</sup>, are transcribed into RNA by T7 RNA polymerase with high efficiency and fidelity (Figure 7).

[00141] Provided herein, in various embodiments, is the site-specific modification of DNA or RNA using dTPT3<sup>L</sup>-dNaM, dTPT3<sup>L</sup>-dMMO2, or dTPT3<sup>L</sup>-(d)DMO (where R is a linker, *e.g.* R = H for dTPT3, R = 3-aminopropyn-1-yl for dTPT3<sup>A</sup>, R = dichloroacetyl-3-aminopropyn-1-yl for dTPT3<sup>PA</sup>, R = 4-oxahept-1,6-diyn-1-yl for dTPT3<sup>CO</sup>, R = hepta-1,6-diyn-1-yl for dTPT3<sup>CC</sup>).

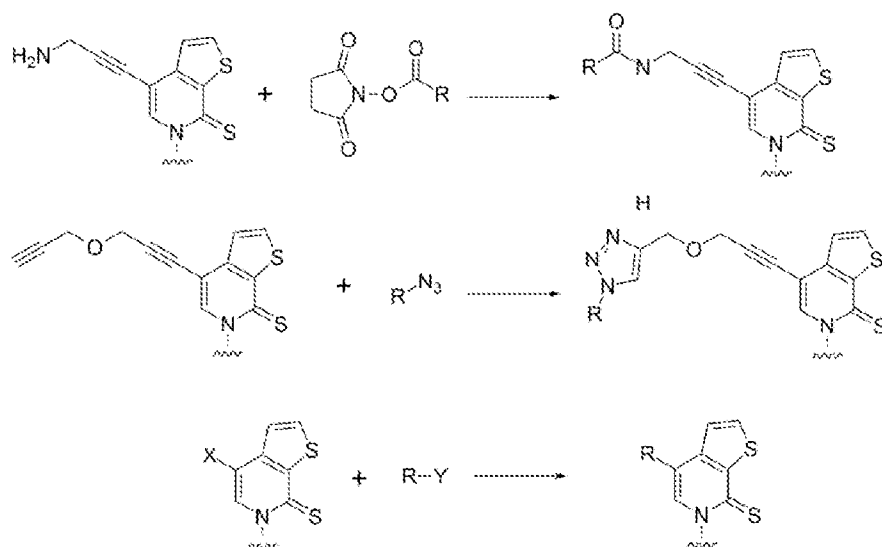
[00142] *Demonstration of general linker strategy with alkynes* (Figure 5). To further demonstrate the feasibility of our system, we have synthesized and characterized the alkynyl functionalized variants of d5SICS, dDMO, and dMMO2, including d5SICS<sup>CO</sup>, d5SICS<sup>CC</sup>, dDMO<sup>CO</sup>, dDMO<sup>CC</sup>, dMMO2<sup>pCO</sup>, dMMO2<sup>pCC</sup> (Figure 2) each of these alkyne-functionalized unnatural nucleotide should be efficiently PCR amplified when present in DNA. Once amplified, the DNA containing, for example, the d5SICS<sup>CO</sup>-dNaM base pair, may be efficiently site-specifically modified with small molecules or one or more proteins possessing azide groups using Click chemistry, *e.g.*, copper-catalyzed click reactions. We have also demonstrated the utility of dDMO, and dFEMO (Figure 2) for incorporation into oligonucleotides and the use of these functionalized oligonucleotides in click chemistry reactions with azides to functionalize the oligonucleotides in a site-specific manner.

[00143] *Demonstration of linker strategy with dTPT3<sup>PA</sup>* (Figure 6). Addition of a linker to the d5SICS and dMMO2 scaffolds significantly reduces the efficiency with which the unnatural nucleotides are enzymatically incorporated in DNA, which is expected to limit their practical applications. However, we have found that the dTPT3 scaffold is much more tolerant to linker addition (Figure 6). For example, dTPT3<sup>PA</sup>TP is incorporated into a primer opposite dNaM in a

template by DNA polymerases with virtually the same efficiency and fidelity as a natural base pair. Accordingly, provided herein, in various embodiments, is the use of unnatural nucleobases based on the dTPT3 scaffold, including dTPT3<sup>PA</sup> (protected amino-functional linker), dTPT3<sup>A</sup> (amino-functional linker), and dTPT3<sup>CO</sup> (alkyne-azide ether linker for derivatization via click chemistry), and dTPT3<sup>CC</sup> (alkyne-azide trimethylene linker for derivatization via click chemistry) in the synthesis of site-specific functionalized oligonucleotides.

[00144] Scheme 1 illustrates examples of dTPT3 with different linkers that could be used to site-specifically modify DNA or RNA.

Scheme 1:



[00145] For clarity only the dTPT3 scaffold nucleobase moieties are shown, but it is understood that they are used as nucleotides. The functionalization reactions can be carried out either prior to or after incorporation of the unnatural nucleobases into an oligonucleotide. Scheme 1, top reaction, illustrates the use of dTPT3<sup>A</sup> comprising a primary amine-bearing linker that is acylated using an activated ester to form an amide, wherein the R group comprises the cargo. The middle reaction of Scheme 1 illustrates the use of dTPT3<sup>CO</sup> (dTPT3<sup>CC</sup> could also be used) comprising an alkynyl-bearing linker reacted with an azide to yield a triazole via Click chemistry, wherein the R group of the triazole that is formed comprises the cargo. The bottom reaction illustrates the most general case for a dTPT3 scaffold derivative bearing a linker group R<sub>1</sub> with a reactive moiety that can selectively form a covalent bond with a R<sub>2</sub> group that includes a reactive moiety complementary to the reactive moiety of the linker, for example, thiol-maleimide, hydrazine-aldehyde, etc.

[00146] In one embodiment, a linker comprising an azide reactive group is useful for attaching an alkyne comprising cargo through a click reaction. In one embodiment, a linker comprising a thiol group can form reversible disulfide bonds or irreversible bonds with a variety of cargo

accepting groups, including, but not limited to, maleimide, bromide, iodide, sulphonyl derivatives, active esters and isothiocyanate derivatives. In one embodiment, a linker comprising an azide group is reactive with a cargo molecule comprising a phosphine group.

[00147] In one embodiment, an oligonucleotide comprising one or more linker-derivatized unnatural nucleobases is configured for use as a molecular beacon. The fluorophore of the molecular beacon is a cargo molecule attached to a reactive center of the linker-derivatized unnatural nucleobase. Exemplary fluorophore cargo molecules include, but are not limited to, 6-FAM, Fluorescein, Cy3<sup>TM</sup>, JOE (6-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein), Cy5<sup>TM</sup>, TAMRA, MAX, TET<sup>TM</sup>, ROX (carboxy-X-rhodamine), TYE<sup>TM</sup> 563, Hexachlorofluorescein, TEX 615, TYE<sup>TM</sup> 665, TYE 705, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 594, Alexa Fluor® 647, Alexa Fluor® 660, Alexa Fluor® 750, IRDye® 800CW, ATTO<sup>TM</sup> 488, ATTO<sup>TM</sup> 532, ATTO<sup>TM</sup> 550, ATTO<sup>TM</sup> 565, ATTO<sup>TM</sup> Rho101, ATTO<sup>TM</sup> 590, ATTO<sup>TM</sup> 633, ATTO<sup>TM</sup> 647N, Rhodamine Green<sup>TM</sup>-X, Rhodamine Red<sup>TM</sup>-X, 5-TAMRA<sup>TM</sup>, Texas Red®-X, Lightcycler® 640, and Dy 750.

[00148] An unnatural base pair, in some embodiments, allows for the site-specific inclusion of different functionalities into DNA for Systematic Evolution of Ligands by Exponential Enrichment (SELEX) applications, including the generation of DNA and/or RNA aptamers. DNA and RNA aptamers have a variety of targets, including nucleic acids, small molecules, peptides, carbohydrates, and cells. SELEX includes the creation of a library of nucleic acid molecules, contacting the library with target molecules to select nucleic acid molecules which bind to the target molecules, and amplifying library members which bound to target molecules. Additional rounds of selection and amplification continue until sufficient aptamers are recovered. An aptamer, in one aspect, includes any unnatural base disclosed herein. In some embodiments, a SELEX experiment, wherein library components comprise unnatural nucleobases, generates an aptamer affinity against a target molecule in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or fewer rounds of selection than a library which does not comprise unnatural nucleobases. In some embodiments, an aptamer comprising one or more unnatural nucleobases has a greater affinity for a target molecule than an aptamer containing only natural nucleobases. The addition of one or more unnatural nucleobases in a SELEX library increases the chemical and structural diversity of the resulting DNA or RNA aptamers. In some embodiments, an unnatural aptamer has at least a nanomolar affinity against its target molecule. In some embodiments, an unnatural aptamer has at least a picomolar affinity against its target molecule. For example, an unnatural aptamer has an affinity for its target molecule which is between 1 and 1,000 pM. In some embodiments, an unnatural aptamer has at least a femtomolar affinity for its target molecule. For example, an unnatural aptamer has an affinity for its target molecule which is between 1

and 1,000 fM. An unnatural aptamer selected using SELEX may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more unnatural nucleobases. In some embodiments, an unnatural aptamer comprises dTPT3 or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\alpha 14a$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\alpha 14b$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\alpha 14c$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\alpha 14d$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\alpha 14e$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\alpha 14f$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\beta 8a$  or a derivative or analog thereof. In some embodiments, an unnatural aptamer comprises a nucleobase having the formula  $\beta 8b$  or a derivative or analog thereof.

**[00149]** Various combinations of the components set forth above in regard to exemplary reaction mixtures and reaction methods can be provided in a kit form. Such a kit can include individual components that are separated from each other, for example, being carried in separate vessels or packages. A kit can include one or more sub-combinations of the components set forth herein, the one or more sub-combinations being separated from other components of the kit. The sub-combinations can be combinable to create a reaction mixture set forth herein (or combined to perform a reaction set forth herein). In particular embodiments, a sub-combination of components that is present in an individual vessel or package is insufficient to perform a reaction set forth herein.

**[00150]** However, the kit as a whole can include a collection of vessels or packages the contents of which can be combined to perform a reaction set forth herein.

**[00151]** A kit can include a suitable packaging material to house the contents of the kit. The packaging material can be constructed by well-known methods, preferably to provide a sterile, contaminant-free environment. The packaging materials employed herein can include, for example, those customarily utilized in commercial kits sold for use with nucleic acid sequencing systems. Exemplary packaging materials include, without limitation, glass, plastic, paper, foil, and the like, capable of holding within fixed limits a component set forth herein.

**[00152]** The packaging material can include a label which indicates a particular use for the components. The use for the kit that is indicated by the label can be one or more of the methods set forth herein as appropriate for the particular combination of components present in the kit.

For example, a label can indicate that the kit is useful for a method of conjugating a cargo molecule to a linker moiety of an unnatural nucleobase in an oligonucleotide.

[00153] Instructions for use of the packaged reagents or components can also be included in a kit. The instructions will typically include a tangible expression describing reaction parameters, such as the relative amounts of kit components and sample to be admixed, maintenance time periods for reagent/sample admixtures, temperature, buffer conditions, and the like.

[00154] It will be understood that not all components necessary for a particular reaction need be present in a particular kit. Rather one or more additional components can be provided from other sources. The instructions provided with a kit can identify the additional component(s) that are to be provided and where they can be obtained.

[00155] In one embodiment, a kit provides one or more unnatural nucleobases or derivatives thereof and reagents configured for performing site-specific functionalization using the one or more unnatural nucleobases or derivatives thereof.

### Examples

[00156] Currently, the free nucleosides and phosphoramidites of d5SICS and dNaM are commercially available from Berry and Associates (Dexter, MI).

#### *Example 1. PCR-based screen to identify unnatural base pairs.*

[00157] The triphosphates of the  $\alpha 6$  group were prepared from the previously reported nucleosides (Kubelka, T., Slavetinska, L., Eigner, V. and Hock, M. Synthesis of 2,6-disubstituted pyridin-3-yl C-2'-deoxyribonucleosides through chemoselective transformations of bromo-chloropyridine C-nucleosides. *Org. Biomol. Chem.*, 11, 4702-4718) according to Ludwig, J. and Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-O-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.*, 54, 631-635. The purity of all other triphosphates was confirmed by MALDI-TOF and UV-VIS. Taq and OneTaq DNA polymerases were purchased from New England Biolabs (Ipswich, MA). A mixture of dNTPs was purchased from Fermentas (Glen Burnie, MD). SYBR Green I Nucleic Acid Gel Stain (10,000 $\times$ ) was purchased from Life Technologies (Carlsbad, CA). The synthesis of the DNA templates, D8 (Malyshev, D.A., Dhami, K., Quach, H.T., Lavergne, T., Ordoukhanian, P., Torkamani, A. and Romesberg, F.E. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc. Natl. Acad. Sci. USA*, 109, 12005-12010), used for screening rounds 1-5, and D6 (Malyshev, D.A., Seo, Y.J., Ordoukhanian, P. and Romesberg, F.E. PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.*, 131, 14620-14621), used for all other amplifications, was described previously. Sanger sequencing was carried out as described previously (Malyshev, D.A., Dhami, K., Quach, H.T., Lavergne, T.,

**[00158]** All PCR amplifications were performed in a CFX Connect Real-Time PCR Detection System (Bio-Rad), in a total volume of 25  $\mu$ L using the following conditions: 1 $\times$  OneTaq reaction buffer, 0.5 $\times$  Sybr Green I, MgSO<sub>4</sub> adjusted to 4.0 mM, 0.2 mM of each dNTP, 50  $\mu$ M of each unnatural triphosphate, 1 mM of Primer1 and Primer2 (See Table 2), and 0.02 U/ $\mu$ L of the DNA polymerase. Other conditions specific for each round of screening are described in Table 3. Amplified products were purified using DNA Clean and Concentrator-5 spin columns from Zymo Research (Irvine, CA). After purification, the PCR products were sequenced on a 3730 DNA Analyzer (Applied Biosystems) to determine the retention of the unnatural base pair as described below. Fidelity was characterized from unnatural base pair (UBP) retention as determined by sequencing with Primer1 on a 3730 DNA Analyzer (Applied Biosystems).

Name	Sequence (5' to 3')	Remarks
Primer1	<u>CACACAGGAAACAGCTATGAC</u>	Primers for PCR
Primer2	<u>GAAATTAATACGACTCACTATAGG</u>	
Primer1- poly-dT	TT TTTTTT <u>CACACAGGAAACAGCTATGAC</u>	Primers for Sanger sequencing
Primer2- poly-dT	TT TTTTTTTT <u>GAAATTAATACGACTCACTATAGG</u>	
D8	<u>CACACAGGAAACAGCTATGACCCGGGTTATTACATGCGCTAGCACTT</u> GGAATTCACAATACT <b>NaM</b> TCTTTAAGGAAACCATAGTAAATCTCCTT CTTAAAGTTAAGCTTAACC <u>CCTATAGTGAGTCGTATTAATTTC</u>	
D6	<u>CACACAGGAAACAGCTATGACCCGGGTTATTACATGCGCTAGCACTT</u> GGAATTCACCAGACGNNN <b>NaM</b> NNNCGGGACCCATAGTAAATCTCCT TCTTAAAGTTAAGCTTAACC <u>CCTATAGTGAGTCGTATTAATTTC</u>	N = randomized natural nucleotide



Table 3. PCR conditions for each consecutive round of the PCR screen.

Reaction Components	Rounds 1-4	Round 5	Round 6	Final PCR characterization	
Buffer	1×OneTaq	1×OneTaq	1×OneTaq	1×OneTaq	1×OneTaq
Enzyme	OneTaq	Taq	OneTaq or Taq	Taq	OneTaq
Template	D8 <sup>a</sup> (0.1 ng)	D8 <sup>a</sup> (0.1 ng)	D6 <sup>a</sup> (0.01 ng)	D6 <sup>a</sup> (0.01 ng)	D6 <sup>a</sup> (0.01 ng)
Thermal conditions					
Initial denaturing	96 °C, 60 s	96 °C, 60 s	96 °C, 60 s	96 °C, 60 s	96 °C, 60 s
Denaturing	96 °C, 10 s	96 °C, 10 s	96 °C, 5 s	96 °C, 5 s	96 °C, 10 s
Annealing	60 °C, 15 s	60 °C, 15 s	60 °C, 5 s	60 °C, 5 s	60 °C, 5 s
Extension	68 °C, 60 s	68 °C, 60 s	68 °C, 10 s	68 °C, 10 s	68 °C, 30 s
# of cycles	16	16	24	16 +16 +20 <sup>b</sup>	16 +16 +20 <sup>b</sup>

<sup>a</sup> See Table 2 for sequences of the templates and primers. <sup>b</sup> Initial amount of template was 0.01 ng. PCR mixture was amplified over 16 cycles, diluted 40,000-fold and amplified over another 16 cycles. The dilution/amplification step was repeated resulting in 52 total cycles of amplification.

**[00159]** Specific PCR assay conditions. PCR with the most promising UBPs was carried out with the conditions as described in Table 3. PCR products were further purified on 2% agarose gels, followed by single band excision and subsequent clean up using the Zymo Research Zymoclean Gel DNA Recovery Kit. After elution with 20 µl of water, the DNA concentration was measured using fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Life Technologies), and purified amplicons were sequenced in triplicate with both Primer1 and Primer2 to determine UBP retention and thus amplification fidelity. Amplification of DNA containing the pairs involving analogs of group α6 was performed with OneTaq polymerase under the following thermal cycling conditions: initial denaturation at 96 °C for 1 min; 16 cycles of 96 °C for 10 s, 60 °C for 15 s, 68 °C for 1 min. Fidelity was determined by sequencing amplicons in the Primer1 direction in triplicate. Amplification of DNA containing the UBPs formed between dTPT3 and d2MN or dDM2 was performed using OneTaq or Taq polymerases for 16 cycles under the following thermal cycling conditions: 1) OneTaq: initial denaturation at 96 °C for 1 min, 96 °C for 10 s, 60 °C for 15 s, 68 °C for 1 min; or 2) Taq: initial denaturation at 96 °C for 1 min, 96 °C for 5 s, 60 °C for 5 s, 68 °C for 10 s. Fidelity was determined by sequencing amplicons in the Primer1 direction in triplicate.

**[00160]** Results. To screen for well replicated UBPs, unnatural deoxynucleoside triphosphates were grouped for analysis into either dMMO2/dNaM- or d5SICS/dTPT3-like analogs, although

the distinction is not completely clear in all cases. In total, 80 dMMO2/dNaM analogs were grouped into twelve “ $\alpha$  groups” ( $\alpha 1 - \alpha 12$ ; Figure 8), and 31 d5SICS/dTPT3 analogs were grouped into six “ $\beta$  groups” ( $\beta 1 - \beta 6$ ; Figure 9). Note that the group designations used here should not be confused with anomer designation (all analogs examined are  $\beta$  glycosides). In addition, to increase the structure-activity relationship (SAR) content of the screen, seven previously reported nucleoside analogs (dTOK576–dTOK588) with substituted pyridyl nucleobases (Kubelka, T., Slavetinska, L., Eigner, V. and Hocekm M. Synthesis of 2,6-disubstituted pyridin-3-yl C-2'-deoxyribonucleosides through chemoselective transformations of bromo-chloropyridine C-nucleosides. *Org. Biomol. Chem.*, 11, 4702-4718.) were phosphorylated and included as group  $\alpha 6$ . For screening, a 134-mer single-stranded DNA template containing a centrally located dNaM (referred to as D8) was PCR amplified in the presence of the natural triphosphates (200  $\mu$ M each), all pairwise combinations of an  $\alpha$  and a  $\beta$  triphosphate group shown in Figures 8 and 9 (50  $\mu$ M each), and 0.02 U/ $\mu$ L DNA polymerase. During the first round of PCR, dNaM templates the incorporation of an  $\alpha$  analog and is then replaced by a  $\beta$  analog when the original strand is copied in the second round, with the resulting UBP amplified in subsequent rounds. The amplification product of each reaction was analyzed by Sanger sequencing. The presence of an unnatural nucleotide results in the abrupt termination of the sequencing chromatogram, allowing the level of UBP retention to be quantified by the amount of read through. The percentage of UBP retained in the DNA after amplification during each round of screening is shown in Figure 14.

**[00161]** The first round of screening employed 0.1 ng of template and 16 cycles of amplification under relatively permissive conditions that included OneTaq polymerase and a 1 min extension time. For this example, OneTaq is considered permissive because it is a mixture of Taq (a family A polymerase) and Deep Vent (a family B polymerase), with the latter possessing exonucleotidic proofreading that allows for the excision of an incorrectly incorporated triphosphate. Under these conditions, only the pairs involving group  $\beta 5$  or  $\beta 6$  showed high retention.

**[00162]** The combinations of  $\beta 5$  or  $\beta 6$  and the  $\alpha$  groups that showed the highest retention were progressed to a second round of screening, wherein they were divided into smaller groups (denoted by a, b, or c; Figures 8 and 9). High retention ( $\geq 97\%$ ) was observed with  $\beta 5a$  and  $\alpha 2c$ ,  $\alpha 9a$ ,  $\alpha 9c$ ,  $\alpha 10a$ ,  $\alpha 10c$ ,  $\alpha 12b$ , or  $\alpha 12c$ ; with  $\beta 5b$  and  $\alpha 9a$ ,  $\alpha 9b$ ,  $\alpha 10c$ , or  $\alpha 12b$ ; and with  $\beta 6b$  and  $\alpha 10c$ . Moderate retention (84 – 96%) was observed with  $\beta 5a$  and  $\alpha 1a$ ,  $\alpha 1b$ ,  $\alpha 6a$ ,  $\alpha 9b$ ,  $\alpha 10b$ , or  $\alpha 12a$ ;  $\beta 5b$  and  $\alpha 1a$ ,  $\alpha 1b$ ,  $\alpha 2c$ ,  $\alpha 6a$ ,  $\alpha 9c$ ,  $\alpha 10a$ ,  $\alpha 12a$ , or  $\alpha 12c$ ;  $\beta 6a$  and  $\alpha 1b$  or  $\alpha 10c$ ; and  $\beta 6b$  and  $\alpha 1a$ ,  $\alpha 6a$ ,  $\alpha 9a-c$ ,  $\alpha 10a$ ,  $\alpha 10b$ , or  $\alpha 12a-c$ .

[00163] For a third round of screening,  $\alpha$  analogs were analyzed in groups of only one to three compounds, and group  $\beta$ 6a was subdivided into its two constituent triphosphates, dTPT1TP and dFPT1TP. The highest retention ( $\geq 90\%$ ) was observed with  $\beta$ 5a and  $\alpha$ 1a,  $\alpha$ 2cII,  $\alpha$ 9a–c,  $\alpha$ 10aI,  $\alpha$ 10aII,  $\alpha$ 10c,  $\alpha$ 12b, or dTfMOTP;  $\beta$ 5b and  $\alpha$ 9a,  $\alpha$ 9c, or  $\alpha$ 10c; dFPT1TP and  $\alpha$ 10aI; and  $\beta$ 6b and  $\alpha$ 1a,  $\alpha$ 9a–c,  $\alpha$ 10aI,  $\alpha$ 10aII,  $\alpha$ 10c,  $\alpha$ 12b, dNMOTP, dTfMOTP, or dCNMOTP. Only slightly less retention (80–89%) was seen with  $\beta$ 5a and  $\alpha$ 2cI,  $\alpha$ 12a, dNMOTP, dQMOTP, or dTOK587TP;  $\beta$ 5b and  $\alpha$ 1a,  $\alpha$ 2cII,  $\alpha$ 10aI,  $\alpha$ 10aII,  $\alpha$ 12b, or dTOK587TP; dFPT1TP and  $\alpha$ 10c; and  $\beta$ 6b and  $\alpha$ 12a, dQMOTP, dFuMO1TP, or dTOK587TP.

[00164] For a fourth round of screening, all of the  $\alpha$  derivatives from Figure 9 were analyzed as individual triphosphates, with the exception of  $\alpha$ 9b and  $\alpha$ 9c, which remained grouped. The highest retention ( $\geq 91\%$ ) was observed with  $\beta$ 5a and  $\alpha$ 9b,  $\alpha$ 9c, dFIMOTP, dIMOTP, dFEMOTP, dMMO2TP, d2OMeTP, dDMOTP, d5FMTP, dNaMTP, dVMOTP, dZMOTP, dCIMOTP, dTfMOTP, dQMOTP, d2MNTP, dDM2TP, or dTOK587TP;  $\beta$ 5b and  $\alpha$ 9b,  $\alpha$ 9c, dFIMOTP, dIMOTP, dFEMOTP, dNaMTP, dZMOTP, dCIMOTP, dQMOTP, dMM1TP, dDM2TP, or dTOK587TP;  $\beta$ 6 analog dFPT1TP and  $\alpha$  analogs d2OMeTP or dNaMTP; and  $\beta$ 6b and  $\alpha$ 9b,  $\alpha$ 9c, dFIMOTP, dIMOTP, dFEMOTP, dMMO2TP, dDMOTP, dTMOTP, dNMOTP, d5FMTP, dNaMTP, dVMOTP, dZMOTP, dCIMOTP, dTfMOTP, dQMOTP, dCNMOTP, d2MNTP, dTOK587TP, or dFuMO2TP.

[00165] To increase the stringency of the screen, a fifth round was performed with Taq polymerase instead of OneTaq, as it lacks exonuclease proofreading activity and thus increases the sensitivity to mispair synthesis. This round also separated all remaining  $\alpha$  and  $\beta$  groups into individual triphosphates. The highest retention ( $\geq 90\%$ ) was seen with dSICSTP and dNaMTP; dSNICSTP and dNaMTP; dTPT2TP and dFDMOTP; dTPT3TP and dFIMOTP, dIMOTP, or dNaMTP; and dFTPT3TP and dFIMOTP, dIMOTP, dFEMOTP, dNMOTP, dNaMTP, dCIMOTP, dTfMOTP, or dCNMOTP.

[00166] To better differentiate between the UBPs, we progressed the sixty-two most promising candidate UBPs to a sixth round of screening in which the template concentration was decreased 10-fold (to 10 pg) to allow for greater amplification, and thereby afford greater discrimination, and the template was changed to D6 (Malyshev, D.A., Seo, Y.J., Ordoukhanian, P. and Romesberg, F.E. PCR with an expanded genetic alphabet. *J. Am. Chem. Soc.*, 131, 14620–14621), where the three flanking nucleotides on either side of the unnatural nucleotide are randomized among the natural nucleotides. Moreover, the denaturation and annealing steps were decreased to 5 s each, and the extension time was decreased to 10 s. Under these conditions, we explored amplification either with OneTaq or with Taq alone. The results with OneTaq showed the highest retention ( $>95\%$ ) with dSICSTP and dNaMTP; dSNICSTP and

dFEMOTP; dTPT3TP and dFIMOTP, dIMOTP, dFEMOTP, dZMOTP, or dNaMTP; and dFTPT3TP and dIMOTP or dFEMOTP. Moderate retention (86% – 94%) was observed with dSICSTP and dFEMOTP or dDM2TP; d5SICSTP and dNaMTP; dSNICSTP or dIMOTP; dTPT2TP and dNaMTP; dTPT3TP and dNMOTP, dCIMOTP, dQMOTP, dCNMOTP or d2MNTP; and dFTPT3TP and dFIMOTP, dNaMTP, dZMOTP, dCIMOTP, dTfMOTP, or dCNMOTP. While retention during Taq-mediated amplification was in general reduced relative to that with OneTaq, the general trends were similar. The highest retention (>96%) was observed with dTPT3TP and dFIMOTP or dIMOTP, and with dFTPT3TP and dFIMOTP. Only slightly lower retention (89%–94%) was observed with dTPT3TP and dFEMOTP, dNaMTP, or dCNMOTP; and dFTPT3TP and dIMOTP, dFEMOTP, dNaMTP, dCIMOTP, dCNMOTP, or d2MNTP.

[00167] Amplification with the most promising combinations of triphosphates, dTPT3TP or dFTPT3TP and dFIMOTP, dIMOTP, dFEMOTP, or dNaMTP, was then performed over 52 cycles with Taq and a 10 s extension time, to explore particularly stringent conditions, or with OneTaq and a 30 s extension time, to explore more practical conditions (Table 4). Both amplified strands were sequenced in triplicate to determine UBP retention with high accuracy. With Taq, dTPT3-dNaM, dTPT3-dFIMO, dFTPT3-dNaM, and dFTPT3-dFIMO showed the highest retention, while the pairs involving dIMO and dFEMO showed somewhat less retention. With OneTaq, dTPT3-dNaM and dFTPT3-dNaM showed the highest retention, followed closely by dFTPT3-dFIMO and dTPT3-dFIMO.

Table 4.

dβTP	dαTP	Amplification, ×10 <sup>12</sup>	Retention, %	Fidelity per Doubling, %
<i>Taq</i> , 10 s Extension				
<b>TPT3</b>	<b>FIMO</b>	8.5	84 ± 3	99.60 ± 0.09
	<b>IMO</b>	6.3	81 ± 5	99.50 ± 0.15
	<b>FEMO</b>	5.0	79 ± 3	99.44 ± 0.09
	<b>NaM</b>	5.8	86.5 ± 0.5	99.66 ± 0.01
<b>FTPT3</b>	<b>FIMO</b>	4.8	84 ± 3	99.60 ± 0.09
	<b>IMO</b>	5.6	82 ± 5	99.54 ± 0.13
	<b>FEMO</b>	5.7	81 ± 4	99.51 ± 0.11
	<b>NaM</b>	3.7	91 ± 6	99.76 ± 0.15
<b>5SICS</b>	<b>NaM</b>	9.3	< 50 <sup>b</sup>	< 85 <sup>b</sup>
<i>OneTaq</i> , 1 min Extension				

WO 2015/021432		PCT/US2014/050423		
<b>TPT3</b>	<b>FIMO</b>	8.7	84.7 ± 1.1	99.61 ± 0.03
	<b>IMO</b>	9.4	82.9 ± 1.7	99.56 ± 0.05
	<b>FEMO</b>	10.4	82.2 ± 1.0	99.55 ± 0.03
	<b>NaM</b>	8.3	91.2 ± 1.3	99.79 ± 0.03
<b>FTPT3</b>	<b>FIMO</b>	8.2	86 ± 3	99.65 ± 0.08
	<b>IMO</b>	7.1	76.8 ± 1.6	99.38 ± 0.05
	<b>FEMO</b>	6.3	72.4 ± 1.4	99.24 ± 0.04
	<b>NaM</b>	7.0	90 ± 2	99.76 ± 0.06
<b>5SICS</b>	<b>NaM</b>	8.1	77.1 ± 0.7	99.00 ± 0.02

<sup>a</sup>Retention and fidelity determined as described in Example 1. <sup>b</sup>UBP retention below 50%, and fidelity is thus estimated to be <85%.

[00168] The screening data suggest that several pairs formed between dTPT3 and the previously unexamined pyridine-based derivatives of  $\alpha 6$  were reasonably well replicated. Thus, we examined in triplicate the amplification of DNA containing these UBPs using OneTaq and 16 amplification cycles with 1 min extension times (Table 5). The pairs formed between dTPT3 and dTOK580, dTOK582, or dTOK586 were poorly replicated. However, the pairs formed between dTPT3 and dTOK588, dTOK581, dTOK576, and dTOK587 were amplified with a retention of 62%, 65%, 85%, and 94%, respectively.

Table 5. Amplification and fidelity data of dTPT3 against pyridine-based derivatives from group  $\alpha 6$ ; DM5, MMO2, DMO, and NaM were also characterized for scaffold comparison.

d $\beta$ TP	d $\alpha$ TP	Amplification	Retention, %	Fidelity, %
<b>TPT3</b>	<b>TOK576</b>	780	85.20 ± 1.12	98.35 ± 0.14
<b>TPT3</b>	<b>TOK580</b>	1056	< 50 <sup>b</sup>	< 85 <sup>b</sup>
<b>TPT3</b>	<b>TOK581</b>	1034	65.07 ± 0.15	95.80 ± 0.02
<b>TPT3</b>	<b>TOK582</b>	1240	< 50 <sup>b</sup>	< 85 <sup>b</sup>
<b>TPT3</b>	<b>TOK586</b>	948	< 50 <sup>b</sup>	< 85 <sup>b</sup>
<b>TPT3</b>	<b>TOK587</b>	818	93.81 ± 1.35	99.34 ± 0.15
<b>TPT3</b>	<b>TOK588</b>	666	61.98 ± 7.09	94.99 ± 1.13
<b>TPT3</b>	<b>DM5</b>	- <sup>a</sup>	- <sup>a</sup>	- <sup>a</sup>
<b>TPT3</b>	<b>MMO2</b>	1096	90.95 ± 3.63	99.06 ± 0.40
<b>TPT3</b>	<b>DMO</b>	864	84.02 ± 1.92	98.23 ± 0.23
<b>TPT3</b>	<b>NaM</b>	1004	99.23 ± 1.12	99.92 ± 0.11

<sup>a</sup> No amplification was detected for this sample. <sup>b</sup> Unnatural base pair retention was below 50% and fidelity is thus estimated less than 85%.

[00169] Finally, the screening data suggested that the pairs formed between dTPT3 and d2MN or dDM2 are reasonably well replicated, despite neither d2MN nor dDM2 possessing a putatively essential ortho H-bond acceptor. Thus, these pairs were further examined via 16 cycles of amplification with OneTaq or Taq alone, and with extension times of either 1 min or 10 s (Table 6). With Taq alone, only poor retention was observed. However, with OneTaq, retention was better for both pairs. Retention of the dTPT3-dDM2 pair is 58% and 69% with 1 min and 10 s extension times, respectively. Remarkably, dTPT3-d2MN is amplified with retentions of 96% and 94% with 1 min and 10 s extension times, respectively.

Table 6. Amplification and fidelity data of dTPT3 against either d2MN or dDM2,  $\alpha$  analogs without an ortho H-bond acceptor.

Enzyme	d $\beta$ TP	d $\alpha$ TP	Extension Time	Amplification	Retention, %	Fidelity per Doubling, %
OneTaq	<b>TPT3</b>	<b>2MN</b>	1 min	880	95.54 $\pm$ 1.55	99.53 $\pm$ 0.17
			10 s	224	93.53 $\pm$ 1.42	99.15 $\pm$ 0.20
	<b>TPT3</b>	<b>DM2</b>	1 min	1420	57.92 $\pm$ 6.02	94.89 $\pm$ 0.94
			10 s	376	68.46 $\pm$ 4.34	95.65 $\pm$ 0.72
Taq	<b>TPT3</b>	<b>2MN</b>	1 min	- <sup>a</sup>	-	-
			10 s	- <sup>a</sup>	-	-
	<b>TPT3</b>	<b>DM2</b>	1 min	334	< 50 <sup>b</sup>	< 85 <sup>b</sup>
			10 s	266	< 50 <sup>b</sup>	< 85 <sup>b</sup>

<sup>a</sup> No amplification was detected for these samples. <sup>b</sup> Unnatural base pair retention was below 50% and fidelity is thus estimated less than 85%.

[00170] Discussion. A PCR-based screen to identify the most promising UBPs was performed herein. To increase the SAR content of the screen, seven novel  $\alpha$  derivatives that are based on a pyridyl scaffold with different substituents at the positions ortho and para to the glycosidic linkage were included.

[00171] Structure-activity relationship data. Even under permissive conditions, where exonucleotidic proofreading activity was present and extension times were 1 min, only mixed groupings of  $\alpha$  analogs with  $\beta$  analogs showed significant levels of retention, demonstrating that

efficient replication requires the pairing of an  $\alpha$  scaffold with a  $\beta$  scaffold. However, the only  $d\beta$  groups that showed high retention were  $\beta 5$  and  $\beta 6$ . This reveals the privileged status of the d5SICS/dTPT3-like scaffold relative to all of the others examined. The dominant contribution to the high retention with group  $\beta 5$  proved to result not from pairs involving d5SICS, but rather from pairs involving dSICS, and to a lesser extent dSNICS. For example, under all conditions, dSICS-dNaM was better replicated than d5SICS-dNaM. d5SICS resulted from the optimization of dSICS for pairing with dMMO2; apparently, the increased bulk of dNaM makes the added methyl group deleterious. Furthermore, dSNICS-dNaM is replicated nearly as well (with OneTaq) or better (with Taq) than d5SICS-dNaM, suggesting that a 6-aza substituent optimizes UBP synthesis by facilitating insertion of the unnatural triphosphate opposite dNaM or by increasing the efficiency with which the unnatural nucleotide templates the insertion of dNaMTP. Finally, dSNICS-dFEMO is also better replicated than d5SICS-dNaM, but only in the presence of proofreading, suggesting that while triphosphate insertion may be less efficient, increased efficiency of extension results in an overall increase in fidelity. The dominant contribution to high fidelity retention with group  $\beta 6$  was provided by dTPT3 and dFTPT3. In general, both paired well with dNaM, dFEMO, dFIMO, or dIMO. dTPT3 paired especially well with dFIMO and dIMO, suggesting that the para iodo substituent mediates favorable interactions, and it also paired well with dFEMO and especially dNaM when exonuclease activity was present. dFTPT3 paired well with either dIMO or dFEMO in the presence of exonuclease activity, as well as with dFIMO and dNaM in its absence.

[00172] While the nitrogen substituent of the pyridine-based  $\alpha$  analogs (group  $\alpha 6$ ) was generally detrimental for replication, a more detailed analysis of the UBPs formed with dTPT3 revealed several trends. A methyl, chloro, or amino substituent at the position ortho to the C-glycosidic linkage resulted in poorly replicated pairs, presumably due to poor extension after incorporation of the unnatural triphosphate. The ortho methoxy substituent of dTOK581 resulted in better replication, presumably due to its ability to both hydrophobically pack with the template during UBP synthesis and accept an H-bond with a polymerase-based H-bond donor during extension. The data also revealed that the methylsulfanyl ortho substituent of dTOK588, dTOK576, and especially dTOK587 results in better replication. This improvement is likely due to more optimized compromise between the ability to hydrophobically pack and the ability to accept an H-bond from the polymerase at the primer terminus. In addition, the para substituent in this series of derivatives can contribute to efficient replication, with a bromo substituent being the best, followed by a second methylsulfanyl group, and then finally a simple methyl group. When dTOK587, with its combination of the ortho methylsulfanyl and para bromo substituents, was paired with dTPT3, the resulting UBP was replicated by OneTaq and 1 min extension times with

a fidelity (calculated from retention level) of 99.3%, which is slightly better than d5SICS-dMMO2 under similar conditions. Clearly, similar ortho methylsulfanyl and para bromo substituents should be examined with the more efficiently replicated  $\alpha$ -like scaffolds, such as dFIMO and dNaM.

[00173] The replication of the pairs formed between dTPT3 and d2MN or dDM2 also merits discussion. DNA containing these pairs is not amplified by Taq alone, but is well amplified by OneTaq. This result was unexpected because neither d2MN nor dDM2 possesses the ortho H-bond acceptor that has been postulated to be essential for extension of the nascent (natural or unnatural) primer terminus. Specifically, when a nucleotide is positioned at the growing primer terminus, the H-bond acceptor is disposed into the developing minor groove where it accepts an H-bond from the polymerase, and this H-bond is thought to be required for proper terminus alignment. When amplified with OneTaq and a 1 min extension time, dTPT3-d2MN is replicated with a fidelity of 99.5%, which only drops to 99.1% when the extension time is reduced to 10 s. The absence of amplification in the absence of proofreading, coupled with the only small decrease observed in the presence of proofreading when extension times were reduced, implies that the surprisingly high fidelity amplification of DNA containing dTPT3-d2MN results from selective extension of the UBP relative to mispairs. This suggests that the absence of an ortho H-bond acceptor is more deleterious for the extension of a mispair than for the extension of the UBPs.

[00174] Efforts toward the expansion of the genetic alphabet. Overall, the data confirms that dTPT3-dNaM is the most promising UBP of those tested. However, the pairs formed between dTPT3 and dFEMO, dFIMO, or dIMO, or between dFTPT3 and dNaM, dFEMO, dFIMO, or dIMO, are also promising. In addition to the most promising UBPs noted above, it is noteworthy that a remarkable number of additional novel pairs are replicated with only a moderately reduced fidelity, or are replicated with a high fidelity when the amplification is performed under less stringent conditions (Table 7). Along with the most efficiently replicated UBPs, these pairs provide a wide range of scaffolds with diverse physicochemical properties for further optimization efforts, where different physicochemical properties are expected to bestow the constituent nucleotides with different pharmacokinetic-like properties.

Table 7.

d $\beta$ TP	d $\alpha$ TP	Retention (%)
<b>SICS</b>	<b>NaM</b>	99 <sup>a</sup>
<b>SICS</b>	<b>FEMO</b>	92 <sup>b</sup>
<b>SNICS</b>	<b>NaM</b>	90 <sup>a</sup>

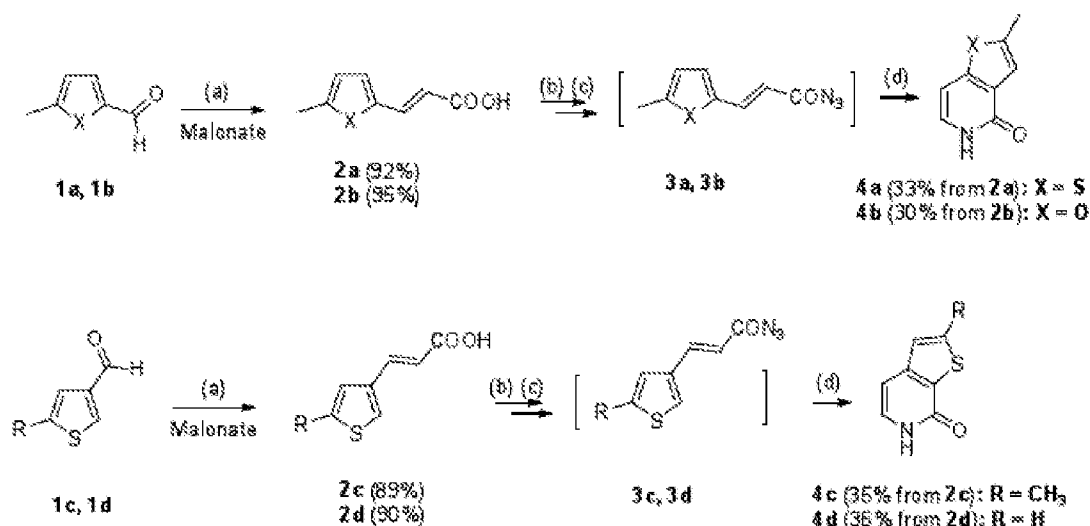


SNICS	FEMO	95 <sup>b</sup>
SNICS	IMO	88 <sup>b</sup>
TPT3	NMO	89 <sup>b</sup>
TPT3	ZMO	86 <sup>b</sup>
TPT3	CIMO	90 <sup>b</sup>
TPT3	QMO	90 <sup>b</sup>
TPT3	CNMO	91 <sup>b</sup>
FTPT3	NMO	94 <sup>a</sup>
FTPT3	ZMO	88 <sup>a</sup>
FTPT3	CIMO	97 <sup>a</sup>
FTPT3	QMO	87 <sup>a</sup>
FTPT3	CNMO	94 <sup>a</sup>

<sup>a</sup>PCR Conditions: 100 pg D8 template amplified for 16 cycles with Taq polymerase under thermocycling conditions: initial denaturation at 96 °C for 1 min, 96 °C for 30 s, 60 °C for 15 s, 68 °C for 60 s. <sup>b</sup>PCR Conditions: 10 pg D6 template amplified for 24 cycles with OneTaq polymerase under thermocycling conditions: initial denaturation at 96 °C for 1 min, 96 °C for 5 s, 60 °C for 5 s, 68 °C for 10 s.

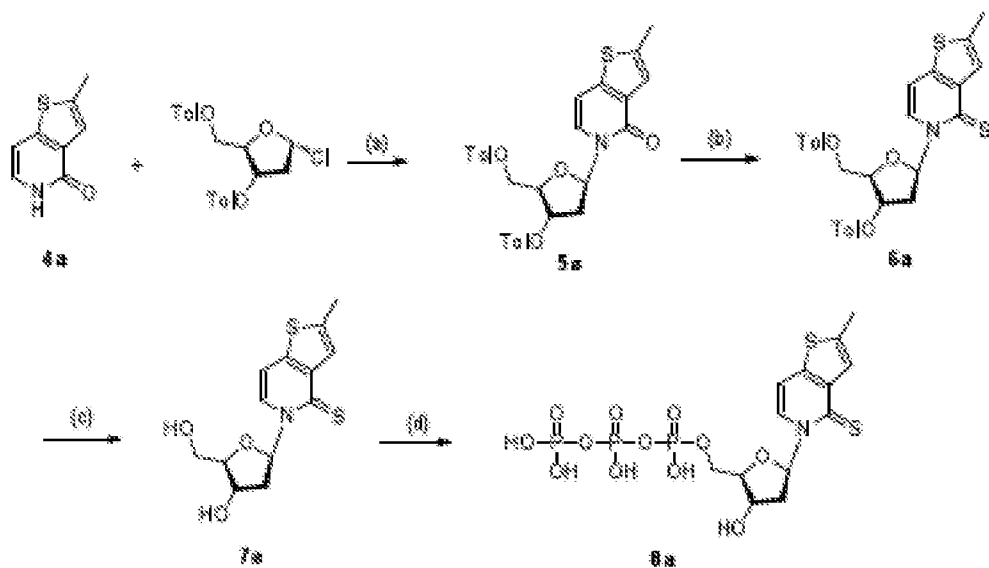
***Example 2. General procedure for triphosphate synthesis.***

[00175] Proton sponge (1.3 equiv) and the free nucleoside derivative (1.0 equiv) were dissolved in dry trimethyl phosphate (40 equiv) and cooled to -15 °C under nitrogen atmosphere. Freshly distilled POCl<sub>3</sub> (1.3 equiv) was added dropwise and the resulting mixture was stirred at -10 °C for 2 h. Tributylamine (6.0 equiv) and a solution of tributylammonium pyrophosphate (5.0 eq.) in dimethylformamide (0.5 M) were added. Over 30 min, the reaction was allowed to warm slowly to 0 °C and then was quenched by addition of 0.5 M aqueous Et<sub>3</sub>NH<sub>2</sub>CO<sub>3</sub> (TEAB) pH 7.5 (2 vol-equiv.). The mixture was diluted two-fold with H<sub>2</sub>O and the product was isolated on a DEAE Sephadex column (GE Healthcare) with an elution gradient of 0 to 1.2 M TEAB, evaporated, and co-distilled with H<sub>2</sub>O (3×). Additional purification by reverse-phase (C18) HPLC (0 - 35% CH<sub>3</sub>CN in 0.1 M TEAB, pH 7.5) was performed, (10% – 31% yield).



Scheme S1. (a) Piperidine, Py, 100 °C, 12 h, then reflux for 1 h; (b) SOCl<sub>2</sub>, DMF, CHCl<sub>3</sub>, reflux, 3 h, (c) NaN<sub>3</sub>, 1,4-dioxane, H<sub>2</sub>O, 5 °C, 0.5 h; (d) diphenyl ether, 230 °C, 1 h.

[00176] The nucleobase analogs 4a, 4b, 4c and 4d were synthesized based on literature methods<sup>1,2</sup> as shown in Scheme S1. Briefly, condensation of the aldehyde (1a–d) with malonic acid at 100 °C in pyridine as a solvent and piperidine as a catalyst for 12 h, followed by a reflux for 1 h, yielded the corresponding acrylic acid intermediates (2a–d). Chlorination of these acids with thionyl chloride in chloroform in the presence of DMF afforded the acyl chlorides, which were not purified but could be used directly in the preparation of the azides (3a–d). Compounds 3a–d were prepared in a biphasic mixture of 1,4-dioxane and water at 5 °C with sodium azide. Crude mixtures of 3a–d in CHCl<sub>3</sub> solutions were added portion-wise to diphenyl ether and heated to 230 °C to give the isocyanates that underwent subsequent intramolecular cyclization to the fused 6-5 bicyclic systems 4a–d.



Scheme S2. (a) N,O-bis(TMS)acetamide, SnCl<sub>4</sub> (1 M in CH<sub>2</sub>Cl<sub>2</sub>), CH<sub>2</sub>Cl<sub>2</sub>, 3 h, 45%; (b)

Lawesson's reagent, toluene, reflux, overnight, 58%; (c) 30% NaOMe, MeOH, rt, 1 h, 85%; (d) Proton sponge, POCl<sub>3</sub>, Bu<sub>3</sub>N, Bu<sub>3</sub>NPPi, (MeO)<sub>3</sub>P, DMF, -20 °C, 31%.

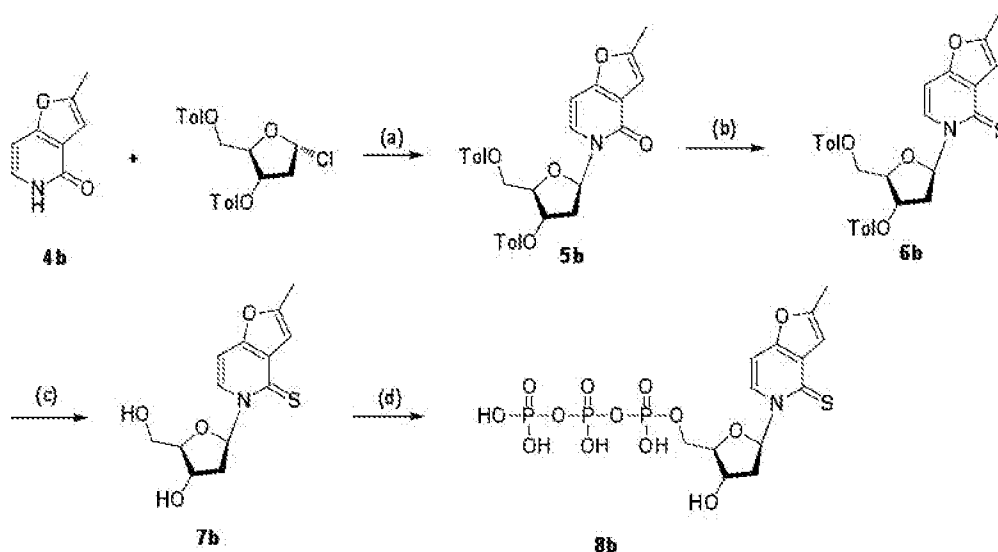
[00177] Compound 5a. To a solution of 4a (54 mg, 0.33 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (83 mg, 0.39 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribosyl chloride (196 mg, 0.39 mmol) was added. The reaction mixture was cooled to 0 °C and SnCl<sub>4</sub> was added dropwise (1.0 M in CH<sub>2</sub>Cl<sub>2</sub>, 160 µL, 0.16 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous NaHCO<sub>3</sub>, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound 5a as white foam (77 mg, 0.15 mmol, 45%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.97-6.82 (m, 11H, Ar-H), 6.44 (d, J = 7.5 Hz, 1H, H-1'), 5.63 (d, J = 6.5 Hz, 1H, H-3'), 4.76-4.68 (m, 2H, H-5'a, 5'b), 4.59 (d, J = 2.5 Hz, H-4'), 2.89 (dd, J = 1.5, 0.5 Hz, 1H, H-2'a), 2.59 (s, 3H, Ar-CH<sub>3</sub>), 2.43 (s, 3H, Ar-CH<sub>3</sub>), 2.43 (s, 3H, Ar-CH<sub>3</sub>), 2.36-2.30 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 166.6, 166.5, 158.5, 147.2, 144.8, 144.6, 140.0, 131.1, 130.3, 130.0, 129.9, 129.7, 127.1, 126.9, 125.6, 122.8, 102.7, 85.9, 83.3, 75.6, 64.8, 39.6, 22.1, 16.1. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>29</sub>H<sub>28</sub>NO<sub>6</sub>S (M+H<sup>+</sup>) 518.1632, found 518.1621.

[00178] Compound 6a. Compound 5a (27 mg, 0.052 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (41.5 mg, 0.10 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound 6a as a yellow foam (16 mg, 0.03 mmol, 58%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.00-7.89 (m, 4H, Ar-H), 7.71 (m, 1H, Ar-H), 7.49-7.48 (m, 1H, H-1'), 7.29-7.21 (m, 4H, Ar-H), 7.65-7.62 (m, 1H, Ar-H), 6.90 (d, J = 7.5 Hz, 1H, Ar-H), 5.64-5.62 (m, 1H, H-4'), 4.85-4.74 (m, 2H, H-5'a), 4.68-4.67 (m, 1H, H-5'b), 3.38-3.34 (m, 1H, H-3'), 2.26 (s, 3H), 2.44 (s, 3H), 2.41 (s, 3H), 2.28-2.22 (m, 1H). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 174.6, 166.6, 144.9, 144.8, 142.9, 142.7, 142.3, 130.3, 130.0, 129.7, 127.9, 127.0, 126.9, 126.8, 108.3, 100.0, 91.4, 84.0, 74.9, 64.5, 39.3, 22.2, 22.1, 16.3. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>29</sub>H<sub>28</sub>NO<sub>5</sub>S<sub>2</sub> (M+H<sup>+</sup>) 534.1403, found 534.1404.

[00179] Compound 7a. To a solution of 6a (20 mg, 0.037 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 7a as yellow foam (9.2 mg, 0.031 mmol, 85%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 8.36 (d, J = 4 Hz, 1H, Ar-H), 7.58 (d, J = 1 Hz, 1H, Ar-H), 7.35 (t, J = 4 Hz, 1H, H-1'), 7.22 (d, J = 8

Hz, 1H, Ar-H), 4.07-4.06 (m, 1H, H-4'), 4.07 (d,  $J = 4$  Hz, 1H, H-3'), 3.80 (dd,  $J = 24, 4$  Hz, 2H, H-5'a, b), 2.79-2.76 (m, 1H, H-2'a), 2.13-2.08 (m, 1H, H-2'b).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  173.60, 143.29, 142.26, 142.23, 129.35, 126.11, 108.01, 91.14, 88.44, 70.37, 61.35, 41.59, 14.81. HRMS (ESI+)  $m/z$  calcd for  $\text{C}_{13}\text{H}_{16}\text{NO}_3\text{S}_2$  ( $\text{M}+\text{H}^+$ ) 298.0566, found 298.0569.

[00180] Compound 8a (dTPT1TP). Compound 8a (11.2 mg, 20.8  $\mu\text{mol}$ , 31%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from 7a (20 mg, 67.3  $\mu\text{mol}$ ).  $^{31}\text{P}$  NMR (162 MHz,  $\text{D}_2\text{O}$ )  $\delta$  -10.3 (d,  $J = 19.8$  Hz,  $\gamma\text{-P}$ ), -10.9 (d,  $J = 20.1$  Hz,  $\alpha\text{-P}$ ), -22.8 (t,  $J = 19.4$  Hz,  $\beta\text{-P}$ ). MS (MALDI-TOF-, matrix: 9-aminoacridine) ( $m/z$ ): [ $\text{M}-\text{H}$ ]-calcd for  $\text{C}_{13}\text{H}_{17}\text{NO}_{12}\text{P}_3\text{S}_2$ , 536.3, found, 536.7.



Scheme S3. (a) N,O-bis(TMS)acetamide,  $\text{SnCl}_4$  (1 M in  $\text{CH}_2\text{Cl}_2$ ),  $\text{CH}_2\text{Cl}_2$ , 3 h, 41%; (b) Lawesson's reagent, toluene, reflux, overnight, 52%; (c) 30% NaOMe, MeOH, rt, 1 h, 85%; (d) Proton sponge,  $\text{POCl}_3$ ,  $\text{Bu}_3\text{N}$ ,  $\text{Bu}_3\text{NPPI}$ ,  $(\text{MeO})_3\text{P}$ , DMF,  $-20^\circ\text{C}$ , 21%.

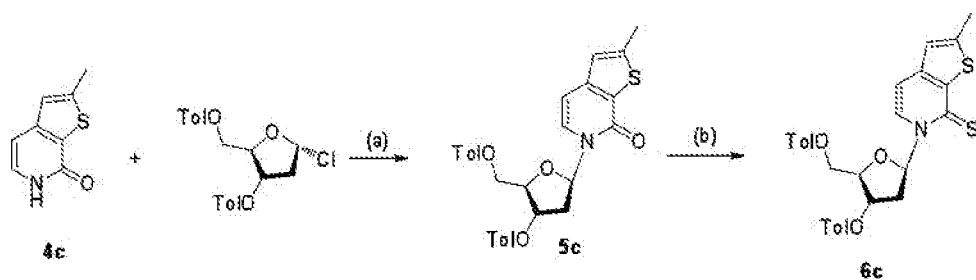
[00181] Compound 5b. To a solution of 4b (100 mg, 0.67 mmol) in  $\text{CH}_2\text{Cl}_2$  (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (165 mg, 0.81 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribose-1-phosphate (292 mg, 0.81 mmol) was added. The reaction mixture was cooled to  $0^\circ\text{C}$  and  $\text{SnCl}_4$  was added dropwise (1.0 M in  $\text{CH}_2\text{Cl}_2$ , 200  $\mu\text{L}$ , 0.2 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous  $\text{NaHCO}_3$ , extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound 5b as white foam (137 mg, 0.27 mmol, 41%).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.99 (d,  $J = 8.1$  Hz, 2H, Ar-H), 7.93 (d,  $J = 8.1$  Hz, 2H, Ar-H), 7.55 (d,  $J = 7.7$  Hz, 1H, Ar-H), 7.33 - 7.28 (m, 2H, Ar-H), 7.27 - 7.20 (m, 2H, Ar-H), 6.82 (dd,  $J = 8.3, 5.6$  Hz, 1H, Ar-H), 6.57 (d,  $J = 0.9$  Hz, 1H, Ar-H), 6.41 (d,  $J = 7.7$  Hz, 1H, H-1'), 5.68 - 5.61 (m, 1H, H-4'), 4.75 (dd,  $J = 12.1, 3.4$  Hz, 2H, H-5'a, b), 4.62 (q,  $J = 3.1$  Hz, 1H, H-3'), 2.94 (ddd,  $J = 14.3, 5.6, 1.7$  Hz, 1H, H-2'a), 2.48 - 2.39 (s, 3x3H, Ar-CH<sub>3</sub>), 2.36 - 2.26

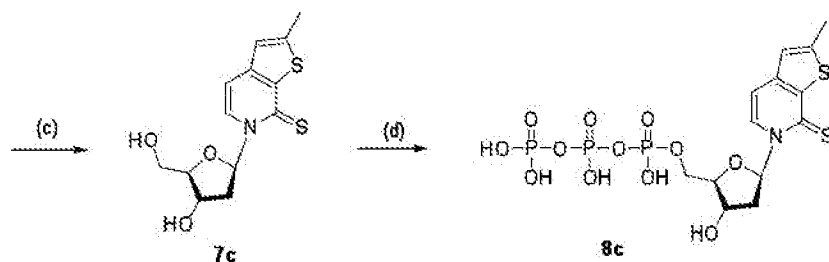
(m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 166.6, 166.5, 159.2, 159.1, 154.5, 144.8, 144.6, 130.2, 130.0, 129.7, 127.5, 127.1, 126.9, 117.5, 103.3, 96.6, 86.0, 83.2, 75.5, 64.7, 39.8, 22.1, 14.1. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>29</sub>H<sub>28</sub>NO<sub>7</sub> (M+H<sup>+</sup>) 502.1860, found 502.1885.

**[00182]** Compound 6b. Compound 5b (29 mg, 0.056 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (41.5 mg, 0.10 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound 6b as a yellow foam (15 mg, 0.029 mmol, 52%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.10-7.89 (m, 5H, Ar-H), 7.52-7.48 (m, 1H, H-1'), 7.29-7.22 (m, 4H, Ar-H), 6.8 (d, J = 1 Hz, 1H, Ar-H), 6.73 (d, J = 7.5 Hz, 1H, Ar-H), 5.65-5.62 (m, 1H, H-4'), 4.84-4.74 (m, 2H, H-5'a, b), 4.67-4.65 (m, 1H, H-3'), 3.36-3.32 (m, 1H, H-2'a), 2.44 (s, 3H, Ar-CH<sub>3</sub>), 2.43 (s, 3H, s, 3H, Ar-CH<sub>3</sub>), 2.41 (s, 3H, s, 3H, Ar-CH<sub>3</sub>), 2.27-2.21 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 166.6, 156.9, 153.9, 144.8, 130.3, 130.0, 129.8, 129.7, 127.9, 106.4, 96.0, 83.9, 56.6, 39.5, 22.1, 12.6. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>29</sub>H<sub>28</sub>NO<sub>6</sub>S (M+H<sup>+</sup>) 518.1632, found 518.1638.

**[00183]** Compound 7b. To a solution of 6b (20 mg, 0.039 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 7b as yellow foam (9.3 mg, 0.033 mmol, 85%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 8.57 (d, J = 5 Hz, 1H, Ar-H), 7.42 (t, J = 4 Hz, 1H, H-1'), 7.13 (d, J = 7.5 Hz, 1H, Ar-H), 6.80 (s, 1H, Ar-H), 4.50-4.47 (m, 1H, H-4'), 4.12 (d, J = 3.5 Hz, 1H, H-3'), 3.95 (dd, J = 30, 3 Hz, 2H, H-5'a, b), 2.81-2.77 (m, 1H, H-2'a), 2.50 (s, 3H, Ar-CH<sub>3</sub>), 2.18-2.14 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD) δ 172.9, 157.2, 154.5, 132.7, 131.3, 105.4, 100.8, 90.9, 88.5, 70.3, 61.3, 41.8, 12.7. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>13</sub>H<sub>16</sub>NO<sub>4</sub>S (M+H<sup>+</sup>) 282.0795, found 282.0790.

**[00184]** Compound 8b (dFPT1TP). Compound 8b (3.7 mg, 7.1 μmol, 10%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from 7b (20 mg, 71.2 μmol). <sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O) δ -10.4 (d, J = 20.0 Hz, γ-P), -10.9 (d, J = 19.4 Hz, α-P), -22.8 (t, J = 20.0 Hz, β<sub>3</sub>-P). MS (MALDI-TOF<sup>-</sup>, matrix: 9-aminoacridine) (m/z): [M-H]<sup>-</sup> calcd for C<sub>13</sub>H<sub>17</sub>NO<sub>13</sub>P<sub>3</sub>S, 520.3, found, 520.1.





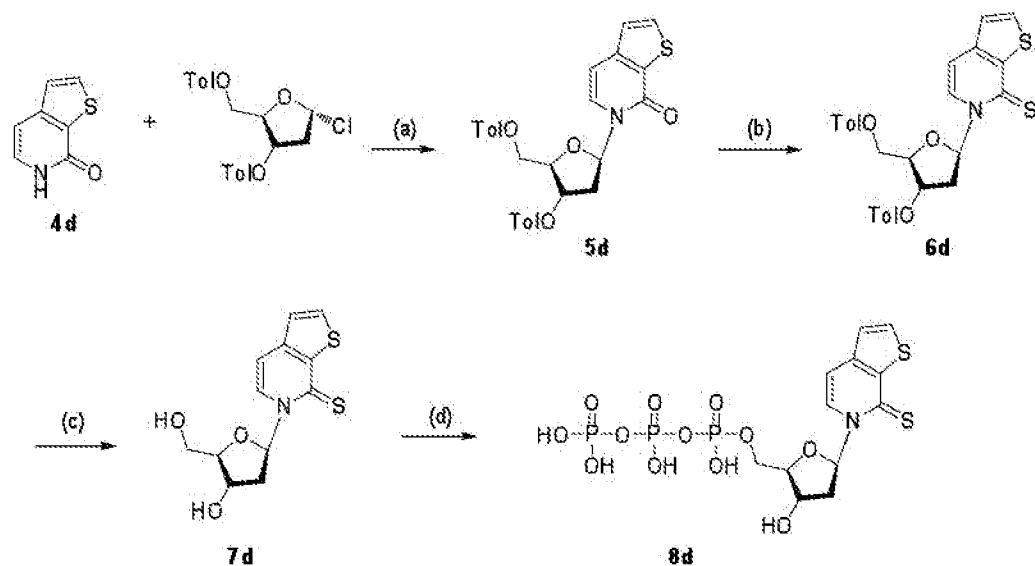
Scheme S4. (a) N,O-bis(TMS)acetamide,  $\text{SnCl}_4$  (1 M in  $\text{CH}_2\text{Cl}_2$ ),  $\text{CH}_2\text{Cl}_2$ , 3 h, 40%; (b) Lawesson's reagent, toluene, reflux, overnight, 31%; (c) 30% NaOMe, MeOH, rt, 1 h, 81%; (d) Proton sponge,  $\text{POCl}_3$ ,  $\text{Bu}_3\text{N}$ ,  $\text{Bu}_3\text{NPPi}$ ,  $(\text{MeO})_3\text{P}$ , DMF,  $-20^\circ\text{C}$ , 15%.

**[00185]** Compound 5c. To a solution of 4c (46 mg, 0.28 mmol) in  $\text{CH}_2\text{Cl}_2$  (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (66 mg, 0.33 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribose chloride (120 mg, 0.33 mmol) was added. The reaction mixture was cooled to  $0^\circ\text{C}$  and  $\text{SnCl}_4$  was added dropwise (1.0 M in  $\text{CH}_2\text{Cl}_2$ , 140  $\mu\text{L}$ , 0.14 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous  $\text{NaHCO}_3$ , extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound 5c as white foam (58 mg, 0.11 mmol, 40%).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  7.98-7.90 (m, 4H, Ar-H), 7.53 (d,  $J = 7.4$  Hz, 1H, Ar-H), 7.27- 7.21 (m, 4H, Ar-H), 6.83-6.82 (m, 2H, Ar-H), 6.44 (d,  $J = 7.5$  Hz, 1H, H-1'), 5.63(d,  $J = 6.5$  Hz, 1H, H-4'), 4.76 - 4.60 (m, 2H, H-5'a, b), 4.59 (d,  $J = 2.5$  Hz, 1H, H-3'), 2.89 (dd,  $J = 13, 5.5$  Hz, H-2'a), 2.59 (s, 3H, Ar-CH<sub>3</sub>), 2.43 (s, 3H, Ar-CH<sub>3</sub>), 2.40 (s, 3H, Ar-CH<sub>3</sub>), 2.37-2.30 (m, 1H, H-2'b).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  166.5, 158.0, 149.9, 146.3, 144.8, 144.6, 130.3, 130.0, 129.7, 128.8, 127.3, 122.8, 103.7, 100.0, 85.8, 83.2, 75.5, 64.8, 39.5, 22.1, 16.7. HRMS (ESI<sup>+</sup>)  $m/z$  calcd for  $\text{C}_{29}\text{H}_{28}\text{NO}_6\text{S}$  ( $\text{M}+\text{H}^+$ ) 518.1632, found 518.1631.

**[00186]** Compound 6c. Compound 5c (50 mg, 0.097 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1.5 mL). Lawesson's reagent (83 mg, 0.20 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound 6c as a yellow foam (16 mg, 0.03 mmol, 31%).  $^1\text{H}$  NMR (500 MHz,  $\text{CDCl}_3$ )  $\delta$  8.13-7.97 (m, 5H, Ar-H), 7.52-7.49 (m, 1H, H-1'), 7.37-7.29 (m, 4H, Ar-H), 6.99 (d,  $J = 1$  Hz, 1H, Ar-H), 6.91 (d,  $J = 7.5$  Hz, 1H, Ar-H), 5.73-5.71 (m, 1H, H-4'), 4.91-4.82 (m, 2H, H-5'a, b), 4.76-4.74 (m, 1H, H-3'), 3.41-3.37 (m, 1H, H-2'a), 2.68 (s, 3H, Ar-CH<sub>3</sub>), 2.52 (s, 3H, Ar-CH<sub>3</sub>), 2.49 (s, 3H, Ar-CH<sub>3</sub>), 2.39-2.34 (m, 1H, H-2'b).  $^{13}\text{C}$  NMR (125 MHz,  $\text{CDCl}_3$ )  $\delta$  172.1, 166.6, 154.0, 144.9, 144.7, 140.4, 130.3, 130.0, 129.7, 129.6, 127.0, 126.8, 122.7, 109.0, 91.2, 83.9, 75.0, 64.5, 39.4, 22.2, 22.1, 17.0. HRMS (ESI<sup>+</sup>)  $m/z$  calcd for  $\text{C}_{29}\text{H}_{28}\text{NO}_5\text{S}_2$  ( $\text{M}+\text{H}^+$ ) 534.1403, found 534.1406.

**[00187]** Compound 7c. To a solution of 6c (20 mg, 0.037 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 7c as yellow foam (8.9 mg, 0.03 mmol, 81%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$  8.48, (d, J = 7.5 Hz, 1H, Ar-H), 7.42 (t, J = 5 Hz, 1H, H-1'), 7.20 (d, J = 5 Hz, 1 H, Ar-H), 7.12 (s, 1H, Ar-H), 4.51-4.48 (m, 1H, H-4'), 4.13 (d, J = 5 Hz, 1H, H-3'), 3.95 (dd, J = 30, 5 Hz, 2H, H-5'a, b), 2.81-2.78 (m, 1H, H-2'a), 2.67 (s, 3H, Ar-CH<sub>3</sub>), 2.21-2.16 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD)  $\delta$  171.1, 154.0, 144.1, 141.1, 131.1, 122.7, 108.8, 90.9, 88.5, 70.5, 61.4, 41.7, 15.4. HRMS (ESI+) m/z calcd for C<sub>13</sub>H<sub>16</sub>NO<sub>3</sub>S<sub>2</sub> (M+H<sup>+</sup>) 298.0566, found 298.0566.

**[00188]** Compound 8c. Compound 8c (10.8 mg, 20.2  $\mu$ mol, 30%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from 7c (20 mg, 67.3  $\mu$ mol). <sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O)  $\delta$  -10.8 (d, J = 19.8 Hz, y-P), -11.5 (d, J = 20.1 Hz, a-P), -23.3 (t, J = 20.1 Hz, f3-P). MS (MALDI-TOF-, matrix: 9-aminoacridine) (m/z): [M-H]<sup>-</sup> calcd for C<sub>13</sub>H<sub>17</sub>NO<sub>12</sub>P<sub>3</sub>S<sub>2</sub>, 536.3, found, 536.1



Scheme S5. (a) N,O-bis(TMS)acetamide, SnCl<sub>4</sub> (1 M in CH<sub>2</sub>Cl<sub>2</sub>), CH<sub>2</sub>Cl<sub>2</sub>, 3 h, 39%; (b) Lawesson's reagent, toluene, reflux, overnight, 33%; (c) 30% NaOMe, MeOH, rt, 1 h, 82%; (d) Proton sponge, POCl<sub>3</sub>, Bu<sub>3</sub>N, Bu<sub>3</sub>NPPi, (MeO)<sub>3</sub>P, DMF, -20 °C, 30%.

**[00189]** Compound 5d. To a solution of 4d (200 mg, 1.32 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (8 mL) at room temperature under nitrogen atmosphere was added bis(trimethylsilyl)acetamide (298 mg, 1.46 mmol). After stirring for 40 min, 3,5-bis(toluoyl)-2-deoxyribose-1-phosphate (563 mg, 1.46 mmol) was added. The reaction mixture was cooled to 0 °C and SnCl<sub>4</sub> was added dropwise (1.0 M in CH<sub>2</sub>Cl<sub>2</sub>, 660  $\mu$ L, 0.66 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous NaHCO<sub>3</sub>, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel

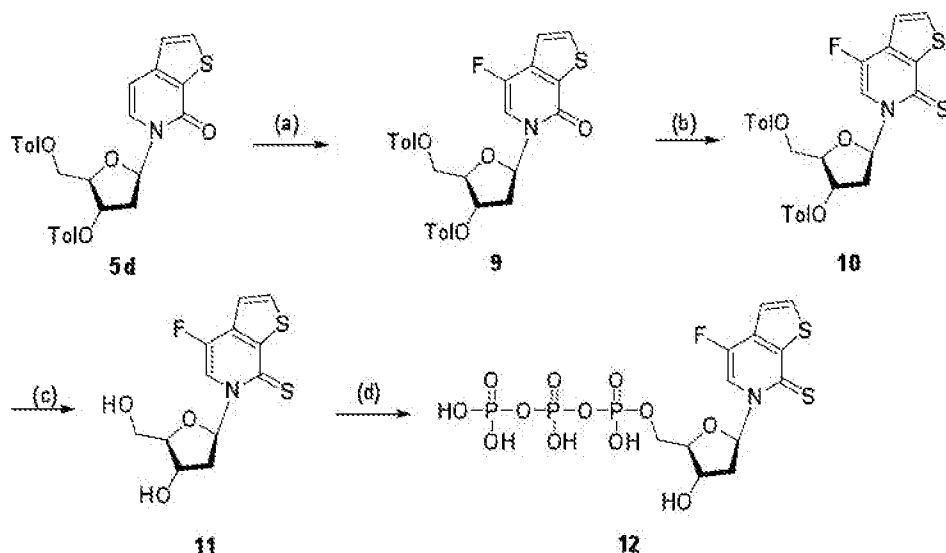
column chromatography (Hexane/EtOAc) to afford compound 5d as a white foam (260 mg, 0.52 mmol, 39%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.98-7.90 (m, 4H, Ar-H), 7.70 (d, J = 6 Hz, 1H, Ar-H), 7.55 (d, J = 9.5 Hz, 1H, Ar-H), 7.28-7.16 (m, 5H, Ar-H), 6.84-6.85 (m, 1H, Ar-H), 6.57 (d, J = 9.5 Hz, H-1'), 5.66-5.64 (m, 1H, H-4'), 4.75-4.72 (m, 2H, H-5'a, b), 4.61 (m, 1H, H-3'), 2.95-2.90 (m, 1H, H-2'a), 2.43 (s, 3H, Ar-CH<sub>3</sub>), 2.40 (s, 3H, Ar-CH<sub>3</sub>), 2.39-2.31 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>)  $\delta$  166.2, 166.1, 158.1, 145.1, 144.4, 144.2, 133.8, 129.9, 129.6, 129.3, 129.1, 126.9, 126.5, 124.2, 103.5, 85.5, 82.9, 75.1, 64.4, 39.2, 21.7. HRMS (ESI+) m/z calcd for C<sub>20</sub>H<sub>20</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>5</sub>S (M+H<sup>+</sup>) 504.1475, found 504.1480.

[00190] Compound 6d. Compound 5d (50 mg, 0.1 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (48 mg, 0.12 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated, and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound 6d as a yellow foam (17 mg, 0.033 mmol, 33%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  8.14-7.82 (m, 7H, Ar-H), 7.51 (dd, J = 7.5, 6.0 Hz, 1H, H-1'), 7.32-7.23 (m, 5H, Ar-H), 6.99 (d, J = 7.2 Hz, 1H, Ar-H), 5.73 (m, 1H, H-4'), 4.92-4.83 (m, 2H, H-5'a, b), 4.78-4.77 (m, 1H, H-3'), 3.43-3.40 (m, 1H, H-2'a), 2.51 (s, 3H, Ar-CH<sub>3</sub>), 2.48 (s, 3H, Ar-CH<sub>3</sub>), 2.39-2.36 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>)  $\delta$  173.5, 166.6, 144.9, 144.8, 139.5, 138.0, 134.5, 130.3, 130.0, 129.7, 129.5, 126.8, 124.7, 109.5, 91.4, 84.0, 75.0, 64.5, 39.4, 22.2, 22.1. HRMS (ESI+) m/z calcd for C<sub>28</sub>H<sub>26</sub>NO<sub>5</sub>S<sub>2</sub> (M+H<sup>+</sup>) 520.1247, found 520.1241.

[00191] Compound 7d. To a solution of 6d (20 mg, 0.039 mmol) in methanol (1.0 mL) was added dropwise 30% NaOMe (8.66 mg, 0.16 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 7d as yellow foam (9.0 mg, 0.032 mmol, 82%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$  8.48 (d, J = 5 Hz, 1H, Ar-H), 8.01 (d, J = 5 Hz, 1H, Ar-H), 7.40-7.38 (m, 2H, Ar-H), 7.29 (d, J = 10 Hz, 1H, H-1'), 4.47-4.46 (m, 1H, H-4'), 4.10 (m, 1H, H-3'), 3.94-3.88 (m, 2H, H-5'a, b), 2.77-2.76 (m, 1H, H-2'a), 2.19-2.14 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD)  $\delta$  171.2, 144.7, 139.6, 137.6, 130.5, 124.2, 108.8, 90.7, 88.2, 70.1, 61.0, 41.3. HRMS (ESI+) m/z calcd for C<sub>12</sub>H<sub>14</sub>NO<sub>3</sub>S<sub>2</sub> (M+H<sup>+</sup>) 284.041, found 284.0410.

[00192] Compound 8d (dTPT3TP). Compound 8d (5.7 mg, 10.9  $\mu$ mol, 31%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from 7d (10 mg, 35.3  $\mu$ mol). <sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O)  $\delta$  -9.3 (d, J = 19.5 Hz,  $\gamma$ -P), -10.8 (d, J = 19.8 Hz,  $\alpha$ -P), -22.4 (t, J = 20.0 Hz,  $\beta$ -P). MS (MALDI-TOF-, matrix: 9-aminoacridine) (m/z): [M-H]<sup>-</sup> calcd for C<sub>12</sub>H<sub>15</sub>NO<sub>12</sub>P<sub>3</sub>S<sub>2</sub><sup>-</sup>, 521.9, found, 521.9.





Scheme S6. (a) i. Selectfluor, MeOH/CH<sub>3</sub>CN, reflux, 3 h; ii. TfOH-CH<sub>2</sub>Cl<sub>2</sub> (1:1 v/v), 1 h, 85%; (b) Lawesson's reagent, toluene, reflux, overnight, 32%; (c) 30% NaOMe, MeOH, rt, 1 h, 85%; (d) Proton sponge, POCl<sub>3</sub>, Bu<sub>3</sub>N, Bu<sub>3</sub>NPPI, (MeO)<sub>3</sub>P, DMF, -20 °C, 10%.

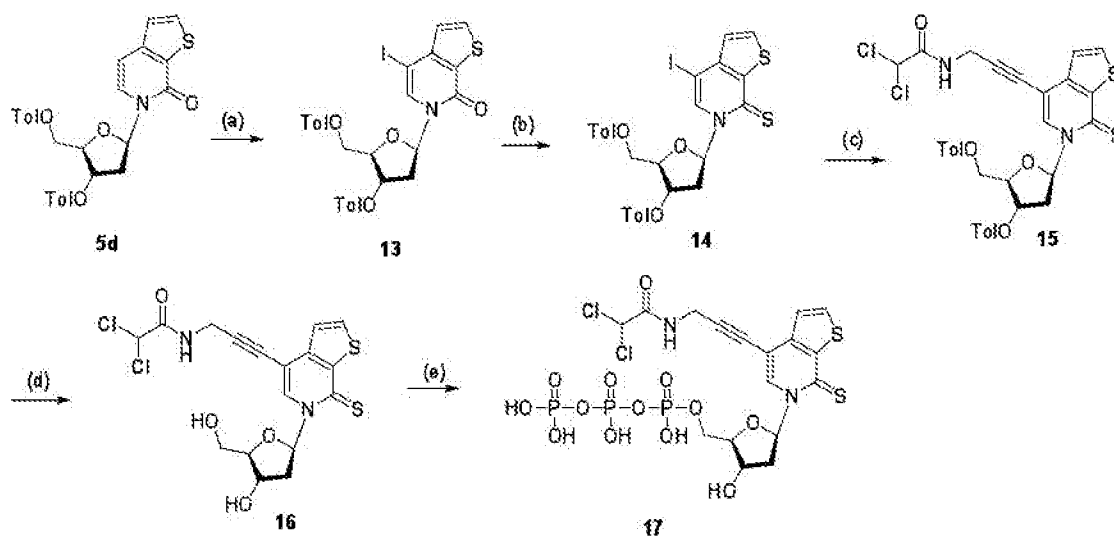
**[00193]** Compound 9. Compound 5d (55 mg, 0.11 mmol) was dissolved in 1.0 mL MeOH-CH<sub>3</sub>CN (1:1 v/v), Selectfluor (42 mg, 0.12 mmol) was added and the mixture was heated at reflux for 3 h, then the solvent was evaporated, the residue was dissolved in EtOAc (20 mL), the organic phase was washed with water three times. Then the organic solvent was evaporated, and the solid residue was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in 1 mL TfOH-CH<sub>2</sub>Cl<sub>2</sub> (1:1 v/v) and the mixture was stirred at room temperature for 1 h, then the mixture was concentrated, and the crude product was subjected to a silica gel column chromatography (hexane/EtOAc) to afford compound 9 as a white solid (49 mg, 0.093 mmol, 85%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  7.98-7.92 (m, 4H, Ar-H), 7.75 (d, J = 5 Hz, 1H, Ar-H), 7.52 (d, J = 7.5 Hz, 1H, Ar-H), 7.32-7.21 (m, 5H, Ar-H), 6.82-6.78 (m, 1H, H-1'), 5.64-5.61 (m, 1H, H-4'), 4.80-4.59 (m, 2H, H-5'a, b), 4.62-4.59 (m, 1H, H-3'), 2.93- 2.87 (m, 1H, H-2'a), 2.43 (s, 3H, Ar-CH<sub>3</sub>), 2.39 (s, 3H, Ar-CH<sub>3</sub>), 2.34-2.27 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>)  $\delta$  166.2, 166.1, 156.4, 144.5, 144.3, 137.7, 137.5, 134.6, 129.9, 129.6, 129.3, 126.6, 126.4, 120.2, 112.1, 111.7, 85.5, 83.1, 75.0, 64.2, 39.1, 21.8, 21.7. <sup>19</sup>F NMR (376 MHz, CDCl<sub>3</sub>)  $\delta$  -151.5. HRMS (ESI+) m/z calcd for C<sub>28</sub>H<sub>25</sub>FNO<sub>6</sub>S (M+H<sup>+</sup>) 522.1381, found 522.1380.

**[00194]** Compound 10. Compound 9 (20 mg, 0.038 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL). Lawesson's reagent (18.5 mg, 0.046 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound 10 as a yellow foam (6.5 mg, 0.012 mmol, 32%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>)  $\delta$  8.11-7.85 (m, 6H, Ar-H), 7.40-7.39(m, 2H,

Ar-H, H-1'), 7.28-7.21 (m, 4H, Ar-H), 5.64-5.63 (m, 1H, H-4'), 4.83 (m, 2H, H-5'a, b), 4.69 (m, 1H, H-3'), 3.34-3.29 (m, 1H, H-2'a), 2.44 (s, 1H, Ar-CH<sub>3</sub>), 2.40 (s, 3H, Ar-CH<sub>3</sub>), 2.30-2.26 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>)  $\delta$  170.9, 166.6, 166.5, 144.9, 144.8, 138.6, 130.3, 130.0, 129.9, 129.7, 129.7, 126.9, 126.7, 120.5, 116.3, 116.0, 100.0, 91.6, 84.3, 74.7, 64.3, 39.2, 22.2, 22.1. <sup>19</sup>F NMR (376 MHz, CDCl<sub>3</sub>)  $\delta$  -142.9. HRMS (ESI<sup>+</sup>) *m/z* calcd for C<sub>28</sub>H<sub>25</sub>FNO<sub>5</sub>S<sub>2</sub> (M+H<sup>+</sup>) 538.1153, found 538.1155.

**[00195]** Compound 11. To a solution of 10 (10 mg, 0.019 mmol) in methanol (1.5 mL) was added dropwise 30% NaOMe (4.33 mg, 0.08 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was then concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 11 as yellow foam (4.9 mg, 0.016 mmol, 85%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$  8.68 (d, *J* = 5 Hz, 1H, Ar-H), 8.12 (d, *J* = 5 Hz, 1H, Ar-H), 7.52 (d, *J* = 5 Hz, 1H, Ar-H), 7.28 (t, *J* = 6.5 Hz, 1H, H-1'), 4.48 (m, 1H, H-4'), 4.10 (m, 1H, H-3'), 3.94 (dd, *J* = 35, 3 Hz, 2H, H-5'a, b), 2.78-2.75 (m, 1H, H-2'a), 2.24-2.19 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD)  $\delta$  170.2, 150.2, 148.3, 139.0, 131.7, 131.6, 119.8, 117.8, 117.4, 91.5, 88.7, 70.1, 61.0, 41.5. <sup>19</sup>F NMR (376 MHz, CD<sub>3</sub>OD)  $\delta$  -145.3. HRMS (ESI<sup>+</sup>) *m/z* calcd for C<sub>12</sub>H<sub>13</sub>FNO<sub>3</sub>S<sub>2</sub> (M+H<sup>+</sup>) 302.0315, found 302.0314.

**[00196]** Compound 12 (dFTPT3TP). Compound 12 (2.0 mg, 3.7  $\mu$ mol, 22%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from 11 (5 mg, 16.6  $\mu$ mol). <sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O)  $\delta$  -10.9 (d, *J* = 20.0 Hz,  $\gamma$ -P), -11.6 (d, *J* = 21.1 Hz,  $\alpha$ -P), -23.3 (t, *J* = 23.1 Hz,  $\beta$ -P). <sup>19</sup>F NMR (376 MHz, D<sub>2</sub>O)  $\delta$  -138.5 (s). MS (MALDI-TOF-, matrix: 9-aminoacridine) (*m/z*): [M-H]<sup>-</sup> calcd for C<sub>12</sub>H<sub>14</sub>FNO<sub>12</sub>P<sub>3</sub>S<sub>2</sub><sup>-</sup>, 539.9, found, 540.1.



Scheme S7. (a) ICl, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C to rt, overnight, 63%; (b) Lawesson's reagent, toluene, reflux, overnight, 27%; (c) 2,2-dichloro-N-prop-2-yn-1-ylacetamide, (PPh<sub>3</sub>)<sub>4</sub>Pd, CuI, Et<sub>3</sub>N, DMF, rt, overnight, 27%; (d) TBAF, THF, rt, overnight, 100%; (e) (PPh<sub>3</sub>)<sub>4</sub>Pd, CuI, Et<sub>3</sub>N, DMF, rt, overnight, 22%.

overnight, 91%; (d) 30% NaOMe, MeOH, rt, 1 h, 74%; (e) Proton sponge, POCl<sub>3</sub>, Bu<sub>3</sub>N, Bu<sub>3</sub>NPPI, (MeO)<sub>3</sub>P, DMF, -20 °C, 25%.

**[00197]** Compound 13. To a solution of 5d (73 mg, 0.145 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (1 mL) at 0 °C under nitrogen atmosphere was added dropwise iodine monochloride (1.0 M in CH<sub>2</sub>Cl<sub>2</sub>, 0.15 mL, 0.15 mmol). The resulting mixture was stirred at room temperature overnight. The reaction mixture was quenched with saturated aqueous NaHCO<sub>3</sub> and saturated aqueous Na<sub>2</sub>S<sub>2</sub>O<sub>3</sub>, extracted with CH<sub>2</sub>Cl<sub>2</sub>, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (Hexane/EtOAc) to afford compound 13 as white foam (57 mg, 0.091 mmol, 63%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 7.98-7.93 (m, 4H, Ar-H), 7.83 (s, 1H, Ar-H), 7.72 (d, J = 5 Hz, 1H, Ar-H), 7.28-7.17 (m, 5H, Ar-H), 6.78-6.75 (m, 1H, H-1'), 5.65-5.63 (m, 1H, H-4'), 4.76 (m, 2H, H-5'a, b), 4.63-4.62 (m, 1H, H-3'), 2.95-2.91 (m, 1H, H-2'a), 2.43 (s, 3H, Ar-CH<sub>3</sub>), 2.35 (s, 3H, Ar-CH<sub>3</sub>), 2.34-2.29 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 166.6, 166.5, 157.6, 147.4, 144.9, 144.6, 133.7, 132.7, 130.3, 130.1, 129.8, 129.7, 129.4, 128.5, 127.0, 126.8, 86.1, 83.8, 75.7, 64.7, 39.9, 22.2, 22.1. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>28</sub>H<sub>25</sub>INO<sub>5</sub>S (M+H<sup>+</sup>) 630.0442, found 630.0440.

**[00198]** Compound 14. Compound 13 (30 mg, 0.048 mmol) was dried by 3 co-evaporations with anhydrous toluene. The residue was dissolved in anhydrous toluene (1 mL), Lawesson's reagent (23 mg, 0.057 mmol) was added and the mixture was heated overnight at reflux. After filtration on cotton, the filtrate was concentrated and the crude product was subjected to a silica gel column chromatography (Hexane/EtOAc) to afford compound 14 as a yellow foam (8.4 mg, 0.013 mmol, 27%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.31 (s, 1H, Ar-H), 7.99-7.82 (m, 5H, Ar-H), 7.39-7.36 (m, 1H, H-1'), 7.29-7.20 (m, 5H, Ar-H), 5.65-5.64 (m, 1H, H-4'), 4.83-4.81 (m, 2H, H-5'a, b), 4.71-4.70 (m, 1H, H-3'), 3.35 (dd, J = 15, 5.5 Hz, 1H, H-2'a), 2.44 (s, 3H, Ar-CH<sub>3</sub>), 2.39 (s, 3H, Ar-CH<sub>3</sub>), 2.27-2.21 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 172.9, 166.6, 144.9, 144.7, 144.6, 141.8, 137.8, 135.0, 130.3, 130.2, 129.8, 129.7, 128.6, 126.9, 126.7, 91.7, 84.5, 75.3, 64.6, 39.5, 22.2, 22.1. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>28</sub>H<sub>25</sub>INO<sub>5</sub>S<sub>2</sub> (M+H<sup>+</sup>) 646.0213, found 646.0219.

**[00199]** Compound 15. To a solution of 14 (10 mg, 0.015 mmol) in DMF (2 mL) under nitrogen atmosphere was added (PPh<sub>3</sub>)<sub>4</sub>Pd (1.7 mg, 0.0015 mmol), CuI (0.57 mg, 0.011 mmol) and Et<sub>3</sub>N (5 μL, 0.030 mmol). The reaction mixture was degassed and a solution of Cl<sub>2</sub>CHCONHCH<sub>2</sub>CCH (3.8 mg, 0.0225 mmol) in DMF (0.5 mL) was added. The reaction mixture was stirred overnight at room temperature and monitored by TLC. The reaction mixture was diluted with EtOAc, quenched with saturated aqueous NaHCO<sub>3</sub>, extracted with EtOAc, dried, filtered and evaporated. The crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 15 as yellow foam (9.2 mg, 0.0135

mmol, 91%). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ 8.26 (s, 1H, Ar-H), 7.99-7.82 (m, 5H, Ar-H), 7.40-7.37 (m, 2H, Ar-H, H-1'), 7.29-7.21 (m, 4 H, Ar-H), 6.71 (br, 1H, NH), 6.95 (s, 1H, CHCl<sub>2</sub>), 5.65-5.64 (m, 1H, H-4'), 4.85-4.79 (m, 2H, H-5'a, b), 4.73 (m, 1H, H-3'), 4.26-4.11 (m, 2H, NHCH<sub>2</sub>), 3.38-3.34 (m, 1H, H-2'a), 2.44 (s, 3H, Ar-CH<sub>3</sub>), 2.40 (s, 3H, Ar-CH<sub>3</sub>), 2.31-2.25 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CDCl<sub>3</sub>) δ 173.3, 166.6, 164.1, 144.8, 139.0, 138.3, 133.4, 130.3, 130.1, 129.8, 129.7, 126.9, 124.3, 104.8, 91.8, 88.3, 84.5, 78.8, 75.2, 66.4, 64.7, 39.6, 31.3, 22.2. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>33</sub>H<sub>29</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>6</sub>S<sub>2</sub> (M+H<sup>+</sup>) 683.0839, found 683.0854.

**[00200]** Compound 16. To a solution of 15 (9.2 mg, 0.0135 mmol) in methanol (1.0 ml) was added dropwise 30% NaOMe (2.92 mg, 0.32 mmol). The reaction mixture was stirred for 1 h at room temperature and monitored by TLC. The reaction mixture was concentrated and the crude product was subjected to silica gel column chromatography (MeOH/CH<sub>2</sub>Cl<sub>2</sub>) to afford compound 16 as yellow foam (4.5 mg, 0.01 mmol, 74%). <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD) δ 8.69 (s, 1H, Ar-H), 8.06 (d, J = 5 Hz, 1H, Ar-H), 7.53 (d, J = 5 Hz, 1H, Ar-H), 7.30 (t, J = 5 Hz, 1H, H-1'), 6.33 (s, 1H, CHCl<sub>2</sub>), 4.47-4.46 (m, 1H, H-4'), 4.36 (s, 2H, NHCH<sub>2</sub>), 4.11-4.08 (m, 1H, H-3'), 3.97 (dd, J = 12, 3Hz, 2H, H-5'a, b), 2.79-2.74 (m, 1H, H-2'a), 2.21-2.16 (m, 1H, H-2'b). <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD) δ 172.9, 138.4, 134.3, 123.9, 122.8, 104.7, 100.0, 91.2, 88.6, 77.1, 70.3, 66.4, 61.1, 41.7, 30.2. HRMS (ESI<sup>+</sup>) m/z calcd for C<sub>17</sub>H<sub>17</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>4</sub>S<sub>2</sub> (M+H<sup>+</sup>) 447.0001, found 447.0020.

**[00201]** Compound 17 (dTPT3<sup>PA</sup>TP). Compound 17 (2.2 mg, 3.1 μmol, 28%) was synthesized using the General Procedure for Triphosphate Synthesis described above starting from 16 (5 mg, 11.2 μmol). <sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O) δ -10.85 (d, J = 19.9 Hz, γ-P), -11.63 (d, J = 20.0 Hz, α-P), -23.07 (s), -23.26 (t, J = 19.7 Hz, β<sub>3</sub>-P). MS (MALDI-TOF<sup>-</sup>, matrix: 9-aminoacridine) (m/z): [M-H]<sup>-</sup> calcd for C<sub>17</sub>H<sub>18</sub>Cl<sub>2</sub>N<sub>2</sub>O<sub>13</sub>P<sub>3</sub>S<sub>2</sub><sup>-</sup>, 684.9, found, 685.0.

**Example 3. General procedure for PCR amplification assay to determine fidelity.**

**[00202]** Materials. Taq and OneTaq DNA polymerases were purchased from New England Biolabs (Ipswich, MA). A mixture of dNTPs was purchased from Fermentas (Glen Burnie, MD). SYBR Green I Nucleic Acid Gel Stain (10,000×) was purchased from Life Technologies (Carlsbad, CA).

**[00203]** DNA oligonucleotides. Complete oligonucleotide sequences are provided in Table 8. Fully natural primers were purchased from Integrated DNA Technologies (Coralville, Iowa). Reagents and solvents for synthesis of unnatural primers 1-3 were obtained from Glen Research (Sterling, VA) and/or Applied Biosystems (Foster City, CA). The oligonucleotides were prepared using standard automated DNA synthesis with ultra-mild natural phosphoramidites (Glen Research) and dNaM phosphoramidite (Berry & Associates, Inc., Dexter, MI) on

controlled pore glass supports (0.20  $\mu\text{mol}$ , 1000 Å, Glen Research) and an ABI Expedite 8905 synthesizer. After automated synthesis, the oligonucleotides were cleaved from the support, deprotected by incubation in conc. aqueous ammonia overnight at room temperature, purified by DMT purification (glen-pak™ cartridge, Glen Research), and desalted over Sephadex G-25 (NAP-25 Columns, GE Healthcare). The concentration of single stranded oligonucleotides was determined by UV absorption at 260 nm.

**[00204]** PCR assay. PCR amplifications were performed in a total volume of 25  $\mu\text{L}$  and with conditions specific for each assay as described in Table 9. After amplification, a 5  $\mu\text{L}$  aliquot was analyzed on a 6% non-denaturing PAGE gel ran along with 50bp ladder (Life Technologies) to confirm amplicon size. The remaining solution was purified by spin-column (DNA Clean and Concentrator-5; Zymo Research, Irvine, CA), followed by 4% agarose gel, recovered with Zymoclean Gel DNA Recovery Kit (Zymo Research), quantified by fluorescent dye binding (Quant-iT dsDNA HS Assay kit, Life Technologies), and sequenced on a 3730 DNA Analyzer (Applied Biosystems). Fidelity was determined as the average %retention of the unnatural base pair per doubling as described below.

**[00205]** Determination of fidelity. The percent retention of an unnatural base pair (F) was measured using raw sequencing data and normalized to fidelities per doubling. Briefly, the presence of an unnatural nucleotide leads to a sharp termination of the sequencing profile, while mutation to a natural nucleotide results in “readthrough”. The extent of the “read-through” is thus inversely correlated with the retention of the unnatural base pair. To use the sequencing data as a quantitative measurement of PCR fidelity, we performed calibration experiments in the range of 50-100% retention of the unnatural base pair. Therefore, low retention (<50%) and high “read-through” make the quantification inaccurate.

**[00206]** Quantification of the high retention (>50%) was performed by adjusting the start and stop points for the Sequencing Analysis software (Applied Biosystems) and then determining the average signal intensity individually for each channel (A, C, G and T) for peaks within those defined points (35-45 nucleotides in length) before (section L) and S36 after (section R) the unnatural nucleotide. The R/L ratio was normalized using sequencing calibration plots to account for both noise in the sequencing chromatograms and the read-through in the control samples. The R/L ratio of after normalization (R/L<sub>norm</sub>) corresponds to the percentage of the natural sequences in the pool. Finally, F was calculated as  $1 - (R/L)_{\text{norm}}$  and the retention of the unnatural base pair per doubling (fidelity, f) was calculated as  $1/(F^{\log_2 A})$ , where A is an amplification and  $\log_2 A$  is the number of doublings. Each sample before and PCR amplification was sequenced in triplicate in each direction to minimize sequencing error. Corresponding data is provided in Table 10. Under standard PCR conditions, DNA containing dTPT3-dNaM was

amplified by OneTaq with an efficiency that is only 4-fold lower than that of DNA containing just the natural base pairs, and with a fidelity in excess of 99.98%. This fidelity corresponds to an error rate of  $10^{-4}$  per nucleotide, which overlaps with the  $10^{-4}$  to  $10^{-7}$  error rate of fully natural DNA with commonly used PCR systems. With Taq polymerase, the efficiency is only 2.5-fold lower than that of a natural base pair, and the fidelity is 99.7%. This fidelity corresponds to an error rate of  $10^{-3}$ , which is similar to that observed with the Taq-mediated amplification of natural DNA.

Table 8. DNA sequences.

Name	Sequence (5' to 3') Primer regions underlined	Remarks
Fend1	<u>CACACAGGAAACAGCTATGAC</u>	Primers for PCR (templates D6 and 134mer)
Fend2	<u>GAAATTAATACGACTCACTATAGG</u>	
Fend1-poly-dT	TT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT <u>CACACAG</u> GAAACAGCTATGAC	Primers for Sanger sequencing (templates D6 and 134mer)
Fend2-poly-dT	TT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT <u>GAAAT</u> TAATACGACTCACTATAGG	
D6	<u>CACACAGGAAACAGCTATGACCCGGGTATTAC</u> ATGCGCTAGCACTTGGAATTCACCAGACGNNN <b>NaM</b> NNNCGGGACCCATAGTAAATCTCCTTCTT AAAGTTAAGCTTAAC <u>CCCTATAGTGAGTCGTATT</u> AATTTC	N = randomized natural nucleotide
134mer	<u>CACACAGGAAACAGCTATGACCCGGGTATTAC</u> ATGCGCTAGCACTTGGAATTCACAATACT <b>NaM</b> TCTTTAAGGAAACCATAGTAAATCTCCTTCTT AAAGTTAAGCTTAAC <u>CCCTATAGTGAGTCGTATT</u> AATTTC	
Primer1	<b>NaM</b> <u>CCTGCGTCAATGTAATGTTC</u>	Primers for PCR with Temp1-3
Primer2	<u>TTCACGGT</u> <b>NaM</b> <u>AGCACGCATAGG</u>	
Primer3	<u>CCAATGTACC</u> <b>NaM</b> <u>TGCGTATGTTC</u>	
Primer-rev	<u>CCCTGCGTTTATCTGCTCTC</u>	
Temp1	<u>CCCTGCGTTTATCTGCTCTCTCGGTCGTTTCGGC</u> TGCGGCGGAACATTACATTGACGCAGG	The nucleotides shown in bold form a mispair with dNaM in the first round of PCR
Temp2	<u>CCCTGCGTTTATCTGCTCTCTCGGTCGTTTCGGC</u> TGCGGCGCTATGCGTGCTTACCGTGAA	
Temp3	<u>CCCTGCGTTTATCTGCTCTCTCGGTCGTTTCGGC</u> TGCCGGAACATACGCATGGTACATTGG	

Table 9. PCR Conditions.

	OneTaq	Taq	PCR for biotin labeling (unnatural base pair centrally located)	PCR for biotin labeling (unnatural base pair positioned at 1, 9, 11 positions)
Buffer	1x OneTaq	1x Taq	1x OneTaq	1x OneTaq
Enzyme, U/ $\mu$ L	OneTaq, 0.02	Taq, 0.02	OneTaq, 0.02	OneTaq, 0.02
Template	D6 (0.01 ng)	D6 (0.01 ng)	134mer (0.5 ng)	60mer (0.5 ng)
dNTPs, $\mu$ M	200	200	200	200
dNaMTP, $\mu$ M	100	100	100	100
dXTP, $\mu$ M	100	100	100 of d5SICS <sup>PA</sup> TP or dTPT3 <sup>PA</sup> TP	100 of d5SICS <sup>PA</sup> TP or dTPT3 <sup>PA</sup> TP
Mg <sup>2+</sup> , mM	3	3	3	3
Primers, $\mu$ M	1	1	1	1
SYBR Green I	0.5x	0.5x	0.5x	0.5x
Thermal conditions				
Initial denaturing	-	-	96 °C, 1 min	96 °C, 1 min
Denaturing	96 °C, 10s	96 °C, 10s	96 °C, 15s	96 °C, 15s
Annealing	60 °C, 15s	60 °C, 15s	60 °C, 30s	64 °C, 30s
Extension	68 °C, 60s	68 °C, 15s	68 °C, 2 min	68 °C, 2 min
# cycles	16+16+16	20	12	12

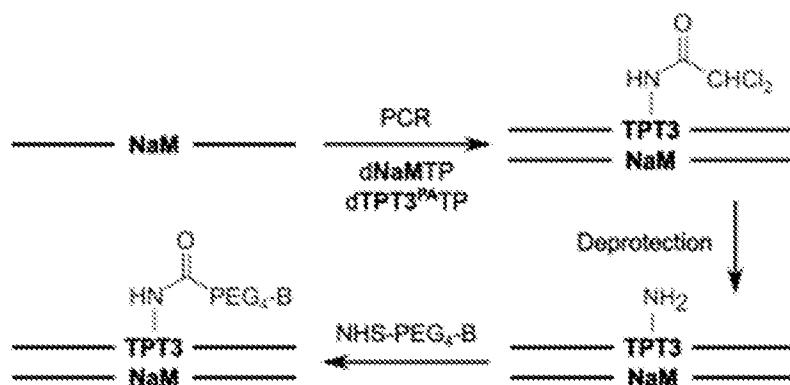
Table 10.

	OneTaq PCR (48 cycles)			Taq PCR (20 cycles)		
dXTP	amplification $\times 10^{12}$	retention, %	fidelity, %	amplification $\times 10^{13}$	retention, %	fidelity, %
5SICS	9.4	96.3 $\pm$ 1.7	99.91 $\pm$ 0.04	7.7	86.7 $\pm$ 1.0	98.90 $\pm$ 0.01
TPT3	12.9	>99	>99.98	11.7	95.6 $\pm$ 1.7	99.66 $\pm$ 0.13
TPT3 <sup>PA</sup>	4.7	98.6 $\pm$ 1.2	99.97 $\pm$ 0.03	3.5	85 $\pm$ 4	98.7 $\pm$ 0.4
5SICS <sup>PA</sup>	9.2	45 $\pm$ 2	98.16 $\pm$ 0.12	6.4	<sup>a</sup> —	<sup>a</sup> —
<sup>a</sup> Unnatural base pair lost during amplification						

**Example 4: Site-specific labeling of TPT3: analysis via streptavidin gel shift assay.**

**[00207]** A 134-mer DNA comprising a centrally positioned dTPT3<sup>PA</sup>-dNaM or d5SICS<sup>PA</sup>-dNaM was synthesized. DNA templates were amplified by PCR under the conditions described in Table 9. Upon completion, NaOH (1 M, 12.5  $\mu$ L) was added directly to PCR samples to a final concentration of 0.2 M and incubated for 5 hr at room temperature. After the addition of NaOAc (3 M, pH 5.5, 7.5  $\mu$ L) and 200  $\mu$ L of cold ethanol, the samples were mixed, incubated on ice overnight, and DNA was precipitated by centrifugation at 10,000 rfu for 30 min at 4 °C. The supernatant was removed and the pellets were carefully washed with 80% ethanol. The

samples were resuspended in 50  $\mu$ L of the annealing buffer (50 mM Na phosphate, pH 7.5, 100 mM NaCl, 1 mM EDTA), heated to 95  $^{\circ}$ C and cooled to room temperature over 30 min. NHS-PEG<sub>4</sub>-biotin (Thermo Scientific) solution in the annealing buffer (40 mM, 50  $\mu$ L) was mixed with the DNA samples and incubated overnight at room temperature. The samples were purified by spin-column (DNA Clean and Concentrator-5, Zymo Research) and eluted in 10  $\mu$ L of elution buffer. Half of the sample (5  $\mu$ L) was mixed with 1  $\mu$ g of streptavidin (Promega) in annealing buffer, incubated for 30 min at 37  $^{\circ}$ C, mixed with 5x non-denaturing loading buffer (Qiagen), and loaded on 6% non-denaturing PAGE. The remaining half was mixed with 5x non-denaturing loading buffer, and loaded directly on the gel as a control. After running the gel at 110 V for 30 min, the gel was soaked in 1x Sybr Gold Nucleic Acid Stain (Life Technologies) for 30 min and visualized using a Molecular Imager Gel Doc XR+ equipped with 520DF30 filter (Bio-Rad). A schematic of the labeling strategy described is shown below.



**Example 5. General procedure for transcription of an unnatural base pair.**

[00208] To characterize the transcription of the unnatural base pairs formed by dTPT3 and dNaM, or analogs or derivatives thereof (wherein derivatives include linker moieties), ribonucleotides and deoxynucleotides are synthesized and converted to the corresponding triphosphates or deoxyphosphoramidites, and the deoxyphosphoramidites are incorporated into DNA templates using automated DNA synthesis. Transcription experiments are conducted with 100 nM DNA substrate, 1x Takara buffer (40 mM Tris-HCl, pH 8.0, 8 mM MgCl<sub>2</sub>, 2 mM spermidine), DEPC-treated and nuclease-free sterilized water (Fisher), T7 polymerase (50 units), 20  $\mu$ M each natural NTP,  $\alpha$ -<sup>32</sup>P-ATP (2.5  $\mu$ Ci, MP Biomedicals), and either 5  $\mu$ M TPT3TP or 5  $\mu$ M NamTP. After incubation for 2 hr at 37  $^{\circ}$ C, the reaction is quenched by the addition of 10  $\mu$ L of gel loading solution (10 M urea, 0.05% bromophenol blue), and the reaction mixture is loaded onto a 20% polyacrylamide-7 M urea gel, subjected to electrophoresis, and analyzed by phosphorimaging. Transcription efficiency is examined by measuring (at low percent conversion) the amount of full-length product formed as a function of time.



**Example 6. General procedure for thermodynamic analysis of a DNA duplex comprising an unnatural base pair.**

[00209] UV melting experiments are carried out using a Cary 300 Bio UV-visible spectrophotometer. The absorbance of a sample (3  $\mu$ L oligonucleotide comprising an unnatural base pair, 10 mM PIPES buffer, pH 7.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub>) is monitored at 260 nm from 21 °C to 80 °C at a heating rate of 0.5 °C per min. Melting temperatures are determined via the derivative method using the Cary Win UV thermal application software.

[00210] Thermodynamic parameters are determined by van't Hoff analysis  $T_m^{-1} = R[\ln([C_T]/4)]\Delta H + \Delta S^\circ/\Delta H^\circ$ , where  $\Delta H^\circ$  and  $\Delta S^\circ$  are the standard enthalpy and entropy changes determined from UV experiments, respectively,  $R$  is the universal gas constant and  $[C_T]$  is the total oligonucleotide strand concentration. The changes in the number of water molecules associated with the melting process,  $\Delta n_w$ , are obtained from the dependence of  $T_m$  on water activity ( $a_w$ ) according to the equation  $\Delta n_w = (-\Delta H/R)[\delta(T_m^{-1})/\delta(\ln a_w)]$ . The slope of the plot of reciprocal temperature (K<sup>-1</sup>) of melting versus the logarithm of water activity at different concentrations (0, 2, 5, 7, 10, 12 and 15% wt%) of ethylene glycol is taken as the value of  $\delta(T_m^{-1})/\delta(\ln a_w)$ .

[00211] CD experiments are performed with an Aviv model 61 DS spectropolarimeter equipped with a Peltier thermoelectric temperature control unit (3  $\mu$ M oligonucleotide concentration, 10 mM PIPES buffer, pH 7.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub>). The data are collected using a 1 cm path length quartz cuvette with scanning from 360 to 220 nm, a time constant of 3s and a wavelength step size of 0.5 nm at 25 °C.

**Example 7. In vitro selection with unnatural nucleobases.**

[00212] An oligonucleotide library comprising unnatural nucleic acids is generated. A sample of the library is subjected to sequential binding and elution from a target molecule, for example, a protein. The pool of binding nucleic acids are amplified by PCR and subjected to another round of selection for binding to the target molecule. This selection process is repeated a number of times. To increase selection pressure, in the last few rounds of selection, the concentration of target molecule and/or incubation time is reduced. Surviving nucleic acids are sequenced as potential aptamers. The binding affinities of potential aptamers are determined using flow-cytometry.

**Example 8. General procedure for DNA Click Reaction.**

[00213] To a DNA solution (0.2 pmol) in 14  $\mu$ L DMSO was added 1  $\mu$ L azide-PEG(3+3)-S-S-Biotin(20 mM in H<sub>2</sub>O), followed by 2  $\mu$ L of ligand (BimC<sub>4</sub>A)<sub>3</sub>(4 mM in H<sub>2</sub>O), 1  $\mu$ L of sodium ascorbate (100 mM in H<sub>2</sub>O), and 1  $\mu$ L of PBS buffer (5 $\times$ ), the mixture was then vortexed and as

the last component, 1  $\mu$ L of a freshly prepared  $\text{CuSO}_4$  solution (4 mM in  $\text{H}_2\text{O}$ ) was added. The solution was shaken for 2 h at 37 °C, and then the resulting product DNA was purified (DNA Clean & Concentrator-5 kit, Zymo Research Corp.). The purified samples were used directly for gel mobility assays (see below).

**Example 9. General procedures for post-amplification DNA labeling (From Seo et al., JACS 2011, 133, 19878).**

[00214] For post-enzymatic synthesis labeling, dsDNA with a free amino group was incubated with 10 mM EZ-Link sulfo-NHS-SS-biotin or EZ-Link NHS-PEG<sub>4</sub>-biotin (Thermo Scientific) for 1 h at rt in phosphate labeling buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1 mM EDTA), and then purified using the Qiagen PCR purification kit. With either dichloroacetyl protected amine derivatives such as dTPT3<sup>PA</sup> or d5SICS<sup>PA</sup>, the amine first required deprotection, which was accomplished by overnight incubation in a concentrated aqueous ammonia solution at rt. Ammonia was removed via a SpeedVac concentrator (water aspirator followed by oil vacuum pump). To cleave the disulfide containing linkers (*i.e.* SS-biotin or SS-PEG<sub>4</sub>-biotin), dsDNA was treated with DTT (final concentration of 30 mM) for 1 hour at 37 °C. For backbone labeling, dsDNA with a backbone phosphorothioate was incubated with 25 mM EZ-Link iodoacetyl-PEG<sub>2</sub>-biotin (Thermo Scientific) in phosphate labeling buffer overnight at 50 °C, and products were purified with Qiagen PCR Purification Kit. All reactions manipulating attached biotin moieties were quantified by streptavidin gel-shift assays.

[00215] *Gel Mobility Assays.* DNA samples (10–50 ng) were mixed with 1  $\mu$ g of streptavidin (Promega) in phosphate labeling buffer (50 mM sodium phosphate, pH 7.5, 150 mM NaCl, 1 mM EDTA), incubated for 30 min at 37°C, mixed with 5 $\times$  nondenaturing loading buffer (Qiagen), and loaded on 6% nondenaturing PAGE. The gel was run at 150 V for 25-40 min, then stained with 1 $\times$  Sybr Gold Nucleic Acid Stain (Life Technologies) in TBE for 30 min and visualized using a Molecular Imager Gel Doc XR+ equipped with 520DF30 filter (Bio-Rad). Strong bands corresponding to dsDNA (at ~150 bp) and the 1:1 complex between dsDNA and streptavidin (at ~400 bp) were apparent. Faint bands corresponding to higher order (slower migrating) complexes of DNA and streptavidin or from unbiotinylated, single-stranded DNA resulting from incomplete annealing after PCR in some cases were also apparent.

[00216] All patents and publications referred to herein are incorporated by reference herein to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference in its entirety.

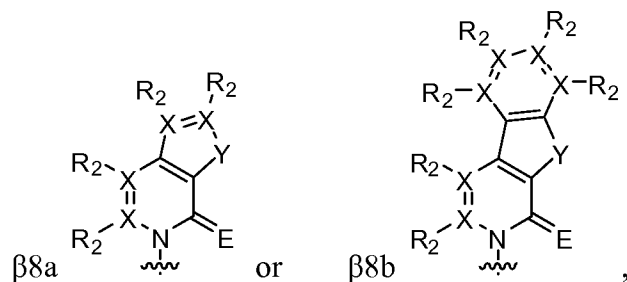
[00217] The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of

excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the embodiments claimed. Thus, it should be understood that although the present embodiments have been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this disclosure as defined by the appended claims.

## CLAIMS

What is claimed is:

1. A compound comprising a nucleobase analog of any of the following formulas:



wherein each X is independently carbon or nitrogen,

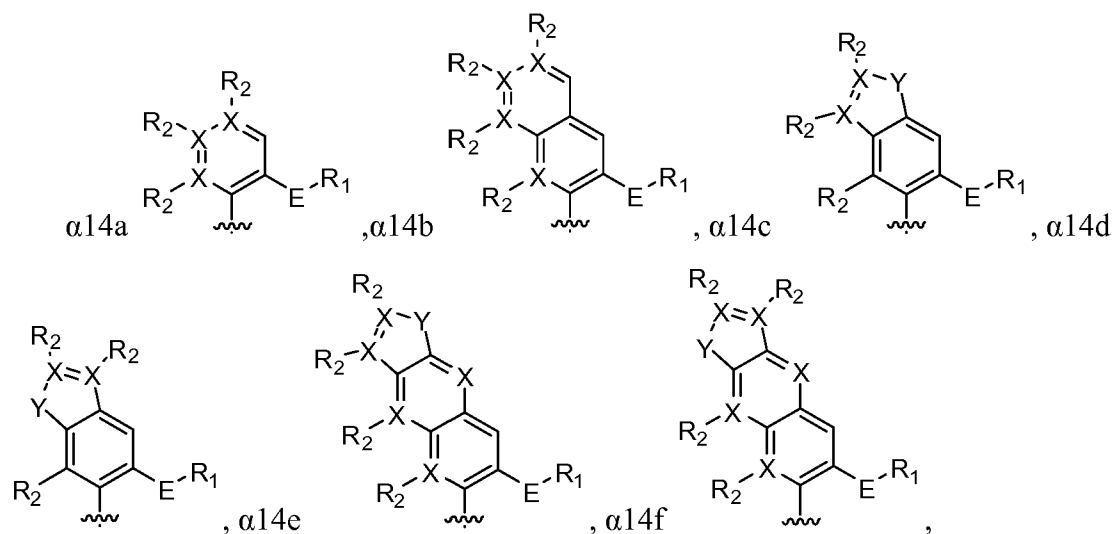
wherein each R<sub>2</sub> is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded,

wherein each Y is independently sulfur, oxygen, selenium, or secondary amine,

wherein each E is independently sulfur or selenium,

and the wavy line indicates a point of bonding to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety or an analog thereof, wherein the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, connected to a mono-phosphate, diphosphate, or triphosphate group, optionally comprising an  $\alpha$ -thiotriphosphate,  $\beta$ -thiophosphate, or  $\gamma$ -thiophosphate group, or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

2. A compound comprising a nucleobase analog of any of the following formulas



wherein each X is independently carbon or nitrogen,

wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded,

wherein each  $R_2$  is optional and when present, is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, coupled linker to which a cargo is bonded,

wherein each Y is independently sulfur, oxygen, selenium, or secondary amine,

wherein each E is independently sulfur or selenium,

and the wavy line indicates a point of bonding to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety or an analog thereof, wherein the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, bound to a mono-phosphate, diphosphate, or triphosphate group, optionally comprising an  $\alpha$ -thiotriphosphate,  $\beta$ -thiophosphate, or  $\gamma$ -thiophosphate group, or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

3. The compound of claim 1, wherein the reactive center comprises an amino group, an acetylenic group, a thiol group, an aldehyde group, or an azide group.

4. The compound of claim 2, wherein the reactive center comprises an amino group, an acetylenic group, a thiol group, an aldehyde group, or an azide group.

5. The compound of claim 1, wherein the ribosyl or deoxyribosyl moiety bears a triphosphate or an  $\alpha$ -thiotriphosphate group bonded to a 5'-hydroxyl thereof.

6. The compound of claim 2, wherein the ribosyl or deoxyribosyl moiety bears a triphosphate or an  $\alpha$ -thiotriphosphate group bonded to a 5'-hydroxyl thereof.

7. The compound of claim 1, wherein the ribosyl or deoxyribosyl moiety is incorporated into a RNA or DNA oligonucleotide chain, respectively, or the ribosyl or deoxyribosyl moiety or analog thereof is incorporated into an RNA or a DNA analog.

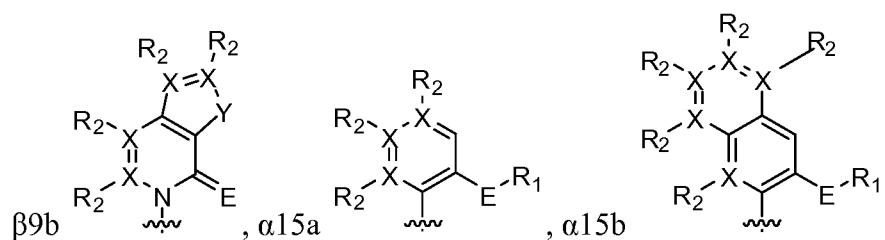
8. The compound of claim 2, wherein the ribosyl or deoxyribosyl moiety is incorporated into a RNA or DNA oligonucleotide chain, respectively, or the ribosyl or deoxyribosyl moiety or analog thereof is incorporated into an RNA or a DNA analog.

9. The compound of claim 1, wherein at least one  $R_2$  independently comprises a group selected from the group consisting of a  $-C\equiv C-CH_2NHR_3$  group, wherein  $R_3$  is hydrogen or is an amino-protecting group; acetylenic group suitable for use in a click reaction with a cargo reagent comprising a cargo and an acetylene-reactive group; thiol group suitable for use in a reaction with a cargo reagent comprising a cargo and a thiol-reactive group; aldehyde group suitable for

use in a reaction with a cargo reagent comprising a cargo and an aldehyde-reactive group; azide group suitable for use in a reaction with a cargo reagent comprising a cargo and an azide-reactive group; and  $-C\equiv C-(CH_2)_n-C\equiv CH$ , wherein  $n$  is 1, 2, 3, 4, 5, or 6; and  $-C\equiv C-(CH_2)_{n1}-O-(CH_2)_{n2}-C\equiv CH$ , wherein  $n1$  and  $n2$  are each independently 1, 2, or 3.

10. The compound of claim 2, wherein at least one  $R_2$  independently comprises a group selected from the group consisting of a  $-C\equiv C-CH_2NHR_3$  group, wherein  $R_3$  is hydrogen or is an amino-protecting group; acetylenic group suitable for use in a click reaction with a cargo reagent comprising a cargo and an acetylene-reactive group; thiol group suitable for use in a reaction with a cargo reagent comprising a cargo and a thiol-reactive group; aldehyde group suitable for use in a reaction with a cargo reagent comprising a cargo and an aldehyde-reactive group; azide group suitable for use in a reaction with a cargo reagent comprising a cargo and an azide-reactive group; and  $-C\equiv C-(CH_2)_n-C\equiv CH$ , wherein  $n$  is 1, 2, 3, 4, 5, or 6; and  $-C\equiv C-(CH_2)_{n1}-O-(CH_2)_{n2}-C\equiv CH$ , wherein  $n1$  and  $n2$  are each independently 1, 2, or 3.

11. A nucleobase pair comprising a first nucleobase analog having the formula  $\beta 9b$ , and a second nucleobase analog having any of the formulas  $\alpha 15a$  or  $\alpha 15b$ :



wherein each  $X$  is independently carbon or nitrogen,

wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded,

wherein each  $R_2$  is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded,

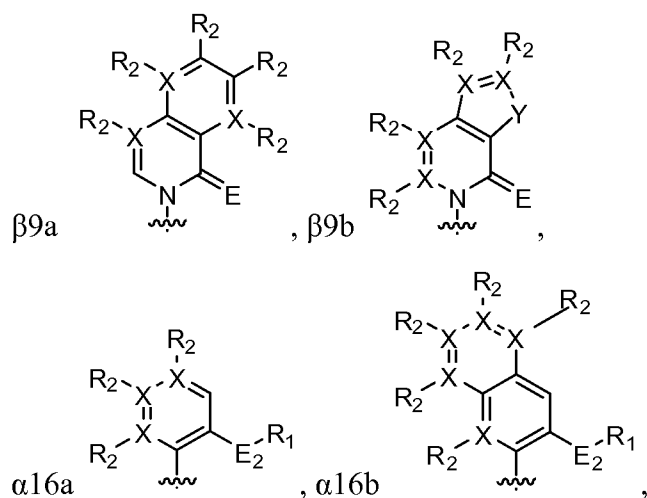
wherein each  $Y$  is independently sulfur, oxygen, selenium, or secondary amine,

wherein each  $E$  is independently oxygen, sulfur or selenium,

and the wavy line indicates a point of bonding to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety or an analog thereof, wherein the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, bound to a mono-phosphate, diphosphate, or triphosphate group, optionally comprising an  $\alpha$ -thiotriphosphate,  $\beta$ -

thiophosphate, or  $\gamma$ -thiophosphate group, or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

12. A nucleobase pair comprising a first nucleobase analog having any of the formulas  $\beta 9a$  or  $\beta 9b$ , and a second nucleobase analog having any of the formulas  $\alpha 16a$  or  $\alpha 16b$ :



wherein each X is independently carbon or nitrogen,

wherein each  $R_1$  is independently hydrogen, alkyl group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded,

wherein each  $R_2$  is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, azide, nitro group, a reactive linker comprising a reactive center adapted to bond to a cargo reagent comprising a cargo and a group of reactivity complementary to the reactive center, or a coupled linker to which a cargo is bonded,

wherein each Y is independently sulfur, oxygen, selenium, or secondary amine,

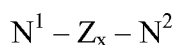
wherein each E is independently oxygen, sulfur or selenium,

wherein each  $E_2$  is independently sulfur or selenium,

and the wavy line indicates a point of bonding to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety or an analog thereof, wherein the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, bound to a mono-phosphate, diphosphate, or triphosphate group, optionally comprising an  $\alpha$ -thiotriphosphate,  $\beta$ -thiophosphate, or  $\gamma$ -thiophosphate group, or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

13. A double stranded oligonucleotide duplex wherein a first oligonucleotide strand comprises a compound of claim 1, and a second complementary oligonucleotide strand comprises a complementary base-pairing nucleobase in a complementary base-pairing site thereof.

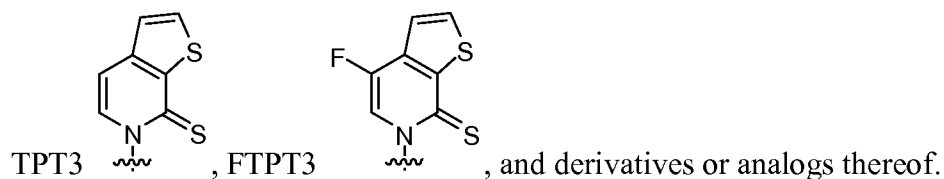
14. A double stranded oligonucleotide duplex wherein a first oligonucleotide strand comprises a compound of claim 2, and a second complementary oligonucleotide strand comprises a complementary base-pairing nucleobase in a complementary base-pairing site thereof.
15. The double stranded oligonucleotide duplex of claim 13, wherein the first strand comprises dTPT3 and the second strand comprises dNaM, dDMO, or dMMO2 in a complementary base-pairing site thereof.
16. A method of carrying out a site-specific functionalization of a double stranded oligonucleotide duplex, comprising: incorporating an unnatural nucleobase comprising a reactive linker comprising a reactive center, the nucleobase being selected from a compound of claim 1, into a first oligonucleotide strand; then, synthesizing a second strand complementary to the first strand, the second strand comprising a nucleobase complementary to the unnatural nucleobase at a site-specific complementary position therein, under conditions such that the first strand and the second strand form a double stranded oligonucleotide duplex; then, contacting the double stranded oligonucleotide duplex incorporating the unnatural nucleobase comprising the reactive linker moiety with a cargo reagent comprising a cargo and a group of complementary reactivity, under conditions suitable for reaction of the reactive linker and the group of complementary reactivity to occur to yield a coupled linker; to provide the functionalized double stranded oligonucleotide duplex with the cargo bonded thereto via a coupled linker.
17. A method of carrying out a site-specific functionalization of a double stranded oligonucleotide duplex, comprising: incorporating an unnatural nucleobase comprising a reactive linker comprising a reactive center, the nucleobase being selected from a compound of claim 2, into a first oligonucleotide strand; then, synthesizing a second strand complementary to the first strand, the second strand comprising a nucleobase complementary to the unnatural nucleobase at a site-specific complementary position therein, under conditions such that the first strand and the second strand form a double stranded oligonucleotide duplex; then, contacting the double stranded oligonucleotide duplex incorporating the unnatural nucleobase comprising the reactive linker moiety with a cargo reagent comprising a cargo and a group of complementary reactivity, under conditions suitable for reaction of the reactive linker and the group of complementary reactivity to occur to yield a coupled linker; to provide the functionalized double stranded oligonucleotide duplex with the cargo bonded thereto via a coupled linker.
18. A structure comprising the formula:





wherein N<sup>1</sup> is a nucleotide or analog thereof, or terminal phosphate group; wherein N<sup>2</sup> is a nucleotide or analog thereof, or terminal hydroxyl group; wherein Z is a compound of claim 1 or claim 2, and wherein x is an integer from 1 to 20.

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



20. The compound of claim 2, selected from the group consisting of:

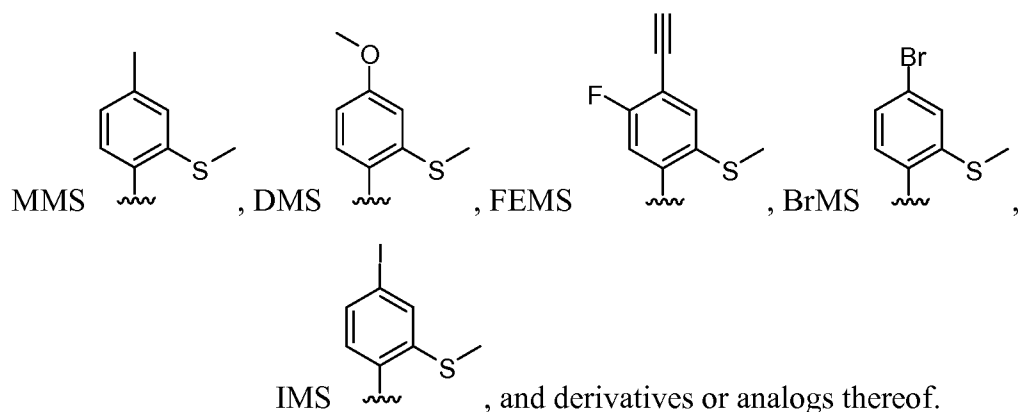


Figure 1

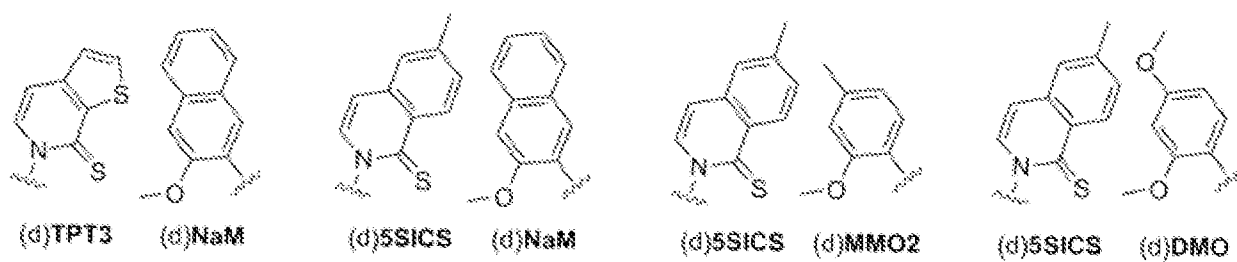


Figure 2

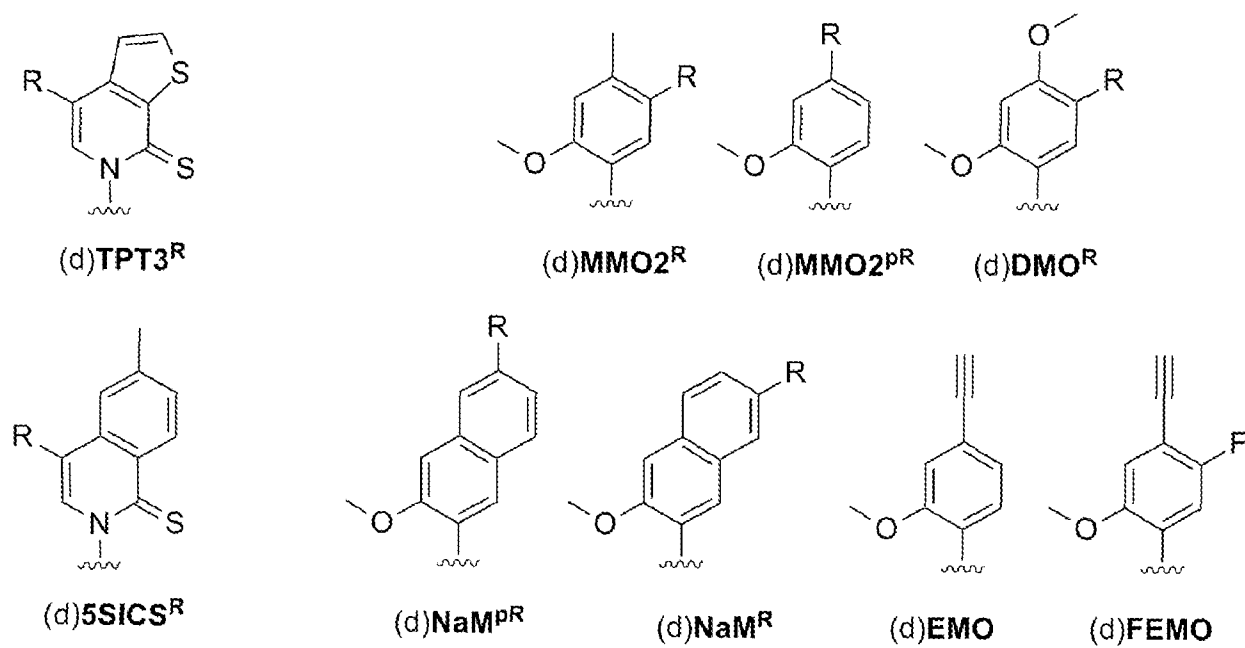


Figure 3

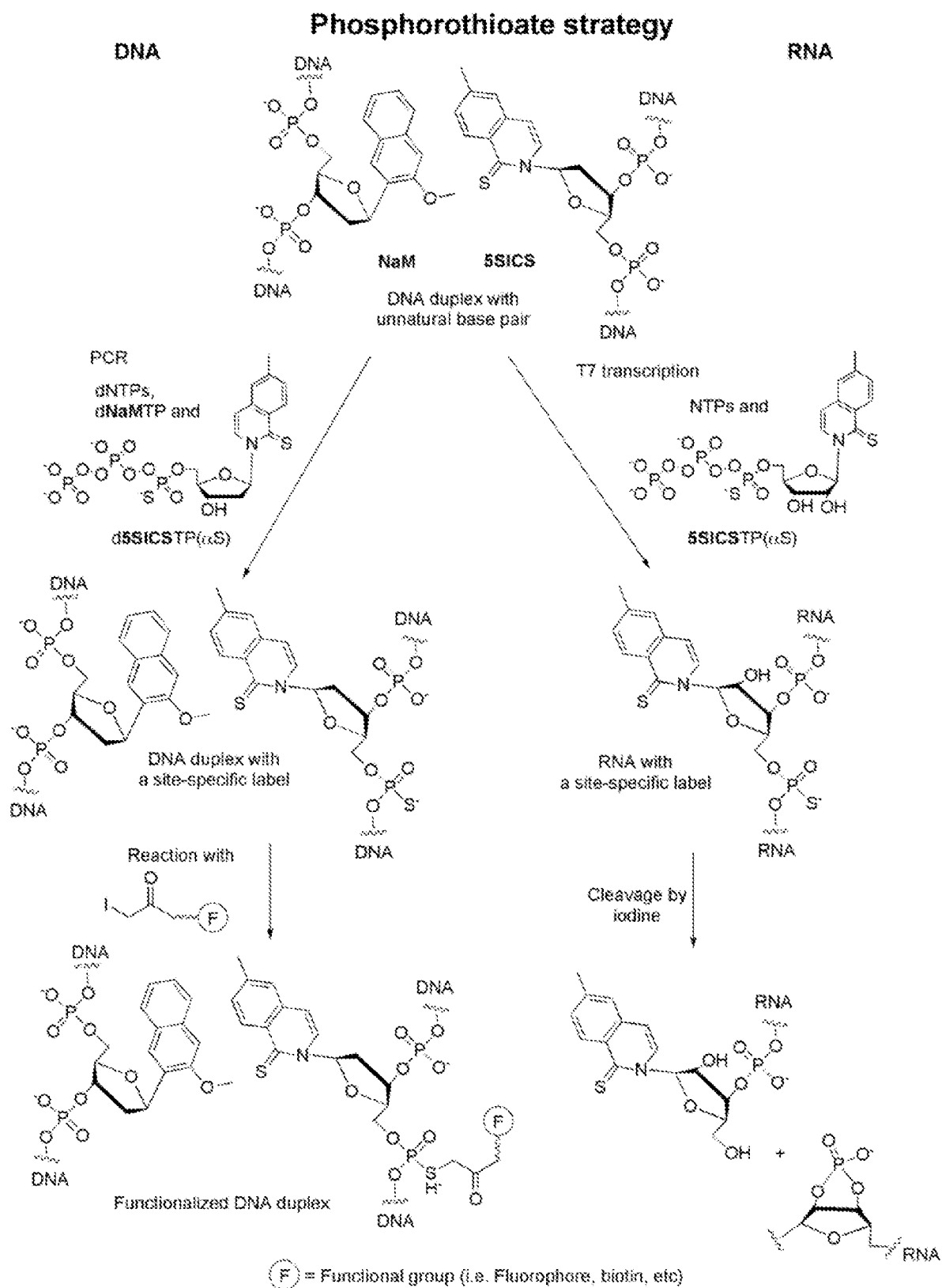


Figure 4

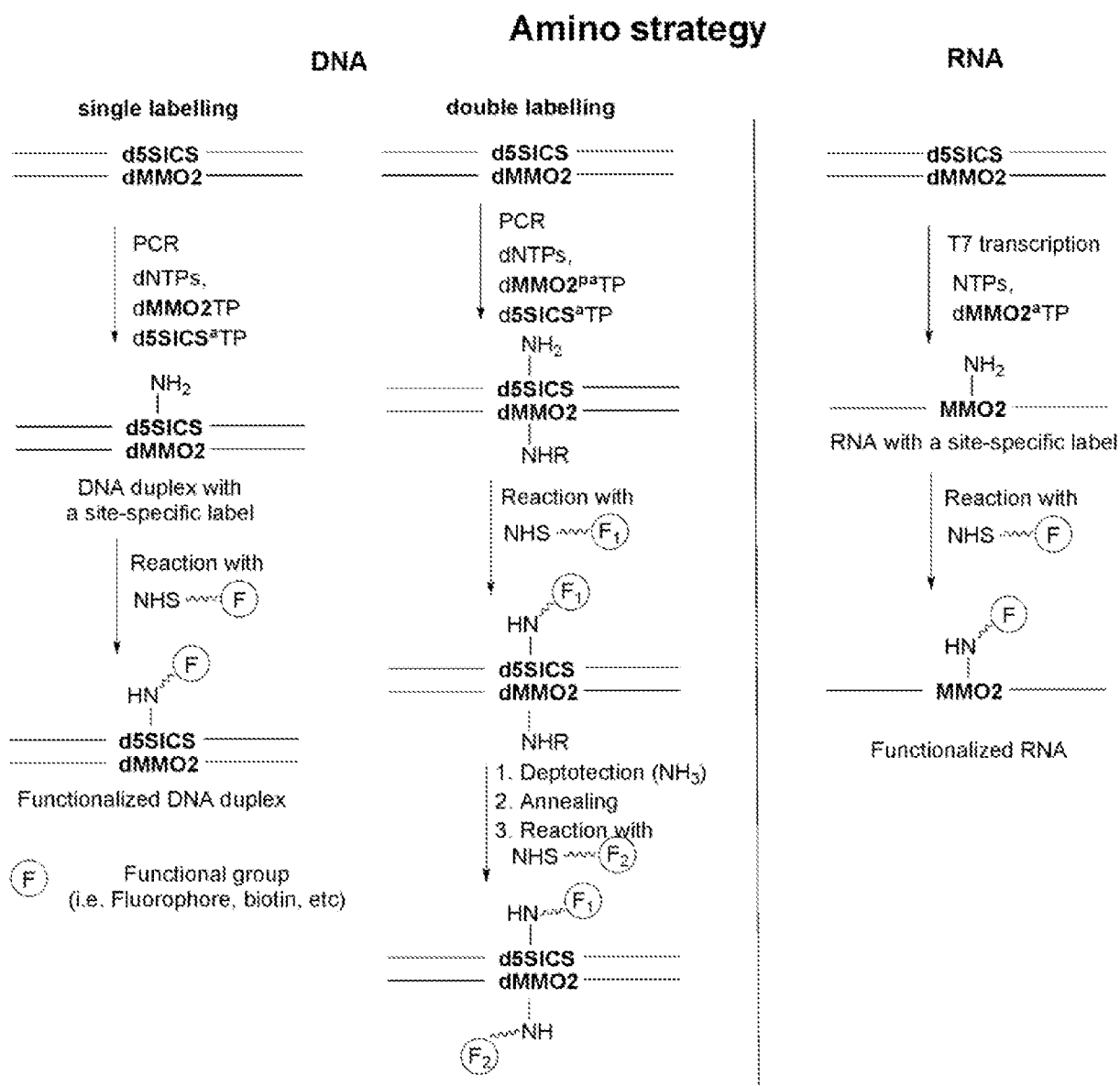


Figure 5

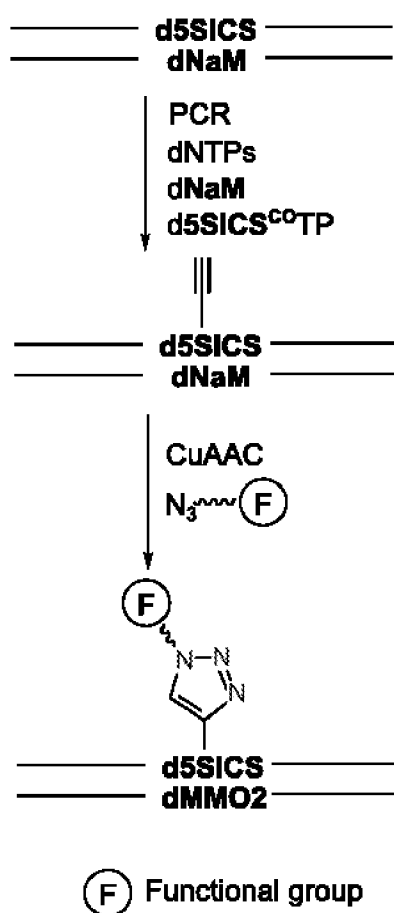


Figure 6

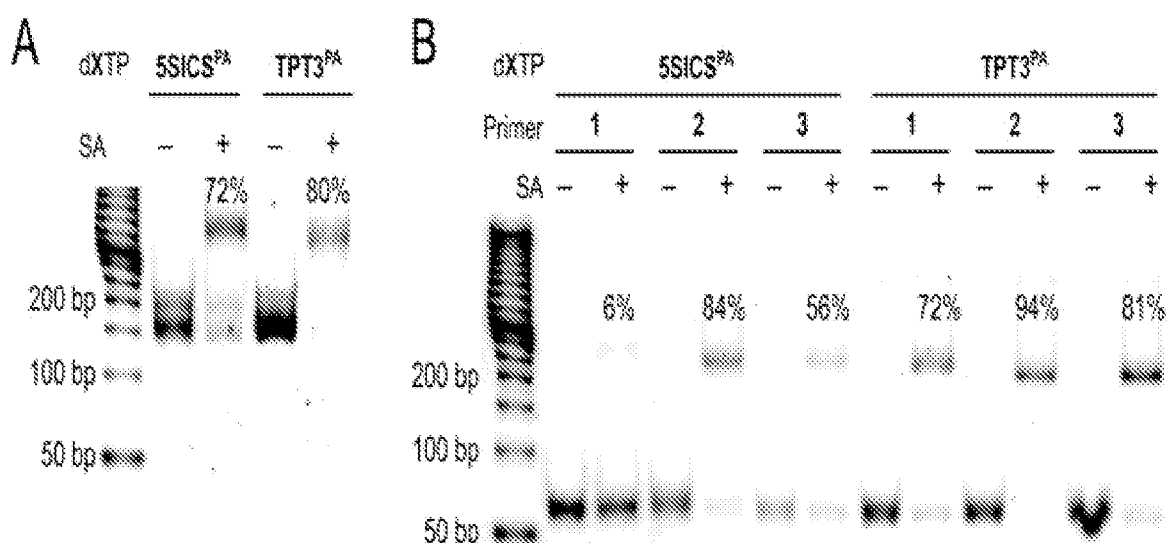




Figure 8

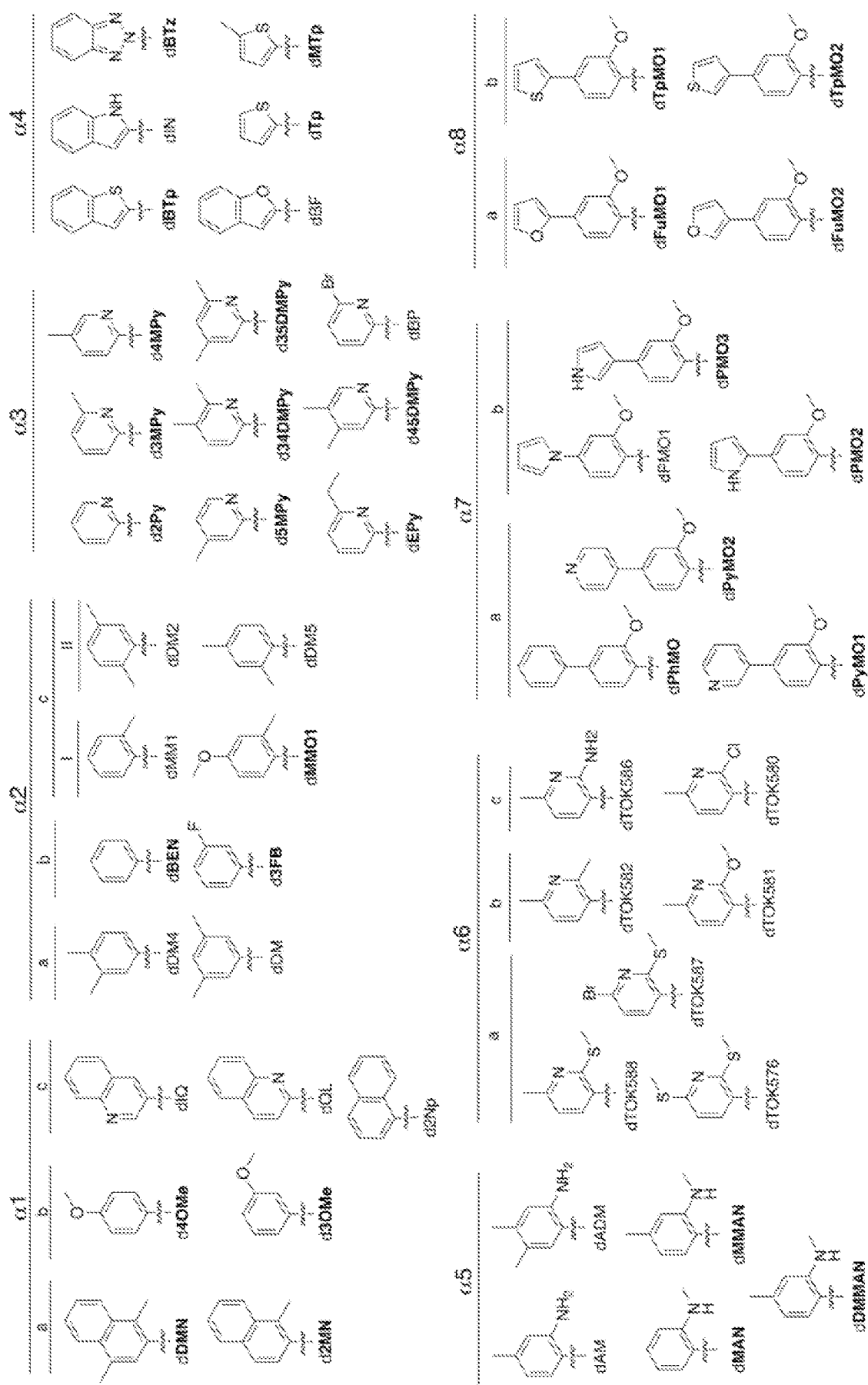


Figure 8  
Continued

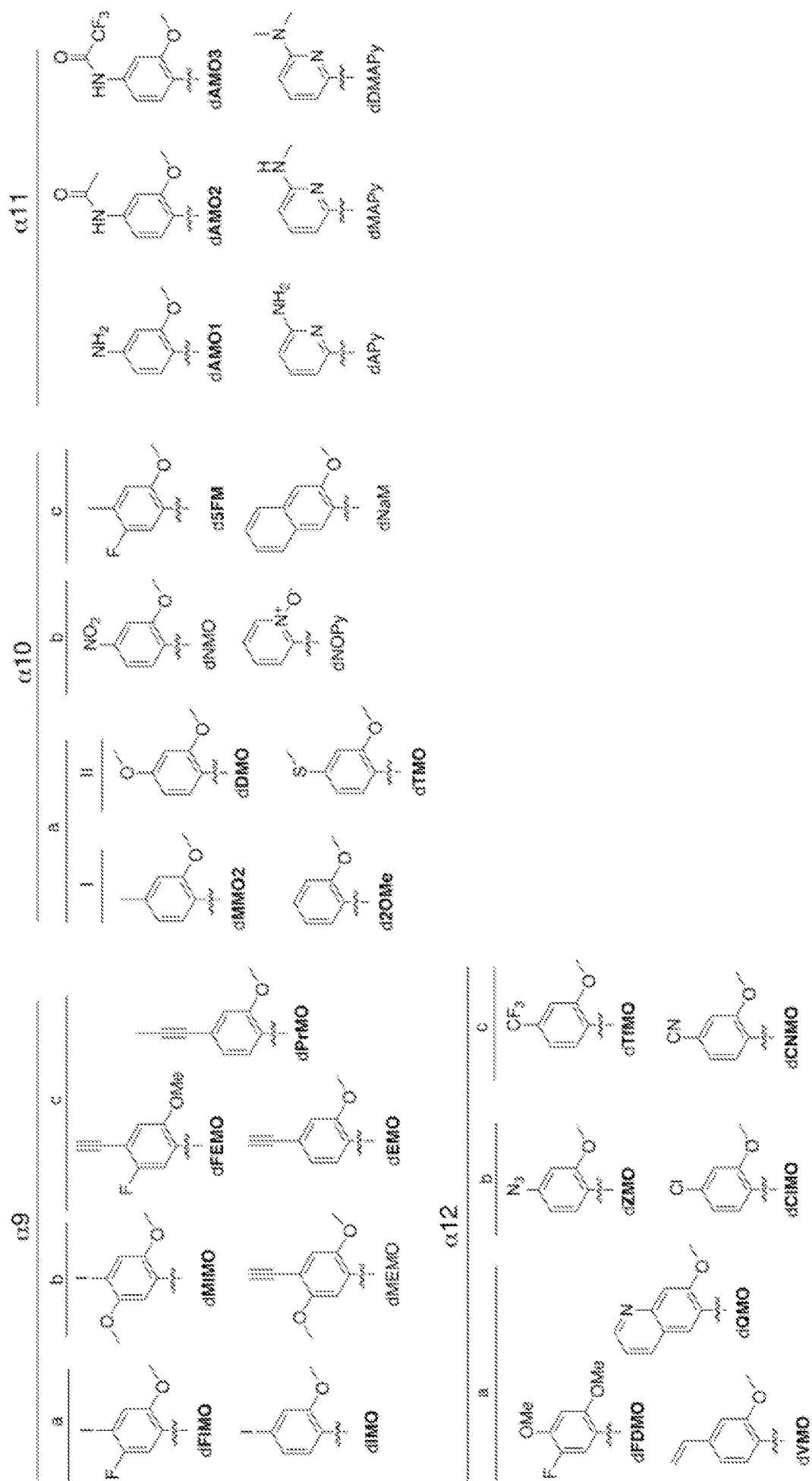




Figure 9

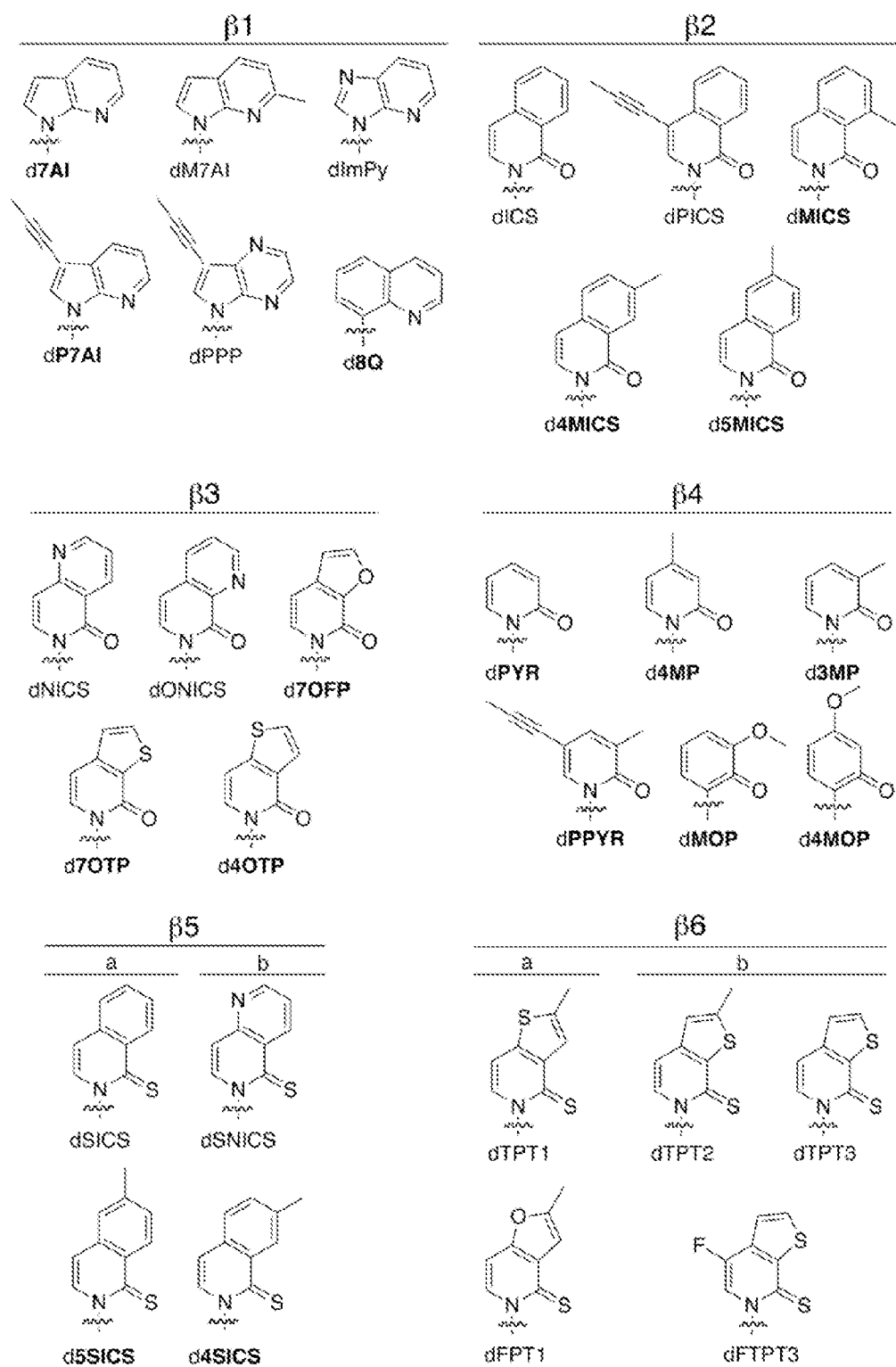


Figure 10

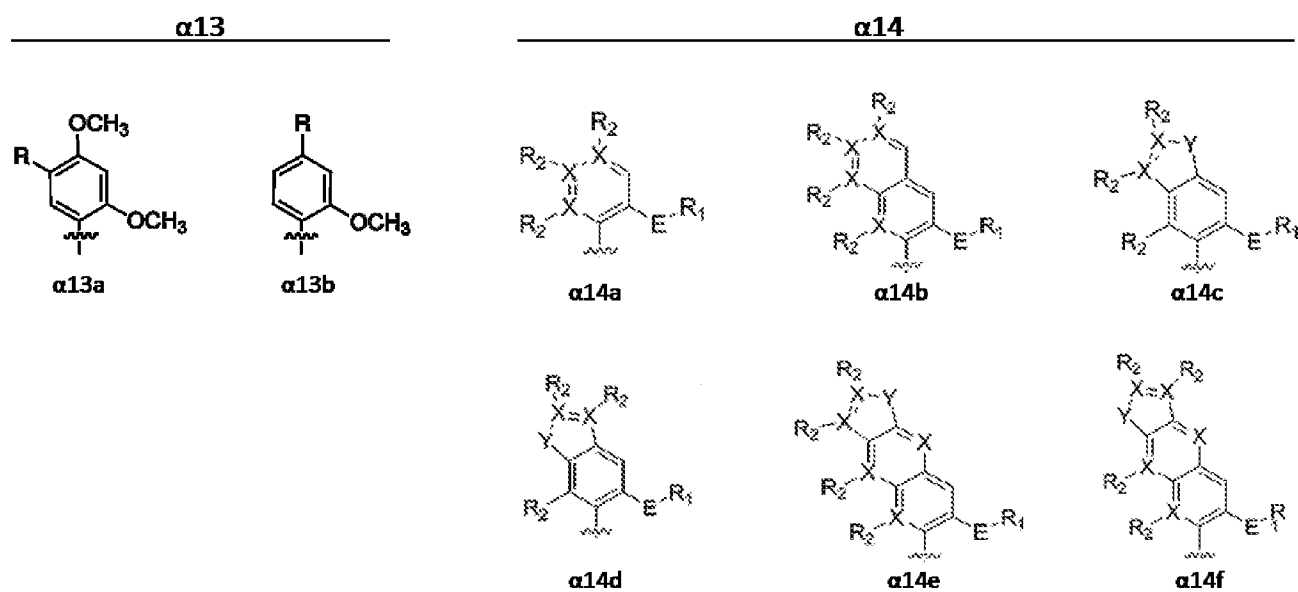


Figure 11

α14a

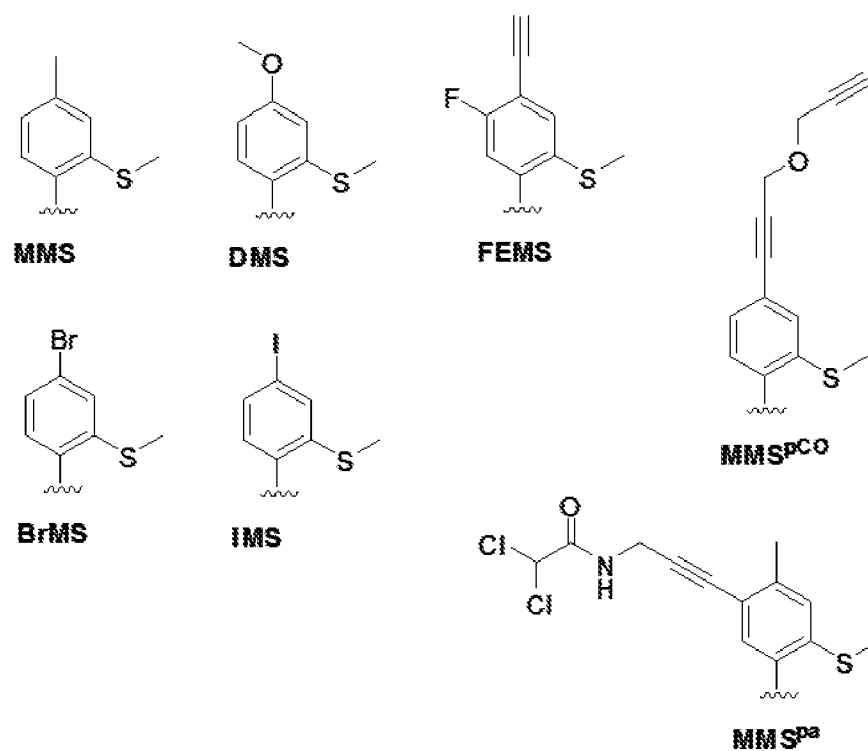


Figure 12

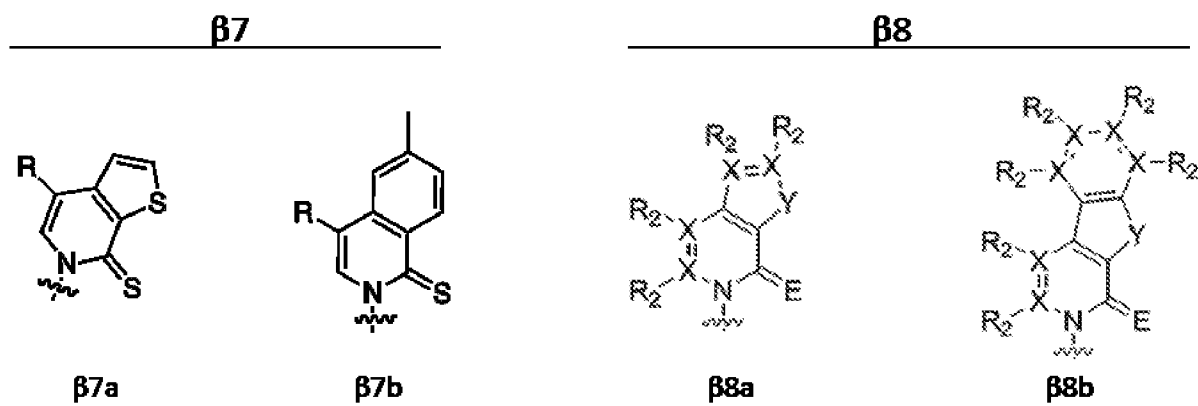


Figure 13

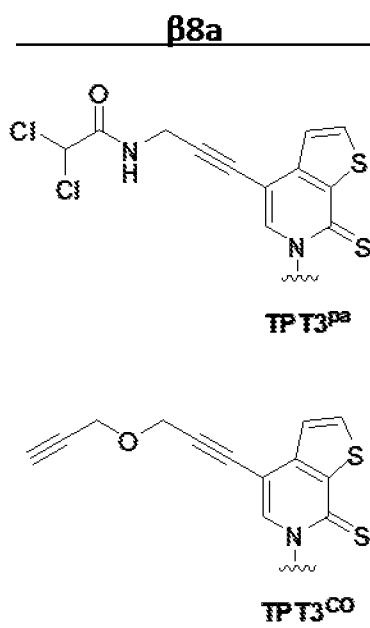


Figure 14

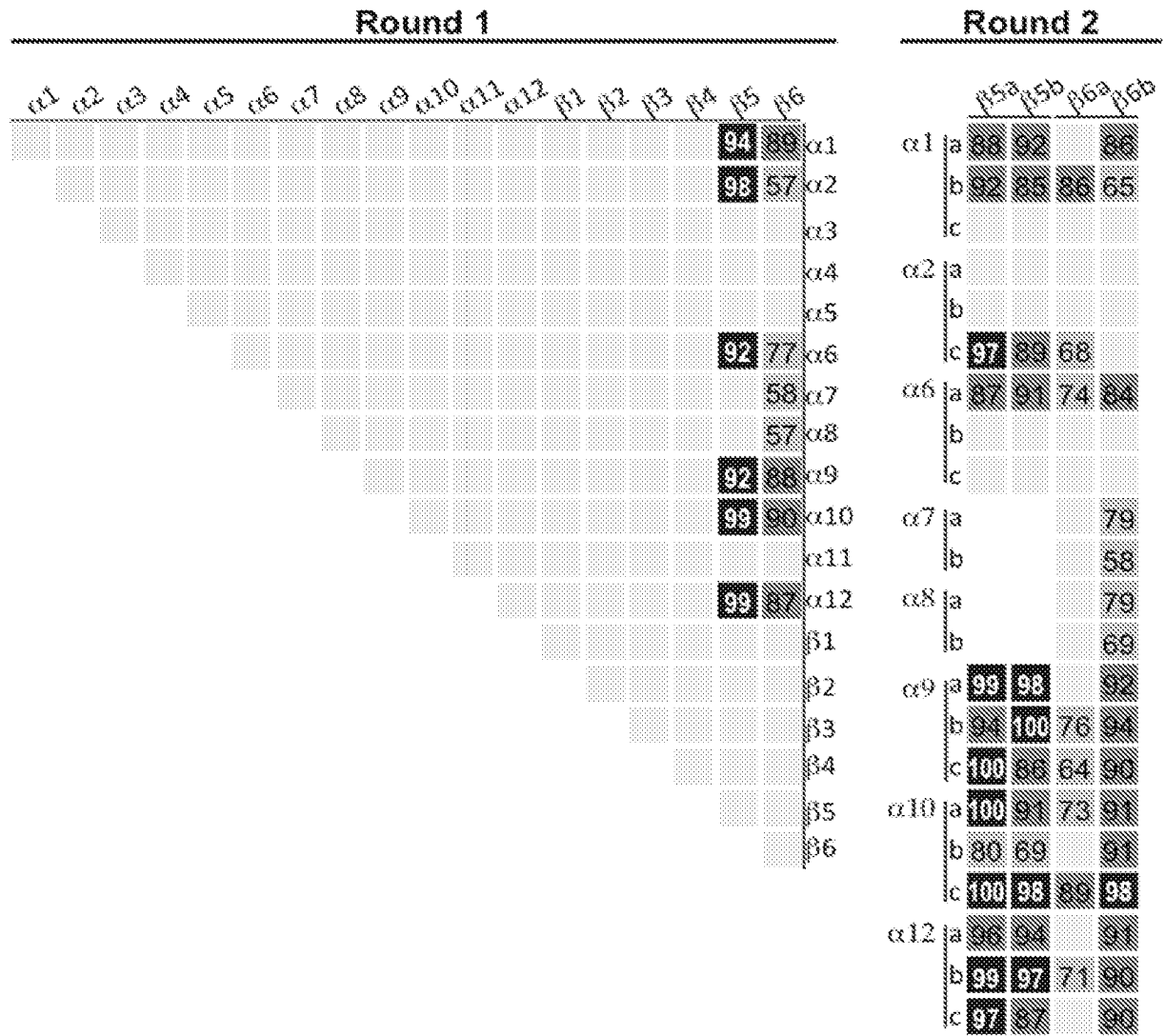
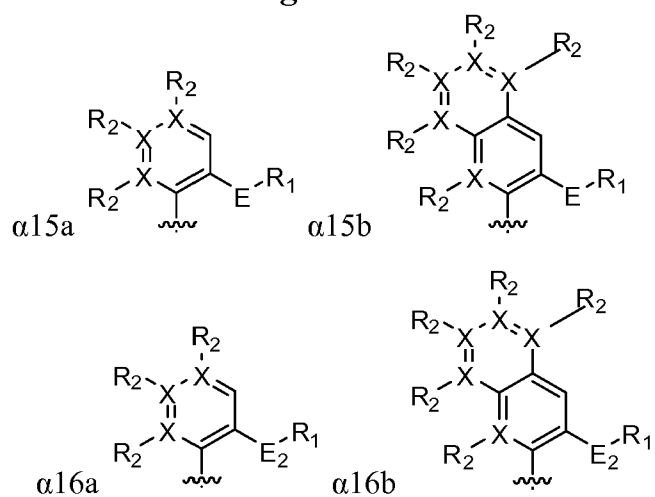


Figure 14  
Continued

Round 3						Round 4					

Round 6									
OneTaq		SICS	SSICS	ASICS	SNICS	TPT2	TPT3	FIPT3	
$\alpha 1$	2MN	60			61		88	70	
	MM1				61	58			
	DM2	88			71	62			
$\alpha 6$	TOK587	78			83		72	81	
$\alpha 9$	FIMO	80	81				97	89	
	IMO	81	75		88		95	97	
	FEMO	92	83		95		95	97	
$\alpha 10$	NMO						89	82	
	NaM	95	86		81	93	97	88	
$\alpha 12$	ZMO						96	87	
	CIMO						90	94	
	TfMO						76	89	
	QMO						90	84	
	CNMO						91	92	
Taq		SICS	SSICS	ASICS	SNICS	TPT2	TPT3	FIPT3	
$\alpha 1$	2MN				71		64	91	
$\alpha 2$	MM1								
	DM2								
$\alpha 6$	TOK587	77			72		74	80	
$\alpha 9$	FIMO				71		99	99	
	IMO				65		96	89	
	FEMO						93	92	
$\alpha 10$	NMO						75	82	
	NaM	79			73		89	92	
$\alpha 12$	ZMO						77	73	
	CIMO						77	89	
	TfMO						69	72	
	QMO						73		
	CNMO						91	94	

**Figure 15****Figure 16**