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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 October 2020 (29.10.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/219853 A1

(51) International Patent Classification:
C12P 19/34 (2006.01) *C12N 15/09* (2006.01)
C12M 1/00 (2006.01) *C12N 15/10* (2006.01)

(21) International Application Number:
PCT/US2020/029780

(22) International Filing Date:
24 April 2020 (24.04.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16/393,510 24 April 2019 (24.04.2019) US

(71) Applicant: **MOLECULAR ASSEMBLIES, INC.** [US/US]; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US).

(72) Inventors: **EFCAVITCH, J., William**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US). **AGARWALLA, Sanjay**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US). **ALBIZATI, Kim**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US). **GRUBBS, Alan, W.**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US). **HOLDEN, Matthew, T.**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US). **HOPKINS, Patrycia, A.**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US). **SINGH, Jay, K.**; 10355 Science Center Drive, Suite 130, San Diego, CA 92121 (US).

(74) Agent: **MEYERS, Thomas, C.**; Brown Rudnick LLP, One Financial Center, Boston, MA 02111 (US).

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

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

Published:
— *with international search report (Art. 21(3))*

WO 2020/219853 A1

(54) Title: HOMOPOLYMER ENCODED NUCLEIC ACID MEMORY

(57) Abstract: Nucleic acid memory strands encoding digital data using a sequence of homopolymer tracts of repeated nucleotides provides a cheaper and faster alternative to conventional digital DNA storage techniques. The use of homopolymer tracts allows for lower fidelity, high throughput sequencing techniques such as nanopore sequencing to read data encoded in the memory strands. Specialized synthesis techniques allow for synthesis of long memory strands capable of encoding large volumes of data despite the reduced data density afforded by homopolymer tracts as compared to conventional single nucleotide sequences.

HOMOPOLYMER ENCODED NUCLEIC ACID MEMORY

RELATED APPLICATIONS

The present application claims priority to U.S. no-provisional application serial number 16/393,510, filed April 24, 2019, the content of which is incorporated by reference herein.

FIELD OF THE INVENTION

The invention relates to methods and apparatus for storing data in a nucleic acid memory strand comprising homopolymer tracts.

BACKGROUND

DNA digital storage is a process of representing digital data using the base sequences of DNA and storing that data through DNA synthesis of polynucleotides corresponding to the base sequence encoding the data. DNA digital storage provides several advantages over conventional data storage methods and targets a market in the tens of billions of dollars. Conventional data storage methods including flash memory and recording on magnetic tapes pose problems related to physical space requirements, reliance on scarce resources, and data integrity. DNA digital storage provides much greater data storage density with significantly lower energy requirements. Current methods rely on high-fidelity sequencing techniques with little tolerance for errors in order to accurately read the data encoded in the DNA. The required sequencing methods are relatively slow and expensive to meet the fidelity requirements. An example of current DNA digital storage techniques is described in U.S. Pat. No. 9,384,320 to Church, et al. (incorporated herein by reference). In order to increase sequencing fidelity, current methods such as those described by Church encode data using sequences that avoid features that are difficult to read or write such as sequence repeats.

The synthesis side of current DNA digital storage techniques further limits adoption of the technology through a lack of speed, production of toxic byproducts, and high costs. Most *de novo* nucleic acid sequences are synthesized using solid phase phosphoramidite-techniques that involve the sequential de-protection and synthesis of sequences built from phosphoramidite reagents corresponding to natural (or non-natural) nucleic acid bases. While inkjet synthesis on array-based formats is capable of very low cost phosphoramidite synthesis, the strands that are made are limited to 100-200 bases in length, must sacrifice some of the length to index

sequences, and are made in sub-femtomolar scale requiring post-synthesis amplification to provide sufficient material for subsequent read-out. Using conventional synthesis techniques, nucleic acids greater than 200 base pairs (bp) in length experience high rates of breakage and side reactions. Additionally, conventional synthesis techniques produce toxic by-products, and the disposal of this waste limits the availability of nucleic acid synthesizers and increases the costs of oligo production. These complications related to synthesis and read-out in DNA digital storage have limited the applications for an otherwise promising technology.

SUMMARY

The invention provides systems and methods for storing data using sequences of homopolymer tracts encoding digital data. Representing each bit in the data sequence using a homopolymer tract of repeated bases (e.g., 2-10 nucleotides) allows for higher throughput and less expensive sequencing techniques to be used. Because the sequence read relies only on discriminating the transition between homopolymer tracts and does not require a faithful read of each individual nucleotide, sequencing techniques such as nanopore sequencing, zero-mode waveguide (ZMW) single molecule sequencing, and mass spectrometry may be used to increase speed and reduce cost.

Recording of data using homopolymer tracts as described herein is most efficiently accomplished using long strands (e.g., 5-10 kb) of nucleic acid. While traditional synthesis techniques are length limited, template-independent polynucleotide synthesis of using, for example, a nucleotidyl transferase are capable of synthesizing long strands at reduced costs and with lower waste production. Enzymatically synthesized ssDNA memory strands only require 50% of the DNA synthesis compared to conventional phosphoramidite approaches because ssDNA strands longer than about 100 – 200 nucleotides in length require complex and costly ligation or PCR techniques and can only produce ssDNA from dsDNA intermediates. See, U.S. Pat. No. 8,808,989 to Efcavitch, et al., incorporated herein by reference. Data encoding can be in numerical base 2, 3, 4 using standard nucleotides or data density can be increased using any number of modified nucleotide analogs to generate base 8, 10, 12, or more encoding schemes.

The limitations on modified nucleotide analogs are only that they can be incorporated using the chosen synthesis technique (e.g., terminal deoxynucleotidyl transferase (TdT)) and can be differentiated from one another using the chosen sequencing analysis. In some embodiments, synthesis may be accomplished using polymerase theta in the presence of Mn^{2+}.

Consistent homopolymer tract length is not essential to the systems and methods of the

invention because it is only the transition between individual tracts that needs to be recognized. Even though the tract length can be allowed to vary, synthesis techniques of the invention can effectively control the average homopolymer tract length by adjusting the ratio of deoxynucleotides (dNTPs) to the oligonucleotide memory strands being synthesized and controlling the exposure time of the dNTPs to the nascent memory strand. The length of the homopolymer tracts can be optimized to the readout technology; the highest data storage density is achieved with single nucleotide readout resolution, but the highest readout speed and accuracy are achievable by expanding the size of the nucleotide bit to the minimum detectable length (e.g., 2-10 nucleotides) for a given sequencing technology.

Systems and methods of the invention that use nanopore sequencing may use specialized memory strand constructs such as stoppers (e.g., hairpins or macromolecular appendages) included on one or both ends of the strand. In other nanopore-based methods, the memory strand may be circularized and threaded between adjacent nanopores.

Certain aspects of the invention include a method of recording data using a nucleic acid memory strand. Steps of the method may include creating an in-silico oligonucleotide sequence that represents a dataset where each nucleotide of the oligonucleotide sequence corresponds to a unit of said dataset. A nucleic acid memory strand can then be synthesized comprising a plurality of homopolymer tracts where each homopolymer tract corresponds to a nucleotide of the oligonucleotide sequence. The plurality of homopolymer tracts may include between 3 and 10 repeated nucleotides. Each unit of said dataset can be represented in base 2, base 3, base 4, or higher as desired for a particular application.

In certain embodiments, the nucleic acid memory strand may be from at least about 200 nucleotides in length to about 5,000 nucleotides in length. The synthesizing step may include controlling homopolymer length by varying dNTP concentration. Steps of the method may include modifying a first end of the nucleic acid memory strand to prevent passage of the first end through a nanopore of a nanopore sequencing system; passing a second end of the nucleic acid memory strand through the nanopore; and modifying the second end of the nucleic acid memory strand to prevent passage of the second end through the nanopore.

Other embodiments may utilize a memory strand be comprised of heteropolymer tracts of a defined stoichiometry or composition to further increase the coding capacity of a set number of nucleotide analogs. Further embodiments may seek to protect the data encoded in a memory strand by using nucleotide analogs that are similar in structure but employ linkers that can be removed under different conditions such as ultraviolet or visible light, oxidizing or

reducing agents, alkaline or acidic pH, or sequence specific nucleases, thereby disguising the data to those without knowledge of the applicable process.

The dataset may be selected from the group consisting of a text file, an image file, and an audio file. The synthesizing step may include template-independent synthesis. In certain embodiments, a nucleotidyl transferase enzyme may be used to catalyze said template-independent synthesis. Polymerase theta can be used to catalyze said template-independent synthesis in some embodiments.

Aspects of the invention may include a method of reading data from a nucleic acid memory strand. Steps of the method can include sequencing a nucleic acid memory strand comprising a plurality of homopolymer tracts; converting the nucleic acid memory strand sequence into digitized data, wherein each of the plurality of homopolymer tracts represents a nucleotide corresponding to a unit of data; and converting the digitized piece of data to a readable format. Steps of the method may include displaying the readable format. The plurality of homopolymer tracts may include between about 2 and about 10 nucleotide repeats. The nucleic acid memory strand may be between at least about 200 nucleotides and about 5,000 nucleotides in length.

In various embodiments, the sequencing step can comprise nanopore sequencing, sequencing by synthesis, or mass spectrometry. The sequencing, translating, and converting steps may be repeated one or more times on the nucleic acid memory strand.

Other aspects of the invention are apparent to the skilled artisan upon consideration of the following figures and detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows an enzymatic synthesis cycle used to form homopolymer tracts.

FIG. 2 shows a method of synthesizing a nucleic acid memory strand with homopolymer tracts.

FIG. 3 shows a method of reading data from a nucleic acid memory strand with homopolymer tracts.

FIG. 4 shows the relationship of strands/GB to homopolymer length and strand length.

FIG. 5 shows the data that can be encoded in a DNA strand comprised of 500 polymer tracts as a function of the number of distinguishable nucleotide analogs in a memory strand.

FIG. 6 shows a nanopore trapped nucleic acid memory strand of the invention.

FIG. 7 shows a scheme for changing the data encoded within a strand of homopolymer tracts in response to treatment conditions.

FIG. 8 shows a system for synthesizing nucleic acid memory strands with homopolymer tracts.

FIG. 9 shows a system for the parallel synthesis of nucleic acid memory strands with homopolymer tracts on an array of nanowells.

FIG. 10 gives a more detailed schematic of components that may appear within a system.

FIG. 11 shows analysis of the enzymatically-mediated synthesis of two different base composition homopolymer tracts.

FIG. 12 shows the process for converting character text into a 12-bit nucleic acid sequence.

FIG. 13 shows polyacrylamide gel electrophoresis (PAGE) analysis of twelve cycles of enzymatic synthesis.

FIG. 14A-L shows the experimentally found homopolymer distribution for each of the 12-bit homopolymer additions.

FIG. 15 illustrates cycle steps for the template independent DNA polymerase mediated synthesis of homopolymer encoded information polymers.

FIG. 16 shows a side view of a typical reaction zone, in a 2D array of reaction zones, showing hydrophobic patterning isolating a liquid reaction droplet.

FIG. 17 shows a side view of a reaction zone with a controlled heater element or electrochemical element.

FIG. 18 shows a reaction zone showing a light transparent top cover.

FIG. 19 shows an exemplary apparatus useful for the addressable light activation of enzymatic DNA synthesis.

FIG. 20 shows kinetics of UV light decaging of 3'-O-(2-nitro)-benzyl dATP.

FIG. 21 shows optically controlled oligonucleotide extension. An enzymatic reaction mixture containing 3'-orthonitrobenzyl dATP was irradiated with a low-power light source for various intervals so that the exposure time controlled the amount of decaging. The amount of usable natural dATP formed in turned controlled the average length of the polynucleotide tract.

FIG. 22 shows structures of 3'-O-caging modifications allowing two-color light-mediated decaging.

FIG. 23 shows gel electrophoresis analysis of each of 12 cycles of homopolymer synthesis using 3'-O-(2-nitro)-benzyl dATP and dCTP, dGTP, dTTP.

FIG. 24 shows electrophoretic analysis of enzymatic synthesis reactions comparing incorporation modulating dNTP analogs versus unmodified dNTP.

FIG. 25 shows electrophoretic analysis of two consecutive cycles of incorporation rate modulating dNTP analogs. Modified analogs were used to form N+1 homopolymers, then followed by N+2 homopolymer additions. Homopolymer synthesis rate modulating modifications were removed from oligonucleotide prior to gel analysis.

FIG. 26 shows an electropherogram of TdT extension reactions showing homopolymer rate modulating analog incorporation followed by second cycle of homopolymer rate modulating analog incorporation. Modified nucleotide followed by consecutive modified nucleotide.

FIG. 27 shows a modified dNTP analog used for information polymer composition.

FIG. 28 shows electrophoretic analysis of an information polymer composed of modified nucleotides.

FIG. 29 shows detection of an information polymer (P71) composed of modified nucleotides by translocation through a nanopore.

FIG. 30 shows detection of an information polymer (86B) composed of modified nucleotides by translocation through a nanopore.

DETAILED DESCRIPTION

The invention provides systems and methods for writing data to and reading data from nucleic acids having homopolymer tracts corresponding to units of digital data. By repeating (e.g., 3-10 times) each nucleotide in the data-encoding sequence, only the transition between homopolymer tracts needs to be observed in sequencing reads allowing for lower fidelity, higher throughput sequencing techniques that can result in cheaper execution in nucleic acid data storage. Advantages of synthesizing nucleic acid homopolymer tract memory strands are: 1) the ability to make very long (5 – 10 kb) strands, which enables the use of high throughput, long read DNA sequencing technologies for readout, 2) the ability to tolerate errors in sequencing readout technologies and 3) the ability to make nucleic acid memory strands with costs far less than that of conventional chemical synthesis methods. The use of homopolymer nucleic acid memory strands is best realized in long (e.g., 5-10 kb) strands that can be efficiently produced using template-independent TdT enzymes or polymerase theta wherein homopolymer tract length can be controlled by altering exposure time and dNTP to polynucleotide memory strand ratio.

The synthesis of homopolymers for encoding data by an enzymatically mediated approach is easily achieved by using natural or modified nucleotide triphosphates that are not terminators,

resulting in the simplest and most rapid method of DNA synthesis possible. One natural or modified nucleotide triphosphate is delivered to a reaction zone with a nucleotidyl transferase, allowed to react and then removed by washing with a buffer thus completing one "write" cycle of data storage as illustrated in FIG. 1. The data strand synthesis occurs in entirely aqueous environment, with no toxic or hazardous chemicals, thus enabling practical devices suitable for large scale data storage centers.

FIG. 2 shows a method 101 of synthesizing a nucleic acid memory strand with homopolymer tracts according to certain embodiments. The method 101 includes creating 103 an in-silico oligonucleotide sequence representing a dataset. The dataset may comprise digitized data that may represent text, an image, a video, an audio, or any other piece of information that may be digitized. The oligonucleotide sequence may comprise any number of natural or modified nucleotides or analogs thereof and may encode the dataset using a base 2, base 3, base 4, or greater scheme depending on the number of unique nucleotides or analogs used in the memory strand. In a simple embodiment, the encoding scheme may correspond to a binary data scheme conventionally represented by a series of 0s and 1s where one or more nucleotides or analogs may correspond to 0s and one or more other nucleotides may correspond to 1s. A nucleic acid memory strand (e.g., RNA, single-stranded, or double-stranded DNA) comprising a series of homopolymer tracts each corresponding to a nucleotide, in order, in the in silico oligonucleotide sequence can then be synthesized 105. In certain embodiments, further steps include modifying 107 one end of the memory strand, threading 109 the memory strand through a nanopore and modifying 111 the other end of the strand to prevent the ends from passing through the nanopore.

FIG. 3 shows a method 203 of reading data from a nucleic acid memory strand with homopolymer tracts. Steps of the method 203 include sequencing 203 a series of homopolymer tracts in a nucleic acid memory strand, converting 205 the sequence to a dataset, converting the dataset into a readable format (e.g., an image, a video, an audio clip, or a piece of text), and optionally displaying 209 that readable format of data (e.g., on a monitor or using a printer or other input/output device).

Preferably, systems and methods of the invention use long strands of DNA (5-10 kb) that are either single stranded or double stranded and may be naturally occurring or generated by chemical or enzymatic synthesis. In certain embodiments, the nucleic acid memory strands may be generated enzymatically using TdT to create a series of homopolymer tracts that may be 2-10, 3-10, 4-10 nucleotides or longer. The homopolymer tracts can each consist of Adenine (A), Guanine (G), Cytosine (C), or Thymine (T). The sequence of alternating homopolymer tracts can be used to

encode the data that is to be stored in the memory strand.

Each nucleotide homopolymer tract can represent various amounts of data depending on the number of bases used. The number of bits required to make one byte (decimal 256) is defined by the following relationship: #bits/byte = 8/(log2(n)), where n = the numerical base that is used. Each tract can correspond to one bit if two base encoding is used or ¼ of a byte if four base encoding is used. In certain embodiments, DNA data strands may be composed of 2-10 nucleotide homopolymer tracts (using a base 2 dataset representation), which would allow 333 bits to 100 bits to be represented in a memory strand between 999 and 1000 bases long. In preferred embodiments, nucleic acid base encoding of data may be such that single homopolymer tracts of one nucleotide are always adjacent to homopolymer tract of a different nucleotide. For example, encoding may be such that a homopolymer tract of Adenine would not be immediately preceded or followed by another Adenine homopolymer tract. In the case where two adjacent homopolymer tracts comprise the same nucleotide, a homopolymer tract may be synthesized that is measurably longer than the average homopolymer tracts representing single nucleotides in the encoded data sequence. Those longer tracts may be created through manipulation of the synthesis reactions described below by, for example, increasing the concentration of dNTPs in the reaction or increasing the reaction time. The exact length of two adjacent identical homopolymer tracts need only be long enough to be unambiguously distinguished from single homopolymer tracts using the readout device (i.e., nanopore sequencer). In certain embodiments, a non-nucleotide homopolymer spacer could be added between A, G, C, or T homopolymer tracts to clearly distinguish adjacent same nucleotide homopolymer tracts from one another. The use of A, G, C, & T homopolymer tracts enables the creation of a four (4) bit encoding space increasing the density of data that can be stored in one contiguous strand rather than simply using two nucleotides (similar to 0s and 1s in binary code). For example, four contiguous homopolymer tracts can encode 256 digits (i.e., one byte) if A, G, C, & T are used in a base 4 scheme. In such an embodiment, there would be 83 bytes or 25 bytes represented in a 996 or 1000 nucleotide long nucleic acid memory strand if three (3) nucleotide long homopolymer tracts or ten (10) nucleotide long homopolymer tracts were used respectively.

In various embodiments, base 8 or even base 12 coding schemes may be employed through the incorporation into the memory strand of homopolymer tracts of uniquely modified nucleotide or non-nucleotide analogs. Those modified nucleotide or non-nucleotide analogs should generate a unique digital signal with a readout device like a nanopore sequencer or a single molecule ZMW sequencer. TdT, as discussed below, can incorporate a wide range of modified dNTP analogs that can enhance the signal provided by a readout device like a nanopore and thus may be used for

generating nucleic acid memory strands with data encoded in them. Homopolymers of modified nucleotides (e.g., A*, G*, C* & T* or A**, G**, C** & T**) can be synthesized using TdT and modified dNTP analogs (e.g., dA*TP or dA**TP) of each of the four bases to generate an eight (8) bit or even a twelve (12) bit encoding scheme. Higher base (n) encoding allows for data compression and results in a reduction in the number of DNA strands that are required to encode a given amount of information. The relationship determining the number of DNA strands per GB of data as a function of the length of the homopolymer tract, the numerical base (n), and the strand length synthesized is defined by: #strands/GB = $(8/(\log2(n))*10^9$*homopolymer length* 1/strand length as illustrated in FIG. 4.

The number of unique homopolymer tracts may be limited only by the ability of the readout technology (i.e., nanopore or ZMW single molecule sequencing) to determine one homopolymer tract from another. There are several reports in the literature of the detection of homopolymers composed of unmodified nucleotides by detecting the change in the ionic current during translocation through nanopores (Venta et al, 2013; Feng et al, 2015). Modifications that alter the dwell time of the DNA in the nanopore will generate a distinguishable and characteristic ionic-current signal. Singer et al 2010 and Morin et al 2016 use non-covalently bound bisPNA or γPNA functionalized with 5 or 10kDa PEG to enhance detection by nanopores. Liu et al 2015 selectively created adamantly 8-oxoG analogs to modify the dwell time and generate a unique signal. Given the tolerance of TdT to incorporating bulky modifications at N6 of dATP, N4 of dCTP, N2 or O6 of dGTP and O4 or N3 of dTTP, acyl or alky modifications at those positions may be screened and chosen to enhance the detection modality of nanopore or ZMW single molecule sequencing technologies. Detection may be improved through modified nucleotides that enhance the differential current blockade in a nanopore or enhance the dwell time of a modified nucleotide in the active site of a DNA polymerase in a ZMW single molecule approach. Other natural and non-natural purine and pyrimidine nucleotide analogs may be used if they generate a unique digital signal with a readout device like a nanopore sequencer or a single molecule ZMW sequencer. Modifications at the C5 or C7 of pyrimidines and purines respectively may be used if they generate a unique digital signal with a readout device like a nanopore sequencer or a single molecule ZMW sequencer. Suitable modified nucleotide triphosphates are chosen to be rapidly incorporated during the enzymatic extension step and provide a substitution-specific dwell time with as short of a homopolymer as possible during the detection step. Examples of modifications to A, G, C, & T bases suitable for expanding the bit encoding space may include but are not limited to N6-benzoyl dA, N6-benzyl dA, N6-alkyl dA, N6-acyl dA, N6-substituted alkyl dA, N6-substituted acyl dA, N6-

aryl acyl dA, N6-substituted aryl acyl dA, N2-alkyl-dG, N2-acyl dG, N2-aryl acyl dG, N2-substituted alkyl dG, N2-substituted acyl dG, N2-substituted aryl acyl dG, O6 alkyl dG, O4 alkyl dT, N3 alkyl dT, N3 acyl dT, O6 substituted alkyl dG, O4 substituted alkyl dT, C5-propargyl amine dT, C5-propargyl amine dC, C7-propargyl amine dA, C7-propargyl amine dG, substituted C5-propargyl amine dT, substituted C5-propargyl amine dC, substituted C7-propargyl amine dA, substituted C7-propargyl amine dG. Preferred embodiments of substitutions include but are not limited to covalent attachments that are completely stable to removal except under the most extreme chemical conditions of pH, temperature and concentration of reactive species. Substitutions that are able to affect a unique current blockade may include but are not limited to alkyl, heteroatom substituted alkyl, aromatic hydrocarbons, alkyl substituted aromatic hydrocarbons, heteroatom substituted alkyl substituted aromatic hydrocarbons, heteroatom substituted aromatic hydrocarbons, benzyl, substituted benzyl or combinations of the such. In some embodiments, the substitutions can be polyethylene glycols composed of 2 to 450 monomer units. In some embodiments, substitutions comprised of peptides or peptoids may be suitable to increase the dwell time of homopolymers in a specific and discernable manner. The efficiency of incorporation of modified nucleotides by template-independent polymerases like TdT may be modulated by the use of different metal ion cofactors such as but not limited to Co++, Zn++, Mg++, Mn++, or mixtures of two or more different metal ions. Each modified nucleotide may require a different metal ion for optimal performance during enzymatic homopolymer synthesis.

Since long term stability of the DNA data strands is essential, there is a distinct advantage to using non-purine based homopolymers since they are subject to depurination at low pH. In some embodiments, the homopolymer bits can be composed of only a single nucleotide type (i.e., Thymine) that is modified with two, three, four or more different chemical groups resulting in homopolymer tracts that each result in a unique current blockade. Thus, one nucleotide labeled with four unique modifiers can substitute for the presence of A, G, C, T. Other embodiments that use only one of the other three nucleotides with two, three, four or more different chemical groups are possible.

If the modified nucleotide analogs are sufficient enough to cause a unique dwell time for the passage of a single nucleotide through a nanopore, another embodiment would use single nucleotide bits instead of homopolymer bits. Single modified nucleotide bits would be advantageous in allowing the maximum density of information per DNA strand thus reducing the cost of DNA based data storage.

The precise length of the homopolymer tract is not critical so long as the sequencing

technology used for the read-out can clearly distinguish the start and stop of one homopolymer tract from another. Although there are distinct synthesis and storage density advantages to increasing the number of unique nucleotides or bases (including modified nucleotide or non-nucleotide analogs) used and therefore reducing the length needed to capture a set amount of data, the lowest cost per DNA data storage synthesis may be achieved by using four natural nucleotide dNTP monomers during enzymatic synthesis since those reagents are widely used in the molecular biology & sequencing fields and are produced in very large batches with the lowest manufacturing cost. The cost of production of dNTP analogs to increase the number of unique homopolymer tracts may be reduced as the use of DNA data storage increases and manufacturing scale of analogs is also increased.

Any method of synthesizing the homopolymer tract segment may be used with systems and methods of the invention but preferred embodiments use the template-independent enzyme TdT. TdT provides certain benefits insofar as it will rapidly and inexpensively generate a homopolymer with a Poisson distribution where the average size of the homopolymer may be strictly controlled by the ratio of the [dNTP] to the nascent oligonucleotide memory strand. In some embodiments, polymerase theta in the presence of Mn^{2+} can be used as a template-independent polymerase to synthesize homopolymer tract nucleic acid memory strands. In another embodiment, the length of the homopolymer tract segments can be controlled by delivering an excess of dNTP to a reaction zone and then removing the reactants after carefully controlled interval of time.

TdT has demonstrated the ability to synthesize homopolymer tracts of a fairly defined length by controlling the ratio of dNTP concentration to the concentration of 3'- ends of the nucleic acid strand desired to be modified. Inkjet synthesis on array-based formats is capable of very low cost phosphoramidite synthesis but the strands that are made are limited to 100-200 bases in length, must sacrifice some of the length to index sequences, are made in sub- femtomolar scale requiring post-synthesis amplification to provide sufficient material for subsequent read-out and are mostly suited for relatively inefficient short read sequencing readout technologies.

Strands of single stranded DNA synthesized according to processes of the invention may benefit from the prevention of hairpins or dsDNA either during the synthesis or during the readout. Hairpin formation can be prevented by modifying the exocyclic amines of one member of a A:T or G:C base pair to prevent the hydrogen bonding necessary for base pairing. In some embodiments, the exocyclic amines may be modified by acylation or alkylation. Any simple and stable modification of the exocyclic amines of A, G or C, which prevents base pairing can be used to

prevent hairpin formation. In certain embodiments, the N6 of deoxyadenosine and the N2 of deoxyguanosine may be acetylated with an acetyl group preventing base pairing. In some embodiments, the N6 of deoxyadenosine and the N4 of deoxycytidine can be modified to prevent base pairing and hairpin formation. In some embodiments, the O6 of deoxyguanosine or the O6 of deoxythymidine can be modified to prevent base pairing and hairpin formation. In some embodiments, the O4 of deoxythymidine or the N3 of deoxythymidine can be modified to prevent base pairing and hairpin formation. In some embodiments, modifications to A, G, C, or T to generate higher order base encoding schemes also serve the purpose of preventing base pairing and hairpin formation. In some embodiments, homopolymer bits can be composed of only a single nucleotide type (i.e., Thymine) that is modified with two, three, four or more different chemical groups that result in a unique current blockade and prevent the formation of intra- or inter-strand double strand regions. In another embodiment, a thermostable version of TdT or another template-independent nucleotidyl transferase can be used to perform strand synthesis at an elevated temperature, thus preventing prevent the formation of intra- or inter-strand double strand regions.

Control of the homopolymer tract length can be optimized for any analogs as described above after determination and calibration of the incorporation rate of the dNTP analog to create a reproducible range of homopolymer tract lengths of 2-10 nucleotides in length. The use of A*, G*, C*, & T* and A**, G**, C** & T** homopolymer tracts allows the creation of an eight (8) bit or twelve (12) bit encoding, increasing the density of data that can be stored in one contiguous strand rather than simply using two nucleotides to encode for a "0" and a "1". Three (3) contiguous homopolymer tracts can encode 256 digits (i.e., one byte) if A, G, C, T, A*, G*, C*, & T* are used. In such embodiments, there would be 111 bytes or 33 bytes in a 999 or 990 nucleotide long nucleic acid memory strand if three (3) nucleotide long homopolymer tracts or ten (10) nucleotide long homopolymer tracts were used respectively. Two (2) contiguous homopolymer tracts can encode 256 digits (i.e., one byte) if A, G, C, T, A*, G*, C*, T*, A**, G**, C**, & T**are used. In those embodiments, there would be 166 bytes or 50 bytes in a 996 or 1000 nucleotide long nucleic acid memory strand if three (3) nucleotide long homopolymer tracts or ten (10) nucleotide long homopolymer tracts were used respectively.

In certain embodiments, data may also be encoded in the memory strand heteropolymer tracts of random sequence and defined composition to achieve a higher level of data compression. The heteropolymer stretches can be generated with enzymatic reactions using mixtures of different dNTPs, where the dNTP stoichiometry is used to control the composition of the heteropolymer tracts. The number and type of heteropolymer tracts is limited only by the combinations of dNTP

analogs and the ability of the detection modality to distinguish the compositions of the different tracts. For m dNTP analogs, there are $(m^2-m)/2$ binary combinations for heteropolymer formation. Detection modalities which can distinguish two different levels of tract compositions for each binary combination (e.g. a tract where analogs A and B are present in an approximately 2:1 ratio respectively and a tract where they are present in a 1:2 ratio) allow data to be encoded at a rate of base m^2 from a set of m analogs, effectively doubling the coding capacity of the memory strands. FIG. 5 illustrates the data that can be stored in a memory strand as a function of the number of available dNTP analogs using either homopolymer and a binary heteropolymer-based encoding scheme with two levels tract composition.

Data encoding strands of the invention may not necessarily require precisely defined homopolymer lengths since they only need to be long enough (ca 2-10 nucleotides) to allow the unambiguous discrimination of the transition between homopolymer tract segments by a high throughput DNA sequencing technology. Existing Nest Generation Sequencing-by-Synthesis (SBS) systems can readily determine the transition between two adjacent homopolymer tracts. Again, the precise length of the homopolymer tract is not important to the accurate detection of a homopolymer bit. The use of tracts of the same nucleotide offers advantages in overcoming the most common errors in current SBS platforms: insertions and deletions. The deletion of one nucleotide in a homopolymer tract >2 nt will still be interpreted as a true homopolymer. Likewise, the insertion of a single nucleotide in a homopolymer tract would not be falsely interpreted as two adjacent homopolymers since the insertion of more than one nucleotide during SBS is an unlikely event. This sequencing error tolerance offers the advantage of decreasing the sequencing depth required to ensure correct decoding of the information stored by the DNA data strand. Existing nanopore systems can easily distinguish homopolymer tracts of A, G, C, or T from each other based on their differential current blockade. In certain embodiments, single molecule ZMW sequencing can be used to determine the linear order of homopolymer tracts on a linear strand. The use of either sequencing technology may require that the DNA initiator has properties that are compatible with the sequencing readout technology like a self-complimentary hairpin at the 5'-end of a synthesized single stranded memory strand to provide a primer for single molecule ZMW sequencing. Nanopore sequencing technology may also require a self-complimentary hairpin at the 5'- end of the strand to provide a "start" data mark. In various embodiments, the readout technology may be any next generation sequencing method such as that offered by Illumina (San Diego, CA). In some embodiments, the readout or sequencing technology may be Mass Spectrometry based. The technology specific error rate of the readout technology is not

important so long as it can unambiguously detect the transition between two different homopolymer tract and/or unambiguously detect the difference between one homopolymer tract length and one 2x in length in the case where two identical homopolymer tracts are adjacent to each other.

Certain read-out technologies may be preferable to others based on the specific application of the invention. Technologies such as nanopore sequencing can be non-destructive and leave the nucleic acid memory strand intact, suitable for multiple read-out cycles. Read-out technologies that are dependent on Sequencing by Synthesis (SBS) like ZMW single molecule and others, generate a copy of the original template strand and may require post-readout manipulation (i.e., strand separation by melting) to remove the complimentary strand and return the original nucleic acid memory strand to its pristine state ready for a subsequent cycle of readout. Other readout technologies, like mass spectrometry, are destructive and would deplete the pool of nucleic acid memory strands after repeated cycles of sampling and readout.

In various embodiments nucleic acid memory strands may include "stoppers". "Stoppers" may be macromolecular constructs which prevent the passage of a single stranded or double stranded nucleic acid through a nanopore (Manrao, et al., 2012, Reading DNA at single- nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase, Nature Biotechnology 30, 349–353, incorporated herein by reference). Proteins like phi29 DNA polymerase are large enough to not be drawn through the larger pore (about 6.3 nm) on the cis side of a protein nanopore. The pore diameter of the smaller side is estimated to be about 1.2 nm wide. Stoppers can be used either on the 5'- or 3'- end of a nucleic acid memory strand of the invention. In some applications, it may be desirable to have a stopper at both the 3'- or 5'- end of a nucleic acid molecule. Stoppers can consist of a hairpin (stem-loop) structure with either a protruding 5'- or 3'- overhang to which the information encoding nucleic acid is covalently attached. If the stopper consists of a hairpin, the length of the ds stem may be sufficiently long to resist any melting force exerted on it by the electric field used to translocate the memory strand through the nanopore. In certain embodiments, one base of the double-stranded stem region can be crosslinked to its cognate base that forms that base-pair so that it is impossible for the double- stranded stem portion of the hairpin to melt under the influence of the force exerted on it by the electric field that translocates the rest of the molecule through a nanopore. To utilize a hairpin stopper for TdT mediated nucleic acid memory synthesis according to certain embodiments, the stopper may have a 3'- overhang of sufficient length (i.e., >10 nucleotides) to allow the binding of TdT for template independent synthesis.

Stoppers may consist of a non-nucleotide macromolecular construct which can be appended

to either the 3'- or 5' end of a nucleic acid molecule. The construct can be synthesized by direct conjugation of a macromolecular species onto the 3'- end of a nucleic acid by a polymerase or transferase like TdT (Sorensen, et al., 2013, Enzymatic Ligation of Large Biomolecules to DNA, ACS Nano, 7(9):8098–8104, incorporated herein by reference) or by the incorporation of a functionalized nucleotide which allows the specific modification of the nucleic acid via that functionality (Winz, et al., 2015, Nucleotidyl transferase assisted DNA labeling with different click chemistries, Nucleic Acids Res. 43(17):e110, incorporated herein by reference). 5'- end stoppers may be readily introduced at the time of chemical synthesis of an oligonucleotide adapter either via direct synthesis of a hairpin or via secondary modification of a functional handle introduced as the last step of the 3' to 5' oligonucleotide synthesis and may be used as an initiator. Alternatively, 5'-end stoppers can be constructed by attaching an oligonucleotide initiator via the 5'-end to a magnetic or non-magnetic bead or particle or nanoparticle, enzymatically synthesizing the homopolymer tract containing memory strand and then leaving the memory strand attached to the magnetic or non-magnetic bead or particle or nanoparticle.

Stoppers can be further modified to allow cleavage of the stopper from the rest of the molecule to allow the nucleic acid strand to either passively diffuse out of the nanopore or to be translocated out of the nanopore through the application of a voltage thus allowing strand to be recovered.

In certain embodiments, template independent polymerases or transferases can be used to modify pre-synthesized strands of nucleic acids to enable the use of nanopore devices as "Write Once, Read Many" types of memory devices. Part of the inherent issues associated with the use of nanopore devices as DNA sequencers is the high error rate they produce because of the poor discrimination of the nanopore. This may be due to the speed of translocation through the pore or the fact that the approximate depth of the nanopore is 8 nm, allowing for multiple bases to be present in the pore at the same time. The homopolymer memory strands of the invention address this issue through the use of homopolymer repeats, decreasing the need for strict sequencing accuracy. In certain embodiments, the shortcomings of nanopore sequencing may be addressed by implementing a hairpin adapter to one end of a double stranded DNA memory strand such that during the translocation and base calling process, each sense of the DNA memory strand could be read such that reading an individual base and its complementary strand could compensate for the error rate of reading each base only once. In certain embodiments, nanopore sequencing fidelity may be increased by appropriate modification of each end (5'- & 3'-) of a single-stranded or double-stranded nucleic acid molecule with a bulky appendage that will not translocate across a

pore (e.g., protein or solid state). The molecule may then be trapped within a pore and translocated forwards and backwards many times to allow multiple reads of the same molecule in the same pore thus reducing the sequencing error rate by the square of the number of reads (if the sequencing read errors are due to stochastic origins).

Transferases like TdT may be used to append large and bulky modified nucleotide analogs to the 3'- end of a DNA molecule. In certain embodiments, a WORM nanopore memory device may be generated using the following steps: (1) generating a single molecule of DNA encoding specific information in any high density encoding scheme as discussed above and covalently modifying the 5'-end with a bulky molecular construct that prevents complete translocation of the DNA molecule through a nanopore; (2) threading the DNA molecule through the nanopore until the 5'-modified end is in contact with the nanopore and it cannot translocate any further; (3) using TdT and a modified nucleotide to covalently add one (or more) bulky nucleotide analogs ("stoppers") to the 3-end of the DNA molecule to effectively trap the molecule within the torus of the nanopore; (4) reversing the polarity of the current to the nanopores to clear out any DNA molecules that are not 3'- modified thus creating a pure population of "trapped" (5'- & 3'-modified) nucleic acid strands; (5) removing any un-trapped nucleic acids from the vicinity of the nanopore through washing or other means; (6) reading the "trapped" DNA strand in either or both directions, using an applied voltage, (potentially reading multiple times to reduce the error rate to an acceptable level). In various embodiments, step 6 may consist of a voltage induced "read" in one direction, and rapid translocation in the opposite direction to "rewind" the data encoding nucleic acid through the nanopore followed by another voltage induced "read" in the original direction. This cycle of "read" – "rewind" – "read" can be repeated as many times as desired.

In some embodiments, the trapped nucleic acid strand may be read during translocation in either direction. In some embodiments, the trapped strand can be translocated to one end of the molecule (either 5'- or 3'-) and read in the opposite direction as such a polarity of reading may provide a higher accuracy read.

In certain embodiments, a circularized nucleic acid memory strand may be generated using synthesis methods described above followed by circularization. The circularized strand may comprise a bulky macromolecule or specific homopolymer sequence where the ends of the synthesized strand were joined in order to designate a start and stop point for data reading. Start and stop homopolymer sequences may also be used in linear nucleic acid strands. The circularized strand 305 may have been threaded between two adjacent nanopores (306 and 309) such that the circular strand 305 is physically trapped between the two nanopores (306 and 309) located on a

single membrane 303 as shown in FIG. 6. The circularized memory strand may encode digital information as either sequences of single nucleotides, homopolymer tract sequences, sequences of modified nucleotide analogs, or some combination thereof. One nanopore 309 may be used to generate an electrical signal as the information-encoded memory strand is translocated though the pore, while the other nanopore 307 may simply act as a portal to allow the DNA molecule to return to the cis side of membrane 303 and first nanopore 309. One advantage of this scheme is that the information encoding strand may be recycled for repeated reading and can be read multiple times to thus reduce any possible read-out error.

Other embodiments may encode data in the memory strands so that it can be accessed only under a specific set of conditions. In such cases the memory strands are at least partially comprised of nucleotides containing modifications attached with a cleavable linker. Modifications (e.g., chemical protecting groups) and linkers can be selected so that if a polymer tract translocates through a nanopore without the correct treatment, the current blockades differ from the sequence that encodes the data. FIG. 7 outlines a scheme using disulfide and amide-linked modifications to a dG nucleotide and illustrates how the current blockade and data encoded in the memory strand may change in response to treatment conditions. G* and G** are structurally similar in size and flexibility and may produce similar current blockades on nanopore platforms, yet are removed under different conditions. G* and G*** are structurally different yet the modifications share the same removal conditions. Other embodiments may employ other modifications or linkers that are cleavable with different treatments such as specific wavelengths of light, acidic or alkaline pH, oxidative or reductive conditions, or sequence-specific nucleases. Some embodiments may use the presence or absence of memory strand modifications for encryption or as a chemical marker of previous access or alteration to the data. Most linker cleavage reactions are effectively irreversible, so this approach is best suited for write-one read many systems where single molecules may be sufficient to encode data without redundancy.

Many possible information encoding schemes which are useful with readout schemes of the invention are possible and may be apparent to one skilled in the art based on the present disclosure.

Synthesis may be accomplished using acoustic delivery of drops into wells of plates (e.g., 1536 well plates of 1.5 μL each). . In various embodiments, nucleic acid memory strands may be synthesized on a bead or a magnetic bead or a surface and either left on the bead or magnetic bead or surface after the full-length synthesis is complete or removed from the synthesis support depending on the application.

In certain embodiments, systems for the synthesis of long (5-10 kb) data strands may use

inkjet delivery to arrays of wells (e.g., multiple nanoliter volume wells). In other embodiments, multiple pneumatically controlled actuators can be positioned above each well to simultaneously deliver reagents to each position of an array. Each actuator would be served by a selector valve that would choose between each of the two or more nucleotides or modified nucleotides formulated with a template-independent polymerase that are used to specify the bits of the DNA data strand. One or more additional selector valve ports would be dedicated to one or more wash reagents if necessary. The array of nanoliter volume wells can be open at both ends, as long as the diameter of the wells is such that delivered liquids are trapped by capillary action within the length of the open-ended wells. After each round of nucleotide-enzyme formulation is delivered to the open-ended well, a rinse reagent or enzymatic reaction stop reagent can be flowed across and through the lower opening of the array of wells such that each well is rinsed of the reaction mixture thus preparing the array for the next cycle of enzymatic synthesis. In other embodiments, a vacuum source is used to rapidly remove one reagent from the capillary nanowells prior to delivery of the next reagent.

Certain embodiments may use highly parallel nanofluidic chambers with valve-controlled reagent deliveries. An exemplary microfluidic nucleic acid memory strand synthesis device is shown in FIG. 8 for illustrative purposes and not to scale. Microfluidic channels 255, including regulators 257, couple reservoirs 253 to a reaction chamber 251 and an outlet channel 259, including a regulator 257 to evacuate waste from the reaction chamber 251. Microfluidic devices for nucleic acid memory strand synthesis may include, for example, channels 255, reservoirs 253, and/or regulators 257. Nucleic acid memory strand synthesis may occur in a microfluidic reaction chamber 251 which may include a number of anchored synthesized nucleotide initiators which may include beads or other substrates anchored or bound to an interior surface of the reaction chamber and capable of optionally releasably bonding a polynucleotide initiator. The reaction chamber 251 may include at least one intake and one outlet channel 259 so that reagents may be added and removed to the reaction chamber 254. The reaction chamber 251 should be temperature controlled to maintain optimal and reproducible enzymatic synthesis conditions. The microfluidic device may include a reservoir 253 for each respective dNTP or analog to be used in the memory chain coding scheme. Each of these reservoirs 253 may also include an appropriate amount of TdT or any other enzyme which elongates DNA or RNA strands in a template- independent manner. Additional reservoirs 253 may contain reagents for washing or other tasks.

The reservoirs 253 can be coupled to the reaction chamber 254 via separate channels 255 and reagent flow through each channel 255 into the reaction chamber 254 may be individually

regulated through the use of gates, valves, pressure regulators, or other means. Flow out of the reaction chamber 254, through the outlet channel 259, may be similarly regulated. The reservoirs 253 may hold dNTPs, modified dNTPs or any analogs thereof described above suspended in a fluid at a known concentration such that the concentration of reagent may be strictly controlled based on the volume of reagent allowed to flow into the reaction chamber 254. Accordingly, the length of each homopolymer tract may be managed through control of the reagent concentration.

In certain instances, reagents may be recycled, particularly the dNTP and enzyme reagents. Reagents may be drawn back into their respective reservoirs 253 from the reaction chamber 254 via the same channels 255 through which they entered by inducing reverse flow using gates, valves, vacuum pumps, pressure regulators or other regulators 257. Alternatively, reagents may be returned from the reaction chamber 254 to their respective reservoirs 253 via independent return channels. The microfluidic device may include a controller capable of operating the gates, valves, pressure, or other regulators 257 described above.

An exemplary microfluidic nucleic acid memory strand synthesis reaction may include flowing a desired dNTP (used throughout to refer reference any component molecule used to encode data in a nucleic acid memory chain of the invention) reagent into the reaction chamber 254 at a predetermined concentration and for a predetermined amount of time (calculated to result in the desired homopolymer length) before removing the NTP reagent from the reaction chamber 254 via an outlet channel 259 or a return channel (not shown); flowing a wash reagent into the reaction chamber 254; removing the wash reagent from the reaction chamber 254 through an outlet channel 259; flowing the next NTP reagent in the desired memory strand sequence under conditions calculated to achieve the desired homopolymer tract ratio; and repeating until the desired nucleic acid memory strand has been synthesized. After the desired nucleic acid memory strand has been synthesized, it may be released from the reaction chamber anchor or substrate and collected via an outlet channel 259 or other means.

Because of the significant number of homopolymer encoded DNA strands required to encode useable amounts of data, highly parallel methods of DNA synthesis are required. In some embodiments, as depicted in FIG. 9, a flow cell containing an array of wells (9-1) are formed on a suitable substrate by patterning horizontal and vertical stripes (9-2) of hydrophobic materials to form a plurality of hydrophilic wells (?) bordered by hydrophobic regions. Typical dimensions of the hydrophilic wells can be 300 x 300 nm to 1000 x 1000 nm. This hydrophilic array forms the floor of a flow cell with a gap of suitable dimensions between the floor and an optically transparent cover. A solution of cold (i.e., below the enzyme -specific temperature optimum) nucleotidyl

transferase, one of natural or modified nucleotide triphosphates and any necessary co-factors, is flowed into the flow cell through an inlet (9-3) such that upon cessation of fluid flow, the enzyme-nucleotide triphosphate solution beads up into spatially defined droplets (9-4) positioned above each hydrophilic region. An IR source (9-5) through a shaping lens (9-6) projects a beam (9-7) onto a DLP (digital light projection) device (9-8), which is used to simultaneously steer IR beams (9-9) to each of the hydrophobically constrained droplets (9-4) that is chosen to have a specific nucleotide added, resulting in the rapid heating of the polymerase extension reaction formulation to the temperature for maximum enzyme activity for a period of time defined to synthesize a homopolymer of the desired length. After some suitably defined reaction time, the IR source is shut off and a cold rinse buffer is rapidly injected into the flow cell and drained through an outlet (9-10), thus quenching the reaction and finishing one "write" cycle. This series of steps is repeated multiple times for each "write" cycle so that each nascent data strand is randomly accessed according to its spatial location and the chosen nucleotide to be added, until the full length homopolymer data strand is completed. In some embodiments, a DLP device with 1920 x 1080 steerable mirrors can be used to simultaneously randomly access ~2M synthesis positions in the synthesis flow cell. In some embodiments, the template-independent polymerase used is thermophilic with a reaction temperature optimum well above room temperature, such that enzymatic activity is minimized in the hydrophobically constrained droplets prior to the rapid heating of the droplet by the IR source. In some embodiments, the bottom surface of the flow cell, bearing the hydrophobically defined hydrophilic wells, is abutted to a cooling device that maintains the droplets in the hydrophobic wells at a reduced temperature to prevent enzymatic activity until the temperature is raised by the IR source.

In another embodiment, a flow cell composed of an array of wells that are formed on a suitable substrate by patterning horizontal and vertical stripes of hydrophobic materials forming a plurality of hydrophilic spots bordered by hydrophobic regions. Typical dimensions of the hydrophilic spots can be 300 x 300 nm to 1000 x 1000 nm. Each hydrophilic spot is positioned over an individually addressable CMOS heater. This hydrophilic-CMOS heater array forms the floor of a flow cell with a gap of suitable dimensions between the floor and an optically transparent cover. A solution of cold (i.e., below the enzyme -specific temperature optimum) nucleotidyl transferase and one of natural or modified nucleotide triphosphates is flowed into the flow cell such that upon cessation of fluid flow, the polymerase extension reaction solution beads up into spatially defined droplets positioned above each hydrophilic region with an associated CMOS heater. Each of the hydrophobically constrained droplet that is chosen to have a specific nucleotide added is

rapidly heated to the temperature for maximum enzyme activity for a period of time defined to synthesize a homopolymer of the desired length. After some suitably defined reaction time, the heater is shut off and a cold rinse buffer is rapidly injected into the flow cell, thus quenching the reaction and finishing one "write" cycle. This series of steps is repeated multiple times for each "write" cycle so that each nascent data strand is randomly accessed according to its spatial location and the chosen nucleotide to be added, until the full length homopolymer data strand is completed. In some embodiments, the enzyme used is thermophilic with a temperature optimum well above room temperature, such that there is a low probability of unwanted nucleotide addition in the hydrophobically constrained drop prior to the rapid heating of the droplet by the CMOS heater.

As one skilled in the art would recognize as necessary or best-suited for the systems and methods of the invention, systems and methods of the invention may include computing devices as shown in FIG. 10 that may include one or more of processor 309 (e.g., a central processing unit (CPU), a graphics processing unit (GPU), etc.), computer-readable storage device 307 (e.g., main memory, static memory, etc.), or combinations thereof which communicate with each other via a bus. Computing devices may include mobile devices 101 (e.g., cell phones), personal computers 901, and server computers 511. In various embodiments, computing devices may be configured to communicate with one another via a network 517.

Computing devices may be used to control the synthesis of memory strands, the reading of sequenced memory strands, and the compiling or translating of data between human or machine-readable formats, digitized data, and nucleic acid sequences among other steps described herein. Computing devices may be used to display the readable format of data.

A processor 309 may include any suitable processor known in the art, such as the processor sold under the trademark XEON E7 by Intel (Santa Clara, CA) or the processor sold under the trademark OPTERON 6200 by AMD (Sunnyvale, CA).

Memory 307 preferably includes at least one tangible, non-transitory medium capable of storing: one or more sets of instructions executable to cause the system to perform functions described herein (e.g., software embodying any methodology or function found herein); data (e.g., data to be encoded in a memory strand); or both. While the computer-readable storage device can in an exemplary embodiment be a single medium, the term "computer-readable storage device" should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the instructions or data. The term "computer-readable storage device" shall accordingly be taken to include, without limit, solid-state memories (e.g., subscriber identity module (SIM) card, secure digital card (SD card), micro SD

card, or solid-state drive (SSD)), optical and magnetic media, hard drives, disk drives, and any other tangible storage media.

Any suitable services can be used for storage 527 such as, for example, Amazon Web Services, memory 307 of server 511, cloud storage, another server, or other computer-readable storage. Cloud storage may refer to a data storage scheme wherein data is stored in logical pools and the physical storage may span across multiple servers and multiple locations. Storage 527 may be owned and managed by a hosting company. Preferably, storage 527 is used to store records 399 as needed to perform and support operations described herein.

Input/output devices 305 according to the invention may include one or more of a video display unit (e.g., a liquid crystal display (LCD) or a cathode ray tube (CRT) monitor), an alphanumeric input device (e.g., a keyboard), a cursor control device (e.g., a mouse or trackpad), a disk drive unit, a signal generation device (e.g., a speaker), a touchscreen, a button, an accelerometer, a microphone, a cellular radio frequency antenna, a network interface device, which can be, for example, a network interface card (NIC), Wi-Fi card, or cellular modem, or any combination thereof.

One of skill in the art will recognize that any suitable development environment or programming language may be employed to allow the operability described herein for various systems and methods of the invention. For example, systems and methods herein can be implemented using Perl, Python, C++, C#, Java, JavaScript, Visual Basic, Ruby on Rails, Groovy and Grails, or any other suitable tool. For a computing device 101, it may be preferred to use native xCode or Android Java.

FIG. 11 shows the polyacrylamide gel electrophoresis analysis of two different single homopolymer tracts made via enzymatic synthesis. Lane A is a sample of a starting 20-mer oligonucleotide that is used in all following lanes. Lane B is a sample from a TdT reaction containing a 20-mer oligonucleotide and the non-reversible terminator ddATP showing the formation of a 21-mer. Lane C is a sample from a TdT reaction containing a 20-mer and the natural nucleotide dATP after 1 minute at 37°C. Lane D is a sample of the same reaction mixture in Lane D after 5 minutes at 37°C. Lane E is a sample of the same reaction mixture in Lane C after 15 minutes at 37°C. Those three lanes illustrate homopolymer length control due to consumption of the input dATP in ~5 minutes during the TdT extension reaction and the observation of no homopolymer growth between 5 and 15 minutes. Lane F is a sample from a TdT reaction containing a 20-mer oligonucleotide and the nucleotide analog N6-benzoyl-dATP after 1 minute at 37°C. Lane G is a sample of the reaction mixture in Lane F after 5 minutes at 37°C. Lane H is a

sample of the same reaction mixture in Lane F after 15 minutes at 37°C. Although there are qualitative differences between the length of dA^{Bz} homopolymers formed in Lanes F-H, the same length control is demonstrated even with an N6-modified dATP analog.

The writing of digital data into a molecular storage format based on molecular approaches offers advantages over currently used storage media like tape or disk. DNA based storage has been sparking interest because of the high information density achievable, the extremely long lifetime and the low energy consumption during dormant periods. To date, most efforts to use synthetic DNA as a storage media have involved chemical synthesis using the popular phosphoramidite method.

DNA based data storage can require vastly larger number of strands than are currently synthesized for the existing research markets. Any synthesis technology (chemical or enzymatic) that depends upon the removal of a nucleotide blocking or terminator imposes additional steps and complexity since reagents must be deliver to an array of synthesis features (i.e., wells or spots) in an addressable fashion to direct the correct nucleotide to the correct location on an array. Array based methods of synthesis can be performed on one of several ways: 1) bulk delivery of activated reactants followed by selective removal of a blocking group or 2) addressable delivery of activated reactants followed by bulk blocking group removal or 3) bulk delivery of inactive reactants followed by addressable activation. The addressable delivery of reagents to large (10^4-10^6) 2-D arrays is generally achieved using inkjet deposition to each desired location. If the data encoding scheme uses four nucleotides, then four separate write heads must be used and indexed in a complex X-Y mechanical fashion. Additionally, the use of inkjet delivery limits the dimensions between each feature (well or spot) to low tens of microns. The challenge in this process is to decrease the step time for each synthesis cycle, no matter how it is achieved.

In certain embodiments, systems and methods of the invention may include delivering an inactive reaction mixture to every feature on a 2-D array, then selectively activating only those feature that require the addition of an A or G or C or T. Addressable methods of delivery, activation or blocking group removal that do not rely on mechanical movement are preferred. A preferred embodiment uses bulk delivery of reactants and selective activation of specific synthesis features by removal of a blocking group from a nucleotide analog that then allows rapid DNA polymerase mediated incorporation and formation of a homopolymer bit as shown in FIG. 15, thus enabling the highly parallel and rapid synthesis of homopolymer encoded nucleic acid memory strands. There are several other advantages to homopolymer encoded nucleic acid information polymers: 1) homopolymer encoded bits overcome the error profiles associated with next

generation sequencing, 2) the resulting polymers are unnatural nucleic acids and thus cannot be repurposed for bioterrorist activities.

Delivery and selective activation of template independent polymerase DNA synthesis can be done serially (deliver A ->addressably activate and initiate homopolymer synthesis -> wash; deliver C ->addressably activate and initiate homopolymer synthesis -> wash; deliver G -> addressably activate and initiate homopolymer synthesis ->wash; deliver T -> addressably activate and initiate homopolymer synthesis -> wash) or in parallel (deliver all four nucleotides to all features simultaneously -> addressably activate A, C, G, T either simultaneously or serially -> wash). In this fashion, the synthesis cycle becomes very efficient and involves only three steps: reactant delivery, incorporation reaction activation, then wash before the next cycle starts. The reaction is halted by either a rapid removal out of the reactants by a gas or a liquid or by the rapid delivery of a quenching reagent. In a preferred embodiment, the quenching reagent is a metal chelator.

A preferred design for an apparatus that can be used to synthesize homopolymer bit encoding memory strands consists of a flow cell that is composed of a 2-D array of hydrophobically patterned wells that are suitably modified to support template independent enzymatic synthesis. After delivery of reactants to the 2-D array of hydrophobic wells, liquids will bead up defining spatially distinct reaction zones as shown in FIG. 16.

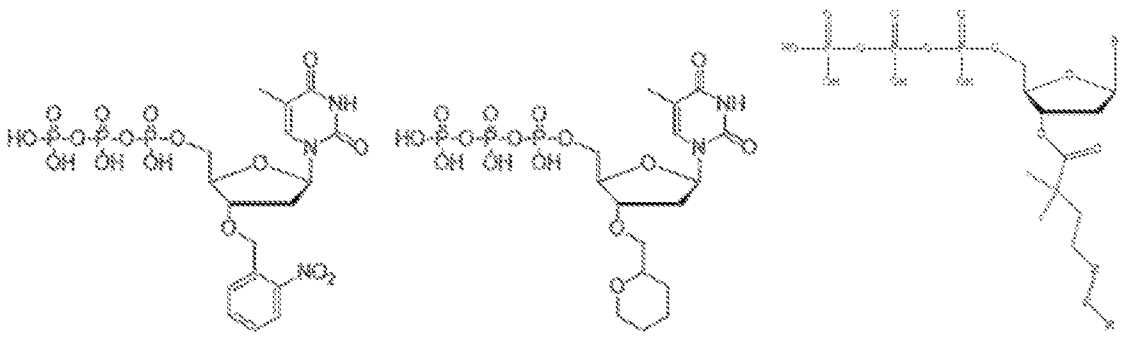
A preferred embodiment is one in which the bottom surface of the well, enclosed on four sides by hydrophobic patterning, is modified with a covalently attached oligonucleotide initiator, in a 5'→3'orientation with the 5'-termini attached to the bottom surface of the well. In another embodiment, the well is physically formed by etching a cavity in the bottom surface of the flow cell, in which case the covalently attached oligonucleotide initiator is covalently attached to the surfaces (bottom and sides) of the well. In some cases, the well is open at the bottom to allow for liquid flow through the well. In other cases, the well is closed at the bottom.

Nucleotide analogs that are protected at the 3'-OH are generally inactive with commercially available or WT TdT enzyme (US 10,059,929). Thus, a mixture of a 3'-blocked dNTP analog, TdT protein and suitable co-factors, can be mixed together in the presence of an initiator oligonucleotide, at 37°C with little to no homopolymer formation. Once the 3'-OH is "decaged" (i.e., unblocked) by removal of the protecting or blocking group, the resulting nucleotide is available for free running incorporation and homopolymer formation. Decaging that can be accomplished by "deliveryless" methods are preferred. Analogs that are constructed to allow addressable 3'-OH decaging by administration or exposure to an activating energy such as light,

heat, electrochemical generation of pH change, or a reducing agent are preferred. Each of these decaging or unblocking reactions can be accomplished in flow-cells that are appropriately constructed to with either mechanically steerable light through light transparent covers (e.g., FIG. 17), individually addressable heater (e.g,. FIG .18), electrochemically induced pH change or the generation of reducing conditions.

Each of the flow cell designs illustrated is compatible with a method of addressable decaging of a nucleotide contained in a droplet constrained by the hydrophobic patterning around it. FIG. 19 illustrates an apparatus that could be used with a flow cell designed for the use of light decaged nucleotide analogs. Other embodiments and configurations known to one skilled in the art are possible for this purpose.

Each decaging mechanism requires nucleotide analogs specifically designed for the physiochemical process selected for use in the system:



In some embodiments, nucleotide analogs suitable for light mediated decaging are 3'-O-(2-nitro-benzyl)-dNTP. In some embodiments, nucleotide analogs suitable for heat mediated decaging are 3'-O-(tetrahydrofuranyl)-dNTP. In some embodiments, nucleotide analogs suitable for reduction mediated decaging are 3'-O-methyl-dithiomethyl-dNTP. Many other 3'-OH protecting groups are suitable, as long as the resulting 3'-OH modified nucleotide analog is not a substrate for a template independent polymerase and is readily removed by "delivery-less" methods like light, heat or electrochemically generated reactants. The composition of these dNTP analogs is different than that described in WO 2016/034807, in so far as the 3'-O modifications described therein are explicitly stated to be substrates for a template independent polymerase and are called reversible terminators. The subject of this invention are 3'-O-blocking groups that are explicitly not substrates for a polymerase and serve to cage the dNTP until removed thus allowing polymerization to proceed. Mathews AS et al (2016) describe the use of 3'-O-(2-nitrobenzyl)-dNTP analogs as reversible terminators for the controlled enzymatic synthesis of natural, non-homopolymer oligonucleotides. They explicitly teach the use of such nucleotides as substrates for DNA synthesis

by using extremely long (ca 1 hour) enzymatic reaction times and furthermore explicitly teach their use as reversible terminators in contrast to the subject of this patent, which teaches the use of these analogs as caging groups to initiate enzymatic polymerization of multiple nucleotides.
Initiation of free running homopolymer synthesis can be achieved by other methods than caged dNTP analogs. Some embodiments can use fluidic pulses of polymerase to start and stop ssDNA synthesis as described in Church 9,928,869 (2018). In Reza et al WO 2017/196783, template dependent enzymatic synthesis is initiated by activation of polymerase by an electrochemically generated pH change. Although modified nucleotide analogs, including 3'-O-reversible terminators are described, neither patent teaches the activation of homopolymer synthesis using decaging of caged dNTP analogs. Church 9,928,869 (2018) describes free running homopolymer synthesis using natural dNTPs and controlling the length of homopolymers formed by the reaction duration. Lee HR et al (2018) describes the use of mechanical delivery to deposit natural nucleotides to a plurality of reaction zones on a 2D array and control the length of homopolymer formation by actively destroying the unreacted unmodified dNTP with apyrase. In some embodiments of the subject of this patent, homopolymer rate modulating modified dNTP analogs with unmodified 3'-OH are used in combination with either fluidic pulses or mechanical XY delivery or pH activated template independent DNA polymerases.

FIGS. 20 and 21 show the kinetics of light mediated decaging. As the 3'-caged dNTP in the template independent polymerase reaction mixture is converted to a substrate for the polymerase, it is polymerized into the nascent homopolymer strand. Complete conversion of the caged dNTP to uncaged (substrate) dNTP is not required as available uncaged dNTP is readily incorporated by the polymerase. Once decaged, the length of the desired homopolymer is controlled by controlling the concentration of decaged dNTP generated and/or the time of the enzymatically mediated incorporation reaction.

FIG. 22 shows examples of two nucleotide analogs that are caged with 3'-O modifications that are able to be decaged with visible wavelengths of light and exhibit tunable photophysical properties (Peterson J.A., et al J. Amer. Chem. Soc. 2018 140:7343-6), this allowing for the simultaneous introduction, decaging and synthesis of two different homopolymer strand sequences. In some embodiments, nucleotide analogs modified with two different 3'-caging species, capable of being decaged by two different wavelengths of light, are delivered simultaneously to a 2D array of synthesis locations. In another embodiment, four dNTP analogs modified with four different light removable, 3'-caging species are delivered simultaneously to a 2D array of synthesis locations and decaged with four different wavelengths of light. Decaging reactions can be carried out

sequentially on one set of locations on the 2D array followed by delivery of the two remaining dNTPs and subsequent sequential decaging, thus completing one round of increasing the length of all information polymers by one nucleotide. In an alternative embodiment, four dNTP analogs modified with four different 3'-caging species can be delivered simultaneously and decaged by the simultaneous exposure of each synthesis feature to one of four different wavelengths of light. A hardware configuration like that shown in FIG. 19 can be used for multicolor decaging by the incorporation of 2 or more wavelength specific light sources and appropriate shuttering mechanism to direct one of the 2 or more light sources onto the 2D light directing system.

FIG. 23 shows the synthesis of a 12-bit information strand consisting of consecutive cycles of homopolymer synthesis using dCTP, dGTP, dTTP interspersed with UV light decaged 3'-oNBn-dATP at cycles #1, 6, 8, 10, 12. The incorporation reaction can be terminated by several methods, including but not limited to the rapid removal of the enzymatic reaction components either by introduction of a bolus of a gas or a liquid. In some embodiments the liquid simply rinses away the reaction components while in other embodiments, the rinse liquid contains active quenching agents like EDTA or other enzymatic inhibitors.

In another embodiment, the method of caging the dNTP can be mediated by a steric mechanism that involves a modification to the nucleotide base instead of the 3'-OH. In a preferred embodiment, the nucleotide, with an unmodified 3'-OH, is modified with a removable steric blocking group (incorporation blocker) at the N6, N4, N2, O4 of A, C, G, or T respectively that cages the dNTP. Homopolymer synthesis can be initiated by either light mediated, heat mediated or reduction mediated cleavage of the steric blocking group, thus "uncaging" a natural nucleotide suitable for free running homopolymer synthesis.

In some embodiments, the modified nucleotide is comprised of a linker containing one portion that renders the dNTP caged and thus un-incorporable until removed and another portion remains covalently bound though out the multiple homopolymer tract synthesis. In some embodiments, the purine or pyrimidine base is modified at two locations, one modification caging the nucleotide and preventing enzymatic incorporation, the other remaining covalently bound throughout the polymer synthesis; each modification removable by different mechanisms allowing selective removal of each. In some embodiments, the purine or pyrimidine base is modified at two locations, each one removable by different mechanisms allowing selective removal of each.

Suitable dNTP analog modifications can be designed that will result in a nucleotide with either no scar or with a scar depending upon the sequence detection modality to be used for reading the homopolymer bit data strand. For readout methods that rely on Sequencing By Synthesis

(SBS), all modifications to the purine or pyrimidines must be removed after homopolymer synthesis but prior to SBS readout. For readout methods that rely on current modulation during polymer translocation through one or more nanopores, modifications to purines and pyrimidines that are readily distinguishable from each other are desired. In some embodiments, modifications are made to the purine and pyrimidine base that modulate the polymerase kinetics during homopolymer synthesis. These "rate modulating" modifications may also act as current modulators for nanopore detection, or they may be removed for SBS detection.

Regardless of the mechanism of activation, controlling the length of free running homopolymer synthesis is important. Ideally, homopolymers of 2 to 4 nucleotides are desirable. In the presence of natural nucleotides, TdT has a propensity for forming homopolymers based on the nucleobase corresponding to the Km of the individual nucleotides (A>T>G>C). As noted by Lee HR et al (2018), it is difficult to limit homopolymer growth without increasing the deletion frequency. The above authors report ~66% del errors in their attempts to limit homopolymer growth to 2-3 nucleotides. Although TdT is reported to operate in a distributive manner, leading to Poisson distributions of homopolymer length in free running synthesis, in practice it is hard to drive complete conversion of the initiator nucleic acid during homopolymer synthesis without long enzymatic extension reaction times. Too short of free running homopolymer synthesis reaction time leads to bit deletion errors as reported above. Too long of a homopolymer synthesis reaction time leads to excessively long homopolymer formation and inefficient data density. One solution is to use modified nucleotide analogs that modulates the rate of multiple base incorporation but don't require removal at every step, thus allowing for complete conversion of the initiating nucleic acid, control of the length of the homopolymer formed, and preserving the simple two step cycle that is the subject of this patent. Ideally, homopolymer tracts in information polymers are greater than two nucleotides but four or less nucleotides in length.

If necessary, the homopolymer length limiting modifications can be removed from all nucleotides at the end of the synthesis, thus generating a natural DNA molecule, suitable for SBS detection. Modifications that are chemically compatible with the decaging conditions described above are most desirable; they should be removable by chemical conditions that are orthogonal to those used for decaging. In some embodiments, the 3'-OH is caged by 3'-O-(2-nitrobenzyl), while the N6 of dA, the N4 of dC, the N2 of dG and the O4 or N3 of dT are modified with non-terminating moieties that modulate multiple additions during free running enzymatic homopolymer synthesis. FIG. 24 shows the comparison of free-running homopolymer synthesis is the presence and absence of rate modulating modifications. In some embodiments, the homopolymer

synthesis rate modulating modifications are the same for all four nucleotide analogs, while in another embodiment, the modifications are different for each of dATP, dCTP, DGTP & dTTP analogs. In some embodiments, the rate limiting modifications remain in place during the entire synthesis.

FIGS. 25 and 26 show the results of two consecutive addition cycles of two differently modified nucleotide analogs. In another embodiment, the nascent information polymer is exposed to periodic or alternating cycles with chemical conditions that result in partial or complete removal of the homopolymer rate limiting modifications. In some embodiments of the use of Class 2 dNTP analogs, previously incorporated rate modulating modifications are removed simultaneously with polymerase extension by the inclusion of a mild reducing agent in alternating cycles of enzymatic extension reaction mixture.

If a non-SBS method of readout is to be used (i.e., nanopores), the homopolymer rate modulating modifications could be covalently attached with non-cleavable linkages and left in place during the readout. In some embodiments, homopolymer rate modulating modifications are also used to encode information during the readout process. In a preferred embodiment, the homopolymer synthesis rate modulating analogs are also designed to modulate the current during nanopore translocation and furthermore, are selected to provide two or more detectable and distinguishable current blockade levels. In another embodiment, only a single type of nucleotide (i.e., A or C or G or T) is used in information strand synthesis and the homopolymer bits are encoded by two or more differential current blockade generating modifications.

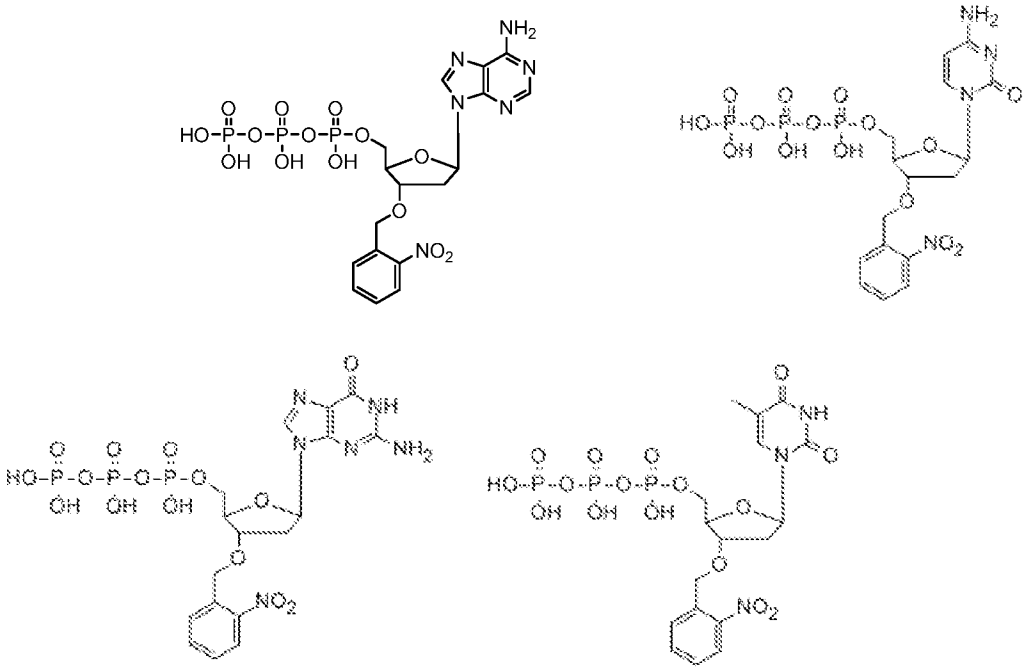
FIGS. 27-30 shows an example of the detection by a solid state nanopore of an enzymatically synthesized homopolymer of peptide-modified dNTP analogs. A dTTP analog modified with a cleavable disulfide linkage to a five amino acid peptide (N-Ac-CYPEE), was used to generate a fully modified molecule via free running homopolymer synthesis of >100 nt in length. Translocation through a 2D silicon nitride 30 nm thick nanopore, at 500 mV resulted in a large signal deflection compared to an unmodified dU homopolymer (courtesy of Goeppert LLC, Philadelphia, PA).

The following table describes six different classes of modified dNTP analogs that are useful for the three different “delivery-less” homopolymer synthesis activation approaches and two different homopolymer encoded nucleic acid memory strand readout technologies.

	SBS or nanopore detection				Only nanopore detection	
	Analog Class I	Analog Class II	Analog Class III	Analog Class IV	Analog Class V	Analog Class VI

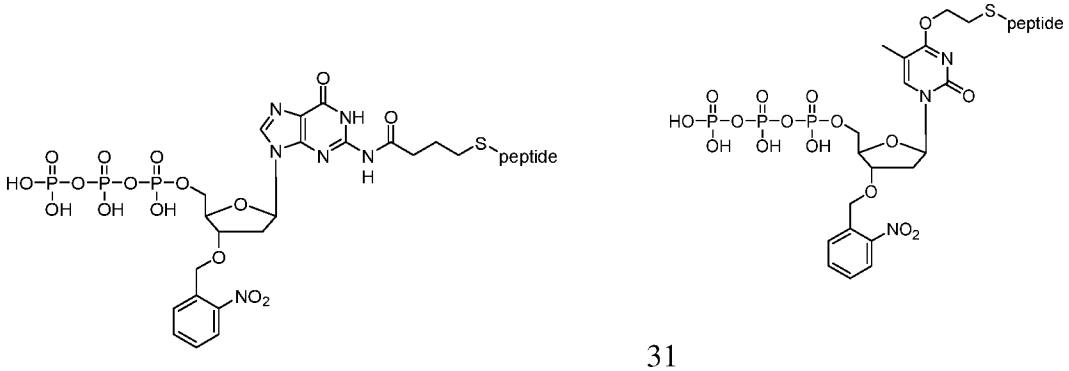
dNTP Analogs	dATP, dCTP, dGTP, dTTP	dATP, dCTP, dGTP, dTTP	dATP, dCTP, dGTP, dTTP	dATP, dCTP, dGTP, dTTP	Only one dNTP	Only one dNTP
3'-OH	Removable incorporation blocker	Removable incorporation blocker	unmodified	unmodified	Removable incorporation blocker	unmodified
Base	unmodified	Removable incorporation rate modulator	Removable incorporation blocker	Removable incorporation blocker	>2 differentially detectable non-removable modifications	>2 differentially detectable non-removable modifications
				orthogonally removable incorporation rate modulator		orthogonally removable incorporation blocker

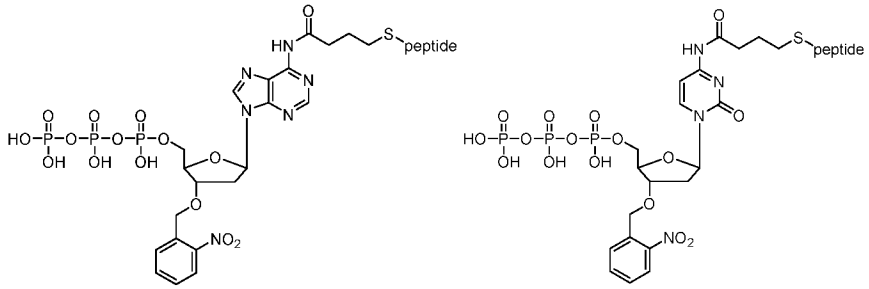
Below are shown examples of four light mediated decaging nucleotide analogs that make up Class I:



This class of analogs demonstrates 3'-O-caged dNTPs that become natural non-modified dNTPs upon light mediated decaging. Once decaged, the rate of homopolymer formation will be no different than that of a natural nucleotide and the length of the homopolymers formed must be modulated by additives to the template independent polymerase reaction formulation. In some embodiments, tetrahydropyranyl, 4-methoxy-tetrahydropyranyl, tetrahydrofuranyl, acetyl, methoxyacetyl, or phenoxyacetyl modifications are useful for the heat-triggered decaging of 3'-O blocked dNTP analogs. In some embodiments, 3'-O-analogs like -O-CH_2-S-S-R are useful for electrochemically mediated decaging by reductive conditions. Class I dNTP analogs are characterized by nucleotides that result in unmodified homopolymers that are suitable for readout by either classical Sequencing by Synthesis or by ratcheting style nanopores. In both instances, accuracy in the length of the homopolymers is not necessary, but the accurate detection of transitions between the homopolymers is required.

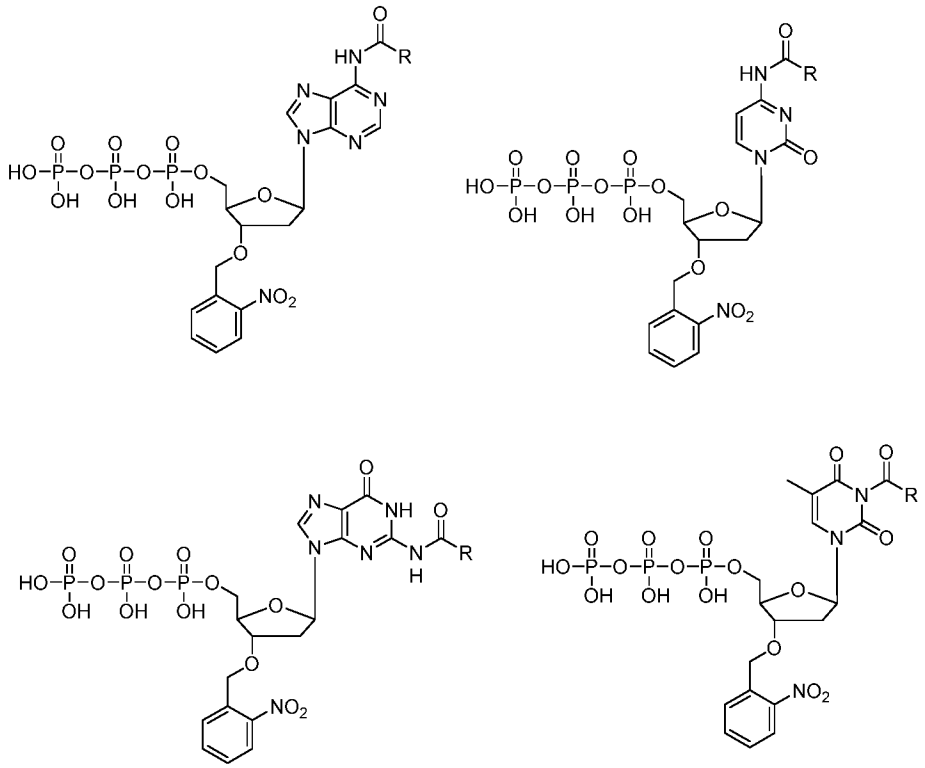
Below are shown examples of four Class II light mediated decaging nucleotide analogs, with orthogonally removable peptide rate modulating modifications:





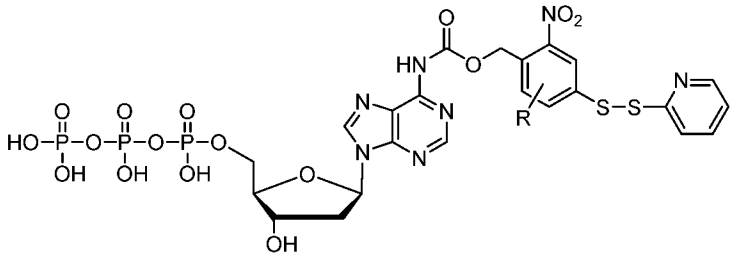
Peptide modifications for use with Class II dNTP analogs, are meant to be removed at the end of the information polymer synthesis for subsequent interrogation by either sequencing methods that rely on SBS or by translocation through a nanopore, sensitive enough to detect the transitions form one homopolymer type (A, C, G, T) to the next. Peptides suitable for use in this application, slow down the rate of incorporation of modified nucleotide analogs, to prevent long homopolymer formation prior to complete conversion of unmodified strands or modified strands of a different composition. In some embodiments, the incorporation rate modulating modifications may consist of 1, 2, 3, or 4 different peptides for the four nucleotide analogs. In some embodiments, the peptides are linked to the nucleotide through a disulfide linker, with a self-immolating scar, cleavable under mild chemical conditions. In some embodiments, the peptides may consist of, but are not limited to, Ac-EECGY, Ac-EEGCGW, Ac-EEGCGGW, Ac-EC-pNA, Ac-CWEE, Ac-CYPEE, Ac-EEGCPPW, Ac-CPYEE, Ac-CPWEE, Ac-CWPEE. Many other peptide sequences and compositions are possible to one skilled in the art. Peptides with an overall anionic composition seem most suitable; peptides with cationic composition accelerate the rate of incorporation of nucleotides, as previously noted by Finn PS et al 2003. Peptides with covalent linkage to the nucleotides other than through a disulfide to a cysteine amino acid are possible so long as the ability to remove them from the nucleotides of the completed homopolymer DNA strands, without leaving a residual scar is maintained. Particularly useful are disulfide cleavage mediated self-immolating linkers that eliminate under mild conditions by the formation of a thiolactone.

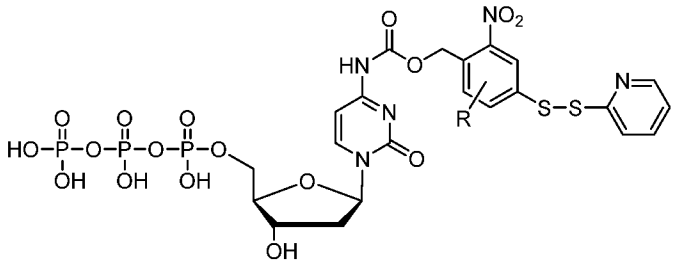
Below are examples of four nucleotides of Class II, with light mediated decaging and orthogonally removable non-peptide, homopolymer synthesis rate modulating modifications:



Non-peptides modifications to the N6 of Adenine, the N4 of Cytidine, the N2 of Guanine and the N3 of Thymine can act as incorporation rate modulators. In some embodiments, acetyl, diacetyl, isobutyryl, proprionyl, pivaloyl, benzoyl, cyclohexyl and other organic modifiers are useful. Modifiers of this type are removed post information polymer synthesis by but not limited to ammonolysis. Class II dNTP analogs are characterized by nucleotides that result in unmodified homopolymers that are suitable for readout by either classical Sequencing by Synthesis or by ratcheting style nanopores. In both instances, accuracy in the length of the homopolymers is not necessary, but the accurate detection of transitions between the homopolymers is required.

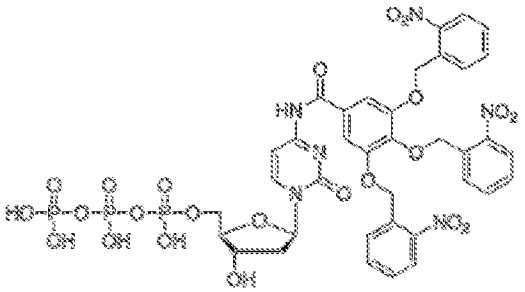
Below are examples of two Class III nucleotide analogs, with unmodified 3'-OH and removable incorporation caging modifications on a purine or pyrimidine:





In some embodiments, the caging modifications to the 2-nitrobenzyl function that are useful are bulky modifications like but not limited to peptides, cyclic peptides, PEG, branched PEG, star PEG, dendrimers, nanoparticles, Class III dNTP analogs are also characterized by nucleotides that result in unmodified homopolymers that are suitable for readout by either classical Sequencing by Synthesis or by ratcheting style nanopores. In both instances, accuracy in the length of the homopolymers is not necessary, but the accurate detection of transitions between the homopolymers is required.

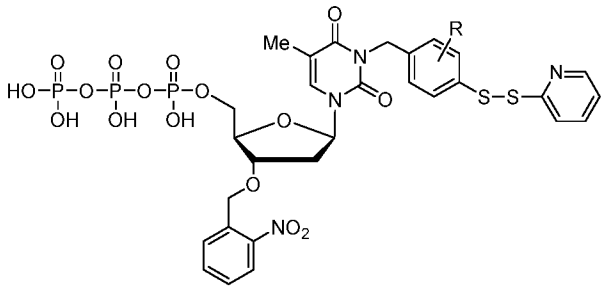
Below is an example of one Class IV nucleotide analog, with unmodified 3'-OH, removable incorporation blockers at positions other than the 3'-OH and orthogonally removable homopolymer synthesis rate modulator modifications also not located at the 3'-OH:

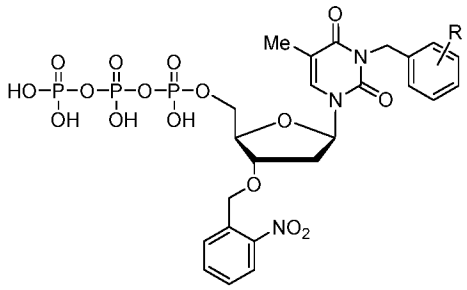


Nucleotide analogs of Class IV are designed to have one or more incorporation blocking modifications that cage the nucleotide from enzymatic incorporation until removed by light, heat or electrochemical means. Large, bulky modifications that render the nucleotide analog inactive to polymerase incorporation are preferred. In some embodiments, the one or more modifications are

derivatives of 2-nitrobenzyl that can be removed by photolysis. Removal of the caging modifications at every cycle, leaves incorporation rate modulating modifications that are removed at the completion of the information polymer synthesis, by a chemical method orthogonal to that used to decage the nucleotide analog. In a preferred embodiment, the incorporation blocking modifications are covalently attached to the incorporation rate modulating modifications. In some embodiments caging modifications to the 2-nitrobenzyl function that are bulky modifications and act as steric blockers are useful. Examples of modifications that can act as steric blockers are, but not limited to, peptides, proteins, peptoids, PEG, branched PEG, star PEG, dendrimers, or nanoparticles.

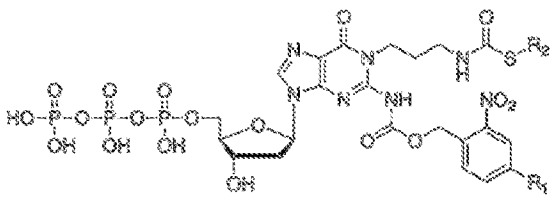
Below are examples of Class V pyrimidine nucleotides, with 3'-O-caging modifications and non-removable base modifications:

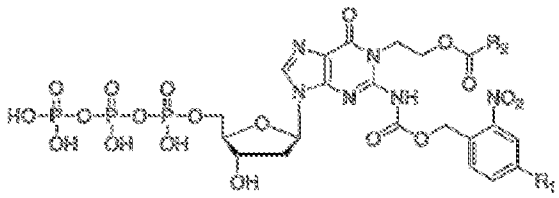




In the case where nanopore detection by direct translocation is desired as the readout method, the nucleotides are doubly modified with a suitable removable 3'-O-caging group and with a covalent, non-removable modification that acts as a blockade current modulator during direct translocation through a non-indexing nanopore. Suitable modifications may be peptides or non-peptides. The ideal embodiment for a blockade current modulating group is the smallest possible modification that gives the most unique signal in the shortest possible homopolymer stretch. A key innovation of this class of dNTP analogs is that only one type of modified nucleotide is required because the "sequence" of the homopolymer is encoded by the sequence of two or more current modulating modifications, not the nucleotide itself.

Below are examples of Class VI nucleotide analogs, with unmodified 3'-OH, removable incorporation blocking (caging) modifications and two or more different types of non-removable current modulating modifications, useful for detection by direct nanopore sequencing. A key innovation of Class VI dNTP analogs is that only one type of modified nucleotide is required, because the "sequence" of the homopolymer is encoded by the sequence of two or more current modulating modifications, not by the nucleotide itself. Encoding is not limited to base 2 or base 4, but only by the number of current modulating modifications that can be made to a single purine or pyrimidine nucleotide.





In some embodiments, the incorporation blocking (caging) modification is removable by either light, heat or electrochemical means, while the two or more nanopore detection elements are non-removable and survive repeated exposure to the decaging conditions used at every cycle during the homopolymer strand synthesis. Caging modifications to the 2-nitrobenzyl that act as steric blockers are bulky modifications, like but not limited to, peptides, proteins, peptoids, PEG, branched PEG, star PEG, dendrimers, or nanoparticles.

Examples

Example 1

N^6-benzoyl-deoxyadenosine triphosphate was prepared by charging a vial with N^6-benzoyl-2'-deoxyadenosine (0.055 g, 0.16 mmol) under dry N_2 and trimethyl phosphate (0.435 mL) was added. To the resulting solution was added tributylamine (0.077 mL, 0.32 mmol) and the reaction mixture was flushed with dry N_2 for 30 min while being held at -5°C. To this vial was added anhydrous phosphorous oxychloride (0.018 mL, 0.19 mmol) via syringe and the reaction mixture was stirred at -5°C for 3 min. A second aliquot of anhydrous phosphorous oxychloride (0.009 mL, 0.10 mmol) was added via syringe and the reaction mixture was stirred at -5°C for 8 min. A second vial was charged with tributylamine pyrophosphate (0.075 g, 0.14 mmol), flushed with dry N_2, and anhydrous acetonitrile was added (0.609 mL), followed by tributylamine (0.231 mL, 0.97 mmol). The prepared tributylamine pyrophosphate mixture was cooled to -20°C and added to the reaction mixture and allowed to react for 10 m. The reaction was quenched by the dropwise addition of H_2O (4.35 mL). The contents of the flask were combined with 0.87 mL of H2O and extracted with dichloromethane (3 x 150mL). The aqueous phase was adjusted to pH 6.5 with concentrated NH_4OH and stirred for 12 h at 4°C. The mixture was transferred to a 250 mL round bottom flask with 50 mL of water, and concentrated under reduced pressure. The residue was dissolved in 40 mL water, and purified via ion-exchange chromatography (AKTA FPLC, Fractogel DEAE 48 mL column volume, stepwise gradient 0 -> 70% TEAB in water, pH 7.5). Fractions containing the desired product were pooled, the concentration by A260, and concentrated under reduced pressure, with removal of residual triethylammonium bicarbonate via iterative

concentration from water (5 x 50 mL) to dryness to provide N^6-benzoyl-2'-deoxyadenosine triphosphate.

Controlled synthesis of homopolymer tracts, comprised of modified nucleotides, by a nucleotidyl transferase, TdT, was conducted in the following manner. Stock solutions of Deoxyadenosine triphosphate (TriLink Biosciences) and N^6-benzoyl-deoxyadenosine triphosphate at 1 mM each were prepared in H_2O.

0.5 µL (500 pmoles) of each of the different triphosphates was separately combined with 1.5 µL (30U) of commercially available TdT (Thermo Scientific), 0.5 µL (50 pmoles) of 5'-TAATAATAATAATTTTT-3' (IDT), 2 µL of commercially available TdT Rxn buffer (Thermo Scientific - 1M potassium cacodylate, 0.125M Tris, 0.05% (v/v) Triton X100, 5 mM C_oCl2 (pH 7.2 at 25°C)) and 4.5 µL of H2O. The reactions were incubated at 37°C. 30 µL aliquots were removed after 1 m, 5 m, & 15 m and each quenched with 20 µL 5 mM EDTA. Each sample was dried down under vacuum and reconstituted in 100 µL H2O. 10 µL of each timepoint was mixed with 10 µL of denaturing load buffer (100% formamide and 0.1% Orange G) and applied to the well of a 1mm x 20 cm x 14 cm 20% polyacrylamide gel. After electrophoresis at 400 V for 3.5 h, the bands were visualized with Sybr Gold (Thermo Scientific) and photographed under UV illumination (UV-blocking Wratten 2A filter 405 nm cutoff, UVP, LLC).

For synthesis of multiple homopolymer tracts, the initiator oligonucleotide may be attached to a bead to allow multiple rounds of enzymatic synthesis interspersed with removal of the previous reactants and washes. 5'-biotin-TAATAATAATAATTTTT-3' (IDT) can be incubated with streptavidin coated magnetic sepharose microbeads (GE Healthcare Life Sciences). Oligonucleotide-charged beads can be prepared by removing an aliquot of bead slurry and transferring to filter cup. The beads can then be washed five times with 1x PBS (using 2x volume of bead slurry) vortex and spin down each rinse. ½ bead slurry volume of 1x PBS may then be added, and biotinylated oligonucleotide can be spiked in at 1/10th published bead binding capacity. The mixture can be incubated at 37C for 2 hours with vortexing every 30min. After 2 hours, a small amount of supernatant may be removed and the A260 measured for any unbound oligonucleotide. Once the A260 shows <10% remaining oligonucleotide, the beads can be washed 5x with MQ water. The washed beads can be brought up to a desired concentration in MQ water.

Homopolymer synthesis can be performed as described above using 2-10x molar equivalents of desired dNTP relative to bead bound oligonucleotide. The reaction mixture containing beads, dNTP, TdT and buffer can be incubated at 37°C for 15 minutes. The reactions may be stopped with 10µl EDTA and rinsed 3x with water. A new cycle of homopolymer synthesis

may be initiated by adding fresh TdT enzyme, dNTP, buffer and incubated at 37°C for 15 minutes. After stopping the reaction with EDTA and rinsing 3x with water; the cycle of homopolymer synthesis can be repeated as many times as desired. After the last EDTA quench and 3x rinsing with water, the support-bound alternating homopolymers can be cleaved from the solid support using 100µl conc ammonium, and the supernatant dried down by Gen-vac, then stored at -20°C until ready to be analyzed using a polyacrylamide gel as described above.

In another experiment, A two-character message, "MA", was converted into binary and then into a base-two nucleotide code as shown in FIG 12. Each letter character was converted into a corresponding number from 1 to 26 (A→1, …Z→26). Each number was then converted into a six-bit binary number by appending two zeros to the normal binary representation of 1 to 26 (e.g., "M" = 001101; "A" = 000001). Each bit of the binary representation was converted into a base-two nucleotide representation according to the following table:

	Odd bit position	Even bit position
"0"	A	C
"1"	T	G

Thus "MA" was translated to 001101 000001 and then to the single nucleotide string, ACTGAGACACAG, which was synthesized as $A_nC_nT_nG_nA_nG_nA_nC_nA_nC_nA_nG_n$, where each nucleotide is synthesized as a variable length homopolymer.

Synthesis of the 12-bit homopolymer encoded nucleic acid was conducted with a 5'-biotinylated 39nt long oligonucleotide initiator attached to 34 um streptavidin-sepharose beads (GE Healthcare) at ~20 pmol/ul beads: 5'Biotin-CAGGTCCTAUCGATATCTGTGAGCTTAATGTCCTTATGT-3'.

The oligonucleotide contains two features for releasing the final product from the solid support used during synthesis: 1) a single deoxyuridine residue that allows cleavage with the USER enzyme system (New England Biolabs) and 2) an Eco R V endonuclease restriction site.

Starting with ~2nmol of bead-bound initiator, each variable length homopolymer was enzymatically synthesized using TdT and one of four modified nucleotide triphosphates. Each reaction was conducted in a total volume of 750 ul containing 40-100 uM modified dNTP (40 uM – A; 100 uM-C; 50 uM-T; 100 uM-G), 20 U TdT (Thermo-Fisher Scientific), 1X TdT Buffer (Thermo-Fisher Scientific) with incubation for 2.5-20 min at 37°C. After each enzymatic extension step, the reaction was quenched by adding 500 ul of 250 mM EDTA in 10 mM Tris buffer (pH 6.8). The beads were recovered by centrifugation at 10000xg and removal of the supernatant.

FIG. 13 shows the PAGE analysis of each cycle of enzymatic synthesis of a 12-bit homopolymer data strand. Each lane is marked with the cycle number, starting with "N", which shows the unreacted 39nt initiator. The black arrow points to a 60nt oligonucleotide size marker.

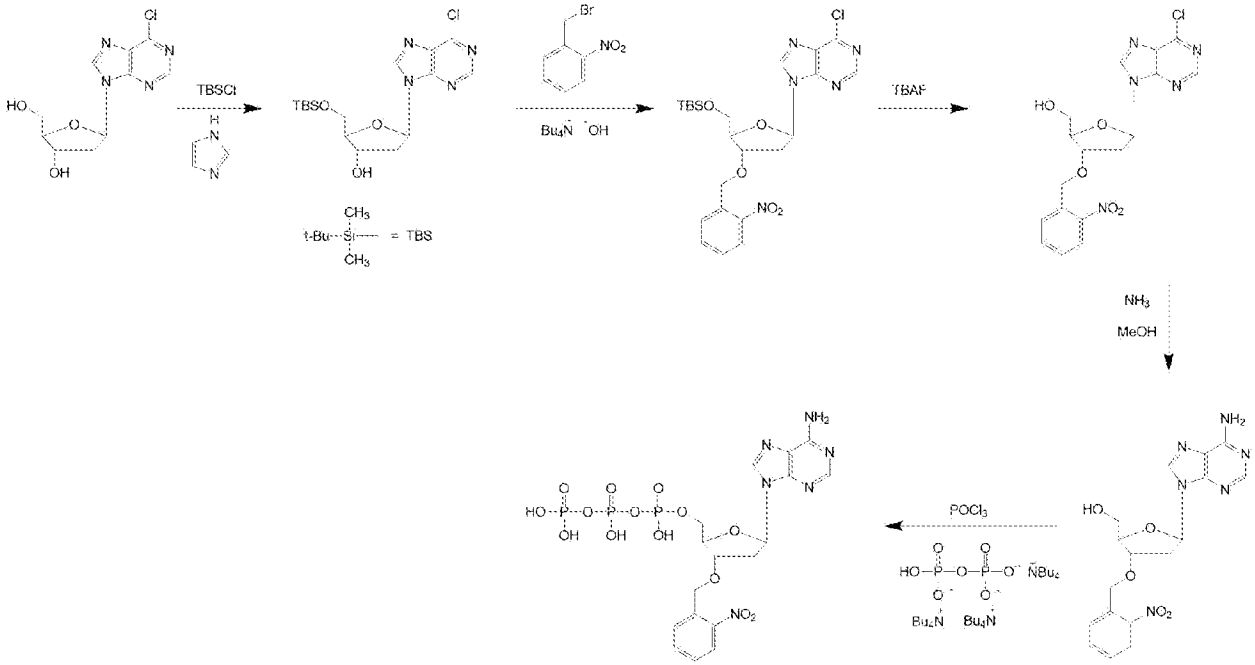
After removal of the full-length data strand from the solid support, NGS library preparation was performed using the ACCEL-NGS® 1S PLUS DNA LIBRARY KIT (Swift Bioscience) following the manufacturer's instructions and subsequently sequenced using an Illumina MiSeq System. FIGS. 14A-L shows histograms of the observed base composition of each of the twelve nucleotide additions (as labelled) generated during the enzymatic synthesis.

Example 2

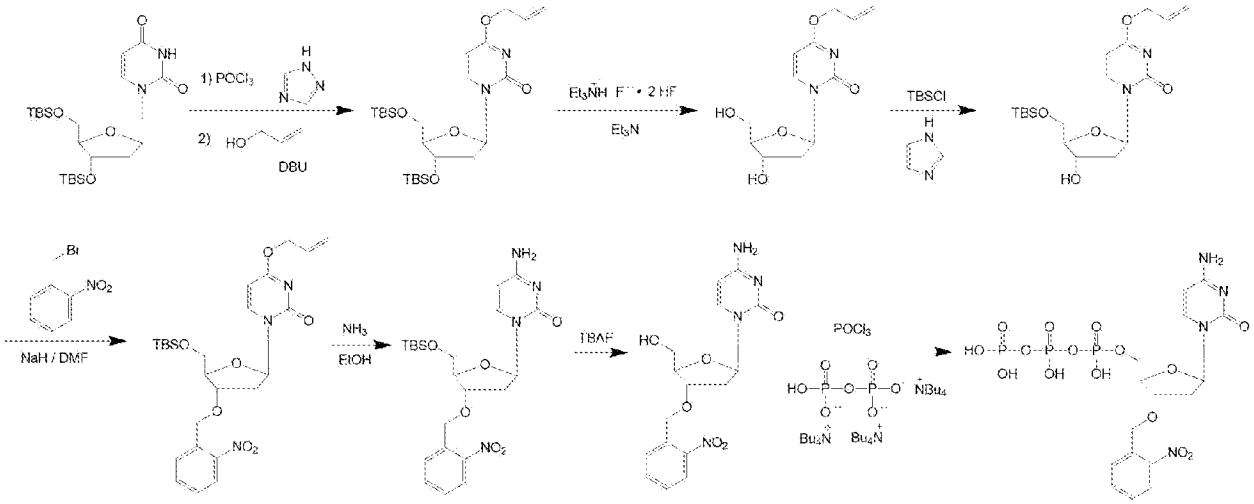
Procedures for synthesizing Class I – Purine & Pyrimidine dNTP analogs

Schemes for synthesis of Class I dNTP analogs

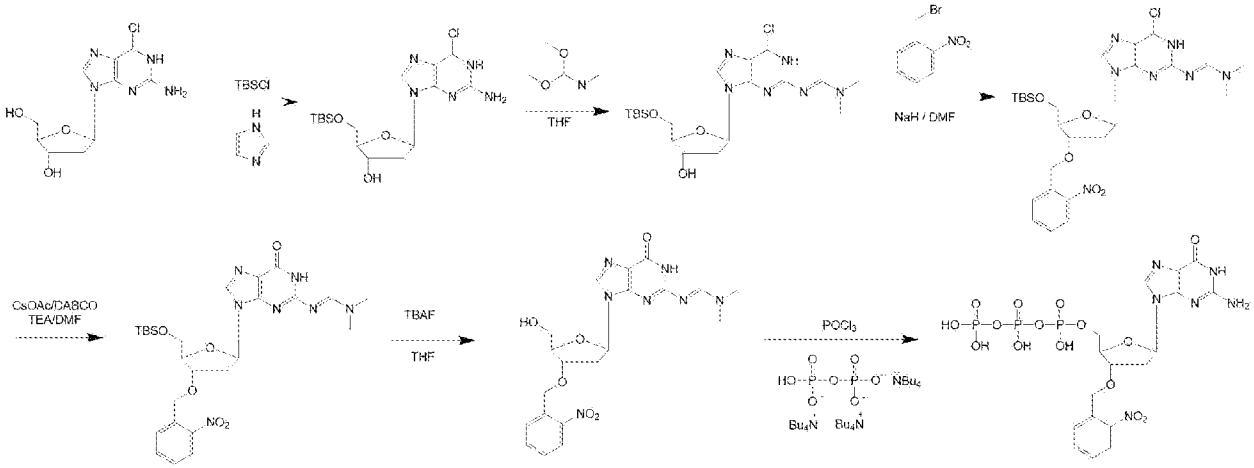
Nitrobenzyl - Adenosine



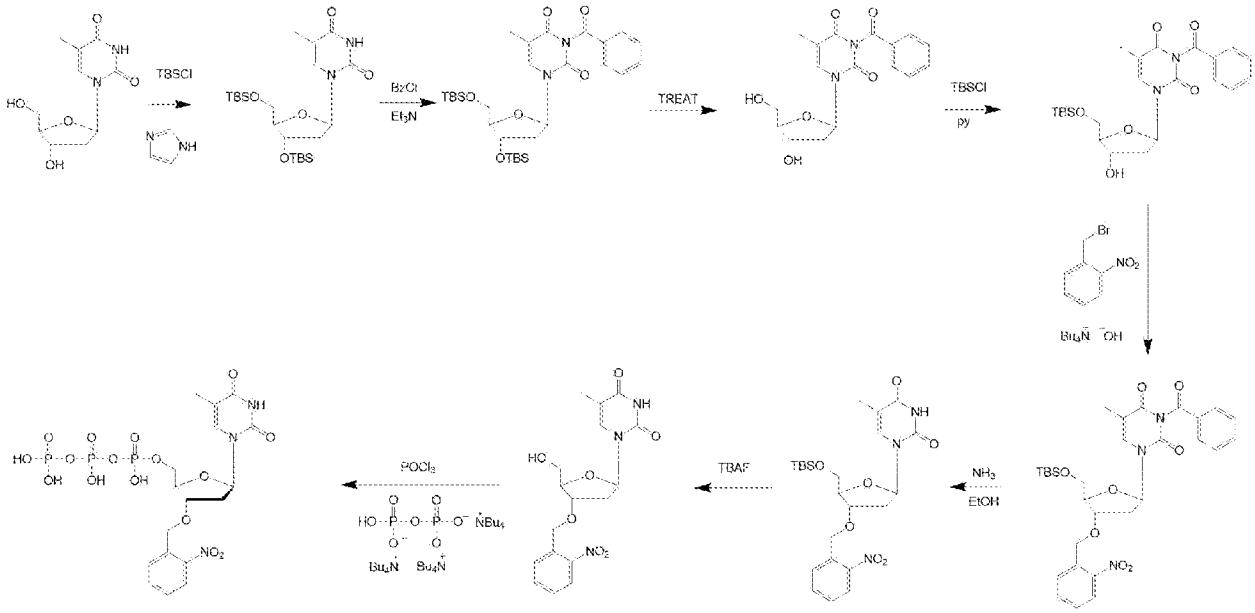
Nitrobenzyl - Cytosine



Nitrobenzyl - Guanosine

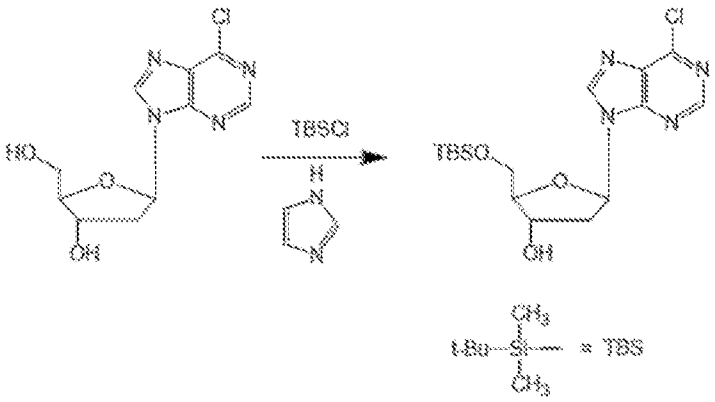


Nitrobenzyl - Thymidine



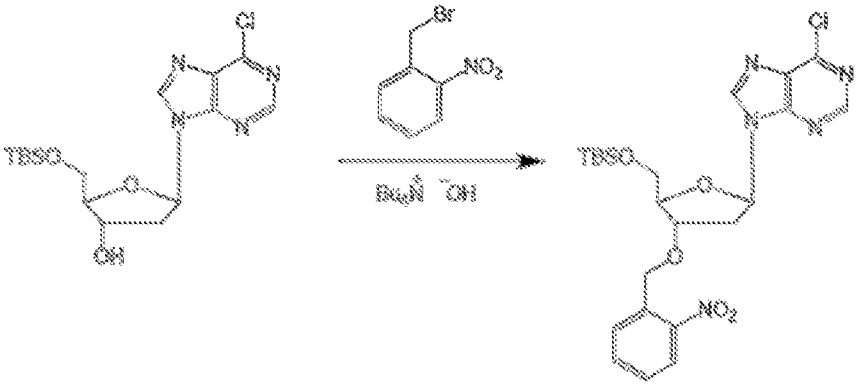
Example 3

Detailed procedures for 3'-O-Nitrobenzyl Deoxyadenosine:

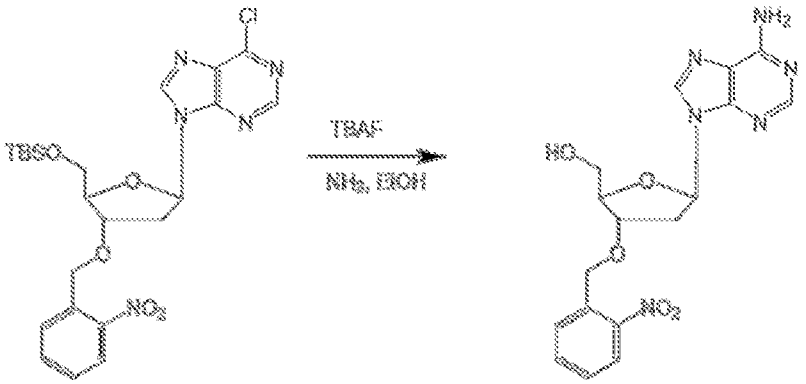


9-[β-D-5'-hydroxy-2'-deoxyribofuranosyl]-6-chloropurine (1.00 g, 3.69mmol) and imidazole (554 mg, 8.12 mmol) were dissolved in anhydrous dimethylformate (18 mL), followed by addition of tert-butyldimethylsilyl chloride (611 mg, 3.93 mmol). The reaction mixture was

stirred at room temperature for 20 hours under argon. The dried residue was impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 2:1) to obtain *9-[β-D-5'*-O-*(*tert-*butyldimethylsilyl)-2'-deoxyribofuranosyl]-6-chloropurine.*

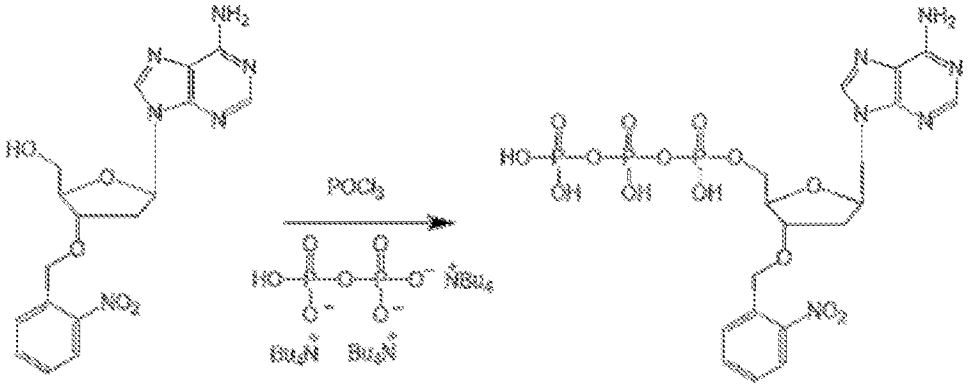


9-[β-D-5'-O-*(*tert-*butyldimethylsilyl)-2'-deoxyribofuranosyl]-6-chloropurine* (1.73 g, 4.48 mmol) was dissolved in anhydrous dichloromethane (135 mL). Tetrabutylammonium bromide (722 mg, 2.24 mmol), 2-nitrobenzyl bromide (2.41 g, 11.2 mmol) and 40% aqueous sodium hydroxide (65 mL) were added to the previously made solution. The reaction mixture was stirred at room temperature for 1h and diluted with ethyl acetate (300 mL). The layers were separate. The aqueous layer was extracted with ethyl acetate (125 mL x 2). The combined organic layers were dried over anhydrous sodium sulfate. The organic layer was impregnated on silica gel, followed by purification with flash column chromatography (hexanes/ethyl acetate, 2:1) to yield *9-[β-D-5'*-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyribofuranosyl]-6-chloropurine.*



9-[β-D-5'-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyribofuranosyl]-6-chloropurine.* (2.22 g, 3.83 mmol) was dissolved in anhydrous tetrahydrofuran (40 mL) and cooled down to 0 °C. Followed by addition of 1.0 M tetrabutylammonium fluoride in tetrahydrofuran solution (4.20 mL, 4.20 mmol) dropwise. The reaction mixture was stirred at room temperature for 1 h. After reaction mixture was dried, the residue was dissolved in dioxane and 7 N ammonia in ethanol (40 mL). The reaction mixture was stirred in the sealed round bottom at 90 °C for 18 h.

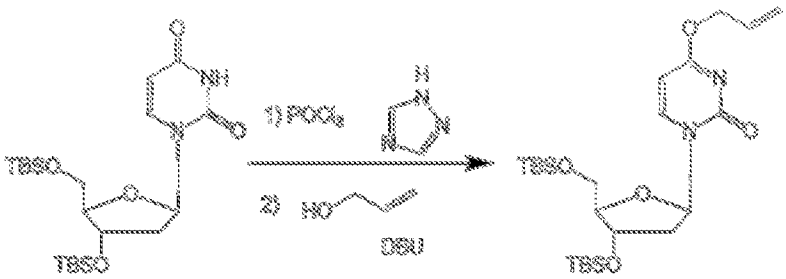
The reaction mixture was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 20:1) to obtain *3'-O-(2-nitrobenzyl)-2'-deoxyadenosine*.



3'-O-(2-nitrobenzyl)-2'-deoxyadenosine (15 mg, 1 Eq, 38 μmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 μL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

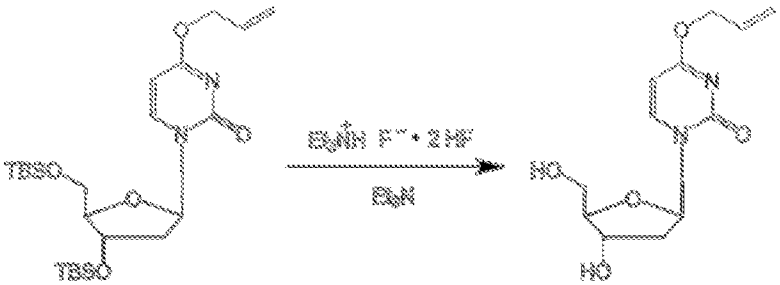
Example 4

Detailed procedures for 3'-O-Nitrobenzyl Deoxycytidine:

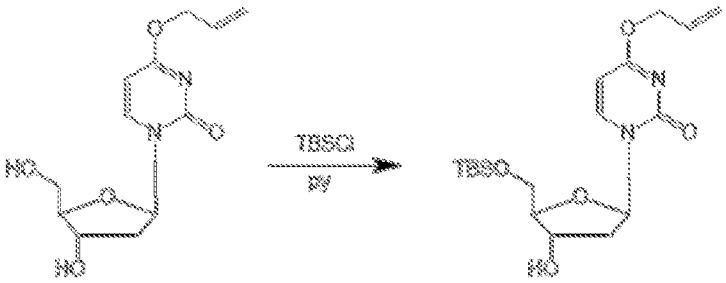


*3',5'-di-*O-(tert-*butyldimethylsilyl)-2'-deoxyuridine* (1.00 g, 2.12 mmol) was dissolved in anhydrous acetonitrile (90 mL) and cooled down to 0 °C under argon. Phosphoryl trichloride (1.49

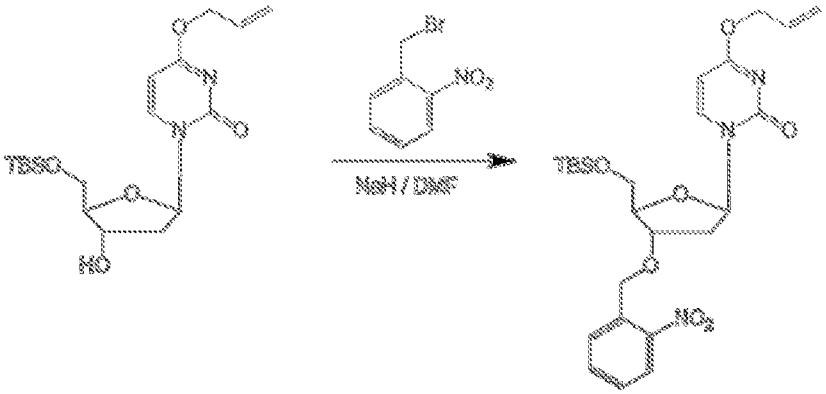
mL, 2.12 mmol) was added dropwise over 2 minutes. After 10 minutes, triethylamine (11.1 mL, 79.7 mmol) was added dropwise over 3 minutes. After 15 minutes, the reaction mixture was stirred at room temperature over 2 h The reaction mixture was cooled to 0 °C and triazole (4.40 g, 63.7 mmol). was added as a solid in one portion. The precipitate was observed, and the suspension was stirred for 30 minutes. After stirring at room temperature for 2 h, the reaction mixture was concentrated to dryness. The residue was dissolved in dichloromethane (30 mL) and washed with saturated solution of sodium bicarbonate (25 mL x 2), brine (25 mL). The organic layer was dried over anhydrous sodium sulfate and concentrated to dryness. The residue was dissolved in dichloromethane, followed by addition allyl alcohol (2.00 mL, 29.4 mmol) and triethylamine (2.67 mL, 18.9 mmol). The reaction mixture was stirred at 0 °C for 15 minutes. DBU (0.33 mL, 2.17 mmol) was added stirred at room temperature for 6 h. The reaction mixture was diluted with dichloromethane (17 mL) and washed with brine (15 mL). The organic layer was dried over anhydrous sodium sulfate. The organic layer was impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 4:1) to yield *4*-O-*allyl-3',5'-di*-O-(tert-*butyldimethylsilyl)-2'-deoxyuridine*.



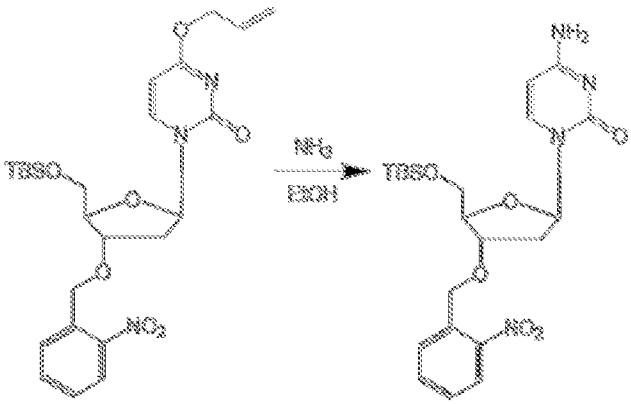
4-O-*allyl-3',5'-di*-O-(tert-*butyldimethylsilyl)-2'-deoxyuridine* (3.37 g, 5.27 mmol) was dissolved in dry tetrahydrofuran (50 mL). Triethylamine (1.98 mL, 14.2 mmol) was added followed by the triethylammonium fluoride dihydrofluoride (2.32 mL, 14.2 mmol) under argon. The reaction mixture was stirred at room temperature for 29 h, followed by concentration. The residue was dissolved in dichloromethane (100 mL) and washed with 1.5 M ammonium carbonate (75 mL x 1), brine (75 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *4*-O-*allyl-2'-deoxyuridine*.



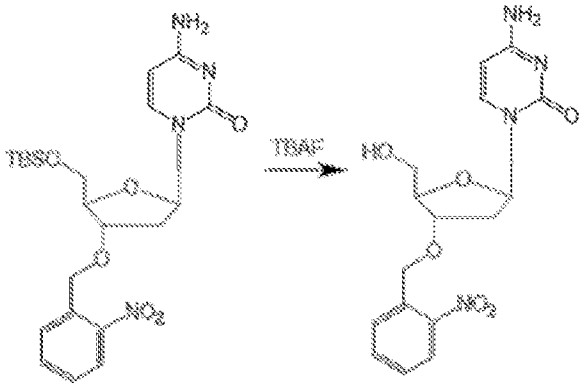
4-O-*allyl-2'-deoxyuridine* (1.18 g, 4.40mmol) was dissolved in anhydrous pyridine (37 mL), followed by addition of tert-butyldimethylsilyl chloride (815 mg, 5.41 mmol) under argon. The reaction mixture was stirred at room temperature for 20 h. After concentration, the residue was dissolved in dichloromethane and impregnated on silica. The crude product was purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *4*-O-*allyl-5'*-O-*(*tert-*butyldimethylsilyl)-2'-deoxyuridine*.



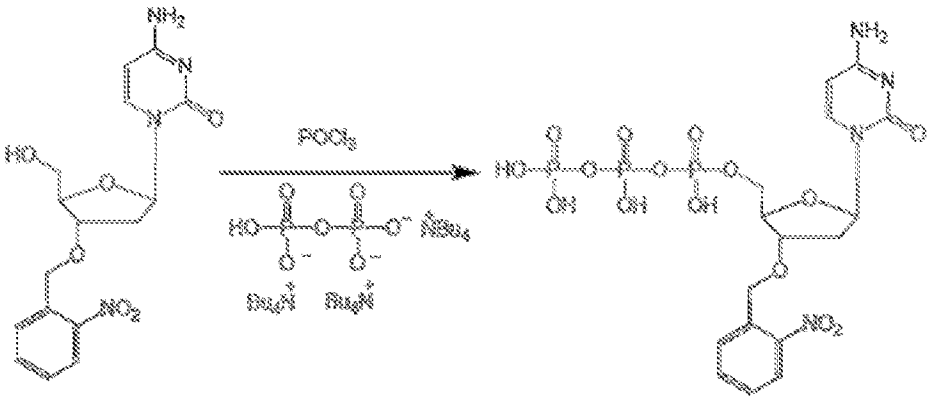
To a mixture of *4*-O-*allyl-5'*-O-*(*tert-*butyldimethylsilyl)-2'-deoxyuridine* (1.28 g, 3.35 mmol), tetrabutylammonium hydroxide (1.5 mL, 55-60% in water) and sodium iodide (50.0 mg, 0.335 mmol) in dichloromethane/water (20mL, 1:1) was added 1.0 M sodium hydroxide solution (10 mL) under argon. The reaction mixture was stirred for 10 minutes at room temperature, followed by addition of 2-nitrobenzyl bromide (1.45 g, 6.70 mmol) in 10 mL dichloromethane over 5 minutes. After stirring at room temperature for 7 h, the reaction mixture was diluted with dichloromethane (150 mL). The organic layer was washed with brine (20 mL) and dried over anhydrous sodium sulfate. The organic layer was impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 1:1) to yield *4*-O-*allyl-5'*-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyuridine*.



4-O-*allyl-5'*-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyuridine* (1.55, 3.00 mmol) was dissolved in 7 N ammonia in ethanol (55 mL) and stirred in the sealed round bottom at 55 °C for 20 h. The reaction mixture was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 20:1) to give *5'*-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine*.



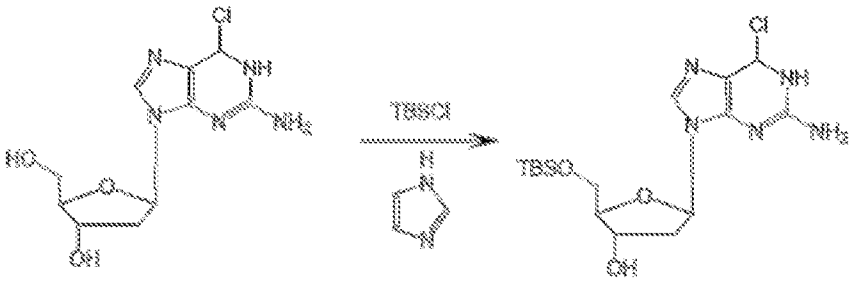
5'-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine* (2.22 g, 3.83 mmol) was dissolved in anhydrous tetrahydrofuran (40 mL) and cooled down to 0 °C. Followed by addition of 1.0 M tetrabutylammonium fluoride in tetrahydrofuran solution (4.20 mL, 4.20 mmol) dropwise. The reaction mixture was stirred at room temperature for 1 h. After reaction mixture was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 8:2) to afford *3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine*.



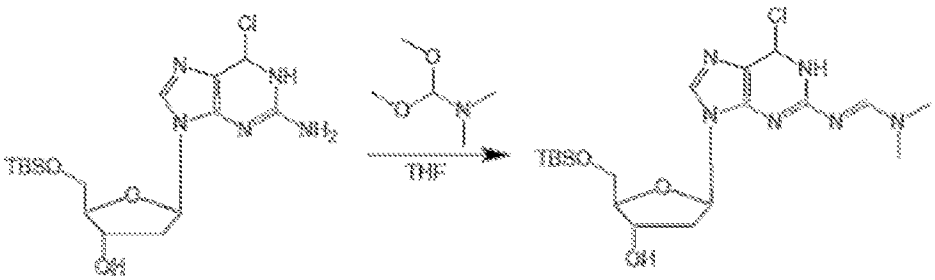
3'-O-(2-nitrobenzyl)-2'-deoxycytidine (14 mg, 1 Eq, 38 µmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 µL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

Example 5

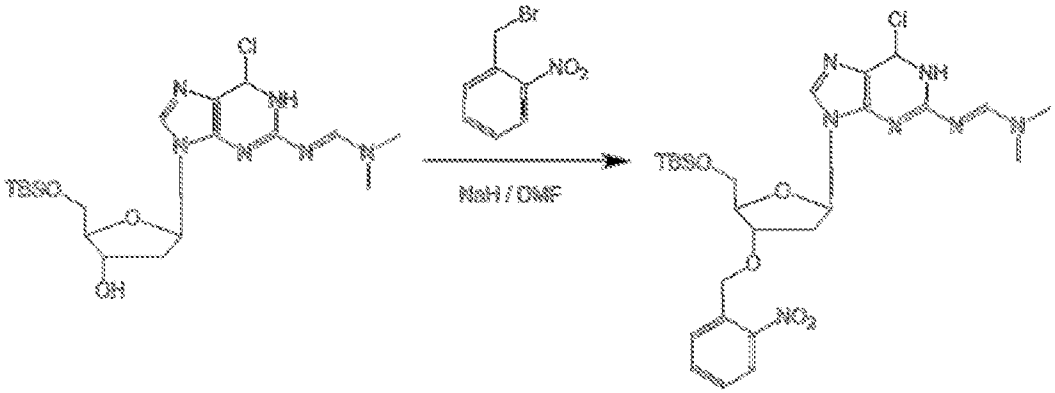
Detailed procedures for 3'-O-Nitrobenzyl Deoxyguanosine:



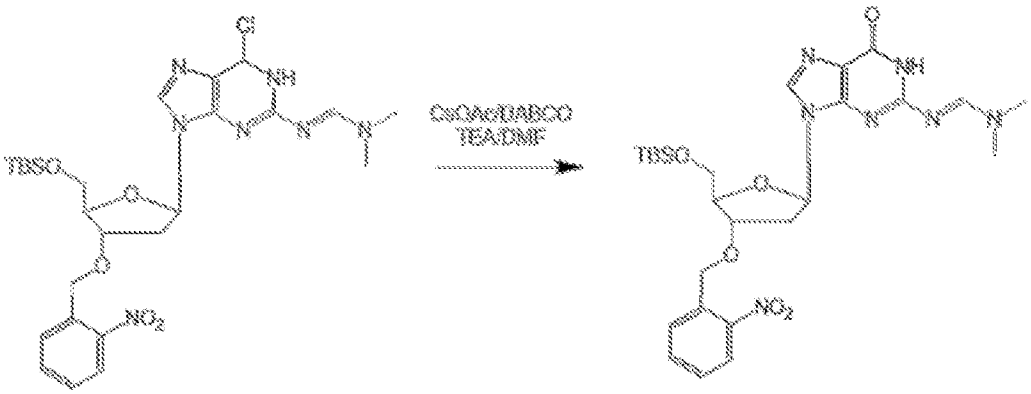
2-Amino-6-chloro-9-[β-D-2'-deoxyribofuranosyl]purine (1.00 g, 3.50mmol) and imidazole (715 mg, 10.50 mmol) were dissolved in anhydrous dimethylformate (18 mL), followed by addition of tert-butyldimethylsilyl chloride (686 mg, 4.60 mmol). The reaction mixture was stirred at room temperature for 12 hours under argon. The dried residue was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to afford *2-amino-6-chloro-9-[β-D-5'*-O-(tert-*butyldimethylsilyl)-2'-deoxyribofuranosyl]purine*.



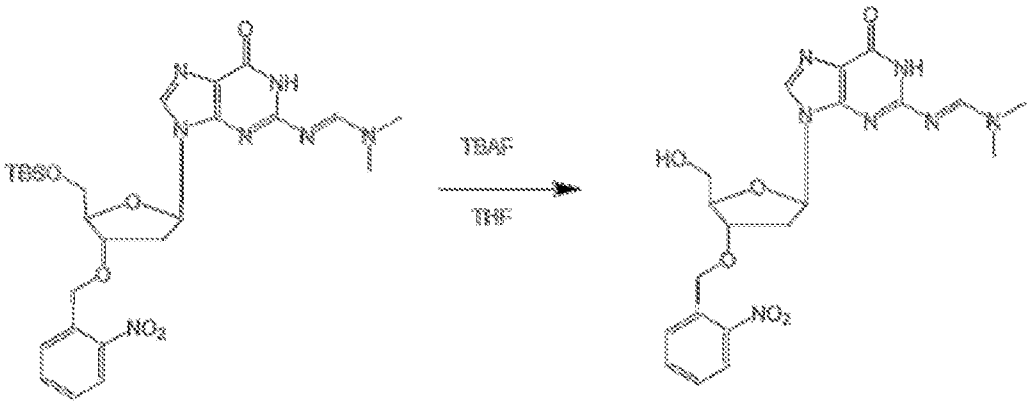
2-Amino-6-chloro-9-[β-D-5'-O-*(*tert-*butyldimethylsilyl)-2'-deoxyribofuranosyl]purine* (1.19 g, 3.00 mmol) was dissolved in anhydrous tetrahydrofuran (8.0 mL), followed by addition of *N,N*-dimethylformamide dimethyl acetal (3.10 mL, 18.0 mmol) at room temperature The reaction mixture was stirred at 40 °C for 3 h. The reaction mixture was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *6-chloro-N^2-[(dimethylaminomethylene)amino]-9-[β-D-5'*-O-*(*tert-*butyldimethylsilyl)-2'-deoxyribofuranosyl]purine.*



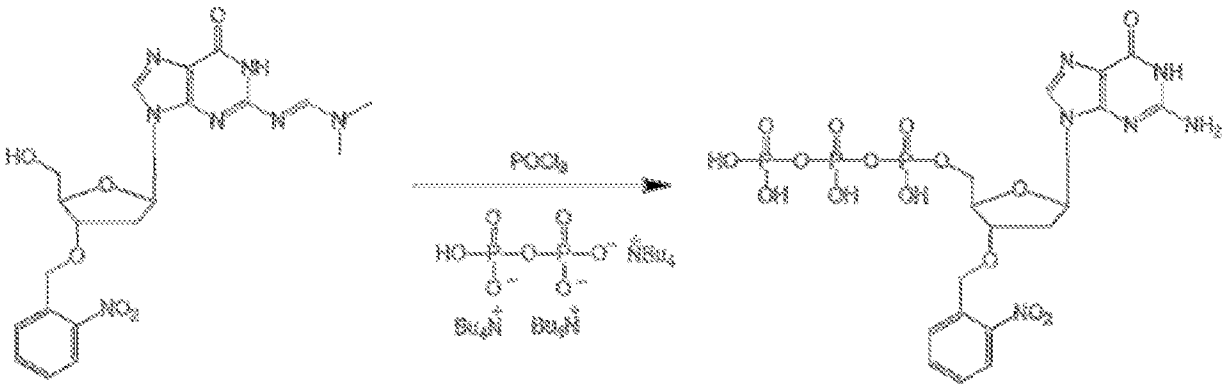
6-Chloro-N^2-[(dimethylaminomethylene)amino]-9-[β-D-5'-O-*(*tert-*butyldimethylsilyl)-2'-deoxyribofuranosyl]purine* (1.09 g, 2.40 mmol) was dissolved in anhydrous acetonitrile (3.5 mL), followed by addition of sodium hydride powder in mineral oil (60%) (122 mg, 4.80 mmol) at 0 °C. After stirring at room temperature for 1 h, solution of 2-nitrobenzyl bromide (1.04 g, 4.80 mmol) in anhydrous acetonitrile (1.5 mL) was added. After stirring at room temperature for 2 h, the reaction mixture was filtrated. The filtrate was dried and obtained residue was dissolved in ethyl acetate (100 mL). The organic layer was washed with saturated solution sodium bicarbonate (50 mL), brine (50 mL) and dried over anhydrous sodium sulfate. The organic layer was impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate 4:6) to give *6-chloro-N^2-[(dimethylaminomethylene)amino]-9-[β-D-5'*-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyribofuranosyl]purine.*



6-Chloro-N^2*-[(dimethylaminomethylene)amino]-9-[*β*-D-5'*-O-(tert-*butyldimethylsilyl)-3'*-O-(2-nitrobenzyl)-2'-deoxyribofuranosyl]purine* (1.02 g, 1.73 mmol) was dissolved in anhydrous dimethylformate (15 mL), followed by addition of cesium acetate (996 mg, 5.19 mmol), 1,4-diazabicyclo[2.2.2]octane (194 mg, 1.73 mmol) and triethylamine (0.72 mL, 5.19 mmol) under argon. The reaction mixture was stirred at room temperature for 18 h. Acetic anhydride (5 mL) was added and stirred for 0.5 h. The reaction mixture was quenched with water (100 mL) and extracted with ethyl acetate (100 mL). The organic layer was dried over anhydrous sodium sulfate and impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 20:1) to give *5'*-O-(tert-*butyldimethylsilyl)*-N^2*-[(dimethylamino)methylene]-3'*-O-*(2-nitrobenzyl)-2'-deoxyguanosine*.



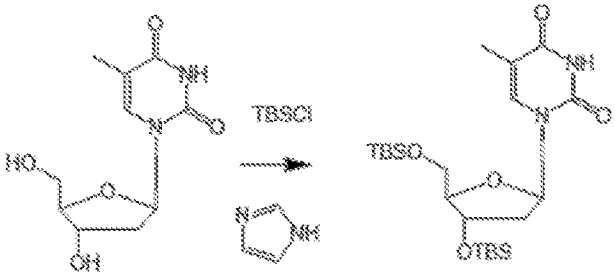
5'-O-(tert-*butyldimethylsilyl)*-N^2*-[(dimethylamino)methylene]-3'*-O-*(2-nitrobenzyl)-2'-deoxyguanosine* (732 mg, 1.28 mmol) was dissolved in anhydrous tetrahydrofuran (8 mL) under argon and cooled down to 0 °C. Followed by addition of 1.0 M tetrabutylammonium fluoride in tetrahydrofuran solution (2.56 mL, 2.56 mmol) dropwise. The reaction mixture was stirred at room temperature for 2 h. The reaction mixture was poured to cold water (50 mL) and extracted with ethyl acetate (50 mL x 2). Organic layers were combined and dried over anhydrous sodium sulfate. The organic layer was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 10:1) to yield N^2*-[(dimethylamino)methylene]-3'*-O-*(2-nitrobenzyl)-2'-deoxyguanosine*.



N^2-*[(dimethylamino)methylene]-3'*-O-*(2-nitrobenzyl)-2'-deoxyguanosine* (17 mg, 1 Eq, 38 µmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 µL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

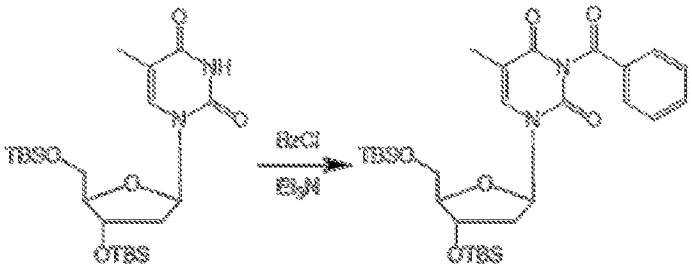
Example 6

Detailed procedures for 3'-O-Nitrobenzyl Deoxythymidine:

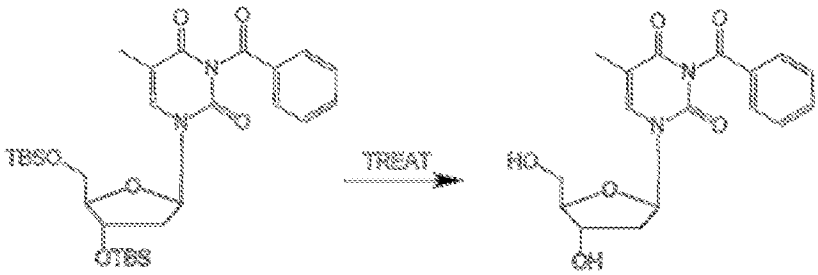


Thymidine (2.50 g, 10.3 mmol) was suspended in dimethylformamide (60 mL) at room temperature under argon. To the suspension imidazole (4.22 g, 61.9 mmol) and tert-butylchlorodimethylsilane (4.66 g, 30.1 mmol) were added. After stirring for 2 h, the reaction

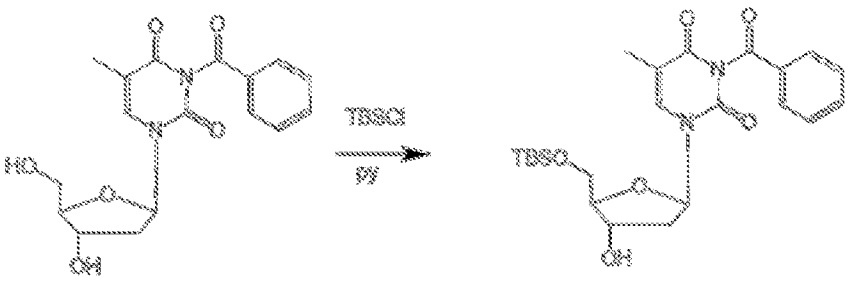
mixture was quenched with methanol (8 mL) and diluted with ethyl acetate (200 mL). The organic layer was washed with water (100 mL x 2), saturated solution of sodium bicarbonate (100 mL) and brine (100 mL). Organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 8:2) to yield *3',5'-di*-O-*(*tert-*butyldimethylsilyl)thymidine*.



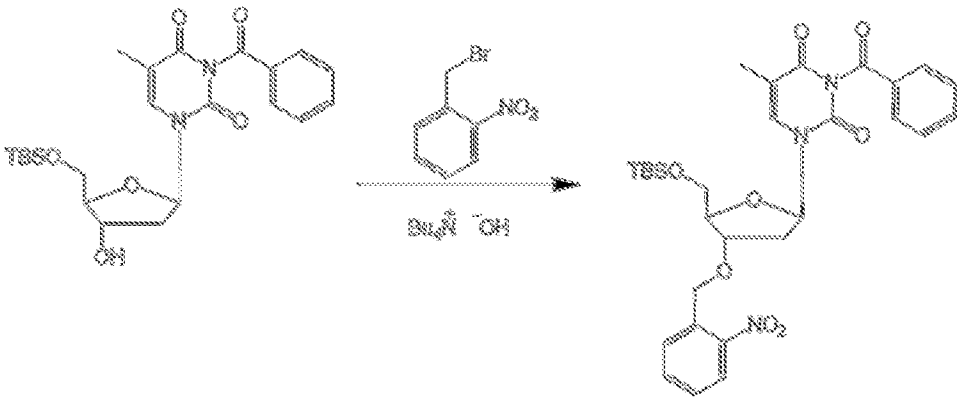
3',5'-di-O-*(*tert-*butyldimethylsilyl)thymidine* (4.34 g, 9.22 mmol) and dimethyl-4-aminopyridine (1.12 g, 9.22 mmol) were dissolved in anhydrous dichloromethane (140 mL). Triethylamine (5.14 mL, 36.9 mmol) was added and the reaction mixture was cooled to 0 °C. Benzyl chloride (3.21 mL, 27.7 mmol) was added dropwise and allowed to warp up to room temperature. After stirring for 14 h, saturated solution of sodium bicarbonate (80 mL) was added and layers were separate. The aqueous layer was extracted with dichloromethane (200 mL x 2). Combined organic layers were washed with water (300 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate (8:2) to afford *3*-N-*benzoyl-3'5'-di*-O-*(*tert-*butyldimethylsilyl)thymidine*.



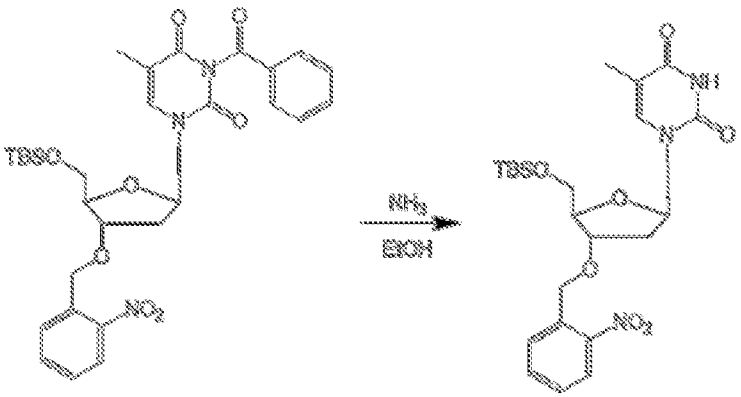
3-N-*benzoyl-3'5'-di*-O-*(*tert-*butyldimethylsilyl)thymidine* (3.03 g, 5.27 mmol) was dissolved in dry tetrahydrofuran (50 mL). Triethylamine (1.98 mL, 14.2 mmol) was added followed by the triethylammonium fluoride dihydrofluoride (2.32 mL, 14.2 mmol) under argon. The reaction mixture was stirred at room temperature for 29 h, followed by concentration. The residue was dissolved in dichloromethane (100 mL) and washed with 1.5 M ammonium carbonate (75 mL), brine (75 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to give *3*-N-*benzoylthymidine*.



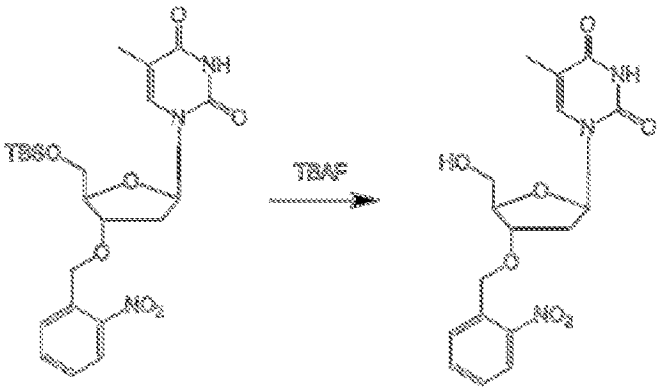
3-N-*benzoylthymidine* (XX g, 4.40mmol) was dissolved in anhydrous pyridine (37 mL), followed by addition of *tert*-butyldimethylsilylchloride (815 mg, 5.41 mmol) under argon. The reaction mixture was stirred at room temperature for 20 h. After concentration, the residue was dissolved in dichloromethane and impregnated on silica. The crude product was purified by flash column chromatography (dichloromethane/methanol, 9:1) *3*-N-*benzoyl-5'*-O-(tert-*butyldimethylsilyl)thymidine*.



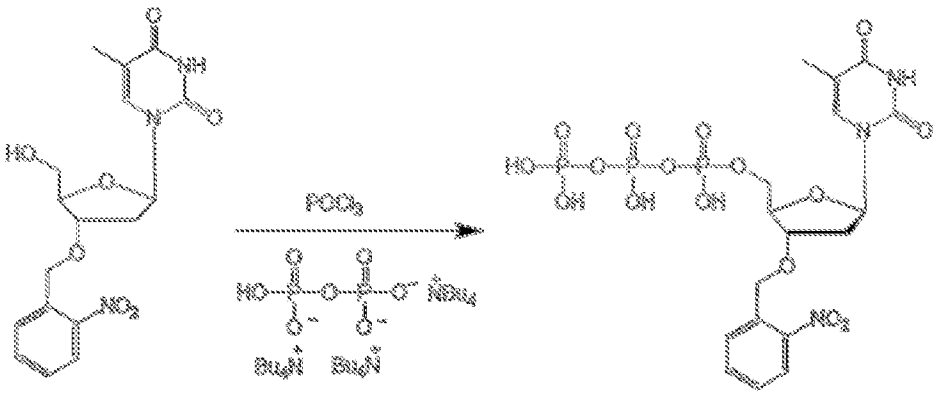
To *3*-N-*benzoyl-5'*-O-(tert-*butyldimethylsilyl)thymidine* (1.18 g, 2.56 mmol) aqueous solution of tetrabutylammonium hydroxide (10 mL, 60%) was added, followed by sodium iodide (76.7 mg, 0.51 mmol), dichloromethane (10 mL), water (10 mL) and aqueous solution of 1M sodium hydroxide (10 mL). This mixture was added dropwise to a solution of 2-nitrobenzyl bromide (718 mg, 3.32 mmol) in dichloromethane (10 mL). After stirring reaction mixture at room temperature for 6h, water (10mL) was added. The aqueous layer was extracted with dichloromethane (50 mL x 3). The organic layers were combined and washed with brine (100 mL), dried over anhydrous sodium sulfate. After filtration and concentration, the residue was dissolved with ethyl acetate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 6:4) to give *3*-N-*benzoyl-5'*-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)thymidine*.



3-N-*benzoyl-5'*-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)thymidine* (1.24 g, 2.08 mmol) was dissolved in ethanol (15 mL), followed by addition of 30% ammonium hydroxide solution (1.5 mL, 12.3 mmol). The reaction mixture was stirred at room temperature for 1 h, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 6:4) to yield *5'*-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)thymidine*.



5'-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)thymidine* (681 mg, 1.39 mmol) was dissolved in anhydrous tetrahydrofuran (12 mL) under argon and cooled down to 0 °C. Followed by addition of 1.0 M tetrabutylammonium fluoride in tetrahydrofuran solution (2.78 mL, 2.78 mmol) dropwise. The reaction mixture was stirred at room temperature for 2 h. The reaction mixture was poured to cold water (50 mL) and extracted with ethyl acetate (50 mL x 3). The organic layers were combined and dried over anhydrous sodium sulfate. The organic layer was impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 10:1) to give *3'*-O-*(2-nitrobenzyl)thymidine*.



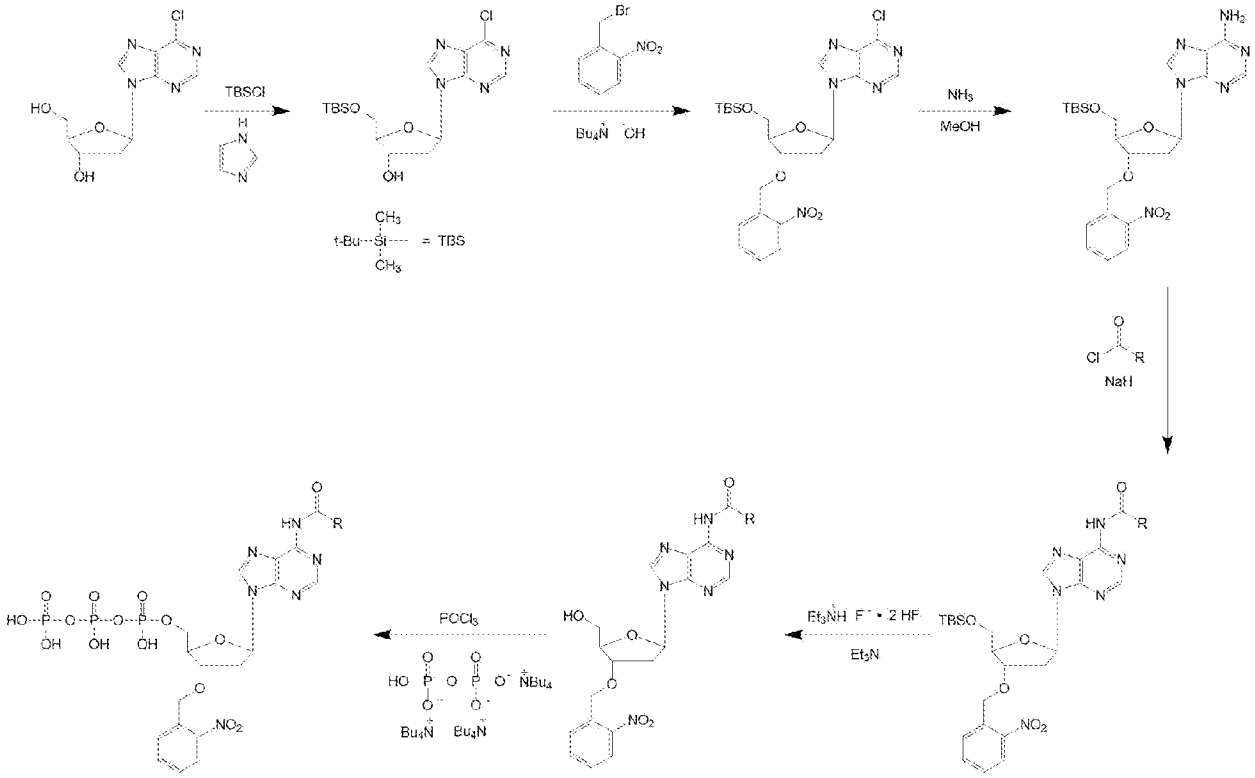
3'-O-*(2-nitrobenzyl)thymidine* (15 mg, 1 Eq, 38 μmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 μL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

Example 6

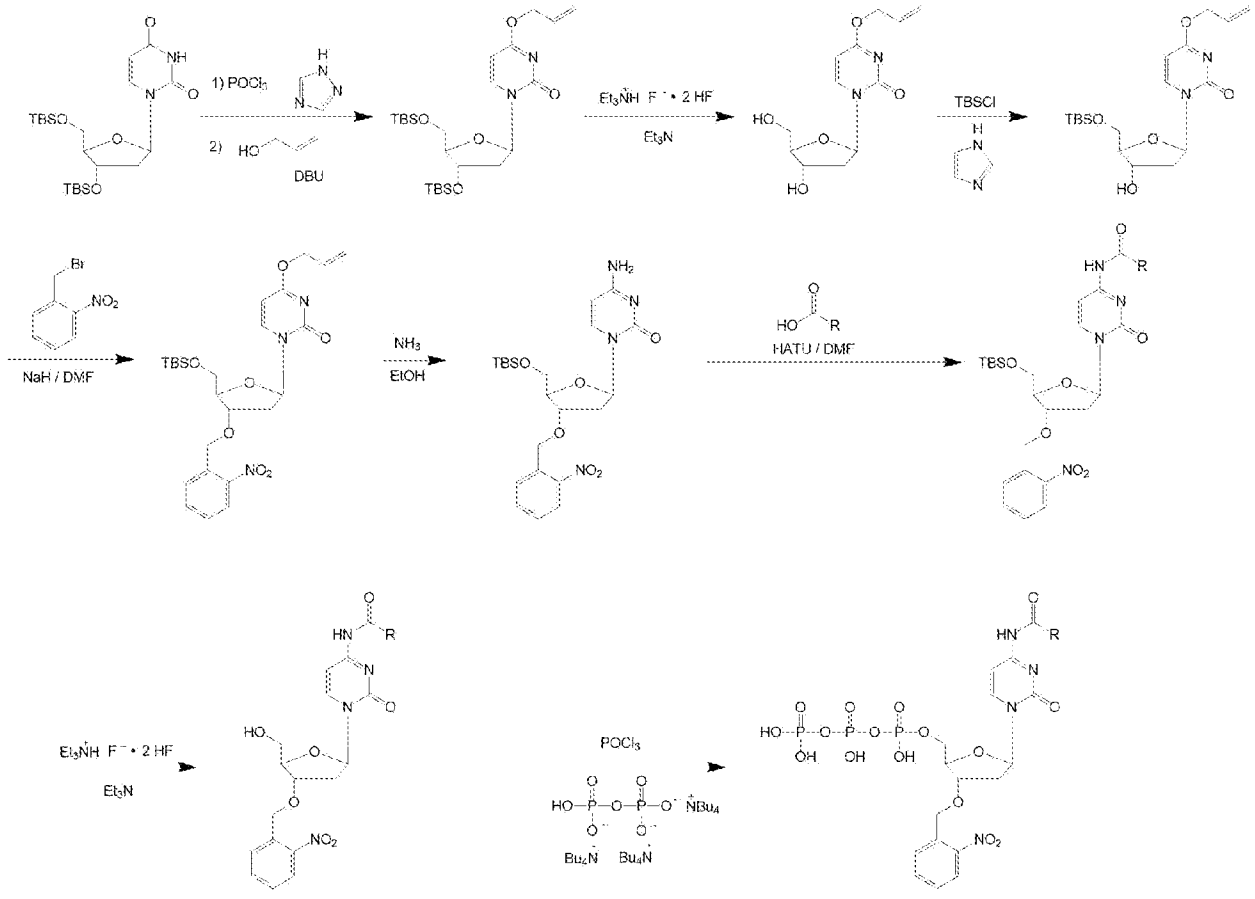
Procedures for synthesizing Class II – Purine & Pyrimidine dNTP analogs.

Schemes for synthesis of Class II non-peptide dNTP analogs

Non Peptidic Precursor - Adenosine

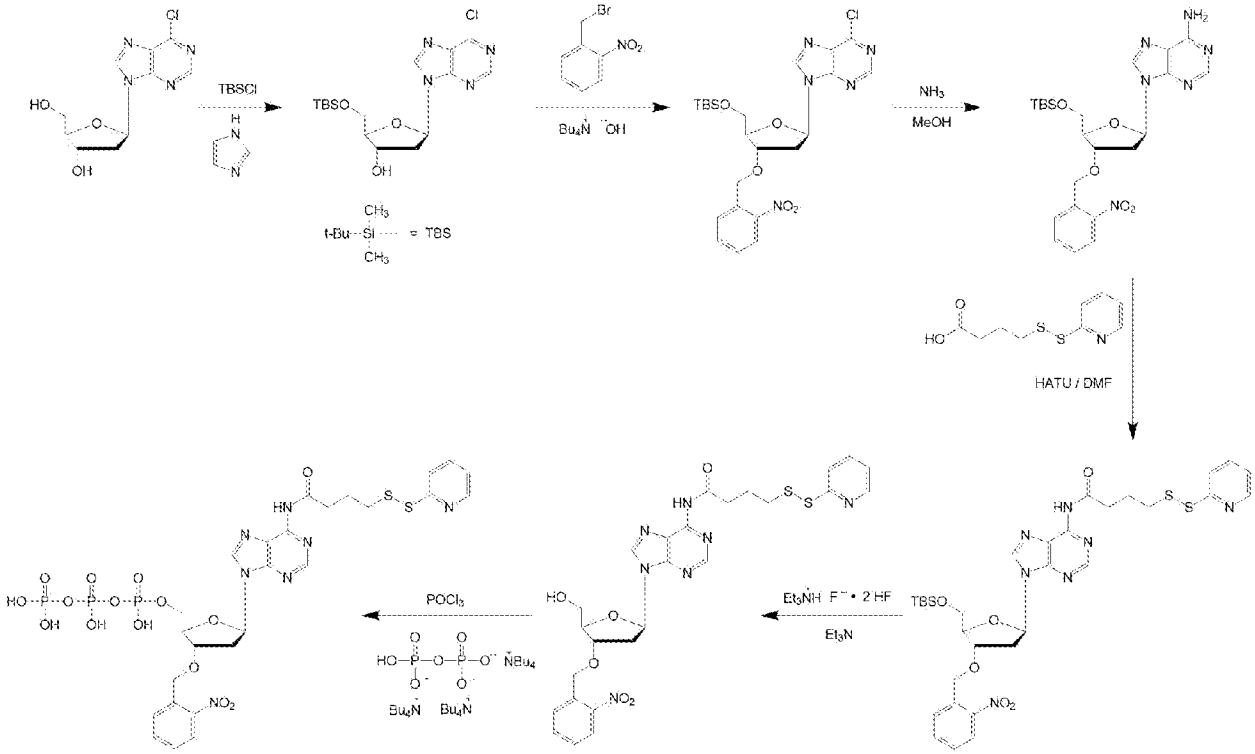


Non Peptidic Precursor - Cytosine

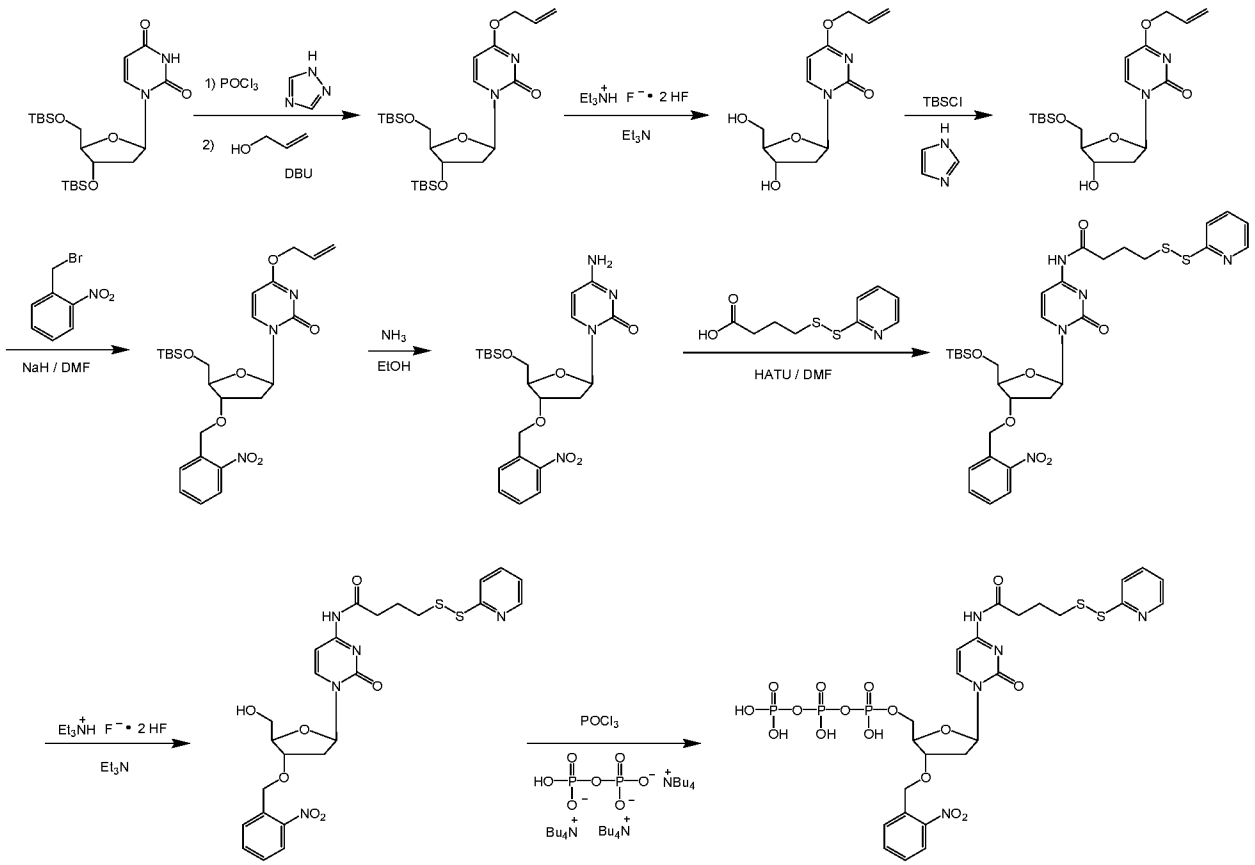


Schemes for synthesis of Class II peptide-dNTP analogs

Disulfide Peptide Precursor - Adenosine

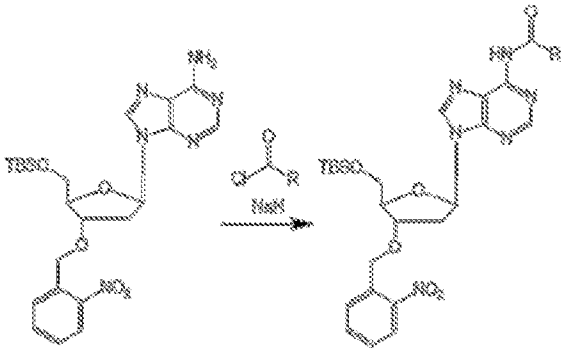


Disulfide Peptide Precursor - Cytosine

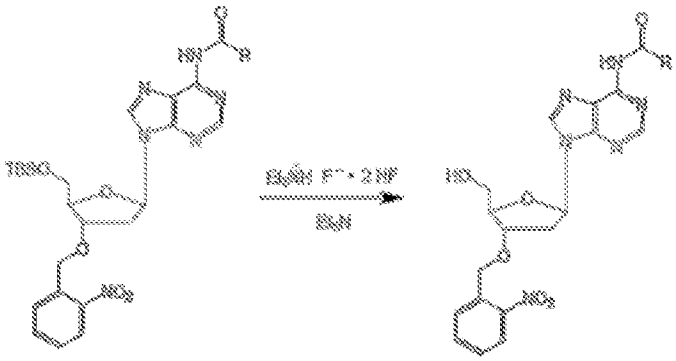


Example 7

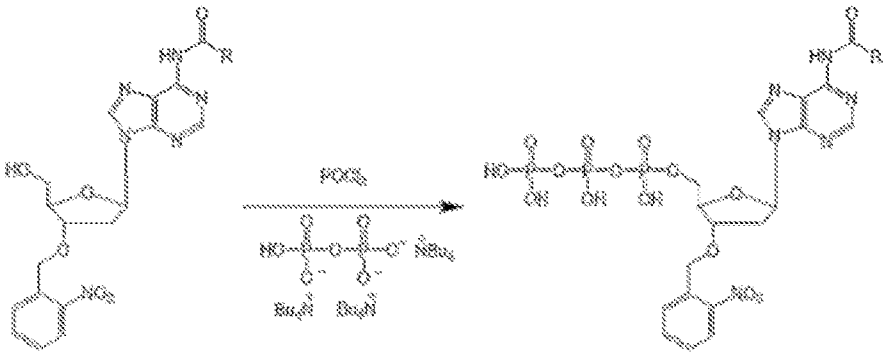
Detailed procedures for synthesis of Class II non-peptide dATP constructs:



9-[β-D-5'-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyadenosine* (0.24 mmol) was dissolved in anhydrous tetrahydrofurane (4 mL) at room temperature under argon. Sodium hydride (48 mg, 1.21 mmol) was added in one portion. After stirring for 1 h, reaction mixture was cooled to 0 °C and acyl chloride (3 eq, 0.723 mmol) was added drop wise. The reaction mixture was stirred for 18 h at room temperature. The reaction mixture was poured into cold solution of saturated sodium bicarbonate and dichloromethane. The layers were separated. Organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 2:8) to yield desired product.



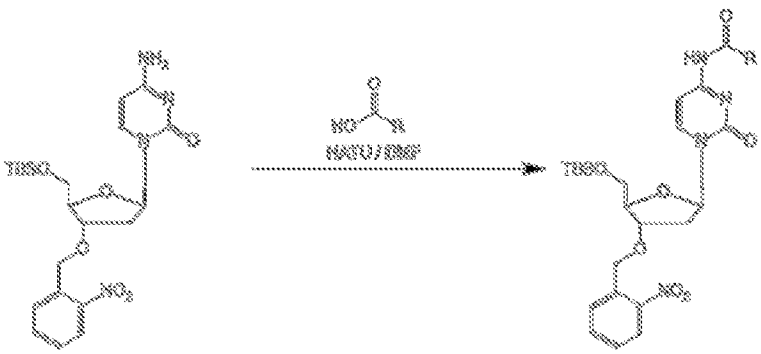
N-substituted nucleoside (5.27 mmol) was dissolved in dry tetrahydrofuran (50 mL). Triethylamine (1.98 mL, 14.2 mmol) was added followed by the triethylammonium fluoride dihydrofluoride (2.32 mL, 14.2 mmol) under argon. The reaction mixture was stirred at room temperature for 29 h, followed by concentration. The residue was dissolved in dichloromethane (100 mL) and washed with 1.5 M ammonium carbonate (75 mL x 1), brine (75 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *N-substituted 3'*-O-*(2-nitrobenzyl)-2'-deoxyadenosine.*



N-substituted 3'-O-*(2-nitrobenzyl)-2'-deoxyadenosine* (1 Eq, 38 μmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 μL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

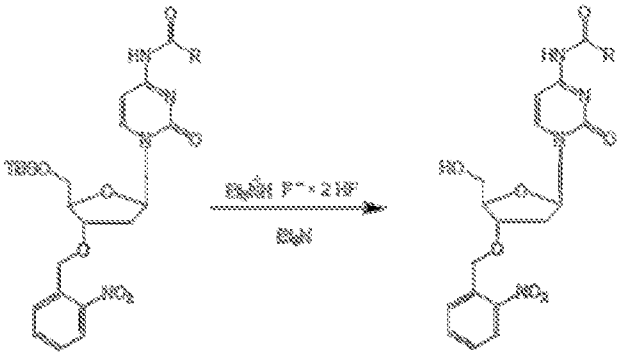
Example 8

Detailed procedure for synthesis of Class II non-peptide dCTP constructs:

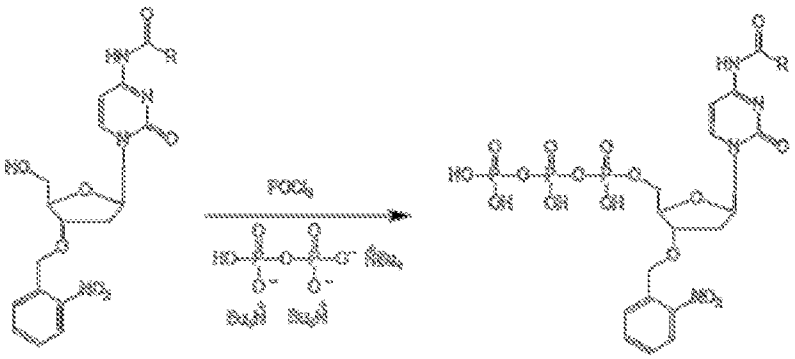


5'-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine* (524.3 mg, 1.10 mmol) and carboxylic acid (1.32 mmol) were dissolved in anhydrous dimethylformate under argon. *N,N*-Diisopropylethylamine (0.48 mL, 2.74 mmol) and 2-(3H-[1,2,3]triazolo[4,5-b]pyridin-3-yl)-1,1,3,3-tetramethylisouronium hexafluorophosphate (V) (417 mg, 1.10 mmol) were added. The

reaction mixture was stirred at room temperature for 18 h and diluted with ethyl acetate (30 mL). The organic layer was washed with saturated sodium bicarbonate solution (30 mL), dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 4:6) to yield *N-substituted 5'*-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine.*



N-substituted 5'-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine* (3.37 g, 5.27 mmol) was dissolved in dry tetrahydrofuran (50 mL). Triethylamine (1.98 mL, 14.2 mmol) was added followed by the triethylammonium fluoride dihydrofluoride (2.32 mL, 14.2 mmol) under argon. The reaction mixture was stirred at room temperature for 29 h, followed by concentration. The residue was dissolved in dichloromethane (100 mL) and washed with 1.5 M ammonium carbonate (75 mL x 1), brine (75 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *N-substituted 3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine.*

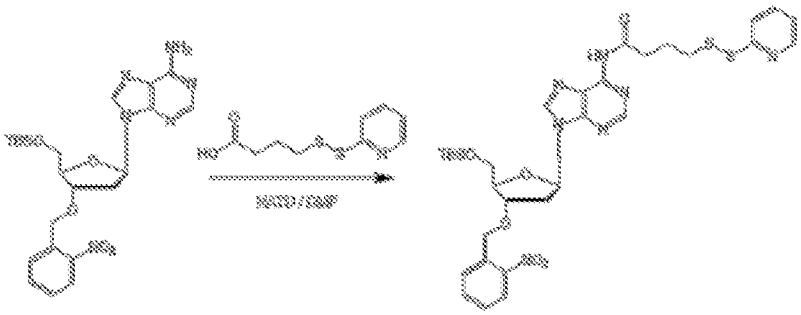


N-substituted 3'-O-*(2-nitrobenzyl)-2'-deoxycytidine* (1 Eq, 38 µmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 µL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was

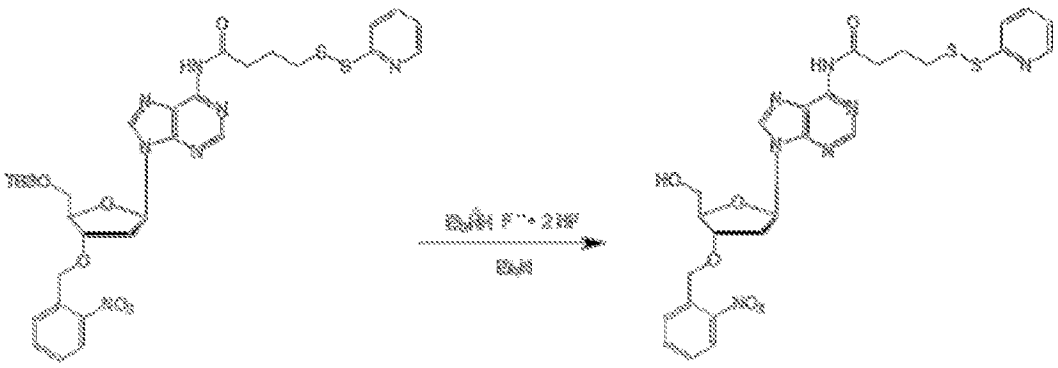
prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

Example 9

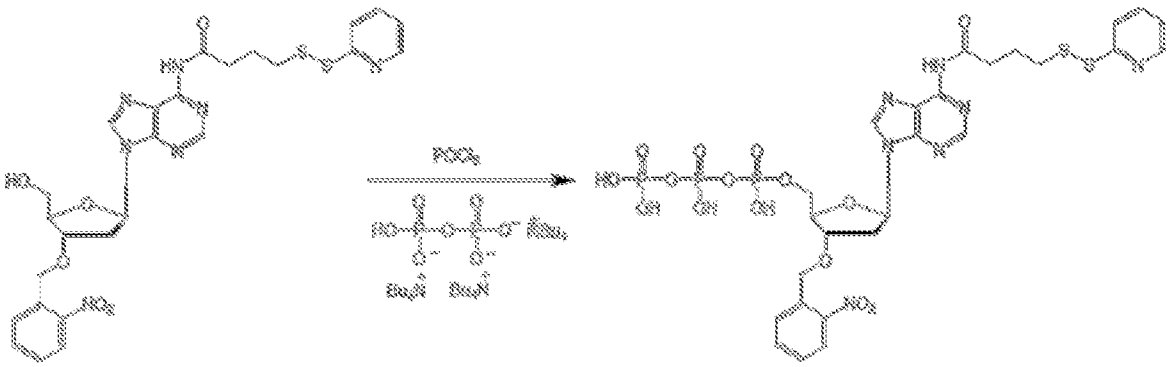
Detailed procedure for synthesis of peptide-dATP conjugates:



9-[β-D-5'-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyadenosine* (524.3 mg, 1.10 mmol) and 4-(pyridin-2-yldisulfaneyl)butanoic acid (302.7mg, 1.32 mmol) were dissolved in anhydrous dimethylformate under argon. *N,N*-Diisopropylethylamine (0.48 mL, 2.74 mmol) and 2-(3H-[1,2,3]triazolo[4,5-b]pyridin-3-yl)-1,1,3,3-tetramethylisouronium hexafluorophosphate (V) (417 mg, 1.10 mmol) were added. The reaction mixture was stirred at room temperature for 18 h and diluted with ethyl acetate (30 mL). The organic layer was washed with saturated sodium bicarbonate solution (30 mL), dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 4:6) to yield *N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-5'*-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyadenosine.*



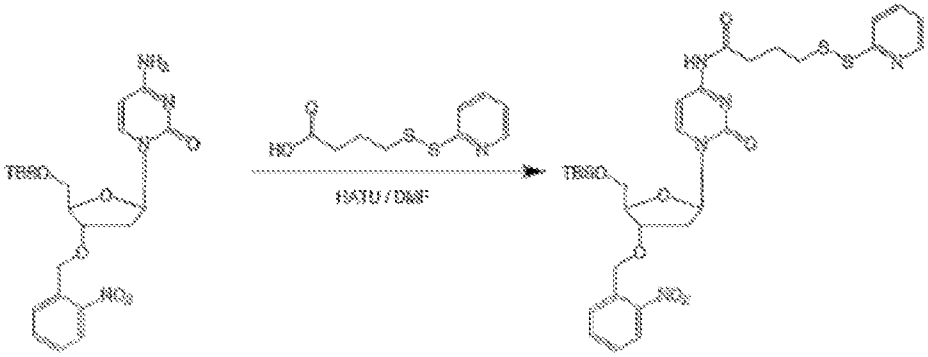
N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-5'-O-(tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyadenosine* (3.75 g, 5.27 mmol) was dissolved in dry tetrahydrofuran (50 mL). Triethylamine (1.98 mL, 14.2 mmol) was added followed by the triethylammonium fluoride dihydrofluoride (2.32 mL, 14.2 mmol) under argon. The reaction mixture was stirred at room temperature for 29 h, followed by concentration. The residue was dissolved in dichloromethane (100 mL) and washed with 1.5 M ammonium carbonate (75 mL x 1), brine (75 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-3'*-O-*(2-nitrobenzyl)-2'-deoxyadenosine*.



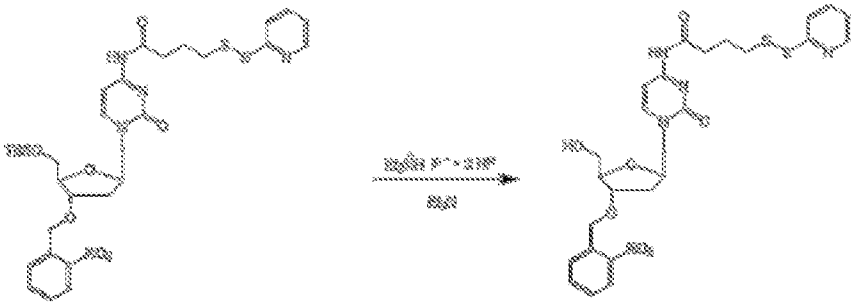
N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-3'-O-*(2-nitrobenzyl)-2'-deoxyadenosine* (23 mg, 1 Eq, 38 μmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 μL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 10 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

Example 10

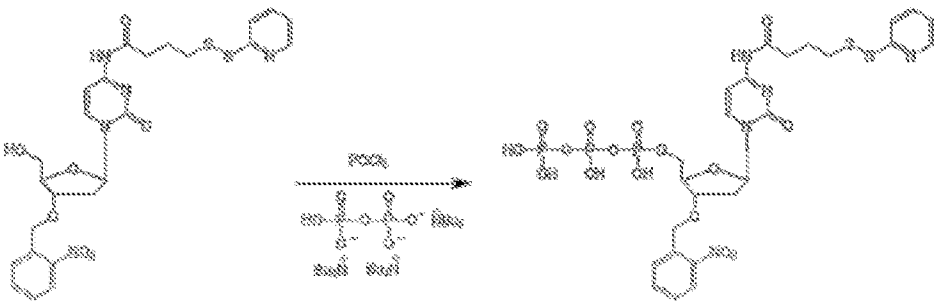
Detailed procedure for synthesis of peptide-dCTP conjugates:



5'-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine* (524.3 mg, 1.10 mmol) and 4-(pyridin-2-yldisulfaneyl)butanoic acid (1.32 mmol) were dissolved in anhydrous dimethylformate under argon. *N,N*-Diisopropylethylamine (0.48 mL, 2.74 mmol) and 2-(3H-[1,2,3]triazolo[4,5-b]pyridin-3-yl)-1,1,3,3-tetramethylisouronium hexafluorophosphate (V) (417 mg, 1.10 mmol) were added. The reaction mixture was stirred at room temperature for 18 h and diluted with ethyl acetate (30 mL). The organic layer was washed with saturated sodium bicarbonate solution (30 mL), dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (hexane/ethyl acetate, 4:6) to yield *N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-5'*-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine.*



N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-5'-O-*(*tert-*butyldimethylsilyl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine* (3.62 g, 5.27 mmol) was dissolved in dry tetrahydrofuran (50 mL). Triethylamine (1.98 mL, 14.2 mmol) was added followed by the triethylammonium fluoride dihydrofluoride (2.32 mL, 14.2 mmol) under argon. The reaction mixture was stirred at room temperature for 29 h, followed by concentration. The residue was dissolved in dichloromethane (100 mL) and washed with 1.5 M ammonium carbonate (75 mL x 1), brine (75 mL). The organic layer was dried over anhydrous sodium sulfate, impregnated on silica and purified by flash column chromatography (dichloromethane/methanol, 9:1) to obtain *N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-3'*-O-*(2-nitrobenzyl)-2'-deoxycytidine.*

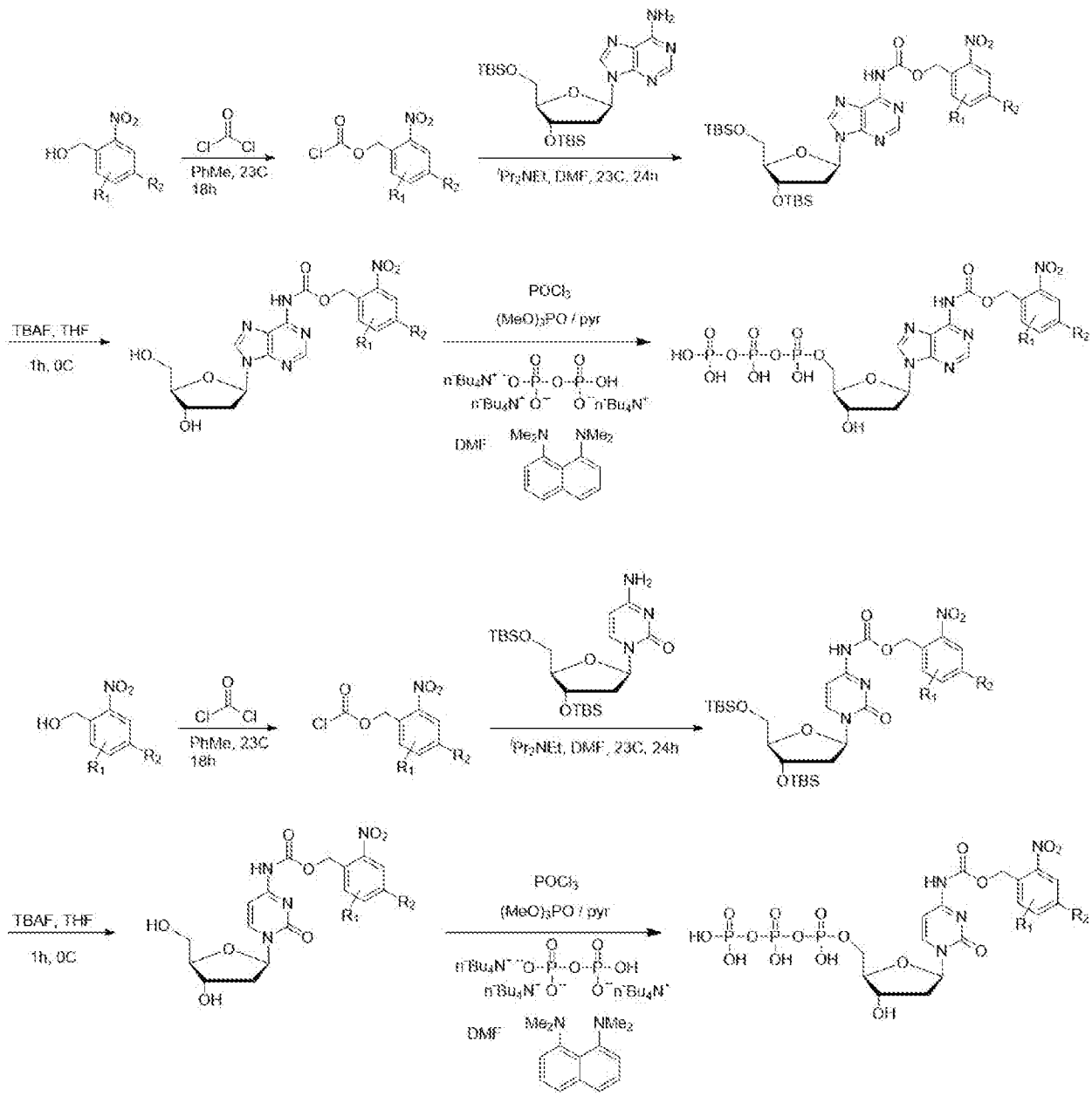


N-(4-(pyridine-2-yldisulfaneyl)butanyryl)-3'-O-*(2-nitrobenzyl)-2'-deoxycytidine* (22 mg, 1 Eq, 38 μmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 μL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 10 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

Example 11

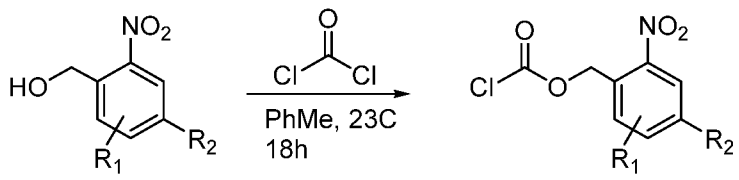
Procedures for synthesizing Class III – Purine & Pyrimidine dNTP analogs.

Schemes for synthesis of Class III dNTP analogs



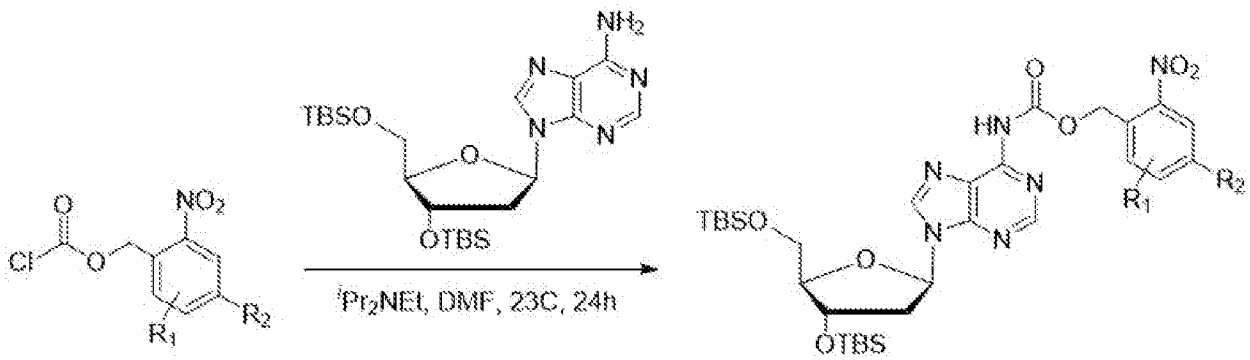
Example 12

Detailed Procedures for Class III – Purine dNTP analogs:

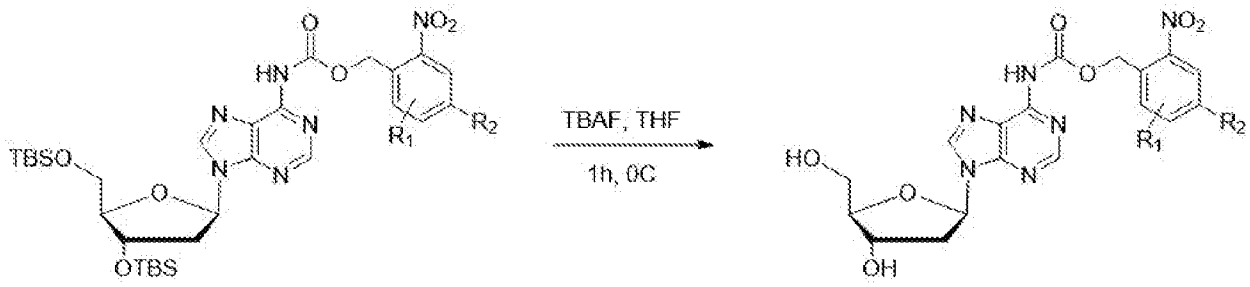


To a solution of phosgene (6.46 g, 2 Eq, 65.3 mmol) in dry Toluene (100 mL) at 23 °C was added (2-nitrophenyl) methanol (5.00 g, 1 Eq, 32.6 mmol) in 20 mL dry THF. The reaction was stirred for 24 hours at 23 °C. The reaction was then concentrated to dryness under a vacuum

trapped with a NaOH aqueous solution. The amber oil that remained was used directly in the reaction without further purification.

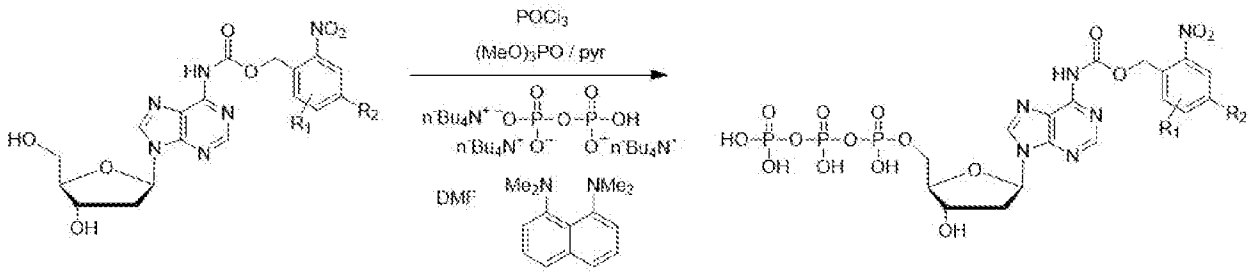


To a solution of 9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl) oxy)methyl) tetrahydrofuran-2-yl)-9H-purin-6-amine (12.2 g, 1.1 Eq, 25.5 mmol) in dry DMF (100 mL) at 0 °C was added N,N-diisopropylethylamine (3.60 g, 4.9 mL, 1.2 Eq, 27.8 mmol). The reaction was stirred for 30 minutes and then charged with 2-nitrobenzyl carbonochloridate (5.00 g, 1 Eq, 23.2 mmol) slowly drop-wise over 30 minutes keeping the temperature below 5C. The reaction was then allowed to warm to room temperature and stirred overnight. The reaction was poured into a cooled solution of 5% Na2CO3 and EtOAc. The EtOAc layer was dried with sodium sulfate and then concentrated to dryness. The crude product was chromatographed on silica gel using hexane / EtOAc mixtures to give a purified product which was used in the next reaction.



2-nitrobenzyl (9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl) oxy)methyl) tetrahydrofuran-2-yl)-9H-purin-6-yl)carbamate (5.00 g, 1 Eq, 7.59 mmol) was dissolved in THF at room temperature and then cooled to OC under a blanket of dry Argon. The mixture was then charged with tetrabutylammonium fluoride (4.96 g, 2.5 Eq, 19.0 mmol). The

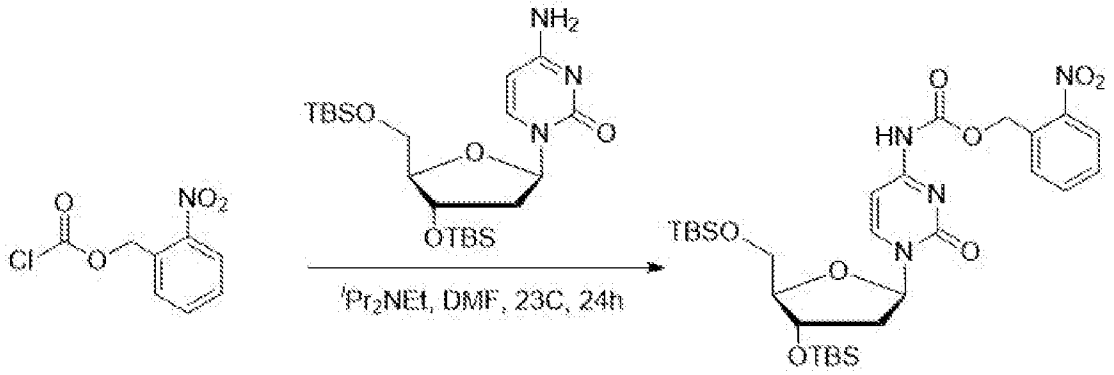
mixture was stirred for 2 hours at 0C and then warmed to 23C for 1 hour. The solution was poured into a cold solution of 10% NaHCO3 and extracted with DCM. The DCM layer was concentrated and the crude product purified on silica gel eluting with 5-50% DCM / MeOH to afford the product suitable for triphosphorylation.



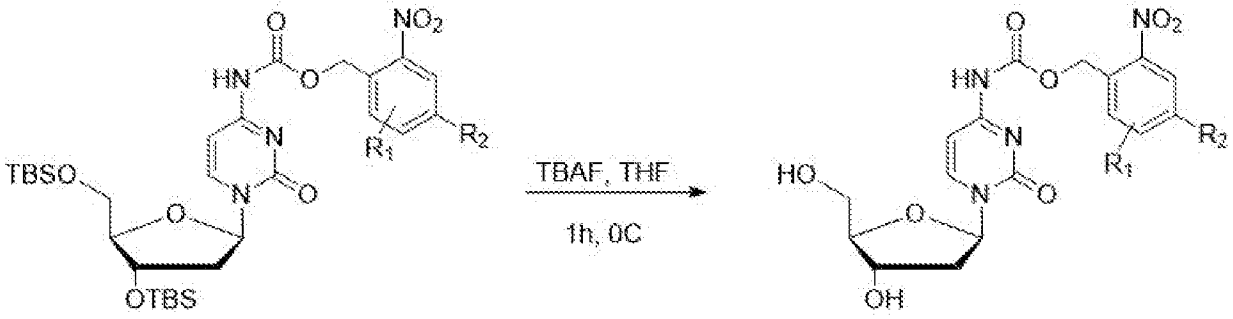
2-nitrobenzyl (9-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-9H-purin-6-yl)carbamate (28 mg, 1 Eq, 65 µmol) was dissolved in trimethyl phosphate (1.5 mL) and 0.60 mL of dry pyridine and cooled in an ice bath under Argon. A first aliquot of phosphoryl trichloride (30 mg, 18 µL, 3 Eq, 0.20 mmol) was added. Five minutes later, a second aliquot of 10 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.23 g, 4 Eq, 0.26 mmol) in dry DMF was prepared under Ar and cooled in an ice bath. This was added to the reaction mixture dropwise over 30 seconds at rxn t = 35 min. Immediately the pre-weighed N1,N1,N8,N8-tetramethyl-naphthalene-1,8-diamine (56 mg, 4 Eq, 0.26 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred for 30 min and then transferred to a separatory funnel. The solution was extracted 1X with 10mL of EtOAc. The aq. layer was transferred to a small tube for FPLC separation.

Example 13

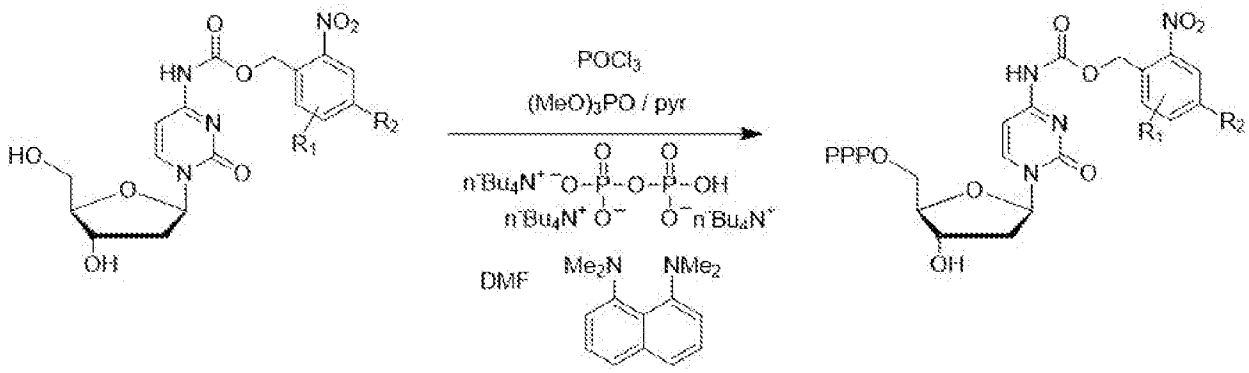
Detailed Procedures for Class III – Pyrimidine dNTP analogs:



To a solution of 4-amino-1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)-methyl)tetrahydrofuran-2-yl)pyrimidin-2(1H)-one (9.30 g, 1.1 Eq, 20.4 mmol) in dry DMF (100 mL) at 0 °C was added N,N-diisopropylethylamine (2.88 g, 3.9 mL, 1.2 Eq, 22.3 mmol) . The reaction was stirred for 30 minutes and then charged with 2-nitrobenzyl carbonochloridate (4.00 g, 1 Eq, 18.6 mmol) slowly dropwise over 30 minutes keeping the temperature below 5C. The reaction was then allowed to warm to room temperature and stirred overnight. The reaction was poured into a cooled solution of 5% Na2CO3 and EtOAc. The EtOAc layer was was dried with sodium sulfate and then concentrated to dryness. The crude product was chromatographed on silica gel using hexane / EtOAc mixtures to give a purified product which was used in the next reaction.



2-nitrobenzyl (1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)-tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-yl)carbamate (5.00 g, 1 Eq, 7.88 mmol) was dissolved in 25 mL THF at room temperature and then cooled to OC under a blanket of dry Argon. The mixture was then charged with tetrabutylammonium fluoride (5.15 g, 2.5 Eq, 19.7 mmol). The mixture was stirred for 2 hours at 0C and then warmed to 23C for 1 hour. The solution was poured into a cold solution of 10% NaHCO3 and extracted with DCM. The DCM layer was concentrated and the crude product purified on silica gel eluting with 5-50% DCM / MeOH to afford the product suitable for triphosphorylation.



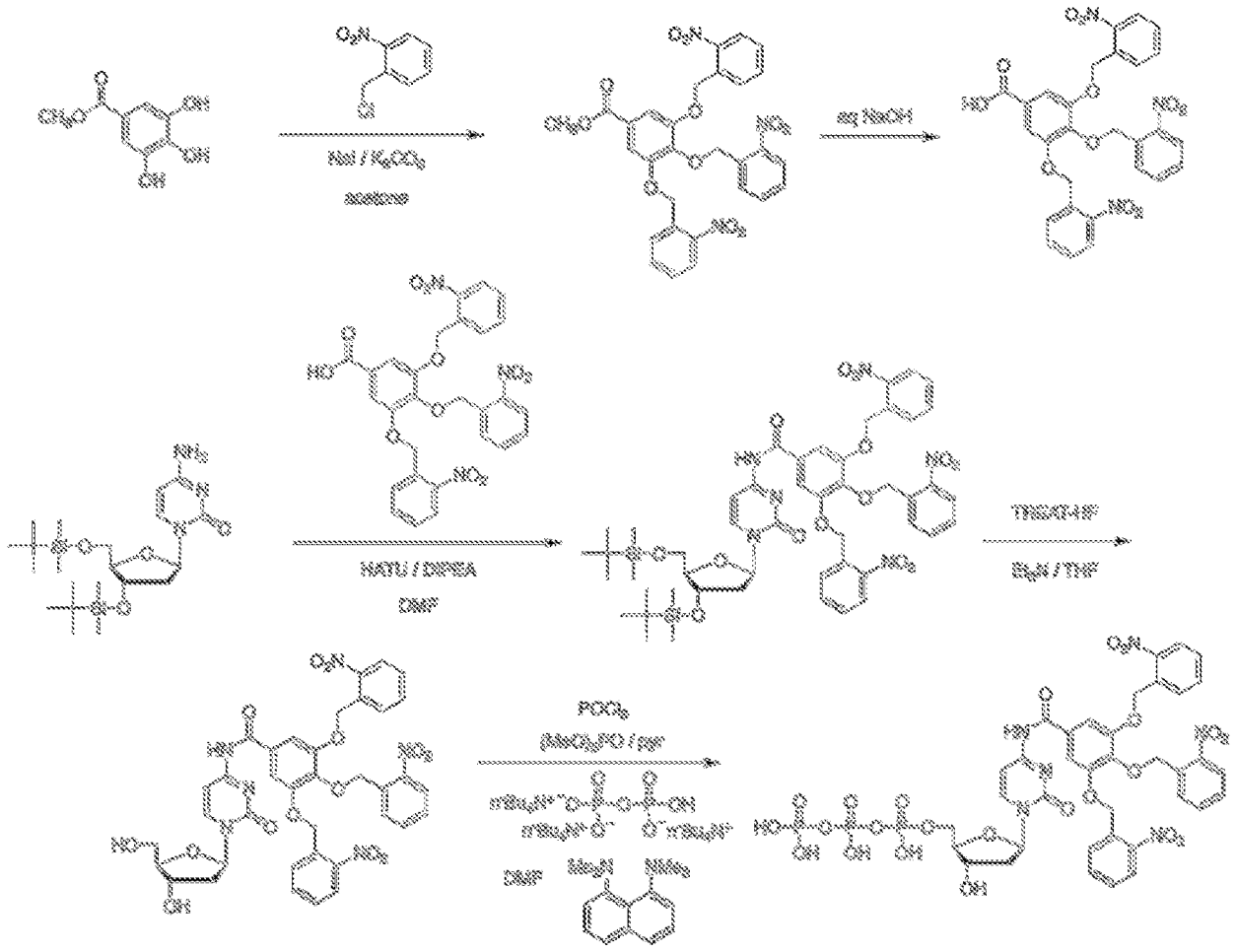
2-nitrobenzyl (1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-yl)carbamate (35.0 mg, 1 Eq, 86.1 µmol)was dissolved in trimethyl

phosphate (1.5 mL) and 0.60 mL of dry pyridine and cooled in an ice bath under Argon. A first aliquot of phosphoryl trichloride (39.6 mg, 3 Eq, 258 µmol) was added. Five minutes later a second aliquot of 10 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (311 mg, 4 Eq, 345 µmol) in dry DMF was prepared under Ar and cooled in an ice bath. This was added to the reaction mixture dropwise over 30 seconds at rxn t = 35 min. Immediately the pre-weighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (73.8 mg, 4 Eq, 345 µmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred for 30 min and then transferred to a separatory funnel. The solution was extracted 1X with 10mL of EtOAc. The aq. layer was transferred to a small tube for FPLC separation.

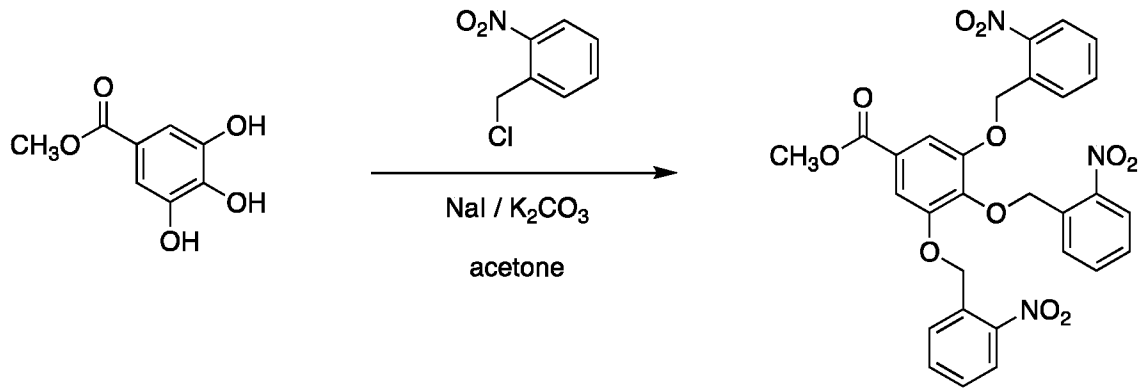
Example 14

Procedures for synthesizing Class IV – dCTP analog.

Scheme for synthesis of Class IV dCTP analog

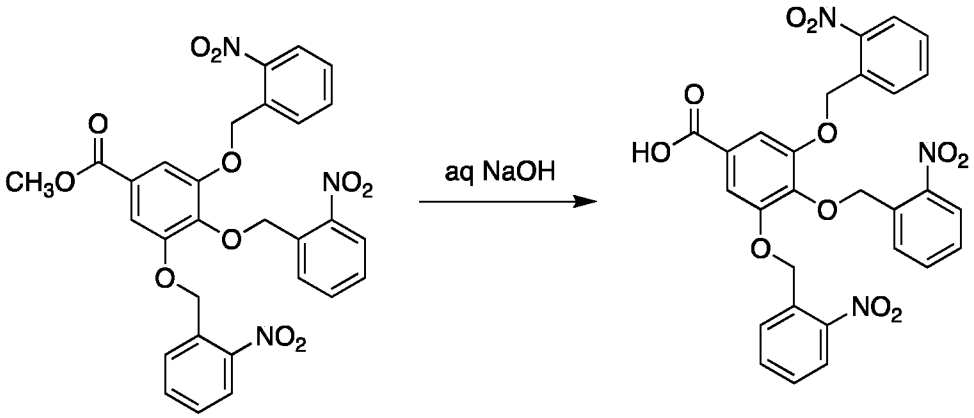


Detailed Procedures for Class IV dCTP analog:

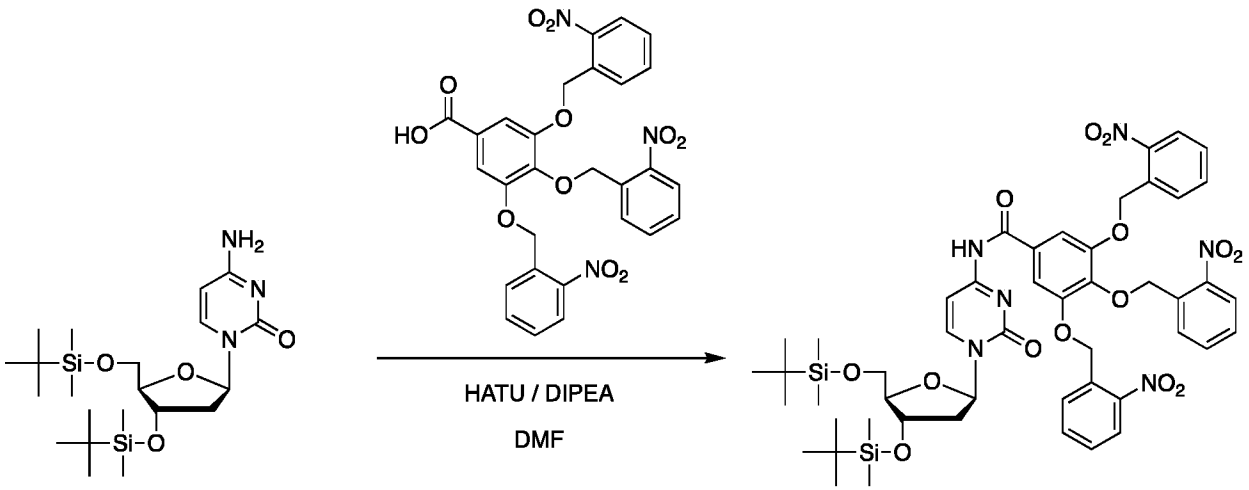


Methyl 3,4,5-trihydroxybenzoate (10 g, 1 Eq, 54 mmol) was dissolved in 50 ml of acetone. Sodium iodide (0.81 g, 0.1 Eq, 5.4 mmol) and potassium carbonate (38 g, 5 eq, 270

mmol) were added as solids at ambient temperature. 1-(Chloromethyl)-2-nitrobenzene (34 g, 3.6 Eq, 0.20 mol) was added dropwise as a solution in 40 mL of acetone over 10 minutes. The mixture was stirred for 1 hour and then heated to 50 °C for 6 hr. The mixture was cooled to ambient temperature and the bulk of the solvent was removed on a rotovap. The residue was suspended in 200 mL of EtOAc and this was washed successively with 200 mL portions of water and saturated aqueous NaCl solution. The EtOAc layer was dried with sodium sulfate and evaporated. The crude product was chromatographed on silica using hexane/EtOAc mixtures to give a purified product which can be used in the next reaction.

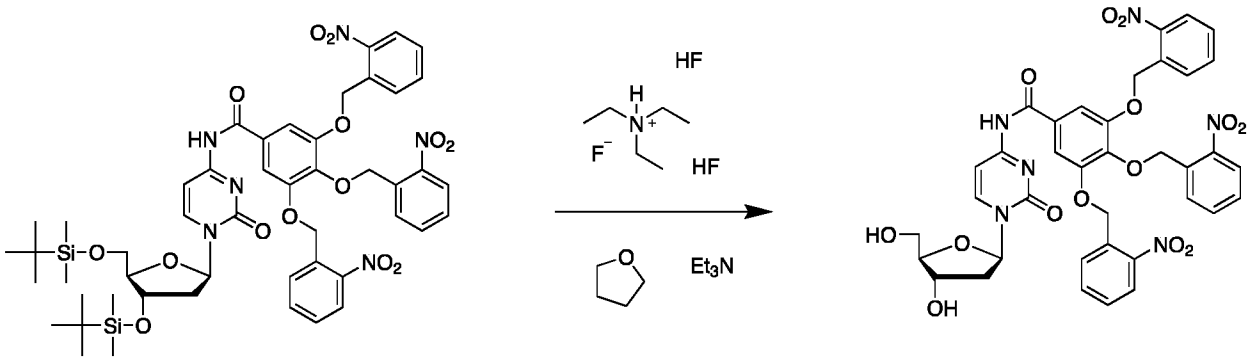


Methyl 3,4,5-tris((2-nitrobenzyl)oxy)benzoate (25 g, 1 Eq, 42 mmol) was dissolved in 300 mL of THF. Aqueous 2M NaOH (105 mL, 210 mmol) was added and the mixture was stirred at ambient T for 18 h. The bulk of the THF was removed on a rotovap and the residue was acidified slowly with 6 M HCl to a pH of 1 or less. The resulting solid was filtered and washed well with water and dried on the filter funnel for 5 hours and then under high vacuum for 18 h. The product was taken into the next reaction without further purification.

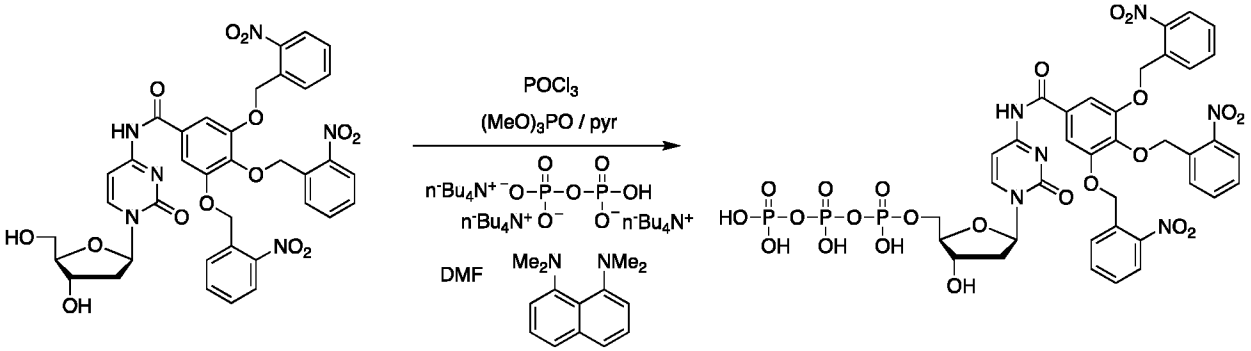


The 4-amino-1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl) oxy)methyl) tetrahydrofuran-2-yl)pyrimidin-2(1H)-one (12 g, 1 Eq, 26 mmol) and 3,4,5-tris((2-

nitrobenzyl) oxy) benzoic acid (15 g, 1 Eq, 26 mmol) were dissolved in 50 mL of dry DMF at ambient temperature under an argon atmosphere. N-Ethyl-N-isopropylpropan-2-amine (5.1 g, 6.8 mL, 1.5 Eq, 39 mmol) was added followed by a solution of 1-((dimethylamino)(dimethyliminio)methyl)-1H-[1,2,3]triazolo[4,5-b]pyridine 3-oxide hexafluorophosphate(V) (12 g, 1.2 Eq, 31 mmol) in 10 mL of dry DMF, added dropwise over 5 minutes at ambient temperature. The mixture was stirred at ambient temperature for 18 h. The mixture was dissolved in 300 mL of EtOAc and this was washed successively with 200 mL portions of water (2X) and saturated aqueous NaCl solution. the EtOAc was dried with sodium sulfate and evaporated first on a rotovap and then under high vacuum for 18 h. The residue was chromatographed on silica using mixtures of dichloromethane and methanol to give the desired product as a colorless foam.



N-(1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl) oxy) methyl) tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-yl)-3,4,5-tris ((2-nitrobenzyl) oxy) benzamide (5 g, 1 Eq, 5 mmol) was dissolved in 25.0 mL of dry THF at ambient T under Argon. triethylamine (4 g, 6 mL, 8 Eq, 4e+1 mmol) was added rapidly followed by triethylammonium fluoride dihydrofluoride (5 g, 5 mL, 6 Eq, 3e+1 mmol) also added rapidly at ambient temperature. The mixture was stirred at ambient T for 24 h. Silica gel (20 g) was added and the mixture was evaporated on a rotovap to a fine powder and then loaded onto a 100g silica column aand eluted with mixtures of dichloromethane and methanol to give the nucleoside as a slightly yellow foam.

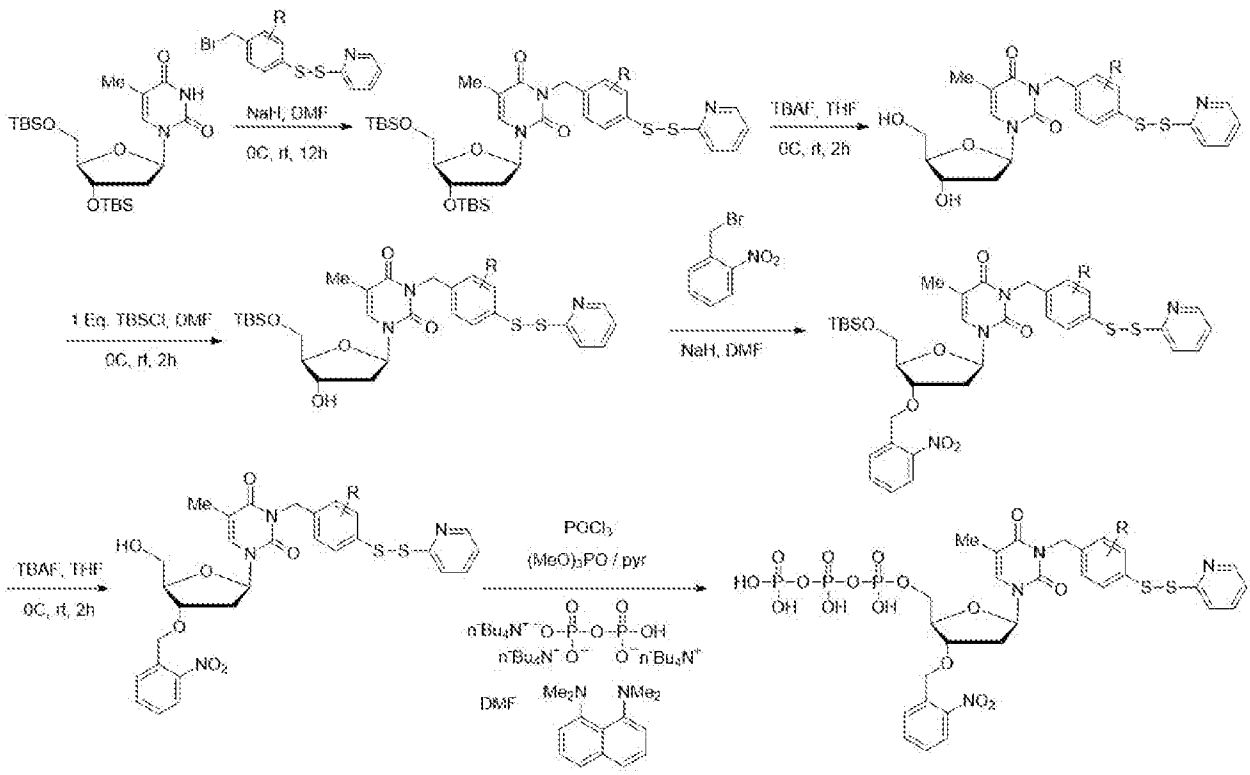


N-(1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-yl)-3,4,5-tris((2-nitrobenzyl)oxy)benzamide (30 mg, 1 Eq, 38 μmol) was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of 6 uL of phosphoryl trichloride (18 mg, 11 μL, 3 Eq, 0.11 mmol) was added. Five minutes later, a second aliquot of 5 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (0.14 g, 4 Eq, 0.15 mmol) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (33 mg, 4 Eq, 0.15 mmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

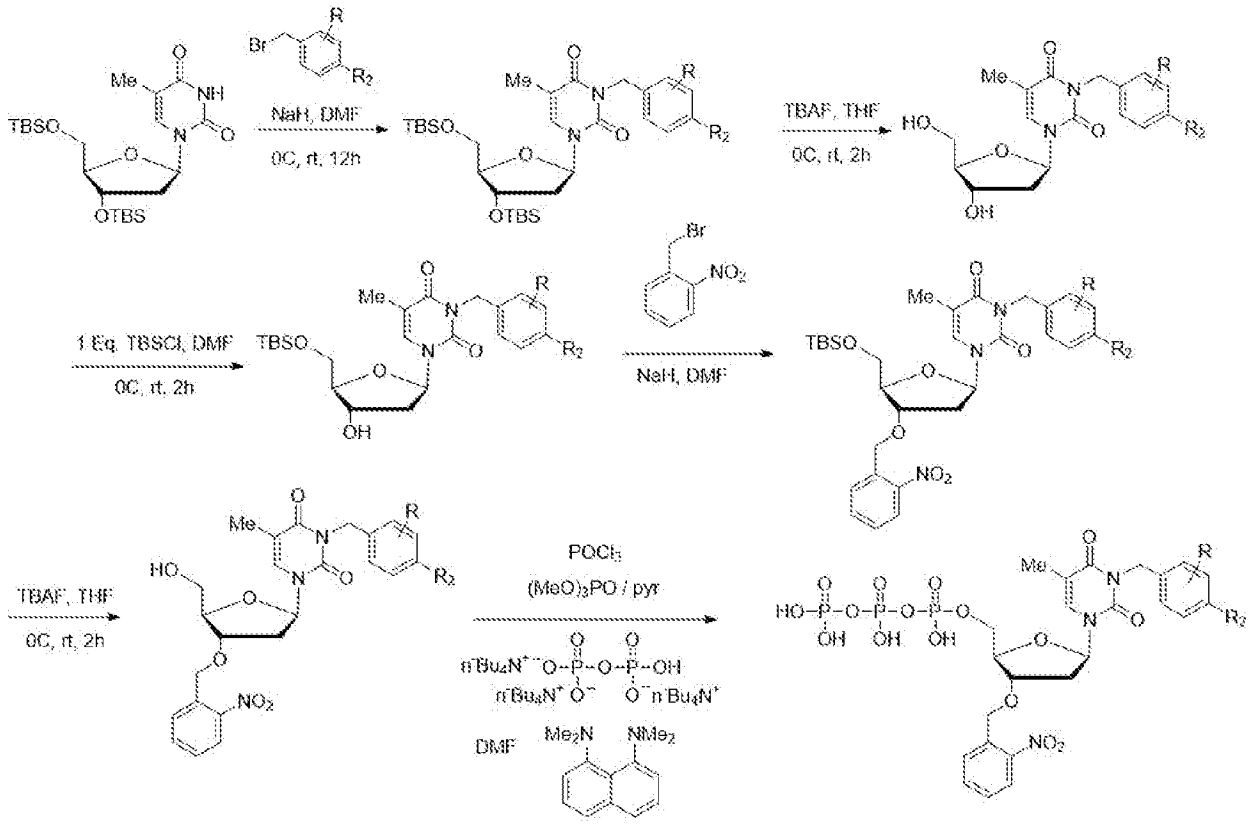
Example 15

Procedures for synthesizing Class V peptide & non-peptide analogs.

Scheme for synthesis of Peptide - Thymidine dNTP conjugates

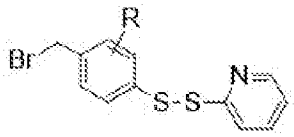


Scheme for synthesis of Non-Peptide - Thymidine dNTP conjugates

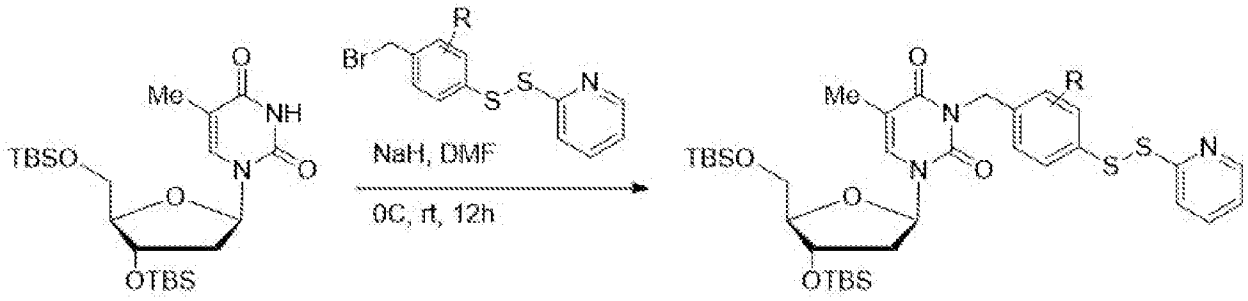


Example 16

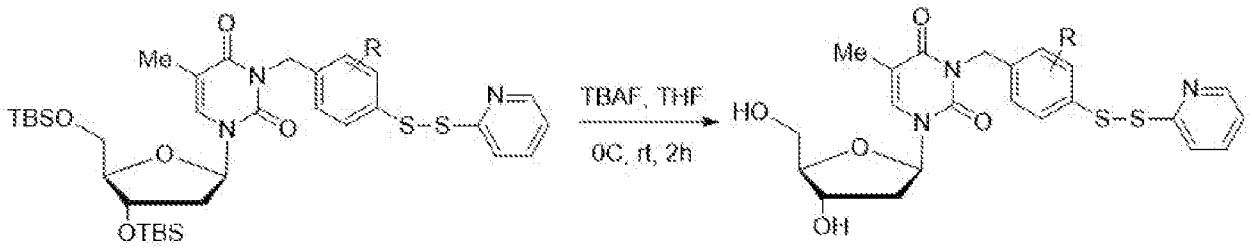
Detailed Procedures for Peptide – dTTP analogs:



4-(bromomethyl)benzenethiol (5.00 g, 1 Eq, 24.6 mmol) was dissolved in Methanol (50 mL) and cooled to 0C. The mixture was then charged with 1,2-di(pyridin-2-yl) disulfane (5.42 g, 1 Eq, 24.6 mmol) and stirred at 0C for 18 hours. The reaction was then concentrated directly and purified on silica get eluting with hexanes / Ethyl acetate (0-100% EtOAc) to afford the product as a white solid which was used directly in the next reaction.



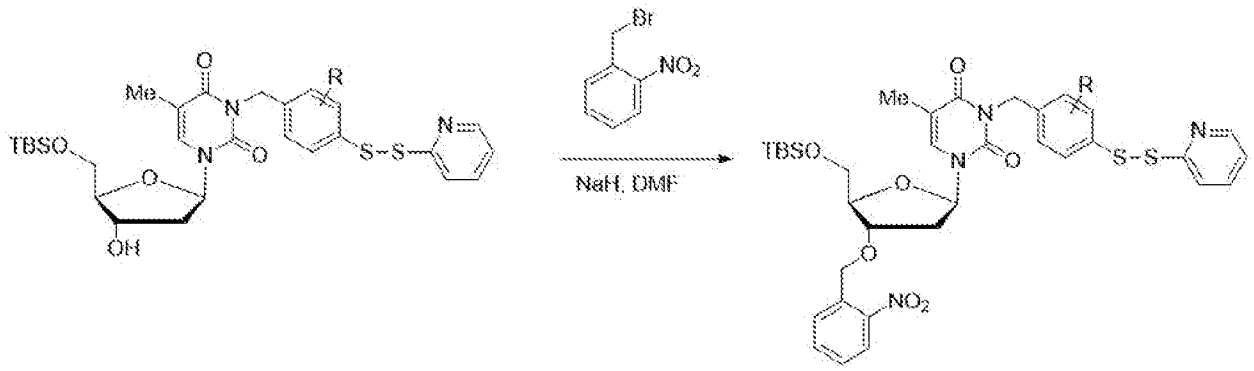
1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione (5.00 g, 1 Eq, 10.6 mmol) was dissolved in 50 mL DMF and then cooled to 0C. The reaction was stirred 30 minutes at 0C and then charged with sodium hydride (306 mg, 1.2 Eq, 12.7 mmol). The reaction was then allowed to stir an additional 30 minutes at 0C and then warmed to 23C. The mixture was then charged with 2-((4-(bromomethyl)phenyl)-disulfaneyl)-pyridine (3.32 g, 1 Eq, 10.6 mmol) and stirring was continued for an additional 2 hours at 23C. The reaction was then poured into a cold solution of 10% NaHCO3 and DCM. The DCM layer was separated and dried over sodium sulfate and concentrated to dryness. The mixture was then purified on silica gel eluting with 0-20% DCM / methanol to afford the desired product.



1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)-methyl)tetrahydrofuran-2-yl)-5-methyl-3-(4-(pyridin-2-yldisulfaneyl)benzyl)pyrimidine-

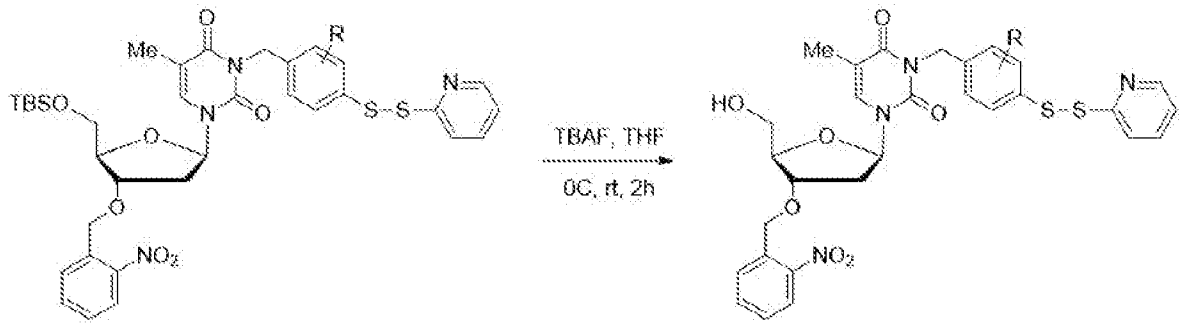
2,4(1H,3H)-dione (2.00 g, 1 Eq, 2.85 mmol) was dissolved in THF and cooled to 0C. The mixture was then charged with tetrabutylammonium fluoride (2.23 mg, 3 Eq, 8.55 mmol) at 0C. The reaction was kept stirring for 2 hours at 0C and then warmed to 23C for an additional hour. The reaction was then cooled again to 0C and charged into a pre-cooled solution of 10% NaHCO3 and DCM at 0C. The DCM layer was then separated and dried over sodium sulfate and concentrated and purified on silica gel eluting with 5-50% DCM / methanol to afford the pure product.

1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methyl-3-(4-(pyridin-2-yldisulfaneyl)benzyl)pyrimidine-2,4(1H,3H)-dione (5.00 g, 1 Eq, 10.6 mmol)was dissolved in 20 mL of THF at 23C. The mixture was then charged with triethylamine (1.07 g, 1.5 mL, 1 Eq, 10.6 mmol) and cooled to 0C. The mixture was then charged with TBS-Cl (1.59 g, 1 Eq, 10.6 mmol) and stirring was continued for an additional 2 hours at 0C. The mixture was then charged to a pre-cooled mixture of 10% aq. NaCl and DCM. The DCM layer was dried over sodium sulfate, concentrated and dried to afford an amber oil. The crude product was then purified on silica gel eluting with 5-50% DCM / methanol to afford the pure product.

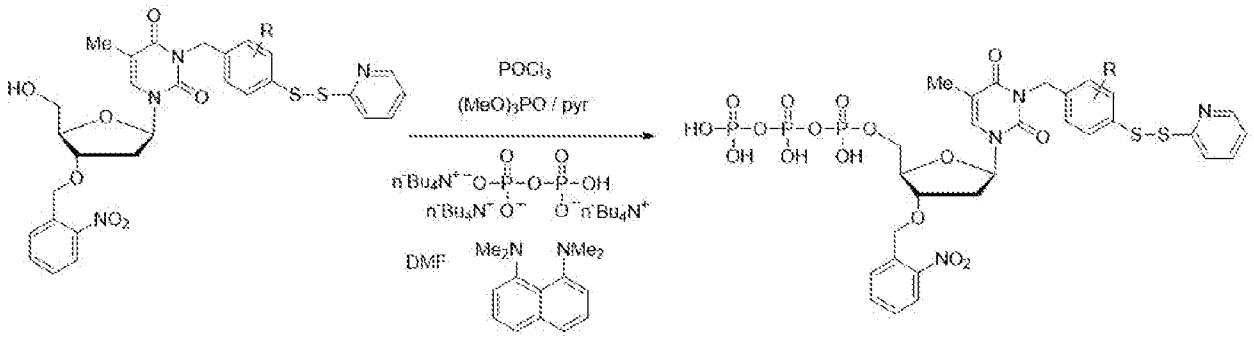


1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-hydroxytetrahydrofuran-2-yl)-5-methyl-3-(4-(pyridin-2-yldisulfaneyl)benzyl)pyrimidine-2,4(1H,3H)-dione (2.00 g, 1 Eq, 3.40 mmol) was dissolved in 20 mL DMF and then cooled to 0C. The mixture was then charged with sodium hydride (98.0 mg, 1.2 Eq, 4.08 mmol) and stirring was continued for an additional 30 minutes at 0C. The reaction was then charged with 1-(bromomethyl)-2-nitrobenzene (735 mg, 1 Eq, 3.40 mmol) and stirring was continued for an additional 1 hour at 0C. The reaction was then charge into a pre-cooled mixture 10% NaCl and EtOAc. The EtOAc layer was separated and dried

over sodium sulfate and concentrated to dryness. The crude product was purified on silica gel eluting with 0-50 hexanes / EtOAc to afford the desired product.



1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-((2-nitrobenzyl)oxy)tetrahydrofuran-2-yl)-5-methyl-3-(4-(pyridin-2-yldisulfaneyl)benzyl)pyrimidine-2,4(1H,3H)-dione (1.00 g, 1 Eq, 1.38 mmol) was dissolved in THF and cooled to 0C. The reaction was then charged with TBAF (362 mg, 1 Eq, 1.38 mmol) at 0C and stirred for 1 hour at 0C then then warmed to rt over the course of 2 hours. The mixture was then poured into a pre-cooled solution of 10% NaHCO3 and DCM. The DCM layer was separated and dried over sodium sulfate. The crude product was purified on silica gel eluting with 5-25% DCM / methanol to afford the desired product.

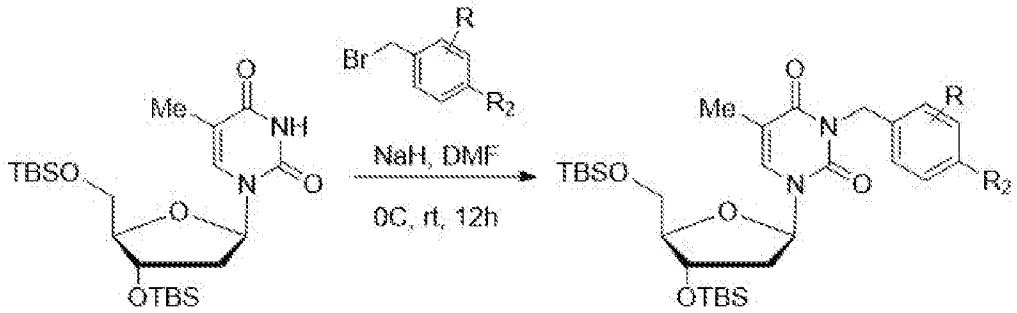


4-(pyridin-2-yl)benzyl (1-((2R,4S,5R)-5-(hydroxymethyl)-4-((2-nitrobenzyl)oxy)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-yl)carbamate (35.0 mg, 1 Eq, 61.0 µmol) was dissolved in trimethyl phosphate (1.5 mL) and 0.60 mL of dry pyridine and cooled in an ice bath under Argon. A first aliquot of phosphoryl trichloride (39.6 mg, 3 Eq, 258 µmol) was added. Five minutes later a second aliquot of 10 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (311 mg, 4 Eq, 345 µmol) in dry DMF was prepared under Ar and cooled in an ice bath. This was added to the reaction mixture dropwise over 30 seconds at rxn t = 35 min. Immediately the pre-weighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (73.8 mg, 4 Eq, 345 µmol) was added

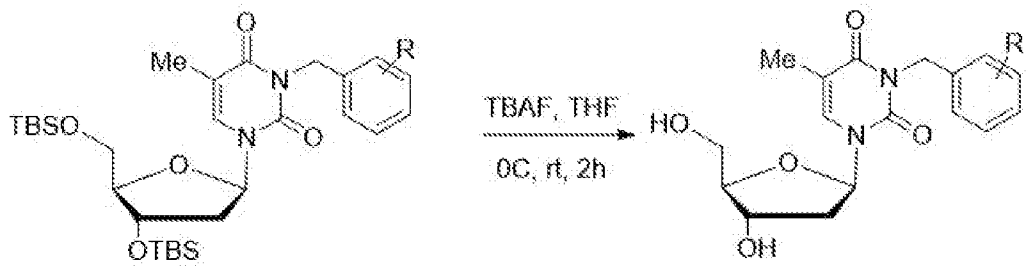
as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred for 30 min and then transferred to a separatory funnel. The solution was extracted 1X with 10mL of EtOAc. The aq. layer was transferred to a small tube for FPLC separation.

Example 17

Detailed Procedures for Non-peptide – dTTP Analog:



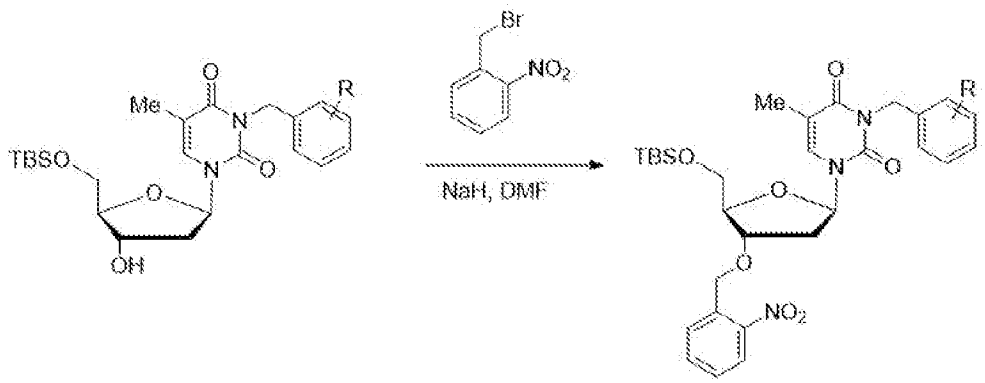
1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)-oxy)methyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione (5.00 g, 1 Eq, 10.6 mmol) was dissolved in 50 mL DMF and then cooled to 0C. The reaction was stirred 30 minutes at 0C and then charged with sodium hydride (306 mg, 1.2 Eq, 12.7 mmol) . The reaction was allowed to stir an additional 30 minutes at 0C and then warmed to 23C. The mixuture was then charged with (bromomethyl)benzene (1.82 g, 1 Eq, 10.6 mmol) and stirring was continued for an additional 2 hours at 23C. The reaction was then poured into a cold solution of 10% NaHCO3 and DCM. The DCM layer was separated and dried over sodium sulfate and concentrated to dryness. The mixture was then purified on silica gel eluting with 0-20% DCM / methanol to afford the desired product.



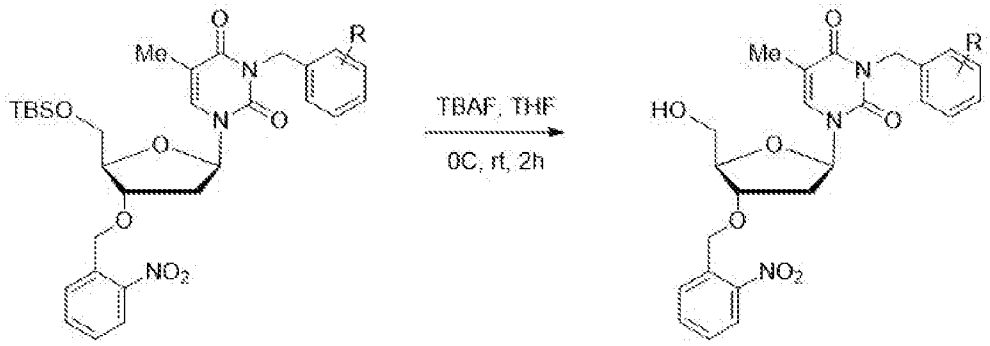
3-benzyl-1-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)-methyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione (4.00 g, 1 Eq, 7.13 mmol)was dissolved in THF and cooled to 0C. The mixture was then charged with tetrabutylammonium fluoride (3.72 g, 2 Eq, 14.26 mmol) at 0C. The reaction was kept stirring for 2 hours at 0C and then warmed to 23C for an additional hour. The reaction was then cooled again to 0C and charged

into a pre-cooled solution of 10% NaHCO3 and DCM at 0C. The DCM layer was then separated and dried over sodium sulfate and concentrated and purified on silica gel eluting with 5-50% DCM / methanol to afford the pure product.

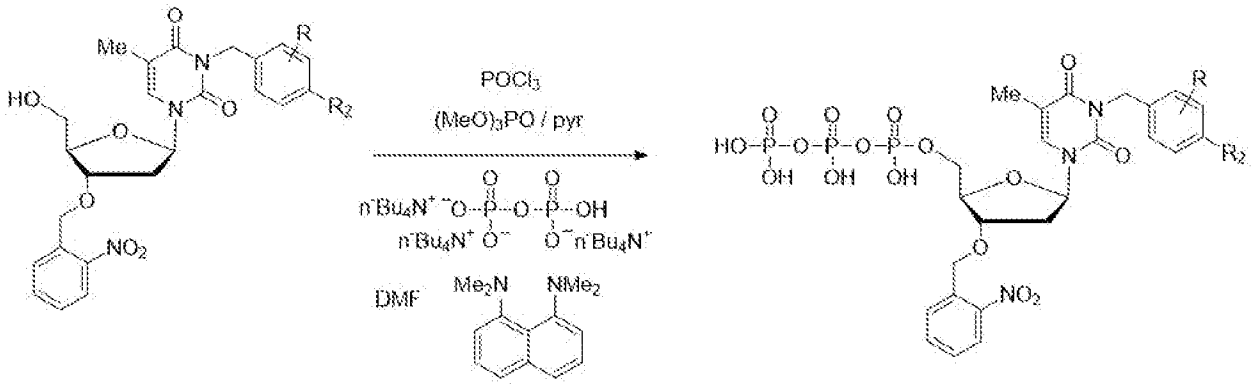
3-benzyl-1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione (1.00 g, 1 Eq, 3.01 mmol) was dissolved in 20 mL of THF at 23C. The mixture was then charged with triethylamine (304 mg, 0.42 mL, 1 Eq, 3.01 mmol) and cooled to 0C. The mixture was then charged with TBS-Cl (453 mg, 1 Eq, 3.01 mmol) and stirring was continued for an additional 2 hours at 0C. The mixture was then charged to a pre-cooled mixture of 10% aq. NaCl and DCM. The DCM layer was dried over sodium sulfate, concentrated and dried to afford an amber oil. The crude product was then purified on silica gel eluting with 5-50% DCM / methanol to afford the pure product.



3-benzyl-1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-hydroxytetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione (6.00 g, 1 Eq, 13.4 mmol) was dissolved in 20 mL DMF and then cooled to 0C. The mixture was then charged with sodium hydride (387 mg, 1.2 Eq, 16.1 mmol) and stirring was continued for an additional 30 minutes at 0C. The reaction was then charged with 1-(bromomethyl)-2-nitrobenzene (2.90 g, 1 Eq, 13.4 mmol) and stirring was continued for an additional 1 hour at 0C. The reaction was then charge into a pre-cooled mixture 10% NaCl and EtOAc. The EtOAc layer was separated and dried over sodium sulfate and concentrated to dryness. The crude product was purified on silica gel eluting with 0-50 hexanes / EtOAc to afford the desired product.



3-benzyl-1-((2R,4S,5R)-5-(((tert-butyldimethylsilyl)oxy)methyl)-4-((2-nitrobenzyl)oxy)-tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione (1.50 g, 1 Eq, 2.58 mmol)was dissolved in THF and cooled to 0C. The reaction was then charged with TBAF (674 mg, 1 Eq, 2.58 mmol) at 0C and stirred for 1 hour at 0C then then warmed to rt over the course of 2 hours. The mixture was then poured into a pre-cooled solution of 10% NaHCO3 and DCM. The DCM layer was separated and dried over sodium sulfate. The crude product was purified on silica gel eluting with 5-25% DCM / methanol to afford the desired product.



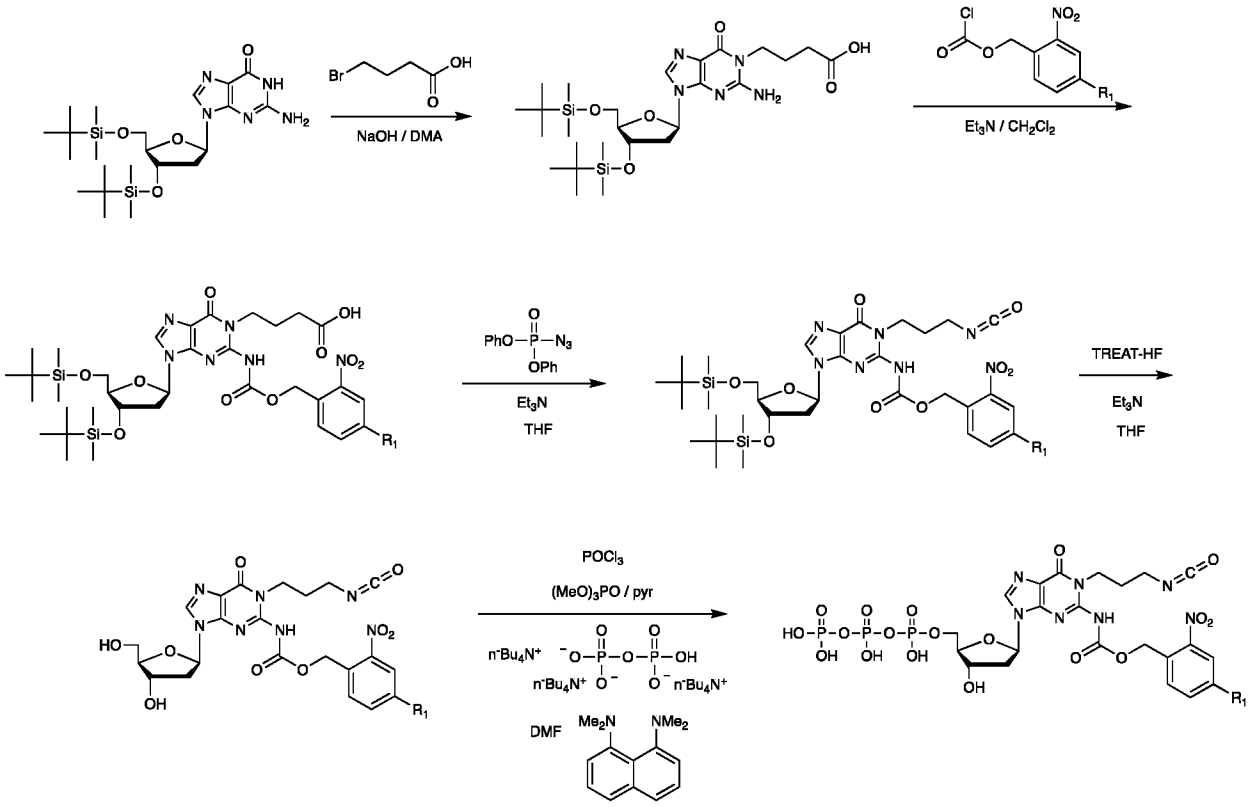
benzyl (1-((2R,4S,5R)-5-(hydroxymethyl)-4-((2-nitrobenzyl)oxy)tetrahydrofuran-2-yl)-2-oxo-1,2-dihydropyrimidin-4-yl)carbamate (35.0 mg, 1 Eq, 70.5 µmol) was dissolved in trimethyl phosphate (1.5 mL) and 0.60 mL of dry pyridine and cooled in an ice bath under Argon. A first aliquot of phosphoryl trichloride (39.6 mg, 3 Eq, 258 µmol) was added. Five minutes later a second aliquot of 10 uL was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (311 mg, 4 Eq, 345 µmol) in dry DMF was prepared under Ar and cooled in an ice bath. This was added to the reaction mixture dropwise over 30 seconds at rxn t = 35 min. Immediately the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (73.8 mg, 4 Eq, 345 µmol) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred for 30 min and then transferred to a separatory funnel. The solution was

extracted 1X with 10mL of EtOAc. The aq. layer was transferred to a small tube for FPLC separation.

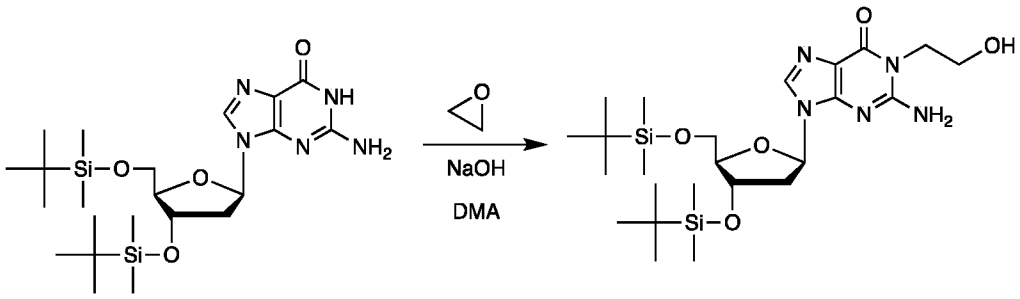
Example 18

Procedures for synthesizing Class VI peptide & non-peptide dGTP analogs.

Schemes for synthesis of Class VI – dGTP constructs

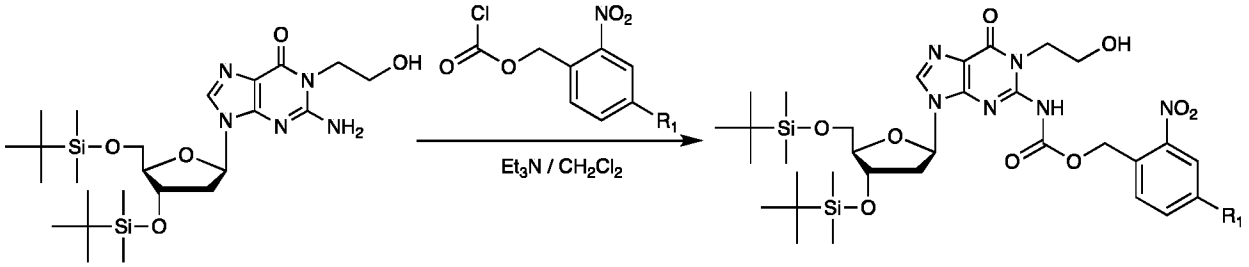


Detailed Procedures for non-peptide dGTP analogs:

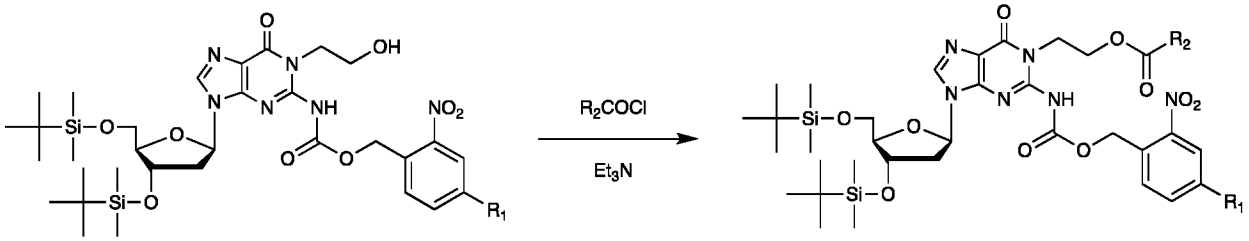


2-Amino-9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tertbutyldimethylsilyl) oxy) methyl) tetrahydrofuran-2-yl)-1,9-dihydro-6H-purin-6-one (0.50 g, 1 Eq, 1.0 mmol) was dissolved in 5.0 mL of dry dimethylacetamide under argon. oxirane (0.13 g, 3 Eq, 3.0 mmol) was added at ambient temperature followed by sodium hydroxide (40 mg, 1 Eq, 1.0 mmol) as a solid. The

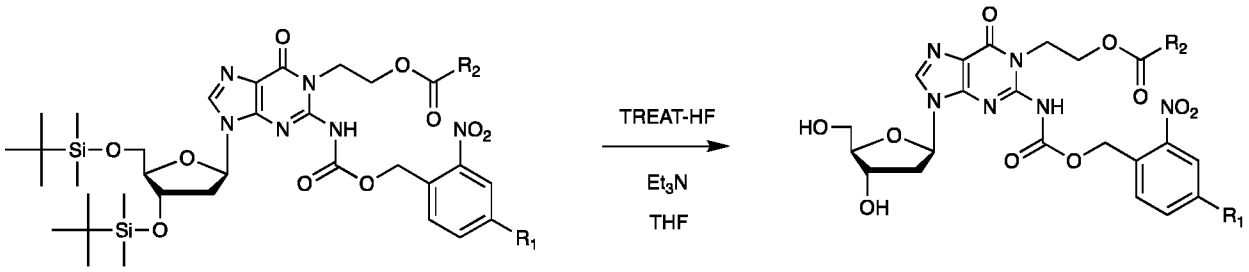
mixture was stirred at ambient temperature for 4 h. The mixture was diluted with 50 mL of EtOAc and this was washed successively with 100 mL of water and 100 mL of brine. The EtOAc layer was dried with sodium sulfate and evaporated to leave a yellow oil. This was chromatogaphed on 40 g of silica using dichloromethane/methanol mixtures as eluent to provide a white foam.



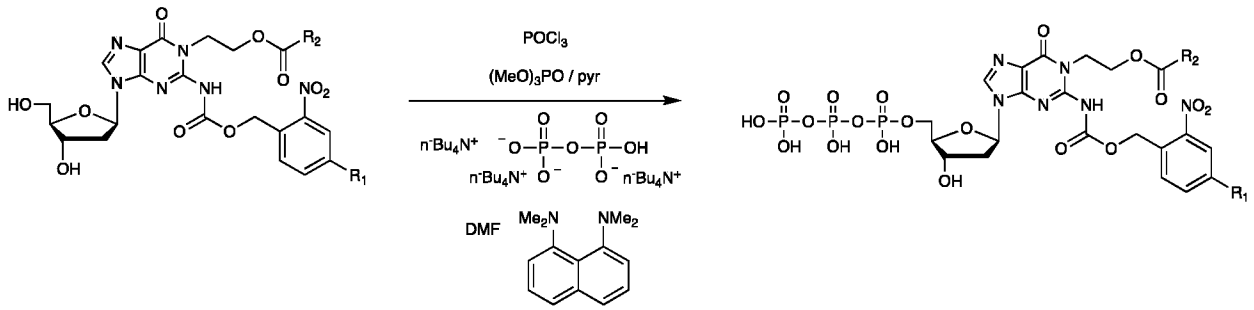
2-Amino-9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-1-(2-hydroxyethyl)-1,9-dihydro-6H-purin-6-one (200 mg, 1 Eq, 370 µmol) was suspended in 5 mL of dry pyridine at ambient temperature under argon. The chloroformate (1 equiv), added as a solid. The mixture was heated to 95 C for 8 h and cooled to ambient T. The solvent was removed in vacuo and the residue was diluted with 50 mL of EtOAc and this was washed successively with 50 mL of water and 50 mL of brine. The EtOAc layer was dried with sodium sulfate and evaporated to leave a yellow oil. This was chromatogaphed on 40 g of silica using dichloromethane/methanol mixtures as eluent to provide a white foam.



The alcohol starting material was dissolved in dry THF at ambient temperature under argon. Two equivalents of triethylamine were added. A solution of the acyl chloride in THF was added dropwise at ambient temperature and the mixture was stirred for 18 h. The solvent was removed in vacuo and the residue was diluted with 50 mL of EtOAc and this was washed successively with 50 mL of water and 50 mL of brine. The EtOAc layer was dried with Na2SO4 and evaporated to leave a light brown solid. This was chromatogaphed on a silica column using dichloromethane/methanol mixtures as eluent to provide the corresponding ester.

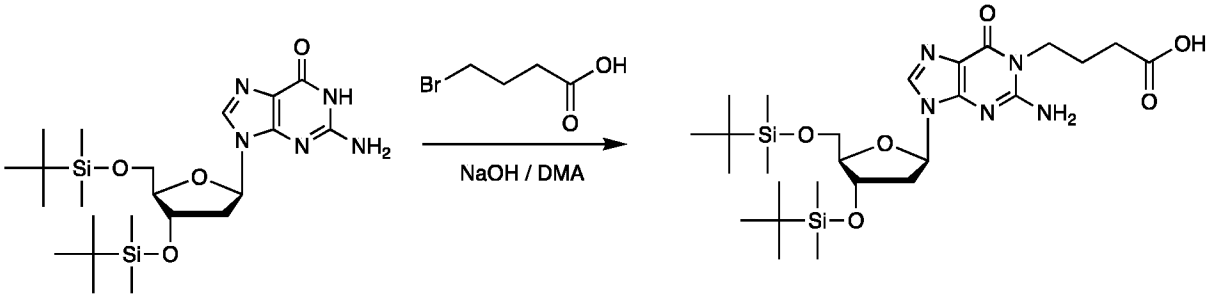


The bis-silyl ether (1 eq) was dissolved in dry THF at ambient T under Argon. Triethylamine (8 Eq) was added rapidly followed by triethylammonium fluoride dihydrofluoride (6 Eq) also added rapidly at ambient temperature. The mixture was stirred at ambient temperature for 24 h. Silica gel was added and the mixture was evaporated on a rotovap to a fine powder and then loaded onto a silica column and eluted with mixtures of dichloromethane and methanol to give the nucleoside as a slightly yellow foam.

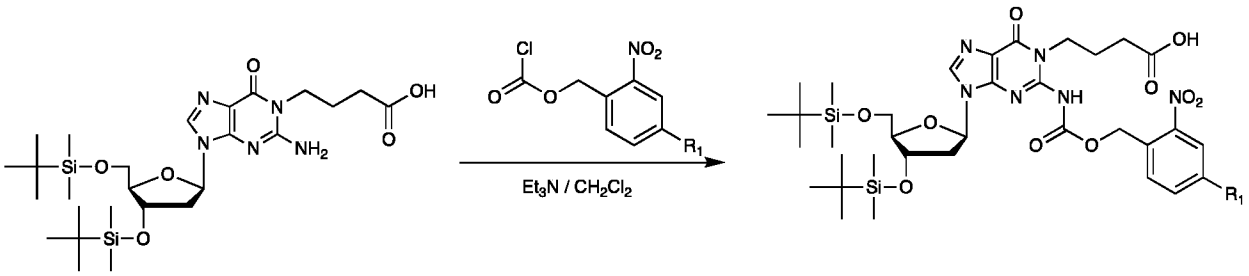


The nucleoside was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of phosphoryl trichloride (1.5 eq) was added. Five minutes later, a second aliquot of 1.5 eq was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (4 Eq) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (4 Eq) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

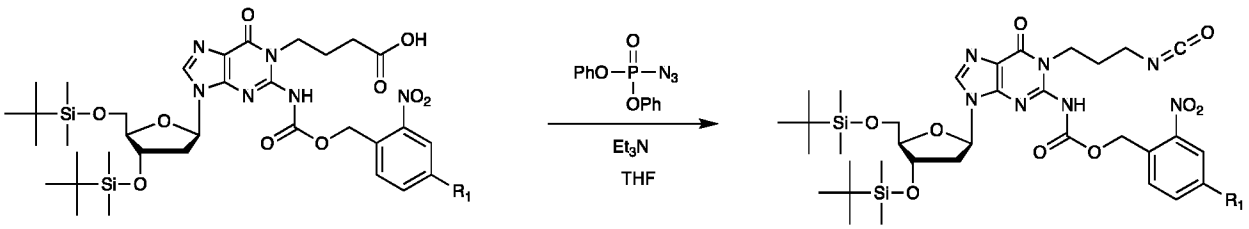
Detailed Procedures for peptide – dGTP analogs:



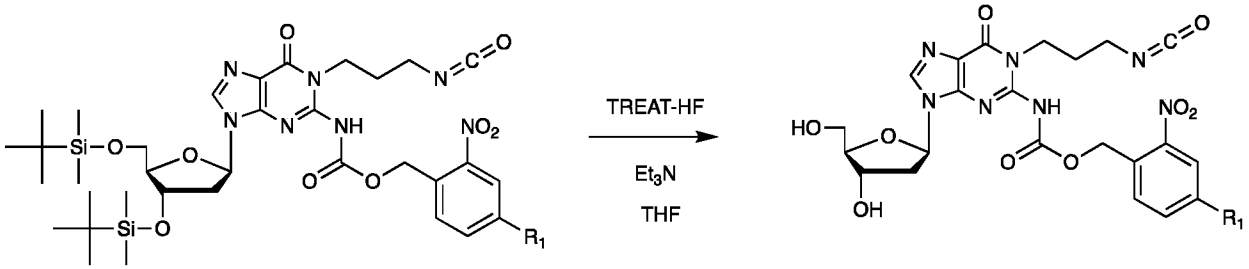
2-Amino-9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-1,9-dihydro-6H-purin-6-one (1.00 g, 1 Eq, 2.02 mmol) was dissolved in 30 mL of dry N,N-dimethylacetamide under argon. 4-bromobutanoic acid (337 mg, 1 Eq, 2.02 mmol) was added at ambient temperature followed by sodium hydroxide (161 mg, 2 Eq, 4.03 mmol), added as a solid. The mixture was heated to 80 C and stirred for 12 h. Th mixture was cooled to ambient temperature and diluted with 100 mL of EtOAc and this was washed successively with 50 mL of water and 50 mL of brine. The EtOAc layer was dried with Na2SO4 and evaporated to leave a light brown solid. This was chromatogaphed on a silica column using dichloromethane/methanol mixtures as eluent to provide 4-(2-amino-9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-6-oxo-6,9-dihydro-1H-purin-1-yl)butanoic acid as a white solid.



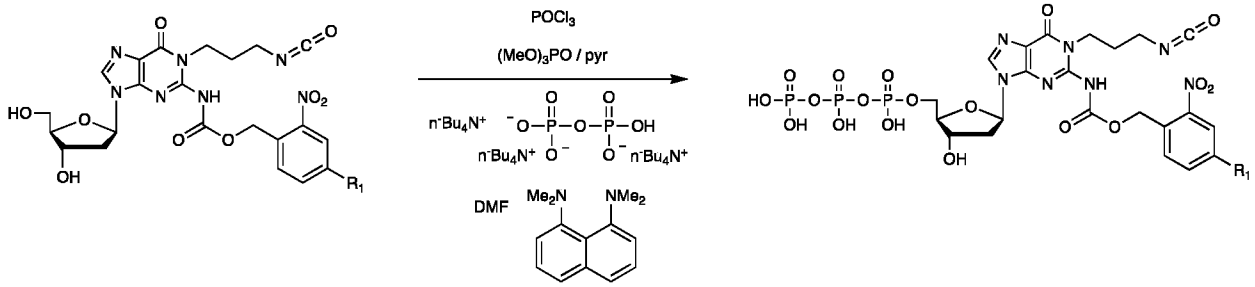
4-(2-amino-9-((2R,4S,5R)-4-((tert-butyldimethylsilyl)oxy)-5-(((tert-butyldimethylsilyl)oxy)methyl)tetrahydrofuran-2-yl)-6-oxo-6,9-dihydro-1H-purin-1-yl)butanoic acid (1 Eq) was suspended in 5 mL of dry pyridine at ambient temperature under argon. The chloroformate (1 equiv), added as a solid. The mixture was heated to 95 C for 8 h and cooled to ambient T. The solvent was removed in vacuo and the residue was diluted with 50 mL of EtOAc and this was washed successively with 50 mL of water and 50 mL of brine. The EtOAc layer was dried with sodium sulfate and evaporated to leave a yellow oil. This was chromatogaphed on 40 g of silica using dichloromethane/methanol mixtures as eluent to provide a white foam.



The carboxylic acid was dissolved or suspended in dry THF at ambient T. To this solution was added 1.3 eq of triethylamine followed by 1.1 eq of diphenylphosphoryl azide. The mixture was heated to reflux for 20 h and cooled to ambient T. Silica gel was added to the mixture and the solvent were evaporated to give a fine powder. This was loaded onto a column of silica gel and eluted with mixtures of EtOAc and dichloromethane to give the desired isocyanate as a colorless oil.



The bis-silyl ether (1 eq) was dissolved in dry THF at ambient T under Argon. Triethylamine (8 Eq) was added rapidly followed by triethylammonium fluoride dihydrofluoride (6 Eq) also added rapidly at ambient temperature. The mixture was stirred at ambient temperature for 24 h. Silica gel was added and the mixture was evaporated on a rotovap to a fine powder and then loaded onto a silica column and eluted with mixtures of dichloromethane and methanol to give the nucleoside as a slightly yellow foam.



The nucleoside was co-evaporated with pyridine (1 mL x 3) and dried on high vac overnight. It was then dissolved in 1.5 mL of trimethylphosphate and 0.60 mL dry pyridine and cooled in an ice bath under argon. A first aliquot of phosphoryl trichloride (1.5 eq) was added. Five minutes later, a second aliquot of 1.5 eq was added. The mixture was stirred an additional 30 min. A solution of tetrabutylammonium hydrogen diphosphate (4 Eq) in 1.5 mL dry DMF was prepared under Ar and cooled in an ice bath. This was added to the rxn mixture

dropwise over 30 sec. Immediately, the preweighed N1,N1,N8,N8-tetramethylnaphthalene-1,8-diamine (4 Eq) was added as a solid in one portion. The mixture was stirred for 30 min after this addition and was quenched with 8 mL of cold 0.1 M TEAB buffer. The mixture was stirred in the ice bath for 10 min and then transferred to a separatory funnel. The solution was extracted 1X with 10 mL of EtOAc. The aq layer was transferred to a small tube for FPLC separation which was conducted immediately after the EtOAc extraction. Final purification was by reverse phase HPLC.

Decaging of a 3'-O-(2-nitro-benzyl)-dATP and homopolymer synthesis is shown in FIGS. 20 and 21. 25 uM of 3'-O-(2-nitro-benzyl)-dATP (TriLink Technologies, San Diego, CA) was mixed with 1uM of an oligonucleotide initiator (5'-biotin-TTTTTTGGCCTTTTUTAATAATAATAATAATTTTT, IDT) with 1X TdT reaction Buffer (Thermo-Fisher), 2U/uL of terminal deoxynucleotidyl transferase (Thermo-Fisher), and 0.002 U/uL inorganic pyrophosphatase (Thermo-Fisher). The reaction volume was the subjected to 20-22 mW/cm2 of light at 365 nm for various intervals and then allowed to sit for 30 minutes at 37oC. After quenching by the addition of 0.1 M EDTA, each timepoint was mixed with an equal volume of 2X Novex TBE-urea gel loading buffer (Thermo-Fisher) and analyzed by polyacrylamide gel electrophoresis (15%), stained with Sybr Gold (Thermo-Fisher) and photographed with a ultraviolet transilluminator.

Incorporation by Reference

References and citations to other documents, such as patents, patent applications, patent publications, journals, books, papers, web contents, have been made throughout this disclosure. All such documents are hereby incorporated herein by reference in their entirety for all purposes.

Equivalents

Various modifications of the invention and many further embodiments thereof, in addition to those shown and described herein, will become apparent to those skilled in the art from the full contents of this document, including references to the scientific and patent literature cited herein. The subject matter herein contains important information, exemplification and guidance that can be adapted to the practice of this invention in its various embodiments and equivalents thereof.

CLAIMS

1. A method of synthesizing a plurality of nucleic acid memory strands, the method comprising:

providing an array of two or more substrate-linked nucleic acids with addressable delivery of activation energy to each of the substrate-linked nucleic acids;

extending one or more of the substrate-linked nucleic acids with a homopolymer tract of two or more repeating nucleotides by delivering addressable activation energy to one or more of the substrate-linked nucleic acids in the presence of a plurality of blocked nucleotide analogs and a template-independent polymerase, wherein the template-independent polymerase incorporates unblocked nucleotide analogs and not blocked nucleotide analogs and wherein the addressable activation energy converts the blocked nucleotide analogs into unblocked nucleotide analogs.

2. The method of claim 1, wherein the blocked nucleotide analog is converted to an unblocked nucleotide analog by removal of a blocking group.

3. The method of claim 1, wherein the addressable activation energy comprises light, reducing conditions, pH change, or heat.

4. The method of claim 2, wherein the removable blocking group is at the 3'-OH of the blocked nucleotide analog.

5. The method of claim 2, wherein the removable blocking group is on the purine or pyrimidine base of the nucleotide analog.

6. The method of claim 1, wherein the blocked nucleotide analog comprises a removable blocking group on a 3'-OH of the deoxyribose or ribose of a nucleotide triphosphate and a non-removable modification on a purine or pyrimidine base of the nucleotide analog.

7. The method of claim 1, wherein the plurality of blocked nucleotide analogs are modified nucleotides of a same nucleobase comprising removable 3'-O-blocking groups and two or more non-removable molecular modifications that allow differentiation between the modified nucleotide analogs of the same nucleobase.

8. The method of claim 1 further comprising:

stopping the extension;

extending the homopolymer tract with an additional homopolymer tract of two or more repeating nucleotides by delivering addressable activation energy to the homopolymer tract in the presence of another plurality of blocked nucleotide analogs and the template-independent polymerase.

9. The method of claim 8, wherein the extension is stopped after a predetermined length of time in order to obtain a desired length for the homopolymer tract.

10. The method of claim 1, wherein rate of extension is modulated by modifications to the blocked nucleotide analogs.

11. The method of claim 10, wherein the rate modulating modifications are removed from the homopolymer tract after extension.

12. The method of claim 1, wherein the rate modulating modifications are removed during extension.

13. The method of claim 1, wherein the repeating nucleotides of the homopolymer tract are between 2 and about 10.

14. The method of claim 8, further comprising repeating the stopping and extending steps to synthesize a nucleic acid memory strand.

15. The method of claim 14, wherein the nucleic acid memory strand is from about 200 nucleotides in length to about 5,000 nucleotides in length.

16. The method of claim 1, wherein a predetermined concentration of the blocked nucleotide analogs is provided in the extending step to obtain a desired length for the homopolymer tract.

17. The method of claim 8, wherein the homopolymer tract and additional homopolymer tract comprise different nucleobases.

18. The method of claim 14 wherein the nucleic acid memory strand encodes a dataset selected from the group consisting of a text file, an image file, and an audio file.

19. The method of claim 18, further comprising displaying a readable format of the dataset.

20. The method of claim 14, wherein a unit of data is represented in base 2.

21. The method of claim 14, wherein a unit of data is represented in base 3.

22. The method of claim 14, wherein a unit of data is represented in base 4.

23. The method of claim 14, wherein a unit of data is represented in greater than base 4.

24. The method of claim 14, wherein the data is represented by the extend of decaging or resultant tract length at individual steps of memory strand synthesis.

25. The method of claim 14, wherein data encoded in the nucleic acid memory strand is retrieved by DNA sequencing.

26. The method of claim 14, wherein data encoded in the nucleic acid memory strand is retrieved by passage of the nucleic acid memory strand through a nanopore.

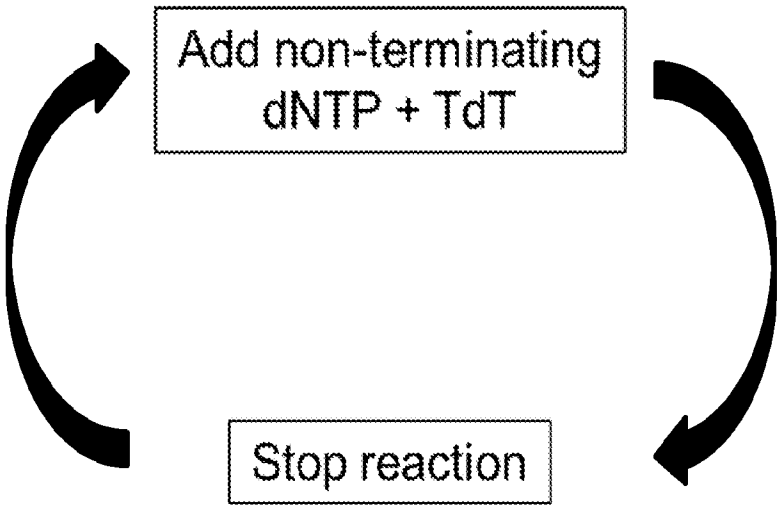


FIG. 1

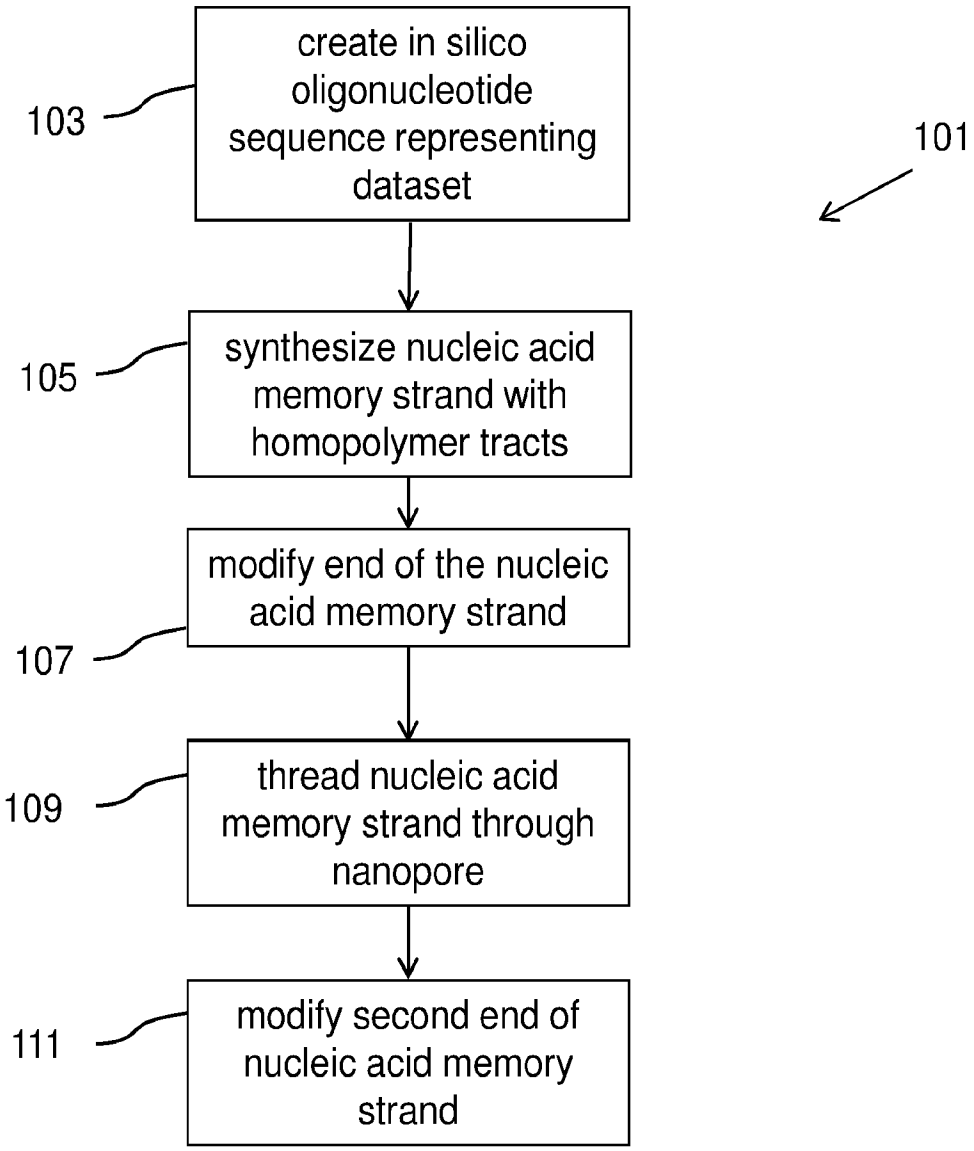

create in silico oligonucleotide sequence representing dataset
103
101
synthesize nucleic acid memory strand with homopolymer tracts
105
modify end of the nucleic acid memory strand
107
thread nucleic acid memory strand through nanopore
109
modify second end of nucleic acid memory strand
111

FIG. 2

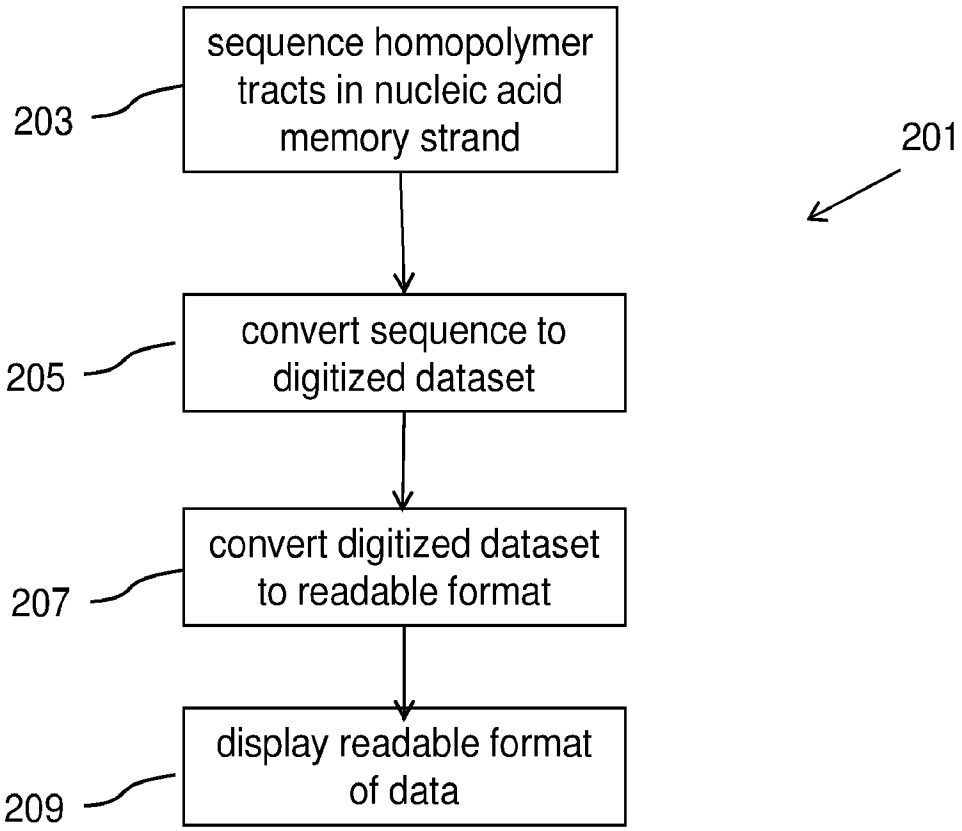

sequence homopolymer tracts in nucleic acid memory strand
203
201
convert sequence to digitized dataset
205
convert digitized dataset to readable format
207
display readable format of data
209

FIG. 3

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	2	500	3.2E+07
4	2	500	1.6E+07
6	2	500	1.2E+07
8	2	500	1.1E+07
12	2	500	8.3E+06
16	2	500	8.3E+06

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	5	500	8.0E+07
4	5	500	4.0E+07
6	5	500	3.1E+07
8	5	500	2.7E+07
12	5	500	2.3E+07
16	5	500	2.0E+07

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	10	500	1.6E+08
4	10	500	8.0E+07
6	10	500	6.2E+07
8	10	500	5.3E+07
12	10	500	4.5E+07
16	10	500	4.0E+07

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	2	1000	1.6E+07
4	2	1000	8.0E+06
6	2	1000	6.2E+06
8	2	1000	5.3E+06
12	2	1000	4.5E+06
16	2	1000	4.0E+06

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	5	1000	4.0E+07
4	5	1000	2.0E+07
6	5	1000	1.5E+07
8	5	1000	1.3E+07
12	5	1000	1.1E+07
16	5	1000	1.0E+07

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	10	1000	8.0E+07
4	10	1000	4.0E+07
6	10	1000	3.1E+07
8	10	1000	2.7E+07
12	10	1000	2.2E+07
16	10	1000	2.0E+07

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	2	10000	1.6E+06
4	2	10000	8.0E+05
6	2	10000	6.2E+05
8	2	10000	5.3E+05
12	2	10000	4.5E+05
16	2	10000	4.0E+05

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	5	10000	4.0E+06
4	5	10000	2.0E+06
6	5	10000	1.5E+06
8	5	10000	1.3E+06
12	5	10000	1.1E+06
16	5	10000	1.0E+06

Base n	Homopolymer Length	Strand Length	# Strands/GB
2	10	10000	8.0E+06
4	10	10000	4.0E+06
6	10	10000	3.1E+06
8	10	10000	2.7E+06
12	10	10000	2.2E+06
16	10	10000	2.0E+06

FIG. 4

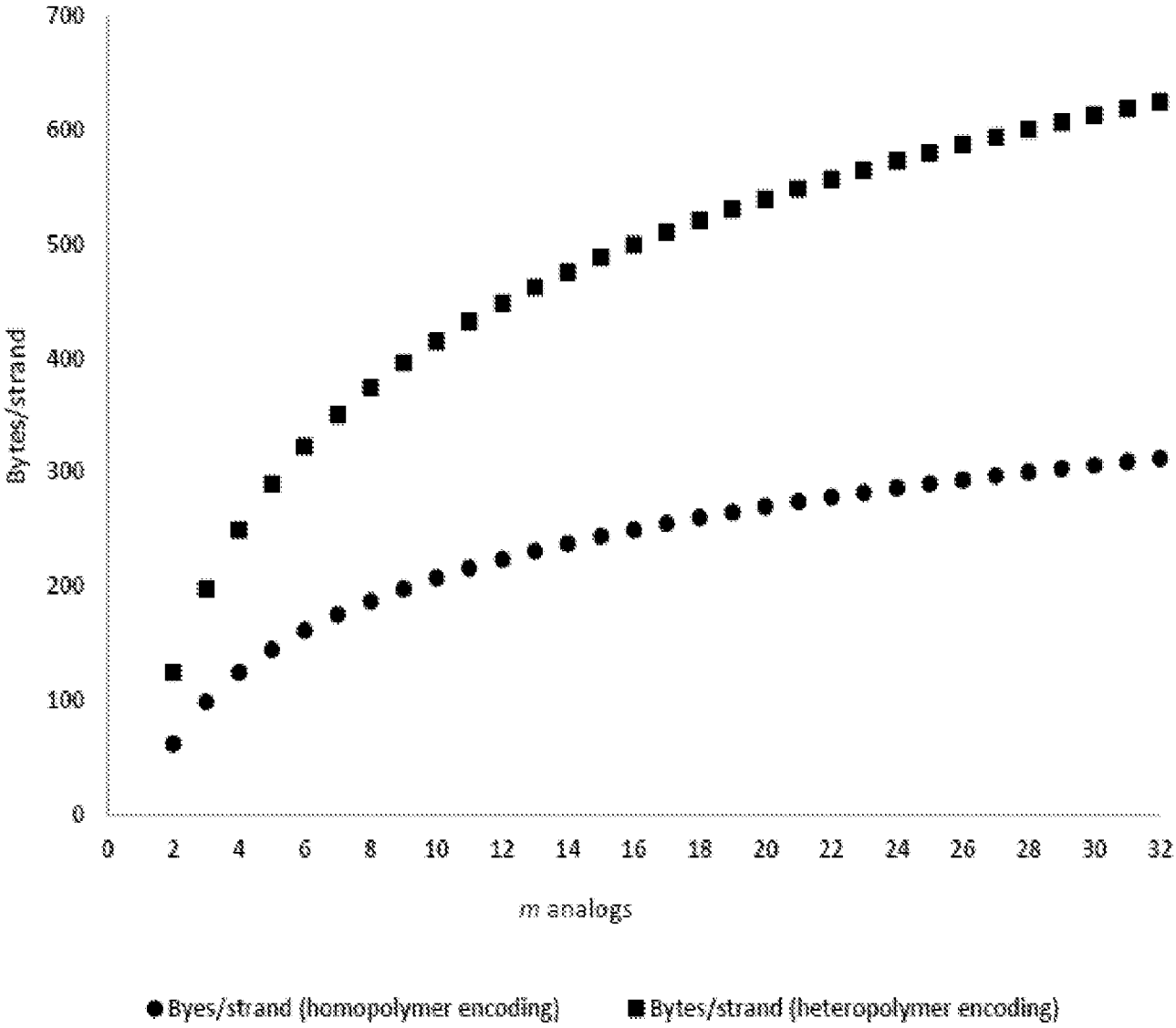

700
600
500
400
300
200
100
0
Bytes/strand
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
m analogs
Byes/strand (homopolymer encoding)
Bytes/strand (heteropolymer encoding)

FIG. 5

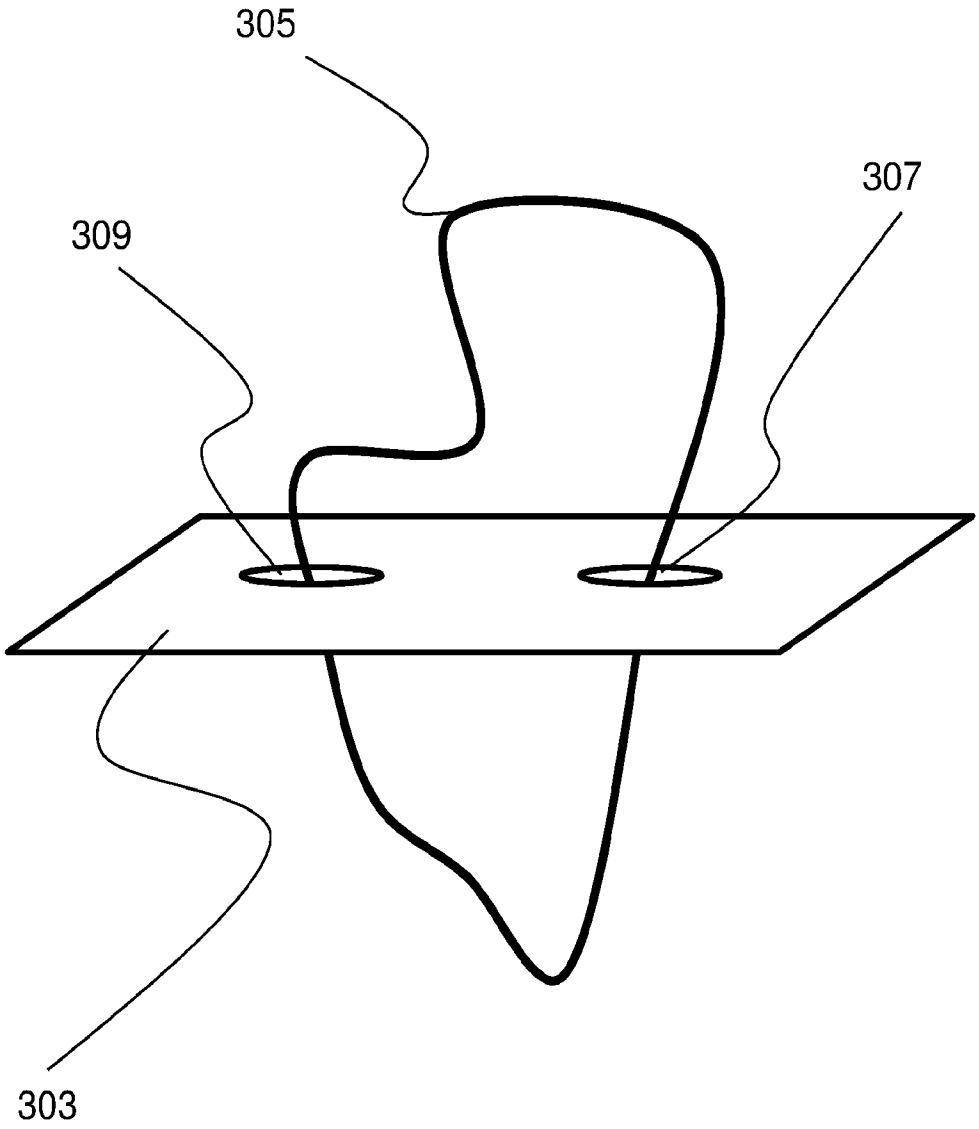
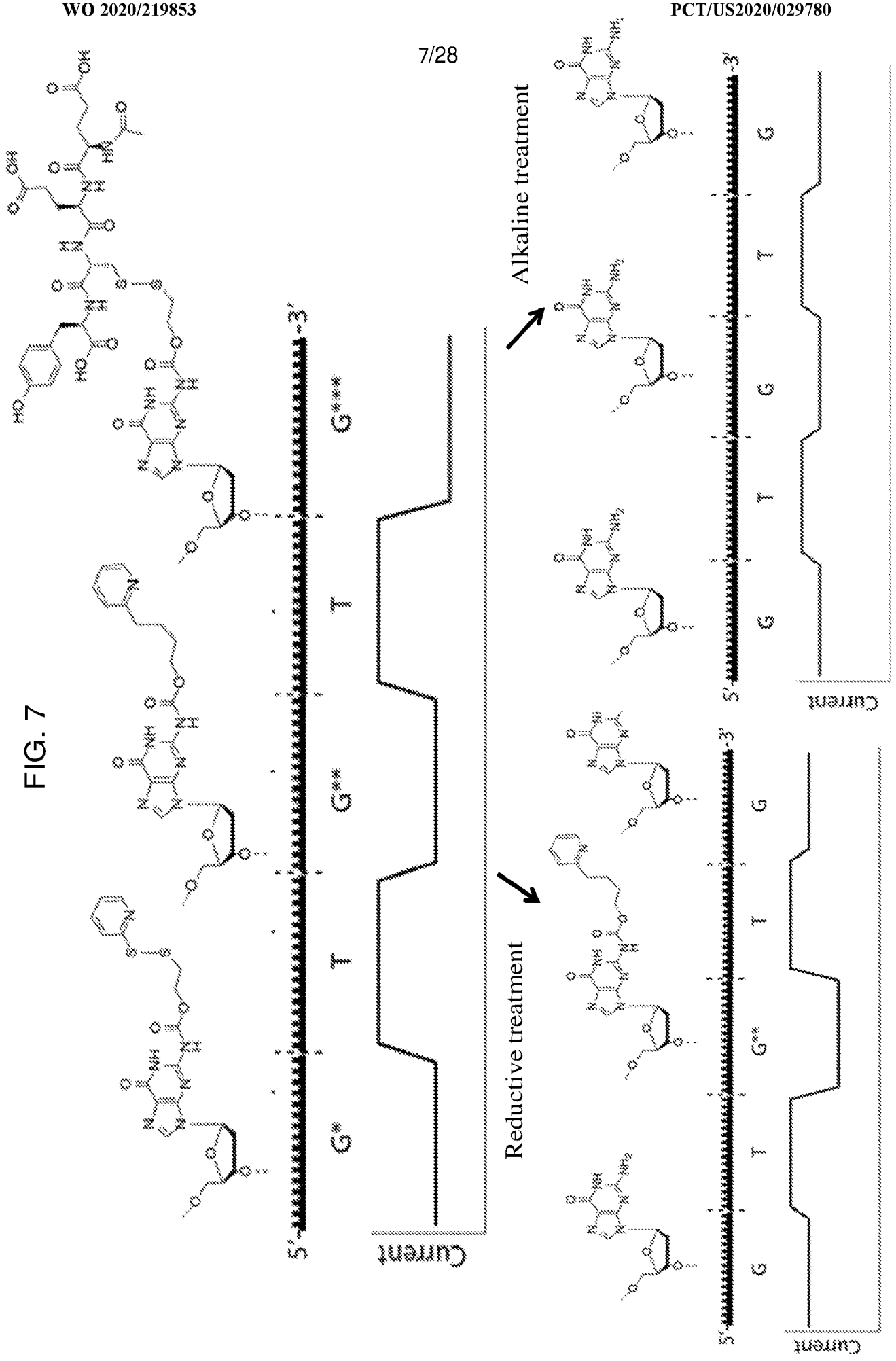


FIG. 6

FIG. 7


5'
3'
G*
T
G**
T
G***
Current
Reductive treatment
Alkaline treatment
G
T
G**
T
G
Current
G
T
G**
T
G
Current

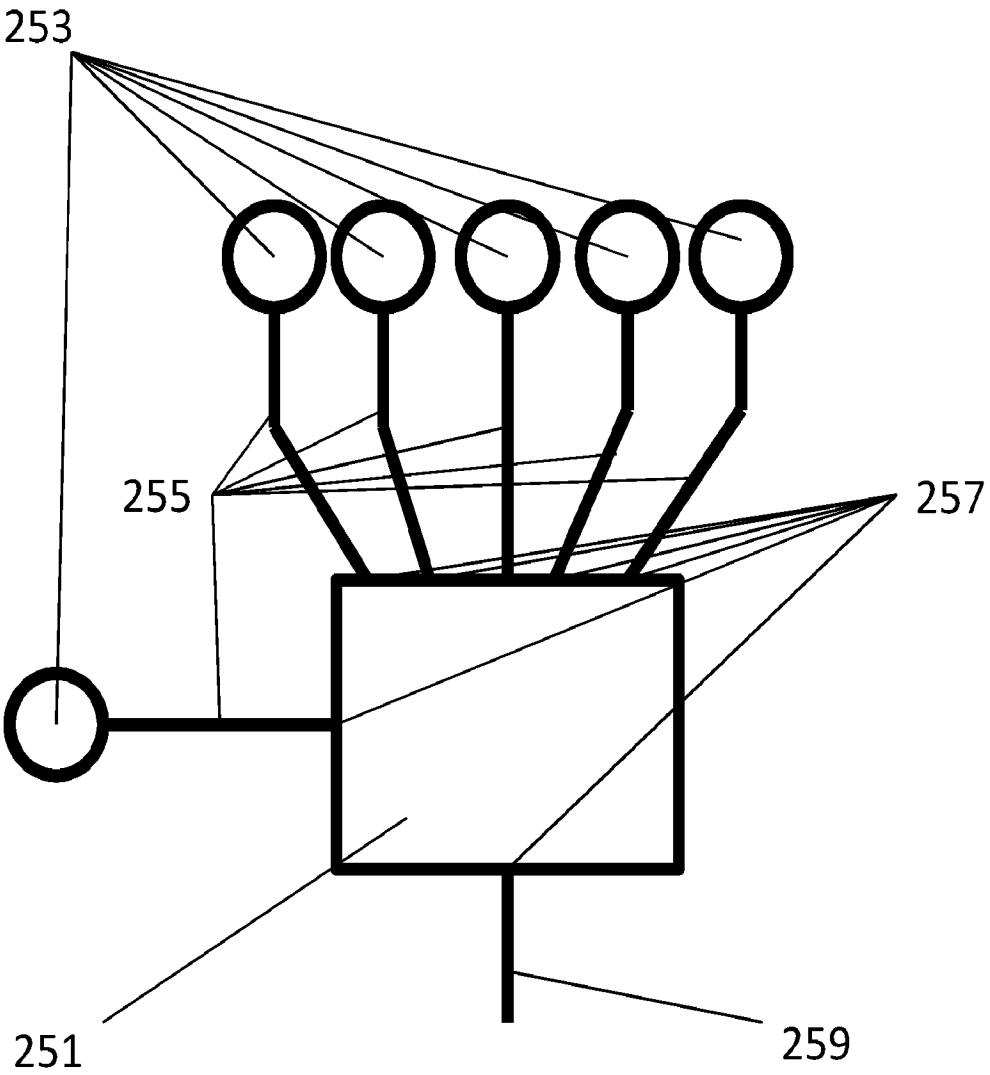


FIG. 8

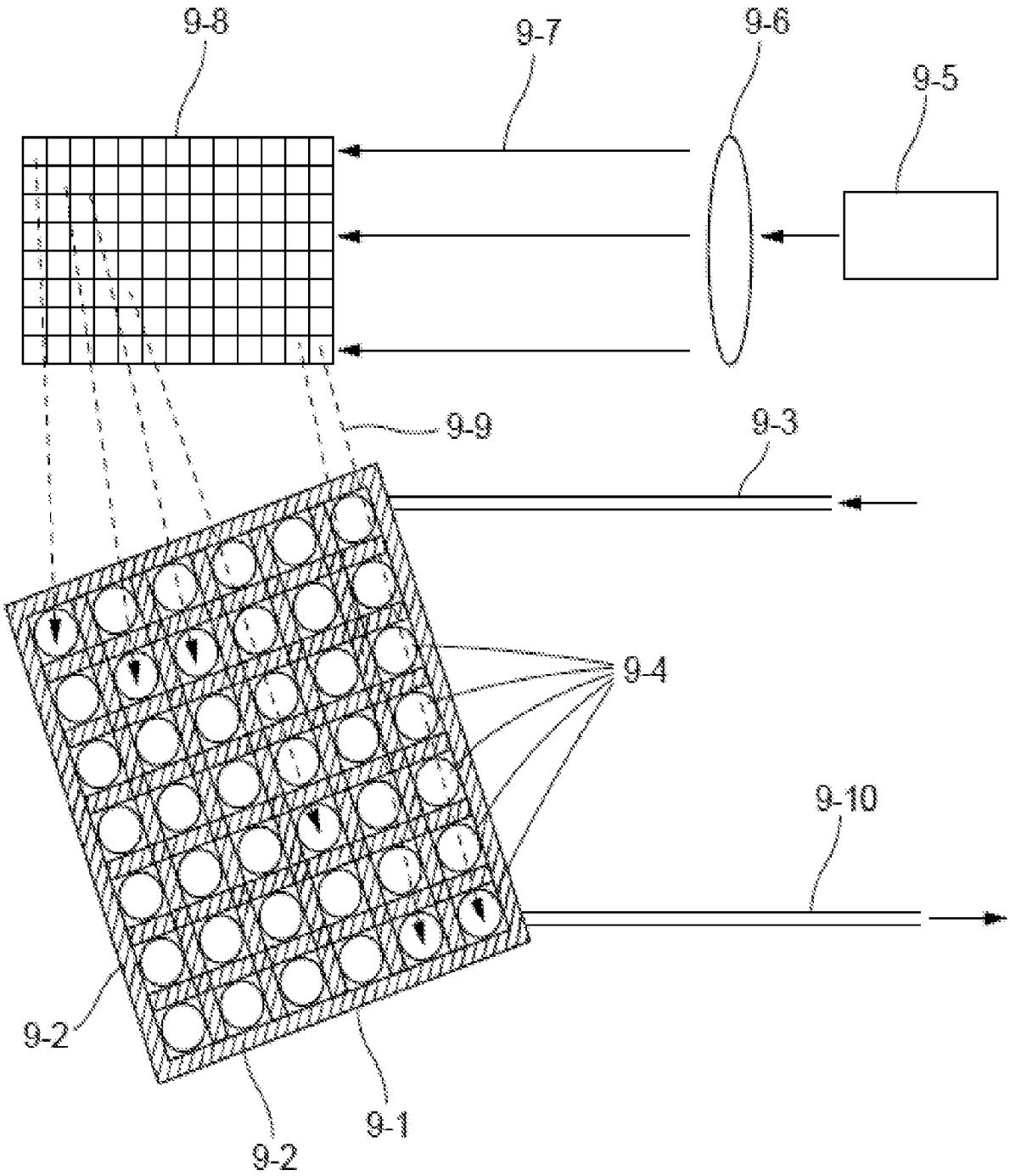


FIG. 9

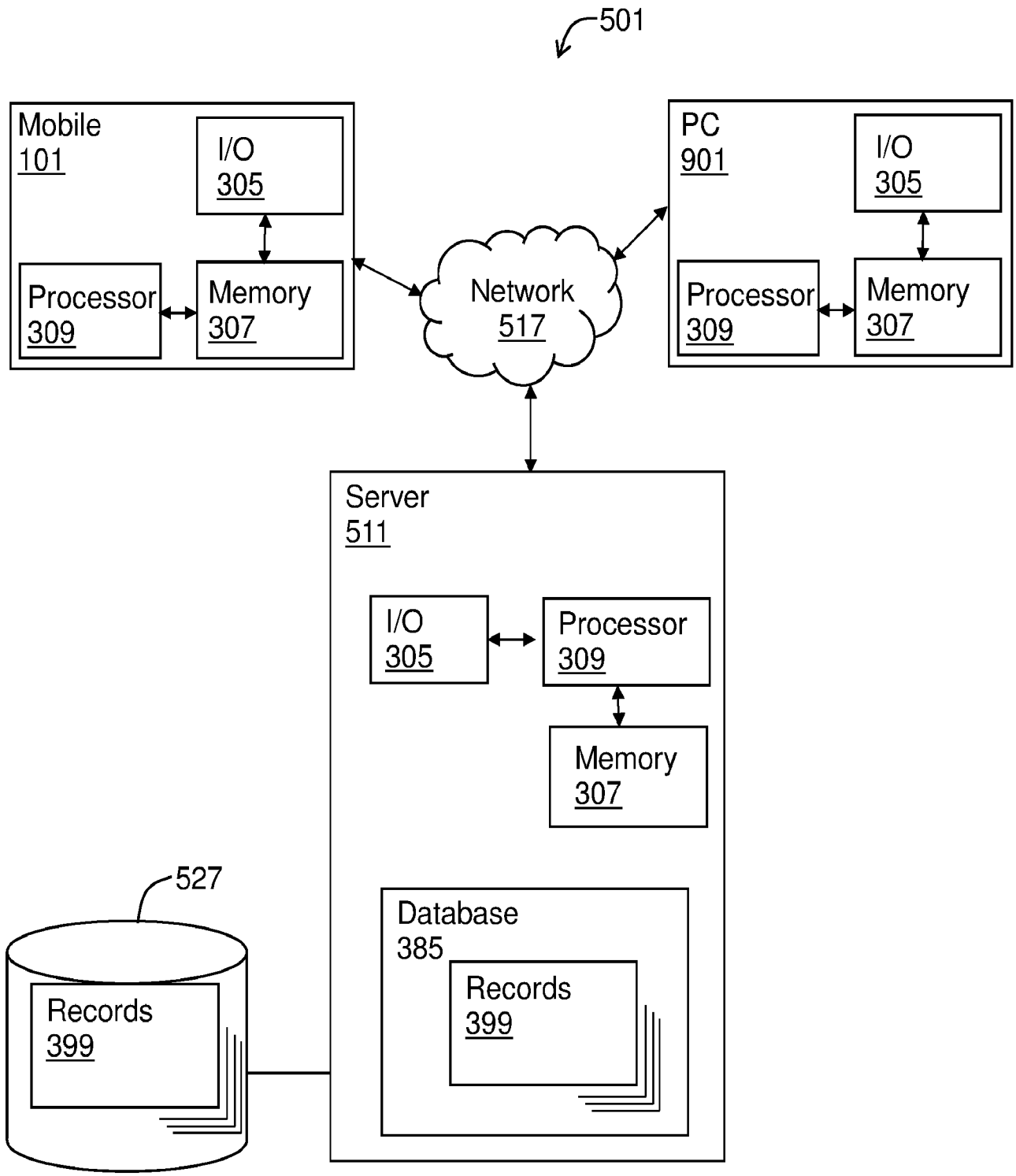

501
Mobile
101
I/O
305
Processor
309
Memory
307
Network
517
PC
901
I/O
305
Processor
309
Memory
307
Server
511
I/O
305
Processor
309
Memory
307
527
Records
399
Database
385
Records
399

FIG. 10

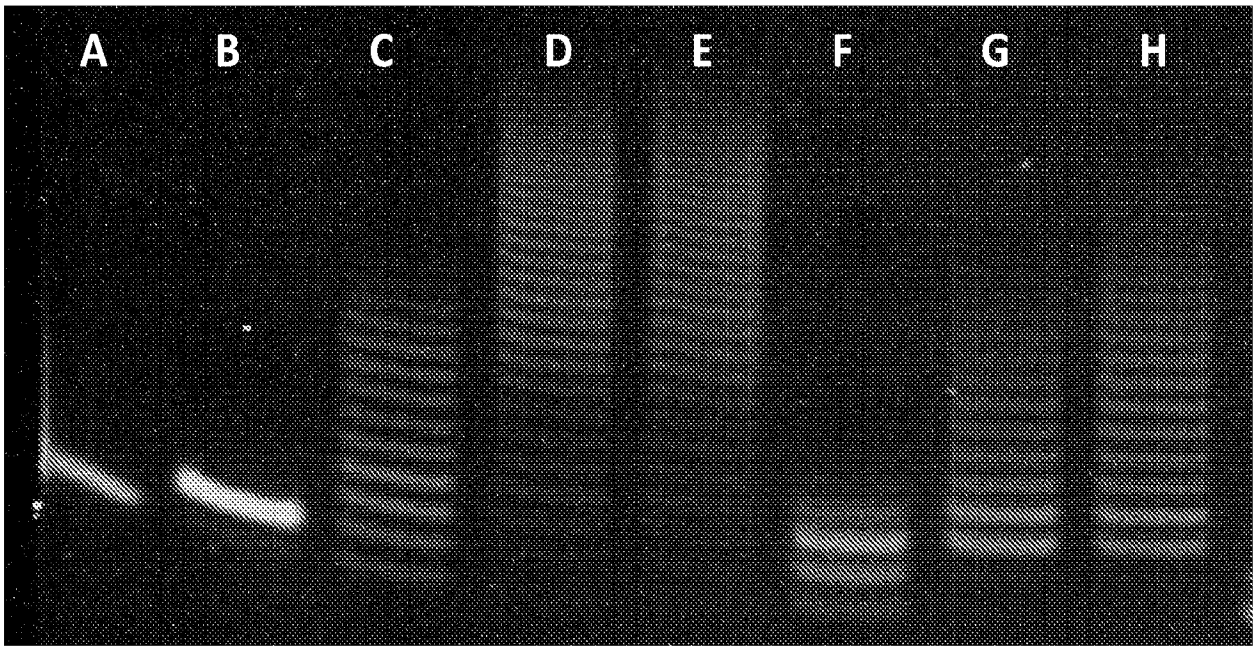


FIG. 11

Message = "MA"
Assign each character a number: A→1;...Z→26
i.e., M→13; A→1
Convert each number to 6 bit binary:
13 = 001101
1 = 000001
Convert binary to nucleotides:
odd positions are encoded "0" = A, "1"=T
even positions are encoded "0"=C, "1"=G
Synthesize homopolymer nucleic acid sequence
Message = "MA"
"MA" = "13 1"
"MA" = 001101 000001
"MA" = ACTGAGACACAG
$A_nC_nT_nG_nA_nG_nA_nC_nA_nC_nA_nG_n$

FIG. 12

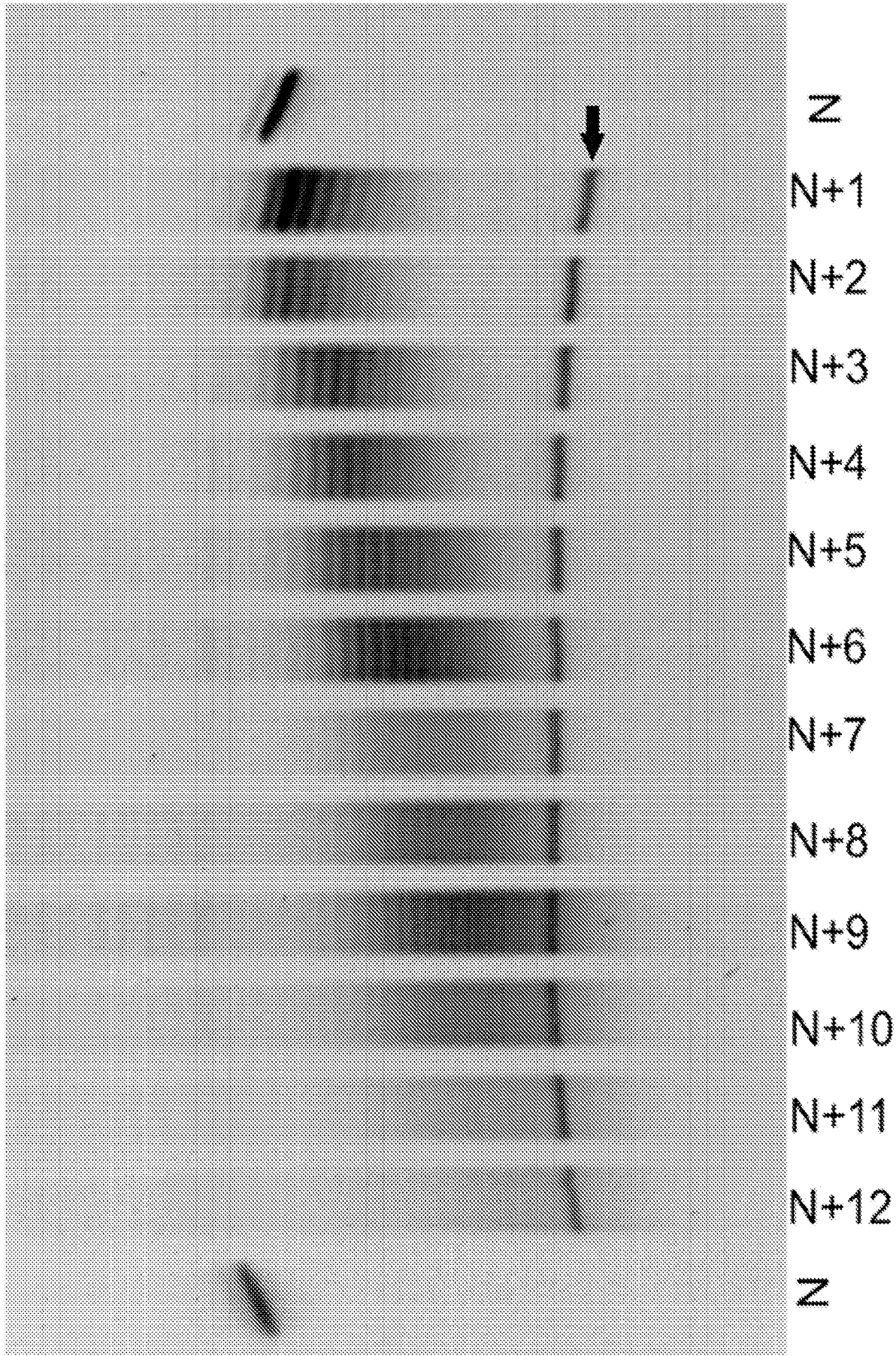


FIG. 13

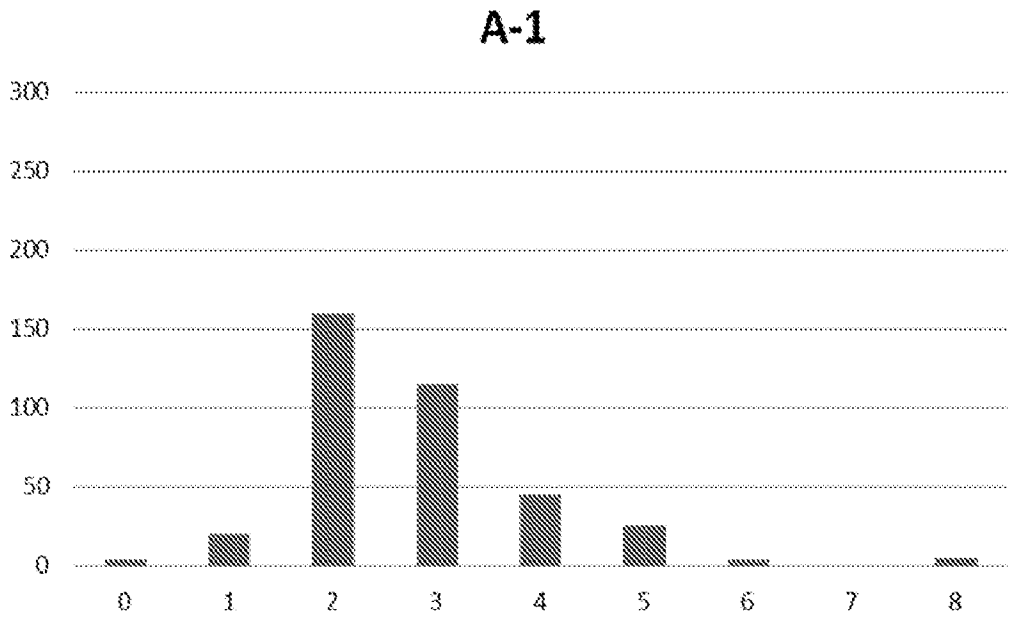


FIG. 14A

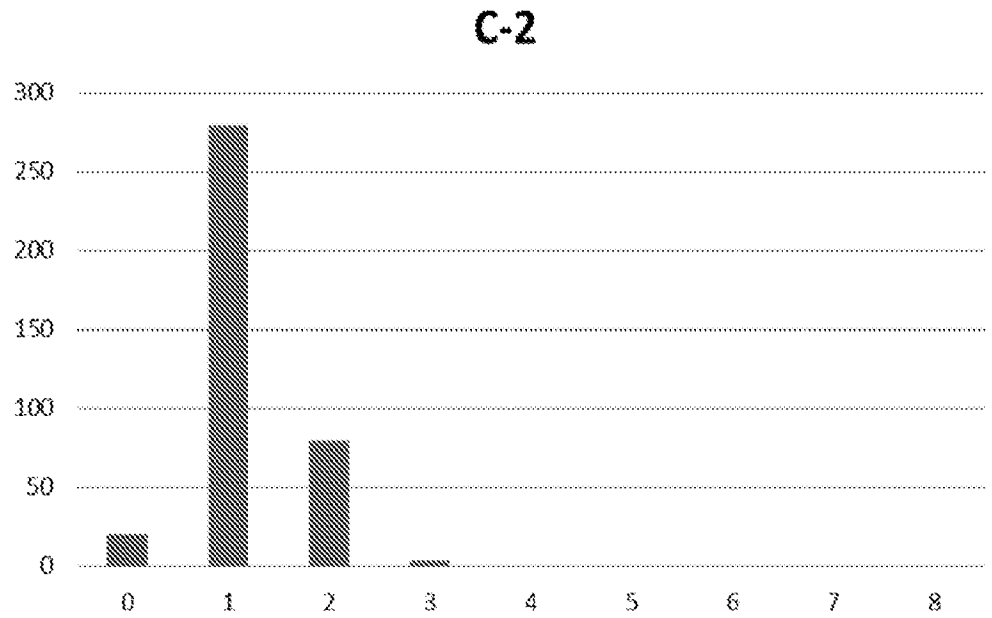


FIG. 14B

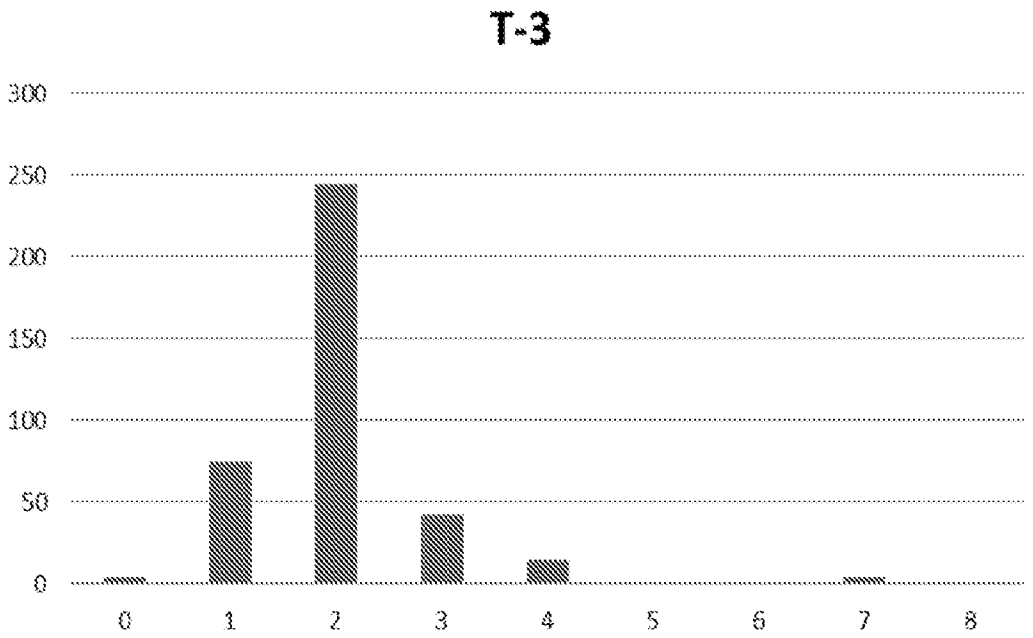


FIG. 14C

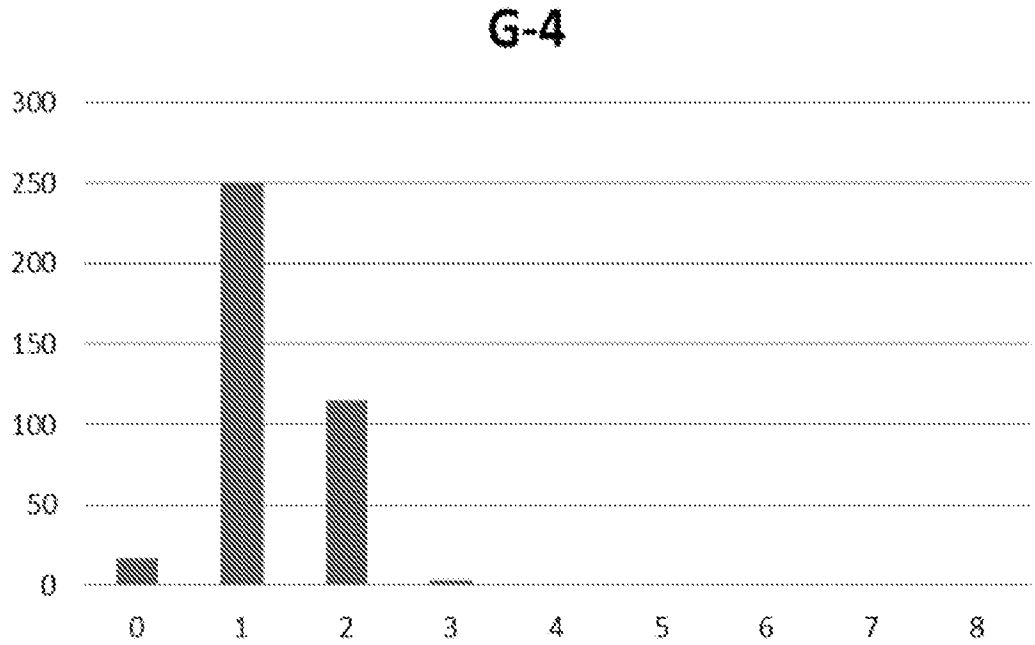


FIG. 14D

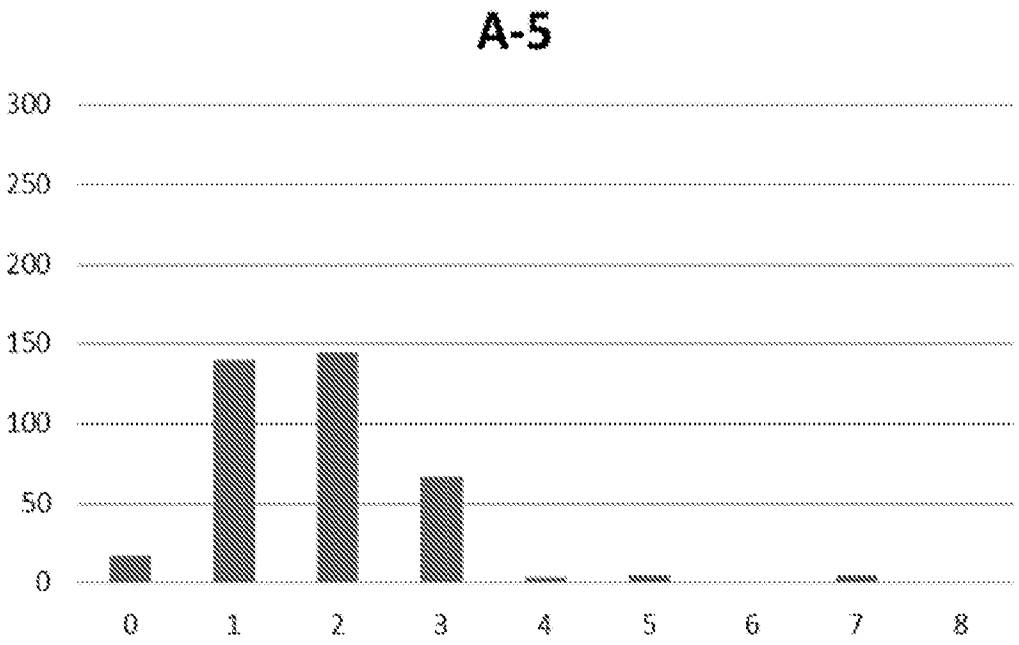


FIG. 14E

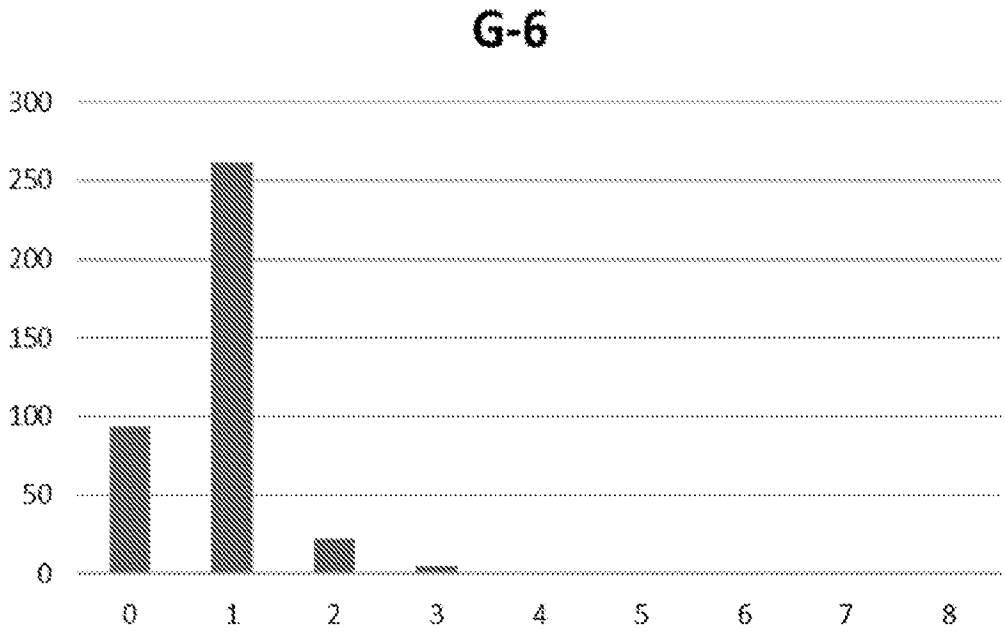


FIG. 14F

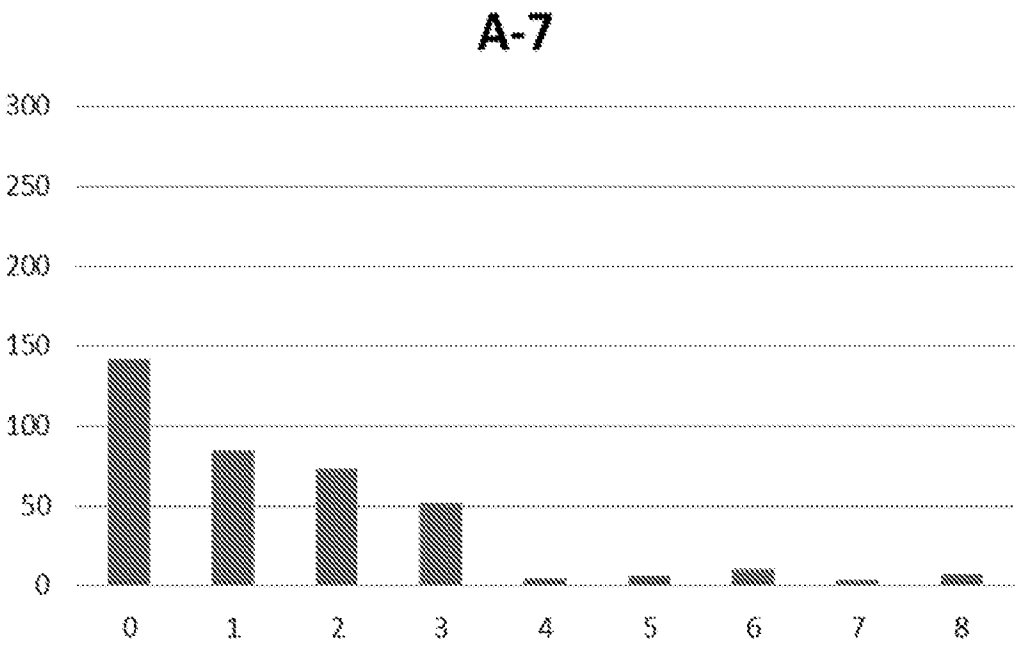


FIG. 14G

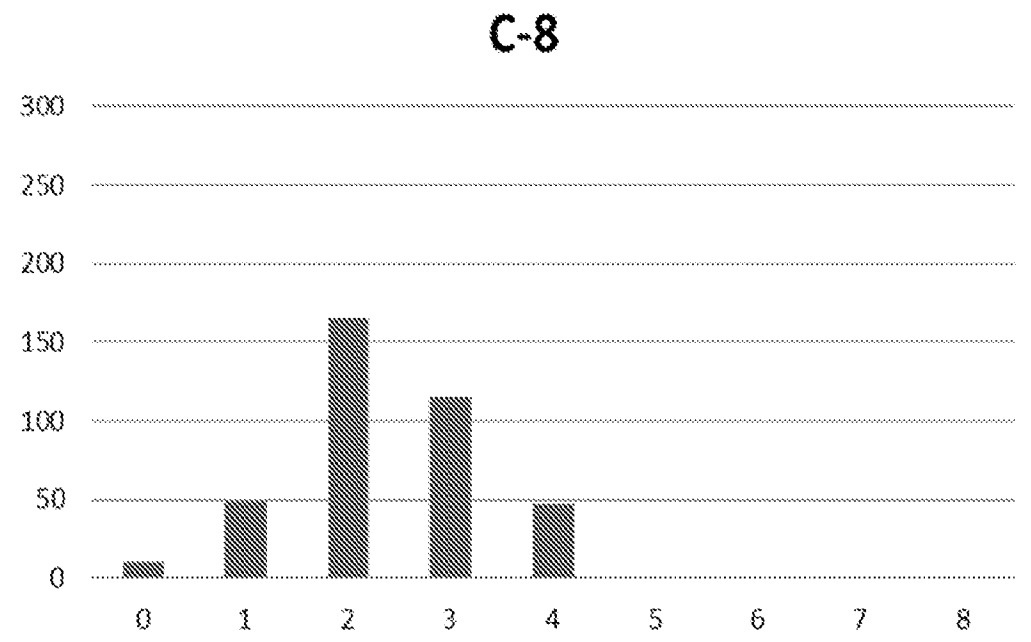


FIG. 14H

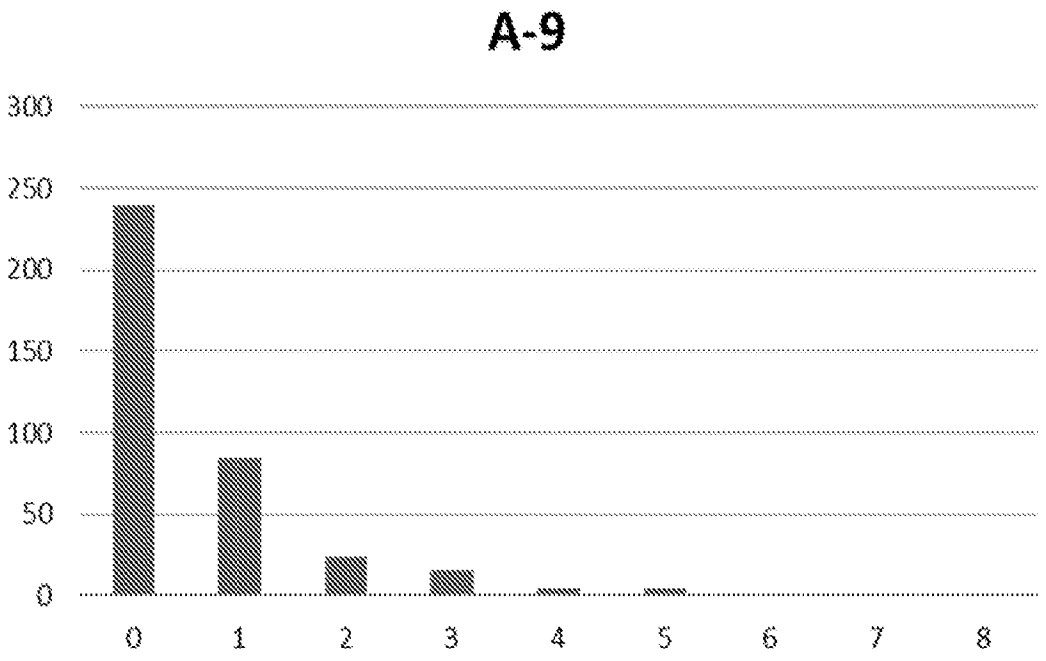


FIG. 14I

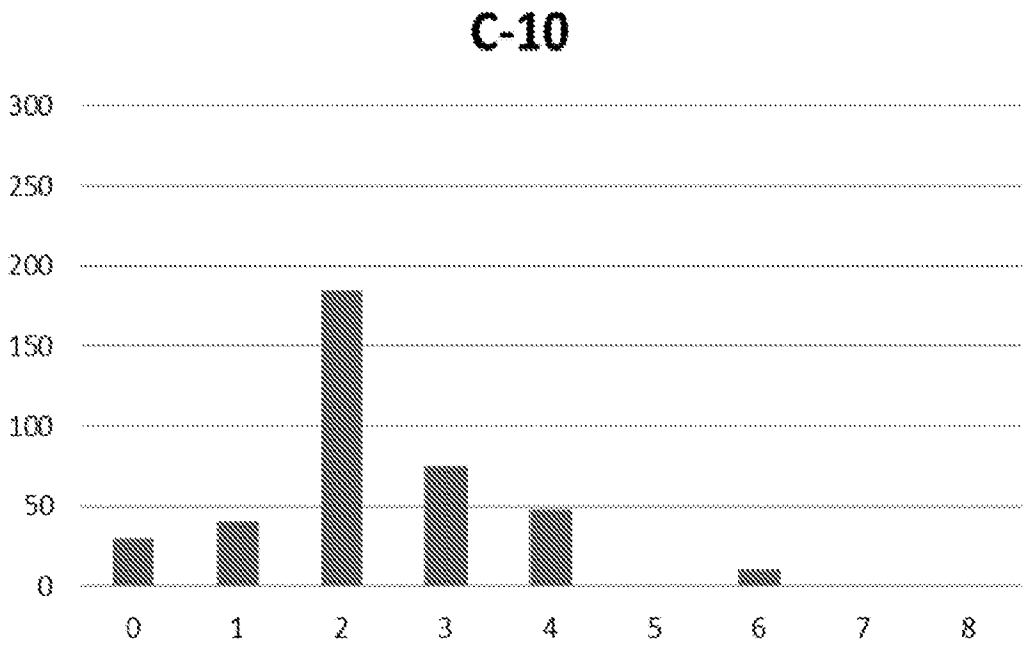


FIG. 14J

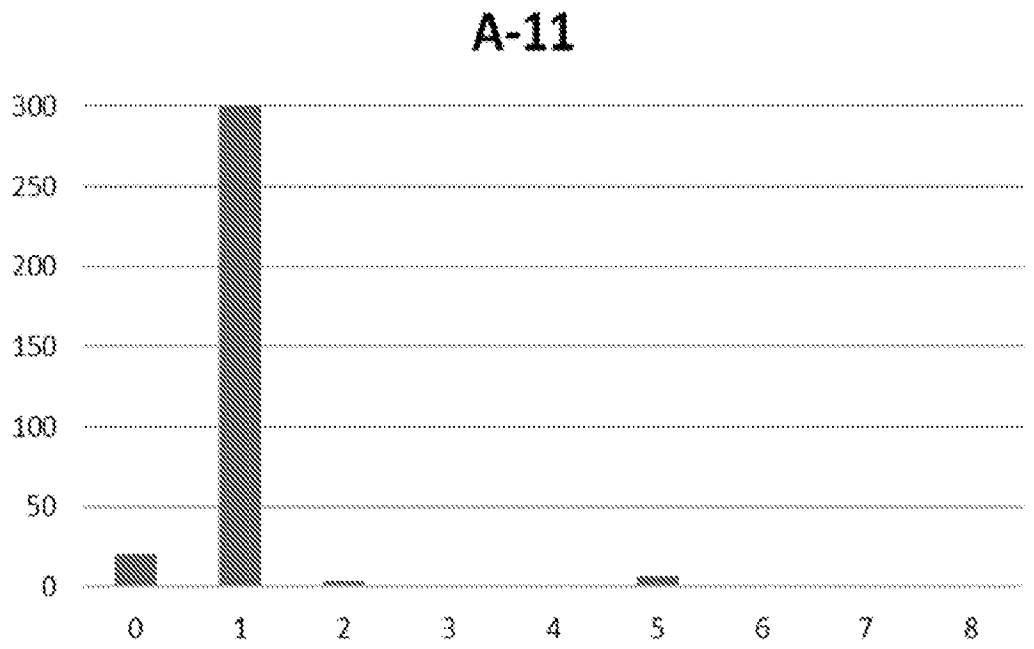


FIG. 14K

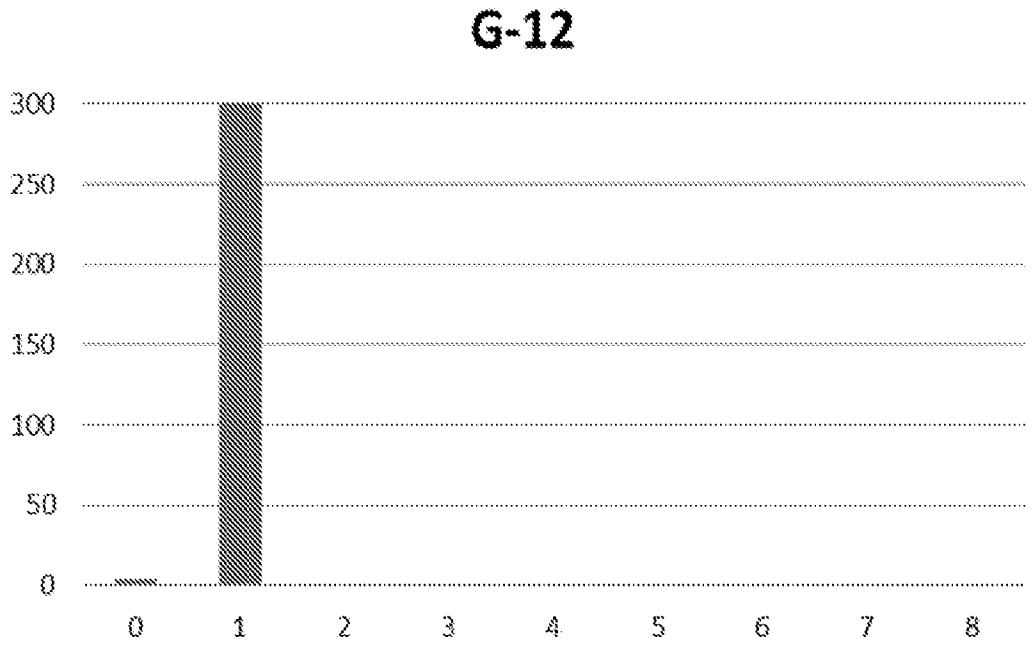


FIG. 14L

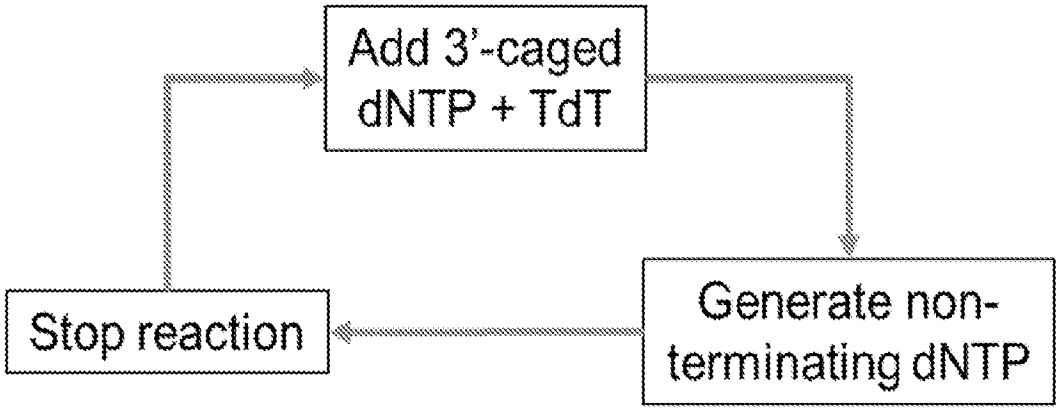


FIG. 15

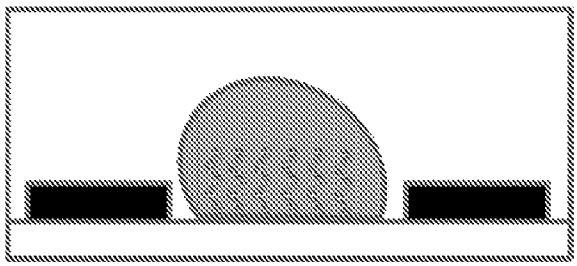


FIG. 16

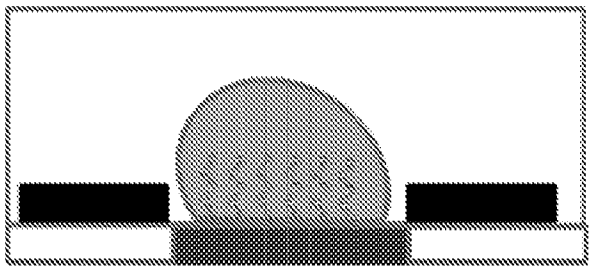


FIG. 17

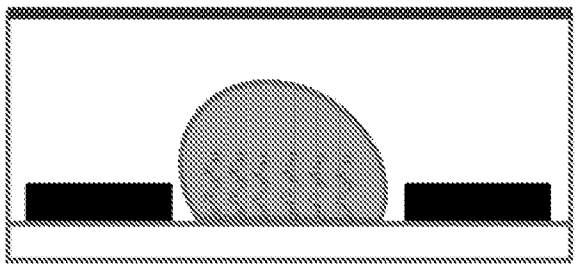


FIG. 18

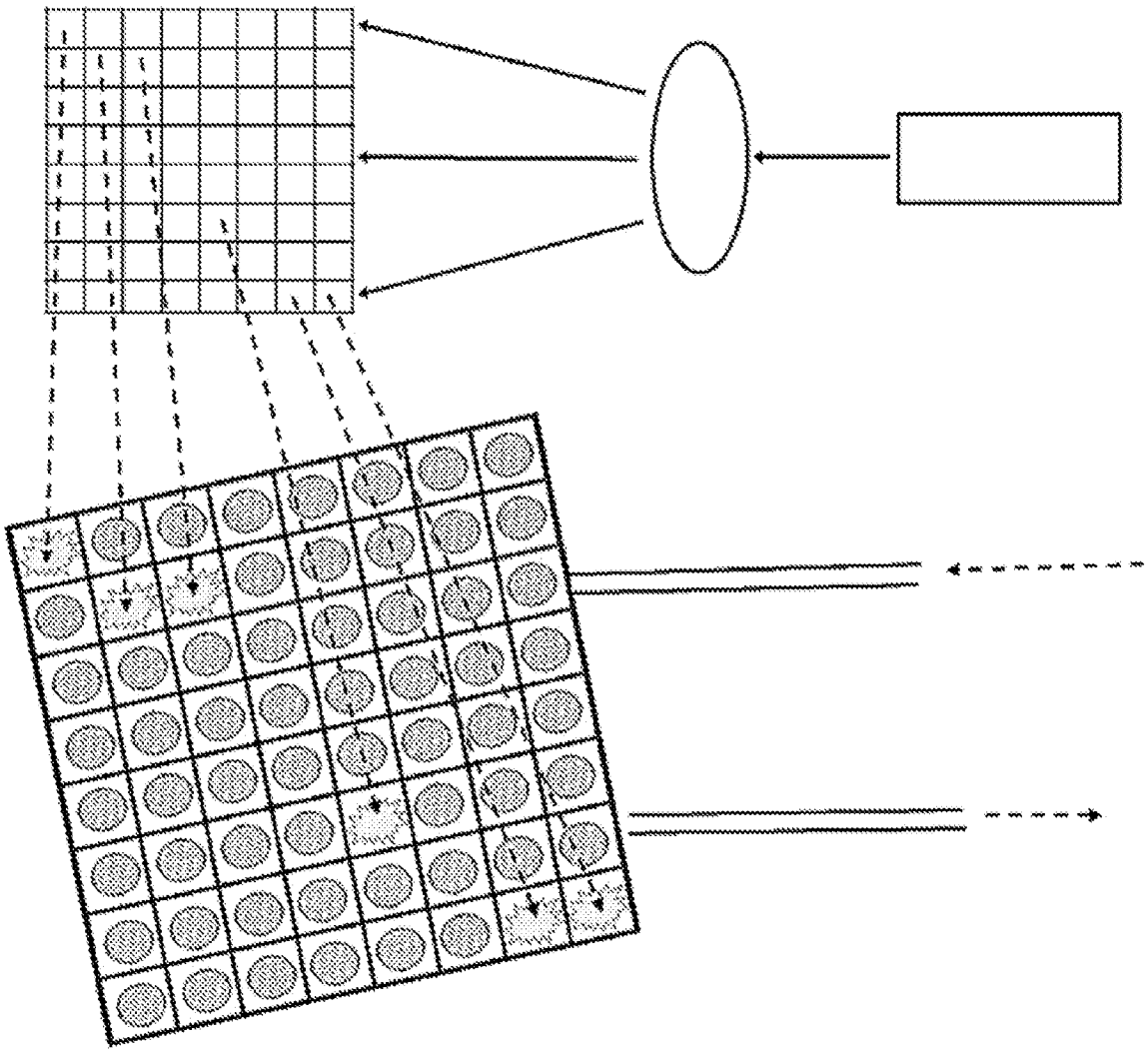


FIG. 19

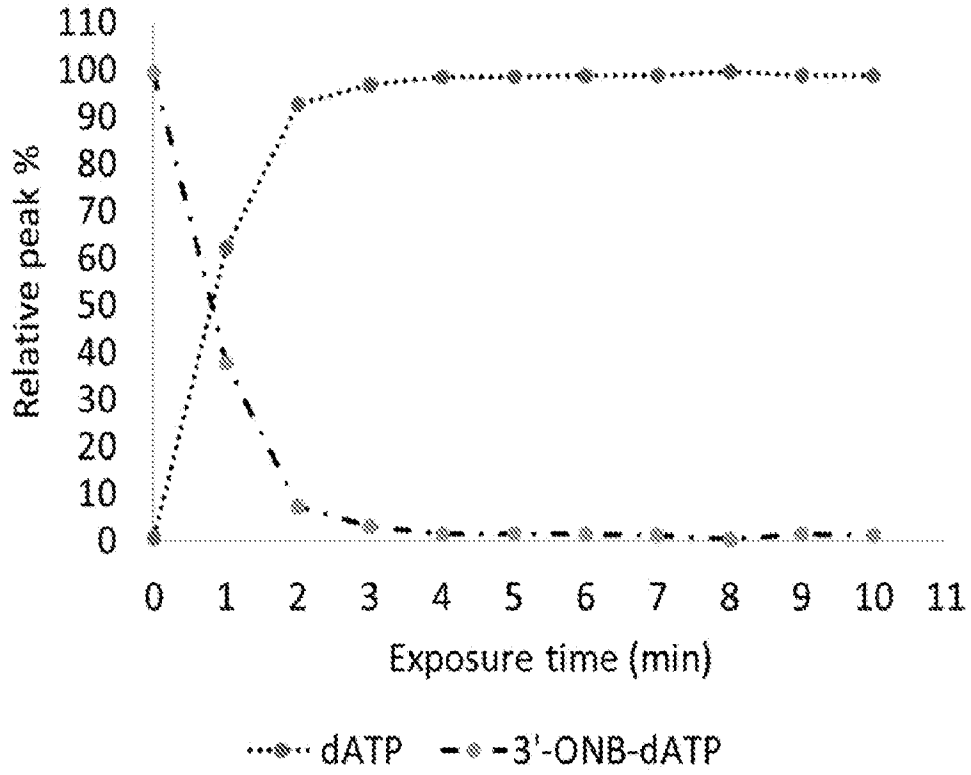
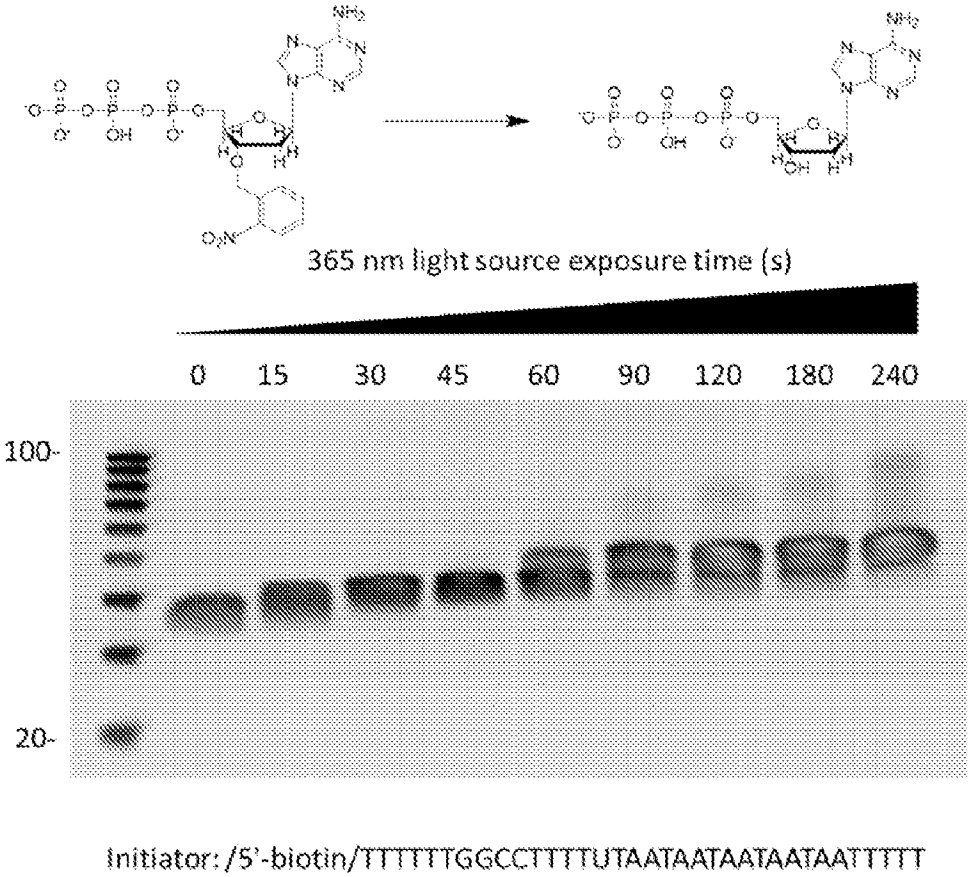
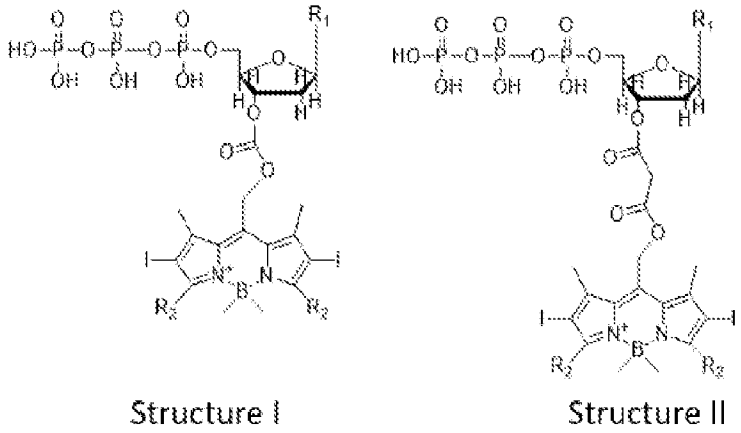


FIG. 20



Initiator: /5'-biotin/TTTTTTGGCCTTTTTUTAATAATAATAATAATTTTT

FIG. 21



Structure I Structure II

FIG. 22

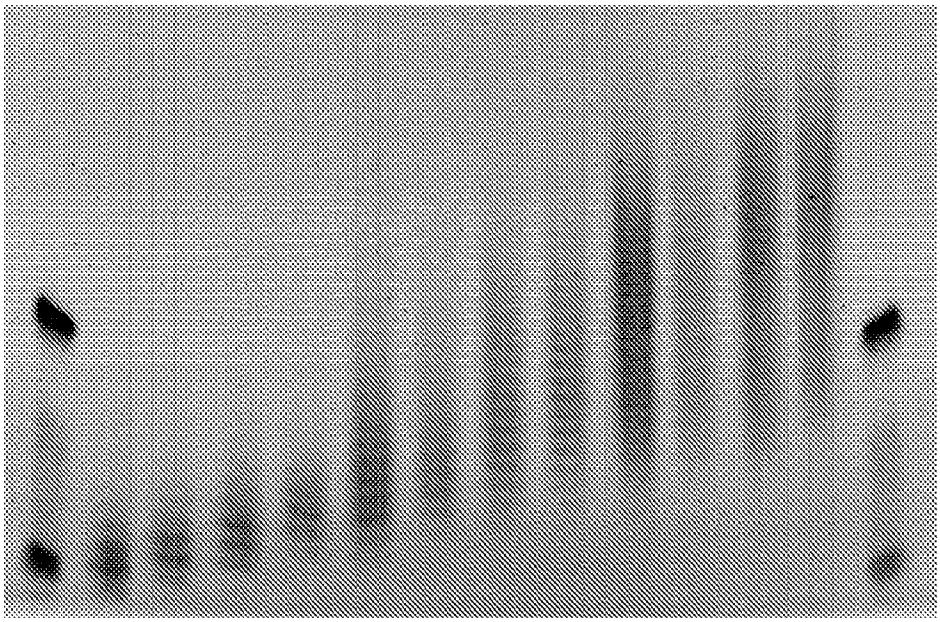


FIG. 23

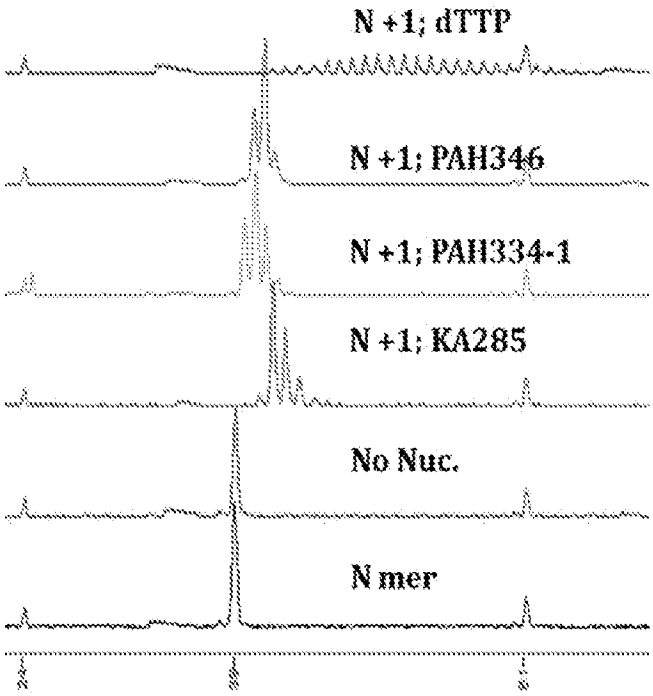
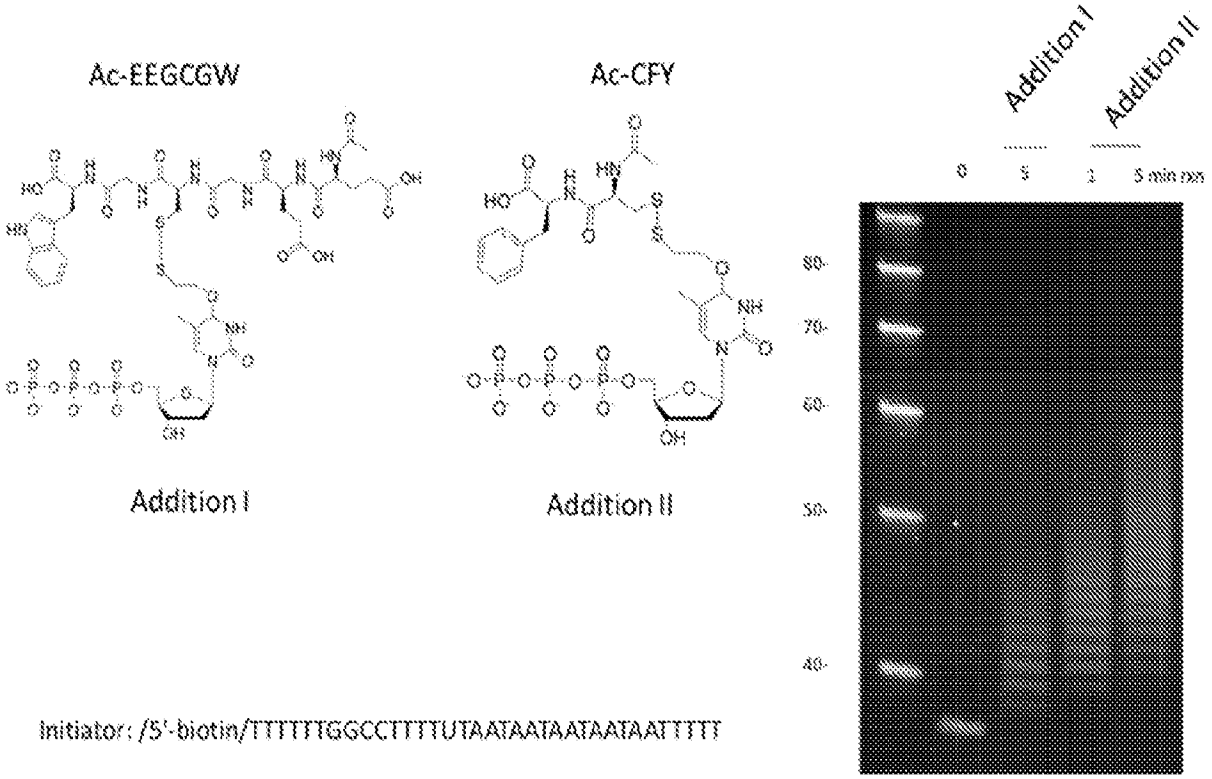


FIG. 24


Ac-EEGCGW
Ac-CFY
Addition I
Addition II
Addition I
Addition II
0
5
1
5 min
80
70
60
50
40
Initiator: /5'-biotin/TTTTTTGGCCTTTTUTAATAATAATAATAATTTTT

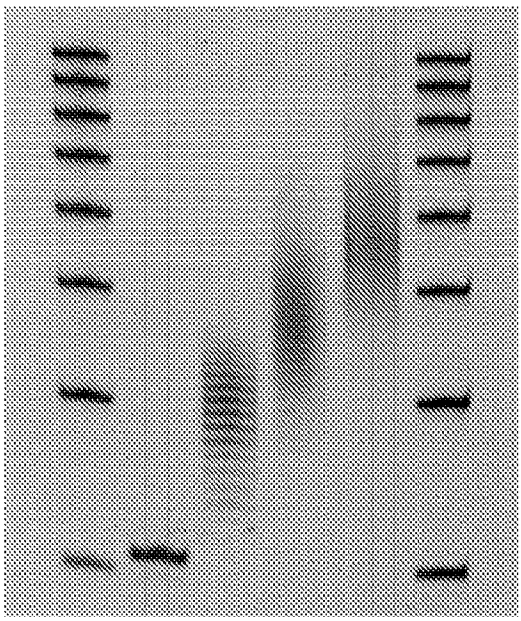


FIG. 25

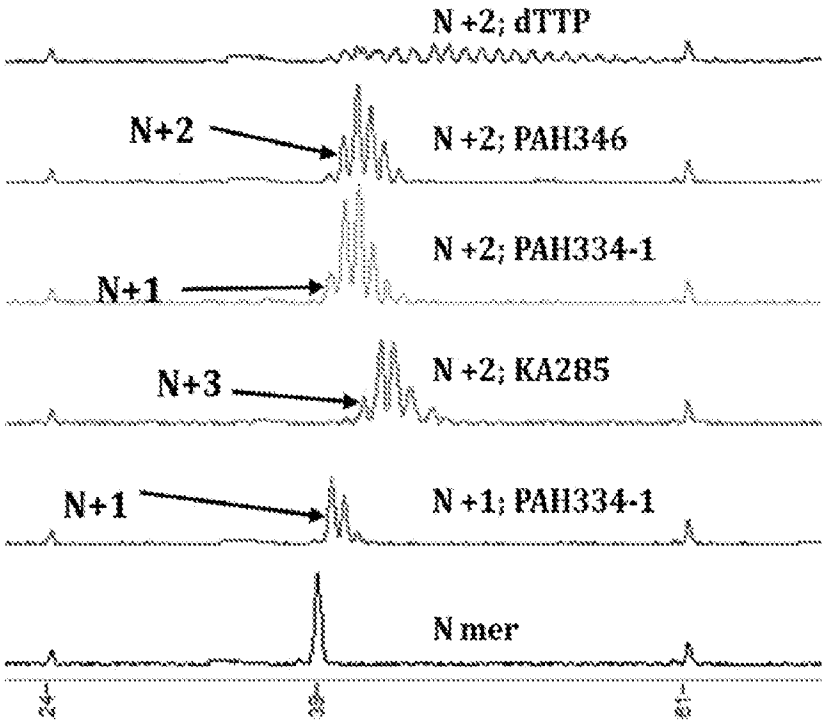

N +2; dTTP
N+2
N +2; PAH346
N +2; PAH334-1
N+1
N+3
N +2; KA285
N+1
N +1; PAH334-1
N mer

FIG. 26

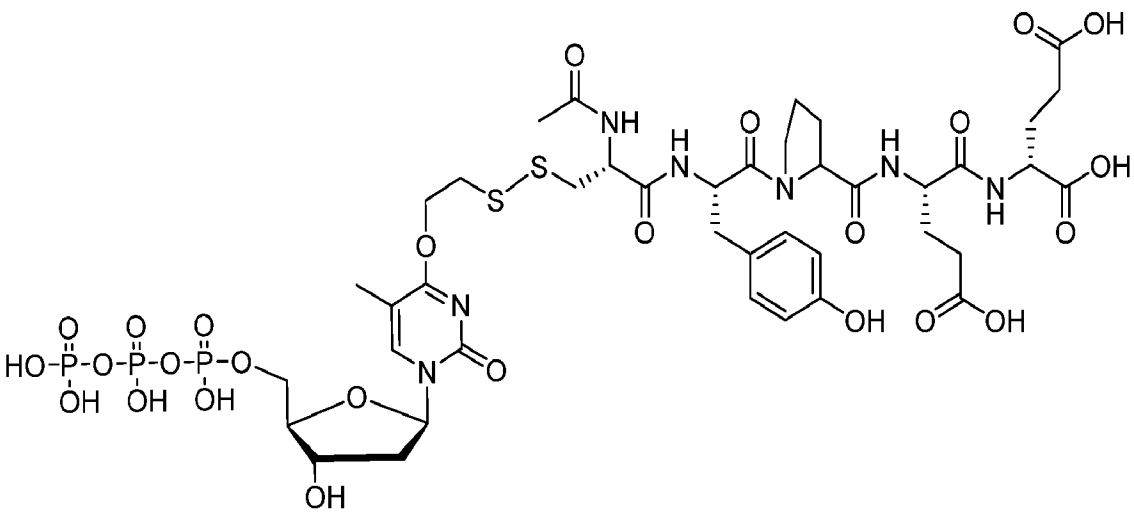


FIG. 27

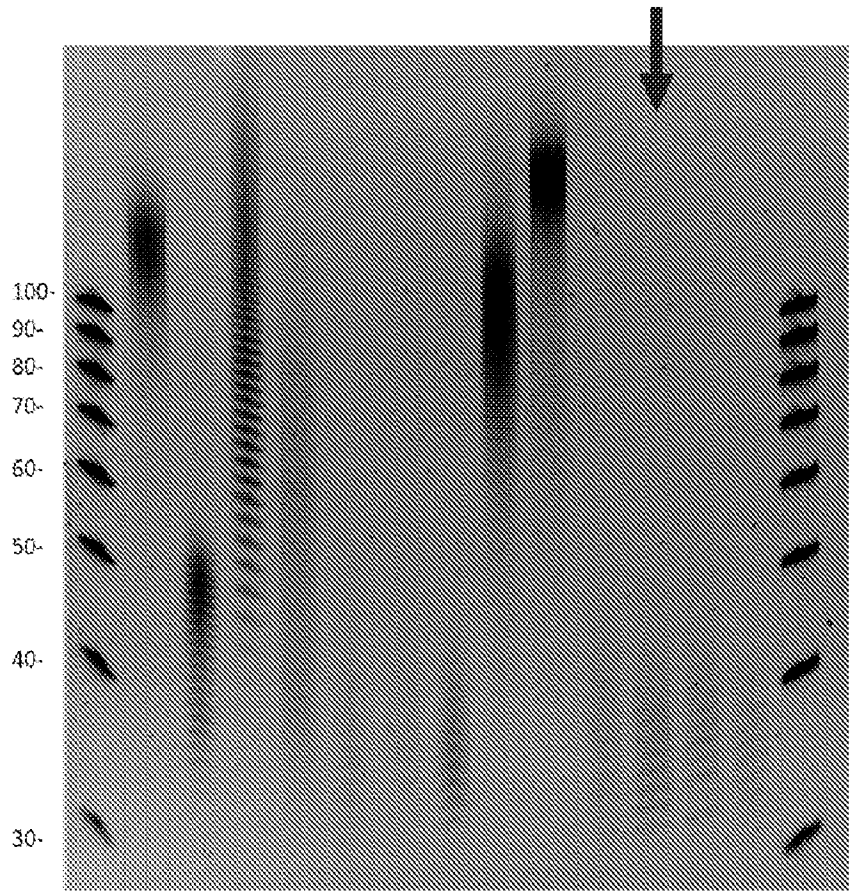


FIG. 28

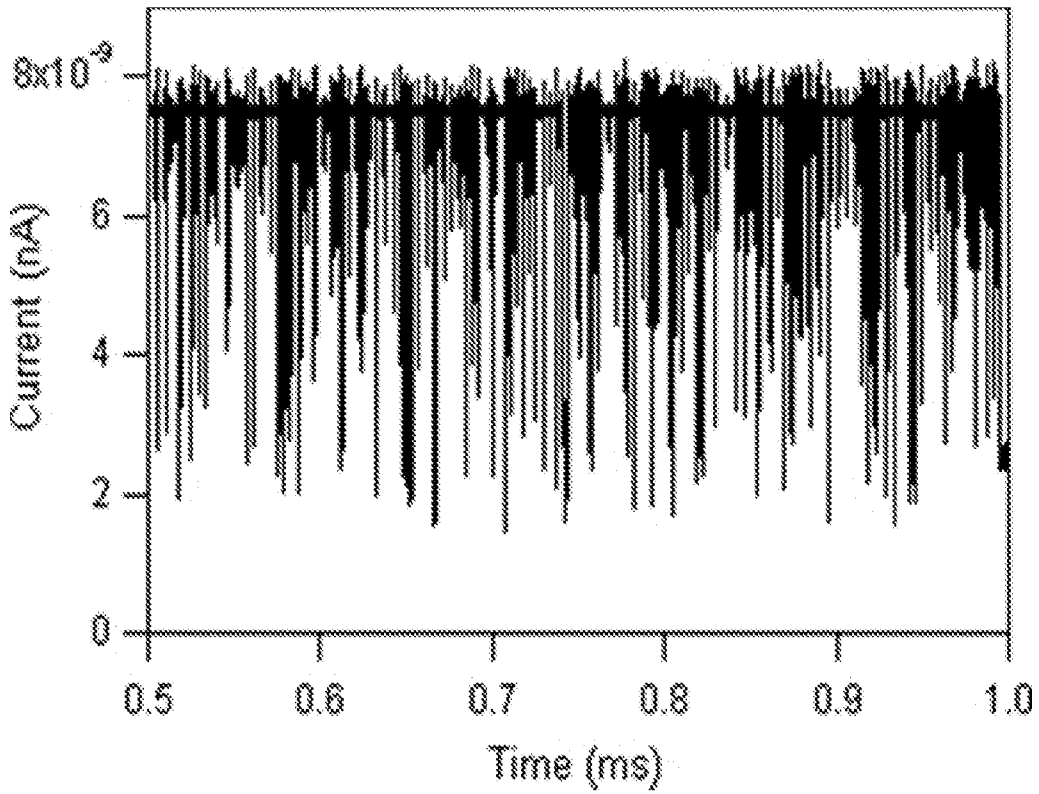

8x10^-9
6
4
2
0
Current (nA)
0.5
0.6
0.7
0.8
0.9
1.0
Time (ms)

FIG. 29

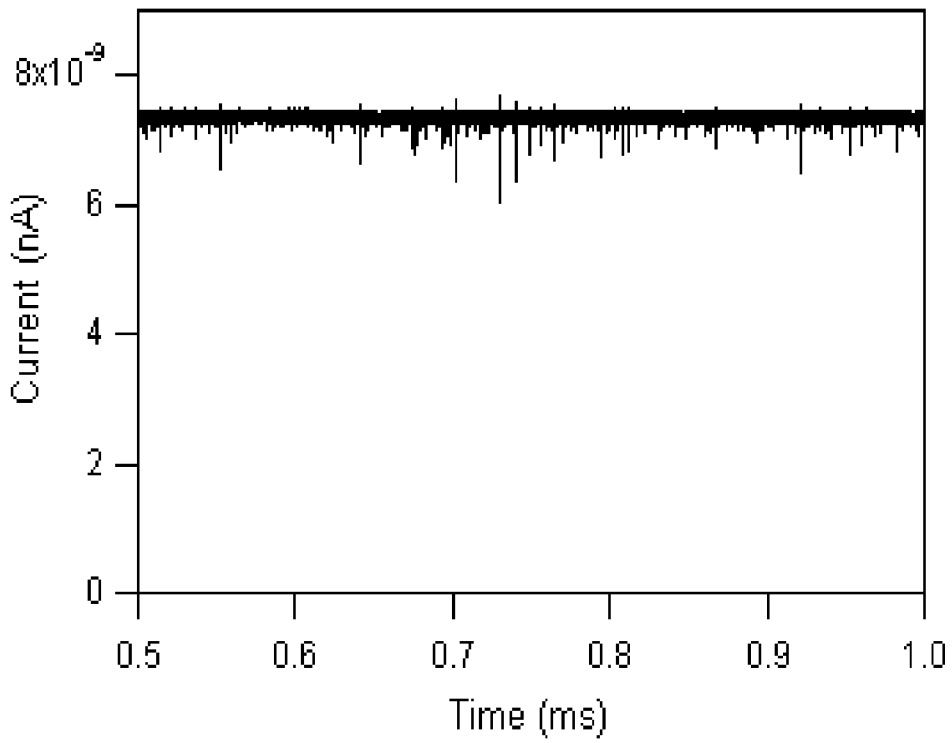

8x10^-9
6
4
2
0
Current (nA)
0.5
0.6
0.7
0.8
0.9
1.0
Time (ms)

FIG. 30

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/29780

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12P 19/34, C12M 1/00, C12N 15/09, C12N 15/10 (2020.01)

CPC - C12N 15/10, C12N 15/1093

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 2019/0040459 A1 (MOLECULAR ASSEMBLIES, INC.) 07 February 2019 (07.02.2019). Especially Abstract, para [0010]-[0014], [0034]-[0035], [0039]-[0040], [0053], [0056], [0061]	1-3, 5, 8-9, 13, 15-26 ---------- 4, 6-7, 10-12, 14
Y	HUTTER et al. Labeled Nucleoside Triphosphates with Reversibly Terminating Aminoalkoxyl Groups, Nucleosides Nucleotides Nucleic Acids., 10 December 2013, Vol 29, No 11, Pages 879 -895. Especially pg 880 para 1-3, pg 881 para 3, pg 883 para 8, pg 884 para 1, pg 885 para 6, 889 para 2	4, 6-7, 10-12, 14
Y	US 2010/0092957 A1 (ZHAO et al.) 15 April 2010 (15.04.2010). Especially para [0045], [0048], [0071]	6-7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 July 2020	10 AUG 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 20/29780

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☒ furnished subsequent to the international filing date for the purposes of international search only:

      ☒ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments: