



## (51) International Patent Classification:

C12Q 1/6869 (2018.01) G11C 13/00 (2006.01)

C07H 21/04 (2006.01)

## (21) International Application Number:

PCT/US2022/038591

## (22) International Filing Date:

27 July 2022 (27.07.2022)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

63/226,720 28 July 2021 (28.07.2021) US

63/269,324 14 March 2022 (14.03.2022) US

(71) Applicant: NAI0, INC. [US/US]; 743 Crestview Drive, San Carlos, California 94070 (US).

(72) Inventor: KOOL, Eric; 743 Crestview Drive, San Carlos, California 94070 (US).

(74) Agent: LI, Wei; Wilson Sonsini Goodrich &amp; Rosati, 650 Page Mill Road, Palo Alto, California 94304-1050 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, 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.

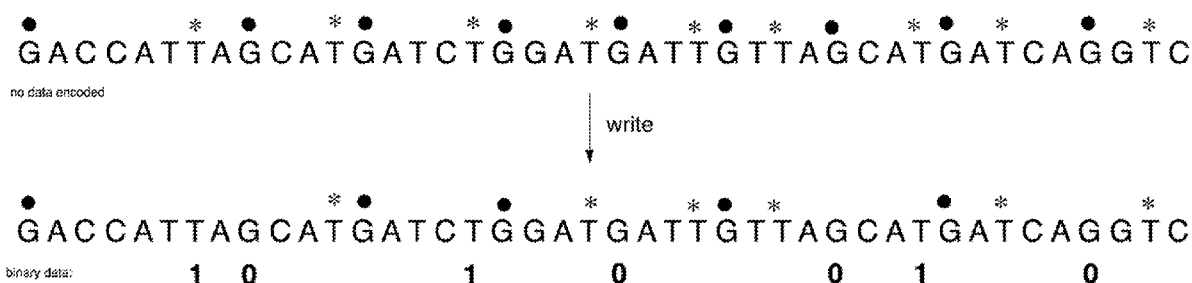
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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))

(54) Title: COMPOSITIONS, SYSTEMS, AND METHODS FOR NUCLEIC ACID DATA STORAGE

Fig. 15



(57) Abstract: Provided herein are writable polymers (e.g., writable nucleic acid polymers) for data storage and related methods. Generally, a writable polymer (e.g., a nucleic acid polymer) contains one or more convertible residues (e.g., convertible nucleobases) that are capable of being converted from a first state to a second state, the first and the second state being different. Various methods can be utilized to generate a writable nucleic polymer, such as polymerase extension via rolling circle reaction or chemical synthesis and ligation. Also provided herein are various methods for writing or encoding a writable nucleic acid polymer by selectively converting nucleobases into the second state. Also provided herein are various methods for reading or decoding a nucleic acid polymer encoded with data.

## **COMPOSITIONS, SYSTEMS, AND METHODS FOR NUCLEIC ACID DATA STORAGE**

### **CROSS REFERENCE TO RELATED APPLICATIONS**

**[0001]** This application claims the benefit of U.S. Provisional Patent Application No. 63/226,720, filed July 28, 2021, and U.S. Provisional Patent Application No. 63/269,324, filed on March 14, 2022, the content of each of which is incorporated by reference herein in its entirety.

### **TECHNICAL FIELD**

**[0002]** The disclosure is generally directed to compositions, systems, and methods for storing data in nucleic acid molecules.

### **BACKGROUND**

**[0003]** As the amount of digital data increases, the complications of storing digital data long term is becoming a rapidly growing issue. Electronically or magnetically archived digital data can easily be manipulated, distorted, and/or lost while in storage. While efficient solid-state electronic methods for archival data storage exist, they are not stable over a period of years, resulting in loss of data unless the data is periodically rewritten or transferred to a new device. Similarly, magnetic tape is commonly used for data archiving, but it also degrades over time. Therefore, ways to efficiently encode and store data, especially over long periods, are being pursued very actively.

**[0004]** Nucleic acid molecules (especially DNA) offer a potential solution for overcoming issues with data storage. With its sequences of repeated bases, nucleic acid polymers are essentially biochemical molecules of digital information, which can be stably stored at high densities for extremely long durations in time. Natural DNA contains digital information encoded in the four bases: A, C, T, and G, and can be used to encode binary data in its sequence in synthesized strands. A single polymer of DNA can be very long (such as in chromosomes) and encodes millions of bits of data. It has been estimated that 1 cubic inch of DNA can encode  $10^{18}$  bytes of data. Furthermore, DNA is relatively stable, and has yielded sequence information even from samples tens or hundreds of thousands of years old. Thus, DNA offers considerable promise for archiving data.

**[0005]** Further, to facilitate the access to stored data in nucleic acid molecules, the stored data can be read rapidly and cheaply via high-throughput sequencing techniques. Advances in sequencing technology have greatly lowered the cost and increased the speed of sequencing,

allowing data in DNA to be read efficiently. Newer long-read single molecule technologies enable rapid reading of bases in single DNA molecules tens of thousands of bases in length. Newer nanopore technologies enable the reading of sequence from single molecules of DNA in seconds to minutes (*see* N Kono and K. Arakawa, *Dev Growth Differ.* 2019; 61:316-326; and Q Chen and Z. Liu, *Sensors (Basel).* 2019; 19:1886; the disclosures of which are each incorporated herein by reference), and can read sequences of strands tens of thousands or base pairs in length or more.

**[0006]** Although nucleic acids are a great potential source of data storage, the process of synthesizing of nucleic acids in particular data-defining sequences is inefficient and thus the process of encoding the nucleic acids is a substantial barrier to utilizing nucleic acids as data storage. Current approaches for storing data in DNA involve chemical or enzymatic synthesis of strands of arbitrary sequences that encode digital information (*see* G. M. Church, Y. Gao, and S. Kosuri *Science.* 2012; 337:1628; X. Chengtao, et al., *Nucleic Acids Res.* 2021; 49:5451-5469; and E. Yoo, et al., *Comput Struct Biotechnol J.* 2021; 19:2468-2476; the disclosures of which are each incorporated herein by reference). Oligonucleotide synthesizers can produce DNAs of length up to roughly 100-200 nucleotides. Specialized synthesizers can produce hundreds or thousands of oligonucleotides at one time, which promises higher throughput of data writing. In addition to chemical DNA synthesis, enzymatic approaches involving polymerases or other enzymes are also under investigation for creating DNAs of arbitrary data-encoding sequence. These involve adding specialized nucleotides one at a time, or short segments of DNA, step by step.

**[0007]** The approach of encoding data in DNA during synthesis is limited by yield, strand length, time, and cost. Current efficient DNA synthesizers produce strands up to roughly 200 nucleotides, and thus encode relatively small amounts of information. Large numbers of different oligonucleotides must be synthesized to compensate for the short sequences. Oligonucleotide synthesis requires excess reagents to achieve high stepwise yields, and requires expensive consumption of reagents and solvents. It also requires time to achieve these high yields for each nucleotide addition (commonly 1-5 min for each step), which implies the need for extended time for encoding larger amounts of data. Common enzymatic approaches under development similarly add nucleotides or groups of nucleotides in stepwise fashion, and have not yet greatly improved on the ability to produce very long strands and encode large amounts of data. Because the enzymatic synthesis approaches also occur stepwise, they also have limits in the speed of data encoding. Further, since both the above chemical and enzymatic strategies typically produce relatively short strands, they may not be

ideal for single molecule sequencing, and instead may rely on sequencing methods that require larger amounts of each written DNA.

### **SUMMARY OF THE DISCLOSURE**

**[0008]** In one aspect, provided herein are polymers for encoding data, comprising:

a plurality of convertible residues iteratively spaced along and covalently linked to the backbone of the polymer,

wherein each of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme;

wherein the plurality of convertible residues are covalently linked to the polymer in the first state and in the second state.

**[0009]** In certain embodiments, the polymer is a nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.

**[00010]** In certain embodiments, the nucleic acid polymer is a single-stranded nucleic acid polymer.

**[00011]** In certain embodiments, the nucleic acid polymer is double-stranded nucleic acid polymer.

**[00012]** In certain embodiments, the nucleic acid polymer comprises Deoxyribonucleic acid (DNA), Ribonucleic acid (RNA), phosphorothioate DNA, glycerol nucleic acids (GNA), threose nucleic acids (TNA), locked nucleic acids (LNA), or a combination thereof.

**[00013]** In certain embodiments, the nucleic acid polymer comprises greater than 10 convertible residues.

**[00014]** In certain embodiments, the ratio of the total number of nucleotides to the convertible residues in the nucleic acid polymer is between 2 to 100.

**[00015]** In certain embodiments, the plurality of convertible nucleobases are non-naturally occurring nucleobases.

**[00016]** In certain embodiments, the plurality of convertible nucleobases are modified naturally occurring nucleobases or derivatives of naturally occurring nucleobases.

**[00017]** In certain embodiments, each of the plurality of convertible nucleobases comprises a chemically modifiable moiety.

**[00018]** In certain embodiments, each of the plurality of convertible nucleobases the chemically modifiable moiety is directly attached to the base of the convertible nucleobases.

**[00019]** In certain embodiments, each of the plurality of convertible nucleobases the chemically modifiable moiety is attached to the base without a linker or a sidechain.

**[00020]** In certain embodiments, the plurality of convertible nucleobases are covalently linked to the backbone of the nucleic acid via the sugar.

**[00021]** In certain embodiments, the chemically modifiable moiety is activatable by light, voltage, enzymatic agent, chemical reagent, or a redox agent, thereby converting from the first state into the second state.

**[00022]** In certain embodiments, the chemically modifiable moiety is activatable by light, thereby converting from the first state into the second state.

**[00023]** In certain embodiments, the conversion from the first state into the second state occurs via an irreversible reaction.

**[00024]** In certain embodiments, the convertible nucleobase becomes a naturally occurring nucleobase after conversion into the second state.

**[00025]** In certain embodiments, the convertible nucleobase becomes guanine, adenine, thymine, uracil or cytosine after conversion into the second state.

**[00026]** In certain embodiments, the backbone of the polymer (e.g., phosphate and sugar in nucleic acid polymer) remain unchanged during the conversion from the first state into the second state.

**[00027]** In certain embodiments, the polymer comprises two or more different sets of convertible residues, each set of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different.

**[00028]** In certain embodiments, each of the plurality of convertible residues comprises a chemically modifiable moiety that can be activated by light.

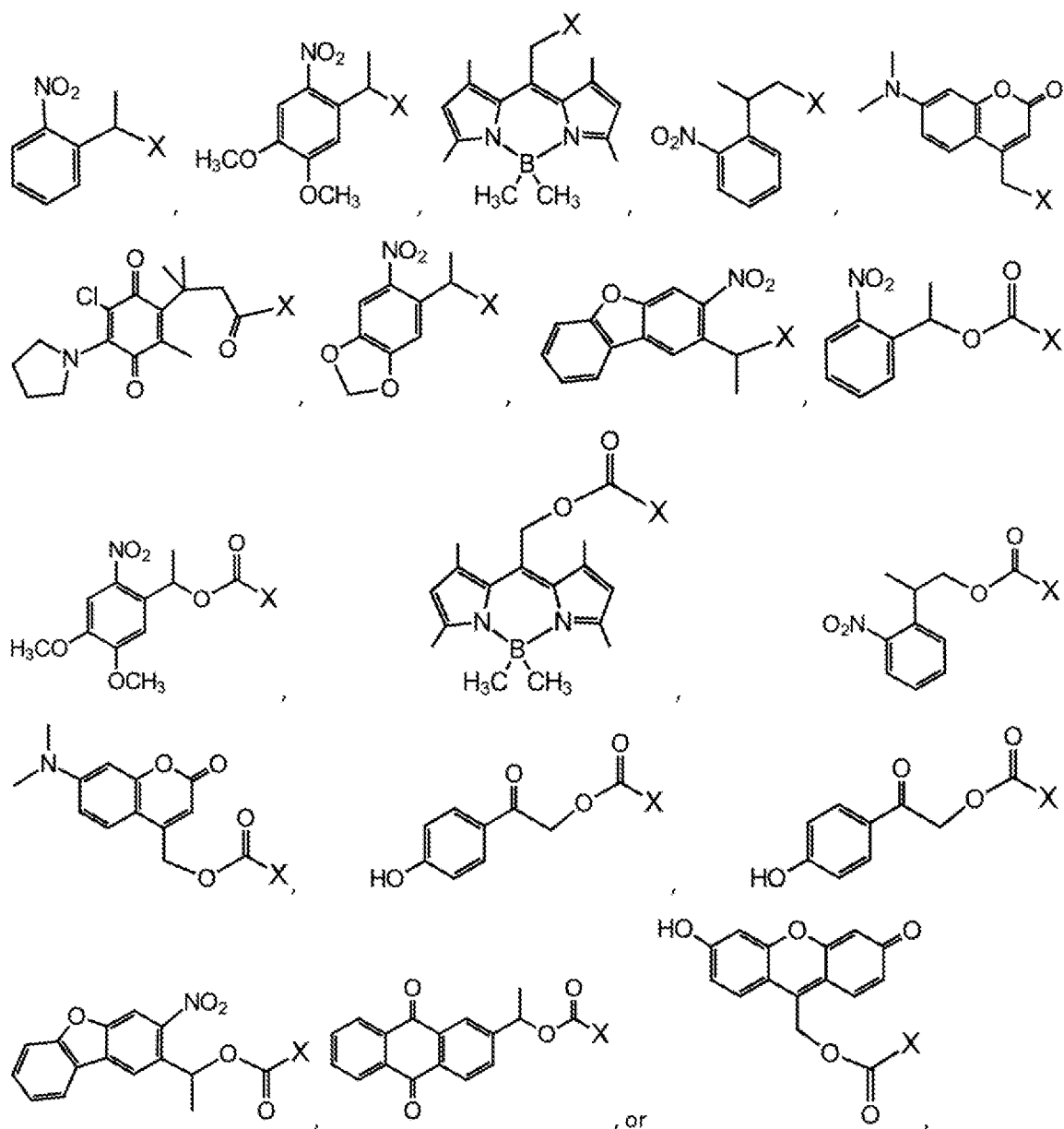
**[00029]** In certain embodiments, the two or more different sets of convertible residues are activatable by light of different wavelengths.

**[00030]** In certain embodiments, a first set of convertible residues is activatable by light of a first wavelength, and a second set of convertible residues is activatable by light of a second wavelength, the first wavelength and the second wavelength being different.

**[00031]** In certain embodiments, the chemically modifiable moiety comprises one or more photo-removable groups.

**[00032]** In certain embodiments, the chemically modifiable moiety is a leaving group.

**[00033]** In certain embodiments, the one or more photo-removable groups are:



wherein X represents NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase to which the photo-removable group is attached.

**[00034]** In certain embodiments, the plurality of convertible nucleobases are capable of being converted by light of a wavelength of 325 nm, 360 nm, or 400 nm.

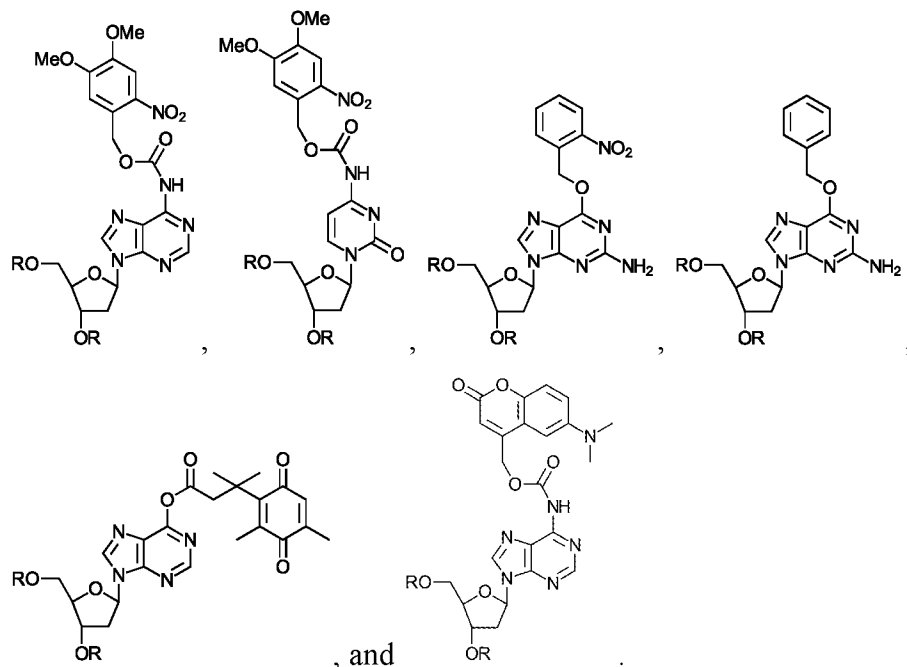
**[00035]** In certain embodiments, the plurality of convertible nucleobases are capable of being converted by light of a wavelength of between 400 nm to 850 nm.

**[00036]** In certain embodiments, each of the plurality of convertible nucleobases comprises a chemically modifiable moiety that is activatable by redox.

**[00037]** In certain embodiments, the chemically modifiable moiety is capable of being activated by localized oxidation.

[00038] In certain embodiments, the chemically modifiable moiety is capable of being activated by oxidation using electrodes.

[00039] In certain embodiments, a nucleotide comprising the convertible nucleobase is selected from the group consisting of:



[00040] In certain embodiments, the convertible nucleobase is selected from the group consisting of O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

[00041] In certain embodiments, the first state and the second state of the plurality of convertible nucleobases are readable by a sequencing method capable of detecting and differentiating non-naturally occurring and/or modified nucleobases.

[00042] In certain embodiments, the first state and the second state of the plurality of convertible nucleobases are readable by nanopore sequencing.

[00043] In certain embodiments, the first state and the second state of the plurality of convertible nucleobases are readable by sequencing by synthesis.

[00044] In certain embodiments, when the plurality of convertible nucleobases are converted to the second state, properties of the plurality of convertible nucleobases are modified (e.g., having reduced size, altered shape, modified H-bonding, and/or modified polymerase substrate ability) as compared to the first state.

[00045] In certain embodiments, one or more of the plurality of convertible nucleobases are capable of being converted from the second state into a third state; wherein the one or more

of the plurality of convertible nucleobases are attached covalently to the nucleic acid polymer in the third state.

**[00046]** In certain embodiments, each of the plurality of convertible residues is capable of being independently and selectively converted.

**[00047]** In certain embodiments, the polymers provided herein further comprise a plurality of spacer residues linked via the backbone of the polymer, wherein each of the plurality of convertible residues are separated by one or more spacer residues of the plurality of spacer residues.

**[00048]** In certain embodiments, the iterative spacing among the plurality of convertible residues conforms to a resolution of a writing mechanism for encoding data on the polymer.

**[00049]** In certain embodiments, the iterative spacing among two adjacent convertible residues is equal to or greater than a resolution of a data encoding mechanism for encoding data into the polymer.

**[00050]** In certain embodiments, the resolution of the writing mechanism is at least 1 nm.

**[00051]** In certain embodiments, the plurality of spacer residues do not interfere with reading of the convertible residues.

**[00052]** In certain embodiments, the plurality of spacer residues in the polymer are the same spacer residues.

**[00053]** In certain embodiments, the plurality of spacer residues comprise two or more different spacer residues (e.g., different nucleobases such as different naturally occurring nucleobases).

**[00054]** In certain embodiments, the polymer consists essentially of spacer residues.

**[00055]** In certain embodiments, each of the plurality of convertible nucleobases are separated by 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50 spacer residues.

**[00056]** In certain embodiments, each of the plurality of convertible nucleobases are separated by 6 spacer residues.

**[00057]** In certain embodiments, the plurality of spacer residues are naturally occurring nucleobases, non-naturally nucleobases, tetrahydrofuran abasic residues, or ethylene glycol residues.

**[00058]** In certain embodiments, the plurality of spacer residues are naturally occurring nucleobases.

**[00059]** In certain embodiments, the polymers provided herein further comprise one or more delimiters linked to the backbone of the polymer.

[00060] In certain embodiments, each of the one or more delimiters comprises one or more naturally occurring nucleobases or non-naturally nucleobases.

[00061] In certain embodiments, the one or more delimiters comprise naturally occurring nucleobases.

[00062] In certain embodiments, the one or more delimiters separate two or more adjacent data fields within the polymer.

[00063] In certain embodiments, the polymers provided herein further comprise one or more data tags.

[00064] In certain embodiments, the one or more data tags comprise one or more naturally occurring nucleobases or non-naturally nucleobases.

[00065] In certain embodiments, the polymer is a nucleic acid polymer and the one or more data tags are present at the 5' or 3' end of the nucleic acid polymer.

[00066] In certain embodiments, the one or more data tags are incorporated to the nucleic acid polymer during the nucleic acid polymer is synthesized, during the plurality of convertible nucleobases are converted to the second state, or via ligation after the plurality of convertible nucleobases are converted to the second state.

[00067] In certain embodiments, the polymer can be stored under standard nucleic acid storage protocols.

[00068] In certain embodiments, the polymer is a nucleic acid polymer that can be stored in appropriate nuclease-free solution at room temperature, or at a lower temperature (e.g., - 20 °C).

[00069] In certain embodiments, the polymer can be stored at room temperature without stabilizer.

[00070] In another aspect, also provided herein are systems for data writing, comprising:

a writable polymer comprising a plurality of convertible residues iteratively spaced along and covalently linked to the backbone of the polymer, wherein each of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; wherein the plurality of convertible residues are attached covalently linked to the polymer in the first state and in the second state; and

a data writing device for writing data on the writable polymer.

**[00071]** In certain embodiments, the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.

**[00072]** In certain embodiments, the data writing device comprises a nanopore.

**[00073]** In certain embodiments, the data writing device comprises a microscope with a light source.

**[00074]** In certain embodiments, the data writing device converts the plurality of convertible nucleobases into the second state by light pulses, voltage pulses, an enzymatic agent, or a redox agent.

**[00075]** In certain embodiments, the data writing device converts the converts the plurality of convertible nucleobases into the second state by light pulses.

**[00076]** In certain embodiments, the data writing device comprises a light irradiation device.

**[00077]** In another aspect, also provided herein are methods for generating a writable nucleic acid polymer, comprising:

providing a circular single-stranded oligonucleotide template, wherein the circular single-stranded oligonucleotide template is complementary to a repeating data field that comprises convertible nucleobases; and

incubating the circular single-stranded oligonucleotide template in the presence of a nucleic acid primer, a polymerase, and triphosphate nucleotides, wherein the triphosphate nucleotides comprise convertible nucleobases in a first state and are capable of being converted from the first state into a second state, the first state and the second state being different.

**[00078]** In certain embodiments, the circular single-stranded oligonucleotide template comprises nucleobases complementary to the convertible nucleobases, and wherein the complementary nucleobases are iteratively spaced such that the incubation of the template with the nucleic acid primer, the polymerase, and the triphosphate nucleotides provides a nucleic acid polymer comprising a plurality of the convertible nucleobases iteratively spaced along and covalently linked via the backbone of the nucleic acid polymer; wherein the plurality of the convertible nucleobases are covalently linked to the nucleic acid polymer in the first state and in the second state.

**[00079]** In certain embodiments, the repeating data field further comprises spacer nucleobases, and wherein the triphosphate nucleotides further comprise triphosphate spacer nucleotides.

**[00080]** In yet another aspect, provided herein are methods for generating a writable nucleic acid polymer, comprising:

chemically synthesizing a plurality of oligomers, each oligomer comprises a plurality of convertible nucleobases iteratively spaced along and linked via the nucleic acid polymer backbone, wherein each of the plurality of convertible nucleobases has a first state and is capable of being converted from the first state into a second state; wherein the plurality of convertible nucleobases are attached covalently to the nucleic acid polymer in the first state and in the second state, the first state and the second state being different; and

ligating the plurality of oligomers to form the writable nucleic acid polymer

**[00081]** In certain embodiments, each of the plurality of oligomers comprises a plurality of spacer residues linked via the backbone of the nucleic acid polymer, wherein each of the plurality of the convertible nucleobases is separated by one or more spacer residues of the plurality of spacer residues.

**[00082]** In certain embodiments, the ligating step is via chemical ligation.

**[00083]** In certain embodiments, the ligating step is via enzymatic ligation.

**[00084]** In certain embodiments, a complementary DNA splint is used in the ligating step.

**[00085]** In certain embodiments, the method further comprises: annealing a plurality of complements to the oligomers prior to the ligating step.

**[00086]** In yet another aspect, provided herein are methods for writing data onto a writable polymer, comprising:

providing a writable polymer that comprises a plurality of convertible residues iteratively spaced along and covalently linked via the backbone of the polymer, wherein each convertible residues of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; and

selectively converting, utilizing a data writing device, one or more of the plurality of convertible residues into the second state such that a data encoded polymer is generated.

**[00087]** In certain embodiments, the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.

**[00088]** In certain embodiments, the data writing device comprises a nanopore, and the method further comprising: passing the writable polymer through the nanopore of the writing device, wherein the nanopore comprises converts one or more of the plurality of convertible residues into the second state.

**[00089]** In certain embodiments, the nanopore is a plasmonic nanopore that provides light pulses or redox energy to selectively convert convertible nucleobases from the first state into the second state.

**[00090]** In certain embodiments, the data writing device comprises a plasmonic well or channel, and the method further comprising: transferring the writable polymer into the plasmonic well or channel of the data encoding device, wherein the plasmonic well or channel provides light pulses or redox energy to selectively convert convertible nucleobases from the first state into the second state.

**[00091]** In certain embodiments, the data writing device selectively converts the convertible residues into the second state by light pulses, voltage pulses, an enzymatic agent, or a redox agent.

**[00092]** In certain embodiments, the data writing device selectively converts the convertible residues into the second state by light pulses.

**[00093]** In certain embodiments, the convertible residues become naturally occurring nucleobases after conversion into the second state.

**[00094]** In certain embodiments, the plurality of convertible residues comprise two or more types of convertible residues, wherein a first type of convertible residues are activatable by light of a first wavelength and a second type of convertible residues are activatable by light of a second wavelength.

**[00095]** In certain embodiments, the iterative spacing among the plurality of the convertible residues conforms to a resolution of the data writing device for selectively converting the convertible residues.

**[00096]** In certain embodiments, the selectively converting step does not require specific positioning of the writable polymer.

**[00097]** In certain embodiments, the conversion of the convertible residues into the second state is non-uniform on the data encoded polymer.

**[00098]** In certain embodiments, the conversion of the convertible residues into the second state is not limited to certain positions on the data encoded polymer.

**[00099]** In certain embodiments, the method further comprises stretching or combing the writable polymer (e.g., a writable DNA) on a solid support.

**[000100]** In certain embodiments, the method further comprises visualizing locations of the convertible residues using a dye.

**[000101]** In certain embodiments, the method further comprises locally illuminating or locally exciting the writable polymer.

**[000102]** In certain embodiments, the locally illuminating or locally exciting uses Stimulated Emission Depletion (STED) laser.

**[000103]** In certain embodiments, the method further comprises joining two or more data fields from two or more writable polymers end-to-end, resulting in a joined polymer comprising two or more data fields.

**[000104]** In certain embodiments, the method further comprises controlling the passage rate of the writable polymer through the nanopore of the writing device.

**[000105]** In certain embodiments, a plurality of writable polymers pass through the data writing device to write the same data (e.g., generating data redundancy).

**[000106]** In yet another aspect, also provided herein are methods for reading data from a polymer encoded with data, comprising:

providing the polymer encoded with data comprising convertible residues iteratively spaced along and covalently linked via the backbone of the polymer, wherein a first subset of the convertible residues are in a first state and a second subset of the convertible residues are in a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; and

passing the writable polymer encoded with data through a data reading device to read the encoded data on the polymer encoded with data.

**[000107]** In certain embodiments, the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.

**[000108]** In certain embodiments, the convertible residues in the first state can be converted into the second state via light.

**[000109]** In certain embodiments, wherein the data reading device comprises a nanopore.

**[000110]** In certain embodiments, wherein the data reading device is a sequencing device.

**[000111]** In certain embodiments, the sequencing device is a sequencing by synthesis device.

**[000112]** In certain embodiments, the method further comprises measuring current flow of electrolytes during passage of the writable polymer.

**[000113]** In certain embodiments, the method further comprises determining whether each of the plurality of convertible residues is in the first state or the second state based on the measured current flow of electrolytes during passage of the writable polymer.

**[000114]** In certain embodiments, the method further comprises re-passing the polymer encoded with data through the data reading device to re-read the encoded data on the polymer encoded with data.

**[000115]** In certain embodiments, the method further comprises validating and correcting the encoded data on the polymer encoded with data by comparing the encoded data on multiple copies of the polymer encoded with data.

**[000116]** In yet another aspect, also provided herein are methods for reading or decoding data from a nucleic acid polymer encoded with data, the method comprising:

providing a plurality of redundant copies of the nucleic acid polymer encoded with data comprising:

a plurality of converted nucleobases, wherein each converted nucleobase comprises a first nucleobase structure, wherein the first converted nucleobase has been converted from a first state into a second state, the first state and the second state being different; and

a plurality of convertible nucleobases, wherein each convertible nucleobase comprising a second nucleobase structure and a directly linked leaving group, and wherein the convertible nucleobase is provided in a first state and is capable of being converted from the first state into a second state by releasing the second leaving group from the second nucleobase structure, the first state and the second state being different;

wherein the converted nucleobases and convertible nucleobases are linked via the nucleic acid polymer backbone; and

sequencing each redundant copy of the plurality redundant copies of the nucleic acid polymer.

In certain embodiments, the method further comprises: detecting the plurality of converted nucleobases and the plurality of convertible nucleobases; and decoding the data based on the detected plurality of converted nucleobases.

**[000117]** In certain embodiments, the plurality of converted nucleobases in the first state and the second state are readable by a polymerase enzyme.

**[000118]** In certain embodiments, the plurality of convertible nucleobases in the first state and the second state are readable by a polymerase enzyme.

**[000119]** In certain embodiments, the plurality of converted nucleobases and the plurality of convertible nucleobases are detected based on the sequencing result of the redundant copies of the nucleic acid polymer encoded with data.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

**[000120]** The description and claims will be more fully understood with reference to the following figures and data graphs, which are presented as exemplary embodiments and should not be construed as a complete recitation of the scope of the disclosure.

**[000121]** FIGS. 1A and 1B provide a schematic of a writable nucleic acid polymer in accordance with various embodiments.

**[000122]** FIGS. 2A and 2B provide a schematic of a data encodable nucleic acid polymer in accordance with various embodiments.

**[000123]** FIGS. 3A-3G show structures of various example convertible nucleobases for use in a writable nucleic acid polymer.

**[000124]** FIG. 4 provides an example of convertible nucleobase O6-nitrobenzyl-guanine in accordance with various embodiments.

**[000125]** FIGS. 5A and 5B show structures of various example nucleotides comprising a convertible nucleobase for use in a writable polymer in accordance with various embodiments.

**[000126]** FIG. 6 provides molecular structure diagrams of various removable groups (e.g., leaving groups) in a convertible nucleobase for use in a writable polymer in accordance with various embodiments.

**[000127]** FIG. 7 provides a schematic of generating a writable nucleic acid polymer utilizing polymerase extension via a rolling circle reaction in accordance with various embodiments.

**[000128]** FIG. 8 provides a schematic of generating a writable nucleic acid polymer utilizing chemical synthesis and ligation in accordance with various embodiments.

**[000129]** FIGS. 9A-9C provide a schematic for encoding data in a writable nucleic acid polymer utilizing a nanopore and light energy in accordance with various embodiments.

**[000130]** FIGS. 10A-10C provide a schematic for encoding data in a data encodable nucleic acid polymer comprising pairs of convertible nucleobases utilizing a nanopore and light energy in accordance with various embodiments.

**[000131]** FIGS. 11A-11C illustrate encoding data in a writable nucleic acid polymer comprising convertible nucleobases utilizing a nanopore and light energy in accordance with

various embodiments. FIG. 11A: a writable nucleic acid polymer comprising convertible nucleobases  $C_a$  and  $C_b$ ; FIG. 11B: the writable nucleic acid polymer passing through a nanopore, certain convertible nucleobases (e.g., a  $C_a$  on the 3' end) has been converted by light energy to converted nucleobases (e.g.,  $C_a'$ ) as the written state; and FIG. 11C: certain convertible nucleobases  $C_a$  and  $C_b$  have been selectively converted to converted nucleobases  $C_a'$  and  $C_b'$ , respectively, resulting in a nucleic acid polymer encoded with data comprising stochastically or irregularly spaced converted nucleobases  $C_a'$  and  $C_b'$ .

**[000132]** FIGS. 12A-12C provide a schematic for encoding data in a writable nucleic acid polymer comprising duads utilizing a nanopore and light energy in accordance with various embodiments.

**[000133]** FIGS. 13A-13C provide molecular structure diagrams of dual-bit convertible nucleobases for use in a writable nucleic acid polymer in accordance with various embodiments.

**[000134]** FIGS. 14A and 14B provide data decoding strategies using a nanopore current-based sequencing (FIG. 14A) and sequencing by synthesis (FIG. 14B) in accordance with various embodiments.

**[000135]** FIG. 15 illustrates an example of encoding a data encodable nucleic acid polymer comprising convertible nucleobases with binary data 1010010, by selectively converting

certain  $\overset{*}{T}$  to T and  $\overset{\bullet}{G}$  to G, respectively. Certain convertible nucleobases in the data encodable nucleic acid polymer are skipped during the data encoding process, and the resulting nucleic acid polymer encoded with data comprises stochastically and/or irregularly spaced converted nucleobases (e.g., T and G).

## DETAILED DESCRIPTION

**[000136]** Provided herein are compositions of data-encodable polymers (e.g., nucleic acid polymers), and methods and systems thereof, for data encoding/decoding (writing/reading) and data storage. Also provided herein are method of making the polymers (e.g., nucleic acid polymers) described herein.

**[000137]** Turning now to the drawings and data, compositions and systems of nucleic acid data storage, methods of use and methods of synthesis, in accordance with various embodiments, are disclosed. In several embodiments, a system of data storage comprises writable (i.e., data-encodable) nucleic acid polymers having one or more nucleobases that are convertible. Accordingly, a writable nucleic acid polymer is akin to a blank tape that is

encodable, wherein the writable nucleic acid polymer is encoded by converting one or more of its nucleobases. Nucleobase conversion can be thought of as a binary code, where each convertible nucleobase is akin to a “bit,” unconverted nucleobases are akin to a “0,” and nucleobases that have been converted are akin to a “1.” It should be understood, however, that a binary code is not the only possibility, and codes can be written in ternary, quaternary, or other numeral system code, which can be done utilizing multiple types of convertible bases or performing multiple writings to further alter the state of a convertible base. In some embodiments, the conversion of a convertible nucleobase is stable, or permanent, which allows for long-term archiving. In some embodiments, the combination of two convertible nucleotides comprises a “bit”.

**[000138]** In some embodiments, a convertible residue (e.g., a convertible nucleobase) is referred to as a writable “bit,” and a converted residue (e.g., a converted nucleobase such as a native nucleobase) is referred to as a written “bit.”

**[000139]** In some embodiment, the terms “writable” and “data-encodable” are used herein interchangeably. In some embodiment, the terms “writing” and “data encoding” are used herein interchangeably.

**[000140]** In some embodiments, the terms “leaving group” and “removable group” are used herein interchangeably. In some embodiment, when referring to convertible nucleobases, the terms “pair” and “duad” are used herein interchangeably. “Duad,” used herein refers to a pair of different convertible nucleobases (e.g., writable bits) that are located close enough relative to one another in the polymers described herein (e.g., nucleic acid polymers) such that both are exposed to a single writing action or event (e.g. the same pulse of light or the same voltage pulse). Thus, the convertible nucleotides that comprise the duad are closer than the resolution of the writing action or event.

**[000141]** In other embodiments of the systems provided herein, the systems comprise two or more sets of convertible nucleobases (e.g., nucleobases having different structures, such as having different chemically modifiable moieties), where nucleobase conversion (e.g., cage group removal off of nucleobase) can be thought of as a binary code, and each convertible nucleobase (or sets of two or more convertible bases) is akin to a writable “bit” of data, and each converted nucleobase (or sets of two or more converted nucleobases) is akin to a written “bit” of data. In some embodiments, convertible nucleobases are utilized to encode a data bit, where conversion of a first nucleobase structure (i.e., a first set of convertible nucleobases) is akin to a “0,” and conversion of a second nucleobase structure (i.e., a second set of convertible nucleobases) of the pair is akin to a “1”, and data can be encoded by selective

conversion of nucleobases along the polymer (e.g., the nucleic acid polymer). In some embodiments, a pair of convertible nucleobases are utilized to encode data in a writable bit, where conversion of one nucleobase of the pair is akin to a “0,” and conversion of both nucleobases of the pair is akin to a “1” and data can be encoded by nucleobase pair conversions along the polymer. It should be understood, however, that a binary code is not the only possibility, and codes can be written in ternary, quaternary, or other numeral system code, which can be done utilizing multiple types of convertible bases or performing multiple writings to further alter the state a convertible base. In some embodiments, the conversion of a convertible nucleobase is stable for long periods, or permanent, which allows for long-term archiving.

**[000142]** In some embodiments, the nucleic acid polymer is a single-stranded nucleic acid polymer or a double-stranded nucleic acid polymer. In some embodiments, the nucleic acid polymer is a single-stranded nucleic acid polymer. In some embodiments, the nucleic acid polymer is a double-stranded nucleic acid polymer.

**[000143]** Some embodiments are directed towards compositions of writable nucleic acid polymers. Any appropriate nucleic acid polymer can be utilized, including (but not limited to) DNA, RNA, phosphorothioate DNA, glycerol nucleic acids (GNA), threose nucleic acids (TNA). Further, a nucleic acid polymer may be single stranded or double stranded. In several embodiments, a writable nucleic acid polymer comprises a plurality of convertible nucleobases that are linked by a polymer backbone. In certain embodiments, convertible nucleobases are spaced apart to provide spatial resolution such that each nucleobase can be independently and selectively converted in accordance with encoding. In some embodiments, spacer residues linked via the polymer backbone are utilized to provide spaces between the convertible nucleobases. In some embodiments, spacer residues are unreactive to the writing mechanism. In various embodiments, a writable nucleic acid polymer can further include delimiters and/or data tags for labeling the data, each of which can be provided by a particular sequence of nucleobases.

**[000144]** In some embodiments, any appropriate nucleic acid polymer can be utilized, including (but not limited to) DNA, RNA, phosphorothioate DNA, glycerol nucleic acids (GNA), threose nucleic acids (TNA), locked nucleic acids (LNA), and combinations thereof.

**[000145]** In some embodiments, the plurality of convertible nucleotides are capable of being incorporated into the nucleic acid polymer by one or more polymerase enzymes.

**[000146]** In some embodiments, the plurality of convertible nucleobases are non-naturally occurring nucleobases. In some embodiments, the plurality of convertible nucleobases are modified naturally occurring nucleobases or derivatives of naturally occurring nucleobases.

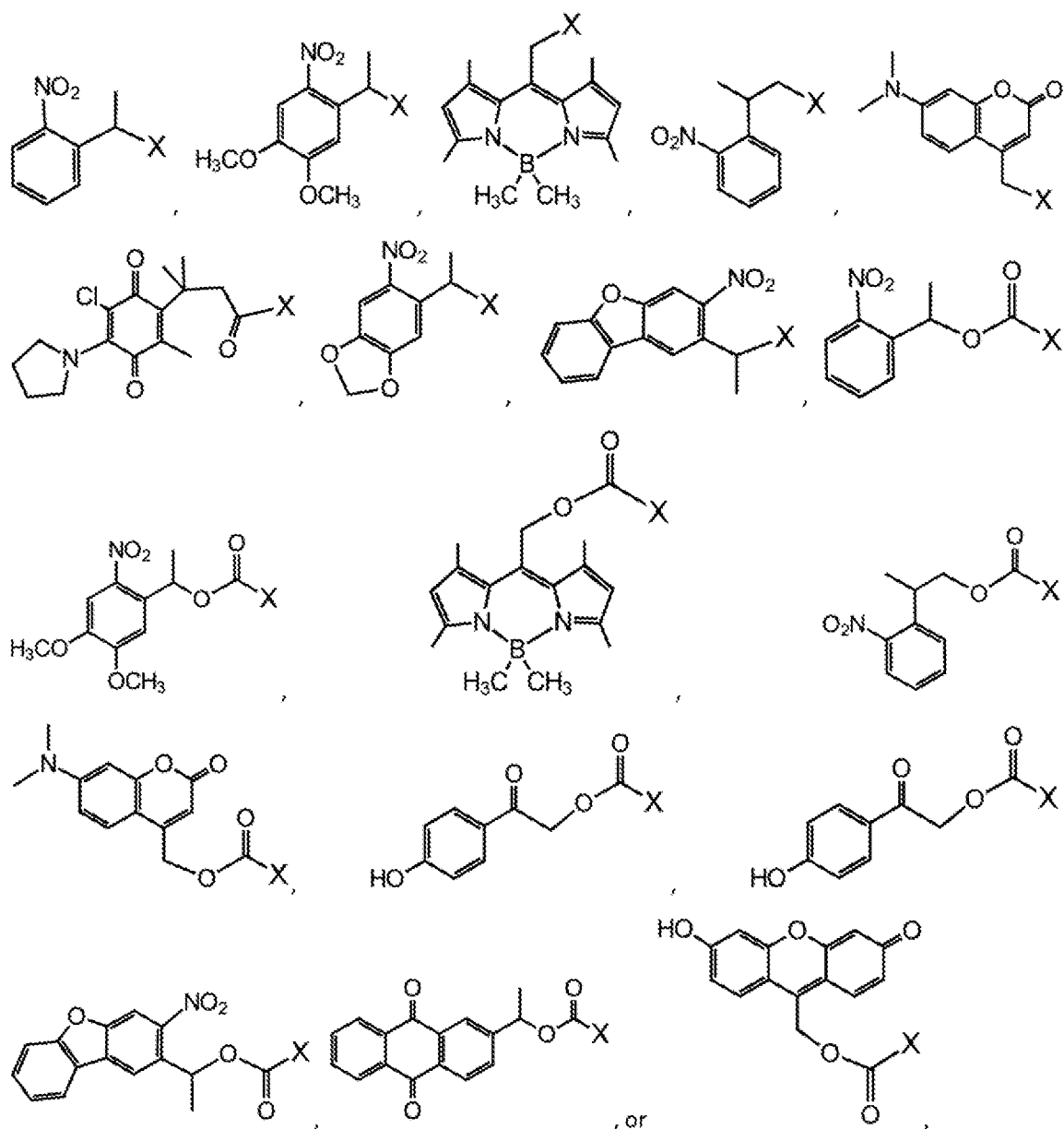
**[000147]** In some embodiments, each of the plurality of convertible nucleobases comprises a chemically modifiable moiety. In some embodiments, each of the plurality of convertible nucleobases the chemically modifiable moiety is directly attached to the base of the convertible nucleobases. In some embodiments, each of the plurality of convertible nucleobases the chemically modifiable moiety is attached to the base without a linker or a sidechain. In some embodiments, the plurality of convertible nucleobases are covalently linked to the backbone of the nucleic acid via a sugar of the backbone of the nucleic acid. In some embodiments, the removable group in the plurality of convertible nucleobases are covalently linked to the backbone of the nucleic acid via the nucleobase.

**[000148]** In some embodiments, the convertible nucleobases are linked to the backbone of the nucleic acid polymer in the same way that a nucleobase in a native nucleotide is linked to the backbone of the nucleic acid polymer (via the sugar in a nucleotide), without an intervening linker or as a sidechain.

**[000149]** In some embodiments, the nucleobase conversion (i.e., from the first state to the second state) is performed by removing one or more removal groups from the nucleobase. In several embodiments, the removable group is a caging group.

**[000150]** In one embodiment, the chemically modifiable moiety is activatable by light, thereby converting from the first state into the second state. In some embodiments, the conversion from the first state into the second state occurs via an irreversible reaction. In some embodiments, the convertible nucleobase becomes a naturally occurring nucleobase after conversion into the second state. In some embodiments, the convertible nucleobase becomes a native nucleobase after conversion into the second state. In one embodiment, the convertible nucleobase becomes guanine, adenine, thymine, uracil, or cytosine after conversion into the second state. In some embodiments, the backbone of the polymer (e.g., phosphate and sugar in nucleic acid polymer) remain unchanged during the conversion from the first state into the second state. In some embodiments, the chemically modifiable moiety is activatable by light, voltage, enzymatic agent, chemical reagent, or a redox agent or redox electrode, thereby converting from the first state into the second state. In some embodiments, the chemically modifiable moiety comprises one or more photo-removable groups.

**[000151]** In some embodiments, the one or more photo-removable groups are:



wherein X represents NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase to which the photo-removable group is attached.

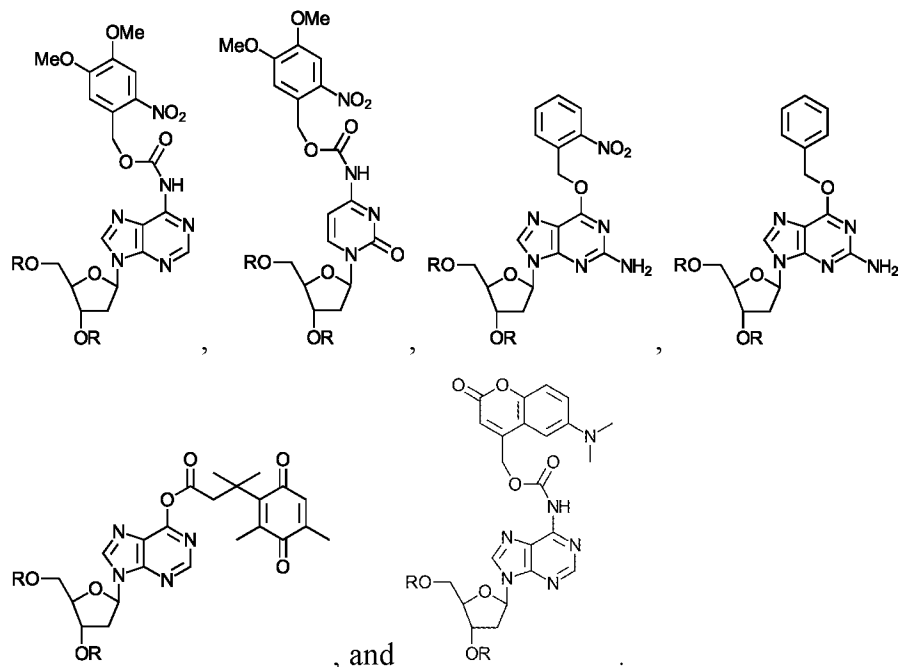
**[000152]** In some embodiments, the plurality of convertible nucleobases are capable of being converted by light of a wavelength of 325 nm, 360 nm, or 400 nm.

**[000153]** In some embodiments, the plurality of convertible nucleobases are capable of being converted by light of a wavelength of between 400 nm to 850 nm.

**[000154]** In some embodiments, each of the plurality of convertible nucleobases comprises a chemically modifiable moiety that is activatable or removable by redox. In some embodiments, the chemically modifiable moiety is capable of being activated by localized

oxidation. In some embodiments, the chemically modifiable moiety is capable of being activated by oxidation or reduction using one or more electrodes.

**[000155]** In some embodiments, a nucleotide comprising the convertible nucleobase is selected from the group consisting of:



**[000156]** In some embodiments, the convertible nucleobase (with a specific substitution position of the removable group) is selected from the group consisting of O6-guanine, O6-thioguanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, O4-uracil, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

**[000157]** In some embodiments, the first state and the second state of the plurality of convertible nucleobases are readable by a sequencing method capable of detecting and differentiating non-naturally occurring and/or modified nucleobases. In some embodiments, the first state and the second state of the plurality of convertible nucleobases are readable by nanopore sequencing. In some embodiments, the first state and the second state of the plurality of convertible nucleobases are readable by sequencing by synthesis. In some embodiments, when the plurality of convertible nucleobases are converted to the second state, properties of the plurality of convertible nucleobases are modified (e.g., having reduced size, altered shape, modified H-bonding, and/or modified polymerase substrate ability and/or polymerase coding) as compared to the first state. In some embodiments, one or more of the plurality of convertible nucleobases are capable of being converted from the second state into a third state; wherein the one or more of the plurality of convertible nucleobases are attached

covalently to the nucleic acid polymer in the third state. In some embodiments, each of the plurality of convertible residues is capable of being independently and selectively converted.

**[000158]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) comprise two or more different sets of convertible residues, each set of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different. In some embodiments, each of the plurality of convertible residues comprises a chemically modifiable moiety that can be activated and/or removed by light, and the two or more different sets of convertible residues are activatable and/or removable by light of different wavelengths. In some embodiments, a first set of convertible residues is activatable by light of a first wavelength, and a second set of convertible residues is activatable by light of a second wavelength, the first wavelength and the second wavelength being different.

**[000159]** In certain embodiments, the convertible nucleobases (or pairs of convertible bases) in the writable nucleic acid polymers described herein are iteratively spaced apart to provide spatial resolution such that each nucleobase (or each set or pair) can be independently and selectively converted in accordance with encoding. In certain embodiments, the convertible nucleobases are regularly or irregularly spaced apart, but data is encoded by identifying and selectively converting certain nucleobases to yield a nucleic acid polymer encoded with data. In some of the embodiments, the data encoding mechanism may skip any convertible nucleobases as necessary until it reaches the right convertible nucleobase in accordance with the code.

**[000160]** In some preferred embodiments, the convertible nucleobases are regularly spaced apart (e.g., by spacers), but data is encoded by identifying and selectively converting certain nucleobases to yield a nucleic acid polymer encoded with data comprising stochastically spaced converted nucleobases (i.e., written bits). One of the advantages of the writable nucleic acid polymers provided herein is no controlling of the position or passing rate of the writable nucleic acid polymers is needed. Certain convertible nucleobases can be skipped.

**[000161]** In several embodiments, a writing procedure is utilized to encode a writable nucleic acid with data. Data encoding can be performed by selectively converting convertible nucleobases of a nucleic acid molecule such that the written nucleic acid molecule contains a sequence of unconverted and converted nucleobases, akin to a binary code of “zeros” and “ones”. Any appropriate mechanism to chemically convert a nucleobase into second structure can be utilized. In accordance with various embodiments, a nucleobase is altered via light, voltage, enzymatic agent, chemical reagent, and/or a redox agent.

**[000162]** In some embodiments, the data written (data-encoded) nucleic acid molecule contains a sequence of converted nucleobases comprising a converted first set of nucleobases and a converted second set of nucleobases, akin to a binary code of “zeros” and “ones”.

**[000163]** In some embodiments, the data written (encoded) nucleic acid polymers are stored in accordance with standard nucleic acid storage protocols. For instance, data written nucleic acid polymers can be stored dry, as a precipitate, or in an appropriate nuclease-free solution at room temperature, or at colder temperatures (e.g., -20°C). Stabilizers such as (for example) alcohol, chelating agents and nuclease inhibitors, may be included with the stored nucleic acid. To read the data on written nucleic acid polymers, any appropriate sequencer capable of reading unnatural and/or altered nucleobases can be utilized, such as Oxford Nanopore Technologies PromethION, MinION, and GridION sequencing platforms (Oxford, UK) or Pacific Bioscience’s Single Molecule, Real-Time (SMRT) sequencing platform (Menlo Park, CA). Alternatively, a nanopore device can be fabricated or manufactured for reading the data. The nanopore can be comprised of solid-state materials, or can contain one or more proteins.

**[000164]** In some embodiments, the use of solid supports to sequester and stabilize the nucleic acid such as polymer beads, glass beads, or mineral solids are also contemplated. In some embodiments, the data on the written (encoded) nucleic acid polymers is decoded or read by sequencing by synthesis (SBS). And in some embodiments, a sequencer capable of reading modified and/or unmodified nucleobases can be utilized to decode or read data, such as Oxford Nanopore Technologies PromethION, MinION, and GridION sequencing platforms (Oxford, UK) or Pacific Bioscience’s Single Molecule, Real-Time (SMRT) sequencing platform (Menlo Park, CA).

**[000165]** The present disclosure overcomes many of the limitations associated with traditional nucleic acid data storage by separating the synthesis and data encoding into distinct steps. The disclosure provides molecular strategies for producing long strands of writable nucleic acids that, in themselves, do not encode data, but rather provide a template with the capacity for being written. Writable nucleic acid polymers can be produced in bulk in advance of data encoding. The disclosure further provides compositions and systems comprising convertible nucleobases (and pairs of convertible nucleobases) that act as writable “bits” of data, which can be switched from a first state into a second state, thus defining “0” and “1” in binary code. The disclosure further provides methods for writing data into the writable nucleic acid polymers provided herein at the single molecule level, thus consuming negligible amounts of material. Data writing may be achieved chemically or physically, utilizing (for example) light pulses or voltage pulses. Finally, because the written nucleic acid

polymers are long, they encode more data per molecule than do short DNAs, and can be efficiently and rapidly read by various sequencers existing within the current market. The compositions, systems, and methods described herein greatly increase the speed and density of nucleic acid data encoding while lowering cost.

### **Writable Polymers for Encoding Data**

**[000166]** In one aspect, provided herein are polymers for encoding data, comprising a plurality of convertible residues, iteratively spaced along and covalently linked to the backbone of the polymer, wherein each of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, and wherein the plurality of convertible residues are covalently linked to the polymer in the first state and in the second state. In some embodiments, the first state and the second state are different (e.g., the convertible residues have different structures when in the first and the second state). In some embodiments, the plurality of convertible residues in the first state and in the second state are readable by a polymerase enzyme. In some embodiments, the plurality of convertible residues are repeatedly spaced along the backbone of the polymer.

**[000167]** In some embodiments, the polymers described herein are nucleic acid polymers and the plurality of convertible residues are convertible nucleobases.

**[000168]** In certain embodiments, the convertible residues are iteratively spaced apart to provide spatial resolution such that each residue can be independently converted. In some embodiments, any appropriate spacer (e.g., non-writable, i.e., unreactive to the data writing mechanism) are between the convertible residues. In some embodiments, residues linked by the polymer backbone can be utilized as spacers. In some embodiments, the spacers spaced between the convertible residues in accordance with the spatial resolution of the writing mechanism and/or writing device. In some embodiments, spacers are residues, which may be unreactive to the writing mechanism. In some embodiments, these spacers are unmodified DNA nucleotides. In various embodiments, the polymer further comprises delimiters and/or data tags for labeling the data.

**[000169]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) further comprise a plurality of spacer residues linked via the backbone of the polymer, wherein each of the plurality of convertible residues are separated by one or more spacer residues of the plurality of spacer residues. In some embodiments, wherein the iterative spacing among the plurality of convertible residues conforms to a resolution of a writing mechanism for encoding data on the polymer. In some embodiments, the iterative spacing

among two adjacent convertible residues is equal to or greater than a resolution of a data encoding mechanism for encoding data into the polymer. In some embodiments, the resolution of the writing mechanism is at least 1 nm. In some embodiments, the plurality of spacer residues do not interfere with reading of the convertible residues. In some embodiments, the plurality of spacer residues in the polymer are the same spacer residues. In some embodiments, the plurality of spacer residues comprise two or more different spacer residues (e.g., different nucleobases such as different naturally occurring nucleobases).

**[000170]** In some embodiments, the polymers described herein are blank tapes. In some embodiments, the polymers described herein are blank tapes of DNA. Blank tape used herein refers to a writable nucleic acid polymer that comprises convertible nucleobases iteratively spaced along the writable nucleic acid polymer, such that conversion of convertible nucleobases from a first state into a second state results in encoding of data. The blank tape itself contains no data, but is capable of being encoded with data by use of an appropriate writing system (e.g., by light) via converting the convertible nucleobases. In some embodiments, the blank tape is writable sequentially from one end to the other end to encode data.

**[000171]** In some embodiments, the blank tape is writable over its entire length. In some embodiments, each convertible nucleobase in the blank tape is independently and individually writable.

**[000172]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) consist essentially of spacer residues.

**[000173]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) comprise no delimiter or data tag.

**[000174]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) consist of spacer residues and convertible residues (e.g., convertible nucleobases).

**[000175]** In some embodiments, each of the plurality of convertible nucleobases are separated by 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50 spacer residues. In some embodiments, each of the plurality of convertible nucleobases are separated by 6 spacer residues. In some embodiments, the plurality of spacer residues are naturally occurring nucleobases, non-naturally nucleobases, tetrahydrofuran abasic residues, or ethylene glycol residues. the plurality of spacer residues are naturally occurring nucleobases.

**[000176]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) further comprise one or more delimiters linked to the backbone of the polymer. In some embodiments, each of the one or more delimiters comprises one or more naturally occurring

nucleobases or non-naturally nucleobases. In some embodiments, the one or more delimiters comprise naturally occurring nucleobases. In some embodiments, the one or more delimiters separate two or more adjacent data fields within the polymer.

**[000177]** In some embodiments, the polymers described herein (e.g., nucleic acid polymers) further comprise one or more data tags. In some embodiments, the one or more data tags comprise one or more naturally occurring nucleobases or non-naturally nucleobases. In some embodiments, the polymer is a nucleic acid polymer and the one or more data tags are present at the 5' or 3' end of the nucleic acid polymer. In some embodiments, the one or more data tags are incorporated to the nucleic acid polymer during the nucleic acid polymer is synthesized, during the plurality of convertible nucleobases are converted to the second state, or via ligation after the plurality of convertible nucleobases are converted to the second state.

**[000178]** In some embodiments, the polymer can have any number or length of monomeric units, for example, from as short as 10 monomeric units to longer than 100,000 monomeric units. In various embodiments, the polymer has greater than 500 monomeric units, greater than 1,000 monomeric units, greater than 5000 monomeric units, greater than 10,000 monomeric units, greater than 50,000 monomeric units, or greater than 100,000 monomeric units.

**[000179]** In some embodiments, the nucleic acid polymer comprises greater than 10 convertible residues. In some embodiments, the nucleic acid polymer comprises greater than 100 convertible residues. In some embodiments, the nucleic acid polymer comprises greater than 500 convertible residues. In some preferred embodiments, the nucleic acid polymer comprises greater than 1,000 convertible residues. In some embodiments, the nucleic acid polymer comprises greater than 10,000 convertible residues. In some embodiments, the nucleic acid polymer comprises greater than 100,000 convertible residues.

**[000180]** In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 2 to 500. In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 2 to 200. In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 2 to 100. In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 2 to 10. In some embodiments, the

ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 10 to 50.

**[000181]** In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 10 to 100. In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 20 to 100. In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is between 20 to 50. In some embodiments, the ratio of the total number of monomeric units (e.g., nucleotides) to the convertible residues (e.g., convertible nucleobases) in the polymer (e.g., nucleic acid polymer) is greater than 100.

### **Writable nucleic acid polymers**

**[000182]** In certain embodiments, the polymers described herein (e.g., writable polymers) are nucleic acid polymers and the plurality of convertible residues are convertible nucleobases. In certain embodiments, the polymers described herein are nucleic acid polymers comprising a plurality of convertible nucleobases iteratively spaced along and covalently linked to the backbone of the nucleic acid polymer, wherein each of the plurality of convertible nucleobases has a first state (e.g., having a first state structure) and is capable of being converted from the first state into a second state (e.g., having a second state structure), the plurality of convertible nucleobases are covalently linked to the nucleic acid polymer in the first state and in the second state. In some embodiments, the first state and the second state are different and are both readable by a polymerase enzyme. In some embodiments, the nucleobase in the second state is a natural nucleobase. In some embodiments, the nucleobase in the second state is scarless (i.e., in native form of nucleobase, such as guanine, adenine, thymine, thiothymine, thioguanine, or 5-methylcytosine, or cytosine).

**[000183]** In some embodiments, the unwritten state is also referred to as the unconverted state, and the written state is also referred to the converted state.

**[000184]** Compounds in accordance with embodiments of the disclosure are based on nucleic acids having a plurality of convertible nucleobases, which are akin to writable data bits. Each convertible nucleobase can exist in two or more states, an unwritten state (e.g., a first state) akin to a "0", and at least a first written state (e.g., a second state of the nucleobase) akin to a

written bit denoting “1”, and in some embodiments a second written state (e.g., a third state of the nucleobase), and/or further written states (i.e., the written bits are further writable). In several embodiments, the writable nucleic acid polymers are synthesized with a plurality of convertible nucleobases in an “unwritten” state that are capable of being converted to “written” state(s). In some embodiments, two different convertible nucleobases are employed as a pair for encoding a single bit; conversion of one encodes a “0” while conversion of the other encodes a “1”. These writable nucleic acids can be created having long lengths (e.g., 5 to 50 kb, or more) and can be produced in bulk, prior to data writing.

**[000185]** In some embodiments, a single convertible nucleobase is utilized to encode a bit of data. In some embodiments, a set of two or more convertible nucleobases is utilized to enable the encoding of a bit of data. In some embodiments, a pair of two different convertible nucleobases are employed as a pair for enabling the encoding of a single bit. In some embodiments utilizing a pair of two different convertible nucleobases, conversion of a first nucleobase encodes a “0” while conversion of the other nucleobase encodes a “1”. In some embodiments utilizing a pair of two different convertible nucleobases, conversion of one nucleobase encodes a “0” while conversion of both of the nucleobases encodes a “1”.

**[000186]** In several embodiments, the writable nucleic acid polymer comprises a plurality of convertible nucleobases that are linked to the polymer backbone. In certain embodiments, convertible nucleobases are iteratively spaced apart to provide spatial resolution such that each nucleobase can be independently converted. In some embodiments, the spatial resolution depends, at least in part, on the writing mechanism. For instance, if an optical light source and device with 1 nm of resolution is used to alter nucleobases, then each convertible base needs to be separated by at least 1 nm. Any appropriate spacer between the alterable nucleobases can be utilized. In some embodiments, residues linked by the polymer backbone can be utilized as spacers. Because the distances between nucleobases in a double-stranded DNA polymer is about 0.34 nm, in accordance with numerous embodiments, three spacers are utilized for each nanometer of spatial resolution of the alteration-inducing source. In some embodiments, spacers are nucleobases, which may be unreactive to the writing mechanism. In various embodiments, a writable nucleic acid polymer can further include delimiters and/or data tags for labeling the data, each of which can be provided by a particular sequence of residues.

**[000187]** In several embodiments, a data encodable nucleic acid polymer comprises a plurality of convertible nucleobases that are linked by the polymer backbone. In certain embodiments, convertible nucleobases are regularly or irregularly spaced apart, but data is

encoded by identifying and selectively converting nucleobases to yield an encoded polymer. In some of the embodiments utilizing regularly or irregularly spaced convertible nucleobases, the data encoding mechanism may skip any convertible nucleobases as necessary until it reaches the right convertible nucleobase in accordance with the code, resulting in a nucleic acid polymer encoded with data comprising stochastically and/or regularly spaced converted nucleobases. In certain embodiments, convertible nucleobases (or sets of nucleobases) are iteratively spaced apart to provide spatial resolution such that each nucleobase (or each set of nucleobases) can be independently converted. The spatial resolution depends, at least in part, on the writing mechanism. For instance, if an optical light source and device with 1 nm of resolution is used to alter nucleobases, then each convertible base (or each set of nucleobases) needs to be separated by at least 1 nm. Any appropriate spacer between the convertible nucleobases (or sets of nucleobases) can be utilized. In some embodiments, residues linked by the polymer backbone can be utilized as spacers. Because the distances between nucleobases in a double-stranded DNA polymer is about 0.34 nm, in accordance with numerous embodiments, three spacers are utilized for each nanometer of spatial resolution of the alteration-inducing source. In some embodiments, spacers are nucleobases, which may be unreactive to the writing mechanism. In various embodiments, a data encodable nucleic acid polymer can further include delimiters and/or data tags for labeling the data, each of which can be provided by a particular sequence of residues.

**[000188]** In some embodiment, the writable nucleic acid polymers provided herein are capable of being written (e.g., convertible nucleobases selectively and sequentially converted to converted (e.g., naturally occurring or native nucleobases)) in both directions (e.g., in either the 5' to 3' direction or the 3' to 5' direction).

**[000189]** FIG. 1A illustrates an example of a writable nucleic acid polymer having a plurality of writable nucleobases. The writable nucleic acid polymer comprises a repeating strand sequence, which can exist as a single-stranded or double-stranded molecule. The repeating unit comprises convertible nucleobases, which may be natural or unnatural, that can undergo chemical changes from a first structure state to a second structure state, akin to a switch from a "0" state to a "1" state. Each of these convertible bases is akin to a "bit" for data encoding. It is understood that the definition of "1" and "0" is arbitrary, and simply meant to signify binary code. Prior to any data writing, convertible nucleobase bases are initially provided in the unconverted state. In some embodiments, the repeating unit of the writable nucleic acid polymer comprises data fields that include a plurality of convertible nucleobases, and may also contain spacers or sequences that delimit or separate bits. FIG. 1B provides another

example of a data field sequence having a plurality of convertible nucleobases separated by spacers. For example, as shown, three spacers are utilized between each convertible nucleobase which would provide 1 nm of spatial resolution. It is understood that longer spacer sequences can be used in cases of lower bit-writing resolution. In some embodiments, a writable nucleic acid polymer includes one or more unique data tag sequences, denoting documentation such as type of data, date, or other information. A unique data tag sequence may be written during the synthesis of the writable DNA, or may be written during the data writing process, or may be added on to an end via a primer, or may be added to the data strand via ligation after data writing.

**[000190]** FIG. 2A illustrates yet another example of a data encodable nucleic acid polymer having a plurality of convertible nucleobases in which each bit is a pair of convertible nucleobases that are iteratively repeated along the polymer. The data encodable nucleic acid polymer can exist as a single-stranded or double-stranded molecule. Each convertible nucleobase contains a removable group such that the nucleobase can be converted from one structure state to a second structure state by removing the removable group via light or redox energy. In reference to FIG. 2A, in some embodiments, conversion of the “C<sub>a</sub>” nucleobase yet maintaining the “C<sub>b</sub>” unconverted yields a “zero” bit and conversion of the “C<sub>b</sub>” nucleobase yet maintaining the “C<sub>a</sub>” unconverted yields “one” bit. In some embodiments, conversion of the “C<sub>a</sub>” nucleobase yet maintaining the “C<sub>b</sub>” unconverted yields a “zero” bit and conversion of both the “C<sub>a</sub>” and “C<sub>b</sub>” nucleobases yields a “one” bit. It is understood that the definition of “zero” and “one” is arbitrary, and simply meant to signify binary code.

**[000191]** FIG. 2B illustrates a further example of a data encodable nucleic acid polymer having a plurality of convertible nucleobases in which each bit is a convertible nucleobase that are spaced along the nucleic acid polymer. The data encodable nucleic acid polymer can exist as a single-stranded or double-stranded molecule. Each convertible nucleobase contains removal group such that the nucleobase can be converted from one structure state to a second structure state by removing the removable group via light or redox energy. As shown in FIG. 2B, in some embodiments, conversion of the “C<sub>a</sub>” nucleobase yields a “zero” bit and conversion of the “C<sub>b</sub>” nucleobase yields “one” bit. In these embodiments, convertible nucleobases can be left unconverted and thus do not contribute to the code of data.

**[000192]** In some embodiments, a data encodable nucleic acid polymer includes one or more unique data tag sequences, denoting documentation such as type of data, date, or other information. A unique data tag sequence may be incorporated during the synthesis of the

encodable polymer, or may be added on to an end via a primer, or may be added to the data strand via ligation after data encoding.

**[000193]** In various embodiments, writable nucleic acid polymers can be any length, for example, from as short as 15 nucleotides to longer than 100 kilobases. In various embodiments, a writable nucleic acid polymer is greater than 500 nucleotides long, is greater than 1000 nucleotides, is greater than 5000 nucleotides, is greater than 10,000 nucleotides, is greater than 50,000 nucleotides, or is greater than 100,000 nucleotides. Maximum lengths are only limited by the stability of the DNA, by the method used to make them, and by the method used to read the written data. In some embodiments, longer strands have the advantage of containing more data per molecule. Notably, current sequencing technologies can handle nucleic acid strands of tens to hundreds of thousands of bases in length (see N Kono and K. Arakawa, *Dev Growth Differ.* 2019; 61:316-326; and Q Chen and Z. Liu, *Sensors (Basel)*. 2019; 19:1886; the disclosures of which are each incorporated herein by reference).

**[000194]** Several embodiments are directed to convertible nucleobases, which can be incorporated into a writable nucleic acid polymer. A convertible nucleobase, in accordance with various embodiments, is a nucleic acid base that is capable of being converted from a first chemical state into a second chemical state by a controlled reaction chemistry. Any appropriate mechanism to convert a nucleobase from a first state into a second state can be utilized, including (but not limited to) light pulses, voltage pulses, enzymatic agent, chemical reagent, and/or redox agent. It is understood that “nucleobases” are not limited to naturally occurring structures, but may also embody unnatural nucleobases, such as designer nucleobases.

**[000195]** In some embodiments, the convertible nucleobases are nucleic acid bases that are capable of being converted from a first structural state into a second structural state by a controlled reaction chemistry. In some embodiments, a convertible nucleobase comprises a removable group that can be removed (e.g., as a leaving group) to provide a structural change. Any appropriate mechanism to convert a nucleobase from a first state into a second state can be utilized, including (but not limited to) light pulses, voltage pulses, enzymatic agent, chemical reagent, and/or redox agent. It is understood that “nucleobases” are not limited to naturally occurring structures, but may also embody unnatural nucleobases, such as designer nucleobases.

**[000196]** In some embodiments, the structural change results in a conversion of a non-natural nucleobase (e.g., nucleobase in the first structural state) to a natural or native nucleobase

(e.g., nucleobase in the second structure state). A natural or native nucleobase in this definition can be identified by standard sequencing methods. In some embodiments, the nucleobase in the second state is a natural nucleobase. In some embodiments, the nucleobase in the second state has no scar. In some embodiments, the nucleobase in the first state comprises a chemically modifiable moiety. In some embodiments, the nucleobase in the first state does not comprise a linker (or a linker moiety) or a sidechain between the base of the nucleobase and the chemically modifiable moiety. In some embodiments, when the nucleobase in the first state is converted to the second state, the chemically modifiable moiety is removed, thereby leaving the nucleobase in the second state a natural or native nucleobase. In some embodiments, the nucleobase in the first state and in the second state are readable or recognizable by polymerase. In some embodiments, the written nucleic acid polymer is readable by various sequencing methods, e.g., sequencing by synthesis (SBS).

**[000197]** In some embodiments, “scar” used herein refers to a group not normally found on naturally occurring DNA (such as a portion of a linker or a sidechain) that remains behind after a covalent bond is cleaved. Scars are frequently observed in some DNA sequencing technologies where a label is released by cleaving a linker during sequencing steps.

**[000198]** Provided in FIGS. 3A-3G are examples of convertible nucleobases in their unconverted and converted states. In several embodiments, convertible nucleobases can encode “bits” of data, enabling conversion from a first structure state to a second structure state, akin to “0” or “1” digital bit designations. In some embodiments, each state of the nucleobase is to be readable by sequencing methods capable of detecting and differentiating unnatural and/or modified bases, such as (for example) sequencing by synthesis or nanopore sequencing. As provided in FIGS. 3A-3G are examples of convertible nucleobases designed to be converted from a first state into a second state by localized pulses of light, which remove caging groups, reducing the size, altering shape or H-bonding of the base. Various photo removable groups can be incorporated into light convertible nucleobases (see, e.g., D. D. Young and A. Deiters, *Org Biomol Chem.* 2007; 5:999-1005; and Y. Wu, Z. Yang, and Y. Lu, *Curr Opin Chem Biol.* 2020; 57:95-104; the disclosures of which are each incorporated herein by reference). While a few examples are provided, it is understood that any appropriate photo-removable group and other nucleobases may be used in accordance with the various embodiments. FIG. 3E provides a convertible nucleobase that can be converted by localized enzymatic activity that removes a group resulting in altered size, shape, and H-bonding (see A. E. Pegg and T. L. Byers, *FASEB J* 1992; 6:2302-10. FIG. 3F provides a convertible nucleobase that is converted by localized oxidation, resulting in an altered shape

and polymerase substrate capability (K. Kino, et al., Genes Environ. 2017; 39:21). FIG. 3G provides a convertible nucleobase that is converted with a redox-removable group, again resulting in an altered size, shape, and/or polymerase substrate ability. In FIGS. 3A-3G, both the unconverted state and converted state of these nucleobase are uniquely identifiable by current sequencing methods.

**[000199]** FIG. 4 illustrates the conversion of a convertible nucleobase O6-nitrobenzyl-guanine to guanine by using light energy to break the bond with the nitrobenzyl group. This conversion can represent a bit of data or can be utilized in combination with one or more other convertible nucleobases to represent a writable bit of data. When decoding data via sequencing by synthesis, an unconverted O6-nitrobenzyl-guanine will be read as a mix of A and G and after conversion, the resulting guanine will be read as >99% G.

**[000200]** FIGS. 5A-5B show more examples of convertible nucleobases that can be converted from a first state into a second state by localized pulses of light, which remove caging groups, yielding natural nucleobase structures. Each exemplary convertible nucleobase includes a caging or removable group, which is denoted as “CG” in the structure drawings. While a few examples are provided, it is understood that any appropriate convertible nucleobase structure that includes a photoremovable caging group may be used in accordance with the various embodiments. In FIGS. 5A-5B, both the unconverted state and converted state of these nucleobase structures are uniquely identifiable by current sequencing methods.

**[000201]** FIG. 6 provides further examples of photoremovable caging groups that can be utilized with the nucleobase structures to provide convertible nucleobases that can be converted from a first state into a second state by localized pulses of light. In various embodiments, any one of the photoremovable caging groups of FIG. 6 can be combined with the nucleobase structures in FIG. 4 and 5A-5B. The photoremovable caging groups include a linker denoted as “X” which connect to the nucleobase structure denoted as R. In addition to the examples provided, various other photoremovable caging groups can be incorporated into light convertible nucleobases (see, e.g., D. D. Young and A. Deiters, Org Biomol Chem. 2007; 5:999-1005; and Y. Wu, Z. Yang, and Y. Lu, Curr Opin Chem Biol. 2020; 57:95-104; the disclosures of which are each incorporated herein by reference).

**[000202]** Numerous embodiments are also directed to a writable nucleic acid polymer further incorporating one or more of spacers, delimiters, and data tags. In accordance with various embodiments, a spacer is molecular residue incorporated within a writable nucleic acid polymer that provides a requisite space between convertible nucleobases in accordance with

spatial resolution of the data writing mechanism. In many embodiments, a spacer will be distinguishable from convertible nucleobases such that when the data is read in a sequencer, the spacer does not interfere with the ability to read the convertible nucleobases. In some embodiments, a spacer is unreactive with the data writing mechanism. In some embodiments, a writable nucleic acid polymer will utilize the same residue repeatedly for each and every spacer. In some embodiments, however, a writable nucleic acid polymer will utilize two or more different residues as spacers. Any appropriate residue that is distinguishable from the convertible nucleobases may be utilized as spacers, including naturally occurring nucleobases, unnatural nucleobases, tetrahydrofuran abasic residues, and/or ethylene glycol residues.

**[000203]** In some embodiments, a spacer is distinguishable from convertible nucleobases and/or converted nucleobases such that when the data is read in a sequencer, the spacer does not interfere with the ability to encode data and decode/read the encoded data. In some embodiments, a spacer is unreactive with the data encoding mechanism.

**[000204]** A delimiter, in accordance with various embodiments, is a residue that signifies a boundary. In some embodiments, a delimiter is utilized to separate two adjacent data fields. Any appropriate residue that is distinguishable from the convertible nucleobases may be utilized as a delimiter, including naturally occurring nucleobases, unnatural nucleobases, tetrahydrofuran abasic residues, and/or ethylene glycol residues.

**[000205]** In several embodiments, a data tag is a string of residues (typically 4 or more residues) that signifies certain data. For instance, a data tag can signify type of data, date, data source, or any other information. Any appropriate residues that are distinguishable from the convertible nucleobases may be utilized as data tag residues, including naturally occurring nucleobases, unnatural nucleobases, tetrahydrofuran abasic residues, and/or ethylene glycol residues.

**[000206]** In another aspect, also provided herein are methods for generating a writable nucleic acid polymer, comprising providing a circular single-stranded oligonucleotide template, wherein the circular single-stranded oligonucleotide template is complementary to a repeating data field that comprises convertible nucleobases; and incubating the circular single-stranded oligonucleotide template in the presence of a nucleic acid primer, a polymerase, and triphosphate nucleotides, wherein the triphosphate nucleotides comprise convertible nucleobases in a first state and are capable of being converted from the first state into a second state, the first state and the second state being different.

**[000207]** In some embodiments, the circular single-stranded oligonucleotide template comprises nucleobases complementary to the convertible nucleobases, and wherein the complementary nucleobases are iteratively spaced such that the incubation of the template with the nucleic acid primer, the polymerase, and the triphosphate nucleotides provides a nucleic acid polymer comprising a plurality of the convertible nucleobases iteratively spaced along and covalently linked via the backbone of the nucleic acid polymer; wherein the plurality of the convertible nucleobases are covalently linked to the nucleic acid polymer in the first state and in the second state.

**[000208]** In some embodiments, the repeating data field further comprises spacer nucleobases, and wherein the triphosphate nucleotides further comprise triphosphate spacer nucleotides.

**[000209]** In another aspect, also provided herein are methods for generating a writable nucleic acid polymer, comprising chemically synthesizing a plurality of oligomers, each oligomer comprises a plurality of convertible nucleobases iteratively spaced along and linked via the nucleic acid polymer backbone, wherein each of the plurality of convertible nucleobases has a first state and is capable of being converted from the first state into a second state; wherein the plurality of convertible nucleobases are attached covalently to the nucleic acid polymer in the first state and in the second state, the first state and the second state being different; and ligating the plurality of oligomers to form the writable nucleic acid polymer.

**[000210]** In some embodiments, each of the plurality of oligomers comprises a plurality of spacer residues linked via the backbone of the nucleic acid polymer, wherein each of the plurality of the convertible nucleobases is separated by one or more spacer residues of the plurality of spacer residues. In some embodiments, the ligating step is via chemical ligation. In some embodiments, the ligating step is via enzymatic ligation. In some embodiments, a complementary DNA splint is used in the ligating step.

**[000211]** In some embodiments, the plurality of oligomers have the same sequence. In some embodiments, the plurality of oligomers are a plurality of copies of the same sequence. In some embodiments, the plurality of oligomers have different sequences.

**[000212]** In some embodiments, the method further comprising annealing a plurality of complements to the oligomers prior to the ligating step.

**[000213]** Writable nucleic acids can be generated by any appropriate method for generating long nucleic acid polymers. Generally, in accordance with various embodiments, polymerase extension or chemical synthesis is utilized to generate writable nucleic acid polymers. If

polymerase extension is utilized, appropriate convertible nucleobases and residues that can be polymerized by the polymerase are to be utilized. If chemical synthesis is utilized, a broader range of convertible nucleobases and residues, but generally synthesis results in shorter nucleic acid strands (e.g., between 10 and 200 residues), which can be ligated together to generate longer nucleic acid polymers. It is understood that both polymerase and ligation methods can construct repeating writable polymers in either single-stranded or double-stranded states.

**[000214]** Illustrated in FIG. 7 is an example of generating a writable nucleic acid utilizing polymerase extension, and in particular, the figure illustrates an enzymatic rolling circle reaction method. In certain embodiments, a circular single-stranded DNA oligonucleotide is utilized as template (M. G. Mohsen and E. T. Kool, *Acc Chem Res.* 2016; 49: 2540–2550, the disclosure of which is incorporated herein by reference). The circular single-stranded DNA oligonucleotide is complementary to the repeating data field that comprises convertible nucleobases. In various embodiments, the circular single-stranded DNA oligonucleotide further comprises spacers, delimiters, and/or data tags. In various embodiments, the circular DNA size is 2-2000 nucleotides in length, preferably 2-200 nucleotides in length, and more preferably 45-95 nucleotides in length.

**[000215]** Once the nucleic acid circular template encoding the repeating data fields is constructed, it is incubated with a nucleic acid primer, a polymerase, a suitable buffer to support polymerase activity, and nucleoside triphosphates suitable for generating the writable nucleic acid. The primer binds the circle and the polymerase then produces a long repeating complement of the circle. Rolling circle nucleic acid synthesis is documented to proceed for many thousands of nucleotides, producing long DNA repeats (*see* M. M. Ali, et al., *Chem Soc Rev.* 2014; 43:3324-41; and M. G. Mohsen and E. T. Kool, *Acc Chem Res.* 2016 Nov 15; 49(11): 2540–2550; the disclosures of which are incorporated herein by reference). In some embodiments, a data tag is utilized, which may be included at the remote 5'-end of the primer, and remains non-complementary to the DNA circle. Rolling circle DNA synthesis in this case will result in the repeating writable nucleic acid with a data tag attached to the 5'-end. If writable nucleic acid polymers are desired to be double-stranded, a primer complementary to the repeating data fields can be used together with a polymerase and nucleotides complementary to the first polymer to generate the complementary strand.

**[000216]** FIG. 8 illustrates a chemical synthesis and ligation method for generating a writable nucleic acid. In some instances, nucleotides for incorporation into a writable nucleic acid are not efficient polymerase substrates, especially many unnatural nucleobases, preventing the

ability to effectively use a polymerase to generate long strands of the nucleic acid polymer. In a chemical synthesis and ligation approach, short writable nucleic acid polymers are constructed on a DNA synthesizer, which can be done utilizing phosphoramidite synthesis protocols, typically resulting in polymer lengths of 10-200 nucleotides. To assist in ligation, in some embodiments, the short-synthesized polymer further comprises a 5'-phosphate group and a native unaltered 3'-hydroxyl group. A DNA ligase enzyme in the presence of ATP (e.g., T4 DNA ligase) will join the short polymers together to generate a long repeating polymer. In some embodiments, a complementary "splint" nucleic acid oligonucleotide that can hybridize to the reactive ends is utilized to assist ligation.

**[000217]** In some embodiments, to generate a double-stranded writable nucleic acid, a nucleic acid complement comprising a 5'-phosphate group is synthesized. Prior to ligation, the complement strand hybridizes with the writable nucleic acid. In some embodiments, hybridization of the complement strand results in a duplex with sticky ends that can be efficiently ligated into a double-stranded writable nucleic acid polymer utilizing a ligase enzyme.

**[000218]** Ligation-derived polymer molecules may result in a range of polymer lengths. In some embodiments, a mixture of polymers with variable lengths is used for data encoding. In some embodiments, a specific length is enriched and/or isolated (e.g., by electrophoresis) and subsequently used for data encoding.

**[000219]** Several embodiments are directed to polymerase expansion of writable nucleic acid polymers via repetitive expansion using a thermostable polymerase (e.g., DNA polymerase from *Thermococcus litoralis*). For more on polymerase expansion of repetitive regions, see J. S. Hartig and E. T. Kool, *Nucleic Acids Res.* 2005; 33:4922-7, the disclosure of which is incorporated herein by reference.

**[000220]** If the ends of the data field DNA to be ligated are inefficient as a ligase enzyme substrate because of poor hybridization or an unnatural structure that interferes with the enzyme, in accordance with various embodiments, natural nucleobases can be added at ligation sites to ensure a good hybridization/ligation. In some embodiments, chemical ligation is utilized to generate a writable nucleic acid polymer. Chemical ligation can be achieved with cyanogen bromide, with carbodiimide reagents, or by nucleophilic reaction of a phosphorothioate group on one nucleic acid polymer strand terminus and a leaving group, such as (for example) iodide, on the other nucleic acid polymer strand terminus. Although chemical ligation involves joining of a phosphate end to a hydroxyl end, the reaction may be carried out with a 5'-phosphate and 3'-hydroxyl, or a 3'-phosphate and a 5'-hydroxyl. Such

methods of chemical ligation have been described (see E. T. Kool, Acc Chem Res. 1998; 31:502–510; C. Obianyor, et al., Chembiochem. 2020; 21:3359-3370; and Y. Xu and E. T. Kool, Nucleic Acids Res. 1999; 27:875-81; the disclosures of which are each incorporated herein by reference).

### **Methods and Systems of Data Writing and Reading**

**[000221]** In another aspect, provided herein are systems and methods for writing or reading the writable or written polymers provided herein (e.g., nucleic acid polymers).

#### Systems

**[000222]** In another aspect, provided herein are systems for data writing, comprising: a writable polymer comprising a plurality of convertible residues iteratively spaced along and covalently linked to the backbone of the polymer, wherein each of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; wherein the plurality of convertible residues are attached covalently linked to the polymer in the first state and in the second state; and a data writing device for writing data on the writable polymer.

**[000223]** In some embodiments, the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases. In some embodiments, the data writing device comprises a nanopore. In some embodiments, the data writing device converts the plurality of convertible nucleobases into the second state by light pulses, voltage pulses, an enzymatic agent, or a redox agent. In some embodiments, the data writing device converts the plurality of convertible nucleobases into the second state by light pulses. In some embodiments, the data writing device comprises a light irradiation device.

#### Methods for Writing/Encoding Writable Polymers

**[000224]** In yet another aspect, provided herein are methods for writing data onto a writable polymer, comprising: providing a writable polymer that comprises a plurality of convertible residues iteratively spaced along and covalently linked via the backbone of the writable polymer, wherein each convertible residues of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; and selectively converting,

utilizing a data writing device, one or more of the plurality of convertible residues into the second state such that a data encoded polymer is generated.

**[000225]** Several embodiments are directed towards writing and reading data on nucleic acid polymers. In many embodiments, a writable nucleic acid polymer is provided having convertible nucleobases iteratively spaced along the writable polymer. The provided writable nucleic acid polymer may also have spacers, delimiters, and data tags, as described herein. To write data upon a nucleic acid polymer, in accordance with various embodiments, an individual strand is passed through a device having a nanopore. The device having a nanopore further provides a means for selectively converting a convertible nucleobase from a first state into a second state. A number of means can be utilized for converting a convertible nucleobase, including (but not limited to) light pulses, voltage pulses, an enzymatic agent, a chemical reagent, and/or a redox agent. An example of a nanopore device for passing DNA through and encoded with localized light pulses is described within the examples provided in the Exemplary Embodiments.

**[000226]** In some embodiments, the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases. In some embodiments, the data writing device comprises a nanopore, and the method further comprising passing the writable polymer through the nanopore of the writing device, wherein the nanopore comprises converts one or more of the plurality of convertible residues into the second state.

**[000227]** In some embodiments, the nanopore is a plasmonic nanopore that provides localized excitation energy to selectively convert convertible nucleobases from the first state into the second state. In some embodiments, the data writing device comprises a plasmonic well or channel, and the method further comprising transferring the writable polymer into the plasmonic well or channel of the data encoding device, wherein the plasmonic well or channel provides local excitation from light pulses to selectively convert convertible nucleobases from the first state into the second state. In some embodiments, the data writing device selectively converts the convertible residues into the second state by light pulses, voltage pulses, an enzymatic agent, or a redox agent. In some embodiments, the data writing device selectively converts the convertible residues into the second state by light pulses.

**[000228]** In some embodiments, the convertible residues become naturally occurring nucleobases after conversion into the second state.

**[000229]** In some embodiments, the starting position and/or the ending positions of the writing on the writable polymer can be any position (i.e., any convertible residue such as

convertible nucleobase) in the writable polymer (e.g., writable nucleic acid polymer) and specific starting and/or ending positions are not needed.

**[000230]** In some embodiments, the selectively converting step starts on either end of the writable polymer (e.g. the 5' or 3' end of a nucleic acid polymer). In some embodiment, the selectively converting step starts on the 5' or the 3' end of the nucleic acid polymer. In some embodiment, the selectively converting step selectively converts the convertible residues (e.g., convertible nucleobases) in either direction of the writable polymer. In some embodiments, the selectively converting step selectively converts the convertible nucleobases (e.g., writable bits) in either the 5' to 3' direction or the 3' to 5' direction. In some embodiment, the selectively converting step starts on the 5' end of the nucleic acid polymer. In some embodiment, the selectively converting step starts on the 3' end of the nucleic acid polymer.

**[000231]** In some embodiments, the writing starts at any position (e.g., any convertible residue such as convertible nucleobase) on the writable polymer. In some embodiments, the writing ends at any position (e.g., any convertible residue such as convertible nucleobase) on the writable polymer. In some embodiments, the writing starts and ends at any position (e.g., any convertible residue such as convertible nucleobase) on the writable polymer.

**[000232]** In some embodiments, the writable polymer is writable over its entire length, and the writing starts at the beginning position (e.g., the 3' end of a nucleic acid polymer) and ends at the end position (e.g., the 5' end of the nucleic acid polymer).

**[000233]** In some embodiments, the plurality of convertible residues comprise two or more types of convertible residues, wherein a first type of convertible residues are activatable by light of a first wavelength and a second type of convertible residues are activatable by light of a second wavelength. In some embodiments, the iterative spacing among the plurality of the convertible residues conforms to a resolution of the data writing device for selectively converting the convertible residues. In some embodiments, the selectively converting step does not require specific positioning of the writable polymer. In some embodiments, the conversion of the convertible residues into the second state is non-uniform on the data encoded polymer. In some embodiments, the conversion of the convertible residues into the second state is not limited to certain positions on the data encoded polymer.

**[000234]** In some embodiments, the writable polymer comprises a plurality of convertible residues regularly spaced along the writable polymer. In some embodiments, the data encoded polymer after the data is written comprises stochastically or irregularly spaced converted nucleobases.

**[000235]** In some embodiments, the plurality of convertible nucleobases are capable of being converted by light of a wavelength of 325 nm, 360 nm, or 400 nm.

**[000236]** In some embodiments, the plurality of convertible nucleobases are capable of being converted by light of a wavelength of between 400 nm to 850 nm.

**[000237]** In some embodiments, the method further comprises stretching or combing the writable polymer (e.g., a writable DNA) on a solid support.

**[000238]** In some embodiments, the method further comprises visualizing locations of the convertible residues using a dye.

**[000239]** In some embodiments, the method further comprises locally illuminating or locally exciting the writable polymer. In some embodiments, the locally illuminating or locally exciting uses Stimulated Emission Depletion (STED) laser.

**[000240]** In some embodiments, the method further comprises joining two or more data fields from two or more writable polymers end-to-end, resulting in a joined polymer comprising two or more data fields.

**[000241]** In some embodiments, the method further comprises controlling the passage rate of the writable polymer through the nanopore of the writing device.

**[000242]** In some embodiments, a plurality of writable polymers pass through the data writing device or multiple devices in parallel to write the same data (e.g., generating data redundancy).

**[000243]** In some embodiments, data encoded polymers generated by selectively converting convertible nucleobases comprises different polymer molecules encoded with the same data. In some embodiments, the data encoded nucleic acid polymers comprise converted nucleobases at different positions along the nucleic acid polymers (e.g., differently and optionally irregularly spaced) but encoding the same data (e.g., the sequential order of the written data bits are the same among different encoded polymer molecules).

**[000244]** In some embodiments, to encode data on a writable nucleic acid polymer provided herein, in accordance with various embodiments, an individual polymer has light energy or redox energy impinged upon the polymer in an iterative fashion such that it can controllably and selectively convert the convertible nucleobases to encode a data code (e.g., a binary data code).

**[000245]** Although a device with a nanopore is described, any device that can controllably and selectively convert the convertible nucleobases in accordance with a data code. In some embodiments, the device utilizes plasmonic channels or plasmonic wells for controllably and selectively converting the convertible nucleobases.

**[000246]** In several embodiments, as a writable nucleic acid polymer passes through the nanopore, the device selectively provides the means for converting the convertible nucleobase. For instance, if a nucleobase is to be converted into a second state via light pulses, as the nucleic acid polymer passes through the nanopore, the device can provide light such that it contacts the convertible nucleobase and converts the convertible nucleobase into the second state. If a nucleobase is to remain in a first state, the device will not provide light such that the convertible nucleobase will pass through the nanopore without conversion. In many embodiments, to ensure a device only converts a single nucleobase, the convertible nucleobase can be flanked with spacers in accordance with the device's writing resolution. For instance, if an optical light source and device with 1 nm of resolution is used to alter nucleobases, then each convertible base needs to be separated by at least 1 nm.

**[000247]** In certain embodiments, if a nucleobase is to be converted into a second state via light pulses, as the nucleic acid polymer passes through the nanopore, the device can provide light such that it only contacts the set of convertible nucleobases to be converted. If a nucleobase is to remain in the initial state, the device will not provide light such that the convertible nucleobase will pass through the nanopore without conversion. In many embodiments, to ensure a device only converts a set of nucleobase, the set of convertible nucleobases can be flanked with spacers in accordance with the device's writing resolution.

**[000248]** In some embodiments, to ensure a device only converts a single nucleobase (or a set of nucleobases), the device utilizes two or more means for converting a nucleobase; a first means being able to convert a first nucleobase structure but not a second nucleobase structure and a second means being able to convert the second nucleobase structure but not the first nucleobase structure. For instance, a device can utilize two wavelengths of light for providing energy such that the first wavelength is able to convert a first nucleobase structure but not a second nucleobase structure and a second wavelength is able to convert the second nucleobase structure but not the first nucleobase structure.

**[000249]** In some embodiments, to ensure a device only converts a single nucleobase (or a set of nucleobases), the device utilizes two or more means for converting a nucleobase; a first means being able to convert a first nucleobase structure but not a second nucleobase structure and a second means being able to convert both the first nucleobase structure and the second nucleobase structure concurrently as a pair. For instance, a device can utilize two wavelengths of light for providing energy such that the first wavelength is able to convert a first nucleobase structure but not a second nucleobase structure and a second wavelength is

able to convert both the first nucleobase structure and the second nucleobase structure concurrently as a pair.

**[000250]** In many embodiments, the writing device is provided a code for writing the data into the nucleic acid polymer. Accordingly, the writing device will selectively convert various nucleobases of the polymer that are akin to being a “1” in binary code, while selectively allowing nucleobases of the polymer to pass through the pore without conversion that are akin to being a “0”. After writing a data code into the nucleic acid polymer, it can be stored by any appropriate means for storing nucleic acid molecules. For instance, data written nucleic acid polymers can be stored dry, as a precipitate, or in an appropriate nuclease-free solution at room temperature, or at colder temperatures (e.g., -20°C). Stabilizers such as (for example) alcohol, chelating agents and nuclease inhibitors, may be included with the stored nucleic acid.

**[000251]** In some embodiments, the polymers provided herein (e.g., nucleic acid polymers) can be stored under standard nucleic acid storage protocols. In some embodiments, the polymer is a nucleic acid polymer that can be stored in appropriate nuclease-free solution at room temperature, or at a lower temperature (e.g., -20 °C). In some embodiments, the polymer can be stored at room temperature without stabilizer.

**[000252]** In many embodiments, the data encoding device is provided a code for writing the data into the nucleic acid polymer. Accordingly, in some embodiments, the encoding device will selectively convert various nucleobases of the polymer that in accordance with the code. In some embodiments that use solitary nucleobases as a bit, a data is encoded by selecting converting some of the nucleobase and selectively not converting the others, resulting in a binary code of converted and unconverted nucleobases. In some embodiments that use solitary nucleobases as a bit, a data is encoded by selectively converting some of the nucleobase into a first converted structure and selectively converting others into a second converted structure, resulting in a binary code of converted nucleobases; any unconverted nucleobases remain unencoded and are not utilized to decode the data code.

**[000253]** In some embodiments that utilize a set of nucleobases to encode a bit, each set will comprise at least two convertible nucleobases and the encoding device will selectively convert a first nucleobase of some of the sets into a converted structure and selectively convert a second nucleobase of other sets into a converted structure, resulting in a binary code. In some embodiments that utilize a set of nucleobases to encode a bit, each set will comprise at least two convertible nucleobases and the encoding device will selectively

convert a first nucleobase of some of the sets into a converted structure and selectively convert both nucleobases of other sets into a converted structure, resulting in a binary code.

**[000254]** In some embodiments, nucleic acid polymers most efficiently store data at the single molecule level, providing the highest potential density of information. In some embodiments, however, if redundancy of data is required for better accuracy of data storage, then a plurality of nucleic acid polymers could be used to redundantly write the same data on each polymer of the plurality. Error correction algorithms are already well developed for digital data storage, and some of these algorithms can be applied in the present approach (see J. Li, et al., IEEE Transactions on Emerging Topics in Computing. 2021; 9:651-663, the disclosure of which is incorporated herein by reference).

**[000255]** In various embodiments in which the encoded data is to be decoded by sequencing by synthesis (SBS), it may be desirable to have a redundancy of data and thus the same data on each polymer of the plurality. For instance, when using a nucleobase structure such as O6-nitrobenzyl-guanine, the structure is read as a mix of A and G using SBS and thus a redundancy of reading the structure would be needed to interpret whether the structure is O6-nitrobenzyl-guanine, guanine, or adenine. In some methods of SBS, the redundancy is inherent to each single sequence being read.

*Methods for Reading/Decoding Writable Polymers*

**[000256]** In another aspect, also provided herein are methods for reading data from a polymer encoded with data, comprising: providing the polymer encoded with data comprising convertible residues iteratively spaced along and covalently linked via the backbone of the polymer, wherein a first subset of the convertible residues are in a first state and a second subset of the convertible residues are in a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; and passing the writable polymer encoded with data through a data reading device to read the encoded data on the polymer encoded with data.

**[000257]** In some embodiments, the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases. In some embodiments, the convertible residues in the first state can be converted into the second state via light. In some embodiments, the data reading device comprises a nanopore. In some embodiments, the data reading device is a sequencing device. In some embodiments, the sequencing device is a sequencing by synthesis device.

**[000258]** In some embodiments, the method further comprising measuring current flow of electrolytes during passage of the writable polymer.

**[000259]** In some embodiments, the method further comprising determining whether each of the plurality of convertible residues is in the first state or the second state based on the measured current flow of electrolytes during passage of the writable polymer.

**[000260]** In some embodiments, the method further comprising re-passing the polymer encoded with data through the data reading device to re-read the encoded data on the polymer encoded with data.

**[000261]** In some embodiments, the method further comprising validating and correcting the encoded data on the polymer encoded with data by comparing the encoded data on multiple copies of the polymer encoded with data.

**[000262]** In another aspect, also provided herein are methods for reading or decoding data from a nucleic acid polymer encoded with data, the method comprising:

providing a plurality of redundant copies of the nucleic acid polymer encoded with data comprising:

a plurality of converted nucleobases, wherein each converted nucleobase comprises a first nucleobase structure, wherein the first converted nucleobase has been converted from a first state into a second state, the first state and the second state being different; and

a plurality of convertible nucleobases, wherein each convertible nucleobase comprising a second nucleobase structure and a directly linked removable group, and wherein the convertible nucleobase is provided in a first state and is capable of being converted from the first state into a second state by releasing the second removable group from the second nucleobase structure, the first state and the second state being different;

wherein the converted nucleobases and convertible nucleobases are linked via the nucleic acid polymer backbone; and

sequencing each redundant copy of the plurality redundant copies of the nucleic acid polymer.

**[000263]** In some embodiments, the method further comprising detecting the plurality of converted nucleobases and the plurality of convertible nucleobases; and decoding the data based on the detected plurality of converted nucleobases.

**[000264]** In some embodiments, the plurality of converted nucleobases in the first state and the second state are readable by a polymerase enzyme. In some embodiments, the plurality of convertible nucleobases in the first state and the second state are readable by a polymerase enzyme. In some embodiments, the plurality of converted nucleobases and the plurality of convertible nucleobases are detected based on the sequencing result of the redundant copies of the nucleic acid polymer encoded with data.

**[000265]** In some embodiments, the sequencing starts on either end of the writable polymer (e.g. the 5' or 3' end of a nucleic acid polymer). In some embodiment, the sequencing starts on the 5' or the 3' end of the nucleic acid polymer. In some embodiment, the sequencing starts on the 5' end of the nucleic acid polymer. In some embodiment, the sequencing starts on the 3' end of the nucleic acid polymer

**[000266]** FIGS. 9A-9C illustrate an example of utilizing a device with a nanopore 501 for writing data into writable nucleic acid polymers 503. The device comprises a substrate 505 that includes a plasmonic nanostructure 507 for providing localized light energy to the writable polymer 503. The writable polymer 503 is controllably passed through a nanopore 501 at a steady rate. The nanopore may be comprised of protein or may be artificial, such as a pore engineered in silicon or other inorganic solid (see N Kono and K. Arakawa, *Dev Growth Differ.* 2019; 61:316-326; and Q Chen and Z. Liu, *Sensors (Basel)*. 2019; 19:1886; the disclosures of which are each incorporated herein by reference). Methods for constructing nanopores, and methods for a controlled rate of passage, have been previously described (see Y. Zhishan, et al., *Nanoscale Res Lett.* 2020; 15: 80, the disclosure of which is incorporated herein by reference). As the writable nucleic acid polymer 503 passes through the nanopore 501 at the controlled rate, the device selectively converts individual convertible nucleobases as they pass through the pore as encoded. As shown in FIG. 9B, a pulse of light 509 can be impinged on the convertible nucleobase via a plasmonic nanostructure 507 locally just as it passes through the pore, which can be appropriately timed due to the controlled rate of passage through the pore. As result of selective nucleobase conversion, binary digital data is encoded into the polymer (FIG. 9C).

**[000267]** FIGS. 10A-10C illustrate another example of utilizing a device with a nanopore 701 for encoding data into encodable nucleic acid polymers 703 comprising a plurality of sets of convertible nucleobases that are iteratively repeated along the polymer. The device comprises a substrate 705 that includes a plasmonic nanostructure 707 for providing localized light energy of multiple wavelengths to the data encodable polymer 703. The polymer 703 is controllably passed through a nanopore 701 at a steady rate. As the data encodable nucleic acid polymer 703 passes through the nanopore 701 at the controlled rate, the device selectively converts one or both convertible nucleobases of each set as the set passes through the pore, as prescribed by a data code. In this example, the data code to be encoded is 1001, where 1 is represented by  $C_a'$  and 0 is represented by  $C_a'C_b'$ . As shown in FIG. 10A, a pulse of light 709 at a first wavelength (e.g., 400 nm) can be impinged on a set via a plasmonic nanostructure 707 locally just as it passes through the pore, which results in conversion of a

single convertible base (as shown it converts base  $C_a$  into  $C_a'$ ). As shown in FIG. 10B, a pulse of light 711 at a second wavelength (e.g., 365 nm) can be impinged on the set via a plasmonic nanostructure 707 locally just as it passes through the pore, which results in conversion of a both convertible bases (as shown it converts bases  $C_a$  and  $C_b$  into  $C_a'$  and  $C_b'$ ). As result of selective nucleobase conversion, binary digital data is encoded into the polymer 703, which is encoded via sets with single nucleobase conversion 713 and sets of dual nucleobase conversion 715 (FIG. 10C).

**[000268]** FIGS. 11A-11C illustrate yet another example of utilizing a device with a nanopore 801 for encoding data into encodable nucleic acid polymers 803 comprising a plurality of two convertible nucleobase structures that are stochastically or irregularly repeated along the polymer. The device comprises a substrate 805 that includes a plasmonic nanostructure 807 for providing localized light energy of one or more wavelengths to the data encodable polymer 803. The polymer 803 is controllably passed through a nanopore 801 at a steady rate. As the data encodable nucleic acid polymer 803 passes through the nanopore 801 at the controlled rate, the device selectively converts one convertible nucleobase structure at a time, as prescribed by a data code. In this example, the data code to be encoded is 10110, where 1 is represented by  $C_a'$  and 0 is represented by  $C_b'$ . As shown in FIG. 11A, a pulse of light 809 can be impinged on a first nucleobase structure via a plasmonic nanostructure 807 locally just as it passes through the pore, which results in conversion of the nucleobase (as shown it converts base  $C_a$  into  $C_a'$ ). As shown in FIG. 11B, a pulse of light 809 can be impinged on a second nucleobase structure via a plasmonic nanostructure 807 locally just as it passes through the pore, which results in conversion of the nucleobase (as shown it converts base  $C_b$  into  $C_b'$ ). Further, as shown in FIGS. 11B and 11C, convertible bases 813, 815, and 817 are skipped, in accordance with the code. As result of selective nucleobase conversion, binary digital data is encoded into the polymer 803, which is encoded by converted nucleobases  $C_a'C_b'Ca'Ca'C_b'$  and skipping any convertible base in accordance with the data code.

**[000269]** Highly localized light excitation can be achieved via specialized sub-wavelength microscopic focusing strategies such as STEDX, or by the use of nanoplasmonic structures such as bow ties or by the use of zero-mode waveguides (see Y. Fang and M Sun, *Light Sci Appl.* 2015; 4:e294; and X. Shi, et al. *Small.* 2018; 14:e1703307; the disclosures of which are each incorporated herein by reference). If redox is to be used for nucleobase conversion, an applied potential of an electrode near or in a nanopore or nanochannel can be used. With a regular rate of passage, timed electronic pulses of voltage potential can result in appropriate spacing of nucleobase conversion. For enzymatic nucleobase conversion, the writable nucleic

acid polymer can be passed through two adjacent nanopores at a controlled rate; as a convertible nucleobase enters the volume between two pores, the enzyme is contacted (e.g. by microfluidics) with the strand at a local moiety/base/bit. Timing of microfluidic flow and controlled passage of the writable polymer can be in concert with appropriate spacing such that data is encoded with fidelity.

**[000270]** Several embodiments are also directed towards positive bit writing with dual bits. Accordingly, in certain embodiments, a writable nucleic acid polymer includes one or more repeated duads of convertible nucleobases, each convertible base of the duad is within the same field of resolution of the writing mechanism. In some embodiments, each convertible nucleobase of a duad is adjacent with other nucleobase of the duad. In some embodiments, each convertible nucleobase of a duad is near enough to the other nucleobase of the duad to be addressed in the same converting signal. In some embodiments, one convertible nucleobase of a duad has different reaction condition for nucleobase conversion than the other nucleobase of the duad. For example, in some embodiments, a first convertible nucleobase of a duad is converted by light at a first wavelength and a second convertible nucleobase of the duad is converted by light at a second wavelength. Thus, in certain embodiments of encoding a writable nucleic acid polymer comprising one or more duads, as each duad enters a nanopore, a particular reaction condition is provided to convert a first convertible nucleobase, or a second convertible nucleobase, or both the first and the second convertible nucleobases in accordance with a code.

**[000271]** FIGS. 12A-12C illustrate an example of utilizing a device with a nanopore 601 for writing data into writable nucleic acid polymers 603 comprising a plurality of duads. The device comprises a substrate 605 that includes a plasmonic nanostructure 607 for providing localized light energy of multiple wavelengths to the writable polymer 603. The writable polymer 603 is controllably passed through a nanopore 601 at a steady rate. As the writable nucleic acid polymer 603 passes through the nanopore 601 at the controlled rate, the device selectively converts individual convertible nucleobases of a duad as the duad pass through the pore as encoded. As shown in FIG. 12A, a pulse of light 609 at a first wavelength (e.g., 400 nm) can be impinged on the duad via a plasmonic nanostructure 607 locally as it passes through the pore, which results in conversion of a single convertible base (as shown it converts base  $W_a$  into  $W_a'$ ). As shown in FIG. 12B, a pulse of light 611 at a second wavelength (e.g., 325 nm) can be impinged on the duad via a plasmonic nanostructure 607 locally as it passes through the pore, which results in conversion of a both convertible bases (as shown it converts bases  $W_a$  and  $W_b$  into  $W_a'$  and  $W_b'$ ). As result of selective nucleobase

conversion, binary digital data is encoded into the polymer 603, which is encoded via duads with single nucleobase conversion 613 and duads of dual nucleobase conversion 615 (FIG. 12C). Examples of convertible nucleobases that are converted at specific wavelengths are provided in FIGS. 13A-13C.

**[000272]** In many embodiments, to read the data on written nucleic acid polymers, any appropriate sequencer capable of reading unnatural and/or altered nucleobases can be utilized. In certain embodiments, a device is capable of writing and reading nucleic acid polymers. In certain embodiments, a nanopore has dual functionality for both writing and reading nucleic acid polymers, however, some devices may include distinct nanopores for performing writing and reading. Examples of commercial nanopore sequencers include Oxford Nanopore Technologies PromethION, MinION, and GridION sequencing platforms (Oxford, UK) and Pacific Bioscience's Single Molecule, Real-Time (SMRT) sequencing platform (Menlo Park, CA). Alternatively, a nanopore device can be fabricated or manufactured for writing and/or reading the data. The nanopore can be comprised of solid-state materials, or can contain one or more proteins.

**[000273]** In many embodiments, to decode the data on encoded nucleic acid polymers, any appropriate sequencer capable of reading unnatural and/or altered nucleobases can be utilized. Examples of sequencing techniques used to decode DNA include (but are not limited to) shotgun sequencing, long-read sequencing, nanopore sequencing, and sequencing by synthesis.

**[000274]** In various embodiments in which the encoded data is to be decoded by sequencing by synthesis (SBS), it may be desirable to have a redundancy of data and thus the same data on each polymer of the plurality. For instance, when using a nucleobase structure such as O6-nitrobenzyl-guanine, the structure is read as a mix of A and G using SBS and thus a redundancy of reading the structure would be needed to interpret whether the structure is O6-nitrobenzyl-guanine, guanine, or adenine.

**[000275]** Provided in FIG. 14A is an example of utilizing a nanopore to read nucleobase sequences of convertible and converted nucleobases. In this example, O4-nitrobenzylthymine (T-4-ONB) is provided as the convertible base and removal of the nitrobenzyl group converts the nucleobases into a thymine. The current reading providing is differentiable between these two structures, as the microcurrent of T-4-ONB has low current and thymine has larger current. Although T-4-ONB is provided in this example, any convertible nucleobases in which an appreciable change in structure size and/or charge can be utilized, including (but not limited to) structures provided in FIGS. 4 and 5A-5B.

**[000276]** In certain embodiments, sequencing by synthesis (SBS) is performed to decode the data within a nucleic acid polymer, which may help in decoding between certain bases that have been converted and/or left unconverted. Standard SBS utilizes a polymerase to read a strand of the DNA sequence and make a complementary copy of the strand. The converted nucleobases should have the ability to serve as polymerase substrates and yield a predictable sequence result, enabling the polymerase to incorporate a base opposite and continue in the synthesis. For example, O6-nitrobenzylguanine (O6NBG) is contemplated as a convertible base, which is a suitable substrate for a DNA polymerase enzyme, thus enabling its reading by SBS. Sequencing of O6NBG nucleobase yields a reading that is a mixture of A and G nucleobases encoded at that position (see, e.g., A. M. Kietrys, W. A. Velema, and E. T. Kool, J Am Chem Soc. 2017; 139:17074-17081, the disclosure of which is incorporated herein by reference). When the nitrobenzyl group is removed to convert into a guanine structure, however, the sequencing reads will have a clear signal of G. When utilizing SBS, sequencing of multiple copies of encoded nucleic acid can help differentiate whether a nucleobase is a converted structure (e.g., guanine) or an unconverted structure (e.g., O6-nitrobenzylguanine) at a given position, thus indicating the presence of whether data has been encoded at that position. Notably, sequencing of multiple copies of encoded nucleic acid may be helpful in distinguishing several convertible/converted nucleobase structures, such as the structures provided in FIGS. 4 and 5A-5B.

**[000277]** Provided in FIG. 14B is an example of utilizing SBS to read nucleobase sequences of convertible and converted nucleobases. In this example, O4-nitrobenzylthymine (T-4-ONB) is provided as the convertible base and removal of the nitrobenzyl group converts the nucleobases into a thymine. SBS of T-4-ONB results in reading of a mixture of bases whereas the removal of the nitrobenzyl group results in a specific reading of thymine (see, e.g., A. M. Kietrys, W. A. Velema, and E. T. Kool, J Am Chem Soc. 2017; 139:17074-17081, the disclosure of which is incorporated herein by reference). Although T-4-ONB is provided in this example, any convertible nucleobases in which the sequencing readings changes as a result of the conversion can be utilized, including (but not limited to) structures provided in FIGS. 4 and 5A-5B.

### **Certain Embodiments**

**[000278]** Embodiment 1. A nucleic acid polymer for encoding data, comprising:

a plurality of pairs of convertible nucleobases, wherein the pairs are iteratively spaced along the nucleic acid polymer and each convertible nucleobase is linked via the nucleic acid polymer backbone,

wherein each convertible nucleobase of each pair comprises a nucleobase structure and a leaving group, the leaving group linked to the nucleobase structure via a linker, and wherein each convertible nucleobase of each pair is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the leaving group from the nucleobase structure.

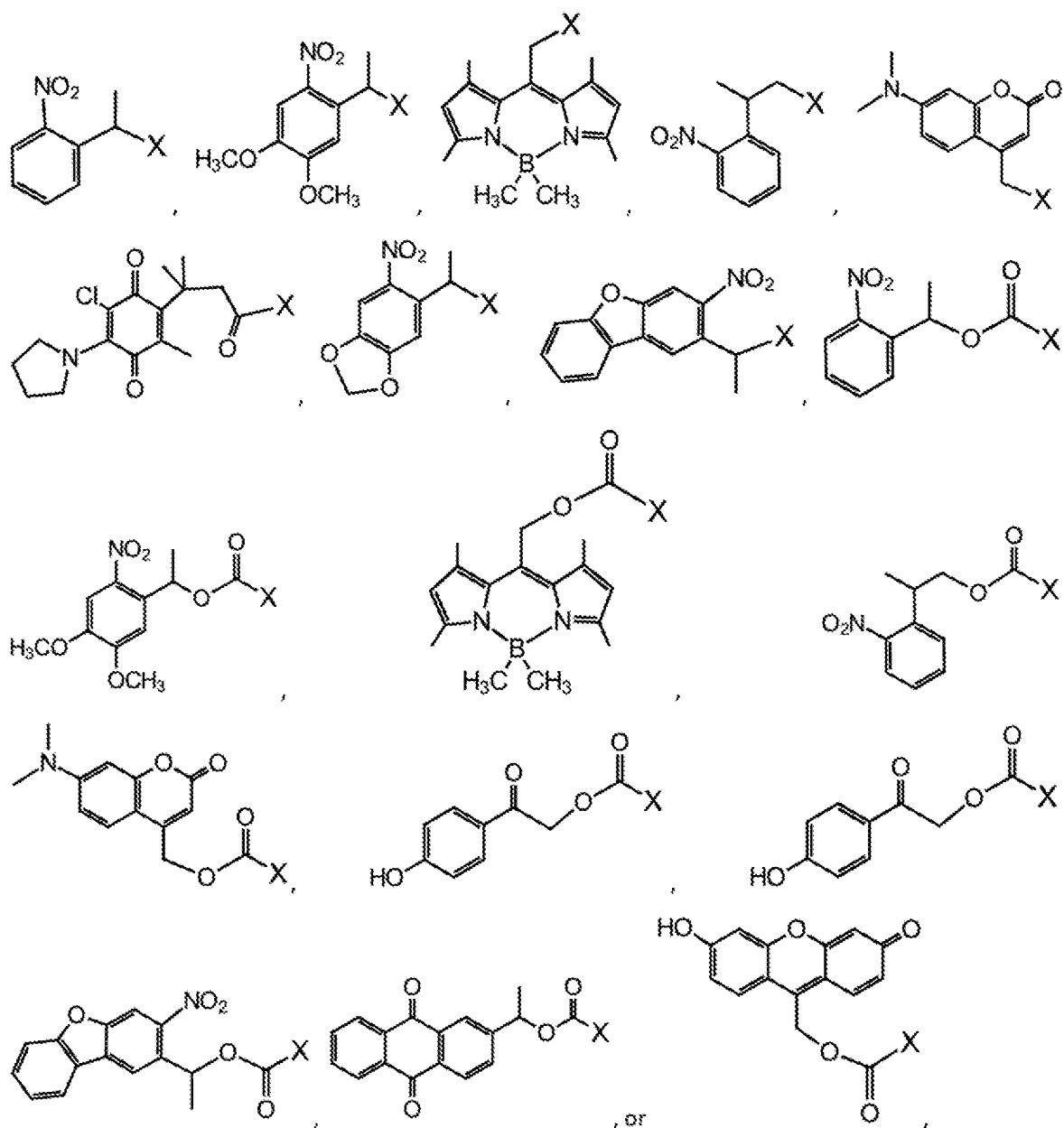
**[000279]** Embodiment 2. The nucleic acid polymer of Embodiment 1 further comprising a first plurality of sets of spacer residues, each spacer residue linked via the nucleic acid polymer backbone, wherein each set of the first plurality comprises two or more spacer residues, wherein each set of the first plurality is provided in-between each pair of the plurality of pairs of convertible nucleobases to provide the iterative spacing among the plurality of pairs of convertible nucleobases.

**[000280]** Embodiment 3. The nucleic acid polymer of Embodiment 2 further comprising a second plurality of sets of spacer residues, each spacer residue linked via the nucleic acid polymer backbone, wherein each set of the second plurality comprises one or more spacer residues, wherein each set of second plurality is provided in-between the convertible nucleobases in each pair of nucleobases, and wherein the number of spacer residues in each set of the second plurality is less than the number of spacer residues in each set of the first plurality.

**[000281]** Embodiment 4. The nucleic acid polymer of Embodiment 1 or 2., wherein the iterative spacing among the pairs of convertible nucleobases is equal to or greater than a resolution of a data encoding mechanism for encoding data into the nucleic acid polymer.

**[000282]** Embodiment 5. The nucleic acid polymer of any one of Embodiments 1-4, wherein each convertible nucleobase comprises one of the following nucleobase structures: O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

**[000283]** Embodiment 6. The nucleic acid polymer of any one of Embodiments 1-5, wherein the leaving group comprises one of:



wherein  $X$  is the linker to the nucleobase structure, wherein the linker is one of:  $NR_2$ ,  $NHR$ ,  $OR$ , or  $SR$ , and wherein  $R$  is the nucleobase structure.

**[000284]** Embodiment 7. The nucleic acid polymer of Embodiment 1, wherein light energy is used to release each leaving group, and wherein a first wavelength of light provides energy capable of converting a first convertible nucleobase of each pair into its second state, and wherein a second wavelength of light provides energy capable of converting a second convertible base of each pair into its second state.

**[000285]** Embodiment 8. The nucleic acid polymer of Embodiment 7, wherein the second wavelength of light provides energy that is further capable of converting the first convertible nucleobase of each pair into its second state.

**[000286]** Embodiment 9. A nucleic acid polymer for encoding data, comprising:

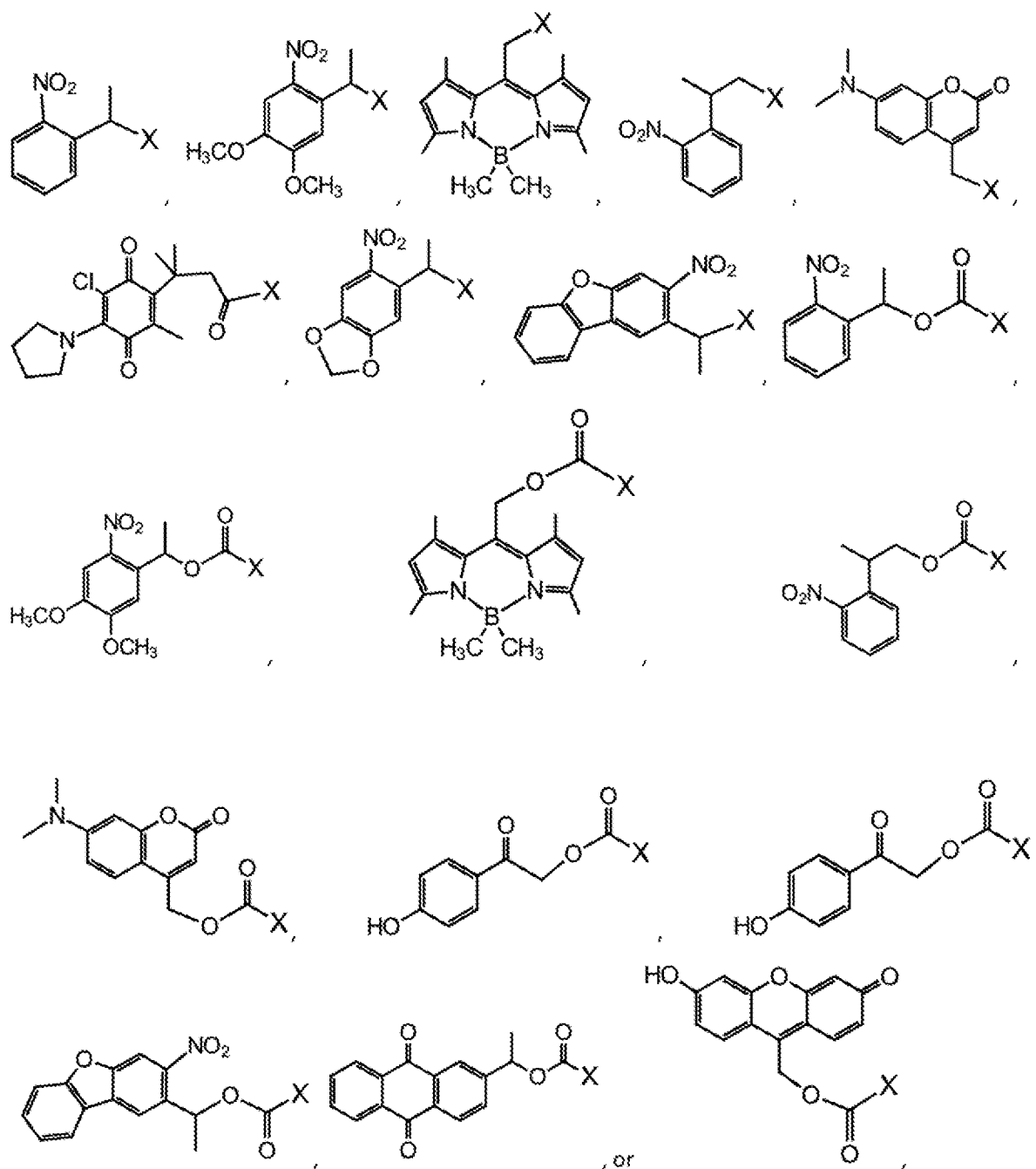
a first plurality of convertible nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the first plurality comprises a first nucleobase structure and a first leaving group, the first leaving group linked to the first nucleobase structure via a first linker, and wherein each convertible nucleobase of the first plurality is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the first leaving group from the first nucleobase structure; and

a second plurality of convertible nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the second plurality comprises a second nucleobase structure and a second leaving group, the second leaving group linked to the second nucleobase structure via a second linker, and wherein each convertible nucleobase of the first plurality is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the second leaving group from the second nucleobase structure.

**[000287]** Embodiment 10. The nucleic acid polymer of Embodiment 9 further comprising a plurality of spacer residues linked via the nucleic acid polymer backbone, wherein spacer residues are stochastically or irregularly provided in between the convertible nucleobases.

**[000288]** Embodiment 11. The nucleic acid polymer of Embodiment 9 or 10, wherein each convertible nucleobase comprises one of the following nucleobase structures: O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

**[000289]** Embodiment 12. The nucleic acid polymer of any one of Embodiments 9-11, wherein the leaving group comprises one of:



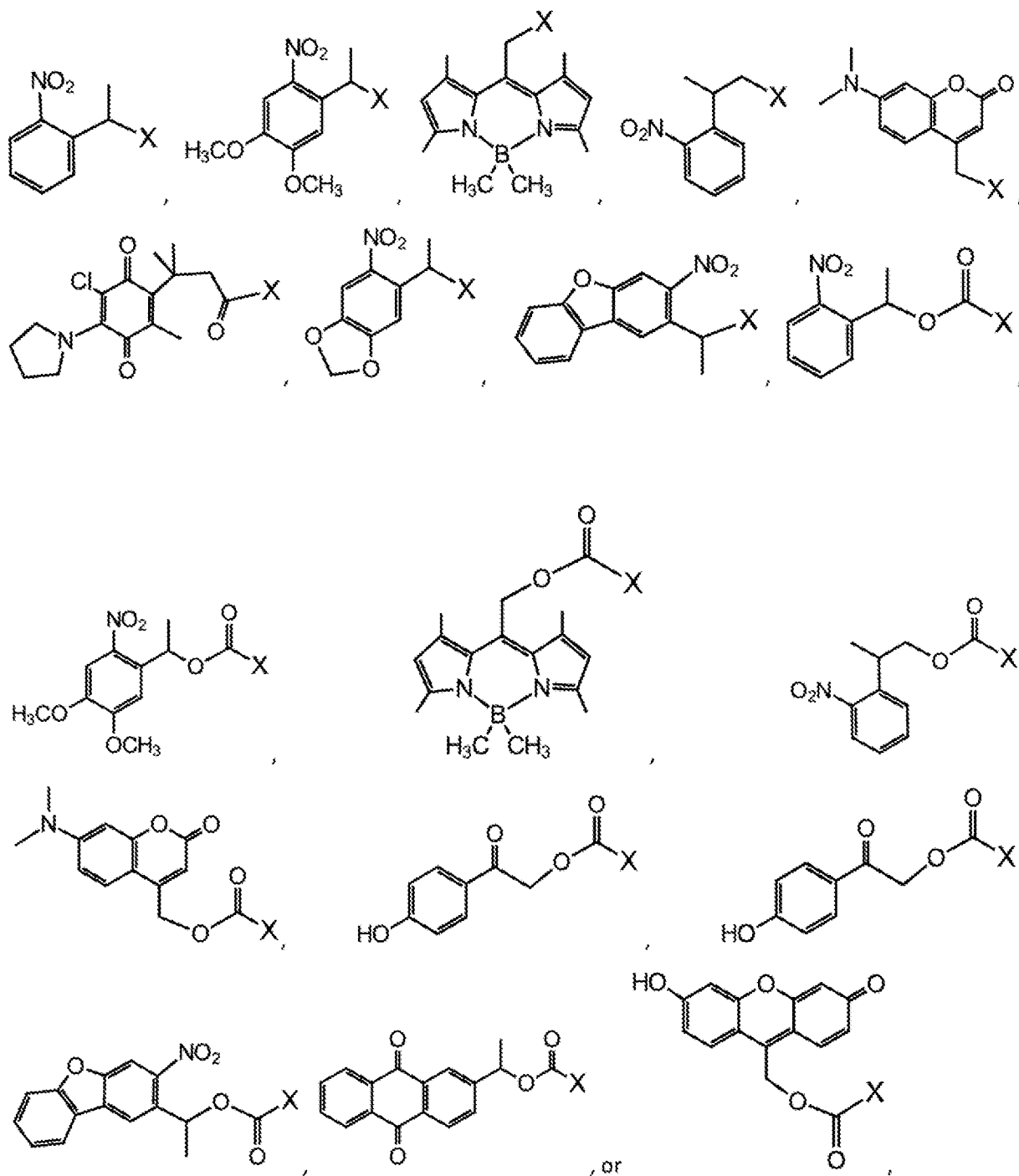
wherein X is the linker to the nucleobase structure, wherein the linker is one of: NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase structure.

**[000290]** Embodiment 13. A convertible nucleobase for use in a data encodable polymer, comprising: a nucleobase structure and a leaving group, wherein the leaving group is linked to the nucleobase structure via a linker, and wherein the leaving group is capable of being removed from the nucleobase structure by light energy or redox energy.

**[000291]** Embodiment 14. The convertible nucleobase of Embodiment 13, wherein the nucleobase structure comprises O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-

adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

**[000292]** Embodiment 15. The convertible nucleobase of Embodiment 13, wherein the leaving group comprises:



wherein X is a linker to the nucleobase structure, wherein the linker is one of: NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase structure.

**[000293]** Embodiment 16. The convertible nucleobase of Embodiment 15, wherein the linker comprises: NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase structure.

**[000294]** Embodiment 17. A data encoded nucleic acid polymer, comprising:

a plurality of pairs of nucleobases, wherein each pair of nucleobases comprises at least a first converted nucleobase, wherein the first converted nucleobase comprises a first nucleobase structure, wherein the first converted nucleobase has been converted from a first state into a second state by light energy or redox energy that released a first leaving group from the first nucleobase structure;

wherein each pair of nucleobases further comprises at least one of:

a convertible nucleobase that comprises a nucleobase structure and a second leaving group, the second leaving group linked to the second nucleobase structure via a linker, and wherein the convertible nucleobase is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the second leaving group from the second nucleobase structure; or

a second converted nucleobase, wherein the second converted nucleobase comprises a second nucleobase structure, wherein the second converted nucleobase has been converted from a first state into a second state by light energy or redox energy that released a second leaving group from the second nucleobase structure;

wherein the pairs of nucleobases are iteratively spaced along the nucleic acid polymer and the nucleobases are linked via the nucleic acid polymer backbone.

**[000295]** Embodiment 18. The nucleic acid polymer of Embodiment 17 further comprising a first plurality of sets of spacer residues, each spacer residue linked via the nucleic acid polymer backbone, wherein each set of the first plurality comprises two or more spacer residues, wherein each set of the first plurality is provided in-between each pair of the plurality of pairs of nucleobases to provide the iterative spacing among the plurality of pairs of nucleobases.

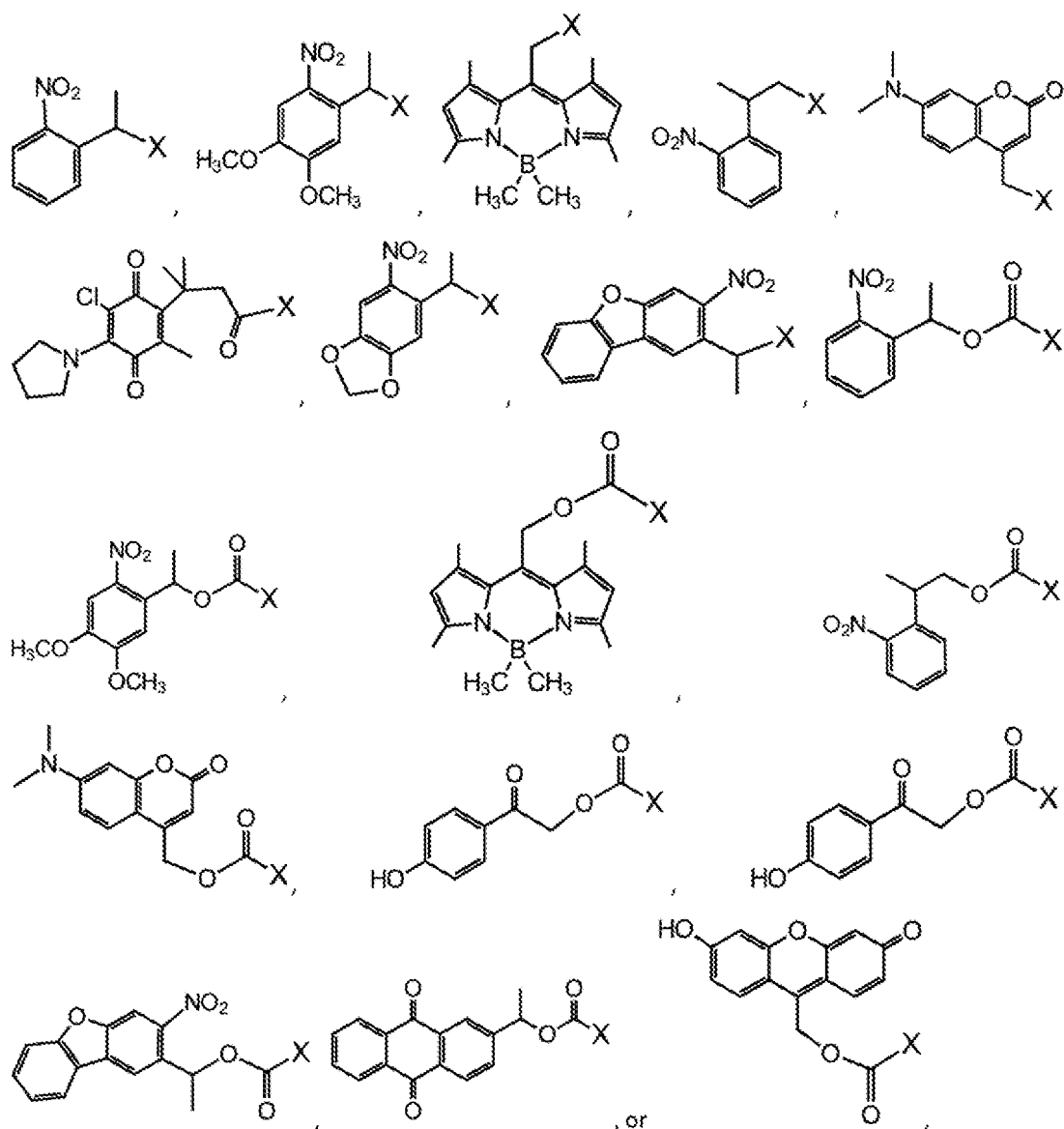
**[000296]** Embodiment 19. The nucleic acid polymer of Embodiment 18 further comprising a second plurality of sets of spacer residues, each spacer residue linked via the nucleic acid polymer backbone, wherein each set of the second plurality comprises one or more spacer residues, wherein each set of second plurality is provided in-between the convertible nucleobases in each pair of nucleobases, and wherein the number of spacer residues in each set of the second plurality is less than the number of spacer residues in each set of the first plurality.

[000297] Embodiment 20. The nucleic acid polymer of Embodiment 17 or 18, wherein the iterative spacing among the pairs of nucleobases is equal to or greater than a resolution of a data encoding mechanism used to encode data into the data encoded nucleic acid polymer.

[000298] Embodiment 21. The nucleic acid polymer of any one of Embodiments 14-20, wherein each converted nucleobase has one following nucleobase structures: guanine, adenine, thymine, or cytosine.

[000299] Embodiment 22. The nucleic acid polymer of any one of Embodiments 14-21, wherein each convertible nucleobase comprises one of the following nucleobase structures: O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

[000300] Embodiment 23. The nucleic acid polymer of any one of Embodiments 14-22, wherein the second leaving group of each convertible nucleobase comprises one of:



wherein X is a linker to the nucleobase structure, wherein the linker is one of: NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase structure.

**[000301]** Embodiment 24. A data encoded nucleic acid polymer, comprising:

a first plurality of converted nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via a nucleic acid polymer backbone, wherein each converted nucleobase of the first plurality comprises a first nucleobase structure, wherein each converted nucleobase of the first plurality has been converted from a first state into a second state by light energy or redox energy that released a first leaving group from the first nucleobase structure; and

a second plurality of converted nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via a nucleic acid polymer backbone, wherein each converted nucleobase of the second plurality comprises a second nucleobase structure, wherein each converted nucleobase of the second plurality has been converted from a first state into a second state by light energy or redox energy that released a second leaving group from the second nucleobase structure.

**[000302]** Embodiment 25. The data encoded nucleic acid polymer of Embodiment 24, further comprising:

a first plurality of convertible nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the first plurality comprises the first nucleobase structure and the first leaving group, wherein the first leaving group is linked to the first nucleobase structure via a first linker; and

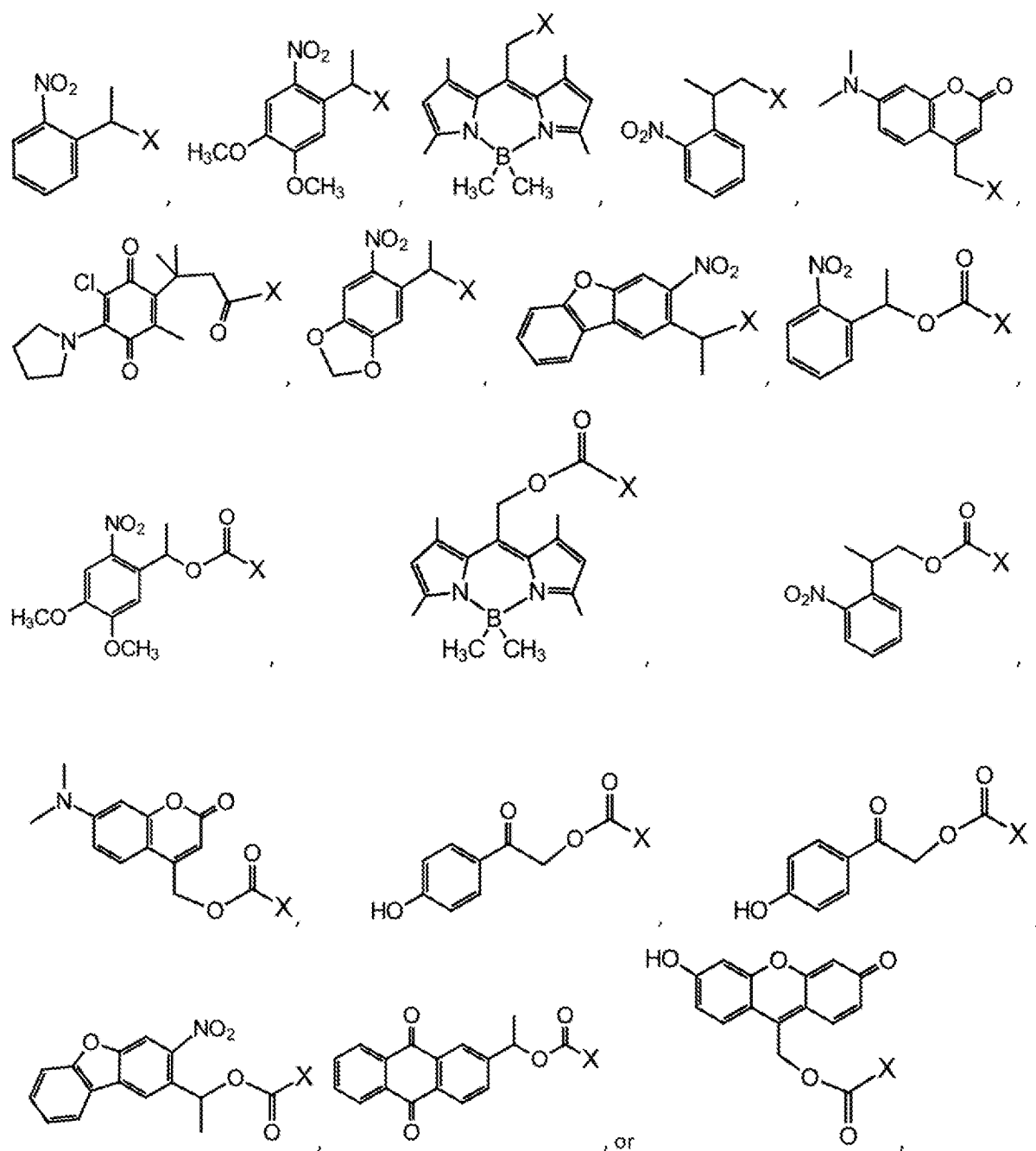
a second plurality of convertible nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the second plurality comprises the second nucleobase structure and the second leaving group, wherein the second leaving group is linked to the second nucleobase structure via a second linker.

**[000303]** Embodiment 26. The nucleic acid polymer of Embodiment 25 further comprising a plurality of spacer residues linked via the nucleic acid polymer backbone, wherein spacer residues are stochastically or irregularly provided in between nucleobases comprising the converted and convertible nucleobases.

**[000304]** Embodiment 27. The nucleic acid polymer of any one of Embodiment 24 to 26, wherein each converted nucleobase has one following nucleobase structures: guanine, adenine, thymine, or cytosine.

**[000305]** Embodiment 28. The nucleic acid polymer of any one of Embodiments 25 to 27, wherein each convertible nucleobase comprises one of the following nucleobase structures: O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.

**[000306]** Embodiment 29. The nucleic acid polymer of any one of Embodiments 25 to 28, wherein the leaving group of each convertible nucleobase comprises one of:



**[000307]** Embodiment 30. A method of encoding data onto a data encodable nucleic acid polymer, comprising:

providing a data encodable nucleic acid polymer that comprises: a plurality of pairs of convertible nucleobases, wherein the pairs are iteratively spaced along the nucleic acid polymer and each convertible nucleobase is linked via the nucleic acid polymer backbone,

wherein each convertible nucleobase of each pair comprises a nucleobase structure and a leaving group, the leaving group linked to the nucleobase structure via a linker, and wherein each convertible nucleobase of each pair is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the leaving group from the nucleobase structure; and

selectively converting, utilizing a data encoding device, at least one nucleobase of each pair of convertible nucleobases into the second state by providing a light energy or redox energy to release the leaving group from the nucleobase structure of the at least one nucleobase.

**[000308]** Embodiment 31. The method of Embodiment 30, wherein the data encoding device comprises a plasmonic nanopore, and the method further comprising: passing the data encodable nucleic acid polymer through the plasmonic nanopore of the data encoding device, wherein the plasmonic nanopore provides the light energy or redox energy to release the leaving group from the nucleobase structure of the at least one nucleobase.

**[000309]** Embodiment 32. The method of Embodiment 31, wherein the data encodable nucleic acid polymer further comprises a first plurality of sets of spacer residues, each spacer residue linked via the nucleic acid polymer backbone, wherein each set of the first plurality comprises two or more spacer residues, wherein each set of the first plurality is provided in-between each pair of the plurality of pairs of convertible nucleobases to provide the iterative spacing among the plurality of pairs of convertible nucleobases.

**[000310]** Embodiment 33. The method of Embodiment 31 or 32, wherein the iterative spacing among the pairs of convertible nucleobases is equal to or greater than the resolution of the data encoding device.

**[000311]** Embodiment 34. The method of Embodiment 30, wherein the data encoding device comprises a plasmonic well or channel, and the method further comprising: transfer the data encodable nucleic acid polymer into the plasmonic well or channel of the data encoding device, wherein the plasmonic well or channel provides the light energy or redox energy to release the leaving group from the nucleobase structure of the at least one nucleobase.

**[000312]** Embodiment 35. The method of Embodiment 30, wherein the data encoding device comprises a STED laser system, and the method further comprising: stretching the data encodable nucleic acid polymer and focusing the STED laser onto the stretched data encodable nucleic acid polymer, wherein the STED laser provides the light energy or redox energy to release the leaving group from the nucleobase structure of the at least one nucleobase.

**[000313]** Embodiment 36. A method of encoding data onto a data encodable nucleic acid polymer, comprising:

providing a data encodable nucleic acid polymer that comprises:

a first plurality convertible nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the first plurality comprises a first nucleobase structure and a first leaving group, the first leaving group linked to the first nucleobase structure via a first linker, and wherein each convertible nucleobase of the first plurality is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the first leaving group from the first nucleobase structure; and

a second plurality of convertible nucleobases stochastically or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the second plurality comprises a second nucleobase structure and a second leaving group, the second leaving group linked to the second nucleobase structure via a second linker, and wherein each convertible nucleobase of the first plurality is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the second leaving group from the second nucleobase structure; and

selectively converting, utilizing a data encoding device, a subset of the convertible nucleobases of the first plurality and the second plurality into the second state by providing a light energy or redox energy to release the leaving group from the nucleobase structure of the convertible nucleobases.

**[000314]** Embodiment 37. The method of Embodiment 36, wherein the subset of the convertible nucleobases of the first plurality and the second plurality that are selectively converted are based on a data code to be encoded.

**[000315]** Embodiment 38. The method of Embodiment 37, wherein the selective conversion of nucleobases yields a nucleic acid polymer comprising convertible nucleobases in between converted nucleobases.

**[000316]** Embodiment 39. The method of Embodiment 36, wherein the data encoding device comprises a plasmonic nanopore, and the method further comprising:

passing the data encodable nucleic acid polymer through the plasmonic nanopore of the data encoding device, wherein the plasmonic nanopore provides the light energy or redox energy to release the leaving group from the nucleobase structure of the convertible nucleobases.

**[000317]** Embodiment 40. The method of Embodiment 30, wherein the data encoding device comprises a plasmonic well or channel, and the method further comprising:

**[000318]** transfer the data encodable nucleic acid polymer into the plasmonic well or channel of the data encoding device, wherein the plasmonic well or channel provides the light energy or redox energy to release the leaving group from the nucleobase structure of the convertible nucleobases.

**[000319]** Embodiment 41. The method of Embodiment 30, wherein the data encoding device comprises a STED laser system, and the method further comprising:

stretching the data encodable nucleic acid polymer and focusing the STED laser energy onto the stretched data encodable nucleic acid polymer, wherein the STED laser provides the light energy or redox energy to release the leaving group from the nucleobase structure of the convertible nucleobases.

**[000320]** Embodiment 42. A method for decoding data from a data encoded nucleic acid polymer, the method comprising:

providing a plurality of redundant copies of a data encoded nucleic acid polymer that comprises:

a plurality of converted nucleobases, wherein each converted nucleobase comprises first nucleobase structure, wherein the first converted nucleobase has been converted from a first state into a second state by light energy or redox energy that released a first leaving group from the first nucleobase structure; and

a plurality of convertible nucleobases, wherein each convertible nucleobase comprises a nucleobase structure and a leaving group, the leaving group linked to the second nucleobase structure via a linker, and wherein the convertible nucleobase is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the second leaving group from the second nucleobase structure

wherein the converted nucleobases and convertible nucleobases are linked via the nucleic acid polymer backbone; and

sequencing each redundant copy of the plurality of redundant copies;  
detecting the plurality converted nucleobases and the plurality of convertible nucleobases;

decoding the data based on the detected plurality of converted nucleobases.

**[000321]** Embodiment 43. The method of Embodiment 42, wherein the plurality of the plurality converted nucleobases and the plurality of convertible nucleobases are detected based on the sequencing result of the redundant copies of the data encoded nucleic acid polymer.

**[000322]** Embodiment 44. The method of Embodiment 43, wherein a sequencing result indicating a mix of nucleobase structures at a particular nucleobase indicates a convertible nucleobase that is not a part of the data code.

**[000323]** Embodiment 45. A nucleic acid polymer for encoding data, comprising:

a first plurality convertible nucleobases regularly or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the first plurality comprises a first nucleobase structure and a first leaving group, the first leaving group linked to the first nucleobase structure via a first linker, and wherein each convertible nucleobase of the first plurality is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the first leaving group from the first nucleobase structure; and

a second plurality of convertible nucleobases regularly or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the second plurality comprises a second nucleobase structure and a second leaving group, the second leaving group linked to the second nucleobase structure via a second linker, and wherein each convertible nucleobase of the first plurality is provided in a first state and is capable of being converted from the first state into a second state by light energy or redox energy that releases the second leaving group from the second nucleobase structure.

**[000324]** Embodiment 46. The nucleic acid polymer of Embodiment 45 further comprising a plurality of spacer residues linked via the nucleic acid polymer backbone, wherein spacer residues are provided in between the convertible nucleobases.

**[000325]** Embodiment 47. A data encoded nucleic acid polymer, comprising:

a first plurality of converted nucleobases regularly or irregularly spaced along the nucleic acid polymer and linked via a nucleic acid polymer backbone, wherein each converted nucleobase of the first plurality comprises a first nucleobase structure, wherein

each converted nucleobase of the first plurality has been converted from a first state into a second state by light energy or redox energy that released a first leaving group from the first nucleobase structure; and

a second plurality of converted nucleobases regularly or irregularly spaced along the nucleic acid polymer and linked via a nucleic acid polymer backbone, wherein each converted nucleobase of the second plurality comprises a second nucleobase structure, wherein each converted nucleobase of the second plurality has been converted from a first state into a second state by light energy or redox energy that released a second leaving group from the second nucleobase structure.

**[000326]** Embodiment 48. The data encoded nucleic acid polymer of Embodiment 47, further comprising:

a first plurality of convertible nucleobases regularly or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the first plurality comprises the first nucleobase structure and the first leaving group, wherein the first leaving group is linked to the first nucleobase structure via a first linker; and

a second plurality of convertible nucleobases regularly or irregularly spaced along the nucleic acid polymer and linked via the nucleic acid polymer backbone, wherein each convertible nucleobase of the second plurality comprises the second nucleobase structure and the second leaving group, wherein the second leaving group is linked to the second nucleobase structure via a second linker.

**[000327]** Embodiment 49. The nucleic acid polymer of Embodiment 48 further comprising a plurality of spacer residues linked via the nucleic acid polymer backbone, wherein spacer residues are provided in between nucleobases comprising the converted and convertible nucleobases.

## **EXEMPLARY EMBODIMENTS**

**[000328]** Described herein are various examples of compositions, systems, and methods for data storage utilizing nucleic acid polymers. Examples of writable nucleic acid polymers, methods to produce such polymers, methods to writing data, and methods for reading data are provided.

**Example 1: Writable DNA polymer with MeNPOC nucleobases**

**[000329]** A writable nucleic acid molecule can be generated to comprise bits, data fields, spacers, delimiters, and/or a terminal identifier tag. In this example, a converted nucleobase (i.e., “1”) is 5-aminopropynyl-deoxyuridine, and an unconverted nucleobase (i.e., “0”) is the same molecule with the amine group substituted by a MeNPOC group, which can be efficiently removed by light (see P. Klan, et al., Chem Rev. 2013; 113:119-91, the disclosure of which is incorporated herein by reference). The writable nucleic acid is constructed with all convertible nucleobases having an MeNPOC-substituted deoxyuridine base, which is denoted “0” in the following example:

Data field: 5’-C-(A)<sub>6</sub>-0-(A)<sub>6</sub>-0-(A)<sub>6</sub>-0-(A)<sub>6</sub>-0-(A)<sub>6</sub>-0-(A)<sub>6</sub>-0-(A)<sub>6</sub>-0-(A)<sub>6</sub>-(C)-3’

**[000330]** The data field contains “0” bits spaced by six adenine nucleotides (A) to allow for spatial resolution for writing via focused light energy. It is shown here with eight bits (one “byte” in 8-bit architecture). The cytosines at the ends can provide a data delimiter function, signifying a break between one 8-bit field and the next. It is understood that spacers and delimiter are not limited to adenosines and cytidines and could be almost any single or multiple natural or unnatural residue that is detectably different from the convertible nucleobases and, preferably, is unreactive to the writing mechanism. It is also understood that a delimiter may not be needed to achieve efficient data encoding. In such a case, a writable nucleic acid contains repeating bits and spacers that are not contained within delimiters. It is also understood that the spacing and number of spacers between bits can be readily altered to reflect the resolution and precision of the writing method.

**[000331]** The writable nucleic acid polymer consists of the data field sequence repeated in a string. The polymer can be tagged at the 5’ or 3’ end by a data tag. This can comprise a sequence of natural bases that denote time, date, type of data, user, or other useful identifying information. It is understood that a data tag may not be necessary for some applications, as identifying information can be written directly into the data fields.

**Example 2: Writable nucleic acid polymer produced by rolling circle reaction**

**[000332]** In this example, a circular DNA oligonucleotide encoding the repeating “data field” in example 1 as described. The circle is chosen to be complementary to the repeating unit, and is chosen in this case to be 57 nucleotides in size, which falls in a size range that is known to act as a good substrate for DNA polymerase-mediated rolling circle synthesis (see M. G. Mohsen and E. T. Kool, Acc Chem Res. 2016 Nov 15; 49(11): 2540–2550, the disclosure of which is incorporated herein by reference). The circle sequence is as follows:

5'-GTTTTTTATTTTTTATTTTTTATTTTTTATTTTTTATTTTTTATTTTTTATTTTTTG-  
3' where the 5' and 3' ends are joined intramolecularly to make a circle.

**[000333]** A DNA primer is constructed with a 3' end complementary to the circle. An example of an effective primer sequence is below:

Primer: 5'-IDsequence-AAAAAATAAAAAACCAAAAAA-3'

**[000334]** The ID sequence is optional. The DNA primer is annealed to the DNA circle in a  $Mg^{2+}$ -containing buffer that supports DNA polymerase activity. The mixture is contacted with nucleoside triphosphates (dNTPs) that will comprise the repeating data field. For the data field in example 1, the necessary dNTPs are 5-nitroveratryl—oxycarbonyl-aminoproynyl deoxyuridine 5'-triphosphate, dATP, and dCTP. Contacting this solution with a suitable DNA polymerase enzyme at a temperature supportive of enzyme activity produces a long repeating writable DNA polymer, comprising repeating data fields, and a DNA data identifier tag at the 5' end. Gel analysis shows that the blank tape is 10,000 to 50,000 nucleotides in length. It is isolated from the smaller polymerase, nucleotides, and circle by size exclusion chromatography, column purification, precipitation, gel electrophoresis, or by other purification methods, and is stored in the dark to avoid stray bit writing.

**[000335]** Various DNA polymerase enzymes for rolling circle synthesis have been described (see S. Ishino and Y. Ishino, *Front Microbiol.* 2014; 5:465, the disclosure of which is incorporated herein by reference). Examples include phi29 and BST3.0 polymerases. A polymerase with high processivity enables longer writable DNA polymers to be produced. A polymerase with the ability to efficiently accept modified nucleotides (such as the modified deoxyuridine described here) as substrates can be used.

### Example 3: Writable nucleic acid polymer produced by synthesis and ligation

**[000336]** In this example, a ligase enzyme is used to assemble single-stranded and/or double-stranded writable DNA polymers containing the convertible nucleobase O6-ortho-nitrobenzylG (see FIG. 3D, denoted X here), which is not efficiently incorporated into DNA by most polymerase enzymes due to its blocked base pairing ability. The designed 8-bit repeating data field sequence is the following:

5'-CCT-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-CGA-3'

**[000337]** A ligatable oligonucleotide comprising the single 8-bit field is synthesized with the following sequence:

5'-pCCT-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-X-(A)6-(CGA)-3'

Where “p” denotes a terminal phosphate group. A splint for ligating this sequence is synthesized with the following sequence:

5'-TTTTTTAGGTCGTTTTTT-3'

**[000338]** Contacting this splint and the data field oligonucleotide with T4 DNA ligase and ATP in a ligase-supporting buffer results in joining of many data field oligomers end-to-end, resulting in a long polymer strand. Gel analysis of this product reveals a ladder of lengths ranging from 5000-50,000 nucleotides in size. If desired, portions of the “data field” DNA product can be split up and ligated at one end separately with different DNA identifiers, to be used separately in data writing. The long data fields are used for writing as a mixture of lengths. Alternatively, use of an electrophoresis gel and cutting out and eluting a specific band results in a blank tape DNA of homogeneous length.

**[000339]** A double-stranded writable DNA polymer is obtained by similar methods. In this case, the first data field oligonucleotide is also employed, but a different complement is used in the formation of a duplex with sticky ends. The sequence of this complementary oligonucleotide is as follows:

5'-

pGTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTT  
TTTTAGGTC-3'

**[000340]** Hybridization of the complementary oligonucleotide with the data field oligonucleotide results in a duplex with sticky ends. Ligation with T4 DNA ligase and ATP results in a long repeating DNA double-stranded polymer. Gel analysis of this product reveals a ladder of lengths ranging from 5000-50,000 base pairs in size. If desired, portions of the data field DNA product can be split up and ligated at one end separately with different DNA identifiers, to be used separately in data writing. The long data fields are used for writing as a mixture of lengths. Alternatively, use of an electrophoresis gel and cutting out and eluting a specific band results in a blank tape DNA of homogeneous length.

### Example 4: Data writing via light

**[000341]** A nanopore device with a plasmonic bow tie on the exit side of the pore is used to write digital data on the writable DNA polymer from example 1. Nanopores with plasmonic bow ties have been described (see X. Shi, et al., *Small*. 2018 May;14(18):e1703307, the disclosure of which is incorporated herein by reference). The writable polymer is dissolved in an electrolyte solution and is moved through the pore at a regular rate via applied potential across the two sides of the pore. The test bit sequence “01100101” is written repeatedly. This

is achieved by flashing a beam of light on the nanoplasmonic structure at spaced time intervals to coincide with the bit spacing in a data field. Subsequent analysis by nanopore sequencing then reveals the sequence of “1” and “0” bits, and the repetition allows the analysis of the precision and errors in bit writing. Statistical analysis and data correction on the repeat units in the sequence confirms the intended bit sequence. Subsequent experiments with longer data strings reveal the ability to encode more data per molecule. Comparison of multiple copies of DNA tapes written with the same data enables sequence comparison and error correction.

#### **Example 5: Writing data via DNA stretching and light**

**[000342]** In this example, data is encoded in the double-stranded writable DNA polymer from Example 3 by DNA stretching or combing, combined with local illumination to write bits. In the stretching/combing technique, flow is used to stretch individual DNA molecules with lengths of tens of thousands of nucleotides on a slide or other solid support, and the locations of the long DNAs are visualized by simple dyes added to solution (see T. F. Chan, et al., *Nucleic Acids Res.* 2006; 34:e113; and S Takahashi, M. Oshige, and S. Katsura, *Molecules.* 2021; 26:1050; the disclosures of which are each incorporated herein by reference). Light is focused progressively along the strand at intended “1” sites along the strand to convert nucleobase bits from the “0” state to the “1” state. The light illumination is achieved at high resolution by the use of the STED technique, which uses two lasers to illuminate locally with high precision (see G. Vicidomini, P. Bianchini, and A. Diaspro, *Nat Methods.* 201; 15:173-182, the disclosure of which is incorporated herein by reference).

**[000343]** The resulting written DNAs can be stored for archiving. When the data is to be retrieved, the stored data can be read by nanopore sequencing of the DNA polymer (see Example 7).

**[000344]** In another embodiment, the bit nucleotide comprises a fluorescent dye linked by a photocleavable linker to a fluorescence quencher. The presence of the quencher keeps the unwritten DNA nonfluorescent. “Localized illumination” of the “stretched DNA” strand results in cleavage of the linker, resulting in loss of the quencher, rendering the local nucleotide fluorescent. Progression of the photoexciting light along the stretched data field DNA results in writing bits at data-encoding intervals. The slide is stored as written data. When the data is to be retrieved, it is read by imaging the strand on the slide and analyzing the “1” bits as fluorescent spots; the spacing denotes the presence and numbers of intervening “0” bits.

**Example 6: Writing data via redox**

[000345] This example describes the writing of data by redox with writable DNA polymers comprising the redox-reactive nucleotide in FIG. 3G. In this experiment, a nanopore device with an electrode at the pore is employed. A DNA blank tape containing redox-reactive nucleobases is passed through the pore at a controlled rate. As the DNA passes through, reductive voltage potential is applied as a pulse at timed spacing. This results in reduction and loss of the group on the “0” bit, switching it to the aminopropyne group which encodes “1”. Spacing in time of applied reduction results in variable but predictable spacing of “1” and “0” groups, which defines the digital data.

**Example 7: Reading written DNA polymers via nanopore sequencing**

[000346] Common nanopore sequencing devices measure current flow of electrolytes during passage of a DNA molecule through the pore. Since DNA bases each differ in size and shape, this slightly alters the current as each different base passes the pore. In this example, an experiment is carried out with a commercial nanopore device, and the readout changes in current over time while a written DNA tape passes through. In this case, the single-stranded written DNA polymer produced in Example 3 and written as in Example 4 is employed. The “1” and “0” bits comprise G and nitrobenzylG, which differ considerably in size. Experiments with DNA tapes having bits in all-“0” state (blank polymer) reveal the lowering of current when the largest nitrobenzylG nucleotides pass through, and can distinguish the differences in current between these “0” bits and the spacers and delimiters. Separately, DNA all-“1” polymers are measured, showing the level of current observed as the “1” (G) bits pass through. These experiments provide calibration for reading and distinguishing current levels that denote “1” and “0” bits. Next, fully written DNA polymers are passed through. Current levels denoting “1” and “0” are read and placed in context of current levels seen for spacers and delimiters. Multiple reads of the same strand are used, if needed, to improve accuracy of data reading.

**Example 8. Dual bit writable nucleic acid polymers**

[000347] This example provides a writable nucleic acid polymer design that enables the writing of both “1” and “0” bits with an active signal. In this design, zeros are not passively included in the data field, but rather require an active switching signal. Photo removable groups can be triggered at distinct wavelengths of light. FIGS. 13A-13C show examples of

nucleotides comprising a group that can be removed by irradiation at 325 nm, and a different group that can be removed at 400 nm irradiation. If these two groups are placed near one another in a data field of the blank DNA tape, a light pulse at 400nm removes only one of the two groups in the pair. On the other hand, a light pulse at 325 nm results in loss of both of the groups. These two outcomes are akin to “0” and “1” for encoding data.

### Example 9: Construction of data encodable DNA

**[000348]** A 141nt DNA strand is synthesized to contain pairs of iteratively repeating convertible nucleobases (X and Y) separated by two spacer nucleobases, with each pair representing a bit of encodable data. Each pair of nucleobases is separated with ten intervening spacer nucleobases. The total number of pairs in the strand is 11, and thus the DNA can encode 11 bits of "one" and "zero" data. The sequence of this 150mer is:

5'-

TCGATTXAYAATTATTCCTXAYAATTATTCCTXAYAATTATTCCTXAYAATTATTC  
CTXAYAATTATTCCTXAYAATTATTCCTXAYAATTATTCCTXAYAATTATTCCTXA  
YAATTATTCCTXAYTTTATCTTATXAYTCGA - 3' = 141

where X denotes O6-nitrobenzylguanine and Y denotes N6-coumarinylmethyl-adenine.

**[000349]** A complementary DNA sequence is synthesized to be complementary to the first strand such that a duplex can be formed. The complementary sequence can be designed to create overhanging sticky ends, and the two strands are further modified with 5' phosphate groups. The sequence of this 141mer is:

5'-

TCGATTCATAAGATAAATTCAGGAATAATTTTCAGGAATAATTTTCAGGAATAAT  
TTTCAGGAATAATTTTCAGGAATAATTTTCAGGAATAATTTTCAGGAATAATTTTC  
AGGAATAATTTTCAGGAATAATTTTCAATCGA - 3'

**[000350]** Note that the bases in this complement are designed to be complementary to the converted versions of bases X and Y. Longer DNAs can store more data per molecule. To generate longer nucleic acid polymers for data storage, the two DNA strands can be mixed in a  $Mg^{2+}$ -containing buffer that supports hybridization and enzymatic ligation. ATP and T4 DNA ligase are added, resulting in end-to-end joining of the 150nt DNAs into longer polymer chains, having lengths of  $\sim 300$ bp and more, including DNAs of  $\sim 1500$ bp as analyzed by agarose gel electrophoresis. Data encodable DNAs of preferred size can be isolated by gel electrophoresis and extracted. Accordingly, data encodable polymers can be

provided and utilized as a mixture of lengths or having specific lengths by excising specific bands.

### Example 10: Data encoding into a polymer

**[000351]** A nanopore device with a plasmonic bow tie on the exit side of the pore is used to write digital data on the data encodable DNA polymer from Example 9. Nanopores with plasmonic bowties have been described (see X. Shi, et al., Small. 2018 May;14(18):e1703307, the disclosure of which is incorporated herein by reference). The data encodable polymer is dissolved in an electrolyte solution and is moved through the pore at a regular rate via applied potential across the two sides of the pore. The data sequence “01100101100” is encoded in the polymer (for the first 150 nucleotides). This is achieved by flashing a beam of light on the nanoplasmonic structure at spaced time intervals to coincide with the paired bit spacing.

**[000352]** To encode a bit of data, light energy can be provided by 400 nm wavelength onto the bit pair to release the coumarinylmethyl group from the N6-coumarinylmethyl-adenine to convert the nucleobase into an adenine. The light energy at 400 nm does not affect the O6-nitrobenzylguanine, leaving the nucleobase unconverted. This bit pair conversion can be denoted a “zero.” Likewise, light energy can be provided by 365 nm wavelength onto the bit pair to release the nitrobenzyl group from the O6-nitrobenzylguanine to convert the nucleobase into a guanine and to release the coumarinylmethyl group from the N6-coumarinylmethyl-adenine to convert the nucleobase into an adenine. This bit pair conversion can be denoted a “one.” Data encoding can continue to yield the data sequence “01100101100,” which structurally would have the following nucleobase sequence:

5'-

TCGATTXAAAATTATTCCTGAAAATTATTCCTGAAAATTATTCCTXAAAATTATTC  
CTXAAAATTATTCCTGAAAATTATTCCTXAAAATTATTCCTGAAAATTATTCCTGA  
AAATTATTCCTXAATTTATCTTATXAATCGA - 3'

where X denotes O6-nitrobenzylguanine and Y denotes N6-coumarinylmethyl-adenine. Notably, multiple copies can be encoded such that decoding can be performed by SBS as unconverted nucleobases will be read as a mixture of bases in the sequencing result.

**Example 11: Decoding data from encoded DNA**

**[000353]** After data has been encoded into a 1500 bp DNA strand by use of a nanopore device combined by use of dual wavelength light pulses, the resulting DNA is ready for decoding ("reading") when the data is to be recovered. The DNA can be encoded with a multiplicity of approximately 10 to 100 copies, the encoded DNA contains enough copies to enable mixtures of outcomes to be decoded. The DNA is sequenced by use of long-read single-molecule sequencing by synthesis (Pacific Biosciences). The sequence output shows that the convertible bases are sequenced as expected, with near 100% fidelity; 98% or better) reading as the bases that were in the original assembly. Where a "zero" is encoded, the coumarinyl group is removed from the N6-coumarinylmethyl-adenine, resulting in formation of adenine. Thus, the signal of "A" is found to be enhanced over that of N6-coumarinylmethyl-adenine at this position. However, the O6-nitrobenzylguanine sequencing signature in the same bit pair reads as mix of G and A. At positions encoded to be "one", both the coumarinyl group and the nitrobenzyl groups are removed, resulting in both the A signal being enhanced at position Y in the bit and adenine signal being enhanced at position X the same bit pair.

**Example 12: Stochastic or irregular data encoding**

**[000354]** In this example, the convertible nucleobases are provided irregularly spaced along the polymer. The data encodable polymer comprises O6-nitrobenzylguanine and O4-nitrobenzylthymine along the strand. Conversion of O6-nitrobenzylguanine into guanine can be denoted as a "zero" and conversion O4-nitrobenzylthymine into thymine can be denoted as a "one." As the polymer pass through the nanopore, data is encoded by selectively converting the appropriate convertible nucleobase in accordance with a data code. Furthermore, convertible nucleobases can be skipped to ensure the correct code is encoded. FIG. 15 illustrates a DNA polymer before and after data encoding, in which a code of "1010010" is encoded. Several convertible nucleobases are skipped and left unconverted in the process. When the encoded data is decoded, only the converted nucleobases are utilized to decipher the data code and the unconverted bases are ignored. When using SBS, multiple redundant encoded DNA polymers can be utilized to decipher whether a particular nucleobase is unconverted (e.g., by providing reads of mixed nucleobase structures) or converted (e.g., by providing reads a singular nucleobase structure).

### **Example 13: Constructing "Writeable" DNA with Modified Convertible Nucleobases at Regular Intervals**

**[000355]** The convertible base O6-coumarinylG (G\*) is synthesized as a deoxynucleoside triphosphate derivative (dG\*TP). It acts as a polymerase substrate when a DNA template is provided to contain a complementary base, such as "benzi" (*see, e.g.*, C. M. N. Aloisi et al., J. Am. Chem. Soc 2020, 142(15):6962-6969). Benzi is known to pair selectively with O6alkylG modified bases.

**[000356]** A circular single-stranded DNA oligonucleotide is constructed having 60 nucleotides in size, with a single "benzi" nucleotide in the sequence. The other 59 nucleotides comprise native A, C, T, and G nucleotides. A DNA primer (20 nt in length) (1  $\mu$ M) complementary to a non-benzi region of the circle is added to a solution of the circle (1  $\mu$ M) in polymerase-supporting buffer. To induce a "rolling circle" DNA synthesis, Phi29 polymerase is added along with five nucleotides at 500  $\mu$ M each (dATP, dGTP, dCTP, dTTP, and dG\*TP), under suitable conditions known for the Phi29 polymerase activity. After 4 hours, the resulting solution has long repeating single-stranded DNAs of varying length but many over 10kB in length as judged by agarose gel electrophoresis with size markers. Sequencing of the single-stranded DNAs in the solution confirms that the repeating sequence contains a G\* base once per repeat, evenly spaced at 60 nucleotides apart.

**[000357]** This solution of single-stranded DNAs is converted to double-stranded form by using a primer complementary to this repeating sequence, along with four native nucleoside triphosphates and phi29 polymerase. The result is a solution of long double-stranded DNAs containing single G\* modified bases every 60 bp.

**[000358]** This polymerase approach, together with modified DNA bases, is used to solve the problem of incorporating photomodifiable groups into nucleobases in a DNA where the photomodifiable groups are not substrates for polymerase enzymes.

**[000359]** To construct a repeating DNAs containing a second modified base, a modification of this strategy is used. The modified base T\* is synthesized as a deoxynucleoside triphosphate derivative. T\* is O4-nitrophenethylT, containing a group NPE that can be removed with light. O4-alkylT is known to pair with polymerases opposite G. *See, e.g.*, M. K. Dosanjh et al., Carcinogenesis 1993, 14(9):1915-1919.

**[000360]** A second circular DNA containing benzi is constructed once in the sequence. In this case, there is also only one C in the sequence, placed 10nt away from benzi; the remainder of the bases are G, C, and T. Using DNA polymerase and primer as described

above, together with the same five nucleotides above (dATP, dGTP, dCTP, dTTP, and dG\*TP) produces long repeating DNAs containing G\* once per repeat and a single G per repeat ten nucleotides away. Use of a DNA primer complementary to this repeat, combined with polymerase and nucleotides (dTTP, dGTP, dATP, dT\*TP, with no dCTP) results in synthesis of long repeating DNA duplexes containing G\* once per repeat and T\* once per repeat, ten bp away from G\* and in the opposite strand.

**[000361]** This example shows that writable DNAs with photo-removable nucleobases at regular intervals can be synthesized using nucleotide with photo-removable nucleobases (e.g., photo-removable nucleobases that will convert to natural nucleobase after conversion by light) in the presence of polymerase. This method can utilize polymerase for controllable production of longer strands of DNAs. DNAs produced using this method are significantly longer than those DNAs can only be synthesized by ligations of synthetic oligos, such as DNAs with backbone modifications.

#### **Example 14: Writing "scarless" Data in DNA and Reading with Long-Read SMRT Sequencing**

**[000362]** A 20kb DNA is constructed to contain two modified convertible nucleobases (X and Y) that can be converted to native DNA nucleobases upon "writing" by photoirradiation. The positions of all modifications are known, and are repeatedly spaced with distance of about 60 base pairs (ca. 20nm) between each occurrence of a given modification. That is, X is located approximately 60 base pairs (bp) from the adjacent X, and Y approximately 60 bp from adjacent Y. Both modifications (X and Y) are within 10 base pairs of each other, such that a given pair or duad of X/Y is simultaneously exposed in a given localized photoexcitation event. This DNA assembly is denoted "DNA blank tape". Mixed polymerases can be used for incorporation of two or more modified nucleobases in the DNA blank tape.

**[000363]** Nucleobase X is guanine modified with an O-nitrophenethyl (NPE) group directly attached without linker or sidechain at O-6. It can be converted to native guanine (i.e., without a scar) by irradiation at 360 nm. In this example, the O-6 modified guanine is the "unwritten" ("blank") form of the nucleotide, and after successful removal by irradiation, the guanine product is considered written, and its interpretation as 1 or 0 depends on the state of a nearby Y modification.

**[000364]** Previous work has shown that guanine modified by an alkyl group at O-6 can be read by a polymerase enzyme via sequencing by synthesis. *See, e.g.,* A. M. Kietrys, J. Am. Chem. Soc. 2017, 139(47);17074-17081. It typically codes for a mixture of A and G among the numerous reads of the sequence. The quantitative percentages of coding depend on which exact modification and which polymerase is used to read it, and this is measured beforehand (in a calibration experiment) by SMRT sequencing of synthetic DNA fragments containing the modification. Consensus reads yield the percentages of base encoding for this modification. For example, one can observe that on rereading the same DNA fragment, one sees that the polymerase inserts C opposite the modified base in 30% of reads (interpreting the base as "G"), and inserts T opposite the base in 64% of reads (interpreting it as "A"). This mixed signal for a single modified base is a signal (a fingerprint) of an unwritten bit. If the base in that single molecule is successfully photoconverted to G, then essentially 100% of reads will interpret it as G.

**[000365]** If there are multiple copies (for example, 1000 copies) of the same DNA molecule containing this modification at one position, and the DNA is irradiated by light in bulk solution at 360 nm to the extent that the NPE group is removed in 50% of the DNAs, then this change remains readable by sequencing by synthesis. Its consensus read is a 50% average between the fingerprint of the modified nucleobase (i.e., O-6 nitrophenethyl substituted guanine) and that of the native nucleobase (i.e., guanine). Thus the user can read data that is encoded by light at less than 100% complete yield.

**[000366]** Also in this example, nucleobase Y is thymine modified with a coumarinyl (Coum) group at O-4. It can be converted to native thymine in a "scarless reaction" by photoirradiation at 360 nm or 400 nm. Similar to the analysis of guanine above, a calibration is done with SMRT sequencing to determine its mixed coding percentages, distinct from that of native thymine. This mixed coding percentage is a fingerprint denoting an unconverted Coum-thymine, such as occurs in an unwritten bit. When Coum-thymine is photoconverted to a native nucleobase thymine (T), it codes as native T, essentially 100% of reads. As for nucleobase X, one can interpret partial conversion among multiple copies of DNA by observing an averaging of the fingerprints of the modified nucleobase Y and native nucleobase T.

**[000367]** In this example, a "0" bit is interpreted as such when T-Coum in a G-NPE/T-Coum pair is converted to T via irradiation at 400nm. If both modifications are removed (using 360nm irradiation), the bit is interpreted as "1". Again, reads of multiple copies of the data can be used to interpret bits that are converted below the 100% maximal yield.

**[000368]** Writing a data “bit” locally makes use of local irradiation or local excitation method such as translocating a STED microscope irradiation beam along the DNA, or translocating the DNA in a zero mode waveguide or through a plasmonic nanopore using methods known in the art.

**[000369]** Note that the blank tape DNA in this example is modified with approximately evenly spaced X and Y everywhere in the DNA sequence. Thus it contains the potential to be written with binary data anywhere. Pairs of X,Y modified groups are simply regarded as lacking data (i.e., unwritten). The identical data can be written starting anywhere in the DNA (assuming there is enough length to complete the writing process). Since the DNA positioning relative to the writing light can stochastically vary, and the speed of translocation can vary, one can still write and read data by interpreting the string of 0 and 1 bits, skipping over "blank" bits. This has the advantage of not requiring careful positioning of the start and stop site of writing, and does not require perfect control over translocation speed. Because there is no need to pause to position bits, the writing method is simpler and faster than methods that function by controlling the translocation and exact position of the DNA polymer through a nanopore.

**[000370]** Data encoding the letter "e" are written into the DNA blank tape at the single molecule level using a superresolution microscope on stretched DNA molecules on a slide. The 8-bit Unicode binary string for letter “e” is 01100101, using eight pulses of 360 nm light (1) and/or 400 nm light (0) from a superresolution microscope at 20 nm resolution. The writing is done 1000 times on 1000 single molecules, collecting the DNAs at the end by washing the slide containing them.

**[000371]** This "written" DNA is submitted to SMRT sequencing. Positions showing the fingerprints of modified nucleobases (as a G-NPE/T-Coum pair) are interpreted as blank and not encoding data. Paired bit positions in which the consensus of reads show an averaging of the fingerprints of modified and unmodified bases are interpreted as data; selectively unblocking of T by removing NPE indicates a "0" and paired bit positions that show substantial conversion of both T and G indicate a written bit of "1". Progressing along the strand in order generates the bit string 01100101, indicating the storage of data (data conversion interprets it as the letter "e").

**[000372]** Note that data correction can be optionally used to correct errors. For example, if most single molecule DNA copies yield a string of 01100101, but other binary strings are also present, comparisons of binary data can lead to the correct conclusion. For example, some missed bits may occur (example 0100101) or the data may run out because the end of

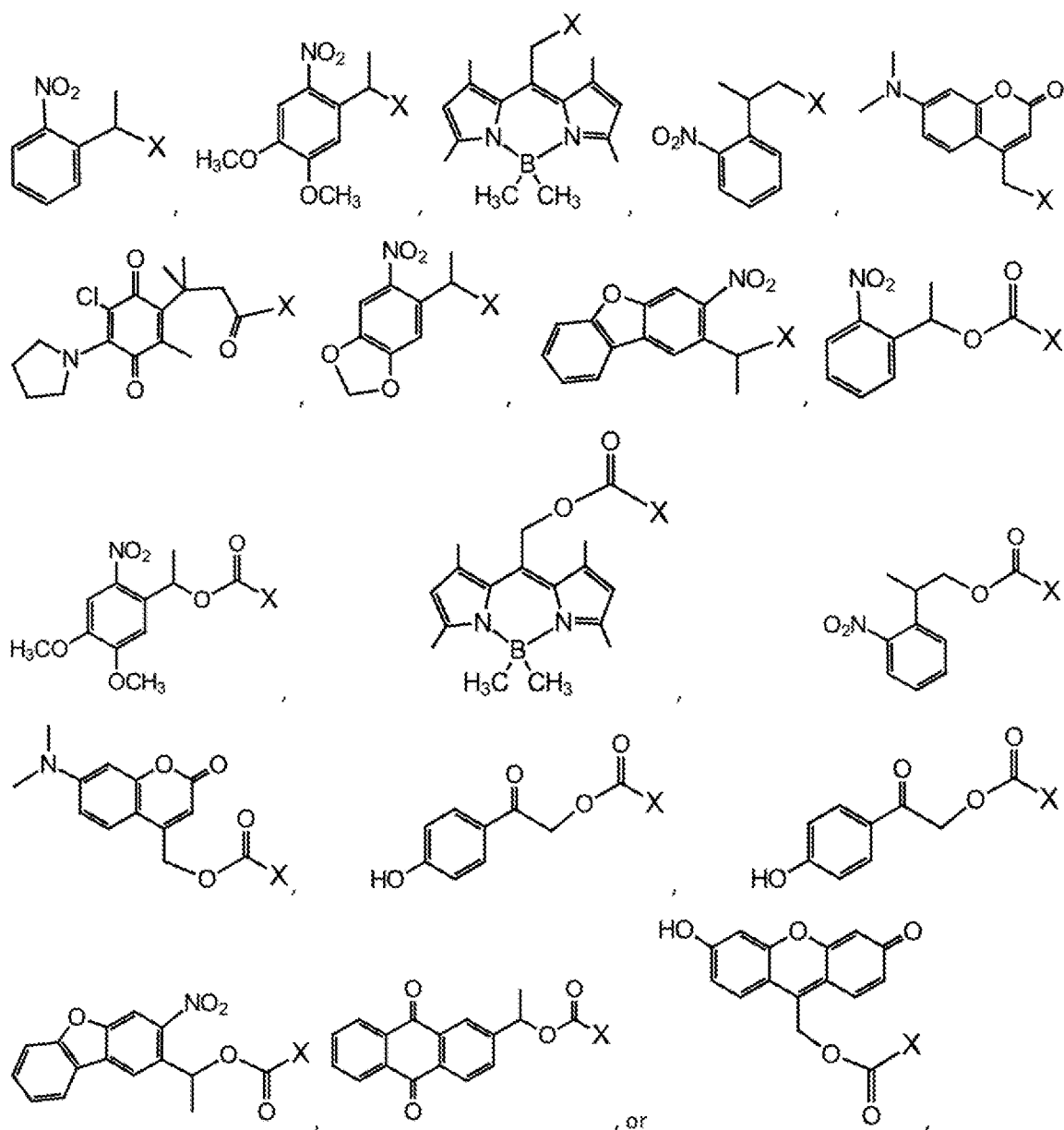
the DNA can be reached (example 01100). However the comparison of these different strings leads to the correct conclusion even with these errors. This dual bit active writing enables the user to write more rapidly than would be possible if specific positioning of the DNA were required.

**WHAT IS CLAIMED IS:**

1. A polymer for encoding data, comprising:  
a plurality of convertible residues iteratively spaced along and covalently linked to the backbone of the polymer,  
wherein each of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme;  
wherein the plurality of convertible residues are covalently linked to the polymer in the first state and in the second state.
2. The polymer of claim 1, wherein the polymer is a nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.
3. The polymer of claim 2, wherein the nucleic acid polymer is a single-stranded nucleic acid polymer.
4. The polymer of claim 2, wherein the nucleic acid polymer is double-stranded nucleic acid polymer.
5. The polymer of any one of claims 2-4, wherein the nucleic acid polymer comprises Deoxyribonucleic acid (DNA), Ribonucleic acid (RNA), phosphorothioate DNA, glycerol nucleic acids (GNA), threose nucleic acids (TNA), locked nucleic acids (LNA), or a combination thereof.
6. The polymer of any one of claims 2-5, wherein the nucleic acid polymer comprises greater than 10 convertible residues.
7. The polymer of any one of claims 2-6, wherein the ratio of the total number of nucleotides to the convertible residues in the nucleic acid polymer is between 2 to 100.
8. The polymer of any one of claims 2-7, wherein the plurality of convertible nucleobases are non-naturally occurring nucleobases.
9. The acid polymer of claim 8, wherein the plurality of convertible nucleobases are modified naturally occurring nucleobases or derivatives of naturally occurring nucleobases.

10. The polymer of any one of claims 2-9, wherein each of the plurality of convertible nucleobases comprises a chemically modifiable moiety.
11. The polymer of any one of claims 2-10, wherein each of the plurality of convertible nucleobases the chemically modifiable moiety is directly attached to the base of the convertible nucleobases.
12. The polymer of any one of claims 2-10, wherein each of the plurality of convertible nucleobases the chemically modifiable moiety is attached to the base without a linker or a sidechain.
13. The polymer of claim 11 or 12, wherein the plurality of convertible nucleobases are covalently linked to the backbone of the nucleic acid via the sugar .
14. The polymer of any one of claims 2-13, wherein the chemically modifiable moiety is activatable by light, voltage, enzymatic agent, chemical reagent, or a redox agent, thereby converting from the first state into the second state.
15. The polymer of claim 14, wherein the chemically modifiable moiety is activatable by light, thereby converting from the first state into the second state.
16. The polymer of claim 14 or 15, wherein the conversion from the first state into the second state occurs via an irreversible reaction.
17. The polymer of any one of claims 2-16, wherein the convertible nucleobase becomes a naturally occurring nucleobase after conversion into the second state.
18. The polymer of claim 17, wherein the convertible nucleobase becomes guanine, adenine, thymine, uracil or cytosine after conversion into the second state.
19. The polymer of any one of the preceding claims, wherein the backbone of the polymer (e.g., phosphate and sugar in nucleic acid polymer) remain unchanged during the conversion from the first state into the second state.
20. The polymer of any one of the preceding claims, wherein the polymer comprises two or more different sets of convertible residues, each set of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different.

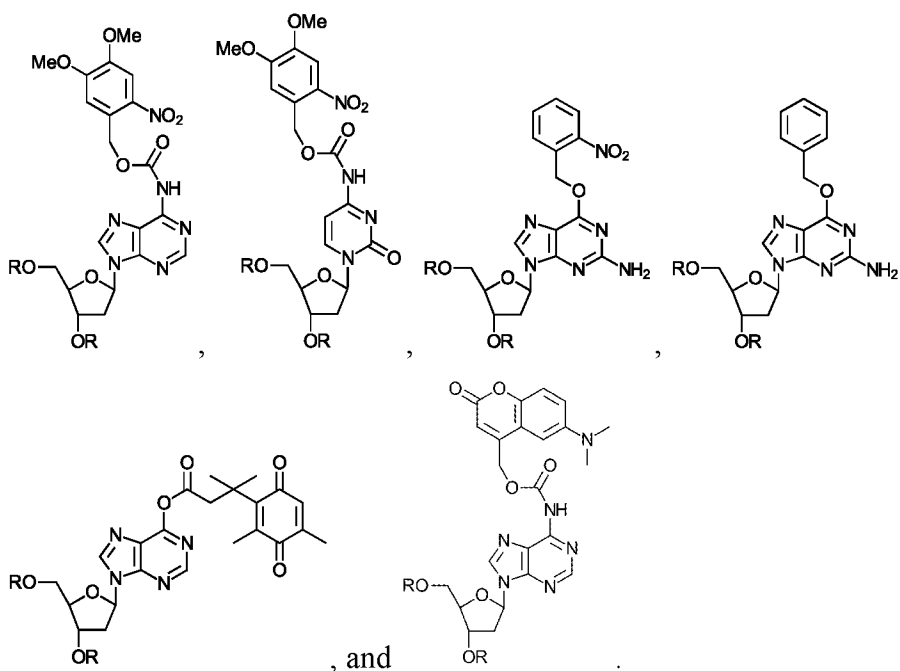
21. The polymer of any one of the preceding claims, wherein each of the plurality of convertible residues comprises a chemically modifiable moiety that can be activated by light.
22. The polymer of claim 20, wherein the two or more different sets of convertible residues are activatable by light of different wavelengths.
23. The polymer of claim 22, wherein a first set of convertible residues is activatable by light of a first wavelength, and a second set of convertible residues is activatable by light of a second wavelength, the first wavelength and the second wavelength being different.
24. The polymer of claim any one of the preceding claims, wherein the chemically modifiable moiety comprises one or more photo-removable groups.
25. The polymer of claim 24, wherein the chemically modifiable moiety is a leaving group.
26. The polymer of claim 24, wherein the one or more photo-removable groups are:



wherein X represents NR<sub>2</sub>, NHR, OR, or SR, and wherein R is the nucleobase to which the photo-removable group is attached.

27. The polymer of any one of claims 2-24, wherein the plurality of convertible nucleobases are capable of being converted by light of a wavelength of 325 nm, 360 nm, or 400 nm.
28. The polymer of any one of claims 2-24, wherein the plurality of convertible nucleobases are capable of being converted by light of a wavelength of between 400 nm to 850 nm.
29. The polymer of any one of claims 2-28, wherein each of the plurality of convertible nucleobases comprises a chemically modifiable moiety that is activatable by redox.

30. The polymer of claim 29, wherein the chemically modifiable moiety is capable of being activated by localized oxidation.
31. The polymer of claim 29, wherein the chemically modifiable moiety is capable of being activated by oxidation using electrodes.
32. The polymer of any one of claims 2-30, wherein a nucleotide comprising the convertible nucleobase is selected from the group consisting of:



33. The polymer of any one of claims 2-30, wherein the convertible nucleobase is selected from the group consisting of O6-guanine, N2-guanine, N7-guanine, N6-adenine, N5-adenine, O4-thymine, N3-thymine, 2-thio-thymine, 4-thio-thymine, N4-cytosine, or N3-cytosine.
34. The polymer of any one of claims 2-32, wherein the first state and the second state of the plurality of convertible nucleobases are readable by a sequencing method capable of detecting and differentiating non-naturally occurring and/or modified nucleobases.
35. The polymer of claim 34, wherein the first state and the second state of the plurality of convertible nucleobases are readable by nanopore sequencing.
36. The polymer of claim 34, wherein the first state and the second state of the plurality of convertible nucleobases are readable by sequencing by synthesis.
37. The polymer of any one of claims 2-36, wherein when the plurality of convertible nucleobases are converted to the second state, properties of the plurality of convertible

- nucleobases are modified (e.g., having reduced size, altered shape, modified H-bonding, and/or modified polymerase substrate ability) as compared to the first state.
38. The polymer of any one of claims 2-37, wherein one or more of the plurality of convertible nucleobases are capable of being converted from the second state into a third state; wherein the one or more of the plurality of convertible nucleobases are attached covalently to the nucleic acid polymer in the third state.
  39. The polymer of any one of claims 1-38, wherein each of the plurality of convertible residues is capable of being independently and selectively converted.
  40. The polymer of any one of claims 1-38, further comprising a plurality of spacer residues linked via the backbone of the polymer, wherein each of the plurality of convertible residues are separated by one or more spacer residues of the plurality of spacer residues.
  41. The polymer of any one of the preceding claims, wherein the iterative spacing among the plurality of convertible residues conforms to a resolution of a writing mechanism for encoding data on the polymer.
  42. The polymer of any one of the preceding claims, wherein the iterative spacing among two adjacent convertible residues is equal to or greater than a resolution of a data encoding mechanism for encoding data into the polymer.
  43. The polymer of claim 41, wherein the resolution of the writing mechanism is at least 1 nm.
  44. The polymer of any one of the preceding claims, wherein the plurality of spacer residues do not interfere with reading of the convertible residues.
  45. The polymer of any one of the preceding claims, wherein the plurality of spacer residues in the polymer are the same spacer residues.
  46. The polymer of any one of claims 1-44, wherein the plurality of spacer residues comprise two or more different spacer residues (e.g., different nucleobases such as different naturally occurring nucleobases).
  47. The polymer of any one of the preceding claims, wherein the polymer consists essentially of spacer residues.
  48. The polymer of any one of claims 2-46, wherein each of the plurality of convertible nucleobases are separated by 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, or 50 spacer residues.

49. The polymer of claim 48, wherein each of the plurality of convertible nucleobases are separated by 6 spacer residues.
50. The polymer of any one of claims 2-49, wherein the plurality of spacer residues are naturally occurring nucleobases, non-naturally nucleobases, tetrahydrofuran abasic residues, or ethylene glycol residues.
51. The polymer of claim 50, wherein the plurality of spacer residues are naturally occurring nucleobases.
52. The polymer of any one of the preceding claims, further comprising one or more delimiters linked to the backbone of the polymer.
53. The polymer of claim 52, wherein each of the one or more delimiters comprises one or more naturally occurring nucleobases or non-naturally nucleobases.
54. The polymer of claim 52 or 53, wherein the one or more delimiters comprise naturally occurring nucleobases.
55. The polymer of any one of claims 52-54, wherein the one or more delimiters separate two or more adjacent data fields within the polymer.
56. The polymer of any one of the preceding claims, further comprising one or more data tags.
57. The polymer of claim 56, wherein the one or more data tags comprise one or more naturally occurring nucleobases or non-naturally nucleobases.
58. The polymer of claim 56 or 57, wherein the polymer is a nucleic acid polymer and the one or more data tags are present at the 5' or 3' end of the nucleic acid polymer.
59. The polymer of any one of claims 56-58, wherein the one or more data tags are incorporated to the nucleic acid polymer during the nucleic acid polymer is synthesized, during the plurality of convertible nucleobases are converted to the second state, or via ligation after the plurality of convertible nucleobases are converted to the second state.
60. The polymer of any one of the preceding claims, wherein the polymer can be stored under standard nucleic acid storage protocols.
61. The polymer of claim 60, wherein the polymer is a nucleic acid polymer that can be stored in appropriate nuclease-free solution at room temperature, or at a lower temperature (e.g., - 20 °C).

62. The polymer of claim 60 or 61, wherein the polymer can be stored at room temperature without stabilizer.
63. A system for data writing, comprising:  
a writable polymer comprising a plurality of convertible residues iteratively spaced along and covalently linked to the backbone of the polymer, wherein each of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; wherein the plurality of convertible residues are attached covalently linked to the polymer in the first state and in the second state; and  
a data writing device for writing data on the writable polymer.
64. The system of claim 63, wherein the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.
65. The system of claim 63 or 64, wherein the data writing device comprises a nanopore.
66. The system of claim 63 or 64, wherein the data writing device comprises a microscope with a light source.
67. The system of any one of claim 63-66, wherein the data writing device converts the plurality of convertible nucleobases into the second state by light pulses, voltage pulses, an enzymatic agent, or a redox agent.
68. The system of claim 67, wherein the data writing device converts the converts the plurality of convertible nucleobases into the second state by light pulses.
69. The system of any one of claim 63-68, wherein the data writing device comprises a light irradiation device.
70. A method for generating a writable nucleic acid polymer, comprising:  
providing a circular single-stranded oligonucleotide template, wherein the circular single-stranded oligonucleotide template is complementary to a repeating data field that comprises convertible nucleobases; and  
incubating the circular single-stranded oligonucleotide template in the presence of a nucleic acid primer, a polymerase, and triphosphate nucleotides, wherein the triphosphate nucleotides comprise convertible nucleobases in a first state and are capable of being converted from the first state into a second state, the first state and the second state being different.

71. The method of claim 70, wherein the circular single-stranded oligonucleotide template comprises nucleobases complementary to the convertible nucleobases, and wherein the complementary nucleobases are iteratively spaced such that the incubation of the template with the nucleic acid primer, the polymerase, and the triphosphate nucleotides provides a nucleic acid polymer comprising a plurality of the convertible nucleobases iteratively spaced along and covalently linked via the backbone of the nucleic acid polymer; wherein the plurality of the convertible nucleobases are covalently linked to the nucleic acid polymer in the first state and in the second state.
72. The method of claim 70 or 71, wherein the repeating data field further comprises spacer nucleobases, and wherein the triphosphate nucleotides further comprise triphosphate spacer nucleotides.
73. A method for generating a writable nucleic acid polymer, comprising:  
chemically synthesizing a plurality of oligomers, each oligomer comprises a plurality of convertible nucleobases iteratively spaced along and linked via the nucleic acid polymer backbone, wherein each of the plurality of convertible nucleobases has a first state and is capable of being converted from the first state into a second state; wherein the plurality of convertible nucleobases are attached covalently to the nucleic acid polymer in the first state and in the second state, the first state and the second state being different; and  
ligating the plurality of oligomers to form the writable nucleic acid polymer
74. The method of claim 73, wherein each of the plurality of oligomers comprises a plurality of spacer residues linked via the backbone of the nucleic acid polymer, wherein each of the plurality of the convertible nucleobases is separated by one or more spacer residues of the plurality of spacer residues.
75. The method of claim 73 or 74, wherein the ligating step is via chemical ligation.
76. The method of claim 73 or 74, wherein the ligating step is via enzymatic ligation.
77. The method of any one of claims 73-76, wherein a complementary DNA splint is used in the ligating step.
78. The method of any one of claims 73-77, further comprising:  
annealing a plurality of complements to the oligomers prior to the ligating step.
79. A method for writing data onto a writable polymer, comprising:

providing a writable polymer that comprises a plurality of convertible residues iteratively spaced along and covalently linked via the backbone of the polymer, wherein each convertible residues of the plurality of convertible residues has a first state and is capable of being converted from the first state into a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme; and

selectively converting, utilizing a data writing device, one or more of the plurality of convertible residues into the second state such that a data encoded polymer is generated.

80. The method of claim 79, wherein the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.
81. The method of claim 79 or 80, wherein the data writing device comprises a nanopore, and the method further comprising:

passing the writable polymer through the nanopore of the writing device, wherein the nanopore comprises converts one or more of the plurality of convertible residues into the second state.
82. The method of claim 81, wherein the nanopore is a plasmonic nanopore that provides light pulses or redox energy to selectively convert convertible nucleobases from the first state into the second state.
83. The method of claim 79 or 80, wherein the data writing device comprises a plasmonic well or channel, and the method further comprising:

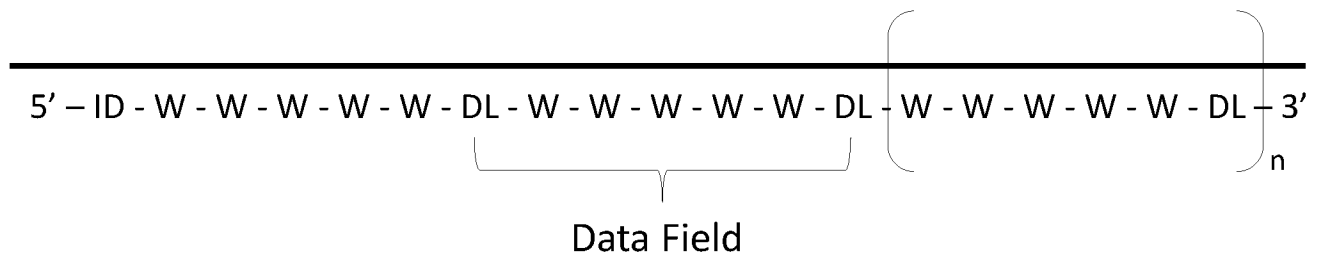
transferring the writable polymer into the plasmonic well or channel of the data encoding device, wherein the plasmonic well or channel provides light pulses or redox energy to selectively convert convertible nucleobases from the first state into the second state.
84. The method of any one of claims 79-81, wherein the data writing device selectively converts the convertible residues into the second state by light pulses, voltage pulses, an enzymatic agent, or a redox agent.
85. The method of any one of claims 79-82, wherein the data writing device selectively converts the converts the convertible residues into the second state by light pulses.

86. The method of any one of claims 79-85, wherein the convertible residues become naturally occurring nucleobases after conversion into the second state.
87. The method of any one of claims 79-86, wherein the plurality of convertible residues comprise two or more types of convertible residues, wherein a first type of convertible residues are activatable by light of a first wavelength and a second type of convertible residues are activatable by light of a second wavelength.
88. The method of any one of claims 79-87, wherein the iterative spacing among the plurality of the convertible residues conforms to a resolution of the data writing device for selectively converting the convertible residues.
89. The method of any one of claims 79-88, wherein the selectively converting step does not require specific positioning of the writable polymer.
90. The method of any one of claims 79-88, wherein the conversion of the convertible residues into the second state is non-uniform on the data encoded polymer.
91. The method of any one of claims 79-88, wherein the conversion of the convertible residues into the second state is not limited to certain positions on the data encoded polymer.
92. The method of any one of claims 79-91, further comprising stretching or combing the writable polymer (e.g., a writable DNA) on a solid support.
93. The method of any one of claims 79-92, further comprising visualizing locations of the convertible residues using a dye.
94. The method of any one of claims 79-93, further comprising locally illuminating or locally exciting the writable polymer.
95. The method of claim 94, wherein the locally illuminating or locally exciting uses Stimulated Emission Depletion (STED) laser.
96. The method of any one of claims 79-95, further comprising joining two or more data fields from two or more writable polymers end-to-end, resulting in a joined polymer comprising two or more data fields.
97. The method of any one of claims 79-96, further comprising controlling the passage rate of the writable polymer through the nanopore of the writing device.

98. The method of any one of claims 79-97, wherein a plurality of writable polymers pass through the data writing device to write the same data (e.g., generating data redundancy).
99. A method for reading data from a polymer encoded with data, comprising:  
providing the polymer encoded with data comprising convertible residues iteratively spaced along and covalently linked via the backbone of the polymer, wherein a first subset of the convertible residues are in a first state and a second subset of the convertible residues are in a second state, the first state and the second state being different and the plurality of convertible residues in the first state and the second state are readable by a polymerase enzyme;  
and  
passing the writable polymer encoded with data through a data reading device to read the encoded data on the polymer encoded with data.
100. The method of claim 99, wherein the writable polymer is a writable nucleic acid polymer and the plurality of convertible residues are convertible nucleobases.
101. The method of claim 99 or 100, wherein the convertible residues in the first state can be converted into the second state via light.
102. The method of any one of claims 99-101, wherein the data reading device comprises a nanopore.
103. The method of any one of claims 99-101, wherein the data reading device is a sequencing device.
104. The method of claim 103, wherein the sequencing device is a sequencing by synthesis device.
105. The method of any one of claims 99-104, further comprising measuring current flow of electrolytes during passage of the writable polymer.
106. The method of any one of claims 99-105, further comprising determining whether each of the plurality of convertible residues is in the first state or the second state based on the measured current flow of electrolytes during passage of the writable polymer.
107. The method of any one of claims 99-106, further comprising re-passing the polymer encoded with data through the data reading device to re-read the encoded data on the polymer encoded with data.

108. The method of any one of claims 99-107, further comprising validating and correcting the encoded data on the polymer encoded with data by comparing the encoded data on multiple copies of the polymer encoded with data.
109. A method for reading or decoding data from a nucleic acid polymer encoded with data, the method comprising:  
providing a plurality of redundant copies of the nucleic acid polymer encoded with data comprising:  
a plurality of converted nucleobases, wherein each converted nucleobase comprises a first nucleobase structure, wherein the first converted nucleobase has been converted from a first state into a second state, the first state and the second state being different; and  
a plurality of convertible nucleobases, wherein each convertible nucleobase comprising a second nucleobase structure and a directly linked leaving group, and wherein the convertible nucleobase is provided in a first state and is capable of being converted from the first state into a second state by releasing the second leaving group from the second nucleobase structure, the first state and the second state being different;  
wherein the converted nucleobases and convertible nucleobases are linked via the nucleic acid polymer backbone; and  
sequencing each redundant copy of the plurality redundant copies of the nucleic acid polymer.
110. The method of claim 109, further comprising:  
detecting the plurality of converted nucleobases and the plurality of convertible nucleobases; and  
decoding the data based on the detected plurality of converted nucleobases.
111. The method of claim 109 or 110, wherein the plurality of converted nucleobases in the first state and the second state are readable by a polymerase enzyme.
112. The method of any one of claims 109-111, wherein the plurality of convertible nucleobases in the first state and the second state are readable by a polymerase enzyme.
113. The method of any one of claims 109-112, wherein the plurality of converted nucleobases and the plurality of convertible nucleobases are detected based on the

sequencing result of the redundant copies of the nucleic acid polymer encoded with data.

**Fig. 1A****Fig. 1B**Key

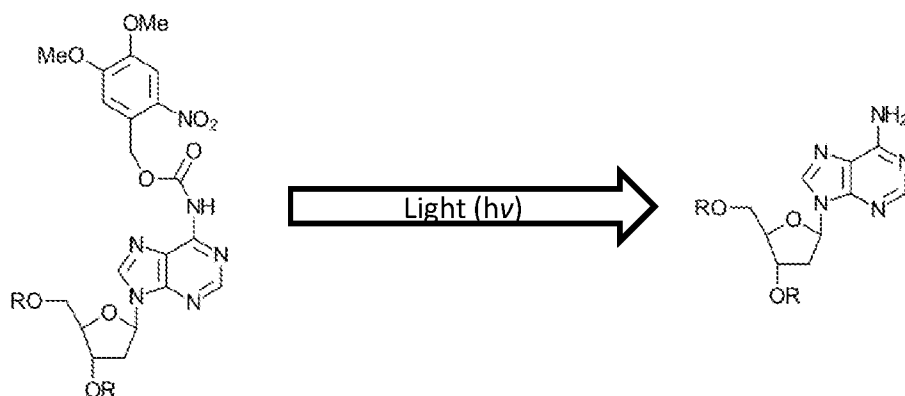
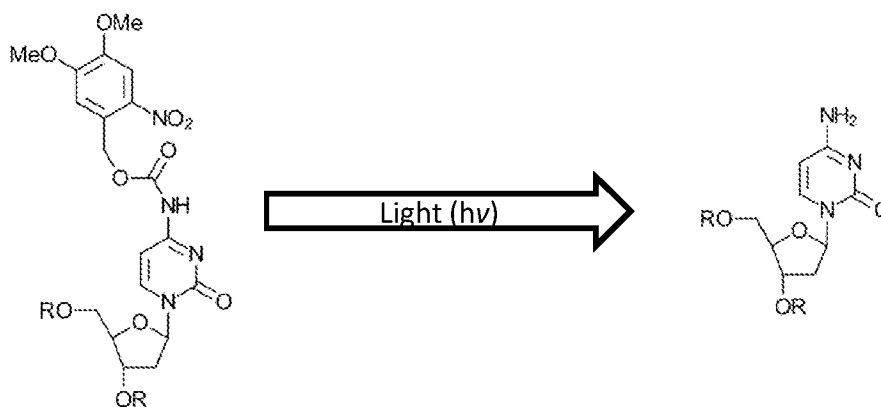
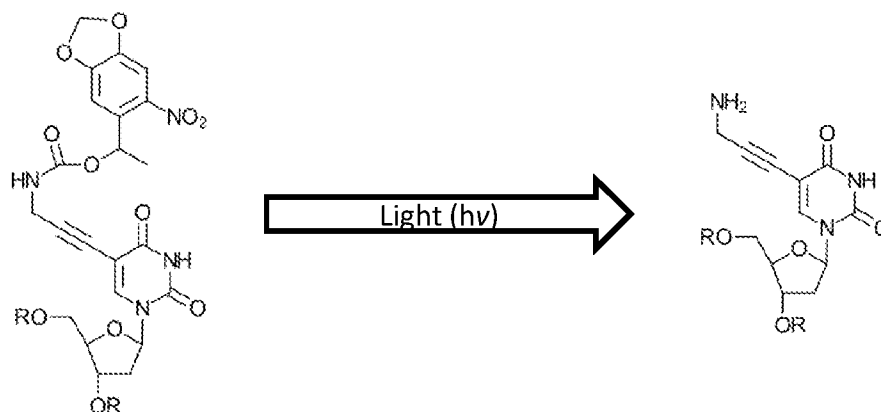
ID – Identifier sequence  
W – Writable nucleobase  
S – Spacer nucleobase  
DL – Delimiter sequence

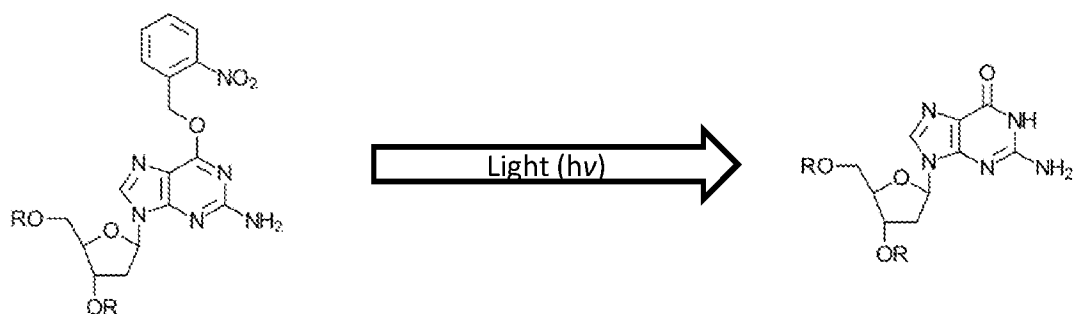
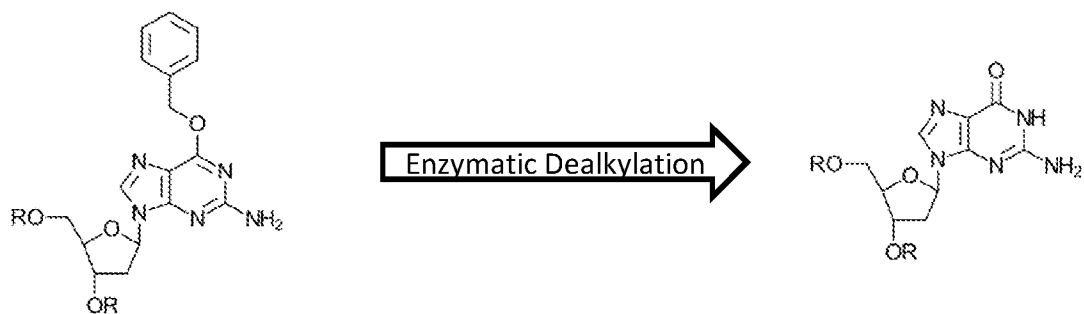
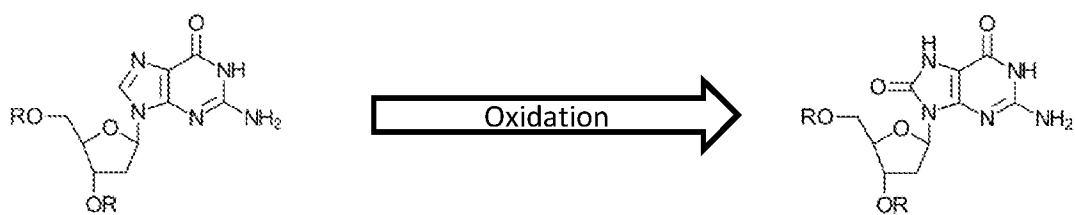
**Fig. 2A****Fig. 2B**Key

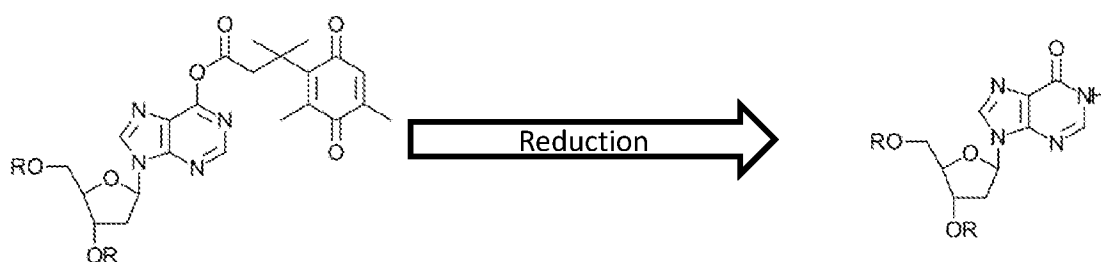
C – Convertible nucleobase

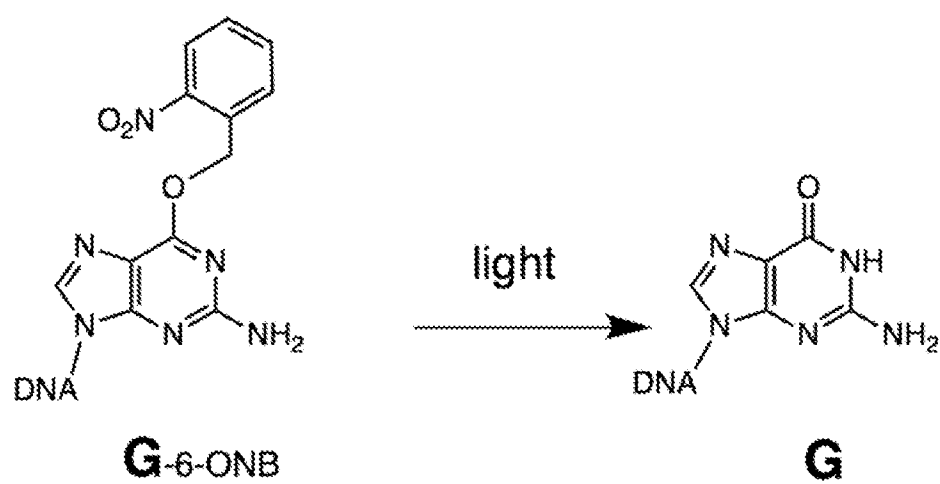
N – Random or spacer nucleobase

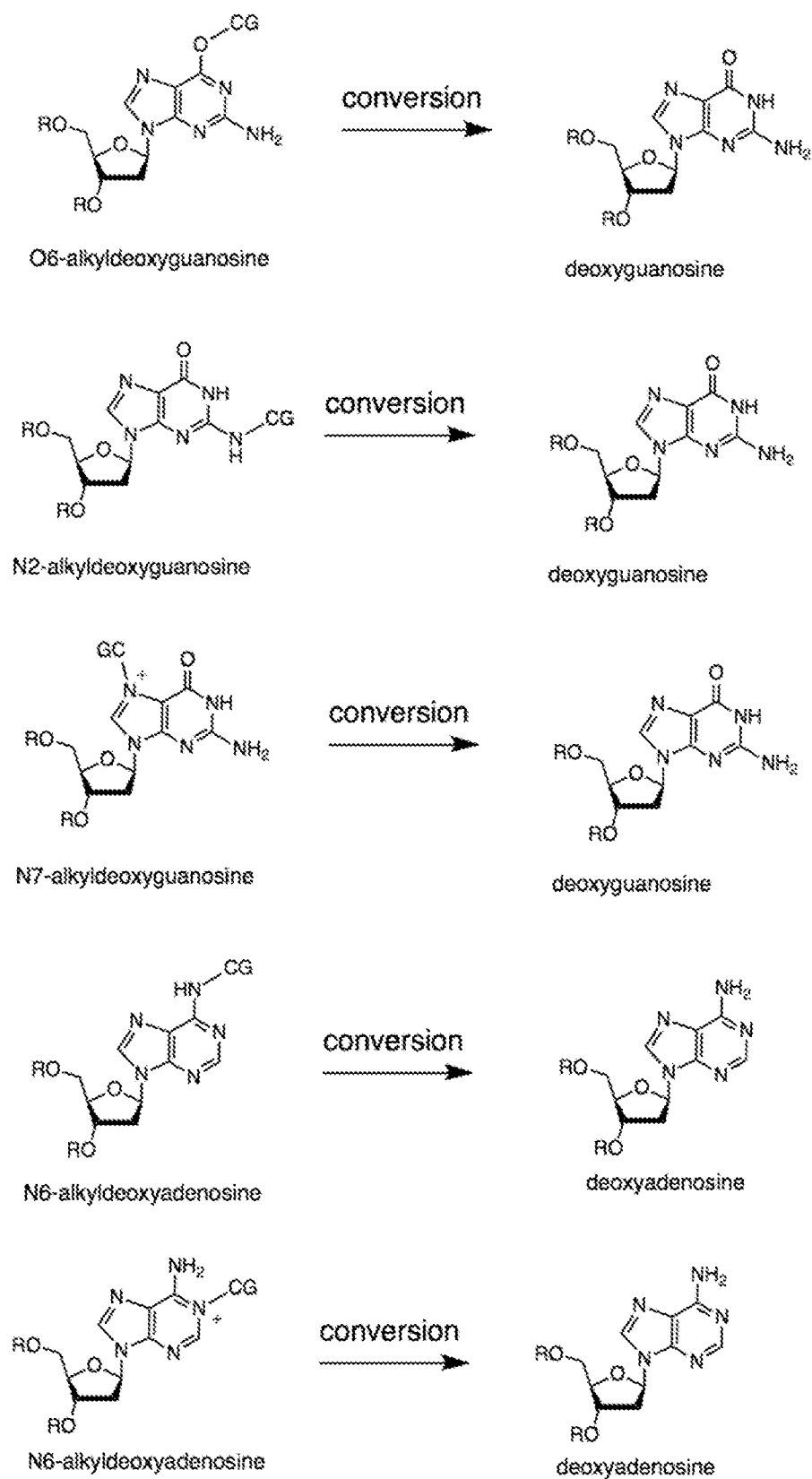


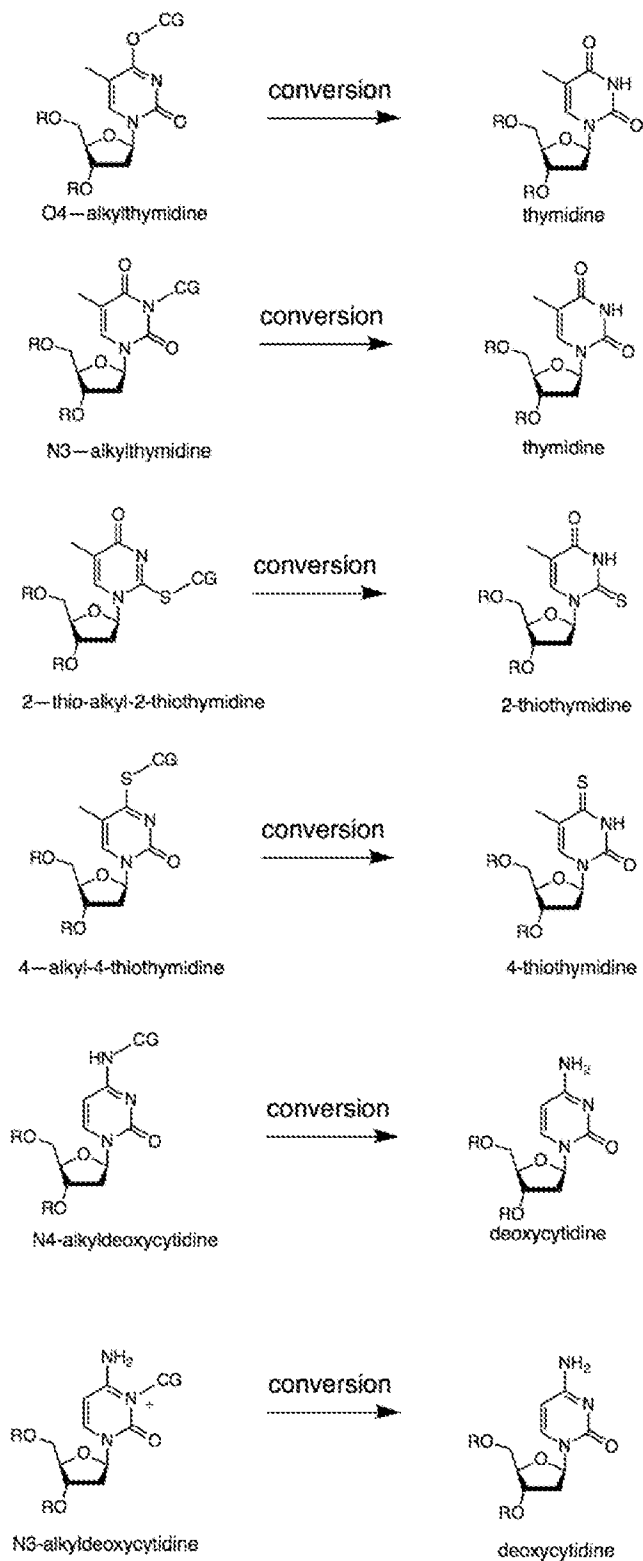
**Fig. 3A****Fig. 3B****Fig. 3C**

**Fig. 3D****Fig. 3E****Fig. 3F**

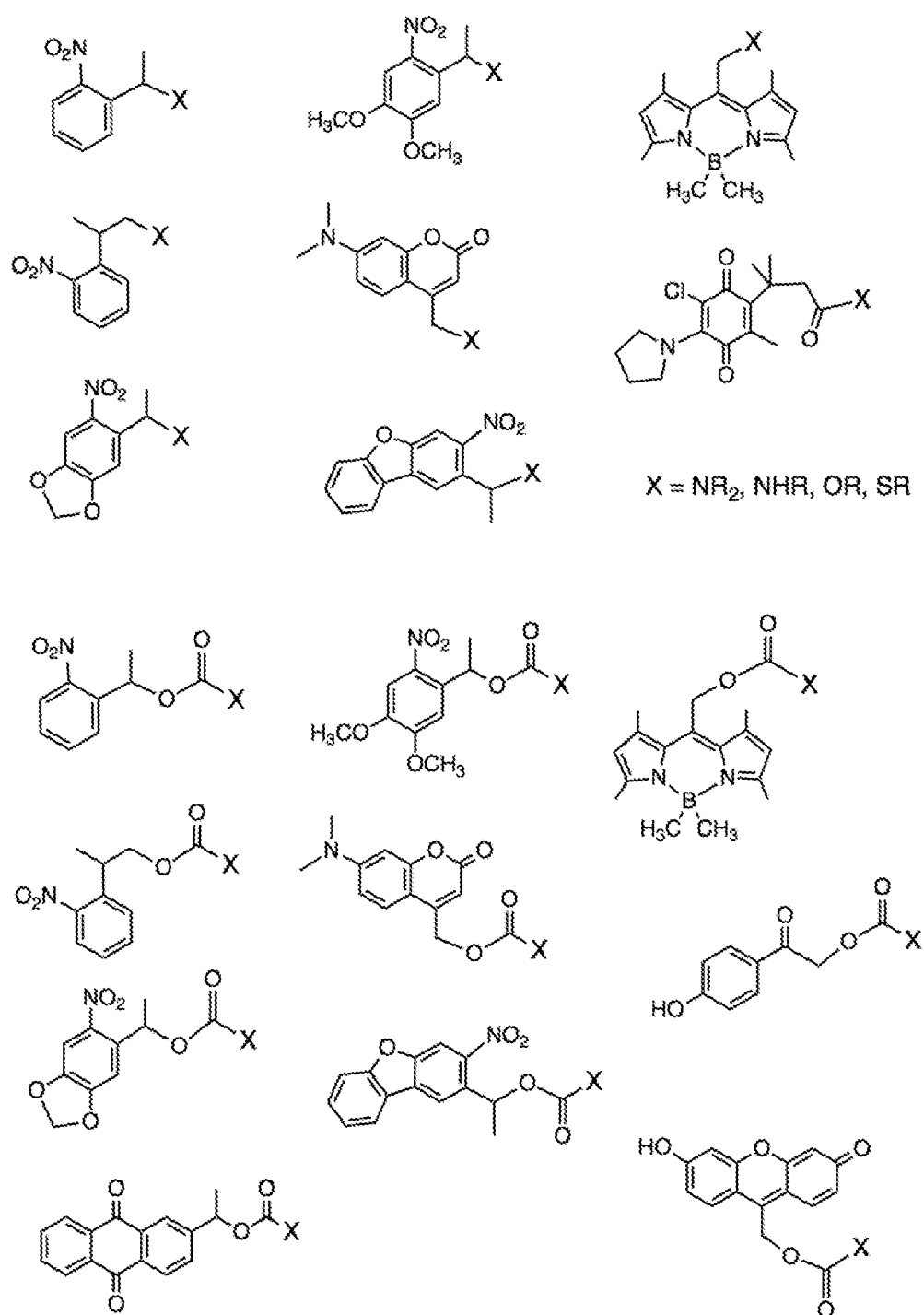
**Fig. 3G**

**Fig. 4**

**Fig. 5A**

**Fig. 5B**

**Fig. 6**



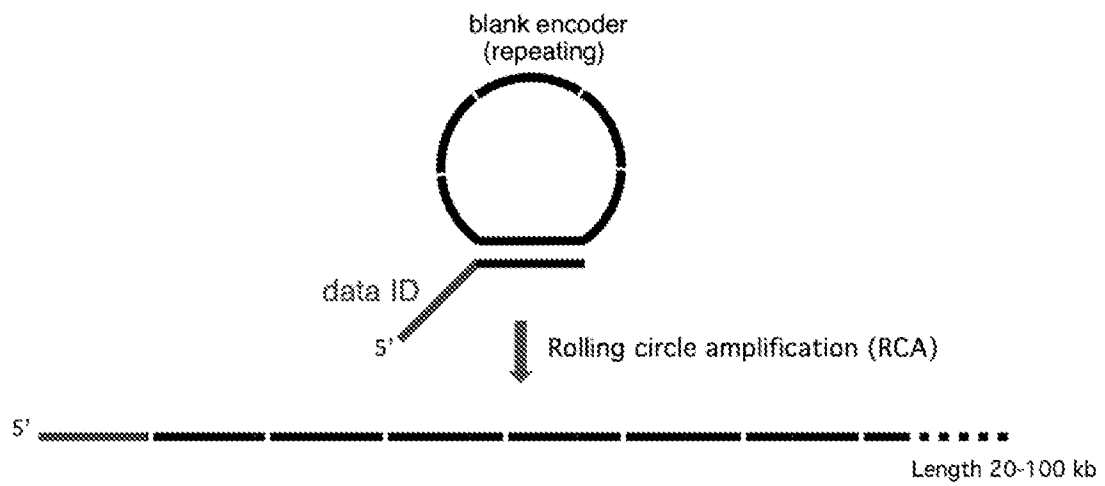
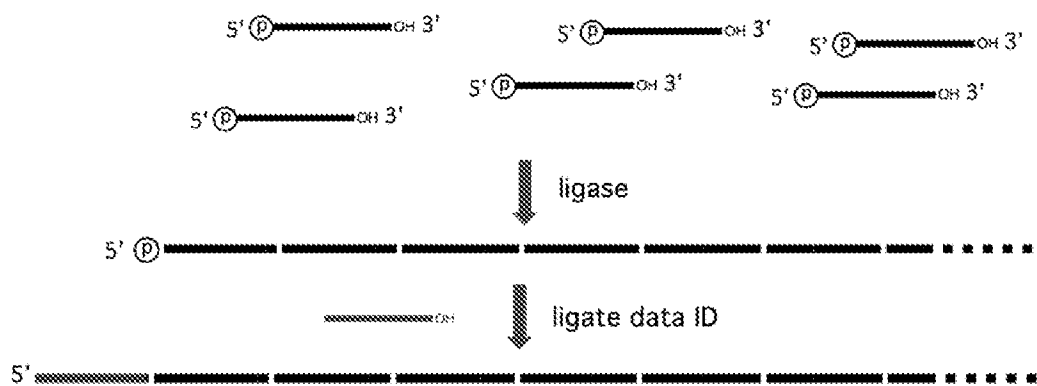
**Fig. 7****Fig. 8**

Fig. 9A

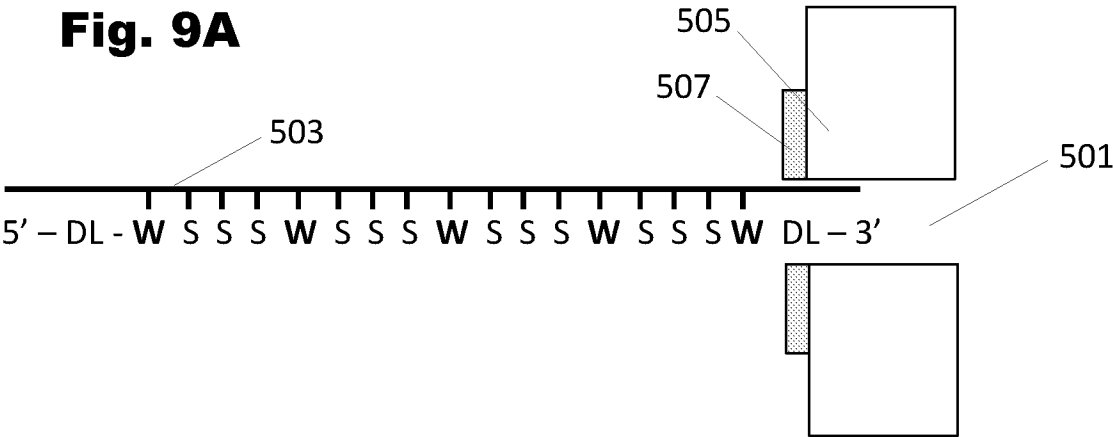


Fig. 9B

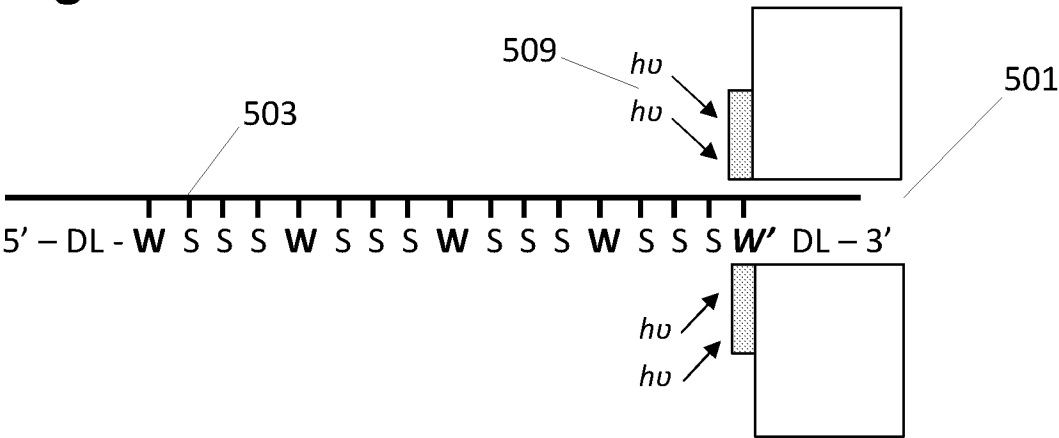
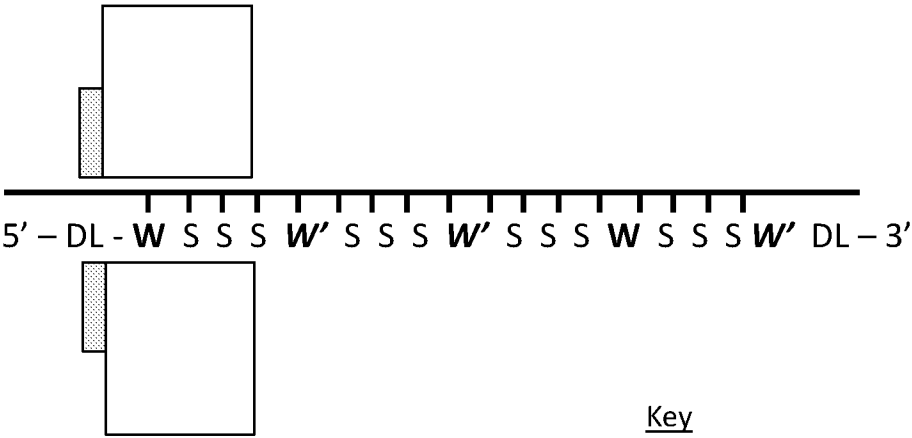


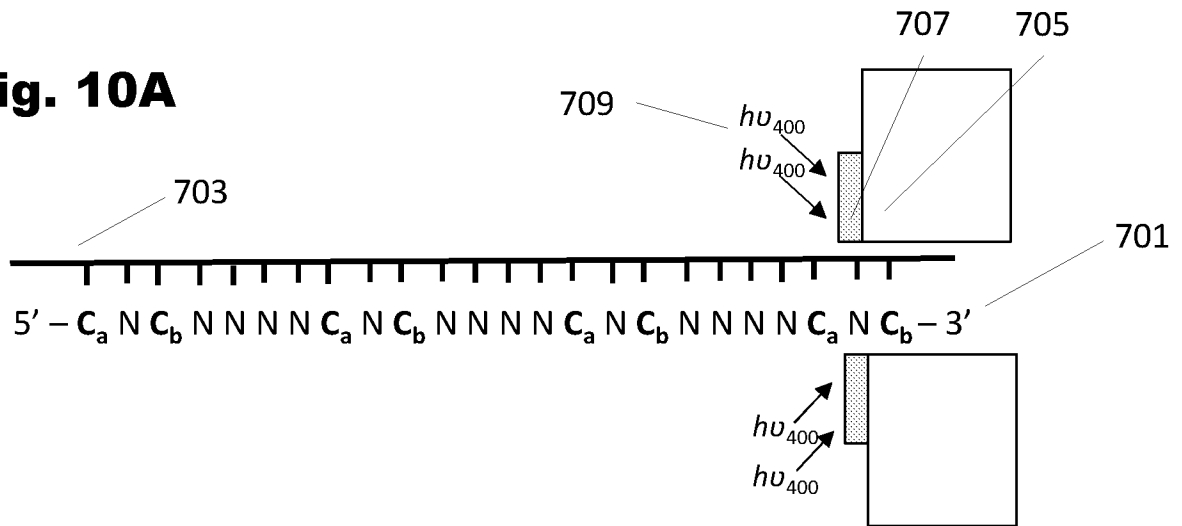
Fig. 9C



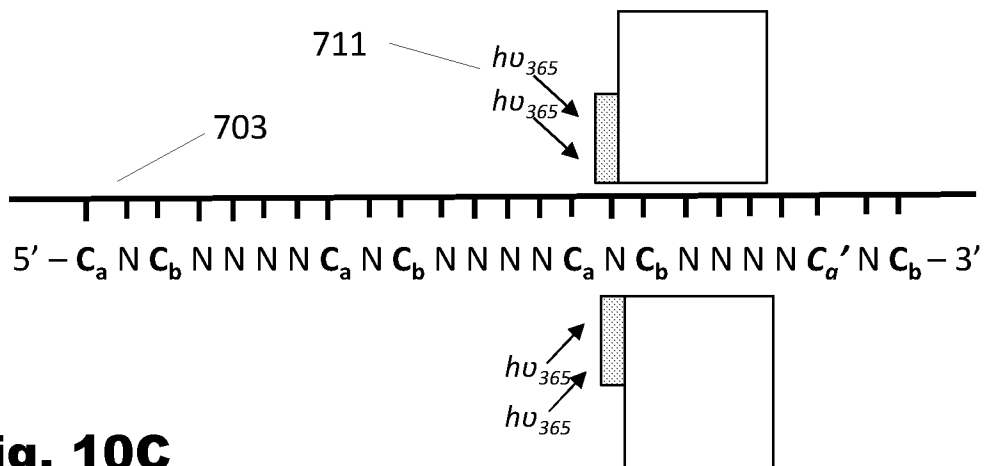
Key

- W – Convertible nucleobase
- W' – Converted nucleobase
- S – Spacer nucleobase
- DL – Delimiter sequence

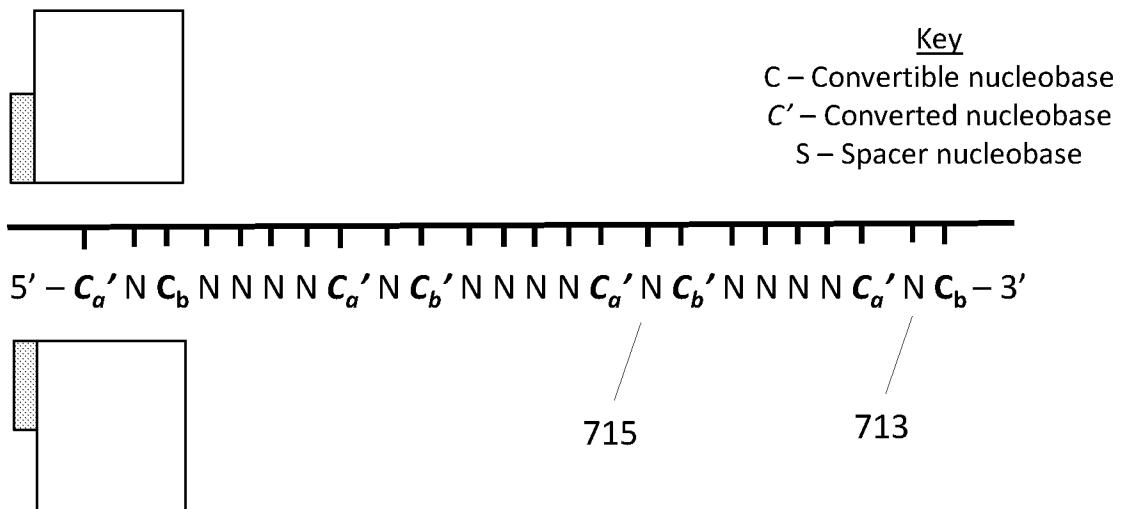
**Fig. 10A**



**Fig. 10B**



**Fig. 10C**



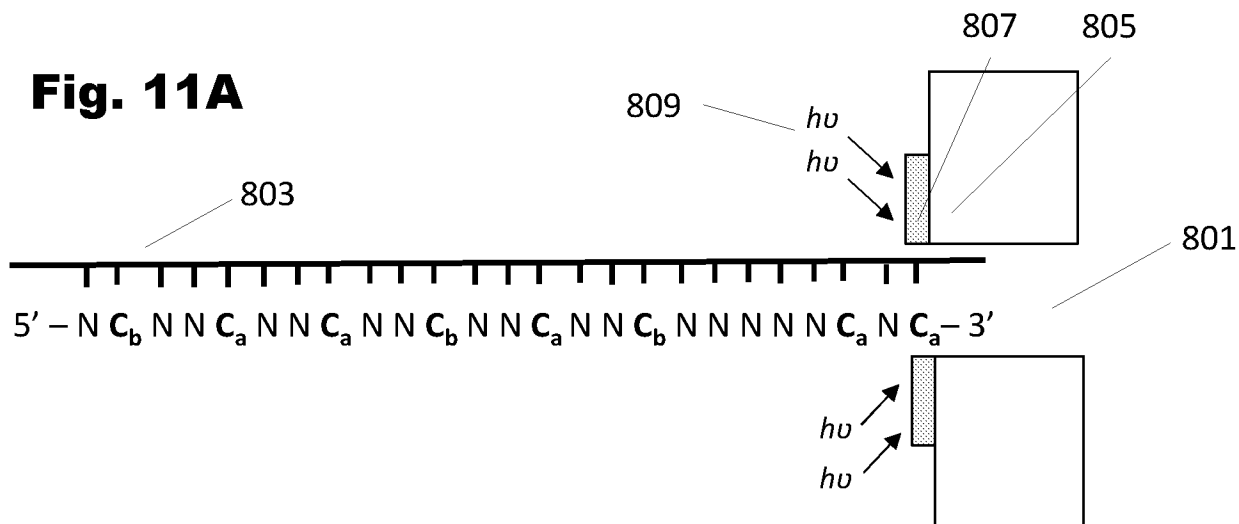
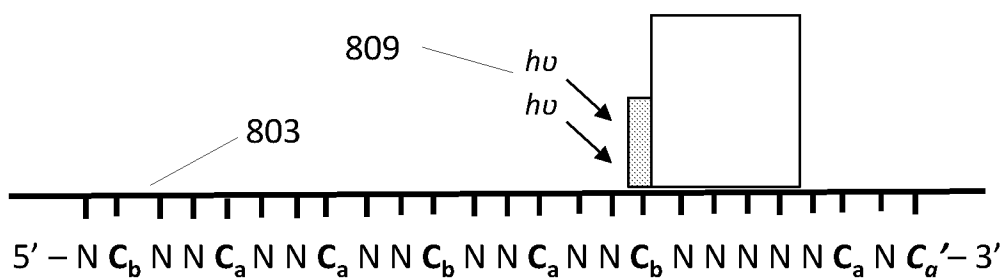
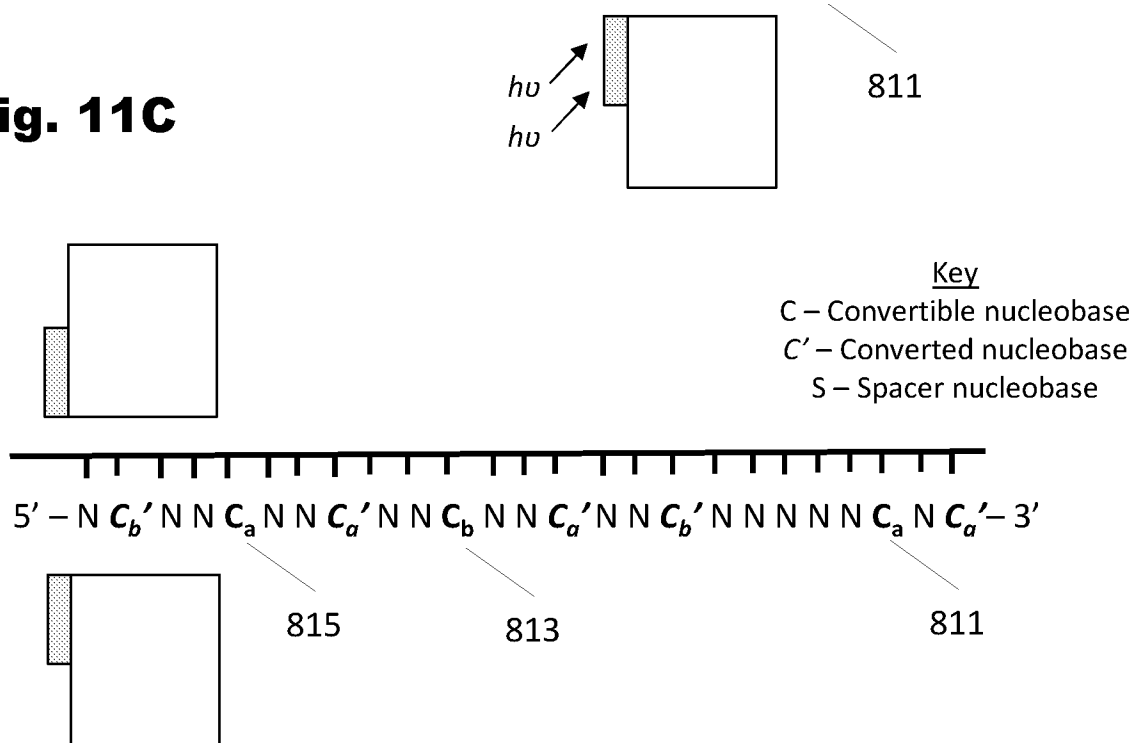
**Fig. 11A****Fig. 11B****Fig. 11C**

Fig. 12A

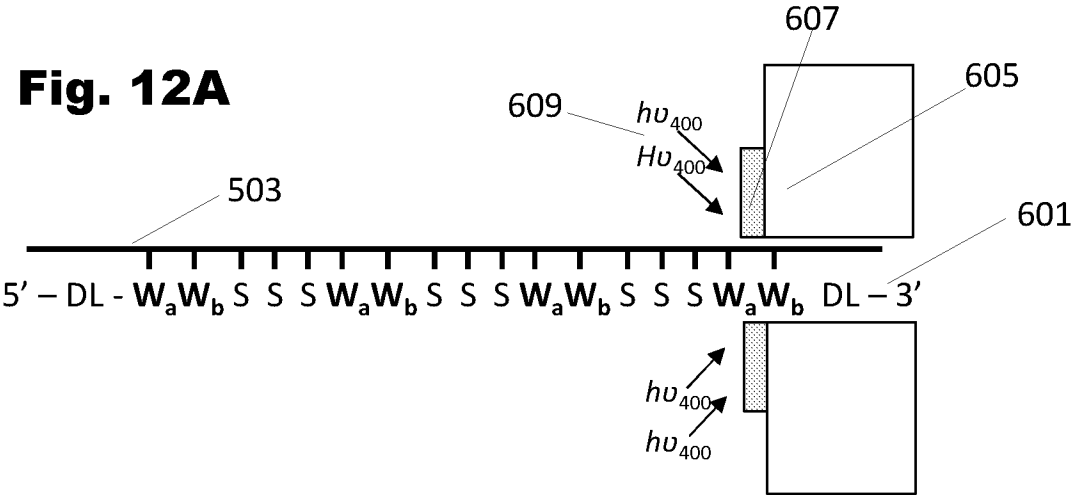


Fig. 12B

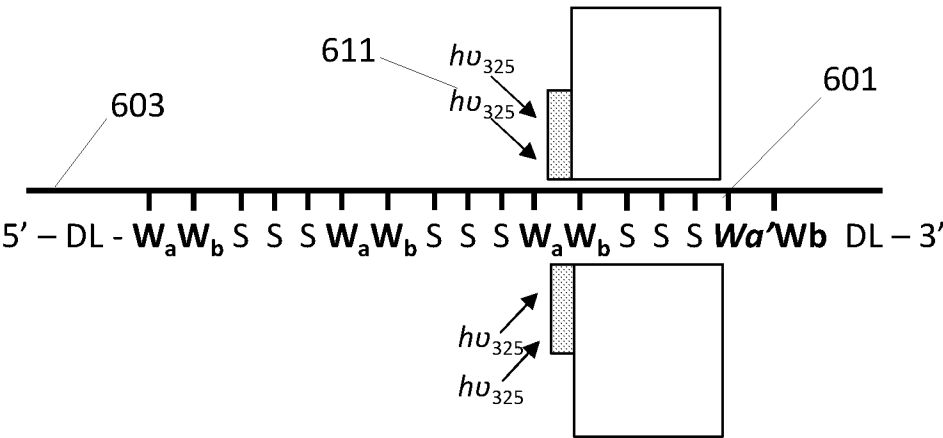
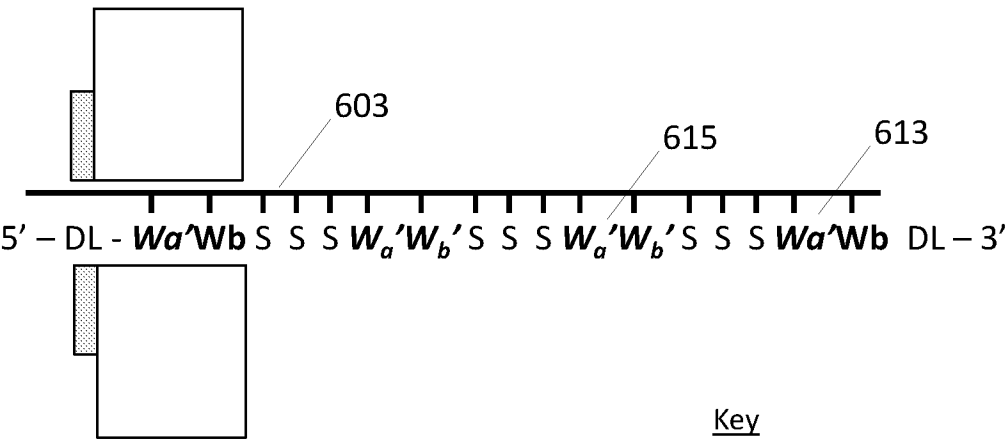
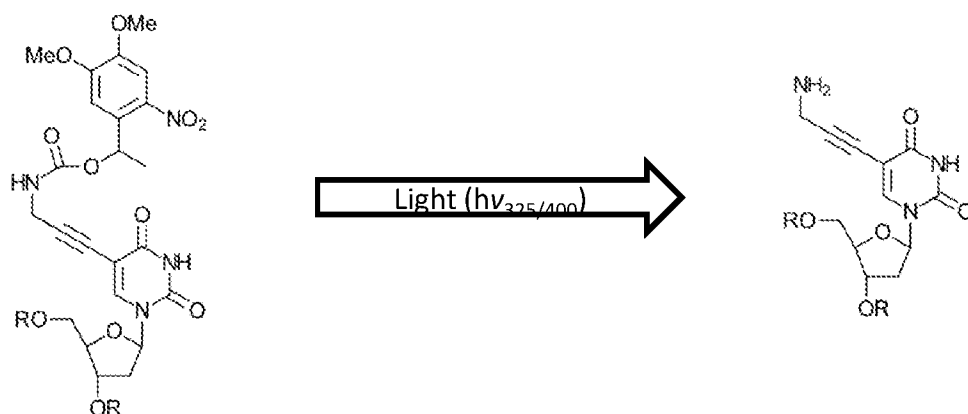
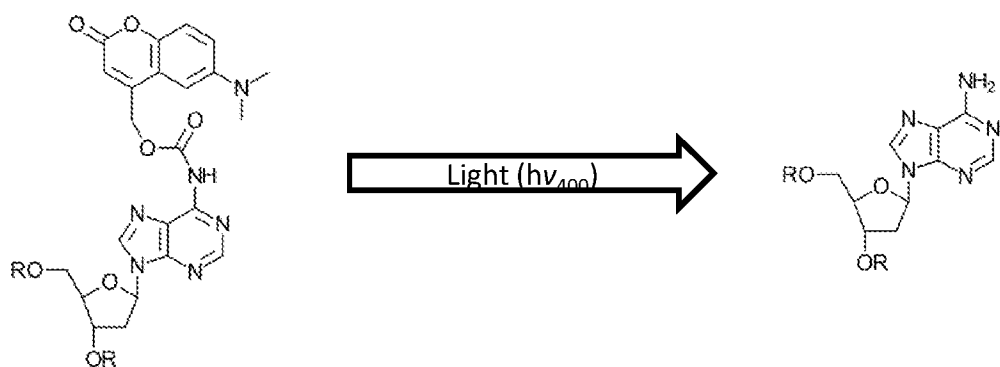
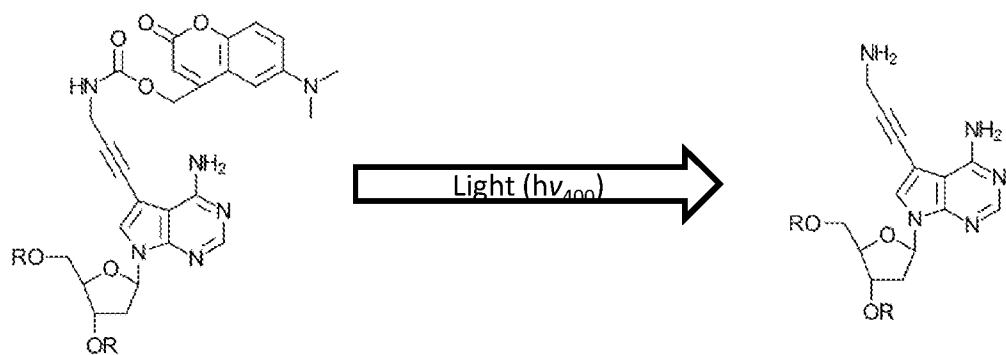
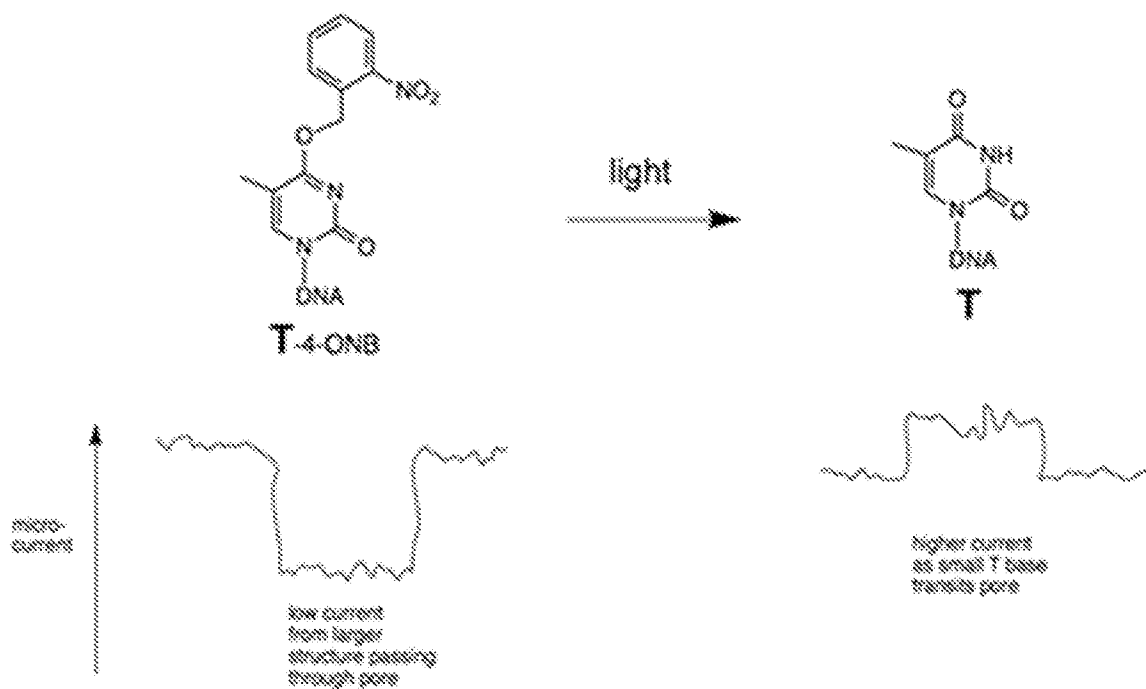
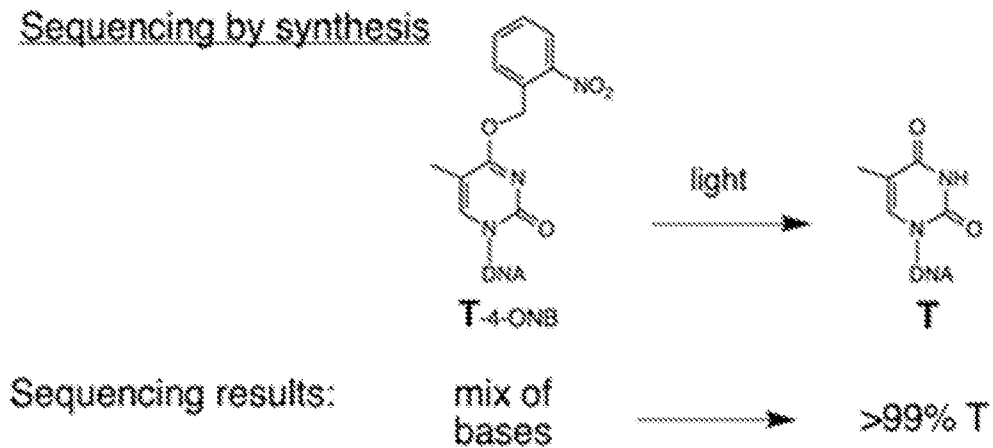


Fig. 12C

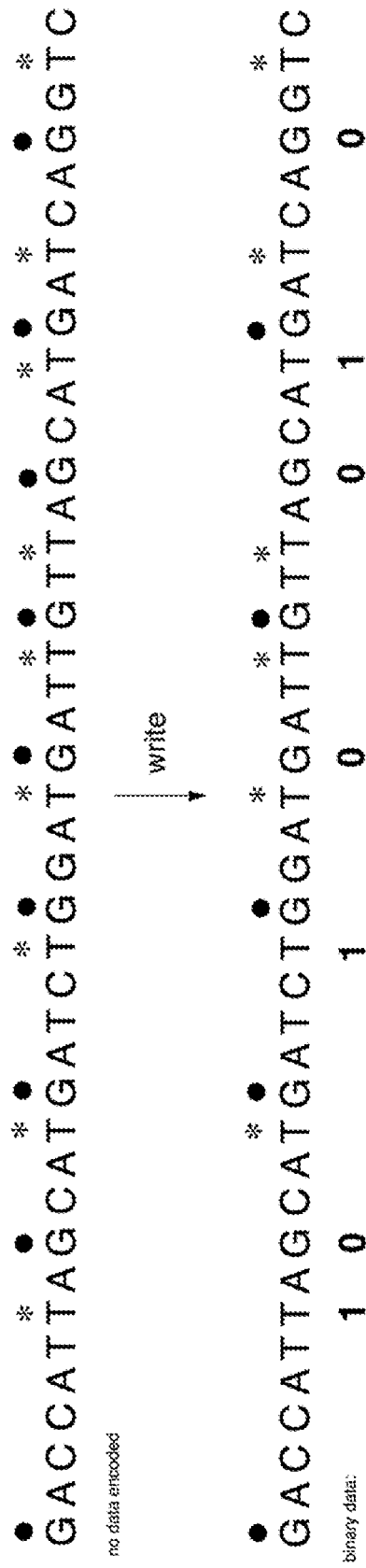


Key  
W – Convertible nucleobase  
 $W'$  – Converted nucleobase  
S – Spacer nucleobase  
DL – Delimiter sequence

**Fig. 13A****Fig. 13B****Fig. 13C**

**Fig. 14A**Nanopore current-based sequencing**Fig. 14B**Sequencing by synthesis

**Fig. 15**



## INTERNATIONAL SEARCH REPORT

International application No.  
**PCT/US2022/038591**

## A. CLASSIFICATION OF SUBJECT MATTER

**C12Q 1/6869 (2018.01) C07H 21/04 (2006.01) G11C 13/00 (2006.01)**

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)  
G11C13/0019 AND Nucleic Acid Polymer, G11C13/0019 AND photocleavage.

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

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

PATENW, CAPLUS, BIOSIS, EMBASE, GOOGLE, GOOGLE PATENTS, GOOGLE SCHOLAR, ESPACENET, & keywords: Nucleic Acid Polymer, Photoremovable Nucleic Acid Polymer, DNA storage and similar keywords.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------|-----------------------|
|           | Documents are listed in the continuation of Box C                                  |                       |

☒ Further documents are listed in the continuation of Box C

☒ See patent family annex

|                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| *<br>"A"<br>"D"<br>"E"<br>"L"<br>"O"<br>"P" | Special categories of cited documents:<br>document defining the general state of the art which is not considered to be of particular relevance<br>document cited by the applicant in the international application earlier application or patent but published on or after the international filing date<br>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)<br>document referring to an oral disclosure, use, exhibition or other means<br>document published prior to the international filing date but later than the priority date claimed | "T"<br>"X"<br>"Y"<br>"&" | 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<br>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<br>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<br>document member of the same patent family |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Date of the actual completion of the international search  
31 October 2022

Date of mailing of the international search report  
31 October 2022

## Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE  
PO BOX 200, WODEN ACT 2606, AUSTRALIA  
Email address: pct@ipaustalia.gov.au

## Authorised officer

Nirvana Nagi  
AUSTRALIAN PATENT OFFICE  
(ISO 9001 Quality Certified Service)  
Telephone No. +61 2 6225 6154

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2022/038591

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
the subject matter listed in Rule 39 on which, under Article 17(2)(a)(i), an international search is not required to be carried out, including
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

**See Supplemental Box for Details**

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

## Supplemental Box

**Continuation of: Box III**

This International Application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

This Authority has found that there are different inventions based on the following features that separate the claims into distinct groups:

- Claims 1-62 pertain to a polymer for encoding data comprising a plurality of convertible residues wherein the residues are capable of being converted from a first state to a second state and are readable by a polymerase enzyme.
- Claims 63-69 pertain to a system for data writing comprising the writable nucleic acid polymer and a data writing device.
- Claims 70-78 pertain to methods of generating writable nucleic acid polymers.
- Claims 79-98 pertain to methods for writing data onto a writable polymer.
- Claims 99-113 pertain to methods of reading or decoding data.

PCT Rule 13.2, first sentence, states that unity of invention is only fulfilled when there is a technical relationship among the claimed inventions involving one or more of the same or corresponding special technical features. PCT Rule 13.2, second sentence, defines a special technical feature as a feature which makes a contribution over the prior art.

When there is no special technical feature common to all the claimed inventions there is no unity of invention.

In the above groups of claims, the identified features may have the potential to make a contribution over the prior art but are not common to all the claimed inventions and therefore cannot provide the required technical relationship. The only feature common to all of the claimed inventions and which provides a technical relationship among them is the writeable polymer comprising a plurality of convertible residues wherein each convertible residue has a first state and is capable of being converted from a first state into a second state and read by a polymerase enzyme.

However this feature does not make a contribution over the prior art because it is disclosed in:

D1 - WO 2020/128517 A1 discloses a polymer suitable for encoding data (See Abstract). The polymer disclosed is a nucleic acid polymer and therefore readable by a polymerase enzyme.

D2 - WO 2018/222853 A1 discloses homopolymer tracts for encoding digital data (See Abstract).

D3 - WO 2017/011492 A1 discloses nucleotide polymers to store binary information (See Abstract and Page 10).

Therefore in the light of this document this common feature cannot be a special technical feature. Therefore there is no special technical feature common to all the claimed inventions and the requirements for unity of invention are consequently not satisfied *a posteriori*.

| INTERNATIONAL SEARCH REPORT                           |                                                                                                                                                                                | International application No.                                                                                                             |
|-------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                | PCT/US2022/038591                                                                                                                         |
| Category*                                             | Citation of document, with indication, where appropriate, of the relevant passages                                                                                             | Relevant to claim No.                                                                                                                     |
| X                                                     | WO 2020/128517 A1 (Oxford Nanopore Technologies LTD) 25 June 2020<br>Abstract; Page 4, lines 15-20; Figures 1, 11 and 19; Page 21, lines 29-31; Example 6, page 90; Example 2. | 1-10, 14-16, 19-22, 24, 28-29, 34-37, 39-40, 44-45, 48, 50, 60-61, 63-65, 67-69, 75-76, 79-82, 85-85, 89-91, 94, 97-103, 105-107, 109-113 |
| X                                                     | WO 2018/222853 A1 (Molecular Assemblies INC) 12 June 2018<br>Abstract; Figures 2, 6, 7; Pages 8, 17.                                                                           | Claims 1-10, 14-21, 24-26, 29-30, 34-37, 39-41, 44, 63-65, 67, 99-107                                                                     |
| X                                                     | WO 2017/011492 A1 (HAYWARD COLLEGE) 19 January 2017<br>Abstract; Pages 10, 17, 21.                                                                                             | 1-6, 8-10, 33-36, 63-65, 99-103                                                                                                           |
| Y                                                     | Abstract; Pages 10, 17, 21.                                                                                                                                                    | 1-113                                                                                                                                     |
| Y                                                     | CHARLES, L. et al., 'Design of Abiological Digital Poly(phosphodiester)s', Acc Chem Res. 2021, Vol. 54, No. 7 Pages 1791-1800.<br>Abstract, Figure 7, Page 1796.               | 1-113                                                                                                                                     |
|                                                       |                                                                                                                                                                                |                                                                                                                                           |

| INTERNATIONAL SEARCH REPORT                                                                                                                                                                                                                                           |                  | International application No. |                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|-------------------------------|------------------|
| Information on patent family members                                                                                                                                                                                                                                  |                  | PCT/US2022/038591             |                  |
| This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information. |                  |                               |                  |
| Patent Document/s Cited in Search Report                                                                                                                                                                                                                              |                  | Patent Family Member/s        |                  |
| Publication Number                                                                                                                                                                                                                                                    | Publication Date | Publication Number            | Publication Date |
| WO 2020/128517 A1                                                                                                                                                                                                                                                     | 25 June 2020     | WO 2020128517 A1              | 25 Jun 2020      |
|                                                                                                                                                                                                                                                                       |                  | CN 113167784 A                | 23 Jul 2021      |
|                                                                                                                                                                                                                                                                       |                  | EP 3899533 A1                 | 27 Oct 2021      |
|                                                                                                                                                                                                                                                                       |                  | US 2022042967 A1              | 10 Feb 2022      |
| WO 2018/222853 A1                                                                                                                                                                                                                                                     | 12 June 2018     | WO 2018222853 A1              | 06 Dec 2018      |
|                                                                                                                                                                                                                                                                       |                  | EP 3630350 A1                 | 08 Apr 2020      |
|                                                                                                                                                                                                                                                                       |                  | EP 3959330 A1                 | 02 Mar 2022      |
|                                                                                                                                                                                                                                                                       |                  | JP 2020522257 A               | 30 Jul 2020      |
|                                                                                                                                                                                                                                                                       |                  | JP 2022532995 A               | 21 Jul 2022      |
|                                                                                                                                                                                                                                                                       |                  | KR 20200027927 A              | 13 Mar 2020      |
|                                                                                                                                                                                                                                                                       |                  | US 2019040459 A1              | 07 Feb 2019      |
|                                                                                                                                                                                                                                                                       |                  | US 10982276 B2                | 20 Apr 2021      |
|                                                                                                                                                                                                                                                                       |                  | US 2019194739 A1              | 27 Jun 2019      |
|                                                                                                                                                                                                                                                                       |                  | US 11174512 B2                | 16 Nov 2021      |
|                                                                                                                                                                                                                                                                       |                  | US 2019344239 A1              | 14 Nov 2019      |
|                                                                                                                                                                                                                                                                       |                  | WO 2020219853 A1              | 29 Oct 2020      |
| WO 2017/011492 A1                                                                                                                                                                                                                                                     | 19 January 2017  | WO 2017011492 A1              | 19 Jan 2017      |
|                                                                                                                                                                                                                                                                       |                  | CN 108026557 A                | 11 May 2018      |
|                                                                                                                                                                                                                                                                       |                  | EP 3322812 A1                 | 23 May 2018      |
|                                                                                                                                                                                                                                                                       |                  | EP 3322812 B1                 | 18 May 2022      |
|                                                                                                                                                                                                                                                                       |                  | JP 2018527900 A               | 27 Sep 2018      |
|                                                                                                                                                                                                                                                                       |                  | JP 6920275 B2                 | 18 Aug 2021      |
|                                                                                                                                                                                                                                                                       |                  | JP 2021058216 A               | 15 Apr 2021      |
|                                                                                                                                                                                                                                                                       |                  | KR 20180030092 A              | 21 Mar 2018      |
|                                                                                                                                                                                                                                                                       |                  | US 2017017436 A1              | 19 Jan 2017      |
|                                                                                                                                                                                                                                                                       |                  | US 9928869 B2                 | 27 Mar 2018      |
|                                                                                                                                                                                                                                                                       |                  | US 2017337324 A1              | 23 Nov 2017      |
|                                                                                                                                                                                                                                                                       |                  | US 10289801 B2                | 14 May 2019      |
| US 2019121936 A1                                                                                                                                                                                                                                                      | 25 Apr 2019      |                               |                  |
| End of Annex                                                                                                                                                                                                                                                          |                  |                               |                  |
| Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.                                                                                                                                             |                  |                               |                  |