



(51) International Patent Classification:

C12Q 1/68 (2006.01) *C12N 9/96* (2006.01)
C07K 14/31 (2006.01) *C12N 9/12* (2006.01)

(21) International Application Number:

PCT/US20 17/028272

(22) International Filing Date:

19 April 2017 (19.04.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/325,669 21 April 2016 (21.04.2016) US

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) Inventors: **CHURCH, George M.**; 218 Kent Street, Brookline, Massachusetts 02446 (US). **PALLA, Mirko**; 35 Cloverdale Road, Apt. 2, Newton, Massachusetts 02461 (US). **STRANGES, Peter Benjamin**; 111 Woodstock Street, Apt. 5, Somerville, Massachusetts 02144 (US). **NI-**

VALA, Jeffrey Matthew; 219 Harvard Avenue, Apt. 1, Allston, Massachusetts 02134 (US).

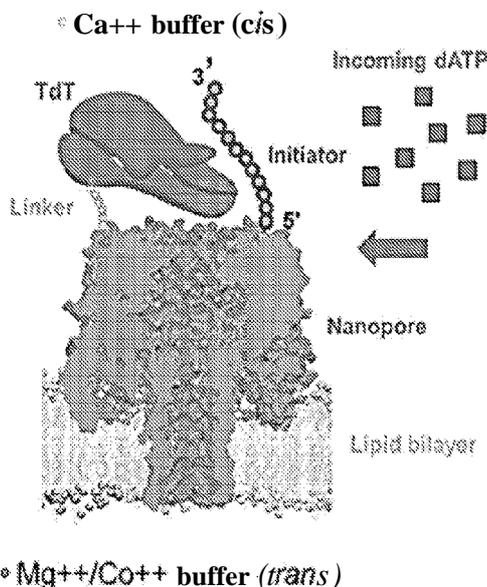
(74) Agent: **IWANICKI, John P.**; Banner & Witcoff, Ltd., 28 State Street, Suite 1800, Boston, Massachusetts 02109 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, 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,

(54) Title: METHOD AND SYSTEM OF NANOPORE-BASED INFORMATION ENCODING

FIG. 3A



(57) Abstract: This invention provides methods and systems of DNA synthesis including providing an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir, providing an electrode over the lipid bilayer wherein the electrode can modulate voltage across the lipid bilayer, wherein the enzyme catalyzes DNA synthesis in response to the voltage..



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))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

METHOD AND SYSTEM OF NANOPORE-BASED INFORMATION ENCODING**RELATED APPLICATION DATA**

This application claims priority to U.S. Provisional Application No. 62/325,669 filed on April 21, 2016 which is hereby incorporated herein by reference in its entirety for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

This invention was made with government support under Grant No. R01HG007415 awarded by National Institutes of Health. The government has certain rights in the invention.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on April 11, 2017, is named 010498_00941-WO_SL.txt and is 12,306 bytes in size.

FIELD

The present invention relates in general to methods of nanopore-based sequencing by synthesis.

BACKGROUND

The information revolution is generating large amounts of complex digital media. This has introduced multiple challenges dealing with archiving, maintenance, and retrieval of information. Nucleic acids such as DNA can be an attractive candidate for long-term information storage because of its high-density encoding capability, optimal archiving

conditions, and already established techniques to decode its stored content (Goldman, N. et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature 494, 77-80 (2013)). There remains a need for efficient and accurate nanopore-based information encoding and storage methods and systems.

SUMMARY

The present disclosure addresses this need and is based on the discovery that a scalable, nanopore-based information encoding method using an enzymatic approach can be used for information encoding. The present disclosure provides a method of DNA synthesis including providing an encoding unit including an enzyme, a single-stranded DNA (ssDNA) and a nanopore, providing a lipid bilayer having on opposite sides a first reservoir and a second reservoir, which may be referred to herein as a *cis* and a *trans* reservoir respectively, with each reservoir having a different buffer composition, wherein the nanopore provides a channel through the lipid bilayer and may be partially within the lipid bilayer and extending within the lipid bilayer, and the enzyme and the ssDNA are in the first or *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer to cause migration of or influx of activating ions from the second reservoir through the nanopore and through the lipid membrane and into the first reservoir, and wherein the enzyme is activated in the presence of the activating ions and catalyzes addition of one or more nucleotides present in the first reservoir to the single-stranded DNA.

The disclosure provides a method for making single stranded DNA using a template independent polymerase wherein activating ions are separate from the template independent polymerase, nucleotides and a template strand for extension which are present in a reaction zone. A voltage is used to cause migration of the activating ions into the reaction zone to

activate the template independent polymerase to add one or more nucleotides to a terminal end of the template strand. According to one embodiment, a nanopore separates the activating ions from the template independent polymerase, nucleotides and a template strand for extension present in a reaction zone. The nanopore is connected to a lipid bilayer which separates the activating ions from the template independent polymerase, nucleotides and a template strand for extension present in a reaction zone. Applying a voltage across the lipid bilayer causes migration of the activating ions through the nanopore and lipid bilayer and into the reaction zone including the template independent polymerase, nucleotides and a template strand thereby activating the template independent polymerase to add one or more nucleotides to the template strand.

The present disclosure further provides a system of DNA synthesis including an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore, a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir, and an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.

The present disclosure provides a method of encoding information including providing an encoding unit comprising a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase and the ssDNA are in the *cis* reservoir, providing nucleotide bases to the *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the

electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux.

The present disclosure further provides an information encoding system including an encoding unit comprising a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore, a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase, the ssDNA and nucleotide bases are in the *cis* reservoir, and an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux.

The present disclosure provides a method of encoding information including providing an encoding unit including a DNA polymerase, a circular single-stranded DNA (ssDNA) template and a protein nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the DNA polymerase and the circular ssDNA template are in the *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via DNA synthesis in response to the voltage.

The present disclosure further provides an information encoding system including an encoding unit including an enzyme, a single-stranded DNA (ssDNA) and a nanopore, a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer

composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir, and an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage. In one embodiment, the ssDNA is circular.

The present disclosure provides a method of single molecule attachment for high throughput nanopore measurement including providing an encoding unit including a single molecule, a single-stranded DNA (ssDNA) and a nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the single molecule and the ssDNA are in the *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the single molecule catalyzes DNA synthesis in response to the voltage induced ion-flux.

The present disclosure further provides a system of single molecule attachment for high throughput nanopore measurement including an encoding unit including a single molecule, a single-stranded DNA (ssDNA) and a nanopore, a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the single molecule and the ssDNA are in the *cis* reservoir, and an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the single molecule catalyzes DNA synthesis in response to the voltage.

The present disclosure provides a composition including an encoding unit comprising a terminal deoxynucleotidyl transferase attached to a protein nanopore, and a lipid bilayer wherein the protein nanopore is within the lipid bilayer.

The present disclosure provides a system including an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore, and a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir. In one embodiment, the system further comprises an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage. In another embodiment, the ssDNA is linear or circular.

The present disclosure further provides kits. The kit includes compositions including an encoding unit comprising a terminal deoxynucleotidyl transferase attached to a protein nanopore, and a lipid bilayer wherein the protein nanopore is within the lipid bilayer, and a reagent to reconstitute the terminal deoxynucleotidyl transferase, the nucleotide bases and the ssDNA. Components of a subject kit can be in separate containers; or can be combined in a single container. The kits can further include one or more additional reagents, where such additional reagents can be selected from: a dilution buffer; a reconstitution solution; a wash buffer; a control reagent; a control enzyme; and the like. A subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, flash drive, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are

provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

The present disclosure provides a method of DNA synthesis including applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises an enzyme, a single-stranded DNA (ssDNA) and a nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are conjugated to the nanopore and are in the *cis* reservoir, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage. In one embodiment, the *cis* reservoir further comprises nucleotide bases. In another embodiment, the voltage is applied via an electrode that modulates the voltage across the lipid bilayer.

The present disclosure provides a method of encoding information including applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase and the ssDNA are conjugated to the nanopore and are in the *cis* reservoir, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux. In one embodiment, the *cis* reservoir further comprises nucleotide bases. In another

embodiment, the voltage is applied via an electrode that modulates the voltage across the lipid bilayer. In one embodiment, the ssDNA is linear.

The present disclosure provides a method of encoding information including applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises a DNA polymerase, a circular single-stranded DNA (ssDNA) template and a protein nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the DNA polymerase and the circular ssDNA template are conjugated to the nanopore and are in the *cis* reservoir, and wherein the information is encoded via DNA synthesis in response to the voltage. In one embodiment, the voltage is applied via an electrode that modulates the voltage across the lipid bilayer. In another embodiment, the *cis* reservoir contains dATP, dGTP, or dTTP, and wherein the *trans* reservoir contains dCTP.

The disclosure provides that the enzyme is conjugated to the nanopore or free in the *cis* reservoir. The disclosure provides that the enzyme is a template dependent or independent DNA polymerase. The disclosure provides that the DNA polymerase is a ϕ 29 DNA polymerase or a terminal deoxynucleotidyl transferase. The disclosure provides that the ssDNA is immobilized to the nanopore. The disclosure provides that the ssDNA is immobilized around the rim of the nanopore at its 5' end. The disclosure provides that the ssDNA is a template or an initiator for DNA synthesis. The disclosure provides that the nanopore is a protein membrane channel. The disclosure provides that the protein membrane channel comprises aHL, MspA, or ϕ 29 connector. The disclosure provides that the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes

base addition during DNA synthesis. The disclosure provides that the method and system further includes providing nucleotide bases to the *cis* reservoir and turning on or otherwise applying the voltage such that catalytic ions (Mg^{2+}/Co^{2+}) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes nucleotide base addition during DNA synthesis. The disclosure provides that the nucleotide bases can be added as a single base or a homopolymer run during DNA synthesis. The disclosure provides that the nucleotide bases comprise dATP, dTTP, dCTP, or dGTP. The disclosure provides that the method and system further includes turning off or discontinuing application of the voltage and flushing the *cis* reservoir with fresh non-catalytic buffer and removing any nucleotide bases and catalytic ions. The disclosure provides that the method and system is capable of single molecule measurement. The disclosure provides that the terminal deoxynucleotidyl transferase is conjugated to the protein nanopore or free in the *cis* reservoir.

The disclosure provides that the information is encoded at single-base resolution or homopolymer resolution. The disclosure provides that the information encoding accommodates 12 possible 2-base state transitions. The disclosure provides that the information encoding accommodates $N*12$ unique N-base state transitions. The disclosure provides that the ssDNA is linear or circular for information encoding or storage. The disclosure provides that the information is encoded and stored in the product DNA. The disclosure provides that the DNA polymerase is conjugated to the protein nanopore or free in the *cis* reservoir. The disclosure provides that circular ssDNA template is immobilized to the protein nanopore. The disclosure provides that the circular ssDNA template includes cytosine, adenine, thymine, and inosine but not guanine. The disclosure provides that the DNA polymerase is a $\phi 29$ DNA polymerase. The disclosure provides that the protein nanopore is a protein membrane channel. The disclosure provides that the *cis* reservoir contains dATP, dGTP, or dTTP, and wherein the *trans* reservoir contains dCTP. The

disclosure provides that the dCTP flows to the *cis* reservoir and is incorporated into a complement strand of DNA when the voltage is positive. The disclosure provides that the dATP, dGTP or dTTP is incorporated into a complement strand of DNA when the voltage is negative. The disclosure provides that the method records voltage polarity applied across the lipid bilayer over time by encoding this information into a complement strand of DNA during DNA synthesis. The disclosure provides that the information is stored in the complement strand in the form of regions of bases either containing or lacking cytosine bases at the sites which complement the inosine bases of the template strand, which translates to the "1" or "0" state of a single bit of information. The disclosure provides that the enzyme is conjugated to the nanopore or free in the *cis* reservoir. The disclosure provides that the nucleotide base can be added as a single base or a homopolymer run during DNA synthesis. The disclosure provides that the system is capable of single molecule measurement. The disclosure provides that the single molecule is a template dependent or independent DNA polymerase. The disclosure provides that the single molecule is conjugated to the nanopore by protein conjugation, such as SpyCatcher/SpyTag protein conjugation. The disclosure provides that the enzyme such as the DNA polymerase is covalently coupled to the α -hemolysin heptamer using protein conjugation, such as a Spy/Catcher/SpyTag conjugation approach. The disclosure provides that the electrode is a complementary metal-oxide-semiconductor (CMOS)-based electrode array for high-throughput electrical recording.

The disclosure provides a composition including an encoding unit comprising a terminal deoxynucleotidyl transferase attached to a protein nanopore, and a lipid bilayer wherein the protein nanopore is within the lipid bilayer. The composition further comprises a single stranded DNA (ssDNA). The disclosure provides that the ssDNA is linear.

It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

Further features and advantages of certain embodiments of the present invention will become more fully apparent in the following description of embodiments and drawings thereof, and from the claims.

INCORPORATION BY REFERENCE

All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee. The foregoing and other features and advantages of the present embodiments will be more fully understood from the following detailed description of illustrative embodiments taken in conjunction with the accompanying drawings in which:

FIGS. 1A-B depict simplified models for template-dependent and template-independent DNA polymerase activity. FIG. 1A shows that most DNA polymerases require double-stranded DNA as a substrate, where the 5'→3' strand is used as a primer and the complementary strand 3'→5' is used as a template. FIG. 1B shows that terminal deoxynucleotidyl transferase is capable of catalyzing phosphoryl transfer in the absence of a template that cannot be accommodated in its active site. Adapted from Motea, E. A. & Berdis, A. J. Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochim. Biophys. Acta - Proteins Proteomics* 1804, 115 1-1 166 (20 10).

FIGS. 2A-B depict schematics of an encoding unit. FIG. 2A shows terminal deoxynucleotidyl transferase (TdT) attachment strategy to aHL nanopore via SpyCather-SpyTag conjugation. 5'-immobilized initiator single-stranded DNA strand is also shown, which is being extended by catalysis in the presence of divalent Mg^{2+}/Co^{2+} ion containing buffer. FIG. 2B shows that each monomeric unit can accommodate a TdT enzyme (assembling into a heptameric unit), which can extend each ssDNA tethered to the pore rim with the same information content based on voltage-mediated ion-flux modulation.

FIGS. 3A-D depict a TdT-based gene synthesis scheme by ion-flux modulation through a nanopore array. FIG. 3A shows to flow in dATP (or other natural nucleotide, one at a time) in non-catalytic buffer (Ca^{2+} or other non- Mg^{2+}/Co^{2+} allowing binding but not covalent bond catalysis). FIG. 3B shows to turn the voltage gradient on across membrane (ON state), such that a plume of catalytic ions (Mg^{++}/Co^{++}) is injected to the *cis* side (from the *trans* side). Addition of a single base or a homopolymer run to the initiator ssDNA will take place due to catalysis. FIG. 3C shows that if the voltage gradient is not applied (OFF state), catalytic ion migration will not occur, which will lead to no addition of base to ssDNA.

FIG. 3D shows to flush *cis* compartment with fresh non-catalytic buffer to remove any nucleotides and catalytic ion residues. This cycle is repeated with the next nucleotide to be encoded into the ssDNA as desired.

FIGS. 4A-C depict numerical simulations of the steady-state ions spatial distribution near the nanopore. FIG. 4A shows a finite element 2D axisymmetric model of a nanopore surrounded by $1\ \mu\text{m}$ of *cis* and *trans* chamber in all directions from the pore. FIGS. 4B-C show the space-dependent concentration of Ca^{2+} ions along the z-axis in the vicinity of the pore, as a function of applied voltage (FIG. 4B, bulk Ca^{2+} concentration in *trans* is 1 M) or as a function of Ca^{2+} concentration at a fixed applied voltage of 0.3 V (FIG. 4C). Adapted from Anderson, B. N. et al. Probing Solid-State Nanopores with Light for the Detection of Unlabeled Analytes. ACS Nano 8, 11836-1 1845 (20 14).

FIGS. 5A-D depict simultaneous electrical and optical readout of ionic current in nanopores. FIG. 5A shows a schematic of the measurement setup. Both chambers of the experimental cell are equipped with Ag/AgCl electrodes for the electrical recording of the ionic current through the pore. FIG. 5B shows a zoomed-in view of the nanopore area. The optical detection of the ionic current is achieved by adding Ca^{2+} ions to the top (*trans*) chamber and Ca^{2+} ionophore to the bottom (*cis*) chamber of the cell. FIG. 5C shows the chemical structure of the ionophore. A voltage across the membrane drives Ca^{2+} ions to the *cis* side where they complex with the ionophore, resulting in fluorescence signal. The fluorescence intensity is proportional to the Ca^{2+} concentration and, thus, the flow of ions through the pore. Subsequently, a drop in the ionic current during a translocation event can be simultaneously detected from the decrease in the fluorescence intensity FIG. 5D. Adapted from Ivankin, A. et al. Label-Free Optical Detection of Biomolecular Translocation through Nanopore Arrays. ACS Nano 8, 10774-1078 1 (2014).

FIG. 6 shows various views of a demonstration of maximum packing density of a single electrode array surface area with aHL nanopores. Not to scale.

FIG. 7 depicts schematics of information encoding and/or decoding of the nanopore-based system.

FIG. 8 depicts schematics of information encoding and/or decoding of the nanopore-based system.

FIG. 9 depicts a schematic of circular inosine template with no Gs (SEQ ID NO: 3). dCTP (only present in the *trans* solution) flows through the pore and is incorporated into the complement strand during polymerization only if the polarity of the applied voltage (*cis* positive) forces it through the pore into the proximity of the polymerase.

FIG. 10 depicts a schematic of a voltage recorder embodiment of the nanopore-based system. Figure 10 discloses SEQ ID NOS 3, 4, and 5, respectively, in order of appearance.

FIG. 11 shows that various types of membrane channels can be used to tether a single enzyme with the proposed strategy for single molecule ion current measurement.

FIGS. 12A-D depict the assembly of the porin-polymerase construct. FIG. 12A shows the protein constructs used to form the porin-polymerase conjugate include unmodified aHL with a Strep-tag, aHL with a C-terminal SpyTag peptide and 6x-His-tag (SEQ ID NO: 1), and 029 with a C-terminal SpyCatcher domain. Assembly steps are shown in FIG. 12B. aHL-SpyTag-6x-His ("6x-His" disclosed as SEQ ID NO: 1) and unmodified aHL are oligomerized with lipid and the 1:6 SpyTagunmodified assembled porin is purified. Addition of 029-SpyCatcher to the 1:6 pore yields one polymerase per aHL pore. A molecular model generated with Rosetta using the determined structures for ϕ 29 polymerase (PDB: 2PYJ), aHL (PDB: 7AHL) and SpyCatcher/SpyTag (PDB: 2X5P) is shown in FIG. 12C. Colors of

the proteins match the schematic representations in FIGS. 23A&B. The expected tag exit site on the polymerase and the opening to the nanopore can be in close proximity with distances as short as 46 Å in some models. The stoichiometry in solution of the porin subunits was confirmed by SDS PAGE without boiling. FIG. 12D. To confirm the assembly, excess $\phi 29$ -SpyCatcher was added to 1:6 pore. The combination yields only pores with one polymerase attached.

FIGS. 13A-D show representative current *versus* time traces for the various stages of the pore assembly. When neither tagged nucleotide nor polymerase is present, only stable open channel current is observable. FIG. 13A. Attachment of polymerase does not change the mean open channel current. FIG. 13B. The current root-mean-square fluctuation (RMSF) increase in FIG. 13B may be an indication of the polymerase coupled to the pore. When no polymerase is attached to the pore and tagged nucleotide is introduced transient events are observed. FIG. 13C. When polymerase-template is attached to the pore and the complementary base dG6P-dT30 ("dT30" disclosed as SEQ ID NO: 2) is added there are prolonged capture events. FIG. 13D.

FIGS. 14A-B show base discrimination of the four tagged nucleotides on a semiconductor chip array. All measurements were taken on a pore-polymerase-template complex under non-catalytic conditions where the first base on the template is complementary to the added tagged nucleotide. FIG. 14A shows current *versus* dwell time (duration of each current blockade) plots for captures of all tagged nucleotides. Capture events cluster into distinct current and dwell time regions for each tagged-nucleotide. FIG. 14B shows representative single pore traces of tagged-nucleotide capture shown in FIG. 14A. Current blockade levels for each are marked in red. The blockades demonstrate unique, single-molecule events corresponding to the four distinct tag captures.

FIGS. 15A-B show representative examples of real-time detection of numerous successive tagged nucleotide incorporations into a self-priming DNA hairpin template catalyzed by nanopore-bound polymerase on the Genia chip. Two base captures of tagged C and G nucleotides with standard A and T nucleotides, are shown in FIG. 15A. Part of the template sequence (SEQ ID NO: 6), with homopolymers grouped into a single base, is shown in red. The only captures observed in the trace match the expected levels for dG6P-T30 ("T30" disclosed as SEQ ID NO: 2) and dC6P-dSp3. Four base sequencing is shown in FIG. 15B. Events were categorized by manually assigning current blockade signatures to their respective tag capture boxes. Of the 12 bases sequenced, 85% match the template strand (SEQ ID NO: 7), with one incorrectly identified base, and one deletion. Figure 15B also discloses SEQ ID NO: 8. (*See, e.g.*, WO2013191793 for general descriptions of real-time detection of successive tagged nucleotide incorporations into a DNA template catalyzed by nanopore-bound polymerase on the chip.)

FIG. 16 shows sequences of ϕ 29 (SEQ ID NO: 9) and α HL (SEQ ID NO: 10) constructs with colored annotations for the various protein sequence regions. Figure 16 also discloses SEQ ID NO: 1.

DETAILED DESCRIPTION

The present disclosure provides a nanopore-based information encoding method using an enzymatic approach for various applications. According to one aspect, the disclosure provides a method of nucleic acid synthesis including applying a voltage to cause migration of activating ions from a first location to a reaction zone including a template independent polymerase, nucleotides and an initiator strand including a terminal nucleotide, wherein the activating ions activate the template independent polymerase to add one or more nucleotides to the terminal nucleotides of the initiator strand. In one embodiment, the template

independent polymerase and the initiator are co-localized on a substrate. In one embodiment, the reaction zone is separated from the first location by a nanopore or nanochannel under influence of an electric field. In another embodiment, the reaction zone is separated from the first location by a nanopore or nanochannel having a portion of the nanopore or nanochannel within a lipid bilayer under influence of an electric field. In one embodiment, the reaction zone is separated from the first location by a nanopore or nanochannel having a portion of the nanopore or nanochannel within a lipid bilayer under influence of an electric field, and wherein application of the electric field causes migration of the activating ions from the first location through the nanopore or nanochannel and into the reaction zone wherein the template independent polymerase is activated to add one or more nucleotides to the terminal nucleotide of the initiator strand.

According to certain aspects, the present disclosure provides methods and systems of DNA synthesis. The methods of DNA synthesis include providing an encoding unit including an enzyme, a single-stranded DNA (ssDNA) and a nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage. In some embodiments, the enzyme can be conjugated to the nanopore or in the *cis* reservoir. In some embodiments, the lipid bilayer has two reservoirs on either side such that each reservoir contains a different composition of various types of buffer, salt, or concentrations known to a skilled in the art. In one embodiment, the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis. In another embodiment,

nucleotide bases are provided to the *cis* reservoir and the voltage is turned on such that catalytic ions (Mg^{2+}/Co^{2+}) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes nucleotide base addition during DNA synthesis.

According to certain other aspects, the present disclosure provides methods and systems of encoding information. The methods of information encoding include providing an encoding unit comprising a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase and the ssDNA are in the *cis* reservoir, providing nucleotide bases to the *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux. In one embodiment, the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition, and the *trans* reservoir contains catalytic buffer (Mg^{2+}/Co^{2+}) that promotes base addition. In another embodiment, the terminal deoxynucleotidyl transferase is activated by catalytic ions (Mg^{2+}/Co^{2+}) that are injected to the *cis* side from the *trans* side when the voltage is turned on and catalyzes the addition of the nucleotide base to the ssDNA. In one embodiment, the methods include providing an encoding unit comprising a DNA polymerase, a circular single-stranded DNA (ssDNA) template and a protein nanopore, providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the DNA polymerase and the circular ssDNA template are in the *cis* reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the

information is encoded via DNA synthesis in response to the voltage. In one embodiment, the *cis* reservoir contains dATP, dGTP, or dTTP, and the *trans* reservoir contains dCTP. In one embodiment, the dCTP flows to the *cis* reservoir and is incorporated into a complement strand of DNA when the voltage is positive. In one embodiment, the dATP, dGTP or dTTP is incorporated into a complement strand of DNA when the voltage is negative. In certain embodiments, the methods and systems can record voltage polarity applied across the lipid bilayer over time by encoding this information into a complement strand of DNA during DNA synthesis. In certain embodiments, the methods and systems of nucleic acid synthesis or information encoding are used in a microfluidic system wherein the *cis* or *trans* buffering reservoir corresponds to microfluidic chamber of the microfluidic system and wherein the activating or inactivating ions and nucleotide bases are flowed to the corresponding microfluidic chamber.

According to certain aspects, a composition is provided that comprises an encoding unit comprising a terminal deoxynucleotidyl transferase attached to a protein nanopore, and a lipid bilayer wherein the protein nanopore is within the lipid bilayer. In one embodiment, the composition further comprises a single stranded DNA (ssDNA). In one embodiment, the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, and the terminal deoxynucleotidyl transferase, the ssDNA and nucleotide bases are in the *cis* reservoir. In one embodiment, an electrode can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the terminal deoxynucleotidyl transferase catalyzes nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux. In some embodiments, the terminal deoxynucleotidyl transferase is conjugated to the protein nanopore or free in the *cis* reservoir. In some embodiments, multiple terminal deoxynucleotidyl transferase molecules are tethered to one or more nanopore subunits.

The term "nanopore," as used herein, generally refers to a pore, nanopore sensor, nanogap, nanochannel, channel or passage formed or otherwise provided in a membrane. A membrane may be an organic membrane, such as a lipid bilayer, or a synthetic membrane, such as a membrane formed of a polymeric material. The membrane may be a polymeric material. The nanopore may be disposed adjacent or in proximity to a sensing circuit or an electrode coupled to a sensing circuit, such as, for example, a complementary metal-oxide semiconductor (CMOS) or field effect transistor (FET) circuit. In some examples, a nanopore has a characteristic width or diameter on the order of 0.1 nanometers (nm) to about 1000 nm. Some nanopores are proteins. Alpha hemolysin is an example of a protein nanopore. Exemplary nanopores include a hole or passage through a membrane formed by a multimeric protein ring. Typically, the passage is 0.2-25 nm wide. Nanopores, as used herein, may include transmembrane structures that may permit the passage of molecules through a membrane. Examples of nanopores include α -hemolysin (*Staphylococcus aureus*) and MspA (*Mycobacterium smegmatis*). Other examples of nanopores may be found in the art describing nanopore sequencing or described in the art as pore-forming toxins, such as the β -PFTs Pantan-Valentine leukocidin S, aerolysin, and Clostridial Epsilon-toxin, the α -PFTs cytolysin A, the binary PFT anthrax toxin, or others such as pneumolysin or gramicidin. Nanopores have become technologically and economically significant with the advent of nanopore sequencing technology. Methods for nanopore sequencing are known in the art, for example, as described in U.S.P.N. 5,795,782, which is incorporated by reference. Briefly, nanopore detection involves a nanopore-perforated membrane immersed in a voltage-conducting fluid, such as an ionic solution including, for example, KCl, NaCl, NiCl, LiCl or other ion forming inorganic compounds known to those of skill in the art. A voltage is applied across the membrane, and an electric current results from the conduction of ions through the nanopore. When the nanopore interacts with polymers, such as DNA, flow

through the nanopore is modulated in a monomer-specific manner, resulting in a change in the current that permits identification of the monomer(s). Nanopores within the scope of the present disclosure include solid state nonprotein nanopores known to those of skill in the art and DNA origami nanopores known to those of skill in the art. Such nanopores provide a nanopore width larger than known protein nanopores which allow the passage of larger molecules for detection, while still being sensitive enough to detect a change in ionic current when the complex passes through the nanopore.

The term "nucleic acid," as used herein, generally refers to a molecule comprising one or more nucleic acid subunits. A nucleic acid may include one or more subunits selected from adenosine (A), cytosine (C), guanine (G), thymine (T), uracil (U) and inosine (I), or modification and variants thereof. A nucleotide can include A, C, G, T, U or I, or variants thereof. A nucleotide can include any subunit that can be incorporated into a growing nucleic acid strand. In some examples, a nucleic acid is deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), or derivatives thereof. A nucleic acid may be single-stranded or double stranded. Nucleic acid molecules may optionally include one or more non-standard nucleotide(s), nucleotide analog(s) and/or modified nucleotides. Examples of modified nucleotides include, but are not limited to diaminopurine, S²T, 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-D46-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-

thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, 2,6-diaminopurine and the like. Nucleic acid molecules may also be modified at the base moiety (e.g., at one or more atoms that typically are available to form a hydrogen bond with a complementary nucleotide and/or at one or more atoms that are not typically capable of forming a hydrogen bond with a complementary nucleotide), sugar moiety or phosphate backbone.

The term "polymerase," as used herein, generally refers to any enzyme capable of catalyzing a polymerization reaction, and variants, mutants, or homologues thereof. Examples of polymerases include, without limitation, a DNA or RNA polymerase, a terminal deoxynucleotidyl transferase (TdT), a transcriptase or a ligase, and variants, mutants, or homologues thereof. A polymerase can be a polymerization enzyme. In certain embodiments, the enzymes capable of catalyzing a polymerization reaction include template-dependent or template-independent polymerases. In exemplary embodiments, suitable template-independent polymerase includes a terminal deoxynucleotidyl transferase (TdT), and variants, mutants, or homologues thereof. In certain exemplary embodiments, mutant TdTs having mutation residues introduced in the enzyme's catalytic sites can be used. In certain other embodiments, a single-stranded DNA binding protein (SSB) and TdT (SSB-TdT) fusion protein having improved processivity can be used. In the SSB-TdT fusion protein, the SSB portion binds with high affinity in a cooperative manner to the single-stranded DNA while the TdT portion of the fusion protein retains the processivity and catalyzes continuous nucleotide base addition to the 3' end of the initiator ssDNA, thus enhancing gene synthesis. In certain other embodiments, suitable template-dependent polymerases include any strand displacing DNA polymerase, such as $\phi 29$ and *Bst* (full length or large fragment), and variants, mutants, or homologues thereof. In some embodiments,

these template-dependent polymerases include mutations at single or multiple residue sites to improve processivity, similar to TdT mutants *via* protein engineering. Selection of suitable DNA polymerases are known to a skilled in the art and are available in various commercial sources such as in the tools and resources section of the website of New England Biolabs, Inc.

Nanopore-Based Ion-Flux Modulator Methods and Systems

Described herein are methods, devices and systems for information encoding (or DNA synthesis) using, or with the aid of, one or more nanopores. The one or more nanopores may be inserted in a membrane (e.g., lipid bi-layer) that is disposed adjacent or in sensing proximity to an electrode that is part of, or coupled to, an integrated circuit.

In certain examples, a nanopore device includes a single nanopore in a membrane that is adjacent or sensing proximity to an electrode. In other examples, a nanopore device includes at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000, or 10,000 nanopores in proximity to a sensor circuit or sensing electrodes. The one or more nanopore may be associated with an individual electrode and sensing integrated circuit or a plurality of electrodes and sensing integrated circuits.

A nanopore-based ion-flux modulator system may include a reaction chamber that includes one or more nanopore devices. A nanopore device may be an individually addressable nanopore device (e.g., a device that is capable of detecting a signal and providing an output independent of other nanopore devices in the system). An individually addressable nanopore can be individually readable. In some cases, an individually addressable nanopore can be individually writable. As an alternative, an individually addressable nanopore can be individually readable and individually writable. The system can include one or more

computer processors for facilitating sample preparation and various operations of the disclosure, such as nucleic acid synthesis. The processor can be coupled to nanopore device.

A nanopore device may include a plurality of individually addressable sensing electrodes. Each sensing electrode can include a membrane adjacent to the electrode, and one or more nanopores in the membrane. A plurality of nanopore devices may be arranged in an array.

An enzyme (e.g., DNA polymerase, RNA polymerase, ligase) may incorporate nucleotides to a growing polynucleotide chain. Enzymes (e.g., polymerases) provided herein can generate polymer chains.

FIGS. 1A-B schematically depict simplified models for template-dependent and template-independent DNA polymerase activity. In the template dependent mode, incoming dNTPs are incorporated into the growing strand that is complementary to the template strand. In the template independent mode, template independent polymerase such as a terminal deoxynucleotidyl transferase catalyzes phosphoryl transfer in the absence of a template.

The present disclosure provides methods and systems for nucleic acid sequencing by synthesis with the aid of a nanopore. The nanopore may be formed or otherwise embedded in a membrane disposed adjacent to a sensing electrode of a sensing circuit, such as an integrated circuit. The integrated circuit may be an application specific integrated circuit (ASIC). In some examples, the integrated circuit is a field effect transistor or a complementary metal-oxide semiconductor (CMOS). The sensing circuit may be situated in a chip or other device having the nanopore, or off of the chip or device, such as in an off-chip configuration. The semiconductor can be any semiconductor, including, without limitation, Group IV (e.g., silicon) and Group III-V semiconductors (e.g., gallium arsenide).

In some cases, the sensing circuit detects an electrical signal associated with the nucleic acid or tag. The nucleic acid may be a subunit of a larger strand. The tag may be a byproduct of a nucleotide incorporation event or other interaction between a tagged nucleic acid and the nanopore or a species adjacent to the nanopore, such as an enzyme that cleaves a tag from a nucleic acid. The tag may remain attached to the nucleotide. Tags for nucleotides are known to a skilled in the art. For example, tags are described in US Patent Application No. 2014/0134616, which is hereby incorporated by reference in its entirety. A detected signal may be collected and stored in a memory location, and later used to construct a sequence of the nucleic acid. The collected signal may be processed to account for any abnormalities in the detected signal, such as errors. FIGS. 2-6 show various examples and schematics of the nanopore-based nucleic acid sequencing by synthesis.

The present disclosure provides enzymes e.g., a DNA or RNA polymerase, a transcriptase or a ligase in the nanopore-based nucleic acid sequencing by synthesis with the tagged nucleotides as described herein. In some cases, the enzyme is any enzyme that creates a nucleic acid strand by phosphate linkage of nucleotides. In some cases, the DNA polymerase is θ N polymerase or a variant thereof, E. Coli DNA polymerase I, Bacteriophage T4 DNA polymerase, Sequenase, Taq DNA polymerase, θ N polymerase (exo-)A485L/Y409V, phi29 DNA Polymerase (ϕ 29 DNA Polymerase), Bst polymerase, or variants, mutants, or homologues thereof. A homologue can have any suitable percentage homology, including without limitation at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or at least about 95% sequence identity.

The polymerase can have kinetic rate profile that is suitable for detection of the tags by the nanopore. The rate profile can refer to the overall rate of nucleotide incorporation

and/or a rate of any step of nucleotide incorporation such as nucleotide addition, enzymatic isomerization such as to or from a closed state, cofactor binding or release, product release, incorporation of nucleic acid into the growing nucleic acid, or translocation.

The kinetics of the enzyme can also be affected and/or controlled by manipulating the composition of the buffer in contact with the enzyme. For example, non-catalytic divalent ions (e.g., ions that do not promote polymerase function such as Ca^{2+}) can be mixed with catalytic divalent ions (e.g., ions that promote polymerase function such as Mg^{2+} /or Co^{2+}) to stop or slow the polymerase down. The buffer conditions that promote the TdT polymerase function are well established in the art and are known to a skilled in the art. In certain exemplary embodiments, the reaction condition that promotes the TdT polymerase function is described according to the commercial vendor (NEB) in the typical DNA tailing reaction (continuous ssDNA extension) as follows:

Step 1. Mix:

- a. 5.0 μl 10X TdT Buffer
- b. 5.0 μl 2.5 mM CoCh solution provided
- c. 5.0 pmols DNA (330 ng for 100 bp, 1 μg for 300 bp, 10 pmols DNA ends)
- d. 0.5 μl 10 mM dNTP (alpha-32P dATP may also be used)
- e. 0.5 μl Terminal Transferase (20 units/ μl) deionized H₂O to a final volume of 50 μl .

Step 2. Incubate at 37°C for 30 minutes, where IX Terminal Transferase Reaction Buffer is as follows:

50 mM Potassium Acetate

20 mM Tris-acetate

10 mM Magnesium Acetate

pH 7.9 @ 25°C

A skilled in the art can optimize and adjust the reaction condition for optimum result.

The disclosure provides that an enzyme may be attached to the nanopore. Suitable methods for attaching the enzyme to the nanopore include cross-linking such as the formation of intra-molecular disulfide bonds. The nanopore and the enzyme may also be a fusion protein (i.e., encoded by a single polypeptide chain). Methods for producing fusion proteins may include fusing the coding sequence for the enzyme in frame and adjacent to the coding sequence for the nanopore (without a stop codon in between) and expressing this fusion sequence from a single promoter. In some examples, the enzyme may be attached or otherwise coupled to the nanopore using molecular staples or protein fingers. In some cases, the enzyme is attached through an intermediate molecule or linker, such as for example biotin conjugated to both the enzyme and the nanopore with streptavidin tetramers linked to both biotins. The enzyme can also be attached to the nanopore with an antibody. In some cases, proteins that form a covalent bond between each other (e.g., the SpyTagTM/SpyCatcherTM system) are used to attach the polymerase to the nanopore. In some cases, phosphatase enzymes or an enzyme that cleaves the tag from the nucleotide are also attached to the nanopore. FIGS. 12A-D show schematics of an embodiment of the conjugation strategy wherein an amide bond is formed between Lys and Asp side chains via SpyCatcher - SpyTag.

Tags of the disclosure may be atoms or molecules, or a collection of atoms or molecules. A tag may provide an optical, electrochemical, magnetic, or electrostatic (e.g., inductive, capacitive) signature, which signature may be detected with the aid of a nanopore. A nucleotide can include a tag (or tag species) that is coupled to any location of the nucleotide including, but not limited to a phosphate (e.g., gamma phosphate), sugar or nitrogenous base moiety of the nucleotide. In some cases, tags are detected while tags are associated with a polymerase during the incorporation of nucleotide tags. In some examples,

the tag can move through or in proximity to the nanopore and be detected with the aid of the nanopore.

The present disclosure provides a method for sequencing a nucleic acid. The method comprises incorporating (e.g., polymerizing) tagged nucleotides. In some cases, a tag associated with an individual nucleotide can be detected by a nanopore without being released from the nucleotide upon incorporation.

The present disclosure provides a chip for a nanopore-based nucleic acid sequencing by synthesis that can comprise a plurality of individually addressable nanopores. An individually addressable nanopore of the plurality can contain at least one nanopore formed in a membrane disposed adjacent to an integrated circuit. Each individually addressable nanopore can be capable of detecting a tag associated with an individual nucleotide. The nucleotide can be incorporated (e.g., polymerized) and the tag may not be released from the nucleotide upon incorporation.

The following examples are set forth as being representative of the present disclosure. These examples are not to be construed as limiting the scope of the present disclosure as these and other equivalent embodiments will be apparent in view of the present disclosure, figures and accompanying claims.

EXAMPLE I

Nanopore-based ion-flux modulator for enzymatic information encoding

Components:

Almost every polymerase known to date uses a DNA or RNA template to catalytically incorporate a mononucleotide into the growing primer strand. Terminal deoxynucleotidyl transferase (TdT), however is a unique DNA polymerase, which performs template-

independent DNA synthesis using a single-stranded DNA (ssDNA) as the substrate (Motea, E. A. & Berdis, A. J. Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochim. Biophys. Acta - Proteins Proteomics* 1804, 1151-1166 (2010)). (FIGS. 1A-B). *In vitro* studies showed that TdT can incorporate all four natural nucleotides into a ssDNA (Kato, K. I., Goncalves, J. M., Houts, G. E. & Bollum, F. J. Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase. *J. Biol. Chem.* 242, 2780-2789 (1967)). Subsequent studies also confirmed that TdT requires an initiator chain at least six nucleotides in length (Chang, L. M., Cassani, G. R. & Bollum, F. J. Deoxynucleotide-polymerizing enzymes of calf thymus gland. VII. Replication of homopolymers. *J. Biol. Chem.* 247, 7718-7723 (1972)). As all polymerases, TdT requires a divalent metal ion to catalyze the phosphoryl transfer reaction to incorporate a nucleotide into a growing DNA strand. It has been shown that Mg^{2+} facilitates dGTP/dATP, while Co^{2+} increases dCTP/dTTP catalytic polymerization efficiency (Chang, L. M. & Bollum, F. J. Multiple roles of divalent cation in the terminal deoxynucleotidyltransferase reaction. *J. Biol. Chem.* 265, 17436-17440 (1990)). TdT catalytic activity has been shown to be regulated by protein-protein interactions. For example, TdT interacting factor (TdiFs) binds to the C-terminus of TdT and increases its polymerase activity by 4-fold (Fujita, K. et al. Terminal deoxynucleotidyltransferase forms a ternary complex with a novel chromatin remodeling protein with 82 kDa and core histone. *Genes Cells* 8, 559-571 (2003)). Thus, primer length, divalent metal ion and protein-protein interactions ("optimization or control" parameters) are critical regulating the catalytic activity of TdT in general.

An encoding unit:

According to an exemplary embodiment, an encoding unit is formed by tethering one or a plurality of TdT to a protein nanopore such as α HL or MspA as well as 5' immobilizing single-stranded DNA (ssDNA) around the pore rim for initiating synthesis (FIG. 2A). In one embodiment, the tethering is *via* the SpyCatcher/SpyTag conjugation method. Or alternatively, only the ssDNA is immobilized as described above, while the TdT is in solution for stochastic access to the initiator strand driven by Brownian motion.

According to an exemplary embodiment, two liquid reservoirs (*cis/trans*) that are separated by a lipid bilayer are formed on the opposite sides of the lipid bilayer. In one embodiment, the *cis* side contains non-catalytic buffer (Ca^{2+}), which will prevent covalent bond catalysis, *i.e.*, base addition to the 3' end of the primer sequence. In one embodiment, the *trans* side contains high concentration of catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$), which promotes covalent bond catalysis, *i.e.*, base addition to the 3' end of the primer sequence. In one embodiment, an encoding unit is inserted into the lipid bilayer over an electrode, which can modulate voltage across the membrane in a self-addressable manner.

Encoding algorithm steps:

According to an exemplary embodiment, in step 1, flowing in dATP (or other nucleotide, one at a time) in non-catalytic buffer (Ca^{2+} or other non- $\text{Mg}^{2+}/\text{Co}^{2+}$ allowing binding but not covalent bond catalysis), referred to as step 1, "NUCLEOTIDE FLOW" (FIG. 3A).

In step 2a, turning the voltage gradient on across membrane (ON state), such that a plume of catalytic ions ($\text{Mg}^{++}/\text{Co}^{++}$) is injected to the *cis* side (from the *trans* side). The polarity must be *cis* side negative polarity, *trans* side positive polarity to force the positively charged ions to migrate to the *cis* compartment upon activation, referred to as step 2a, "ENZYME ACTIVATION" (FIG. 3B).

A highly localized, hemispheric ion gradient will be generated at this instance immediately outside of the nanopore (α HL), which may be finely controlled *via* the voltage magnitude and pulse duration. This control is demonstrated theoretically with COMSOL simulations (Anderson, B. N. et al. Probing Solid-State Nanopores with Light for the Detection of Unlabeled Analytes. ACS Nano 8, 11836-1 1845 (20 14)) (FIGS. 4A-C), as well as experimentally using Ca^{2+} sensitive fluorescent activation (Ivankin, A. et al. Label-Free Optical Detection of Biomolecular Translocation through Nanopore Arrays. ACS Nano 8, 10774-1078 1 (20 14)) (FIG. 5). Essentially, it is desirable to generate a locally-saturated reaction volume around the TdT enzyme to promote its catalytic activity, *i.e.*, addition of a single base or a homopolymer run to the ssDNA.

The enzymatic activity of the TdT in the immediate vicinity of each pore is proportional to the catalytic divalent $\text{Mn}^{2+}/\text{Co}^{2+}$ ion flux through the pore and remains constant as long as the steady flow of ions is maintained. The catalytic ion distribution around the nanopore can be modulated/regulated and thus the enzymatic activity is voltage-tunable.

In step 2b, if the voltage gradient is not applied (OFF state), catalytic ion migration will not occur, which will lead to no addition - only transient binding - of a single base to ssDNA, referred to as step 2b, "EZNYME PASSIVATION" (FIG. 3C).

In step 3, flushing the *cis* compartment with fresh non-catalytic buffer to remove any nucleotides and catalytic ion residues, referred to as step 3, "FLUSHING" (FIG. 3D). Note, that the flow rate of this "FLUSHING" step must be optimized such that it does not destroy the lipid bilayer.

Steps 1-3 constitute the gene synthesis cycle for a particular nucleotide. This cycle is repeated with subsequent dTTP, dGTP and dCTP for "NUCLEOTIDE FLOW" iterations

with either steps 2a for TdT activation (base or homopolymer addition) or 2b for TdT passivation (no base addition).

More specifically, since the electrodes are self-addressable, at each "NUCLEOTIDE FLOW" step the ON/OFF state can be controlled, *i.e.*, the particular (single or multiple) base incorporation per encoding cycle. In an exemplary embodiment, DNA can be synthesized in a controlled manner (with either at single-base resolution, if there is a very good pulse control; or with homopolymer-resolution, which can be regarded as a distinct states of {A, C, T, G}).

Encoding schemes:

According to an exemplary embodiment, this scheme can be used to encode information *via* the base case: 12 possible 2-base state transitions (see below) using standard techniques (Goldman, N. et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature 494, 77-80 (2013)), while read out can be implemented by subsequent DNA sequencing (Illumina or nanopore-based) methods. Depending on the alphabet size to be encoded the state-transition map can be extended with multiple-base transitions (for example a 3-base state transition: A- \rightarrow C- \rightarrow T to accommodate more characters. Every hierarchical dimension can be extended with 3 other states, so it has a multiplicity of 3. Specifically, a 2-state transition map has 12 states, a 3-state transition map has $3 * 12 = 36$, while an N-state transition map has $N * 12$ unique states.

01. A^C
02. A^T
03. A^G
04. C^A
05. C^T
06. C^G
07. T^A
08. T^C
09. T^G
10. G^C

11. G^AT
12. G^AA

If single-base synthesis accuracy is achieved *via* the fine voltage-tunable TdT control, this method can be utilized as a high-throughput gene synthesis.

It was shown that template dependent nanopore-coupled polymerase activity may be modulated in the presence of various cations, such as non-catalytic Ca²⁺, which prevents polymerase catalysis, in which case ion current signal corresponding to base-specific binding events can be observed only as in "static capture" (see FIGS. 4A-C & FIGS. 14A-B). With catalytic Mn²⁺, which promotes polymerase catalysis, in which case ion current signal corresponding to template progression can be observed as in "real sequencing" (see FIGS. 5A-D).

It has been demonstrated that each case has the expected enzymatic effects when in saturated concentration regimes (3 mM Ca²⁺, 0.1 mM Mn²⁺).

Using the already established nanopore-array platform, specific nucleotides or homopolymer runs can be encoded into self-priming DNA hairpins in a template-dependent manner. Voltage functions can be established to appropriately control the ion-flux through the nanopore, where optimal voltage magnitude and pulse duration parameters can be established. After establishing these fine-tuning parameters, TdT enzymes instead of $\phi 29$ polymerase can be coupled to the nanopore as shown in FIG. 2. Alternatively, the synthesis efficiency and accuracy can be evaluated using TdT in solution - without tethering it/them to the pore. Enzymatic activity can also be monitored *via* continuous imaging of incorporating fluorescently labeled nucleotides inhibited or promoted by the non-catalytic (Ca²⁺) or catalytic (Mn²⁺/Co²⁺) ion gradient, while running the already established voltage modulation

protocols for activation/deactivation. Processivity of TdT may be boosted with the fusion of single-strand binding proteins (SSB), such as TdiF.

High-throughput scaling:

According to an exemplary embodiment, an electrode area is a circle with diameter, $D_e = 5 \mu\text{m}$, the electrode surface area is, $A_e = 20 \mu\text{m}^2$. In one embodiment, the aHL pore area is assumed such that the pore outer diameter is $D_p = 80 \text{ \AA} = 0.008 \mu\text{m}$, the pore surface area is, $A_p = 5 \times 10^{-5} \mu\text{m}^2$.

In one embodiment, about 400,000 pores can be fitted above an electrode with maximum packing density, but it can be assumed that in reality there will be a 10% insertion rate. This translates to about 40,000 pores per electrode (FIG. 6). Each pore can accommodate seven TdT molecules in theory (FIG. 2B), so each can synthesize a strand with the similar information (A,C,T,G) content based on ion-flux modulation. So, each electrode area can accommodate about 40,000 **encoding units** of the same type (like a local "amplification" or **encoding cluster**).

In one embodiment, for a 128,000 electrode-containing chip, it may synthesize 128,000 different species of DNA (= gene or a string of information with a 40,000-fold amplification - can help with minimizing error).

EXAMPLE II

Information encoding with a DNA polymerase-coupled nanopore

DNA polymerases (e.g. phi-29 DNAP) preferentially incorporate dCTP when copying a template containing inosine bases. In the absence of dCTP, the polymerase will incorporate dATP, dGTP, or dTTP (in order of decreasing preference). This example describes an information encoding system comprising a DNA polymerase (e.g. phi 29 DNAP)-coupled

nanopore (alpha hemolysin) and a circular ssDNA template strand containing some sequence of cytosine, adenine, thymine, and inosine bases (template-CATI) (FIG. 9). The polymerase and ssDNA template is tethered on the *cis* side of the nanopore. Also present in the *cis* solution are dATP, dGTP, and dTTP (in addition to other cofactors and salts required for polymerase activity). In the *trans* solution, only dCTP is present (in addition to the other cofactors and salts). The information encoding method described herein employs standard DNA sequencing conditions, the co-factors and salts are well-established in the art and are known to a skilled in the art. When a voltage is applied across the nanopore (*cis* side positive), the negatively charged dCTP molecules are electrophoretically forced through the pore and into the proximity of the polymerase-template complex. If the polymerase is actively copying the template-CATI strand to produce the complement strand during this time, cytosine is preferentially incorporated into the nascent complement strand during synthesis at sites that complement the inosine of the template strand. However, when the polarity of the voltage is reversed (i.e. *cis* side negative), dCTP no longer flows through the pore, and the polymerase is no longer able to incorporate cytosine into the complement strand. That is, dATP, dGTP, or dTTP must be incorporated instead. Thus, this system "records" the voltage polarity applied across the nanopore over time by encoding this information into a complement strand of DNA during polymerization of the template strand. Information is stored in the complement strand in the form of regions of bases either containing or lacking cytosine bases at the sites which complement the inosine bases of the template strand, and this translates to the "1" or "0" state of a single bit of information (FIG. 10).

EXAMPLE III

Single Molecule Attachment Strategy for High-Throughput Nanopore Measurement

This example describes a novel design and experimental characterization of an attachment strategy to precisely couple a ϕ 29 polymerase molecule (or potentially any other enzyme) to a aHL nanopore (or potentially any other type of membrane channel, such as MspA, ϕ 29 connector, etc. - see FIG. 11) for single molecule (SM) measurement (FIGS. 12A-D). In this work, a specific application of this strategy - namely DNA sequencing - was demonstrated by observing the SM incorporation events associated with the catalytic step of the polymerase during complementary (FIGS. 13A-D and FIGS. 14A-B) and successive (FIGS. 15A-B) tagged nucleotide incorporation into a template DNA. The method of isolating a 1:6 aHL nanopore, followed by addition of a single covalently bound polymerase could be extended to other methods of single molecule detection via a nanopore. Single molecule enzyme activity, or protein-protein interactions could be observed by coupling the desired molecular event to the alteration of current through the pore. This technology could serve as the basis for the design of a host of new, high-throughput molecular sensors.

One attachment pair or moiety for the attachment of the polymerase to the pore is SpyTag/SpyCatcher. The method is described in US Patent Application No. 13/578,070, published as US20 130053544. The method is not specific to nanopore/polymerase/sequencing. Another general strategy of nucleic acid sequencing is described in US Patent Application No. 14/073,445, published as US20 140 1346 16.

In this example, the SpyCatcher/SpyTag (Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M. Peptide Tag Forming a Rapid Covalent Bond to a Protein, through Engineering a Bacterial Adhesin. Proc. Natl. Acad. Sci. 2012, 109, E690-E697) protein conjugation system was adopted to couple a single polymerase with one aHL heptamer. Previous work had demonstrated that it was possible to purify a 1:6 heptameric pore, where one subunit contains a C-terminal 6x-histidine tag (6x-His-tag) (SEQ

ID NO: 1) and the other six contain neutral Strep-tags (Davis, R.; Chen, R.; Bibillo, A.; Korenblum, D.; Dorwart, M. Nucleic Acid Sequencing Using Tags. US20 140 1346 16, 2014). An aHL pore coupled to a single polymerase molecule was made from three proteins: aHL with a C-terminal Strep-tag, aHL with a C-terminal SpyTag peptide followed by a 6x-His-tag (SEQ ID NO: 1), and ϕ 29 with a C-terminal SpyCatcher (FIG. 12A). The whole porin-polymerase conjugate was assembled stepwise, by first forming and purifying the 1:6 (SpyTag:unmodified) aHL pore, followed by addition of (ϕ 29-SpyCatcher (FIG. 12B). Amino acid linker lengths between aHL-SpyTag and (ϕ 29-SpyCatcher were chosen based on assembling the structures of these proteins (Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. Structure of Staphylococcal Alpha-Hemolysin, a Heptameric Transmembrane Pore. Science 1996, 274, 1859-1866; Berman, A. J.; Kamtekar, S.; Goodman, J. L.; Lazaro, J. M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T. a. Structures of phi29 DNA Polymerase Complexed with Substrate: The Mechanism of Translocation in B-Family Polymerases. EMBO J. 2007, 26, 3494-3505) into a porin-polymerase conjugate followed by macromolecular modeling of the linkers using Rosetta (Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; et al. ROSETTA3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. Methods Enzymol. 2011, 487, 545-574.) FIG. 12C. Protein sequences are detailed in FIG. 16.

Experimental Methods

Protein expression and purification

The ϕ 29 DNA polymerase- SpyCatcher construct with an N-terminal Strep-tag was expressed in BL21 DE3 Star cells by growing them in Magic Media (Invitrogen, Grand Island, NY) at 37°C until OD -0.6, followed by overnight growth at 25°C. Cells were

resuspended and lysed by sonication in Polymerase Buffer (PolBuff): 50 mM Tris pH 7.5, 150 mM NaCl, 0.1 mM EDTA, 0.05% (v/v) Tween-20, 5 mM 2-mercaptoethanol. Benzolase nuclease was added post cell lysis to remove excess bound DNA. The protein was purified using Streptactin columns per manufacturer's instructions (IBA, Goettingen, Germany). Purified protein was eluted with PolBuff with added desthiobiotin. Both aHL-Strep-tag and aHL-SpyTag-6x-His ("6x-His" disclosed as SEQ ID NO: 1) were expressed in BL21 DE3 Star pLys-S cells grown in Magic Media for 8 hours at 37°C. Each was lysed by sonication in 50 mM Tris pH 8.0, 200 mM NaCl. Strep-tagged aHL was purified on Streptactin columns, and eluted in the same buffer with desthiobiotin. His-tagged aHL was purified with a cobalt column and eluted with 300 mM imidazole.

1:6 Porin assembly formation and isolation

To form a 1:6 SpyTag:unmodified aHL pore, purified aHL proteins were mixed in a ratio of 1:9 SpyTag construct:unmodified. The lipid 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) was added to a final concentration of 5 mg/mL, followed by incubation at 40°C for 30 mins. Lipid vesicles were subsequently popped by adding n-octyl- β -D-glucoside (*fiOG*) to 5% (v/v). Fully formed oligomers were separated from vesicles and monomers by size exclusion chromatography (SEC) in 20 mM HEPES pH 7.5, 75 mM KCl and 30 mM β OG. Oligomeric protein obtained from the SEC was then run on a MonoS column in 20 mM MES pH 5.0, 0.1% Tween-20 and eluted with a linear gradient to 2.0 M NaCl. The desired 1:6 assembly eluted after the 0:7 porin since the 1:6 assembly contains a 6x-His-tag (SEQ ID NO: 1). The 1:6 composition was confirmed by adding SpyCatcher protein and observing a size shift of the conjugate on an SDS polyacrylamide gel indicative of only one SpyCatcher molecule per assembled pore.

Polymerase and template attachment

Purified $\phi 29$ and the desired template were bound to the pore by incubating two molar equivalents of polymerase and four equivalents of DNA template per 1:6 pore overnight at 4°C. The full tertiary complex was isolated by SEC in 20 mM HEPES pH 7.5, 150 mM KCl, 0.01% Tween-20, 5mM TCEP. Isolated fractions were characterized by SDS PAGE to confirm the presence of $\phi 29$ and α HL conjugate. Formed complexes were tested for polymerase function by rolling circle amplification.

Lipid bilayer formation

Synthetic lipid 1,2-di-O-phytanyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids, Alabaster, AL) was diluted in tridecane (Sigma-Aldrich, St. Louis, MO) to a final concentration of 15 mg/mL. A single lipid bilayer was formed on a silanized CMOS chip surface containing an array of 264 Ag/AgCl electrodes. The automated lipid spreading protocol utilized an iterative buffer and air bubble flow to mechanically thin the membrane. During this step, current was applied across the lipid bilayer in order to detect its capacitance, which directly correlates to structural integrity of the membrane. An empirically determined capacitance threshold value of 5 fF/ μm^2 was used to classify the properly formed, single lipid bilayer to conclude the thinning protocols.

Pore insertion

The automated pore insertion method consisted of two voltage protocols: (1) initially constant DC voltage was applied at 160 mV for 1 min, immediately followed by (2) a linearly increasing voltage ramp from 50 mV to 600 mV with a 1 mV/s incremental step. The smoothly increasing voltage gradient amplified the electrical driving force guiding the nanopores into the lipid bilayer. If a cell became active, *i.e.*, started to record current in the pA regime, we considered this event a pore insertion, due to the measured increase in

conductance across the bilayer. Immediately after this event, this cell was turned off to prevent additional pore insertions. In this way, the probability of multiple pore insertions above the same electrode array element was minimized.

Nanopore experiments

All ternary complex capture (TCC) experiments were performed in a buffer containing 300 mM NaCl, 3 mM CaCh - providing the non-catalytic divalent cations to probe nucleotide binding/unbinding events - and 20 mM HEPES pH 7.5. For sequencing experiments, this buffer was modified by replacing CaCh with 0.1 mM MnCh as a catalytic cation source during the polymerase extension reaction to initiate and sustain sequential nucleotide additions along the template DNA. Purified porin-polymerase-template conjugates were diluted in buffer to a final concentration of 2 nM. After pumping a 5 μ L aliquot to the *cis* compartment, single pores were embedded in the planar lipid bilayer that separates two compartments (denoted *cis* and *trans*) each containing ~3 μ L of buffer solution. Experiments were conducted at 27°C with 5 μ M tagged nucleotides added to the *cis* well.

Data acquisition

The ionic current through the nanopore was measured between individually addressable Ag/AgCl electrodes coupled to a silicon substrate integrated electrical circuit. This consisted of an integrating patch clamp amplifier (Genia Technologies, Mountain View CA), which provided a constant 100 mV potential across the lipid bilayer in voltage clamp mode. Data were recorded at a 1 kHz bandwidth in an asynchronous configuration at each cell using circuit-based analog-to-digital conversion and noise filtering (Genia Technologies, Mountain View CA), which allows independent sequence reads at each pore complex. During the various experimental steps, a precision syringe pump (Tecan, Mannedorf, Switzerland)

was utilized in an automated fashion to deliver reagents into the microfluidic chamber of the CMOS chip at a flow rate of 1 μ L/s. Software control was implemented in Python, which interfaced with the pump *via* an RS 232 communication protocol.

Event detection and data analysis

Ionic current blockade events were identified using a custom event detection algorithm implemented in MATLAB (2014b, MathWorks, Natick, MA). Briefly, an event was identified by selecting segments that deflected from open channel current ($I_o = -30$ pA at 100 mV in 300 mM NaCl, 3 mM CaCh and 20 mM HEPES pH 7.5) below a cutoff value of 70% of I_o (21 pA) to a stable current level (I_B) with a minimum dwell time of >10 ms. For each nanopore experiment, event searches were performed to obtain the average residual current level (with respect to open channel) for each capture event (I_{RES}). Statistical analysis was performed to determine the mean, median and standard deviation of each capture event by fitting a Gaussian to a histogram of I_B values. The residual current blockade was defined as: $I_{RES}\% = I_{RES}/I_o$, while the duration of the event in the deflected segment corresponds to the dwell time. Mean dwell time and residual current of each event in an experimental set was accumulatively quantified using scatter plots and box-and-whisker plots. On each box plot, the central red mark represents the median, while the bottom and top blue edges of the box are the 1st and 3rd quartile median values respectively. The whiskers extend to the lowest and highest values within 1.5 IQR of the 1st and 3rd quartile medians. Alternatively, average dwell time/residual current probability histograms were generated by plotting each bin normalized by the total number of observed events.

Classification of capture events

As a conservative classification method, we have identified the ternary complex capture (TCC) events as all events clustered inside the tag capture box defined by a mean dwell time interval of 10^{-2} to 10^{+1} seconds and a normalized current blockade (or residual current) region bounded by the 1st and 3rd quartile median values (lower/upper bounds) of the normalized current blockage boxplots - for a particular tagged nucleotide - respectively (FIGS. 14A-B).

What is claimed is:

1. A method of DNA synthesis comprising

providing an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA)
and a nanopore,

providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each
having a different buffer composition, wherein the nanopore is within the lipid bilayer and
the enzyme and the ssDNA are in the *cis* reservoir, and

applying a voltage across the lipid bilayer via an electrode wherein the electrode can
modulate the voltage across the lipid bilayer, and wherein the enzyme is activated and
catalyzes DNA synthesis in response to the voltage.
2. The method of claim 1, wherein the enzyme is conjugated to the nanopore or free in
the *cis* reservoir.
3. The method of claim 1, wherein the enzyme is a template dependent or independent
DNA polymerase.
4. The method of claim 3, wherein the DNA polymerase is a ϕ 29 DNA polymerase or a
terminal deoxynucleotidyl transferase.
5. The method of claim 1, wherein the ssDNA is immobilized to the nanopore.
6. The method of claim 1, wherein the ssDNA is immobilized around rim of the
nanopore at its 5' end.
7. The method of claim 1, wherein the ssDNA is a template or an initiator for DNA
synthesis.

8. The method of claim 1, wherein the nanopore is a protein membrane channel.
9. The method of claim 8, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.
10. The method of claim 1, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.
11. The method of claim 10, further comprising providing nucleotide bases to the *cis* reservoir and turning on the voltage such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes nucleotide base addition during DNA synthesis.
12. The method of claim 11, wherein the nucleotide bases can be added as a single base or a homopolymer run during DNA synthesis.
13. The method of claim 11, wherein the nucleotide bases comprise dATP, dTTP, dCTP, or dGTP.
14. The method of claim 11, further comprising turning off the voltage and flushing the *cis* reservoir with fresh non-catalytic buffer and removing any nucleotide bases and catalytic ions.
15. The method of claim 1, wherein the method is capable of single molecule measurement.
16. A system of DNA synthesis comprising

an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore,

a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir, and

an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.

17. The system of claim 16, wherein the enzyme is conjugated to the nanopore or free in the *cis* reservoir.
18. The system of claim 16, wherein the enzyme is a template dependent or independent DNA polymerase.
19. The system of claim 16, wherein the DNA polymerase is a ϕ 29 DNA polymerase or a terminal deoxynucleotidyl transferase.
20. The system of claim 16, wherein the ssDNA is immobilized to the nanopore.
21. The system of claim 16, wherein the ssDNA is immobilized around rim of the nanopore at its 5' end.
22. The system of claim 16, wherein the ssDNA is a template or an initiator for DNA synthesis.
23. The system of claim 16, wherein the nanopore is a protein membrane channel.

24. The system of claim 23, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.
25. The system of claim 16, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.
26. The system of claim 16, wherein nucleotide bases are added to the *cis* reservoir, which are incorporated into product DNA when the voltage is turned on such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes the incorporation of the nucleotide bases into the product DNA.
27. The system of claim 16, wherein the nucleotide bases can be added as a single base or a homopolymer run during DNA synthesis.
28. The system of claim 16, wherein the system is capable of single molecule measurement.
29. A method of encoding information comprising
- providing an encoding unit comprising a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore,
- providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase and the ssDNA are in the *cis* reservoir,
- providing nucleotide bases to the *cis* reservoir, and

applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux.

30. The method of claim 29, wherein the terminal deoxynucleotidyl transferase is conjugated to the protein nanopore or free in the *cis* reservoir.
31. The method of claim 29, wherein the ssDNA is immobilized to the protein nanopore.
32. The method of claim 29, wherein the ssDNA is immobilized around rim of the protein nanopore at its 5' end.
33. The method of claim 29, wherein the ssDNA is an initiator for nucleotide base addition.
34. The method of claim 29, wherein the protein nanopore is a protein membrane channel.
35. The method of claim 34, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.
36. The method of claim 29, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition.
37. The method of claim 29, wherein the terminal deoxynucleotidyl transferase is activated by catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) that are injected to the *cis* side from the *trans* side when the voltage is turned on and catalyzes the addition of the nucleotide base to the ssDNA.

38. The method of claim 29, wherein the nucleotide bases can be added as a single base or a homopolymer run during DNA synthesis.
39. The method of claim 29, wherein the nucleotide bases comprise dATP, dTTP, dCTP or dGTP.
40. The method of claim 29, further comprising turning off the voltage and flushing the *cis* reservoir with fresh non-catalytic buffer and removing any nucleotide bases and catalytic ions.
41. The method of claim 29, wherein the method is capable of single molecule measurement.
42. The method of claim 29, wherein the information is encoded at single-base resolution or homopolymer resolution.
43. The method of claim 29, wherein the information encoding accommodates 12 possible 2-base state transitions.
44. The method of claim 29, wherein the information encoding accommodates N^{*12} unique N-base state transitions.
45. The method of claim 29, wherein the information is encoded and stored in the product DNA.
46. The method of claim 29, wherein the ssDNA is linear.
47. An information encoding system comprising
an encoding unit comprising a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore,

a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase, the ssDNA and nucleotide bases are in the *cis* reservoir, and

an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux.

48. The system of claim 47, wherein the terminal deoxynucleotidyl transferase is conjugated to the protein nanopore or free in the *cis* reservoir.
49. The system of claim 47, wherein the ssDNA is immobilized to the protein nanopore.
50. The system of claim 47, wherein the ssDNA is immobilized around rim of the protein nanopore at its 5' end.
51. The system of claim 47, wherein the ssDNA is as an initiator for DNA synthesis.
52. The system of claim 47, wherein the protein nanopore is a protein membrane channel.
53. The system of claim 52, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.
54. The system of claim 47, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.

55. The system of claim 47, wherein nucleotide bases are added to the *cis* reservoir, which are incorporated into product DNA when the voltage is turned on such that catalytic ions (Mg^{2+}/Co^{2+}) are injected to the *cis* side from the *trans* side, thereby activating the terminal deoxynucleotidyl transferase that catalyzes the incorporation of the nucleotide bases into the product DNA.

56. The system of claim 47, wherein the nucleotide bases can be added as a single base or a homopolymer run during DNA synthesis.

57. The system of claim 47, wherein the nucleotide bases comprise dATP, dTTP, dCTP or dGTP.

58. The system of claim 47, wherein the system is capable of single molecule measurement.

59. The system of claim 47, wherein the information is encoded at single-base resolution or homopolymer resolution.

60. The system of claim 47, wherein the information encoding accommodates 12 possible 2-base state transitions.

61. The system of claim 47, wherein the information encoding accommodates $N * 12$ unique N-base state transitions.

62. The system of claim 47, wherein the information is encoded and stored in the product DNA.

63. The system of claim 47, wherein the ssDNA is linear.

64. A method of encoding information comprising
- providing an encoding unit comprising a DNA polymerase, a circular single-stranded DNA (ssDNA) template and a protein nanopore,
- providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the DNA polymerase and the circular ssDNA template are in the *cis* reservoir,
- and
- applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the information is encoded via DNA synthesis in response to the voltage.
65. The method of claim 64, wherein the DNA polymerase is conjugated to the protein nanopore or free in the *cis* reservoir.
66. The method of claim 64, wherein the circular ssDNA template is immobilized to the protein nanopore.
67. The method of claim 64, wherein the circular ssDNA template includes cytosine, adenine, thymine, and inosine but not guanine.
68. The method of claim 64, wherein the DNA polymerase is a ϕ 29 DNA polymerase.
69. The method of claim 64, wherein the protein nanopore is a protein membrane channel.
70. The method of claim 69, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.

71. The method of claim 64, wherein the *cis* reservoir contains dATP, dGTP, or dTTP, and wherein the *trans* reservoir contains dCTP.

72. The method of claim 71, wherein the dCTP flows to the *cis* reservoir and is incorporated into a complement strand of DNA when the voltage is positive.

73. The method of claim 71, wherein the dATP, dGTP or dTTP is incorporated into a complement strand of DNA when the voltage is negative.

74. The method of claim 64, wherein the method records voltage polarity applied across the lipid bilayer over time by encoding this information into a complement strand of DNA during DNA synthesis.

75. The method of claim 74, wherein the information is stored in the complement strand in the form of regions of bases either containing or lacking cytosine bases at the sites which complement the inosine bases of the template strand, which translates to the "1" or "0" state of a single bit of information.

76. An information encoding system comprising

an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore,

a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir, and

an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.

77. The system of claim 76, wherein the enzyme is conjugated to the nanopore or free in the *cis* reservoir.
78. The system of claim 76, wherein the enzyme is a template dependent or independent DNA polymerase.
79. The system of claim 76, wherein the DNA polymerase is a ϕ 29 DNA polymerase or a terminal deoxynucleotidyl transferase.
80. The system of claim 76, wherein the ssDNA is immobilized to the nanopore.
81. The system of claim 76, wherein the ssDNA is a template or an initiator for DNA synthesis.
82. The system of claim 76, wherein the nanopore is a protein membrane channel.
83. The system of claim 82, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.
84. The system of claim 76, wherein the *cis* reservoir contains dATP, dGTP, or dTTP, and wherein the *trans* reservoir contains dCTP.
85. The system of claim 84, wherein the dCTP flows to the *cis* reservoir and is incorporated into a complement strand of DNA when the voltage is positive.
86. The system of claim 76, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.
87. The system of claim 76, wherein nucleotide bases are added to the *cis* reservoir, which are incorporated into product DNA when the voltage is turned on such that catalytic

ions (Mg^{2+}/Co^{2+}) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes the incorporation of the nucleotide bases into the product DNA.

88. The system of claim 76, wherein the nucleotide base can be added as a single base or a homopolymer run during DNA synthesis.

89. The system of claim 76, wherein the system is capable of single molecule measurement.

90. The system of claim 76, wherein the information is encoded and stored in the product DNA.

91. The system of claim 76, wherein the ssDNA is circular.

92. A method of single molecule attachment for high throughput nanopore measurement comprising

providing an encoding unit comprising a single molecule, a single-stranded DNA (ssDNA) and a nanopore,

providing a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the single molecule and the ssDNA are in the *cis* reservoir, and

applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the single molecule catalyzes DNA synthesis in response to the voltage induced ion-flux.

93. The method of claim 92, wherein the single molecule is a template dependent or independent DNA polymerase.

94. The method of claim 92, wherein the DNA polymerase is a ϕ 29 DNA polymerase or a terminal deoxynucleotidyl transferase.
95. The method of claim 92, wherein the ssDNA is immobilized to the nanopore.
96. The method of claim 92, wherein the ssDNA is as a template or an initiator for DNA synthesis.
97. The method of claim 92, wherein the single molecule is conjugated to the nanopore by SpyCatcher/SpyTag protein conjugation.
98. The method of claim 92, wherein the nanopore is a protein membrane channel.
99. The method of claim 98, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.
100. The method of claim 92, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.
101. The method of claim 92, further comprising providing nucleotide bases to the *cis* reservoir and turning on the voltage such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes the addition of the nucleotide bases during DNA synthesis.
102. The method of claim 92, wherein the nucleotide base can be added as a single base or a homopolymer run during DNA synthesis.

103. The method of claim 92, further comprising turning off the voltage, flushing the *cis* reservoir with fresh non-catalytic buffer and removing any nucleotide bases and catalytic ions.

104. The method of claim 92, wherein the method is capable of single molecule measurement.

105. The method of claim 92, wherein the information is encoded and stored in the product DNA.

106. A system of single molecule attachment for high throughput nanopore measurement comprising

an encoding unit comprising a single molecule, a single-stranded DNA (ssDNA) and a nanopore,

a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the single molecule and the ssDNA are in the *cis* reservoir, and

an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the single molecule catalyzes DNA synthesis in response to the voltage.

107. The system of claim 106, wherein the single molecule is a template dependent or independent DNA polymerase.

108. The system of claim 106, wherein the DNA polymerase is a ϕ 29 DNA polymerase or a terminal deoxynucleotidyl transferase.

109. The system of claim 106, wherein the ssDNA is immobilized to the nanopore.
110. The system of claim 106, wherein the ssDNA is a template or an initiator for DNA synthesis.
111. The system of claim 106, wherein the single molecule is conjugated to the nanopore by SpyCatcher/SpyTag protein conjugation.
112. The system of claim 106, wherein the nanopore is a protein membrane channel.
113. The system of claim 106, wherein the protein membrane channel comprises aHL, MspA, or $\phi 29$ connector.
114. The system of claim 106, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.
115. The system of claim 106, wherein nucleotide bases are added to the *cis* reservoir, which are incorporated into product DNA when the voltage is turned on such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes the incorporation of the nucleotide bases into the product DNA.
116. The system of claim 106, wherein the nucleotide base can be added as a single base or a homopolymer run during DNA synthesis.
117. The system of claim 106, wherein the system is capable of single molecule measurement.
118. The system of claim 107, wherein the DNA polymerase is a $\phi 29$ DNA polymerase.
119. The system of claim 106, wherein the ssDNA is immobilized to the nanopore.

120. The system of claim 106, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.
121. The system of claim 120, further comprising turning on the voltage such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating the DNA polymerase that catalyzes nucleotide addition during DNA synthesis.
122. The system of claim 107, wherein the DNA polymerase is covalently coupled to the α -hemolysin heptamer using a SpyCatcher/SpyTag conjugation approach.
123. The system of claim 106, wherein the electrode is a complementary metal-oxide-semiconductor (CMOS)-based electrode array for high-throughput electrical recording.
124. The system of claim 106, wherein information is encoded and stored in the product DNA.
125. A composition comprising
- an encoding unit comprising a terminal deoxynucleotidyl transferase attached to a protein nanopore, and
 - a lipid bilayer wherein the protein nanopore is within the lipid bilayer.
126. The composition of claim 125, further comprising a single stranded DNA (ssDNA).
127. The composition of claim 125, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, and wherein the terminal deoxynucleotidyl transferase, the ssDNA and nucleotide bases are in the *cis* reservoir.

128. The composition of claim 125, wherein an electrode can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the terminal deoxynucleotidyl transferase catalyzes nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux.

129. The composition of claim 127, wherein the terminal deoxynucleotidyl transferase is conjugated to the protein nanopore or free in the *cis* reservoir.

130. The composition of claim 125, wherein the ssDNA is immobilized to the protein nanopore.

131. The composition of claim 125, wherein the ssDNA is immobilized around rim of the protein nanopore at its 5' end.

132. The composition of claim 125, wherein the ssDNA is an initiator for DNA synthesis.

133. The composition of claim 125, wherein the protein nanopore is a protein membrane channel.

134. The composition of claim 133, wherein the protein membrane channel comprises (xHL, MspA, or ϕ 29 connector).

135. The composition of claim 127, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.

136. The composition of claim 127, wherein nucleotide bases are added to the *cis* reservoir, which are incorporated into product DNA when the voltage is turned on such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating

the terminal deoxynucleotidyl transferase that catalyzes the incorporation of the nucleotide bases into the product DNA.

137. The composition of claim 136, wherein the nucleotide bases can be added as a single base or a homopolymer run during DNA synthesis.

138. The composition of claim 136, wherein the nucleotide bases comprise dATP, dTTP, dCTP or dGTP.

139. The composition of claim 128, wherein the composition is capable of single molecule measurement.

140. The composition of claim 125, wherein the ssDNA is linear.

141. A system comprising

an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore, and

a lipid bilayer having on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the *cis* reservoir.

142. The system of claim 141, further comprising an electrode that can apply a voltage across the lipid bilayer, wherein the electrode can modulate the voltage across the lipid bilayer, wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.

143. The system of claim 141, wherein the enzyme is conjugated to the nanopore or free in the *cis* reservoir.

144. The system of claim 141, wherein the enzyme is a template dependent or independent DNA polymerase.

145. The system of claim 144, wherein the DNA polymerase is a ϕ 29 DNA polymerase or a terminal deoxynucleotidyl transferase.

146. The system of claim 141, wherein the ssDNA is immobilized to the nanopore.

147. The system of claim 141, wherein the ssDNA is a template or an initiator for DNA synthesis.

148. The system of claim 141, wherein the nanopore is a protein membrane channel.

149. The system of claim 148, wherein the protein membrane channel comprises aHL, MspA, or ϕ 29 connector.

150. The system of claim 141, wherein the *cis* reservoir contains dATP, dGTP, or dTTP, and wherein the *trans* reservoir contains dCTP.

151. The system of claim 150, wherein the dCTP flows to the *cis* reservoir and is incorporated into a complement strand of DNA when the voltage is positive.

152. The system of claim 141, wherein the *cis* reservoir contains non-catalytic buffer (Ca^{2+}) that prevents base addition during DNA synthesis, and wherein the *trans* reservoir contains catalytic buffer ($\text{Mg}^{2+}/\text{Co}^{2+}$) that promotes base addition during DNA synthesis.

153. The system of claim 141, wherein nucleotide bases are added to the *cis* reservoir, which are incorporated into product DNA when the voltage is turned on such that catalytic ions ($\text{Mg}^{2+}/\text{Co}^{2+}$) are injected to the *cis* side from the *trans* side, thereby activating the enzyme that catalyzes the incorporation of the nucleotide bases into the product DNA.

154. The system of claim 141, wherein the nucleotide base can be added as a single base or a homopolymer run during DNA synthesis.
155. The system of claim 141, wherein the system is capable of single molecule measurement.
156. The system of claim 141, wherein the ssDNA is circular or linear.
157. A kit comprising the composition of claim 125, and
a reagent to reconstitute the terminal deoxynucleotidyl transferase, the nucleotide bases and the ssDNA.
158. A method of nucleic acid synthesis comprising
applying a voltage to cause migration of activating ions from a first location to a reaction zone including a template independent polymerase, nucleotides and an initiator strand including a terminal nucleotide, wherein the activating ions activate the template independent polymerase to add one or more nucleotides to the terminal nucleotides of the initiator strand.
159. The method of claim 158 wherein the template independent polymerase and the initiator are co-localized on a substrate.
160. The method of claim 158 wherein the reaction zone is separated from the first location by a nanopore or nanochannel under influence of an electric field.
161. The method of claim 158 wherein the reaction zone is separated from the first location by a nanopore or nanochannel having a portion of the nanopore or nanochannel within a lipid bilayer under influence of an electric field.

162. The method of claim 158 wherein the reaction zone is separated from the first location by a nanopore or nanochannel having a portion of the nanopore or nanochannel within a lipid bilayer under influence of an electric field, and wherein application of the electric field causes migration of the activating ions from the first location through the nanopore or nanochannel and into the reaction zone wherein the template independent polymerase is activated to add one or more nucleotides to the terminal nucleotide of the initiator strand.

163. A method of DNA synthesis comprising

applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises an enzyme, a single-stranded DNA (ssDNA) and a nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are conjugated to the nanopore and are in the *cis* reservoir, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.

164. The method of claim 163, wherein the *cis* reservoir further comprises nucleotide bases.

165. The method of claim 163, wherein the voltage is applied via an electrode that modulates the voltage across the lipid bilayer.

166. A method of DNA synthesis comprising

applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises an enzyme, a single-stranded DNA (ssDNA) and a nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the nanopore is within the lipid

bilayer and the enzyme and the ssDNA are conjugated to the nanopore and are in the *cis* reservoir, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.

167. The method of claim 166, wherein the *cis* reservoir further comprises nucleotide bases.

168. The method of claim 166, wherein the voltage is applied via an electrode that modulates the voltage across the lipid bilayer.

169. A method of encoding information comprising

applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises a terminal deoxynucleotidyl transferase, a single-stranded DNA (ssDNA) and a protein nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the terminal deoxynucleotidyl transferase and the ssDNA are conjugated to the nanopore and are in the *cis* reservoir, and wherein the information is encoded via terminal deoxynucleotidyl transferase catalyzed nucleotide base addition to the 3' end of the ssDNA in response to the voltage induced ion-flux.

170. The method of claim 169, wherein the *cis* reservoir further comprises nucleotide bases.

171. The method of claim 169, wherein the voltage is applied via an electrode that modulates the voltage across the lipid bilayer.

172. The method of claim 169, wherein the ssDNA is linear.

173. A method of encoding information comprising

applying a voltage across a lipid bilayer having one or more encoding units associated therewith, wherein each encoding unit comprises a DNA polymerase, a circular single-stranded DNA (ssDNA) template and a protein nanopore, wherein the lipid bilayer has on opposite sides a *cis* and a *trans* reservoir each having a different buffer composition, wherein the protein nanopore is within the lipid bilayer and wherein the DNA polymerase and the circular ssDNA template are conjugated to the nanopore and are in the *cis* reservoir, and wherein the information is encoded via DNA synthesis in response to the voltage.

174. The method of claim 173, wherein the voltage is applied via an electrode that modulates the voltage across the lipid bilayer.

175. The method of claim 173, wherein the *cis* reservoir contains dATP, dGTP, or dTTP, and wherein the *trans* reservoir contains dCTP.

FIG. 1A

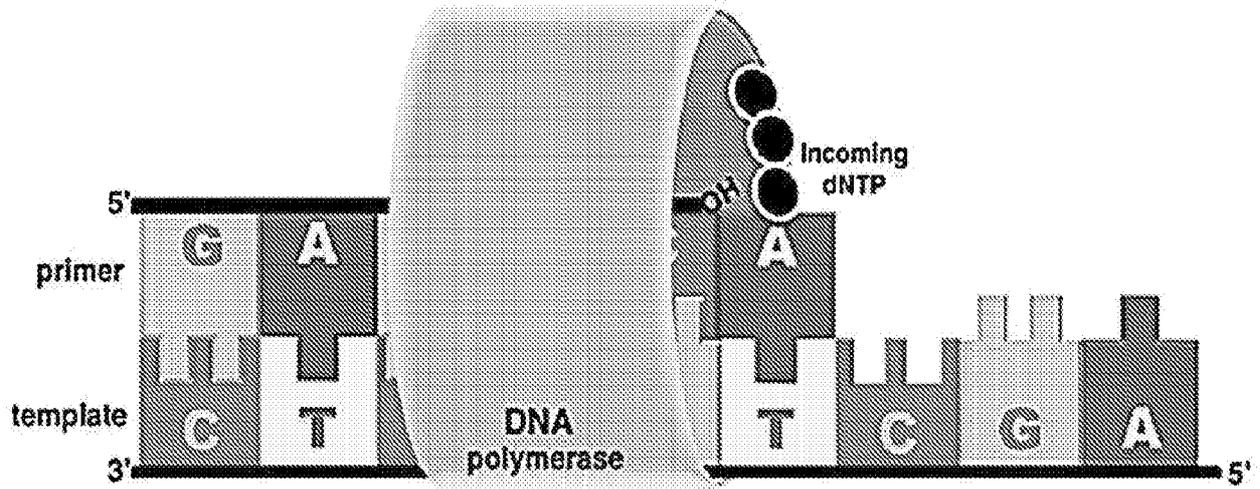


FIG. 1B

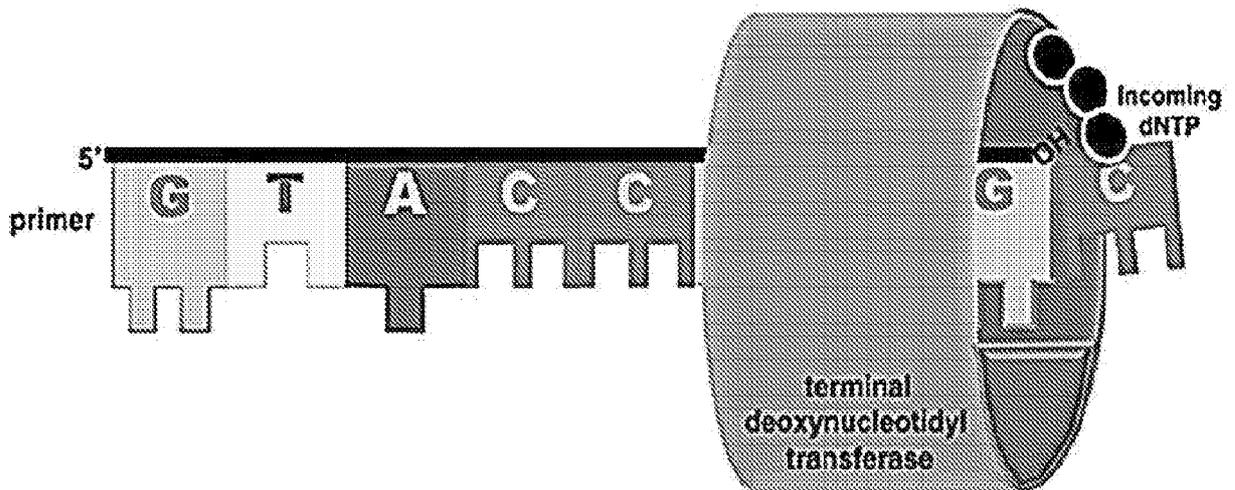


FIG. 2A

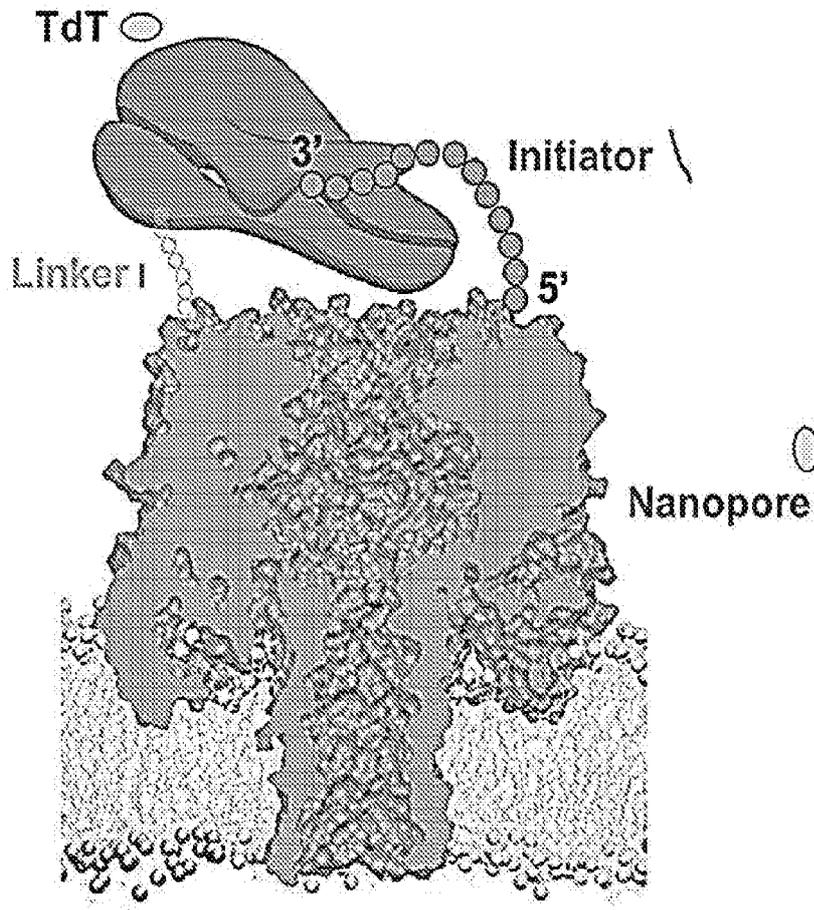


FIG. 2B

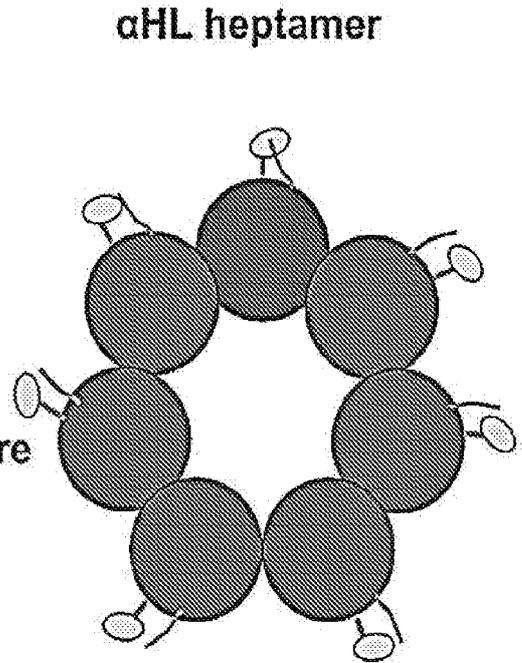


FIG. 3A

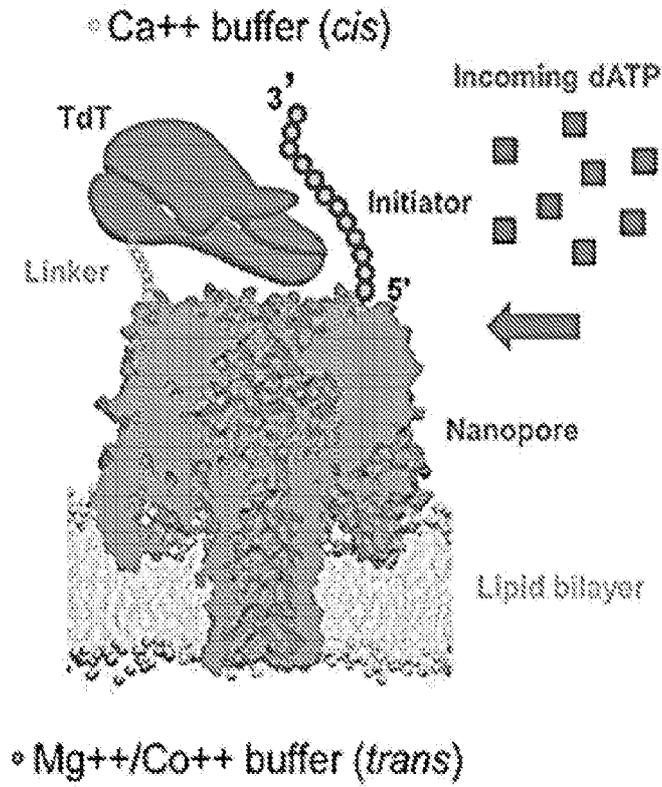


FIG. 3B

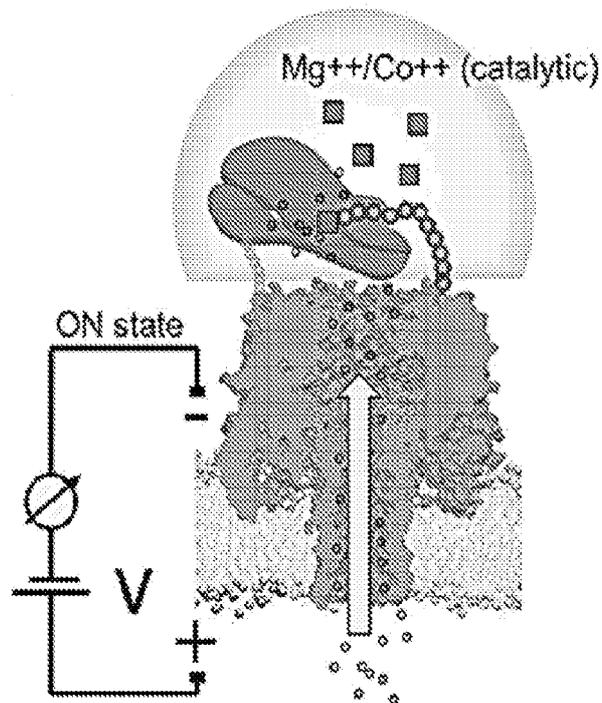


FIG. 3C

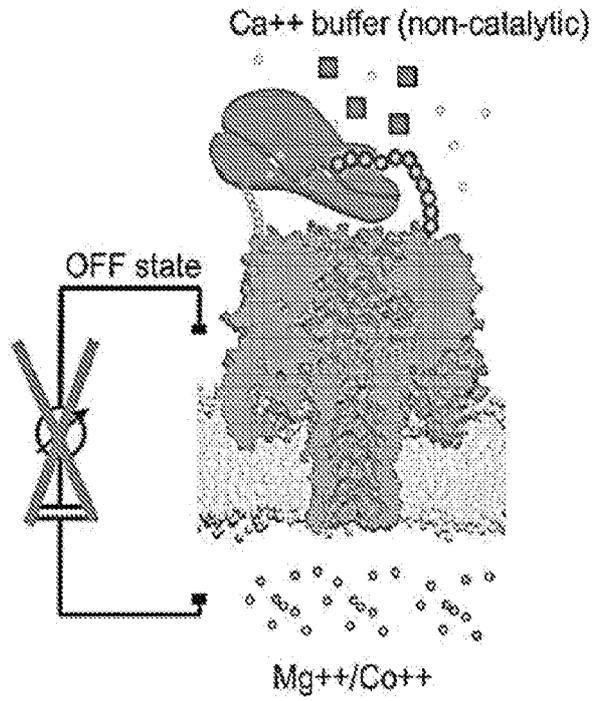


FIG. 3D

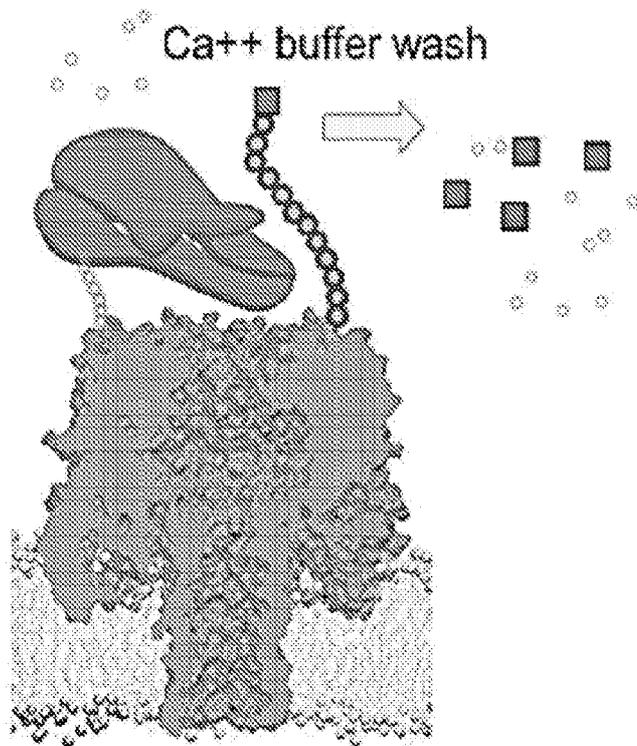


FIG. 4A

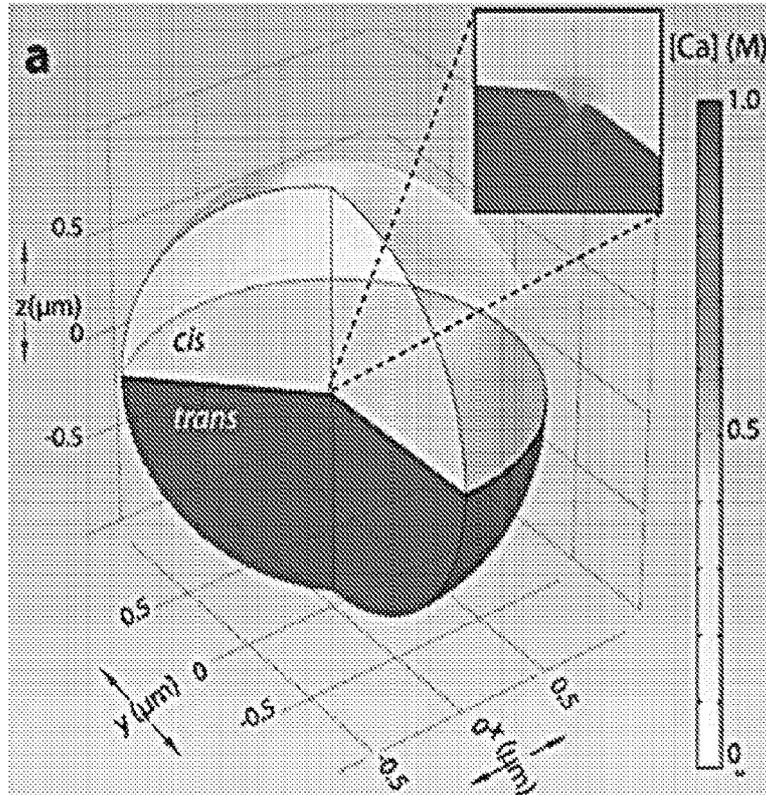


FIG. 4B

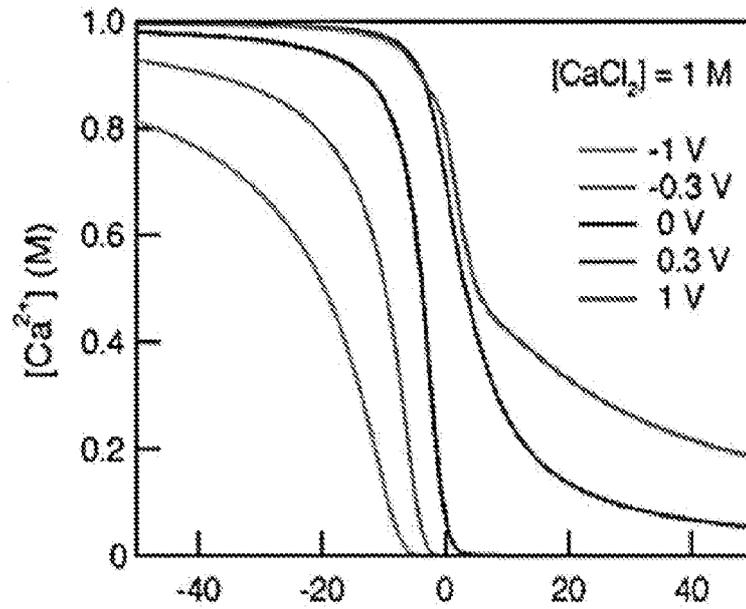


FIG. 4C

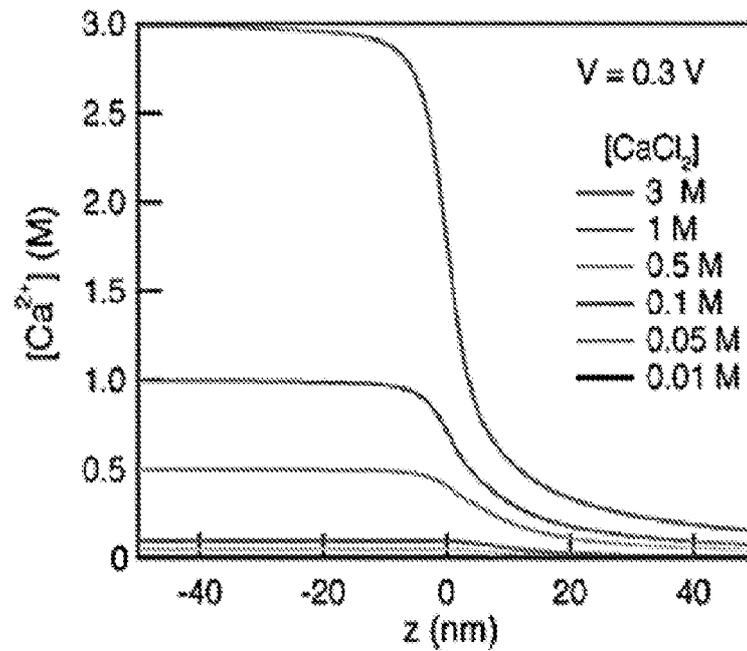


FIG. 5A

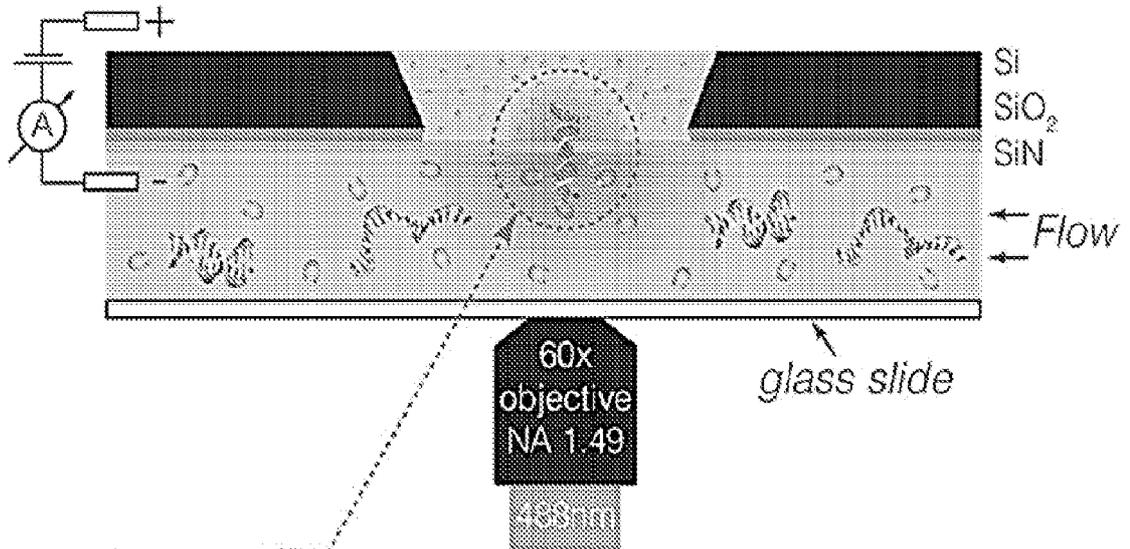


FIG. 5B

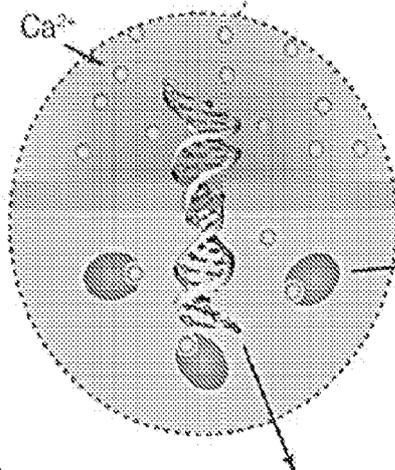


FIG. 5C

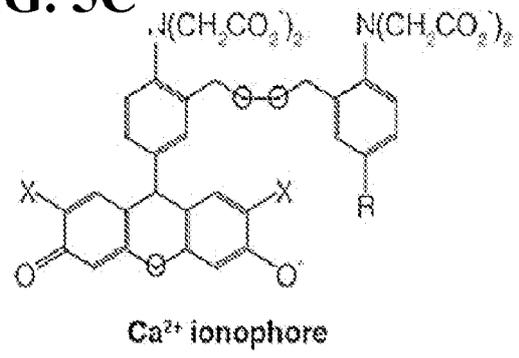


FIG. 5D

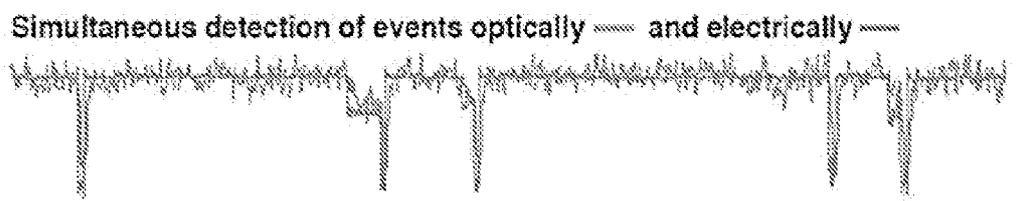


FIG. 6

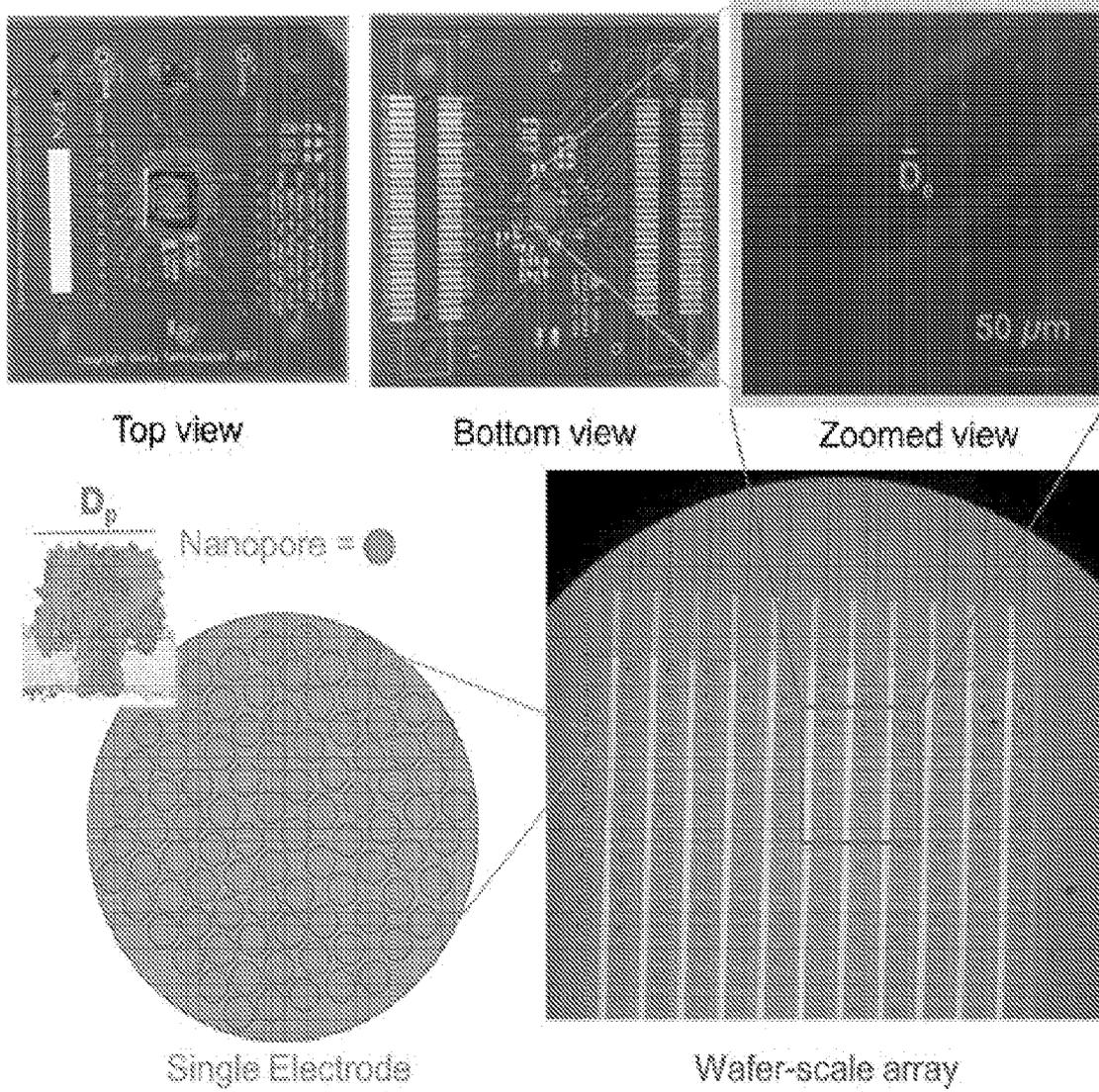
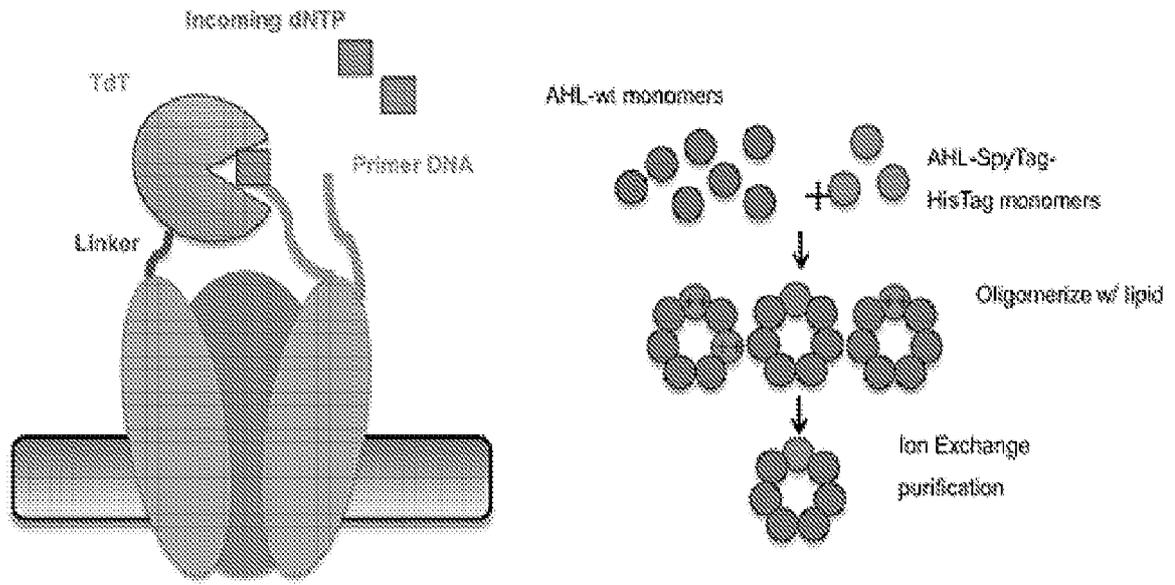


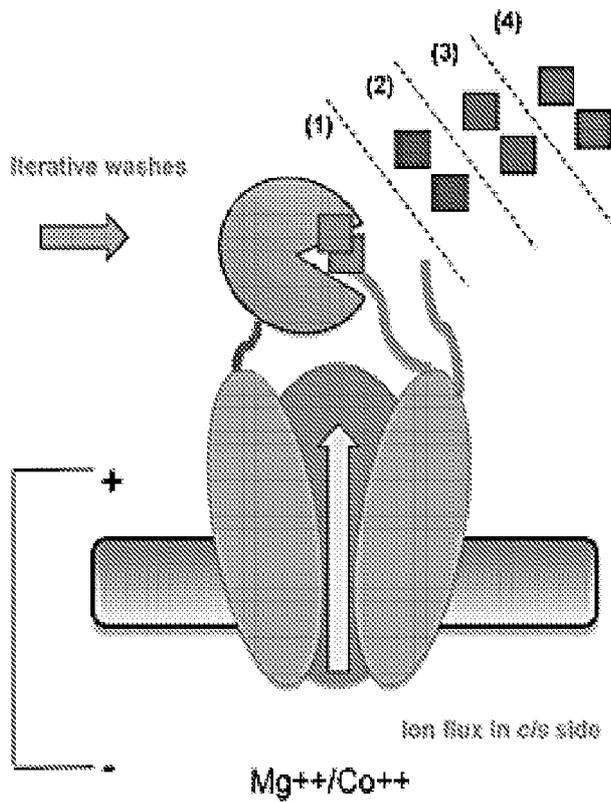
FIG. 7



Terminal deoxynucleotidyl transferase (TdT) catalyze phosphoryl transfer in the absence of a template

- 7 TdTs per pore → does not have to be purified
- May have multiple pores per electrode array
- Can accommodate 100s!

FIG. 8



Encoding Algorithm:

Flow cell with 5' immobilized primers -- flowing in alternating pulses of:

dATP-cation, then wash, then Mg⁺⁺/Co⁺⁺, then
 dCTP-cation, then wash, then Mg⁺⁺/Co⁺⁺, then
 dGTP-cation, then wash, then Mg⁺⁺/Co⁺⁺, then
 dTTP-cation, then wash, then Mg⁺⁺/Co⁺⁺.



(where cation = Ca⁺⁺ or other non-Mg⁺⁺ allowing binding but not covalent bond catalysis)

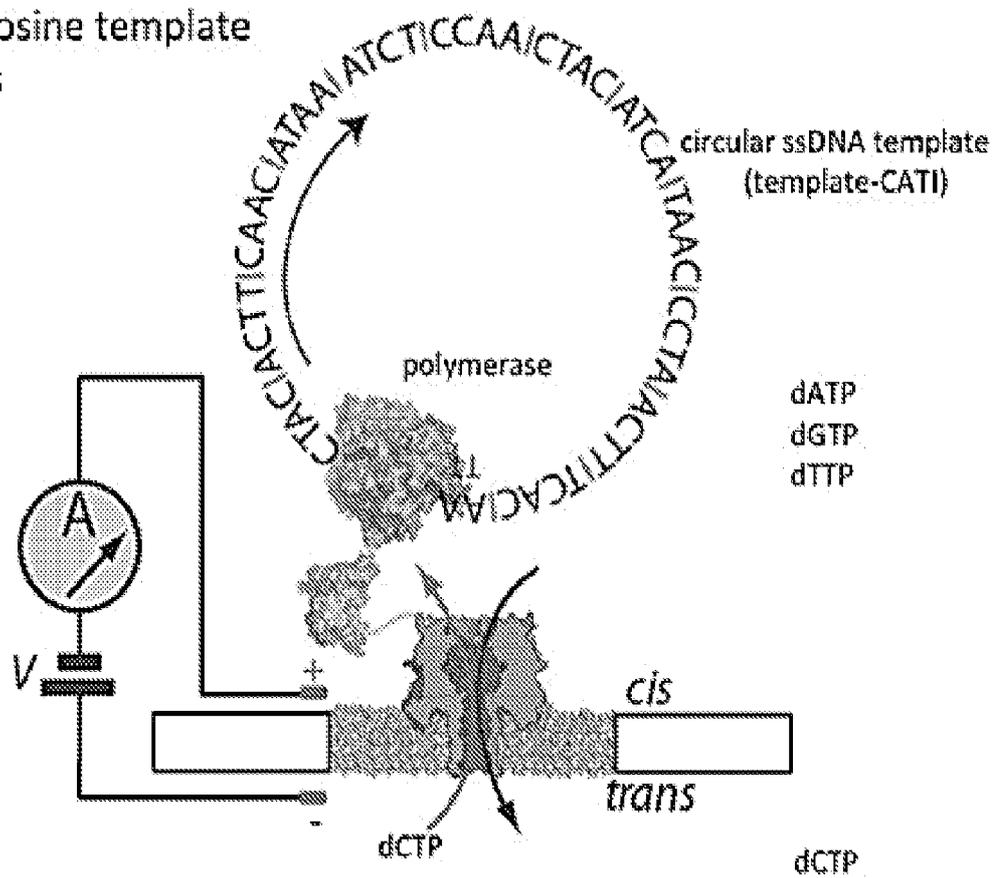
Decoding Algorithm:

State-transition based nanopore sequencing

- Ion flux modulation via voltage gradient
- 128k individually addressed DNA code
- Finite state machine based calculations

FIG. 9

Circular inosine template
with no Gs



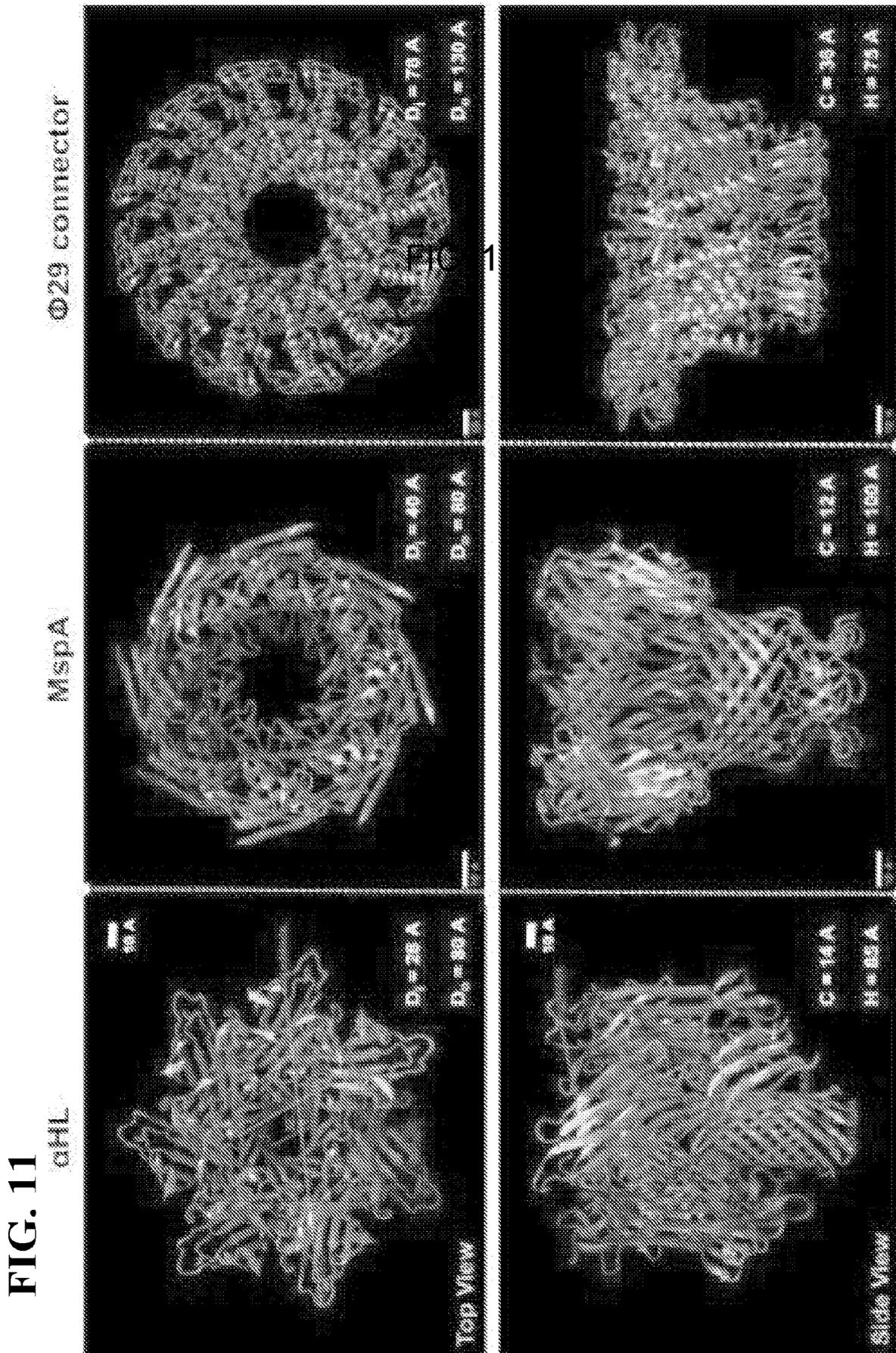


FIG. 12A

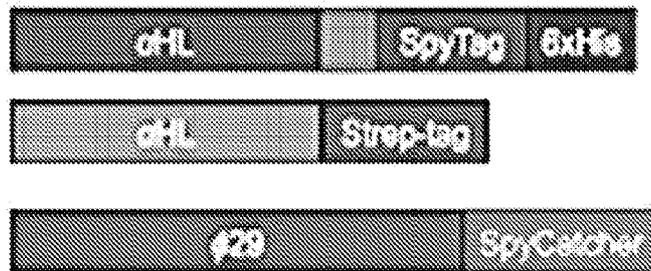


FIG. 12B

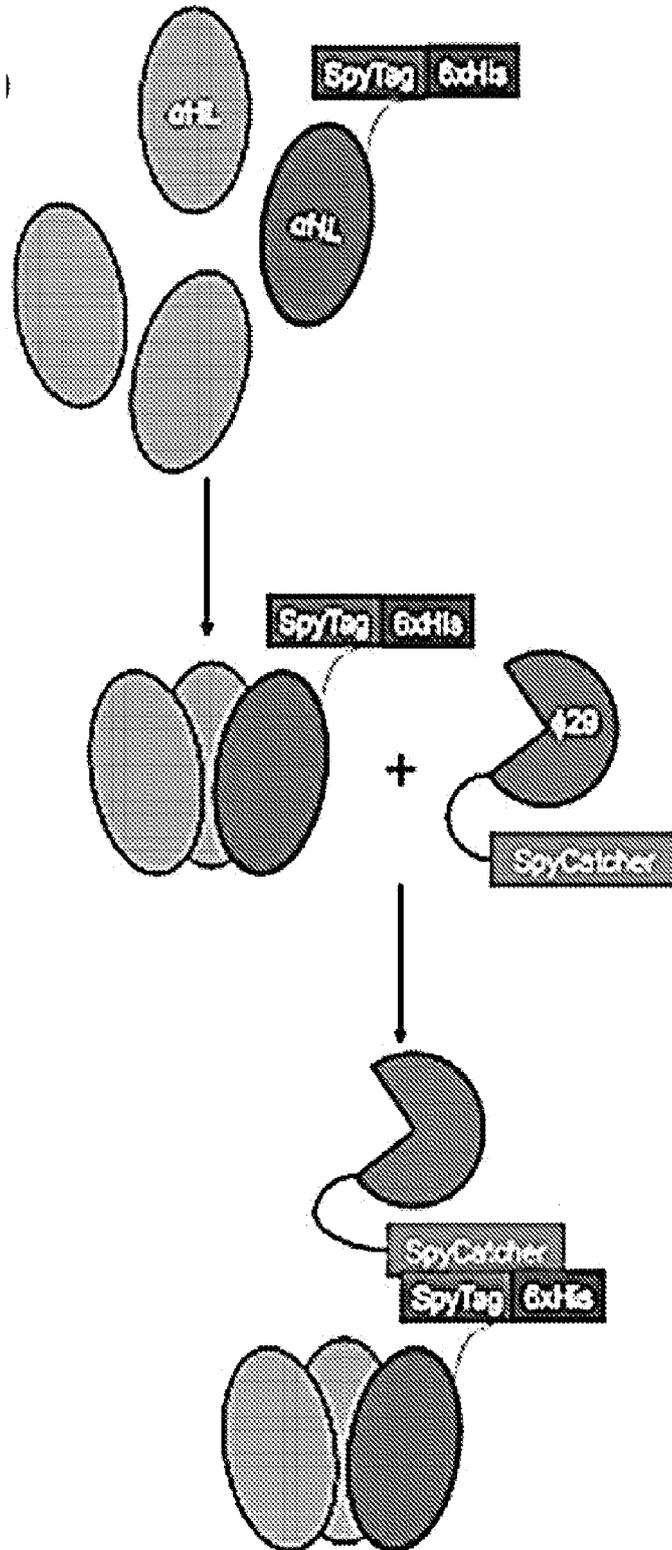


FIG. 12C

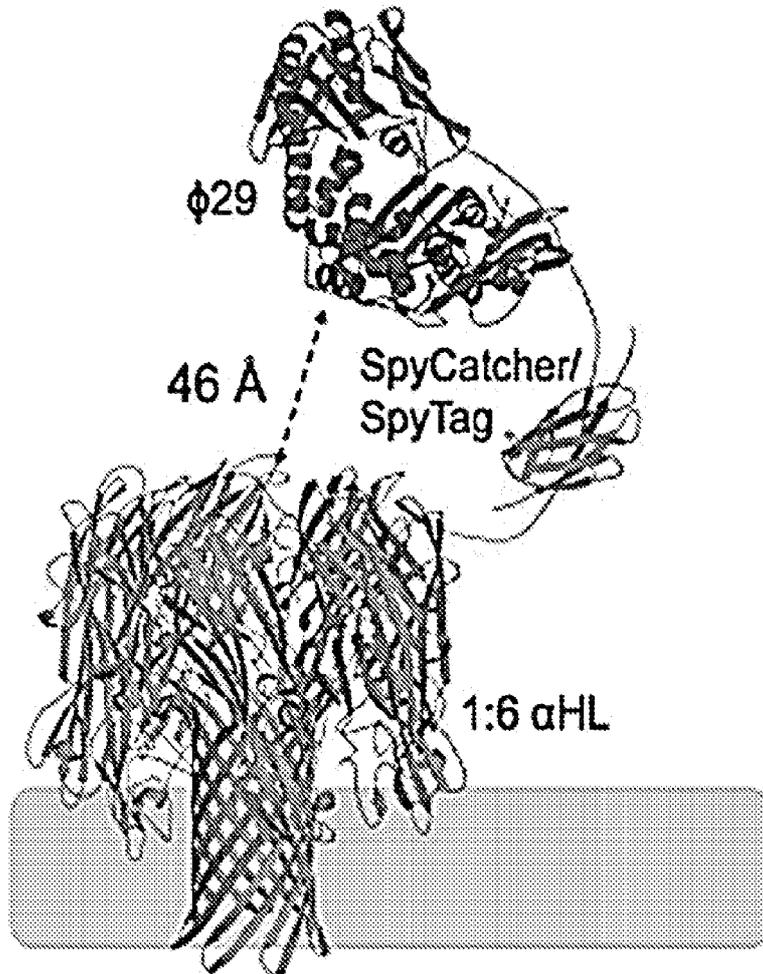


FIG. 12D

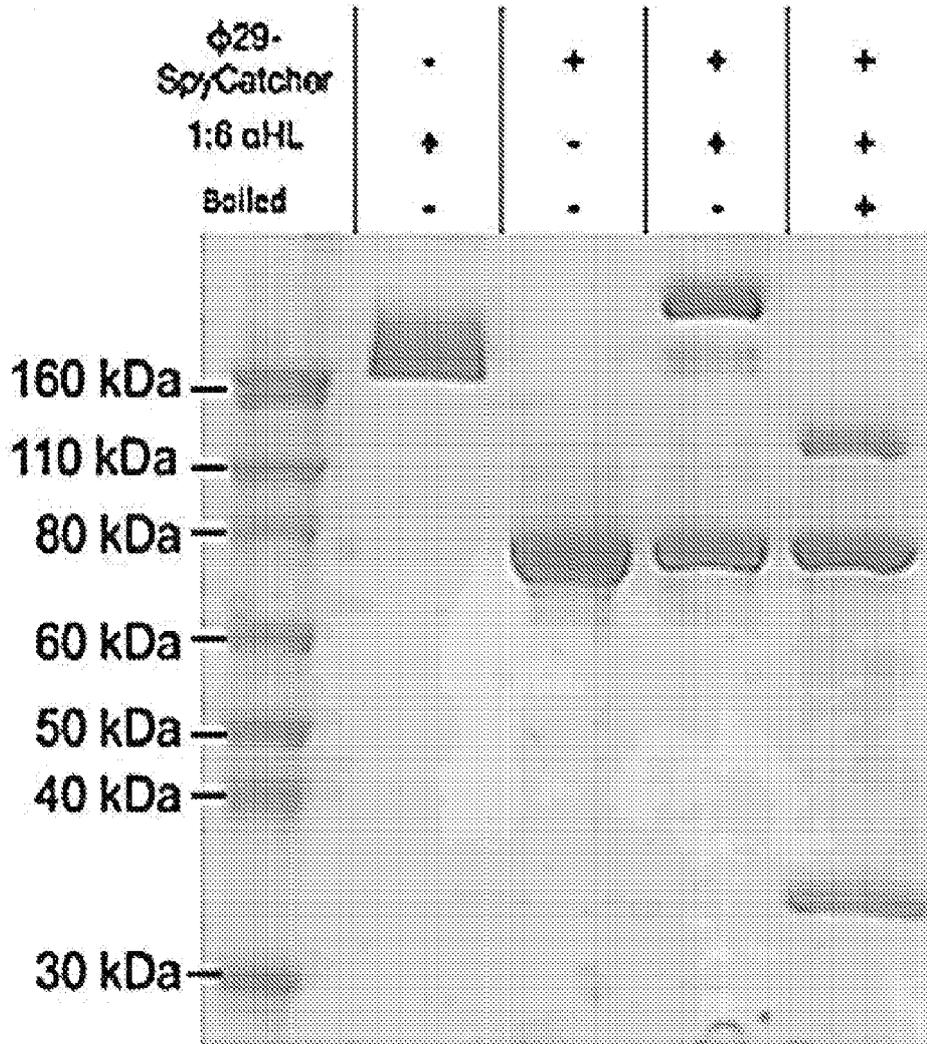


FIG. 13A

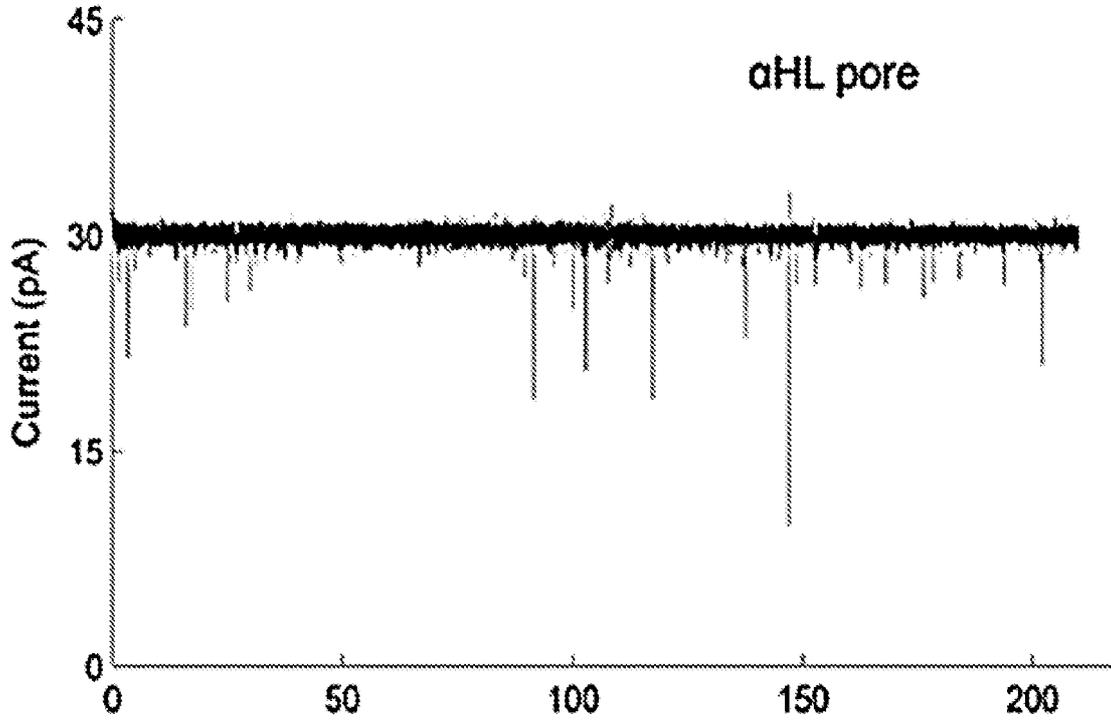


FIG. 13B

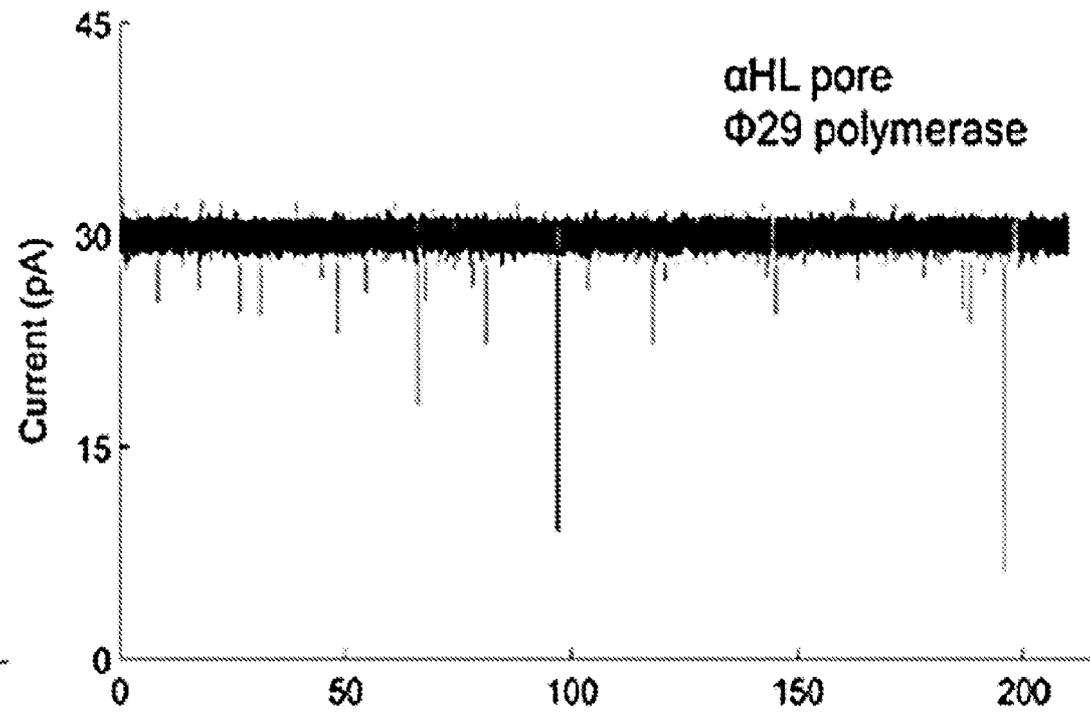


FIG. 13C

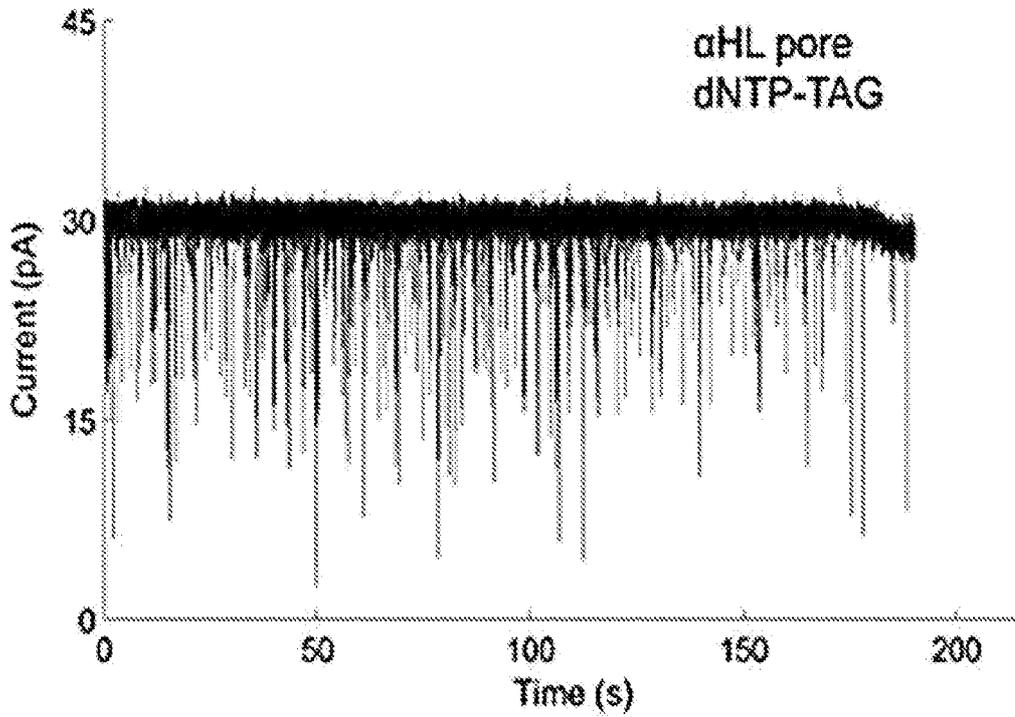


FIG. 13D

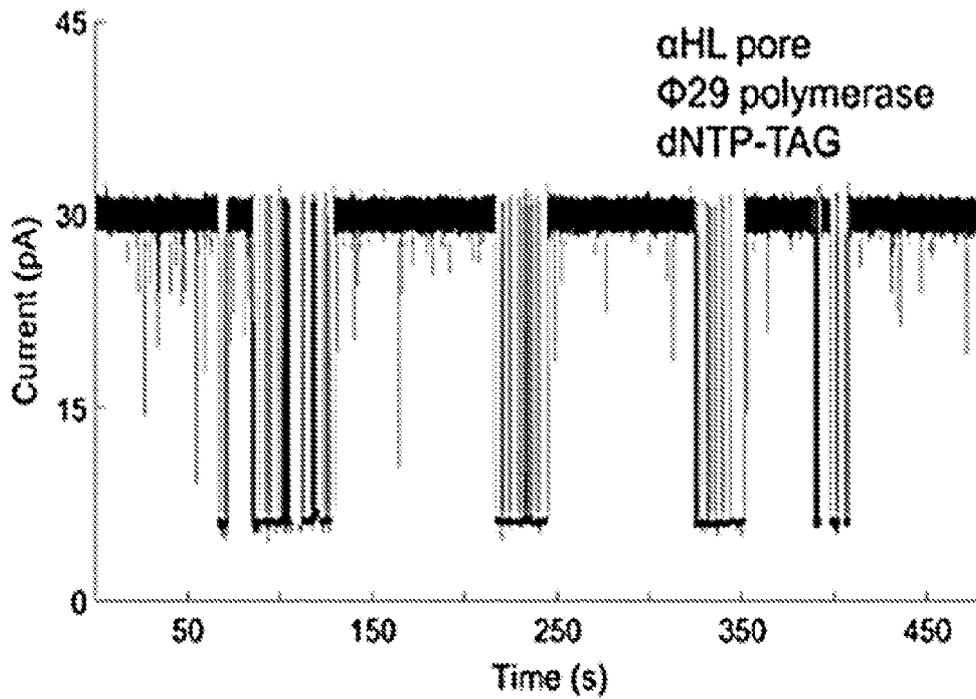


FIG. 14A

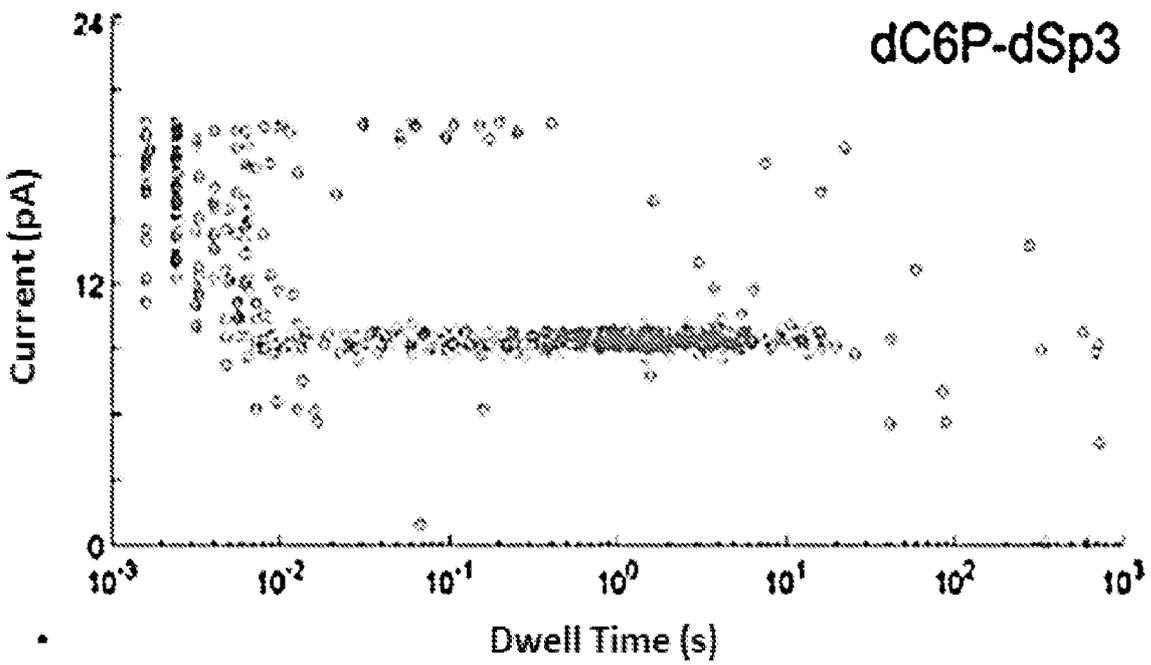
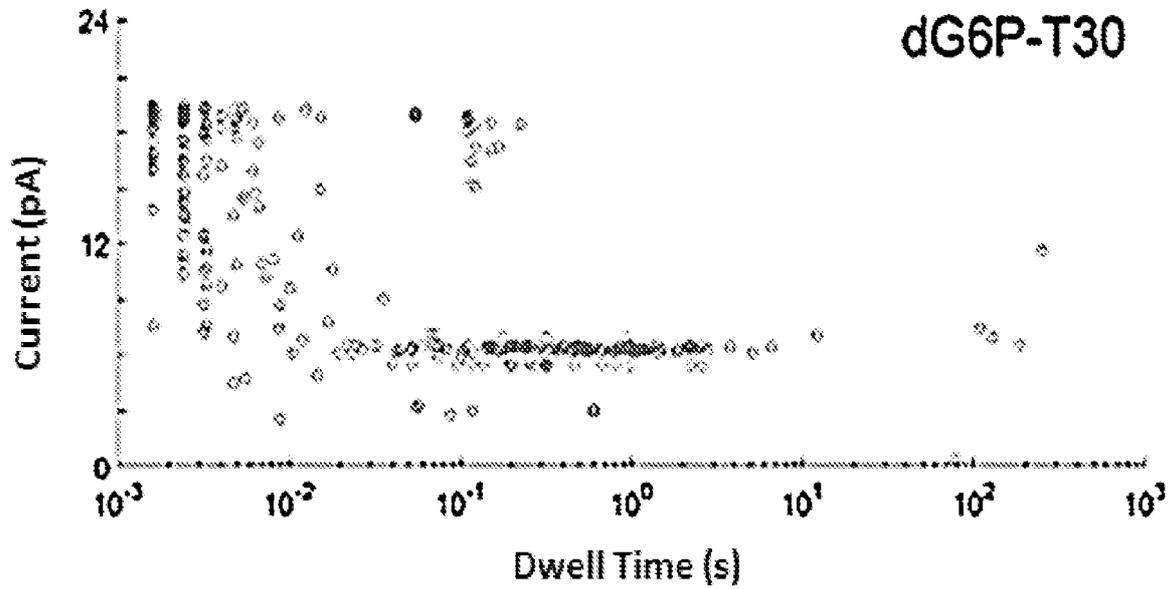


FIG. 14A (cont.)

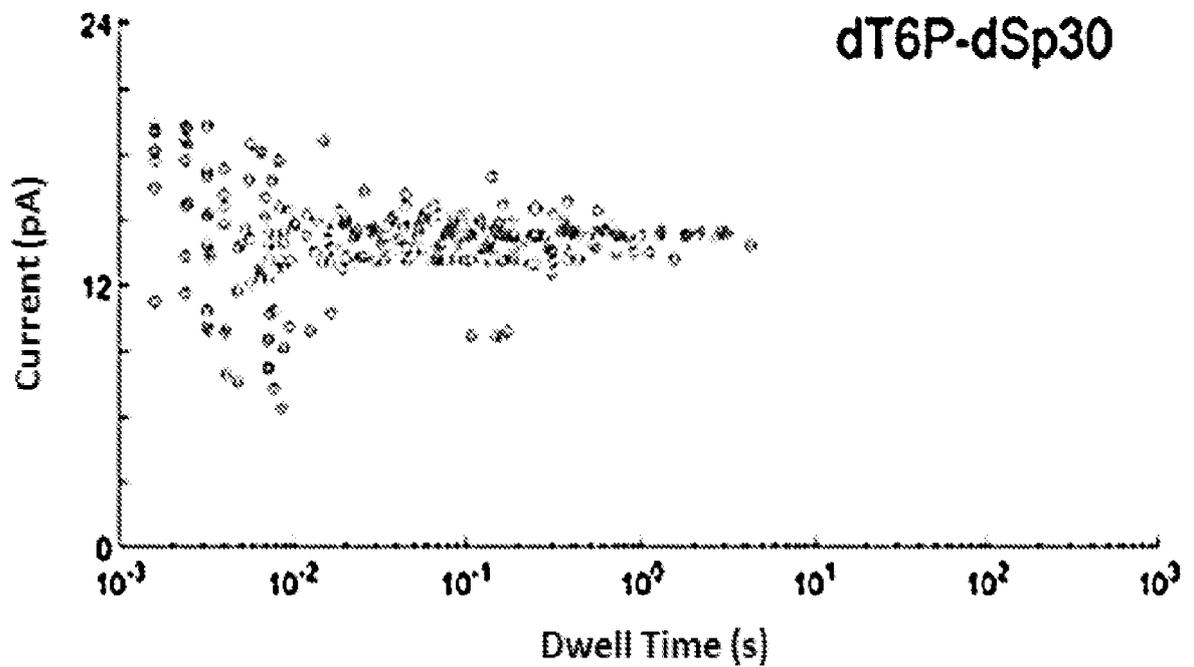
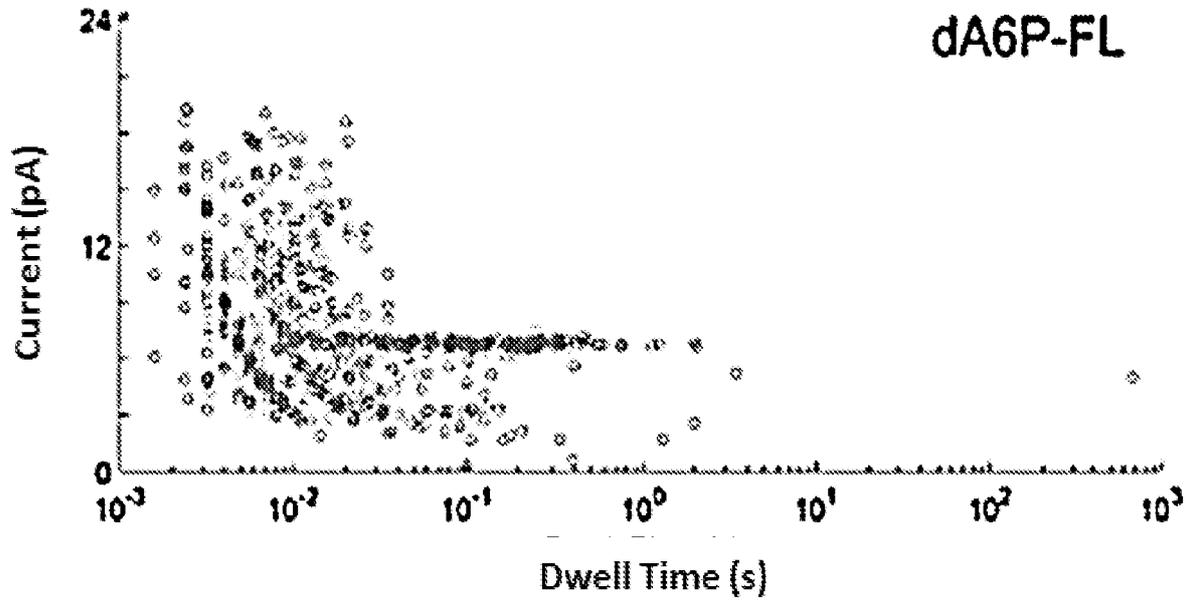


FIG. 14B

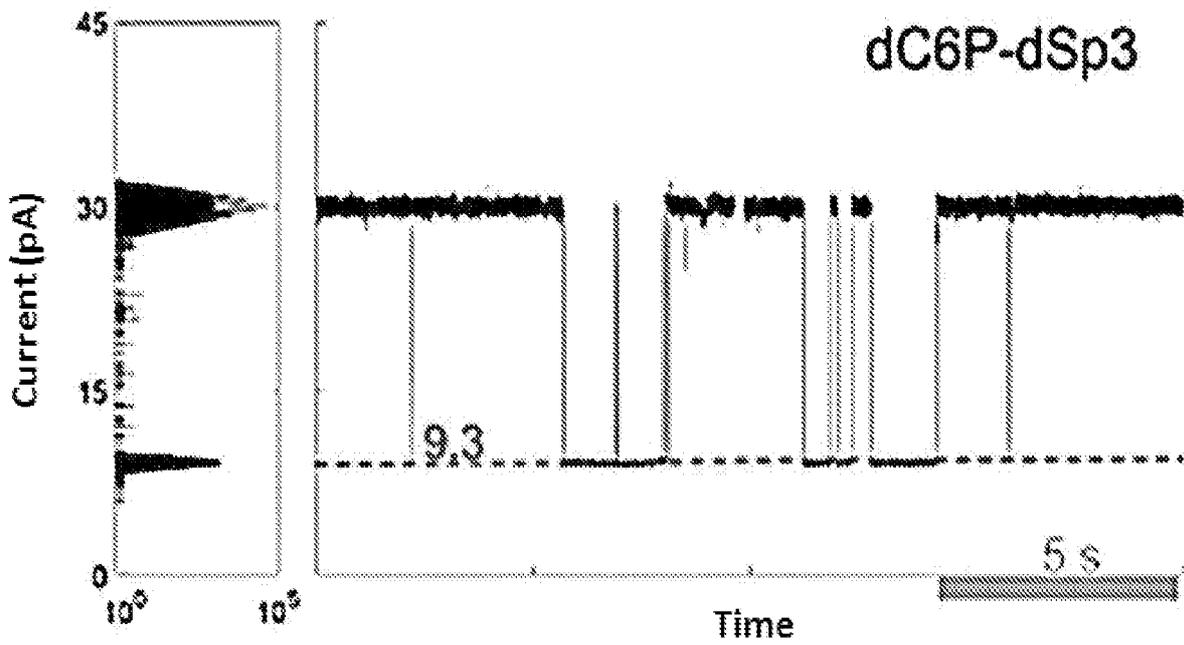
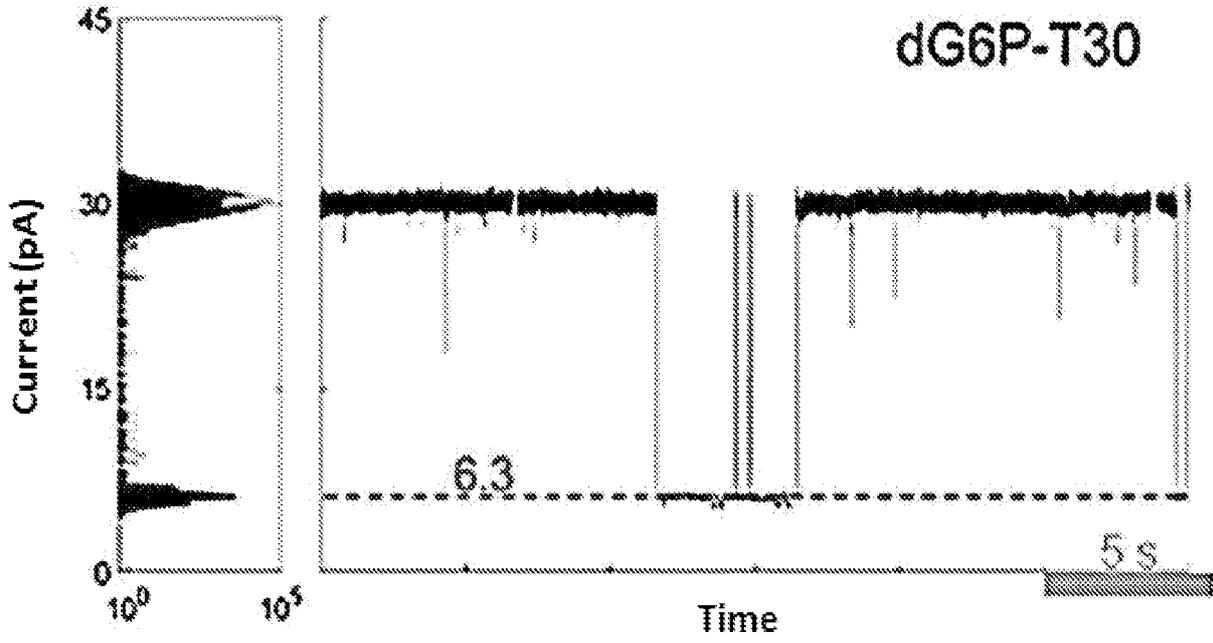


FIG. 14B (cont.)

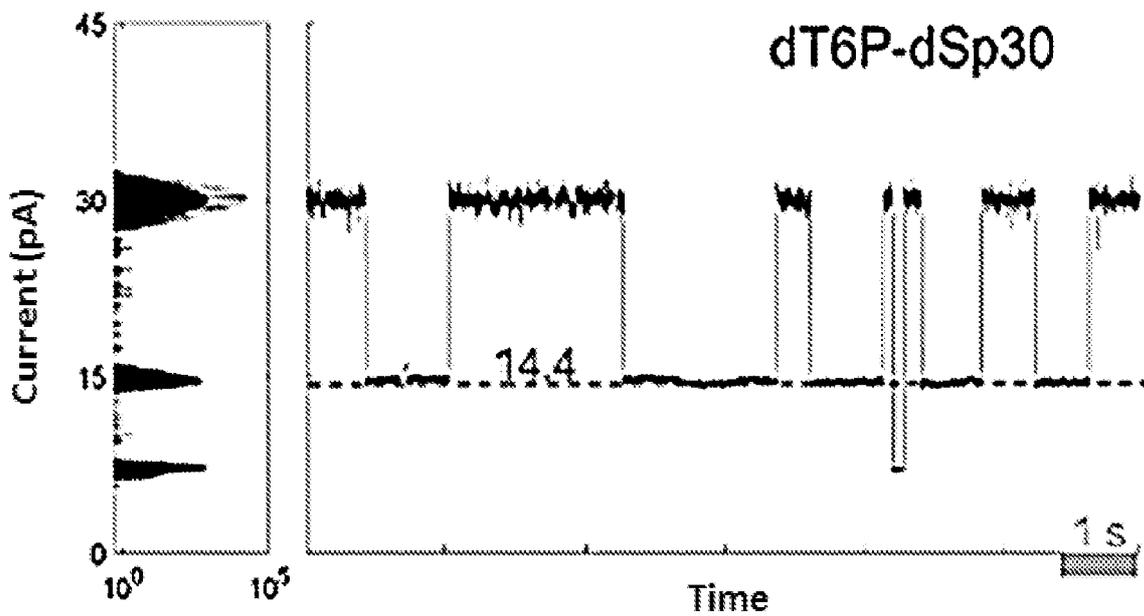
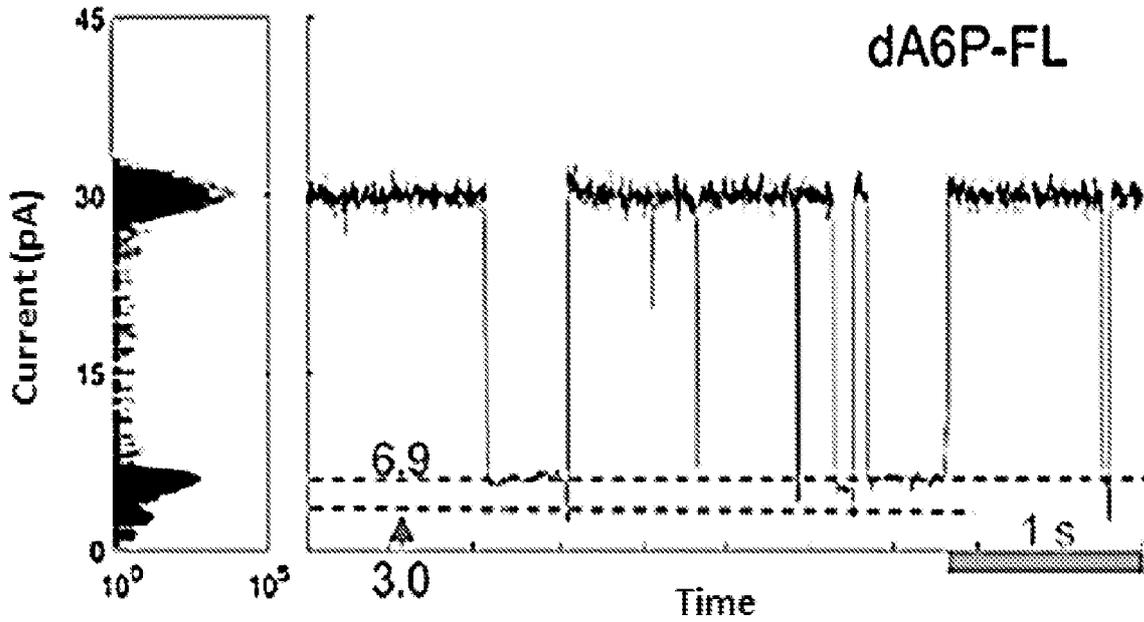


FIG. 15A

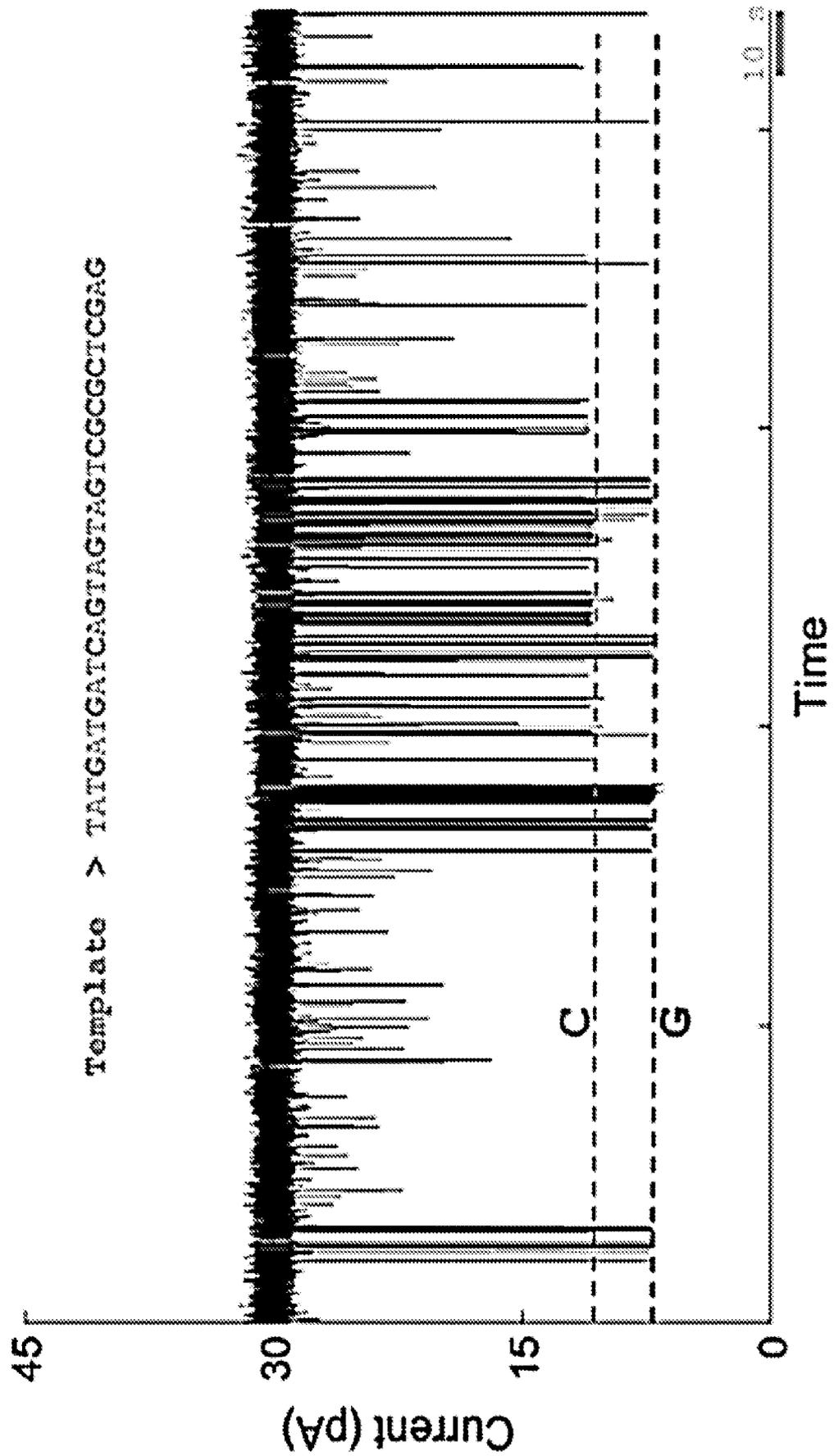


FIG. 15B

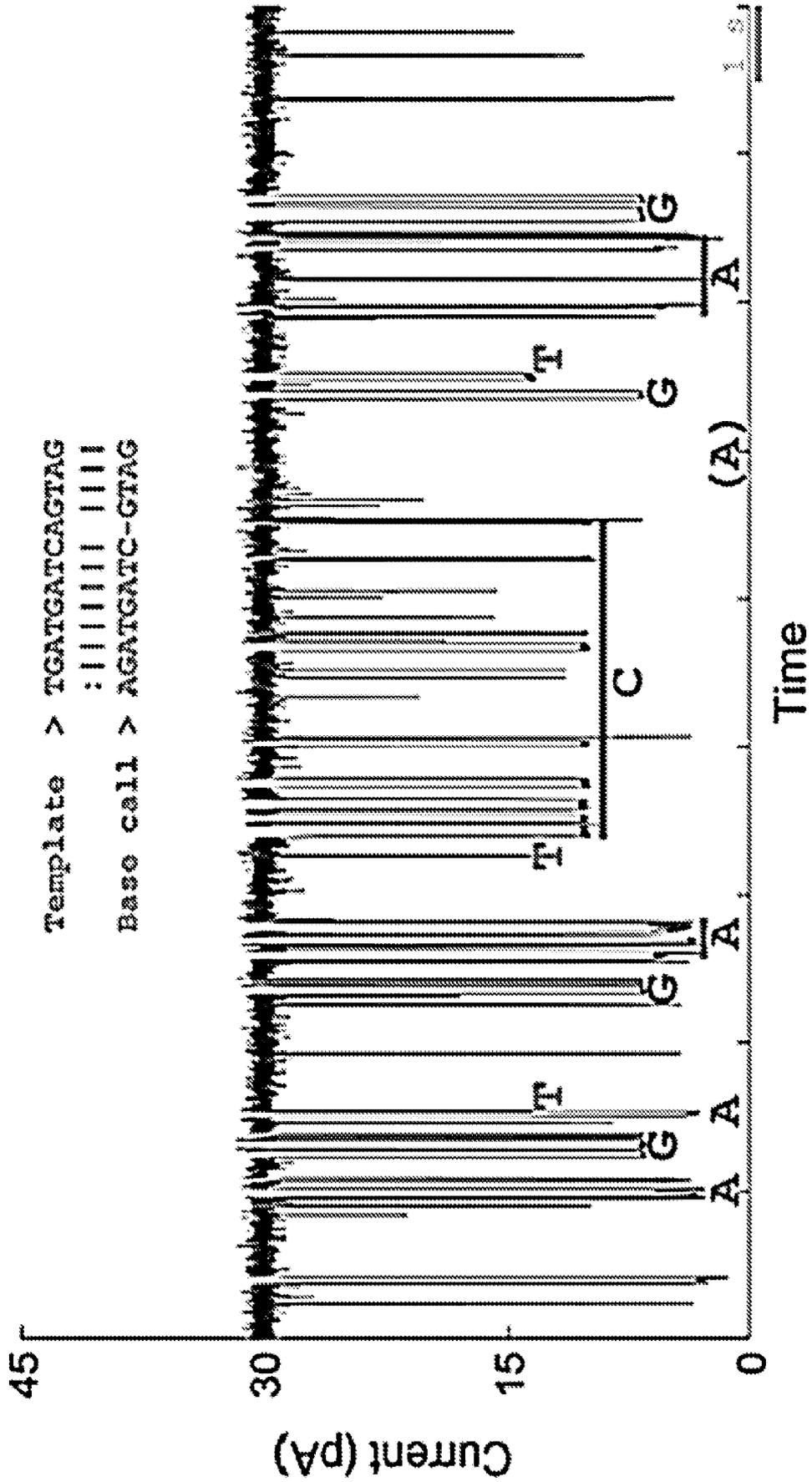


FIG. 16

> ϕ 29-SpyCatcher

MASWSHPQFEKGAETHMPRKMYSCDFETTTKVEDCRVWAYGYMNIEDHSEY
 KIGNSLDEFMAWVLKVQADLYFHNLKFDGAFIINWLERNGFKWSADGLPNT
 YNTIISRMGQWYMIDICLGKGRKIHTVIYDSLKKLPPVKKIAKDFKLT
 VLKGDIDYHKERPVGKITYPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAG
 SDSLKGFKDIITTKKFKKVFPTLSLGLDKEVRYAYRGGFTWLNDRFKEKEI
 GEGMVFVNSLYPAQMYSRLLPYGEPVFEFGKYVWDEDYPLHIQHRCFEF
 LKEGYIPTIQIKRSRFYKNEYLYKSSGGGEIADLWLSNVLDLELMKEHYDLYN
 VEYISGLKFKATTGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKFAS
 NPDVTGKVPYLNKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTITAAQ
 ACYDRIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKRAKYLRQ
 KTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVKCAGMTDKIKKEVTFENFKV
 GFSRKMKPKPVQVPGGVVLVDDFTTIKSGDYDIPTTENLYFOGAMVDTLS
 GLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELRDSSGKTIIST
 WISDGQVKDFYLYPGKYTEVETAAPDGYEVATAITFTVNEQGQVTVNGKAT
 KGDAHI

Color assignment: StrepTag ϕ 29 linker
 SpyCatcher

> α HL-SpyTag-His

MADEDINIKTGTTDIGSNIYVKTGELVTYDKENGMHKVYYSFLDDKNHMK
 KLLVIRTYGTIAGQYRVYSEEGANKSGLAWPSAFKVLQQLPDNEVAQIEDY
 YPRNSIDTKEYMSTLTYGFNGNVTGDDTCKIGGLIGANVEIGHTLKYVQPD
 FKTLLESPTDKKVGWVIFNNMVMQNWSPYDEDCWNPVYGNQLFMKTPNGS
 MKAANFLDPNKASSLLSSGFSPDPATVITMDRKAASKOOTLIDVLYEPVRD
 DYQLHWTSENWKGNTKDKWIDPSSERYKIDWEKKEEMTNGGSSGGSSGGAH
 IVMVDAYKPTKKGHHHHHH

Colors assignment: α HL linker SpyTag 6xHis

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/28272

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68, C07K 14/31 ; C12N 9/96, 9/12 (2017.01)
CPC - C12N 9/1252, 9/96; C07K 14/31 ; C12Y 207/07007

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

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

See Search History Document

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

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y --- A	US 2010/0035260 A1 (Olasagasti et al.) 11 February 2010 (11.02.2010). Especially para [0027], [0049], [0098], [0099], [0100], [0208], [0213], [0213]	1-28, 76-90, 92-96, 98-110, 112-121, 124, 141-156 ----- 64-74, 91, 97, 111, 122, 123, 163-168, 173-175 ----- 75
Y --- A	Nelson et al. phi29 DNA polymerase based rolling circle amplification of templates for DNA sequencing. Biotechniques June 2002 Supplement pages 44-47. Especially abstract, pg 44 col 2 para 1, pg 45 col 1 para 3	64-74, 91,173-175 ----- 75
Y ---	Fuller et al. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. Proc Nat Acad Sci ePub 18 April 2016 Vol 113 no 19 Pages 5233-5238. Especially Supporting information [online] pg 1 col 2 para 2	97, 111, 122, 163-168, 173-175
Y	US 2012/0091011 A1 (Graham et al.) 19 April 2012 (19.04.2012). Especially para [0048]	123
Y ---	Alseth et al. Inosine in DNA and RNA. Curr Opin Genet Dev June 2014 Vol 26 Pages 116-123. Especially pg 117 fig 1.	67

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

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

Date of the actual completion of the international search 18 August 2017	Date of mailing of the international search report 2.5 SEP 2017
---	---

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774
---	--

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/28272

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 I7(2)(a) for the following reasons:

- 1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

- 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:
— Go to Extra Sheet for continuation—

- 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.:
1-28, 64-124, 141-156, 163-168, 173-175 limited to phi-29 DNA polymerase
- 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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 17/28272

C (Continuation). DOCUMENTS CONSIDERED \emptyset BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2013/0053544 A1 (Howarth) 28 February 2013 (28.02.2013). Especially para [0090]	163-168, 173-175
A	US 2015/0269313 A1 (President and Fellows of Harvard College) 24 September 2015 (24.09.2015). Especially claim 1, para [0019]	75

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 17/28272

— continuation of Box III (Lack of Unity of Invention)—

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I+: Claims 1-15, 29-46, 64-75, 92-105, 158-175, drawn to a method of DNA synthesis comprising applying a voltage across a lipid bilayer wherein an enzyme is activated and catalyzes DNA synthesis in response to the voltage. The method of DNA synthesis will be searched to the extent that the enzyme encompasses phi-29 DNA polymerase (e.g., claim 4). It is believed that claims 1-15, 64-75, 92-105, 163-168, 173-175 read on this first named invention and thus these claims will be searched without fee to the extent that they encompass phi-29 DNA polymerase. Additional enzymes will be searched upon payment of additional fees. Applicant must specify the claims that encompass any additional elected enzymes. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be: a terminal deoxynucleotidyl transferase (TdT) (claims 1-15, 29-46, 92-105, 158-175).

Group II+: Claims 16-28, 47-63, 76-91, 106-157, drawn to a composition (system) comprising an enzyme, ssDNA, and a nanopore in a lipid bilayer. Group II+ will be searched upon payment of additional fee(s). The composition (system) may be searched, for example, to the extent that the enzyme encompasses phi-29 DNA polymerase (e.g., claim 19) for an additional fee and election as such. It is believed that claims 16-28, 76-91, 106-124, 141-156 read on this exemplary invention. Additional enzymes will be searched upon the payment of additional fees. Applicants must indicate, if applicable, which claims read on this named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the '+' group(s) will result in only the first named invention to be searched/examined. An exemplary election would be terminal deoxynucleotidyl transferase (TdT)(claims 16-28, 47-63, 76-91, 106-117, 119-157).

The inventions listed as Groups I+ and II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features:

Group I+ has the special technical feature of a method of DNA synthesis or encoding information, not required by Group II+.

Group II+ has the special technical feature of a composition (system) comprising an enzyme, ssDNA, and a nanopore in a lipid bilayer, not required by Group I+.

Among the inventions listed as Groups I+ and II+ are the specific enzymes, recited therein. Each invention requires a specific enzyme, not required by any of the other inventions.

Common Technical Feature:

1. Groups I+ inventions share the common technical feature of a method comprising providing an encoding unit comprising an enzyme, a single-stranded DNA (ssDNA) and a nanopore, providing a lipid bilayer having on opposite sides a cis and a trans reservoir each having a different buffer composition, wherein the nanopore is within the lipid bilayer and the enzyme and the ssDNA are in the cis reservoir, and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer, and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage.
2. Groups I+ and II+ share the common technical feature of a protein nanopore embedded in a lipid bilayer.
3. Groups I+ and II+ share the common technical feature of a [template dependent] phi-29 DNA polymerase or [template independent] terminal deoxynucleotidyl transferase (TdT)
4. Group II+ inventions share the common technical feature of terminal deoxynucleotidyl transferase covalently attachment to a protein nanopore in a bilayer.

However, said common technical features do not represent a contribution over the prior art, and is obvious over US 2010/0035260 A1 to Olasagasti et al. (hereinafter "Olasagasti"), in view of WO 2015167972 A1 (Tsavachidou), in view of US 2010/0297638 A1 to Bayley et al. (hereinafter "Bayley").

— continued on next sheet-----

— continuation of Box III (Lack of Unity of Invention)—

As to common technical features # 1 and #2, Olasagasti teaches providing an encoding unit comprising an enzyme (para [0100]; "Enzymes that interact with polynucleotides are known to those of skill in the art and can include, but are not limited to, DNA polymerase such as a DNA polymerase selected from...Phi-29 DNA polymerase"), a single-stranded DNA (ssDNA) (para [0099]; ssDNA template) and a nanopore, providing a lipid bilayer (para [0243]; In the biological nanopore setup, a planar lipid bilayer is created across a 20 nm TELON aperture in a KCl solution. A single alpha-hemolysin protein channel is inserted into the planar lipid"); having on opposite sides a cis and a trans reservoir each having a different buffer composition (para [0098]; Mg²⁺ at greater than millimolar concentrations are added to the trans compartment. The cis compartment comprises all the other reagents, enzymes, and substrates necessary for catalysis. The cis compartment also comprises trace concentrations of EDTA (at about 0.1 mM) to ensure that free [Mg²⁺] on the cis side is effectively zero in bulk phase. Since Mg²⁺ is a divalent cation under physiological conditions, an applied voltage that attracts a polynucleotide into the nanopore (trans side +) would drive Mg²⁺ in the opposite direction towards the cis compartment. Thus, in the volume (area of medium) immediately adjacent to the pore aperture, the free [Mg²⁺] is a function of the voltage-driven flux from the trans side to the cis side across the nanopore minus the Mg²⁺), wherein the nanopore is within the lipid bilayer (para [0243]) and the enzyme and the ssDNA are in the cis reservoir (para [0098]), and applying a voltage across the lipid bilayer via an electrode wherein the electrode can modulate the voltage across the lipid bilayer (para [0098]), and wherein the enzyme is activated and catalyzes DNA synthesis in response to the voltage (para [0208]; "the cis compartment contains all components necessary for DNA polymerase activity except for Mg²⁺. Thus, no catalysis can take place, (b) When voltage is applied (trans-side +), Mg²⁺ is driven across the pore into the cis compartment. (c) When a DNA-polymerase complex is captured by the pore, the Mg²⁺ concentration in the volume immediately adjacent to the pore is sufficiently high to permit Mg²⁺ occupation of the two critical loci in the enzyme's catalytic site. Polymerization of the copied strand can then occur").

As to common technical feature #3, Tsavachidou teaches terminal deoxynucleotidyl transferase (pg 64 para 1; "exposing the nucleic acid molecule and its parts to conditions to cause polymerization, and to a template-independent polymerization reaction solution comprising terminal deoxynucleotidyl transferase (TdT) molecules"; pg 97 para 3; "In a certain embodiment, tail tags comprising labels that alter conductivity when passed through a suitable nanopore device") and Olasagasti teaches phi-29 DNA polymerase (para [0100]).

As to common technical feature #4, Bayley teaches an adaptor reporter molecule covalently attached to an alpha-hemolysin nanopore in a bilayer (claim 1; 1. A transmembrane protein pore for use in detecting an analyte in a sample, which comprises a molecular adaptor that facilitates an interaction between the pore and the analyte, wherein the adaptor is covalently attached to the pore in an orientation that allows the analyte to be detected using the pore, wherein the pore is modified to facilitate the orientation of the adaptor"). Bayley does not teach covalent attachment of TdT. However, it would have been obvious to one of ordinary skill in the art that conjugation methods well-known in the art could have been utilized to covalently attach TdT to a protein nanopore, where the nanopore was possibly mutated specifically for the site-specific conjugation. Furthermore, an artisan would have been motivated to attach the polymerase to the nanopore, based on the discussion set forth in common technical feature #1, where it is indicated that the essential enzyme-activating cofactor (e.g. Mg²⁺) concentration would be at its highest near the nanopore (Olasagasti para [0098]; "in the volume (area of medium) immediately adjacent to the pore aperture, the free [Mg²⁺] is a function of the voltage-driven flux from the trans side to the cis side across the nanopore minus the Mg²⁺").

As the common technical features were known in the art at the time of the invention, they cannot be considered common special technical features that would otherwise unify the groups.

Therefore, Groups I+ and II+ lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.