

METHODS FOR SEQUENCING BIOPOLYMERS

GOVERNMENT SUPPORT

[0001] This invention was made with government support under R21 HG010522 awarded by the National Institutes of Health. The government has certain rights in the invention.

RELATED APPLICATIONS

[0002] This application claims priority to and the benefit of U.S. Provisional Patent Application No. 63/018,352 filed April 30, 2020, which is incorporated herein by reference in its entirety for all purposes.

FIELD

[0003] The present disclosure provides devices, systems, and methods related to sequencing a biopolymer. In particular, the present disclosure relates to methods for sequencing a polynucleotide using a bioelectronic device that obtains a bioelectronic signature of polymerase activity based on current fluctuations as complementary nucleotidepolyphosphate monomers having distinctive charges are incorporated into the template polynucleotide.

BACKGROUND

[0004] As proteins perform their various functions, movements are generated that underlie these functions. The ability to develop devices, systems, and methods that measure the electrical characteristics corresponding to the fluctuations generated by an active protein can be a basis for label-free detection and analysis of protein function. For example, monitoring the functional fluctuations of an active enzyme may provide a rapid and simple method of screening candidate drug molecules that affect the enzyme's function. In other cases, the ability to monitor the fluctuations of proteins that process biopolymers (e.g., carbohydrates, polypeptides, nucleic acids, and the like) may reveal new information about their conformational changes and how those changes are linked to function. Additionally, diagnostic and analytical devices can be developed to take advantage of the electrical characteristics produced by active proteins, providing new ways to leverage biomechanical properties for practical use.

SUMMARY

[0005] Embodiments of the present disclosure include a method for sequencing a polynucleotide using a bioelectronic device. In accordance with these embodiments, the method includes introducing a template polynucleotide to a bioelectronic device. In some embodiments, the bioelectronic device comprises a polymerase functionally coupled to at least one of a first electrode and a second electrode. In some embodiments, the method further includes introducing a solution comprising four nucleotidepolyphosphate monomers to the device comprising the template polynucleotide. In some embodiments, at least three of the four nucleotidepolyphosphate monomers present in the solution comprises a distinctive charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, the method further includes obtaining a bioelectronic signature of polymerase activity based on current fluctuations as each complementary nucleotidepolyphosphate monomer is incorporated into the template polynucleotide. In some embodiments, at least one characteristic of the bioelectronic signature identifies each of the complementary nucleotidepolyphosphate monomers incorporated into the template polynucleotide.

[0006] In some embodiments, the bioelectronic signature comprises a closed period corresponding to the polymerase being in a closed state.

[0007] In some embodiments, the at least one characteristic of the bioelectronic signature comprises current amplitude level.

[0008] In some embodiments, at least one of the nucleotidepolyphosphate monomers present in the solution comprises a distinctive negative charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, at least two of the nucleotidepolyphosphate monomers present in the solution comprise distinctive negative charges relative to their corresponding standard nucleotidetriphosphate monomer. In some embodiments, at least three of the nucleotidepolyphosphate monomers present in the solution comprise distinctive negative charges relative to their corresponding standard nucleotidetriphosphate monomer. In some embodiments, the negative charge is conferred by at least one additional phosphate group relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, the negative charge is conferred by 1 to 20 additional phosphate groups. In some embodiments, a distinctive negative charge on the at least one nucleotidepolyphosphate monomer corresponds to an increased or decreased current amplitude level relative to the current amplitude level of its corresponding standard nucleotidetriphosphate monomer.

[0009] In some embodiments, at least one of the nucleotidetriphosphate monomers present in the solution comprises a distinctive positive charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, at least two of the nucleotidetriphosphate monomers present in the solution comprises a distinctive positive charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, at least three of the nucleotidetriphosphate monomers present in the solution comprises a distinctive positive charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, a distinctive positive charge on the at least one nucleotidetriphosphate monomer corresponds to an increased or decreased current amplitude level relative to the current amplitude level of its corresponding standard nucleotidetriphosphate monomer.

[0010] In some embodiments, the exonuclease activity of the polymerase is disabled.

[0011] In some embodiments, the polymerase is functionally coupled to the first and second electrodes using a linker comprising streptavidin.

[0012] In some embodiments, linker is attached to a region of the polymerase that is inactive.

[0013] In some embodiments, the method includes applying a voltage bias between the first and second electrodes that is 100mV or less.

[0014] In some embodiments, the standard nucleotidetriphosphate monomers include adenine (dATP), cytosine (dCTP), guanine (dGTP), and thymine (dTTP).

[0015] In some embodiments, the first and/or the second electrode comprises gold, palladium, platinum, silver, copper, or any alloys thereof.

[0016] In some embodiments, the device comprises a dielectric layer at least partially covering a top surface of the first and/or second electrode.

[0017] In some embodiments, the first electrode and second electrode are positioned so that about a 1 nm to about a 50 nm gap is formed between the two electrodes.

[0018] Embodiments of the present disclosure also include a bioelectronic device that includes a first electrode and a second electrode separated by a gap, and a protein attached to the first and second electrodes via a linker comprising a distinctive electrical charge. In some embodiments, the distinctive electrical charge modulates conductance through the protein.

[0019] In some embodiments, the linker comprises a peptide or polypeptide. In some embodiments, the linker comprises streptavidin. In some embodiments, the protein is biotinylated.

[0020] In some embodiments, the linker comprises a distinctive negative charge. In some embodiments, the distinctive negative charge is conferred by addition of a glutamate moiety, an aspartate moiety, or a combination thereof, coupled to the streptavidin. In some embodiments, the distinctive negative charge increases the conductance through the protein.

[0021] In some embodiments, the linker comprises a distinctive positive charge. In some embodiments, the distinctive positive charge is conferred by addition of an arginine moiety, a histidine moiety, a lysine moiety, or a combination thereof, coupled to the streptavidin. In some embodiments, the distinctive positive charge increases or decreases the conductance through the protein.

[0022] In some embodiments, the first and/or the second electrode comprises gold, palladium, platinum, silver, copper, or any alloys thereof.

[0023] In some embodiments, the protein is selected from the group consisting of a polymerase, a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease.

[0024] In some embodiments, the linker is attached to an inactive region of the protein.

[0025] In some embodiments, the linker comprises streptavidin coupled to a polyglutamate moiety, the protein comprises a polymerase, and the first and second electrodes comprise palladium, platinum, or any alloys thereof.

[0026] Embodiments of the present disclosure also include a method of modulating electronic conductance through a protein using any of the bioelectronic devices described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] FIG. 1: Representations of examples of deoxynucleotidepolyphosphates in which varying numbers of phosphates contribute different amounts of charge to each deoxynucleotidepolyphosphate.

[0028] FIG. 2: Representative graph showing how the electronic conductance of a polymerase molecule depends upon local charge (expressed in terms of surface potential).

[0029] FIGS. 3A-3B: Representative graphs of current vs. time as measured through a polymerase incorporating deoxynucleotidetriphosphates. FIG. 3A demonstrates that the time duration of the various features depends upon the nucleotide being incorporated, but the amplitude of the current does not. FIG. 3B demonstrates that different amounts of charge carried by each of the four nucleotides results in different current levels in the closed state of the polymerase. The sequence readout is shown above the trace.

[0030] FIG. 4: Representative diagram of the backbone structure of a hexagluamate mutant streptavidin showing the charged tails (E6) attached to the C termini of each of the streptavidin monomers.

[0031] FIGS. 5A-5B: Representative conductance distributions for wildtype (FIG. 5A) and the hexagluamate mutant streptavidin (FIG. 5B) measured at a 2.5 nm gap using Pd electrodes.

[0032] FIGS. 6A-6B: Representative diagrams of streptavidin-linked Φ 29 polymerases bridging electrode pairs using wildtype streptavidin (FIG. 6A) and the hexagluamate mutant streptavidin (FIG. 6B). The corresponding conductance distributions measured at a 4.5 nm gap with Pd electrodes are provided in the graphs below each diagram.

DETAILED DESCRIPTION

[0001] In the following detailed description, reference is made to the accompanying drawings that form a part hereof, and in which are shown by way of illustration embodiments that may be practiced. It is to be understood that other embodiments may be utilized and structural or logical changes may be made without departing from the scope. Therefore, the following detailed description is not to be taken in a limiting sense, and the scope of embodiments is defined by the appended claims and their equivalents.

[0002] Various operations may be described as multiple discrete operations in turn, in a manner that may be helpful in understanding embodiments; however, the order of description should not be construed to imply that these operations are order dependent.

[0033] Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology can be found in Benjamin Lewin, *Genes* IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); and other similar references.

[0034] Suitable methods and materials for the practice or testing of this disclosure are described below. Such methods and materials are illustrative only and are not intended to be limiting. Other methods and materials similar or equivalent to those described herein can be used. For example, conventional methods well known in the art to which this disclosure pertains are described in various general and more specific references, including, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, 1989; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3d ed.,

Cold Spring Harbor Press, 2001; Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates, 1992 (and Supplements to 2000); Ausubel *et al.*, *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, 4th ed., Wiley & Sons, 1999; Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1990; and Harlow and Lane, *Using Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1999. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

[0035] Section headings as used in this section and the entire disclosure herein are merely for organizational purposes and are not intended to be limiting.

1. Definitions

[0036] To facilitate review of the various embodiments of the disclosure, the following explanations of specific terms are provided below. All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

[0037] The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

[0038] The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

[0039] As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one

element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of” “only one of” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

[0040] As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

[0041] In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

[0042] As used herein, “biological sample” generally refers to a biological specimen containing genomic DNA, RNA (such as mRNA), protein, or combinations thereof, obtained from a subject. Examples include, but are not limited to, saliva, peripheral blood, urine, tissue biopsy, surgical specimen, and autopsy material. In embodiments, the biological sample is a bodily fluid, such as blood, or a component thereof, such as plasma or serum.

[0043] As used herein, “biopolymer” generally refers to polymers (e.g., produced by living organisms or synthetically generated). Biopolymers contain monomeric units that are covalently bonded to form larger structures. There are three main classes of biopolymers,

classified according to the monomeric units used and the structure of the biopolymer formed: polynucleotides (RNA and DNA), which are long polymers composed of 13 or more nucleotide monomers; polypeptides, which are short polymers of amino acids; and polysaccharides, which are often linear bonded polymeric carbohydrate structures. Other examples of biopolymers include rubber, suberin, melanin and lignin.

[0044] As used herein, an “isolated” biological component (e.g., such as a nucleic acid molecule, protein, or cell) has been substantially separated or purified away from other biological components in the cell of the organism, or the organism itself, in which the component naturally occurs, such as other chromosomal and extra-chromosomal DNA and RNA, proteins and cells. Nucleic acid molecules and proteins that have been “isolated” may be understood to have been purified by standard purification methods. The term also embraces nucleic acid molecules and proteins prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acid molecules and proteins.

[0045] As used herein, “burst” generally refers to a section of a current stream in which the measured current changes between two levels more frequently relative to a pause, typically being in the high state for half the time of a pause. The measured current of the peaks in a burst is greater than about 20% of the baseline current passing through the molecule. Typically, a burst is observed when a nucleotide is being incorporated into a template sequence.

[0046] As used herein, “modification,” “chemical modification,” or “chemically modified” generally refers to any of a number of various processes involving the alteration of the chemical constituent or structure of molecules. For example, a polymerase can be chemically modified to form a chemical bond with a first electrode and a second electrode. In one example, a chemically-modified electrode is an electrode that has a surface chemically converted to change the electrode’s properties, such as its physical, chemical, electrochemical, optical, electrical, and/or transport characteristics. As provided herein, the chemical modification can also involve chemically altering a polymerase so that it is compatible with a linker that binds to an electrode (e.g., biotin/streptavidin, HaloTag, and the like). In other embodiments, a modification can be generated via protein synthesis. For example, a polymerase can be designed to comprise one or more modifications (e.g., a linker) when synthesized from a polynucleotide that encodes the protein and the modification (e.g. linker).

[0047] As used herein, “contact” and “contacting” can include placement in direct physical association, including both a solid and liquid form. “Contacting” can include a specific chemical contact between two different substances (e.g., covalent bond, or non-covalent bond having specific ligand interaction with specific amino acid residues).

[0048] As used herein, “complementarity” or “complementary” generally refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). “Perfectly complementary” generally indicates that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. “Substantially complementary” as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions.

[0049] As used herein, “current stream” generally refers to a current signal generated over time, such as from the bioelectronic devices of the present disclosure.

[0050] As used herein, a “label” generally refers to an agent capable of detection, for example by ELISA, spectrophotometry, flow cytometry, or microscopy. For example, a label can be attached to a nucleic acid molecule or protein (indirectly or directly), thereby permitting detection of the nucleic acid molecule or protein. Examples of labels include, but are not limited to, radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent agents, fluorophores, haptens, enzymes, and combinations thereof. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed for example in Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, New York, 1989) and Ausubel et al. (In Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1998).

[0051] As used herein, the term “linker” or “linked” means joined together, either directly or indirectly. For example, a first moiety may be covalently or noncovalently (e.g., electrostatically) linked to a second moiety. This includes, but is not limited to, covalently bonding one molecule to another molecule, noncovalently bonding one molecule to another (e.g., electrostatically bonding), non-covalently bonding one molecule to another molecule by hydrogen bonding, non-covalently bonding one molecule to another molecule by van der Waals forces, and any and all combinations of such couplings. Indirect attachment is possible, such as by using a “linker” (a molecule or group of atoms positioned between two moieties). In several embodiments, linked components are associated in a chemical or physical manner so

that the components are not freely dispersible from one another. For example, two components may be covalently bound to one another so that the two components are incapable of separately dispersing or diffusing. In several embodiments, linked components are associated in a chemical or physical manner so that the components are not freely dispersible from one another. For example, two components may be covalently bound to one another so that the two components are incapable of separately dispersing or diffusing.

[0052] As used herein, the terms “non-naturally occurring” and “engineered” interchangeably indicate the involvement of the hand of man. These terms, when referring to nucleic acid molecules or polypeptides, generally indicate that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

[0053] As used herein, “nucleic acid” generally refers to a deoxyribonucleotide or ribonucleotide polymer, which can include analogues of natural nucleotides that hybridize to nucleic acid molecules in a manner similar to naturally occurring nucleotides. In one example, a nucleic acid molecule is a single stranded (ss) DNA or RNA molecule, such as a probe or primer. In another example, a nucleic acid molecule is a double stranded (ds) nucleic acid. In another example, a nucleic acid is a modified DNA or RNA molecule, such as a xenonucleic acid (XNA). The term “nucleotide” generally refers to a base-sugar-phosphate combination and includes ribonucleoside triphosphates ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dTTP, dUTP, dGTP, dTTP, or derivatives thereof. As described further herein, a nucleic acid can include deoxynucleotidetriphosphate monomers (dNTPs), such as those having altered charges as compared to their corresponding standard deoxynucleotidetriphosphate monomers.

[0054] As used herein, “polypeptide,” “peptide,” and “protein” generally refer to a polymer in which the monomers are amino acid residues that are joined together through amide bonds. When the amino acids are alpha-amino acids, either the L-optical isomer or the D-optical isomer can be used, the L-isomers being preferred in nature. The term polypeptide is specifically intended to cover naturally occurring proteins, as well as those that are recombinantly or synthetically produced. A substantially purified polypeptide as used herein refers to a polypeptide that is substantially free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In one embodiment, the polypeptide is at least 50%, for example at least 80% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In another embodiment, the polypeptide is at least 90% free of other proteins, lipids, carbohydrates or other materials with which it is naturally

associated. In yet another embodiment, the polypeptide is at least 95% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated.

[0055] Conservative amino acid substitution tables providing functionally similar amino acids are well known to one of ordinary skill in the art. The following six groups are examples of amino acids that are considered to be conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0056] A non-conservative amino acid substitution can result from changes in: (a) the structure of the amino acid backbone in the area of the substitution; (b) the charge or hydrophobicity of the amino acid; or (c) the bulk of an amino acid side chain. Substitutions generally expected to produce the greatest changes in protein properties are those in which: (a) a hydrophilic residue is substituted for (or by) a hydrophobic residue; (b) a proline is substituted for (or by) any other residue; (c) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine; or (d) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl. Variant amino acid sequences may, for example, be 80, 90 or even 95 or 98% identical to the native amino acid sequence. Programs and algorithms for determining percentage identity can be found at the NCBI website.

[0057] As used herein, “probe” generally refers to a short sequence of nucleotides, such as at least 8, at least 10, at least 15, at least 20, or at least 21 nucleotides in length, which can be used to detect the presence of a complementary sequence by molecular hybridization. In particular examples, oligonucleotide probes include a label that permits detection of oligonucleotide probe:target sequence hybridization complexes. Laboratory standards and values can be set based on a known or determined population value and can be supplied in the format of a graph or table that permits comparison of measured, experimentally determined values.

[0058] As used herein, “primer” generally refers to a short nucleic acid molecule, for instance DNA oligonucleotides 10 -100 nucleotides in length, such as 5, 6, 7, 8, 9, 10, 11, 12, or more in length. Primers can be annealed to a complementary target nucleic acid strand by nucleic acid hybridization to form a hybrid between the primer and the target nucleic acid strand. Primers can be used for amplification of a nucleic acid sequence, such as by PCR or other nucleic acid amplification methods known in the art.

[0059] As used herein, the term “purified” does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the protein referred to is more pure than the protein in its natural environment within a cell. For example, a preparation of a protein is purified such that the protein represents at least 50% of the total protein content of the preparation. Similarly, a purified oligonucleotide preparation is one in which the oligonucleotide is purer than in an environment including a complex mixture of oligonucleotides. Purity of a compound may be determined, for example, by high performance liquid chromatography (HPLC) or other conventional methods.

[0060] As used herein, “recombinant” generally refers to recombinant nucleic acid or protein that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or by the artificial manipulation of isolated segments of nucleic acids, for example, by genetic engineering techniques. The term recombinant includes nucleic acids and proteins that have been altered solely by addition, substitution, or deletion of a portion of a natural nucleic acid molecule or protein.

[0061] As used herein, the term “subject” includes human and non-human animals. “Patient” and “subject” are used interchangeably herein.

[0062] As used herein, the terms, “substantial identity” or “substantially identical” generally refer to a nucleic acid or fragment thereof, that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), refers to a nucleotide sequence having at least about 95% sequence identity, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed below. A nucleic acid molecule having substantial identity to a reference nucleic acid molecule may, in certain instances, encode a polypeptide having the same or substantially similar amino acid sequence as the polypeptide encoded by the reference nucleic acid molecule.

[0063] As applied to polypeptides, the term “substantial similarity” or “substantially similar” means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 95% sequence identity, even more preferably at least 98% or 99% sequence identity. Preferably, residue positions which are not identical differ by conservative amino acid substitutions. A “conservative amino acid substitution” is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other

by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson (1994) *Methods Mol. Biol.* 24: 307-331, herein incorporated by reference. Examples of groups of amino acids that have side chains with similar chemical properties include (1) aliphatic side chains: glycine, alanine, valine, leucine and isoleucine; (2) aliphatic-hydroxyl side chains: serine and threonine; (3) amide-containing side chains: asparagine and glutamine; (4) aromatic side chains: phenylalanine, tyrosine, and tryptophan; (5) basic side chains: lysine, arginine, and histidine; (6) acidic side chains: aspartate and glutamate, and (7) sulfur-containing side chains are cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, glutamate-aspartate, and asparagine-glutamine. Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.* (1992) *Science* 256: 1443-1445, herein incorporated by reference. A “moderately conservative” replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

[0064] Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG software contains programs such as Gap and Bestfit which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutant thereof. *See, e.g.*, GCG Version 6.1. Polypeptide sequences also can be compared using FASTA using default or recommended parameters, a program in GCG Version 6.1. FASTA (e.g., FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (2000) *supra*). Another preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially BLASTP or TBLASTN, using default parameters. *See, e.g.*, Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410 and Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-402, each herein incorporated by reference.

[0065] As used herein, “variant” generally refers to a peptide or polypeptide that differs in amino acid sequence by the insertion, deletion, or conservative substitution of amino acids, but retain at least one biological activity. “SNP” refers to a variant that is a single nucleotide

polymorphism. Representative examples of “biological activity” include the ability to be bound by a specific antibody or to promote an immune response. Variant is also used herein to describe a protein with an amino acid sequence that is substantially identical to a referenced protein with an amino acid sequence that retains at least one biological activity. A conservative substitution of an amino acid, *i.e.* replacing an amino acid with a different amino acid of similar properties (*e.g.*, hydrophilicity, degree, and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes can be identified, in part, by considering the hydropathic index of amino acids, as understood in the art. Kyte *et al.*, *J. Mol. Biol.* 157:105-132 (1982). The hydropathic index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydropathic indexes can be substituted and still retain protein function. In one aspect, amino acids having hydropathic indexes of ± 2 are substituted. The hydrophilicity of amino acids can also be used to reveal substitutions that would result in proteins retaining biological function. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide, a useful measure that has been reported to correlate well with antigenicity and immunogenicity. Substitution of amino acids having similar hydrophilicity values can result in peptides retaining biological activity, for example immunogenicity, as is understood in the art. Substitutions may be performed with amino acids having hydrophilicity values within ± 2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties.

[0066] As used herein, “pause” generally refers to a section of a current stream in which the fluctuations in measured current are interrupted by a slower feature of about twice the duration of the neighboring features. Typically, a pause is observed before and after a nucleotide has been incorporated into a template sequence, and the duration of the pause relative to the neighboring pulses of current increases as the concentration of nucleotide triphosphates is lowered.

[0067] As used herein, a “polymerase” generally refers to an enzyme that synthesizes long chains of polymers or nucleic acids. DNA polymerase and RNA polymerase are used to assemble DNA and RNA molecules, respectively, by copying a DNA template strand using base-pairing interactions or RNA by half ladder replication.

[0068] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present disclosure. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[0069] As noted herein, the disclosed embodiments have been presented for illustrative purposes only and are not limiting. Other embodiments are possible and are covered by the disclosure, which will be apparent from the teachings contained herein. Thus, the breadth and scope of the disclosure should not be limited by any of the above-described embodiments but should be defined only in accordance with claims supported by the present disclosure and their equivalents. Moreover, embodiments of the subject disclosure may include methods, compositions, systems and apparatuses/devices which may further include any and all elements from any other disclosed methods, compositions, systems, and devices, including any and all elements corresponding to detecting protein activity. In other words, elements from one or another disclosed embodiments may be interchangeable with elements from other disclosed embodiments. Moreover, some further embodiments may be realized by combining one and/or another feature disclosed herein with methods, compositions, systems and devices, and one or more features thereof, disclosed in materials incorporated by reference. In addition, one or more features/elements of disclosed embodiments may be removed and still result in patentable subject matter (and thus, resulting in yet more embodiments of the subject disclosure). Furthermore, some embodiments correspond to methods, compositions, systems, and devices which specifically lack one and/or another element, structure, and/or steps (as applicable), as compared to teachings of the prior art, and therefore represent patentable subject matter and are distinguishable therefrom (i.e. claims directed to such embodiments may contain negative limitations to note the lack of one or more features prior art teachings).

2. Methods for Sequencing a Biopolymer

[0070] Embodiments of the present disclosure include devices, systems, and methods related to sequencing a biopolymer. In particular, the present disclosure relates to methods for sequencing a polynucleotide using a bioelectronic device that obtains a bioelectronic signature (e.g., current amplitude levels) of polymerase activity based on current fluctuations as

complementary nucleotidepolyphosphate monomers (e.g., having distinctive charges) are incorporated into the template polynucleotide.

[0071] As described further herein, the devices, systems, and methods of the present disclosure can be used to generate a bioelectronic signature of an enzyme-of-interest, which can be used to determine the sequence of any biopolymer (e.g., polynucleotide). In some embodiments, the enzyme-of-interest can be a polymerase, and various aspects of a bioelectronic signature of a polymerase as it adds nucleotide monomers to a template polynucleotide strand can be used to determine the sequence of that template polynucleotide. For example, a bioelectronic signature of polymerase activity can be based on current fluctuations as each complementary nucleotide monomer is incorporated into the template polynucleotide. In some embodiments, a bioelectronic signature is generated based on current amplitude levels corresponding to polymerase activity as complementary nucleotidepolyphosphate monomers having distinctive charges are incorporated into a template polynucleotide. In some embodiments, the bioelectronic device used to generate a bioelectronic signature comprises a polymerase functionally coupled to both a first electrode and a second electrode. The term “nucleotide” generally refers to a base-sugar-phosphate combination and includes ribonucleoside triphosphates ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof.

[0072] In some embodiments, methods for sequencing a biopolymer include identifying a nucleotide base incorporated into a template strand of a nucleic acid (e.g., DNA) using a bioelectronic device. In accordance with these embodiments, the method includes introducing a template polynucleotide to the bioelectronic device. In some embodiments, the bioelectronic device comprises a polymerase functionally coupled to at least one of a first electrode and a second electrode. In some embodiments, the method further includes introducing a solution comprising four nucleotidepolyphosphate monomers (e.g., dNxPs) to the device comprising the template polynucleotide. In some embodiments, at least three of the four nucleotidepolyphosphate monomers present in the solution comprises a distinctive charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, the method further includes obtaining a bioelectronic signature of polymerase activity based on current fluctuations as each complementary nucleotidepolyphosphate monomer is incorporated into the template polynucleotide. In some embodiments, at least one characteristic of the bioelectronic signature identifies each of the complementary nucleotidepolyphosphate monomers incorporated into the template polynucleotide.

[0073] As one of ordinary skill in the art will readily recognize and appreciate after having benefited from the teachings of the present disclosure, the methods described herein can be used with any bioelectronic device that senses the duration of the open and closed states of an enzyme (e.g., polymerase). Exemplary devices include, but are not limited to, the bioelectronic devices and systems disclosed in U.S. Patent No. 10,422,787 and PCT Appln. No. PCT/US2019/032707, both of which are herein incorporated by reference in their entirety and for all purposes. Additionally, it will be readily recognized and appreciated by those of ordinary skill in the art based on the present disclosure that the forgoing embodiments apply equally to (and include) sequencing RNAs with the substitution of rNTPs for dNTPs and the use of an RNA polymerase.

[0074] Further, one of ordinary skill in the art would readily recognize and appreciate that the methods described herein can be used in conjunction with other methods involving the sequencing of a biopolymer. In particular, the various embodiments disclosed in PCT Application No. PCT/US21/19428, which is herein incorporated by reference in its entirety, describes the interpretation of current fluctuations generated by a DNA polymerase as it actively extends a template, and how signal features (e.g., bioelectronic signature) may be interpreted in terms of the nucleotide being incorporated, and thus, how these signals can read the sequence of the template. This approach utilizes features of the signal that vary in time. For example, the time that the polymerase stays in a low current state reflects the concentration of the nucleotidetriphosphate in solution. If the concentration of a particular nucleotide triphosphate is low, then the polymerase must stay open for a longer time in order to capture the correct nucleotide, and since the open conformation of the polymerase corresponds to a lower current, the dip in current associated with the open state lasts for longer. Additionally, the various embodiments disclosed in PCT Application No. PCT/US20/38740, which is herein incorporated by reference in its entirety, describes how the base-stacking polymerization rate constant differences are reflected in the closed-state (high current states) so that the duration of these states may also be used as an indication of which one of the four nucleotides is being incorporated. It can be desirable to be able to use the amplitude of the signal as yet an additional contribution to determining sequence. Further, the various embodiments disclosed in PCT Application No. PCT/US21/17583, which is herein incorporated by reference in its entirety, describes methods that utilize a defined electrical potential to maximize electrical conductance of a protein-of-interest (e.g., polymerase), which can serve as a basis for the fabrication of enhanced bioelectronic devices for the direct measurement of protein activity.

[0075] In accordance with the above embodiments, the present disclosure includes methods of sequencing using nucleotidepolyphosphates (e.g., dN_xPs) having distinctive charges (e.g., more positive or more negative relative to their corresponding standard nucleotidetriphosphate monomer) to generate a bioelectronic signature that can be used to identify each of the complementary nucleotidepolyphosphate monomers (see, e.g., FIG. 1) incorporated into a template polynucleotide. In some embodiments, the bioelectronic signature comprises a closed period corresponding to the polymerase being in a closed state (e.g., nucleotide incorporation). In some embodiments, at least one characteristic of the bioelectronic signature comprises current amplitude level, such as a change (e.g., increase or decrease) in the current amplitude level based on the incorporation of a particular nucleotidepolyphosphate monomer having a distinctive charge relative to its corresponding standard nucleotidetriphosphate monomer (FIGS. 3A-3B). As described further herein, the electrical conductance of a polymerase varies depending on local charge (expressed in terms of surface potential; see FIG. 2), which can be altered by nucleotidepolyphosphate monomers having distinctive charges, and which provide a basis for identifying which monomer has been incorporated into a particular template nucleic acid sufficient to determine the sequence of that template nucleic acid.

[0076] In some embodiments, at least one of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) has a distinctive negative charge relative to its corresponding standard nucleotidetriphosphate monomer. In some embodiments, at least two of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleic acid) have distinctive negative charges relative to their corresponding standard nucleotidetriphosphate monomer. In some embodiments, at least three of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) have distinctive negative charges relative to their corresponding standard nucleotidetriphosphate monomer. In some embodiments, all four of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) have distinctive negative charges relative to their corresponding standard nucleotidetriphosphate monomer.

[0077] In some embodiments, the negative charge is conferred by at least one additional phosphate group relative to its corresponding standard nucleotidetriphosphate monomer (FIG. 1). In some embodiments, the negative charge is conferred by 1 to 20 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 19 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 18 additional phosphate groups.

groups. In some embodiments, the negative charge is conferred by 1 to 17 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 16 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 15 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 14 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 13 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 12 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 11 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 10 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 9 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 8 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 7 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 6 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 5 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 4 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 to 3 additional phosphate groups. In some embodiments, the negative charge is conferred by 1 or 2 additional phosphate groups.

[0078] As would be recognized by one of ordinary skill in the art based on the present disclosure, the additional of other chemical moieties can also be used to alter the charge of the nucleotidepolyphosphate monomers (e.g., to be more negative than their corresponding standard nucleotidepolyphosphate monomers), provided they do not interfere with the function of the polymerase. Additionally, as disclosed in Dellafiore, M.A. et al., *Front. Chem.*, 04 May 2016 (<https://doi.org/10.3389/fchem.2016.00018>) and by Duffy, K. et al., *BMC Biology*, 02 September 2020 (<https://doi.org/10.1186/s12915-020-00803-6>) and McKenzie, L.K. et al., *Chem. Soc. Rev.*, 01 Mar 2021 (<https://doi.org/10.1039/D0CS01430C>) any modified nucleic acids (e.g. modified nucleobases, sugars, or phosphates) can be obtained according to various strategies. A phosphate charge may be eliminated with triazole linkages as demonstrated by Fujino, T. et al, *J. Org. Chem.* 31 August 2016 (<https://doi.org/10.1021/acs.joc.6b01618>) and a positively charged uridine may be incorporated as demonstrated by Vaish, N.K. et al., *Biochemistry*, 04 July 2003 (<https://doi.org/10.1021/bi027354i>).

[0079] The design strategies of various modified nucleotides suitable for sequencing methodologies generally require that the modified nucleotides not disturb the base pair interactions (Watson-Crick and Hoogsteen); that the modified nucleotides be substrates of the corresponding DNA or RNA polymerases; the introduction of the modified nucleotide be

efficient at any position of the sequence; and that the modified sequence be a template for the corresponding polymerases.

[0080] In some embodiments, a distinctive negative charge on the at least one nucleotidepolyphosphate monomer corresponds to an increased or decreased current amplitude level relative to the current amplitude level of its corresponding standard nucleotidetriphosphate monomer. For example, as shown in FIG. 2, depending on the peak conductance level, additional negative charges on a nucleotidepolyphosphate monomer (e.g., reduced potential) can correspond to an increase in current amplitude level in a bioelectronic signature (e.g., right side of **201** in FIG. 2), or can correspond to a decrease in current amplitude level in a bioelectronic signature (e.g., left side of **201** in FIG. 2). Thus, in some embodiments, a distinctive negative charge on at least one deoxynucleotidepolyphosphate monomer corresponds to an increased current amplitude level relative to the current amplitude level of its corresponding standard deoxynucleotidetriphosphate monomer. In other embodiments, a distinctive negative charge on at least one deoxynucleotidepolyphosphate monomer corresponds to a decreased current amplitude level relative to the current amplitude level of its corresponding standard deoxynucleotidetriphosphate monomer.

[0081] Similarly, in some embodiments, at least one of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) has a distinctive positive charge relative to its corresponding standard nucleotidepolyphosphate monomer. In some embodiments, at least two of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) have distinctive positive charges relative to their corresponding standard nucleotidepolyphosphate monomers. In some embodiments, at least three of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) have distinctive positive charges relative to their corresponding standard nucleotidepolyphosphate monomers. In some embodiments, all four of the nucleotidepolyphosphate monomers present in a solution (e.g., a solution comprising a bioelectronic device and a template nucleotide) have distinctive positive charges relative to their corresponding standard nucleotidepolyphosphate monomers. As would be recognized by one of ordinary skill in the art based on the present disclosure, the addition of any chemical moiety can be used to alter the charge of the nucleotidepolyphosphate monomers (e.g., to be more positive than its corresponding standard nucleotidepolyphosphate monomer), provided it does not interfere with the function of the polymerase.

[0082] In some embodiments, a distinctive positive charge on the at least one nucleotidetriphosphate monomer corresponds to an increased or decreased current amplitude level relative to the current amplitude level of its corresponding standard nucleotidetriphosphate monomer. For example, as shown in FIG. 2, depending on the conductance peak value, additional positive charges on a nucleotidetriphosphate monomer (e.g., increased potential) can correspond to a decrease in current amplitude level in a bioelectronic signature (e.g., right side of **201** in FIG. 2), or can correspond to an increase in current amplitude level in a bioelectronic signature (e.g., left side of **201** in FIG. 2). Thus, in some embodiments, a distinctive positive charge on at least one nucleotidetriphosphate monomer corresponds to an decreased current amplitude level relative to the current amplitude level of its corresponding standard nucleotidetriphosphate monomer. In other embodiments, a distinctive positive charge on at least one nucleotidetriphosphate monomer corresponds to an increased current amplitude level relative to the current amplitude level of its corresponding standard nucleotidetriphosphate monomer.

[0083] In accordance with these embodiments, examples of modified nucleotides, each carrying a different charge, include but are not limited to, deoxynucleotidetriphosphates (dNTPs). The structure of four exemplary dNTPs are provided in FIG. 1. In this example, the chemical structure of deoxyadenosinetriphosphate **101**, the normal triphosphate form is shown. Here, “x” refers to the number of phosphates attached to the nucleotides; for **101**, x=3. Reference numbers **102**, **103** and **104** correspond to deoxynucleotides with additional phosphate groups. In this example, **102** corresponds to a deoxythymidine tetraphosphate (x=4) carrying one additional electronic charge relative to **101**; **103** corresponds to a deoxycytidine hexylphosphate (x=6) carrying three additional charges relative to **101**; and **104** corresponds to a deoxyguanosine nonylphosphate (x=9) carrying six additional charges relative to **101**. These structures provided in FIG. 1 are exemplary illustrations only. As would be recognized by one of ordinary skill in the art based on the present disclosure, there are many other ways of modifying the charge on the triphosphate of a normal nucleotide, including but not limited to, adding amine groups to induce a positive charge or sulfates or carboxylates to produce a negative charge at or near pH 7.0. Further variations are also possible, for example, by altering the pH so as to ionize moieties attached to the deoxynucleotides.

[0084] An exemplary mechanism of modulation of signal amplitude is illustrated in FIG. 2. This example demonstrates the conductance in nanoSiemens (nS) of a Φ 29 polymerase as a function of its potential in mV on the Normal Hydrogen Electrode (NHE) scale. The solid line is a fit to a Lorentzian function:

$$\text{[0085]} \quad G = A \frac{\Gamma^2}{(E_0 - E)^2 + \Gamma^2}$$

[0086] where E is the potential, $E_0 = 259$ mV and $2\Gamma = 183$ mV. Away from the peak **201** or the tails **202** of the Lorentzian, the conductance is strongly dependent on potential. The region of maximum dependence of conductance on potential, $\left(\frac{dG}{dE}\right)_{MAX}$, **203** on the positive potential side of the peak occurs at

$$\text{[0087]} \quad E_1 = E_0 + \frac{1}{\sqrt{3}}\Gamma$$

[0088] or about 320 mV, as marked **204**. In this region, the slope, $\left(\frac{dG}{dE}\right)_{MAX} \approx 0.05$ nS/mV.

The slope is similar in the region **205**, which marks the potential when the polymerase is contacted by a Pd electrode and an Au electrode, and the region **206** which marks the potential when the polymerase is contacted by a Pt electrode and an Au electrode.

[0089] The change in potential ΔV of a polymerase binding a charged molecule can be estimated from the Coulomb charging energy of a sphere of radius a and dielectric constant ϵ . Taking the charge to be N units of e , the charge on an electron,

$$\text{[0090]} \quad \Delta V = \frac{Ne}{8\pi\epsilon\epsilon_0 a}$$

[0091] where ϵ_0 is the permittivity of free space, 8.85×10^{-12} F/m. The phosphate chain will be bound at the catalytic site of the polymerase where the relative permittivity, ϵ , (also known as the dielectric constant) will be approximately 4. Taking the radius of a $\Phi 29$ polymerase to be 3.5 nm, the potential shift per unit added electronic charge is ~ 50 mV/e. The corresponding shift in conductance, ΔG is

$$\text{[0092]} \quad \Delta G = \Delta V \times \left(\frac{dG}{dE}\right)_{MAX}$$

[0093] leading to a conductance change of about 2.5 nS. This is a readily measurable change on going from a triphosphate (as in **101** in FIG. 1) to a tetrakisphosphate (as in **102** in FIG. 1). The shifts for larger numbers of phosphates will likely be smaller than given by this equation using the value of four for the relative dielectric constant. For example, a long polyphosphate such as **104** would protrude out into the surrounding solvent where the dielectric constant would approach 80, reducing the potential change by a factor $80/4=20$. In this case, the potential shift per added phosphate would be reduced to 50mV/20 or 2.5 mV, so that the last phosphates in the chain would contribute only 0.135 nS, a much smaller, but still a measurable amount.

[0094] In some embodiments, a bioelectronic device of the present disclosure includes a polymerase operated with a gold and a platinum electrode, poised at the potential **206**. When

the polymerase captures a dN_xP with $x > 3$, the negative charge on the polymerase increases relative to the case where it captures a regular dNTP, shifting its potential to lower values; and therefore, increasing the conductance during the closed conformation of the polymerase (during which the dN_xP complementary to the template is bound). Thus, as long as the shifted potential at the most extreme values lay in a range **207** on one side of the peak **201**, the effect of added negative charge distinguishes each nucleotide by an increased conductance in the order deoxyguanosine nonylphosphate ($x=9$) > deoxycytidine hexylphosphate ($x=6$) > deoxythymidine tetraphosphate ($x=4$) > deoxyadenosine triphosphate ($x=3$), using the examples of dN_xPs shown in FIG. 1.

[0095] Additionally, FIG. 3 illustrates a signal **301** taken with all four deoxynucleotide triphosphates present. Here, the chemical identity of nucleotides is only evident in the duration of the open **302** or closed **303** states, as disclosed in PCT Application No. PCT/US21/19428. However, when each dN_xP carries a distinctive charge, the current level in the closed state signals its identity as illustrated in **304**. In the example just given, highest conductance state signals incorporation of a “g” base **305**, the next level, **306**, incorporation of a “c” base, the next, **307**, a “t” base and the lowest, **308**, an “a” base.

3. Bioelectronic Devices and Systems

[0096] In accordance with the above methods, the bioelectronic devices and systems of the present disclosure generally include a first electrode and a second electrode that are configured for contact with a sample (e.g., biopolymer sample) to be analyzed. In some embodiments, such as when the electrodes are planar, the first and/or second electrode do not require a dielectric layer. In other embodiments, the first and/or second electrode can have a dielectric layer. In some embodiments, the first and/or second electrode comprise a metal selected from gold, silver, copper, platinum, palladium, and ruthenium (or any alloys thereof). In some embodiments, the metal is palladium. In some embodiments, the methods of the present disclosure include applying a voltage bias between the first and second electrodes that is 100mV or less. It will be recognized by one of ordinary skill in the art based on the present disclosure that the bioelectronic devices and systems of the present disclosure, according to the various methods described herein, can be used to sequence DNA and RNA polymers by similar techniques (e.g., sequence RNA polymers using an RNA dependent RNA polymerase and four different ribonucleotides each carrying a distinctive charge).

[0097] In some embodiments, the gap between the first and second electrode is a width of about 1.0 nm to about 50.0 nm. In some embodiments, the gap between the first and second

electrode is a width of about 1.0 nm to about 40.0 nm. In some embodiments, the gap between the first and second electrode is a width of about 1.0 nm to about 30.0 nm. In some embodiments, the gap between the first and second electrode is a width of about 1.0 nm to about 20.0 nm. In some embodiments, the gap has a width of about 1.0 nm to about 10.0 nm. In some embodiments, the gap has a width of about 1.0 nm to about 7.5 nm. In some embodiments, the gap has a width of about 1.0 nm to about 5.0 nm. In some embodiments, the gap has a width of about 4.0 nm to about 5.0 nm.

[0098] In some embodiments, the polymerase can be attached to one electrode in one embodiment and to both electrodes in another embodiment. The polymerase can be attached to the electrode(s) either directly or indirectly. In some embodiments, the polymerase is attached to the electrode(s) via a linker. In some embodiments, the polymerase is attached to the electrode indirectly via interactions with a ligand attached to the electrode. In some embodiments, the polymerase is modified to incorporate a ligand-binding site. In some embodiments, the polymerase is a biotinylated polymerase. In some embodiments, the polymerase comprises an Avitag. In some embodiments, the polymerase is a biotinylated polymerase and is attached to the electrode via streptavidin. In some embodiments, the polymerase is modified to incorporate an amino acid residue that allows for click-chemistry attachment of other chemical groups to the electrodes (e.g., 4-Azido-L-phenylalanine.). In some embodiments, the exonuclease activity of the polymerase is disabled. In some embodiments, linker is attached to a region of the polymerase that is inactive.

[0099] When the polymerase is attached to both electrodes, the distance between the two attachment points is at least about 1 nm to about the overall size of the polymerase. In one embodiment, the distance is from about 1 nm to about 10 nm. In another embodiment, the distance is from about 3 nm to about 7 nm. In another embodiment, the distance is from about 5 nm to about 6 nm. In some embodiments, the distance is from about 2 to about 8 nm.

[0100] When the polymerase is attached to both electrodes, the two attachment points must not move relative to each other, when the polymerase undergoes open-to-closed conformational changes. The crystal structures of many polymerases are currently available (see, e.g., [//www.rcsb.org/](http://www.rcsb.org/)) in both open and closed forms. Thus, when choosing the two attachment points, the two residues must be separated from each other by distances that are similar to the gap between the electrodes used to contact the polymerase, such as between about 1 nm and about 10 nm. For example, in some embodiments, this distance is between 2 nm and 8 nm. In embodiments, the two attachment points have the same atomic coordinates in both the open and closed forms, to within half a nanometer.

[0101] In some embodiments, the polymerase incorporates an inserted flexible sequence. Any peptide can be used as the flexible sequence as long as (a) it does not form an alpha helix or a beta sheet and (b) the residues in the sequence do not substantially change the pI relative to the pI of the unmodified polymerase. It is contemplated that additional polymerases may be used so long as the modified polymerase functions efficiently which can be determined by a rolling-circle amplification assay.

[0102] In some embodiments, the device further comprises a nucleic acid template. The nucleic acid template can be a DNA template in one embodiment and an RNA template in another embodiment. For the polymerase to operate at maximum speed, and for the electrical signals to be readily processed, it is understood that the polymerase not be stalled or obstructed by secondary structures formed by the DNA template, such as when the template contains single stranded regions.

[0103] As described further herein, the various features or characteristics of a bioelectronic signature of an active protein-of-interest can be used to determine the sequence of a biopolymer. As would be recognized by one of skill in the art based on the present disclosure, the methods of obtaining a bioelectronic signature and extracting various characteristics described herein can be used to determine the sequence of any biopolymer using any corresponding enzyme-of-interest, including but not limited to a polymerase, a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease. In accordance with these embodiments, the present disclosure also provides bioelectronic devices and systems in which the charge on a protein-of-interest is altered in order to modulate the overall conductance of a protein complex (e.g., a protein-of-interest and corresponding linker). For example, the conductance of a given protein complex in a bioelectronic device configured with platinum electrodes can be modulated (e.g., increased) to be similar to the conductance provided by gold electrodes by altering the charge of the protein complex (e.g., shifting the potential). Thus, embodiments of the present disclosure include a bioelectronic device that includes a first electrode and a second electrode separated by a gap, and a protein (e.g., a polymerase) attached to the first and second electrodes via a linker comprising a distinctive electrical charge. In some embodiments, the distinctive electrical charge modulates conductance through the protein in a manner that enhances the function of the bioelectronic device (e.g., sequencing of a biopolymer).

[0104] In some embodiments, the linker used to generate a bioelectronic device of the present disclosure includes a peptide or polypeptide. In some embodiments, the linker comprises streptavidin. In some embodiments, the linker comprises streptavidin that has been

modified to have a positive or negative charge (e.g., with polyglutamate). In some embodiments, the protein-of-interest is biotinylated. In some embodiments, the linker comprises a distinctive negative charge. In some embodiments, the distinctive negative charge is conferred by addition of a glutamate moiety, an aspartate moiety, or a combination thereof (including derivatives, variants, and polymers), coupled to the streptavidin. In some embodiments, the distinctive negative charge increases the conductance through the protein. In some embodiments, the linker comprises a distinctive positive charge. In some embodiments, the distinctive positive charge is conferred by addition of an arginine moiety, a histidine moiety, a lysine moiety, or a combination thereof (including derivatives, variants, and polymers), coupled to the streptavidin. In some embodiments, the distinctive positive charge increases or decreases the conductance through the protein.

[0105] As described further herein, the influence of charge on electronic transmission was demonstrated by comparing conductance obtained with a wild-type streptavidin with that of a streptavidin molecule containing an additional hexaglutamate (six negative charges) tail at the C terminus of each of the four monomers that comprise the streptavidin tetramer. FIG. 4 includes a representative diagram of the backbone structure of a hexaglutamate mutant streptavidin showing the charged tails (E6) attached to the C termini of each of the streptavidin monomers labeled (1, 2, 3, 4) next to the point where the biotin ligand emerges from the protein monomers. The four hexaglutamate tails **401**, **402**, **403**, and **404** are labeled E6.

[0106] FIGS. 5A-5B include representative conductance distributions for wild-type (FIG. 5A) and the hexaglutamate mutant streptavidin (FIG. 5B) measured at a 2.5 nm gap using Pd electrodes. The highest conductance peaks in each case (**501**, **502**) correspond to the desired maximum bonding between the electrodes. These measurements were made with Pd electrodes. For the wild-type, the conductance of the highest peak is 6.8 nS. For the E6 mutant, it is 21.38 nS. Thus, the conductance has increased by 14.58 nS with the addition of 24 electronic charges. This is about 0.6 nS per electronic charge, less than the maximum estimated above, but large enough to be readily measured.

[0107] Charge modulation of conductance confers several advantages, including but not limited to, increasing the conductance of a complex using electrodes whose rest potentials place them far from the peak of the conductance vs. potential curve (see, e.g., FIG. 2). Peak conductances can be obtained with gold or gold-palladium combinations of electrodes, but gold has various limitations as an electrode material in many semiconductor processes. Therefore, it is advantageous to use other metals, such as Pt or Pd or other noble metals, but doing so can

lead to a reduced signal. Thus, as provided herein, the signal can be restored or enhanced by altering the charge of the protein complex (e.g., using a modified streptavidin as a linker).

[0108] FIGS. 6A-6B includes representative diagrams of streptavidin-linked Φ 29 polymerases bridging electrode pairs using wildtype streptavidin (FIG. 6A) and the hexaglutamate mutant streptavidin (FIG. 6B). The corresponding conductance distributions measured at a 4.5 nm gap with Pd electrodes are provided in the graphs below each diagram. The graphs in FIGS. 6A and 6B include conductance distributions for a doubly-biotinylated Φ 29 **503** polymerase coupled to biotinylated electrodes **501** by wildtype streptavidin **502**. The peak conductance **506** is 6.61 nS for the Pd electrodes used in this example. FIG. 6B shows the conductance distribution measured for the same complex on Pd electrodes, but with the hexaglutamate mutant streptavidin used as connectors **504**. The conductance peak **507** is increased to 11.22 nS. While this effect is demonstrated here for streptavidin molecules, it readily can be applied to any other type of protein molecule used as a connector, as would be recognized by one of ordinary skill in the art based on the present disclosure.

[0109] In accordance with the above embodiments, a polymerase can be functionally coupled to a first and second electrodes using a linker comprising streptavidin. In some embodiments, the polymerase is biotinylated. In some embodiments, the linker is attached to a region of the polymerase that is inactive. In some embodiments, the polymerase and the first and second electrodes are biotinylated, and the linker comprises a streptavidin molecule comprising at least two biotin binding sites. In some embodiments, the exonuclease activity of the polymerase is disabled. In some embodiments, the gap has a width of about 1.0 nm to about 20.0 nm. In some embodiments, the first and second electrodes are separated by a dielectric layer. In some embodiments, the method comprises applying a voltage bias between the first and second electrodes that is 100mV or less.

[0110] Embodiments of the present disclosure also include a system for direct electrical measurement of polymerase activity. In accordance with these embodiments, the system includes any of the bioelectronic devices described herein, a means for introducing dNTPs capable of interacting with the polymerase, a means for applying a voltage bias between the first and second electrodes that is 100mV or less, and a means for monitoring fluctuations that occur as the dNTPs are incorporated into a template polynucleotide by the polymerase.

[0111] As persons of ordinary skill in the art will readily recognize and appreciate after having benefited from the teachings of the present disclosure, the methods described herein can be used with any bioelectronic device that senses the duration of the open and closed states of an enzyme (e.g., polymerase). Exemplary devices include, but are not limited to, the

bioelectronic devices and systems disclosed in U.S. Patent No. 10,422,787 and PCT Appln. No. PCT/US2019/032707, both of which are herein incorporated by reference in their entirety and for all purposes.