

DIRECT ELECTRICAL READOUT OF NUCLEIC ACID SEQUENCES**CROSS REFERENCE TO RELATED APPLICATION**

[0001] This disclosure claims the priority benefit of the earlier filing date of U.S. Provisional Patent Application No. 62/864,174 filed on June 20, 2019, which is hereby incorporated by reference in its entirety.

ACKNOWLEDGMENT OF GOVERNMENT SUPPORT CLAUSE

[0002] This invention was made with government support under R21 HG010522 and R01 HG009180 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO A SEQUENCE LISTING

[0003] This application incorporates by reference the Sequence Listing submitted in Computer Readable Form created on June 19, 2020 and containing 6 kilobytes.

FIELD

[0004] This disclosure relates to sequencing, and in particular to devices, systems and methods for direct measurement of polymerase activity.

SUMMARY

[0005] Disclosed herein are devices, systems and methods for direct measurement of polymerase activity.

[0006] In some embodiments, a device is provided, the device comprising: at least a first electrode and a second electrode, the first and second electrode being separated by a gap; and a polymerase comprising two sites for attachment to the first and second electrodes, wherein the two attachment sites are separated by a distance of at least about 1 nm and the distance does not significantly change (that is a change in overall distance of less than 10%) with conformational changes of the polymerase, such as the two attachment sites have the same atomic coordinates when the polymerase is in an open or closed conformation.

[0007] In some embodiments, a system is provided, the system comprising a device as described herein and configured to practice one or more methods disclosed herein.

[0008] In some embodiments, a method of detecting incorporation of a nucleotide is provided, the method comprising: (a) introducing a solution comprising a nucleic acid

template to a device as described herein; (b) measuring a first current generated when a bias is applied to the device; (c) introducing a solution comprising a dNTP to the device under conditions that allow for incorporation of the dNTP complementary to the nucleic acid template; (d) measuring a second current generated in step (c); (e) removing the solution comprising the unincorporated dNTP; (f) repeating steps (c) to (e); (g) obtaining a signal stream from the current generated in step (b) and step (d); (h) identifying a section of current signal that fluctuates between two levels of current for an amount of time greater than about 40 ms; and (i) detecting that the nucleotide has been incorporated if the signal stream comprises a pause.

[0009] In some embodiments, a method of determining the number of nucleic acids repeated in a nucleic acid template sequence is provided, the method comprising: (a) introducing a solution comprising a nucleic acid template to a device as described herein; (b) measuring a first current generated when a bias is applied to the device; (c) introducing a solution comprising a dNTP to the device under conditions that allow for incorporation of the dNTP complementary to the nucleic acid template; (d) measuring a second current generated in step (c); (e) removing the solution comprising the unincorporated dNTP; (f) repeating steps (c) to (e); (g) obtaining a signal stream from the current generated in step (b) and step (d), wherein the signal stream comprises at least one pause and at least one burst; (h) identifying sections of current signal that fluctuate between two levels of current an amount of time of greater than about 40 ms; and (i) determining that the nucleotide has been repeated in the nucleic acid template sequence if the signal stream comprises a pause that is about twice the length of a plurality of pauses in the signals stream.

[0010] In some embodiments, a method of determining the identity of a nucleotide being incorporated into a nucleic acid template sequence is provided, the method comprising: (a) introducing a solution comprising a nucleic acid template to a device as described herein; (b) measuring a first current generated when a bias is applied to the device; (c) introducing a solution comprising a dNTP to the device under conditions that allow for incorporation of the dNTP complementary to the nucleic acid template; (d) measuring a second current generated in step (c); (e) removing the solution comprising the unincorporated dNTP; (f) repeating steps (c) to (e); (g) obtaining a signal stream from the current generated in step (b) and step (d), wherein the signal stream comprises at least one pause and at least

one burst; and (h) determining the identity of the incorporated nucleotide from the burst signal.

[0011] In some embodiments, a method of sequencing a nucleic acid sequence is provided, the method comprising: (a) introducing a solution comprising a nucleic acid template to a device as described herein, wherein the nucleic acid template comprises all possible 16 combinations of two sequential bases in the sequence; (b) introducing a solution comprising a dNTP to the device under conditions that allow for incorporation of the dNTP complementary to the nucleic acid template; (c) obtaining a signal stream from repeat sequences in the nucleic acid template, wherein the signal stream comprises at least one pause and at least one burst; (d) using a signal analysis to characterize the burst signal of each incorporated dNTP; and (e) applying the signal analysis for sequencing an unknown nucleic acid sequence.

[0012] The foregoing and other features of the disclosure will become more apparent from the following detailed description, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF DRAWINGS

[0013] Embodiments will be readily understood by the following detailed description in conjunction with the accompanying drawings and the appended claims. Embodiments are illustrated by way of example and not by way of limitation in the figures of the accompanying drawings.

[0014] **FIGS. 1Ai – 1Aiii** show schematics diagram of a polymerase having at least one contact. FIG. 1A-i shows Φ 29 polymerase with a single Avitag at the N terminus (sequence N Contact 1). FIG. 1A-ii shows Φ 29 polymerase with a second Avitag inserted between E279 and D280 (sequence Contact 2). FIG. 1Ai-iii shows Φ 29 polymerase with the two Avitags and an additional flexible linker in the N terminus sequence (sequence N Contact 2).

[0015] **FIG. 1B** shows an SDS-PAGE gel showing the Gen III modified polymerase in lane 2. On incubating with streptavidin, the major product (red arrow) is a complex with a mass approximately equal to the polymerase plus two streptavidin molecules. Lane 4 is WT streptavidin.

[0016] **FIG. 1C** shows an agarose DNA gel showing elongation products from a rolling circle polymerization assay using (lane 1) the Gen III polymerase and (lane 4) the polymerase complexed with streptavidin. The rolling circle template is 5'-P-

CCGTACGATTCGTATCTACTATCGTTTCGATTCGCATCATCTA -3' (SEQ ID NO: 4)

and the primer for forming the double stranded region is 5'-

GGCATGCTAAGCATAGATGAT -3' (SEQ ID NO: 5).

[0017] **FIG. 2** shows the sequences of the DNA templates used and their folded structures (SEQ ID NOs: 6 – 14).

[0018] **FIGS. 3A and 3B** show urea-PAGE gel (denaturing gel (20% polyacrylamide - 8M Urea) for a homopolymer sequence (FIG. 3A) and a non-homopolymer sequence (FIG. 3B). The templates all contain hairpins, so, on a denaturing gel (20% polyacrylamide - 8M Urea), the product shows up as a sharp higher molecular weight band if fully extended. FIG. 3A shows that A10 is sometimes fully extended as shown by the presence of some fully extended product in the first lane; C10 does not extend at all; and A10C10 may form a small amount of extended product, but polymerization usually halts abruptly at two locations. The homopolymer templates fail to extend fully. FIG. 3B shows that the non-homopolymer sequence (ATC)15 is readily extended to completion.

[0019] **FIG. 4** shows a DNA sequence (SEQ ID NO: 15) and the primers that form nicks of 3, 6 and 9 base pairs (SEQ ID NOs: 16-18).

[0020] **FIG. 5A** shows a schematic diagram of a pair of electrodes bridged by a polymerase.

[0021] **FIG. 5B** shows a schematic diagram of a pair of electrodes bridged by three polymerase molecules.

[0022] **FIG. 6** shows a plot of current vs. time obtained from a device as described herein comprising one (61), two (62) or three (63) polymerase molecules attached to both electrodes. Each polymerase molecule generates a signal.

[0023] **FIGS. 7A-7E** show examples of chaotic and processive signals.

[0024] **FIGS. 8A – 8D** show data the electrical signal obtained when a polymerase is processing a repeated (ATC) sequence. FIG. 8A shows raw data plotted in current vs. time. FIG. 8B shows a plot of derivative vs. time. This plot shows the irregular speed of the polymerase. FIG. 8C shows a high-resolution plot of current vs. time. This plot shows regular gaps in the signal at a mean interval of about 11.6 ms. FIG. 8D shows a pulse width

distribution for a short segment of the data showing the gaps (width 0.5 ms) and the faster peaks between (width 0.2 ms).

[0025] **FIG. 9** shows plots of current vs. time for a DNA template consisting of (ATC) repeats. Repeated motifs are observed in panels i and iv and ii and v, consistent with the threefold sequence repeat.

[0026] **FIGS. 10A and 10B** shows conductance distributions for 3D RNA polymerase in the absence (FIG. 10A) and presence (FIG. 10B) of ribonucleotide triphosphates.

[0027] **FIG. 10C** shows the sequence of the RNA template being processed (SEQ ID NO. 19).

[0028] **FIGS. 11A and 11B** show current vs. time recordings for 3D RNA polymerase in the absence (FIG. 11A) and presence (FIG. 11B) of ribonucleotide triphosphates.

[0029] **FIG. 12** is a schematic illustrating polymerase with streptavidin contacts bridging a pair of electrodes in accordance with embodiments provided herein.

[0030] **FIG. 13** shows current-voltage curves for different connections between the molecular complex and the electrodes.

[0031] **FIGS. 14A-14D** show conductance distributions for the three generations of polymerase. (FIG. 14A) Gen I - one contact yields 2 peaks presumably because of additional contacts via the surface cysteines on $\Phi 29$. (FIG. 14B) Gen II connected via thiolated streptavidins. (FIG. 14C) Gen III connected via WT streptavidin connected to the electrodes by thiolated biotin molecules. The red arrow points to the additional high conductance peak that appears as a consequence of the second contact point. (FIG. 14D) When a bare tip is used (Gen II, thio-streptavidin on the substrate) Peak 3 is no longer observed.

[0032] **FIGS. 15A-15B** shows closed (15A) to open (15B) transition changes polymerase conductance. (FIG. 15A) Distribution in the absence of deoxynucleoside triphosphate (dNTPs) but with bound template. The polymerase is largely open. (FIG. 15B) In the presence of bound non-hydrolyzable dNTPs - the polymerase is locked in the close configuration. The highest conductance peak moves from 5.6 to ~15 nS.

[0033] **FIGS. 16A-16F** show characterizing the signals generated when polymerase is activated. FIG. 16A illustrates the current through the polymerase changes markedly over time as the STM probe drifts (curve **1602**). This drifting baseline current is fitted with an asymmetric least squares (ALS) procedure to yield a smoothed background current (curve **1601** superimposed on curve **1602**). (FIG. 16B) Subtraction of the fitted background shows the rapid changes in current that occur in an activated polymerase. Typically the dynamic signals occur in bursts (b-**1603**) interspersed with pauses (p-**1604**). FIG. 16C shows the

current recorded vs time in the absence of the complementary nucleotide triphosphate **1605**, and the ALS fit **1606**. These data were chosen to have about the same DC conductance as the active polymerase in (FIG. 16A) and (FIG. 16B). FIG. 16D has the baseline subtracted signal shows that there is noise **1607** in the inactive system. The noise coincides with the formation of a contact to the polymerase in all cases, so it is not random. FIG. 16E shows an expanded portion of the noise signal measured from an activated polymerase showing two distinct noise components - large fluctuations (LF **1608**) and small fluctuations (SF **1609**). FIG. 16F is a histogram comparing the distribution of spike heights from the active (+dTTP **1610**) and inactive (-dTTP **1611**) polymerase.

[0034] FIGS. 17A-17E show small and large fluctuations defined in terms of baseline current. Larger signals are obtained with the higher-conductance contacts. To quantify this, the baseline levels in a given run are quantified by fitting a Gaussian model to the distribution of measured currents, characterized by Gaussian peak current, I_p (FIG. 17A, 17C). A two exponential fit (I_S , I_L) is used to model the distribution of noise signal amplitudes. When the complementary base is absent (FIG. 17B) the fit converges to one value, $I_S = I_L$ in all cases. (FIG. 17D) In the presence of the complementary nucleotide, 9 out of 13 molecules showed a bimodal distribution of spike heights with $I_S \ll I_L$. (e) Values of I_L (line **1701**) and I_S (line **1702**) plotted vs the associated value of baseline current (I_p). Values of the single exponential distribution fitted for the -dTTP experiments are shown around line **1703**. Lines are linear fits yielding $L_F = 0.25(\pm 0.03)I_p$ and $L_S = 0.07(\pm 0.01)I_p$. The small fluctuations have essentially the same dependence on I_p in the control experiments (-dTTP).

[0035] FIGS. 18A-18H illustrate comparing noise characteristics from two sequences. Representative raw (**1801**) and background subtracted signals (**1802**) for d(A)₁₀ (FIG. 18A) and d(ATC)₅ (FIG. 18E). 160 ms samples of noise are shown in (FIG. 18B) and (FIG. 18F). These are raw data selected for having a constant background current. They illustrate the "25% rule" approximately, though small fluctuations appear to be absent in the run of d(ATC)₅ data shown. (FIG. 18C) and (FIG. 18G) FWHM distributions for ~ 1s of signal traces (exponential fits are shown as the thin lines). (FIG. 18D) and (FIG. 18H) Interval between signals spikes, fitted with log-normal distributions (lines).

[0036] FIGS. 19A-19D illustrate effects of biotin binding on the conductance of streptavidin. FIG. 19A shows the conductance distribution for thiolated streptavidin measured with a bare probe at a gap of 2.5 nm. Almost no data is collected when the gap is increased to 3.5 nm (see FIGS 20A and 20B). FIG. 19B shows that after the addition of the monobiotinylated Gen I polymerase, a quite different conductance distribution is obtained at

a larger gap (3.5 nm). FIG. 19C shows the conductance distributions for the Gen III bisbiotinylated polymerase contacted via thiolated streptavidin to the electrodes and FIG. 19D as bound to WT streptavidin connected to the electrodes via thio-biotin ligands.

[0037] **FIGS. 20A and 20B** illustrate conductance distributions for Φ 29 and streptavidin. FIG. 20A shows conductance distributions as a function of gap size for Gen I monobiotinylated Φ 29 bound to streptavidin on the electrodes. FIG. 20B shows conductance distributions as a function of gap size for streptavidin alone.

[0038] **FIG. 21** illustrates a repeated measurement of the conductance distribution for Φ 29 in the closed form.

[0039] **FIG. 22** illustrates conductance distributions for an active polymerase. FIG. 22 shows a conductance distribution for Φ 29 in the presence of template and dNTPs. The highest conductance feature has moved from ~6nS (-dNTPs) to 12 nS. This increases to 15 to 20 nS when the polymerase is frozen in the closed configuration with non-hydrolyzable dNTPs.

[0040] **FIG. 23** provides a current-voltage curve for an active polymerase. FIG. 23 shows an example of an IV curve obtained from an active polymerase (see FIG. 22). Arrows point to large noise spikes in the normally quiet region below ± 100 mV (compare to FIG. 13).

[0041] **FIGS. 24A-24E** illustrate signals obtained from a monobiotinylated polymerase. Noise signals obtained with the Gen I monobiotinylated polymerase. FIG. 24A provides current vs time with bound template, but no dNTPs. The highest current region is expanded in FIG. 24B. There are many jumps in the baseline, but no obvious two-level telegraph noise. FIG. 24 C is a current-time recording with dNTPs added. Noise spikes are more evident and, when an expanded trace is plotted FIG. 24D clear bursts of two-level random telegraph noise are present. FIG. 24E is a histogram of burst duration, clearly longer for the +dNTP case. Note: in these plots current is increasing in the downward direction.

[0042] **FIGS. 25A-25D** illustrate the effect of background subtraction on polymerase signals. Effect of ALS background subtraction (smoothing factor = 0.1 ms). FIGS. 25A and 25C are raw and subtracted data for a region of relatively flat baseline. FIGS. 25B (raw) and FIG. 25D (subtracted) illustrate how the noise features are well-preserved even in the presence of large baseline variations.

[0043] **FIGS. 26A and 26B** show homopolymer templates fail to extend fully. In particular, FIGS. 26A and 26 B show the difficulty of starting the polymerase on homopolymer templates. The templates all contain hairpins, so, on a denaturing gel (20% polyacrylamide - 8M Urea), the product shows up as a sharp higher molecular weight band if

fully extended. FIG. 26A, the homopolymers: A₁₀ is sometimes fully extended as shown by the presence of some fully extended product in the first lane. C₁₀ does not extend significantly. A₁₀C₁₀ may form a small amount of extended product, but polymerization usually halts abruptly at two locations. FIG. 26B the non-homopolymer sequence (ATC)₁₅ is readily extended to completion. The length of the extended product was verified by running a gel with a synthetic control of the extended length.

[0044] FIGS. 27A-27C show how noise signals from d(C)₁₀ are dense despite incomplete polymerization. FIG. 27A is a typical noise sample in a dense signal burst. FIG. 27B shows pulses that are closely spaced (~1ms) and contain both fast (0.05 ms) and slower (0.2 ms) features in the distribution of pulse widths shown in FIG. 27C.

[0045] FIGS. 28A and 28B illustrate fluctuations for active and inactive polymerases. Values of I_L (2801) and I_S (2802) plotted vs the associated value of baseline current (I_p) for 38 Φ29 molecules actively transcribing the d(ATC)₅ template FIG. 28A and for 25 molecules transcribing d(C)₁₀ (FIG. 28B). The analysis follows that presented in FIGS. 17A-17E for d(A)₁₀. Lines are linear fits yielding $i_L = 0.27(\pm 0.03)I_p$ and $i_S = 0.04(\pm 0.01)I_p$ for d(ATC)₅ and $i_L = 0.32(\pm 0.03)I_p$ and $i_S = 0.05(\pm 0.007)I_p$ for d(C)₁₀.

[0046] FIG. 29 shows Fourier Transform Infrared (FTIR) scans showing biotin functionalization of Pd.

[0047] FIGS. 30A-30D shows CV scans of functionalized Pd. The range of operating potentials for the electrodes was determined by cyclic voltammetry.

[0048] FIGS. 31A-31D show sweeps to negative potentials destabilize thiol-bound adsorbates illustrating how electrode potentials that destabilize thiol-bound adsorbate molecular films were determined.

[0049] FIGS. 32A-32D show electrodes are passivated by biomolecular adsorbates in the potential range used for electronic measurements.

DETAILED DESCRIPTION OF DISCLOSED EMBODIMENTS

[0050] 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.

[0051] 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.

[0052] For the purposes of the description, a phrase in the form “A/B” or in the form “A and/or B” means (A), (B), or (A and B). For the purposes of the description, a phrase in the form “at least one of A, B, and C” means (A), (B), (C), (A and B), (A and C), (B and C), or (A, B and C). For the purposes of the description, a phrase in the form “(A)B” means (B) or (AB) that is, A is an optional element.

[0053] The term “a” or “an” may mean more than one of an item.

[0054] The description may use the terms “embodiment” or “embodiments,” which may each refer to one or more of the same or different embodiments. Furthermore, the terms “comprising,” “including,” “having,” and the like, as used with respect to embodiments, are synonymous, and are generally intended as "open" terms (e.g., the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc.).

[0055] The term “about” means within plus or minus 10% of a stated value. For example, “about 100” would refer to any number between 90 and 110.

[0056] With respect to the use of any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[0057] 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.

[0058] 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. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

[0059] To facilitate review of the various embodiments of the disclosure, the following explanations of specific terms are provided:

[0060] **Biological sample:** 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.

[0061] **Burst:** The term “burst” 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. In embodiments, 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 in that the high state of the signal occurs at the incorporation of a nucleotide.

[0062] **Chemical Modification:** A number of various processes involving the alteration of the chemical constitution or structure of molecules. 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.

[0063] **Contacting:** Placement in direct physical association, including both a solid and liquid form.

[0064] **Complementarity:** 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" means 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.

[0065] Current stream: The term “current stream” refers to the current signal generated over time from a device described herein.

[0066] Isolated: An “isolated” biological component (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.

[0067] Label: An agent capable of detection, 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).

[0068] Linked or linker: The term “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).

[0069] 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.

[0070] **Nick:** A discontinuity in a double stranded DNA molecule where there is no phosphodiester bond between adjacent nucleotides of one strand typically through damage or enzyme action. Nicks allow for release of torsion in the strand during DNA replication. Nicks may also play a role in the DNA mismatch repair mechanisms that fix errors on both the leading and lagging daughter strands. By creating nicks, the DNA forms a circular shape.

[0071] **Non-naturally occurring or engineered:** Terms used herein as interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean 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.

[0072] **Nucleic acid:** 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 a particular example, a nucleic acid molecule is a single stranded (ss) DNA or RNA molecule, such as a probe or primer. In another particular example, a nucleic acid molecule is a double stranded (ds) nucleic acid, such as a target nucleic acid. The term “**nucleotide**” 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.

[0073] **Optional:** “Optional” or “optionally” means that the subsequently described event or circumstance can but need not occur, and that the description includes instances where said event or circumstance occurs and instances where it does not.

[0074] **Pause:** The term “pause” 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.

[0075] **Polymerase:** 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.

[0076] **Primer:** 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.

[0077] Probe: 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, 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.

[0078] Protein: The terms “protein,” “peptide,” “polypeptide” refer, interchangeably, to a polymer of amino acids and/or amino acid analogs that are joined by peptide bonds or peptide bond mimetics. The twenty naturally-occurring amino acids and their single-letter and three-letter designations are as follows: Alanine A Ala; Cysteine C Cys; Aspartic Acid D Asp; Glutamic acid E Glu; Phenylalanine F Phe; Glycine G Gly; Histidine H His; Isoleucine I He; Lysine K Lys; Leucine L Leu; Methionine M Met; Asparagine N Asn; Proline P Pro; Glutamine Q Gln; Arginine R Arg; Serine S Ser; Threonine T Thr; Valine V Val; Tryptophan w Trp; and Tyrosine Y Tyr. In one embodiment, a peptide is an antibody or fragment or part thereof, for example, any of the fragments or antibody chains listed above. In some embodiments, the peptide may be post-translationally modified.

[0079] Under conditions sufficient to: A phrase that is used to describe any environment that permits the desired activity.

[0080] II. Devices, Systems and Methods of Use

[0081] The present disclosure provides devices, systems and methods for direct measurement of polymerase activity. In some embodiments, a device for direct measurement of polymerase activity is provided. In embodiments, the polymerase is a DNA polymerase. In some embodiments, the polymerase is an RNA polymerase.

[0082] In embodiments, the device comprises a first and a second electrode, the first and second electrode being separated by a gap; and a polymerase attached to one or both

electrodes; wherein the first electrode and the second electrode have an opening formed therethrough.

[0083] In embodiments, the device comprises a first and a second electrode, the first and second electrode being separated by a gap; and a polymerase attached to one or both electrodes.

[0084] In some embodiments, the device comprises:

- (a) a dielectric substrate;
- (b) a first electrode disposed on the dielectric substrate;
- (c) an insulating dielectric layer disposed on the first electrode;
- (d) a second electrode disposed on the insulating dielectric layer;
- (e) a passivation layer disposed on the second electrode; and
- (f) a polymerase attached to one or both the electrodes;

wherein the first electrode, the insulating dielectric layer, the second electrode and passivation layer have an opening formed therethrough.

[0085] In some embodiments, the device comprises:

- (a) a dielectric substrate;
- (b) a first electrode disposed on the dielectric substrate;
- (c) a second electrode disposed on the insulating dielectric layer;
- (d) a passivation layer disposed on top of the electrodes; and
- (e) a polymerase attached to one or both the electrodes;

wherein the passivation layer has an opening formed therethrough.

[0086] In some embodiments, the device comprises:

- (a) a first and a second electrode, the first and second electrode being co-planar and separated by a gap and lying in a plane together with the second electrode;

- (b) a polymerase attached to at least one of the first or second electrodes, such as attaching to the first electrode by a first attachment site and to the second electrode at a second attachment site so that the first attachment site and the second attachment site do not move relative to each other when the polymerase undergoes open-to-closed conformational changes; and

wherein the first electrode and the second electrodes are configured for contact with a sample to be analyzed.

[0087] In embodiments in which the electrodes are planar, the device advantageously does not require a dielectric layer. Devices requiring dielectric layers can suffer from drawbacks. Dielectric layers require adhesion layers to adhere to the electrodes. These adhesion layers can oxidize upon exposure to air, which, in effect, increases the size of the gap between the electrodes. To compensate for this effect, the dielectric layer can be made thinner. However, a thin dielectric layer is susceptible to pinholes, which can be difficult to eliminate.

[0088] In each of the device embodiments described herein, the first and/or second electrode comprise a metal selected from the group consisting of gold, platinum, palladium, and ruthenium or alloy thereof. In some embodiments, the metal is palladium.

[0089] In some embodiments, the gap has 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.

[0090] *Attachment Points on the Polymerase*

[0091] The polymerase can be attached to one electrode in one embodiment and to both electrodes in a second embodiment. The polymerase can be attached to the electrode(s) either directly or indirectly. In some embodiments, the polymerase is attached to the first electrode at a first attachment site and to the second electrode at a second attachment site. In some embodiments, the polymerase is attached to an electrode at a first site of the polymerase, such as at the n-terminus of the polymerase and a second site which is in the inactive region of the polymerase. In some embodiments, a first attachment site is specifically modified for attachment at one electrode and a second attachment site that is non-specific, but within an inactive region in the polymerase. 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 aspects, the polymerase is a biotinylated polymerase. In some aspects, the polymerase comprises an Avitag. In some aspects, the polymerase is a biotinylated polymerase and is attached to the electrode via streptavidin. In some embodiments, the polymerase is modified to incorporate a Streptag.

[0092] 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). 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 20 nm. In embodiments, the distance is about 1 nm to about 10 nm. In embodiments, the distance is about 1 nm to about 5 nm. In embodiments, 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 embodiments, the distance is about 1 nm, about 2 nm, about 3 nm, about 4 nm, about 5 nm, about 6 nm, about 7 nm, about 8 nm, about 9 nm, about 10 nm. In embodiments, the distances is no more than about 10 nm, such as no more than 10 nm, 9 nm, 8 nm, 7 nm, 6 nm, 5 nm, 4 nm or 3 nm.

[0093] 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 structure of many polymerases is available (see, for example, www.rcsb.org/) in both open and closed forms. Thus, when selecting the two attachment points, the two residues (e.g., attachment points) must be separated from each other and have the same atomic coordinates in both the open and closed forms. In embodiments, the attachment sites are in the inactive region of the polymerase.

[0094] When the polymerase comprises two Avitag linkers, the ability to bind two streptavidin molecules is demonstrated by the formation of the appropriate product on a protein gel, as shown in **FIG. 1B**.

[0095] The polymerase, as modified, must function efficiently, as shown by a rolling-circle amplification assay in **FIG. 1C**. Note that this assay shows that the modified polymerase, attached to the streptavidin molecular wires, performs continuously and

repeatedly polymerization on a circular template. Therefore, circular templates can be used for repeated sequencing of a given molecule.

[0096] In some embodiments, the polymerase incorporates an inserted flexible sequence as shown in **FIG. 1A**. It is contemplated that 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 (i.e., a change is smaller than one pH point). An exemplary flexible sequence is GASSGNSTNGTSNGSS (SEQ ID NO: 20).

[0097] *The Polymerase Template*

[0098] In some embodiments, the device further comprises a nucleic acid template. The nucleic acid template is a DNA template in one embodiment and an RNA template in a second embodiment, such as a nucleic acid template with 9 or more nucleotides or bases, such as 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 etc. nucleotides or bases.

[0099] For the polymerase to operate at maximum speed, and for the electrical signals to be readily processed, it is a requirement that the polymerase not be stalled or obstructed by secondary structures formed by the DNA template. This is particularly true in the case where the template contains single stranded regions. **FIG. 2** as well as the Example section provides exemplary primer templates that can be used. **FIG. 3** shows the products of polymerization reaction on a denaturing gel. Because the templates are self-primed by hairpins (**FIG. 2**), full extension of the template by a polymerase results in only one product at the molecular weight of the fully polymerized chain. **FIG. 3** shows that this is only the case for the homopolymer d(ATC)₅.

[00100] In order for the technique to work on genomic DNA, templates must be double stranded to remove the secondary structures formed by single-stranded DNA. It is expected that obstructions will, nonetheless remain. Methods to identify those regions from the signal are discussed below. Polymerase Φ 29 has excellent strand displacement activity but does not start efficiently at a nick of only one nucleotide. To address this, primer libraries with larger nicks can be used. Exemplary primers are provided in **FIG. 4**.

[00101] The present disclosure also provides a system for direct measurement of a polymerase activity. The system comprises a device as described herein; a means for introducing a nucleic acid template and dNTPs; a means for applying a bias between the first

and second electrode of value; and a means for monitoring fluctuations that occur as a dNTP interacts with a polymerase.

[00102] *Methods of Use*

[00103] The present disclosure provides methods of using the devices described herein. In a first embodiment, the present disclosure provides a method for detecting the incorporation of a nucleotide. In a second embodiment, the present disclosure provides a method of determining the number of nucleotides repeated in a nucleic acid template. In a third embodiment, the present disclosure provides a method of determining the identity of a nucleotide being incorporated into a nucleic acid template sequence. In a fourth embodiment, the present disclosure provides a method of sequencing a nucleotide. Also disclosed are methods for sequencing a homopolymer sequence by counting the number of repeats of a pause-burst-pause sequence in the presence of a single nucleotide. Further, methods to analyze current signals in the case where more than one molecule is present in the gap/junction are disclosed in which signal levels are counted, such as a signal level of 2 indicates 1 molecule is present in the junction, a signal level of 3 indicates 2 molecules are present etc.

[00104] The signature of an active polymerase is the presence of current fluctuations on the millisecond timescale that are 25% or more of the DC current flowing through the polymerase. *See* “Engineering an Enzyme for Direct Electrical Monitoring of Activity”. Small fluctuations, on the order of 10% of the DC current are always present, even in the absence of components critical to the function of the polymerase. As such, fluctuations of 10% or less indicate the polymerase is inactive. 50 to 70% of the molecules studied are active when wired into a circuit. A second limitation occurs in functionalizing “wider” electrodes so that that there is sufficient space for more than one polymerase molecule to bridge the electrode gap and attach to both electrodes. For example, if the electrode lengths are 10 nm, only one polymerase could attach to both electrodes. However, if the electrode lengths are 20 nm, 2 polymerase molecules could attach to both electrodes.

[00105] Referring to **FIG. 5**, it is possible to ensure that only one molecule (**2**) bridges a gap in electrode (**1**) by making the electrodes sufficiently narrow. However, functionalizing narrow electrodes is difficult because the probability that a suitable coupling geometry exists in a narrow gap is small. However, in wider electrodes (**3**) more than one polymerase may bind. However, as discussed below, the preferred signals generated by an active polymerase consisting of two well defined signal levels. With more than one molecule in the gap, signals are additive, as illustrated in **FIG. 6** which shows representative signals for one (**61**), two

(62), and three (63) molecules in one device. Because the currents from each molecule consist of two levels, identifying devices that contain more than one molecule is straightforward. Furthermore, the number of molecules in each device can be counted by counting the discrete current levels (64). For a few (2, 3 or 4) molecules, a machine-learning algorithm³ can be trained, using known template sequences, to recognize these more complicated signals from multiple molecules in the gap.

[00106] *Filtering Signals*

[00107] The large signals are only observed for active polymerases, but the nature of the signal changes drastically, depending on the ability of the polymerase to process a template. As shown in **FIG. 3A**, the d(C)10 templates are hardly extended at all and the d(A)10 templates are extended sporadically. The d(ATC)5 templates are extended efficiently (**FIG. 3B**). Correspondingly, signals from d(A)10 templates and d(C)10 templates are irregular, fluctuating in amplitude and pulse width, signals labeled as “chaotic” in **FIGS. 7A** and **7B**. In contrast, signals obtained with the efficiently processed d(ATC)5 are uniform in amplitude with time (**FIG. 7C**) and labeled as “Processive”. Processive signals are those in which the polymerase is working its way through a template whereas chaotic signals indicate the polymerase is attempting to mount a template. The disclosed devices, systems and methods allow the differences between the two to be evaluated.

[00108] Processive signals continue over extended times if the template is efficiently processed as shown by the longer run in **FIG. 7E**. In templates that are at least partially fully processed (i.e. d(A)10 – **FIG. 3A**) chaotic signals are interspersed with processive signals (**FIG. 7E**). Sequencing is greatly simplified on double stranded templates where many types of sequence are efficiently processed by Φ 29 polymerase, resulting in long runs of processive signals. Even in double stranded DNA, secondary structure will stall a polymerase, for example in a run of G bases where G tetraplexes form. These structures are readily identified by a return to the chaotic signal type. In embodiments, the chaotic spikes are removed/rejected.

[00109] *Interpreting Signals*

[00110] Processive runs are straightforward to interpret. A high time resolution signal is shown in **FIG. 8**. There are distinct pauses at intervals of about 12 ms, corresponding to the known nucleotide incorporation rate of Φ 29 polymerase⁴ in the saturating concentrations of dNTP used here (marked by the arrows in **FIG. 8C**). Thus, the processive signals allow for a straightforward way of counting each nucleotide incorporation. This is extremely advantageous in interpreting signals from runs of repeated sequence, notoriously difficult to

analyze with present methods. More interesting yet is the rich detail inside each pulse. **FIG. 9** shows a time sequence of pulses from the d(ATC)₅ polymer, acquired in the time sequence shown by the Roman numerals. There is clearly a threefold repeat in the signal (labeled 1,2,3).

[00111] The presence of a threefold repeat in a polymer with a threefold repeat in its sequence is not, by itself, enough to identify bases in a heteropolymer. A major factor in polymerase selectivity is stacking between the penultimate base and the dNTP being incorporated so that contacts with the polymerase do not change much between the four nucleotides - that is to say, polymerase-nucleotide contacts (which could generate a nucleotide specific signal) are not themselves very nucleotide specific. There are sequence-specific differences in incorporation kinetics whereas measurable differences in the incorporation kinetics of the various combinations of dNTP and penultimate base have been shown, as a consequence of base-stacking interactions in the polymerase binding pocket as the incoming dNTP is loaded on top of the previously incorporated nucleotide. Each pulse is a nucleotide incorporation.

[00112] The method for decoding these signals comprises:

- (a) constructing double stranded templates comprising repeats of all possible 16 base stacking combination (e.g., A on A, A on T, A on C, A on G and so on);
- (b) introducing a solution comprising a dNTP to the device under conditions that allow for incorporation of the dNTP complementary to the nucleic acid template;
- (c) obtaining a signal stream from all 16 combinations of repeat sequences, wherein the signal stream comprises at least one pause and at least one burst;
- (d) training a machine learning algorithm (Chang, S.; Huang, S.; Liu, H.; Zhang, P.; Liang, F.; Akahori, R.; Li, S.; Gyrfas, B.; Shumway, J.; Ashcroft, B.; He, J.; Lindsay, S., Chemical recognition and binding kinetics in a functionalized tunnel junction. *Nanotechnology* **2012**, 23 (23), 235101 which is hereby incorporated by reference in its entirety) to recognize the burst associated with each of the 16 types of signal; and
- (e) using the algorithm to identify each nucleotide in the signal from a heteropolymer by finding the best fit to the training model developed with the 16 possible stacking

combinations. The starting point, that is to say the nucleotide immediately before the first new nucleotide incorporated is known because the sequence of the primer template is known.

[00113] The algorithm analyzes the signal stream to characterize the bursts in between each pause, by the number of pulses, their durations, sharpness as characterized by Fourier transforms, the intervals between them, and then training a machine-learning algorithm to associate signal features with a particular nucleotide or nucleic acid sequence.

[00114] OTHER NUCLEIC ACIDS

[00115] **FIG. 10** shows conductance distributions measured on an RNA dependent RNA polymerase, polio virus 3D pol (RSCPDB structure 1RDR). This polymerase has been modified to include the peptide sequence known as a Strep tag at its C-terminus. This binds streptavidin, allowing a single specific connection to an electrode, as described in the Examples below. **FIG. 11** shows conductance distributions for 3D pol primed with a hairpin RNA template (**FIG. 10C**). The conductance of the polymerase is increased dramatically, greater than 25% of the baseline current passing through the polymerase when it is active, as previously observed for DNA polymerase. Similarly, large current fluctuations are only observed when the polymerase is polymerizing the RNA template, as shown in **FIG. 11**.

EXAMPLES

Example 1

Materials and Methods for Example 2 and Representative Embodiments

[00116] This Example provides the materials and methods utilized for the studies disclosed in Example 2 and for representative embodiments disclosed herein.

[00117] **Expression and purification of Avitag ϕ 29 DNA polymerase.** *E. coli* strain BL21 (DE3) (Novagen) was transformed with different versions of the ϕ 29 gene embedded in pET15b plasmid and grown on an LB agar plate (1% tryptone, 0.5% yeast extract, 0.5% NaCl, and 1.5% agar) containing ampicillin (50 μ g/ml) to select the transformants. Cells were grown in LB medium (20 ml) at 37°C with shaking for 12 hours. A part of the culture was diluted 1:1000 into fresh LB medium (1L) and grown at 37°C with shaking. The 1L culture was induced by adding 0.5 mM IPTG when the OD₆₀₀ reached 0.6 and kept shaking overnight at 18°C. Cells were harvested in a 1L centrifuge bottle by centrifugation at 5,710g for 20 minutes at 4°C and stored at -80°C until required.

[00118] Biotinylation of Avitag ϕ 29 DNA polymerase. For subsequent BirA biotinylation, the purified protein was exchanged to a buffer containing 20 mM potassium phosphate buffer pH 7.0, 200 mM L-Glutamic acid potassium salt and 1 mM DTT. The *in vitro* enzymatic biotinylation was performed by incubating 100 μ g of polymerase in the same buffer with 10 mM ATP, 10 mM Mg(OAc)₂, 50 mM biotin and 15 units of BirA (Avidity) for 1 hour at 30°C. Free biotin was removed by a desalting column (GE Life Sciences).

[00119] Rolling circle replication (RCR) ssDNA template and primer. Linear single strand oligonucleotides RCR (5'-p - CCGTACGATTCGTATCTACTATCGTTTCGATTTCGCATCATCTA -3'; SEQ ID NO: 21) was used to form circular RCR template by enzymatic self-ligation with Circligase (Epicentre). 0.1 nmol linear single strand RCR DNA was mixed with 100 Units Circligase in 1 X reaction buffer containing 50 μ M ATP and 2.5 mM MnCl₂. After 2 hours incubation at 60°C, the product was heated up to 80°C for 10 minutes to inactivate the Circligase. The linear ssDNA left in the solution was digested by Exo I (NEB). The RCR template was analyzed by electrophoresis on a denaturing gel containing 8 M urea and 20% polyacrylamide for quality control. 2.5 pmols of RCR template was annealed with 50 pmols RCR primer (5'-GGCATGCTAAGCATAGATGAT -3'; SEQ ID NO: 22) by heating up to 95°C for 5 minutes and gradually cooling down to room temperature (decreasing 0.1°C/s) and stored at -20°C for later use.

[00120] Activity assay of Avitag ϕ 29 DNA polymerase. A rolling circle replication reaction was performed for the activity test of all versions of ϕ 29 DNA polymerase. 1.25 pmols RCR template and primer complex was mixed with 500 μ M dNTP and 4 pmols ϕ 29 DNA polymerase in 1 X reaction buffer containing 50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM (NH₄)₂SO₄, 4 mM DTT. The mixture was incubated at 30°C for 1 hour. The product was visualized on 0.8% agarose gel by GelRed (Biotium) staining.

[00121] Functionalizing substrates and STM probes. Palladium substrates for STM measurement were prepared by evaporating a 200 nm palladium film onto a silicon wafer using an electron-beam evaporator (Lesker PVD 75), with a 10 nm titanium adhesion layer. The substrates were treated with a hydrogen flame immediately before functionalizing and then immersed in solutions of thiolated streptavidin (ProteinMods) or thiolated biotin overnight. The thiolated biotin was prepared and dissolved in freshly degassed pure ethanol to a final concentration of 50 μ M. 1 μ M thiolated streptavidin solutions in 1 mM PB buffer were used for substrate functionalization. All the buffers and solutions were prepared in Milli-Q water with a conductivity of 18.2 M Ω . For all measurements, the 1 mM PB buffer

(pH 7.4) was degassed with argon to avoid interference from oxygen. The polymerization buffer was 1mM phosphate buffer, pH=7.4, 4 mM TCEP, 10mM MgCl₂ with 1 mM dNTPs and 1 μ M template. Sequences of the self-priming templates used are shown in FIG. 1Ai-1Aiii. Substrate functionalization was characterized by Fourier transform infrared (FTIR) spectroscopy. STM probes were etched from a 0.25mm Pd wire (California Fine Wires) by an AC electrochemical method. To avoid current leakage, probes were insulated with high-density polyethylene following the method described earlier for gold probes. Each probe was tested by STM in 1 mM phosphate buffer (pH 7.4) at +0.5 V bias to ensure the leakage current was <1pA. For functionalization, the probe was immersed in ligand solutions for 4 h or overnight. After that, it was taken out, rinsed with water, gently blown dry with nitrogen gas, and used immediately. Cyclic voltammetry was used to check that the potential regions used (+50 to -50mV vs Ag/AgCl, 10 mM KCl) were free from Faradaic currents in the presence of the various components of the assembly.

[00122] STM Measurements. STM measurements were carried out on a PicoSPM scanning probe microscope (Agilent Technologies), using a DAQ card (PCI-6821 or PCIE-7842R, National Instruments) for data acquisition. The Teflon cell, in which buffer solution and analytes were added, was cleaned with Piranha solution and then sonicated in Milli-Q water three times to remove residues (Note that Piranha solution is highly corrosive and must be handled with extreme care). In order to better control the surface potential, an Ag/AgCl reference electrode with a 10 mM KCl salt bridge was connected onto the substrate. The probe was firstly engaged to 4 pA setpoint current with a bias of -0.2 V and then left to stabilize for 2 h before measurement. For STM IV sweep measurements, the servo system was first turned off and the probe was retracted by ΔZ nm with a speed of 1 nm/s. After that, the probe was suspended at that height for 1 minute, during which a custom Labview program was used to monitor the current change. Once the current exceeded a threshold of 50 pA, we considered it as a binding event and IV sweeps were started from -0.2 V to +0.2 V and then back, with a sweep rate of 1 V/s, followed by a 0.2 s resting. Subsequently, the current was checked again. If the current was still more than twice the noise level (6 pA), IV curves were continuously recorded until the bound protein molecule escaped. After one-minute of measurement, the servo system was turned on to re-engage the probe and then the whole process repeated. In each measurement, at least 1000 IV curves were collected, from which curves that reproduced on the up sweep and down sweep were selected (80% of the total) to construct the conductance distribution histogram. Current vs. time traces were recorded by another Labview program with a similar procedure except that the bias was held

constant during the probe holding process. The analog-to-digital sampling rate is 50 KHz, about 5 times the intrinsic response of the STM electronics.

Example 2

[00123] Proteins have been shown to be electrically-conductive if tethered to an electrode by means of a specific binding agent, opening the possibility of building electronic devices that exploit the remarkable chemical versatility of enzymes by direct read-out of activity. Single-molecule conductance increases tenfold if two specific contacts are made (as is possible with bivalent antibodies). Here, the problem of forming contacts with proteins that, unlike antibodies, do not possess a multiplicity of native binding sites was addressed. In particular, contact points were engineered into a Φ 29 polymerase, with the points chosen so as to leave the active site free. Contacts are formed by introducing biotinylatable peptide sequences into Φ 29 and functionalizing electrodes with streptavidin. Φ 29 connected by one biotinylated contact and one non-specific contact showing rapid small fluctuations in current when activated. Signals were greatly enhanced with two specific contacts, a feature in the conductance distribution changing by > 9 nS between the open and closed conformations of the polymerase. Polymerase activity is accompanied by large changes in conductance over the millisecond timescale.

[00124] Proteins are widely assumed to be insulators. Reports of metallic conduction in bacterial wires and long range conduction in protein multilayers were thought to be exceptions. However, a number of proteins, chosen only for their *redox inactivity* (i.e. no electron transport function, no redox active centers) conduct very well if contacted by binding agents that can inject charge carriers into their interiors. This conductance is limited to electrons (or holes) because the proteins we have studied to date act to passivate electrodes in cyclic voltammetry studies. Furthermore, the internal conductance of the protein is sensitive to conformational changes. Key to these experiments was the use of antibodies together with electrodes that were functionalized with an epitope to which the two binding domains of the antibody could bind. The spread of conductances (owing to the range of contact geometries) was measured by recording current-voltage (IV) curves from approximately 1000 molecules, and plotting histograms of the conductances derived from the IV curves. Distributions for antibodies had two peaks, one at ~ 0.2 nS (similar to the value of the single peak observed for proteins bound by only one ligand) with a second peak at $\sim 10\times$ the conductance (~ 2 nS). The second peak was interpreted as owing to specific binding events at each binding domain of the antibody, attached so that the antibody bridges the electrode gap. This interpretation was

confirmed using Fab fragments (one binding head) or functionalizing only one of the two electrodes with an epitope. It was also found that the internal conductance of proteins contacted via specific bindings agent is higher than the conductance associated with contact regions, so that that the measured electronic properties do not change much with gap size, until contact is lost with the protein. Proteins are remarkably versatile molecular machines, capable of molecular recognition, highly selective catalysis, directional energy transfer, directed polymer synthesis and many other functions, so integration of proteins into bioelectronic devices has been a long sought-after goal, and here the problem of making specific contacts to proteins that do not have multiple ligand binding sites was addressed, with the goal of demonstrating a device that transduces enzyme activity into electrical signals directly.

[00125] Φ 29 polymerase was selected to be evaluated. It is a preferred polymerase for DNA sequencing applications because it is highly processive, has a high strand displacement activity (needing only a nick to initiate polymerization) and is very accurate (1 part in 10^6). It is also fast, turning over 100 nucleotides per second at saturating concentrations of nucleotidetriphosphates (dNTP) and template. The major motion of the enzyme involves the closing of a "finger" domain into the "palm" and "thumb" once template and matching dNTP are assembled and ready for incorporation. At some point over the open-closed-open transition, the enzyme translocates the template so that the next available site is positioned for binding of the complementary dNTP.

[00126] Here, the engineering of specific binding sites for electrical contacts into Φ 29 polymerase is reported using insertion of the Avitag sequence into a clone for the polymerase, and subsequently biotinylating the lysine within the Avitag peptide sequence using the BirA enzyme. This then allows for strong, and specific binding to Streptavidin. Initially a Φ 29 biotinylated was tested just at the N terminus (Gen I), and bound to electrodes that had been covered with thiolated Streptavidin. Biotin-bound Streptavidin makes an excellent molecular wire and also serves to keep the Φ 29 (with its seven surface cysteines) away from the metal electrodes. A second contact was introduced about 5 nm distant from the first (Gen II), at a site in the (inactivated) exonuclease domain. A third type of polymerase (Gen III) was made by incorporating a flexible linker next to the N-terminal Avitag. Finally, direct tethering of a thiolated Streptavidin to the electrodes was compared, with indirect tethering, using a thiolated biotin molecule to functionalize the electrodes. Both conductance distributions (by recording current-voltage, IV, curves) were measured and dynamic responses by recording current vs. time ($I(t)$) for bound polymerases in a gap of nominally

constant size. The introduction of two binding sites introduced a new high conductance feature in the conductance distribution, analogous to the second peak found naturally for antibodies with two binding domains. It was also found that the indirect connection of streptavidin via thiolated biotin gives higher conductance than direct connection via surface thiolation of the streptavidin. The conductance of the $\Phi 29$ was found to increase greatly in the closed conformation. The change in DC conductance measured the between open and closed states is consistent with the scale of large dynamic swings in the current recorded through the enzyme as a function of time when the enzyme is active.

[00127] RESULTS AND DISCUSSION

[00128] Recombinant $\Phi 29$ DNA polymerase constructs with inserted Avitag

[00129] The starting enzyme was a $\Phi 29$ DNA polymerase, rendered exonuclease-deficient with D12A and D66A mutations. A Q5 site-directed mutagenesis kit (NEB) was used to insert the Avitag DNA sequence into a pET15b plasmid containing the mutant polymerase gene. The equivalent inserted peptide sequence is shown with flanking linker sequences, in FIGS. 1Ai-1C. The epsilon-amine of the central lysine (K, marked by an arrow on FIG. 1Aii-iii) was biotinylated using the BirA enzyme. Three generations of modified enzyme were tested. The first (FIG. 1Ai) was biotinylated only at the N terminus of the protein. The second (FIG. 1Aii) contained a second Avitag approximately 5nm from the N terminus between E279 and D280. This second site is located in the deactivated exonuclease domain and was chosen because its position does not change with respect to the N terminus over the open to closed transition. The third (FIG. 1Aiii) contained an additional flexible linking sequence (GNSTNGTSNGSS; SEQ ID NO: 22) adjacent to the N-terminal Avitag to allow for greater flexibility in contact geometry. Biotinylation was verified with a SDS-PAGE gel analysis of the free- and streptavidin bound polymerases. FIG. 1B shows the increase in molecular weight that occurs as two streptavidins are bound to the Gen III polymerase (yet higher molecular weight features are likely polymer aggregates of alternating polymerase and streptavidin). The activity of the modified polymerases was verified in vitro using a rolling circle amplification assay (FIG. 1C).

[00130] Scanning Tunneling Microscope Conductance Measurements.

[00131] Measurements were made using an electrochemical scanning tunneling microscope (Pico STM, Agilent) with insulated palladium (Pd) probes and a Pd substrate, both held under potential control using a salt-bridged reference electrode (FIG. 12). Electrodes were modified with either a thiolated biotin (SH-biotin) or thiolated streptavidin, and then incubated with a solution of the biotinylated polymerase (Methods). Measurements

were made in a reaction buffer containing MgCl_2 and tris(2-carboxyethyl)phosphine (TCEP) to prevent polymerase oxidation. Nucleotide triphosphates were added to activate the polymerases. Current-voltage (IV) characteristics were measured using a fixed gap (no servo control). The gap remained constant to within about 0.1 nm over approximately 1 minute, and the bias was swept between -0.2 and +0.2 V and back again at a rate of 1 V/s. No attempt was made to fish for proteins or to push or pull them. Thus, the trapped proteins that yielded signals were molecules that bound across the gap in equilibrium, or near equilibrium conformations. After 1 minute, the gap was returned to the set-point value. This cycle was repeated to obtain further IV sweeps. 80% of these sweeps reproduced exactly on reversing the sweep direction. (In a number of runs, the current was recorded before re-engaging the servo to check that there had been no significant drift.)

[00132] In the present example, both probe and substrate are functionalized with streptavidin. In consequence, two streptavidin molecules can contact each other across the gap. When contacts are successfully made with streptavidin, they recapitulate the distribution observed for single streptavidin molecules and no signals are observed from streptavidin alone for gaps larger than 3.5 nm. However, when biotinylated polymerase is allowed to complex with the streptavidin, significant currents are measured out to gap distances >4.5 nm with conductance distributions that are quite distinct from that of streptavidin. The data reported in this Example were obtained with a gap of 4.5 nm unless noted otherwise.

[00133] FIG. 12 shows polymerase with streptavidin contacts bridging a pair of electrodes. As illustrated in FIG. 12, an STM probe 1200 is held ~4.5 nm above a conducting substrate 1201, immersed in electrolyte and under potential control via a salt bridge (SB) 1202 connected to an Ag/AgCl reference 1203. The electrodes are functionalized with thiolated biotin 1204 which captures streptavidin molecules (SA) 1205 which then trap a polymerase ($\Phi 29$) 1206. A bias V_b 1207 is applied between the electrodes. Potentials are maintained so as to avoid Faradaic currents.

[00134] FIG. 13 shows a selection of IV curves for a Gen I and Gen II polymerase. Forward and reverse sweeps are superimposed demonstrating reproducibility. Typical current-voltage curves from a trapped single molecule of polymerase. Each trace is for a sweep from -0.2 V to +0.2 V and then back to -0.2 V. The conductances for individual molecules are obtained from the slope of these traces. Currents are increased by about an

order of magnitude for Gen III polymerase (2 biotin contacts FIG. 1Aiii - scans are labeled 1301) compared to Gen I polymerase (1 biotin contact FIG. 1Ai - scans are labeled 1302).

[00135] The response is linear, with the exception of noise spikes that appear above $\pm 100\text{mV}$ as a consequence of voltage-induced contact fluctuations. The slope of each trace was used to calculate a conductance, $G=I/V$ and histograms of conductance (based on ~ 1000 measurements each) are shown for the three generations of polymerase in FIGS. 14A-14D.

[00136] The Gen I polymerase forms only one specific contact, but in contrast to what we have observed in the past for proteins with a single binding site, there are two peaks in the conductance distribution, one at about 0.2 nS (characteristic of a weak, non-specific binding) and a second peak at $\sim 1\text{ nS}$. It is believed that the second peak is a consequence of additional surface linkages owing to the seven surface cysteines in this polymerase (conductance via a thiolated streptavidin is highly unlikely at this gap size). This second peak only appears when a polymerase is present, and the current distribution is quite different from that measured through streptavidin alone at a smaller gap size. As observed for other proteins, these distributions change very little with gap size until contact with the protein is lost. FIG. 14D shows the conductance distribution measured for a Gen II with thiolated streptavidin on the surface but with a bare (unfunctionalized) probe. The 3 nS peak observed in FIG. 14B is now absent. Taken together, these experiments show that an additional high conductance feature occurs in the conductance distribution when (a) the polymerase has two biotinylated contacts and (b) streptavidin is present on both the probe and substrate. Therefore, this highest conductance peak is attributed to biotin-mediated binding events that bridge the electrode gap.

[00137] FIG. 14C shows a conductance histogram for a Gen III polymerase, using a thiolated biotin molecule to attach Streptavidin to the electrode surface. The highest conductance feature has moved out to 5.6 nS . This is a consequence of the biotin-mediated attachment of the streptavidin to the electrodes. FIGS. 19A-19C show conductance distributions for the Gen III polymerase as connected by thiolated streptavidin (19C - larger peak = 3.7 nS), or wild type (WT) streptavidin bound to biotinylated electrodes (19D - peak 3 = 7 nS) showing the increase in conductance when a specific ligand is used for a contact. In addition, it is concluded that the introduction of the additional flexible linker has reduced the overall conductance of the complex a little for a given coupling chemistry (5.6 nS for Gen III – FIG. 14C, vs 7 nS for Gen II – FIG. 19D - with biotin mediated coupling to the electrodes

in both cases). These data demonstrate the successful engineering of a bis-functionalized polymerase that retains polymerase activity while showing increased electrical conductivity.

[00138] The conformational sensitivity of the electronic signals was next evaluated. It was shown earlier that ligand binding changes the conductance distribution of a streptavidin molecule significantly: is the same true of a polymerase? In order to probe this, a Gen III polymerase was generated to be stabilized in the closed form by incubating a template-bound polymerase with nonhydrolyzable nucleotide triphosphates (NH-dNTPs). The NH-dNTP binds the template-polymerase complex, closing the polymerase, which then remains in the closed configuration because of the non-hydrolyzable triphosphate. Conductance distributions for the normal, open (-dNTP) and closed (+NHdNTP) forms are shown in FIGS. 15A and 15B. There is a very large change in the highest conductance feature (from 5.6 to approximately 15 nS).

[00139] The highest conductance feature is quite small, so the experiment was repeated to verify its existence (FIG. 21). Interestingly, IV curves taken on the active form (using normal dNTPs) yield an intermediate conductance (FIG. 22), implying that an active polymerase may stay in a high conductance state for long enough to yield increased conductance in an IV measurement (which takes 0.4 s). Taken together, these measurements indicate that the conductance distribution is affected by the conformation of the polymerase. One interesting consequence of measuring IV curves on an active enzyme (as in FIG. 22) is that noise now appears on the curves in the usually quiet region below 100mV (FIG. 23).

[00140] Taken together, these observations indicate that it may be possible to observe signs of polymerase activity in a recording of current taken over time ($I(t)$), as the polymerase incorporates incoming nucleotides. $I(t)$ was recorded at a constant 50mV bias (below the threshold for contact noise) for a single stranded template primed by a 15 base pair hairpin (Methods). To do this, the same procedure outline above was followed for obtaining IV curves, but holding the bias constant, and recording current for 90s before withdrawing the probe and reestablishing servo control. A gap of 2.5 nm under servo control was started with, increasing the gap to 6 nm, and then brought the tip down to 4.5 nm and then recorded current for 60 s. Typically, no current was recorded for the first 10-20 s, after which a contact formed and an $I(t)$ curve was obtained. Contacts were formed with molecules in >50% of these "fishing" attempts. The currents jumped suddenly on contact with the molecule, but then changes substantially as the contact point drifts. A typical current-time trace is shown in FIG. 16A. By comparing the distribution of currents measured in current-time traces with the conductance distributions obtained from many repeated contacts with the molecule, we have

shown that the variation in the current-time trace can be accounted for by variations in the contact point.

[00141] Noise Analysis.

[00142] The Gen I polymerase with a single contact was first studied. Bursts of noise were clearly visible when the polymerase was activated on addition of the complementary dNTPs (FIG. 24A). However, currents were generally low (c.f., the conductance distribution in FIG. 14A) as were the signal levels. It was for this reason that we constructed the bis-functionalized polymerases that yield the higher conductance features seen in FIGS. 14B and 14C. FIG. 16A shows a typical $I(t)$ trace for an $d(A)_{10}$ single stranded template in the presence of all four dNTPs (black trace, mainly obscured by the overlaid red trace) in the polymerization buffer (Methods). Large changes in the background current obscure the fluctuations, so the background was removed using an asymmetric least squares (ALS) fit. The ALS accurately follows the background without distorting the noise signals (FIGS 25A-25D). (Attempts to remove the background with a median filter distorted the noise signals.) The ALS fit to the raw data is superimposed on it in FIG. 16A. The flattened noise trace, obtained by subtracting the ALS fit from the raw data, is shown in FIG. 16B. The noise signals are quite large, in excess of a nA in this case, and occur in bursts with pauses in between them, though no discernable repeat pattern was found. A control signal of similar overall DC conductance was selected (FIG. 16C), taken in a buffer containing only dATP, dCTP and dGTP, lacking the necessary dTTP. Following the same subtraction procedure outlined for the active molecule, a noise trace is obtained, as shown in FIG. 16D. The result is clearly not random noise, because it coincides exactly with the period for which the molecule is contacted. The same behavior was seen in eight other control experiments. So clearly, both active and inactive polymerases produce electrical noise. Given the variable background signal and the stochastic nature of the fluctuations, can we distinguish active from inactive polymerase? Inspection of the details of the noise signals at a higher time resolution offers a clue. FIG. 16E shows a trace containing both large (LF) and small (SF) fluctuations. If the size of the fluctuations were compared from the control experiment (-dTTP) with the size distribution of the fluctuations for the active polymerase (+dTTP) (measured at about the same background current), the current fluctuations in the active polymerase are much larger (FIG. 16F). The runs of similar background current were compared, because inspection of

many traces indicates that the size of the current fluctuations increases with the background current.

[00143] In order to quantify the relationship between the size of the fluctuations and the background current, the distributions of both were studied for a number of molecules with, and without dTTP present in the buffer solution. For each molecule, the ALS fitted baseline currents were binned as shown by the examples in FIGS. 17A, 17C. Many of these distributions could be fitted by a Gaussian. However, some could not - for example, the background can jump between two or three levels. In these cases, the largest peak that was clear of the background was fitted. The peak of the fitted Gaussian, I_p , was then used to characterize the baseline for that run. Example of the binned noise signals are given in FIGS. 17B, 17D. To characterize these a double exponential distribution was used:

[00144]
$$N(i) = A_1 \exp\left(-\frac{i}{i_s}\right) + A_2 \exp\left(-\frac{i}{i_L}\right)$$

[00145] where i is the current in a given bin of the histogrammed current distribution. In the case of the experiments in which dTTP was withheld (e.g., FIG. 17B), the fits all converged to a single exponential ($i_s = i_L$). For recordings with dTTP present, most of the fits converged on the double exponential distribution with $i_s < i_L$ (4 of 13 molecules showed the small peaks only). The results are summarized in FIG. 17E. Activated molecules (+dTTP) showed both large and small fluctuations. The controls (-dTTP) showed only small fluctuations, essentially equal to the small fluctuations also seen in the active polymerases. There is an approximately linear relation between background current and fluctuation amplitude as shown by the three linear fits. For the large fluctuations, characteristic of active polymerases, $i_L = (0.25 \pm 0.026)i_p$ (i is the 1/e value of the fit, so larger and smaller spikes are observed also). For the small fluctuations, present in both active and inactive polymerases, $i_s = (0.06 \pm 0.01)i_p$. Thus, the active state can be identified by the presence of fluctuations that are about 25% of the baseline current, while fluctuations in the inactive state are about 10% of the baseline current. Not all polymerase molecules contacted were active as indicated by the lack of large fluctuations in 4 of the 13 molecules studied. On the other hand, none of the eight -dTTP control runs showed large fluctuations (Table 1).

[00146] This analysis was repeated using data obtained in 38 runs in which d(ATC)₅ was used as the template, with the results shown in FIGS. 28A-28B. The fitted amplitude distributions for the large fluctuations (LF) show considerable variation, but the trends

observed for $(dA)_{10}$ (FIGS. 17A-17E) are reproduced well with $i_L = 0.27(\pm 0.03)I_p$ and $i_s = 0.04(\pm 0.01)I_p$ for the small amplitudes.

[00147] In order to confirm the association of large fluctuations (by which we now mean 25%, or more, of the baseline current) with polymerase activity experiments were carried out with different templates in different conditions. In each case, the DNA template consisted of a 15 base double-stranded region connected by a 9 base loop with the ssDNA template extending from the 5' end of the hairpin. The ssDNA templates were $(dA)_{10}$, $(dC)_{10}$, $(dA)_{10}(dC)_{10}$ and $(dTAC)_5$. Positive experiments were carried out in the reaction buffer (1mM phosphate buffer, pH=7.4, 4 mM TCEP, 10mM $MgCl_2$ with 1 mM dNTPs and 1 μ M template). In the control experiments, one critical ingredient was withheld as listed below in Table 1 (where they are labeled as control, C). In addition, measurements were carried out using nonhydrolyzable dNTPs. Each of the runs was analyzed as described above for $(dA)_{10} \pm dTTP$.

Experimental Conditions	Control/Active	Fraction Large Fluctuations
$(dA)_{10}$ +dNTPs + Mg^{2+}	A	0.69
$(dA)_{10}$ +dATP, + dCTP, +dGTP, -dTTP, + Mg^{2+}	C	0
$(dC)_{10}$ +dNTPs + Mg^{2+}	A	0.52
$(dA)_{10}(dC)_{10}$ +dNTPs + Mg^{2+}	A	0.44
$(dATC)_5$ +dNTPs + Mg^{2+}	A	0.52
$(dA)_{10}(dC)_{10}$ - dNTPs + Mg^{2+}	C	0
$(dA)_{10}(dC)_{10}$ +dNTPs - Mg^{2+}	C	0
$(dA)_{10}(dC)_{10}$ +dTTP -dGTP, + Mg^{2+}	C	0.32
$(dA)_{10}(dC)_{10}$ NH-dNTPs + Mg^{2+}	C	0
No template +dNTPs + Mg^{2+}	C	0

Table 1: Occurrence of large fluctuations (fraction of measured molecules) for various experimental conditions. NH = non-hydrolyzable dNTPs. Large fluctuations are identified by a machine fit converging on a two component distribution of fluctuations.

[00148] Withholding any one of the ingredients critical to polymerase function appears to abolish the large fluctuations. An interesting exception was $(dA)_{10}(dC)_{10}$ in the presence of dTTP only. One expects that the polymerase would reach the end of the A tract and then stall for want of the missing dGTP nucleotide, but that is not the case. In order to investigate the behavior of the polymerase further, denaturing gels of the polymerization reaction for all four templates were run. Because the primer is terminated in a hairpin, the

denatured polymerization product produces a higher molecular weight feature on the gel, with no feature at the molecular weight of the primer if the reaction goes to completion. The results are seen in FIGS. 26A and 26B. Though single-stranded homopolymers have been used for studies of a different polymerase motion previously it is known that polymerases tend to stall as they encounter secondary structure in a single-stranded DNA template. The gels show that the only template that is fully extended is d(ATC)₅. d(A)₁₀ is fully extended in a significant fraction of events. d(C)₁₀ does not appear to be extended at all, which is quite surprising because the noise signals from this template are quite dense (FIGS. 27A-27C). The d(A)₁₀d(C)₁₀ is extended in a fraction of cases, but only to the end of the A tract. Thus, it is likely that the polymerase falls off this template, allowing for polymerization to start again in the presence dTTP only, accounting for the observation of activity in the absence of dGTP (Table 1). Longer pauses are observed in this experiment as might be consistent with the polymerase falling off and having to bind a new template.

[00149] The details of the noise for two different sequences were studied to investigate whether any of the noise features changed as the template sequence was changed. Homopolymer templates could not be compared because of the poor activity of the enzyme on C tracts, so we have compared d(A)₁₀ and d(ATC)₅. FIG. 18A shows raw data for d(A)₁₀ (a) and d(ATC)₅ (e) with the extracted noise signature plotted below. Although the flattening algorithm does not distort the signal significantly, the amplitude of fluctuations is affected by changes in the baseline current, as discussed above. For this reason, regions where the baseline was constant and the noise signal is dense were looked for. Two representative samples are shown in FIGS. 18B and 18F. The difference in these noise signals is noticeable. All of the runs like this in the A or C homopolymers had significant amplitude fluctuations. Since the baseline current (and therefore the contact) was constant in the regions examined, these fluctuations reflect the behavior of the polymerase. On the other hand, the ATC polymer signals have many regions where the fluctuation amplitude (at constant baseline current) was constant, for times of up to 0.5 s, the longest constant baseline signal obtained. This difference may be related to the difference in processivity observed in the biochemical measurements (FIGS. 26A and 26B). It is striking that signals obtained from the d(C)₁₀ template are similar to those from d(A)₁₀ (FIGS. 27A-27C) despite the fact that d(C)₁₀ is barely extended. So the rather chaotic signals observed in both A and C homopolymers may reflect stochastic motion of the polymerase on these templates. In contrast, the readily processed d(ATC)₅ yields regions of steady signal that might prove easier to interpret. The peak widths (FIGS. 18C and 18G) and the distribution of intervals between pulses (FIGS.

18D and 18H) show that there are some differences in the peak characteristics and peak frequencies between the two templates.

[00150] Additionally, FIG. 29 shows Fourier Transform Infrared (FTIR) scans showing biotin functionalization of Pd. Further, FIGS. 30A-30D shows CV scans of functionalized Pd. The range of operating potentials for the electrodes was determined by cyclic voltammetry. Pd substrates were cut into 0.5 cm x 4.0 cm in size and used as the working electrode, with an active cell area of about 0.5 cm x 1.0 cm. The substrate was treated with a hydrogen flame before functionalization. Cyclic voltammetry was performed on a potentiostat (Model AFCBP1, Pine Instruments), using a Pt wire as the counter electrode and an Ag/AgCl (3M KCl) as the reference electrode. The maximum sweep range is from -0.5 V to +0.5 V, with a sweep rate of 10 mV/s, starting from ± 0.05 V and increasing the scan range on 0.05V increments. (FIG. 30A) Bare Pd. (FIG. 30B)) Functionalized with SH-Biotin (50 μ M, overnight). (FIG. 30C) With streptavidin added (1 μ M, 0.5 h). (FIG. 30D) And after addition of Gen III Φ 29 polymerase (1 μ M, 2 hours). Two repeated sweeps are shown at the highest bias, showing some instability in the adsorbate.

[00151] FIGS. 31A-31D show sweeps to negative potentials destabilize thiol-bound adsorbates illustrating how electrode potentials that destabilize thiol-bound adsorbate molecular films were determined. FIG. 31A shows repeated sweeps on bare Pd reproduce well (HA is the hydrogen adsorption peak). On functionalization with biotin (FIG. 31B), biotin plus streptavidin (FIG. 31C) and biotin plus streptavidin plus Φ 29 polymerase (FIG. 31D), the HA peak, initially moved to more negative potentials relative to bare Pd, moves up towards the potential observed in bare Pd, consistent with stripping of the adsorbate by thiol reduction.

[00152] FIGS. 32A-32D show electrodes are passivated by biomolecular adsorbates in the potential range used for electronic measurements. In particular, FIGS. 32A-32D show how cyclic voltammetry was used to determine a range of potentials over which electrodes are passivated by biomolecular adsorbates. (FIG. 32A) Repeated sweeps on bare Pd. On functionalization with biotin (FIG. 32B), biotin plus streptavidin (FIG. 32C) and biotin plus streptavidin plus Φ 29 polymerase (FIG. 32D) the electrode becomes increasingly passivated. These measurements illustrate that these electronically conductive films do not exchange charge with ions in solution.

[00153] CONCLUSIONS

[00154] Engineering two contact points into a polymerase yields features in the conductance distribution that are approximately 3 to 10 times larger than those observed with

only one engineered contact and a second, non-specific, contact. The conductance of the complex of streptavidin and doubly biotinylated $\Phi 29$ is further increased if biotin is used to anchor the streptavidin to the electrodes in place of the usual anchoring by means of thiolating surface lysines. There are significant changes in the conductivity as polymerase $\Phi 29$ undergoes the open to closed transition and these changes can be detected dynamically as the occurrence of rapid noise spikes that have an amplitude of about 25% (or more) of the background current. There are similar, but much smaller, signals generated when the polymerase is not active. These small signals are also proportional to the background current (approximately 6%) and are present in both resting and active polymerase. This observation of a proportionality between the signal amplitude and the background current has implications for models of the mechanism of signal generation. Measurements of the electronic decay length in protein molecules show that the overall conductance has a large contribution from a small contact conductance, which limits the overall conductance significantly. The measurements of electronic decay length also show that the decay length is on the order of the protein size, so that it is impossible to distinguish between a slow exponential decay and a resistive model in which the decay of current scales with $1/(\text{length})$. The two models predict different dependencies of the measured current on internal changes in the molecule. In a resistor model, where the contact resistance is R_c , the internal molecular resistance is R and the change associated with a conformational transition is δR , the fractional change in current on undergoing the transition is $\delta i/i = \delta R/(R_c + R + \delta R)$. In a model based on transition probability, the current fluctuation is directly proportional to the change associated with the conformational transmission multiplied by the probability of charge injection, so again proportional to the baseline current. However, the resistive model implies a non-linear relationship between δi and i because of the dependence of the slope on R_c .

[00155] Although certain embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that a wide variety of alternate and/or equivalent embodiments or implementations calculated to achieve the same purposes may be substituted for the embodiments shown and described without departing from the scope. Those with skill in the art will readily appreciate that embodiments may be implemented in a very wide variety of ways. This application is intended to cover any adaptations or variations of the embodiments discussed herein. Therefore, it is manifestly intended that embodiments be limited only by the claims and the equivalents thereof.