

DEVICE, SYSTEM AND METHOD FOR DIRECT ELECTRICAL MEASUREMENT OF ENZYME ACTIVITY

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under HG910080 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

Electrical readout of motions that underlie functions of a native protein might enable many new types of analytical measurement without labeling. For example, monitoring functional fluctuations of an enzyme would provide a rapid and simple way of screening candidate drug molecules. Monitoring the fluctuations of proteins that process biopolymers would reveal information about their composition and conformation.

Electrical readout of enzyme function was demonstrated by Choi et al. (Choi, Moody et al. 2012) who showed that telegraph noise, induced in a carbon nanotube field effect transistor, reflected the functional motion of the enzyme lysozyme when acting on its substrate, peptidoglycan. It was realized that monitoring the fluctuations of precessive enzymes, such as DNA polymerase might thus give a method for sequencing DNA. One example was given in a controversial paper, in which the Huang group (Chen, Lee et al. 2013) claimed to measure electrical fluctuations in a polymerase as nucleotides were incorporated into an extending chain, the signals reporting the sequence of the template being extended with high accuracy. The paper was subsequently retracted (*Nature Nanotechnology* 8, 452–458 (2013); published online 5 May 2013; corrected after print 11 July 2013 and 28 August 2013; retracted after print 3 June 2015) but illustrates what might be possible if the structural fluctuations of a protein could be monitored by an electrical readout. More significantly, a working realization of this proposal was demonstrated around the same time by the Collins group who used a carbon nanotube field effect transistor to which a polymerase was tethered (Olsen, Choi et al. 2013). The signals consisted of telegraph noise that were shown to be associated with the opening and closing of the polymerase as nucleotides were incorporated. Importantly, the characteristics of the

noise reflected the specific nucleotide that was being incorporated, opening the way to electrical single-molecule readout of DNA sequences.

In Olsen, Choi et al. 2013, fluctuations of the protein were detected indirectly via the electric field fluctuations they generate, the field fluctuations being sensed by a field effect transistor channel in close proximity to the polymerase or lysozyme. **FIG. 1** illustrates the proposal of Chen et al. A polymerase **1** bound with a primed DNA template **2** in the presence of nucleotide triphosphates in solution **3** is bound by means of antibodies **4** to gold beads **5** that span the gap between the source **6** and drain **7** of a field effect transistor, the channel of which **8** is formed over a gate electrode **9**. In view of the retraction of the original report, it is not clear that this invention actually worked, but it contains many of the elements of the successful invention of the Collins group. This is illustrated in **FIG. 2A** where the source **21** and drain **22** of a field effect transistor are joined by a carbon nanotube **23** that forms the channel of the transistor. An enzyme (lysozyme in this case) **24** is attached to the carbon nanotube. A semiconductor back-gate **25** is used to set the transistor to its most sensitive operating point, midway between turn-on and turn off, and enzyme activity is detected via fluctuations in the FET current. In a later paper, the same group showed (**FIG. 2B**) how a polymerase **26** attached to the carbon nanotube channel **23** and bound by a primed DNA template **27** generated noise spikes, each one of which was associated with the incorporation of a nucleotide by the polymerase. Two examples of the train of signals obtained are given in **FIG. 2C** for a poly(dA) **28** and a poly(dC) **29**. Clear differences in the signals show that sequencing is possible though the noise background **30** from the CNT FET is significant compared to the signal level **31**. A rather similar method has been proposed by Merriman and Mola (Merriman and Mola, 2016). This is illustrated in **FIG. 2C**. A polymerase **436** is chemically linked via a linker **437** to a molecular wire **433** that is connected to the source **438** and drain **439** of a field effect transistor. A gate electrode **440** is placed below the molecular wire. The data presented in their patent application seems to indicate even lower signal levels than those obtained in the device of Olsen et al.

Clearly, it would be desirable to make a more direct electrical connection to the enzyme under test. We have developed a technology called recognition tunneling and have used recognition molecules to bind a protein to at least one of a pair of closely spaced electrodes (Zhang, Song et al. 2017). This approach is illustrated in **FIG. 3**. The device consists of a first electrode **41** and a second electrode **42** separated by a thin dielectric layer **43**. Recognition molecules **44** are strongly attached to each of the electrodes by, for

example, thiol-metal bonds. The molecules are chosen to be specifically recognized and bound by a target protein **45**. These recognition events are generally reversible, and so unsuitable for holding a protein of interest in the gap for studies of its function. Thus, our previous technology of recognition tunneling is inapplicable to the present problem and irrelevant to it, as here we need to keep a known protein connected to the measuring device, rather than use the measuring device to detect the arrival of an unknown protein. More limiting still, is the requirement that these devices be operated at a high enough bias (V_b , **46**) such that the bias itself drives the protein into a mode where telegraph noise fluctuations in current (**47**) are generated. In the case of this published work, where the protein was integrin and the recognition molecule was an RGD peptide, protein binding is detected by operating the device at a relatively high bias ($>100\text{mV}$) and observing the fluctuations induced by the applied bias. The use of voltage-induced fluctuations as a detector of protein binding entirely precludes the use of this device to measure fluctuations in a protein's structure and conformation that enable the protein's critical biomolecular function, because these voltage-induced fluctuations occur in all proteins exposed to a high enough potential difference, regardless of functional motions. Therefore, in the art previously discussed, signals from these critical functional motions cannot be distinguished from voltage-induced fluctuations. Accordingly, it is desired to find a system and method for detecting the binding of a protein molecule across a pair of electrodes in conditions that eliminate voltage-induced conductance fluctuations and keep the protein in place while it is exposed to chemical stimuli that generate key and measurable functional fluctuations and to measure the response of those fluctuations to other chemicals, such as candidate drugs or biopolymers.

Citation of any reference in this section is not to be construed as an admission that such reference is prior art to the present disclosure.

SUMMARY

The present disclosure relates to devices, systems and methods for direct electrical measurement of protein activity. In some embodiments, a device is provided, the device comprising: a first and a second electrode, the first and second electrode being separated by a gap; a protein attached to one or both electrodes; wherein the first electrode and the second electrode are configured for contact with a sample to be analyzed.

In some embodiments, the protein is attached to one electrode. In other embodiments, the protein is attached to two electrodes.

In some embodiments, the device further comprises an insulating dielectric layer disposed within the gap.

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

In some embodiments, the protein is attached to one electrode. In other embodiments, the protein is attached to both electrodes.

In some embodiments, the protein is attached to the electrode via a linker.

In some embodiments, the device comprises: a first and a second electrode, the first and second electrode being separated by a gap; a protein attached to one or both electrodes; wherein a current fluctuation is produced when the protein interacts with a chemical entity.

In some embodiments, a device is provided, the device comprising:

- (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;
- (f) a protein 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.

In some embodiments, the device comprises: a first and a second electrode, the first and second electrode being separated by a gap; a protein attached to one or both electrodes; wherein the first electrode and the second electrode have an opening formed therethrough.

In some embodiments, the device comprises: a first and a second electrode, the first and second electrode being co-planar and separated by a gap; a protein attached to one or both electrodes; wherein the first electrode and the second electrodes are configured for contact with a sample to be analyzed.

In some embodiments, the device further comprises an insulating dielectric layer disposed within the gap.

In some embodiments, the protein is attached to one electrode. In some aspects of this embodiment, the protein is a polymerase. In some aspects of this embodiment, the polymerase is attached to the electrode via a linker. In some aspects, the polymerase is a biotinylated polymerase. In some aspects, the polymerase is a biotinylated polymerase and is attached to the electrode via streptavidin.

In some embodiments of the device, the first and/or second electrode comprise a metal selected from the group consisting of gold, platinum, palladium, and ruthenium. In some embodiments, the metal is palladium.

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 2.0 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. In some embodiments, the gap has a width of about 5.0 nm to about 6.0 nm.

In some embodiments, the device can be used to detect a single molecule.

In some embodiments, a system is provided, the system comprising a device as described herein; a means for introducing a chemical entity that is capable of interacting with the protein; a means for applying a bias between the first and second electrode of value; and a means for monitoring fluctuations that occur as a chemical entity interacts with the protein.

In some embodiments, the bias is between 1 mV and 50mV.

In some embodiments, the bias is between 1 mV and 100mV.

In some embodiments, a method is provided, the method comprising (a) providing a system as described herein; (b) contacting the protein with a chemical entity; (c) applying a bias between the first and second electrode of value such that spontaneous fluctuations of the current between the electrodes do not occur; (d) detecting fluctuations that occur as the chemical entity interacts with the protein.

The methods of the disclosure can be used to detect the activity of a single protein molecule. The methods can also be used to sequence a biopolymer. The methods can also be used in drug screening assays. Advantageously, the methods require no labels or special chemistries.

The methods of sequencing a biopolymer provide for long reads (>10kB), and polymerase runs at the speed of native polymerase (100nt/s).

The devices of the disclosure have simple device geometries, which allows for easy scale up.

The present disclosure relates to an array, system and method for sequencing a biopolymer by direct electrical measurements on single processive protein.

In one embodiment, the present disclosure provides an array for sequencing a biopolymer comprising: an arrangement of a plurality of devices, as described herein. In one aspect of this embodiment, the array is for sequencing DNA.

The present disclosure provides a system for direct measurement of protein activity. The system comprises: (a) an array as described herein; (b) optionally a means for introducing and removing a solution to the array; (c) a means for applying a bias between the first and second electrode; and (d) a means for monitoring the current generated between the first and second electrodes. In one aspect of this embodiment, the system is for direct measurement of polymerase activity.

The present disclosure also provides a method for sequencing a biopolymer. In one embodiment, the method is for sequencing DNA, the method comprises: (a) introducing a solution comprising a DNA template to a system as described herein; (b) measuring a first current generated when a bias is applied to a system as described herein; (b) introducing a solution comprising a dNTP to the system under conditions that allow for incorporation of the dNTP complementary to the DNA template; (c) measuring a second current generated in step (b); (d) removing the solution comprising unincorporated dNTP; (e) repeating steps (b) through (d) with each of the remaining three types of dNTPs not used in step (b); (f) repeating steps (b) through (e); wherein the DNA is sequenced from the generated current signals.

In another embodiment, the method comprises: (a) introducing a solution comprising a DNA template to a system as described herein; (b) measuring a first current generated

when a bias is applied to a system as described herein; (b) introducing a solution comprising at least two types of dNTPs to the system under conditions that allow for incorporation of the dNTP complementary to the DNA template, wherein the types of dNTPs are present in the solution at different concentrations; (c) measuring a second current generated in step (b); (d) removing the solution comprising the unincorporated dNTPs; (e) repeating steps (b) through (d) with the remaining types of dNTPs not used in step (b); (f) repeating steps (b) through (e); wherein the DNA is sequenced from the generated current signals.

BRIEF DESCRIPTION OF FIGURES

FIG. 1 shows a known detection system for polymerase fluctuations according to Chen et al.

FIG. 2A shows a known detection system for Lysozyme fluctuations according to Choi et al. **FIG. 2B** shows a known detection system for polymerase fluctuations according to Olsen et al. **FIG. 2C** shows nucleotide sequence dependent data according to Olson et al. **FIG 2D** shows a known detection system for polymerase fluctuations according to Merriman and Mola.

FIG. 3 shows a known device for protein detection.

FIG. 4 shows a schematic diagram of an embodiment of the disclosure.

FIG. 5A shows the linear current voltage characteristic obtained when a protein is bound at two points as shown in **FIG. 4**. **FIG. 5B** shows the distribution of slopes of the linear region for a large number of single molecule measurements.

FIG. 6 shows signals representative of protein fluctuations.

FIG. 7 shows a schematic diagram of an embodiment of the disclosure in which the protein is a polymerase.

FIG. 8A shows the principle of the fluorescence activity assay used to measure polymerase activity. FAM (F) fluorescence is quenched unless polymerase activity is present. **FIG. 8B** shows the fluorescent intensity of FAM on the substrate after 60-minute incubation with wild type, exonuclease-free (D12A, E14A) or exonuclease-free enzyme attached to streptavidin.

FIG. 9 shows distribution of conductances measured for (left) thiolated streptavidin and (right) after attachment of biotinylated phi29 polymerase. The streptavidin measurements were made at a gap of 2.5 nm and the streptavidin plus phi29 measurements were made at a gap of 3.5 nm.

FIG. 10 shows conductance changes with conformation. Left panel is for thio-streptavidin and the right panel is data taken from the same film of molecules after biotin was added. The conductance changes showing that it is sensitive to the conformational changes induced by biotin binding.

FIG. 11 shows high contrast, high time-resolution recordings of protein fluctuations in (left) and STM gap and (right) a solid-state chip. The sample is anti-DNP IgE with DNP on the electrodes. The devices are operated at 200 mV with a gap of 4.6 nm.

FIG. 12 shows current collected at constant gap (as marked lower left on each panel) and 50mV bias on phi29-streptavidin complex showing just contact fluctuations in the absence of template DNA and dNTPs (A,B,C) and the additional telegraph noise that appears when template DNA and dNTPs are added (D,E,F). Insets to right show full runs over 20 to 40s duration. The red circle denotes the high current point expanded in the traces on the left.

FIG. 13 shows distributions of measured conductances (log scale) for three antibodies and integrin bound to peptide ligands (as listed in Table 1) on Pd electrodes for a gap of about 4.5 nm, with the exception of the data for streptavidin coupled by thiol bonds to the electrodes where data was taken at a gap of about 2.5 nm. The distributions are arbitrarily displaced vertically for clarity. The insets illustrate how the antibodies, with two binding sites for peptide attached to the electrodes can bridge the gap, resulting the second, higher peak in conductance (labeled "2"). Single connections can also occur ("1"). These data show how high conductance (about 2 nS) can be obtained over long distances (the 13 nm between the binding sites of the antibodies) if two chemical connections are made between the protein and the electrodes, one to each electrode.

FIG. 14 shows gap distance-dependence of conductance distributions for (left) streptavidin and right, a streptavidin-polymerase complex taken at different gap sizes as marked. The conductance distributions change little with gap size, showing that conduction in these proteins is by a delocalized transport mechanism. Inset shows estimates of the protein heights - the streptavidin is about 4 nm high, and complex of phi29 bound to streptavidin is about 9nm high. Signals were obtained from gaps as big as

5.5 nm showing that the conduction path must be, in part, through the phi29, although the probe is almost certainly in contact with the polymerase interior in these data sets.

FIG. 15 shows attachment of a polymerase via two biotinylated sites on the polymerase, separated by 5 nm, to two streptavidin molecules on the electrodes. The streptavidin molecules are coupled to the electrodes via thiol moieties.

FIG. 16 shows an array of polymerase molecules bound with DNA templates, each polymerase being wired into an individually addressed pair of electrodes.

FIG. 17 shows a sequence of exposure to nucleotide triphosphates and rinses for determining the sequence of each template molecule at each site.

FIG. 18 shows two telegraph noise bursts characteristic of the sequential incorporation of two identical nucleotides at sites containing identical bases

DETAILED DESCRIPTION

The disclosure includes at least the following:

- (1.) A device substantially as shown and described.
- (2.) A system for direct electrical measurement of protein activity as shown and described.
- (3.) A method for detecting protein activity as shown and described.
- (4.) A method of sequencing a biopolymer as shown and described.
- (5.) A device for direct measurement of protein activity, the device comprising a first and a second electrode, the first and second electrode being separated by a gap; and a protein attached to one or both electrodes; wherein the first electrode and the second electrode have an opening formed therethrough.
- (6.) The device of the above (5.), wherein the gap has a width of about 1.0 nm to about 20.0 nm.
- (7.) A device for direct measurement of protein activity, the device comprising:
 - (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;
- (f) a protein 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.

(8.) A device for direct measurement of protein activity, the device comprising:

- (a) a first and a second electrode, the first and second electrode being co-planar and separated by a gap;
- (b) a protein attached to at least one electrode;

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

(9.) The device of any of the above (5.) to (8.), wherein the protein is selected from the group consisting of a polymerase, a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease.

(10.) The device of the above (9.), wherein the protein is a polymerase.

(11.) The device of the above (10.), wherein the polymerase is attached to one electrode.

(12.) The device of the above (10.) or (11.), wherein the polymerase is attached to the electrode via a linker.

(13.) The device of the any of the above (10.) to (12.), wherein the protein is a biotinylated polymerase.

(14.) The device of the above (13.), wherein the biotinylated polymerase is attached to the electrode via a thio-streptavidin linker.

(15.) The device of the any of the above (5.) to (14.), wherein the first and/or second electrode comprise a metal selected from the group consisting of gold, platinum, palladium, and ruthenium.

(16.) The device of the above (15.), wherein the first and/or second electrode comprise palladium.

(17.) A device for direct measurement of protein activity, the device comprising:

- (a) a dielectric substrate;
- (b) a first electrode disposed on the dielectric substrate;
- (c) a second electrode disposed on the dielectric substrate, wherein the first and second electrode being separated by a gap between 1 and 10 nm;
- (d) a passivation layer disposed on top of the electrodes; and
- (e) a protein attached to one or both the electrodes;

wherein the passivation layer has an opening formed therethrough positioned to allow a sample to pass to the gap between the first and second electrode.

(18.) A system for direct electrical measurement of protein activity comprising

- (a) a device of any of the above (5.) to (17.);
- (b) a means for introducing a chemical entity that is capable of interacting with the protein;
- (c) a means for applying a bias between the first and second electrode; and
- (d) a means for monitoring fluctuations that occur as the chemical entity interacts with the protein.

(19.) The system of the above (18.), wherein the protein is a polymerase.

(20.) The system of the above (18.), wherein the protein is an exonuclease, proteasome, or glycan.

(21.) The system of the above (18.), wherein the protein is a kinase.

(22.) A method of detecting the activity of a single protein molecule, the method comprising

(a) introducing a chemical entity that is capable of interacting with the protein molecule to the system of the above (18.);

(b) applying a bias between the two electrodes chosen so that a steady DC current is observed; and

(c) observing fluctuations in current between the two electrodes that arise when the chemical entity interacts with the protein.

(23.) The method of the above (22.), wherein the protein is selected from the group consisting of a polymerase, a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease.

(24.) The method of the above (22.) or (23.), wherein the chemical entity is selected from the group consisting of a nucleotide triphosphate, a nucleic acid, a peptide, a glycan and a kinase.

(25.) A method of sequencing DNA, said method comprising

(a) introducing a primed DNA template to the system of the above (18.);

(b) introducing a solution comprising the four dNTPs;

(c) applying a bias between the two electrodes chosen so that a steady DC current is observed;

(d) detecting fluctuations in current between the two electrodes that arise when each new nucleotide is incorporated to the primer; and

(e) determining the identity of each of nucleotides being incorporated.

(26.) The method of the above (25.), wherein the solution comprises the four dNTPs at about the same concentration relative to each other.

(27.) The method of the above (25.) or (26.), wherein the concentrations of the dNTPs are about equal to or above the saturation concentration of the template-bound polymerase.

(28.) The method of any of the above (25.) through (27.), wherein step (d) comprises detecting the presence of one or more current spike(s).

(29.) The method of any of the above (25.) through (27.), wherein step (e) comprises using the characteristics of each spike.

(30.) A method of sequencing a biopolymer, said method comprising

- (a) introducing a biopolymer to the system of the above (18.);
- (b) applying a bias between the two electrodes chosen so that a steady DC current is observed;
- (c) detecting fluctuations in current between the two electrodes that arise when a monomer is removed from the end of the biopolymer; and
- (d) determining the identity of each monomer removed from the biopolymer.

(31.) The method of the above (30.), wherein the biopolymer is DNA, a peptide, or a glycan.

(32.) A method of detecting the activity of kinase, the method comprising

- (a) introducing a candidate kinase inhibitor molecule to the system of the above (20.);
- (b) applying a bias between the two electrodes chosen so that a steady DC current is observed;
- (c) detecting fluctuations in current between the two electrodes that arise when the kinase interacts with the candidate kinase inhibitor molecule; and
- (d) determining whether the kinase has activity in the presence of the candidate kinase inhibitor molecule.

(33.) An array for sequencing a biopolymer as herein described.

(34.) An array for sequencing DNA as herein described.

(35.) An array for sequencing DNA comprising:

an arrangement of a plurality of devices, wherein each 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 molecule attached to the first and second electrode,

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

(36.) An array for sequencing DNA comprising:

an arrangement of a plurality of devices, wherein the device comprises:

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

wherein the passivation layer has an opening formed therethrough.

(37.) An array for sequencing DNA comprising:

an arrangement of a plurality of devices, wherein the device comprises:

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

- (b) a polymerase attached to at least one electrode;

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

- (38.) The array of any of the above (33.) to (37.), wherein the arrangement is a grid.

- (39.) A system for direct measurement of polymerase activity comprising:

- (a) an array as described herein;
- (b) optionally a means for introducing and removing a solution to the array;
- (c) a means for applying a bias between the first and second electrode; and
- (d) a means for monitoring the current generated between the first and second electrodes.

- (40.) A method for sequencing DNA, the method comprising:

- (a) introducing a solution comprising a DNA template to a system as described herein;
- (b) measuring a first current generated when a bias is applied to a system as described herein;
- (c) introducing a solution comprising a dNTP to the system under conditions that allow for incorporation of the dNTP complementary to the DNA template;
- (d) measuring a second current generated in step (c);
- (e) removing the solution comprising unincorporated dNTP;
- (f) repeating steps (c) through (e) with each of the remaining three types of dNTPs not used in step (c); and
- (g) repeating steps (c) through (f);

wherein the DNA is sequenced from the generated current signals.

(41.) A method for sequencing DNA, the method comprising:

(a) introducing a solution comprising a DNA template to a system as described herein;

(b) measuring a first current generated when a bias is applied to a system as described herein;

(c) introducing a solution comprising at least two types of dNTPs to the system under conditions that allow for incorporation of the dNTP complementary to the DNA template, wherein the types of dNTPs are present in the solution at different concentrations;

(d) measuring a second current generated in step (b);

(e) removing the solution comprising the unincorporated dNTPs;

(f) repeating steps (c) through (e) with the remaining types of dNTPs not used in step (c); and

(g) repeating steps (c) through (f);

wherein the DNA is sequenced from the generated current signals.

(42.) The method of the above (41.), wherein the solution in step (c) comprises four types of dNTPs.

(43.) The method of the above (41.), wherein the solution in step (c) comprises at least two types of dNTPs.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as those commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. The materials, methods and examples are illustrative only, and are not intended to be limiting. All publications, patents and other documents mentioned herein are incorporated by reference in their entirety.

Throughout this specification, the word “comprise” or variations such as “comprises” or “comprising” will be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers.

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

The terms “and” and “or” may refer to either the conjunctive or disjunctive and mean “and/or”.

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.

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.

DEVICE AND SYSTEM FOR DIRECT MEASUREMENT OF PROTEIN ACTIVITY

The present disclosure provides a device for direct measurement of protein activity. In one embodiment, the device comprises a first and a second electrode, the first and second electrode being separated by a gap; and a protein attached to one or both electrodes; wherein the first electrode and the second electrode have an opening formed therethrough.

In another embodiment, the device comprises a first and a second electrode, the first and second electrode being separated by a gap; and a protein attached to one or both electrodes.

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;
- (f) a protein 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.

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 protein attached to one or both the electrodes;

wherein the passivation layer has an opening formed therethrough.

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 protein attached to at least one electrode;

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

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.

In each of the device embodiments described herein, the protein is selected from the group consisting of a polymerase, a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease.

The protein can be attached to one electrode directly or indirectly. In some embodiments, the protein is attached to the electrode via a linker. In some embodiments, the protein is

attached to the electrode indirectly via interactions with a ligand attached to the electrode. In some embodiments, the protein is modified to incorporate a ligand-binding site.

In one embodiment, the device comprises: a first and a second electrode, the first and second electrode being separated by a gap; a polymerase attached to one or both electrodes; wherein the first electrode and the second electrode have an opening formed therethrough.

In some embodiments, the polymerase is attached to one electrode. In some aspects of this embodiment, the polymerase is attached to the electrode via a linker. In some aspects, the polymerase is a biotinylated polymerase. In some aspects, the polymerase is a biotinylated polymerase and is attached to the electrode via streptavidin.

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. In some embodiments, the metal is palladium.

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.

In some embodiments, the device can be used to detect a single molecule.

The present disclosure also provides a system for direct measurement of protein activity. The system comprises a device as described herein; a means for introducing a chemical entity that is capable of interacting with the protein; a means for applying a bias between the first and second electrode; and a means for monitoring the current generated between the first and second electrodes as the chemical entity interacts with the protein.

In each of the system embodiments described herein, the protein is selected from the group consisting of a polymerase, a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease. In one embodiment, the protein is a polymerase.

When the protein is a polymerase, the polymerase is attached to one electrode, or preferably both electrodes. In some aspects of this embodiment, the polymerase is attached to the electrodes via one or more linkers. In some aspects, the polymerase is a

biotinylated polymerase. In some aspects, the polymerase is a biotinylated polymerase and is attached to the electrode via streptavidin.

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

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.

FIG. 4 shows a schematic diagram of a system according to an embodiment of the disclosure. A protein molecule **51** is covalently modified at certain sites **53** and **54**. Such sites can be surface cysteine residues modified by reaction with a maleimide, lysine residues modified by means of an NHS ester or the insertion of a histidine tag at the N or C terminus of the protein to bind nitrilotriacetic acid, or by biotinylation of the protein and attachment via thiolated streptavidin molecules attached to the electrodes. Other means of attachment such as Myc tags or GST tags may be used as is well known in the art. The critical and unique design aspects of this embodiment are that the protein itself is utilized as the detector, to which end strong and permanent chemical tethers are used to attach the protein to the electrodes. The modified sites are coupled to flexible linkers, which may be short (1 to 10 repeats) alkane oligomers or polyethylene glycol oligomers or short peptide chains incorporated into the protein recombinantly. These are in turn terminated by reactive groups **56** that tether the linker molecules to the metal electrodes **57** and **58**. While a variety of linkages are possible, thiol linkages are preferred, and amines may also be used, as can biotin-streptavidin linkages. A bias **59** is applied between the two electrodes and the current passing between the electrodes is monitored **60**. The system is immersed in a buffer solution containing ions necessary for enzyme function and the effects of introducing chemical entity **61** (such as a substrate for the enzyme, and/or a drug) is recorded.

The basis of the present disclosure lies in a remarkably unexpected and very recent observation about the behavior of a protein in a large (approximately 4.5 nm) gap when the protein is strongly tethered to two electrodes as described above. We find that below the critical bias voltage previously reported for the onset of telegraph noise signals, a

simple linear (Ohmic) response is found. This is completely unexpected because proteins are believed to be molecular solids in which the mode of electron transport should only be tunneling. However, tunneling cannot account for the large currents with linear current-voltages observed when proteins are tethered in the manner described in **FIG. 4**. An example of a typical current-voltage curve measured on a single protein molecule is shown in **FIG. 5**. Large noise fluctuations are observed above about ± 100 mV as previously reported, but below ± 100 mV (boxed area labeled **71** in **FIG. 5A**) there is a linear region which implies a remarkably high DC conductance for the protein, even over this large (4.5 nm) distance. When this linear region is fitted and a distribution of fitted conductances obtained, the conductance distribution can be fitted by an exponential distribution as shown in **FIG. 5B** (solid line is the fit). Note that in this case the mean conductance, K , has a value of 1.5 nS.

A larger collection of measurements reveals a more complex distribution of conductances as shown in **FIG 13**. This figure plots histograms of the frequency of a given conductance versus conductance for a series of proteins coupled to the electrodes by various means. The conductance scale is logarithmic (base 10) and the peaks that are fitted as shown by the lines are Gaussian, so these conductances are distributed according to a log-normal distribution. (The distributions have been arbitrarily displaced vertically for clarity.) The proteins are tethered to the electrodes either by binding to a specific ligand for a particular protein, or, in the case of streptavidin via thiol linkages. The ligands were chemically attached to the electrodes via thiol (cysteine) linkages. The various proteins, their ligands, the dissociation constant for the protein-ligand complex, controls used to verify specific binding and the peaks values of conductance obtained by fitting the distributions are listed in Table 1.

Protein	Ligand	K_D	Control	Peak Conductance (nS)
IgE Anti-DNP	Thiolated-dinitrophenol	65 nM	IgE isotype	0.266, 1.95
IgG Anti-HIV	CHNTPVYKLDISEATQV	240 nM	IgG isotype	0.334, 2.21
IgG Anti-Ebola	CALDRWEKIRLR	1400 nM	IgG isotype	0.260, 2.09
$\alpha_v\beta_3$ Integrin	Cyclic RGD-C	~ 10 nM	$\alpha_4\beta_1$ Integrin	0.341
Thio-Streptavidin	NA	NA	Add Biotin	0.336

Table 1: Proteins used in the conductance study

No conductance was observed when electrodes were exposed to the control molecules listed, showing that specific chemical tethering of the protein to the electrodes is required for electronic conductance to be observed.

In the case of the three antibodies, two binding sites are available, one at each of the two binding domains, separated by 13 nm. As a consequence, a second, higher conductance peak is observed in the distributions for these three molecules. This yields the second peak conductance listed for these molecules in Table 1. Consequently, high conductance can be obtained over long distances (13 nm) if proteins are chemically tethered to both electrodes.

This threshold for the onset of voltage-driven fluctuations of about 100mV has been found for a number of proteins studied to date. Thus, by operating the junction below this threshold for the onset of spontaneous fluctuations (i.e., $V < V_C$ in **FIG. 4**, where V_C is about 100mV) a DC current serves to indicate that a protein is trapped in its quiescent state. This signal is quite large: on average 75 pA at 50 mV bias for the example just given, and substantially more if the protein is chemically tethered so as to bridge the electrode gap.

Protein fluctuations open up additional channels for electron transport. Thus, when a protein is biased below V_C but stimulated by introducing a substrate molecule, large current fluctuations can occur. An example of the current signals induced by protein fluctuations is shown in **FIG. 6**. Note the greatly improved signal **82** to noise **81** compared to that shown in **FIG. 2C**. A further example of induced protein fluctuations is given in **FIG. 12**. This shows data collected by an STM probe held at a constant 3.5 or 4.5 nm height above a monolayer of phi29 polymerase coupled to a palladium electrode via a biotinylated N-terminus binding thiolated streptavidin molecules bound on the electrode surface. Panels A, B and C show the fluctuations in current that occur as a result of fluctuations in the contact point with the polymerase. Over time (insets to right) this can lead to large changes in current but ms-timescale telegraph noise is not observed if the bias is below V_C . The expanded current-time traces on the left are taken from the peak current regions (circled in red on the long-time scale plots inset on the right). When primed, single-stranded DNA template and dNTPs are added, ms-timescale telegraph noise is induced, as shown in panels D, E and F.

The signals shown in **FIG. 12** were taken with a single chemical attachment point for the polymerase. In consequence, the connection to the second electrode is highly variable, as

shown by the current fluctuations in **FIGS. 12 A, B and C**. Another important drawback of a single chemical contact is the poor performance of the physical contact between the protein and the second electrode. This is illustrated in **FIG. 14**. This shows conductance distributions measured over a streptavidin monolayer (left side) and over the same monolayer after binding biotinylated phi29 polymerase (right side) as the electrode gap is increased in 1 nm increments for gap values as marked (the distribution curves are displace vertically for clarity). For the case of streptavidin, very few curves are recorded at a gap distance of 3.5 nm. For the case where phi29 is bound to the streptavidin, curves are recorded out to 4.5 nm (with a few recordings at 5.5 nm - not shown). However, this is substantially smaller than the overall height of the polymerase-streptavidin complex of about 9 nm (shown in the inset). This stands in sharp contrast to the antibody data shown in **FIG. 13** where high conductances were obtained over the 13 nm path that separates the two binding domains. This demonstrates the desirability of forming two chemically well-defined contacts, one to each electrode in the pair.

Methods of Making a Device of the Disclosure

A device of the disclosure can be readily fabricated by depositing a layer of a noble metal such as Au, Pt or Pd onto a silicon, glass or sapphire wafer (or other dielectric substrate), then depositing a thin (typically 1 nm) layer of a reactive metal for adhesion (such as chrome or titanium), and then a layer of 1 to 10, or 1 to 20 or 1 to 50 nm of the noble metal. This bottom electrode is then covered with an insulating dielectric layer, preferably alumina, though other oxides such as SiO₂ or hafnium oxide can be used. The layer should be between 2 and 10 nm in thickness. A 2 nm layer can be deposited by coating the bottom noble metal electrode with 1 to 1.5 nm of aluminum, and allowing it to oxidize in air, thereby producing a 2 to 3 nm thick layer of Al₂O₃. If a greater thickness of dielectric is required, further Al₂O₃ can be added by atomic layer deposition with water/trimethylaluminum cycles as is well known in the art.

A second noble metal electrode is then deposited, again using a thin adhesion layer (chrome or titanium) but of a maximum of 1nm so as not to alter significantly the gap presented at the edge of the device where this adhesion layer will oxidize.

Finally, a passivation layer is placed on top of the top electrode. This can be alumina, SiO₂, hafnium oxide or a resist material such as PMMA or SU8.

In order to make a cavity small enough to ensure that the exposed electrode area is such that only one polymerase is attached, a small opening is then made using Reactive Ion Etching (RIE) as is well known in the art. This opening may be between 10 and 500 nm in diameter with about 50 nm preferred. The depth of the opening should be large enough so that it cuts through the passivation layer, the top electrode, the dielectric layer separating the electrodes, and into the bottom electrode.

A second way to limit the amount of exposed electrode area is to make one of the electrodes (top or bottom, with top preferred) a thin wire of 50 to 100 nm in width. The RIE opening can then be much larger (e.g., micron sized permitting conventional lithography) because the exposed electrode will be limited by the small width of the electrode.

A third way is to control the functionalization chemistry by controlling the amount of time that the junction is exposed to polymerase molecules and/or the concentration of polymerase. The loading of each junction can be tested by monitoring the telegraph noise that is induced by contact fluctuations when the applied bias is above 100mV. The presence of 2 signal levels indicates that just a single molecule is trapped. The presence of three levels indicates that two molecules are trapped and so on. In this way the concentration and exposure time can be adjusted an ideal Poisson loading wherein about 30% of the sites are singly occupied.

After cleaning with an oxygen plasma, the exposed area of the electrodes in the opening can be functionalized with protein. This may be either directly, using the native thiols on the surface of the protein, or via chemical modifications that attach sulfhydryl groups to the protein, or indirectly, by attaching a thiolated streptavidin and then capturing a biotinylated protein.

A second approach to making the device is to form the two electrodes in the same plane with a small gap between them. This can be done by opening a trench across a single wire using e-beam lithography and lift-off or reactive ion etching (RIE) as is well known in the art. Other approaches are to use helium ion milling, or angled deposition of metal over a step edge so that a gap is naturally formed. The electrode pair are then covered with a passivation layer and an opening formed by RIE such that the electrode gap is exposed. The electrodes can then be functionalized as described above.

The width of the exposed electrodes is important, because the devices described here generally rely on connecting to just one molecule. In the case of extracting sequence information, this requirement of a single molecule signal is particularly important. If the electrodes are not much wider than a single molecule (5 to 15 nm) then attachment of multiple molecules across the gap is not possible. However, reliable functionalization and fabrication of such small electrodes is very difficult. In practice, we have found that electrodes of up to 100 nm width are unlikely to capture more than one protein molecule. In particular, when the probability of binding in the desired (bridging) configuration is small, it may even be desirable to have even wider electrodes, of 200, 300, 400, 500 or even 1000 nm width. Protein molecules that are bound to just one electrode, rather than bridging the pair of electrodes, will contribute relatively small amounts of current.

Methods of Attaching a Polymerase to the Electrodes

When the protein is a polymerase, it should be a polymerase with high processivity, such as the phi29 polymerase, and its exonuclease function should be disabled as described below.

The wild-type (WT) polymerase requires modification to (a) remove its exonuclease activity and (b) add a chemical attachment point if so desired. This modification is achieved by recombinant DNA using an E. Coli expression system to produce the modified polymerase.

Exonuclease activity of phi29 requires the following acidic amino acids: D12, E14, D66 and D169. Mutating any one of these will eliminate the exonuclease activity, and we have mutated D12 and E14 to alanine.

The clone we used has both the his-tag and the avitag (for biotinylation) at the N-terminus, ie His-Avitag-Phi29. As a result, the following sequence was added to the N-terminus of the enzyme:

MGSSHHHHHHSSGLVPRGSLNDIFEAQKIEWHEGASS.

The six histidine residues are the his tag (used for purification of the desired enzyme product) and the GLNDIFEAQKIEWHE is the Avitag. The biotin is attached to the K in the avitag by the biotin ligase BirA (Avidity, Lansing, MI). Activity assays show that the biotinylated enzyme attached to streptavidin is still active (**FIG. 8B**).

Another useful and unexpected feature of the present disclosure is that both streptavidin and polymerase are conductive proteins, so, as we show below, the polymerase can be attached to the electrode indirectly. First a streptavidin, modified with thiols, is attached to the electrode, and then biotinylated polymerase introduced. This binds to the streptavidin, providing a conductive path to the electrode.

The polymerase can also be modified at the C-terminus by the same recombinant methods. In addition to the avitag, other peptide-based binding tags can be used such as GST tags, Myc tags and His tags. These tags can all be incorporated at either the N- or C-terminus of the polymerase. Incorporation at the C terminus places the tag site close to the site at which a primed template is captured, so oligoalanine or glycine-glycine-serine spacer sequences can be incorporated between the C terminus and the tag to reduce interference with the template capture activity of the polymerase.

The same technology can also be used to attach other proteins whose activity is to be monitored using the methods of the present disclosure (such as kinases, proteases or molecules that process glycans).

In addition to modification at the N- or C- termini, there are seven cysteines in phi29. None are disulfide bonded, so all have the potential for forming disulfide bonds, offering additional sites for attachment to electrodes. Based on the structure of phi29 with template and primer C448, C106 and C22 are most surface exposed and look like good candidates for either biotinylation through maleimide or direct attachment to heavy metals that bind sulfhydryl groups. The problem is how to control specificity. We tried to mutate out all but one cysteine once, but the result is insoluble protein. However, up to four may be removed without affecting solubility, leaving C448, C106 and C22 as targets for attachment points.

Although the present disclosure works with just one chemical attachment site to one electrode, the second contact being made by physical contact between the protein and the metal, it is desirable to make two chemically well-defined contacts in a manner that spans the gap between the two electrodes. The C terminus is separated from the N terminus by a distance of ~ 5nm, and if biotinylation via an avitag is used, attachment to the same streptavidin molecule by both the N- and C- termini of the same polymerase is improbable. Thus, with both electrodes functionalized with thio-streptavidin, there is an opportunity for bridging structures to form in which the N terminus is connected to one electrode and the C terminus to the second electrode.

Another approach is to use two attachment points that are widely spaced, but in an inactive region of the protein. In the case of phi29 polymerase, where the exonuclease domain has been disabled by mutations of D12 and E14, two points in the exonuclease domain spaced by > 5nm are found between G111 and K112, and between E279 and D280. Accordingly, with the Avitag sequence GLNDIFEAQKIEWHE inserted at these two points, and the Avitag lysine biotinylated by BirA, the polymerase can be bound across a pair of electrodes as shown in **FIG. 15**. This shows a phi29 polymerase **1501** that is biotinylated using Avitags place between G111 and K112, and between E279 and D280 (**1502** and **1503**). These biotins bind thiolated streptavidin molecules **1504** and **1505** that are attached via sulfur linkages **1506** and **1507** to the electrodes **1508** and **1509**. Since gap between these attachment points (double headed arrow on **FIG. 15**) is a little over 5nm, this attachment geometry requires a gap **1510** of a little under 5 nm.

Making deterministic contacts between polymerases and electrodes by adding 'conducting whiskers' to the polymerase.

At the N terminus add recombinantly:

CGSSHHHHHSSFTLIELLIVVAIIIGILAAIAIPQFSAYRVKAYNSAASSDRLNLKTA
LESADFADDQTYPPESGLVPRGSGASS-f29

The terminal C is the cysteine for attachment to the first electrode.

The his tag is for protein extraction and purification.

The 61-amino acid sequence following SS is the sequence of the pilus protein from geobacter sulferreductans, which acts as a metallic wire.

At the C terminus add recombinantly:

f29-
AAFTLIELLIVVAIIIGILAAIAIPQFSAYRVKAYNSAASSDRLNLKTALESADFADDQT
YPPEC

The 61-amino acid sequence following AA is the sequence of the pilus protein from geobacter sulferreductans.

The terminal C is the cysteine for attachment to the first electrode.

Conductivity of the complex

Key to the present disclosure is that good electronic conductance can be obtained through the polymerase. **FIG. 9** shows (left panel) a distribution of conductances through streptavidin connected to the substrate via thiol bonds. These data were taken with an electrode gap of 2.5 nm. If the gap is increased to 3.5 nm, very few reads occur. If the gap is further increased to 4.5 nm, no reads at all are obtained. This is consistent with the fact that a streptavidin molecule, lying flat on the surface, is about 4 nm high. The right-hand side of **FIG. 9** shows data for a complex of biotinylated phi29 polymerase with streptavidin on a Pd electrode. These data were obtained at a gap size of 3.5 nm. Similar data were recorded at a gap size of 4.5 nm for the complex (a gap where no conductance was recorded for the streptavidin). Since these distances are larger than the gaps at which the streptavidin alone gave robust signals, they show that conduction is occurring through the polymerase wired in series with the streptavidin. Note how the distribution of conductances is altered in the complex compared to streptavidin alone.

Conductivity changes with conformation

FIG. 10 shows a measured distribution of conductances for streptavidin alone (left panel) and streptavidin after exposure to biotin. The distribution of measured conductances changes after the streptavidin binds biotin, showing that the conductance of the protein is sensitive to changes in protein conformation.

Dynamic monitoring of protein conformation

FIG. 11 shows a recording of voltage-induced conformational fluctuations recorded (left panel) with a scanning tunneling microscope and (right panel) with a solid-state chip using a gap of 4.5 nm. The bias was 200mV, 100mV above VC. The protein is an anti-dinitrophenol molecule binding dinitrophenol (DNP) attached to the electrodes. Very large changes in conductance on the millisecond timescale are readily recorded with excellent signal to noise ratio. **FIGS. 12D, E and F** show the fluctuations in current that occur when the bias is reduced below VC (to 50mV in this case) and the protein activated by addition of substrate. Telegraph noise is generated on ms timescales with very high signal-to-noise ratio.

METHODS OF USE OF THE DEVICES AND SYSTEMS

The present disclosure provides methods for sequencing a biopolymer. We illustrate this for the specific case of a nucleic acid chain being extended by a polymerase in FIG. 7. Here, a polymerase 96, is modified at two points 53, 52. A specific example would be the two surface-exposed cysteines available in phi29 polymerase. Another example is the biotin functionalization shown in FIG. 15. Maleimide modified alkanes or PEG linkers 54, 55, are used to attach the polymerase to the electrodes 57, 58 via strong thiol bonds (for example) 56. The binding of the polymerase in the electrode gap is verified by means of a steady DC current 60 when $V < V_C$ 59. When the polymerase 96 is complexed with a primed template 91 and exposed to a solution comprising each of the four nucleotide triphosphates 92 93 94 95, characteristic current fluctuations will signal the incorporation of a given nucleotide. In an alternative embodiment, surface cysteines could be used to make one of the attachments directly.

In use, once the device is prepared, it should be rinsed and then exposed to the primed template DNA to be sequenced. This DNA is prepared from the sample to be sequenced by ligating hairpin primers as well known in the art. A buffer solution comprising the four dNTPs and Mg^{2+} should be introduced to the device. The dNTPs are present in the buffer solution in about equal concentration. In one embodiment, the concentrations of dNTPs are about equal to the saturation concentration of template-bound polymerase, at the saturation of concentration of template-bound polymerase in a second embodiment, and above the saturation concentration in third embodiment. When the concentrations of dNTPs are at or above the saturation concentration, the polymerase runs fast (i.e., 100 nucleotide incorporations per second).

When the buffer solution comprising the dNTPs is introduced to the device, a polymerization reaction will be initiated and the captured template will be copied to the primer, producing a series of current spikes. Each spike (or cluster of spikes) occurs as each new nucleotide is incorporated to the primer, and the characteristics of each spike (duration, amplitude, shape) used to decode the identity of the nucleotide being incorporated. A typical sequencing speed at saturation concentration ($>30 \mu M$) of nucleotides is about 100 nucleotides per second. The saturation concentration of template is about 10 nM with a new template incorporated almost immediately after completion of the previous template. Each molecule will continue to turn over template so long as there

are templates available in solution. Therefore, one molecule can sequence continuously for as long as the device is operated.

The device geometry is extremely simple with no need to separate fluidic compartments for each junction, so one junction would only occupy about a micron². Allowing for interconnects, isolation and on-chip processing electronics, a single reading device could readily be fitted into an area of 100 microns by 100 microns, so an active chip area of 1 cm² would accommodate 10,000 devices. A chip with 10,000 junctions on a chip operated for 1 hour would sequence an entire human genome ($10000 \times 3600 \times 100 = 3.6 \times 10^9$). A denser device geometry or a larger chip would accommodate even a significant fraction of inactive devices and still permit genome-scale sequencing on one small device in times of an hour or less.

The preceding example illustrates the sequencing of nucleic acid polymers, but it also can be applied such that other enzymes that process polymers could be used. For example, current fluctuations in an exonuclease will reflect the composition of the nucleic acid they are degrading. The same would be true of proteasomes that digest peptides. An example is the proteasome 20S CP, and proteasomes like this could likely be used for single molecule peptide sequencing by incorporating them into the system of **FIG. 4** and monitoring the electrical signals generated when they are fed peptide molecules and biased below VC.

Similar enzymes, called glycosidases, exist for digesting glycans. The incorporation of a glycosidase into the device of **FIG. 4** would allow electronic sequencing of glycans. The variation in bonding of glycans might preclude a direct linear read out of sequence, but the organization of cutting events in time may allow for identification of the glycans.

In yet another embodiment, the present disclosure provides a method for detecting kinase activity. In this embodiment, a kinase is incorporated into the device of **FIG. 4**, exposed to its substrate, and kinase activity signaled by the generation of large current fluctuations when biased below VC. The system could then be exposed to candidate kinase inhibitor drugs as the fluctuations are monitored, to discover which drugs "kill" the activity of the kinase. In the present art, the use of fluorescence labeling methods may interfere with a protein enzyme interacting with its substrate, and the present disclosure removes the requirement for labeling of proteins.

In all of these methods, the use of a simple junction (as opposed to a FET structure) greatly simplifies both manufacture and enables scale up to large parallel arrays of devices. The device of the disclosure may be prepared in massively parallel fabrication using methods for scalable fabrication of junction devices, as described below.

ARRAYS AND SYSTEMS FOR SEQUENCING DNA OR OTHER POLYMERS

The present disclosure provides an array for sequencing biopolymers using any of the enzymes that interact processively with molecular templates such as a nuclease, a proteasome, a glycopeptidase, a glycosidase, a kinase and an endonuclease. The embodiments below illustrate an array for sequencing DNA using a polymerase. It should be understood that any processive enzyme can be substituted for the polymerase in the arrays.

The array comprises an arrangement of a plurality of devices. The device used in the arrays of the present disclosure include the following.

In one embodiment, 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.

In another embodiment the device comprises a first and a second electrode, the first and second electrode being separated by a gap; and a polymerase attached to both the first and second electrode.

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;
- (f) a polymerase molecule attached to the first and second electrodes;

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

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 dielectric substrate;
- (d) a passivation layer disposed on top of the electrodes; and
- (e) a polymerase molecule attached to one or both the electrodes;

wherein the passivation layer has an opening formed therethrough.

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;
- (b) a protein attached to at least one electrode;

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

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. In some embodiments, the metal is palladium.

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.

The array of devices can be arranged in any suitable manner, e.g., in a grid.

In some embodiments, the array comprises polymerase molecules bound with template DNA. Such templates can be made by ligating genomic DNA fragments (generated by

sonication, for example) with primer sequences containing a nick for binding by polymerase, as is well known in the art. The result is a library of templates spanning an entire genome, if needed. Each template will then randomly bind one polymerase in the array.

Referring now to **FIG. 16**, a grid of contacts **101**, **102** is formed from two layers of contact metals separated by a dielectric as is well known in the art, and then covered by a passivation layer. Each intersection is then exposed by selectively removing the passivation, and polymerase molecules **103** bound at each junction. These biomolecular junctions can be prepared by methods discussed in the previous section. By making the electrodes narrow enough (about one micron in width) usually only one polymerase (or none) will bind. In the case where two or a few more bind, the signals can still be deconvolved because they consist of two current levels for one molecule, three for two, and so on.

The present disclosure also provides a system of arrays for direct measurement of polymerase activity. The system comprises an array as described herein; optionally a means for introducing and removing a solution to the array; a means for applying a bias between the first and second electrode; and a means for monitoring the current generated between the first and second electrodes.

Referring back to **FIG. 16**, means **106** are provided for applying a small bias (typically 50mV) and reading the current from each junction so that it may be recorded by a computer storage system.

METHODS FOR DNA SEQUENCING USING A SYSTEM OF ARRAYS

The present disclosure also provides a method for DNA sequencing using a system of arrays as described herein.

The sequencing proceeds by introducing a solution comprising one nucleotide monophosphate (e.g., one from among dATP, dGTP, dCTP, dTTP) together with magnesium to the array. Each polymerase bound with a complementary nucleotide will generate a signal, which is read by the unique pair of electrodes to which each polymerase is bound. For example, if the added nucleotide is dCTP, then every template bound at a G base will incorporate a C into the extending chain, generating a signal, whereas the other sequences will generate distinctly different signals. For example, a polymerase presented with a non-matching base will generate a brief burst of signal as the mis-matched base is

captured, but the signal train will terminate prematurely as the mismatch is rejected and the process of polymerization and translocation is aborted. In contrast, incorporation of a matching base results in a much longer train of pulses as the process of incorporation and translocation is completed. As the second step, the array is rinsed to remove excess nucleotide and a solution comprising the next nucleotide is introduced (for example dATP, so that all site with a T now generate a signal). The cycle is continued until all four dNTPs have been cycled through the device, after which the cycle is repeated. This cycling can be repeated until all the template DNA is exhausted, thus generating sequence data for the entire library of fragments.

FIG. 17 shows a method for sequencing DNA. On adding dATP to the array **201**, the molecule shown will generate a signal **204** as the A is incorporated into the extending chain. The array is then rinsed to remove dATP **202**, and the next nucleotide (shown as dTTP here) added **203**. The cycle is continued as shown. The addition of dATP, then dCTP and then dATP all give signals of base incorporation **204** whereas the presence of the other nucleotides do not **205**. Accordingly, the sequence at that particular location will be recorded as TGT.

It will be recognized that this approach has two major advantages over current sequencing strategies that use cycling of dNTPs. One is that, by utilizing a single molecule read-out at a time-scale faster than the base-incorporation rate of a polymerase, it now becomes straightforward to count repeats of the same base. So the sequence AAAAAA would give 5 distinct bursts of signal in the presence of dTTP, and so on. The second is that, in contrast to known optical readout schemes, the length of template that is read is not constrained by distance from the mounting substrate, so that the potential read length is as high as the processivity of the polymerase (10kB for phi29).

FIG. 18 shows the ability to read individual sequential incorporations, and hence read homopolymeric runs of sequence. A current vs. time recording **300** shows the incorporation of two sequential bases (here, electrical data are shown for dTTP being incorporated at two successive A sites). Each signal consists of a burst **301** of telegraph noise of duration about 20 ms, separated by gaps between bursts, also, on average of about 10 to 20 ms at micromolar concentrations of nucleotide triphosphate. Thus, on any given step of the chemical cycling illustrated in **FIG. 17**, where just one nucleotide triphosphate is present, the number of repeated bases of its complement in the template strand may be counted simply by recording the number of such bursts of telegraph noise.

In another embodiment, the solution comprises more than one dNTP, with the dNTPs present in different concentrations. For example, the solution comprises 1 mM dATP, 100 μ M dGTP 10 μ M dCTP and 0.1 μ M dTTP. In this embodiment, the polymerases in the array would generate signals continuously as each template is extended. At points where T is present in the template DNA, the signal of incorporation would follow the previous burst of telegraph noise rapidly (generally within 10 ms). The template DNAs in which the next base was a C would show a more delayed burst of telegraph noise because of the slower arrival of dGTP owing to its lower concentration. Similarly, templates containing a G would be preceded by a longer delay because of even lower concentration of dCTP. The longest delays would precede A bases because the concentration of dTTP is lowest.

In yet another embodiment, the two approaches can be combined, using 2 cycles of rinsing, using one pair of nucleotides in unequal concentration in the first cycle, and then the other two, also in unequal concentration in the second cycle.

While the preceding section describes methods of sequencing DNA using a system of arrays comprising a polymerase, the system of arrays can be easily modified to sequence other polymers as well.

Although the invention has been described and illustrated in the foregoing illustrative embodiments, it is understood that the present disclosure has been made only by way of example, and that numerous changes in the details of implementation of the invention can be made without departing from the spirit and scope of the invention. Features of the disclosed embodiments can be combined and rearranged in various ways. All publications, patents and other documents mentioned herein are incorporated by reference in their entirety.

REFERENCES

The following references are hereby incorporated by reference in their entireties:

Chen, Y.-S., C.-H. Lee, M.-Y. Hung, H.-A. Pan, J.-C. Chiou and G. S. Huang (2013). "DNA sequencing using electrical conductance measurements of a DNA polymerase." Nature Nano Technology 8: 452-458.

Choi, Y., I. S. Moody, P. C. Sims, S. R. Hunt, B. L. Corso, I. Perez, G. A. Weiss and P. G. Collins (2012). "Single-molecule lysozyme dynamics monitored by an electronic circuit." Science 335(6066): 319-324.

Merriman, B. and P. Mola, *Biomolecular Sensors and Methods*. US patent application number: WO 2016210386A1, 2016

Olsen, T. J., Y. Choi, P. C. Sims, O. T. Gul, B. L. Corso, C. Dong, W. A. Brown, P. G. Collins and G. A. Weiss (2013). "Electronic measurements of single-molecule processing by DNA polymerase I (Klenow fragment)." J Am Chem Soc **135**(21): 7855-7860.

Zhang, B., W. Song, P. Pang, Y. Zhao, P. Zhang, I. Csabai, G. Vattay and S. Lindsay (2017). "Observation of Giant Conductance Fluctuations in a Protein." Nano Futures **1**(3).