Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2864824 C 2020/06/16

(11)(21) 2 864 824

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2013/02/15
(87) Date publication PCT/PCT Publication Date: 2013/08/22
(45) Date de délivrance/Issue Date: 2020/06/16
(85) Entrée phase nationale/National Entry: 2014/08/15
(86) N° demande PCT/PCT Application No.: US 2013/026414
(87) N° publication PCT/PCT Publication No.: 2013/123379
(30) Priorités/Priorities: 2012/02/16 (US61/599,754); 2012/10/12 (US61/713,163)

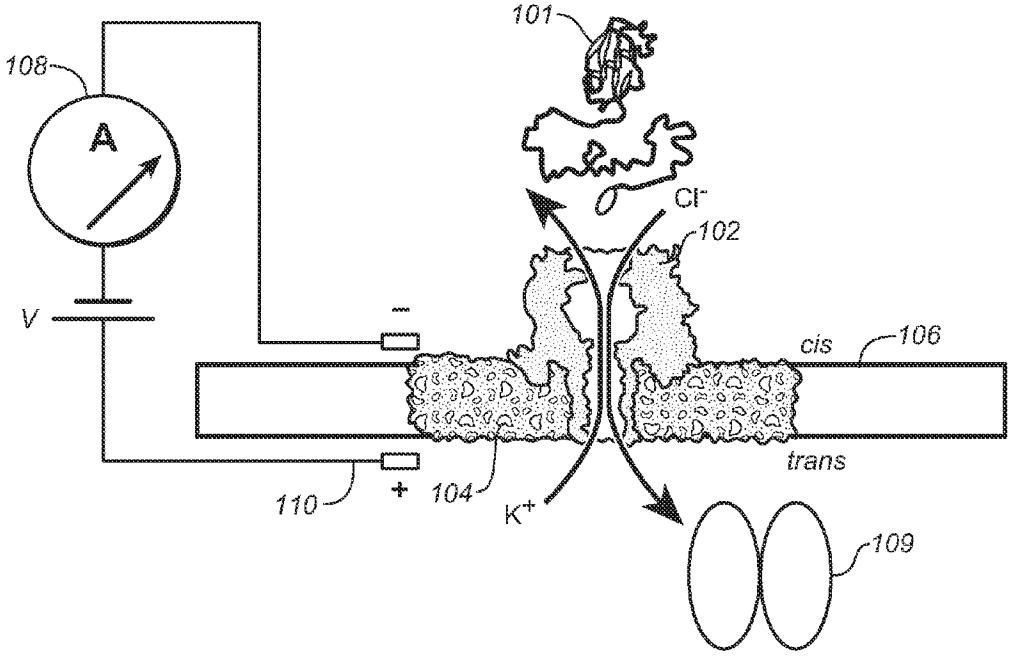
(51) Cl.Int./Int.Cl. *C12Q 1/00* (2006.01), *C07K 1/34* (2006.01), *C12M 1/40* (2006.01), *G01N 33/483* (2006.01)

(72) Inventeurs/Inventors:
NIVALA, JEFFREY M., US;
MARKS, DOUGLAS B., US;
AKESON, MARK A., US

(73) Propriétaire/Owner:
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, US

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : DETECTEUR DE NANOPORE POUR LA TRANSLOCATION DE PROTEINE A MEDIATION PAR ENZYME
(54) Title: NANOPORE SENSOR FOR ENZYME-MEDIATED PROTEIN TRANSLOCATION



(57) Abrégé/Abstract:
Described herein is a device and method for translocating a protein through a nanopore and monitoring electronic changes caused by different amino acids in the protein. The device comprises a nanopore in a membrane, an amplifier for providing a voltage between the cis side and trans side of the membrane, and an NTP driven unfoldase which processed the protein to be translocated. The exemplified unfoldase is the ClpX unfoldase from E. coli.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2013 (22.08.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/123379 A2

(51) International Patent Classification:
C12M 1/00 (2006.01)

(21) International Application Number:
PCT/US2013/026414

(22) International Filing Date:
15 February 2013 (15.02.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/599,754 16 February 2012 (16.02.2012) US
61/713,163 12 October 2012 (12.10.2012) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin St., 12th Floor, Oakland, CA 94607-5200 (US).

(72) Inventors; and
(71) Applicants : **NIVALA, Jeffrey, M.** [US/US]; 205 Baskin Engineering, University Of California, Sata Cruz, Santa Cruz, CA 95064 (US). **MARKS, Douglas, B.** [US/US]; 217 Baskin Engineering, University Of California, Santa Cruz, Santa Cruz, CA 95064 (US). **AKESON, Mark, A.** [US/US]; MS SOE2, University Of California, Santa Cruz, Santa Cruz, CA 95064 (US).

(74) Agents: **ASTON, David, J.** et al.; Peters Verny, LLP, 425 Sherman Avenue, Suite 230, Palo Alto, CA 94306 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: NANOPORE SENSOR FOR ENZYME-MEDIATED PROTEIN TRANSLOCATION

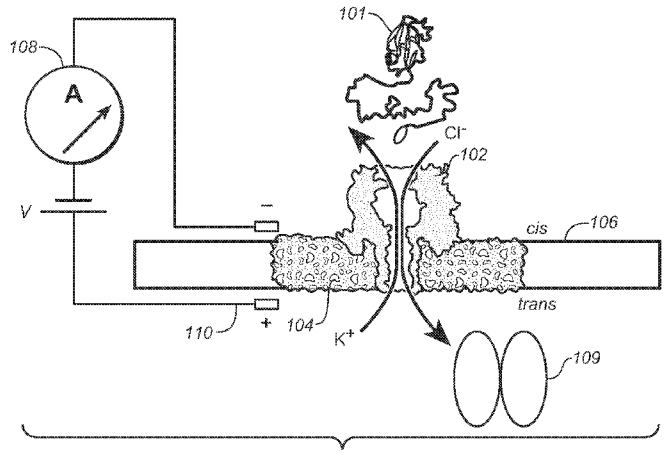


FIG. 1A

(57) Abstract: Described herein is a device and method for translocating a protein through a nanopore and monitoring electronic changes caused by different amino acids in the protein. The device comprises a nanopore in a membrane, an amplifier for providing a voltage between the *cis* side and *trans* side of the membrane, and an NTP driven unfoldase which processed the protein to be trans-located. The exemplified unfoldase is the ClpX unfoldase from *E. coli*.

WO 2013/123379 A2

Nanopore Sensor for Enzyme-Mediated Protein Translocation

Inventors: Jeffrey Nivala, Douglas Marks, Mark Akeson

STATEMENT OF GOVERNMENTAL SUPPORT

This invention was made with Government support under contract R01HG006321 awarded by the National Institutes of Health and contract 24033-444071 awarded by the National Human Genome Research institute. The Government has certain rights in the invention.

REFERENCE TO SEQUENCE LISTING, COMPUTER PROGRAM, OR COMPACT DISK

In accordance with "Legal Framework for EFS-Web," (06 April 11) Applicants submit herewith a sequence listing as an ASCII text file. The text file will serve as both the paper copy required by 37 CFR 1.821(c) and the computer readable form (CRF) required by 37 CFR 1.821(e). The date of creation of the file was 2/13/2013, and the size of the ASCII text file in bytes is 24,576.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates to the field of single molecule protein analysis and also to the field of nanopore analysis.

RELATED ART

Presented below is background information on certain aspects of the present invention as they may relate to technical features referred to in the detailed description, but not necessarily described in detail. That is, individual compositions or methods used in the present invention may be described in greater detail in the publications and patents discussed below, which may provide further guidance to those skilled in the art for making or using certain aspects of the present invention as claimed. The discussion below should not be construed as an admission as to the relevance or the prior art effect of the patents or publications described.

Nanopores have been used for various biosensing applications, the most popular of which has been DNA analysis (e.g. sequencing). Similarly, nanopore sequencing of proteins has also been envisioned. However, unlike nucleic acids, proteins are generally not uniformly charged (making it difficult to drive translocation via an applied voltage) and they also fold into complex, large, and stable structures that cannot transverse a nanopore's aperture. More specifically, the reasons that protein sequencing is technically more challenging than DNA sequencing include: i) twenty different natural amino acids are found in proteins compared to four nucleotides for DNA sequencing (not including post-translational and epigenetic modifications); ii) both tertiary and secondary structures must be unfolded to allow the denatured protein to thread through the nanopore sensor in single file order; and iii) processive unidirectional translocation of the denatured polypeptide through the nanopore electric field must be achieved despite non-uniform charge along the polypeptide backbone.

The use of nanopores to sequence biopolymers was proposed more than a decade ago (Pennisi, E. Search for pore-fection. *Science* 336, 534-537 (2012), Church, G.M., Deamer, D.W., Branton, D., Baldarelli, R. & Kasianowicz, J. Characterization of individual polymer molecules based on monomer-interface interaction.)

Recent advances in enzyme-based control of DNA translocation (Cherf, G.M., Lieberman, K.R., Rashid, Hytham, R., Lam, C.E., Karplus, K. & Akeson, M. Automated "Forward and reverse ratcheting of DNA in a nanopore at 5-Å precision," Nat. Biotechnol. 30, 344-348 (2012).), and in DNA nucleotide resolution using modified biological pores, (Manrao, et al. "DNA at single-nucleotide resolution with a mutant MspA nanopore and

phi29 DNA polymerase," Nat. Biotechnol. 30, 349-353 (2012)), have set the stage for a nanopore DNA sequencing instrument anticipated for commercial release in late 2012 (Hallam, K. "Oxford nanopore to sell tiny DNA sequencer," Bloomberg, published online 17 February 2012, Hayden, E. Nanopore genome sequencer makes its debut. Nature, published online 17 February 2012).

Although protein movement through nanopores has been established (Mohammadet al..," Controlling a single protein in a nanopore through electrostatic traps," *J. Am. Chem. Soc.* 130, 4081-4088 (2008), Talaga, D.S. & Li, J. "Single-molecule protein unfolding in solid state nanopores," *J. Am. Chem. Soc.* 131, 9287-9297 (2009), Merstorf, et al. " Wild type, mutant protein unfolding and phase transition detected by single-nanopore recording," *ACS Chem. Biol.* 7, 652-658 (2012)), a technique to unfold proteins for controlled, sequential translocation has until now not been demonstrated.

BRIEF SUMMARY OF THE INVENTION

The following brief summary is not intended to include all features and aspects of the present invention, nor does it imply that the invention must include all features and aspects discussed in this summary.

The present invention provides a device for translocating a protein through a nanopore, comprising: a membrane having nanopore therein, said membrane separating a chamber into a *cis* side and a *trans* side, wherein the protein is to be added to the *cis* side and translocated through the nanopore to the *trans* side; and a protein translocase enzyme, on one side of said chamber, which binds to and translocates the protein through the nanopore. Translocation will occur in a sequential order, that is, in a defined sequence of amino acid residues passing into the nanopore, which will generally follow the primary amino acid sequence of the protein. Multiple proteins can be translocated one at a time.

The present invention also provides a device for translocating a protein through a nanopore, comprising: a nanopore in a membrane, said membrane separating a fluidic chamber into a *cis* side and a *trans* side, wherein a protein to be translocated is added to the *cis* side and is translocated through the nanopore to the *trans* side; a circuit for providing a voltage gradient between the *cis* side and the *trans* side and for measuring ionic current flowing through the nanopore; and a specific enzyme, such as a protein translocase and/or an

NTP driven unfoldase added to the fluid chamber, e.g. by being allowed to become attached to said nanopore on the *cis* side, or by addition in solution to the *trans* side.

In one embodiment of the present invention, the nanopore is defined by a pore protein. In one preferred embodiment, the pore protein is α-hemolysin.

In another embodiment of the present invention, the protein translocase is an NTP driven unfoldase which operates on the protein molecule to be translocated. In one preferred embodiment, the NTP driven unfoldase is an AAA+ enzyme. In another preferred embodiment, the AAA+ enzyme is a combination of subunits of *E. coli* ClpX.

In another embodiment of the present invention, the circuit for detection of protein translocation comprises a patch clamp amplifier applying a positive voltage to the *trans* side. The patch clamp amplifier maintains a constant voltage and measures changes in current. In a preferred embodiment, the device comprises a computer, attached to the patch clamp amplifier, for rapidly recording changes in ionic current through the nanopore. As the protein passes through the nanopore an ionic current signature is obtained which can detect on the order of 1 to 100,000 fluctuations per second, providing information about individual amino acids translocating through the pore. For example, recording at 100kHz can be used to produce one data point every 10 μS. The data can be correlated to structural features of the protein being translocated.

In another embodiment of the present invention, a system for translocating a protein through a nanopore is provided, comprising a nanopore in a membrane separating a fluidic chamber into a *cis* side and a *trans* side, wherein a protein to be translocated is added to the *cis* side and is translocated through the nanopore to the *trans* side; said fluidic chamber comprising an ionic buffer containing an enzyme cofactor such as NTP (nucleoside 5'-triphosphate, e.g. ATP and/or GTP) and a non-denatured protein to be translocated on the *cis* side; a circuit for providing a voltage between the *cis* side and the *trans* side and measuring ionic current flowing through the nanopore; and a protein translocase such as an NTP driven unfoldase in solution in the chamber on the *cis* side.

In an alternative embodiment of the present invention, the nanopore is defined by a pore protein such as a multimeric pore protein and the protein translocase such as an NTP driven unfoldase is attached to the multimeric pore protein. The protein translocase may be

covalently or non-covalently attached to the pore protein, and may be on the *cis* side, the *trans* side or both sides of the membrane and pore protein.

In another embodiment of the present invention, the protein to be translocated is a non-denatured protein (i.e. in its native state) and, further, comprises an exogenous sequence comprising a targeting domain for the protein to be targeted to pass through the nanopore and contact the NTP driven unfoldase. In a preferred embodiment, the NTP driven unfoldase is ClpX and the nanopore protein is α- hemolysin. The targeting domain in the exogeneous sequence serves to guide the protein to the nanopore. The targeting domain may be configured to be affected by the voltage across the nanopore. In one preferred embodiment, the targeting domain comprises at about 5 negatively charged amino acids or at least about 5-30 negatively charged amino acids and is drawn to the positive side of the chamber by a voltage gradient applied between the *cis* side and the *trans* side.

The present invention also provides a method for translocating a non-denatured protein through a nanopore, comprising the steps of: providing a device for translocating a protein through a nanopore, said device comprising a nanopore in a membrane separating a fluidic chamber into a *cis* side and a *trans* side, wherein a protein to be translocated is added to the *cis* side and is translocated through the nanopore to the *trans* side; a circuit for providing a voltage between the *cis* side and the *trans* side and measuring ionic current flowing through the nanopore; and a protein translocase in solution on the *trans* side; adding to said fluidic chamber a buffer containing NTPs (where the translocase is NTP-driven); optionally adding a non-denatured protein to the *cis* side; allowing the non-denatured protein to be captured or threaded through the nanopore (e.g. by charge) so that it can contact the protein translocase; and measuring ionic current changes caused by translocation of the non-denatured protein through the nanopore.

In one embodiment of the present invention, the step of measuring current changes comprises measuring current changes for states of (i) open channel in the nanopore, (ii), capture of the nondenatured protein by the nanopore, and (iii) passage of a protein from (ii) through the nanopore. In a preferred embodiment, the measuring comprises detecting differences between states (i), (ii) and (iii). In another preferred embodiment, the measuring comprises measuring differences during state (iii) caused by amino acid structure of the protein passing through the nanopore. The method may further comprise the step of measuring a state of binding of the NTP driven unfoldase to the nondenatured protein and

translocation of the unfoldase toward the nanopore, which occurs as a state between states (ii) and (iii). This would result in measuring four states. As described and illustrated, e.g. in Figure 2, there may be a final state (v) measured when the translocation is complete and the nanopore returns to initial state.

In another preferred embodiment, the nanopore is defined by a pore protein. In another preferred embodiment, the pore protein is α- hemolysin. In another preferred embodiment, the NTP driven unfoldase is attached to the pore protein. In another preferred embodiment, the NTP driven unfoldase is an AAA+ enzyme. In another preferred embodiment, the AAA+ enzyme is ClpX. In another embodiment, the circuit comprises a patch clamp amplifier applying a constant voltage between the *cis* chamber and the *trans* chamber.

In certain aspects of the present invention, the nanopore is defined by α- hemolysin or another multimeric pore protein. The pore protein does not need to be functionalized; that is it may be used as the protein exits in its native environment; it does not need to have any molecular structures added to it to attach or bind the protein being translocated. That is, it may be generic for translocation of any protein sequence, and does not specifically bind to or recognize the protein to be translocated. The protein to be translocated is also preferably in a native form. It may, in certain embodiments of the present invention, have attached to it a molecular structure for improving "threading" of the protein through the nanopore. In certain aspects of the present invention, the protein to be translocated comprises an exogeneous sequence comprising a targeting domain for the protein translocase. The targeting domain may comprise at least about 5-30 amino acids, or 10-30 amino acids. The amino acids may be negatively charged amino acids, e.g. 30-100 glutamate or aspartate residues, or other negatively charged synthetic monomers, e.g. dextran sulfate, located at an amino or carboxy terminus of the protein to be translocated. The amino acids may also be positively charged, e.g. arginine or lysine, if the voltage polarity was reversed from that exemplified below.

In certain aspects of the present invention, the protein to be translocated is translocated in its native state, that is, without being denatured or otherwise unfolded; the translocase serves to unfold the protein as it traverses the voltage gradient across the nanopore.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a diagrammatic drawing (cartoon) of a nanopore sensor with a single AHL (α-hemolysin) pore embedded in a lipid bilayer.

Figure 1B is a diagrammatic drawing that shows a protein captured in the nanopore.

Figure 1C is a diagrammatic drawing that shows the engineered proteins used in the present examples for translocation.

Figure 2A, 2B is a diagrammatic representation and plot that shows the ionic current traces during ClpX-mediated protein translocation.

Figure 2C is a trace that shows the ionic current traces during protein S2-35 translocation.

Figure 2D is a trace that shows the ionic current traces during protein S2-148 translocation.

Figure 3A, 3B and 3C is a series of bar graphs that shows the comparison of ionic current state dwell times for the three model proteins.

Figure 4A-D is a diagrammatic representation and current trace that shows the ionic current traces during voltage-mediated protein translocation without the presence of ClpX in the *trans* solution. After a highly variable capture duration (< 5 sec – > 2 min), all substrates tested will eventually unfold and translocate due to the applied voltage. No ramping states are observed, detailed signal features are lost from S2-35 and S2-148 linker states, and all states have more widely distributed durations as compared to ClpX-mediated events. **Figure 4A** shows the ionic current traces during voltage-mediated S1 translocation. Compared to **Fig. 2A**, state iii is absent and state iv has a longer and more variable duration on average. **Figure 4B** illustrates the model of voltage-mediated protein translocation. **Figure 4B** shows four cartoon structures. i. through iv. Cartoons i-iv correspond to ionic current states i-iv in **Fig. 4A. Figure 4C** shows the ionic current traces during voltage-mediated S2-35 translocation. Ramping of states iii and vi are absent and resolution of state v (**Fig. 2C**) is diminished. **Figure 4D** shows the ionic current traces during voltage-mediated S2-148 exhibits similar behavior to S2-35 with the corresponding states omitted (**Fig. 2D**).

Figure 5 is a frequency bar graph that demonstrates the comparison of ionic current state dwell times of ClpX-dependent (with ClpX present in the *trans* side) ramping state iii for the three model proteins. S1 n=45, S2-35 n=62, S2-148 n=66.

Figure 6 is a frequency bar graph that shows the comparison of ionic current state dwell times of the putative second Smt3 domain translocation state vii of the S2-35 (n=42) and S2-148 (n=41) proteins in events that included ClpX-dependent ramping states iii and vi.

Figure 7 is a frequency bar graph that shows the comparison of ionic current dwell times of ClpX-dependent ramping state vi of proteins S2-35 (n=44) and S2-148 (n=44).

Figure 8 is a frequency bar graph that shows the comparison of ionic current dwell times of the putative Smt3 domain translocation states iv and vii for the three model proteins. The black bars represent dwell times for events that included ramping state iii (ClpX-driven). The gray bars represent events that did not include the ramping state (not ClpX-driven). With ramping n=254, without ramping n=183.

Figure 9 is a frequency bar graph that shows state v dwell times for S2-35 translocation events. The black bars represent dwell times for events that included ramping state iii (ClpX-driven). The gray bars represent events that did not include the ramping state (not ClpX-driven). With ramping n=50, without ramping n=45.

Figure 10 is a diagrammatic representation that illustrates a fusion protein of ClpP/α-HL embedded within a lipid membrane.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

OVERVIEW

Described herein is a system for translocating an individual protein through a nanopore so as to enable information about the protein, e.g. amino acid content of the protein, to be obtained through electronic signals reflecting passage of the protein through the nanopore. By providing a nanopore in a membrane, a voltage between the *cis* side and *trans* side of the membrane, and a protein translocase, the present device achieves the enzyme-controlled unfolding and translocation of native proteins through a nanopore sensor using the protein translocase in such a way that circuitry between the *cis* side and the *trans* side can monitor and record signals indicative of the amino acid content of the protein, e.g. amino acid

sequence. For practical purposes, an array of nanopores and circuits can be provided. These can be in a single chamber or in multiple chambers.

Referring now to **Figure 1A**, the present device operates to translocate a substrate protein 101 and comprises a pore protein 102 (α-hemolysin, or "AHL") embedded in a lipid bilayer 104 that is comprised in an ~ 25 μm aperture in a membrane 106 separating a fluid compartment into a cis side, containing the protein 101, and a trans side, to which the protein 101 is going to be translocated through the pore 102. The device includes a controllable amplifier 108 for applying a constant voltage between a positive electrode 110 on the trans side and a negative electrode on the cis side. A protein translocase 109, exemplified below as ClpX, is present on the trans side of the chamber. Amplifier 108 also provides a circuitry for detecting and, preferably, recording changes in ionic current (i.e. flow of ions such as the depicted Cl- and K+) that take place very rapidly as the protein 101 translocates. In the examples, data were collected at 100 kHz, but high speed data sampling devices are known and may be used (e.g. 200MHz Model 7150 from Pentek, Inc). Figure 1B shows a detailed view of the AHL pore protein 102 in the lipid bilayer 104 and also shows the protein translocase 109 which is on the *trans* side and which is acting on protein 101 which in the cartoon is threaded through and is on both sides of the pore 102. As shown in Figure 1C, a model substrate protein bearing a Smt3 domain at its amino-terminus is coupled by a charged flexible linker to an ssrA tag at its carboxy-terminus. The charged, flexible tag is threaded through the nanopore into the trans-side solution, while the folded Smt3 domain at this point prevents complete translocation of the captured protein. ClpX present in the trans solution binds the C-terminal ssrA sequence of the substrate protein. Fueled by ATP hydrolysis, ClpX translocates along the protein tail toward the channel, and subsequently catalyzes unfolding and translocation of the Smt3 domain(s) through the pore. The Smt3 domains are folded, while the linker(s) are not. Smt3 is further described in US 2009/0280535, "SUMO Fusion Protein Expression System for Producing Native Proteins."Demonstrated in the examples below is enzymatic control of protein unfolding and translocation through the α-hemolysin nanopore. Segments of each substrate protein were discerned based on amino acid composition as they passed through the circa 50-Angstrom-long trans-membrane pore lumen (nanopore). The translocase enzyme used is selected and controlled to enable the device to provide protein sequence information. The enzyme is selected from a class of enzymes termed generally herein "protein translocases," referring to

the ability of such enzymes to cause physical movement relative to a substrate. Included within the term as used herein is a class of enzymes often referred to as "unfoldases," in that they catalyze the unfolding of a native protein without affecting the primary structure, i.e. the primary sequence of the protein.

In certain embodiments, the substrate protein is tagged for recognition by the translocase. One way to do this is the use of an ssrA tag. Various ssrA tags are known, as this is the mechanism used in several bacterial species for marking proteins to be degraded by a ClpX protease system. In the examples, the ssrA tag is a C-terminal 11 residue AA sequence (shown at the end of SEQ ID NO: 5) of which subsets of this sequence are recognized by ClpA or ClpX uniquely. As noted, other sequences may be used. IN certain embodiments a protein nanopore protein or a chimeric nanopore as shown in **Figure 10** may be embedded in a thin insulting membrane (for example, a lipid bilayer or a graphene sheet) separating two conductive aqueous solutions of differential voltage. Sensing would be imparted by the flow of ionic current through the nanopore; as the protein translocated or otherwise interacted with the pore, blockades of ion flow would occur, providing an electronic signal for subsequent analysis. This protein nanopore or chimeric nanopore could also be utilized in arrays or lab-on-chip devices for paralleled separation and/or purification of target proteins in mass.

The invention may be carried out in various apparatus for nanopore analysis, such as an array or a chip. The apparatus may be any of those described in International Application No. PCT/GB08/004127 (published as WO 2009/077734, entitled "Formation of layers of amphiphilic molecules"), PCT/GB10/000789 (published as WO 2010/122293 entitled "Lipid bilayer sensor array"), International Application No. PCT/GB10/002206 (published as WO 00/28132 entitled "Biochemical analysis instrument") or International Application No. PCT/US99/25679 (published as WO 2000/28312 entitled "Coupling method"). As will become apparent from the description below, the protein is translocated as a single polypeptide sequence, wherein individual amino acids pass sequentially through the pore. A number of proteins may be translocated serially.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Generally, nomenclatures utilized in connection with, and techniques of, cell and molecular biology and chemistry are those well-known and commonly used in the art. Certain experimental techniques, not specifically defined, are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification. For purposes of clarity, the following terms are defined below.

Ranges: For conciseness, any range set forth is intended to include any sub-range within the stated range, unless otherwise stated. As a non-limiting example, a range of 120 to 250 is intended to include a range of 120-121, 120-130, 200-225, 121-250 etc. The term "about" has its ordinary meaning of approximately and may be determined in context by experimental variability. In case of doubt, "about" means plus or minus 5% of a stated numerical value.

The term "nanopore" is used herein in its conventional sense to refer to any small hole or channel of the order of 0.5 to 10 nanometers in internal diameter. The term "nanopore" includes both biological (e.g. α-hemolysin) or artificial nanopores. The present nanopores can vary in dimensions, for example it can have a diameter of between about 0.5 nm and 10 nm in size. For example, the diameter can be about 0.5 nm, 1 nm, 1.25 nm, 1.5 nm, 1.75 nm, 2 nm, 2.25 nm, 2.5 nm, 2.75 nm, 3 nm, 3.5 nm, 4 nm, 4.5 nm, 5 nm, 6 nm, 7 nm, 8 nm, 9nm, 10 nm, or any dimension there between. Biological nanopores can be created by pore proteins. Artificial nanopores can be made by micromolding or drilling. They also can be made by etching a somewhat larger hole (several tens of nanometers) in a piece of silicon, and then gradually filling it in using ion-beam sculpting methods which results in a much smaller diameter hole.

The term "pore protein" is used herein in its conventional sense to refer to pore-forming proteins (PFPs) which assemble into ring-like structures in the vicinity of the target membrane to expose sufficient hydrophobicity to drive spontaneous bilayer insertion. Pore

proteins are typically (but not exclusively) produced by bacteria, such *C. septicum* and *S. aureus*. PFPs can be alpha-pore-forming toxins, such as Cytolysin A of *E. coli*; or beta-pore-forming toxins, such as α-hemolysin and Panton-Valentine leukocidin (PVL); or binary toxins, such as Anthrax toxin; or cholesterol-dependent cytolysins (CDCs), such as Pneumolysin; or Small pore-forming toxins, such as Gramicidin A. A preferred pore protein is α-hemolysin (AHL).

The term "α-hemolysin" is used herein in its conventional sense to refer to a pore-forming toxin from the bacterium, *Staphylococcus aureus*. α-hemolysin consists mostly of beta-sheets (68%) with only about 10% alpha-helices. The *hla* gene on the *S. aureus* chromosome encodes the 293 residue protein monomer, which forms heptameric units on the cellular membrane to form a complete beta-barrel pore. This structure allows the toxin to perform its major function, development of pores in the cellular membrane.

The term "membrane" is used herein in its conventional sense to refer to a thin, film-like structure. The membrane separating the *cis* and *trans* chambers comprises at least one pore or channel. Membranes can be generally classified into synthetic membranes and biological membranes. Any membrane may be used in accordance with the invention. Suitable membranes are well-known in the art. The membrane is preferably an amphiphilic layer. An amphiphilic layer is a layer formed from amphiphilic molecules, such as phospholipids, which have both at least one hydrophilic portion and at least one lipophilic or hydrophobic portion. The amphiphilic layer may be a monolayer or a bilayer. The amphiphilic molecules may be synthetic or naturally occurring. Non-naturally occurring amphiphiles which form a monolayer are known in the art and include, for example, block copolymers (Gonzalez-Perez et al., Langmuir, 2009, 25, 10447-10450).

The term "lipid bilayer" is used herein in its conventional sense to refer to a thin polar membrane made of two layers of lipid molecules, arranged so that the hydrophilic phosphate heads point "out" to the water on either side of the bilayer and the hydrophobic tails point "in" the core of the bilayer. The lipid bilayers are usually a few nanometers in width, and they are impermeable to most charged water-soluble molecules. Lipid bilayers are large enough structures to have some of the mechanical properties of liquids or solids. The area compression modulus Ka, bending modulus Kb, and edge energy, can be used to describe them. Solid lipid bilayers also have a shear modulus, but like any liquid, the shear modulus is

zero for fluid bilayers. Lipid bilayers can also be supported by solid substrates having apertures, such as heat shrink tubing, fused silica, borosilicate glass, mica, and oxidized silicon. Lipids may be applied, e.g., through Langmuir-Blodgett technique, vesicle fusion processes or the combination of the two.

The term "NTP" is used herein in its conventional sense to refer to nucleoside triphosphate, a molecule containing a nucleoside bound to three phosphates, making it a nucleotide. NTP can be adenosine triphosphate (ATP), guanosine triphosphate (GTP), cytidine triphosphate (CTP), 5-methyluridine triphosphate (m^5UTP), uridine triphosphate (UTP), deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), deoxycytidine triphosphate (dCTP), deoxythymidine triphosphate (dTTP) or deoxyuridine triphosphate (dUTP). "NTP" also refers to other less abundant NTPs, such as intermediates of nucleotide metabolism, including less common natural

The term "NTP driven unfoldase" is used herein in its conventional sense to refer to an NTP-dependent enzyme that catalyzes protein unfolding. The very common NTP driven unfoldases are ATP-dependent proteases, such as proteasomal ATPases, AAA proteases, or AAA+ enzymes (defined below); membrane fusion proteins, such as NSF (N-Ethylmaleimide-sensitive fusion protein)/Sac18p (N-Ethylmaleimide-sensitive fusion protein homologue in yeast) or p97 / VCP / Cdc48p (97-kDa valosin-containing protein); Pex1p and Pex6p (peroxisomal ATPase); Katanin and SKD1 (Vps4p homolog in mouse)/Vps4p (Vacuolar protein sorting 4 homolog in yeast); Dynein (motor protein); DNA replication proteins, such as ORC (origin recognition complex), Cdc6 (cell division control protein 6), MCM (minichromosome maintenance protein), DnaA, or RFC (replication factor C)/clamp-loader; RuvB (holliday junction ATP-dependent DNA helicase RuvB, EC=3.6.4.12); TIP49a/TIP49 and TIP49b/TIP48 (eukaryotic RuvB-like protein).

The term "AAA+ enzyme" is used herein in its conventional sense to refer to the AAA+ superfamily of enzymes. AAA+ is an abbreviation for ATPases Associated with diverse cellular Activities. They share a common conserved module of approximately 230 amino acid residues. This is a large, functionally diverse protein family belonging to the AAA+ superfamily of ring-shaped P-loop NTPases, which exert their activity through the energy-dependent remodeling or translocation of macromolecules. Examples include ClpAP, ClpXP, ClpCP, HslVU and Lon in bacteria and their homologues in mitochondria and chloroplasts. With the exception of Lon, AAA+ enzymes (sometimes referred to as

unfoldases or proteases) consist of regulatory (ATPase) and proteolytic subunits, while Lon is a single polypeptide containing both regulatory and proteolytic domains. ClpX and ClpA dock with ClpP to form ClpXP and ClpAP proteases, whereas HslU docks with HslV to form another protease, HslVU. ClpA and ClpX form hexamers, in contrast to ClpP which forms heptamers. HslU and HslV each form hexamers, although HslU heptamers have also been reported. The regulatory subunits ClpA, ClpX and HslU function as chaperones. Further details on ClpX may be found in Maillard et al., "ClpX(P) generates mechanical force to unfold and translocate its protein substrates," Cell 145:459-4669 (April 29, 2011). As reported there, the ClpX motor shares is basic design with other AAA+ enzymes, including prokaryotic ClpA, ClpB, HsIu, FtsH or Lon. The AAA+ enzyme is also referred to as an "AAA+ molecular motor". Further description of the AAA+ superfamily is found in Ogura et al. "AAA+ superfamily ATPases" common structure-diverse function," Genes to Cells, 6:575-597 (2001). AS described there, the AAA+ family members associated with mitochondria are Bcs1p, Lon/Pim1p, ClpX and Hsp78.

The term "HsIU" is used herein in its conventional sense to refer to ATP-dependent protease ATPase subunit HsIU, also called unfoldase HsIU. HsIU is a member of the Hsp100 and Clp family of ATPase. It can also form complex with HsIV to act as an unfoldase (See, Bochtler et al., "The structures of HsIU and the ATP-dependent protease HsIU-HsIV," Nature 403(6771):800-805 (2000).

The term "Lon protease" is used herein in its conventional sense to refer to a family of proteases found in archaea, bacteria and eukaryotes. Lon proteases are ATP-dependent serine peptidases belonging to the MEROPS peptidase family S16 (lon protease family, clan SF). In the eukaryotes the majority of the Lon proteases are located in the mitochondrial matrix. In yeast, the Lon protease PIM1 is located in the mitochondrial matrix. It is required for mitochondrial function, it is constitutively expressed but is increased after thermal stress, suggesting that PIM1 may play a role in the heat shock response.

The term "protein translocase" is used herein in its conventional sense to mean a protein-binding polypeptide, such as a polypeptide which is able to control movement of a protein substrate, for example an enzyme, enzyme complex, or a part of an enzyme complex that operates on a protein substrate and moves it relative to the enzyme in a processive manner, i.e. as a function of enzymatic activity. The term "processive" is understood in the art to refer to a stepwise activity in which the enzyme "processes" the substrate in a number

of steps. In the present case, the protein translocase generally processes the protein to be translocated in a sequential manner, that is, moving along the primary amino acid sequence. For convenience, a number of enzymes also commonly called "unfoldases" are included in this definition, in particular NTP driven unfoldases. Also specifically included in this definition is the AAA+ enzyme superfamily and the ClpX member of this superfamily.

Also included as examples of the general term "protein translocase" are proteases such as Lon protease and HsIU, which enzymes are either modified to eliminate the enzymatic cleavage activity of the enzyme or arranged so that cleavage occurs after the sequence is translocated through the nanopore.

Other exemplary protein translocases are related to ClpX, (which is also an unfoldase), e.g. ClpA, mitochondrial protein translocases TOM (translocase of the outer membrane) or other TOM and TIM proteins. The chosen protein translocase can also be any part of the mitochondrial protein translocase complex, such as the chaperones, TOM import receptor, TOM channel complex, and "motor" proteins.

The term "ClpX enzyme" of "ClpX" is used herein in its conventional sense to refer to a member of the HSP (heat-shock protein) 100 family having the Uniprot designation clpX and having the 424 amino acid sequence given there, processed into mature form, as a subunit. ClpX subunits associate to form a six-membered (homohexameric) ring that is stabilized by binding of ATP or nonhydrolysable analogs of ATP. The N-terminal domain of ClpX is a C4-type zinc binding domain (ZBD) involved in substrate recognition. ZBD forms a very stable dimer that is essential for promoting the degradation of some typical ClpXP substrates such as lO and MuA. It is described further in Wawrzynow et al, "The ClpX heat-shock protein of Escherichia coli, the ATP-dependent substrate specificity component of the ClpP-ClpX protease, is a novel molecular chaperone," EMBO J. 1995 May 1; 14(9): 1867–1877, An amino acid sequence is also given at eclowiki.net under "clpX: gene products".

Similarly, ClpA refers to the UniProt/Swiss-Prot entry clpA, which has a 758 amino acid sequence given there for the ClpA subunit. It forms a complex of six ClpA subunits assembled into a hexameric ring in the presence of ATP. It is a component of the ClpAP complex composed of six ClpA subunits assembled into a hexameric ring in the presence of ATP, and fourteen ClpP subunits arranged in two heptameric rings. Binds to ClpS.

The term "non-denatured protein" is used herein in its conventional sense, i.e. a protein that is at least partially folded into a native secondary and tertiary structure, with any native cysteine bonds, hydrogen bonding and multimeric form essentially intact. This is contrasted with a denatured protein, which usually is insoluble and aggregated.

The term "negatively charged amino acids" is used herein in its conventional sense, i.e. meaning proteins that have surfaces rich with negatively charged amino acids like glutamate and aspartate.

GENERAL METHOD AND APPARATUS

Translocation of proteins through a nanopore sensor device offers a number of possible applications, including sequencing, structure/fold analysis, purification/separation, intracellular protein delivery, and insight into the mechanics of enzymes driving the translocating polypeptide. Unlike nucleic acids, proteins are generally not uniformly charged (making it difficult to drive translocation via an applied voltage) and fold into complex, large, and stable structures that cannot transverse a nanopore's aperture. To address these issues, unfolding and translocation of natively folded proteins through a protein nanopore may be accomplished via a variety of enzymes, exemplified by the *E. coli* ClpX (or other types of protein translocases/unfoldases).

The present methods and devices may be used to measure one or more characteristics of the protein being translocated.

A variety of different types of measurements may be made. This includes without limitation: electrical measurements and optical measurements. Possible electrical measurements include: current measurements, impedance measurements, tunnelling measurements (Ivanov AP et al., Nano Lett. 2011 Jan 12;11(1):279-85), and FET measurements (International Application WO 2005/124888). Optical measurements may be combined with electrical measurements (Soni GV et al., Rev Sci Instrum. 2010 Jan;81(1):014301). The measurement may be a transmembrane current measurement such as measurement of ionic current flowing through the pore.

Electrical measurements may be made using standard single channel recording equipment as described in Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7, Lieberman KR et al, J Am Chem Soc. 2010;132(50):17961-72, and International Application

WO-2000/28312. Alternatively, electrical measurements may be made using a multi-channel system, for example as described in International Application WO-2009/077734 and International Application WO-2011/067559.

The signal measurement is typically indicative of the identity of the protein or the amino acids in the protein. The signal can therefore be used to characterize, such as sequence, the protein as discussed above.

1. Enzymes used for translocation

E. coli ClpX was used in the working example of the present device; it was selected for initial work because it generates sufficient mechanical force (>20 pN) to denature stable protein folds, and because it translocates along proteins at a suitable rate for primary sequence analysis by nanopore sensors (up to 80 amino acids per second). ClpX is part of the ClpXP proteasome-like complex. ClpP is composed of a diheptameric cylinder-like protease that binds at one or both ends a regulatory hexameric ATP-dependent unfoldase/translocase complex (e.g. ClpX). ClpX acts as a gate that allows for tagged proteins to enter into the inner lumen of the ClpP protease complex for subsequent degradation. The ATP-dependent unfoldase/translocase activity of the hexameric protein complex, ClpX, is employed to unfold and thread proteins through a nanopore.

ClpX may be prepared (and was prepared here) as described in Martin et al. "Rebuilt AAA+ Motors reveal operating principles for ATP-fuelled machines," Nature 437:1115-1120 (2005). A variety of alternative enzymes may serve the function of the protein translocase in the present method and device. As described there, combinations of 2, 3 or 6 ClpX-deltaN subunits, lacking N terminal amino acids 1-60, linked with a 20 amino acid long linker were prepared as a single polypeptide chain.

The repertoire of cellular functions involving AAA+ ATPases is diverse. A subset of AAA+ proteins is not active as ATPases and some do not even bind ATP. It seems however, that these proteins form complexes with other family members which do serve as ATPases. However, the ATPase subunits or domains of all known ATP-dependent proteases belong to the AAA+ family.

One example of a suitable AAA+ enzyme is Clp/Hsp100 ATPases. Clp/Hsp100 ATPases are responsible for selecting protein targets. For example, the two different bacterial

ATPases ClpX and ClpA impart distinct substrate preferences to the ClpP peptidase. The ssrA degradation sequence, an 11-residue peptide that is appended to polypeptides stalled on the ribosome, is recognized by both ClpX and ClpA. Mutational analysis of the ssrA sequence revealed that this same tag is recognized by the two unfolding enzymes via different residues, further confirming the distinct binding preferences of each ATPase.

Using the energy from ATP-hydrolysis, Clp/Hsp100 enzymes actively direct structural changes in their substrates. These ATP-driven structural changes result in two distinct biological outcomes for the protein substrates: degradation or remodeling. ClpA, based on its ability to degrade casein, was the first prokaryotic Clp/ Hsp100 protein functionally identified. Accordingly, the degradation pathway for the Clp/Hsp100 proteins is the better characterized of the two processes. During Clp/Hsp100-facilitated protein degradation, first, the Clp/Hsp100 component recognizes and selects a target protein. The enzyme binds to a short peptide sequence (e.g., the ssrA degradation tag) usually located near either the C or N terminus of the substrate. Then, in a reaction that requires multiple cycles of ATP-hydrolysis, the enzyme unfolds and directionally translocates the target substrate to the peptidase chamber where it is degraded.

Mitochondrial protein translocases may also be used. These may be translocases TOM or TIM from human or eukaryotic cells, such as TOMM20 (translocase of outer mitochondrial membrane 20 homolog), TOMM22 (mitochondrial import receptor subunit 22 homolog), TOMM40 (translocase of outer mitochondrial membrane 40 homolog), TOM7 (translocase of mitochondrial outer membrane 7), TOMM7 (translocase of outer mitochondrial membrane 7 homolog), TIMM8A (translocase of inner mitochondrial membrane 8 homolog A), TIMM50 (translocase of inner mitochondrial membrane 50 homolog). For example, TOMM40 is embedded into outer membranes of mitochondria and is required for the movement of proteins into mitochondria. More, precisely, TOMM40 is the channel-forming subunit of a translocase of the mitochondrial outer membrane (TOM) that is essential for protein transport into mitochondria

Another alternative protein translocase may be prepared from the Sec family of translocases. These include SecB (chaperone protein), SecA (ATPase), SecY (internal membrane complex in prokaryotes), SecE (interal membrane complex in prokaryotes), SecG (internal membrane complex in prokaryotes) or Sec61 (internal membrane complex in eukaryotes), SecD (membrane protein), and SecF (membrane protein).

Another alternative protein translocase is Type III Secretion System (TTS) Translocase, such as HrcN and any of the 20 subunits of the TTS translocases, or Sec-independent periplasmic protein translocase TatC.

Other alternative protein translocases are chaperones. These are proteins that assist the non-covalent folding or unfolding and the assembly or disassembly of other macromolecular structures, but do not occur in these structures when the structures are performing their normal biological functions having completed the processes of folding and/or assembly. Many chaperones are heat shock proteins, that is, proteins expressed in response to elevated temperatures or other cellular stresses. Hsp 70, as is known, refers to 70-kDa heat shock proteins (Hsp70s), such as DnaK, HscA (Hsc66), and HscC (Hsc62) in prokaryotes, and Hsc70, Hsp70, BiP or Grp78 (binding immunoglobulin protein), mtHsp70 or Grp75 in eukaryotic organisms, and human Hsp70 proteins, such as Hsp70, Hsp70-2, Hsp70-4, Hsp70-4L, Hsp70-5, Hsp70-6, Hsp70-7, Hsp70-8, Hsp70-9, Hsp70-12a, Hsp70-14. Hsp70 proteins are central components of the cellular network of molecular chaperones and folding catalysts. Hsp70s assist a wide range of folding processes, including the folding and assembly of newly synthesized proteins, refolding of misfolded and aggregated proteins, membrane translocation of organellar and secretory proteins, and control of the activity of regulatory protein. ATP binding and hydrolysis are essential in vitro and in vivo for the chaperone activity of Hsp70 proteins.

Hsp70 chaperone families are recognized as most common remodeling enzyme together with Hsp60 chaperone families. Hsp70s and Hsp60s prevent off-pathway interactions during protein folding by providing an isolated environment for the folding protein. In contrast, the Clp/Hsp100 unfolding enzymes actively direct the structural changes in their substrates. Clp/Hsp100s act on folded and assembled complexes, as well as improperly folded and aggregated proteins.

HSP90 aids the delivery of the mitochondrial preprotein to the TOM complex in an ATP-dependent process.

Hsp100 (Clp family in *E. coli*) proteins have been studied in vivo and in vitro for their ability to target and unfold tagged and misfolded proteins. Proteins in the Hsp100/Clp family form large hexameric structures with unfoldase activity in the presence of ATP. These proteins are thought to function as chaperones by processively threading client proteins

through a small 20 Å (2 nm) pore, thereby giving each client protein a second chance to fold. Some of these Hsp100 chaperones, like ClpA and ClpX, associate with the double-ringed tetradecameric serine protease ClpP; instead of catalyzing the refolding of client proteins, these complexes are responsible for the targeted destruction of tagged and misfolded proteins.

Hsp104, the Hsp100 of Saccharomyces cerevisiae, is essential for the propagation of many yeast prions. Deletion of the HSP104 gene results in cells that are unable to propagate certain prions.

The enzyme used in the working examples has the sequence:

MGSSHHHHHHSSHMSALPTPHEIRNHLDDYVIGQEQAKKVLAVAVYNHYKRLRNG
DTSNGVELGKSNILLIGPTGSGKTLLAETLARLLDVPFTMADATTLTEAGYVGEDVE
NIIQKLLQKCDYDVQKAQRGIVYIDEIDKISRKSDNPSITRDVSGEGVQQALLKLIEGT
VAAVPPQGGRKHPQQEFLQVDTSKILFICGGAFAGLDKVISHRVETGSGIGFGATVKA
KSDKASEGELLAQVEPEDLIKFGLIPEFIGRLPVVATLNELSEEALIQILKEPKNALTKQ
YQALFNLEGVDLEFRDEALDAIAKKAMARKTGARGLRSIVEAALLDTMYDLPSMED
VEKVVIDESVIDGQSKPLLIYGKPEAQQASGEASGAGGSEGGGSEGGTSGATMSALP
TPHEIRNHLDDYVIGQEQAKKVLAVAVYNHYKRLRNGDTSNGVELGKSNILLIGPTG
SGKTLLAETLARLLDVPFTMADATTLTEAGYVGEDVENIIQKLLQKCDYDVQKAQR
GIVYIDEIDKISRKSDNPSITRDVSGEGVQQALLKLIEGTVAAVPPQGGRKHPQQEFLQ
VDTSKILFICGGAFAGLDKVISHRVETGSGIGFGATVKAKSDKASEGELLAQVEPEDLI
KFGLIPEFIGRLPVVATLNELSEEALIQILKEPKNALTKQYQALFNLEGVDLEFRDEAL
DAIAKKAMARKTGARGLRSIVEAALLDTMYDLPSMEDVEKVVIDESVIDGQSKPLLI
YGKPEAQQASGEASGAGGSEGGGSEGGSSGATMSALPTPHEIRNHLDDYVIGQEQA
KKVLAVAVYNHYKRLRNGDTSNGVELGKSNILLIGPTGSGKTLLAETLARLLDVPFT
MADATTLTEAGYVGEDVENIIQKLLQKCDYDVQKAQRGIVYIDEIDKISRKSDNPSIT
RDVSGEGVQQALLKLIEGTVAAVPPQGGRKHPQQEFLQVDTSKILFICGGAFAGLDK
VISHRVETGSGIGFGATVKAKSDKASEGELLAQVEPEDLIKFGLIPEFIGRLPVVATLN
ELSEEALIQILKEPKNALTKQYQALFNLEGVDLEFRDEALDAIAKKAMARKTGARGL
RSIVEAALLDTMYDLPSMEDVEKVVIDESVIDGQSKPLLIYGKPEAQQASGE. (SEQ ID NO: 8)

It is a synthetically designed trimer of ClpX subunits that is expressed as a single chain. As noted above, a variety of translocase constructs may be used in the present system

2. Enzymes may be coupled to the nanopore, free in solution on one side, and/or present on both sides of the nanopore

Translocase on the *cis* side (same side as the substrate protein)

In certain embodiments, for example as shown in **Figure 10**, ClpA or Clp X may be coupled to the nanopore. An engineered alpha- hemolysin/ClpP fusion protein pore may be assembled to form an active heptameric protein nanopore covalently fused at its N-terminal cap domain to the ClpX-binding domain of the ClpP heptamer complex. Fusion of the ClpX-binding domain of ClpP to the top of the nanopore will enable ClpX to assemble in solution, attach to the ClpP domain, and function on the top of nanopore

Figure 10 illustrates a fusion protein comprising the α-hemolysin pore protein subunits fused to subunits of a ClpP protein. The ClpX protein translocase can then non-covalently "dock" onto the ClpP subunits. In this embodiment the protein translocase is on the *cis* side of, and fused to, the nanopore.

In the embodiment of **Figure 10**, the axial pores of the translocase and the nanopore must be aligned in the correct orientation; that is, ClpX must be bound to the nanopore in such a way that as the protein substrate is captured from solution and driven through the ClpX central cavity, it then directly enters into the AHL (alpha hemolysin) upper lumen and is eventually forced through the entire nanopore (See **Figure 10**). To achieve this goal, a ClpX-binding domain may be fused to the head of AHL. In nature, the ClpX hexamer naturally binds to the head domains at each opening of the tetradecameric (double heptameric rings) protease ClpP and acts as a gate that only allows tagged proteins to enter the proteolytic chamber. By fusing the ClpX-binding domain of ClpP onto the head of AHL, it will enable ClpX to directly bind with high affinity atop the nanopore via a natural protein-protein interaction. Fortunately, both AHL and ClpP assemble into homoheptameric rings with strikingly similar diameters; thus, fusion of the ClpX binding domain of the ClpP monomer to the head of each AHL monomer will create a heptameric nanopore complex composed of these ClpP/AHL fusion monomers. A previous study has investigated the fusion of AHL monomers, showing that AHL is indeed tolerant of fusions at both the N and C termini. In addition, a study investigated deletions at the C-terminus of ClpP, showing that it is not a region critical for heptamer formation or ClpX binding. This data suggests that ClpP would be tolerant of an AHL fusion to its C-terminus, while fusion to the ClpP N-terminus

would almost certainly inhibit binding of ClpX as the N-terminal loops have been shown to be critical for such activity. Based on these previous studies, the AHL/ClpP fusion monomer is designed with a single AHL monomer fused to the C-terminal of a truncated ClpP monomer (separated by a flexible 5-15 Gly-Ser linker to allow each protein monomer to fold properly). These ClpP— peptide linker—AHL fusion protein DNA sequences may constructed via PCR assembly, His-tagged, and inserted inside pT7-SC1 expression vectors. Expression of the fusion protein may be done through coupled in-vitro transcription and translation using the T7-S30 expression system, used previously to express AHL fusions, and purified with Ni-NTA affinity chromatography.

Another embodiment where the translocase is on the *cis* side of the device involves the use of accessory proteins. In this case, proteins that bind to the substrate protein are used to facilitate control of movement of the substrate protein to or though the nanopore. For example, the protein present in the *trans* side (as in **Figure 1**) need not be an active translocase/unfoldase but rather another protein (for example trigger factor) that non-specifically binds to unfolded portions of polypeptides. Trigger factor is a ribosome-associated molecular chaperone and is the first chaperone to interact with a nascent polypeptide. It acts as a chaperone by maintaining the newly synthesized protein in an open conformation. Other chaperonins or heat shock proteins could be used.

This *trans* side protein (e.g. trigger factor) would thereby sequentially capture the unfolded substrate protein as it is translocated into the trans solution by the *cis* side translocase/unfoldase, preventing the substrate protein from moving back into the *cis* side.

In another embodiment, the substrate protein is provided with a factor that blocks its unfolding by the unfoldase until a predetermined state is reached. In this embodiment, both substrate protein and unfoldase are added to the *cis* side. Substrate proteins are tagged at one terminus with an unfoldase-binding motif (for example the ssrA tag for ClpX), a pore-targeting domain (for example a charged poly-peptide tail that will be pulled into the pore under an applied voltage), and an unfoldase-resistant protein (for example dihydrofolate reductase or barnase in presence of stabilizing ligands are resistant to unfolding by ClpX; see, Hoskins, JR et al. ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. PNAS August 20, 2002 vol. 99 no. 17 11037-11042). The substrate protein therefore is protected from the unfoldase with this "blocking domain" between the its folded domain and the pore targeting and unfoldase-binding motif domains.

The blocking domain protein (fused with the pore-targeting domain and unfoldase-targeting motif domain) may be chemically or enzymatically attached to the substrate protein post-translationally.

In the bulk *cis* solution, the unfoldase binds to the unfoldase-binding motif and translocates along the pore-targeting domain. Translocation of the unfoldase along the tagged substrate stops once the enzyme approaches the unfoldase-resistant protein (the "blocking domain"). When this tagged-substrate-protein/unfoldase complex is captured by the nanopore, unfolding and translocation of the blocking domain is initiated. It is catalyzed by the extra destabilizing forces imparted by the voltage at the pore and/or other protein translocases/unfoldases present in the *trans* solution that interact with the tagged substrate protein tail after capture and threading of the substrate tail through the pore. Unfolding and translocation of the entire substrate protein by the cis-bound unfoldase through the nanopore is then possible after the pore-activated unfolding of the blocking domain is catalyzed.

An example of a blocking domain sequence strategy example sequence is as follows:

N-terminus—substrate protein—blocking domain—charged tail—unfoldase-binding motif—C-terminus.

Translocase on the *trans* side

ClpX complexes are placed in solution on the *trans* side of the nanopore, and the substrate protein dissolved on *cis* side of the nanopore will be forced to thread through the nanopore beginning at the N or C terminus (for example, voltage-driven by engineering a few charged amino acids into the protein terminus, such as 5-10 Asp, Lys or Arg residues) where the ClpX complex will capture this tagged polypeptide tail, and begin mechanically pulling/translocating the substrate down through the nanopore. Native ClpX binding, unfolding, and translocation activity of tagged proteins is used to control the movement of proteins through a nanopore sensor for subsequent analysis. A wild type protein nanopore or other solid-state nanopore may be utilized if ClpX or another translocase such as ClpA is placed on the *trans* side of the device, in solution, where it is allowed to capture tagged protein tails that were threaded from the *cis* side through the nanopore to the *trans* side solution (that is, "fishing" for the ClpX). Upon capture of this tagged-tail, the ClpX complex would be able to begin mechanically pulling on the protein from across the pore, until it is eventually able to unfold and thread the entire polypeptide through the pore and into the *trans* side solution. The initial

threading of the tagged-tail could be accomplished via the addition of several charged residues proximal to the tag sequence on the N or C terminus of the target protein, wherein a voltage differential would drive the charged tail through the nanopore, making it available to "fish" for ClpX.

As shown, e.g. in Example 1 and Example 2 below, the protein translocase is added in solution to the *trans* side of the chamber; it serves to unfold the protein to be translocated by pulling it through the nanopore, which is too narrow to permit passage of the folded protein. The same or a different protein translocase may be located on the *cis* side of the nanopore for unfolding the protein.

3. Nanopore sensing

The method of sensing in the present method and device may involve measuring one, two, three, four or five or more characteristics of the protein. The one or more characteristics are preferably selected from (i) the length of the protein, (ii) the identity of the protein, (iii) the sequence of the protein, (iv) the secondary or tertiary structures of the protein and (v) whether or not the protein is modified. Any combination of (i) to (v) may be measured in accordance with the invention.

For characteristic (i), the length of the protein may be measured using the number of interactions between the protein and the pore, and/or the dwell time of the protein as it translocates through the pore.

For characteristic (ii), the identity of the protein may be measured in a number of ways. The identity of the protein may be measured in conjunction with measurement of the sequence of the protein or without measurement of the sequence of the protein. The former is straightforward; the protein is sequenced and thereby identified. The latter may be done in several ways. For instance, the presence of a particular motif in the protein may be measured (without measuring the remaining sequence of the protein). Alternatively, the measurement of a particular electrical and/or optical signal in the method may identify the protein as coming from a particular source.

For characteristic (iii), the sequence of the protein can be determined as described herein. The sequence may be determined on an individual amino acid residue-by-residue basis, or may be read in blocks of amino acids, which may be mapped to known protein

sequences, in a manner analogous to re-sequencing of DNA. Thus the method need not resolve each individual amino acid, but rather one could just resolve "words" or blocks/chunks of amino acids (e.g. 2 to 10 aa) that would still enable identification of the protein/polypeptide sequence.

For characteristic (iv), the secondary and tertiary structures may be measured in a variety of ways. For instance, if the method involves an electrical measurement, the secondary structure (e.g. detection of an alpha helix region versus a loop region or a beta sheet region) may be measured using a change in dwell time or a change in current flowing through the pore.

For characteristic (v), the presence or absence of any modification may be measured. The method preferably comprises determining whether or not the protein is modified by methylation phosphorylation, oxidation, by damage (e.g. misfolding or covalent modification of an amino acid), by glycosylation or with one or more labels, tags or spacers. Specific modifications will result in specific interactions with the nanopore which can be measured using the methods described below.

As discussed below, the present methods and device can extract protein sequence information from the protein being translocated by analysis of the ionic current measured through the nanopore. This is partly dependent on the use of sensitive electronics such as a patch clamp amplifier, a finite state machine, and signal processing such as weighted averaging, all described below. In addition, the present methods involve analysis of various current states that have now been found to be associated with the transit of the protein through the nanopore and the concomitant current blockage, unblockage or modulation. As described below, the signal detected by enzyme-mediated protein traversal will enter a so-called "ramping state" which is characteristic of binding of the enzyme to the substrate protein, then a series of separate amplitude transitions as the protein translocates through the nanopore, followed by an open state equivalent to current through the pore prior to translocation. In summary, using the exemplified experimental setup, one can observe distinct states (see **Fig. 2**) of

(i) open channel current, prior to translocation – about 30-35 pA, or 32-36 pA;

(ii) decrease to about 11-15 (e.g. about 13-15) pA upon capture of the protein to be translocated (i.e. the protein blocks the nanopore opening;

(iii) decrease to or below 10 pA and various amplitude changes (including a "ramping" effect discussed below), as the enzyme binds the protein and translocates along the protein tail toward the nanopore (i.e. the enzyme blocks more of the nanopore opening);

(iv) after unfolding of the protein, the entire protein is translocated through the nanopore, generating a unique current pattern;

(v) return to open channel state.

Importantly, state (iv) (**Fig. 2A and 2B**) presents a current pattern that can be correlated to protein structure. As described below, the artificial linkers used in the examples showed different current amplitudes and duration. When analyzing a protein to determine an unknown feature such as sequence, one may correlate observed changes in amplitude and RMS noise to the amino acid-dependent features of the translocated protein. This includes but is not limited to tertiary and secondary structures, amino acid sequence and post-translational modifications

The nanopore biosensing technology is based on the blockage of ionic current that occurs when a molecule translocates through and/or interacts with a pore under an applied voltage. The current blockade is thus dependent upon the applied voltage and the properties of the interacting molecule (for example, charge, size).

Time-dependent transport properties of the nanopore aperture may be measured by any suitable technique. The transport properties may be a function of the medium used to transport the polypeptide, solutes (for example, ions) in the liquid, the polypeptide (for example, chemical structure of the monomers), or labels on the polypeptide. Exemplary transport properties include current, conductance, resistance, capacitance, charge, concentration, optical properties (for example, fluorescence and Raman scattering), and chemical structure. Desirably, the transport property is current.

Exemplary means for detecting the current between the *cis* and the *trans* chambers have been described in WO 00/79257, U.S. Patent Nos. 6,46,594, 6,673 6,673,615, 6,627,067, 6,464,842, 6,362,002, 6,267,872, 6,015,714, 6,428,959, 6,617,113 and 5,795,782 and U.S. Publication Nos. 2004/0121525, 2003/0104428, and 2003/0104428, and can include, but are not limited to, electrodes directly associated with the channel or pore at or near the pore aperture, electrodes placed within the *cis* and the *trans* chambers, ad insulated

glass micro-electrodes. The electrodes may be capable of, but not limited to, detecting ionic current differences across the two chambers or electron tunneling currents across the pore aperture or channel aperture. In another embodiment, the transport property is electron flow across the diameter of the aperture, which may be monitored by electrodes disposed adjacent to or abutting on the nanopore circumference. Such electrodes can be attached to an Axopatch 200B amplifier for amplifying a signal.

In one embodiment, the medium is electrically conductive. In another preferred embodiment, the medium is an aqueous solution. In another preferred embodiment, the method further comprises the steps of measuring the electrical current between the two pools; comparing the electrical current value (I_1) obtained at the first time the first polarity was induced with the electrical current value (I_2) obtained at the time the second time the first polarity was induced; and determining the difference between I_1 and I_2 thereby obtaining a difference value δI. In another preferred embodiment the method further comprises the steps of measuring the electrical current between the two pools; comparing the electrical current value (I_1) obtained at the first time the first polarity was induced with the electrical current value (I_2) obtained at a later time and determining the difference between I1 and I_2 thereby obtaining a difference value δI.

In an alternative embodiment, the method further comprises the steps of providing reagents that initiate enzyme activity; introducing the reagents to the pool comprising the polypeptide complex; and incubating the pool at a suitable temperature. In another preferred embodiment, the reagents are selected from the group consisting of an activator and a cofactor

4. Manufacture of Nanopore Thin Film Devices

Single-channel (nanopore) thin film devices and methods for using the same are provided. The subject devices typically comprise a mixed-signal semiconductor wafer, at least one electrochemical layer, the electrochemical layer comprising a semiconductor material, such as silicon dioxide or the like, wherein the semiconductor material further comprises a surface modifier, such as a hydrocarbon, wherein the electrochemical layer defines a plurality of orifices, the orifices comprising a chamber and a neck and wherein the chamber of the orifices co-localize with a first metal composition of the mixed-signal

semiconductor wafer, wherein a portion of the orifice is plugged with a second metal, for example, silver, wherein the second metal is in electronic communication with the first metal, and wherein the orifice further comprises a thin film, such as a phospholipid bilayer, the thin film forming a solvent-impermeable seal at the neck of the orifice, the thin film further comprising a pore, and wherein the orifice encloses an aqueous phase and a gas phase.

In another preferred embodiment, the semiconductor material is selected from the group consisting of silicon dioxide (SiO2), silicon oxy nitride (SiON), silicon nitride (SiN), metal oxide, and metal silicate. In another preferred embodiment, the semiconductor material is silicon dioxide. In another preferred embodiment, the surface modifier is a hydrocarbon. In another preferred embodiment, the metallization composition is selected from the group consisting of nickel, gold, copper, and aluminum. In a most preferred embodiment, the metal is silver. In a preferred embodiment, the thin film is a molecular bilayer. In another preferred embodiment, the thin film is a phospholipid bilayer. In one alternative embodiment, the orifice is between 0.5 and 3 μm in size. In a preferred embodiment, the orifice is between 1 and 2 μm in size. In a most preferred embodiment, the orifice is between 1.25 and 1.5 μm in size. In another preferred embodiment, the pore is a biological molecule. In another preferred embodiment, the biological molecule is selected from the group consisting of an ion channel, a nucleoside channel, a peptide channel, a sugar transporter, a synaptic channel, a transmembrane receptor, and a nuclear pore. In a most preferred embodiment, the biological molecule is alpha-hemolysin. In a preferred embodiment, the pore aperture is between about 1 and 10 nm in size. In another preferred embodiment, the pore aperture is between about 1 and 4 nm in size. In a most preferred embodiment, the pore aperture is between about 1 and 2 nm in size. In an alternative most preferred embodiment the pore aperture is between about 2 and 4 nm in size.

Biological nanopores have utility in detection of polypeptides but, due to the low current used (approximately in the tens of picoamps). Detection using high-through put of a single nanopore sequencing device may be limited to approximately 1000 amino acid residues per second. Manufacturing arrays of biological nanopores that can operate independently of each other, such as used in the manufacture of very large arrays of integrated circuits, allow a very large scale array of nanopores to perform millions of biochemical reactions and analyses in a single second.

A variety of nanopores may be used in the present system. In one embodiment, the

pore or channel is shaped and sized having dimensions suitable for passaging a polymer. In another embodiment, the pore or channel accommodates a substantial portion of the polymer. In a preferred embodiment, the polymer is a polypeptide. The pore or channel may also be a pore molecule or a channel molecule and comprise a biological molecule, or a synthetic modified molecule, or altered biological molecule, or a combination thereof. Such biological molecules are, for example, but not limited to, an ion channel, such as α-hemolysin ,MspA, a nucleoside channel, a peptide channel, a sugar transporter, a synaptic channel, a transmembrane receptor, such as GPCRs and the like, a receptor tyrosine kinase and the like, a T-cell receptor, a MHC receptor, a nuclear receptor, such as a steroid hormone receptor, a nuclear pore, synthetic variants, chimeric variants, or the like.

In one preferred embodiment the biological molecule is α-hemolysin. In another preferred embodiment the biological molecule is MspA (Mycobacteria smegmatis porin A). In yet another preferred embodiment the pore is a solid-state pore.

In an alternative, the pore or channel comprises non-enzyme biological activity. The pore or channel having non-enzyme biological activity can be, for example, but not limited to, proteins, peptides, antibodies, antigens, nucleic acids, peptide nucleic acids (PNAs), locked nucleic acids (LNAs), morpholinos, sugars, lipids, glycosyl phosphatidyl inositols, glycophosphoinositols, lipopolysaccharides or the like. The compound can have antigenic activity. The compound can have selective binding properties whereby the polymer binds to the compound under a particular controlled environmental condition, but not when the environmental conditions are changed. Such conditions can be, for example, but not limited to, change in $[H^+]$, change in environmental temperature, change in stringency, change in hydrophobicity, change in hydrophilicity, or the like.

In another embodiment, the pore or channel further comprises a linker molecule, the linker molecule selected from the group consisting of a thiol group, a sulfide group, a phosphate group, a sulfate group, a cyano group, a piperidine group, an Fmoc group, and a Boc group. In another embodiment the compound is selected from the group consisting of a bifunctional alkyl sulfide and gold.

In one embodiment, the pore is sized and shaped to allow passage of an activator, wherein the activator is selected from the group consisting of ATP, NAD^+, $NADP^+$, diacylglycerol, phosphatidylserine, eicosinoids, retinoic acid, calciferol, ascorbic acid,

neuropeptides, enkephalins, endorphins, 4-aminobutyrate (GABA), 5-hydroxytryptamine (5-HT), catecholamines, acetyl CoA, S-adenosylmethionine, and any other biological activator. In another embodiment the pore is sized and shaped to allow passage of a cofactor, wherein the cofactor is selected from the group consisting of Mg^{2+}, Mn^{2+}, Ca^{2+}, ATP, NAD^{+}, $NADP^{+}$, and any other biological cofactor.

The array elements may be manufactured in a step-wise parallel manner, similar to the manufacture of transistors on integrated circuits. All, or most, of the similar layers of each array element are created in a sequence of single process steps that simultaneously take place on all, or most, of the array elements.

There appears to be no simple way to synchronize the activities of separate molecules of biological reagents, so each element in the array should be able to act independently of the other elements. This may be accomplished by including a digital logic circuit with each single biological nanopore that implements a finite state machine that controls and senses the biochemical state of the complex off single (or multiple) molecules associated with the biological nanopore. The finite state machine allows low latency control of the complex of molecules associated with the biological nanopore and at the same time can store information gathered for retrieval at another time.

In order that each of the biological nanopore elements in an array may be in communication with one another using a minimum number of wired connections, a serial interface and addressable logic can be used to multiplex the large amount of data entering and exiting the array.

Not all of the array elements may have a thin film or bilayer across their respective orifice. The capacitance of the membrane present in the nanopore as measured by the finite state machine can be used to detect the presence of non-functional array elements. If it is subsequently determined that a proportion of array elements lack a thin film or bilayer is greater when compared with a proportion that is preferred, then the step of overlaying the membrane such as TEFLON film and lipid coat can be repeated.

An electrode, for example a grounded macroscopic AgCl electrode, may be placed in contact with second solution. When membranes such as bilayers are positioned in place across all the functionable orifices, no ion current will flow from second solution to first solution. A predetermined amount of pore molecule or channel molecule, such as for

example, alpha-hemolysin toxin or MspA, is added to second solution. The concentration of pore molecule or channel molecule is sufficient to form a single channel in any of the thin films or bilayers in approximately, for example, fifteen minutes. The time to form such channels can be for example, between one-half minute and one hour, for example, about one-half minute, one minute, two minutes, three minutes, four minutes, five minutes, seven minutes, ten minutes, fifteen minutes, twenty minutes, twenty five minutes, thirty minutes, thirty five minutes, forty minutes, forty five minutes, fifty minutes, fifty five minutes, sixty minutes, or any time therebetween. The time for formation can be altered by an operator by several factors or parameters, for example, increasing or decreasing the ambient or incubation temperature, increasing or decreasing the concentration of salt in second solution or first solution, placing a potential difference between the first solution and the second solution, or other methods know to those of skill in the art. The finite state machine can detect and/or sense formation of a single channel in its corresponding bilayer by reacting to the flow of current (ions) through the circuit, the circuit comprising the macroscopic electrode, the second solution, the single nanopore or channel molecule, first solution, and the metal electrode for any given array element.

Formation of biological channels is a stochastic process. Once a single channel has formed in a given array element bilayer, it is preferred that the chance that a second channel so forming therein is reduced or preferably, eliminated. The probability of second channel insertion can be modulated with applied potential, that is, potential difference, across the bilayer. Upon sensing a single channel, a finite state machine may adjust the potential on the metal electrode to decrease the possibility of second channel insertion into the same bilayer.

Despite the precautions taken in the previous step(s) a second channel may form in a given bilayer. The finite state machine can detect the formation of the second channel. A pulse of precisely controlled low pressure can force one out of two channels allowing a single channel to remain embedded in the bilayer.

In the course of using the biological nanopore for biochemical actuation and detection, the pore may become permanently obstructed. A finite state machine can detect and sense this obstructed state and can remove the blocked channel from the bilayer by inactivating the heating element thereby applying suction (reduced pressure) upon the bilayer.

In an alternative embodiment, each array element may comprise a gold electrode surrounding the orifice. This gold electrode may serve to activate chemical reagents using reduction or oxidation reactions and that can act specifically at the location of a specific orifice.

The finite state machine can be created for example using state-of-the-art commercially available 65nm process technology, for example from Taiwan Semiconductor Manufacturing Company, Taiwan). A 600 x 600 array of nanopores can perform 360,000 biochemical reaction and detection/sensing steps at a rate of 1000 Hz. This may enable sequencing of polynucleotides, for example, to proceed at a rate of 360 million baser per second per 1 cm x 1 cm die cut from the semiconductor wafer.

Exemplary means for applying an electric field between the *cis*- and *trans*-chambers are, for example, electrodes comprising an immersed anode and an immersed cathode, that are connected to a voltage source. Such electrodes can be made from, for example silver chloride, or any other compound having similar physical and/or chemical properties.

5. Equipment

In the working examples, a patch-clamp amplifier, Molecular Devices AxoPatch 200B, regulates the applied voltage and measures the ionic current through the channel. The data are recorded using the Molecular Devices Digidata 1440A digitizer, sampled at 50 kHz and low-pass filtered at 5 kHz with a four-pole Bessel filter. One of the station uses a different patch clamp, the A-M systems Model 2400.

Other equipment may be used, as follows:

Control Logic: Hardware and Software

The voltage control logic is programmed using a finite state machine (FSM) within the LabVIEW 8 software. The FSM logic is implemented on a field-programmable gate array (FPGA) hardware system, National Instruments PCI-7831R. An FPGA is a reconfigurable hardware platform that permits fast measurement and voltage reaction times (1 μsec output sample time). An FSM is a logic construct in which program execution is broken up into a series of individual states. Each state has a command associated with it, and transitions between states are a function of system measurements. Measurements of the pore

current are processed and passed to the FSM as inputs. Changes in the FSM control logic are made as necessary, without the need to re-compile and re-route the design to run on the FPGA. This achieves a balance between speed and flexibility, by enabling the system to react to events on the order of a microsecond, while also allowing for the control logic to be reconfigured as necessary between experiments.

The finite state machine can be used to detect and control binding of a molecule to a polymer. The molecule is a protein, preferably the protein is an enzyme. The finite state machine can also detect a polymer compound having a structural element that inhibits transposition of the polymer compound through a nanopore. In one embodiment, the polymer compound comprises a peptide nucleic acid.

The finite state machine can control binding of a molecule to a polymer at a rate of between about 5 Hz and 2000 Hz. The finite state machine can control binding of a molecule to a polymer at, for example, about 5 Hz, at about 10 Hz, at about 15 Hz, at about 20 Hz, at about 25 Hz, at about 30 Hz, at about 35 Hz, at about 40 Hz, at about 45 Hz, at about 50 Hz, at about 55 Hz, at about 60 Hz, at about 65 Hz, at about 70 Hz, at about 75 Hz, at about 80 Hz, at about 85 Hz, at about 90 Hz, at about 95 Hz, at about 100 Hz, at about 110 Hz, at about 120 Hz, at about 125 Hz, at about 130 Hz, at about 140 Hz, at about 150 Hz, at about 160 Hz, at about 170 Hz, at about 175 Hz, at about 180 Hz, at about 190 Hz, at about 200 Hz, at about 250 Hz, at about 300 Hz, at about 350 Hz, at about 400 Hz, at about 450 Hz, at about 500 Hz, at about 550 Hz, at about 600 Hz, at about 700 Hz, at about 750 Hz, at about 800 Hz, at about 850 Hz, at about 900 Hz, at about 950 Hz, at about 1000 Hz, at about 1125 Hz, at about 1150 Hz, at about 1175 Hz, at about 1200 Hz, at about 1250 Hz, at about 1300 Hz, at about 1350 Hz, at about 1400 Hz, at about 1450 Hz, at about 1500 Hz, at about 1550 Hz, at about 1600 Hz, at about 1700 Hz, at about 1750 Hz, at about 1800 Hz, at about 1850 Hz, at about 1900 Hz, at about 950 Hz, and at about 2000 Hz. In a preferred embodiment, the finite state machine can control binding of a molecule to a polymer at a rate of between about 25 Hz and about 250 Hz. In amore preferred embodiment the finite state machine can control binding of a molecule to a polymer at a rate of between about 45 Hz and about 120 Hz. In a most preferred embodiment the finite state machine can control binding of a molecule to a polymer at a rate of about 50 Hz.

Moving Average Filter

Every 5.3 μsec, the FPGA samples the ionic current and computes a windowed mean amplitude, using a window size of 0.75 msec. If the mean enters a chosen threshold range, the FPGA detects entry and continues to monitor the mean, re-checking the threshold every 0.2 msec. If the mean remains within the threshold range for four consecutive checks, the FSM logic diagnoses the blockade as an event type known to be consistent with the chosen threshold.

In the absence of a change in voltage, the expected time delay between the start of an event and diagnosis of an event is 1.35 msec; 0.75 msec for the windowed mean to first enter the threshold, and 0.6 msec for three more confirmed tests. In practice, the diagnosis time ranges from 1.1 to 2.5 msec. The mean filter was implemented in our invention's initial demonstration

Exponentially-Weighted Moving Average Filter

To improve the FSM's robustness to false detections of terminal steps, an exponentially-weighted moving average (EWMA) filter may be used to replace the mean filter. The EWMA filter represents a digital implementation of an analog RC filter commonly used for signal smoothing in electrical engineering applications. The filter calculates a moving average that places exponentially less significance on past samples and allows the filtered signal to better track the real signal. EWMA filtering also performs signal smoothing more efficiently than a simple moving average due to its recursive implementation:

$$i^{bar}(t) = (1-a)i^{bar}(t) + ai(t-1), \qquad (1)$$

where i and i^{bar} are unfiltered and filtered current signals, respectively, and t is the sample number. Filtering the data from the terminal step detection experiments offline, with $\alpha = 0.9$, showed a substantial improvement in robustness to false positives over the mean filter. As with the mean filter, four consecutive threshold tests will be used for event diagnosis, waiting 0.2 msec between threshold tests.

In the absence of a change in voltage, the expected time delay between the start of an event and diagnosis of an event is 0.7 msec; 0.1 msec for the EWMA to first enter the

threshold, and 0.6 msec for three more confirmed tests. More rigorous evaluation of EWMA detection times will be part of our ongoing work.

Voltage Control Using FSM/FPGA

The nanopore system can be setup in a 0.3 mM KCl solution. A patch-clamp amplifier, Molecular Devices AxoPatch 200B, regulates the applied voltage and measures the ionic current through the channel. The data are recorded using the Molecular Devices Digidata 1440A digitizer, sampled at 50 kHz and low-pass filtered at 5 kHz with a four-pole Bessel filter.

The voltage control logic is programmed using a FSM within the LabVIEW 8 software. The FSM logic is implemented on a field-programmable gate array (FPGA) hardware system, National Instruments PCI-7831R. An FPGA is a reconfigurable hardware platform that permits fast measurement and voltage reaction times (1 μsec output sample time). An FSM is a logic construct where program execution is broken up into a series of individual states. Each state has a command associated with it, and transitions between states are a function of system measurements. Measurements of the pore current are processed and passed to the FSM as inputs. Changes in the FSM control logic are made as necessary, without the need to re-compile and reroute the design to run on the FPGA. This achieves a balance between speed and flexibility, by enabling the system to react to events on the order of a microsecond, while also allowing for the control logic to be reconfigured as necessary between experiments.

6. Exemplary applications

Applications and/or uses of the invention disclosed herein may include, but not be limited to the following: 1). Assay of protein-nucleic acid complexes in mRNA, rRNA, and DNA. 2). Assay of the presence of microbe or viral content in food and environmental samples via peptide analysis. 3). Identification of microbe or viral content in food and environmental samples via peptide analysis. 4). Identification of pathologies via peptide analysis in plants, human, microbes, and animals. 5). Assay of peptides in medical diagnosis. 6). Forensic assays.

The present nanopore device can be used to monitor the turnover of enzymes such as proteases, kinases, and phosphatases, which have important applications in cell proliferation.

The present nanopore device can function as a biosensor to monitor the interaction between soluble substances such as enzyme substrates or signaling molecules. Examples include blood components such as glucose, uric acid and urea, hormones such as steroids and cytokines, and pharmaceutical agents that exert their function by binding to receptor molecules.

The present nanopore device can monitor in real time the function of important biological structures such as ribosomes, and perform this operation with a single functional unit.

The present methods and devices may also be used to detect and quantify altered protein expression, absence/presence versus excess, expression of proteins or to monitor protein levels during therapeutic intervention. The amount of protein in a given sample may be estimated using an array of nanopore devices according to the present invention. Polypeptides or proteins to be translocated can also be utilized as markers of treatment efficacy against the diseases noted above and other brain disorders, conditions, and diseases over a period ranging from several days to months. Qualitative or quantitative methods for this comparison are well known in the art.

Diagnostics

The polypeptides, fragments, oligopeptides, and PNAs that may be translocated by the present system may be used to detect and quantify altered protein expression, absence/presence versus excess, expression of proteins or to monitor protein levels during therapeutic intervention. Conditions, diseases or disorders associated with altered expression include idiopathic pulmonary arterial hypertension, secondary pulmonary hypertension, a cell proliferative disorder, particularly anaplastic oligodendroglioma, astrocytoma, oligoastrocytoma, glioblastoma, meningioma, ganglioneuroma, neuronal neoplasm, multiple sclerosis, Huntington's disease, breast adenocarcinoma, prostate adenocarcinoma, stomach adenocarcinoma, metastasizing neuroendocrine carcinoma, nonproliferative fibrocystic and proliferative fibrocystic breast disease, gallbladder cholecystitis and cholelithiasis, osteoarthritis, and rheumatoid arthritis; acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis,

amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, benign prostatic hyperplasia, bronchitis, Chediak-Higashi syndrome, cholecystitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, chronic granulomatous diseases, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polycystic ovary syndrome, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, severe combined immunodeficiency disease (SCID), Sjogren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, hemodialysis, extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infection; a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, gynecomastia; actinic keratosis, arteriosclerosis, bursitis,cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal, nocturnal hemoglobinuria, polycythemia vera, primary thrombocythemia, complications of cancer, cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In another aspect, the polynucleotide of the invention.

The polypeptides, fragments, oligopeptides, and PNAs, or fragments thereof, may be used to detect and quantify altered protein expression, absence/presence versus excess, expression of proteins or to monitor protein levels during therapeutic intervention. Disorders associated with altered expression include akathesia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, ataxias, bipolar disorder, catatonia, cerebral palsy,

cerebrovascular disease Creutzfeldt-Jakob disease, dementia, depression, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, muscular dystrophy, neuralgias, neurofibromatosis, neuropathies, Parkinson's disease, Pick's disease, retinitis pigmentosa, schizophrenia, seasonal affective disorder, senile dementia, stroke, Tourette's syndrome and cancers including adenocarcinomas, melanomas, and teratocarcinomas, particularly of the brain. These polypeptides or proteins can also be utilized as markers of treatment efficacy against the diseases noted above and other brain disorders, conditions, and diseases over a period ranging from several days to months. Qualitative or quantitative methods for this comparison are well known in the art.

For example, the polypeptide or peptide may be labeled by standard methods and added to a biological sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed and the amount of label (or signal) is quantified and compared with a standard value. If the amount of label in the patient sample is significantly altered in comparison to the standard value, then the presence of the associated condition, disease or disorder is indicated.

In order to provide a basis for the diagnosis of a condition, disease or disorder associated with protein expression, a normal or standard expression profile is established. This may be accomplished by combining a biological sample taken from normal subjects, either animal or human, with a peptide tag under conditions for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a substantially purified target sequence is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular condition is used to diagnose that condition.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies and in clinical trial or to monitor the treatment of an individual patient. Once the presence of a condition is established and a treatment protocol is initiated, diagnostic assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate the level that is observed in a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

Example 1: Construction of a nanopore device for monitoring protein translocation

The nanopore device of this example is diagrammed in **Figure 1A,** which represents the set up used. A nanopore sensor was prepared with a single AHL pore embedded in a lipid bilayer separating two Teflon® PTFE polymer wells each containing 100 µl of 0.2 M KCl solution (30°C). Voltage is applied between the wells (*trans* side +180 mV), causing ionic current flow through the channel. Current diminishes in the presence of a captured protein molecule.

Briefly, for each experiment a single AHL nanopore was inserted into a 30 µm diameter lipid bilayer that separates two wells (termed *cis* and *trans*) that each contained 100µl of PD buffer (pH7.6). A covalently-linked trimer of an N-terminal truncated ClpX variant (ClpX-ΔN_3) was used for all ClpX nanopore experiments. The ClpX-ΔN_3 BLR expression strain was obtained from Andreas Martin (UC Berkeley). ClpX protein expression was induced at an OD 600 of ~1 by addition of 0.5 mM IPTG, and incubated at 23°C with shaking for 3-4 hours. Cultures were pelleted, resuspended in lysis buffer (50 mM NaH2PO4 pH 8, 300 mM NaCl, 100 mM KCl, 20 mM imidazole, 10% glycerol, 10 mM BME) and lysed via vortexing with glass beads. After centrifugation and filtration of the lysate, the protein was purified on a Ni2+ -NTA affinity column (Thermo) and an Uno-Q anion exchange column (Bio-Rad).

Both *cis* compartment and *trans* compartment are filled with 100 µl of a 200 mM KCl buffer optimized for ClpX function. This buffer was supplemented with 5 mM ATP as indicated. A patch-clamp amplifier (Axopatch 200B, Molecular Devices) applied a constant 180 mV potential between two Ag/AgCl electrodes (*trans* side +) and recorded ionic current through the nanopore as a function of time. Substrate proteins were added to the *cis* solution at ~1 µM final concentration, while ~100 nM ClpX was present in the *trans* solution.

A constant 180 mV potential was applied across the bilayer and ionic current was measured through the nanopore between Ag/AgCl electrodes in series with an integrating patch clamp amplifier (Axopatch 200B, Molecular Devices) in voltage clamp mode. Data were recorded using an analog-to-digital converter (Digidata 1440A, Molecular Devices) at 100 kHz bandwidth in whole-cell configuration (β=1) then filtered at 5 kHz using an analog

low-pass Bessel filter. Experimental conditions were prepared by the daily preparation of PD/ATP 5 mM and PD/ATP 4 mM. ClpX was diluted 1:10 in PD/ATP 5 mM for a final concentration of 30-100 nM ClpX in 4.5 mM ATP final. ClpX solution was used to fill the entire system before isolation of a single AHL nanopore. Upon insertion, the *cis* well was perfused with ~6 mL PD/ATP 4 mM. Experiments were conducted at 30 °C with 1-2 μM substrate added to the *cis* well. Protein substrate capture events were ejected with reserve polarity due to pore clogs or after a predetermined duration. Voltage-induced translocations were frequently ejected to prevent clogging and to increase the efficiency of data collection. A single nanopore experiment is defined as the time during which ionic current data were acquired from one AHL nanopore in an intact bilayer before termination by bilayer rupture, loss of channel conductance or completion of a preset number of translocation events.

Example 2: Engineering protein S1, S2-35 and S2-148 for translocation

The substrate proteins used for translocation are schematically illustrated in **Figure 1C**, which shows (i) S1, a protein bearing a single N-terminal Smt3-domain coupled to a 65-amino-acid-long charged flexible segment capped at its carboxy-terminus with the 11 amino acid ClpX-targeting domain (ssrA tag); (ii) S2-35, similar to S1 but appended at its N-terminus by a 35 amino acid linker and a second Smt3 domain; (iii) S2-148, identical to S2-35 except for an extended 148 amino acid linker between the Smt3 domains.

For our initial experiments, we used a modified version of the ubiquitin-like protein Smt3. Smt3 is comprised of ~100 amino acids arranged into four ß-strands and a single α-helix. We further engineer Smt3 into S1. To construct substrate protein S1, DNA encoding the 76 amino acid tail (GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSDGDGDSDGDDAANDENYALAA) (SEQ ID NO: 1) was constructed by polymerase chain reaction (PCR) and cloned into pET-SUMO vector (Invitrogen) at the T/A-cloning site, fusing the tail sequence onto the Smt3 sequence 3' end.

The 76 amino acid tail contained about 10 negatively charged residues and the 11 amino acid Clpx binding domain ssrA.

To facilitate nanopore analysis, the engineered Smt3 protein, S1, was modified in two ways, 1) It was appended with a 65-amino-acid-long glycine/serine tail including 13

interspersed negatively charged aspartate residues (SEQ ID NO: 6). This unstructured polyanion was designed to promote capture and retention of S1 in the electric field across the nanopore. Based on its crystal structure[13], the Smt3 folded domain is predicted to sit on top of the AHL vestibule (**Fig. 1B**). 2) The appended polyanion was capped at its C-terminus with the ssrA tag, an 11 amino acid ClpX-targeting motif[14]. This ssrA peptide tag allowed ClpX to specifically bind to the C-terminus of the protein when it threaded through the pore into the *trans* compartment.

S2-35 (SEQ ID NO:5) was constructed by PCR-based addition of DNA encoding the 35 amino acid linker (GGSGSGGSGSGGSGSQNEYRSGGSGSGGSGSGGSG) (SEQ ID NO: 2) to the 5' end of the S1 Smt-3 sequence. This linker-modified S1 gene was then cloned into pE-SUMO vector (LifeSensors) at the BsaI site, fusing the added linker and S1 sequence to the 3' end of the pE-SUMO Smt3 sequence.

DNA for the S2-148 (SEQ ID NO: 6) linker addition (GGSGSAGSGASGSSGSEGSGASGSAGSGSAGSRGSGASGSAGSGSAGSGGAEAAKE AAKEAAKEAAKEAAKAGGSGSAGSAGSASSGSDGSGASGSAGSGSAGSKGSGASGS AGSGSSGS) (SEQ ID NO:3) was constructed by PCR, and cloned into the S2-35 vector within the 35 amino acid linker region by the Gibson assembly method. These engineered proteins were expressed in E. coli strain BL21 (DE3)*. Expression was induced at ~0.6 OD 600 by addition of 0.5 mM IPTG, and incubated at 37°C with shaking for 4-6 hours. Cultures were pelleted, resuspended in lysis buffer and lysed via vortexing with glass beads. After centrifugation and filtration of the lysate, the protein was purified on a Ni^{2+} -NTA affinity column (Thermo).

Example 3: Detection of translocation of protein S1

A representative ionic current trace for capture and translocation of protein S1 in the presence of ClpX and ATP is shown in **Figure 2A**. **Fig. 2A** shows the ionic current traces during S1 translocation. (i) Open channel current through the AHL nanopore under standard conditions (~34 ± 2 pA, RMS noise 1.2 ± 0.1 pA). (ii) Capture of the S1 substrate. Upon protein capture, the ionic current drops to ~14 pA (~0.7 pA RMS noise). (iii) ClpX-mediated ramping state. The ionic current decreases to below 10 pA and is characterized by one or more gradual amplitude transitions. This pattern is only observed in the presence of ClpX

and ATP (*trans* compartment). (iv) Smt3 domain unfolding and translocation through the nanopore (~3.8 pA, 1.7 pA RMS noise). (v) Return to open channel current upon completion of substrate translocation to the *trans* compartment, From the open channel current of ~34 ± 2 pA (**Fig. 2A**, i), S1 capture resulted in a current drop to ~14 pA (**Fig. 2A**, ii). This stable current lasted for tens of seconds and was observed in the presence or absence of ClpX and ATP added to the *trans* compartment (**Fig. 4**). This is consistent with the Smt3 structure held stationary atop the pore vestibule by electrical force acting on the charged polypeptide tail in the pore electric field. In the presence of ClpX and ATP, this initial current state was often followed by a progressive downward current ramp reaching an average of ~10 pA with a median duration of 4.3 seconds (**Fig 2A**, iii and **Fig. 5**). This current ramp was observed with protein S1 a total of 45 times over ~5.5 hours of experimentation when ClpX and ATP were present; in contrast, the ramp was never observed following state ii when ClpX and ATP were absent from the *trans* solution over ~2.3 hours of experimentation. In a majority of events, the ClpX-dependent ramping state terminated with an abrupt ionic current decrease to about 3pA (**Fig. 2A**, iv). The median duration for state iv was ~700 ms (**Fig. 3A**) before it ended in a rapid increase to open channel current (**Fig. 2A**, v).

Based on these data we hypothesized that ClpX served as a molecular machine that used chemical energy derived from ATP hydrolysis to pull the S1 protein through the nanopore. This process was intermittently assisted by electrical force as charged amino acids entered the pore electric field.

Figure 2B illustrates the working model of ClpX-mediated translocation of S1. Cartoons i-v correspond to ionic current states i-v in panel a. Proposed steps in this process are diagrammed in **Figure 2B**: i) open channel; ii) Protein S1 capture by the pore with the Stm3 segment perched above the vestibule with the slender, charged polypeptide tail segment extended into the pore lumen, and the ssrA tag in the *trans* compartment. In this ionic current state ClpX is not bound to S1 or, alternatively, has bound but is still distant from the pore; iii) ClpX advances along the S1 strand toward the *trans*-side orifice of the AHL pore until it makes contact. Ionic current decreases due to proximity of ClpX to the pore; iv) under combined force exerted by ClpX and the pore electric field, the Stm3 structure atop the pore is sequentially denatured thus allowing the polypeptide to advance relative to the nanopore. In this state, the ionic current has decreased because larger amino acids (or Smt3 secondary structures) have entered the pore lumen. This ionic current state persists until the S1 protein is

completely pulled into the *trans* compartment resulting in a return to the open channel current v.

Example 4: Detection of translocation of proteins S2-35 and S2-148

This model makes a testable prediction. If the observed current states are due to processive movement of polypeptide segments into the pore lumen driven in part by ClpX, then changing the protein primary structure should result in sequential changes in the ionic current pattern that are ClpX/ATP-dependent. In particular, addition of a second Smt3 domain should result in a second ramping state (**Fig. 2A**, iii) followed by a second state centered at 3pA (**Fig. 2A**, iv). As a test, we fused a flexible glycine/serine-rich 35 amino acid linker to the N terminus of the S1 protein and capped this with a second Smt3 domain (protein S2-35, **Fig 1C**, ii and Sequence List, S2-35). Thus, the single folded-component sequence of S1 (C-terminus>charged flexible tail>Smt3>N-terminus) is repeated twice in S2-35 (C-terminus>charged flexible tail>Smt3>flexible linker>Smt3>N-terminus).

When protein S2-35 was captured in the nanopore with ClpX/ATP present in the *trans*-compartment, an ionic current pattern with eight reproducible states was observed (**Fig. 2C**).

Figure 2C shows the ionic current traces during protein S2-35 translocation. Open channel current (state i) is not shown. States ii–iv are identical to states ii-iv in panel a. (v) Gradual increase in ionic current to about 10 pA. In our working model this corresponds to a transition from Smt3 domain translocation to linker region translocation. vi) A second putative ramping state that closes resembles ramping state iii. vii) A second putative Smt3 translocation state with ionic current properties that closely resemble state iv. viii) Return to open channel current. The first four states (**Fig. 2C**, i-iv) were identical to states i-iv caused by S1 translocation (compare **Fig. 2A** and **2C**). This similarity included ramping state iii that is diagnostic for ClpX engagement, and the Smt3-dependent state iv. However, beginning at state v, the S2-35 pattern diverged from the S1 pattern (compare **Fig. 2A and 2C**). That is, following Smt3 state iv, a typical S2-35 ionic current trace did not proceed to the open channel current, but instead transitioned to a ~6 pA state with a median duration of 1.5 seconds (**Fig. 2C**, v and **Fig. 3B**). This was followed by a ~8.5 pA state (**Fig. 2C**, vi) that closely resembled ramping state iii, and a subsequent ionic current state that closely

resembled the putative Smt3 translocation state iv (**Fig. 2C**, vii). In other words, consistent with our model, the putative ClpX-bound and Smt3-dependent states that were observed once during S1 events (**Fig. 2A**) were observed twice during S2-35 events (**Fig. 2C**). These analogous states for the two constructs shared nearly identical amplitudes, RMS noise values, and durations (**Fig. 3A**, **Fig. 5**, and **Fig. 6**).

This dependence of ionic current on protein structure is consistent with ClpX-driven protein translocation through the nanopore. As an additional test, we re-examined ionic current state v observed during S2-35 translocation. This state is consistent with movement of the 35 amino acid linker through the nanopore based on two observations: 1) its average ionic current is measurably higher than surrounding states (**Fig. 2C**) as expected for an amino acid sequence with few bulky side chains; and 2) in the time domain, ionic current state v occurs between Smt3-dependent ionic current states iv and vi as expected given its position along the S2-35 primary sequence (**Fig 1C**, ii and related sequences).

If state v corresponds to translocation of the polypeptide linker under ClpX control, then changes in the length and composition of this linker should result in duration and current amplitude changes. For this test, we designed a third protein in which the S2-35 linker region was appended with an additional 113 amino acids, yielding a final construct consisting of two Smt3 domains separated by an extended 148 amino acid flexible linker (protein S2-148, **Fig 1C**, iii, and sequences S2-148).

Figure 2D shows the ionic current traces during protein S2-148 translocation. Ionic current states i-v and vi-viii were nearly identical to those states for S2-35 translocation (panel c). (v) In our working model, this ionic current state corresponds to translocation of the 148 amino acid linker. Its amplitude is ~3 pA higher than the S2-35 linker amplitude (~9 pA), and it has a median duration ~2.5 fold longer than the comparable S2-35 state v. Translocation events that included ramping state iii were observed 62 times for protein S2-35 (7.3 hours of experimentation), and 66 times for protein S2-148 (4.3 hours of experimentation), when ClpX and ATP were present. In the absence of ClpX/ATP, these ramping states were never observed for S2-35 (1.7 hours of experimentation), nor for S2-148 (1.2 hours of experimentation).

Figure 3A shows state iv (putative Smt3 translocation state). These events included only those that manifest the ClpX-dependent ramping state iii. S1 n=45, S2-35 n=60, S2-148

n=65. **Figure 3B** shows Comparison of linker region state v dwell times for the S2-35 and S2-148 proteins. Events included in these histograms manifest ramping state iii. S2-35 n=50, S2-148 n=50. **Figure 3C** shows state v translocation dwell times for S2-148 translocation events. The black bars represent dwell times for events that included ramping state iii (ClpX-driven). The gray bars represent events that did not include the ramping state (not ClpX-driven). With ramping n=50, without ramping n=20.

As predicted, when this protein was captured in the nanopore under standard conditions in the presence of ClpX/ATP, eight reproducible states similar to S2-35 events were observed (**Fig. 2d**, **Fig. 3A**, **Fig. 5**, **6**, and **7**).

Importantly, however, the S2-35 and S2-148 events differed significantly at state v (compare **Fig. 2c** and **2d**). That is, the S2-148 state v had a higher mean residual current than did S2-35 (~9 vs ~6 pA, respectively), and a median duration ~2.5 fold longer than that of S2-35 state v (**Fig. 3B**). The increased duration for S2-148 state v relative to S2-35 is expected as it should take ClpX longer to process the additional amino acids, while the increased current level is likely due to differences in linker amino acid composition between the two proteins (S2-35 linker: 51% Gly, 34% Ser, 15% other; S2-148 linker: 34% Gly, 32% Ser, 19% Ala, 15% other). However, we cannot exclude relative proximity of the proteins' Smt3 domains to the nanopore orifice that necessarily must differ due to the linker lengths.

We note that voltage-driven translocation of all three model proteins was observed absent ClpX/ATP (**Fig. 4**). However, these ClpX-minus translocation events lacked the diagnostic ramping states shown in **Figure 2**, and they were significantly longer and more variable in duration than were ClpX-mediated translocation events (**Fig. 3C**, **Fig. 8**, and **Fig. 9**). This is consistent with an unregulated translocation process dependent upon random structural fluctuations of the captured protein molecule and intermittent electrical force acting on a polymer with variable charge density. Thiscontrasts with the relatively constant ATP hydrolysis rate and mechanical force imparted by the ClpX motor.

SEQUENCE LIST:

SEQ ID NO 1:

GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSD
GDGDSDGDDAANDENYALAA

SEQ ID NO 2:

GGSGSGGSGSGGSGSQNEYRSGGSGSGGSGSGGSG

SEQ ID NO 3:

GGSGSAGSGASGSSGSEGSGASGSAGSGSAGSRGS
GASGSAGSGSAGSGGAEAAKEAAKEAAKEAAKEAAKAGGSGSAGSAGSASSGSDG
SGASGSAGSGSAGSKGSGASGSAGSGSSGS

SEQ ID NO 4 (S1): MGSSHHHHHHGSG ← affinity purification tag region

LVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG ← Smt3
domain

GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSD
GDGDSDGDD ← charged tail

AANDENYALAA ← ssrA tag

SEQ ID NO 5 (S2-35):

MGHHHHHHGS ← affinity purification tag region

LQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGK
EMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGGGGSGSGGSGSGGSGSQNE
YRSGGSGSGGSGSGGSG ← Smt3 domain

MGSSHHHHHHGSG ← affinity purification tag region

LVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG ← Smt3
domain

GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSD
GDGDSDGDD ← charged tail

AANDENYALAA ← ssrA tag

SEQ ID NO 6 (S2-148):

MGHHHHHHGS ← affinity purification tag region

LQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGK
EMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG ← Smt3 domain

GGSGSGGSGSGGSGSQNEYRSGGGGSGSAGSGASGSSGSEGSGASGSAGSGSAGSRG
SGASGSAGSGSAGSGGAEAAKEAAKEAAKEAAKEAAKAGGSGSAGSAGSASSGSD
GSGASGSAGSGSAGSKGSGASGSAGSGSSGSSGGSG ← linker region

MGSSHHHHHHGSG ← affinity purification tag region

LVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG ← Smt3
domain

GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSD
GDGDSDGDD ← charged tail

AANDENYALAA ← ssrA tag

SEQ ID NO: 7 (S1-RQA):
MGSSHHHHHHGSG ← affinity purification tag region

LVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME
AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG ← Smt3
domain

GGSSGGSGGSGSSGDGGSSGGSGGSGSSGDGGSSGGSGGDGSSGDGGSDGDSDGSD
GDGDSDGDD ← charged tail

AANDENYALAA ← ssrA tag

RQA ← additional residues added to obscure the ssrA tag

CONCLUSION

The above specific description is meant to exemplify and illustrate the invention and should not be seen as limiting the scope of the invention, which is defined by the literal and equivalent scope of the appended claims. Any patents or publications mentioned in this specification are intended to convey details of methods and materials useful in carrying out certain aspects of the invention which may not be explicitly set out but which would be understood by workers in the field.

What is claimed is:

1. A device for translocating a protein through a nanopore, comprising:
   (a) a membrane having a nanopore therein, said membrane separating a chamber into a *cis* side and a *trans* side, wherein the protein is to be added to the *cis* side and translocated through the nanopore to the *trans* side; and
   (b) a protein translocase, on at least one side of said chamber, which binds to and translocates the protein through the translocase and through the nanopore in a sequential manner.

2. The device of claim 1 further comprising: a circuit for providing a voltage between the *cis* side and the *trans* side of the membrane and for measuring a signal based on ionic current flowing through the nanopore during a period of time of translocation, wherein the circuit detects changes in the signal that reflect characteristics of the protein as it is translocated.

3. The device of claim 1 or 2 wherein the nanopore is defined by a pore protein.

4. The device of claim 3 wherein the pore protein is α-hemolysin.

5. The device of claim 3 wherein the pore protein does not specifically bind the protein.

6. The device of any one of claims 1 to 5, wherein the protein translocase is an NTP-driven unfoldase.

7. The device of claim 6 wherein the NTP-driven unfoldase is an AAA+ enzyme.

8. The device of claim 7 wherein the AAA+ enzyme is ClpX.

9. The device of claim 2 wherein the circuit comprises an amplifier applying a

positive voltage to the *trans* side.

10. The device of claim 9 comprising a computer, attached to a patch clamp amplifier, for recording changes in ionic current through the nanopore.

11. The device of claim 10 wherein said computer is programmed with instructions comparing the changes in ionic current to structural features of the protein to be translocated.

12. The device of claim 2 wherein the circuit comprises a sensor for sensing changes in ionic current that occur at 100 mHz.

13. A system for translocating a protein through a nanopore, comprising:
(a) a nanopore in a membrane separating a fluidic chamber into a *cis* side and a *trans* side, wherein a protein to be translocated is added to the *cis* side and is translocated through the nanopore to the *trans* side;
(b) said fluidic chamber containing on the *cis* side (i) an ionic buffer comprising NTPs and (ii) a non-denatured protein to be translocated;
(c) a circuit for providing a voltage between the *cis* side and the *trans* side and measuring ionic current flowing through the nanopore;
(d) a protein translocase in the chamber arranged so that it can translocate the protein through the translocase and through the nanopore in a sequential manner.

14. The system of claim 13 wherein the nanopore is defined by a multimeric pore protein and the protein translocase is an NTP-driven unfoldase.

15. The system of claim 13 wherein the protein to be translocated comprises an exogenous sequence comprising a targeting domain for the protein translocase.

16. The system of claim 15 wherein the targeting domain comprises at least 5

negatively charged amino acids.

17. The system of claim 13 wherein the protein translocase is ClpX and the nanopore is defined by α-hemolysin.

18. A method for translocating a protein through a nanopore, comprising the steps of:

(a) providing a device comprising a nanopore in a membrane separating a fluidic chamber into a *cis* side and a *trans* side;

(b) adding to said fluidic chamber a buffer containing NTPs and an NTP-driven translocase;

(c) adding a protein to be translocated to the *cis* side; and

(d) allowing the protein to be translocated by the NTP-driven translocase through the NTP-driven translocase and through the nanopore in a sequential manner.

19. A method for translocating a protein through a nanopore, comprising the steps of:

(a) providing a device for translocating a protein through a nanopore, comprising:

(i) a nanopore in a membrane separating a fluidic chamber into a *cis* side and a *trans* side, wherein a protein to be translocated is added to the *cis* side and is translocated through the nanopore to the *trans* side; and

(ii) a circuit for providing a voltage between the *cis* side and the *trans* side and for measuring ionic current flowing through the nanopore;

(b) adding to said fluidic chamber a buffer containing NTPs and an NTP-driven translocase;

(c) adding a protein to be translocated to the *cis* side;

(d) allowing the protein to contact the NTP-driven translocase;

(e) allowing the protein to be translocated by the NTP-driven translocase through the NTP-driven translocase and through the nanopore; and

(f) measuring ionic current changes caused by translocation of the protein through the nanopore.

20. The method of claim 19 wherein the step of measuring current changes comprises measuring current changes for states of (i) open channel, (ii) capture of the protein by the nanopore, and (iii) passage of a protein from (ii) through the nanopore.

21. The method of claim 20 wherein the measuring comprises detecting differences between states (i), (ii) and (iii).

22. The method of claim 20 wherein the measuring comprises measuring differences during state (iii) caused by amino acid structure of the protein passing through the nanopore.

23. The method of claim 21 or 22 further comprising the step of measuring a state of binding of the NTP-driven translocase to the protein and translocation of the protein toward the nanopore, which occurs as a state between states (ii) and (iii).

24. The method of claim 19 wherein the nanopore is defined by a pore protein.

25. The method of claim 24 wherein the pore protein is α-hemolysin.

26. The method of claim 19 or 20 wherein the NTP-driven translocase is attached to the nanopore.

27. The method of claim 19 wherein the NTP-driven translocase is an AAA+ enzyme.

28. The method of claim 27 wherein the AAA+ enzyme is ClpX.

29. The method of claim 19 wherein the circuit comprises a patch clamp amplifier

applying a constant voltage between the *cis* chamber and the *trans* chamber.

30. The method of claim 19 or 20 wherein the protein is added in a non-denatured state.

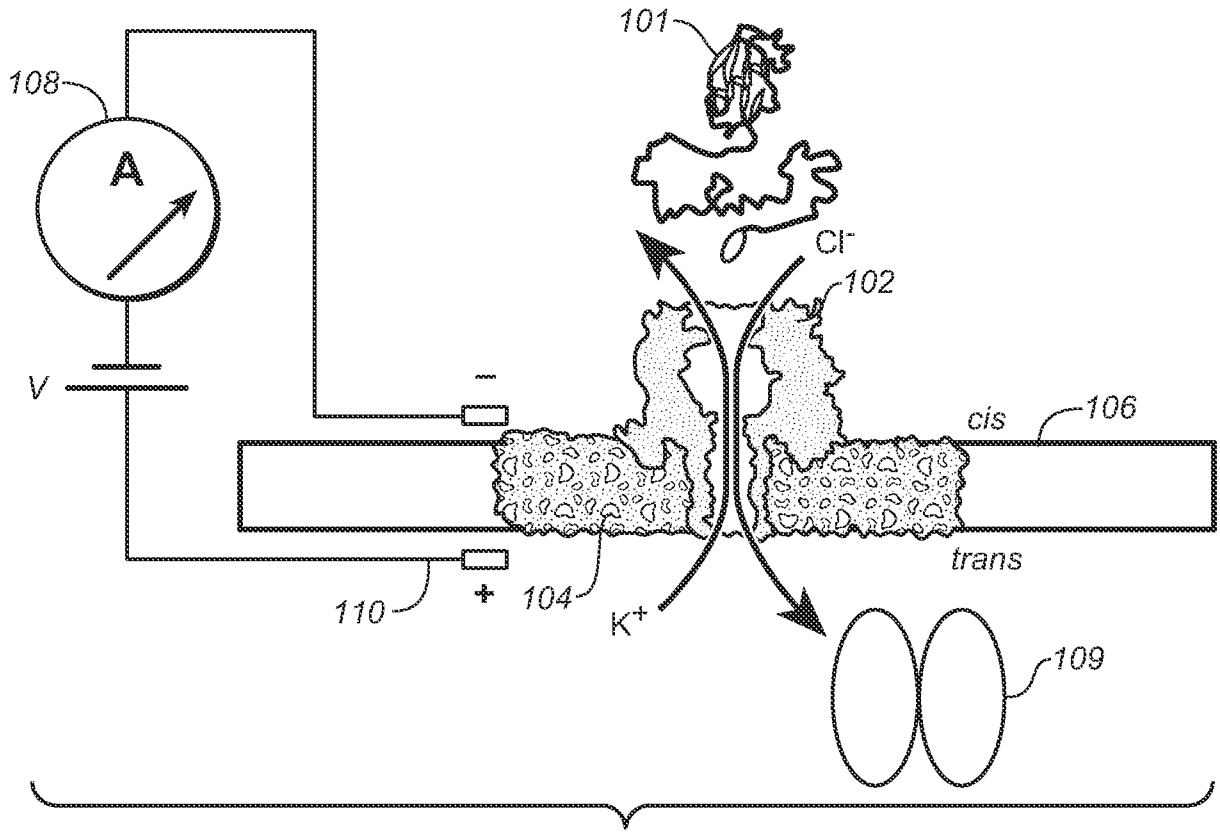


FIG. 1A

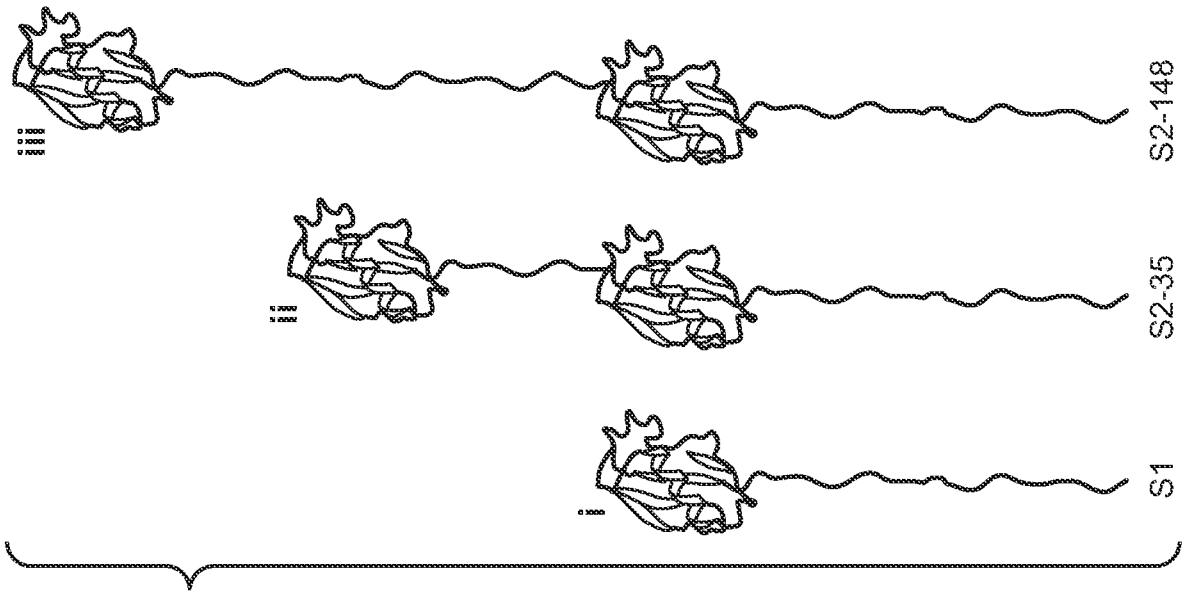


FIG. 1C

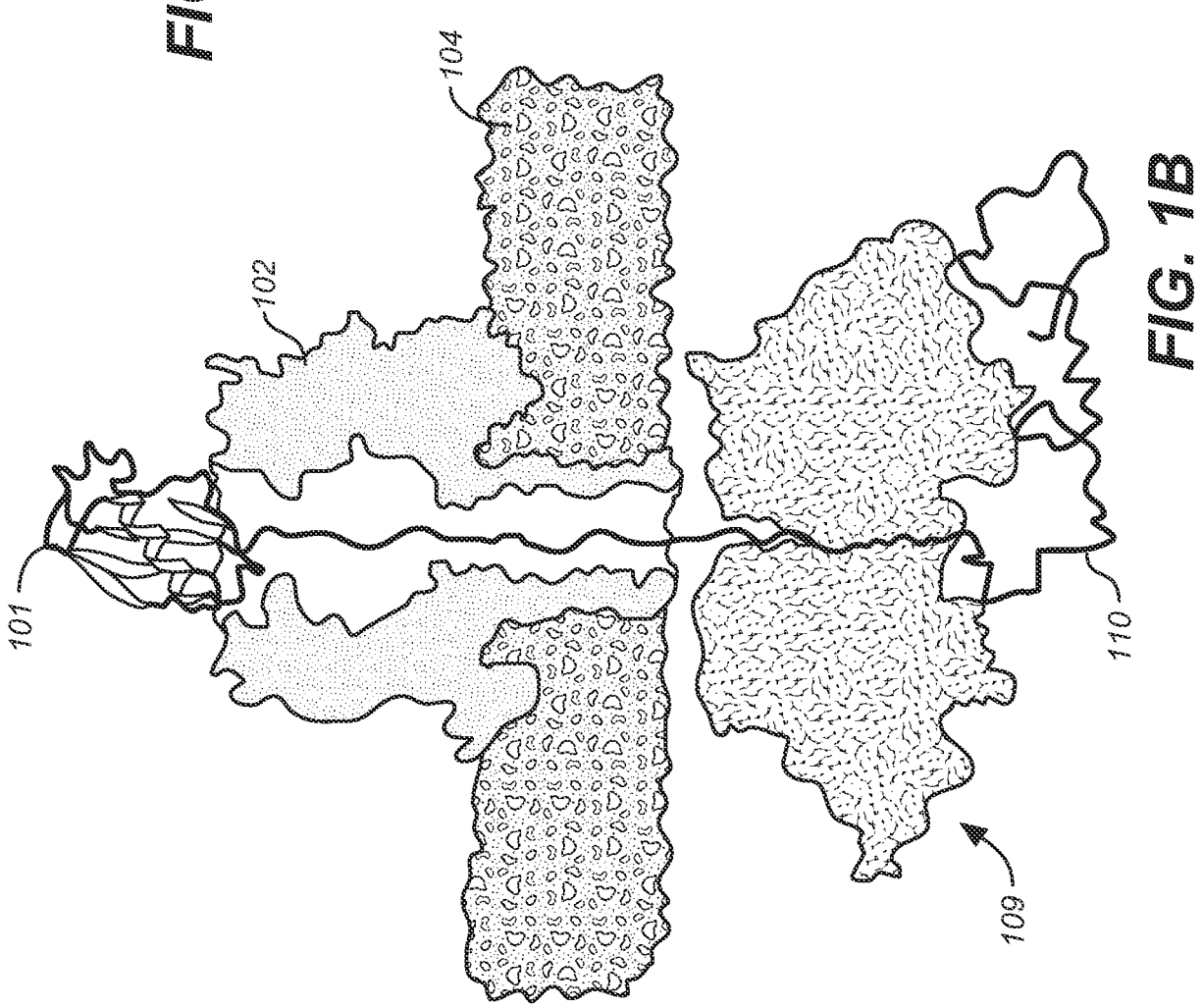


FIG. 1B

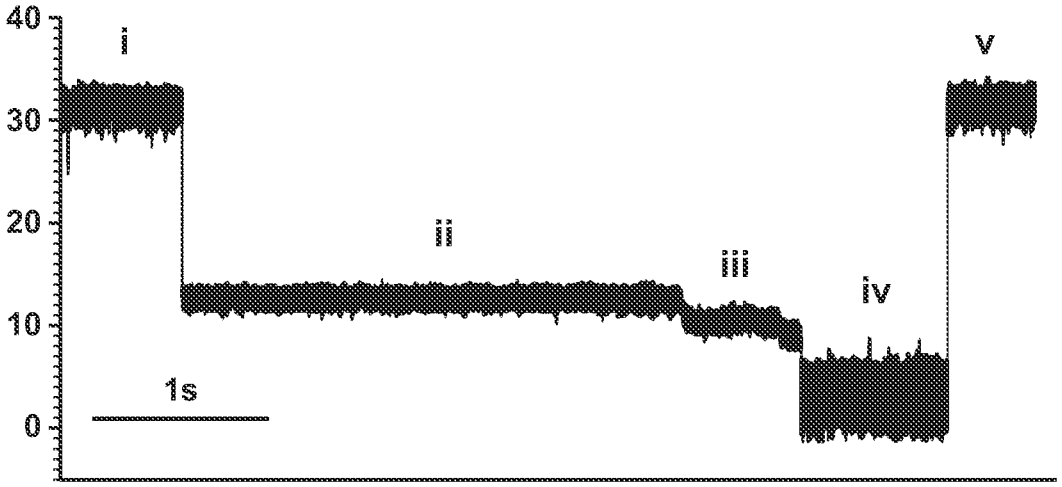


FIG. 2A

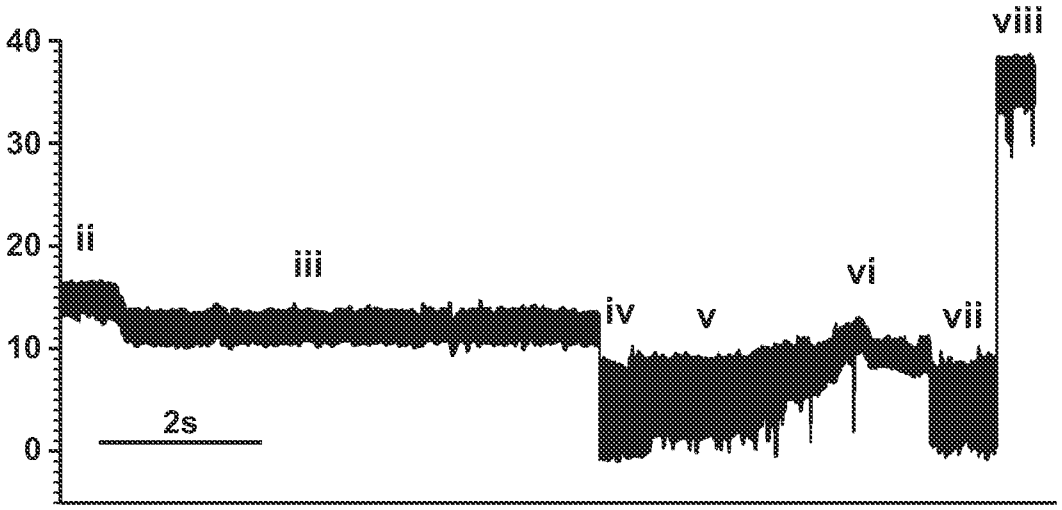


FIG. 2C

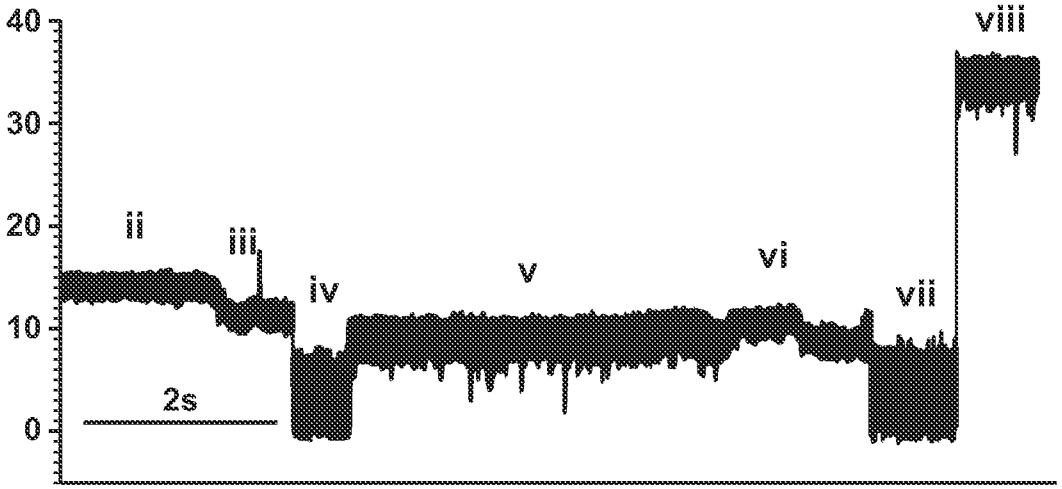


FIG. 2D

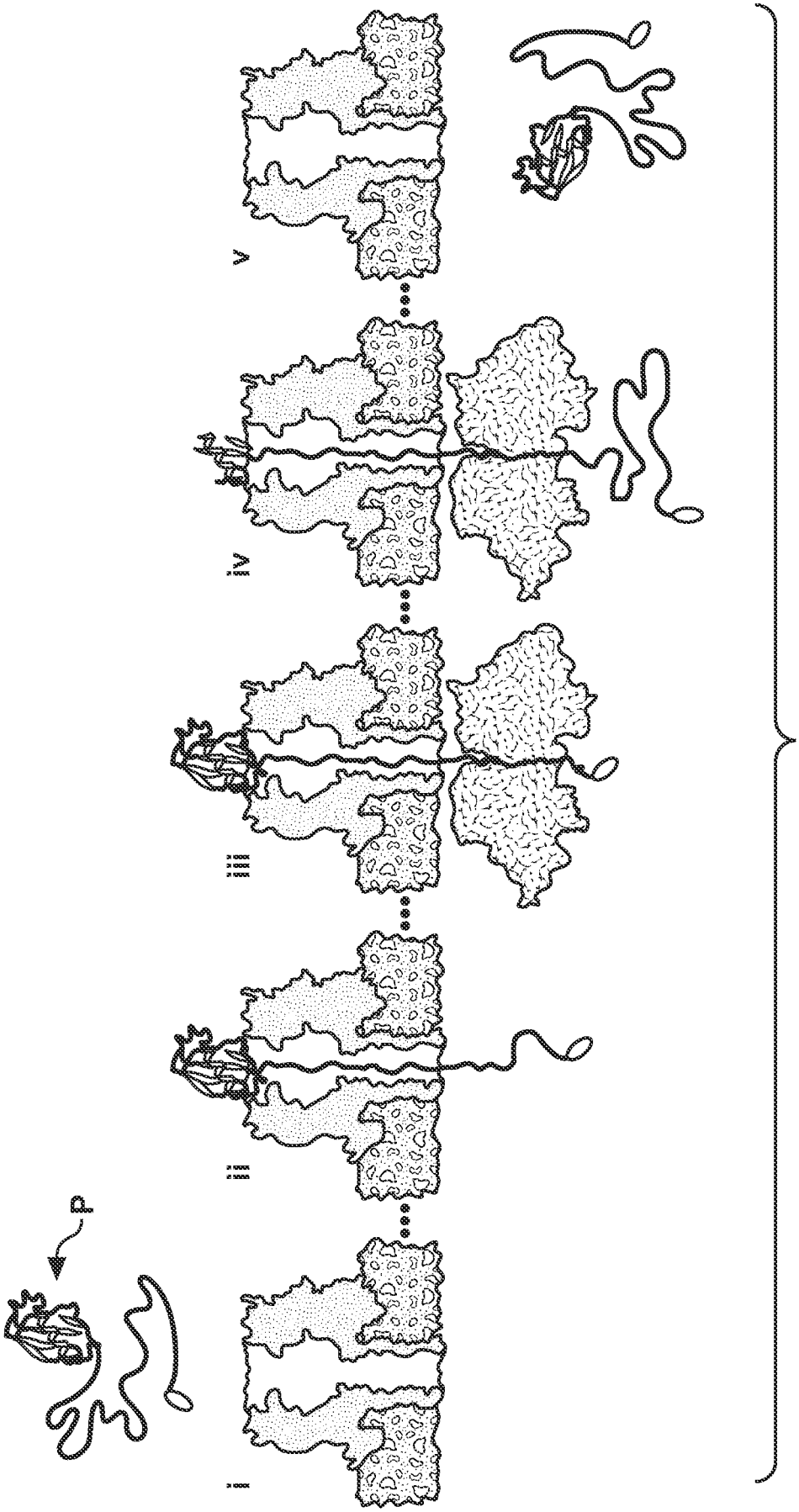


FIG. 2B

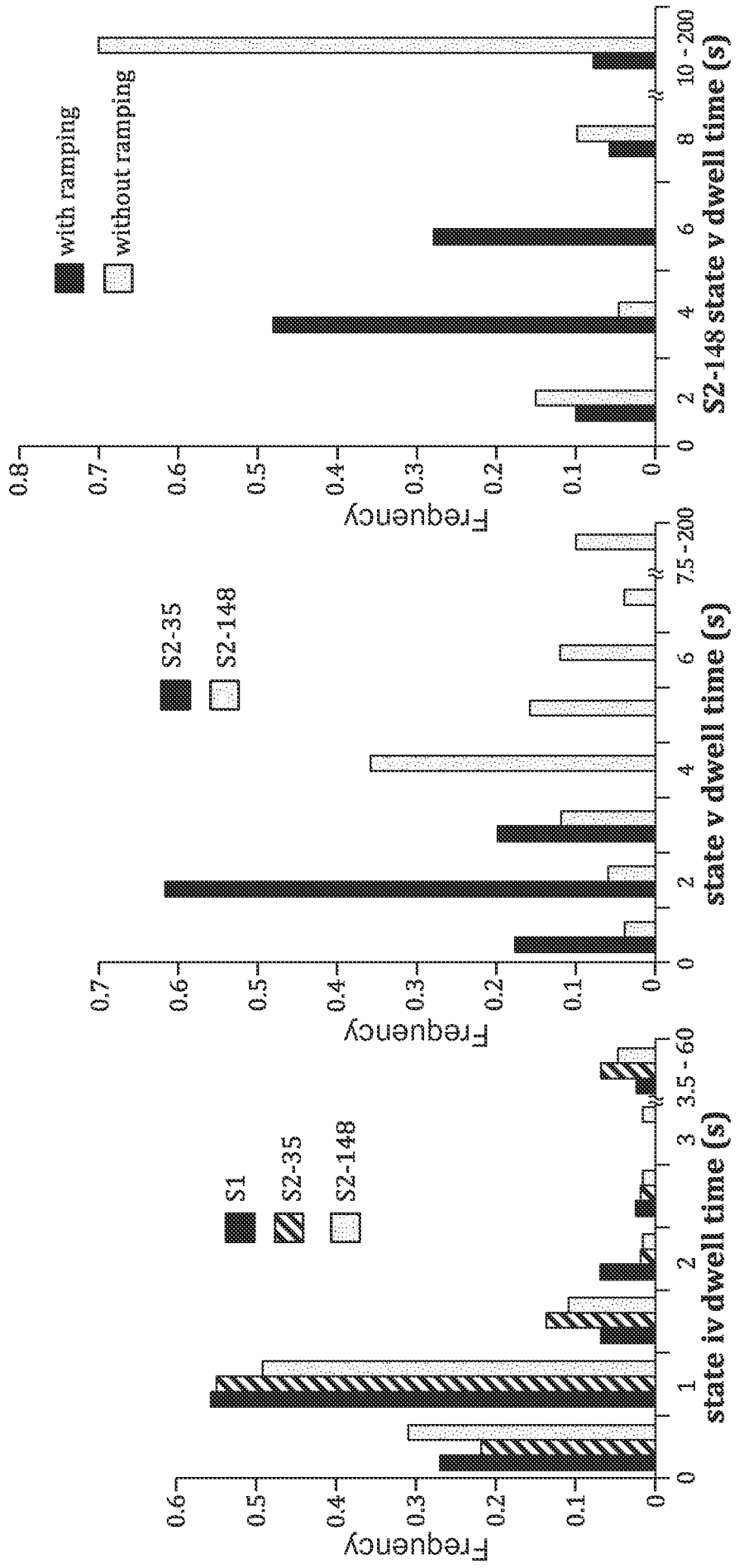


FIG. 3C

FIG. 3B

FIG. 3A

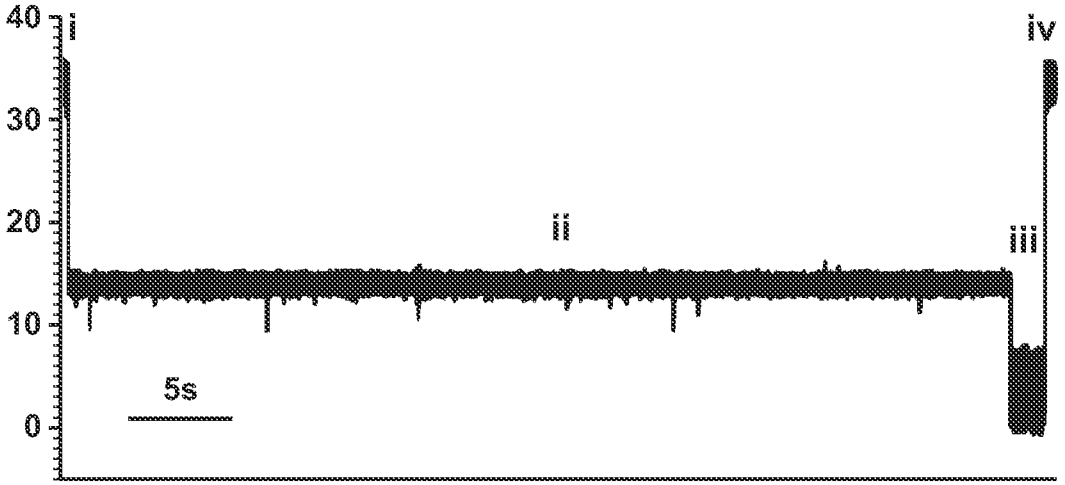


FIG. 4A

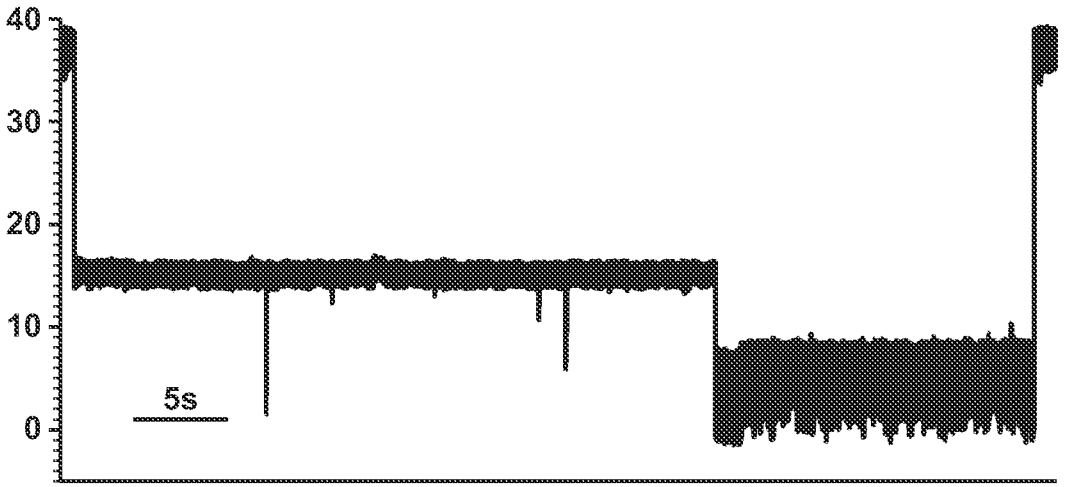


FIG. 4C

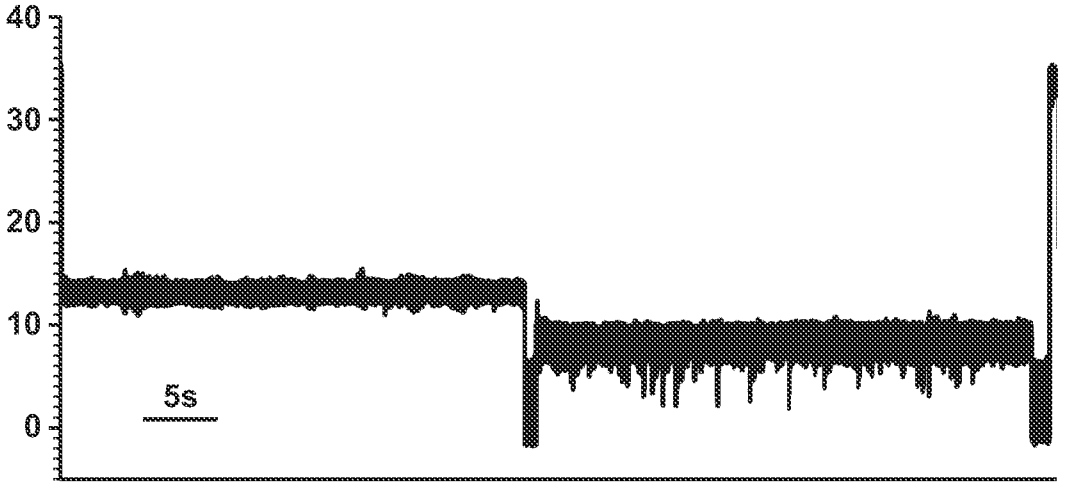


FIG. 4D

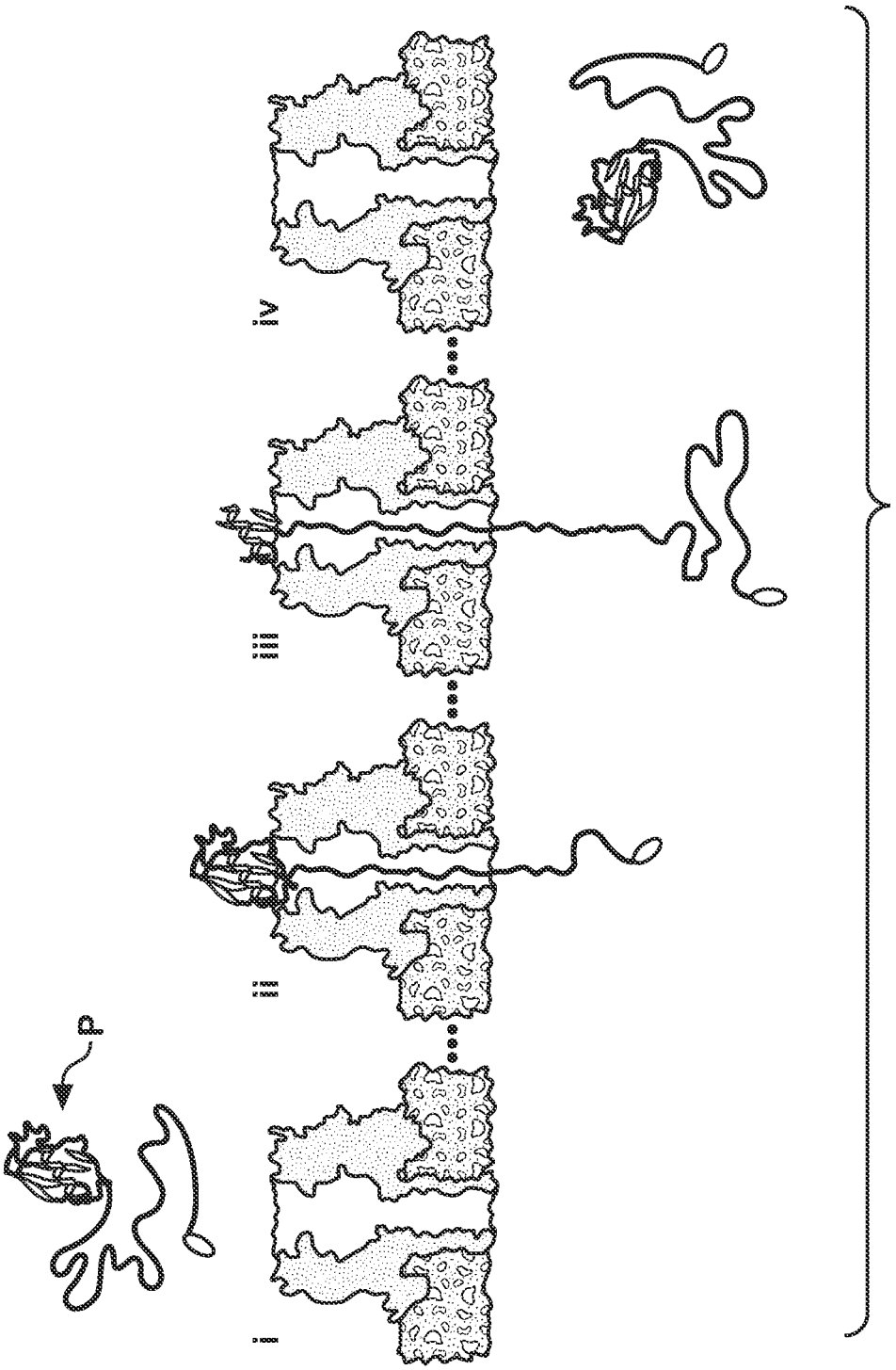


FIG. 4B

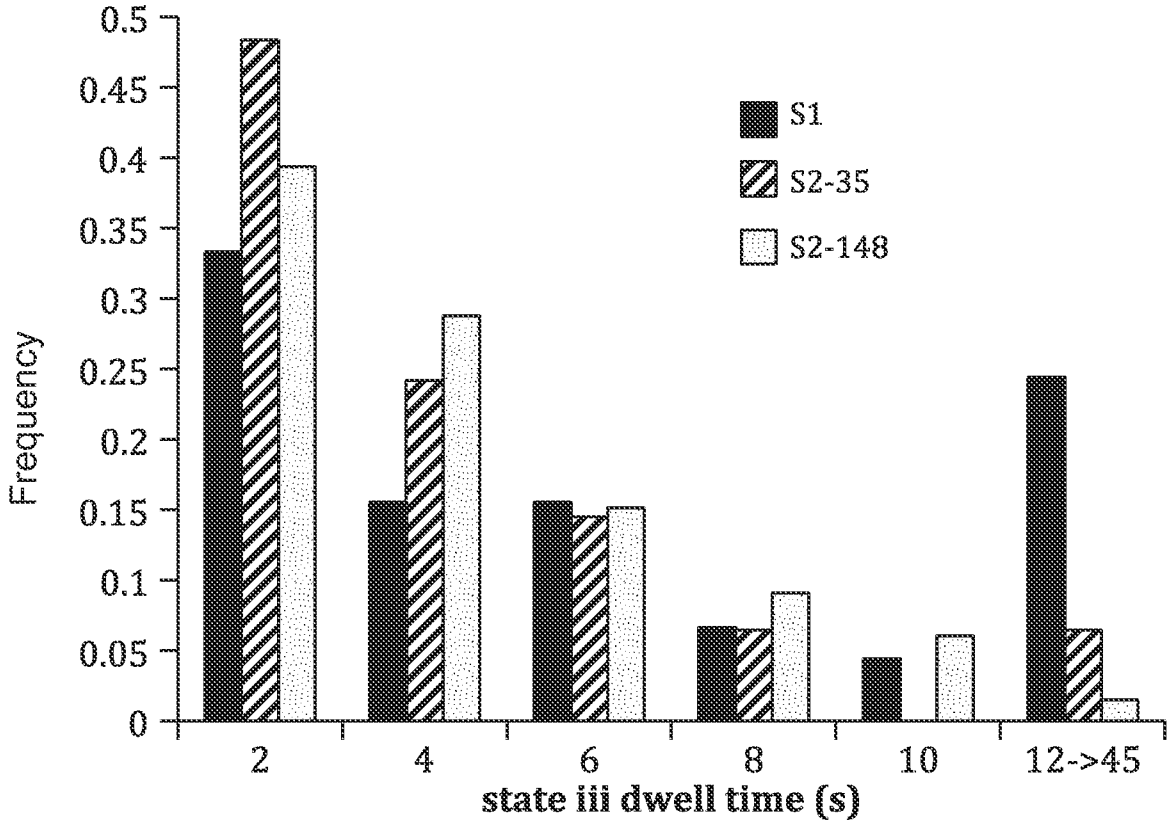

0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Frequency
S1
S2-35
S2-148
2
4
6
8
10
12->45
state iii dwell time (s)

FIG. 5

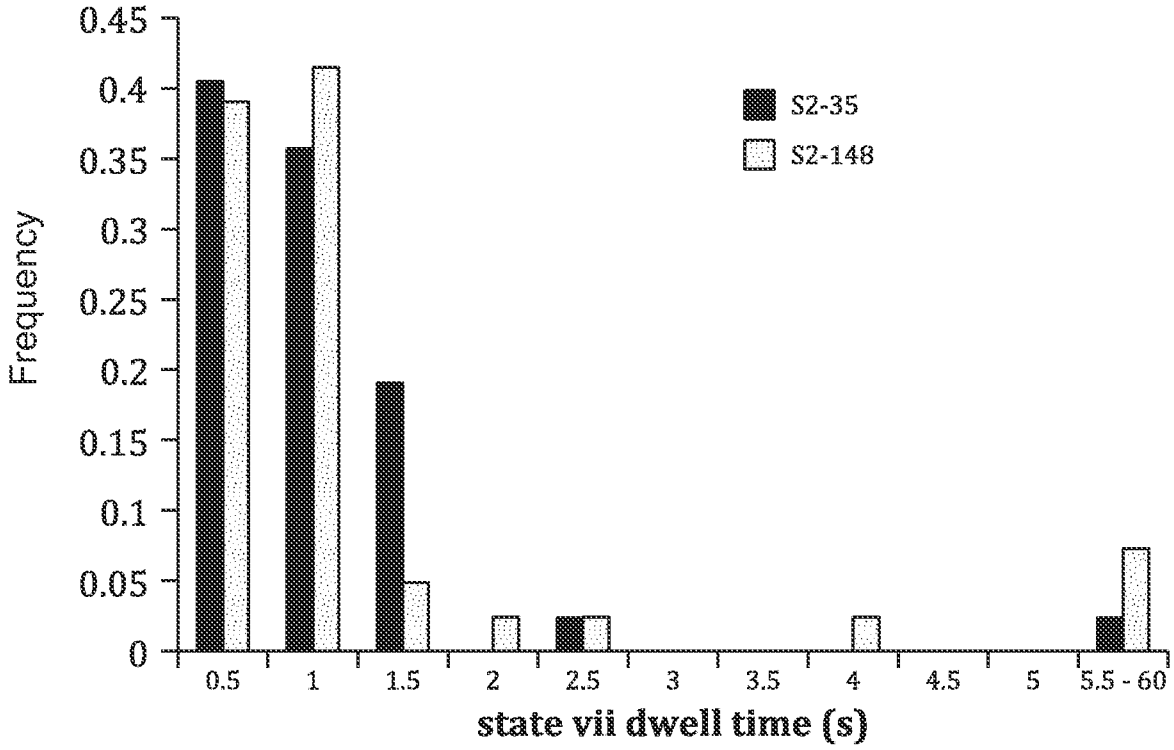

0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Frequency
S2-35
S2-148
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5 - 60
state vii dwell time (s)

FIG. 6

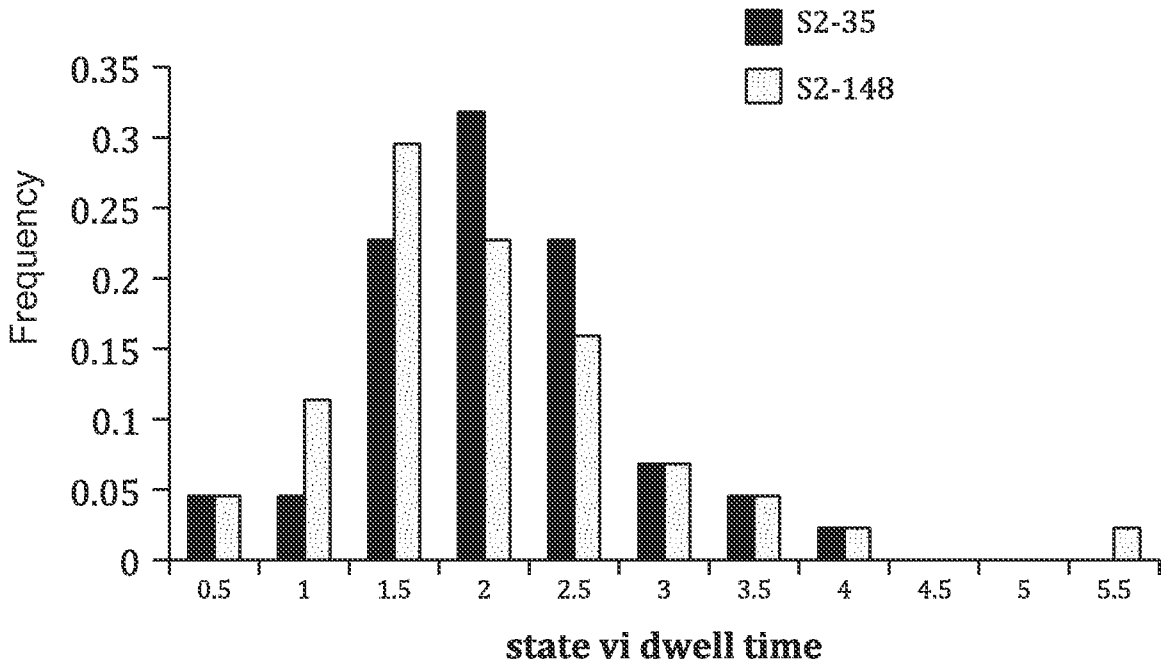

S2-35
S2-148
Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
state vi dwell time

FIG. 7

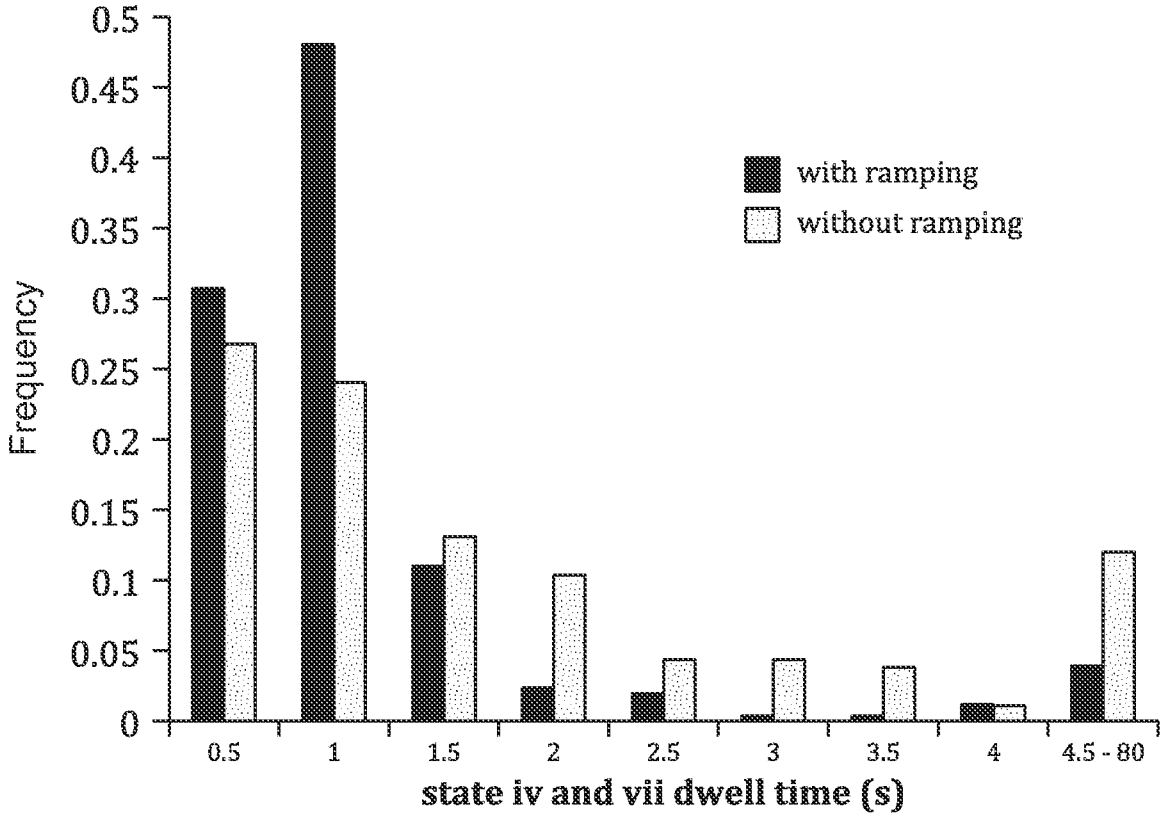

with ramping
without ramping
Frequency
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5 - 80
state iv and vii dwell time (s)

FIG. 8

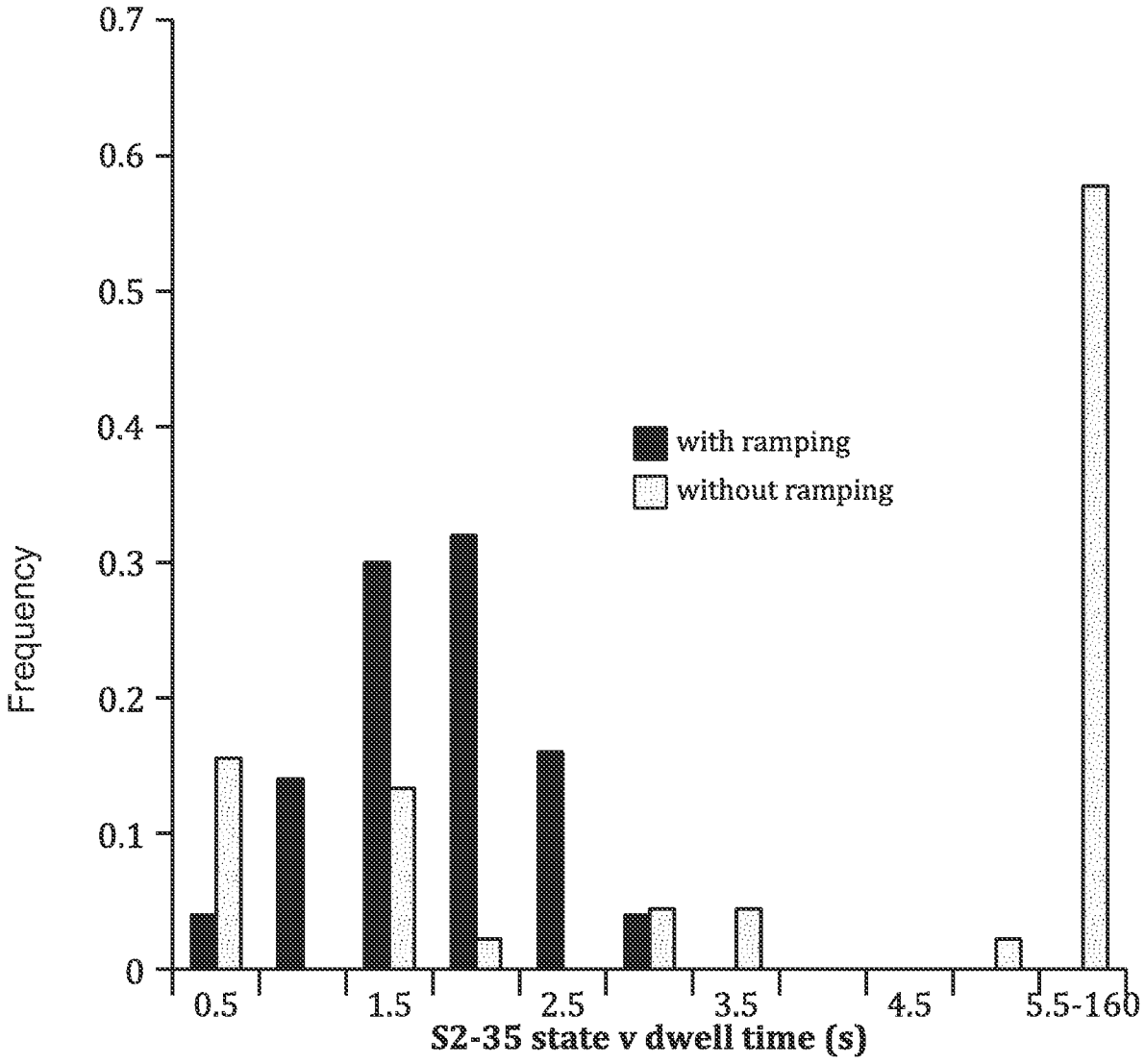

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Frequency
with ramping
without ramping
0.5
1.5
2.5
3.5
4.5
5.5-160
S2-35 state v dwell time (s)

FIG. 9

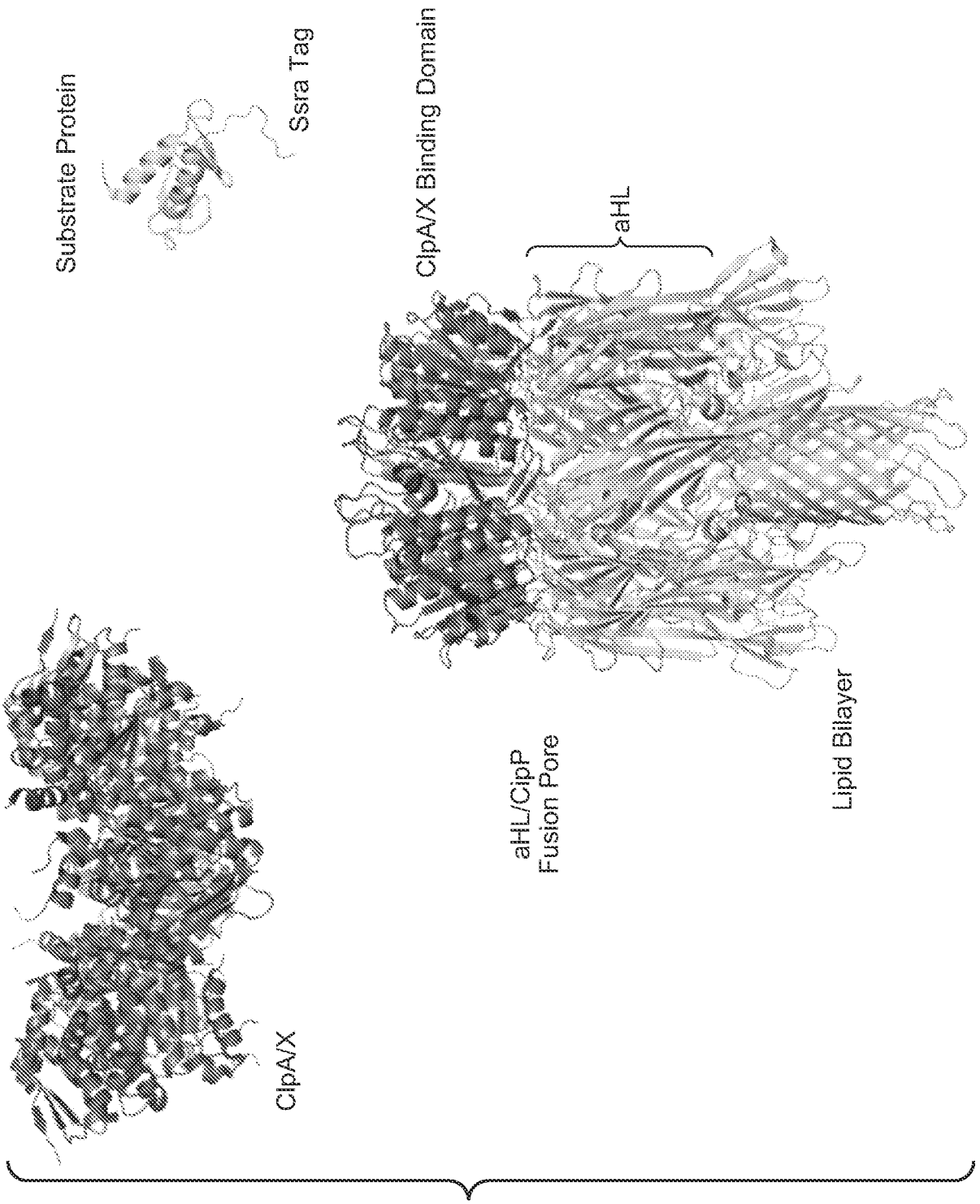
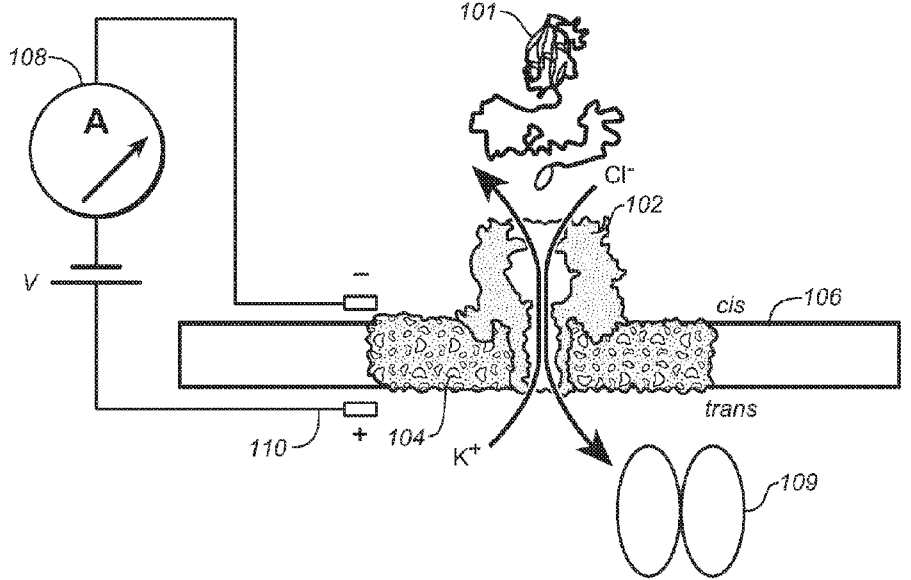

Substrate Protein
Ssra Tag
ClpA/X Binding Domain
aHL
ClpA/X
aHL/ClpP
Fusion Pore
Lipid Bilayer

FIG. 10


101
108
A
V
Cl⁻
102
cis
106
trans
110
+
104
K⁺
109