US005795782A

United States Patent [19]

Church et al.

[11] Patent Number: 5,795,782

[45] Date of Patent: Aug. 18, 1998

[54] CHARACTERIZATION OF INDIVIDUAL POLYMER MOLECULES BASED ON MONOMER-INTERFACE INTERACTIONS

[75] Inventors: **George Church**, Brookline, Mass.; **David W. Deamer**, Santa Cruz, Calif.; **Daniel Branton**, Lexington; **Richard Baldarelli**, Natick, both of Mass.; **John Kasianowicz**, Darnestown, Md.

[73] Assignees: **President & Fellows of Harvard College**, Cambridge, Mass.; **The Regents of the University of California**, Oakland, Calif.

[21] Appl. No.: **405,735**

[22] Filed: **Mar. 17, 1995**

[51] **Int. Cl.**[6] **G01N 33/483**

[52] **U.S. Cl.** **436/2**; 436/151; 435/4

[58] **Field of Search** 435/6, 4, 5; 436/2, 436/151

[56] **References Cited**

PUBLICATIONS

Auld et al., "A Neutral Amino Acid Change in Segment IIS4 Dramatically Alters the Gating Properties of the Voltage-Dependent Sodium Channel", 1990, *Proc. Natl. Acad. Sci. USA*, 87:323–27.

Bensimon, A., et al., "Alignment and Sensitive Detection of DNA by a Moving Interface", 1994, *Science*, 265:2096–98.

Benz et al., "Pore Formation by LamB of *Escherichia Coli* in Lipid Bilayer Membranes", 1986, *J. Bacteriology*, 165(3):978–86.

Benz et al., "Mechanism of Sugar Transplant through the Sugar-Specific LamB Channel of *Escherichia Coli* Outer Membrane", 1987, *J. Membrane Biol.*, 100:21–29.

Bezrukov et al., "Counting Polymers Moving Through a Single Ion Channel", 1994, *Nature*, 370:279–81.

Boulain et al., "Mutagenesis by Random Linker Insertion into the LamB Gene of *Escherichia Coli* K12", 1986, *Mol. Gen. Genet.*, 205:339–48.

Boulanger et al., "Characterization of Ion Channels Involved in the Penetration of Phage T4 DNA into *Escherichia Coli* Cells", 1988, *J. Biolog. Chem.*, 263(20):9767–75.

Boulanger et al., "Ion Channels Are Likely to Be Involved in the Two Steps of Phage T5 DNA Penetration into *Escherichia Coli* Cells", 1992, *J. Biolog. Chem.*, 267(5):3168–72.

Boyd et al., "Determinants of Membrane Protein Topology", 1987, *Proc. Natl. Acad. Sci. USA*, 84:8525–29.

Charbit et al., "Permissive Sites and Topology of an Outer Membrane Protein with a Reporter Epitope", 1991, *J. Bacteriology*, 173(1):262–75.

Dargent et al., "Selectivity for Maltose and Maltodextrins of Maltoporin, a Pore-Forming Protein of *E. Coli* Outer Membrane", 1987, *FEB Letters*, 220(1):136–42.

Dargent et al., "Effect of Point Mutations on the in-Vitro Pore Properties of Maltoporin, a Protein of *Escherichia Coli* Outer Membrane", 1988, *J. Mol. Biol.*, 201:497–506.

DeBlois et al., "Electrokinetic Measurements with Submicron Particles and Pores by the Resistive Pulse Technique", 1977, *J. Colloid and Interface Science*, 61(2):323–35.

Ehrmann et al., "Genetic Analysis of Membrane Protein Topology by a Sandwich Gene Fusion Approach", 1990, *Proc. Natl. Acad. Sci. USA*, 87:7574–78.

Ferenci et al., "Channel Architecture in Maltoporin: Dominance Studies with LamB Mutations Influencing Maltodextrin Binding Provide Evidence for Independent Selectivity Filters in Each Subunit", 1989, *J. Bacteriology*, 171(2):855–61.

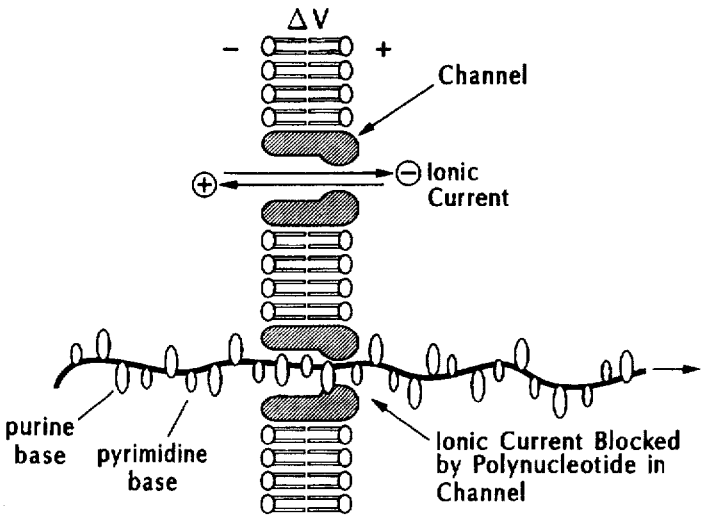
(List continued on next page.)

Primary Examiner—Charles L. Patterson, Jr.
Attorney, Agent, or Firm—Fish & Richardson P.C.

[57] **ABSTRACT**

A method is disclosed for characterizing a linear polymer molecule by measuring physical changes across an interface between two pools of media as the linear polymer traverses the interface and monomers of the polymer interact with the interface, where the physical changes are suitable to identify characteristics of the polymer.

15 Claims, 6 Drawing Sheets



PUBLICATIONS

Ghadiri et al., "Artificial Transmembrane Ion Channels From Self-Assembling Peptide Nanotubes", 1994, *Nature*, 369:301-304.

Hall et al., "Alamethicin: A Rich Model for Channel Behavior", 1984, *J. Biophys.*, 45:233-47.

Hamill et al., "Improved Patch-Clamp Techniques for High-Resolution Current Recording from Cells and Cell-Free Membrane Patches", 1981, *Pflügers Archiv. Eur. J. Physiology*, 391(2):85-100.

Harrington et al., "The F Pilus of *Escherichia Coli* Appears to Support Stable DNA Transfer in the Absence of Wall-to-Wall Contact Between Cells", 1990, *J. Bacteriology*, 172(12):7263-64.

Heinemann et al., "Open Channel Noise IV: Estimation of Rapid Kinetics of Formamide Block in Gramicidin A Channels", 1988, *J. Biophys.*, 54:757-64.

Heinemann et al., "Open Channel Noise V: Fluctuating Barriers to Ion Entry in Gramicidin A Channels", 1990, *J. Biophys.*, 57:499-514.

Henry et al., "Blockade of a Mitochondrial Cationic Channel by an Addressing Peptide: An Electrophysiological Study", 1989, *J. Membranee Biol.*, 112:139-47.

Hoshi et al., "Biophysical and Molecular Mechanisms of Shaker Potassium Channel Inactivation", 1990, *Science*, 250:533-38.

Hoshi et al., "Two Types of Inactivation in Shaker K^+ Channels: Effects of Alterations in the Carboxy-Terminal Region", 1991, *Neuron*, 7:547-56.

Kubitschek, "Electronic Counting and Sizing of Bacteria", *Nature*, 1958, 182:234-35.

Lakey et al., "The Voltage-Dependent Activity of *Escherichia Coli* Porins in Different Planar Bilayer Reconstitutions", 1989, *Eur. J. Biochem.*, 186:303-308.

Lopez et al., "Hydrophobic Substitution Mutations in the S4 Sequence Alter Voltage-Dependent Gating in Shaker K^+ Channels", 1991, *Neuron*, 7:327-36.

Moellerfeld et al., "Improved Stability of Black Lipid Membranes by Coating with Polysaccharide Derivatives Bearing Hydrophibic Anchor Groups", 1986, *Biochimica et Biophysica Acta*, 857:265-70.

Nath et al., "Transcription by T7 RNA Polymerase Using benzo[a]pyrene-modified templates", 1991, *Carcinogenesis*, 12(6):973-76.

Neher et al., "Single-Channel Currents Recorded from Membrane of Denervated Frog Muscle Fibres", 1976, *Nature*, 260:799-801.

Novick et al., "Fluorescence Measurement of the Kinetics of DNA Injection by Bacteriophage 1 into Liposomes", 1988, *Biochemistry*, 27:7919-24.

Ollis et al., "Domain of *E. Coli* DNA Polymerase I Showing Sequence Homology to T7 DNA Polymerase", 1985, *Nature*, 313:818-19.

Ollis et al., "Structure of Large Fragment of *Escherichia Coli* DNA Polymerase I Complexed with dTMP", 1985, *Nature*, 313:762-66.

Ovchinnikov et al., 3. The Cyclic Peptides: Structure, Conformation, and Function: P. Gramicidin S. (851), Its Analogs and Tyrocidines A-C (904-906), 1982, *The Proteins, Third Edition*, 5:547-55.

Ovchinnikov et al., 3. The Cyclic Peptides: Structure, Conformation, and Function: T. Valinomycin (913), 1982, *The Proteins, Third Edition*, 5:563-73.

Patton et al., "Amino Acid Residues Required for Fast Na^+-channel Inactivation: Charge Neutralizations and Deletions in the III-IV Linker", 1992, *Proc. Natl. Acad. Sci. USA*, 89:10905-909.

Shiver et al., "On the Explanation of the Acidic pH Requirement for In Vitro Activity of Colicin E1", 1987, *J. Biological Chem.*, 262(29):14273-281.

Sigworth et al., "Open Channel Noise: III. High Resolution Recordings Show Rapid Current Fluctations in Gramicidin A and Four Chemical Analogues", 1987, *J. Biophys.*, 52:1055-64.

Simon et al., "A Protein Conducting Channel in the Endoplasmic Reticulum", *Cell*, 65:371-80, (1991).

Taylor et al., "'Reversed' Alamethicin Conductance in Lipid Bilayers", 1991, *J. Biophys.*, 59:873-79.

Weiss et al., "Molecular Architecture and Electrostatic Properties of a Bacterial Porin", 1991, *Science*, 254:1627-30.

West et al., "A Cluster of Hydrophobic Amino Acid Residues Required for Fast Na^+-channel Inactivation", 1992, *Proc. Natl. Acad. Sci. USA*, 89:10910-14.

Wonderlin et al., "Optimizing Planar Lipid Bilayer Single-Channel Recordings for High Resolution with Rapid Voltage Steps", 1990, *J. Biophys.*, 58:289-97.

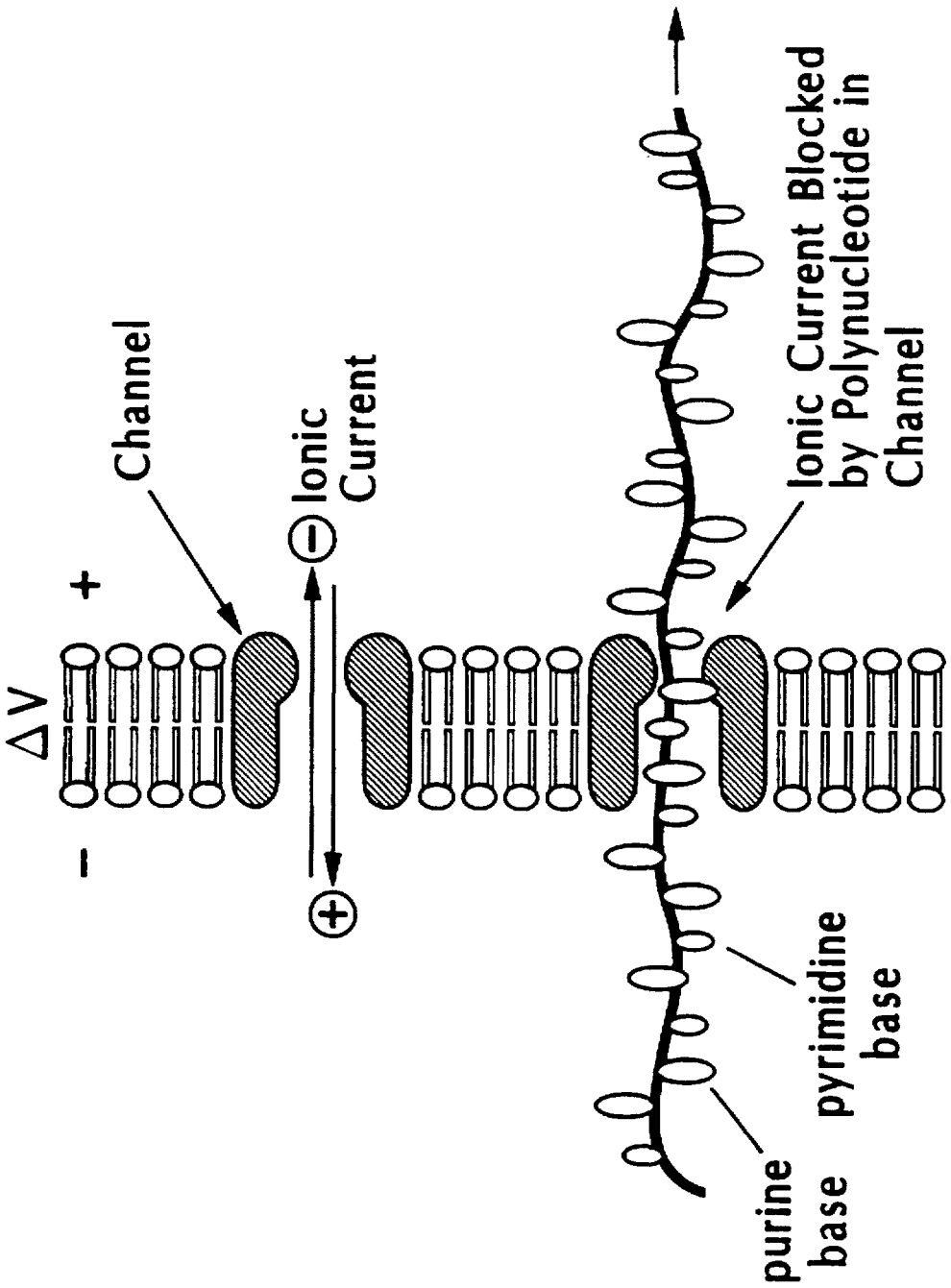

ΔV
+
-
Channel
Ionic Current
Ionic Current Blocked by Polynucleotide in Channel
pyrimidine base
purine base

FIG. 1

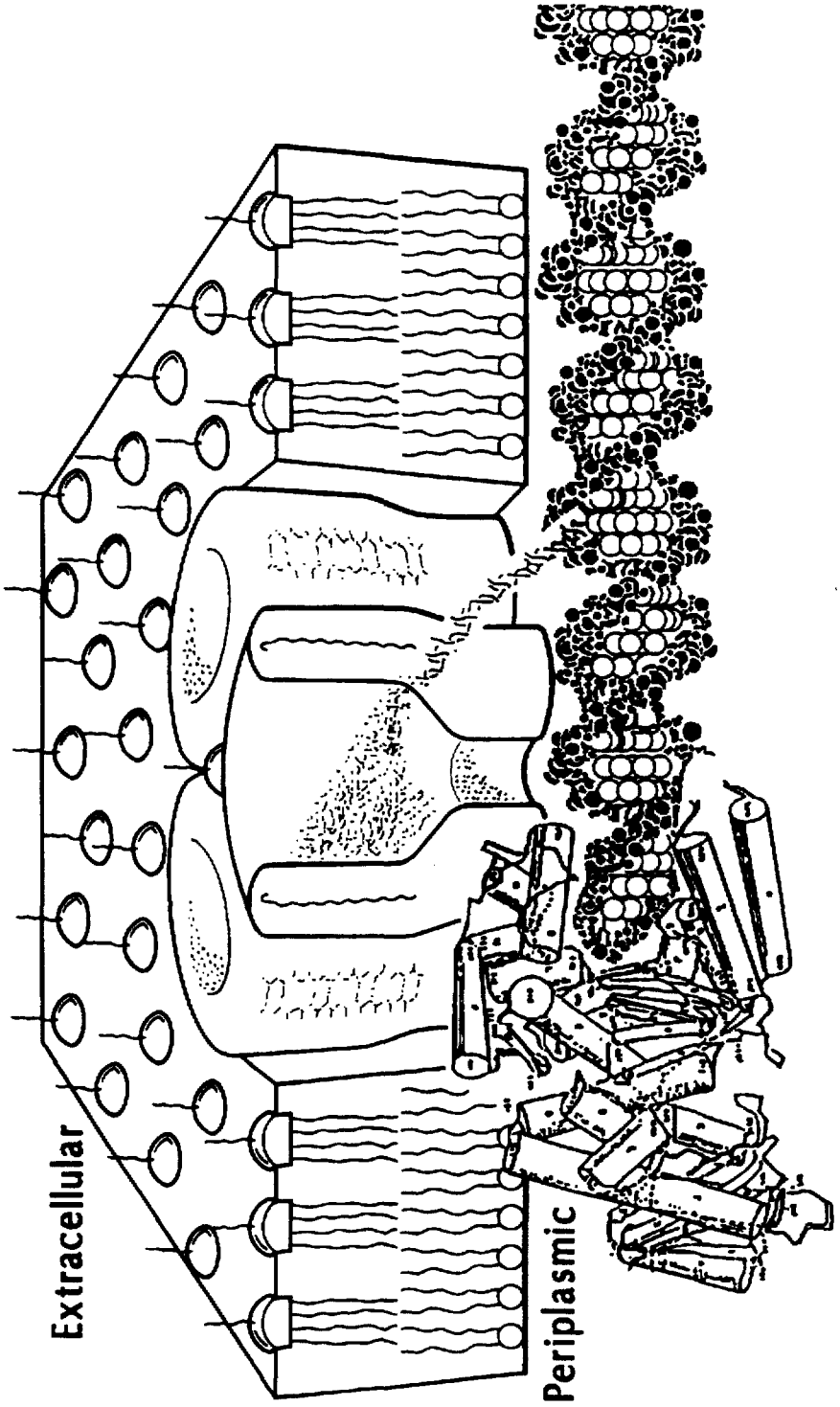


FIG. 2

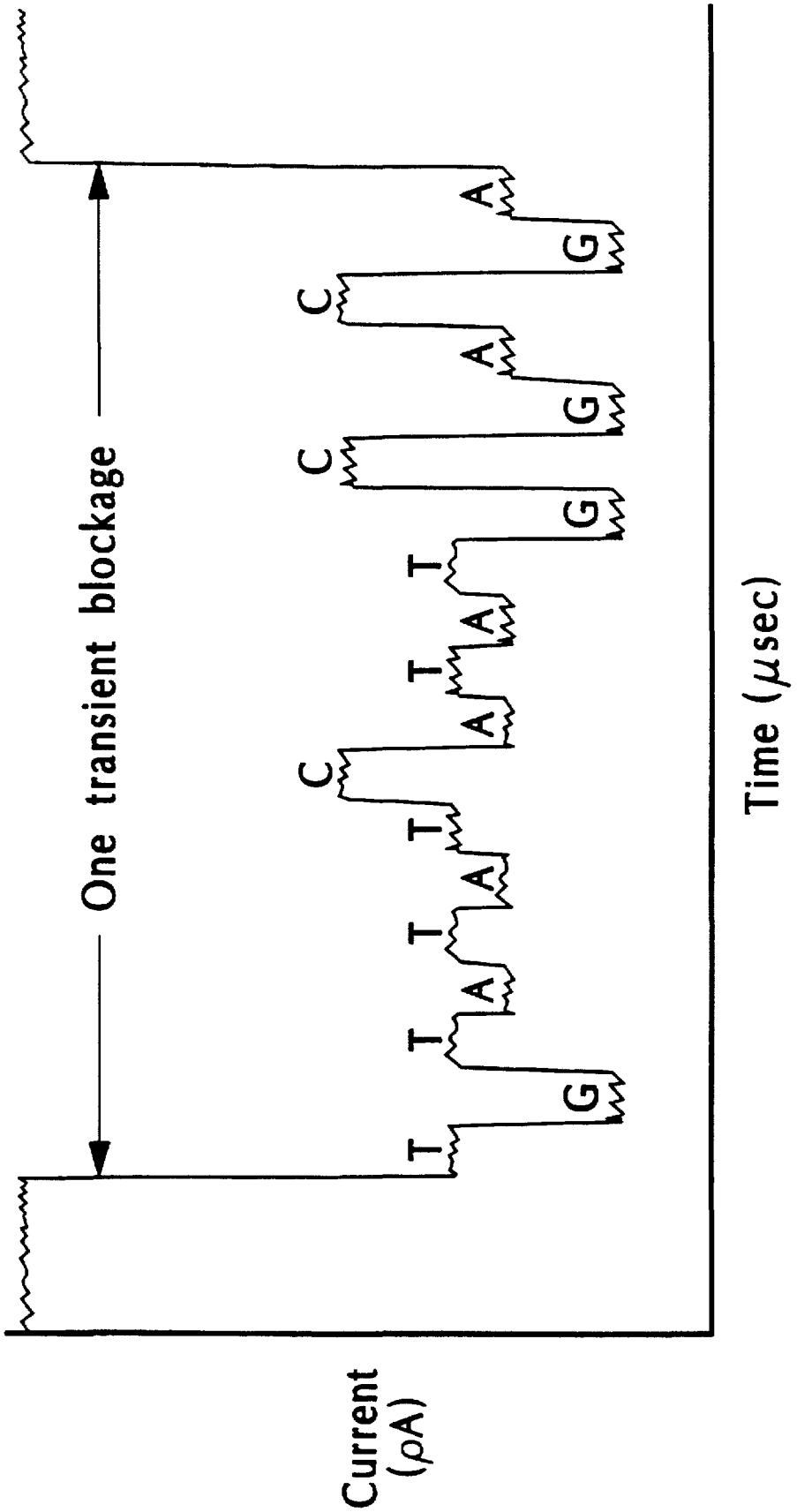

One transient blockage
T
T
A
T
A
T
A
C
T
A
T
A
C
G
C
A
G
A
G
Current (ρA)
Time (μsec)

FIG. 3

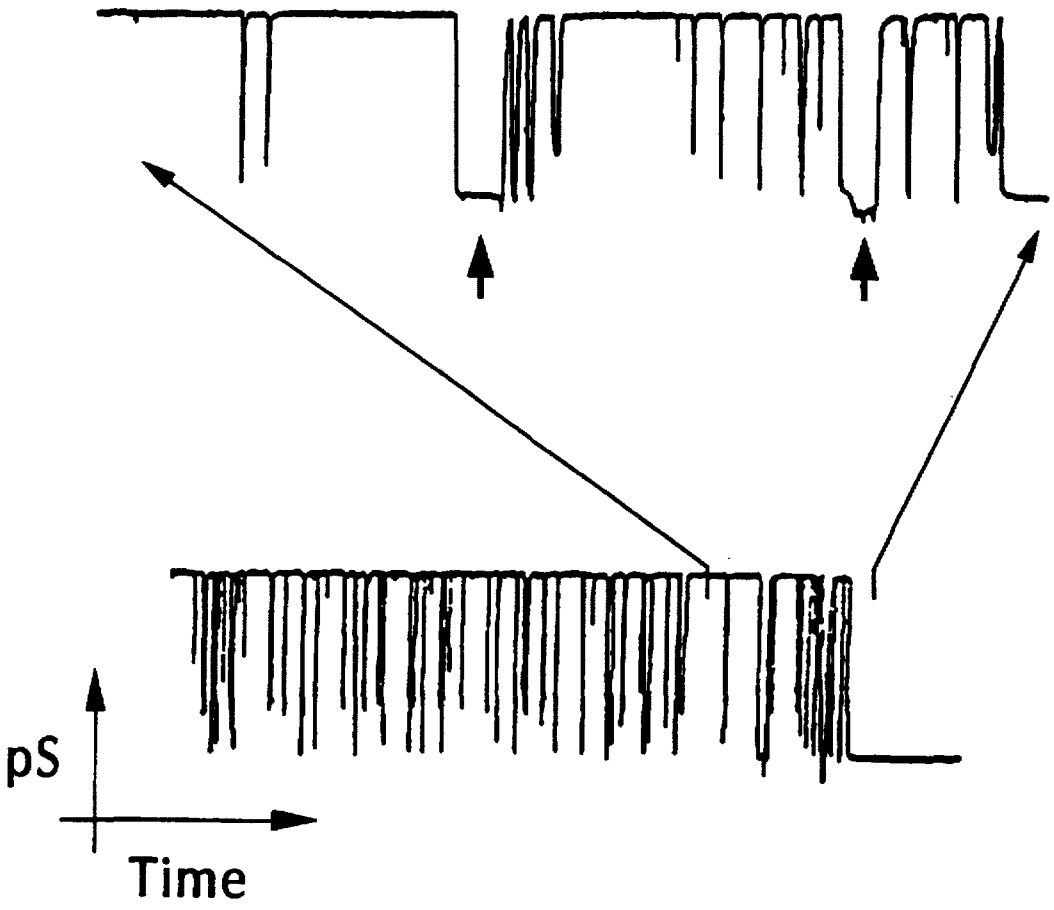


FIG. 4

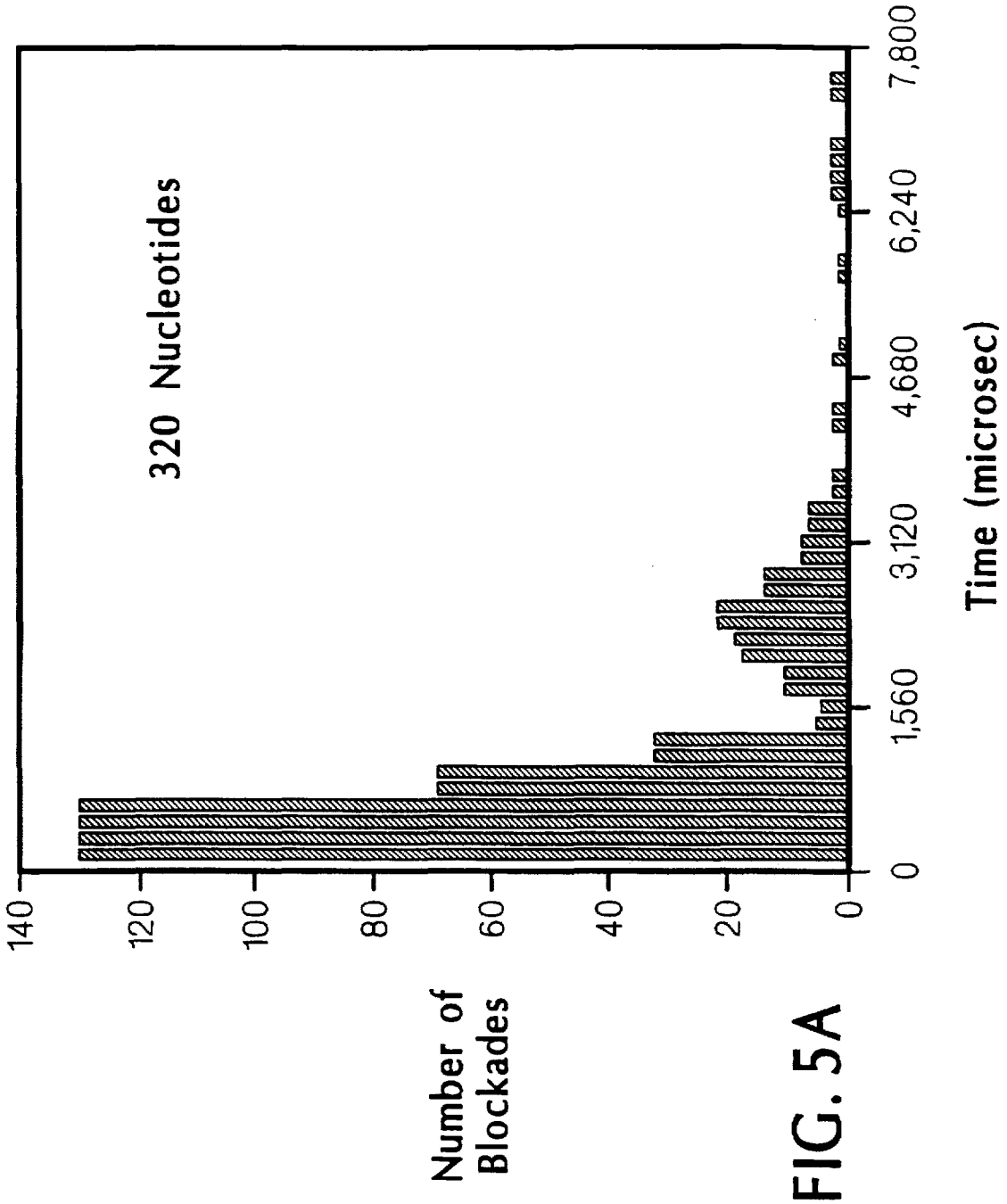

320 Nucleotides
Number of Blockades
140
120
100
80
60
40
20
0
0
1,560
3,120
4,680
6,240
7,800
Time (microsec)

FIG. 5A

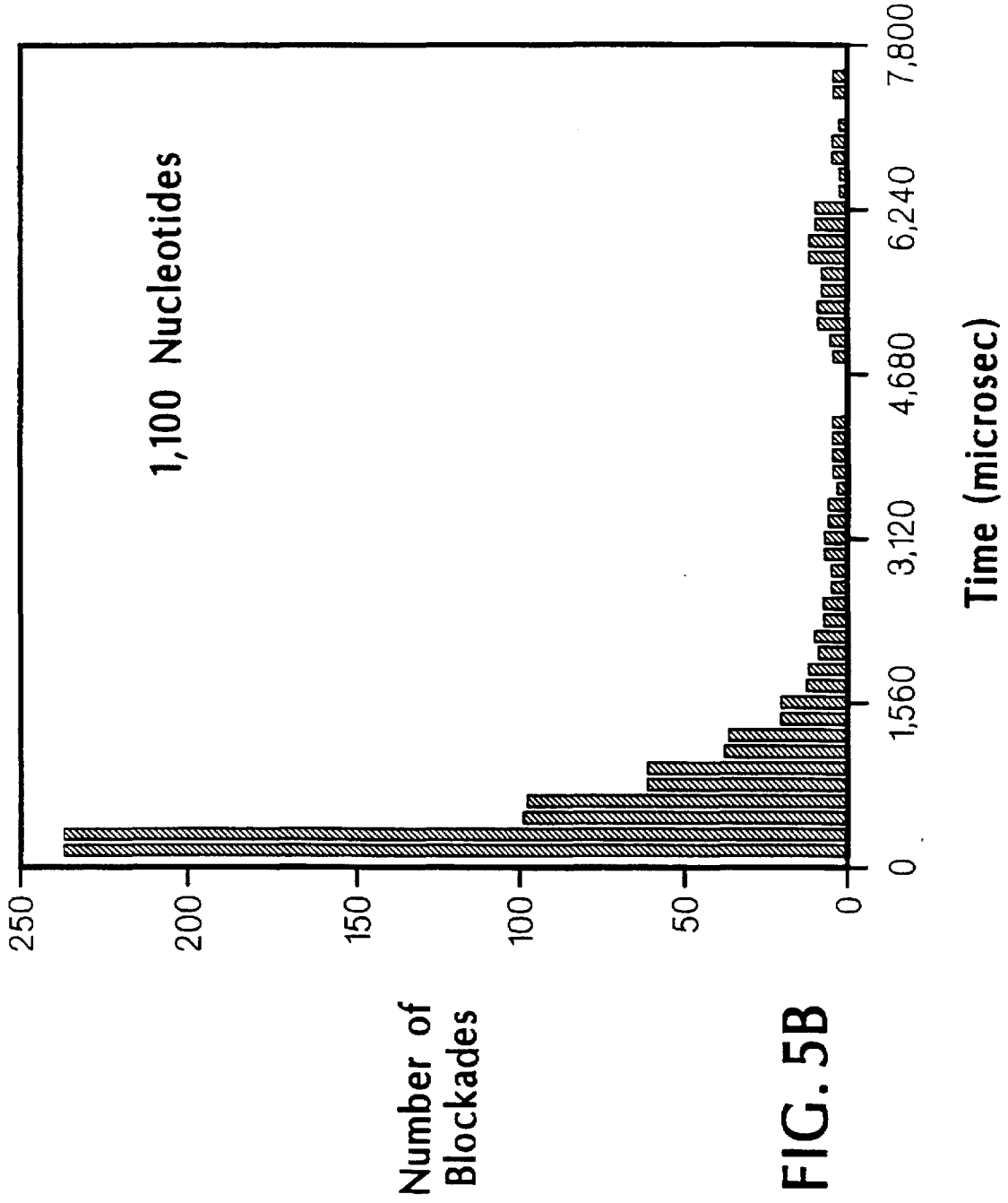

1,100 Nucleotides
Number of Blockades
250
200
150
100
50
0
0
1,560
3,120
4,680
6,240
7,800
Time (microsec)

FIG. 5B

CHARACTERIZATION OF INDIVIDUAL POLYMER MOLECULES BASED ON MONOMER-INTERFACE INTERACTIONS

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

This invention was made with Government support under NIH grant 1R21HG00811-01 (George Church) awarded by the Public Health Service and grant NSF #MCB-9421831 (Daniel Branton) awarded by the National Science Foundation. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

The general field of the invention is polymer characterization.

Rapid, reliable, and inexpensive characterization of polymers, particularly nucleic acids, has become increasingly important. One notable project, known as the Human Genome Project, has as its goal sequencing the entire human genome, over three billion nucleotides.

Typical current nucleic acid sequencing methods depend either on chemical reactions that yield multiple length DNA strands cleaved at specific bases, or on enzymatic reactions that yield multiple length DNA strands terminated at specific bases. In each of these methods, the resulting DNA strands of differing length are then separated from each other and identified in strand length order. The chemical or enzymatic reactions, as well as the technology for separating and identifying the different length strands, usually involve tedious, repetitive work. A method that reduces the time and effort required would represent a highly significant advance in biotechnology.

SUMMARY OF THE INVENTION

We have discovered a method for rapid, easy characterization of individual polymer molecules, for example polymer size or sequence determination. Individual molecules in a population may be characterized in rapid succession.

Stated generally, the invention features a method for evaluating a polymer molecule which includes linearly connected (sequential) monomer residues. Two separate pools of liquid-containing medium and an interface between the pools are provided. The interface between the pools is capable of interacting sequentially with the individual monomer residues of a single polymer present in one of the pools. Interface-dependent measurements are continued over time, as individual monomer residues of a single polymer interact sequentially with the interface, yielding data suitable to infer a monomer-dependent characteristic of the polymer. Several individual polymers, e.g., in a heterogenous mixture, can be characterized or evaluated in rapid succession, one polymer at a time, leading to characterization of the polymers in the mixture.

The method is broadly useful for characterizing polymers that are strands of monomers which, in general (if not entirely), are arranged in linear strands. Any polymer whose monomer units interact with the interface can be characterized. The method is particularly useful for characterizing biological polymers such as deoxyribonucleic acids, ribonucleic acids, polypeptides, and oligosaccharides, although other polymers may be evaluated. In some embodiments, a polymer which carries one or more charges (e.g., nucleic acids, polypeptides) will facilitate implementation of the invention.

The monomer-dependent characterization achieved by the invention may include identifying physical characteristics

such as the number and composition of monomers that make up each individual molecule, preferably in sequential order from any starting point within the polymer or its beginning or end. A heterogenous population of polymers may be characterized, providing a distribution of characteristics (such as size) within the population. Where the monomers within a given polymer molecule are heterogenous, the method can be used to determine their sequence.

The pools of medium used in the invention may be any fluid that permits adequate polymer mobility for interface interaction. Typically, the pools will be liquids, usually aqueous solutions or other liquids or solutions in which the polymers can be distributed.

The interface between the pools is designed to interact sequentially with the monomers of one polymer molecule at a time. As described in greater detail below, the useful portion of the interface may be a passage in or through an otherwise impermeable barrier, or it may be an interface between immiscible liquids. It is preferable that only one passage is present or functional in the impermeable barrier.

The interface-dependent measurements made according to the invention can be any measurement, e.g., physical or electrical, that varies with polymer-interface interaction. For example, physical changes the monomers cause as they interact sequentially with the interface may be measured. Current changes resulting from the polymer's interference with ion flow at the interface may be measured. The measurements may reflect the sequential interaction of the monomers with the interface, so as to permit evaluation of monomer-dependent characteristics of the polymer molecule (e.g., size or mass of individual monomers or of the entire polymer, or the sequence or identity of individual monomers which make up the polymer).

In one embodiment, the pools include electrically conductive medium which can be of the same or different compositions. The pools with conducting media are separated by an impermeable barrier containing an ion-permeable passage, and measurements of the interface characteristics include establishing an electrical potential between the two pools such that ionic current can flow across the ion permeable passage. When the polymer interacts sequentially with the interface at the ion permeable passage, the ionic conductance of the passage will change (e.g., decrease or increase) as each monomer interacts, thus indicating characteristics of the monomers (e.g., size, identity) and/or the polymer as a whole (e.g., size).

In a different embodiment, the concentration of polymers in a solution can be determined, using the conducting medium and ion-permeable passage described above. As a voltage differential is applied across the pools, the polymer molecules interact with the ion-permeable passage. The number of interactions (conductance change events) per unit time is proportional to the number of polymer molecules in the solution. This measurement is preferably made under relatively low resolution recording conditions, e.g., below the level of resolution of individual monomer/pore interactions.

The conducting medium used can be any medium, preferably a solution, more preferably an aqueous solution, which is able to carry electrical current. Such solutions generally contain ions as the current conducting agents, e.g., sodium, potassium, chloride, calcium, cesium, barium, sulfate, phosphate. Conductance (g) across the pore or channel is determined by measuring the flow of current across the pore or channel via the conducting medium. A voltage difference can be imposed across the barrier between

the pools by conventional means, e.g., via a voltage source which injects or administers current to at least one of the pools to establish a potential difference. Alternatively, an electrochemical gradient may be established by a difference in the ionic composition of the two pools, either with different ions in each pool, or different concentrations of at least one of the ions in the solutions or media of the pools. In this embodiment of the invention, conductance changes are measured and are indicative of monomer-dependent characteristics.

The term "ion permeable passages" used in this embodiment of the invention includes ion channels, ion-permeable pores, and other ion-permeable passages, and all are used herein to include any local site of transport through an otherwise impermeable barrier. For example, the term includes naturally occurring, recombinant, or mutant proteins which permit the passage of ions under conditions where ions are present in the medium contacting the channel or pore. Synthetic pores are also included in the definition. Examples of such pores can include, but are not limited to, chemical pores formed, e.g., by nystatin, ionophores (e.g., A23187; Pressman et al., *Ann. Rev. Biochem.* 45:501, 1976), or mechanical perforations of a membranous material. Proteinaceous ion channels can be voltage-gated or voltage independent, including mechanically gated channels (e.g., stretch-activated K^+ channels), or recombinantly engineered or mutated voltage dependent channels (e.g., Na^+ or K^+ channels constructed as is known in the art).

Another preferred type of passage is a protein which includes a portion of a bacteriophage receptor which is capable of binding all or part of a bacteriophage ligand (either a natural or functional ligand) and transporting bacteriophage DNA from one side of the interface to the other. The polymer to be characterized includes a portion which acts as a specific ligand for the bacteriophage receptor, so that it may be injected across the barrier/interface from one pool to the other.

The protein channels or pores of the invention can include those translated from one or more natural and/or recombinant DNA molecule(s) which includes a first DNA which encodes a channel or pore forming protein and a second DNA which encodes a monomer-interacting portion of a monomer polymerizing agent (e.g., a nucleic acid polymerase). The expressed protein or proteins are capable of non-covalent association or covalent linkage (any linkage herein referred to as forming an "assemblage" of "heterologous units"), and when so associated or linked, the polymerizing portion of the protein structure is able to polymerize monomers from a template polymer, close enough to the channel forming portion of the protein structure to measurably affect ion conductance across the channel. Alternatively, assemblages can be formed from unlike molecules, e.g., a chemical pore linked to a protein polymerase, but these assemblages still fall under the definition of a "heterologous" assemblage.

The invention also includes the recombinant fusion protein(s) translated from the recombinant DNA molecule(s) described above, so that a fusion protein is formed which includes a channel forming protein linked as described above to a monomer-interacting portion of a nucleic acid polymerase. Preferably, the nucleic acid polymerase portion of the recombinant fusion protein is capable of catalyzing polymerization of nucleotides. Preferably, the nucleic acid polymerase is a DNA or RNA polymerase, more preferably T7 RNA polymerase.

The polymer being characterized may remain in its original pool, or it may cross the passage. Either way, as a given polymer molecule moves in relation to the passage, individual monomers interact sequentially with the elements of the interface to induce a change in the conductance of the passage. The passages can be traversed either by polymer transport through the central opening of the passage so that the polymer passes from one of the pools into the other, or by the polymer traversing across the opening of the passage without crossing into the other pool. In the latter situation, the polymer is close enough to the channel for its monomers to interact with the passage and bring about the conductance changes which are indicative of polymer characteristics. The polymer can be induced to interact with or traverse the pore, e.g., as described below, by a polymerase or other template-dependent polymer replicating catalyst linked to the pore which draws the polymer across the surface of the pore as it synthesizes a new polymer from the template polymer, or by a polymerase in the opposite pool which pulls the polymer through the passage as it synthesizes a new polymer from the template polymer. In such an embodiment, the polymer replicating catalyst is physically linked to the ion-permeable passage, and at least one of the conducting pools contains monomers suitable to be catalytically linked in the presence of the catalyst. A "polymer replicating catalyst," "polymerizing agent" or "polymerizing catalyst" is an agent that can catalytically assemble monomers into a polymer in a template dependent fashion—i.e., in a manner that uses the polymer molecule originally provided as a template for reproducing that molecule from a pool of suitable monomers. Such agents include, but are not limited to, nucleotide polymerases of any type, e.g., DNA polymerases, RNA polymerases, tRNA and ribosomes.

The characteristics of the polymer can be identified by the amplitude or duration of individual conductance changes across the passage. Such changes can identify the monomers in sequence, as each monomer will have a characteristic conductance change signature. For instance, the volume, shape, or charges on each monomer will affect conductance in a characteristic way. Likewise, the size of the entire polymer can be determined by observing the length of time (duration) that monomer-dependent conductance changes occur. Alternatively, the number of monomers in a polymer (also a measure of size) can be determined as a function of the number of monomer-dependent conductance changes for a given polymer traversing a passage. The number of monomers may not correspond exactly to the number of conductance changes, because there may be more than one conductance level change as each monomer of the polymer passes sequentially through the channel. However, there will be a proportional relationship between the two values which can be determined by preparing a standard with a polymer of known sequence.

The mixture of polymers used in the invention does not need to be homogenous. Even when the mixture is heterogenous, only one molecule interacts with a passage at a time, yielding a size distribution of molecules in the mixture, and/or sequence data for multiple polymer molecules in the mixture.

In preferred embodiments, the passage is a natural or recombinant bacterial porin molecule. In other preferred embodiments, the passage is a natural or recombinant voltage-sensitive or voltage gated ion channel, preferably one which does not inactivate (whether naturally or through recombinant engineering as is known in the art). "Voltage sensitive" or "gated" indicates that the channel displays activation and/or inactivation properties when exposed to a particular range of voltages. Preferred channels for use in the invention include the α-hemolysin toxin from *S. aureus* and maltoporin channels.

In an alternate embodiment of the invention, the pools of medium are not necessarily conductive, but are of different compositions so that the liquid of one pool is not miscible in the liquid of the other pool and the interface is the immiscible interface between the pools. In order to measure the characteristics of the polymer, a polymer molecule is drawn through the interface of the liquids, resulting in an interaction between each sequential monomer of the polymer and the interface. The sequence of interactions as the monomers of the polymer are drawn through the interface is measured, yielding information about the sequence of monomers that characterize the polymer. The measurement of the interactions can be by a detector that measures the deflection of the interface (caused by each monomer passing through the interface) using reflected or refracted light, or a sensitive gauge capable of measuring intermolecular forces. Several methods are available for measurement of forces between macromolecules and interfacial assemblies, including the surface forces apparatus (Israelachvili, *Intermolecular and Surface Forces*, Academic Press, New York, 1992), optical tweezers (Ashkin et al., *Oppt. Lett.*, 11:288, 1986; Kuo and Sheetz, *Science*, 260:232, 1993; Svoboda et al., *Nature* 365:721, 1993), and atomic force microscopy (Quate, *F. Surf. Sci.* 299:980, 1994; Mate et al., *Phys. Rev. Lett.* 59:1942, 1987; Frisbie et al., *Science* 265:71, 1994; all hereby incorporated by reference).

The interactions between the interface and the monomers in the polymer are suitable to identify the size of the polymer, e.g., by measuring the length of time during which the polymer interacts with the interface as it is drawn across the interface at a known rate, or by measuring some feature of the interaction (such as deflection of the interface, as described above) as each monomer of the polymer is sequentially drawn across the interface. The interactions can also be sufficient to ascertain the identity of individual monomers in the polymer.

This invention offers advantages particularly in nucleotide sequencing, e.g., reduction in the number of sequencing steps, and increasing the speed of sequencing and the length of molecule capable of being sequenced. The speed of the method and the size of the polymers it can sequence are particular advantages of the invention. The linear polymer may be very large, and this advantage will be especially useful in reducing template preparation time, sequencing errors and analysis time currently needed to piece together small overlapping fragments of a large gene or stretch of polymer.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic representation of an embodiment of DNA characterization by the method of the invention. The unobstructed ionic current (illustrated for the channel at the top of the diagram), is reduced as a polymeric molecule begins its traversal through the pore (illustrated for the channel at the bottom of the diagram). The monomeric units of the polymer (drawn as different sized ovals on the strand) interfere sequentially and differentially with the flow of ions through the channel.

FIG. 2 is a schematic representation of an implementation of DNA sequencing by the method of the invention. In this embodiment, the polymer is drawn across the opening of the channel, but is not drawn through the channel. The channel, e.g., a porin, is inserted in the phospholipid bilayer. A polymerase domain is fused by its N-terminus to the C-terminus of one of the porin monomers (the porin C-termini are on the periplasmic side of the membrane in both *Rhodobacter capsulatus* and LamB porins). Fusions on the other side of the membrane can also be made. Malto-oligosaccharides can bind and block current from either side. The polymerase is shown just prior to binding to the promoter. A non-glucosylated base is shown near a pore opening, while a penta-glucosylated cytosine is shown 10 bp away. The polymerase structure represented is that of DNA polymerase I (taken from Ollis et al., 1985, *Nature*, 313:762–66), and the general porin model is from Jap (1989, *J. Mol. Biol.*, 205:407–19).

FIG. 3 is a schematic representation of DNA sequencing results by the method of the invention. The schematic depicts, at very high resolution, one of the longer transient blockages such as can be seen in FIG. 4. The monomeric units of DNA (bases G, A, T, and C) interfere differentially with the flow of ions through the pore, resulting in discrete conductance levels that are characteristic of each base. The order of appearance of the conductance levels sequentially identifies the monomers of the DNA.

FIG. 4 is a recording of the effect of polyadenylic acid (poly A) on the conductance of a single α-hemolysin channel in a lipid bilayer between two aqueous compartments containing 1 M NaCl, 10 mM Tris, pH 7.4. Before addition of RNA, the conductance of the channel was around 850 pS. The cis compartment, to which poly A is added, is −120 mV with respect to the trans compartment. After adding poly A to the cis compartment, the conductance of the α-hemolysin channel begins to exhibit transient blockages (conductance decreases to about 100 pS) as individual poly A molecules are drawn across the channel from the cis to the trans compartment. When viewed at higher resolution (expanded time scale, at top), the duration of each transient blockage is seen to vary between less than 1 msec up to 10 msec. Arrows point to two of the longer duration blockages. See FIGS. 5A and 5B for histograms of blockage duration.

FIGS. 5A and 5B are comparisons of blockage duration with purified RNA fragments of 320nt (FIG. 5A) and 1100nt (FIG. 5B) lengths. The absolute number of blockades plotted in the two histograms are not comparable because they have not been normalized to take into account the different lengths of time over which the data in the two graphs were collected.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

As summarized above, we have determined a new method for rapidly analyzing polymers such as DNA and RNA. We illustrate the invention with two primary embodiments. In one embodiment, the method involves measurements of ionic current modulation as the monomers (e.g., nucleotides) of a linear polymer (e.g., nucleic acid molecule) pass through or across a channel in an artificial membrane. During polymer passage through or across the channel, ionic currents are reduced in a manner that reflects the properties of the polymer (length, concentration of polymers in solution, etc.) and the identities of the monomers. In the second embodiment, an immiscible interface is created between two immiscible liquids, and, as above, polymer passage through the interface results in monomer interactions with the interface which are sufficient to identify characteristics of the polymer and/or the identity of the monomers.

The description of the invention will be primarily concerned with sequencing nucleic acids, but this is not

intended to be limiting. It is feasible to size and sequence polymers other than nucleic acids by the method of the invention, including linear protein molecules which include monomers of amino acids. Other linear arrays of monomers, including chemicals (e.g., biochemicals such as polysaccharides), may also be sequenced and characterized by size.

I. Polymer Analysis Using Conductance Changes Across An Interface

Sensitive single channel recording techniques (i.e., the patch clamp technique) can be used in the invention, as a rapid, high-resolution approach allowing differentiation of nucleotide bases of single DNA molecules, and thus a fast and efficient DNA sequencing technique or a method to determine polymer size or concentration (FIGS. 1 and 2). We will describe methods to orient DNA to a pore molecule in two general configurations (see FIGS. 1 and 2) and record conductance changes across the pore (FIG. 3). One method is to use a pore molecule such as the receptor for bacteriophage lambda (LamB) or α-hemolysin, and to record the process of DNA injection or traversal through the channel pore when that channel has been isolated on a membrane patch or inserted into a synthetic lipid bilayer (FIG. 1). Another method is to fuse a DNA polymerase molecule to a pore molecule and allow the polymerase to move DNA over the pore's opening while recording the conductance across the pore (FIG. 2). A third method is to use a polymerase on the trans side of the membrane/pore divider to pull a single stranded nucleic acid through the pore from the cis side (making it double stranded) while recording conductance changes. A fourth method is to establish a voltage gradient across a membrane containing a channel (e.g., α-hemolysin) through which a single stranded or double stranded DNA is electrophoresed.

The apparatus used for this embodiment includes 1) an ion-conducting pore or channel, perhaps modified to include a linked or fused polymerizing agent, 2) the reagents necessary to construct and produce a linear polymer to be characterized, or the polymerized molecule itself, and 3) an amplifier and recording mechanism to detect changes in conductance of ions across the pore as the polymer traverses its opening.

A variety of electronic devices are available which are sensitive enough to perform the measurements used in the invention, and computer acquisition rates and storage capabilities are adequate for the rapid pace of sequence data accumulation.

A. Characteristics Identified by the Methods

1) Size/Length of Molecules The size or length of a polymer can be determined by measuring its residence time in the pore or channel, e.g., by measuring duration of transient blockade of current. The relationship between this time period and the length of the polymer can be described by a reproducible mathematical function which depends on the experimental condition used. The function is likely a linear function for a given type of polymer (e.g., DNA, RNA, polypeptide), but if it is described by another function (e.g., sigmoidal or exponential), accurate size estimates may be made by first preparing a standard curve using known sizes of like linear molecules.

2) Identity of Residues/Monomers

The chemical composition of individual monomers is sufficiently variant to cause characteristic changes in channel conductance as each monomer traverses the pore due to physical configuration, size/volume, charge, interactions with the medium, etc. For example, our experimental data suggest that poly A reduces conductance more than does poly U, indicating a physical difference between purines and pyrimidines that forms the basis of nucleotide identification in this invention.

The nucleotide bases of DNA will influence pore conductance during traversal, but if the single channel recording techniques are not sensitive enough to detect differences between normal bases in DNA, it is practical to supplement the system's specificity by using modified bases. The modifications should be asymmetrical (on only one strand of double stranded template), to distinguish otherwise symmetrical base pairs.

Modified bases are readily available. These include: 1) methylated bases (lambda can package and inject DNA with or without methylated A's and C's), 2) highly modified bases found in the DNAs of several bacteriophage (e.g. T4, SP15), many of which involve glycosylations coupled with other changes (Warren, 1980, *Ann. Rev. Microbiol.*, 34:137–58), and 3) the modified nucleotide triphosphates that can be incorporated by DNA polymerase (e.g. biotinylated, digoxigenated, and fluorescently tagged triphosphates).

In order to identify the monomers, conditions should be appropriate to avoid secondary structure in the polymer to be sequenced (e.g., nucleic acids); if necessary, this can be achieved by using a recording solution which is denaturing. Using single stranded DNA, single channel recordings can be made in up to 40% formamide and at temperatures as high as 45° C. using e.g., the αhemolysin toxin protein in a lipid bilayer. These conditions are not intended to exclude use of any other denaturing conditions. One skilled in the art of electrophysiology will readily be able to determine suitable conditions by 1) observing incorporation into the bilayer of functional channels or pores, and 2) observing transient blockades of conductance uninterrupted by long-lived blockades caused by polymers becoming stuck in the channel because of secondary structure. Denaturing conditions are not always necessary for the polymerase-based methods or for double stranded DNA methods of the invention. They may not be necessary for single stranded methods either, if the pore itself is able to cause denaturation, or if the secondary structure does not interfere.

3) Concentration of Polymers in Solutions

Concentration of polymers can be rapidly and accurately assessed by using relatively low resolution recording conditions and analyzing the number of conductance blockade events in a given unit of time. This relationship should be linear and proportional (the greater the concentration of polymers, the more frequent the current blockage events), and a standardized curve can be prepared using known concentrations of polymer.

B. Principles and Techniques

1) Recording Techniques

The conductance monitoring methods of the invention rely on an established technique, single-channel recording, which detects the activity of molecules that form channels in biological membranes. When a voltage potential difference is established across a bilayer containing an open pore molecule, a steady current of ions flows through the pore from one side of the bilayer to the other. The nucleotide bases of a DNA molecule, for example, passing through or over the opening of a channel protein, disrupt the flow of ions through the pore in a predictable way. Fluctuations in the pore's conductance caused by this interference can be detected and recorded by conventional single-channel recording techniques. Under appropriate conditions, with modified nucleotides if necessary, the conductance of a pore can change to unique states in response to the specific bases in DNA.

This flux of ions can be detected, and the magnitude of the current describes the conductance state of the pore. Multiple conductance states of a channel can be measured in a single recording as is well known in the art. By recording the fluctuations in conductance of the maltoporin (LamB) pore, for example, when DNA is passed through it by phage lambda injection or over its opening by the action of a polymerase fused to the surface of the LamB protein, we estimate that a sequencing rate of 100–1000 bases/sec/pore can be achieved.

The monitoring of single ion channel conductance is an inexpensive, viable method that has been successful for the last two decades and is in very wide spread current use. It directly connects movements of single ions or channel proteins to digital computers via amplifiers and analog to digital (A to D, A/D) converters. Single channel events taking place in the range of a few microseconds can be detected and recorded (Hamill et al., 1981, *Pfluegers Arch. Eur. J. Physiol.*, 391:85–100). This level of time resolution ranges from just sufficient to orders of magnitude greater than the level we need, since the time frame for movement of nucleotide bases relative to the pore for the sequencing method is in the range of microseconds to milliseconds. The level of time resolution required depends on the voltage gradient or the enzyme turnover number if the polymer is moved by an enzyme. Other factors controlling the level of time resolution include medium viscosity, temperature, etc.

The characteristics and conductance properties of any pore molecule that can be purified can be studied in detail using art-known methods (Sigworth et al., supra; Heinemann et al., 1988, *Biophys. J.*, 54:757–64; Wonderlin et al., 1990, *Biophys. J.*, 58:289-97). These optimized methods are ideal for our polymer sequencing application. For example, in the pipette bilayer technique, an artificial bilayer containing at least one pore protein is attached to the tip of a patch-clamp pipette by applying the pipette to a preformed bilayer reconstituted with the purified pore protein in advance. Due to the very narrow aperture diameter of the patch pipette tip (2 microns), the background noise for this technique is significantly reduced, and the limit for detectable current interruptions is about 10 microseconds (Sigworth et al., supra; Heinemann et al., 1990, *Biophys. J.*, 57:499–514). Purified channel protein can be inserted in a known orientation into preformed lipid bilayers by standard vesicle fusion techniques (Schindler, 1980, *FEBS Letters*, 122:77–79), or any other means known in the art, and high resolution recordings are made. The membrane surface away from the pipette is easily accessible while recording. This is important for the subsequent recordings that involve added DNA. The pore can be introduced into the solution within the patch pipette rather than into the bath solution.

An optimized planar lipid bilayer method has recently been introduced for high resolution recordings in purified systems (Wonderlin et al., supra). In this method, bilayers are formed over very small diameter apertures (25–80 microns) in plastic. This technique has the advantage of allowing access to both sides of the bilayer, and involves a slightly larger bilayer target for reconstitution with the pore protein. This optimized bilayer technique is an alternative to the pipette bilayer technique.

Instrumentation is needed which can apply a variable range of voltages from about +400 mV to –400 mV across the channel/membrane, assuming that the trans compartment is established to be 0 mV; a very low-noise amplifier and current injector, analog to digital (A/D) converter, data acquisition software, and electronic storage medium (e.g., computer disk, magnetic tape). Equipment meeting these criteria is readily available, such as from Axon Instruments, Foster City, Calif. (e.g., Axopatch 200A system; pclamp 6.0.2 software).

Preferred methods of large scale DNA sequencing involve translating from base pairs to electronic signals as directly and as quickly as possible in a way that is compatible with high levels of parallelism, miniaturization and manufacture. The method should allow long stretches (even stretches over 40 kbp) to be read so that errors associated with assembly and repetitive sequence can be minimized. The method should also allow automatic loading of (possibly non-redundant) fresh sequences.

2) Channels and Pores Useful in the Invention Any channel protein which has the characteristics useful in the invention (e.g., minimum pore size around 2 Å, maximum around 9 nm; conducts current) may be employed. Pore sizes across which polymers can be drawn may be quite small and do not necessarily differ for different polymers. Pore sizes through which a polymer is drawn will be e.g., approximately 0.5 –2.0 nm for single stranded DNA; 1.0 –3.0 nm for double stranded DNA; and 1.0 – 4.0 nm for polypeptides. These values are not absolute, however, and other pore sizes might be equally functional for the polymer types mentioned above.

Examples of bacterial pore-forming proteins which can be used in the invention include Gramicidin (e.g., Gramicidin A, B, C, D, or S, from *Bacillus brevis;* available from Fluka, Ronkonkoma, N.Y.); Valinomycin (from *Streptomyces fulvissimus;* available from Fluka), LamB (maltoporin), OmpF, OmpC, or PhoE from *Escherichia coli*, Shigella, and other Enterobacteriaceae, alpha-hemolysin (from *S. aureus*), Tsx, the F-pilus, and mitochondrial porin (VDAC). This list is not intended to be limiting.

A modified voltage-gated channel can also be used in the invention, as long as it does not inactivate quickly, e.g., in less than about 500 msec (whether naturally or following modification to remove inactivation) and has physical parameters suitable for e.g., polymerase attachment (recombinant fusion proteins) or has a pore diameter suitable for polymer passage. Methods to alter inactivation characteristics of voltage gated channels are well known in the art (see e.g., Patton, et al., *Proc. Natl. Acad. Sci. USA*, 89:10905–09 (1992); West, et al., *Proc. Natl. Acad. Sci. USA*, 89:10910–14 (1992); Auld, et al., *Proc. Natl. Acad. Sci. USA*, 87:323–27 (1990); Lopez, et al., *Neuron*, 7:327–36 (1991); Hoshi, et al., *Neuron*, 7:547–56 (1991); Hoshi, et al., *Science*, 250:533–38 (1990), all hereby incorporated by reference).

Appropriately sized physical or chemical pores may be induced in a water-impermeable barrier (solid or membranous) up to a diameter of about 9 nm, which should be large enough to accommodate most polymers (either through the pore or across its opening). Any methods and materials known in the art may be used to form pores, including track etching and the use of porous membrane templates which can be used to produce pores of the desired material (e.g., scanning-tunneling microscope or atomic force microscope related methods).

Chemical channels or pores can be formed in a lipid bilayer using chemicals (or peptides) such as Nystatin, as is well known in the art of whole-cell patch clamping ("perforated patch" technique); ionophores such as A23187 (Calcimycin), ETH 5234, ETH 157 (all chemicals available from Fluka, Ronkonkoma, N.Y.; this list is not intended to be limiting), peptide channels such as Alamethicin, etc.

To produce pores linked with polymerase, synthetic/recombinant DNA coding for a fusion protein can be tran-

scribed and translated, then inserted into an artificial membrane in vitro. For example, the C-terminus of *E. coli* DNA polymerase I (and by homology, T7 polymerase) is very close to the surface of the major groove of the newly synthesized DNA. If the C-terminus of a polymerase is fused to the N-terminus of a pore forming protein such as colicin E1 and the colicin is inserted into an artificial membrane, one opening of the colicin pore should face the DNA's major groove and one should face the opposite side of the lipid bilayer. For example, the colicin molecule can be modified to achieve a pH optimum compatible with the polymerase as in Shiver et al. (*J. Biol. Chem.*, 262:14273–14281 1987, hereby incorporated by reference). Both pore and polymerase domain can be modified to contain cysteine replacements at points such that disulfide bridges form to stabilize a geometry that forces the pore opening closer to the major groove surface and steadies the polymer as it passes the pore opening. The loops of the pore domain at this surface can be systematically modified to maximize sensitivity to changes in the DNA sequence.

C. General Considerations for Conductance Based Measurements

1) Electrical/Channel Optimization

The conductance of a pore at any given time is determined by its resistance to ions passing through the pore (pore resistance) and by the resistance to ions entering or leaving the pore (access resistance). For a pore's conductance to be altered in discrete steps, changes in one or both of these resistance factors will occur by unit values. The base pairs of a DNA molecule represent discrete units that are distinct from each other along the phosphate backbone. As long as the orientation of DNA to the pore remains relatively constant, and the membrane potential does not change, as each base pair passes by (or through) the pore, it is likely to interfere with a reproducible number of ions. Modifications made to the individual bases would influence the magnitude of this effect.

To resolve stretches of repeating identical bases accurately, and to minimize reading errors in general, it may be useful for the pore to register a distinct (probably higher) level of conductance in between the bases. This can take place naturally in the pore-polymerase system with helix rotation during polymerization, or in the phage system between entry of base pairs into the pore, or when the regions in between base pairs pass by a rate limiting site for ion flux inside the pore. Modified bases used to distinguish nucleotide identities may also contribute significantly to this issue, because they should magnify the conductance effect of the bases relative to the effect of regions in between the bases. With single strand passage through a pore, charged phosphates may punctuate the passage of each base by brief, higher conductance states. Also, if the rate of movement is constant, then punctuation between bases may not be required to resolve stretches of repeating identical bases.

Altered conductance states have been described for many channels, including some LamB mutants (Dargent et al., 1988, supra). A mutant may be a valuable alternative to a wild type channel protein if its fluctuation to a given state is sensitive to nucleotide bases in DNA. Alternative systems can also be developed from other channel proteins that are known to have multiple single channel conductance states. Examples of these are the alamethicin channel, which under certain conditions fluctuates through at least 20 discrete states (Taylor et al., 1991, *Biophys. J.*, 59:873–79), and the OmpF porin, which shows gating of its individual monomers giving rise to four discrete states (Lakey et al., 1989, *Eur. J. Biochem.*, 186:303–308).

Since channel events can be resolved in the microsecond range with the high resolution recording techniques available, the limiting issue for sensitivity with the techniques of our invention is the amplitude of the current change between bases. Resolution limits for detectable current are in the 0.2 pA range (1 pA=6.24×10^6 ions/sec). Each base affecting pore current by at least this magnitude is detected as a separate base. It is the function of modified bases to affect current amplitude for specific bases if the bases by themselves are poorly distinguishable.

One skilled in the art will recognize that there are many possible configurations of the sequencing method described herein. For instance, lipid composition of the bilayer may include any combination of non-polar (and polar) components which is compatible with pore or channel protein incorporation. Any configuration of recording apparatus may be used (e.g., bilayer across aperture, micropipette patches, intra-vesicular recording) so long as its limit of signal detection is below about 0.5 pA, or in a range appropriate to detect monomeric signals of the polymer being evaluated. If polymeric size determination is all that is desired, the resolution of the recording apparatus may be much lower.

A Nernst potential difference, following the equation

$$E_{ion}=(RT/ZF)\log_e([ion]_o/[ion]_i)$$

where E_{ion} is the solvent ion (e.g., potassium ion) equilibrium potential across the membrane, R is the gas constant, T is the absolute temperature, z is the valency of the ion, F is Faraday's constant, $[ion]_o$ is the outside and $[ion]_i$ is the inside ionic concentration (or trans and cis sides of the bilayer, respectively),

can be established across the bilayer to force polymers across the pore without supplying an external potential difference across the membrane. The membrane potential can be varied ionically to produce more or less of a differential or "push." The recording and amplifying apparatus is capable of reversing the gradient electrically to clear blockages of pores caused by secondary structure or cross-alignment of charged polymers.

2) Optimization of Methods In an operating system of the invention, one can demonstrate that the number of transient blockades observed is quantitatively related to the number of polymer molecules that move through the channel from the cis to the trans compartment. By sampling the trans compartment solution after observing one to several hundred transient blockades and using quantitative, competitive PCR assays (e.g., as in Piatak et al., 1993, *BioTechniques*, 14:70–79) it is possible to measure the number of molecules that have traversed the channel. Procedures similar to those used in competitive PCR can be used to include an internal control that will distinguish between DNA that has moved through the channel and contaminating or aerosol DNA.

Further steps to optimize the method may include:

1. Slowing the passage of polynucleotides so that individual nucleotides can be sensed. Since the blockade durations we observed are in the millisecond range, each nucleotide in a one or two thousand monomer-long polynucleotide occupies the channel for just a few microseconds. To measure effects of individual nucleotides on the conductance, substantially reducing the velocity may offer substantial improvement. Approaches to accomplish this include: (a) increasing the viscosity of the medium, (b) establishing the lower limit of applied potential that will move polynucleotides into the channel (c) use of high processivity polymerase in the trans compartment to "pull" DNA through the pore in place of voltage gradients. Using

enzymes to pull the DNA through the pore may also solve another potential problem (see 3, below).

2. Making a channel in which an individual nucleotide modulates current amplitude. While a-toxin may give rise to distinguishable current amplitudes when different mono-polynucleotides pass through the channel, 4–5 nucleotides in the strand necessarily occupy the length of its approximately 50Å long channel at any given time. Ionic current flow may therefore reflect the sum of the nucleotide effects, making it difficult to distinguish monomers. To determine current modulation attributable to individual monomers, one may use channels containing a limiting aperture that is much shorter than the full length of the overall channel (Weiss et al., supra). Artificia channels based on self-assembled cylindrical B-sheet peptides (Ghadiri et al., 1994, *Nature*, 369:301–304) are useful in that their diameter can readily be engineered to desired dimensions.

3. Enhancing movement of DNA in one direction. If a DNA molecule is being pulled through a channel by a voltage gradient, the probability of its moving backward against the gradient will be given by

$$e^{-}$$

(energy to move against the voltage gradient/kT)

where kT is energy associated with thermal fluctuations. For example, using reasonable assumptions for the effective charge density of the DNA polyelectrolyte in buffer (Manning, 1969, *J. Chem. Phys.*, 51:924–33), at room temperature the probability of thermal energy moving the DNA molecule backward 10Åagainst a 100 mV voltage gradient $_{-e}^{-4}$, or about one in fifty. Should this problem exist, some kind of ratchet mechanism, possibly a polymerase or other DNA binding protein, may be useful in the trans chamber to prevent backward movements of the DNA.

3) Advantages of Single Channel Sequencing

The length of continuous DNA sequence obtainable from the methods described herein will only be limited in certain embodiments (e.g., by the packaging limit of phage lambda heads (~50 kb) or by the size of the template containing polymerase promoter sequences). Other embodiments (e.g., voltage gradients) have no such limitation and should even make it possible to sequence DNA directly from tissue samples, since the technique is not limited to cloned DNA. Having large contiguous sequence as primary input data will substantially reduce the complexity of sequence assembly, particularly in the case of repetitive DNA. There are other applications if consistent conductance behaviors can be correlated with particular properties of given molecules (i.e. shape).

D. Specific Methods and Examples of Current Based Characterization

The following specific examples of current based polymer characterization are presented to illustrate, not limit the invention.

1) The LamB pore Maltoporin (LamB) is an outer membrane protein from *E. coli* that functions as a passive diffusion pore (porin) for small molecules and as a specific transport pore for passage of maltose and maltodextrins (Szmelcman et al., 1975, *J. Bacteriol.*, 124:112–18). It is also the receptor for bacteriophage lambda (Randall-Hazelbauer and Schwartz, 1973, *J. Bacteriol.* 116:1436–1446). Three identical copies of the LamB gene product assemble to form the native pore. Each subunit (MW ~48,000) is composed of predominantly beta-structure and is a pore in itself, though it is thought that the three pores fuse into one at the periplasmic side of the membrane (Lepault et al., 1988, *EMBO, J.*, 7:261–68).

A protein folding model for LamB is available that predicts which portions of the mature protein reside on the external and periplasmic surfaces of the membrane (Charbit et al., 1991, *J. Bacteriol.*, 173:262–75). Permissive sites in the protein have been mapped to several extramembranous loops that tolerate the insertion of foreign polypeptides without significantly disrupting pore properties (Boulain et al., 1986, *Mol. Gen. Genet.*, 205:339–48; Charbit et al., 1986, *EMBO J.*, 5:3029–37; Charbit et al., 1991, supra). The LamB protein has been crystallized and a high resolution structure derived (3.1 Å) (Schirmer et al., 1995, *Science*, 267:512–514).

The pore properties of wild type LamB and a few mutant proteins have been studied at low resolution in planar lipid bilayer single channel recordings (Benz et al., 1986, *J. Bacteriol.*, 165:978–86; Benz et al., 1987, *J. Membrane Biol.*, 100:21–29; Dargent et al., 1987, *FEBS Letters*, 220:136–42; Dargent et al., 1988, *J. Mol. Biol.*, 201:497–506). The pore has a very stable conductance of 150 pS in 1M NaCl, and shows selectivity for maltose and maltodextrins. These molecules effectively block conductance of the pore. One LamB mutant (Tyr^{163}→Asp) exhibits distinct sublevels of conductance (30 pS each).

The LamB pore is extremely stable, and high time resolution recordings can be made for use in this invention. The time resolution of channel conductance measurements with the conventional planar lipid bilayer technique is limited because of the background noise associated with the high electrical capacitance of bilayers formed on large diameter apertures (100–200 microns), but smaller apertures or insulated glass microelectrodes can improve the resolution of LamB channel recordings. Preferably, improved LamB conductance recordings will use the pipette bilayer technique (Sigworth et al., supra).

Example 1: Conductance Measurements of Lambda DNA Injection

Bacteriophage lambda injects its DNA through the LamB pore at a rate of about 1000 bp/sec (Novick et al., 1988, *Biochemistry*, 27:7919–24). Lambda will inject its DNA into liposomes reconstituted with purified *E. coli* LamB protein. Alcohol or chloroform may be useful in this system (Randall-Hazelbauer and Schwartz, 1973, supra), but if the solvents disrupt current recordings, one can use either the LamB protein from a closely related species, e.g., *Shigella sonnei* 3070, which allows spontaneous lambda DNA injection into liposomes (Schwartz et al., 1975, *J. Virol.*, 15:679–85; Roessner et al., 1983, *J. Biol. Chem.*, 258:643–48), or a hybrid protein containing portions of LamB from both species, which behaves similarly (Roessner et al., 1987, *J. Mol. Biol.*, 195:963–66).

The conductance of single LamB pores is monitored during the addition of phage to the medium bathing the bilayer. An initial change in conductance upon phage binding will be followed by a drop in conductance as DNA enters the pore. Any sustained conductance fluctuations that follow are indicative of base pairs passing through the pore during injection. The fluctuations should be in the millisecond range, and the period of fluctuation will generally last for about 60 sec (the time required for injection). The conductance should then go up again to a level even higher than the original pre-phage state, since post-injection phage/porin complexes have been observed to allow molecules larger than the normal LamB exclusion limit to pass through (Roessner et al., 1986, *J. Biol. Chem.*, 261:386–90).

Asymmetrically modified DNA produced by annealing modified and unmodified complementary strands or by

custom primed DNA synthesis, can be ligated to lambda vector DNA and packaged in vitro. Modified DNA that is packaged efficiently and can be injected into bacterial cells will be appropriate for the LamB sequencing system.

Example 2: Conductance Measurements of Pore-Polymerase Complexes

Alternatively, the pore makes use of a polymerase molecule to pass DNA over the pore's opening one base pair at a time. Nucleotide bases of DNA will affect ion flux through the pore as they are passed over it by the polymerase, and the corresponding conductance fluctuations can be detected by high resolution single-channel recording techniques. The polymerase is held in position at the pore's surface as part of a fusion protein with the pore (see FIG. 2).

Fusion proteins are constructed (e.g., LamB and T7 RNA polymerase) such that both pore and polymerase are functional. The permissive sites in LamB (or any other desired pore protein) that can accommodate insertion of polypeptide sequences without significantly disrupting pore properties are targeted for these fusions. Accommodation of an entire active protein has been demonstrated for *E. coli* membrane proteins (Boyd et al., 1987, *Proc. Natl. Acad. Sci. USA*, 84:8525–29; Ehrmann et al., 1990, *Proc. Natl. Acad. Sci. USA*, 87:7574–78; both hereby incorporated by reference). At least some of the LamB permissive sites (12 have been described) will support insertion of the polymerase. T7 RNA polymerase is best fused to the C-terminal end of LamB, since the polymerase is known to function in protein fusions with this orientation (Ostrander et al., 1990, *J. Bacteriol.*, 116:1436–46; hereby incorporated by reference).

Fusion constructions can be screened for LamB function on MacConkey agar plates containing maltooligosaccharides. This screen is sensitive enough to detect a range of partial LamB functions. Fusion proteins with even partially functional pores should have high conductance states in bilayer experiments prior to DNA addition. Purified pore polymerase fusion proteins can be assayed for T7 RNA polymerase activity or used directly to reconstitute liposomes in preparation for channel recordings.

FIG. 2 shows a schematic representation of a possible pore-polymerase fusion complex with template DNA. The orientation of the base pairs to the pore opening is likely to alternate through several potential angles, due to the polymerase having some freedom of movement. For this reason, nucleotide base modifications that take into account the specific properties of the pore will be helpful to observe consistent conductance effects. For instance, the selectivity of LamB for maltose and maltooligosaccharides is a promising area of nucleotide modification for this system, particularly since these molecules have a dramatic effect on LamB conductance in vitro (Benz et al., 1986, supra; Benz et al., 1987, supra; Dargent et al., 1987, supra).

The affinity of maltooligosaccharides for LamB increases in proportion to the number of glucose residues attached (up to five) (Benz et al., 1987, supra), thus, nucleotides attached to maltose or maltooligosaccharides are likely to block the pore more effectively than unsubstituted nucleotides. Furthermore, the number of glucose molecules attached to a substituted nucleotide may significantly influence the effect it has on LamB conductance. By substituting specific bases with modified nucleotides linked to a given number of oligosaccharide units, it should be possible to relate those substituted bases with predictable conductance states of the pore.

Example 3: Testing pore fusions

The above system can be tested with a short oligonucleotide containing the T7 RNA polymerase promoter attached to phage T4 DNA that is asymmetrically modified at C residues with oligoglucose chains. The oligonucleotide template is made with the modified bases on the displaced strand, because T7 RNA polymerase has been shown to function when nucleotide analogs are present on this strand (Nath et al., 1991, *Carcinocenesis*, 12:973–76). The conductance of pore-polymerase complexes is monitored while adding this template and ribonucleotide triphosphates (NTPs) to the system. By adding the NTPs sequentially, one can sample the polymerase at four positions relative to the start site. This provides conductance information pertaining to the modified cytosine closest to the promoter at several distances from (and orientations to) the pore. By adding all four NTPs, the effects of each modified base as it passes the pore can be sampled during RNA synthesis.

The conductance profiles for this modified fragment are compared with control fragments containing no modifications to correlate given conductance shifts with the modified nucleotides. Uniform lengths for the oligoglucosyl moieties on given bases can be obtained by HPLC purifying the modified dNTPs. Additional pore-polymerase geometries and defined glucose chain lengths on modified bases are also within the scope of the invention.

Example 4: Recording from Shigella LamB

We have made recordings from Shigella LamB channels. We modified the pipette bilayer technique of Sigworth et al. (supra) by adding purified Shigella LamB to the solution inside the patch pipette, rather than to the bath solution. This modification has provided more consistent pore activity. We estimate the single channel conductance of this pore to be 120 pS (recording in bilateral 0.83M KCl, +25 mV). This is similar to the conductance of LamB from *E. coli* (Benz et al., 1986, supra; Dargent et al., 1987, supra; Benz et al., **1987**, supra; Dargent et al., 1988, supra). Our equipment and technique have the potential for recording at very high resolution. To our knowledge, we have made the first patch-clamp recordings of LamB molecules and the first recordings of any kind for the Shigella protein.

We tested the effect of a maltooligosaccharide mixture on the conductance of Shigella LamB, since maltooligosaccharides are known to inhibit the conductance of *E. coli* LamB in planar lipid bilayer experiments (Dargent et al., 1987, supra; Benz et al., 1987, supra) and can be used for nucleotide modifications. Our recordings show that the conductance of *Shigella LamB* molecules is inhibited by the addition of maltooligosaccharides to the bath. In the same recordings, we have shown that the conductance of these pores increases as the concentration of maltooligosaccharide in the bath is reduced. This reversible inhibition of conductance by maltooligosaccharides is similar to that observed for the *E.* coli protein (Dargent et al., 1987, supra).

In our experiments testing the effects of lambda DNA injection on the conductance of LamB pores, the Shigella protein was chosen because lambda will spontaneously inject its DNA in vitro when bound to this receptor, as opposed to the *E. coli* receptor, which requires the presence of organic compounds (ethanol or chloroform) for lambda injection. LamB conductance was altered when lambda injected its DNA through the pore, and the conductance changes were detectable during a patch-clamp recording. We have obtained several classes of response when lambda is added to the bath during patch-clamp recordings of Shigella LamB, ranging from no response at all to nearly complete inhibition of conductance, to rapidly fluctuating conductance levels. We observed that under the bath conditions

used for patch-clamp recording, our preparation of Shigella LamB is routinely capable on inducing lambda DNA injection in vitro.

Multiple pores in the patch membrane at one time make it difficult to interpret the lambda response, and it is important to obtain single pores in the patch membrane.

Advantages of using phage lambda to orient DNA to pore:

1) The system is simple to set up.
2) Lambda injection is efficient and fast.
3) Lambda vectors are used extensively to construct genomic and cDNA libraries, thus there is a tremendous resource of potential sequence information readily available for direct application of this technique.
4) The average insert size for cosmid libraries is about 45 kb; this sets the average "read" size of contiguous DNA sequence. While this is less than the potential read size for a functional pore-polymerase complex (see below), it is still about 100 times the average read size for most conventional sequencing.

Ion flux can take place through phage receptor pores that contain phage DNA. T5 provides an alterative phage system (as do T3, T4 and P1, all of which have efficient in vitro packaging systems).

The membrane spanning length of LamB pores is estimated to be 30 Å (Benz et al., 1987, supra). Thus, at any given time during DNA injection, 8–9 base pairs of DNA are present in the pore. For the effects of single base pairs on pore current to be more easily measurable, it is preferable to have a region of the pore that is rate limiting for ion flux past one or two base pairs. Such a region may take the form of an "eyelet" structure as seen in the *R. capsulatus* porin (Weiss et al., supra), where steric interactions are limiting, or, depending on the exact amino acids involved, hydration, electrostatic, as well as steric interactions may produce a rate limiting site. Alternatively, since it is likely that the bacteriophage tail fiber widens the LamB pore upon phage attachment, and it is possible that the phage DNA passes through the fiber during injection, the bottle neck for ion flow may be at some position along the inside of the tail fiber.

Advantages of the pore-polymerase system for orienting DNA to the pore

1) This system offers some flexibility in orienting DNA to the pore's opening, and thus provides the mechanism to optimize this orientation.
2) The average size of contiguous sequence obtainable by this technique is very large; it is limited by the processivity of T7 RNA polymerase since there is no obvious template size restriction. T7 RNA polymerase is highly processive in vitro (Golomb et al., *Proc. Natl. Acad. Sci. USA*, 71:760–64; Niles et al., supra; Oakley et al., 1975, *Biochemistry*, 14:4684–91).
3) The rate of sequencing with this system is also very high, limited only by the rate of polymerase activity when fused to the pore. The rate of T7 RNA polymerase is ≈300 bases/sec (Martin et al., 1987, *Biochemistry*, 26:2690–96). This provides an estimate of the sequencing rate for this system.
4) In principal, any source of DNA can be used as template for this system, provided it contains the T7 RNA polymerase promoter. This includes high molecular weight DNA from tissue samples which is ligated to a T7 promoter oligonucleotide.

The fusion proteins constructed must have at least partial pore function to ensure a high single-channel conductance, so that there will be "room" for lower conductance states when DNA is added.

The polymerase portion of the fusion can be considered an external protein domain of the pore. This polymerase domain must demonstrate activity when the complex is inserted into bilayers. T7 RNA polymerase is known to function when fused to the C-terminus of other proteins (Ostrander et al., supra). Thus, this orientation to LamB should be productive, provided the C-terminal amino acids of LamB remain intact (Boulain et al., supra). The permissive sites of LamB are particularly attractive for polymerase insertion, since most of these sites map to regions predicted to form extramembranous loops (Charbit et al., 1991, supra), and several proteins have been shown to function when inserted at such sites in other membrane proteins (Boyd et al., supra; Ehrmann et al., 1990, supra). Our cloning scheme is designed to allow variation in the length of peptide linkers at either end of the polymerase insertion site.

The LamB protein forms a trimer, so expressing the pore-polymerase construct in cells that otherwise lack LamB protein will result in a pore with three polymerases. Such complexes may be unstable or nonfunctional. This problem can be avoided by producing heterotrimers between normal LamB monomers and pore-polymerase monomers. Functional LamB heterotrimers between normal and mutant forms have been observed (Ferenci et al., 1989, *J. Bacteriol.*, 171:855–61; hereby incorporated by reference).

The orientation of the polymerase to the pore's opening must be such that during polymerization, nucleotide bases are positioned close enough to affect ion flux through the pore. The mouth of the pore is small compared to the polymerase or DNA, thus, polymerase activity at the pore's surface will affect ion access. Some configurations, however, may be better suited for sequencing, in that they may provide more consistent conductance readings.

Several insertion target sites for LamB have been described, although random insertion is an option with our screening approach. It is also possible to express domains of the polymerase at different sites in LamB, such that they fold together on the pore's surface into an optimal configuration. Alternatively, by using two polymerase molecules per trimer, the template might be anchored over the pore in a more optimal position. Information provided by the progressing crystallographic investigations of these two proteins (Sousa et al., 1989, *Proteins: Struct. Funct. Genet.*, 5:266–70; Stauffer et al., 1990, *J. Mol. Biol.*, 211:297–99) may be useful to help design the geometry of the fusion to suit the needs of the project. We are not limited to these particular proteins, however, since in principle any channel molecule and any processive nucleic acid translocation molecule could potentially suffice. An example is the conjugation process in *E. coli*, where genome-sized (single-stranded) DNA is transported through the F-pilus at a rate of about 780 bp/sec (Rees et al., 1989, *J. Bacteriol.*, 171:3152–57; Harrington et al., 1990, *J. Bacteriol.*, 172:7263–64). This process can be monitored in situ with the patch-clamp technique.

The oligoglucosyl chains attached to modified bases are expected to have considerable rotational freedom. If each LamB monomer has a binding site for maltooligosaccharides, which is the present model (Ferenci et al., supra), then conductance readings from one base pair may be obscured by the side chains from adjacent base pairs interacting with other pores in the trimer. This problem can be overcome by using heterotrimers containing LamB mutant monomers that have low maltose binding affinity complexed with wild type monomers (or a hyperbinding mutant) in a ratio of 2:1. It has been shown that low maltose affinity monomers do not prevent maltodextrin transport when present in heterotrimers (Ferenci et al., supra).

2) The alpha-hemolysin pore forming protein

Discussion and examples of the invention using the bacterial pore-forming protein α-hemolysin toxin (α-toxin or α-hemolysin) are below. This system operates as shown in FIG. 1; nucleic acid polymers are threaded through the atoxin pore as shown, and the monomeric charges and physical obstruction alter ionic conductance through the pore. Because the purine and pyrimidine bases in the polynucleotide have differing molecular sizes and chemical properties, a specific ionic current will flow as each nucleotide enters and passes through the channel, thus electrosensing the monomer sequence in the linear polymer.

Example 5: Bilayer recordings from α-hemolysin

For these experiments, the bacterial pore-forming protein from *S. aureus*, α-hemolysin, spontaneously embeds in lipid bilayers to produce a large, heptameric, currentconducting channel. α-hemolysin forms a robust channel which has the appropriate diameter to admit a single stranded DNA polymer. Furthermore, it can remain open for indefinite time periods when subjected to a continuous voltage gradient. Diphytanoyl phosphatidylcholine was used to form lipid bilayer membranes across 0.2 mm holes in a Teflon film separating two compartments containing buffer solution of the following composition: 1M NaCl, 10 mM Tris, pH 7.4 (Montal et al., 1972, *PNAS*, 69:3561). In initial, multichannel experiments, α-hemolysin was added to the cis side of the bilayer and approximately 10 channels were allowed to incorporate into the bilayer before excess α-hemolysin was removed. Voltage applied across the bilayer was then varied from 0 mV to 140 mV. Under the buffer conditions used, the channels were continuously open before addition of polynucleotide. After addition of poly A to the cis chamber, the channels began to exhibit transient blockades at potentials greater than 100 mV. Similar effects were seen with poly C and poly U polymer additions. Significantly, the blockades only occurred when the voltage was applied in the direction expected to produce electrophoretic movements of a poly-anion like RNA from the cis to the trans side of the channel, i.e., only when the trans side was positive.

Further experiments with single channels demonstrated many well-resolved individual channel blockades in the presence of poly A, poly C, or poly U molecules (for example, see FIG. 4). Qualitatively, the number of transient blockades was proportional to the concentration of polynucleotide. Typical current blockades exhibited 85–90% reductions of current amplitude and lasted up to several milliseconds. Because the polynucleotide preparations used in these experiments contain a range of molecular weights, we could not quantitatively relate blockade duration to polynucleotide length. But qualitatively, average blockade duration was greater when using solutions containing longer RNA polymers (MW 140 kb– 1700kb) than when using solutions containing shorter polymers (MW 77 kb–160 kb). occasionally, long-lived blockades of several seconds or more were observed. These often cleared spontaneously, but could always be cleared by briefly reversing the voltage polarity. Again, there was virtually no effect on the magnitude of channel conductance when the trans side was negative. To verify that the polynucleotides were producing the long-lived blockades, RNAse was added to the RNA in the cis chamber to gradually hydrolyze it. When RNAse was added to poly U in the cis chamber while transient blockades were being observed, the duration of the transient blockades, but not their amplitude, gradually decreased over a period of several minutes, eventually becoming too short to be detectable.

From these experiments, it is apparent that polynucleotides are not simply binding to the channel and causing it to partially close, because if that were true, the current blockades would not depend on the polarity of the voltage gradient. Our interpretation is that ionic current through a channel can be modulated by passage of single polymer strands. This interpretation is supported by the fact that ribonuclease decreases the duration but not the amplitude of the current blockades. It is also consistent with our observation that circular single-stranded molecules appear to produce virtually no blockades and that double-stranded molecules with single stranded ends produce only indefinitely long-lived blockades.

Example 6: Relationship between polymer length and channel blockade duration

To determine the relation between chain length and duration of the current blockade, we used samples of synthetic short (~ 320 nt) and long (~ 1,100 nt) poly U acid that we size-selected by gel electrophoresis. These experiments have been repeated, with several independently purified polyuridylic acid samples which gave consistent results. Using polymers whose chain length centered around 320 nt, about 35% of the recorded current blockades had lifetimes of around 2.1 msec at 120 mV (FIG. 5A), and around 1.7 msec at 140 mV (data not shown), with the remaining signals having short lifetimes of < 1 msec. We presume that the short duration blockades represent polymers that interact with the channel (e.g., loops of polymer that come to lie on the channel aperture, without fully entering and traversing the channel). We attribute the clear peak of blockades centered around 2.1 msec or 1.7 msec (depending on applied voltage) to polymers that have traversed the channel, because: 1) Based on the consistency of the peak position from run to run, the shift in peak position from 2.1 msec at 120 mV to 1.7 msec at 140 mV is statistically significant and hard to explain by any model other than a polymer being threaded through the channel; 2) When RNA that had not been size-selected (e.g., RNA containing the full range of polymer lengths from 250 nt to 1600 nt) was used, we detected the corresponding full range of blockade durations rather than durations that exhibited a "narrow" peak (as seen in FIGS. 5A and 5B); and 3) Experiments with the ca 1,100 nt polymers have shown a peak centered around 5.8–61 msec at 140 mV (FIG. 5B). If one assumes a linear relationship between polymer size and blockade duration, it can be seen that 1100 nt/320 nt=3.4 and that 3.4×1.7 msec =5.8 msec, lending credibility to the accuracy of the methods of the invention for measuring polymer length by measuring signal duration.

II. Polymer Analysis By Detection of Monomers At An Interface

A different embodiment of the invention includes a method of characterizing a linear polymer using 1) an interface, generally created by two immiscible liquids, and 2) a monitoring device such as a force transducer or deflection gauge (e.g., using light) to monitor each monomer of the polymer as it passes across the interface. This embodiment of the sequencing invention is encompassed by pulling a single molecule through the interface formed by two immiscible liquids by either mechanical or electrophoretic means. The force required to pull each successive monomer in the polymer through the interface can be measured, or it may be desirable to monitor physical deflections of the interface or other modifications/interactions of the interface by the monomers to register each successive monomer's move through the interface.

Multiple polymer strands have commonly been pulled through a liquid-air interface as a means of separating and purifying polymers from their surrounding liquor. We propose that polymer chains can be similarly pulled through the

interface formed by two immiscible liquids. Both the atomic force microscope and optical tweezers are now routinely used in liquid environments to measure atomic and molecular scale forces and movements. A suitably fine probe attached to a force transducer such as those used in atomic force microscopy or in optical tweezers would advance in Angstrom size steps, and have been used to pull double stranded DNA through DNA solutions. If the force required to pull the different monomers of a polymer through the interface differs from one monomer to another, then measuring the force required to pull each successive monomer through the interface will provide a direct determination of the sequence of monomers in the polymer.

The force required to move a chemical group from one phase to a second, immiscible phase is related to its partition coefficient. For the pertinent polymers that could be sequence by this technique, the partition coefficient for its monomers would differ from each other. For example, the logarithm of the partition coefficients of the DNA monomers adenine, thymine, guanine, and cytosine in a butanol:water system are, respectively, 2.44, 1.12, 0.45, and −0.68.

A chemical embodiment of this aspect of the invention could be a butanol:water interface, with the underlying aqueous phase containing the polymer to be sequenced. If a fine ceramic, plastic, or metallic probe bearing a suitable charge or chemical group at its tip (e.g., to attract DNA, a positively charged tip; to attract mRNA, oligo dT moieties) is driven through the overlying butanol into the underlying aqueous phase, polymer will stick to the probe tip and be pulled through the interface as the tip is withdrawn from the aqueous phase into the butanol phase. Although the initial strand of material that is pulled through the interface may contain multiple individual polymers, inevitably one single polymer strand will be longer than the others or will have stuck to the probe tip in such a fashion that it will be pulled last, and singly, through the interface. Refinements to the probe tip to increase the likelihood of selecting only one polymer may include decreasing the charge or number of chemical moieties.

Because the energy to pull each of the different monomers of a single polymer chain through the interface will reflect the properties of the monomer, recording the force required to pull a single stranded DNA molecule, for instance, through an interface while maintaining a constant slow movement will in effect record the sequence of the polymer.

An alternative method of measuring the transit of monomers from one phase to the other may be the use of optical means as are known in the art to detect the deflection of the interface caused by each monomer. Due to varying physical properties of the monomer (e.g., size, mass, volume), light may be scattered off the interface in predictable ways for each monomer. For instance, directing a laser at the interface and observing the optical deflection using a bi-cell detector may identify individual monomers by their characteristic deflections of light. Alternatively, pulsed laser techniques may be used, with pulses on the order of 10^{-9} to 10^{-12} seconds directed at the interface and recorded using a time dependent detector.

Other details of this second aspect of the invention are set forth in applicable portions of the detailed description of the first aspect of the invention.

Other embodiments are within the following claims.

What is claimed is:

1. A method for evaluating a single polymer molecule or multiple polymer molecules one at a time, the polymer molecule(s) including linearly connected monomer residues, the method comprising:

a) providing two separate pools of liquid containing medium and an interface between the two pools,

b) providing candidate polymer molecules in one of the pools, and

c) making interface-dependent measurements over time as individual monomer residues of a single polymer molecule interact sequentially with the interface, yielding data suitable to determine a monomer-dependent characteristic of the polymer molecule.

2. The method of claim **1** in which several single polymer molecules are evaluated in rapid succession.

3. The method of claim **1** wherein:

a) the pools comprise electrically conductive medium;

b) the interface between the pools comprises an impermeable barrier containing an ion-permeable passage; and

c) making the interface-dependent measurements comprises measuring conductance of the ion-permeable passage in the presence of a voltage differential between the pools, while monomers of the single polymer molecule interact sequentially with the passage.

4. The method of claim **3** wherein the voltage differential is externally applied across the pools.

5. The method of claim **3** wherein the voltage differential between the pools results at least in part from at least one difference in the composition of one pool in comparison to the composition of the other pool.

6. The method of claim **3** wherein the measurement is made as the polymer traverses the passage, crossing from one of the conducting pools to the other.

7. The method of claim **3**, or claim **7**, wherein a template-dependent polymer replicating catalyst is physically linked to the ion-permeable passage, and at least one of the conducting pools contains monomers suitable to be catalytically linked in the presence of the catalyst and the single polymer molecule.

8. The method of claim **3** wherein the amplitude or duration of individual conductance changes across the passage are indicative of the sequential identity of monomers of the polymer.

9. The method of claim **3** wherein the number of changes in conductance states is indicative of the number of monomers in the polymer.

10. The method of claim **3** wherein the duration of the polymer interaction with the interface is indicative of the number of monomers in the polymer.

11. The method of claim **1**, claim **3**, claim **10**, or claim **10** wherein multiple molecules of a heterogeneous mixture of individual polymer molecules are characterized to provide a size distribution of polymers in the mixture.

12. The method of claim **3** wherein the passage is a bacterial porin channel.

13. The method of claim **3** wherein the passage is a voltage-sensitive channel.

14. The method of claim **13** wherein the voltage-sensitive channel does not inactivate.

15. The method of claim **3** wherein the interface comprises a bacteriophage receptor segment and the polymer comprises a specific ligand for the bacteriophage receptor.

* * * * *