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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 April 2009 (09.04.2009)

PCT

(10) International Publication Number
WO 2009/044170 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/68* (2006.01)

(21) International Application Number:
PCT/GB2008/003372

(22) International Filing Date: 6 October 2008 (06.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0719531.6 5 October 2007 (05.10.2007) GB

(71) Applicant *(for all designated States except US)*: **ISIS INNOVATION LIMITED** [GB/GB]; Ewert House, Ewert Place, Summertown, Oxford 0X2 7SG (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BAYLEY, John Hagan** [GB/GB]; University of Oxford, Chemical Biology Department - Bayley Group, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA (GB). **WU, Haichen** [CN/GB]; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA (GB). **MAGLIA, Giovanni** [IT/GB]; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA (GB). **ASTIER, Yann** [FR/PT]; Universidade Nova de Lisboa, Instituto de Technologia, Quimica e Biologica, Aptd. 12 Oeiras, 2781-901 (PT).

(74) Agents: **CHAPMAN, Lee Phillip** et al.; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

(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, BR, BW, BY, BZ, CA, CH, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(54) Title: MOLECULAR ADAPTORS

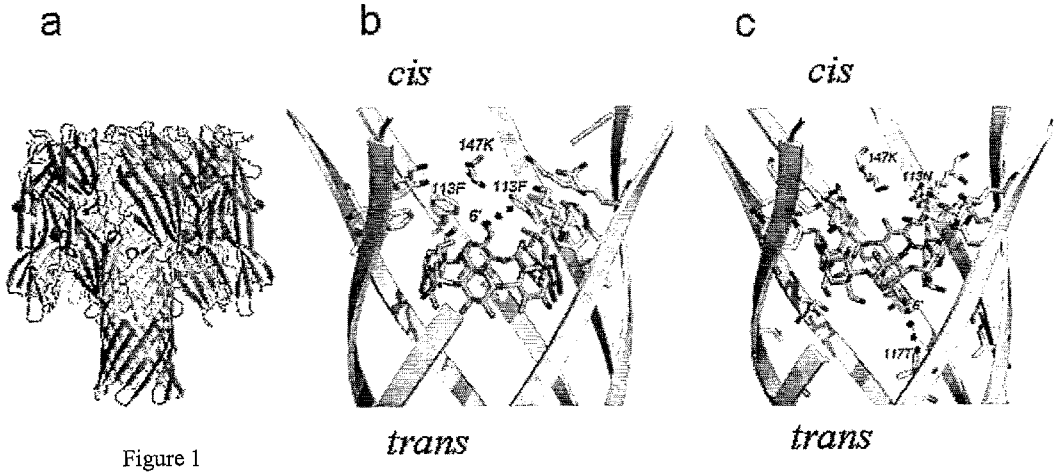


Figure 1

(57) Abstract: The invention relates to transmembrane protein pore for use in detecting a analyte in a sample. The pore comprises a molecular adaptor that facilitates an interaction between the pore and the analyte. The adaptor is covalently attached to the pore in an orientation that allows the analyte to be detected using the pore.

WO 2009/044170 A1

MOLECULAR ADAPTORS

Field of the invention

The invention relates to transmembrane protein pores that are useful for detecting analytes by stochastic sensing. A molecular adaptor is covalently attached to the pore in an orientation that allows detection of the analyte.

Background of the invention

Stochastic detection is an approach to sensing that relies on the observation of individual binding events between analyte molecules and a receptor. Stochastic sensors can be created by placing a single pore of nanometer dimensions in an insulating membrane and measuring voltage-driven ionic transport through the pore in the presence of analyte molecules. The frequency of occurrence of fluctuations in the current reveals the concentration of an analyte that binds within the pore. The identity of an analyte is revealed through its distinctive current signature, notably the duration and extent of current block (Braha, O., Walker, B., Cheley, S., Kasianowicz, J. J., Song, L., Gouaux, J. E., and Bayley, H. (1997) *Chem.Biol.* **4**, 497-505; and Bayley, H., and Cremer, P. S. (2001) *Nature* **413**, 226-230).

Engineered versions of the bacterial pore-forming toxin α-hemolysin (αHL) have been used for stochastic sensing of many classes of molecules (Bayley, H., and Cremer, P. S. (2001) *Nature* **413**, 226-230; Shin, S.-H., Luchian, T., Cheley, S., Braha, O., and Bayley, H. (2002) *Angew.Chem.Int.Ed.* **41**, 3707-3709; and Guan, X., Gu, L.-Q., Cheley, S., Braha, O., and Bayley, H. (2005) *ChemBioChem* **6**, 1875-1881). In the course of these studies, it was found that attempts to engineer αHL to bind small organic analytes directly can prove taxing, with rare examples of success (Guan, X., Gu, L.-Q., Cheley, S., Braha, O., and Bayley, H. (2005) *ChemBioChem* **6**, 1875-1881). Fortunately, a different strategy was discovered, which utilized non-covalently attached molecular adaptors, notably cyclodextrins (Gu, L.-Q., Braha, O., Conlan, S., Cheley, S., and Bayley, H. (1999) *Nature* **398**, 686-690), but also cyclic peptides (Sanchez-Quesada, J., Ghadiri, M. R., Bayley, H., and Braha, O. (2000) *J.Am.Chem.Soc.* **122**, 11758-11766) and cucurbiturils (Braha, O., Webb, J., Gu, L.-Q., Kim, K., and Bayley, H. (2005) *ChemPhysChem* **6**, 889-892). Cyclodextrins

become transiently lodged in the αHL pore and produce a substantial but incomplete channel block. Organic analytes, which bind within the hydrophobic interiors of cyclodextrins, augment this block allowing analyte detection (Gu, L.-Q., Braha, O., Conlan, S., Cheley, S., and Bayley, H. (1999) *Nature* **398**, 686-690).

Summary of the invention

The inventors have surprisingly demonstrated that a molecular adaptor can be covalently attached to a transmembrane protein pore in a specific orientation. The inventors have also surprisingly demonstrated that a molecular adaptor can be covalently attached to a transmembrane pore in an orientation that allows an analyte to affect the current flowing through the pore in a manner specific for that analyte. A molecular adaptor can therefore be covalently attached to the pore in an orientation that allows the pore to be used to detect the analyte via stochastic sensing. The pore is preferably modified to facilitate the orientation of the molecular adaptor.

The fixed nature of the molecular adaptor means that a distinctive current flows through the pore whenever the analyte interacts with the pore. As a result, the transmembrane protein pores of the invention are useful tools for stochastic sensing, especially for detecting nucleotides or sequencing nucleic acids.

Accordingly, the invention provides a transmembrane protein pore for use in detecting an analyte in a sample, which comprises a molecular adaptor that facilitates an interaction between the pore and the analyte, wherein the adaptor is covalently attached to the pore in an orientation that allows the analyte to be detected using the pore.

The invention also provides:

- a method of producing a transmembrane protein pore of the invention, comprising:

(a) covalently attaching to a transmembrane protein pore a molecular adaptor that facilitates an interaction between the pore and an analyte; and

(b) determining whether or not the adaptor is attached to the pore in an orientation that allows the analyte to be detected using the pore;

- a method of determining presence or absence of an analyte, comprising:

(a) contacting the analyte with a transmembrane protein pore of the invention so that the analyte interacts with the pore; and

(b) measuring the current passing through the pore during the interaction and thereby determining the presence or absence of the analyte;

- a method of identifying an individual nucleotide, comprising:

(a) contacting the nucleotide with a transmembrane protein pore of the invention so that the nucleotide interacts with the pore; and

(b) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide;

- a method of sequencing a target nucleic acid sequence, comprising:

(a) digesting an individual nucleotide from one end of the target sequence using a processive exonuclease;

(b) contacting the nucleotide with a transmembrane protein pore of the invention so that the nucleotide interacts with the adaptor;

(c) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide; and

(d) repeating steps (a) to (c) at the same end of the nucleic acid sequence and thereby determining the sequence of the nucleic acid; and

- a kit for sequencing a nucleic acid, comprising a transmembrane protein pore according to the invention and a processive exonuclease.

Description of the Figures

Figure 1 shows X-ray structures of $(M113F\text{-}RL2)_7$•βCD and $(M113N\text{-}RL2)_7$•βCD generated from the pdb files (Montoya and Gouaux, unpublished). a) Side view of the αHL pore with amino acids 108 to 120 and 138 to 150 highlighted in yellow. b) $(M113F)_7$•βCD. Amino acids 108 to 120 of αHL are depicted in yellow and βCD is shown in blue sticks. The side chains of Phe-113 and Lys-147 are also shown in stick form. The distance between the O atom of a primary hydroxyl of βCD and the center of the nearest phenyl ring of Phe-113 is about 6.5 Å (dotted line). c) $(M113N\text{-}RL2)_7$•βCD. Amino acids 108 to 120 of αHL are depicted in yellow and βCD is shown in blue sticks. The side chains of Asn-113, Lys-147 and

Thr-117 are also shown in stick form. The distance between the O atom of a primary hydroxyl of βCD and the nearest O atom of Thr-117 is about 6.0 Å (dotted line).

Figure 2 shows the chemical structure and a stick representation of βCD-PDP. The molecule was drawn in ChemDraw 3D and then rendered in PyMol.

Figure 3 shows a comparison of the properties of the non-covalent complex αHL (M113F-RL2)$_7$•βCD and the covalent adduct αHL (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD. Each experiment was repeated at least 6 times. a) Cartoon representation of amino acids 108 to 120 of the mutant αHL pore (M113F-RL2)$_7$. Phe-113 residues are represented in stick form. Single-channel current trace from a (M113F-RL2)$_7$ pore in 25 mM Tris·HCl, pH 8.0, 1 M KCl, +100mV. b) Cartoon representation of αHL (M113F-RL2)$_7$ with βCD bound as revealed by the X-ray structure of (M113F-RL2)$_7$•βCD. The arrows indicate that βCD is free to move in and out of the pore. The current trace shows an initially unoccupied pore followed by the binding of βCD and its subsequent dissociation (mean τ_{off} = 316 ± 62 ms). c) Cartoon representation of amino acids 108 to 120 of αHL (M113F-RL2)$_6$(M113C-D8RL2)$_1$. Phe-113 residues are represented in stick form and Cys-113 is colored in brown and also represented in sticks. An enlarged view of (c) can be found in Figure 4. The arrow indicates that when βCD-PDP enters the pore the primary hydroxyls interact with the side chains of Phe-113, and Cys-113 reacts with the pyridyl disulfide of βCD-PDP to form a disulfide bond. The current trace shows the current amplitude changes during the reaction. The transient state with 85% current block is marked (*). d) Cleavage of (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD with dithiothreitol (DTT). The arrow indicates that the cyclodextrin derivative detaches from position 113 after cleavage of the disulfide bond with DTT. The current trace shows the current amplitude changes during cleavage. The transient state with 98% current block is marked (*). e) Single-channel I-V curves for (M113F-RL2)$_7$•βCD (τ) and αHL (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD (■.). Conditions: 25 mM Tris·HCl, pH 8.0, 1 M KCl.

Figure 4 shows an enlarged view of Figure 3c. Cartoon representation of amino acids 108 to 120 of αHL (M113F-RL2)$_6$(M113C-D8RL2)$_1$. Phe-113 residues are represented in stick form and Cys-113 is colored in brown and also represented in

sticks.

Figure 5 shows a comparison of the properties of the non-covalent complex αHL (M113N-RL2)$_7$-βCD and the covalent adduct αHL (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD. Each experiment was repeated at least 6 times. a) Cartoon representation of amino acids 108 to 120 of the mutant αHL pore (M113N-RL2)$_7$. Asn-113 residues are represented in stick form. Single-channel current trace from a (M113N-RL2)$_7$ pore in 25 mM Tris·HCl, pH 8.0, 1 M KCl, -100mV. b) Cartoon representation of αHL (M113N-RL2)$_7$ with βCD bound as revealed by the X-ray structure of (M113N-RL2)$_7$•βCD. The arrows indicate that βCD is free to move in and out of the pore. The current trace shows an initially unoccupied pore followed by the binding of βCD and its subsequent dissociation (mean $\tau_{off} = 10.7 \pm 1.5$ s). c) Cartoon representation of amino acids 108 to 120 of αHL (M113N-RL2)$_6$(T117C-D8RL3)$_1$. Asn-113 residues are represented in stick form and Cys-117 is colored in brown and also represented in sticks. An enlarged view of Figure 5c can be found in Figure 6. The arrow indicates that when βCD-PDP enters the pore the secondary hydroxyls interact with the side chains of Asn-113, and Cys-117 can react with the pyridyl disulfide of βCD-PDP to form a disulfide bond. The current trace shows the current amplitude change when the reaction takes place between Cys-117 and the pyridyl disulfide of βCD-PDP. The transient state with 80% current block is marked (*). d) Cleavage of (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD with DTT. The arrow indicates that the cyclodextrin derivative detaches from position 117 after cleavage of the disulfide bond. The current trace shows the current amplitude changes during cleavage. The transient state with 90% current block is marked (*). e) Single-channel I-V curves for (M113N-RL2)$_7$•βCD (τ) and αHL (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD (■). Conditions: 25 mM Tris.HCl, pH 8.0, 1 M KCl.

Figure 6 shows an enlarged view of Figure 5c. Cartoon representation of amino acids 108 to 120 of αHL (M113N-RL2)$_6$(T117C-D8RL3)$_1$. Asn-113 residues are represented in stick form and Cys-117 is colored in brown and also represented in sticks.

Figure 7 shows the detection of 2-adamantanamine with αHL pores containing a covalently-attached βCD. a) Current trace from a single (M113F-

RL2)$_6$(M113C-D8RL2)$_1$-βCD pore in 25 mM Tris·HCl, pH 8.0, 1 M KCl, at +100mV. b) Current trace in the presence of 30 μM 2-adamantanamine in the *cis* chamber. c) Current trace with 30 μM 2-adamantanamine in the *trans* chamber. d) Current trace from a single (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD pore in 25 mM Tris·HCl, pH 8.0, 1 M KCl, at +100mV. e) Current trace with 30 μM 2-adamantanamine in the *cis* chamber. f) Current trace with 30 μM 2-adamantanamine in the *trans* chamber.

Figure 8 shows long binding events of *cis*-lithocholic acid with (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD. a) αHL with attached molecular adaptor (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD and structure of *cis*-lithocholic acid. b) Characteristic binding event of *cis*-lithocholic acid (50 μM) with (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD in 25 mM Tris·HCl, pH 8.0, 1 M KCl, at -100mV. Mean $\tau_{off} = 8.6 \pm 7.8$ s (number of events = 60). Level 1: (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD; level 2: (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD•lithocholate, 92% block; level 3: (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD•lithocholate, 98% block.

Figure 9 shows that both (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD and (M113F-RL2)$_7$ in the presence of 50 μM βCD show closures in the absence of *cis*-lithocholate that might be confused with *cis*-lithocholate binding, but the events are shorter. a) (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD in the absence of *cis*-lithocholic acid under the same conditions as in Figure 6. Level 1: (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD; level 2: spontaneous gating of the pore. The events varied in duration from 30 to 300 ms. b) (M113F-RL2)$_7$ in the presence of 50 μM βCD in 25 mM Tris·HCl, pH 8.0, 1 M KCl, at -100mV. $\tau_{off} = 0.68 \pm 0.12$ s. Level 1: (M113F-RL2)$_7$; level 2: (M113F-RL2)$_7$•βCD; level 3: spontaneous gating of the pore complex(M113F-RL2)$_7$•βCD.

Figure 10 shows the variation of the single-channel current with temperature for (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD in 25 mM Tris·HCl, pH 8.0, 1 M KCl at +100 mV. The values shown in the plot are the average values for three different experiments. The single-channel currents depend linearly on the temperature. I (pA) = 14.307 + 0.5402T(°C).

Figure 11 shows the ion selectivity of (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD

and $(M113N)_6(T117C\text{-}D8RL3)_1$-βCD. a) I-V curves for $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ ( ) and $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD (■) based on recordings made in 25 mM Tris·HCl, pH 8.0 with *cis*: 200 mM KCl; *trans*: 1000 mM KCl (n = 3). Reversal potentials (*Vr*) are marked by arrows. b) I-V curves for $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ ( ) and $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD (■) based on recordings made in 25 mM Tris·HCl, pH 8.0 with *cis*: 1000 mM KCl; *trans*: 200 mM KCl (n = 3). Reversal potentials (*Vr*) are marked by arrows.

Figure 12 shows the sequence encoding αHL-D8RL3 (including the octa-aspartate tail). This sequence is shown in SEQ ID NO: 18. The sequence encoding αHL-D8RL3 (SEQ ID NO: 18) is identical to nucleotides 13 to 918 of SEQ ID NO: 15, which encode T117C-D8RL3 used in the Example, except that SEQ ID NO: 15 has TGC (rather than ACG) at residues 361 to 363.

Description of the Sequences

SEQ ID NO: 1 shows the polynucleotide sequence that encodes one subunit of wild-type α-hemolysin.

SEQ ID NO: 2 shows the amino acid sequence of one subunit of wild-type α-hemolysin.

SEQ ID NO: 3 shows the polynucleotide sequence that encodes one subunit of α-hemolysin M113H-RL2.

SEQ ID NO: 4 shows the amino acid sequence of one subunit of α-hemolysin M113H-RL2.

SEQ ID NO: 5 shows the polynucleotide sequence that encodes one subunit of α-hemolysin M113K-RL2.

SEQ ID NO: 6 shows the amino acid sequence of one subunit of α-hemolysin M113K-RL2.

SEQ ID NO: 7 shows the polynucleotide sequence that encodes one subunit of α-hemolysin M113R-RL2.

SEQ ID NO: 8 shows the amino acid sequence of one subunit of α-hemolysin M113R-RL2.

SEQ ID NO: 9 shows the polynucleotide sequence that encodes one subunit of α-hemolysin M113F-RL2 used in the Example.

SEQ ID NO: 10 shows the amino acid sequence of one subunit of α-hemolysin M113F-RL2 used in the Example.

SEQ ID NO: 11 shows the polynucleotide sequence that encodes one subunit of α-hemolysin M113N-RL2 used in the Example.

SEQ ID NO: 12 shows the amino acid sequence of one subunit of α-hemolysin M113N-RL2 used in the Example.

SEQ ID NO: 13 shows the polynucleotide sequence that encodes one subunit of α-hemolysin M113C-D8RL2 used in the Example.

SEQ ID NO: 14 shows the amino acid sequence of one subunit of α-hemolysin M113C-D8RL2 used in the Example.

SEQ ID NO: 15 shows the amino acid sequence of one subunit of α-hemolysin T117C-D8RL3 used in the Example.

SEQ ID NO: 16 shows the polynucleotide sequence that encodes one subunit of α-hemolysin T117C-D8RL3 used in the Example.

SEQ ID NO: 17 shows the amino acid sequence of lambda exonuclease. The sequence is one of three identical subunits that assemble into a trimer.

SEQ ID NO: 18 shows the sequence encoding αHL-D8RL3 (the sequence in Figure 12).

Detailed description of the invention

It is to be understood that different applications of the disclosed products and methods may be tailored to the specific needs in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

In addition as used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "an analyte" includes "analytes", reference to "a transmembrane protein pore" includes two or more such pores, reference to "a molecular adaptor" includes two or more such adaptors, and the like.

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

Transmembrane protein pores

The present invention relates to a transmembrane protein pore for use in detecting an analyte. The transmembrane protein pore comprises a molecular adaptor that facilitates an interaction with the analyte. The adaptor is covalently attached to the pore in a specific orientation. The adaptor is oriented such that, during the interaction between the analyte and the pore, the analyte affects the current flowing through the pore in a manner specific for that analyte. The adaptor is therefore covalently attached to the pore in an orientation that allows the pore to be used to detect the analyte via stochastic sensing.

The transmembrane protein pore of the invention is a useful tool for stochastic sensing. The fixed nature of the molecular adaptor means that the signal obtained from the pore is entirely dependent on the presence of the analyte in the barrel or channel of the pore and is not affected by dissociation of the adaptor from the pore. In other words, the fixed nature of the adaptor means that a distinctive current will flow through the pore whenever an analyte interacts with the pore. This results in a more sensitive system that allows: (1) the detection of rare interaction events, such as when the concentration of the analyte is low; (2) the detection of common interaction events, such as when the analyte concentration is high; (3) the detection of long-lived interaction events, such as when the affinity of the analyte for the adaptor is high; and (4) the measurement of the dwell time of the analyte, so that the affinity of the analyte for the adaptor can be measured. Most importantly, the lack of any interruption in the signal means that every analyte that enters the pore is detected. Such an efficient detection system facilitates the sequencing of nucleic acids. This is discussed further below.

The transmembrane protein pore of the invention has additional benefits. As described in the Example, the transmembrane protein pore of the invention is capable of functioning at elevated temperatures, such as up to 55°C or up to 100°C. This means that the pore is suitable for detecting analytes under extreme conditions.

The fixed nature of the molecular adaptor means that it is not free to interact with the analyte independently of the pore. This reduces interference and means that a more accurate measurement of the concentration of the analyte can be obtained.

As described in the Example, the attachment of the molecular adaptor to the pore permanently alters its ion selectivity. This allows the production of a transmembrane protein pore having a particular ion selectivity.

Finally, the fixed nature of the molecular adaptor means that the pore and adaptor can be stored together, thereby allowing the production of a ready-to-use sensor.

Transmembrane protein pore

A transmembrane protein pore is a polypeptide that permits ions to flow from one side of the membrane to the other along an applied potential. The pore preferably permits the analyte to flow from one side of the membrane to the other along the applied potential.

The pore is typically an oligomer. The pore is preferably made up of several repeating subunits, such as 6, 7 or 8 subunits. The pore is more preferably heptameric. The pore typically comprises a barrel or channel through which the ions may flow. The subunits of the pore typically surround a central axis and contribute strands to a transmembrane β barrel or channel or a transmembrane α-helix bundle or channel.

The barrel or channel of the pore typically comprises amino acids that facilitate interaction with the analyte. These amino acids are preferably located near the constriction of the barrel or channel. A pore for use in detecting nucleotides or nucleic acids typically comprises one or more positively charged amino acids, such as arginine, lysine or histidine. These amino acids typically facilitate the interaction between the pore and a nucleotide by interacting with the phosphate groups in the nucleotide or by π-cation interaction with the base in the nucleotide. A pore for use in detecting nucleotides or nucleic acids preferably has a ring of positively charged amino acids, such as arginine, lysine or histidine, located near the constriction of the barrel or channel. Each positively charged amino acid is typically provided by each of the pore subunits.

Pores for use in accordance with the invention can be β-barrel pores or α-helix bundle pores. β-barrel pores comprise a barrel or channel that is formed from β-sheets. Suitable β-barrel pores include, but are not limited to, β-toxins, such as α-hemolysin and leukocidins, and outer membrane proteins/porins of bacteria, such as *Mycobacterium smegmatis* porin A (MspA), outer membrane porin F (OmpF), outer membrane porin G (OmpG), outer membrane phospholipase A and *Neisseria* autotransporter lipoprotein (NalP). α-helix bundle pores comprise a barrel or channel that is formed from α-helices. Suitable α-helix bundle pores include, but are not limited to, inner membrane proteins and α outer membrane proteins, such as WZA.

The most preferred pore for use in the invention is α-hemolysin or a variant thereof. The α-hemolysin pore is formed of seven identical subunits (i.e. it is heptameric). The sequence of one subunit of α-hemolysin is shown in SEQ ID NO: 2. A variant is a heptameric pore in which one or more of the seven subunits has an amino acid sequence which varies from that of SEQ ID NO: 2 and which retains pore activity. A variant may include modifications that facilitate covalent attachment of the adaptor or orientation of the adaptor as discussed below.

1, 2, 3, 4, 5, 6 or 7 of the subunits in a variant α-hemolysin may have an amino acid sequence that varies from that of SEQ ID NO: 2. All seven subunits within a variant pore are may be identical but are typically different, particularly if one or more of the subunits has been modified to facilitate covalent attachment of the adaptor or orientation of the adaptor as discussed below.

Preferred variants of α-hemolysin for use in detecting nucleotides or nucleic acids have one or more positively charged amino acids, such as arginine, lysine or histidine, located near the constriction of the barrel or channel. The pore preferably has a ring of 4, 5, 6 or preferably 7 positively charged amino acids, such as arginine, lysine or histidine, located near the constriction of the barrel or channel. Each amino acid in the ring is typically provided by each of the variant subunits. Suitable variants include a positively charged amino acid at position 113 of each subunit. The pore for use in detecting nucleotides or nucleic acids is preferably α-hemolysin (M113K-RL2)$_7$ which comprises seven subunits as shown in SEQ ID NO: 4 or preferably α-

hemolysin $(M113H\text{-}RL2)_7$ which comprises seven subunits as shown in SEQ ID NO: 6 or most preferably α-hemolysin $(M113R\text{-}RL2)_7$ which comprises seven subunits as shown in SEQ ID NO: 8.

Other preferred variants of α-hemolysin have one or more uncharged amino acids, such as asparagine, or one or more aromatic amino acids, such as phenylalanine, located near the constriction of the barrel or channel. This has the purpose of binding and orienting β-CD-PDP for the covalent attachment reaction, and maintaining orientation of β-CD-PDP after covalent attachment. The pore preferably has a ring of 4, 5, 6 or preferably 7 uncharged or aromatic amino acids located near the constriction of the barrel or channel. Each amino acid in the ring is typically provided by each of the variant subunits. Suitable variants include an uncharged or aromatic amino acid at position 113 of each subunit. The pore for use in accordance with the invention is preferably α-hemolysin $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$.

The variant may be a naturally-occurring variant which is expressed by an organism, for instance by a *Staphylococcus* bacterium. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the amino acid sequence of SEQ ID NO: 2, a subunit of a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the subunit polypeptide may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of SEQ ID NO: 2 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 200 or more, for example 230, 250, 270 or 280 or more, contiguous amino acids ("hard homology").

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions may be made, for example, according to Table 1 below.

Table 1 – Conservative substitutions

Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other.

NON-AROMATIC	Non-polar	G A P
		I L V
	Polar – uncharged	C S T M
		N Q
	Polar – charged	D E
		H K R
AROMATIC		H F W Y

One or more amino acid residues of the amino acid sequence of SEQ ID NO: 2 may alternatively or additionally be deleted. Up to 1, 2, 3, 4, 5, 10, 20 or 30 residues may be deleted, or more.

Variants may include subunits made of fragments of SEQ ID NO: 2. Such fragments retain pore forming activity. Fragments may be at least 50, 100, 200 or 250 amino acids in length. Such fragments may be used to produce chimeric pores. A fragment preferably comprises the pore forming domain of SEQ ID NO: 2.

Variants include chimeric protein pores comprising fragments or portions of SEQ ID NO: 2. Chimeric protein pores are formed from subunits each comprising fragments or portions of SEQ ID NO: 2. The pore or channel part of a chimeric protein pore is typically formed by the fragments or portions of SEQ ID NO: 2.

One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the N-terminus or C-terminus of the amino acid sequence of SEQ ID NO: 2 or polypeptide variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids. A carrier protein may be fused to an amino acid sequence according to the invention.

Standard methods in the art may be used to determine homology. For example the UWGCG Package provides the BESTFIT program which can be used to

calculate homology, for example used on its default settings (Devereux *et al* (1984) *Nucleic Acids Research* **12**, p387-395). The PILEUP and BLAST algorithms can be used to calculate homology or line up sequences (such as identifying equivalent residues or corresponding sequences (typically on their default settings)), for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S.F *et al* (1990) J Mol Biol 215:403-10.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold (Altschul *et al*, supra). These initial neighbourhood word hits act as seeds for initiating searches to find HSP's containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci.* USA 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence

is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

A transmembrane protein pore may be modified for example by the addition of histidine or aspartic acid residues to assist their identification or purification or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence.

A pore may be labelled with a revealing label. The revealing label may be any suitable label which allows the pore to be detected. Suitable labels include, but are not limited to, fluorescent molecules, radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

The pore may be isolated from a pore-producing organism, such as *Staphylococcus aureus*, or made synthetically or by recombinant means. For example, the pore may be synthesized by *in vitro* translation and transcription. The amino acid sequence of the pore may be modified to include non-naturally occurring amino acids or to increase the stability of the compound. When the pores are produced by synthetic means, such amino acids may be introduced during production. The pores may also be altered following either synthetic or recombinant production.

The pores may also be produced using D-amino acids. This is conventional in the art for producing such proteins or peptides.

As discussed in detail above and below, the pore can contain one or more specific modifications to facilitate attachment and orientation of the adaptor. The pore may also contain other non-specific modifications as long as they do not interfere with the attachment and orientation of the adaptor. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the pores. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with $NaBH_4$, amidination with methylacetimidate or acylation with acetic anhydride.

A recombinant transmembrane pore can be produced using standard methods known in the art. Nucleic acid sequences encoding a pore or a pore subunit may be isolated and replicated using standard methods in the art. Nucleic acid sequences encoding a pore or a pore subunit may be expressed in a bacterial host cell using

standard techniques in the art. The pore or a pore subunit may be produced in a cell by *in situ* expression of the polypeptide from a recombinant expression vector. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

Nucleic acid sequences encoding a pore or a pore subunit may be isolated and replicated using standard methods in the art. Chromosomal DNA may be extracted from a pore-producing organism, such as *Staphylococcus aureus*. The gene encoding the pore or a pore subunit may be amplified using PCR involving specific primers. The amplified sequence may then be incorporated into a recombinant replicable vector such as a cloning vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus nucleic acid sequences encoding a pore or a pore subunit may be made by introducing a polynucleotide encoding a pore or a pore subunit into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells for cloning of polynucleotides encoding a pore or a pore subunit are known in the art and described in more detail below.

The nucleic acid sequence encoding a pore or a pore subunit may be cloned into suitable expression vector. In an expression vector, the nucleic acid sequence encoding a pore or a pore subunit is typically operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell. Such expression vectors can be used to express a pore or a pore subunit.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. Multiple copies of the same or different pore or pore subunit genes may be introduced into the vector.

The expression vector may then be introduced into a suitable host cell. Thus, in the case of a monomeric pore or a oligomeric pore comprising only one type of subunit, the method of the invention may be carried out on a cell produced by introducing a nucleic acid sequence encoding a pore or a pore subunit into an

expression vector, introducing the vector into a compatible bacterial host cell, and growing the host cell under conditions which bring about expression of the nucleic acid sequence encoding the pore or the pore subunit. The recombinantly-expressed pore subunits will self-assemble into a pore in the host cell membrane. Alternatively, the recombinant pore produced in this manner may be isolated from the host cell and inserted into another membrane. In the case of an oligomeric pore comprising at least two different subunits, the different subunits may be expressed separately in different host cells as described above, removed from the host cells and assembled into a pore in a separate membrane, such as a rabbit cell membrane.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said nucleic acid sequence and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example a tetracycline resistance gene. Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. A T7, *trc*, *lac*, *ara* or λ_L promoter is typically used.

The host cell typically expresses the pore or the pore subunit at a high level. Host cells transformed with a nucleic acid sequence encoding a pore or a pore subunit will be chosen to be compatible with the expression vector used to transform the cell. The host cell is typically bacterial and preferably *Escherichia coli*. Any cell with a λ DE3 lysogen, for example C41 (DE3), BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, Origami and Origami B, can express a vector comprising the T7 promoter.

A pore or a pore subunit may be produced in large scale following purification by any protein liquid chromatography system from pore-producing organisms or after recombinant expression as described above. Typical protein liquid chromatography systems include FPLC, AKTA systems, the Bio-Cad system, the Bio-Rad BioLogic system and the Gilson HPLC system. The naturally-occurring or recombinantly-produced pore or or pore subunit may then be inserted into a naturally-occurring or artificial membrane for use in accordance with the invention.

Pores or pore subunits can be inserted into a membrane using any method known in the art. For instance, pores or pore subunits can be introduced into a

membrane using the "pick and place" method described in International Application No. PCT/GB2006/001057 (published as WO 2006/100484).

Molecular adaptor

The transmembrane pore comprises a molecular adaptor that facilitates the interaction between the pore and the analyte. The presence of the adaptor improves the host-guest chemistry of the pore and the analyte. The principles of host-guest chemistry are well-known in the art. The adaptor has an effect on the physical or chemical properties of the pore that improves its interaction with the analyte. The adaptor typically alters the charge of the barrel or channel of the pore or specifically interacts with or binds to the analyte thereby facilitating its interaction with the pore.

Preferably, at least part of the adaptor is inside the barrel or channel of the pore. More preferably, all of the adaptor is inside the barrel or channel of the pore.

The adaptor mediates the interaction between the analyte and the pore. The analyte preferably reversibly binds to the pore via or in conjunction with the adaptor. The analyte most preferably reversibly binds to the pore via or in conjunction with the adaptor as it passes through the pore across the membrane. The analyte can also reversibly bind to the barrel or channel of the pore via or in conjunction with the adaptor as it passes through the pore across the membrane. The adaptor preferably constricts the barrel or channel so that it may interact with the analyte. The adaptor more preferably constricts the barrel or channel and restricts flow of ions through the pore. Flow of ions through the pore is restricted if the presence of the adaptor interferes with movement of ions through the pore and across the membrane. Flow of ions through the pore is required for analyte detection. As discussed above, the interaction between the analyte and the pore affects the current of ions flowing through the pore in a manner specific for that analyte. The adaptor is typically cyclic. The adaptor preferably has the same symmetry as the pore. An adaptor having seven-fold symmetry is typically used if the pore is heptameric (e.g. has seven subunits around a central axis that contribute 14 strands to a transmembrane β barrel). Likewise, an adaptor having six-fold symmetry is typically used if the pore is hexameric (e.g. has six subunits around a central axis that contribute 12 strands to a transmembrane β barrel, or is a 12-stranded β barrel).

The adaptor typically interacts with, such as binds to, the analyte. Interaction, such as binding, between the adaptor and the analyte increases the likelihood that the analyte interacts with the pore. The adaptor typically interacts with the analyte via host-guest chemistry. A portion of the adaptor is typically capable of interacting with the analyte. The portion comprises one or more chemical groups that are capable of interacting with the analyte. The one or more chemical groups preferably interact with the analyte by non-covalent interactions, such as hydrophobic interactions, hydrogen bonding, Van der Waal's forces, π-cation interactions and/or electrostatic forces. The one or more chemical groups that are capable of interacting with the analyte preferably comprise amino groups and/or hydroxyl groups. The amino groups or hydroxyl groups can be attached to primary, secondary or tertiary carbon atoms.

The portion of the adaptor that is capable of interacting with the analyte preferably comprises an aromatic ring that is capable of interacting with the analyte via π-cation interactions. The portion of the adaptor that is capable of interacting with the analyte more preferably comprises a ring of amino groups or hydroxyl groups, such as a ring of 6, 7 or 8 amino groups or hydroxyl groups. The portion most preferably comprises a ring of seven amino groups or hydroxyl groups. A ring of protonated amino groups may interact with negatively charged groups in an analyte, such as phosphate groups in a nucleotide. A ring of amino groups may interact with an analyte in combination with a ring of positively charged amino acids in the constriction of the barrel or channel of the pore. A ring of hydroxyl groups may interact with the analyte via hydrogen bonding. A ring of hydroxyl groups may interact with the analyte in combination with a ring of uncharged amino acids in the constriction of the barrel or channel of the pore.

The adaptor most preferably (1) constricts the barrel or channel and thereby restricts flow of ions through the pore and (2) interacts with or binds to the analyte.

As discussed in more detail below, the correct orientation of the adaptor within the barrel or channel of the pore can be facilitated by host-guest chemistry between the adaptor and the pore. The adaptor preferably comprises one or more chemical groups that are capable of interacting with one or more amino acids in the pore. The adaptor more preferably comprises one or more chemical groups that are

capable of interacting with one or more amino acids in the pore via non-covalent interactions, such as hydrophobic interactions, hydrogen bonding, Van der Waal's forces, π-cation interactions and/or electrostatic forces. The chemical groups that are capable of interacting with one or more amino acids in the pore are typically hydroxyls. The hydroxyl groups can be attached to primary, secondary or tertiary carbon atoms. The hydroxyl groups may form hydrogen bonds with uncharged amino acids in the pore, preferably in the barrel or channel of the pore. As discussed in more detail below below, the interaction of one or more chemical groups in the adaptor with one or more amino acids in the pore can be used to hold the adaptor in the correct orientation. The one or more chemical groups in the adaptor that are capable of interacting with one or more amino acids in the pore to hold the adaptor in the correct orientation may be the same one or more chemical groups that are capable of interacting with the analyte or may be different from the one or more chemical groups that are capable of interacting with the analyte.

Any adaptor that that facilitates the interaction between the pore and the analyte can be used. Suitable adaptors include, but are not limited to, cyclodextrins, cyclic peptides and cucurbiturils. The adaptor is preferably heptakis-6-amino-β-cyclodextrin (am_7-β-CD) or 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1β-CD). Table 2 below shows preferred combinations of pores and adaptors.

Table 2 - Suitable combinations of pores and adaptors

Pore	**Number of strands in the transmembrane β-barrel**	**Adaptor**
Leukocidin	16	γ-cyclodextrin (γ-CD)
OmpF	16	γ-cyclodextrin (γ-CD)
α-hemolysin (or a variant thereof discussed above)	14	β-cyclodextrin (β-CD) 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1β-CD) heptakis-6-amino-β-cyclodextrin (am_7-β-CD)

OmpG	14	β-cyclodextrin (β-CD) 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1β-CD) heptakis-6-amino-β-cyclodextrin (am_7-β-CD)
NalP	12	α-cyclodextrin (α-CD)
OMPLA	12	α-cyclodextrin (α-CD)

Preferably, the adaptor is not a linear polymer of repeating subunits, such as a peptide, polypeptide, protein, nucleic acid or polynucleotide. More preferably, the adaptor is not a protein kinase inhibitor peptide.

Covalent attachment

The adaptor is covalently attached to the pore. The adaptor can be covalently attached to the pore using any method known in the art. The adaptor may be attached directly to the pore. The adaptor is preferably attached to the pore using a bifunctional crosslinker. Suitable crosslinkers are well-known in the art. A preferred crosslinker is succinimidyl 3-(2-pyridyldithio)propionate (SPDP). Typically, the adaptor is covalently attached to the bifunctional crosslinker before the adaptor/crosslinker complex is covalently attached to the pore but it is also possible to covalently attach the bifunctional crosslinker to the pore before the bifunctional crosslinker/pore complex is attached to the adaptor.

The site of covalent attachment is selected such that the adaptor facilitates interaction of the analyte with the pore and thereby allows detection of the analyte. The site of covalent attachment is selected such that the analyte affects the current flowing through the pore in a manner specific for that analyte.

Preferably, the adaptor is covalently attached to the barrel or channel of the pore. The adaptor can be covalently attached at any site in the barrel or channel as long as the adaptor facilitates interaction of the analyte with the pore and thereby allows detection of the analyte. Preferably, the adaptor is covalently attached to an amino acid in the barrel or channel that is close to the site at which the interaction between the adaptor and the barrel or channel takes place. If the adaptor is

covalently attached to an amino acid in the barrel or channel that is far from the site at which the interaction between the adaptor and the barrel or channel takes place, a bifunctional crosslinker of suitable length may be used so that the adaptor can reach the site in the barrel or channel at which the interaction between the adaptor and the barrel or channel takes place. The adaptor does not have to be attached to the site within the barrel or channel with which it reversibly interacts when contacted with the pore.

The pore is preferably modified to facilitate the covalent attachment of the molecular adaptor with the bifunctional crosslinker. The barrel or channel of the pore is more preferably modified to facilitate the covalent attachment of the molecular adaptor with the bifunctional crosslinker. The pore may be modified using any method known in the art. Any of the variants discussed above may be used to facilitate the covalent attachment of the molecular adaptor or the bifunctional crosslinker.

The modification typically involves the introduction of one or more, such as 1, 2, 3, 4, 5, 6 or 7, amino acids into the pore, preferably the barrel or channel of the pore, to facilitate covalent attachment of the adaptor with the bifunctional crosslinker. The one or more amino acids may be introduced into the same or different subunits of the pore. Any amino acid that is capable of forming a covalent bond, such as cysteine, can be introduced. The amino acid may be naturally-occurring or non-naturally occurring. The one or more amino acids are preferably introduced by substitutions. Table 3 below shows an example of the types of amino acids that can be introduced into the pore to facilitate the covalent attachment of the adaptor or the bifunctional crosslinkers via specific chemical groups.

Table 3 - Modification of the pore to facilitate covalent attachment of the adaptor

Chemical group(s) in the adaptor or bifunctional crosslinker	**Example(s) of adaptors containing the chemical groups**	**Example(s) of crosslinker containing the chemical group(s)**	**Amino acid(s) that can be introduced into the pore**
Pyridyl disulfide group		succinimidyl 3-(2-pyridyldithio)propionate (SPDP)	Cysteine

In a preferred embodiment, a cysteine residue is introduced into a heptameric pore using one of the subunits shown in SEQ ID NOs: 14 and 16.

Orientation of the adaptor

The adaptor is covalently attached to the pore in an orientation that allows the analyte to be detected using the pore. The adaptor is oriented such the analyte affects the current flowing through the pore in a manner specific for that analyte. The adaptor is oriented so this it improves the host-guest chemistry of the pore and analyte. The adaptor is oriented so that it affects the physical or chemical properties of the pore and improves its interaction with the analyte. If the adaptor is capable of specifically interacting with or binding to the analyte, the adaptor is oriented so that it specifically interacts with or binds to the analyte. In the latter embodiment, the portion of the adaptor that interacts with the analyte is preferably oriented towards the end of the pore through which the analyte enters. More preferably, one or more chemical groups in the adaptor that are capable of interacting with the analyte are oriented towards the end of the pore through which the analyte enters. The groups are preferably amino groups or hydroxyl groups. The end of the pore through which the analyte enters may be the *cis* end or the *trans* end. The end is preferably the *trans* end.

The orientation of the adaptor is determined by the covalent attachment and/or the host-guest chemistry between the adaptor and the pore. The covalent attachment may be designed so that the adaptor is correctly oriented. For instance, the site (e.g. amino acid) at which the adaptor is covalently attached to the pore may be designed so that the adaptor is correctly oriented and/or a bifunctional crosslinker may be used so that the adaptor is correctly oriented.

Preferably, the pore is modified to facilitate orientation of the adaptor. Modification of the pore can improve the host-guest chemistry between the adaptor and the pore and thereby hold the adaptor in the correct orientation. More preferably, the barrel or channel of the pore is modified to facilitate orientation of the adaptor. The barrel or channel of the pore may be modified at any site that facilitates orientation of the adaptor. Preferably, the barrel or channel of the pore is modified at the site(s) at which host-guest chemistry occurs between the covalently-attached

adaptor and the pore. Most preferably, one or more amino acids in the barrel or channel of the pore that are capable of interacting with the covalently-attached adaptor via non-covalent interactions are modified to facilitate orientation of the adaptor. The amino acids that are modified may be one or more of the amino acids discussed above that facilitate interaction of the pore with the analyte. The pore may be modified using any method known in the art. Any of the variants discussed above may be used to facilitate the orientation of the molecular adaptor.

The modification typically involves the introduction into the pore, or preferably the barrel or channel, of one or more, such as 1, 2, 3, 4, 5, 6 or 7, amino acids to facilitate the orientation of the adaptor. The one or more amino acids may be introduced into the same or different subunits. The one or amino acids are preferably amino acids that hold the adaptor in the correct orientation. The one or amino acids are more preferably amino acids that hold the adaptor in the correct orientation via non-covalent interactions, such as hydrophobic interactions, hydrogen bonding, Van der Waal's forces, π-cation interactions and/or electrostatic forces. The one or more amino acids may be naturally-occurring or non-naturally occurring. The one or more amino acids are preferably uncharged amino acids. The uncharged amino acids may be polar, such as asparagine, or nonpolar, such as phenylalanine. Uncharged amino acids in the pore can orientate the adaptor by interacting with hydroxyl groups in the adaptor via hydrogen bonding. The one or more amino acids are preferably introduced by substitutions. Table 4 below shows the types of amino acids that can be introduced into the pore to correctly orient adaptors containing different chemical groups.

Table 4 - Modification of the pore to orient the adaptor

Chemical group(s) in the adaptor	**Example(s) of adaptors containing the chemical groups**	**Amino acid(s) that can be introduced into the pore to orient the adaptor**	**Interaction**
Hydroxyl groups attached to primary carbon atoms	6-monodeoxy-6-monoamino-β-cyclodextrin (am_1βCD)	Phenylalanine	Hydrophobic interactions
Hydroxyl groups attched to	6-monodeoxy-6-monoamino-β-	Asparagine	Hydrogen bonding

secondary carbon atoms	cyclodextrin ($am_1\beta CD$)		

In a preferred embodiment, a heptameric pore comprises six phenylalanine residues that are capable of orienting an adaptor containing hydroxyl groups attched to primary carbon atoms, such as 6-monodeoxy-6-monoamino-β-cyclodextrin ($am_1\beta CD$), towards the *cis* side of the pore. The heptameric pore can comprise six of the subunits shown in SEQ ID NO: 10.

In another preferred embodiment, a heptameric pore comprises six asparagine residues that are capable of orienting an adaptor containing secondary hydroxyl groups, such as 6-monodeoxy-6-monoamino-β-cyclodextrin ($am_1\beta CD$), towards the *trans* side of the pore. The heptameric pore can comprise six of the subunits shown in SEQ ID NO: 12.

Most preferably, the pore is modified to facilitate the covalent attachment and to facilitate the orientation of the adaptor. In such an embodiment, the spatial relationship between the site of covalent attachment and site(s) at which the pore is modified to facilitate the orientation of the adaptor is designed to ensure that the adaptor is held in the correct orientation. The modification can be made anywhere within the pore, preferably in the barrel or channel, as long as the adaptor is held in the correct orientation. Preferably, the modification that facilitates the covalent attachment and the modification(s) that facilitate the orientation of the adaptor are spatially proximate within the pore. More preferably, the modification that facilitates the covalent attachment and the modification(s) that facilitate the orientation of the adaptor are made in the same ring of amino acids within the barrel or channel of the pore. One amino acid in the ring may be modified to facilitate covalent attachment of the adaptor, while one or more, such as 1, 2, 3, 4, 5 or 6, of the other amino acids in the ring can be modified to facilitate orientation of the adaptor. If the modification that facilitates the covalent attachment is distant from the modification(s) that facilitate the orientation of the adaptor, a bifunctional crosslinker of appropriate length may be used to ensure that the adaptor is positioned close to the modification(s) that facilitate the orientation of the adaptor and thereby correctly oriented. Any of the modifications discussed above can be used.

In the most preferred embodiment, a heptameric pore is produced using six of the subunits shown in SEQ ID NO: 10 and one of the subunits shown in SEQ ID NO: 14 or using six of the subunits shown in SEQ ID NO: 12 and one of the subunits shown in SEQ ID NO: 16.

Method of producing the transmembrane protein pores of the invention

The invention also relates to a method of producing a transmembrane protein pore of the invention. The method comprises covalently attaching to a transmembrane protein pore a molecular adaptor that facilitates an interaction between the pore and an analyte. The adaptor can be covalently attached to the pore using any method known in the art.

Any of the pores, adaptors and bifunctional crosslinkers discussed above can be used in the method. The site of covalent attachment is selected as discussed above.

The pore is preferably modified to facilitate the covalent attachment of the adaptor and/or to facilitate the orientation of the adaptor. This is discussed in more detail above.

The method also comprises determining whether or not the adaptor is attached to the pore in an orientation that allows the analyte to be detected using the pore. This involves determining whether or not the transmembrane protein pore can be used to determine the presence or absence of the analyte. This can be done as described in more detail below. If the presence or absence of the analyte can be determined, the adaptor is in the correct orientation and a transmembrane protein pore of the invention has been produced. If the presence or absence of the analyte cannot be determined, the adaptor is likely to be in an incorrect orientation and a transmembrane protein pore of the invention has not been produced.

Method of detecting analytes

The present invention also relates to a method of determining the presence or absence of an analyte. The method comprises contacting the analyte with a transmembrane protein pore of the invention so that the analyte interacts with the pore and measuring the current passing through the pore during the interaction and thereby determining the presence or absence of the analyte. Any of the

transmembrane protein pores of the invention can be used. The benefits associated with using a transmembrane protein pore of the invention to detect an analyte is discussed above.

The analyte is present if the current flows through the pore in a manner specific for the analyte (i.e. if a distinctive current associated with the analyte is detected flowing through the pore). The analyte is absent if the current does not flow through the pore in a manner specific for the analyte.

The invention therefore involves stochastic sensing of an analyte. The invention can be used to differentiate analytes of similar structure on the basis of the different effects they have on the current passing through a transmembrane protein pore. The invention can also be used to measure the concentration of a particular analyte in a sample.

The invention may also be used in a sensor that uses many or thousands of pores of the invention in bulk sensing applications.

The method may be carried out using any suitable membrane/pore system in which a transmembrane protein pore of the invention is inserted into a membrane. The method is typically carried out using (i) an artificial membrane comprising a transmembrane protein pore of the invention, (ii) an isolated, naturally-occurring membrane comprising a transmembrane protein pore of the invention, or (iii) a cell expressing a transmembrane protein pore that has been modified in accordance with the invention. The method is preferably carried out using an artificial membrane. The membrane may comprise other transmembrane and/or intramembrane proteins as well as other molecules in addition to the transmembrane protein pore of the invention.

The method of the invention is typically carried out *in vitro*.

Membrane

The membrane forms a barrier to the flow of ions and analytes. The membrane is preferably a lipid bilayer. Lipid bilayers suitable for use in accordance with the invention can be made using methods known in the art. For example, lipid bilayer membranes can be formed using the method of Montal and Mueller (1972). The method of the invention may be carried out using lipid bilayers formed from any

membrane lipid including, but not limited to, phospholipids, glycolipids, cholesterol and mixtures thereof. The lipid bilayer is preferably formed from 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine.

Methods are known in the art for inserting pores into membranes, such as lipid bilayers. For example, the pore may be suspended in a purified form in a solution containing a lipid bilayer such that it diffuses to the lipid bilayer and is inserted by binding to the lipid bilayer and assembling into a functional state. Alternatively, the pore may be directly inserted into the membrane using the method described in M.A. Holden, H. Bayley. J. Am. Chem. Soc. 2005, 127, 6502-6503.

Interaction between the pore and analyte

The analyte may be contacted with the pore on either side of the membrane. The analyte may be introduced to the pore on either side of the membrane. The analyte may be contacted with the side of the membrane that allows the analyte to pass through the pore to the other side of the membrane. For example, the analyte is contacted with an end of the pore, which in its native environment allows the entry of ions or small molecules, such as analytes, into the barrel or channel of the pore such that the analyte may pass through the pore. In such cases, the analyte interacts with the pore and/or adaptor as it passes across the membrane through the barrel or channel of the pore. Alternatively, the analyte may be contacted with the side of the membrane that allows the analyte to interact with the pore via or in conjunction with the adaptor, dissociate from the pore and remain on the same side of the membrane. The present invention provides pores in which the orientation of the adaptor is fixed. As a result, the analyte is preferably contacted with the end of the pore towards which the adaptor is oriented. Most preferably, the analyte is contacted with the end of the pore towards which the portion of the adaptor that interacts with the analyte is orientated.

The analyte may interact with the pore in any manner and at any site. As discussed above, the analyte preferably reversibly binds to the pore via or in conjunction with the adaptor. The analyte most preferably reversibly binds to the pore via or in conjunction with the adaptor as it passes through the pore across the membrane. The analyte can also reversibly bind to the barrel or channel of the pore

via or in conjunction with the adaptor as it passes through the pore across the membrane.

During the interaction between the analyte and the pore, the analyte affects the current flowing through the pore in a manner specific for that analyte. For example, a particular analyte will reduce the current flowing through the pore for a particular mean time period and to a particular extent. In other words, the current flowing through the pore is distinctive for a particular analyte. Control experiments may be carried out to determine the effect a particular analyte has on the current flowing through the pore. Results from carrying out the method of the invention on a test sample can then be compared with those derived from such a control experiment in order to identify a particular analyte in the sample or determine whether a particular analyte is present in the sample. The frequency at which the current flowing through the pore is affected in a manner indicative of a particular analyte can be used to determine the concentration of that analyte in the sample.

Apparatus

The method may be carried out using any apparatus that is suitable for investigating a membrane/pore system in which a transmembrane protein pore is inserted into a membrane. The method may be carried out using any apparatus that is suitable for stochastic sensing. For example, the apparatus comprises a chamber comprising an aqueous solution and a barrier that separates the chamber into two sections. The barrier has an aperture in which the membrane containing the pore is formed. The analyte may be contacted with the pore by introducing the analyte into the chamber. The analyte may be introduced into either of the two sections of the chamber.

The method of the invention involves measuring the current passing through the pore during interaction with the analyte. Therefore the apparatus also comprises an electrical circuit capable of applying a potential and measuring an electrical signal across the membrane and pore. The method may be carried out using a patch clamp or a voltage clamp. The method preferably involves the use of a voltage clamp. The Example discloses one way to carry out a voltage clamp method.

Analyte

The analyte can be any substance in a sample. Suitable analytes include, but are not limited to, metal ions, inorganic salts, polymers, such as a polymeric acids or bases, dyes, bleaches, pharmaceuticals, diagnostic agents, recreational drugs, explosives and environmental pollutants.

The analyte can be an analyte that is secreted from cells. Alternatively, the analyte can be an analyte that is present inside cells such that the analyte must be extracted from the cells before the invention can be carried out.

The analyte is preferably an amino acid, peptide, polypeptide or a protein. The amino acid, peptide, polypeptide or protein can be naturally-occurring or non-naturally-occurring. The polypeptide or protein can include within it synthetic or modified amino acids. A number of different types of modification to amino acids are known in the art. For the purposes of the invention, it is to be understood that the analyte can be modified by any method available in the art.

The protein can be an enzyme, antibody, hormone, growth factor or growth regulatory protein, such as a cytokine. The cytokine may be selected from an interleukin, preferably IFN-1, IL-2, IL-4, IL-5, IL-6, IL-10, IL-12 or IL-13, an interferon, preferably IL-γ or other cytokines such as TNF-α. The protein may be a bacterial protein, fungal protein, virus protein or parasite-derived protein. Before it is contacted with the pore, the protein may be unfolded to form a polypeptide chain and thereby allow it to enter the barrel or channel of the pore and interact with the pore via or in conjunction with the adaptor.

The detection of nucleotides and nucleic acids is discussed in more detail below.

The analyte is present in any suitable sample. The invention is typically carried out on a sample that is known to contain or suspected to contain the analyte. The invention may be carried out on a sample that contains one or more analytes whose identity is unknown. Alternatively, the invention may be carried out on a sample to confirm the identity of one or more analytes whose presence in the sample is known or expected.

The sample may be a biological sample. The invention may be carried out *in vitro* on a sample obtained from or extracted from any organism or microorganism.

The organism or microorganism is typically prokaryotic or eukaryotic and typically belongs to one the five kingdoms: plantae, animalia, fungi, monera and protista. The invention may be carried out *in vitro* on a sample obtained from or extracted from any virus. The sample is preferably a fluid sample. The sample typically comprises a body fluid of the patient. The sample may be urine, lymph, saliva, mucus or amniotic fluid but is preferably blood, plasma or serum. Typically, the sample is human in origin, but alternatively it may be from another mammal animal such as from commercially farmed animals such as horses, cattle, sheep or pigs or may alternatively be pets such as cats or dogs.

The sample may be a non-biological sample. The non-biological sample is preferably a fluid sample. Examples of a non-biological sample include surgical fluids, water such as drinking water, sea water or river water, and reagents for laboratory tests.

The sample is typically processed prior to being assayed, for example by centrifugation or by passage through a membrane that filters out unwanted molecules or cells, such as red blood cells. The sample may be measured immediately upon being taken. The sample may also be typically stored prior to assay, preferably below -70°C.

Conditions

The method of the invention involves the measuring of a current passing through the pore during interaction with the analyte. Suitable conditions for measuring ionic currents through transmembrane protein pores are known in the art and disclosed in the Example. The method is carried out with a voltage applied across the membrane and pore. The voltage used is typically from –250mV to +250mV. The voltage used is preferably in a range having a lower limit selected from -200 mV, -150 mV, -100 mV, -50 mV, -20mV and 0 mV and an upper limit independently selected from +10 mV, + 20 mV, +50 mV, +100 mV, +150 mV and +200 mV.

The method is carried out in the presence of any alkali metal chloride salt. In the exemplary apparatus discussed above, the salt is present in the aqueous solution in the chamber. Potassium chloride (KCl), sodium chloride (NaCl) or caesium

chloride (CsCl) is typically used. KCl is preferred. The salt concentration is typically from 0.1 to 2.5M, from 0.3 to 1.9M, from 0.5 to 1.8M, from 0.7 to 1.7M, from 0.9 to 1.6M or from 1M to 1.4M. The salt concentration is preferably about 1M.

The method is typically carried out in the presence of a buffer. In the exemplary apparatus discussed above, the buffer is present in the aqueous solution in the chamber. Any buffer may be used in the method of the invention. One suitable buffer is Tris-HCl buffer. The method is typically carried out at a pH of from 4.0 to 10.0, from 4.5 to 9.5, from 5.0 to 9.0, from 5.5 to 8.8, from 6.0 to 8.7 or from 7.0 to 8.8 or 7.5 to 8.5. The pH used is preferably about 8.0.

The method is typically carried out at from 0°C to 100°C, from 15°C to 95°C, from 16°C to 90°C, from 17°C to 85°C, from 18°C to 80°C, 19°C to 70°C, or from 20°C to 60°C. The method is preferably carried out at room temperature.

Method of identifying an individual nucleotide

The present invention also relates to a method of identifying an individual nucleotide. The method comprises contacting the nucleotide with a transmembrane protein pore of the invention so that the nucleotide interacts with the pore and measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide. The invention therefore involves stochastic sensing of an individual nucleotide. The invention can be used to differentiate nucleotides of similar structure on the basis of the different effects they have on the current passing through a transmembrane protein pore. Individual nucleotides can be identified at the single molecule level from their current amplitude when they interact with the pore. The invention can also be used to determine whether or not a particular nucleotide is present in a sample. The invention can also be used to measure the concentration of a particular nucleotide in a sample.

Individual nucleotide

An individual nucleotide in accordance with the invention is a single nucleotide. An individual nucleotide is one which is not bound to another

polynucleotide by a nucleotide bond. A nucleotide bond involves one of the phosphate groups of a nucleotide being bound to the sugar group of another nucleotide. An individual nucleotide is typically one which is not bound by a nucleotide bond to another polynucleotide sequence of at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 500, at least 1000 or at least 5000 nucleotides. For example, the individual nucleotide has been digested from a target polynucleotide sequence, such as a DNA or RNA strand.

The method of the invention may be used to identify any nucleotide. The nucleotide can be naturally-occurring or artificial. A nucleotide typically contains a nucleobase, a sugar and at least one phosphate group. The nucleobase is typically heterocyclic. Suitable nucleobases include purines and pyrimidines and more specifically adenine, guanine, thymine, uracil and cytosine. The sugar is typically a pentose sugar. Suitable sugars include, but are not limited to, ribose and deoxyribose. The nucleotide is typically a ribonucleotide or deoxyribonucleotide. The nucleotide typically contains a monophosphate, diphosphate or triphosphate.

Suitable nucleotides include, but are not limited to, adenosine monophosphate (AMP), adenosine diphosphate (ADP), adenosine triphosphate (ATP), guanosine monophosphate (GMP), guanosine diphosphate (GDP), guanosine triphosphate (GTP), thymidine monophosphate (TMP), thymidine diphosphate (TDP), thymidine triphosphate (TTP), uridine monophosphate (UMP), uridine diphosphate (UDP), uridine triphosphate (UTP), cytidine monophosphate (CMP), cytidine diphosphate (CDP), cytidine triphosphate (CTP), cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), deoxyadenosine monophosphate (dAMP), deoxyadenosine diphosphate (dADP), deoxyadenosine triphosphate (dATP), deoxyguanosine monophosphate (dGMP), deoxyguanosine diphosphate (dGDP), deoxyguanosine triphosphate (dGTP), deoxythymidine monophosphate (dTMP), deoxythymidine diphosphate (dTDP), deoxythymidine triphosphate (dTTP), deoxyuridine monophosphate (dUMP), deoxyuridine diphosphate (dUDP), deoxyuridine triphosphate (dUTP), deoxycytidine monophosphate (dCMP), deoxycytidine diphosphate (dCDP) and deoxycytidine triphosphate (dCTP). The nucleotide is preferably AMP, TMP, GMP, UMP, dAMP, dTMP, dGMP or dCMP.

The nucleotide may be derived from the digestion of a nucleic acid sequence such as ribonucleic acid (RNA) or deoxyribonucleic acid. Individual nucleotides from a single nucleic acid sequence may be contacted with the pore in a sequential manner in order to sequence the whole or part of the nucleic acid. Sequencing nucleic acids in accordance with the second embodiment of the invention is discussed in more detail below.

The nucleotide is typically unmodified, such as when the nucleotide is derived from the digestion of a nucleic acid sequence. Alternatively, the nucleotide may be modified or damaged. The nucleotide is typically methylated. The nucleotide may be labelled with a revealing label. The revealing label may be any suitable label which allows the nucleotide to be detected. Suitable labels include fluorescent molecules, radioisotopes, e.g. ^{125}I, ^{35}S, and linkers such as biotin.

The nucleotide is typically present in any suitable biological sample. Suitable biological samples are discussed above.

Carrying out the method

All of the discussion above concerning detecting analytes, and in particular concerning the pores, membranes, apparatus and conditions that may be used, equally applies to this method.

In terms of conditions, the voltage used is preferably about +130 mV for deoxy-ribo nucleotides 5′ monophosphate, such as dAMP, dTMP, dGMP and dCMP, and +110 mV for ribo nucleotides 5′ monophosphate, such as AMP, TMP, GMP and UMP. The method is preferably carried out at +130mV at pH 8.0, 1M KCl for deoxy-ribo nucleotides 5′ monophosphate, such as dAMP, dTMP, dGMP and dCMP, and at +110mV at pH 8.0, 1M KCl for ribo nucleotides 5′ monophosphate, such as AMP, TMP, GMP and UMP.

Method of sequencing nucleic acids

The present invention also relates to a method of sequencing a target nucleic acid sequence. The method comprises (a) digesting an individual nucleotide from one end of the target sequence using a processive exonuclease; (b) contacting the nucleotide with a transmembrane protein pore of the invention so that the nucleotide

interacts with the pore; (c) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide; and (d) repeating steps (a) to (c) at the same end of the nucleic acid sequence and thereby determining the sequence of the nucleic acid. Hence, the method involves stochastic sensing of each single nucleotide of a nucleic acid sequence in a successive manner in order to sequence the nucleic acid.

A transmembrane protein pore of the invention is particularly suited to this method. In order to effectively sequence the nucleic acid, it is important to ensure that every nucleotide in the nucleic acid is identified in a successive manner. The fixed nature of the adaptor in a transmembrane protein pore of the invention means that a distinctive current will flow through the pore whenever each successive nucleotide interacts with the pore.

The whole or only part of the nucleic acid may be sequenced using this method. The nucleic acid can be any length. For example, the nucleic acid can be at least 10, at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 400 or at least 500 nucleotides in length. The nucleic acid can be naturally-occurring or artificial. For instance, the method may be used to verify the sequence of a manufactured oligonucleotide. The method is typically carried out *in vitro*.

Steps (b) and (c) of this method are generally identical to the steps carried out in the method of identifying nucleotides discussed above. All of the discussion above concerning detecting analytes, and in particular concerning the pores, membranes, apparatus and conditions that may be used, equally applies to this method. The nucleic acid is typically present in any biological sample as discussed above.

Processive exonuclease

The method of sequencing a target nucleic acid involves contacting the nucleic acid with a processive exonuclease to release individual nucleotides from one end of the nucleic acid. Processive exonucleases are enzymes that typically latch onto one end of a nucleic acid and digest the sequence one nucleotide at a time from that end. The processive exonuclease can digest the nucleic acid in the 5′ to 3′ direction or 3′ to 5′ direction. The end of the nucleic acid to which the processive

exonuclease binds is typically determined through the choice of enzyme used and/or using methods known in the art. Hydroxyl groups or cap structures at either end of the nucleic acid sequence may typically be used to prevent or facilitate the binding of the processive exonuclease to a particular end of the nucleic acid sequence.

Any processive exonuclease enzyme may be used in the method. The preferred enzyme for use in the method is lambda exonuclease. The sequence of one subunit of lambda exonuclease is shown in SEQ ID NO: 17. Three identical subunits interact to form a trimer exonuclease. Variants of lambda exonuclease are enzymes formed of polypeptide subunits which have an amino acid sequence which varies from that of SEQ ID NO: 17 and which retain processive exonuclease activity. The variants may vary from SEQ ID NO: 17 in the same manner and to the same extent as discussed for variants of SEQ ID NO: 2 above. A variant preferably comprises the domains responsible for binding to the nucleic acid and for digesting the nucleic acid (catalytic domain). A variant preferably has an increased or reduced rate of enzyme activity as required and/or higher salt tolerance compared to the wild-type enzyme. The processive exonuclease may be produced using any of the methods discussed above for the production of transmembrane protein pores.

The method of sequencing a target nucleic acid sequence involves contacting the nucleic acid sequence with the processive exonuclease so that the nucleotides are digested from the end of the nucleic acid at a rate that allows identification of each individual nucleotide as discussed above. Methods for doing this are well known in the art. For example, Edman degradation is used to successively digest single amino acids from the end of polypeptide such that they may be identified using High Performance Liquid Chromatography (HPLC). A homologous method may be used in the present invention.

The processive exonuclease is preferably covalently attached to the transmembrane protein pore. Methods for covalently attaching the processive exonuclease to the pore are well known in the art.

The rate at which the processive exonuclease must function is typically slower than the optimal rate of a wild-type processive exonuclease. A suitable rate of activity of the processive exonuclease in the method of the second embodiment involves digestion of from 0.5 to 1000 nucleotides per second, from 0.6 to 500

nucleotides per second, 0.7 to 200 nucleotides per second, from 0.8 to 100 nucleotides per second, from 0.9 to 50 nucleotides per second or 1 to 20 or 10 nucleotides per second. The rate is preferably 1, 10, 100, 500 or 1000 nucleotides per second. A suitable rate of processive exonuclease activity can be achieved in various ways. For example, variant processive exonucleases with a reduced optimal rate of activity may be used in accordance with the invention.

The activity of processive exonucleases is typically pH dependent such that their activity falls as pH is reduced. Hence, the method of the second embodiment is typically carried out at a pH of from 7.5 to 8.0 or from 7.7 to 8.0. The pH used is preferably about 8.0.

The rate of activity of processive exonucleases typically falls as salt concentration rises. However, very high salt concentrations typically have a detrimental effect on the activity of the enzyme. Another way of limiting the rate of the enzyme is to carry out the method at a salt concentration that reduces the rate of the activity of the enzyme without adversely affecting its activity. For example, the method may be carried out at a salt concentration of from 0.5 to 1M. The salt concentration is preferably about 1M.

Kits

The present invention also relates to kits that may be used to carry out the method of sequencing a target nucleic acid sequence. The kits are therefore suitable for sequencing nucleic acids. The kits comprise a transmembrane pore of the invention and a processive exonuclease.

The kit may additionally comprise one or more other reagents or instruments which enable any of the embodiments of the method mentioned above to be carried out. Such reagents or instruments include one or more of the following: suitable buffer(s) (aqueous solutions), means to obtain a sample from a subject (such as a vessel or an instrument comprising a needle), means to amplify nucleic acid sequences, a membrane as defined above or voltage or patch clamp apparatus. Reagents may be present in the kit in a dry state such that a fluid sample resuspends the reagents. The kit may also, optionally, comprise instructions to enable the kit to

be used in the method of the invention or details regarding which patients the method may be used for. The kit may, optionally, comprise nucleotides.

The following Example illustrates the invention:

Example

In this Example, we covalently attach β-cyclodextrin (β-CD) to the αHL pore, using the mutations at position 113 to orient the adaptors for covalent bonding and stabilize them within the pore lumen after reaction. This approach has two important outcomes. First, the adaptor cannot dissociate from the nanopore. Second, the orientation of the cyclodextrin can be controlled with either the primary hydroxyls or the secondary hydroxyls facing the *trans* entrance of the pore.

1. Materials and methods

1.1 αHL pores

The α-hemolysin mutant pores $(M113F\text{-}RL2)_7$ (WT background; SEQ ID NO: 10), $(M113N\text{-}RL2)_7$ (WT background; SEQ ID NO: 12), $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ (SEQ ID NOs: 10 and 14), $(M113N)_6(T117C\text{-}D8RL3)_1$ (SEQ ID NOs: 12 and 16) were expressed, assembled and purified as previously described.

RL2 is the product of a semisynthetic gene that was devised to permit cassette mutagenesis of the sequence encoding the transmembrane β barrel (Cheley, S., Braha, O., Lu, X., Conlan, S., and Bayley, H. (1999) *Protein Sci.* **8**, 1257-1267). It contains six silent restriction sites and five altered amino acids in the encoded polypeptide sequence (K8A, V124L, G130S, N139Q and I142L). D8RL2 is RL2 with an octa-aspartate tail.

RL3 is identical to the WT αHL polypeptide at the amino acid level, but the gene contains six silent mutations in the region encoding the loop that aid cassette mutagenesis (S. Cheley, unpublished). D8RL3 is RL3 with an octa-aspartate tail. Figure 12 shows the complete polynucleotide sequence of pT7-RL3-D8.

1.2 Chemicals

Reagents were obtained as follows: 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids), pentane (JT Baker), hexadecane (99+%, Sigma-Aldrich), 6-monodeoxy-6-monoamino-β-cyclodextrin hydrochloride (99%, Cyclolabs Budapest, Hungary), Trizma base (99.9%, Sigma-Aldrich), concentrated HCl (analytical reagent grade, Fisher Scientific), potassium chloride (99%, Sigma-Aldrich), 2-adamantanamine·HCl (99%, Aldrich), succinimidyl 3-(2-pyridyldithio)propionate (SPDP, 95%, Fluka), triethylamine (99.5%, Fluka).

1.3 Synthesis of 6-monodeoxy-6-[3-(2-pyridyldithio)propionyl]monoamino-β-cyclodextrin (βCD-PDP)

6-Monodeoxy-6-monoamino-β-cyclodextrin (am_1βCD) hydrochloride (11.7 mg, 0.01 mmol) and succinimidyl 3-(2-pyridyldithio)propionate (SPDP, 6.24 mg, 0.02 mmol) were dissolved in 5 mL MeOH / H_2O (1:1). Triethylamine (100 μL) was added over 5 min and the mixture was stirred overnight. The desired product appeared as a white precipitate, which was filtered off and washed successively with cold water and acetone. An analytical sample was obtained by recrystallization from DMSO and water. ^{1}H NMR (DMSO-*d*6): δ 8.61 (m, 1H), 7.82 (m, 2H), 7.25 (ddd, *J* 7.2, 4.8, 1.2 Hz, 1H), 5.75 (m, 14H), 4.83 (m, 7H), 4.67 (m, 7H), 3.64 (m, 35H), 2.99 (t, *J* 7.2Hz, 2H), 2.55 (t, *J* 7.2Hz, 2H). ESI-MS $[M+Na]^+$ 1353.37; theoretical distribution 1353.37, 100%; 1354.38, 52.3%; 1355.38, 26.3%; 1356.38, 8.7%, 1357.38, 2.7%; 1358.38, 0.8%. Found 1353.37, 100%; 1354.38, 58.6%, 1355.37, 33.1%, 1356.38, 12.6%, 1357.38, 4.3%, 1358.38, 1.2%.

1.4 Single channel current recording

A bilayer of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) was formed across an aperture of 100-150 μm in diameter in a polycarbonate film (20-μm thickness from Goodfellow, Malvern, PA) that divided a planar bilayer chamber into two compartments, *cis* and *trans*. Both compartments contained 1 mL of buffer. Engineered protein pores were added to the *cis* compartment, which was connected to ground. βCD-PDP was added to the *trans* compartment, which was connected to the head-stage of the amplifier.

All experiments were carried out at ±100 mV in 25 mM Tris·HCl, 1 M KCl, pH 8.0, at 22.5 ± 2°C, unless otherwise stated. Fresh frozen aliquots of protein and βCD-PDP were used each day. Currents were recorded with a patch clamp amplifier (Axopatch 200B; Axon instruments, Foster City, CA), low pass filtered with a built-in 4-pole Bessel filter at 10 kHz, and sampled at 20 kHz by a computer equipped with a Digidata 1200 A/D converter (Axon instruments).

1.5 Data analysis

Current traces were analysed with pClamp 9.0 software (Axon Instruments). Events were detected using the Event Detection feature, and used to construct amplitude and dwell time histograms. Origin (Microcal, Northampton, MA) was used for curve fitting and graph presentation.

1.6 Molecular models

The pdb files of the heptameric α-hemolysin pores $(M113F\text{-}RL2)_7$ and $(M113N\text{-}RL2)_7$ provided by Michelle Montoya and Eric Gouaux were displayed in PyMOL (DeLano Scientific, San Carlos, CA). The βCD-PDP structure was also constructed in PyMOL by using the building tools.

2. Results

2.1 Design of the linker and derivatization of engineered αHL pores with βCD-PDP

The X-ray structures of $(M113F\text{-}RL2)_7$•βCD and $(M113N)_7$•βCD (Figure 1, Montoya and Gouaux, unpublished) reveal the positions of non-covalently bound βCD within the β barrel of the αHL pore. In $(M113F\text{-}RL2)_7$•βCD, the seven phenyl rings of Phe-113 take part in hydrophobic interactions with the seven primary 6-hydroxyl groups of βCD, resulting in a specific orientation of the βCD within the pore, in which the primary hydroxyls point towards the *cis* mouth (Figure 1b). In contrast, in $(M113N\text{-}RL2)_7$•βCD, the seven amide groups of Asn-113 and the ε-amino groups of Lys-147 are, respectively, within hydrogen bonding distance of the seven 2- and the seven 3-secondary hydroxyl groups of βCD. In this orientation, the

primary hydroxyl groups of βCD point towards the *trans* entrance of the pore (Figure 1c).

By molecular modeling, we identified suitable positions in the lumen of the αHL barrel at which to attach βCD. In $(M113F\text{-}RL2)_7$•βCD, the distance between the center of a phenyl ring of Phe-113 and the nearest O atom of a primary hydroxyl groups of βCD is 6.5 ± 0.5 Å (average of the seven positions) (Figure 1b). In $(M113N\text{-}RL2)_7$•βCD, the distance between the O atom of a primary hydroxyl groups of βCD and the closest O atom of Thr-117 is on average 6.0 ± 0.5 Å (Figure 1c). We chose a bifunctional crosslinker succinimidyl 3-(2-pyridyldithio)propionate (SPDP) to link 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1βCD, in which a single primary hydroxyl group of βCD is substituted with an amino group) to a cysteine residue at position-113 or position-117. We first coupled am_1βCD with SPDP to form βCD-PDP (Figure 2). In βCD-PDP, the pyridyl disulfide at the end of the linker can be attacked by the free thiol of a cysteine residue in the αHL pore to form a disulfide bond. The length of the linker was measured by building βCD-PDP in PyMol; the distance between the N atom in the amide bond of am_1βCD and the pyridyl S atom is approximately 7.0 Å in an extended conformation. Therefore, the linker is slightly longer than required for linking am_1βCD to a cysteine at position-113 with the βCD in the orientation found in $(M113F\text{-}RL2)_7$ or to a cysteine at position-117 with the βCD in the orientation found in $(M113N\text{-}RL2)_7$.

On the basis of these modeling studies, a heteroheptameric αHL pore $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ was engineered with six phenylalanine residues at position-113 and a single cysteine at the seventh position. The cysteine residue in the M113C subunit was designed to be able to react with the pyridyl disulfide of βCD-PDP presented from the *trans* side of the bilayer in such a way that the primary hydroxyls of the cyclodextrin would remain in the proximity of the six Phe residues of the M113F subunits. After reaction, βCD would be anchored inside the αHL pore as a permanent molecular adaptor. In another construct, we placed a cysteine residue at position 117 in the engineered pore $(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$. This pore was designed to react with βCD-PDP while maintaining the stabilizing interactions

of the secondary hydroxyls hydrogen bonded to residues Asn-113 and Lys-147 in the M113N subunits.

We portrayed the desired covalent complexes $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD and $(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$-βCD in PyMOL. In $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD, the distance between the amide N atom of am_1βCD and the S atom of the cysteine at position 113 is 5.6 Å (Figures 3c and 4). In $(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$-βCD, the distance between the amide N atom of am_1βCD and the S of cysteine at position 117 is 6.2 Å (Figures 5c and 6). These values are close to the corresponding measurements taken from the X-ray structures of the non-covalent complexes (see above) and confirm that the choice of the linker length is correct.

2.2 Electrical characterization of $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD

When a single $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ pore was introduced into the lipid bilayer from the *cis* chamber, we measured an ionic current of 98 ± 5 pA (+100 mV, 25 mM Tris·HCl, pH 8.0, 1 M KCl, n = 6). To observe the reaction with the pore, βCD-PDP was added to the *trans* chamber, and the potential was held at +100 mV. Two types of transient blocking events were observed, amounting to 70% and 95% of the total current. By contrast, when βCD interacts with $(M113F\text{-}RL2)_7$, only 70% current blockades are observed. Therefore, the 70% block is most likely generated when βCD-PDP binds within the pore with its primary hydroxyls interacting with the Phe-113 side chains and the linker protruding towards the *cis* side. The 95% block may occur when βCD-PDP enters the pore with the pyridyl ring lodged inside the cavity of the cyclodextrin, thereby reducing the current flow. Eventually, after a period ranging from 5 min to 1 h, the current level became permanently locked into a state of 70% block. This event was preceded by an 85% block that lasted for 48 ± 7 ms (n = 5) (Figures 3c and 4). No recovery of the current from the 70% blocked state to that of the unmodified pore was observed during a total recording period of 24 h (over 16 experiments). By comparison, the mean τ_{off} of βCD non-covalently bound to $(M113F\text{-}RL2)_7$ is 316 ± 62 ms under the same conditions. The blocked state is presumed to represent the αHL pore with βCD

covalently attached through a disulfide bond at position 113 (Figures 3c and 4). In keeping with this idea, the disulfide bond was cleaved after the addition of 2 mM DTT to the *cis* chamber. During cleavage a transient intermediate state with a lifetime of 150 ± 12 ms (n = 5) and 98% current block was observed (Figure 3d). The 70% block in $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD was also maintained after the applied potential was ramped up and down within the range of ± 250mV. It might be noted that the binding of neutral molecules such as βCD within the αHL pore is voltage-dependent, owing to the effects of electrosmosis (Gu, L.-Q., Cheley, S., and Bayley, H. (2003) *Proc.Natl.Acad.Sci.USA* **100**, 15498-15503). Covalent binding therefore prevents βCD dissociation at potentials where the non-covalent complex is short-lived.

Current-voltage (I-V) curves were measured for $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD and the non-covalent complex $(M113F)_7$•βCD. The I-V curves are not significantly different between +150 and -60 mV. However, when the applied potential is below -60 mV, the I-V curve for $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD is more rectifying than that of $(M113F\text{-}RL2)_7$•βCD (Figure 3e).

2.3 Electrical characterization of $(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$

When a single $(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$ protein pore was introduced into the lipid bilayer from the *cis* chamber, we measured an ionic current of -80 ± 5 pA (-100 mV, 25 mM Tris·HCl, pH 8.0, 1 M KCl, n = 6). After βCD-PDP was added to the *trans* compartment, only one type of transient blocking event was observed with a 70% current block, which is close to the 65% block observed with $(M113N\text{-}RL2)_7$•βCD. After a period ranging from 1 to 20 min, an intermediate state with a lifetime of 5 ± 2 ms (n = 5) and 80% current block was observed, which was followed by a permanent 70% current block (Figures 4 and 5c). As in the case of $(M113F\text{-}RL2)_6$ $(M113C\text{-}D8RL2)_1$-βCD, this 70% block was not reversed during a total of 24 h (over 16 experiments), while the mean τ_{off} for $(M113N\text{-}RL2)_7$•βCD is only 10.7 ± 1.5 s under the same conditions. Again, the applied potential could be ramped up and down within the range of ±250 mV without unblocking the pore. Therefore, this state is presumed to represent the αHL pore with βCD covalently

attached through a disulfide bond to position 117 (Figures 4 and 5c). Again, the disulfide bond was cleaved after the addition of 2 mM DTT to the *cis* chamber (Figure 5d). During cleavage, a transient intermediate state with 80% current block and a lifetime of 300 ± 20 ms (n = 5) was observed (Figure 5d).

Current-voltage (I-V) curves were measured for the adduct (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD. he I-V curve of this complex is similar to that of (M113N-RL2)$_7$•βCD within the range of ± 100 mV, diverging slightly at both high positive and negative potentials (Figure 5e).

2.4 Attempts to link βCD in reverse orientations

To further investigate and confirm the proposed orientation of the covalently-attached βCD, we prepared the heteroheptamers (M113N-RL2)$_6$(M113C-D8RL3)$_1$ and (M113F-RL2)$_6$(T117C-D8RL2)$_1$, where covalent bond formation between βCD-PDP and the protein would in each case produce an orientation of the cyclodextrin opposite to that which the 113 mutations would normally support, as evidenced in the X-ray structures (Figure 1).

In the case of (M113N-RL2)$_6$(M113C-D8RL2)$_1$, covalent attachment of βCD-PDP to Cys-113 failed to occur after ten attempts of 30 min each. By contrast, the heteroheptamer (M113F-RL2)$_6$(T117C-D8RL3)$_1$ did react with βCD-PDP as indicated by a permanent 80% current block. Like the two cases described earlier, the attached βCD resisted potential ramps of ±250mV and could be detached by the addition of 2 mM DTT to the *cis* chamber (not shown). The presumed covalent adduct, (M113F-RL2)$_6$(T117C-D8RL3)$_1$-βCD, was examined by electrical recording in 25 mM Tris·HCl, pH 8.0, 1 M KCl, at ±100 mV (n = 3). The single-channel current (24.0 pA) was lower than that of (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD (40.0 pA) and displayed frequent full blockades at both positive and negative potentials (data not shown). In addition, the current passed by (M113F-RL2)$_6$(T117C-D8RL3)$_1$-βCD was far noisier than that observed with (M113N-RL2)$_6$(T117C-D8RL3)$_1$-βCD, precluding its use for the stochastic sensing of organic analytes (see below).

2.5 Analyte detection using engineered αHL pores with covalently-attached βCD adaptors

To evaluate the newly created constructs as biosensor detection elements, we carried out stochastic sensing of 2-adamantanamine (2-$AdNH_2$), a model analyte (12,15). The βCD adaptor was attached to $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ or $(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$ *in situ* as described above. 2-Adamantanamine·HCl was then added to either the *cis* or *trans* chamber and the transient current substates were observed (Figure 7). These events represent the formation and dissociation of individual 2-$AdNH_2$•βCD complexes within the engineered αHL pore. The mean dwell times (τ_{off}) were similar to those observed when 2-$AdNH_2$ was added to the non-covalent complexes $(M113F\text{-}RL2)_7$•βCD and $(M113N\text{-}RL2)_7$•βCD (Table 5 below). The small differences in τ_{off} probably arise from interference by the linker.

Table 5 - Stochastic sensing of 2-adamantanamine (2-$AdNH_2$) with covalent and non-covalent adducts of βCD and αHL

Construct	$\tau_{off} \pm$ SD (ms)	
	2-$AdNH_2$ (*cis*)	2-$AdNH_2$ (*trans*)
$(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD	0.70 ± 0.11 (n = 7)	0.99 ± 0.05 (n = 9)
$(M113F\text{-}RL2)_7$•βCD	1.20 ± 0.08 (n = 6)	1.25 ± 0.07 (n = 6)
$(M113N\text{-}RL2)_6(T117C\text{-}D8RL3)_1$-βCD	1.15 ± 0.09 (n = 7)	1.16 ± 0.07 (n = 9)
$(M113N\text{-}RL2)_7$•βCD	1.12 ± 0.06 (n = 6)	1.56 ± 0.08 (n = 8)

Conditions: 25 mM Tris·HCl, pH 8.0, 1 M KCl, +100 mV, with 2-AdNH$_2$·HCl (*cis* or *trans*) at 30 μM. The number of individual experiments is in parentheses.

2.6 Examination of long binding events using αHL pores with covalently-attached adaptors

With covalently-attached adaptors it is possible to study the kinetics of two types of events that are more difficult to examine with non-covalently attached adaptors: long-lived and rare events. *Cis*-Lithocholic acid is the tightest binding molecule known for βCD (Yang, Z., and Breslow, R. (1997) *Tetrahedron Lett.* **38**, 6171-6172). Therefore, we chose this molecule to test the newly constructed pores for studying binding events of long duration. (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD was selected because of its quieter background signal in the absence of analyte (see Figure 7, for example). When *cis*-lithocholic acid (50 μM, Figure 8a) was introduced into the *cis* chamber in 25 mM Tris·HCl, pH 8.0, 1 M KCl, at +100 mV (n = 3), surprisingly, no signal was observed. This is perhaps due to steric hindrance to binding to the *cis* side of (M113F-RL2)$_6$(M113C-D8RL2)$_1$-βCD by the linker that attaches the βCD to the barrel wall. When *cis*-lithocholic acid (50 μM) was introduced into the *trans* chamber at -100 mV (n = 3), we observed many very long events. The events showed a 98% current block, with a 92% block for 200 ± 100 ms at the beginning of each event and for 150 ± 100 ms at the end of the event (Figure 8b, Figure 9). The mean τ_{off} of the events was 8.6 ± 7.8 s (number of events = 60; range 0.5 s to 80 s), which is considerably longer than the τ_{off} (0.68 ± 0.12 s) for βCD in the (M113F-RL2)$_7$•βCD complex under the same conditions, i.e. βCD would most often dissociate from the non-covalent (M113F-RL2)$_7$•βCD during a lithocholic acid binding event. This shows the advantage of the covalently-attached adaptor over the non-covalent complexes for examining long-lived binding events.

2.7 αHL pores with covalently-attached adaptors at high temperatures

We speculated that αHL pores with covalently attached CDs might work better than underivatized pores at high temperatures when the dwell time of CDs is reduced (Kang, X., Gu, L.-Q., Cheley, S., and Bayley, H. (2005)

Angew.Chem.Int.Ed.Engl., 1495-1499). Single-channel current traces of $(M113F)_6(M113C\text{-}D8RL2)_1$-βCD were obtained at up to 55°C. The single-channel currents depended linearly on the temperature (Figure 10). I (pA) = 14.3 + 0.54T(°C) in 25 mM Tris.HCl, pH 8.0, 1 M KCl at +100 mV, which is a similar temperature dependence to that observed previously with an αHL pore with non-covalently bound βCD (Kang, X., Gu, L.-Q., Cheley, S., and Bayley, H. (2005) *Angew.Chem.Int.Ed.Engl.*, 1495-1499). No dissociation of βCD from the $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD pore was observed during these experiments. For example, the current passed by $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD remained constant at 55°C during a 5 min measurement, while the τ_{off} for βCD bound to $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ was only 56 ± 3.8 ms under the same conditions. These data show that αHL pores with covalently attached CDs permit stochastic sensing in aqueous solution under extreme conditions.

2.8 Permanent alteration of ion selectivity in αHL pores with covalently-attached adaptors

Non-covalent molecular adaptors lodged within the αHL pore can drastically alter the charge selectivity for ion transport (Gu, L.-Q., Dalla Serra, M., Vincent, J. B., Vigh, G., Cheley, S., Braha, O., and Bayley, H. (2000) *Proc.Natl.Acad.Sci.USA* **97**, 3959-3964). For example, at pH 7.5, the charge selectivity (P_{K^+}/P_{Cl^-}, KCl: *cis* 200 mM, *trans* 1000 mM) of WT-αHL changes from 0.55 ± 0.02 to 0.25 ± 0.01 after βCD becomes lodged in the lumen, i.e. the pore becomes more anion selective (Gu, L.-Q., Dalla Serra, M., Vincent, J. B., Vigh, G., Cheley, S., Braha, O., and Bayley, H. (2000) *Proc.Natl.Acad.Sci.USA* **97**, 3959-3964). Therefore, we expected that βCD, covalently attached inside the β barrel, would alter the ion selectivity of an αHL pore permanently. We took $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ and $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD as a test case and constructed I-V curves for single-channel currents recorded under both *cis/trans* and *trans/cis* KCl gradients (Figure 11). The measured conductance values (g) and reversal potentials (*Vr*), and the charge selectivity (P_{K^+}/P_{Cl^-}) calculated from the Goldman-Hodgkin-Katz equation (Hille, B. (2001) *Ion channels of excitable membranes, 3rd edition,*, Sinauer,

Sunderland, MA, USA), of both the underivatized and derivatized pores are shown in Table 6 (below).

Table 6 - Reversal potentials (Vr), calculated charge selectivities (P_{K+}/P_{Cl-}) and conductance values (g) of $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$ and $(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD

Pore	Buffer*	*Vr* (mV)	P_{K+}/P_{Cl-}	*g* (pS)†
$(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$	*cis* 200 / *trans* 1000	11.4 ± 0.1	0.47 ± 0.01	973 ± 10
	cis 1000 / *trans* 200	-5.0 ± 0.1	0.73 ± 0.02	
$(M113F\text{-}RL2)_6(M113C\text{-}D8RL2)_1$-βCD	*cis* 200 / *trans* 1000	28.3 ± 0.2	0.10 ± 0.01	295 ± 5
	cis 1000 / *trans* 200	-26.0 ± 0.2	0.14 ± 0.01	

For each entry, three separate experiments were performed. The reversal potentials (*Vr*) are mean values (n = 3) under the conditions stated. Charge selectivities (P_{K+}/P_{Cl-}, n = 3) and conductance values (n = 3) are quoted as the mean ± SD. P_{K+}/P_{Cl-} was calculated with the Goldman-Hodgkin-Katz equation by using activities and the experimental V_r value (20).

*25 mM Tris·HCl, pH 8.0. The salt concentrations (KCl) are given in mM.

†+100 mV, 1 M KCl, 25 mM Tris·HCl, pH 8.0, in both chambers.

The covalently-attached molecular adaptor indeed altered the charge selectivity of the αHL pore and this effect was permanent. Again the pore becomes more anion selective and the observed changes in P_{K+}/P_{Cl-} are similar to those seen with non-covalently bound βCD (Gu, L.-Q., Dalla Serra, M., Vincent, J. B., Vigh, G., Cheley, S., Braha, O., and Bayley, H. (2000) *Proc.Natl.Acad.Sci.USA* **97**, 3959-3964). This feature should be useful in the de novo design of membrane channels both for basic studies of ion permeation and for applications in biotechnology, where a varied assemblage of pores with modified properties would be of great utility (Hwang, W. L., Holden, M. A., White, S., and Bayley, H. (2007) *submitted for*

publication; and Holden, M. A., Needham, D., and Bayley, H. (2007) *J Am Chem Soc*, in press).

3. Conclusion

Engineering of the αHL pore to allow the binding of organic analytes for stochastic detection has been successful only rarely (Guan, X., Gu, L.-Q., Cheley, S., Braha, O., and Bayley, H. (2005) *ChemBioChem* **6**, 1875-1881), but non-covalent molecular adaptors have proved useful in this regard, despite certain limitations imposed by their continual association with and dissociation from the pore. To remedy the latter problem, we have now covalently attached the βCD adaptor inside the αHL pore. Further, the attachment has been performed with control over the orientation of the βCD, i.e. with the primary hydroxyls of the βCD facing either the *trans* or the *cis* entrance of the pore. Covalent attachment means that there are no gaps in detection, and control over the βCD orientation is important because it is likely that certain analytes bind through only one of the two entrances to the βCD cavity or that they bind in different ways depending upon the side of entry (Kang, X. F., Cheley, S., Guan, X., and Bayley, H. (2006) *J Am Chem Soc* **128**(33), 10684-10685). To exemplify the utility of αHL pores with covalently-attached βCD in stochastic sensing, we show here that the pores can be used to detect analytes with long dwell times within βCD. In this case, the βCD would normally dissociate from the pore before the analyte dissociates from the βCD, degrading the information content of the signal (Braha, O., Webb, J., Gu, L.-Q., Kim, K., and Bayley, H. (2005) *ChemPhysChem* **6**, 889-892). The effects of continual CD association and dissociation are exacerbated at high temperatures (Kang, X., Gu, L.-Q., Cheley, S., and Bayley, H. (2005) *Angew.Chem.Int.Ed.Engl.*, 1495-1499), and as demonstrated here covalent attachment also remedies this problem. We also show that αHL pores with covalently-attached βCD can have permanently altered ion selectivity, which might be useful, for example, for building nanobatteries (Hwang, W. L., Holden, M. A., White, S., and Bayley, H. (2007) *submitted for publication*; and Holden, M. A., Needham, D., and Bayley, H. (2007) *J Am Chem Soc*, in press).

Single molecule DNA sequencing (Bayley, H. (2006) *Curr Opin Chem Biol* **10**, 628-637) will also benefit from the use of protein nanopores containing

covalently attached βCDs. In one approach, deoxyribonucleoside 5'-monophosphates are released from an individual DNA strand by a processive exonuclease (Jett, J. H., Keller, R. A., Martin, J. C., Marrone, B. L., Moyzis, R. K., Ratliff, R. L., Seitzinger, N. K., Shera, E. B., and Stewart, C. C. (1989) *J.Biomol.Struct.Dynam.* **7**, 301-309). In the original manifestation, fluorescent base analogs would be released from a fully-substituted transcribed DNA strand, which has proved very difficult to implement. However, we have recently shown that the four deoxyribonucleoside 5'-monophosphates can be distinguished by using the non-covalent αHL $(M113R\text{-}RL2)_7$•$am_7βCD$ complex, where $am_7βCD$ is heptakis-(6-deoxy-6-amino)-β-cyclodextrin (Astier, Y., Braha, O., and Bayley, H. (2006) *J Am Chem Soc* **128**(5), 1705-1710). While this is a step forward, for successful sequencing, every base released by an exonuclease must be captured by the detection element and identified. Therefore if the αHL pore with a molecular adaptor is to be used, the adaptor being covalently attached will help to avoid gaps in sequencing.

Tables 7 to 16 below set out the sequence information

Table 7 – SEQ ID NOs: 1 and 2 - Wild-type α-hemolysin from *Staphylococcus aureus*
SEQ ID NO: 2 is encoded by residues 1 to 879 of SEQ ID NO: 1

```
gcagattctg atattaatat taaaaccggt actacagata ttggaagcaa tactacagta    60
aaaacaggtg atttagtcac ttatgataaa gaaaatggca tgcacaaaaa agtattttat   120
agttttatcg atgataaaaa tcacaataaa aaactgctag ttattagaac gaaaggtacc   180
attgctggtc aatatagagt ttatagcgaa gaaggtgcta acaaaagtgg tttagcctgg   240
ccttcagcct ttaaggtaca gttgcaacta cctgataatg aagtagctca aatatctgat   300
tactatccaa gaaattcgat tgatacaaaa gagtatatga gtactttaac ttatggattc   360
aacggtaatg ttactggtga tgatacagga aaaattggcg gccttattgg tgcaaatgtt   420
tcgattggtc atacactgaa atatgttcaa cctgatttca aaacaatttt agagagccca   480
actgataaaa aagtaggctg gaaagtgata tttaacaata tggtgaatca aaattgggga   540
ccatatgata gagattcttg gaacccggta tatggcaatc aacttttcat gaaaactaga   600
aatggttcta tgaaagcagc agataacttc cttgatccta acaaagcaag ttctctatta   660
tcttcagggt tttcaccaga cttcgctaca gttattacta tggatagaaa agcatccaaa   720
caacaaacaa atatagatgt aatatacgaa cgagttcgtg atgattacca attgcattgg   780
acttcaacaa attggaaagg taccaatact aaagataaat ggacagatcg ttcttcagaa   840
agatataaaa tcgattggga aaaagaagaa atgacaaatt aa                      882

ADSDINIKTG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT    60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYMSTLTYGF   120
NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG   180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK   240
QQTNIDVIYE RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN          293
```

Table 8 – SEQ ID NOs: 3 and 4 - α-hemolysin M113H-RL2
SEQ ID NO: 4 is encoded by residues 39 to 920 of SEQ ID NO: 3
n = a, c, g or t
X at location 5 stands for Leu or Phe
X at location 7 stands for Tyr, Trp, Cys, Ser, Leu or Phe

```
gttctgttta actttaagaa gggagatata catatgagca gattctgata ttnacntnng    60
cgaccggtac tacagatatt ggaagcaata ctacagtaaa aacaggtgat ttagtcactt   120
atgataaaga aaatggcatg cacaaaaaag tattttatag ttttatcgat gataaaaatc   180
acaataaaaa actgctagtt attagaacaa aaggtaccat tgctggtcaa tatagagttt   240
atagcgaaga aggtgctaac aaaagtggtt tagcctggcc ttcagccttt aaggtacagt   300
tgcaactacc tgataatgaa gtagctcaaa tatctgatta ctatccgcgg aattcgattg   360
atacaaaaga gtatcacagt acgttaacgt acggattcaa cggtaacctt actggtgatg   420
atactagtaa aattggaggc cttattgggg cccaggtttc cctaggtcat acacttaagt   480
atgttcaacc tgatttcaaa acaattctcg agagcccaac tgataaaaaa gtaggctgga   540
aagtgatatt taacaatatg gtgaatcaaa attggggacc atacgatcga gattcttgga   600
acccggtata tggcaatcaa cttttcatga agactagaaa tggttctatg aaagcagcag   660
ataacttcct tgatcctaac aaagcaagtt ccctattatc ttcagggttt tcaccagact   720
tcgctacagt tattactatg gatagaaaag catccaaaca acaaacaaat atagatgtaa   780
tatacgaacg agttcgtgat gattaccaat tgcattggac ttcaccaaat tggaaaggta   840
ccaatactaa agataaatgg acagatcgtt cttcagaaag atataaaatc gattgggaaa   900
aagaagaaat gacaaattaa tgtaanttat ttgtacatgt acaaataaat ataatttata   960
actttagccg aagctggatc cggctgctac naancccnaa ngnagctgan ttgnctgctg  1020
ccccccctgac natactagca naccccttgg gnccctaacg ggtctgnggg gtttttgctg  1080
aangngnact tttccgnnan tcnnccccgn ccccccnggt gaaatccnaa nccccnaacn  1140
ggngntgnta ncaantttan tggnncntna ntttnnaaan cnnntaantt ngnaancccc  1200
nttttncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn  1260
annnccnttn aangggcccc cnnggggnaa nnntnggggn                        1300

QILIXTXATG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT    60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYHSTLTYGF   120
NGNLTGDDTS KIGGLIGAQV SLGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG   180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK   240
QQTNIDVIYE RVRDDYQLHW TSPNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN          293
```

Table 9 – SEQ ID NOs: 5 and 6 - α-hemolysin M113K-RL2
SEQ ID NO: 6 is encoded by residues 39 to 920 of SEQ ID NO: 5
n = a, c, g or t
X at location 5 stands for Leu or Phe
X at location 7 stands for Tyr, Trp, Cys, Ser, Leu or Phe

```
gttctgttta actttaagaa gggagatata catatgagca gattctgata ttnacntnng      60
cgaccggtac tacagatatt ggaagcaata ctacagtaaa aacaggtgat ttagtcactt     120
atgataaaga aaatggcatg cacaaaaaag tattttatag ttttatcgat gataaaaatc     180
acaataaaaa actgctagtt attagaacaa aaggtaccat tgctggtcaa tatagagttt     240
atagcgaaga aggtgctaac aaaagtggtt tagcctggcc ttcagccttt aaggtacagt     300
tgcaactacc tgataatgaa gtagctcaaa tatctgatta ctatccgcgg aattcgattg     360
atacaaaaga gtataaaagt acgttaacgt acggattcaa cggtaacctt actggtgatg     420
atactagtaa aattggaggc cttattgggg cccaggtttc cctaggtcat acacttaagt     480
atgttcaacc tgatttcaaa acaattctcg agagcccaac tgataaaaaa gtaggctgga     540
aagtgatatt taacaatatg gtgaatcaaa attggggacc atacgatcga gattcttgga     600
acccggtata tggcaatcaa cttttcatga agactagaaa tggttctatg aaagcagcag     660
ataacttcct tgatcctaac aaagcaagtt ccctattatc ttcagggttt tcaccagact     720
tcgctacagt tattactatg gatagaaaag catccaaaca acaaacaaat atagatgtaa     780
tatacgaacg agttcgtgat gattaccaat tgcattggac ttcaccaaat tggaaaggta     840
ccaatactaa agataaatgg acagatcgtt cttcagaaag atataaaatc gattgggaaa     900
aagaagaaat gacaaattaa tgtaanttat ttgtacatgt acaaataaat ataatttata     960
actttagccg aagctggatc cggctgctac naancccnaa ngnagctgan ttgnctgctg    1020
ccccccctgac natactagca naccccttgg gnccctaacg ggtctgnggg gtttttgctg    1080
aangngnact tttccgnnan tcnnccggn ccccccnggt gaaatccnaa nccccnaacn    1140
ggngntgnta ncaantttan tggnncntna ntttnnaaan cnnntaantt ngnaancccc    1200
nttttncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn    1260
annnccnttn aangggcccc cnnggggnaa nnntnggggn                          1300

QILIXTXATG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT      60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYKSTLTYGF     120
NGNLTGDDTS KIGGLIGAQV SLGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG     180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK     240
QQTNIDVIYE RVRDDYQLHW TSPNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN            293
```

Table 10 – SEQ ID NOs: 7 and 8 - α-hemolysin M113R-RL2
SEQ ID NO: 8 is encoded by residues 39 to 920 of SEQ ID NO: 7
n = a, c, g or t
X at location 5 stands for Leu or Phe
X at location 7 stands for Tyr, Trp, Cys, Ser, Leu or Phe

```
gttctgttta actttaagaa gggagatata catatgagca gattctgata ttnacntnng      60
cgaccggtac tacagatatt ggaagcaata ctacagtaaa aacaggtgat ttagtcactt     120
atgataaaga aaatggcatg cacaaaaaag tattttatag ttttatcgat gataaaaatc     180
acaataaaaa actgctagtt attagaacaa aaggtaccat tgctggtcaa tatagagttt     240
atagcgaaga aggtgctaac aaaagtggtt tagcctggcc ttcagccttt aaggtacagt     300
tgcaactacc tgataatgaa gtagctcaaa tatctgatta ctatccgcgg aattcgattg     360
atacaaaaga gtatagaagt acgttaacgt acggattcaa cggtaacctt actggtgatg     420
atactagtaa aattggaggc cttattgggg cccaggtttc cctaggtcat acacttaagt     480
atgttcaacc tgatttcaaa acaattctcg agagcccaac tgataaaaaa gtaggctgga     540
aagtgatatt taacaatatg gtgaatcaaa attggggacc atacgatcga gattcttgga     600
acccggtata tggcaatcaa cttttcatga agactagaaa tggttctatg aaagcagcag     660
ataacttcct tgatcctaac aaagcaagtt ccctattatc ttcagggttt tcaccagact     720
tcgctacagt tattactatg gatagaaaag catccaaaca acaaacaaat atagatgtaa     780
tatacgaacg agttcgtgat gattaccaat tgcattggac ttcaccaaat tggaaaggta     840
ccaatactaa agataaatgg acagatcgtt cttcagaaag atataaaatc gattgggaaa     900
aagaagaaat gacaaattaa tgtaanttat ttgtacatgt acaaataaat ataatttata     960
actttagccg aagctggatc cggctgctac naancccnaa ngnagctgan ttgnctgctg    1020
ccccccctgac natactagca naccccttgg gnccctaacg ggtctgnggg gtttttgctg    1080
aangngnact tttccgnnan tcnncccggn ccccccnggt gaaatccnaa nccccnaacn    1140
ggngntgnta ncaantttan tggnncntna ntttnnaaan cnnntaantt ngnaancccc    1200
nttttncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn    1260
annnccnttn aangggcccc cnnggggnaa nnntnggggn                          1300

QILIXTXATG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT      60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYRSTLTYGF     120
NGNLTGDDTS KIGGLIGAQV SLGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG     180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK     240
QQTNIDVIYE RVRDDYQLHW TSPNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN            293
```

Table 11 – SEQ ID NOs: 9 and 10 - α-hemolysin M113F-RL2
SEQ ID NO: 10 is encoded by residues 39 to 920 of SEQ ID NO: 9
n = a, c, g or t
X at location 5 stands for Leu or Phe
X at location 7 stands for Tyr, Trp, Cys, Ser, Leu or Phe

```
gttctgttta actttaagaa gggagatata catatgagca gattctgata ttnacntnng      60
cgaccggtac tacagatatt ggaagcaata ctacagtaaa aacaggtgat ttagtcactt     120
atgataaaga aaatggcatg cacaaaaaag tattttatag ttttatcgat gataaaaatc     180
acaataaaaa actgctagtt attagaacaa aaggtaccat tgctggtcaa tatagagttt     240
atagcgaaga aggtgctaac aaaagtggtt tagcctggcc ttcagccttt aaggtacagt     300
tgcaactacc tgataatgaa gtagctcaaa tatctgatta ctatccgcgg aattcgattg     360
atacaaaaga gtatttcagt acgttaacgt acggattcaa cggtaacctt actggtgatg     420
atactagtaa aattggaggc cttattgggg cccaggtttc cctaggtcat acacttaagt     480
atgttcaacc tgatttcaaa acaattctcg agagcccaac tgataaaaaa gtaggctgga     540
aagtgatatt taacaatatg gtgaatcaaa attggggacc atacgatcga gattcttgga     600
acccggtata tggcaatcaa cttttcatga agactagaaa tggttctatg aaagcagcag     660
ataacttcct tgatcctaac aaagcaagtt ccctattatc ttcagggttt tcaccagact     720
tcgctacagt tattactatg gatagaaaag catccaaaca acaaacaaat atagatgtaa     780
tatacgaacg agttcgtgat gattaccaat tgcattggac ttcaccaaat tggaaaggta     840
ccaatactaa agataaatgg acagatcgtt cttcagaaag atataaaatc gattgggaaa     900
aagaagaaat gacaaattaa tgtaanttat ttgtacatgt acaaataaat ataatttata     960
actttagccg aagctggatc cggctgctac naancccnaa ngnagctgan ttgnctgctg    1020
ccccccctgac natactagca naccccttgg gnccctaacg ggtctgnggg gtttttgctg    1080
aangngnact tttccgnnan tcnncccggn ccccccnggt gaaatccnaa nccccnaacn    1140
ggngntgnta ncaantttan tggnncntna ntttnnaaan cnnntaantt ngnaancccc    1200
nttttncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn    1260
annnccnttn aangggcccc cnnggggnaa nnntnggggn agatataaaa tcgattggga    1320
aaaagaagaa atgacaaatt aa                                             1342

QILIXTXATG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT      60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYFSTLTYGF     120
NGNLTGDDTS KIGGLIGAQV SLGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG     180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK     240
QQTNIDVIYE RVRDDYQLHW TSPNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN            293
```

Table 12 – SEQ ID NOs: 11 and 12 - α-hemolysin M113N-RL2
SEQ ID NO: 12 is encoded by residues 39 to 920 of SEQ ID NO: 11
n = a, c, g or t
X at location 5 stands for Leu or Phe
X at location 7 stands for Tyr, Trp, Cys, Ser, Leu or Phe

```
gttctgttta actttaagaa gggagatata catatgagca gattctgata ttnacntnng    60
cgaccggtac tacagatatt ggaagcaata ctacagtaaa aacaggtgat ttagtcactt   120
atgataaaga aaatggcatg cacaaaaaag tattttatag ttttatcgat gataaaaatc   180
acaataaaaa actgctagtt attagaacaa aaggtaccat tgctggtcaa tatagagttt   240
atagcgaaga aggtgctaac aaaagtggtt tagcctggcc ttcagccttt aaggtacagt   300
tgcaactacc tgataatgaa gtagctcaaa tatctgatta ctatccgcgg aattcgattg   360
atacaaaaga gtataacagt acgttaacgt acggattcaa cggtaacctt actggtgatg   420
atactagtaa aattggaggc cttattgggg cccaggtttc cctaggtcat acacttaagt   480
atgttcaacc tgatttcaaa acaattctcg agagcccaac tgataaaaaa gtaggctgga   540
aagtgatatt taacaatatg gtgaatcaaa attggggacc atacgatcga gattcttgga   600
acccggtata tggcaatcaa cttttcatga agactagaaa tggttctatg aaagcagcag   660
ataacttcct tgatcctaac aaagcaagtt ccctattatc ttcagggttt tcaccagact   720
tcgctacagt tattactatg gatagaaaag catccaaaca acaaacaaat atagatgtaa   780
tatacgaacg agttcgtgat gattaccaat tgcattggac ttcaccaaat tggaaaggta   840
ccaatactaa agataaatgg acagatcgtt cttcagaaag atataaaatc gattgggaaa   900
aagaagaaat gacaaattaa tgtaanttat ttgtacatgt acaaataaat ataatttata   960
actttagccg aagctggatc cggctgctac naancccnaa ngnagctgan ttgnctgctg  1020
ccccccctgac natactagca naccccttgg gnccctaacg ggtctgnggg gtttttgctg  1080
aangngnact tttccgnnan tcnncccggn ccccccnggt gaaatccnaa ncccccnaacn  1140
ggngntgnta ncaantttan tggnncntna ntttnnaaan cnnntaantt ngnaancccc  1200
nttttncnan ggcnaannnn nancctttna naaaaaancc nnnggggggg tttcnntnnn  1260
annnccnttn aangggcccc cnnggggnaa nnntnggggn                        1300

QILIXTXATG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT    60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYNSTLTYGF   120
NGNLTGDDTS KIGGLIGAQV SLGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG   180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK   240
QQTNIDVIYE RVRDDYQLHW TSPNWKGTNT KDKWTDRSSE RYKIDWEKEE MTN          293
```

Table 13 – SEQ ID NOs: 13 and 14 - α-hemolysin M113C-D8RL2
SEQ ID NO: 14 is encoded by residues 39 to 944 of SEQ ID NO: 13
n = a, c, g or t
X at location 5 stands for Leu or Phe
X at location 7 stands for Tyr, Trp, Cys, Ser, Leu or Phe

```
gttctgttta actttaagaa gggagatata catatgagca gattctgata ttnacntnng      60
cgaccggtac tacagatatt ggaagcaata ctacagtaaa aacaggtgat ttagtcactt     120
atgataaaga aaatggcatg cacaaaaaag tattttatag ttttatcgat gataaaaatc     180
acaataaaaa actgctagtt attagaacaa aaggtaccat tgctggtcaa tatagagttt     240
atagcgaaga aggtgctaac aaaagtggtt tagcctggcc ttcagccttt aaggtacagt     300
tgcaactacc tgataatgaa gtagctcaaa tatctgatta ctatccgcgg aattcgattg     360
atacaaaaga gtattgcagt acgttaacgt acggattcaa cggtaacctt actggtgatg     420
atactagtaa aattggaggc cttattgggg cccaggtttc cctaggtcat acacttaagt     480
atgttcaacc tgatttcaaa acaattctcg agagcccaac tgataaaaaa gtaggctgga     540
aagtgatatt taacaatatg gtgaatcaaa attggggacc atacgatcga gattcttgga     600
acccggtata tggcaatcaa cttttcatga agactagaaa tggttctatg aaagcagcag     660
ataacttcct tgatcctaac aaagcaagtt ccctattatc ttcagggttt tcaccagact     720
tcgctacagt tattactatg gatagaaaag catccaaaca acaaacaaat atagatgtaa     780
tatacgaacg agttcgtgat gattaccaat tgcattggac ttcaccaaat tggaaaggta     840
ccaatactaa agataaatgg acagatcgtt cttcagaaag atataaaatc gattgggaaa     900
aagaagaaat gacaaatgat gacgatgatg acgacgatga ttaatgtaan ttatttgtac     960
atgtacaaat aaatataatt tataacttta gccgaagctg gatccggctg ctacnaancc    1020
cnaangnagc tganttgnct gctgcccccc tgacnatact agcanacccc ttgggnccct    1080
aacgggtctg nggggttttt gctgaangng nacttttccg nnantcnncc cggncccccc    1140
nggtgaaatc cnaanccccn aacnggngnt gntancaant ttantggnnc ntnantttnn    1200
aaancnnnta anttngnaan ccccntttтn cnanggcnaa nnnnnancct ttnanaaaaa    1260
anccnnnggg ggggtttcnn tnnn                                          1284

QILIXTXATG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT      60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYCSTLTYGF     120
NGNLTGDDTS KIGGLIGAQV SLGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG     180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK     240
QQTNIDVIYE RVRDDYQLHW TSPNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNDDDDDDD     300
D                                                                     301
```

Table 14 – SEQ ID NOs: 15 and 16 - α-hemolysin T117C-D8RL3
SEQ ID NO: 16 is encoded by residues 13 to 918 of SEQ ID NO: 15

```
gatatacata tggcagattc tgatattaat attaaaaccg gtactacaga tattggaagc   60
aatactacag taaaaacagg tgatttagtc acttatgata aagaaaatgg catgcacaaa  120
aaagtatttt atagttttat cgatgataaa aatcacaata aaaaactgct agttattaga  180
acaaaaggta ccattgctgg tcaatataga gtttatagcg aagaaggtgc taacaaaagt  240
ggtttagcct ggccttcagc ctttaaggta cagttgcaac tacctgataa tgaagtagct  300
caaatatctg attactatcc gcggaattcg attgatacaa aagagtatat gagtacgtta  360
tgctacggat tcaacggtaa tgttactggt gatgatacag gaaaaattgg aggccttatt  420
ggtgcaaatg tttcgattgg tcatacactt aagtatgttc aacctgattt caaaacaatt  480
ctcgagagcc caactgataa aaaagtaggc tggaaagtga tatttaacaa tatggtgaat  540
caaaattggg gaccatacga tcgagattct tggaacccgg tatatggcaa tcaacttttc  600
atgaaaacta gaaatggttc tatgaaagca gcagataact tccttgatcc taacaaagca  660
agttctctat tatcttcagg gttttcacca gacttcgcta cagttattac tatggataga  720
aaagcatcca aacaacaaac aaatatagat gtaatatacg aacgagttcg tgatgattac  780
caattgcatt ggacttcaac aaattggaaa ggtaccaata ctaaagataa atggacagat  840
cgttcttcag aaagatataa aatcgattgg gaaaaagaag aaatgacaaa tgatgacgat  900
gatgacgacg atgattgata agcttggatc cggctgc                           937

ADSDINIKTG TTDIGSNTTV KTGDLVTYDK ENGMHKKVFY SFIDDKNHNK KLLVIRTKGT   60
IAGQYRVYSE EGANKSGLAW PSAFKVQLQL PDNEVAQISD YYPRNSIDTK EYMSTLCYGF  120
NGNVTGDDTG KIGGLIGANV SIGHTLKYVQ PDFKTILESP TDKKVGWKVI FNNMVNQNWG  180
PYDRDSWNPV YGNQLFMKTR NGSMKAADNF LDPNKASSLL SSGFSPDFAT VITMDRKASK  240
QQTNIDVIYE RVRDDYQLHW TSTNWKGTNT KDKWTDRSSE RYKIDWEKEE MTNDDDDDDD  300
D                                                                   301
```

Table 15 – SEQ ID NO: 17 - lambda exonuclease

```
SHMTPDIILQ RTGIDVRAVE QGDDAWHKLR LGVITASEVH NVIAKPRSGK KWPDMKMSYF   60
HTLLAEVCTG VAPEVNAKAL AWGKQYENDA RTLFEFTSGV NVTESPIIYR DESMRTACSP  120
DGLCSDGNGL ELKCPFTSRD FMKFRLGGFE AIKSAYMAQV QYSMWVTRKN AWYFANYDPR  180
MKREGLHYVV IERDEKYMAS FDEIVPEFIE KMDEALAEIG FVFGEQWR               228
```

Table 16 – SEQ ID NO: 18 - α-hemolysin-D8RL3

```
gcagattctg atattaatat taaaaccggt actacagata ttggaagcaa tactacagta   60
aaaacaggtg atttagtcac ttatgataaa gaaaatggca tgcacaaaaa agtattttat  120
agttttatcg atgataaaaa tcacaataaa aaactgctag ttattagaac aaaaggtacc  180
attgctggtc aatatagagt ttatagcgaa gaaggtgcta acaaaagtgg tttagcctgg  240
ccttcagcct ttaaggtaca gttgcaacta cctgataatg aagtagctca aatatctgat  300
tactatccgc ggaattcgat tgatacaaaa gagtatatga gtacgttaac gtacggattc  360
aacggtaatg ttactggtga tgatacagga aaaattggag gccttattgg tgcaaatgtt  420
tcgattggtc atacacttaa gtatgttcaa cctgatttca aaacaattct cgagagccca  480
actgataaaa aagtaggctg gaaagtgata tttaacaata tggtgaatca aaattgggga  540
ccatacgatc gagattcttg gaacccggta tatggcaatc aacttttcat gaaaactaga  600
aatggttcta tgaaagcagc agataacttc cttgatccta acaaagcaag ttctctatta  660
tcttcagggt tttcaccaga cttcgctaca gttattacta tggatagaaa agcatccaaa  720
caacaaacaa atatagatgt aatatacgaa cgagttcgtg atgattacca attgcattgg  780
acttcaacaa attggaaagg taccaatact aaagataaat ggacagatcg ttcttcagaa  840
agatataaaa tcgattggga aaaagaagaa atgacaaatg atgacgatga tgacgacgat  900
gattgataag cttggatccg gctgc                                        925
```

CLAIMS

1. A transmembrane protein pore for use in detecting an analyte in a sample, which comprises a molecular adaptor that facilitates an interaction between the pore and the analyte, wherein the adaptor is covalently attached to the pore in an orientation that allows the analyte to be detected using the pore.

2. A transmembrane protein pore according to claim 1, wherein a portion of the adaptor that is capable of interacting with the analyte is oriented towards the end of the pore through which the analyte enters.

3. A transmembrane protein pore according to claim 2, wherein the portion comprises one or more chemical groups that are capable of interacting with the analyte.

4. A transmembrane protein pore according to claim 3, wherein the chemical groups are amino groups or hydroxyl groups.

5. A transmembrane protein pore according to any one of claims 2 to 4, wherein the end of the pore is the *cis* end or the *trans* end.

6. A transmembrane protein pore according to any one of the preceding claims, wherein the pore is modified to facilitate the covalent attachment of the adaptor.

7. A transmembrane protein pore according to any one of the preceding claims, wherein the pore is modified to facilitate the orientation of the adaptor.

8. A transmembrane protein pore according to any one of the preceding claims, wherein the pore comprises SEQ ID NO: 2 or a variant thereof.

9. A transmembrane protein pore according to any one of the preceding claims, wherein the molecular adaptor is a cyclodextrin.

10. A transmembrane protein pore according to claim 9, wherein the cyclodextrin is heptakis-6-amino-β-cyclodextrin (am_7-β-CD) or 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1βCD).

11. A transmembrane pore according to any one of the preceding claims, wherein the analyte is an individual nucleotide.

12. A transmembrane protein pore according to any one of the preceding claims, wherein the adaptor is covalently attached to the pore via a bifunctional crosslinker.

13. A method of producing a transmembrane protein pore according to any one of claims 1 to 12, comprising:

(a) covalently attaching to a transmembrane protein pore a molecular adaptor that facilitates an interaction between the pore and an analyte; and

(b) determining whether or not the adaptor is attached to the pore in an orientation that allows the analyte to be detected using the pore.

14. A method according to claim 13, further comprising modifying the pore to facilitate the covalent attachment and/or to facilitate the orientation of the adaptor.

15. A method of determining presence or absence of an analyte, comprising:

(a) contacting the analyte with a transmembrane protein pore according to any one of claims 1 to 12 so that the analyte interacts with the pore; and

(b) measuring the current passing through the pore during the interaction and thereby determining the presence or absence of the analyte.

16. A method of identifying an individual nucleotide, comprising:

(a) contacting the nucleotide with a transmembrane protein pore according to any one of claims 1 to 12 so that the nucleotide interacts with the pore; and

(b) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide.

17. A method of sequencing a target nucleic acid sequence, comprising:

(a) digesting an individual nucleotide from one end of the target sequence using a processive exonuclease;

(b) contacting the nucleotide with a transmembrane protein pore according to any one of claims 1 to 12 so that the nucleotide interacts with the adaptor;

(c) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide; and

(d) repeating steps (a) to (c) at the same end of the nucleic acid sequence and thereby determining the sequence of the nucleic acid.

18. A kit for sequencing a nucleic acid, comprising a transmembrane protein pore according to any one of claims 1 to 12 and a processive exonuclease.

Figures

Figure 1

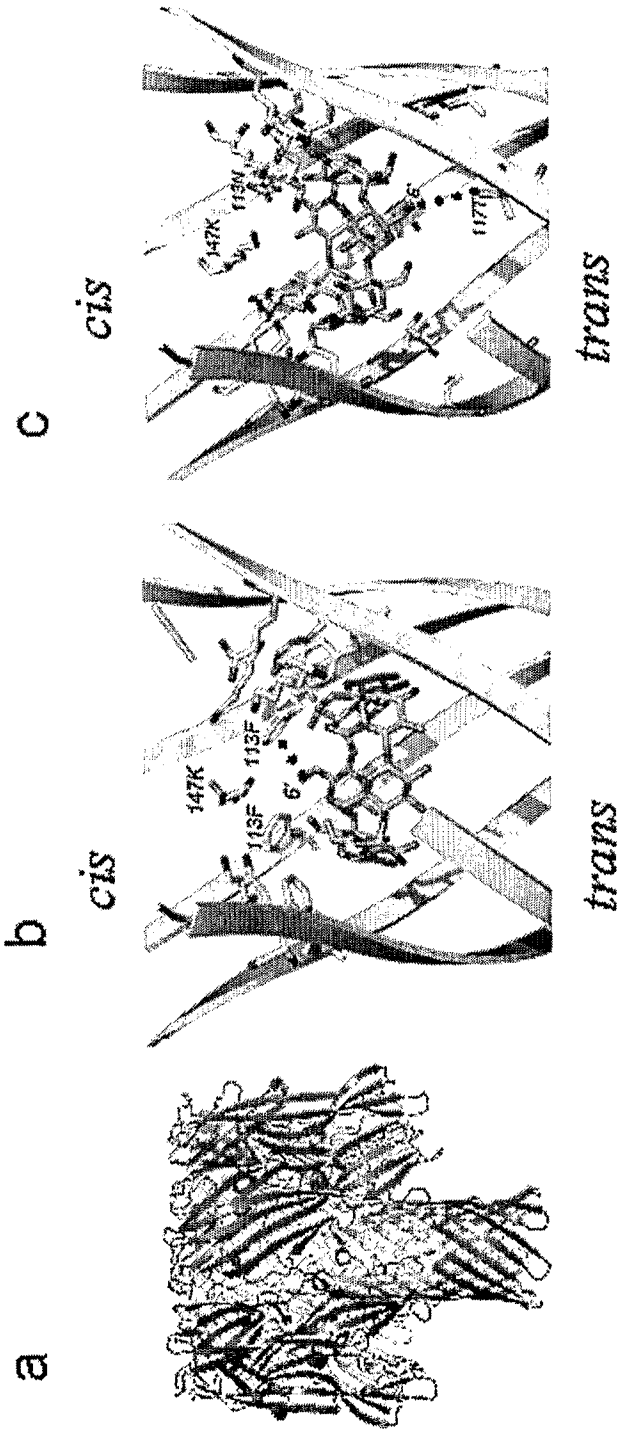


Figure 2

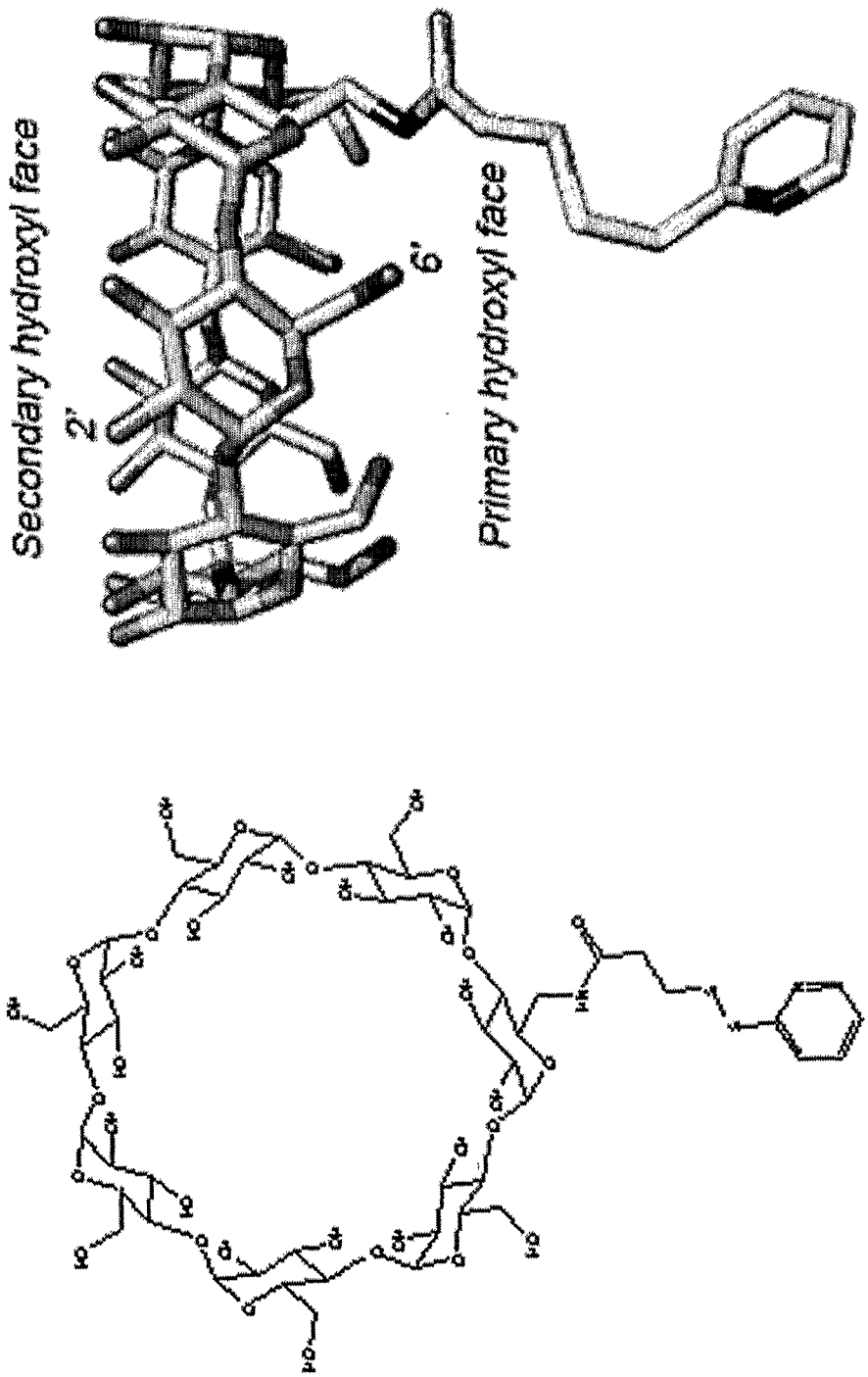


Figure 3

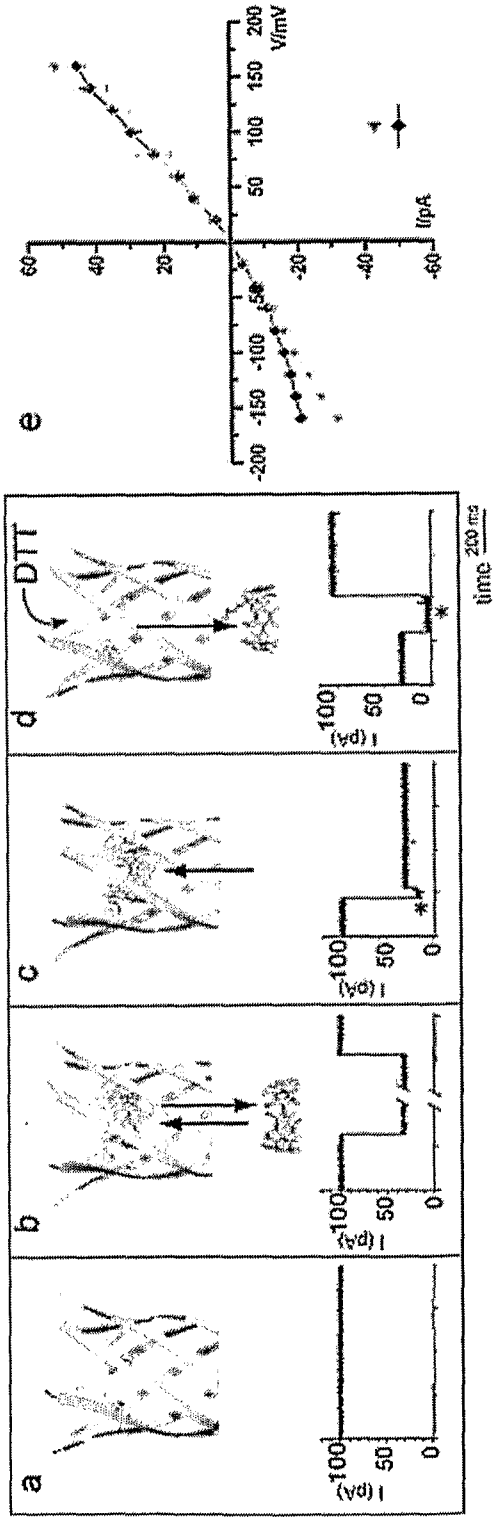
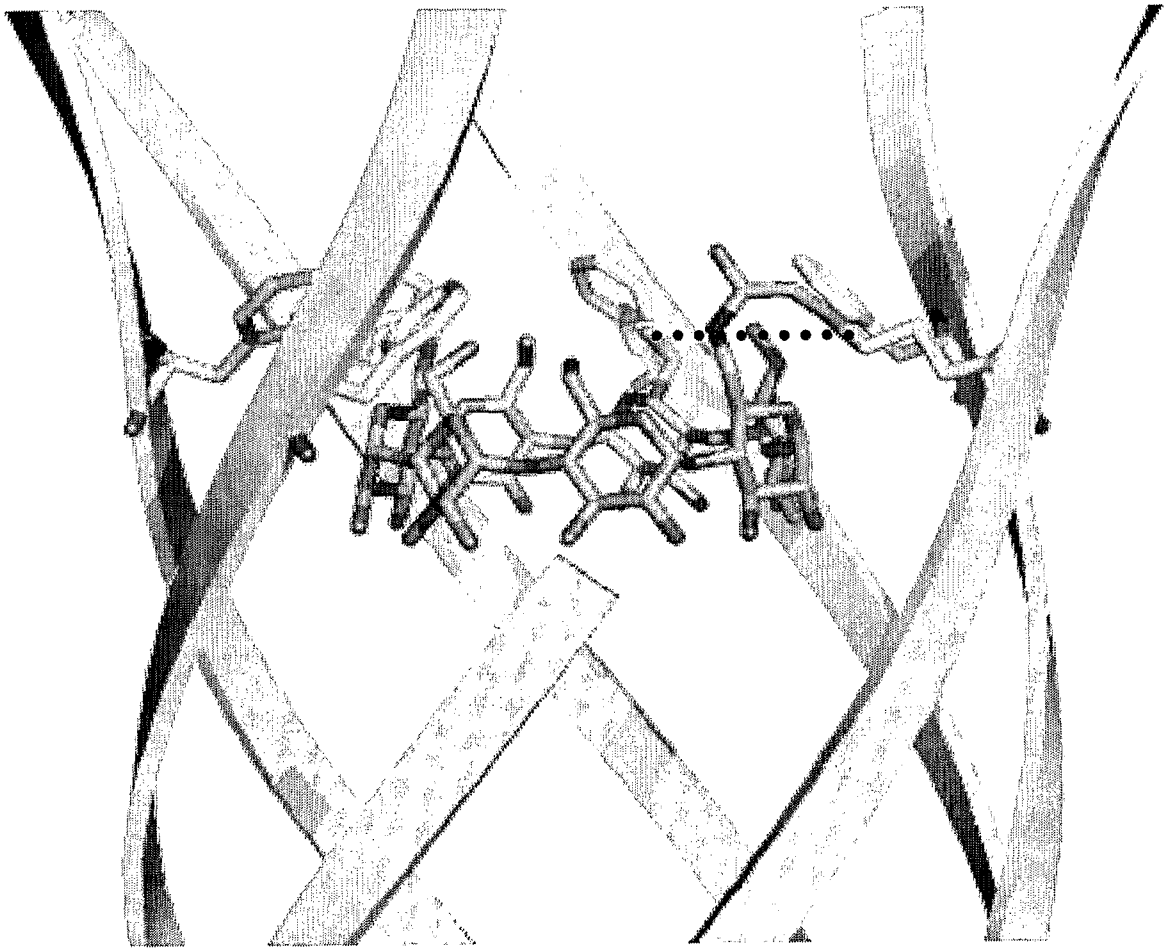

e
V/mV
200
150
100
50
-50
-100
-150
-200
60
40
20
-20
-40
-60
I/pA
d
DTT
c
b
a
I (pA)
100
50
0
time 200 ms

Figure 4

cis



trans

Figure 5

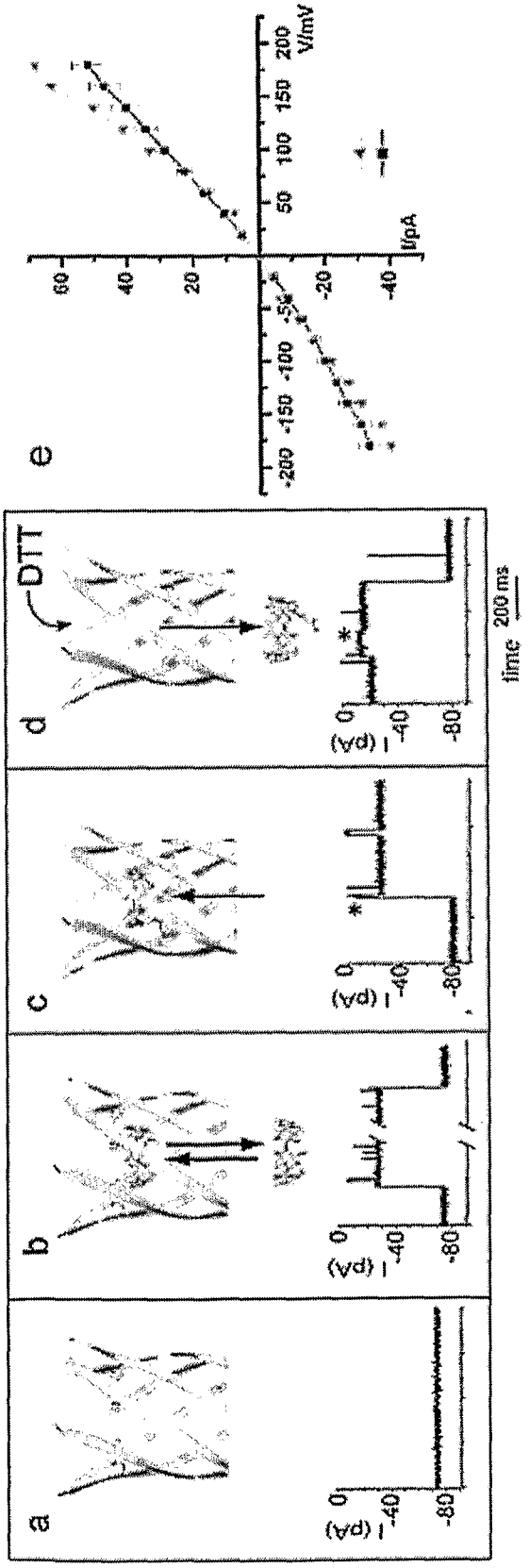
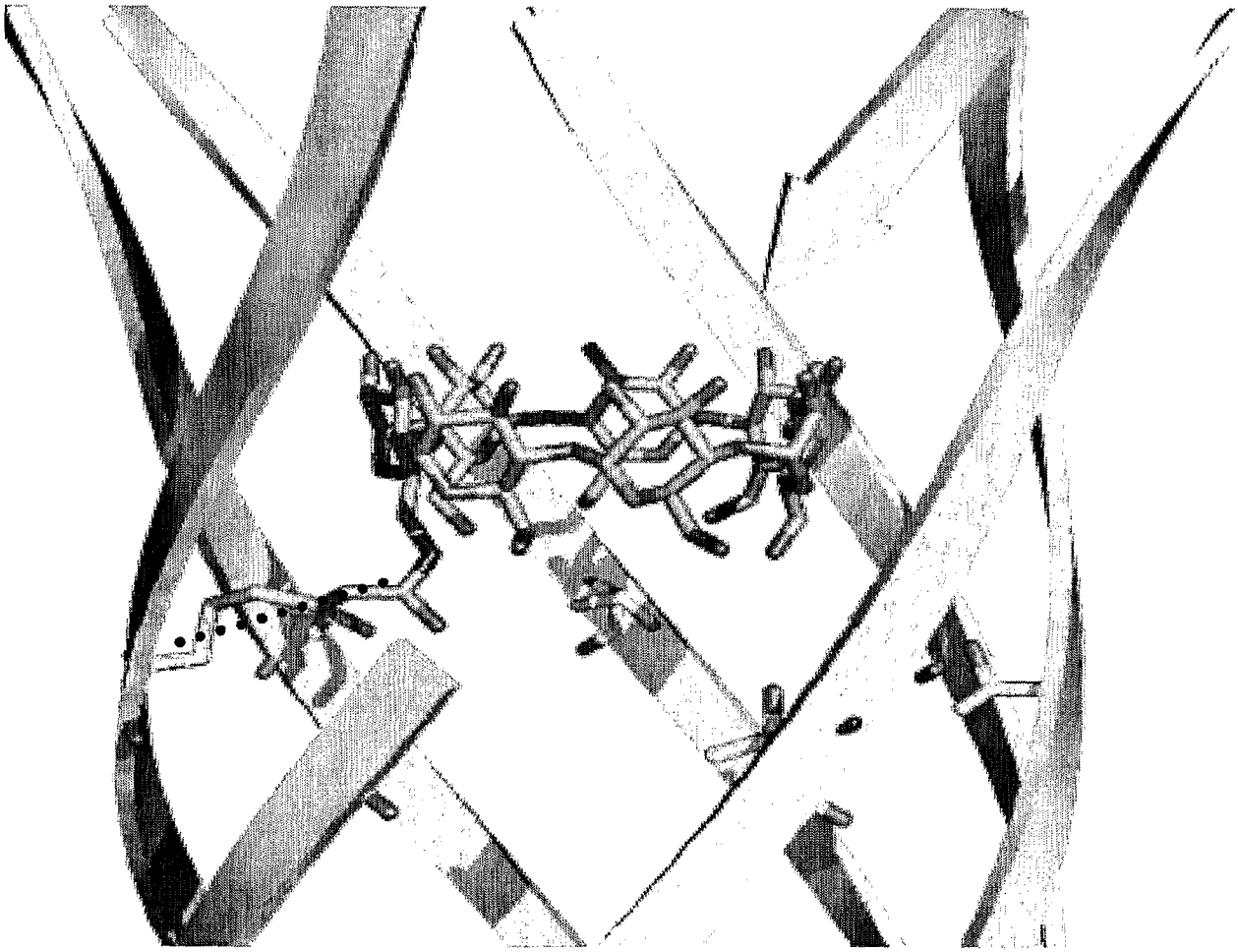

a
b
c
d
e
DTT
I (pA)
0
-40
-80
time 200 ms
V/mV
I/pA
200
150
100
50
-50
-100
-150
-200
60
40
20
-20
-40

Figure 6

cis



trans

Figure 7

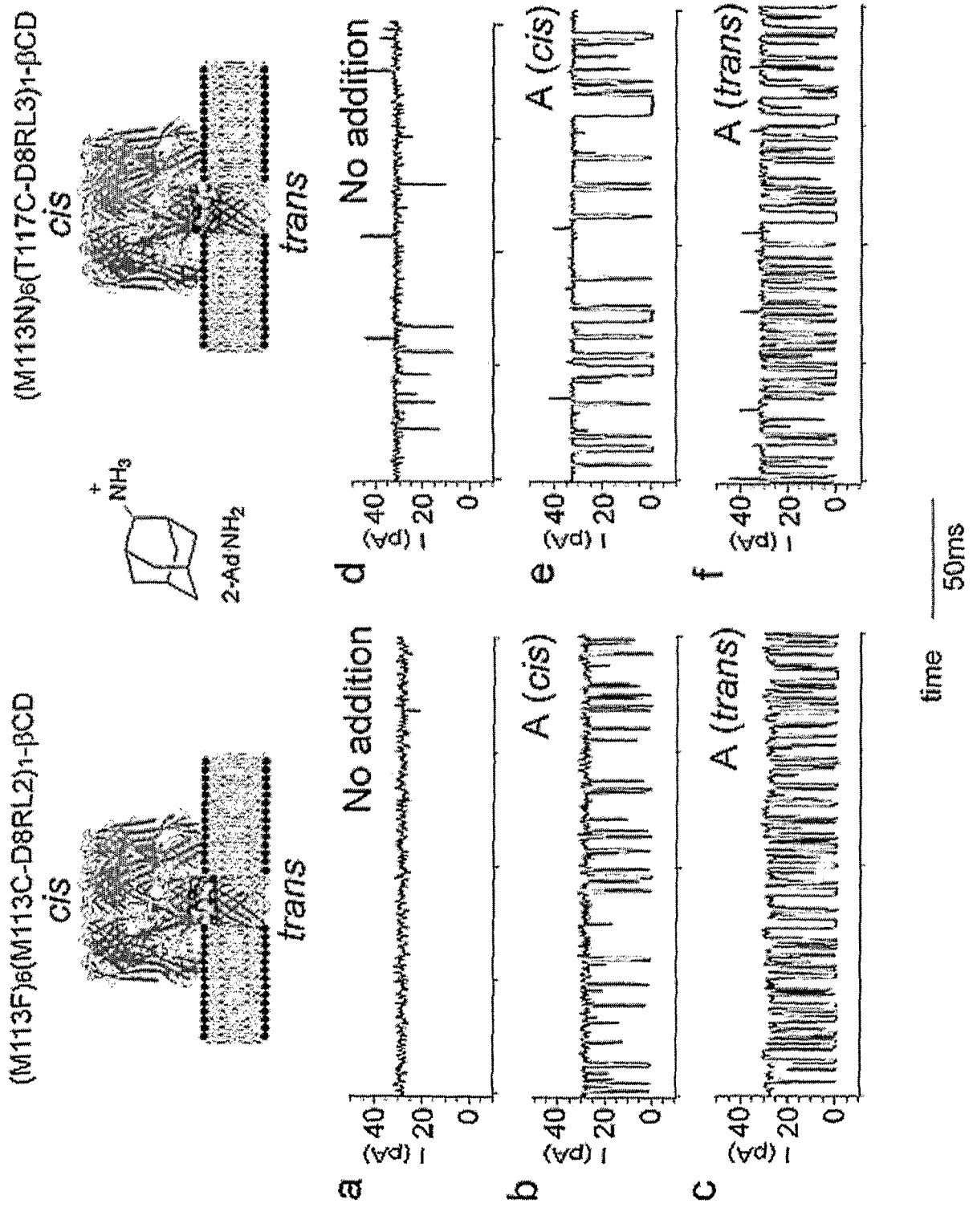

(M113F)6(M113C-D8RL2)1-βCD
cis
trans
(M113N)6(T117C-D8RL3)1-βCD
cis
trans
2-Ad NH2
No addition
A (cis)
A (trans)
a
b
c
d
e
f
I (pA)
40
20
0
time
50ms

Figure 8

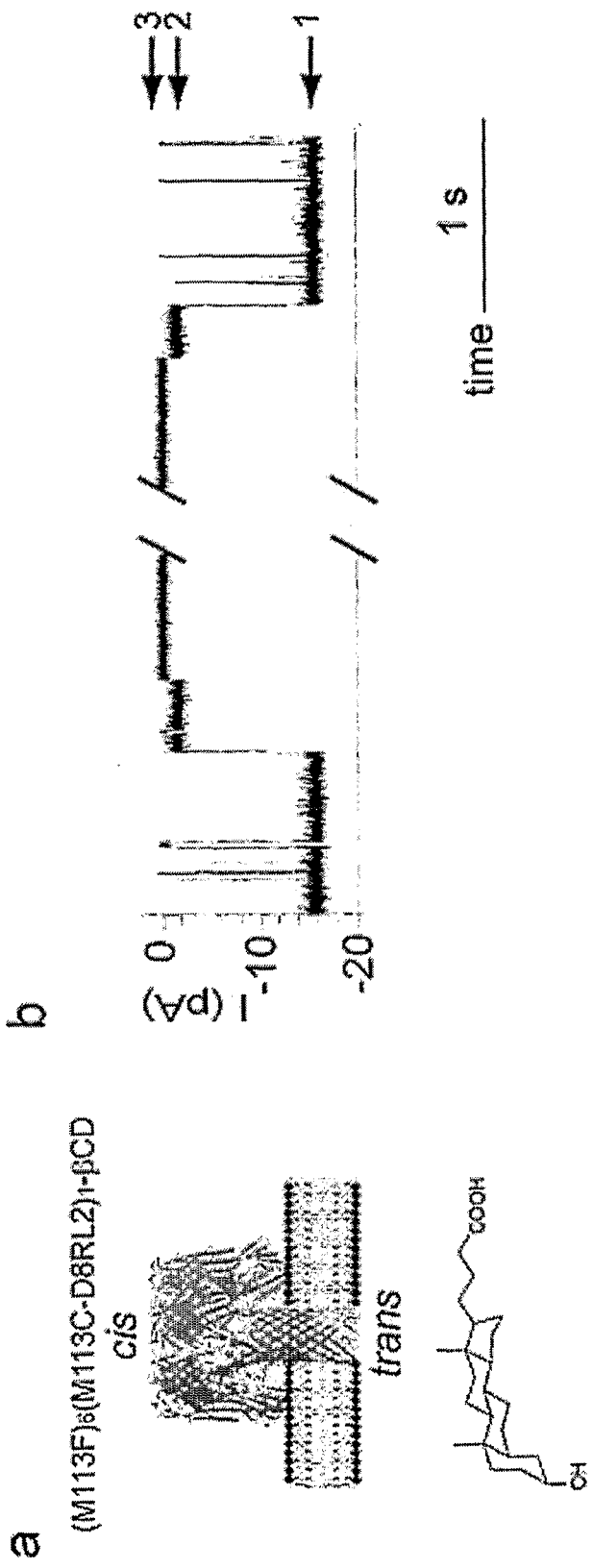

a
(M113F)6(M113C-D8RL2)1-βCD
cis
trans
COOH
δ
b
3
2
1
I (pA)
0
-10
-20
time
1 s

Figure 9

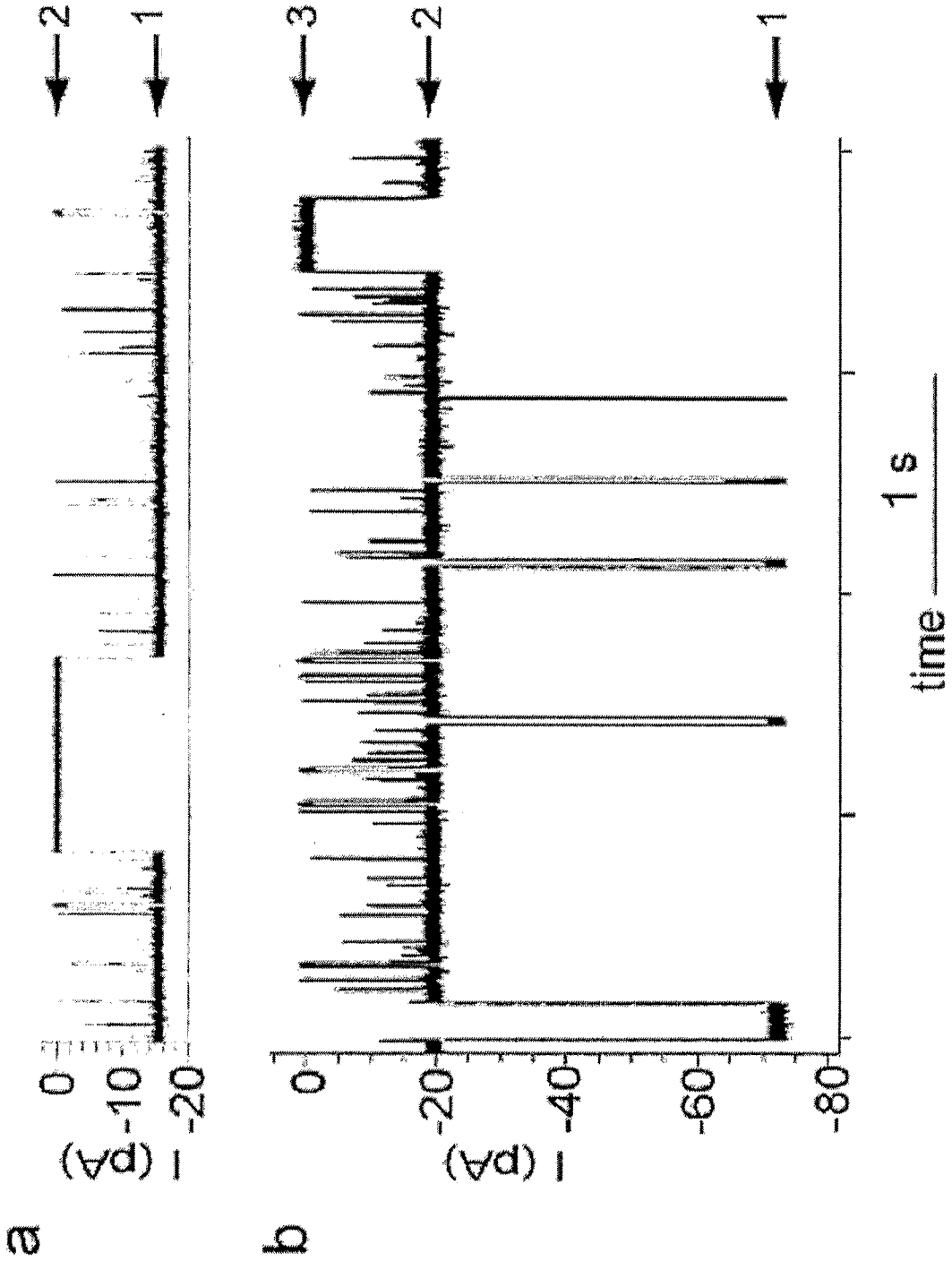


Figure 10

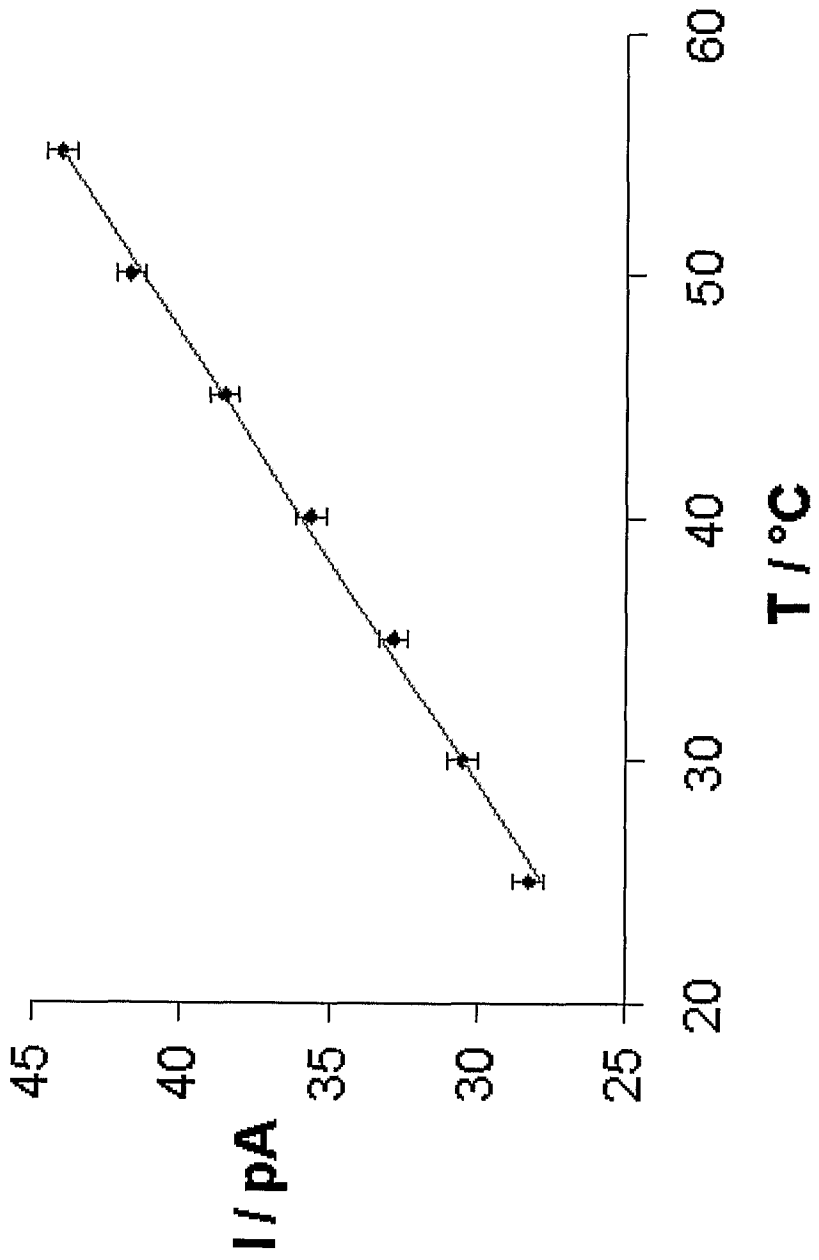


Figure 11a

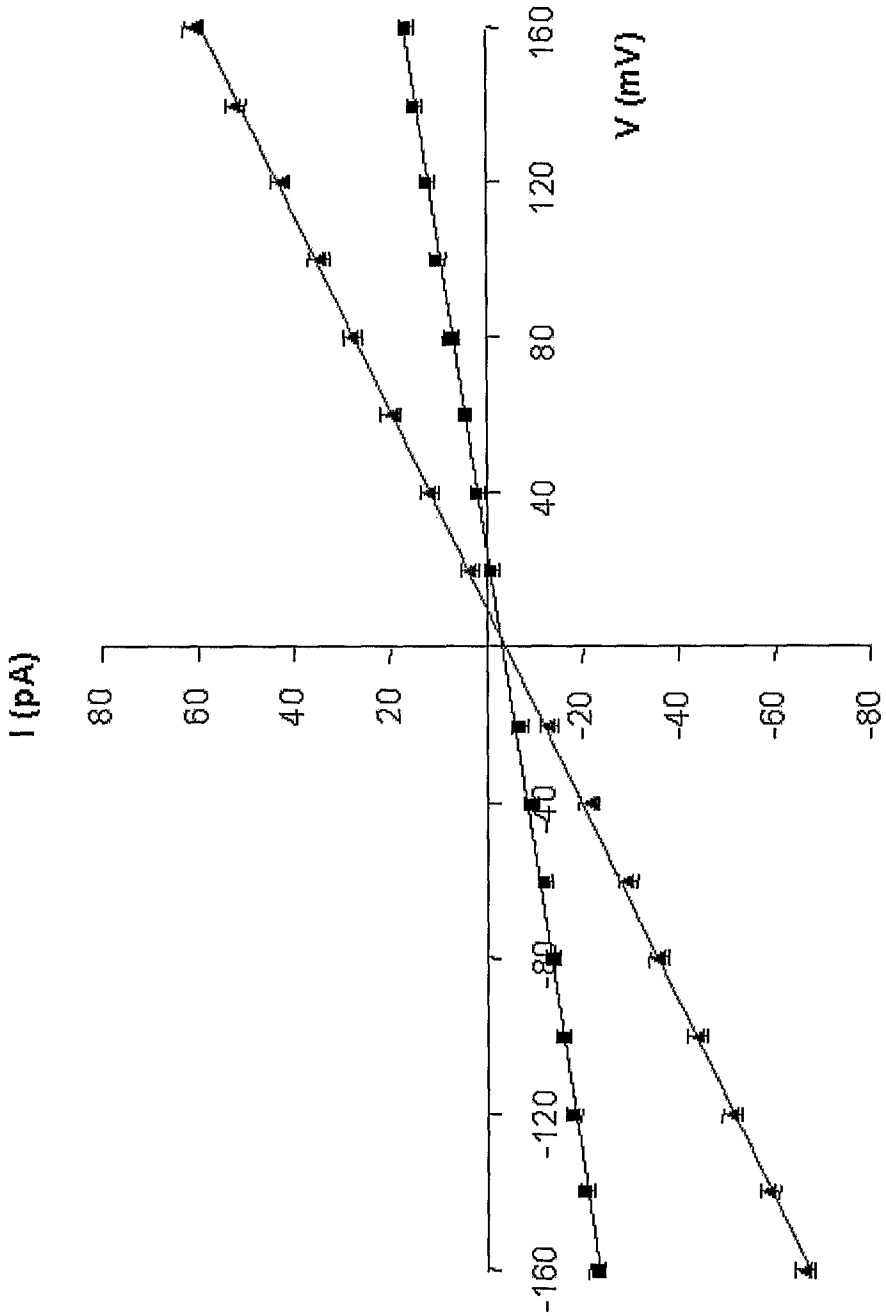


Figure 11b

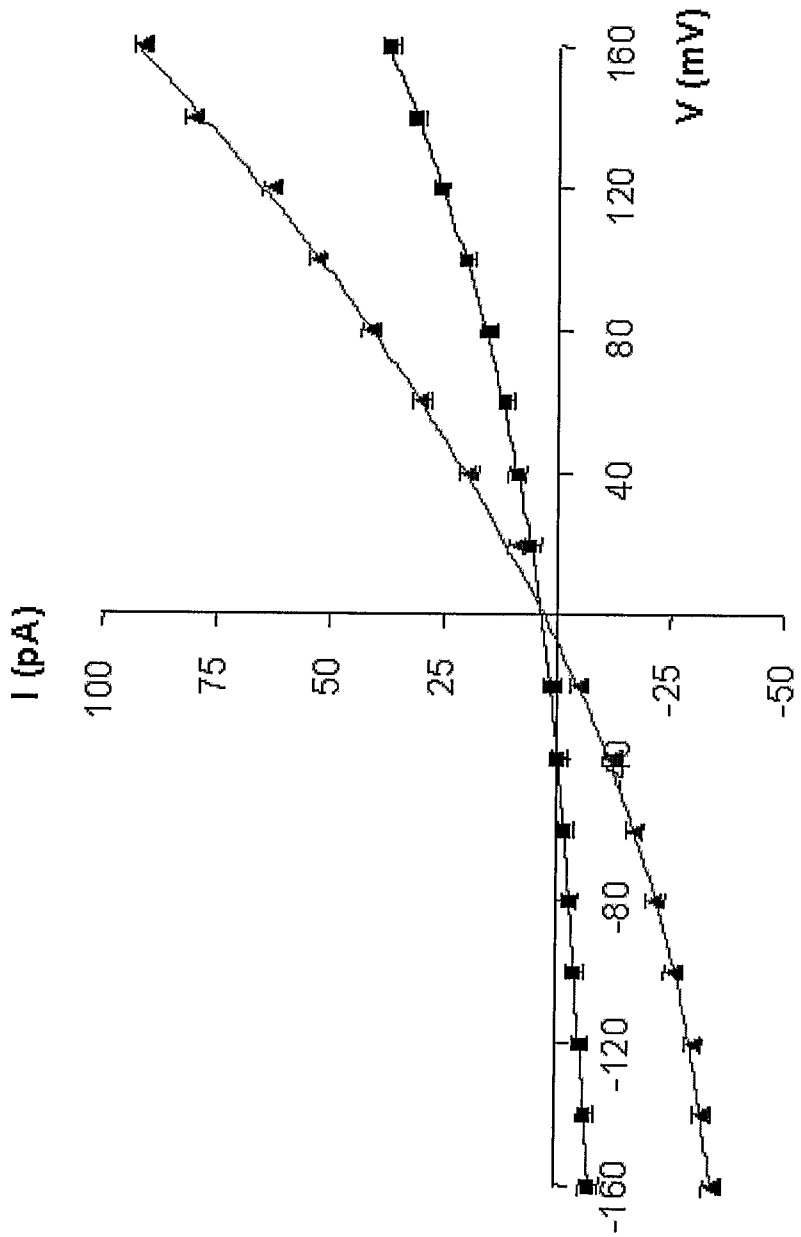
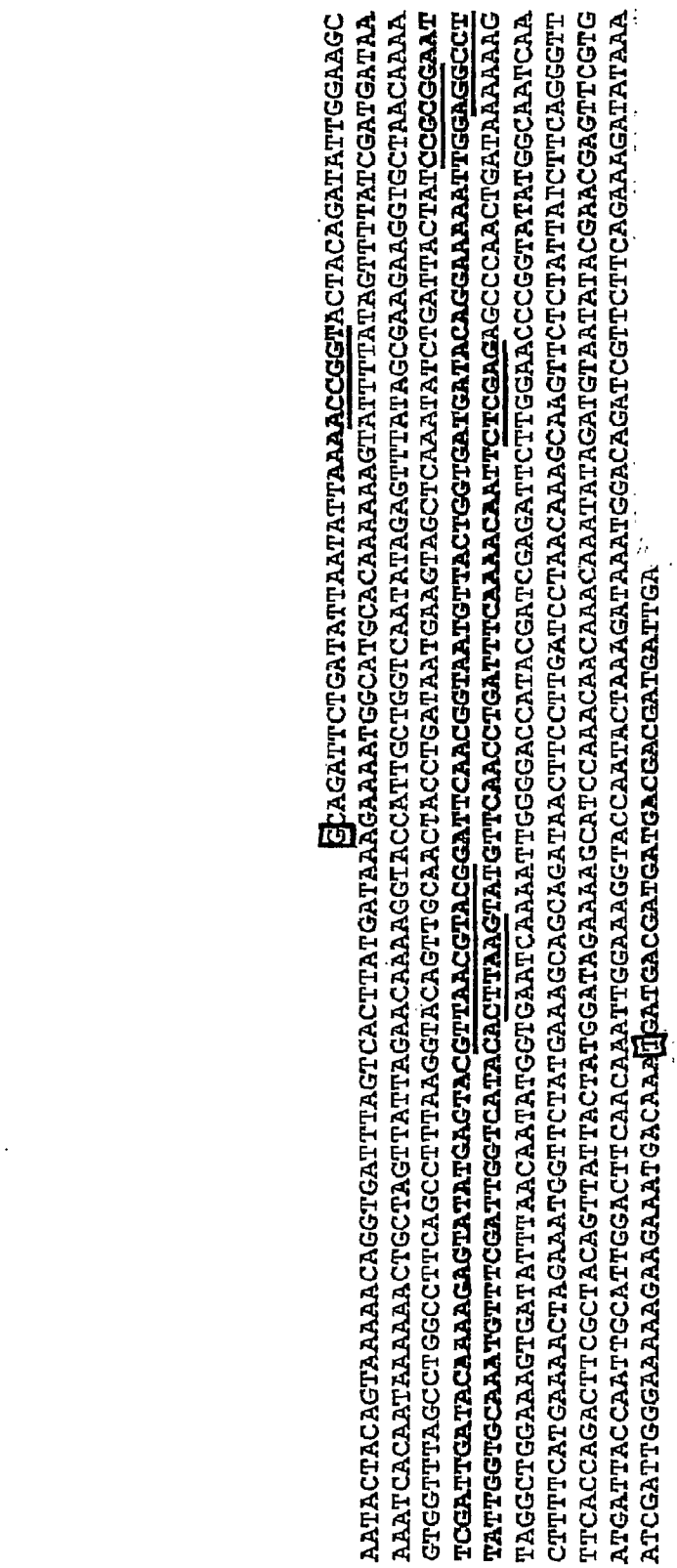


Figure 12


GCAGATTCTGATATTAATATTAAAACCGGTACTACAGATATTGGAAGC
AATACTACAGTAAAAACAGGTGATTTAGTCACTTATGATAAGAAAATGGCATGCACAAAAAAGTATTTTATAGTTTTATCGATGATAA
AAATCACAATAAAAAACTGCTAGTTATTAGAACAAAAGGTACCATTGCTGGTCAATATAGAGTTTATAGCGAAGAGGTGCTAACAAAA
GTGGTTTAGCCTGGCCTTCAGCCTTTAAGGTACAGTTGCAACTACCTGATAATGAAGTAGCTCAAATATCTGATTACTATCTATCCGCGGAAT
TCGATTGATACAAAAGAGTATATGAGTACGTTAACGTTAACGGATTCAACGGTAATGTTACTGGTGATGATACAGGAAAAATTGGAGGCCT
TATTGGTGCAAATGTTTTCGATTGGTCATACACTTAAGTATGTTCAACCTGATTTCAAACAATTCTCGAGAGCCCAACTGATAAAAAAG
TAGGCTGGAAAGTGATATTTAACAATATGGTGAATCAAAATTGGGACCATACGATCGAGATTCTTGGAACCCGGTATATGGCAATCAA
CTTTTCATGAAACTAGAAATGGTTCTATGAAAGCAGCAGATAACTTCCTTGATCCTAACAAAGCAAGTTCTCTATTATCTTCAGGGTT
TTCACCAGACTTCGCTACAGTTATTACTATGGATAGAAAAGCATCCAACAACAAACAAATATAGATGTAATATACGAACGAGTTCGTG
ATGATTACCAATTGCATTGGACTTCAACAAATTGGAAAGGTACCAATACTAAAGATAATGGACAGATCGTTCTTCAGAAAGATATAAA
ATCGATTGGGAAAAAGAAGAAATGACAAATGATGACGACGATGATTGA

INTERNATIONAL SEARCH REPORT

CORRECTED VERSION

International application No
PCT/GB2008/003372

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data, PAJ, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 01/42782 A (HARVARD COLLEGE [US]) 14 June 2001 (2001-06-14) page 5	1-18
X	WO 2007/057668 A (ISIS INNOVATION [GB]; BAYLEY HAGAN [GB]; ASTIER YANN [GB]; BRAHA ORIT) 24 May 2007 (2007-05-24) pages 15-16	1-18
X	US 6 426 231 B1 (BAYLEY HAGAN [US] ET AL) 30 July 2002 (2002-07-30) columns 9-10	1-18
X	WO 01/59453 A (TEXAS A & M UNIV SYS [US] TEXAS A & M UNIV SYS [US]; BAYLEY HAGAN P [U) 16 August 2001 (2001-08-16) page 2	1-18

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 February 2009	18. 02. 2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Lunter, Pim

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2003/099951 A1 (AKESON MARK [US] ET AL AKESON MARK [US] ET AL) 29 May 2003 (2003-05-29) paragraph [0161]	1-18
X	US 2002/094526 A1 (BAYLEY HAGAN P [US] ET AL) 18 July 2002 (2002-07-18) paragraph [0012]	1-18
X	WO 03/095669 A (TEXAS A & M UNIV SYS [US]) 20 November 2003 (2003-11-20) pages 13-14	1-18
X	HOWORKA STEFAN ET AL: "Sequence-specific detection of individual DNA strands using engineered nanopores" NATURE BIOTECHNOLOGY, vol. 19, no. 7, July 2001 (2001-07), pages 636-639, XP002510816 ISSN: 1087-0156 page 636 left column abstract	1-18
X	MOVILEANU L ET AL: "Detecting protein analytes that modulate transmembrane movement of a polymer chain within a single protein pore" NATURE BIOTECHNOLOGY, NATURE PUBLISHING GROUP, NEW YORK, NY, US, vol. 18, 1 September 2000 (2000-09-01), pages 1091-1095, XP002188230 ISSN: 1087-0156 page 1091 right column abstract	1-18
X	XIE H ET AL: "Single-Molecule Observation of the Catalytic Subunit of cAMP-Dependent Protein Kinase Binding to an Inhibitor Peptide" CHEMISTRY AND BIOLOGY, CURRENT BIOLOGY, LONDON, GB, vol. 12, no. 1, 1 January 2005 (2005-01-01), pages 109-120, XP004722469 ISSN: 1074-5521 pages 109-110	1-18

-/--

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HOWORKA STEFAN ET AL: "Kinetics of duplex formation for individual DNA strands within a single protein nanopore" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, vol. 98, no. 23, 2001, pages 12996-13001, XP002510817 ISSN: 0027-8424 abstract ----	1-18
X	HOWORKA STEFAN ET AL: "DNA duplex formation of individual DNA strands with a single protein pore" BIOPHYSICAL JOURNAL CONF- 46TH ANNUAL MEETING OF THE BIOPHYSICAL SOCIETY; SAN FRANCISCO, CALIFORNIA, USA; FEBRUARY 23-27, 2002,, vol. 82, no. 1 PART 2, 1 February 2002 (2002-02-01), page 508A, XP008100272 abstract ----	1-18

INTERNATIONAL SEARCH REPORT
CORRECTED VERSION

International application No
PCT/GB2008/003372

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0142782 A	14-06-2001	AU 2076401 A	18-06-2001
		EP 1238275 A1	11-09-2002
WO 2007057668 A	24-05-2007	AU 2006314296 A1	24-05-2007
		CA 2629009 A1	24-05-2007
		CN 101356288 A	28-01-2009
		EP 1951898 A1	06-08-2008
		KR 20080083281 A	17-09-2008
		US 2008311582 A1	18-12-2008
US 6426231 B1	30-07-2002	US 6927070 B1	09-08-2005
WO 0159453 A	16-08-2001	AU 4147401 A	20-08-2001
		EP 1255772 A2	13-11-2002
US 2003099951 A1	29-05-2003	NONE	
US 2002094526 A1	18-07-2002	US 2005208574 A1	22-09-2005
WO 03095669 A	20-11-2003	AU 2003245272 A1	11-11-2003
		EP 1504114 A1	09-02-2005