Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2629009 C 2016/02/02

(11)(21) **2 629 009**

(12) **BREVET CANADIEN**
CANADIAN PATENT

(13) **C**

(86) Date de dépôt PCT/PCT Filing Date: 2006/11/15
(87) Date publication PCT/PCT Publication Date: 2007/05/24
(45) Date de délivrance/Issue Date: 2016/02/02
(85) Entrée phase nationale/National Entry: 2008/05/08
(86) N° demande PCT/PCT Application No.: GB 2006/004265
(87) N° publication PCT/PCT Publication No.: 2007/057668
(30) Priorité/Priority: 2005/11/15 (GB0523282.2)

(51) Cl.Int./Int.Cl. *C12Q 1/68* (2006.01)

(72) Inventeurs/Inventors:
BAYLEY, HAGAN, GB;
ASTIER, YANN, GB;
BRAHA, ORIT, US

(73) Propriétaire/Owner:
ISIS INNOVATION LIMITED, GB

(74) Agent: OSLER, HOSKIN & HARCOURT LLP

(54) Titre : PROCEDES UTILISANT DES PORES
(54) Title: METHODS USING PORES

(57) Abrégé/Abstract:
The invention relates to a method of identifying an individual nucleotide, comprising (a) contacting the nucleotide with a transmembrane protein pore 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. The invention also relates to a method of sequencing nucleic acid sequences and kits related thereto.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 May 2007 (24.05.2007)

PCT

(10) International Publication Number
WO 2007/057668 A1

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

(21) International Application Number:
PCT/GB2006/004265

(22) International Filing Date:
15 November 2006 (15.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0523282.2 15 November 2005 (15.11.2005) GB

(71) Applicant *(for all designated States except US)*: **ISIS INNOVATION LIMITED** [GB/GB]; Ewert House, Ewert Place, Summertown, Oxford OX2 7SG (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BAYLEY, Hagan** [GB/GB]; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA (GB). **ASTIER, Yann** [FR/GB]; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA (GB). **BRAHA, Orit**; University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford OX1 3TA (US).

(74) Agents: **ROQUES, Sarah, Elizabeth** 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, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, LV, LY, MA, MD, 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, 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, HU, IE, IS, IT, LT, LU, LV, MC, NL, 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*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2007/057668 A1

(54) **Title:** METHODS USING PORES

(57) **Abstract:** The invention relates to a method of identifying an individual nucleotide, comprising (a) contacting the nucleotide with a transmembrane protein pore 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. The invention also relates to a method of sequencing nucleic acid sequences and kits related thereto.

METHODS USING PORES

Field of the invention

The invention relates to the identification of individual nucleotides and other phosphate containing moieties using transmembrane pores. In particular, the invention relates to the sequencing of target nucleic acids using transmembrane pores.

Background of the invention

The current method for sequencing DNA involves a number of costly reagents such as fluorescent ddXTPs, dXTPs, primers and polymerase. This method requires sophisticated equipment, which needs to be operated by a qualified technician. Also, this method is limited to sequences of less than one thousand nucleotides in length.

Other sequencing methods have been considered in order to reduce cost, simplify the method, and allow sequencing to take place out of the lab. Cycle extension, polymerase reading, exonuclease sequencing, and DNA micro-arrays are methods that have been considered (Braslavsky, I., B. Herbert, et al. (2003), PNAS **100**(7): 3960-3964). These methods have been comprehensively reviewed (Marziali, A. and M. Akeson (2001), Ann. Rev. Biomed. Eng. **3**: 195-223).

One potential method of sequencing DNA is based on threading a single strand of DNA through a nanopore and identifying its sequence from the variation in the ionic current flowing through the pore as the strand is threaded (Kasianowicz, J. J., E. Brandin, et al. (1996), Proc. Natl. Acad. Sci. **93**: 13770-13773). A second potential approach is exonuclease sequencing (Chan, E. Y. (2005), Mutat. Res. **573**: 13-40). This method involves digesting the DNA one nucleotide at a time (Dapprich, J. (1999), Cytomet. **36**: 163-168; and Matsuura, S.-I., J. Komatsu, et al. (2001), Nuc. Ac. Res. **29**(16): e79) and then identifying each of the released nucleotides. However, these methods require modification of the DNA before digestion or modification of the nucleotides once they have been released from the DNA by exonuclease. The development of exonuclease sequencing is currently being held back by the difficulty in identifying the nucleotides at the single molecular level as they are released by the enzyme. Investigators have tried to identify the nucleotides using fluorescent labeling with limited success.

Stochastic sensing involves placing a nanometer sized pore in an insulating lipid bilayer membrane and measuring the ionic transport through the pore. When an analyte interacts with a binding site within the pore, a change in the ionic current is detected (Braha, O., B. Walker, et al. (1997), Chem. & Biol. **4**: 497-505; and Bayley, H. and P. S. Cremer (2001), Nature **413**: 226-230). The extent and duration of the current block resulting from each binding event can reveal the identity of the analyte. The frequency of the binding events can reveal the analyte concentration. Various binding sites can be created within the pore by way of protein mutation, chemical modification, and by use of molecular adaptors and carriers (Gu, L.-Q., O. Braha, et al. (1999), Nature **398**: 686-690; and Braha, O., J. Webb, et al. (2005), Chem. Phys. Chem. **6**: 889-892).

Summary of the invention

It has been surprisingly demonstrated that individual nucleotides can be identified at the single molecule level from their current amplitude when they interact with a transmembrane pore. Hence, stochastic sensing may be used to identify individual nucleotides and to sequence nucleic acid sequences via exonuclease sequencing.

Accordingly, the invention provides a method of identifying an individual nucleotide, comprising:

(a) contacting the nucleotide with a transmembrane protein pore 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.

The invention further provides:

- 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 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; and

- a kit for sequencing a nucleic acid, comprising:
  - a cyclodextrin; and
  - a processive exonuclease.

The method of sequencing of the invention is a rapid and simple DNA sequencing method at the single molecule level. It is also a cheap method of sequencing DNA because it does not involve the use of expensive reagents, such as fluorophores.

<u>Description of the Figures</u>

Figure 1 shows the α-hemolysin $(M113R)_7$ mutant and heptakis-6-amino-β-cyclodextrin (am_7-βCD). A - sagittal cut through the α-hemolysin structure, position 113 is indicated by the arrow. B - spacefilled structure of am_7-βCD. C -possible interaction of am_7-βCD with α-hemolysin $(M113R)_7$

Figure 2A shows dCMP detection. A - Current trace of single $(M113R)_7$ mutant inserted in a phospholipid bilayer at +130 mV. L1 identifies the current of the unoccupied protein nanopore. B - in the presence of 40 μM am_7-βCD in the trans chamber. L2 indicates the current level observed when am_7-βCD binds temporarily inside the nanopore. C - dCMP 5 μM is now added to the cis chamber. L3 shows the current level that is observed when dCMP binds to the temporary complex $(M113R)_7$/ am_7-βCD.

Figure 2B showns the nteraction of the α-hemolysin (αHL) pore with heptakis-(6-deoxy-6-amino)-β-cyclodextrin (am_7βCD) and dCMP. A - Model of the heptameric αHL pore (7AHL), in which Met-113 has been substituted with Arg. A model of am_7βCD in cross-section generated in ChemDraw Ultra has been positioned manually at van der Waals distances from the Arg side chains, which block the passage of the cyclodextrin when it enters the pore from the *trans* side.

When $am_7\beta CD$ is present inside the pore, two rings of positive charge, one ring of seven primary amino groups contributed by the cyclodextrin, and a second ring of seven arginine side-chains, are separated by ~10Å. Aminocyclodextrins have previously been shown to bind nucleoside monophosphates with the phosphate group in an ionic interaction with the protonated amino groups. It is possible that the overall stability of such complexes is enhanced by p-cation interactions between the nucleotide bases and the Arg side chains. The dCMP molecule is positioned so that the phosphate group interacts with the protonated amines of $am_7\beta CD$ and the cytosine ring interacts with the guanidinium groups of the Arg side chains. B - Current trace from a single $(M113R)_7$ pore at +130 mV. L1 identifies the current flowing through the unoccupied protein nanopore, which is shown as a model on the right. C - Current trace after the addition of 40 μM $am_7\beta CD$ to the *trans* chamber. L2 indicates the current level observed when $am_7\beta CD$ is bound inside the nanopore. D - Current trace after the addition of 5 μM dCMP to the *cis* chamber. L3 shows the current level that is observed when dCMP binds to the $(M113R)_7\bullet am_7\beta CD$ complex.

Figure 3 shows dXMP current amplitudes. Current trace of single $(M113R)_7$ pore inserted in a phospholipid bilayer, at +130 mV potential. 40 μM am_7-βCD is present in the trans chamber. A - dGMP 5 μM is added to the cis chamber. The all points histogram of the current trace is shown on the right together with the structures of dGMP, dTMP (B), dAMP (C), and dCMP (D).

Figure 4 shows cyclodextrin current levels. Current trace of single $(M113R)_7$ mutant inserted in a phospholipid bilayer with 40 μM am_7-βCD present in the trans chamber at +130 mV. L1 and L1′ indicate two current levels of the unoccupied nanopore, and L2 and L2′ show two current levels resulting from the binding of am_7-βCD to $(M113R)_7$. The insert shows the amplitude histogram of the current trace with the peaks corresponding to the current levels L1, L1′, L2, and L2′.

Figure 5 shows single event analysis. A shows a single event analysis histogram of the L3 current level from all four dXMP in the same solution. B shows a single event analysis histogram of L3 originating from L2 only. 5 μM of dGMP, dAMP, dCMP, and 10 μM of dTMP are present in the *cis* chamber.

Figure 6 shows simultaneous detection of dXMP. A shows the current trace of a single $(M113R)_7$ mutant inserted in a phospholipid bilayer, +130 mV potential is applied between the Ag/AgCl electrodes. The buffer is Tris-HCl 25 mM pH 8.0 with 1M KCl. 40 μM am_7-βCD is present in the trans chamber. 5 μM of dGMP, dTMP, dAMP, and dCMP are added to the cis chamber. The colored bands illustrate the amplitude distribution of each dXMP. B displays an all point histogram from a current trace of 8000 binding events. Each peak is super-imposed with the statistical distribution of each dXMP.

Figure 7 shows the statistical method. Two Gaussian distributions A and B overlap at the point of intersection I. The area of Gaussian A beyond the point of intersection I is integrated and represents the probability of population A to be identified as population B.

Description of the Sequence Listing

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

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

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

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

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

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

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

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

SEQ ID NO: 9 shows the amino acid sequence of lambda exonuclease. The sequence is one of three identical subunits that assemble into a trimer.

Detailed description of the invention

Method of identifying an individual nucleotide

In a first embodiment, the present invention relates to a method of identifying an individual nucleotide comprising contacting the nucleotide with a transmembrane protein pore 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. 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.

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 individual nucleotide may however be bonded or attached to other chemical groups, such as fluorescent molecules or chemical groups containing radioisotopes, e.g. ^{125}I, ^{35}S. The types of nucleotides for identification in accordance with the invention are discussed in more detail below.

The method may be carried out using any suitable membrane/pore system in which a transmembrane protein pore is inserted into a membrane. The method is typically carried out using (i) an artificial membrane comprising a naturally-occurring or recombinant transmembrane protein pore, (ii) an isolated, naturally-occurring membrane comprising a recombinant transmembrane protein pore, (iii) an isolated, naturally-occurring membrane comprising a transmembrane protein pore or

(iv) a cell expressing a naturally-occurring or recombinant transmembrane protein pore. 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.

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

Membrane

The membrane forms a barrier to the flow of ions and nucleotides. The membrane is preferably a lipid bilayer. Lipid bilayers suitable for use in accordance with 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.

Transmembrane protein pore

The method of the invention is carried out using a 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 electrochemical gradient. The pore preferably permits the nucleotide to flow from one side of the membrane to the other along an electrochemical gradient.

The pore is typically an oligomer. The pore is preferably made up of several repeating subunits. The pore is preferably pentameric or heptameric. The pore typically comprises a barrel or channel through which the ions may flow.

The barrel or channel of the pore typically comprises amino acids that facilitate interaction with the nucleotide. A pore for use in accordance with the invention typically comprises one or more positively charged amino acids, such as arginine, lysine or histidine. These positively charged amino acids are preferably located near the constriction of the barrel or channel. These amino acids typically facilitate the interaction between the pore and the nucleotide by interacting with the phosphate groups in the nucleotide or by p-cation interaction with the base in the nucleotide. The pore 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.

Suitable pores for use in accordance with the invention include, but are not limited to, α-hemolysin, porins and leukocidins.

The preferred pore for use in the invention is α-hemolysin or a variant thereof. The α-hemolysin pore is formed of seven identical subunits (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. 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. The seven subunits within a variant pore are typically identical but may be different.

A preferred variant of α-hemolysin has 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. Variants typically include a positively charged amino acid at position 113 of each subunit. The pore for use in the invention is preferably α-hemolysin $(M113K)_7$ which comprises seven subunits as shown in SEQ ID NO: 4 or preferably α-hemolysin $(M113H)_7$ which comprises seven subunits as shown in SEQ ID NO: 6 or most preferably α-hemolysin $(M113R)_7$ which comprises seven subunits as shown in SEQ ID NO: 8.

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 the following table. 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.

Pores used in accordance with the invention may be modified for example by the addition of histidine 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. It may be desirable to provide the polypeptides in a form suitable for attachment to a solid support. For example, the pore may be attached to a solid support in order to insert the pore into the membrane.

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, polynucleotides and linkers 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 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 modified following either synthetic or recombinant production.

The pores may also be produced using D-amino acids. In such cases the amino acids will be linked in reverse sequence in the C to N orientation. This is conventional in the art for producing such proteins or peptides.

A number of side chain modifications are known in the art and may be made to the side chains of the pores. Such modifications include, for example, modifications of amino acids by reductive alkylation 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 may be isolated and replicated using standard methods in the art. Nucleic acid sequences encoding a pore may be expressed in a bacterial host cell using standard techniques in the art. The pore may be introduced into 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 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 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 may be made by introducing a polynucleotide encoding a pore 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 are known in the art and described in more detail below.

The nucleic acid sequence encoding a pore may be cloned into suitable expression vector. In an expression vector, the nucleic acid sequence encoding a pore 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.

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 genes may be introduced into the vector.

The expression vector may then be introduced into a suitable host cell. Thus the method of the invention may be carried out on a cell produced by introducing a nucleic acid sequence encoding a pore 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. Alternatively, the recombinant pore produced in this manner may be isolated from the bacterial host cell and inserted into another 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 at a high level. Host cells transformed with a nucleic acid sequence encoding a pore 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 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 may then be inserted into a naturally-occurring or artificial membrane for use in accordance with the invention.

The method of the invention may employ any one of the pores described above.

Interaction between the pore and nucleotide

The nucleotide may be contacted with the pore on either side of the membrane. The nucleotide may be introduced to the pore on either side of the membrane. The nucleotide is preferably contacted with the pore on a side of the membrane that allows ions to enter the pore and flow across the membrane along an electrochemical gradient. The nucleotide is preferably contacted with a side of the membrane that allows the nucleotide to pass through the pore to the other side of the membrane. For example, the nucleotide is contacted with an end of the pore which in its native environment allows the entry of ions or small molecules, such as nucleotides, into the barrel or channel of the pore such that the ions or small molecules may pass through the pore.

The nucleotide may interact with the pore in any manner and at any site. The nucleotide preferably reversibly binds to the pore. The nucleotide more preferably reversibly binds to the barrel or the channel of the pore. The nucleotide most preferably reversibly binds to the channel or barrel of the pore as it passes through the pore across the membrane.

During the interaction between the nucleotide and the pore, the nucleotide affects the current flowing through the pore in a manner specific for that nucleotide. For example, a particular nucleotide will reduce the current flowing through the pore for a particular time period and to a particular extent. Control experiments may be carried out to determine the effect a particular nucleotide 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 nucleotide in the sample or determine whether a particular nucleotide is present in the sample. The frequency at which the current flowing through the pore is affected in a manner indicative of a particular nucleotide can be used to determine the concentration of that nucleotide 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 comprising the pore is formed. The nucleotide may be contacted with the pore by introducing the nucleotide into the chamber. The nucleotide 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 nucleotide. Therefore the apparatus also comprises an electrical circuit capable of applying 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 patch clamp. The Example discloses one way of carry out a patch clamp method.

Molecular adaptor

The transmembrane pore preferably comprises a molecular adaptor that facilitates the interaction between the pore and the nucleotide. The adaptor typically has an effect on the physical or chemical properties of the pore that improves its interaction with the nucleotide. The adaptor typically alters the charge of the barrel or channel of the pore or specifically interacts with or binds to the nucleotide thereby facilitating its interaction with the pore. The adaptor preferably interacts with one or more phosphate groups on the nucleotide or interacts with the base in the nucleotide by p-cation interaction. The adaptor may mediate the interaction between the nucleotide and the pore. For instance, the nucleotide may reversibly bind to the pore via the adaptor. Alternatively, the adaptor may interact with the nucleotide in conjunction with the pore. For instance, the nucleotide may reversibly bind to both the pore and the adaptor. The adaptor preferably constricts the barrel or channel so that it may interact with the nucleotide.

The adaptor itself may reversibly interact with the pore and may therefore move in and out of the barrel or channel of the pore. Alternatively, the adaptor may be covalently attached to the barrel or channel of the pore so that it cannot leave.

The adaptor typically has a ring of amino groups. The adaptor preferably has a ring of seven amino groups. This ring of amino groups may interact with the nucleotide in combination with a ring of positively charged amino acids in the constriction of the barrel or channel of the pore.

One suitable adaptor is cyclodextrin. The adaptor is preferably heptakis-6-amino-β-cyclodextrin (am_7-β-CD).

Nucleotide

The method of the invention may be used to identify any nucleotide. The nucleotide can be naturally-occuring 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 monosphosphate (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. The invention is typically carried out on a sample that is known to contain or suspected of containing one or more nucleotides. The invention may be carried out on a sample that contains one or more nucleotides whose identity is unknown. Alternatively, the invention may be carried out on a sample to confirm the identity of one or more nucleotides whose presence in the sample is known or expected. 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 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 nucleotide. 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 +50 mV to +200 mV. The voltage used is preferably from +70 mV to +150 mV, from +85 mV to +145 mV or from +100 mV to +140 mV. 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 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 2M, 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 7.5 to 12.0, from 7.6 to 11.0, from 7.7 to 10.0, from 7.8 to 9.5, from 8.0 to 9.0 or from 8.0 to 8.5. The pH used is preferably about 8.0.

The method is typically carried out at from 14°C to 100°C, from 15°C to 90°C, from 16°C to 80°C, from 17°C to 70°C, from 18°C to 60°C, 19°C to 50°C, or from 20°C to 40°C. The method is preferably carried out at room temperature.

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

In a second embodiment, the invention relates to 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 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 second embodiment involves stochastic sensing of each single nucleotide of a nucleic acid sequence in a successive manner in order to sequence the nucleic acid. The whole or only part of the nucleic acid may be sequenced using the method of the second embodiment. The nucleic acid can be naturally-occuring or artificial. For instance, the method of the second embodiment may be used to verify the sequence of a manufactured oligonucleotide. The method of the second embodiment is typically carried out *in vitro*.

Steps (b) and (c) of the method of the second embodiment are generally identical to the steps carried out in the method of the first embodiment discussed above. All of the discussion above concerning the first embodiment, and in particular concerning the membranes, apparatus, pores, molecular adaptors, nucleotides and conditions that may be used in the first embodiment, equally applies to the second embodiment. The nucleic acid in the second embodiment is typically present in any biological sample as discussed above for the first embodiment.
The method of the second embodiment may be carried out on a sample which contains one or more nucleic acids whose sequence is unknown. Alternatively the

method of the second embodiment may be carried out on a sample to confirm the identity of nucleic acids whose presence in the sample is known or is expected. The nucleic acid sequence is typically amplified prior to being sequenced using the method of the second embodiment.

Processive exonuclease

The method of the second embodiment involves contacting the nucleic acid sequence 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 sequence 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 of the invention. The preferred enzyme for use in the method of the invention is lambda exonuclease. The sequence of one subunit of lambda exonuclease is shown in SEQ ID NO: 9. 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: 9 and which retain processive exonuclease activity. The variants may vary from SEQ ID NO: 9 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 a reduced rate of enzyme activity 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 the second embodiment 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 in accordance with the first embodiment of the invention. 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 in the method of the second embodiment 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 of the second embodiment at a salt concentration that reduces the rate of the activity of the enzyme without adversely affecting its activity. For example, the method of the second embodiment may be carried out at a salt concentration of from 0.5 to 1M. The salt concentration is preferably about 1M.

Kits

In a third embodiment, the invention also relates to kits that may be used to carry out the second embodiment of the invention. The kits are therefore suitable for sequencing nucleic acids. The kits comprises a cyclodextrin and a processive exonuclease. The cyclodextrin is preferably heptakis-6-amino-*β*-cyclodextrin. The processive exonuclease may be any of those discussed above with reference to the second embodiment. The kit preferably further comprises a transmembrane protein pore. The pore may be any of those discussed above with reference to the first embodiment.

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 order to bring the size of the ionic conducting path of the α-hemolysin $(M113R)_7$ mutant (Figure 1A) closer to the size of the nucleotide to be detected, the diameter of the nanopore was reduced by fitting a cyclodextrin near the constriction of the pore. The heptakis-6-amino-β-cyclodextrin (am_7-βCD)(Figure 1B), which has seven amino groups in the primary positions, was used. When the cyclodextrin is inside the pore (Figure 1C), in conjunction with the seven arginines in position 113 on the protein mutant, one ring of seven amino groups on one side, and a second ring of seven arginine groups are present within a short distance from each other in the narrowest area of the passage through the pore. This amino/arginine ring structure has the property of binding phosphate groups reversibly thus immobilising the XMP and dXMP in the pore for 5 to 30 ms. These binding events are clearly detectable through the resulting change in current amplitude.

1. Material and methods

α-hemolysin mutant $(M113R)_7$ was expressed and purified as previously described (Cheley, S., L.-Q. Gu, et al. (2002), Chem. & Biol. **9**: 829-838).

Chemicals

1,2-diphytanoyl-*sn*-glycero-3-phosphocholine from Avanti Polar Lipids Inc. Pentane was purchased from JT Baker, and hexadecane 99+% from Sigma-Aldrich. Heptakis(6-deoxy-6-amino)-β-cyclodextrin.HCl >99% was purchased from CYCLOLAB Ltd Budapest, Hungary. 2-deoxy-guanosine 5′ monophosphate sodium salt 99% was purchased from Acros, 2-deoxy-cytosine 5′ monophosphate di-sodium salt >95%, 2-deoxy-thymidine 5′ monophosphate di-sodium salt >97%, and 2-deoxy-adenosine 5′ monophosphate di-sodium salt >95% from Fluka. Uridine 5′ monophosphate di-sodium salt 99%, and cytosine 5′ monophosphate acid >98% were bought from Fluka. Adenosine 5′ monophosphate acid 99%, and guanosine 5′ monophosphate di-sodium salt 97% were purchased from Acros. Trizma Base 99.9% was purchased from Sigma-Aldrich, and concentrated HCl analytical reagent grade from Fisher Scientific. Potassium chloride 99%, and sodium chloride 99.9% were purchased from Sigma-Aldrich. Potassium bromide 99.5% and cesium chloride 99% were purchased from Fluka.

Equipment

A patch clamp amplifier Axopatch 200B from Axon instruments was used with a computer equipped with a Digidata 1200 A/D converter (Axon instruments). A Teflon chamber was used. Data was collected in pClamp 9.2, and analyzed in Clampfit 9.0. Plots and graphs were obtained with Microcal Origin 6.0, and integration were run on a personal calculator.

Experimental conditions

Lipid bilayer membranes were formed from 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine by the method of Montal and Mueller (1972), on 100-150 μm diameter orifice in a 20 μm polycarbonate film (20 μm thickness from Goodfellow, Malvern PA) separating the *trans* and the *cis* chamber. The *cis* side of the chamber was at ground, and the *trans* side of the chamber was connected to the head stage. The potential refers to the potential value of the *trans* side electrode. The adaptor molecule was added to the *trans* side, the α-hemolysin mutant and the analyte molecules were added to the *cis* side. dXMP experiments were carried out at +130mV, XMP experiments at +110mV. All experiments reported here were obtained at pH 8.0 Tris-HCl 25 mM in 1 M KCl. Fresh aliquots of nucleotide solution were used everyday. Experiments were carried out at room temperature 22.5 +/-2°C unless stated otherwise.

2. Results

2-deoxy-nucleotide 5′ monophosphates partially block homoheptameric pores formed by $(M113R)_7$/ heptakis 6 amino β-cyclodextrin

Single-channel recordings were carried out on the homo-heptameric pores formed from αHL-M113R with am_7-βCD applied from the *trans* side (Figure 2A and 2B). In the absence of am_7-βCD, the pore remained permanently open (Figure 2A and 2B, B) with a unitary current (L1) of 145±5 pA (+130 mV) in 1M KCl in pH 8.0 Tris-HCl 25 mM buffer. The addition of 40 μM am_7-βCD in the *trans* chamber alone leads to reversible blocking events to a current level of 65±5 pA (L2 in Figure 2A, B and Figure 2B, C). Upon addition of 5 μM dCMP to the *cis* chamber, a third current level is observed at 22±1 pA (L3 in Figure 2A, C and Figure 2B, D) originating from current level L2. L3 represents the binding of dCMP to the complex of $(M113R)_7$ with am_7-βCD. Addition of dXMP at up to 300 μM to the *trans* instead of the *cis* side of the chamber did not lead to any alteration of the cyclodextrin binding conductance states (not shown).

In the experimental conditions described above, current blocking events due to cyclodextrin binding were observed when unmodified β-cyclodextrin was added to the *trans* chamber (40 μM) in the presence of a αHL-M113R single nanopore. However, no further current blocking events were observed, when dXMPs (up to 300 μM) were added either to the *trans* or the *cis* chamber (not shown).

In the absence of am$_7$-βCD in the *trans* chamber, blocking events (<1ms) were observed when minimum concentrations of 300 μM dGMP or dTMP were added to the *cis* chamber (not shown). These events are not observed over the timescale of the experiment at 5 μM dXMP or XMP.

Adding am$_7$-βCD in the *trans* chamber while measuring the current through a wild type α-HL single channel led to cyclodextrin binding events, but no further alterations of the current were observed when adding dXMP to either the *cis* or the *trans* chamber.

<u>2-deoxy-nucleotide 5′ monophosphate can be identified from the amplitude of the partial block of homoheptameric pores formed by (M113R)$_7$/ heptakis 6 amino β-cyclodextrin</u>

The partial block of the transient complex (M113R)$_7$/ am$_7$-βCD differed in amplitude depending on which dXMP was added to the *cis* chamber (Figure 3). Addition of dGMP (5 μM) to the *cis* side displayed an average blocking to a current level of 16 pA (Figure 3A). The all points amplitude histogram of the trace shown in Figure 3A is shown to the right of the trace together with the structure of dGMP. The other nucleotides all display different amplitudes as shown in Figure 3B for dTMP, 3C for dAMP, and 3D for dCMP. Out of the four dXMP, dGMP blocks the most current.

The current amplitudes from independent experiments displayed some variations that originated from individualities of the protein nanopore. The average current, at +130 mV, for a (M113R)$_7$ is 139 pA, but some channels display currents as high as 147 pA and as low as 131 pA. Therefore, to compare current traces current traces were normalized from different experiments between 0 current and the (M113R)$_7$/am$_7$-βCD current level set to 65 pA.

The dwell time (τ_{off}) of each dXMP was calculated over 500 events from each of 3 independent experiments (Table 1).

Table 1. Dwell time in ms of dGMP, dTMP, dAMP, and dCMP averaged from three independent measurements.

	dGMP	dTMP	dAMP	dCMP
τ_{off} (ms)	9.8 ±0.2	19.8 ±0.8	7.1 ±0.2	10.5 ±0.4

Cyclodextrin current levels

At pH 8.0, the mutant $(M113R)_7$ exhibits two current levels L1/L1′ when the protein channel is unoccupied (Figure 4). The cyclodextrin adapter can bind to the protein regardless of which current level L1/L1′ the protein is in.

Two current levels are observed when recording the current level of the $(M113R)_7$ nanopore (Figure 4). L1 is the main current level, as shown in the insert of Figure 4. The binding of am_7-βCD to the nanopore leads to two current levels represented by L2 and L2′ (three levels are observed at pH 7.5, not shown). The binding of am_7-βCD to the protein nanopore occurs independently of current level L1 or L1′. L2 is the main conductance level observed when am_7-βCD is bound to $(M113R)_7$ and it originates from both L1 and L1′ conductance levels of the empty nanopore with no apparent correlation. The current level L2′ observed when am_7-βCD is bound to $(M113R)_7$ accounts for less than 15% of the conductance when the cyclodextrin adapter is bound (see insert Figure 4).

The nucleotide binding events sometimes vary in amplitude as a result of which current level L2 and L2′ they originate from. A 0.5pA shift was observed depending on which of L2 or L2′ the dXMP binding event originates from (not shown). It leads to an increased overlap of the nucleotide binding event histogram (Figure 5).

It is possible to analyze manually each recording in order to remove each analyte binding event stemming from the bound cyclodextrin current level L2′ described in Figure 3. Figure 5 shows the difference between the single event analysis histogram obtained from an unmodified dXMP detection current recording (Figure 5A), and the same recording where analyte binding events stemming off level L2′ (Figure 4) have been removed. The two histograms display the same four peaks corresponding to dGMP, dTMP, dAMP, and dCMP. The amplitude of the peaks in Figure 5A is larger than in 5B because the analyte binding events stemming off L2′ have been removed, therefore the histogram originates from fewer events. The separation between each peak seems better in 5B than in 5A. However, removing these events from the recording did not yield a complete separation of each peak (Figure 5B). As a result, the cyclodextrin current levels L2 and L2′ shown in Figure 4 were not taken into account in the single event analysis histograms, and resulting statistics reported hereafter.

2-deoxy nucleotide 5′ monophosphates can be identified from the amplitude of the partial block of homoheptameric pores formed by the transient complex $(M113R)_7$/ heptakis 6 amino β-cyclodextrin

The partial block of $(M113R)_7$/am_7-βCD pores proved to differ in amplitude depending on which dXMP was added to the *cis* chamber. The different amplitudes could be resolved when dGMP, dTMP, dAMP, and dCMP were added to the *cis* chamber simultaneously (Figure 5).

Figure 6A shows the current amplitude for a mixed solution of all 4 nucleotides from a single experiment. Colored bands are superimposed onto the recorded current trace in order to illustrate the amplitude distribution of each dXMP. Figure 6B shows an amplitude histogram of a current trace of 8000 events assumed to be the amplitude distribution generated by each nucleotide. The amplitude histogram is superimposed with Gaussian distributions. The fit is obtained from the peak current value given by this experiment as the distribution mean value and the σ value that was obtained from fitting and averaging the distribution of each nucleotide independently (Table 2). From fitting the current traces with Gaussian distributions, the probabilities of identification for each nucleotide was established.

Statistical methods

Current traces of $(M113R)_7$ in the presence of am_7-βCD on the *trans* side and one of the analyte nucleotides on the *cis* side were digitally filtered at 300Hz (low pass Gaussian filter), and an all points amplitude histogram was constructed. These histograms display a large peak corresponding to the current amplitude that is observed when the am_7-βCD binds to the $(M113R)_7$ α-hemolysin mutant(corresponding to L2 in Figure 1). This current amplitude varies between protein channels within 5% from one experiment to another. For this reason, the all points amplitude histograms were normalised between 0 current, and the main cyclodextrin peak set at 65 pA. In the normalised histogram, the nucleotide peak was fitted to a Gaussian distribution. The mean and sigma (σ) values of the same nucleotide were averaged from at least 3 independent experiments each containing 1000 events (values listed in Table 2).

Table 2. Average values of the distributions of each nucleotide from three independent experiments all normalised between 0 and 65 pA.

	Average
G	16.0, σ=0.64
T	17.4, σ=0.41
A	18.4, σ=0.54
C	20.0, σ=0.51

The probabilities for the reading of each base were determined from experiments with all 4 nucleotides present simultaneously over at least 3000 nucleotide binding events in each trace. The traces were filtered (300Hz low pass Gaussian digital filter) and normalised between 0 and the cyclodextrin peak at 65 pA as described above for the individual nucleotide experiments. The peak values of each nucleotide were averaged from 5 independent experiments (Table 3).

Table 3. Peak values of each nucleotide where all 4 nucleotides are present from 5 independent experiments. The last column displays the average value of each peak for which the overlap of the Gaussian distributions are integrated.

	Average
G(pA)	16.2±0.5
T(pA)	17.6±0.6
A(pA)	18.6±0.6
C(pA)	20.2±0.5

In experiments where all 4 nucleotides are present, the Gaussians from all 4 nucleotides have an overlap. The statistics were calculated for the binding signal from one nucleotide to be identified as itself or another nucleotide from the level of overlap between the Gaussian distribution of this nucleotide and that of its neighbors.

The point of intersection between the two Gaussians is calculated form the respective peak positions (given by the experiment with mixed nucleotides), and σ values for each distribution (given by the fitting of individual nucleotide experiments). The accuracy probability is given by integrating the area of the Gaussian that is beyond the intercept value with the neighboring Gaussian (Figure 6, Table 4). The first column of Table 4 is the nucleotide that interacts with the nanopore, and the first row is what is read from the corresponding current amplitude.

Table 4. Shows the probability of the added nucleotide (vertical) to be detected as itself or another nucleotide (horizontal).

	G_{read}	T_{read}	A_{read}	C_{read}
G_{added}	0.88	0.12	0	0
T_{added}	0.06	0.83	0.11	0
A_{added}	0	0.19	0.74	0.07
C_{added}	0	0	0.06	0.94

3. Conclusion

The results presented indicate that stochastic sensing is a promising alternative for the identification of single nucleotides. It also means that exonuclease sequencing can be used as a cheap, rapid, and simple DNA sequencing method at the single molecule level. Exonucleas sequencing is also a cheap method of sequencing DNA because it does not require expensive reagents, such as fluorophores.

All points histograms are a sufficient analysis method to identify each nucleotide with an accuracy ranging from 74 to 94% (Table 4). The dwell time values of the XMP and dXMP are too similar in the conditions to further differenciate each analyte. The statistics drawn from the amplitude histograms can be further improved by compensating for the cyclodextrin current levels as shown in Figure 4. The current amplitude difference between each dXMP is about 1 pA. This resolution depends on a number of parameters as follows.

Voltage dependence

The binding events are voltage dependent. At 50 mV, very few binding events are observed, suggesting that a minimum field is required to drive the dXMP and XMP to the binding site. At +150 and +200 mV the amplitudes no longer allow to differentiate the nucleotides. +130 mV proved to be the best voltage for deoxy-ribo nucleotides 5′ monophosphate, and +110 mV yielded the best resolution for ribo nucleotides 5′ monophosphate.

Salt concentration

Tris-HCl pH 8 buffer 0.5, 1, and 2M KCl were tested. From the all points amplitude histograms, the best resolution between the peaks was obtained at 1M KCl.

pH dependence

The current amplitudes are pH dependent, Tris-HCl buffer 1M KCl at pH 7.5, 8.0, 8.2, 8.5, 9.0, and 9.5 were tested. At pH 8.0 and above two current levels are observed upon binding of am_7-βCD (Figure 4). At pH 7.5 the heptakis 6 amino β-cyclodextrin displays a third current level (not shown). It causes dTMP to display two types of events with different amplitudes, one of which is within the range of dGMP events, thus leading to a loss of resolution. At pH 9.5, the nucleotide binding events are no longer observed. The best peak separation is obtained at pH 8.0.

Salt dependence

The resolution between dXMPs and XMPs is better with KCl than with NaCl or CsCl. 1M KCl yields better resolution than 2M KCl. The use of KBr did not allow the identification of the different nucleotides as each binding event led to a complete block of the transient complex $(M113R)_7$/am_7-βCD.

Temperature dependence

Lowering the temperature to 14°C or increasing it to 50°C did not interfere with the detection of the dXMP/XMP. However, it did not improve the resolution of the amplitude histograms.

Other α-hemolysin mutants

$(M113N)_7$ was seen to bind am_7-βCD but no nucleotide detection was observed. $(M113F)_7$ and $(M113F/147K)_7$ didn't yield detection whether am_7-βCD was added or not. $(M113K)_7$ was tested in the same conditions. In this case, the recording is very similar to that of Figure 1. The nucleotide binding was detected with $(M113K)_7$ mutant and am_7-βCD but the peak separation between the nucleotides was smaller than when $(M113R)_7$ was used.

Ribonucleotides 5′ monophosphate

XMP were successfully identified with this method, the resolution between each base was inferior to that of dXMP with peak separations smaller than 1 pA for U, A, and C. The all point histogram current amplitudes appear in the same order as those of dXMPs, with GMP displaying the lowest current (largest blocking), followed by UMP, AMP, and CMP with the highest current amplitude (smallest blocking). The optimal voltage for XMP identification was found to be +110mV at pH 8.0 1M KCl.

Mechanism

The $(M113R)_7/am_7$-βCD transient complex has also shown to bind and differentiate glucose phosphates (glucose 1P and galactose 1P). It suggests a strong interaction between the arginine ring on one side, the phosphate group and the amine ring from the am_7-βCD on the other side. Unmodified β-cyclodextrin does not yield any detection. Little difference is observed between XMP and dXMP suggesting that the hydroxyl groups do not play a large role in the binding

The embodiments of the present invention for which an exclusive property or privilege is claimed are defined as follows:

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

(a) contacting the nucleotide with a transmembrane protein pore, which comprises a cyclodextrin that facilitates an interaction between the nucleotide and the pore; and

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

2. A method according to claim 1, wherein the interaction involves the nucleotide reversibly binding to the channel of the pore.

3. A method according to claim 1 or 2, wherein the pore is (a) α-hemolysin formed of seven identical subunits as shown in SEQ ID NO: 2; or (b) a variant thereof in which one or more of the seven subunits has at least 50% amino acid identity to SEQ ID NO: 2 over the entire sequence and retains pore activity.

4. A method according to claim 3, wherein the variant is $(M113R)_7$.

5. A method according to any one of claims 1-4, wherein the cyclodextrin is a heptakis-6-amino-β-cyclodextrin (am_7-β-CD).

6. A method according to any one of claims 1-5, wherein the individual nucleotide is a monophosphate, diphosphate or triphosphate.

7. A method according to any one of claims 1-6, wherein the individual nucleotide is a ribonucleotide.

8. A method according to claim 7, further comprising before step (a) digesting a ribonucleic acid (RNA) sequence to provide the individual nucleotide.

9. A method according to any one of claims 1-6, wherein the individual nucleotide is a deoxyribonucleotide.

10. A method according to claim 9, further comprising before step (a) digesting a deoxyribonucleic acid (DNA) sequence to provide the individual nucleotide.

11. A method according to claim 8 or 10, wherein more than one of the individual nucleotides of the RNA or DNA sequence are contacted with the pore in a sequential manner such that the identity of the whole or part of the sequence may be determined.

12. 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, which comprises a cyclodextrin that facilitates an interaction between the nucleotide and 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.

13. A kit for sequencing a nucleic acid, comprising:

a cyclodextrin;

a processive exonuclease; and

a transmembrane protein pore.

Figures

Fig. 1

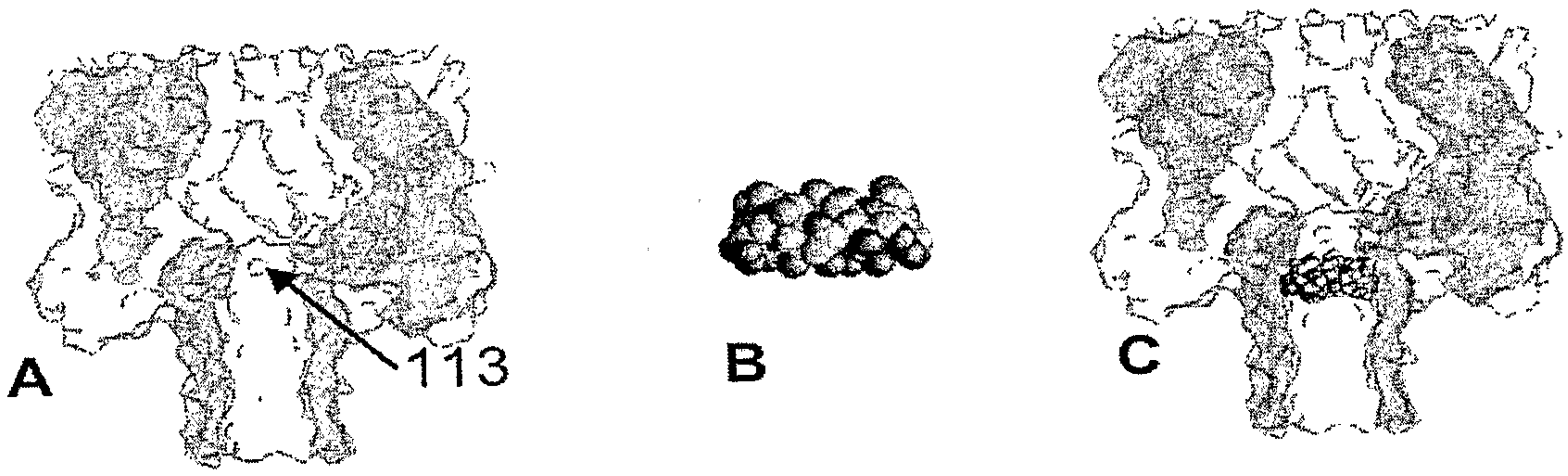


Fig.2A.

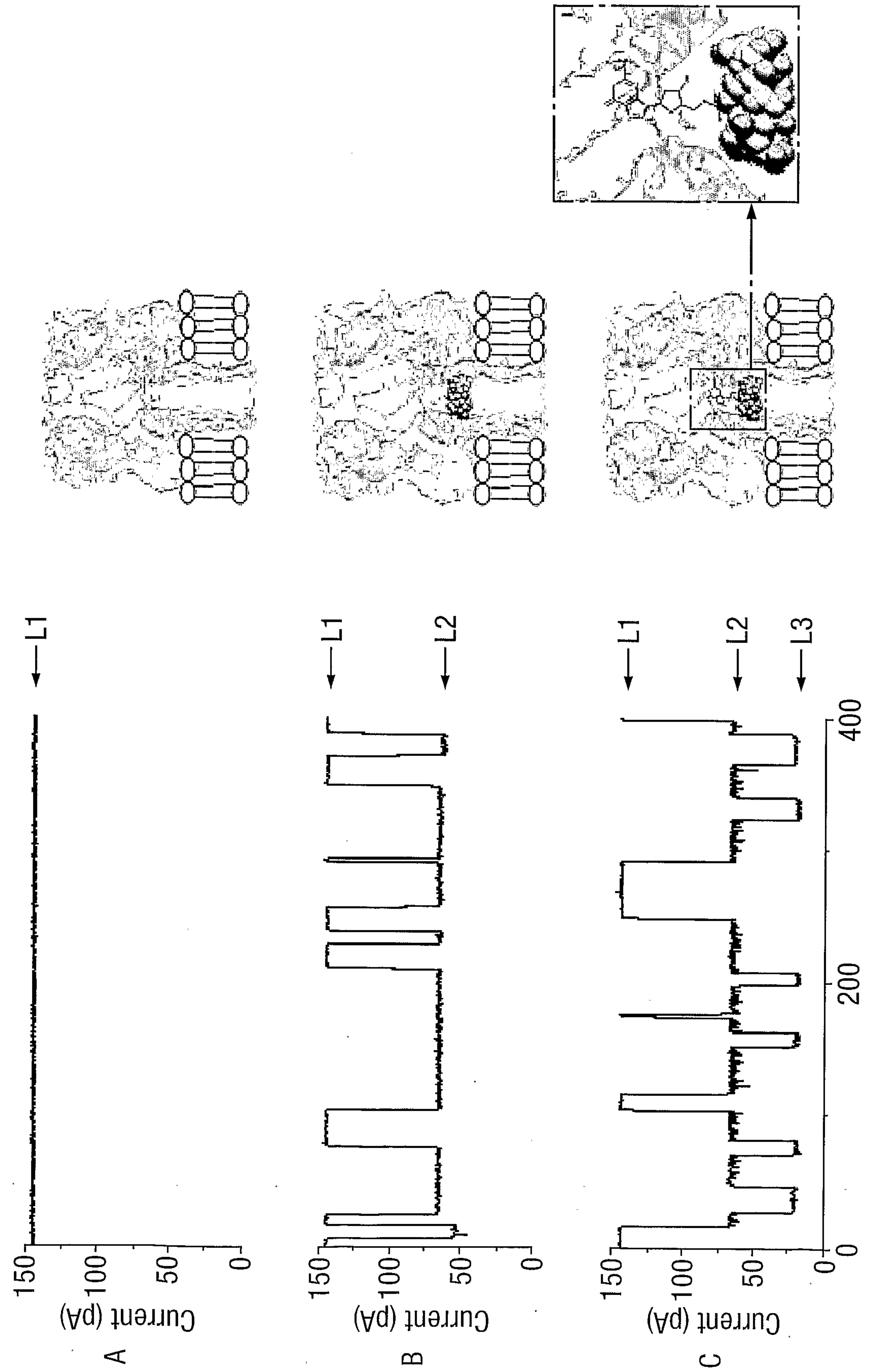
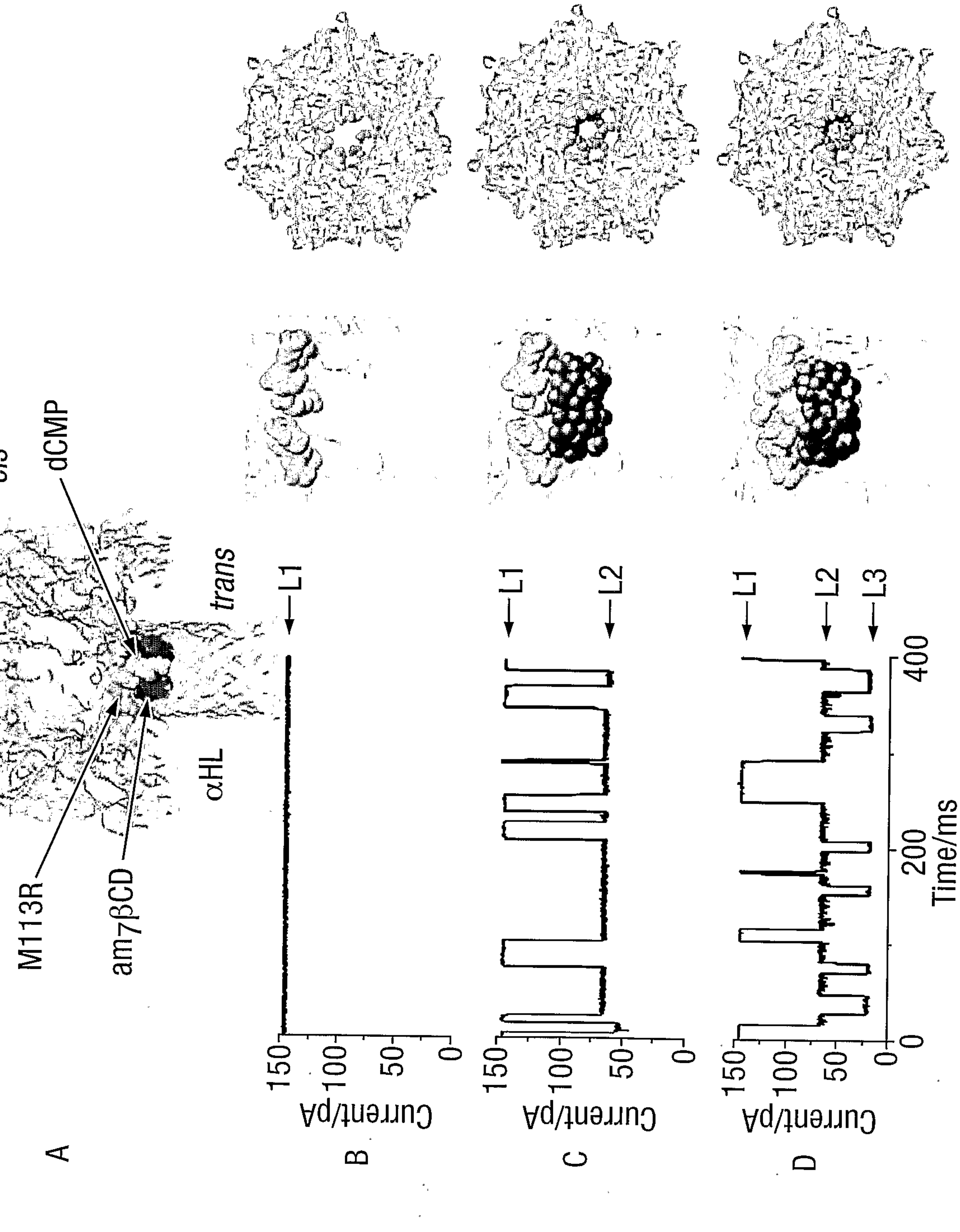


Fig.2B.


A
M113R
am7βCD
dCMP
cis
trans
αHL
B
C
D
Current/pA
150
100
50
0
L1
L2
L3
0
200
400
Time/ms

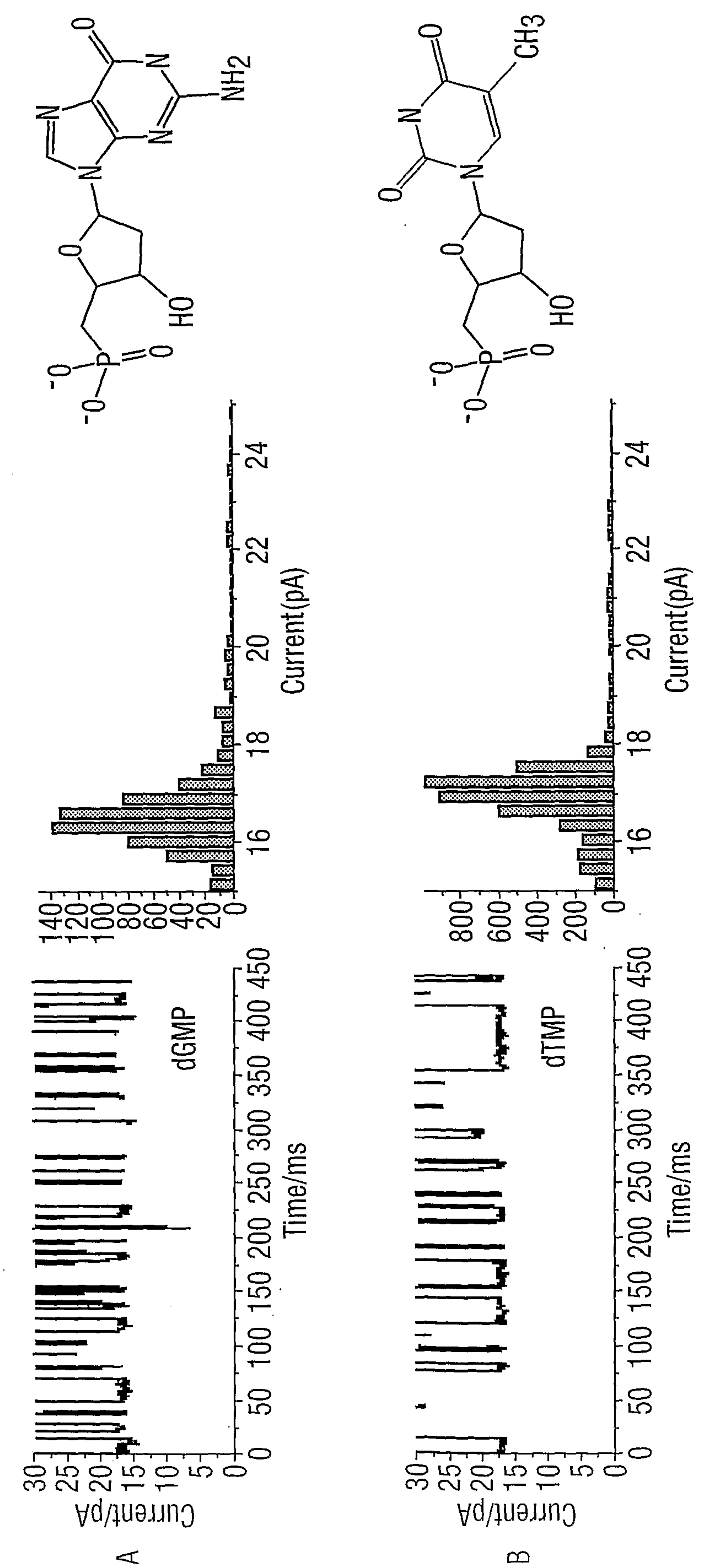


Fig.3.

Fig.3 (Cont).

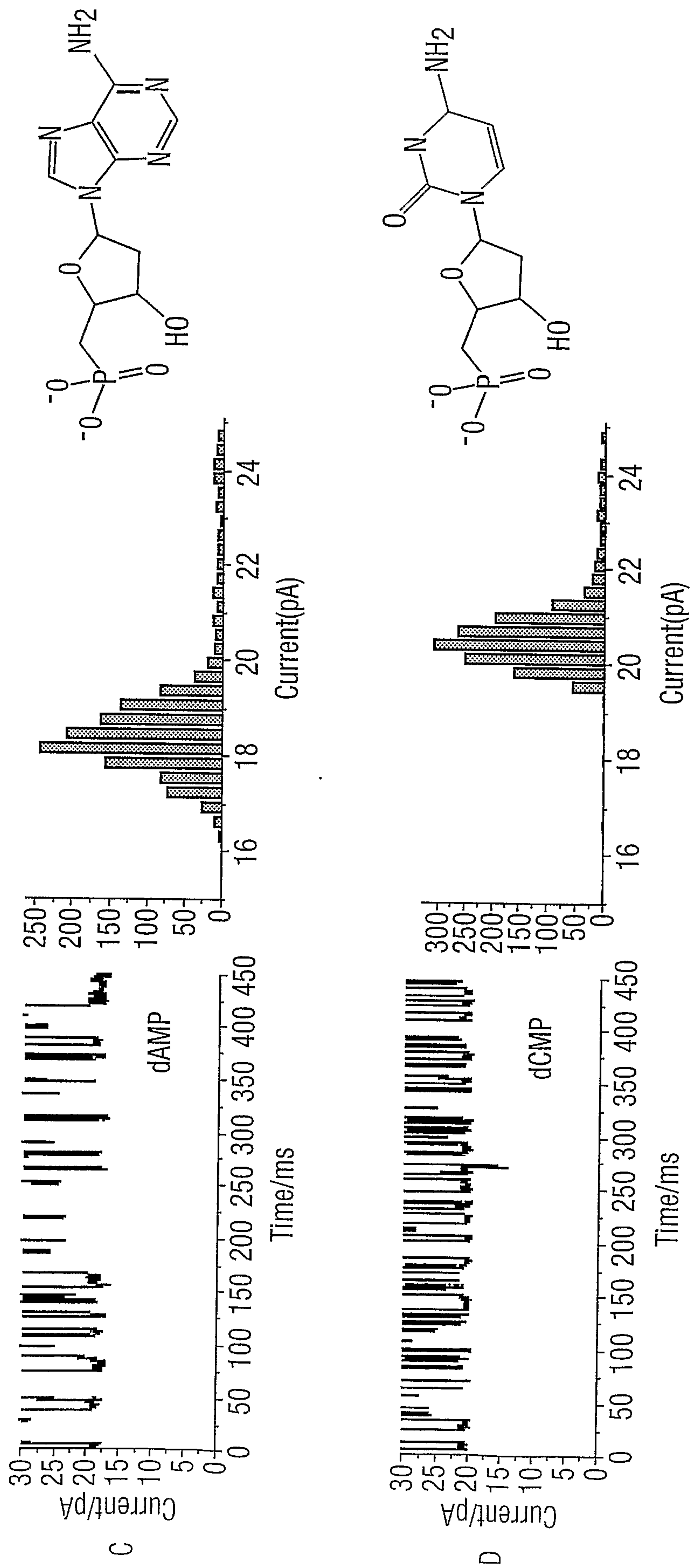


Fig. 4

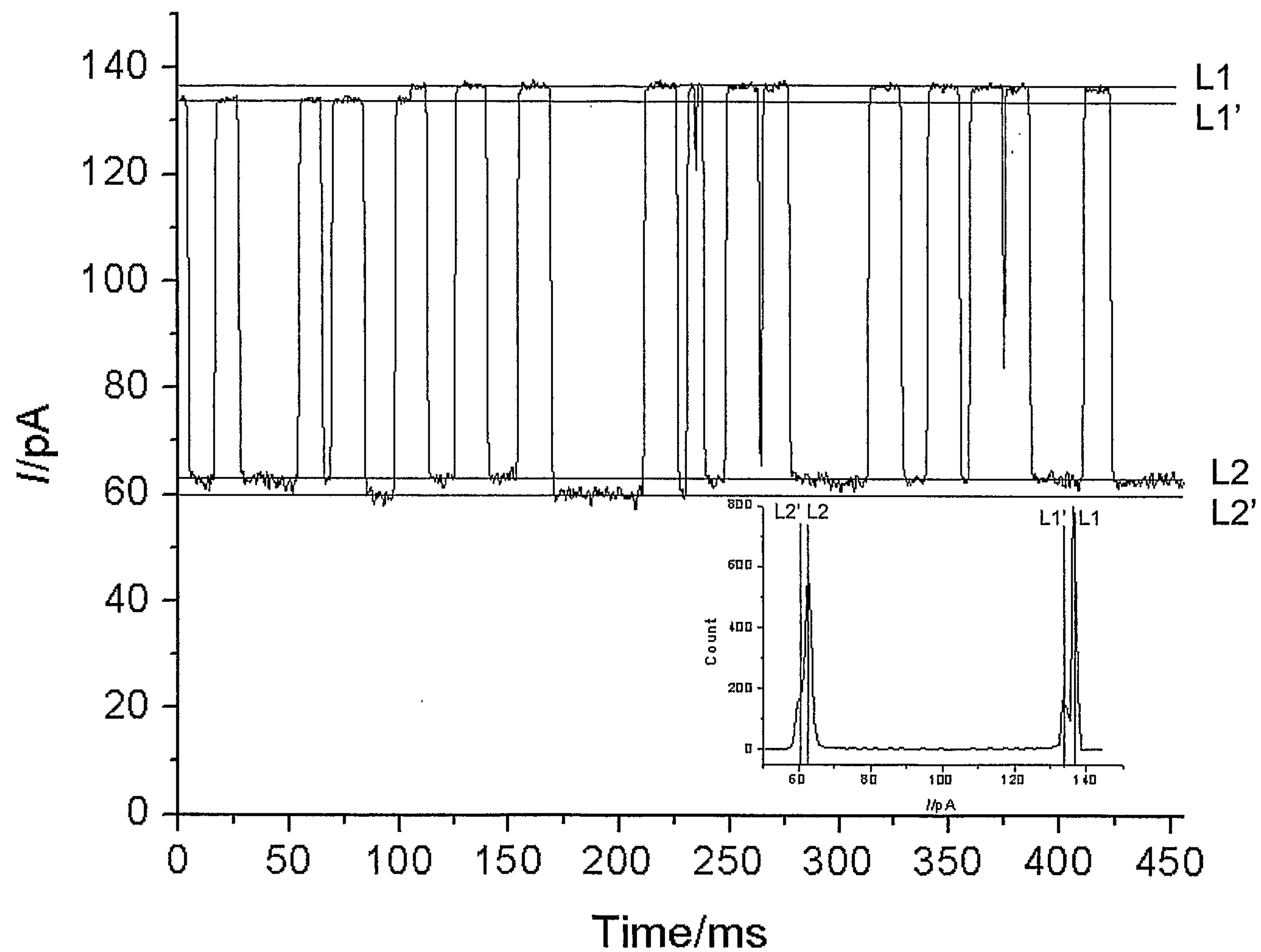

140
120
100
80
60
40
20
0
I/pA
0 50 100 150 200 250 300 350 400 450
Time/ms
L1
L1'
L2
L2'
Count
800
600
400
200
60 80 100 120 140
I/pA
L2' L2
L1' L1

Fig. 5

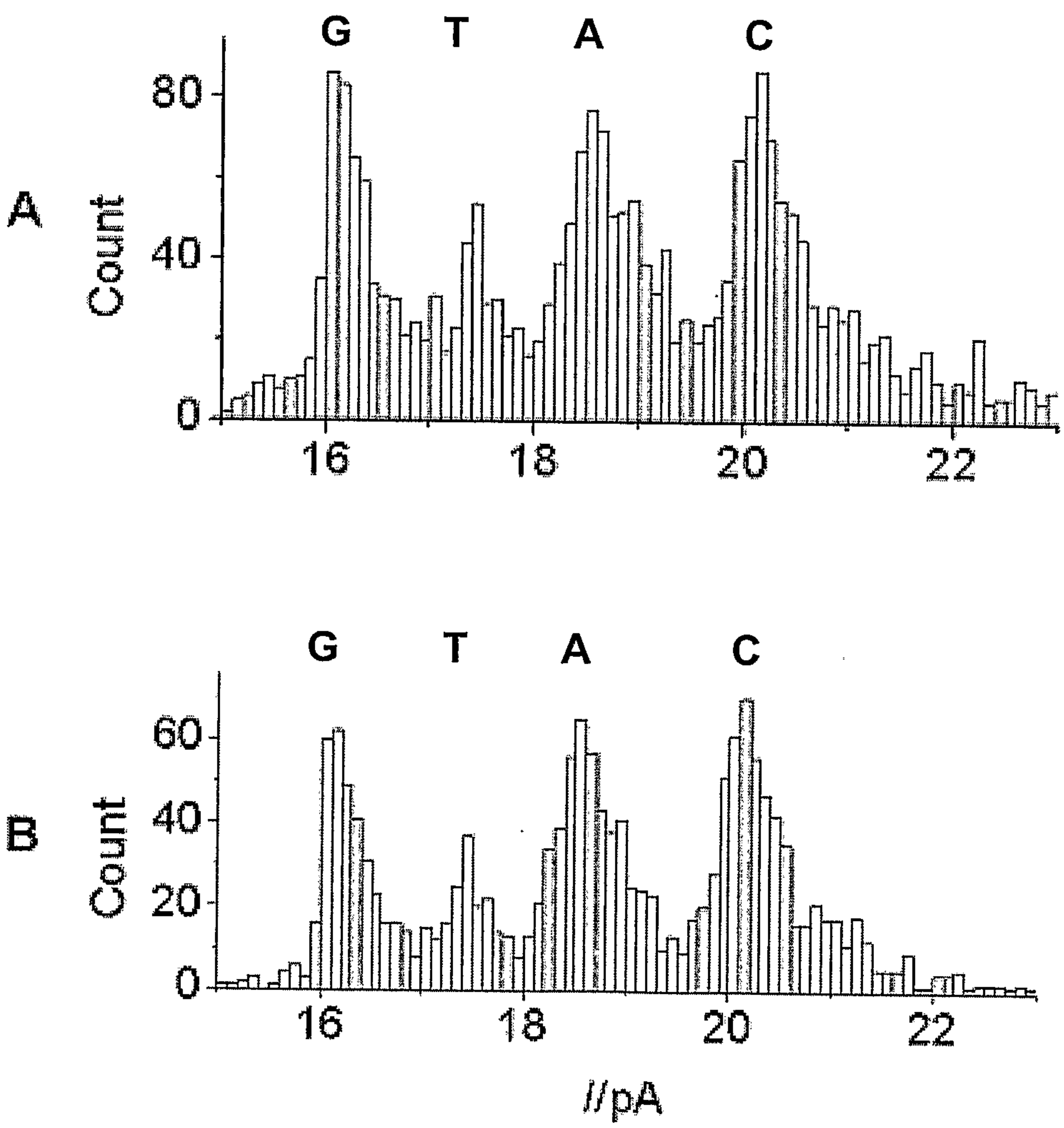
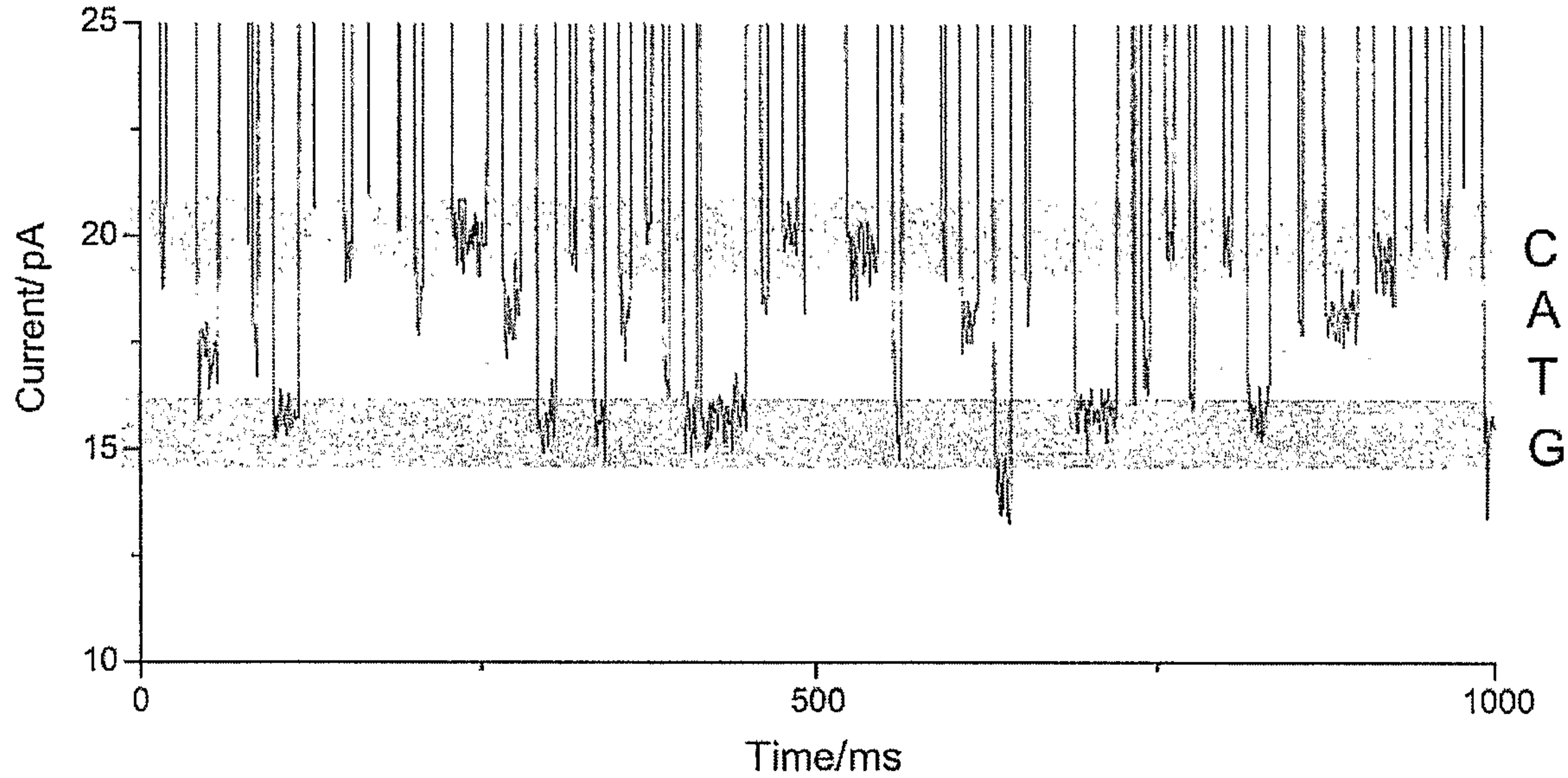


Fig. 6

A


25
20
15
10
Current/pA
0
500
1000
Time/ms
C
A
T
G

B

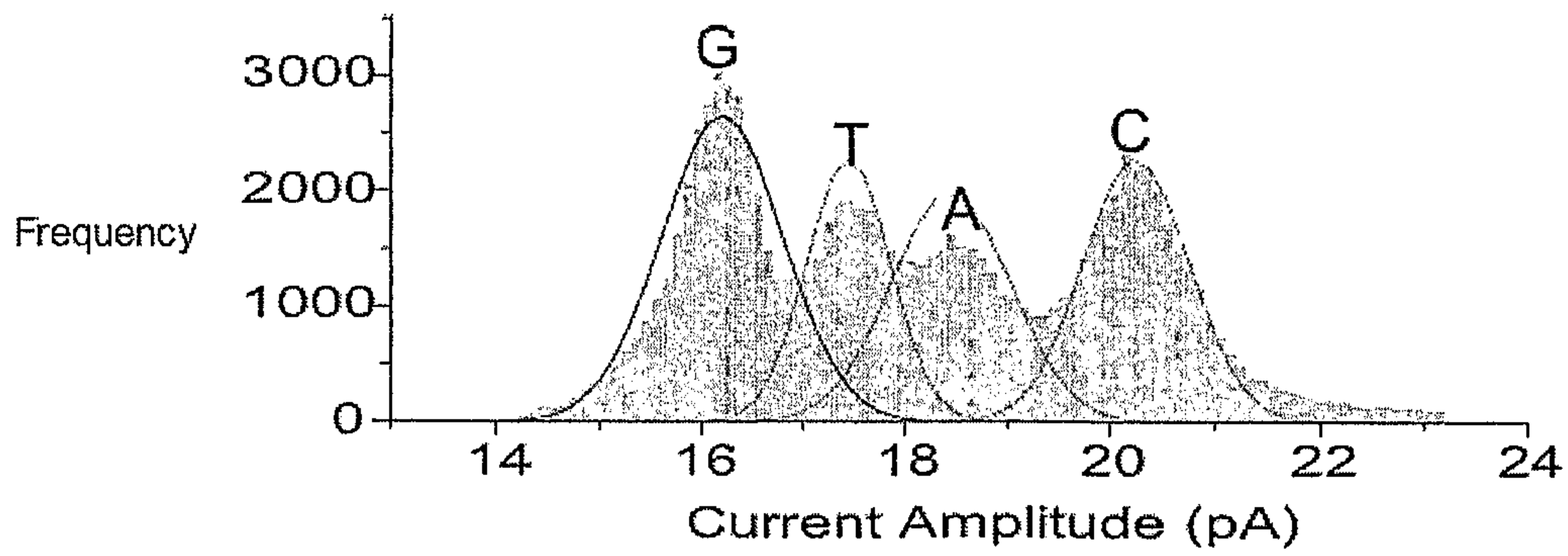

Frequency
3000
2000
1000
0
G
T
A
C
14
16
18
20
22
24
Current Amplitude (pA)

Fig. 7