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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 August 2012 (16.08.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/107778 A2

(51) International Patent Classification:
G01N 33/487 (2006.01)

(21) International Application Number:
PCT/GB2012/050301

(22) International Filing Date:
10 February 2012 (10.02.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/441,718 11 February 2011 (11.02.2011) US

(71) Applicant *(for all designated States except US)*: **OXFORD NANOPORE TECHNOLOGIES LIMITED** [GB/GB]; Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford, Oxfordshire OX4 4GA (GB).

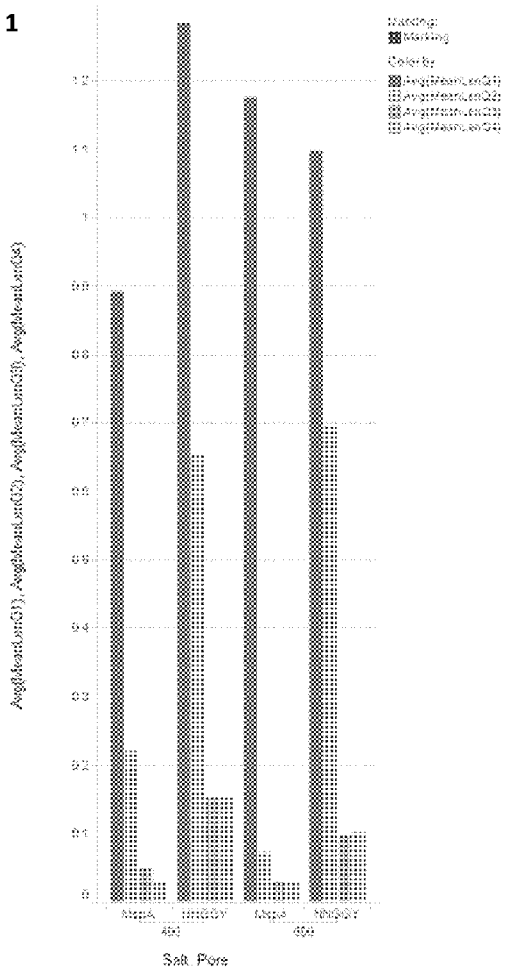
(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CLARKE, James** [GB/GB]; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **HERON, Andrew John** [GB/GB]; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **JAYASINGHE, Lakmal** [LK/GB]; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WALLACE, Jayne** [GB/GB]; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert

[Continued on next page]

(54) Title: MUTANT PORES

Fig. 1



(57) Abstract: The invention relates to mutant forms of Msp. The invention also relates to nucleic acid characterisation using Msp.

WO 2012/107778 A2

Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WHITE, James** [GB/GB]; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB).

(74) Agent: **JA 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, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

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

MUTANT PORES

Field of the invention

The invention relates to mutant forms of Msp. The invention also relates to nucleic acid characterisation using Msp.

Background of the invention

Nanopore sensing is an approach to sensing that relies on the observation of individual binding events between analyte molecules and a receptor. Nanopore 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 identity of an analyte is revealed through its distinctive current signature, notably the duration and extent of current block and the variance of current levels.

There is currently a need for rapid and cheap nucleic acid (e.g. DNA or RNA) sequencing technologies across a wide range of applications. Existing technologies are slow and expensive mainly because they rely on amplification techniques to produce large volumes of nucleic acid and require a high quantity of specialist fluorescent chemicals for signal detection. Nanopore sensing has the potential to provide rapid and cheap nucleic acid sequencing by reducing the quantity of nucleotide and reagents required.

Two of the essential components of sequencing nucleic acids using nanopore sensing are (1) the control of nucleic acid movement through the pore and (2) the discrimination of nucleotides as the nucleic acid polymer is moved through the pore. In the past, to achieve nucleotide discrimination the nucleic acid has been passed through a mutant of hemolysin. This has provided current signatures that have been shown to be sequence dependent. It has also been shown that a large number of nucleotides contribute to the observed current, making a direct relationship between observed current and nucleic acid sequence challenging.

While the current range for nucleotide discrimination has been improved through mutation of the hemolysin pore, a sequencing system would have higher performance if the current differences between nucleotides could be improved further. In addition, it has been observed that when the nucleic acids are moved through a pore, some current states show high variance. It has also been shown that some mutant hemolysin pores exhibit higher variance than others. While the variance of these states may contain sequence specific information, it is desirable to produce pores that have low variance to simplify the system. It is also desirable to reduce the number of nucleotides that contribute to the observed current.

The different forms of Msp are porins from *Mycobacterium smegmatis*. MspA is a 157 kDa octameric porin from *Mycobacterium smegmatis*. The structure of MspA has been well documented by researchers (Gundlach, Proc Natl Acad Sci U S A. 2010 Sep 14; 107(37):16060-5. Epub 2010 Aug 26). Some key residues have been identified and modified to enhance the properties of the pore. These mutations have been performed to allow DNA to transition through the MspA pore. MspB, C and D are also known forms of Msp.

Summary of the invention

The inventors have surprisingly demonstrated that novel mutants of Msp display improved properties for estimating the characteristics, such as the sequence of nucleic acids. The mutants surprisingly display improved nucleotide discrimination. In particular, the mutants surprisingly display an increased current range, which makes it easier to discriminate between different nucleotides, and a reduced variance of states, which increases the signal-to-noise ratio. In addition, the number of nucleotides contributing to the current as the nucleic acid moves through the pore is decreased. This makes it easier to identify a direct relationship between the observed current as the nucleic acid moves through the pore and the nucleic acid sequence.

The inventors have also surprisingly shown that Msp shows improved sequencing properties when the movement of the nucleic acid through the pore is controlled by a Phi29 DNA polymerase. In particular, the coupling of Msp and Phi29 DNA polymerase results in three unexpected advantages. First, the nucleic acid moves through the pore at a rate that is commercially viable yet allows effective sequencing. Second, an increased current range is observed as the nucleic acid moves through the pore allowing the sequence to be determined more easily. Third, a decreased current variance is observed thereby increasing the signal-to-noise ratio.

Accordingly, the invention provides a mutant Msp monomer comprising a variant of the sequence shown in SEQ ID NO: 2, wherein the variant comprises at least one of the following mutations:

(a) asparagine (N), serine (S), glutamine (Q) or threonine (T) at position 88;

(b) serine (S), glutamine (Q) or tyrosine (Y) at position 90;

(c) leucine (L) or serine (S) at position 105;

(d) arginine (R) at position 126;

(e) serine (S) at position 75;

(f) serine (S) at position 77;

(g) arginine (R) at position 59;

(h) glutamine (Q) , asparagine (N) or threonine (T) at position 75;

(i) glutamine (Q) , asparagine (N) or threonine (T) at position 77;

(j) leucine (L) at position 78;

(k) asparagine (N) at position 81;

(l) asparagine (N) at position 83;

(m) serine (S) or threonine (T) at position 86;

(n) phenylalanine (F), valine (V) or leucine (L) at position 87;

(o) tyrosine (Y), phenylalanine (F), valine (V), arginine (R), alanine (A), glycine (G) or cysteine (C) at position 88;

(p) phenylalanine (F), valine (V) or leucine (L) at position 89;

(q) leucine (L), phenylalanine (F), tryptophan (W), histidine (H), threonine (T), glycine (G), alanine (A), valine (V), arginine (R), lysine (K), asparagine (N) or cysteine (C) at position 90;

(r) serine (S), glutamine (Q), leucine (L), methionine (M), isoleucine (I), alanine (A), valine (V), glycine (G), phenylalanine (F), tryptophan (W), tyrosine (Y), histidine (H), threonine (T), arginine (R), lysine (K), asparagine (N) or cysteine (C) at position 91;

(s) alanine (A) or serine (S) at position 92;

(t) serine (S), alanine (A), threonine (T), glycine (G) at position 93;

(u) leucine (L) at position 94;

(v) valine (V) at position 95;

(w) arginine (R), aspartic acid (D), valine (V), asparagine (N), serine (S) or threonine (T) at position 96;

(x) serine (S) at position 97;

(y) serine (S) at position 98;

(z) serine (S) at position 99;

(aa) serine (S) at position 100;

(bb) phenylalanine (F) at position 101;

(cc) lysine (K), serine (S) or threonine (T) at position 102;

(dd) alanine (A), glutamine (Q), asparagine (N), glycine (G) or threonine (T) at position 103;

(ee) isoleucine at position 104;

(ff) tyrosine (Y), alanine (A), glutamine (Q), asparagine (N), threonine (T), phenylalanine (F), tryptophan (W), histidine (H), glycine (G), valine (V), arginine (R), lysine (K), proline (P), or cysteine (C) at position 105;

(gg) phenylalanine (F), isoleucine (I), valine (V) or serine (S) at position 106;

(hh) proline (P) or serine (S) at position 108;

(ii) asparagine (N) at position 118;

(jj) serine (S) or cysteine (C) at position 103; and

(kk) cysteine at one or more of positions 10 to 15, 51 to 60, 136 to 139 and 168 to 172.

The invention also provides:

- a construct comprising two or more covalently attached monomers derived from Msp;
- a polynucleotide which encodes a mutant of the invention or a construct of the invention;
- a homo-oligomeric pore derived from Msp comprising identical mutant monomers of the invention;
- a hetero-oligomeric pore derived from Msp comprising at least one mutant monomer of the invention, wherein at least one of the eight monomers differs from the others;
- a method of characterising a target nucleic acid sequence, comprising:

(a) contacting the target sequence with a pore of the invention and a nucleic acid binding protein so that the protein controls the movement of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore; and

(b) measuring the current passing through the pore during each interaction and thereby characterising the target sequence;

- a kit for sequencing a target nucleic acid sequence comprising (a) a pore of the invention and (b) a nucleic acid handling enzyme;
- an apparatus for sequencing target nucleic acid sequences in a sample, comprising (a) a plurality of pores of the invention and (b) a plurality of nucleic acid handling enzymes;
- a method of characterising a target nucleic acid sequence, comprising:

(a) contacting the target sequence with a pore derived from Msp and a Phi29 DNA polymerase such that the polymerase controls the movement of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore; and

(b) measuring the current passing through the pore during each interaction and thereby characterising the target sequence, wherein steps (a) and (b) are carried out with a voltage applied across the pore;

- a method of forming a sensor for characterising a target nucleic acid sequence, comprising:

(a) contacting a pore derived from Msp with a Phi29 DNA polymerase in the presence of the target nucleic acid sequence; and

(b) applying a voltage across the pore to form a complex between the pore and the polymerase; and thereby forming a sensor for characterising the target nucleic acid sequence;

- a method of increasing the rate of activity of a Phi29 DNA polymerase, comprising:

(a) contacting the Phi29 DNA polymerase with a pore derived from Msp in the presence of a nucleic acid sequence; and

(b) applying a voltage across the pore to form a complex between the pore and the polymerase; and thereby increasing the rate of activity of a Phi29 DNA polymerase;

- a kit for characterising a target nucleic acid sequence comprising (a) a pore derived from Msp and (b) a Phi29 DNA polymerase; and
- an apparatus for characterising target nucleic acid sequences in a sample, comprising a plurality of pores derived from Msp and a plurality of Phi29 DNA polymerases.

Description of the Figures

Fig. 1 shows the average dwell time of individual current levels as a single DNA strand translocates the nanopore. The data is collated from a number of single molecules and is split into quartiles by current levels.

Fig. 2 shows current levels and variance obtained from using Phi29 in Unzipping mode to move a DNA strand (SEQ ID NO: 15) through the MS-$(NNNRRK)_8$ nanopore.

Fig. 3 shows current levels and variance obtained from using Phi29 in Unzipping mode to move a DNA strand (SEQ ID NO: 15) through the HL-$(mutant)_7$ nanopore.

Fig. 4 shows the current levels for a single MspA channel recorded at a range of applied potentials (-200 mV to 200 mV).

Fig. 5 shows the IV curve of open pore levels for the baseline MspA mutant, MS-(B1)8. Each line represents a single pore.

Fig.6 shows the IV curve of open pore levels for the MspA mutant, MS-(B1-I105Y)8. Each line represents a single pore.

Fig. 7 shows the IV curve of open pore levels for the MspA mutant, MS-(B1-I105N)8. Each line represents a single pore.

Fig. 8 shows the change in current between a high conductance state (275 pA) and a low conductance state (150 pA) for the MS-(B1-I105A)8 pore at 180 mV.

Fig. 9 shows the current levels produced when DNA is unzipped through the baseline MS-$(B1)_8$ pore. Current range for these events is ~ 30 pA.

Fig. 10 shows the current levels produced when DNA is unzipped through the baseline MS-$(B1\text{-}I105A)_8$ pore. Current range for these events is ~ 40 pA.

Fig. 11 shows the DNA substrate design used in Examples 9 and 12 and 15.

Fig. 12 shows the DNA substrate design used in Examples 10 and 11.

Fig. 13 shows how the sequencing profile changes, for the same DNA sequence, when point mutations are made in the MspA monomer sequence. These plots show the average of the profile of the levels obtained from multiple polynucleotides. A) This graph shows the sequencing profile for the MS-(B1)8 pore. B) This graph shows the sequencing profile for the MS-(B1-D90Q-D93S-I105A)8 pore. C) This graph shows the sequencing profile for the MS-(B1-D90Q-Q126R)8 pore. D) This graph shows the sequencing profile for the MS-(B1-L88N-D90Q-D91M)8 pore. E) This graph shows the sequencing profile for the MS-(B1-L88N-D90Q-D91S)8 pore. F) This graph shows the sequencing profile for the MS-(B1-G75S-G77S-L88N-Q126R)8 pore.

Fig. 14 shows the DNA substrate design used in Example 13.

Fig. 15 shows an example event trace for the controlled translocation of RNA, mediated by Phi29 DNA polymerase, through the MspA mutant pore MS-(B1)8. An expanded view, of the region highlighted in the upper trace, is shown below.

Fig. 16 shows pore insertion into the lipid bilayer. A) Shows pore insertion of the MS-(B1)8 oligomerised from the monomer. B) Shows pore insertion of the MS-(B1-B1)4 oligomerised from the dimer.

Fig. 17 shows an example event trace for the controlled translocation of DNA, mediated by a helicase, through the MS-(B1)8 mutant pore which was produced by oligomerisation of the monomer. An expanded view, of the region highlighted in the upper trace, is shown below.

Fig. 18 shows an example event trace for the controlled translocation of DNA, mediated by a helicase, through the MS-(B1-B1)4 mutant pore which was produced by oligomerisation of the dimer. An expanded view, of the region highlighted in the upper trace, is shown below.

Fig. 19 shows the DNA substrate design used in Example 16.

Fig. 20 shows an example event trace for the controlled translocation of DNA containing both cytosine and 5-methylcytosine, mediated by a helicase, through the MS-(B1-L88N)8 mutant pore. An expanded view of the region highlighted in the upper trace is shown below.

Description of the Sequence Listing

SEQ ID NO: 1 shows the polynucleotide sequence encoding the NNN-RRK mutant MspA monomer.

SEQ ID NO: 2 (also referred to as "B1") shows the amino acid sequence of the mature form of the NNN-RRK mutant of the MspA monomer. The mutant lacks the signal sequence and the amino terminal methionine (encoded by the start codon) and includes the following mutations: D90N, D91N, D93N, D118R, D134R and E139K. These mutations allow DNA transition through the MspA pore.

SEQ ID NO: 3 shows the polynucleotide sequence encoding the Phi29 DNA polymerase.

SEQ ID NO: 4 shows the amino acid sequence of the Phi29 DNA polymerase.

SEQ ID NO: 5 shows the codon optimised polynucleotide sequence derived from the *sbcB* gene from *E. coli*. It encodes the exonuclease I enzyme (EcoExo I) from *E. coli*.

SEQ ID NO: 6 shows the amino acid sequence of exonuclease I enzyme (EcoExo I) from *E. coli*.

SEQ ID NO: 7 shows the codon optimised polynucleotide sequence derived from the *xthA* gene from *E. coli*. It encodes the exonuclease III enzyme from *E. coli*.

SEQ ID NO: 8 shows the amino acid sequence of the exonuclease III enzyme from *E. coli*. This enzyme performs distributive digestion of 5' monophosphate nucleosides from one strand of double stranded DNA (dsDNA) in a 3' – 5' direction. Enzyme initiation on a strand requires a 5' overhang of approximately 4 nucleotides.

SEQ ID NO: 9 shows the codon optimised polynucleotide sequence derived from the *recJ* gene from *T. thermophilus*. It encodes the RecJ enzyme from *T. thermophilus* (*Tth*RecJ-cd).

SEQ ID NO: 10 shows the amino acid sequence of the RecJ enzyme from *T. thermophilus* (*Tth*RecJ-cd). This enzyme performs processive digestion of 5' monophosphate nucleosides from ssDNA in a 5' – 3' direction. Enzyme initiation on a strand requires at least 4 nucleotides.

SEQ ID NO: 11 shows the codon optimised polynucleotide sequence derived from the bacteriophage lambda *exo* (*redX*) gene. It encodes the bacteriophage lambda exonuclease.

SEQ ID NO: 12 shows the amino acid sequence of the bacteriophage lambda exonuclease. The sequence is one of three identical subunits that assemble into a trimer. The enzyme performs highly processive digestion of nucleotides from one strand of dsDNA, in a 5'-3'direction (http://www.neb.com/nebecomm/products/productM0262.asp). Enzyme initiation on a strand preferentially requires a 5' overhang of approximately 4 nucleotides with a 5' phosphate.

SEQ ID NOs: 13 to 15 show the sequences used in Example 2.

SEQ ID NOs: 16 to 18 show the amino acid sequences of the mature forms of the MspB, C and D mutants respectively. The mature forms lack the signal sequence.

SEQ ID NOs: 19 and 20 show the sequences used in Examples 9, 12 and 15.

SEQ ID NOs: 21 to 23 show the sequences used in Examples 10 and 11.

SEQ ID NOs: 24 to 27 show the sequences used in Example 13.

SEQ ID NO: 28 shows the DNA sequence of the dimer of the mature form of the NNN-RRK mutant of the MspA monomer used in Example 14.

SEQ ID NO: 29 shows the protein sequence of the dimer of the mature form of the NNN-RRK mutant of the MspA monomer used in Example 14.

SEQ ID NO: 30, 31 and 32 show the sequences used in Example 16.

SEQ ID NO: 33 shows the linker sequence shown used in the construct shown in SEQ ID NO: 29.

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 "a mutant" includes "mutants", reference to "a substitution" includes two or more such substitutions, reference to "a pore" includes two or more such pores, reference to "a nucleic acid sequence" includes two or more such sequences, and the like.

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

Mutant Msp monomers

The present invention provides mutant Msp monomers. The mutant Msp monomers may be used to form the pores of the invention. A mutant Msp monomer is a monomer whose sequence varies from that of a wild-type Msp monomer and which retains the ability to form a pore. Methods for confirming the ability of mutant monomers to form pores are well-known in the art and are discussed in more detail below.

The mutant monomers have improved nucleotide reading properties i.e. display improved nucleotide capture and discrimination. In particular, pores constructed from the mutant monomers capture nucleotides and nucleic acids more easily than the wild type. In addition, pores constructed from the mutant monomers display an increased current range, which makes it easier to discriminate between different nucleotides, and a reduced variance of states, which increases the signal-to-noise ratio. In addition, the number of nucleotides contributing to the current as the nucleic acid moves through pores constructed from the mutants is decreased. This makes it easier to identify a direct relationship between the observed current as the nucleic acid moves through the pore and the nucleic acid sequence. The improved nucleotide reading properties of the mutants are achieved via five main mechanisms, namely by changes in the:

- sterics (increasing or decreasing the size of amino acid residues);
- charge (e.g. introducing +ve charge to interact with the nucleic acid sequence);
- hydrogen bonding (e.g. introducing amino acids that can hydrogen bond to the base pairs);
- pi stacking (e,g, introducing amino acids that interact through delocalised electron pi systems); and/or
- alteration of the structure of the pore (e.g. introducing amino acids that increase the size of the vestibule and/or constriction).

Any one or more of these five mechanisms may be responsible for the improved properties of the pores of the invention. For instance, a pore of the invention may display improved nucleotide reading properties as a result of altered sterics, altered hydrogen bonding and an altered structure.

The introduction of bulky residues, such as phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H), increases the sterics of the pore. The introduction of aromatic residues, such as phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H), also increases the pi staking in the pore. The introduction of bulky or aromatic residues also alters the structure of the pore, for instance by opening up the pore and increasing the size of the vestibule and/or constriction. This is described in more detail below.

A mutant monomer of the invention comprises a variant of the sequence shown in SEQ ID NO: 2. SEQ ID NO: 2 is the NNN-RRK mutant of the MspA monomer. It includes the following mutations: D90N, D91N, D93N, D118R, D134R and E139K. A variant of SEQ ID NO: 2 is a polypeptide that has an amino acid sequence which varies from that of SEQ ID NO: 2 and which retains its ability to form a pore.

The variant comprises at least one of the following mutations:

(a) asparagine (N), serine (S), glutamine (Q) or threonine (T) at position 88;
(b) serine (S), glutamine (Q) or tyrosine (Y) at position 90;
(c) leucine (L) or serine (S) at position 105;
(d) arginine (R) at position 126;
(e) serine (S) at position 75;
(f) serine (S) at position 77;
(g) arginine (R) at position 59;
(h) glutamine (Q) , asparagine (N) or threonine (T) at position 75;
(i) glutamine (Q) , asparagine (N) or threonine (T) at position 77;
(j) leucine (L) at position 78;
(k) asparagine (N) at position 81;

(l) asparagine (N) at position 83;

(m) serine (S) or threonine (T) at position 86;

(n) phenylalanine (F), valine (V) or leucine (L) at position 87;

(o) tyrosine (Y), phenylalanine (F), valine (V), arginine (R), alanine (A), glycine (G) or cysteine (C) at position 88;

(p) phenylalanine (F), valine (V) or leucine (L) at position 89;

(q) leucine (L), phenylalanine (F), tryptophan (W), histidine (H), threonine (T), glycine (G), alanine (A), valine (V), arginine (R), lysine (K), asparagine (N) or cysteine (C) at position 90;

(r) serine (S), glutamine (Q), leucine (L), methionine (M), isoleucine (I), alanine (A), valine (V), glycine (G), phenylalanine (F), tryptophan (W), tyrosine (Y), histidine (H), threonine (T), arginine (R), lysine (K), asparagine (N) or cysteine (C) at position 91;

(s) alanine (A) or serine (S) at position 92;

(t) serine (S), alanine (A), threonine (T), glycine (G) at position 93;

(u) leucine (L) at position 94;

(v) valine (V) at position 95;

(w) arginine (R), aspartic acid (D), valine (V), asparagine (N), serine (S) or threonine (T) at position 96;

(x) serine (S) at position 97;

(y) serine (S) at position 98;

(z) serine (S) at position 99;

(aa) serine (S) at position 100;

(bb) phenylalanine (F) at position 101;

(cc) lysine (K), serine (S) or threonine (T) at position 102;

(dd) alanine (A), glutamine (Q), asparagine (N), glycine (G) or threonine (T) at position 103;

(ee) isoleucine at position 104;

(ff) tyrosine (Y), alanine (A), glutamine (Q), asparagine (N), threonine (T), phenylalanine (F), tryptophan (W), histidine (H), glycine (G), valine (V), arginine (R), lysine (K), proline (P), or cysteine (C) at position 105;

(gg) phenylalanine (F), isoleucine (I), valine (V) or serine (S) at position 106;

(hh) proline (P) or serine (S) at position 108;

(ii) asparagine (N) at position 118;

(jj) serine (S) or cysteine (C) at position 103; and

(kk) cysteine at one or more of positions 10 to 15, 51 to 60, 136 to 139 and 168 to 172.

In wild-type MspA, residues 88 and 105 in each monomer form a hydrophobic ring in the inner constriction of the pore. The hydrophobic residues at positions L88 and I105 sit just above the main constriction of the pore, facing into the aqueous channel. Mutation of these residues produces pores that have significantly higher open pore currents to the baseline (SEQ ID NO: 2). The current differences observed when mutations are made at these positions are significantly higher than would be expected from making a single mutation. This surprising result implies that mutations at these positions may have an effect on the structure of the channel rather than just the local environment at these residues. Although the SEQ ID NO: 2 baseline has been reported to exhibit a wide range of pore conductance, the reason for this is not well understood. Mutations to positions L88 and I105 result in the dominant pore current level being significantly higher than the baseline pore. In addition, this higher conductance state is the dominant conformation of the mutant, which is desirable for a large current range and increased signal to noise.

The introduction of N, S, Q or T at position 88 (i.e. mutation (a) above) introduces into the inner constriction of the pore an amino acid that can hydrogen bond to the nucleotides in a nucleic acid.

Residues 90 and 91 in each monomer also form part of the inner constriction of the pore. Residue 118 in each monomer is present within the vestibule of the pore. Residue 134 in each monomer is part of the entrance to the pore.

The introduction of S, Q or Y at position 90 (i.e. mutation (b) above) introduces into the inner constriction of the pore an amino acid that can hydrogen bond to the nucleotides in a nucleic acid.

The variant may include any number of mutations (a) to (kk), such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more of the mutations. Preferred combinations of mutations are discussed below. The amino acids introduced into the variant may be naturally-occuring or non-naturally occurring derivatives thereof. The amino acids introduced into the variant may be D-amino acids.

Any number of cysteines may be introduced into the variant. Cysteines are preferably introduced at one or more, such as two or all of, positions 90, 91 and 103. These positions may be useful for chemical attachment of a molecular adaptor as discussed in more detail below. Any number of cysteines, such as 2, 3, 4, 5, 6 or more cysteines, may be introduced at positions 10 to 15, 51 to 60, 136 to 139 and 168 to 172. These positions are present in non-conserved loop regions of the pore and so are useful for chemically attaching a nucleic acid binding protein to the pore as discussed in more detail below.

In a preferred embodiment, the variant comprises one or more of the substitutions shown in (A) to (Z) below. The variant may include any number of the substitutions in A to Z, such as 1, 2, 3, 4 or 5.

(A) The introduction of one or more of (i) serine (S) at positon 75, (ii) serine (S) at position 77, (iii) asparagine (N) at position 88, (iv) glutamine (Q) at position 90 and (v) arginine (R) at position 126. The variant may include 1, 2, 3, 4 or 5 of these substitutions. The advantages of homo-octameric pores including all four substitutions in each monomer are shown in Table 3 below.

(B) The introduction of one or more of (i) glutamine (Q) at position 90 and (ii) arginine (R) at position 126. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 3 below.

(C) The introduction of one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) arginine (R) at position 126. The variant may include 1, 2 or 3 of these substitutions. The advantages of homo-octameric pores including all three of these substitutions in each monomer are shown in Table 3 below.

(D) The introduction of one or more of (i) serine (S) at position 88 and (ii) glutamine (Q) at position 90. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 3 below.

(E) The introduction of one or more of (i) asparagine (N) at position 88 and (ii) glutamine (Q) at position 90. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 3 below.

(F) The introduction of one or more of (i) glutamine (Q) at position 90 and (ii) alanine (A) at position 105. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(G) The introduction of one or more of (i) serine (S) at position 90 and (ii) serine (S) at position 92. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(H) The introduction of one or more of (i) threonine (T) at position 88 and (ii) serine (S) at position 90. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(I) The introduction of one or more of (i) glutamine (Q) at position 87 and (ii) serine (S) at position 90. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(J) The introduction of one or more of (i) tyrosine (Y) at position 89 and (ii) serine (S) at position 90. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(K) The introduction of one or more of (i) asparagine (N) at position 88 and (ii) phenylalanine (F) at position 89. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(L) The introduction of one or more of (i) asparagine (N) at position 88 and (ii) tyrosine (Y) at position 89. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(M) The introduction of one or more of (i) serine (S) at position 90 and (ii) alanine (A) at position 92. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(N) The introduction of one or more of (i) serine (S) at position 90 and (ii) asparagine (N) at position 94. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(O) The introduction of one or more of (i) serine (S) at position 90 and (ii) isoleucine (I) at position 104. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(P) The introduction of one or more of (i) aspartic acid (D) at position 88 and (ii) lysine (K) at position 105. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(Q) The introduction of one or more of (i) asparagine (N) at position 88 and (ii) arginine (R) at position 126. The variant may include 1 or 2 of these substitutions. The advantages of homo-octameric pores including both substitutions in each monomer are shown in Table 2 below.

(R) The one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) arginine (R) at position 91. The variant may include 1, 2 or 3 of these substitutions. The advantages of homo-octameric pores including all three substitutions in each monomer are shown in Table 2 below.

(S) The introduction of or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) serine (S) at position 91. The variant may include 1, 2 or 3 of these substitutions. The advantages of homo-octameric pores including all three substitutions in each monomer are shown in Table 2 below.

(T) The introduction of one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) valine (V) at position 105. The variant may include 1, 2 or 3 of these substitutions. The advantages of homo-octameric pores including all three substitutions in each monomer are shown in Table 2 below.

(U) The introduction of one or more of (i) glutamine (Q) at position 90, (ii) serine (S) at position 93 and (iii) alaine (A) at position 105. The variant may include 1, 2 or 3 of these substitutions. The advantages of homo-octameric pores including all three substitutions in each monomer are shown in Table 2 below.

(V) The introduction of one or more of (i) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) at position 90, (ii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) at position 91 and (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) at position 105. The variant may include 1, 2 or 3 of these substitutions. The introduction of these bulky, aromatic residues increases the sterics and pi stacking in the vestibule and/or constriction of the pore. They also increase the size of the vestibule and/or constriction (i.e. open up the pore).

(W) The introduction of one or more of (i) serine (S), threonine (T), glycine (G), alanine (A) or valine (V) at position 90, (ii) serine (S), threonine (T), glycine (G), alanine (A) or valine (V) at position 91 and (iii) serine (S), threonine (T), glycine (G), alanine (A) or valine (V) at position 105. The variant may include 1, 2 or 3 of these substitutions. The introduction of smaller residues decreases the sterics in the vestibule and/or constriction of the pore.

(X) The introduction of serine (S), arginine (R), lysine (K) or histidine (H) at position 90 and/or serine (S), arginine (R), lysine (K) or histidine (H) at position 91. The introduction of positively-charged residues (R, K or H) increases the interactions between the constriction of the pore and the nucleic acid sequence.

(Y) The introduction of serine (S), threonine (T), asparagine (N), glutamine (Q), tyrosine (Y) or histidine (H) at position 90 and/or serine (S), threonine (T), asparagine (N), glutamine (Q), tyrosine (Y) or histidine (H) at position 91. The introduction of these residues

increases the hydrogen bonding that occurs between the constriction of the pore and the nucleic acid sequence. They also increase the size of the vestibule and/or constriction (i.e. open up the pore).

(Z) The introduction of cysteine at one or more of positions 90, 91and 103. This allows chemical groups to be attached to the pore via cysteine linkage. This is discussed in more detail above and below.

Preferred variants include, but are not limited to, those comprising at least one of the following substitution(s): L88N; L88S; L88Q; L88T; D90S; D90Q; D90Y; I105L; I105S; Q126R; G75S; G77S; G75S, G77S, L88N and Q126R; G75S, G77S, L88N, D90Q and Q126R; D90Q and Q126R; L88N, D90Q and Q126R; L88S and D90Q; L88N and D90Q; E59R; G75Q; G75N; G75S; G75T; G77Q; G77N; G77S; G77T; I78L; S81N; T83N; N86S; N86T; I87F; I87V; I87L; L88N; L88S; L88Y; L88F; L88V; L88Q; L88T; I89F; I89V; I89L; N90S; N90Q; N90L; N90Y; N91S; N91Q; N91L; N91M; N91I; N91A; N91V; N91G; G92A; G92S; N93S; N93A; N93T; I94L; T95V; A96R; A96D; A96V; A96N; A96S; A96T; P97S; P98S; F99S; G100S; L101F; N102K; N102S; N102T; S103A; S103Q; S103N; S103G; S103T; V104I; I105Y; I105L; I105A; I105Q; I105N; I105S; I105T; T106F; T106I; T106V; T106S; N108P; N108S; D90Q and I105A; D90S and G92S; L88T and D90S; I87Q and D90S; I89Y and D90S; L88N and I89F; L88N and I89Y; D90S and G92A; D90S and I94N; D90S and V104I; L88D and I105K; L88N and Q126R; L88N, D90Q and D91R; L88N, D90Q and D91S; L88N, D90Q and I105V; D90Q, D93S and I105A; N91Y; N90Y and N91G; N90G and N91Y; N90G and N91G; I05G; N90R; N91R; N90R and N91R; N90K; N91K; N90K and N91K; N90Q and N91G; N90G and N91Q; N90Q and N91Q; R118N; N91C; N90C; N90W; N91W; N90K; N91K; N90R; N91R; N90S and N91S; N90Y and I105A; N90G and I105A; N90Q and I105A; N90S and I105A; L88A and I105A; L88S and I105S; L88N and I105N; N90G and N93G; N90G; N93G; N90G and N91A; I105K; I105R; I105V; I105P; I105W; L88R; L88A; L88G; L88N; N90R and I105A; N90S and I105A; L88A and I105A; L88S and I105S; L88N and I105N; L88C; S103C; and I105C..

A particularly preferred variant comprises I105N. Pores constructed from mutant monomers comprising I105N have a residual current that is increased by approximately 80%. The change in current in relation to different nucleotides is also increased. This reflects a change in structure of pores constructed from mutant monomers comprising I105N. Such pores therefore have an improved ability to discriminate nucleotides.

Preferred single mutants and their advantages when used in homo-octameric pores are shown in Table 1 below.

Table 1

Position	Substitution	Advantage(s)
E59	R	Increased DNA capture
G75	Q	Less variance
G75	N	Less variance
G75	S	Less variance
G75	T	Less variance
G77	Q	Less variance
G77	N	Less variance
G77	S	Less variance
G77	T	Less variance
I78	L	Tighter distribution of pore sizes
S81	N	More stable beta barrel
T83	N	More stable beta barrel
N86	S	Larger DNA range
N86	T	Larger DNA range
I87	F	Less variance
I87	V	Less variance
I87	L	Less variance
L88	N	Less variance
L88	S	Less variance
L88	Y	Altered DNA-noise profile
L88	F	Altered DNA-noise profile
L88	V	Less variance
L88	Q	Less variance
L88	T	Larger DNA range Less variance
I89	F	Larger DNA range Altered DNA recognition
I89	V	Tighter distribution of pore sizes
I89	L	Tighter distribution of pore sizes
N90	S	Less variance Altered DNA recognition
N90	Q	Increased pore current Increased DNA range Altered DNA recognition
N90	L	Altered DNA recognition
N90	Y	Altered DNA recognition
N91	S	Altered DNA recognition
N91	Q	Altered DNA recognition
N91	L	Altered DNA recognition
N91	M	Altered DNA recognition
N91	I	Altered DNA recognition
N91	A	Altered DNA recognition
N91	V	Altered DNA recognition
N91	G	Altered DNA recognition
G92	A	Larger DNA range.
G92	S	Stabilises D90S mutations
N93	S	Larger DNA range.
N93	A	Larger DNA range.

N93	T	Larger DNA range Altered DNA discrimination
I94	L	Larger DNA range.
T95	V	Stable open pore current level Altered DNA range
A96	R	Increased DNA capture
A96	D	Altered DNA recognition
A96	V	Good pores Altered pore variance
A96	N	Good pores Altered pore variance
A96	S	Good pores Altered pore variance
A96	T	Good pores Altered pore variance
P97	S	Good pores Altered pore variance
P98	S	Good pores Altered pore variance
F99	S	Good pores Altered pore variance
G100	S	Good pores Altered pore variance
L101	F	Good pores Altered pore variance
N102	K	Increased DNA capture
N102	S	Altered pore variance profil Larger DNA range
N102	T	Altered DNA discrimination.
S103	A	Altered DNA recognition
S103	Q	Larger DNA range Altered DNA recognition
S103	N	Larger DNA range Altered DNA recognition
S103	G	Larger DNA range Altered DNA recognition
S103	T	Altered DNA discrimination.
V104	I	Altered DNA discrimination.
I105	Y	Larger DNA range Altered DNA recognition
I105	L	Larger DNA range Less variance
I105	A	Larger DNA range.
I105	Q	Altered DNA recognition
I105	N	Larger DNA range
I105	S	Altered DNA recognition
I105	T	Altered DNA recognition
T106	F	Stable open pore current level Altered DNA variance
T106	I	Altered DNA recognition
T106	V	Altered DNA recognition

T106	S	Larger DNA range
N108	P	Stable open pore current level Altered DNA variance
N108	S	Stable open pore current level Altered DNA variance
Q126	R	Increased DNA capture

Preferred multiple mutants and their advantages when used in homo-octameric pores are shown in Table 2 below.

Table 2

Mutant	**Advantage(s)**
D90Q/I105A	Altered DNA recognition
D90S/G92S	Altered DNA recognition
L88T/D90S	Altered DNA recognition
I87Q/D90S	Altered DNA recognition
I89Y/D90S	Altered DNA recognition
L88N/I89F	Altered DNA recognition.
L88N/I89Y	Altered DNA recognition Larger DNA range
D90S/G92A	Altered DNA recognition
D90S/I94N	Altered DNA recognition
D90S/V104I	Altered DNA recognition
L88D/I105K	Altered DNA recognition
L88N/Q126R	Less variance Increased DNA capture
L88N/D90Q/D91R	Altered DNA discrimination Increased DNA capture
L88N/D90Q/D91S	Altered DNA discrimination
L88N/D90Q/I105V	Altered DNA discrimination
D90Q-D93S-I105A	Altered DNA discrimination

The most preferred mutants and their advantages when used in homo-octameric pores are shown in the Table 3 below.

Table 3 – Most preferred mutants and their advantages

Mutant	**Advantage(s)**
G75S/G77S/L88N/Q126R (ONT Ref: B2C)	Stable open pore current Increased DNA capture Less variance
G75S/G77S/L88N/D90Q/Q126R	Stable open pore current Increased DNA capture Less variance Altered DNA recognition
L88N	Less variance

D90Q/Q126R	Increased pore current Increased DNA range Altered DNA recognition Increased DNA capture
L88N/D90Q/Q126R	Increased pore current Increased DNA range Altered DNA recognition Increased DNA capture
L88S/D90Q	Less variance Altered DNA recognition
D90S	Less variance Altered DNA recognition
D90Q	Increased pore current Increased DNA range Altered DNA recognition
L88S	Less variance
L88Q	Less variance
L88N/D90Q	Increased pore current Increased DNA range Altered DNA recognition
I105L	Less variance. Large DNA range.
I105S	Large DNA range.

In addition to the specific mutations discussed above, the variant may include other mutations. Over the entire length of the amino acid sequence of SEQ ID NO: 2, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant 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 100 or more, for example 125, 150, 175 or 200 or more, contiguous amino acids ("hard homology").

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.

SEQ ID NO: 2 is the NNN-RRK mutant of the MspA monomer. The variant may comprise any of the mutations in the MspB, C or D monomers compared with MspA. The mature forms of MspB, C and D are show in SEQ ID NOs: 16 to 18. In particular, the variant may comprise the following substitution present in MspB: A138P. The variant may comprise one or more of the following substitutions present in MspC: A96G, N102E and A138P. The variant may comprise one or more of the following mutations present in MspD: Deletion of G1, L2V, E5Q, L8V, D13G, W21A, D22E, K47T, I49H, I68V, D91G, A96Q, N102D, S103T, V104I, S136K and G141A. The variant may comprise combinations of one or more of the mutations and substitutions from Msp B, C and D.

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions replace amino acids with other amino acids of similar chemical structure, similar chemical properties or similar side-chain volume. The amino acids introduced may have similar polarity, hydrophilicity, hydrophobicity, basicity, acidity, neutrality or charge

to the amino acids they replace. Alternatively, the conservative substitution may introduce another amino acid that is aromatic or aliphatic in the place of a pre-existing aromatic or aliphatic amino acid. Conservative amino acid changes are well-known in the art and may be selected in accordance with the properties of the 20 main amino acids as defined in Table 4 below. Where amino acids have similar polarity, this can also be determined by reference to the hydropathy scale for amino acid side chains in Table 5.

Table 4 – Chemical properties of amino acids

Ala	aliphatic, hydrophobic, neutral	Met	hydrophobic, neutral
Cys	polar, hydrophobic, neutral	Asn	polar, hydrophilic, neutral
Asp	polar, hydrophilic, charged (-)	Pro	hydrophobic, neutral
Glu	polar, hydrophilic, charged (-)	Gln	polar, hydrophilic, neutral
Phe	aromatic, hydrophobic, neutral	Arg	polar, hydrophilic, charged (+)
Gly	aliphatic, neutral	Ser	polar, hydrophilic, neutral
His	aromatic, polar, hydrophilic, charged (+)	Thr	polar, hydrophilic, neutral
Ile	aliphatic, hydrophobic, neutral	Val	aliphatic, hydrophobic, neutral
Lys	polar, hydrophilic, charged(+)	Trp	aromatic, hydrophobic, neutral
Leu	aliphatic, hydrophobic, neutral	Tyr	aromatic, polar, hydrophobic

Table 5 - Hydropathy scale

Side Chain	Hydropathy
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5

Asn	-3.5
Lys	-3.9
Arg	-4.5

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

Variants may include fragments of SEQ ID NO: 2. Such fragments retain pore forming activity. Fragments may be at least 50, 100, 150 or 200 amino acids in length. Such fragments may be used to produce the pores of the invention. A fragment preferably comprises the pore forming domain of SEQ ID NO: 2. Fragments must include one of residues 88, 90, 91, 105, 118 and 134 of SEQ ID NO: 2. Typically, fragments include all of residues 88, 90, 91, 105, 118 and 134 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 amino terminal or carboxy terminal 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. Other fusion proteins are discussed in more detail below. A variant may have a methionine at the amino terminal of SEQ ID NO: 2.

As discussed above, a variant is a polypeptide that has an amino acid sequence which varies from that of SEQ ID NO: 2 and which retains its ability to form a pore. A variant typically contains the regions of SEQ ID NO: 2 that are responsible for pore formation. The pore forming ability of Msp, which contains a β-barrel, is provided by β-sheets in each subunit. A variant of SEQ ID NO: 2 typically comprises the regions in SEQ ID NO: 2 that form β-sheets. One or more modifications can be made to the regions of SEQ ID NO: 2 that form β-sheets as long as the resulting variant retains its ability to form a pore. A variant of SEQ ID NO: 2 preferably includes one or more modifications, such as substitutions, additions or deletions, within its α-helices and/or loop regions.

The mutant monomers may be modified to assist their identification or purification, for example by the addition of histidine residues (a his tag), aspartic acid residues (an asp tag), a streptavidin tag or a flag tag, 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. An alternative to introducing a genetic tag is to chemically react a tag onto a native or engineered position on the pore. An example of this would be to react a gel-shift reagent to a cysteine engineered on the

outside of the pore. This has been demonstrated as a method for separating hemolysin hetero-oligomers (Chem Biol. 1997 Jul;4(7):497-505).

The mutant monomer 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 mutant monomer may be made synthetically or by recombinant means. For example, the pore may be synthesized by *in vitro* translation and transcription (IVTT). The amino acid sequence of the mutant monomer may be modified to include non-naturally occurring amino acids or to increase the stability of the monomer. When the mutant monomer is produced by synthetic means, such amino acids may be introduced during production. The mutant monomer may also be altered following either synthetic or recombinant production.

The mutant monomer may also be produced using D-amino acids. For instance, the mutant monomer may comprise a mixture of L-amino acids and D-amino acids. This is conventional in the art for producing such proteins or peptides.

The mutant monomer contains one or more specific modifications to facilitate nucleotide discrimination. The mutant monomer may also contain other non-specific modifications as long as they do not interfere with pore formation. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the mutant monomer. 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.

The mutant monomer can be produced using standard methods known in the art. Polynucleotide sequences encoding a mutant monomer may be derived and replicated using standard methods in the art. Such sequences are discussed in more detail below. Polynucleotide sequences encoding a mutant monomer may be expressed in a bacterial host cell using standard techniques in the art. The mutant monomer 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.

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

in accordance with the invention. Methods for inserting pore into membranes are discussed below.

In some embodiments, the mutant monomer is chemically modified. The mutant monomer can be chemically modified in any way and at any site. The mutant monomer is preferably chemically modified by attachment of a molecule to one or more cysteines (cysteine linkage), attachment of a molecule to one or more lysines, attachment of a molecule to one or more non-natural amino acids, enzyme modification of an epitope or modification of a terminus. Suitable methods for carrying out such modifications are well-known in the art. The mutant monomer may be chemically modified by the attachment of any molecule. For instance, the mutant monomer may be chemically modified by attachment of a dye or a fluorophore.

In some embodiments, the mutant monomer is chemically modified with a molecular adaptor that facilitates the interaction between a pore comprising the monomer and a target nucleotide or target nucleic acid sequence. The presence of the adaptor improves the host-guest chemistry of the pore and the nucleotide or nucleic acid sequence and thereby improves the sequencing ability of pores formed from the mutant monomer. 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 nucleotide or nucleic acid sequence. The adaptor may alter the charge of the barrel or channel of the pore or specifically interact with or bind to the nucleotide or nucleic acid sequence thereby facilitating its interaction with the pore.

The molecular adaptor is preferably a cyclic molecule, a cyclodextrin, a species that is capable of hybridization, a DNA binder or interchelator, a peptide or peptide analogue, a synthetic polymer, an aromatic planar molecule, a small positively-charged molecule or a small molecule capable of hydrogen-bonding.

The adaptor may be cyclic. A cyclic adaptor preferably has the same symmetry as the pore. The adaptor preferably has eight-fold symmetry since Msp typically has eight subunits around a central axis. This is discussed in more detail below.

The adaptor typically interacts with the nucleotide or nucleic acid sequence via host-guest chemistry. The adaptor is typically capable of interacting with the nucleotide or nucleic acid sequence. The adaptor comprises one or more chemical groups that are capable of interacting with the nucleotide or nucleic acid sequence. The one or more chemical groups preferably interact with the nucleotide or nucleic acid sequence 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 nucleotide or nucleic acid sequence are preferably positively charged. The

one or more chemical groups that are capable of interacting with the nucleotide or nucleic acid sequence more preferably comprise amino groups. The amino groups can be attached to primary, secondary or tertiary carbon atoms. The adaptor even more preferably comprises a ring of amino groups, such as a ring of 6, 7 or 8 amino groups. The adaptor most preferably comprises a ring of eight amino groups. A ring of protonated amino groups may interact with negatively charged phosphate groups in the nucleotide or nucleic acid sequence.

The correct positioning of the adaptor within the pore can be facilitated by host-guest chemistry between the adaptor and the pore comprising the mutant monomer. 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 or amines. 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. Any adaptor that that facilitates the interaction between the pore and the nucleotide or nucleic acid sequence can be used.

Suitable adaptors include, but are not limited to, cyclodextrins, cyclic peptides and cucurbiturils. The adaptor is preferably a cyclodextrin or a derivative thereof. The cyclodextrin or derivative thereof may be any of those disclosed in Eliseev, A. V., and Schneider, H-J. (1994) *J. Am. Chem. Soc.* 116, 6081-6088. The adaptor is more preferably heptakis-6-amino-β-cyclodextrin (am_7-βCD), 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1-βCD) or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu_7-βCD). The guanidino group in gu_7-βCD has a much higher pKa than the primary amines in am_7-βCD and so it more positively charged. This gu_7-βCD adaptor may be used to increase the dwell time of the nucleotide in the pore, to increase the accuracy of the residual current measured, as well as to increase the base detection rate at high temperatures or low data acquisition rates.

If a succinimidyl 3-(2-pyridyldithio)propionate (SPDP) crosslinker is used as discussed in more detail below, the adaptor is preferably heptakis(6-deoxy-6-amino)-6-N-mono(2-pyridyl)dithiopropanoyl-β-cyclodextrin (am_6amPDP_1-βCD).

More suitable adaptors include γ-cyclodextrins, which comprise 8 sugar units (and therefore have eight-fold symmetry). The γ-cyclodextrin may contain a linker molecule or may be modified to comprise all or more of the modified sugar units used in the β-cyclodextrin examples discussed above.

The molecular adaptor is preferably covalently attached to the mutant monomer. The adaptor can be covalently attached to the pore using any method known in the art. The adaptor is typically attached via chemical linkage. If the molecular adaptor is attached via cysteine linkage, the one or more cysteines have preferably been introduced to the mutant by substitution. The mutant monomers of the invention can of course comprise a cysteine residue at one or more of positions 88, 90, 91, 103 and 105. The mutant monomer may be chemically modified by attachment of a molecular adaptor to one or more, such as 2, 3, 4 or 5, of these cysteines. Alternatively, the mutant monomer may be chemically modified by attachment of a molecule to one or more cysteines introduced at other positions. The molecular adaptor is preferably attached to one or more of positions 90, 91 and 103 of SEQ ID NO: 2.

The reactivity of cysteine residues may be enhanced by modification of the adjacent residues. For instance, the basic groups of flanking arginine, histidine or lysine residues will change the pKa of the cysteines thiol group to that of the more reactive S^- group. The reactivity of cysteine residues may be protected by thiol protective groups such as dTNB. These may be reacted with one or more cysteine residues of the mutant monomer before a linker is attached. The molecule may be attached directly to the mutant monomer. The molecule is preferably attached to the mutant monomer using a linker, such as a chemical crosslinker or a peptide linker.

Suitable chemical crosslinkers are well-known in the art. Preferred crosslinkers include 2,5-dioxopyrrolidin-1-yl 3-(pyridin-2-yldisulfanyl)propanoate, 2,5-dioxopyrrolidin-1-yl 4-(pyridin-2-yldisulfanyl)butanoate and 2,5-dioxopyrrolidin-1-yl 8-(pyridin-2-yldisulfanyl)octananoate. The most preferred crosslinker is succinimidyl 3-(2-pyridyldithio)propionate (SPDP). Typically, the molecule is covalently attached to the bifunctional crosslinker before the molecule/crosslinker complex is covalently attached to the mutant monomer but it is also possible to covalently attach the bifunctional crosslinker to the monomer before the bifunctional crosslinker/monomer complex is attached to the molecule.

The linker is preferably resistant to dithiothreitol (DTT). Suitable linkers include, but are not limited to, iodoacetamide-based and Maleimide-based linkers.

In other embodiment, the monomer may be attached to a nucleic acid binding protein. This forms a modular sequencing system that may be used in the methods of sequencing of the invention. Nucleic acid binding proteins are discussed below.

The nucleic acid binding protein is preferably covalently attached to the mutant monomer. The protein can be covalently attached to the pore using any method known in the art. The monomer and protein may be chemically fused or genetically fused. The monomer and protein are genetically fused if the whole construct is expressed from a single polynucleotide

sequence. Genetic fusion of a pore to a nucleic acid binding protein is discussed in International Application No. PCT/GB09/001679 (published as WO 2010/004265).

If the nucleic acid binding protein is attached via cysteine linkage, the one or more cysteines have preferably been introduced to the mutant by substitution. The mutant monomers of the invention can of course comprise cysteine residues at one or more of positions 10 to 15, 51 to 60, 136 to 139 and 168 to 172. These positions are present in loop regions which have low conservation amongst homologues indicating that mutations or insertions may be tolerated. They are therefore suitable for attaching a nucleic acid binding protein. The reactivity of cysteine residues may be enhanced by modification as described above.

The nucleic acid binding protein may be attached directly to the mutant monomer or via one or more linkers. The molecule may be attached to the mutant monomer using the hybridization linkers described in International Application No. PCT/GB10/000132 (published as WO 2010/086602). Alternatively, peptide linkers may be used. Peptide linkers are amino acid sequences. The length, flexibility and hydrophilicity of the peptide linker are typically designed such that it does not to disturb the functions of the monomer and molecule. Preferred flexible peptide linkers are stretches of 2 to 20, such as 4, 6, 8, 10 or 16, serine and/or glycine amino acids. More preferred flexible linkers include $(SG)_1$, $(SG)_2$, $(SG)_3$, $(SG)_4$, $(SG)_5$ and $(SG)_8$ wherein S is serine and G is glycine. Preferred rigid linkers are stretches of 2 to 30, such as 4, 6, 8, 16 or 24, proline amino acids. More preferred rigid linkers include $(P)_{12}$ wherein P is proline.

The mutant monomer may be chemically modified with a molecular adaptor and a nucleic acid binding protein.

Constructs

The invention also provides a construct comprising two or more covalently attached monomers derived from Msp. The construct of the invention retains its ability to form a pore. One or more constructs of the invention may be used to form pores for characterising, such as sequencing, nucleic acids sequences. The construct may comprise 2, 3, 4, 5, 6, 7, 8, 9 or 10 monomers. The two or more monomers may be the same or different.

The monomers do not have to be mutant monomers of the invention. For instance, at least one monomer may comprise the sequence shown in SEQ ID NO: 2. Alternatively, at least one monomer may comprise a variant of SEQ ID NO: 2 which is at least 50% homologous to SEQ ID NO: 2 over its entire sequence based on amino acid identity, but does not include any of the specific mutations required by the mutant monomers of the invention. More preferably, the variant 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.In a preferred embodiment, at least one monomer in the construct is a mutant monomer of the invention. All of the monomers in the construct may be a mutant monomer of the invention. The mutant monomers may be the same or different. In a more preferred embodiment, the construct comprises two monomers and at least one of the monomers is a mutant monomer of the invention.

The monomers are preferably genetically fused. Monomers are genetically fused if the whole construct is expressed from a single polynucleotide sequence. The coding sequences of the monomers may be combined in any way to form a single polynucleotide sequence encoding the construct.

The monomers may be genetically fused in any configuration. The monomers may be fused via their terminal amino acids. For instance, the amino terminus of the one monomer may be fused to the carboxy terminus of another monomer. If the construct is formed from the genetic fusion of two or more monomers each comprising the sequence shown in SEQ ID NO: 2 or a variant thereof, the second and subsequent monomers in the construct (in the amino to carboxy direction) may comprise a methionine at their amino terminal ends (each of which is fused to the carboxy terminus of the previous monomer). For instance, if M is a monomer comprising the sequence shown in SEQ ID NO: 2 or a variant (without an amino terminal methionine) and mM is a monomer comprising the sequence shown in SEQ ID NO: 2 or a variant with an amino terminal methionine, the construct may comprise the sequence M-mM, M-mM-mM or M-mM-mM-mM. The presences of these methionines typically results from the expression of the start codons (i.e. ATGs) at the 5' end of the polynucleotides encoding the second or subsequent monomers within the polynucleotide encoding entire construct. The first monomer in the construct (in the amino to carboxy direction) may also comprise a methionine (e.g. mM-mM, mM-mM-mM or mM-mM-mM-mM).

The two or more monomers may be genetically fused directly together. The monomers are preferably genetically fused using a linker. The linker may be designed to constrain the mobility of the monomers. Preferred linkers are amino acid sequences (i.e. peptide linkers). Any of the peptide linkers discussed above may be used. The construct preferably comprises the sequence shown in SEQ ID NO: 29 or a variant thereof. Each monomer in SEQ ID NO: 29 comprises the sequence shown in SEQ ID NO: 2 or a variant thereof. The second monomer also comprises a methionine at its amino terminus as described above. The two monomers are linked by a peptide linker. A variant of SEQ ID NO: 29 may vary from SEQ ID NO: 29 in any of the ways discussed above with reference to variants of SEQ ID NO: 2. The linker may also be modified or replaced with a peptide linker discussed above.

In another preferred embodiment, the monomers are chemically fused. A subunit is chemically fused to an enzyme if the two parts are chemically attached, for instance via a chemical crosslinker. Any of the chemical crosslinkers discussed above may be used. The linker may be attached to one or more cysteine residues introduced into a mutant monomer of the invention. Alternatively, the linker may be attached to a terminus of one of the monomers in the construct.

If a construct contains different monomers, crosslinkage of monomers to themselves may be prevented by keeping the concentration of linker in a vast excess of the monomers. Alternatively, a "lock and key" arrangement may be used in which two linkers are used. Only one end of each linker may react together to form a longer linker and the other ends of the linker each react with a different monomers. Such linkers are described in International Application No. PCT/GB10/000132 (published as WO 2010/086602).

Polynucleotides

The present invention also provides polynucleotide sequences which encode a mutant monomer of the invention. The mutant monomer may be any of those discussed above. The polynucleotide sequence preferably comprises a sequence at least 50%, 60%, 70%, 80%, 90% or 95% homologous based on nucleotide identity to the sequence of SEQ ID NO: 1 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95% nucleotide identity over a stretch of 300 or more, for example 375, 450, 525 or 600 or more, contiguous nucleotides ("hard homology"). Homology may be calculated as described above. The polynucleotide sequence may comprise a sequence that differs from SEQ ID NO: 1 on the basis of the degeneracy of the genetic code.

The present invention also provides polynucleotide sequences which encode any of the genetically fused constructs of the invention. The polynucleotide preferably comprises two or more sequences as shown in SEQ ID NO: 1 or a variant thereof as described above. The polynucleotide sequence preferably comprises the sequence of SEQ ID NO: 28 or a sequence at least 50%, 60%, 70%, 80%, 90% or 95% homologous based on nucleotide identity to the sequence of SEQ ID NO: 28 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95% nucleotide identity over a stretch of 600 or more, for example 750, 900, 1050 or 1200 or more, contiguous nucleotides ("hard homology"). Homology may be calculated as described above. The polynucleotide sequence may comprise a sequence that differs from SEQ ID NO: 28 on the basis of the degeneracy of the genetic code.

Polynucleotide sequences may be derived and replicated using standard methods in the art. Chromosomal DNA encoding wild-type Msp may be extracted from a pore producing

organism, such as *Mycobacterium smegmatis*. The gene encoding the pore subunit may be amplified using PCR involving specific primers. The amplified sequence may then undergo site-directed mutagenesis. Suitable methods of site-directed mutagenesis are known in the art and include, for example, combine chain reaction. Polynucleotides encoding a construct of the invention can be made using well-known techniques, such as those described in Sambrook, J. and Russell, D. (2001). Molecular Cloning: A Laboratory Manual, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The resulting polynucleotide sequence may then be incorporated into a recombinant replicable vector such as a cloning vector. The vector may be used to replicate the polynucleotide in a compatible host cell. Thus polynucleotide sequences may be made by introducing a polynucleotide 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 are known in the art and described in more detail below.

The polynucleotide sequence may be cloned into suitable expression vector. In an expression vector, the polynucleotide sequence 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 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 polynucleotide sequences may be introduced into the vector.

The expression vector may then be introduced into a suitable host cell. Thus, a mutant monomer or construct of the invention can be produced by inserting a polynucleotide sequence 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 polynucleotide sequence. The recombinantly-expressed monomer or construct may self-assemble into a pore in the host cell membrane. Alternatively, the recombinant pore produced in this manner may be removed from the host cell and inserted into another membrane. When producing pores 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 polynucleotide 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 subunit at a high level. Host cells transformed with a polynucleotide sequence 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. In addition to the conditions listed above any of the methods cited in Proc Natl Acad Sci U S A. 2008 Dec 30;105(52):20647-52 may be used to express the Msp proteins.

Pores

The invention also provides various pores. The pores of the invention are ideal for charcterising, such as sequencing, nucleic acid sequences because they can discriminate between different nucleotides with a high degree of sensitivity. The pores can surprisingly distinguish between the four nucleotides in DNA and RNA. The pores of the invention can even distinguish between methylated and unmethylated nucleotides. The base resolution of pores of the invention is surprisingly high. The pores show almost complete separation of all four DNA nucleotides. The pores further discriminate between deoxycytidine monophosphate (dCMP) and methyl-dCMP based on the dwell time in the pore and the current flowing through the pore.

The pores of the invention can also discriminate between different nucleotides under a range of conditions. In particular, the pores will discriminate between nucleotides under conditions that are favourable to the characterising, such as sequencing, of nucleic acids. The extent to which the pores of the invention can discriminate between different nucleotides can be controlled by altering the applied potential, the salt concentration, the buffer, the temperature and the presence of additives, such as urea, betaine and DTT. This allows the function of the pores to be fine-tuned, particularly when sequencing. This is discussed in more detail below. The pores of the invention may also be used to identify nucleic acid polymers from the interaction with one or more monomers rather than on a nucleotide by nucleotide basis.

A pore of the invention may be isolated, substantially isolated, purified or substantially purified. A pore of the invention is isolated or purified if it is completely free of any other components, such as lipids or other pores. A pore is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a pore is substantially isolated or substantially purified if it present in a form that comprises less than

10%, less than 5%, less than 2% or less than 1% of other components, such as lipids or other pores. Alternatively, a pore of the invention may be present in a lipid bilayer.

A pore of the invention may be present as an individual or single pore. Alternatively, a pore of the invention may be present in a homologous or heterologous population of two or more pores.

Homo-oligomeric pores

The invention also provides a homo-oligomeric pore derived from Msp comprising identical mutant monomers of the invention. The homo-oligomeric pore preferably comprises one of the mutants shown in Tables 1, 2 and 3. The homo-oligomeric pore of the invention is ideal for characterising, such as sequencing, nucleic acids. The homo-oligomeric pore of the invention may have any of the advantages discussed above. The advantages of specific homo-oligomeric pores of the invention are indicated in Tables 1, 2 and 3.

The homo-oligomeric pore may contain any number of mutant monomers. The pore typically comprises 7, 8, 9 or 10 identical mutant monomers. The pore preferably comprises eight identical mutant monomers. One or more, such as 2, 3, 4, 5, 6, 7, 8, 9 or 10, of the mutant monomers is preferably chemically modified as discussed above.

Methods for making pores are discussed in more detail below.

Hetero-oligomeric pores

The invention also provides a hetero-oligomeric pore derived from Msp comprising at least one mutant monomer of the invention, wherein at least one of the eight monomers differs from the others. The hetero-oligomeric pore of the invention is ideal for characterising, such as sequencing, nucleic acids. Hetero-oligomeric pores can be made using methods known in the art (e.g. Protein Sci. 2002 Jul;11(7):1813-24).

The hetero-oligomeric pore contains sufficient monomers to form the pore. The monomers may be of any type. The pore typically comprises 7, 8, 9 or 10 monomers. The pore preferably comprises eight monomers.

The pore may comprise at least one monomer comprising (a) the sequence shown in SEQ ID NO: 2 or (b) a variant thereof which does not have a mutation required by the mutant monomers of the invention. Suitable variants are discussed above. In this embodiment, the remaining monomers are preferably mutant monomers of the invention. Hence, the pore may comprise 9, 8, 7, 6, 5, 4, 3, 2 or 1 mutant monomers of the invention.

In a preferred embodiment, the pore comprises (a) one mutant monomer and (b) seven identical monomers, wherein the mutant monomer in (a) is different from the identical

monomers in (b). The identical monomers in (b) preferably comprise (i) the sequence shown in SEQ ID NO: 2 or (ii) a variant thereof which does not have a mutation present in the mutant monomers of the invention.

Preferred pores include, but are not limited to, any of the following:

(a) Seven monomers comprising the sequence shown in SEQ ID NO: 2 and one mutant monomer comprising the substitution N90R, N90K, N90Y, N90Q, N90W or N90C. These pores have a single steric amino acid (Y or W), a single charged amino acid (K or R) or a single reactive amino acid (C) introduced into the inner constriction.

(b) Seven monomers comprising the sequence shown in SEQ ID NO: 2 and one mutant monomer comprising the substitution N91R, N91K, N91Y, N91Q, N91W or N91C. These pores have a single steric amino acid (Y or W), a single charged amino acid (K or R) or a single reactive amino acid (C) introduced into the inner constriction.

(c) Seven monomers comprising the sequence shown in SEQ ID NO: 2 and one mutant monomer comprising the substitution L88C, S103C or I105C. These pores have a reactive amino acid introduced into the pore.

In another preferred embodiment, all of the monomers (i.e. 10, 9, 8 or 7 of the monomers) are mutant monomers of the invention and at least one of them differs from the others. In a more preferred embodiment, the pore comprises eight mutant monomers of the invention and at least one of them differs from the others.

In all the embodiments discussed above, one or more, such as 2, 3, 4, 5, 6, 7, 8, 9 or 10, of the mutant monomers is preferably chemically modified as discussed above. Preferred pores (a) to (c) above are preferably chemically modified by attachment of a molecule to one or more of the introduced cysteines.

Methods for making pores are discussed in more detail below.

Construct-containing pores

The invention also provides a pore comprising at least one construct of the invention. A construct of the invention comprises two or more covalently attached monomers derived from Msp. In other words, a construct must contain more than one monomer. The pore contains sufficient constructs and, if necessary, monomers to form the pore. For instance, an octameric pore may comprise (a) two constructs each comprising four monomers or (b) one construct comprising two monomers and six monomers that do not form part of a construct. At least two of the monomers in the pore are in the form of a construct of the invention. The monomers may

be of any type. The pore typically comprises 7, 8, 9 or 10 monomers in total (at least two of which must be in a construct). The pore preferably comprises eight monomers (at least two of which must be in a construct).

A pore typically contains (a) one construct comprising two monomers and (b) 5, 6, 7 or 8 monomers. The construct may be any of those discussed above. The monomers may be any of those discussed above, including mutant monomers of the invention.

Another typical pore comprises more than one construct of the invention, such as two, three or four constructs of the invention. Such pores further comprise sufficient monomers to form the pore. The monomer may be any of those discussed above. A further pore of the invention comprises only constructs comprising 2 monomers, for example a pore may comprise 4, 5, 6, 7 or 8 constructs comprising 2 monomers. A specific pore according to the inventions comprises four constructs each comprising two monomers. The constructs may oligomerise into a pore with a structure such that only one monomer of a construct contributes to the barrel or vestibule of the pore. Typically the other monomers of the construct will be on the outside of the barrel or vestibule of the pore. For example, pores of the invention may comprise 5, 6, 7 or 8 constructs comprising 2 monomers where the barrel or vestibule comprises 8 monomers.

Mutations can be introduced into the construct as described above. The mutations may be alternating, i.e. the mutations are different for each monomer within a two monomer construct and the constructs are assembled as a homo-oligomer resulting in alternating modifications. In other words, monomers comprising MutA and MutB are fused and assembled to form an A-B:A-B:A-B:A-B pore. Alternatively, the mutations may be neighbouring, i.e. identical mutations are introduced into two monomers in a construct and this is then oligomerised with different mutant monomers. In other words, monomers comprising MutA are fused follow by oligomerisation with MutB-containing monomers to form A-A:B:B:B:B:B:B:B.

One or more of the monomers of the invention in a construct-containing pore may be chemically-modified as discussed above.

Methods of identifying an individual nucleotide

The present invention also provides methods of characterising an individual nucleotide. The methods comprise contacting the nucleotide with a 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 characterising the nucleotide. The invention therefore involves nanopore sensing of an individual nucleotide. The invention also provides methods of identifying an individual nucleotide comrpsining measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide. Any of the pores of the

invention can be used. The pore of the invention is preferably chemically modified with a molecular adaptor as discussed above.

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

The methods may be carried out using any suitable membrane/pore system in which a pore of the invention is inserted into a membrane. The methods are typically carried out using (i) an artificial membrane comprising a pore of the invention, (ii) an isolated, naturally occurring membrane comprising a pore of the invention, or (iii) a cell expressing a pore that has been modified in accordance with the invention. The methods are 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 pore of the invention.

The membrane forms a barrier to the flow of ions, nucleotides and nucleic acids. Any membrane may be used in accordance with the invention. Suitable membranes are well-known in the art. The membrane is preferably an amphiphilic layer. An amphiphilic layer is a layer formed from amphiphilic molecules, such as phospholipids, which have both hydrophilic and lipophilic properties. The amphiphiles may be synthetic or naturally occurring. The amphiphilic layer may be a monolayer or a bilayer. Non-naturally occurring amphiphiles and amphiphiles which form a monolayer are known in the art and include, for example, block copolymers (Gonzalez-Perez *et al.*, Langmuir, **2009**, 25, 10447-10450).

The membrane may be 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). Lipid bilayers can also be formed using the method described in International Application No. PCT/GB08/000563.

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, mycolic acid and mixtures thereof. Any of the lipids described in International Application No. PCT/GB08/000563 may be used.

In another preferred embodiment, the membrane is a solid state layer. A solid-state layer is not of biological origin. In other words, a solid state layer is not derived from or isolated from a biological environment such as an organism or cell, or a synthetically manufactured version of a biologically available structure. Solid state layers can be formed from both organic and inorganic materials including, but not limited to, microelectronic materials, insulating materials such as Si3N4, A1203, and SiO, organic and inorganic polymers such as polyamide, plastics such as Teflon® or elastomers such as two-component addition-cure silicone rubber, and glasses. The solid state layer may be formed from monatomic layers, such as graphene, or layers that are only a few atoms thick. Suitable graphene layers are disclosed in International Application No. PCT/US2008/010637 (published as WO 2009/035647). An amphiphilic layer may be formed across a solid state pore. This may be described in the art as hybrid pore formation (Hall *et al.*, Nat Nanotechnol., **2010**, 5, 874-877).

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 "pick and place" method described in M.A. Holden, H. Bayley. J. Am. Chem. Soc. 2005, 127, 6502-6503 and International Application No. PCT/GB2006/001057 (published as WO 2006/100484).

The methods of the invention are typically carried out *in vitro*.

Individual nucleotide

An individual nucleotide is a single nucleotide. An individual nucleotide is one which is not bound to another nucleotide or nucleic acid 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 nucleic acid 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 methods 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. Nucleic acid sequences can be digested using any method known in the art. Suitable methods include, but are not limited to, those using enzymes or catalysts. Catalytic digestion of nucleic acids is disclosed in Deck *et al.*, Inorg. Chem., 2002; 41: 669-677.

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 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 or oxidised. 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.

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 may be contacted with the 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 nucleotide may pass through the pore. In such cases, the nucleotide interacts with the pore and/or adaptor as it passes across the membrane through the barrel or channel of the pore. Alternatively, the nucleotide may be contacted with the side of the membrane that allows the nucleotide 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 position of the adaptor is fixed. As a result, the nucleotide is preferably contacted with the end of the pore which allows the adaptor to interact with the nucleotide.

The nucleotide may interact with the pore in any manner and at any site. As discussed above, the nucleotide preferably reversibly binds to the pore via or in conjunction with the adaptor. The nucleotide most preferably reversibly binds to the pore via or in conjunction with the adaptor as it passes through the pore across the membrane. The nucleotide 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 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 mean time period and to a particular extent. In other words, the current flowing through the pore is distinctive for a particular nucleotide. 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. The ratio of different nucleotides within a sample can also be calculated. For instance, the ratio of dCMP to methyl-dCMP can be calculated.

Apparatus

The methods may be carried out using any apparatus that is suitable for investigating a membrane/pore system in which a pore of the invention is inserted into a membrane. The method may be carried out using any apparatus that is suitable for nanopore 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 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 methods may be carried out using the apparatus described in International Application No. PCT/GB08/000562.

The methods of the invention involve measuring the current passing through the pore during interaction with the nucleotide. Therefore the apparatus also comprises an electrical circuit capable of applying a potential and measuring an electrical signal across the membrane and pore. The methods may be carried out using a patch clamp or a voltage clamp. The methods preferably involve the use of a voltage clamp.

Sample

The nucleotide 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 nucleotide. 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 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. Alternatively a sample of plant origin is typically obtained from a commercial crop, such as a cereal, legume, fruit or vegetable, for example wheat, barley, oats,

canola, maize, soya, rice, bananas, apples, tomatoes, potatoes, grapes, tobacco, beans, lentils, sugar cane, cocoa, cotton, tea, coffee.

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 methods of the invention involve 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 –400mV to +400mV. The voltage used is preferably in a range having a lower limit selected from -400 mV, -300mV, -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, +200 mV, +300 mV and +400 mV. The voltage used is more preferably in the range 100mV to 240mV and most preferably in the range of 160mV to 240mV. It is possible to increase discrimination between different nucleotides by a pore of the invention by using an increased applied potential.

The methods are typically 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 from 150mM to 1M. High salt concentrations provide a high signal to noise ratio and allow for currents indicative of the presence of a nucleotide to be identified against the background of normal current fluctations. Lower salt concentrations may be used if nucleotide detection is carried out in the presence of an enzyme, such as when sequencing nucleic acids. This is discussed in more detail below.

The methods are 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

methods are typically carried out at a pH of from 4.0 to 12.0, from 4.5 to 10.0, 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 7.5.

The methods are 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 methods may be carried out at room temperature. The methods are preferably carried out at a temperature that supports enzyme function, such as about 37°C.

Methods of characterising nucleic acids

The present invention also provides methods of characterising a target nucleic acid sequence. One or more characteristics of the target nucleic acid sequence may be deteremined. The method may involve measuring two, three, four or five or more characteristics of the target nucleic acid sequence. The one or more characteristics are preferably selected from (i) the length of the target nucleic acid sequence, (ii) the identity of the target nucleic acid sequence, (iii) the sequence of the target nucleic acid sequence, (iv) the secondary structure of the target nucleic acid sequence and (v) whether or not the target nucleic acid sequence is modified. Any combination of (i) to (v) may be determined in accordance with the invention.

For (i), the length of the nucleic acid sequence may be measured using the number of interactions between the target nucleic acid sequence and the pore.

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

For (iii), the sequence of the nucleic acid sequence can be determined as described previously. Suitable sequencing methods, particularly those using electrical measurements, are described in Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7, Lieberman KR et al, J Am Chem Soc. 2010;132(50):17961-72, and International Application WO 2000/28312.

For (iv), the secondary structure may be measured in a variety of ways. For instance, the secondary structure may be measured using a change in dwell time or a change in current flowing through the pore.

The invention also provides a method of estimating the sequenceof a target nucleic acid sequence. The invention further provides a method of sequencing a target nucleic acid sequence.

A nucleic acid is a macromolecule comprising two or more nucleotides. The nucleotides may be any of those discussed above.

In one embodiment, the method comprises (a) contacting the target sequence with a pore of the invention and a nucleic acid binding protein so that protein controls the movement of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore and (b) measuring the current passing through the pore during each interaction and thereby charcterising, such as estimating the sequence of or sequencing, the target sequence. Hence, the method involves nanopore sensing of a proportion of the nucleotides in a target nucleic acid sequence as the nucleotides pass through the barrel or channel in order to characterising, such as sequencing, the target sequence.

In another embodiment, the method comprises (a) contacting the target sequence with a pore of the invention and an exonuclease such that the exonuclease digests an individual nucleotide from one end of the target sequence; (b) contacting the nucleotide with the pore so that the nucleotide interacts with the adaptor; (c) measuring the current passing through the pore during the interaction and thereby characterising the nucleotide; and (d) repeating steps (a) to (c) at the same end of the target sequence and thereby characterining of the target sequence. Hence, the method involves nanopore sensing of a proportion of the nucleotides in a target nucleic acid sequence in a successive manner in order to characterise the target sequence. In a preferred embodiment, the method concerns sequencing the target nucleic acid sequence and step (a) comprises determining the identity of the nucleotide. Individual nucleotides are described above.

The pores of the invention are particularly suited to these methods because they display improved nucleotide discrimination. In particular, they display an increased current range, which makes it easier to discriminate between different nucleotides, and a reduced variance of states, which increases the signal-to-noise ratio. In addition, in relation to the former embodiment, the number of nucleotides contributing to the current as the nucleic acid moves through the pore is decreased. This makes it easier to identify a direct relationship between the observed current as the nucleic acid moves through the pore and the nucleic acid sequence. The pores of the invention are preferably chemically modified with (1) a molecular adaptor and/or (2) the nucleic acid binding protein or exonuclease as discussed above.

The whole or only part of the target nucleic acid sequence may be characterised, such as sequenced, using this method. The nucleic acid sequence can be any length. For example, the nucleic acid sequence 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 sequence can be 1000 or more nucleotides or 5000 or more nucleotides in length. The nucleic acid sequence can be naturally occurring or artificial. For instance, the method may be used to verify the sequence of a manufactured oligonucleotide. The methods are typically carried out *in vitro*.

The methods may be carried out using any suitable membrane/pore system in which a pore is inserted into a membrane. The methods are typically carried out using any of the systems, apparatus or conditions disclosed above.

As mentioned above, good nucleotide discrimination can be achieved at low salt concentrations if the temperature is increased. In addition to increasing the solution temperature, there are a number of other strategies that can be employed to increase the conductance of the solution, while maintaining conditions that are suitable for enzyme activity. One such strategy is to use the lipid bilayer to divide two different concentrations of salt solution, a low salt concentration of salt on the enzyme side and a higher concentration on the opposite side. One example of this approach is to use 200 mM of KCl on the *cis* side of the membrane and 500 mM KCl in the *trans* chamber. At these conditions, the conductance through the pore is expected to be roughly equivalent to 400 mM KCl under normal conditions, and the enzyme only experiences 200 mM if placed on the *cis* side. Another possible benefit of using asymmetric salt conditions is the osmotic gradient induced across the pore. This net flow of water could be used to pull nucleotides into the pore for detection. A similar effect can be achieved using a neutral osmolyte, such as sucrose, glycerol or PEG. Another possibility is to use a solution with relatively low levels of KCl and rely on an additional charge carrying species that is less disruptive to enzyme activity.

The target sequence being analysed can be combined with known protecting chemistries to protect the sequence from being acted upon by the binding protein or exonuclease while in the bulk solution. The pore can then be used to remove the protecting chemistry. This can be achieved either by using protecting groups that are unhybridised by the pore, binding protein or enzyme under an applied potential (WO 2008/124107) or by using protecting chemistries that are removed by the binding protein or enzyme when held in close proximity to the pore (J Am Chem Soc. 2010 Dec 22;132(50):17961-72).

Strand sequencing

Strand sequencing involves the controlled and stepwise translocation of nucleic acid polymers through a pore. Pores of the invention can be used in strand sequencing. One method of the invention uses a nucleic acid binding protein to control the movement of the target sequence through the pore. Examples of such proteins include, but are not limited to, nucleic

acid handling enzymes, such as nucleases, polymerases, topoisomerases, ligases and helicases, and non-catalytic binding proteins such as those classified by SCOP (Structural Classification of Proteins) under the Nucleic acid-binding protein superfamily (50249). The binding protein may be single strand binding protein (SSB).

A nucleic acid is a macromolecule comprising two or more nucleotides. The nucleic acid bound by the protein may comprise any combination of any nucleotides. The nucleotides may be any of those discussed above. The nucleic acid can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The nucleic acid may be any synthetic nucleic acid known in the art, such as peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or other synthetic polymers with nucleotide side chains. The nucleic acid bound by the protein may be single stranded, such as cDNA, RNA, GNA, TNA or LNA, or double stranded, such as DNA. Proteins that bind single stranded nucleic acids may be used to sequence double stranded DNA as long as the double stranded DNA is dissociated into a single strand before it is bound by the protein.

The nucleic acid binding protein is preferably a nucleic acid handling enzyme. A nucleic acid handling enzyme is a polypeptide that is capable of interacting with and modifying at least one property of a nucleic acid. The enzyme may modify the nucleic acid by cleaving it to form individual nucleotides or shorter chains of nucleotides, such as di- or trinucleotides. The enzyme may modify the nucleic acid by orienting it or moving it to a specific position. The nucleic acid handling enzyme does not need to display enzymatic activity as long as it is capable of binding the target sequence and controlling its movement through the pore. For instance, the enzyme may be modified to remove its enzymatic activity or may be used under conditions which prevent it from acting as an enzyme. Such conditions are discussed in more detail below.

The nucleic acid handling enzyme is preferably derived from a nucleolytic enzyme. The nucleic acid handling enzyme used in the construct of the enzyme is more preferably derived from a member of any of the Enzyme Classification (EC) groups 3.1.11, 3.1.13, 3.1.14, 3.1.15, 3.1.16, 3.1.21, 3.1.22, 3.1.25, 3.1.26, 3.1.27, 3.1.30 and 3.1.31. The enzyme may be any of those disclosed in International Application No. PCT/GB10/000133 (published as WO 2010/086603).

Preferred enzymes are polymerases, exonucleases, helicases and topoisomerases, such as gyrases. Suitable enzymes include, but are not limited to, exonuclease I from *E. coli* (SEQ ID NO: 6), exonuclease III enzyme from *E. coli* (SEQ ID NO: 8), RecJ from *T. thermophilus* (SEQ ID NO: 10) and bacteriophage lambda exonuclease (SEQ ID NO: 12) and variants thereof. Three subunits comprising the sequence shown in SEQ ID NO: 10 or a variant thereof interact to form a trimer exonuclease. The enzyme is preferably based on Phi29 DNA polymerase (SEQ ID NO: 4).

A variant of SEQ ID NOs: 4, 6, 8, 10 or 12 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 4, 6, 8, 10 or 12 and which retains nucleic acid binding ability. The variant may include modifications that facilitate binding of the nucleic acid and/or facilitate its activity at high salt concentrations and/or room temperature.

Over the entire length of the amino acid sequence of SEQ ID NO: 4, 6, 8, 10 or 12, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant 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: 4, 6, 8, 10 or 12 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"). Homology is determined as described above. The variant may differ from the wild-type sequence in any of the ways discussed above with reference to SEQ ID NO: 2. The enzyme may be covalently attached to the pore as discussed above.

The enzyme is not required to be in as close a proximity to the pore lumen as for individual nucleotide sequencing as there is no potential for disorder in the series in which nucleotides reach the sensing moiety of the pore.

The two strategies for single strand DNA sequencing are the translocation of the DNA through the nanopore, both *cis* to *trans* and *trans* to *cis*, either with or against an applied potential. The most advantageous mechanism for strand sequencing is the controlled translocation of single strand DNA through the nanopore under an applied potential. Exonucleases that act progressively or processively on double stranded DNA can be used on the *cis* side of the pore to feed the remaining single strand through under an applied potential or the *trans* side under a reverse potential. Likewise, a helicase that unwinds the double stranded DNA can also be used in a similar manner. There are also possibilities for sequencing applications that require strand translocation against an applied potential, but the DNA must be first "caught" by the enzyme under a reverse or no potential. With the potential then switched back following binding the strand will pass *cis* to *trans* through the pore and be held in an extended conformation by the current flow. The single strand DNA exonucleases or single strand DNA dependent polymerases can act as molecular motors to pull the recently translocated single strand back through the pore in a controlled stepwise manner, *trans* to *cis*, against the applied potential.

Exonuclease-based methods

In one embodiment, the method of characterising a target nucleic acid sequence involves contacting the target sequence with an exonuclease enzyme. Any of the exonuclease enzymes discussed above may be used in the method. The exonuclease releases individual nucleotides from one end of the target sequence. The enzyme may be covalently attached to the pore as discussed above.

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 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 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 exonuclease to a particular end of the nucleic acid sequence.

The method involves contacting the nucleic acid sequence with the exonuclease so that the nucleotides are digested from the end of the nucleic acid at a rate that allows characterisation or identification of a proportion of nucleotides 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 rate at which the exonuclease functions is typically slower than the optimal rate of a wild-type exonuclease. A suitable rate of activity of the exonuclease in the method of the invention 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 exonuclease activity can be achieved in various ways. For example, variant exonucleases with a reduced optimal rate of activity may be used in accordance with the invention.

Msp and Phi29 DNA polymerase

In a preferred embodiment, characterisation, such as strand sequencing, is carried out using a pore derived from Msp and a Phi29 DNA polymerase. The method comprises (a) contacting the target sequence with a pore derived from Msp and a Phi29 DNA polymerase such that the polymerase controls the movement of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore and (b) measuring the

current passing through the pore during each interaction and thereby characterising, such as determining the sequence, of the target sequence, wherein steps (a) and (b) are carried out with a voltage applied across the pore. When the target sequence is contacted with a Phi29 DNA polymerase and a pore derived from Msp, the target sequence firstly forms a complex with the Phi29 DNA polymerase. When the voltage is applied across the pore, the target sequence/Phi29 DNA polymerase complex forms a complex with the pore and controls the movement of the target sequence through the pore.

This embodiment has three unexpected advantages. First, the target sequence moves through the pore at a rate that is commercially viable yet allows effective sequencing. The target sequence moves through the Msp pore more quickly than it does through a hemolysin pore. Second, an increased current range is observed as the nucleic acid moves through the pore allowing the sequence to be determined more easily. Third, a decreased current variance is observed when the specific pore and polymerase are used together thereby increasing the signal-to-noise ratio.

Any nucleic acid sequence described above may be characterised or sequenced. At least a portion of the nucleic acid sequence is preferably double stranded.

The pore may be any of the pores discussed above. The pore is preferably a pore of the invention. The pore may comprise eight monomers comprising the sequence shown in SEQ ID NO: 2, 16, 17 or 18 or a variant thereof. The pore does not have to include any of the mutations of the invention.

Wild-type Phi29 DNA polymerase has polymerase and exonuclease activity. It may also unzip double stranded nucleic acids under the correct conditions. Hence, the enzyme may work in three modes. This is discussed in more detail below.

The Phi29 DNA polymerase may comprise the sequence shown in SEQ ID NO: 4 or a variant thereof. A variant of SEQ ID NOs: 4 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 4 and which retains nucleic acid binding activity. The variant must work in at least one of the three modes discussed below. Preferably, the variant works in all three modes. The variant may include modifications that facilitate handling of the nucleic acid and/or facilitate its activity at high salt concentrations and/or room temperature.

Over the entire length of the amino acid sequence of SEQ ID NO: 4, a variant will preferably be at least 40% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 50%, 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: 4 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"). Homology is determined as described above. The variant may differ from the wild-type sequence in any of the ways discussed above with reference to SEQ ID NO: 2. The enzyme may be covalently attached to the pore as discussed above.

Any of the systems, apparatus or conditions discussed above may be used in accordance with this preferred embodiment. The salt concentration is typically from 0.15M to 0.6M. The salt is preferably KCl.

The method may be carried out in one of three preferred ways based on the three modes of the Phi29 DNA polymerase. Each way includes a method of proof reading the sequence. First, the method is preferably carried out using the Phi29 DNA polymerase as a polymerase. In this embodiment, steps (a) and (b) are carried out in the presence of free nucleotides and an enzyme cofactor such that the polymerase moves the target sequence through the pore against the field resulting from the applied voltage. The target sequence moves in the 5' to 3' direction. The free nucleotides may be one or more of any of the individual nucleotides discussed above. The enzyme cofactor is a factor that allows the Phi29 DNA polymerase to function either as a polymerase or an exonuclease. The enzyme cofactor is preferably a divalent metal cation. The divalent metal cation is preferably Mg^{2+}, Mn^{2+}, Ca^{2+} or Co^{2+}. The enzyme cofactor is most preferably Mg^{2+}. The method preferably further comprises (c) removing the free nucleotides such that the polymerase moves the target sequence through the pore with the field resulting from the applied voltage (i.e. in the 3' and 5' direction) and a proportion of the nucleotides in the target sequence interacts with the pore and (d) measuring the current passing through the pore during each interaction and thereby proof reading the sequence of the target sequence obtained in step (b), wherein steps (c) and (d) are also carried out with a voltage applied across the pore.

Second, the method is preferably carried out using the Phi29 DNA polymerase as an exonuclease. In this embodiment, wherein steps (a) and (b) are carried out in the absence of free nucleotides and the presence of an enzyme cofactor such that the polymerase moves the target sequence through the pore with the field resulting from the applied voltage. The target sequence moves in the 3' to 5' direction. The method preferably further comprises (c) adding free nucleotides such that the polymerase moves the target sequence through the pore against the field resulting from the applied voltage (i.e. in the 5' to 3' direction) and a proportion of the nucleotides in the target sequence interacts with the pore and (d) measuring the current passing through the pore during each interaction and thereby proof reading the sequence of the target sequence obtained in step (b), wherein steps (c) and (d) are also carried out with a voltage applied across the pore.

Third, the method is preferably carried out using the Phi29 DNA polymerase in unzipping mode. In this embodiment, steps (a) and (b) are carried out in the absence of free nucleotides and the absence of an enzyme cofactor such that the polymerase controls the movement of the target sequence through the pore with the field resulting from the applied voltage (as it is unzipped). In this embodiment, the polymerase acts like a brake preventing the target sequence from moving through the pore too quickly under the influence of the applied voltage. The method preferably further comprises (c) lowering the voltage applied across the pore such that the target sequence moves through the pore in the opposite direction to that in steps (a) and (b) (i.e. as it re-anneals) and a proportion of the nucleotides in the target sequence interacts with the pore and (d) measuring the current passing through the pore during each interaction and thereby proof reading the sequence of the target sequence obtained in step (b), wherein steps (c) and (d) are also carried out with a voltage applied across the pore.

The invention also provides a method of forming a sensor for sequencing a target nucleic acid sequence, comprising (a) contacting a pore derived from Msp with a Phi29 DNA polymerase in the presence of the target nucleic acid sequence and (b) applying a voltage across the pore to form a complex between the pore and the polymerase and thereby forming a sensor for sequencing the target nucleic acid sequence. The invention further provides a method of increasing the rate of activity of a Phi29 DNA polymerase, comprising contacting the Phi29 DNA polymerase with a pore derived from Msp in the presence of a nucleic acid sequence and applying a voltage across the pore to form a complex between the pore and the polymerase and thereby increasing the rate of activity of a Phi29 DNA polymerase.

Kits

The present invention also provides kits for characterising, such as sequencing, a target nucleic acid sequence. One kit comprises (a) a pore of the invention and (b) a nucleic acid handling enzyme. Another kit comprises (a) a pore derived from Msp and (b) a Phi29 DNA polymerase. Any of the embodiments discussed above with reference to the methods of the invention are equally applicable to the kits of the invention.

The kits of the invention may additionally comprise one or more other reagents or instruments which enable any of the embodiments 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 and/or express polynucleotide 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.

Apparatus

The invention also provides an apparatus for characterising, such as sequencing, target nucleic acid sequences in a sample. The apparatus may comprise (a) a plurality of pores of the invention and (b) a plurality of nucleic acid handling enzymes. Alternatively, the invention may comprise a plurality of pores derived from Msp and a plurality of Phi29 DNA polymerases. The apparatus may be any conventional apparatus for analyte analysis, such as an array or a chip.

The apparatus preferably comprises:

a sensor device that is capable of supporting the plurality of pores and being operable to perform nucleic acid characterising or sequencing using the pores and enzymes;

- at least one reservoir for holding material for performing the characterising or sequencing;
- a fluidics system configured to controllably supply material from the at least one reservoir to the sensor device; and
- a plurality of containers for receiving respective samples, the fluidics system being configured to supply the samples selectively from the containers to the sensor device. The apparatus may be any of those described in International Application No. PCT/GB10/000789 (published as WO 2010/122293), International Application No. PCT/GB10/002206 (not yet published) or International Application No. PCT/US99/25679 (published as WO 00/28312).

The following Examples illustrate the invention:

Example 1

Homo-oligomers are pores where all the monomer units are identical. As the monomer units will self-assemble, these are the simplest constructs to produce. Our strategies for improving the base reader properties can be split into categories:

- Sterics *(increasing or decreasing the size of amino acid residues)*
- Charge *(introducing +ve charge to interact with DNA)*
- Hydrogen bonding *(residues that can hydrogen bond to the base pairs)*
- Pi Stacking *(amino acids that interact through delocalised electron pi systems)*

Increase Sterics / Pi Stacking (all NNN-RRK background):

Sterics – substitution for residues with bulk *(e.g. Phenylalanine, Tryptophan, Tyrosine, Histidine)*

Pi Stacking – substitution for aromatic residues *(e.g. Phenylalanine, Tryptophan, Tyrosine, Histidine)*

In all the following tables (6-11), the mutations made to SEQ ID NO: 2 are shown. B1 = SEQ ID NO: 2.

Table 6

MS-(B1-D91Y)$_8$	Bulky Tyrosine at the constriction.
MS-(B1-D90G/D91Y)$_8$	Bulky Tyrosine at the constriction, reduced barrel.
MS-(B1-D90Y/D91G)$_8$	Bulky Tyrosine at the constriction, reduced barrel.
MS-(B1-I05Y)$_8$	Mutation just above barrel to increase size.

Decreased Sterics – substitution for residues with-smaller size *(e.g. Serine, Threonine, Glycine, Alanine, Valine)*

Table 7

MS-(B1-D90G/D91G)$_8$	Reduction of sterics in the barrel.
MS-(B1-I05A)$_8$	Mutation just above barrel to reduce size.
MS-(B1-I05G)$_8$	Mutation just above barrel to reduce size.

Charge –substitution for residues with positive charge *(e.g. Arginine, Lysine, Histidine)*

Table 8

MS-(B1-D90R)$_8$	Charged Arginine at the constriction.
MS-(B1-D91R)$_8$	Charged Arginine at the constriction.
MS-(B1-D90R/D91R)$_8$	Double Arginine at the constriction.
MS-(B1-D90K)$_8$	Charged Lysine at the constriction.
MS-(B1-D91K)$_8$	Charged Lysine at the constriction.
MS-(B1-D90K/D91K)$_8$	Double Lysine at the constriction.

Hydrogen Bonding – substitution for residues with-bonding capacity *(e.g. Asparagine, Glutamine, Tyrosine, Histidine)*

Table 9

MS-(B1-D90Q)$_8$	Glutamine at the constriction.
MS-(B1-D91Q)$_8$	Glutamine at the constriction.

MS-(B1-D90Q/D91G)$_8$	Glutamine at the constriction, size reduction.
MS-(B1-D90G/D91Q)$_8$	Glutamine at the constriction, size reduction.
MS-(B1-D90Q/D91Q)$_8$	Double Glutamine at the constriction.

Table 10

MS-(B1- D118N)$_8$	Removal of the charge in the middle of the lumen.
MS-(B1-D118A)$_8$	Removal of the charge in the middle of the lumen.

Homo-oligomers can also be modified to contain reactive group, which can then be chemically modified.

Table 11

MS-(B1-D91C)$_8$	Addition of Cysteine at the barrel.
MS-(B1-D90C)	Addition of Cysteine at the barrel.

Example 2

Different monomer units can be combined to create novel oligomer pores. When the oligomer contains more than one different subunit *(e.g. MS-(MutA)$_6$(MutB)$_1$(MutC)$_1$)*, the pore is a hetero-oligomer. Hetero-oligomers typically only have one unit modified *(e.g. MS-(MutA)$_7$(MutB)$_1$)*. Other ratios of hetero-oligomers could also be formed *(e.g. MS-(MutA)$_6$(MutB)$_2$)*. Subunits may also include SEQ ID NO: 2.

The advantage of hetero-oligomers is that a single chemical change can be made to the pore (rather than introducing a change to every monomer-unit). This is a less drastic change to the structure than a homo-oligomer and may allow residues to be introduced into the pore at a position which did not work for a homo-oligomer. A single residue interacting with the DNA may be beneficial compared to multiple units *(e.g. a single Arg on a hetero-octamer, compared to eight Arg on an octamer)*. Mutants can also be combine to produce different effects at the same residue, an example of this would be to reduce the size of seven units, while increasing the size of one *(e.g. MS-(D90G)$_8$(D90Y)$_1$)*.

Mutant design rules will be similar to those presented above for homo-oligomers.

Introduction of a Single Steric Residue

Table 12

MS-(B1)$_7$ (B1-D90Y)$_1$	Addition of single Tyrosine in the barrel.
MS-(B1)$_7$ (B1-D91Y)$_1$	Addition of single Tyrosine in the barrel.
MS-(B1)$_7$	Addition of single Tryptophan in the barrel.

$(B1\text{-}D90W)_1$	
$MS\text{-}(B1)_7$ $(B1\text{-}D91W)_1$	Addition of single Tryptophan in the barrel.

Introduction of a Single Charged Residue

Table 13

$MS\text{-}(B1)_7$ $(B1\text{-}D90K)_1$	Addition of single Lysine in the barrel.
$MS\text{-}(B1)_7$ $(B1\text{-}D91K)_1$	Addition of single Lysine in the barrel.
$MS\text{-}(B1)_7$ $(B1\text{-}D90R)_1$	Addition of single Arginine in the barrel.
$MS\text{-}(B1)_7$ $(B1\text{-}D91R)_1$	Addition of single Arginine in the barrel.

Introduction of a Single Reactive Residue

Table 14

$MS\text{-}(B1)_7$ $(B1\text{-}D90C)_1$	Addition of single Cysteine in the barrel.
$MS\text{-}(B1)_7$ $(B1\text{-}D91C)_1$	Addition of single Cysteine in the barrel.

Example 3

Introduction of a Single Reactive Residue for chemical modification.

Table 15

$MS\text{-}(B1)_7$ $(B1\text{-}D90C)_1$	Addition of single Cysteine in the barrel.
$MS\text{-}(B1)_7$ $(B1\text{-}D118C)_1$	Addition of single Cysteine in the lumen.
$MS\text{-}(B1)_7$ $(B1\text{-}G54C)_1$	Addition of single Cysteine for nucleic acid binding protein attachment.

Example 4

The following Tables summarize the mutant pores of the invention. The first concerns homo-oligomers and the second concerns hetero-oligomers.

Table 16

Design_type_name	Mutant_Short	Mutant_full
Mutant (Homo)	MS-(B1)8	MS-(D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90Y)8	MS-(D90Y/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90R)8	MS-(D90R/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90K)8	MS-(D90K/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D118N)8	MS-(D90N/D91N/D93N/D118N/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90Q)8	MS-(D90Q/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D91Q)8	MS-(D90N/D91Q/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D91Y)8	MS-(D90N/D91Y/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90G/D91Y)8	MS-(D90G/D91Y/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D91R)8	MS-(D90N/D91R/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90R/D91R)8	MS-(D90R/D91R/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D91K)8	MS-(D90N/D91K/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90K/D91K)8	MS-(D90K/D91K/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105A)8	MS-(D90N/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105G)8	MS-(D90N/D91N/D93N/I105G/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105Y)8	MS-(D90N/D91N/D93N/I105Y/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105N)8	MS-(D90N/D91N/D93N/I105N/D118R/D134R/E139K)8
Chemical (Homo)	MS-(B1-D91C)8	MS-(D90N/D91C/D93N/D118R/D134R/E139K)8
Chemical (Homo)	MS-(B1-D90C)8	MS-(D90C/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90Y/D91G)8	MS-(D90Y/D91G/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90G/D91G)8	MS-(D90G/D91G/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90G/D93G)8	MS-(D90G/D91N/D93G/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90G)8	MS-(D90G/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D91G)8	MS-(D90N/D91G/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D93G)8	MS-(D90N/D91N/D93G/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90G/D91A)8	MS-(D90G/D91A/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90S)8	MS-(D90S/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D91S)8	MS-(D90N/D91S/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90S/D91S)8	MS-(D90S/D91S/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105K)8	MS-(D90N/D91N/D93N/I105K/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105R)8	MS-(D90N/D91N/D93N/I105R/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105V)8	MS-(D90N/D91N/D93N/I105V/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105L)8	MS-(D90N/D91N/D93N/I105L/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105P)8	MS-(D90N/D91N/D93N/I105P/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105W)8	MS-(D90N/D91N/D93N/I105W/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105S)8	MS-(D90N/D91N/D93N/I105S/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105T)8	MS-(D90N/D91N/D93N/I105T/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-I105Q)8	MS-(D90N/D91N/D93N/I105Q/D118R/D134R/E139K)8

Mutant (Homo)	MS-(B1-L88R)8	MS-(L88R/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88A)8	MS-(L88A/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88Y)8	MS-(L88Y/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88G)8	MS-(L88G/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88N)8	MS-(L88N/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88Q)8	MS-(L88Q/D90N/D91N/D93N/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90Y/I105A)8	MS-(D90Y/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90G/I105A)8	MS-(D90G/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90Q/I105A)8	MS-(D90Q/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90R/I105A)8	MS-(D90R/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-D90S/I105A)8	MS-(D90S/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88A/I105A)8	MS-(L88A/D90N/D91N/D93N/I105A/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88S/I105S)8	MS-(L88S/D90N/D91N/D93N/I105S/D118R/D134R/E139K)8
Mutant (Homo)	MS-(B1-L88N/I105N)8	MS-(L88N/D90N/D91N/D93N/I105N/D118R/D134R/E139K)8

Table 17

Design_type	**Mutant_Short**	**Mutant_full**
Mutant (Hetero)	MS-(B1)7(B1-D90K)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90K/D91N/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D90R)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90R/D91N/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D90Y)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90Y/D91N/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D90Q)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90Q/D91N/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D91Q)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91Q/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D91K)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91K/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D91R)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91R/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D91Y)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91Y/D93N/D118R/D134R/E139K)1
Mutant (Hetero)	MS-(B1)7(B1-D90W)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90W/D91N/D93N/D118R/D134R/E139K)1

Mutant (Hetero)	MS-(B1)7(B1-D91W)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91W/D93N/D118R/D134R/E139K)1
Chemical (Hetero)	MS-(B1)7(B1-D90C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90C/D91N/D93N/D118R/D134R/E139K)1
Chemical (Hetero)	MS-(B1)7(B1-D91C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1
Chemical (Hetero)	MS-(B1)7(B1-L88C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1
Chemical (Hetero)	MS-(B1)7(B1-S103C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1
Chemical (Hetero)	MS-(B1)7(B1-I105C)1	MS-(D90N/D91N/D93N/D118R/D134R/E139K)7(D90N/D91C/D93N/D118R/D134R/E139K)1

Example 5 – MspA compared with HL

We have combined Phi29 DNA polymerase (DNAP) as a molecular motor with a mutant MspA nanopore to allow controlled movement of a DNA strand through the pore. A voltage was applied across the pore and a current was generated from the movement of ions in a salt solution on either side of the nanopore. As the DNA moved through the pore, the ionic flow through the pore changed with respect to the DNA. This information has been shown to be sequence dependent.

We compared a mutant form of hemolysin with MspA, in particular MS-$(B1)_8$. The current range is higher for MspA compared with hemolysin (HL). In addition, the current range is also larger for MspA when a strand of DNA is threaded into the pore.

We have shown that there are a number of surprising features with MspA that were not anticipated by bringing the MspA and the Phi29 DNAP together. The main differences are:

1. Faster strand movement (Unzipping mode) compared to HL.
2. Increase current range when moving a strand through the pore.
3. Decreased variance of the current levels compared to HL mutants.

Faster Strand Movement

A 134mer ssDNA template (SEQ ID NO: 13) was hybridised to a 84mer ssDNA (SEQ ID NO: 14) to form a 84mer dsDNA template with a 50mer ssDNA 5' overhang. This strand moved through the MS-$(B1)_8$ MspA mutant and the hemolysin mutant using Phi29 DNAP in Unzipping mode. Two runs were acquired; one at 400 mM KCl and the other 600 mM KCl, all at room temperature with 10 mM Hepes, pH 8.0, 1 mM EDTA, 1mM DTT. The applied

potential was optimised for each mutant construct; HL was ran at 220 mV and the MspA at 180 mV.

Current levels were extracted as events from the DNA in the enzyme bound state these events were indexed and the current level, duration and variance of the event recorded.

For all the unzipping runs, the speed of unzipping was not consistent through the strand. This can be shown by calculating the average of the event duration, split by quarters of event index (Fig. 1). The first quarter provided events that had a much longer duration that the following quarters, this was true for both HL and MspA. For the first quarter, the average event length was shortest for MspA at 400 mM KCl and shortest for HL at 600 mM. However, in Q2, Q3 and Q4, the MspA produced shorter events for both salt conditions. Assuming the signal to noise is sufficient, short events are desirable as they indicate a rapid movement of the DNA strand through the pore, thus increasing the experimental throughput.

Increased Current Range and Reduced Variance

In the nanopore experiments described here, the current levels are mainly dependent on the salt concentration, the applied voltage, and temperature. The HL and the MS-$(B1)_8$ MspA mutants were compared in Unzipping mode using Phi29 DNA polymerase with set physical conditions of: 600 mM KCl, 10 mM Hepes, 1 mM EDTA, 1 mM DTT, pH 8.0, +220 mV. The DNA used in this experiment was a 100mer hairpin with a 34mer single stranded 5' overhang (SEQ ID NO: 15). The runs were conducted at room temperature.

Current levels were extracted as events from the DNA in the enzyme bound state these events were indexed and the current level, duration and variance of the event recorded (Figs. 2 and 3).

It is clear from these experiments, that the MspA mutant gives a significantly larger current range of approximately 50 pA compared with the HL mutant where the range is approximately 20 pA (Figs. 2 and 3). A large current range is advantageous as it will provide a greater signal to noise and make it easier to distinguish distinct current states. This is of particular benefit for sequencing applications, when N bases may contribute to the current signal, leading to 4^N possible current states.

The variance of the states is also reduced for the MspA mutant compared to the HL. This is shown by the standard deviation of the events in the traces above (Figs. 2 and 3). For the strands above, the average of the standard deviation across all events for the MspA strand was 3.6 compared to 4.5 for HL. Low variance of states is desirable to allow accurate estimations of the event current level.

Example 6 - Open pore current comparison of the MS-(B1)8 baseline to the MS-(B1-I105)8 mutants

The current levels of MspA pores can be controlled by mutating the I105 position in the protein. We demonstrate that the open pore current can be increased by over 80% as a result of making a single mutation to the MspA monomer.

Single channels were inserted into a lipid membrane under the following conditions: 400 mM KCl, 10 mM Hepes, pH 8.0, room temperature. The open pore current level was recorded over a range of applied potentials from -200 mV to 200 mV to produce an IV curve. The experiment was repeated for a number of pores to assess the distribution of the sample. An example of the data from an IV curve run can be seen (Fig. 4).

In our experiments, the baseline MS-(B1)8 mutant produces pore that have an open pore current of approximately 150 pA at +160 mV (Fig. 5).

The experiment was repeated with the MS-(B1-I105Y)8 mutant which exhibited a large number of pores with a higher residual current. For these channels, the open pore current was approximately 200 pA at +160 mV (Fig. 6).

The experiment was repeated with the MS-(B1-I105N)8 mutant which exhibited two main distributions of current levels. Ten out of sixteen pores gave a higher residual current in a tight distribution. For these channels, the open pore current was approximately 280 pA at +160 mV (Fig. 7).

Example 7 - a MS-(B1-I105A)8 pore that spontaneously changes conductance

MspA mutant pores have been observed to spontaneously change conductance during electrical recording experiments.

Electrical measurements were acquired as described in example 6, using the MS-(B1-I105A)8 mutant pore.

A single MspA mutant pore is capable of interchanging between high and low conductance states spontaneously (Fig. 8). This suggests that the mutations to the MspA allow conformational changes that are rarely observed in the baseline MS-(B1)8 pore. It is possible that mutations at the I105 position stabilise the high conductance state of the pore.

Example 8 - Comparison DNA currents when moving DNA through the baseline MS-(B1)8 pores compared to the MS-(B1-I105A)8 pores

The $MS\text{-}(B1)_8$ pore and the $MS\text{-}(B1\text{-}I105N)_8$ pores were compared in unzipping mode using Phi29 DNA polymerase with set physical conditions of: 400 mM KCl, 10 mM Hepes, 1 mM EDTA, 1 mM DTT, pH 8.0, +180 mV. The DNA used in this experiment was a 100mer

hairpin with a 34mer single stranded 5' overhang (SEQ ID NO: 15). The runs were conducted at room temperature.

Current levels were extracted as events from the DNA in the enzyme bound state these events were indexed and the current level, duration and variance of the event recorded.

The spread of current levels from the DNA strand moving through the MS-$(B1)_8$ mutant was ~ 30 pA under these conditions (Fig. 9). The same experiment was repeated using the MS-$(B1\text{-}I105A)_8$ mutant, the current levels exhibited a range of ~ 40 pA for the same DNA strand (Fig. 10). The larger current range of the MS-$(I105A)_8$ mutant is desirable to discriminate combinations of nucleotides within the nanopore.

Example 9 – Signal noise comparison of the MS-(B1)8 baseline to the MS-(B1-L88N)8 mutants

The noise levels of MspA pores can be controlled by mutating the L88 position in the MspA monomer sequence. It was demonstrated that the noise level can be reduced by 19% as a result of making a single mutation to the MspA monomer.

This example compares the MS-(B1)8 pore and the MS-(B1-L88N)8 pores in translocating mode, by using a helicase to control the movement of intact DNA strands through a nanopore.

Materials

Primers were designed to amplify a ~400 bp fragment of PhiX174. Each of the 5'-ends of these primers included a 50 nucleotide non-complimentary region, either a homopolymeric stretch or repeating units of 10 nucleotide homopolymeric sections. These serve as identifiers for controlled translocation of the strand through a nanopore, as well as determining the directionality of translocation. In addition, the 5'-end of the forward primer was "capped" to include four 2'-O-Methyl-Uracil (mU) nucleotides and the 5'-end of the reverse primer was chemically phosphorylated. These primer modifications then allow for the controlled digestion of predominantly only the antisense strand, using lambda exonuclease. The mU capping protects the sense strand from nuclease digestion whilst the PO4 at the 5' of the antisense strand promotes it. Therefore after incubation with lambda exonuclease only the sense strand of the duplex remains intact, now as single stranded DNA (ssDNA). The generated ssDNA was then PAGE purified as previously described.

The DNA substrate design used in this experiment is shown in Fig. 11 (SEQ ID NOs: 19 and 20 (sequences and tags presented below)). The DNA substrate consists of a 400base section of ssDNA from PhiX, with a 50T 5'-leader to aid capture by the nanopore. Annealed to this

strand just after the 50T leader is a primer containing a 3' cholesterol tag (3' Cholesteryl-TEG) to enrich the DNA on the surface of the bilayer, and thus improve capture efficiency.

SEQ ID NO: 19

mUmUmUmUTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TGGTTGTTTCTGTTGGTGCTGATATTGCTTTTGATGCCGACCCTAAATTTTTTGCCTGT
TTGGTTCGCTTTGAGTCTTCTTCGGTTCCGACTACCCTCCCGACTGCCTATGATGTTT
ATCCTTTGAATGGTCGCCATGATGGTGGTTATTATACCGTCAAGGACTGTGTGACTA
TTGACGTCCTTCCCCGTACGCCGGGCAATAACGTTTATGTTGGTTTCATGGTTTGGTC
TAACTTTACCGCTACTAAATGCCGCGGATTGGTTTCGCTGAATCAGGTTATTAAAGA
GATTATTTGTCTCCAGCCACTTAAGTGAGGTGATTTATGTTTGGTGCTATTGCTGGCG
GTATTGCTTCTGCTCTTGCTGGTGGCGCCATGTCTAAATTGTTTGGAGGCGGTC

SEQ ID NO: 20 (plus 3' Cholesteryl-TEG tag)

GCAATATCAGCACCAACAGAAACAACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TT/3CholTEG/

Experimental Method

Buffered solution: 400 mM NaCl, 10 mM Hepes pH 8.0, 1 mM ATP, 1 mM $MgCl_2$, 1 mM DTT

Nanopores: MS(B1)8 MspA;

MS(B1-L88N)8 MspA

Enzyme: Helicase

Electrical measurements were acquired from single MspA nanopores inserted in 1,2-diphytanoyl-glycero-3-phosphocholine lipid (Avanti Polar Lipids) bilayers. Bilayers were formed across ~100 μm diameter apertures in 20 μm thick PTFE films (in custom Delrin chambers) via the Montal-Mueller technique, separating two 1 mL buffered solutions. All experiments were carried out in the stated buffered solution. Single-channel currents were measured on Axopatch 200B amplifiers (Molecular Devices) equipped with 1440A digitizers. Ag/AgCl electrodes were connected to the buffered solutions so that the *cis* compartment (to which both nanopore and enzyme/DNA are added) is connected to the ground of the Axopatch headstage, and the *trans* compartment is connected to the active electrode of the headstage.

After achieving a single pore of either MS(B1)8 or MS(B1-L88N)8 in the bilayer, DNA polynucleotide (SEQ ID NOs: 19 and 20) and helicase were added to 100 μL of buffer and pre-incubated for 5 mins (DNA = 1.5 nM, Enzyme = 1 μM). This pre-incubation mix was added to 900 μL of buffer in the cis compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 0.15 nM, Enzyme = 0.1 μM). Helicase ATPase activity was initiated as required by the addition of

divalent metal (1 mM $MgCl_2$) and NTP (1 mM ATP) to the cis compartment. Experiments were carried out at a constant potential of +140 mV. Current levels were extracted as events from the DNA in the enzyme bound state these events were indexed and the current level, duration and variance of the event recorded.

Using the MspA pore MS-(B1)8, 31.08% of the detected events had a standard deviation >2.0 at an applied potential of +140 mV (additional data summarized in Table 18). The experiment was repeated with the MS-(B1-L88N)8 mutant where only 12.38% of the detected events exhibited a standard deviation of >2.0 at an applied potential of +140 mV (additional data summarized in Table 18). Therefore, the point mutation at L88 in the MspA monomer sequence has reduced the observed noise range by 19%

Table 18

Property	Pore	
	MS-(B1)8	MS-(B1-L88N)8
Mean S.D.	2.30	1.79
Median S.D.	1.57	1.48
% of S.D. > 2	31.08	12.38

Example 10 - Signal noise comparison of the MS-(B1)8 baseline to the MS-(B1-L88N)8, MS-(B1-L88S)8 and MS-(B1-L88Q)8 mutants

The noise levels of MspA pores can be altered by mutating the L88 position in the protein. It was demonstrated that the noise level can be reduced as a result of making a single mutation to the MspA monomer.

This example compares the MS-(B1)8 pore to the MS-(B1-L88N)8, MS-(B1-L88S)8 and MS-(B1-L88Q)8 pores in unzipping mode, by using Phi29 DNA polymerase to control the movement of intact DNA strands through a nanopore. The DNA substrate design used in all the experiments described in this example is shown in Fig. 12 (SEQ ID NOs: 21, 22 and 23). SEQ ID NO: 23 was tagged with an IDT Int Spacer 9 (iSp9) and 3' Cholesteryl-TEG (3CholTEG) as shown below. The runs were conducted at room temperature at an applied potential of +180 mV.

SEQ ID NO: 23:

CAGCGATGGAGATAC/iSp9//3CholTEG/

Experimental Method

Buffered solution: 400 mM KCl, 10 mM Hepes pH 8.0, 1 mM EDTA, 1 mM DTT

Nanopores: MS(B1)8 MspA;

MS(B1-L88N)8 MspA;

MS(B1-L88S)8 MspA;

MS(B1-L88Q)8 MspA;

Enzyme: Phi29 DNA polymerase SEQ ID NO: 4

Electrical measurements were acquired as described in example 9. After achieving a single pore of either MS(B1)8, MS(B1-L88N)8, MS(B1-L88S)8 or MS(B1-L88Q)8 in the bilayer, DNA polynucleotide (SEQ ID NOs: 21, 22 and 23) and Phi29 DNA polymerase were added to 100 μL of buffer and pre-incubated for 5 mins. This pre-incubation mix was added to 900 μL of buffer in the cis compartment of the electrophysiology chamber to initiate capture of the polymerase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 0.5 nM, Enzyme = 0.1 μM). Experiments were carried out at a constant potential of +180 mV. Current levels observed when the DNA is in the enzyme bound state were indexed and the current level, its duration and variance were recorded.

In the experiments, the baseline MS-(B1)8 mutant exhibited high levels of noise (76.15% of standard deviations > 2.0, see Table 19) at +180 mV. The other three mutants tested, (MS-(B1-L88N)8, MS-(B1-L88S)8 and MS-(B1-L88Q)8) which had a single point mutation at position L88, all observed lower levels of noise (see Table 19) than the baseline pore over the same DNA strand sequence. Therefore, it was possible to reduce signal noise by applying point mutations at position L88 in MspA monomer sequence.

Table 19

Pore	Mean S.D	Median S.D	% of S.D. > 2
MS-(B1)8	3.26	2.89	76.15
MS-(B1-L88N)8	3.22	2.60	74.18
MS-(B1-L88S)8	3.12	2.33	71.71
MS-(B1-L88Q)8	3.30	2.46	74.19

Example 11 - Overall signal range comparison of the MS-(B1)8 baseline to other MspA mutants

The signal range of MspA pores can be increased by mutating various positions within the MspA protein monomer sequence.

This example compares the MS-(B1)8 pore to the following pores - MS-(B1-D90Q)8, MS-(B1-I105L)8, MS-(B1-I105Y)8, MS-(B1-I89Y-D90S)8, MS-(B1-N86T)8 and MS-(B1-S103G)8 - pores in unzipping mode, by using a Phi29 DNA polymerase to control the movement of intact DNA strands through a nanopore. The DNA substrate design, used in all the experiments described in this example, is shown in Fig. 12 (SEQ ID NOs: 21, 22 and 23). SEQ ID NO: 23, tagged with iSp9 and 3CholTEG is shown above. The runs were conducted at room

temperature at an applied potential of +180 mV. Current levels observed when the DNA is in the enzyme bound state were indexed and the current level, its duration and variance were recorded.

Experimental Method

Buffered solution: 400 mM KCl, 10 mM Hepes pH 8.0, 1 mM EDTA, 1 mM DTT

Nanopores: MS(B1)8 MspA;

MS(B1-D90Q)8 MspA;

MS-(B1-I105L)8 MspA;

MS-(B1-I105Y)8 MspA;

MS-(B1-I89Y-D90S)8 MspA;

MS-(B1-N86T)8 MspA;

MS-(B1-S103G)8 MspA;

Enzyme: Phi29 DNA polymerase SEQ ID NO: 4

Electrical measurements were acquired as described in example 10. After achieving a single pore of either MS(B1)8, MS(B1-D90Q)8, MS(B1-I105L)8, MS(B1-I105Y)8, MS-(B1-I189Y-D90S)8, MS-(B1-N86T)8 or MS-(B1-S103G)8 in the bilayer, DNA polynucleotide (SEQ ID NOs: 21, 22 and 23) and Phi29 DNA polymerase were added to 100 μL of buffer and pre-incubated for 5 mins. This pre-incubation mix was added to 900 μL of buffer in the cis compartment of the electrophysiology chamber to initiate capture of the polymerase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 0.5 nM, Enzyme = 0.1 μM). Experiments were carried out at a constant potential of +180 mV. Current levels observed when the DNA is in the enzyme bound state were indexed and the current level, its duration and variance were recorded.

In the experiments, the baseline MS-(B1)8 mutant exhibited a maximum range of 35 pA at +180 mV (Table 20). The other 6 mutants tested (MS-(B1-D90Q)8, MS-(B1-I105L)8, MS-(B1-I105Y)8, MS-(B1-I89Y-D90S)8, MS-(B1-N86T)8 and MS-(B1-S103G)8) all observed a greater maximum range than the baseline pore (See Table 20) over the same DNA strand sequence. Therefore, it was possible to increase signal range by applying point mutations at various locations in the MspA monomer sequence.

Table 20

Entry No.	Mutant Pore	Range (pA)
1	MS(B1)8	34
2	MS-(B1-D90Q)8	70
3	MS-(B1-I105L)8	42
4	MS-(B1-I105Y)8	45
5	MS-(B1-I89Y-D90S)8	67

6	MS-(B1-N86T)8	58
7	MS-(B1-S103G)8	54

Example 12 - Overall sequencing profile comparison of the MS-(B1)8 baseline to other MspA mutants

The sequencing profile of MspA pores can be controlled by mutating a variety of positions in the MspA protein monomer sequence.

This example compares the MS-(B1)8 pore to MS-(B1-D90Q-D93S-I105A)8, MS-(B1-D90Q-Q126R)8, MS-(B1-L88N-D90Q-D91M)8, MS-(B1-L88N-D90Q-D91S)8 and MS-(B1-G75S-G77S-L88N-Q126R)8 pores in translocating mode, by using a helicase to control the movement of intact DNA strands through a nanopore.

Experimental Method

Buffered solution: 400 mM NaCl, 10 mM Hepes pH 8.0, 1 mM ATP, 1 mM $MgCl_2$, 1 mM DTT

Nanopores: MS(B1)8 MspA;

MS(B1-D90Q-D93S-I105A)8 MspA;

MS(B1-D90Q-Q126R)8 MspA;

MS(B1-L88N-D90Q-D91M)8 MspA;

MS(B1-L88N-D90Q-D91S)8 MspA;

MS(B1-G75S-G77S-L88N-Q126R)8 MspA;

Enzyme: Helicase

The experimental set-up was carried out as described in Example 9. After achieving a single pore of either MS-(B1)8, MS-(B1-D90Q-D93S-I105A)8, MS-(B1-D90Q-Q126R), MS-(B1-L88N-D90Q-D91M)8, MS-(B1-L88N-D90Q-D91S)8 or MS-(B1-G75S-G77S-L88N-Q126R)8 in the bilayer, DNA polynucleotide (SEQ ID NOs: 19 and 20 (sequence and tags shown above)) and helicase were added to 100 μL of buffer and pre-incubated for 5 mins (DNA = 1.5 nM, Enzyme = 1 μM). This pre-incubation mix was added to 900 μL of buffer in the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 0.15 nM, Enzyme = 0.1 μM). Helicase ATPase activity was initiated as required by the addition of divalent metal (1 mM $MgCl_2$) and NTP (1 mM ATP) to the cis compartment. Experiments were carried out at a constant potential of +140 mV. Current levels observed when the DNA is in the enzyme bound state were indexed and the current level, its duration and variance were recorded.

In the experiments, the baseline MS-(B1)8 mutant produced the sequencing profile shown in Fig. 13a. The experiment was repeated with the following mutants MS-(B1-D90Q-D93S-I105A)8, MS-(B1-D90Q-Q126R), MS-(B1-L88N-D90Q-D91M)8, MS-(B1-L88N-D90Q-

D91S)8 and MS-(B1-G75S-G77S-L88N-Q126R)8, which exhibited a variety of different sequencing profiles (see Fig. 13 b-f). Therefore, by making point mutations at a variety of positions within MspA monomer sequence it is possible to alter the sequencing profile that is detected.

Example 13 – Analysis of an RNA strand sequence using the MS-(B1)8 baseline pore

This example describes how the MspA baseline pore MS-(B1)8 combined with the Phi29 DNA polymerase can be used to sequence a strand of RNA.

This example uses the MS-(B1)8 pore in unzipping mode, by using a Phi29 DNA polymerase to control the movement of intact RNA strands through a nanopore. The RNA/DNA hybrid substrate design used in this experiment is shown in Fig. 14 (SEQ ID NOs: 24 and 25). SEQ ID NOs: 24 and 25 are presented below (RNA in bold). The runs were conducted at room temperature at an applied potential of +180 mV.

SEQ ID NO: 24:

5'OH-
CCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCC**CUAUUCUGUUUAUGUUUC UUGUUUGU** – 3'OH

SEQ ID NO: 25 (plus cholesterol tag):

5'Phos-
UAUUCUGUUUAUGUUUCUUGUUUGUUAGCCCCCUUUGAUAAGACAAAUA CAAAGAACAAA-3'Chol

Materials

In order to synthesize the RNA/DNA hybrid strand (120 mer in length), it was necessary to ligate SEQ ID NOs: 24 and 25 together. This was achieved by using the complementary DNA adapter strand SEQ ID NO: 26 to bring the two strands into close proximity, where they were subsequently ligated together forming the 120mer DNA/RNA hybrid SEQ ID NO: 27.

SEQ NO: 27 (plus cholesterol tag; RNA in bold):

5'OH-
CCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCC**CUAUUCUGUUUAUGUUUC UUGUUUGUUAUUCUGUUUAUGUUUCUUGUUUGUUAGCCCCCUUUGAUAA GACAAAUACAAAGAACAA**A-3'Chol

Experimental Method

Buffered solution: 400 mM KCl, 10 mM Hepes pH 8.0, 1 mM EDTA, 1 mM DTT

Nanopore: MS(B1)8 MspA;

Enzyme: Phi29 DNA polymerase SEQ ID NO: 4

Electrical measurements were acquired as described in example 10. After achieving a single pore of MS(B1)8 in the bilayer, DNA polynucleotide (SEQ ID NOs: 24 and 25) and Phi29 DNA polymerase were added to 100 μL of buffer and pre-incubated for 5 mins. This pre-incubation mix was added to 900 μL of buffer in the cis compartment of the electrophysiology chamber to initiate capture of the polymerase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 0.2 nM, Enzyme = 0.2 μM). Experiments were carried out at a constant potential of +180 mV. Current levels were extracted as events from the DNA in the enzyme bound state these events were indexed and the current level, duration and variance of the event recorded.

In the experiments, the baseline MS-(B1)8 mutant, combined with Phi29 DNA polymerase as a molecular motor, was observed to detect distinct current levels as the RNA strand was threaded through the pore. These current signals were then used to determine the sequence of the target. Typical RNA translocation events, in Phi29 DNA polymerase unzipping mode, are shown in Fig. 15.

Example 14 – MspA dimer and oligomerisation to form pores

This example describes the preparation and oligomerisation of the MspA dimer.

Preparation of Dimer

MspA NNNRRK monomeric protein consists of 184 amino acid residues. A single polypeptide was designed to make a dimeric version of MspA-NNNRRK protein.

DNA sequence encoding the 184 residue MspA-NNNRRK polypeptide was linked to a second DNA sequence encoding the identical polypeptide chain via a short DNA linker sequence. The linker DNA sequence encodes SGSGSGDDDDDDDDSGSGSS (SEQ ID NO: 33; shown as $-(SG)_3-D_8-(SG)_2(SS)-$). An initiator codon (ATG) was added just before the first base and a DNA encoding two stop codons (TAATAG) was added after the last base. Therefore, the entire DNA sequence encoding MspA-NNNRRK-$(SG)_3-D_8-(SG)_2$(SS)-MspA-NNNRRK is shown in SEQ ID NO: 28.

The DNA was synthesised at GenScript USA Inc and cloned into a pT7 vector for expression purposes.

Protein was generated by coupled in vitro transcription and translation (IVTT) by using an E. coli T7-S30 extract system for circular DNA (Promega).

The complete 1 mM amino acid mixture minus cysteine and the complete 1 mM amino acid mixture minus methionine were mixed in equal volumes to obtain the working amino acid solution required to generate high concentrations of the proteins. The amino acid mix (2.5.0 μL), premix solution (10 μL), [35S]L-methionine (0.5μL) and rifampicin (2 μL, 50 mg/mL) were mixed with plasmid DNA (4 μL,400 ng/mL) and T7 S30 extract (7.5 μL). Synthesis was carried out for 90 min at 37 °C to generate 25 μL of IVTT proteins for MspA-NNNRRK monomer and dimer. After the reaction, samples were centrifuged at 25,000g for 10 mins and the supernatant was discarded. The pellet was washed with 100 μL MBSA (10mM MOPS, 150mM NaCl, pH 7.4 cotaining 1mg/mL BSA) and resuspended in 25 μL Lamellae sample buffer. Samples were subjected to SDS-PAGE on a 10% gel. The gel was dried at 80 °C for 45 mins and exposed to X-ray film for 2 hours. The gel showed 2 distinct bands, one corresponding to the MspA dimer and one to the MspA monomer.

Oliomerisation of Monomer and Dimer

Expression of the dimer and, separately, the monomer was carried out in the presence of synthetic lipid vesicles to facilitate oligomerisation. A five component lipid mixture was used (PS: SM: PE: PC: Cholesterol in 10:10:20:30:30 ratio, 25mg/mL). 50 μL of lipid mixture was centrifuged at 25,000g for 10 mins in a 1.5 mL eppendorf tube and the supernatant was discarded. The complete 1 mM amino acid mixture minus cysteine and the complete 1 mM amino acid mixture minus methionine were mixed in equal volumes to obtain the working amino acid solution required to generate high concentrations of the proteins. The membrane pellet was resuspended with amino acid mix (10.0 μL), premix solution (40 μL), [35S]L-methionine and rifampicin (2μL, 50 mg/mL). Plasmid DNA (16 μL,400 ng/mL) and T7 S30 extract (30.0 μL) were added to initiate synthesis. Synthesis was carried out for 90 min at 37 °C to generate 100 μL of IVTT protein. IVTT reaction sample was centrifuged (25,000g, 10 mins) and the resulting membrane pellet was washed with MBSA and subjected to SDS–polyacrylamide gel electrophoresis in a 7.5% gel. The gel was dried on a watman 3M paper at 50 °C for 3 hours and exposed to X-ray film for 2 hours. The gel showed 8 distinct bands for the oligomerised MspA dimer, all of which migrated more slowly in SDS PAGE than the oligomerised monomer.

Protein purification for bilayer experiments

Three protein bands from the dimer oligomerisation experiment were excised from the gel and purified. Using the autoradiogram as the template, bands were cut and rehydrated in buffer (150 to 200 μL of 25 mM Tris.HCl, pH 8.0). The paper was removed and the gel piece was crushed using a pestle. The slurry was filtered through a QIAshredder column (Qiagen) by

centrifugation at 25,000 x gfor 10 min. The resulting protein from the third band from the monomer level was then used in the electrophysiology experiments described in Example 15.

Example 15 – Comparison of the MS-(B1)8 oligomerised from the Monomer with the MS-(B1-B1)4 oligomerised from the Dimer

This example compares the MS-(B1)8 pore oligomerised from the monomer (SEQ ID NO: 2) with the MS-(B1-B1)4 pore oligomerised from the dimer (SEQ ID NO: 29) in translocating mode, by using a helicase to control the movement of intact DNA strands (SEQ ID NOs: 19 and 20 (sequence and tags shown above)) through a nanopore.

Experimental Method

Buffered solution: 400 mM NaCl, 10 mM Hepes pH 8.0, 1 mM ATP, 1 mM $MgCl_2$, 1 mM DTT

Nanopores: MS-(B1)8;

MS-(B1-B1)4

Enzyme: Helicase

Electrical measurements were acquired using 128 well silicon chips (format 75 μm diameter, 20 μm depth and 250 μm pitch) which were silver plated (WO 2009/077734). Chips were initially washed with 20 mL ethanol, then 20 mL dH_2O, then 20 mL ethanol prior to CF4 plasma treatment. The chips used were then pre-treated by dip-coating, vacuum-sealed and stored at 4 °C. Prior to use the chips were allowed to warm to room temperature for at least 20 minutes.

Bilayers were formed by passing a series of slugs of 3.6 mg/mL 1,2-diphytanoyl-glycero-3-phosphocholine lipid (DPhPC, Avanti Polar Lipids, AL, USA) dissolved in 1 M KCl, 10 mM Tris, pH 7.5, at 0.45 μL/s across the chip. Initially a lipid slug (250 μL) was flowed across the chip, followed by a 100 μL slug of air. Two further slugs of 155 μL and 150 μL of lipid solution, each separated by a 100 μL slug of air were then passed over the chip. After bilayer formation the chamber was flushed with 3 mL of buffer at a flow rate of 3 μl/s. Electrical recording of the bilayer formation was carried out at 10 kHz with an integration capacitance of 1.0 pF.

A solution of the biological nanopore was prepared using either the MS-(B1)8 pore oligomerised from the monomer or the MS-(B1-B1)4 pore oligomerised from the dimer in 10 mM Tris, 1 mM EDTA, pH 8.0. A holding potential of +180 mV was applied and the solution flowed over the chip and pores were allowed to enter bilayers. The sampling rate and the integration capacitance were then maintained at 10 kHz and 1.0 pF respectively and the applied potential reduced to zero.

A control programme, which applied a holding potential of +180 mV, was run. DNA polynucleotide (SEQ ID NOs: 19 and 20) and helicase were pre-incubated for 5 mins. This pre-incubation mix (which included $MgCl_2$ and ATP) was then flowed over the chip to initiate capture of the helicase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 1.5 nM, Enzyme = 10 nM). Experiments were carried out at a constant potential of +180 mV. Current levels were extracted as events from the DNA in the enzyme bound state. These events were indexed and the current level, duration and variance of the event recorded.

In the experiments, the baseline MS-(B1-B1)4 mutant pore formed from oligomerisation of the dimer inserted into lipid bilayers as effectively as the MS(B1)8 pore formed from oligomerisation of the monomer (see Fig. 16 showing pore insertion for the MS(B1)8 and MS-(B1-B1)4). When the monomer and dimer oligomerised pores were combined with a helicase as a molecular motor, it was possible to detect distinct current levels as the DNA strand was threaded through the pore. Typical DNA translocation events, in helicase translocating mode, are shown in Fig. 17 for the MS-(B1)8 pore formed from oligomerisation of the monomer and Fig. 18 for the MS-(B1-B1)4 pore formed from oligomerisation from the dimer. Therefore, the MS-(B1-B1)4 pore mutant oligomerised from the dimer unit was found to be as good a pore as the MS-(B1)8 pore mutant oligomerised from the monomer unit.

Example 16 – Use of the MS-(B1-L88N)8 mutant MspA pore to distinguish 5-methylcytosine from cytosine

This example describes how the MS-(B1-L88N)8 mutant pore of MspA can be used to distinguish cytosine from its epigenetically modified base 5-methylcytosine. The DNA substrate designs used in this experiment are shown in Fig. 19 and have the following sequences:

TTTTTTTTT/idSp/TTTTTTTTmCTTTTTTTTCTTTTTTTTmCGTTTTTTTTCGTTTTTTTTGTATCTCCATCGCTGCCCCCTTTTTCCCCCTTTTT (which is SEQ ID NO: 30 with 9 T nucleotidesand an IDT Int d Spacer (idSp) at the 5' end). mC represents 5-methylcytosine

GGCAGCGATGGAGATACTTGAGGCGAGCGGTCAA (SEQ ID NO: 31) and

5CholTEG/TTGACCGCTCGCCTC (SEQ ID NO: 32 with a 5' Cholesteryl-TEG tag).

Materials

In order to form the DNA strand construct shown in Fig. 19 it was necessary to hybridise SEQ ID NO: 30, 31 and 32 together. This was carried out by pre-incubating all three strands at the same time.

Experimental Method

Buffered solution: 1 M KCl, 10 mM Hepes pH 8.0, 1 mM ATP, 1 mM $MgCl_2$, 1 mM DTT

Nanopores: MS(B1-L88N)8 MspA

Enzyme: Helicase

The experimental set-up was carried out as described in Example 9. After achieving a single pore of MS-(B1-L88N)8, in the bilayer, DNA polynucleotide (SEQ ID NOs: 30, 31 and 32) and helicase were added to 50 µL of buffer and pre-incubated for 5 mins (DNA = 5 nM, Enzyme = 100 nM). This pre-incubation mix was added to 950 µL of buffer in the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 5 nM, Enzyme = 100 nM). Helicase ATPase activity was initiated as required by the addition of divalent metal (1 mM $MgCl_2$) and NTP (1 mM ATP) to the cis compartment. Experiments were carried out at a constant potential of +120 mV. Current levels were extracted as events from the DNA in the enzyme bound state. These events were indexed and the current level, duration and variance of the event recorded.

In the experiments it was observed that cytosine and 5-methylcytosine produced different current levels when translocated through the MS-(B1-L88N)8 pore under the control of a helicase (see Fig. 20). Therefore, using this mutated form of MspA it was possible to distinguish cytosine from its epigenetically modified base 5-methylcytosine.

CLAIMS

1. A mutant Msp monomer comprising a variant of the sequence shown in SEQ ID NO: 2, wherein the variant comprises at least one of the following mutations:

(a) asparagine (N), serine (S), glutamine (Q) or threonine (T) at position 88;

(b) serine (S), glutamine (Q) or tyrosine (Y) at position 90;

(c) leucine (L) or serine (S) at position 105;

(d) arginine (R) at position 126;

(e) serine (S) at position 75;

(f) serine (S) at position 77;

(g) arginine (R) at position 59;

(h) glutamine (Q) , asparagine (N) or threonine (T) at position 75;

(i) glutamine (Q) , asparagine (N) or threonine (T) at position 77;

(j) leucine (L) at position 78;

(k) asparagine (N) at position 81;

(l) asparagine (N) at position 83;

(m) serine (S) or threonine (T) at position 86;

(n) phenylalanine (F), valine (V) or leucine (L) at position 87;

(o) tyrosine (Y), phenylalanine (F), valine (V), arginine (R), alanine (A), glycine (G) or cysteine (C) at position 88;

(p) phenylalanine (F), valine (V) or leucine (L) at position 89;

(q) leucine (L), phenylalanine (F), tryptophan (W), histidine (H), threonine (T), glycine (G), alanine (A), valine (V), arginine (R), lysine (K), asparagine (N) or cysteine (C) at position 90;

(r) serine (S), glutamine (Q), leucine (L), methionine (M), isoleucine (I), alanine (A), valine (V), glycine (G), phenylalanine (F), tryptophan (W), tyrosine (Y), histidine (H), threonine (T), arginine (R), lysine (K), asparagine (N) or cysteine (C) at position 91;

(s) alanine (A) or serine (S) at position 92;

(t) serine (S), alanine (A), threonine (T), glycine (G) at position 93;

(u) leucine (L) at position 94;

(v) valine (V) at position 95;

(w) arginine (R), aspartic acid (D), valine (V), asparagine (N), serine (S) or threonine (T) at position 96;

(x) serine (S) at position 97;

(y) serine (S) at position 98;

(z) serine (S) at position 99;

(aa) serine (S) at position 100;

(bb) phenylalanine (F) at position 101;

(cc) lysine (K), serine (S) or threonine (T) at position 102;

(dd) alanine (A), glutamine (Q), asparagine (N), glycine (G) or threonine (T) at position 103;

(ee) isoleucine at position 104;

(ff) tyrosine (Y), alanine (A), glutamine (Q), asparagine (N), threonine (T), phenylalanine (F), tryptophan (W), histidine (H), glycine (G), valine (V), arginine (R), lysine (K), proline (P), or cysteine (C) at position 105;

(gg) phenylalanine (F), isoleucine (I), valine (V) or serine (S) at position 106;

(hh) proline (P) or serine (S) at position 108;

(ii) asparagine (N) at position 118;

(jj) serine (S) or cysteine (C) at position 103; and

(kk) cysteine at one or more of positions 10 to 15, 51 to 60, 136 to 139 and 168 to 172.

2. A mutant according to claim 1, wherein the variant comprises one or more of the following substitutions:

(a) one or more of (i) serine (S) at positon 75, (ii) serine (S) at position 77, (iii) asparagine (N) at position 88 , (iv) glutamine (Q) at position 90 and (v) arginine (R) at position 126;

(b) one or more of (i) glutamine (Q) at position 90 and (ii) arginine (R) at position 126;

(c) one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) arginine (R) at position 126;

(d) one or more of (i) serine (S) at position 88 and (ii) glutamine (Q) at position 90;

(e) one or more of (i) asparagine (N) at position 88 and (ii) glutamine (Q) at position 90;

(f) one or more of (i) glutamine (Q) at position 90 and (ii) alanine (A) at position 105;

(g) one or more of (i) serine (S) at position 90 and (ii) serine (S) at position 92;

(h) one or more of (i) threonine (T) at position 88 and (ii) serine (S) at position 90;

(i) one or more of (i) glutamine (Q) at position 87 and (ii) serine (S) at position 90;

(j) one or more of (i) tyrosine (Y) at position 89 and (ii) serine (S) at position 90;

(k) one or more of (i) asparagine (N) at position 88 and (ii) phenylalanine (F) at position 89;

(l) one or more of (i) asparagine (N) at position 88 and (ii) tyrosine (Y) at position 89;

(m) one or more of (i) serine (S) at position 90 and (ii) alanine (A) at position 92;

(n) one or more of (i) serine (S) at position 90 and (ii) asparagine (N) at position 94;

(o) one or more of (i) serine (S) at position 90 and (ii) isoleucine (I) at position 104;

(p) one or more of (i) aspartic acid (D) at position 88 and (ii) lysine (K) at position 105;

(q) one or more of (i) asparagine (N) at position 88 and (ii) arginine (R) at position 126;

(r) one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) arginine (R) at position 91;

(s) one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) serine (S) at position 91;

(t) one or more of (i) asparagine (N) at position 88, (ii) glutamine (Q) at position 90 and (iii) valine (V) at position 105;

(u) one or more of (i) glutamine (Q) at position 90, (ii) serine (S) at position 93 and (iii) alaine (A) at position 105;

(v) one or more of (i) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) at position 90, (ii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) at position 91 and (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) at position 105;

(w) one or more of (i) serine (S), threonine (T), glycine (G), alanine (A) or valine (V) at position 90, (ii) serine (S), threonine (T), glycine (G), alanine (A) or valine (V) at position 91 and (iii) serine (S), threonine (T), glycine (G), alanine (A) or valine (V) at position 105;

(x) serine (S), arginine (R), lysine (K) or histidine (H) at position 90 and/or serine (S), arginine (R), lysine (K) or histidine (H) at position 91;

(y) serine (S), threonine (T), asparagine (N), glutamine (Q), tyrosine (Y) or histidine (H) at position 90 and/or serine (S), threonine (T), asparagine (N), glutamine (Q), tyrosine (Y) or histidine (H) at position 91; and

(z) cysteine at one or more of positions 90, 91 and 103.

3. A mutant according to claim 1 or 2, wherein the variant comprises at least one of the following substitution(s):

(i) L88N;
(ii) L88S;
(iii)L88Q;
(iv)L88T;
(v) D90S;
(vi)D90Q;
(vii) D90Y;
(viii) I105L;
(ix)I105S;
(x) Q126R;
(xi)G75S;
(xii) G77S;
(xiii) G75S, G77S, L88N and Q126R;
(xiv) G75S, G77S, L88N, D90Q and Q126R;

(xv) D90Q and Q126R;
(xvi) L88N, D90Q and Q126R;
(xvii) L88S and D90Q;
(xviii) L88N and D90Q;
(xix) E59R;
(xx) G75Q;
(xxi) G75N;
(xxii) G75S;
(xxiii) G75T;
(xxiv) G77Q;
(xxv) G77N;
(xxvi) G77S;
(xxvii) G77T;
(xxviii) I78L;
(xxix) S81N;
(xxx) T83N;
(xxxi) N86S;
(xxxii) N86T;
(xxxiii) I87F;
(xxxiv) I87V;
(xxxv) I87L;
(xxxvi) L88N;
(xxxvii) L88S;
(xxxviii) L88Y;
(xxxix) L88F;
(xl) L88V;
(xli) L88Q;
(xlii) L88T;
(xliii) I89F;
(xliv) I89V;
(xlv) I89L;
(xlvi) N90S;
(xlvii) N90Q;
(xlviii) N90L;
(xlix) N90Y;
(l) N91S;
(li) N91Q;
(lii) N91L;
(liii) N91M;
(liv) N91I;
(lv) N91A;
(lvi) N91V;
(lvii) N91G;
(lviii) G92A;
(lix) G92S;
(lx) N93S;
(lxi) N93A;
(lxii) N93T;
(lxiii) I94L;
(lxiv) T95V;
(lxv) A96R;
(lxvi) A96D;
(lxvii) A96V;
(lxviii) A96N;
(lxix) A96S;
(lxx) A96T;
(lxxi) P97S;
(lxxii) P98S;
(lxxiii) F99S;
(lxxiv) G100S;
(lxxv) L101F;
(lxxvi) N102K;
(lxxvii) N102S;
(lxxviii) N102T;
(lxxix) S103A;
(lxxx) S103Q;
(lxxxi) S103N;
(lxxxii) S103G;
(lxxxiii) S103T;
(lxxxiv) V104I;
(lxxxv) I105Y;
(lxxxvi) I105L;
(lxxxvii) I105A;
(lxxxviii) I105Q;
(lxxxix) I105N;
(xc) I105S;
(xci) I105T;
(xcii) T106F;
(xciii) T106I;
(xciv) T106V;
(xcv) T106S;
(xcvi) N108P;
(xcvii) N108S;
(xcviii) D90Q and I105A;
(xcix) D90S and G92S;
(c) L88T and D90S;
(ci) I87Q and D90S;

(cii) I89Y and D90S;
(ciii) L88N and I89F;
(civ) L88N and I89Y;
(cv) D90S and G92A;
(cvi) D90S and I94N;
(cvii) D90S and V104I;
(cviii) L88D and I105K;
(cix) L88N and Q126R;
(cx) L88N, D90Q and D91R;
(cxi) L88N, D90Q and D91S;
(cxii) L88N, D90Q and I105V;
(cxiii) D90Q, D93S and I105A;
(cxiv) N91Y;
(cxv) N90Y and N91G;
(cxvi) N90G and N91Y;
(cxvii) N90G and N91G;
(cxviii) I05G;
(cxix) N90R;
(cxx) N91R;
(cxxi) N90R and N91R;
(cxxii) N90K;
(cxxiii) N91K;
(cxxiv) N90K and N91K;
(cxxv) N90Q and N91G;
(cxxvi) N90G and N91Q;
(cxxvii) N90 Q and N91Q;
(cxxviii) R11 8N;
(cxxix) N91C;
(cxxx) N90C;
(cxxxi) N90W;
(cxxxii) N91 W;
(cxxxiii) N90 K;
(cxxxiv) N91 K;
(cxxxv) N90 R;
(cxxxvi) N91 R;
(cxxxvii) N90 S and N91S;
(cxxxviii) N90 Y and I105A;
(cxxxix) N90 G and I105A;
(cxl) N90Q and I105A;
(cxli) N90S and I105A;
(cxlii) L88A and I105A;
(cxliii) L88S and I105S;
(cxliv) L88N and I105N;
(cxlv) N90G and N93G;
(cxlvi) N90G;
(cxlvii) N93G;
(cxlviii) N90 G and N91A;
(cxlix) I105K;
(cl) I105R;
(cli) I105V;
(clii) I105P;
(cliii) I105W;
(cliv)
(clv) L88R;
(clvi) L88A;
(clvii) L88G;
(clviii) L88N;
(clix) N90R and I105A;
(clx) N90S and I105A;
(clxi) L88A and I105A;

(clxii) L88S and I105S;

(clxiii) L88N and I105N;

(clxiv) L88C;

(clxv) S103C; and

(clxvi) I105C.

4. A mutant according to any one of the preceding claims, wherein the mutant is chemically modified.

5. A mutant according to claim 4, wherein the mutant is chemically modified by attachment of a molecule to one or more cysteines, attachment of a molecule to one or more lysines, attachment of a molecule to one or more non-natural amino acids, enzyme modification of an epitope or modification of a terminus.

6. A mutant according to claim 5, wherein the one or more cysteines have been introduced to the mutant by substitution.

7. A mutant according to claim 5 or 6, wherein the molecule is (a) a molecular adaptor that facilitates the interaction between a pore comprising the monomer and a target nucleotide or target nucleic acid sequence or (b) a nucleic acid binding protein..

8. A mutant according to any one or claims 5 to 7, wherein the attachment is via a linker.

9. A mutant according to any one of claims 5 to 8, wherein the molecule is attached to one or more of positions 90, 91 and 103 of SEQ ID NO: 2.

10. A construct comprising two or more covalently attached monomers derived from Msp.

11. A construct according to claim 10, wherein the two or more monomers are the same or different.

12. A construct according to claim 10 or 11, wherein at least one monomer comprises the sequence shown in SEQ ID NO: 2.

13. A construct according to any one of claims 10 to 12, wherein at least one of the monomers is a mutant monomer as defined in any one of claims 1 to 8.

14. A construct according to any one of claims 10 to 13, wherein the construct comprises two monomers and at least one of the monomers is a mutant as defined in any one of claims 1 to 8.

15. A construct according to any one of claims 10 to 14, wherein the monomers are genetically fused.

16. A construct according to any one of claims 10 to 15, wherein the monomers are attached via a linker.

17. A polynucleotide which encodes a mutant according to any one of claims 1 to 3 or a construct according to claim 15.

18. A homo-oligomeric pore derived from Msp comprising identical mutant monomers according to any one of claims 1 to 3.

19. A homo-oligomeric pore according to claim 18, wherein the pore comprises eight identical mutant monomers according to any one of claims 1 to 3.

20. A hetero-oligomeric pore derived from Msp comprising at least one mutant monomer according to any one of claims 1 to 3, wherein at least one of the eight monomers differs from the others.

21. A hetero-oligomeric pore according to claim 20, wherein the pore comprises eight mutant monomers according to claim 1 and at least one of them differs from the others.

22. A hetero-oligomeric pore according to claim 21, wherein the pore comprises at least one monomer comprising the sequence shown in SEQ ID NO: 2.

23. A hetero-oligomeric pore according to claim 21 or 22, wherein the pore comprises (a) one mutant monomer and (b) seven identical monomers, wherein the mutant monomer in (a) is different from the identical monomers in (b).

24. A hetero-oligomeric pore according to any one of claims 21 to 23, wherein the pore comprises:

(a) seven monomers comprising the sequence shown in SEQ ID NO: 2 and one mutant monomer comprising the substitution N90R, N90K, N90Y, N90Q, N90W or N90C;

(b) seven monomers comprising the sequence shown in SEQ ID NO: 2 and one mutant monomer comprising the substitution N91R, N91K, N91Y, N91Q, N91W or N91C;

or

(c) seven monomers comprising the sequence shown in SEQ ID NO: 2 and one mutant monomer comprising the substitution L88C, S103C or I105C.

25. A pore according to any one of claims 18 to 24, wherein at least one of the mutant monomers is chemically-modified as defined in claims 4 to 9.

26. A pore comprising at least one construct according to claims 10 to 16.

27. A pore according to claim 26, which comprises (a) one construct as defined in claim 13 and (b) six monomers each comprising (i) the sequence shown in SEQ ID NO: 2 or (ii) a variant of SEQ ID NO: 2 as defined in claim 1 or 2.

28. A pore according to claim 26, which comprises four contructs as defined in claim 14.

29. A pore according to any one of claims 26 to 28, wherein at least one of the constructs is chemically-modified as defined in claims 4 to 9.

30. A method of characterising a target nucleic acid sequence, comprising:

(a) contacting the target sequence with a pore according to any one of claims 18 to 29 and a nucleic acid binding protein so that the protein controls the movement of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore; and

(b) measuring the current passing through the pore during each interaction and thereby characterising the target sequence.

31. A method according to claim 30, wherein characterising the target nucleic acid sequence comprises estimating the sequence of or sequencing the target nucleic acid sequence.

32. A kit for characterising a target nucleic acid sequence comprising (a) a pore according to any one of claims 18 to 29 and (b) a nucleic acid handling enzyme.

33. An apparatus for characterising target nucleic acid sequences in a sample, comprising (a) a plurality of pores according to claims 18 to 29 and (b) a plurality of nucleic acid handling enzymes.

34. An apparatus according to claim 32, wherein the apparatus comprises:

a sensor device that is capable of supporting the plurality of pores and being operable to perform nucleic acid characterisation using the pores and enzymes;

at least one reservoir for holding material for performing the characterisation;

a fluidics system configured to controllably supply material from the at least one reservoir to the sensor device; and

a plurality of containers for receiving respective samples, the fluidics system being configured to supply the samples selectively from the containers to the sensor device.

35. A method of characterising a target nucleic acid sequence, comprising:

(a) contacting the target sequence with a pore derived from Msp and a Phi29 DNA polymerase such that the polymerase controls the movement of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore; and

(b) measuring the current passing through the pore during each interaction and thereby characterising the target sequence, wherein steps (a) and (b) are carried out with a voltage applied across the pore.

36. A method according to claim 35, wherein characterising the target nucleic acid sequence comprises estimating the sequence of or sequencing the target nucleic acid sequence.

37. A method according to claim 35 or 36, wherein steps (a) and (b) are carried out in the presence of free nucleotides and an enzyme cofactor such that the polymerase moves the target sequence through the pore against the field resulting from the applied voltage.

38. A method according to claim 37, wherein the method further comprises:

(c) removing the free nucleotides such that the polymerase moves the target sequence through the pore in the opposite direction to that in steps (a) and (b) and a proportion of the nucleotides in the target sequence interacts with the pore; and

(d) measuring the current passing through the pore during each interaction and thereby proof reading the sequence of the target sequence obtained in step (b), wherein steps (c) and (d) are also carried out with a voltage applied across the pore.

39. A method according to claim 35 or 36, wherein steps (a) and (b) are carried out in the absence of free nucleotides and the presence of an enzyme cofactor such that the polymerase moves the target sequence through the pore with the field resulting from the applied voltage.

40. A method according to claim 39, wherein the method further comprises:

(c) adding free nucleotides such that the polymerase moves the target sequence through the pore in the opposite direction to that in steps (a) and (b) and a proportion of the nucleotides in the target sequence interacts with the pore; and

(d) measuring the current passing through the pore during each interaction and thereby proof reading the sequence of the target sequence obtained in step (b), wherein steps (c) and (d) are also carried out with a voltage applied across the pore.

41. A method according to claim 35 or 36, wherein steps (a) and (b) are carried out in the absence of free nucleotides and the absence of an enzyme cofactor such that the polymerase controls the movement of the target sequence through the pore with the field resulting from the applied voltage.

42. A method according to claim 41, wherein the method further comprises:

(c) lowering the voltage applied across the pore such that the target sequence moves through the pore in the opposite direction to that in steps (a) and (b) and a proportion of the nucleotides in the target sequence interacts with the pore; and

(d) measuring the current passing through the pore during each interaction and thereby proof reading the sequence of the target sequence obtained in step (b), wherein steps (c) and (d) are also carried out with a voltage applied across the pore.

43. A method of forming a sensor for characterising a target nucleic acid sequence, comprising:

(a) contacting a pore derived from Msp with a Phi29 DNA polymerase in the presence of the target nucleic acid sequence; and

(b) applying a voltage across the pore to form a complex between the pore and the polymerase;

and thereby forming a sensor for characterising the target nucleic acid sequence.

44. A method of increasing the rate of activity of a Phi29 DNA polymerase, comprising:

(a) contacting the Phi29 DNA polymerase with a pore derived from Msp in the presence of a nucleic acid sequence; and

(b) applying a voltage across the pore to form a complex between the pore and the polymerase;

and thereby increasing the rate of activity of a Phi29 DNA polymerase.

45. A method according to claim 43 or 44, which further comprises increasing the applied voltage across the pore to increase the rate of activity of the Phi29 DNA polymerase.

46. A method according to any one of claims 35 to 45, wherein at least a portion of the nucleic acid sequence is double stranded.

47. A method according to any one of claims 35 to 46, wherein the pore is as defined in any one of claims 18 to 29.

48. A method according to any one of claims 35 to 46, wherein the pore comprises eight monomers comprising the sequence shown in SEQ ID NO: 2 or a variant thereof.

49. A method according to any one of claims 35 to 48, wherein the Phi29 DNA polymerase comprises the sequence shown in SEQ ID NO: 4 or a variant thereof having at least 50% homology to SEQ ID NO: 4 based on amino acid identity over the entire sequence and retains enzyme activity.

50. A kit for characterising a target nucleic acid sequence comprising (a) a pore derived from Msp and (b) a Phi29 DNA polymerase.

51. An apparatus for characterising target nucleic acid sequences in a sample, comprising a plurality of pores derived from Msp and a plurality of Phi29 DNA polymerases.

52. An apparatus according to claim 51, wherein the analysis apparatus is as defined in claim 34.

Fig. 1

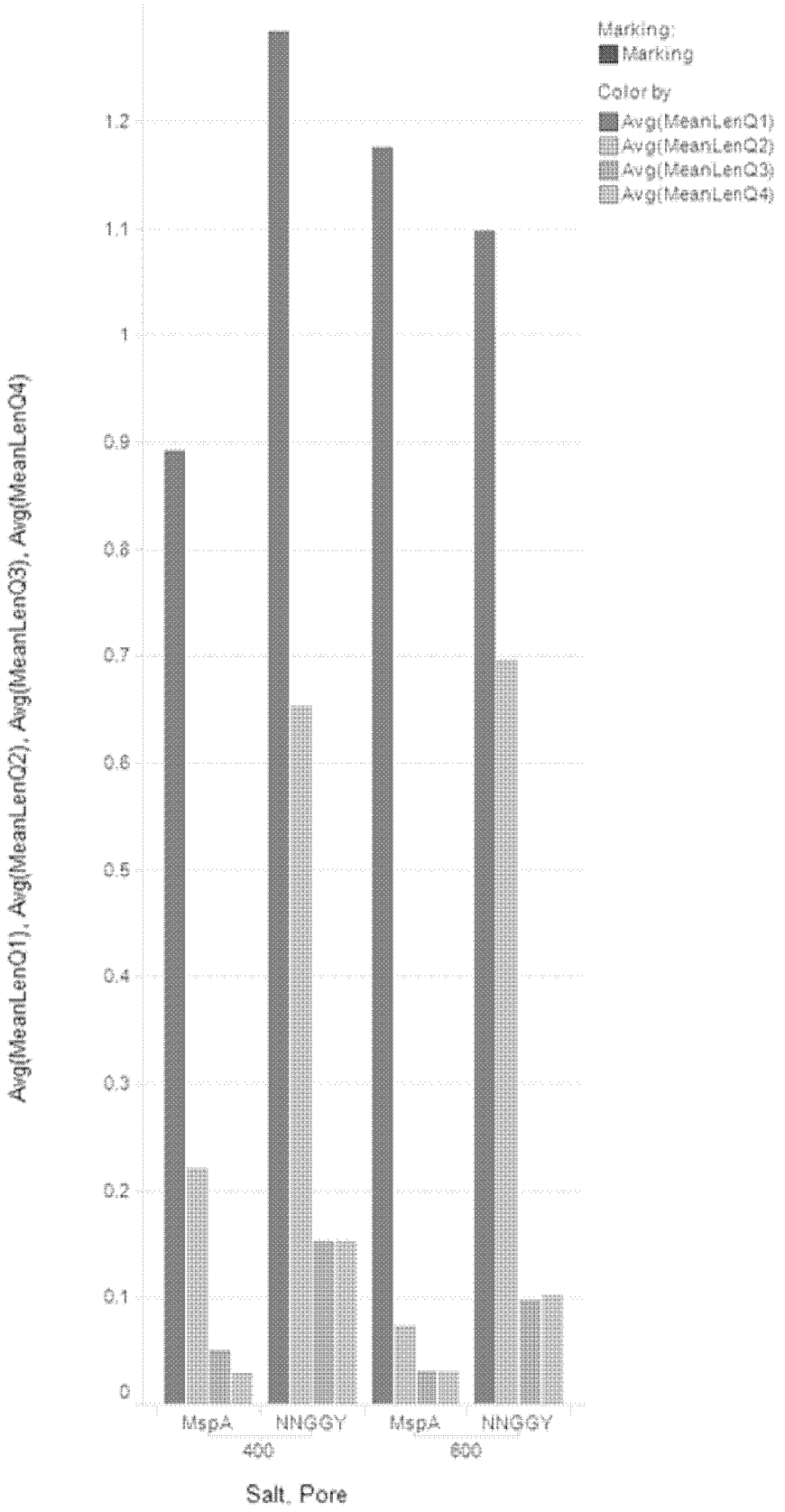

Marking:
Marking
Color by
Avg(MeanLenQ1)
Avg(MeanLenQ2)
Avg(MeanLenQ3)
Avg(MeanLenQ4)
Avg(MeanLenQ1), Avg(MeanLenQ2), Avg(MeanLenQ3), Avg(MeanLenQ4)
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
MspA
NNGGY
MspA
NNGGY
400
600
Salt, Pore

Fig. 2

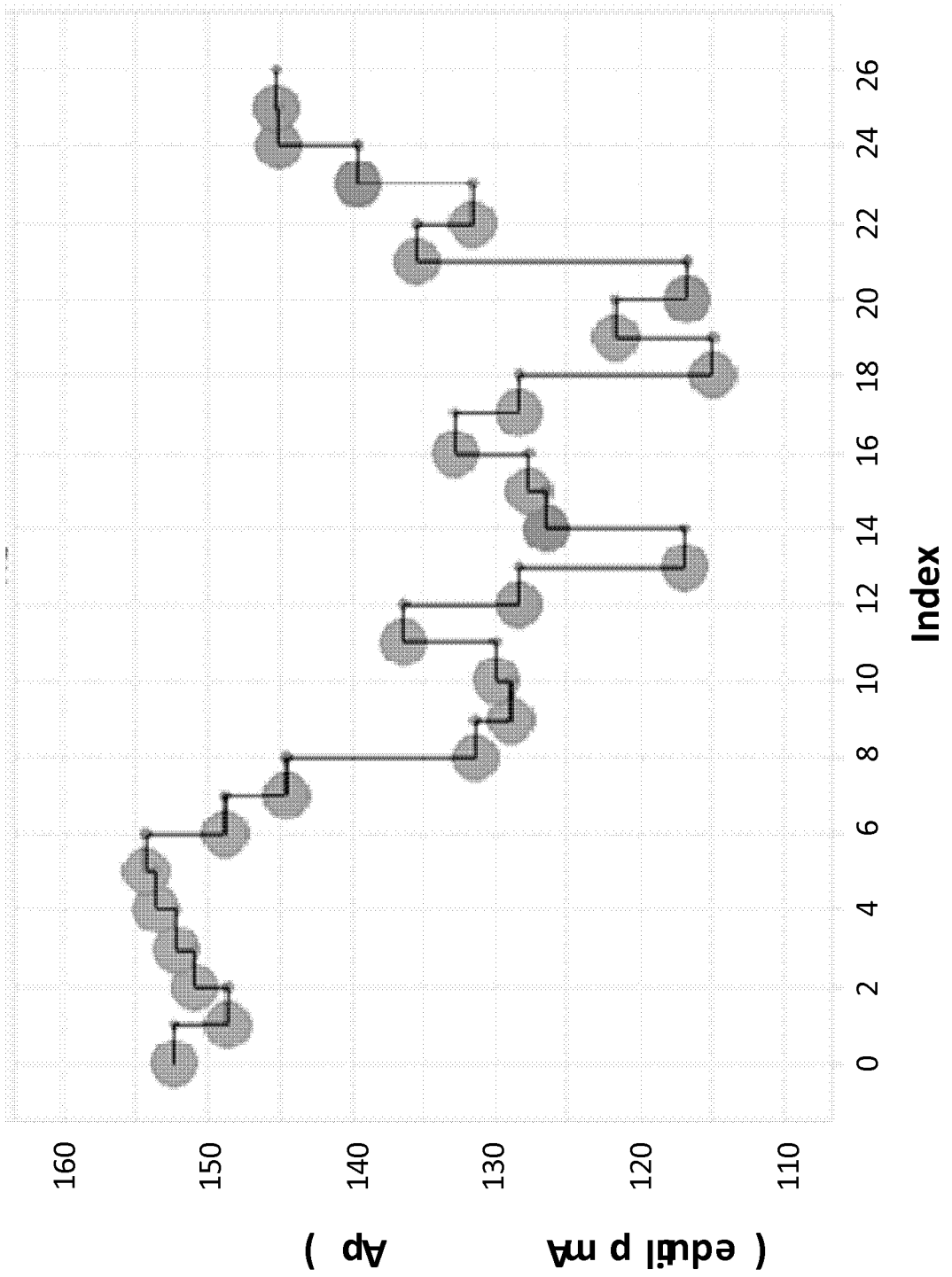


Fig. 3

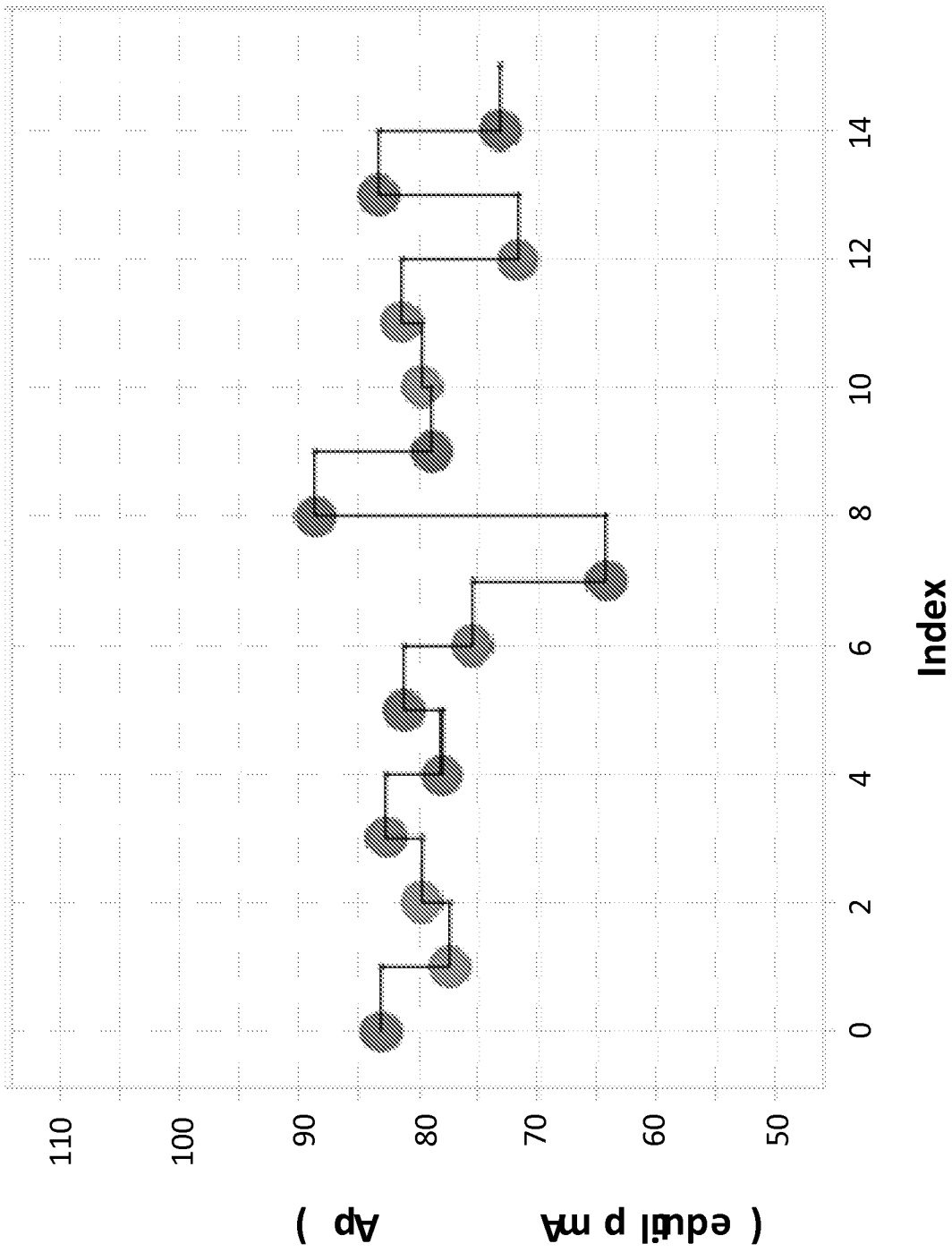


Fig. 4

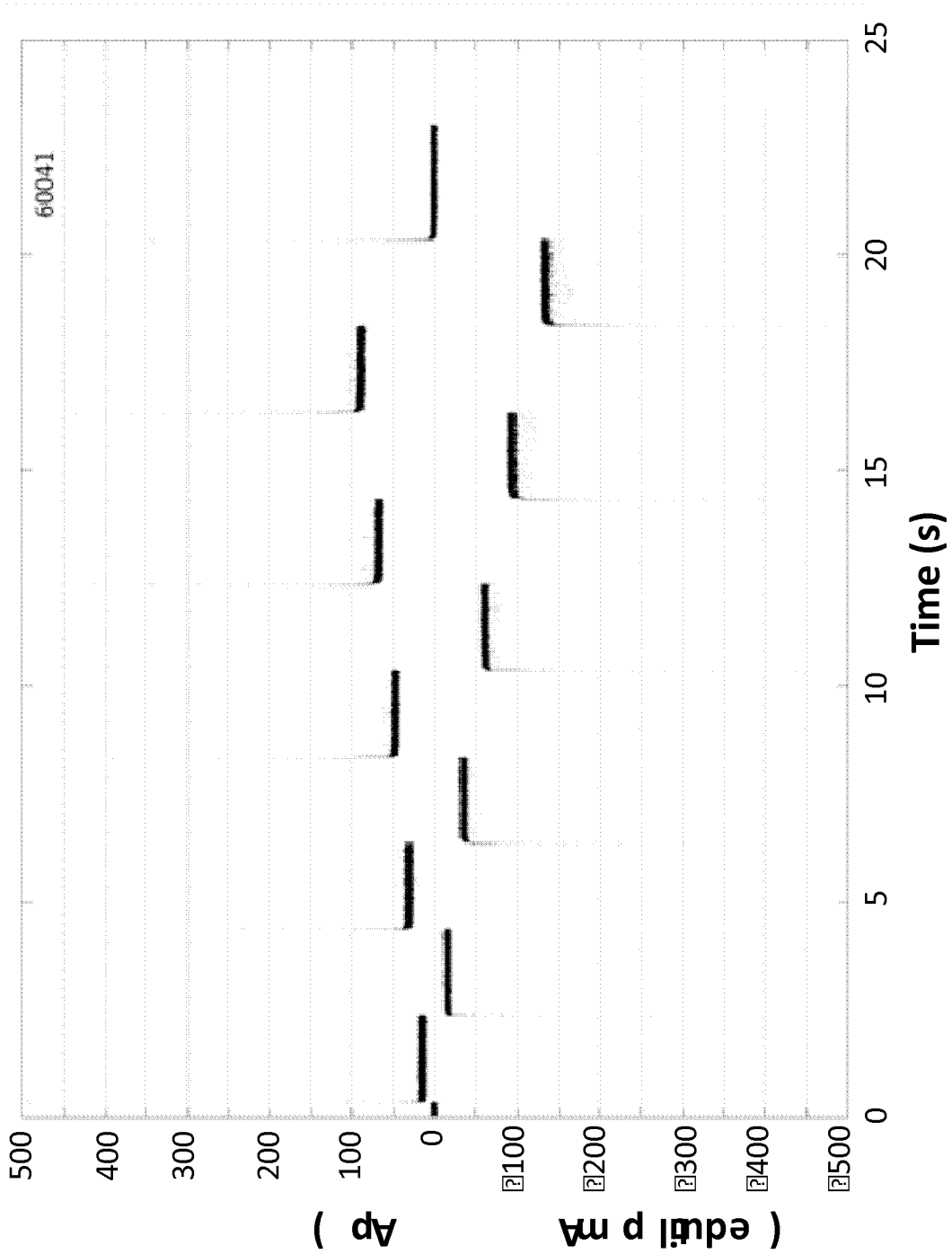

60041
500
400
300
200
100
0
100
200
300
400
500
ΔP)
(edutil p mA
0
5
10
15
20
25
Time (s)

Fig. 5

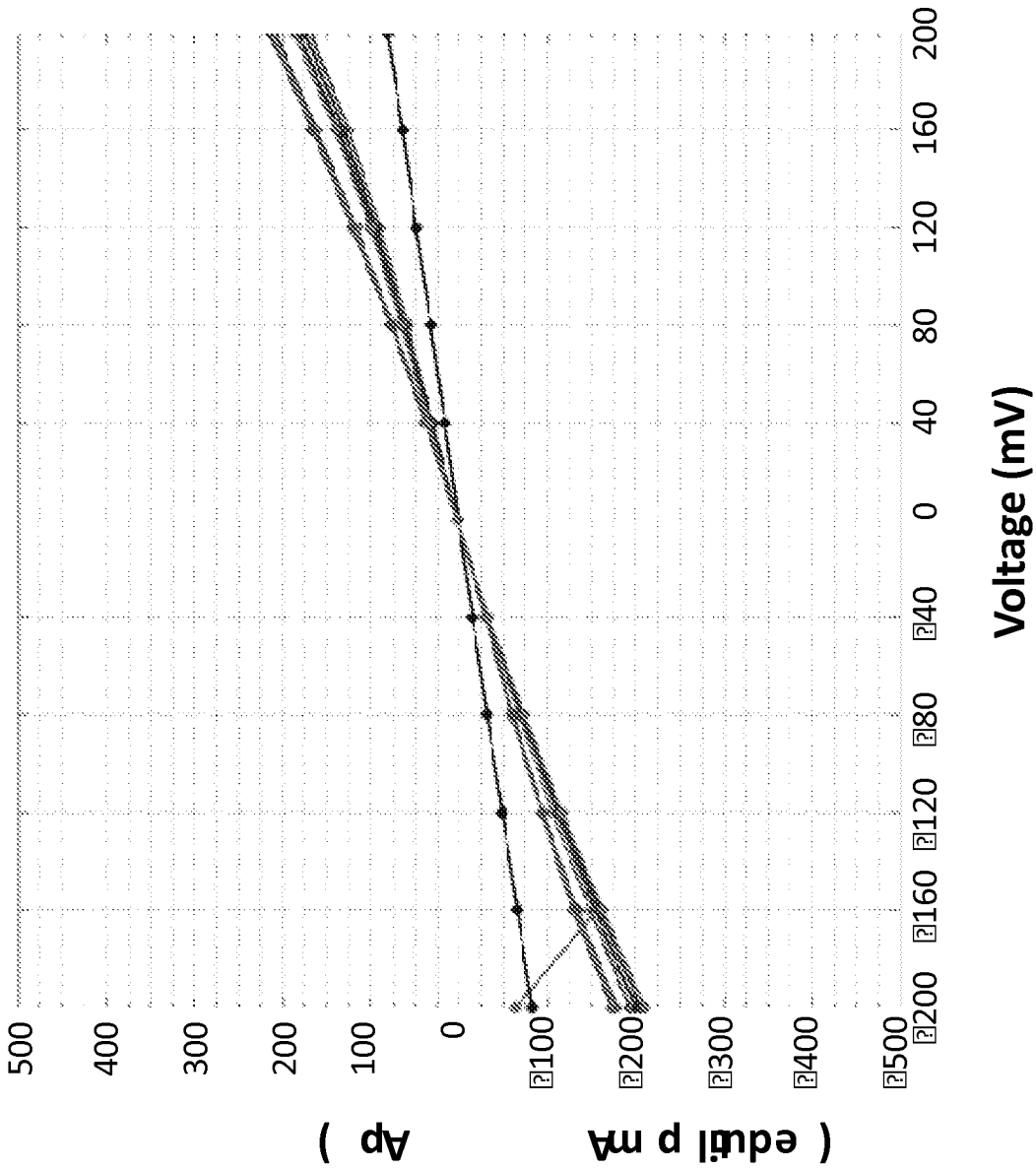

Voltage (mV)
-200
-160
-120
-80
-40
0
40
80
120
160
200
-500
-400
-300
-200
-100
0
100
200
300
400
500
Ap)
(edúíl p mA

Fig. 6

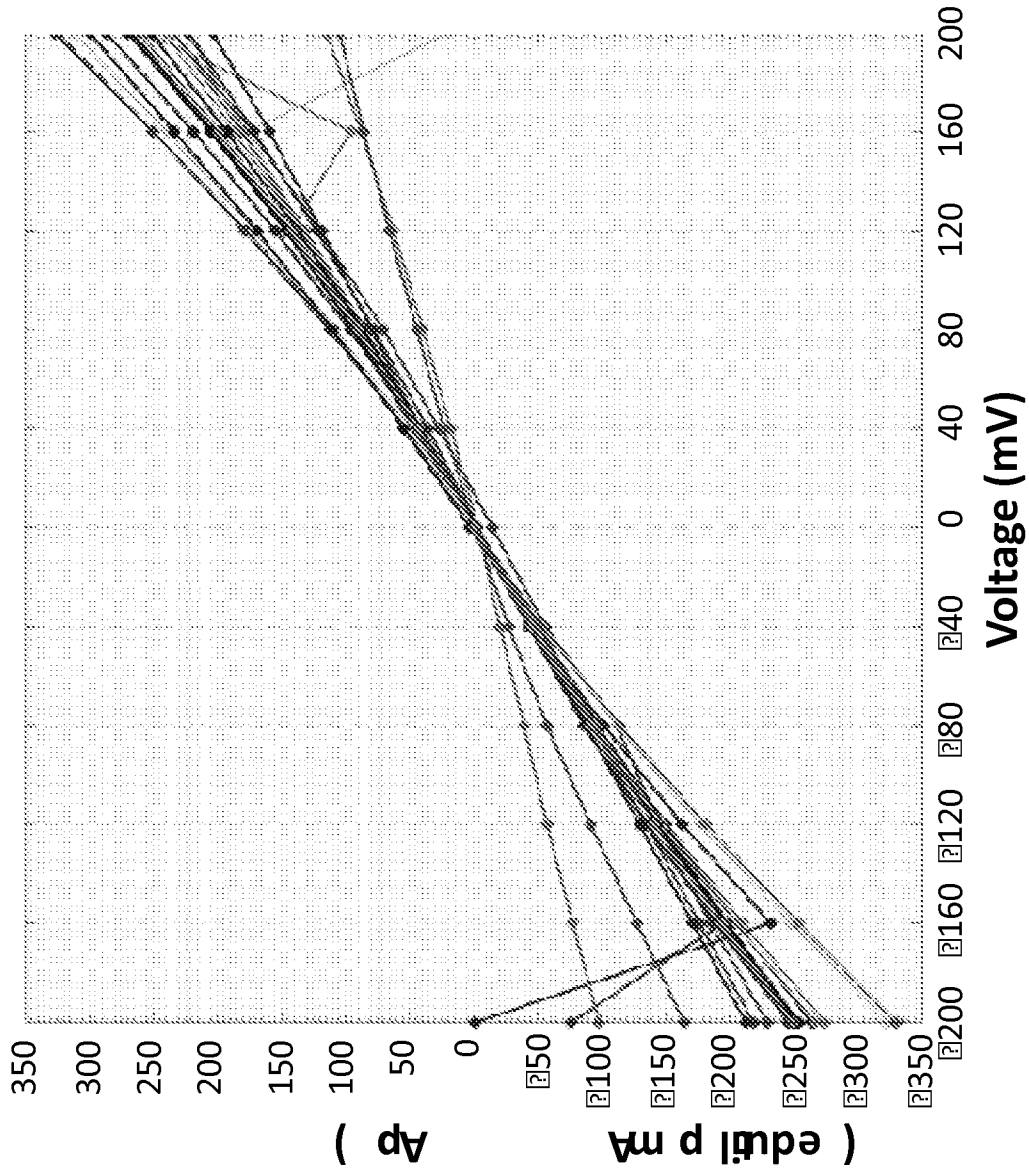

Voltage (mV)
-200
-160
-120
-80
-40
0
40
80
120
160
200
-350
-300
-250
-200
-150
-100
-50
0
50
100
150
200
250
300
350
(edúl p mA
Ap)

Fig. 7

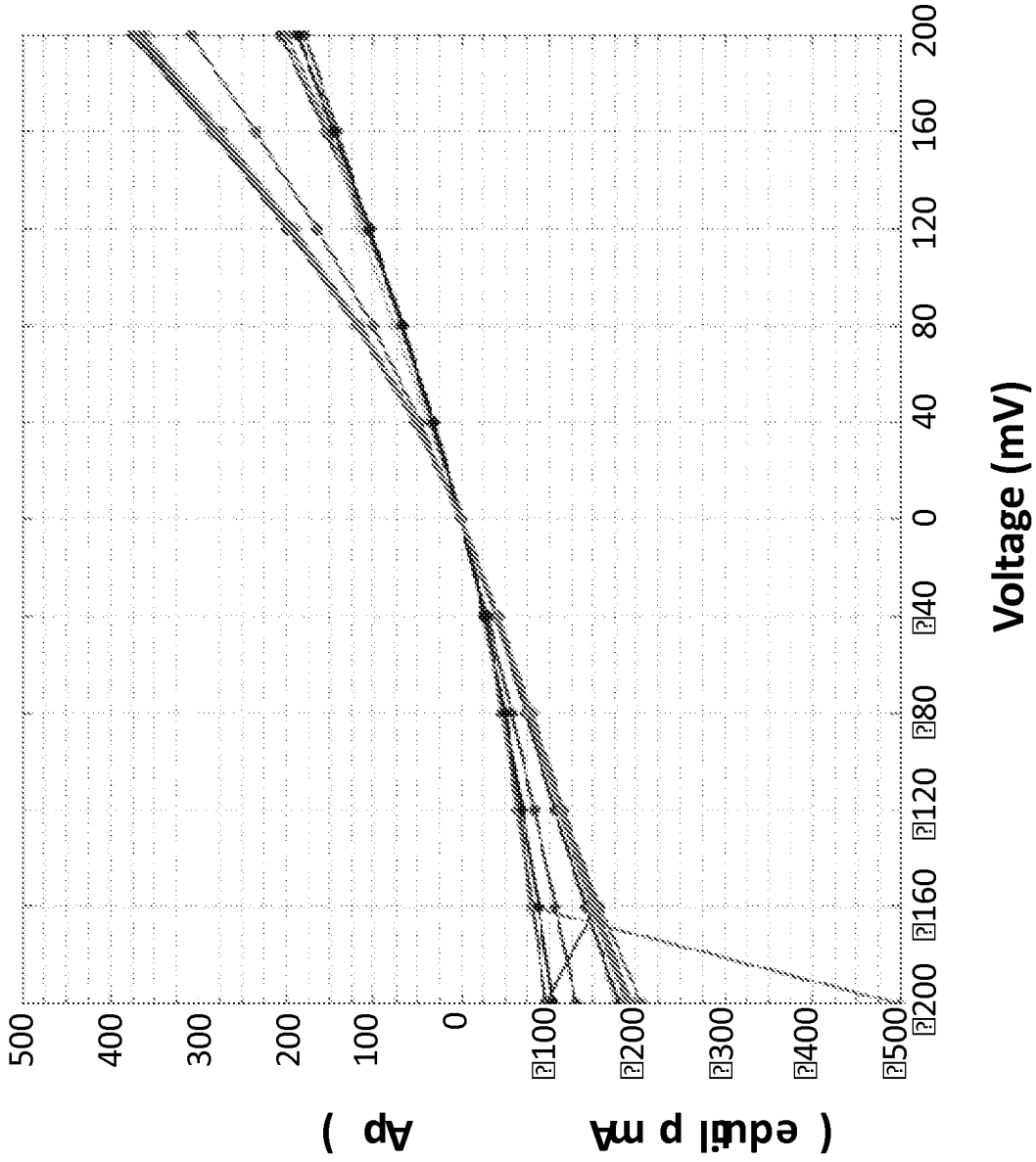

500
400
300
200
100
0
100
200
300
400
500
200
160
120
80
40
0
40
80
120
160
200
Voltage (mV)
Δp)
(edúíl p mA

Fig. 8

8/20

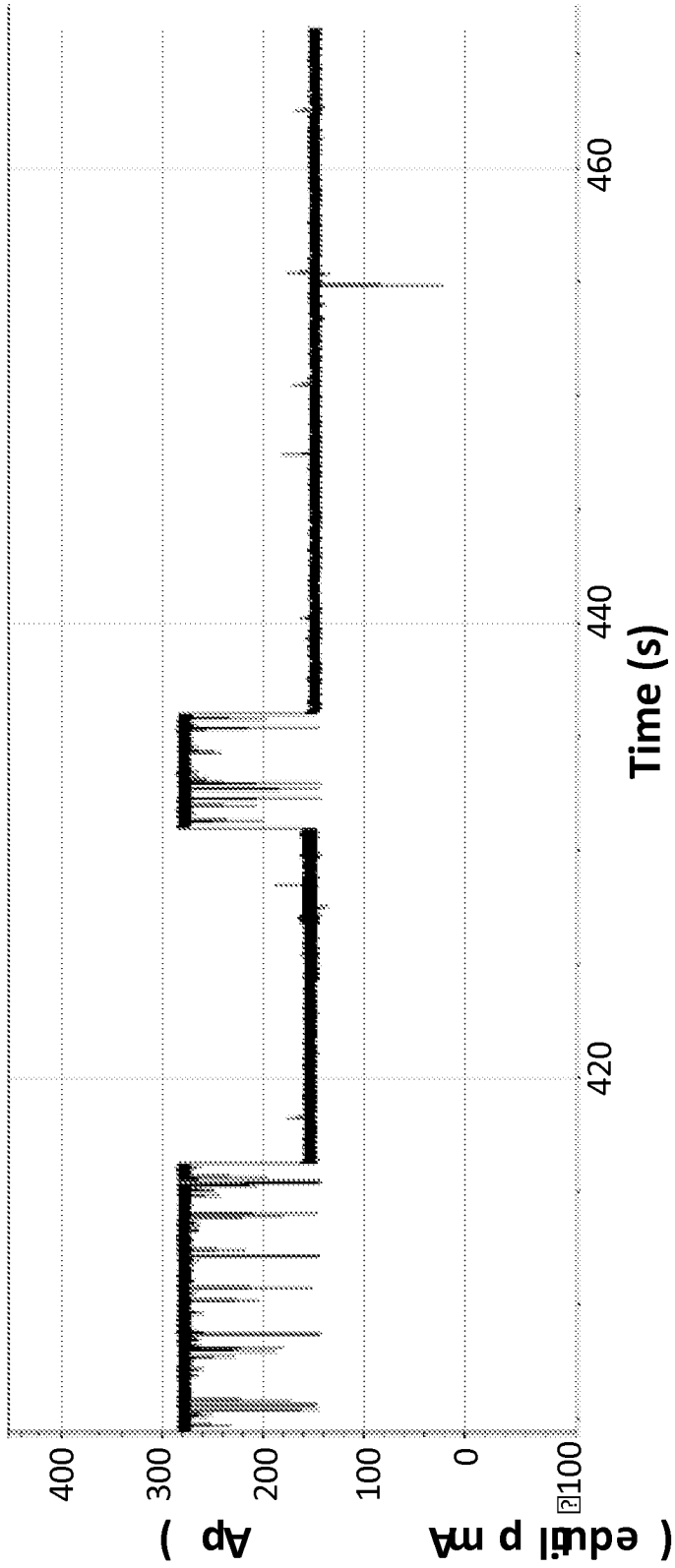


Fig. 9

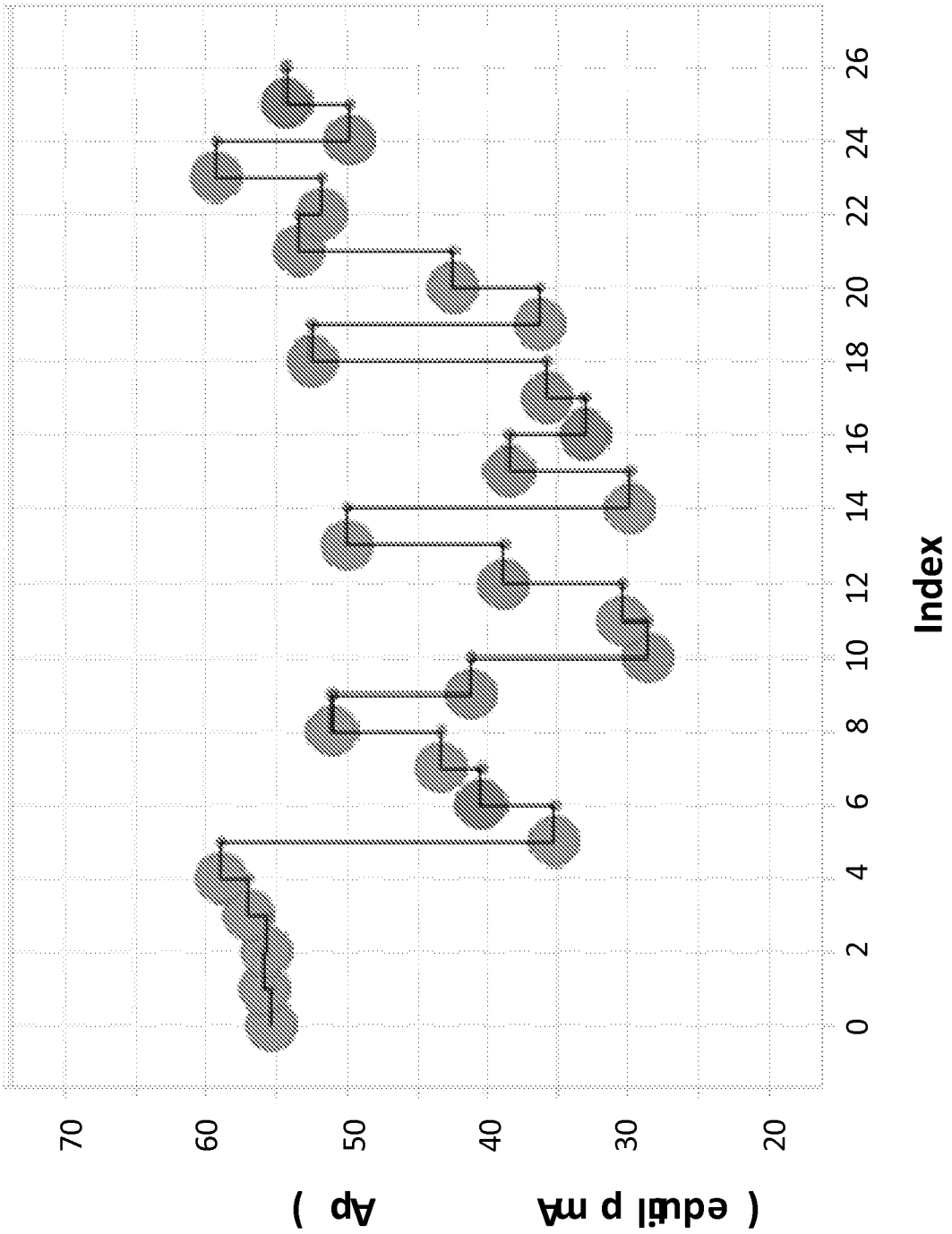


Fig. 10

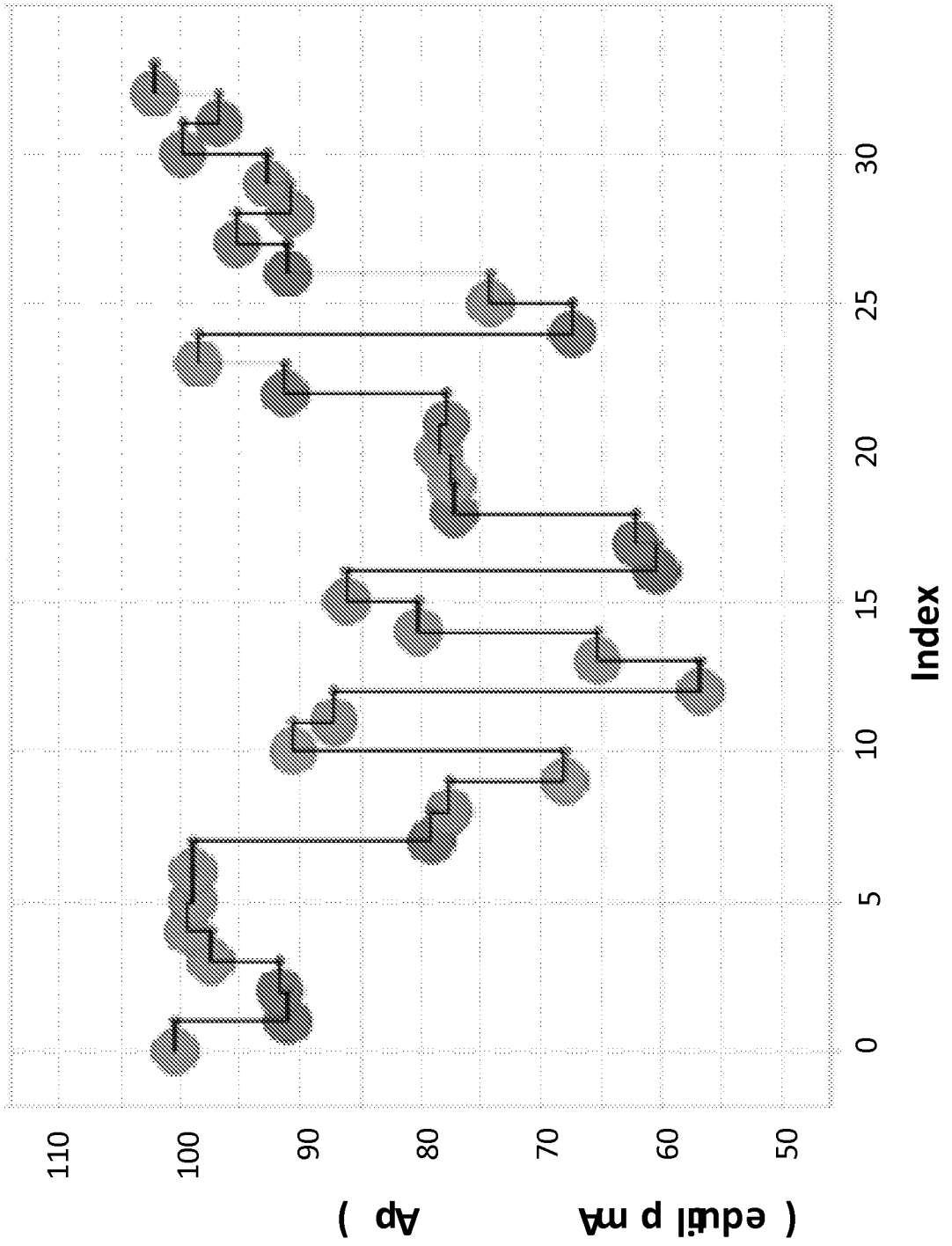


Fig. 11 11/20

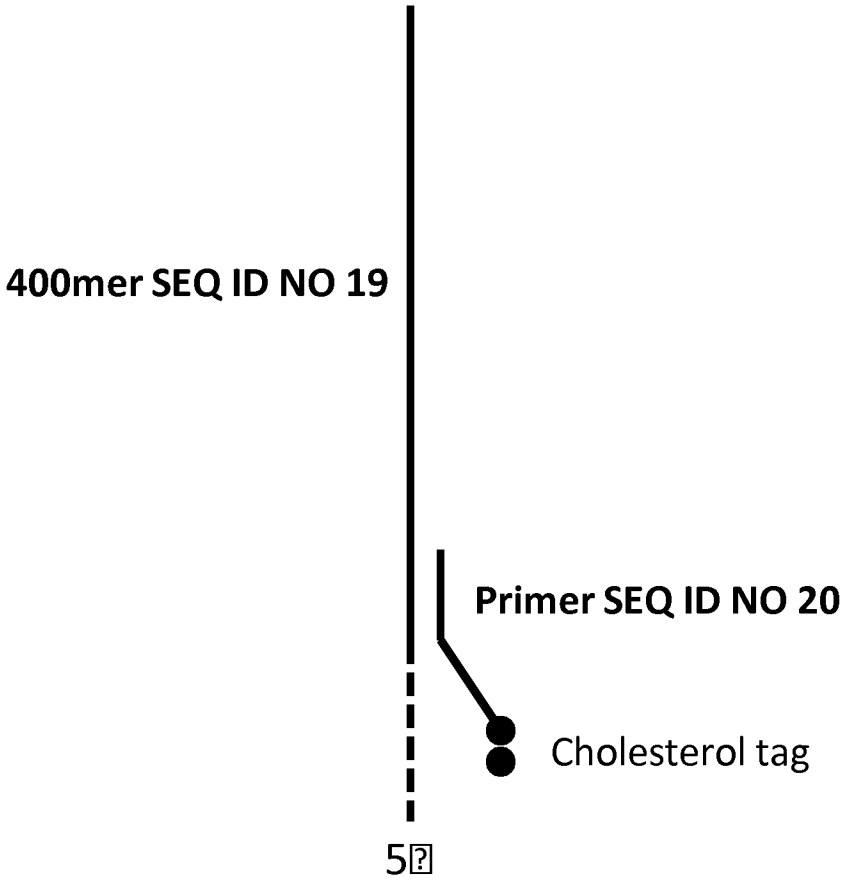


Fig. 12 12/20

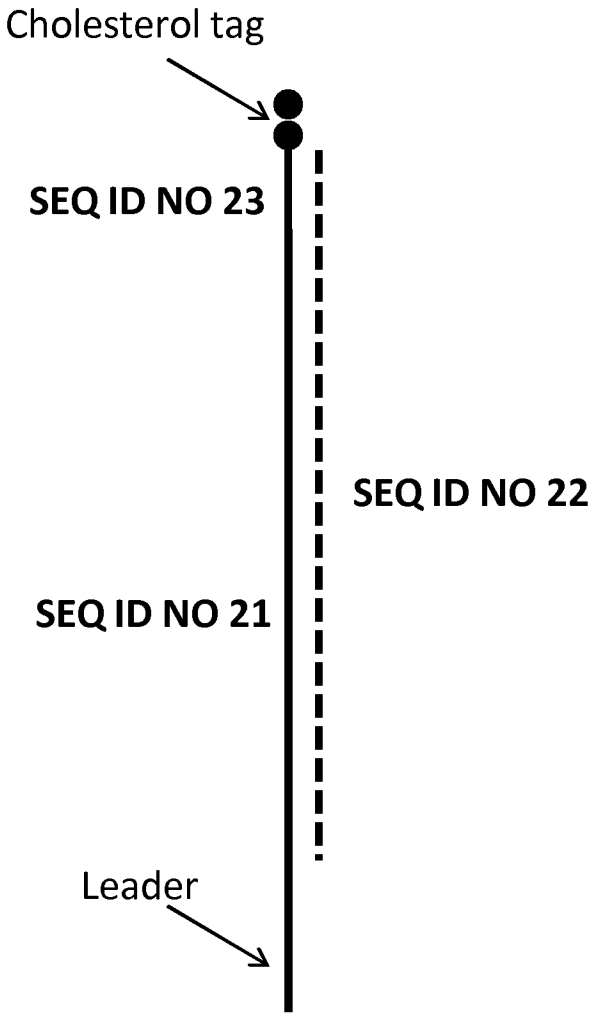


Fig. 13

13/20

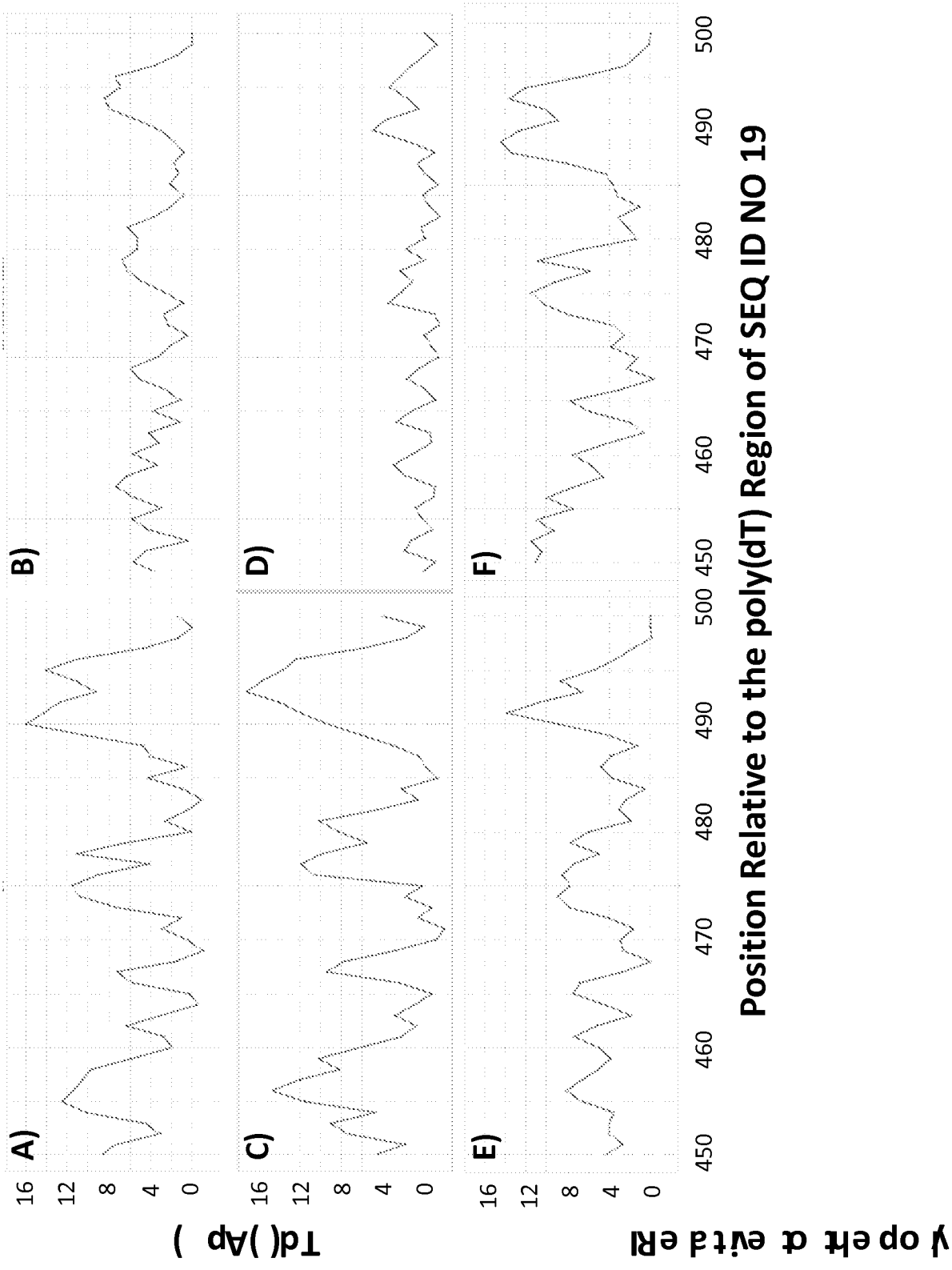

A)
B)
C)
D)
E)
F)
0
4
8
12
16
450
460
470
480
490
500
Position Relative to the poly(dT) Region of SEQ ID NO 19
Td( )Ap
Fluorescence Intensity

Fig. 14

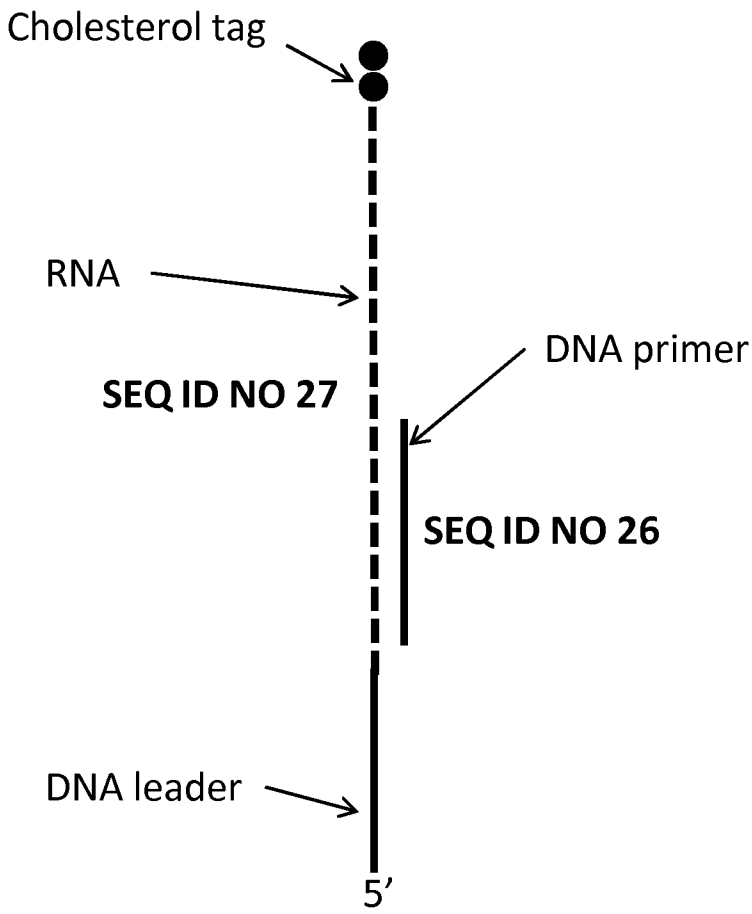

Cholesterol tag
RNA
SEQ ID NO 27
DNA primer
SEQ ID NO 26
DNA leader
5'

Fig. 15

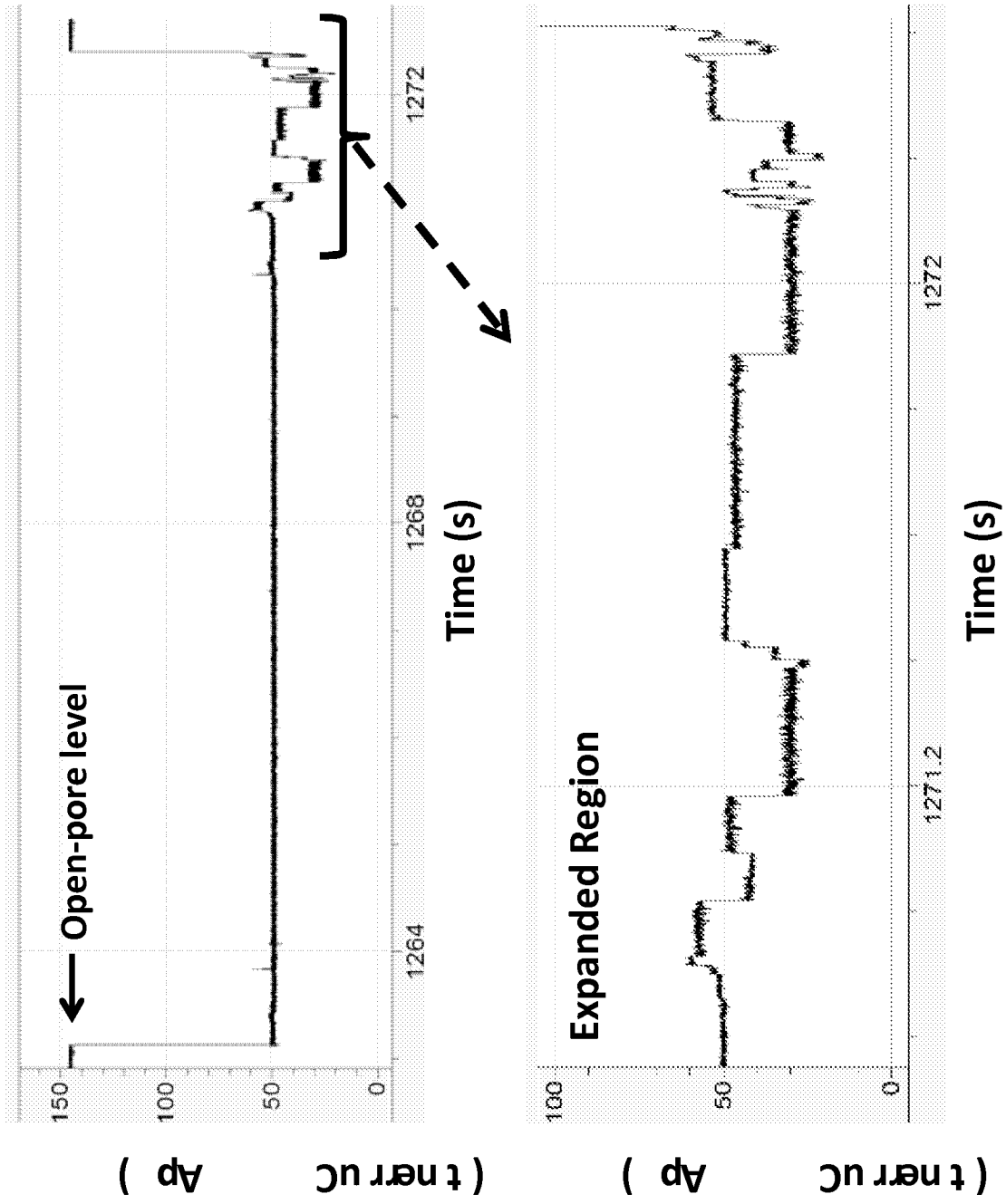

Open-pore level
Time (s)
Current (pA)
Expanded Region
Time (s)
Current (pA)

Fig. 16

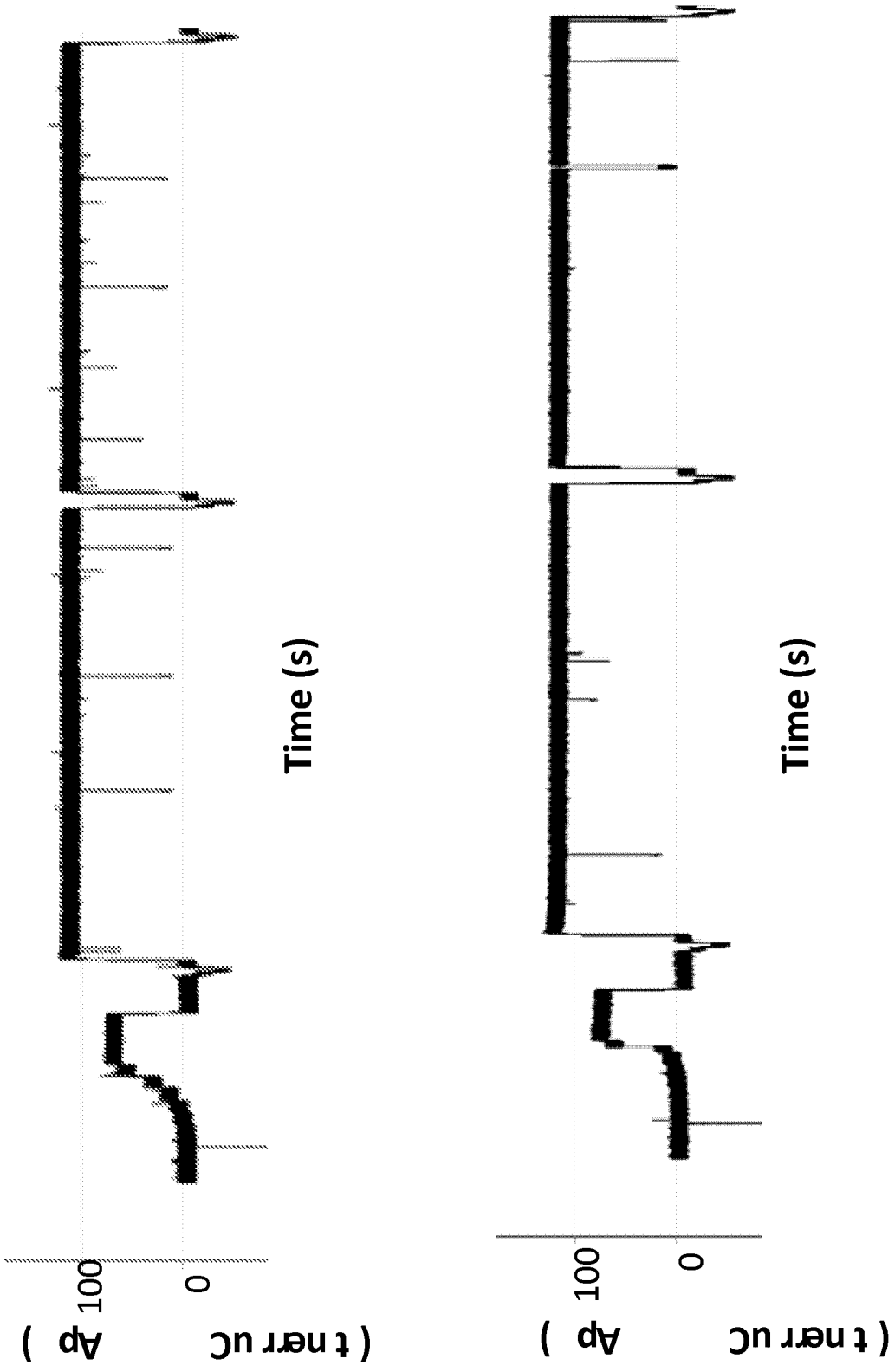


Fig. 17

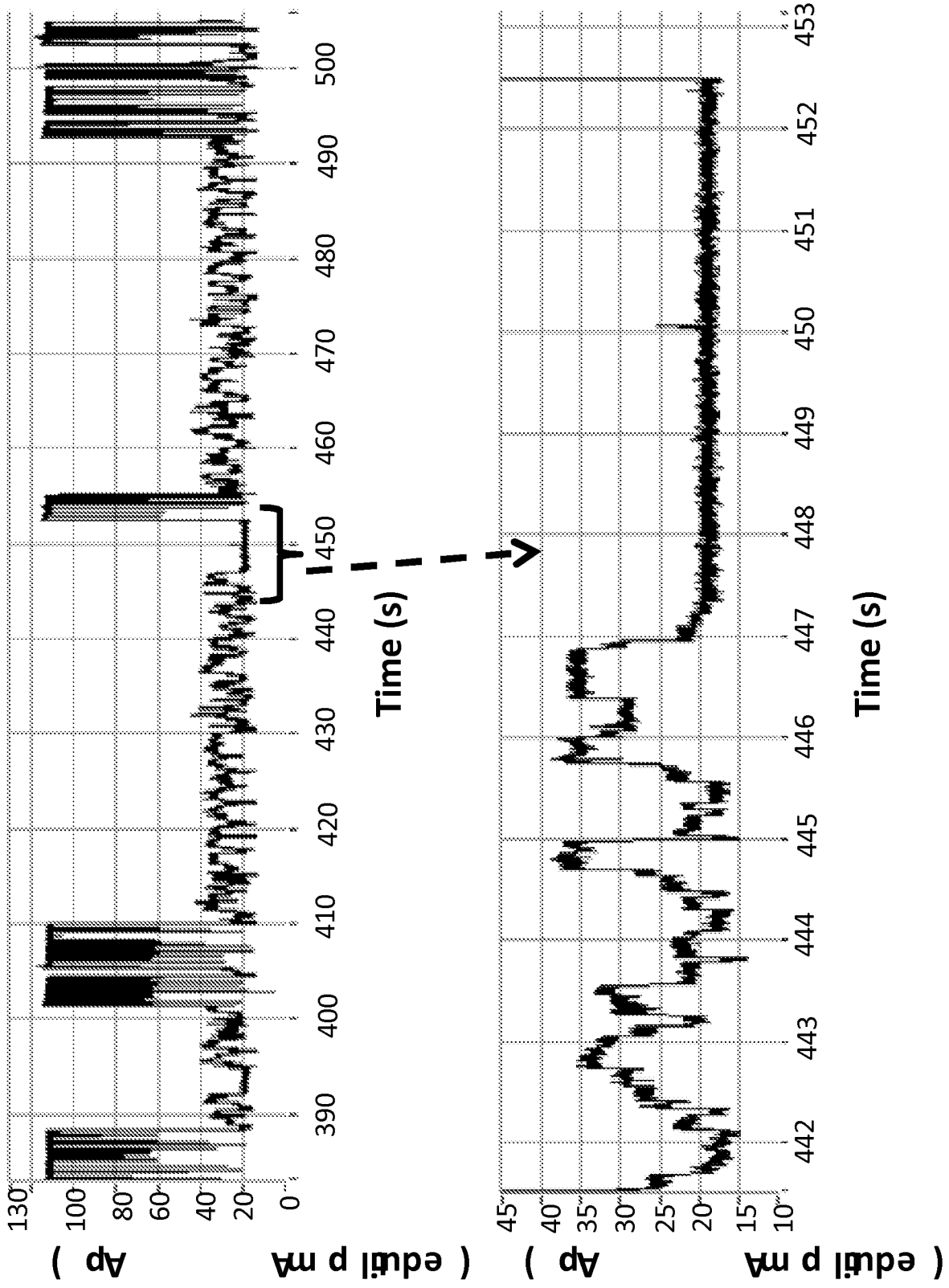

(edutil p mA
Ap)
Time (s)
130
100
80
60
40
20
0
390
400
410
420
430
440
450
460
470
480
490
500
(edutil p mA
Ap)
Time (s)
45
40
35
30
25
20
15
10
442
443
444
445
446
447
448
449
450
451
452
453

Fig. 18

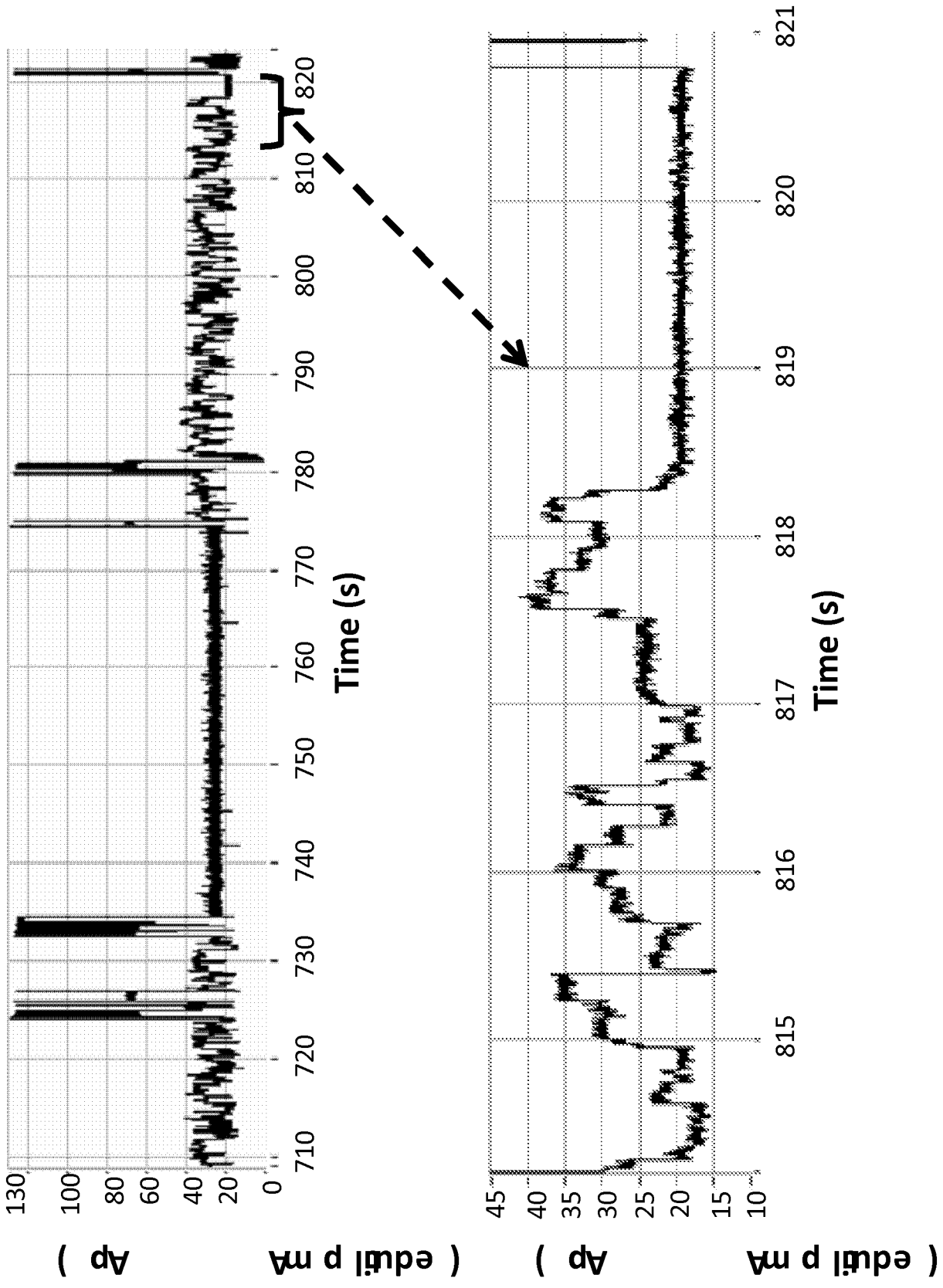
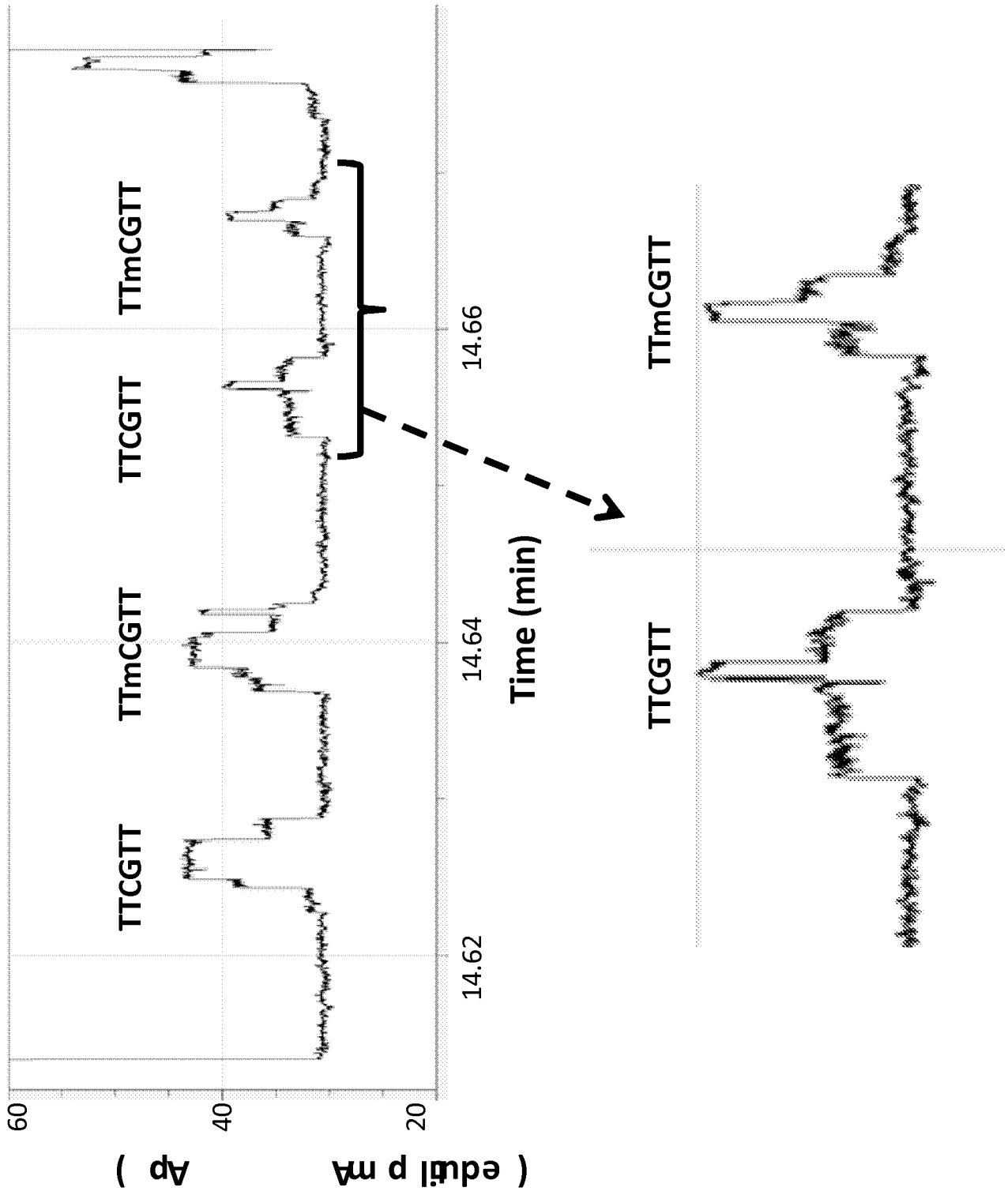

130
100
80
60
40
20
0
710
720
730
740
750
760
770
780
790
800
810
820
Time (s)
(edutil p mA
Ap)
45
40
35
30
25
20
15
10
815
816
817
818
819
820
821
Time (s)
(edutil p mA
Ap)

Fig. 19

Fig. 20


TTCGTT
TTmCGTT
TTCGTT
TTmCGTT
14.62
14.64
14.66
Time (min)
60
40
20
(edúil p mA Δp)
TTCGTT
TTmCGTT