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

**(19) World Intellectual Property Organization**
International Bureau

**(43) International Publication Date**
**5 November 2015 (05.11.2015)**

WIPO | PCT

**(10) International Publication Number**
**WO 2015/166276 A1**

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

**(21) International Application Number:**
PCT/GB2015/051291

**(22) International Filing Date:**
1 May 2015 (01.05.2015)

**(25) Filing Language:** English

**(26) Publication Language:** English

**(30) Priority Data:**

| | | |
|---|---|---|
| 1407809.1 | 2 May 2014 (02.05.2014) | GB |
| 1417708.3 | 7 October 2014 (07.10.2014) | GB |
| 1417712.5 | 7 October 2014 (07.10.2014) | GB |

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

**(72) Inventors: JAYASINGHE, Lakmal**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WALLACE, Elizabeth Jayne**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **PUGH, Jonathan Bankes**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **HAMBLEY, Richard George**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WOOD, Neil Roger**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **BROWN, Clive Gavin**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WHITE, James**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB).

**(74) Agent: CHAPMAN, Lee Phillip**; 14 South Square, Gray's Inn, London Greater London WC1R 5JJ (GB).

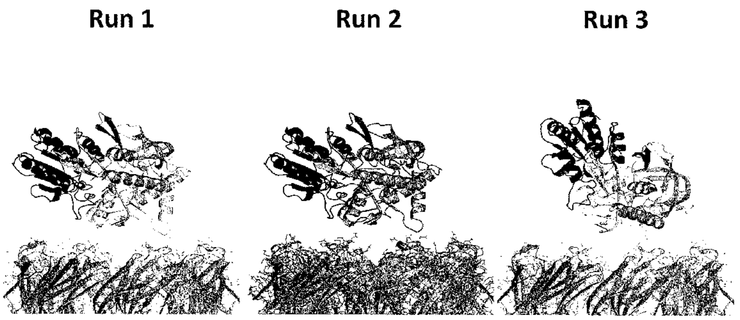
**(81) Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

*[Continued on next page]*

# (54) Title: METHOD OF IMPROVING THE MOVEMENT OF A TARGET POLYNUCLEOTIDE WITH RESPECT TO A TRANSMEMBRANE PORE

Figure 1



**(57) Abstract**: The invention relates to improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein. The invention also relates to improved transmembrane pores and polynucleotide binding proteins.

WO 2015/166276 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published**:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

# METHOD

## Field of the invention

The invention relates to improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein. The invention also relates to improved transmembrane pores and polynucleotide binding proteins.

## Background of the invention

There is currently a need for rapid and cheap polynucleotide (e.g. DNA or RNA) sequencing and identification 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 polynucleotide and require a high quantity of specialist fluorescent chemicals for signal detection.

Transmembrane pores (nanopores) have great potential as direct, electrical biosensors for polymers and a variety of small molecules. In particular, recent focus has been given to nanopores as a potential DNA sequencing technology.

When a potential is applied across a nanopore, there is a change in the current flow when an analyte, such as a nucleotide, resides transiently in the barrel for a certain period of time. Nanopore detection of the nucleotide gives a current change of known signature and duration. In the strand sequencing method, a single polynucleotide strand is passed through the pore and the identities of the nucleotides are derived. Strand sequencing can involve the use of a polynucleotide binding protein to control the movement of the polynucleotide through the pore.

The different forms of Msp are porins from *Mycobacterium smegmatis.* MspA is a 157 kDa octameric porin from *Mycobacterium smegmatis.* Wild-type MspA does not interact with DNA in a manner that allows the DNA to be characterised or sequenced. The structure of MspA and the modifications required for it to interact with and characterise DNA have been well documented (Butler, 2007, Nanopore Analysis of Nucleic Acids, Doctor of Philosophy Dissertation, University of Washington; Gundlach, Proc Natl Acad Sci U S A. 2010 Sep 14; 107(37):16060-5. Epub 2010 Aug 26; and International Application No. PCT/GB2012/050301 (published as WO/2012/107778).

## Summary of the invention

The inventors have surprisingly demonstrated that polynucleotide binding protein-controlled movement of a target polynucleotide with respect to a transmembrane pore is improved by modifying a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore.

Accordingly, the invention provides a method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein, comprising modifying a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore and thereby improving the movement of the target polynucleotide with respect to the transmembrane pore.

The invention also provides:

- a method of moving a target polynucleotide with respect to a transmembrane pore using a polynucleotide binding protein, comprising

a) providing a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified; and

b) contacting the transmembrane pore and polynucleotide binding protein provided in a) with the target polynucleotide such that the protein controls the movement of the polynucleotide with respect to the transmembrane pore;

- a method of characterising a target polynucleotide, comprising:

a) providing a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified;

b) contacting the transmembrane pore and polynucleotide binding protein provided in (a) with the target polynucleotide such that the protein controls the movement of the polynucleotide with respect to the transmembrane pore; and

c) taking one or more measurements as the polynucleotide moves with respect to the transmembrane pore, wherein the measurements are indicative of one or more characteristics of the polynucleotide, and thereby characterising the target polynucleotide;

- a transmembrane pore in which a part of the transmembrane pore which interacts with a polynucleotide binding protein has been modified;
- a mutant Msp monomer comprising a variant of SEQ ID NO: 2 in which a part of the monomer which interacts with a polynucleotide binding protein has been modified;
- a construct comprising two or more covalently attached MspA monomers, wherein at least one of the monomers is a mutant monomer of the invention;
- a homo-oligomeric pore derived from Msp comprising identical mutant monomers of the invention or identical constructs of the invention;
- a hetero-oligomeric pore derived from Msp comprising at least one mutant monomer of the invention or at least one construct of the invention;
- a polynucleotide binding protein in which a part of the protein which interacts with a transmembrane pore has been modified;
- a combination of a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified;
- a kit for characterising a target polynucleotide comprising (a) a transmembrane pore of the invention and (b) the components of a membrane;
- a kit for characterising a target polynucleotide comprising (a) a polynucleotide binding protein of the invention and (b) a polynucleotide adaptor to which the polynucleotide binding protein is optionally bound;
- an apparatus for characterising target polynucleotides in a sample, comprising (a) a plurality of transmembrane pores of the invention or a plurality of combinations of the invention and (b) a plurality of membranes; and
- a method of characterising a target polynucleotide, comprising:

    a) providing a transmembrane pore and a polymerase in which a part of the transmembrane pore which interacts with the polymerase and/or a part of the polymerase which interacts with the transmembrane pore has been modified;

    b) contacting the target polynucleotide with the transmembrane pore and polymerase provided in a) and labelled nucleotides such that phosphate labelled species are sequentially added to the target polynucleotide by the polymerase, wherein the phosphate species contain a label specific for each nucleotide; and

c) detecting the phosphate labelled species using the transmembrane pore and thereby characterising the polynucleotide.

## Description of the Figures

Figure 1 shows the three different initial simulation orientations of T4 Dda – E94C/A360C/C109A/C136A (SEQ ID NO: 24 with mutations E94C/A360C/C109A/C136A and then (ΔM1)G1G2) with respect to either MspA - (G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K = MspA mutant 1) or MspA – ((Del-L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119 = MspA mutant 2). The difference between run 1 and run 2 was that both the enzyme and pore had different side chain conformations despite the pore and enzyme being in the same position. In run three the enzyme has been tilted slightly with respect to the nanopore.

Figure 2 shows a plot (y-axis label = number of pore/enzyme contacts, x-axis label = pore amino acid residue number) of the interaction points of the nanopore MspA mutant 1 with T4 Dda – E94C/A360C/C109A/C136A. Each row of the plot shows the interaction points for the different enzyme/nanopore orientations e.g. runs 1-3.

Figure 3 shows a plot (y-axis label = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) of the interaction points of the enzyme T4 Dda – E94C/A360C/C109A/C136A with MspA mutant 1. Each row of the plot shows the interaction points for the different enzyme/nanopore orientations e.g. runs 1-3.

Figure 4 shows a plot (y-axis label = number of pore/enzyme contacts, x-axis label = pore amino acid residue number) of the interaction points of the nanopore MspA mutant 2 with T4 Dda – E94C/A360C/C109A/C136A. Each row of the plots shows the interaction points for the different enzyme/nanopore orientations e.g. runs 1-3.

Figure 5 shows a plot (y-axis label = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) of the interaction points of the enzyme T4 Dda – E94C/A360C/C109A/C136A with MspA mutant 2. Each row of the plot shows the interaction points for the different enzyme/nanopore orientations e.g. runs 1-3.

Figure 6 (A) shows two regions of a plot (y-axis label = pore amino acid residue number, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore

(MspA mutant 2) interact with particular amino acids in the enzyme (T4 Dda – E94C/A360C/C109A/C136A) from run 1. Figure 6 (B) shows a region of a plot (y-axis label (a1) = pore amino acid residue number, y-axis label (a2) = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore (MspA mutant 2) interact with particular amino acids in the enzyme (T4 Dda – E94C/A360C/C109A/C136A) from run 3. The grey bands in the plots indicate an interaction between amino acids. The darkness of the grey band corresponds to the number of interactions between enzyme/pore, with dark grey = many interactions and light grey = fewer interactions. The first amino acid in each box corresponds to the interacting amino acid in the MspA mutant 2 and the second amino acid corresponds to the interacting amino acids in T4 Dda – E94C/A360C/C109A/C136A.

Figure 7 shows DNA construct X used in Example 2. Section A corresponded to thirty iSpC3 spacers. Section B corresponded to SEQ ID NO: 28. Label C corresponded to the enzyme T4 Dda – E94C/C109A/C136A/A360C (SEQ ID NO: 24 with mutations E94C/C109A/C136A/A360C). Section D corresponded to four iSp18 spacers. Section E corresponded to SEQ ID NO: 29. Section F corresponded to four i5NitInd groups (IDT). Section G corresponded to SEQ ID NO: 30. Section H corresponded to four iSpC3 spacers. Section J corresponded to SEQ ID NO: 31. Section K corresponded to SEQ ID NO: 32. Section L corresponded to six iSp18 spacers and two thymine residues. Section M corresponded to a 3' cholesterol tether.

Figure 8 shows DNA construct Y used in Example 2. Section A corresponded to SEQ ID NO: 33. Label B corresponded to the enzyme T4 Dda – E94C/C109A/C136A/A360C (SEQ ID NO: 24 with mutations E94C/C109A/C136A/A360C). Section C corresponded to four iSpC3 spacers. Section D corresponded to SEQ ID NO: 27. Section E corresponded to four i5NitInd groups (IDT). Section F corresponded to SEQ ID NO: 34. Section G corresponded to SEQ ID NO: 32. Section H corresponded to six iSp18 spacers and two thymine residues. Section I corresponded to a 3' cholesterol tether.

Figure 9 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56W/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56W/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 10 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)E59Y/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E59Y/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 11 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 12 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56Y/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations )D56Y/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 13 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E57D/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E57D/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 14 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59T/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59T/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 15 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59Q/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59Q/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 16 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 17 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 18 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 19 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134N/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134N/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 20 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59W/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2

with mutations D56N/E59W/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 21 shows the three different initial simulation orientations of Phi29 DNA polymerase-(D12A/D66A) (SEQ ID NO: 9 with mutations D12A/D66A) with respect to αHL -(E111N/K147N)8 (SEQ ID NO: 4). The difference between run 2 and run 3 was that both the enzyme and pore had different side chain conformations despite the pore and enzyme being in the same position. In run one the enzyme has been tilted slightly with respect to the nanopore.

Figure 22 shows a plot (y-axis label = number of pore/enzyme contacts, x-axis label = pore amino acid residue number) of the interaction points of the nanopore αHL -(E111N/K147N)8 with Phi29 DNA polymerase-(D12A/D66A). Each row of the plot shows the interaction points for the different enzyme/nanopore orientations e.g. runs 1-3.

Figure 23 shows a plot (y-axis label = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) of the interaction points of the enzyme Phi29 DNA polymerase-(D12A/D66A) with αHL -(E111N/K147N)8 . Each row of the plot shows the interaction points for the different enzyme/nanopore orientations e.g. runs 1-3.

Figure 24 shows a zoomed in region of a plot (y-axis label (a1) = pore amino acid residue number, y-axis label (a2) = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore (αHL -(E111N/K147N)8 ) interact with particular amino acids in the enzyme (Phi29 DNA polymerase-(D12A/D66A)) from run 1. The grey bands in the plot indicate an interaction between amino acids. The darkness of the grey band corresponds to the number of interactions between enzyme/pore, with dark grey = many interactions and light grey = fewer interactions. The first amino acid in each box corresponds to the interacting amino acid in the αHL -(E111N/K147N)8 and the second amino acid corresponds to the interacting amino acids in Phi29 DNA polymerase-(D12A/D66A).

Figure 25 shows a zoomed in region of a plot (y-axis label (a1) = pore amino acid residue number, y-axis label (a2) = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore (αHL -(E111N/K147N)8 ) interact with particular amino acids in the enzyme (Phi29 DNA polymerase-(D12A/D66A)) from run 1. The black bands in the plot indicate an interaction between amino acids. The darkness of the grey band corresponds to the number of interactions between enzyme/pore, with dark grey = many interactions and light grey = fewer interactions. The first amino acid in each box corresponds to the interacting amino acid in the αHL -(E111N/K147N)8 and the second amino acid corresponds to the interacting amino acids in Phi29 DNA polymerase-(D12A/D66A).

Figure 26 shows two zoomed in regions of a plot (y-axis label (a1) = pore amino acid residue number, y-axis label (a2) = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore (αHL -(E111N/K147N)8 interact with particular amino acids in the enzyme (Phi29 DNA polymerase-(D12A/D66A) from run 2. The grey bands in the plot indicate an interaction between amino acids. The darkness of the grey band corresponds to the number of interactions between enzyme/pore, with dark grey = many interactions and light grey = fewer interactions. The first amino acid in each box corresponds to the interacting amino acid in the αHL -(E111N/K147N)8 and the second amino acid corresponds to the interacting amino acids in Phi29 DNA polymerase-(D12A/D66A).

Figure 27 shows a zoomed in region of a plot (y-axis label (a1) = pore amino acid residue number, y-axis label (a2) = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore (αHL -(E111N/K147N)8 interact with particular amino acids in the enzyme (Phi29 DNA polymerase-(D12A/D66A)) from run 2. The grey bands in the plot indicate an interaction between amino acids. The darkness of the grey band corresponds to the number of interactions between enzyme/pore, with dark grey = many interactions and light grey = fewer interactions. The first amino acid in each box corresponds to the interacting amino acid in the αHL -(E111N/K147N)8 and the second amino acid corresponds to the interacting amino acids in Phi29 DNA polymerase-(D12A/D66A).

Figure 28 shows a zoomed in region of a plot (y-axis label (a1) = pore amino acid residue number, y-axis label (a2) = number of pore/enzyme contacts, x-axis label = enzyme amino acid residue number) which shows which amino acids in the pore (αHL -(E111N/K147N)8 interact with particular amino acids in the enzyme (Phi29 DNA polymerase-(D12A/D66A) from run 3. The grey bands in the plot indicate an interaction between amino acids. The darkness of the grey band corresponds to the number of interactions between enzyme/pore, with dark grey = many interactions and light grey = fewer interactions. The first amino acid in each box corresponds to the interacting amino acid in the αHL -(E111N/K147N)8 and the second amino acid corresponds to the interacting amino acids in Phi29 DNA polymerase-(D12A/D66A).

Figure 29 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-

L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 30 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 31 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/K199L/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 32 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56L/E59L/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56L/E59L/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 33 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)G1A/D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations G1A/D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 34 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/K199L/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119) G1A/D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations G1A/D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 35 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for all three traces) of when a helicase (T4 Dda - E94C/C109A/C136A/A360C) controlled the translocation of the DNA construct X through the MspA nanopore MspA – ((Del-L74/G75/D118/L119)D56N/E59Y/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59Y/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Sections B and C show zoomed in regions of current trace A.

Figure 36 shows a bar chart of various enzyme/pore combinations which were investigated in order to determine the number of slips forward per kilobase and the % bases missed when construct X was translocated through the nanopore under the control of the enzyme (x-axis label = pore and enzyme combinations 1-7 (see table 12) and y-axis label A = slips forward per kilobase and y-axis label B = % bases missed ). The bar's labelled with a black star correspond to the % bases missed and those labelled with a circle correspond to the slips forward per kilobase.

Figure 37 shows a cartoon representation of the wild-type MspA nanopore. Region 1 corresponds to the cap forming region and includes residues 1-72 and 122-184. Region 2 corresponds to the barrel forming region and includes residues 73-82 and 112-121. Region 3 corresponds to the constriction and loops region and includes residues 83-111.

## Description of the Sequence Listing

SEQ ID NO: 1 shows the codon optimised polynucleotide sequence encoding the wild-type MspA monomer. This mutant lacks the signal sequence.

SEQ ID NO: 2 shows the amino acid sequence of the mature form of the wild-type MspA monomer. This mutant lacks the signal sequence.

SEQ ID NO: 3 shows the polynucleotide sequence encoding one monomer of α-hemolysin-E111N/K147N (α-HL-NN; Stoddart *et al.*, PNAS, 2009; 106(19): 7702-7707).

SEQ ID NO: 4 shows the amino acid sequence of one monomer of α-HL-NN.

SEQ ID NOs: 5 to 7 show the amino acid sequences of MspB, C and D.

SEQ ID NO: 8 shows the polynucleotide sequence encoding the Phi29 DNA polymerase.

SEQ ID NO: 9 shows the amino acid sequence of the Phi29 DNA polymerase.

SEQ ID NO: 10 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: 11 shows the amino acid sequence of exonuclease I enzyme (EcoExo I) from *E. coli*.

SEQ ID NO: 12 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: 13 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: 14 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: 15 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: 16 shows the codon optimised polynucleotide sequence derived from the bacteriophage lambda *exo* (*redX*) gene. It encodes the bacteriophage lambda exonuclease.

SEQ ID NO: 17 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 NO: 18 shows the amino acid sequence of Hel308 Mbu.

SEQ ID NO: 19 shows the amino acid sequence of Hel308 Csy.

SEQ ID NO: 20 shows the amino acid sequence of Hel308 Tga.

SEQ ID NO: 21 shows the amino acid sequence of Hel308 Mhu.

SEQ ID NO: 22 shows the amino acid sequence of TraI Eco.

SEQ ID NO: 23 shows the amino acid sequence of XPD Mbu.

SEQ ID NO: 24 shows the amino acid sequence of Dda 1993.

SEQ ID NO: 25 shows the amino acid sequence of Trwc Cba.

SEQ ID NO: 26 shows positions 72 to 82 of SEQ ID NO: 2.

SEQ ID NO: 27 shows positions 111 to 121 of SEQ ID NO: 2.

SEQ ID NO: 28 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 29 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 30 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 31 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 32 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 33 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 34 shows a polynucleotide sequence used in Example 2.

SEQ ID NO: 35 shows the polynucleotide sequence encoding the lysenin monomer.

SEQ ID NO: 36 shows the amino acid sequence of the lysenin monomer.

## 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 polynucleotide” includes two or more polynucleotides, reference to “a polynucleotide binding protein includes two or more such proteins, reference to “a helicase” includes two or more helicases, reference to “a monomer” refers to two or more monomers, reference to “a pore” includes two or more pores and the like.

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

### Modification methods

The present invention provides a method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein. The method is preferably for improving the movement of a target polynucleotide through a transmembrane pore when the movement is controlled by a polynucleotide binding protein. Target polynucleotides are discussed in more detail below.

The method comprises modifying a part of the transmembrane pore which interacts with the polynucleotide binding protein or a part of the polynucleotide binding protein which interacts with the transmembrane pore. The method may comprise modifying both the transmembrane pore and the polynucleotide binding protein.

Methods of modifying pores and proteins, such as via amino acid introductions and/or substitutions, are known in the art and are discussed in more detail below.

Characterisation, such as sequencing, of a polynucleotide using a transmembrane pore typically involves analysing polymer units made up of k nucleotides where k is a positive integer (i.e. 'k-mers'). This is discussed in International Application No. PCT/GB2012/052343 (published as WO 2013/041878). As the target polynucleotide moves with respect to, or through the pore, different k-mers within the polynucleotide are analysed, typically by measuring the current flowing through the pore. The movement of the polynucleotide with respect to, such as through, the pore can be viewed as movement from one k-mer to another or from k-mer to k-mer.

The method of the invention preferably provides more consistent movement of the target polynucleotide with respect to, such as through, the transmembrane pore. The method preferably provides more consistent movement from one k-mer to another or from k-mer to k-mer as the target polynucleotide moves with respect to, such as through, the pore. The method preferably allows the target polynucleotide to move with respect to, such as through, the transmembrane pore more smoothly. The method preferably provides more regular or less irregular movement of the target polynucleotide with respect to, such as through, the transmembrane pore.

The method preferably reduces the amount of slipping forward associated with the movement of the target polynucleotide with respect to, such as through, the pore. Some helicases including the Dda helicase used in the Example move along the polynucleotide in a 5' to 3' direction. When the 5'end of the polynucleotide (the end away from which the helicase moves) is captured by the pore, the helicase works with the direction of the field resulting from the applied potential and moves the threaded polynucleotide into the pore and into the *trans* chamber. Slipping forward involves the DNA moving forwards relative to the the pore (i.e. towards its 3' and away from it 5' end) at least 4 consecutive nucleotides and typically more than 10 consecutive nucleotides. Slipping forward may involve movement forward of 100 consecutive nucleotides or more and this may happen more than once in each strand.

The method of the invention preferably reduces the noise associated with the movement of the target polynucleotide with respect to, such as through, the transmembrane pore. Unwanted movement of the target polynucleotide in any dimension as a k-mer is being analysed typically results in noise in the current signature or level for the k-mer. The method of the invention may reduce this noise by reducing unwanted movement associated with one or more k-mers, such as each k-mer, in the target polynucleotide. The method of the invention may reduce the noise associated with the current level or signature for one or more k-mers, such as each k-mer, in the target polynucleotide.

In a preferred embodiment, the target polynucleotide is double stranded and the method reduces the noise associated with the movement of the complement strand to a greater degree than it reduces the noise associated with the movement of the template strand and/or the method increases the consistency of the movement of the complement strand to a greater degree than it increases the consistency of the movement of the template strand. This is advantageous for strand sequencing of double stranded target polynucleotides. The two stands of the double stranded polynucleotide are preferably linked by a bridging moiety, such as a hairpin loop or hairpin loop adaptor. This is discussed in more detail below.

### *Transmembrane pore*

A transmembrane pore is a structure that crosses the membrane to some degree. It permits hydrated ions driven by an applied potential to flow across or within the membrane. The transmembrane pore typically crosses the entire membrane so that hydrated ions may flow from one side of the membrane to the other side of the membrane. However, the transmembrane pore does not have to cross the membrane. It may be closed at one end. For instance, the pore may be a well, gap, channel, trench or slit in the membrane along which or into which hydrated ions may flow.

Any transmembrane pore may be used in the invention. The pore may be biological or artificial. Suitable pores include, but are not limited to, protein pores, polynucleotide pores and solid state pores. The pore may be a polynucleotide origami pore, such as a DNA origami pore (Langecker *et al.*, Science, 2012; 338: 932-936).

The transmembrane pore is preferably a transmembrane protein pore. A transmembrane protein pore is a polypeptide or a collection of polypeptides that permits hydrated ions, such as polynucleotides, to flow from one side of a membrane to the other side of the membrane. In the present invention, the transmembrane protein pore is capable of forming a pore that permits hydrated ions driven by an applied potential to flow from one side of the membrane to the other. The transmembrane protein pore preferably permits analytes such as nucleotides or polynucleotides to flow from one side of the membrane, such as a triblock copolymer membrane, to the other. The transmembrane protein pore typically allows a polynucleotide, such as DNA or RNA, to be moved through the pore.

The transmembrane protein pore may be a monomer or an oligomer. The pore is preferably made up of several repeating subunits, such as at least 6, at least 7, at least 8 or at least 9 subunits. The pore is preferably a hexameric, heptameric, octameric or nonameric pore. The pore may be a homo-oligomer or a hetero-oligomer.

The transmembrane protein pore typically comprises a barrel or channel through which the ions may flow. The subunits of the pore typically surround a central axis and contribute strands to a transmembrane β barrel or channel or a transmembrane α-helix bundle or channel.

The barrel or channel of the transmembrane protein pore typically comprises amino acids that facilitate interaction with analyte, such as nucleotides, polynucleotides or nucleic acids. These amino acids are preferably located near a constriction of the barrel or channel. The transmembrane protein pore typically comprises one or more positively charged amino acids, such as arginine, lysine or histidine, or aromatic amino acids, such as tyrosine or tryptophan. These amino acids typically facilitate the interaction between the pore and nucleotides, polynucleotides or nucleic acids.

Transmembrane protein pores for use in accordance with the invention can be derived from β-barrel pores or α-helix bundle pores. β-barrel pores comprise a barrel or channel that is formed from β-strands. Suitable β-barrel pores include, but are not limited to, β pore forming toxins, such as α-hemolysin, anthrax toxin and leukocidins, and outer membrane proteins/porins of bacteria, such as *Mycobacterium smegmatis* porin (Msp), for example MspA, MspB, MspC or MspD, outer membrane porin F (OmpF), outer membrane porin G (OmpG), outer membrane phospholipase A and *Neisseria* autotransporter lipoprotein (NalP), and other pores, such as lysenin. α-helix bundle pores comprise a barrel or channel that is formed from α-helices. Suitable α-helix bundle pores include, but are not limited to, inner membrane proteins and α outer membrane proteins, such as WZA and ClyA toxin. The transmembrane pore may be derived from lysenin. Suitable pores derived from lysenin are disclosed in International Application No. PCT/GB2013/050667 (published as WO 2013/153359). The transmembrane pore may be derived from Msp, such as MspA, or from α-hemolysin (α-HL).

The unmodified transmembrane pore used in the invention preferably comprises seven or more monomers comprising the sequence shown in SEQ ID NO: 2 or a variant thereof. The unmodified transmembrane pore more preferably comprises 8 or 9 monomers comprising the sequence shown in SEQ ID NO: 2 or a variant thereof. SEQ ID NO: 2 and variants thereof are discussed in more detail below. The pores modified in accordance with the invention may comprise any of the variants discussed below, especially the variants described with reference to the mutant Msp monomers of the invention.

In SEQ ID NO: 2 or a variant thereof, the part of the transmembrane pore which interacts with the polynucleotide binding protein typically comprises the amino acids at positions 12, 14, 48, 52, 53, 54, 55, 56, 57, 58, 59, 60, 134, 135, 136, 137, 138, 139, 169 and 170. These numbers correspond to the relevant positions in SEQ ID NO: 2 and may need to be altered in the case of variants where one or more amino acids have been inserted or deleted compared with SEQ ID NO: 2. A skilled person is capable of determining the corresponding position in a variant of SEQ ID NO: 2. For instance, position 4 in SEQ ID NO: 2 becomes position 6 in a variant having two amino acids added at the amino (N) terminus. If the variant is formed only by substitution of amino acids in SEQ ID NO: 2 (i.e. no amino acids are added to or deleted from SEQ ID NO: 2), the corresponding positions in the variant typically have the same numbering at the positions in SEQ ID NO: 2. The same is true for SEQ ID NOs: 4, 9, 24 and 36.

The part of the transmembrane pore which interacts with the polynucleotide binding protein preferably comprises the amino acids at positions:

(a) 12, 14, 52, 54, 56, 57, 59, 134, 136, 138, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in a variant thereof;

(b) 12, 14, 56, 57, 59, 134, 136, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in a variant thereof;

(c) 56, 57, 59, 134, 136, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in a variant thereof; or

(d) 56, 57, 59, 134 and 139 in SEQ ID NO: 2 or at the corresponding positions in a variant thereof.

The part of the transmembrane pore which interacts with the polynucleotide binding protein preferably comprises the amino acid at position 56 in SEQ ID NO: 2 or at the corresponding position in the variant thereof. The amino acid at position 56 (aspartic acid; D) may be replaced with asparagine (N), arginine (R), phenylalanine (F), tyrosine (Y) or leucine (L).

The part of the transmembrane pore which interacts with the polynucleotide binding protein preferably comprises the amino acid at position 59 in SEQ ID NO: 2 or at the corresponding position in a variant thereof. The amino acid at position 59 (glutamic acid; E) may be replaced with asparagine (N), arginine (R), phenylalanine (F), tyrosine (Y) or leucine (L).

The transmembrane protein pore may also be derived from α-hemolysin (α-HL). The wild type α-HL pore is formed of seven identical monomers or subunits (i.e. it is heptameric). The sequence of one monomer or subunit of α-hemolysin-NN is shown in SEQ ID NO: 4. The transmembrane protein pore preferably comprises seven monomers each comprising the sequence shown in SEQ ID NO: 4 or a variant thereof. Amino acids 1, 7 to 21, 31 to 34, 45 to 51, 63 to 66, 72, 92 to 97, 104 to 111, 124 to 136, 149 to 153, 160 to 164, 173 to 206, 210 to 213, 217, 218, 223 to 228, 236 to 242, 262 to 265, 272 to 274, 287 to 290 and 294 of SEQ ID NO: 4 form loop regions. Residues 113 and 147 of SEQ ID NO: 4 form part of a constriction of the barrel or channel of α-HL.

In such embodiments, a pore comprising seven proteins or monomers each comprising the sequence shown in SEQ ID NO: 4 or a variant thereof are preferably used in the method of the invention. The seven proteins may be the same (homo-heptamer) or different (hetero-heptamer).

In SEQ ID NO: 4, the part of the transmembrane pore which interacts with the polynucleotide binding protein typically comprises the amino acids at positions 16, 17, 18, 19, 21, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 281, 283, 285, 287, 288 and 293. These numbers correspond to the relevant positions in SEQ ID NO: 4 and may need to be altered in the case of variants where one or more amino acids have been inserted or deleted compared with SEQ ID NO: 4. A skilled person can determine the corresponding positions in a variant as discussed above.

The part of the transmembrane pore which interacts with the polynucleotide binding protein preferably comprises the amino acids at positions:

(a) 17, 18, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 287, 288 and 293 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof;

(b) 17, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242 and 287 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof; or

(c) 17, 19, 46, 93, 236, 237, 239, 240, 287 and 288 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof.

The amino acids at any of these positions may be replaced with phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) in accordance with the invention.

A variant of SEQ ID NO: 4 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 4 and which retains its pore forming ability. The ability of a variant to form a pore can be assayed using any method known in the art. For instance, the variant may be inserted into an amphiphilic layer, such as a triblock copolymer membrane, along with other appropriate subunits and its ability to oligomerise to form a pore may be determined. Methods are known in the art for inserting subunits into amphiphilic layers, such as triblock copolymer membranes. Suitable methods are discussed above.

The variant may include modifications that facilitate covalent attachment to or interaction with the polynucleotide binding protein. The variant preferably comprises one or more reactive cysteine residues that facilitate attachment to the polynucleotide binding protein. For instance, the variant may include a cysteine at one or more of positions 8, 9, 17, 18, 19, 44, 45, 50, 51, 237, 239 and 287 and/or on the amino or carboxy terminus of SEQ ID NO: 4. Preferred variants comprise a substitution of the residue at position 8, 9, 17, 237, 239 and 287 of SEQ ID NO: 4 with cysteine (A8C, T9C, N17C, K237C, S239C or E287C). The variant is preferably any one of the variants described in International Application No. PCT/GB09/001690 (published as WO 2010/004273), PCT/GB09/001679 (published as WO 2010/004265) or PCT/GB10/000133 (published as WO 2010/086603).

The variant may also include modifications that facilitate any interaction with nucleotides.

The variant may be a naturally occurring variant which is expressed naturally by an organism, for instance by a *Staphylococcus* bacterium. Alternatively, the variant may be expressed *in vitro* or recombinantly by a bacterium such as *Escherichia coli*. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the amino acid sequence of SEQ ID NO: 4, 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 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 can be determined as discussed below.

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 4 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions may be made as discussed above.

One or more amino acid residues of the amino acid sequence of SEQ ID NO: 4 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 be fragments of SEQ ID NO: 4. Such fragments retain pore-forming activity. Fragments may be at least 50, 100, 200 or 250 amino acids in length. A fragment preferably comprises the pore-forming domain of SEQ ID NO: 4. Fragments typically include residues 119, 121, 135. 113 and 139 of SEQ ID NO: 4.

One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the amino terminus or carboxy terminus of the amino acid sequence of SEQ ID NO: 4 or a 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 a pore or variant.

As discussed above, a variant of SEQ ID NO: 4 is a subunit that has an amino acid sequence which varies from that of SEQ ID NO: 4 and which retains its ability to form a pore. A variant typically contains the regions of SEQ ID NO: 4 that are responsible for pore formation. The pore forming ability of α-HL, which contains a β-barrel, is provided by β-strands in each subunit. A variant of SEQ ID NO: 4 typically comprises the regions in SEQ ID NO: 4 that form β-strands. The amino acids of SEQ ID NO: 4 that form β-strands are discussed above. One or more modifications can be made to the regions of SEQ ID NO: 4 that form β-strands as long as the resulting variant retains its ability to form a pore. Specific modifications that can be made to the β-strand regions of SEQ ID NO: 4 are discussed above.

A variant of SEQ ID NO: 4 preferably includes one or more modifications, such as substitutions, additions or deletions, within its α-helices and/or loop regions. Amino acids that form α-helices and loops are discussed above.

The variant of SEQ ID NO: 4 may be modified to assist its identification or purification as discussed below.

The transmembrane protein pore may also be derived from lysenin. The unmodified transmembrane pore used in the invention preferably comprises at least one monomer comprising the sequence shown in SEQ ID NO: 36 or a variant thereof.

In SEQ ID NO: 36 or a variant thereof, the part of the transmembrane pore which interacts with the polynucleotide binding protein typically comprises the amino acids at positions (i) 31 (serine; S), (ii) 33 (serine; S), (iii) 108 (proline; P), (iv) 109 (proline; P), (v) 110 (threonine) and (vi) 138 (proline; P). These numbers correspond to the relevant positions in SEQ ID NO: 36 and may need to be altered in the case of variants where one or more amino acids have been inserted or deleted compared with SEQ ID NO: 36. A skilled person can determine the correspondin positions in a variant as discussed above.

The part of the transmembrane pore which interacts with the polynucleotide binding protein preferably comprises any number and combination of these amino acids in SEQ ID NO: 36 or the variant thereof. The part of the transmembrane pore which interacts with the polynucleotide binding protein preferably comprises the amino acids at positions {i}, {ii}, {iii}, {iv}, {v}, {vi}, {i,ii}, {i,iii}, {i,iv}, {i,v}, {i,vi}, {ii,iii}, {ii,iv}, {ii,v}, {ii,vi}, {iii,iv}, {iii,v}, {iii,vi}, {iv,v}, {iv,vi}, {v,vi}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,ii,vi}, {i,iii,iv}, {i,iii,v}, {i,iii,vi}, {i,iv,v}, {i,iv,vi}, {i,v,vi}, {ii,iii,iv}, {ii,iii,v}, {ii,iii,vi}, {ii,iv,v}, {ii,iv,vi}, {ii,v,vi}, {iii,iv,v}, {iii,iv,vi}, {iii,v,vi}, {iv,v,vi}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iii,vi}, {i,ii,iv,v}, {i,ii,iv,vi}, {i,ii,v,vi}, {i,iii,iv,v}, {i,iii,iv,vi}, {i,iii,v,vi}, {i,iv,v,vi}, {ii,iii,iv,v}, {ii,iii,iv,vi}, {ii,iii,v,vi}, {ii,iv,v,vi}, {iii,iv,v,vi}, {i,ii,iii,iv,v}, {i,ii,iii,iv,vi}, {i,ii,iii,v,vi}, {i,ii,iv,v,vi}, {i,iii,iv,v,vi}, {ii,iii,iv,v,vi} or {i,ii,iii,iv,v,vi} in SEQ ID NO: 36 or at the corresponding positions in the variant thereof.

Any number and combination of (i) to (vi) as set out above may be replaced with phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) in accordance with the invention.

A variant of SEQ ID NO: 36 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 36 and which retains its pore forming ability. The ability of a variant to form a pore can be assayed using any method known in the art as discussed above.

Over the entire length of the amino acid sequence of SEQ ID NO: 36, 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: 36 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 can be determined as discussed below.

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 36 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions may be made as discussed above.

One or more amino acid residues of the amino acid sequence of SEQ ID NO: 36 may additionally be deleted from the variants 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: 36. Such fragments retain pore forming activity. This may be assayed as described above. Fragments may be at least 50, 100, 150, 200 or 250 amino acids in length. Such fragments may be used to produce the pores of the invention. Since the region of from about position 44 to about position 126 of SEQ ID NO: 36 can be modified by one or more deletions in accordance with the invention, a fragment does not have to contain the entire region. Hence, fragments shorter than the length of the unmodified region are envisaged by the invention. A fragment preferably comprises the pore forming domain of SEQ ID NO: 36. A fragment more preferably comprises the region from about position 44 to about position 126 of SEQ ID NO: 36.

One or more amino acids may be alternatively or additionally added to the variants described above. An extension may be provided at the amino terminal or carboxy terminal of the amino acid sequence of the variant of SEQ ID NO: 36, including a 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.

As discussed above, a variant is a polypeptide that has an amino acid sequence which varies from that of SEQ ID NO: 36 and which retains its ability to form a pore. A variant typically contains the region of SEQ ID NO: 36 that is responsible for pore formation, namely from about position 44 to about position 126. It may contain a fragment of this region as discussed above.

The variant of SEQ ID NO: 36 may be any of those disclosed in International Application No. PCT/GB2013/050667 (published as WO 2013/153359).

In some embodiments, the transmembrane protein pore is chemically modified. The pore can be chemically modified in any way and at any site. Suitable modifications are discussed below with reference to the mutant Msp monomers. Such modifications can be applied to any of the pores used in the invention.

*Polynucleotide binding protein*

Suitable polynucleotide binding proteins are discussed below. The unmodified polynucleotide binding protein used in the invention preferably comprises the sequence shown in SEQ ID NO: 24 or a variant thereof. Suitable variants of SEQ ID NO: 24 are discussed below.

In SEQ ID NO: 24 or a variant thereof, the part of the polynucleotide binding protein which interacts with the transmembrane pore typically comprises the amino acids at positions 1, 2, 3, 4, 5, 6, 51, 176, 177, 178, 179, 180, 181, 185, 189, 191, 193, 194, 195, 197, 198, 199, 200, 201, 202, 203, 204, 207, 208, 209, 210, 211, 212, 213, 216, 219, 220, 221, 223, 224, 226, 227, 228, 229, 247, 254, 255, 256, 257, 258, 259, 260, 261, 298, 300, 304, 308, 318, 319, 321, 337, 347, 350, 351, 405, 415, 422, 434, 437, 438. These numbers correspond to the relevant positions in SEQ ID NO: 24 and may need to be altered in the case of variants where one or more amino acids have been inserted or deleted compared with SEQ ID NO: 24. A skilled person can determine the correspondin positions in a variant as discussed above. The part of the polynucleotide binding protein which interacts with the transmembrane pore preferably comprises the amino acids at

(a) positions 1, 2, 4, 51, 177, 178, 179, 180, 185, 193, 195, 197, 198, 199, 200, 202, 203, 204, 207, 208, 209, 210, 211, 212, 216, 221, 223, 224, 226, 227, 228, 229, 254, 255, 256, 257, 258, 260, 304, 318, 321, 347, 350, 351, 405, 415, 422, 434, 437 and 438 in SEQ ID NO: 24 or at the corresponding positions in the variant thereof; or

(b) positions 1, 2, 178, 179, 180, 185, 195, 197, 198, 199, 200, 202, 203, 207, 209, 210, 212, 216, 221, 223, 226, 227, 255, 258, 260, 304, 350 and 438 in SEQ ID NO: 24 or at the corresponding positions in the variant thereof.

The part of the polynucleotide binding protein which interacts with the transmembrane pore preferably comprises one or more of, such as 2, 3 or 4 of, the amino acids at positions 195, 198, 199 and 258 in SEQ ID NO: 24 or the variant thereof. The part of the polynucleotide binding protein which interacts with the transmembrane pore preferably comprises the amino acid at position 195, 198, 199 or 258 in SEQ ID NO: 24 or at the corresponding positions in the variant thereof. The modified polynucleotide binding protein of the invention preferably comprises a variant of SEQ ID NO: 24 which comprises one or more of the following modifications (a) W195A, (b) D198V, (c) K199L or (d) E258L. The variant may comprise {a};

{b}; {c}; {d}; {a,b}; {a,c}; {a,d}; {b,c}; {b,d}; {c,d}; {a,b,c}; {a,b,d}; {a,c,d}; {b,c,d}; or {a,b,c,d}. The variant of SEQ ID NO: 24 may further comprise any of the additional modifications discussed below. The modifications set out in this paragraph are preferred when the modified polynucleotide binding protein interacts with a pore derived from MspA, particularly any of the modified pores of the invention.

The part of the polynucleotide binding protein which interacts with the transmembrane pore preferably comprises the amino acid at position 199 of SEQ ID NO: 24 or at the corresponding position in the variant thereof. The modified polynucleotide binding protein of the invention preferably comprises a variant of SEQ ID NO: 24 which comprises K199A, K199V, K199F, K199D, K199S, K199W or K199L.

The unmodified polynucleotide binding protein used in the invention preferably comprises the sequence shown in SEQ ID NO: 9 or a variant thereof. A variant of SEQ ID NO: 9 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 9 and which retains its ability to bind a polynucleotide. The ability of a variant to bind a polynucleotide can be assayed using any method known in the art.

Over the entire length of the amino acid sequence of SEQ ID NO: 9, 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: 9 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 can be determined as discussed below.

The variant may be modified in any of the ways discussed above with reference to other variants. Preferred variants of SEQ ID NO: 9 lack amino acids 1 to 4.

In SEQ ID NO: 9 or variant thereof, the part of the polynucleotide binding protein which interacts with the transmembrane pore typically comprises the amino acids at positions 80, 81, 82, 84, 85, 205, 206, 209, 215, 216, 220, 221, 224, 236, 240, 241, 267, 270, 272, 278, 287, 289, 293, 296, 307, 308, 309, 310, 320, 321, 322, 323, 327, 349, 415, 418 and 419. These numbers correspond to the relevant positions in SEQ ID NO: 9 and may need to be altered in the case of variants where one or more amino acids have been inserted or deleted. A skilled person can determine the correspondin positions in a variant as discussed above.

The part of the polynucleotide binding protein which interacts with the transmembrane pore preferably comprises the amino acids at

(a) positions 80, 84, 205, 209, 215, 216, 221, 224, 236, 241, 267, 272, 289, 296, 307, 308, 309, 320, 321, 322, and 419 in SEQ ID NO: 9 or at the corresponding positions in a variant thereof;

(b) positions 80, 84, 209, 215, 216, 221, 267, 272, 289, 307, 308, 309, 321 and 322 in SEQ ID NO: 9 or at the corresponding positions in a variant thereof; or

(b) positions 215, 267, 272, 307, 308 and 322 in SEQ ID NO: 9 or at the corresponding positions in a variant thereof.

Any of these positions may be replaced with phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) in accordance with the invention.

*Parts which interact*

The part of the transmembrane pore which may be modified in accordance with the invention is typically the part of the transmembrane pore which interacts with or contacts the polynucleotide binding protein when the protein controls the movement of the target polynucleotide with respect to the pore. The part may comprise one or more amino acids which interact with or contact one or more amino acids in the polynucleotide binding protein when the protein controls the movement of the target polynucleotide with respect to the pore. Specific amino acids are discussed in more detail below.

The part of the polynucleotide binding protein which may be modified in accordance with the invention is typically the part of the protein which interacts with or contacts the transmembrane pore when the protein controls the movement of the target polynucleotide with respect to the pore. The part may comprise one or more amino acids which interact with or contact one or more amino acids in the transmembrane pore when the protein controls the movement of the target polynucleotide with respect to the pore. Specific amino acids are discussed in more detail below.

The part of the transmembrane pore which interacts with the polynucleotide binding protein and/or the part of the polynucleotide binding protein which interacts with the transmembrane pore can be identified using any method known in the art. The part(s) may be identified using protein modelling, x-ray diffraction measurement of the protein in a crystalline state (Rupp B (2009). Biomolecular Crystallography: Principles, Practice and Application to

Structural Biology. New York: Garland Science.), nuclear magnetic resonance (NMR) spectroscopy of the protein in solution (Mark Rance; Cavanagh, John; Wayne J. Fairbrother; Arthur W. Hunt III; Skelton, Nicholas J. (2007). Protein NMR spectroscopy: principles and practice (2nd ed.). Boston: Academic Press.) or cryo-electron microscopy of the protein in a frozen-hydrated state (van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A (2000). "Single-particle electron cryo-microscopy: towards atomic resolution." Q Rev Biophys. 33: 307–69. Structural information of proteins determined by above mentioned methods are publicly available from the protein bank (PDB) database.

Protein modelling exploits the fact that protein structures are more conserved than protein sequences amongst homologues. Hence, producing atomic resolution models of proteins is dependent upon the identification of one or more protein structures that are likely to resemble the structure of the query sequence. In order to assess whether a suitable protein structure exists to use as a “template” to build a protein model, a search is performed on the protein data bank (PDB) database. A protein structure is considered a suitable template if it shares a reasonable level of sequence identity with the query sequence. If such a template exists, then the template sequence is “aligned” with the query sequence, i.e. residues in the query sequence are mapped onto the template residues. The sequence alignment and template structure are then used to produce a structural model of the query sequence. Hence, the quality of a protein model is dependent upon the quality of the sequence alignment and the template structure.

The part(s) may also be identified using molecular simulations or energy minimisations (Kalli AC, Campbell ID, Sansom MSP, (2013) “Conformational Changes in Talin on Binding to Anionic Phopholipid Membranes Facilitate Signaling by Integrin Transmembrane Helices” PLOS Computational Biology, 9, 10, e1003316 and Durrieu M, Lavery R, Baaden M, (2008) “Interactions between Neuronal Fusion Proteins Explored by Molecular Dynamics”, Biophysical Journal, 94, 3436-3446).

*Surface*

The method preferably comprises making one or modifications to the surface of the transmembrane pore which interacts with the polynucleotide binding protein and/or to the surface of the polynucleotide binding protein which interacts with the transmembrane pore. Any number of modifications can be made, such as 2 or more, 3 or more, 5 or more, 10 or more, 15 or more, 20 or more, 25 or more, 30 or more, 50 or more or 100 or more modifications.

The surface of the transmembrane pore and/or the polynucleotide binding protein may be identified using any method in the art, including any of the methods disclosed above. For instance, the surface may be identified using protein modeling, molecular simulations or energy minimisations. Protein modelling is a general term to describe e.g. making models of proteins of unknown structure, simulating proteins to investigate dynamics, molecular docking.

### *Pore entrance*

The method comprises making one or modifications to the entrance of the transmembrane pore which interacts with the polynucleotide binding protein. Any number of modifications may be made as discussed above. The entrance of the pore may be identified as discussed above.

### *Modifications*

Any modifications may be made in accordance with the invention. The method may involve making one or more modifications which (a) alter the charge, (b) alter the sterics, (c) alter the hydrogen bonding, (d) alter the π stacking or (e) alter the structure of the part of the transmembrane pore which interacts with the polynucleotide binding protein and/or the part of the polynucleotide binding protein which interacts with the transmembrane pore. Any number and combination of these may be altered. For instance, the method may involve making one or more modification which {a}; {b}; {c}; {d}; {e}; {a,b}; {a,c}; {a,d}; {a,e}; {b,c}; {b,d}; {b,e}; {c,d}; {c,e}; {d,e}; {a,b,c}; {a,b,d}; {a,b,e}; {a,c,d}; {a,c,e}; {a,d,e}; {b,c,d}; {b,c,e}; {b,d,e}; {c,d,e}; {a,b,c,d}; {a,b,c,e}; {a,b,d,e}; {a,c,d,e}; {b,c,d,e}; or {a,b,c,d,e}.

When modifying a protein, the one or more modifications typically involve introducing or replacing one or more amino acids. The invention typically involves making one or more amino acid substitutions.

Modifications which alter the charge may involve increasing the net negative charge or decreasing the net negative charge. The method preferably comprises making one or more modifications which decrease the net negative charge of the part of the transmembrane pore which interacts with the polynucleotide binding protein. Modifications which decrease net negative charge are discussed in more detail below with reference to mutant Msp monomers. Any of these modifications may be made in other transmembrane pores and/or polynucleotide binding proteins. In a preferred embodiment, the transmembrane pore comprises seven or more monomers, such as 8 or 9 monomers, comprising SEQ ID NO: 2 of a variant thereof and the method preferably comprises modifying one or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, so they do not comprise aspartic acid (D) or glutamic acid (E) at one or more of positions 56, 57, 59, 134 and 139 in SEQ ID NO: 2 or at one or more of the corresponding positions in the variant thereof. The method more preferably comprises modifying one or more of the monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, so they comprise one or more of (a) D56N, D56R, D56F, D56Y, D56L D56K, and/or (b) E57N, E57R, E57F, E57Y, E57L or E57K, and/or (c) E59N, E59R, E59F, E59Y, E59L, E59K, and/or (d) D134N, D134R, D134F, D134Y, D134L, D134K, and/or (e) E139N, E139R, E139F, E139Y, E139L or E139K. One or more of the monomers may comprise any number and combination of these modifications. For instance, one or more of the monomers may comprise {a}; {b}; {c}; {d}; {e}; {a,b}; {a,c}; {a,d}; {a,e}; {b,c}; {b,d}; {b,e}; {c,d}; {c,e}; {d,e}; {a,b,c}; {a,b,d}; {a,b,e}; {a,c,d}; {a,c,e}; {a,d,e}; {b,c,d}; {b,c,e}; {b,d,e}; {c,d,e}; {a,b,c,d}; {a,b,c,e}; {a,b,d,e}; {a,c,d,e}; {b,c,d,e}; or {a,b,c,d,e}. One or more of the monomers may comprise D56N and E59R, D56F and E59R, D56N and E59F, D56N and E59Y or D56L and E59L. One or more of the monomers may comprise D56N and E59R, D56F and E59R or D56N and E59F.

The modified transmembrane pore (i.e. the transmembrane pore resulting from the modification method of the invention) preferably does not comprise one or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 monomers, which are variants of SEQ ID NO: 2 comprising or consisting of E59R, D90N, D91N, D93N, D118R, D134R and E139K.

Modifications which alter the sterics may involve increasing or decreasing the size of amino acid residues, for instance by substitution. For instance, sterics can be increased by the introduction of one or more bulky amino acids, such as phenylalanine (F), tryptophan (W), tyrosine (Y) and histidine (H).

Modifications which alter the hydrogen bonding may involve the introduction or replacement of one or more amino acids which can hydrogen bond.

Modifications which alter the π stacking may involve the introduction or replacement of amino acids that interact through delocalised electron π systems. For instance, π stacking can be increased by the introduction of one or more aromatic amino acids, such as phenylalanine (F), tryptophan (W), tyrosine (Y) and histidine (H).

In a preferred embodiment, the transmembrane pore comprises one or more monomers comprising SEQ ID NO: 36 of a variant thereof and the method preferably comprises modifying one or more of the monomers so they comprise phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions (i) 31, (ii) 33, (iii) 108, (iv) 109, (v) 110 and (vi) 138 in SEQ ID NO: 36 or at one or more of the corresponding positions in the variant thereof, such as {i}, {ii}, {iii}, {iv}, {v}, {vi}, {i,ii}, {i,iii}, {i,iv}, {i,v}, {i,vi}, {ii,iii}, {ii,iv}, {ii,v}, {ii,vi}, {iii,iv}, {iii,v}, {iii,vi}, {iv,v}, {iv,vi}, {v,vi}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,ii,vi}, {i,iii,iv}, {i,iii,v}, {i,iii,vi}, {i,iv,v}, {i,iv,vi}, {i,v,vi}, {ii,iii,iv}, {ii,iii,v}, {ii,iii,vi}, {ii,iv,v}, {ii,iv,vi}, {ii,v,vi}, {iii,iv,v}, {iii,iv,vi}, {iii,v,vi}, {iv,v,vi}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iii,vi}, {i,ii,iv,v}, {i,ii,iv,vi}, {i,ii,v,vi}, {i,iii,iv,v}, {i,iii,iv,vi}, {i,iii,v,vi}, {i,iv,v,vi}, {ii,iii,iv,v}, {ii,iii,iv,vi}, {ii,iii,v,vi}, {ii,iv,v,vi}, {iii,iv,v,vi}, {i,ii,iii,iv,v}, {i,ii,iii,iv,vi}, {i,ii,iii,v,vi}, {i,ii,iv,v,vi}, {i,iii,iv,v,vi}, {ii,iii,iv,v,vi} or {i,ii,iii,iv,v,vi}.

In a preferred embodiment, the transmembrane pore comprises seven monomers comprising SEQ ID NO: 4 of a variant thereof and the method preferably comprises modifying one or more of the monomers, such as 1, 2, 3, 4, 5, 6 or 7 of the monomers, so they comprise phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions 16, 17, 18, 19, 21, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 281, 283, 285, 287, 288 and 293 in SEQ ID NO: 4 or at one or more of the corresponding positions in the variant thereof.

In a more preferred embodiment, the transmembrane pore comprises seven monomers comprising SEQ ID NO: 4 of a variant thereof and the method preferably comprises modifying one or more of the monomers, such as 1, 2, 3, 4, 5, 6 or 7 of the monomers, so they comprise phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions

(a) 17, 18, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 287, 288 and 293 in SEQ ID NO: 4 or at one or more of the corresponding positions in the variant thereof;

(b) 17, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242 and 287 in SEQ ID NO: 4 or at one or more of the corresponding positions in the variant thereof; or

(c) 17, 19, 46, 93, 236, 237, 239, 240, 287 and 288 in SEQ ID NO: 4 or at one or more of the corresponding positions in the variant thereof.

## Modified pores

The method also provides a transmembrane pore modified in accordance with the invention. A part of the transmembrane pore which interacts with a polynucleotide binding protein has been modified. The part of the transmembrane pore which interacts with a polynucleotide binding protein when the polynucleotide binding protein is used to control the movement of a target polynucleotide with respect to, or through, the pore has been modified. Any of the modifications discussed above may be made in the pores of the invention.

The transmembrane pore preferably comprises seven or more monomers, such as 8 or 9 monomers, comprising the sequence shown in SEQ ID NO: 2 or a variant thereof. One or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, preferably comprises a variant of SEQ ID NO: 2 which comprises one or more of (a) D56N, D56R, D56F, D56Y or D56L, (b) E57N or E57R, (c) E59N, E59R, E59F, E59Y or E59L, (d) D134N or D134R and (e) E139N, E139R or E139K. Any number and combination of these modifications may be made in a single monomer as discussed above.

One or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, preferably comprises a variant of SEQ ID NO: 2 which comprises D56N, D56R, D56F, D56Y or D56L.

One or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, preferably comprises a variant of SEQ ID NO: 2 which comprises E59N, E59R, E59F, E59Y or E59L.

One or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, preferably comprises a variant of SEQ ID NO: 2 which comprises D56N and E59R, D56F and E59R, D56N and E59F, D56N and E59Y or D56L and E59L.

One or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, preferably comprises a variant of SEQ ID NO: 2 which comprises D56N and E59R, D56F and E59R or D56N and E59F.

One or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, are not variants of SEQ ID NO: 2 which comprise or consist of E59R, D90N, D91N, D93N, D118R, D134R and E139K.

The transmembrane pore preferably comprises one or more monomers comprising a variant of SEQ ID NO: 36 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions (i) 31, (ii) 33, (iii) 108, (iv) 109, (v) 110 and (vi) 138 or at one or more the corresponding positions thereof, such as {i}, {ii}, {iii}, {iv}, {v}, {vi}, {i,ii}, {i,iii}, {i,iv}, {i,v}, {i,vi}, {ii,iii}, {ii,iv}, {ii,v}, {ii,vi}, {iii,iv}, {iii,v}, {iii,vi}, {iv,v}, {iv,vi}, {v,vi}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,ii,vi}, {i,iii,iv}, {i,iii,v}, {i,iii,vi}, {i,iv,v}, {i,iv,vi}, {i,v,vi}, {ii,iii,iv}, {ii,iii,v}, {ii,iii,vi}, {ii,iv,v}, {ii,iv,vi}, {ii,v,vi}, {iii,iv,v}, {iii,iv,vi}, {iii,v,vi}, {iv,v,vi}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iii,vi}, {i,ii,iv,v}, {i,ii,iv,vi}, {i,ii,v,vi}, {i,iii,iv,v}, {i,iii,iv,vi}, {i,iii,v,vi}, {i,iv,v,vi}, {ii,iii,iv,v}, {ii,iii,iv,vi}, {ii,iii,v,vi}, {ii,iv,v,vi}, {iii,iv,v,vi}, {i,ii,iii,iv,v}, {i,ii,iii,iv,vi}, {i,ii,iii,v,vi}, {i,ii,iv,v,vi}, {i,iii,iv,v,vi}, {ii,iii,iv,v,vi} or {i,ii,iii,iv,v,vi}.

The transmembrane pore preferably comprises seven monomers comprising SEQ ID NO: 4 or a variant thereof in which one or more of the monomers, such as 1, 2, 3, 4, 5, 6 or 7 of the monomers, is a variant of SEQ ID NO: 4 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions 16, 17, 18, 19, 21, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 281, 283, 285, 287, 288 and 293 or at one or more of the corresponding positions thereof, such as one or more of positions:

(a) 17, 18, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 287, 288 and 293 or at one or more of the corresponding positions thereof;

(b) 17, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242 and 287 in SEQ ID NO: 4 or at one or more of the corresponding positions thereof; or

(c) 17, 19, 46, 93, 236, 237, 239, 240, 287 and 288 in SEQ ID NO: 4 or at one or more of the corresponding positions thereof.

## Mutant Msp monomers

The present invention also 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 polynucleotide reading properties i.e. display improved polynucleotide capture and nucleotide discrimination. In particular, pores constructed from the mutant monomers capture nucleotides and polynucleotides 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 polynucleotide moves with respect to, or through, pores constructed from the mutants is decreased. This makes it easier to identify a direct relationship between the observed current as the polynucleotide moves with respect to, or through, the pore and the polynucleotide sequence.

A mutant monomer of the invention comprises a variant of the sequence shown in SEQ ID NO: 2. SEQ ID NO: 2 is the wild-type MspA monomer. 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 ability of a variant to form a pore can be assayed using any method known in the art. For instance, the variant may be inserted into an amphiphilic layer along with other appropriate subunits and its ability to oligomerise to form a pore may be determined. Methods are known in the art for inserting subunits into membranes, such as amphiphilic layers. For example, subunits may be suspended in a purified form in a solution containing a triblock copolymer membrane such that it diffuses to the membrane and is inserted by binding to the membrane and assembling into a functional state.

The mutant Msp monomer is modified in accordance with the invention. A part of the monomer which interacts with a polynucleotide binding protein has been modified. The part of the monomer which interacts with a polynucleotide binding protein when the polynucleotide binding protein is used to control the movement of a target polynucleotide with respect to, or through, a pore comprising the monomer has been modified. Any of the modifications discussed above may be made in the monomer of the invention.

The monomer preferably comprises a variant of SEQ ID NO: 2 which comprises one or more of (a) D56N, D56R, D56F, D56Y or D56L, (b) E57N or E57R, (c) E59N, E59R, E59F, E59Y or E59L, (d) D134N or D134R and (e) E139N, E139R or E139K. Any number and combination of these modifications may be made in the monomer as discussed above. The variant preferably comprises D56N and E59R, D56F and E59R, D56N and E59F, D56N and E59Y or D56L and E59L. The variant preferably comprises D56N and E59R, D56F and E59R or D56N and E59F. The monomer preferably does not comprise a variant of SEQ ID NO: 2 which comprises or consists of E59R, D90N, D91N, D93N, D118R, D134R and E139K.

*Barrel deletions*

In the variant (of SEQ ID NO: 2), 2, 4, 6, 8 or 10 of the amino acids at positions 72 to 82 of SEQ ID NO: 2 may have been deleted. 2, 4, 6, 8 or 10 of the amino acids at positions 111 to 121 of SEQ ID NO: 2 may have also been deleted. In other words, 2, 4, 6, 8 or 10 amino acids may be deleted from the downward strand (positions 72 to 82) and the upward strand (positions 111 to 121) of the barrel region of SEQ ID NO: 2. These deletions and their advantages are discussed in more detail in UK Application No. 1417708.3 co-filed with this application.

The number of amino acids deleted from positions 72 to 82 may be different from the number of amino acids deleted from positions 111 to 121. The number of amino acids deleted from positions 72 to 82 is preferably the same as the number of amino acids deleted from positions 111 to 121.

Any combination of amino acids from positions 72 to 82 and amino acids from positions 111 to 121 may be deleted. The majority of the amino acids in the downward and upwards strands of the barrel in SEQ ID NO: 2 alternate between hydrophobic and hydrophilic. The hydrophobic amino acids are selected from tryptophan (W), leucine (L), valine (V), isoleucine (I), phenylalanine (F) and tyrosine (Y). The hydrophilic amino acids are selected from serine (S), glycine (G), asparagine (N), proline (P) and aspartic acid (D).

Positions 72 to 82 of SEQ ID NO: 2 correspond to W-S-L-G-V-G-I-N-F-S-Y (SEQ ID NO: 26 with the hydrophobic amino acids underlined). Positions 111 to 121 of SEQ ID NO: 2 correspond to P-G-V-S-I-S-A-D-L-G-N (SEQ ID NO: 27 with the hydrophobic amino acids underlined). This alternation between hydrophobic and hydrophilic amino acids results in the beta-sheet which forms part of the barrel of the pore.

The amino acids from positions 72 to 82 remaining after deletion (*i.e.* after 2, 4, 6, 8 or 10 amino acids have been deleted from positions 72 to 82) preferably comprise 3, 5, 7 or 9 consecutive amino acids which alternate between hydrophobic and hydrophilic.

The amino acids from positions 111 to 121 remaining after deletion (*i.e.* after 2, 4, 6, 8 or 10 amino acids have been deleted from positions 111 to 121) preferably comprise 3, 5, 7 or 9 consecutive amino acids which alternate between hydrophobic and hydrophilic.

The amino acids deleted from positions 72 to 82 may correspond to the amino acids deleted from positions 111 to 121 as shown in Table 1 below. For instance, if L74 and G75 are deleted from positions 72 to 82, D118 and L119 may be deleted from positions 111 to 121.

**Table 1 – Corresponding amino acids in the barrel of SEQ ID NO: 2**

| Position in (a) | Corresponding position in (b) |
|---|---|
| W72 | N121 |
| S73 | G120 |
| L74 | L119 |
| G75 | D118 |
| V76 | A117 |

| | |
|---|---|
| G77 | S116 |
| I78 | I115 |
| N79 | S114 |
| F80 | V113 |
| S81 | G112 |
| Y82 | P111 |

One or more positions of the amino acids that have been deleted from positions 72 to 82 may not correspond to the one or more positions of the amino acids that have been deleted from positions 111 to 121 as shown in Table 1. For instance, if L74 and G75 are deleted from positions 72 to 82, A117 and D118 may be deleted from positions 111 to 121.

The positions of (all of) the amino acids that have been deleted from positions 72 to 82 may not correspond to the positions of (all of) the amino acids that have been deleted from positions 111 to 121 as shown in Table 1. For instance, if L74 and G75 are deleted from positions 72 to 82, I115 and S116 may be deleted from positions 111 to 121.

The amino acids deleted from positions 72 to 82 are preferably consecutive. The amino acids deleted from positions 111 to 121 are preferably consecutive. The amino acids deleted from positions 72 to 82 and the amino acids deleted from positions 111 to 121 are preferably consecutive.

The invention preferably provides mutant monomers comprising a variant of the sequence shown in SEQ ID NO: 2, wherein in the variant (i) L74, G75, D118 and L119 have been deleted, (ii) G75, V76, A117 and D118 have been deleted, (iii) V76, G77, S116 and A117 have been deleted, (iv) G77, I78, I115 and S116 have been deleted, (v) I78, N79, S114 and I115 have been deleted, (vi) N79, F80, V113 and S114 have been deleted or (vii) F80, S81, G112 and V113 have been deleted. The invention preferably comprises a variant of the sequence shown in SEQ ID NO: 2, wherein in the variant L74, G75, V76, G77, S116, A117, D118 and L119 have been deleted. The invention preferably comprises a variant of the sequence shown in SEQ ID NO: 2, wherein in the variant L74, G75, N79, F80, V113, S114, D118 and L119 or L74, G75, F80, S81, G112, V113, D118 and L119.

The skilled person can identify other combinations of amino acids that may be deleted in accordance with the invention. The following discussion using the numbering of residues in SEQ ID NO: 2 (i.e. before any amino acids have been deleted as defined above).

*Positions 90 and 91*

In wild-type MspA, amino acids 90 and 91 are both aspartic acid (D). These amino acids in each monomer form part of an inner constriction of the pore. The variant preferably does not comprise aspartic acid (D) at position 90. The variant preferably does not comprise aspartic acid (D) or glutamic acid (E) at position 90. The variant preferably does not have a negatively charged amino acid at position 90.

The variant preferably does not comprise aspartic acid (D) at position 91. The variant preferably does not comprise aspartic acid (D) or glutamic acid (E) at position 91. The variant preferably does not have a negatively charged amino acid at position 91.

The variant preferably comprises 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 90 and/or position 91. Any combinations of these amino acids at positions 90 and 91 are envisaged by the invention. The variant preferably comprises asparagine (N) at position 90 and/or position 91. The variant more preferably comprises asparagine (N) at position 90 and position 91. These amino acids are preferably inserted at position 90 and/or 91 by substitution.

*Position 93*

In wild-type MspA, amino acid 93 is aspartic acid (D). This amino acid in each monomer also forms part of an inner constriction of the pore.

The variant preferably comprises aspartic acid (D) or glutamic acid (E) at position 93. The variant preferably has a negative charge at position 93. The glutamic acid (E) is preferably introduced by substitution.

*Cap forming region*

In wild-type MspA, amino acids 1 to 72 and 122 to 184 form the cap of the pore. Of these amino acids, V9, Q12, D13, R14, T15, W40, I49, P53, G54, D56, E57, E59, T61, E63, Y66, Q67, I68, F70, P123, I125, Q126, E127, V128, A129, T130, F131, S132, V133, D134, S136, G137, E139, V144, H148, T150, V151, T152, F163, R165, I167, S169, T170 and S173 face inwards into the channel of the pore.

*Barrel forming region*

In wild-type MspA, amino acids 72 to 82 and 112 to 121 form the barrel of the pore. Of these amino acids, S73, G75, G77, N79, S81, G112, S114, S116, D118 and G120 face inwards into the channel of the pore. S73, G75, G77, N79, S81 face inwards in the downwards strand and G112, S114, S116, D118 and G120 face inwards in the upwards strand.

*Decreased net negative charge*

The variant preferably comprises one or more modifications which decrease the net negative charge of the inward facing amino acids in the cap forming region and/or the barrel forming region of the monomer. The variant preferably comprises two or more modifications which decrease the net negative charge of the inward facing amino acids in the cap forming region and the barrel forming region of the monomer. Any such modifications to the barrel forming region are in addition to the deletions of the invention discussed above.

The variant may comprise any number of modifications, such as 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 15 or more, 20 or more, 30 or more, or 40 or more modifications.

The net negative charge may be decreased by any means known in the art. The net negative charge is decreased in a manner that does not interfere with the ability of the mutant monomer to form a pore. This can be measured as discussed above.

The net negative charge of the inward facing amino acids in the cap forming region and/or the barrel forming region may be decreased. This means that the inward facing amino acids in the cap forming region and/or the barrel forming region comprise fewer negatively charged amino acids than in SEQ ID NO: 2 and/or comprises more positively charged amino acids than in SEQ ID NO: 2. The one or more modifications may lead to a net positive charge in the inward facing amino acids in the cap forming region and/or the barrel forming region

The net charge can be measured using methods known in the art. For instance, the isolectric point may be used to define the net charge of the inward facing amino acids in the cap forming region and/or the barrel forming region.

The one or more modifications are preferably one or more deletions of negatively charged amino acids. Removal of one or more negatively charged amino acids reduces the net negative charge of the inward facing amino acids in the cap forming region and/or barrel forming region. A negatively charged amino acid is an amino acid with a net negative charge. Negatively charged amino acids include, but are not limited to, aspartic acid (D) and glutamic acid (E). Methods for deleting amino acids from proteins, such as MspA monomers, are well known in the art.

The one or more modifications are preferably one or more substitutions of negatively charged amino acids with one or more positively charged, uncharged, non-polar and/or aromatic amino acids. A positively charged amino acid is an amino acid with a net positive charge. The positively charged amino acid(s) can be naturally-occurring or non-naturally-occurring. The positively charged amino acid(s) may be synthetic or modified. For instance, modified amino acids with a net positive charge may be specifically designed for use in the invention. A number of different types of modification to amino acids are well known in the art.

Preferred naturally-occurring positively charged amino acids include, but are not limited to, histidine (H), lysine (K) and arginine (R). Any number and combination of H, K and/or R may be substituted for the inward facing amino acids in the cap forming region and/or barrel forming region.

The uncharged amino acids, non-polar amino acids and/or aromatic amino acids can be naturally occurring or non-naturally-occurring. They may be synthetic or modified. Uncharged amino acids have no net charge. Suitable uncharged amino acids include, but are not limited to, cysteine (C), serine (S), threonine (T), methionine (M), asparagines (N) and glutamine (Q). Non-polar amino acids have non-polar side chains. Suitable non-polar amino acids include, but are not limited to, glycine (G), alanine (A), proline (P), isoleucine (I), leucine (L) and valine (V). Aromatic amino acids have an aromatic side chain. Suitable aromatic amino acids include, but are not limited to, histidine (H), phenylalanine (F), tryptophan (W) and tyrosine (Y). Any number and combination of these amino acids may be substituted into the inward facing amino acids in the cap forming region and/or the barrel forming region.

The one or more negatively charged amino acids are preferably substituted with alanine (A), valine (V), asparagine (N) or glycine (G). Preferred substitutions include, but are not limited to, substitution of D with A, substitution of D with V, substitution of D with N and substitution of D with G.

The one or more modifications are preferably one or more introductions of positively charged amino acids. The introduction of positive charge decreases the net negative charge. The one or more positively charged amino acids may be introduced by addition or substitution. Any amino acid may be substituted with a positively charged amino acid. One or more uncharged amino acids, non-polar amino acids and/or aromatic amino acids may be substituted with one or more positively charged amino acids. Any number of positively charged amino acids may be introduced.

Wild-type MspA comprises a polar glutamine (Q) at position 126. The one or more modifications preferably reduce the net negative charge at position 126. The one or more modifications preferably increase the net positive charge at positions 126. This can be achieved by replacing the polar amino acid at position 126 or an adjacent or a nearby inward facing amino acid with a positively charged amino acid. The variant preferably comprises a positively charged amino acid at position 126. The variant preferably comprises a positively charged amino acid at one or more of positions 123, 125, 127 and 128. The variant may comprise any number and combination of positively charged amino acids at positions 123, 125, 127 and 128. The positively charged amino acid(s) may be introduced by addition or substitution.

The one or more modifications are preferably one or more introductions of positively charged amino acids which neutralise one or more negatively charged amino acids. The neutralisation of negative charge decreases the net negative charge. The one or more positively charged amino acids may be introduced by addition or substitution. Any amino acid may be substituted with a positively charged amino acid. One or more uncharged amino acids, non-polar amino acids and/or aromatic amino acids may be substituted with one or more positively charged amino acids. Any number of positively charged amino acids may be introduced. The number is typically the same as the number of negatively charged amino acids being neutralised.

The one or more positively charged amino acids may be introduced at any position in the cap forming region and/or the barrel forming region as long as they neutralise the negative charge of the one or more inward facing negatively charged amino acids. To effectively neutralise the negative charge in the cap forming region, there is typically 5 or fewer amino acids in the variant between each positively charged amino acid that is introduced and the negatively charged amino acid it is neutralising. There are preferably 4 or fewer, 3 or fewer or 2 or fewer amino acids in the cap forming region of the variant between each positively charged amino acid that is introduced and the negatively charged amino acid it is neutralising. There is more preferably two amino acids in the cap forming region of the variant between each positively charged amino acid that is introduced and the negatively charged amino acid it is neutralising. Each positively charged amino acid is most preferably introduced adjacent in the cap forming region of the variant to the negatively charged amino acid it is neutralising.

To effectively neutralise the negative charge in the barrel forming region, there is typically 5 or fewer inward facing amino acids between each positively charged amino acid that is introduced and the negatively charged amino acid it is neutralising. There is preferably 4 or fewer, 3 or fewer or 2 or fewer inward facing amino acids in the barrel forming region of the variant between each positively charged amino acid that is introduced and the negatively charged amino acid it is neutralising. There is more preferably one inward facing amino acid in the barrel forming region of the variant between each positively charged amino acid that is introduced and the negatively charged amino acid it is neutralising. Each positively charged amino acid is most preferably introduced at the inward facing position adjacent in the barrel forming region of the variant to the negatively charged amino acid it is neutralising.

Wild-type MspA comprises aspartic acid (D) at positions 118 and 134 and glutamic acid (E) at position 139. Amino acid 118 in each monomer is present within the barrel of the pore (Figure 37). The variant preferably comprises a positively charged amino acid at one or more of positions 114, 116, 120, 123, 70, 73, 75, 77 and 79. Positive charges at one or more of these positions neutralise the negative charge at position 118. Positively charged amino acids may present at any number and combination of positions 114, 116, 120, 123, 70, 73, 75, 77 and 79. The amino acids may be introduced by addition or substitution.

Amino acids 134 and 139 in each monomer are part of the cap. The variant comprises a positively charged amino acid at one or more of positions 129, 132, 136, 137, 59, 61 and 63. Positive charges at one or more of these positions neutralise the negative charge at position 134. Positively charged amino acids may be present at any number and combination of positions 129, 132, 136, 137, 59, 61 and 63. The amino acids may be introduced by addition or substitution.

The variant preferably comprises a positively charged amino acid at one or more of positions 137, 138, 141, 143, 45, 47, 49 and 51. Positive charges at one or more of these positions neutralise the negative charge at position 139. Positively charged amino acids may be present at any number and combination of positions 137, 138, 141, 143, 45, 47, 49 and 51. The amino acids may be introduced by addition or substitution.

### *Positions 118, 126, 134 and 139*

The one or more modifications preferably reduce the net negative charge at one or more of positions 118, 126, 134 and 139. The one or more modifications preferably reduce the net negative charge at 118; 126; 134; 139; 118 and 126; 118 and 134; 118 and 139; 126 and 134; 126 and 139; 134 and 139; 118, 126 and 134; 118, 126 and 139; 118, 134 and 139; 126, 134 and 139; or 118, 126, 134 and 139.

The variant preferably does not comprise aspartic acid (D) or glutamic acid (E) at one or more of positions 118, 126, 134 and 139. The variant preferably does not comprise aspartic acid (D) or glutamic acid (E) at any of the combination of positions 118, 126, 134 and 139 disclosed above. The variant more preferably comprises arginine (R), glycine (G) or asparagine (N) at one or more of positions 118, 126, 134 and 139, such as any of the combinations of positions 118, 126, 134 and 139 disclosed above. The variant most preferably comprises D118R, Q126R, D134R and E139K.

Methods for introducing or substituting naturally-occurring amino acids are well known in the art. For instance, methionine (M) may be substituted with arginine (R) by replacing the codon for methionine (ATG) with a codon for arginine (CGT) at the relevant position in a polynucleotide encoding the mutant monomer. The polynucleotide can then be expressed as discussed below.

Methods for introducing or substituting non-naturally-occurring amino acids are also well known in the art. For instance, non-naturally-occurring amino acids may be introduced by including synthetic aminoacyl-tRNAs in the IVTT system used to express the mutant monomer. Alternatively, they may be introduced by expressing the mutant monomer in *E. coli* that are auxotrophic for specific amino acids in the presence of synthetic (i.e. non-naturally-occurring) analogues of those specific amino acids. They may also be produced by naked ligation if the mutant monomer is produced using partial peptide synthesis.

The one or more modifications are preferably one or more chemical modifications of one or more negatively charged amino acids which neutralise their negative charge. For instance, the one or more negatively charged amino acids may be reacted with a carbodiimide.

*Other modifications*

The variant preferably comprises one or more of:

(a) serine (S) at position 75;

(b) serine (S) at position 77; and

(c) asparagine (N) or lysine (K) at position 88.

The variant may comprise any number and combination of (a) to (c), including (a), (b), (c), (a) and (b), (b) and (c), (a) and (c) and (a), (b) and (c). The variant preferably comprises G75S, G77S and L88N.

The variant most preferably comprises (a) D90N, D91N, D93N, D118R, D134R and E139K, (b) L88N, D90N, D91N, D93N, D118R, D134R and E139K, (c) G75S, G77S, L88N, D90N, D91N, D93N, D118R, Q126R, D134R and E139K or (d) G75S, G77S, L88N, D90N, D91N, D118R, Q126R, D134R and E139K.

The variant preferably further comprises one or more of:

(e) phenylalanine (F) at position 89;
(f) glutamic acid (E) at position 95 and lysine (K) at position 98;
(g) aspartic acid (D) at position 96;
(h) glycine (G) at position 102;
(i) alanine (A) at position 103; and
(j) alanine (A), serine (S) or proline (P) at position 108.

The may comprise any number and combination of (e) to (j).

*Variants*

In addition to the specific mutations discussed above, the variant of SEQ ID NO: 2 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/).

SEQ ID NO: 2 is the mature form of the wild-type 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 shown in SEQ ID NOs: 5 to 7. 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 2 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 3.

Table 2 – 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 3 - 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. 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.

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 monomers derived from Msp may be modified to assist their identification or purification, for example by the addition of a streptavidin tag or by the addition of a signal sequence to promote their secretion from a cell where the monomer does not naturally contain such a sequence. Other suitable tags are discussed in more detail below. The monomer may be labelled with a revealing label. The revealing label may be any suitable label which allows the monomer to be detected. Suitable labels are described below.

The monomer derived from Msp may also be produced using D-amino acids. For instance, the monomer derived from Msp 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 monomer derived from Msp contains one or more specific modifications to facilitate nucleotide discrimination. The monomer derived from Msp 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 monomer derived from Msp. 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 monomer derived from Msp can be produced using standard methods known in the art. The monomer derived from Msp may be made synthetically or by recombinant means. For example, the monomer may be synthesized by *in vitro* transcription and translation (IVTT). Alternatively the monomer may be synthesized by recombinant protein expression in *E.coli*. Suitable methods for producing pores and monomers are discussed in International Application Nos. PCT/GB09/001690 (published as WO 2010/004273), PCT/GB09/001679 (published as WO 2010/004265) or PCT/GB10/000133 (published as WO 2010/086603). Methods for inserting pores into membranes are discussed.

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 polynucleotide sequence. The presence of the adaptor improves the host-guest chemistry of the pore and the nucleotide or polynucleotide 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 polynucleotide 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 polynucleotide 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 polynucleotide sequence via host-guest chemistry. The adaptor is typically capable of interacting with the nucleotide or polynucleotide sequence. The adaptor comprises one or more chemical groups that are capable of interacting with the nucleotide or polynucleotide sequence. The one or more chemical groups preferably interact with the nucleotide or polynucleotide sequence by non-covalent interactions, such as hydrophobic interactions, hydrogen bonding, Van der Waal's forces, $\pi$-cation interactions and/or electrostatic forces. The one or more chemical groups that are capable of interacting with the nucleotide or polynucleotide sequence are preferably positively charged. The one or more chemical groups that are capable of interacting with the nucleotide or polynucleotide 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 polynucleotide 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 facilitates the interaction between the pore and the nucleotide or polynucleotide 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 is 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 polynucleotide binding protein. This forms a modular sequencing system that may be used in the methods of sequencing of the invention. Polynucleotide binding proteins are discussed below.

The polynucleotide binding protein is preferably covalently attached to the mutant monomer. The protein can be covalently attached to the monomer 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 monomer to a polynucleotide binding protein is discussed in International Application No. PCT/GB09/001679 (published as WO 2010/004265).

If the polynucleotide 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 polynucleotide binding protein. The reactivity of cysteine residues may be enhanced by modification as described above.

The polynucleotide 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 polynucleotide binding protein.

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 monomer before a linker is attached.

The molecule (with which the monomer is chemically modified) may be attached directly to the monomer or attached via a linker as disclosed in International Application Nos. PCT/GB09/001690 (published as WO 2010/004273), PCT/GB09/001679 (published as WO 2010/004265) or PCT/GB10/000133 (published as WO 2010/086603).

Any of the proteins described herein, such as the mutant monomers and pores of the invention, 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, a flag tag, a SUMO tag, a GST tag or a MBP 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 protein. An example of this would be to react a gel-shift reagent to a cysteine engineered on the outside of the protein. This has been demonstrated as a method for separating hemolysin hetero-oligomers (Chem Biol. 1997 Jul;4(7):497-505).

Any of the proteins described herein, such as the mutant monomers and pores of the invention, may be labelled with a revealing label. The revealing label may be any suitable label which allows the protein 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.

Any of the proteins described herein, such as the monomers or pores of the invention, may be made synthetically or by recombinant means. For example, the protein may be synthesized by *in vitro* translation and transcription (IVTT). The amino acid sequence of the protein may be modified to include non-naturally occurring amino acids or to increase the stability of the protein. When a protein is produced by synthetic means, such amino acids may be introduced during production. The protein may also be altered following either synthetic or recombinant production.

Proteins may also be produced using D-amino acids. For instance, the protein 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 protein may also contain other non-specific modifications as long as they do not interfere with the function of the protein. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the protein(s). 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.

Any of the proteins described herein, including the monomers and pores of the invention, can be produced using standard methods known in the art. Polynucleotide sequences encoding a protein may be derived and replicated using standard methods in the art. Polynucleotide sequences encoding a protein may be expressed in a bacterial host cell using standard techniques in the art. The protein 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. These methods are described in Sambrook, J. and Russell, D. (2001). Molecular Cloning: A Laboratory Manual, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Proteins may be produced in large scale following purification by any protein liquid chromatography system from protein producing organisms or after recombinant expression. Typical protein liquid chromatography systems include FPLC, AKTA systems, the Bio-Cad system, the Bio-Rad BioLogic system and the Gilson HPLC system.

## Constructs

The invention also provides a construct comprising two or more covalently attached MspA monomers, wherein at least one of the monomers is a mutant monomer of the invention. The construct of the invention retains its ability to form a pore. This may be determined as discussed above. One or more constructs of the invention may be used to form pores for characterising, such as sequencing, polynucleotides. The construct may comprise at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9 or at least 10 monomers. The construct preferably comprises two monomers. The two or more monomers may be the same or different.

At least one monomer in the construct is a mutant monomer of the invention. 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more or 10 or more monomers in the construct may be mutant monomers of the invention. All of the monomers in the construct are preferably mutant monomers of the invention. The mutant monomers may be the same or different. In a preferred embodiment, the construct comprises two mutant monomers of the invention.

The mutant monomers of the invention in the construct are preferably approximately the same length or are the same length. The barrels of the mutant monomers of the invention in the construct are preferably approximately the same length or are the same length. Length may be measured in number of amino acids and/or units of length.

The construct may comprise one or more monomers which are not mutant monomers of the invention. MspA mutant monomers which are non mutant monomers of the invention include monomers comprising SEQ ID NO: 2 or a comparative variant of SEQ ID NO: 2. At least one monomer in the construct may comprise SEQ ID NO: 2 or a comparative variant of the sequence shown in SEQ ID NO: 2 which comprises one or more of, preferably all of, D90N, D91N, D93N, D118R, D134R and E139K. At least one monomer in the construct may be any of the monomers disclosed in International Application No. PCT/GB2012/050301 (published as WO/2012/107778), including those comprising a comparative variant of the sequence shown in SEQ ID NO: 2 which comprises G75S, G77S, L88N, D90N, D91N, D93N, D118R, Q126R, D134R and E139K. A comparative variant of SEQ ID NO: 2 is at least 50% homologous to SEQ ID NO: 2 over its entire sequence based on amino acid identity. More preferably, the comparative 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.

The monomers in the construct 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. 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 (without an amino terminal methionine) and mM is a monomer 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.

In another preferred embodiment, the monomers are chemically fused. Two monomers are chemically fused 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).

Mutant αHL and lysenin monomers

The present invention also provides mutant αHL monomers. The mutant Msp monomers may be used to form the pores of the invention. A mutant αHL monomer is a monomer whose sequence varies from that of a wild-type αHL 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.

A mutant monomer of the invention comprises a variant of the sequence shown in SEQ ID NO: 4. Variants are discussed above. The mutant αHL monomer is modified in accordance with the invention. A part of the monomer which interacts with a polynucleotide binding protein has been modified. The part of the monomer which interacts with a polynucleotide binding protein when the polynucleotide binding protein is used to control the movement of a target polynucleotide with respect to, or through, a pore comprising the monomer has been modified. Any of the modifications discussed above may be made in the monomer of the invention.

The monomer preferably comprises a variant of SEQ ID NO: 4 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions:

(a) 16, 17, 18, 19, 21, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 281, 283, 285, 287, 288 and 293 or at one or more of the corresponding positions thereof;

(b) 17, 18, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 287, 288 and 293 or the corresponding positions thereof;

(c) 17, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242 and 287 or the corresponding positions thereof; or

(d) 17, 19, 46, 93, 236, 237, 239, 240, 287 and 288 or the corresponding positions thereof.

The present invention also provides mutant lysenin monomers. The mutant lysenin monomers may be used to form the pores of the invention. A mutant lysenin monomer is a monomer whose sequence varies from that of a wild-type lysenin 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.

A mutant monomer of the invention comprises a variant of the sequence shown in SEQ ID NO: 36. Variants are discussed above. The mutant lysenin monomer is modified in accordance with the invention. A part of the monomer which interacts with a polynucleotide binding protein has been modified. The part of the monomer which interacts with a polynucleotide binding protein when the polynucleotide binding protein is used to control the movement of a target polynucleotide with respect to, or through, a pore comprising the monomer has been modified. Any of the modifications discussed above may be made in the monomer of the invention.

The monomer preferably comprises a variant of SEQ ID NO: 36 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions 31, 33, 108, 109, 110 and 138 or the corresponding positions thereof.

The mutant αHL monomers and mutant lysenin monomers of the invention may be used to form constructs, homo-oligomeric pores and hetero-oligomeric pores in the same way as mutant Msp monomers discussed above. Such pores typically comprise seven mutant αHL monomers.

## Modified polynucleotide binding proteins

The method also provides a polynucleotide binding protein modified in accordance with the invention. A part of the polynucleotide binding protein which interacts with a transmembrane pore has been modified. The part of the polynucleotide binding protein which interacts with a transmembrane pore when the polynucleotide binding protein is used to control the movement of a target polynucleotide with respect to, or through, the pore has been modified. Any of the modifications discussed above may be made in the proteins of the invention.

The polynucleotide binding protein preferably comprises a variant of the sequence shown in SEQ ID NO: 24. Preferred variants are discussed above. The part of the variant of SEQ ID NO: 24 that may be modified in accordance with the invention is discussed above. The modified polynucleotide binding protein of the invention preferably comprises a variant of SEQ ID NO: 24 which comprises K199A, K199V, K199F, K199D, K199S, K199W or K199L. The variant of SEQ ID NO: 24 preferably further comprises any of the modifications discussed below.

The polynucleotide binding protein preferably comprises a variant of the sequence shown in SEQ ID NO: 9. Preferred variants are discussed above. The part of the variant of SEQ ID NO: 9 that may be modified in accordance with the invention is discussed above. The modified polynucleotide binding protein of the invention preferably comprises a variant of SEQ ID NO: 9 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions 80, 81, 82, 84, 85, 205, 206, 209, 215, 216, 220, 221, 224, 236, 240, 241, 267, 270, 272, 278, 287, 289, 293, 296, 307, 308, 309, 310, 320, 321, 322, 323, 327, 349, 415, 418 and 419 or the corresponding positions thereof, such as at one or more of:

(a) positions 80, 84, 205, 209, 215, 216, 221, 224, 236, 241, 267, 272, 289, 296, 307, 308, 309, 320, 321, 322, and 419 or the corresponding positions thereof;

(b) positions 80, 84, 209, 215, 216, 221, 267, 272, 289, 307, 308, 309, 321 and 322 or the corresponding positions thereof; or

(b) positions 215, 267, 272, 307, 308 and 322 or the corresponding positions thereof.

<u>Polynucleotides</u>

The present invention provides polynucleotide sequences which encode a modified transmembrane pore of the invention. The modified pore may be any of those discussed above or below.

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 variants of the sequence shown in SEQ ID NO: 1. The polynucleotide sequence preferably comprises two or more sequences having at least 50%, 60%, 70%, 80%, 90% or 95% homology to SEQ ID NO: 1 based on nucleotide identity 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 present invention also provides polynucleotide sequences which encode any of the modified polynucleotide binding proteins of the invention. 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: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 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: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 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 monomers or constructs, the different monomers or constructs 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 $\lambda_L$ promoter is typically used.

The host cell typically expresses the monomer or construct 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.

The invention also comprises a method of producing a mutant monomer of the invention or a construct of the invention. The method comprises expressing a polynucleotide of the invention in a suitable host cell. The polynucleotide is preferably part of a vector and is preferably operably linked to a promoter.

## Pores

The invention also provides various pores. The movement of target polynucleotides with respect to, such as through, the pores of the invention is typically more consistent. The pores preferably allow more consistent movement from k-mer to k-mer as the target polynucleotide moves with respect to, such as through, the pore. The pores preferably allow the target polynucleotide to move with respect to, such as through, the pore more smoothly. The pore preferably reduces the amount of stuttering associated with the movement of the target polynucleotide with respect to, such as through, the pore. The pores preferably provide more regular or less irregular movement of the target polynucleotide with respect to, such as through, the pore.

The noise associated with the movement of a target polynucleotide with respect to, such as through, the pore of the invention is typically reduced. The pores of the invention may reduce this noise by reducing unwanted movement associated with one or more k-mers, such as each k-mer, in the target polynucleotide. The pores of the invention may reduce the noise associated with the current level or signature for one or more k-mers, such as each k-mer, in the target polynucleotide.

If the target polynucleotide is double stranded, the noise associated with movement of the complement strand relative to the template strand is reduced and/or the movement of the complement strand relative to the template strand is more consistent using the pores of the invention. This is advantageous for strand sequencing of double stranded target polynucleotides.

The pores of the invention are ideal for characterising, such as sequencing, polynucleotide 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 polynucleotide 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 is present in a form that comprises less than 10%, less than 5%, less than 2% or less than 1% of other components, such as triblock copolymers, lipids or other pores. Alternatively, a pore of the invention may be present in a membrane. Suitable membranes are discussed below.

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.

### *Modified pores*

The invention provides a transmembrane pore in which a part of the transmembrane pore which interacts with a polynucleotide binding protein has been modified. The pore may be any of those discussed above. Where the pore comprises one or more variants of SEQ ID NO: 2, the one or more variants preferably do not comprise or consist of E59R, D90N, D91N, D93N, D118R, D134R and E139K.

### *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 may comprise any of the mutants of the invention. The homo-oligomeric pore of the invention is ideal for characterising, such as sequencing, polynucleotides. The homo-oligomeric pore of the invention may have any of the advantages discussed above.

The homo-oligomeric pore may contain any number of mutant monomers. The pore typically comprises at least 7, at least 8, at least 9 or at least 10 identical mutant monomers, such as 7, 8, 9 or 10 mutant monomers. The pore preferably comprises eight or nine 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.

A preferred homo-oligomeric pore comprises eight or nine subunits each comprising a variant of SEQ ID NO: 2 in which L74, G75, D118 and L119 have been deleted and which comprises D56N, E59R, L88N, D90N, D91N, Q126R, D134R and E139K.

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. The hetero-oligomeric pore of the invention is ideal for characterising, such as sequencing, polynucleotides. 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 at least 7, at least 8, at least 9 or at least 10 monomers, such as 7, 8, 9 or 10 monomers. The pore preferably comprises eight or nine monomers.

In a preferred embodiment, all of the monomers (such as 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 or nine mutant monomers of the invention and at least one of them differs from the others. They may all differ from one another.

The mutant monomers of the invention in the pore are preferably approximately the same length or are the same length. The barrels of the mutant monomers of the invention in the pore are preferably approximately the same length or are the same length. Length may be measured in number of amino acids and/or units of length. The mutant monomers of the invention in the pore preferably have the same number of amino acids deleted from positions 72 to 82 and/or positions 111 to 121.

In another preferred embodiment, at least one of the mutant monomers is not a mutant monomer of the invention. 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. Any number of the monomers in the pore may not be a mutant monomer of the invention. The pore preferably comprises seven or eight mutant monomers of the invention and a monomer which is not a monomer of the invention. The mutant monomers of the invention may be the same or different.

The mutant monomers of the invention in the construct are preferably approximately the same length or are the same length. The barrels of the mutant monomers of the invention in the construct are preferably approximately the same length or are the same length. Length may be measured in number of amino acids and/or units of length.

The pore may comprise one or more monomers which are not mutant monomers of the invention. MspA mutant monomers which are non mutant monomers of the invention include monomers comprising SEQ ID NO: 2 or a comparative variant of SEQ ID NO: 2. At least one monomer in the pore may comprise SEQ ID NO: 2 or a comparative variant of the sequence shown in SEQ ID NO: 2 which comprises one or more of, preferably all of, D90N, D91N, D93N, D118R, D134R and E139K. At least one monomer in the pore may be any of the monomers disclosed in International Application No. PCT/GB2012/050301 (published as WO/2012/107778), including those comprising a comparative variant of the sequence shown in SEQ ID NO: 2 which comprises G75S, G77S, L88N, D90N, D91N, D93N, D118R, Q126R, D134R and E139K. A comparative variant of SEQ ID NO: 2 is at least 50% homologous to SEQ ID NO: 2 over its entire sequence based on amino acid identity. More preferably, the comparative 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 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.

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 wherein at least one of the monomers is a mutant monomer of the invention. 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) four constructs each comprising two constructs, (b) two constructs each comprising four monomers or (b) one construct comprising two monomers and six monomers that do not form part of a construct. For instance, an nonameric pore may comprise (a) four constructs each comprising two constructs and one monomer that does not form part of a construct, (b) two constructs each comprising four monomers and a monomer that does not form part of a construct or (b) one construct comprising two monomers and seven monomers that do not form part of a construct. Other combinations of constructs and monomers can be envisaged by the skilled person.

At least two of the monomers in the pore are in the form of a construct of the invention. The construct, and hence the pore, comprises at least one mutant monomer of the invention. The pore typically comprises at least 7, at least 8, at least 9 or at least 10 monomers, such as 7, 8, 9 or 10 monomers, in total (at least two of which must be in a construct). The pore preferably comprises eight or nine monomers (at least two of which must be in a construct).

The construct containing pore may be a homo-oligomer (i.e. include identical constructs) or be a hetero-oligomer (i.e. where at least one construct differs from the others).

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, monomers comprising SEQ ID NO: 2 and mutant monomers comprising a comparative variant of SEQ ID NO: 2 as discussed above.

Another typical pore comprises more than one construct of the invention, such as two, three or four constructs of the invention. If necessary, such pores further comprise sufficient additional monomers or constructs to form the pore. The additional monomer(s) may be any of those discussed above, including mutant monomers of the invention, monomers comprising SEQ ID NO: 2 and mutant monomers comprising a comparative variant of SEQ ID NO: 2 as discussed above. The additional construct(s) may be any of those discussed above or may be a construct comprising two or more covalently attached MspA monomers each comprising a monomer comprising SEQ ID NO: 2 or a comparative variant of SEQ ID NO: 2 as 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. At least one construct is a construct of the invention, i.e. at least one monomer in the at least one construct, and preferably each monomer in the at least one construct, is a mutant monomer of the invention. All of the constructs comprising 2 monomers may be constructs of the invention.

A specific pore according to the invention comprises four constructs of the invention each comprising two monomers, wherein at least one monomer in each construct, and preferably each monomer in each construct, is a mutant monomer of the invention. The constructs may oligomerise into a pore with a structure such that only one monomer of each construct contributes to the channel of the pore. Typically the other monomers of the construct will be on the outside of the channel of the pore. For example, pores of the invention may comprise 5, 6, 7 or 8 constructs comprising 2 monomers where the channel 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 or constructs. In other words, monomers comprising MutA are fused followed by oligomerisation with MutB-containing monomers to form A-A: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.

## Combinations

The invention also provides a combination of a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified. The pore may be any of those discussed above. The polynucleotide binding protein may be any of those discussed above.

The pore in the combination preferably comprises seven or more monomers, such as 8 or 9 monomers, comprising the sequence shown in SEQ ID NO: 2 or a variant thereof, wherein one or more of the seven or more monomers, such as 2, 3, 4, 5, 6, 7, 8 or 9 of the monomers, comprises a variant of SEQ ID NO: 2 which comprises (a) D56N, D56R, D56F, D56Y or D56L and/or (b) E59N, E59R, E59F, E59Y or E59L (preferably D56N and E59R, D56F and E59R, D56N and E59F, D56N and E59Y or D56L and E59L or more preferably D56N and E59R, D56F and E59R or D56N and E59F) and the modified polynucleotide binding protein in the combination preferably comprises a variant of SEQ ID NO: 24 which comprises K199A, K199V, K199F, K199D, K199S, K199W or K199L. The variant of SEQ ID NO: 24 preferably further comprises any of the modifications discussed below:

Preferred combinations of variants are shown in each row below.

| **SEQ ID NO: 2 comprising** | **SEQ ID NO: 24 comprising** |
|---|---|
| D56N and E59R | K199L |
| D56F and E59R | K199L |
| D56N and E59F | K199L |

## Polynucleotide characterisation

The invention provides a method of characterising a target polynucleotide. The method involves measuring one or more characteristics of the target polynucleotide. The target polynucleotide may also be called the template polynucleotide or the polynucleotide of interest.

### *Polynucleotide*

A polynucleotide, such as a nucleic acid, is a macromolecule comprising two or more nucleotides. The polynucleotide or nucleic acid may comprise any combination of any nucleotides. The nucleotides can be naturally occurring or artificial. One or more nucleotides in the polynucleotide can be oxidized or methylated.

The polynucleotide may be single stranded or double stranded. At least a portion of the polynucleotide is preferably double stranded.

The polynucleotide can be a nucleic acid, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The polynucleotide can comprise one strand of RNA hybridised to one strand of DNA. The polynucleotide 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 PNA backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds. The GNA backbone is composed of repeating glycol units linked by phosphodiester bonds. The TNA backbone is composed of repeating threose sugars linked together by phosphodiester bonds.

LNA is formed from ribonucleotides as discussed above having an extra bridge connecting the 2' oxygen and 4' carbon in the ribose moiety.

The polynucleotide is most preferably ribonucleic nucleic acid (RNA) or deoxyribonucleic acid (DNA).

The polynucleotide can be any length. For example, the polynucleotide 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 or nucleotide pairs in length. The polynucleotide can be 1000 or more nucleotides or nucleotide pairs, 5000 or more nucleotides or nucleotide pairs in length or 100000 or more nucleotides or nucleotide pairs in length.

Any number of polynucleotides can be investigated. For instance, the method of the invention may concern characterising 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100 or more polynucleotides. If two or more polynucleotides are characterized, they may be different polynucleotides or two instances of the same polynucleotide.

The polynucleotide can be naturally occurring or artificial. For instance, the method may be used to verify the sequence of a manufactured oligonucleotide. The method is typically carried out *in vitro*.

### *Sample*

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

The first sample and/or second sample may be a biological sample. The invention may be carried out *in vitro* using at least one sample obtained from or extracted from any organism or microorganism. The first sample and/or second sample may be a non-biological sample. The non-biological sample is preferably a fluid sample. Examples of non-biological samples include surgical fluids, water such as drinking water, sea water or river water, and reagents for laboratory tests.

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

*Characterisation*

The method may involve measuring two, three, four or five or more characteristics of the polynucleotide. The one or more characteristics are preferably selected from (i) the length of the polynucleotide, (ii) the identity of the polynucleotide, (iii) the sequence of the polynucleotide, (iv) the secondary structure of the polynucleotide and (v) whether or not the polynucleotide is modified. Any combination of (i) to (v) may be measured in accordance with the invention, such as {i}, {ii}, {iii}, {iv}, {v}, {i,ii}, {i,iii}, {i,iv}, {i,v}, {ii,iii}, {ii,iv}, {ii,v}, {iii,iv}, {iii,v}, {iv,v}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,iii,iv}, {i,iii,v}, {i,iv,v}, {ii,iii,iv}, {ii,iii,v}, {ii,iv,v}, {iii,iv,v}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iv,v}, {i,iii,iv,v}, {ii,iii,iv,v} or {i,ii,iii,iv,v}. Different combinations of (i) to (v) may be measured for the first polynucleotide compared with the second polynucleotide, including any of those combinations listed above.

For (i), the length of the polynucleotide may be measured for example by determining the number of interactions between the polynucleotide and the pore or the duration of interaction between the polynucleotide and the pore.

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

For (iii), the sequence of the polynucleotide 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, if the method involves an electrical measurement, the secondary structure may be measured using a change in dwell time or a change in current flowing through the pore. This allows regions of single-stranded and double-stranded polynucleotide to be distinguished.

For (v), the presence or absence of any modification may be measured. The method preferably comprises determining whether or not the polynucleotide is modified by methylation, by oxidation, by damage, with one or more proteins or with one or more labels, tags or spacers. Specific modifications will result in specific interactions with the pore which can be measured using the methods described below. For instance, methylcyotsine may be distinguished from cytosine on the basis of the current flowing through the pore during its interaction with each nucleotide.

The target polynucleotide is contacted with a pore of the invention. The pore is typically present in a membrane. Suitable membranes are discussed below. The method may be carried out using any apparatus that is suitable for investigating a membrane/pore system in which a pore is present in a membrane. The method may be carried out using any apparatus that is suitable for transmembrane pore sensing. For example, the apparatus comprises a chamber comprising an aqueous solution and a barrier that separates the chamber into two sections. The barrier typically has an aperture in which the membrane containing the pore is formed. Alternatively the barrier forms the membrane in which the pore is present.

The method may be carried out using the apparatus described in International Application No. PCT/GB08/000562 (WO 2008/102120).

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

Electrical measurements may be made using standard single channel recording equipment as describe in Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7, Lieberman KR et al, J Am Chem Soc. 2010;132(50):17961-72, and International Application WO 2000/28312. Alternatively, electrical measurements may be made using a multi-channel system, for example as described in International Application WO 2009/077734 and International Application WO 2011/067559.

The method is preferably carried out with a potential applied across the membrane. The applied potential may be a voltage potential. Alternatively, the applied potential may be a chemical potential. An example of this is using a salt gradient across a membrane, such as an amphiphilic layer. A salt gradient is disclosed in Holden *et al.*, J Am Chem Soc. 2007 Jul 11;

129(27):8650-5. In some instances, the current passing through the pore as a polynucleotide moves with respect to the pore is used to estimate or determine the sequence of the polynucleotide. This is strand sequencing.

The method may involve measuring the current passing through the pore as the polynucleotide moves with respect to, such as through, the pore. Therefore the apparatus used in the method may also comprise 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.

The method of the invention may involve the measuring of a current passing through the pore as the polynucleotide moves with respect to the pore. Suitable conditions for measuring ionic currents through transmembrane protein pores are known in the art and disclosed in the Example. The method is typically carried out with a voltage applied across the membrane and pore. The voltage used is typically from +5 V to -5 V, such as from +4 V to -4 V, +3 V to -3 V or +2 V to -2 V. The voltage used is typically from -600 mV to +600mV or -400 mV to +400 mV. The voltage used is preferably in a range having a lower limit selected from -400 mV, -300 mV, -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 100 mV to 240 mV and most preferably in the range of 120 mV to 220 mV. It is possible to increase discrimination between different nucleotides by a pore by using an increased applied potential.

The method is typically carried out in the presence of any charge carriers, such as metal salts, for example alkali metal salt, halide salts, for example chloride salts, such as alkali metal chloride salt. Charge carriers may include ionic liquids or organic salts, for example tetramethyl ammonium chloride, trimethylphenyl ammonium chloride, phenyltrimethyl ammonium chloride, or 1-ethyl-3-methyl imidazolium chloride. In the exemplary apparatus discussed above, the salt is present in the aqueous solution in the chamber. Potassium chloride (KCl), sodium chloride (NaCl), caesium chloride (CsCl) or a mixture of potassium ferrocyanide and potassium ferricyanide is typically used. KCl, NaCl and a mixture of potassium ferrocyanide and potassium ferricyanide are preferred. The charge carriers may be asymmetric across the membrane. For instance, the type and/or concentration of the charge carriers may be different on each side of the membrane.

The salt concentration may be at saturation. The salt concentration may be 3 M or lower and is typically from 0.1 to 2.5 M, from 0.3 to 1.9 M, from 0.5 to 1.8 M, from 0.7 to 1.7 M, from

0.9 to 1.6 M or from 1 M to 1.4 M. The salt concentration is preferably from 150 mM to 1 M. The method is preferably carried out using a salt concentration of at least 0.3 M, such as at least 0.4 M, at least 0.5 M, at least 0.6 M, at least 0.8 M, at least 1.0 M, at least 1.5 M, at least 2.0 M, at least 2.5 M or at least 3.0 M. 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 fluctuations.

The method is typically carried out in the presence of a buffer. In the exemplary apparatus discussed above, the buffer is present in the aqueous solution in the chamber. Any buffer may be used in the method of the invention. Typically, the buffer is phosphate buffer. Other suitable buffers are HEPES and 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 method may be 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 are typically carried out at room temperature. The methods are optionally carried out at a temperature that supports enzyme function, such as about 37 °C.

*Polynucleotide binding protein*

Step (a) preferably comprises contacting the polynucleotide with a polynucleotide binding protein such that the protein controls the movement of the polynucleotide with respect to, such as through, the pore.

More preferably, the method comprises (a) contacting the polynucleotide with the pore of the invention and a polynucleotide binding protein such that the protein controls the movement of the polynucleotide through the pore and (b) taking one or more measurements as the polynucleotide moves with respect to the pore, wherein the measurements are indicative of one or more characteristics of the polynucleotide, and thereby characterising the polynucleotide.

More preferably, the method comprises (a) contacting the polynucleotide with the pore of the invention and a polynucleotide binding protein such that the protein controls the movement of the polynucleotide through the pore and (b) measuring the current through the pore as the polynucleotide moves with respect to the pore, wherein the current is indicative of one or more characteristics of the polynucleotide, and thereby characterising the polynucleotide.

The polynucleotide binding protein may be any protein that is capable of binding to the polynucleotide and controlling its movement with respect to, or through, the pore. It is straightforward in the art to determine whether or not a protein binds to a polynucleotide. The protein typically interacts with and modifies at least one property of the polynucleotide. The protein may modify the polynucleotide by cleaving it to form individual nucleotides or shorter chains of nucleotides, such as di- or trinucleotides. The protein may modify the polynucleotide by orienting it or moving it to a specific position, i.e. controlling its movement. The polynucleotide binding protein is preferably a modified polynucleotide binding protein of the invention.

The polynucleotide binding protein is preferably derived from a polynucleotide handling enzyme. A polynucleotide handling enzyme is a polypeptide that is capable of interacting with and modifying at least one property of a polynucleotide. The enzyme may modify the polynucleotide by cleaving it to form individual nucleotides or shorter chains of nucleotides, such as di- or trinucleotides. The enzyme may modify the polynucleotide by orienting it or moving it to a specific position. The polynucleotide handling enzyme does not need to display enzymatic activity as long as it is capable of binding the polynucleotide and controlling its movement with respect to, or 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 polynucleotide handling enzyme is preferably derived from a nucleolytic enzyme. The polynucleotide 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: 11), exonuclease III enzyme from *E. coli* (SEQ ID NO: 13), RecJ from *T. thermophilus* (SEQ ID NO: 15) and bacteriophage lambda exonuclease (SEQ ID NO: 17) , TatD exonuclease and variants thereof. Three subunits comprising the sequence shown in SEQ ID NO: 15 or a variant thereof interact to form a trimer exonuclease. The polymerase may be PyroPhage® 3173 DNA Polymerase (which is commercially available from Lucigen® Corporation), SD Polymerase (commercially available from Bioron®) or variants thereof. The enzyme is preferably Phi29 DNA polymerase (SEQ ID NO: 9) or a variant thereof. The topoisomerase is preferably a member of any of the Moiety Classification (EC) groups 5.99.1.2 and 5.99.1.3.The enzyme is most preferably derived from a helicase, such as Hel308 Mbu (SEQ ID NO: 18), Hel308 Csy (SEQ ID NO: 19), Hel308 Mhu (SEQ ID NO: 21), TraI Eco (SEQ ID NO: 22), XPD Mbu (SEQ ID NO: 23) or a variant thereof. Any helicase may be used in the invention. The helicase may be or be derived from a Hel308 helicase, a RecD helicase, such as TraI helicase or a TrwC helicase, a XPD helicase or a Dda helicase. The helicase may be any of the helicases, modified helicases or helicase constructs disclosed in International Application Nos. PCT/GB2012/052579 (published as WO 2013/057495); PCT/GB2012/053274 (published as WO 2013/098562); PCT/GB2012/053273 (published as WO2013098561); PCT/GB2013/051925 (published as WO 2014/013260); PCT/GB2013/051924 (published as WO 2014/013259); PCT/GB2013/051928 (published as WO 2014/013262) and PCT/GB2014/052736 (published as WO/2015/055981).

The helicase preferably comprises the sequence shown in SEQ ID NO: 25 (Trwc Cba) or a variant thereof, the sequence shown in SEQ ID NO: 18 (Hel308 Mbu) or a variant thereof or the sequence shown in SEQ ID NO: 24 (Dda) or a variant thereof. Variants may differ from the native sequences in any of the ways discussed below for transmembrane pores. A preferred variant of SEQ ID NO: 24 comprises (a) E94C and A360C or (b) E94C, A360C, C109A and C136A and then optionally (ΔM1)G1G2 (i.e. deletion of M1 and then addition G1 and G2).

Any number of helicases may be used in accordance with the invention. For instance, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more helicases may be used. In some embodiments, different numbers of helicases may be used.

The method of the invention preferably comprises contacting the polynucleotide with two or more helicases. The two or more helicases are typically the same helicase. The two or more helicases may be different helicases.

The two or more helicases may be any combination of the helicases mentioned above. The two or more helicases may be two or more Dda helicases. The two or more helicases may be one or more Dda helicases and one or more TrwC helicases. The two or more helicases may be different variants of the same helicase.

The two or more helicases are preferably attached to one another. The two or more helicases are more preferably covalently attached to one another. The helicases may be attached in any order and using any method. Preferred helicase constructs for use in the invention are described in International Application Nos. PCT/GB2013/051925 (published as WO 2014/013260); PCT/GB2013/051924 (published as WO 2014/013259); PCT/GB2013/051928 (published as WO 2014/013262) and PCT/GB2014/052736 (published as WO/2015/055981).

A variant of SEQ ID NOs: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 and which retains polynucleotide binding ability. This can be measured using any method known in the art. For instance, the variant can be contacted with a polynucleotide and its ability to bind to and move along the polynucleotide can be measured. The variant may include modifications that facilitate binding of the polynucleotide and/or facilitate its activity at high salt concentrations and/or room temperature. Variants may be modified such that they bind polynucleotides (i.e. retain polynucleotide binding ability) but do not function as a helicase (i.e. do not move along polynucleotides when provided with all the necessary components to facilitate movement, e.g. ATP and $Mg^{2+}$). Such modifications are known in the art. For instance, modification of the $Mg^{2+}$ binding domain in helicases typically results in variants which do not function as helicases. These types of variants may act as molecular brakes (see below).

Over the entire length of the amino acid sequence of SEQ ID NO: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25, 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: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 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, 280, 300, 400, 500, 600, 700, 800, 900 or 1000 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 and 4 above. The enzyme may be covalently attached to the pore. Any method may be used to covalently attach the enzyme to the pore.

A preferred molecular brake is TrwC Cba-Q594A (SEQ ID NO: 25 with the mutation Q594A). This variant does not function as a helicase (i.e. binds polynucleotides but does not move along them when provided with all the necessary components to facilitate movement, e.g. ATP and $Mg^{2+}$).

In strand sequencing, the polynucleotide is translocated through the pore either with or against an applied potential. Exonucleases that act progressively or processively on double stranded polynucleotides 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. A polymerase may also be used. 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.

Any helicase may be used in the method. Helicases may work in two modes with respect to the pore. First, the method is preferably carried out using a helicase such that it moves the polynucleotide through the pore with the field resulting from the applied voltage. In this mode the 5’ end of the polynucleotide is first captured in the pore, and the helicase moves the polynucleotide into the pore such that it is passed through the pore with the field until it finally translocates through to the *trans* side of the membrane. Alternatively, the method is preferably carried out such that a helicase moves the polynucleotide through the pore against the field resulting from the applied voltage. In this mode the 3’ end of the polynucleotide is first captured in the pore, and the helicase moves the polynucleotide through the pore such that it is pulled out of the pore against the applied field until finally ejected back to the *cis* side of the membrane.

The method may also be carried out in the opposite direction. The 3’ end of the polynucleotide may be first captured in the pore and the helicase may move the polynucleotide into the pore such that it is passed through the pore with the field until it finally translocates through to the *trans* side of the membrane.

When the helicase is not provided with the necessary components to facilitate movement or is modified to hinder or prevent its movement, it can bind to the polynucleotide and act as a brake slowing the movement of the polynucleotide when it is pulled into the pore by the applied field. In the inactive mode, it does not matter whether the polynucleotide is captured either 3’ or 5’ down, it is the applied field which pulls the polynucleotide into the pore towards the trans side with the enzyme acting as a brake. When in the inactive mode, the movement control of the polynucleotide by the helicase can be described in a number of ways including ratcheting, sliding and braking. Helicase variants which lack helicase activity can also be used in this way.

The polynucleotide may be contacted with the polynucleotide binding protein and the pore in any order. It is preferred that, when the polynucleotide is contacted with the polynucleotide binding protein, such as a helicase, and the pore, the polynucleotide firstly forms a complex with the protein. When the voltage is applied across the pore, the polynucleotide/protein complex then forms a complex with the pore and controls the movement of the polynucleotide with respect to, such as through, the pore.

Any steps in the method using a polynucleotide binding protein are typically carried out in the presence of free nucleotides or free nucleotide analogues and an enzyme cofactor that facilitates the action of the polynucleotide binding protein. The free nucleotides may be one or more of any of the individual nucleotides discussed above. The free 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 free nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP or dCMP. The free nucleotides are preferably adenosine triphosphate (ATP). The enzyme cofactor is a factor that allows the construct to function. 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 $Mg^{2+}$.

### *Helicase(s) and molecular brake(s)*

In a preferred embodiment, the method comprises:

(a) providing the polynucleotide with one or more helicases and one or more molecular brakes attached to the polynucleotide;

(b) contacting the polynucleotide with a pore of the invention and applying a potential across the pore such that the one or more helicases and the one or more molecular brakes are brought together and both control the movement of the polynucleotide through the pore;

(c) taking one or more measurements as the polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the polynucleotide and thereby characterising the polynucleotide.

This type of method is discussed in detail in the International application No PCT/GB2014/052737.

The one or more helicases may be any of those discussed above. The one or more molecular brakes may be any compound or molecule which binds to the polynucleotide and slows the movement of the polynucleotide through the pore. The one or more molecular brakes preferably comprise one or more compounds which bind to the polynucleotide. The one or more compounds are preferably one or more macrocycles. Suitable macrocycles include, but are not limited to, cyclodextrins, calixarenes, cyclic peptides, crown ethers, cucurbiturils, pillararenes, derivatives thereof or a combination 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 agent 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 one or more molecular brakes are preferably not one or more single stranded binding proteins (SSB). The one or more molecular brakes are more preferably not a single-stranded binding protein (SSB) comprising a carboxy-terminal (C-terminal) region which does not have a net negative charge or (ii) a modified SSB comprising one or more modifications in its C-terminal region which decreases the net negative charge of the C-terminal region. The one or more molecular brakes are most preferably not any of the SSBs disclosed in International Application No. PCT/GB2013/051924 (published as WO 2014/013259).

The one or more molecular brakes are preferably one or more polynucleotide binding proteins. The polynucleotide binding protein may be any protein that is capable of binding to the polynucleotide and controlling its movement through the pore. It is straightforward in the art to determine whether or not a protein binds to a polynucleotide. The protein typically interacts with and modifies at least one property of the polynucleotide. The protein may modify the polynucleotide by cleaving it to form individual nucleotides or shorter chains of nucleotides, such as di- or trinucleotides. The moiety may modify the polynucleotide by orienting it or moving it to a specific position, i.e. controlling its movement.

The polynucleotide binding protein is preferably derived from a polynucleotide handling enzyme. The one or more molecular brakes may be derived from any of the polynucleotide handling enzymes discussed above. Modified versions of Phi29 polymerase (SEQ ID NO: 8) which act as molecular brakes are disclosed in US Patent No. 5,576,204. The one or more molecular brakes are preferably derived from a helicase.

Any number of molecular brakes derived from a helicase may be used. For instance, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more helicases may be used as molecular brakes. If two or more helicases are be used as molecular brakes, the two or more helicases are typically the same helicase. The two or more helicases may be different helicases.

The two or more helicases may be any combination of the helicases mentioned above. The two or more helicases may be two or more Dda helicases. The two or more helicases may be one or more Dda helicases and one or more TrwC helicases. The two or more helicases may be different variants of the same helicase.

The two or more helicases are preferably attached to one another. The two or more helicases are more preferably covalently attached to one another. The helicases may be attached in any order and using any method. The one or more molecular brakes derived from helicases are preferably modified to reduce the size of an opening in the polynucleotide binding domain through which in at least one conformational state the polynucleotide can unbind from the helicase. This is disclosed in WO 2014/013260.

Preferred helicase constructs for use in the invention are described in International Application Nos. PCT/GB2013/051925 (published as WO 2014/013260); PCT/GB2013/051924 (published as WO 2014/013259); PCT/GB2013/051928 (published as WO 2014/013262) and PCT/GB2014/052736.

### *Spacers*

The one or more helicases may be stalled at one or more spacers as discussed in International Application No. PCT/GB2014/050175. Any configuration of one or more helicases and one or more spacers disclosed in the International Application may be used in this invention.

### *Membrane*

The pore of the invention may be present in a membrane. In the method of the invention, the polynucleotide is typically contacted with the pore of the invention in a membrane. 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 amphiphilic molecules may be synthetic or naturally occurring. 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). Block copolymers are polymeric materials in which two or more monomer sub-units are polymerized together to create a single polymer chain. Block copolymers typically have properties that are contributed by each monomer sub-unit. However, a block copolymer may have unique properties that polymers formed from the individual sub-units do not possess. Block copolymers can be engineered such that one of the monomer sub-units is hydrophobic (i.e. lipophilic), whilst the other sub-unit(s) are hydrophilic whilst in aqueous media. In this case, the block copolymer may possess amphiphilic properties and may form a structure that mimics a biological membrane. The block copolymer may be a diblock (consisting of two monomer sub-units), but may also be constructed from more than two monomer sub-units to form more complex arrangements that behave asamphiphiles. The copolymer may be a triblock, tetrablock or pentablock copolymer. The membrane is preferably a triblock copolymer membrane.

Archaebacterial bipolar tetraether lipids are naturally occurring lipids that are constructed such that the lipid forms a monolayer membrane. These lipids are generally found in extremophiles that survive in harsh biological environments, thermophiles, halophiles and acidophiles. Their stability is believed to derive from the fused nature of the final bilayer. It is straightforward to construct block copolymer materials that mimic these biological entities by creating a triblock polymer that has the general motif hydrophilic-hydrophobic-hydrophilic. This material may form monomeric membranes that behave similarly to lipid bilayers and encompass a range of phase behaviours from vesicles through to laminar membranes. Membranes formed from these triblock copolymers hold several advantages over biological lipid membranes. Because the triblock copolymer is synthesized, the exact construction can be carefully controlled to provide the correct chain lengths and properties required to form membranes and to interact with pores and other proteins.

Block copolymers may also be constructed from sub-units that are not classed as lipid sub-materials; for example a hydrophobic polymer may be made from siloxane or other non-hydrocarbon based monomers. The hydrophilic sub-section of block copolymer can also possess low protein binding properties, which allows the creation of a membrane that is highly resistant when exposed to raw biological samples. This head group unit may also be derived from non-classical lipid head-groups.

Triblock copolymer membranes also have increased mechanical and environmental stability compared with biological lipid membranes, for example a much higher operational temperature or pH range. The synthetic nature of the block copolymers provides a platform to customize polymer based membranes for a wide range of applications.

The membrane is most preferably one of the membranes disclosed in International Application No. PCT/GB2013/052766 (published as WO/2014/06443) or PCT/GB2013/052767(published as WO/2014/064444).

The amphiphilic molecules may be chemically-modified or functionalised to facilitate coupling of the polynucleotide.

The amphiphilic layer may be a monolayer or a bilayer. The amphiphilic layer is typically planar. The amphiphilic layer may be curved. The amphiphilic layer may be supported.

Amphiphilic membranes are typically naturally mobile, essentially acting as two dimensional fluids with lipid diffusion rates of approximately $10^{-8}$ cm s-1. This means that the pore and coupled polynucleotide can typically move within an amphiphilic membrane.

The membrane may be a lipid bilayer. Lipid bilayers are models of cell membranes and serve as excellent platforms for a range of experimental studies. For example, lipid bilayers can be used for *in vitro* investigation of membrane proteins by single-channel recording. Alternatively, lipid bilayers can be used as biosensors to detect the presence of a range of substances. The lipid bilayer may be any lipid bilayer. Suitable lipid bilayers include, but are not limited to, a planar lipid bilayer, a supported bilayer or a liposome. The lipid bilayer is preferably a planar lipid bilayer. Suitable lipid bilayers are disclosed in International Application No. PCT/GB08/000563 (published as WO 2008/102121), International Application No. PCT/GB08/004127 (published as WO 2009/077734) and International Application No. PCT/GB2006/001057 (published as WO 2006/100484).

### *Coupling*

The polynucleotide is preferably coupled to the membrane comprising the pore of the invention. The method may comprise coupling the polynucleotide to the membrane comprising the pore of the invention. The polynucleotide is preferably coupled to the membrane using one or more anchors. The polynucleotide may be coupled to the membrane using any known method.

Each anchor comprises a group which couples (or binds) to the polynucleotide and a group which couples (or binds) to the membrane. Each anchor may covalently couple (or bind) to the polynucleotide and/or the membrane. If a Y adaptor and/or a hairpin loop adaptors are used, the polynucleotide is preferably coupled to the membrane using the adaptor(s).

The polynucleotide may be coupled to the membrane using any number of anchors, such as 2, 3, 4 or more anchors. For instance, a polynucleotide may be coupled to the membrane using two anchors each of which separately couples (or binds) to both the polynucleotide and membrane.

The one or more anchors may comprise the one or more helicases and/or the one or more molecular brakes.

If the membrane is an amphiphilic layer, such as a copolymer membrane or a lipid bilayer, the one or more anchors preferably comprise a polypeptide anchor present in the membrane and/or a hydrophobic anchor present in the membrane. The hydrophobic anchor is preferably a lipid, fatty acid, sterol, carbon nanotube, polypeptide, protein or amino acid, for example cholesterol, palmitate or tocopherol. In preferred embodiments, the one or more anchors are not the pore.

The components of the membrane, such as the amphiphilic molecules, copolymer or lipids, may be chemically-modified or functionalised to form the one or more anchors. Examples of suitable chemical modifications and suitable ways of functionalising the components of the membrane are discussed in more detail below. Any proportion of the membrane components may be functionalized, for example at least 0.01%, at least 0.1%, at least 1%, at least 10%, at least 25%, at least 50% or 100%.

The polynucleotide may be coupled directly to the membrane. The one or more anchors used to couple the polynucleotide to the membrane preferably comprise a linker. The one or more anchors may comprise one or more, such as 2, 3, 4 or more, linkers. One linker may be used to couple more than one, such as 2, 3, 4 or more, polynucleotides to the membrane.

Preferred linkers include, but are not limited to, polymers, such as polynucleotides, polyethylene glycols (PEGs), polysaccharides and polypeptides. These linkers may be linear, branched or circular. For instance, the linker may be a circular polynucleotide. The polynucleotide may hybridise to a complementary sequence on the circular polynucleotide linker.

The coupling may be permanent or stable. In other words, the coupling may be such that the polynucleotide remains coupled to the membrane when interacting with the pore.

The coupling may be transient. In other words, the coupling may be such that the polynucleotide may decouple from the membrane when interacting with the pore.

Coupling of polynucleotides to a linker or to a functionalised membrane can also be achieved by a number of other means provided that a complementary reactive group or an anchoring group can be added to the polynucleotide. The addition of reactive groups to either end of a polynucleotide has been reported previously. The one or more anchors preferably couple the polynucleotide to the membrane via hybridisation. Hybridisation in the one or more anchors allows coupling in a transient manner as discussed above. The one or more anchors may comprise a single stranded or double stranded polynucleotide. One part of the anchor may be ligated to a single stranded or double stranded polynucleotide. Ligation of short pieces of ssDNA have been reported using T4 RNA ligase I (Troutt, A. B., M. G. McHeyzer-Williams, et al. (1992). "Ligation-anchored PCR: a simple amplification technique with single-sided specificity." Proc Natl Acad Sci U S A 89(20): 9823-5). Alternatively, either a single stranded or double stranded polynucleotide can be ligated to a double stranded polynucleotide and then the two strands separated by thermal or chemical denaturation. If the polynucleotide is a synthetic strand, the one or more anchors can be incorporated during the chemical synthesis of the polynucleotide. For instance, the polynucleotide can be synthesised using a primer having a reactive group attached to it.

Ideally, the polynucleotide is coupled to the membrane without having to functionalise the polynucleotide. This can be achieved by coupling the one or more anchors, such as a polynucleotide binding protein or a chemical group, to the membrane and allowing the one or more anchors to interact with the polynucleotide or by functionalizing the membrane. The one or more anchors may be coupled to the membrane by any of the methods described herein. In particular, the one or more anchors may comprise one or more linkers, such as maleimide functionalised linkers.

In this embodiment, the polynucleotide is typically RNA, DNA, PNA, TNA or LNA and may be double or single stranded. This embodiment is particularly suited to genomic DNA polynucleotides.

Where the one or more anchors comprise a protein, they may be able to anchor directly into the membrane without further functonalisation, for example if it already has an external hydrophobic region which is compatible with the membrane. Examples of such proteins include, but are not limited to, transmembrane proteins, intramembrane proteins and membrane proteins.

According to a preferred embodiment, the one or more anchors may be used to couple a polynucleotide to the membrane when the polynucleotide is attached to a leader sequence which preferentially threads into the pore. Leader sequences are discussed in more detail below. Preferably, the polynucleotide is attached (such as ligated) to a leader sequence which preferentially threads into the pore. Such a leader sequence may comprise a homopolymeric polynucleotide or an abasic region. The leader sequence is typically designed to hybridise to the one or more anchors either directly or via one or more intermediate polynucleotides (or splints). In such instances, the one or more anchors typically comprise a polynucleotide sequence which is complementary to a sequence in the leader sequence or a sequence in the one or more intermediate polynucleotides (or splints). In such instances, the one or more splints typically comprise a polynucleotide sequence which is complementary to a sequence in the leader sequence.

*Double stranded polynucleotide*

The polynucleotide may be double stranded. If the polynucleotide is double stranded, the method preferably further comprises before the contacting step ligating a bridging moiety, such as a hairpin loop, to one end of the polynucleotide. The two strands of the polynucleotide may then be separated as or before the polynucleotide is contacted with the pore in accordance with the invention. The two strands may be separated as the polynucleotide movement through the pore is controlled by a polynucleotide binding protein, such as a helicase, or molecular brake. This is described in International Application No. PCT/GB2012/051786 (published asWO 2013/014451).

Linking and interrogating both strands on a double stranded construct in this way increases the efficiency and accuracy of characterization.

*Round the corner sequencing*

In a preferred embodiment, a target double stranded polynucleotide is provided with a bridging moiety, such as a hairpin loop, at one end and the method comprises contacting the polynucleotide with the pore of the invention such that both strands of the polynucleotide move through the pore and taking one or more measurements as the both strands of the polynucleotide move with respect to the pore wherein the measurements are indicative of one or more characteristics of the strands of the polynucleotide and thereby characterising the target double stranded polynucleotide. Any of the embodiments discussed above equally apply to this embodiment.

*Leader sequence*

Before the contacting step, the method preferably comprises attaching to the polynucleotide a leader sequence which preferentially threads into the pore. The leader sequence facilitates the method of the invention. The leader sequence is designed to preferentially thread into the pore of the invention and thereby facilitate the movement of polynucleotide through the pore. The leader sequence can also be used to link the polynucleotide to the one or more anchors as discussed above.

*Double coupling*

The method of the invention may involve double coupling of a double stranded polynucleotide. In a preferred embodiment, the method of the invention comprises:

(a) providing the double stranded polynucleotide with a Y adaptor at one end and a bridging moiety adaptor, such as a hairpin loop adaptor, at the other end, wherein the Y adaptor comprises one or more first anchors for coupling the polynucleotide to the membrane, wherein the bridging moiety adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the bridging moiety adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane;

(b) contacting the polynucleotide provided in step (a) with the pore the invention such that the polynucleotide moves through the pore; and

(c) taking one or more measurements as the polynucleotide moves with respect to the pore, wherein the measurements are indicative of one or more characteristics of the polynucleotide, and thereby characterising the target polynucleotide.

This type of method is discussed in detail in International Application No. PCT/GB2015/050991.

*Adding hairpin loops and leader sequences*

Before provision, a double stranded polynucleotide may be contacted with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors comprising the leader sequence and wherein a proportion of the substrates in the population are hairpin loop adaptors. The transposase fragments the double stranded polynucleotide analyte and ligates MuA substrates to one or both ends of the fragments. This produces a plurality of modified double stranded polynucleotides comprising the leader sequence at one end and the hairpin loop at the other. The modified double stranded polynucleotides may then be investigated using the method of the invention.

These MuA based methods are disclosed in International Application No. PCT/GB2014/052505 (published as WO/2015/022544). They are also discussed in detail in International Application No. PCT/GB2015/050991.

One or more helicases may be attached to the MuA substrate Y adaptors before they are contacted with the double stranded polynucleotide and MuA transposase. Alternatively, one or more helicases may be attached to the MuA substrate Y adaptors before they are contacted with the double stranded polynucleotide and MuA transposase.

One or more molecular brakes may be attached to the MuA substrate hairpin loop adaptors before they are contacted with the double stranded polynucleotide and MuA transposase. Alternatively, one or more molecular brakes may be attached to the MuA substrate hairpin loop adaptors before they are contacted with the double stranded polynucleotide and MuA transposase.

### *Uncoupling*

The method of the invention may involve characterising multiple target polynucleotides and uncoupling of the at least the first target polynucleotide.

In a preferred embodiment, the invention involves characterising two or more target polynucleotides. The method comprises:

(a) providing a first polynucleotide in a first sample;

(b) providing a second polynucleotide in a second sample;

(c) coupling the first polynucleotide in the first sample to a membrane using one or more anchors;

(d) contacting the first polynucleotide with the pore of the invention such that the polynucleotide moves through the pore;

(e) taking one or more measurements as the first polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the first polynucleotide and thereby characterising the first polynucleotide;

(f) uncoupling the first polynucleotide from the membrane;

(g) coupling the second polynucleotide in the second sample to the membrane using one or more anchors;

(h) contacting the second polynucleotide with the pore of the invention such that the second polynucleotide moves through the pore; and

(i) taking one or more measurements as the second polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the second polynucleotide and thereby characterising the second polynucleotide.

This type of method is discussed in detail in International Application No. PCT/GB2015/050992. If one or more anchors comprise a hydrophobic anchor, such as cholesterol, the agent is preferably a cyclodextrin or a derivative thereof or a lipid. 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 agent 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). Any of the lipids disclosed herein may be used.

*Modified polynucleotides*

Before characterisation, a target polynucleotide may be modified by contacting the polynucleotide with a polymerase and a population of free nucleotides under conditions in which the polymerase forms a modified polynucleotide using the target polynucleotide as a template, wherein the polymerase replaces one or more of the nucleotide species in the target polynucleotide with a different nucleotide species when forming the modified polynucleotide. The modified polynucleotide may then be provided with one or more helicases attached to the polynucleotide and one or more molecular brakes attached to the polynucleotide. This type of modification is described in International Application No. PCT/GB2015/050483. Any of the polymerases discussed above may be used. The polymerase is preferably Klenow or 9o North.

The template polynucleotide is contacted with the polymerase under conditions in which the polymerase forms a modified polynucleotide using the template polynucleotide as a template. Such conditions are known in the art. For instance, the polynucleotide is typically contacted with the polymerase in commercially available polymerase buffer, such as buffer from New England Biolabs®. The temperature is preferably from 20 to 37°C for Klenow or from 60 to 75°C for 9o North. A primer or a 3' hairpin is typically used as the nucleation point for polymerase extension.

Characterisation, such as sequencing, of a polynucleotide using a transmembrane pore typically involves analyzing polymer units made up of k nucleotides where k is a positive integer (i.e. 'k-mers'). This is discussed in International Application No. PCT/GB2012/052343 (published as WO 2013/041878). While it is desirable to have clear separation between current measurements for different k-mers, it is common for some of these measurements to overlap. Especially with high numbers of polymer units in the k-mer, i.e. high values of k, it can become difficult to resolve the measurements produced by different k-mers, to the detriment of deriving information about the polynucleotide, for example an estimate of the underlying sequence of the polynucleotide.

By replacing one or more nucleotide species in the target polynucleotide with different nucleotide species in the modified polynucleotide, the modified polynucleotide contains k-mers which differ from those in the target polynucleotide. The different k-mers in the modified polynucleotide are capable of producing different current measurements from the k-mers in the target polynucleotide and so the modified polynucleotide provides different information from the target polynucleotide. The additional information from the modified polynucleotide can make it easier to characterise the target polynucleotide. In some instances, the modified polynucleotide itself may be easier to characterise. For instance, the modified polynucleotide may be designed to include k-mers with an increased separation or a clear separation between their current measurements or k-mers which have a decreased noise.

The polymerase preferably replaces two or more of the nucleotide species in the target polynucleotide with different nucleotide species when forming the modified polynucleotide. The polymerase may replace each of the two or more nucleotide species in the target polynucleotide with a distinct nucleotide species. The polymerase may replace each of the two or more nucleotide species in the target polynucleotide with the same nucleotide species.

If the target polynucleotide is DNA, the different nucleotide species in the modified typically comprises a nucleobase which differs from adenine, guanine, thymine, cytosine or methylcytosine and/or comprises a nucleoside which differs from deoxyadenosine, deoxyguanosine, thymidine, deoxycytidine or deoxymethylcytidine. If the target polynucleotide is RNA, the different nucleotide species in the modified polynucleotide typically comprises a nucleobase which differs from adenine, guanine, uracil, cytosine or methylcytosine and/or comprises a nucleoside which differs from adenosine, guanosine, uridine, cytidine or methylcytidine. The different nucleotide species may be any of the universal nucleotides discussed above.

The polymerase may replace the one or more nucleotide species with a different nucleotide species which comprises a chemical group or atom absent from the one or more nucleotide species. The chemical group may be a propynyl group, a thio group, an oxo group, a methyl group, a hydroxymethyl group, a formyl group, a carboxy group, a carbonyl group, a benzyl group, a propargyl group or a propargylamine group.

The polymerase may replace the one or more nucleotide species with a different nucleotide species which lacks a chemical group or atom present in the one or more nucleotide species. The polymerase may replace the one or more of the nucleotide species with a different nucleotide species having an altered electronegativity. The different nucleotide species having an altered electronegativity preferably comprises a halogen atom.

The method preferably further comprises selectively removing the nucleobases from the one or more different nucleotides species in the modified polynucleotide.

Other characterisation methods

In another embodiment, a polynucleotide is characterised by detecting labelled species that are released as a polymerase incorporates nucleotides into the polynucleotide. The polymerase uses the polynucleotide as a template. Each labelled species is specific for each nucleotide. The polynucleotide is contacted with a pore of the invention, a polymerase and labelled nucleotides such that phosphate labelled species are sequentially released when nucleotides are added to the polynucleotide(s) by the polymerase, wherein the phosphate species contain a label specific for each nucleotide. The polymerase may be any of those discussed above. The phosphate labelled species are detected using the pore and thereby characterising the polynucleotide. This type of method is disclosed in European Application No. 13187149.3 (published as EP 2682460). Any of the embodiments discussed above equally apply to this method.

In another embodiment, the invention also provides a method of characterising a target polynucleotide, comprising

a) contacting the target polynucleotide with a transmembrane pore and a polynucleotide binding protein selected from TatD exonuclease, PyroPhage® 3173 DNA

Polymerase, SD Polymerase and variants thereof such that the protein controls the movement of the polynucleotide with respect to the transmembrane pore; and

c) taking one or more measurements as the polynucleotide moves with respect to the transmembrane pore, wherein the measurements are indicative of one or more characteristics of the polynucleotide, and thereby characterising the target polynucleotide. In this embodiment, the transmembrane pore and/or the polynucleotide binding protein are preferably not modified in accordance with the invention. Any of the embodiments discussed above with reference to characterisation of polynucleotides equally applies to this embodiment. The transmembrane pore may be any of those discussed above.

## Kits

The present invention also provides a kit for characterising a target polynucleotide. In one embodiment, the kit comprises a pore of the invention and the components of a membrane. The membrane is preferably formed from the components. The pore is preferably present in the membrane. The kit may comprise components of any of the membranes disclosed above, such as an amphiphilic layer or a triblock copolymer membrane. The kit may further comprise a polynucleotide binding protein, preferably a modified polynucleotide binding protein of the invention.

In another embodiment, the kit comprises a modified polynucleotide binding protein of the invention and a polynucleotide adaptor. The polynucleotide binding protein is preferably bound to the adaptor. The adaptor may be any of those discussed above.

The kit may further comprise one or more anchors for coupling the polynucleotide to the membrane.

The kit is preferably for characterising a double stranded polynucleotide and preferably comprises a Y adaptor and a hairpin loop adaptor. The Y adaptor preferably has one or more helicases attached and the hairpin loop adaptor preferably has one or more molecular brakes attached. The one or more helicases and/or the one or more molecular brakes may be modified in accordance with the invention. The Y adaptor preferably comprises one or more first anchors for coupling the polynucleotide to the membrane, the hairpin loop adaptor preferably comprises one or more second anchors for coupling the polynucleotide to the membrane and the strength of coupling of the hairpin loop adaptor to the membrane is preferably greater than the strength of coupling of the Y adaptor to the membrane.

The kit 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 polynucleotides 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 for which organism the method may be used.

## Apparatus

The invention also provides an apparatus for characterising a target polynucleotide. The apparatus comprises a plurality of pores of the invention or a plurality of combinations of the invention. The apparatus also comprises a plurality of membranes. The plurality of pores are preferably present in the plurality of membranes. The number of pores and membranes is preferably equal. Preferably, a single pore is present in each membrane.

The apparatus preferably further comprises instructions for carrying out the method of the invention. The apparatus may be any conventional apparatus for polynucleotide analysis, such as an array or a chip. Any of the embodiments discussed above with reference to the methods of the invention are equally applicable to the apparatus of the invention. The apparatus may further comprise any of the features present in the kit of the invention.

The apparatus is preferably set up to carry out the method of the invention.

The apparatus preferably comprises:

a sensor device that is capable of supporting the plurality of pores and membranes and being operable to perform polynucleotide characterisation using the pores and membranes; and

at least one port for delivery of the material for performing the characterisation.

Alternatively, the apparatus preferably comprises:

a sensor device that is capable of supporting the plurality of pores and membranes being operable to perform polynucleotide characterisation using the pores and membranes; and

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

The apparatus more preferably comprises:

a sensor device that is capable of supporting the membrane and plurality of pores and membranes and being operable to perform polynucleotide characterising using the pores and membranes;

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

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

one or more containers for receiving respective samples, the fluidics system being configured to supply the samples selectively from one or more containers to the sensor device.

The apparatus may be any of those described in International Application No. No. PCT/GB08/004127 (published as WO 2009/077734), PCT/GB10/000789 (published as WO 2010/122293), International Application No. PCT/GB10/002206 (published as WO 2011/067559) or International Application No. PCT/US99/25679 (published as WO 00/28312).

The following Example illustrates the invention.

## Example 1

This example describes the simulations which were run to investigate the interaction between MspA - (G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K = MspA mutant 1) or MspA – ((Del-L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119 = MspA mutant 2) with T4 Dda – E94C/A360C/C109A/C136A (SEQ ID NO: 24 with mutations E94C/A360C/C114A/C171A/C421D and then (ΔM1)G1G2)).

Simulations were performed using the GROMACS package version 4.0.5, with the GROMOS 53a6 forcefield and the SPC water model.

The MspA mutant 1 and MspA mutant 2 models were based on the crystal structure of MspA found in the protein data bank, accession code 1UUN. The relevant mutations were made using PyMOL, and in the case of MspA mutant 2 the residues L74/G75/D118/L119 were deleted from the barrel. The resultant pore models were then energy minimised using the steepest descents algorithm. The T4 Dda – E94C/A360C/C109A/C136A model was based on the Dda1993 structure found in the protein data bank, accession code 3UPU. Again, relevant mutations were made using PyMOL, and the model was energy minimised using the steepest descents algorithm.

The T4 Dda – E94C/A360C/C109A/C136A model was then placed above MspA mutant 1 and MspA mutant 2. Three simulations were performed for the T4 Dda – E94C/A360C/C109A/C136A /MspA mutant 1 and T4 Dda – E94C/A360C/C109A/C136A /MspA mutant 2 systems, with the orientation of T4 Dda – E94C/A360C/C109A/C136A differing in each simulation (See Figure 1 for cartoon representations of the three different simulation orientations). The pore was placed into a lipid membrane comprising DPPC molecules and the simulation box was solvated. Throughout the simulation, restraints were applied to the backbone of the pore. However, the enzyme was unrestrained. The system was simulated in the NPT ensemble for 40 ns, using the Berendsen thermostat and Berendsen barostat to 300 K.

The contacts between the enzyme and pore were analysed using both GROMACS analysis software and also locally written code. Figures 2 to 5 showed the amino acid residues which interacted in MspA mutant 1 (Figures 2 and 3) and MspA mutant 2 (Figures 4 and 5) with the enzyme T4 Dda – E94C/A360C/C109A/C136A. The tables below show the number of contacts observed for both pore and enzyme amino acids (Table 4 shows the MspA mutant 1 amino acid contact points observed when the interactions were measured between MspA mutant 1 and T4 Dda – E94C/A360C/C109A/C136A, Table 5 shows the T4 Dda – E94C/A360C/C109A/C136A amino acid contact points observed when the interactions were measured between MspA mutant 1 and T4 Dda – E94C/A360C/C109A/C136A, Table 6 shows the MspA mutant 2 amino acid contact points observed when the interactions were measured between MspA mutant 2 and T4 Dda – E94C/A360C/C109A/C136A, Table 7 shows the T4 Dda – E94C/A360C/C109A/C136A amino acid contact points observed when the interactions were measured between MspA mutant 2 and T4 Dda – E94C/A360C/C109A/C136A). Figure 6 shows which amino acids in the pore (MspA mutant 2) interacted with particular amino acids in the enzyme (T4 Dda – E94C/A360C/C109A/C136A). The data obtained from the simulations showed that a greater number of interaction points were detected between MspA mutant 2 and T4 Dda – E94C/A360C/C109A/C136A than were detected between MspA mutant 1 and T4 Dda – E94C/A360C/C109A/C136A.

| **Run 1** | | **Run 2** | | **Run 3** | |
|---|---|---|---|---|---|
| **Pore Amino Acid Residue** | **Number of Contacts** | **Pore Amino Acid Residue** | **Number of Contacts** | **Pore Amino Acid Residue** | **Number of Contacts** |
| 57 | 5304 | 56 | 5271 | 57 | 2068 |
| 59 | 4806 | 57 | 2262 | 136 | 1800 |
| 136 | 1515 | 136 | 1697 | 59 | 1419 |
| 134 | 1443 | 139 | 1053 | 134 | 975 |
| 56 | 1402 | 52 | 720 | 56 | 817 |
| 54 | 382 | 134 | 215 | 12 | 581 |
| 12 | 263 | 138 | 196 | 139 | 180 |
| 169 | 49 | 55 | 5 | 58 | 87 |
| 14 | 17 | 59 | 1 | 137 | 32 |
| 58 | 8 | | | 14 | 8 |
| 55 | 4 | | | 48 | 5 |
| 52 | 4 | | | 169 | 3 |
| 138 | 2 | | | | |
| 139 | 1 | | | | |
| 137 | 1 | | | | |

**Table 4**

| **Run 1** | | **Run 2** | | **Run 3** | |
|---|---|---|---|---|---|
| **Enzyme Amino Acid Residue** | **Number of Contacts** | **Enzyme Amino Acid Residue** | **Number of Contacts** | **Enzyme Amino Acid Residue** | **Number of Contacts** |
| 2 | 5702 | 180 | 3619 | 255 | 2365 |
| 180 | 3644 | 199 | 2104 | 216 | 2126 |
| 179 | 2205 | 202 | 1909 | 221 | 1027 |
| 178 | 1550 | 1 | 1378 | 227 | 929 |
| 227 | 513 | 4 | 981 | 351 | 239 |
| 4 | 390 | 51 | 678 | 321 | 223 |
| 177 | 297 | 434 | 282 | 254 | 199 |

| | | | | | |
|---|---|---|---|---|---|
| 212 | 275 | 179 | 153 | 258 | 198 |
| 1 | 169 | 178 | 101 | 224 | 177 |
| 194 | 75 | 177 | 84 | 257 | 137 |
| 204 | 58 | 197 | 71 | 256 | 115 |
| 176 | 56 | 5 | 19 | 223 | 109 |
| 213 | 46 | 201 | 19 | 212 | 54 |
| 3 | 37 | 181 | 19 | 308 | 25 |
| 216 | 33 | 200 | 2 | 207 | 21 |
| 211 | 28 | 6 | 1 | 350 | 11 |
| 202 | 28 | | | 228 | 5 |
| 224 | 26 | | | 210 | 4 |
| 223 | 26 | | | 319 | 3 |
| 191 | 17 | | | 304 | 2 |
| 199 | 12 | | | 209 | 2 |
| 201 | 8 | | | 347 | 1 |
| 434 | 4 | | | 261 | 1 |
| 405 | 1 | | | 260 | 1 |
| 255 | 1 | | | 247 | 1 |

**Table 5**

| **Run 1** | | **Run 2** | | **Run 3** | |
|---|---|---|---|---|---|
| **Pore Amino Acid Residue** | **Number of Contacts** | **Pore Amino Acid Residue** | **Number of Contacts** | **Pore Amino Acid Residue** | **Number of Contacts** |
| 59 | 26063 | 59 | 7271 | 56 | 9681 |
| 57 | 10231 | 57 | 4828 | 59 | 7422 |
| 134 | 6034 | 169 | 3039 | 57 | 3640 |
| 136 | 5757 | 134 | 499 | 136 | 3160 |
| 169 | 3357 | 136 | 28 | 12 | 2083 |
| 56 | 1689 | 56 | 17 | 14 | 1132 |
| 137 | 374 | 54 | 1 | 134 | 432 |

| | | | | | |
|---|---|---|---|---|---|
| 58 | 134 | 14 | 1 | 54 | 44 |
| 14 | 10 | 12 | 1 | 169 | 8 |
| 135 | 9 | | | 53 | 2 |
| 60 | 6 | | | | |
| 170 | 5 | | | | |

**Table 6**

| **Run 1** | | **Run 2** | | **Run 3** | |
|---|---|---|---|---|---|
| **Enzyme Amino Acid Residue** | **Number of Contacts** | **Enzyme Amino Acid Residue** | **Number of Contacts** | **Enzyme Amino Acid Residue** | **Number of Contacts** |
| 350 | 7013 | 202 | 8318 | 199 | 4908 |
| 258 | 6277 | 180 | 3505 | 197 | 3828 |
| 223 | 4829 | 179 | 1297 | 185 | 3158 |
| 195 | 4081 | 212 | 1089 | 198 | 2873 |
| 198 | 3990 | 258 | 617 | 207 | 1998 |
| 438 | 3642 | 211 | 324 | 202 | 1645 |
| 260 | 3113 | 198 | 236 | 223 | 1559 |
| 207 | 2781 | 265 | 57 | 180 | 1427 |
| 226 | 2563 | 260 | 55 | 209 | 1309 |
| 304 | 2489 | 259 | 37 | 210 | 1152 |
| 200 | 2116 | 255 | 24 | 203 | 1150 |
| 227 | 1307 | 1 | 22 | 204 | 646 |
| 347 | 845 | 200 | 19 | 437 | 466 |
| 321 | 831 | 300 | 18 | 200 | 431 |
| 422 | 818 | 203 | 14 | 211 | 347 |
| 318 | 740 | 261 | 12 | 405 | 176 |
| 415 | 733 | 216 | 10 | 227 | 97 |
| 210 | 639 | 177 | 10 | 258 | 94 |
| 229 | 555 | 213 | 9 | 212 | 72 |
| 255 | 552 | 207 | 6 | 256 | 68 |
| 224 | 492 | 337 | 2 | 216 | 55 |

| | | | | | |
|---|---|---|---|---|---|
| 228 | 461 | 204 | 2 | 189 | 42 |
| 208 | 395 | 434 | 1 | 228 | 22 |
| 193 | 307 | 298 | 1 | 220 | 18 |
| 256 | 256 | | | 219 | 17 |

**Table 7**

## Example 2

This example describes how a helicase - T4 Dda – E94C/C109A/C136A/A360C (SEQ ID NO: 24 with mutations E94C/C109A/C136A/A360C) was used to control the movement of DNA construct X or Y (shown in Figures 7 and 8) through a number of different MspA nanopores. All of the nanopores tested exhibited changes in current as the DNA translocated through the nanopore. The mutant nanopores tested exhibited either more consistent movement of the target polynucleotide or reduced noise associated with the movement of the target polynucleotide as it translocated through the nanopore or both.

## Materials and Methods

Prior to setting up the experiment, DNA construct X or Y (final concentration 0.1 nM) was pre-incubated at room temperature for five minutes with T4 Dda – E94C/C109A/C136A/A360C (final concentration added to the nanopore system 10 nM, which was provided in buffer (253 mM KCl, 50 mM potassium phosphate, pH 8.0, 2 mM EDTA)). After five minutes, TMAD (100 µM final concentration added to the nanopore system) was added to the pre-mix and the mixture incubated for a further 5 minutes. Finally, MgCl2 (2 mM final concentration added to the nanopore system), ATP (2 mM final concentration added to the nanopore system) and KCl (500 mM final concentration added to the nanopore system) were added to the pre-mix.

Electrical measurements were acquired from single MspA nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide (II), 150mM Potassium Ferricyanide (III), pH 8.0). After achieving a single pore inserted in the block co-polymer, then buffer (2 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide (II), 150mM Potassium Ferricyanide (III), pH 8.0) was flowed through the system to remove any excess MspA nanopores. 150uL of 500mM KCl, 25mM K Phosphate, pH8.0 was then flowed through the system. After 10 minutes a further 150uL of 500mM KCl, 25mM K Phosphate, pH8.0 was flowed through the system and then the enzyme (T4 Dda – E94C/C109A/C136A/A360C, 10 nM final concentration), DNA construct X or Y (0.1 nM final concentration), fuel (MgCl2 2 mM final concentration, ATP 2 mM final concentration) pre-mix (150 µL total) was then flowed into the single nanopore experimental system. The experiment was run at - 120 mV and helicase-controlled DNA movement monitored.

**Results**

A number of different nanopores were investigated in order to determine the effect of mutations to regions of the transmembrane pore which were thought to interact with the helicase T4 Dda – E94C/C109A/C136A/A360C. The mutant pores which were investigated are listed below with the baseline nanopore with which they were compared. A number of different parameters were investigated in order to identify improved nanopores 1) the average noise of the signal (where noise is equal to the standard deviation of all events in a strand, calculated over all strands) which in an improved nanopore would be lower than the baseline, 2) the average current range which was a measure of the spread of current levels within a signal and which in an improved nanopore would be higher than the baseline, 3) the average signal to noise quoted in the table is the signal to noise (average current range divided by average noise of the signal) over all strands and in an improved nanopore would be higher than the baseline and 4) the percentage of complement slipping forwards which in an improved nanopore would be lower than the baseline.

The measurement of complement slipping forwards was calculated using the following procedure 1) the helicase controlled DNA movements were mapped to a model, 2) the helicase-controlled DNA movements were then subjected to filtering, 3) the mapped helicase controlled DNA movements were checked to ensure accurate mapping, 4) the transitions that were classified as a slipping forward movement of at least four consecutive nucleotides were then added together and a percentage based on the total number of transitions was calculated.

In table 8 below, MspA mutant 3 (which contained the additional mutations D56N/E59R) was compared to MspA mutant 1 (baseline). MspA mutant 3 exhibited a lower mean noise of the signal, a higher mean current range and a higher average signal to noise than MspA mutant 1. Therefore, the D56N/E59R mutations which were made to improve the interaction between the nanopore and the enzyme resulted in reduced noise associated with the movement of the target polynucleotide through the nanopore.

In table 8 below, a number of MspA mutants 4-7 and 24 were compared to MspA mutant 1 (baseline). MspA mutants 4-7 and 24 all differed from MspA mutant 1 in that residues had been deleted and that mutations had been made in order to effect how the enzyme and the nanopore interacted (the mutations which the nanopores had in common were L88N/D90N/D91N/Q126R/D134R/E139K). MspA mutants 4-7 and 24 exhibited an improvement in at least one of the measured parameters (mean noise of the signal, median noise of the signal, mean current range, average signal to noise and percentage of complement slipping forwards) when compared to MspA mutant 1. However, the measured improvements were attributed to the combination of changes made to the nanopores (MspA mutants 4-7 and 24) e.g. deletions and the mutations which were made in order to effect how the enzyme and the nanopore interacted.

**Pore ID's**

MspA mutant 1 = MspA - (G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K).

MspA Mutant 3 = MspA – (D56N/E59R/G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K).

MspA mutant 4 = MspA – ((Del-L74/G75/D118/L119)E57R/E59N/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E57R/E59N/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 5 = MspA – ((Del-L74/G75/D118/L119)D56W/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56W/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 9.

MspA mutant 6 = MspA – ((Del-L74/G75/D118/L119)E59Y/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E59Y/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 10.

MspA mutant 7 = MspA – ((Del-L74/G75/D118/L119)D56N/E59S/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59S/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 24 = MspA – ((Del-L74/G75/D118/L119)D56N/E59W/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59W/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 20.

| **Pore ID** | **Mean Noise of thee Signal** | **Median Noise of the Signal** | **Standard Deviation of the Noise of** | **Mean Current Range (pA)** | **Median Current Range (pA)** | **Standard Deviation of the** | **Average Signal to Noise (S2N)** | **Median S2N** | **Standard Deviation of S2N** | **Percentage of complement** |
|---|---|---|---|---|---|---|---|---|---|---|

| | | | **the Signal** | | | **Current Range** | | | | **slipping forwards** |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.35 | 1.35 | 0.14 | 15.48 | 15.78 | 2.23 | 11.44 | | | |
| 3 | 1.31 | 1.24 | 0.26 | 15.64 | 15.60 | 1.76 | 11.94 | | | |
| 4 | 1.40 | 1.22 | 0.45 | 14.74 | 14.64 | 1.88 | 10.56 | | | |
| 5 | 1.35 | 1.28 | 0.26 | 14.38 | 14.26 | 1.79 | 10.69 | 11.95 | 0.33 | 0.25 |
| 6 | 1.36 | 1.32 | 0.23 | 14.96 | 14.89 | 1.72 | 11.00 | 11.89 | 1.17 | 0.27 |
| 7 | 1.31 | 1.27 | 0.23 | 14.96 | 14.91 | 1.73 | 11.39 | 11.26 | 0.29 | 0.43 |
| 24 | 1.38 | 1.32 | 0.26 | 15.16 | 15.09 | 1.85 | 10.97 | | | |

**Table 8**

In table 9 below, a number of MspA mutants 8-19 and 23 were compared to MspA mutant 2 (baseline). MspA mutants 8-19 and 23 all had the same residues deleted (L74/G75/D118/L119) and the following mutations (D90N/D91N/Q126R/D134R/E134K) as MspA mutant 2 but they differed from MspA mutant 2 in the fact that they had been mutated at a range of positions which effected how the enzyme and the nanopore interacted. Of the various parameters which were investigated and measured - mean noise of the signal, mean current range, average signal to noise and percentage of complement slipping forwards MspA mutants 8-19 and 23 exhibited an improvement in at least one of these parameters when compared to the baseline nanopore MspA mutant 2. Therefore, the mutations which were made to improve the interaction between the nanopore and the enzyme resulted either in reduced noise associated with the movement of the target polynucleotide through the pore or more consistent movement of the target through the pore.

In table 9 below, a number of MspA mutants 20-22 were compared to MspA mutant 2 (baseline). MspA mutants 20-22 all differed from MspA mutant 2 in the residues which had been deleted and the mutations which were made in order to effect how the enzyme and the nanopore interacted (the mutations which the nanopores had in common were L88N/D90N/D91N/Q126R/D134R/E139K). MspA mutants 20-22 exhibited an improvement in at least one of the measured parameters (mean noise of the signal, mean current range, average signal to noise and percentage of complement slipping forwards) when compared to MspA mutant 2. However, the measured improvements in noise and movement consistency were attributed to the combination of changes made to the nanopores (MspA mutants 20-22) e.g. deletions and the mutations which were made in order to effect how the enzyme and the nanopore interacted.

**Pore ID's**

MspA mutant 2 = MspA – ((Del-L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 11.

MspA mutant 8 = MspA – ((Del-L74/G75/D118/L119)D56N/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 9 = MspA – ((Del-L74/G75/D118/L119)E59N/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E59N/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 10 = MspA – ((Del-L74/G75/D118/L119)D56N/E57N/E59N/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E57N/E59N/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 11 = MspA – ((Del-L74/G75/D118/L119)E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 12 = MspA – ((Del-L74/G75/D118/L119)D56Y/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations )D56Y/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 12.

MspA mutant 13 = MspA – ((Del-L74/G75/D118/L119)D56N/E57D/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E57D/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 13.

MspA mutant 14 = MspA – ((Del-L74/G75/D118/L119)D56N/E59T/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59T/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 14.

MspA mutant 15 = MspA – ((Del-L74/G75/D118/L119)D56N/E59Q/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59Q/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 15.

MspA mutant 16 = MspA – ((Del-L74/G75/D118/L119)E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 16.

MspA mutant 17 = MspA – ((Del-L74/G75/D118/L119)D56N/E59R/L88N/D90N/D91N/Q126R/D134N/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/L88N/D90N/D91N/Q126R/D134N/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 19.

MspA mutant 18 = MspA – ((Del-L74/G75/D118/L119)D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 17.

MspA mutant 19 = MspA – ((Del-L74/G75/D118/L119)D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119). Example helicase controlled DNA movement shown in Figure 18.

MspA mutant 20 = MspA – ((Del-F80/S81/G112/V113)D56N/E59R/G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K and deletion of the amino acids F80/S81/G112/V113)

MspA mutant 21 = MspA – ((Del-G75/V76/A117/D118)D56N/E59R/G77S/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/G77S/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids G75/V76/A117/D118)

MspA mutant 22 = MspA – ((Del-N79/F80/V113/V114)D56N/E59R/G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59R/G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K and deletion of the amino acids N79/F80/V113/V114)

MspA mutant 23 = MspA – ((Del-L74/G75/D118/L119)D56N/E59Y/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59Y/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119).

| **Pore ID** | **Mean Noise of the Signal** | **Median Noise of the Signal** | **Standard Deviation of the Noise of the Signal** | **Mean Current Range (pA)** | **Median Current Range (pA)** | **Standard Deviation of the Current Range (pA)** | **Average Signal to Noise (S2N)** | **Median S2N** | **Standard Deviation of S2N** | **Percentage of complement slipping forwards** |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.22 | 1.17 | 0.20 | 15.31 | 15.32 | 1.66 | 12.50 | | | 0.49 |
| 8 | 1.20 | 1.17 | 0.15 | 14.74 | 14.62 | 1.66 | 12.28 | | | |
| 9 | 1.11 | 1.08 | 0.19 | 14.74 | 14.71 | 2.36 | 13.28 | | | |
| 10 | 1.46 | 1.35 | 0.36 | 19.02 | 19.16 | 2.07 | 13.03 | | | |
| 11 | 1.26 | 1.22 | 0.24 | 16.08 | 16.16 | 1.84 | 12.76 | | | |
| 12 | 1.24 | 1.19 | 0.22 | 15.93 | 15.92 | 1.83 | 12.88 | 13.34 | 1.12 | 0.60 |
| 13 | 1.19 | 1.16 | 0.19 | 15.45 | 15.36 | 1.77 | 12.95 | 11.95 | 0.33 | 0.70 |
| 14 | 1.31 | 1.26 | 0.24 | 16.04 | 16.09 | 1.87 | 12.29 | 12.61 | 0.87 | 0.80 |
| 15 | 1.22 | 1.16 | 0.25 | 15.55 | 15.53 | 1.72 | 12.76 | 13.33 | 0.86 | 0.55 |
| 16 | 1.37 | 1.33 | 0.22 | 15.68 | 15.68 | 1.81 | 11.44 | 11.79 | 0.64 | 0.30 |
| 17 | 1.31 | 1.27 | 0.21 | 16.04 | 16.05 | 1.93 | 12.25 | 12.03 | 0.57 | 0.65 |
| 18 | 1.24 | 1.17 | 0.27 | 15.77 | 15.66 | 1.86 | 12.76 | 12.76 | 0.12 | 0.25 |
| 19 | 1.33 | 1.28 | 0.24 | 15.61 | 15.60 | 1.92 | 11.74 | 12.29 | 0.59 | 0.38 |
| 20 | 1.32 | 1.22 | 0.32 | 17.11 | 17.15 | 2.11 | 12.96 | | | |
| 21 | 1.27 | 1.21 | 0.24 | 15.81 | 15.84 | 1.69 | 12.48 | | | |
| 22 | 1.21 | 1.13 | 0.28 | 15.82 | 15.82 | 1.90 | 13.13 | | | |
| 23 | 1.48 | 1.43 | 0.24 | 15.90 | 15.93 | 1.92 | 10.72 | 10.37 | 0.66 | 0.30 |

**Table 9**

## Example 3

This example describes the simulations which were run to investigate the interaction between α- hemolysin-(E111N/K147N)8 (SEQ ID NO: 4) with Phi29 DNA polymerase-(D12A/D66A) (SEQ ID NO: 9 with mutations D12A/D66A).

Simulations were performed using the GROMACS package version 4.0.5, with the GROMOS 53a6 forcefield and the SPC water model.

The αHL-(E111N/K147N)8 model was based on the crystal structure of αHL wild-type found in the protein data bank, accession code 7AHL. The relevant mutations were made using PyMOL and the resultant pore model was then energy minimised using the steepest descents algorithm. The Phi29 DNA polymerase-(D12A/D66A) (SEQ ID NO: 9 with mutations D12A/D66A) model was based on the crystal structure of Phi29 DNA polymerase-(D12A/D66A) found in the protein data bank, accession code 2PYL.

The Phi29 DNA polymerase-(D12A/D66A) model was then placed above αHL-(E111N/K147N)8. Three simulations were performed for the Phi29 DNA polymerase-(D12A/D66A)/ αHL-(E111N/K147N)8 system, with the orientation of Phi29 DNA polymerase-(D12A/D66A) differing in each simulation (See Figure 21 for cartoon representations of the three different simulation orientations). The pore was placed into a lipid membrane comprising DPPC molecules and the simulation box was solvated. Throughout the simulation, restraints were applied to the backbone of the pore. However, the enzyme was unrestrained. The system was simulated in the NPT ensemble for 40 ns, using the Berendsen thermostat and Berendsen barostat to 300 K.

The contacts between the enzyme and pore were analysed using both GROMACS analysis software and also locally written code. Figures 22 and 23 show the amino acid residues which interacted in αHL -(E111N/K147N)8 with Phi29 DNA polymerase-(D12A/D66A). The tables below show the number of contacts observed for both pore and enzyme amino acids (Table 10 shows the αHL -(E111N/K147N)8 amino acid contact points observed when the interactions were measured between αHL -(E111N/K147N)8 and Phi29 DNA polymerase-(D12A/D66A), Table 11 shows the Phi29 DNA polymerase-(D12A/D66A) amino acid contact points observed when the interactions were measured between αHL -(E111N/K147N)8 and Phi29 DNA polymerase-(D12A/D66A). Table 10 shows all the amino acids residues in αHL-

E111N/K147N)8 that made more than 100 contacts with Phi29 DNA polymerase-(D12A/D66A) and Table 11 shows all the amino acid residues in Phi29 DNA polymerase-(D12A/D66A) that made more than 100 contacts with αHL-(E111N/K147N)8. Figures 24-28 show which amino acids in the pore (αHL -(E111N/K147N)8 ) interacted with particular amino acids in the enzyme (Phi29 DNA polymerase-(D12A/D66A) in runs 1-3.

| **Run 1** | | **Run 2** | | **Run 3** | |
|---|---|---|---|---|---|
| **Pore Amino Acid Residue** | **Number of Contacts** | **Pore Amino Acid Residue** | **Number of Contacts** | **Pore Amino Acid Residue** | **Number of Contacts** |
| S239 | 3802 | T19 | 5932 | S239 | 6420 |
| N93 | 2785 | N17 | 5125 | E287 | 4255 |
| K240 | 1217 | K240 | 2679 | K237 | 2618 |
| Q242 | 1159 | K46 | 2481 | R236 | 2568 |
| K237 | 958 | E287 | 2243 | A238 | 1796 |
| K288 | 706 | Q241 | 1802 | K240 | 1791 |
| E287 | 337 | N47 | 1496 | Q242 | 1032 |
| D285 | 107 | T18 | 535 | N293 | 967 |
| | | S239 | 442 | K288 | 775 |
| | | K21 | 398 | Q241 | 609 |
| | | K288 | 361 | R281 | 256 |
| | | S16 | 163 | K283 | 118 |
| | | K237 | 136 | N17 | 112 |

**Table 10**

| **Run 1** | | **Run 2** | | **Run 3** | |
|---|---|---|---|---|---|
| **Enzyme Amino Acid Residue** | **Number of Contacts** | **Enzyme Amino Acid Residue** | **Number of Contacts** | **Enzyme Acid Residue** | **Number of Contacts** |
| E322 | 4107 | E272 | 6820 | R308 | 9172 |
| F309 | 1675 | E267 | 3135 | E272 | 5732 |
| G321 | 1642 | S215 | 2941 | S307 | 2304 |

| | | | | | |
|---|---|---|---|---|---|
| R289 | 798 | E221 | 1778 | F309 | 1409 |
| G320 | 764 | D84 | 1738 | R289 | 1064 |
| E241 | 697 | L216 | 1346 | E221 | 699 |
| R236 | 528 | K209 | 1126 | E296 | 643 |
| K240 | 420 | K80 | 1048 | Y224 | 511 |
| G323 | 359 | E419 | 788 | E322 | 363 |
| R308 | 180 | K205 | 633 | E293 | 308 |
| | | E418 | 463 | H287 | 275 |
| | | V270 | 383 | W327 | 166 |
| | | G85 | 213 | K220 | 148 |
| | | R415 | 177 | S349 | 129 |
| | | W81 | 167 | E418 | 113 |
| | | D278 | 147 | Y310 | 108 |
| | | S82 | 122 | | |
| | | K206 | 107 | | |

**Table 11**

## Example 4

This example describes how a number of different helicases were used to control the movement of DNA construct X (see Figure 7) through a number of different MspA nanopores. All of the nanopores tested exhibited changes in current as the DNA translocated through the nanopore. This example investigates the number of slips forward per kilobase and the % bases missed in construct X for a number of pore/enzyme combinations. The helicases investigated in the example moved along the polynucleotide in a 5’ to 3’ direction. When the 5’end of the polynucleotide (the end away from which the helicase moves) was captured by the pore, the helicase worked with the direction of the field resulting from the applied potential and moved the threaded polynucleotide into the pore and into the *trans* chamber. In this Example, slipping forward involved the DNA moving forwards relative to the the pore (i.e. towards its 3’ and away from it 5’ end) at least 4 consecutive nucleotides.

## Materials and Methods

Prior to setting up the experiment, DNA construct X (final concentration 0.1 nM) was pre-incubated at room temperature for five minutes with the appropriate enzyme (either T4 Dda – E94C/C109A/C136A/A360C or T4 Dda – E94C/C109A/C136A/K199L/A360C (final concentration added to the nanopore system 10 nM, which was provided in buffer (253 mM KCl, 50 mM potassium phosphate, pH 8.0, 2 mM EDTA)). After five minutes, TMAD (100 µM final concentration added to the nanopore system) was added to the pre-mix and the mixture incubated for a further 5 minutes. Finally, MgCl2 (2 mM final concentration added to the nanopore system), ATP (2 mM final concentration added to the nanopore system) and KCl (500 mM final concentration added to the nanopore system) were added to the pre-mix.

Electrical measurements were acquired from single MspA nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide (II), 150mM Potassium Ferricyanide (III), pH 8.0) as described in Example 2. The appropriate MspA nanopore was selected from the list below (MspA mutants 2, 19, 25, 26 or 27).

## **Results**

A number of different nanopores/enzyme combinations were investigated in order to determine the affect of mutations to regions of the transmembrane pore and enzyme which were thought to interact with each other. These mutation positions were identified in the molecular modeling experiment described in Example 1. Two different parameters were investigated in order to identify pore and enzyme combinations which exhibited improved helicase controlled DNA translocation 1) the number of slips forward per kilobase and 2) the % bases missed in construct X.

The measurement of slips forward per kilobase was calculated using the following procedure 1) the helicase controlled DNA movements were mapped to a model, 2) the helicase-controlled DNA movements were then subjected to filtering, 3) the mapped helicase controlled DNA movements were checked to ensure accurate mapping, 4) the transitions that were classified as a slipping forward movement of at least four consecutive nucleotides were determined per kilobase. The % bases missed in construct X is a measure of the number of bases in construct X which are missed as a result of slips forward along DNA construct X expressed as a percentage.

Table 12 below shows the different pore and enzyme combinations tested, the corresponding figure number which shows a number of example current traces when the helicase controlled the movement of construct X through the nanopore and the appropriate column reference for figure 36 which shows the data relating to 1) the number of slips forward per kilobase and 2) the % bases missed in construct X. All of the pore/enzyme combinations show less than 5 slips forward per kilobase and less than 12% bases missed in construct X. However, the combination of MspA mutant 26 with T4 Dda – E94C/C109A/C136A/K199L/A360C produced the lowest slips forward per kilobase and the lowest % bases missed in construct X. Therefore, this was a particularly preferred combination which was predicted from the modeling experiment in Example 1 to produce a pore and enzyme with a particularly favourable interaction and more consistent movement.

**Pore ID's**

MspA mutant 25 = MspA - ((Del-L74/G75/D118/L119)D56L/E59L/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56L/E59L/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119)

MspA mutant 26 = MspA – ((Del-L74/G75/D118/L119)G1A/D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations G1A/D56N/E59F/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119).

MspA mutant 27 = MspA – ((Del-L74/G75/D118/L119)D56N/E59Y/L88N/D90N/D91N/Q126R/D134R/E139K)8 (SEQ ID NO: 2 with mutations D56N/E59Y/L88N/D90N/D91N/Q126R/D134R/E139K and deletion of the amino acids L74/G75/D118/L119).

| **Nanopore** | **Enzyme** | **Figure Showing Example Current Trace** | **Column in Figure 36 Corresponding to this Combination** |
|---|---|---|---|
| MspA Mutant 2 | T4 Dda – E94C/C109A/C136A/A360C | 29 | 1 |
| MspA Mutant 19 | T4 Dda – E94C/C109A/C136A/A360C | 30 | 2 |
| MspA Mutant 19 | T4 Dda – E94C/C109A/C136A/K199L/A360C | 31 | 3 |
| MspA Mutant 25 | T4 Dda – E94C/C109A/C136A/A360C | 32 | 4 |
| MspA Mutant 26 | T4 Dda – E94C/C109A/C136A/A360C | 33 | 5 |
| MspA Mutant 26 | T4 Dda – E94C/C109A/C136A/K199L/A360C | 34 | 6 |
| MspA Mutant 27 | T4 Dda – E94C/C109A/C136A/A360C | 35 | 7 |

**Table 12**

## CLAIMS

1. A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein, comprising modifying a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore and thereby improving the movement of the target polynucleotide with respect to the transmembrane pore.

2. A method according to claim 1, wherein the method provides more consistent movement of the target polynucleotide.

3. A method according to claim 2, wherein the method reduces the noise associated with the movement of the target polynucleotide.

4. A method according to any one of claims 1 to 3, wherein the target polynucleotide is double stranded and wherein the method reduces the noise associated with the movement of the complement strand to a greater degree than it reduces the noise associated with the movement of the template strand and/or the method increases the consistency of the movement of the complement strand to a greater degree than it increases the consistency of the movement of the template strand.

5. A method according to any one of the preceding claims, wherein the method further comprises contacting the transmembrane pore and the polynucleotide binding protein with the target polynucleotide such that the protein controls the movement of the polynucleotide with respect to the transmembrane pore.

6. A method according to any one of the preceding claims, wherein the method comprises making one or more modifications to the surface of the transmembrane pore which interacts with the polynucleotide binding protein and/or to the surface of the polynucleotide binding protein which interacts with the transmembrane pore.

7. A method according to any one of the preceding claims, wherein the method comprises making one or more modifications to the entrance of the transmembrane pore which interacts with the polynucleotide binding protein.

8. A method according to any one of the preceding claims, wherein the method comprises making one or more modifications which alter the charge, sterics, hydrogen bonding, π stacking or structure of the part of the transmembrane pore which interacts with the polynucleotide binding protein and/or the part of the polynucleotide binding protein which interacts with the transmembrane pore.

9. A method according to claim 8, wherein the method comprises making one or more modifications which decrease the net negative charge of the part of the transmembrane pore which interacts with the polynucleotide binding protein.

10. A method according to claim 9, wherein the one or more modifications are one or more deletions of negatively charged amino acids or one or more substitutions of negatively charged amino acids with one or more positively charged, uncharged, non-polar and/or aromatic amino acids.

11. A method according to any one of the preceding claims, wherein the method comprises modifying a transmembrane pore which comprises seven or more monomers comprising the sequence shown in SEQ ID NO: 2 or a variant thereof.

12. A method according to claim 11, wherein the part of the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acids at positions

(a) 12, 14, 48, 52, 53, 54, 55, 56, 57, 58, 59, 60, 134, 135, 136, 137, 138, 139, 169 and170 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof;

(b) 12, 14, 52, 54, 56, 57, 59, 134, 136, 138, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof;

(c) 12, 14, 56, 57, 59, 134, 136, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof;

(d) 56, 57, 59, 134, 136, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof; or

(e) 56, 57, 59, 134 and 139 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof.

13. A method according to claim 12, wherein the method comprises modifying one or more of the seven or more monomers so they do not comprise aspartic acid (D) or glutamic acid (E) at one or more of positions 56, 57, 59, 134 and 139 of SEQ ID NO: 2 or at one or more of the corresponding positions in the variant thereof.

14. A method according to claim 13, the method comprises modifying one or more of the seven or more monomers so they comprise one or more of (a) D56N or D56R, (b) E57N or E57R, (c) E59N or E59R, (d) D134N or D134R and (e) E139N, E139R or E139K.

15. A method according to any one of claims 11 to 14, wherein the variant of SEQ ID NO: 2 comprises D90N, D91N, D118R, D134R and E139K and optionally D93N or comprises (a) L88N, D90N, D91N, D93N, D118R, D134R and E139K, (b) G75S, G77S, L88N, D90N, D91N, D93N, D118R, Q126R, D134R and E139K, or (c) G75S, G77S, L88N, D90N, D91N, D118R, Q126R, D134R and E139K.

16. A method according to any one of claims 11 to 14, wherein in the variant of SEQ ID NO: 2 (a) 2, 4, 6, 8 or 10 of the amino acids at positions 72 to 82 of SEQ ID NO: 2 have been deleted and (b) 2, 4, 6, 8 or 10 of the amino acids at positions 111 to 121 of SEQ ID NO: 2 have been deleted.

17. A method according to any one of the claims 11 to 16, wherein the modified transmembrane pore does not comprise one or more monomers which are variants of SEQ ID NO: 2 comprising or consisting of E59R, D90N, D91N, D93N, D118R, D134R and E139K.

18. A method according to any one of claims 1 to 10, wherein the method comprises modifying a transmembrane pore which comprises seven monomers comprising the sequence shown in SEQ ID NO: 4 or a variant thereof.

19. A method according to claim 18, wherein the part of the the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acids at positions

(a) 16, 17, 18, 19, 21, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 281, 283, 285, 287, 288 and 293 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof;

(b) 17, 18, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 287, 288 and 293 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof;

(c) 17, 19, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242 and 287 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof; or

(d) 17, 19, 46, 93, 236, 237, 239, 240, 287 and 288 in SEQ ID NO: 4 or at the corresponding positions in the variant thereof.

20. A method according to any one of claims 1 to 10, wherein the method comprises modifying a transmembrane pore which comprises at least one monomer comprising the sequence shown in SEQ ID NO: 36 or a variant thereof.

21. A method according to claim 20, wherein the part of the the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acids at positions 31, 33, 108, 109, 110 and 138 in SEQ ID NO: 36 or at the corresponding positions in the variant thereof.

22. A method according to any one of the preceding claims, wherein the method comprises modifying a polynucleotide binding protein which comprises the sequence shown in SEQ ID NO: 24 or a variant thereof.

23. A method according to claim 22, wherein the part of the polynucleotide binding protein which interacts with the transmembrane pore comprises the amino acids at positions

(a) 1, 2, 3, 4, 5, 6, 51, 176, 177, 178, 179, 180, 181, 185, 189, 191, 193, 194, 195, 197, 198, 199, 200, 201, 202, 203, 204, 207, 208, 209, 210, 211, 212, 213, 216, 219, 220, 221, 223, 224, 226, 227, 228, 229, 247, 254, 255, 256, 257, 258, 259, 260, 261, 298, 300, 304, 308, 318, 319, 321, 337, 347, 350, 351, 405, 415, 422, 434, 437, 438 in SEQ ID NO: 24 or at the corresponding positions in the variant thereof;

(b) positions 1, 2, 4, 51, 177, 178, 179, 180, 185, 193, 195, 197, 198, 199, 200, 202, 203, 204, 207, 208, 209, 210, 211, 212, 216, 221, 223, 224, 226, 227, 228, 229, 254, 255, 256, 257, 258, 260, 304, 318, 321, 347, 350, 351, 405, 415, 422, 434, 437 and 438 in SEQ ID NO: 24 or at the corresponding positions in the variant thereof; or

(c) positions 1, 2, 178, 179, 180, 185, 195, 197, 198, 199, 200, 202, 203, 207, 209, 210, 212, 216, 221, 223, 226, 227, 255, 258, 260, 304, 350 and 438 in SEQ ID NO: 24 or at the corresponding positions in the variant thereof.

24. A method according to claims 22 or 23, wherein the variant of SEQ ID NO: 24 comprises (a) E94C and A360C or (b) E94C, A360C, C109A and C136A.

25. A method according to any one of claims 1 to 21, wherein the method comprises modifying a polynucleotide binding protein which comprises the sequence shown in SEQ ID NO: 9 or a variant thereof.

26. A method according to claim 25, wherein the part of the polynucleotide binding protein which interacts with the transmembrane pore comprises the amino acids at positions

(a) 80, 81, 82, 84, 85, 205, 206, 209, 215, 216, 220, 221, 224, 236, 240, 241, 267, 270, 272, 278, 287, 289, 293, 296, 307, 308, 309, 310, 320, 321, 322, 323, 327, 349, 415, 418 and 419 in SEQ ID NO: 9 or at the corresponding positions in the variant thereof;

(b) positions 80, 84, 205, 209, 215, 216, 221, 224, 236, 241, 267, 272, 289, 296, 307, 308, 309, 320, 321, 322, and 419 in SEQ ID NO: 9 or at the corresponding positions in the variant thereof;

(c) 80, 84, 209, 215, 216, 221, 267, 272, 289, 307, 308, 309, 321 and 322 in SEQ ID NO: 9 or at the corresponding positions in the variant thereof; or

(d) positions 215, 267, 272, 307, 308 and 322 in SEQ ID NO: 9 or at the corresponding positions in the variant thereof.

27. A method of moving a target polynucleotide with respect to a transmembrane pore using a polynucleotide binding protein, comprising

a) providing a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified; and

b) contacting the transmembrane pore and polynucleotide binding protein provided in a) with the target polynucleotide such that the protein controls the movement of the polynucleotide with respect to the transmembrane pore.

28. A method of characterising a target polynucleotide, comprising:

a) providing a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified;

b) contacting the transmembrane pore and polynucleotide binding protein provided in (a) with the target polynucleotide such that the protein controls the movement of the polynucleotide with respect to the transmembrane pore; and

c) taking one or more measurements as the polynucleotide moves with respect to the transmembrane pore, wherein the measurements are indicative of one or more characteristics of the polynucleotide, and thereby characterising the target polynucleotide.

29. A method according to claim 27 or 28, wherein the transmembrane pore and/or the polynucleotide binding protein has been modified as defined in any one of claims 6 to 26.

30. A transmembrane pore in which a part of the transmembrane pore which interacts with a polynucleotide binding protein has been modified.

31. A transmembrane pore according to claim 30, wherein the transmembrane pore is modified as defined in any one of claims 11 to 21.

32. A transmembrane pore according to claim 30 or 31, wherein the transmembrane pore comprises seven or more monomers comprising the sequence shown in SEQ ID NO: 2 or a variant thereof and one or more of the seven or more monomers comprises a variant of SEQ ID NO: 2 which comprises one or more of (a) D56N, D56R, D56F, D56Y or D56L, (b) E57N or E57R, (c) E59N, E59R, E59F, E59Y or E59L, (d) D134N or D134R and (e) E139N, E139R or E139K.

33. A transmembrane pore according to claim 32, wherein one or more of the seven or more monomers are not variants of SEQ ID NO: 2 which comprise or consist of E59R, D90N, D91N, D93N, D118R, D134R and E139K.

34. A transmembrane pore according to claim 30 or 31, wherein the transmembrane pore comprises

(a) one or more monomers comprising a variant of SEQ ID NO: 36 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions (i) 31, (ii) 33, (iii) 108, (iv) 109, (v) 110 and (vi) 138 in SEQ ID NO: 36 or at one or more of the corresponding positions; or

(b) seven monomers comprising SEQ ID NO: 4 or a variant thereof in which one or more of the monomers, such as 1, 2, 3, 4, 5, 6 or 7 of the monomers, is a variant of SEQ ID NO: 4 which comprises phenylalanine (F), tryptophan (W), isoleucine (I), leucine (L), valine (V), alanine (A), arginine (R), lysine (K), aspartic acid (D), glutamic acid (E) or tyrosine (Y) at one or more of positions 16, 17, 18, 19, 21, 46, 47, 93, 236, 237, 238, 239, 240, 241, 242, 281, 283, 285, 287, 288 and 293 in SEQ ID NO: 4 or at one or more corresponding positions.

35. A mutant monomer comprising a variant of SEQ ID NO: 2, 4 or 36 in which a part of the monomer which interacts with a polynucleotide binding protein has been modified.

36. A mutant monomer according to claim 30, wherein the part of the monomer which interacts with a polynucleotide binding protein is defined in claim 12, 19 or 21 and/or the monomer is modified as defined in claim 32 or 34.

37. A construct comprising two or more covalently attached monomers, wherein at least one of the monomers is a mutant monomer according to claim 35 or 36.

38. A construct according to claim 36, wherein the two or more mutant monomers are genetically fused.

39. A homo-oligomeric pore comprising identical mutant monomers according to claim 35 or 36 or identical constructs according to claim 37 or 38.

40. A hetero-oligomeric pore comprising at least one mutant monomer according to claim 35 or 36 or at least one construct according claim 37 or 38.

41. A polynucleotide binding protein in which a part of the protein which interacts with a transmembrane pore has been modified.

42. A polynucleotide binding protein according to claim 41, wherein the polynucleotide binding protein comprises a variant of the sequence shown in SEQ ID NO: 9 or 24.

43. A polynucleotide binding protein according to claim 42, wherein the part of the polynucleotide binding protein which interacts with a transmembrane pore is defined in claim 23 or 26.

44. A polynucleotide encoding a transmembrane pore according to any one of claims 30 to 34, 39 or 40, a mutant monomer according to any one of claims 35 to 36, a construct according to claim 38 or a polynucleotide binding protein according to any one of claims 41 to 543.

45. A combination of a transmembrane pore and a polynucleotide binding protein in which a part of the transmembrane pore which interacts with the polynucleotide binding protein and/or a part of the polynucleotide binding protein which interacts with the transmembrane pore has been modified.

46. A combination according to claim 45, wherein the combination comprises a transmembrane pore according to any one of claim 31 to 34, 39 or 40 and/or a polynucleotide binding protein according to claim 42 or 43.

47. A kit for characterising a target polynucleotide comprising (a) a transmembrane pore according to any one of claims 30 to 34, 39 or 40 and (b) the components of a membrane.

48. A kit for characterising a target polynucleotide comprising (a) a polynucleotide binding protein according to any one of claims 41 to 43 and (b) a polynucleotide adaptor to which the polynucleotide binding protein is optionally bound.

49. A kit according to claim 48, wherein the kit further comprises means to couple a target polynucleotide to a membrane.

50. An apparatus for characterising target polynucleotides in a sample, comprising (a) a plurality of transmembrane pores according to any one of claims 35 to 40 or a plurality of combinations according to claim 53 or 54 and (b) a plurality of membranes.

51. A method of characterising a target polynucleotide, comprising:

a) providing a transmembrane pore and a polymerase in which a part of the transmembrane pore which interacts with the polymerase and/or a part of the polymerase which interacts with the transmembrane pore has been modified;

b) contacting the target polynucleotide with the transmembrane pore and polymerase provided in a) and labelled nucleotides such that phosphate labelled species are sequentially added to the target polynucleotide by the polymerase, wherein the phosphate species contain a label specific for each nucleotide; and

c) detecting the phosphate labelled species using the transmembrane pore and thereby characterising the polynucleotide.

Figure 1

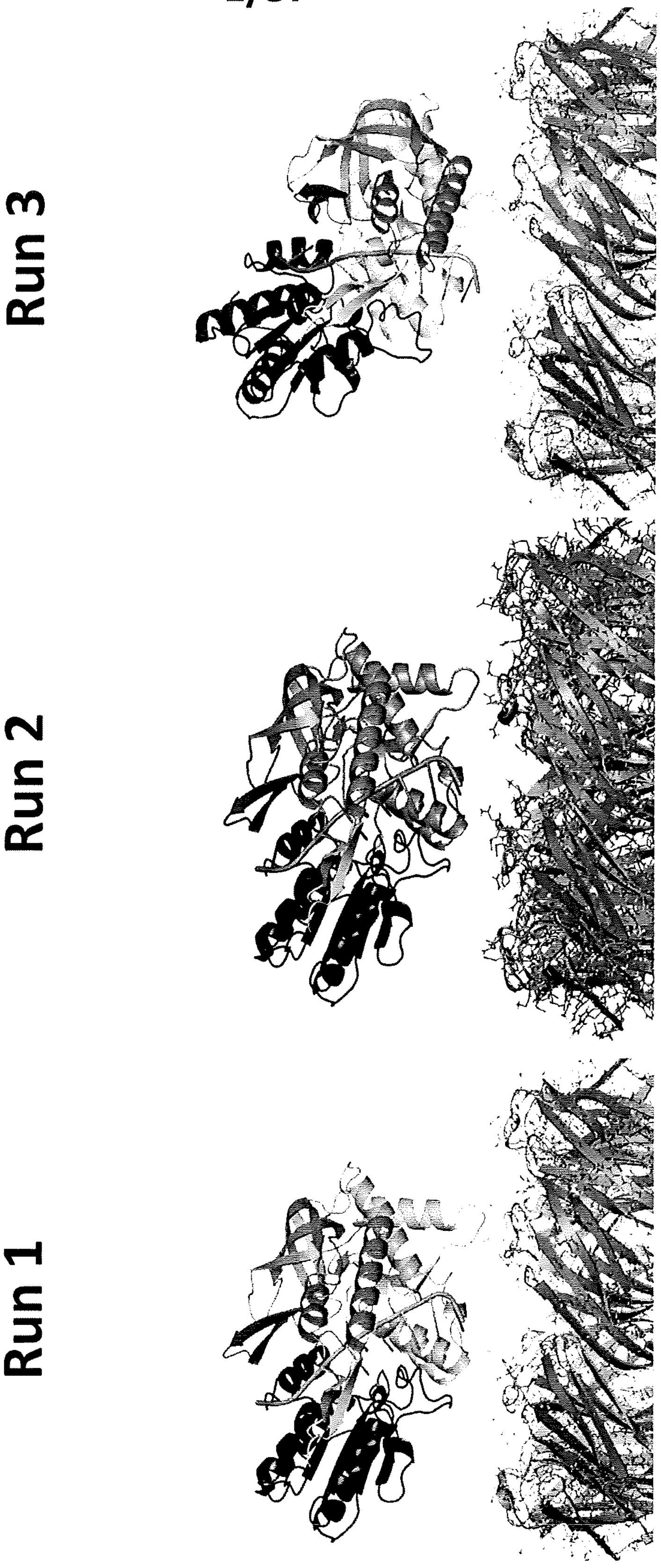


**Figure 2**

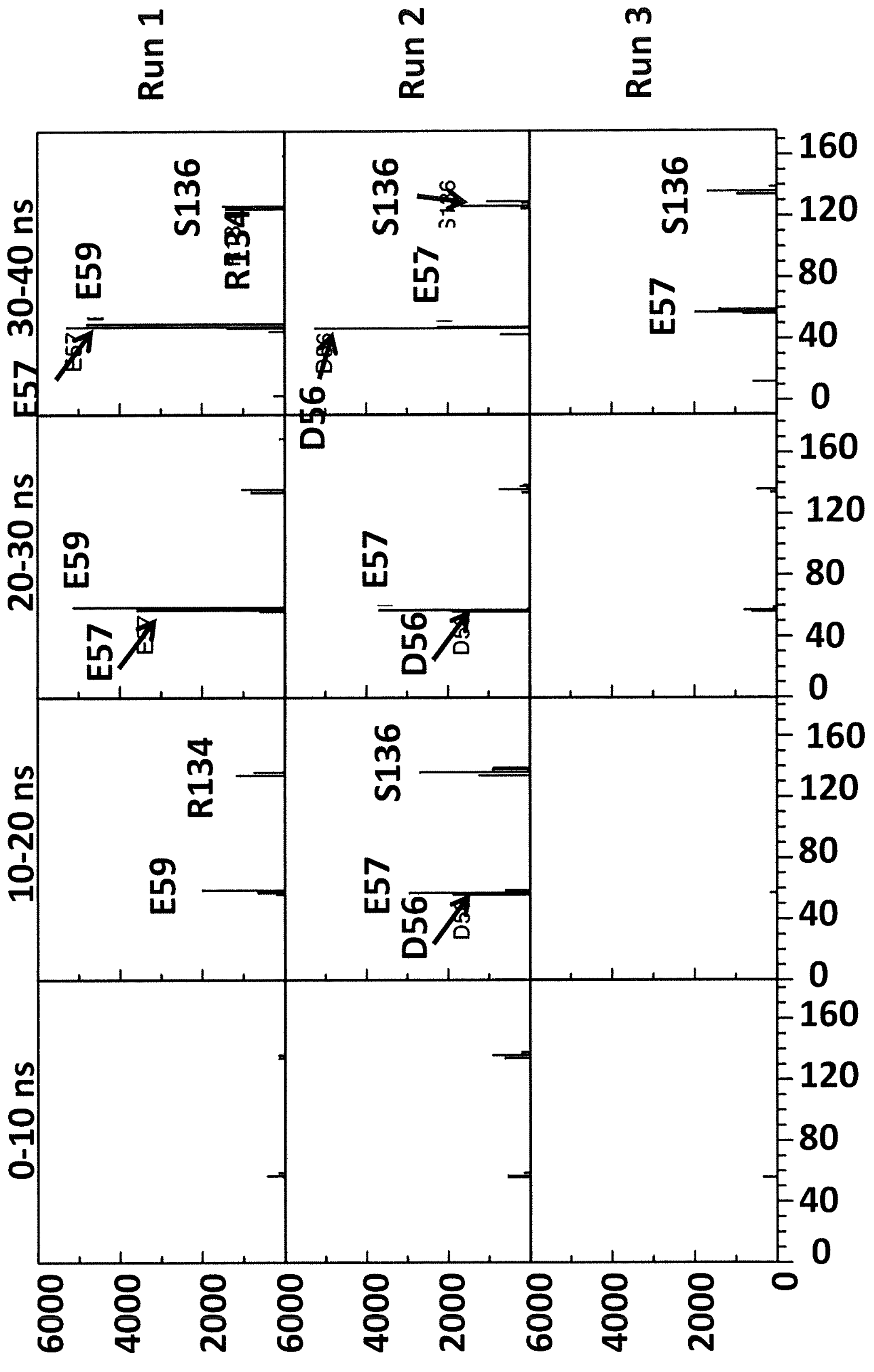

Run 1
Run 2
Run 3
0-10 ns
10-20 ns
20-30 ns
30-40 ns
E57
E59
S136
R134
D56
E57
S136
E57
S136
E59
R134
E57
D56
S136
E59
E57
E57
D56
0
40
80
120
160
6000
4000
2000
0

Figure 3

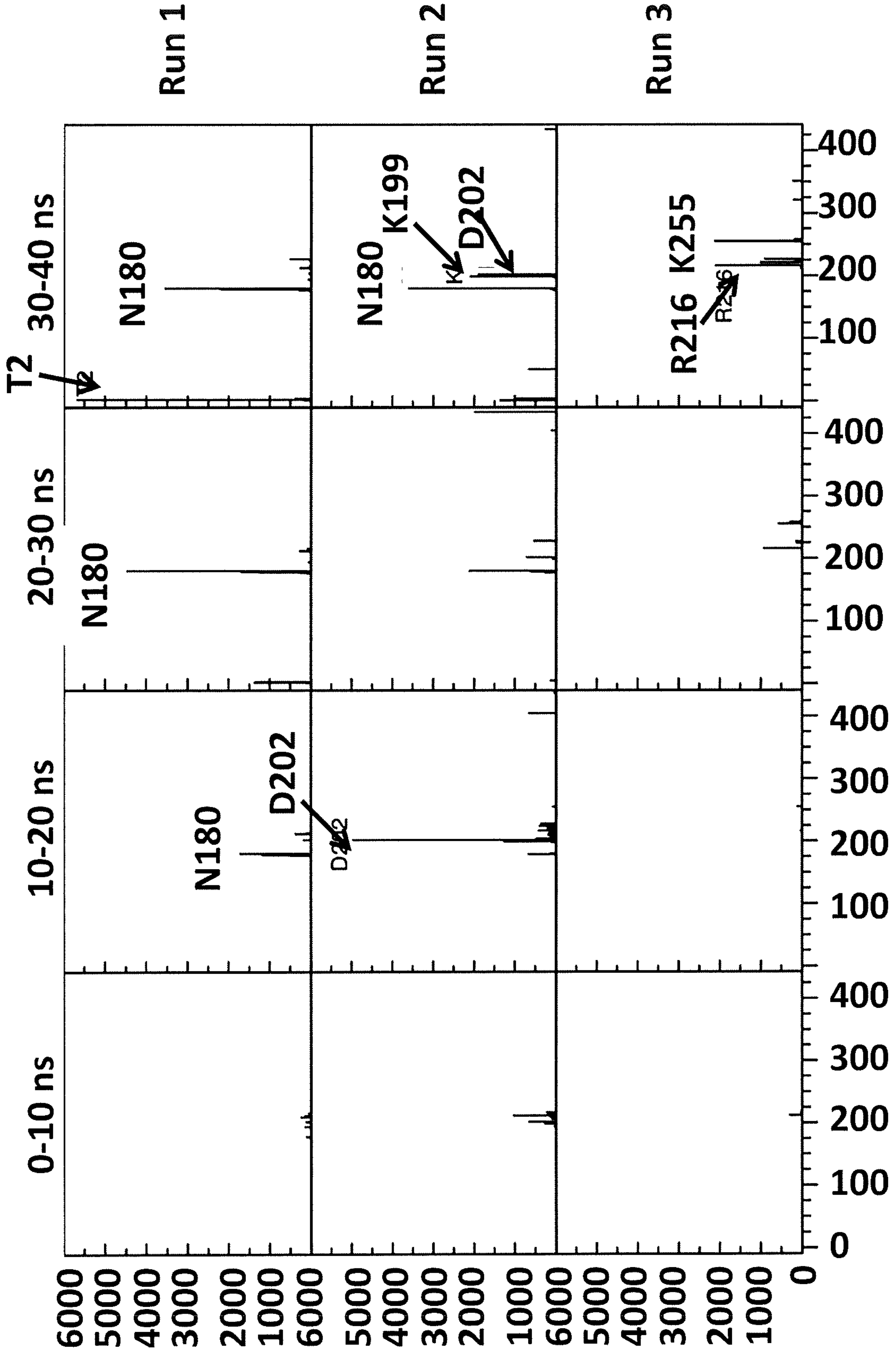

Run 1
Run 2
Run 3
0-10 ns
10-20 ns
20-30 ns
30-40 ns
T2
N180
N180
N180
N180
D202
D202
D202
K199
K
R216
R216
K255
0
100
200
300
400
0
1000
2000
3000
4000
5000
6000

Figure 4

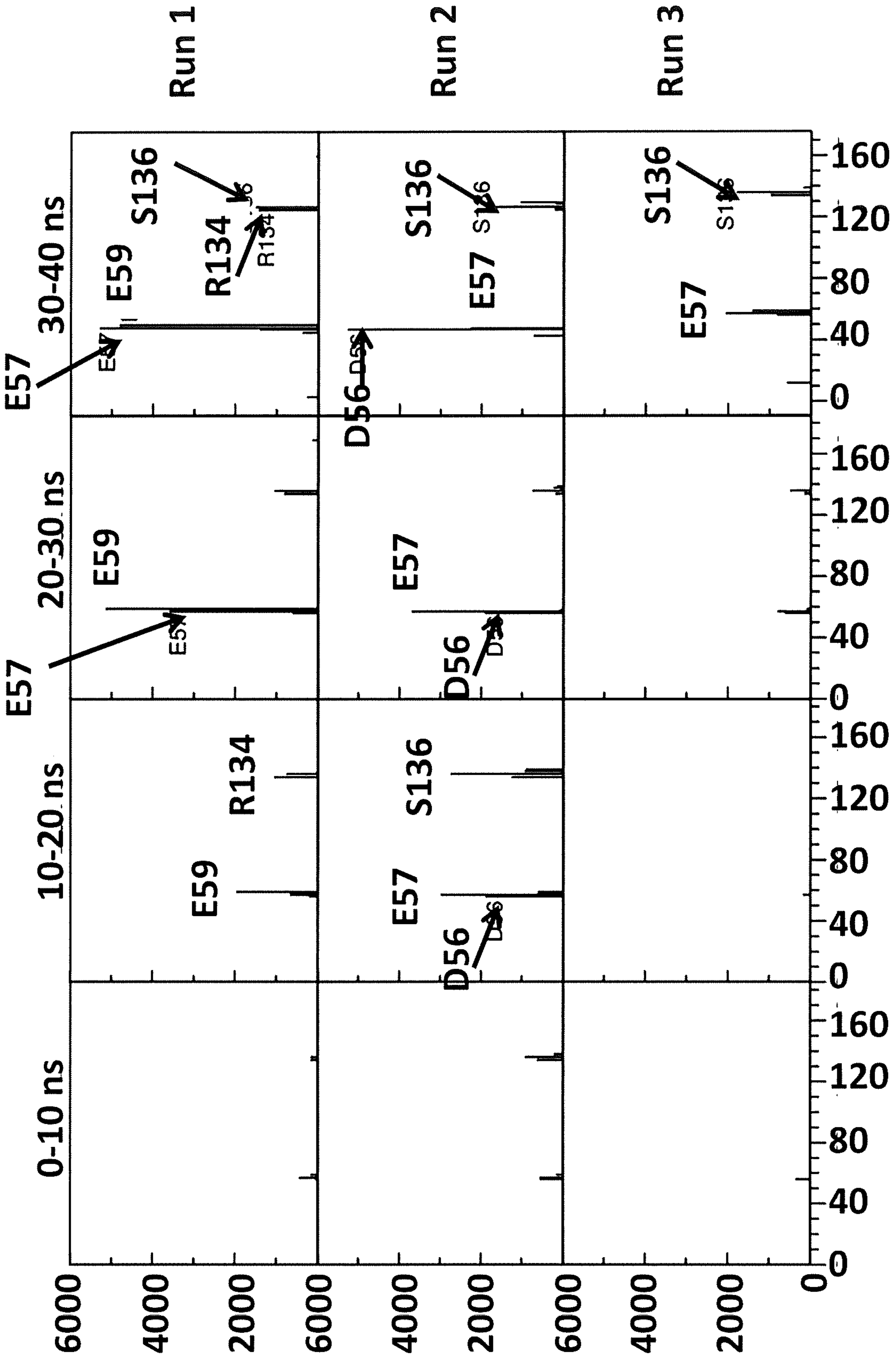

Run 1
Run 2
Run 3
0-10 ns
10-20 ns
20-30 ns
30-40 ns
E57
E59
R134
S136
D56
0
40
80
120
160
6000
4000
2000
0

**Figure 5**

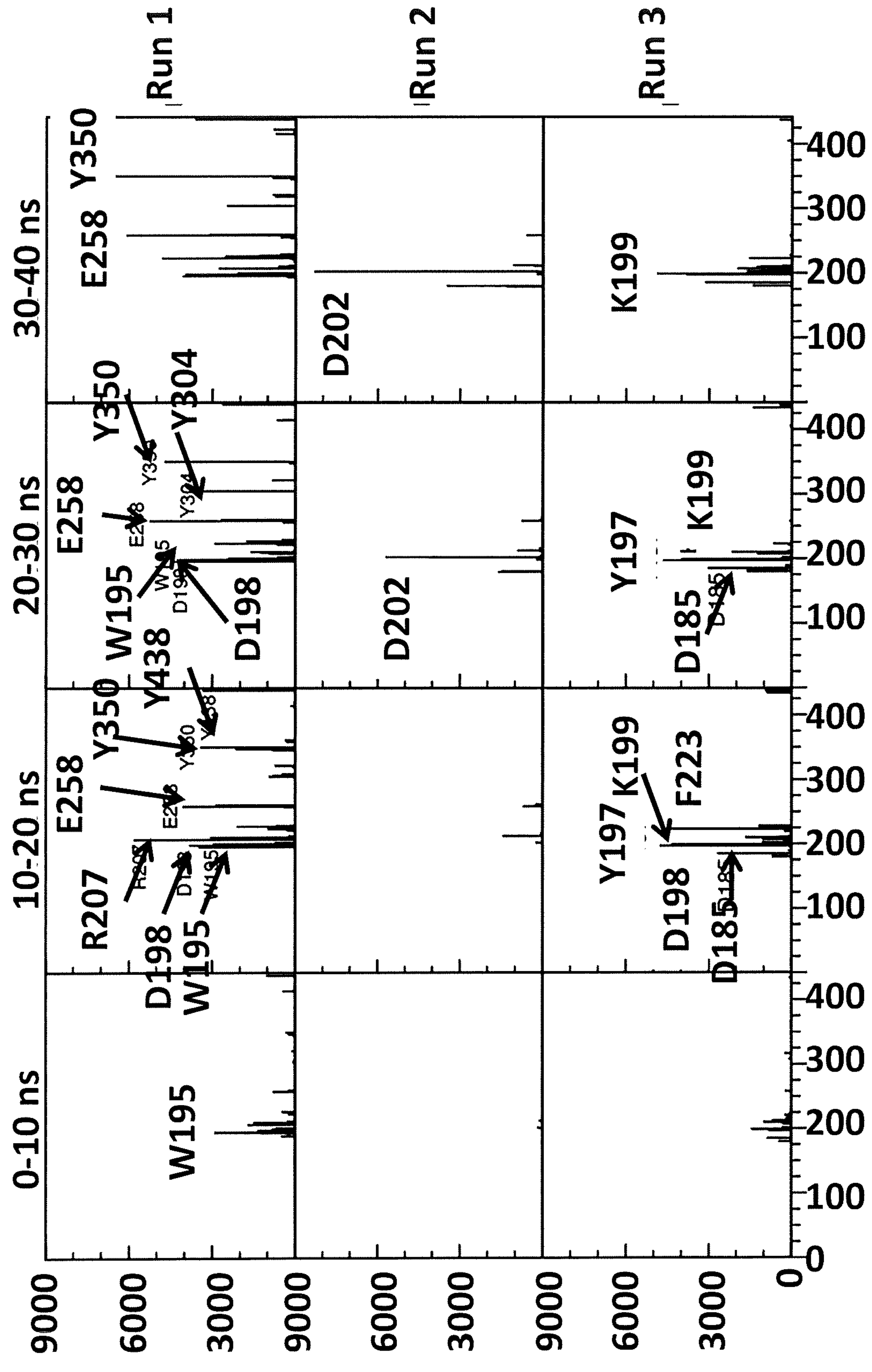

Run 1
Run 2
Run 3
0-10 ns
10-20 ns
20-30 ns
30-40 ns
W195
R207
D198
W195
E258
Y350
Y438
W195
D198
E258
Y350
Y304
E258
Y350
D202
D202
Y197
K199
F223
D198
D185
Y197
K199
D185
K199
0
100
200
300
400
9000
6000
3000
0

Figure 6

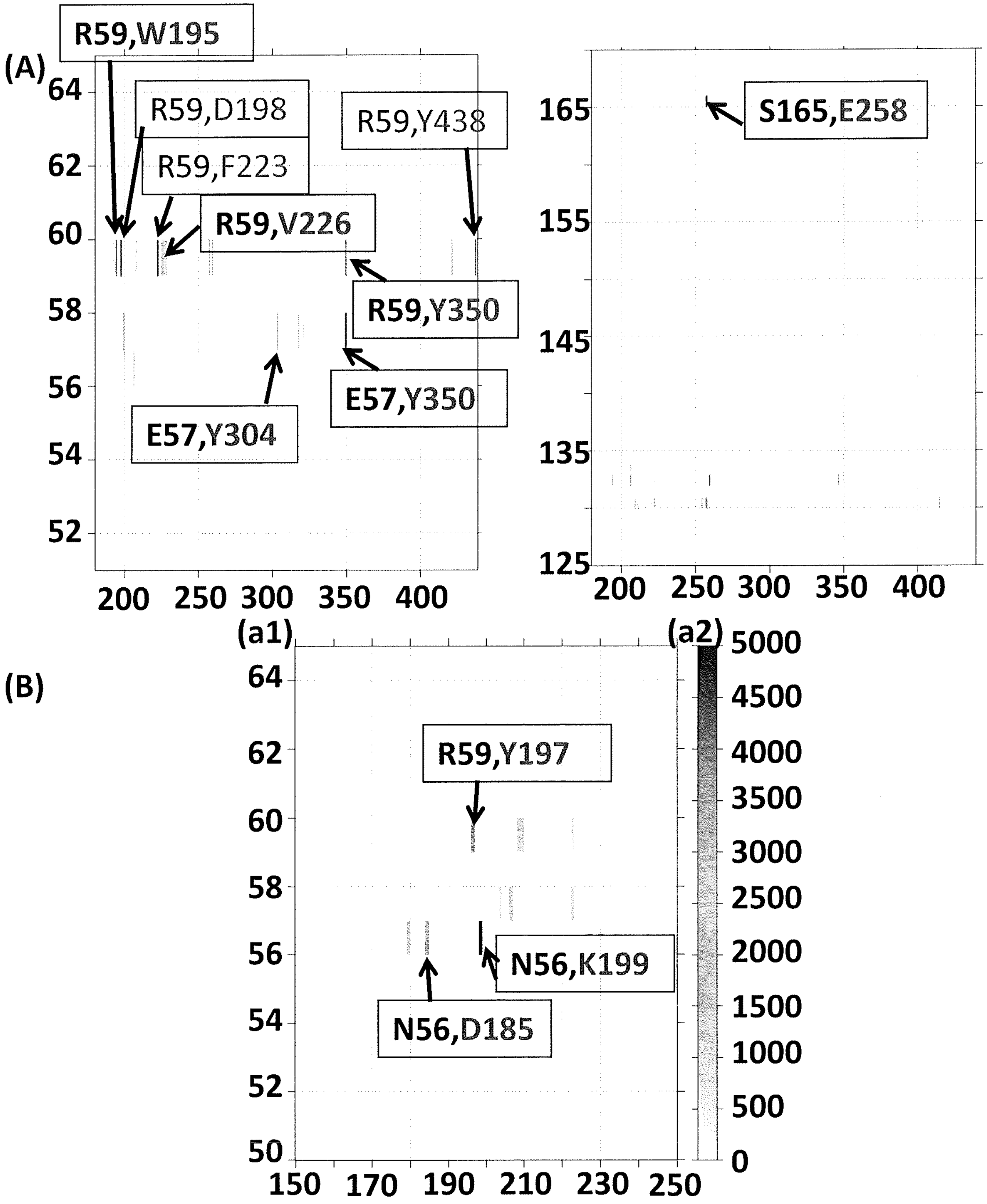

(A)
R59,W195
R59,D198
R59,F223
R59,V226
R59,Y438
R59,Y350
E57,Y350
E57,Y304
64
62
60
58
56
54
52
200
250
300
350
400
S165,E258
165
155
145
135
125
200
250
300
350
400
(B)
(a1)
(a2)
R59,Y197
N56,K199
N56,D185
64
62
60
58
56
54
52
50
150
170
190
210
230
250
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0

Figure 7

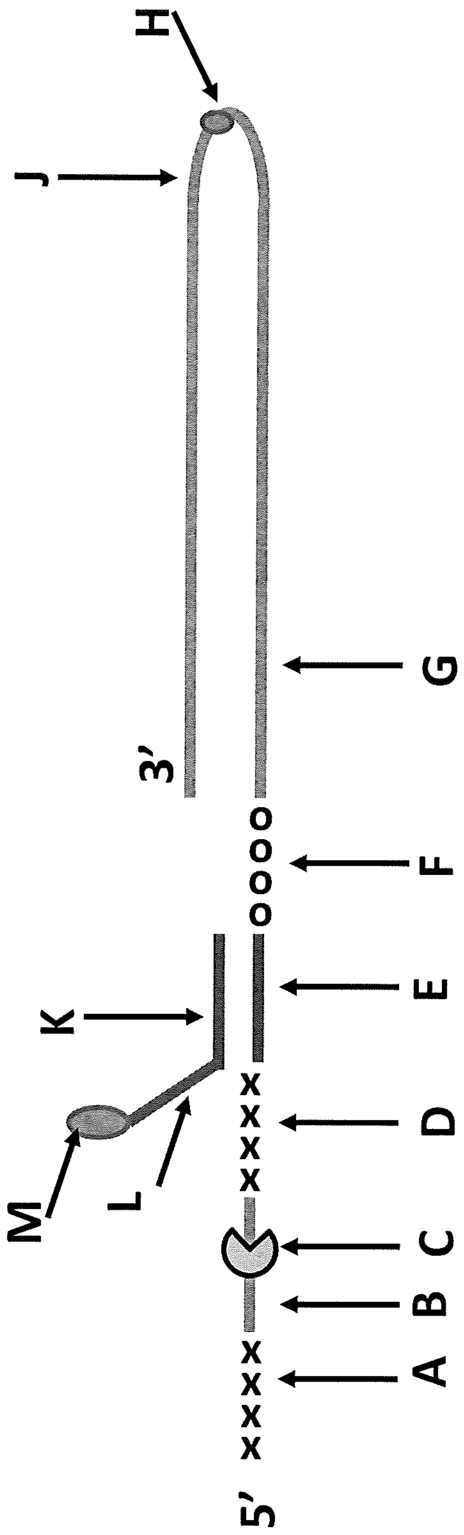


Figure 8

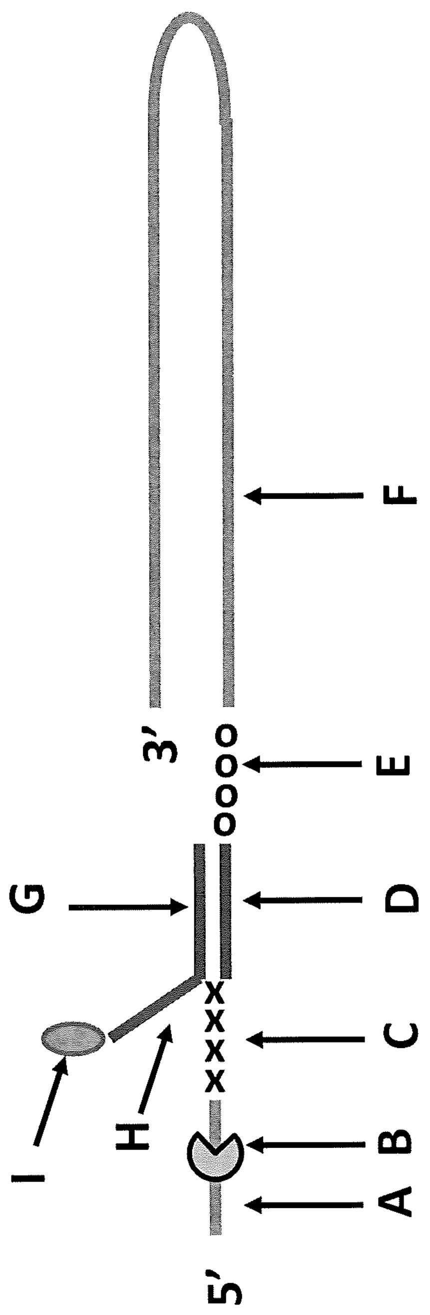


Figure 9

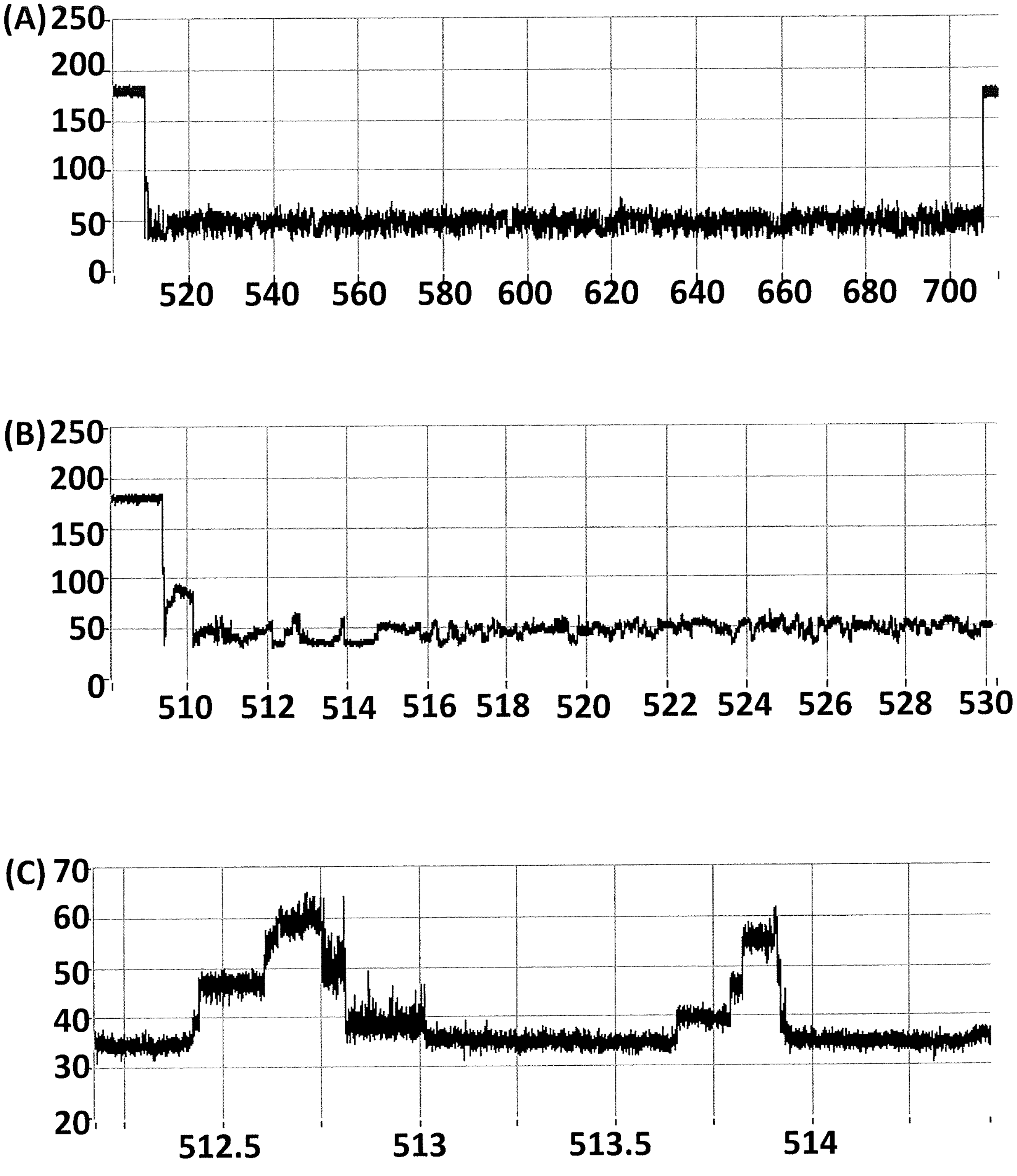

(A)
250
200
150
100
50
0
520
540
560
580
600
620
640
660
680
700
(B)
250
200
150
100
50
0
510
512
514
516
518
520
522
524
526
528
530
(C)
70
60
50
40
30
20
512.5
513
513.5
514

Figure 10

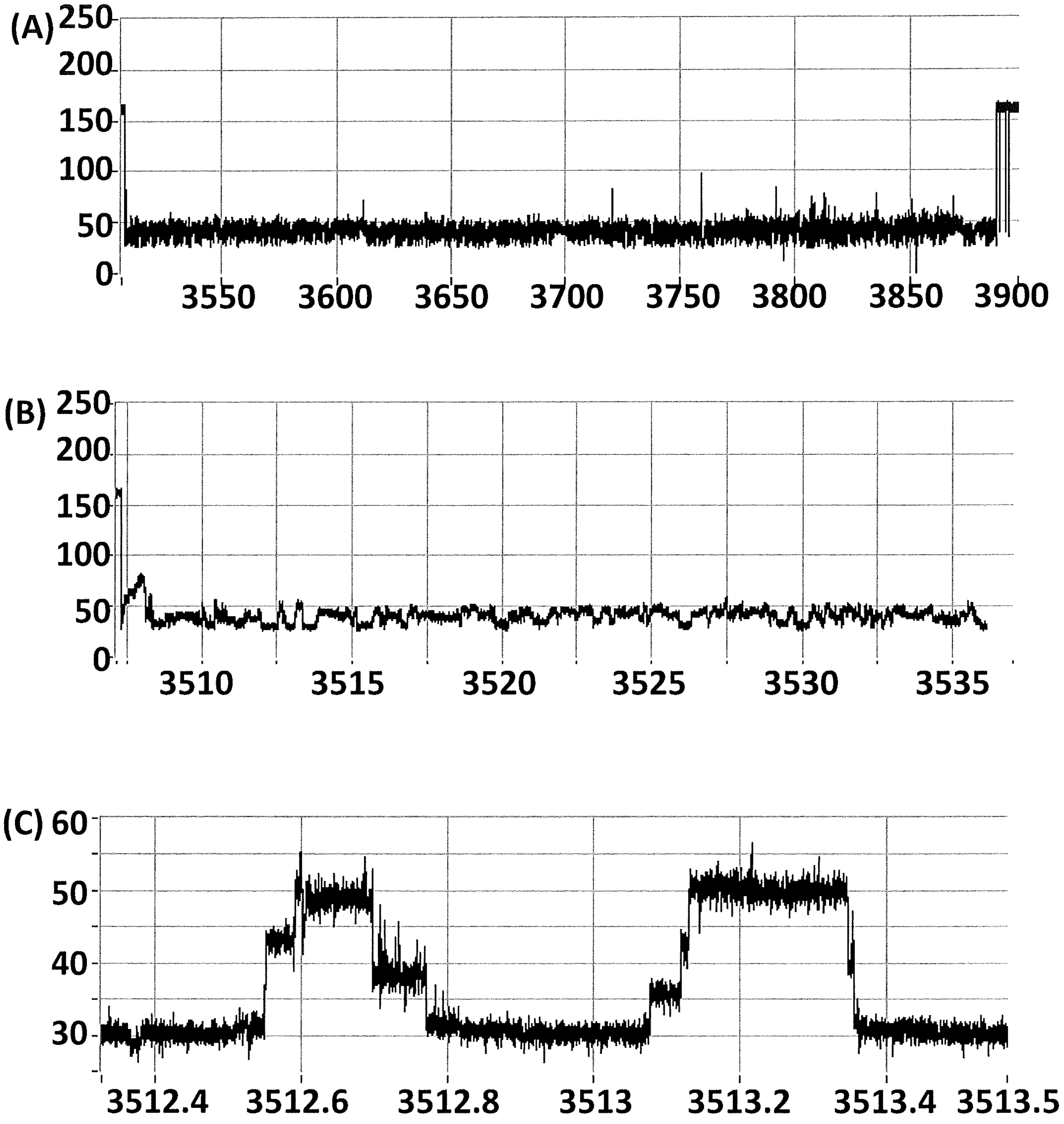

(A)
250
200
150
100
50
0
3550
3600
3650
3700
3750
3800
3850
3900
(B)
250
200
150
100
50
0
3510
3515
3520
3525
3530
3535
(C)
60
50
40
30
3512.4
3512.6
3512.8
3513
3513.2
3513.4
3513.5

Figure 11

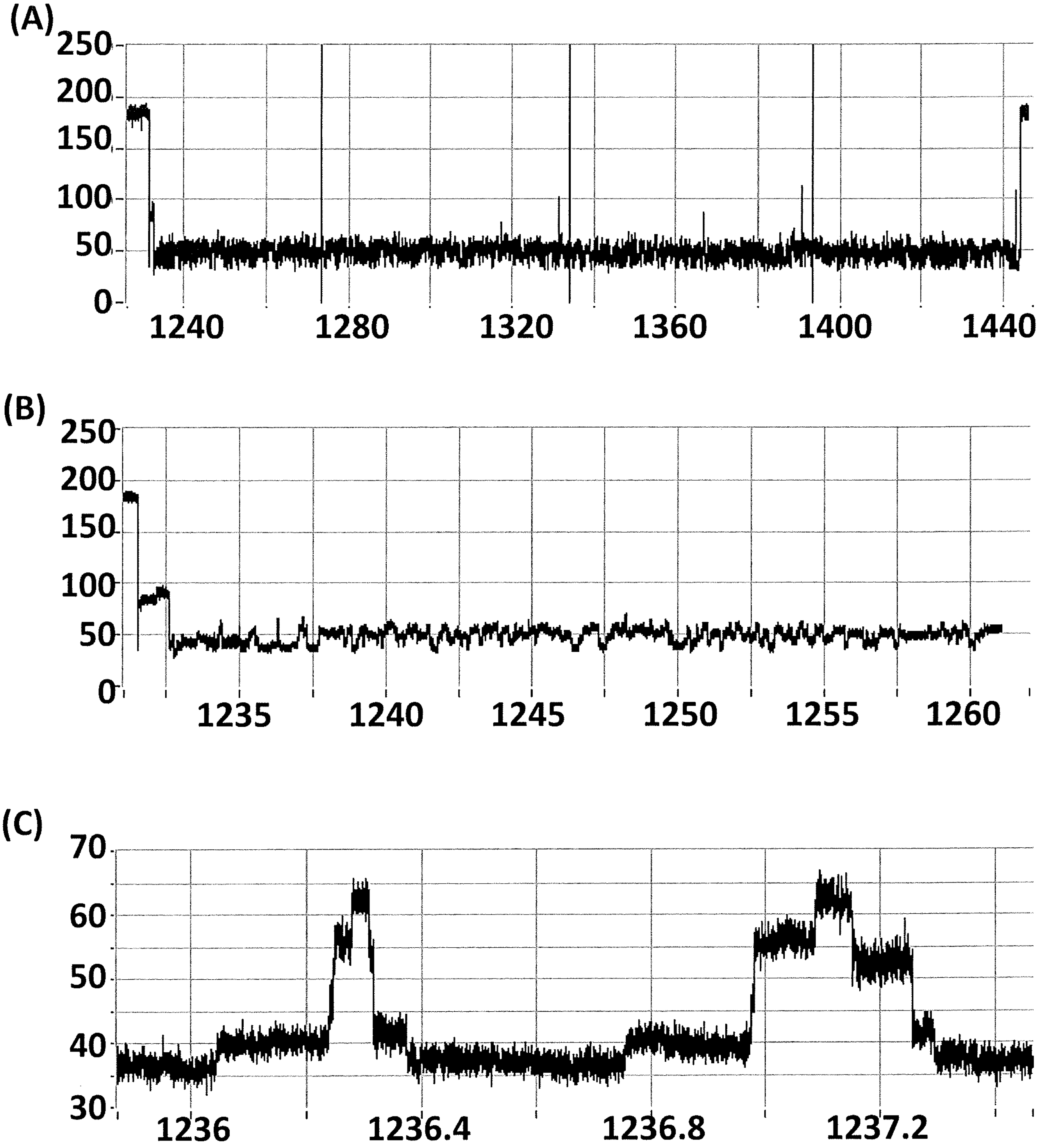

(A)
250
200
150
100
50
0
1240
1280
1320
1360
1400
1440
(B)
250
200
150
100
50
0
1235
1240
1245
1250
1255
1260
(C)
70
60
50
40
30
1236
1236.4
1236.8
1237.2

Figure 12

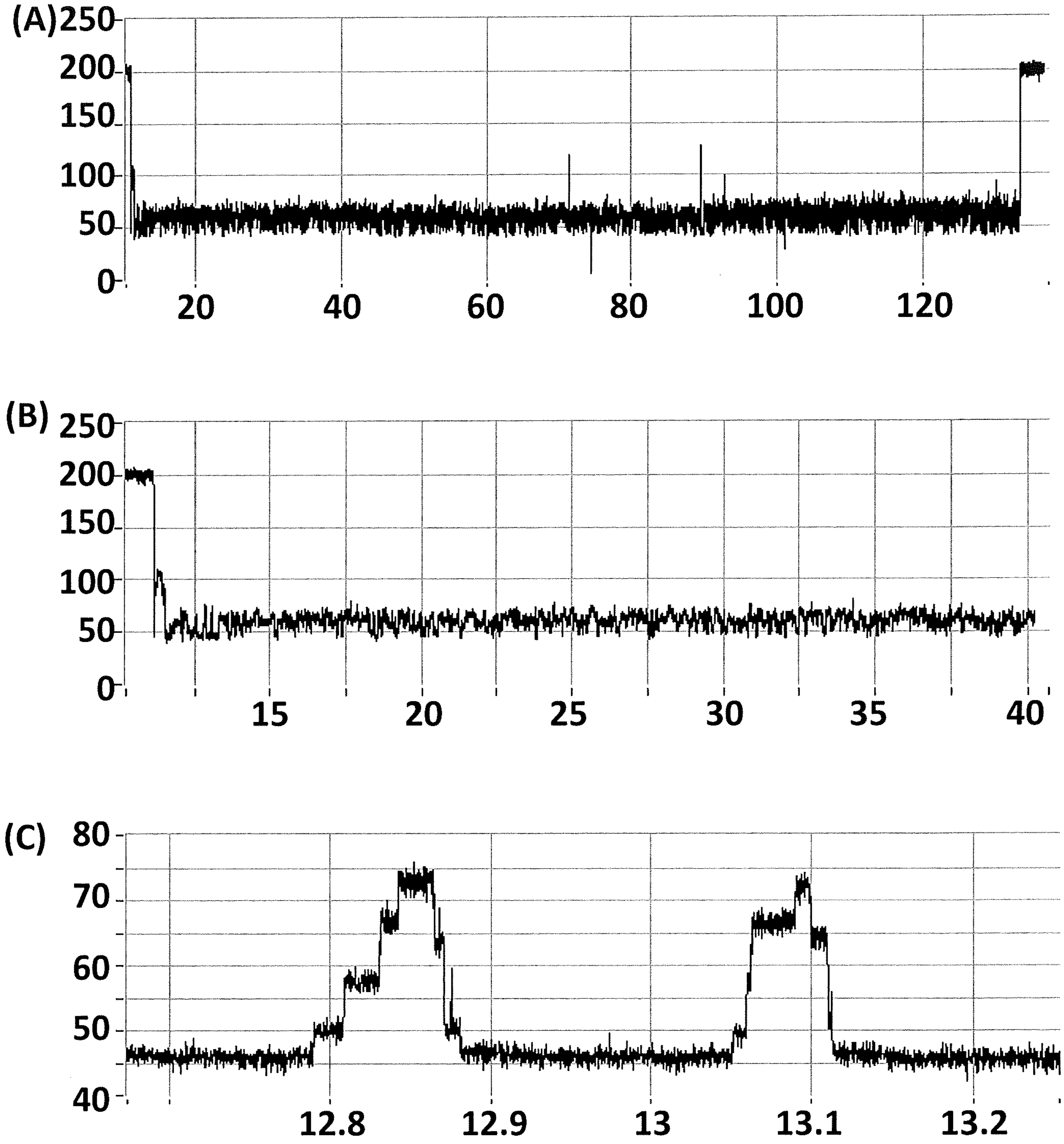

(A)
250
200
150
100
50
0
20
40
60
80
100
120
(B)
250
200
150
100
50
0
15
20
25
30
35
40
(C)
80
70
60
50
40
12.8
12.9
13
13.1
13.2

Figure 13

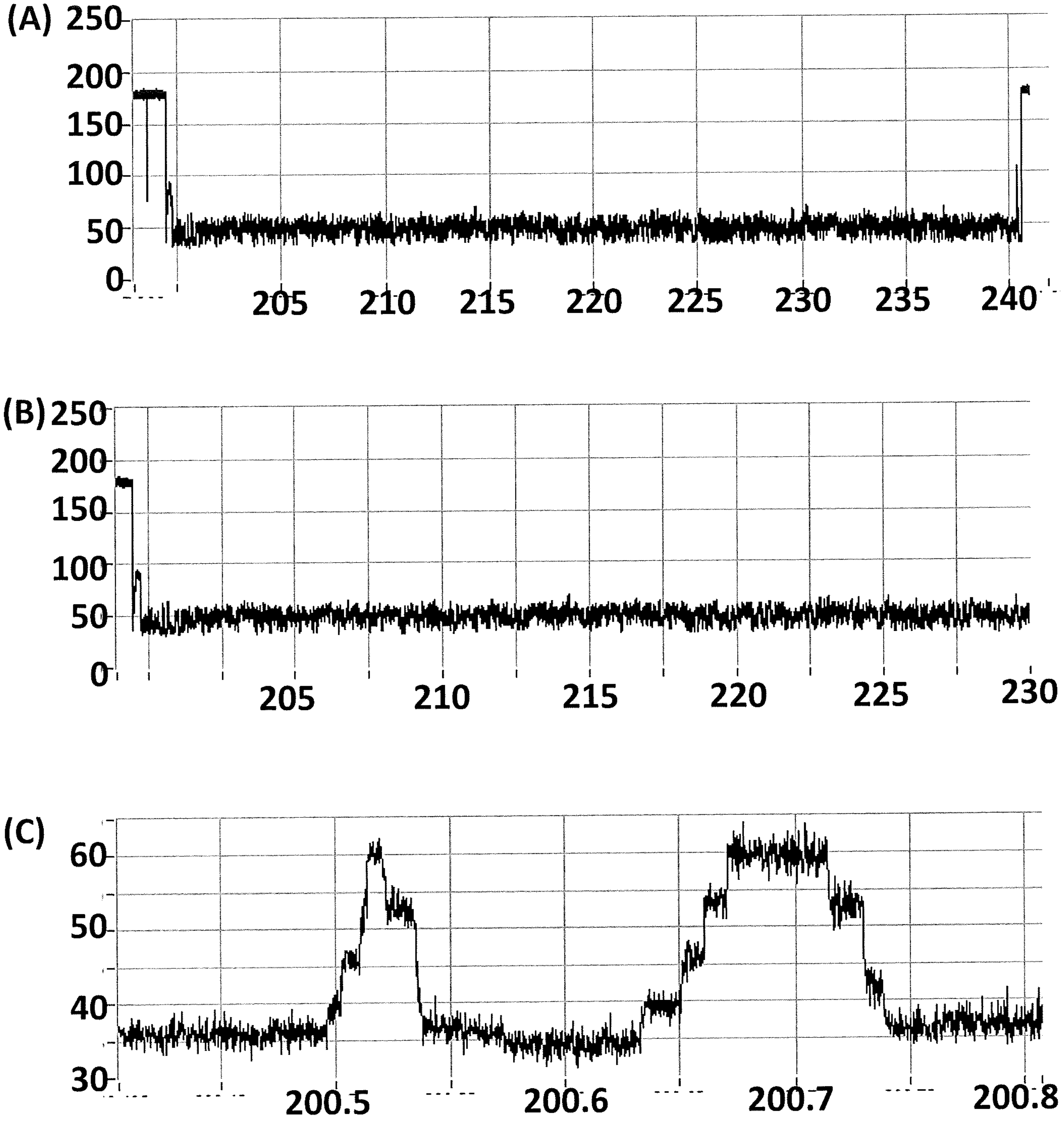

(A)
250
200
150
100
50
0
205
210
215
220
225
230
235
240
(B)
250
200
150
100
50
0
205
210
215
220
225
230
(C)
60
50
40
30
200.5
200.6
200.7
200.8

Figure 14

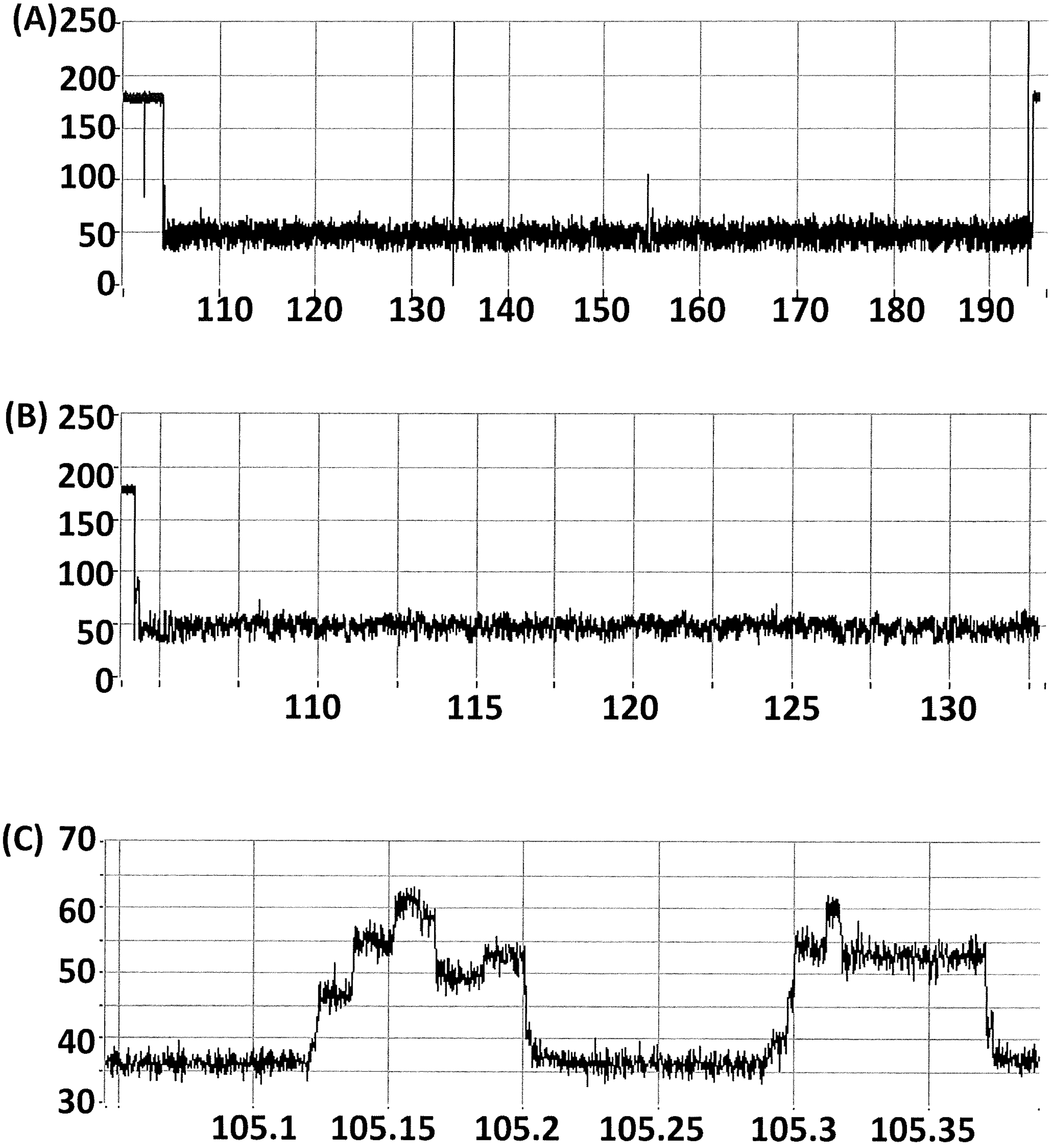

(A)
250
200
150
100
50
0
110
120
130
140
150
160
170
180
190
(B)
250
200
150
100
50
0
110
115
120
125
130
(C)
70
60
50
40
30
105.1
105.15
105.2
105.25
105.3
105.35

Figure 15

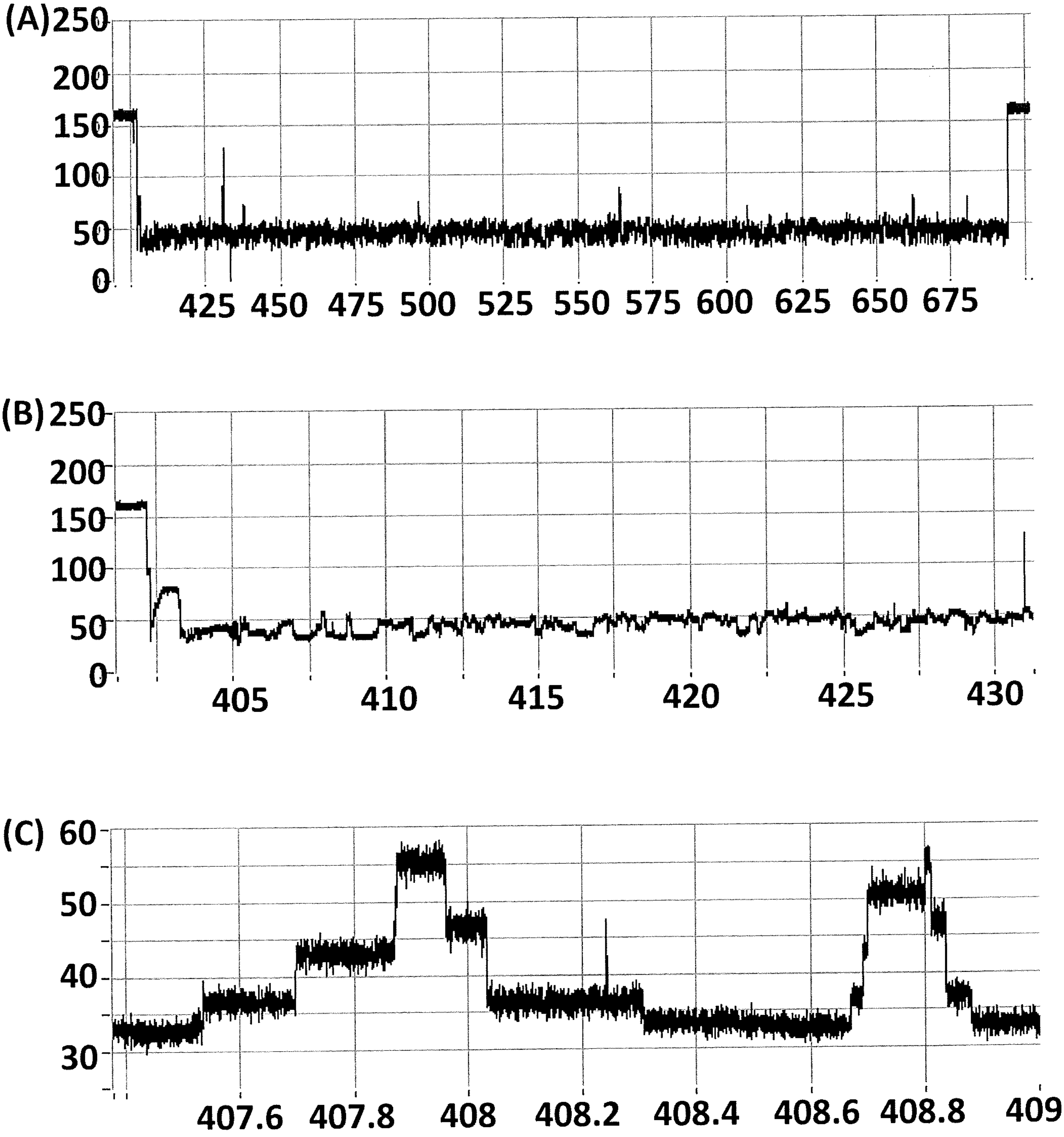

(A)
250
200
150
100
50
0
425
450
475
500
525
550
575
600
625
650
675
(B)
250
200
150
100
50
0
405
410
415
420
425
430
(C)
60
50
40
30
407.6
407.8
408
408.2
408.4
408.6
408.8
409

Figure 16

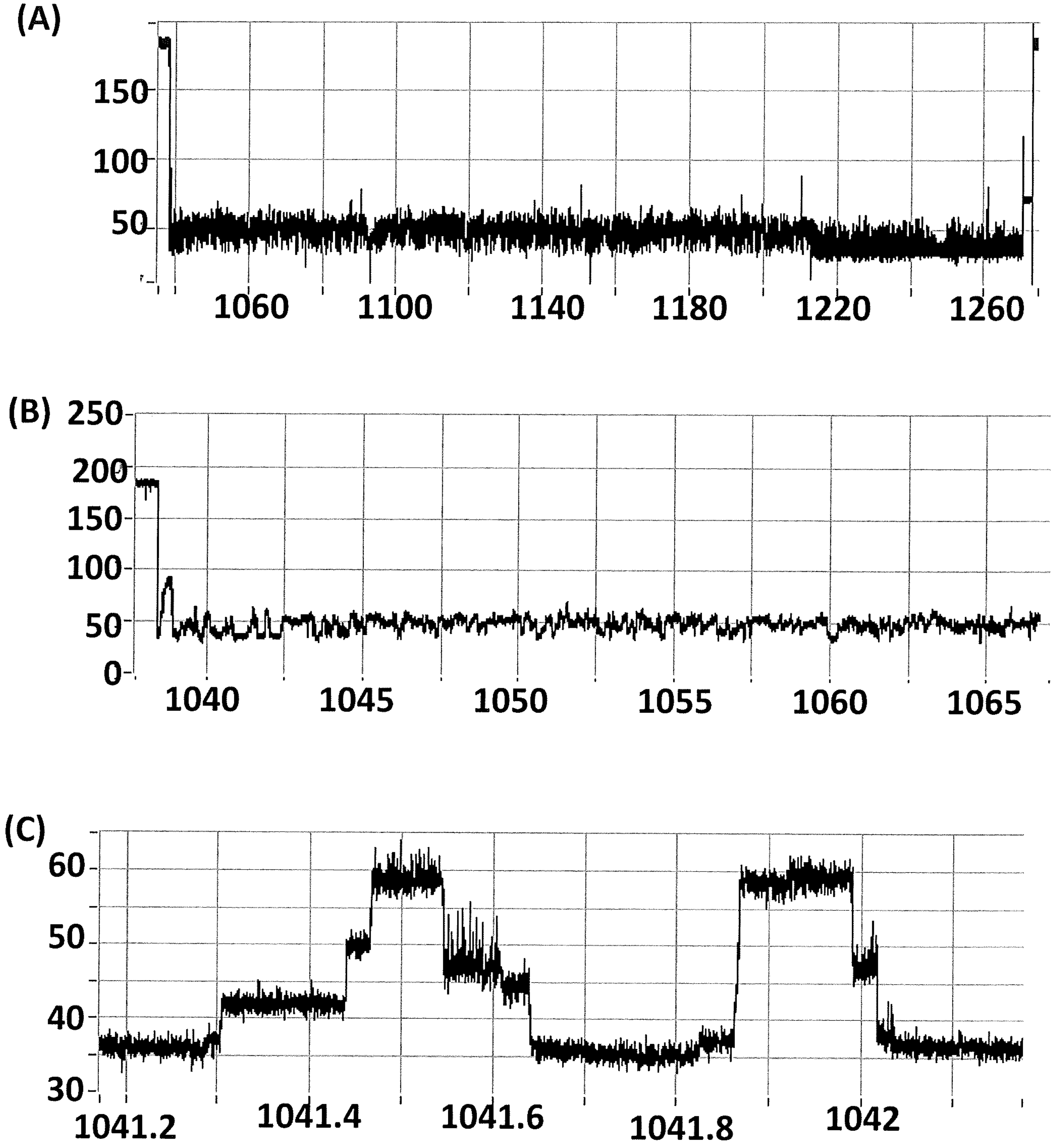

(A)
150
100
50
1060
1100
1140
1180
1220
1260
(B)
250
200
150
100
50
0
1040
1045
1050
1055
1060
1065
(C)
60
50
40
30
1041.2
1041.4
1041.6
1041.8
1042

Figure 17

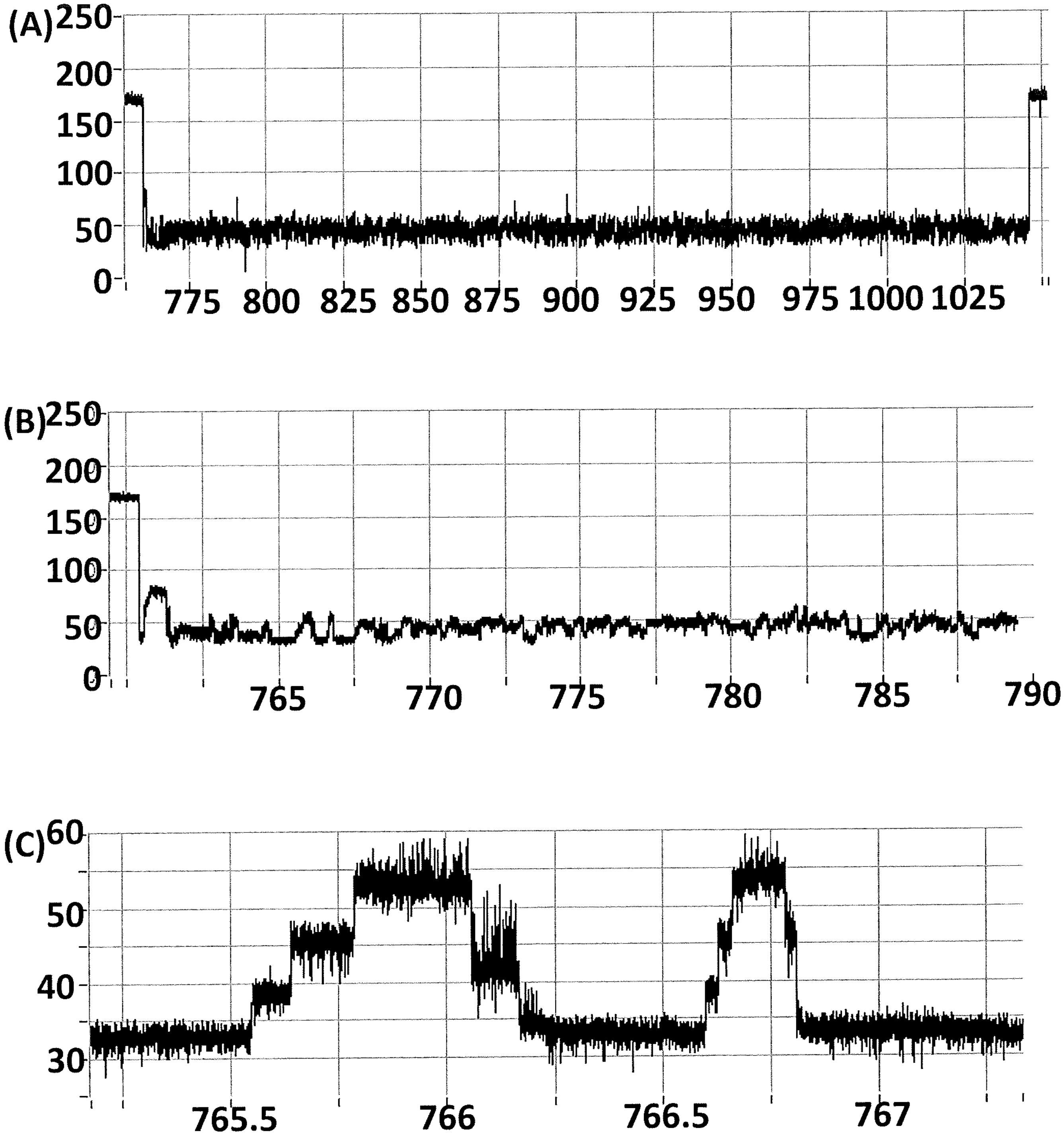

(A)
250
200
150
100
50
0
775
800
825
850
875
900
925
950
975
1000
1025
(B)
250
200
150
100
50
0
765
770
775
780
785
790
(C)
60
50
40
30
765.5
766
766.5
767

Figure 18

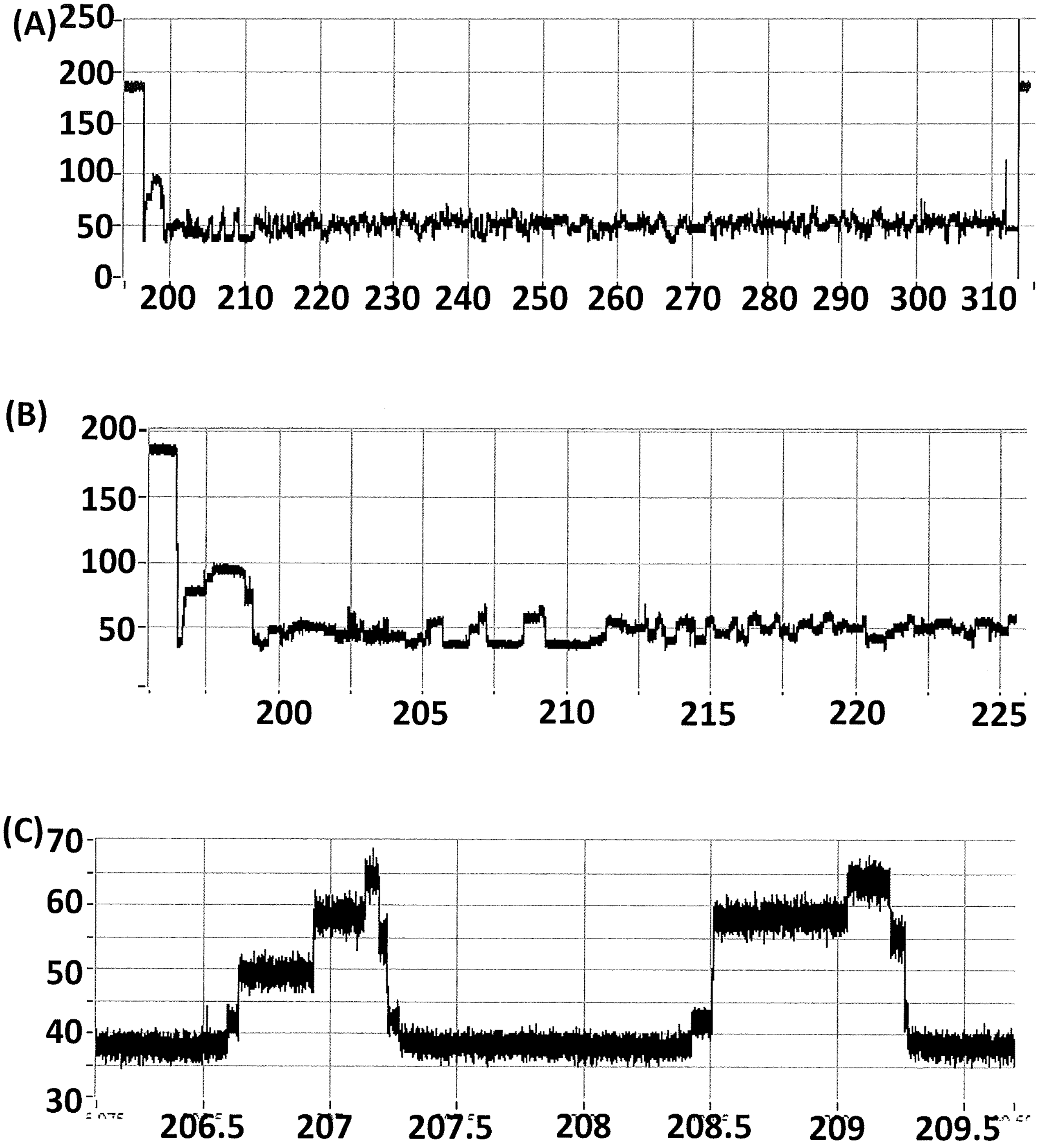

(A)
250
200
150
100
50
0
200
210
220
230
240
250
260
270
280
290
300
310
(B)
200
150
100
50
200
205
210
215
220
225
(C)
70
60
50
40
30
206.5
207
207.5
208
208.5
209
209.5

Figure 19

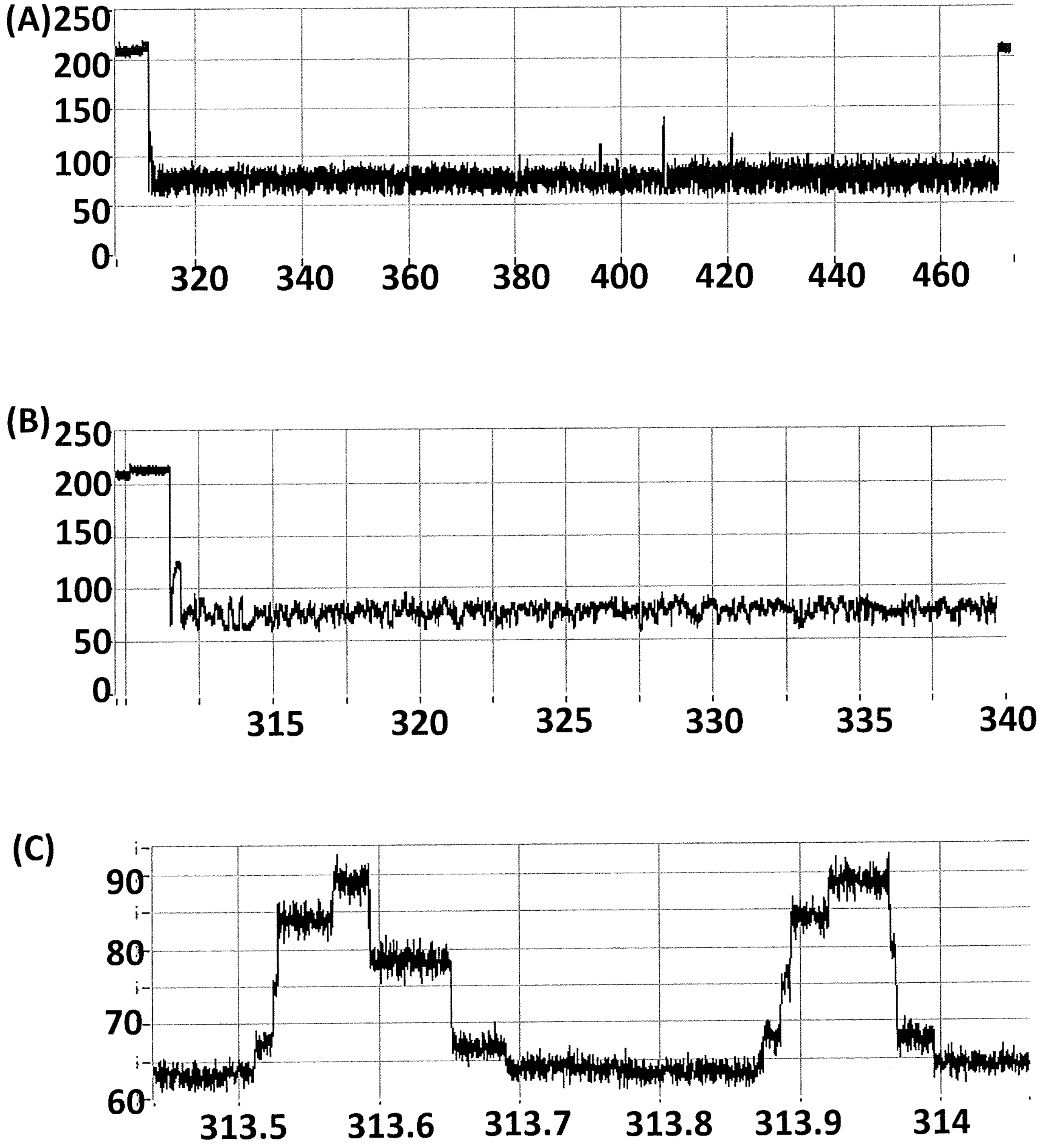

(A)
250
200
150
100
50
0
320
340
360
380
400
420
440
460
(B)
250
200
150
100
50
0
315
320
325
330
335
340
(C)
90
80
70
60
313.5
313.6
313.7
313.8
313.9
314

Figure 20

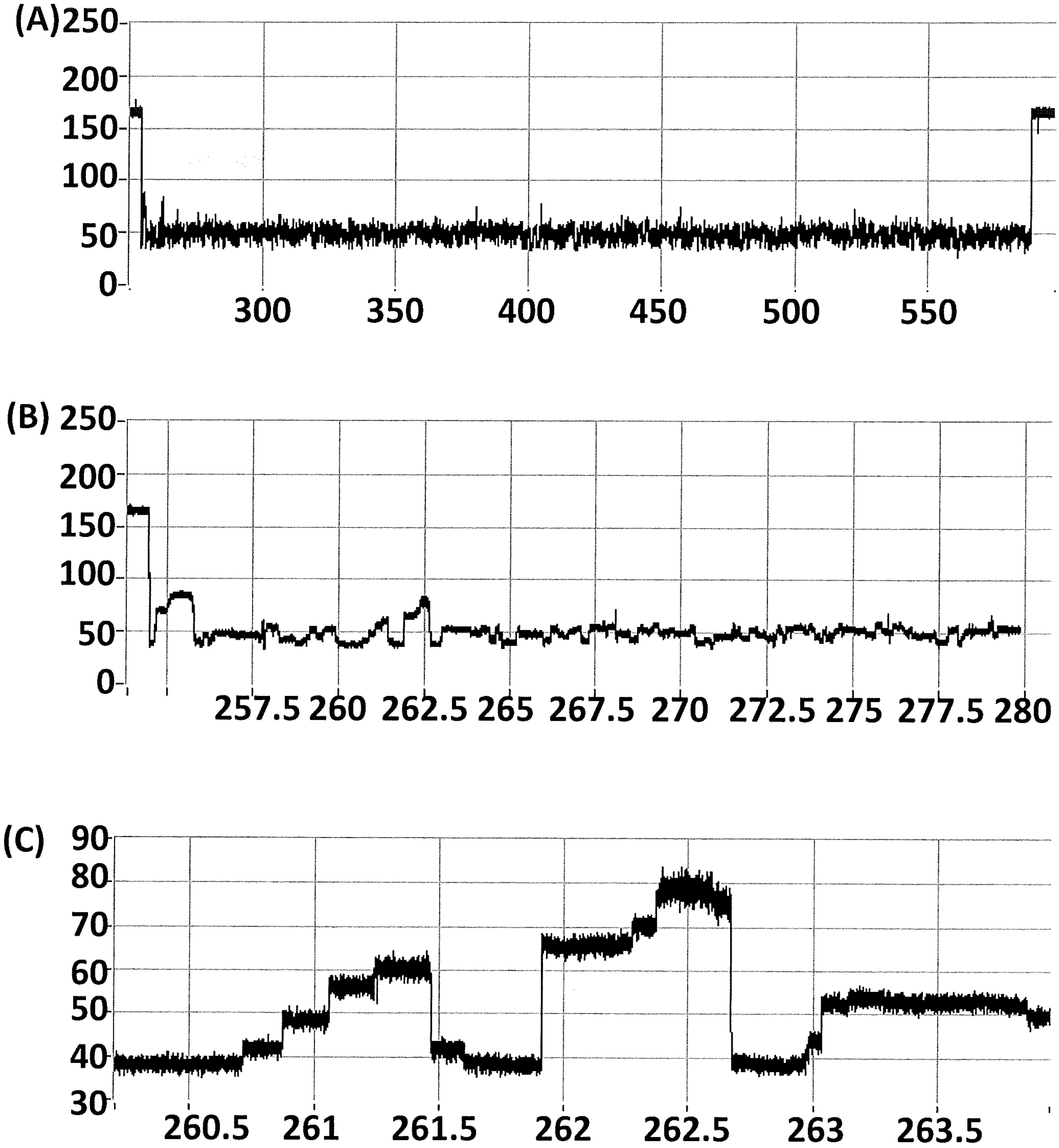

(A)
250
200
150
100
50
0
300
350
400
450
500
550
(B)
250
200
150
100
50
0
257.5
260
262.5
265
267.5
270
272.5
275
277.5
280
(C)
90
80
70
60
50
40
30
260.5
261
261.5
262
262.5
263
263.5

Figure 21

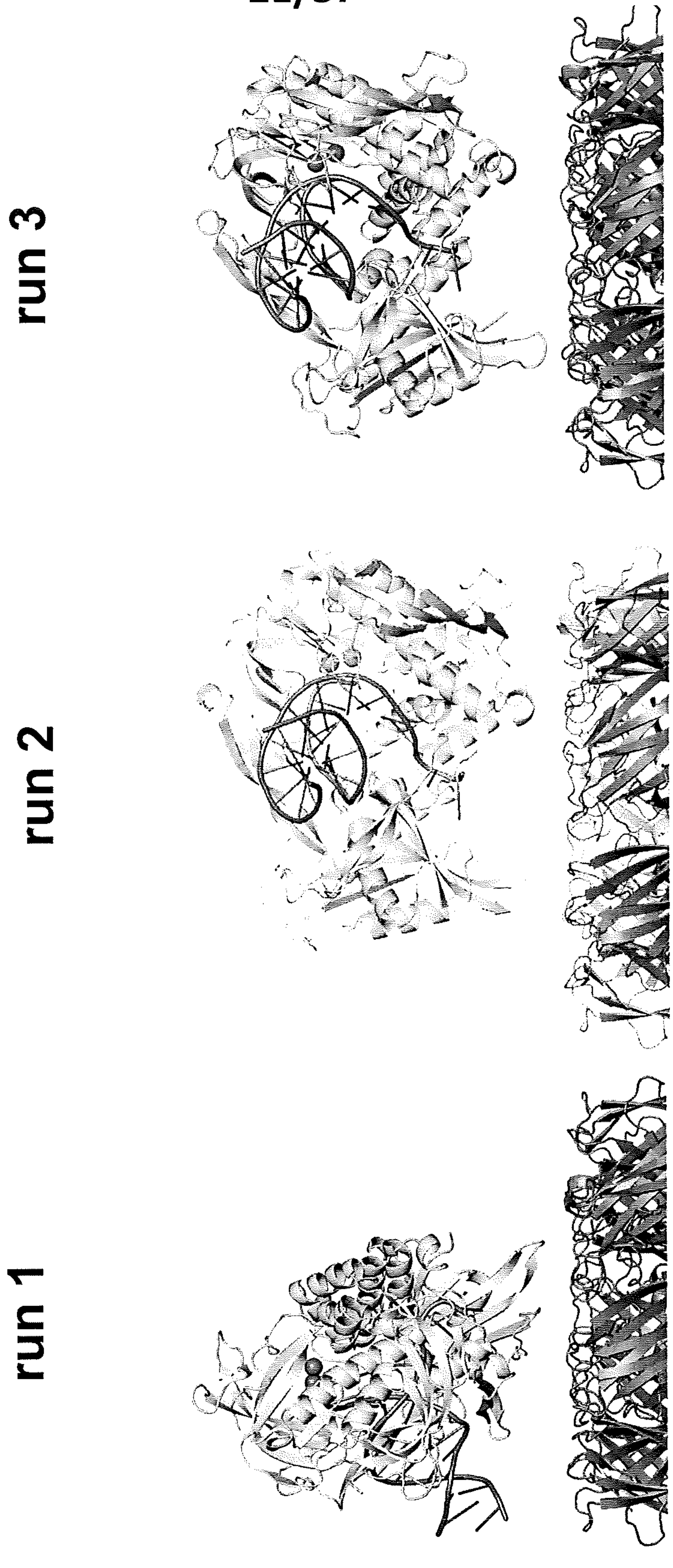


Figure 22

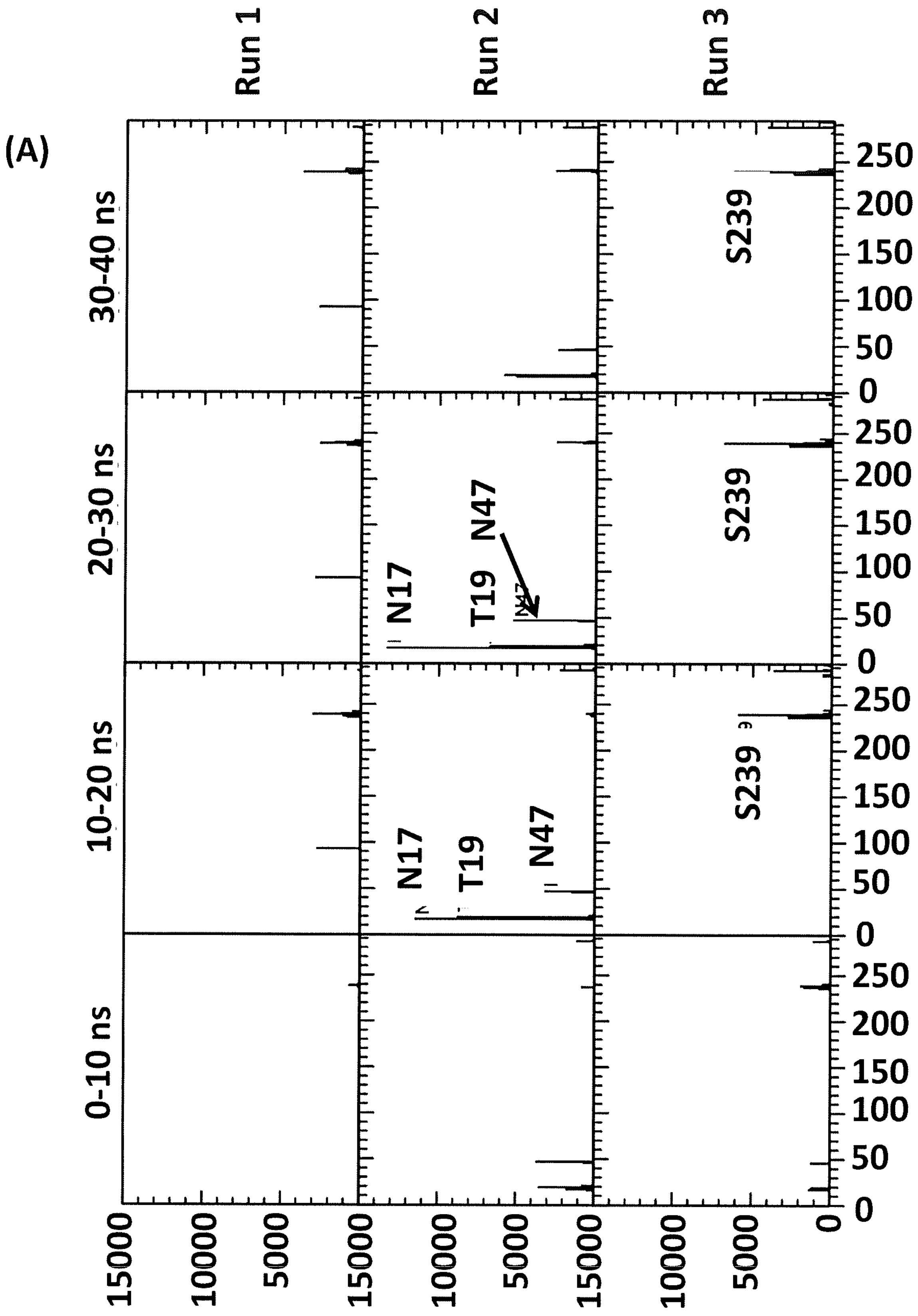

(A)
Run 1
Run 2
Run 3
0-10 ns
10-20 ns
20-30 ns
30-40 ns
N17
T19
N47
S239
0
50
100
150
200
250
15000
10000
5000
0

Figure 23

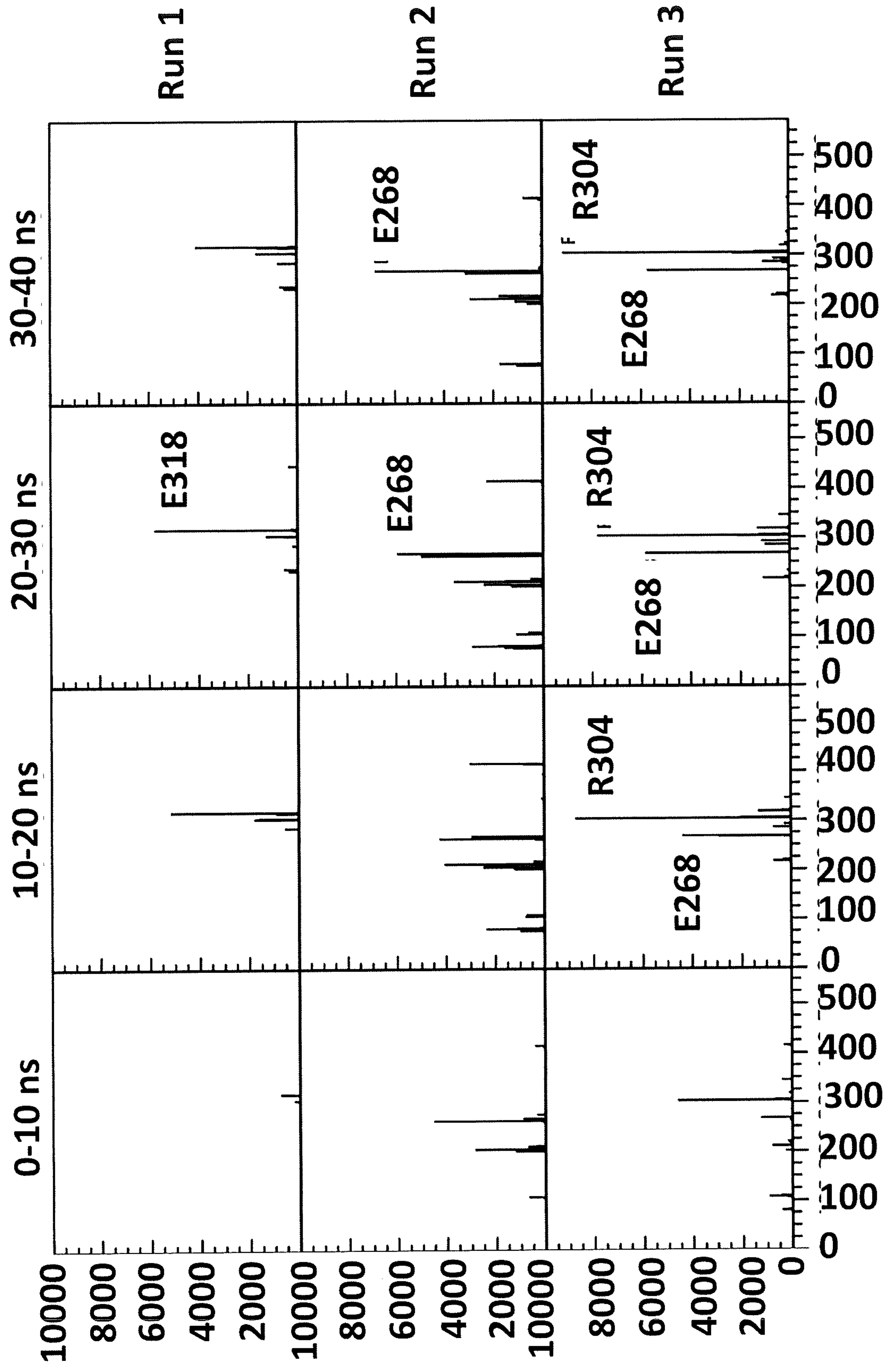

Run 1
Run 2
Run 3
0-10 ns
10-20 ns
20-30 ns
30-40 ns
E318
E268
E268
E268
R304
R304
R304
E268
E268
E268
0
100
200
300
400
500
10000
8000
6000
4000
2000
0

Figure 24

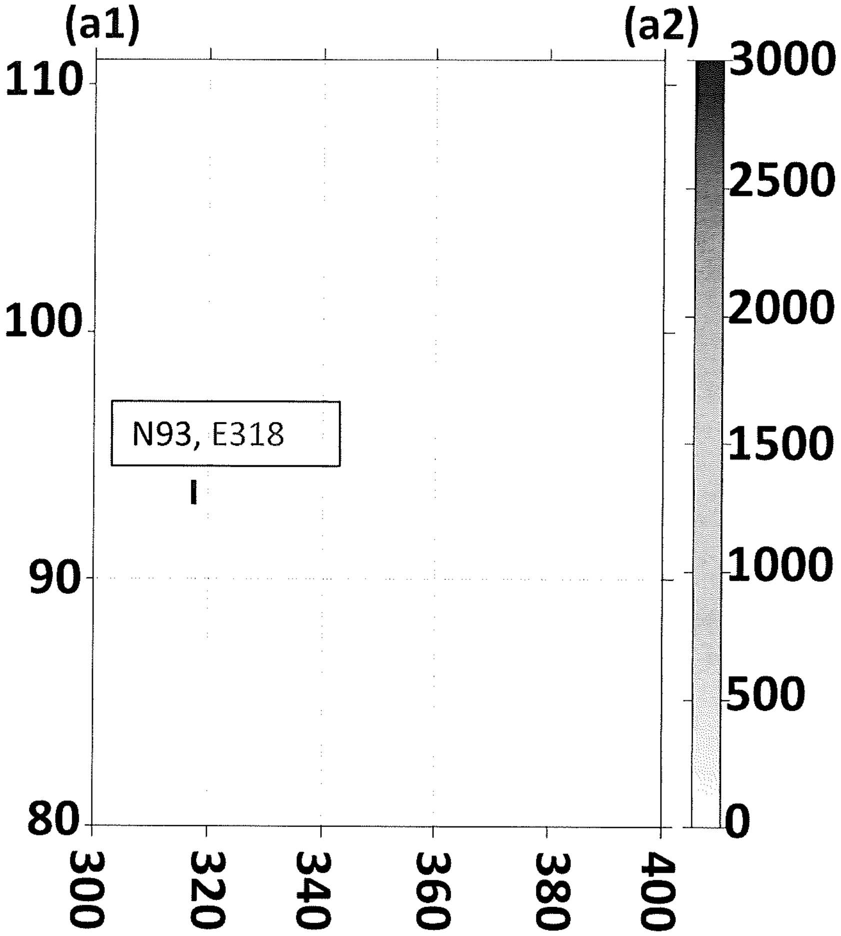

(a1)
(a2)
110
100
90
80
N93, E318
300
320
340
360
380
400
3000
2500
2000
1500
1000
500
0

Figure 25

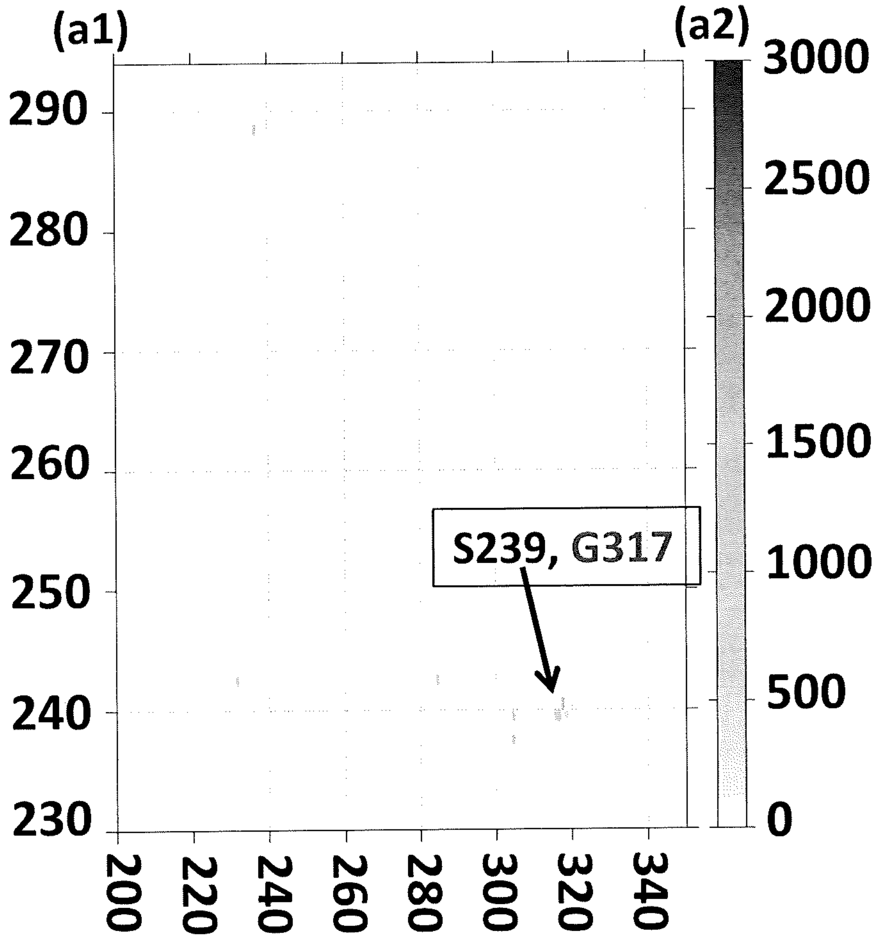

(a1)
(a2)
290
280
270
260
250
240
230
200
220
240
260
280
300
320
340
3000
2500
2000
1500
1000
500
0
S239, G317

Figure 26

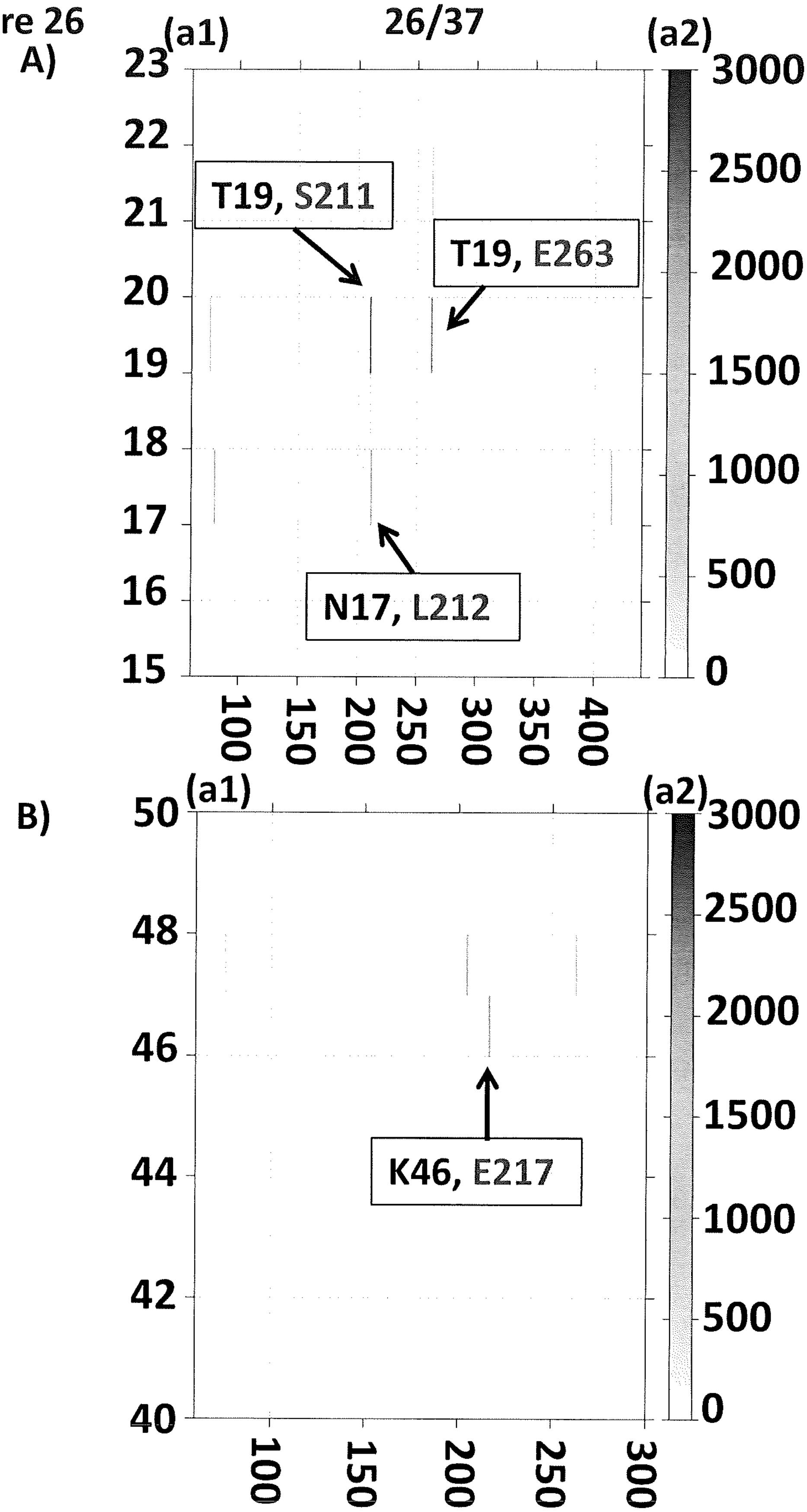

A)
(a1)
(a2)
23
22
21
20
19
18
17
16
15
T19, S211
T19, E263
N17, L212
100
150
200
250
300
350
400
3000
2500
2000
1500
1000
500
0
B)
(a1)
(a2)
50
48
46
44
42
40
K46, E217
100
150
200
250
300
3000
2500
2000
1500
1000
500
0

Figure 27

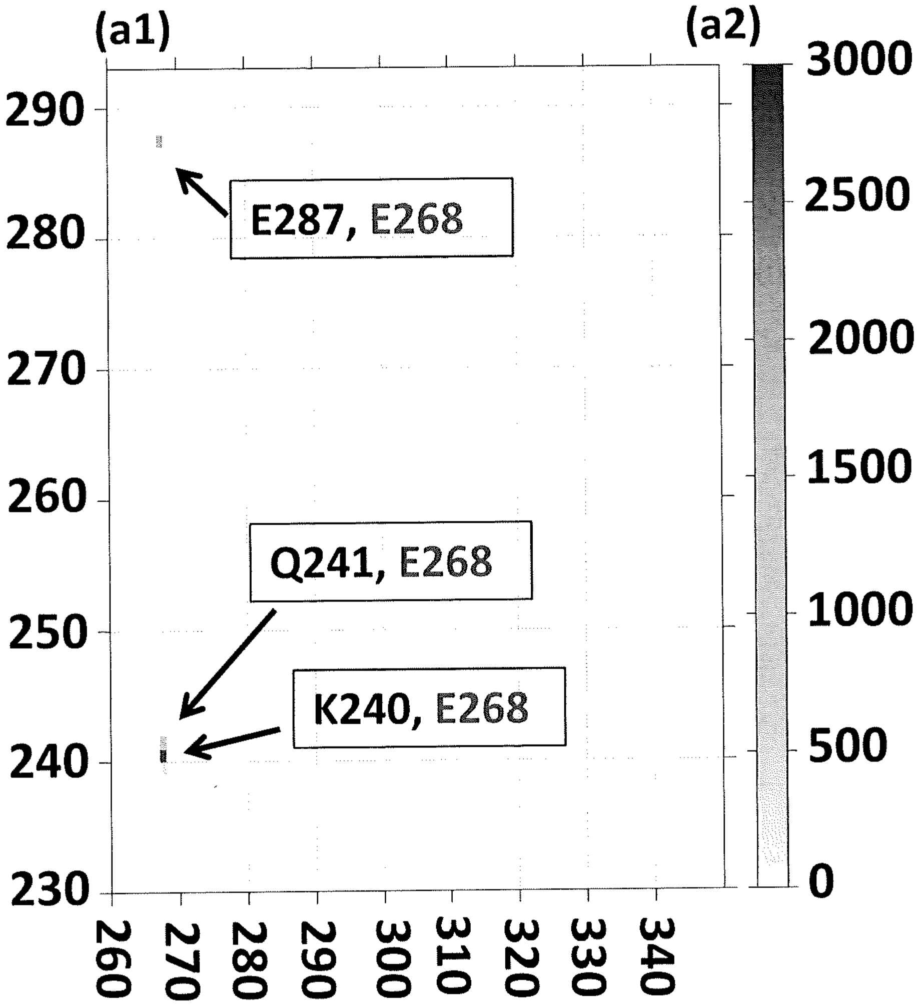

(a1)
(a2)
E287, E268
Q241, E268
K240, E268
290
280
270
260
250
240
230
260
270
280
290
300
310
320
330
340
3000
2500
2000
1500
1000
500
0

Figure 28

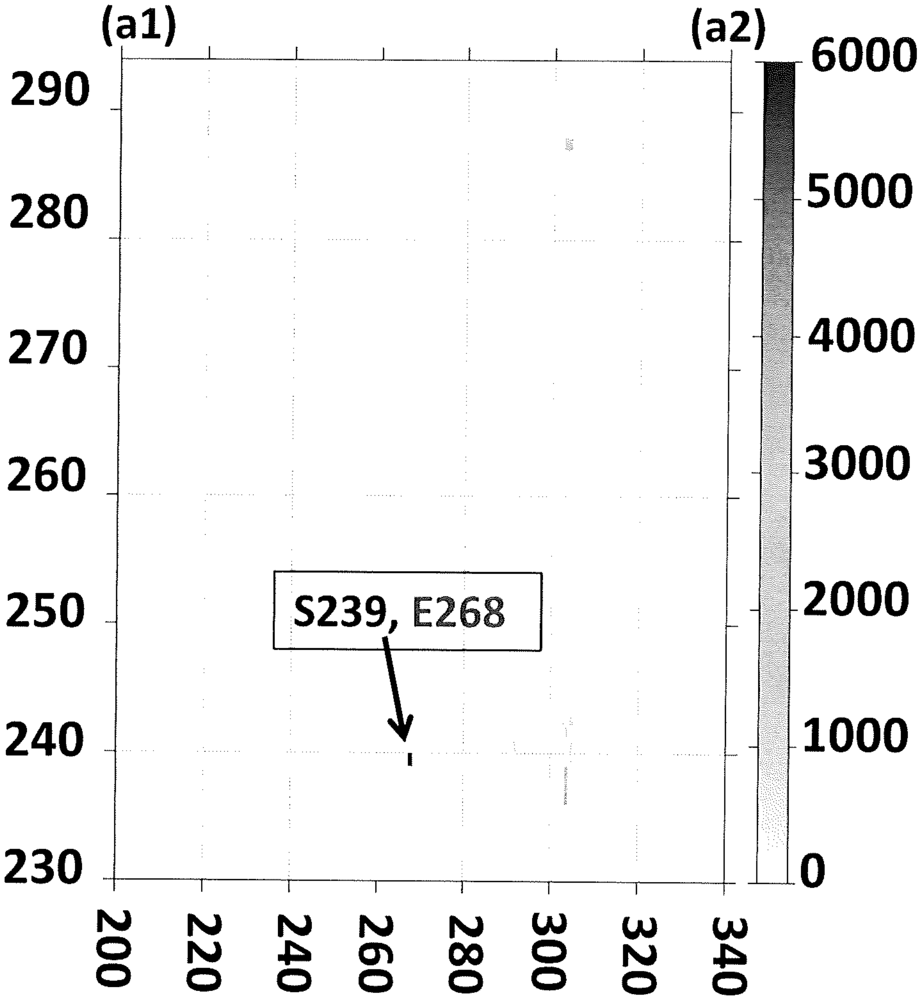

(a1)
(a2)
290
280
270
260
250
240
230
200
220
240
260
280
300
320
340
6000
5000
4000
3000
2000
1000
0
S239, E268

Figure 29

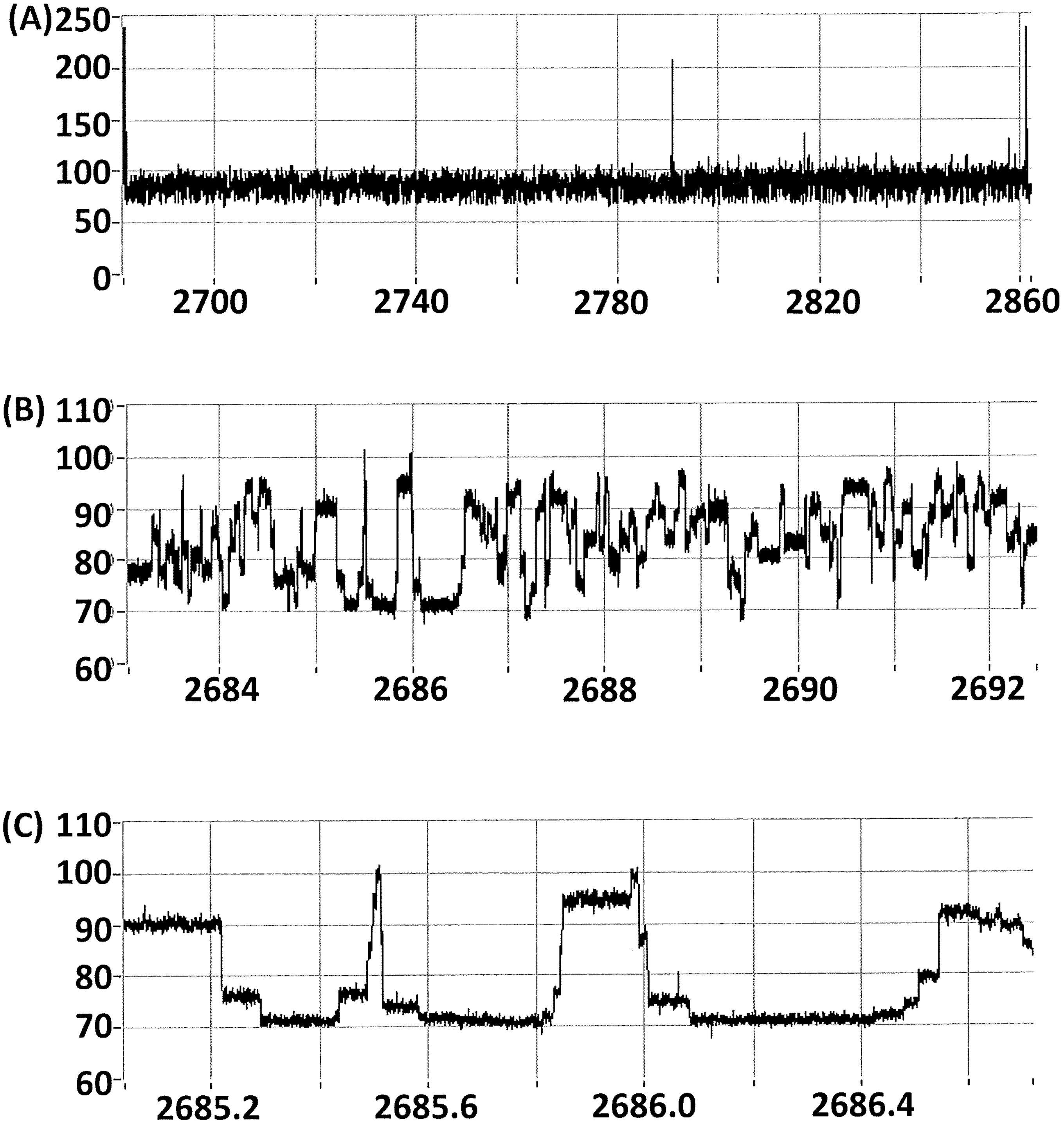

(A)
250
200
150
100
50
0
2700
2740
2780
2820
2860
(B)
110
100
90
80
70
60
2684
2686
2688
2690
2692
(C)
110
100
90
80
70
60
2685.2
2685.6
2686.0
2686.4

Figure 30

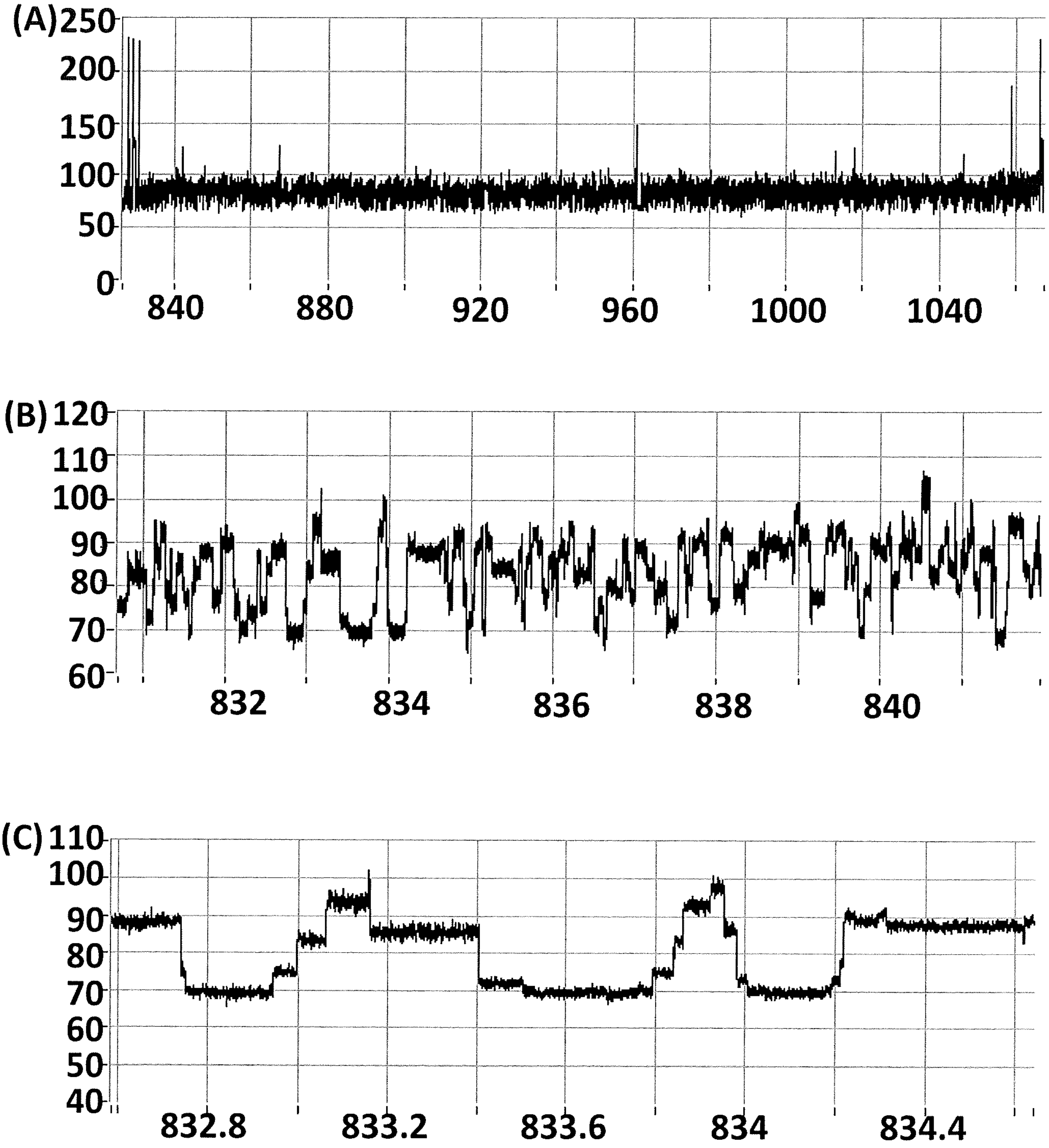

(A)
250
200
150
100
50
0
840
880
920
960
1000
1040
(B)
120
110
100
90
80
70
60
832
834
836
838
840
(C)
110
100
90
80
70
60
50
40
832.8
833.2
833.6
834
834.4

Figure 31

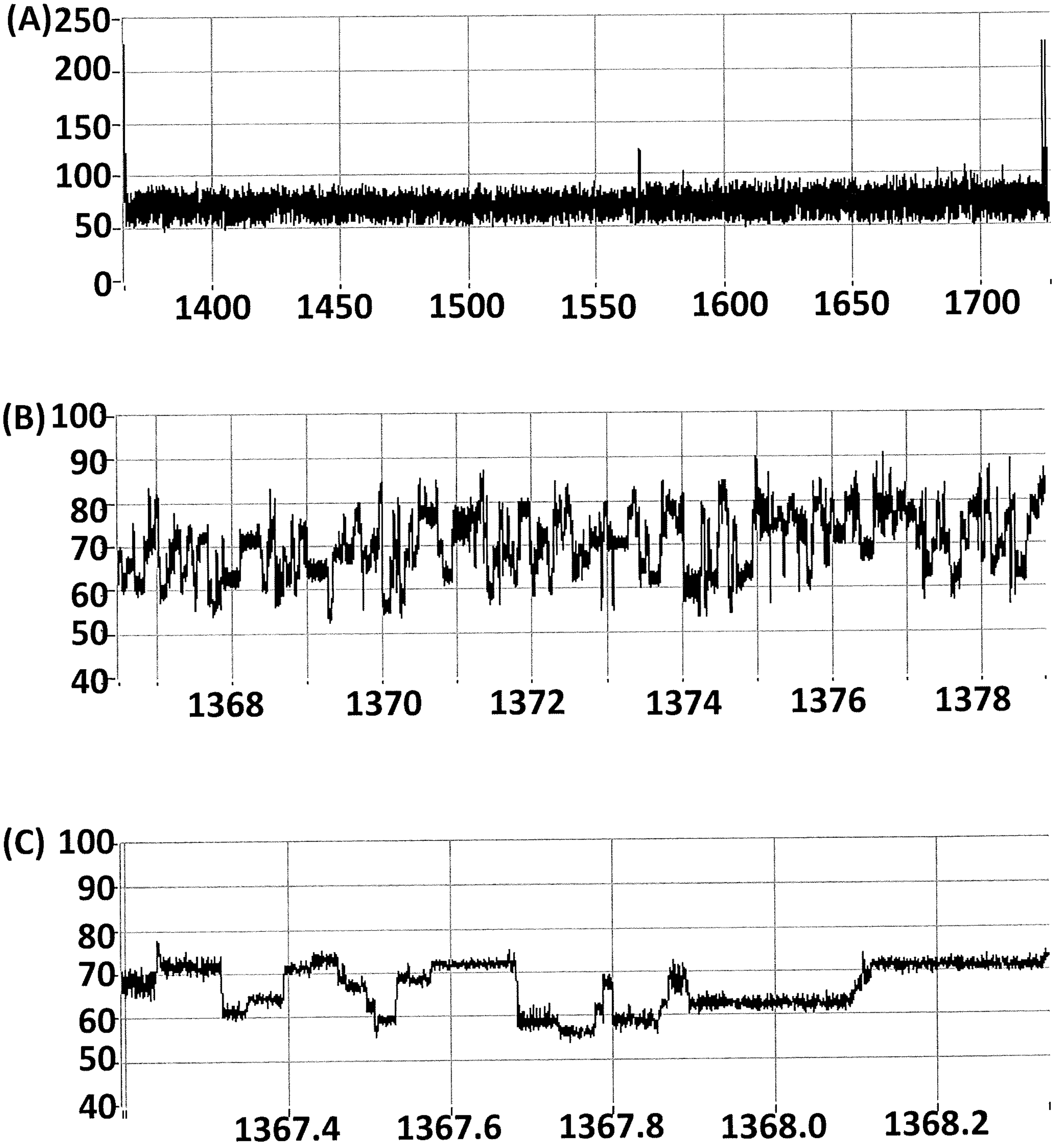

(A)
250
200
150
100
50
0
1400
1450
1500
1550
1600
1650
1700
(B)
100
90
80
70
60
50
40
1368
1370
1372
1374
1376
1378
(C)
100
90
80
70
60
50
40
1367.4
1367.6
1367.8
1368.0
1368.2

Figure 32

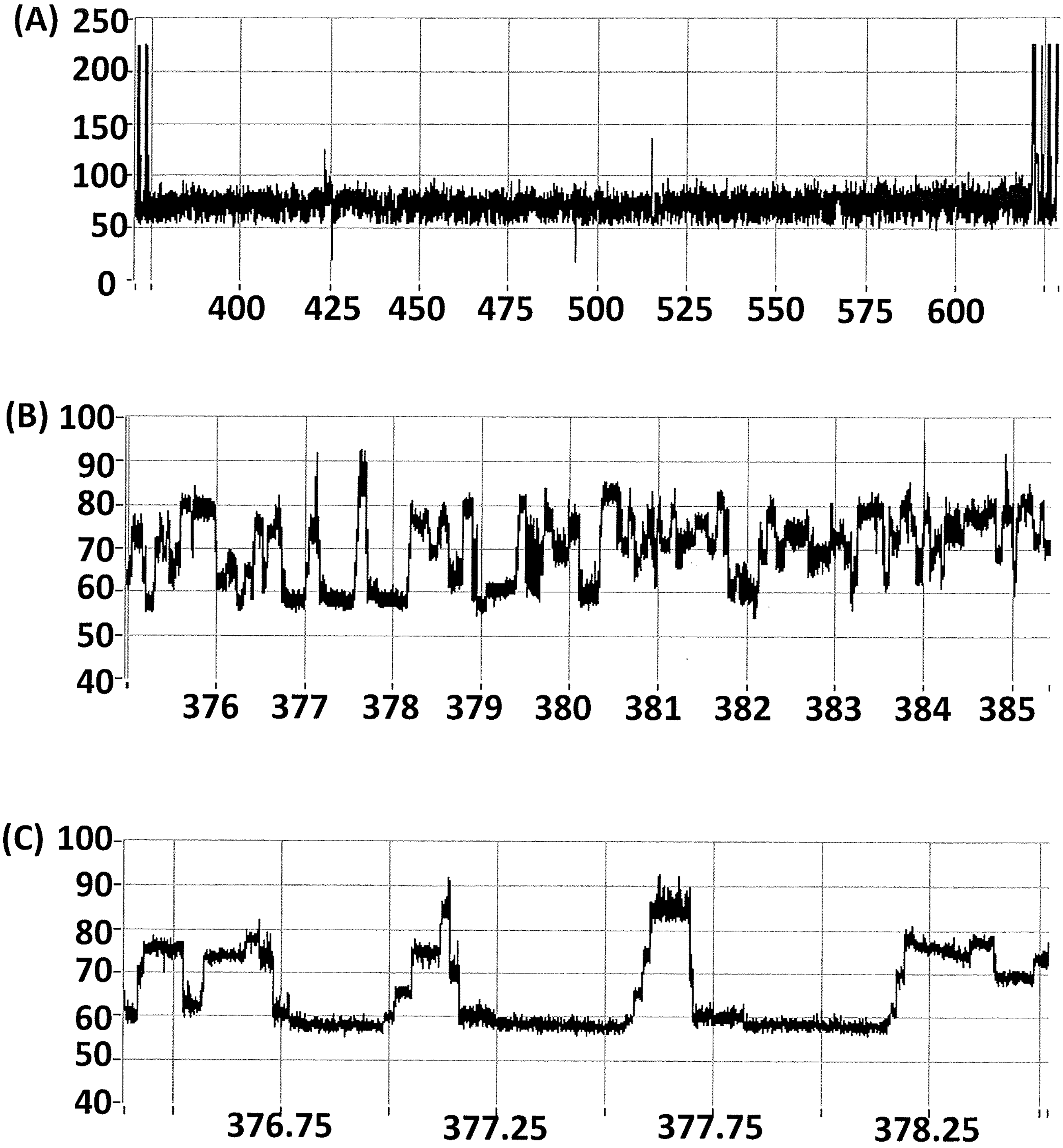

(A)
250
200
150
100
50
0
400
425
450
475
500
525
550
575
600
(B)
100
90
80
70
60
50
40
376
377
378
379
380
381
382
383
384
385
(C)
100
90
80
70
60
50
40
376.75
377.25
377.75
378.25

**Figure 33**

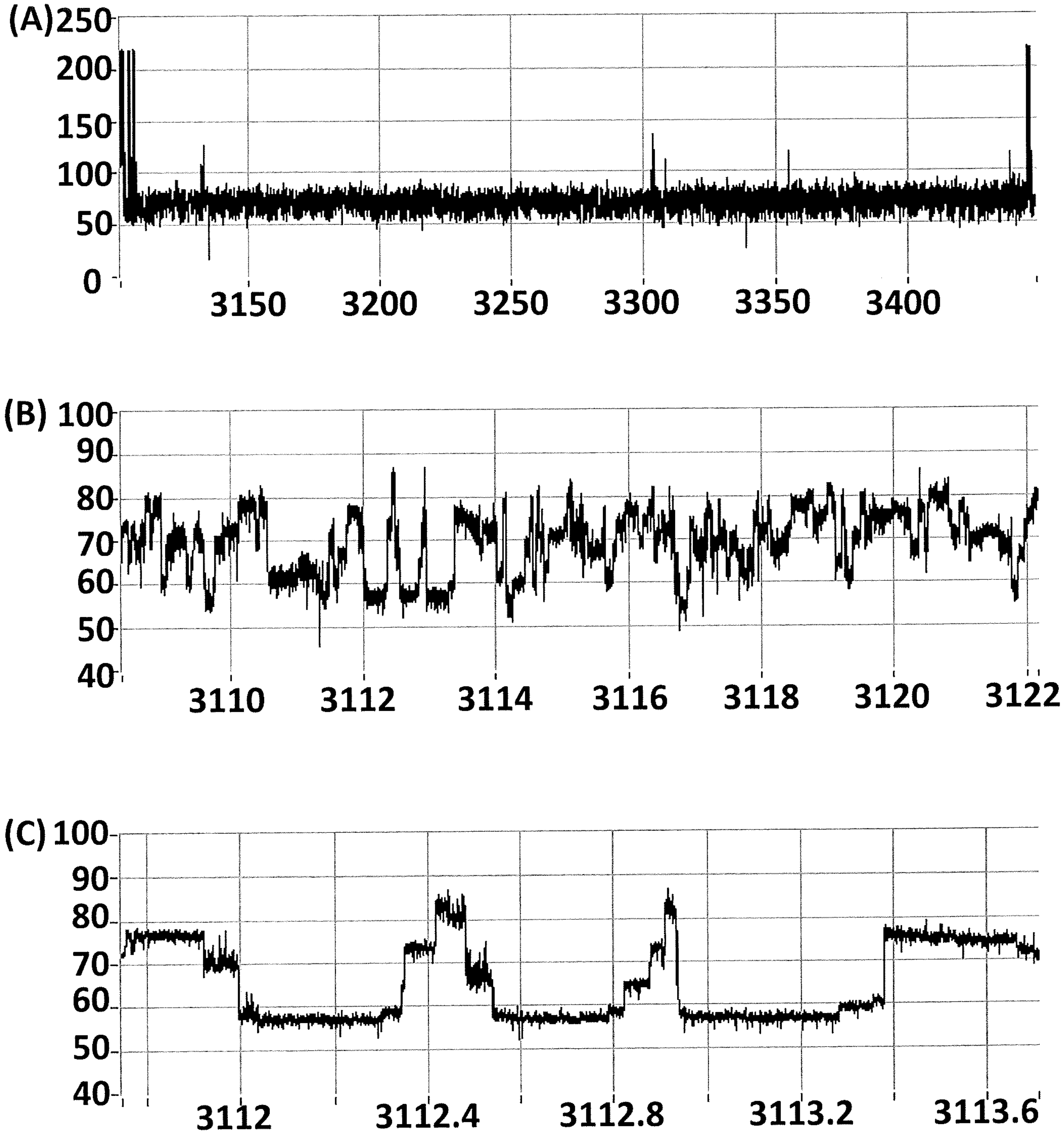

(A)
250
200
150
100
50
0
3150
3200
3250
3300
3350
3400
(B)
100
90
80
70
60
50
40
3110
3112
3114
3116
3118
3120
3122
(C)
100
90
80
70
60
50
40
3112
3112.4
3112.8
3113.2
3113.6

Figure 34

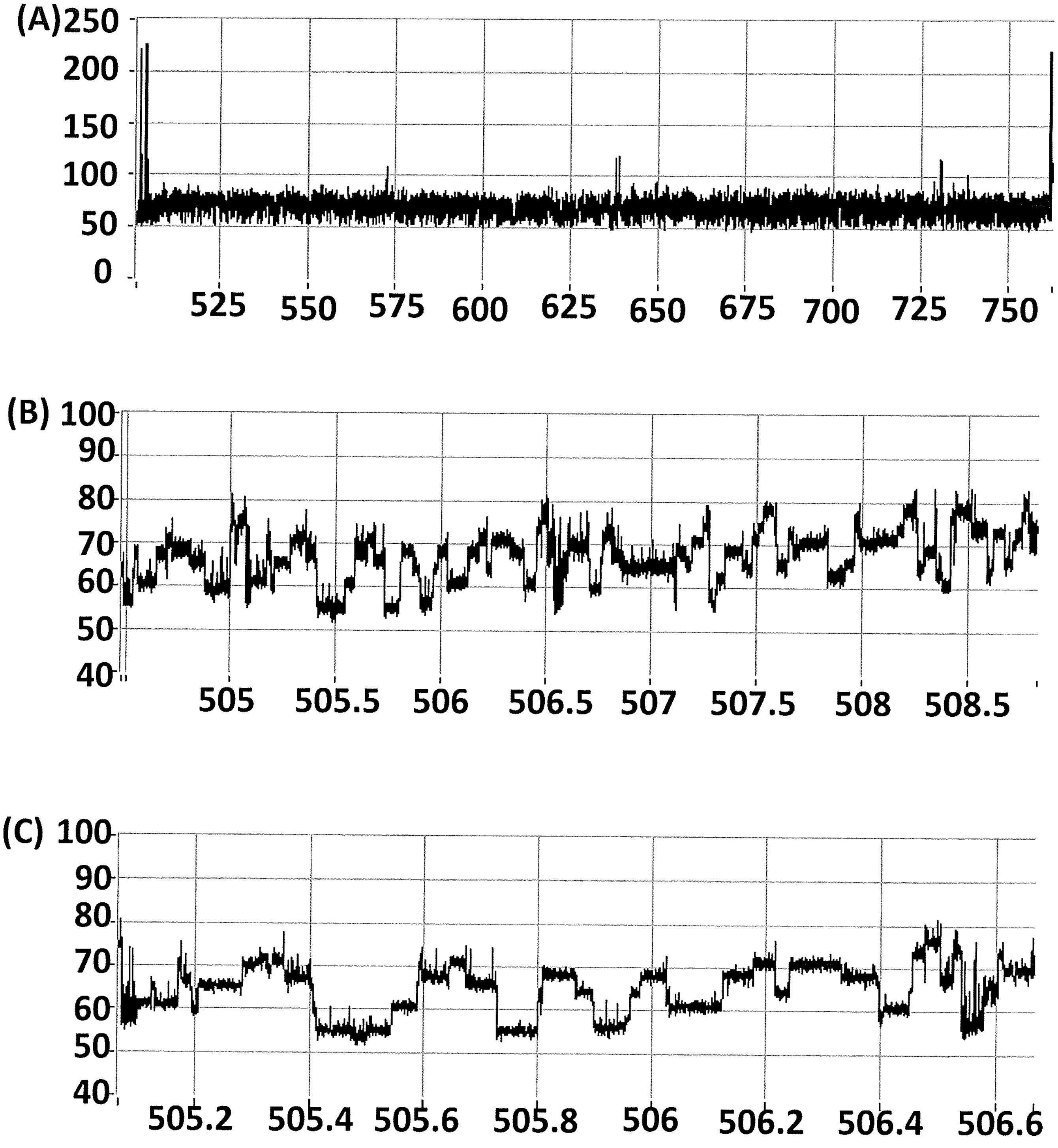

(A)
250
200
150
100
50
0
525
550
575
600
625
650
675
700
725
750
(B)
100
90
80
70
60
50
40
505
505.5
506
506.5
507
507.5
508
508.5
(C)
100
90
80
70
60
50
40
505.2
505.4
505.6
505.8
506
506.2
506.4
506.6

Figure 35

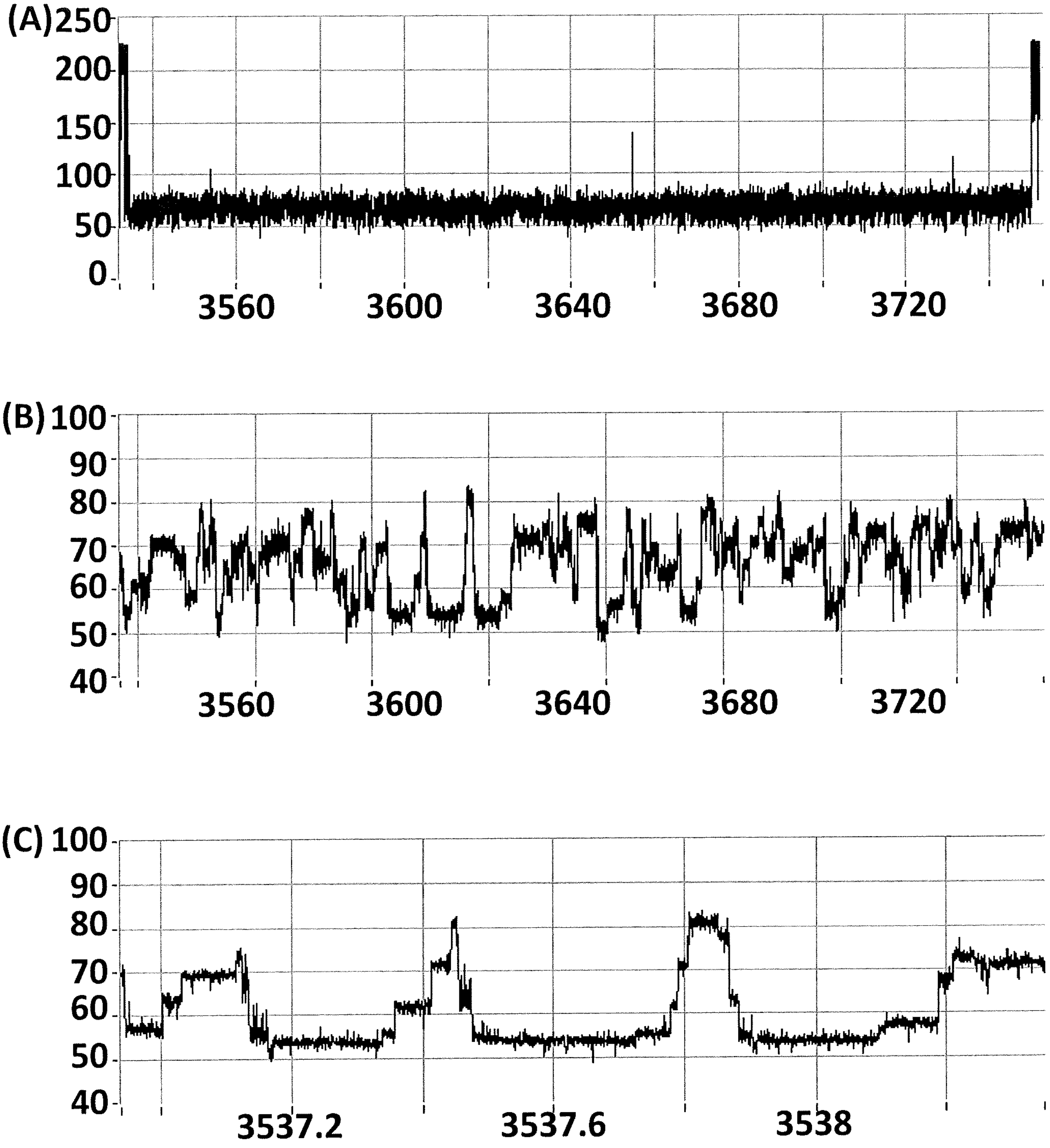

(A)
250
200
150
100
50
0
3560
3600
3640
3680
3720
(B)
100
90
80
70
60
50
40
3560
3600
3640
3680
3720
(C)
100
90
80
70
60
50
40
3537.2
3537.6
3538

Figure 36

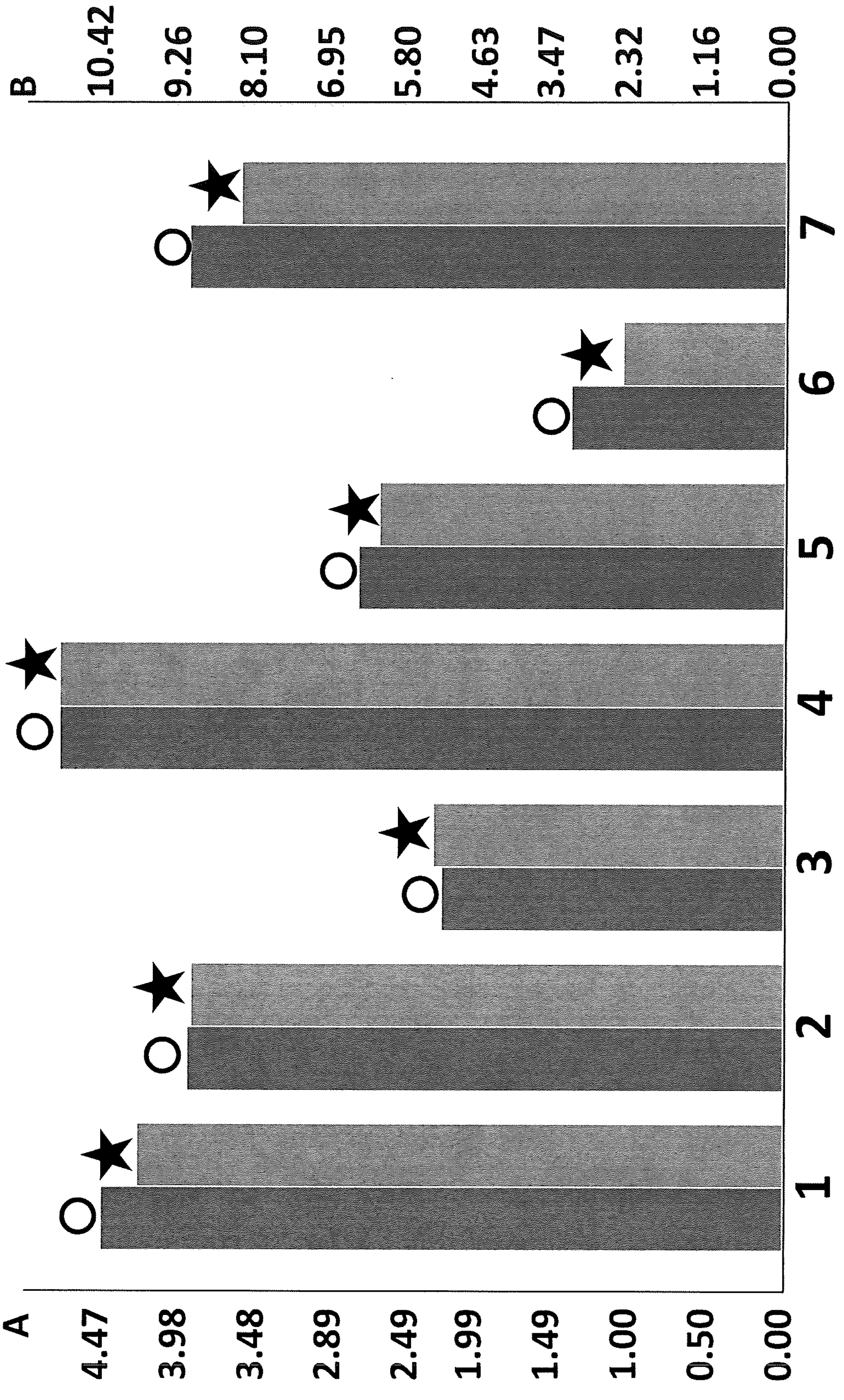
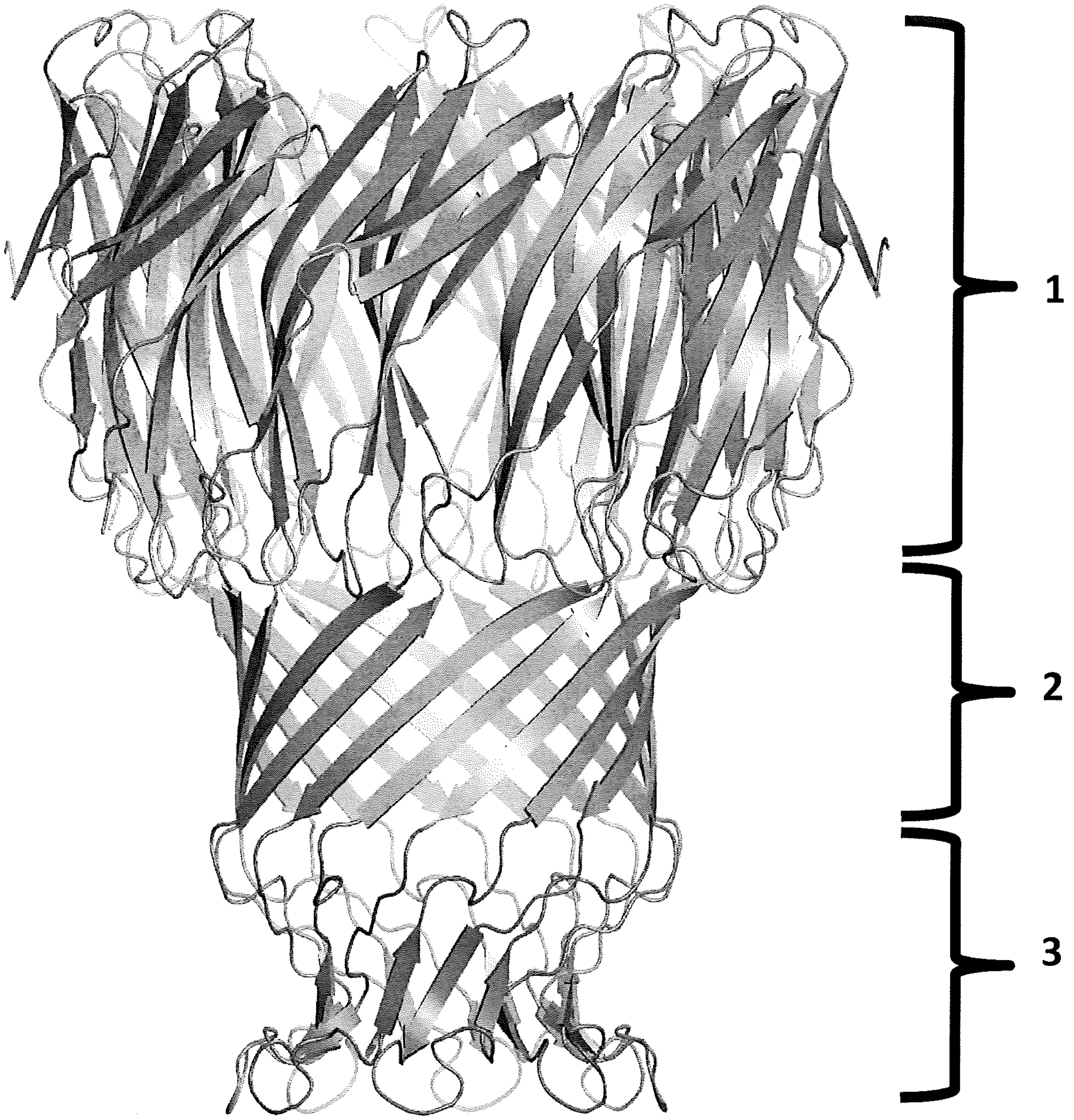

B
10.42
9.26
8.10
6.95
5.80
4.63
3.47
2.32
1.16
0.00
A
4.47
3.98
3.48
2.89
2.49
1.99
1.49
1.00
0.50
0.00
1
2
3
4
5
6
7

Figure 37



## INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/051291

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

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | ELIZABETH A MANRAO ET AL: "Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase", NATURE BIOTECHNOLOGY,<br>vol. 30, no. 4, 1 April 2012 (2012-04-01), pages 349-353, XP055028414,<br>ISSN: 1087-0156, DOI: 10.1038/nbt.2171<br>abstract<br>page 349, left-hand column, paragraph 2 - right-hand column, paragraph 1<br>page 350, left-hand column, paragraph 3<br>figure 1<br>-----<br>-/-- | 1-17, 27-33, 35-40, 44-47, 50,51 |

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

| | |
|---|---|
| Date of the actual completion of the international search<br>28 July 2015 | Date of mailing of the international search report<br>14/10/2015 |
| Name and mailing address of the ISA/<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Authorized officer<br>Surdej, Patrick |

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 2012/107778 A2 (OXFORD NANOPORE TECH LTD [GB]; CLARKE JAMES [GB]; HERON ANDREW JOHN [G) 16 August 2012 (2012-08-16)<br>page 2, line 9 - line 16<br>page 2, line 17 - line 24<br>page 6, line 31 - line 35<br>page 8, line 26 - page 9, line 8<br>page 9, line 19 - line 23<br>page 11, line 18 - line 20<br>page 46, line 30 - page 47, line 7<br>page 56, line 3 - line 8<br>example 4<br>table 16<br>----- | 1-17, 27-33, 35-40, 44-47, 50,51 |
| X | WO 2013/098562 A2 (OXFORD NANOPORE TECH LTD [GB]) 4 July 2013 (2013-07-04)<br>page 1, line 4 - line 6<br>page 5, line 9 - line 14<br>page 14, line 9 - line 22<br>page 55, line 20 - line 23<br>page 57, line 28 - page 59, line 14<br>page 61, line 17 - line 24<br>example 1<br>----- | 1-17, 27-33, 35-40, 44-47, 50,51 |
| X | WO 2014/013262 A1 (OXFORD NANOPORE TECH LTD [GB]) 23 January 2014 (2014-01-23)<br>page 1, line 4 - line 8<br>page 1, line 27 - page 2, line 10<br>page 5, line 30 - page 6, line 12<br>page 13, line 24 - line 29<br>page 68, line 28 - page 69, line 8<br>page 95, line 23 - page 97, line 15<br>example 9<br>----- | 1-17, 27-33, 35-40, 44-47, 50,51 |
| A | WO 2010/034018 A2 (UNIV WASHINGTON [US]; UAB RESEARCH FOUNDATION [US]; GUNDLACH JENS H [U) 25 March 2010 (2010-03-25)<br>page 4, line 19 - line 28<br>page 20, line 31 - page 21, line 16<br>page 23, line 3 - line 20<br>page 53, line 13 - page 54, line 22<br>page 66, line 29 - page 67, line 19<br>example 6<br>figure 1<br>----- | 1-17, 27-33, 35-40, 44-47, 50,51 |

International application No.
PCT/GB2015/051291

# INTERNATIONAL SEARCH REPORT

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

**Remark on Protest**

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

   A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein the part of the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acid at positions 12, 14, 48, 52, 53, 54, 55, 56, 57, 58, 59, 60, 134, 135, 136, 137, 138, 139, 169 and170 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof

   \-\-\-

2. claims: 1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

   A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein the part of the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acid at positions 12, 14, 52, 54, 56, 57, 59, 134, 136, 138, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof

   \-\-\-

3. claims: 1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

   A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein the part of the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acid at positions 12, 14, 56, 57, 59, 134, 136, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof

   \-\-\-

4. claims: 1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

   A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein the part of the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acid at positions 56, 57, 59, 134, 136, 139 and 169 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof

   \-\-\-

5. claims: 1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

   A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein the part of the transmembrane pore which interacts with the polynucleotide binding protein comprises the amino acid at positions 56, 57, 59, 134 and 139 in SEQ ID NO: 2 or at the corresponding positions in the variant thereof

---

6-7. claims: 1-17, 27-33, 35-40, 44-47, 50, 51(all partially)

A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein, respectively, said transmembrane pore is a variant of SEQ ID NO: 2 (a) 2, 4, 6, 8 or 10 of the amino acids at positions 72 to 82 of SEQ ID NO: 2 have been deleted or (b) 2, 4, 6, 8 or 10 of the amino acids at positions 111 to 121 of SEQ ID NO: 2 have been deleted

---

8. claims: 18, 19(completely); 1-10, 27-31, 34-40, 44-47, 50, 51(partially)

A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein said transmembrane pore which comprises seven monomers comprising the sequence shown in SEQ ID NO: 4 or a variant thereof

---

9. claims: 20, 21(completely); 1-10, 27-31, 34-40, 44-47, 50, 51(partially)

A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or transmembrane pore, wherein said transmembrane pore which comprises at least one monomer comprising the sequence shown in SEQ ID NO: 36 or a variant thereof

---

10. claims: 22-24(completely); 1-10, 27-31, 41-46, 48, 49(partially)

A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or a polynucleotide binding protein, wherein said polynucleotide binding protein comprises comprises the sequence shown in SEQ ID NO: 24 or a variant thereof

---

**FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210**

11. claims: 25, 26(completely); 1-10, 27-31, 41-46, 48, 49, 51(partially)

A method of improving the movement of a target polynucleotide with respect to a transmembrane pore when the movement is controlled by a polynucleotide binding protein or a polynucleotide binding protein, wherein said polynucleotide binding protein comprises the sequence shown in SEQ ID NO: 9 or a variant thereof

---

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2015/051291

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2012107778 A2 | 16-08-2012 | CA 2826374 A1 | 16-08-2012 |
| | | CN 103460040 A | 18-12-2013 |
| | | EP 2673638 A2 | 18-12-2013 |
| | | JP 2014506575 A | 17-03-2014 |
| | | KR 20140049511 A | 25-04-2014 |
| | | US 2014186823 A1 | 03-07-2014 |
| | | WO 2012107778 A2 | 16-08-2012 |
| WO 2013098562 A2 | 04-07-2013 | AU 2012360244 A1 | 24-07-2014 |
| | | CA 2861808 A1 | 04-07-2013 |
| | | CN 104126018 A | 29-10-2014 |
| | | EP 2798084 A2 | 05-11-2014 |
| | | JP 2015503917 A | 05-02-2015 |
| | | KR 20140108706 A | 12-09-2014 |
| | | US 2014335512 A1 | 13-11-2014 |
| | | WO 2013098562 A2 | 04-07-2013 |
| WO 2014013262 A1 | 23-01-2014 | AU 2013291765 A1 | 05-02-2015 |
| | | CA 2879355 A1 | 23-01-2014 |
| | | EP 2875152 A1 | 27-05-2015 |
| | | US 2015218629 A1 | 06-08-2015 |
| | | WO 2014013262 A1 | 23-01-2014 |
| WO 2010034018 A2 | 25-03-2010 | CA 2774710 A1 | 25-03-2010 |
| | | CN 102216783 A | 12-10-2011 |
| | | CN 104710519 A | 17-06-2015 |
| | | EP 2344891 A2 | 20-07-2011 |
| | | US 2012055792 A1 | 08-03-2012 |
| | | US 2014256913 A1 | 11-09-2014 |
| | | US 2014308662 A1 | 16-10-2014 |
| | | US 2014309402 A1 | 16-10-2014 |
| | | WO 2010034018 A2 | 25-03-2010 |