



- (51) International Patent Classification:
C07K 14/245 (2006.01) C12Q 1/6869 (2018.01)
- (21) International Application Number:
PCT/GB2018/051191
- (22) International Filing Date:
03 May 2018 (03.05.2018)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
1707122.6 04 May 2017 (04.05.2017) GB
- (71) Applicants: **OXFORD NANOPORE TECHNOLOGIES LIMITED** [GB/GB]; Gosling Building, Edmund Halley Road, Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB). **VIB VZW** [BE/BE]; Rijnvisschestraat 120, B-9052 Gent (BE). **VRIJE UNIVERSITEIT BRUSSEL** [BE/BE]; Pleinlaan 2, B-1050 Brussels (BE).
- (72) Inventors: **JAYASINGHE, Lakmal Nishantha**; Oxford Nanopore Technologies Limited, Gosling Building, Edmund Halley Road Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB). **WALLACE, Elizabeth Jayne**; Oxford Nanopore Technologies Limited, Gosling Building, Edmund Halley Road Oxford Science Park, Oxford Ox-

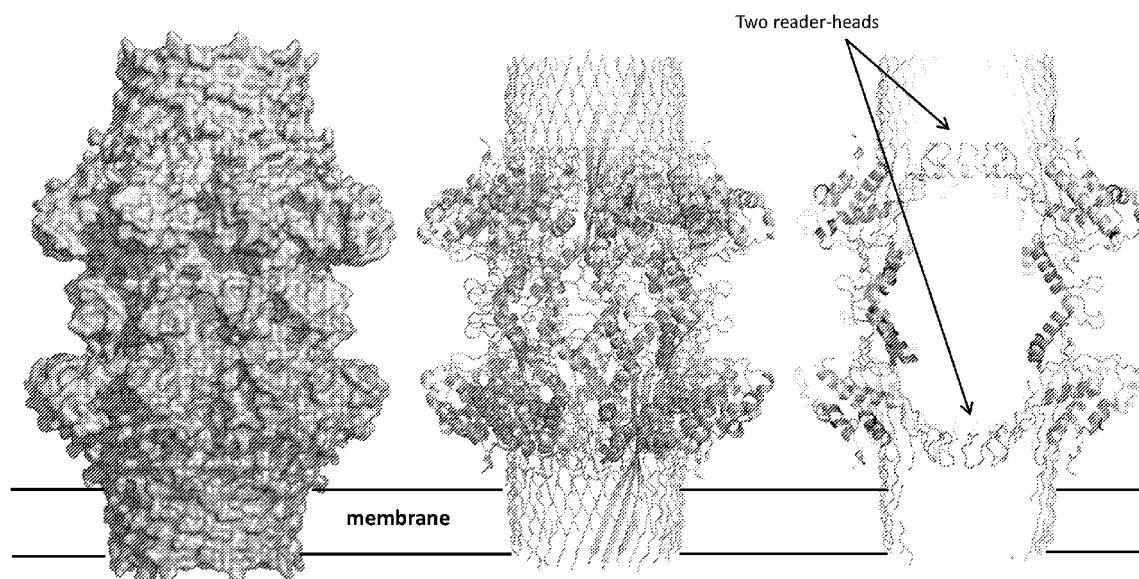
fordshire OX4 4DQ (GB). **SINGH, Pratik Raj**; Oxford Nanopore Technologies Limited, Gosling Building, Edmund Halley Road Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB). **HAMBLEY, Richard George**; Oxford Nanopore Technologies Limited, Gosling Building, Edmund Halley Road Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB). **JORDAN, Michael**; Oxford Nanopore Technologies Limited, Gosling Building, Edmund Halley Road Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB). **REMAUT, Han**; VIB vzw, Rijnvisschestraat 120, 9052 Zwijnaarde (BE).

(74) Agent: **TUXWORTH, Pamela Mary**; J A KEMP, 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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, 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,

(54) Title: TRANSMEMBRANE PORE CONSISTING OF TWO CSGG PORES

Figure 1



(57) Abstract: Provided is a method of characterising a polynucleotide using a transmembrane pore, wherein the pore is a double pore comprising a first Csg G pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof.



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

TRANSMEMBRANE PORE CONSISTING OF TWO CSGG PORES

Field

5 The invention relates to CsgG pores and to methods of analyte detection and characterisation using CsgG pores.

Background

10 Nanopore sensing is an approach to sensing that relies on the observation of individual binding or interaction events between analyte molecules and a receptor. Nanopore sensors can be created by placing a single pore of nanometer dimensions in an insulating membrane and measuring voltage-driven ionic transport through the pore in the presence of analyte molecules. The identity of an analyte is revealed through its distinctive current signature, notably the duration and extent of current block and the variance of current levels. Such nanopore sensors are commercially available, such as the MinION™ device sold by Oxford Nanopore
15 Technologies Ltd, comprising an array of nanopores integrated with an electronic chip. Nanopore sensing has the potential to provide rapid and cheap nucleic acid sequencing by reducing the quantity of nucleotide and reagents required.

Two of the essential components of sequencing nucleic acids using nanopore sensing are (1) the control of nucleic acid movement through the pore and (2) the discrimination of
20 nucleotides as the nucleic acid polymer is moved through the pore.

CsgG is a pore from *Escherichia coli* that has been proposed for use as a nanopore for detecting and characterising analytes. Mutations to the wild-type CgG pore that improve the properties of the pore in this context have also been disclosed (WO2016/034591).

Summary

25 The inventors have demonstrated that double pores consisting of two CsgG pores in series can be used to detect and characterise analytes, such as polynucleotides. The double pores may be used in particular to facilitate characterization of polynucleotides that contain at least one homopolymeric stretch, i.e. several consecutive copies of the same nucleotide. Such double
30 pores are provided, as are novel monomers that may be used to produce double pores for use in such methods. The monomers comprise amino acid residues that when assembled into a double pore strengthen the interactions between the two pores, amino acid residues that inhibit ions from exiting the pore at the junction between the two pores, amino acid residues that facilitate

interaction of a negatively charged analyte, such as a polynucleotide, with one or both of the barrel regions of the pore and/or amino acid residues that increase the length of one or both of the two narrowest parts of the double pore, i.e. the constrictions in the barrels of the double pore. Also provided are CsgG pores comprising the novel monomers and the use of such pores in methods of characterising analytes.

Previously, several mutations in CsgG pores have sharpened the reader head in order to call bases accurately with available algorithms and analysis tools. Advancements in the algorithms and analysis tools means, we are not restricted to sharp reader heads anymore. The inventors have recognised that new algorithms can deal with longer reader heads which provide additional information that may be lacking in short sharp reader heads. Therefore, the inventors have reengineered the CsgG constriction to elongate the reader head. This may also help to call homopolymers with better accuracies.

In particular, the following are provided:

- A method of characterising a polynucleotide using a transmembrane pore, wherein the pore is a double pore comprising a first CsgG pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof.

- A double pore comprising a first CsgG pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof, wherein:

(i) the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a homooligomer, and the first CsgG pore, or homologue thereof, comprises monomers that have a different amino acid sequence from the monomers of which the second CsgG pore, or homologue thereof, is comprised;

(ii) the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a homooligomer, and the first CsgG pore, or homologue thereof, and/or the second CsgG pore, or homologue thereof is not a wild-type pore;

(iii) the first CsgG pore, or homologue thereof, is a heterooligomer and the second CsgG pore, or homologue thereof, is a homooligomer;

(iv) the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a heterooligomer; or

(v) the first CsgG pore, or homologue thereof, is a heterooligomer and the second CsgG pore, or homologue thereof, is a heterooligomer.

- A CsgG monomer, or a monomer of a CsgG homologue, comprising:
 - (i) a cysteine residue at a position corresponding to R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2;
 - (ii) a residue at a position corresponding to any one or more of R97, Q100, I107, R110, E101, N102 and L113 of SEQ ID NO: 2, which residue is more hydrophobic than the residue present at the corresponding position of SEQ ID NO: 2, or in the amino acid sequence of a wild type CsgG homologue, wherein the residue at the position corresponding to R97 and/or I107 is M, the residue at the position corresponding to R110 is I, L, V, M, W or Y, and/or the residue at the position corresponding to E101 or N102 is V or M;
 - (iii) a residue at a position corresponding to any one or more of A98, A99, T104, V105, L113, Q114 and S115 of SEQ ID NO: 2 which is bulkier than the residue present at the corresponding position of SEQ ID NO: 2, or in the amino acid sequence of a wild type CsgG homologue, wherein the residue at the position corresponding to T104 is L, M, F, W, Y, N, Q, D or E, the residue at the position corresponding to L113 is M, F, W, Y, N, G, D or E and/or the residue at the position corresponding to S115 is M, F, W, Y, N, Q or E; and/or
 - (iv) a residue in the barrel region of the pore at a position corresponding to any one or more of D149, E185, D195, E210 and E203 which has less negative charge than the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer, wherein the residue at the position corresponding to D149, E185, D195 and/or E203 is K.
- A construct comprising two or more covalently attached CsgG monomers, wherein at least one of the monomers is a monomer as disclosed herein;
- A polynucleotide which encodes a monomer as disclosed herein or a construct as disclosed herein.
- A pore comprising at least one monomer as disclosed herein or a construct as disclosed herein.
- A method for determining the presence, absence or one or more characteristics of a target analyte, comprising: contacting the target analyte with a double pore as disclosed herein or a pore as disclosed herein such that the target analyte moves with respect to the pore; and taking one or more measurements as the analyte moves with respect to the pore and thereby determining the presence, absence or one or more characteristics of the analyte.

- Use of a double pore as disclosed herein or a pore as disclosed herein to determine the presence, absence or one or more characteristics of a target analyte.
 - A kit for characterising a target analyte comprising (a) a double pore as disclosed herein or a pore as disclosed herein and (b) the components of a membrane.
- 5 - An apparatus for characterising target analytes in a sample, comprising an array of double pores as disclosed herein or an array of pores as disclosed herein in a plurality of membranes.
- A method of producing a monomer as disclosed herein or a construct as disclosed herein, comprising expressing a polynucleotide as disclosed herein in a suitable host cell and thereby
- 10 producing a monomer as disclosed herein or a construct as disclosed herein.

Description of the Figures

Figure 1 shows the structure of a double pore comprising two CsgG pores in a tail to tail orientation. The two reader heads are indicated.

15 Figure 2 shows holes in the walls of a CsgG double pore. The inventors have produced data suggesting that double pore current is less than half the single pore current (at higher voltages). The inventors have proposed that this could be due to current leak from side pockets at the interface of the two pores. These gaps can be filled in by changing one or more amino acid residues in this area to bulkier amino acid residues.

20 Figure 3 shows the structure of part of the interface between two CsgG pores in a double pore (dimer). The mutations are shown in a pore that comprises Y51A and F56Q mutations (AQ = CP1-(WT-Y51A/F56Q-StrepII(C))₉). The indicated Cys mutant pairs may form S-S bonds.

Figure 4 shows the structure of part of a double pore with a single stranded DNA molecule inserted in the pore. There are approximately 15 nucleotides between the two

25 constrictions (reader heads). The two reader-heads are separated by a non-DNA interacting region.

Figure 5A shows the cross section of a CsgG pore showing the constriction (reader head) with a single stranded DNA inserted.

Figure 5B shows the cross section of a wild type CsgG pore in which the three main amino acid residues, F56 (red), N55 (blue) and Y51 (green), are indicated. The constriction is

30 located within the barrel (at the top) in a relatively unstructured loop. The reader head can be elongated either by mutations at existing positions or by inserting additional amino acid residues.

For example, the reader head can be broadened by mutations at each of the three indicated positions and/or by mutations at the 52, 53 and 54 positions.

Figure 5B shows the positions of the residues from K49 to F56 in a monomer of the CsgG pore. 51 can be moved further down by increasing the length of the loop in between 51 and 55. New amino acid residues can be inserted between 51 and 52, 52 and 53, 53 and 54 or 54 and 55. For example, 1, 2, 3 or more amino acid residues may be inserted. To keep the flexible nature of the loop, A/S/G/T can be inserted. To add a kink to the loop P can be inserted. New A amino acid residues could contribute to the signal (e.g. S/T/N/Q/M/F/W/Y/V/I). Similarly, new amino acids can be inserted between 55 and 56 (1 or 2 or more). They can be any of the above amino acids. Y51 can also move downwards by inserting amino acids to both sides of the loop above Y51. For example S or G or SG or SGG or SGS or GS or GSS or GSG or other suitable amino acid (1 or 2 or more) can be inserted (i) between (49 and 50) and between (52 and 53); (ii) between (50 and 51) and between (51 and 52); (iii) combinations of 1 and 2; or (iv) any of (i) to (iii) can be combined with other insertions (e.g. insertions between 55 and 56).

Figure 6 shows results obtained with the sharpened reader head of a baseline pore comprising monomers having the sequence shown in SEQ ID NO: 2 in which the following substitutions have been made: Y51A; F56Q; K94Q; R97W; and R192D, and in which V105 to I107 have been deleted. The Y51A and F56Q sharpen the reader head. **A.** Discrimination of bases against the reader head position. When dominant current level contributions are approximated to 5 bases within the reader head, middle base (3rd) shows the major contribution to the signal whilst 1st, 2nd, 4th and 5th bases contribute at a relatively lower level (1st base is towards trans side of the pore and the 5th base is towards the cis side of the pore). **B.** Separation of the four nucleotides at each reader head position when dominant current level contributions are approximated to 5 bases within the reader head. **C.** Example squiggle from the baseline pore.

Figure 7 shows results obtained with the broader reader head of the baseline pore containing an additional N55V substitution. **A.** Discrimination of bases against the reader head position. When dominant current level contributions are approximated to 5 bases within the reader head, 3rd and 4th bases show a major contribution to the signal whilst 1st, 2nd and 5th bases contribute at a relatively lower level (1st base is towards trans side and the 5th base is towards the cis side). **B.** Separation of the four nucleotides at each reader head position when dominant current level contributions are approximated to 5 bases within the reader head. **C.** Example squiggle from the N55V pore.

Figure 8 shows results obtained with the broader reader head of the baseline pore containing an A51Q substitution. **A.** Discrimination of bases against the reader head position. When dominant current level contributions are approximated to 5 bases within the reader head, 2nd and 3rd bases show the biggest contribution to the signal whilst 1st, 4th and 5th bases contribute at a relatively lower level (1st base is towards trans side of the pore and the 5th base is towards the cis side of the pore). **B.** Separation of the four nucleotides at each reader head position when dominant current level contributions are approximated to 5 bases within the reader head. **C.** Example squiggle from the A51Q pore.

Figure 9 shows results obtained with the broader reader head of the baseline pore containing a Q56V substitution. **A.** Discrimination of bases against the reader head position. When dominant current level contributions are approximated to 5 bases within the reader head, 2nd and 3rd bases show the biggest contribution to the signal whilst 1st, 4th and 5th bases contribute at a relatively lower level (1st base is towards trans side of the pore and the 5th base is towards the cis side of the pore). **B.** Separation of the four nucleotides at each reader head position when dominant current level contributions are approximated to 5 bases within the reader head. **C.** Example squiggle from the Q56V pore.

Figure 10A is a comparison of the discrimination of bases against the reader head position between the baseline pore (Figure 6) and the N55V pore (Figure 7).

Figure 10B is a comparison of the example squiggles produced using the baseline pore (Figure 6) and the N55V pore (Figure 7) and shows that mutants give different squiggles to baseline.

Figure 11 shows the structures and reader heads of the baseline CsgG pore used in the Examples (A), a CsgG pore with an elongated reader head (B) and a double CsgG pore (C). Homopolymer basecalling is improved compared to the baseline when the elongated reader head pore or the double pore is used.

Figure 12 shows sequence alignments of the 21 CsgG homologues having the amino acid sequences shown in SEQ ID Nos 2 to 7 and 9 to 23. The C-terminal serine (S) of each of SEQ ID NOs: 2, 3, 5, 6, 7, 9 and 10 is not included in the alignment.

Figure 13 shows the same relative sequence alignments as Figure 12 with predicted alpha helical secondary structure regions additionally shaded.

Figure 14 shows the same relative sequence alignments as Figure 12 with predicted beta sheet secondary structure regions additionally shaded.

Description of the Sequence Listing

SEQ ID NO: 1 shows the codon optimised polynucleotide sequence encoding the wild-type CsgG monomer from *Escherchia coli* Str. K-12 substr. MC4100. This monomer lacks the signal sequence.

5 SEQ ID NO: 2 shows the amino acid sequence of the mature form of the wild-type CsgG monomer from *Escherchia coli* Str. K-12 substr. MC4100. This monomer lacks the signal sequence. The abbreviation used for this CsgG = CsgG-Eco.

10 SEQ ID NO: 3 shows the amino acid sequence of YP_001453594.1: 1-248 of hypothetical protein CKO_02032 [*Citrobacter koseri* ATCC BAA-895], which is 99% identical to SEQ ID NO: 2.

SEQ ID NO: 4 shows the amino acid sequence of WP_001787128.1: 16-238 of curli production assembly/transport component CsgG, partial [*Salmonella enterica*], which is 98% to SEQ ID NO: 2.

15 SEQ ID NO: 5 shows the amino acid sequence of KEY44978.1|: 16-277 of curli production assembly/transport protein CsgG [*Citrobacter amalonaticus*], which is 98% identical to SEQ ID NO: 2.

SEQ ID NO: 6 shows the amino acid sequence of YP_003364699.1: 16-277 of curli production assembly/transport component [*Citrobacter rodentium* ICC168], which is 97% identical to SEQ ID NO: 2.

20 SEQ ID NO: 7 shows the amino acid sequence of YP_004828099.1: 16-277 of curli production assembly/transport component CsgG [*Enterobacter asburiae* LF7a], which is 94% identical to SEQ ID NO: 2.

SEQ ID NO: 8 shows the amino acid sequence of WP_006819418.1: 19-280 of transporter [*Yokenella regensburgei*], which is 91% identical to SEQ ID NO: 2.

25 SEQ ID NO: 9 shows the amino acid sequence of WP_024556654.1: 16-277 of curli production assembly/transport protein CsgG [*Cronobacter pulveris*], which is 89% identical to SEQ ID NO: 2.

30 SEQ ID NO: 10 shows the amino acid sequence of YP_005400916.1 :16-277 of curli production assembly/transport protein CsgG [*Rahnella aquatilis* HX2], which is 84% identical to SEQ ID NO: 2.

SEQ ID NO: 11 shows the amino acid sequence of KFC99297.1: 20-278 of CsgG family curli production assembly/transport component [*Kluyvera ascorbata* ATCC 33433], which is 82% identical to SEQ ID NO: 2.

SEQ ID NO: 12 shows the amino acid sequence of KFC86716.1|:16-274 of CsgG family curli production assembly/transport component [*Hafnia alvei* ATCC 13337], which is 81% identical to SEQ ID NO: 2.

5 SEQ ID NO: 13 shows the amino acid sequence of YP_007340845.1|:16-270 of uncharacterised protein involved in formation of curli polymers [*Enterobacteriaceae* bacterium strain FGI 57], which is 76% identical to SEQ ID NO: 2.

SEQ ID NO: 14 shows the amino acid sequence of WP_010861740.1: 17-274 of curli production assembly/transport protein CsgG [*Plesiomonas shigelloides*], which is 70% identical to SEQ ID NO: 2.

10 SEQ ID NO: 15 shows the amino acid sequence of YP_205788.1 : 23-270 of curli production assembly/transport outer membrane lipoprotein component CsgG [*Vibrio fischeri* ES114], which is 60% identical to SEQ ID NO: 2.

SEQ ID NO: 16 shows the amino acid sequence of WP_017023479.1: 23-270 of curli production assembly protein CsgG [*Aliivibrio logei*], which is 59% identical to SEQ ID NO: 2.

15 SEQ ID NO: 17 shows the amino acid sequence of WP_007470398.1: 22-275 of Curli production assembly/transport component CsgG [*Photobacterium* sp. AK15], which is 57% identical to SEQ ID NO: 2.

20 SEQ ID NO: 18 shows the amino acid sequence of WP_021231638.1: 17-277 of curli production assembly protein CsgG [*Aeromonas veronii*], which is 56% identical to SEQ ID NO: 2.

SEQ ID NO: 19 shows the amino acid sequence of WP_033538267.1: 27-265 of curli production assembly/transport protein CsgG [*Shewanella* sp. ECSMB14101], which is 56% identical to SEQ ID NO: 2.

25 SEQ ID NO: 20 shows the amino acid sequence of WP_003247972.1: 30-262 of curli production assembly protein CsgG [*Pseudomonas putida*], which is 54% identical to SEQ ID NO: 2.

SEQ ID NO: 21 shows the amino acid sequence of YP_003557438.1: 1-234 of curli production assembly/transport component CsgG [*Shewanella violacea* DSS12], which is 53% identical to SEQ ID NO: 2.

30 SEQ ID NO: 22 shows the amino acid sequence of WP_027859066.1: 36-280 of curli production assembly/transport protein CsgG [*Marinobacterium jannaschii*], which is 53% identical to SEQ ID NO: 2.

SEQ ID NO: 23 shows the amino acid sequence of CEJ70222.1: 29-262 of Curli production assembly/transport component CsgG [*Chryseobacterium oranimense* G311], which is 50% identical to SEQ ID NO: 2.

SEQ ID NO: 24 shows the amino acid sequence of StrepII(C).

5 SEQ ID NO: 25 shows the DNA sequence encoding the polypeptide Pro-CP1-Eco-*(Mutant-StrepII(C))*.

SEQ ID NO: 26 shows the amino acid sequence of the polypeptide Pro-CP1-Eco-*(Mutant-StrepII(C))*.

10 **Detailed Description**

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 only, and is not intended to be limiting.

15 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
20 monomers, reference to “a pore” includes two or more pores and the like.

In all of the discussion herein, the standard one letter codes for amino acids are used. These are as follows: alanine (A), arginine (R), asparagine (N), aspartic acid (D), cysteine (C), glutamic acid (E), glutamine (Q), glycine (G), histidine (H), isoleucine (I), leucine (L), lysine (K), methionine (M), phenylalanine (F), proline (P), serine (S), threonine (T), tryptophan (W),
25 tyrosine (Y) and valine (V). Standard substitution notation is also used, i.e. Q42R means that Q at position 42 is replaced with R.

In the paragraphs herein where different amino acids at a specific position are separated by the / symbol, the / symbol means “or”. For instance, Q87R/K means Q87R or Q87K.

In the paragraphs herein where different positions are separated by the / symbol, the / symbol means “and” such that Y51/N55 is Y51 and N55.
30

All amino-acid substitutions, deletions and/or additions disclosed herein are with reference to a mutant CsgG monomer comprising a variant of the sequence shown in SEQ ID NO: 2, unless stated to the contrary.

Reference to a mutant CsgG monomer comprising a variant of the sequence shown in SEQ ID NO: 2 encompasses mutant CsgG monomers comprising variants of sequences as set out in the further SEQ ID NOS as disclosed below. Amino-acid substitutions, deletions and/or additions may be made to CsgG monomers comprising a variant of the sequence other than shown in SEQ ID NO:2 that are equivalent to those substitutions, deletions and/or additions disclosed herein with reference to a mutant CsgG monomer comprising a variant of the sequence shown in SEQ ID NO:2.

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

Double Pores

Provided is a double pore comprising a first CsgG pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof.

The double pore may comprise a first CsgG pore, or homologue thereof, which is a homooligomer and a second CsgG pore, or homologue thereof, which is a homooligomer, wherein the first CsgG pore, or homologue thereof, comprises monomers that have a different amino acid sequence from the monomers of which the second CsgG pore, or homologue thereof, is comprised.

The double pore may comprise a first CsgG pore, or homologue thereof, which is a homooligomer and a second CsgG pore, or homologue thereof, which is a homooligomer, wherein the first CsgG pore, or homologue thereof, and/or the second CsgG pore, or homologue thereof, comprises monomers that comprises a non-naturally occurring amino acid sequence, i.e. wherein the sequence of the first pore and/or the second pore is not a wild-type sequence.

The double pore may comprise a first CsgG pore, or homologue thereof, which is a heterooligomer and a second CsgG pore, or homologue thereof, which is a homooligomer.

The double pore may comprise a first CsgG pore, or homologue thereof, which is a homooligomer and a second CsgG pore, or homologue thereof, which is a heterooligomer.

The double pore may comprise a first CsgG pore, or homologue thereof, which is a heterooligomer and a second CsgG pore, or homologue thereof, which is a heterooligomer.

The homooligomer may contain any number of monomers having identical amino acid sequences. The homooligomer 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 homooligomer preferably comprises eight or nine identical monomers. One or more, such as 2, 3, 4, 5, 6, 7, 8, 9

or 10, of the monomers may be chemically modified as discussed herein. The homooligomer may comprise any of the modified monomers, such as a monomer of the invention, or wild type monomers. The monomers may have the amino acid sequence shown in any one of SEQ ID NOs: 2 to 23 or may be a variant of any one of these sequences. The variant may have any one or more of the mutations described herein. Where the homooligomer comprises wild type monomers, the double pore provided herein comprises only one such homooligomer. The homooligomer may be paired with a different homooligomer in the double pore, which may comprise wild type or mutant monomers, such as the modified monomers, such as a monomer of the invention. The homooligomer may be paired with a heterooligomer in the double pore.

10 The heterooligomer may contain any number of monomers sufficient to form a pore. 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. The heterooligomer may comprise 2, 3, 4, 5, 6, 7, 8, 9 or 10 different monomers. At least one of the monomers in the heterooligomer differs from the others. All of the monomers in the heterooligomers may be different from one another. For example, each monomer may comprise a different mutation or group of mutations.

The heterooligomer may comprise at least one modified monomer, such as a monomer of the invention. All of the monomers (such as 10, 9, 8 or 7 of the monomers) may be monomers of the invention, wherein at least one of the monomers differs from the others. For example, the heterooligomer may comprise eight or nine monomers of the invention, wherein at least one of them differs from the others. They may all differ from one another.

At least one of the monomers may not be a monomer of the invention. For example, at least one of the monomers in the heterooligomer may be a mutant monomer which comprises mutations other than the mutations present in the monomers of the present invention. Suitable mutant monomers are known in the art, for example in WO2016/034591, PCT/GB2017/050569, PCT/GB2017/050570 and PCT/GB2017/050571. The mutant monomers disclosed in WO2016/034591, PCT/GB2017/050569, PCT/GB2017/050570 and PCT/GB2017/050571 are incorporated herein.

Any number of the monomers in the pore may not be a monomer of the invention, from 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 monomers. All of the monomers in at least one of the pores making up the double pore of the invention may not be monomers of the invention. The heterooligomer is preferably a nonomer that comprises seven or eight monomers of the invention, that may be the same or different, and one or two monomers which are not monomers of the invention, or is

an octamer that comprises six or seven monomers of the invention, that may be the same or different, and one or two monomers which is not a monomer of the invention.

The heterooligomer may comprise roughly equal numbers of a first monomer and a second monomer, such as 5 of one and 4 of the other, 4 of each, or 3 of one and 4 of the other.

5 The first and second monomers may be different modified monomers, such as different monomers of the invention, only one of the first and second monomers may be a monomer of the invention, or both the first and second monomers may be monomers that are not monomers of the invention.

10 The monomers in the heterooligomer are preferably approximately the same length or are the same length. The barrels of the 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 pore may comprise one or more monomers which are not monomers of the invention. CsgG monomers, and monomers of CsgG homologues, which are not monomers of the invention include wild type monomers comprising the amino acid sequence shown in SEQ
15 ID NO: 2 or in any one of SEQ ID NOs: 3 to 24. Variants of SEQ ID NOs: 2 to 23 may also be used. Such variants are typically at least 50% homologous to one or more of SEQ ID NOs: 2 to 23 over its entire 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
20 least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of any one of SEQ ID NOs: 2 to 23 over the entire sequence.

One or more, such as 2, 3, 4, 5, 6, 7, 8, 9 or 10, of the monomers may be chemically modified.

25 The first and second pores in the double pore typically both comprise the same number of monomers. For example, where the first pore is nonomeric, the second pore is typically also nonomeric. The double pore may comprise two pores each comprising 2, 3, 4, 5, 6, 7, 8, 9 or 10 monomers, preferably 7, 8 or 10 monomers, more preferably 9 monomers.

30 The double pore is typically a transmembrane pore. The double pore may be associated with the membrane such the first pore and the second pore form a continuous channel through the membrane. Typically the first pore present in the double pore crosses the membrane and the second pore is located on the *cis* or *trans* side of the membrane. In this embodiment, the barrel of the first pore is typically inserted in the membrane and the tail of the first pore protrudes from the membrane. The first pore and the second pore may be orientated in opposite directions. The

tail region of the first CsgG pore, or homologue thereof, may be adjacent to the tail region of the second CsgG pore, or homologue thereof. The tail of the first pore typically associates with the tail of the second pore.

5 In the double pore, the first CsgG pore, or homologue thereof, may be attached to the second CsgG pore, or homologue thereof, by hydrophobic interactions and/or by one or more disulphide bond. One or more, such as 2, 3, 4, 5, 6, 8, 9, for example all, of the monomers in the first pore and/or the second pore may be modified to enhance such interactions. This may be achieved in any suitable way.

10 At least one cysteine residue in the amino acid sequence of the first pore at the interface between the first and second pores may be disulphide bonded to at least one cysteine residue in the amino acid sequence of the second pore at the interface between the first and second pores. The cysteine residue in the first pore and/or the cysteine residue in the second pore may be a cysteine residue that is not present in the wild type CsgG monomer and/or in a wild type CsgG homologue monomer. Multiple disulphide bonds, such as from 2, 3, 4, 5, 6, 7, 8 or 9 to 16, 18,
15 24, 27, 32, 36, 40, 45, 48, 54, 56 or 63, may form between the two pores in the double pore. One or both the first or second pore may comprise at least one monomer, such as up to 8, 9 or 10 monomers, that comprises a cysteine residue at the interface between the first and second pores at a position corresponding to R97, I107, R110, Q100, E101, N102 and/or L113 of SEQ ID NO:
2.

20 At least one monomer in the first pore and/or at least one monomer in the second pore may comprise at least one residue at the interface between the first and second pores, which residue is more hydrophobic than the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer. For example, from 2 to 10, such as 3, 4, 5, 6, 7, 8 or 9, residues in the first pore and/or the second pore may be more
25 hydrophobic than the residues at the same positions in the corresponding wild type CsgG monomer, or wild type CsgG homologue. Such hydrophobic residues strengthen the interaction between the two pores in the double pore. The at least one residue at the interface between the first and second pores may be at a position corresponding to R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2. Where the residue at the interface in the wild type CsgG
30 monomer, or the wild type CsgG homologue monomer is R, Q, N or E, the hydrophobic residue is typically I, L, V, M, F, W or Y. Where the residue at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is I, the hydrophobic residue is typically L, V, M, F, W or Y. Where the residue at the interface in the wild type CsgG monomer, or the

wild type CsgG homologue monomer is L, the hydrophobic residue is typically I, V, M, F, W or Y.

The double pore may comprise one or more monomer that comprises one or more cysteine residue at the interface between the pores and one or more monomer that comprises one or more introduced hydrophobic residue at the interface between the pores, or may comprise one or more monomer that comprises such cysteine residues and such hydrophobic residues. For example, one or more, such as any 2, 3, or 4, of the positions in the monomer corresponding to the positions at R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2 may comprise a cysteine (C) residue and one or more, such as any 2, 3 or 4, of the positions in the monomer corresponding to the positions at R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2 may comprise a hydrophobic residue, such as I, L, V, M, F, W or Y.

The present inventors have identified holes in the join between two wild-type CsgG pores that assemble in a tail to tail orientation to form a double pore. The inventors have recognised that any gaps present between the first and second pore may allow current, i.e. ions, to pass into or out of the pore at the junction between the first and second pore. Such passage of ions would be detrimental when the pore is used to detect or characterise an analyte. Accordingly, these holes may be closed in a double pore, i.e. the double pore typically does not include any holes at the join between the first and second pore. The channel provided by the double pore typically does not leak ions. It is a channel with continuous walls. The structure of the double pore provides a solid wall around the channel formed by the first and second pores.

The double pore according may contain bulky residues at one or more, such as 2, 3, 4, 5, 6 or 7, positions in the tail region, which residues are typically at the interface between the first and second pores and are bulkier than the residues present at the corresponding positions in the wild type CsgG monomer, or the wild type CsgG homologue monomer. The bulk of these residues prevents holes from forming in the walls of the pore at the interface between the first and second pore in the double pore. The at least one bulky residue at the interface between the first and second pores is typically at a position corresponding to A98, A99, T104, V105, L113, Q114 or S115 of SEQ ID NO: 2. Where the residue at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is A, the bulky residue is typically I, L, V, M, F, W, Y, N, Q, S or T. Where the residue present at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is T, the bulky residue is typically L, M, F, W, Y, N, Q, R, D or E. Where the residue present at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is V, the bulky residue is typically I, L,

M, F, W, Y, N, Q. Where the residue present at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is L, the bulky residue is typically M, F, W, Y, N, Q, R, D or E. Where the residue present at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is Q, the bulky residue is typically F, W or Y. Where the residue present at the interface in the wild type CsgG monomer, or the wild type CsgG homologue monomer is S, the bulky residue is typically M, F, W, Y, N, Q, E or R.

Particularly where the second pore is located outside the membrane, the second pore, and optionally the first pore, preferably comprises residues in the barrel region of the pore that reduce the negative charge inside the barrel compared to the charge in the barrel of the wild type CsgG pore, or homologue. These mutations make the barrel more hydrophilic. At least one monomer in the first pore and/or at least one monomer in the second pore of the double pore may comprise at least one residue in the barrel region of the pore, which residue has less negative charge than the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer. The charge inside the barrel is sufficiently neutral or positive such that negatively charged analytes, such as polynucleotides, are not repelled from entering the pore by electrostatic charges. At least one residue, such as 2, 3, 4 or 5 residues, in the barrel region of the pore at a position corresponding to D149, E185, D195, E210 and/or E203 of SEQ ID NO: 2 may be a neutral or positively charged amino acid. At least one residue, such as 2, 3, 4 or 5 residues, in the barrel region of the pore at a position corresponding to D149, E185, D195, E210 and/or E203 of SEQ ID NO: 2 is preferably N, Q, R or K.

Particular examples of charge-removing mutations in SEQ ID NO: 2 or SEQ ID NO: 26 include the following: E185N/E203N; D149N/E185R/D195N/E201R/E203N, D149N/E185R/D195N/E201N/E203N, D149R/E185N/D195N/E201N/E203N, D149R/E185N/E201N/E203N, D149N/E185N/D195/E201N/E203N, D149N/E185N/E201N/E203N, D149N/E185N/E203N, D149N/E185N/E201N, D149N/E203N, D149N/E201N/D195N, D149N/E201N, D195N/E201N/E203N, E201N/E203N, D195N/E203, E203R, E203N, E201R, E201N, D195R, D195N, E185R, E185N, D149R and D149N.

At least one monomer in the first CsgG pore, or homologue thereof, may comprise at least one residue in the constriction of the barrel region of the first pore, which residue decreases, maintains or increases the length of the constriction compared to the wild type CsgG pore, or the wild type CsgG homologue pore, and/or at least one monomer in the second CsgG pore, or homologue thereof, may comprise at least one residue in the constriction of the barrel region of the second pore, which residue decreases, maintains or increases the length of the constriction

compared to the wild type CsgG pore, or the wild type CsgG homologue pore. Preferably, the length of the constriction in the first pore and/or the length of the constriction in the second pore is at least as long as in the wild-type pore and more preferably longer.

The length of the pore may be increased by inserting residues into the region corresponding to the region between positions K49 and F56 of SEQ ID NO: 2. From 1 to 5, such as 2, 3, or 4 amino acid residues may be inserted at any one or more of the following positions defined by reference to SEQ ID NO: 2: K49 and P50, P50 and Y51, Y51 and P52, P52 and A53, A53 and S54, S54 and N55 and/or N55 and F56. Preferably from 1 to 10, such as 2 to 8, or 3 to 5 amino acid residues in total are inserted into the sequence of a monomer. Preferably, all of the monomers in the first pore and/or all of the monomers in the second pore have the same number of insertions in this region. The inserted residues may increase the length of the loop between the residues corresponding to Y51 and N55 of SEQ ID NO: 2. The inserted residues may be any combination of A, S, G or T to maintain flexibility; P to add a kink to the loop; and/or S, T, N, Q, M, F, W, Y, V and/or I to contribute to the signal produced when an analyte interacts with the barrel of the pore under an applied potential difference. The inserted amino acids may be any combination of S, G, SG, SGG, SGS, GS, GSS and/or GSG.

In the double pore, the constriction in the barrel of the first pore and/or the second pore may comprise at least one residue, such as 2, 3, 4 or 5 residues, which influences the properties of the pore when used to detect or characterise an analyte compared to when a first pore or a second pore with a wild-type constriction is used, wherein the at least one residue in the constriction of the barrel region of the pore is at a position corresponding to Y51, N55, Y51, P52 and/or A53 of SEQ ID NO: 2. The at least one residue may be Q or V at a position corresponding to F56 of SEQ ID NO: 2; A or Q at a position corresponding to Y51 of SEQ ID NO: 2; and/or V at a position corresponding to N55 of SEQ ID NO: 2.

The double pore may comprise at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the second CsgG pore, or homologue thereof, which monomer comprises two or more of the mutations defined above.

Mutant CsgG Monomers

Provided are novel mutant CsgG monomers, and monomers of CsgG homologues. The monomer may be an isolated monomer. The mutant CsgG monomers may be used to form pores and double pores. A mutant CsgG monomer is a monomer whose sequence varies from that of a wild-type CsgG monomer and which retains the ability to form a pore. A mutant monomer of a

CsgG homologue is a monomer whose sequence varies from that of a wild-type CsgG 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. 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.

Provided is a CsgG monomer, or a monomer of a CsgG homologue, comprising a cysteine residue at a position corresponding to R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2.

Provided is a CsgG monomer, or a monomer of a CsgG homologue, comprising a residue at a position corresponding to any one or more of R97, Q100, I107, R110, E101, N102 and L113 of SEQ ID NO: 2, which residue is more hydrophobic than the residue present at the corresponding position of SEQ ID NO: 2, or in the amino acid sequence of a wild type CsgG homologue, such as the corresponding position of any one of SEQ ID NOs: 3 to 23, wherein the residue at the position corresponding to R97 and/or I107 is M, the residue at the position corresponding to R110 is I, L, V, M, W or Y, and/or the residue at the position corresponding to E101 or N102 is V or M. The residue at a position corresponding to Q100 is typically I, L, V, M, F, W or Y; and or the residue at a position corresponding to L113 is typically I, V, M, F, W or Y.

Particular monomers may have the sequence shown in SEQ ID NO 2 or SEQ ID NO: 26 comprising Y51A, F56Q substitutions and R97I/V/L/M/F/W/Y, I107L/V/M/F/W/Y, R110I/V/L/M/F/W/Y, Q100I/V/L/M/F/W/Y, E101I/V/L/M/F/W/Y, N102I/V/L/M/F/W/Y and L113CI/V/L/M/F/W/Y in combination, R97I/V/L/M/F/W/Y and N102I/V/L/M/F/W/Y in combination and/or R97I/V/L/M/F/W/Y and E101I/V/L/M/F/W/Y in combination. I107 may already form hydrophobic interactions between two pores.

Provided is a CsgG monomer, or a monomer of a CsgG homologue, comprising a residue at a position corresponding to any one or more of A98, A99, T104, V105, L113, Q114 and S115 of SEQ ID NO: 2 which is bulkier than the residue present at the corresponding position of SEQ ID NO: 2, or in the amino acid sequence of a wild type CsgG homologue, such as the corresponding position of any one of SEQ ID NOs: 3 to 23, wherein the residue at the position corresponding to T104 is L, M, F, W, Y, N, Q, D or E, the residue at the position corresponding

to L113 is M, F, W, Y, N, G, D or E and/or the residue at the position corresponding to S115 is M, F, W, Y, N, Q or E. The residue at a position corresponding to A98 or A99, is typically I, L, V, M, F, W, Y, N, Q, S or T. The residue at a position corresponding to V105 is I, L, M, F, W, Y, N or Q. The residue at a position corresponding to Q114 is F, W or Y. The residue at a position corresponding to E210 is N, Q, R or K.

Particular monomers may have the sequence shown in SEQ ID NO 2 or SEQ ID NO: 26 comprising Y51A, F56Q substitutions and 1, 2, 3, 4, 5, 6 or all of the following substitutions: A98I/L/V/M/F/W/Y/N/Q/S/T; A99I/L/V/M/F/W/Y/N/Q/S/T; T104N/Q/L/R/D/E/M/F/W/Y; V105I/L/M/F/W/Y/N/Q; L113M/F/W/Y/N/Q/D/E/L/R; Q114Y/F/W; and S115N/Q/M/F/W/Y/E/R.

Provided is a CsgG monomer, or a monomer of a CsgG homologue, comprising a residue in the barrel region of the pore at a position corresponding to any one or more of D149, E185, D195, E210 and E203 less negative charge than the residue present at the corresponding position of SEQ ID NO: 2, or in the amino acid sequence of a wild type CsgG homologue, such as the corresponding position of any one of SEQ ID NOs: 3 to 23, wherein the residue at the position corresponding to D149, E185, D195 and/or E203 is K.

Provided is a CsgG monomer, or a monomer of a CsgG homologue, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the second CsgG pore, or homologue thereof, comprises at least one residue in the constriction of the barrel region of the pore, which residue increases the length of the constriction compared to the wild type CsgG pore, or the wild type CsgG homologue pore. The at least one residue is additional to the residues present in the constriction of the wild type CsgG pore, or the wild type CsgG homologue pore.

The length of the pore may be increased by inserting residues into the region corresponding to the region between positions K49 and F56 of SEQ ID NO: 2. From 1 to 5, such as 2, 3, or 4 amino acid residues may be inserted at any one or more of the following positions defined by reference to SEQ ID NO: 2: K49 and P50, P50 and Y51, Y51 and P52, P52 and A53, A53 and S54, S54 and N55 and/or N55 and F56. Preferably from 1 to 10, such as 2 to 8, or 3 to 5 amino acid residues in total are inserted into the sequence of the monomer. The inserted residues may increase the length of the loop between the residues corresponding to Y51 and N55 of SEQ ID NO: 2. The inserted residues may be any combination of A, S, G or T to maintain flexibility; P to add a kink to the loop; and/or S, T, N, Q, M, F, W, Y, V and/or I to contribute to the signal produced when an analyte interacts with the barrel of the pore under an applied

potential difference. The inserted amino acids may be any combination of S, G, SG, SGG, SGS, GS, GSS and/or GSG.

5 Provided is a CsgG monomer, or a monomer of a CsgG homologue, comprising at least one residue in the constriction of the barrel region of the pore at a position corresponding to N55, P52 and/or A53 of SEQ ID NO: 2 that is different from the residue present in the corresponding wild type monomer, wherein the residue at a position corresponding to N55 is V.

10 Any two or more of the above described residues may be present in the same monomer. In particular the monomer may comprise at least one said cysteine residue, at least one said hydrophobic residue, at least one said bulky residue, at least one said neutral or positively charged residue and/or at least one said residue that increases the length of the constriction.

15 The monomer may additionally comprise one or more, such as 2, 3, 4 or 5 residues, which influence the properties of the pore when used to detect or characterise an analyte compared to when a first pore or a second pore with a wild-type constriction is used, wherein the at least one residue in the constriction of the barrel region of the pore is at a position corresponding to Y51, N55, Y51, P52 and/or A53 of SEQ ID NO: 2. The at least one residue may be Q or V at a position corresponding to F56 of SEQ ID NO: 2; A or Q at a position corresponding to Y51 of SEQ ID NO: 2; and/or V at a position corresponding to N55 of SEQ ID NO: 2.

20 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.

25 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.

30

Variants

In addition to the specific residues discussed above, the monomer may include other mutations of the wild type sequence.

5 The monomers, and at least one of the monomers in the pores and double pores, may comprise aspartic acid (D), glutamine (Q), phenylalanine (F), serine (S) or threonine (T) at a position corresponding to arginine (R) 192 of SEQ ID NO: 2. Such monomers, and in particular a monomer comprising a D at this position, are much easier to express than monomers comprising a R at a position corresponding to 192 of SEQ ID NO: 2.

10 The monomers, and at least one of the monomers in the pores and double pores, may comprise aspartic acid (N), glutamine (Q), arginine (R), phenylalanine (F), tyrosine (Y), tryptophan (W), leucine (L) or serine (S) in which the lysine (K) at a position corresponding to lysine (K) 192 of SEQ ID NO: 2. Such monomers, and in particular a monomer comprising a Q or N at this position, are less noisy than otherwise identical monomers that comprise a K at a position corresponding to 94 of SEQ ID NO: 2.

15 The monomers, and at least one of the monomers in the pores and double pores, may comprise cis and/or trans loop mutations. Such mutations may play a vital role in double pores, such as in facilitating pore-pore interactions (cis side) and/or pore-enzyme interactions (trans side).

20 The monomers, and at least one of the monomers in the pores and double pores, may comprise one or more of the following amino acid residues, wherein the position of the amino acid residue is defined by reference to SEQ ID NO: 2: D43S (a serine residue at the position corresponding to D43 of SEQ ID NO: 2), E44S (a serine residue at the position corresponding to E44 of SEQ ID NO: 2), F48S/N/Q/Y/W/I/V/H/R/K (a S, N, Q, Y, W, I, V, H, R or K residue at the position corresponding to F48 of SEQ ID NO: 2), Q87N/R/K (a N, R or K residue at the position corresponding to Q87 of SEQ ID NO: 2), N91K/R (a K or R residue at the position corresponding to N91 of SEQ ID NO: 2), K94N/Q/R/F/Y/W/L/S/N (a N, Q, R, F, Y, W, L, S or N residue at the position corresponding to K94 of SEQ ID NO: 2), R97F/Y/W/V/I/K/S/Q/H (a F, Y, W, V, I, K, S, Q or H residue at the position corresponding to R97 of SEQ ID NO: 2), E101I/L/A/H (a L, A or H residue at the position corresponding to E101 of SEQ ID NO: 2), N102K/Q/L/I/V/S/H, R110F/G/N (a K, Q, L, I, V, S or H residue at the position corresponding to N102 of SEQ ID NO: 2), Q114R/K (a R or K residue at the position corresponding to Q114 of SEQ ID NO: 2), R142Q/S (a Q or S residue at the position corresponding to R142 of SEQ ID NO: 2), T150Y/A/V/L/S/Q/N (a Y, A, V, L, S, Q or N residue at the position corresponding to

30

T150 of SEQ ID NO: 2), R192D/Q/F/S/T (a D, Q, F, S or T residue at the position corresponding to R192 of SEQ ID NO: 2) and/or D248S/N/Q/K/R (a S, N, Q, K or R residue at the position corresponding to D248 of SEQ ID NO: 2).

The monomers, and at least one of the monomers in the pores and double pores, may
5 comprise one or more of the following deletions, wherein the position of the deletion is defined by reference to SEQ ID NO: 2: one or more of the residues at positions corresponding to R192, F193, I194, D195, Y196, Q197, R198, L199, L200 and E201 of SEQ ID NO: 2 and/or deletion of one or more of the residues corresponding to V139, G140, D149, T150, V186, Q187, V204 and/or G205 of SEQ ID NO: 2, G137, G138, Q151, Y152, Y184, E185, Y206 and/or T207 of
10 SEQ ID NO: 2 and/or A141, R142, G147, A148, A188, G189, G202 and/or E203 of SEQ ID NO: 2.

The monomers, and at least one of the monomers in the pores and double pores, may comprise:

- a W residue at the position corresponding to R97 of SEQ ID NO: 2;
- 15 - a D, Q, F, S or T residue at the position corresponding to R192 of SEQ ID NO: 2;
- a Y residue at the position corresponding to R97 of SEQ ID NO: 2 and/or a W or Y residue at the position corresponding to R93 of SEQ ID NO: 2;
- a Q or N residue at the position corresponding to K94 of SEQ ID NO: 2;
- a K or R residue at the position corresponding to G103 and/or T104 of SEQ ID NO: 2;
- 20 and/or
- a T residue at the position corresponding to F191 of SEQ ID NO: 2, deletion of the residues corresponding to V105, A106 and I107 of SEQ ID NO: 2 and/or deletion of the residues corresponding to F193, I194, D195, Y196, Q197, R198 and L199 of SEQ ID NO: 2.

The monomers, and at least one of the monomers in the pores and double pores, may
25 comprise an A at the position corresponding to Y51 of SEQ ID NO: 2 and /or a Q at the position corresponding to F56Q of SEQ ID NO: 2.

The monomers, and at least one of the monomers in the pores and double pores, may comprise mutations corresponding to the following mutations in SEQ ID NO: 2:

- (1) Y51A, F56Q and R192D;
- 30 (2) Y51A, F56Q and R97W.
- (3) Y51A, F56Q, R192D and R97W;
- (4) Y51A, F56Q, R192D and R93W;
- (5) Y51A, F56Q, R192D, R93Y and R97Y; or

- (6) Y51A, F56Q, R192D and R93W.
- (7) the mutations of any one of (1)-(6) and:
- (a) deletion of V105, A106 and I107.
 - (b) K94Q or K94N;
 - 5 (c) deletion of D195, Y196, Q197, R198 and L199 or deletion of F193, I194, D195, Y196, Q197, R198 and L199; and/or
 - (d) F191T.
- (8) the mutations of any one of (1)-(6) and:
- (i) K94Q and deletion of V105, A106 and I107;
 - 10 (ii) K94N and deletion of V105, A106 and I107;
 - (iii) F191T and deletion of V105, A106 and I107;
 - (iv) K94Q and F191T;
 - (v) K94N and F191T;
 - (vi) K94Q, F191T and deletion of V105, A106 and I107; or
 - 15 (vii) K94N, F191T and deletion of V105, A106 and I107.
- (9) the mutations of any one of (1)-(8) and:
- T104K or T104R;
 - L90R;
 - N91R;
 - 20 - I95R;
 - A99R;
 - E101K, E101N, E101Q, E101T or E101H;
 - E44N or E44Q; and/or
 - Q42K.
- 25 The monomer may be a variant of any one of SEQ ID NOs: 2 to 23, preferably of SEQ ID NO: 2. Over the entire length of the amino acid sequence of any one of SEQ ID NOs: 2 to 23, 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%,
30 97% or 99% homologous based on amino acid identity to the amino acid sequence of any one of SEQ ID NO: 2 to 23 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/>).

10 SEQ ID NO: 2 is the amino acid sequence of the wild-type CsgG monomer from *Escherichia coli* Str. K-12 substr. MC4100. At least one of the monomers comprised in the pores or double pores may comprise the amino acid sequence shown in SEQ ID NO: 2. At least one of the monomers comprised in the pores or double pores may comprise the amino acid sequence shown in any one of SEQ ID NOs: 3 to 23.

15 The monomers, and at least one of the monomers comprised in the pores or double pores may comprise an amino acid sequence that is a variant of the amino acid sequence shown in SEQ ID NO: 2. The variant of SEQ ID NO: 2 may comprise any of the substitutions present in another CsgG homologue. Preferred CsgG homologues are shown in SEQ ID NOs: 3 to 23. The variant may comprise combinations of one or more of the substitutions present in SEQ ID NOs:
20 3 to 23 compared with SEQ ID NO: 2. For example, mutations may be made at any one or more of the positions in SEQ ID NO: 2 that differ between SEQ ID NO: 2 and any one of SEQ ID NOs: 3 to 23. Such a mutation may be a substitution of an amino acid in SEQ ID NO: 2 with an amino acid from the corresponding position in any one of SEQ ID NOs: 3 to 23. Alternatively, the mutation at any one of these positions may be a substitution with any amino acid, or may be
25 a deletion or insertion mutation, such as deletion or insertion of 1 to 10 amino acids, such as of 2 to 8 or 3 to 6 amino acids. Other than the mutations disclosed herein, the amino acids that are conserved between SEQ ID NO: 2 and all of SEQ ID NOs: 3 to 23 are preferably present in a variant. However, conservative mutations may be made at any one or more of these positions that are conserved between SEQ ID NO: 2 and all of SEQ ID NOs: 3 to 23.

30 Provided is a pore-forming CsgG mutant monomer, or monomer of a CsgG homologue, that comprises any one or more of the amino acids described herein as being substituted into a specific position of SEQ ID NO: 2 at a position in the structure of the monomer that corresponds to the specific position in SEQ ID NO: 2. Corresponding positions may be determined by

standard techniques in the art. For example, the PILEUP and BLAST algorithms mentioned above can be used to align the sequence of a CsgG monomer, or monomer of a CsgG homologue, with SEQ ID NO: 2 and hence to identify corresponding residues.

5 In particular, a pore-forming CsgG mutant monomer that comprises any one or more of the following is provided:

- a W at a position corresponding to R97 in SEQ ID NO:2;
- a W at a position corresponding to R93 in SEQ ID NO:2;
- a Y at a position corresponding to R97 in SEQ ID NO: 2;
- a Y at a position corresponding to R93 in SEQ ID NO: 2;
- 10 - a Y at each of the positions corresponding to R93 and R97 in SEQ ID NO: 2;
- a D at the position corresponding to R192 in SEQ ID NO:2;
- deletion of the residues at the positions corresponding to V105-I107 in SEQ ID NO:2;
- deletion of the residues at one or more of the positions corresponding to F193 to L199 in SEQ ID NO: 2;
- 15 - deletion of the residues the positions corresponding to F195 to L199 in SEQ ID NO: 2;
- deletion of the residues the positions corresponding to F193 to L199 in SEQ ID NO: 2;
- a T at the position corresponding to F191 in SEQ ID NO: 2;
- 20 - a Q at the position corresponding to K49 in SEQ ID NO: 2;
- a N at the position corresponding to K49 in SEQ ID NO: 2;
- a Q at the position corresponding to K42 in SEQ ID NO: 2;
- a Q at the position corresponding to E44 in SEQ ID NO: 2;
- a N at the position corresponding to E44 in SEQ ID NO: 2;
- 25 - a R at the position corresponding to L90 in SEQ ID NO: 2;
- a R at the position corresponding to L91 in SEQ ID NO: 2;
- a R at the position corresponding to I95 in SEQ ID NO: 2;
- a R at the position corresponding to A99 in SEQ ID NO: 2;
- a H at the position corresponding to E101 in SEQ ID NO: 2;
- 30 - a K at the position corresponding to E101 in SEQ ID NO: 2;
- a N at the position corresponding to E101 in SEQ ID NO: 2;
- a Q at the position corresponding to E101 in SEQ ID NO: 2;
- a T at the position corresponding to E101 in SEQ ID NO: 2;

- a K at the position corresponding to Q114 in SEQ ID NO: 2.

The CsgG pore-forming monomer may further comprise an A at the position corresponding to Y51 in SEQ ID NO: 2 and/or a Q at the position corresponding to F56 in SEQ ID NO: 2.

5 The pore-forming mutant monomer typically retains the ability to form the same 3D structure as the wild-type CsgG monomer, such as the same 3D structure as a CsgG monomer having the sequence of SEQ ID NO: 2. The 3D structure of CsgG is known in the art and is disclosed, for example, in Cao et al (2014) PNAS E5439-E5444. Any number of mutations may be made in the wild-type CsgG sequence in addition to the mutations described herein provided
10 that the CsgG mutant monomer retains the improved properties imparted on it by the mutations.

Typically the CsgG monomer will retain the ability to form a structure comprising three alpha-helices and five beta-sheets. The present inventors have shown in particular that mutations may be made at least in the region of CsgG which is N-terminal to the first alpha helix (which starts at S63 in SEQ ID NO:2), in the second alpha helix (from G85 to A99 of SEQ ID
15 NO: 2), in the loop between the second alpha helix and the first beta sheet (from Q100 to N120 of SEQ ID NO: 2), in the fourth and fifth beta sheets (S173 to R192 and R198 to T107 of SEQ ID NO: 2, respectively) and in the loop between the fourth and fifth beta sheets (F193 to Q197 of SEQ ID NO: 2) without affecting the ability of the CsgG monomer to form a transmembrane pore, which transmembrane pore is capable of translocating polypeptides. Therefore, it is
20 envisaged that further mutations may be made in any of these regions in any CsgG monomer without affecting the ability of the monomer to form a pore that can translocate polynucleotides. It is also expected that mutations may be made in other regions, such as in any of the alpha helices (S63 to R76, G85 to A99 or V211 to L236 of SEQ ID NO: 2) or in any of the beta sheets (I121 to N133, K135 to R142, I146 to R162, S173 to R192 or R198 to T107 of SEQ ID NO: 2)
25 without affecting the ability of the monomer to form a pore that can translocate polynucleotides. It is also expected that deletions of one or more amino acids can be made in any of the loop regions linking the alpha helices and beta sheets and/or in the N-terminal and/or C-terminal regions of the CsgG monomer without affecting the ability of the monomer to form a pore that can translocate polynucleotides.

30 Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2 in addition to those discussed herein, 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 |
| 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 or more residues may be deleted.

The monomers may be fragments of SEQ ID NO: 2, and pores and the double pores may comprise fragments of SEQ ID NO:2 or of a homologue of SEQ ID NO: 2. Such fragments retain pore forming activity. Fragments may be at least 50, at least 100, at least 150, at least 200 or at least 250 amino acids in length. Such fragments may be used to produce the pores. A fragment preferably comprises the membrane spanning domain corresponding to K135-Q153 and S183-S208 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 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. Other fusion proteins are discussed in more detail below.

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. A variant typically contains the regions of SEQ ID NO: 2 that are responsible for pore formation. The pore forming ability of CsgG, 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, namely K135-Q153 and S183-S208. 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 CsgG 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 CsgG may also be produced using D-amino acids. For instance, the monomer derived from CsgG 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 CsgG contains one or more specific modifications to facilitate nucleotide discrimination. The monomer derived from CsgG 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 CsgG. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with NaBH₄, amidation with methylacetimidate or acylation with acetic anhydride.

The monomer derived from CsgG can be produced using standard methods known in the art. The monomer derived from CsgG may be made synthetically or by recombinant means. For example, the monomer may be synthesised by *in vitro* translation and transcription (IVTT). Suitable methods for producing pores and monomers are discussed in WO 2010/004273, WO 2010/004265 and WO 2010/086603. Methods for inserting pores into membranes are also discussed.

In some embodiments, the monomer, which may optionally be comprised in a pore or double pore is chemically modified. The mutant monomer can be chemically modified in any way and at any site. The 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 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 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.

5 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 or nine-fold symmetry since CsgG typically has
10 eight or nine 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
15 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, π -cation interactions and/or electrostatic forces. The one or more chemical groups that are capable of
20 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.

25 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
30 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 (am₇- β CD), 6-monodeoxy-6-monoamino- β -cyclodextrin (am₁- β CD) or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu₇- β CD). The guanidino group in gu₇- β CD has a much higher pK_a than the primary amines in am₇- β CD and so it is more positively charged. This gu₇- β CD 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 (am₆amPDP₁- β CD).

More suitable adaptors include γ -cyclodextrins, which comprise 9 sugar units (and therefore have nine-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 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, for instance in the barrel, by substitution. The monomer may be chemically modified by attachment of a molecular adaptor to one or more cysteines in the monomer. The one or more cysteines may be naturally-occurring, i.e. at positions 1 and/or 215 in SEQ ID NO: 2. Alternatively, the mutant monomer may be chemically modified by attachment of a molecule to one or more cysteines introduced at other positions. The cysteine at position 215 may be removed, for instance by substitution, to ensure that the molecular adaptor does not attach to that position rather than the cysteine at position 1 or a cysteine introduced at another position.

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 may be attached directly to the monomer. The molecule is preferably attached to the 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-yl)disulfanyl)propanoate, 2,5-dioxopyrrolidin-1-yl 4-(pyridin-2-yl)disulfanyl)butanoate and 2,5-dioxopyrrolidin-1-yl 8-(pyridin-2-yl)disulfanyl)octanoate. The most preferred crosslinker is succinimidyl 3-(2-pyridyl)disulfanyl)propionate (SPDP). Typically, the molecule is covalently attached to the bifunctional crosslinker before the molecule/crosslinker complex is covalently attached to the 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 methods of sequencing. Polynucleotide binding proteins are discussed below.

The polynucleotide binding protein is preferably covalently attached to the 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 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 one or more cysteines are preferably introduced into 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. In such embodiments, the naturally-occurring cysteine at position 251 may be removed. The reactivity of cysteine residues may be enhanced by modification as described above.

The polynucleotide binding protein may be attached directly to the monomer or via one or more linkers. The molecule may be attached to the monomer using the hybridization linkers

described in 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 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 monomer may be chemically modified with a molecular adaptor and a polynucleotide binding protein.

The molecule (with which the monomer is chemically modified) may be attached directly to the monomer or attached via a linker as disclosed in WO 2010/004273, WO 2010/004265 or WO 2010/086603.

Any of the proteins described herein, such as the monomers and pores, 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 monomers and pores, 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. ¹²⁵I, ³⁵S, enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

Any of the proteins described herein, such as the monomers or pores, may be made synthetically or by recombinant means. For example, the protein may be synthesised 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 NaBH₄, amidation with methylacetimidate or acylation with acetic anhydride.

Any of the proteins described herein, including the monomers and pores, 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

Also provided is a construct comprising two or more covalently attached CsgG monomers, or monomer of CsgG homologues, wherein at least one of the monomers is a modified monomer, such as a monomer of the invention. The construct retains its ability to form a pore. This may be determined as discussed above. One or more constructs 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 modified monomer, such as a monomer of the invention. Two 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 monomers of the invention. All of the monomers in the construct are preferably monomers of the invention. The monomers may be the same or different. In a preferred embodiment, the construct comprises two monomers of the invention.

The monomers in the construct are preferably approximately the same length or are the same length. The barrels of the monomers 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 monomers of the invention. CsgG monomers which are non mutant monomers of the invention include monomers comprising any one of SEQ ID NOs: 2 to 23 or a variant of any one of SEQ ID NOs: 2 to 23 in which none of the amino acids/positions discussed above have been mutated. At least one monomer in the construct may comprise SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 or a variant of the sequence shown in SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23. A variant of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 is at least 50% homologous to SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 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, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 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
5 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
10 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
15 attached to one or more cysteine residues introduced into a mutant monomer. 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
20 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 WO 2010/086602.

Polynucleotides

Also provided are polynucleotide sequences which encode a modified monomer, such as
25 a monomer 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: 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
30 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.

Also provided are polynucleotide sequences which encode any of the genetically fused constructs. 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.

Polynucleotide sequences may be derived and replicated using standard methods in the art. Chromosomal DNA encoding wild-type CsgG may be extracted from a pore producing organism, such as *Escherichia coli*. 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 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 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 or a synthetic membrane.

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

The host cell typically expresses the 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 Cao et al, 2014, PNAS, Structure of the nonameric bacterial amyloid secretion channel, doi – 1411942111 and Goyal et al, 2014, Nature, 516, 250-253 structural and mechanistic insights into the bacterial amyloid secretion channel CsgG may be used to express the CsgG 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 in addition to the double pores described above. The pores 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 can even distinguish between methylated and unmethylated nucleotides. The base resolution of the pores 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.

Provided is a pore comprising at least one monomer according to the invention or a construct according to the invention. The pore may be homooligomeric or heterooligomeric.

The pores 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 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 may also be used to identify polynucleotide polymers from the interaction with one or more monomers rather than on a nucleotide by nucleotide basis.

The pore may be isolated, substantially isolated, purified or substantially purified. A pore 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, the pore may be present in a membrane.

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

The use of the term pore in the present disclosure is intended to encompass both pores and double pores.

Homo-oligomeric Pores

Also provided is a homo-oligomeric pore comprising identical monomers as disclosed herein. The homo-oligomeric pore may comprise any of the monomers of the invention. The homo-oligomeric pore is ideal for characterising, such as sequencing, polynucleotides.

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

Provided are homo-oligomeric double pores.

Hetero-oligomeric Pores

Also provided is a hetero-oligomeric pore comprising at least one monomer as disclosed herein. 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 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 monomers of the invention and at least one of them differs from the others. They may all differ from one another.

The monomers of the invention in the pore are preferably approximately the same length or are the same length. The barrels of the 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.

In another preferred embodiment, at least one of the monomers is not a monomer of the invention. In this embodiment, the remaining monomers are preferably monomers of the invention. Hence, the pore may comprise 9, 8, 7, 6, 5, 4, 3, 2 or 1 monomers of the invention. Any number of the monomers in the pore may not be a monomer of the invention. The pore preferably comprises seven or eight monomers of the invention and a monomer which is not a monomer of the invention. The monomers of the invention may be the same or different.

The monomers, such as the monomers of the invention, in the pore are preferably approximately the same length or are the same length. The barrels of the monomers 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 pore may comprise one or more monomers which are not monomers of the invention. CsgG monomers which are not monomers of the invention include monomers comprising SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 or a variant of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 in which none of the amino acids/positions discussed above in relation to the invention have been mutated/substituted. A comparative variant of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 is typically at least 50% homologous to SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 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, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 or 23 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.

Provided are hetero-oligomeric double pores.

20 *Construct-containing Pores*

Also provided is a pore comprising at least one construct of the invention. The pore may be a double pore. The construct comprises two or more covalently attached monomers derived from CsgG wherein at least one of the monomers is a monomer of the invention. In other words, the 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 may be in the form of a construct of the invention. The construct, and hence the pore, comprises at least one 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
5 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
10 those discussed above, including monomers of the invention, monomers comprising SEQ ID NO: 2 to 23 and monomers comprising a variant of SEQ ID NO: 2 to 23.

Another typical pore comprises more than one construct, such as more than one construct of the invention, for example two, three or four constructs of the invention. If necessary, such pores further comprise sufficient additional monomers or constructs to form the pore. The
15 additional monomer(s) may be any of those discussed above, including monomers of the invention, monomers comprising SEQ ID NO: 2 to 23 and monomers comprising a variant of SEQ ID NO: 2 to 23 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 CsgG monomers each comprising a monomer comprising SEQ ID NO: 2 to 23 or a variant of SEQ ID
20 NO: 2 to 23.

A further pore 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 in such a pore 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 monomer of the invention. All of the
25 constructs comprising 2 monomers may be constructs of the invention.

A specific pore 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 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.
30 Typically the other monomers of the construct will be on the outside of the channel of the pore. For example, the pore may comprise 7, 8, 9 or 10 constructs comprising 2 monomers where the channel comprises 7, 8, 9 or 10 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 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 follow by oligomerisation with MutB-containing monomers to form A-A:B:B:B:B:B:B.

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

Also provided is a method of producing a monomer according to the invention or a construct according to the invention, comprising expressing a polynucleotide according to the invention in a suitable host cell and thereby producing the monomer or the construct.

15 *Analyte Characterization*

The double pore is particularly suited for characterising, such as sequencing, polynucleotides. Double pores comprising two wild type CsgG, or two CsgG homologues, may also be used in a method of characterizing an analyte, such as a polynucleotide. Double pores have two constrictions that can function as read heads in a method of characterizing a
20 polynucleotide, particularly when sequencing a polynucleotide. Having two read heads is particularly beneficial when sequencing homopolymeric regions of a polynucleotide.

Accordingly, provided is a method of characterising a polynucleotide using a transmembrane pore, wherein the pore is a double pore comprising a first CsgG pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof. In a preferred
25 embodiment, the polynucleotide comprises a homopolymeric region. Also provided is the use of a double pore or a pore as disclosed herein to determine the presence, absence or one or more characteristics of a target analyte.

Provided is a method of sequencing a series of identical nucleotides in a polynucleotide, which method comprises contacting the a polynucleotide with a double pore or a pore as
30 disclosed herein, such that the target analyte moves with respect to the pore; and taking one or more measurements as the analyte moves with respect to the pore and thereby determining the identity and number of identical nucleotides in the polynucleotide.

The pore may be one that has an elongated read head (barrel constriction) compared to the CsgG pores used for sequencing in the prior art, particularly compared to a CsgG pore comprising SEQ ID NO: 2 with Y51A and F56Q substitutions. The read head may be comparable to (the same, or approximately the same length), or longer than, the constriction in a wild-type CsgG pore comprising SEQ ID NO: 2. An elongated constriction that can function as a read head is advantageous in a method of characterizing a polynucleotide, particularly when sequencing a polynucleotide. Having a longer read head is particularly beneficial when sequencing homopolymeric regions of a polynucleotide.

In the method, the double pore may be any of the double pores, or the first CsgG pore, or homologue thereof may be any homooligomer and the second CsgG pore, or homologue thereof, may be any homooligomer, including wild-type homooligomers, and the first CsgG pore, or homologue thereof, and the second CsgG pore, or homologue thereof, may be identical to one another.

Provided is a method for determining the presence, absence or one or more characteristics of a target analyte, comprising:

- (a) contacting the target analyte with a double pore or a pore as disclosed herein such that the target analyte moves with respect to the pore; and
- (b) taking one or more measurements as the analyte moves with respect to the pore and thereby determining the presence, absence or one or more characteristics of the analyte.

In a preferred embodiment, the analyte is a polynucleotide. In a more preferred embodiment the polynucleotide is a polynucleotide comprising a homopolymeric region.

The method may comprise determining one or more characteristics 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. The one or more characteristics of the analyte are typically measured by electrical measurement and/or optical measurement.

Provided is a method of determining the presence, absence or one or more characteristics of a target analyte. The method involves contacting the target analyte with a pore as disclosed herein such that the target analyte moves with respect to, such as through, the pore and taking one or more measurements as the analyte moves with respect to the pore and thereby determining the presence, absence or one or more characteristics of the analyte. The target analyte may also be called the template analyte or the analyte of interest.

Steps (a) and (b) are preferably carried out with a potential applied across the pore. As discussed in more detail below, the applied potential typically results in the formation of a complex between the pore and a polynucleotide binding protein. 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 an amphiphilic layer. A salt gradient is disclosed in Holden *et al.*, J Am Chem Soc. 2007 Jul 11;129(27):8650-5.

The method is for determining the presence, absence or one or more characteristics of a target analyte. The method may be for determining the presence, absence or one or more characteristics of at least one analyte. The method may concern determining the presence, absence or one or more characteristics of two or more analytes. The method may comprise determining the presence, absence or one or more characteristics of any number of analytes, such as 2, 5, 10, 15, 20, 30, 40, 50, 100 or more analytes. Any number of characteristics of the one or more analytes may be determined, such as 1, 2, 3, 4, 5, 10 or more characteristics.

The target analyte is preferably a metal ion, an inorganic salt, a polymer, an amino acid, a peptide, a polypeptide, a protein, a nucleotide, an oligonucleotide, a polynucleotide, a dye, a bleach, a pharmaceutical, a diagnostic agent, a recreational drug, an explosive or an environmental pollutant. The method may concern determining the presence, absence or one or more characteristics of two or more analytes of the same type, such as two or more proteins, two or more nucleotides or two or more pharmaceuticals. Alternatively, the method may concern determining the presence, absence or one or more characteristics of two or more analytes of different types, such as one or more proteins, one or more nucleotides and one or more pharmaceuticals.

The target analyte can be secreted from cells. Alternatively, the target analyte can be an analyte that is present inside cells such that the analyte must be extracted from the cells.

The analyte is preferably an amino acid, a peptide, a polypeptides and/or a protein. The amino acid, peptide, polypeptide or protein can be naturally-occurring or non-naturally-occurring. The polypeptide or protein can include within them synthetic or modified amino acids. A number of different types of modification to amino acids are known in the art. Suitable amino acids and modifications thereof are above. It is to be understood that the target analyte can be modified by any method available in the art.

The protein can be an enzyme, an antibody, a hormone, a growth factor or a growth regulatory protein, such as a cytokine. The cytokine may be selected from interleukins, preferably IFN-1, IL-1, IL-2, IL-4, IL-5, IL-6, IL-10, IL-12 and IL-13, interferons, preferably IL-

γ , and other cytokines such as TNF- α . The protein may be a bacterial protein, a fungal protein, a virus protein or a parasite-derived protein.

The target analyte is preferably a nucleotide, an oligonucleotide or a polynucleotide. Nucleotides and polynucleotides are discussed below. Oligonucleotides are short nucleotide
5 polymers which typically have 50 or fewer nucleotides, such as 40 or fewer, 30 or fewer, 20 or fewer, 10 or fewer or 5 or fewer nucleotides. The oligonucleotides may comprise any of the nucleotides discussed below, including the abasic and modified nucleotides.

The polynucleotide is preferably a polynucleotide that comprises a homopolymeric region, i.e, the polynucleotide comprises a series of repeated nucleotides such as two or more
10 adjacent A, G, C, T or U bases, such as 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 100 or more consecutive As, Gs, Cs, Ts or Us. The double pores and pores, particularly the pores in which the length of the narrowest part of the pore is the same as or longer than in the wild type CsgG pore, are particularly suited for determining the sequence of such homopolymeric polynucleotide sequences.

15 The target analyte, such as a target polynucleotide, may be present in any suitable sample. Examples of suitable samples are discussed below.

The pore is typically present in a membrane as discussed below. The target analyte may be coupled or delivered to the membrane using of the methods discussed below.

20 Any of the measurements discussed below can be used to determine the presence, absence or one or more characteristics of the target analyte. The method preferably comprises contacting the target analyte with the pore such that the analyte moves with respect to, such as moves through, the pore and measuring the current passing through the pore as the analyte moves with respect to the pore and thereby determining the presence, absence or one or more characteristics of the analyte.

25 The target analyte is present if the current flows through the pore in a manner specific for the analyte (i.e. if a distinctive current associated with the analyte is detected flowing through the pore). The analyte is absent if the current does not flow through the pore in a manner specific for the nucleotide. Control experiments can be carried out in the presence of the analyte to determine the way in which it affects the current flowing through the pore.

30 The method can be used to differentiate analytes of similar structure on the basis of the different effects they have on the current passing through a pore. Individual analytes can be identified at the single molecule level from their current amplitude when they interact with the pore. The method can also be used to determine whether or not a particular analyte is present in

a sample. The method can also be used to measure the concentration of a particular analyte in a sample. Analyte characterisation using pores other than CsgG is known in the art.

Characterisation, such as sequencing, of a polynucleotide using a transmembrane pore may be carried out such as disclosed in WO2013/041878. As the target polynucleotide moves with respect to, or through the pore, the analyte may be characterised from the distinctive ion current signature produced, typically by measuring the ion current flow through the pore. The level of current measured at any particular time is typically dependent on multiple polymer (for example nucleotide) units. The number of polymer units contributing to the current at any time will depend on the structure of the polymer, particularly on the structure of the barrel constriction(s). For example, from about 3 to about 20, such a 5, 6, 7, 8, 9, 10, 12, 13, 14 or 15 polymer units may influence the current level at any one time.

Analytical techniques to characterise the polynucleotide may for example involve the use of an HMM, a neural network and for example a Forwards Backwards algorithm or Viterbi algorithm to determine the likelihood of the series of measurements corresponding to a particular sequence. Alternatively the polynucleotide may be characterised by determining a feature vector and comparing the feature vector to another feature vector, which may be known, such as disclosed in WO 2013/121224. However, the analytical techniques used to characterise the polynucleotide are not necessarily restricted to the above examples.

20 *Polynucleotide Characterisation*

Provided is a method of characterising a target polynucleotide, such as sequencing a polynucleotide. There are two main strategies for characterising or sequencing polynucleotides using nanopores, namely strand characterisation/sequencing and exonuclease characterisation/sequencing. The method may concern either method.

25 In strand sequencing, the DNA is translocated through the nanopore either with or against an applied potential. Exonucleases that act progressively or processively on double stranded DNA can be used on the *cis* side of the pore to feed the remaining single strand through under an applied potential or the *trans* side under a reverse potential. Likewise, a helicase that unwinds the double stranded DNA can also be used in a similar manner. A polymerase may also be used.
30 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.

In one embodiment, the method of characterising a target polynucleotide involves contacting the target sequence with a pore as disclosed herein and a helicase enzyme. Any helicase may be used in the method. Suitable helicases are discussed below. Helicases may work in two modes with respect to the pore. First, the method is preferably carried out using a helicase such that it controls movement of the target sequence through the pore with the field resulting from the applied voltage. In this mode the 5' end of the DNA is first captured in the pore, and the enzyme controls movement of the DNA into the pore such that the target sequence is passed through the pore with the field until it finally translocates through to the trans side of the bilayer. Alternatively, the method is preferably carried out such that a helicase enzyme controls movement of the target sequence through the pore against the field resulting from the applied voltage. In this mode the 3' end of the DNA is first captured in the pore, and the enzyme controls movement of the DNA through the pore such that the target sequence is pulled out of the pore against the applied field until finally ejected back to the cis side of the bilayer.

In exonuclease sequencing, an exonuclease releases individual nucleotides from one end of the target polynucleotide and these individual nucleotides are identified as discussed below. In another embodiment, the method of characterising a target polynucleotide involves contacting the target sequence with a pore and an exonuclease enzyme. Any of the exonuclease enzymes discussed below may be used in the method. The enzyme may be covalently attached to the pore as discussed below.

Exonucleases are enzymes that typically latch onto one end of a polynucleotide and digest the sequence one nucleotide at a time from that end. The exonuclease can digest the polynucleotide in the 5' to 3' direction or 3' to 5' direction. The end of the polynucleotide to which the exonuclease binds is typically determined through the choice of enzyme used and/or using methods known in the art. Hydroxyl groups or cap structures at either end of the polynucleotide may typically be used to prevent or facilitate the binding of the exonuclease to a particular end of the polynucleotide.

The method involves contacting the polynucleotide with the exonuclease so that the nucleotides are digested from the end of the polynucleotide at a rate that allows characterisation or identification of a proportion of nucleotides as discussed above. Methods for doing this are well known in the art. For example, Edman degradation is used to successively digest single

amino acids from the end of polypeptide such that they may be identified using High Performance Liquid Chromatography (HPLC). A homologous method may be used.

The rate at which the exonuclease functions is typically slower than the optimal rate of a wild-type exonuclease. A suitable rate of activity of the exonuclease in the method involves digestion of from 0.5 to 1000 nucleotides per second, from 0.6 to 500 nucleotides per second, 0.7 to 200 nucleotides per second, from 0.8 to 100 nucleotides per second, from 0.9 to 50 nucleotides per second or 1 to 20 or 10 nucleotides per second. The rate is preferably 1, 10, 100, 500 or 1000 nucleotides per second. A suitable rate of exonuclease activity can be achieved in various ways. For example, variant exonucleases with a reduced optimal rate of activity may be used.

In the strand characterisation embodiment, the method comprises contacting the polynucleotide with a pore as disclosed herein such that the polynucleotide moves with respect to, such as through, the pore and 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.

In the exonucleotide characterisation embodiment, the method comprises contacting the polynucleotide with a pore as disclosed herein and an exonuclease such that the exonuclease digests individual nucleotides from one end of the target polynucleotide and the individual nucleotides move with respect to, such as through, the pore and taking one or more measurements as the individual nucleotides move with respect to the pore, wherein the measurements are indicative of one or more characteristics of the individual nucleotides, and thereby characterising the target polynucleotide.

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

The individual nucleotides may interact with the pore in any manner and at any site. The nucleotides preferably reversibly bind to the pore via or in conjunction with an adaptor as discussed above. The nucleotides most preferably reversibly bind to the pore via or in

conjunction with the adaptor as they pass through the pore across the membrane. The nucleotides can also reversibly bind to the barrel or channel of the pore via or in conjunction with the adaptor as they pass through the pore across the membrane.

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

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.

This embodiment also uses a pore as disclosed herein. Any of the pores and embodiments discussed above with reference to the target analyte may be used.

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. One or more nucleotides in the polynucleotide may be damaged. For instance, the polynucleotide may comprise a pyrimidine dimer. Such dimers are typically associated with damage by ultraviolet light and are the primary cause of skin melanomas. One or more nucleotides in the polynucleotide may be modified, for instance with a label or a tag. Suitable labels are described below. The polynucleotide may comprise one or more spacers.

A nucleotide typically contains a nucleobase, a sugar and at least one phosphate group. The nucleobase and sugar form a nucleoside.

The nucleobase is typically heterocyclic. Nucleobases include, but are not limited to, purines and pyrimidines and more specifically adenine (A), guanine (G), thymine (T), uracil (U) and cytosine (C).

The sugar is typically a pentose sugar. Nucleotide sugars include, but are not limited to, ribose and deoxyribose. The sugar is preferably a deoxyribose.

The polynucleotide preferably comprises the following nucleosides: deoxyadenosine (dA), deoxyuridine (dU) and/or thymidine (dT), deoxyguanosine (dG) and deoxycytidine (dC).

The nucleotide is typically a ribonucleotide or deoxyribonucleotide. The nucleotide typically contains a monophosphate, diphosphate or triphosphate. The nucleotide may comprise more than three phosphates, such as 4 or 5 phosphates. Phosphates may be attached on the 5' or 3' side of a nucleotide. Nucleotides include, but are not limited to, adenosine monophosphate (AMP), guanosine monophosphate (GMP), thymidine monophosphate (TMP), uridine monophosphate (UMP), 5-methylcytidine monophosphate, 5-hydroxymethylcytidine monophosphate, cytidine monophosphate (CMP), cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), deoxyadenosine monophosphate (dAMP), deoxyguanosine monophosphate (dGMP), deoxythymidine monophosphate (dTMP), deoxyuridine monophosphate (dUMP), deoxycytidine monophosphate (dCMP) and deoxymethylcytidine monophosphate. The nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP, dCMP and dUMP.

A nucleotide may be abasic (i.e. lack a nucleobase). A nucleotide may also lack a nucleobase and a sugar (i.e. is a C3 spacer).

The nucleotides in the polynucleotide may be attached to each other in any manner. The nucleotides are typically attached by their sugar and phosphate groups as in nucleic acids. The nucleotides may be connected via their nucleobases as in pyrimidine dimers.

The polynucleotide may be single stranded or double stranded. The polynucleotide is preferably single stranded. Single stranded polynucleotide characterization is referred to as 1D in the Examples. At least a portion of the polynucleotide may be 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.

5 LNA is formed from ribonucleotides as discussed above having an extra bridge connecting the 2' oxygen and 4' carbon in the ribose moiety. Bridged nucleic acids (BNAs) are modified RNA nucleotides. They may also be called constrained or inaccessible RNA. BNA monomers can contain a five-membered, six-membered or even a seven-membered bridged structure with a “fixed” C3'-endo sugar puckering. The bridge is synthetically incorporated at the 2', 4'-position
10 of the ribose to produce a 2', 4'-BNA monomer.

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
15 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 may concern characterising 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100 or more polynucleotides. If two
20 or more polynucleotides are characterised, 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*.

25

Sample

The polynucleotide is typically present in any suitable sample. The method is typically carried out on a sample that is known to contain or suspected to contain the polynucleotide. Alternatively, the method may be carried out on a sample to confirm the identity of a
30 polynucleotide whose presence in the sample is known or expected.

The sample may be a biological sample. The method may be carried out *in vitro* using a sample obtained from or extracted from any organism or microorganism. The organism or microorganism is typically archaeal, prokaryotic or eukaryotic and typically belongs to one of

the five kingdoms: plantae, animalia, fungi, monera and protista. The method may be carried out *in vitro* on a sample obtained from or extracted from any virus. The sample is preferably a fluid sample. The sample typically comprises a body fluid of the patient. The sample may be urine, lymph, saliva, mucus or amniotic fluid but is preferably blood, plasma or serum.

5 Typically, the sample is human in origin, but alternatively it may be from another mammal animal such as from commercially farmed animals such as horses, cattle, sheep, fish, chickens or pigs or may alternatively be pets such as cats or dogs. Alternatively, the sample may be of plant origin, such as a sample obtained from a commercial crop, such as a cereal, legume, fruit or vegetable, for example wheat, barley, oats, canola, maize, soya, rice, rhubarb, bananas,
10 apples, tomatoes, potatoes, grapes, tobacco, beans, lentils, sugar cane, cocoa, cotton.

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

Characterisation

20 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 method, such
25 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.

30 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, methylcytosine 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 as disclosed herein. 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 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 (WO 2005/124888).

5 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
10 KR et al, J Am Chem Soc. 2010;132(50):17961-72, and WO 2000/28312. Alternatively, electrical measurements may be made using a multi-channel system, for example as described in WO 2009/077734 and 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
15 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.

20 The method may involve measuring the current passing through the pore as the polynucleotide moves with respect to 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.

25 The method 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 -
30 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. 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

The strand characterisation method 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.

5 More preferably, the method comprises (a) contacting the polynucleotide with a pore as disclosed herein and a polynucleotide binding protein such that the protein controls the movement of the polynucleotide with respect to, such as 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
10 characterising the polynucleotide.

More preferably, the method comprises (a) contacting the polynucleotide with a pore as disclosed herein and a polynucleotide binding protein such that the protein controls the movement of the polynucleotide with respect to, such as through, the pore and (b) measuring the current through the pore as the polynucleotide moves with respect to the pore, wherein the
15 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 through the pore. It is straightforward in the art to determine whether or not a protein binds to a polynucleotide. The protein typically interacts
20 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 derived from a polynucleotide handling
25 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
30 enzymatic activity as long as it is capable of binding the polynucleotide and controlling its movement through the pore. For instance, the enzyme may be modified to remove its enzymatic activity or may be used under conditions which prevent it from acting as an enzyme. Such conditions are discussed in more detail below.

The 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
5 of those disclosed in 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*, exonuclease III enzyme from *E. coli*, RecJ from *T. thermophilus* and bacteriophage lambda exonuclease, TatD exonuclease and variants thereof. The polymerase may be PyroPhage® 3173
10 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 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, Hel308
15 Csy, Hel308 Tga, Hel308 Mhu, TraI Eco, XPD Mbu or a variant thereof. Any helicase may be used. 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: WO 2013/057495; WO 2013/098562; WO2013098561; WO 2014/013260; WO 2014/013259; WO 2014/013262 and
20 WO2015/055981.

Any number of helicases may be used. 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 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
25 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
30 helicases are more preferably covalently attached to one another. The helicases may be attached in any order and using any method. Preferred helicase constructs are described in WO 2014/013260; WO 2014/013259; WO 2014/013262 and WO2015/055981.

A variant of a naturally occurring helicase which retains polynucleotide binding ability may be used. Polynucleotide binding ability 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). A preferred molecular brake is TrwC Cba-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+}).

The enzyme may be covalently attached to the pore. Any method may be used to covalently attach the enzyme to the pore.

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 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 providing the polynucleotide with one or more helicases and one or more molecular brakes attached to the polynucleotide; contacting the polynucleotide with a double pore or a pore as disclosed herein 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 with respect to, such as through, the pore; and 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 disclosed in WO2015/110777.

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 (am₇- β CD), 6-monodeoxy-6-monoamino- β -cyclodextrin or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu₇- β CD).

The one or more molecular brakes are preferably one or more single stranded binding proteins (SSB). The one or more molecular brakes are more preferably 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 one of the SSBs disclosed in 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 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 are described in WO 2014/013260, WO 2014/013259, WO 2014/013262 and WO2015/055981.

If the one or more helicases are used in the active mode (i.e. when the one or more helicases are provided with all the necessary components to facilitate movement, e.g. ATP and Mg^{2+}), the one or more molecular brakes are preferably (a) used in an inactive mode (i.e. are used in the absence of the necessary components to facilitate movement or are incapable of active movement), (b) used in an active mode where the one or more molecular brakes move in the opposite direction to the one or more helicases or (c) used in an active mode where the one or more molecular brakes move in the same direction as the one or more helicases and more slowly than the one or more helicases.

If the one or more helicases are used in the inactive mode (i.e. when the one or more helicases are not provided with all the necessary components to facilitate movement, e.g. ATP and Mg^{2+} or are incapable of active movement), the one or more molecular brakes are preferably (a) used in an inactive mode (i.e. are used in the absence of the necessary components to facilitate movement or are incapable of active movement) or (b) used in an active mode where the one or more molecular brakes move along the polynucleotide in the same direction as the polynucleotide through the pore.

The one or more helicases and one or more molecular brakes may be attached to the polynucleotide at any positions so that they are brought together and both control the movement of the polynucleotide through the pore. The one or more helicases and one or more molecular brakes are at least one nucleotide apart, such as at least 5, at least 10, at least 50, at least 100, at least 500, at least 1000, at least 5000, at least 10,000, at least 50,000 nucleotides or more apart. If the method concerns characterising a double stranded polynucleotide provided with a Y adaptor at one end and a hairpin loop adaptor at the other end, the one or more helicases are preferably attached to the Y adaptor and the one or more molecular brakes are preferably attached to the hairpin loop adaptor. In this embodiment, the one or more molecular brakes are preferably one or more helicases that are modified such that they bind the polynucleotide but do not function as a helicase. The one or more helicases attached to the Y adaptor are preferably stalled at a spacer as discussed in more detail below. The one or more molecular brakes attach to the hairpin loop adaptor are preferably not stalled at a spacer. The one or more helicases and the one or more molecular brakes are preferably brought together when the one or more helicases reach the hairpin loop. The one or more helicases may be attached to the Y adaptor before the Y adaptor is attached to the polynucleotide or after the Y adaptor is attached to the polynucleotide.

The one or more molecular brakes may be attached to the hairpin loop adaptor before the hairpin loop adaptor is attached to the polynucleotide or after the hairpin loop adaptor is attached to the polynucleotide.

5 The one or more helicases and the one or more molecular brakes are preferably not attached to one another. The one or more helicases and the one or more molecular brakes are more preferably not covalently attached to one another. The one or more helicases and the one or more molecular brakes are preferably not attached as described in WO 2014/013260, WO 2014/013259, WO 2014/013262 and WO2015/055981.

10 *Spacers*

The one or more helicases may be stalled at one or more spacers as discussed in WO2014/135838. Any configuration of one or more helicases and one or more spacers disclosed in WO2014/135838 may be used.

15 The one or more spacers are preferably part of the polynucleotide, for instance they interrupt(s) the polynucleotide sequence. The one or more spacers are preferably not part of one or more blocking molecules, such as speed bumps, hybridised to the polynucleotide.

20 There may be any number of spacers in the polynucleotide, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more spacers. There are preferably two, four or six spacers in the polynucleotide. There may be one or more spacers in different regions of the polynucleotide, such as one or more spacers in the Y adaptor and/or hairpin loop adaptor.

25 The one or more spacers may comprise any molecule or combination of molecules that stalls the one or more helicases. The one or more spacers may comprise any molecule or combination of molecules that prevents the one or more helicases from moving along the polynucleotide. It is straightforward to determine whether or not the one or more helicases are stalled at one or more spacers in the absence of a transmembrane pore and an applied potential. For instance, the ability of a helicase to move past a spacer and displace a complementary strand of DNA can be measured by PAGE.

30 The one or more spacers typically comprise a linear molecule, such as a polymer. The one or more spacers typically have a different structure from the polynucleotide. For instance, if the polynucleotide is DNA, the one or more spacers are typically not DNA. In particular, if the polynucleotide is deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), the one or more spacers preferably comprise peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or a synthetic polymer with nucleotide side

chains. The one or more spacers may comprise one or more nucleotides in the opposite direction from the polynucleotide. For instance, the one or more spacers may comprise one or more nucleotides in the 3' to 5' direction when the polynucleotide is in the 5' to 3' direction. The nucleotides may be any of those discussed above.

5 The one or more spacers preferably comprises one or more nitroindoles, such as one or more 5-nitroindoles, one or more inosines, one or more acridines, one or more 2-aminopurines, one or more 2-6-diaminopurines, one or more 5-bromo-deoxyuridines, one or more inverted thymidines (inverted dTs), one or more inverted dideoxy-thymidines (ddTs), one or more dideoxy-cytidines (ddCs), one or more 5-methylcytidines, one or more 5-
10 hydroxymethylcytidines, one or more 2'-O-Methyl RNA bases, one or more Iso-deoxycytidines (Iso-dCs), one or more Iso-deoxyguanosines (Iso-dGs), one or more iSpC3 groups (i.e. nucleotides which lack sugar and a base), one or more photo-cleavable (PC) groups, one or more hexandiol groups, one or more spacer 9 (iSp9) groups, one or more spacer 18 (iSp18) groups, a
15 polymer or one or more thiol connections. The one or more spacers may comprise any combination of these groups. Many of these groups are commercially available from IDT® (Integrated DNA Technologies®).

 The one or more spacers may contain any number of these groups. For instance, for 2-aminopurines, 2-6-diaminopurines, 5-bromo-deoxyuridines, inverted dTs, ddTs, ddCs, 5-methylcytidines, 5-hydroxymethylcytidines, 2'-O-Methyl RNA bases, Iso-dCs, Iso-dGs, iSpC3
20 groups, PC groups, hexandiol groups and thiol connections, the one or more spacers preferably comprise 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more. The one or more spacers preferably comprise 2, 3, 4, 5, 6, 7, 8 or more iSp9 groups. The one or more spacers preferably comprise 2, 3, 4, 5 or 6 or more iSp18 groups. The most preferred spacer is four iSp18 groups.

 The polymer is preferably a polypeptide or a polyethylene glycol (PEG). The
25 polypeptide preferably comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more amino acids. The PEG preferably comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more monomer units.

 The one or more spacers preferably comprise one or more abasic nucleotides (i.e. nucleotides lacking a nucleobase), such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more abasic nucleotides. The nucleobase can be replaced by -H (idSp) or -OH in the abasic nucleotide.
30 Abasic spacers can be inserted into polynucleotides by removing the nucleobases from one or more adjacent nucleotides. For instance, polynucleotides may be modified to include 3-methyladenine, 7-methylguanine, 1,N6-ethenoadenine inosine or hypoxanthine and the nucleobases may be removed from these nucleotides using Human Alkyladenine DNA

Glycosylase (hAAG). Alternatively, polynucleotides may be modified to include uracil and the nucleobases removed with Uracil-DNA Glycosylase (UDG). In one embodiment, the one or more spacers do not comprise any abasic nucleotides.

If linear molecule spacers are used, the polynucleotide may be provided with a double stranded region of polynucleotide adjacent to the end of each spacer past which the one or more helicases are to be moved. The double stranded region typically helps to stall the one or more helicases on the adjacent spacer. The presence of the double stranded region(s) is particularly preferred if the method is carried out at a salt concentration of about 100 mM or lower. Each double stranded region is typically at least 10, such as at least 12, nucleotides in length. If the polynucleotide is single stranded, a double stranded region may be formed by hybridising a shorter polynucleotide to a region adjacent to a spacer. The shorter polynucleotide is typically formed from the same nucleotides as the polynucleotide, but may be formed from different nucleotides. For instance, the shorter polynucleotide may be formed from LNA.

If linear molecule spacers are used, the polynucleotide may be provided with a blocking molecule at the end of each spacer opposite to the end past which the one or more helicases are to be moved. This can help to ensure that the one or more helicases remain stalled on each spacer. It may also help retain the one or more helicases on the polynucleotide in the case that it/they diffuse(s) off in solution. The blocking molecule may be any of the chemical groups discussed below which physically cause the one or more helicases to stall. The blocking molecule may be a double stranded region of polynucleotide.

The one or more spacers preferably comprise one or more chemical groups which physically cause the one or more helicases to stall. The one or more chemical groups are preferably one or more pendant chemical groups. The one or more chemical groups may be attached to one or more nucleobases in the polynucleotide. The one or more chemical groups may be attached to the polynucleotide backbone. Any number of these chemical groups may be present, such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more. Suitable groups include, but are not limited to, fluorophores, streptavidin and/or biotin, cholesterol, methylene blue, dinitrophenols (DNPs), digoxigenin and/or anti-digoxigenin and dibenzylcyclooctyne groups.

Different spacers in the polynucleotide may comprise different stalling molecules. For instance, one spacer may comprise one of the linear molecules discussed above and another spacer may comprise one or more chemical groups which physically cause the one or more helicases to stall. A spacer may comprise any of the linear molecules discussed above and one

or more chemical groups which physically cause the one or more helicases to stall, such as one or more abasics and a fluorophore.

Suitable spacers can be designed depending on the type of polynucleotide and the conditions under which the method is carried out. Most helicases bind and move along DNA and so may be stalled using anything that is not DNA. Suitable molecules are discussed above.

The method is preferably carried out in the presence of free nucleotides and/or the presence of a helicase cofactor. This is discussed in more detail below. In the absence of the transmembrane pore and an applied potential, the one or more spacers are preferably capable of stalling the one or more helicases in the presence of free nucleotides and/or the presence of a helicase cofactor.

If the method is carried out in the presence of free nucleotides and a helicase cofactor as discussed below (such that the one or more helicases are in the active mode), one or more longer spacers are typically used to ensure that the one or more helicases are stalled on the polynucleotide before they are contacted with the transmembrane pore and a potential is applied. One or more shorter spacers may be used in the absence of free nucleotides and a helicase cofactor (such that the one or more helicases are in the inactive mode).

The salt concentration also affects the ability of the one or more spacers to stall the one or more helicases. In the absence of the transmembrane pore and an applied potential, the one or more spacers are preferably capable of stalling the one or more helicases at a salt concentration of about 100 mM or lower. The higher the salt concentration used in the method, the shorter the one or more spacers that are typically used and *vice versa*.

Preferred combinations of features are shown in the Table below.

| Polynucleotide | Spacer composition* | Spacer length (i.e. number of *) | Salt [] | Free nucleotides? | Helicase cofactor? |
|----------------|---------------------|----------------------------------|--------------|-------------------|--------------------|
| DNA | iSpC3 | 4 | 1 M | Yes | Yes |
| DNA | iSp18 | 4 | 100-1000 mM | Yes | Yes |
| DNA | iSp18 | 6 | <100-1000 mM | Yes | Yes |
| DNA | iSp18 | 2 | 1 M | Yes | Yes |
| DNA | iSpC3 | 12 | <100-1000 mM | Yes | Yes |
| DNA | iSpC3 | 20 | <100-1000 mM | Yes | Yes |
| DNA | iSp9 | 6 | 100-1000 mM | Yes | Yes |

| | | | | | |
|-----|------|---|-----|-----|-----|
| DNA | idSp | 4 | 1 M | Yes | Yes |
|-----|------|---|-----|-----|-----|

The method may involve moving two or more helicases past a spacer. In such instances, the length of the spacer is typically increased to prevent the trailing helicase from pushing the leading helicase past the spacer in the absence of the pore and applied potential. If the method concerns moving two or more helicases past one or more spacers, the spacer lengths discussed above may be increased at least 1.5 fold, such 2 fold, 2.5 fold or 3 fold. For instance, if the method concerns moving two or more helicases past one or more spacers, the spacer lengths in the third column of Table 4 above may be increased 1.5 fold, 2 fold, 2.5 fold or 3 fold.

10 *Membrane*

The double pore or pore as disclosed herein may be present in a membrane. In the method, the polynucleotide is typically contacted with the double pore or pore in a membrane. Any membrane may be used. 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 that 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 as amphiphiles. The copolymer may be a triblock, tetrablock or pentablock copolymer. The membrane is preferably a triblock copolymer membrane.

30 Archaeobacterial 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 synthesised, 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 customise polymer based membranes for a wide range of applications.

The membrane is most preferably one of the membranes disclosed in International Application No. WO2014/064443 or WO2014/064444.

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 WO 2008/102121, WO 2009/077734 and WO 2006/100484.

5 Methods for forming lipid bilayers are known in the art. Lipid bilayers are commonly formed by the method of Montal and Mueller (Proc. Natl. Acad. Sci. USA., 1972; 69: 3561-3566), in which a lipid monolayer is carried on aqueous solution/air interface past either side of an aperture which is perpendicular to that interface. The lipid is normally added to the surface of an aqueous electrolyte solution by first dissolving it in an organic solvent and then allowing a drop of the solvent to evaporate on the surface of the aqueous solution on either side of the
10 aperture. Once the organic solvent has evaporated, the solution/air interfaces on either side of the aperture are physically moved up and down past the aperture until a bilayer is formed. Planar lipid bilayers may be formed across an aperture in a membrane or across an opening into a recess.

The method of Montal & Mueller is popular because it is a cost-effective and relatively
15 straightforward method of forming good quality lipid bilayers that are suitable for protein pore insertion. Other common methods of bilayer formation include tip-dipping, painting bilayers and patch-clamping of liposome bilayers.

Tip-dipping bilayer formation entails touching the aperture surface (for example, a pipette tip) onto the surface of a test solution that is carrying a monolayer of lipid. Again, the lipid
20 monolayer is first generated at the solution/air interface by allowing a drop of lipid dissolved in organic solvent to evaporate at the solution surface. The bilayer is then formed by the Langmuir-Schaefer process and requires mechanical automation to move the aperture relative to the solution surface.

For painted bilayers, a drop of lipid dissolved in organic solvent is applied directly to the
25 aperture, which is submerged in an aqueous test solution. The lipid solution is spread thinly over the aperture using a paintbrush or an equivalent. Thinning of the solvent results in formation of a lipid bilayer. However, complete removal of the solvent from the bilayer is difficult and consequently the bilayer formed by this method is less stable and more prone to noise during electrochemical measurement.

30 Patch-clamping is commonly used in the study of biological cell membranes. The cell membrane is clamped to the end of a pipette by suction and a patch of the membrane becomes attached over the aperture. The method has been adapted for producing lipid bilayers by clamping liposomes which then burst to leave a lipid bilayer sealing over the aperture of the pipette. The

method requires stable, giant and unilamellar liposomes and the fabrication of small apertures in materials having a glass surface.

Liposomes can be formed by sonication, extrusion or the Mozafari method (Colas *et al.* (2007) *Micron* 38:841–847).

5 In a preferred embodiment, the lipid bilayer is formed as described in International Application No. WO 2009/077734. Advantageously in this method, the lipid bilayer is formed from dried lipids. In a most preferred embodiment, the lipid bilayer is formed across an opening as described in WO2009/077734.

10 A lipid bilayer is formed from two opposing layers of lipids. The two layers of lipids are arranged such that their hydrophobic tail groups face towards each other to form a hydrophobic interior. The hydrophilic head groups of the lipids face outwards towards the aqueous environment on each side of the bilayer. The bilayer may be present in a number of lipid phases including, but not limited to, the liquid disordered phase (fluid lamellar), liquid ordered phase, solid ordered phase (lamellar gel phase, interdigitated gel phase) and planar bilayer crystals
15 (lamellar sub-gel phase, lamellar crystalline phase).

Any lipid composition that forms a lipid bilayer may be used. The lipid composition is chosen such that a lipid bilayer having the required properties, such surface charge, ability to support membrane proteins, packing density or mechanical properties, is formed. The lipid composition can comprise one or more different lipids. For instance, the lipid composition can
20 contain up to 100 lipids. The lipid composition preferably contains 1 to 10 lipids. The lipid composition may comprise naturally-occurring lipids and/or artificial lipids.

The lipids typically comprise a head group, an interfacial moiety and two hydrophobic tail groups which may be the same or different. Suitable head groups include, but are not limited to, neutral head groups, such as diacylglycerides (DG) and ceramides (CM); zwitterionic head
25 groups, such as phosphatidylcholine (PC), phosphatidylethanolamine (PE) and sphingomyelin (SM); negatively charged head groups, such as phosphatidylglycerol (PG); phosphatidylserine (PS), phosphatidylinositol (PI), phosphatic acid (PA) and cardiolipin (CA); and positively charged headgroups, such as trimethylammonium-Propane (TAP). Suitable interfacial moieties include, but are not limited to, naturally-occurring interfacial moieties, such as glycerol-based or
30 ceramide-based moieties. Suitable hydrophobic tail groups include, but are not limited to, saturated hydrocarbon chains, such as lauric acid (*n*-Dodecanolic acid), myristic acid (*n*-Tetradecanoic acid), palmitic acid (*n*-Hexadecanoic acid), stearic acid (*n*-Octadecanoic) and arachidic (*n*-Eicosanoic); unsaturated hydrocarbon chains, such as oleic acid (*cis*-9-

Octadecanoic); and branched hydrocarbon chains, such as phytanoyl. The length of the chain and the position and number of the double bonds in the unsaturated hydrocarbon chains can vary. The length of the chains and the position and number of the branches, such as methyl groups, in the branched hydrocarbon chains can vary. The hydrophobic tail groups can be linked to the interfacial moiety as an ether or an ester. The lipids may be mycolic acid.

The lipids can also be chemically-modified. The head group or the tail group of the lipids may be chemically-modified. Suitable lipids whose head groups have been chemically-modified include, but are not limited to, PEG-modified lipids, such as 1,2-Diacyl-sn-Glycero-3-Phosphoethanolamine-N-[Methoxy(Polyethylene glycol)-2000]; functionalised PEG Lipids, such as 1,2-Distearoyl-sn-Glycero-3 Phosphoethanolamine-N-[Biotinyl(Polyethylene Glycol)2000]; and lipids modified for conjugation, such as 1,2-Dioleoyl-sn-Glycero-3-Phosphoethanolamine-N-(succinyl) and 1,2-Dipalmitoyl-sn-Glycero-3-Phosphoethanolamine-N-(Biotinyl). Suitable lipids whose tail groups have been chemically-modified include, but are not limited to, polymerisable lipids, such as 1,2-bis(10,12-tricosadiynoyl)-sn-Glycero-3-Phosphocholine; fluorinated lipids, such as 1-Palmitoyl-2-(16-Fluoropalmitoyl)-sn-Glycero-3-Phosphocholine; deuterated lipids, such as 1,2-Dipalmitoyl-D62-sn-Glycero-3-Phosphocholine; and ether linked lipids, such as 1,2-Di-O-phytanoyl-sn-Glycero-3-Phosphocholine. The lipids may be chemically-modified or functionalised to facilitate coupling of the polynucleotide.

The amphiphilic layer, for example the lipid composition, typically comprises one or more additives that will affect the properties of the layer. Suitable additives include, but are not limited to, fatty acids, such as palmitic acid, myristic acid and oleic acid; fatty alcohols, such as palmitic alcohol, myristic alcohol and oleic alcohol; sterols, such as cholesterol, ergosterol, lanosterol, sitosterol and stigmasterol; lysophospholipids, such as 1-Acyl-2-Hydroxy-sn-Glycero-3-Phosphocholine; and ceramides.

In another preferred embodiment, the membrane comprises a solid state layer. Solid state layers can be formed from both organic and inorganic materials including, but not limited to, microelectronic materials, insulating materials such as Si_3N_4 , Al_2O_3 , and SiO_2 , organic and inorganic polymers such as polyamide, plastics such as Teflon® or elastomers such as two-component addition-cure silicone rubber, and glasses. The solid state layer may be formed from graphene. Suitable graphene layers are disclosed in WO 2009/035647. If the membrane comprises a solid state layer, the pore is typically present in an amphiphilic membrane or layer contained within the solid state layer, for instance within a hole, well, gap, channel, trench or slit within the solid state layer. The skilled person can prepare suitable solid state/amphiphilic

hybrid systems. Suitable systems are disclosed in WO 2009/020682 and WO 2012/005857. Any of the amphiphilic membranes or layers discussed above may be used.

5 The method is typically carried out using (i) an artificial amphiphilic layer comprising a pore, (ii) an isolated, naturally-occurring lipid bilayer comprising a pore, or (iii) a cell having a pore inserted therein. The method is typically carried out using an artificial amphiphilic layer, such as an artificial triblock copolymer layer. The layer may comprise other transmembrane and/or intramembrane proteins as well as other molecules in addition to the pore. Suitable apparatus and conditions are discussed below. The method is typically carried out *in vitro*.

10 *Coupling*

The polynucleotide is preferably coupled to the membrane comprising the pore. The method may comprise coupling the polynucleotide to the membrane comprising the pore. 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.

15 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).

20 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.

25 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.

30 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 functionalised, for example at least 0.01%, at least 0.1%, at least 1%, at least 10%, at least 25%, at least 50% or 100%.

5 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 couple more than one, such as 2, 3, 4 or more, polynucleotides to the membrane.

10 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 one or more anchors or one or more linkers may comprise a component that can be cut to broken down, such as a restriction site or a photolabile group.

15 Functionalised linkers and the ways in which they can couple molecules are known in the art. For instance, linkers functionalised with maleimide groups will react with and attach to cysteine residues in proteins. In the present context, the protein may be present in the membrane or may be used to couple (or bind) to the polynucleotide. This is discussed in more detail below.

20 Crosslinkage of polynucleotides can be avoided using a “lock and key” arrangement. Only one end of each linker may react together to form a longer linker and the other ends of the linker each react with the polynucleotide or membrane respectively. Such linkers are described in WO 2010/086602.

25 The use of a linker is preferred in the sequencing embodiments discussed below. If a polynucleotide is permanently coupled directly to the membrane in the sense that it does not uncouple when interacting with the pore (i.e. does not uncouple in step (b) or (e)), then some sequence data will be lost as the sequencing run cannot continue to the end of the polynucleotide due to the distance between the membrane and the pore. If a linker is used, then the polynucleotide can be processed to completion.

30 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.

For certain applications, such as aptamer detection, the transient nature of the coupling is preferred. If a permanent or stable linker is attached directly to either the 5' or 3' end of a polynucleotide and the linker is shorter than the distance between the membrane and the transmembrane pore's channel, then some sequence data will be lost as the sequencing run cannot continue to the end of the polynucleotide. If the coupling is transient, then when the coupled end randomly becomes free of the membrane, then the polynucleotide can be processed to completion. Chemical groups that form permanent/stable or transient links are discussed in more detail below. The polynucleotide may be transiently coupled to an amphiphilic layer or triblock copolymer membrane using cholesterol or a fatty acyl chain. Any fatty acyl chain having a length of from 6 to 30 carbon atom, such as hexadecanoic acid, may be used.

In preferred embodiments, a polynucleotide, such as a nucleic acid, is coupled to an amphiphilic layer such as a triblock copolymer membrane or lipid bilayer. Coupling of nucleic acids to synthetic lipid bilayers has been carried out previously with various different tethering strategies. These are summarised in Table 5 below.

15

Table 5

| Anchor comprising | Type of coupling | Reference |
|---|-------------------------|--|
| Thiol | Stable | Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." <u>J Am Chem Soc</u> 125 (13): 3696-7. |
| Biotin | Stable | Nikolov, V., R. Lipowsky, et al. (2007). "Behavior of giant vesicles with anchored DNA molecules." <u>Biophys J</u> 92 (12): 4356-68 |
| Cholesterol | Transient | Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies." <u>J Am Chem Soc</u> 126 (33): 10224-5 |
| Surfactant (e.g. Lipid, Palmitate, etc) | Stable | van Lengerich, B., R. J. Rawle, et al. "Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." <u>Langmuir</u> 26 (11): 8666-72 |

Synthetic polynucleotides and/or linkers may be functionalised using a modified phosphoramidite in the synthesis reaction, which is easily compatible for the direct addition of suitable anchoring groups, such as cholesterol, tocopherol, palmitate, thiol, lipid and biotin groups. These different attachment chemistries give a suite of options for attachment to polynucleotides. Each different modification group couples the polynucleotide in a slightly

20

different way and coupling is not always permanent so giving different dwell times for the polynucleotide to the membrane. The advantages of transient coupling are discussed above.

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. A thiol group can be added to the 5' of ssDNA or dsDNA using T4 polynucleotide kinase and ATP γ S (Grant, G. P. and P. Z. Qin (2007). "A facile method for attaching nitroxide spin labels at the 5' terminus of nucleic acids." Nucleic Acids Res **35**(10): e77). An azide group can be added to the 5'-phosphate of ssDNA or dsDNA using T4 polynucleotide kinase and γ -[2-Azidoethyl]-ATP or γ -[6-Azidoethyl]-ATP. Using thiol or Click chemistry a tether, containing either a thiol, iodoacetamide OPSS or maleimide group (reactive to thiols) or a DIBO (dibenzocyclooctyne) or alkyne group (reactive to azides), can be covalently attached to the polynucleotide. A more diverse selection of chemical groups, such as biotin, thiols and fluorophores, can be added using terminal transferase to incorporate modified oligonucleotides to the 3' of ssDNA (Kumar, A., P. Tchen, et al. (1988). "Nonradioactive labeling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase." Anal Biochem **169**(2): 376-82). Streptavidin/biotin and/or streptavidin/desthiobiotin coupling may be used for any other polynucleotide. The Examples below describes how a polynucleotide can be coupled to a membrane using streptavidin/biotin and streptavidin/desthiobiotin. It may also be possible that anchors may be directly added to polynucleotides using terminal transferase with suitably modified nucleotides (e.g. cholesterol or palmitate).

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 hybridisation may be present in any part of the one or more anchors, such as between the one or more anchors and the polynucleotide, within the one or more anchors or between the one or more anchors and the membrane. For instance, a linker may comprise two or more polynucleotides, such as 3, 4 or 5 polynucleotides, hybridised together. The one or more anchors may hybridise to the polynucleotide. The one or more anchors may hybridise directly to the polynucleotide or directly to a Y adaptor and/or leader sequence attached to the polynucleotide or directly to a hairpin loop adaptor attached to the polynucleotide (as discussed below). Alternatively, the one or more anchors may be hybridised to one or more, such as 2 or 3, intermediate polynucleotides (or "splints") which are hybridised to the polynucleotide, to a Y

adaptor and/or leader sequence attached to the polynucleotide or to a hairpin loop adaptor attached to the polynucleotide (as discussed below).

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. To a double stranded polynucleotide, it is possible to add either a piece of single stranded polynucleotide to one or both of the ends of the duplex, or a double stranded polynucleotide to one or both ends. For addition of single stranded polynucleotides to the a double stranded polynucleotide, this can be achieved using T4 RNA ligase I as for ligation to other regions of single stranded polynucleotides. For addition of double stranded polynucleotides to a double stranded polynucleotide then ligation can be "blunt-ended", with complementary 3' dA / dT tails on the polynucleotide and added polynucleotide respectively (as is routinely done for many sample prep applications to prevent concatemer or dimer formation) or using "sticky-ends" generated by restriction digestion of the polynucleotide and ligation of compatible adapters. Then, when the duplex is melted, each single strand will have either a 5' or 3' modification if a single stranded polynucleotide was used for ligation or a modification at the 5' end, the 3' end or both if a double stranded polynucleotide was used for ligation.

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.

Adenylated polynucleotides are intermediates in ligation reactions, where an adenosine-monophosphate is attached to the 5'-phosphate of the polynucleotide. Various kits are available for generation of this intermediate, such as the 5' DNA Adenylation Kit from NEB. By substituting ATP in the reaction for a modified nucleotide triphosphate, then addition of reactive groups (such as thiols, amines, biotin, azides, etc) to the 5' of a polynucleotide can be possible. It may also be possible that anchors could be directly added to polynucleotides using a 5' DNA adenylation kit with suitably modified nucleotides (e.g. cholesterol or palmitate).

A common technique for the amplification of sections of genomic DNA is using polymerase chain reaction (PCR). Here, using two synthetic oligonucleotide primers, a number

of copies of the same section of DNA can be generated, where for each copy the 5' of each strand in the duplex will be a synthetic polynucleotide. Single or multiple nucleotides can be added to 3' end of single or double stranded DNA by employing a polymerase. Examples of polymerases which could be used include, but are not limited to, Terminal Transferase, Klenow and *E. coli* Poly(A) polymerase). By substituting ATP in the reaction for a modified nucleotide triphosphate then anchors, such as a cholesterol, thiol, amine, azide, biotin or lipid, can be incorporated into double stranded polynucleotides. Therefore, each copy of the amplified polynucleotide will contain an anchor.

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 functionalising 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.

The one or more anchors can comprise any group that couples to, binds to or interacts with single or double stranded polynucleotides, specific nucleotide sequences within the polynucleotide or patterns of modified nucleotides within the polynucleotide, or any other ligand that is present on the polynucleotide.

Suitable binding proteins for use in anchors include, but are not limited to, *E. coli* single stranded binding protein, P5 single stranded binding protein, T4 gp32 single stranded binding protein, the TOPO V dsDNA binding region, human histone proteins, *E. coli* HU DNA binding protein and other archaeal, prokaryotic or eukaryotic single stranded or double stranded polynucleotide (or nucleic acid) binding proteins, including those listed below.

The specific nucleotide sequences could be sequences recognised by transcription factors, ribosomes, endonucleases, topoisomerases or replication initiation factors. The patterns of modified nucleotides could be patterns of methylation or damage.

The one or more anchors can comprise any group which couples to, binds to, intercalates with or interacts with a polynucleotide. The group may intercalate or interact with the polynucleotide via electrostatic, hydrogen bonding or Van der Waals interactions. Such groups

include a lysine monomer, poly-lysine (which will interact with ssDNA or dsDNA), ethidium bromide (which will intercalate with dsDNA), universal bases or universal nucleotides (which can hybridise with any polynucleotide) and osmium complexes (which can react to methylated bases). A polynucleotide may therefore be coupled to the membrane using one or more
5 universal nucleotides attached to the membrane. Each universal nucleotide may be coupled to the membrane using one or more linkers. The universal nucleotide preferably comprises one of the following nucleobases: hypoxanthine, 4-nitroindole, 5-nitroindole, 6-nitroindole, formylindole, 3-nitropyrrole, nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole, 5-nitroindazole, 4-aminobenzimidazole or phenyl (C6-aromatic ring). The universal nucleotide
10 more preferably comprises one of the following nucleosides: 2'-deoxyinosine, inosine, 7-deaza-2'-deoxyinosine, 7-deaza-inosine, 2-aza-deoxyinosine, 2-aza-inosine, 2-O'-methylinosine, 4-nitroindole 2'-deoxyribonucleoside, 4-nitroindole ribonucleoside, 5-nitroindole 2'-deoxyribonucleoside, 5-nitroindole ribonucleoside, 6-nitroindole 2'-deoxyribonucleoside, 6-nitroindole ribonucleoside, 3-nitropyrrole 2'-deoxyribonucleoside, 3-nitropyrrole ribonucleoside,
15 an acyclic sugar analogue of hypoxanthine, nitroimidazole 2'-deoxyribonucleoside, nitroimidazole ribonucleoside, 4-nitropyrazole 2'-deoxyribonucleoside, 4-nitropyrazole ribonucleoside, 4-nitrobenzimidazole 2'-deoxyribonucleoside, 4-nitrobenzimidazole ribonucleoside, 5-nitroindazole 2'-deoxyribonucleoside, 5-nitroindazole ribonucleoside, 4-aminobenzimidazole 2'-deoxyribonucleoside, 4-aminobenzimidazole ribonucleoside, phenyl C-
20 ribonucleoside, phenyl C-2'-deoxyribosyl nucleoside, 2'-deoxynebularine, 2'-deoxyisoguanosine, K-2'-deoxyribose, P-2'-deoxyribose and pyrrolidine. The universal nucleotide more preferably comprises 2'-deoxyinosine. The universal nucleotide is more preferably IMP or dIMP. The universal nucleotide is most preferably dPMP (2'-Deoxy-P-nucleoside monophosphate) or dKMP (N6-methoxy-2, 6-diaminopurine monophosphate).

25 The one or more anchors may couple to (or bind to) the polynucleotide via Hoogsteen hydrogen bonds (where two nucleobases are held together by hydrogen bonds) or reversed Hoogsteen hydrogen bonds (where one nucleobase is rotated through 180° with respect to the other nucleobase). For instance, the one or more anchors may comprise one or more nucleotides, one or more oligonucleotides or one or more polynucleotides which form Hoogsteen hydrogen
30 bonds or reversed Hoogsteen hydrogen bonds with the polynucleotide. These types of hydrogen bonds allow a third polynucleotide strand to wind around a double stranded helix and form a triplex. The one or more anchors may couple to (or bind to) a double stranded polynucleotide by forming a triplex with the double stranded duplex.

In this embodiment at least 1%, at least 10%, at least 25%, at least 50% or 100% of the membrane components may be functionalised.

Where the one or more anchors comprise a protein, they may be able to anchor directly into the membrane without further functionalisation, 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. Alternatively the protein may be expressed with a genetically fused hydrophobic region which is compatible with the membrane. Such hydrophobic protein regions are known in the art.

The one or more anchors are preferably mixed with the polynucleotide before contacting with the membrane, but the one or more anchors may be contacted with the membrane and subsequently contacted with the polynucleotide.

In another aspect the polynucleotide may be functionalised, using methods described above, so that it can be recognised by a specific binding group. Specifically the polynucleotide may be functionalised with a ligand such as biotin (for binding to streptavidin), amylose (for binding to maltose binding protein or a fusion protein), Ni-NTA (for binding to poly-histidine or poly-histidine tagged proteins) or a peptides (such as an antigen).

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.

An example of a molecule used in chemical attachment is EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride). Reactive groups can also be added to the 5' of polynucleotides using commercially available kits (Thermo Pierce, Part No. 22980). Suitable methods include, but are not limited to, transient affinity attachment using histidine residues and

Ni-NTA, as well as more robust covalent attachment by reactive cysteines, lysines or non natural amino acids.

Double stranded polynucleotide

5 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. The two strands may be separated as the polynucleotide movement through the pore is controlled by a
10 polynucleotide binding protein, such as a helicase, or molecular brake.

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

 The bridging moiety is capable of linking the two strands of the target polynucleotide. The bridging moiety typically covalently links the two strands of the target polynucleotide. The
15 bridging moiety can be anything that is capable of linking the two strands of the target polynucleotide, provided that the bridging moiety does not interfere with movement of the single stranded polynucleotide through the transmembrane pore.

 The bridging moiety may be linked to the target polynucleotide by any suitable means known in the art. The bridging moiety may be synthesised separately and chemically attached or
20 enzymatically ligated to the target polynucleotide. Alternatively, the bridging moiety may be generated in the processing of the target polynucleotide.

 The bridging moiety is linked to the target polynucleotide at or near one end of the target polynucleotide. The bridging moiety is preferably linked to the target polynucleotide within 10 nucleotides of the end of the target polynucleotide

25 Suitable bridging moieties include, but are not limited to a polymeric linker, a chemical linker, a polynucleotide or a polypeptide. Preferably, the bridging moiety comprises DNA, RNA, modified DNA (such as abasic DNA), RNA, PNA, LNA or PEG. The bridging moiety is more preferably DNA or RNA.

 The bridging moiety is most preferably a hairpin loop or a hairpin loop adaptor. Suitable
30 hairpin adaptors can be designed using methods known in the art. The hairpin loop may be any length. The hairpin loop is typically 110 or fewer nucleotides, such as 100 or fewer nucleotides, 90 or fewer nucleotides, 80 or fewer nucleotides, 70 or fewer nucleotides, 60 or fewer nucleotides, 50 or fewer nucleotides, 40 or fewer nucleotides, 30 or fewer nucleotides, 20 or

fewer nucleotides or 10 or fewer nucleotides, in length. The hairpin loop is preferably from about 1 to 110, from 2 to 100, from 5 to 80 or from 6 to 50 nucleotides in length. Longer lengths of the hairpin loop, such as from 50 to 110 nucleotides, are preferred if the loop is involved in the differential selectability of the adaptor. Similarly, shorter lengths of the hairpin loop, such as from 1 to 5 nucleotides, are preferred if the loop is not involved in the selectable binding as discussed below.

The hairpin adaptor may be ligated to either end of the first and/or second polynucleotide, i.e. the 5' or the 3' end. The hairpin adaptor may be ligated to the first and/or second polynucleotide using any method known in the art. The hairpin adaptor may be ligated using a ligase, such as T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase and 9°N DNA ligase.

The two strands of the polynucleotide may be separated using any method known in the art. For instance, they may be separated by a polynucleotide binding protein or using conditions which favour dehybridisation (examples of conditions which favour dehybridisation include, but are not limited to, high temperature, high pH and the addition of agents that can disrupt hydrogen bonding or base pairing, such as formamide and urea).

The hairpin adaptor preferably comprises a selectable binding moiety. This allows the first and/or second polynucleotide to be purified or isolated. A selectable binding moiety is a moiety that can be selected on the basis of its binding properties. Hence, a selectable binding moiety is preferably a moiety that specifically binds to a surface. A selectable binding moiety specifically binds to a surface if it binds to the surface to a much greater degree than any other moiety used. In preferred embodiments, the moiety binds to a surface to which no other moiety binds.

Suitable selective binding moieties are known in the art. Preferred selective binding moieties include, but are not limited to, biotin, a polynucleotide sequence, antibodies, antibody fragments, such as Fab and scFv, antigens, polynucleotide binding proteins, poly histidine tails and GST tags. The most preferred selective binding moieties are biotin and a selectable polynucleotide sequence. Biotin specifically binds to a surface coated with avidins. Selectable polynucleotide sequences specifically bind (i.e. hybridise) to a surface coated with homologous sequences. Alternatively, selectable polynucleotide sequences specifically bind to a surface coated with polynucleotide binding proteins.

The hairpin adaptor and/or the selectable binding moiety may comprise a region that can be cut, nicked, cleaved or hydrolysed. Such a region can be designed to allow the first and/or

second polynucleotide to be removed from the surface to which it is bound following purification or isolation. Suitable regions are known in the art. Suitable regions include, but are not limited to, an RNA region, a region comprising desthiobiotin and streptavidin, a disulphide bond and a photocleavable region.

5 The double stranded target polynucleotide preferably comprises a leader sequence at the opposite end of the bridging moiety, such as a hairpin loop or hairpin loop adaptor. Leader sequences are discussed in more detail below.

Round the Corner Sequencing

10 In a preferred embodiment, a target double stranded polynucleotide is provided with a bridging moiety, such as a hairpin loop or hairpin loop adaptor, at one end and the method comprises contacting the polynucleotide with the pore 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
15 one or more characteristics of the strands of the polynucleotide and thereby characterising the target double stranded polynucleotide. In another preferred embodiment, a target double stranded polynucleotide is provided with a bridging moiety, such as a hairpin loop or hairpin loop adaptor, at one end and the method comprises contacting the polynucleotide with the pore and exonuclease such that both strands of the polynucleotide are digested to form individual
20 nucleotides. Any of the embodiments discussed above equally apply to this embodiment.

Leader Sequence

 Before the contacting step in the strand characterisation/sequencing method, the method preferably comprises attaching to the polynucleotide a leader sequence which preferentially
25 threads into the pore. The leader sequence facilitates the method. The leader sequence is designed to preferentially thread into the pore 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.

 The leader sequence typically comprises a polymer. The polymer is preferably
30 negatively charged. The polymer is preferably a polynucleotide, such as DNA or RNA, a modified polynucleotide (such as abasic DNA), PNA, LNA, polyethylene glycol (PEG) or a polypeptide. The leader preferably comprises a polynucleotide and more preferably comprises a single stranded polynucleotide. The leader sequence can comprise any of the polynucleotides

discussed above. The single stranded leader sequence most preferably comprises a single strand of DNA, such as a poly dT section. The leader sequence preferably comprises the one or more spacers.

5 The leader sequence can be any length, but is typically 10 to 150 nucleotides in length, such as from 20 to 150 nucleotides in length. The length of the leader typically depends on the transmembrane pore used in the method.

The leader sequence is preferably part of a Y adaptor as defined below.

Double Coupling

10 The method may involve double coupling of a double stranded polynucleotide. In a preferred embodiment, the method 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
15 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 a pore, for example a double
20 pore, as disclosed herein such that the polynucleotide moves with respect to, such as 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.

25 This type of method is discussed in detail in the UK Application No. 1406147.7.

The double stranded polynucleotide is provided with a Y adaptor at one end and a bridging moiety adaptor at the other end. The Y adaptor and/or the bridging moiety adaptor are typically polynucleotide adaptors. They may be formed from any of the polynucleotides discussed above.

30 The Y adaptor typically comprises (a) a double stranded region and (b) a single stranded region or a region that is not complementary at the other end. The Y adaptor may be described as having an overhang if it comprises a single stranded region. The presence of a non-complementary region in the Y adaptor gives the adaptor its Y shape since the two strands

typically do not hybridise to each other unlike the double stranded portion. The Y adaptor comprises the one or more first anchors. Anchors are discussed in more detail above.

The Y adaptor preferably comprises a leader sequence which preferentially threads into the pore. This is discussed above.

5 The bridging moiety adaptor preferably comprises a selectable binding moiety as discussed above. The bridging moiety adaptor and/or the selectable binding moiety may comprise a region that can be cut, nicked, cleaved or hydrolysed as discussed above.

If one or more helicases and one or more molecular brakes are used as discussed above, the Y adaptor preferably comprises the one or more helicases and the bridging moiety adaptor
10 preferably comprises the one or more molecular brakes.

The Y adaptor and/or the bridging moiety adaptor may be ligated to the polynucleotide using any method known in the art. One or both of the adaptors may be ligated using a ligase, such as T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase and 9°N DNA ligase. Alternatively, the adaptors may be added to the polynucleotide using the methods
15 discussed below.

In a preferred embodiment, step a) of the method comprises modifying the double stranded polynucleotide so that it comprises the Y adaptor at one end and the bridging moiety adaptor at the other end. Any manner of modification can be used. The method preferably comprises modifying the double stranded polynucleotide. This is discussed in more detail
20 below. The methods of modification and characterisation may be combined in any way.

The strength of coupling (or binding) of the bridging moiety adaptor to the membrane is greater than the strength of coupling (or binding) of the Y adaptor to the membrane. This can be measured in any way. A suitable method for measuring the strength of coupling (or binding) is disclosed in the Examples of the UK Application No. 1406147.7.

25 The strength of coupling (or binding) of the bridging moiety adaptor is preferably at least 1.5 times the strength of coupling (or binding) of the Y adaptor, such as at least twice, at least three times, at least four times, at least five or at least ten times the strength of coupling (or binding) of the anchor adaptor. The affinity constant (Kd) of the bridging moiety adaptor for the membrane is preferably at least 1.5 times the affinity constant of the Y adaptor, such as at least
30 twice, at least three times, at least four times, at least five or at least ten times the strength of coupling of the Y adaptor.

There are several ways in which the bridging moiety adaptor couples (or binds) more strongly to the membrane than the Y adaptor. For instance, the bridging moiety adaptor may

comprise more anchors than the Y adaptor. For instance, the bridging moiety adaptor may comprise 2, 3 or more second anchors whereas the Y adaptor may comprise one first anchor.

The strength of coupling (or binding) of the one or more second anchors to the membrane may be greater than the strength of coupling (or binding) of the one or more first anchors to the membrane. The strength of coupling (or binding) of the one or more second anchors to the bridging moiety adaptor may be greater than the strength of coupling (or binding) of the one or more first anchors to the Y adaptor. The one or more first anchors and the one or more second anchors may be attached to their respective adaptors via hybridisation and the strength of hybridisation is greater in the one or more second anchors than in the one or more first anchors. Any combination of these embodiments may also be used. Strength of coupling (or binding) may be measure using known techniques in the art.

The one or more second anchors preferably comprise one or more groups which couple(s) (or bind(s)) to the membrane with a greater strength than the one or more groups in the one or more first anchors which couple(s) (or bind(s)) to the membrane. In preferred embodiments, the bridging moiety adaptor/one or more second anchors couple (or bind) to the membrane using cholesterol and the Y adaptor/one or more first anchors couple (or bind) to the membrane using palmitate. Cholesterol binds to triblock copolymer membranes and lipid membranes more strongly than palmitate. In an alternative embodiment, the bridging moiety adaptor/one or more second anchors couple (or bind) to the membrane using a mono-acyl species, such as palmitate, and the Y adaptor/one or more first anchors couple (or bind) to the membrane using a diacyl species, such as dipalmitoylphosphatidylcholine.

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.

Each substrate in the population preferably comprises at least one overhang of universal nucleotides such that the transposase fragments the template polynucleotide and ligates a

substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs and wherein the method further comprises ligating the overhangs to the fragments in the constructs and thereby producing a plurality of modified double stranded polynucleotides. Suitable universal nucleotides are discussed above. The overhang is preferably five nucleotides in length.

Alternatively, each substrate in population preferably comprises (i) at least one overhang and (ii) at least one nucleotide in the same strand as the at least one overhang which comprises a nucleoside that is not present in the template polynucleotide such that the transposase fragments the template polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs, and wherein the method further comprises (a) removing the overhangs from the constructs by selectively removing the at least one nucleotide and thereby producing a plurality of double stranded constructs comprising single stranded gaps and (b) repairing the single stranded gaps in the constructs and thereby producing a plurality of modified double stranded polynucleotides. The polynucleotide typically comprises the nucleosides deoxyadenosine (dA), deoxyuridine (dU) and/or thymidine (dT), deoxyguanosine (dG) and deoxycytidine (dC). The nucleoside that is not present in the polynucleotide is preferably abasic, adenosine (A), uridine (U), 5-methyluridine (m⁵U), cytidine (C) or guanosine (G) or comprises urea, 5, 6 dihydroxythymine, thymine glycol, 5-hydroxy-5 methylhydantoin, uracil glycol, 6-hydroxy-5, 6-dihydrothymine, methyltartronylurea, 7, 8-dihydro-8-oxoguanine (8-oxoguanine), 8-oxoadenine, fapy-guanine, methy-fapy-guanine, fapy-adenine, aflatoxin B1-fapy-guanine, 5-hydroxy-cytosine, 5-hydroxy-uracil, 3-methyladenine, 7-methylguanine, 1,N6-ethenoadenine, hypoxanthine, 5-hydroxyuracil, 5-hydroxymethyluracil, 5-formyluracil or a cis-syn-cyclobutane pyrimidine dimer. The at least one nucleotide preferably is 10 nucleotides or fewer from the overhang. The at least one nucleotide is the first nucleotide in the overhang. All of the nucleotides in the overhang preferably comprise a nucleoside that is not present in the template polynucleotide.

These MuA based methods are disclosed in International Application No. PCT/GB2014/052505. They are also discussed in detail in the UK Application No. 1406147.7.

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

In a preferred embodiment, the method 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 a pore as disclosed herein such that the polynucleotide moves with respect to, such as 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 as disclosed herein such that the second polynucleotide moves with respect to, such as 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 the UK Application No. 1406155.0.

Step (f) (i.e. uncoupling of the first polynucleotide) may be performed before step (g) (i.e. before coupling the second polynucleotide to the membrane). Step (g) may be performed before step (f). If the second polynucleotide is coupled to the membrane before the first polynucleotide is uncoupled, step (f) preferably comprises selectively uncoupling the first

polynucleotide from the membrane (i.e. uncoupling the first polynucleotide but not the second polynucleotide from the membrane). A skilled person can design a system in which selective uncoupling is achieved. Steps (f) and (g) may be performed at the same time. This is discussed in more detail below.

5 In step (f), at least 10% of the first polynucleotide is preferably uncoupled from the membrane. For instance, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 95% of the first polynucleotide may be uncoupled from the membrane. Preferably, all of the first polynucleotide is uncoupled from the membrane. The amount of the first polynucleotide uncoupled from the membrane can be
10 determined using the pore. This is disclosed in the Examples.

The first polynucleotide and second polynucleotide may be different from one another. Alternatively, the first and second polynucleotides may be different polynucleotides. In such instances, there may be no need to remove at least part of the first sample before adding the second polynucleotide. This is discussed in more detail below. If the method concerns
15 investigating three or more polynucleotides, they may all be different from one another or some of them may be different from one another.

The first polynucleotide and the second polynucleotide may be two instances of the same polynucleotide. The first polynucleotide may be identical to the second polynucleotide. This allows proof reading. If the method concerns investigating three or more polynucleotides, they
20 may all be three or more instances of the same polynucleotide or some of them may be separate instances of the same polynucleotide.

The first sample and second sample may be different from one another. For instance, the first sample may be derived from a human and the second sample may be derived from a virus. If the first and second samples are different from one another, they may contain or be suspected
25 of containing the same first and second polynucleotides. If the method concerns investigating three or more samples, they may all be different from one another or some of them may be different from one another.

The first sample and the second sample are preferably two instances of the same sample. The first sample is preferably identical to the second sample. This allows proof reading. If the
30 method concerns investigating three or more samples, they may all be three or more instances of the same sample or some of them may be separate instances of the same sample.

Any number of polynucleotides can be investigated. For instance, the method may concern characterising 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100 or more polynucleotides. If three or

more polynucleotides are investigated using the method, the second polynucleotide is also uncoupled from the membrane and the requisite number of steps are added for the third polynucleotide. The same is true for four or more polynucleotides.

The method involves uncoupling the first polynucleotide from the membrane. The method may involve uncoupling the second polynucleotide from the membrane if three or more polynucleotides are being investigated.

The first polynucleotide can be uncoupled from the membrane using any known method. The first polynucleotide is preferably not uncoupled from the membrane in step (f) using the transmembrane pore. The first polynucleotide is preferably not uncoupled from the membrane using a voltage or an applied potential.

Step (f) preferably comprises uncoupling the first polynucleotide from the membrane by removing the one or more anchors from the membrane. If the anchors are removed, the second polynucleotide is coupled to the membrane using other (or separate) anchors. The anchors used to couple the second polynucleotide may be the same type of anchors used to couple the first polynucleotide or different type of anchors.

Step (f) more preferably comprises contacting the one or more anchors with an agent which has a higher affinity for the one or more anchors than the anchors have for the membrane. A variety of protocols for competitive binding or immunoradiometric assays to determine the specific binding capability of molecules are well known in the art (see for example Maddox *et al*, J. Exp. Med. 158, 1211-1226, 1993). The agent removes the anchor(s) from the membrane and thereby uncouples the first polynucleotide. The agent is preferably a sugar. Any sugar which binds to the one or more anchors with a higher affinity than the one or more anchors have for the membrane may be used. The sugar may be a cyclodextrin or derivative thereof as discussed below.

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 (am₇- β CD), 6-monodeoxy-6-monoamino- β -cyclodextrin (am₁- β CD) or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu₇- β CD). Any of the lipids disclosed herein may be used.

If an anchor comprise(s) streptavidin, biotin or desthiobiotin, the agent is preferably biotin, desthiobiotin or streptavidin. Both biotin and desthiobiotin bind to streptavidin with a higher affinity than streptavidin binds to the membrane and *vice versa*. Biotin has a stronger

affinity for streptavidin than desthiobiotin. An anchor comprising streptavidin may therefore be removed from the membrane using biotin or streptavidin and *vice versa*.

If an anchor comprises a protein, the agent is preferably an antibody or fragment thereof which specifically binds to the protein. An antibody specifically binds to a protein if it binds to the protein with preferential or high affinity, but does not bind or binds with only low affinity to other or different proteins. An antibody binds with preferential or high affinity if it binds with a K_d of 1×10^{-6} M or less, more preferably 1×10^{-7} M or less, 5×10^{-8} M or less, more preferably 1×10^{-8} M or less or more preferably 5×10^{-9} M or less. An antibody binds with low affinity if it binds with a K_d of 1×10^{-6} M or more, more preferably 1×10^{-5} M or more, more preferably 1×10^{-4} M or more, more preferably 1×10^{-3} M or more, even more preferably 1×10^{-2} M or more. Any method may be used to detect binding or specific binding. Methods of quantitatively measuring the binding of an antibody to a protein are well known in the art. The antibody may be a monoclonal antibody or a polyclonal antibody. Suitable fragments of antibodies include, but are not limited to, Fv, F(ab') and F(ab')₂ fragments, as well as single chain antibodies. Furthermore, the antibody or fragment thereof may be a chimeric antibody or fragment thereof, a CDR-grafted antibody or fragment thereof or a humanised antibody or fragment thereof.

Step (f) preferably comprises contacting the one or more anchors with an agent which reduces ability of the one or more anchors to couple to the membrane. For instance, the agent could interfere with the structure and/or hydrophobicity of the one or more anchors and thereby reduce their ability to couple to the membrane. If an anchor comprises cholesterol, the agent is preferably cholesterol dehydrogenase. If an anchor comprises a lipid, the agent is preferably a phospholipase. If an anchor comprises a protein, the agent is preferably a proteinase or urea. Other combination of suitable anchors and agents will be clear to a person skilled in the art.

Step (f) preferably comprises uncoupling the first polynucleotide from the membrane by separating the first polynucleotide from the one or more anchors. This can be done in any manner. For instance, the linker could be cut in an anchor comprising a linker. This embodiment is particularly applicable to anchors which involve linkage via hybridisation. Such anchors are discussed above.

Step (f) more preferably comprises uncoupling the first polynucleotide from the membrane by contacting the first polynucleotide and the one or more anchors with an agent which competes with the first polynucleotide for binding to one or more anchors. Methods for determining and measuring competitive binding are known in the art. The agent is preferably a polynucleotide which competes with the first polynucleotide for hybridisation to the one or more

anchors. For instance, if the first polynucleotide is coupled to the membrane using one or more anchors which involve hybridisation, the polynucleotide can be uncoupled by contacting the one or more anchors with a polynucleotide which also hybridises to the site of hybridisation. The polynucleotide agent is typically added at a concentration that is higher than the concentration of the first polynucleotide and one or more anchors. Alternatively, the polynucleotide agent may hybridise more strongly to the one or more anchors than the first polynucleotide.

Step (f) more preferably comprises (i) contacting the first polynucleotide and the one or more anchors with urea, tris(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT), streptavidin or biotin, UV light, an enzyme or a binding agent; (ii) heating the first polynucleotide and the one or more anchors; or (iii) altering the pH. Urea, tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT) are capable of disrupting anchors and separating the first polynucleotide from the membrane. If an anchor comprises a streptavidin-biotin link, then a streptavidin agent will compete for binding to the biotin. If an anchor comprises a streptavidin-desthiobiotin link, then a biotin agent will compete for binding to the streptavidin. UV light can be used to breakdown photolabile groups. Enzymes and binding agents can be used to cut, breakdown or unravel the anchor. Preferred enzymes include, but are not limited to, an exonuclease, an endonuclease or a helicase. Preferred binding agents include, but are not limited to, an enzyme, an antibody or a fragment thereof or a single-stranded binding protein (SSB). Any of the enzymes discussed below or antibodies discussed above may be used. Heat and pH can be used to disrupt hybridisation and other linkages.

If the first polynucleotide is uncoupled from the membrane by separating the first polynucleotide from the one or more anchors, the one or more anchors will remain in the membrane. Step (g) preferably comprises coupling the second polynucleotide to the membrane using the one or more anchors that was separated from the first polynucleotide. For instance, the second polynucleotide may also be provided with one or more polynucleotides which hybridise(s) to the one or more anchors that remain in the membrane. Alternatively, step (g) preferably comprises coupling the second polynucleotide to the membrane using one or more separate anchors from the ones separated from the first polynucleotide (i.e. one or more other anchors). The one or more separate anchors may be the same type of anchors used to couple the first polynucleotide to the membrane or may be different types of anchors. Step (g) preferably comprises coupling the second polynucleotide to the membrane using one or more different anchors from the one or more anchors separated from the first polynucleotide.

In a preferred embodiment, steps (f) and (g) comprise uncoupling the first polynucleotide

from the membrane by contacting the membrane with the second polynucleotide such that the second polynucleotide competes with the first polynucleotide for binding to the one or more anchors and replaces the first polynucleotide. For instance, if the first polynucleotide is coupled to the membrane using one or more anchors which involve hybridisation, the first polynucleotide
5 can be uncoupled by contacting the anchors with the second polynucleotide attached to polynucleotides which also hybridise to the sites of hybridisation in the one or more anchors. The second polynucleotide is typically added at a concentration that is higher than the concentration of the first polynucleotide and the one or more anchors. Alternatively, the second polynucleotide may hybridise more strongly to the one or more anchors than the first
10 polynucleotide.

Removal or Washing

Although the first polynucleotide is uncoupled from the membrane in step (f), it is not necessarily removed or washed away. If the second polynucleotide can be easily distinguished
15 from the first polynucleotide, there is no need to remove the first polynucleotide.

Between steps (f) and (g), the method preferably further comprises removing at least some of the first sample from the membrane. At least 10% of the first sample may be removed, such as at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% of the first sample may be removed.

20 The method more preferably further comprises removing all of the first sample from the membrane. This can be done in any way. For instance, the membrane can be washed with a buffer after the first polynucleotide has been uncoupled. Suitable buffers are discussed below.

Modified Polynucleotides

25 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.
30 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 UK Application No. 1403096.9. Any of the polymerases discussed above may be used. The polymerase is preferably Klenow or 90 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
5 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
10 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
15 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.

20

Analyte delivery

The target analyte is preferably attached to a microparticle which delivers the analyte towards the membrane. This type of delivery is disclosed in UK Application No. 1418469.1. Any type of microparticle and attachment method may be used.

25

Other Characterisation Method

In another embodiment, a polynucleotide is characterised by detecting labelled species that are added to the target polynucleotide by a polymerase and then released. The polymerase uses the polynucleotide as a template. Each labelled species is specific for each nucleotide. The
30 polynucleotide is contacted with a pore as disclosed herein, such as a double pore, a polymerase and labelled nucleotides such that phosphate labelled species are sequentially added to the polynucleotide by the polymerase, wherein the phosphate species contain a label specific for each nucleotide. The labelled species may be detected using the pore before they are released

from the nucleotides (i.e. as they are added to the target polynucleotide) or after they are released from the nucleotides.

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 EP-A-2682460. Any of the embodiments discussed above equally apply to this method.

Examples of labelled species include, but are not limited to, polymers, polyethylene glycols, sugars, cyclodextrins, fluorophores, drugs, metabolites, peptides. A non-limiting example of such tags can be found in the work of Kumar et al. Sci Rep. 2012;2:684. Epub 2012 Sep 21.

Methods of Forming Sensors

Also provided is a method of forming a sensor for characterising a target polynucleotide. The method comprises forming a complex between a pore as disclosed herein and a polynucleotide binding protein, such as a helicase or an exonuclease. The complex may be formed by contacting the pore and the protein in the presence of the target polynucleotide and then applying a potential across the pore. The applied potential may be a chemical potential or a voltage potential as described above. Alternatively, the complex may be formed by covalently attaching the pore to the protein. Methods for covalent attachment are known in the art and disclosed, for example, in WO 2010/004265 and WO 2010/086603. The complex is a sensor for characterising the target polynucleotide. The method preferably comprises forming a complex between a pore as disclosed herein and a helicase. Any of the embodiments discussed above equally apply to this method.

Also provided is a sensor for characterising a target polynucleotide. The sensor comprises a complex between a pore as disclosed herein and a polynucleotide binding protein. Any of the embodiments discussed above equally apply to the sensor.

Kits

Also provided is a kit for characterising a target polynucleotide. The kit comprises a double pore and/or a pore as disclosed herein 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. Any of the polynucleotide binding proteins discussed above may be used.

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

5 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 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
10 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 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
15 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 as disclosed herein or
20 details regarding for which organism the method may be used.

Apparatus

Also provided is an apparatus for characterising a target analyte, such as a target polynucleotide. The apparatus comprises a plurality of the pores as disclosed herein and a
25 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 for characterising target analytes, may comprise or an array of pores as disclosed herein, for example an array of double pores, in a plurality of membranes.

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

the apparatus of the invention. The apparatus may further comprise any of the features present in the kit as disclosed herein.

The apparatus is preferably set up to carry out a method as disclosed herein.

The apparatus preferably comprises:

- 5 a sensor device that is capable of supporting the plurality of pores and membranes and being operable to perform analyte 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:

- 10 a sensor device that is capable of supporting the plurality of pores and membranes being operable to perform analyte characterisation using the pores and membranes; and at least one reservoir for holding material for performing the characterisation.

The apparatus more preferably comprises:

- 15 a sensor device that is capable of supporting the membrane and plurality of pores and membranes and being operable to perform analyte 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
20 configured to supply the samples selectively from one or more containers to the sensor device.

The apparatus may be any of those described in WO 2009/077734, WO 2010/122293, WO 2011/067559 or WO 00/28312.

Example 1: Double Pore Production

- 25 DNA (SEQ ID NO: 25) encoding the polypeptide Pro-CP1-Eco-(*Mutant-StrepII(C)*) (SEQ ID NO: 26) was cloned into a pT7 vector containing ampicillin resistance gene. Concentration of DNA solution was adjusted to 400µg/uL. 1µl of DNA was used to transform the cell line ONT001 which is Lemo BL21 DE3 cell line in which the gene coding for CsgG protein is replaced with DNA responsible for kanamycin resistance. Cells were then plated out
30 on LB agar containing ampicillin (0.1mg/ml) and kanamycin (0.03mg/ml) and incubated for approximately 16 hours at 37°C.

Bacterial colonies grown on LB plates containing ampicillin and kanamycin can be assumed to have incorporated the CP1 plasmid with no endogenous production. One such colony

was used to inoculate a starter culture of LB media (100mL) containing both carbenicillin (0.1mg/ml) and kanamycin (0.03mg/ml). The starter culture was grown at 37°C with agitation, until OD600 was reached to 1.0 – 1.2. The starter culture was used to inoculate a fresh 500ml culture to and OD600 of 0.1. LB media containing the following additives - carbenicillin (0.1mg/ml), kanamycin (0.03mg/ml), 500µM Rhamnose, 15mM MgSO₄ and 3mM ATP. The culture was grown at 37°C with agitation until stationary phase was entered and held for a further hour – stationary phase ascertained by plateau of measured OD600. Temperature of the culture was then adjusted to 18°C and glucose was added to a final concentration of 0.2%. Once culture was stable at 18 °C induction was initiated by the addition of lactose to a final concentration of 1%. Induction was carried out for approximately 18 hours with agitation at 18°C.

Following induction, the culture was pelleted by centrifugation at 6,000g for 30 minutes. The pellet was resuspended in 50mM Tris, 300mM NaCl, containing Protease Inhibitors (Merck Millipore 539138), Benzonase Nuclease (Sigma E1014), 1X Bugbuster (Merck Millipore 70921) and 0.1% Brij 58 pH8.0 (approximately 10ml of buffer per gram of pellet). The suspension was mixed well until it is fully homogeneous, sample was then transferred to roller mixer at 4°C for approximately 5 hours. Lysate was pelleted by centrifugation at 20,000g for 45 minutes and the supernatant was filtered through 0.22µm PES syringe filter. Supernatant which contains CP1 was taken forward for purification by column chromatography.

Sample was applied to a 5ml Strep Trap column (GE Healthcare). Column was washed with 25mM Tris, 150mM NaCl, 2mM EDTA, 0.1% Brij 58 pH8 until a stable baseline of 10 column volumes was maintained. Column was then washed with 25mM Tris, 2M NaCl, 2mM EDTA, 0.1% Brij 58 pH8 before being returned to 150mM buffer. Elution was carried out with 10mM desthiobiotin. Elution peak was pooled and carried forward for ion exchange purification on a 1ml Q HP column (GE Healthcare) using 25mM Tris, 150mM NaCl, 2mM EDTA, 0.1% Brij 58 pH8 as the binding buffer and 25mM Tris, 500mM NaCl, 2mM EDTA, 0.1% Brij 58 pH8 as the elution buffer. Flowthrough peak was observed to contain both dimer and monomer protein, elution peak at approx. 400ms/sec was observed to contain monomeric pore. Flowthrough peak was concentrated via vivaspin column (100kd MWCO) and carried forward for size exclusion chromatography on 24ml S200 increase column (GE Healthcare) with the buffer 25mM Tris, 150mM NaCl, 2mM EDTA, 0.1% Brij 58, 0.1% SDS pH8. Dimeric (double) pore eluted at 9ml while the monomeric pore eluted at 10.5ml.

Example 2: Elongation of the constriction (read head) in a CsgG pore

Materials and methods

Electrical measurements were acquired from a variety of single CsgG 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, buffer (1mL 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide(II), 150mM Potassium Ferricyanide(III), pH 8.0) was flowed through the system to remove any excess CsgG nanopores. A platform QC was then run to determine the number of channels containing single nanopores. To do this 1mL of buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide(II), 150mM Potassium Ferricyanide(III), pH 8.0) containing 240nM TBA was flown through the inlet. Whilst platform QC was in progress, RBF1 (940mM KCL, 50mM HEPES, 20mM MgCl₂ and 22mM ATP) was diluted to 1x using nuclease free water. 800µl of 1x RBF1 was then flushed through the inlet. After 10 minutes the SpotON valve was opened and 200µL of 1x RBF1 was flown through the inlet. The system was now ready for sequencing.

Meanwhile, DNA sample was prepared for sequencing using the following method: 1µg of DNA analyte was incubated with the 40nM of adapter mix (containing a T4 Dda helicase enzyme prebound to the adapter) and blunt TA ligase for 10 minutes. The ligation mixture was then purified to remove unligated free adapter using Spri purification. The final ligated mixture was eluted in 25 µL elution buffer containing 40 mM CAPS at pH10, 40 mM KCl and 400 nM cholesterol tether. For each chip, 12µl of DNA-adapter ligated mix was mixed with 37.5 µL RBF1 and 25.5µL water (final volume of 75µL) and added to Flow Cell via the SpotON port for sequencing. The experiment was then run for 6 hours at -180mV.

Results

A list of squiggles was assembled for each pore mutant, and using these squiggles a Hidden Markov Model (HMM) model was made that describes the signal characteristics of all sequence contexts observed in the DNA analyte. Using these models, we were able to gain insight into the length and shape of the read head. The left-hand plots in Figures 6 to 9 show the current levels versus read head position grouped by base. For example, at read head position 3, the median current of all sequence contexts with either an A or C or G or T at the 3rd position are averaged. The right-hand plots in Figures 6 to 9 show discrimination versus read head position, where discrimination is defined as the variance in the current attributable to variation in the base at that position, averaged over all sequence contexts.

Figure 6 shows the discrimination profile for a baseline pore comprising monomers having the sequence shown in SEQ ID NO: 2 in which the following substitutions have been

made: Y51A; F56Q; K94Q; R97W; and R192D, and in which V105 to I107 have been deleted. This pore has one dominant position in the read head at the 3rd position when 5 bases are present in the pore read head at a given time. Figures 7 to 9 show the discrimination profile for pores comprising the above baseline pore mutations in SEQ ID NO: 2 and one of the following
5 additional substitutions: N55V, A51Q and Q56V. These mutants all demonstrate an elongation of the read head compared to the baseline pore, with two positions in the read head now contributing to the signal.

CLAIMS

1. A method of characterising a polynucleotide using a transmembrane pore, wherein the pore is a double pore comprising a first CsgG pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof.
2. A method according to claim 1, wherein the polynucleotide comprises a homopolymeric region.
3. A method according to claim 1 or 2, wherein the first CsgG pore, or homologue thereof, comprises from six to ten monomers and/or the second CsgG pore, or homologue thereof comprises from six to ten monomers.
4. A method according to any one of the preceding claims, wherein the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a homooligomer.
5. A method according to any one of the preceding claims, wherein the first CsgG pore, or homologue thereof, and the second CsgG pore, or homologue thereof, are identical.
6. A double pore comprising a first CsgG pore, or a homologue thereof, and a second CsgG pore, or a homologue thereof, wherein:
 - (i) the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a homooligomer, and the first CsgG pore, or homologue thereof, comprises monomers that have a different amino acid sequence from the monomers of which the second CsgG pore, or homologue thereof, is comprised;
 - (ii) the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a homooligomer, and the first CsgG pore, or homologue thereof, and/or the second CsgG pore, or homologue thereof is not a wild-type pore;
 - (iii) the first CsgG pore, or homologue thereof, is a heterooligomer and the second CsgG pore, or homologue thereof, is a homooligomer;
 - (iv) the first CsgG pore, or homologue thereof, is a homooligomer and the second CsgG pore, or homologue thereof, is a heterooligomer;
 - (v) the first CsgG pore, or homologue thereof, is a heterooligomer and the second CsgG pore, or homologue thereof, is a heterooligomer.
7. A double pore according to claim 6 (ii) or (v), wherein the first CsgG pore, or homologue thereof, and the second CsgG pore, or homologue thereof are identical.

8. A double pore according to claim 6, wherein the first CsgG pore, or homologue thereof, and the second CsgG pore, or homologue thereof, are different.
9. A double pore according to any one of claims 6 to 8, wherein the first CsgG pore, or homologue thereof, and the second CsgG pore, or homologue thereof, are in opposite orientations.
10. A double pore according to claim 9, wherein the tail region of the first CsgG pore, or homologue thereof, is adjacent to the tail region of the second CsgG pore, or homologue thereof.
11. A double pore according to any one of claims 6 to 10, wherein the first CsgG pore, or homologue thereof, is attached to the second CsgG pore, or homologue thereof, by hydrophobic interactions and/or by one or more disulphide bond.
12. A double pore according to claim 11, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the second CsgG pore, or homologue thereof, comprises at least one cysteine residue at the interface between the first and second pores, which cysteine residue is not present in the wild type CsgG monomer, or the wild type CsgG homologue monomer.
13. A double pore according to claim 12, wherein the cysteine residue at the interface between the first and second pores is at a position corresponding to R97, I107, R110, Q100, E101, N102 and/or L113 of SEQ ID NO: 2.
14. A double pore according to any one of claims 11 to 13, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the second CsgG pore, or homologue thereof, comprises at least one residue at the interface between the first and second pores, which residue is more hydrophobic than the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer.
15. A double pore according to claim 14, wherein the at least one residue at the interface between the first and second pores is at a position corresponding to R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2.
16. A double pore according to claim 14 or 15, wherein the at least one residue at the interface between the first and second pores is:
- (i) I, L, V, M, F, W, Y, where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is R, Q, N or E;

- (ii) L, V, M, F, W, Y, where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is I; and/or
- (iii) C, I, V, M, F, W, Y, where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is L.

5

17. A double pore according to any one of claims 6 to 16, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the second CsgG pore, or homologue thereof, comprises at least one residue at the interface between the first and second pores, which residue is bulkier than the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer.

10

18. A double pore according to claim 17, wherein the at least one residue at the interface between the first and second pores is at a position corresponding to A98, A99, T104, V105, L113, Q114 or S115 of SEQ ID NO: 2.

15

19. A double pore according to claim 17 or 18, wherein the at least one residue at the interface between the first and second pores is:

- (i) I, L, V, M, F, W, Y, N, Q, S or T where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is A;
- (ii) L, M, F, W, Y, N, Q, R, D or E where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is T;
- (iii) I, L, M, F, W, Y, N, Q where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is V;
- (iv) M, F, W, Y, N, Q, R, D or E where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is L;
- (v) F, W, Y where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is Q; and/or
- (vi) M, F, W, Y, N, Q, E or R where the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer is S.

20

25

30

20. A double pore according to any one of the preceding claims, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the

second CsgG pore, or homologue thereof, comprises at least one residue in the barrel region of the pore, which residue has less negative charge than the residue present at the corresponding position in the wild type CsgG monomer, or the wild type CsgG homologue monomer.

21. A double pore according to claim 20, wherein the at least one residue in the barrel region of the pore is at a position corresponding to D149, E185, D195, E210 and/or E203 of SEQ ID NO: 2.
22. A double pore according to claim 20 or 21, wherein the at least one residue in the barrel region of the pore is N, Q, R or K.
23. A double pore according to any one of the preceding claims, wherein at least one monomer in the first CsgG pore, or homologue thereof, comprises at least one residue in the constriction of the barrel region of the first pore, which residue decreases, maintains or increases the length of the constriction compared to the wild type CsgG pore, or the wild type CsgG homologue pore, and/or at least one monomer in the second CsgG pore, or homologue thereof, comprises at least one residue in the constriction of the barrel region of the second pore, which residue decreases, maintains or increases the length of the constriction compared to the wild type CsgG pore, or the wild type CsgG homologue pore.
24. A double pore according to claim 23, wherein the at least one residue in the constriction of the barrel region of the pore is a residue that is additional to the residues present in the constriction of the wild type CsgG pore, or the wild type CsgG homologue pore.
25. A double pore according to claim 23 or 24, wherein the at least one residue is between the residues corresponding to K49 and P50, P50 and Y51, Y51 and P52, P52 and A53, A53 and S54, S54 and N55 and/or N55 and F56 of SEQ ID NO: 2.
26. A double pore according to claim 23, wherein the at least one residue increases the length of the loop between the residues corresponding to Y51 and N55 of SEQ ID NO: 2.
27. A double pore according to any one of claims 23 to 26, wherein the at least one residue is:
- (i) A, S, G or T;
 - (ii) P; and/or
 - (iii) S, T, N, Q, M, F, W, Y, V and/or I.
28. A double pore according to claim 27, wherein the at least one residue is S, G, SG, SGG, SGS, GS, GSS and/or GSG.

29. A double pore according to claim 23, wherein the at least one residue in the constriction of the barrel region of the pore is at a position corresponding to Y51, N55, Y51, P52 and/or A53 of SEQ ID NO: 2.
30. A double pore according to claim 29, wherein the at least one residue is:
- 5 (i) Q or V at a position corresponding to F56 of SEQ ID NO: 2;
- (ii) A or Q at a position corresponding to Y51 of SEQ ID NO: 2; and/or
- (iii) V at a position corresponding to N55 of SEQ ID NO: 2.
31. A double pore according to any one of the preceding claims, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the
- 10 second CsgG pore, or homologue thereof, comprises two or more of the mutations defined in any one of claims 12 to 30.
32. A method according to any one of claims 1 to 5, wherein the double pore is a pore as defined in any one of claims 6 to 31.
33. A CsgG monomer, or a monomer of a CsgG homologue, comprising:
- 15 (i) a cysteine residue at a position corresponding to R97, I107, R110, Q100, E101, N102 and or L113 of SEQ ID NO: 2;
- (ii) a residue at a position corresponding to any one or more of R97, Q100, I107, R110, E101, N102 and L113 of SEQ ID NO: 2, which residue is more hydrophobic than the residue present at the corresponding position of SEQ ID
- 20 NO: 2, or in the amino acid sequence of a wild type CsgG homologue, wherein the residue at the position corresponding to R97 and/or I107 is M, the residue at the position corresponding to R110 is I, L, V, M, W or Y, and/or the residue at the position corresponding to E101 or N102 is V or M;
- (iii) a residue at a position corresponding to any one or more of A98, A99, T104,
- 25 V105, L113, Q114 and S115 of SEQ ID NO: 2 which is bulkier than the residue present at the corresponding position of SEQ ID NO: 2, or in the amino acid sequence of a wild type CsgG homologue, wherein the residue at the position corresponding to T104 is L, M, F, W, Y, N, Q, D or E, the residue at the position corresponding to L113 is M, F, W, Y, N, G, D or E and/or the residue at the
- 30 position corresponding to S115 is M, F, W, Y, N, Q or E; and/or
- (iv) a residue in the barrel region of the pore at a position corresponding to any one or more of D149, E185, D195, E210 and E203 which has less negative charge than the residue present at the corresponding position in the wild type CsgG monomer,

or the wild type CsgG homologue monomer, wherein the residue at the position corresponding to D149, E185, D195 and/or E203 is K.

34. A monomer according to claim 33, wherein:
- (i) the residue at a position corresponding to Q100 is I, L, V, M, F, W or Y;
 - 5 (ii) the residue at a position corresponding to L113 is I, V, M, F, W or Y;
 - (iii) the residue at a position corresponding to A98 or A99, is I, L, V, M, F, W, Y, N, Q, S or T;
 - (iv) the residue at a position corresponding to V105 is I, L, M, F, W, Y, N or Q;
 - (v) the residue at a position corresponding to Q114 is F, W or Y; and/or
 - 10 (vi) the residue at a position corresponding to E210 is N, Q, R or K.
35. A CsgG monomer, or a monomer of a CsgG homologue, wherein at least one monomer in the first CsgG pore, or homologue thereof, and/or at least one monomer in the second CsgG pore, or homologue thereof, comprises at least one residue in the constriction of the barrel region of the pore, which residue increases the length of the constriction compared to the wild type
- 15 CsgG pore, or the wild type CsgG homologue pore.
36. A monomer according to claim 35, wherein the at least one residue is additional to the residues present in the constriction of the wild type CsgG pore, or the wild type CsgG homologue pore.
37. A monomer according to claim 36, wherein the at least one residue is between the
- 20 residues corresponding to K49 and P50, P50 and Y51, Y51 and P52, P52 and A53, A53 and S54, S54 and N55 and/or N55 and F56 of SEQ ID NO: 2.
38. A monomer according to claim 36 or 37, wherein residue increases the length of the loop between the residues corresponding to Y51 and N55 of SEQ ID NO: 2.
39. A monomer according to any one of claims 35 to 38, wherein the at least one residue is:
- 25 (i) A, S, G or T;
- (ii) P; and/or
- (iii) S, T, N, Q, M, F, W, Y, V and/or I.
40. A monomer according to claim 39, wherein the at least one residue is S, G, SG, SGG, SGS, GS, GSS and/or GSG.
- 30 41. A monomer according to claim 35, wherein the at least one residue in the constriction of the barrel region of the pore is at a position corresponding to N55, P52 and/or A53 of SEQ ID NO: 2, wherein the residue at a position corresponding to N55 is V.

42. A construct comprising two or more covalently attached CsgG monomers, wherein at least one of the monomers is a monomer according to any one of claims 33 to 41.
43. A polynucleotide which encodes a monomer according to any one of claims 33 to 41 or a construct according to claim 42.
- 5 44. A pore comprising at least one monomer according to any one of claims 35 to 41 or a construct according to claim 42.
45. A pore according to claim 44, which is homooligomeric or heterooligomeric.
46. A pore according to claim 44 or 45, which comprises nine monomers.
47. A method for determining the presence, absence or one or more characteristics of a target
10 analyte, comprising:
- (a) contacting the target analyte with a double pore according to any one of claims 6 to 31 or a pore according to any one of claims 44 to 47 such that the target analyte moves with respect to the pore; and
 - (b) taking one or more measurements as the analyte moves with respect to the pore
15 and thereby determining the presence, absence or one or more characteristics of the analyte.
48. A method according to claim 47, wherein the analyte is a polynucleotide.
49. A method according to any one of claims 1 to 5 and claim 48, comprising determining one or more characteristics selected from (i) the length of the polynucleotide, (ii) the identity of
20 the polynucleotide, (iii) the sequence of the polynucleotide, (iv) the secondary structure of the polynucleotide and (v) whether or not the polynucleotide is modified.
50. A method according to any one of claim 1 to 5 and 47 to 49, wherein one or more characteristics of the polynucleotide are measured by electrical measurement and/or optical measurement.
- 25 51. Use of a double pore according to any one of claims 6 to 31 or a pore according to any one of claims 44 to 47 to determine the presence, absence or one or more characteristics of a target analyte.
52. A kit for characterising a target analyte comprising (a) a double pore according to any one of claims 6 to 31 or a pore according to any one of claims 44 to 47 and (b) the components
30 of a membrane.
53. An apparatus for characterising target analytes in a sample, comprising an array of double pores according to any one of claims 6 to 31 or an array of pores according to any one of claims 44 to 47 in a plurality of membranes.

54. An apparatus according to claim 53, wherein the apparatus comprises:
- (a) a sensor device that is capable of supporting the plurality of pores and membranes being operable to perform analyte characterisation using the pores and membranes; and
 - 5 (b) at least one port for delivery of the material for performing the characterisation and/or at least one reservoir for holding material for performing the characterisation.
55. An apparatus according to claim 54, wherein the apparatus further comprises:
- 10 (a) a fluidics system configured to controllably supply material from the at least one reservoir to the sensor device; and
 - (b) a plurality of containers for receiving respective samples, the fluidics system being configured to supply the samples selectively from the containers to the sensor device.
56. A method of producing a monomer according to any one of claims 33 to 41 or a construct
15 according to claim 42, comprising expressing a polynucleotide according to claim 43 in a suitable host cell and thereby producing a monomer according to any one of claims 33 to 41 or a construct according to claim 42.

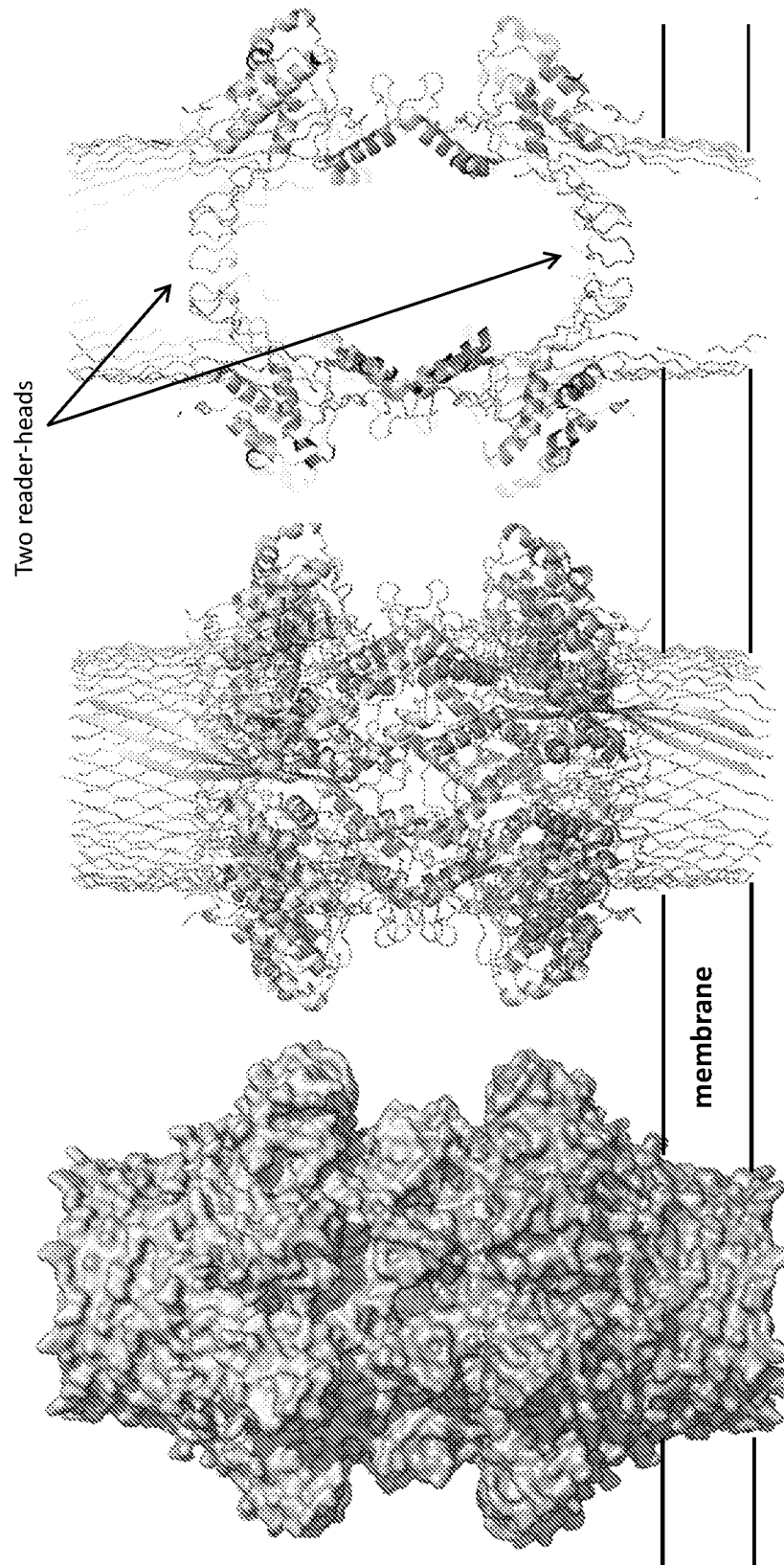


Figure 1

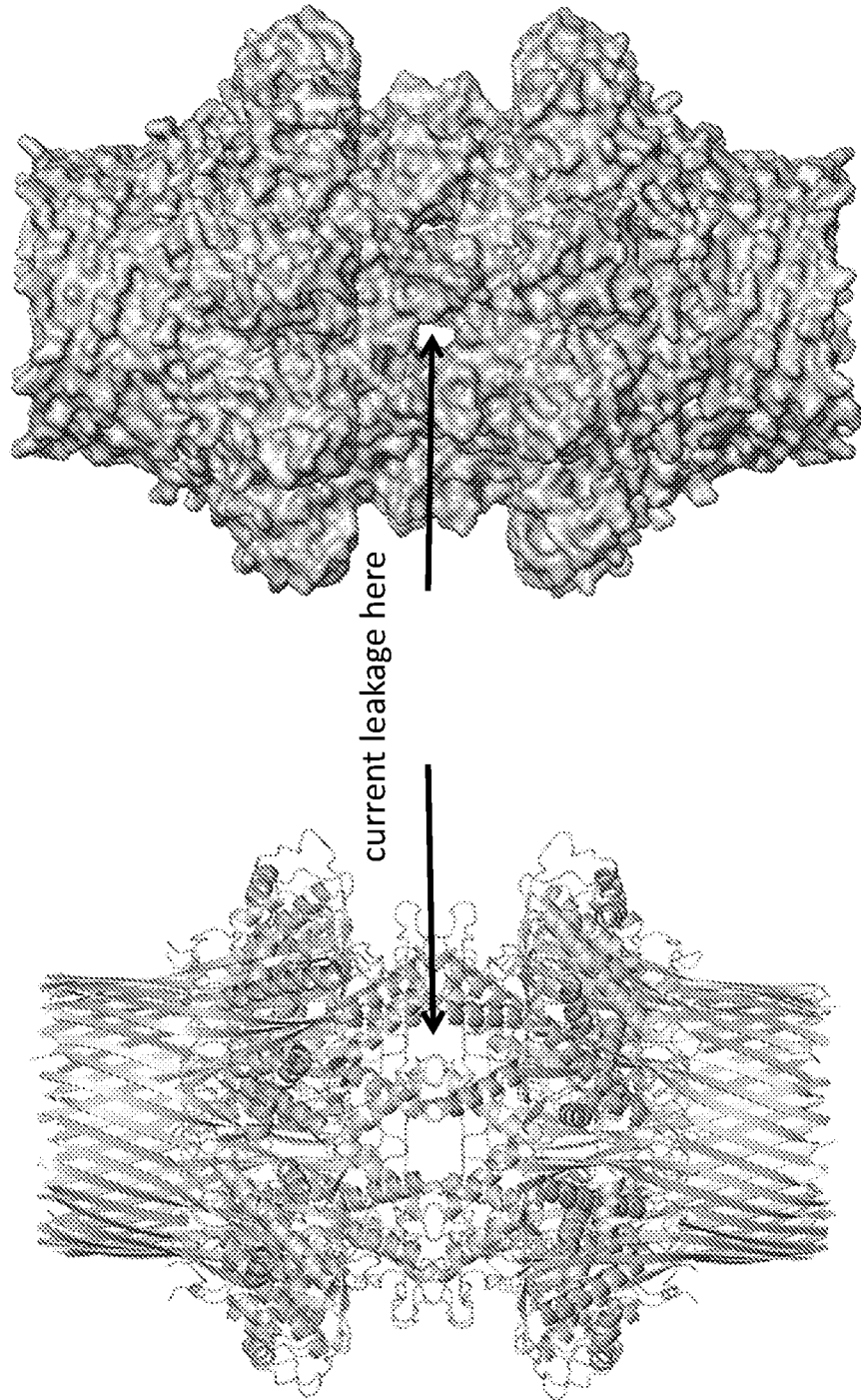
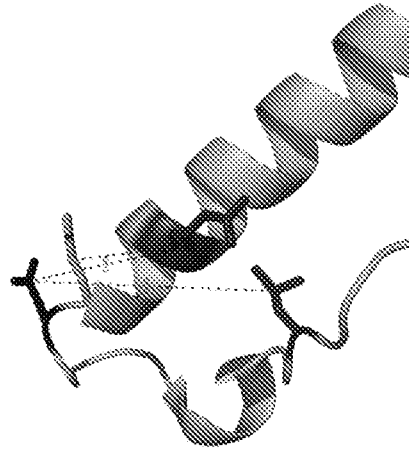


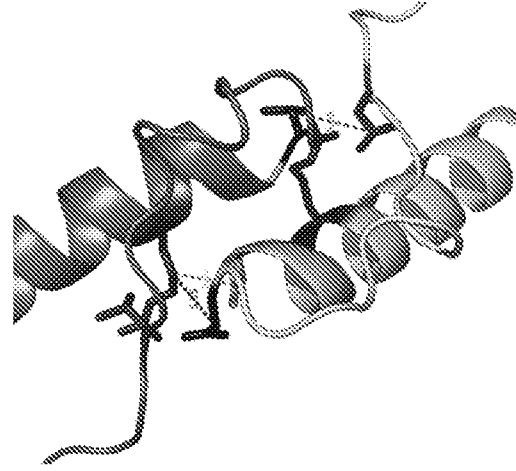
Figure 2

Figure 3

- CP1-(AQ-R97C-StrepII(C))9
- CP1-(AQ-I107C-StrepII(C))9
- CP1-(AQ-R110C-StrepII(C))9
- CP1-(AQ-Q100C-StrepII(C))9
- CP1-(AQ-E101C-StrepII(C))9
- CP1-(AQ-N102C/L113C-StrepII(C))9
- CP1-(AQ-R97C/N102C-StrepII(C))9
- CP1-(AQ-R97C/E101C-StrepII(C))9



Monomer
From the prepore



Dimer
From the x-structure

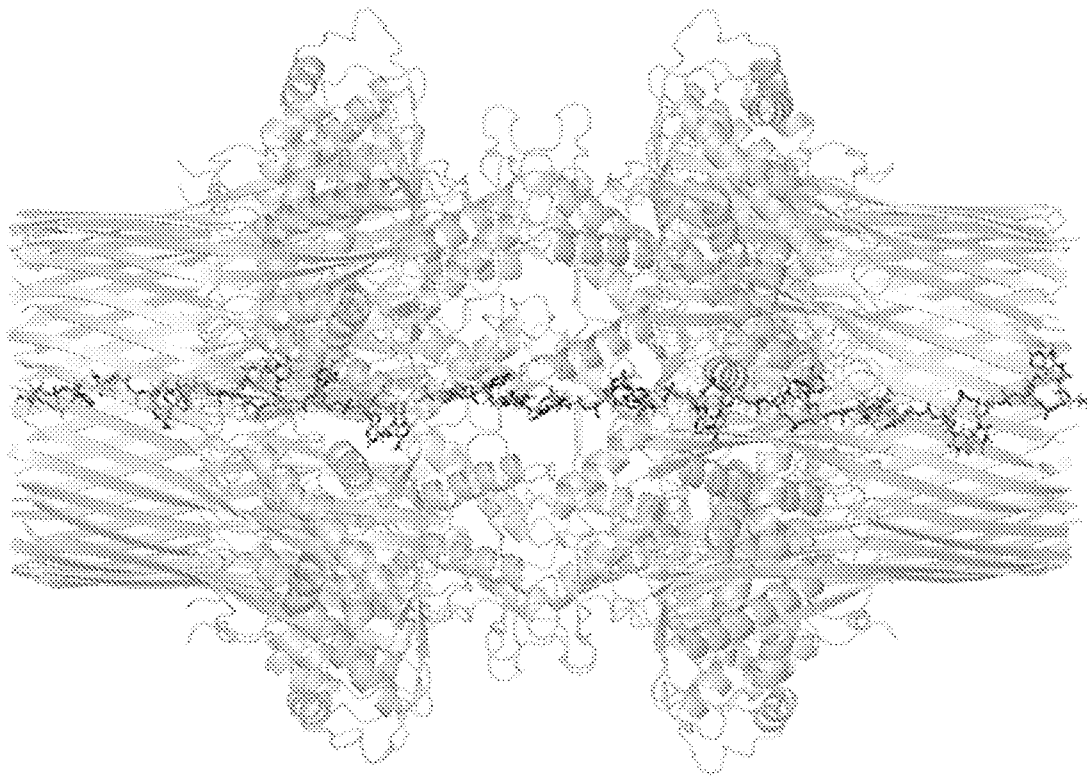
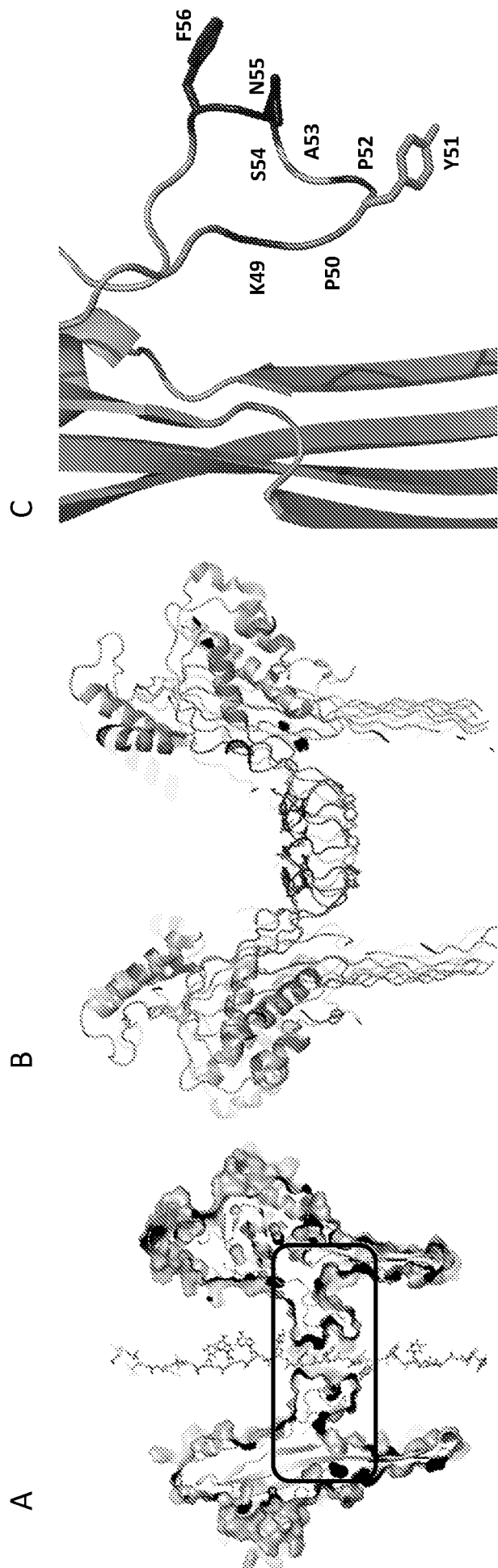


Figure 4

Figure 5



6/17

F56Q/N55/Y51A

Fig. 6A

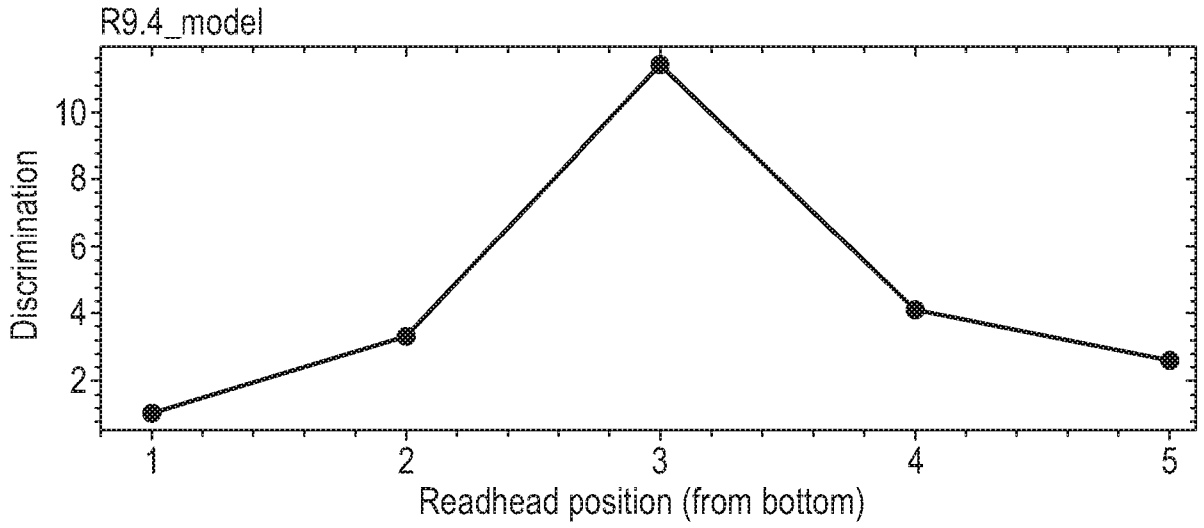


Fig. 6B

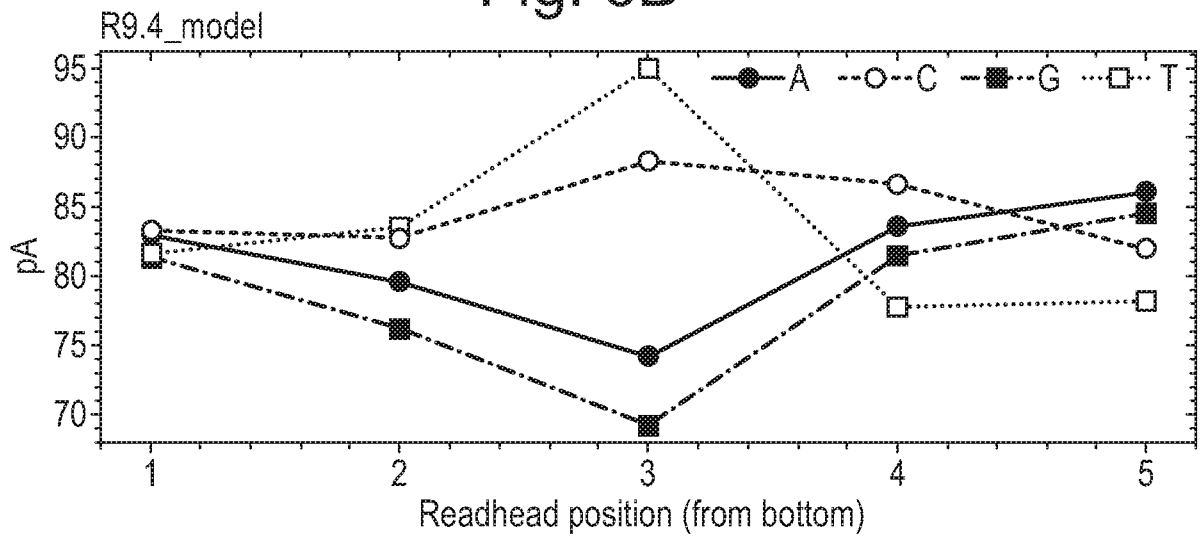
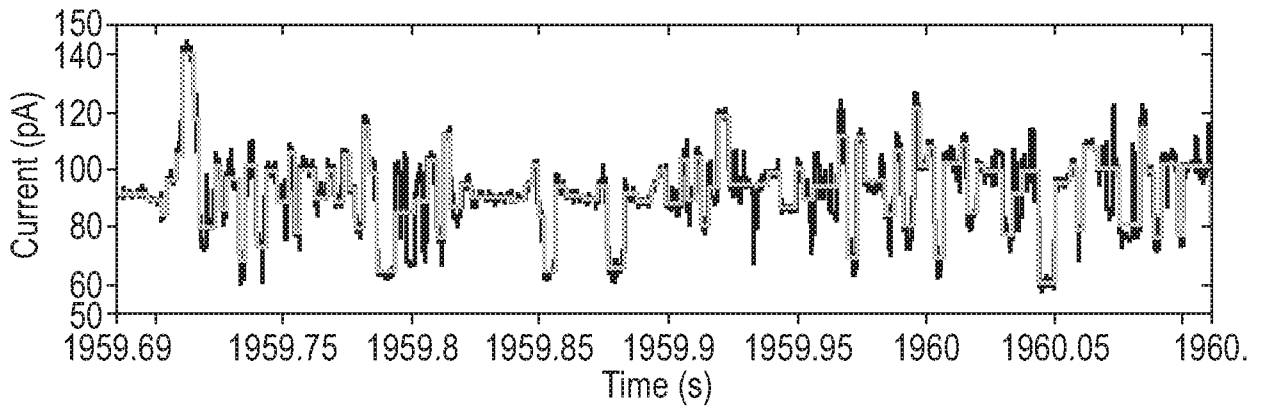


Fig. 6C



7/17

N55V

Fig. 7A

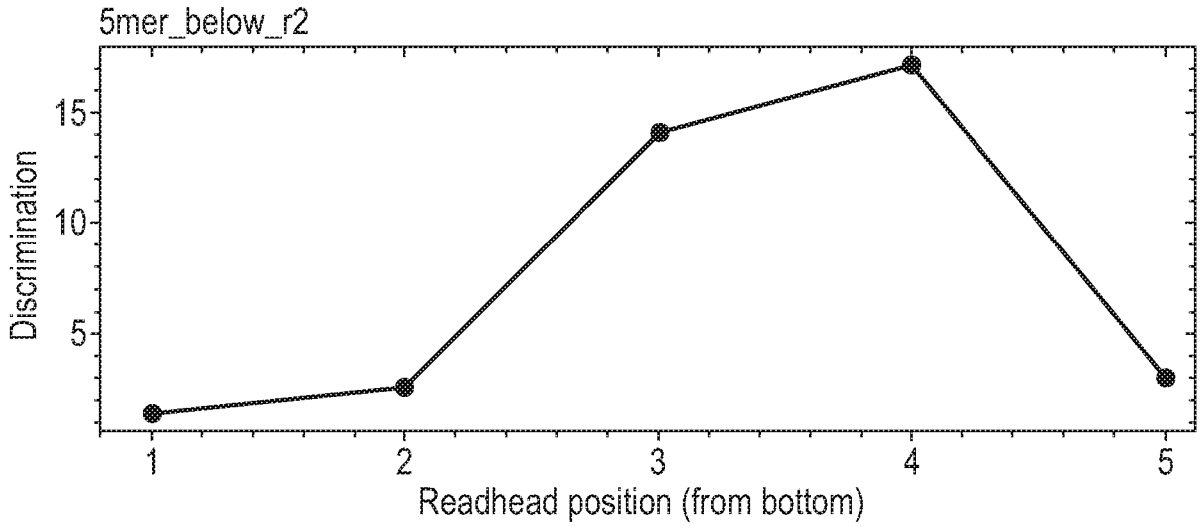


Fig. 7B

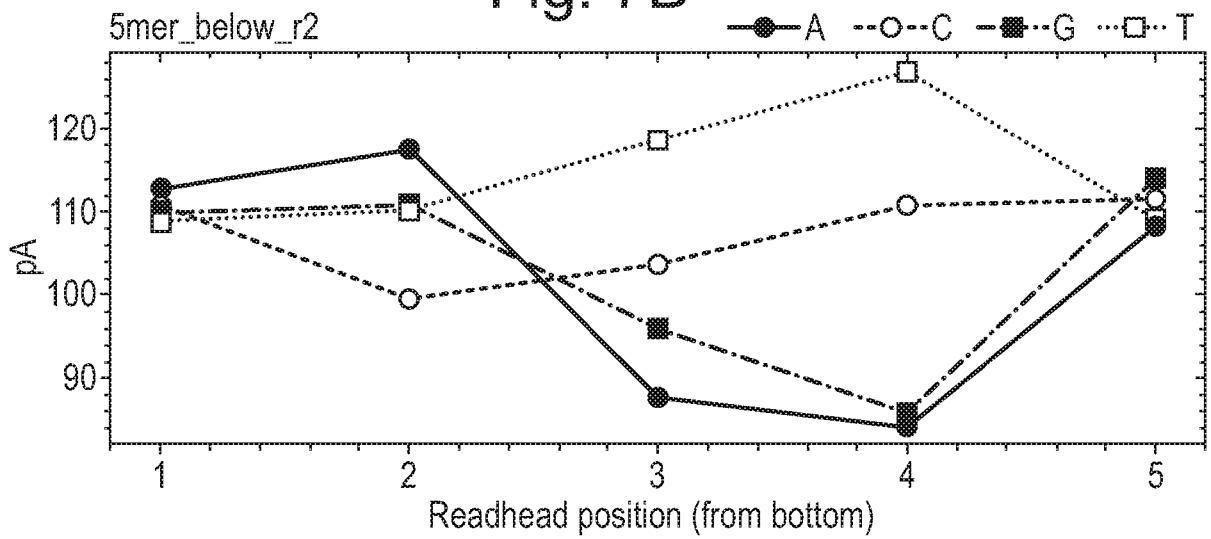
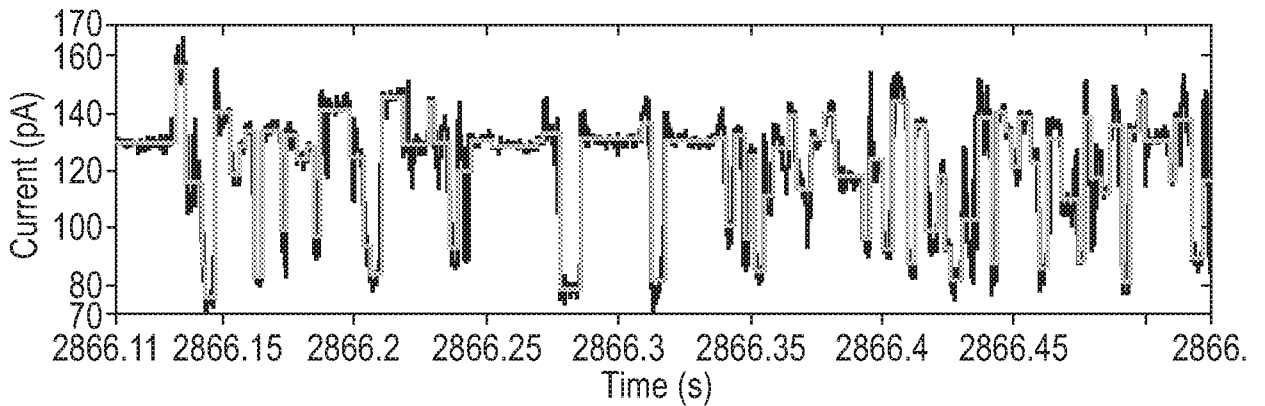


Fig. 7C



A51Q

Fig. 8A

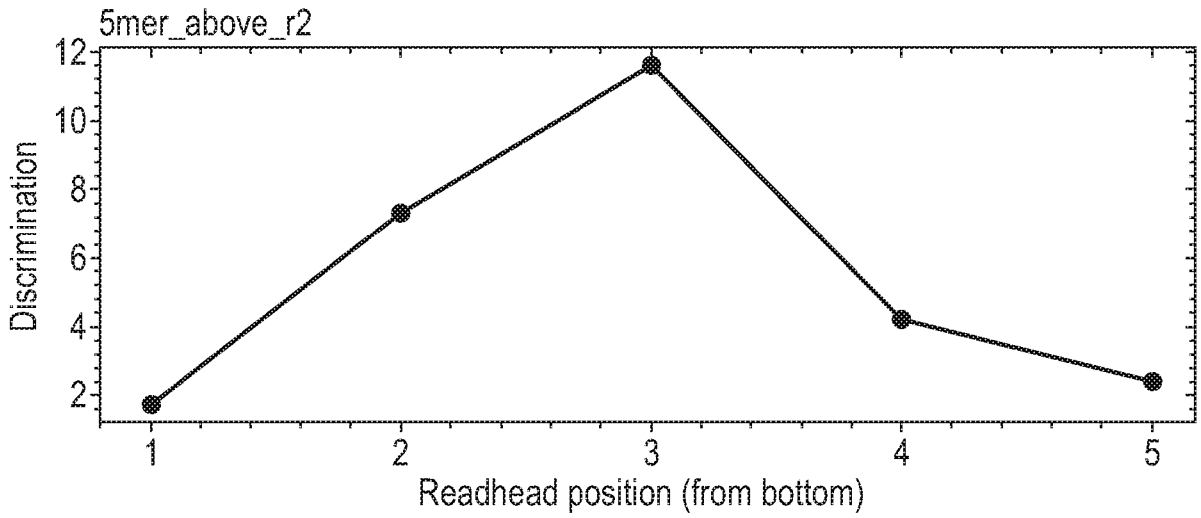


Fig. 8B

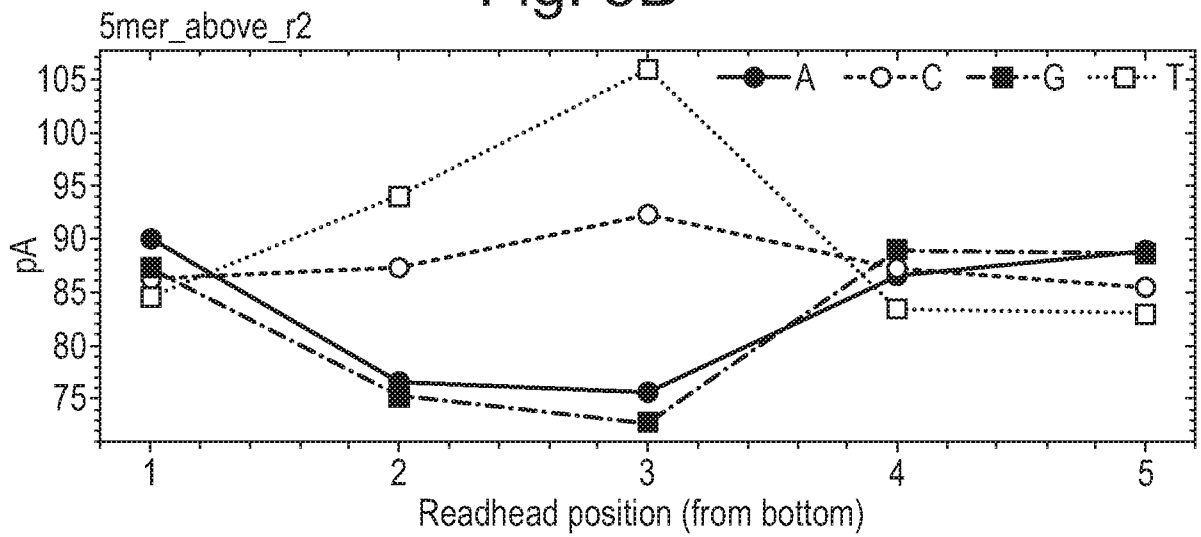
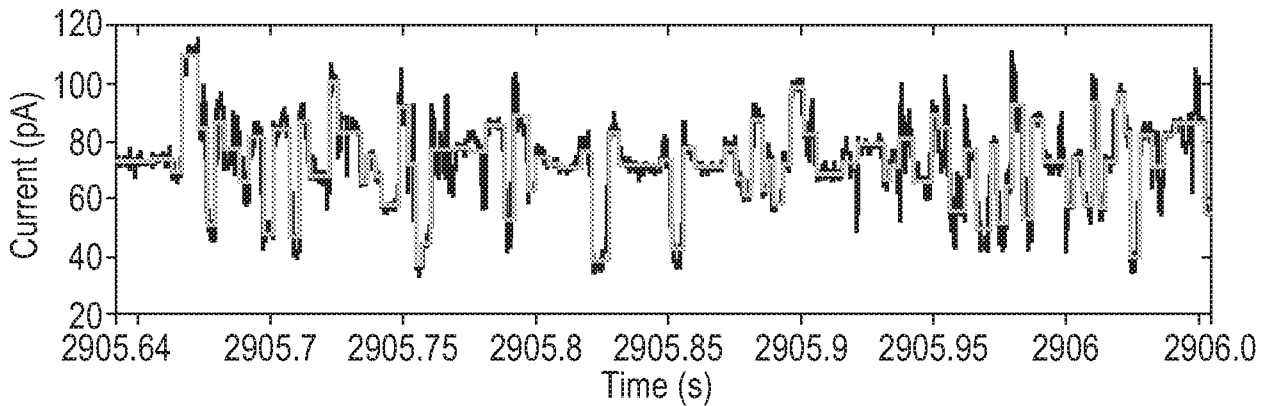


Fig. 8C



9/17

Fig. 9A

Q56V

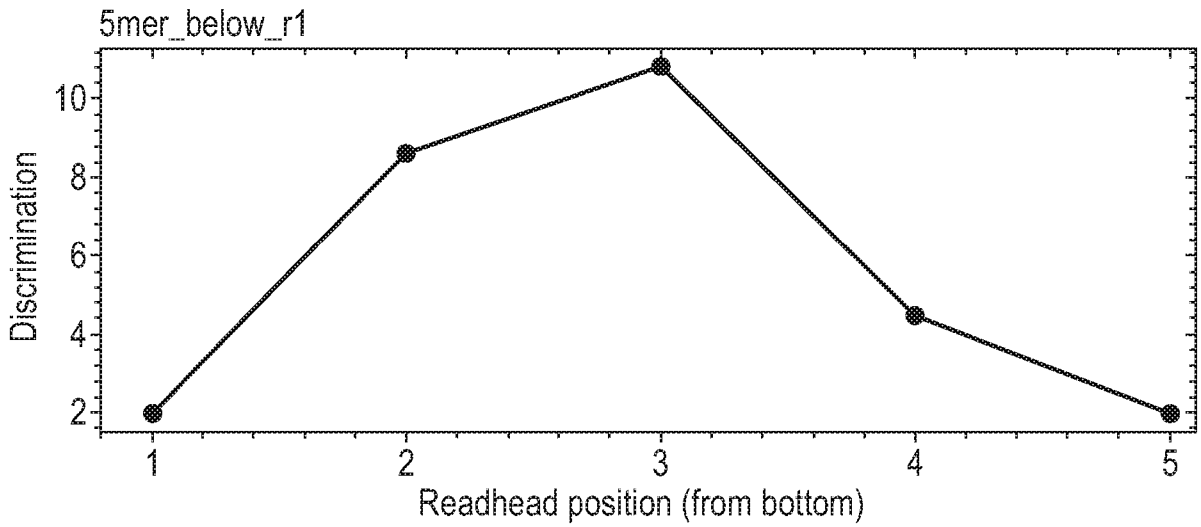


Fig. 9B

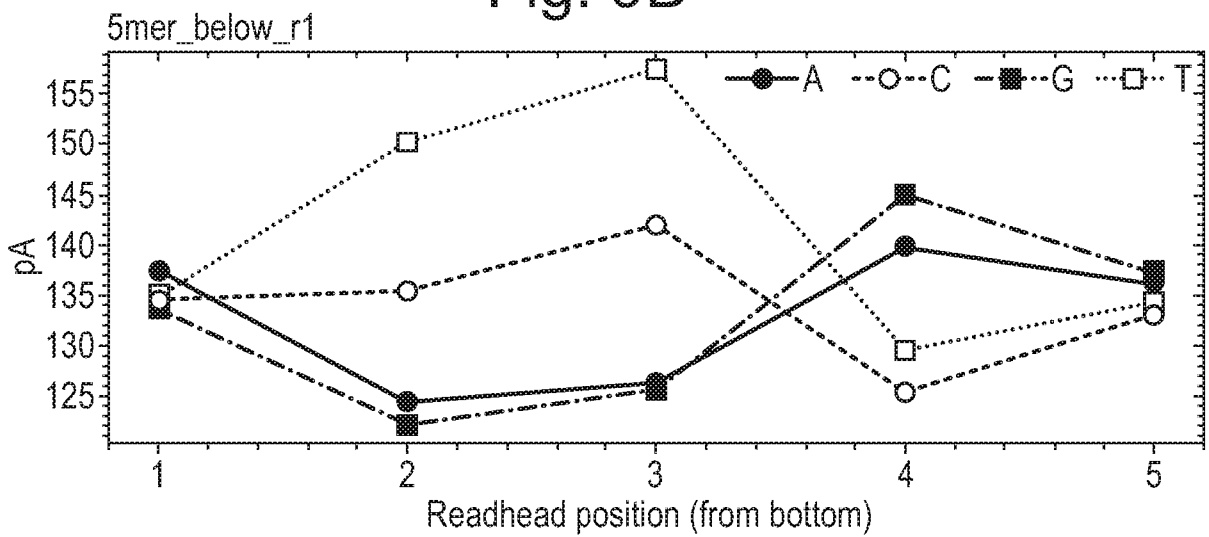


Fig. 9C

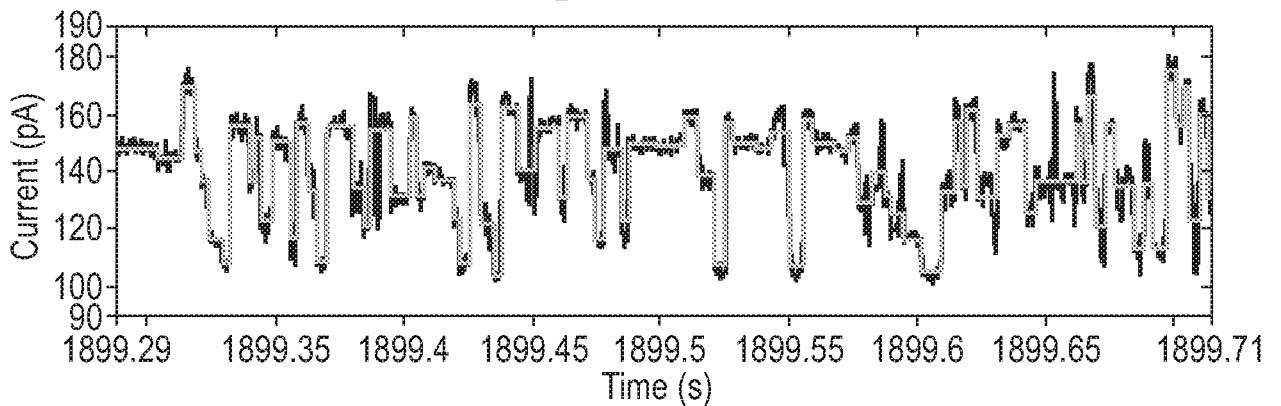


Fig. 10A

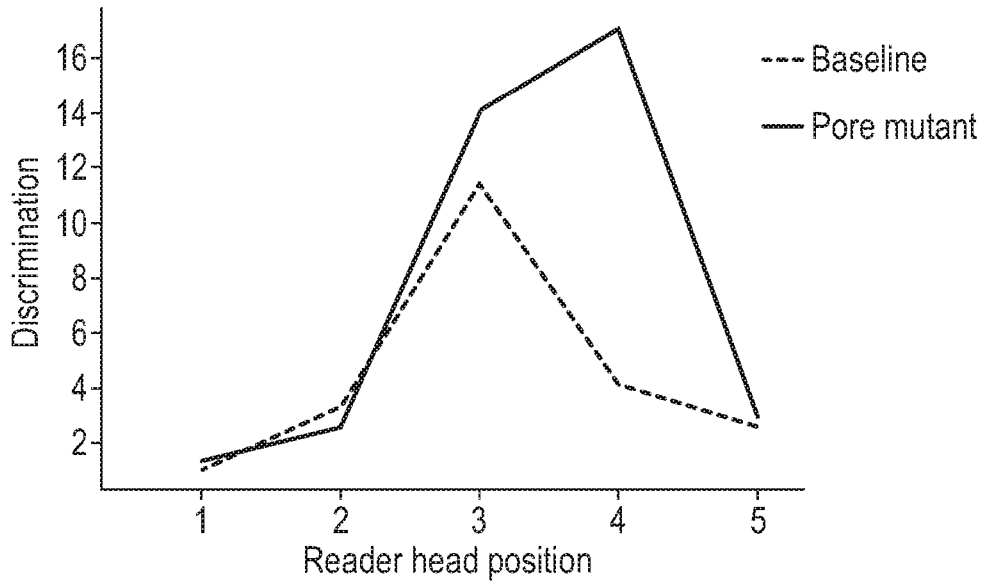
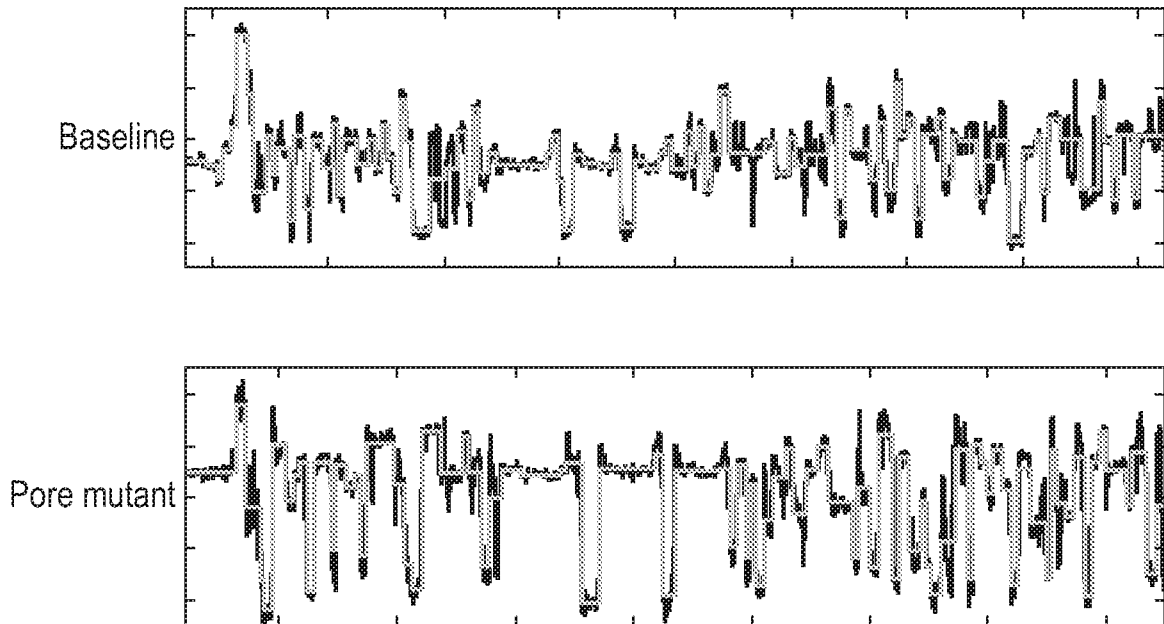


Fig. 10B



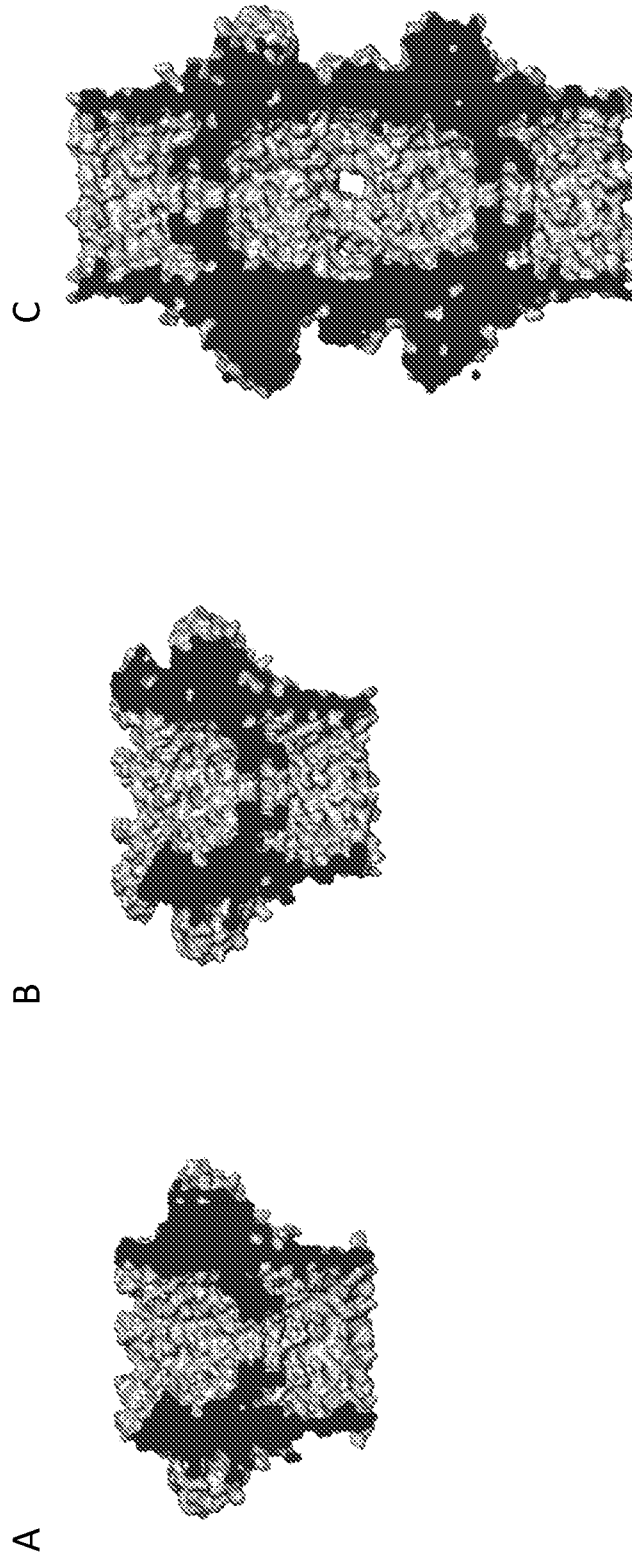


Figure 11

Fig. 12

R97 *

SEQ_ID_NO_2 1CUT--APPKAAKPTLMPRAQSYKDLHLPAFTCKTIVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVTALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-PLQSLTAANIVVEGSIIG 129

SEQ_ID_NO_5 1CUT--APPKAAKPTLMPRAQSYKDLHLPIPTCKTIVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVTALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-PLQSLTAANIVVEGSIIG 129

SEQ_ID_NO_6 1CUT--FPPKAAKPTLMPRAQSYKDLHLPPVTGKIVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVTALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-PLPSTLTAANIVVEGSIIG 129

SEQ_ID_NO_7 1CUT--APPKAAKPTLMPRAQSYRDLHLPAFTCKTIVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVTALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-PLQSLAAANVVEGSIIG 129

SEQ_ID_NO_9 1CUT--APPKAAKPTLMPRAQSYRDLNLPDPKGLFVSVVNIQDETGFQKPYPA-SNFSTAVPQSATMLV TALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAENNRN-PLQSLVAANVVEGSIIG 129

SEQ_ID_NO_10 1CUT--AAPKEAARPTLPPRAPSVDLHLPSQGRIFVSVVNIQDETGFQKPYPA-CNFSTAVPQSATAMLVSALKDSRFVPLERQGLQNLNKRKIIRAAQENGSVAINNQR-PLSSLVAANVVEGSIIG 129

SEQ_ID_NO_11 1CUT--SPPQAAKPTLPPASQYDLHLPEPQGRFVSVVNIQDETGFQKPYPA-SNFSTVPSQATAMLVSALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-QLPSLVAANVVEGSIIG 129

SEQ_ID_NO_12 1CUT--APPKQAAKPTLMPRAQSYQDLHLPEPAGKLFVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVSALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNQH-QLSSLVAANVVEGSIIG 129

SEQ_ID_NO_14 1-IT--EVPKEAARPTLMPRASTYKDLVALPKPNGKIVSVVNIQDETGFQKPYPA-SNFSTAVPQSGNAMLTSALKDSRFVPLEREGLOL NKRKIIRAAQENGTVAANNQ-PLPSLLSANVVEGSIIG 128

SEQ_ID_NO_18 1CATHIGSPVADEKATLMPRASVYKELISLPKPKGKIIVAVYDFRDQTGOVLPAPA-SNFSTAVTQGGVAMLSTALWDSQRFVPLEREGLOL NTKIIRAAQENGTVAINNRI-QLPSLVAANVVEGSIIG 131

SEQ_ID_NO_3 1-----MPRAQSYKDLHLPMPTCKTIVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVTALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-PLQSLTAANIVVEGSIIG 115

SEQ_ID_NO_17 1-----PEFSKEPTLMARGTAVQDLVSLPKPKGVSVVDFRDQTGOVLPQPN-SNFSTAVPQCGAALLTALDSTRMFLEREGLOLTKIIRAAQKDEIPTNHCV-HLPSLASANVVEGSIIG 124

SEQ_ID_NO_13 1CUT--TTPQAAKPTLPRDATYKDLVSLPQPKKIIVAVVNIQDETGFQKPYPA-SNFSTVPSQATAMLVSSLKDSRFVPLERQGLNLLNKRKIIRAAQENGTVAINNRI-PLPSLYSANVVEGSIIG 129

SEQ_ID_NO_22 1-----LTRRMSTYQDLIDMPAPRGKIIVTAVVSFRDQSGQVYKPPAPS-SSFSTAVTQGAAMLVVNLNDSRFVPLEREGLOLTKIIRAAKLDNDVPVWNSA-GLPSLLAAANVVEGSIIG 116

SEQ_ID_NO_15 1-----PEFSEPTLMQGANVIDLISLPKPKGKIIVSVYDFRDQTGOVLPQPN-SNFSTAVPQCGTALLTMALLDSEMFVPLERQGLQNLTKIIRAAQKQESISNHGS-TLPSLLSANVVEGSIIG 124

SEQ_ID_NO_16 1-----PDASEPTLMQGATYLDLISLPKPKGKIIVSVYDFRDQTGOVLPQPN-SNFSTAVPQCGTALLTMALLDSEMFVPLERQGLQNLTKIIRAAQKQESISNHGS-TLPSLLSANVVEGSIIG 124

SEQ_ID_NO_19 1-----ASSLMPKGESYVDLNLPAQGVMLAAVYDFRDQTGOVLPAPS-SNFSTAVPQSGTAFLLAQALNDSRFVPLEREGLOLTKIIRVRAGLK-----GDAN-KLPQLNSAQIKWEGSIIG 114

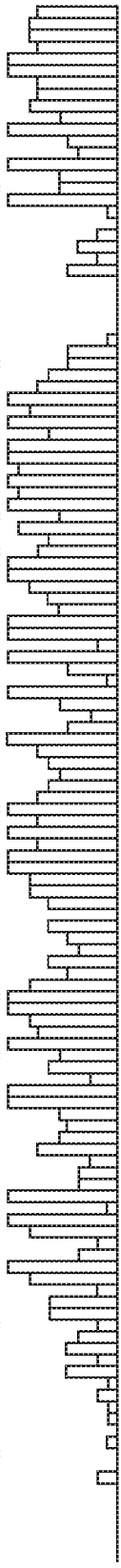
SEQ_ID_NO_21 1-----MPKSDTYDGLGCLPHPQGSMLAAVYDFRDQTGOVKAIPS-SNFSTAVPQSGTAFLLAQALNDSRFVPLEREGLOLTKIIRVRAGLK-----GEAN-QLPQLSSAQILMEGSIIG 109

SEQ_ID_NO_20 1-----QDSETPLETPRASTYDNLNMPREKGRMLMAVVYDFRDQTGOVLPAPA-SSFSTVTPQGAASMLMDALSASGWFVPLEREGLOLTKIIRASOKPQDVAENING-ELPPLQANLMEGSIIG 123

SEQ_ID_NO_23 1-----PSPPERSTMGELTPSTAELRNLPNEKIVIGVYKFRDQTGOVYKPSBNGNNSAVPQSGPTTILIKALEDSRFVPLERENIANLNERQIIRSTRQETMKDACKNSQSLPPLLYAGILLEGSIIG 126

SEQ_ID_NO_4 1CUT--APPKQAAKPTLMPRAQSYKDLHLPAFTCKTIVSVVNIQDETGFQKPYPA-SNFSTAVPQSATAMLVTALKDSRFVPLERQGLQNLNKRKIIRAAQENGTVAINNRI-PLQSLTAANIVVEGSIIG 129

Conservation



-----200101121525427729*429*1*4436856**376*89**957457-799**8*8*87678*536*15*2**679**87+6**+**7*9*87551000005242-1*66*45*6080**8998

Fig. 12 (Cont.)

R192

*

SEQ_ID NO 2 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 261

SEQ_ID NO 5 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 261

SEQ_ID NO 6 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 261

SEQ_ID NO 7 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-AQNFVLVKYRHSVTPPE 261

SEQ_ID NO 9 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 261

SEQ_ID NO 10 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 261

SEQ_ID NO 11 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 259

SEQ_ID NO 12 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 259

SEQ_ID NO 14 129 YSDDIKTGGCAGARYFCIGADCKYKVDQAVNLRADVDTGCVLLSVNTSKTILSSELSAGVRFIEYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 258

SEQ_ID NO 18 132 YDSNVRTGGCAGARYFCIGASGEYKVDQVTVNLRADVDIRSRILNSVFTSKTVMGQVQAGVRFVYKRLLEAEAGFSTNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAD-RQNDILVKYRHSVTPPE 261

SEQ_ID NO 3 116 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 247

SEQ_ID NO 17 125 YDNIQTGGCAGARYLGGGQVPRADQVTVNLRADVDTGCVLLSVNTSKTILSKEIQVGFVDFKDLLEAEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 254

SEQ_ID NO 13 130 YASNVRTGGCAGARYFCIGGSTQVQLDQAVNLRIVNVTGCVLLSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 255

SEQ_ID NO 22 117 YDSNIHTGGCAGARYFCIGASEKRYRVDVTVNLRADVDIRTCRILHSVLFSKKILSREIRSDVYRFIEFKHLLLEMEAGITINDPAQLCVLSAIESAVAHILVDFGVIKKSMSKADPNE-LNSPVIQAYQQQRI 245

SEQ_ID NO 15 125 YDSNIHTGGCAGARYLGGGQVPRADQVTVNLRADVDTGCVLLSVNTSKTILSKEIQVGFVDFKDLLEAEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 248

SEQ_ID NO 16 125 YDSNIHTGGCAGARYLGGGQVPRADQVTVNLRADVDTGCVLLSVNTSKTILSKEIQVGFVDFKDLLEAEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 248

SEQ_ID NO 19 115 YDNIHTGGCAGARYLGGGQVPRADQVTVNLRADVDTGCVLLSVNTSKTILSKEIQVGFVDFKDLLEAEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 239

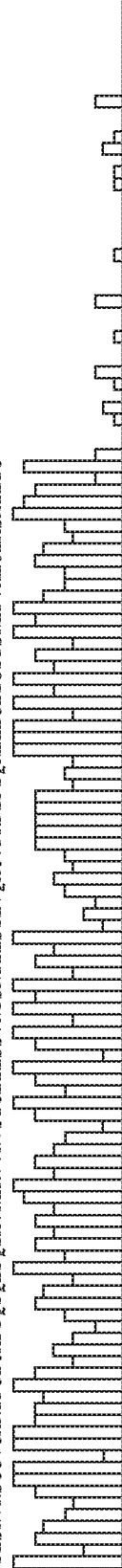
SEQ_ID NO 21 110 YDNIHTGGCAGARYLGGGQVPRADQVTVNLRADVDTGCVLLSVNTSKTILSKEIQVGFVDFKDLLEAEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 234

SEQ_ID NO 20 124 YDNIHTGGCAGARYLGGGQVPRADQVTVNLRADVDTGCVLLSVNTSKTILSKEIQVGFVDFKDLLEAEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 233

SEQ_ID NO 23 127 YDSNVTGGCAGARYFCIGASTQVPRQDRTIYLRVAVTLNGELIKLVYTSKTIKSTVNGSPFRYIDYERLLEAEVGLTQNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 234

SEQ_ID NO 4 130 YESNVSGGCGARYFCIGADTQVQLDQIAVNLRVNVTSTGEILSSVNTSKTILSYEVQAGVRFIDYQRLLLEGEVGYTSNEPVMCLMSAIEFTGVIFLINDGIDRGLNDLQNKAE-RQNDILVKYRHSVTPPE 223

Conservation



*498659**2**998*957536985*69797+*797629*69*29*5*9*597*243675699999565**5*7*795*9*8669856*49+300120130010030001000-011021003

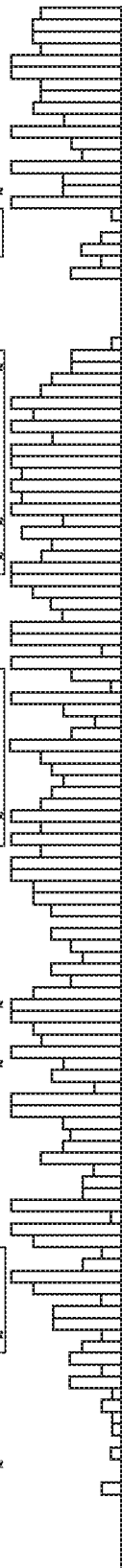
Fig. 13

R97

*

| | | | | | | | | | | | | | | | |
|--------|----|----|---|----------|---------|-----|-------|-------|--------|-------|------|----|-----|------|-----|
| SEQ_ID | NO | 2 | 1CUT--APPKEAARPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 5 | 1CUT--APPKEAARPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 6 | 1CUT--TPPKAAKPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 7 | 1CUT--APPKEAARPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 9 | 1CUT--APPKEAARPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 10 | 1CUT--AAPKEAARPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-CNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 11 | 1CUT--SPPKQAAKPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 12 | 1CUT--APPKQAAKPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 14 | 1-TT--EVPKEAARPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 128 |
| SEQ_ID | NO | 18 | 1CATHIGSPVADEKATLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 131 |
| SEQ_ID | NO | 3 | 1-----MPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 115 |
| SEQ_ID | NO | 17 | 1-----PETSKEPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 124 |
| SEQ_ID | NO | 13 | 1CUT--TPPKAAKPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |
| SEQ_ID | NO | 22 | 1-----LTRRMSTVQDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 116 |
| SEQ_ID | NO | 15 | 1-----PETSSEPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 124 |
| SEQ_ID | NO | 16 | 1-----PDASEPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 124 |
| SEQ_ID | NO | 19 | 1-----ASSLMPKESYDLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 114 |
| SEQ_ID | NO | 21 | 1-----MPKSDTYDLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 109 |
| SEQ_ID | NO | 20 | 1-----QDSEPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 123 |
| SEQ_ID | NO | 23 | 1-----PSDPERSTWGHLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 126 |
| SEQ_ID | NO | 4 | 1CUT--APPKQAAKPTLMPRAQSYKDLTLELPAPTGKIFVSVYNIODETQFKPYA-SNFSTAVPQSATAMLVTKD | SRWFIPLE | QGLNLLN | TKI | IRAAQ | ENGTV | AINNRI | -PLOS | LTAA | NI | VEG | STIG | 129 |

Conservation



-----201011121525427729*429*1*4438656**376*89**957457-799**8*8*876*8*536*15*2*6*79**87*6***7*9*87551000085242-1*66*45*6988**8998

Fig. 13 (Cont.)

R192

SEQ_ID NO 2 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 5 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 6 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 7 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 9 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 10 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 11 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 12 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 14 129 YDSDIKTGGAGARVFCGADGKRVVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 18 132 YDSNVRTGGAGAKYFCGASCEYRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAD-RQNDLLVKYRHMISVYPE 261

SEQ_ID NO 3 116 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 247

SEQ_ID NO 17 125 YDTNITQGGAGARVFCGASGQYRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 254

SEQ_ID NO 13 130 YASNVTGFGARVFCGASTQVQLDQVAVNLRVVNVSTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 248

SEQ_ID NO 22 117 YDSNLETGGAGARVFCGASEKYRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 245

SEQ_ID NO 15 125 YDSNITKGGAGARVFCGSGQYRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 248

SEQ_ID NO 16 125 YDSNITKGGAGARVFCGSGQYRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 248

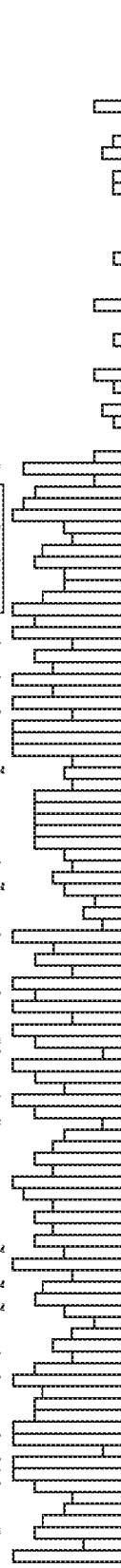
SEQ_ID NO 19 115 YDTNVRTGGAGARVFCGAAATQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 239

SEQ_ID NO 21 110 YDTNITKGGAGARVFCGWSKFRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 234

SEQ_ID NO 20 124 YDTNVRSGGVCARVFCGIDISREYRVDQVAVNLRAVDVNTGEVLLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 233

SEQ_ID NO 23 127 YDSNVTGFGARVFCGASTQVRODRITIVRAVSTLNGEILLKTVYTSKTIILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 234

SEQ_ID NO 4 130 YESNVKSGGVCARVFCGADTQVQLDQIAVNLRVVNVSTGEILLSSVNTSKTILSYEQAGVFRFDYQRLLEGEVGYTSNEPVMICLSALETGVIFLINDGIERGLMDLQNKAE-RQNDLLVKYRHMISVYPE 223



Conservation

*498659**2**998*957536985*69797**797629*69*29*5*9*597*243675699999565**5*7*795*9*866985*93*300120130010930001000-011021003-----

R97

*

Fig. 14

| | | | | | | | | | | | |
|--------------|---|-------|---|--------------------|----------------------------|--------------|--------|--------------|----------|----------|-----|
| SEQ_ID_NO_2 | 1 | ICLT | --APPKEAARPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLTAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_5 | 1 | ICLT | --APPKEAARPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLTAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_6 | 1 | ICLT | --TPPKAAKPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLPSLTAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_7 | 1 | ICLT | --APPKEAARPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLTAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_9 | 1 | ICLT | --APPKEAARPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLVAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_10 | 1 | ICLT | --AAPKEAARPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLSLSVAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_11 | 1 | ICLT | --SPPKQAAKPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-QLPSLVAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_12 | 1 | ICLT | --APPKQAAKPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-QLSLSVAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_14 | 1 | IT | --EVPKEAARPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLSLSVAANI | WVEGSIIG | 128 | |
| SEQ_ID_NO_18 | 1 | ICATH | IGSPVADEKATLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-QLPSLVAANI | WVEGSIIG | 131 | |
| SEQ_ID_NO_3 | 1 | --- | -----MPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLTAANI | WVEGSIIG | 115 | |
| SEQ_ID_NO_17 | 1 | --- | -----PETSKEPTLMARGTAVQDLSLPEPKKVIIVSVVDFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLSLSVAANI | WVEGSIIG | 124 | |
| SEQ_ID_NO_13 | 1 | ICLT | --TPPKAAKPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLSLSVAANI | WVEGSIIG | 129 | |
| SEQ_ID_NO_22 | 1 | --- | -----LTRRMSTYQDILMPAPRGKIVTAVYSFRDQSGQVKPAP | -SSFSTAVTQSAATAMLV | VNLDSSGFIPIERQGLONLLNERKI | IIRAAKRDV | VPVNSA | -GLP | SLSAANI | WVEGSIIG | 116 |
| SEQ_ID_NO_15 | 1 | --- | -----PETSSEPTLMQAGANIIDLISLPEPKKVIIVSVVDFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLSLSVAANI | WVEGSIIG | 124 | |
| SEQ_ID_NO_16 | 1 | --- | -----PDASESPTLMQAGAVIIDLISLPEPKKVIIVSVVDFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLSLSVAANI | WVEGSIIG | 124 | |
| SEQ_ID_NO_19 | 1 | --- | -----ASSLMPKGESYDILNLPAPQGMIAAVDFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLVAANI | WVEGSIIG | 114 | |
| SEQ_ID_NO_21 | 1 | --- | -----MPASDITYDILGLPFPQGSMLAAVDFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLVAANI | WVEGSIIG | 109 | |
| SEQ_ID_NO_20 | 1 | --- | -----QDSEPTLTPRASTYDILNLPAPQGMIAAVDFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLVAANI | WVEGSIIG | 123 | |
| SEQ_ID_NO_23 | 1 | --- | -----PSDPERSTWGLPTPSTAEILNLPENKIVIGVKKFRDTGQFKPYP | -SNFSTAVPQSAATAMLV | SALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLVAANI | WVEGSIIG | 126 | |
| SEQ_ID_NO_4 | 1 | ICLT | --APPKQAAKPTLMPRAQSYKDLIHLPAFTGKIFVSVVNIODETGQFKPYP | -SNFSTAVPQSAATAMLV | TALKDSRWFIPIERQGLONLLNERKI | IIRAAQENGTVA | INNR | I-PLQSLTAANI | WVEGSIIG | 129 | |



Conservation

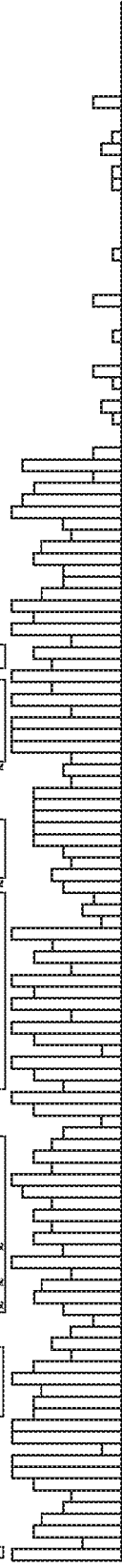
.....20010112152542729*1*4438656**376*89**957457-799**8*8*87678*536*15*2**679**87*6**+**7*9*87551000005242-1*66*45*6988**898

Fig. 14 (Cont.)

R.192

*

| | | | | |
|--------|-------|-----|---|-----|
| SEQ_ID | NO_2 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAE-RQNDILVVKYRHSVPPE | 261 |
| SEQ_ID | NO_5 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-RQNDILVVKYRHSVPPE | 261 |
| SEQ_ID | NO_6 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-RQNDILVVKYRHSVPPE | 261 |
| SEQ_ID | NO_7 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-AQNPVIVKYRHSVPPE | 261 |
| SEQ_ID | NO_9 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-AQNPVIVKYRHSVPPE | 261 |
| SEQ_ID | NO_10 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-AQNPVIVKYRHSVPPE | 261 |
| SEQ_ID | NO_11 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 259 |
| SEQ_ID | NO_12 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 259 |
| SEQ_ID | NO_14 | 129 | MSDIKTTGGAGARYFCIGADGKRVVDQVAVNLRAVNVSTCEVLLSSVNTSKTILSSELSAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 258 |
| SEQ_ID | NO_18 | 132 | MSNVRTGGAGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 261 |
| SEQ_ID | NO_3 | 116 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAE-RQNDILVVKYRHSVPPE | 247 |
| SEQ_ID | NO_17 | 125 | YDNTIQTGGAGARYLIGGSGQYRQDQVTVNLRAVNVSTCEILSSVNTSKTILSKEIQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 254 |
| SEQ_ID | NO_13 | 130 | MSNVRTGGAGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 255 |
| SEQ_ID | NO_22 | 117 | MSNIFTCGAGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSKEIQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 245 |
| SEQ_ID | NO_15 | 125 | YDSNITKTCGAGARYLIGGSGQYRQDQVTVNLRAVNVSTCEILSSVNTSKTILSVEVSAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 248 |
| SEQ_ID | NO_16 | 125 | YDSNITKTCGAGARYLIGGSGQYRQDQVTVNLRAVNVSTCEILSSVNTSKTILSVEVSAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 248 |
| SEQ_ID | NO_19 | 115 | YDTNVRTCGAGARYLIGGSGQYRQDQVTVNLRAVNVSTCEILSSVNTSKTILSKEIQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 239 |
| SEQ_ID | NO_21 | 110 | YDNTNITKTCGAGARYLIGGSGQYRQDQVTVNLRAVNVSTCEILSSVNTSKTILSVEVSAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 234 |
| SEQ_ID | NO_20 | 124 | YDTNVRTSGGEGARYLIGGSGQYRQDQVTVNLRAVNVSTCEILSSVNTSKTILSVEVSAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 233 |
| SEQ_ID | NO_23 | 127 | YDSNVRTCGGAGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSVEVSAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 234 |
| SEQ_ID | NO_4 | 130 | MSNVKSGGCGARYFCIGADTQYQDQIAVNLRAVNVSTCEILSSVNTSKTILSYEQAGVFRFDYQRLLEGEVETSTNEPVMCLMSAIEGVIIFLINDGIDRGLNDLQNKAD-INSFVLEKYSIIVP | 223 |



Conservation

*498659*2**998*957536985*69797**797629*69*29*5*9*597*24367569999565***5*7*795*9*8669858*493*300120130810030001000-011021003-*****

pctgb2018051191-seq1.txt
SEQUENCE LISTING

<110> OXFORD NANOPORE TECHNOLOGIES LIMITED

<120> PORE

<130> N410779W0

<150> GB1707122.6

<151> 2017-05-04

<160> 26

<170> PatentIn version 3.5

<210> 1

<211> 786

<212> DNA

<213> Escherichia coli

<400> 1

```
tgtctgaccg caccgccgaa agaagcggca cgtccgacc tgatgccgcg tgcacagtct      60
tataaagatc tgacccatct gccggctccg acgggcaaaa tttttgtag cgtctataac      120
atccaggacg aaaccggtca atttaaaccg taccggcgga gtaatttctc cacggccggt      180
ccgcagagtg caaccgctat gctggtcacg gcactgaaag attcccgttg gttcattccg      240
ctggaacgcc agggcctgca aaacctgctg aatgaacgta aaattatccg cgagctcag      300
gaaaacggta ccgtggccat taacaatcgt attccgctgc aaagcctgac cgccgcaaac      360
atcatggttg aaggctctat catcggttac gaatcaaacg tcaaacggg cgggtgtgggc      420
gcacgttatt ttggcattgg tgctgatacc cagtaccaac tggaccagat cgagttaac      480
ctgcgcgtgg ttaatgtcag caccggcgaa attctgagct ctgtgaatac cagcaaaacg      540
atcctgtctt acgaagtgca ggctgggtgtt tttcgtttca ttgattatca acgcctgctg      600
gaaggcgaag tcggttacac ctcaaacgaa ccggtgatgc tgtgtctgat gtcggcgatt      660
gaaacgggtg ttattttctt gatcaatgat ggcatcgacc gtggtctgtg ggatctgcag      720
aacaagccg aacgtcaaaa tgacattctg gtgaaatacc gccacatgag tgttccgccg      780
gaatcc      786
```

<210> 2

<211> 262

pctgb2018051191-seq1.txt

<212> PRT

<213> Escherichia coli

<400> 2

Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Arg Pro Thr Leu Met Pro
 1 5 10 15

Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Ala Pro Thr Gly
 20 25 30

Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
 35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
 50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro
 65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
 85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg Ile Pro
 100 105 110

Leu Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile
 115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
 130 135 140

Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
 145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser Val Asn
 165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
 180 185 190

pctgb2018051191-seql.txt

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Val Gly Tyr Thr Ser
195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val
210 215 220

Ile Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp Leu Gln
225 230 235 240

Asn Lys Ala Glu Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg His Met
245 250 255

Ser Val Pro Pro Glu Ser
260

<210> 3
<211> 248
<212> PRT
<213> Citrobacter koseri

<400> 3

Met Pro Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Met Pro
1 5 10 15

Thr Gly Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly
20 25 30

Gln Phe Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln
35 40 45

Ser Ala Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe
50 55 60

Ile Pro Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys
65 70 75 80

Ile Ile Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg
85 90 95

pctgb2018051191-seql.txt

Ile Pro Leu Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser
100 105 110

Ile Ile Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg
115 120 125

Tyr Phe Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala
130 135 140

Val Asn Leu Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser
145 150 155 160

Val Asn Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val
165 170 175

Phe Arg Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr
180 185 190

Thr Ser Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr
195 200 205

Gly Val Ile Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp
210 215 220

Leu Gln Asn Lys Ala Glu Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg
225 230 235 240

His Met Ser Val Pro Pro Glu Ser
245

<210> 4
<211> 223
<212> PRT
<213> Salmonella enterica

<400> 4

Cys Leu Thr Ala Pro Pro Lys Gln Ala Ala Lys Pro Thr Leu Met Pro
1 5 10 15

Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Ala Pro Thr Gly

pctgb2018051191-seql.txt

20

25

30

Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
 35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
 50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro
 65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
 85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Met Asn Asn Arg Ile Pro
 100 105 110

Leu Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile
 115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
 130 135 140

Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
 145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser Val Asn
 165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
 180 185 190

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Ser
 195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly
 210 215 220

<210> 5

pctgb2018051191-seq1.txt

<211> 262

<212> PRT

<213> Citrobacter amalonaticus

<400> 5

Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Lys Pro Thr Leu Met Pro
1 5 10 15

Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Ile Pro Thr Gly
20 25 30

Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Val Pro
65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg Ile Pro
100 105 110

Leu Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile
115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
130 135 140

Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser Val Asn
165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
180 185 190

pctgb2018051191-seql.txt

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Ser
195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val
210 215 220

Ile Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp Leu Gln
225 230 235 240

Asn Lys Ala Asp Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg His Met
245 250 255

Ser Val Pro Pro Glu Ser
260

<210> 6
<211> 262
<212> PRT
<213> Citrobacter rodentium

<400> 6

Cys Leu Thr Thr Pro Pro Lys Glu Ala Ala Lys Pro Thr Leu Met Pro
1 5 10 15

Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Val Pro Thr Gly
20 25 30

Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro
65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

pctgb2018051191-seql.txt

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg Ile Pro
100 105 110

Leu Pro Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile
115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Ala Gly Ala Arg Tyr Phe
130 135 140

Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser Val Asn
165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
180 185 190

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Ser
195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val
210 215 220

Ile Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp Leu Gln
225 230 235 240

Asn Lys Ala Asp Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg Gln Met
245 250 255

Ser Val Pro Pro Glu Ser
260

- <210> 7
- <211> 262
- <212> PRT
- <213> Enterobacter asburiae

- <400> 7

pctgb2018051191-seq1.txt

Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Lys Pro Thr Leu Met Pro
 1 5 10 15

Arg Ala Gln Ser Tyr Arg Asp Leu Thr His Leu Pro Ala Pro Thr Gly
 20 25 30

Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
 35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
 50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser His Trp Phe Ile Pro
 65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
 85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Asn Asn Asn Arg Met Pro
 100 105 110

Leu Gln Ser Leu Ala Ala Ala Asn Val Met Ile Glu Gly Ser Ile Ile
 115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
 130 135 140

Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
 145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Val Leu Ser Ser Val Asn
 165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
 180 185 190

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Ser
 195 200 205

pctgb2018051191-seq1.txt

Asn Glu Pro Val Met Met Cys Leu Met Ser Ala Ile Glu Thr Gly Val
210 215 220

Ile Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp Leu Gln
225 230 235 240

Asn Lys Ala Asp Ala Gln Asn Pro Val Leu Val Lys Tyr Arg Asp Met
245 250 255

Ser Val Pro Pro Glu Ser
260

<210> 8
<211> 262
<212> PRT
<213> Yokenella regensburgei

<400> 8

Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Lys Pro Thr Leu Met Pro
1 5 10 15

Arg Ala Gln Ser Tyr Arg Asp Leu Thr His Leu Pro Leu Pro Ser Gly
20 25 30

Lys Val Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
50 55 60

Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Val Pro
65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Asp Asn Asn Arg Ile Pro
100 105 110

Leu Gln Ser Leu Thr Ala Ala Asn Val Met Ile Glu Gly Ser Ile Ile

115

120

125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
 130 135 140

Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
 145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Val Leu Ser Ser Val Asn
 165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
 180 185 190

Phe Val Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Ser
 195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val
 210 215 220

Ile Tyr Leu Ile Asn Asp Gly Ile Glu Arg Gly Leu Trp Asp Leu Gln
 225 230 235 240

Gln Lys Ala Asp Val Asp Asn Pro Ile Leu Ala Arg Tyr Arg Asn Met
 245 250 255

Ser Ala Pro Pro Glu Ser
 260

<210> 9
 <211> 262
 <212> PRT
 <213> Cronobacter pulveris

<400> 9

Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Lys Pro Thr Leu Met Pro
 1 5 10 15

Arg Ala Gln Ser Tyr Arg Asp Leu Thr Asn Leu Pro Asp Pro Lys Gly
 20 25 30

pctgb2018051191-seql.txt

Lys Leu Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
50 55 60

Thr Ser Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro
65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Glu Asn Asn Arg Met Pro
100 105 110

Leu Gln Ser Leu Val Ala Ala Asn Val Met Ile Glu Gly Ser Ile Ile
115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
130 135 140

Gly Ile Gly Gly Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Val Leu Ser Ser Val Asn
165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
180 185 190

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Ala
195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val
210 215 220

Ile His Leu Ile Asn Asp Gly Ile Asn Arg Gly Leu Trp Glu Leu Lys
225 230 235 240

pctgb2018051191-seql.txt

Asn Lys Gly Asp Ala Lys Asn Thr Ile Leu Ala Lys Tyr Arg Ser Met
245 250 255

Ala Val Pro Pro Glu Ser
260

<210> 10
<211> 262
<212> PRT
<213> Rahnella aquatilis

<400> 10

Cys Leu Thr Ala Ala Pro Lys Glu Ala Ala Arg Pro Thr Leu Leu Pro
1 5 10 15

Arg Ala Pro Ser Tyr Thr Asp Leu Thr His Leu Pro Ser Pro Gln Gly
20 25 30

Arg Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
35 40 45

Lys Pro Tyr Pro Ala Cys Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
50 55 60

Thr Ala Met Leu Val Ser Ala Leu Lys Asp Ser Lys Trp Phe Ile Pro
65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

Arg Ala Ala Gln Glu Asn Gly Ser Val Ala Ile Asn Asn Gln Arg Pro
100 105 110

Leu Ser Ser Leu Val Ala Ala Asn Ile Leu Ile Glu Gly Ser Ile Ile
115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe
130 135 140

pctgb2018051191-seql.txt

Gly Ile Gly Ala Ser Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
145 150 155 160

Leu Arg Ala Val Asp Val Asn Thr Gly Glu Val Leu Ser Ser Val Asn
165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
180 185 190

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Leu Gly Tyr Thr Thr
195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Ser Gly Val
210 215 220

Ile Tyr Leu Val Asn Asp Gly Ile Glu Arg Asn Leu Trp Gln Leu Gln
225 230 235 240

Asn Pro Ser Glu Ile Asn Ser Pro Ile Leu Gln Arg Tyr Lys Asn Asn
245 250 255

Ile Val Pro Ala Glu Ser
260

<210> 11
<211> 259
<212> PRT
<213> Kluyvera ascorbata

<400> 11

Cys Ile Thr Ser Pro Pro Lys Gln Ala Ala Lys Pro Thr Leu Leu Pro
1 5 10 15

Arg Ser Gln Ser Tyr Gln Asp Leu Thr His Leu Pro Glu Pro Gln Gly
20 25 30

Arg Leu Phe Val Ser Val Tyr Asn Ile Ser Asp Glu Thr Gly Gln Phe
35 40 45

pctgb2018051191-seq1.txt

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ser Val Pro Gln Ser Ala
 50 55 60

Thr Ala Met Leu Val Ser Ala Leu Lys Asp Ser Asn Trp Phe Ile Pro
 65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
 85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Val Asn Asn Arg Thr Gln
 100 105 110

Leu Pro Ser Leu Val Ala Ala Asn Ile Leu Ile Glu Gly Ser Ile Ile
 115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Ala Gly Ala Arg Tyr Phe
 130 135 140

Gly Ile Gly Ala Ser Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
 145 150 155 160

Leu Arg Val Val Asn Val Ser Thr Gly Glu Val Leu Ser Ser Val Asn
 165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Phe Gln Ala Gly Val Phe Arg
 180 185 190

Tyr Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Val Gly Tyr Thr Val
 195 200 205

Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val
 210 215 220

Ile Tyr Leu Val Asn Asp Gly Ile Ser Arg Asn Leu Trp Gln Leu Lys
 225 230 235 240

Asn Ala Ser Asp Ile Asn Ser Pro Val Leu Glu Lys Tyr Lys Ser Ile
 245 250 255

pctgb2018051191-seql.txt

Ile Val Pro

<210> 12
<211> 259
<212> PRT
<213> Hafnia alvei

<400> 12

Cys Leu Thr Ala Pro Pro Lys Gln Ala Ala Lys Pro Thr Leu Met Pro
1 5 10 15

Arg Ala Gln Ser Tyr Gln Asp Leu Thr His Leu Pro Glu Pro Ala Gly
20 25 30

Lys Leu Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
35 40 45

Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala
50 55 60

Thr Ala Met Leu Val Ser Ala Leu Lys Asp Ser Gly Trp Phe Ile Pro
65 70 75 80

Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

Arg Ala Ala Gln Glu Asn Gly Thr Ala Ala Val Asn Asn Gln His Gln
100 105 110

Leu Ser Ser Leu Val Ala Ala Asn Val Leu Val Glu Gly Ser Ile Ile
115 120 125

Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Ala Gly Ala Arg Phe Phe
130 135 140

Gly Ile Gly Ala Ser Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn
145 150 155 160

Leu Arg Val Val Asp Val Asn Thr Gly Gln Val Leu Ser Ser Val Asn

pctgb2018051191-seql.txt

165

170

175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg
 180 185 190

Tyr Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ile Gly Tyr Thr Thr
 195 200 205

Asn Glu Pro Val Met Leu Cys Val Met Ser Ala Ile Glu Thr Gly Val
 210 215 220

Ile Tyr Leu Val Asn Asp Gly Ile Asn Arg Asn Leu Trp Thr Leu Lys
 225 230 235 240

Asn Pro Gln Asp Ala Lys Ser Ser Val Leu Glu Arg Tyr Lys Ser Thr
 245 250 255

Ile Val Pro

- <210> 13
- <211> 255
- <212> PRT
- <213> Enterobacteriaceae bacterium strain FGI 57

<400> 13

Cys Ile Thr Thr Pro Pro Gln Glu Ala Ala Lys Pro Thr Leu Leu Pro
 1 5 10 15

Arg Asp Ala Thr Tyr Lys Asp Leu Val Ser Leu Pro Gln Pro Arg Gly
 20 25 30

Lys Ile Tyr Val Ala Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe
 35 40 45

Gln Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ser Val Pro Gln Ser Ala
 50 55 60

Thr Ala Met Leu Val Ser Ser Leu Lys Asp Ser Arg Trp Phe Val Pro
 65 70 75 80

pctgb2018051191-seql.txt

Leu Glu Arg Gln Gly Leu Asn Asn Leu Leu Asn Glu Arg Lys Ile Ile
85 90 95

Arg Ala Ala Gln Gln Asn Gly Thr Val Gly Asp Asn Asn Ala Ser Pro
100 105 110

Leu Pro Ser Leu Tyr Ser Ala Asn Val Ile Val Glu Gly Ser Ile Ile
115 120 125

Gly Tyr Ala Ser Asn Val Lys Thr Gly Gly Phe Gly Ala Arg Tyr Phe
130 135 140

Gly Ile Gly Gly Ser Thr Gln Tyr Gln Leu Asp Gln Val Ala Val Asn
145 150 155 160

Leu Arg Ile Val Asn Val His Thr Gly Glu Val Leu Ser Ser Val Asn
165 170 175

Thr Ser Lys Thr Ile Leu Ser Tyr Glu Ile Gln Ala Gly Val Phe Arg
180 185 190

Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Ala Gly Phe Thr Thr
195 200 205

Asn Glu Pro Val Met Thr Cys Leu Met Ser Ala Ile Glu Glu Gly Val
210 215 220

Ile His Leu Ile Asn Asp Gly Ile Asn Lys Lys Leu Trp Ala Leu Ser
225 230 235 240

Asn Ala Ala Asp Ile Asn Ser Glu Val Leu Thr Arg Tyr Arg Lys
245 250 255

<210> 14
<211> 258
<212> PRT
<213> Plesiomonas shigelloides

<400> 14

pctgb2018051191-seq1.txt

Ile Thr Glu Val Pro Lys Glu Ala Ala Lys Pro Thr Leu Met Pro Arg
 1 5 10 15

Ala Ser Thr Tyr Lys Asp Leu Val Ala Leu Pro Lys Pro Asn Gly Lys
 20 25 30

Ile Ile Val Ser Val Tyr Ser Val Gln Asp Glu Thr Gly Gln Phe Lys
 35 40 45

Pro Leu Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Gly Asn
 50 55 60

Ala Met Leu Thr Ser Ala Leu Lys Asp Ser Gly Trp Phe Val Pro Leu
 65 70 75 80

Glu Arg Glu Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile Arg
 85 90 95

Ala Ala Gln Glu Asn Gly Thr Val Ala Ala Asn Asn Gln Gln Pro Leu
 100 105 110

Pro Ser Leu Leu Ser Ala Asn Val Val Ile Glu Gly Ala Ile Ile Gly
 115 120 125

Tyr Asp Ser Asp Ile Lys Thr Gly Gly Ala Gly Ala Arg Tyr Phe Gly
 130 135 140

Ile Gly Ala Asp Gly Lys Tyr Arg Val Asp Gln Val Ala Val Asn Leu
 145 150 155 160

Arg Ala Val Asp Val Arg Thr Gly Glu Val Leu Leu Ser Val Asn Thr
 165 170 175

Ser Lys Thr Ile Leu Ser Ser Glu Leu Ser Ala Gly Val Phe Arg Phe
 180 185 190

Ile Glu Tyr Gln Arg Leu Leu Glu Leu Glu Ala Gly Tyr Thr Thr Asn
 195 200 205

pctgb2018051191-seq1.txt

Glu Pro Val Met Met Cys Met Met Ser Ala Leu Glu Ala Gly Val Ala
210 215 220

His Leu Ile Val Glu Gly Ile Arg Gln Asn Leu Trp Ser Leu Gln Asn
225 230 235 240

Pro Ser Asp Ile Asn Asn Pro Ile Ile Gln Arg Tyr Met Lys Glu Asp
245 250 255

Val Pro

<210> 15
<211> 248
<212> PRT
<213> Vibrio fischeri

<400> 15

Pro Glu Thr Ser Glu Ser Pro Thr Leu Met Gln Arg Gly Ala Asn Tyr
1 5 10 15

Ile Asp Leu Ile Ser Leu Pro Lys Pro Gln Gly Lys Ile Phe Val Ser
20 25 30

Val Tyr Asp Phe Arg Asp Gln Thr Gly Gln Tyr Lys Pro Gln Pro Asn
35 40 45

Ser Asn Phe Ser Thr Ala Val Pro Gln Gly Gly Thr Ala Leu Leu Thr
50 55 60

Met Ala Leu Leu Asp Ser Glu Trp Phe Tyr Pro Leu Glu Arg Gln Gly
65 70 75 80

Leu Gln Asn Leu Leu Thr Glu Arg Lys Ile Ile Arg Ala Ala Gln Lys
85 90 95

Lys Gln Glu Ser Ile Ser Asn His Gly Ser Thr Leu Pro Ser Leu Leu
100 105 110

pctgb2018051191-seq1.txt

Ser Ala Asn Val Met Ile Glu Gly Gly Ile Val Ala Tyr Asp Ser Asn
115 120 125

Ile Lys Thr Gly Gly Ala Gly Ala Arg Tyr Leu Gly Ile Gly Gly Ser
130 135 140

Gly Gln Tyr Arg Ala Asp Gln Val Thr Val Asn Ile Arg Ala Val Asp
145 150 155 160

Val Arg Ser Gly Lys Ile Leu Thr Ser Val Thr Thr Ser Lys Thr Ile
165 170 175

Leu Ser Tyr Glu Val Ser Ala Gly Ala Phe Arg Phe Val Asp Tyr Lys
180 185 190

Glu Leu Leu Glu Val Glu Leu Gly Tyr Thr Asn Asn Glu Pro Val Asn
195 200 205

Ile Ala Leu Met Ser Ala Ile Asp Ser Ala Val Ile His Leu Ile Val
210 215 220

Lys Gly Val Gln Gln Gly Leu Trp Arg Pro Ala Asn Leu Asp Thr Arg
225 230 235 240

Asn Asn Pro Ile Phe Lys Lys Tyr
245

<210> 16
<211> 248
<212> PRT
<213> Aliivibrio logei

<400> 16

Pro Asp Ala Ser Glu Ser Pro Thr Leu Met Gln Arg Gly Ala Thr Tyr
1 5 10 15

Leu Asp Leu Ile Ser Leu Pro Lys Pro Gln Gly Lys Ile Tyr Val Ser
20 25 30

Val Tyr Asp Phe Arg Asp Gln Thr Gly Gln Tyr Lys Pro Gln Pro Asn

35

40

45

Ser Asn Phe Ser Thr Ala Val Pro Gln Gly Gly Thr Ala Leu Leu Thr
 50 55 60

Met Ala Leu Leu Asp Ser Glu Trp Phe Tyr Pro Leu Glu Arg Gln Gly
 65 70 75 80

Leu Gln Asn Leu Leu Thr Glu Arg Lys Ile Ile Arg Ala Ala Gln Lys
 85 90 95

Lys Gln Glu Ser Ile Ser Asn His Gly Ser Thr Leu Pro Ser Leu Leu
 100 105 110

Ser Ala Asn Val Met Ile Glu Gly Gly Ile Val Ala Tyr Asp Ser Asn
 115 120 125

Ile Lys Thr Gly Gly Ala Gly Ala Arg Tyr Leu Gly Ile Gly Gly Ser
 130 135 140

Gly Gln Tyr Arg Ala Asp Gln Val Thr Val Asn Ile Arg Ala Val Asp
 145 150 155 160

Val Arg Ser Gly Lys Ile Leu Thr Ser Val Thr Thr Ser Lys Thr Ile
 165 170 175

Leu Ser Tyr Glu Leu Ser Ala Gly Ala Phe Arg Phe Val Asp Tyr Lys
 180 185 190

Glu Leu Leu Glu Val Glu Leu Gly Tyr Thr Asn Asn Glu Pro Val Asn
 195 200 205

Ile Ala Leu Met Ser Ala Ile Asp Ser Ala Val Ile His Leu Ile Val
 210 215 220

Lys Gly Ile Glu Glu Gly Leu Trp Arg Pro Glu Asn Gln Asn Gly Lys
 225 230 235 240

Glu Asn Pro Ile Phe Arg Lys Tyr

<210> 17
 <211> 254
 <212> PRT
 <213> Photobacterium sp. AK15

<400> 17

Pro Glu Thr Ser Lys Glu Pro Thr Leu Met Ala Arg Gly Thr Ala Tyr
 1 5 10 15

Gln Asp Leu Val Ser Leu Pro Leu Pro Lys Gly Lys Val Tyr Val Ser
 20 25 30

Val Tyr Asp Phe Arg Asp Gln Thr Gly Gln Tyr Lys Pro Gln Pro Asn
 35 40 45

Ser Asn Phe Ser Thr Ala Val Pro Gln Gly Gly Ala Ala Leu Leu Thr
 50 55 60

Thr Ala Leu Leu Asp Ser Arg Trp Phe Met Pro Leu Glu Arg Glu Gly
 65 70 75 80

Leu Gln Asn Leu Leu Thr Glu Arg Lys Ile Ile Arg Ala Ala Gln Lys
 85 90 95

Lys Asp Glu Ile Pro Thr Asn His Gly Val His Leu Pro Ser Leu Ala
 100 105 110

Ser Ala Asn Ile Met Val Glu Gly Gly Ile Val Ala Tyr Asp Thr Asn
 115 120 125

Ile Gln Thr Gly Gly Ala Gly Ala Arg Tyr Leu Gly Val Gly Ala Ser
 130 135 140

Gly Gln Tyr Arg Thr Asp Gln Val Thr Val Asn Ile Arg Ala Val Asp
 145 150 155 160

Val Arg Thr Gly Arg Ile Leu Leu Ser Val Thr Thr Ser Lys Thr Ile
 165 170 175

pctgb2018051191-seq1.txt

Leu Ser Lys Glu Leu Gln Thr Gly Val Phe Lys Phe Val Asp Tyr Lys
180 185 190

Asp Leu Leu Glu Ala Glu Leu Gly Tyr Thr Thr Asn Glu Pro Val Asn
195 200 205

Leu Ala Val Met Ser Ala Ile Asp Ala Ala Val Val His Val Ile Val
210 215 220

Asp Gly Ile Lys Thr Gly Leu Trp Glu Pro Leu Arg Gly Glu Asp Leu
225 230 235 240

Gln His Pro Ile Ile Gln Glu Tyr Met Asn Arg Ser Lys Pro
245 250

<210> 18
<211> 261
<212> PRT
<213> Aeromonas veronii

<400> 18

Cys Ala Thr His Ile Gly Ser Pro Val Ala Asp Glu Lys Ala Thr Leu
1 5 10 15

Met Pro Arg Ser Val Ser Tyr Lys Glu Leu Ile Ser Leu Pro Lys Pro
20 25 30

Lys Gly Lys Ile Val Ala Ala Val Tyr Asp Phe Arg Asp Gln Thr Gly
35 40 45

Gln Tyr Leu Pro Ala Pro Ala Ser Asn Phe Ser Thr Ala Val Thr Gln
50 55 60

Gly Gly Val Ala Met Leu Ser Thr Ala Leu Trp Asp Ser Gln Trp Phe
65 70 75 80

Val Pro Leu Glu Arg Glu Gly Leu Gln Asn Leu Leu Thr Glu Arg Lys
85 90 95

pctgb2018051191-seql.txt

Ile Val Arg Ala Ala Gln Asn Lys Pro Asn Val Pro Gly Asn Asn Ala
100 105 110

Asn Gln Leu Pro Ser Leu Val Ala Ala Asn Ile Leu Ile Glu Gly Gly
115 120 125

Ile Val Ala Tyr Asp Ser Asn Val Arg Thr Gly Gly Ala Gly Ala Lys
130 135 140

Tyr Phe Gly Ile Gly Ala Ser Gly Glu Tyr Arg Val Asp Gln Val Thr
145 150 155 160

Val Asn Leu Arg Ala Val Asp Ile Arg Ser Gly Arg Ile Leu Asn Ser
165 170 175

Val Thr Thr Ser Lys Thr Val Met Ser Gln Gln Val Gln Ala Gly Val
180 185 190

Phe Arg Phe Val Glu Tyr Lys Arg Leu Leu Glu Ala Glu Ala Gly Phe
195 200 205

Ser Thr Asn Glu Pro Val Gln Met Cys Val Met Ser Ala Ile Glu Ser
210 215 220

Gly Val Ile Arg Leu Ile Ala Asn Gly Val Arg Asp Asn Leu Trp Gln
225 230 235 240

Leu Ala Asp Gln Arg Asp Ile Asp Asn Pro Ile Leu Gln Glu Tyr Leu
245 250 255

Gln Asp Asn Ala Pro
260

<210> 19
<211> 239
<212> PRT
<213> Shewanella sp. ECSMB14101

<400> 19

pctgb2018051191-seql.txt

Ala Ser Ser Ser Leu Met Pro Lys Gly Glu Ser Tyr Tyr Asp Leu Ile
 1 5 10 15

Asn Leu Pro Ala Pro Gln Gly Val Met Leu Ala Ala Val Tyr Asp Phe
 20 25 30

Arg Asp Gln Thr Gly Gln Tyr Lys Pro Ile Pro Ser Ser Asn Phe Ser
 35 40 45

Thr Ala Val Pro Gln Ser Gly Thr Ala Phe Leu Ala Gln Ala Leu Asn
 50 55 60

Asp Ser Ser Trp Phe Ile Pro Val Glu Arg Glu Gly Leu Gln Asn Leu
 65 70 75 80

Leu Thr Glu Arg Lys Ile Val Arg Ala Gly Leu Lys Gly Asp Ala Asn
 85 90 95

Lys Leu Pro Gln Leu Asn Ser Ala Gln Ile Leu Met Glu Gly Gly Ile
 100 105 110

Val Ala Tyr Asp Thr Asn Val Arg Thr Gly Gly Ala Gly Ala Arg Tyr
 115 120 125

Leu Gly Ile Gly Ala Ala Thr Gln Phe Arg Val Asp Thr Val Thr Val
 130 135 140

Asn Leu Arg Ala Val Asp Ile Arg Thr Gly Arg Leu Leu Ser Ser Val
 145 150 155 160

Thr Thr Thr Lys Ser Ile Leu Ser Lys Glu Ile Thr Ala Gly Val Phe
 165 170 175

Lys Phe Ile Asp Ala Gln Glu Leu Leu Glu Ser Glu Leu Gly Tyr Thr
 180 185 190

Ser Asn Glu Pro Val Ser Leu Cys Val Ala Ser Ala Ile Glu Ser Ala
 195 200 205

pctgb2018051191-seql.txt

Val Val His Met Ile Ala Asp Gly Ile Trp Lys Gly Ala Trp Asn Leu
210 215 220

Ala Asp Gln Ala Ser Gly Leu Arg Ser Pro Val Leu Gln Lys Tyr
225 230 235

<210> 20
<211> 233
<212> PRT
<213> Pseudomonas putida

<400> 20

Gln Asp Ser Glu Thr Pro Thr Leu Thr Pro Arg Ala Ser Thr Tyr Tyr
1 5 10 15

Asp Leu Ile Asn Met Pro Arg Pro Lys Gly Arg Leu Met Ala Val Val
20 25 30

Tyr Gly Phe Arg Asp Gln Thr Gly Gln Tyr Lys Pro Thr Pro Ala Ser
35 40 45

Ser Phe Ser Thr Ser Val Thr Gln Gly Ala Ala Ser Met Leu Met Asp
50 55 60

Ala Leu Ser Ala Ser Gly Trp Phe Val Val Leu Glu Arg Glu Gly Leu
65 70 75 80

Gln Asn Leu Leu Thr Glu Arg Lys Ile Ile Arg Ala Ser Gln Lys Lys
85 90 95

Pro Asp Val Ala Glu Asn Ile Met Gly Glu Leu Pro Pro Leu Gln Ala
100 105 110

Ala Asn Leu Met Leu Glu Gly Gly Ile Ile Ala Tyr Asp Thr Asn Val
115 120 125

Arg Ser Gly Gly Glu Gly Ala Arg Tyr Leu Gly Ile Asp Ile Ser Arg
130 135 140

Glu Tyr Arg Val Asp Gln Val Thr Val Asn Leu Arg Ala Val Asp Val

pctgb2018051191-seql.txt

Ser Ser Ala Gln Ile Leu Met Glu Gly Gly Ile Val Ala Tyr Asp Thr
100 105 110

Asn Ile Lys Thr Gly Gly Ala Gly Ala Arg Tyr Leu Gly Ile Gly Val
115 120 125

Asn Ser Lys Phe Arg Val Asp Thr Val Thr Val Asn Leu Arg Ala Val
130 135 140

Asp Ile Arg Thr Gly Arg Leu Leu Ser Ser Val Thr Thr Thr Lys Ser
145 150 155 160

Ile Leu Ser Lys Glu Val Ser Ala Gly Val Phe Lys Phe Ile Asp Ala
165 170 175

Gln Asp Leu Leu Glu Ser Glu Leu Gly Tyr Thr Ser Asn Glu Pro Val
180 185 190

Ser Leu Cys Val Ala Gln Ala Ile Glu Ser Ala Val Val His Met Ile
195 200 205

Ala Asp Gly Ile Trp Lys Arg Ala Trp Asn Leu Ala Asp Thr Ala Ser
210 215 220

Gly Leu Asn Asn Pro Val Leu Gln Lys Tyr
225 230

<210> 22
<211> 245
<212> PRT
<213> Marinobacterium jannaschii

<400> 22

Leu Thr Arg Arg Met Ser Thr Tyr Gln Asp Leu Ile Asp Met Pro Ala
1 5 10 15

Pro Arg Gly Lys Ile Val Thr Ala Val Tyr Ser Phe Arg Asp Gln Ser
20 25 30

pctgb2018051191-seql.txt

Gly Gln Tyr Lys Pro Ala Pro Ser Ser Ser Phe Ser Thr Ala Val Thr
 35 40 45

Gln Gly Ala Ala Ala Met Leu Val Asn Val Leu Asn Asp Ser Gly Trp
 50 55 60

Phe Ile Pro Leu Glu Arg Glu Gly Leu Gln Asn Ile Leu Thr Glu Arg
 65 70 75 80

Lys Ile Ile Arg Ala Ala Leu Lys Lys Asp Asn Val Pro Val Asn Asn
 85 90 95

Ser Ala Gly Leu Pro Ser Leu Leu Ala Ala Asn Ile Met Leu Glu Gly
 100 105 110

Gly Ile Val Gly Tyr Asp Ser Asn Ile His Thr Gly Gly Ala Gly Ala
 115 120 125

Arg Tyr Phe Gly Ile Gly Ala Ser Glu Lys Tyr Arg Val Asp Glu Val
 130 135 140

Thr Val Asn Leu Arg Ala Ile Asp Ile Arg Thr Gly Arg Ile Leu His
 145 150 155 160

Ser Val Leu Thr Ser Lys Lys Ile Leu Ser Arg Glu Ile Arg Ser Asp
 165 170 175

Val Tyr Arg Phe Ile Glu Phe Lys His Leu Leu Glu Met Glu Ala Gly
 180 185 190

Ile Thr Thr Asn Asp Pro Ala Gln Leu Cys Val Leu Ser Ala Ile Glu
 195 200 205

Ser Ala Val Ala His Leu Ile Val Asp Gly Val Ile Lys Lys Ser Trp
 210 215 220

Ser Leu Ala Asp Pro Asn Glu Leu Asn Ser Pro Val Ile Gln Ala Tyr
 225 230 235 240

pctgb2018051191-seql.txt

Gln Gln Gln Arg Ile
245

<210> 23
<211> 234
<212> PRT
<213> Chryseobacterium oranimense G311

<400> 23

Pro Ser Asp Pro Glu Arg Ser Thr Met Gly Glu Leu Thr Pro Ser Thr
1 5 10 15

Ala Glu Leu Arg Asn Leu Pro Leu Pro Asn Glu Lys Ile Val Ile Gly
20 25 30

Val Tyr Lys Phe Arg Asp Gln Thr Gly Gln Tyr Lys Pro Ser Glu Asn
35 40 45

Gly Asn Asn Trp Ser Thr Ala Val Pro Gln Gly Thr Thr Thr Ile Leu
50 55 60

Ile Lys Ala Leu Glu Asp Ser Arg Trp Phe Ile Pro Ile Glu Arg Glu
65 70 75 80

Asn Ile Ala Asn Leu Leu Asn Glu Arg Gln Ile Ile Arg Ser Thr Arg
85 90 95

Gln Glu Tyr Met Lys Asp Ala Asp Lys Asn Ser Gln Ser Leu Pro Pro
100 105 110

Leu Leu Tyr Ala Gly Ile Leu Leu Glu Gly Gly Val Ile Ser Tyr Asp
115 120 125

Ser Asn Thr Met Thr Gly Gly Phe Gly Ala Arg Tyr Phe Gly Ile Gly
130 135 140

Ala Ser Thr Gln Tyr Arg Gln Asp Arg Ile Thr Ile Tyr Leu Arg Ala
145 150 155 160

pctgb2018051191-seq1.txt

Val Ser Thr Leu Asn Gly Glu Ile Leu Lys Thr Val Tyr Thr Ser Lys
165 170 175

Thr Ile Leu Ser Thr Ser Val Asn Gly Ser Phe Phe Arg Tyr Ile Asp
180 185 190

Thr Glu Arg Leu Leu Glu Ala Glu Val Gly Leu Thr Gln Asn Glu Pro
195 200 205

Val Gln Leu Ala Val Thr Glu Ala Ile Glu Lys Ala Val Arg Ser Leu
210 215 220

Ile Ile Glu Gly Thr Arg Asp Lys Ile Trp
225 230

<210> 24
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> StrepII(C)

<400> 24

Ser Ala Trp Ser His Pro Gln Phe Glu Lys
1 5 10

<210> 25
<211> 861
<212> DNA
<213> Artificial Sequence

<220>
<223> DNA sequence encoding
Pro-CP1-Eco-(WT-Y51A/F56Q/D149N/E185N/E201N/E203N-StrepII(C))

<400> 25
atgcagcgtc tgtttctgct ggtcgcggtg atgctgctga gcggttgctc gaccgcaccg 60
ccgaaagaag cggcacgtcc gaccctgatg ccgcgtgcac agagctataa agatctgacc 120
catctgccgg ctccgacggg caaatcttc gtttctgtct acaacatcca ggacgaaacc 180
ggatcaattta aaccagctcc tgcgtcaaat caatcgactg ccgttccgca gtcagcaacc 240

pctgb2018051191-seq1.txt

gctatgctgg tcacggcact gaaagattcg cgttggttca ttccgctgga acgccagggc 300
ctgcaaaacc tgctgaatga acgtaaaatt atccgcgcag ctcaggaaaa cggtagcctg 360
gccattaaca atcgcacccc gctgcaaagt ctgacggcgg ccaacatcat ggttgaaggc 420
tccattatcg gttatgaaag caatgtcaaa tctggcgggtg tgggcgcacg ttatttcggc 480
attggtgcta ataccagta ccaactggac cagatcgcag ttaacctgcg cgtggttaat 540
gtcagcaccg gcgaaattct gagctctgtg aataccagta aaacgatcct gtcctacaac 600
gtgcaggctg gtgtttttcg tttcattgat tatcaacgcc tgctgaatgg caacgtcggc 660
tacaccagca acgaaccggt gatgctgtgt ctgatgtctg cgattgaaac ggggtgttatt 720
tttctgatca atgatggcat cgaccgtggc ctgtgggatc tgcagaacaa agcggaacgt 780
caaatgaca ttctggtgaa ataccgccac atgtcagttc cgccggaaag ttccgcatgg 840
agccacccgc agttcgaaaa a 861

<210> 26

<211> 287

<212> PRT

<213> Artificial Sequence

<220>

<223> Amino acid sequence of
Pro-CP1-Eco-(WT-Y51A/F56Q/D149N/E185N/E201N/E203N-StrepII(C))

<400> 26

Met Gln Arg Leu Phe Leu Leu Val Ala Val Met Leu Leu Ser Gly Cys
1 5 10 15

Leu Thr Ala Pro Pro Lys Glu Ala Ala Arg Pro Thr Leu Met Pro Arg
20 25 30

Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Ala Pro Thr Gly Lys
35 40 45

Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe Lys
50 55 60

Pro Ala Pro Ala Ser Asn Gln Ser Thr Ala Val Pro Gln Ser Ala Thr
65 70 75 80

pctgb2018051191-seql.txt

Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro Leu
 85 90 95

Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile Arg
 100 105 110

Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg Ile Pro Leu
 115 120 125

Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile Gly
 130 135 140

Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe Gly
 145 150 155 160

Ile Gly Ala Asn Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn Leu
 165 170 175

Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser Val Asn Thr
 180 185 190

Ser Lys Thr Ile Leu Ser Tyr Asn Val Gln Ala Gly Val Phe Arg Phe
 195 200 205

Ile Asp Tyr Gln Arg Leu Leu Asn Gly Asn Val Gly Tyr Thr Ser Asn
 210 215 220

Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val Ile
 225 230 235 240

Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp Leu Gln Asn
 245 250 255

Lys Ala Glu Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg His Met Ser
 260 265 270

Val Pro Pro Glu Ser Ser Ala Trp Ser His Pro Gln Phe Glu Lys
 275 280 285

pctgb2018051191-seq1.txt