



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION
(13) **A1**

(86) **Date de dépôt PCT/PCT Filing Date:** 2018/02/12
(87) **Date publication PCT/PCT Publication Date:** 2018/08/16
(85) **Entrée phase nationale/National Entry:** 2019/08/08
(86) **N° demande PCT/PCT Application No.:** GB 2018/050379
(87) **N° publication PCT/PCT Publication No.:** 2018/146491
(30) **Priorité/Priority:** 2017/02/10 (US62/457,483)

(51) **Cl.Int./Int.Cl. C07K 14/195** (2006.01)
(71) **Demandeur/Applicant:**
OXFORD NANOPORE TECHNOLOGIES LIMITED, GB
(72) **Inventeurs/Inventors:**
JAYASINGHE, LAKMAL, GB;
WALLACE, ELIZABETH JAYNE, GB;
SINGH, PRATIK RAJ, GB
(74) **Agent:** RIDOUT & MAYBEE LLP

(54) **Titre : NANOPORES MODIFIES, COMPOSITIONS LES COMPRENANT ET LEURS UTILISATIONS**
(54) **Title: MODIFIED NANOPORES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF**

(57) **Abrégé/Abstract:**

Provided herein relate to modified or mutant forms of secretin and compositions comprising the same. In particular, the modified or mutant forms of secretin permits efficient capture and/or translocation of an analyte through the modified or mutant secretin nanopores. Methods for using unmodified secretin or the modified or mutant forms of secretin and compositions, for example, for characterizing an analyte, e.g., a target polynucleotide, are also provided.

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
16 August 2018 (16.08.2018)



(10) International Publication Number
WO 2018/146491 A1

(51) International Patent Classification:
C07K 14/195 (2006.01)

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

(21) International Application Number:
PCT/GB2018/050379

Published:
— with international search report (Art. 21(3))
— with sequence listing part of description (Rule 5.2(a))

(22) International Filing Date:
12 February 2018 (12.02.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/457,483 10 February 2017 (10.02.2017) US

(71) Applicant: **OXFORD NANOPORE TECHNOLOGIES LIMITED** [GB/GB]; Gosling Building, Edmund Halley Road, Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB).

(72) Inventors: **JAYASINGHE, Lakmal**; 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 Oxfordshire OX4 4DQ (GB). **SINGH, Pratik Raj**; Oxford Nanopore Technologies Limited, Gosling Building, Edmund Halley Road, Oxford Science Park, Oxford Oxfordshire OX4 4DQ (GB).

(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, 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,

(54) Title: MODIFIED NANOPORES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

(57) Abstract: Provided herein relate to modified or mutant forms of secretin and compositions comprising the same. In particular, the modified or mutant forms of secretin permits efficient capture and/or translocation of an analyte through the modified or mutant secretin nanopores. Methods for using unmodified secretin or the modified or mutant forms of secretin and compositions, for example, for characterizing an analyte, e.g., a target polynucleotide, are also provided.

WO 2018/146491 A1

MODIFIED NANOPORES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

Technical field

5 Provided herein are modified or mutant forms of secretin and compositions comprising the same. Methods for using the modified or mutant forms of secretin and compositions, for example, for characterizing a target analyte, *e.g.*, a target polynucleotide, are also provided. Also provided herein are compositions comprising secretin and an enzyme provided within the secretin lumen.

10

Background

Transmembrane pores (*e.g.*, nanopores) have been used to identify small molecules or folded proteins and to monitor chemical or enzymatic reactions at the single molecule level. The electrophoretic translocation of DNA across nanopores reconstituted into artificial membranes
15 holds great promise for practical applications such as DNA sequencing, and biomarker recognition. However, translocation of double-stranded or single-stranded DNA through nanopores having internal surface facing negatively charged amino acids are not efficient.

Summary

20 The disclosure relates generally to analyte detection using secretins as nanopores. The disclosure generally relates to modified nanopores. In some embodiments, the disclosure provides modified secretin nanopores and subunit polypeptides, compositions or apparatuses comprising the same, and uses thereof. In some embodiments, modified secretin nanopores provided herein are useful for analyte detection and analysis because they promote efficient
25 capture and/or translocation of an analyte, *e.g.*, a negatively-charged or hydrophobic biopolymer such as a polynucleotide or protein, across the nanopores. Accordingly, secretin nanopores, *e.g.*, modified secretin nanopores as described herein can be used for characterizing an analyte, *e.g.*, a target polynucleotide or polypeptide, and other suitable applications. Accordingly, in further embodiments, described herein are methods and compositions for characterizing an analyte, *e.g.*,
30 a target polynucleotide or polypeptide.

One aspect of the present disclosure features a modified secretin nanopore, for example, disposed in a membrane. The modified secretin nanopore comprises a luminal surface defining a lumen that extends through the membrane between a *cis*-opening and a *trans*-opening, wherein the luminal surface comprises one or more amino acid modifications. Examples of the amino
35 acid modifications include, but are not limited to charge-altering modifications (*e.g.*,

substitutions of negatively-charged amino acids with positively-charged amino acids), amino acid modifications that change its hydrophobicity (*e.g.*, substitutions of neutral amino acids with hydrophobic amino acids), amino acid modifications that change the size of an opening, *e.g.* a constriction or gate, in the secretin (*e.g.* substitution of one or more amino acid having a smaller or larger side group than the naturally occurring amino acid(s), or deletion of one or more amino acids that constrict an opening), amino acid modifications that inhibit or prevent gate opening (such as substitution of one or more flexible amino acid with more rigid amino acid(s)), and a combination thereof.

The *cis*-opening and *trans*-opening of the modified secretin nanopores may have a diameter of any size that suits the need of an application (*e.g.*, detection and/or analysis of an analyte such as a target polynucleotide). In some embodiments, the *cis*-opening of the modified secretin nanopores may have a diameter in a range of 60 Å to 120 Å. In some embodiments, the *trans*-opening of the modified secretin nanopores may have a diameter in a range of 40 Å to 100 Å. In some embodiments, the constriction of the modified secretin nanopores may have a diameter of about 7.5 Å to 25 Å.

Any types of secretin may be used to produce the modified secretin nanopores described herein. For example, in some embodiments, the secretin may be of a type II secretion system (*e.g.*, but not limited to GspD). In some embodiments, the secretin may be of a type III secretion system (*e.g.*, but not limited to YscC and InvG). In some embodiments, the secretin may be of a type IV secretion system (*e.g.*, but not limited to PilQ).

In some embodiments where the secretin is an InvG, the modified secretin nanopore may further comprise a subunit polypeptide having an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 1 (corresponding to the amino acid sequence of InvG without N1 or N0 domain). In these embodiments, the luminal surface may further define a constriction within the lumen, the constriction having one or more amino acid modifications (*e.g.*, charge-altering modifications) at amino acids D28, E225, R226, and/or E231 of SEQ ID NO: 1. Examples of such amino acid modifications include but are not limited to (i) D28N/Q/T/S/G/R/K; (ii) E225N/Q/T/A/S/G/P/H/F/Y/R/K; (iii) R226N/Q/T/A/S/G/P/H/F/Y/K/V; (iv) Deletion of E225; (v) Deletion of R226; and (vi) E231N/Q/T/A/S/G/P/H/R/K. In some embodiments, the modified secretin nanopore, the luminal surface may comprise a capture portion having one or more amino acid modifications at amino acids E41, Q45 or E114, examples of which include, but are not limited to (i) Q45R/K; (ii) E41N/Q/T/S/G/R/K; and (iii) E114N/Q/T/S/G/R/K.

The modified secretin nanopore can be homo-multimeric (*e.g.*, all subunits within the nanopore are the same) or hetero-multimeric (*e.g.*, at least one subunit is different from others

within the nanopore). The modified secretin nanopore may comprise any number of subunit polypeptides that are sufficient to form a lumen large enough to permit a target analyte (*e.g.*, polynucleotide) to pass through. In some embodiments, the modified secretin nanopore may comprise 9-20 subunit polypeptides, wherein at least one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or up to all) of the subunit polypeptides comprises one or more of the amino acid modifications as described herein.

Accordingly, modified secretin nanopore subunit polypeptide and polynucleotides comprising nucleotide sequences encoding the modified secretin nanopore subunit polypeptides are also provided herein.

For example, in one aspect the modified GspD secretin nanopore comprises a subunit polypeptide comprising a secretin domain having an amino acid sequence that is at least 95% identical to the amino acid sequence of the secretin domain set forth in SEQ ID NO: 36.

The secretin domain of GspD from *Vibrio cholerae* and from *Escherichia coli* ETEC contains a cap gate. Other Type II secretion system secretin subunit polypeptides, including some GspD subunit polypeptides, such as *Escherichia coli* K12, do not comprise a cap gate. The modified secretin nanopore may, in one aspect be one that does not comprise a cap gate. The secretin domain set in out SEQ ID NO: 36 comprises a cap gate between positions 56 and 77. For example, the secretin domain set forth in SEQ ID NO: 36 may be modified to delete all or part of the cap gate, *e.g.* all or some of the amino acids from D55 or T56 to T77 of SEQ ID NO: 36 may be deleted or substituted. Alternatively, the modified GspD secretin nanopore may naturally lack a cap gate. The amino acids from D55 or T56 to T77 of SEQ ID NO: 36 correspond to the amino acids from D371 or T372 to T393 of SEQ ID NO: 32.

The central gate of GspD may be modified to replace an amino acid with an amino acid having a smaller side group and/or to replace a negatively charged amino acid with a neutral or positively charged amino acid. The secretin domain set in out SEQ ID NO: 36 comprises a central gate between positions 144 to 157, which correspond to positions 460 and 473 of SEQ ID NO: 32. The secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least 95% identical to an amino acid sequence as set forth in SEQ ID NO: 36, wherein: (i) all or some of the amino acids from D55 or T56 to T77 are deleted or substituted, one or more of K60, D64, R71 and E73 is substituted with an uncharged amino acid and/or one or more of D55, T56, T77 and K78 is substituted with P; and/or (ii) F156 is substituted with a smaller amino acid, N151 and/or N152 is/are substituted with a smaller amino acid, D153 is substituted with an uncharged amino acid, G137 and G165 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y63 to R71 may be deleted and/or substituted with GSG or SGS, F156

may be substituted with A, D153 may be substituted with S, and/or N151 and N152 may each independently be substituted with G or S. D55, T56, K60, Y63, D64, R71, E73, T77, K78, G137, N151, N152, D153, F156 and G165 of SEQ ID NO: 36 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32. The modified secretin GspD nanopore may in one aspect comprise a subunit polypeptide comprising an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 33, 34 and/or 35.

The secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least 95% identical to an amino acid sequence as set forth in SEQ ID NO: 35, wherein: (i) all or some of the amino acids from D55 or T56 to T77 are deleted or substituted, one or more of K60, D64, R71 and E73 is substituted with an uncharged amino acid and/or one or more of D55, T56, T77 and K78 is substituted with P; and/or (ii) F156 is substituted with a smaller amino acid, N151 and/or N152 is/are substituted with a smaller amino acid, D153 is substituted with an uncharged amino acid, G137 and G165 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y63 to R71 may be deleted and/or substituted with GSG or SGS, F156 may be substituted with A, D153 may be substituted with S, and/or N151 and N152 may each independently be substituted with G or S. D55, T56, K60, Y63, D64, R71, E73, T77, K78, G137, N151, N152, D153, F156 and G165 of SEQ ID NO: 35 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32.

The secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least 95% identical to an amino acid sequence as set forth in SEQ ID NO: 34, wherein: (i) all or some of the amino acids from D117 or T118 to T139 are deleted or substituted, one or more of K122, D126, R133 and E135 is substituted with an uncharged amino acid and/or one or more of D117, T118, T139 and K140 is substituted with P; and/or (ii) F218 is substituted with a smaller amino acid, N213 and/or N214 is/are substituted with a smaller amino acid, D215 is substituted with an uncharged amino acid, G199 and G227 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y125 to R133 may be deleted and/or substituted with GSG or SGS, F218 may be substituted with A, D215 may be substituted with S, and/or N213 and N214 may each independently be substituted with G or S. D117, T118, K122, Y125, D126, R133, E135, T139, K140, G199, N213, N214, D215, F218 and G227 of SEQ ID NO: 34 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the

full length GspD amino acid sequence set forth in SEQ ID NO: 32.

The secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least 95% identical to an amino acid sequence as set forth in SEQ ID NO: 33, wherein: (i) all or some of the amino acids from D132 or T133 to T154 are deleted or substituted, one or more of K137, D141, R148 and E150 is substituted with an uncharged amino acid and/or one or more of D132, T133, T154 and K155 is substituted with P; and/or (ii) F233 is substituted with a smaller amino acid, N228 and/or N229 is/are substituted with a smaller amino acid, D230 is substituted with an uncharged amino acid, G214 and G242 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y140 to R148 may be deleted and/or substituted with GSG or SGS, F233 may be substituted with A, D230 may be substituted with S, and/or N228 and N229 may each independently be substituted with G or S. D132, T133, K137, Y140, D141, R148, E150, T154, K155, G214, N228, N229, D230, F233 and G242 of SEQ ID NO: 33 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32. For example, in one aspect, a modified InvG nanopore subunit polypeptide comprises an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 1 (corresponding to the amino acid sequence of InvG without N1 or N0 domain), wherein the modified InvG nanopore subunit polypeptide comprises one or more amino acid modifications (*e.g.*, charge-altering amino acid modifications) at amino acid(s) selected from D28, E41, E114, Q45, E225, R226, and E231 of SEQ ID NO: 1. The one or more amino acid modifications (*e.g.*, charge-altering amino acid modifications) may comprise one or more of the following: (i) D28N/Q/T/S/G/R/K; (ii) E225N/Q/T/A/S/G/P/H/F/Y/R/K; (iii) R226N/Q/T/A/S/G/P/H/F/Y/K/V; (iv) Deletion of E225; (v) Deletion of R226; and (vi) E231N/Q/T/A/S/G/P/H/R/K. Other amino acid modifications may include, but are not limited to (i) Q45R/K; (ii) E41N/Q/T/S/G/R/K; and/or (iii) E114N/Q/T/S/G/R/K. Such amino acid modifications may enhance capture of an analyte, *e.g.* a polynucleotide, by the nanopore (*e.g.* mutations at D28, E41, E114 and/or Q45) and/or improve the interaction of an analyte, *e.g.* a polynucleotide, with the constriction of the nanopore (*e.g.* mutations at E225 and/or R226). In another aspect, a modified InvG nanopore subunit polypeptide comprises an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 2 (corresponding to the amino acid sequence of WT InvG including N1 and N0 domains), wherein the modified InvG nanopore subunit polypeptide comprises one or more amino acid modifications (*e.g.*, charge-altering amino acid modifications) at amino acid(s) selected from D199, E212, E285, Q216, E396, R397, and E402 of SEQ ID NO: 2. Non-limiting examples of such amino acid modifications include: (i) D199N/Q/T/S/G/R/K;

(ii) E396N/Q/T/A/S/G/P/H/F/Y/R/K; (iii) R397N/Q/T/A/S/G/P/H/F/Y/K/V; (iv) Deletion of E396; (v) Deletion of R397; (vi) E402N/Q/T/A/S/G/P/H/R/K. Other amino acid modifications may include, but are not limited to (i) Q216R/K; (ii) E212N/Q/T/S/G/R/K; and (iii) E285N/Q/T/S/G/R/K. Such amino acid modifications may enhance capture of an analyte, e.g. a polynucleotide, by the nanopore (e.g. mutations at D199, E212, E285 and/or Q216) and/or improve the interaction of an analyte, e.g. a polynucleotide, with the constriction of the nanopore (e.g. mutations at E396 and/or R397).

A further aspect features a modified InvG nanopore subunit polypeptide that comprises an endopeptidase cleavage site. In this aspect, the modified InvG nanopore subunit polypeptide comprises an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 2 (corresponding to the amino acid sequence of WT InvG including N1 and N0 domains), wherein an endopeptidase cleavage site is inserted between positions 170 and 171 or 171 and 172 of SEQ ID NO: 2. In some embodiments, the modified InvG nanopore subunit polypeptide may further comprise one or more amino acid modifications (e.g., charge-altering amino acid modifications) at amino acid(s) selected from D199, E212, E285, Q216, E396, R397, and E402 of SEQ ID NO: 2. Non-limiting examples of such amino acid modifications include: (i) D199N/Q/T/S/G/R/K; (ii) E396N/Q/T/A/S/G/P/H/F/Y/R/K; (iii) R397N/Q/T/A/S/G/P/H/F/Y/K/V; (iv) Deletion of E396; (v) Deletion of R397; (vi) E402N/Q/T/A/S/G/P/H/R/K. Other amino acid modifications may include, but are not limited to (i) Q216R/K; (ii) E212N/Q/T/S/G/R/K; and (iii) E285N/Q/T/S/G/R/K.

A further aspect of the present disclosure provides a composition comprising a secretin nanopore and an enzyme provided within the lumen of the nanopore. The composition may be disposed within a membrane.

Also within the scope of the present disclosure are apparatuses, for example, for use in characterizing a target analyte, e.g., a target polynucleotide. The apparatus may comprise a chamber housing an aqueous solution having disposed therein a membrane comprising any embodiment of the secretin nanopores described herein.

In some embodiments, the apparatus may further comprise an analyte present in the aqueous solution. Exemplary analytes include, but are not limited to polynucleotides, polypeptides, and/or ligands. In some embodiments where the apparatus comprises a polynucleotide in the aqueous solution, the apparatus can further comprise a polynucleotide binding protein, including, e.g., but not limited to a helicase, exonuclease, or polymerase, which is optionally bound to the polynucleotide. The polynucleotide binding protein may be on the *cis*-side or *trans*-side of the membrane, for example, being in contact (via, e.g., ionic and/or hydrophobic interactions) with or covalently attached to the *cis*-opening or *trans*-opening of the

nanopore.

The modified secretin nanopores and apparatuses as described herein can be used for various biosensor or analyte detection applications, but not limited to polynucleotide sequencing and/or protein detection. Accordingly, methods for using the modified secretin nanopores and apparatuses are also provided herein. For example, the method comprises obtaining an embodiment of the apparatus as described herein and adding an analyte to the aqueous solution on the *cis*-side or the *trans*-side of the membrane disposed in the apparatus. In some embodiments, the method further comprises inducing ionic current flow through the nanopore by applying a voltage gradient across the membrane. In some embodiments, the method further comprises detecting ionic current flow through the nanopore under the applied voltage gradient, which can be used to determine the presence of the analyte.

Where the method is used for polynucleotide characterization, the method can further comprise adding a polynucleotide binding protein (*e.g.*, a helicase, exonuclease, and/or polymerase) in the aqueous solution on the *cis*-side or the *trans*-side of the membrane. In some embodiments, the polynucleotide binding protein may be bound to the polynucleotide analyte and optionally interact with the *cis*-opening or *trans*-opening of the nanopore via, for example non-covalent interactions (*e.g.*, ionic and/or hydrophobic interactions) and/or covalent attachment.

The details of one or more embodiments of the disclosure are set forth in the description below. Other features or advantages of the present disclosure will be apparent from the following drawings and detailed description of several embodiments, and also from the appended claims.

Brief Description of the Drawings

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present disclosure, which can be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

Figure 1A shows Cry-EM structures of the injectisome basal body and isolated secretin. (Left panel) Central slice view of basal body reconstruction (dark-grey contoured as in a and light-grey contoured at lower level to highlight less-ordered features) and isolated secretin (blue). The domain annotation of PrgH, PrgK and InvG is overlaid on the left and the structures of the monomeric domains previously solved on the right. The PrgH cytoplasmic D1 domain (green, bottom left) is not present in the PrgH₁₃₀₋₃₉₂ mutant used in this study and its precise location with respect to the basal body is unclear. The transmembrane helices of PrgH (N-terminal) and

PrgK (C-terminal) and the PrgK N-terminal lipidation are present but diffusely ordered. (Right panel) Refined structures for InvG₁₇₂₋₅₅₇ (blue), PrgH₁₇₁₋₃₆₄ (green), PrgK₂₀₋₂₀₃ (orange) and Rosetta-modeled InvG₃₄₋₁₇₁ (pale blue). One monomer encompassing InvG₃₄₋₅₅₇ is colored according to structural domains: medium blue, N0-N1 domains; cobalt blue, N3 domain; cyan, outer β -sheet; green, inner β -sheet; orange, secretin domain lip; red, S domain (note the displaced interaction with the β -sheet of the $i + 1$ and $i + 2$ promoters).

Figure 1B shows secondary structure topology of a wild-type InvG₁₇₂₋₅₅₇ secretin. β -strands of the secretin domain are numbered, with 1, 3a/3b, 8 and 9 forming the outer β -barrel; 4-7 forming the inner β -barrel; and 1, 2 and 3a forming the lip of the β -barrel. Strand 3 is broken into 3a and 3b by the conserved residue Pro371. The numerical values indicated at both ends of each domain define the first and last amino acid positions of the domain based on SEQ ID NO: 2.

Figure 1C shows secondary structure topologies of a wild-type InvG secretin from *Salmonella enterica* (e.g., based on SEQ ID NO: 2) and a wild-type GspD secretin from *Vibrio cholerae* from positions 97-646 of SEQ ID NO: 10. The figure shows different domains and dimensions of the *cis* and *trans* openings of the InvG nanopore and GspD nanopore. The orientation of the nanopores is such that the OM region of the nanopores (as in the native state) is situated in the membrane as described herein.

Figure 1D shows structures of GspD from *Vibrio cholerae* (PDB: 5wq8) and *E. coli* (PDB: 5wq7). One subunit of each GspD structure is colored in cyan.

Figure 2 shows a comparison of a CsgG nanopore with an InvG nanopore. The top row shows the top view of CsgG and InvG nanopores, while the bottom row shows the side view of CsgG and InvG nanopores. A CsgG nanopore has 9 monomers or subunits and an InvG nanopore has 15 monomers or subunits. However, both CsgG and InvG nanopores have a constriction within the lumen that is roughly the same in diameter.

Figure 3 shows the InvG and CsgG nanopore profiles. The X axis shows the internal pore radius profiles of InvG and CsgG nanopores: -60 (membrane side/*trans* opening) and +60 (*cis* opening) are arbitrary numbers for the height of the pore. 0 is the mid-point. The Y axis shows the actual radius of the lumen of the pore in angstrom for each position of the X axis.

Figure 4 shows a comparison of the constrictions of CsgG and InvG nanopores. The top row shows the side view of the CsgG and InvG nanopores. The bottom row shows the amino acids present within the constriction of CsgG and InvG pores. While both CsgG and InvG nanopores have a constriction of roughly the same in diameter, the constriction of the CsgG nanopore has 3 amino acids at positions 51, 55, and 56 (based on the wild type sequence) and the

InvG nanopore constriction has two amino acids at position 396 and 397 (based on SEQ ID NO: 2).

Figure 5 shows the relative size of a polynucleotide binding protein (*e.g.*, a DNA binding enzyme such as a helicase or polymerase) versus CsgG and InvG nanopores. Since the opening of the InvG nanopore is much wider than that of the CsgG nanopore, the polynucleotide binding protein (*e.g.*, a DNA binding enzyme such as a helicase or polymerase) may interact with the InvG and CsgG nanopores in different orientations.

Figure 6 shows the top views (from different perspectives) of a polynucleotide binding protein (*e.g.*, a DNA binding enzyme such as a helicase or polymerase) interacting with an InvG nanopore. In the left panel, the inner dotted line corresponds to the lower dotted line in the InvG (right panel) of Figure 4 and the outer dotted line corresponds to the upper dotted line in the InvG (right panel) of Figure 4.

Figure 7 shows exemplary combinations of mutations in InvG subunit polypeptide that can be used to form a nanopore. The amino acid positions indicated in the figure are based on SEQ ID NO: 2.

Figure 8 shows the relative positions of the mutations as shown in Figure 7 in an InvG nanopore. While the nanopore does not have N0 or N1 domains, the amino acid positions indicated in the figure are based on SEQ ID NO: 2.

Figure 9 shows the structural homology between GspD and InvG secretin nanopores.

Figure 10 shows the amino acid sequences of GspD from *Vibrio cholerae* and highlights the regions of amino acid sequence that are missing from the crystal structure, i.e. for which the crystal structure has not been determined in the art. The amino acid positions indicated in the figure are based on SEQ ID NO: 32.

Figure 11 shows the domain structure of GspD from *Vibrio cholerae*. The amino acid positions indicated in the figure are based on SEQ ID NO: 32.

Figure 12 shows the structure of GspD from *Vibrio cholerae* and highlights the positions of the N3 constriction site and the cap and central gates. The amino acid positions indicated in the figure are based on SEQ ID NO: 32.

Figure 13 shows the kink formed by G453 and G481 in the amino acid sequence of GspD from *Vibrio cholerae*.

Figure 14 shows electrophysiological characteristics of the GspD-Vch-(WT-del(1-239)/(265-282)-H6(C) mutant which was used as a baseline. A) Open pore current at -180mV in 500mM KCl, 25mM Phosphate buffer, pH 7. B) IV curve ranging from -25mV to -200mV and 25mV to 200mV in 25mV alternating potential steps.

Figure 15 shows IV characteristics of different GspD mutants ranging from -25mV to -200mV and 25mV to 200mV in 25mV alternating potential steps. A) GspD-Vch-(WT-del(1-239)/(265-282)-H6(C). B) GspD-Vch-(WT-Del((N1-K239)/(N265-SGS-E282)/(Y379-GSG-R387)). C) GspD-Vch-(WT-F472A-Del((N1-K239)/(N265-SGS-E282))). D) GspD-Vch-(WT-D469S-Del((N1-K239)/(N265-SGS-E282))). E) GspD-Vch-(WT-N467G/N468S-Del((N1-K239)/(N265-SGS-E282))). F) GspD-Vch-(WT-N467S/N468G-Del((N1-K239)/(N265-SGS-E282))). G) GspD-Vch-(WT-N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282))).

Figure 16 shows DNA translocation through the GspD-Vch-(WT-del(1-239)/(265-282)-H6(C) mutant. A) Open pore current at -180mV in 470mM KCL, 25mM HEPES, 11mM ATP and 10mM MgCl₂, pH8.0. B) Addition of Lambda 3.6kb DNA ligated to adapter shows clear noisy patterns in the current trace. There is an increase in current spikes when DNA is inside the pore. C) Zoomed in image of the noisy pattern show a drop in open pore current which is the DNA translocating through the pore.

Figure 17 is a model of biotinylated static strands bound monovalent streptavidin inside the GspD pore. A) Streptavidin molecule in top of the pore. B) Streptavidin molecule inside the pore above the constriction gate.

Figure 18 shows the capture of streptavidin bound biotinylated static strands by the GspD-Vch-(WT-N467G/N468S-Del((N1-K239)/(N265-SGS-E282))). A) Static strands experiment run for 1 hour in single GspD pore starting with control open pore experiment for 15 minutes and flushing three static strands, ONLA19798, AH71 and AH72 respectively after 15 minutes through the chip. B) Open pore control trace with current around 250pA. C) Addition of ONLA19798 shows the capture of static strand from the open pore instantly. D) Addition of AH71 shows the capture of static strand from the open pore. E) Addition of AH72 also shows the capture of static strand.

Brief Description of the Sequences

SEQ ID NO: 1 is the 391 amino acid sequence of truncated InvG from *Salmonella enterica* (Full length InvG without N0 and N1 domains).

SEQ ID NO: 2 is the 572 amino acid sequence of Wild-type InvG from *Salmonella enterica* (Full length InvG including N0 and N1 domains). The first 171 amino acids correspond to the N0 and N1 domains.

SEQ ID NO: 3 is the amino acid sequence of wild-type InvG from *Salmonella enterica* in which a TEV cleavage site (ENLYFQG) has been added at amino acids 172 to 178 after the N1 and N2 domains (the first 171 amino acids).

SEQ ID NO: 4 is the amino acid sequence of GspD from *Escherichia coli* (strain K12) (>sp|P45758|GSPD_ECOLI type II secretion system protein D OS=*Escherichia coli* (strain K12) GN=gspD PE=2 SV=2).

5 SEQ ID NO: 5 is the amino acid sequence of >tr|Q7BRZ9|Q7BRZ9_YEREN Secretin YscC OS=*Yersinia enterocolitica* GN=yscC PE=3 SV=1.

SEQ ID NO: 6 is the amino acid sequence of >sp|Q04641|MXID_SHIFL Outer membrane protein MxiD OS=*Shigella flexneri* GN=mxid PE=1 SV=1.

10 SEQ ID NO: 7 is the amino acid sequence of >tr|A0A1C6ZHG5|A0A1C6ZHG5_PSEAI Type III secretion outer membrane protein PscC OS=*Pseudomonas aeruginosa* GN=pscC PE=3 SV=1.

SEQ ID NO: 8 is the amino acid sequence of >tr|B7UMB3|B7UMB3_ECO27 T3SS structure protein EscC OS=*Escherichia coli* O127:H6 (strain E2348/69 / EPEC) GN=escC PE=1 SV=1.

15 SEQ ID NO: 9 is the amino acid sequence of >sp|D0ZWR9|SPIA_SALT1 Type III secretion system outer membrane protein SpiA OS=*Salmonella typhimurium* (strain 14028s / SGSC 2262) GN=spiA PE=2 SV=1.

SEQ ID NO: 10 is the amino acid sequence of >tr|A0A1E4UJH6|A0A1E4UJH6_VIBCL Type II secretion system protein GspD OS=*Vibrio cholerae* GN=BFX10_13405 PE=4 SV=1.

20 SEQ ID NO: 11 is the amino acid sequence of >sp|P15644|GSPD_KLEPN Type II secretion system protein D OS=*Klebsiella pneumoniae* GN=pulD PE=1 SV=1.

SEQ ID NO: 12 is the amino acid sequence of >tr|X5F782|X5F782_NEIME Type IV pilus assembly protein PilQ OS=*Neisseria meningitidis* GN=pilQ PE=3 SV=1.

25 SEQ ID NO: 13 is the amino acid sequence of >WP_071651540.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Salmonella enterica*] - 97% identity to SEQ ID NO: 2.

SEQ ID NO: 14 is the amino acid sequence of >WP_038392434.1 type III secretion system outer membrane pore InvG [*Salmonella bongori*] - 94% identity to SEQ ID NO: 2.

30 SEQ ID NO: 15 is the amino acid sequence of >WP_043640872.1 type III secretion system outer membrane pore InvG [*Chromobacterium haemolyticum*] - 69% identity to SEQ ID NO: 2.

SEQ ID NO: 16 is the amino acid sequence of >WP_059765897.1 type III secretion system outer membrane pore InvG [*Burkholderia ubonensis*] - 67% identity to SEQ ID NO: 2.

SEQ ID NO: 17 is the amino acid sequence of >WP_036979259.1 type III secretion system outer membrane pore InvG [*Providencia alcalifaciens*] - 64% identity to SEQ ID NO: 2.

SEQ ID NO: 18 is the amino acid sequence of >WP_051238518.1 type III secretion system outer membrane pore InvG [*Pseudogulbenkiania ferrooxidans*] - 61% identity to SEQ ID NO: 2.

5 SEQ ID NO: 19 is the amino acid sequence of >WP_070981539.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Chromobacterium vaccinii*] - 61% identity to SEQ ID NO: 2.

SEQ ID NO: 20 is the amino acid sequence of >WP_052429256.1 type III secretion system outer membrane pore InvG [*Salmonella enterica*] - 60% identity to SEQ ID NO: 2.

10 SEQ ID NO 21 is the amino acid sequence of >WP_021564153.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Escherichia coli*] - 59% identity to SEQ ID NO: 2.

SEQ ID NO: 22 is the amino acid sequence of >WP_024250244.1 type III secretion system outer membrane pore InvG [*Shigella dysenteriae*] - 56% identity to SEQ ID NO: 2.

15 SEQ ID NO: 23 is the amino acid sequence of >WP_000694679.1 type III secretion system outer membrane pore InvG [*Escherichia coli*] - 53% identity to SEQ ID NO: 2.

SEQ ID NO: 24 is the amino acid sequence of >WP_061203566.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Stenotrophomonas rhizophila*] - 52% identity to SEQ ID NO: 2.

20 SEQ ID NO: 25 is the amino acid sequence of >WP_016498773.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Pseudomonas putida*] - 52% identity to SEQ ID NO: 2.

SEQ ID NO: 26 is the amino acid sequence of >ANI31722.1 type III secretion system outer membrane pore InvG [*Yersinia entomophaga*] - 50% identity to SEQ ID NO: 2.

25 SEQ ID NO: 27 is the amino acid sequence of >WP_053215251.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Yersinia nurmii*] - 49% identity to SEQ ID NO: 2.

SEQ ID NO: 28 is the amino acid sequence of >WP_034249407.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Arsenophonus nasoniae*] - 46% identity to SEQ ID NO: 2.

30 SEQ ID NO: 29 is the amino acid sequence of >WP_006122201.1 EscC/YscC/HrcC family type III secretion system outer membrane ring protein [*Pantoea stewartii*] - 42% identity to SEQ ID NO: 2.

SEQ ID NO: 30 is the amino acid sequence of >KJO55878.1 type III secretion system protein [[*Enterobacter*] *aerogenes*] - 41% identity to SEQ ID NO: 2.

SEQ ID NO: 31 is the amino acid sequence of GspD of *Vibrio cholerae*, including the leader sequence.

SEQ ID NO: 32 is the mature amino acid sequence of GspD of *Vibrio cholerae*.

SEQ ID NO: 33 is the sequence of the N3, secretin and S domains of GspD of *Vibrio cholerae* (amino acids 1 to 239 of SEQ ID NO: 32 deleted).

SEQ ID NO: 34 is the sequence of the N3, secretin and S domains of GspD of *Vibrio cholerae* in which the construction in the N3 domain has been removed by substituting amino acids Y379 to R387 of SEQ ID NO: 32 with the amino acids GSG.

SEQ ID NO: 35 is the sequence of the secretin and S domains of GspD of *Vibrio cholerae*.

SEQ ID NO: 36 is the sequence of the secretin domain of GspD of *Vibrio cholerae*.

SEQ ID NO: 37 is the sequence of >tr|A7ZRJ5|A7ZRJ5_ECO24 General secretion pathway protein D OS=*Escherichia coli* O139:H28 (strain E24377A / ETEC) GN=gspD PE=1 SV=1.

SEQ ID NO: 38 is the sequence of >sp|P31780|GSPD_AERHY Type II secretion system protein D OS=*Aeromonas hydrophila* GN=exeD PE=3 SV=2.

SEQ ID NO: 39 is the sequence of >sp|P35818|GSPD_PSEAE Type II secretion system protein D OS=*Pseudomonas aeruginosa* (strain ATCC 15692 / DSM 22644 / CIP 104116 / JCM 14847 / LMG 12228 / 1C / PRS 101 / PAO1) GN=xcpQ PE=1 SV=1.

SEQ ID NO: 40 is the sequence of >tr|A0A181X688|A0A181X688_KLEOX General secretion pathway protein D OS=*Klebsiella oxytoca* GN=pulD PE=3 SV=1.

Detailed Description of the Invention

Certain transmembrane pores (*e.g.*, protein nanopores or solid state nanopores) are useful as sensors to detect or characterize a biopolymer. The structure of the transmembrane pore, particularly the lumen of the pore, affects the interaction between the biopolymer and the pore and hence the information that can be derived from a signal generated as the biopolymer interacts with the pore. Accordingly, there is a need to identify new transmembrane nanopores that are capable of capturing and translocating an analyte, *e.g.*, a negatively-charged or hydrophobic biopolymer such as a polynucleotide or protein. The present disclosure provides, for the first time, that secretin nanopores are useful for practical applications such as polynucleotide mapping or sequencing, or protein detection.

While transmembrane pores (*e.g.*, protein nanopores or solid state nanopores) are useful as sensors to detect or characterize a biopolymer, translocation of a biopolymer, *e.g.*, a polynucleotide through certain nanopores could be challenging, *e.g.*, because of a large

electrostatic barrier for the entry of a biopolymer into the nanopore. Accordingly, there is a need to engineer transmembrane nanopores that permit more efficient capture and/or translocation of an analyte, *e.g.*, a negatively-charged or hydrophobic biopolymer such as a polynucleotide or protein, across the nanopores, which can be useful for practical applications such as

5 polynucleotide mapping or sequencing or protein detection.

The present disclosure relates to modified secretin nanopores and its subunit polypeptides, compositions or apparatuses comprising the same, and uses thereof. In some aspects, the present disclosure provides modified secretin nanopore subunit polypeptide (*e.g.*, for forming a modified secretin nanopore) and nanopores comprising the same. The secretin

10 nanopores and modified secretin nanopores as described herein can be used for various practical applications such as characterizing an analyte, *e.g.*, a target polynucleotide or polypeptide. Accordingly, described herein are also methods and compositions for characterizing an analyte, *e.g.*, a target polynucleotide or polypeptide.

In some embodiments of any aspects described herein, the *cis* and *trans* openings of the

15 secretin nanopores are of a size such that an enzyme may be able to enter the lumenal cavity. The enzyme may be immobilized within the cavity, for example, by binding or attaching to the lumenal surface of the nanopore or otherwise provided within the lumenal cavity in a non-immobilized fashion. Thus, one aspect of the present disclosure also relates to compositions comprising a secretin nanopore and an enzyme provided within the lumen. The secretin nanopore

20 may be of the wild type or a mutant or modified form as described in more detail below. The enzyme may be present in the *cis* vestibule or the *trans* vestibule of the nanopore, wherein the *cis* vestibule may be defined as the part of the lumen extending from the *cis* opening to the constriction of the nanopore and wherein the *trans* vestibule may be defined as the part of the lumen extending from the *trans* opening to the constriction of the nanopore. Such compositions

25 may be used to detect small molecules that bind to or otherwise interact with the enzyme. The interaction of such small molecules with the enzyme may result in a change in ion current flow through the nanopore, for example by a change of conformation of the enzyme.

Modified secretin nanopore subunit polypeptides

30 Some aspects of the present disclosure provide modified secretin nanopore subunit polypeptides. A modified secretin nanopore subunit polypeptide is a polypeptide whose sequence varies from that of a reference secretin amino acid sequence. The amino acid sequence of the modified secretin nanopore subunit polypeptide comprises (i) a *cis* opening-forming amino acid sequence, (ii) a lumen-forming amino acid sequence, and (iii) a *trans* opening-forming

35 amino acid sequence. The *cis* opening-forming amino acid sequence is one or more portions of

the amino acid sequence that forms part of a *cis* opening of a nanopore when the modified secretin nanopore subunit polypeptide interacts with other subunit polypeptides to form the nanopore in a membrane. The lumen-forming amino acid sequence is one or more portions of the amino acid sequence that forms part of a lumen of the nanopore when the modified secretin nanopore subunit polypeptides interacts with other subunit polypeptides to form the nanopore in a membrane. The *trans* opening-forming amino acid sequence is one or more portions of the amino acid sequence that forms part of a *trans* opening of a nanopore when the modified secretin nanopore subunit polypeptide interacts with other subunit polypeptides to form the nanopore in a membrane. Methods to identify portions of the secretin amino acid sequence that form the *cis* opening, lumen, and *trans* opening of a secretin nanopore are known in the art. For example, a nanopore, a portion of which is embedded into a membrane can be constructed by homology modelling from a known secretin structure using VMD, *e.g.*, as described in Humphrey et al., "VMD: Visual Molecular Dynamics" J. Mol. Graphics (1996) 14: 33-38; and NAMD, *e.g.*, as described in Phillips et al., "Scalable Molecular Dynamics with NAMD" J. Comput. Chem. (2005) 26: 1781-1802. See, *e.g.*, FIG. 1D shows structures of GspD from *Vibrio cholerae* (PDB: 5wq8) and *E. coli* (PDB: 5wq7); and FIG. 8 shows a structure of an InvG nanopore and its different protein domains as well as the corresponding positions of example amino acid modifications within the lumen of the nanopore.

As used herein, the term "reference secretin amino acid sequence" refers to a known amino acid sequence of a secretin nanopore subunit. Various forms of secretin nanopore subunits are known in the art, including, *e.g.*, but not limited to any secretin subunit of a type II, type III, or type IV secretion system. Non-limiting examples of a type II secretion system include GspD, PulD, and pIV. Examples of a type III secretion system include, but are not limited to InvG, MxiD, YscC, PscC, EscC, and SpiA. Non-limiting examples of a type IV secretion system include PilQ. A reference secretin amino acid sequence can be a known amino acid sequence of a member of a type II, type III, or type IV secretion system or a portion thereof. For example, a reference secretin amino acid sequence may be an amino acid sequence corresponding to at least a portion of wild type GspD, PulD, pIV, PilQ, InvG, MxiD, YscC, PscC, EscC, SpiA, ExeD or XcpQ wherein the portion comprises one or more of a secretin domain, a S domain, a N2 domain, a N3 domain and/or another related domain. For example, in some embodiments, the portion may comprise a secretin domain, a S domain, and a N3 domain. In some embodiments, the portion may comprise a secretin domain, a S domain, a N3 domain, and a N2 domain. In some embodiments, the portion may comprise a secretin domain and a S domain. Different domains of secretin nanopores are known in the art. For example, FIG. 1C shows different domains of an InvG from *Salmonella enterica* and GspD from *Vibrio cholerae*. In some

embodiments, a reference secretin amino acid sequence may be an amino acid sequence corresponding to a full-length wild type GspD (*e.g.*, as set forth in SEQ ID NO: 4, SEQ ID NO: 10, SEQ ID NO: 31 or SEQ ID NO: 37 (all including signal sequences), or SEQ ID NO: 32 (without leader peptide)), PulD (*e.g.*, as set forth in SEQ ID NO: 11 or SEQ ID NO: 40), pIV, PilQ (*e.g.*, as set forth in SEQ ID NO: 12), InvG (*e.g.*, as set forth in SEQ ID Nos: 2 and 13-30), MxiD (*e.g.*, as set forth in SEQ ID NO: 6), YscC (*e.g.*, as set forth in SEQ ID NO: 5), PscC (*e.g.*, as set forth in SEQ ID NO: 7), EscC (*e.g.*, as set forth in SEQ ID NO: 8), SpiA (*e.g.*, as set forth in SEQ ID NO: 9), ExeD (*e.g.* as set forth in SEQ ID NO: 38), or XcpQ (*e.g.* as set forth in SEQ ID NO: 39) as known in the art. In some embodiments, a reference secretin amino acid sequence may be an amino acid sequence as set forth in SEQ ID Nos. 1-40. In some embodiments, a reference secretin amino acid sequence may be an amino acid sequence of a wild-type InvG nanopore subunit polypeptide or a mutant thereof, *e.g.*, as described in Worrall et al. "Near-Atomic-Resolution Cryo-EM analysis of the Salmonella T3S Injectisome Basal Body" Nature (2016) 540: 597–601. In some embodiments, a reference secretin amino acid sequence may be an amino acid sequence of a GspD nanopore subunit polypeptide or a mutant thereof, *e.g.*, as described in Yan et al. "Structural insights into the secretin translocation channel in the type II secretion system" Nature Structural & Molecular Biology (2017) doi:10.1038/nsmb.3350. Any natural secretin sequences or variant thereof that are known in the art can be used as a reference secretin amino acid sequence.

In some embodiments, the reference secretin amino acid sequence may be an amino acid sequence corresponding to the secretin domain, secretin and S domains, or secretin, S and N3 domains of the secretin, such as wild type GspD (*e.g.*, as set forth in SEQ ID NO: 36, SEQ ID NO: 35 or SEQ ID NO: 33) or an amino acid sequence corresponding to the secretin, S and N3 domains of GspD in which the constriction site in the N3 domain is deleted or substituted (*e.g.*, as set forth in SEQ ID NO: 34). Any natural truncated secretin sequences or variants thereof that form a pore can be used as a reference secretin amino acid sequence.

Accordingly, in some embodiments, a modified secretin nanopore subunit polypeptide has an amino acid sequence that is different from an amino acid sequence of any natural secretin, for example any of the reference secretin amino acid sequences (*e.g.*, any of SEQ ID NOs: 1-40) and comprises one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40) amino acid modifications relative to the selected natural secretin, for example relative to any of the reference secretin amino acid sequence (*e.g.*, relative to any one of SEQ ID NOs: 1-40). For example, a modified secretin nanopore subunit polypeptide may comprise an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at

least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to an amino acid sequence of a natural secretin, for example, any of the reference secretin amino acid sequence (*e.g.*, any of SEQ ID Nos: 1-40) or any structural or functional fragment thereof (*e.g.*, any fragment, portion, or domain of a secretin described herein, *e.g.*, any fragment, portion, or domain as illustrated in FIG. 1A or 1B), and includes at least one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40) amino acid modifications relative to a natural secretin, for example relative to a reference secretin (*e.g.*, relative to any one of SEQ ID NOs: 1-40) or any structural or functional fragment thereof (*e.g.*, any fragment, portion, or domain of a secretin described herein, *e.g.*, any fragment, portion, or domain as illustrated in FIG. 1A or 1B). The amino acid modification(s) can be selected, for example, to promote membrane integration, promote oligomerization, promote subunit synthesis, promote nanopore stability, promote analyte capture, promote analyte release, promote analyte translocation through a nanopore, improve analyte detection or signal quality, facilitate polymer analysis (*e.g.*, polynucleotide sequences), *etc.* In some embodiments, the amino acid modification(s) may comprise modification(s) to promote analyte capture into a nanopore, to promote analyte translocation through a nanopore, and/or to improve analyte detection such as to improve signal quality. Examples of such amino acid modification(s) include but are not limited to positively-charged substitutions and hydrophobic amino acid substitutions as described herein.

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/>). Sequence identity may be determined by using a pairwise sequence alignment. Global alignment techniques such as the Needleman –Wunsch algorithm, or local alignment methods such as the Smith–Waterman algorithm may be used to determine sequence alignments. Various techniques exist to determine structural homology such as DALI, a distance matrix alignment for constructing structural alignments http://ekhidna.biocenter.helsinki.fi/dali_server/start or SSAP (sequential structure alignment program), a dynamic programming-based method of structural alignment. An example of the latter is CATH <http://www.cathdb.info/>.

In some embodiments, the modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence corresponding to at least a portion of wild-type InvG secretin comprising the secretin domain, S domain, and N3 domain. In some embodiments, the InvG secretin can be obtained from any species, including, *e.g.*, but not limited to bacteria such as *Salmonella*, *Chromobacterium*, *Burkholderia*, *Providencia*, *Pseudogulbenkiania*, *Escherichia*, *Shigella*, *Stenotrophomonas*, *Pseudomonas*, *Yersinia*, *Arsenophonus*, *Pantoea*, and *Enterobacter*. The amino acid sequences of a full-length InvG secretin (including N0 and N1 domains) from different species are set forth in SEQ ID Nos: 2 and 13-32 and 37-40. In one embodiment, the InvG secretin can be obtained from *Salmonella*. For example, in some embodiments, the modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 1, which corresponds to the wild-type InvG secretin from *Salmonella* without N1 or N0 domain; and includes at least one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40) amino acid modifications (*e.g.*, as described herein) relative to the amino acid sequence as set forth in SEQ ID NO: 1. Alternatively, the modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 2, which corresponds to the wild-type full-length InvG secretin; and includes at least one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40) amino acid modifications (*e.g.*, as described herein) relative to the amino acid sequence as set forth in SEQ ID NO: 2. Without wishing to be bound by theory, removing the N1 and N0 domain of InvG secretin can improve signal-to-noise ratio of the modified secretin nanopores when they are used for detecting or characterizing an analyte, *e.g.*, a target polynucleotide or polypeptide.

In some embodiments, the modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at

least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence corresponding to at least a portion of a wild-type GspD secretin comprising the secretin domain, S domain, and N3 domain . In some embodiments, the GspD secretin can be obtained from any species, including, *e.g.*, but not limited to bacteria such as *Vibrio*, *Escherichia*, *Aeromonas*, *Pseudomonas*, and *Klebsiella*. The amino acid sequences of a full-length GspD secretin (including N0 and N1 domains) from different species are set forth in SEQ ID Nos: 4, 10, 31, 32 and 37). In one embodiment, the GspD secretin can be obtained from *Vibrio cholerae*. For example, in some embodiments, the modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 32, 33, 34, 35 or 36. The modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that corresponds to the amino acid sequence as set forth in SEQ ID NO: 32, 33, 34, 35 or 36 and includes at least one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40) amino acid modifications (*e.g.*, as described herein) relative to the amino acid sequence as set forth in SEQ ID NO: 32, 33, 34, 35 or 36. Alternatively, the GspD secretin can be obtained from *E. coli*, or the type II secretin can be PulD, *e.g.* from *Klebsiella oxytoca*, XcpQ, *e.g.* from *Pseudomonas aeruginosa*, or ExeD, *e.g.* from *Aeromonas hydrophila*. For example, in some embodiments, the modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that is at least about 40% (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in the mature portion of SEQ ID NO: 37, 38, 39 or 40, the N3, secretin and S domains of the amino acid sequences set forth in SEQ ID NO: 37, 38, 39 or 40, the secretin and S domains of the amino acid sequences set forth in SEQ ID NO: 37, 38, 39 or 40, or the secretin domain of the amino acid sequences set forth in SEQ ID NO: 37, 38, 39 or 40. The modified secretin nanopore may comprise a subunit polypeptide having an amino acid sequence that corresponds to the amino acid sequence as set forth in SEQ ID NO: 32, 33, 34, 35 or 36 and includes at least one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40) amino acid modifications (*e.g.*, as described herein) relative to the amino acid sequence as set forth in SEQ ID NO: 37, 38, 39 or 40, the N3, the secretin and S domains of the amino acid sequences set forth in SEQ ID NO: 37, 38, 39 or 40, the secretin and S domains of

the amino acid sequences set forth in SEQ ID NO: 37, 38, 39 or 40, or the secretin domain of the amino acid sequences set forth in SEQ ID NO: 37, 38, 39 or 40. The amino acids of the N3, secretin and S domains in these SEQ ID NOs can be determined by aligning the sequence with SEQ ID NO: 31 (as in the supplementary notes to Yan et al. "Structural insights into the secretin translocation channel in the type II secretion system" Nature Structural & Molecular Biology (2017) doi:10.1038/nsmb.3350).

FIG. 1B shows the secondary structure topology of the wild-type InvG secretin from positions 172-557 of SEQ ID NO: 2, where the numbered domains correspond to β -strands and the regions between two numbered domains (shown as a line with an arrowhead in FIG. 1B) correspond to loop regions. By way of example only, domain 4 (amino acids 381-393) and domain 5 (amino acids 400-417) correspond to β strands, and the region (amino acids 393-400) between the domains 4 and 5 corresponds to a loop region.

FIG. 11 shows the secondary structure topology of the wild-type GspD secretin (SEQ ID NO: 32), showing the β -strands, α -helicies and loop regions.

FIG. 1C shows the secondary structure topologies of a wild-type InvG secretin and a wild-type GspD secretin from *Vibrio cholerae* (from positions 97-646 of SEQ ID NO: 10, or from SEQ D NO: 32). In the *Vibrio cholerae* GspD amino acid sequence shown in SEQ ID NO: 32, amino acids 1 to 99 form the N0 domain, amino acids 100 to 163 form the N1 domain, amino acids 164 to 238 form the N2 domain, amino acids 239 to 314 form the N3 domain, amino acids 317 to 588 form the secretin domain and amino acids 589 to 650 form the S domain.

In some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more (*e.g.*, 1, 2, 3, or 4) β -strands of the secretin domains that form the outer β -barrel ("outer β -barrel-forming domains"), *e.g.*, domains numbered 1, 3a/3b, 8, and 9 as shown in FIG. 1B, or β 10, β 11, β 14, β 15, β 20 and β 21 as shown in FIG. 11. In some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more (*e.g.*, 1, 2, 3, or 4) loop regions between the outer β -barrel-forming domains, for example, as shown in FIG. 1B or Fig 11. In some embodiments, at least one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more (*e.g.*, 1, 2, 3, or 4) β -strands of the secretin domains that form the inner β -barrel, ("inner β -barrel-forming domains"), *e.g.*, domains numbered 4, 5, 6, and 7 as shown in FIG. 1B, or β 16, β 17, β 18 and β 19 as shown in FIG. 11. In some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,

12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more (*e.g.*, 1, 2, 3, or 4) loop regions between the inner β -barrel-forming domains. For example, in some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, or 8 amino acid modifications) can be made to the loop region between the inner β -barrel-forming domains 4 and 5 as shown in FIG. 1B. For example, in some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, or 8 amino acid modifications) can be made to the loop region between the inner β -barrel-forming β 16 and β 17 that forms the central gate as shown in FIG. 11.

In some embodiments, at least one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more (*e.g.*, 1, 2, or 3) domains that form the lips of the β -barrel (" β -barrel lip-forming domains"), *e.g.*, domains numbered 1, 2, and 3a as shown in FIG. 1B, or β 12, β 13, α 7, α 8 in FIG. 11, which correspond to the β -strands that form the *trans*-opening portion of the modified secretin nanopore described herein. In some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more (*e.g.*, 1, 2, 3, or 4) loop regions between the β -barrel lip-forming domains. For example, in some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, or 5 amino acid modifications) can be made to the loop region (amino acids 331-335) between the β -barrel lip-forming domains 1 and 2 as shown in FIG 1B or the loop between β 12 and β 13 (cap gate) in FIG. 11, which forms at least part of the *trans*-opening of the modified secretin nanopore described herein.

In some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more β -strands and/or loop regions within the N3 domain as defined in FIG. 1B or FIG 11. For example, in some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, or 5 amino acid modifications) can be made to the loop region defined by amino acids 216-268 of SEQ ID NO: 2. For example, in some embodiments, one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, or 5 amino acid modifications) can be made to the constriction site in the N3 domain of GspD (*e.g.* amino acids N265 to E282 in SEQ ID NO: 32).

In some embodiments, at least one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) can be made to one or more β -strands and/or loop regions within the S domain as defined in FIG. 1B or FIG 11.

Accordingly, in some embodiments, the modified nanopore secretin nanopore may comprise a subunit polypeptide having (i) outer β -barrel-forming domains of InvG or GspD secretin and/or loop regions there between; (ii) inner β -barrel-forming domains of InvG or GspD secretin and/or loop regions there between; (iii) β -barrel lip-forming domains of InvG or GspD secretin and/or loop regions there between, (iv) S domain of InvG or GspD secretin and/or loop regions there between; and (v) N3 domain of InvG or GspD secretin and/or loop regions there between, in which the β -strands and/or loop regions may have different numbers and/or types of amino acid modifications, *e.g.*, depending on their locations within the nanopore and/or its degree of interaction with an analyte and/or an enzyme. For example, the amino acid sequence of each of the outer β -barrel-forming domains and/or loop regions there between may be each independently at least about 50% (including, *e.g.*, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher, including 100%) identical to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37.

The amino acid sequence of each of the inner β -barrel-forming domains and/or loop regions there between may be each independently at least about 50% (including, *e.g.*, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher, including 100%) identical to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37. The amino acid sequence of each of the β -barrel lip-forming domains and/or loop regions there between may be each independently at least about 50% (including, *e.g.*, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher, including 100%) identical to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37. The amino acid sequence of each domain within the S domain and/or loop regions there between may be at least about 50% (including, *e.g.*, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher, including 100%) identical to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37. The amino acid sequence of each domain within the N3 domain and/or loop regions there between may be at least about 50% (including, *e.g.*, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher, including 100%) identical to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37. Each domain may have different percentages of amino acid

identity provided that the resulting modified domain does not adversely affect the capture and/or translocation of an analyte through the lumen of the modified nanopore. For example, in some embodiments, the outer β -barrel forming domains may permit a larger number of amino acid mutations that result in less than 80% or lower (including, *e.g.*, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, or lower) amino acid identity to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37, while the inner β -barrel forming domains maintain a higher amino acid identity, for example, the amino acid sequence of the inner β -barrel forming domains may be each independently at least about 80% or higher (including at least about 85%, at least about 90%, at least about 95% or higher, including 100%). In some embodiments, at least one loop region of the N3 domain (*e.g.*, a loop region defined by amino acids 216-268 of SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37) may permit a larger number of amino acid mutations (*e.g.*, to improve enzyme/nanopore interaction) that result in less than 80% or lower (including, *e.g.*, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, or lower) amino acid identity to the amino acid sequence of the corresponding domain as set forth in SEQ ID NO: 2, or SEQ ID NO: 4, 32 or 37, while the inner β -barrel forming domains maintain a higher amino acid identity, for example, the amino acid sequence of the inner β -barrel forming domains may be each independently at least about 80% or higher (including at least about 85%, at least about 90%, at least about 95% or higher, including 100%).

One of ordinary skill in the art will readily recognize that various types of modifications to the secretin nanopores as described herein (*e.g.*, but not limited to amino acid modifications to different domains of secretin nanopores) can be applied to any other secretin nanopores that have a high structural homology to secretin nanopores as described herein. By way of example only, SEQ ID Nos: 4 and 37, and 10, 31 and 32 relate to GspD from *Escherichia coli* and *Vibrio cholerae*, respectively. The sequence identity between SEQ ID NO: 4 and SEQ ID NO: 10, for example, is 41.6%, the sequence identity between the secretin domains of SEQ ID NO: 4 and SEQ ID NO: 10, for example, is 44.2% and the similarity is 62.7% as calculated by pairwise alignment using the EMBOSS Needle nucleotide alignment algorithm provided by EMBL-EBI http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html. While the sequence identities between the two structures may be low, they share a high structural homology because they both have similar structural domains, including, *e.g.*, secretin domain, S domain, N3 domain, N2 domain, and N1 domain.

Truncated secretin subunit polypeptides that lack the N-terminal domains are capable of forming pores. Therefore the modified secretin nanopore of the invention is, in some embodiments, a truncated secretin nanopore. The truncated secretin nanopore may typically

comprise an N3 domain, a secretin domain and an S domain, a secretin domain and an S domain, or a secretin domain.

Thus, in some embodiments, the secretin nanopore subunit polypeptide comprises a secretin domain comprising a beta barrel forming domain comprising an inner barrel forming subdomain and an outer barrel forming subdomain, each subdomain being composed of β -sheets, the outer barrel typically comprising about six β -sheets and/or the inner barrel typically comprising about four β -sheets. The outer beta barrel may further comprise two α -helices, typically between two of the β -sheets, for example as shown in Figure 11. In a secretin nanopore, the outer barrel typically spans the membrane and the inner barrel typically abuts the lumen of the pore. The inner barrel typically comprises a central gate. The central gate is typically a loop between two β -sheets that form the inner barrel. The central gate typically extends into the pore to narrow the size of the pore. The central gate can be modified by altering amino acids present in the central gate loop as described herein to alter the properties of the pore. The central gate may be flexible, for example the central gate may be capable of opening. The central gate may be rigid to maintain a constant constriction size, e.g. the central gate loop may be closed or partially closed. The beta barrel of the secretin nanopore may also comprise lips, wherein a first lip protrudes from the membrane on the opposite side of the membrane to the inner beta barrel. The second lip may be on the other side of the inner beta barrel to the first lip. The first lip of the beta barrel is typically composed of two α -helices and two β -sheets. The β -sheets may be joined by a loop region that forms a cap gate, or the loop joining the β -sheets may be short and not form a gate. The cap gate may be flexible, for example the cap gate may be capable of opening. The cap gate may be rigid to maintain a constant constriction size, e.g. the cap gate may be closed or partially closed. In some embodiments, the first lip of the beta barrel may comprise no β -sheets and comprise two α -helices that are joined by a loop. In these embodiments the subunit polypeptide forms a nanopore which does not comprise a cap gate. The second lip of the beta barrel may comprise two α -helices.

In some embodiments, the secretin nanopore subunit polypeptide may, in addition to the secretin domain, comprise an S domain. The S-domain may comprise two α -helices. One of the α -helices typically interacts with the beta-barrel of the secretin nanopore. The S-domain is typically located on the outside of the pore (i.e. away from the lumen of the pore).

In some embodiments, the secretin nanopore subunit polypeptide may, in addition to the secretin domain, and optionally the S domain, comprise an N3 domain. The N3 domain is typically composed of β -barrels and α -helices, e.g. from 3 to 6 β -barrels and from 2 to 3 α -helices, such as 3 β -barrels and 2 α -helices as shown in FIG. 11 or 6 β -barrels and 3 α -helices as shown in FIG. 1B. The N3 domain may form a constriction in the lumen of the pore. The N3

domain may be modified so that it does not constrict the pore. The N3 domain may be modified to increase or decrease the size of the constriction.

In some embodiments, the amino acid sequence of the modified secretin nanopore subunit polypeptide comprises one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) at positions within the lumen-forming amino acid sequence. The amino acid modifications are selected to provide improved frequency of capture and/or translocation of an analyte (*e.g.*, a polynucleotide such as double stranded or single stranded DNA) through the nanopore, as compared to a reference secretin amino acid sequence.

In some embodiments, the amino acid modifications may be charge-altering modifications. In some embodiments, the amino acid modification is a positively-charged amino acid substitution. The term "positively-charged amino acid substitution" as used herein refers to a modification to a reference amino acid that increases the net positive charge, or decreases the net negative charge, of the reference amino acid, *e.g.*, as detected at pH 7.0-8.0 (*e.g.*, at pH 8.0) and at room temperature, *e.g.*, at 20-25°C. For example, a positively-charged amino acid substitution can include, but is not limited to, (i) replacement of a negatively-charged amino acid with a less negatively charged amino acid, neutral amino acid, or positively-charged amino acid, (ii) replacement of a neutral amino acid with a positively-charged amino acid, or (iii) replacement of a positively charged amino acid with a more positively-charged amino acid. In some embodiments, a positively-charged amino acid substitution may include deletion of a negatively-charged amino acid or addition of a positively-charged amino acid. In some embodiments, a positively-charged amino acid substitution may include one or more chemical modifications of one or more negatively charged amino acids which neutralize their negative charge. For instance, the one or more negatively charged amino acids may be reacted with a carbodiimide.

A positively-charged amino acid is an amino acid having an isoelectric point (pI) that is higher than the pH of a solution so that the amino acid in the solution carries a net positive charge. For example, examples of a positively-charged amino acid as detected at pH 7.0-8.0 (*e.g.*, at pH 8.0) and at room temperature, *e.g.*, at 20-25°C, include, but are not limited to arginine (R), histidine (H), and lysine (K). A negatively-charged amino acid is an amino acid having a pI that is lower than the pH of a solution so that the amino acid in the solution carries a net negative charge. Examples of a negatively-charged amino acid as detected at pH 7.0-8.0 (*e.g.*, at pH 8.0) and at room temperature, *e.g.*, at 20-25°C, include, but are not limited to aspartic acid (D), glutamic acid (E), serine (S), glutamine (Q). A neutral amino acid is an amino acid having an isoelectric point (pI) that is same as the pH of a solution so that the amino acid in the solution carries no net charge. A neutral amino acid can be a polar, non-polar, or hydrophobic

amino acid. The pI values of amino acids are known in the art. By comparing the pI value of an amino acid of interest to the pH of a solution, one of ordinary skill in the art will readily determine whether the amino acid present in the solution is a positively charged amino acid, a neutral amino acid, or a negatively-charged amino acid. An amino acid can be a naturally-
 5 occurring or synthetic amino acid.

In some embodiments, the amino acid modification may be a modification to change the hydrophobicity of the amino acid. Such a modification includes a modification to a reference amino acid that changes its hydrophobicity, *e.g.*, as detected at pH 7.0-8.0 (*e.g.*, at pH 8.0) and at room temperature, *e.g.*, at 20-25°C. For example, the amino acid modification may be a
 10 substitution of a reference amino acid with a hydrophobic amino acid, *e.g.*, an amino acid with a hydrophobic side chain. Examples of hydrophobic amino acids include glycine (G), alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), phenylalanine (F), methionine (M), tyrosine (Y), and tryptophan (W). For example, the amino acid modification may be a substitution of a neutral amino acid with a hydrophobic amino acid. The hydropathy index of amino acids are
 15 known in the art. Hydrophobicity scales are values that define relative hydrophobicity of amino acid residues. The more positive the value, the more hydrophobic are the amino acids located in that region of the protein. An amino acid can be an naturally-occurring or synthetic amino acid.

In some embodiments, the amino acid modification may be a modification to change the size of the amino acid. Such a modification includes a modification to a reference amino acid
 20 that changes its size, *e.g.*, the size of the side chain. For example, the amino acid modification may be a substitution of a reference amino acid having a large side chain with an amino acid having a smaller side chain. Examples of very large amino acids include phenylalanine (F), tryptophan (W) and tyrosine (Y). Examples of large amino acids include isoleucine (I), leucine (L), methionine (M), lysine (K) and arginine(R). Examples of medium sized amino acids
 25 include valine (V), histidine (H), glutamic acid (E) and glutamine (Q). Examples of small amino acids include cysteine (C), proline (P), threonine (T), aspartic acid (D) and asparagine (N). Examples of very small amino acids include serine (S), glycine (G) and alanine (A). For example, the amino acid modification may be a substitution of a very large amino acid with a large, medium, small or very small amino acid. For example, the amino acid modification may
 30 be a substitution of a large amino acid with a medium, small or very small amino acid. For example, the amino acid modification may be a substitution of a medium amino acid with a small or very small amino acid. The smaller amino acid can be an naturally-occurring or synthetic amino acid.

In some embodiments, the modified secretin nanopore subunit polypeptide is a modified
 35 InvG nanopore subunit polypeptide comprising an amino acid sequence that is at least about 40%

(including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 1 (corresponding to the amino acid sequence of InvG without N1 or N0 domain), wherein the modified InvG nanopore subunit polypeptide comprises one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, or 7 amino acid modifications) at amino acid(s) selected from D28, E41, E114, Q45, E225, R226, and E231 of SEQ ID NO: 1. The amino acid modification can be a positively-charged amino acid substitution or a modification to change the hydrophobicity of a reference amino acid. In some embodiments, the amino acid modification may comprises one or more (*e.g.*, 1, 2, 3, 4, 5, or 6) of the following: (i) D28N/Q/T/S/G/R/K; (ii) E225 N/Q/T/A/S/G/P/H/F/Y/R/K; (iii) R226N/Q/T/A/S/G/P/H/F/Y/K/V; (iv) deletion of E225; (v) deletion of R226; and (vi) E231N/Q/T/A/S/G/P/H/R/K. In some embodiments, the modified InvG nanopore subunit polypeptide may comprise one or more amino acid modifications at amino acid(s) selected from Q45, E41, and E114 of SEQ ID NO: 1. For example, the modified InvG nanopore subunit polypeptide may comprise one or more (*e.g.*, 1, 2, or 3) of the following amino acid modifications: (i) Q45R/K; (ii) E41N/Q/T/S/G/R/K; and (iii) E114N/Q/T/S/G/R/K of SEQ ID NO: 1. The "/" symbol between amino acids X and Y means that a reference amino acid may be modified to amino acid X or amino acid Y. It should be understood that the amino acid positions based on SEQ ID NO: 1 will shift accordingly if modifications (*e.g.*, amino acid addition or deletion) are made to the N-terminus of or within the amino acid sequence as set forth in SEQ ID NO: 1. By way of example only, SEQ ID NO: 2 differs from SEQ ID NO: 1 in that the N-terminus of SEQ ID NO: 2 contains additional 171 amino acids that correspond to the N0 and N1 domains of an InvG nanopore, which are missing from the N-terminus of SEQ ID NO: 1. Thus, one of ordinary skill in the art will readily recognize that the amino acid positions D28, E41, E114, Q45, E225, R226, and E231 in SEQ ID NO: 1 correspond to amino acid positions D199, E212, E285, Q216, E396, R397, and E402 in SEQ ID NO: 2.

In some embodiments, the modified InvG nanopore subunit polypeptide comprises an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 1 and one or any combinations of the amino acid modifications as shown in FIG. 7. For example, in some embodiments, the modified InvG nanopore subunit polypeptide may comprise amino acid substitution E225N/Q/T/A/S/G/P/H/F/Y/R/K and deletion of R226 of SEQ ID NO: 1. In some embodiments, the modified InvG nanopore subunit polypeptide may comprise a deletion of E225 amino acid

and amino acid substitutions E231N/Q/T/A/S/G/P/H/R/K and Q45R/K. It should be noted that the amino acid positions as shown in FIG. 7 (based on SEQ ID NO: 2) are adjusted to correspond to the amino acid positions in SEQ ID NO: 1. Methods of aligning two amino acid sequences are known in the art. Thus, one of ordinary skill in the art can readily identify the corresponding amino acid positions in SEQ ID NO: 1 based on the amino acid positions provided in SEQ ID NO: 2.

In some embodiments, the modified InvG nanopore subunit polypeptide comprises an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 2 and one or any combinations of the amino acid modifications as shown in FIG. 7.

In another aspect, provided herein is a modified secretin nanopore subunit polypeptide comprising an amino acid sequence that is at least 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence of a secretin nanopore subunit polypeptide, *e.g.*, the amino acid sequence as set forth in SEQ ID NOs: 2, or 4-30 (corresponding to the amino acid sequence of wild-type (WT) secretin including N1 and N0 domains), wherein an endopeptidase cleavage site is inserted upstream of the N3 domain of the secretin nanopore subunit polypeptide. In some embodiments, the endopeptidase cleavage site is inserted between the N1 domain and N3 domain of the secretin nanopore subunit polypeptide (*e.g.*, an InvG nanopore subunit polypeptide). In other embodiments, the endopeptidase cleavage site is inserted between the N1 domain and N2 domain (*e.g.*, a GspD or PulD nanopore subunit polypeptide). Such a modified secretin nanopore subunit polypeptide allows removal of N1 and/or N0 domains using an endopeptidase that targets the corresponding endopeptidase cleavage site after expression of the polypeptide. For example, cleavage of N0 and N1 domains can be done by treating full length protein that are expressed and purified with an appropriate endopeptidase.

As used herein, the term "endopeptidase cleavage site" refers to a peptide sequence that is recognized and cleaved by an endopeptidase, which is a proteolytic enzyme that breaks or cleaves bonds of nonterminal amino acids (*e.g.*, within the molecule). Various endopeptidases and their corresponding cleavage sites are known in the art. For example, such information can be assessed online at web.expasy.org/peptide_cutter/peptidecutter_enzymes.html. Non-limiting examples of endopeptidases include, but are not limited to, Trypsin, Chymotrypsin, Elastase, Thermolysin, Pepsin, Glutamyl endopeptidase, Neprilysin, Caspase 1-10, CNBr, Enterokinase,

Proteinase K, Factor Xa Protease, Bovine Alpha Thrombin, and Tobacco Etch Virus (TEV) protease. In one embodiment, the endopeptidase cleavage site inserted into the modified secretin nanopore subunit polypeptide may be recognized by a TEV protease. TEV protease recognizes a linear epitope of the general form E-Xaa-Xaa-T-Xaa-Q-(G/S), with cleavage occurring between
 5 Q and G or Q and S. An exemplary TEV protease cleavage sequence may be ENLYFQG. In one embodiment, the endopeptidase cleavage site inserted into the modified secretin nanopore subunit polypeptide may be recognized by a Factor Xa Protease. Factor Xa cleaves after the arginine residue in its preferred cleavage site Ile-(Glu or Asp)-Gly-Arg. It will sometimes cleave at other basic residues, depending on the conformation of the protein substrate. In another
 10 embodiment, the endopeptidase cleavage site inserted into the modified secretin nanopore subunit polypeptide may be recognized by a bovine alpha thrombin. Thrombin recognizes the consensus sequence Leu-Val-Pro-Arg-Gly-Ser, cleaving the peptide bond between Arg and Gly.

In one aspect, provided herein is a modified InvG nanopore subunit polypeptide comprising an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least
 15 about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 2 (corresponding to the amino acid sequence of WT InvG including N1 and N0 domains), wherein an endopeptidase cleavage site is inserted between positions 170 and 171 or 171 and 172 of SEQ ID NO: 2. In one
 20 embodiment, an endopeptidase cleavage site is inserted between D171 and G172 of SEQ ID NO: 2. Such a modified InvG nanopore subunit polypeptide allows removal of N1 and N0 domains using an endopeptidase that targets the corresponding endopeptidase cleavage site after expression of the polypeptide. Any appropriate endopeptidase cleavage site (*e.g.*, as described herein) can be used. In one embodiment, the endopeptidase cleavage site inserted into the
 25 modified InvG nanopore subunit polypeptide may be recognized by a TEV protease. An exemplary TEV protease cleavage sequence may be ENLYFQG. Example 2 provides an exemplary method for expression and purification of a modified secretin nanopore subunit polypeptide comprising a TEV protease cleavage site.

In some embodiments, the modified InvG nanopore subunit polypeptide comprising an
 30 endopeptidase cleavage site may comprise one or more (*e.g.*, 1, 2, 3, 4, 5, 6, or 7) of the amino acid modifications as shown in FIG. 7. For example, in some embodiments, the modified InvG nanopore subunit polypeptide may comprise amino acid substitution E396N/Q/T/A/S/G/P/H/F/Y/R/K and deletion of R397 of SEQ ID NO: 2. In some embodiments, the modified InvG nanopore subunit polypeptide may comprise a deletion of E396 amino acid
 35 and amino acid substitutions E402N/Q/T/A/S/G/P/H/R/K and Q216R/K.

For example, in one aspect the modified GspD secretin nanopore comprises a subunit polypeptide comprising a secretin domain having an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence of the secretin domain set forth in SEQ ID NO: 36.

The secretin domain of GspD from *Vibrio cholerae* and from *Escherichia coli* ETEC contain a cap gate. Other Type II secretion system subunit polypeptides, including some GspD secretin subunit polypeptides, such as *Escherichia coli* K12, do not comprise a cap gate. The modified GspD secretin nanopore may, in one aspect be one that does not comprise a cap gate. The secretin domain set in out SEQ ID NO: 36 comprises a cap gate between positions 56 and 77. For example, the secretin domain set forth in SEQ ID NO: 36 may be modified to delete all or part of the cap gate, *e.g.* all or some of the amino acids from D55 or T56 to T77 of SEQ ID NO: 36 may be deleted or substituted. Alternatively, the modified GspD secretin nanopore may naturally lack a cap gate.

The central gate of GspD may be modified to replace an amino acid with an amino acid having a smaller side group and/or to replace a negatively charged amino acid with a neutral or positively charged amino acid. The secretin domain set in out SEQ ID NO: 36 comprises a central gate between positions 144 to 157, which correspond to positions 460 and 473 of SEQ ID NO: 32. The secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to an amino acid sequence as set forth in SEQ ID NO: 36, wherein: (i) all or some of the amino acids from D55 or T56 to T77 are deleted or substituted, one or more of K60, D64, R71 and E73 is substituted with an uncharged amino acid and/or one or more of D55, T56, T77 and K78 is substituted with P; and/or (ii) F156 is substituted with a smaller amino acid, N151 and/or N152 is/are substituted with a smaller amino acid, D153 is substituted with an uncharged amino acid, G137 and G165 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y63 to R71 may be deleted and/or substituted with GSG or SGS, F156 may be substituted with A, D153 may be substituted with S, and/or N151 and N152 may each independently be substituted with G or S. D55, T56, K60, Y63, D64, R71, E73, T77, K78, G137, N151, N152, D153, F156 and G165 of SEQ ID NO: 36 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32.

The modified secretin GspD nanopore may comprise a modified secretin domain as defined above with reference to SEQ ID NO 36, an N3 domain and an S domain. The modified secretin GspD nanopore may in one aspect comprises a subunit polypeptide comprising an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to the amino acid sequence as set forth in SEQ ID NO: 33, 34 and/or or 35. SEQ ID NO: 35 comprises a secretin domain and an S domain. SEQ ID NO: 34 comprises a secretin domain, an S domain and a modified N3 domain. SEQ ID NO: 34 comprises a secretin domain, an S domain and an N3 domain. The amino acid modifications referred to with reference to SEQ ID NO: 36 may be made at the corresponding positions of any one of SEQ ID NOs: 31 to 35. The amino acid modifications referred to with reference to SEQ ID NO: 36 may also be made at the corresponding positions of any one of SEQ ID NOs: 4 and 37 to 40, or to a truncated subunit polypeptide comprising a portion of any one of SEQ ID NOs: 4 and 37 to 40, *e.g.* a truncated subunit polypeptide comprising the secretin domain, secretin and S domains or secretin, S and N3 domains of any one of SEQ ID NOs: 4 and 37 to 40.

For example, the secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to an amino acid sequence as set forth in SEQ ID NO: 34, wherein: (i) all or some of the amino acids from D117 or T118 to T139 are deleted or substituted, one or more of K122, D126, R133 and E135 is substituted with an uncharged amino acid and/or one or more of D117, T118, T139 and K140 is substituted with P; and/or (ii) F218 is substituted with a smaller amino acid, N213 and/or N214 is/are substituted with a smaller amino acid, D215 is substituted with an uncharged amino acid, G199 and G227 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y125 to R133 may deleted and/or substituted with GSG or SGS, F218 may be substituted with A, D215 may be substituted with S, and/or N213 and N214 may each independently be substituted with G or S. D117, T118, K122, Y125, D126, R133, E135, T139, K140, G199, N213, N214, D215, F218 and G227 of SEQ ID NO: 34 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32.

For example, the secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least about 40% or higher (including,

e.g., at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to an amino acid sequence as set forth in SEQ ID NO: 35, wherein: (i) all or some of the amino acids from D55 or T56 to T77 are deleted or substituted, one or more of K60, D64, R71 and E73 is substituted with an uncharged amino acid and/or one or more of D55, T56, T77 and K78 is substituted with P; and/or (ii) F156 is substituted with a smaller amino acid, N151 and/or N152 is/are substituted with a smaller amino acid, D153 is substituted with an uncharged amino acid, G137 and G165 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y63 to R71 may be deleted and/or substituted with GSG or SGS, F156 may be substituted with A, D153 may be substituted with S, and/or N151 and N152 may each independently be substituted with G or S. D55, T56, K60, Y63, D64, R71, E73, T77, K78, G137, N151, N152, D153, F156 and G165 of SEQ ID NO: 35 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32.

For example, the secretin domain of the modified GspD secretin nanopore may comprise a secretin domain having an amino acid sequence that is at least about 40% or higher (including, *e.g.*, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or higher) identical to an amino acid sequence as set forth in SEQ ID NO: 33, wherein: (i) all or some of the amino acids from D132 or T133 to T154 are deleted or substituted, one or more of K137, D141, R148 and E150 is substituted with an uncharged amino acid and/or one or more of D132, T133, T154 and K155 is substituted with P; and/or (ii) F233 is substituted with a smaller amino acid, N228 and/or N229 is/are substituted with a smaller amino acid, D230 is substituted with an uncharged amino acid, G214 and G242 are each independently unmodified or substituted with A or V. For example, in the modified secretin GspD nanopore Y140 to R148 may be deleted and/or substituted with GSG or SGS, F233 may be substituted with A, D230 may be substituted with S, and/or N228 and N229 may each independently be substituted with G or S. D132, T133, K137, Y140, D141, R148, E150, T154, K155, G214, N228, N229, D230, F233 and G242 of SEQ ID NO: 33 correspond to D371, T372, K376, Y379, D380, R387, E389, T393, K394, G453, N467, N468, D469, F472 and G481 of the full length GspD amino acid sequence set forth in SEQ ID NO: 32.

In any aspects of the modified secretin nanopore subunit polypeptide described herein, additional amino acid substitutions (other than the amino acid modifications described above), may be made to a reference secretin amino acid sequence, for example up to 1, 2, 3, 4, 5, 6, 7, 8,

9, 10, 15, 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 1 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 2.

Table 1 – 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 2 - Hydropathy scale

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

	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
5	Arg	-4.5

One or more amino acid residues of the reference amino acid sequence (*e.g.*, as set forth in SEQ ID Nos: 1-10) may additionally be deleted from the polypeptides described above. Up to 1, 2, 3, 4, 5, 10, 20 or 30 residues may be deleted, or more.

10 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 reference amino acid sequence (*e.g.*, as set forth in SEQ ID NO: 1 or 2) or polypeptide variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids. A
15 carrier protein may be fused to an amino acid sequence, *e.g.*, an amino acid sequence of a modified secretin nanopore subunit polypeptide. Other fusion proteins are discussed in more detail below.

Methods for modifying amino acids (*e.g.*, by substitution, addition, or deletion) are well known in the art. For instance, a reference amino acid may be substituted with a target amino
20 acid by replacing the codon for the reference amino acid with a codon for the target amino acid at the relevant position in a polynucleotide encoding the modified secretin nanopore subunit polypeptide. The polynucleotide can then be expressed as discussed below. If the amino acid is a non-naturally-occurring amino acid, it may be introduced by including synthetic aminoacyl-tRNAs in the IVTT system used to express the modified secretin nanopore subunit polypeptide.
25 Alternatively, it may be introduced by expressing the modified secretin nanopore subunit polypeptide 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 modified secretin nanopore subunit polypeptide is produced using partial peptide synthesis.

30 The modified secretin nanopore subunit polypeptides described herein may be used to form a homo-multimeric nanopore or hetero-multimeric nanopore as described herein. Accordingly, in some embodiments, the modified secretin nanopore subunit polypeptide retains the ability to form a nanopore with other subunit polypeptides. Methods for assessing the ability of modified monomers to form nanopores are well-known in the art. For instance, a modified
35 secretin nanopore subunit polypeptide may be inserted into an amphiphilic layer along with other appropriate subunits and its ability to oligomerize 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. Alternatively, subunits may be directly

5 Bayley, J. Am. Chem. Soc. 2005, 127, 6502-6503 and International Application No. PCT/GB2006/001057 (published as WO 2006/100484), the contents of which are incorporated herein by reference.

The modified secretin nanopore subunit polypeptides may contain non-specific modifications as long as they do not interfere with nanopore formation. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the
10 amino acids. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with NaBH₄, amidination with methylacetimidate or acylation with acetic anhydride.

The modified secretin nanopore subunit polypeptides can be produced using standard
15 methods known in the art. The modified secretin nanopore subunit polypeptides may be made synthetically or by recombinant means. Exemplary methods for expression and purification of the modified secretin nanopore subunit polypeptides according to some embodiments described herein are provided in Examples 1 and 2. Alternatively, the modified secretin nanopore subunit polypeptides may be synthesized by *in vitro* translation and transcription (IVTT). Suitable
20 methods for producing pores and modified secretin nanopore subunit polypeptides are discussed in International Application Nos. PCT/GB09/001690 (published as WO 2010/004273), PCT/GB09/001679 (published as WO 2010/004265) or PCT/GB10/000133 (published as WO 2010/086603), the contents of each of which are incorporated herein by reference.

The modified secretin nanopore subunit polypeptides as described herein may be
25 produced using D-amino acids. For instance, the modified secretin nanopore subunit polypeptides as described herein may comprise a mixture of L-amino acids and D-amino acids. This is conventional in the art for producing such proteins or peptides.

In some embodiments, the modified secretin nanopore subunit polypeptides may be chemically modified. The modified secretin nanopore subunit polypeptides can be chemically
30 modified in any way and at any site. For instance, the modified secretin nanopore subunit polypeptides may be chemically modified by attachment of a dye or a fluorophore. In some embodiments, the modified secretin nanopore subunit polypeptide may be 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.

In some embodiments, the modified secretin nanopore subunit polypeptide may be chemically modified with a molecular adaptor that facilitates the interaction between a nanopore comprising the modified secretin nanopore subunit polypeptide and a target nucleotide or target polynucleotide sequence. The presence of the adaptor improves the host-guest chemistry of the nanopore and the nucleotide or polynucleotide sequence and thereby improves the sequencing ability of pores formed from the modified secretin nanopore subunit polypeptides. 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 nanopore 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.

In some embodiments, the molecular adaptor may be 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.

In some embodiments, the molecular adaptor can be covalently attached to the modified secretin nanopore subunit polypeptide. The adaptor can be covalently attached to the nanopore using any method known in the art. The adaptor is typically attached via chemical linkage. If the molecular adaptor is attached via cysteine linkage, one or more cysteines can be introduced to the modified secretin nanopore subunit polypeptide by substitution.

In other embodiment, the modified secretin nanopore subunit polypeptide may be attached or coupled to an enzyme such as a polynucleotide binding protein, *e.g.*, helicases, exonucleases, and polymerases. In some embodiments, the modified secretin nanopore subunit polypeptide may be attached or coupled to a helicase, *e.g.*, a DNA helicase. Examples of helicases, exonucleases, and polymerases that are suitable for use in nanopore sequencing are known in the art. In some embodiments, the modified secretin nanopore subunit polypeptide may be attached or coupled to a helicase, *e.g.*, a DNA helicase, a Hel308 helicase (*e.g.*, as described in WO 2013/057495), a RecD helicase (*e.g.*, as described in WO2013/098562), a XPD helicase (*e.g.*, as described in WO201/098561), or a Dda helicase (*e.g.*, as described in WO2015/055981). This forms a modular sequencing system that may be used in the methods of characterizing a target polynucleotide. Polynucleotide binding proteins are discussed below. The translocation speed control may be determined by the type of polynucleotide binding protein and/or amount of fuel (ATP) added to the system. For example, the rate of translocation of the double stranded

DNA analyte may be controlled by a double stranded DNA translocase such as FtsK. Depending upon the fuel (ATP) added to the system, the translocation speed of a target polynucleotide can be between about 30 B/s and 1000 B/s or about 30 B/s and 2000 B/s.

5 In some embodiments, the polynucleotide binding protein can be covalently attached to the modified secretin nanopore subunit polypeptide. The polynucleotide binding protein can be covalently attached to the modified secretin nanopore subunit polypeptide using any method known in the art. The modified secretin nanopore subunit polypeptide and the polynucleotide binding protein may be chemically fused or genetically fused. The modified secretin nanopore subunit polypeptide and the polynucleotide binding protein are genetically fused if the whole
10 construct is expressed from a single polynucleotide sequence. Genetic fusion of a modified secretin nanopore subunit polypeptide to a polynucleotide binding protein is discussed in International Application No. PCT/GB09/001679 (published as WO 2010/004265), the contents of which are incorporated herein by reference.

The modified secretin nanopore subunit polypeptide may be chemically modified with a
15 molecular adaptor and a polynucleotide binding protein.

Any of the proteins described herein, such as the modified secretin nanopore subunit polypeptides and nanopores described herein, 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
20 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
25 Jul;4(7):497-505).

Any of the proteins described herein, such as the modified secretin nanopore subunit polypeptide and nanopores described herein, may be labelled with a detectable label. The detectable 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,
30 enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

Any of the proteins described herein, including the modified secretin nanopore subunit polypeptide described herein, 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
35 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.

5 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.

10 ***Polynucleotides encoding the modified secretin nanopore subunit polypeptides***

Provided herein are also polynucleotide sequences encoding any one of the modified secretin nanopore subunit polypeptides as described herein.

Polynucleotide sequences may be derived and replicated using standard methods in the art. Chromosomal DNA encoding wild-type secretin may be extracted from a pore producing
15 organism, such as *Salmonella typhi*. 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 any one of the modified secretin nanopore subunit polypeptides can be made using well-known techniques, such as those
20 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
25 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.

Another aspect of the disclosure includes a method of producing a modified secretin
30 nanopore subunit polypeptide or a construct described herein. The method comprises expressing a polynucleotide encoding any embodiment of the modified secretin nanopore subunit polypeptides in a suitable host cell. The polynucleotide is preferably part of a vector and is preferably operably linked to a promoter.

35

Modified secretin nanopores

One aspect of the present disclosure features a modified secretin nanopore, for example, that is disposed in a membrane and permits capture of an analyte, *e.g.*, a target polynucleotide or polypeptide, into the modified secretin nanopore and/or translocation of the analyte through the modified secretin nanopore. The modified secretin nanopore, *e.g.*, as disposed in a membrane, comprises a luminal surface defining a lumen that extends, *e.g.*, through the membrane, between a *cis*-opening and a *trans*-opening, in which the luminal surface comprises one or more amino acid modifications. As used herein, the term "luminal surface" refers to the internal surface of a nanopore, which surface comprises a set of amino acids of multiple nanopore subunits, that defines a lumen that is exposed to a solution.

In some embodiments, the secretin nanopore comprises a secretin domain comprising a beta barrel comprising an inner barrel subdomain and an outer barrel subdomain, each composed of β -sheets, with each subunit typically contributing about six β -sheets and/or the inner barrel typically comprising about four β -sheets to the outer barrel. Each subunit may further contribute two α -helices, typically between two of the β -sheets, to the outer beta barrel, for example as shown in Figure 11. The outer barrel typically spans the membrane. The inner barrel typically abuts the lumen of the pore. The inner barrel typically comprises a central gate. The central gate is typically formed from loops between two β -sheets that form the inner barrel in each subunit. The central gate typically extends into the pore to narrow the size of the pore. The central gate can be modified by altering amino acids present in the central gate loop as described herein to alter the properties of the pore. The central gate may be flexible, for example the central gate may be capable of opening. The central gate may be rigid to maintain a constant constriction size, *e.g.* the central gate loop may be closed or partially closed. The beta barrel of the secretin nanopore wherein a first lip protrudes from the membrane on the opposite side of the membrane to the inner beta barrel. The lips of the beta barrel are typically composed of two α -helices and two β -sheets from each subunit polypeptide. The β -sheets in each subunit may be joined by a loop region and the loop regions form a cap gate. Alternatively, the loop joining the β -sheets may be short and not form a gate. The cap gate may be flexible, for example the cap gate may be capable of opening. The cap gate may be rigid to maintain a constant constriction size, *e.g.* the cap gate may be closed or partially closed. In some embodiments, the first lip of the beta barrel may comprise no β -sheets and comprise from each subunit two α -helices that are joined by a loop. In these embodiments the nanopore does not comprise a cap gate. The second lip may be on the other side of the inner beta barrel to the first lip. The second lip of the beta barrel may comprise two α -helices in each subunit.

In some embodiments, the secretin nanopore may in addition to the secretin domain, comprise an S domain. The S-domain may comprise two α -helices. One of the α -helices typically interacts with the beta-barrel of the secretin nanopore. The S-domain is typically located on the outside of the pore (i.e. away from the lumen of the pore).

5 In some embodiments, the secretin nanopore may, in addition to the secretin domain, and optionally the S domain, comprises an N3 domain. The N3 domain is typically composed of β -barrels and α -helices, e.g. from 3 to 6 β -barrels and from 2 to 3 α -helices, such as 3 β -barrels and 2 α -helices as shown in FIG. 11 or 6 β -barrels and 3 α -helices as shown in FIG. 1B. The N3 domain may form a constriction in the lumen of the pore. The N3 domain may be modified
10 so that it does not constrict the pore. The N3 domain may be modified to increase or decrease the size of the constriction.

When used as a nanopore to detect or characterize an analyte, the central gate, cap gate and/or N3 constriction may function as a read-head, i.e. interaction of the analyte with one, two or all of the central gate, cap gate and N3 constriction may alter the signal obtained as an analyte
15 interacts with the pore and thus enable information about the analyte to be derived. Accordingly, the secretin nanopore may comprise one, two or three read-heads.

The amino acid modifications can be selected to improve translocation of an analyte through the modified secretin nanopore, to improve capture of an analyte into the modified secretin nanopore, and/or improve signal quality during detection of an analyte as it moves
20 through the nanopore. Examples of the amino acid modifications are described in detail in the section "Modified secretin nanopore subunit polypeptide" above. While a modified secretin nanopore generally comprises one or more amino acid modifications (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more and up to 40 amino acid modifications) of a luminal surface, it should be appreciated that a modified
25 secretin nanopores may have any of a variety of different modifications. For example, a modified secretin nanopore may have amino acid modifications (luminal or non-luminal) (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, or more and up to 100 amino acid modifications) that promote membrane integration, promote oligomerization, promote subunit synthesis, promote nanopore stability, promote analyte capture, promote analyte release, improve
30 analyte detection, facilitate polymer analysis (e.g., polynucleotide sequences), *etc.*

By way of example only, FIG. 5 shows that an enzyme may interact with CsgG and InvG nanopores in different orientations due to the larger *cis*-opening of the InvG nanopore. Without wishing to be bound by theory, due to the size difference of the enzyme and the nanopore opening (also see FIG. 6), the enzyme may wedge into the nanopore. Similar to CsgG nanopores
35 of which the *cis*-opening was engineered to improve its interaction with an enzyme such as a

polynucleotide binding problem, in some embodiments, the modified secretin nanopores described herein (*e.g.*, the *cis*-opening or capture portion as described herein) can be engineered to facilitate a preferred orientation of an enzyme (*e.g.*, a polynucleotide binding protein) such that it reduces the noise and improves the signal and accuracy.

5 In some embodiments, the *cis*-opening may have a diameter of at least about 30 Å, at least about 40 Å, at least about 50 Å, at least about 60 Å, at least about 70 Å, at least about 80 Å, at least about 90 Å, at least about 100 Å, or higher. In some embodiments, the *cis*-opening may have a diameter of no more than about 150 Å, no more than about 140 Å, no more than about 130 Å, no more than about 120 Å, no more than about 110 Å, no more than about 100 Å, no
10 more than about 90 Å, no more than about 80 Å, no more than about 70 Å, no more than about 60 Å, no more than about 50 Å, or lower. Combinations of the above-referenced ranges are also possible. For example, in some embodiments, the *cis*-opening may have a diameter in a range of about 30 Å to about 120 Å. In some embodiments, the *cis*-opening may have a diameter in a range of about 60 Å to about 120 Å. In some embodiments, the *cis*-opening may have a diameter
15 in a range of about 60 Å to about 100 Å. In some embodiments, the *cis*-opening may have a diameter in a range of about 30 Å to about 80 Å. In one embodiment, the *trans*-opening may have a diameter of about 80 Å.

In some embodiments, the *trans*-opening may have a diameter of at least about 30 Å, at least about 40 Å, at least about 50 Å, at least about 60 Å, at least about 70 Å, at least about 80 Å,
20 at least about 90 Å, at least about 100 Å, or higher. In some embodiments, the *trans*-opening may have a diameter of no more than about 150 Å, no more than about 140 Å, no more than about 130 Å, no more than about 120 Å, no more than about 110 Å, no more than about 100 Å, no more than about 90 Å, no more than about 80 Å, no more than about 70 Å, no more than about 60 Å, no more than about 50 Å, or lower. Combinations of the above-referenced ranges are
25 also possible. For example, in some embodiments, the *trans*-opening may have a diameter in a range of about 30 Å to about 100 Å. In some embodiments, the *trans*-opening may have a diameter in a range of about 40 Å to about 100 Å. In some embodiments, the *trans*-opening may have a diameter in a range of about 60 Å to about 100 Å. In some embodiments, the *trans*-opening may have a diameter in a range of about 30 Å to about 80 Å. In one embodiment, the
30 *trans*-opening may have a diameter of about 80 Å.

In some embodiments, the luminal surface may further define a constriction within the lumen. The diameter of the lumen can vary along an axis that extends between the *cis*-opening and *trans*-opening of the nanopore. As an illustration only, Fig. 3 shows the radius profile of the lumen of an InvG nanopore along the nanopore axis (extending between the *cis*-opening and
35 *trans*-opening), in which the lumen comprises a constriction. As used herein, the term

"constriction" refers to a portion of the lumen having a diameter that is smaller than the diameter of both the *cis*-opening and the *trans*-opening. For example, the constriction may have a diameter that is about 5%-20% (inclusive) of the diameter of the *cis*-opening and/or the diameter of the *trans*-opening. For example, in some embodiments, the constriction may have a diameter of at least about 5 Å, at least about 6 Å, at least about 7 Å, at least about 8 Å, at least about 9 Å, at least about 10 Å, at least about 15 Å, at least about 20 Å, at least about 25 Å, or higher. In some embodiments, the constriction may have a diameter of no more than about 30 Å, no more than about 25 Å, no more than about 20 Å, no more than about 15 Å, no more than about 10 Å, or lower. Combinations of the above-referenced ranges are also possible. For example, in some embodiments, the constriction may have a diameter in a range of about 5 Å to about 25 Å. In some embodiments, the constriction may have a diameter in a range of about 7 Å to about 25 Å. In some embodiments, the constriction may have a diameter in a range of about 10 Å to about 25 Å. In one embodiment, the constriction may have a diameter of about 15 Å.

The constriction may be located about halfway between the *cis*-opening and *trans*-opening. In some embodiments, the constriction may be located at a distance of about 30 Å to about 60 Å away from the *cis*-opening. In some embodiments, the constriction may be located at a distance of about 30 Å to about 60 Å away from the *trans*-opening.

In some embodiments, the modified secretin nanopores described herein may comprise a luminal surface defining a lumen that exhibits the radius profile of a natural secretin nanopore, for example, as shown in FIG. 3.

Any forms of secretin found in a microorganism (*e.g.*, bacteria) may be used to produce the modified secretin nanopore described herein. In some embodiments, the secretin may be any member of a type II, type III, or type IV secretion system. Non-limiting examples of a type II secretion system include GspD, PulD, and pIV,. Examples of a type III secretion system include, but are not limited to InvG, MxiD, YscC, PscC, EscC, and SpiA. An exemplary type IV secretion system includes, but is not limited to PilQ. Accordingly, in some embodiments, the modified secretin nanopore may comprise any embodiment of a modified secretin subunit polypeptide described herein, *e.g.*, in the section "Modified secretin nanopore subunit polypeptide" above.

In some embodiments, one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) of the amino acid modifications described herein (*e.g.*, but not limited to a positively-charged amino acid substitution and/or hydrophobic amino acid substitution) may be present in a portion of the luminal surface that defines the constriction. For example, one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) of the amino acid modifications described herein (*e.g.*, but not limited to a positively-charged amino acid substitution and/or hydrophobic amino acid substitution) may be

present in the portion of the luminal surface that defines the constriction of a modified secretin nanopore, *e.g.*, a modified InvG nanopore. As an example only, FIG. 1A shows the location of a constriction (labelled as "periplasmic gate" in the figure) of a wild-type InvG nanopore. In some embodiments, the constriction of the modified InvG secretin nanopore may have one or more amino acid modifications for improving translocation of an analyte through the nanopore and/or improving detection signal quality as the analyte moves through the nanopore. For example, the constriction of the modified InvG secretin nanopore may comprise amino acid modifications at amino acids D28, E225, R226, and/or E231 of SEQ ID NO: 1. In some embodiments, the constriction of the modified InvG secretin nanopore may comprise one or more (*e.g.*, 1, 2, 3, 4, 5, or 6) of the following amino acid modifications: (i) D28N/Q/T/S/G/R/K; (ii) E225N/Q/T/A/S/G/P/H/F/Y/R/K; (iii) R226N/Q/T/A/S/G/P/H/F/Y/K/V; (iv) deletion of E225; (v) deletion of R226; and (vi) E231N/Q/T/A/S/G/P/H/R/K.

In some embodiments, the luminal surface may further comprise a capture portion (*e.g.*, an analyte capture portion (*e.g.*, a polynucleotide capture portion)). As used herein, the term "capture portion" refers to a portion of a luminal surface of a nanopore that favourably interacts, via one or more amino acids of one or more pore subunits, with a target analyte to permit or facilitate binding of the analyte to, and/or translocation of the analyte through, the nanopore. The capture portion may be located between the *cis*-opening and the constriction of the modified secretin nanopore. In some embodiments, the capture portion may correspond to a N3 domain of a secretin nanopore (*e.g.*, a type II, III, or IV secretion system). For example, the capture portion may correspond to a N3 domain of an InvG nanopore, *e.g.*, as shown in FIG. 1A, or a portion of such a domain. FIG. 1B shows the peptide domains (with corresponding amino acid positions in SEQ ID NO: 2) that encompass the N3 domain of an InvG nanopore. In some embodiments, a capture portion of a luminal surface comprises one or more amino acids of one or more pore subunits (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or more amino acids) on a *cis*-opening side of a constriction.

In some embodiments, the capture portion may correspond to a N3 domain of an InvG nanopore, *e.g.*, as shown in FIG. 1A, or a portion of such a domain and include a "periplasmic constriction" as shown in FIG. 1A, which may act like a second constriction. Thus, in some embodiments, the modified secretin nanopore (*e.g.*, a modified InvG nanopore) may comprise two constrictions – one located about halfway between the *cis*-opening and *trans*-opening as described above and another located close to the *cis*-opening of the nanopore. Such a modified secretin nanopore may act like a two reader nanopore in which an analyte (*e.g.*, a polynucleotide) interacts with the pore lumen at the two constriction sites that are distant from each other.

In some embodiments, the capture portion of the luminal surface may comprise one or more amino acid modifications (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, or more and up to 25 amino acid modifications) for improving capture of a target analyte, *e.g.*, a target polynucleotide. By way of example only, the capture portion of the modified InvG secretin nanopore may comprise amino acid modifications at amino acids E41, Q45, and/or E114 of SEQ ID NO: 1. In some embodiments, the capture portion of the modified InvG secretin nanopore may comprise one or more (*e.g.*, 1, 2, or 3) of the following amino acid modifications: (i) Q45R/K; (ii) E41N/Q/T/S/G/R/K; and (iii) E114N/Q/T/S/G/R/K.

Any of the modified secretin nanopores described herein can be homo-multimeric (*e.g.*, all subunits within the nanopore are the same) or hetero-multimeric (*e.g.*, at least one subunit is different from others within the nanopore). The modified secretin nanopore may comprise any number of subunit polypeptides that are sufficient to form a lumen large enough to permit a target polymer (*e.g.*, polynucleotide) pass through. In some embodiments, the modified secretin nanopore may comprise about 9 to about 20 subunit polypeptides (*e.g.*, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 subunit polypeptides), wherein at least one or more of the subunit polypeptides comprises one or more amino acid substitutions (*e.g.*, positively-charged amino acid substitutions and/or hydrophobic amino acid modifications) as described herein.

The modified secretin nanopores may be isolated, substantially isolated, purified or substantially purified. The modified secretin nanopores can be 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, one or more of the modified secretin nanopores may be present in a membrane. Suitable membranes are discussed below.

The modified secretin nanopore may be present as an individual or single pore. Alternatively, the modified secretin nanopores may be present in a homologous or heterologous population of two or more pores. In some embodiments, the modified secretin nanopores may be arranged in an array, *e.g.*, each nanopore disposed in a membrane present in a microwell. In some embodiments, the array may comprise the modified secretin nanopores and at least one or more non-secretin nanopore known in the art, *e.g.*, but not limited to CsgG nanopores (*e.g.*, as described in WO 2016/034591); α -hemolysin nanopores (*e.g.*, as described in WO 2010/004273); lysenin nanopores (*e.g.*, as described in WO 2013/153359); Msp nanopores (*e.g.*, as described in WO 2012/107778; WO 2015/166275; and WO 2016/055778).

The modified secretin nanopores described herein can provide improved analyte detection and/or analysis. For illustration only, FIG. 4 shows that while both CsgG and InvG nanopores have a constriction of approximately the same in diameter, the constriction of the CsgG nanopore has 3 amino acids at positions 51, 55, and 56 (based on wild type sequence),
 5 respectively, and the InvG nanopore constriction has two amino acids at position 396 and 397 (based on SEQ ID NO: 2), respectively. Further, the amino acid 51 at the constriction of the CsgG nanopore is also a little far from amino acid 55. In contrast, the amino acids 396 and 397 at the constriction of the InvG nanopore are located next to each other, thus providing a sharper reader head. Therefore, in some embodiments, the modified secretin nanopores can provide a
 10 sharper reader head for analyte detection and/or analysis.

Homo-multimeric secretin nanopores

Homo-multimeric nanopores comprising identical modified secretin nanopore subunit polypeptides are also provided herein. The homo-multimeric nanopore may comprise any
 15 embodiment of the modified secretin nanopore subunit polypeptides described herein. The homo-multimeric nanopore can be used for characterizing an analyte, *e.g.*, a target polynucleotide and/or a target polypeptide. The homo-multimeric nanopore described herein may have any of the advantages discussed above.

The homo-multimeric pore may contain any number of modified secretin nanopore
 20 subunit polypeptides. The pore typically comprises at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20 identical modified secretin nanopore subunit polypeptides, such as 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 identical modified secretin nanopore subunit polypeptides.

Hetero-multimeric secretin nanopores

Hetero-multimeric nanopores comprising at least one modified secretin nanopore subunit polypeptides are also provided herein. The hetero-multimeric nanopores can be used for characterizing a target analyte, *e.g.*, a target polynucleotide and/or a target polypeptide. Hetero-multimeric nanopores can be made using methods known in the art (*e.g.*, Protein Sci. 2002 Jul;
 30 11(7):1813-24).

The hetero-multimeric pore contains sufficient subunit polypeptide to form the pore. The subunit polypeptides may be of any type. The pore typically comprises at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20 subunit polypeptides, such as 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
 35 subunit polypeptides.

In some embodiments, all of the subunit polypeptides (such as 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 subunit polypeptides) are modified secretin nanopore subunit polypeptides and at least one of them differs from the others.

In some embodiments, at least one of the subunit polypeptides is not a modified secretin nanopore subunit polypeptide as described herein. In this embodiment, the remaining monomers may be any one of the modified secretin nanopore subunit polypeptides described herein. Hence, the pore may comprise 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 modified secretin nanopore subunit polypeptide(s). The modified secretin nanopore subunit polypeptide(s) that form the nanopore can be the same or different.

Exemplary uses of the secretin nanopores described herein

The modified secretin nanopores can be used for characterizing or detecting an analyte, *e.g.*, a target polynucleotide (*e.g.*, a double stranded polynucleotide and/or a single stranded polynucleotide) and/or a target polypeptide. Accordingly, methods for detecting and/or characterizing an analyte in a sample are also provided herein. The method comprises: providing an aqueous solution comprising any embodiment of the modified secretin nanopores described herein and a membrane, wherein the modified secretin nanopore is disposed in the membrane; and adding an analyte to the aqueous solution on the *cis*-side or *trans*-side of the membrane. In some embodiments, an enzyme such as a polynucleotide binding protein, *e.g.*, helicases, exonucleases, and/or polymerase, can also be added to the aqueous solution on the *cis*-side or *trans*-side of the membrane. The enzyme such as a polynucleotide binding protein may enter the lumen or be in contact (via, *e.g.*, but not limited to ionic and/or hydrophobic interactions) or covalently attached to the *cis*-opening or *trans*-opening, of the modified secretin nanopores. In some embodiments, the analyte may bind to the enzyme such as a polynucleotide binding protein. An analyte may be a target polynucleotide, polypeptide, ligand, or hydrophobic molecule.

In some embodiments, the secretin nanopores may be used to detect molecules that bind to or otherwise interact with an enzyme provided within the *cis* or *trans* vestibule that give rise to a change in conformation of the enzyme. The change in conformation can give rise to a change in ion current flow through the nanopore. Examples of such molecules are drugs, antibodies, peptides, polynucleotides and so on. Examples of enzymes that interact with small molecules such as drugs include but are not limited to Cytochrome p450 enzymes.

In some embodiments, the method may further comprise applying a potential across the membrane. The applied potential may be a voltage potential. Alternatively, the applied potential may be a chemical potential. An example of this is using a salt gradient across a membrane,

such as an amphiphilic layer. A salt gradient is disclosed in Holden et al., J Am Chem Soc. 2007 Jul 11; 129(27):8650-5. The method may be carried out with a voltage applied across the membrane and nanopore. The voltage used may vary from +5 V to -5 V, such as from +4 V to -4 V, +3 V to -3 V or +2 V to -2 V. In some embodiments, the voltage used may be from -600 mV to +600mV or -400 mV to +400 mV. In some embodiments, the voltage used may be 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. In some embodiments, the voltage used may be in the range of 100 mV to 240 mV or 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.

In some embodiments, the method may further comprise, upon application of a potential across the membrane, detecting a signal in response to an analyte passing through the nanopore. The signal may be an electrical measurement and/or an optical measurement. Possible electrical measurements include: current measurements, impedance measurements, tunnelling measurements (Ivanov AP et al., Nano Lett. 2011 Jan 12;11(1):279-85), and FET measurements (International Application WO 2005/124888). Optical measurements may be combined with electrical measurements (Soni GV et al., Rev Sci Instrum. 2010 Jan;81(1):014301). The measurement may be a transmembrane current measurement such as measurement of ionic current flowing through the pore. Alternatively the measurement may be a fluorescence measurement indicative of ion flow through the channel such as disclosed by Heron et al, J. Am. Chem. Soc., 2009, 131 (5), 1652-1653 or measurement of a voltage across the membrane using a FET. In some embodiments, the method may further comprise, upon application of a potential across the membrane, detecting an ionic current flow through the nanopore as an analyte (*e.g.*, but not limited to a target polynucleotide) interacts and/or moves through the nanopore. In some embodiments, the methods may be carried out using a patch clamp or a voltage clamp. In some embodiments, the methods may be carried out using a voltage clamp. Electrical measurements may be made using standard single channel recording equipment as describe in Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7, Lieberman KR et al, J Am Chem Soc. 2010;132(50):17961-72, and International Application WO 2000/28312. Alternatively, electrical measurements may be made using a multi-channel system, for example as described in International Application WO 2009/077734 and International Application WO 2011/067559.

In alternative embodiments, the method may further comprise, upon application of a potential across the membrane, detecting an analyte by measuring the movement or conformational change of an enzyme (*e.g.*, a polynucleotide binding protein or a ligand binding

protein) upon binding to the analyte. In some embodiments, at least a portion of the enzyme may reside within the lumen of the modified secretin nanopore when the analyte is bound to the enzyme. In these embodiments, an ionic current passing through the nanopore may vary with the movement or conformational change of the enzyme bound to an analyte, as compared to an enzyme with no analyte bound thereto. Thus, the presence and/or type of an analyte can be detected by measuring changes in the level of the ionic current and/or current signature generated across the nanopore.

In any of the methods described herein, the aqueous solution in which the modified secretin nanopore and the membrane are disposed may comprise 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 herein, 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 may be used. 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.

In any of the methods described herein, the aqueous solution in which the modified secretin nanopore and the membrane are disposed may comprise salt. 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.

In some embodiments, the aqueous solution may be a low ionic strength solution. As used herein, the term "low ionic strength solution" refers to a solution with an ionic strength of less than 2 M, including, *e.g.*, less than 1 M, less than 900 mM, less than 800 mM, less than 700 mM, less than 600 mM, less than 500 mM, less than 400 mM, less than 300 mM, less than 200 mM, less than 150 mM, or lower. In some embodiments, a lower ionic strength solution has an ionic strength of at least about 50 mM, at least about 100 mM, at least about 150 mM, at least

about 200 mM, at least about 300 mM, at least about 400 mM, at least about 500 mM, at least about 600 mM, at least about 700 mM, at least about 800 mM, at least about 900 mM, at least about 1 M, or higher. Combinations of the above-references ranges are also encompassed. For example, a low ionic strength solution may have an ionic strength of about 100 mM to about 600 mM, or about 150 mM to about 300 mM. Any salt can be used to yield a solution with appropriate ionic strength. In some embodiments, alkaline salt (*e.g.*, but not limited to potassium chloride or sodium chloride) can be used in the low ionic strength solution.

The methods described herein are typically carried out in the presence of a buffer. In the exemplary apparatus discussed herein, the buffer is present in the aqueous solution in the chamber. Any buffer may be used in the methods described herein. 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 or 8.0.

The methods described herein 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.

In some embodiments, the methods described herein can be used to discriminate between different nucleotides under a range of conditions, which is further described in detail in the section "Polynucleotide characterization" below. For example, the methods described herein can be used to discriminate between nucleotides under conditions that are favourable to the characterizing, such as sequencing, of nucleic acids. The extent to which the modified secretin nanopores used in the methods 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 modified secretin nanopores may also be used to identify polynucleotide polymers from the interaction with one or more monomers rather than on a nucleotide by nucleotide basis. In some embodiments, the modified secretin nanopores can also be used to distinguish modified bases, *e.g.*, between methylated and unmethylated nucleotides.

FIG. 2 shows that while a CsgG nanopore has 9 monomers or subunits and an InvG nanopore has 15 monomers or subunits, both nanopores have a constriction of approximately the same in diameter. Unlike CsgG nanopores (*e.g.*, as described in WO 2016/034591), in some embodiments, the modified secretin nanopores (*e.g.*, but not limited to InvG nanopores) can be used to sequence DNA and/or RNA.

In some embodiments, the methods described herein can be used to characterize and/or detect or characterize a molecule or a ligand. For example, the modified secretin nanopores used in the methods described herein may be used for characterizing ligand-enzyme interactions (*e.g.*, nucleic acid-protein interactions or protein-protein interactions). In some embodiments, the nanopores can be used to interrogate ligand-enzyme interactions (*e.g.*, protein-nucleic acid interaction or protein-protein interaction) using different sensing modes such as, for example, by scanning and mapping the locations of binding sites along a ligand (*e.g.*, nucleic acid or polypeptide) and/or by probing the strength of interactions between a ligand and an enzyme (*e.g.*, between a protein and nucleic acid or between a protein and a protein). In some embodiments, native charges of a nucleic acid or protein may be leveraged to apply an electrophoretic force to a nucleic acid-protein complex or a protein-protein complex. For example, in some embodiments, DNA-protein interactions may be evaluated using voltage-driven threading of single DNA molecules through a protein nanopore. In such embodiments, electrical force applied to an individual DNA-protein complex (*e.g.*, a DNA-exonuclease I complex, a DNA-helicase complex, a DNA-clamp complex) may pull the two molecules apart, while at the same time ion current changes may be used to evaluate the dissociation rate of the complex. In some embodiments, modified secretin nanopores provided herein may be used for detection and characterization of nucleic acid-protein interactions involving nucleic acid and other nucleic acid binding proteins such as transcription factors, enzymes, DNA packaging proteins and others. In some embodiments, modified secretin nanopores provided herein may be used for detection and characterization of protein-protein interactions involving a ligand and other ligand binding proteins.

In some embodiments, at least a portion of an enzyme (*e.g.*, but not limited to polynucleotide binding protein) can enter the lumen of the modified secretin nanopores, for example, as shown in FIG. 6. Localization of the enzyme inside the nanopore may restrict undesirable movements of the enzyme and thus result in improved signals. For example, as shown with ClyA nanopores (*e.g.*, as described in International Patent Application Publications WO 2014/153625 and WO 2016/166232), the modified secretin nanopores as described herein, in some embodiments, can be used to detect an analyte by measuring the movement of its binding to an enzyme, at least a portion of which is present inside the nanopore. Since the constriction of secretin nanopores such as InvG nanopores is much smaller than that of ClyA nanopores, signal generating from such an event may be more pronounced with secretin nanopores such as InvG nanopores. Thus, in some embodiments, the modified secretin nanopores and the methods described herein can provide a new area of molecular testing.

Polynucleotide characterization

Another aspect of the present disclosure provides a method of characterizing a target polynucleotide. The method comprises: (a) providing in an aqueous solution a modified secretin nanopore according to any embodiment described herein and a membrane, wherein the modified
 5 secretin nanopore is present in the membrane; (b) adding in the aqueous solution of step (a) the target polynucleotide; and (c) measuring, during application of a potential across the nanopore, ion flow through the modified secretin nanopore, wherein the ion flow measurements are indicative of one or more characteristics of the target polynucleotide. In some embodiments, the target polynucleotide is added to the *cis* side of the aqueous solution. In some embodiments, the
 10 target polynucleotide is added to the *trans* side of the aqueous solution. In some embodiments, the aqueous solution is present in an embodiment of an apparatus described herein.

The target polynucleotide may also be called the template polynucleotide or the polynucleotide of interest.

Polynucleotide

15 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
 20 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.
 25 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,
 30 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
 35 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.

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. At least a portion of the polynucleotide is preferably double stranded.

The polynucleotide can be a nucleic acid, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The polynucleotide can comprise one strand of RNA hybridized to one strand of DNA. The polynucleotide may be any synthetic nucleic acid known in the art, such as peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or other synthetic polymers with nucleotide side chains. The PNA backbone is composed of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds. The GNA backbone is composed of repeating glycol units linked by phosphodiester bonds. The TNA backbone is composed of repeating threose sugars linked together by phosphodiester bonds. LNA is formed from ribonucleotides as discussed above having an extra bridge connecting the 2' oxygen and 4' carbon in the ribose moiety.

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

The polynucleotide can be any length. For example, the polynucleotide can be at least 10, at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 400 or at least 500 nucleotides or nucleotide pairs in length. The polynucleotide can be 1000 or more nucleotides or nucleotide pairs, 5000 or more nucleotides or nucleotide pairs in length or 100000 or more nucleotides or nucleotide pairs in length.

Any number of polynucleotides can be investigated. For instance, the method described herein may concern characterizing 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100 or more

polynucleotides. If two or more polynucleotides are characterized, they may be different polynucleotides or two instances of the same polynucleotide.

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

The polynucleotide may comprise an attached species such as a protein or analyte. The polynucleotide may comprise a hybridized probe.

Characterization

The method for polynucleotide characterization 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 methods described herein, such as {i}, {ii}, {iii}, {iv}, {v}, {i,ii}, {i,iii}, {i,iv}, {i,v}, {ii,iii}, {ii,iv}, {ii,v}, {iii,iv}, {iii,v}, {iv,v}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,iii,iv}, {i,iii,v}, {i,iv,v}, {ii,iii,iv}, {ii,iii,v}, {ii,iv,v}, {iii,iv,v}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iv,v}, {i,iii,iv,v}, {ii,iii,iv,v} or {i,ii,iii,iv,v}. Different combinations of (i) to (v) may be measured for the first polynucleotide compared with the second polynucleotide, including any of those combinations listed above.

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

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

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

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

5 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
10 cytosine on the basis of the current flowing through the pore during its interaction with each nucleotide.

The target polynucleotide is contacted with any one of the modified secretin nanopores described 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
15 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
20 the pore is present.

The method may be carried out using the apparatus described in International Application No. PCT/GB08/000562 (published as WO 2008/102120), the contents of which are incorporated herein by reference.

A variety of different types of measurements may be made. This includes without
25 limitation: electrical measurements and optical measurements. Possible electrical measurements include: current measurements, impedance measurements, tunneling measurements (Ivanov AP et al., Nano Lett. 2011 Jan 12;11(1):279-85), and FET measurements (International Application WO 2005/124888). Optical measurements may be combined with electrical measurements (Soni GV et al., Rev Sci Instrum. 2010 Jan;81(1):014301). The measurement may be a transmembrane
30 current measurement such as measurement of ionic current flowing through the pore. Alternatively the measurement may be a fluorescence measurement indicative of ion flow through the channel such as disclosed by Heron et al, J. Am. Chem. Soc., 2009, 131 (5), 1652-1653 or measurement of a voltage across the membrane using a FET.

Electrical measurements may be made using standard single channel recording

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

The method can be carried out with a potential applied across the membrane. The applied potential may be a voltage potential. Alternatively, the applied potential may be a chemical potential. An example of this is using a salt gradient across a membrane, such as an amphiphilic layer. A salt gradient is disclosed in Holden et al., J Am Chem Soc. 2007 Jul 11;
 10 129(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 may be described as strand sequencing.

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
 15 also comprise an electrical circuit capable of applying a potential and measuring an electrical signal across the membrane and pore. The methods may be carried out using a patch clamp or a voltage clamp. The methods preferably involve the use of a voltage clamp.

The method 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
 20 through transmembrane protein pores are known in the art and also provided herein.

Enzymes such as Polynucleotide Binding Protein

In some embodiments, the method for characterizing an analyte (*e.g.*, a target polynucleotide or polypeptide) may include adding an enzyme such as a polynucleotide binding
 25 protein in an aqueous solution comprising an analyte such that the enzyme binds to the analyte (*e.g.*, target polynucleotide or polypeptide). In some embodiments, the binding of the analyte (*e.g.*, target polynucleotide) to the enzyme such as a polynucleotide binding protein controls the movement of the analyte (*e.g.*, target polynucleotide) through the modified secretin nanopore, thereby characterizing the analyte (*e.g.*, target polynucleotide). In some embodiments, the
 30 movement of an analyte (*e.g.*, target polypeptide or ligand) binding to an enzyme such as a ligand-binding protein can be measured to detect the analyte and/or characterize the interaction of the analyte with the enzyme.

Polynucleotide binding protein: The polynucleotide binding protein may be any protein that is capable of binding to the polynucleotide and controlling its movement through the pore.
 35 Examples of the polynucleotide binding proteins include, but are not limited to helicases,

polymerases, exonucleases, DNA clamps, etc. 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.

Helicase(s) and molecular brake(s)

In one embodiment, the method comprises:

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

(b) adding the polynucleotide in the low ionic strength solution that comprises a modified secretin nanopore present in a membrane, and applying a potential across the pore such that the one or more helicases and the one or more molecular brakes are brought together and both control the movement of the polynucleotide through the pore;

(c) measuring, during application of a potential across the nanopore, ion flow through the modified secretin nanopore, as the polynucleotide moves with respect to the pore wherein the ion flow measurements are indicative of one or more characteristics of the polynucleotide and thereby characterizing the polynucleotide. This type of method is discussed in detail in International Application No. PCT/GB2014/052737 (published as WO 2015/110777), the contents of which are incorporated herein by reference.

Membrane

The modified secretin nanopores described herein may be present in a membrane. In the method of characterizing an analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand), the analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand) is typically contacted with a modified secretin nanopore 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 or 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.

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 synthesized, the exact construction can be carefully controlled to provide the correct chain lengths and properties to form membranes and to interact with pores and other proteins.

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

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

The membrane is most preferably one of the membranes disclosed in International Application No. PCT/GB2013/052766 (published as WO 2014/064443) or PCT/GB2013/052767 (published as WO 2014/064444), the contents of each of which are incorporated herein by reference.

5 The amphiphilic molecules may be chemically-modified or functionalized to facilitate coupling of the analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand).

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.

10 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 analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand) can typically move within an amphiphilic membrane.

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

25 In some embodiments, the analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand) can be coupled to the membrane comprising any one of the modified secretin nanopores described herein. The method may comprise coupling the analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand) to the membrane comprising any one of the modified secretin nanopores described herein. The analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand) is preferably coupled to the membrane using one or more anchors. The analyte (*e.g.*, a target polynucleotide, polypeptide, or a ligand) may be coupled to the membrane using any known method.

Double stranded polynucleotide sequencing

In some embodiments, the polynucleotide may be double stranded. If the polynucleotide is double stranded, the method may further comprises before the contacting step ligating a hairpin adaptor to one end of the polynucleotide. The two strands of the polynucleotide may then be separated as or before the polynucleotide is contacted or interacted with a modified secretin nanopore as described herein. The two strands may be separated as the polynucleotide movement through the pore is controlled by a polynucleotide binding protein, such as a helicase, or molecular brake. This is described in International Application No. PCT/GB2012/051786 (published as WO 2013/014451), the contents of which are incorporated herein by reference. Linking and interrogating both strands on a double stranded construct in this way increases the efficiency and accuracy of characterization.

Round the corner sequencing

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

Leader sequence

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

Modified polynucleotides

Before characterization, a target polynucleotide may be modified by contacting the polynucleotide with a polymerase and a population of free nucleotides under conditions in which the polymerase forms a modified polynucleotide using the target polynucleotide as a template, wherein the polymerase replaces one or more of the nucleotide species in the target polynucleotide with a different nucleotide species when forming the modified polynucleotide.

The modified polynucleotide may then be provided with one or more helicases attached to the polynucleotide and one or more molecular brakes attached to the polynucleotide. This type of modification is described in International Application No. PCT/GB2015/050483, the contents of which are incorporated herein by reference. Any of the polymerases discussed herein may be used.

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®. A primer or a 3' hairpin is typically used as the nucleation point for polymerase extension.

Characterization, 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), the contents of which are incorporated herein by reference. 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. Various algorithms may be employed to characterize the sequence, such as use of a Hidden Markov Model or recurrent neural network. The sequence may be aligned to a reference sequence using methods such as disclosed in International Patent Application Nos. PCT/GB2015/050776 (published as WO 2015/140535) and PCT/GB2015/053083 (published as WO 2016/059427), the contents of each of which are incorporated herein by reference.

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 characterize the target polynucleotide. In some instances, the modified polynucleotide itself may be easier to characterize. 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

5 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

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

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

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

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

Other characterization method

In another embodiment, a polynucleotide is characterized by detecting labelled species

30 that are released as a polymerase incorporates nucleotides into the polynucleotide. The polymerase uses the polynucleotide as a template. Each labelled species is specific for each nucleotide. The polynucleotide is contacted with a modified secretin nanopore described herein, a polymerase and labelled nucleotides such that phosphate labelled species are sequentially released when nucleotides are added to the polynucleotide(s) by the polymerase, wherein the

35 phosphate species contain a label specific for each nucleotide. The polymerase may be any of

those discussed above. The phosphate labelled species are detected using the pore and thereby characterizing the polynucleotide. This type of method is disclosed in European Application No. 13187149.3 (published as EP 2682460). Any of the embodiments discussed above equally apply to this method.

5

Sample

Any suitable sample comprising an analyte to be detected or characterized may be subjected to any of the methods described herein. The methods described herein can be carried out on two or more samples that are known to contain or suspected to contain the analytes.

10 Alternatively, the method may be carried out on two or more samples to confirm the identity of two or more analytes whose presence in the samples is known or expected. In some embodiments, the method may be carried out on samples to distinguish double stranded polynucleotides from single-stranded polynucleotides.

The first sample and/or second sample may be a biological sample. The methods
15 described herein may be carried out in vitro using at least one sample obtained from or extracted from any organism or microorganism. The first sample and/or second sample may be a non-biological sample. The non-biological sample can be 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.

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

25

Kits

Another aspect of the present disclosure also provides a kit, for example, for characterizing a target analyte such as a target polynucleotide, polypeptide, or ligand. The kit comprises any one of the modified secretin nanopores described herein and the components of a
30 membrane. The membrane is preferably formed from the components. The modified secretin nanopore 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 an enzyme such as a polynucleotide binding protein or a ligand binding protein.

The kit may further comprise one or more anchors for coupling the analyte (*e.g.*, polynucleotide, polypeptide, or ligand) 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
 5 may include one or more of the following: suitable buffer(s) (aqueous solutions), means to obtain a sample, *e.g.*, from a subject (such as a vessel or an instrument comprising a needle), means to amplify polynucleotides and/or express proteins or polypeptides, 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
 10 any one of the methods described herein or details regarding for which organism the method may be used.

Apparatus

Another aspect described herein also provides an apparatus, for example, for
 15 characterizing a target analyte such as a polynucleotide, polypeptide, or ligand. The apparatus comprises a plurality of modified secretin nanopores as described herein and a plurality of membranes. In some embodiments, the plurality of the modified secretin nanopores are present in the plurality of membranes. In some embodiments, the numbers of modified secretin nanopores and membranes are equal. In one embodiment, a single modified secretin nanopore is
 20 present in each membrane.

In some embodiments, an apparatus comprises a chamber (*e.g.*, a microwell) containing an aqueous solution having disposed therein a membrane comprising a modified secretin nanopore as described herein. In some embodiments, an apparatus may comprise an array of chambers (*e.g.*, an array of microwells), each of which contains an aqueous solution having
 25 disposed therein a membrane comprising a modified secretin nanopore as described herein. In some embodiments, an apparatus may comprise an array of chambers (*e.g.*, an array of microwells), each of which contains an aqueous solution having disposed therein a membrane comprising a nanopore. In these embodiments, at least one nanopore is a modified secretin nanopore as described herein, and the remaining nanopores may be a non-secretin nanopore
 30 known in the art, *e.g.*, but not limited to CsgG nanopores (*e.g.*, as described in WO 2016/034591); α -hemolysin nanopores (*e.g.*, as described in WO 2010/004273); lysenin nanopores (*e.g.*, as described in WO 2013/153359); Msp nanopores (*e.g.*, as described in WO 2012/107778; WO 2015/166275; and WO 2016/055778). Thus, more than one type of nanopores can be present in such an array.

In some embodiments, the apparatus may further comprise an analyte in the aqueous solution. In some embodiments where the analyte is a polynucleotide, the apparatus may further comprise a polynucleotide binding protein, *e.g.*, a helicase, exonuclease, or polymerase. The polynucleotide binding protein may be bound to the polynucleotide. In some embodiments, the polynucleotide binding protein may be on the *cis*-side of the membrane and the polynucleotide binding protein may be in contact (via *e.g.*, ionic and/or hydrophobic interactions) with or covalently attached to the *cis*-opening of the nanopore. In some embodiments, the polynucleotide binding protein may be on the *trans*-side of the membrane and the polynucleotide binding protein may be in contact (via *e.g.*, ionic and/or hydrophobic interactions) with or covalently attached to the *trans*-opening of the nanopore.

The apparatus can further comprises instructions for carrying out any of the methods as described herein. The apparatus may be any conventional apparatus for polynucleotide analysis, such as an array or a chip. Any of the embodiments discussed above with reference to the methods, *e.g.*, for characterizing a target polynucleotide, are equally applicable to the apparatus described herein. The apparatus may further comprise any of the features present in the kit described herein.

In some embodiments, the apparatus is set up to carry out any of the methods described herein, *e.g.*, for characterizing a target analyte such as a target polynucleotide.

In one embodiment, the apparatus comprises: (a) a sensor device that is capable of supporting the plurality of modified secretin nanopores and membranes and that is operable to perform polynucleotide characterization using the nanopores and membranes; and (b) at least one port for delivery of material for performing the characterization.

Alternatively, the apparatus may comprise: (a) a sensor device that is capable of supporting the plurality of modified secretin nanopores and membranes and that is operable to perform polynucleotide characterization using the nanopores and membranes; and (b) at least one reservoir for holding material for performing the characterization.

In another embodiment, the apparatus may comprise: (a) a sensor device that is capable of supporting the membrane and plurality of modified secretin nanopores and membranes and that is operable to perform polynucleotide characterizing using the pores and membranes; (b) at least one reservoir for holding material for performing the characterizing; (c) a fluidics system configured to controllably supply material from the at least one reservoir to the sensor device; and (d) one or more containers for receiving respective samples, the fluidics system being configured to supply the samples selectively from one or more containers to the sensor device.

The apparatus may be any of those described in International Application No. No. PCT/GB08/004127 (published as WO 2009/077734), PCT/GB10/000789 (published as WO

2010/122293), International Application No. PCT/GB10/002206 (published as WO 2011/067559) or International Application No. PCT/US99/25679 (published as WO 00/28312), the contents of each of which are incorporated herein by reference.

Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present disclosure to its fullest extent. The following specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. All publications cited herein are incorporated by reference for the purposes or subject matter referenced herein.

Example 1. Exemplary method for expression and purification of a modified secretin nanopore subunit polypeptide, e.g., a modified InvG nanopore subunit polypeptide

Ampicillin-resistant pT7 vector containing the gene encoding a modified secretin nanopore subunit polypeptide (*e.g.*, an amino acid sequence as set forth in SEQ ID NO: 1 or 2 with one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, or more and up to 50 amino acid modifications described herein)) with a C terminal hexa-histidine (His) tag and kanamycin-resistant pRham vector containing the gene encoding InvH protein (20Kd protein that would enhance the expression of InvG) were co-transformed into C43 DE3 pLysS cells and plated out on agar plates containing both Ampicillin (100µg/ml) and Kanamycin (30µg/ml) and grown overnight at 37°C. A single colony was used to inoculate a 100ml starter culture of TB media containing both Ampicillin (100µg/ml) and Kanamycin (30µg/ml). The culture was grown at 250rpm at 37°C for 18 hours. 15 ml of starter culture was used to inoculate 500ml of TB media containing both Ampicillin (100µg/ml) and Kanamycin (30µg/ml), and the culture was grown at 250 rpm at 37°C until OD at 600nm reached 0.6. The temperature was reduced to 18°C and the culture was allowed to equilibrate to the reduced temperature for 1 hour. IPTG was added to the final concentration of 0.5mM, and Rhamnose was added to 0.2% to induce protein production. The culture was allowed to incubate for 18 hrs at 250 rpm at 18°C. The culture was harvested by centrifugation at 6000g for 20 minutes. The cell pellet was lysed by resuspending in 7.5ml per 1 g pellet of 25mM HEPES, 500mM NaCl, 15mM Imidazole, Protease inhibitors, 25unit/ml Benzonase Nuclease, 0.01% DDM pH7.5 and mixed to homogeneity. The resuspended pellet was then lysed by sonication (15 cycles of 20 seconds on/20 seconds off for 15 cycles). The lysate was separated by centrifugation at 50,000g for 1 hour. The supernatant was filtered through a 0.22µm filter and applied to a 1mL His trap crude column. The protein was purified by the AKTA system as per manufacturer's instructions, using 25mM HEPES, 500mM NaCl, 15mM Imidazole, 0.01% DDM pH7.5 as the loading buffer; 25mM HEPES, 500mM NaCl,

75mM Imidazole, 0.01% DDM pH7.5 as the wash buffer; and 25mM HEPES, 500mM NaCl, 500mM Imidazole, 0.01% DDM pH7.5 as the elution buffer.

SDS Page was carried out to ascertain that the correct protein was present. Eluted fractions were then pooled and concentrated, for example, via 30kD MWCO Amicon spin column. The protein was carried forward for SEC chromatography of S200 increase column, as per manufacturer's instructions, using 25mM HEPES, 500mM NaCl, 0.001% DDM pH7.5 as buffer A. The protein of interest was eluted as a single peak in an appropriate molecular weight fraction. For example, a modified InvG nanopore subunit polypeptide may have a molecular weight of about 40 kDa to about 70 kDa (which can vary depending on the elution conditions of the SEC chromatography). The elution fractions were pooled and incubated with lecithin liposomes for 3 hours at 37°C with gentle mixing in a thermoshaker. The sample was then spun at 20,000G for 20 minutes, the supernatant discarded, and the pellet resuspended in 25mM HEPES, 500mM NaCl, 0.1% SDS pH7.5. Following resuspension, the sample was heated at 60°C for 15 minutes and spun at 20,000g for 10 minutes. The supernatant was carried forward for SEC on SW TOSOH G4000 column as per manufactures instructions to select for oligomer.

Example 2. Exemplary method for expression and purification of a modified secretin nanopore subunit polypeptide comprising an endopeptidase cleavage site

Ampicillin-resistant pT7 vector containing the gene encoding a modified secretin nanopore subunit polypeptide that comprises an endopeptidase cleavage site such as Tobacco Etch Virus (TEV) protease cleavage site (*e.g.*, an amino acid sequence as set forth in SEQ ID NO: 3 with one or more (*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, or more and up to 50 amino acid modifications described herein)) with a C terminal hexa-histidine tag was transformed into C43 DE3 pLysS cells and plated out on agar plates containing Ampicillin (100µg/ml). A single colony was used to inoculate a 100ml starter culture of TB media containing Ampicillin (100µg/ml). The culture was grown at 250 rpm at 37°C for 18 hours. 15 ml of starter culture was used to inoculate 500ml of TB media containing Ampicillin (100µg/ml) and the culture was grown at 250 rpm at 37°C until the OD at 600nm reached 0.6. The temperature was reduced to 18°C and the culture was allowed to equilibrate to the reduced temperature for 1 hour. IPTG was added to the final concentration of 0.5mM to induce protein production. The culture was allowed to incubate for 18hrs at 250 rpm at 18°C and harvested by centrifugation at 6000g for 20 minutes. The cell pellet was lysed by resuspending in 7.5ml per 1 g pellet of 25mM HEPES, 500mM NaCl, 15mM Imidazole, Protease inhibitors, 25 unit/ml Benzonase Nuclease, and 0.01% DDM pH 7.5, and mixed to homogeneity. The resuspended pellet was then lysed by sonication (15 cycles of 20 seconds on/20 seconds off for 15 cycles).

The lysate was separated by centrifugation at 50,000g for 1 hour. The supernatant was filtered through 0.22µm filter and applied to a 1ml His trap crude column. The protein was purified by the AKTA system as per manufacturer's instructions, using 25mM HEPES, 500mM NaCl, 15mM Imidazole, and 0.01% DDM pH 7.5 as the loading buffer; 25mM HEPES, 500mM NaCl, 75mM Imidazole, and 0.01% DDM pH 7.5 as the wash buffer, and 25mM HEPES, 500mM NaCl, 500mM Imidazole, and 0.01% DDM pH 7.5 as the elution buffer.

SDS Page was carried out to ascertain that the correct protein was present. Eluted fractions were then pooled and concentrated via a 30kD MWCO Amicon spin column. The protein was carried forward for SEC chromatography of S200 increase column as per manufacturer's instructions using 25mM HEPES, 500mM NaCl, and 0.001% DDM pH 7.5 as buffer A. The protein of interest was eluted as a single peak in an appropriate molecular weight fraction. For example, a modified InvG nanopore subunit polypeptide may have a molecular weight of about 40 kDa to about 70 kDa (which can vary depending on the elution conditions of the SEC chromatography). The elution fractions were pooled. His-tagged TEV Protease was added to a final concentration of 0.2mg/ml and the sample was allowed to incubate at 4°C for 18 hours to remove peptide domains that were located upstream of the endopeptidase cleavage site within the vector sequence (*e.g.*, N0 and N1 domains of an InvG protein) from the rest of the modified secretin nanopore subunit. The sample was reapplied to a trap column and the flow-through was collected. Flow-through fractions were incubated with lecithin liposomes for 3 hours at 37°C with gentle mixing in a thermoshaker. The sample was then spun at 20,000G for 20 minutes, the supernatant discarded, and the pellet resuspended in 25mM HEPES, 500mM NaCl, and 0.1% SDS pH 7.5. Following resuspension, the sample was heated at 60°C for 15 minutes and spun at 20,000g for 10 minutes. The supernatant was carried forward for SEC on a SW TOSOH G4000 column as per manufacturer's instructions to select for the oligomer.

Example 3. Design, expression and purification of GspD mutants

GspD mutants were designed as shown in the Tables below using the *Vibrio cholerae* GspD sequence shown in SEQ ID NO: 32 as the starting sequence.

Table 1: DNA Capture Mutants

Mutation	Mutant Position
GspD-Vch-(WT-E253Q/E257Q/E264Q/D290N-Del((N1-K239)/(N265-SGS-E282)))	N3 Domain
GspD- Vch-(WT-E253Q/E257Q/E264K/D290N-Del((N1-K239)/(N265-SGS-E282)))	N3 Domain
GspD- Vch-(WT-E253Q/E257K/E264Q-Del((N1-K239)/(N265-SGS-E282)))	N3 Domain
GspD- Vch-(WT-E257K/E264K-Del((N1-K239)/(N265-SGS-E282)))	N3 Domain

GspD- Vch-(WT-E253R/E257K/E264Q-Del((N1-K239)/(N265-SGS-E282)))	N3 Domain
GspD- Vch-(WT-E454Q/D469S/E479K-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-E454Q/E455N/D469S/E479K-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-E455N/D469S/E479K-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-E454Q/D469S/E479T-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-E253Q/E257Q/E264Q/D290N/E454Q/E479K)-Del((N1-K239)/(N265-SGS-E282)))	Central Gate and N3 Domain
GspD- Vch-(WT-E253R/E257K/E264Q/E454Q/E479K-Del((N1-K239)/(N265-SGS-E282)))	Central Gate and N3 Domain
GspD- Vch-(WT-E253Q/E257K/E264Q/D290N/E454Q/E455N-Del((N1-K239)/(N265-SGS-E282)))	Central Gate and N3 Domain
GspD- Vch-(WT-E253Q/E257Q/E264Q/D290N/E454Q/E455N/D469S/E479K-Del((N1-K239)/(N265-SGS-E282)))	Central Gate and N3 Domain
GspD- Vch-(WT-E253Q/E257K/E264K/D290N/E454Q/E455N/D469S/E479K-Del((N1-K239)/(N265-SGS-E282)))	Central Gate and N3 Domain

Table 2: Increasing constriction size of central gate

Mutation	Mutant Position
GspD- Vch-(WT-F472A-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-Q473S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-N467S/N468G-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-N467G/N468S-Del((N1-K239)/(N265-SGS-E282)))*	Central Gate
GspD- Vch-(WT-N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282)))*	Central Gate
GspD- Vch-(WT-N467G/D469S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD- Vch-(WT-Del((N1-K239)/(N265-SGS-E282))/(T372-SGS-T393)/(T463-N470)))15	Central Gate

*Selected as backgrounds for further mutant designs

5 Table 3: Stabilizing central gate

Mutation	Mutant Position
GspD-Vch-(WT-G453A/N467G/N468S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-G453P/N467G/N468S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-G453V/N467G/N468S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-G453S/N467G/N468S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-N467G/N468S/G481P-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-N467G/N468S/G481V-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-N467G/N468S/G481A-Del((N1-K239)/(N265-SGS-E282)))	Central Gate

GspD-Vch-(WT-N467G/N468S/G481S-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-G453P/N467G/N468S/G481P-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-G453A/N467G/N468S/G481P-Del((N1-K239)/(N265-SGS-E282)))	Central Gate
GspD-Vch-(WT-G453P/N467G/N468S/G481A-Del((N1-K239)/(N265-SGS-E282)))	Central Gate

Table 4: Stabilizing cap gate: charge removal

Mutation	Mutant Position
GspD- Vch-(WT-D380N/N467G/N468S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380S/N467G/N468S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-R387S/N467G/N468S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-R387N/N467G/N468S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380S/R387S/N467G/N468S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380N/R387S/N467G/N468S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380N/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380S/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-R387S/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-R387N/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380S/R387S/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D380N/R387S/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-E367Q/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-E368Q/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D396N/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-K376S/D380S/R387S/E389Q/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate
GspD- Vch-(WT-D371N/K376S/D380S/R387S/E389Q/K394S/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate

Table 5: Stabilizing cap gate

Mutation	Mutant position
GspD- Vch-(WT-N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282))/(D371-SGS-K394)))	Cap Gate deletion
GspD- Vch-(WT-N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282))/(T372-SGS-T393)))	Cap Gate deletion
GspD- Vch-(WT-D371P/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate:

	Proline substitution
GspD- Vch-(WT-K394P/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate: Proline substitution
GspD- Vch-(WT-D371P/K394P/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate: Proline substitution
GspD- Vch-(WT-T372P/T393P/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate: Proline substitution
GspD- Vch-(WT-D371P/T372P/T393P/K394P/N467G/N468S/D469S-Del(N1-K239)/(N265-SGS-E282)))	Cap Gate: Proline substitution

Table 6: Constriction and Cap extreme mutants

Mutation	Mutant position
GspD- Vch-(WT-N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282)/(D371-T393))))15	Cap Gate deletion
GspD- Vch-(WT-G453A/N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282)/(D371-T393))))15	Cap Gate deletion and Central gate mutation
GspD- Vch-(WT-G453A/N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282)/(T372-SGS-T393))))15	Cap Gate deletion and Central gate mutation
GspD- Vch-(WT-N467G/N468S/D469S-Del((N1-K239)/(N265-SGS-E282)/(T372-SGS-T393)/T463-QTT-S466))))15	Cap Gate deletion and Central gate mutation
GspD- Vch-(WT-N467G-Del((N1-K239)/(N265-SGS-E282)/(T372-SGS-T393)/(N468-D469))))15	Cap Gate deletion and Central gate mutation
GspD- Vch-(WT-Del((N1-K239)/(N265-SGS-E282)/(T372-SGS-T393)/(N467-D469))))15	Cap Gate deletion and

	Central gate mutation
GspD- Vch-(WT-Del((N1-K239)/(N265-SGS-E282)/(T372-SGS-T393)/(T463-N470)))15	Central gate deletion

GspD mutants were expressed and purified *in vitro* using NEB pure express Kit. The reaction was setup as shown below.

5 **Table 7: Reaction Mixture**

Component	Volume (μL)
Solution A	10
Solution B	7.5
35S methionine	1
Rifampicin	0.8
Water	4
Lecithin vesicles	20μl (spun as pellet)
DNA	1.5

The volume of the initial reaction mix was 25μL. The reaction mixture was incubated for 3 hours at 37°C in a thermomixer. After incubation, the tube was centrifuged for 10 min at 22000g, of which the supernatant was discarded. The protein present in the pellet was re-suspended in 1x laemmli buffer and run in 5% Tris-HCl gel overnight at 55V. The gel was then dried and exposed to Carestream® Kodak® BioMax® MR film overnight. The film was then processed and the protein in the gel visualized. The Oligomeric band of the protein was cut from the gel and re-suspended in 100mM Tris, 50mM NaCl, 0.1% zwittergent , pH 8.

15 **Example 4. Electrophysiology Setup**

Setting up the experiment involved two separate steps, i) preparing the chips containing multiple wells of bulk co-polymer membrane to have single GspD mutant nanopores inserted and ready for sequencing and ii) DNA sample prep, which is added to the chip for sequencing. Materials and methods for both the steps are explained below.

20 GspD mutants were expressed and purified in-vitro and stored in buffer with 100mM Tris, 50mM NaCl, 0.1% zwittergent, pH 8. These mutant pores were diluted to 1:1000 using the 25mM K Phosphate, 150mM Potassium Ferrocyanide (II), 150mM Potassium Ferricyanide (III),

pH 8.0 buffer and added to the chips to obtain single pores in each wells. After pore insertion, the chips were washed with 1 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide (II), 150mM Potassium Ferricyanide (III), pH 8.0 buffer to remove excess GspD pores. IV curve measurement were performed when required using a script which records current at different potentials, ranging from -25mV to -200mV and 25mV to 200mV in 25mV alternating potential steps. The chips was flushed twice with 500mL of sequencing mix containing 470mM KCL, 25mM HEPES, 11mM ATP and 10mM MgCl₂, pH8.0. The chip is now ready for sequencing.

Meanwhile, for 3.6kb experiment, DNA sample was prepped for sequencing. 1µg of DNA analyte was incubated with the 40nM of Adapter mix (containing E8 helicase enzyme prebound to the adapter) and blunt TA ligase for 10 minutes. The ligation mixture was then purified of unligated free adapter using Spri purification. The final ligated mixture was eluted in 25 µL elution buffer containing 40mM CAPS pH10, 40mM KCl, 400nM cholesterol tether. For each chip, 6µl of DNA-adapter ligated mix was mixed with the sequencing mix (final volume of 75µL) and added to chip for sequencing. The experiment was then run for 6 hours at 180mV.

For the static strands experiment, Biotinylated static strands were incubated with monovalent streptavidin in ratio 1:1 for 10 minutes. The static strands were made to 1mM final concentration in 470mM KCL, 25mM HEPES, pH8. 150µl of strand was then added to the chip for static strand experiment. The pore used for the static strand experiment was GspD- Vch-(WT-N467G/N468S-Del((N1-K239)/(N265-SGS-E282))).

The results are shown in the Tables below and in Figures 14 to 19. The baseline pore is GspD-Vch-(WT-del(N1-K239)/(N265-SGS-E282))). This pore was chosen as a baseline. This pore expresses in IVTT even after deletion of two domains (1-238) as well as the constriction site 265-282 from the N3 domain. It is an open pore at -180mV around 200pA in C13 buffer, although frequent spikes in increases in open pore current are visible. It has an asymmetric IV-curve asymmetric, such that the pore remains open in negative potential and closed in positive potential. It has a non-linear IV curve with increasing open pore current with increasing potential.

Table 8: Characteristics of Mutant GspD Pores

Mutant <i>(Mutation from baseline, using numbering of SEQ ID NO: 32)</i>	Location of Mutant	Change in characteristic
Baseline	Removal of top constriction and N0,N1	Open pore around 200pA at -100mV. Asymmetric IV curve with pore open at

	Domain	negative potential and closed at positive.
Y379-GSG-R387	Removal of cap gate	Increases the open pore current (400-500pA in -180mV) at high voltages. However, the pore still has asymmetric IV which was open in negative potential and closed in positive potential.
F472A	Mutating larger residues to smaller ones in the central gate	Increase in open pore current. Makes the pore open at both negative and positive potential.
D469S	Removing charge in central gate	Small pores with open pore around 60pA at -180mV. However, some large pores with asymmetric IV curves were also seen.
G453 and G481	Central Gate	Important for protein expression and oligomerization and does not express when mutated to other residues apart from A and V.
N467G/N468S*	Mutating larger residues to smaller ones in the central gate	Slight increase in open pore current slightly from 200pA to 300pA at -180mV. Pore start to open even at positive potential.
N467S/N468G	Mutating larger residues to smaller ones in the central gate	Slight increases open pore current slightly from 200pA to 300pA at -180mV. Pore start to open even at positive potential.
N467G/N468S/D469S*	Mutating larger residues to smaller ones in the central gate and removal of Charge in the central gate	Decrease in open pore current to 80pA at -180mV. Pores are symmetrical which are open in both negative and positive potential.

*Selected as backgrounds for further mutant designs

Table 9: Characteristics of Mutant GspD Pores

Mutant <i>(Mutation from baseline)</i>	Comparing IV curves and Current levels
Baseline	100pA at -150mV. Open in negative potential and close in positive.
Y379-GSG-R387	80p at -150mV but open pore spikes up to 200pA (due to the increased cap gate diameter). Still has asymmetric IV curve which is open in negative potential and close in positive.

F472A	Triggers saturation in IV curve at higher potential. Current is around 200pA at -100mV. Open in both negative and positive potential.
D469S	Small pore with 40pA current at -150mV has asymmetric IV. Larger pore has symmetric IV and current of 200pA at -150mV.
N467G/N468S*	Mutating larger residues to smaller ones in the central gate
N467S/N468G	Mutating larger residues to smaller ones in the central gate
N467G/N468S/D469S*	Mutating larger residues to smaller ones in the central gate and removal of Charge in the central gate

*Selected as backgrounds for further mutant designs

Other Embodiments

All of the features disclosed in this specification may be combined in any combination.

- 5 Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

From the above description, one skilled in the art can easily ascertain the essential characteristics of the present disclosure, and without departing from the spirit and scope thereof, can make

- 10 various changes and modifications of the disclosure to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

Equivalents

- While several inventive embodiments have been described and illustrated herein, those of
15 ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described
20 herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented
25 by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and

claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within
5 the inventive scope of the present disclosure.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

All references, patents and patent applications disclosed herein are incorporated by
10 reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase “and/or,” as used herein in the specification and in the claims, should be
15 understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically
20 identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

25 As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or,
30 when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e., “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of
35 patent law.

As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

CLAIMS

1. A modified secretin nanopore disposed in a membrane, the modified secretin
5 nanopore comprising a lumenal surface defining a lumen that extends through the membrane between a *cis*-opening and a *trans*-opening, wherein the lumenal surface comprises one or more amino acid modifications.
2. The modified secretin nanopore of claim 1, wherein the one or more amino acid modifications comprise a charge-altering modification.
- 10 3. The modified secretin nanopore of claim 2, wherein the charge-altering modification is a substitution of a negatively-charged amino acid with a positively-charged amino acid.
4. The modified secretin nanopore of claim 1 or 2, wherein the one or more amino acid modifications comprise a substitution of a neutral amino acid with a hydrophobic amino
15 acid.
5. The modified secretin nanopore of any one of the preceding claims, wherein the *cis*-opening has a diameter in a range of 60 Å to 120 Å.
6. The modified secretin nanopore of any one of the preceding claims, wherein the *trans*-opening has a diameter in a range of 40 Å to 100 Å.
- 20 7. The modified secretin nanopore of any one of the preceding claims, wherein the secretin nanopore comprises a constriction having a diameter of about 7.5 Å to 25 Å.
8. The modified secretin nanopore of any one of the preceding claims, wherein the secretin is of a type II secretion system.
9. The modified secretin nanopore of claim 8, wherein the secretin is GspD.
- 25 10. The modified secretin nanopore of claim 9, which comprises a subunit polypeptide comprising a secretin domain having an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 36.
11. The modified secretin nanopore of claim 9 or 10, which comprises a subunit polypeptide having an amino acid sequence that is at least 95% identical to the amino acid
30 sequence as set forth in SEQ ID NO: 34 or 35.
12. The modified secretin nanopore of any one of claims 9 to 11, wherein the GspD does not comprise a cap gate.
13. The modified secretin nanopore of any one of claims 9 to 12, wherein the central

gate of GspD is modified to replace an amino acid with an amino acid having a smaller side group and/or to replace a negatively charged amino acid with a neutral or positively charged amino acid.

14. The modified secretin nanopore of claim 13, which comprises a subunit polypeptide comprising a secretin domain having an amino acid sequence that is at least 95% identical to an amino acid sequence as set forth in SEQ ID NO: 36, wherein: (i) all or some of the amino acids from D55 or T56 to T77 are deleted or substituted, one or more of K60, D64, R71 and E73 is substituted with an uncharged amino acid and/or one or more of D55, T56, T77 and K78 is substituted with P; and/or (ii) F156 is substituted with a smaller amino acid, N151 and/or N152 is/are substituted with a smaller amino acid, D153 is substituted with an uncharged amino acid, G137 and G165 are each independently unmodified or substituted with A or V.

15. The modified secretin nanopore of claim 14, wherein Y63 to R71 are deleted and/or substituted with GSG or SGS, F156 is substituted with A, D153 is substituted with S, and/or N151 and N152 are each independently substituted with G or S.

16. The modified secretin nanopore of any one of claims 1 to 7, wherein the secretin is of a type IV secretion system.

17. The modified secretin nanopore of claim 16, wherein the secretin is PilQ.

18. The modified secretin nanopore of any one of claims 1 to 7, wherein the secretin is of a type III secretion system.

19. The modified secretin nanopore of claim 18, wherein the secretin is YscC.

20. The modified secretin nanopore of claim 18, wherein the secretin is InvG.

21. The modified secretin nanopore of claim 20, which comprises a subunit polypeptide having an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 1 (corresponding to the amino acid sequence of InvG without N1 or N0 domain).

22. The modified secretin nanopore of claim 20 or 21, wherein the luminal surface further defines a constriction within the lumen, the constriction having one or more amino acid modifications (*e.g.*, charge-altering modifications) at amino acids D28, E225, R226, and/or E231 of SEQ ID NO: 1.

23. The modified secretin nanopore of claim 22, wherein the one or more amino acid modifications (*e.g.*, charge-altering modifications) of the constriction comprise one or more of the following:

- i. D28N/Q/T/S/G/R/K;
- ii. E225N/Q/T/A/S/G/P/H/F/Y/R/K;
- iii. R226N/Q/T/A/S/G/P/H/F/Y/K/V;
- iv. Deletion of E225;
- 5 v. Deletion of R226; and
- vi. E231N/Q/T/A/S/G/P/H/R/K.

24. The modified secretin nanopore of any one of claims 20 to 22, wherein the luminal surface comprises a capture portion having one or more amino acid modifications at amino acids E41, Q45 or E114.

10 25. The modified secretin nanopore of claim 24, wherein the capture portion comprises one or more of the following amino acid modifications:

- i. Q45R/K;
- ii. E41N/Q/T/S/G/R/K; and
- iii. E114N/Q/T/S/G/R/K.

15 26. A modified InvG nanopore subunit polypeptide comprising an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 1 (corresponding to the amino acid sequence of InvG without N1 or N0 domain), wherein the modified InvG nanopore subunit polypeptide comprises one or more amino acid modifications (*e.g.*, charge-altering amino acid modifications) at amino acid(s) selected from D28, E41, E114,
20 Q45, E225, R226, and E231 of SEQ ID NO: 1.

27. The modified InvG nanopore subunit polypeptide of claim 26, wherein the one or more amino acid modifications (*e.g.*, charge-altering amino acid modifications) comprise one or more of the following:

- i. D28N/Q/T/S/G/R/K;
- 25 ii. E225N/Q/T/A/S/G/P/H/F/Y/R/K;
- iii. R226N/Q/T/A/S/G/P/H/F/Y/K/V;
- iv. Deletion of E225;
- v. Deletion of R226; and
- vi. E231N/Q/T/A/S/G/P/H/R/K.

30 28. The modified InvG nanopore subunit polypeptide of claim 26 or 27, comprising one or more of the following amino acid modifications:

- i. Q45R/K;
- ii. E41N/Q/T/S/G/R/K; and
- iii. E114N/Q/T/S/G/R/K.

29. A modified InvG nanopore subunit polypeptide comprising an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 2 (corresponding to the amino acid sequence of WT InvG including N1 and N0 domains), wherein the modified InvG nanopore subunit polypeptide comprises one or more amino acid
 5 modifications (*e.g.*, charge-altering amino acid modifications) at amino acid(s) selected from D199, E212, E285, Q216, E396, R397, and E402 of SEQ ID NO: 2 and/or an endopeptidase cleavage site is inserted between positions 170 and 171 or 171 and 172 of SEQ ID NO: 2.

30. The modified InvG nanopore subunit polypeptide of claim 29, wherein the one or more amino acid modifications (*e.g.*, charge-altering amino acid modifications) comprise one or
 10 more of the following:

- i. D199N/Q/T/S/G/R/K;
- ii. E396N/Q/T/A/S/G/P/H/F/Y/R/K;
- iii. R397N/Q/T/A/S/G/P/H/F/Y/K/V;
- iv. Deletion of E396;
- 15 v. Deletion of R397; and
- vi. E402N/Q/T/A/S/G/P/H/R/K.

31. The modified InvG nanopore subunit polypeptide of claim 29 or 30, comprising one or more of the following amino acid modifications:

- i. Q216R/K;
- 20 ii. E212N/Q/T/S/G/R/K; and
- iii. E285N/Q/T/S/G/R/K.

32. A modified GspD nanopore subunit polypeptide comprising a secretin domain having an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 34.

25 33. The modified GspD nanopore subunit polypeptide of claim 32, which comprises a subunit polypeptide having an amino acid sequence that is at least 95% identical to the amino acid sequence as set forth in SEQ ID NO: 33.

34. The modified GspD nanopore subunit polypeptide of claim 32 or 33, wherein the GspD does not comprise a cap gate.

30 35. The modified GspD nanopore subunit polypeptide of any one of claims 32 to 34, wherein the central gate of GspD is modified to replace an amino acid with an amino acid having a smaller side group and/or to replace a negatively charged amino acid with a neutral or positively charged amino acid.

36. The modified GspD nanopore subunit polypeptide of any one of claims 32 to 35, which comprises a secretin domain having an amino acid sequence that is at least 95% identical to an amino acid sequence as set forth in SEQ ID NO: 34, wherein: (i) all or some of the amino acids from D55 or T56 to T77 are deleted or substituted, one or more of K59, K60, R63, Q66, Q69, K71, Q74 and Q75 is substituted with an uncharged amino acid and/or one or more of D55, T56, T57 and K394 is substituted with P; and/or (ii) F157 is substituted with a smaller amino acid, N152 and/or N153 is/are substituted with a smaller amino acid, D154 is substituted with an uncharged amino acid, G453 and G481 are each independently unmodified or substituted with A or V.

37. The modified secretin nanopore of claim 14, wherein Y379-GSG-R387 or Y379-SGS-R387, F472A, D469S, N467G/N468S, N467S468G, N467G/N468S/D469S.

38. A polynucleotide comprising a nucleotide sequence encoding the modified InvG nanopore subunit polypeptide of any one of claims 26 to 31 or the modified GspD nanopore subunit polypeptide of any one of claims 32 to 37.

39. A modified secretin nanopore comprising a plurality of modified secretin nanopore subunit polypeptides of any one of claims 26 to 31 or a plurality of modified GspD nanopore subunit polypeptides of any one of claims 32 to 37.

40. The modified secretin nanopore of claim 39, comprising 9 to 20 of the modified secretin nanopore subunit polypeptides.

41. The modified secretin nanopore of claim 40, wherein each monomer is a modified secretin nanopore subunit polypeptide of any one of claims 26 to 31 or 32 to 37.

42. The modified secretin nanopore of claim 40, comprising different monomers, at least one of which monomers is a modified secretin nanopore subunit polypeptides of any one of claims 26 to 31 or 32 to 37.

43. An apparatus comprising a chamber housing an aqueous solution having disposed therein a membrane comprising a secretin nanopore, the secretin nanopore comprising a luminal surface defining a lumen that extends through the membrane between a *cis*-opening and a *trans*-opening.

44. The apparatus of claim 43, wherein the secretin nanopore is a modified secretin nanopore and the luminal surface comprises one or more amino acid modifications.

45. The apparatus of claim 44, wherein the modified secretin nanopore is a modified secretin nanopore of any one of claims 1 to 25 and 39 to 42.

46. The apparatus of any one of claims 43 to 45, further comprising an analyte present in the aqueous solution.

47. The apparatus of claim 46, wherein the analyte is a polynucleotide.

48. The apparatus of claim 47, further comprising a polynucleotide binding protein bound to the polynucleotide.

49. The apparatus of claim 48, wherein the polynucleotide binding protein is a
5 helicase, exonuclease, or polymerase.

50. The apparatus of claim 48 or 49, wherein the polynucleotide binding protein is on the *cis*-side of the membrane.

51. The apparatus of claim 50, wherein the polynucleotide binding protein is in contact with or covalently attached to the *cis*-opening of the nanopore.

10 52. The apparatus of claim 48 or 49, wherein the polynucleotide binding protein is on the *trans*-side of the membrane.

53. The apparatus of claim 52, wherein the polynucleotide binding protein is in contact with or covalently attached to the *trans*-opening of the nanopore.

54. A method comprising obtaining an apparatus of any one of claims 43 to 45 and
15 adding an analyte to the aqueous solution on the *cis*-side or the *trans*-side of the membrane.

55. The method of claim 54, wherein the analyte is a polynucleotide.

56. The method of claim 55, further comprising adding a polynucleotide binding protein to the aqueous solution on the *cis*-side of the membrane, wherein the polynucleotide binding protein binds to the polynucleotide.

20 57. The method of claim 56, wherein the polynucleotide binding protein is in contact with or covalently attached to the *cis*-opening of the nanopore, or is within the *cis* vestibule of the nanopore.

58. The method of claim 55, further comprising adding a polynucleotide binding protein to the aqueous solution on the *trans*-side of the membrane, wherein the polynucleotide
25 binding protein binds to the polynucleotide.

59. The method of claim 58, wherein the polynucleotide binding protein is in contact with or covalently attached to the *trans*-opening of the nanopore, or is within the *trans* vestibule of the nanopore.

60. The method of any one of claims 54 to 59, wherein the polynucleotide binding
30 protein is a helicase, exonuclease, or polymerase.

61. The method of any one of claims 54 to 60, further comprising inducing ionic current flow through the nanopore by applying a voltage gradient across the membrane.

62. The method of claim 61, further comprising detecting ionic current flow through the nanopore under the applied voltage gradient.

35 63. The method of claim 62, further comprising determining presence of the analyte

based on the detected ionic current flow.

64. A composition comprising a secretin nanopore and an enzyme provided within the *cis* or *trans* vestibule of the secretin nanopore.

5 65. The composition of claim 64, wherein the secretin nanopore is a wild-type secretin nanopore.

66. The composition of claim 64, wherein the secretin nanopore is a modified secretin nanopore.

67. The composition of claim 66, wherein the modified secretin nanopore is a modified secretin nanopore of any one of claims 1 to 25 and 39 to 42.

10 68. The composition of any one of claims 64 to 67, wherein the nanopore is provided within a membrane and extends through the membrane between a *cis*-opening and a *trans*-opening.

69. Use of a secretin nanopore for analyte detection and/or characterisation.

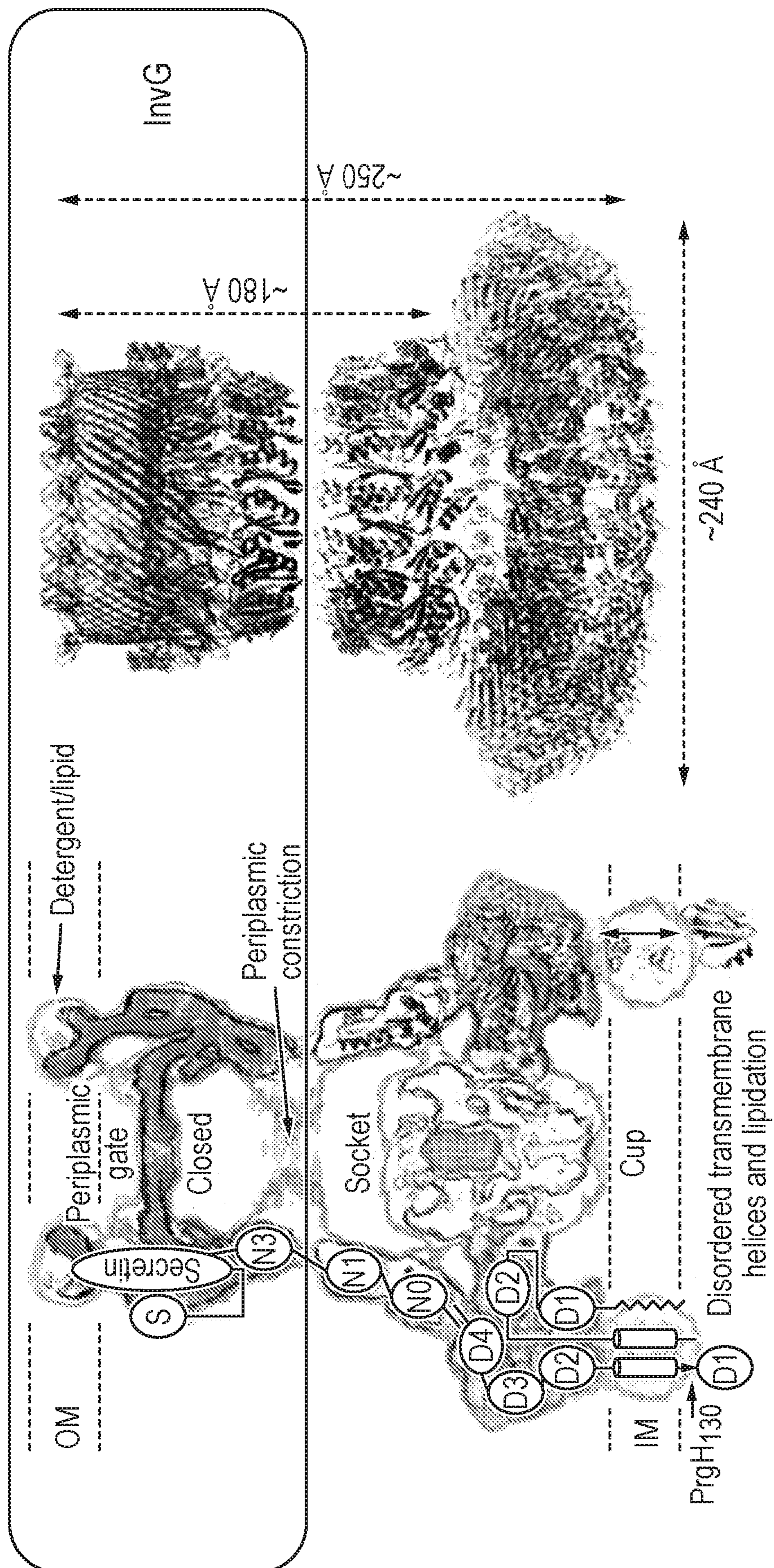
15 70. The use of claim 69, wherein the secretin nanopore is a modified secretin nanopore.

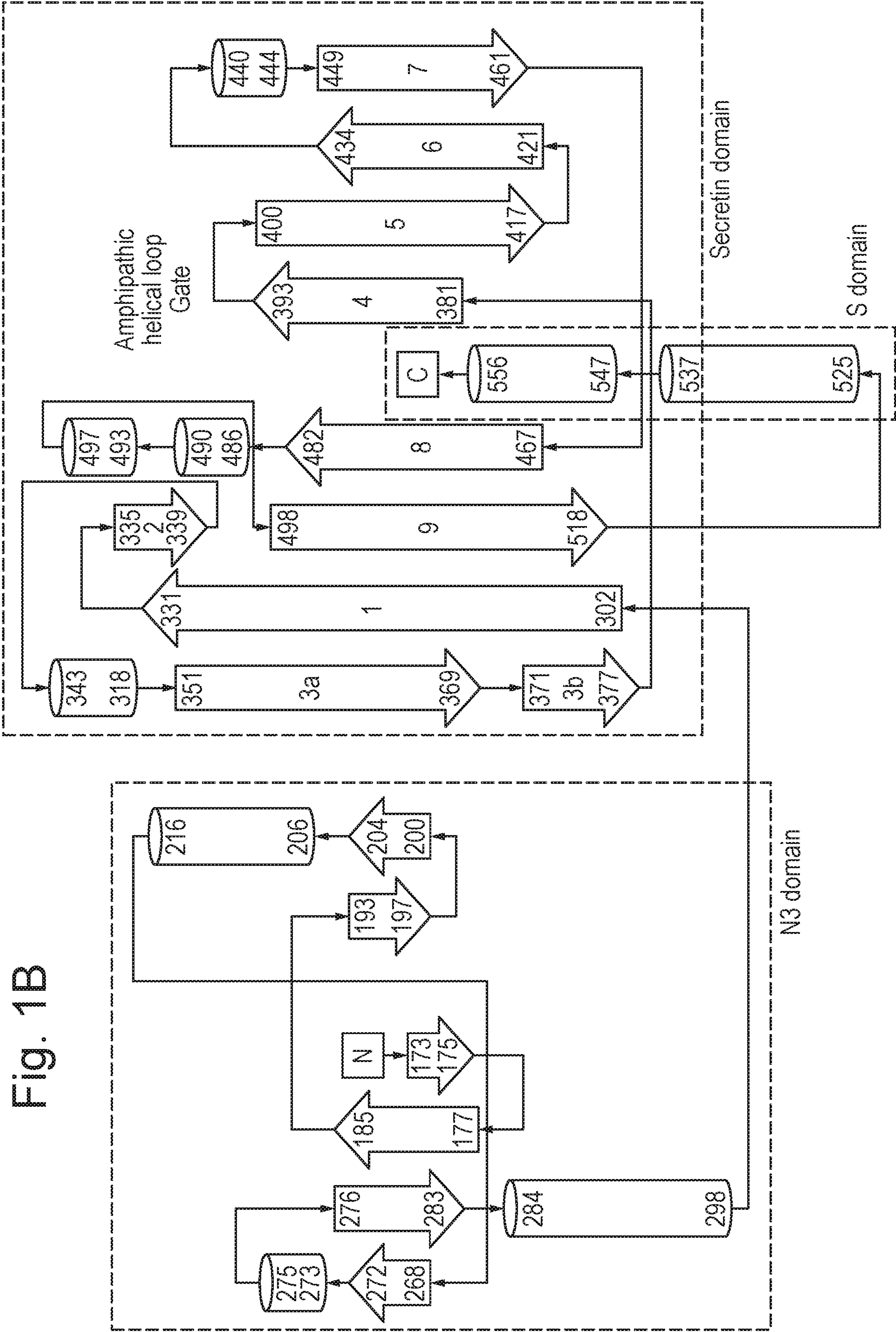
71. The use of claim 70, wherein the modified secretin nanopore is a modified secretin nanopore of any one of claims 1 to 25 and 39 to 42.

20 72. A method of detecting the presence or absence of an analyte in a sample and/or characterising an analyte in a sample, the method comprising contacting a secretin nanopore disposed in a membrane with the sample, applying a voltage gradient across the membrane and detecting ionic current flow through the nanopore.

73. The method of claim 72, wherein the secretin nanopore is a modified secretin nanopore.

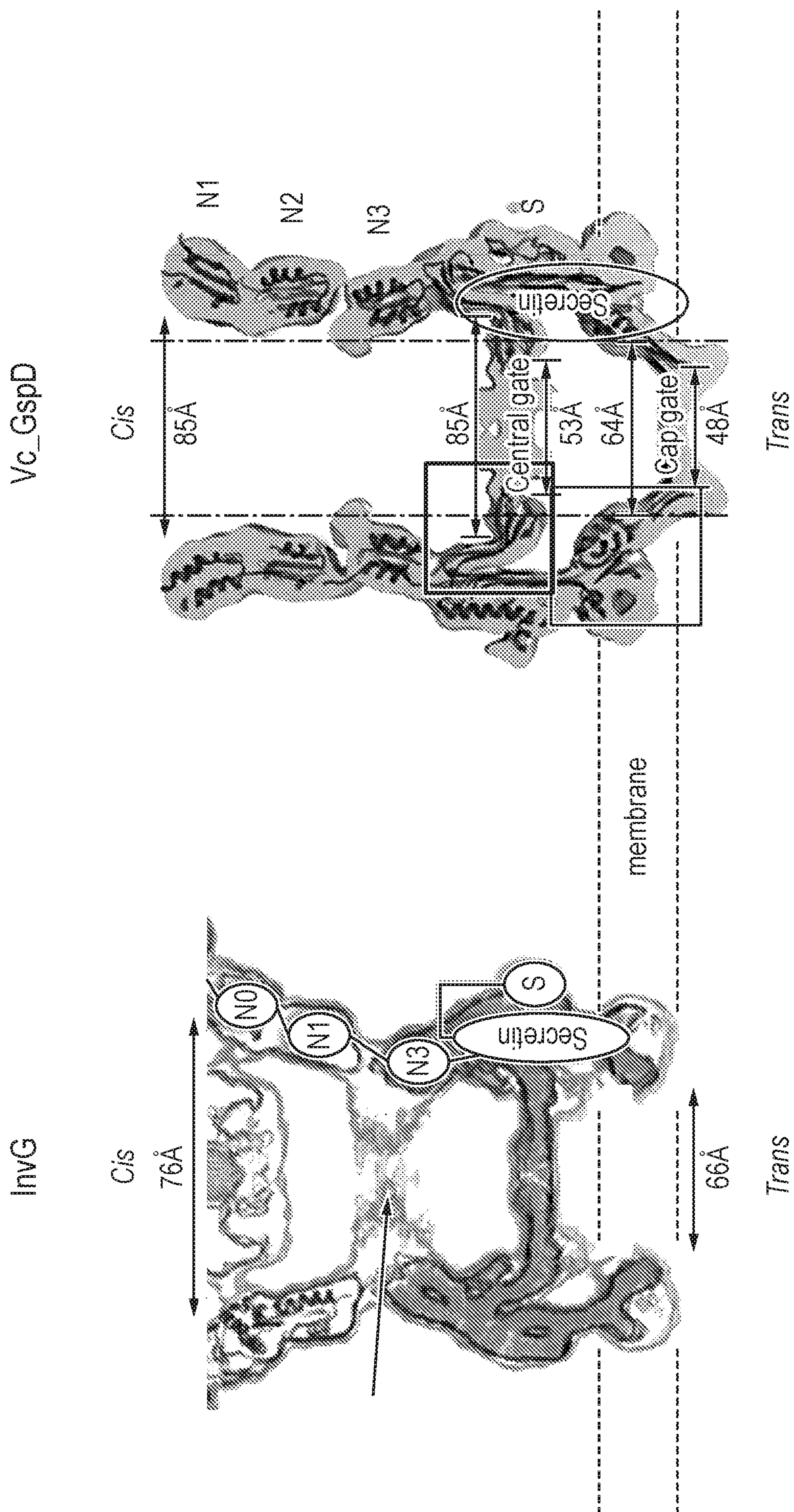
25 74. The use of claim 73, wherein the modified secretin nanopore is a modified secretin nanopore of any one of claims 1 to 25 and 39 to 42.





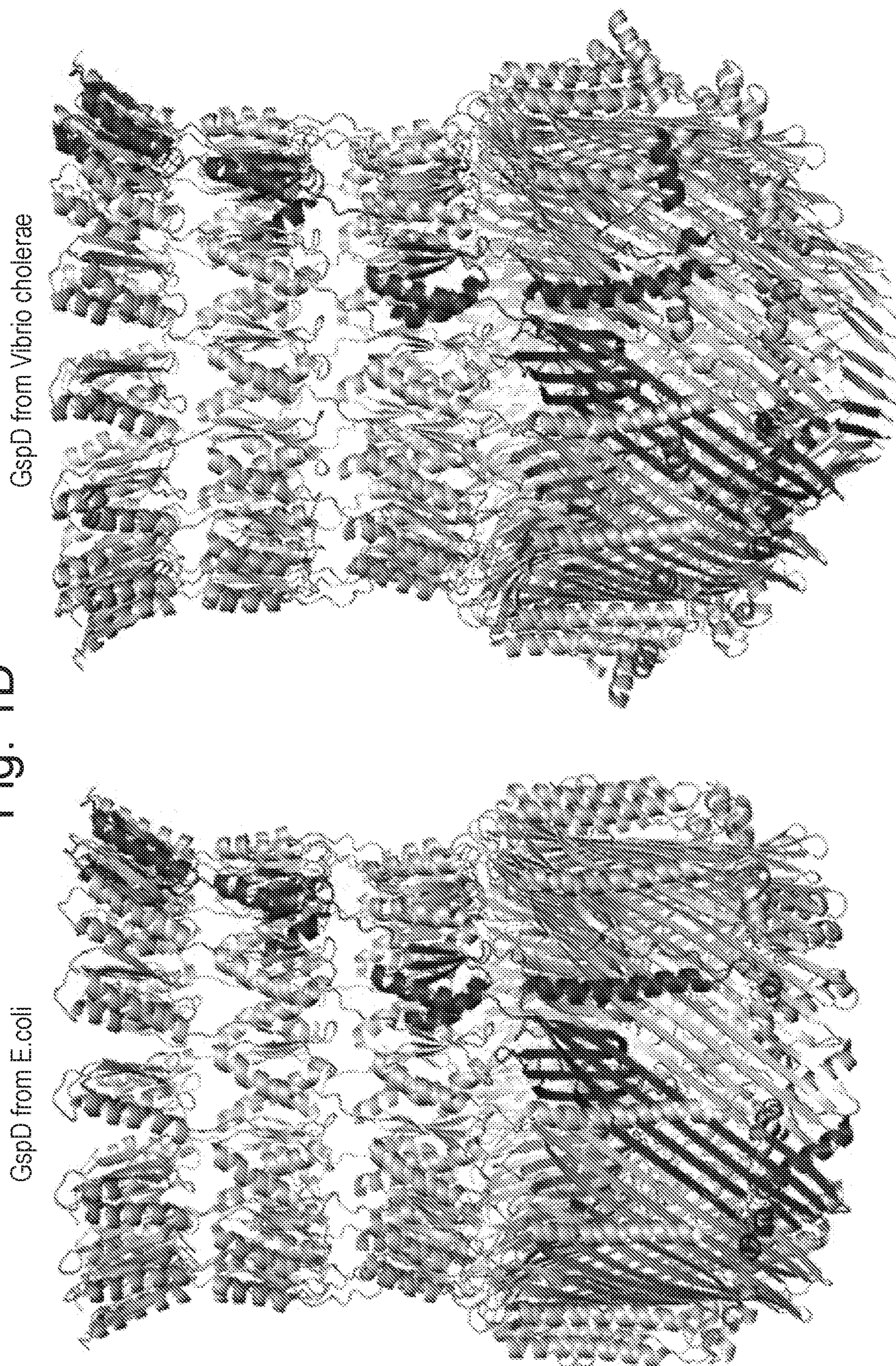
3/28

Fig. 1C



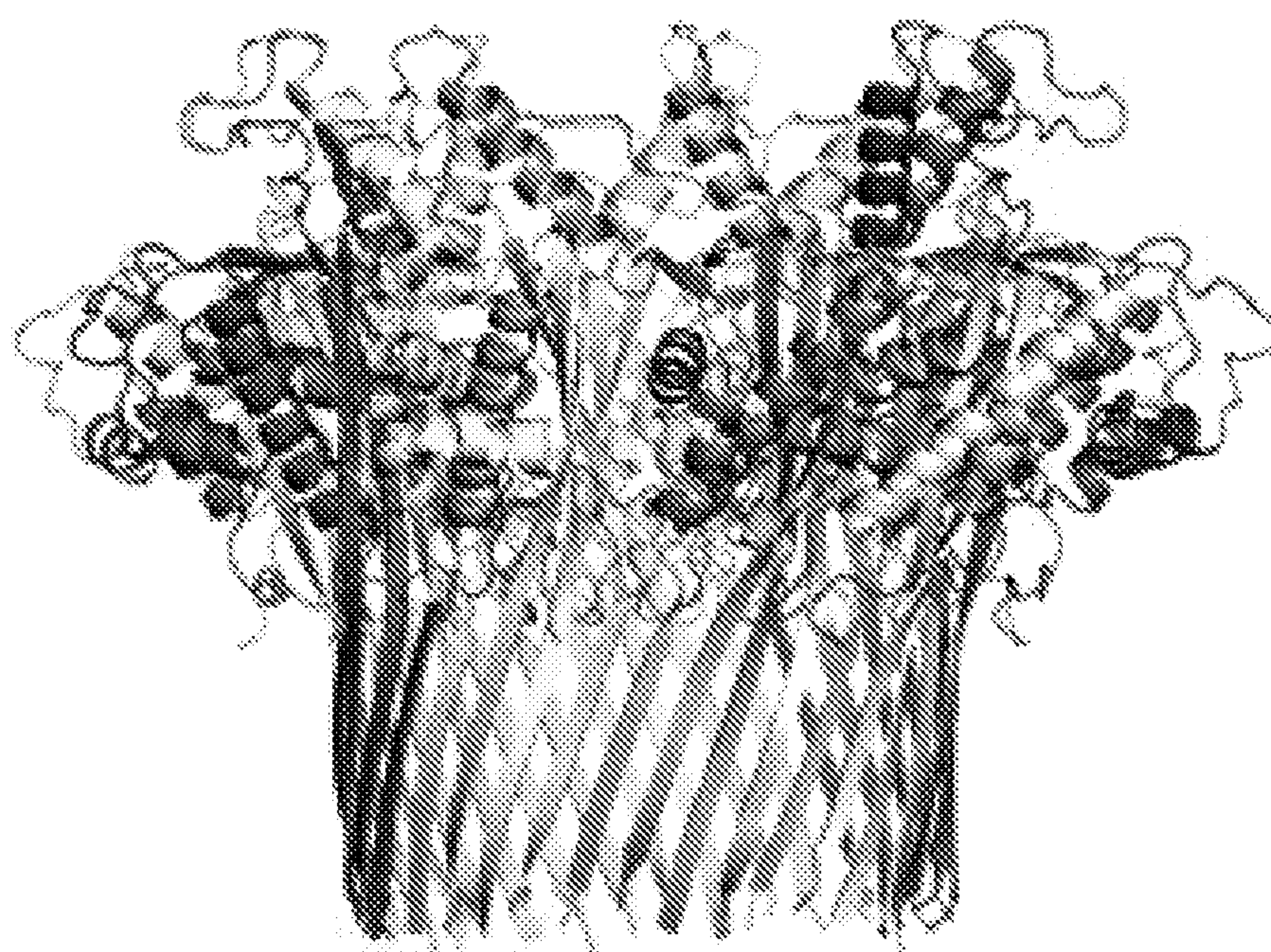
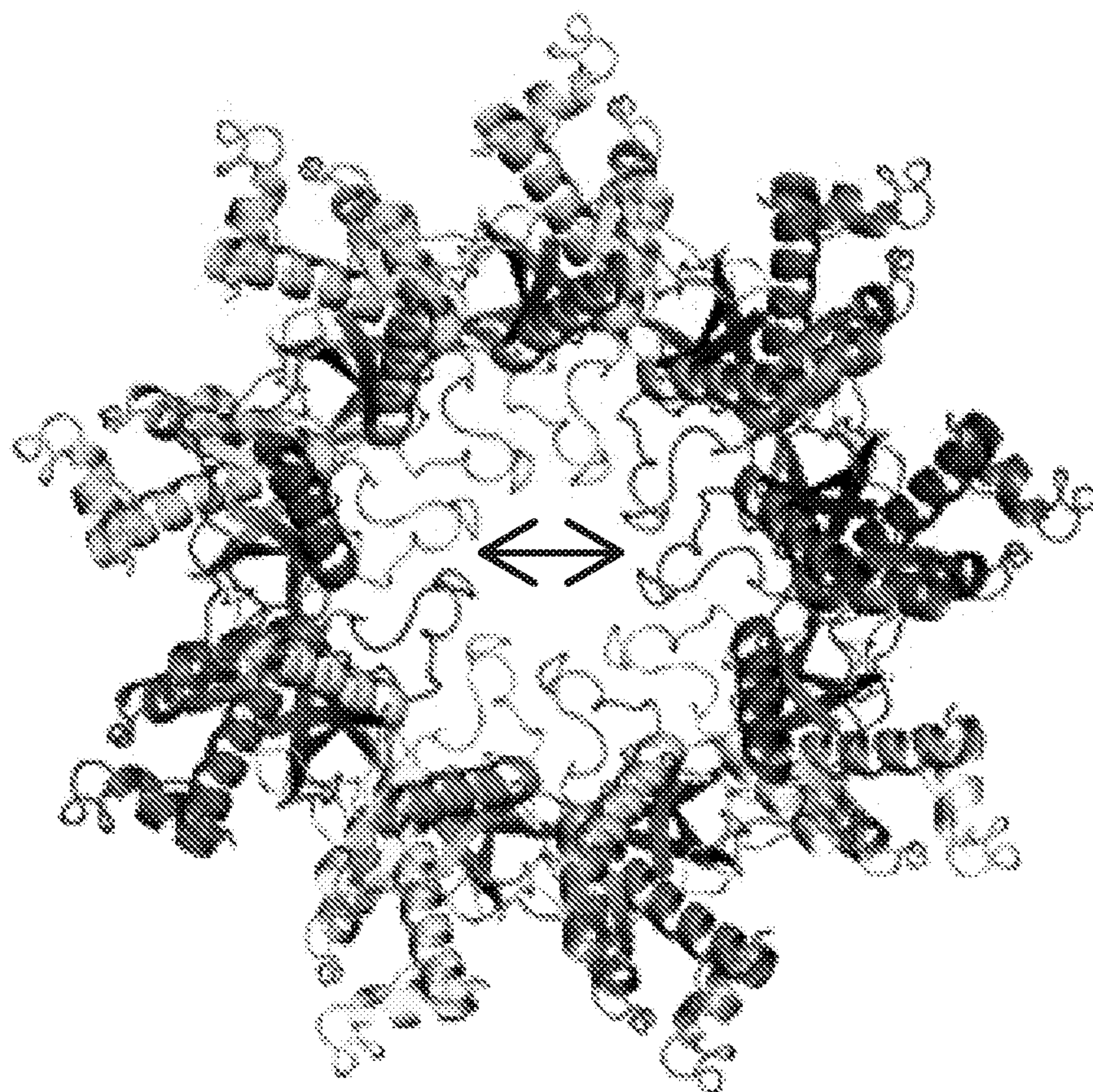
4/28

Fig. 1D



5/28

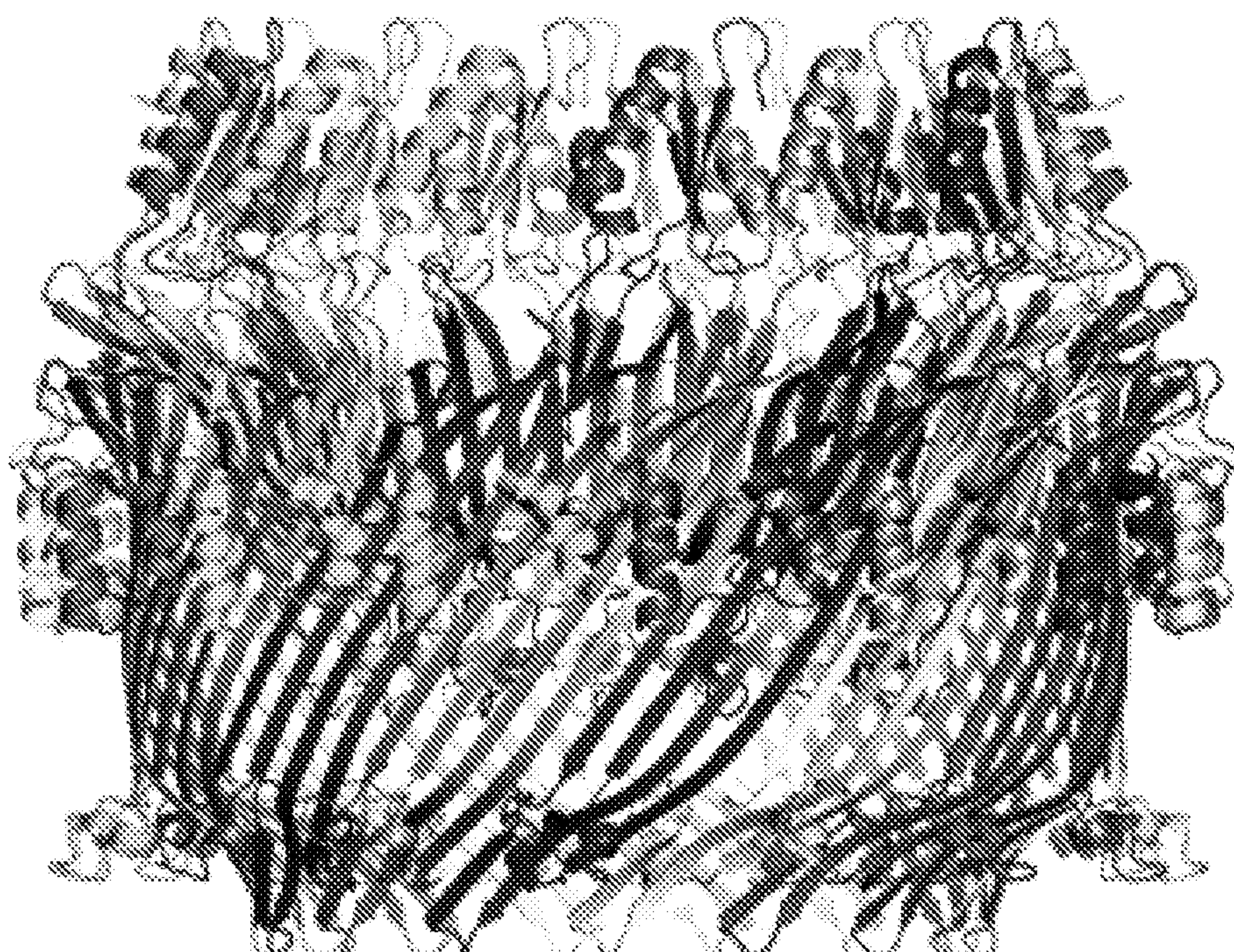
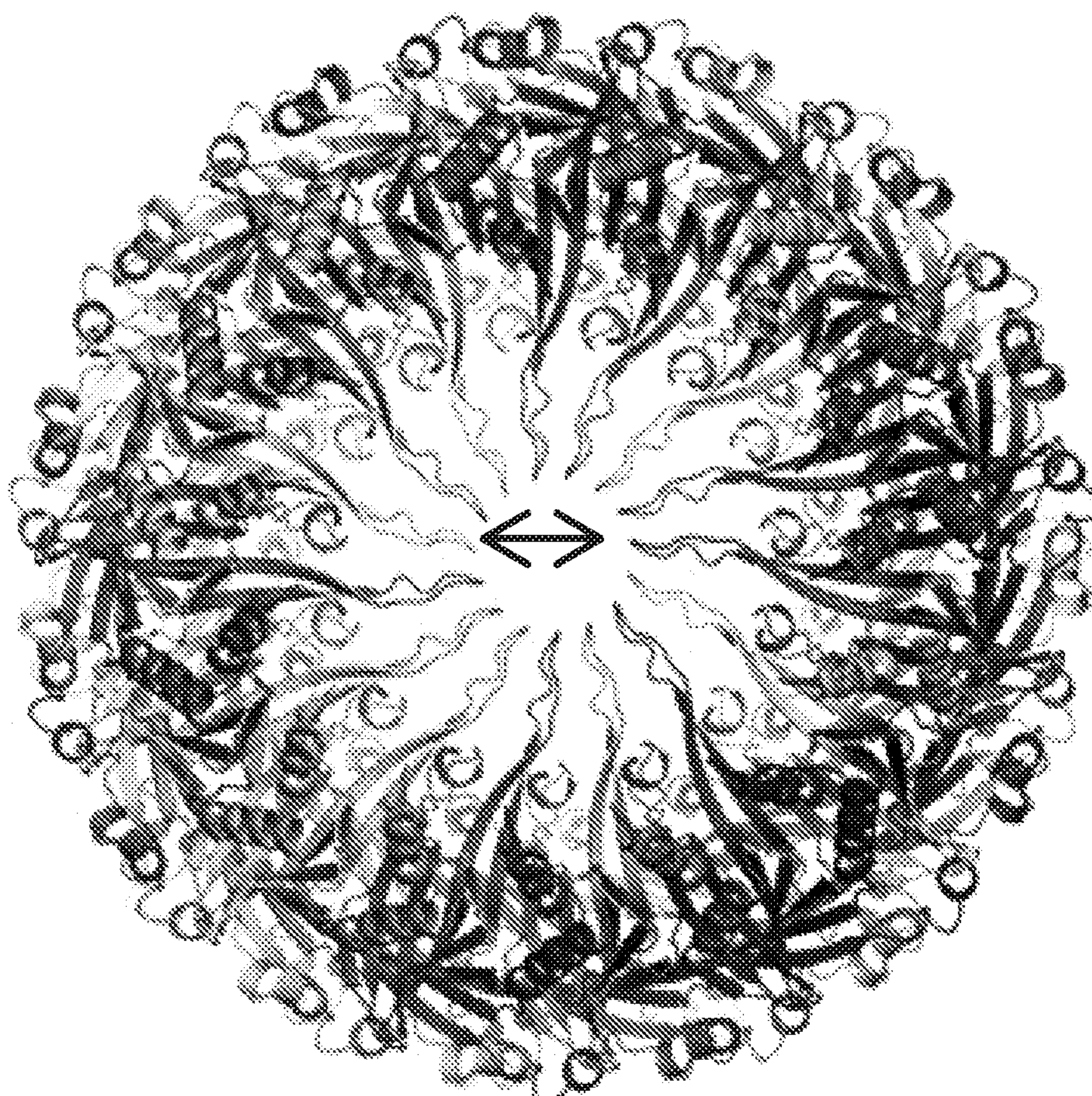
Fig. 2



CsgG

6/28

Fig. 2 (Cont.)



InvG

7/28

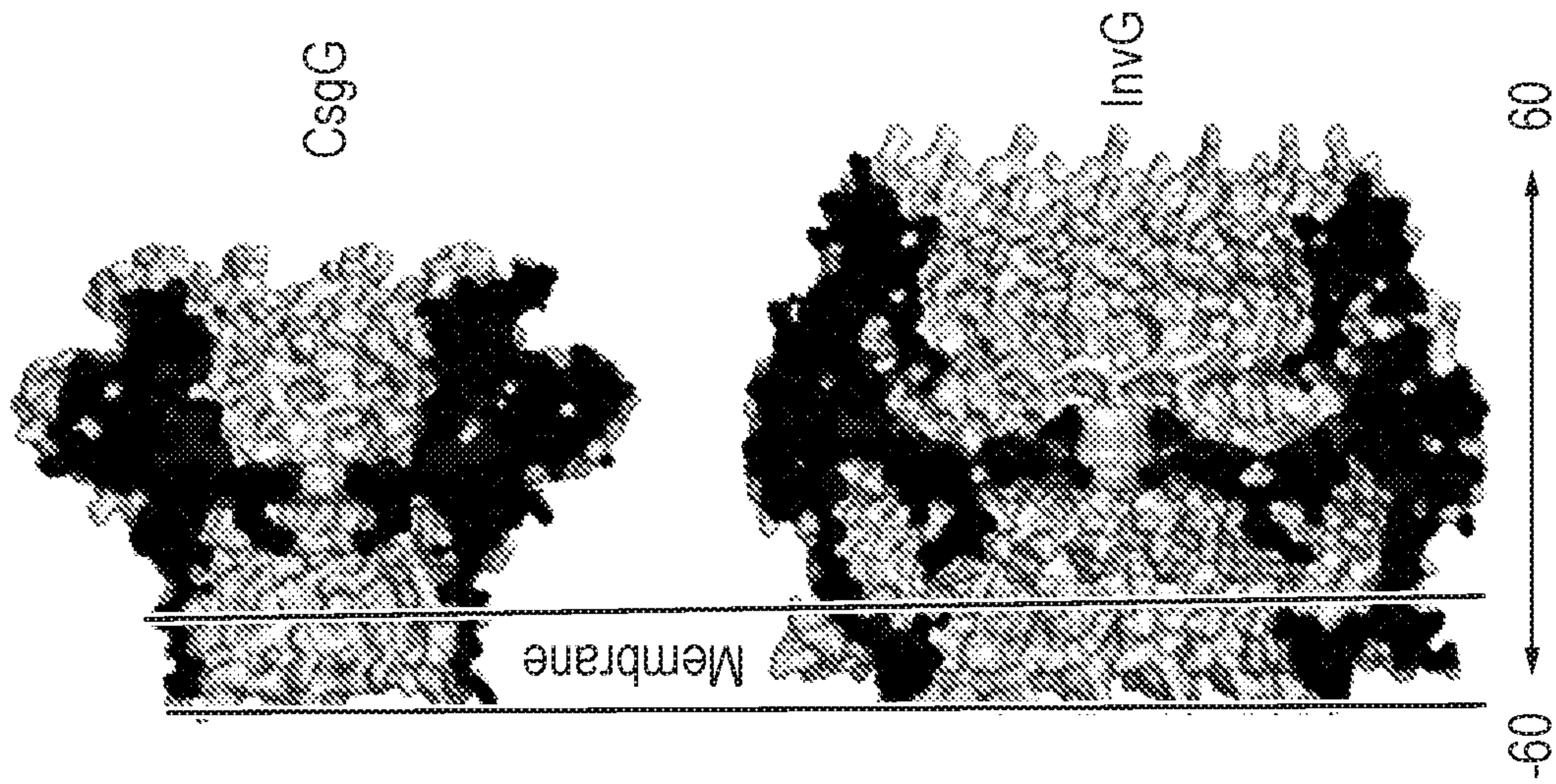


Fig. 3

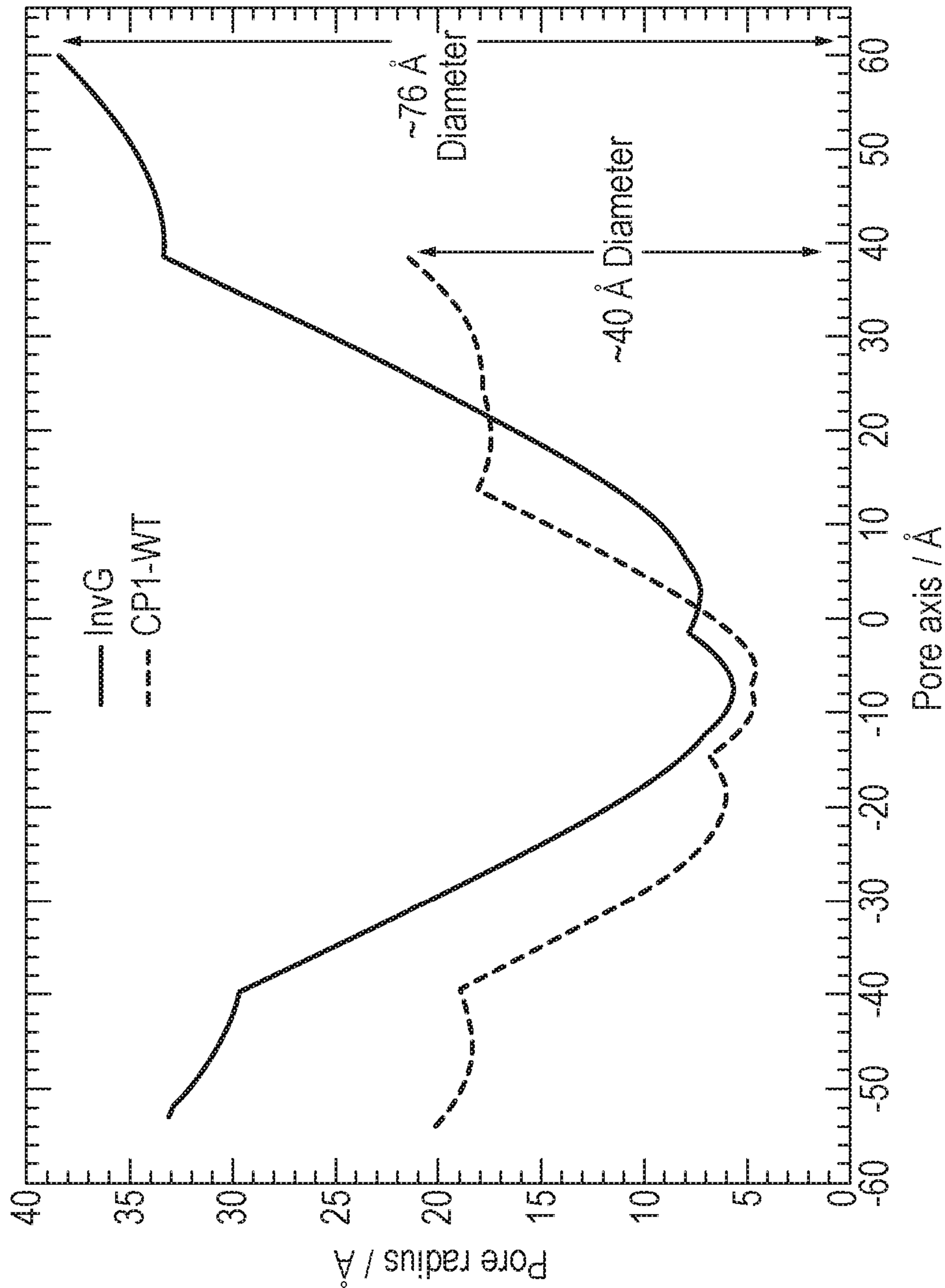
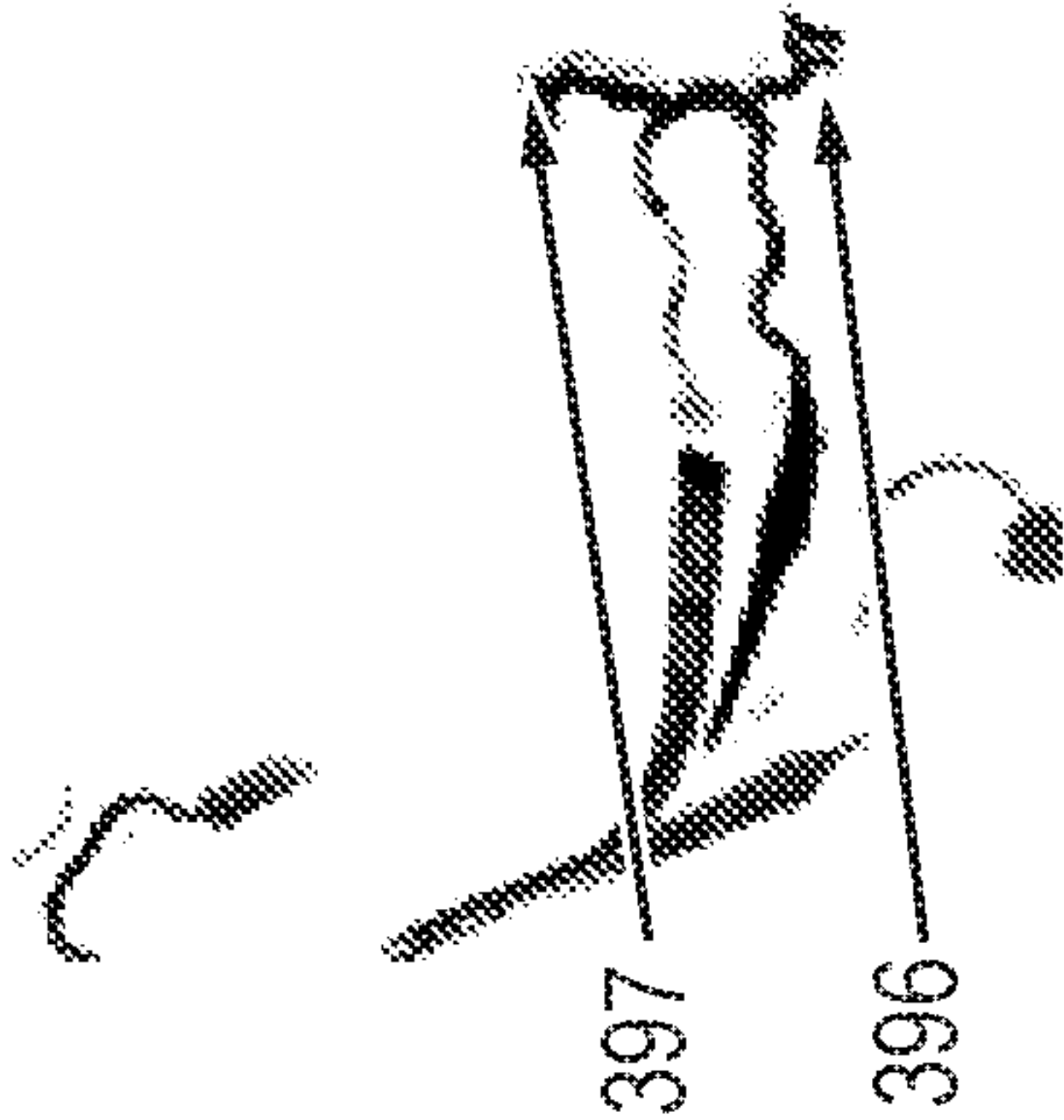
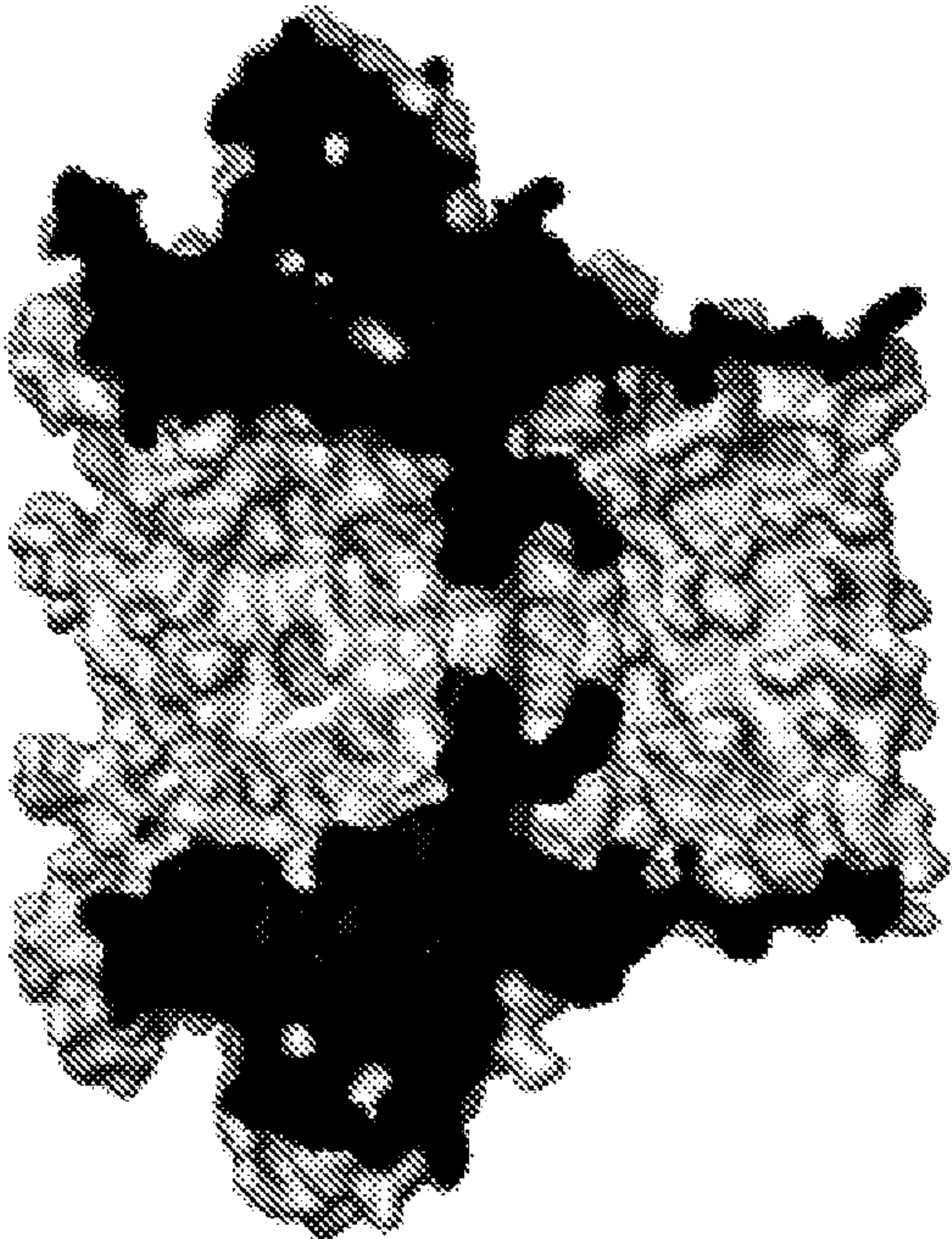
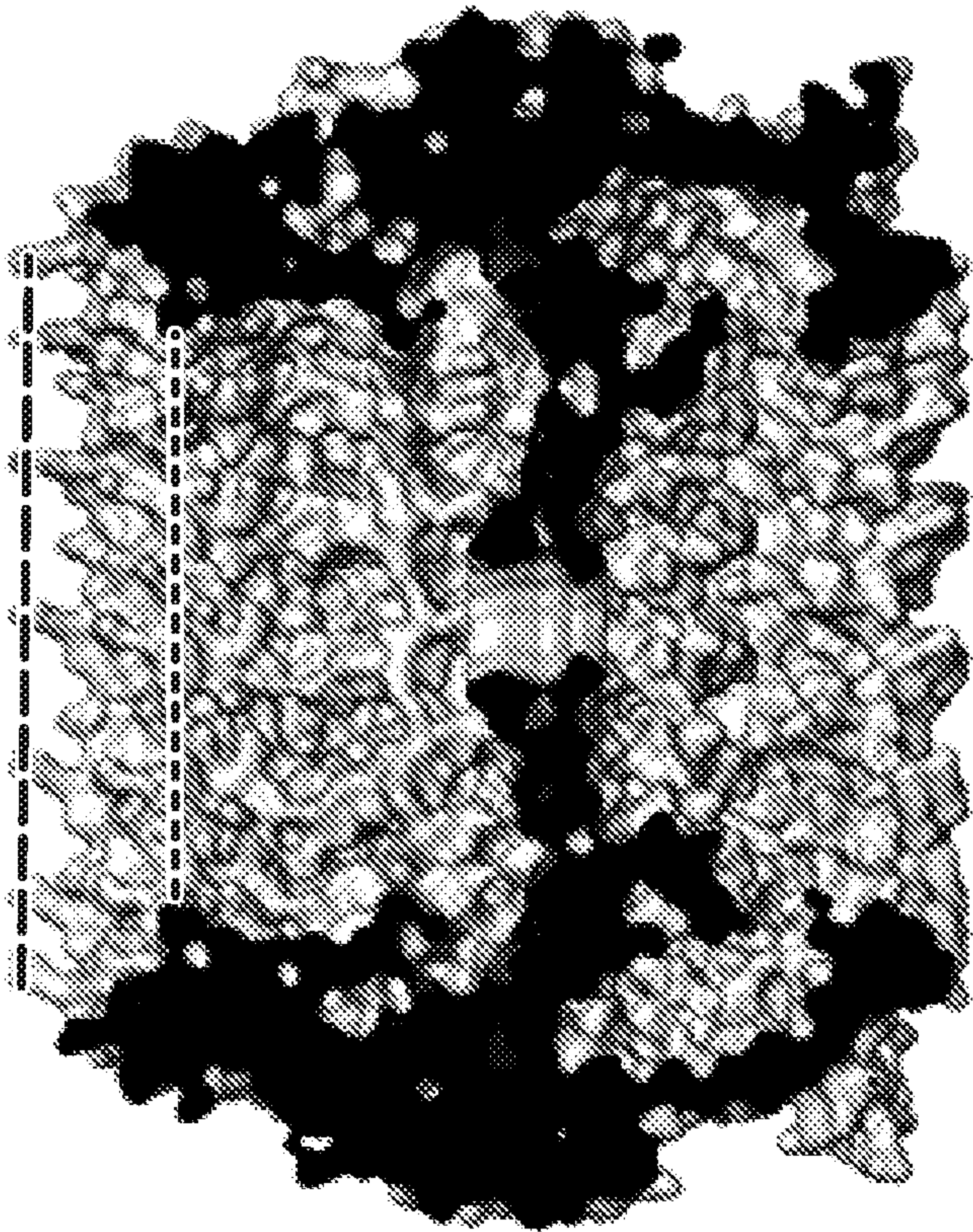
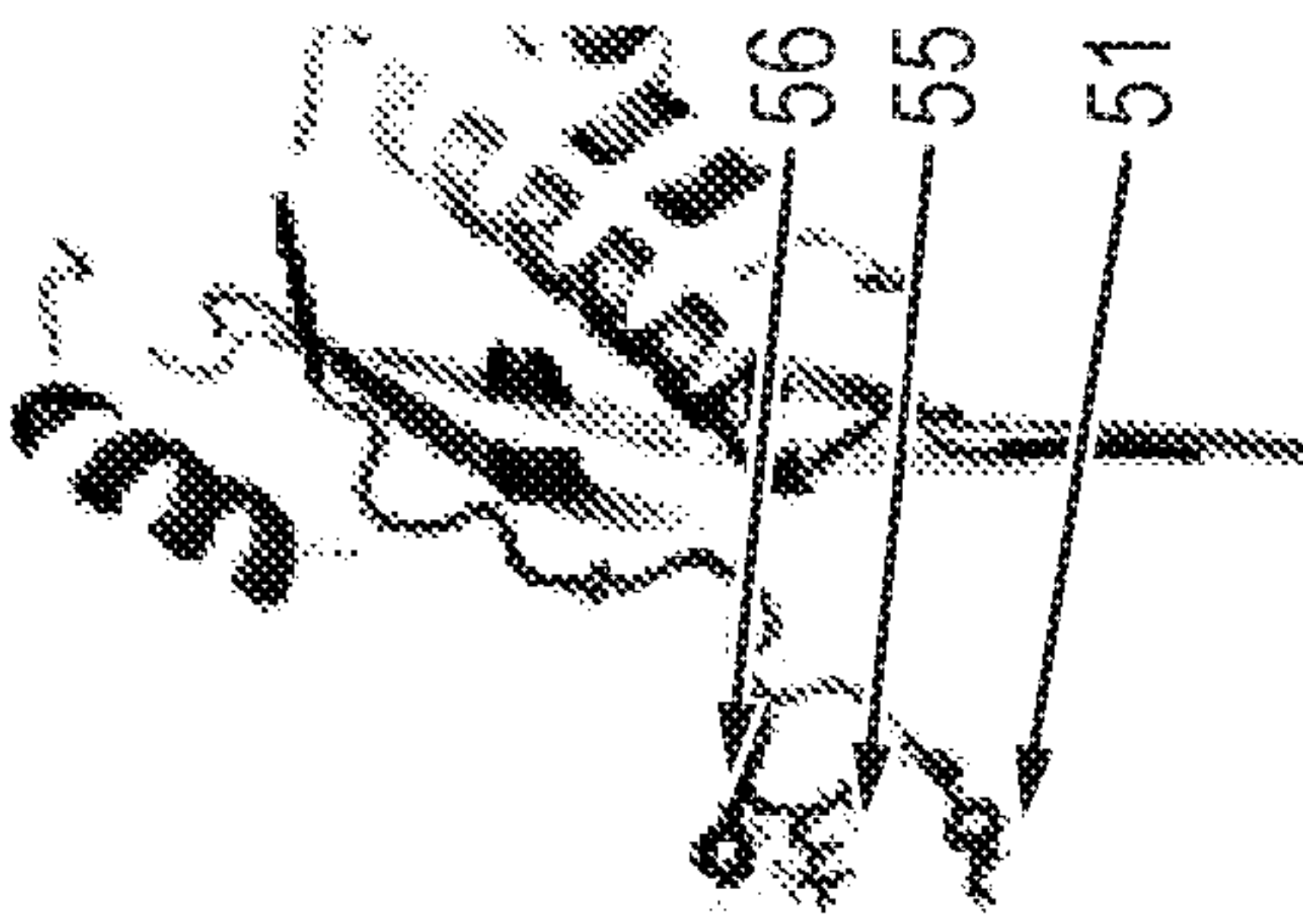


Fig. 4



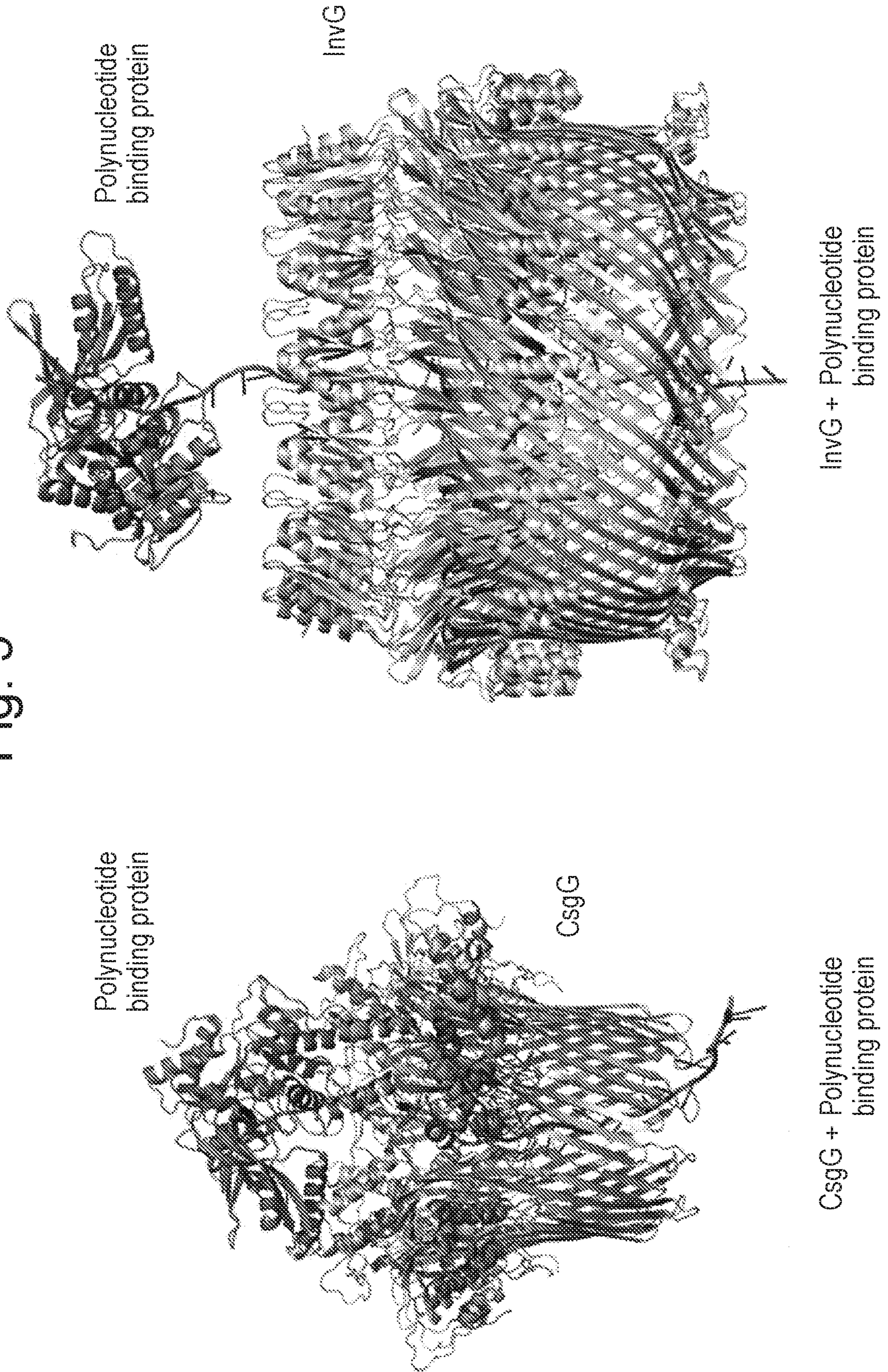
InvG



CsgG

9/28

Fig. 5



10/28

InvG

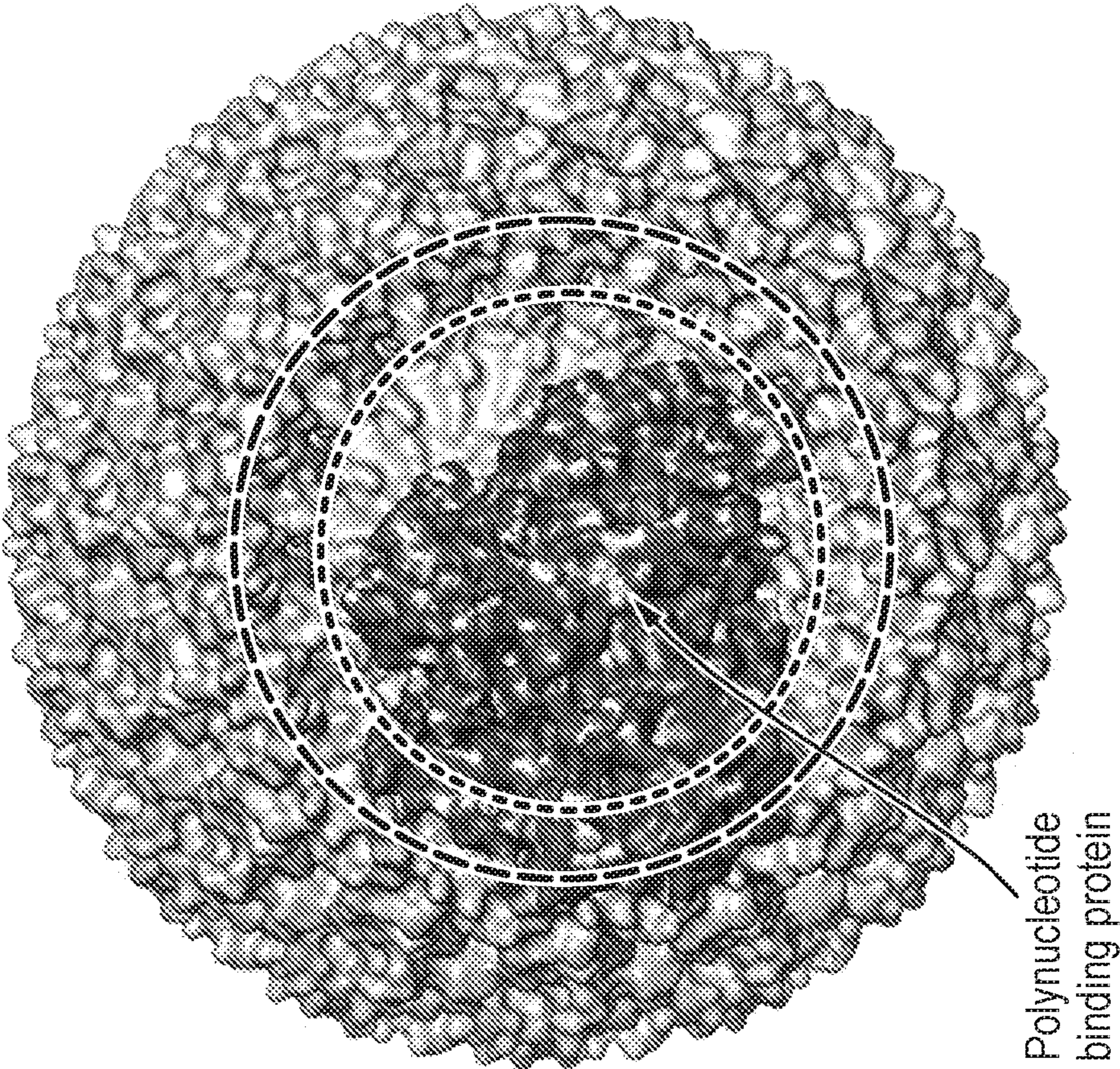
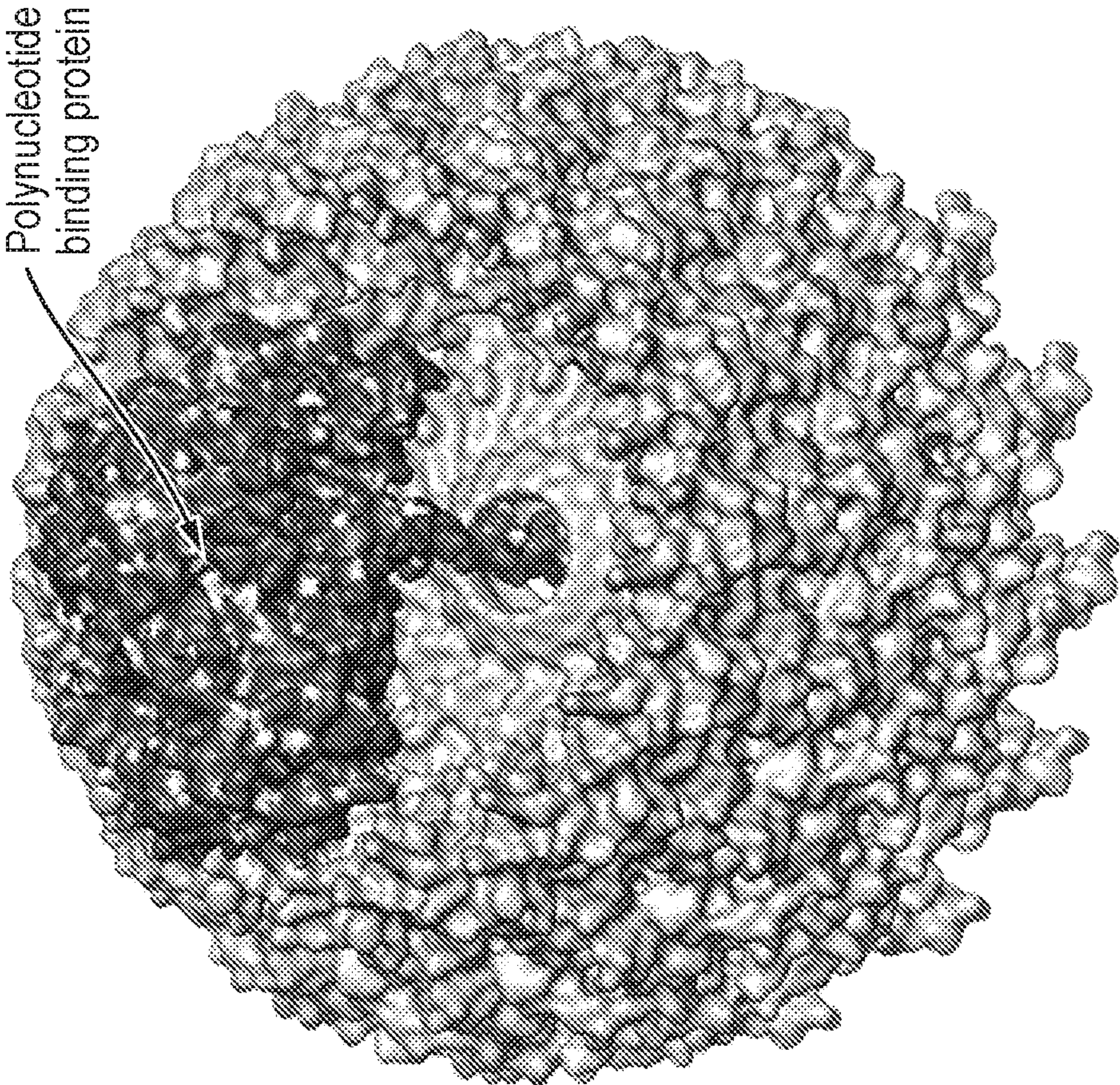


Fig. 6

11/28

Fig. 7

Constriction

1. InvG-(WT-E396N/Q/T/A/S/G/P/H/F/Y/R/K)
2. InvG-(WT-R397N/Q/T/A/S/G/P/H/F/Y/K/V)
3. Combinations / permutations of 1 and 2
4. InvG-(Wt-Del-396)
5. InvG-(Wt-Del-396/397)

Capture

6. InvG-(WT-E402N/Q/T/A/S/G/P/H/R/K)
7. InvG-(WT-Q216R/K)
8. Combinations of 7 with 6
9. Combinations of 8 with 1-5
10. InvG-(WT-D199N/Q/T/S/G/R/K)
11. InvG-(WT-E212N/Q/T/S/G/R/K)
12. InvG-(WT-E285N/Q/T/S/G/R/K)
13. 10 – 12 with 9 (if too many combinations, this could be omitted)

12/28

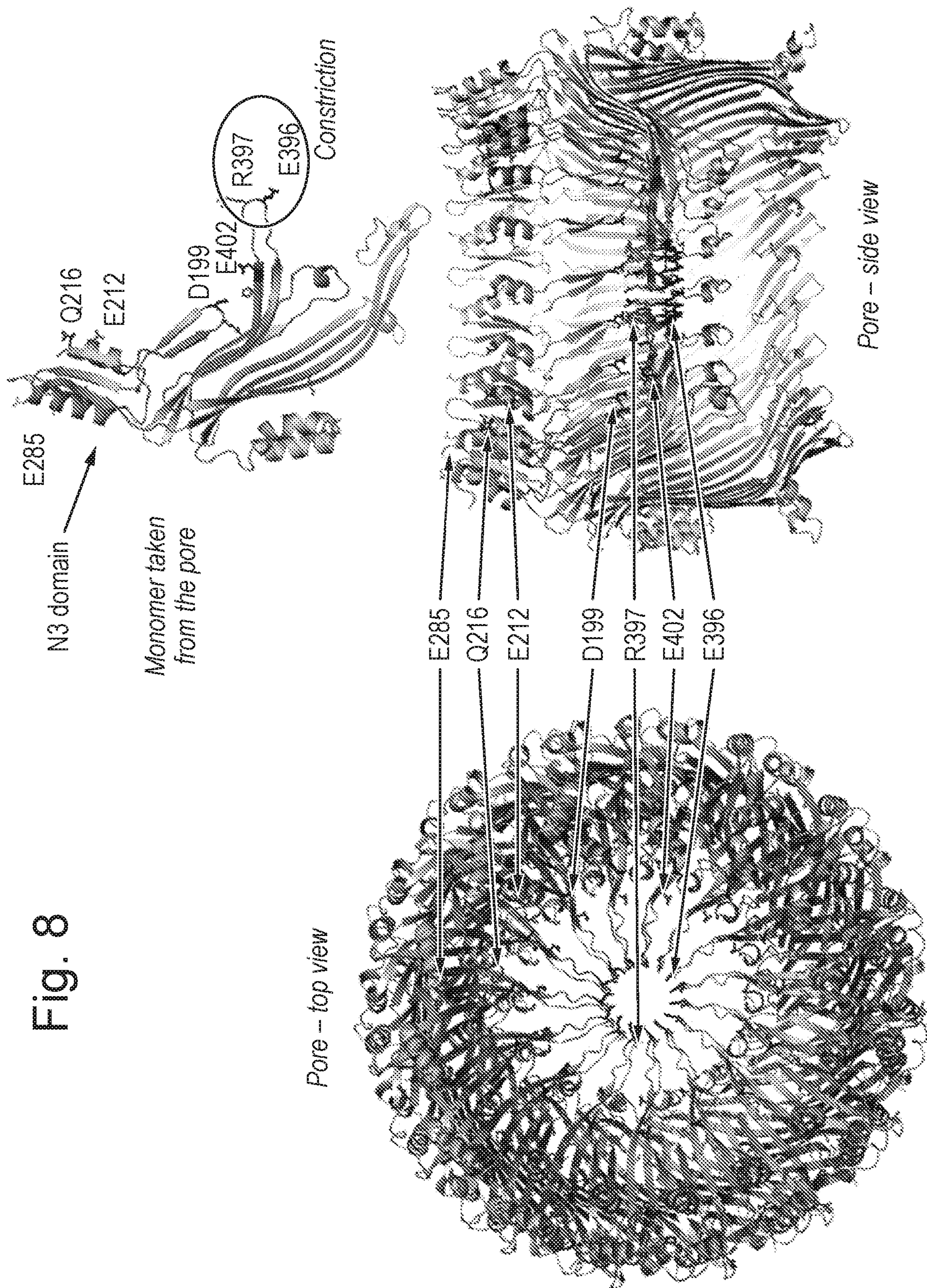


Fig. 8

13/28

Fig. 9

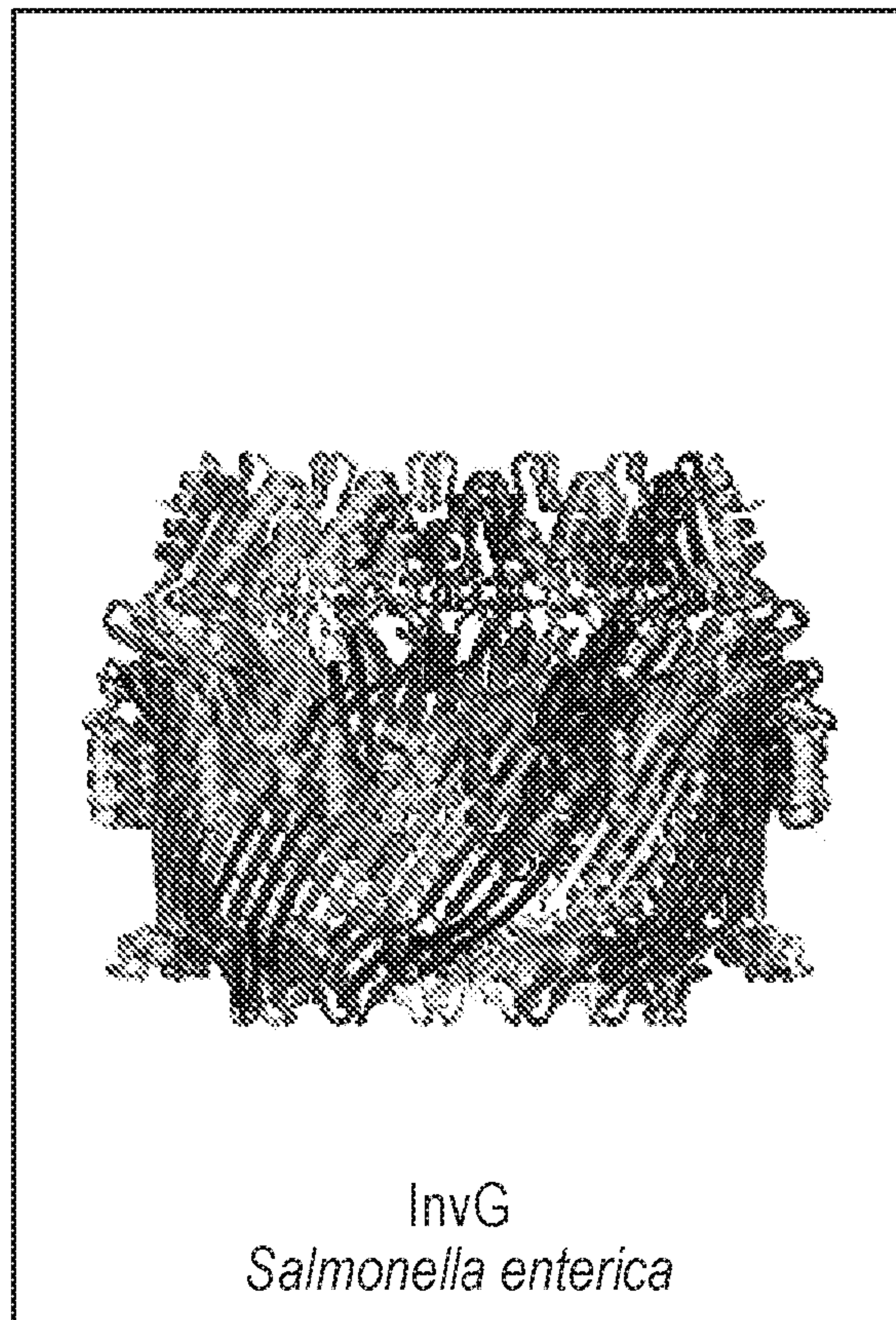
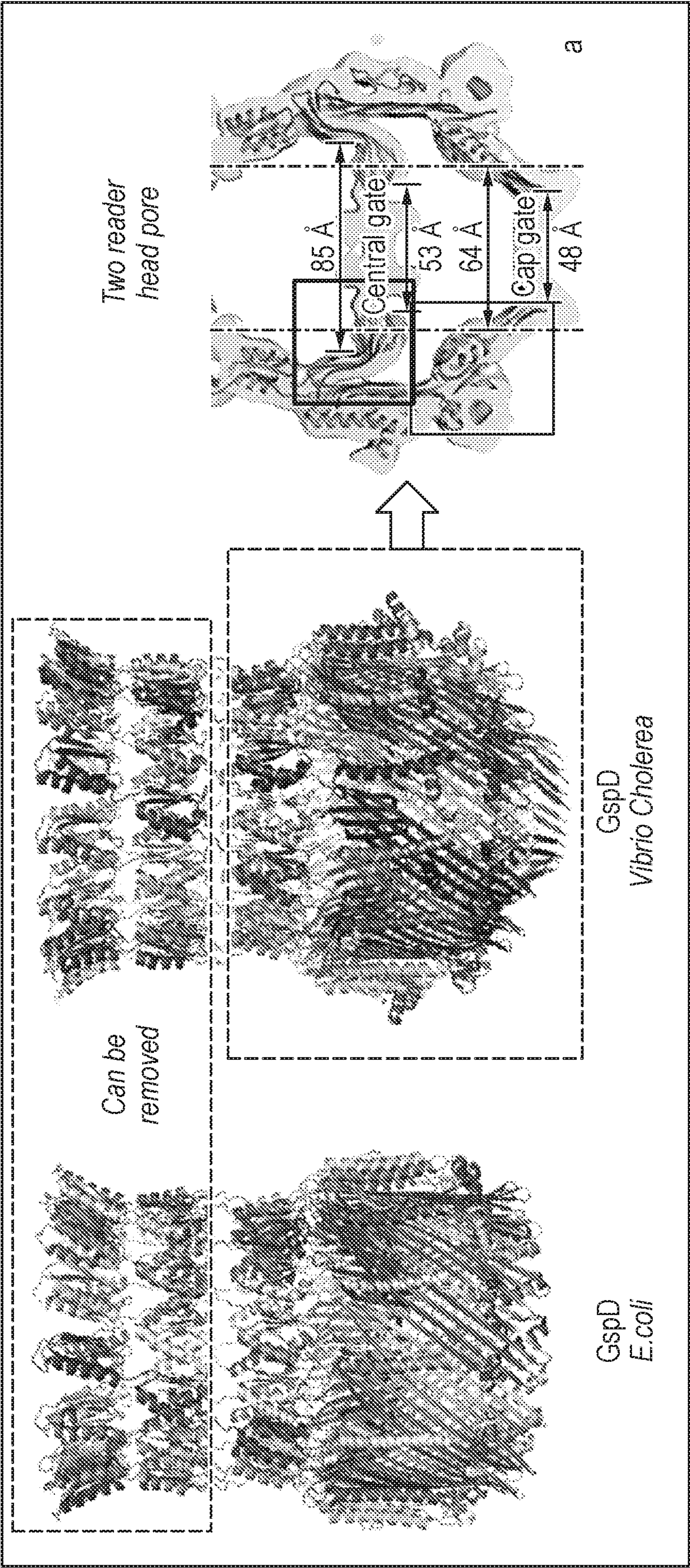


Fig. 9 (Cont.)



GspD sequences - *Vibrio Cholera* **Fig. 10**

PDB - 5WQ8:

NEFSASFKGTDIQEFINIVGRNLEKTIIVDPSVRGKVDVRSFDTLNEEQYYSFFLSVLEVYGFVAVVEMDNGVLKVIKSKDAKTSaipVLSGEERANGDEVITQVAV
KNVSVRELSPLLRQLIDNAGAGNVVHYDPANIILITGRAAVVNRRLAEIIRRVDOAGDKEIEVVVELNNASAAEMVRIVEALNKTIDAQNTPEFLKPKFVADERTNS
ILISGDPKVRERLKRLIKQLDVEMAAKGNRRVVYLKYAKAEDLVEVLKGVSENLQAEKGTGQPTTSKRNEVMIAAHADTNSSLVLTAPQDIMNAMLEVIGQLDIR
RAQVLEALIVEMAEGDGINLGVOQWGSLESGSVIQYGN TGASIGNVMIGLEEAKDTTQTKAVYDTNNNNFLRNETTTTKGDYTKLASALSSIQGAAVSIAMGDW
TALINAVSNDSSSNISSPSITVMDNGEASFIVGEEVPVITGSTAGSNNDNPFQTVDRKEVGIKLVVPQINEGNSVQLNIEQEVSNVLGANGAVDVRFAKRQLN
TSVMVQDGMVLVGLGLDERALESESKVPLLGDIPLLGQLFRSTSSQVEKKNLMVFIKPTIIRDGVTADGITQRKYNVIRAEQLFRAEKGLRLDDASVPVLPKFG
DDRRHSPEIQAFIEQMEAKQ

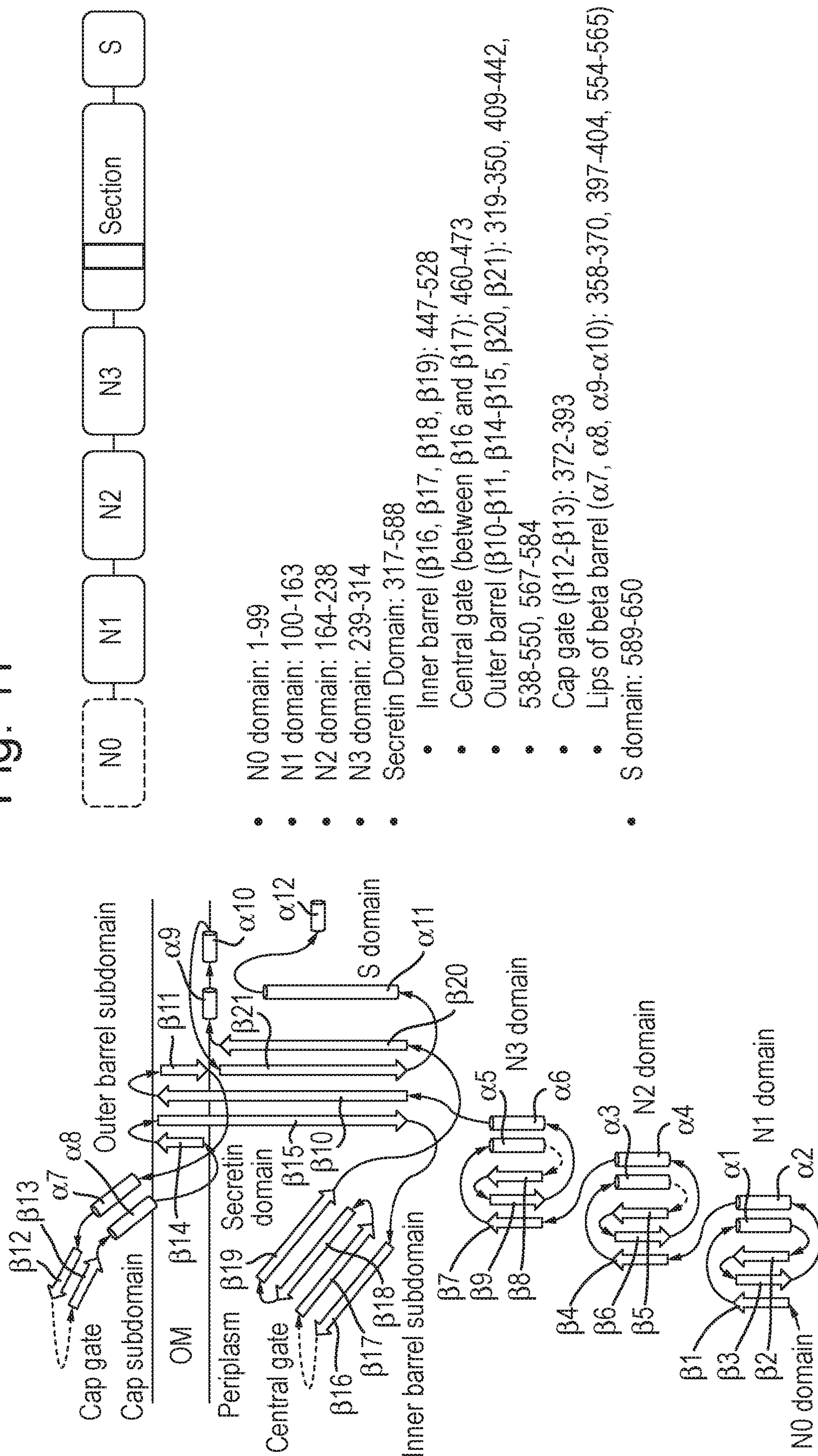
The following regions are missing in the crystal structure:

- N terminus N0 domain (1-96)
- N1 and N2 domains (97-239)
- Constriction loop (461-473)(GSTAGSNNDNPFQ)
- Trans tip loop (370-379) (YDTNNNFLRN)
- Gate loop in N3 (265-282) (NLQAEKGTGQPTTSKRNE)

Uniport - P45779 - GSPD_VIBCH Type II secretion system protein

MIKYWLKSSWLLAGSLLSTPLAMANEFSASFKGTDIQEFINIVGRNLEKTIIVDPSVRGKVDVRSFDTLNEEQYYSFFLSVLEVYGFVAVVEMDNGVLKVIKSKDAK
TSAIPVLSGEERANGDEVITQVAVKNVSVRELSPLLRQLIDNAGAGNVVHYDPANIILITGRAAVVNRRLAEIIRRVDOAGDKEIEVVVELNNASAAEMVRIVEALN
KTIDAQNTPEFLKPKFVADERTNSILISGDPKVRERLKRLIKQLDVEMAAKGNRRVVYLKYAKAEDLVEVLKGVSENLQAEKGTGQPTTSKRNEVMIAAHADTN
SLVLTAPQDIMNAMLEVIGQLDIRRAQVLEALIVEMAEGDGINLGVOQWGSLESGSVIQYGN TGASIGNVMIGLEEAKDTTQTKAVYDTNNNFLRNETTTTKG
DYTKLASALSSIQGAAVSIAMGDWTALINAVSNDSSSNISSPSITVMDNGEASFIVGEEVPVITGSTAGSNNDNPFQTVDRKEVGIKLVVPQINEGNSVQLNIE
QEVSNVLGANGAVDVRFAKRQLNTSVMVQDGMVLVGLGLDERALESESKVPLLGDIPLLGQLFRSTSSQVEKKNLMVFIKPTIIRDGVTADGITQRKYNVIRAE
QLFRAEKGLRLDDASVPVLPKFGDDRRHSPEIQAFIEQMEAKQ

Signal peptide (1-24) underlined.



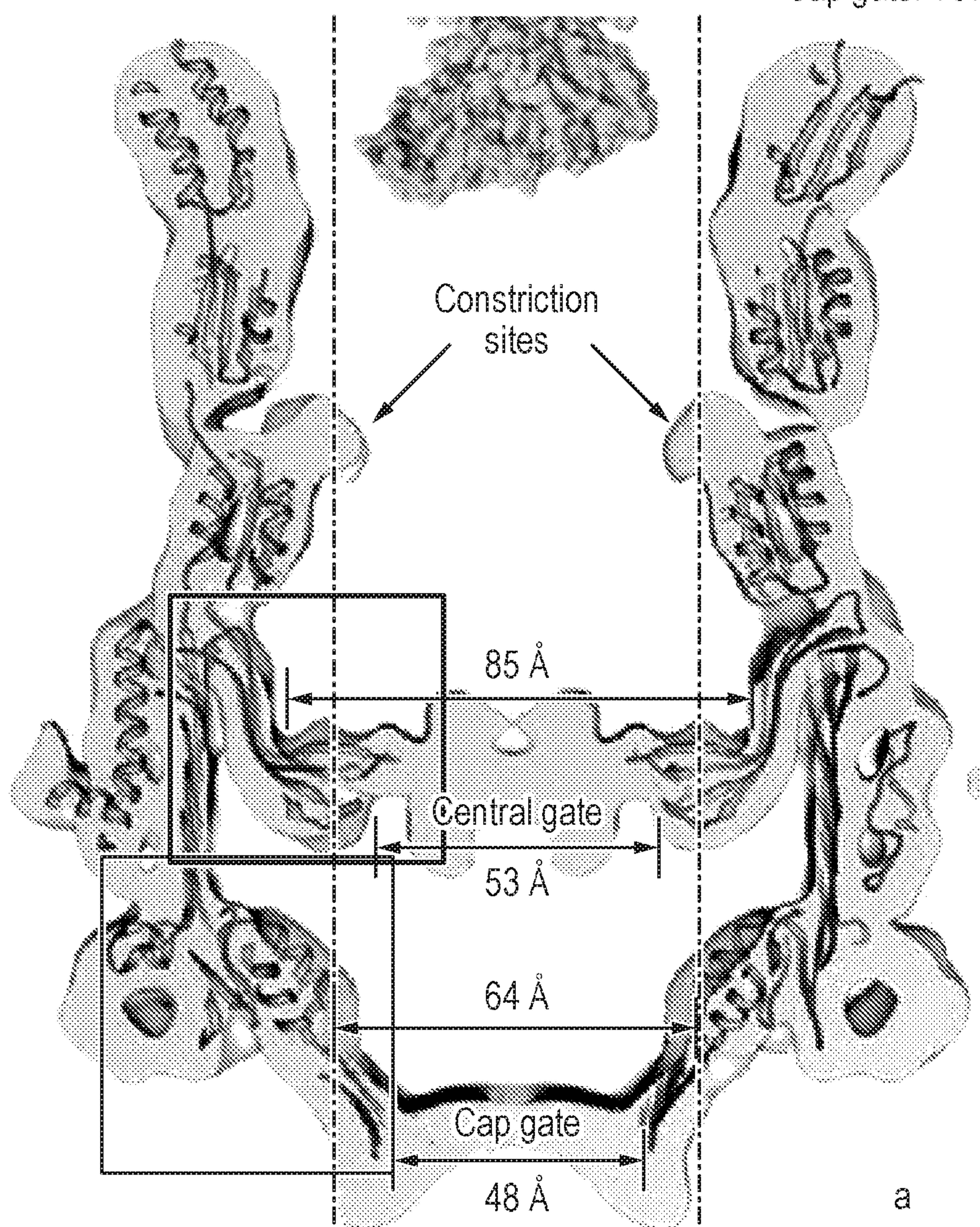
17/28

Fig. 12

GspD from *Vibrio Cholera*

Missing in Crystal structure:

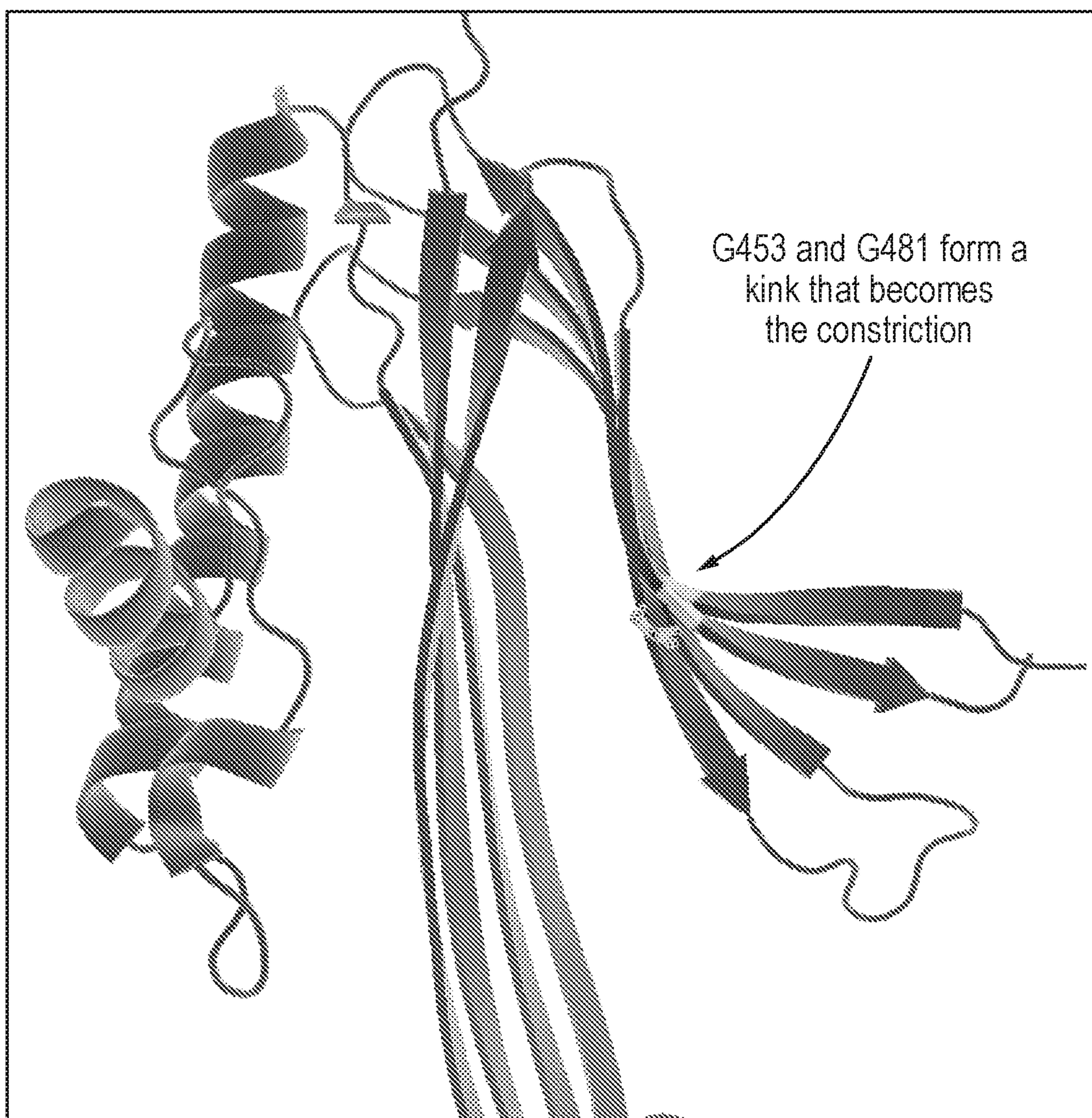
- Constriction site: N265-E282
- Central gate: G461-Q473
- Cap gate: Y379-R387



Yan et.al. Nature structure and Molecular biology (2017)

18/28

Fig. 13



19/28

Fig. 14A

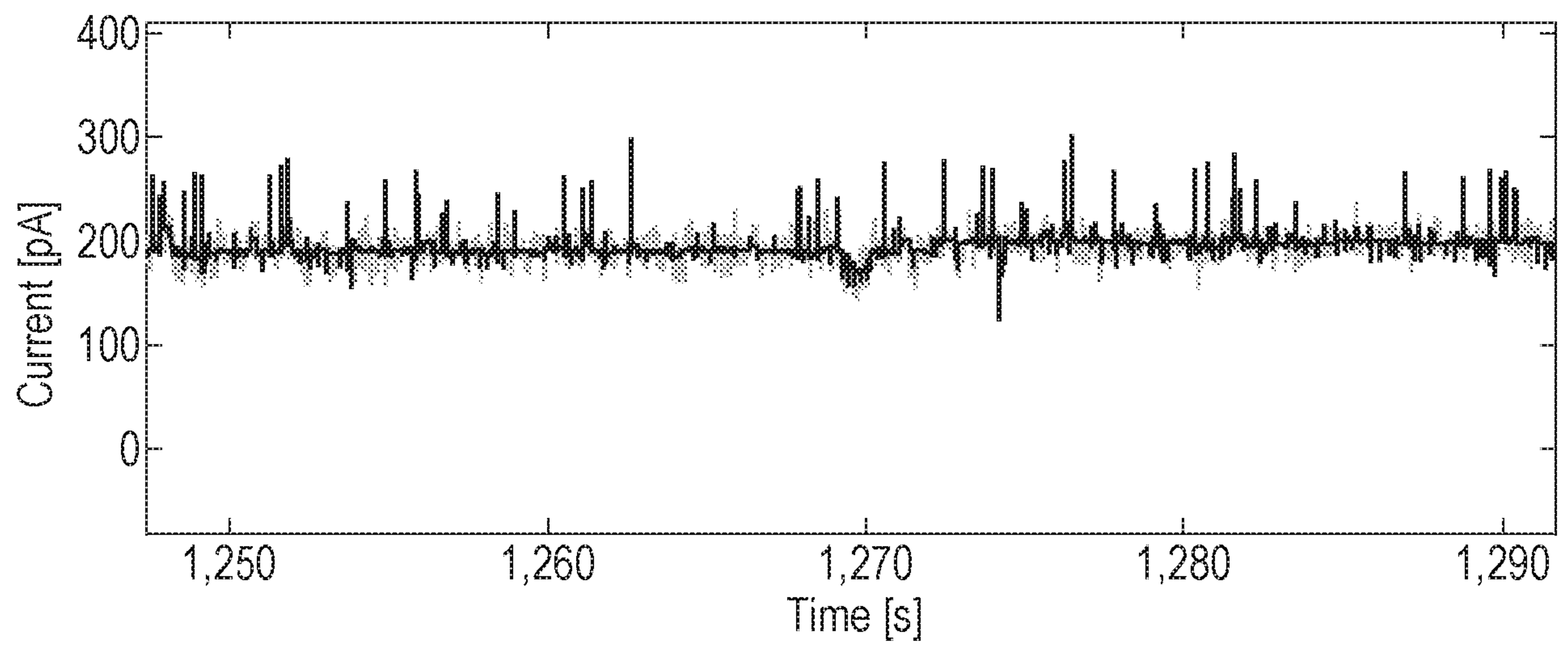
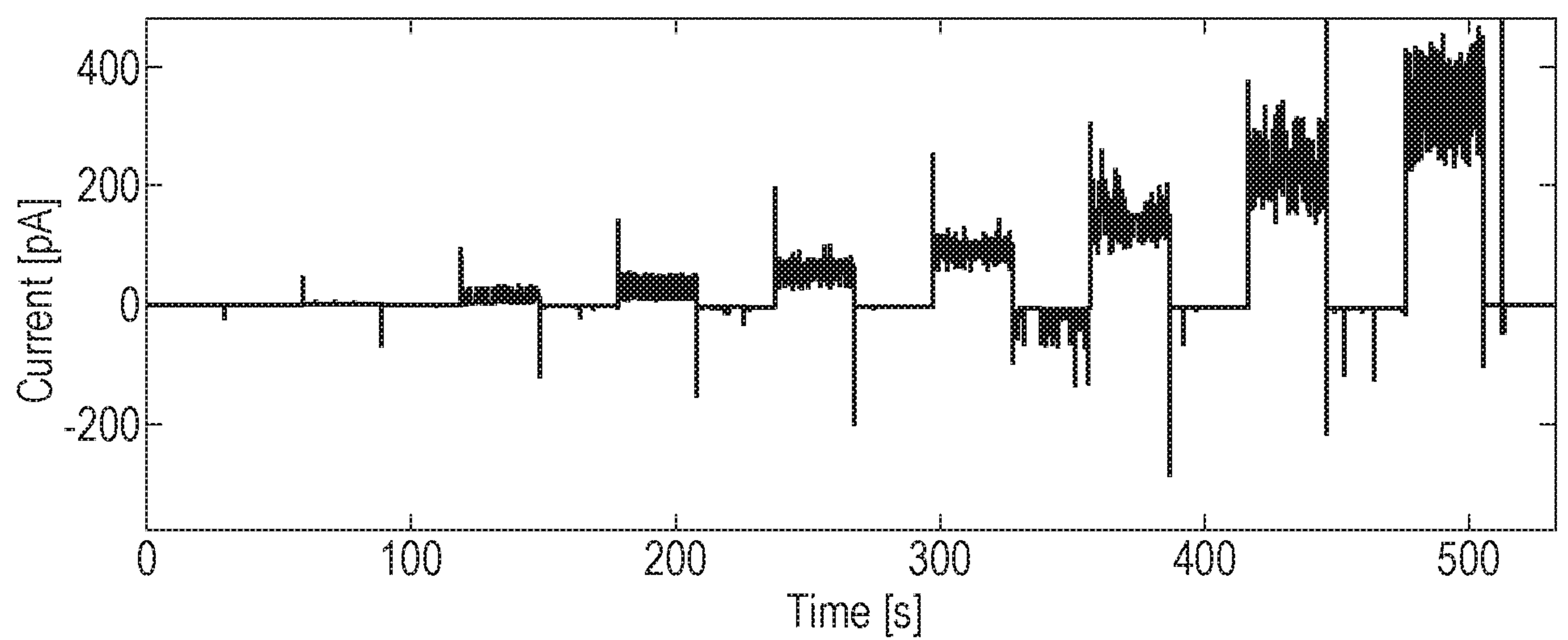


Fig. 14B



20/28

Fig. 15A

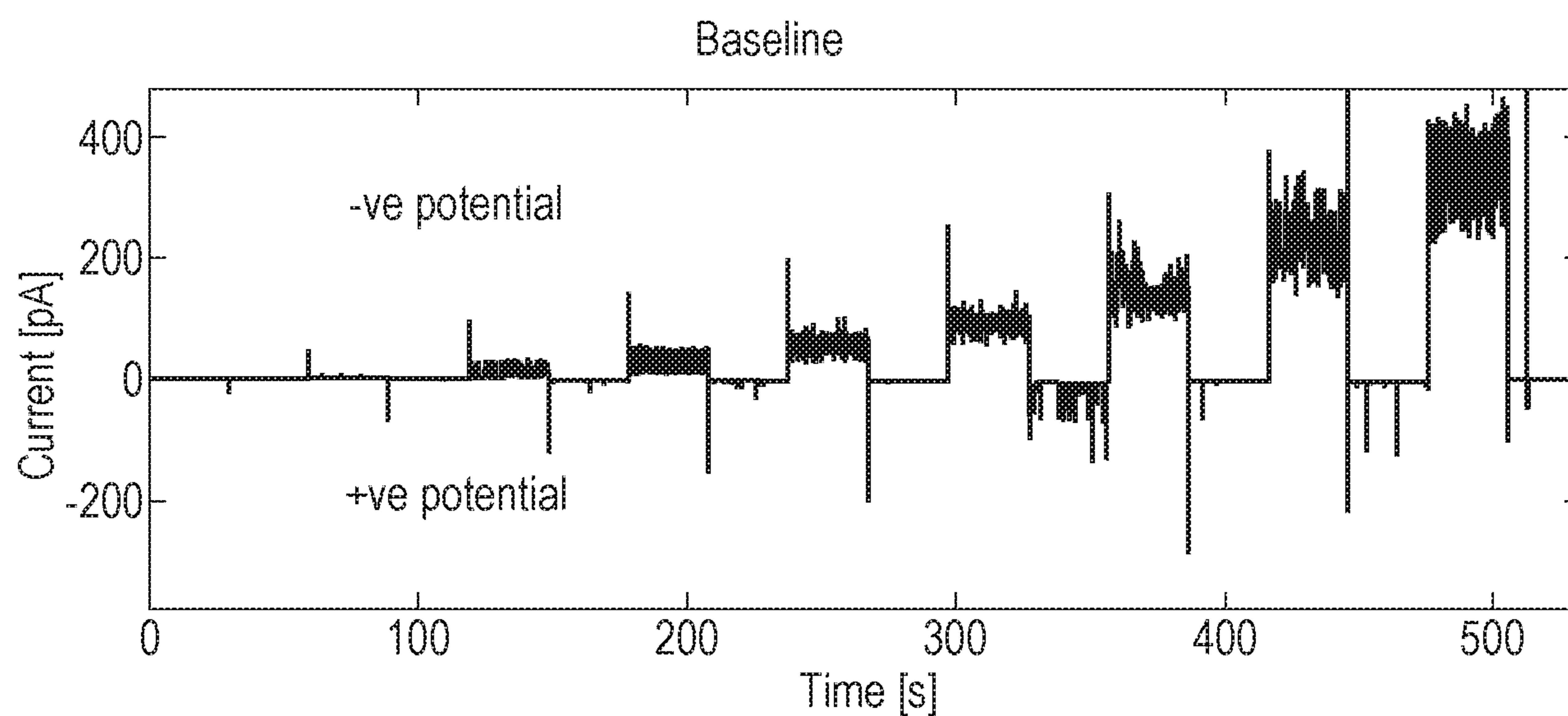
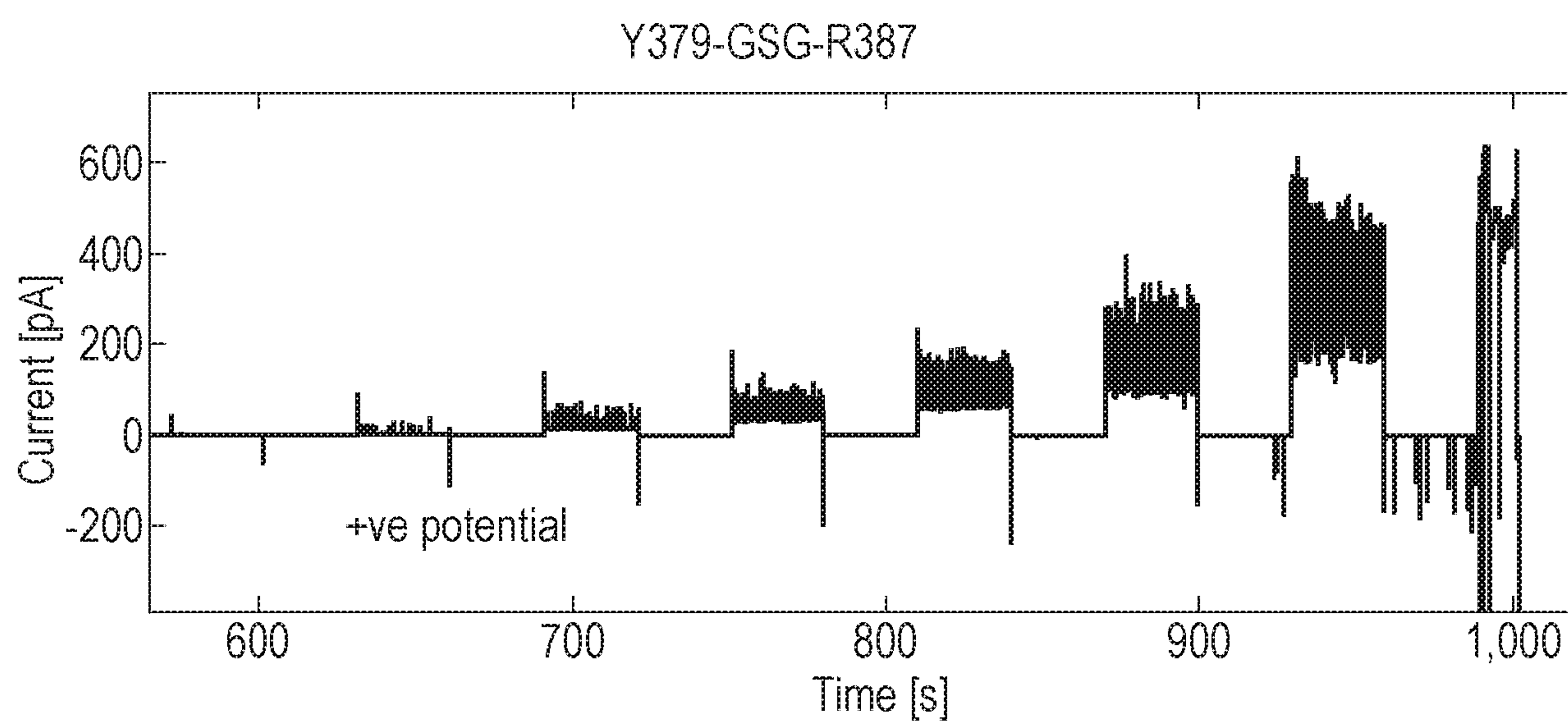


Fig. 15B



21/28

Fig. 15C

F472A

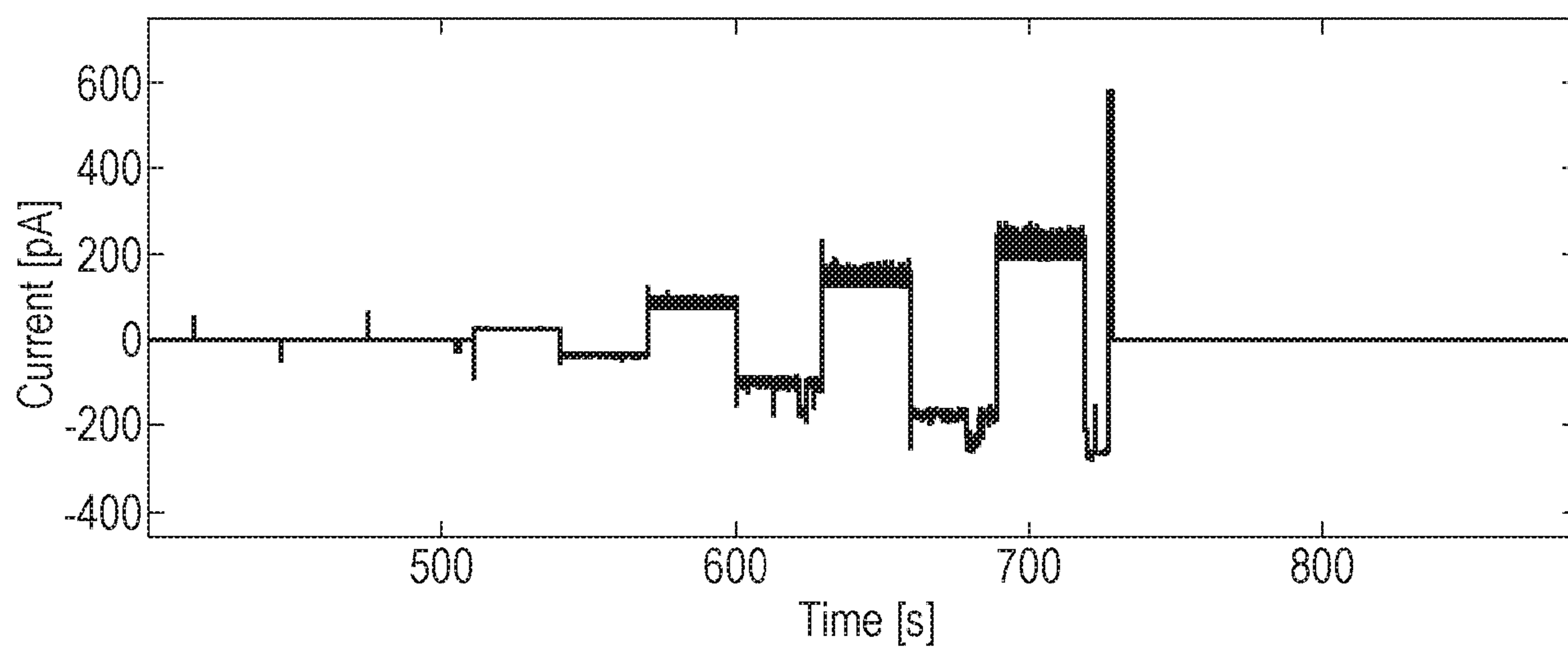
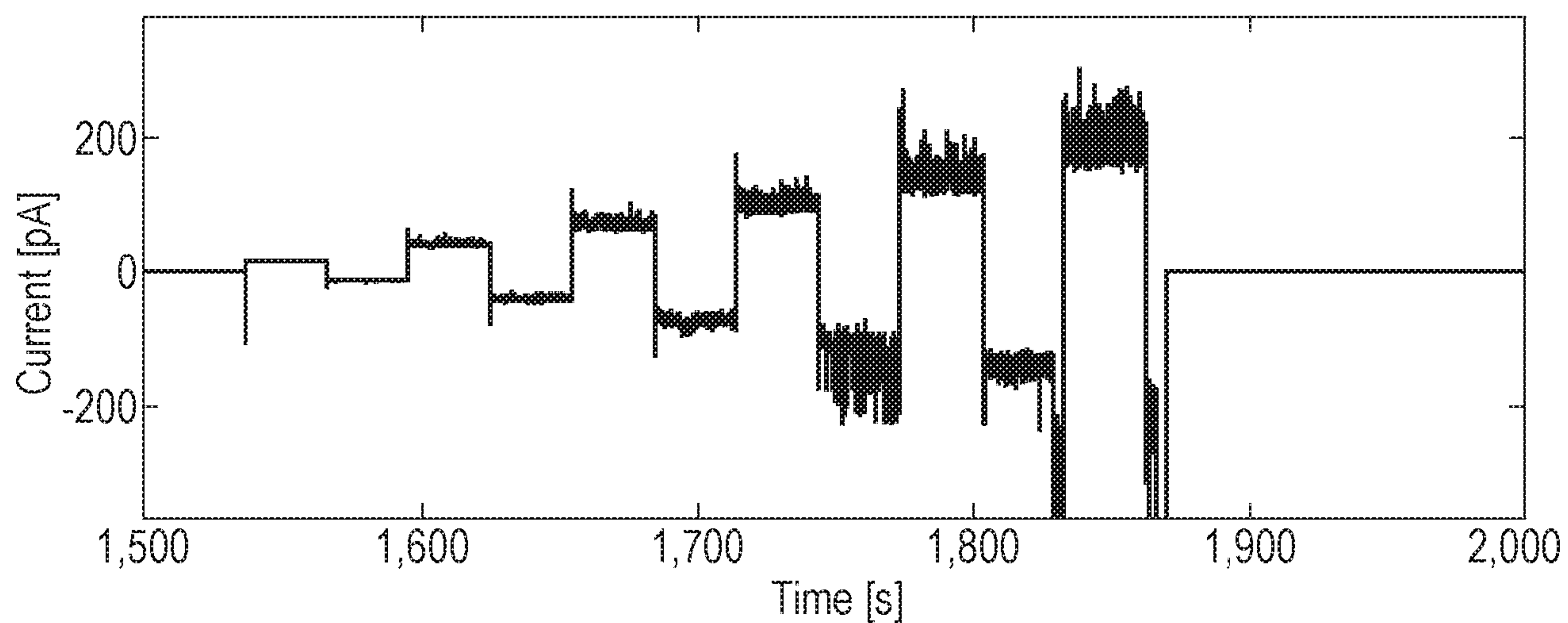


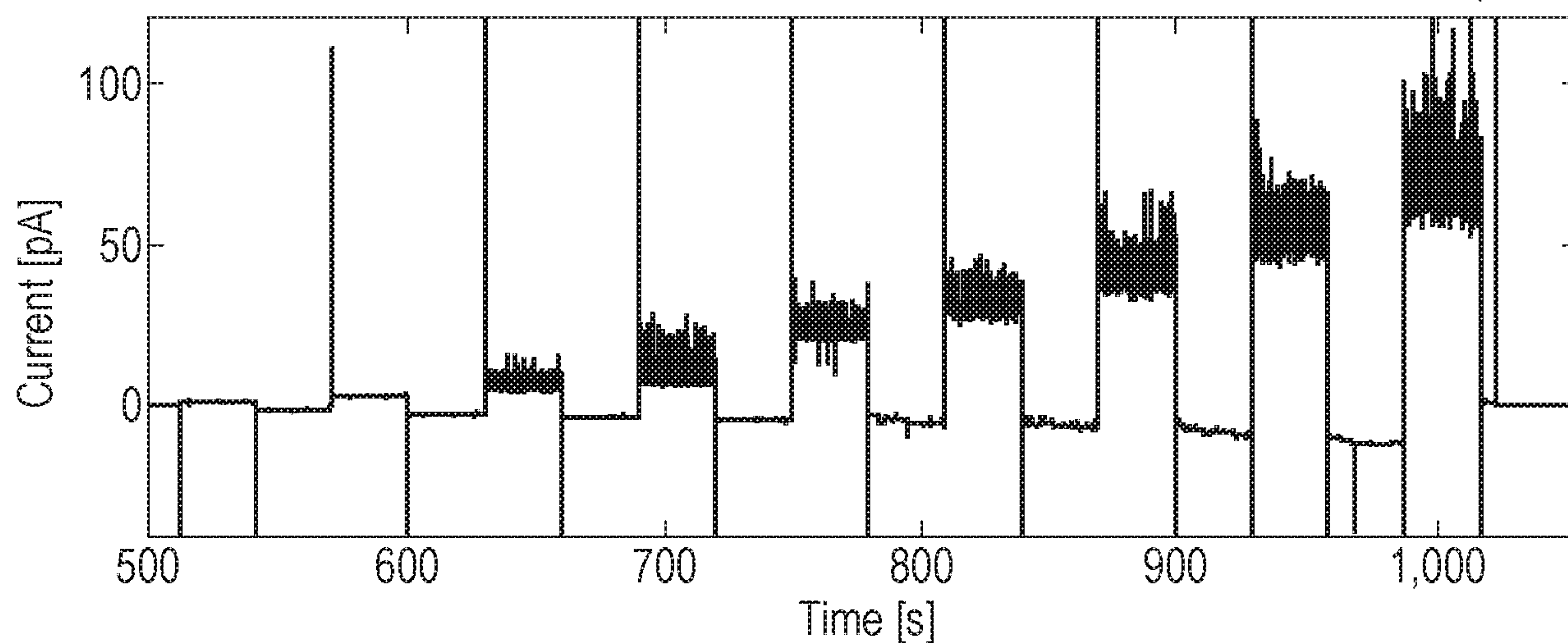
Fig. 15D

D469S

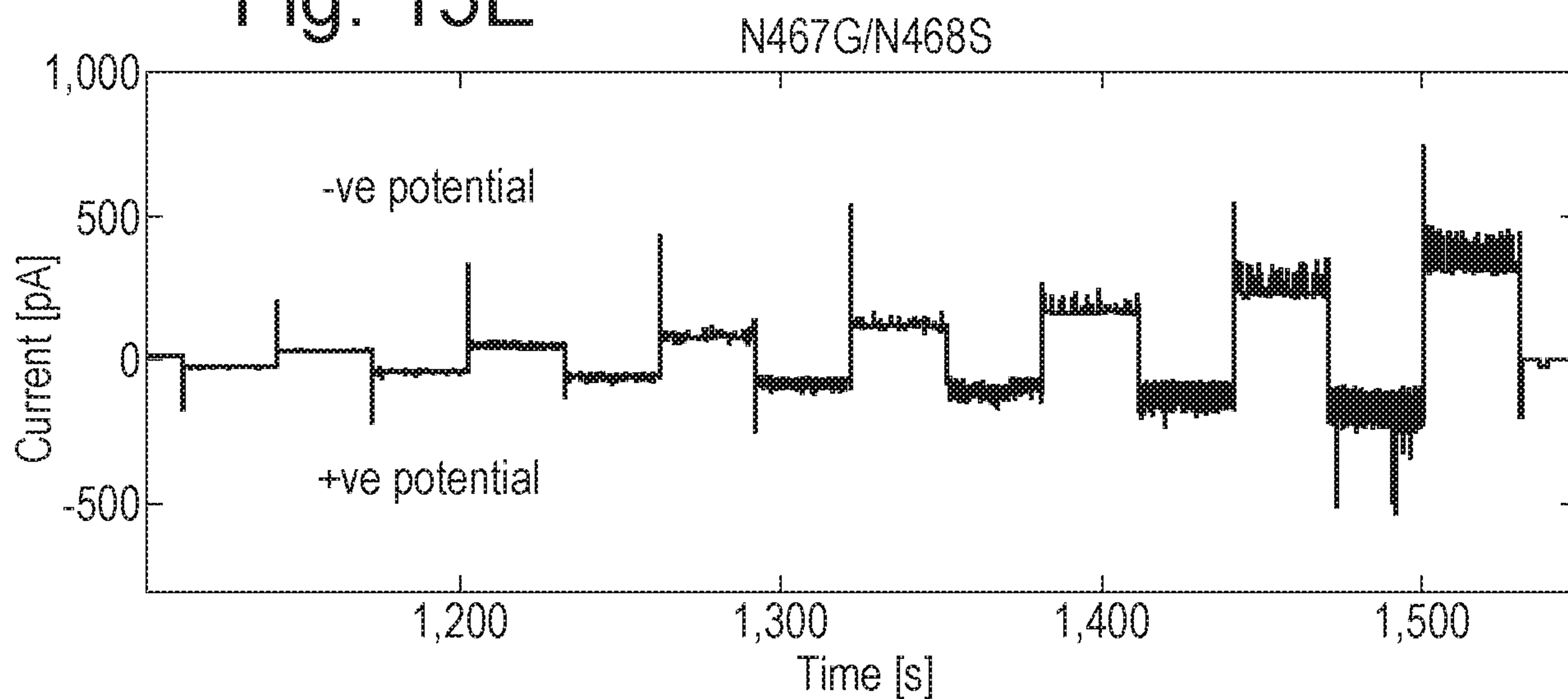
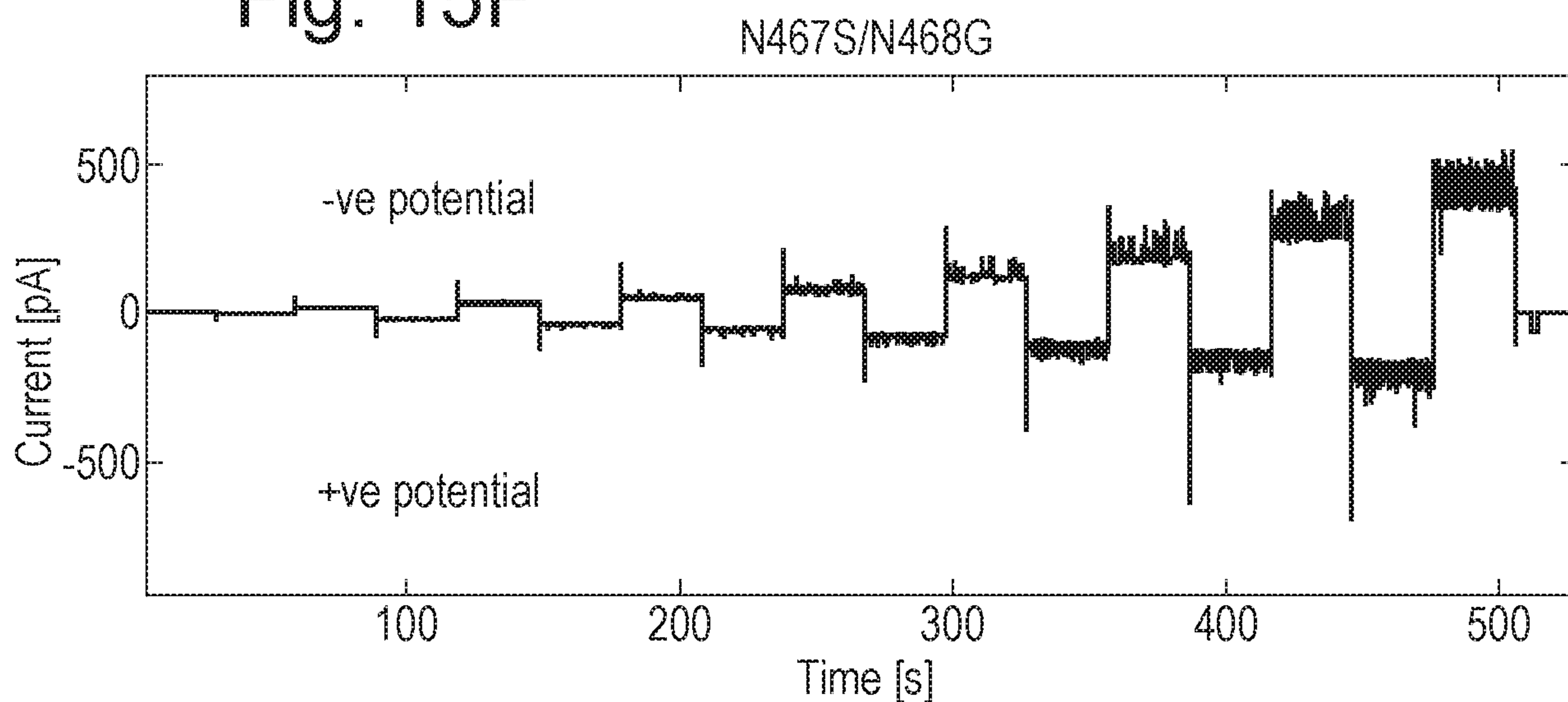
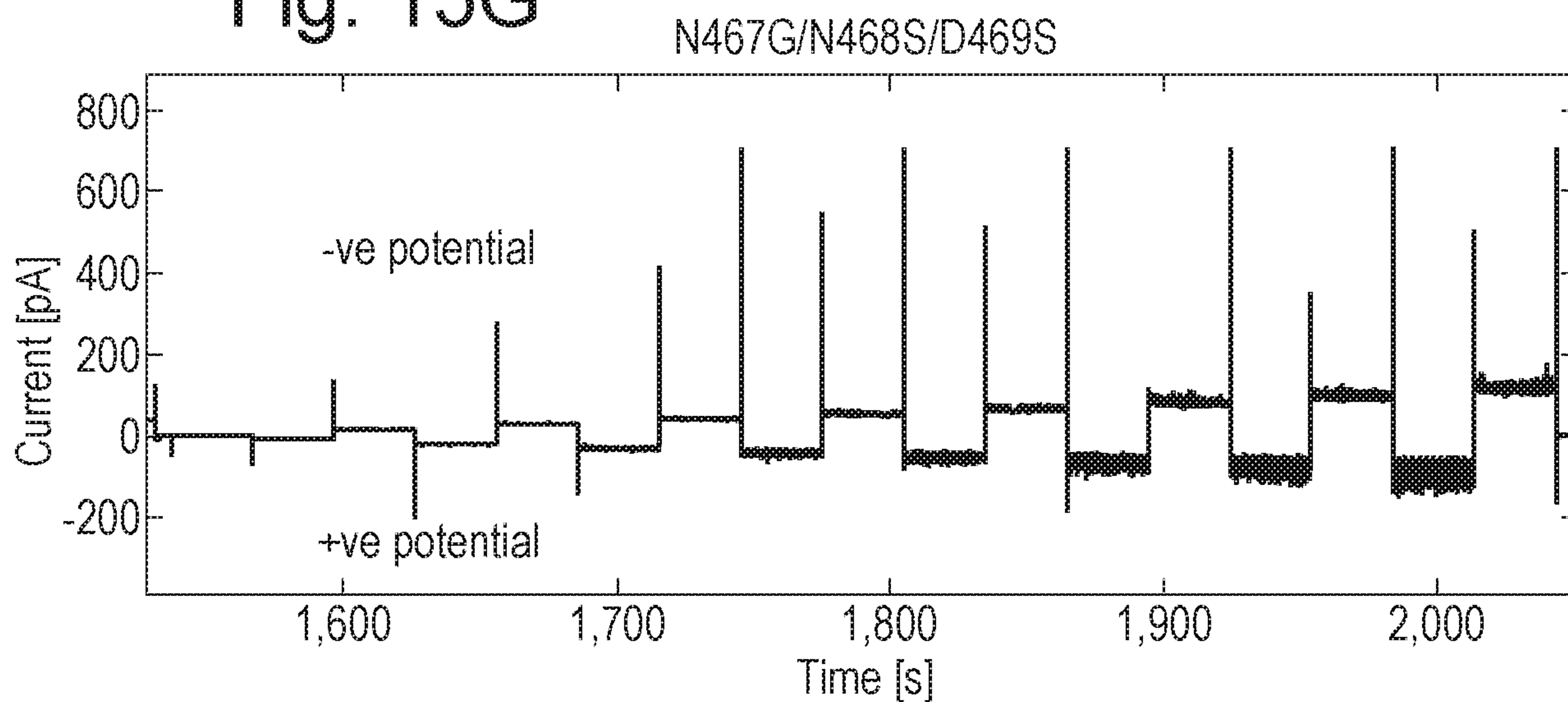
Bigger pore



Smaller pore



22/28

Fig. 15E**Fig. 15F****Fig. 15G**

23/28

Fig. 16A

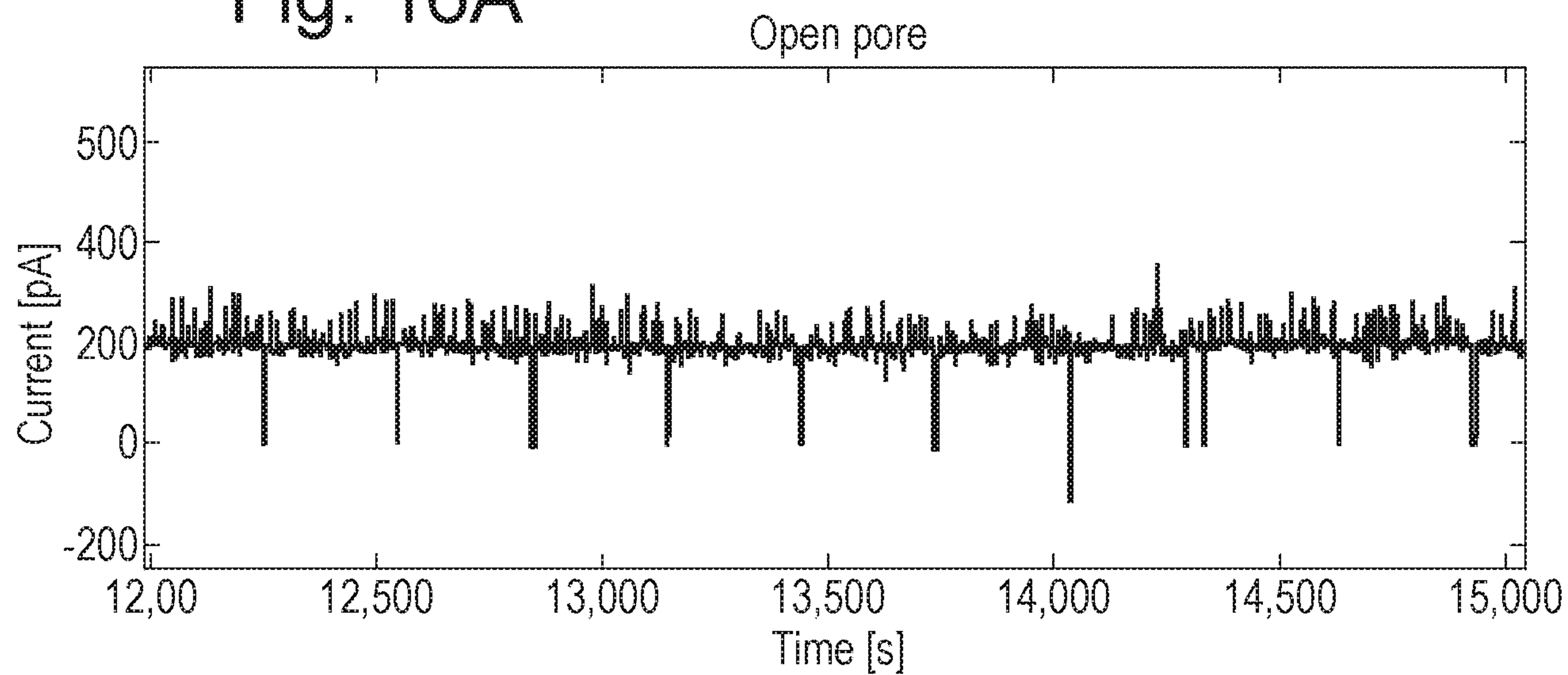


Fig. 16B

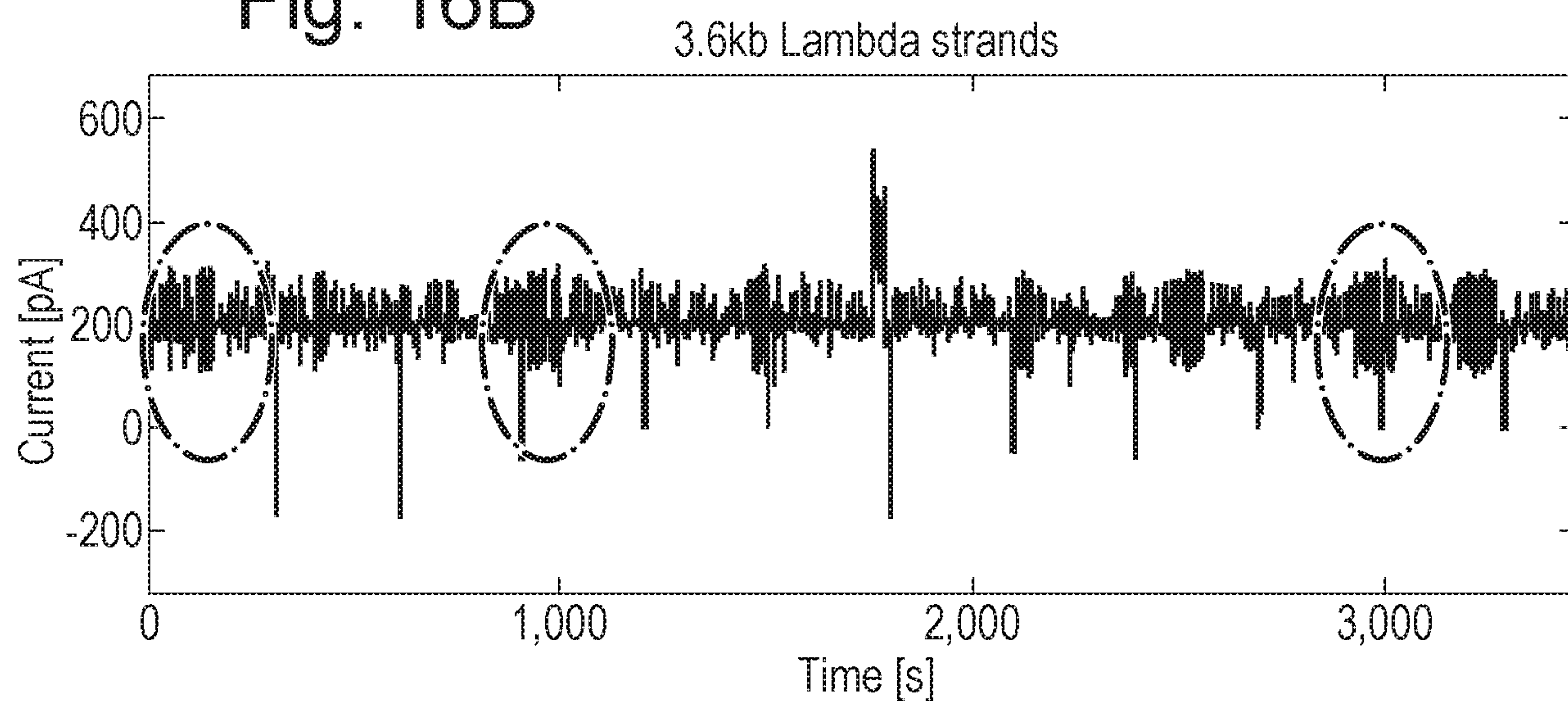
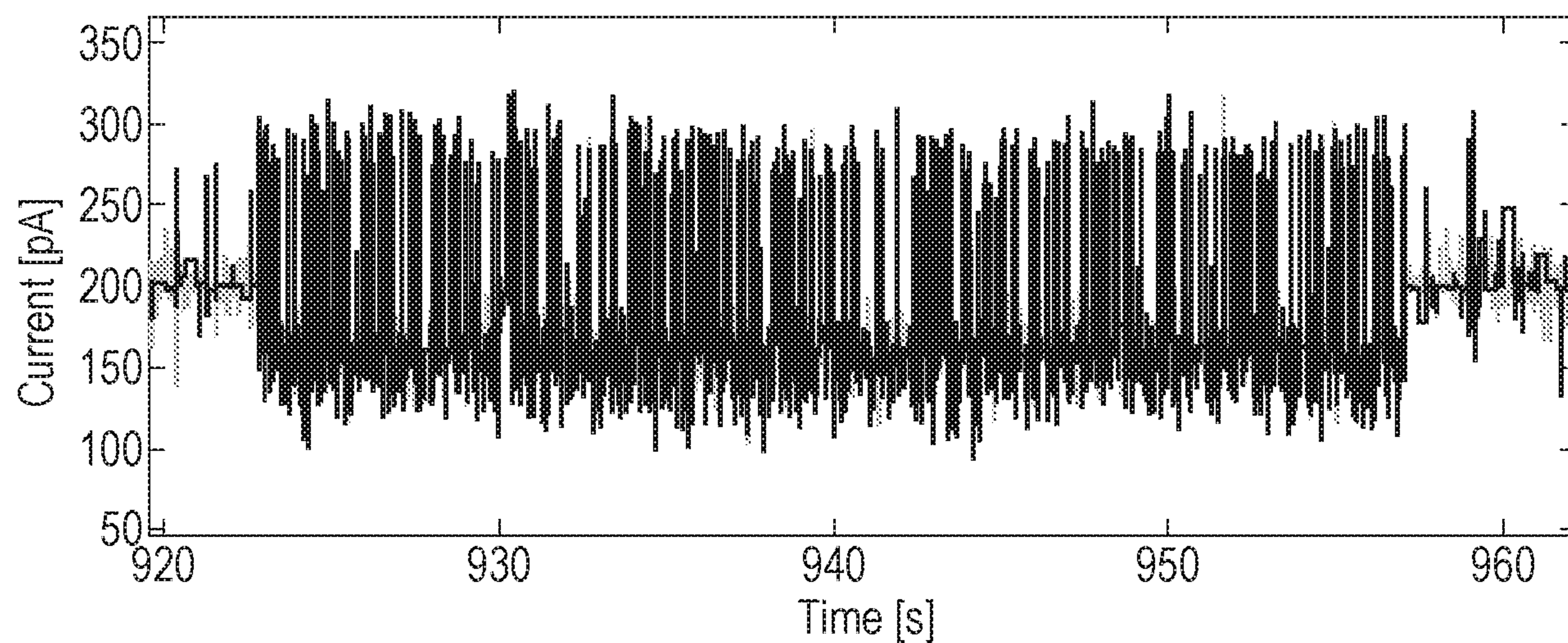
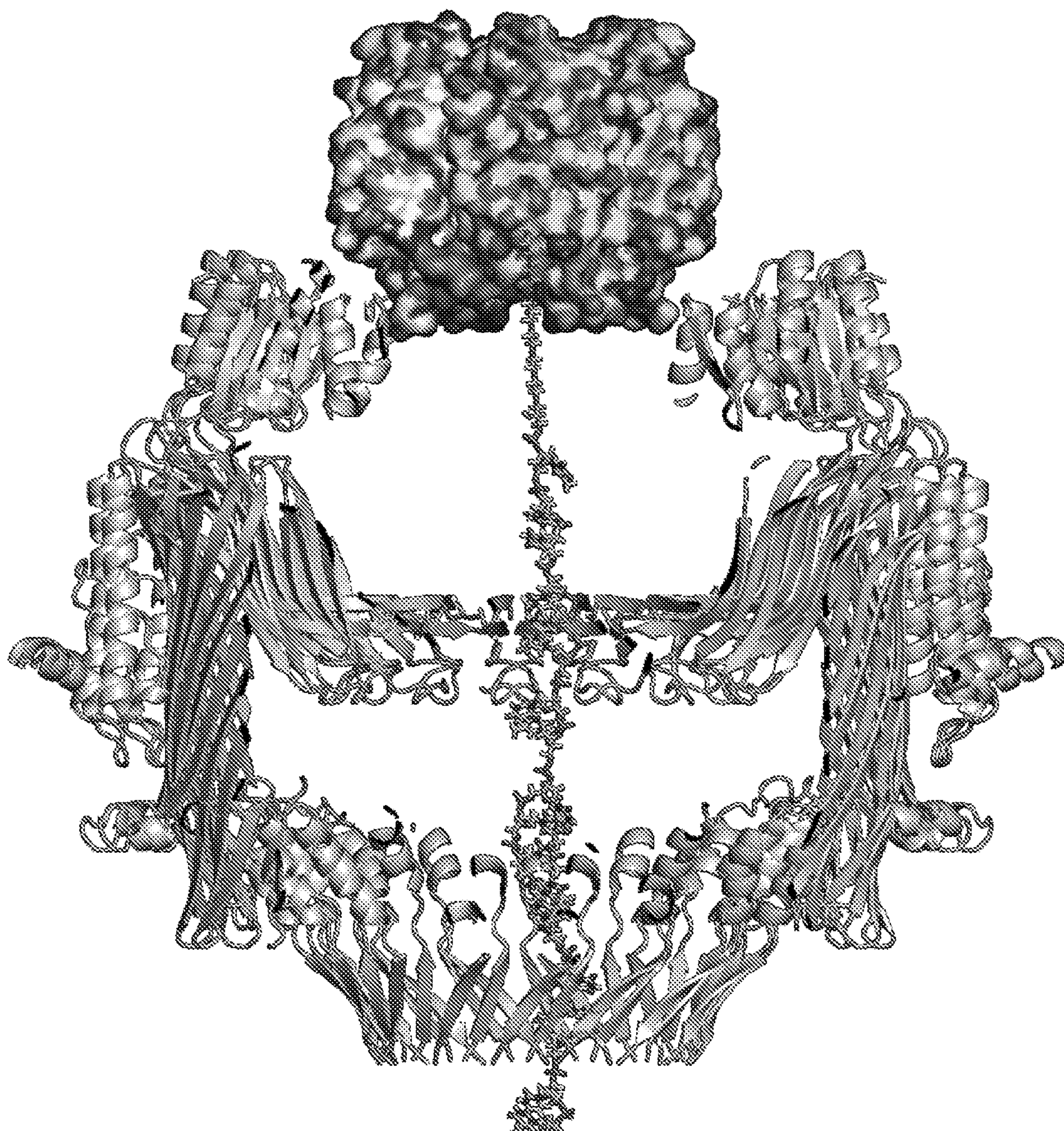


Fig. 16C



24/28

Fig. 17A

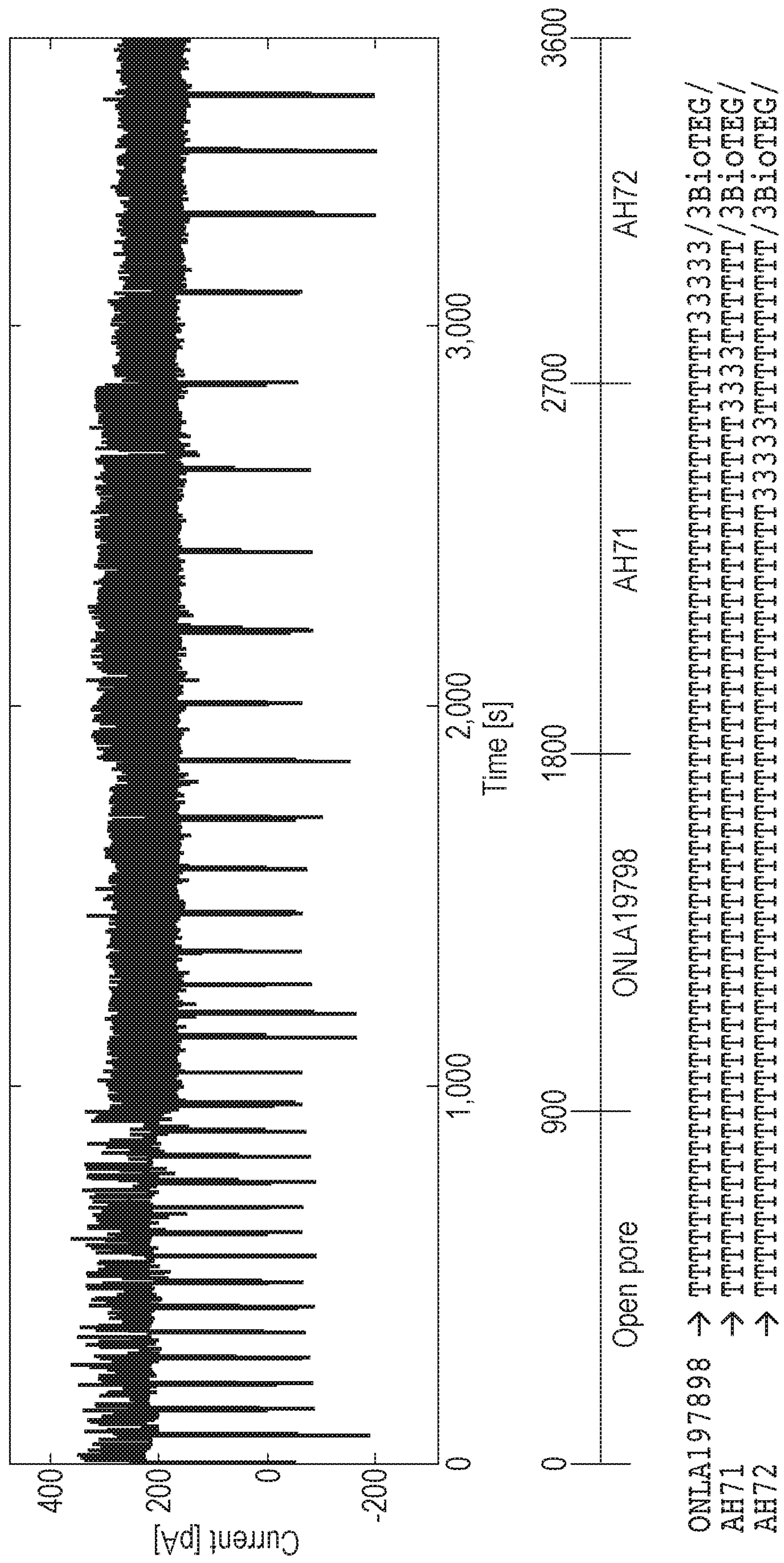


25/28

Fig. 17B



18A



27/28

Fig. 18B

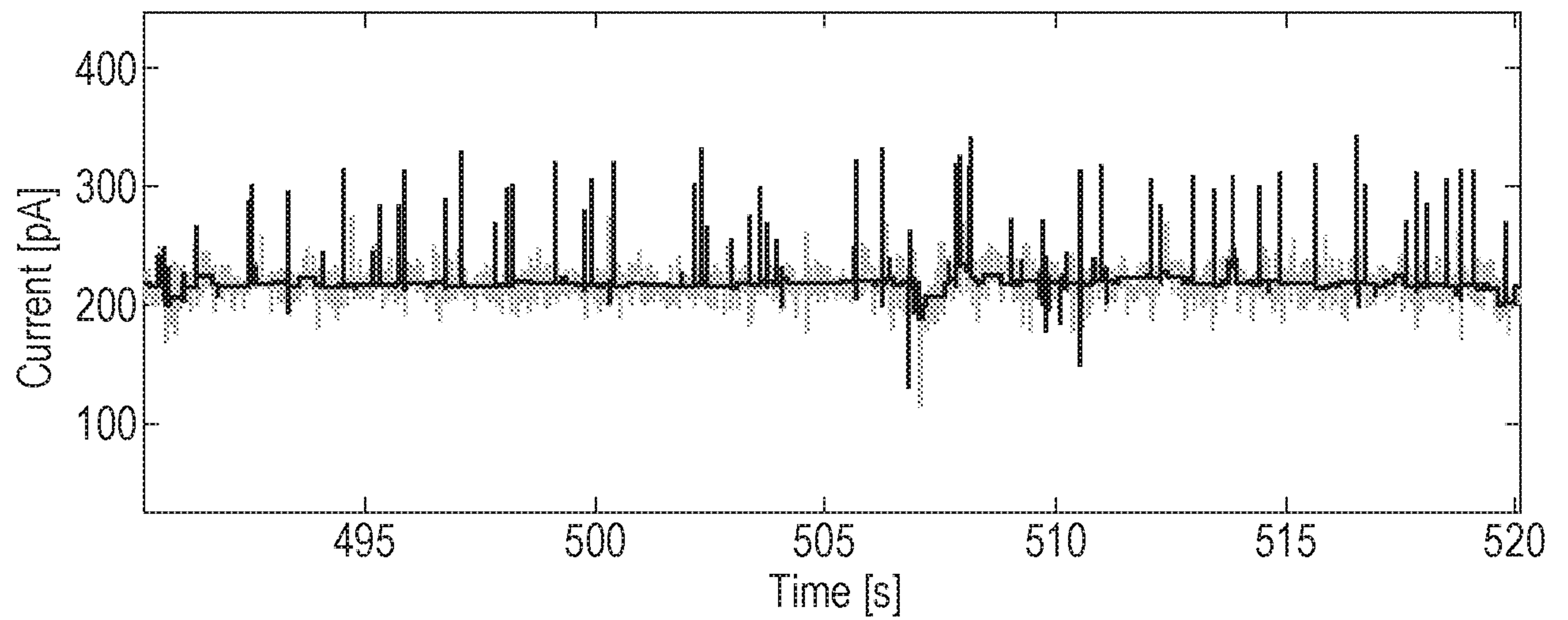
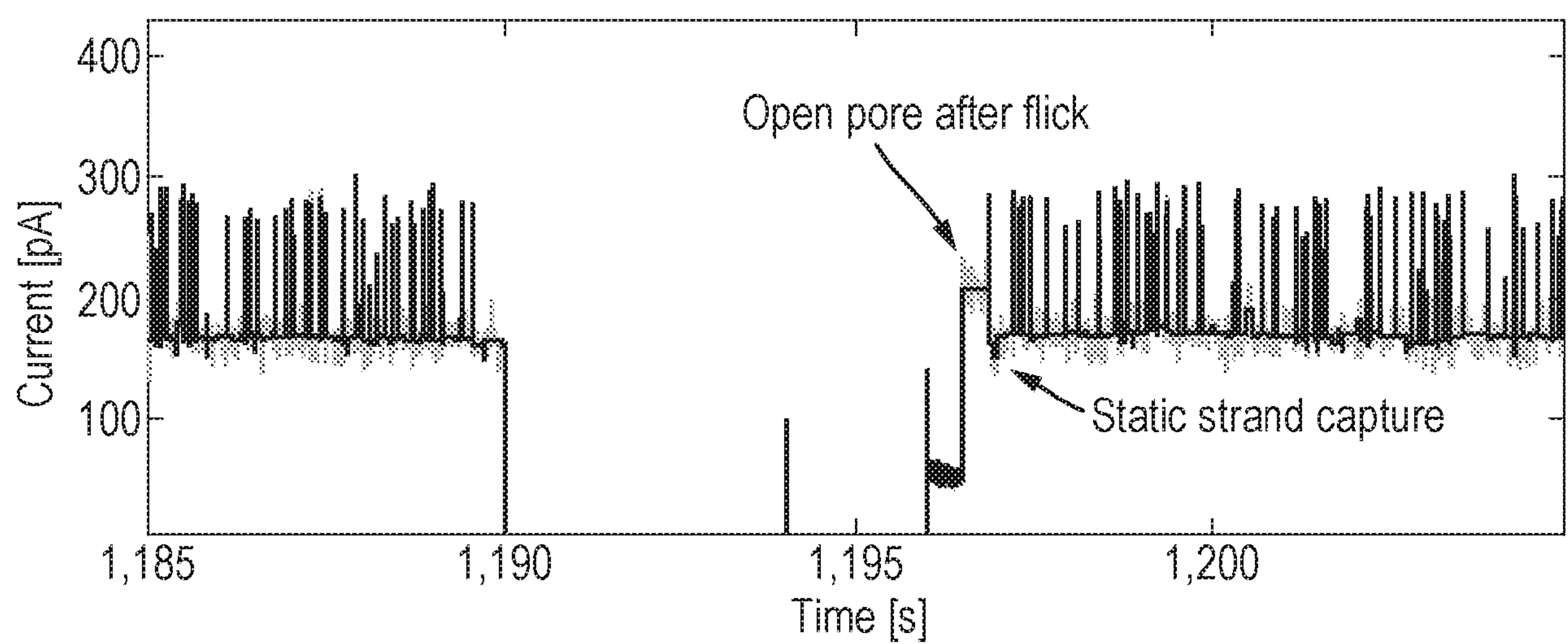


Fig. 18C



28/28

Fig. 18D

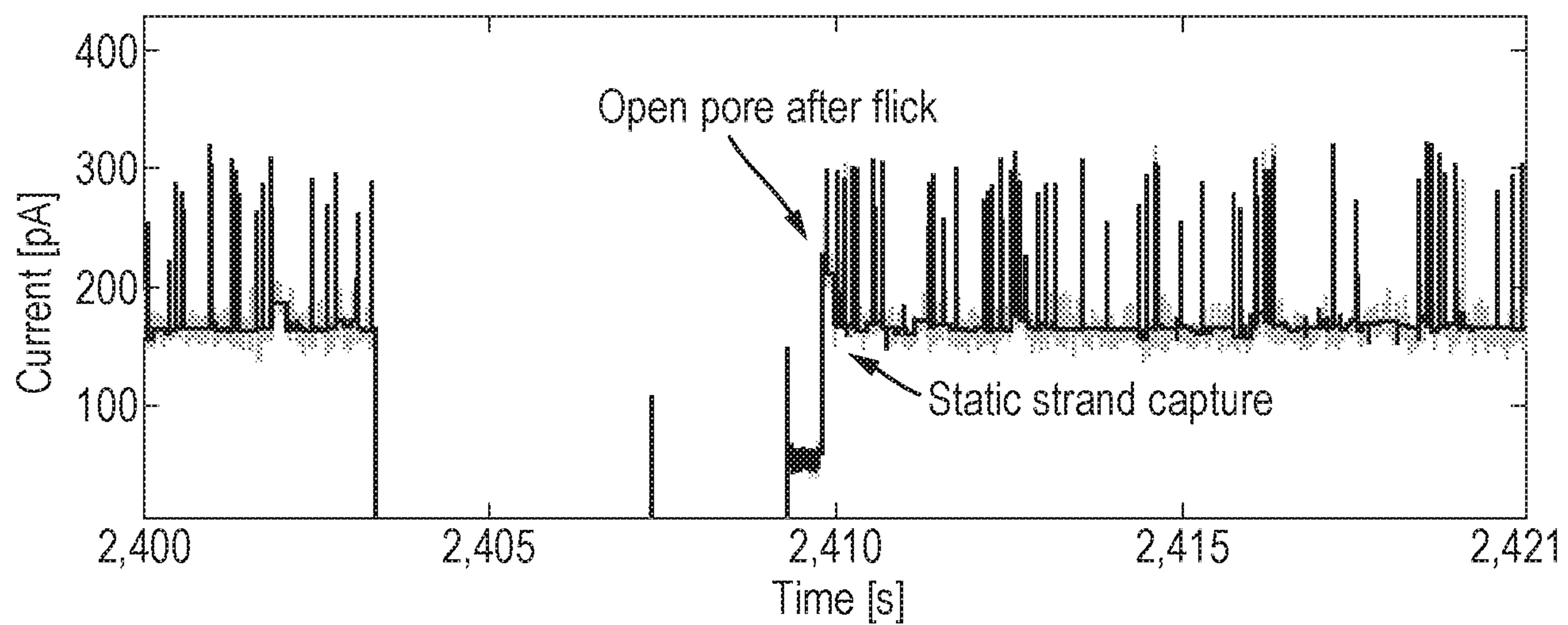


Fig. 18E

