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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 October 2015 (08.10.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/150786 A1

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

(21) International Application Number:
PCT/GB2015/050991

(22) International Filing Date:
31 March 2015 (31.03.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1406147.7 4 April 2014 (04.04.2014) GB
1407815.8 2 May 2014 (02.05.2014) GB
PCT/GB2014/052737 10 September 2014 (10.09.2014) GB

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

(72) Inventors: **CRAWFORD, Marion Louise**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WHITE, James**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB).

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

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

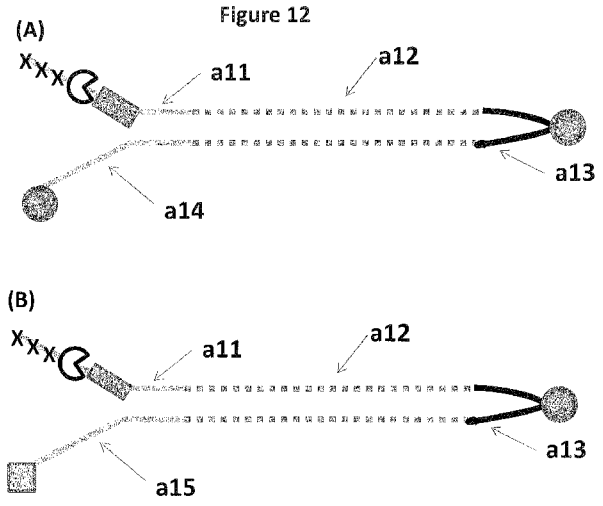
Published:

— *with international search report (Art. 21(3))*

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

(54) Title: METHOD FOR CHARACTERISING A DOUBLE STRANDED NUCLEIC ACID USING A NANO-PORE AND ANCHOR MOLECULES AT BOTH ENDS OF SAID NUCLEIC ACID

Figure 12



(57) Abstract: The invention relates to a method for method of characterising, such as sequencing, a target double stranded polynucleotide. The polynucleotide is coupled to a membrane using at least two adaptors with different strengths of coupling to the membrane.

WO 2015/150786 A1

METHOD FOR CHARACTERISING A DOUBLE STRANDED NUCLEIC ACID USING A NANO-PORE AND ANCHOR MOLECULES AT BOTH ENDS OF SAID NUCLEIC ACID

Field of the invention

The invention relates to a method of characterising, such as sequencing, a target double stranded polynucleotide. The polynucleotide is coupled to a membrane using at least two adaptors with different strengths of coupling to the membrane.

Background of the invention

There is currently a need for rapid and cheap polynucleotide (e.g. DNA or RNA) sequencing and identification technologies across a wide range of applications. Existing technologies are slow and expensive mainly because they rely on amplification techniques to produce large volumes of polynucleotide and require a high quantity of specialist fluorescent chemicals for signal detection.

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

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

It has previously been demonstrated that ultra low concentration analyte delivery can be achieved by coupling the analyte to a membrane in which the relevant detector is present. This lowers by several orders of magnitude the amount of analyte required in order to be detected (WO 2012/164270).

It has also been shown that double stranded polynucleotides can be effectively characterised using strand sequencing if they are modified to include a Y adaptor (a double stranded stem and two non-complementary arms) containing a leader sequence and a hairpin loop adaptor (WO 2013/014451). It is preferred that that Y adaptor containing the leader sequence is attached to one end of the polynucleotide and the hairpin loop adaptor is attached to the other end. The leader sequence preferentially threads into the nanopore and the hairpin loop connecting the two strands of the polynucleotide allows both strands to be investigated as the polynucleotide unzips and moves through the pore. This is advantageous because it doubles the amount of information obtained from a single double stranded polynucleotide. Moreover,

because the sequences in the two strands are complementary, the information from the two strands can be combined informatically. This mechanism provides an orthogonal proof-reading capability that provides higher confidence observations. When Y adaptors and hairpin loops are used together in this way, the Y adaptor typically contains an anchor which couples the polynucleotide to the membrane containing the nanopore. In some instances, double stranded polynucleotides having Y adaptors at both ends are produced in the sample preparation. The presence of the leader sequence and the anchor in the Y adaptors means that the system is typically biased towards characterising these polynucleotides. However, the lack of the hairpin loop linking the two strands in such polynucleotides means that only one strand is investigated.

Summary of the invention

The inventors have surprisingly demonstrated that it is possible to avoid a bias towards double stranded polynucleotides comprising a Y adaptor at both ends by including two anchors in the system. The inventors have used a Y adaptor comprising one or more first anchors for coupling the polynucleotide to the membrane and a hairpin loop adaptor comprising one or more second anchors for coupling the polynucleotide to the membrane. The hairpin loop adaptor couples the polynucleotide to the membrane with more strength than the Y adaptor.

Accordingly, the invention provides a method of characterising a target double stranded polynucleotide using a transmembrane pore in a membrane, comprising:

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

b) contacting the polynucleotide provided in step a) with the transmembrane pore such that at least one strand of the polynucleotide moves through the pore; and

c) taking one or more measurements as the at least one strand of the polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the at least one strand of the polynucleotide and thereby characterising the double stranded target polynucleotide.

The invention also provides:

- a method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising ligating a Y adaptor to one end of the polynucleotide and ligating a hairpin loop adaptor to the other end of the polynucleotide, wherein

the Y adaptor comprises one or more first anchors for coupling the polynucleotide to the membrane, wherein the hairpin loop adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby providing a modified target double stranded polynucleotide;

- a method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising: (a) ligating a Y adaptor to one end of the polynucleotide and ligating a hairpin loop adaptor to the other end of the polynucleotide; and (b) attaching to the Y adaptor one or more first anchors for coupling the polynucleotide to the membrane, attaching to the hairpin loop adaptor one or more second anchors for coupling the polynucleotide to the membrane and thereby providing a modified target double stranded polynucleotide; wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane;

- a target double stranded polynucleotide modified using the method of the invention;

- a method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising contacting the target polynucleotide with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein a proportion of the substrates in the population are hairpin loop adaptors comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby producing a plurality of modified double stranded polynucleotides;

- a method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising: (a) contacting the target polynucleotide with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors and wherein a proportion of the substrates in the population are hairpin loop adaptors, such that the transposase fragments the target polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs; and (b) attaching to the Y adaptors in the plurality of fragment/substrate constructs one or more first anchors and attaching to the hairpin loop adaptors in the plurality of fragment/substrate constructs one or more second anchors and thereby producing a plurality of modified double stranded polynucleotides; wherein the strength of coupling of each hairpin loop adaptor to the membrane is greater than the strength of coupling of each Y adaptor to the membrane;

- a plurality of polynucleotides modified using the method of the invention;
- a method of characterising a target double stranded polynucleotide using a transmembrane pore in a membrane, comprising:

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

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

c) detecting the phosphate labelled species using the pore and thereby characterising the double stranded target polynucleotide;

- a pair of adaptors for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, wherein one adaptor is a Y adaptor comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein the other adaptor is a hairpin loop adaptor comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane;
- a population of adaptors for modifying a target polynucleotide for characterisation using a transmembrane pore in a membrane, wherein a proportion of the adaptors are Y adaptors comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein a proportion of the adaptors are hairpin loop adaptors comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptors to the membrane is greater than the strength of coupling of the Y adaptors to the membrane; and
- a kit for modifying a target polynucleotide comprising (a) a pair of adaptors or a population of adaptors of the invention and (b) a MuA transposase.

Description of the Figures

Figure 1 shows in sections (A) and (B) cartoon representations of the DNA constructs used in Example 2 – four iSpC3 spacers are shown as a cross and four 5-nitroindoles as a black

0, the cholesterol coupling agent as a grey oval and the palmitate coupling agent as a grey square; labels a1-10 are described in full in Example 1.

Figure 2 shows in sections (A) and (B) cartoon representations of the DNA constructs used in Example 3 – four iSpC3 spacers are shown as a cross and four 5-nitroindoles as a black 0, the cholesterol coupling agent as a grey oval and the palmitate anchor as a grey square; labels a1-9 are described in full in Example 1.

Figure 3 shows the proportion of helicase-controlled DNA movements detected for construct 1 where strand a1 translocated through the nanopore and construct 2 where strand a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 1) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 2) helicase controlled DNA movements (labelled a2). This was observed to be approximately 50:50 when a single cholesterol was used to couple the constructs to the membrane. This control showed that if the same coupling agent was used to couple different constructs to the membrane, the number of helicase controlled DNA movements would have been approximately equal for each construct e.g. no bias was observed.

Figure 4 shows the proportion of helicase-controlled DNA movements detected for construct 2 where strand a2 translocated through the nanopore and construct 3 where strand a1 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a2 (construct 2) helicase-controlled DNA movements (labelled a2) to strand a1 (construct 3) helicase controlled DNA movements (labelled a1). This was observed to be approximately 5:95 when a single cholesterol was used to couple construct 2 to the membrane in comparison to when a cholesterol and a palmitate were used to couple construct 3 to the membrane. This experiment showed that when two coupling agents were used to couple a construct to the membrane (where one was a stronger coupling agent than the other e.g. cholesterol was stronger than palmitate) in comparison to a single coupling agent (cholesterol) the number of helicase-controlled DNA movements was strongly biased towards the doubly coupled construct over the singly coupled construct.

Figure 5 shows the proportion of helicase-controlled DNA movements detected for construct 4 where strand a1 translocated through the nanopore and construct 5 where strand a1 or

a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 4 or 5) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 5 only) helicase controlled DNA movements (labelled a2). This was observed to be approximately 33:66 (a2:a1 helicase controlled DNA movements) when two cholesterols were used to couple both constructs to the membrane. The 2:1 bias towards a1 helicase controlled DNA movements was predicted because a2 movements would only be detected if construct 4 was captured by the nanopore, whereas a1 movements would be detected from construct 4 and 5, therefore, twice as many a1 movements were expected.

Figure 6 shows the proportion of helicase-controlled DNA movements detected for construct 3 where strand a1 translocated through the nanopore and construct 6 where strand a1 or a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 3 or 6) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 6 only) helicase controlled DNA movements (labelled a2). This was observed to be approximately 5:95 (a2:a1 helicase controlled DNA movements) when the combination of a cholesterol and a palmitate were used to couple construct 3 to the membrane and two palmitates were used to couple construct 6 to the membrane. The bias towards a1 helicase controlled DNA movements in comparison to the control experiment shown in Figure 5 illustrated that by using the combination of two different coupling agents of differing strength it was possible to select for this construct over the construct which contained two coupling agents of similar strength.

Figure 7 shows in sections (A) and (B) cartoon representations of the DNA constructs used in Example 4 – four iSpC3 spacers are shown as a cross and four 5-nitroindoles as a black 0, the cholesterol coupling agent as a grey oval and the alternative coupling agents investigated in this example as a black triangle; labels a1-10(a-c) are described in full in Example 1 and 4.

Figure 8 shows the proportion of helicase-controlled DNA movements detected for construct 1 (coupled to the membrane using a single cholesterol) where strand a1 translocated through the nanopore and construct 2 (also coupled to the membrane using a single cholesterol) where strand a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they

were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 1) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 2) helicase controlled DNA movements (labelled a2). This was observed to be approximately 50:50 when a single cholesterol was used to couple each construct to the membrane. This control showed that if the same coupling agent was used to couple different constructs to the membrane, the number of helicase controlled DNA movements would have been approximately equal for each construct e.g. no bias was observed.

Figure 9 shows the proportion of helicase-controlled DNA movements detected for construct 1 (coupled to the membrane using a single cholesterol) where strand a1 translocated through the nanopore and construct 7a (coupled to the membrane using a fragment of DNA which had two cholesterol TEG's attached, each shown as a black triangle) where strand a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 1) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 7a) helicase controlled DNA movements (labelled a2). This was observed to be approximately 50:50 when a single cholesterol was used to couple construct 1 to the membrane in comparison to when two cholesterols were used to couple construct 7a to the membrane. This experiment showed that when two cholesterols attached to the same fragment of DNA were used to couple construct 7a to the membrane, in comparison to a single cholesterol, the number of helicase-controlled DNA movements detected showed no bias towards either construct. Therefore, the strength of coupling observed for two cholesterols in the same DNA fragment was similar to a single cholesterol.

Figure 10 shows the proportion of helicase-controlled DNA movements detected for construct 1 (coupled to the membrane using a single cholesterol) where strand a1 translocated through the nanopore and construct 7b (coupled to the membrane using a tocopherol, shown as a black triangle) where strand a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 1) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 7b) helicase controlled DNA movements (labelled a2). This was observed to be approximately 35:65 when a single cholesterol was used to couple construct

1 to the membrane in comparison to the single tocopherol used to couple construct 7b to the membrane. This experiment showed that when tocopherol was used to couple construct 7b to the membrane in comparison to a single cholesterol the number of helicase-controlled DNA movements detected showed a bias towards the construct coupled using tocopherol. Therefore, the strength of coupling observed for tocopherol was slightly stronger than that observed for cholesterol.

Figure 11 shows the proportion of helicase-controlled DNA movements detected for construct 1 (coupled to the membrane using a single cholesterol) where strand a1 translocated through the nanopore and construct 7c (coupled to the membrane using a palmitate, shown as a black triangle) where strand a2 translocated through the nanopore. The x-axis shows the position along the lambda DNA sequence to which that the helicase-controlled DNA movements mapped (they were all concentrated in the same region as strands a1 and a2 corresponded to the template and template complement of the region of lambda between 45,042 bp and 48,487 bp). The Y-axis shows the proportion of strand a1 (construct 1) helicase-controlled DNA movements (labelled a1) to strand a2 (construct 7c) helicase controlled DNA movements (labelled a2). This was observed to be approximately 70:30 when a single cholesterol was used to couple construct 1 to the membrane in comparison to the single palmitate used to couple construct 7c to the membrane. This experiment showed that when palmitate was used to couple construct 7c to the membrane, in comparison to a single cholesterol, the number of helicase-controlled DNA movements detected showed a bias towards the construct coupled using cholesterol. Therefore, the strength of coupling observed for palmitate was weaker than that observed for cholesterol.

Figure 12 shows in sections (A) and (B) cartoon representations of the DNA constructs used in Example 6 – twenty-five iSpC3 spacers are represented as three black crosses, four iSp18's as a grey rectangle, a cholesterol coupling agent as a grey circle and the palmitate coupling agent as a grey square; labels a11-15 are described in full in Example 5.

Figure 13 shows the possible constructs produced (as described in Example 5) after fragmentation and adapter attachment. Construct 10 has two hairpins adapters and cannot be captured by the nanopore. Construct 8 is the desired construct which has one Y-adapter and one hairpin, strands X and Y will both translocate through the nanopore upon capture. Construct 11 has two Y-adapters and, therefore, can capture and translocate only strand X or strand Y.

Figure 14 shows the possible constructs produced (as described in Example 5) after fragmentation and adapter attachment. Construct 10 has two hairpins adapters and cannot be captured by the nanopore. Construct 9 is the desired construct which has one Y-adapter and one hairpin, strands X and Y will both translocate through the nanopore upon capture. Construct 12 has two Y-adapters and, therefore, can capture and translocate only strand X or strand Y.

Figure 15 shows in sections (A) and (B) cartoon representations of two of the DNA constructs (constructs 13 and 14) used in Example 7 – thirty iSpC3 spacers are represented as four black crosses, four iSp18's as a grey rectangle, a cholesterol coupling agent as a grey circle and the palmitate coupling agent as a grey square; labels a12, a16-21 are described in full in Example 7. Two regions of the DNA construct are labelled R1 and R2.

Figure 16 shows in sections (A) and (B) cartoon representations of two of the DNA constructs (constructs 15 and 16) used in Example 7 – thirty iSpC3 spacers are represented as four black crosses, four iSp18's as a grey rectangle, a cholesterol coupling agent as a grey circle labels a12, a16-a18, a20, a22-23 are described in full in Example 7. Two regions of the DNA construct are labelled R1 and R2.

Figure 17 shows a graph of the % of the helicase-controlled DNA movements detected which corresponded to controlled translocation of both regions R1 and R2 through the nanopore as a proportion of all helicase-controlled DNA movements observed. The y-axis label = percentage (%) and the x-axis label = DNA construct number (13-16). Each point on the graph corresponds to a separate experiment (n = 3).

Figure 18 shows a graph of the number of helicase-controlled DNA movements detected per nanopore for each of the DNA constructs. The y-axis label = helicase-controlled DNA movements per nanopore and the x-axis label = DNA construct number (13-16). Each point on the graph corresponds to a separate experiment (n = 3).

Description of the Sequence Listing

SEQ ID NO: 1 shows the codon optimised polynucleotide sequence encoding the MS-B1 mutant MspA monomer. This mutant lacks the signal sequence and includes the following mutations: D90N, D91N, D93N, D118R, D134R and E139K.

SEQ ID NO: 2 shows the amino acid sequence of the mature form of the MS-B1 mutant of the MspA monomer. This mutant lacks the signal sequence and includes the following mutations: D90N, D91N, D93N, D118R, D134R and E139K.

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

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

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

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

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

SEQ ID NO: 10 shows the codon optimised polynucleotide sequence derived from the *sbcB* gene from *E. coli*. It encodes the exonuclease I enzyme (EcoExo I) from *E. coli*.

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

SEQ ID NO: 12 shows the codon optimised polynucleotide sequence derived from the *xthA* gene from *E. coli*. It encodes the exonuclease III enzyme from *E. coli*.

SEQ ID NO: 13 shows the amino acid sequence of the exonuclease III enzyme from *E. coli*. This enzyme performs distributive digestion of 5' monophosphate nucleosides from one strand of double stranded DNA (dsDNA) in a 3' – 5' direction. Enzyme initiation on a strand requires a 5' overhang of approximately 4 nucleotides.

SEQ ID NO: 14 shows the codon optimised polynucleotide sequence derived from the *recJ* gene from *T. thermophilus*. It encodes the RecJ enzyme from *T. thermophilus* (*Tth*RecJ-cd).

SEQ ID NO: 15 shows the amino acid sequence of the RecJ enzyme from *T. thermophilus* (*Tth*RecJ-cd). This enzyme performs processive digestion of 5' monophosphate nucleosides from ssDNA in a 5' – 3' direction. Enzyme initiation on a strand requires at least 4 nucleotides.

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

SEQ ID NO: 17 shows the amino acid sequence of the bacteriophage lambda exonuclease. The sequence is one of three identical subunits that assemble into a trimer. The enzyme performs highly processive digestion of nucleotides from one strand of dsDNA, in a 5'-3'direction (http://www.neb.com/nebecomm/products/productM0262.asp). Enzyme initiation on a strand preferentially requires a 5' overhang of approximately 4 nucleotides with a 5' phosphate.

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

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

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

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

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

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

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

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

SEQ ID NO: 26 shows the polynucleotide sequence of the double stranded portion of a MuA substrate of the invention.

SEQ ID NO: 27 shows the polynucleotide sequence of the double stranded portion of a MuA substrate of the invention. This sequence is complementary to SEQ ID NO: 26 except that it contains a U at the 3' end.

SEQ ID NO: 28 shows polynucleotide sequence of the overhang strand of the double stranded MuA substrate of the invention.

SEQ ID NOs: 29-32 show polynucleotide sequences used in Example 1.

SEQ ID NO: 33-41 show polynucleotide sequences used in Example 1, 2 and 3.

SEQ ID NO: 42 shows a polynucleotide sequence used in Example 4.

SEQ ID NO: 43 shows the polynucleotide sequence of the Enterobacteria phage λ. The sequence contains an additional 12 base overhang attached at the 5' end of the template strand. The sequence shown here is that of the template strand only (the template complement is not shown). This sequence is used in Example 5.

SEQ ID NO: 44-48 shows the polynucleotide sequences used in Example 5 and 6.

SEQ ID NO: 49 shows a polynucleotide sequence used in Example 7. At the 5' end the sequence contains a phosphate group.

SEQ ID NO: 50 shows a polynucleotide sequence used in Example 7. At the 3' end the final thymine in the sequence has a phosphothioate group.

SEQ ID NO: 51 shows a polynucleotide sequence used in Example 7.

SEQ ID NO: 52 shows a polynucleotide sequence used in Example 7.

SEQ ID NO: 53 shows a polynucleotide sequence used in Example 7. At the 5' end the sequence contains a phosphate group.

SEQ ID NO: 54 shows a polynucleotide sequence used in Example 7.

Detailed description of the invention

It is to be understood that different applications of the disclosed products and methods may be tailored to the specific needs in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

In addition as used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes two or more polynucleotides, reference to "an anchor" refers to two or more anchors, reference to "a helicase" includes two or more helicases, reference to "a transmembrane pore" includes two or more pores and the like.

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

Characterisation method of the invention

When Y adaptors and hairpin loop adaptors are used together to characterise double stranded polynucleotides, the Y adaptor generally comprises the anchor for coupling the polynucleotide to the membrane and a leader sequence which preferentially threads into the pore. The anchor and more importantly the leader typically result in a bias towards the characterisation of double stranded polynucleotides having Y adaptors at both ends.

When a population of double stranded polynucleotides are modified using approximately equal amounts of Y adaptors and hairpin loop adaptors, three distinct groups of modified polynucleotides are produced. Approximately, 50% of the modified polynucleotides have a Y adaptor at one end and a hairpin loop adaptor at the other end (*i.e.* are "round the corner" or RTC polynucleotides). It is preferred that these RTC polynucleotides are characterised using the method disclosed in WO 2013/014451. Approximately, 25% of the modified polynucleotides have a Y adaptor at each end (*i.e.* are double Y polynucleotides) and approximately 25% of the modified polynucleotides have a hairpin loop adaptor at each end (*i.e.* are double hairpin polynucleotides). If the Y adaptor comprises the anchor, there is a bias towards characterisation of the double Y polynucleotides because they have two anchors and therefore couple to the membrane more strongly. If the Y adaptor comprises the leader, there is a bias towards characterisation of the double Y polynucleotides because they have two leaders and therefore preferentially thread into the pore. The presence of the leader generally results in a greater bias than the presence of the anchor. If the Y adaptor comprises both the anchor and the leader, there is the greatest bias towards characterisation of the double Y polynucleotides. The inventors have seen biases of approximately 50-fold towards double Y polynucleotides. The double hairpin polynucleotides are not typically characterised because they do not have a single stranded region capable of threading into the nanopore used for characterisation.

The inventors have shown that using one or more anchors on the hairpin loop adaptor not only reduces characterisation of the double Y polynucleotides, but also reduces the observed throughput of the double Y polynucleotides significantly. The inventors conclude from this that the throughput is dependent on the presence of an anchor at the end of the modified polynucleotide which threads into the nanopore (i.e. the Y adaptor end of a RTC polynucleotide).

The inventors have also surprisingly shown that the bias towards double Y polynucleotides can be overcome by using a Y adaptor and a hairpin loop adaptor which have different strength of coupling to the membrane. If the hairpin loop adaptor couples with more strength than the Y adaptor, a bias towards the RTC polynucleotides is achieved. The "strong" hairpin loop adaptor effectively couples the RTC polynucleotides to the membrane and this

allows them to out compete the double Y adaptors which contain two "weaker" tethers. High throughput is maintained for the RTC polynucleotides because of the "weaker" Y adaptor.

The invention concerns characterising a target double stranded polynucleotide using a transmembrane pore in a membrane. The target double stranded polynucleotide is provided with a Y adaptor at one end and a hairpin loop adaptor at the other end. Methods of doing this are discussed in detail below. The Y adaptor comprises one or more first anchors for coupling the polynucleotide to the membrane and the hairpin loop adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane. The strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane.

The polynucleotide provided with the two adaptors is contacted with the transmembrane pore such that at least one strand of the polynucleotide moves through the pore. One or more measurements are taken as the at least one strand of the polynucleotide moves with respect to the pore. The measurements are indicative of one or more characteristics of the at least one strand of the polynucleotide.

The invention preferably concerns sequencing or estimating the sequence of a target double stranded polynucleotide. Only part of the target double stranded polynucleotide may be characterised or sequenced as discussed in more detail below.

Target double stranded polynucleotide

The method of the invention characterises a target double stranded polynucleotide. The target double stranded polynucleotide may also be called the template double stranded polynucleotide or the double stranded polynucleotide of interest.

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

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

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

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

The target double stranded 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 is preferably a deoxyribonucleotide. The nucleotide typically contains a monophosphate, diphosphate or triphosphate. 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) and deoxycytidine monophosphate (dCMP). The nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP, dCMP and dUMP. The nucleotides are most preferably selected from dAMP, dTMP, dGMP, dCMP and dUMP.

The target double stranded polynucleotide preferably comprises the following nucleotides: dAMP, dUMP and/or dTMP, dGMP and dCMP.

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

The nucleotides in the target 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 target polynucleotide is double stranded. The target polynucleotide may contain some single stranded regions, but at least a portion of the target polynucleotide is double stranded.

The target polynucleotide can be a nucleic acid, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The target polynucleotide can comprise one strand of RNA hybridised to one strand of DNA. The polynucleotide may be any synthetic nucleic acid known in the art, such as peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA),

locked nucleic acid (LNA), bridged nucleic acid (BNA) 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. Bridged nucleic acids (BNAs) are modified RNA nucleotides. They may also be called constrained or inaccessible RNA. BNA monomers can contain a five-membered, six-membered or even a seven-membered bridged structure with a "fixed" C3'-endo sugar puckering. The bridge is synthetically incorporated at the 2', 4'-position of the ribose to produce a 2', 4'-BNA monomer.

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

The target 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 nucleotide pairs in length. The polynucleotide can be 1000 or more nucleotide pairs, 5000 or more nucleotide pairs in length or 100000 or more nucleotide pairs in length.

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

The sample may be a biological sample. The invention may be carried out *in vitro* on a sample obtained from or extracted from any organism or microorganism. The organism or microorganism is typically archaeal, prokaryotic or eukaryotic and typically belongs to one of the five kingdoms: plantae, animalia, fungi, monera and protista. The invention may be carried out *in vitro* on a sample obtained from or extracted from any virus. The sample is preferably a fluid sample. The sample typically comprises a body fluid of the patient. The sample may be urine, lymph, saliva, mucus or amniotic fluid but is preferably blood, plasma or serum. Typically, the sample is human in origin, but alternatively it may be from another mammal animal such as from commercially farmed animals such as horses, cattle, sheep, fish, chickens or pigs or may alternatively be pets such as cats or dogs. Alternatively, the sample may be of plant origin, such as a sample obtained from a commercial crop, such as a cereal, legume, fruit or vegetable, for example wheat, barley, oats, canola, maize, soya, rice, rhubarb, bananas, apples, tomatoes, potatoes, grapes, tobacco, beans, lentils, sugar cane, cocoa, cotton.

The sample may be a non-biological sample. The non-biological sample is preferably a fluid sample. Examples of a non-biological sample include surgical fluids, water such as drinking water, sea water or river water, and reagents for laboratory tests.

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

Providing the polynucleotide with two adaptors

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

The Y adaptor typically comprises a double stranded region and a single stranded region or a region that is not complementary at the other end. The Y adaptor may be described as having an overhang if it comprises a single stranded region. The presence of a non-complementary region in the Y adaptor gives the adaptor its Y shape since the two strands typically do not hybridise to each other unlike the double stranded portion. The Y adaptor comprises the one or more first anchors. Anchors are discussed in more detail below.

The Y adaptor preferably comprises a leader sequence which preferentially threads into the pore. The leader sequence facilitates the method of the invention. The leader sequence is designed to preferentially thread into the transmembrane pore and thereby facilitate the movement of polynucleotide through the pore. The leader sequence can also be used to link the polynucleotide to the anchor(s) as discussed below.

The leader sequence typically comprises a polymer. The polymer is preferably negatively charged. The polymer is preferably a polynucleotide, such as DNA or RNA, a modified polynucleotide (such as abasic DNA), PNA, LNA, polyethylene glycol (PEG) or a polypeptide. The leader preferably comprises a polynucleotide and more preferably comprises a single stranded polynucleotide. The leader sequence can comprise any of the polynucleotides discussed above. The single stranded leader sequence most preferably comprises a single strand of DNA, such as a poly dT section. The leader sequence preferably comprises the one or more spacers.

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

Suitable hairpin loop adaptors can be designed using methods known in the art. The hairpin loop may be any length. The hairpin loop is typically 110 or fewer nucleotides, such as 100 or fewer nucleotides, 90 or fewer nucleotides, 80 or fewer nucleotides, 70 or fewer nucleotides, 60 or fewer nucleotides, 50 or fewer nucleotides, 40 or fewer nucleotides, 30 or fewer nucleotides, 20 or fewer nucleotides or 10 or fewer nucleotides, in length. The hairpin loop is preferably from about 1 to 110, from 2 to 100, from 5 to 80 or from 6 to 50 nucleotides in length. Longer lengths of the hairpin loop, such as from 50 to 110 nucleotides, are preferred if the loop is involved in the differential selectability of the adaptor. Similarly, shorter lengths of the hairpin loop, such as from 1 to 5 nucleotides, are preferred if the loop is not involved in the selectable binding as discussed below.

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

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

The hairpin loop adaptor and/or the selectable binding moiety may comprise a region that can be cut, nicked, cleaved or hydrolysed. Such a region can be designed to allow the first and/or second polynucleotide to be removed from the surface to which it is bound following purification or isolation. Suitable regions are known in the art. Suitable regions include, but are not limited to, an RNA region, a region comprising desthiobiotin and streptavidin, a disulphide bond and a photocleavable region.

The Y adaptor and/or the hairpin loop adaptor may be ligated to the polynucleotide using any method known in the art. One or both of the adaptors may be ligated using a ligase, such as T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase and 9°N DNA ligase.

Alternatively, the adaptors may be added to the polynucleotide using the methods of the invention discussed below.

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

Membrane

Any membrane may be used in accordance with the invention. Suitable membranes are well-known in the art. The membrane is preferably an amphiphilic layer. An amphiphilic layer is a layer formed from amphiphilic molecules, such as phospholipids, which have both hydrophilic and lipophilic properties. The amphiphilic molecules may be synthetic or naturally occurring. Non-naturally occurring amphiphiles and amphiphiles which form a monolayer are known in the art and include, for example, block copolymers (Gonzalez-Perez et al., Langmuir, 2009, 25, 10447-10450). Block copolymers are polymeric materials in which two or more monomer sub-units that are polymerized together to create a single polymer chain. Block copolymers typically have properties that are contributed by each monomer sub-unit. However, a block copolymer may have unique properties that polymers formed from the individual sub-units do not possess. Block copolymers can be engineered such that one of the monomer sub-units is hydrophobic (i.e. lipophilic), whilst the other sub-unit(s) are hydrophilic whilst in aqueous media. In this case, the block copolymer may possess amphiphilic properties and may form a structure that mimics a biological membrane. The block copolymer may be a diblock (consisting of two monomer sub-units), but may also be constructed from more than two monomer sub-units to form more complex arrangements that behave as amphipiles. The copolymer may be a triblock, tetrablock or pentablock copolymer. The membrane is preferably a triblock copolymer membrane.

Archaebacterial bipolar tetraether lipids are naturally occurring lipids that are constructed such that the lipid forms a monolayer membrane. These lipids are generally found in extremophiles that survive in harsh biological environments, thermophiles, halophiles and acidophiles. Their stability is believed to derive from the fused nature of the final bilayer. It is straightforward to construct block copolymer materials that mimic these biological entities by creating a triblock polymer that has the general motif hydrophilic-hydrophobic-hydrophilic. This material may form monomeric membranes that behave similarly to lipid bilayers and

encompass a range of phase behaviours from vesicles through to laminar membranes. Membranes formed from these triblock copolymers hold several advantages over biological lipid membranes. Because the triblock copolymer is synthesized, the exact construction can be carefully controlled to provide the correct chain lengths and properties required to form membranes and to interact with pores and other proteins.

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

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

In a preferred embodiment, the invention provides a method of characterising a target double stranded polynucleotide using a transmembrane pore in a membrane comprising a triblock copolymer, optionally wherein the membrane is modified to facilitate the coupling.

The membrane is most preferably one of the membranes disclosed in International Application No. PCT/GB2013/052766 or PCT/GB2013/052767.

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

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

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

The membrane may be a lipid bilayer. Lipid bilayers are models of cell membranes and serve as excellent platforms for a range of experimental studies. For example, lipid bilayers can be used for *in vitro* investigation of membrane proteins by single-channel recording. Alternatively, lipid bilayers can be used as biosensors to detect the presence of a range of substances. The lipid bilayer may be any lipid bilayer. Suitable lipid bilayers include, but are not limited to, a planar lipid bilayer, a supported bilayer or a liposome. The lipid bilayer is preferably a planar lipid bilayer. Suitable lipid bilayers are disclosed in International

Application No. PCT/GB08/000563 (published as WO 2008/102121), International Application No. PCT/GB08/004127 (published as WO 2009/077734) and International Application No. PCT/GB2006/001057 (published as WO 2006/100484).

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

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

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

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

Patch-clamping is commonly used in the study of biological cell membranes. The cell membrane is clamped to the end of a pipette by suction and a patch of the membrane becomes attached over the aperture. The method has been adapted for producing lipid bilayers by clamping liposomes which then burst to leave a lipid bilayer sealing over the aperture of the pipette. The method requires stable, giant and unilamellar liposomes and the fabrication of small apertures in materials having a glass surface.

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

In a preferred embodiment, the lipid bilayer is formed as described in International Application No. PCT/GB08/004127 (published as WO 2009/077734). Advantageously in this method, the lipid bilayer is formed from dried lipids. In a most preferred embodiment, the lipid bilayer is formed across an opening as described in WO2009/077734 (PCT/GB08/004127).

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

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

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

interfacial moiety as an ether or an ester. The lipids may be mycolic acid.

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

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

In another preferred embodiment, the membrane is a solid state layer. Solid state layers can be formed from both organic and inorganic materials including, but not limited to, microelectronic materials, insulating materials such as Si_3N_4, Al_2O_3, and SiO, organic and inorganic polymers such as polyamide, plastics such as Teflon® or elastomers such as two-component addition-cure silicone rubber, and glasses. The solid state layer may be formed from graphene. Suitable graphene layers are disclosed in International Application No. PCT/US2008/010637 (published as WO 2009/035647).

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

Coupling

The polynucleotide may be coupled to the membrane using any known method. The polynucleotide is coupled to the membrane using at least two anchors.

The Y adaptor comprises one or more first anchors. The hairpin loop adaptor comprises one or more second adaptors. Each anchor comprises a group which couples (or binds) to the adaptor and a group which couples (or binds) to the membrane. Each anchor may covalently couple (or bind) to the adaptor and/or the membrane.

The Y adaptor may contain any number of first anchors, such as 2, 3, 4 or more anchors. The hairpin loop adaptor may contain any number of second anchors, such as 2, 3, 4 or more anchors. For instance, one or both adaptors may comprise two anchors each of which separately couples (or binds) to both the adaptor(s) and membrane.

The one or more first anchors and/or the one or more second anchors may comprise one or more polynucleotide binding proteins. Each anchor may comprise one or more polynucleotide binding proteins. The polynucleotide binding protein(s) may be any of those discussed above.

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

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

The Y adaptor and/or the hairpin loop adaptor may be coupled directly to the membrane. The one or more first anchors and/or the one or more second anchors preferably comprise a linker. The one or more first anchors and/or the one or more second anchors may comprise one or more, such as 2, 3, 4 or more, linkers. One linker may be used to couple more than one, such as 2, 3, 4 or more, polynucleotides or adaptors to the membrane.

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

polynucleotide to be characterised may hybridise to a complementary sequence on the circular polynucleotide linker.

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

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

Crosslinkage of polynucleotides can be avoided using a "lock and key" arrangement. Only one end of each linker may react together to form a longer linker and the other ends of the linker each react with the polynucleotide or membrane respectively. Such linkers are described in International Application No. PCT/GB10/000132 (published as WO 2010/086602).

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

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

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

Chemical groups that form permanent/stable or transient links are discussed in more detail below. The polynucleotide may be transiently coupled to an amphiphilic layer or triblock copolymer membrane using cholesterol or a fatty acyl chain. Any fatty acyl chain having a length of from 6 to 30 carbon atom, such as hexadecanoic acid, may be used.

In preferred embodiments, a polynucleotide, such as a nucleic acid, is coupled to an amphiphilic layer such as a triblock copolymer membrane or lipid bilayer. Coupling of nucleic acids to synthetic lipid bilayers has been carried out previously with various different tethering strategies. These are summarised in Table 1 below.

Table 1

Anchor comprising	**Type of coupling**	**Reference**
Thiol	Stable	Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." J Am Chem Soc

		125(13): 3696-7.
Biotin	Stable	Nikolov, V., R. Lipowsky, et al. (2007). "Behavior of giant vesicles with anchored DNA molecules." Biophys J **92**(12): 4356-68
Cholesterol	Transient	Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucletides to lipid membrane assemblies." J Am Chem Soc **126**(33): 10224-5
Surfactant (e.g. Lipid, Palmitate, etc)	Stable	van Lengerich, B., R. J. Rawle, et al. "Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." Langmuir **26**(11): 8666-72

Synthetic polynucleotides and/or linkers may be functionalised using a modified phosphoramidite in the synthesis reaction, which is easily compatible for the direct addition of suitable anchoring groups, such as cholesterol, tocopherol, palmitate, thiol, lipid and biotin groups. These different attachment chemistries give a suite of options for attachment to polynucleotides. Each different modification group couples the polynucleotide in a slightly different way and coupling is not always permanent so giving different dwell times for the polynucleotide to the membrane. The advantages of transient coupling are discussed above.

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

The one or more first anchors and/or the one or more second anchors preferably couple the target double stranded polynucleotide to the membrane via hybridisation. The hybridisation may be present in any part of the one or more anchors, such as between the one or more anchors and the polynucleotide, within the one or more anchors or between the one or more anchors and the membrane. Hybridisation in the one or more anchors allows coupling in a transient manner as discussed above. For instance, a linker may comprise two or more polynucleotides, such as 3, 4 or 5 polynucleotides, hybridised together. The one or more first anchors and/or the one or more second anchors may hybridise to the target double stranded polynucleotide. The one or more first anchors may hybridise directly to the target double stranded polynucleotide or directly to the Y adaptor and/or leader sequence. The one or more second anchors may hybridise directly to the target double stranded polynucleotide or directly to the hairpin adaptor. Alternatively, the one or more first anchors may be hybridised to one or more, such as 2 or 3, intermediate polynucleotides (or "splints") which are hybridised to the target double stranded polynucleotide or to the Y adaptor and/or leader sequence. Alternatively, the one or more second anchors may be hybridised to one or more, such as 2 or 3, intermediate polynucleotides (or "splints") which are hybridised to the target double stranded polynucleotide or to the hairpin loop adaptor. The hybridisation of the one or more anchors to one or more splints may form one or more rigid double stranded polynucleotide linkers as discussed in more detail below.

The one or more first anchors and/or the one or more second anchors may comprise a single stranded or double stranded polynucleotide. One part of the anchor may be ligated to the double stranded polynucleotide. Ligation of short pieces of ssDNA have been reported using T4 RNA ligase I (Troutt, A. B., M. G. McHeyzer-Williams, et al. (1992). "Ligation-anchored PCR: a simple amplification technique with single-sided specificity." Proc Natl Acad Sci U S A 89(20): 9823-5). Alternatively, either a single stranded or double stranded polynucleotide can be ligated to a double stranded polynucleotide and then the two strands separated by thermal or chemical denaturation. To a double stranded polynucleotide, it is possible to add either a piece of single stranded polynucleotide to one or both of the ends of the duplex, or a double stranded polynucleotide to one or both ends. For addition of single stranded polynucleotides to the a double stranded polynucleotide, this can be achieved using T4 RNA ligase I as for ligation to other regions of single stranded polynucleotides. For addition of double stranded polynucleotides to a double stranded polynucleotide then ligation can be "blunt-ended", with complementary 3' dA / dT tails on the target and added polynucleotides respectively (as is routinely done for many sample prep applications to prevent concatemer or dimer formation) or using "sticky-ends" generated by restriction digestion of the polynucleotide and ligation of compatible adapters. Then, when the duplex is melted, each single strand will have either a 5' or

3' modification if a single stranded polynucleotide was used for ligation or a modification at the 5' end, the 3' end or both if a double stranded polynucleotide was used for ligation.

If the polynucleotide is a synthetic strand, the adaptors and anchors can be incorporated during the chemical synthesis of the polynucleotide. For instance, the polynucleotide can be synthesised using a primer having a reactive group attached to it.

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

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

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

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

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

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

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

triplex. The one or more anchors may couple to (or bind to) a double stranded polynucleotide by forming a triplex with the double stranded duplex.

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

Where the one or more first anchors and/or the one or more second anchors comprise a protein, it may be able to anchor directly into the membrane without further functonalisation, for example if it already has an external hydrophobic region which is compatible with the membrane. Examples of such proteins include, but are not limited to, transmembrane proteins, intramembrane proteins and membrane proteins. Alternatively the protein may be expressed with a genetically fused hydrophobic region which is compatible with the membrane. Such hydrophobic protein regions are known in the art.

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

An example of a molecule used in chemical attachment is EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride). Reactive groups can also be added to the 5' of polynucleotides using commercially available kits (Thermo Pierce, Part No. 22980). Suitable methods include, but are not limited to, transient affinity attachment using histidine residues and Ni-NTA, as well as more robust covalent attachment by reactive cysteines, lysines or non natural amino acids.

Strength of coupling

The strength of coupling (or binding) of the hairpin loop adaptor to the membrane is greater than the strength of coupling (or binding) of the Y adaptor to the membrane. This can be measured in any way. A suitable method for measuring the strength of coupling (or binding) is disclosed in the Examples.

The strength of coupling (or binding) of the hairpin loop adaptor is preferably at least 1.5 times the strength of coupling (or binding) of the Y-adaptor, such as at least twice, at least three times, at least four times, at least five or at least ten times the strength of coupling (or binding) of the Y-adaptor. The affinity constant (Kd) of the hairpin loop adaptor for the membrane is preferably at least 1.5 times the affinity constant of the Y adaptor, such as at least twice, at least three times, at least four times, at least five or at least ten times the strength of coupling of the Y adaptor.

There are several ways in which the hairpin loop adaptor couples (or binds) more strongly to the membrane than the Y adaptor. For instance, the hairpin loop adaptor may comprise more anchors than the Y adaptor. For instance, the hairpin loop adaptor may comprise 2, 3 or more second anchors whereas the Y adaptor may comprise one first anchor.

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

The one or more second anchors preferably comprise one or more groups which couples(s) (or bind(s)) to the membrane with a greater strength than the one or more groups in the one or more first anchors which couple(s) (or bind(s)) to the membrane. In preferred embodiments, the hairpin loop adaptor/one or more second anchors couple (or bind) to the membrane using cholesterol and the Y adaptor/one or more first anchors couple (or bind) to the membrane using palmitate. Cholesterol binds to triblock copolymer membranes and lipid membranes more strongly than palmitate. In an alternative embodiment, the hairpin loop adaptor/one or more second anchors couple (or bind) to the membrane using a mono-acyl species, such as palmitate, and the Y adaptor/one or more first anchors couple (or bind) to the membrane using a diacyl species, such as dipalmitoylphosphatidylcholine. Other factors which may affect the ability of the one or more groups to couple (or bind) to the membrane include, but are not limited to, the charge or hydrophobicity of the one or more anchors or one or more linkers or the entropic cost. For instance, the one or more first and second anchors may comprise cholesterol, but the charge of the one or more second anchors may increase the ability of their cholesterol to couple (or bind) to the membrane or the charge of the one or more first anchors may decrease the ability of their cholesterol to couple (or bind) to the membrane.

The one or more first anchors may comprise one or more rigid linkers and the one or more second anchors may comprise one or more flexible linkers. Flexible linkers may allow the one or more groups which couple (or bind) to the membrane to move relative to the target double stranded polynucleotide and thereby increase the likelihood of coupling. In this embodiment, the one or more groups which couple (or bind) to the membrane may be the same in the first and second anchors. For instance, the one or more first anchors may comprise one or more rigid

linkers and cholesterol and the one or more second anchors may comprise one or more flexible linkers and cholesterol. Suitable linkers are discussed above. Rigid linkers are those which do not substantially bend or flex and do not allow substantial rotation either along their length or at their attachments points. Rigid linkers are those which do not permit substantial movement of the one or more groups which couple (or bind) to the membrane relative to the target double stranded polynucleotide. In the case where a rigid linker comprises a polymeric or oligomeric section, the polymeric or oligomeric section is shorter than the persistence length, a measure of rigidity, of the corresponding polymer or oligomer. Persistence length can be measured or calculated using methods known in the art. For instance, the persistence length of double-stranded DNA is approximately 50nm or 500 Angstroms or 147 nucleotide pairs. A stretch of double-stranded DNA of DNA less than 147 nucleotide pairs in length is considered to be rigid. Rigid linkers include, but are not limited to, those comprising double stranded polynucleotides, including DNA, conjugated organic moieties or organic moieties with bulky side chains that restrict conformational freedom. The one or more rigid linkers preferably comprise one or more double stranded polynucleotides. Any of the polynucleotides discussed above may be used. The preferred length of the one or more rigid linkers is 5, 10, 15, 20, 25, 27, 30, 35, 40 or more nucleotide pairs.

Flexible linkers are those which substantially bend or flex or rotate. Flexible linkers may bend or flex along their length or at one or both of their attachment points. Flexible linkers permit substantial movement of the one or more groups which couple (or bind) to the membrane relative to the target double stranded polynucleotide. Flexible linkers are those which permit substantial variation of the orientation of the target double stranded polynucleotide. Flexible linkers may allow movement of the one or more groups by flexing in one or two dimensions. Alternatively or in addition, flexible linkers may allow movement of the one or more groups that is a rotation about an axis. In the case where a flexible linker comprises a polymeric or oligomeric section, the polymeric or oligomeric section contributes significant flexibility if it is similar to or longer than the persistence length of the corresponding polymer or oligomer. Flexible linkers include, but are not limited to, those comprising a single-stranded oligonucleotide or polynucleotide, a short carbon spacer (e.g. alkanes and alkenes), a polypeptide, a polyhistidine tag, nucleoside derivatives, a short polysaccharide or combinations thereof. Flexible linkers typically also allow rotation of the one or more groups which couple (or bind) to the membrane. Flexible linkers may permit rotation in at least one axis. The axis is preferably the longitudinal axis of the linker. The one or more flexible linkers preferably comprise one or more spacer 9 (iSp9) groups or one or more spacer 18 (iSp18) groups. The one

or more flexible linkers preferably comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more spacer 9 (iSp9) groups and/or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more spacer 18 (iSp18) groups.

Preferably (a) the hairpin loop adaptor comprises more anchors than the Y adaptor, (b) the strength of coupling of one or more second anchors to the membrane is greater than the strength of coupling of the one or more first anchors to the membrane, (c) the strength of coupling of the one or more second anchors to the hairpin loop adaptor is greater than the strength of coupling of the one or more first anchors to the Y adaptor, (d) the one or more first anchors and one or more second anchors couple their respective adaptors to the membrane via hybridisation and the strength of hybridisation is greater in the one or more second anchors than in the one or more first anchors, (e) the one or more first anchors comprise one or more rigid linkers and the one or more second anchors comprise one or more flexible linkers or (f) any combination of (a) to (e). In other words, the invention preferably comprises a combination of (a) and (b), a combination of (b) and (c), a combination of (c) and (d), a combination (a) and (c), a combination of (b) and (d), a combination of (a) and (d), a combination of (a) and (e), a combination of (b) and (e), a combination of (c) and (e), a combination of (d) and (e), a combination of (a), (b) and (c), a combination of (b), (c) and (d), a combination of (a), (b) and (d), a combination of (a), (c) and (d), a combination of (a), (b) and (e), a combination of (a), (c) and (e), a combination of (a), (d) and (e), a combination of (b), (c) and (e), a combination of (b), (d) and (e), a combination of (c), (d) and (e), a combination of (a), (b), (c) and (d), a combination of (a), (b), (c) and (e), a combination of (a), (b), (d) and (e), a combination of (a), (c), (d) and (e), a combination of (b), (c), (d) and (e) or a combination of (a), (b), (c), (d) and (e).

Transmembrane pore

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

Electrical measurements may be made using standard single channel recording equipment as describe in Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7, Lieberman KR et al, J Am Chem Soc. 2010;132(50):17961-72, and International Application

WO 2000/28312. Alternatively, electrical measurements may be made using a multi-channel system, for example as described in International Application WO 2009/077734 and International Application WO 2011/067559.

The method is preferably carried out with a potential applied across the membrane. The applied potential may be a voltage potential. Alternatively, the applied potential may be a chemical potential. An example of this is using a salt gradient across a membrane, such as an amphiphilic layer. A salt gradient is disclosed in Holden *et al.*, J Am Chem Soc. 2007 Jul 11; 129(27):8650-5. In some instances, the current passing through the detector (or pore) as a polynucleotide moves with respect to the pore is used to estimate or determine the sequence of the polynucleotide. This is strand sequencing.

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

Step (c) preferably comprises taking one or more measurements of the current passing through the pore as the at least one strand moves with respect to the pore wherein the one or more current measurements are indicative of one or more characteristics of the at least one strand.

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

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

The transmembrane protein pore may be a monomer or an oligomer. The pore is preferably made up of several repeating subunits, such as at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13 or at least 14 subunits, such as 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 subunits. The pore is preferably a hexameric, heptameric, octameric or nonameric pore. The pore may be a homo-oligomer or a hetero-oligomer.

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

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

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

The transmembrane protein pore is preferably derived from Msp, preferably from MspA. Such a pore will be oligomeric and typically comprises 7, 8, 9 or 10 monomers derived from Msp. The pore may be a homo-oligomeric pore derived from Msp comprising identical monomers. Alternatively, the pore may be a hetero-oligomeric pore derived from Msp

comprising at least one monomer that differs from the others. Preferably the pore is derived from MspA or a homolog or paralog thereof.

A monomer derived from Msp typically comprises the sequence shown in SEQ ID NO: 2 or a variant thereof. SEQ ID NO: 2 is the MS-(B1)8 mutant of the MspA monomer. It includes the following mutations: D90N, D91N, D93N, D118R, D134R and E139K. A variant of SEQ ID NO: 2 is a polypeptide that has an amino acid sequence which varies from that of SEQ ID NO: 2 and which retains its ability to form a pore. The ability of a variant to form a pore can be assayed using any method known in the art. For instance, the variant may be inserted into an amphiphilic layer along with other appropriate subunits and its ability to oligomerise to form a pore may be determined. Methods are known in the art for inserting subunits into membranes, such as amphiphilic layers. For example, subunits may be suspended in a purified form in a solution containing a triblock copolymer membrane such that it diffuses to the membrane and is inserted by binding to the membrane and assembling into a functional state. Alternatively, subunits may be directly inserted into the membrane using the "pick and place" method described in M.A. Holden, H. Bayley. J. Am. Chem. Soc. 2005, 127, 6502-6503 and International Application No. PCT/GB2006/001057 (published as WO 2006/100484).

Over the entire length of the amino acid sequence of SEQ ID NO: 2, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of SEQ ID NO: 2 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 100 or more, for example 125, 150, 175 or 200 or more, contiguous amino acids ("hard homology").

Standard methods in the art may be used to determine homology. For example the UWGCG Package provides the BESTFIT program which can be used to calculate homology, for example used on its default settings (Devereux *et al* (1984) *Nucleic Acids Research* **12**, p387-395). The PILEUP and BLAST algorithms can be used to calculate homology or line up sequences (such as identifying equivalent residues or corresponding sequences (typically on their default settings)), for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S.F *et al* (1990) J Mol Biol 215:403-10. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).

SEQ ID NO: 2 is the MS-(B1)8 mutant of the MspA monomer. The variant may comprise any of the mutations in the MspB, C or D monomers compared with MspA. The

mature forms of MspB, C and D are shown in SEQ ID NOs: 5 to 7. In particular, the variant may comprise the following substitution present in MspB: A138P. The variant may comprise one or more of the following substitutions present in MspC: A96G, N102E and A138P. The variant may comprise one or more of the following mutations present in MspD: Deletion of G1, L2V, E5Q, L8V, D13G, W21A, D22E, K47T, I49H, I68V, D91G, A96Q, N102D, S103T, V104I, S136K and G141A. The variant may comprise combinations of one or more of the mutations and substitutions from Msp B, C and D. The variant preferably comprises the mutation L88N. A variant of SEQ ID NO: 2 has the mutation L88N in addition to all the mutations of MS-B1 and is called MS-(B2)8. The pore used in the invention is preferably MS-(B2)8. A variant of SEQ ID NO: 2 has the mutations G75S/G77S/L88N/Q126R in addition to all the mutations of MS-B1 and is called MS-B2C. The pore used in the invention is preferably MS-(B2)8 or MS-(B2C)8.

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

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

Variants may include fragments of SEQ ID NO: 2. Such fragments retain pore forming activity. Fragments may be at least 50, 100, 150 or 200 amino acids in length. Such fragments may be used to produce the pores. A fragment preferably comprises the pore forming domain of SEQ ID NO: 2. Fragments must include one of residues 88, 90, 91, 105, 118 and 134 of SEQ ID NO: 2. Typically, fragments include all of residues 88, 90, 91, 105, 118 and 134 of SEQ ID NO: 2.

One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the amino terminal or carboxy terminal of the amino acid sequence of SEQ ID NO: 2 or polypeptide variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids. A carrier protein may be

fused to an amino acid sequence according to the invention. Other fusion proteins are discussed in more detail below.

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

The monomer derived from Msp contains one or more specific modifications to facilitate nucleotide discrimination. The monomer derived from Msp may also contain other non-specific modifications as long as they do not interfere with pore formation. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the monomer derived from Msp. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with $NaBH_4$, amidination with methylacetimidate or acylation with acetic anhydride.

The monomer derived from Msp can be produced using standard methods known in the art. The monomer derived from Msp may be made synthetically or by recombinant means. For example, the pore may be synthesized by *in vitro* translation and transcription (IVTT). Suitable methods for producing pores are discussed in International Application Nos. PCT/GB09/001690 (published as WO 2010/004273), PCT/GB09/001679 (published as WO 2010/004265) or PCT/GB10/000133 (published as WO 2010/086603). Methods for inserting pores into membranes are discussed.

The transmembrane protein pore is also preferably derived from α-hemolysin (α-HL). The wild type α-HL pore is formed of seven identical monomers or subunits (i.e. it is heptameric). The sequence of one monomer or subunit of α-hemolysin-NN is shown in SEQ ID NO: 4.

In some embodiments, the transmembrane protein pore is chemically modified. The pore can be chemically modified in any way and at any site. The transmembrane protein pore is preferably chemically modified by attachment of a molecule to one or more cysteines (cysteine linkage), attachment of a molecule to one or more lysines, attachment of a molecule to one or more non-natural amino acids, enzyme modification of an epitope or modification of a terminus.

Suitable methods for carrying out such modifications are well-known in the art. The transmembrane protein pore may be chemically modified by the attachment of any molecule. For instance, the pore may be chemically modified by attachment of a dye or a fluorophore.

Any number of the monomers in the pore may be chemically modified. One or more, such as 2, 3, 4, 5, 6, 7, 8, 9 or 10, of the monomers is preferably chemically modified as discussed above.

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

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

Any of the proteins described herein, such as the transmembrane protein pores, can be produced using standard methods known in the art. Polynucleotide sequences encoding a pore or construct may be derived and replicated using standard methods in the art. Polynucleotide sequences encoding a pore or construct may be expressed in a bacterial host cell using standard techniques in the art. The pore 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.

The pore 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.

Polynucleotide characterisation

The method of the invention involves measuring one or more characteristics of the target double stranded polynucleotide.

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

The polynucleotides can be naturally occurring or artificial. For instance, the method may be used to verify the sequence of two or more manufactured oligonucleotides. The methods are typically carried out *in vitro*.

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

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

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

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

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

For (v), the presence or absence of any modification may be measured. The method preferably comprises determining whether or not the polynucleotide is modified by methylation, by oxidation, by damage, with one or more proteins or with one or more labels, tags or spacers.

Specific modifications will result in specific interactions with the pore which can be measured using the methods described below. For instance, methylcytosine may be distinguished from cytosine on the basis of the current flowing through the pore during its interaction with each nucleotide.

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

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

The methods may involve measuring the current passing through the pore as the polynucleotide moves with respect to the pore. Therefore the apparatus may also comprise an electrical circuit capable of applying a potential and measuring an electrical signal across the membrane and pore. The methods may be carried out using a patch clamp or a voltage clamp. The methods preferably involve the use of a voltage clamp.

The methods of the invention may involve the measuring of a current passing through the pore as the polynucleotide moves with respect to the pore. Suitable conditions for measuring ionic currents through transmembrane protein pores are known in the art and disclosed in the Example. The method is typically carried out with a voltage applied across the membrane and pore. The voltage used is typically from +5 V to -5 V, such as from +4 V to -4 V, +3 V to -3 V or +2 V to -2 V. The voltage used is typically from -600 mV to +600mV or -400 mV to +400 mV. The voltage used is preferably in a range having a lower limit selected from -400 mV, -300 mV, -200 mV, -150 mV, -100 mV, -50 mV, -20mV and 0 mV and an upper limit independently selected from +10 mV, + 20 mV, +50 mV, +100 mV, +150 mV, +200 mV, +300 mV and +400 mV. The voltage used is more preferably in the range 100 mV to 240 mV and most preferably in the range of 120 mV to 220 mV. It is possible to increase discrimination between different nucleotides by a pore by using an increased applied potential.

The methods are typically carried out in the presence of any charge carriers, such as metal salts, for example alkali metal salt, halide salts, for example chloride salts, such as alkali metal chloride salt. Charge carriers may include ionic liquids or organic salts, for example tetramethyl ammonium chloride, trimethylphenyl ammonium chloride, phenyltrimethyl ammonium chloride, or 1-ethyl-3-methyl imidazolium chloride. In the exemplary apparatus

discussed above, the salt is present in the aqueous solution in the chamber. Potassium chloride (KCl), sodium chloride (NaCl), caesium chloride (CsCl) or a mixture of potassium ferrocyanide and potassium ferricyanide is typically used. KCl, NaCl and a mixture of potassium ferrocyanide and potassium ferricyanide are preferred. The charge carriers may be asymmetric across the membrane. The type and/or concentration of the charge carriers may be different on each side of the membrane.

The salt concentration may be at saturation. The salt concentration may be 3 M or lower and is typically from 0.1 to 2.5 M, from 0.3 to 1.9 M, from 0.5 to 1.8 M, from 0.7 to 1.7 M, from 0.9 to 1.6 M or from 1 M to 1.4 M. The salt concentration is preferably from 150 mM to 1 M. The method is preferably carried out using a salt concentration of at least 0.3 M, such as at least 0.4 M, at least 0.5 M, at least 0.6 M, at least 0.8 M, at least 1.0 M, at least 1.5 M, at least 2.0 M, at least 2.5 M or at least 3.0 M. High salt concentrations provide a high signal to noise ratio and allow for currents indicative of the presence of a nucleotide to be identified against the background of normal current fluctuations.

The methods are typically carried out in the presence of a buffer. In the exemplary apparatus discussed above, the buffer is present in the aqueous solution in the chamber. Any buffer may be used in the method of the invention. Typically, the buffer is phosphate buffer. Other suitable buffers are HEPES and Tris-HCl buffer. The methods are typically carried out at a pH of from 4.0 to 12.0, from 4.5 to 10.0, from 5.0 to 9.0, from 5.5 to 8.8, from 6.0 to 8.7 or from 7.0 to 8.8 or 7.5 to 8.5. The pH used is preferably about 7.5.

The methods 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.

Step (b) preferably comprises contacting the polynucleotide provided in step a) with a polynucleotide binding protein such that the protein controls the movement of the at least one strand of the polynucleotide through the pore.

More preferably, the method comprises b) contacting the polynucleotide provided in step a) with a transmembrane pore and a polynucleotide binding protein such that at least one strand of the polynucleotide moves through the pore and the protein controls the movement of the at least one strand of the polynucleotide through the pore; and c) measuring the current passing through the pore as the at least one strand of the polynucleotide moves with respect to the pore wherein the current is indicative of one or more characteristics of the at least one strand of the polynucleotide and thereby characterising the double stranded target polynucleotide.

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

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

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

Preferred enzymes are polymerases, exonucleases, helicases and topoisomerases, such as gyrases. Suitable enzymes include, but are not limited to, exonuclease I from *E. coli* (SEQ ID NO: 11), exonuclease III enzyme from *E. coli* (SEQ ID NO: 13), RecJ from *T. thermophilus* (SEQ ID NO: 15) and bacteriophage lambda exonuclease (SEQ ID NO: 17) and variants thereof. Three subunits comprising the sequence shown in SEQ ID NO: 15 or a variant thereof interact to form a trimer exonuclease. The enzyme is preferably Phi29 DNA polymerase (SEQ ID NO: 9) or a variant thereof. The topoisomerase is preferably a member of any of the Moiety Classification (EC) groups 5.99.1.2 and 5.99.1.3.

The enzyme is most preferably derived from a helicase, such as Hel308 Mbu (SEQ ID NO: 18), Hel308 Csy (SEQ ID NO: 19), Hel308 Mhu (SEQ ID NO: 21), TraI Eco (SEQ ID NO: 22), XPD Mbu (SEQ ID NO: 23) or a variant thereof. Any helicase may be used in the invention. The helicase may be or be derived from a Hel308 helicase, a RecD helicase, such as TraI helicase or a TrwC helicase, a XPD helicase or a Dda helicase. The helicase may be any of

the helicases, modified helicases or helicase constructs disclosed in International Application Nos. PCT/GB2012/052579 (published as WO 2013/057495); PCT/GB2012/053274 (published as WO 2013/098562); PCT/GB2012/053273 (published as WO2013098561); PCT/GB2013/051925 (published as WO 2014/013260); PCT/GB2013/051924 (published as WO 2014/013259); PCT/GB2013/051928 (published as WO 2014/013262) and PCT/GB2014/052736.

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

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

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

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

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

A variant of SEQ ID NOs: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 and which retains polynucleotide binding ability. This can be measured using any method known in the art. For instance, the variant can be contacted with a polynucleotide and its ability to bind to and move along the polynucleotide can be measured. The variant may include modifications that facilitate binding of the polynucleotide and/or facilitate its activity at high salt concentrations and/or room temperature. Variants may be

modified such that they bind polynucleotides (i.e. retain polynucleotide binding ability) but do not function as a helicase (i.e. do not move along polynucleotides when provided with all the necessary components to facilitate movement, e.g. ATP and Mg^{2+}). Such modifications are known in the art. For instance, modification of the Mg^{2+} binding domain in helicases typically results in variants which do not function as helicases. These types of variants may act as molecular brakes.

Over the entire length of the amino acid sequence of SEQ ID NO: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of SEQ ID NO: 9, 11, 13, 15, 17, 18, 19, 20, 21, 22, 23, 24 or 25 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 200 or more, for example 230, 250, 270, 280, 300, 400, 500, 600, 700, 800, 900 or 1000 or more, contiguous amino acids ("hard homology"). Homology is determined as described above. The variant may differ from the wild-type sequence in any of the ways discussed above with reference to SEQ ID NO: 2 and 4 above. The enzyme may be covalently attached to the pore. Any method may be used to covalently attach the enzyme to the pore.

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

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

Helicase(s) and molecular brake(s)

In a preferred embodiment, the invention provides a method of controlling the movement of a double stranded polynucleotide through a transmembrane pore, comprising:

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

(b) contacting the target polynucleotide provided in step (a) with the pore; and

(c) 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 target polynucleotide through the pore.

This type of method is discussed in detail in the International Application PCT/GB2014/052737.

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

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

Spacers in the Y adaptor

The one or more helicases attached to the Y adaptor may be stalled at one or more spacers as discussed in International Application No. PCT/GB2014/050175 (published as WO

2014/135838). Any configuration of one or more helicases and one or more spacers disclosed in the International Application may be used in this invention.

RTC sequencing

In a preferred embodiment, step b) comprises contacting the polynucleotide provided in step a) with a transmembrane pore such that both strands of the polynucleotide move through the pore and step c) comprises taking one or more measurements as the both strands of the polynucleotide move with respect to the pore wherein the measurements are indicative of one or more characteristics of the strands of the polynucleotide and thereby characterising the target polynucleotide. Any of the embodiments discussed above equally apply to this embodiment.

Uncoupling

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

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

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

(b) providing a second target double stranded polynucleotide in a second sample with a second Y adaptor at one end and a second hairpin loop adaptor at the other end, wherein the second Y adaptor comprises one or more third anchors for coupling the polynucleotide to the membrane, wherein the second hairpin loop adaptor comprises one or more fourth anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the second hairpin loop adaptor to the membrane is greater than the strength of coupling of the second Y adaptor to the membrane;

(c) coupling the first polynucleotide provided in step (a) to a membrane;

(d) contacting the first polynucleotide coupled in step (c) with a transmembrane pore such that at least one strand of the first polynucleotide moves through the pore;

(e) taking one or more measurements as the at least one strand of the first polynucleotide moves with respect to the pore wherein the measurements are indicative of one or

more characteristics of the at least one strand of the first polynucleotide and thereby characterising the first polynucleotide;

(f) uncoupling the first polynucleotide from the membrane;

(g) coupling the second polynucleotide provided in step (b) to the membrane;

(h) contacting the second polynucleotide coupled in step (g) with a transmembrane pore such that at least one strand of the second polynucleotide moves through the pore; and

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

This type of method is discussed in detail in the UK Application 1406155.0 and in the International application being filed concurrently with this application (ONT IP 055). Any of the embodiments discussed therein are applicable to this method.

Step (f) (i.e. uncoupling of the first polynucleotide) may be performed before step (g) (i.e. before coupling the second polynucleotide to the membrane). Step (g) may be performed before step (f). If the second polynucleotide is coupled to the membrane before the first polynucleotide is uncoupled, step (f) preferably comprises selectively uncoupling the first polynucleotide from the membrane (i.e. uncoupling the first polynucleotide but not the second polynucleotide from the membrane). A skilled person can design a system in which selective uncoupling is achieved. Steps (f) and (g) may be performed at the same time. This is discussed in more detail below.

Removal or washing

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

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

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

Modified target double stranded polynucleotide

Before step a), a target double stranded polynucleotide may be contacted with a polymerase and a population of free nucleotides under conditions in which the polymerase forms a modified target double stranded polynucleotide using the target double stranded 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 target double stranded polynucleotide may then be provided as in step a). This type of modification is described in the International Application No. PCT/GB2015/050483. Any of the polymerases discussed above may be used. The polymerase is preferably Klenow or 9o North.

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

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

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

Other characterisation method

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

Modification methods of the invention

The invention also provides methods of producing the modified target polynucleotide provided in step a) of the characterisation method.

The invention provides a method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane. The method involves ligating a Y adaptor to one end of the polynucleotide and ligating a hairpin loop adaptor to the other end of the polynucleotide. As above, the Y adaptor comprises one or more first anchors for coupling the polynucleotide to the membrane, the hairpin loop adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane and the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane.

The invention provides an alternative method in which the anchors are attached to the adaptors after the adaptors have been ligated to the target double stranded polynucleotide. The method involves ligating a Y adaptor to one end of the polynucleotide and ligating a hairpin loop adaptor to the other end of the polynucleotide. One or more first anchors are then attached to the Y adaptor and one or more second anchors are attached to the hairpin loop adaptor. As above, the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane. The one or more anchors may be attached to the adaptor is any way, including any of those discussed above. The one or more anchors are preferably attached to the adaptor via hybridisation.

Any methods of ligation may be used. Suitable methods are disclosed above.

The Y adaptor and hairpin loop adaptor may be any of those discussed above.

MuA-based methods

The invention also provides a method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising contacting the target polynucleotide with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein a proportion of the substrates in the population are hairpin loop adaptors comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby producing a plurality of modified double stranded polynucleotides.

The invention provides an alternative method in which the anchors are attached to the substrates after the target double stranded polynucleotide has been fragmented by the MuA transposase and the substrates have been ligated to the fragments. The method involves contacting the target polynucleotide with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors and wherein a proportion of the substrates in the population are hairpin loop adaptors. The transposase fragments the target polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs. The transposase preferably fragments the target polynucleotide and ligates a substrate to both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs. The transposase preferably produces a plurality of fragment/substrate constructs each comprising a Y adaptor at one end and a hairpin loop adaptor at the other end. The method also involves attaching to the Y adaptors in the plurality of fragment/substrate constructs one or more first anchors and attaching to the hairpin loop adaptors in the plurality of fragment/substrate constructs one or more second anchors and thereby producing a plurality of modified double stranded polynucleotides. The strength of coupling of each hairpin loop adaptor to the membrane is greater than the strength of coupling of each Y adaptor to the membrane.

The Y adaptor and hairpin loop adaptor may be any of those discussed above.

MuA-based fragmentation a target double stranded polynucleotide is disclosed in International Application No. PCT/GB2014/052505 (published as WO 2015/022544).

The target polynucleotide is contacted with a MuA transposase. This contacting occurs under conditions which allow the transposase to function, i.e. to fragment the target polynucleotide and to ligate MuA substrates to the one or both ends of the fragments. MuA transposase is commercially available, for instance from Thermo Scientific (Catalogue Number F-750C, 20μL (1.1μg/μL)). Conditions under which MuA transposase will function are known in the art.

The target polynucleotide is contacted with a population of double stranded MuA substrates. The double stranded substrates are polynucleotide substrates and may be formed from any of the nucleotides, polynucleotides or nucleic acids discussed above.

Each substrate typically comprises a double stranded portion which provides its activity as a substrate for MuA transposase. The double stranded portion is typically the same in each substrate. The population of substrates may comprise different double stranded portions. Each substrate preferably comprises a double stranded portion which comprises the sequence shown in SEQ ID NO: 26 hybridised to a sequence which is complementary to the sequence shown in SEQ ID NO: 26. The at least one overhang is preferably at the 5' end of the sequence which is complementary to the sequence shown in SEQ ID NO: 26.

In a preferred embodiment, each substrate in the population comprises at least one overhang of universal nucleotides such that the transposase fragments the target polynucleotide and ligates a substrate to one or both ends, preferably both ends, of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs and wherein the method further comprises ligating the overhangs to the fragments in the constructs and thereby producing a plurality of modified double stranded polynucleotides. The transposase preferably produces a plurality of fragment/substrate constructs each comprising a Y adaptor at one end and a hairpin loop adaptor at the other end.

Each substrate preferably comprises only one overhang. The only one overhang is preferably at the 5' end of one strand of the double stranded portion.

The overhang may be at least 3, at least 4, at least 5, at least 6 or at least 7 nucleotides in length. The overhang is preferably 5 nucleotides in length.

A universal nucleotide is one which will hybridise to some degree to all of the nucleotides in the target polynucleotide. Suitable universal nucleotides are described in International Application No. PCT/GB2014/052505 (published as WO 2015/022544).

The overhang(s) of universal nucleotides may further comprise a reactive group, preferably at the 5' end. The reactive group may be used to ligate the overhangs to the fragments in the constructs as discussed below. The reactive group may be used to ligate the fragments to

the overhangs using click chemistry. Suitable reactive groups are disclosed in International Application No. PCT/GB2014/052505 (published as WO 2015/022544).

In a further embodiment, the modification method uses a MuA transposase and a population of MuA substrates each comprising at least one overhang comprising a reactive group. The overhang(s) may be any length and may comprise any combination of any nucleotide(s). Suitable lengths and nucleotides are disclosed above. Suitable reactive groups are discussed above.

In another embodiment, the method comprises contacting the target polynucleotide with a population of double stranded MuA substrates each comprising (i) at least one overhang and (ii) at least one nucleotide in the same strand as the at least one overhang which comprises a nucleoside that is not present in the target polynucleotide such that the transposase fragments the target polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs. A skilled person can identify nucleosides not present in the target polynucleotide as described in International Application No. PCT/GB2014/052505 (published as WO 2015/022544). The transposase preferably fragments the target polynucleotide and ligates a substrate to both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs. The transposase preferably produces a plurality of fragment/substrate constructs each comprising a Y adaptor at one end and a hairpin loop adaptor at the other end. The overhangs can be removed from the constructs by selectively removing the at least one nucleotide to produce a plurality of double stranded constructs comprising single stranded gaps. The single stranded gaps in the constructs can be repaired to produce a plurality of modified double stranded polynucleotides.

One strand of the double stranded portion preferably comprises the sequence shown in SEQ ID NO: 26 and the other strand of the double stranded portion preferably comprises a sequence which is complementary to the sequence shown in SEQ ID NO: 26 and which is modified to include at least one nucleotide that is not present in the target polynucleotide. This "other strand" further comprises the overhang. In a more preferred embodiment, one strand of the double stranded portion comprises the sequence shown in SEQ ID NO: 26 and the other strand of the double stranded portion comprises the sequence shown in SEQ ID NO: 27 (see below). In SEQ ID NO: 27, the dA in the dC and dA dinucleotide at the 3' end had been replaced with dU. This double stranded portion (shown below) may be used when the target polynucleotide comprises deoxyadenosine (dA), thymidine (dT), deoxyguanosine (dG) and deoxycytidine (dC), but not deoxyuridine (dU).

5′-GTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA-3′ (SEQ 26)

3′-CAAAAGCGTAAATAGCACTTTGCGAAAGCGCAAAAAGCACGCGGCGAAGU-5′ (SEQ 27)

In a most preferred embodiment, one strand of the substrate comprises the sequence shown in SEQ ID NO: 26 and the other strand of the substrate comprises the sequence shown in SEQ ID NO: 28 (see below). This substrate (shown below) may be used when the target polynucleotide comprises deoxyadenosine (dA), thymidine (dT), deoxyguanosine (dG) and deoxycytidine (dC), but not deoxyuridine (dU).

5′-GTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTCGTGCGCCGCTTCA-3′ (SEQ 26)

3′-CAAAAGCGTAAATAGCACTTTGCGAAAGCGCAAAAAGCACGCGGCGAAGUCTAG-5′ (SEQ 28)

In all of the embodiments above, the proportion of one type of substrate may be any proportion, such as at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90% or at least about 95%. The remaining proportion of substrates in the population is typically the other type of substrate. For instance, the population may comprise about 40% of the substrates comprising a hairpin loop and about 60% of the Y substrates. The proportion of both types of substrate is preferably about 50%.

Methods for ligating the overhangs to the fragments, selectively removing the nucleotide(s) which comprise(s) a nucleoside that is not present in the target polynucleotide from the ligated constructs and repairing the single stranded gaps in the double stranded constructs are disclosed in International Application No. PCT/GB2014/052505 (published as WO 2015/022544).

Products of the invention

The invention also provides a target double stranded polynucleotide modified using a method of the invention. The modified polynucleotide comprises a Y adaptor attached at one end of the polynucleotide and a hairpin loop adaptor attached at the other end, wherein the Y adaptor comprises one or more first anchors for coupling the polynucleotide to the membrane, wherein the hairpin loop adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby providing a modified target double stranded polynucleotide.

The invention also provides a plurality of polynucleotides modified using the MuA-based methods of the invention. These modified polynucleotides preferably comprise a Y adaptor attached at one end and a hairpin loop adaptor attached at the other end, wherein the Y adaptor comprises one or more first anchor for coupling the polynucleotide to the membrane, wherein the hairpin loop adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby providing a modified target double stranded polynucleotide. Some of the polynucleotides may have Y adaptors as defined above at both ends or hairpin loop adaptors as defined above at both ends.

The invention also provides a pair of adaptors for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, wherein one adaptor is a Y adaptor comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein the other adaptor is a hairpin loop adaptor comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptors to the membrane is greater than the strength of coupling of the Y adaptor to the membrane.

The invention also provides population of adaptors for modifying a target polynucleotide for characterisation using a transmembrane pore in a membrane, wherein a proportion of the adaptors are Y adaptors comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein a proportion of the adaptors are hairpin loop adaptors comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane.

Each adaptor in the pair or population of the invention preferably comprises a double stranded MuA substrate. The substrates may be any of those described above. The substrates preferably comprise a double stranded portion as defined above. The double stranded portion preferably comprises SEQ ID NOs: 26 and 27 as discussed above. The double stranded portion more preferably comprises SEQ ID NOs: 26 and 28 as discussed above.

The proportion values given above in connection with the method of the invention are equally
applicable to the populations of the invention.

Any of the embodiments discussed above with reference to the methods of the invention are equally applicable to the polynucleotides, pairs and populations of the invention.

The population or plurality may be isolated, substantially isolated, purified or substantially purified. A population or plurality is isolated or purified if it is completely free of

any other components, such as the target polynucleotide, lipids or pores. A population or plurality is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a population or plurality 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 lipids or pores.

Kits

The present invention also provides a kit for modifying a target polynucleotide comprising (a) a pair of adaptors of the invention or a population of adaptors of the invention and (b) a MuA transposase. Each adaptor in the pair or population preferably comprises a double stranded MuA substrate.

Any of the embodiments discussed above with reference to the methods and products of the invention equally apply to the kits.

The kit may further comprise the components of a membrane, such as the components of an amphiphilic layer or a triblock copolymer membrane. The kit may further comprise a transmembrane pore or the components of a transmembrane pore. The kit may further comprise a polynucleotide binding protein. Suitable membranes, pores and polynucleotide binding proteins are discussed above.

The kit of the invention may additionally comprise one or more other reagents or instruments which enable any of the embodiments mentioned above to be carried out. Such reagents or instruments include one or more of the following: suitable buffer(s) (aqueous solutions), means to obtain a sample from a subject (such as a vessel or an instrument comprising a needle), means to amplify and/or express polynucleotides, a membrane as defined above or voltage or patch clamp apparatus. Reagents may be present in the kit in a dry state such that a fluid sample resuspends the reagents. The kit may also, optionally, comprise instructions to enable the kit to be used in the method of the invention or details regarding which patients the method may be used for. The kit may, optionally, comprise nucleotides.

The following Example illustrates the invention.

Example 1

This example describes the sample preparation procedure for the DNA constructs 1-6 shown in Figures 1 and 2 and used in Examples 2 and 3.

Materials and Methods

The strands used in this study are from a region of the lambda genome, between 45,042 bp and 48,487 bp. Analytes were made by the polymerase PCR method to include hybridisation sites at defined ends of each of the template and template complement strands as desired. PCR was carried out from lambda genomic DNA.

This template (SEQ ID NO: 29 hybridised to 30) was made using KAPA HiFi 2x Master mix, lambda DNA (NEB) and primers SEQ ID NO: 31 and SEQ ID NO: 32. Reactions were cycled 20 times and product of the correct size was purified by Gel Filtration on Sephacryl S1000 column and concentrated to 0.25mg/ml using Millipore Ultracel 15 50kDa concentrators.

DNA constructs (1, 2, 3, 4, 5 and 6) for electrophysiology experiments were all made according to the same reaction mix; 2x LongAmp Taq master mix, 300 nM of primers 1 and 2 or 3 and 4 (strand labelled a1 in Figures 1 and 2 was produced using primer 1 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 34 which is attached at its 3' end to four iSpC3 spacers which are attached to four 5-nitroindoles which are attached at the opposite end to the 5' end of SEQ ID NO: 36 and primer 2 = SEQ ID NO: 37 attached at the 3' end to four 5-nitroindoles which are attached at the opposite end to the 5' end of SEQ ID NO: 38, strand labelled a2 in Figures 1 and 2 was produced using primer 3 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37 which is attached at its 3' end to four iSpC3 spacers which are attached to four 5-nitroindoles which are attached at the opposite end to the 5' end of SEQ ID NO: 38 and primer 4 = SEQ ID NO: 34 attached at the 3' end to four 5-nitroindoles which are attached at the opposite end to the 5' end of SEQ ID NO: 36) and 1.2 ng ul^{-1} DNA template (SEQ ID NO: 29 hybridised to 30). DNA constructs were all amplified according to the same cycling program; 94°C for 2 mins, [94°C for 15 secs, 58°C for 30 secs, 65°C for 2 mins]$_{12}$ and 65°C for 5 mins. DNA constructs were all purified from a 0.8 % agarose gel according to manufacturer's instructions (Qiagen Gel Extraction kit) and then SPRI purified (Agencourt AMPure beads) according to manufacturer's instructions.

Finally, the DNA strands produced were hybridised to complementary strands of DNA, some of which contained anchors e.g. cholesterol or palmitate. The complementary strands SEQ ID NO: 39 and 40 (with and without attached anchors) were annealed at a five-fold excess at room temperature for ten minutes in 25 mM potassium phosphate buffer, 151 mM potassium chloride, pH 8.0.

DNA construct 1 was made up of four different strands hybridised together as shown in Figure 1(A) – a1 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 34; SEQ ID NO: 34 is attached at the 3' end to

four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 35, a3 = SEQ ID NO: 40 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG, a4 = SEQ ID NO: 39 and a5 = complementary sequence to a3 and part of a1.

DNA construct 2 was made up of four different strands hybridised together as shown in Figure 1(A) and (B) – a2 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37; SEQ ID NO: 37 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 41, a6 = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG, a7 = SEQ ID NO: 40 and a8 = complementary sequence to a6 and part of a2.

DNA construct 3 was made up of four different strands hybridised together as shown in Figure 1(B) and 2(B) – a1 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 34; SEQ ID NO: 34 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 35, a3 = SEQ ID NO: 40 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG, a9 = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' palmitate and a5 = complementary sequence to a3 and part of a1.

DNA construct 4 was made up of four different strands hybridised together as shown in Figure 2(A) – a1 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 34; SEQ ID NO: 34 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 35, a3 = SEQ ID NO: 40 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG, a6 = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG and a5 = complementary sequence to a3 and part of a1.

DNA construct 5 was made up of four different strands hybridised together as shown in Figure 2(A) – a1 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached

at the opposite end to the 5' end of SEQ ID NO: 34; SEQ ID NO: 34 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 35, a3 = SEQ ID NO: 40 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG, a2 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37; SEQ ID NO: 37 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 41 and a6 = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG.

DNA construct 6 was made up of four different strands hybridised together as shown in Figure 2(B) – a1 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 34; SEQ ID NO: 34 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 35, a10 = SEQ ID NO: 40 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' palmitate, a2 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37; SEQ ID NO: 37 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 41 and a9 = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG.

Example 2

This example compares the use of a single anchor to couple constructs to a membrane to that of two anchors of differing strengths again used to couple constructs to a membrane. Two anchors were employed in order to bias, the helicase-controlled DNA movements detected by the nanopore system, towards the construct which was doubly coupled to the membrane.

Materials and Methods

Prior to setting up the experiment, the DNA constructs either 1 and 2 or 2 and 3 (stock concentration 20 nM, final concentration added to nanopore system 0.1 nM) were separately pre-incubated at room temperature for five minutes with T4 Dda – E94C/A360C (stock concentration 250 nM, final concentration added to nanopore system 1 nM , SEQ ID NO: 24 with mutations

E94C/A360C) in buffer (151 mM KCl, 25 mM phosphate, 2 mM EDTA, pH8.0). After five minutes, TMAD (500 μM) was added to the pre-mix and the mixture incubated for a further 5 minutes. Finally, MgCl2 (10 mM final concentration), ATP (2.5 mM final concentration) and buffer (150 mM potassium ferrocyanide (II), 150 mM potassium ferricyanide and 25 mM potassium phosphate pH 8.0) were added to the pre-mix.

Electrical measurements were acquired from single MspA nanopores (MspA –B2C) inserted in block co-polymer in buffer (25 mM potassium phosphate, 150 mM potassium ferrocyanide (II), 150 mM potassium ferricyanide (III), pH 8.0). After achieving a single pore inserted in the block co-polymer, then buffer (2 mL, 25 mM potassium phosphate pH 8.0, 150 mM potassium ferrocyanide (II) and 150 mM potassium ferricyanide (III)) was flowed through the system to remove any excess MspA nanopores. The enzyme (T4 Dda – E94C/A360C, 1 nM final concentration), DNA construct either 1 and 2 or 2 and 3 (0.1 nM final concentration), fuel (MgCl2 10 mM final concentration, ATP 2.5 mM final concentration) pre-mix (150 μL total) was then added to the single nanopore experimental system and the experiment run at a holding potential of 120 mV for 2 hours and helicase-controlled DNA movement monitored.

Results and Discussion

For the control experiment, helicase-controlled DNA movement of constructs 1 and 2 was monitored. The cartoon at the top of Figure 3 shows where the helicase can bind to constructs 1 and 2. Helicase controlled DNA movements corresponding to each construct were identified and the proportion of movements corresponding to strand a1 (construct 1) and strand a2 (construct 2) were compared (see the bottom half of Figure 3). As each construct was coupled to the membrane using the same anchor (a cholesterol anchor) then, of the helicase-controlled DNA movements observed, approximately 50% corresponded to strand a1 and 50% to strand a2. This illustrated that by using the same single anchor, there was no bias towards helicase-controlled translocation movements for one construct over the other.

The experiment also compared the use of a single anchor (cholesterol) to couple construct 2 to the membrane, to the use of two different anchors (of differing strength – cholesterol and palmitate) to couple construct 3 to the membrane. The cartoon at the top of Figure 4 shows where the helicase can bind to strand a1 (constructs 3) and strand a2 (construct 2). Helicase controlled DNA movements corresponding to each construct were identified and the proportion of movements corresponding to a1 and a2 (construct 3 and 2 respectively) were compared (see the bottom half of Figure 4). In this case more than 95% of the helicase-controlled DNA movements detected corresponded to strand a1 (construct 3), which was coupled to the membrane using a

cholesterol and a palmitate. Less than 5% of the helicase-controlled DNA movements corresponded to strand a2 (construct 2). This experiment illustrated that by using two different anchors it was possible to strongly bias the helicase-controlled DNA movements detected by the nanopore system towards the doubly coupled construct over the singly coupled construct.

Example 3

This example compares the use of two anchors of equal strength (double cholesterol constructs 4 and 5 or double palmitate construct 6) to two anchors of differing strengths (palmitate and cholesterol construct 3).

Materials and Methods

DNA constructs 4 and 5 or 3 and 6 were pre-incubated with T4 Dda – E94C/A360C helicase as described in Example 2 above.

Electrical measurements were acquired from single MspA nanopores (MspA –B2C) inserted in block co-polymer in buffer as described in Example 2 and helicase-controlled DNA movements for constructs 4 and 5 or 3 and 6 were monitored.

Results and Discussion

For the control experiment, helicase-controlled DNA movement of constructs 4 and 5 was monitored. The cartoon at the top of Figure 5 shows where the helicase can bind to constructs 4 (only on a1) and 5 (on a1 and a2). Strand a1 of construct 4 was capable of being captured by the nanopore, whereas, strand a1 or strand a2 of construct 5 were capable of being captured by the nanopore. Helicase controlled DNA movements corresponding to each construct were identified and the proportion of movements corresponding to strand a1 (construct 4 and 5) and strand a2 (construct 5 only) were compared (see the bottom half of Figure 5). As each construct was coupled to the membrane using the same anchor (two cholesterol anchors) then of the helicase-controlled DNA movements observed, approximately 66% corresponded to strand a1 (construct 4 and 5) and 33% to strand a2 (construct 5 only). The 2:1 bias towards the detection of a1 helicase controlled DNA movements was expected because a2 movements would only be detected if construct 4 was captured by the nanopore, whereas a1 movements would be detected from capture of both construct 4 and 5, therefore, twice as many a1 movements were expected (as equivalent concentrations of constructs 4 and 5 were added to the nanopore experimental system).

The experiment also compared the use of two coupling agents of differing strengths (cholesterol and palmitate) to couple construct 3 to the membrane, to the use of two identical

coupling agents (both palmitate) to couple construct 6 to the membrane. The cartoon at the top of Figure 6 shows where the helicase can bind to strand a1 (constructs 3 and 6) and strand a2 (construct 6 only). Strand a1 of construct 3 was capable of being captured by the nanopore, whereas, strand a1 or strand a2 of construct 6 were capable of being captured by the nanopore. Helicase controlled DNA movements corresponding to each construct were identified and the proportion of movements corresponding to a1 (construct 2 and 3) and a2 (construct 2 only) were compared (see the bottom half of Figure 6). In this case more than 95% of the helicase-controlled DNA movements detected corresponded to strand a1 (construct 3 and 6), which were coupled to the membrane using either a cholesterol and a palmitate or two palmitates. Less than 5% of the helicase-controlled DNA movements corresponded to strand a2 (construct 6 only). This experiment illustrated that by using two different coupling agents it was possible to strongly bias the helicase-controlled DNA movements detected by the nanopore system towards the construct coupled by coupling agents of differing strength rather than the construct coupled by two identical coupling agents of equal strength. This also shows that by using the combination of a stronger and weaker anchor on a construct that can only be captured by the nanopore at one end it was possible to bias towards this construct over a construct that has two weaker anchors on a construct that can be captured by the nanopore at either end.

Example 4

This example shows how the relative strengths of a number of coupling agents can be compared. This would allow selection of appropriate strength coupling agents for coupling of a construct to a membrane.

Materials and Methods

DNA constructs 1, 2 and 7(a-c) were produced using the method described in Example 1. Constructs 1 and 2 are described in detail above and constructs 7(a-c) are described below.

DNA construct 7a was made up of four different strands hybridised together as shown in Figure 7(B) – a2 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37; SEQ ID NO: 37 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 41, a10(a) = SEQ ID NO: 42 which has six iSp18 spacers attached to its 5' end which are attached at the opposite end to two thymines and a 5' cholesterol TEG; SEQ ID NO: 42 has a further six iSp18 spacers attached to its 3' end which are attached at

the opposite end to two thymines and a 3' cholesterol TEG, a7 = SEQ ID NO: 40 and a8 = complementary sequence to A10a and part of a2 .

DNA construct 7b was made up of four different strands hybridised together as shown in Figure 7(B) – a2 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37; SEQ ID NO: 37 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 41, a10(b) = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' tocopherol, a7 = SEQ ID NO: 40 and a8 = complementary sequence to A10b and part of a2 .

DNA construct 7c was made up of four different strands hybridised together as shown in Figure 7(B) – a2 = SEQ ID NO: 33 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 37; SEQ ID NO: 37 is attached at the 3' end to four iSpC3 spacers which are attached at the opposite end to four 5-nitroindoles which are attached to the 5' end of SEQ ID NO: 41, a10(c) = SEQ ID NO: 39 which is attached at its 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' palmitate, a7 = SEQ ID NO: 40 and a8 = complementary sequence to A10c and part of a2 .

DNA constructs 1, 2 and 7(a-c) were pre-incubated with T4 Dda – E94C/A360C helicase as described in Example 2 above.

Electrical measurements were acquired from single MspA nanopores (MspA –B2C) inserted in block co-polymer in buffer as described in Example 2 and helicase-controlled DNA movements for constructs 1, 2 and 7(a-c) were monitored.

Results and Discussion

For the control experiment, helicase-controlled DNA movement of constructs 1 and 2 was monitored. The cartoon at the top of Figure 8 shows where the helicase can bind to constructs 1 and 2. Helicase controlled DNA movements corresponding to each construct were identified and the proportion of movements corresponding to strand a1 (construct 1) and strand a2 (construct 2) were compared (see the bottom half of Figure 8). As each construct was coupled to the membrane using the same coupling agent cholesterol) then, of the helicase-controlled DNA movements observed, approximately 50% corresponded to strand a1 and 50% to strand a2. This illustrated that by using the same single coupling agent, there was no bias towards helicase-controlled translocation movements for one construct over the other.

The experiment also compared the use of a variety of different coupling agents to couple constructs 7(a-c) to the membrane to that of cholesterol which was used to couple construct 1 to the membrane. The cartoons at the top of Figures 9-11 show where the helicase could bind to strand a1 (constructs 1) and strand a2 (construct 7(a-c)). Helicase controlled DNA movements corresponding to each construct were identified and the proportion of movements corresponding to a1 and a2 (construct 1 and 7(a-c) respectively) were compared (see the bottom half of Figures 9-11). The use of two cholesterols in the same strand of DNA (A10a) was observed to have an equivalent coupling strength to that of the single cholesterol (see Figure 9). Whereas, the use of a tocopherol was observed to be a stronger coupling agent than cholesterol (see Figure 10), as a greater number of DNA controlled helicase movements were observed for the construct anchored by tocopherol than cholesterol. Furthermore, palmitate was observed to be a weaker coupling agent than cholesterol because fewer helicase controlled DNA movements were observed for the construct anchored by palmitate than cholesterol (see Figure 11). Therefore, by comparing the coupling strengths of various coupling agents using this method it would be possible to select two different strength coupling agents and bias the system towards selection of a desired construct (see example 6 for further details).

Example 5

This example describes the sample preparation procedure for the DNA constructs 8 and 9 shown in Figure 12 and used in Examples 6.

Materials and Methods

Lambda DNA (1 µg, SEQ ID NO: 43) was restriction enzyme digested for 1 hour at 37°C in 1x NEBuffer 4, with 5 U of SnaBI and 20 U of BamHI-HF. Following digestion, the DNA was purified using 1.5x Agencourt AMPure beads, according to the manufacturer's protocol. The sample was then eluted in 50 uL of nuclease free water. The DNA was then end-repaired using NEB's NEBNext End-repair system, following the manufacturer's protocol. The sample was again purified using 1.5x Agencourt AMPure beads and the DNA eluted in 42 uL of nuclease free water. Following the end-repair step, the purified DNA was dA-tailed using NEB's NEBNext dA-tailing system, according to the manufacturer's protocol. The DNA was again purified using 1.5x Agencourt AMPure beads, eluting in 20 uL of nuclease free water. The DNA was quantified at A260 nm and adapters ligated at a 25-fold excess per 5' end, using NEB's NEBNext quick ligase module following the manufacturer's protocol. Construct 8 was produced by ligating the following adapters to the fragments of DNA (labelled as a12 in Figure 12) (adapter 1 (labelled a11

in Figure 12) = 25 iSpC3 spacers attached at to the 5' to SEQ ID NO: 44 which is attached at its 3' end to four iSp18 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 45 which is attached at the 3' end to the sequence AACCT (which are joined together by phosphorothioate bonds) where the final T in the sequence is attached to a further non-modified T; adapter 2 (labelled as a14 in Figure 12) = has a phosphate group at the 5' end of sequence GGTT (which are joined together by phosphorothioate bonds), this sequence is attached at its 3' end to the 5' end of SEQ ID NO: 46 which is attached at the 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' cholesterol TEG and adapter 3 (labelled as a13 in Figure 12) = has a phosphate group at the 5' end of SEQ ID NO: 47 and which has an internal cholesterol TEG attached to the 3' end of SEQ ID NO: 47, the internal cholesterol is also attached to the 5' end of SEQ ID NO: 48 which is attached at its 3' end to a G with a phosphorothioate bond and a non-modified T). Construct 9 was produced by ligating the following adapters to the fragments of DNA (labelled as a12 in Figure 12) (adapter 1 (labelled as a11 in Figure 12) = 25 iSpC3 spacers attached at to the 5' to SEQ ID NO: 44 which is attached at its 3' end to four iSp18 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 45 which is attached at the 3' end to the sequence AACCT (which are joined together by phosphorothioate bonds) where the final T in the sequence is attached to a further non-modified T; adapter 4 (labelled as a15 in Figure 12) = has a phosphate group at the 5' end of sequence GGTT (which are joined together by phosphorothioate bonds), this sequence is attached at its 3' end to the 5' end of SEQ ID NO: 46 which is attached at the 3' end to six iSp18 spacers which are attached at the opposite end to two thymines which are attached at the opposite end to a 3' palmitate and adapter 3 (labelled as a13 in Figure 12) = has a phosphate group at the 5' end of SEQ ID NO: 47 and which has an internal cholesterol TEG attached to the 3' end of SEQ ID NO: 47, the internal cholesterol is also attached to the 5' end of SEQ ID NO: 48 which is attached at its 3' end to a G with a phosphorothioate bond and a non-modified T). Following digestion, the ligated DNA was purified using 0.4x Agencourt AMPure beads, eluting in 16 uL of nuclease free water. The DNA was quantified at A260 nm and 4 uL of 757.5 mM KCl, 125 mM potassium phosphate buffer (pH 7) and 5 mM EDTA was added.

Example 6

This example compares the helicase controlled DNA movements detected for constructs 8 and 9. Construct 8 was coupled to the membrane using two cholesterols (both strong coupling agents). Construct 9 was coupled to the membrane using cholesterol in the hairpin and a palmitate

in the Y-adaptor. The constructs both contained lambda DNA which had been fragmented by the process described above.

Materials and Methods

DNA constructs 8 and 9 were each separately pre-incubated with T4 Dda – E94C/A360C/C109A/C136A helicase using a similar procedure to that described in Example 2 above.

Electrical measurements were acquired from single MspA nanopores (MspA –B2C) inserted in block co-polymer in buffer as described in Example 2 and helicase-controlled DNA movements for constructs 8 and 9 were monitored.

Results and Discussion

During the fragmentation and adapter attachment process (described in Example 5) the desired constructs shown in Figure 12 were produced. However, other constructs were also produced during that process which had either two Y-adapters attached (constructs 11 and 12) or two hairpin adapters attached (construct 10) (see Figures 13 and 14). In order for the nanopore to be able to capture DNA, the DNA must have had at least one Y-adapter so that there was a free end to be captured by the pore. Therefore, the DNA constructs with two hairpins attached (construct 10) were not be observed in the nanopore experiments. The constructs which had a Y-adapter on both ends of the lambda DNA fragments could be captured by the nanopore by either end of the DNA (constructs 11 and 12). However, as there was no hairpin attached to the DNA, then only strand x or strand y was translocated through the nanopore. However, if constructs 8 or 9 were captured by the nanopore (via the Y-adapter end) then both strand X and strand Y translocated through the nanopore owing to the hairpin attached at the opposite end. Therefore, it was desired that the constructs which contained a Y adapter and a hairpin were preferentially captured by the nanopore. It was not possible to separate out the various constructs (10, 8 and 11 or 10, 9 and 12) in order to isolate just the desired constructs 8 and 9.

It was investigated as to whether constructs which contained a Y-adapter and a hairpin adapter could be preferentially selected and captured by the nanopore by utilising different combinations of coupling agents. In this example coupling of two cholesterols (construct 8) was compared to coupling of cholesterol and a palmitate (construct 9). Both of the constructs tested exhibited high numbers of helicase controlled DNA movements of either the hairpin/Y-adapter constructs (8 and 9) or the double Y-adapter constructs (11 and 12). When the same coupling agent (cholesterol) was used in both the Y-adapter and the hairpin (construct 8) then 25% of the

helicase-controlled DNA movements detected corresponded to construct 8 and 75% to construct 11. Therefore, only a quarter of the movements detected corresponded to the desired construct. When a strong coupling agent (cholesterol in this case) was used in the hairpin and a weaker coupling agent was used in the Y-adapter (palmitate in this case) 46% of the helicase-controlled DNA movements detected corresponded to construct 9 and 56% to construct 12. This is a significant improvement towards selection of the desired construct 9. This improvement in detection of the desired construct was owing to the use of two coupling agents of differing strengths. The double Y-adapter construct 12 was coupled to the membrane using two weaker coupling agents in comparison to the Y-adapter/hairpin construct 9 which was coupled to the membrane using one weak and one strong anchor. As the desired construct employed one anchor which was stronger than those used in the double Y-adapter construct this was preferentially coupled to the membrane and therefore a larger proportion of the helicase-controlled DNA movements detected corresponded to desired construct 9.

Example 7

This example compares the coupling efficiency of a number of different DNA constructs 13, 14, 15 and 16 which were coupled to the membrane by a number of means. Construct 13 was coupled to the membrane using cholesterol in the hairpin region and a palmitate hybridised by a rigid DNA linker in the Y-adaptor region. Construct 14 was coupled to the membrane using cholesterol in the hairpin region and a palmitate hybridised by a flexible DNA linker in the Y-adaptor region. Construct 15 was coupled to the membrane using cholesterol in the hairpin region and a cholesterol hybridised by a rigid DNA linker in the Y-adaptor region. Construct 16 was coupled to the membrane using cholesterol in the hairpin region and a cholesterol hybridised by a flexible DNA linker in the Y-adaptor region.

Materials and Methods

The constructs all contained lambda DNA fragments (labelled in Figures 15 and 16 as a12) which had been fragmented by the process described above in Example 5, however, different adapters were then ligated to the DNA constructs. The adapters ligated to form DNA constructs 13-16 (shown in Figures 15 and 16) are described in detail below -

Constructs 13, 14, 15 and 16 were produced by ligating the following adapters to the fragments of lambda DNA (labelled as a12 in Figure 15 and 16) (adapter 1 (labelled a16 in Figure 15) = 30 iSpC3 spacers attached to the 5' to SEQ ID NO: 51 which is attached at its 3' end to four iSp18 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 52; adapter 2

(labelled as a17 in Figure 15) = has a phosphate group at the 5' end of SEQ ID NO: 49, this sequence is attached at its 3' end to four iSpC3 spacer which are attached at the opposite end to the 5' end of SEQ ID NO: 50 which is attached at the 3' end to a Thymine via a phosphothioate bond; and adapter 3 (labelled as a18 in Figure 15) = has a phosphate group at the 5' end of SEQ ID NO: 53.

After purification of the above constructs as described in Example 5 then either rigid or flexible DNA linkers with either palmitate or cholesterol attached at the 5' end were then hybridised to the constructs under the following conditions - 0.2ul of each DNA linker (10uM) was added to 13ul of ligated DNA (stock conc. approximately 20-30 nM) and incubated at room temperature for 10 mins.

Construct 13 was hybridised to the adaptor 4 (labelled in Figure 15 as a20) = SEQ ID NO: 40 is attached at its 5' end to six iSp18 spacers which are attached at the opposite end to two thymines and a 5' cholesterol TEG and adaptor 5 (labelled in Figure 15 as a19) = SEQ ID NO: 54 which is attached at the 5' end to a palmitate.

Construct 14 was hybridised to the adaptor 4 (labelled in Figure 15 as a20) = SEQ ID NO: 40 is attached at its 5' end to six iSp18 spacers which are attached at the opposite end to two thymines and a 5' cholesterol TEG and adaptor 6 (labelled in Figure 15 as a21) = SEQ ID NO: 54 which is attached at its 5' end to six iSp18 spacers which are attached at the opposite end to two thymines and a palmitate.

Construct 15 was hybridised to the adaptor 4 (labelled in Figure 15 as a20) = SEQ ID NO: 40 is attached at its 5' end to six iSp18 spacers which are attached at the opposite end to two thymines and a 5' cholesterol TEG and adaptor 5 (labelled in Figure 15 as a22) = SEQ ID NO: 54 which is attached at the 5' end to a cholesterol TEG.

Construct 16 was hybridised to the adaptor 4 (labelled in Figure 15 as a20) = SEQ ID NO: 40 is attached at its 5' end to six iSp18 spacers which are attached at the opposite end to two thymines and a 5' cholesterol TEG and adaptor 6 (labelled in Figure 15 as a23) = SEQ ID NO: 54 which is attached at its 5' end to six iSp18 spacers which are attached at the opposite end to two thymines and a cholesterol TEG.

Prior to setting up the experiment, the DNA constructs 13-16 (final concentration added to nanopore system 0.2 nM) were separately pre-incubated at room temperature for five minutes with T4 Dda – E94C/A360C/C109A/C136A (final concentration added to nanopore system 10 nM) in buffer (151 mM KCl, 25 mM phosphate, 5% glycerol, pH7.0). After five minutes, TMAD (500 μM) was added to the pre-mix and the mixture incubated for a further 5 minutes.

Finally, MgCl2 (2 mM final concentration), ATP (2.5 mM final concentration) and buffer (500 mM KCl and 25 mM potassium phosphate pH 8.0) were added to the pre-mix.

Electrical measurements were acquired from MspA nanopores (MspA –B2C) inserted in block co-polymer in buffer as described in Example 2 and helicase-controlled DNA movements for constructs 13-16 were monitored.

Results and Discussion

This example compares the coupling efficiency of a number of different DNA constructs 13, 14, 15 and 16 by investigating the number of helicase controlled DNA movements which were observed for each nanopore over the course of an experiment and the percentage of these helicase-controlled movements which corresponded to translocation of both region R1 and R2 when compared to all other helicase-controlled DNA movements detected.

For all four of the different constructs tested, helicase controlled DNA movements were observed. In this experiment we have compared the strengths of two different types of anchor – palmitate and cholesterol. It was also possible to compare how the flexibility of two different types of double-stranded polynucleotide linkers affected the strength of the coupling. In all four constructs, the coupling agent in the hairpin a20 (flexible cholesterol tether) remained constant and the coupling agent used in the Y-adapter was varied between palmitate/cholesterol and flexible/rigid linkers.

The constructs used in this experiment were made by the same fragmentation and adapter attachment process described in Example 5. This produced the desired constructs shown in Figures 15 and 16 as well as constructs analogous to those shown in Figures 13 and 14 which had either two Y-adapters attached or two hairpin adapters attached. As described in more detail in example 6 above, in order for the nanopore to be able to capture DNA, the DNA must have had at least one Y-adapter so that there was a free end which was able to be captured by the pore. Therefore, DNA constructs with two hairpins attached were not observed in the nanopore experiments. The constructs which had a Y-adapter on both ends of the DNA fragments were captured by the nanopore by either end of the DNA. However, as there was no hairpin attached to the DNA, then only R1 or R2 would have been translocated through the nanopore. However, if the desired constructs 13-16 were captured by the nanopore (via the Y-adapter end) then regions R1 and R2 would have both translocated through the nanopore. Therefore, it was desired that the constructs which contained a Y adapter and a hairpin were preferentially captured by the nanopore.

The data shown in Figure 17 illustrates the % of the total number of helicase controlled DNA movements detected which corresponded to the translocation of both R1 and R2 regions through the nanopore (which are produced from the desired constructs which contained a Y-adapter and a hairpin adapter). Other helicase-controlled DNA movements which were observed corresponded to translocation of constructs which had two Y-adapters attached to the fragments of lambda DNA rather than a Y-adapter and a hairpin adapter. The highest percentage of helicase-controlled DNA movements of both R1 and R2 regions through the nanopore (and corresponded to the desired construct) was observed when a palmitate was used with either a rigid or flexible linker (constructs 13 and 14). The cholesterol with the rigid tether had a higher percentage of helicase-controlled DNA movements which corresponded to translocation of both R1 and R2 through the nanopore than the cholesterol with the flexible tether.

It is also clear from the data shown in Figure 18 that for both palmitate (constructs 13 and 14) and cholesterol (constructs 15 and 16) the rigid arrangement was a slightly weaker coupling arrangement than the flexible arrangement as constructs 13 and 15 resulted in fewer helicase-controlled DNA movements detected per nanopore. The weakest coupling arrangement tested was in construct 15 where a rigid linker attached to a palmitate was used.

The preferred coupling system would be one which resulted in a high number of helicase-controlled DNA movements per nanopore which also had the highest percentage of helicase-controlled DNA movements which corresponded to translocation of regions R1 and R2. The experiments showed that the highest % of helicase controlled DNA movements for R1 and R2 resulted in the lowest number of helicase-controlled DNA movements per nanopore and vice versa. Therefore, the preferred construct design for optimum overall throughput of the desired helicase controlled DNA movements was construct 15 which had a high number of helicase-controlled DNA movements per nanopore with a percentage of helicase-controlled DNA movements which corresponded to translocation of regions R1 and R2 of just under 50%.

CLAIMS

1. A method of characterising a target double stranded polynucleotide using a transmembrane pore in a membrane, comprising:

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

b) contacting the polynucleotide provided in step a) with the transmembrane pore such that at least one strand of the polynucleotide moves through the pore; and

c) taking one or more measurements as the at least one strand of the polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the at least one strand of the polynucleotide and thereby characterising the double stranded target polynucleotide.

2. A method according to claim 1, wherein (a) the hairpin loop adaptor comprises more anchors than the Y adaptor, (b) the strength of coupling of one or more second anchors to the membrane is greater than the strength of coupling of the one or more first anchors to the membrane (c) the strength of coupling of the one or more second anchors to the hairpin loop adaptor is greater than the strength of coupling of the one or more first anchors to the Y adaptor, (d) the one or more first anchors and one or more second anchors couple their respective adaptors to the membrane via hybridisation and the strength of hybridisation is greater in the one or more second anchors than in the one or more first anchors, (e) the one or more first anchors comprise one or more rigid linkers and the one or more second anchors comprise one or more flexible linkers or (f) any combination of (a) to (e).

3. A method according to claim 2, wherein (i) the one or more second anchors couple to the membrane using cholesterol and the one or more first anchors couple to the membrane using palmitate, (ii) the one or more second anchors couple to the membrane using a mono-acyl species and the one or more first anchors couple to the membrane using a diacyl species, (iii) the one or more first anchors comprise one or more polynucleotide linkers and the one or more second anchors comprise one or more flexible linkers comprising one or more one or more

spacer 9 (iSp9) groups or one or more spacer 18 (iSp18), (iv) a combination of (i) and (iii) or (v) a combination of (ii) and (iii).

4. A method according to any one of the preceding claims, wherein step a) comprises modifying the target double stranded polynucleotide so that it comprises the Y adaptor at one end and the hairpin loop adaptor at the other end.

5. A method according to any one of the preceding claims, wherein step b) comprises contacting the polynucleotide provided in step a) with a transmembrane pore such that both strands of the polynucleotide move through the pore and step c) comprises taking one or more measurements as the both strands of the polynucleotide move with respect to the pore wherein the measurements are indicative of one or more characteristics of the strands of the polynucleotide and thereby characterising the target polynucleotide.

6. A method according to any one of the preceding claims, wherein the one or more characteristics are 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.

7. A method according to any one of the preceding claims, wherein the one or more characteristics of the polynucleotide are measured by electrical measurement and/or optical measurement.

8. A method according to claim 7, wherein the electrical measurement is a current measurement, an impedance measurement, a tunnelling measurement or a field effect transistor (FET) measurement.

9. A method according to any one of the preceding claims, wherein step b) further comprises contacting the polynucleotide provided in step a) with a polynucleotide binding protein such that the protein controls the movement of the at least one strand of the polynucleotide through the pore.

10. A method according to claim 9, wherein the method comprises:

b) contacting the polynucleotide provided in step a) with a transmembrane pore and a polynucleotide binding protein such that at least one strand of the polynucleotide moves through

the pore and the protein controls the movement of the at least one strand of the polynucleotide through the pore; and

c) measuring the current passing through the pore as the at least one strand of the polynucleotide moves with respect to the pore wherein the current is indicative of one or more characteristics of the at least one strand of the polynucleotide and thereby characterising the double stranded target polynucleotide.

11. A method according to claim 9 or 10, wherein the polynucleotide binding protein is derived from a helicase.

12. A method according to any one of the preceding claims, wherein the membrane is an amphiphilic layer or a solid state layer.

13. A method according to any one of the preceding claims, wherein the transmembrane pore is a transmembrane protein pore.

14. A method according to claim 13, wherein the transmembrane protein pore is derived from *Mycobacterium smegmatis* porin (Msp), α-hemolysin (α-HL) or lysenin.

15. A method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising ligating a Y adaptor to one end of the polynucleotide and ligating a hairpin loop adaptor to the other end of the polynucleotide, wherein the Y adaptor comprises one or more first anchors for coupling the polynucleotide to the membrane, wherein the hairpin loop adaptor comprises one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby providing a modified target double stranded polynucleotide.

16. A method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising

(a) ligating a Y adaptor to one end of the polynucleotide and ligating a hairpin loop adaptor to the other end of the polynucleotide; and

(b) attaching to the Y adaptor one or more first anchors for coupling the polynucleotide to the membrane, attaching to the hairpin loop adaptor one or more second anchors for coupling the

polynucleotide to the membrane and thereby providing a modified target double stranded polynucleotide;

wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane.

17. A method according to claim 15 or 16, wherein the adaptors and/or anchors are as defined in claim 2 or 3.

18. A target double stranded polynucleotide modified using a method according to any one of claims 15 to 17.

19. A method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising contacting the target polynucleotide with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein a proportion of the substrates in the population are hairpin loop adaptors comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane, and thereby producing a plurality of modified double stranded polynucleotides.

20. A method for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, comprising:

(a) contacting the target polynucleotide with a MuA transposase and a population of double stranded MuA substrates, wherein a proportion of the substrates in the population are Y adaptors and wherein a proportion of the substrates in the population are hairpin loop adaptors, such that the transposase fragments the target polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs; and

(b) attaching to the Y adaptors in the plurality of fragment/substrate constructs one or more first anchors and attaching to the hairpin loop adaptors in the plurality of fragment/substrate constructs one or more second anchors and thereby producing a plurality of modified double stranded polynucleotides;

wherein the strength of coupling of each hairpin loop adaptor to the membrane is greater than the strength of coupling of each Y adaptor to the membrane.

21. A method according to claim 19 or 20, wherein each substrate in the population comprises at least one overhang of universal nucleotides such that the transposase fragments the target polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs and wherein the method further comprises ligating the overhangs to the fragments in the constructs and thereby producing a plurality of modified double stranded polynucleotides.

22. A method according to claim 19 or 20, wherein each substrate in population comprises (i) at least one overhang and (ii) at least one nucleotide in the same strand as the at least one overhang which comprises a nucleoside that is not present in the target polynucleotide such that the transposase fragments the target polynucleotide and ligates a substrate to one or both ends of the double stranded fragments and thereby produces a plurality of fragment/substrate constructs, and wherein the method further comprises:

(a) removing the overhangs from the constructs by selectively removing the at least one nucleotide and thereby producing a plurality of double stranded constructs comprising single stranded gaps; and

(b) repairing the single stranded gaps in the constructs and thereby producing a plurality of modified double stranded polynucleotides.

23. A plurality of polynucleotides modified using a method according to any one of the claims 19 to 22.

24. A method of characterising a target double stranded polynucleotide using a transmembrane pore in a membrane, comprising:

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

b) contacting the polynucleotide provided in step a) with the transmembrane pore, a polymerase and labelled nucleotides such that phosphate labelled species are sequentially

released when nucleotides are added to the polynucleotide by the polymerase, wherein the phosphate species contain a label specific for each nucleotide; and

c) detecting the phosphate labelled species using the pore and thereby characterising the double stranded target polynucleotide.

25. A pair of adaptors for modifying a target double stranded polynucleotide for characterisation using a transmembrane pore in a membrane, wherein one adaptor is a Y adaptor comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein the other adaptor is a hairpin loop adaptor comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptor to the membrane is greater than the strength of coupling of the Y adaptor to the membrane.

26. A population of adaptors for modifying a target polynucleotide for characterisation using a transmembrane pore in a membrane, wherein a proportion of the adaptors are Y adaptors comprising one or more first anchors for coupling the polynucleotide to the membrane, wherein a proportion of the adaptors are hairpin loop adaptors comprising one or more second anchors for coupling the polynucleotide to the membrane and wherein the strength of coupling of the hairpin loop adaptors to the membrane is greater than the strength of coupling of the Y adaptors to the membrane.

27. A pair of adaptors according to claim 25 or a population of adaptors according to claim 26, wherein each adaptor comprises a double stranded MuA substrate.

28. A pair of adaptors or a population of adaptors according to claim 27, wherein each adaptor is as defined in claim 21 or 22.

29. A kit for modifying a target polynucleotide comprising (a) a pair of adaptors or a population of adaptors according to claim 28 and (b) a MuA transposase.

Figure 1 1/18

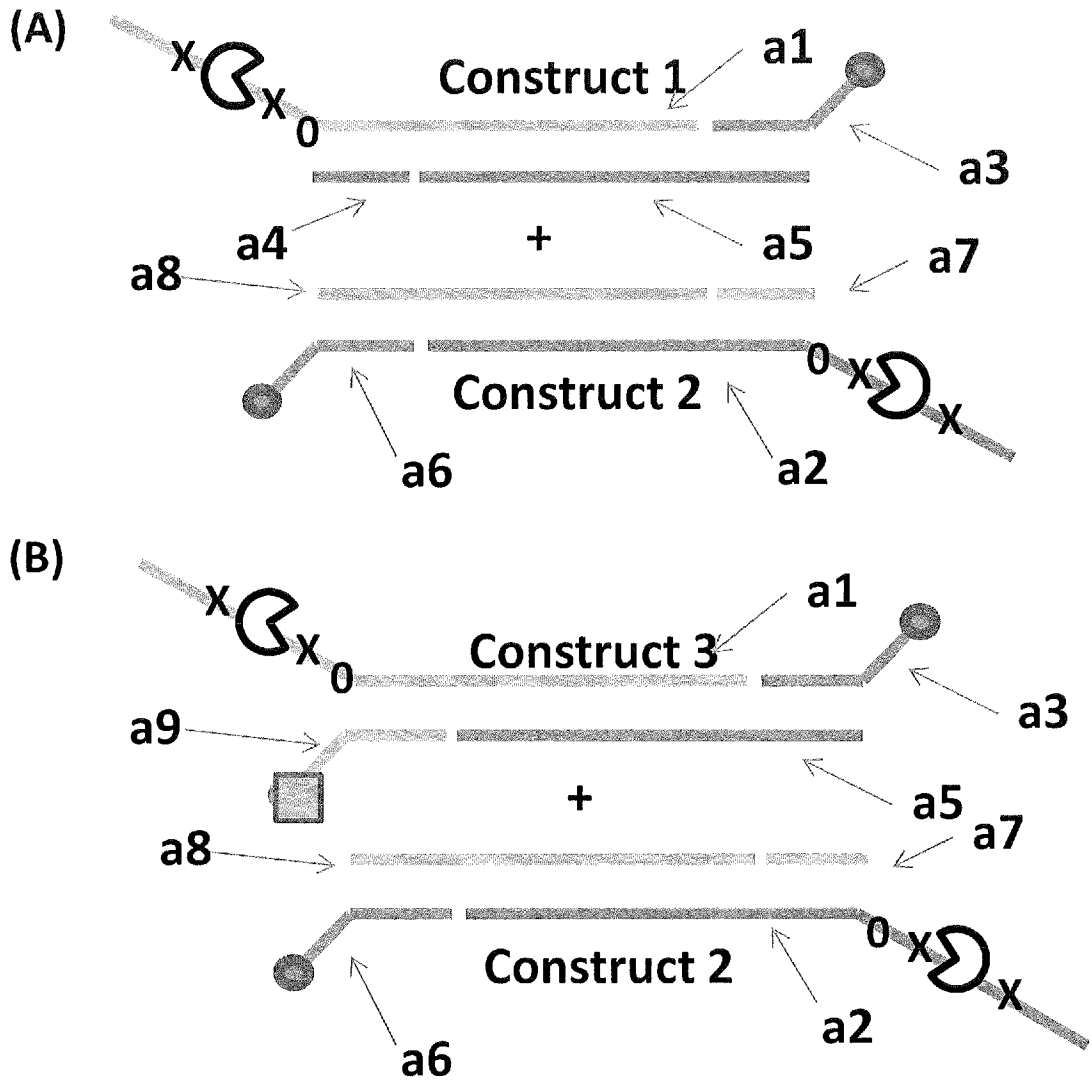

(A)
X
X
0
Construct 1
a1
a3
a4
+
a5
a7
a8
0
X
X
Construct 2
a6
a2
(B)
X
X
0
Construct 3
a1
a3
a9
+
a5
a7
a8
Construct 2
0
X
X
a6
a2

Figure 2 2/18

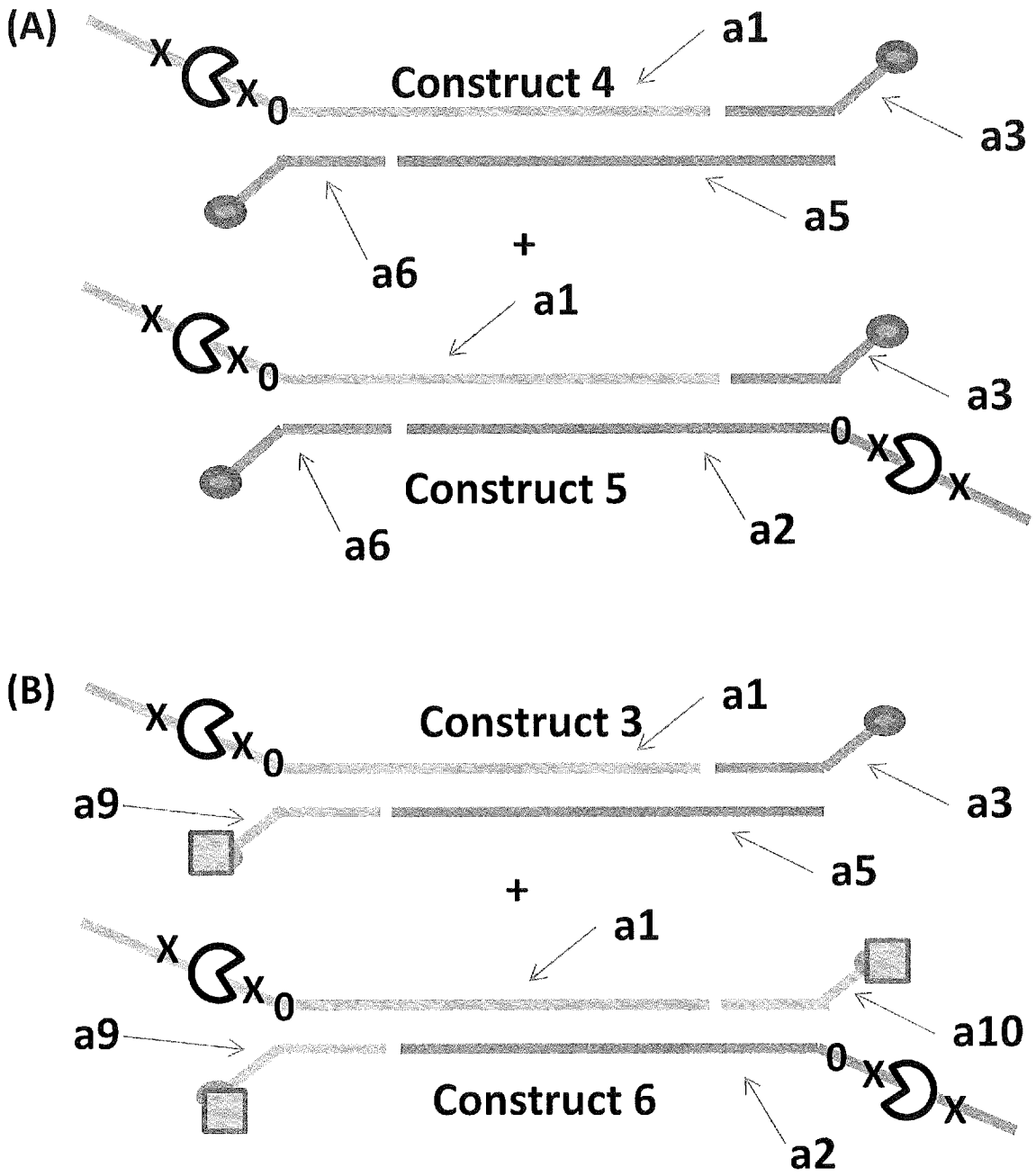

(A)
X
X0
Construct 4
a1
a3
a5
a6
+
a1
X
X0
a3
0
X
X
Construct 5
a6
a2
(B)
X
X0
Construct 3
a1
a3
a9
a5
+
a1
X
X0
a10
a9
0
X
X
Construct 6
a2

Figure 3 3/18

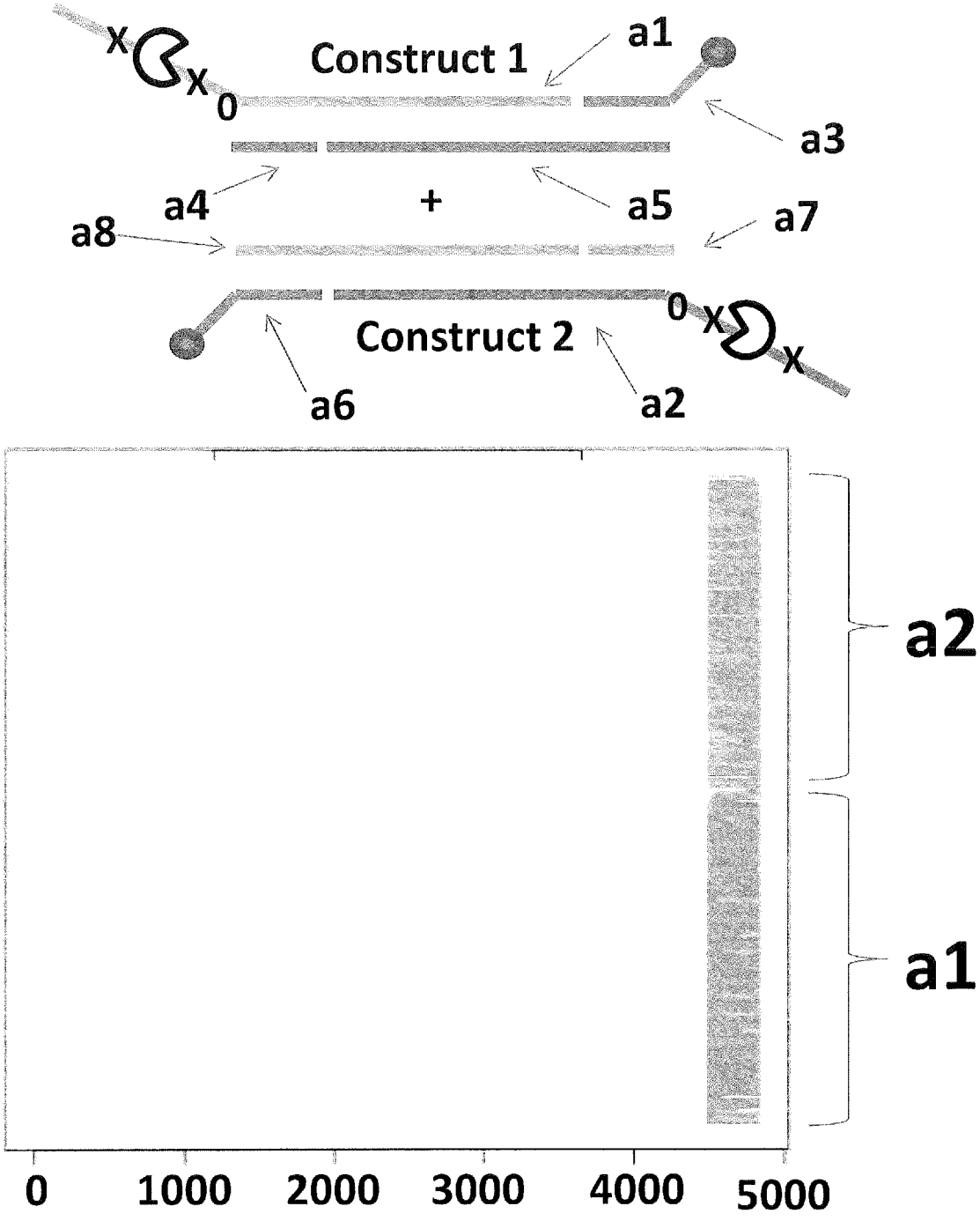
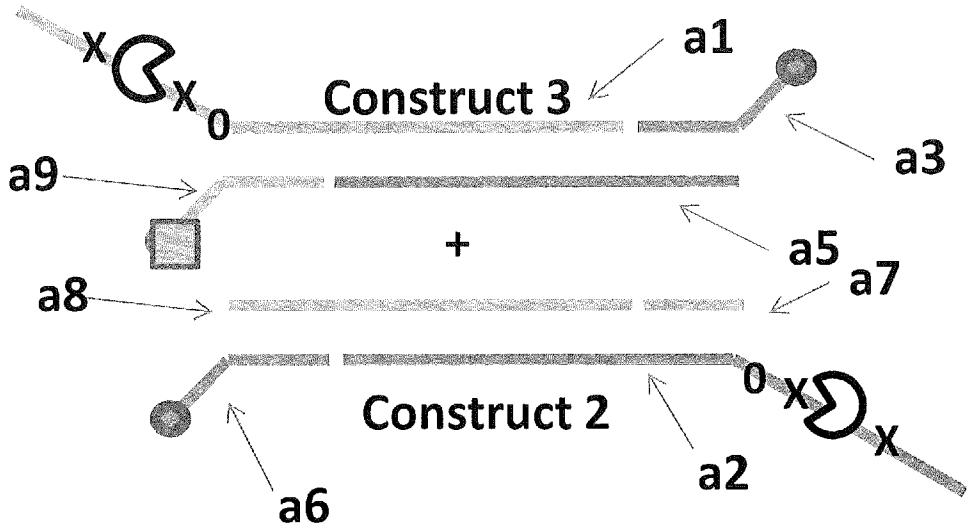

X
X
0
Construct 1
a1
a3
a4
+
a5
a8
a7
0
X
X
Construct 2
a6
a2
a2
a1
0
1000
2000
3000
4000
5000

Figure 4 4/18


x
x
0
Construct 3
a1
a3
a9
+
a5
a7
a8
Construct 2
0
x
x
a2
a6

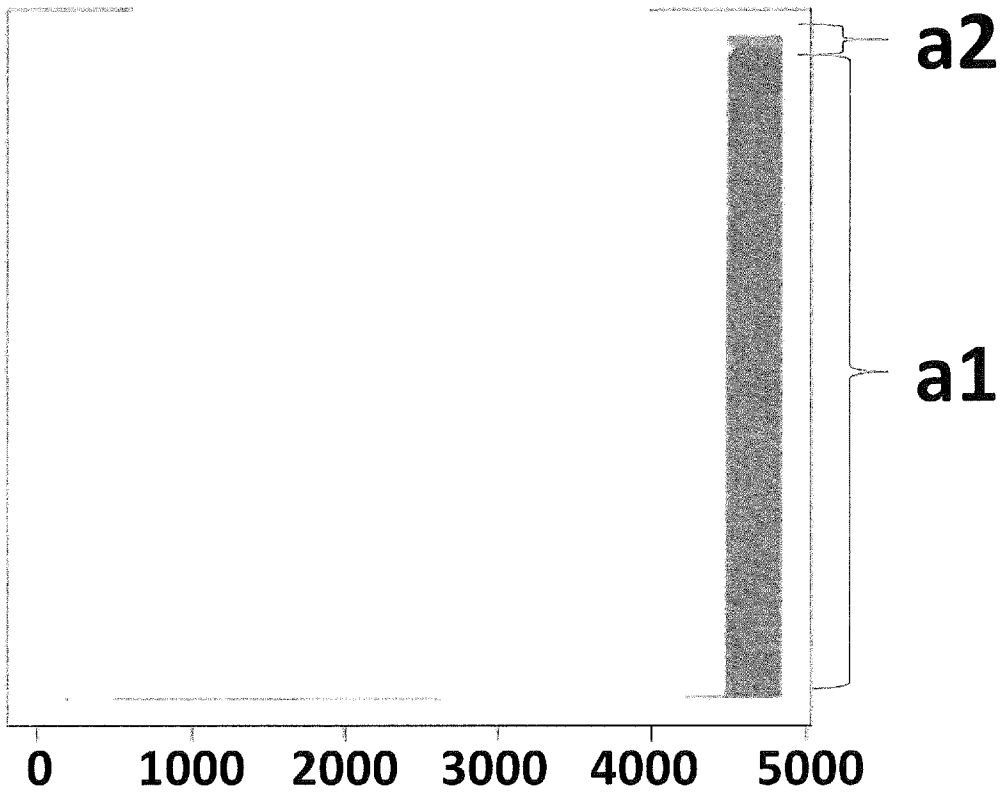

a2
a1
0
1000
2000
3000
4000
5000

Figure 5 5/18

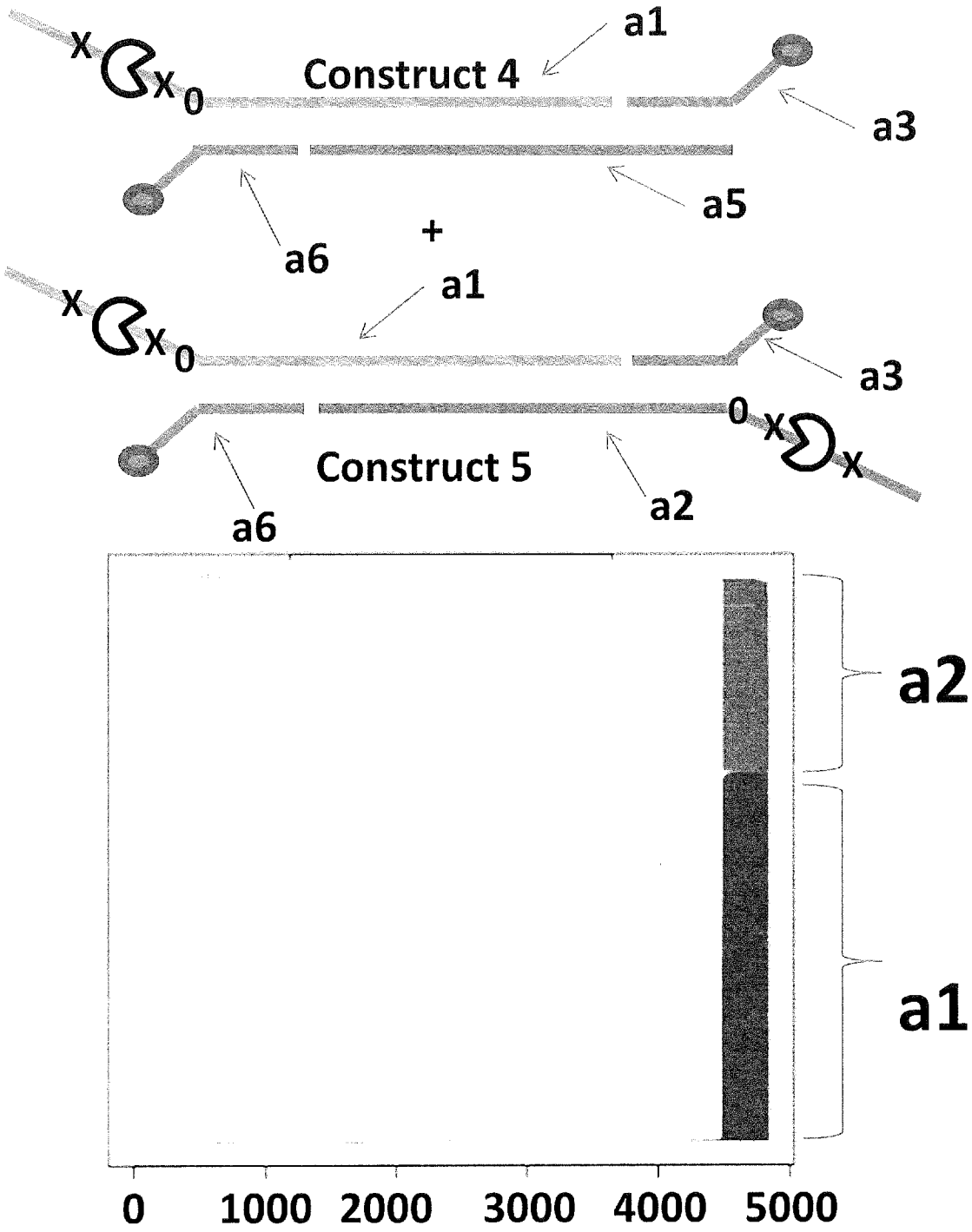

a1
Construct 4
a3
a5
a6
+
a1
a3
Construct 5
a2
a6
a2
a1
0
1000
2000
3000
4000
5000

Figure 6 6/18

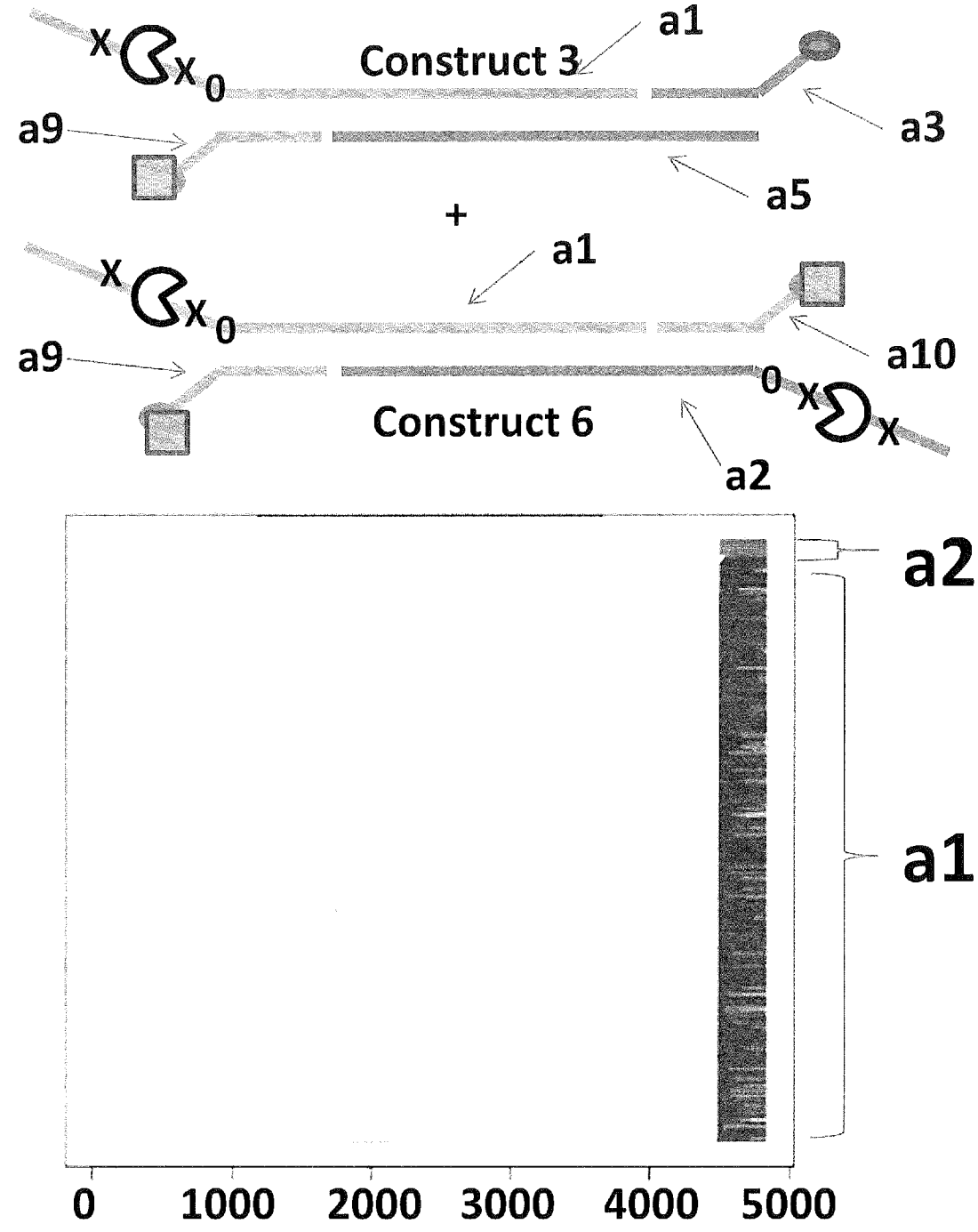

X
X0
Construct 3
a1
a9
a3
a5
+
X
X0
a1
a9
a10
0
X
X
Construct 6
a2
a2
a1
0
1000
2000
3000
4000
5000

Figure 7 7/18

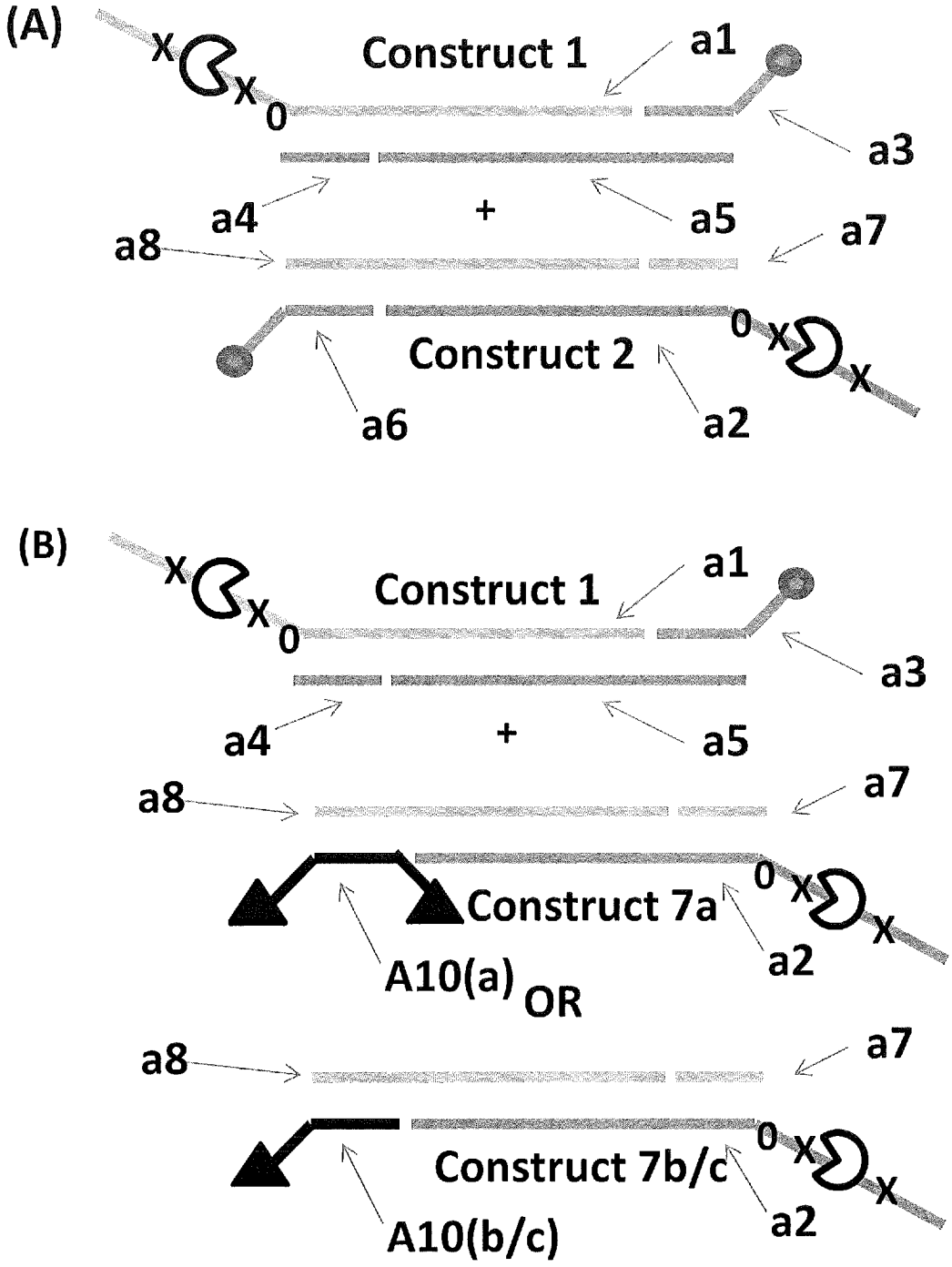
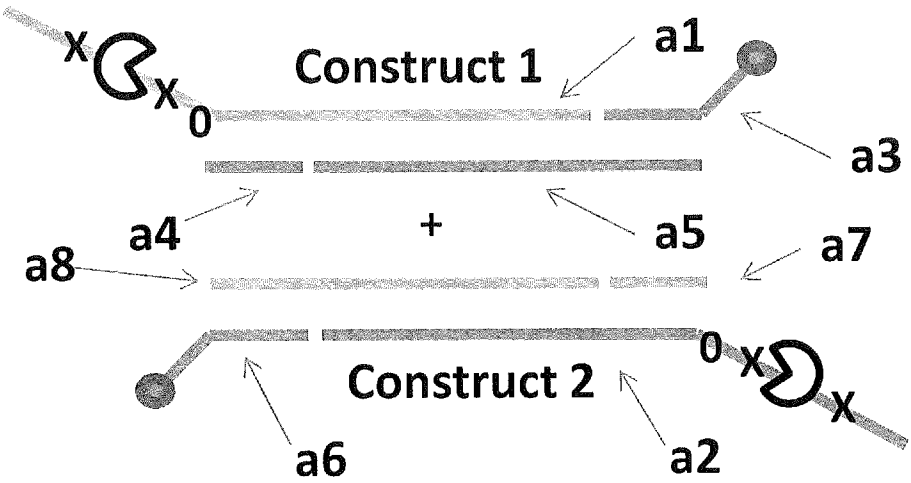

(A)
X
X
0
Construct 1
a1
a3
a4
+
a5
a7
a8
Construct 2
0
X
X
a6
a2
(B)
X
X
0
Construct 1
a1
a3
a4
+
a5
a7
a8
Construct 7a
0
X
X
A10(a)
OR
a2
a8
a7
Construct 7b/c
0
X
X
A10(b/c)
a2

Figure 8 8/18


X
X
0
Construct 1
a1
a3
a4
+
a5
a7
a8
0
X
X
Construct 2
a6
a2

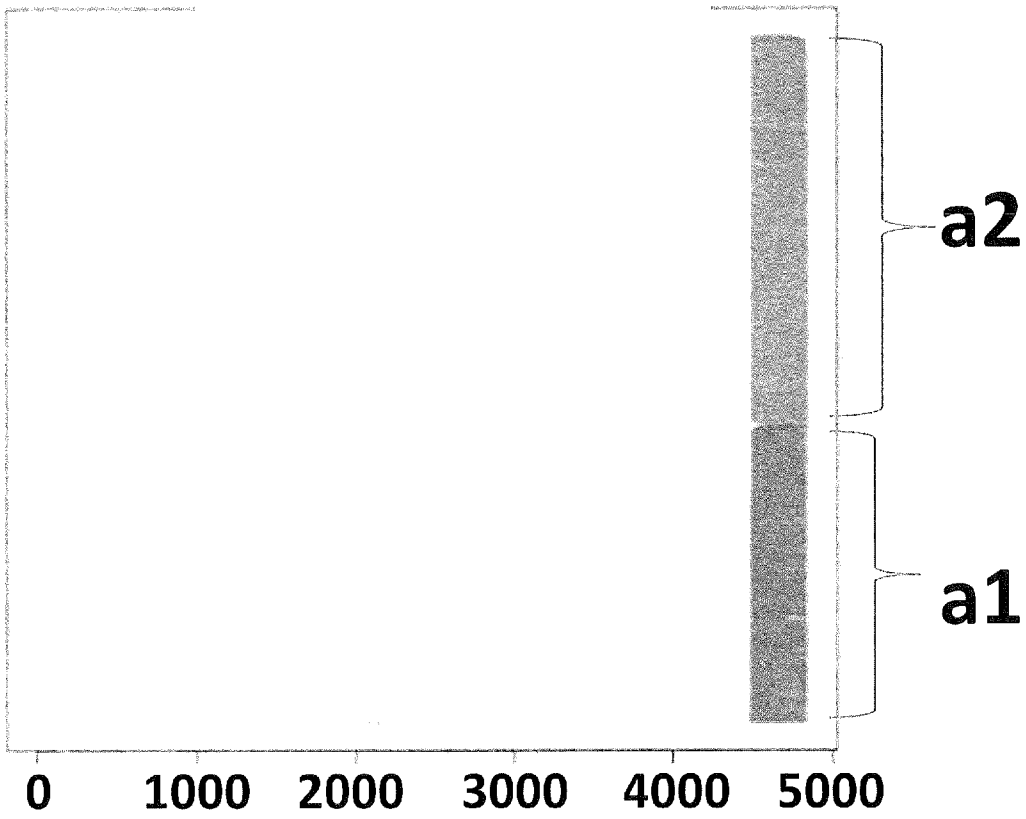
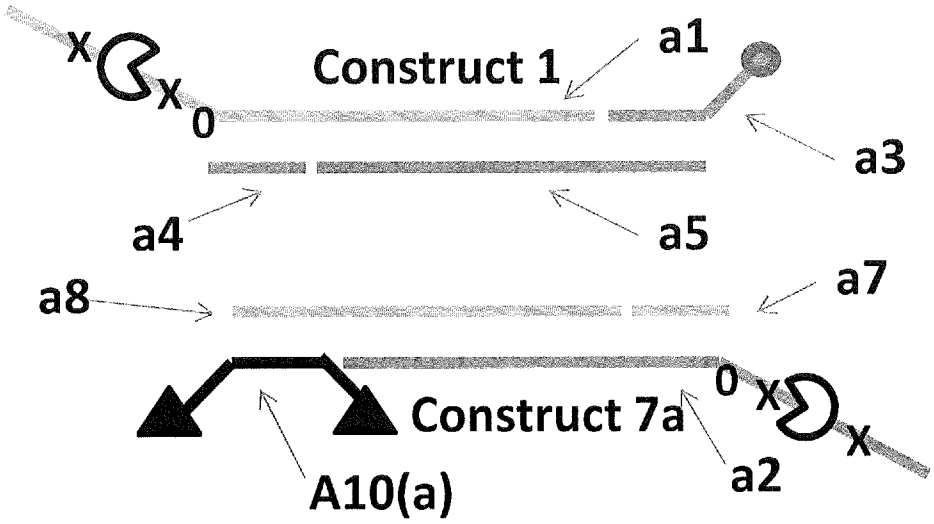

a2
a1
0
1000
2000
3000
4000
5000

Figure 9 9/18


X
X
0
Construct 1
a1
a3
a4
a5
a8
a7
Construct 7a
0
X
X
A10(a)
a2

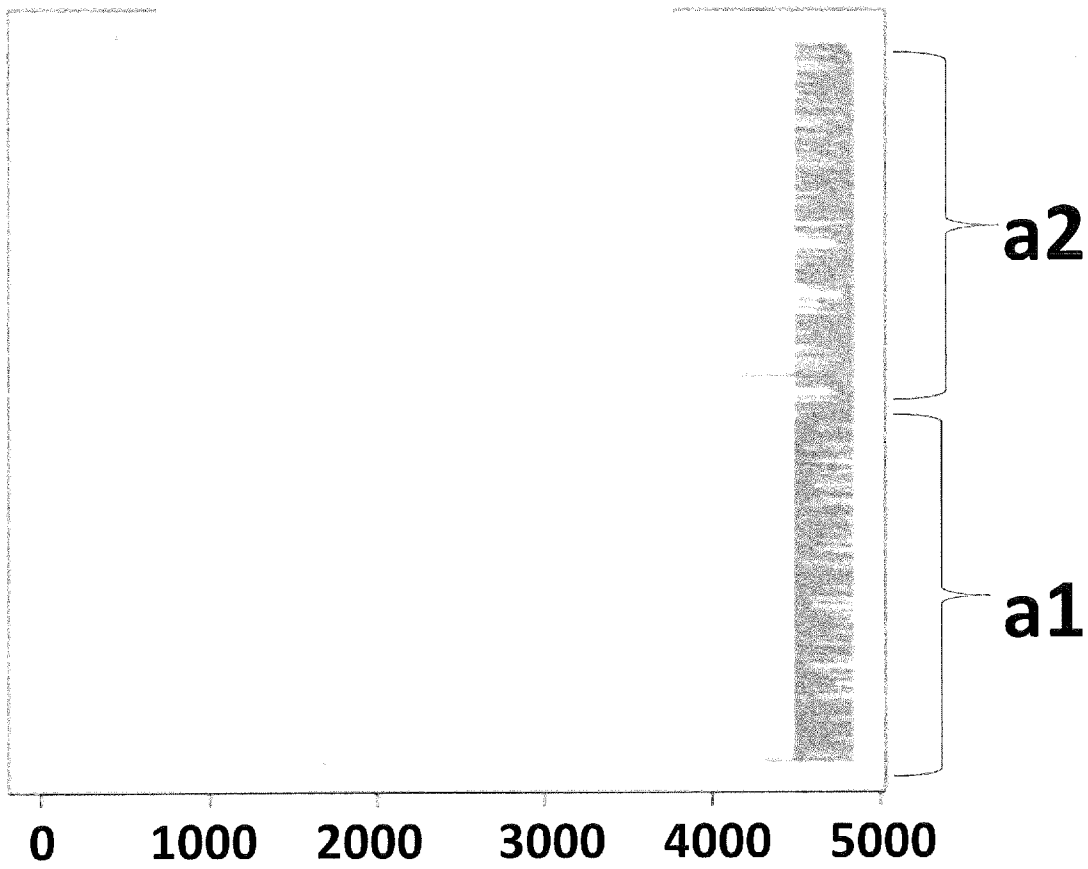

a2
a1
0
1000
2000
3000
4000
5000

Figure 10 10/18

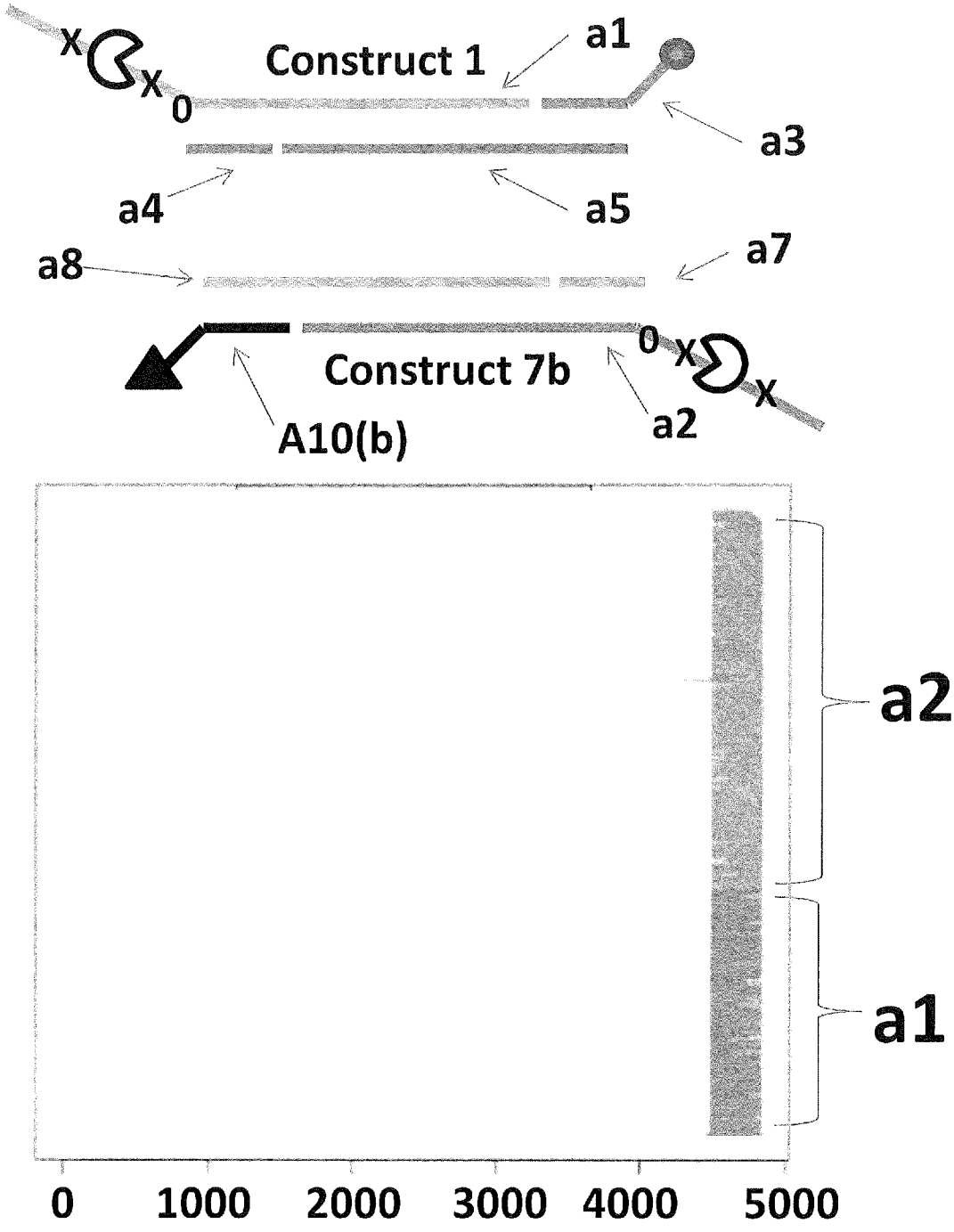


Figure 11 11/18

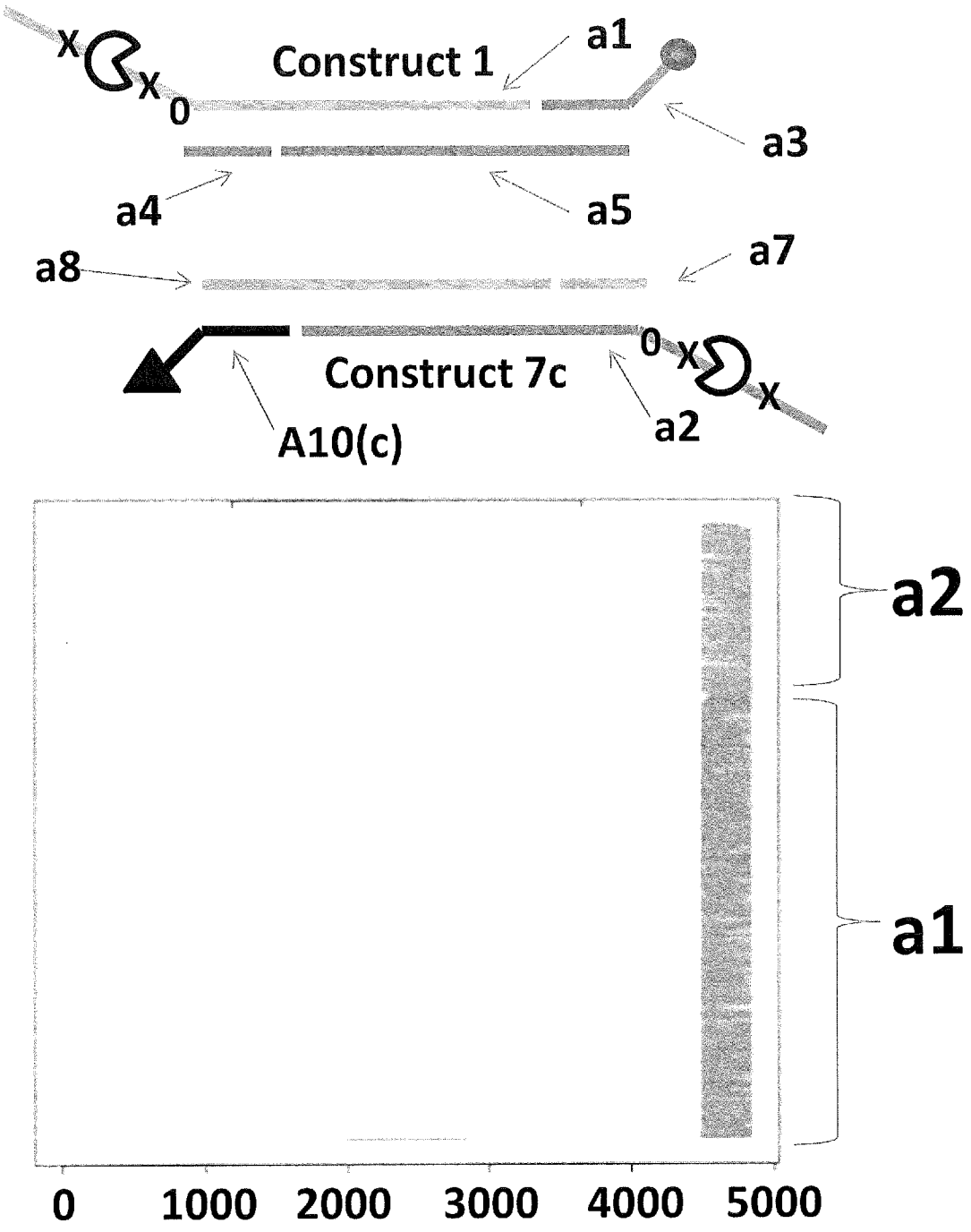


Figure 12 12/18

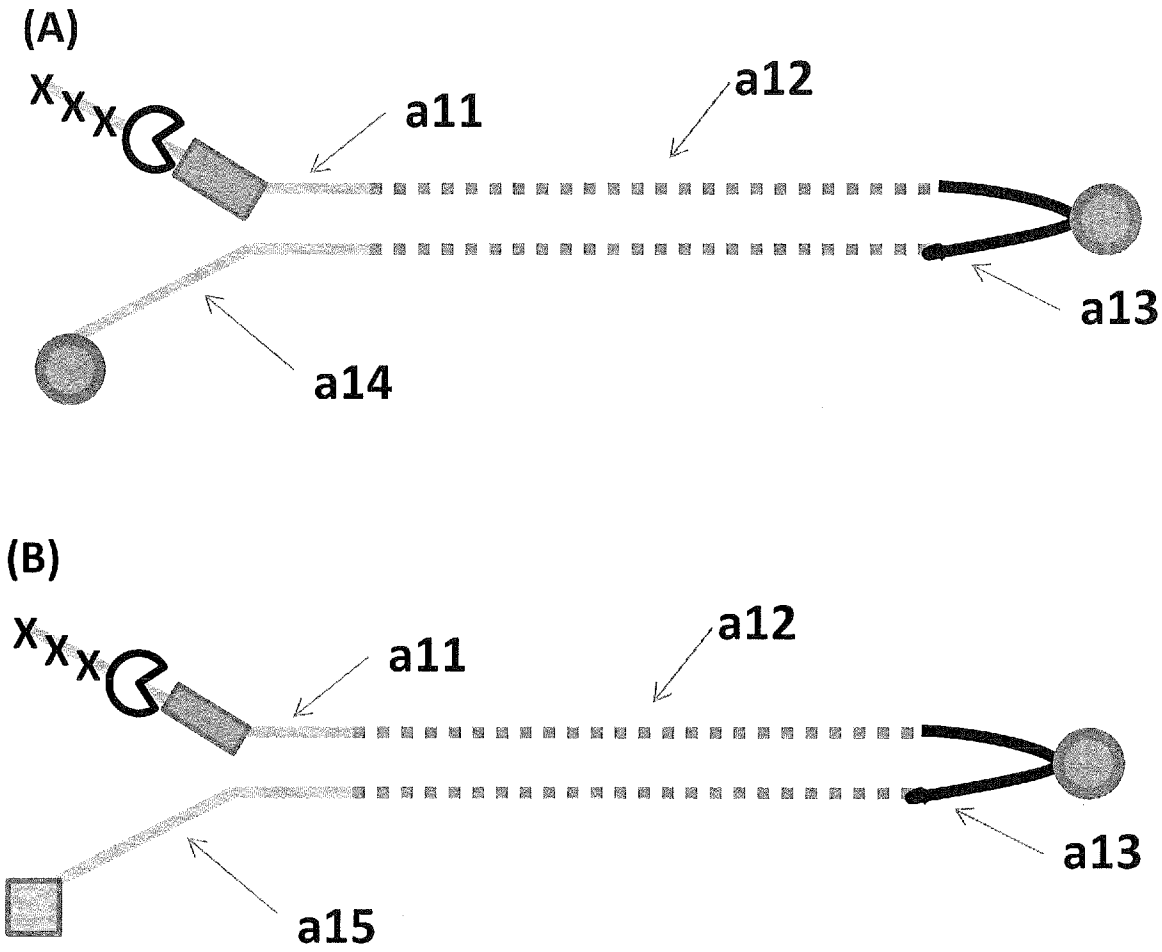


Figure 13 13/18

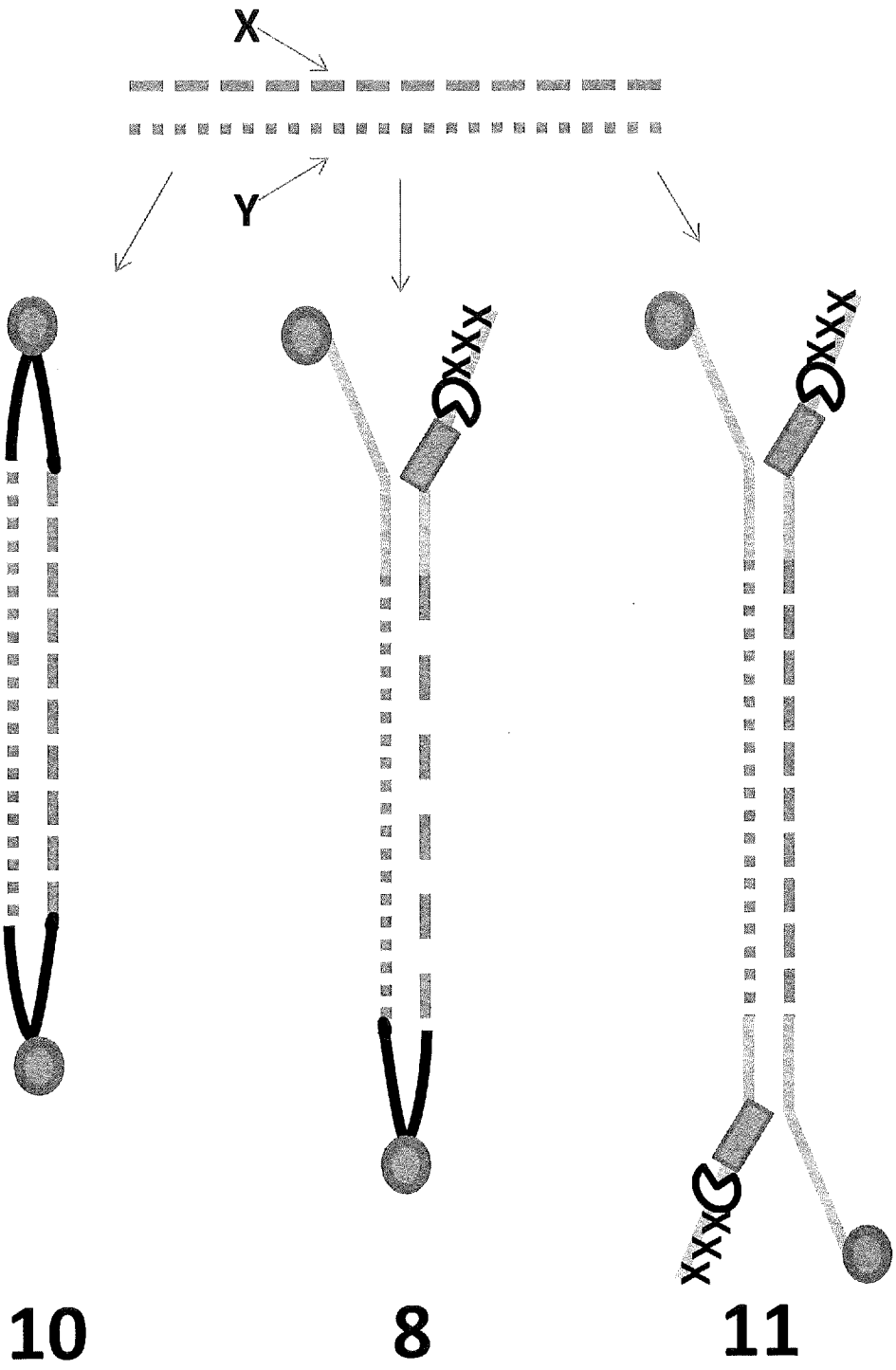


Figure 14

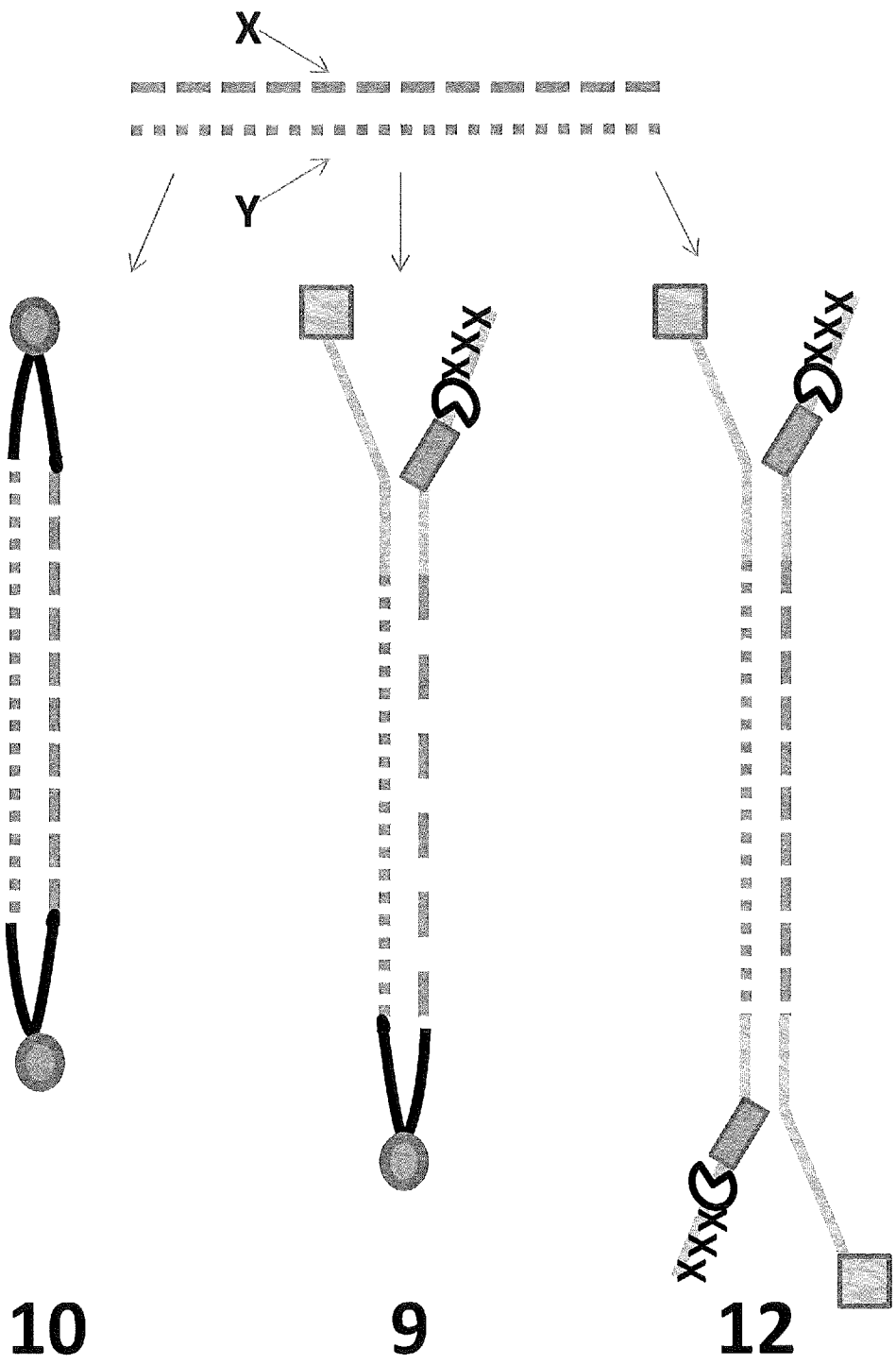


Figure 15

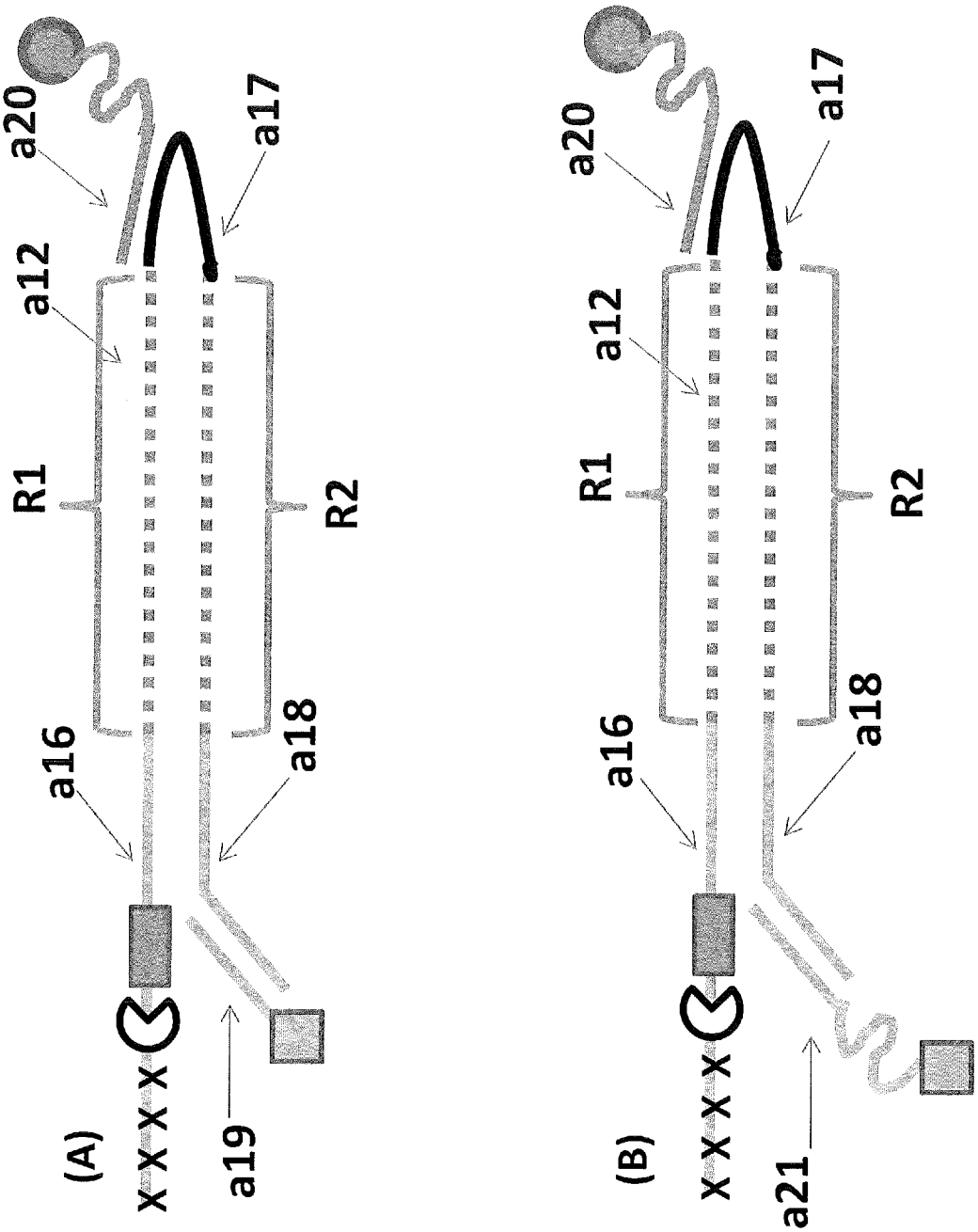

a20
a17
a12
R1
R2
a16
a18
(A)
X X X X
a19
a20
a17
a12
R1
R2
a16
a18
(B)
X X X X
a21

Figure 16

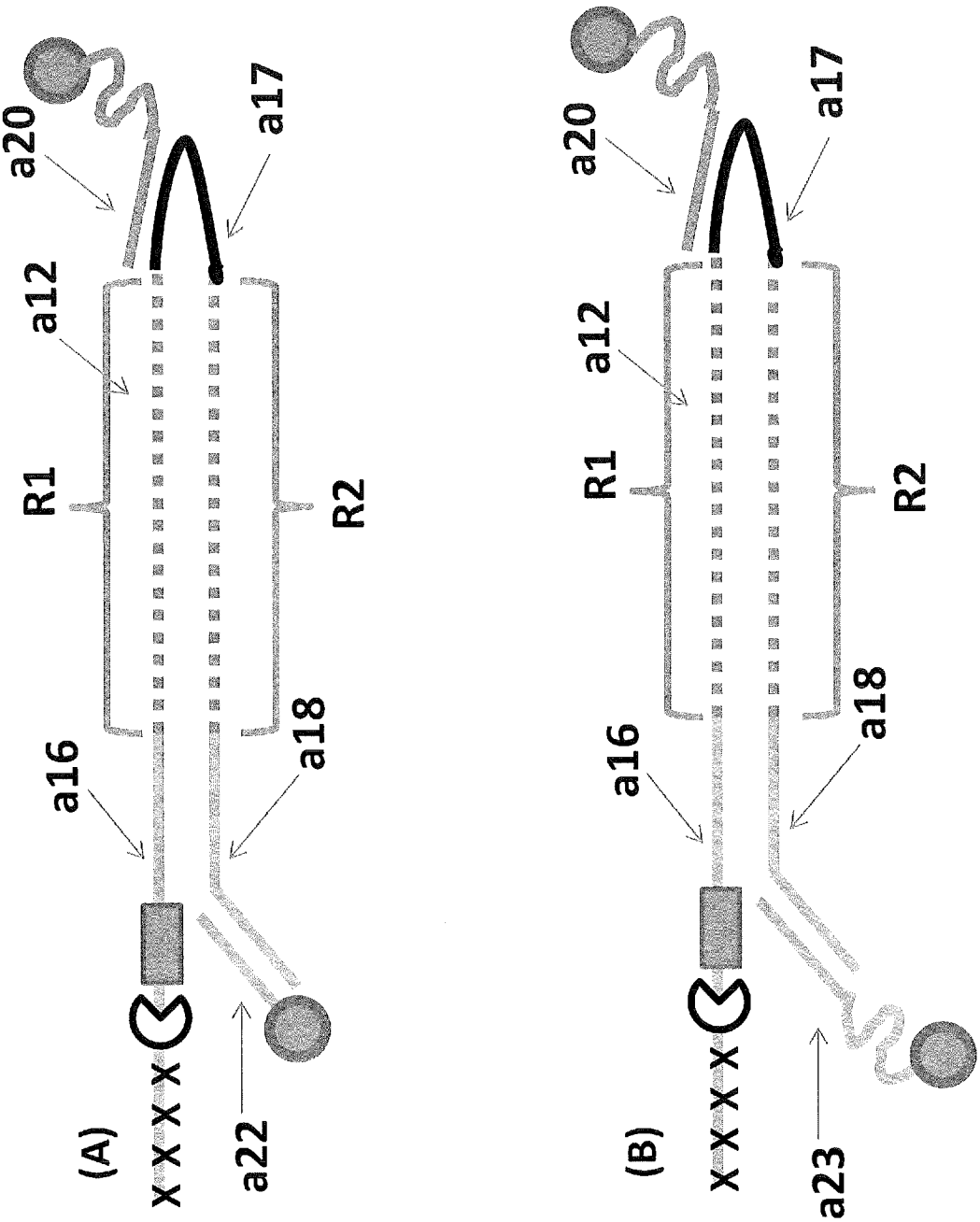

a20
a17
a12
R1
R2
a18
a16
(A)
X X X X
a22
a20
a17
a12
R1
R2
a18
a16
(B)
X X X X
a23

Figure 17 17/18

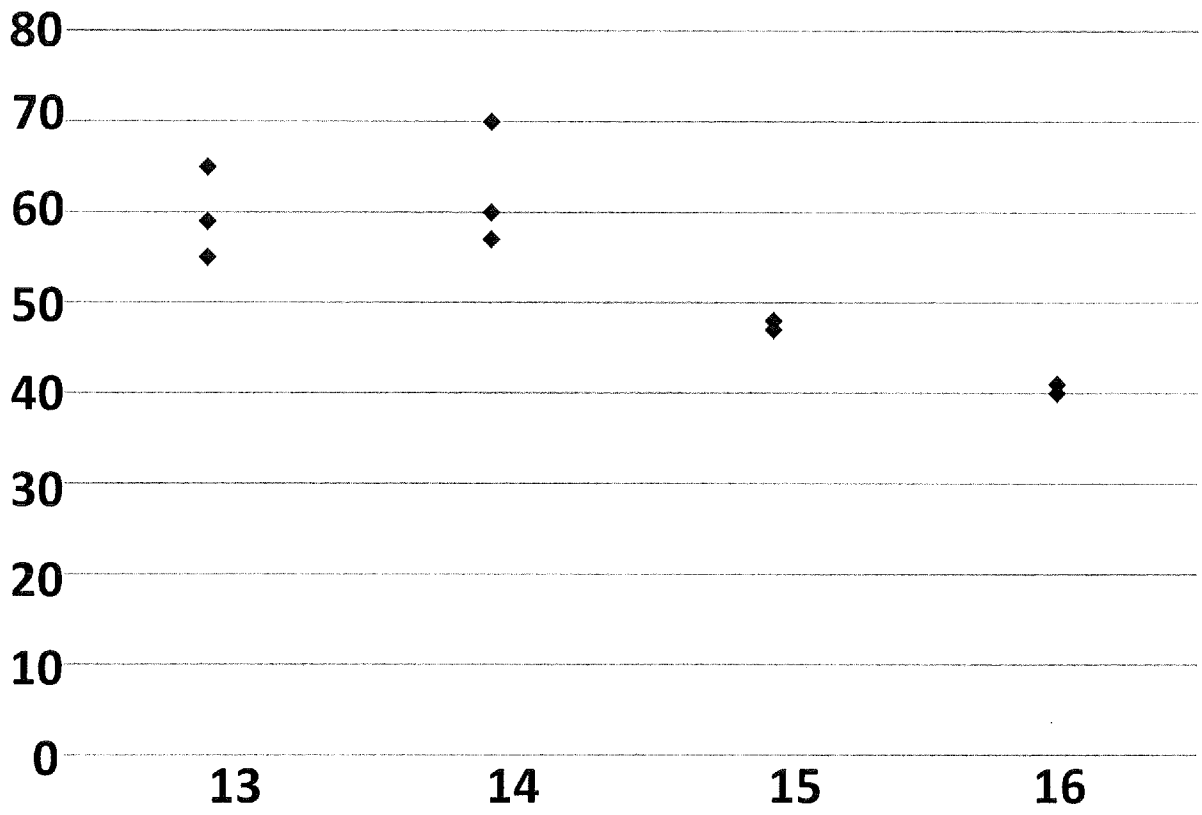
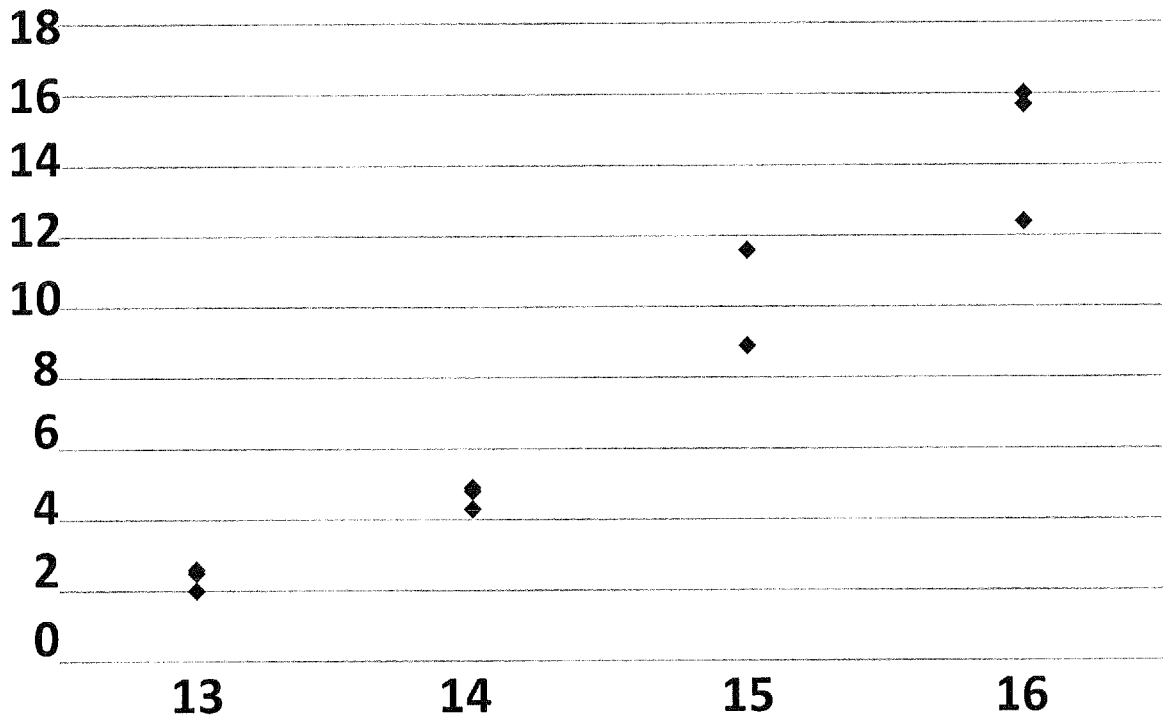


Figure 18 18/18



INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/050991

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

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, CHEM ABS Data, EMBASE, FSTA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2013/014451 A1 (OXFORD NANOPORE TECH LTD [GB]; BROWN CLIVE [GB]; CLARKE JAMES [GB]; HA) 31 January 2013 (2013-01-31) Fig. 9, 10, 17 -----	1-29
A	JUNGKON KIM ET AL: "Elucidation of bacterial genome complexity using next-generation sequencing", BIOTECHNOLOGY AND BIOPROCESS ENGINEERING, THE KOREAN SOCIETY FOR BIOTECHNOLOGY AND BIOENGINEERING, HEIDELBERG, vol. 17, no. 5, 24 October 2012 (2012-10-24), pages 887-899, XP035128702, ISSN: 1976-3816, DOI: 10.1007/S12257-012-0374-X Fig. 2 -----	1-29

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

☒ See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 May 2015	15/06/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Leber, Thomas

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2015/050991

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013014451 A1	31-01-2013	AU 2012288629 A1	06-03-2014
		CA 2843136 A1	31-01-2013
		CN 103827320 A	28-05-2014
		EP 2737084 A1	04-06-2014
		JP 2014531196 A	27-11-2014
		KR 20140050067 A	28-04-2014
		WO 2013014451 A1	31-01-2013