

ENZYME STALLING METHOD**Field of the invention**

The invention relates to new methods of moving helicases past spacers on
5 polynucleotides and controlling the loading of helicases on polynucleotides. The invention also
relates to new methods of characterising target polynucleotides using helicases.

Background of the invention

There is currently a need for rapid and cheap polynucleotide (e.g. DNA or RNA)
10 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
15 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
20 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
nucleotide handling protein, such as a helicase, to control the movement of the polynucleotide
through the pore.

Summary of the invention

Spacers in polynucleotides are typically capable of stalling helicases, i.e. preventing
helicases from moving further along the polynucleotides past the spacers. The inventors have
surprisingly demonstrated that it is possible to move one or more stalled helicases past a spacer
in a polynucleotide by contacting the helicase and polynucleotide with a transmembrane pore
30 and applying a potential. Since the helicase is typically too large to fit through the pore, the
force of the polynucleotide moving through the pore along the potential moves the helicase past
the spacer. This has important applications for controlling the movement of polynucleotides and
characterising, such as sequencing, polynucleotides. The inventors have also surprisingly
demonstrated that it is possible to control the loading of one or more helicases on a
35 polynucleotide using one or more spacers.

The invention therefore provides a method of moving one or more stalled helicases past one or more spacers in a target polynucleotide, comprising contacting (a) the one or more stalled helicases and the target polynucleotide with a transmembrane pore and (b) applying a potential across the pore and thereby moving the one or more helicases past the one or more spacers on the target polynucleotide.

The invention also provides:

- a method of controlling the movement of a target polynucleotide through a transmembrane pore, comprising (a) providing the target polynucleotide with one or more spacers; (b) contacting the target polynucleotide with one or more helicases such that the one or more helicases stall at the one or more spacers; (c) contacting the target polynucleotide and the one or more stalled helicases with the pore; and (d) applying a potential across the pore such that the one or more helicases move past the one or more spacers and control the movement of the target polynucleotide through the pore;
- a method of characterising a target polynucleotide, comprising (a) carrying out the method of controlling the movement of a target polynucleotide through a transmembrane pore of the invention; and (b) taking one or more measurements as the polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the polynucleotide and thereby characterising the target polynucleotide;
- a method of controlling the loading of one or more helicases on a target polynucleotide, comprising (a) providing the polynucleotide with one or more spacers; and (b) contacting the polynucleotide provided in (a) with the one or more helicases such that the one or more helicases bind to the polynucleotide and stall at each spacer;
- an adaptor for controlling the movement of a target polynucleotide, wherein the adaptor comprises (a) (L-S-D)_n or (D-S-L)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer and D is a double stranded polynucleotide and wherein n is a whole number and (b) one or more helicases stalled on each adaptor; and
- a kit for controlling the movement of a target polynucleotide, wherein the kit comprises (a) one or more spacers, (b) one or more helicases and (c) a transmembrane pore.

Description of the Figures

Figure 1 shows a diagram of the lambda DNA construct used in Example 1. SEQ ID NO: 9 (labelled A) is attached at its 3' end to four iSpC3 spacers (labelled B). The four iSpC3 spacers are attached to the 5' end of SEQ ID NO: 10 (labelled C). SEQ ID NO: 10 is attached to four

iSpC3 spacers (labelled D) which are attached to SEQ ID NO: 11 (labelled E) at its 5' end. SEQ ID NO: 10 is hybridised to SEQ ID NO: 12 (labelled F, which has a 3' cholesterol tether).

Figure. 2 shows an example current trace (y-axis label = Current (pA, 20 to 120), x-axis label = Time (s, 3500 to 8000)) of when a helicase (T4 Dda – E94C/A360C (SEQ ID NO: 8 with mutations E94C/A360C and then (Δ M1)G1G2))) controls the translocation of the Lambda DNA construct (0.2 nM as described and illustrated in Figure 1) through a nanopore (MS(B1-G75S/G77S/L88N/Q126R)8 MspA (MspA - B2C) (SEQ ID NO: 2 with mutations G75S/G77S/L88N/Q126R)).

Figure. 3 shows zoomed in regions of the helicase-controlled DNA movement shown in the current trace in Figure 2 (y-axis label = Current (pA, upper trace 20 to 80, lower trace 20 to 60), x-axis label = Time (s, upper trace 2995 to 3020, lower trace 8140 to 8170) for both the upper and lower traces). A) shows the beginning of the helicase-controlled DNA movement and B) shows the end of the helicase controlled DNA movement. The arrow labelled 1 corresponds to when the first four iSpC3 spacers (which is used to stall the movement of the enzyme prior to capture by the nanopore) moves through the nanopore. The arrow labelled 2 corresponds to a second group of four iSpC3 spacers moving through the nanopore.

Figure 4 (a) shows the hairpin and Y-shaped MuA substrate designs used in Example 2. The dUMP in SEQ ID NO: 19 is highlighted as a triangle and the iSpC3 spacers are shown as x's. Figure 4 (b) shows the Lambda DNA construct produced during the sample preparation procedure detailed in Example 2. The 5-10 kB fragment of Lambda DNA is labelled X, the fragment of DNA filled in by the polymerase and joined to the rest of the construct by the ligase is labelled y (and is shown as a dotted line) and the iSpC3 spacers are shown as x's. A tether sequence (SEQ ID NO: 16) is hybridised to the DNA construct as shown. Attached to the 3' end of SEQ ID NO: 16 is six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG (shown as a grey circle).

Figure 5 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for both upper and lower traces) of when a helicase (Trwc Cba (SEQ ID NO: 9) controls the translocation of the Lambda DNA construct (shown in Figure 4b) through a nanopore (MS(B1-G75S/G77S/L88N/Q126R)8 MspA (SEQ ID NO: 2 with mutations G75S/G77S/L88N/Q126R)). The upper trace shows helicase controlled DNA movement of the entire lambda DNA construct through the nanopore, the first iSpC3 spacer labelled X1 produces the spike in current labelled 1 and the second iSpC3 spacers labelled X2 produces the spike in current labelled 2. The lower trace shows a zoomed in region of the end of the helicase controlled DNA movement through the nanopore, the third iSpC3 spacer labelled X3 produces the spike in current labelled 3.

Figure 6 (a) shows the hairpin and Y-shaped MuA substrates designs used in Example 3. The 5' phosphate is labelled as a circle, the inosines in SEQ ID NO: 18 are highlighted as a rectangle and the iSpC3 spacers are shown as x's. Figure 6 (b) shows the Lambda DNA construct produced during the sample preparation procedure detailed in Example 3. The 5-10 kB fragment of Lambda DNA is labelled X, the inosines which have now been attached to x are labelled as a rectangle and the iSpC3 spacers are shown as x's. A tether sequence (SEQ ID NO: 16) is hybridised to the DNA construct as shown. Attached to the 3' end of SEQ ID NO: 16 is six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG (shown as a grey circle).

Figure 7 shows example current traces (y-axis label = Current (pA), x-axis label = Time (s) for both upper and lower traces) of when a helicase (Trwc Cba (SEQ ID NO: 17) controls the translocation of the Lambda DNA construct (shown in Figure 6b) through a nanopore (MS(B1-G75S/G77S/L88N/Q126R)8 MspA (SEQ ID NO: 2 with mutations G75S/G77S/L88N/Q126R)). The upper trace shows helicase controlled DNA movement of the entire lambda DNA construct through the nanopore, the first iSpC3 spacer labelled X1 produces the spike in current labelled 1, the second iSpC3 spacer labelled X2 produces the spike in current labelled 2 and the third iSpC3 spacer labelled X3 produces the spike in current labelled 3. The lower trace shows a zoomed in region of the second half of the helicase controlled DNA movement through the nanopore, the second iSpC3 spacer labelled X2 produces the spike in current labelled 2 and the third iSpC3 spacer labelled X3 produces the spike in current labelled 3.

Figure 8 Fluorescence assay for testing enzyme activity. A custom fluorescent substrate was used to assay the ability of the helicase (labelled a) to displace hybridised dsDNA. 1) The fluorescent substrate strand (48.75 nM final, SEQ ID NO: 25 and 26) has a 3' ssDNA overhang (20 bases), and a 40 base section of hybridised dsDNA. The upper strand (b) has a carboxyfluorescein base (c) at the 5' end of SEQ ID NO: 25, and the hybridised complement (d) has a black-hole quencher (BHQ-1) base (e) at the 3' end of SEQ ID NO: 26. When hybridised, the fluorescence from the fluorescein is quenched by the local BHQ-1, and the substrate is essentially non-fluorescent. 0.975 μ M of a capture strand (f, SEQ ID NO: 27) that is part-complementary to the lower strand of the fluorescent substrate is included in the assay. 2) In the presence of ATP (0.975 mM) and $MgCl_2$ (10 mM), helicase Hel308 Mbu (12 nM, SEQ ID NO: 28) binds to the 3' overhang of the fluorescent substrate, moves along the upper strand, and displaces the complementary strand (d) as shown. 3) Once the complementary strand with BHQ-1 is fully displaced the fluorescein on the major strand fluoresces (shown as a star shape). 4) The displaced lower strand (d) preferentially anneals to an excess of capture strand (f) to prevent re-annealing of initial substrate and loss of fluorescence.

Figure 9 Fluorescence assay for testing enzyme activity. Two possible fluorescent substrates were used to assay the ability of the helicase (labelled a) to displace hybridised dsDNA, these are shown in Figure 9 (a) (labelled 1C-3C) has four Sp9 spacers (shown as a black triangle) connecting the 3' end of SEQ ID NO: 27 to the 3' end of SEQ ID NO: 29 and (b) (labelled 1B-3B) has one Sp9 spacer (shown as a black triangle) connecting the 3' end of SEQ ID NO: 27 to the 3' end of SEQ ID NO: 29. The fluorescent substrate strand (48.75 nM final either a) or b) described previously) has a 3' ssDNA overhang (20 bases), and a 40 base section of hybridised dsDNA. The upper strand (b) has a carboxyfluorescein base (c) at the 5' end of SEQ ID NO: 27, and the hybridised complement (d) has a black-hole quencher (BHQ-1) base (e) at the 3' end of SEQ ID NO: 26. When hybridised, the fluorescence from the fluorescein is quenched by the local BHQ-1, and the substrate is essentially non-fluorescent. 1 μ M of a capture strand (f, SEQ ID NO: 27) that is part-complementary to the lower strand of the fluorescent substrate is included in the assay. 2 a and b) In the presence of ATP (0.975 mM) and $MgCl_2$ (10 mM), helicase Hel308 Mbu (SEQ ID NO: 28) binds to the 3' overhang of the fluorescent substrate, moves along the upper strand up to the Sp9 group(s). The sp9 group (1 in B and 4 in C) stops the helicase from moving past it and the helicase does not displace the complementary strand (SEQ ID NO: 26). Therefore, the fluorescein and black hole quencher remain in close proximity to each other and fluorescence is not observed.

Figure 10 Graph (y-axis = Relative dsDNA turnover, x-axis = fluorescent substrate) of the relative dsDNA turnover rate in buffer solutions (100 mM HEPES pH8, 0.975 mM ATP, 10 mM $MgCl_2$, 1 mg/mL BSA, 48.75 nM fluorescent substrate DNA (A = SEQ ID NOs: 25 and 26, B = SEQ ID NO: 27 attached at its 3' end by one Sp9 spacers to the 5' end of SEQ ID NO: 29 and hybridised to SEQ ID NO: 26, C = SEQ ID NO: 27 attached at its 3' end by four Sp9 spacers to the 5' end of SEQ ID NO: 29 and hybridised to SEQ ID NO: 26), 0.975 μ M capture DNA (SEQ ID NO: 27)) for the Hel308 Mbu helicase (labeled A, SEQ ID NO: 28) at 400 mM of KCl.

Figure 11 Graph (y-axis = Relative dsDNA turnover, x-axis = fluorescent substrate) of the relative dsDNA turnover rate in buffer solutions (100 mM HEPES pH8, 0.975 mM ATP, 10 mM $MgCl_2$, 1 mg/mL BSA, 48.75 nM fluorescent substrate DNA (D = SEQ ID NOs: 32 and 26, E = SEQ ID NO: 27 attached at its 3' end by one idSp groups to the 5' end of SEQ ID NO: 30 and hybridised to SEQ ID NO: 26, F = SEQ ID NO: 27 attached at its 3' end by four idSp groups to the 5' end of SEQ ID NO: 31 and hybridised to SEQ ID NO: 26 and G = SEQ ID NO: 33 hybridised to SEQ ID NO: 26), 0.975 μ M capture DNA (SEQ ID NO: 27)) for the Hel308 Mbu helicase (labeled A, SEQ ID NO: 28) at 400 mM of KCl.

Figure 12 shows the experimental steps for the control strand which does not contain iSpC3 or iSp18 spacers which stall the helicase (labelled 1). The control strand (SEQ ID NO: 34)

contains no spacers or blocking groups and is hybridised to a shorter complementary strand of DNA (SEQ ID NO: 35, which has a carboxyfluorescein attached to its 5' end, shown as a grey circle). This produces a partially double stranded construct which has a 50 nucleotide overhang. The construct (SEQ ID NO: 34 hybridised to SEQ ID NO: 35) is pre incubated with T4 Dda –
 5 E94C/A360C to allow the enzyme to bind to the overhang. Owing to the length of the overhang more than one enzyme can bind to it as shown. The enzyme is then provided with the necessary components to promote helicase movement (ATP and MgCl₂). Additional capture strand (SEQ ID NO: 37) is also added with the ATP and MgCl₂. The helicase then translocates along the control strand, displacing the shorter complementary strand, leaving the control strand with no
 10 helicase or complementary strand bound (labelled A). The complementary strand then forms a hairpin so that it cannot re-anneal to the control strand (labelled B). Helicase which is free in solution or has moved along the control DNA and fallen off at the end is then bound by the excess of capture strand (SEQ ID NO: 37, complex labelled C) as shown. Any DNA which does not bind a helicase remains intact (remains as species F). The sample mixture is then run on a gel
 15 and the separate species identified by the bands they produce on the gel.

Figure 13 shows the experimental steps for a strand which contains iSpC3 or iSp18 spacers (shown as X's in the figure) in order to stall the helicase (labelled 1). The same procedure was carried out as shown in Figure 12, the enzyme was pre-incubated with the DNA construct (the shorter complementary strand has a carboxyfluorescein attached to its 5' end). The
 20 spacer groups are located in front of the double stranded region and the helicase is capable of binding to the DNA as shown. Upon the addition of ATP, MgCl₂ and capture DNA (labelled 2) the helicase is then provided with the necessary components to promote helicase movement. If the spacers are capable of stalling the helicase then it will remain bound to the DNA construct which still contains the short complementary strand (labelled D). If the spacers are not capable of
 25 stalling the helicase then it will move past the spacer and displace the complementary strand (labelled B) leaving species A with no helicase bound. The free enzyme will then bind to the excess capture strand (labelled C) and the displaced complementary strand will form a hairpin (labelled B). If the spacers are able to stall one helicase but not two helicases then the first helicase will be pushed past the spacers by the helicase behind and will displace the
 30 complementary stand. However, the second helicase will not be able to move past the spacers resulting in the complex labelled E. Any DNA which does not bind a helicase remains intact with no helicase bound (species F) The sample mixture is then run on a gel and the separate species identified by the bands they produce on the gel.

Figure 14 shows the gel assay which was run for the control strand (1 in table 11). The
 35 lane labelled M shows a DNA ladder for reference (bands correspond from lowest mass (bottom

of the gel) to highest mass (top of the gel) 200 bp (base pairs), 300 bp, 400 bp, 500/517 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp, 1200 bp and 1517 bp). Lane 1 contains annealed DNA only (SEQ ID NO: 34 hybridised to SEQ ID NO: 35). Lane 2 contains the helicase (T4 Dda – E94C/A360C) pre-bound to control strand (no fuel added). Lane 3 shows the control strand after fuel (ATP and MgCl₂) has been added in buffer 1. Lane 4 shows the control strand after fuel (ATP and MgCl₂) has been added in buffer 2. Band X corresponds to SEQ ID NO: 34 only and band Y corresponds to SEQ ID NO: 34 hybridised to SEQ ID NO: 35. The region labelled 1Y corresponds to one helicase bound to SEQ ID NO: 34 hybridised to SEQ ID NO: 35. The region labelled 2Y corresponds to two helicases bound to SEQ ID NO: 34 hybridised to SEQ ID NO: 35. The region labelled 3Y corresponds to three helicases bound to SEQ ID NO: 34 hybridised to SEQ ID NO: 35. The region labelled 4Y corresponds to four helicases bound to SEQ ID NO: 34 hybridised to SEQ ID NO: 35. The region labelled 5Y corresponds to five helicases bound to SEQ ID NO: 34 hybridised to SEQ ID NO: 35.

Figure 15 shows the gel assay which was run for the DNA constructs containing 3 (7 in table 11, corresponding lanes = 1-4), 4 (8 in table 11, corresponding lanes = 5-8) and 5 iSp18 spacers (9 in table 11, corresponding lanes = 9-12) at the junction between ssDNA and dsDNA. The lanes labelled M show a DNA ladder for reference (bands correspond from lowest mass (bottom of the gel) to highest mass (top of the gel) 200 bp (base pairs), 300 bp, 400 bp, 500/517 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp, 1200 bp and 1517 bp). Lane 1 contains annealed DNA only (SEQ ID NO: 9 attached at its 3' end to three iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). Lane 2 contains the helicase (T4 Dda – E94C/A360C) pre-bound to the DNA construct with no fuel added (SEQ ID NO: 9 attached at its 3' end to three iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). Lane 3 shows the DNA construct (SEQ ID NO: 9 attached at its 3' end to three iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35) after fuel (ATP and MgCl₂) has been added in buffer 1. Lane 4 shows the DNA construct (SEQ ID NO: 9 attached at its 3' end to three iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35) after fuel (ATP and MgCl₂) has been added in buffer 2. Lane 5 contains annealed DNA only (SEQ ID NO: 9 attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). Lane 6 contains the helicase (T4 Dda – E94C/A360C) pre-bound to the DNA construct with no fuel added (SEQ ID NO: 9 attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). Lane 7 shows the DNA construct (SEQ ID NO: 9 attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35) after fuel (ATP and MgCl₂) has been

added in buffer 1. Lane 8 shows the DNA construct (SEQ ID NO: 9 attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35) after fuel (ATP and MgCl₂) has been added in buffer 2. Lane 9 contains annealed DNA only (SEQ ID NO: 9 attached at its 3' end to five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). Lane 10 contains the helicase (T4 Dda – E94C/A360C) pre-bound to the DNA construct with no fuel added (SEQ ID NO: 9 attached at its 3' end to five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). Lane 11 shows the DNA construct (SEQ ID NO: 9 attached at its 3' end to five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35) after fuel (ATP and MgCl₂) has been added in buffer 1. Lane 12 shows the DNA construct (SEQ ID NO: 9 attached at its 3' end to five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35) after fuel (ATP and MgCl₂) has been added in buffer 2. Band X corresponds to ssDNA construct only (e.g. SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36) only and Band Y corresponds to the dsDNA construct only (SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35). The region labelled 1X corresponds to one helicase bound to SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to SEQ ID NO: 36. The region labelled 1Y corresponds to one helicase bound to SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to SEQ ID NO: 36 hybridised to SEQ ID NO: 35. The region labelled 2Y corresponds to two helicases bound to SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to SEQ ID NO: 36 hybridised to SEQ ID NO: 35. The region labelled 3Y corresponds to three helicases bound to SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to SEQ ID NO: 36 hybridised to SEQ ID NO: 35. The region labelled 4Y corresponds to four helicases bound to SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to SEQ ID NO: 36 hybridised to SEQ ID NO: 35. The region labelled 5Y corresponds to five helicases bound to SEQ ID NO: 9 attached at its 3' end to three, four or five iSp18 spacers which are attached to SEQ ID NO: 36 hybridised to SEQ ID NO: 35.

Figure 16 shows the standard DNA construct used in Example 9. The x's represent spacer groups onto which the helicase cannot bind. The length of the region labelled 1 can be altered in order to control the number of helicases which can bind to that region. The figure shows, as an example, the binding of one or two helicases in region 1.

Figure 17 shows an example gel assay of the DNA construct labelled 3 in Table 12 (Five iSpC3 spacers attached to the 5' end of SEQ ID NO: 9, which is attached at its 3' end to four

iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35). The lane labelled M shows a DNA ladder for reference (bands correspond from lowest mass (bottom of the gel) to highest mass (top of the gel) 200 bp (base pairs), 300 bp, 400 bp, 500/517 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp, 1200 bp and 1517 bp). Lanes 1-6 correspond to different concentrations of T4 Dda – E94C/A360C – 1 = 5000 nM, 2 = 2500 nM, 3 = 1250 nM, 4 = 625 nM, 5 = 312.5 nM and 6 = 0 nM). The band observed at level X corresponds to the unbound dsDNA construct. The numbers on the left-hand side of the gel relate to the number of enzymes bound to the DNA. For this DNA construct it was possible to bind up to 6 helicases at the highest concentration of enzyme added. The numbers shown at the top of the gel correspond to the concentration of T4 Dda – E94C/A360C) added.

Figure 18 shows the DNA construct used in example 9. Regions labelled A and B are short strands of DNA (SEQ ID NO: 37) which have been designed so that one helicase can bind to each region (as shown on the right-hand side). Region 1 corresponds to 25 SpC3 spacers, region 2 corresponds to two iSp18 spacers, region 3 corresponds to two iSp18 spacers, region 4 is a section of DNA (SEQ ID NO: 10) which hybridises to another strand of DNA which may or may not be made up of forked DNA (e.g. SEQ ID NO: 42 = non-forked (missing fragment labelled 7) and SEQ ID NO: 12 attached to six iSp18 spacers (shown as fragment labelled 7) = forked), region 5 corresponds to four 5-nitroindoles and region 6 corresponds to another region of DNA (SEQ ID NO: 41) which is hybridised to its complement (SEQ ID NO: 43).

Figure 19 shows a gel assay of the DNA construct described and shown in Figure 18. The band observed at level X corresponds to the unbound dsDNA construct. The lane labelled M shows a DNA ladder for reference (bands correspond from lowest mass (bottom of the gel) to highest mass (top of the gel) 200 bp (base pairs), 300 bp, 400 bp, 500/517 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1000 bp, 1200 bp and 1517 bp). The numbers on the left-hand side of the gel relate to the number of enzymes bound to the DNA. For this DNA construct it was possible to observe binding of two helicases (one at region A and the second at region B as shown in Figure 18) from a concentration of 475 nM enzyme and above. The numbers shown at the top of the gel correspond to the concentration of T4 Dda – E94C/A360C) added.

Figure 20 shows the DNA construct used in Example 10 (referred to as DNA construct X1). There are 25 SpC3 spacers (shown as x's) attached to the 5' end of SEQ ID NO: 38 which is attached at its 3' end to 4 iSp18 spacers (shown as black squares). The four iSp18 spacers are attached to the 5' end of SEQ ID NO: 10 which is attached at its 3' end to four 5-nitroindoles (shown as grey triangles). The four nitroindoles are then attached to the 5' end of SEQ ID NO: 41. The complementary DNA strand which hybridises to SEQ ID NO: 41 is SEQ ID NO: 43.

The complementary DNA strand which hybridises to SEQ ID NO: 10 is SEQ ID NO: 12

(attached to SEQ ID NO: 12's 3' end is six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG (shown as a grey circle)). The region labelled A corresponds to the region of the construct where T4 Dda – E94C/A360C is able to bind.

Figure 21 shows an example current trace (y-axis label = Current (pA, 50 to 250), x-axis label = Time (s, 238 to 252)) of when a helicase (T4 Dda – E94C/A360C) controls the translocation of the DNA construct (0.1 nM, see Figure 20 description) through a nanopore (MspA – B2C). The region labelled 1 shows the helicase controlled translocation of the polyT region (SEQ ID NO: 38, that the enzyme bound onto) through the nanopore. The region labelled 2 corresponds to the helicase controlled translocation of the iSp18 spacers through the nanopore.

Figure 22 shows a diagram of the DNA construct used in Example 4. Two thymines are attached at the 3' end to 28 iSpC3 spacers (labelled A). The 28 iSpC3 spacers are attached at the other end to the 5' end of SEQ ID NO: 23 (sequence corresponds to region B = polyT section and region C = sequence complementary to the tether sequence (SEQ ID NO: 12)). The 3' end of SEQ ID NO: 23 is attached to four iSpC3 spacers (labelled D). The other end of the 4 iSpC3 spacers is attached to the 5' end of SEQ ID NO: 24. The tether sequence (SEQ ID NO: 12) is attached at its 3' end to six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG.

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 amino acid sequence of the helicase Dda 1993 from Enterobacteria phage T4.

SEQ ID NO: 9 shows a polynucleotide sequence used in Example 1, 2, 3, 7 and 8.

SEQ ID NO: 10 shows a polynucleotide sequence used in Example 1 and 9.

SEQ ID NO: 11 shows a polynucleotide sequence used in Example 1. SEQ ID NO: 11 is attached by its 5' end to three iSpC3 spacers which are attached to the 3' end of SEQ ID NO: 10.

SEQ ID NO: 12 shows a polynucleotide sequence used in Examples 1 and 9. In Example 1 SEQ ID NO: 12 is attached at its 3' end to six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG. In Example 9 SEQ ID NO: 12 is attached at its 3' end to six iSp18 spacers only.

5 SEQ ID NO: 13 shows the polynucleotide sequence of the Enterobacteria phage λ . The sequence contains an additional 12 base overhang attached at the 5' end of the sense strand. The sequence shown here is that of the sense strand only.

SEQ ID NO: 14 shows a polynucleotide sequence used in Examples 2 and 3. SEQ ID NO: 14 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units.

10 SEQ ID NO: 15 shows a polynucleotide sequence used in Examples 2 and 3. SEQ ID NO: 15 is attached at its 5' end to the 3' end of SEQ ID NO: 14 by four iSpC3 spacer units.

SEQ ID NO: 16 shows a polynucleotide sequence used in Example 2 and 3 which at the 3' end of the sequence has six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG.

15 SEQ ID NO: 17 shows the amino acid sequence of the Trwc Cba helicase.

SEQ ID NO: 18 shows a polynucleotide sequence used in Example 3. SEQ ID NO: 18 is attached at its 3' end to the 5' end of SEQ ID NO: 9 by four iSpC3 spacer units. This sequence has a phosphate attached to its 5' end and 5 deoxyinosines at positions 1 to 5.

SEQ ID NO: 19 shows a polynucleotide sequence used in Example 2.

20 SEQ ID NO: 20 shows a polynucleotide sequence used in Example 2.

SEQ ID NOs: 21 and 22 are placeholders to maintain the numbering of the following sequences.

SEQ ID NO: 23 shows the polynucleotide sequence used in Example 4. Attached to the 5' end of this sequence is 28 iSpC3 spacers units the last of which has an additional two T's
25 attached to the 5' end of the spacer group. Attached to the 3' end of this sequence is four iSpC3 spacer units which are attached to the 5' end of SEQ ID NO: 24.

SEQ ID NO: 24 shows the polynucleotide sequence used in Example 9. Attached to the 5' end of this sequence is four iSpC3 spacer units, the last of which is attached to SEQ ID NO: 23. Attached to the 5' end of SEQ ID NO: 23 is 28 iSpC3 spacer units the last of which has an
30 additional two T's attached to the 5' end of the spacer group.

SEQ ID NO: 25 shows a polynucleotide sequence used in Example 5. It has a carboxyfluorescein (FAM) base at its 5' end.

SEQ ID NO: 26 shows a polynucleotide sequence used in Example 5 and 6. It has a black-hole quencher (BHQ-1) base at its 3' end.

35 SEQ ID NO: 27 shows a polynucleotide sequence used in Examples 5 and 6.

SEQ ID NO: 28 shows the amino acid sequence of Hel308 Mbu.

SEQ ID NO: 29 shows a polynucleotide sequence used in Example 5. This sequence is connected to SEQ ID NO: 27 at its 5' end by either one or four iSp9 spacer groups.

5 SEQ ID NO: 30 shows a polynucleotide sequence used in Example 6. This sequence is connected to SEQ ID NO: 27 at its 5' end by one idSp group.

SEQ ID NO: 31 shows a polynucleotide sequence used in Example 6. This sequence is connected to SEQ ID NO: 27 at its 5' end by four idSp groups.

SEQ ID NO: 32 shows a polynucleotide sequence used in Example 6. It has a carboxyfluorescein (FAM) base at its 5' end.

10 SEQ ID NO: 33 shows a polynucleotide sequence used in Example 6. It has a carboxyfluorescein (FAM) base at its 5' end.

SEQ ID NO: 34 shows a polynucleotide sequence used in Example 7.

SEQ ID NO: 35 shows a polynucleotide sequence used in Example 7 and 8. It has a carboxyfluorescein (FAM) base at its 5' end.

15 SEQ ID NO: 36 shows a polynucleotide sequence used in Example 7 and 8.

SEQ ID NO: 37 shows a polynucleotide sequence used in Example 7, 8 and 9.

SEQ ID NO: 38 shows a polynucleotide sequence used in Example 8 and 10.

SEQ ID NO: 39 shows a polynucleotide sequence used in Example 8.

SEQ ID NO: 40 shows a polynucleotide sequence used in Example 8.

20 SEQ ID NO: 41 shows a polynucleotide sequence used in Example 9.

SEQ ID NO: 42 shows a polynucleotide sequence used in Example 9.

SEQ ID NO: 43 shows a polynucleotide sequence used in Example 9.

Detailed description of the invention

25 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.

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

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

Methods of moving helicases past spacers

The invention provides a method of moving one or more stalled helicases past one or more spacers in a target polynucleotide. A helicase is stalled if it has stopped moving along the polynucleotide. Each spacer typically stalls the one or more helicases. Methods for determining whether or not one or more helicases are stalled are discussed below. The one or more helicases may be stalled before a spacer. The one or more helicases may be stalled by a spacer. The one or more helicases may be stalled on a spacer. The invention concerns moving the one or more stalled helicases past, i.e. beyond, the one or more spacers.

The one or more stalled helicases and the target polynucleotide are contacted with a transmembrane pore and a potential is applied. As described in more detail below, the target polynucleotide moves through the pore with the field resulting from the applied potential. The one or more helicases are typically too large to move through the pore. When a part of the target polynucleotide enters the pore and moves through the pore along the field resulting from the applied potential, the one or more helicases are moved past the spacer by the pore as the target polynucleotide moves through the pore. This is because the target polynucleotide (including the one or more spacers) moves through the pore and the one or more helicases remain on top of the pore.

This allows the position of the one or more helicases on the target polynucleotide to be controlled. Before the one or more stalled helicases and the target polynucleotide are contacted with a transmembrane pore and the potential is applied, the one or more helicases remain in the position where they are stalled. Even in the presence of the necessary components to facilitate helicase movement (e.g. ATP and Mg^{2+}), the one or more helicases will not move past a spacer on the target polynucleotide and will not move along the portion of the target polynucleotide on other side of the spacer until they are in the presence of the transmembrane pore and the applied potential.

The one or more helicases will also remain in the position where they are stalled in the presence of the transmembrane pore, but in the absence of the applied potential. In this instance, the application of a potential moves the one or more helicases past a spacer. The application of a potential can therefore be used to instigate the movement of the one or more helicases past a spacer and along the portion of the target polynucleotide on other side of the spacer. For instance, an increase in voltage may be used to move the one or more helicases past the spacer.

The invention also provides a method of controlling the movement of a target polynucleotide through a transmembrane pore. The target polynucleotide is provided with one or more spacers. The target polynucleotide is contacted with one or more helicases and the one or more helicases stall at the one or more spacers. This ensures that the one or more helicases

remain at one or more specific positions on the polynucleotide. This is discussed in more detail below. The target polynucleotide and the one or more stalled helicases are contacted with a transmembrane pore. Once a potential is applied, the one or more helicases move past the one or more spacers and along the portion of the polynucleotide on other side of the spacer(s). This
5 allows the one or more helicases to control the movement of the polynucleotide through the pore. The potential is also typically used to thread the polynucleotide into the pore.

Helicases can control the movement of polynucleotides in at least two active modes of operation (when the helicase is provided with all the necessary components to facilitate movement, e.g. ATP and Mg^{2+}) and one inactive mode of operation (when the helicase is not
10 provided with the necessary components to facilitate movement). When provided with all the necessary components to facilitate movement, the helicase moves along the polynucleotide in a 5' to 3' or a 3' to 5' direction (depending on the helicase), but the orientation of the polynucleotide in the pore (which is dependent on which end of the polynucleotide is captured by the pore) means that the helicase can be used to either move the polynucleotide out of the
15 pore against the applied field or move the polynucleotide into the pore with the applied field. When the end of the polynucleotide towards which the helicase moves is captured by the pore, the helicase works against the direction of the field resulting from the applied potential and pulls the threaded polynucleotide out of the pore and into the *cis* chamber. However, when the end away from which the helicase moves is captured in the pore, the helicase works with the
20 direction of the field resulting from the applied potential and pushes the threaded polynucleotide into the pore and into the *trans* chamber.

When the helicase is not provided with the necessary components to facilitate movement it can bind to the polynucleotide and act as a brake slowing the movement of the polynucleotide when it is pulled into the pore by the field resulting from the applied potential. In the inactive
25 mode, it does not matter which end of the polynucleotide is captured, it is the applied field which pulls the polynucleotide into the pore towards the *trans* side with the helicase acting as a brake. When in the inactive mode, the movement control of the polynucleotide by the helicase can be described in a number of ways including ratcheting, sliding and braking.

In the method of the invention, the one or more helicases preferably control the
30 movement of the target polynucleotide through the pore with the field resulting from the applied potential. In one preferred embodiment, the one or more helicases are used in the active mode and the end away from which the one or more helicases move is captured by the pore such that the one or more helicases work with the field resulting from the applied potential and push the polynucleotide through the pore. If the one or more helicases move in the 5' to 3' direction, the
35 5' end of the target polynucleotide is preferably captured by the pore. In such embodiments, the

one or more helicases are moved past the one or more spacers in the 5' to 3' direction. If the one or more helicases move in the 3' to 5' direction, the 3' end of the target polynucleotide is preferably captured by the pore. In such embodiments, the one or more helicases are moved past the one or more spacers in the 3' to 5' direction.

5 In another preferred embodiment, the one or more helicases are used in the inactive mode such that the applied field pulls the target polynucleotide through the pore and the one or more helicases act as a brake. In the method of the invention, the one or more helicases preferably slow or brake the movement of the target polynucleotide through the pore with the field resulting from the applied potential. In either case, the one or more helicases are typically too large to
10 move through the pore and the pore pushes the one or more helicases past the one or more spacers on the polynucleotide as the polynucleotide moves through the pore with the field resulting from the applied potential.

The method of controlling the movement of a target polynucleotide through a transmembrane pore can be helpful during characterisation of the polynucleotide using the pore,
15 for instance during strand sequencing. The invention also provides a method of characterising a target polynucleotide. The target polynucleotide is provided with one or more spacers. The target polynucleotide is contacted with one or more helicases and the one or more helicases stall at the one or more spacers. This ensures that the one or more helicases remain at one or more specific positions on the polynucleotide. This is discussed in more detail below. The target
20 polynucleotide and the one or more stalled helicases are contacted with a transmembrane pore. Once a potential is applied, the one or more helicases move past the one or more spacers and along the portion of the polynucleotide on other side of the spacer(s). This allows the one or more helicases to control the movement of the polynucleotide through the pore. The method also comprises taking one or more measurements as the polynucleotide moves with respect to the
25 pore. The measurements are indicative of one or more characteristics of the polynucleotide.

The ability to stall one or more helicases on the target polynucleotide and move the one or more helicases past the one or more spacers using a transmembrane pore and an applied potential is advantageous because it allows effective characterisation, such as sequencing, of the target polynucleotide. For instance, the one or more helicases can be stalled towards one end of
30 the target polynucleotide in a leader sequence which is designed to be captured by the pore and does not need to be characterised (as described below). The stalling of the one or more helicases in the leader sequence means that the one or more helicases do not move away from the leader sequence along the part of the polynucleotide to be characterised until it/they are contacted with the pore and a potential is applied. Once the one or more helicases and the polynucleotide are
35 contacted with the pore and a potential is applied, the leader sequence is typically captured by

the pore and moves through the pore. This movement moves the one or more helicases past the spacer(s) and along the part of polynucleotide to be characterised (as described above). The one or more helicases may then control the movement of the part of the polynucleotide to be characterised.

5 If the one or more helicases are not stalled in the leader sequence, it/they would move along the polynucleotide away from the leader sequence and along the part of the polynucleotide to be characterised. When the one or more helicases and the polynucleotide are contacted with the pore and a potential is applied under these circumstances, the leader sequence and some, if not all, of the polynucleotide to be characterised will move in an uncontrolled manner through
10 the pore along the field resulting from the applied potential. Only once the one or more helicases come into contact with the pore will it/they begin to control the movement of the part of the polynucleotide to be characterised as discussed above. Any part of the polynucleotide which moves through the pore in an uncontrolled manner cannot be characterised as described below. If the one or more helicases move away from the leader sequence and along most of the rest of
15 the target polynucleotide little, if any, of the polynucleotide will be characterised.

The use of one or more spacers in accordance with the invention also allows the number and position of the one or more helicases on the target polynucleotide to be controlled as discussed in more detail below. For instance, a specific number of helicases may be stalled at specific positions on adaptors which may be ligated to the target polynucleotide before
20 characterisation. Such adaptors are provided by the invention and may be provided in a kit for characterisation. The use of one or more spacers ensures that the helicases remain where they supposed to be until the characterisation is begun, even if the adaptor and/or target polynucleotide before characterisation are in the presence of the components necessary to facilitate helicase movement (e.g. ATP and Mg^{2+}).

Polynucleotide

A polynucleotide, such as a nucleic acid, is a macromolecule comprising two or more nucleotides. The polynucleotide or nucleic acid may comprise any combination of any nucleotides. The nucleotides can be naturally occurring or artificial. One or more nucleotides in
30 the polynucleotide can be oxidized or methylated. One or more nucleotides in the polynucleotide may be damaged. For instance, the polynucleotide may comprise a pyrimidine dimer. Such dimers are typically associated with damage by ultraviolet light and are the primary cause of skin melanomas. One or more nucleotides in the polynucleotide may be modified, for instance with a label or a tag. Suitable labels are described below.

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 nucleotide in the polynucleotide is typically a ribonucleotide or deoxyribonucleotide. The polynucleotide may comprise the following nucleosides: adenosine, uridine, guanosine and cytidine. The nucleotide is preferably a deoxyribonucleotide. The polynucleotide preferably comprises the following nucleosides: deoxyadenosine (dA), deoxyuridine (dU) and/or thymidine (dT), deoxyguanosine (dG) and deoxycytidine (dC).

The nucleotide typically contains a monophosphate, diphosphate or triphosphate. Phosphates may be attached on the 5' or 3' side of a nucleotide.

Suitable nucleotides include, but are not limited to, adenosine monophosphate (AMP), guanosine monophosphate (GMP), thymidine monophosphate (TMP), uridine monophosphate (UMP), 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 polynucleotide preferably comprises the following nucleotides: dAMP, dUMP and/or dTMP, dGMP and dCMP.

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

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

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

The polynucleotide preferably does not comprise any abasic nucleotides (i.e. nucleotides which lack a nucleobase), except in the one or more spacers. The polynucleotide preferably does not comprise any C3 spacers (i.e. nucleotide which lack a nucleobase and a sugar), except in the one or more spacers.

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

The helicase may move along the whole or only part of the target polynucleotide in the method of the invention. The whole or only part of the target polynucleotide may be characterised using the method of the invention.

The target polynucleotide may be single stranded. At least a portion of the target polynucleotide is preferably double stranded. Helicases typically bind to single stranded polynucleotides. If at least a portion of the target polynucleotide is double stranded, the target polynucleotide preferably comprises a single stranded region or a non-hybridised region. The one or more helicases are capable of binding to the single stranded region or one strand of the non-hybridised region. The target polynucleotide preferably comprises one or more single stranded regions or one or more non-hybridised regions.

The one or more spacers are preferably included in the single stranded region or the non-hybridised region of the target polynucleotide. The target polynucleotide may comprise more than one single stranded region or more than one non-hybridised region. The target polynucleotide may comprise a single stranded region or a non-hybridised region within its sequence and/or at one or both ends. The one or more spacers may be included in the double stranded region of the target polynucleotide.

If the one or more helicases used in the method move in the 5' to 3' direction, the target polynucleotide preferably comprises a single stranded region or a non-hybridised region at its 5' end. If the one or more helicases used in the method move in the 3' to 5' direction, the target polynucleotide preferably comprises a single stranded region or a non-hybridised region at its 3' end. If the one or more helicases are used in the inactive mode (i.e. as a brake), it does not matter where the single stranded region or the non-hybridised region is located.

The single stranded region 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 target polynucleotide through the pore. 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
5 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,
10 such as from 20 to 150 nucleotides in length. The length of the leader typically depends on the transmembrane pore used in the method.

If at least a portion of the target polynucleotide is double stranded, the two strands of the double stranded portion are preferably linked using a bridging moiety, such as a hairpin. This facilitates characterisation method of the invention. Linking the two strands of the target
15 polynucleotide by a bridging moiety allows both strands of the polynucleotide to be characterised, such as sequenced, by the transmembrane pore. The two strands dehybridise as the polynucleotide moves through the pore as a single stranded polynucleotide. This method is advantageous because it doubles the amount of information obtained from a single double stranded target polynucleotide. Moreover, because the sequence in the complementary 'anti-
20 sense' strand is necessarily orthogonal to the sequence of the 'sense' strand, the information from the two strands can be combined informatively. Thus, this mechanism provides an orthogonal proof-reading capability that provides higher confidence observations.

Any of the embodiments disclosed in International Application No. PCT/GB2012/051786 (published as WO 2013/014451) may be used. The bridging moiety
25 typically covalently links the two strands of the polynucleotide. The bridging moiety can be anything that is capable of linking the two strands of the target polynucleotide, provided that the bridging moiety does not interfere with movement of the polynucleotide through the transmembrane pore. Suitable bridging moieties include, but are not limited to a polymeric linker, a chemical linker, a polynucleotide or a polypeptide. Preferably, the bridging moiety
30 comprises DNA, RNA, modified DNA (such as abasic DNA), RNA, PNA, LNA or polyethylene glycol (PEG). The bridging moiety is more preferably DNA or RNA. The bridging moiety may comprise the one or more spacers.

The bridging moiety is most preferably a hairpin loop. The hairpin loop may be formed from any of the polynucleotides disclosed above. The hairpin loop or the loop of the hairpin

loop is typically from about 4 to about 100 nucleotides in length, preferably from about 4 to about 8 nucleotides in length.

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

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

The one or more spacers are preferably positioned such that it/they stall(s) the one or more helicases and prevents it/them from moving along the target polynucleotide to be controlled or characterised. For instance, the one or more spacers are preferably located between a leader sequence and the target polynucleotide to be controlled or characterised, for instance within a leader sequence at one end of the polynucleotide. The leader sequence typically enters the pore with the field resulting from the applied potential and the one or more helicases are moved past the one or more spacers as the polynucleotide moves through the pore. The one or more helicases may then control the movement of the remainder of the target polynucleotide through the pore and facilitate its characterisation.

In the most preferred embodiment, the target polynucleotide comprises a double stranded portion which is linked at one end by a bridging moiety, such as a hairpin loop, and a single stranded portion at the other end from the bridging moiety which comprises a leader sequence. The one or more spacers may be present in the leader sequence and/or the bridging moiety.

The target polynucleotide is present in any suitable sample. The invention is typically carried out on a sample that is known to contain or suspected to contain the target polynucleotide. 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 or pigs or may alternatively be pets such as cats or dogs. Alternatively a sample of plant origin is typically obtained from a commercial crop, such as a cereal, legume, fruit or vegetable, for example wheat, quinoa, barley, oats, canola, maize, soya, rice, 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 assayed, for example by centrifugation or by passage through a membrane that filters out unwanted molecules or cells, such as red blood cells. The sample may be measured immediately upon being taken. The sample may also be typically stored prior to assay, preferably below -70°C.

Spacer(s)

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

There may be any number of spacers in the target polynucleotide, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more spacers. There are preferably two, four or six spacers in the target polynucleotide. There may be spacer in different regions of the target polynucleotide, such as a spacer in the leader sequence and a spacer in the hairpin loop.

The one or more spacers each provides an energy barrier which the one or more helicases cannot overcome even in the active mode. The one or more spacers may stall the one or more more helicases by by reducing the traction of the helicase (for instance by removing the bases from the nucleotides in the target polynucleotide) or physically blocking movement of the one or more helicases (for instance using a bulky chemical group) .

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Preferred combinations of features are shown in Table 1 below.

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

As discussed in more detail below, the method may concern moving two or more helicases past a spacer. In such instances, the length of the spacer is typically increased to prevent the trailing helicase from pushing the leading helicase past the spacer in the absence of the pore and applied potential. If the method concerns moving two or more helicases past one or more spacers, the spacer lengths discussed above may be increased at least 1.5 fold, such 2 fold, 2.5 fold or 3 fold. For instance, if the method concerns moving two or more helicases past one or more spacers, the spacer lengths in the third column of Table 1 above may be increased 1.5 fold, 2 fold, 2.5 fold or 3 fold.

The two or more helicases may also be separated such that each has its own one or more spacers. This is discussed in more detail below.

Helicase(s)

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; PCT/GB2013/051924 and PCT/GB2013/051928; and in UK Application No. 1318464.3 filed on 18 October 2013).

The helicase preferably comprises the sequence shown in SEQ ID NO: 17 (TrwC Cba) or as variant thereof, the sequence shown in SEQ ID NO: 28 (Hel308 Mbu) or a variant thereof or the sequence shown in SEQ ID NO: 8 (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: 8 comprises E94C/A360C and then (Δ M1)G1G2 (i.e. deletion of M1 and then addition G1 and G2).

Any number of helicases may be moved past the one or more spacers in accordance with the invention. For instance, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more helicases may be moved past the one or more spacer. In some embodiments, different numbers of helicases may be moved past each spacer. For instance, if two helicases are stalled using two separate spacers, one helicase (the first helicase) may be moved past the first spacer, but two helicases (the first and second helicases) may be moved past the second spacer.

The method of the invention preferably comprises moving two or more, such as three or more or four or more, stalled helicases past one or more spacers. The two or more helicases are typically the same helicases. 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.

5 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; PCT/GB2013/051924 and PCT/GB2013/051928; and in UK Application No. 1318464.3 filed on 18 October 2013.

10 *Conditions and transmembrane pores*

The method comprises applying a potential across the pore. The applied potential may be a voltage potential. The method may comprise applying a voltage potential across the pore. The method may comprise increasing the voltage applied across the pore. In this embodiment, the
15 initial voltage potential is typically not sufficient to move the one or more helicases past the one or more spacer and the increased voltage potential is typically sufficient to move the one or more helicases past the one or more spacers. Alternatively, the applied potential may be a chemical potential. An example of this is using a salt gradient across an amphiphilic layer. A salt gradient is disclosed in Holden *et al.*, J Am Chem Soc. 2007 Jul 11;129(27):8650-5. In some
20 instances, the current passing through the pore as the target polynucleotide moves with respect to the pore is used to determine the sequence of the target polynucleotide. This is strand sequencing.

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
25 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 in the membrane along which or into which hydrated ions may flow.

Any transmembrane pore may be used in the invention. The pore may be biological or
30 artificial. Suitable pores include, but are not limited to, protein pores, polynucleotide pores and solid state pores.

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 at least
35 one hydrophilic portion and at least one lipophilic or hydrophobic portion. The amphiphilic

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

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

The amphiphilic layer is typically 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 Examples. 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 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.

In a preferred embodiment, the lipid bilayer is formed as described in International Application No. PCT/GB08/004127 (published as WO 2009/077734).

In another preferred embodiment, the membrane is a solid state layer. A solid-state layer is not of biological origin. In other words, a solid state layer is not derived from or isolated from a biological environment such as an organism or cell, or a synthetically manufactured version of a biologically available structure. Solid state layers can be formed from both organic and inorganic materials including, but not limited to, microelectronic materials, insulating materials such as Si_3N_4 , Al_2O_3 , and SiO_2 , organic and inorganic polymers such as polyamide, plastics such as Teflon® or elastomers such as two-component addition-cure silicone rubber, and glasses. The solid state layer may be formed from monatomic layers, such as graphene, or layers that are only a few atoms thick. Suitable graphene layers are disclosed in International Application No. PCT/US2008/010637 (published as WO 2009/035647).

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 lipid bilayer. 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*.

The target polynucleotide is preferably coupled to the membrane. This may be done using any known method. If the membrane is an amphiphilic layer, such as a lipid bilayer (as discussed in detail above), the target polynucleotide is preferably coupled to the membrane via a polypeptide present in the membrane or a hydrophobic anchor present in the membrane. The hydrophobic anchor is preferably a lipid, fatty acid, sterol, carbon nanotube or amino acid.

The target polynucleotide may be coupled directly to the membrane. It may be coupled to the membrane using any of the ways disclosed in International Application Number No. PCT/GB2012/051191 (published as WO 2012/164270). The target polynucleotide is preferably coupled to the membrane via a linker. Preferred linkers include, but are not limited to, polymers, such as polynucleotides, polyethylene glycols (PEGs) and polypeptides. If a target polynucleotide is coupled directly to the membrane, then some data will be lost as the characterising run cannot continue to the end of the polynucleotide due to the distance between the membrane and the pore and/or helicase. If a linker is used, then the target polynucleotide can be processed to completion. If a linker is used, the linker may be attached to the target polynucleotide at any position. The linker is typically attached to the target polynucleotide at the tail polymer.

The coupling may be stable or transient. For certain applications, the transient nature of the coupling is preferred. If a stable coupling molecule were attached directly to either the 5' or 3' end of a polynucleotide, then some data will be lost as the characterising run cannot continue to the end of the polynucleotide due to the distance between the membrane and the pore and/or helicase. If the coupling is transient, then when the coupled end randomly becomes free of the membrane, then the polynucleotide can be processed to completion. Chemical groups that form stable or transient links with the membrane are discussed in more detail below. The polynucleotide may be transiently coupled to an amphiphilic layer, such as a lipid bilayer using cholesterol or a fatty acyl chain. Any fatty acyl chain having a length of from 6 to 30 carbon atoms, such as hexadecanoic acid, may be used.

In preferred embodiments, the polynucleotide is coupled to an amphiphilic layer. Coupling of polynucleotides to synthetic lipid bilayers has been carried out previously with various different tethering strategies. These are summarised in Table 2 below.

Table 2

Attachment group	Type of coupling	Reference
Thiol	Stable	Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." <i>J Am Chem Soc</i> 125 (13): 3696-7.
Biotin	Stable	Nikolov, V., R. Lipowsky, et al. (2007). "Behavior of giant vesicles with anchored DNA molecules." <i>Biophys J</i> 92 (12): 4356-68
Cholesterol	Transient	Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies." <i>J Am Chem Soc</i> 126 (33): 10224-5
Lipid	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." <i>Langmuir</i> 26 (11): 8666-72

Polynucleotides may be functionalized using a modified phosphoramidite in the synthesis reaction, which is easily compatible for the addition of reactive groups, such as thiol, cholesterol, lipid and biotin groups. These different attachment chemistries give a suite of attachment options for polynucleotides. Each different modification group tethers 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 can also be achieved by a number of other means provided that a reactive group can be added to the polynucleotide. The addition of reactive groups to either end of DNA has been reported previously. A thiol group can be added to the 5' of ssDNA using polynucleotide kinase and ATP γ S (Grant, G. P. and P. Z. Qin (2007). "A facile method for

attaching nitroxide spin labels at the 5' terminus of nucleic acids." *Nucleic Acids Res* **35**(10): e77). 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).

Alternatively, the reactive group could be considered to be a short region in the polynucleotide to one already coupled to the membrane, so that attachment can be achieved via hybridisation. The region could be part of the polynucleotide or ligated to it. 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).

Most preferably, the polynucleotide is coupled to the membrane using a cholesterol-tagged polynucleotide which hybridises to the polynucleotide.

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 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 lipid bilayer, to the other. The transmembrane protein pore allows a polynucleotide or nucleic acid, 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 6, 7, 8 or 9 subunits. The pore is preferably a hexameric, heptameric, octameric or nonameric pore.

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 lipid bilayer such that it diffuses to the lipid bilayer and is inserted by binding to the lipid bilayer and assembling into a functional state. Alternatively, 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)8 and is called MS-(B2)8. The pore used in the invention is preferably MS-(B2)8. The further preferred variant comprises the mutations G75S/G77S/L88N/Q126R. The variant of SEQ ID NO: 2 has the mutations G75S/G77S/L88N/Q126R in addition to all the mutations of MS-(B1)8 and is called MS-(B2C)8. 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. Conservative amino acid changes are well-known in the art and may be selected in accordance with the properties of the 20 main amino acids as defined in Table 3 below. Where amino acids have similar polarity, this can also be determined by reference to the hydropathy scale for amino acid side chains in Table 4.

Table 3 – Chemical properties of amino acids

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

Table 4 - Hydropathy scale

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

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

5

One or more amino acid residues of the 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.

10 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.

15 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
20 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
25 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,
30 within its α -helices and/or loop regions.

The monomers derived from Msp may be modified to assist their identification or purification, for example by the addition of histidine residues (a his tag), aspartic acid residues (an asp tag), a streptavidin tag or a flag tag, or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence. An
35 alternative to introducing a genetic tag is to chemically react a tag onto a native or engineered position on the pore. An example of this would be to react a gel-shift reagent to a cysteine

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

The monomer derived from Msp may be labelled with a revealing label. The revealing label may be any suitable label which allows the pore to be detected. Suitable labels are described below.

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

The monomer derived from Msp contains one or more specific modifications to facilitate nucleotide discrimination. The monomer derived from Msp may also contain other non-specific modifications as long as they do not interfere with pore formation. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the monomer derived from Msp. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with NaBH₄, 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. The transmembrane protein pore preferably comprises seven monomers each comprising the sequence shown in SEQ ID NO: 4 or a variant thereof. Amino acids 1, 7 to 21, 31 to 34, 45 to 51, 63 to 66, 72, 92 to 97, 104 to 111, 124 to 136, 149 to 153, 160 to 164, 173 to 206, 210 to 213, 217, 218, 223 to 228, 236 to 242, 262 to 265, 272 to 274, 287 to 290 and 294 of SEQ ID NO: 4 form loop regions. Residues 113 and 147 of SEQ ID NO: 4 form part of a constriction of the barrel or channel of α -HL.

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

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

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

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

The variant may be a naturally occurring variant which is expressed naturally by an organism, for instance by a *Staphylococcus* bacterium. Alternatively, the variant may be expressed *in vitro* or recombinantly by a bacterium such as *Escherichia coli*. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the amino acid sequence of SEQ ID NO: 4, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of SEQ ID NO: 4 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 200 or more, for example 230, 250, 270 or 280 or more, contiguous amino acids ("hard homology"). Homology can be determined as discussed above.

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

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

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

10 One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the amino terminus or carboxy terminus of the amino acid sequence of SEQ ID NO: 4 or a variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids. A carrier protein may be fused to a pore or variant.

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

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

The variant may be modified to assist its identification or purification as discussed above.

Pores derived from α -HL can be made as discussed above with reference to pores derived from Msp.

30 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).

The helicase may be covalently attached to the pore. The helicase is preferably not covalently attached to the pore.

Any of the proteins described herein may be modified to assist their identification or purification, for example by the addition of histidine residues (a his tag), aspartic acid residues (an asp tag), a streptavidin tag, a flag tag, a SUMO tag, a GST tag or a MBP tag, or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence. An alternative to introducing a genetic tag is to chemically react a tag onto a native or engineered position on the helicase or pore. An example of this would be to react a gel-shift reagent to a cysteine engineered on the outside of the pore. This has been demonstrated as a method for separating hemolysin hetero-oligomers (Chem Biol. 1997 Jul; 4 (7):497-505).

The target polynucleotide, helicase or pore may be labelled with a revealing label. The revealing label may be any suitable label which can be detected. Suitable labels include, but are not limited to, fluorescent molecules, radioisotopes, e.g. ¹²⁵I, ³⁵S, enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

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

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

The proteins used in the invention may also contain other non-specific modifications as long as they do not interfere with the proteins' function. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the protein(s). Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with NaBH₄, amidination with methylacetimidate or acylation with acetic anhydride.

Polynucleotide sequences encoding a protein may be derived and replicated using standard methods in the art. Polynucleotide sequences encoding a protein may be expressed in a bacterial host cell using standard techniques in the art. The protein may be produced in a cell by *in situ* expression of the polypeptide from a recombinant expression vector. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

These methods are described in Sambrook, J. and Russell, D. (2001). *Molecular Cloning: A Laboratory Manual*, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The gene encoding the sequence of interest may be amplified using PCR involving specific primers. The amplified sequences may then be incorporated into a recombinant replicable vector such as a cloning vector. The vector may be used to replicate the polynucleotide in a compatible host cell. Thus polynucleotide sequences may be made by introducing a polynucleotide encoding the sequence of interest into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells for cloning of polynucleotides are known in the art and described in more detail below.

The polynucleotide sequence may be cloned into a suitable expression vector. In an expression vector, the polynucleotide sequence is typically operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell. Such expression vectors can be used to express a construct.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. Multiple copies of the same or different polynucleotide may be introduced into the vector.

The expression vector may then be introduced into a suitable host cell. Thus, a construct can be produced by inserting a polynucleotide sequence encoding a construct into an expression vector, introducing the vector into a compatible bacterial host cell, and growing the host cell under conditions which bring about expression of the polynucleotide sequence.

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

The host cell typically expresses the construct at a high level. Host cells transformed with a polynucleotide sequence will be chosen to be compatible with the expression vector used to transform the cell. The host cell is typically bacterial and preferably *E. coli*. Any cell with a λ DE3 lysogen, for example Rosetta2(DE3)pLys, C41 (DE3), BL21 (DE3), JM109 (DE3), B834
15 (DE3), TUNER, Origami and Origami B, can express a vector comprising the T7 promoter.

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

20 The method of the invention involves measuring one or more characteristics of the target polynucleotide. The method may involve measuring two, three, four or five or more characteristics of the target polynucleotide. The one or more characteristics are preferably selected from (i) the length of the target polynucleotide, (ii) the identity of the target polynucleotide, (iii) the sequence of the target polynucleotide, (iv) the secondary structure of the
25 target polynucleotide and (v) whether or not the target polynucleotide is modified. Any combination of (i) to (v) may be measured in accordance with the invention.

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

30 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 target polynucleotide or without measurement of the sequence of the target 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
35 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 target 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 target 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.

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

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

The 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
5 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.

10 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 Examples. The method is typically carried out with a voltage applied across the membrane and
15 pore. The voltage used is typically from +2 V to -2 V, typically -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
20 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
25 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
30 ferrocyanide and potassium ferricyanide are preferred. 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. Hel308, XPD, RecD, TraI and Dda helicases surprisingly work under high salt concentrations. The method is preferably carried out
35 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.

5 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
10 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 helicase function, such as about 37 °C.

15 The method may be carried out in the presence of free nucleotides or free nucleotide analogues and/or a helicase cofactor that facilitates the action of the helicase. The method may also be carried out in the absence of free nucleotides or free nucleotide analogues and in the absence of a helicase cofactor. The free nucleotides may be one or more of any of the individual nucleotides discussed above. The free nucleotides include, but are not limited to, adenosine
20 monophosphate (AMP), adenosine diphosphate (ADP), adenosine triphosphate (ATP), guanosine monophosphate (GMP), guanosine diphosphate (GDP), guanosine triphosphate (GTP), thymidine monophosphate (TMP), thymidine diphosphate (TDP), thymidine triphosphate (TTP), uridine monophosphate (UMP), uridine diphosphate (UDP), uridine triphosphate (UTP), cytidine monophosphate (CMP), cytidine diphosphate (CDP), cytidine triphosphate (CTP), cyclic
25 adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), deoxyadenosine monophosphate (dAMP), deoxyadenosine diphosphate (dADP), deoxyadenosine triphosphate (dATP), deoxyguanosine monophosphate (dGMP), deoxyguanosine diphosphate (dGDP), deoxyguanosine triphosphate (dGTP), deoxythymidine monophosphate (dTMP), deoxythymidine diphosphate (dTDP), deoxythymidine triphosphate (dTTP), deoxyuridine
30 monophosphate (dUMP), deoxyuridine diphosphate (dUDP), deoxyuridine triphosphate (dUTP), deoxycytidine monophosphate (dCMP), deoxycytidine diphosphate (dCDP) and deoxycytidine triphosphate (dCTP). The free nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP or dCMP. The free nucleotides are preferably adenosine triphosphate (ATP). The helicase cofactor is a factor that allows the helicase or construct to

function. The helicase cofactor is preferably a divalent metal cation. The divalent metal cation is preferably Mg^{2+} , Mn^{2+} , Ca^{2+} or Co^{2+} . The helicase cofactor is most preferably Mg^{2+} .

Methods of controlling the loading of one or more helicases on a target polynucleotide

5 The invention also provides a method of controlling the loading of one or more helicases on a target polynucleotide. The method comprises providing the target polynucleotide with one or more spacers. The method preferably comprises modifying the target polynucleotide so that it comprises one or more spacers. All of the spacer embodiments discussed above equally apply to this method.

10 The method also comprises contacting the target polynucleotide with the one or more helicases such that the one or more helicases bind to the target polynucleotide and one or more helicases stall at each spacer. The stalling of helicases at spacers may be assayed as discussed above.

 The target polynucleotide may comprise any number of spacers as discussed above. The
15 target polynucleotide preferably comprises two or more spacers, such as 3, 4, 5, 6, 7, 8, 9, 10 or more spacers. Any number of helicases may be stalled at each spacer as discussed above. In this way, it is possible to control where and how many helicases are loaded on the target polynucleotide and thereby facilitate characterisation of the target polynucleotide. The one or more helicases may be moved past the one or more spacers using any of the methods discussed
20 above.

 The target polynucleotide is preferably provided with one or more spacers S and one or more single stranded regions or one or more non-hybridised regions L (L is for loading site). The length of each region L depends on the number of helicases that should bind to each L and be stalled at each spacer S. The one or more spacers S and one or more regions L may be
25 adjacent to (i.e. next to) one another or may be separated by part of the target polynucleotide. Each spacer is typically located at or near the end of each region L towards which the helicase moves. For instance, if the helicase is a 5' to 3' helicase, each spacer is typically located at or near the 3' end of each region, i.e. 5'-L-S-3'. If the helicase is a 3' to 5' helicase, each spacer is typically located at or near the 5' end of each region, i.e. 5'-S-L-3'.

30 The target polynucleotide is preferably provided with (L-S)_n or (S-L)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4. The 5' to 3' direction refers to the target polynucleotide.

The target polynucleotide is preferably provided with one or more single stranded regions or one or more non-hybridised regions L each of which has a spacer S at or near either end, i.e. provided with (S-L-S)_n.

In a preferred embodiment, the spacer is adjacent to a double stranded region D as discussed above, i.e. 5'-L-S-D-3' for 5' to 3' helicases or helicases used in the inactive mode or 5'-D-S-L-3' for 3' to 5' helicases or helicases used in the inactive mode. The target polynucleotide is preferably provided with (L-S-D)_n or (D-S-L)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer, D is a double stranded polynucleotide and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4. L may be the same type of polynucleotide as D or may be a different type of polynucleotide from D. L and/or D may be the same type of polynucleotide as the target polynucleotide or may be a different type of polynucleotide from the target polynucleotide.

In a preferred embodiment, a blocking molecule B is provided at the end of each spacer opposite to the end past which the one or more helicases are to be moved, i.e. 5'-B-L-S-3' for 5' to 3' helicases or helicases used in the inactive mode or 5'-S-L-B-3' for 3' to 5' helicases. The target polynucleotide is preferably provided with (B-L-S)_n or (S-L-B)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer, B is blocking molecule and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4.

In the most preferred embodiment, the target polynucleotide is provided with both D and B, i.e. 5'-B-L-S-D-3' for 5' to 3' helicases or helicases used in the inactive mode or 5'-D-S-L-B-3' for 3' to 5' helicases. The target polynucleotide is most preferably provided with (B-L-S-D)_n or (D-S-L-B)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer, B is blocking molecule, D is a double stranded polynucleotide and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4.

The target polynucleotide may be provided with any number of these spacer-containing units. For instance, the target polynucleotide may be provided with (5'-L-S-3')_n, (5'-S-L-3')_n, (S-L-S)_n, (5'-L-S-D-3')_n, (5'-D-S-L-3')_n, (5'-B-L-S-3')_n, (5'-S-L-B-3')_n, (5'-B-L-S-D-3')_n or (5'-D-S-L-B-3')_n, where n is 2 or more, such as 3, 4, 5, 6, 7, 8, 9, 10 or more. Such embodiments allow multiple helicases to be stalled on the target polynucleotide.

The target polynucleotide may be provided with all of the embodiments discussed above with reference to L, S, D and B by ligating an adaptor of the invention to the target polynucleotide.

In a preferred embodiment, the target polynucleotide is contacted with the one or more helicases such that one helicase (i.e. only one helicase) stalls at each spacer. This can be achieved by providing the target polynucleotide with one or more spacers S and one or more single stranded regions or one or more non-hybridised regions L^1 each of which is only long enough for one helicase to bind. The target polynucleotide is preferably provided with $(L^1-S)_n$ or $(S-L^1)_n$ in the 5' to 3' direction, wherein L^1 is a single stranded polynucleotide or a non-hybridised polynucleotide which is only long enough for one helicase to bind, S is a spacer and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4.

The length of region L^1 depends on the footprint of the helicase and can be calculated in a straightforward manner. Region L^1 may be part of the target polynucleotide or may be added to the target polynucleotide, for instance as part of an adaptor of the invention. Region L^1 is typically 8, 9, 10, 12, 13, 14 or 15 nucleotides in length. The one or more spacers S and one or more L^1 regions may be adjacent to (i.e. next to) one another or may be separated by part of the target polynucleotide. Each spacer S is typically located at or near the end of each region L^1 towards which the helicase moves. For instance, if the helicase is a 5' to 3' helicase, each spacer S is typically located at or near the 3' end of each region L^1 , i.e. 5'- L^1 -S-3'. If the helicase is a 3' to 5' helicase, each spacer S is typically located at or near the 5' end of each region L^1 , i.e. 5'-S- L^1 -3'. The target polynucleotide is preferably provided with one or more single stranded regions or one or more non-hybridised regions L^1 each of which is only long enough for one helicase to bind and each of which has a spacer S at or near either end, i.e. $(S-L^1-S)_n$.

The target polynucleotide may be provided with $(5'-L^1-S-3')_n$, $(5'-S-L^1-3')_n$, $(S-L^1-S)_n$, $(5'-L^1-S-D-3')_n$, $(5'-D-S-L^1-3')_n$, $(5'-B-L^1-S-3')_n$, $(5'-S-L^1-B-3')_n$, $(5'-B-L^1-S-D-3')_n$ or $(5'-D-S-L^1-B-3')_n$, where n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4. Such embodiments allow n helicases to be stalled on the target polynucleotide. One helicase is stalled by each spacer.

In another preferred embodiment, the target polynucleotide is contacted with the one or more helicases such that two helicases (i.e. only two helicases) stall at each spacer. This can be achieved by providing the target polynucleotide with one or more spacers S and one or more single stranded regions or one or more non-hybridised regions L^2 each of which is only long enough for two helicases to bind. The target polynucleotide is preferably provided with $(L^2-S)_n$ or $(S-L^2)_n$ in the 5' to 3' direction, wherein L^2 is a single stranded polynucleotide or a non-hybridised polynucleotide which is only long enough for two helicases to bind, S is a spacer and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4.

The length of region L^2 depends on the footprint of the helicases and can be calculated in a straightforward manner. Region L^2 may be part of the target polynucleotide or may be added

to the target polynucleotide, for instance as part of an adaptor of the invention. Region L^2 is typically 16, 17, 18, 19, 20, 21 or 22 nucleotides in length. The one or more spacers S and one or more regions L^2 may be adjacent to (i.e. next to) one another or may be separated by part of the polynucleotide. Each spacer is typically located at or near the end of each region towards
 5 which the helicase moves. For instance, if the helicase is a 5' to 3' helicase, each spacer is typically located at or near the 3' end of each region. The polynucleotide is preferably provided with one or more single stranded regions or one or more non-hybridised regions L^2 each of which is only long enough for two helicases to bind and each of which has a spacer S at or near either end, i.e. (S- L^2 -S) $_n$.

10 The target polynucleotide may be provided with (5'- L^2 -S-3') $_n$, (5'-S- L^2 -3') $_n$, (S- L^2 -S) $_n$, (5'- L^2 -S-D-3') $_n$, (5'-D-S- L^2 -3') $_n$, (5'-B- L^2 -S-3') $_n$, (5'-S- L^2 -B-3') $_n$, (5'-B- L^2 -S-D-3') $_n$ or (5'-D-S- L^2 -B-3') $_n$, where n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4. Such embodiments allow 2n helicases to be stalled on the target polynucleotide. Two helicases are stalled by each spacer.

15 The two helicases stalled at each spacer are preferably different from one another. This can be controlled in several ways. For instance, two different helicases may be attached to one another, such as covalently attached to one another, and then stalled at each spacer. Suitable constructs are discussed above. Alternatively, blocking polynucleotides may be used to ensure that different helicases are stalled by each spacer. If the method comprises providing the
 20 polynucleotide with one or more spacers S and one or more single stranded regions or one or more non-hybridised regions L^2 each of which is only long enough for two helicases to bind, the method preferably comprises hybridising a blocking polynucleotide to part of each region L^2 so that the remaining (i.e. non-blocked) part of each region is only long enough to bind one helicase. Blocking polynucleotides are typically 2, 3, 4, 5, 6, 7 or 8 nucleotides in length. The
 25 blocking polynucleotide prevents two helicases from binding to the same region at the same time. The polynucleotide comprising the blocking polynucleotides is preferably contacted with one or more helicases such that one helicase binds to the remaining (i.e. non-blocked) part of each region L^2 . Each helicase may then be used to remove each blocking polynucleotide. The one or more bound helicases are preferably provided with free nucleotides and a helicase
 30 cofactor such that they remove each blocking polynucleotide and stall at each spacer S. The polynucleotide produced in this way is then preferably contacted with one or more helicases which are different from the helicases used earlier in the method such that one different helicase binds to each region and is stalled by the spacer and the other stalled helicase.

The method preferably comprises (a) providing the target polynucleotide with (L^2 -S) $_n$ or
 35 (S- L^2) $_n$ in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-

hybridised polynucleotide which is only long enough for two helicases to bind, S is a spacer and n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more; (b) hybridising a blocking polynucleotide to part of each region L^2 so that the remaining part of each region L^2 is only long enough to bind one helicase; (c) contacting the target polynucleotide produced in (b) with one or more helicases such that one helicase binds to the remaining part of each region L^2 ; (d) providing the one or more bound helicases in (c) with free nucleotides and a helicase cofactor such that they remove each blocking polynucleotide and stall at each spacer S; and (e) contacting the target polynucleotide produced in (d) with one or more helicases which are different from those used in (c) such that one different helicase binds to each region L^2 and is stalled by each spacer and each helicase stalled in (d). n is preferably 1, 2, 3 or 4. Other arrangements of S and L^2 , such as $(S-L^2-S)_n$, $(5'-L^2-S-D-3')_n$, $(5'-D-S-L^2-3')_n$, $(5'-B-L^2-S-3')_n$, $(5'-S-L^2-B-3')_n$, $(5'-B-L^2-S-D-3')_n$ and $(5'-D-S-L^2-B-3')_n$ as discussed above, may also be used in this embodiment.

As discussed above, the length of a spacer may be used to control the number of helicases that are stalled and/or the number of helicases which may be moved past the spacer. Longer spacers may be used to stall more helicases. Trains of two or more helicases, such as 3, 4 or 5 helicases, may also move past longer spacers because trailing helicases may push leading helicases past the spacer. The embodiments with reference to L^1 and L^2 above can be modified such that 3, 4 or 5 helicases are stalled at each spacer. For instance, the polynucleotide may be provided with one or more spacers S and one or more single stranded regions or one or more non-hybridised regions each of which is only long enough for three (L^3), four (L^4) or five (L^5) helicases to bind.

Adaptor

The invention also provides an adaptor for controlling the movement of a target polynucleotide. The adaptor is preferably for characterising a target polynucleotide. The adaptor comprises (a) $(L-S-D)_n$ or $(D-S-L)_n$ in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer and D is a double stranded polynucleotide and wherein n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more, and (b) one or more helicases stalled on each adaptor. n is preferably 1, 2, 3 or 4. The 5' to 3' direction refers to the direction of the L and D polynucleotides in the adaptor.

The one or more helicases may be stalled before the spacer S, by the spacer S or on the spacer S.

The adaptor may be ligated to a target polynucleotide such that the target polynucleotide may be used in any of the method discussed above.

L may be L^1 or L^2 as discussed above. An adaptor may comprise a combination of L^1 and L^2 .

All of the spacer embodiments discussed above equally apply to this method.

Any of the embodiments discussed above with reference to L, S and D equally apply to the adaptors of the invention. The adaptor may comprise $(5'-L^1-S-D-3')_n$, $(5'-D-S-L^1-3')_n$, $(5'-B-L^1-S-D-3')_n$ or $(5'-D-S-L^1-B-3')_n$, $(5'-L^2-S-D-3')_n$, $(5'-D-S-L^2-3')_n$, $(5'-B-L^2-S-D-3')_n$ or $(5'-D-S-L^2-B-3')_n$ in the 5' to 3' direction, where n is a whole number, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more. n is preferably 1, 2, 3 or 4. L^1 or L^2 may be replaced with L^3 , L^4 or L^5 .

Most preferably n is 1 and one or two helicases are stalled one the adaptor.

Kit

The invention also provides a kit for controlling the movement of a target polynucleotide. The kit is preferably for characterising a target polynucleotide. The kit comprises (a) one or more spacers, (b) one or more helicases and (c) a transmembrane pore. All of the spacer embodiments discussed above with reference to the methods of the invention equally apply to the kits of the invention. For instance, the one or more spacers may be part of a polynucleotide adaptor, preferably a single stranded polynucleotide adaptor, which may be ligated to the target polynucleotide and which comprises a leader sequence which preferentially threads into the pore. The kit may comprise any of the helicases and pores discussed above.

The one or more spacers and the one or more helicases may be part of an adaptor of the invention.

The kit may further comprise the components of a membrane, such as the phospholipids needed to form an amphiphilic layer, such as a lipid bilayer.

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 the components necessary to facilitate helicase movement (e.g. ATP and Mg^{2+}). The following Examples illustrate the invention.

Examples**Example 1**

This example describes how a T4 Dda – E94C/A360C (SEQ ID NO: 8 with mutations E94C/A360C and then (Δ M1)G1G2) helicase can control the movement of intact DNA strands through a single MspA nanopore (MS(B1- G75S/G77S/L88N/Q126R)8 (MspA – B2C) (SEQ ID NO: 2 with mutations G75S/G77S/L88N/Q126R). The iSpC3 spacers in the lambda DNA construct (SEQ ID NO: 9 attached by its 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 10 which is attached to three iSpC3 spacers which are attached to the 3' end to SEQ ID NO: 11, the SEQ ID NO: 10 region of this construct is hybridised to SEQ ID NO: 12 (which has attached to its 3' end, six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG)) are used to stall the enzyme until the construct is captured by the nanopore. Upon capture the force of the applied potential moves the enzyme T4 Dda – E94C/A360C past the stalling spacer and allows enzyme controlled DNA movement of the lambda construct through the nanopore.

Materials and Methods

Prior to setting up the experiment, the Lambda DNA construct (SEQ ID NO: 9 attached by its 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 10 which is attached at its 3' end to SEQ ID NO: 11, the SEQ ID NO: 10 region of this construct is hybridised to SEQ ID NO: 12 (which has a 3' cholesterol tether)) and T4 Dda – E94C/A360C were pre-incubated together for 15 minutes at 23 °C in buffer (20 mM CAPS, pH 10.0, 500 mM NaCl, 5% Glycerol, 2 mM DTT).

Electrical measurements were acquired at 20 °C (by placing the experimental system on a cooler plate) from single MspA nanopores (MspA - B2C) inserted in block co-polymer in buffer (600 mM KCl, 25 mM potassium phosphate, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). After achieving a single pore inserted in the block co-polymer, then buffer (1 mL, 600 mM KCl, 25 mM potassium phosphate, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) was flowed through the system to remove any excess MspA nanopores (MspA - B2C) and finally experimental buffer was flowed into the system (2 mL 960 mM KCl, 25 mM potassium phosphate, 3 mM Potassium Ferrocyanide (II), 1 mM Potassium ferricyanide (III), pH 8). MgCl₂ (10 mM final concentration) and ATP (1 mM final concentration) were mixed together with buffer (960 mM KCl, 25 mM potassium phosphate, 3 mM Potassium Ferrocyanide (II), 1 mM Potassium ferricyanide (III), pH 8) and then added to the Lambda DNA construct (0.2 nM final concentration), T4 Dda –

E94C/A360C (10 nM final concentration) buffer (20 mM CAPS, pH 10.0, 500 mM NaCl, 5% Glycerol, 2 mM DTT) pre-mix. The pre-mix was then added to the single nanopore experimental system. Experiments were carried out for four hours following a potential flip process (+100 mV for 2 s, then 0 V for 2 s, then -120 mV for 14500s applied at the cis side) and helicase-controlled DNA movement was monitored.

Results and Discussion

The DNA construct is shown in Figure 1. The T4 Dda – E94C/A360C is able to bind on to the region of the construct labelled A (SEQ ID NO: 9) when pre-incubated with the DNA construct. However, the enzyme is not able to motor past the stalling groups when in free solution (labelled B and D in Figure 1). Therefore, the enzyme is stalled at the iSpC3 spacers (labelled B in Figure 1) until the DNA construct is captured by the nanopore. Once captured by the nanopore, the force of the applied potential moves the enzyme T4 Dda – E94C/A360C past the stalling spacer and allows enzyme controlled DNA movement of the lambda construct through the nanopore. An example of a helicase-controlled DNA movement is shown in Figure 2. The helicase-controlled DNA movement was 5170 seconds long and corresponds to the translocation of approximately 30 kB of the lambda construct through the nanopore. Figure 3 shows zoomed in regions of the beginning (a) and end (b) of the helicase-controlled DNA movement, 1 and 2 show when the iSpC3 spacers translocate through the nanopore under control of the helicase.

Example 2

The DNA construct used in this example was produced by fragmentation of Lambda DNA into ~5-10 kB fragments using MuA. The fragments which were produced by the sample prep were then passed through a nanopore, with their movement controlled by a helicase enzyme. The helicase was moved past the dsDNA region (where the tether hybridises to the construct) and the spacers by the force of the applied potential across the nanopore. The observance of characteristic blocks produced by the helicase controlled movement of the markers through the nanopore showed the sample preparation procedure had been successful and that the enzyme had been stalled as shown in Figure 4 (b). This meant that the enzyme had a defined start point and was unable to move past the dsDNA region (where the tether hybridises to the construct) and spacers until captured by the nanopore.

Materials and Methods

2.1 Anneal of DNA strands to form Y-shaped and Hairpin MuA Substrates

The Y-shaped and hairpin MuA substrates were prepared as shown in Table 5 below.

The sample mixtures which contained the DNA to form the Y-shaped and hairpin MuA

substrates were then heated to 95 °C for 2 minutes and then cooled to 16 °C at a rate of 2 °C per minute. This allowed SEQ ID NOs: 14 and 15 (where SEQ ID NO: 14 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units) to anneal to SEQ ID NO: 19 and 9 (where SEQ ID NO: 19 is attached at its 3' end to the 5' end of SEQ ID NO: 9 by four iSpC3 spacer units) to form the Y-shaped MuA substrate and for SEQ ID NO: 19 and 20 (where SEQ ID NO: 19 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units) to form a hairpin loop MuA substrate. The DNA substrate designs of the two MuA substrates formed are shown in Figure 4 (a).

Reagent	Y-shaped	Hairpin	Final Concentrations
Water	12 uL	14 uL	
0.5 M NaCl	2 uL	2 uL	50 mM
0.1 M Tris pH 7.5	2 uL	2 uL	10 mM
SEQ ID NO: 14 and 15 (where SEQ ID NO: 14 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units) (100 uM)	2 uL		10 uM
SEQ ID NO: 19 and 9 (where SEQ ID NO: 19 is attached at its 3' end to the 5' end of SEQ ID NO: 9 by four iSpC3 spacer units) (100 uM)	2 uL		10 uM
SEQ ID NO: 19 and 20 (where SEQ ID NO: 19 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units) (100 uM)		2 uL	10 uM
Total	20 uL	20 uL	

Table 5

2.2 Fragmentation of the DNA Template using the MuA Transposase

Double-stranded Lambda DNA (SEQ ID NO: 13 corresponds to the sequence of the sense strand) was fragmented into approximately 5-10 kB length strands using a MuA transposase. The MuA transposase inserted the MuA substrates (the Y-shaped and the hairpin

MuA substrates) which were annealed in section 3.1. The sample was prepared as shown in Table 6 below. The sample was then incubated at 30 °C for 1 hour and heat inactivated at 75 °C for 10 minutes. The sample was then further purified using a QIAquick™ PCR Purification kit (Qiagen) and eluted in 26 µL.

5

Reagent	Sample Volume 1	Final Concentrations
Water	17.7 uL	
Lambda DNA (90 ng/uL) (SEQ ID NO: 13 shows the sense strand sequence only)	22.3 uL	2 µg
Y-shaped MuA substrate (1 uM)	8 uL	100 nM
Hairpin MuA substrate (1 uM)	8 uL	100 nM
5x Buffer (125 mM Tris (pH 8.0), 50 mM MgCl ₂ , 550 mM NaCl, 0.25% Triton X-100 and 50% Glycerol)	16 uL	1x
MuA (4 uM, Thermo, catalogue No. F-750C)	8 uL	400 nM
Total	80 uL	

Table 6

2.3 USER Digest of Fragmented Lambda DNA with Inserted MuA Substrates

Purified sample volume 1 from step 3.2 was then treated with USER™ digest in order to remove the dUMP from SEQ ID NOs: 19. See Table 7 below for appropriate volumes and concentrations. The sample was then incubated at 37 °C for 30 minutes before it was cooled in an ice block.

10

Reagent	Sample Volume 2	Final Concentrations
Sample Volume 1	26 uL	8 pmol of U
10x DNA ligase buffer	3 uL	1x
USER (1 U/uL)	1 uL	1 U
Total	30 uL	

Table 7

2.4 Repair of single-stranded Gap in the double-stranded Lambda DNA construct fragments

Sample Volume 2 produced after treatment with USER™ was then treated with DNA polymerase and ligase in order to close the single-stranded gap. Sample volume 3 (see table 8 below for appropriate volumes and concentrations) was incubated for 30 minutes at 16 °C and then EDTA (0.5 M, 10 µL) was added to sample volume 3. A QIAquick™ PCR Purification kit was then used to purify each sample, which was eluted in 50 µL of water. An aliquot of the purified sample (1 µL) was run on an Agilent 12000 chip to quantify the sample and Tris-HCl and NaCl (pH 7.5) until were added to the rest of the sample until the concentrations were 10 mM and 50 mM respectively. Finally, SEQ ID NO: 16 (3' end of the sequence has six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG, 0.5 µM) was annealed to the purified sample.

Reagent	Sample Volume 3	Final Concentrations
Water	6.2 uL	
Sample Volume 2	30 uL	
10x DNA ligase buffer	1 uL	1x
dNTPs (10 mM)	0.8 uL	200 uM
T4 DNAP exo(□) (Lucigen)	1 uL	
Ligase (NEB; M0202M)	1 uL	1x
Total	40 uL	

Table 8

2.5 Electrophysiology experiment showing helicase controlled DNA movement of the purified and fragmented Lambda DNA construct

Prior to setting up the experiment, the Lambda DNA construct (0.2 nM, 5-10 kB fragments of Lambda DNA which have had the Y-shaped MuA substrates and the hairpin MuA substrates attached to either end of the fragments by the MuA transposase (see Fig. 4 (b) for an example construct)) and Trwc Cba (SEQ ID NO: 9, 1 µM) were pre-incubated together for 1 hour in buffer (50 mM CAPS, pH 10.0 (pH altered to pH 10.0 by addition of NaOH), 100 mM NaCl).

Electrical measurements were acquired from single MspA nanopores (MspA - B2C) inserted in block co-polymer in buffer (600 mM KCl, 25 mM KH₂PO₄, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). After achieving a single pore in the bilayer, then buffer (1 mL, 600 mM KCl, 25 mM KH₂PO₄, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) was flowed through the system to remove any excess MspA nanopores (MspA - B2C) and the experimental system was placed on a cooler plate set to 8 °C which gave a system temperature of ~15 °C. MgCl₂ (10 mM) and dTTP (5 mM) were mixed together with buffer (600 mM KCl, 25 mM KH₂PO₄, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) and then added to the Lambda DNA construct (0.2 nM), Trwc Cba (SEQ ID NO: 9, 1 µM) buffer (50 mM CAPS, pH 10.0 (pH altered to pH 10.0 by addition of NaOH), 100 mM NaCl) pre-mix. The pre-mix was then added to the single nanopore experimental system. Experiments were carried out for two hours following a potential flip process (+120 mV for 30 mins, then -100 mV for 2 seconds and then 0 mV for 2 seconds) and helicase-controlled DNA movement was monitored.

2.6 Results and Discussion

Helicase controlled DNA movement was observed for the Lambda DNA construct, an example of a helicase-controlled DNA movement is shown in Figure 5. The iSpC3 spacers present in the Lambda DNA construct produced a characteristic block level highlighted by the numbers 1, 2 and 3 in Figure 5. The Y-shaped MuA substrate has four iSpC3 spacers in either strand and the hairpin MuA substrate also has four iSpC3 spacers, each iSpC3 spacer allows more current to flow as that region of the Lambda DNA construct translocates through the nanopore. If the sample preparation has occurred successfully then the iSpC3 spacer events will be observed at the beginning of the Lambda DNA construct, in the middle (marking the transition between sense and antisense sequences) and at the end. Figure 5 clearly shows three instances of increased current flow which correspond to the iSpC3 spacer regions. Therefore, the sample preparation procedure effectively introduced MuA substrates into the Lambda DNA to produce the Lambda DNA constructs shown in Figure 4 (b). Upon capture of the DNA by the nanopore, the enzyme (labelled A) is moved past the dsDNA region (where the tether hybridises to the construct) and the spacers by the force of the applied potential across the nanopore.

Example 3

The DNA construct used in this example was produced by fragmentation of Lambda DNA into ~5-10 kB fragments using MuA. This example is similar to the one described in

Example 2, however, the sample preparation procedure is different (steps 2.3 and 2.4 as described above are not required) as the transposase sequences contain inosines in this example. The enzyme was moved past the dsDNA region and the spacers by the force of the applied potential across the nanopore.

5

Materials and Methods

3.1 Anneal of DNA strands to form Y-shaped and Hairpin MuA Substrates

The Y-shaped 2 and hairpin 2 MuA substrates were prepared as described in Example 2.1 above. Volumes, concentrations and sequences that were used in this example are detailed in table 9 below. The DNA substrate designs of the two constructs formed are shown in Figure 6 (a).

10

Reagent	Y-shaped 2	Hairpin 2	Final Concentrations
Water	12 uL	14 uL	
0.5 M NaCl	2 uL	2 uL	50 mM
0.1 M Tris pH 7.5	2 uL	2 uL	10 mM
SEQ ID NO: 14 and 15 (where SEQ ID NO: 14 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units) (100 uM)	2 uL		10 uM
SEQ ID NO: 18 and 9 (where SEQ ID NO: 18 is attached at its 3' end to the 5' end of SEQ ID NO: 9 by four iSpC3 spacer units) (100 uM)	2 uL		10 uM
SEQ ID NO: 18 and 15 (where SEQ ID NO: 18 is attached at its 3' end to the 5' end of SEQ ID NO: 15 by four iSpC3 spacer units) (100 uM)		2 uL	10 uM
Total	20 uL	20 uL	

Table 9

3.2 Fragmentation of the DNA Template using the MuA Transposase

Double-stranded Lambda DNA (SEQ ID NO: 13 shows the sequence of the sense strand only) was fragmented into approximately 5-10 kB length strands using a MuA transposase. The MuA transposase inserted the MuA substrates (the Y-shaped 2 and the hairpin 2 MuA substrates) which were annealed in section 3.1. The sample was prepared by an analogous procedure as that described in Section 2.2 and table 6 above except the MuA substrates used were the Y-shaped 2

15

and the hairpin 2 MuA substrates. In this case the purified sample X was eluted in a volume of 20 μ L.

3.3 Nick repair in the double-stranded Lambda DNA construct fragments

Once the Y-shaped 2 and the hairpin 2 MuA substrates have been inserted into the fragmented Lambda DNA it is necessary to repair the nick in the strand and join the inosines to the Lambda DNA fragment to produce a complete double-stranded Lambda DNA fragment. One reaction was assembled on ice as described in Table 10 below. The sample was incubated at 16 °C for 60 mins before EDTA (10 μ L, 0.5 M) was added to the sample. The resultant sample mixture was purified using a QiaQuick™ purify and was eluted in 50 μ L of water. An aliquot of the purified sample (1 μ L) was run on an Agilent 12000 chip to quantify the sample and Tris-HCl and NaCl (pH 7.5) were added to the rest of the sample until the concentrations were 10 mM and 50 mM respectively. Finally, SEQ ID NO: 16 (3' end of the sequence has six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG, 0.5 μ M) was annealed to the purified Lambda DNA construct.

Reagent	Sample Z	Final Concentrations
Sample X	16 μ L	
2x DNA ligase buffer	20 μ L	1x
Ligase (NEB; M0202M)	4 μ L	
Total	40 μ L	

Table 10

3.4 Electrophysiology experiment showing helicase controlled DNA movement of the purified and fragmented Lambda DNA construct

Prior to setting up the experiment, the Lambda DNA construct (0.2 nM, 5-10 kB fragments of Lambda DNA which have had the Y-shaped 2 and the hairpin 2 MuA substrates attached to either end of the fragments by the MuA transposase (see Figure 6 (b) for an example construct)) and Trwc Cba (SEQ ID NO: 17, 1 μ M) were pre-incubated together for 1 hour in buffer (50 mM CAPS, pH 10.0 (pH altered to pH 10.0 by addition of NaOH), 100 mM NaCl).

Electrical measurements were acquired from single MspA nanopores (MspA - B2C) inserted in block co-polymer in buffer (600 mM KCl, 25 mM KH₂PO₄, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). After achieving a single pore in the bilayer, then buffer (3 mL, 600 mM KCl, 25 mM KH₂PO₄, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) was flowed through the system to remove any excess MspA nanopores (MspA - B2C) and the experimental system was placed on a cooler

plate set to 8 °C which gave a system temperature of ~15 °C. MgCl₂ (10 mM) and dTTP (5 mM) were mixed together with buffer (600 mM KCl, 25 mM KH₂PO₄, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) and then added to the Lambda DNA construct (0.2 nM), TrwC Cba (SEQ ID NO: 17, 1 µM) buffer (50 mM CAPS, pH 10.0 (pH altered to pH 10.0 by addition of NaOH), 100 mM NaCl) pre-mix. The pre-mix was then added to the single nanopore experimental system. Experiments were carried out for two hours following a potential flip process (+120 mV for 30 mins, then -100 mV for 2 seconds and then 0 mV for 2 seconds) and helicase-controlled DNA movement was monitored.

3.5 Results and Discussion

Helicase controlled DNA movement was observed for the Lambda DNA construct, an example of a helicase-controlled DNA movement is shown in Figure 7. The iSpC3 spacers present in the Lambda DNA construct produced a characteristic block level highlighted by numbers 1-3 in Figure 7. The Y-shaped 2 MuA substrate has four iSpC3 spacers in either strand and the hairpin 2 MuA substrate also has four iSpC3 spacers, each iSpC3 spacer allows more current to flow as that region of the Lambda DNA construct translocates through the nanopore. If the sample preparation occurred successfully then the iSpC3 spacer events will be observed at the beginning of the Lambda DNA construct, in the middle (making the transition between the sense and antisense sequences) and at the end. Figure 7 clearly shows three instances of increased current flow which correspond to the iSpC3 spacer regions. Therefore, the sample preparation procedure effectively introduced MuA substrates into the Lambda DNA to produce the Lambda DNA constructs shown in Figure 6 (b). Upon capture of the DNA by the nanopore, the enzyme (labelled A) is past the dsDNA region (where the tether hybridises to the construct) and the spacers by the force of the applied potential across the nanopore. Owing to the use of the inosines in the Y-shaped 2 and the hairpin 2 MuA substrates, the steps in the sample preparation procedure were reduced as once the MuA substrates had been inserted all that was necessary was to close the nicks in the double-stranded DNA constructs.

Example 4

This Example compares the ability of a TrwC Cba monomer (SEQ ID NO: 17), to control the movement of intact DNA strands (attached to the 5' end of SEQ ID NO: 23 is 28 iSpC3 spacers units the last of which has an additional two T's attached to the 5' end of the spacer group, attached to the 3' end of SEQ ID NO: 23 is a further four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 24, where SEQ ID NO: 12 is hybridised to a region of

SEQ ID NO: 23) through a nanopore, to that of the TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 17 with the mutation Q276C, with one monomer unit being linked to the other via position 276 of each monomer unit using a 3.4 kDa PEG linker). The DNA construct used in this example is shown in Figure 22 (the helicase is capable of binding to the region labelled B). When the DNA construct is captured by the nanopore, the applied potential across the nanopore moves the enzyme past the dsDNA region (where the tether hybridises to the construct) and spacers and helicase controlled DNA movement is observed.

Upon comparison of the helicase controlled movement of the monomer with the dimer, it was observed that the dimer resulted in a greater percentage of long dwell helicase-controlled DNA movement (a long dwell movement is a helicase-controlled DNA movement which is more than three standard deviations away from the mean of the major population of helicase-controlled DNA movements) than the monomer.

Materials and Methods

Prior to setting up the experiment, the DNA (1 nM, attached to the 5' end of SEQ ID NO: 23 is 28 iSpC3 spacers units the last of which has an additional two T's attached to the 5' end of the spacer group, attached to the 3' end of SEQ ID NO: 23 is a further four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 24, where SEQ ID NO: 12 is hybridised to a region of SEQ ID NO: 23) and the enzyme (either a TrwC Cba monomer (1 nM, SEQ ID NO: 17) or TrwC Cba Q276C-3.4kDa dimer (0.3 nM, where each monomer unit comprises SEQ ID NO: 17 with the mutation Q276C, with one monomer unit being linked to the other via position 276 of each monomer unit using a 3.4 kDa PEG linker)) were pre-incubated together for >16 hours.

Electrical measurements were acquired from single MspA nanopores MS(G75S/G77S/L88N/Q126R)8 MspA (SEQ ID NO: 2 with the mutations G75S/G77S/L88N/Q126R) inserted in block copolymer in buffer (625 mM KCl, 100 mM Hepes, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). MgCl₂ (10 mM) and dTTP (5 mM) were mixed together with buffer (625 mM KCl, 100 mM Hepes, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) and then added to the DNA (construct described previously), enzyme pre-mix (either a TrwC Cba monomer (1 nM, SEQ ID NO: 17) or TrwC Cba Q276C-3.4kDa dimer (1 nM, where each monomer unit comprises SEQ ID NO: 17 with the mutation Q276C, with one monomer unit being linked to the other via position 276 of each monomer unit using a 3.4 kDa PEG linker)). After achieving a single pore in the bilayer, the pre-mix was added to the single nanopore experimental system. Experiments

were carried out at a constant potential of +120 mV and helicase-controlled DNA movement was monitored.

Results and Discussion

Helicase controlled DNA movement was observed for the helicase TrwC Cba monomer (SEQ ID NO: 17) and TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 17 with the mutation Q276C, with one monomer unit being linked to the other via position 276 of each monomer unit using a 3.4 kDa PEG linker). Upon capture of the DNA construct by the nanopore the helicase was moved past the dsDNA region (where the tether hybridises to the construct) and the spacers and helicase controlled movement was observed.

Of the helicase-controlled DNA movements observed there is a major population which accounts for around 95% of movements detected, however, there is a small percentage of movements which are significantly longer in dwell time (more than three standard deviations away from the mean of the major population of helicase-controlled DNA movements). These longer movements allow improved data analysis. When the TrwC Cba Q276C-3.4kDa dimer (1 nM) was used to control DNA movement then a much higher percentage (20% for the TrwC Cba Q276C-3.4kDa dimer in comparison to and 5% for the TrwC Cba monomer) of these longer dwell time movements (more than three standard deviations away from the mean of the major population of helicase-controlled DNA movements) was observed. The use of the dimer helicase provides an advantage over the monomer as it allows improved data analysis in the nanopore sequencing system.

Example 5

This Example illustrates that Sp9 spacer units can be used to stall the movement of Hel308 Mbu (SEQ ID NO: 28) (when provided with both ATP and MgCl₂) in a fluorescence based assay for testing enzyme activity.

Materials and Methods

Three different custom fluorescent substrates (A = (control strands containing no spacers) SEQ ID NOs: 25 and 26, B = (strand containing a single Sp9 spacer) SEQ ID NO: 27 attached at its 3' end by one Sp9 spacers to the 5' end of SEQ ID NO: 29 and hybridised to SEQ ID NO: 26, C = (strand containing four Sp9) SEQ ID NO: 27 attached at its 3' end by four Sp9 spacers to the 5' end of SEQ ID NO: 29 and hybridised to SEQ ID NO: 26) were used to assay the ability of Hel308 Mbu (SEQ ID NO: 28) to displace hybridised dsDNA. FAM labelled DNA (for fluorescent substrate A = SEQ ID NO: 25, B = SEQ ID NO: 27 attached by its 3' end to one

sp9 spacer which is attached to the 3' end of SEQ ID NO: 29, C = SEQ ID NO: 27 attached by its 3' end to four sp9 spacers which are attached to the 3' end of SEQ ID NO: 29) is annealed to the partially complementary strand which has a black-hole quencher attached to its 3' end (SEQ ID NO: 26) in a one to one ratio (1 μ M of each strand) in 400mM KCl, 100mM HEPES pH8, 10mM MgCl₂, 1mg/ml BSA. The strands were annealed at room temperature for 30 minutes. The annealed DNA (A = SEQ ID NOs: 25 and 26, B = SEQ ID NO: 27 attached at its 3' end by one Sp9 spacers to the 5' end of SEQ ID NO: 29 and hybridised to SEQ ID NO: 26, C = SEQ ID NO: 27 attached at its 3' end by four Sp9 spacers to the 5' end of SEQ ID NO: 29 and hybridised to SEQ ID NO: 26) was diluted to 50 nM in 400 mM KCl, 100 mM HEPES pH8, 10 mM MgCl₂, 1 mg/ml BSA, 1 mM ATP (1 μ M capture DNA (SEQ ID NO: 27 also present). A sample of Hel308 Mbu (SEQ ID NO: 28) was diluted to 475nM in 400mM KCl, 100mM HEPES pH8, 10mM MgCl₂, 1mg/ml BSA. Hel308 Mbu (12 nM) was then assayed (as described below and shown in Figures 8 and 9) against 48.75 nM annealed DNA in 400mM KCl, 100mM HEPES pH8, 10mM MgCl₂, 1mg/ml BSA, 0.975 mM ATP (0.975 μ M capture DNA also present).

The control strand A is shown in Fig. 8, where in 1A) the fluorescent substrate strand (48.75 nM final) has a 3' ssDNA overhang, and a 40 base section of hybridised dsDNA. The upper strand, containing the 3' ssDNA overhang, has a carboxyfluorescein base attached to the 5' end of SEQ ID NO: 25, and the hybridised complement has a black-hole quencher (BHQ-1) base attached to the 3' end of SEQ ID NO: 26. When hybridised the fluorescence from the fluorescein is quenched by the local BHQ-1, and the substrate is essentially non-fluorescent. 1 μ M of a capture strand (SEQ ID NO: 27) that is part-complementary to the lower strand of the fluorescent substrate is included in the assay. As shown in 2A), in the presence of ATP (0.975 mM) and MgCl₂ (10 mM), helicase (12 nM) added to the substrate binds to the 3' tail of the fluorescent substrate, moves along the upper strand, and displaces the complementary strand. As shown in 3A), once the complementary strand with BHQ-1 is fully displaced the fluorescein on the major strand fluoresces. As shown in 4A), the displaced strand preferentially anneals to an excess of capture strand to prevent re-annealing of initial substrate and loss of fluorescence. Figure 9 shows the steps of the assay for strands B (1B-3B) and C (1C-3C), for these strands the Sp9 spacers stall the helicase and prevent it from separating the strand with the fluorescein attached from the strand with the black hole quencher attached.

Results and Discussion

The graph in Figure 10 shows the initial rate of activity in buffer solution (100 mM Hepes pH8.0, 0.975 mM ATP, 10 mM MgCl₂, 1 mg/ml BSA, 48.75 nM fluorescent substrate DNA (substrates A, B and C as discussed above), 0.975 μ M capture DNA (SEQ ID NO: 27)) for

the Hel308 Mbu (labeled A in Figure 8 and 9; SEQ ID NO: 28) at 400 mM of KCl. At the salt concentration investigated the Hel308 Mbu (SEQ ID NO: 28) exhibited dsDNA turnover of the control strand A. However, Figure 10 clearly indicates that for both constructs which have Sp9 spacers in the sequence (B (one Sp9 spacer) and C (four Sp9 spacers)) the dsDNA turnover was abolished. This indicates that in the presence of ATP and MgCl₂ the Sp9 spacers stall the Hel308 Mbu enzyme in free solution.

Example 6

This Example illustrates that idSp groups can be used to stall the movement of Hel308 Mbu (SEQ ID NO: 28) (when provided with both ATP and MgCl₂) in a fluorescence based assay for testing enzyme activity.

Materials and Methods

Four different custom fluorescent substrates (D = (control strand containing no spacers) SEQ ID NOs: 32 and 26, E = (strand containing a single idSp spacer) SEQ ID NO: 27 attached at its 3' end by one idSp group to the 5' end of SEQ ID NO: 30 and hybridised to SEQ ID NO: 26, F = (strand containing four idSp) SEQ ID NO: 27 attached at its 3' end by four idSp groups to the 5' end of SEQ ID NO: 31 and hybridised to SEQ ID NO: 26 and G = (second control strand containing no spacers) SEQ ID NOs: 33 and 26) were used to assay the ability of Hel308 Mbu (SEQ ID NO: 28) to displace hybridised dsDNA. FAM labelled DNA (for fluorescent substrate D = SEQ ID NO: 32, E = SEQ ID NO: 27 attached by its 3' end to one idSp group which is attached to the 3' end of SEQ ID NO: 30, F = SEQ ID NO: 27 attached by its 3' end to four idSp groups which are attached to the 3' end of SEQ ID NO: 31 and G = SEQ ID NO: 33) is annealed to the partially complementary strand which has a black-hole quencher attached to its 3' end (SEQ ID NO: 26) in a 1 to 1.2 ratio (1:1.2 μ M) in 400mM KCl, 100 mM HEPES pH8, 10 mM MgCl₂, 1 mg/ml BSA. The strands were annealed at room temperature for 15 minutes. The annealed DNA (D = SEQ ID NOs: 32 and 26, E = SEQ ID NO: 27 attached at its 3' end by one idSp group to the 5' end of SEQ ID NO: 30 and hybridised to SEQ ID NO: 26, F = SEQ ID NO: 27 attached at its 3' end by four idSp groups to the 5' end of SEQ ID NO: 31 and hybridised to SEQ ID NO: 26 and G = SEQ ID NOs: 33 and 26) was diluted to 50 nM in 400 mM KCl, 100 mM HEPES pH8, 10 mM MgCl₂, 1 mg/ml BSA, 1 mM ATP (1 μ M capture DNA (SEQ ID NO: 27 also present). Hel308 Mbu (12 nM) was then assayed (as described previously in Example 5 (except the DNA constructs are different and contain idSp groups instead of Sp9 spacers) and shown in Figures 8 and 9 (again where the Sp9 groups in these figures are replaced with idSp

groups)) against 48.75 nM annealed DNA in 400 mM KCl, 100 mM HEPES pH8, 10 mM MgCl₂, 1 mg/ml BSA, 0.975 mM ATP (0.975 μ M capture DNA also present).

Results and Discussion

The graph in Figure 11 shows the initial rate of activity in buffer solution (100 mM Hepes pH 8.0, 0.975 mM ATP, 10 mM MgCl₂, 1 mg/ml BSA, 48.75 nM fluorescent substrate DNA (substrates D, E, F and G as discussed above), 0.975 μ M capture DNA (SEQ ID NO: 27)) for the Hel308 Mbu (labelled A in Fig. 8 and 9; SEQ ID NO: 28) at 400 mM of KCl. At the salt concentration investigated the Hel308 Mbu (SEQ ID NO: 28) exhibited dsDNA turnover of the control strands D and G. However, Figure 11 clearly indicates that for the construct which has four idSp spacers in the sequence (F) the dsDNA turnover was abolished. This indicates that in the presence of ATP and MgCl₂ the four idSp groups stall the Hel308 Mbu enzyme in free solution.

Example 7

This Example illustrates a gel based assay that was used to measure the ability of iSpC3 spacers and iSp18 spacers to stall the movement of T4 Dda – E94C/A360C.

Materials and Methods

The annealed DNA complexes (sequences tested are shown in table 11 below) were mixed in a ratio of (1:1, v/v) with T4 Dda – E94C/A360C in 25 mM phosphate pH 8.0, 200 mM KCl giving final concentrations of T4 Dda – E94C/A360C (2000 nM) and DNA (100 nM). The helicase was allowed to bind to the DNA for 2 hours at ambient temperature. Capture strand (SEQ ID NO: 37, 20 μ M) was added to each sample to bind any unbound enzyme and the samples incubated at ambient temperature for 30 mins. Buffer was added to the samples (DNA construct from table 11 = 50 nM, capture DNA (SEQ ID NO: 37) = 10 μ M and T4 Dda – E94C/A360C = 1000 nM) and they were incubated at ambient temperature for one hour (either Buffer 1 = 25 mM phosphate pH 8.0, 200 mM KCl, 20 mM MgCl₂, 10 mM ATP or Buffer 2 = 25 mM phosphate pH 8.0, 1 M KCl, 25 mM potassium ferricyanide(III), 75 mM potassium ferrocyanide, 20 mM MgCl₂, 10 mM ATP). Loading buffer (25mM Phosphate pH8.0, 151.5 mM KCl, 25% Glycerol, 125mM EDTA) is added to each sample to quench the helicase activity. The samples were loaded onto 4-20% TBE gel and the gel run at 160 V for 1.5 hours. The Gel was then stained with SYBR gold in order to observe the DNA bands.

DNA construct Number	DNA sequences which make up the construct
1 (positive control, no spacers)	SEQ ID NO: 34 hybridised to SEQ ID NO: 35
2 (4 x iSpC3 spacers)	SEQ ID NO: 9 attached at its 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
3 (8 x iSpC3 spacers)	SEQ ID NO: 9 attached at its 3' end to eight iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
4 (12 x iSpC3 spacers)	SEQ ID NO: 9 attached at its 3' end to twelve iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
5 (1 x iSp18 spacer)	SEQ ID NO: 9 attached at its 3' end to one iSp18 spacer which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
6 (2 x iSp18 Spacer)	SEQ ID NO: 9 attached at its 3' end to two iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
7 (3 x iSp18 spacer)	SEQ ID NO: 9 attached at its 3' end to three iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
8 (4 x iSp18 spacer)	SEQ ID NO: 9 attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35
9 (5 x iSp18 spacer)	SEQ ID NO: 9 attached at its 3' end to five iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 hybridised to SEQ ID NO: 35

Table 11

Results and Discussion

Figure 12 show the gel of the positive control experiment where the DNA construct contains no spacers to stall the helicase (T4 Dda – E94C/A360C). Lane 2 of Figure 14 shows that the up to 5 helicase enzymes can bind onto the single-stranded section of SEQ ID NO: 34.

65

Upon the addition of ATP and MgCl₂, the higher bands corresponding to the dsDNA construct bound to multiple enzymes disappears and band corresponding to the ssDNA construct (labelled X and corresponding to SEQ ID NO: 34 only) increases in intensity. This indicates that when the helicase is provided with fuel it moves along the DNA and displaces the hybridised

5 complementary strand (SEQ ID NO: 35).

Figure 15 shows an example of a gel showing the enzyme activity experiment for DNA constructs 7-9 of table 11 (e.g. DNA which has three, four or five iSp18 spacers just before the ssDNA/dsDNA junction). Lanes 1-4 correspond to three iSp18 spacers, lanes 5-7 correspond to four iSp18 spacers and lanes 9-12 correspond to five iSp18 spacers. Lanes 2, 6 and 10 show that, similarly to the control, up to five helicase enzymes can bind onto the single-stranded section of the DNA constructs 7-9 in table 11. Upon the addition of ATP and MgCl₂ the helicases will be provided with the necessary components in order to move along the DNA strand. Lanes 3-4, 7-8 and 11-12 show the various DNA constructs under two different buffer conditions (buffer 1 and 2) after the addition of ATP and MgCl₂. Two bands are clearly visible at the region labelled 1Y and 1X. 1Y corresponds to the dsDNA construct with one helicase stalled and bound. This shows that three, four and five iSp18 spacers are able to stall the helicase under the conditions tested. 1X corresponds to the ssDNA construct with one helicase stalled and bound. This band has resulted from multiple helicases binding to the DNA construct and the helicases pushing the front helicase past the stalling groups, thus displacing the short complementary strand (SEQ ID NO: 35). However, this band still has one helicase bound as a single helicase cannot move past the iSp18 spacers (e.g. species E in figure 13 results). In the case of 4 and 5 iSp18 spacers being used to stall the helicase there are faint bands at level 2Y. This shows that under the conditions tested, when 4/5 iSp18 spacers are used it is possible to stall the movement of up to two helicases.

25 Of the other spacer combinations investigated (entries 2-6 of table 11) in at least one of the buffer conditions tested both iSpC3 and iSp18 spacers were capable of stalling one helicase. Of the two buffer conditions tested, generally more efficient stalling was observed for buffer 2 than buffer 1. The greater the number of spacers included the more efficient the stalling of the helicases under the conditions tested.

30

Example 8

This Example investigates the number of bases needed to control the binding of only one or two T4 Dda – E94C/A360C helicases in a particular region.

35

Materials and Methods

DNA constructs (1 μ M or 100 nM final concentration) detailed below in Table 12 were incubated in appropriate buffer with serially diluted T4 Dda – E94C/A360C. The samples were then loaded on 4-20% TBE gels and run at 160 V for 90 minutes. The gels containing entries 1-6 were then stained using SYBR. Figure 16 shows the type of DNA construct used in this experiment. The length of the region labelled 1 is varied from 2 to 50 to optimise conditions so that it is possible to control the number of T4 Dda – E94C/A360C helicases which can bind to the DNA.

Entry	DNA Construct	Buffer	Incubation Conditions	Concentrations of T4 Dda – E94C/A360C
1	1 μ M, Five iSpC3 spacers attached to the 5' end of SEQ ID NO: 37, which is attached at its 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH 8.0	overnight at room temperature	5000 nM, 2500 nM, 1250 nM, 625 nM, 312.5 nM, 0 M
2	1 μ M, Five iSpC3 spacers attached to the 5' end of SEQ ID NO: 38, which is attached at its 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH 8.0	overnight at room temperature	5000 nM, 2500 nM, 1250 nM, 625 nM, 312.5 nM, 0 M
3	1 μ M, Five iSpC3 spacers attached to the 5' end of SEQ ID NO: 9, which is attached at its 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH 8.0	overnight at room temperature	5000 nM, 2500 nM, 1250 nM, 625 nM, 312.5 nM, 0 M
4	1 μ M, Five iSpC3 spacers attached to the 5' end of two thymines, which are attached at the 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH 8.0	1.5 hours at room temperature	3750 nM, 1875 nM, 938 nM, 469 nM, 235 nM, 0 M
5	1 μ M, Five iSpC3 spacers attached to the 5' end of four thymines, which are attached at the 3' end to	25 mM phosphate, 151.5 mM	1.5 hours at room temperature	3750 nM, 1875 nM, 938 nM, 469 nM, 235 nM, 0 M

68

	four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	KCl pH 8.0		
6	1 μ M, Five iSpC3 spacers attached to the 5' end of eight thymines, which are attached at the 3' end to four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH 8.0	1.5 hours at room temperature	3750 nM, 1875 nM, 938 nM, 469 nM, 235 nM, 0 M
7	100 nM, Five iSpC3 spacers attached to the 5' end of SEQ ID NO: 39, which is attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH8.0	Overnight at room temperature	2700 nM, 1350 nM, 675 nM, 337.5 nM, 169 nM, 0 M
8	100 nM, Five iSpC3 spacers attached to the 5' end of SEQ ID NO: 40, which is attached at its 3' end to four iSp18 spacers which are attached to the 5' end of SEQ ID NO: 36 which is hybridised to SEQ ID NO: 35.	25 mM phosphate, 151.5 mM KCl pH 8.0	Overnight at room temperature	2700 nM, 1350 nM, 675 nM, 338 nM, 169 nM, 0 M

Table 12

Results and Discussion

Each of the DNA constructs listed in Table 12 were investigated to determine how many enzymes can bind to region 1 (shown in Figure 16) which has been varied from 2 bases to 50 bases. For each of the DNA constructs tested only a single band for unbound DNA was observed when no helicase was added. Under the conditions investigated, construct 4 (which has a binding region of 2 thymine bases) and construct 5 (which has a binding region of 4 thymine bases) were not observed to allow any helicases to bind at any concentration (see table 14). Construct 6 (which has a binding region of 8 thymine bases see table 14) and construct 1 (which had a

binding region of 10 thymine bases) was observed to bind one helicase only (see table 13).

Construct 2 which has a binding region of 20 thymine bases allowed two enzymes to bind at the higher concentrations tested and construct 3 which has a binding region of 50 thymine bases allowed up to 6 enzymes to bind at the highest concentrations tested (see Figure 17 for the gel showing this experiment see table 13). Constructs 1-6 used iSpC3 spacers to prevent binding of the helicase at the junction of the ssDNA/dsDNA region. Constructs 7 and 8 used iSp18 spacers to prevent binding of the helicase at the junction of the ssDNA/dsDNA region. Construct 7 (which has a binding region of 16 thymines) and construct 8 (which has a binding region of 18 thymines) both allowed up to two helicases to bind under the conditions tested (see table 15).

Therefore, binding regions greater than 8 thymine bases long allow at least one enzyme to bind to the DNA constructs.

DNA Construct Entry No	Concentration of T4 Dda - E94C/A360C				
	5000	2500	1250	625	312.5
1	One enzyme bound only	One enzyme bound only	One enzyme bound only	One enzyme bound only. Faint band for unbound DNA.	One enzyme bound only. Unbound DNA
2	Two enzymes bound.	Two enzymes bound.	Two enzymes bound.	Two enzymes bound. One enzyme bound.	One enzyme bound only. Unbound DNA
3 (shown in Figure 17)	Six enzymes bound (lane 1 Figure 17).	Five enzymes bound (lane 2 Figure 17).	Faint band for two enzymes bound. Three enzymes bound. Four enzymes bound. Five enzymes bound (lane 3	One enzyme bound. Two enzymes bound. Three enzymes bound. Faint band for four enzymes bound (lane 4 Figure 17).	One enzyme bound. Two enzymes bound. Faint band for three enzymes bound (lane 5 Figure 17).

Table 13

DNA Construct Entry No	Concentration of T4 Dda - E94C/A360C				
	3750	1875	938	469	235
4	Unbound DNA only.	Unbound DNA only.	Unbound DNA only.	Unbound DNA only.	Unbound DNA only.
5	Unbound DNA only.	Unbound DNA only.	Unbound DNA only.	Unbound DNA only.	Unbound DNA only.
6	One enzyme bound. Faint band for unbound DNA	One enzyme bound. Unbound DNA	Faint band for one enzyme bound. Unbound DNA	Faint band for one enzyme bound. Unbound DNA	Unbound DNA only.

Table 14

DNA Construct Entry No	Concentration of T4 Dda - E94C/A360C				
	2700	1350	675	338	169
7	Two enzymes bound.	Two enzymes bound. One enzyme bound. Faint band for unbound DNA	Faint band for two enzymes bound. One enzyme bound. Faint band for unbound DNA	One enzyme bound. Unbound DNA.	Faint band for one enzyme bound. Unbound DNA
8	Two enzymes bound.	Two enzymes bound. Faint band for one enzyme bound.	Faint band for two enzymes bound. One enzyme bound.	Faint band for one enzyme bound. Unbound DNA	Unbound DNA only.

				71	
			Unbound DNA		

Table 15

Example 9

This Example investigates the concentration of T4 Dda – E94C/A360C helicase which when added to the DNA construct X (described and shown in Figure 18) results in the binding of two helicases.

Materials and Methods

Two DNA constructs were tested one which has a complementary strand of DNA which is not forked (SEQ ID NO: 42 is hybridised to the DNA construct shown in Figure 18) and one which is forked (SEQ ID NO: 12 (which has 6 iSp18 spacers attached to the 3' end) is hybridised to the DNA construct shown in Figure 18). The DNA strands shown in Table 16 were annealed at 1 μ M in 25 mM phosphate pH 8.0, 1515.5 mM KCl (a 10% excess of complementary strands SEQ ID NO: 43 (in entries 9 and 10 below) and SEQ ID NO: 12 (entry 10 below) or SEQ ID NO: 42 (entry 9 below) was used).

DNA Construct Entry No:	DNA Strands Annealed
9	25 SpC3 spacers are attached to the 5' end of SEQ ID NO: 37 which is attached at its 3' end to two iSp18 spacers. The iSp18 spacers are attached to another DNA fragment of SEQ ID NO: 37 which again is attached at its 3' end to the another two iSp18 spacers. The second instance of iSp18 spacers is attached to the 5' end of SEQ ID NO: 10 which is attached its 3' end to four nitroindoles. The four nitroindoles are then attached to the 5' end of SEQ ID NO: 41. SEQ ID NO: 41 is hybridised to the complementary strand SEQ ID NO: 43. SEQ ID NO: 10 is hybridised to the non-forked complementary strand SEQ ID NO: 42.
10	25 SpC3 spacers are attached to the 5' end of SEQ ID NO: 37 which is attached at its 3' end to two iSp18 spacers. The iSp18 spacers are attached to another DNA fragment of SEQ ID NO: 37 which again is attached at its 3' end to the another two iSp18 spacers. The second instance of iSp18 spacers is attached to the 5' end of SEQ ID NO: 10 which is attached its 3' end to four nitroindoles. The four nitroindoles are then attached to the 5' end of SEQ ID NO: 41. SEQ ID NO: 41 is hybridised to the complementary strand SEQ ID NO: 43. SEQ ID NO: 10 is hybridised to the forked complementary strand SEQ ID NO: 12 which has 6 iSp18 spacers attached at its 3' end.

Table 16

T4 Dda – E94C/A360C was buffer exchanged into 25 mM phosphate pH 8.0, 151.5 mM KCl and serially diluted. The helicase and DNA were then mixed (1:1, v/v) with the DNA construct samples 9 and 10 described above (final concentration DNA = 100 nM, helicase concentrations investigated = 3800 nM, 1900 nM, 950 nM, 475 nM, 238 nM, 0 nM). The DNA and enzyme volumes were then incubated at ambient temperature for 1.5 hours. Dye free loading buffer (5x, 7.5 µL) was added to each sample (30 µL). Each sample (37.5 µL) was then loaded onto 4-20% TBE gel and run at 160 V for 90 minutes. The gel was then stained using SYBR.

10 **Results and Discussion**

The two DNA constructs listed in Table 16 were investigated to determine what concentration of T4 Dda – E94C/A360C helicase is required in order to promote binding of two helicases. For each of the DNA constructs tested only a single band for unbound DNA was observed when no helicase was added. Under the conditions investigated, both constructs 9 (non-

forked construct) and 10 (forked) observed binding of one helicase from 238 nM helicase and two enzymes from 475 nM and higher. As the concentration of enzyme was increased the band corresponding to two enzymes bound increased in intensity. The design of the DNA construct shown in Figure 18 allows the binding of only two enzymes at concentrations as high as 3800 nM. Therefore, the spacers investigated do not allow the helicase to bind to them when the constructs are pre-incubated.

Example 10

This Example shows how the T4 Dda – E94C/A360C is stalled by four iSp18 spacers in free solution until the construct (DNA construct X1) is captured by the nanopore. Upon capture the force of the applied potential moves the enzyme T4 Dda – E94C/A360C past the stalling spacer and allows enzyme controlled DNA movement of the lambda construct through the nanopore.

Materials and Methods

Prior to setting up the experiment, the DNA construct X1 (0.13 μ L, 100nM) and T4 Dda – E94C/A360C (15.6 μ L, 250nM) were pre-incubated together for 1 hour at room temperature in buffer (50 mM potassium phosphate, 253 mM KCl, pH 8.0).

Electrical measurements were acquired at 30 °C (by placing the experimental system on a cooler plate) from single MspA nanopores (MspA - B2C) inserted in block co-polymer in buffer (600 mM KCl, 25 mM potassium phosphate, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). After achieving a single pore inserted in the block co-polymer, then buffer (1 mL, 600 mM KCl, 25 mM potassium phosphate, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8) was flowed through the system to remove any excess MspA nanopores (MspA - B2C). Potassium ferricyanide (III) (200 μ M final concentration) was added to the DNA (0.1 nM final concentration) enzyme (3 nM final concentration) pre-mix and left to incubate for one minute before $MgCl_2$ (10 mM final concentration) and ATP (1 mM final concentration) were mixed together with buffer (1260 μ L, 600 mM KCl, 25 mM potassium phosphate, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). This experimental mix was then added to the single nanopore experimental system. Experiments were carried out for six hours following a potential flip process (+180 mV for 2 s, then 0 V for 2 s, then -120 mV for 3600s (x6 repeats) applied at the cis side) and helicase-controlled DNA movement was monitored.

Results and Discussion

The DNA construct is shown in Figure 20. The T4 Dda – E94C/A360C is able to bind on to the region of the construct labelled A (SEQ ID NO: 9) when pre-incubated with the DNA construct. However, the enzyme is not able to motor past the stalling groups (four iSp18 spacers) when in free solution (labelled as black boxes). Therefore, the enzyme is stalled at the iSp18 spacers (labelled B in Figure 1) until the DNA construct is captured by the nanopore. Once captured by the nanopore, the force of the applied potential moves the enzyme T4 Dda – E94C/A360C past the stalling spacer (four iSp18 spacers) and allows enzyme controlled DNA movement of the DNA construct X1 construct through the nanopore. An example of a helicase-controlled DNA movement is shown in Figure 21. The section labelled 1 showed when the pT region of SEQ ID NO 38 translocated through the nanopore under control of the helicase. The section labelled 2 showed when the iSp18 spacer translocated through the nanopore under the control of the helicase. The helicase was observed to have a defined start point (resulting from the stalling of the helicase by the iSp18 spacers) as the helicase events observed showed the pT region and the iSp18 signal at the beginning of the helicase controlled DNA movements.

CLAIMS

1. A method of moving one or more stalled helicases past one or more spacers in a polynucleotide, comprising contacting (a) the one or more stalled helicases and the polynucleotide with a transmembrane pore and (b) applying a potential across the pore and thereby moving the one or more helicases past the one or more spacers on the polynucleotide.
2. A method of controlling the movement of a target polynucleotide through a transmembrane pore, comprising:
 - (a) providing the target polynucleotide with one or more spacers;
 - (b) contacting the target polynucleotide with one or more helicases such that the one or more helicases stall at the one or more spacers;
 - (c) contacting the target polynucleotide and the one or more stalled helicases with the pore; and
 - (d) applying a potential across the pore such that the one or more helicases move past the one or more spacers and control the movement of the target polynucleotide through the pore.
3. A method of characterising a target polynucleotide, comprising:
 - (a) carrying out the method of claim 2; and
 - (b) taking one or more measurements as the polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the polynucleotide and thereby characterising the target polynucleotide.
4. A method according to claim 2 or 3, wherein providing the target polynucleotide with one or more spacers comprises modifying the target polynucleotide so that it comprises one or more spacers.
5. A method according to claim 3 or 4, wherein the one or more characteristics are selected from (i) the length of the target polynucleotide, (ii) the identity of the target polynucleotide, (iii) the sequence of the target polynucleotide, (iv) the secondary structure of the target polynucleotide and (v) whether or not the target polynucleotide is modified.
6. A method according to claim 5, wherein the target polynucleotide is modified by methylation, by oxidation, by damage, with one or more proteins or with one or more labels, tags or spacers.

7. A method according to any one of claims 3 to 6, wherein the one or more characteristics of the target 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 the one or more helicases control the movement of the target polynucleotide through the pore with the field resulting from the applied potential.
10. A method according to any one of the preceding claims, wherein the one or more spacers have a different structure from the polynucleotide.
11. A method according to any one of the preceding claims, wherein the polynucleotide is desoxyribonucleic acid (DNA) or ribonucleic acid (RNA) and the one or more spacers comprise peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or a synthetic polymer with nucleotide side chains.
12. A method according to any one of the preceding claims, wherein the one or more spacers comprise one or more nitroindoles, one or more inosines, one or more acridines, one or more 2-aminopurines, one or more 2-6-diaminopurines, one or more 5-bromo-deoxyuridines, one or more inverted thymidines (inverted dTs), one or more inverted dideoxy-thymidines (ddTs), one or more dideoxy-cytidines (ddCs), one or more 5-methylcytidines, one or more 5-hydroxymethylcytidines, one or more 2'-O-Methyl RNA bases, one or more Iso-deoxycytidines (Iso-dCs), one or more Iso-deoxyguanosines (Iso-dGs), one or more C3 groups, one or more photo-cleavable (PC) groups, one or more hexandiol groups, one or more spacer 9 (iSp9) groups, one or more spacer 18 (iSp18) groups, a polymer or one or more thiol connections.
13. A method according to claim 12, wherein the polymer is a polypeptide or a polyethylene glycol (PEG).
14. A method according to any one of claims 1 to 11, wherein the one or more spacers comprise one or more abasic nucleotides.

15. A method according to any one of the preceding claims, wherein the one or more spacers comprise one or more chemical groups which cause the one or more helicases to stall.
16. A method according to claim 15, wherein the one or more chemical groups are one or more fluorophores, streptavidin and/or biotin, cholesterol, methylene blue, a dinitrophenol (DNP), digoxigenin and/or anti-digoxigenin or a dibenzylcyclooctyne group.
17. A method according to any one of the preceding claims, wherein the one or more spacers are capable of stalling the one or more helicases in the presence of free nucleotides and/or the presence of a helicase cofactor.
18. A method according to any one of the preceding claims, wherein the the one or more spacers are capable of stalling the one or more helicases in a salt concentration of about 100 mM or lower.
19. A method according to any one of the preceding claims, wherein the method comprises moving two or more stalled helicases past each spacer.
20. A method according to claim 19, wherein the two or more helicases are attached to one another.
21. A method according to claim 20, wherein the two or more helicases are covalently attached to one another.
22. A method according to any one of the preceding claims, wherein the one or more spacers are included in the polynucleotide and/or are not part of one or more blocking molecules hybridised to the polynucleotide.
23. A method according to any one of the preceding claims, wherein at least a portion of the polynucleotide is double stranded.
24. A method according to claim 23, wherein the one or more spacers are included in a single stranded region or a non-hybridised region of the polynucleotide.

25. A method according to claim 24, wherein the single stranded region comprises a leader sequence which preferentially threads into the pore.
26. A method according to any one of claims 23 to 25, wherein the two strands of the double stranded portion are linked using a bridging moiety.
27. A method according to claim 26, wherein the one or more spacers are included in the bridging moiety.
28. A method according to any one of the preceding claims, wherein the one or more helicases are (a) one or more Hel308 helicases, one or more RecD helicases, one or more XPD helicases or one or more Dda helicases (b) one or more helicases derived from any of the helicases in (a); or (c) a combination of any of the helicases in (a) and/or (b).
29. A method according to any one of the preceding claims, wherein the pore is a transmembrane protein pore or a solid state pore.
30. A method according to claim 29, wherein the transmembrane protein pore is derived from a hemolysin, leukocidin, *Mycobacterium smegmatis* porin A (MspA), MspB, MspC, MspD, lysenin, outer membrane porin F (OmpF), outer membrane porin G (OmpG), outer membrane phospholipase A, *Neisseria* autotransporter lipoprotein (NalP) and WZA.
31. A method according to claim 30, wherein the transmembrane protein pore is:
(a) formed of eight identical subunits as shown in SEQ ID NO: 2 or (b) a variant thereof in which one or more of the eight subunits has at least 50% homology to SEQ ID NO: 2 based on amino acid identity over the entire sequence and retains pore activity; or
(c) α -hemolysin formed of seven identical subunits as shown in SEQ ID NO: 4 or (d) a variant thereof in which one or more of the seven subunits has at least 50% homology to SEQ ID NO: 4 based on amino acid identity over the entire sequence and retains pore activity.
32. Use of a transmembrane pore and an applied potential to move one or more stalled helicases past one or more spacers on a polynucleotide.
33. Use of one or more spacers to stall one or more helicases on a polynucleotide before the polynucleotide is contacted with a transmembrane pore.

34. A method of controlling the loading of one or more helicases on a target polynucleotide, comprising:

(a) providing the target polynucleotide with one or more spacers; and

(b) contacting the target polynucleotide provided in (a) with the one or more helicases such that the one or more helicases bind to the target polynucleotide and stall at each spacer.

35. A method according to claim 34, wherein step (a) comprises modifying the target polynucleotide so that it comprises one or more spacers.

36. A method according to claim 34 or 35, wherein the target polynucleotide is provided with (L-S)_n or (S-L)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer and n is a whole number.

37. A method according to any one of claims 34 to 36, wherein step (b) comprises contacting the target polynucleotide provided in (a) with the one or more helicases such that the one or more helicases bind to the target polynucleotide and one helicase stalls at each spacer.

38. A method according to claim 37, wherein step (a) comprises providing the target polynucleotide with (L¹-S)_n or (S-L¹)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide which is only long enough for one helicase to bind, S is a spacer and n is a whole number.

39. A method according to any one of claims 34 to 36, wherein step (b) comprises contacting the target polynucleotide provided in (a) with the one or more helicases such that the one or more helicases bind to the target polynucleotide and two helicases stall at each spacer.

40. A method according to claim 39, wherein step (a) comprises providing the target polynucleotide with (L²-S)_n or (S-L²)_n in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide which is only long enough for two helicases to bind, S is a spacer and n is a whole number..

41. A method according to claim 39 or 40, wherein the two helicases stalled at each spacer are different from one another.

42. A method according to any one of claims 39 to 41, wherein the two helicases stalled at each spacer are attached to one another or covalently attached to one another.
43. A method according to claim 41, wherein the method comprises:
- (a) providing the target polynucleotide with $(L^2-S)_n$ or $(S-L^2)_n$ in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide which is only long enough for two helicases to bind, S is a spacer and n is a whole number;
 - (b) hybridising a blocking polynucleotide to part of each region L^2 so that the remaining part of each region L^2 is only long enough to bind one helicase;
 - (c) contacting the target polynucleotide produced in (b) with one or more helicases such that one helicase binds to the remaining part of each region L^2 ;
 - (d) providing the one or more bound helicases in (c) with free nucleotides and a helicase cofactor such that they remove each blocking polynucleotide and stall at each spacer S; and
 - (e) contacting the target polynucleotide produced in (d) with one or more helicases which are different from those used in (c) such that one different helicase binds to each region L^2 and is stalled by each spacer and each helicase stalled in (d).
44. A method according to any one of claims 34 to 43, wherein the target polynucleotide is provided with two or more spacers.
45. A method according to any one of claims 34 to 44, wherein the one or more stalled helicases are moved past the one or more spacers using a method according to any one of claims 1 to 31.
46. An adaptor for controlling the movement of a target polynucleotide, wherein the adaptor comprises (a) $(L-S-D)_n$ or $(D-S-L)_n$ in the 5' to 3' direction, wherein L is a single stranded polynucleotide or a non-hybridised polynucleotide, S is a spacer and D is a double stranded polynucleotide and wherein n is a whole number and (b) one or more helicases stalled on each adaptor.
47. An adaptor according to claim 46, wherein L is capable of binding only one helicase (L^1) or capable of binding only two helicases (L^2).
48. An adaptor according to claim 47, wherein n is 1 and one or two helicases are stalled on the adaptor.

49. A kit for controlling the movement of a target polynucleotide, wherein the kit comprises (a) one or more spacers, (b) one or more helicases and (c) a transmembrane pore.
50. A kit according to claim 49, wherein the one or more spacers are as defined in any one of claims 15 to 18.
51. A kit according to claim 49 or 50, wherein the kit comprises an adaptor according to any one of claims 46 to 48.