Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2879355 C 2021/09/21

(11)(21) 2 879 355

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2013/07/18
(87) Date publication PCT/PCT Publication Date: 2014/01/23
(45) Date de délivrance/Issue Date: 2021/09/21
(85) Entrée phase nationale/National Entry: 2015/01/15
(86) N° demande PCT/PCT Application No.: GB 2013/051928
(87) N° publication PCT/PCT Publication No.: 2014/013262
(30) Priorités/Priorities: 2012/07/19 (US61/673,446); 2013/03/08 (US61/774,694)

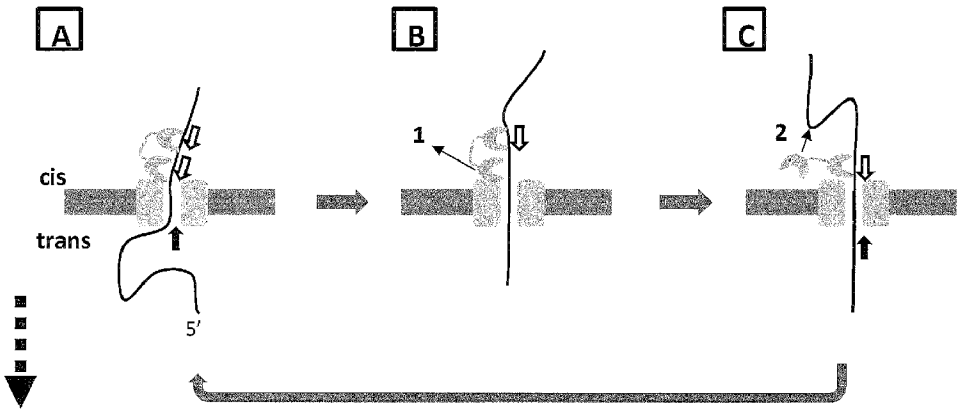
(51) Cl.Int./Int.Cl. *C12Q 1/68* (2018.01), *C12M 1/34* (2006.01), *C12N 15/55* (2006.01), *C12N 9/14* (2006.01), *C12Q 1/6869* (2018.01)

(72) Inventeurs/Inventors:
HERON, ANDREW, GB;
CLARKE, JAMES, GB;
MOYSEY, RUTH, GB;
WALLACE, ELIZABETH JAYNE, GB;
BRUCE, MARK, GB;
JAYASINGHE, LAKMAL, GB;
...

(73) Propriétaire/Owner:

(54) Titre : CONSTRUCTION COMPRENANT UNE HELICASE ET SON UTILISATION DANS LA CARACTERISATION DES POLYNUCLEOTIDES

(54) Title: HELICASE CONSTRUCT AND ITS USE IN CHARACTERISING POLYNUCLEOTIDES



(57) Abrégé/Abstract:
The invention relates to methods using constructs comprising a helicase and an additional polynucleotide binding moiety. The helicase is attached to the polynucleotide binding moiety and the construct has the ability to control the movement of a

CA 2879355 C 2021/09/21
(11)(21) 2 879 355
(13) C

(72) Inventeurs(suite)/Inventors(continued): CAPROTTI, DOMENICO, GB; SOEROES, SZABOLCS, GB; MCNEILL, LUKE, GB; MISCA, MIHAELA, GB

(73) Propriétaires(suite)/Owners(continued):OXFORD NANOPORE TECHNOLOGIES LIMITED, GB

(74) Agent: RIDOUT & MAYBEE LLP

(57) Abrégé(suite)/Abstract(continued):
polynucleotide. The constructs can be used to control the movement of polynucleotides and are particularly useful for sequencing polynucleotides.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 January 2014 (23.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/013262 A1

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

(21) **International Application Number**:
PCT/GB2013/051928

(22) **International Filing Date**:
18 July 2013 (18.07.2013)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/673,446 19 July 2012 (19.07.2012) US
61/774,694 8 March 2013 (08.03.2013) US

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

(72) **Inventors**: **HERON, Andrew**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **CLARKE, James**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **MOYSEY, Ruth**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **WALLACE, Jayne**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **BRUCE, Mark**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **JAYASINGHE, Lakmal**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **CAPROTTI, Domenico**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **SOROS, Szabolcs**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **MCNEILL, Luke**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB). **MISCA, Mihaela**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford Oxfordshire OX4 4GA (GB).

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

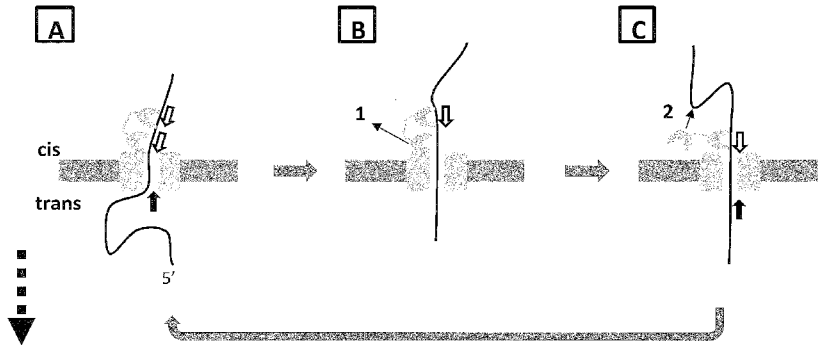
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP,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

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,

[Continued on next page]

(54) **Title**: ENZYME CONSTRUCT

Fig. 11



(57) **Abstract**: The invention relates to methods using constructs comprising a helicase and an additional polynucleotide binding moiety. The helicase is attached to the polynucleotide binding moiety and the construct has the ability to control the movement of a polynucleotide. The constructs can be used to control the movement of polynucleotides and are particularly useful for sequencing polynucleotides.

WO 2014/013262 A1

WO 2014/013262 A1

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

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

HELICASE CONSTRUCT AND ITS USE IN CHARACTERISING POLYNUCLEOTIDES

Field of the invention

The invention relates to methods using constructs comprising a helicase and an additional polynucleotide binding moiety. The helicase is attached to the polynucleotide binding moiety and the construct has the ability to control the movement of a polynucleotide. The constructs can be used to control the movement of polynucleotides and are particularly useful for sequencing polynucleotides.

Background of the invention

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

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

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

Summary of the invention

The inventors have surprisingly demonstrated that attaching an additional polynucleotide binding moiety to a helicase, such as attaching two or more helicases together, results in a construct that has an improved ability to control the movement of a polynucleotide. In particular, the inventors have surprisingly demonstrated that such constructs will strongly bind to a long polynucleotide, such as a polynucleotide comprising 400 nucleotides or more, and will control the movement of most of, if not all of, the polynucleotide without disengaging. This allows the effective control of the movement of the polynucleotide, especially during Strand Sequencing.

Date Reçue/Date Received 2020-09-29

Accordingly, the invention provides a method of characterising a target polynucleotide, comprising:

(a) contacting the target polynucleotide with a transmembrane pore and a construct comprising a helicase and an additional polynucleotide binding moiety, wherein the helicase is attached to the polynucleotide binding moiety and the construct has the ability to control the movement of a polynucleotide, such that the construct controls the movement of the target polynucleotide through the pore; and

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

The invention also provides:

- a construct comprising two or more helicases, wherein the helicases are attached together and the construct has the ability to control the movement of a polynucleotide;
- a polynucleotide sequence which encodes a construct of the invention, wherein the two or more helicases are genetically fused;
- a method of controlling the movement of a polynucleotide, comprising contacting the polynucleotide with a construct of the invention and thereby controlling the movement of the polynucleotide;
- a method of forming a sensor for characterising a target polynucleotide, comprising forming a complex between a pore and a construct as defined above and thereby forming a sensor for characterising the target polynucleotide;
- a sensor for characterising a target polynucleotide, comprising a complex between a pore and a construct as defined above;
- use of a construct as defined above to control the movement of a target polynucleotide through a pore;
- a kit for characterising a target polynucleotide comprising (a) a pore and (b) a construct as defined above;
- an apparatus for characterising target polynucleotides in a sample, comprising a plurality of pores and a plurality of constructs as defined above; and
- a method of producing a construct of the invention, comprising attaching two or more helicases together and thereby producing the construct.

Description of the Figures

Fig. 1 shows a gel of the monomer and dimers of a number of Hel308 Mbu helicase constructs. Lanes 1 and 6 show an appropriate protein ladder. Lane 2 corresponds to the Hel308 Mbu (R687A/A700C) monomer (SEQ ID NO: 10 with the mutations R687A/A700C), Lane 3 corresponds to Hel308 Mbu(R681A/R687A/A700C) monomer (SEQ ID NO: 10 with the mutations R681A/R687A/A700C), Lane 4 corresponds to Hel308 Mbu(R687A/A700C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutations R687A/A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker) and Lane 5 corresponds to Hel308 Mbu(R681A/R687A/A700C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutations R681A/R687A/A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker). The band labelled A corresponds to the monomer and the band labelled B corresponds to the dimer.

Fig. 2 shows a gel of the Hel308 Mhu multimer. Lane 1 shows an appropriate protein ladder and Lane 2 corresponds to the Hel308 Mhu multimer (multiple units of SEQ ID NO: 19).

Fig. 3 shows a gel of the Hel308 Tga(R657A/N674C)-2kDa dimer and monomer at various stages during formation and purification (Lane 1 = protein ladder, Lane 2 = Hel308 Tga(R657A/N674C)-2kDa dimer after heating at 90 °C for 10 min, Lane 3 = Hel308 Tga(R657A/N674C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 16 with the mutations R657A/N674C, with one monomer unit being linked to the other via position 674 of each monomer unit using a 2 kDa PEG linker), Lane 4 = elution peak from a Strep-Tactin Sepharose™ purification, Lane 5 = initial reaction mixture and Lane 6 = Hel308 Tga(R657A/N674C) monomer (SEQ ID NO: 16 with the mutations R657A/N674C)). The band labelled A corresponds to the monomer and the band labelled B corresponds to the dimer.

Fig. 4 shows a fluorescence assay for testing helicase/DNA binding. A custom fluorescent substrate was used to assay the ability of various helicases to bind to single-stranded DNA. The 88 nt single-stranded DNA substrate (1 nM final, SEQ ID NO: 69, labelled A) has a carboxyfluorescein (FAM) base at its 5' end (circle labelled B). As the helicase (labelled C) binds to the oligonucleotide in buffered solution (400 mM NaCl, 10 mM Hepes, pH8.0, 1 mM $MgCl_2$), the fluorescence anisotropy (a property relating to the rate of free rotation of the oligonucleotide in solution) increases. The lower the amount of helicase needed to effect an increase in anisotropy, the tighter the binding affinity between the DNA and helicase. Situation 1 with no enzyme bound has a faster rotation and low anisotropy, whereas, situation 2 with

enzyme bound has slower rotation and high anisotropy. The black bar labelled X corresponds to increasing helicase concentration (the thicker the bar the higher the helicase concentration).

Fig. 5 shows the change in anisotropy of the DNA oligonucleotide (SEQ ID NO: 69, which has a carboxyfluorescein base at its 5' end) with increasing amounts of various Hel308 Mbu constructs (y-axis label = Anisotropy (blank subtracted), x-axis label = Protein Concentration (nM)). The data with black square points correspond to the Hel308 Mbu monomer (SEQ ID NO: 10). The data with the empty circles correspond to the Hel308 Mbu A700C 2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker). A lower concentration of the Hel308 Mbu A700C 2kDa dimer is required to effect an increase in anisotropy, therefore, the dimer has a higher binding affinity for the DNA than the monomer.

Fig. 6 shows the change in anisotropy of the DNA oligonucleotide (SEQ ID NO: 69, which has a carboxyfluorescein base at its 5' end) with increasing amounts of various Hel308 (Mbu) constructs (y-axis label = Anisotropy (blank subtracted), x-axis label = Protein Concentration (nM)). The data with black square points correspond to the Hel308 Mbu monomer (SEQ ID NO: 10). The data with the empty circles correspond to Hel308 Mbu-GTGSGA-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a HhH2 domain (SEQ ID NO: 75)) and the data with the empty triangles correspond to Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a (HhH)2-(HhH)2 domain (SEQ ID NO: 76)). The Hel308 Mbu helicases with additional helix-hairpin-helix binding domains attached show an increase in anisotropy at a lower concentration than the monomer. This indicates that the Hel308 Mbu constructs with additional binding domains have a stronger binding affinity for DNA than the monomer. The Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2, which has four HhH domains, was observed to bind DNA more tightly than Hel308 Mbu-GTGSGA-(HhH)2 which only has two HhH domains.

Fig. 7 shows the change in anisotropy of the DNA oligonucleotide (SEQ ID NO: 69, which has a carboxyfluorescein base at its 5' end) with increasing amounts of various Hel308 (Mbu) constructs (y-axis label = Anisotropy (blank subtracted), x-axis label = Protein Concentration (nM)). The data with black square points correspond to the Hel308 Mbu monomer (SEQ ID NO: 10). The data with the empty circles correspond to Hel308 Mbu-GTGSGA-UL42HV1-I320Del (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to UL42HV1-I320Del (SEQ ID NO: 63)), the data with the empty triangles

pointing up correspond to Hel308 Mbu-GTGSGA-gp32RB69CD (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to gp32RB69CD (SEQ ID NO: 64)) and the data with empty triangles pointing down correspond to Hel308 Mbu-GTGSGA-gp2.5T7-R211Del (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to gp2.5T7-R211Del (SEQ ID NO: 65)). All of the Hel308 Mbu helicases which have additional binding domains attached (Hel308 Mbu-GTGSGA-UL42HV1-I320Del, Hel308 Mbu-GTGSGA-gp32RB69CD and Hel308 Mbu-GTGSGA-gp2.5T7-R211Del) show an increase in anisotropy at a lower concentration than the monomer. This indicates that the Hel308 Mbu constructs with additional binding domains have a stronger binding affinity for DNA than the monomer.

Fig. 8 shows the change in anisotropy of the DNA oligonucleotide (SEQ ID NO: 69, which has a carboxyfluorescein base at its 5' end) with increasing amounts of various Hel308 (Mbu) constructs (y-axis label = Anisotropy (blank subtracted), x-axis label = Protein Concentration (nM)). The data with black square points correspond to the Hel308 Mbu monomer. The data with the empty circles correspond to (gp32RB69CD)-Hel308 Mbu (where the gp32RB69CD (SEQ ID NO: 64) is attached by the linker sequence GTGSGT to the helicase monomer unit (SEQ ID NO: 10)). The (gp32RB69CD)-Hel308 Mbu helicase construct shows an increase in anisotropy at a lower concentration than the monomer, indicating tighter binding to the DNA was observed in comparison to the monomer.

Fig. 9 shows relative equilibrium dissociation constants (K_d) (with respect to the Hel308 Mbu monomer) for various Hel308 (Mbu) constructs, obtained through fitting *two phase dissociation* binding curves through the data shown in Fig.'s 5-8 using Graphpad Prism software (y-axis label = Relative K_d, x-axis label = Ref. Number). The reference numbers correspond to the following Hel308 (Mbu) constructs – 3614 = Hel308 (Mbu), 3694 = (gp32-RB69CD)-Hel308 Mbu, 3733 = Hel308 (Mbu)-A700C 2kDa PEG dimer, 4401 = Hel308 (Mbu)-GTGSGA-(HhH)2, 4402 = Hel308 (Mbu)-GTGSGA-(HhH)2-(HhH)2, 4394 = Hel308 (Mbu)-GTGSGA-gp32RB69CD, 4395 = Hel308 (Mbu)-GTGSGA-gp2.5T7-R112Del and 4396 = Hel308 (Mbu)-GTGSGA-UL42HV1-I320Del. All of the helicase constructs with additional binding domains attached show a lower equilibrium dissociation constant than the Hel308 Mbu monomer alone.

Fig. 10 shows a schematic of helicase monomers controlling the movement of a polynucleotide through a nanopore in a membrane. A) Under an applied field (the direction of the applied field is indicated by the dashed black arrow), DNA in the *cis* compartment is captured by the nanopore, and translocates through the nanopore until the first helicase (grey semi-circle) contacts the top of the nanopore. After this point, the helicase will move along the

DNA (in the prescence of dNTP and suitable metal ion) controlling the movement of the DNA through the nanopore (the direction of the enzyme movement is indicated by the non-filled arrow). In the implementation shown, the DNA strand is captured by the 5' end by the pore, and the enzymes move 3' to 5' pulling the DNA against the field. As long as the enzyme does not dissociate, the strands will all end in the same way at the 5' end and will finally be ejected back to the *cis* side. Alternatively, strands captured by the 3' end by the nanopore will be fed into the pore by the enzyme moving 3' to 5' along the DNA, and will ultimately be ejected into the *trans* side. With enzymes that show 5' to 3' polarity, these modes are reversed. B) With enzyme monomers, if one of the enzymes dissociates (indicated by arrow 1) the DNA will start to translocate through the pore in the opposite direction due to the applied field pulling the DNA into the *cis* compartment. C) The DNA will continue to move with the applied field until a second helicase (black outlined semi-circle) contacts the top of the nanopore.

Fig. 11 shows a schematic of a helicase-helicase dimer controlling the movement of a polynucleotide through a nanopore in a membrane. A) Under an applied field (the direction of the applied field is indicated by the dashed black arrow), DNA in the *cis* compartment is captured by the nanopore, and translocates through the nanopore until the helicase contacts the top of the nanopore. After this point, the helicase will move along the DNA (in the prescence of dNTP and suitable metal ion) controlling the movement of the DNA through the nanopore. In the implementation shown, the DNA strand is captured at the 5' end by the pore, and the enzymes move 3' to 5' pulling the DNA against the field. As long as the enzyme does not dissociate, the strands will all end in the same way at the 5' end and will finally be ejected back to the *cis* side. Alternatively, strands captured at the 3' end by the nanopore will be fed into the pore by the enzyme moving 3' to 5' along the DNA, and will ultimately be ejected into the *trans* side. With 5' to 3' enzymes these modes are reversed. B) With enzyme dimers, if one of the enzymes dissociates (indicated by arrow 1) it will remain attached to the other enzyme and thus remain local to the DNA. C) This enhances rebinding of the dissociated enzyme to the DNA, where it can continue to move along the DNA. This enhanced rebinding improves the chances that the dimer construct will remain on the DNA and eventually move to the end of the DNA, thus improving overall processivity. The enzyme re-binding is indicated by arrow 2. It is possible to observe transitions from C) back to A) if the enzyme on top of the pore dissociates from the DNA. The DNA will be pulled back through the pore by the applied field until it reaches the attached trailing enzyme. The dissociated enzyme can then re-attach itself. This process is highlighted by the pale grey arrow going from C) back to A).

Fig. 12 shows the DNA substrate design used in Examples 5, 6 and 7. Strand A corresponds to SEQ ID NO: 70 (a 400mer) and Strand B corresponds to SEQ ID NO: 71 (primer which has a cholesterol tag at the 3' end (indicated by the two black circles)).

Fig. 13 shows that helicase monomers are able to move DNA through a nanopore in a controlled fashion, producing stepwise changes in current as the DNA moves through the nanopore. Example of current traces (y-axis label = Current (pA), x-axis label = Time (min) for traces A and B) observed when a monomer helicase controls the translocation of DNA (120 mV, 1 M KCl, 10 mM Hepes pH 8.0, 0.15 nM 400 mer DNA, 100 nM Hel308 Mbu monomer (SEQ ID NO: 10), 1 mM DTT, 1 mM ATP, 10 mM $MgCl_2$) through an MS(B1-L88N)8 MspA nanopore (8 monomer units as shown in SEQ ID NO: 2, with the mutation L88N). A) Section of current vs. time acquisition of a Hel308 Mbu monomer controlled 400 mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~260 pA) to a DNA level (~20-60 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore. B) Expanded view of the helicase-controlled DNA movements ending in the characteristic polyT level.

Fig. 14 shows that helicase-helicase dimers are able to move DNA through a nanopore in a controlled fashion, producing stepwise changes in current as the DNA moves through the nanopore. Example of current traces (y-axis label = Current (pA), x-axis label = Time (min) for traces A and B) observed when a dimer helicase controls the translocation of DNA (120 mV, 1 M KCl, 10 mM Hepes pH 8.0, 0.15 nM 400 mer DNA, 10 nM Hel308 Mbu A700C 2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker), 1 mM DTT, 1 mM ATP, 10 mM $MgCl_2$) through an MS(B1-L88N)8 MspA nanopore (SEQ ID NO: 2, with the mutation L88N). A) Section of current vs. time acquisition of a Hel308 Mbu A700C 2kDa dimer controlled 400 mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~260 pA) to a DNA level (~20-60 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore. B) Expanded view of the helicase-controlled DNA movements ending in the characteristic polyT level.

Fig. 15 shows the overall length of the strand movements (number of states, which corresponds to the number of bases moved) for an experiment using monomer Hel308 Mbu (SEQ ID NO: 10) to control DNA movement through a nanopore (+120 mV, 1 M KCl, 10 mM Hepes pH 8.0, 0.15 nM 400 mer DNA, 100 nM Hel308 Mbu monomer, 1 mM DTT, 1 mM ATP, 10 mM $MgCl_2$, MS(B1-L88N)8 MspA, y-axis label = Number of States, x-axis label = Strand). A dotted line highlights the number of states corresponding to 500. For the monomer run 37% of the helicase-controlled DNA movements measured reached the polyT at the end of the DNA strand.

Fig. 16 shows the overall length of the strand movements (number of states, which corresponds to the number of bases moved) for an experiment using Hel308 Mbu A700C 2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit via a 2 kDa PEG linker) to control DNA movement through a nanopore (+120 mV, 1 M KCl, 10 mM Hepes pH 8.0, 0.15 nM 400 mer DNA, 10 nM Hel308 Mbu A700C 2kDa dimer, 1 mM DTT, 1 mM ATP, 10 mM $MgCl_2$, MS(B1-L88N)8 MspA, y-axis label = Number of States, x-axis label = Strand). DNA movements controlled by a dimer are typically much longer than those controlled by a monomer helicase (a dotted line highlights the number of states corresponding to 500). This indicates enzyme rebinding and therefore reduced enzyme dissociation. For the dimer run 47% of the helicase-controlled DNA movements measured reached the polyT, showing the reduced dissociation and improved processivity of the dimer.

Fig. 17 shows six examples of the position in the known DNA sequence (y-axis label = Position in Sequence) of the state-fitted data for the Hel308 Mbu monomer controlled strand movements as a function of the state index x-axis label = State Index). The Hel308 Mbu monomer (SEQ ID NO: 10) data show processive linear movement through the sequence, with periodic dislocations back to previous parts of the sequence, which are the result of enzyme dissociation and the DNA slipping back under the applied field until encountering a trailing enzyme. Many of the helicase-controlled DNA movements do not make it to the end of the sequence due to enzyme dissociation.

Fig. 18 shows six examples of the position in the known DNA sequence (y-axis label = Position in Sequence) of the state-fitted data for the Hel308 Mbu A700C 2kDa homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker) controlled strand movements as a function of the state index x-axis label = State Index). The dimer data show processive linear movement through the sequence, with periodic

dislocations back to previous parts of the sequence, which are the result of enzyme dissociation. However, unlike the monomer data the enzymes proceed to control the movement of the DNA for much longer, and after dissociation the enzyme re-binds to the DNA.

Fig. 19 shows helicase-helicase dimers are able to move DNA through a nanopore in a controlled fashion, producing stepwise changes in current as the DNA moves through the nanopore. Example current traces (y-axis label = Current (pA), x-axis label = Time (min) for traces A and B) observed when a dimer helicase controls the translocation of DNA (180 mV, 400 mM KCl, 10 mM Hepes pH 8.0, 0.15 nM 400 mer DNA, approximately 1 nM Hel308 Mbu Q442C 2 kDa linker homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation Q442C, with one monomer unit being linked to the other via position 442 of each monomer unit using a 2 kDa PEG linker) or 1 nM Hel308 Mbu Q442C 3.4 kDa linker homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation Q442C, with one monomer unit being linked to the other via position 442 of each monomer unit using a 3.4 kDa PEG linker), 1 mM DTT, 1 mM ATP, 1 mM $MgCl_2$) through an MS(B1-L88N)8 MspA nanopore. A) Section of current vs. time acquisition of a Hel308 Mbu Q442C 2 kDa linker homodimer controlled 400mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~170 pA) to a DNA level (~40-80 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore. B) Section of current vs. time acquisition of a Hel308 Mbu Q442C 3.4 kDa linker homodimer controlled 400mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~170 pA) to a DNA level (~40-80 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore.

Fig. 20 shows helicases attached to an additional binding domain are able to move DNA through a nanopore in a controlled fashion, producing stepwise changes in current as the DNA moves through the nanopore. Example current traces (y-axis label = Current (pA), x-axis label = Time (min) for traces A and B) observed when a helicase-controls the translocation of DNA (140 mV, 400 mM NaCl, 10 mM Hepes pH 8.0, 0.6 nM 400 mer DNA, 100 nM Hel308 Mbu + 5^{th} domain Hel308 Hla (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5^{th} domain of Hel308 Hla (SEQ ID NO: 66)) or 100 nM Hel308 Mbu + 5^{th} domain Hel308 Hvo (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5^{th} domain of Hel308 Hvo

(SEQ ID NO: 67))., 1 mM DTT, 1 mM ATP, 1 mM $MgCl_2$) through an MS(B1-L88N)8 MspA nanopore. A) Section of current vs. time acquisition of a Hel308 Mbu + 5th domain Hel308 Hla controlled 400 mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~100 pA) to a DNA level (~10-40 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore. B) Section of current vs. time acquisition of a Hel308 Mbu + 5th domain Hel308 Hvo controlled 400 mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~100 pA) to a DNA level (~10-40 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore.

Fig. 21 shows the DNA substrate design used in Example 8. Strand A corresponds to SEQ ID NO: 72 (a 900mer) and Strand B corresponds to SEQ ID NO: 73 (anti-sense sequence minus a 4 base-pair leader). Strand C corresponds to SEQ ID NO: 74 (primer which has a cholesterol tag at the 3' end (indicated by the two black circles)).

Fig. 22 shows helicases attached to additional helix-harpin-helix binding domains are able to move DNA through a nanopore in a controlled fashion, producing stepwise changes in current as the DNA moves through the nanopore. Example current traces (y-axis label = Current (pA), x-axis label = Time (min) for traces A and B) observed when a helicase controls the translocation of DNA(140 mV, 400 mM NaCl, 100 mM Hepes pH 8.0, 0.1 nM 900 mer DNA, 100 nM Hel308 Mbu-GTGSGA-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a HhH2 domain (SEQ ID NO: 75)) or 100 nM Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a (HhH)2-(HhH)2 domain (SEQ ID NO: 76)), 10 mM potassium ferrocyanide, 10 mM potassium ferricyanide, 1 mM ATP, 1 mM $MgCl_2$) through an MS(B1-G75S-G77S-L88N-Q126R)8 MspA nanopore (8 monomer units as shown in SEQ ID NO: 2 with the mutations G75S/G77S/L88N/Q126R)). A) Section of current vs. time acquisition of a Hel308 Mbu-GTGSGA-(HhH)2 controlled 900 mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in the current from the open-pore level (~110 pA) to a DNA level (~10-40 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level

before exiting the nanopore. B) Section of current vs. time acquisition of a Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 controlled 900 mer DNA movement. Under an applied potential DNA with helicase bound is captured by the nanopore. This produces blocks in current from the open-pore level (~110 pA) to a DNA level (~10-40 pA). The DNA level then shows stepwise changes in current as the enzyme moves the DNA through the pore. The example helicase-controlled DNA movement shown ends in a characteristic long polyT level before exiting the nanopore.

Fig. 23 Fluorescence assay for testing enzyme activity. A custom fluorescent substrate was used to assay the ability of the helicase/helicase dimer (a) to displace hybridised dsDNA. 1) The fluorescent substrate strand (50 nM final, SEQ ID NO: 91 and 92) has both a 3' and 5' ssDNA overhang, and a 44 base section of hybridised dsDNA. The upper strand (b) has a carboxyfluorescein base (c) at the 5' end (labeled 5 in SEQ ID NO: 91), and the hybridised complement (d) has a black-hole quencher (BHQ-1) base (e) at the 3' end (labeled 6 in SEQ ID NO: 92). When hybridised, the fluorescence from the fluorescein is quenched by the local BHQ-1, and the substrate is essentially non-fluorescent. 1 μM of a capture strand (f, SEQ ID NO: 93) that is part-complementary to the lower strand of the fluorescent substrate is included in the assay. 2) In the presence of ATP (1 mM) and $MgCl_2$ (1 mM), helicase (100 nM) added to the substrate binds to the 3' tail 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. 4) Displaced lower strand (d) preferentially anneals to an excess ofcapture strand (f) to prevent re-annealing of initial substrate and loss of fluorescence.

Fig. 24 Graph (y-axis = 1000x dsDNA turnover (molecules/min/enz), x-axis = enzyme) of the initial rate of activity in buffer solutions (10 mM Hepes pH 8.0, 1 mM ATP, 1 mM $MgCl_2$, 50 nM fluorescent substrate DNA (SEQ ID NOs: 91 and 92), 1 μM capture DNA (SEQ ID NO: 93)) for the TrwC Cba monomer (labeled A, SEQ ID NO: 87) and the TrwC Cba-TopoV Mka (labeled B, where TrwC Cba is attached by the linker AYDVGA to domains H-L of Topoisomerase V Mka full sequence shown in SEQ ID NO: 90) at 400 mM of NaCl.

Fig. 25 shows a gel of the TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) and TrwC Cba Q276C monomer (where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) at various stages during formation and purification. Lane M = protein ladder, Lane 1 = E3-Q276C monomer starting material, Lanes 2 and 3 = reaction mix. The band which corresponds to TrwC

Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) is indicated by a grey arrow.

Fig. 26 shows a gel of the TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) and TrwC Cba Q276C monomer (where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) at various stages during formation and purification. Lane M = protein ladder, Lane X = reference lane for TrwC Cba Q276C-3.4kDa dimer, lanes 4-14 contain purified fractions from the elution of TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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 band which corresponds to TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) is indicated by a grey arrow.

Fig. 27 shows a cartoon representation of the chemical reaction steps which are necessary to form the Hel308Mbu-A577Faz-PEG4 linker-TrwC Cba Q276C dimer (where the Hel308Mbu monomer unit (labeled 2) comprises SEQ ID NO: 10 with the amino acid at position 577 mutated to a 4-azido-L-phenylalanine (Faz), which is attached by PEG4 linker (black dotted line) to TrwC Cba monomer unit (labeled 1) SEQ ID NO: 87 with the mutation Q276C, where the linker is attached to each monomer at position 577 on Hel 308 Mbu monomer and position 276 on TrwC Cba). Step one reacts the cysteine at position 276 on the surface of TrwC Cba with the maleimide functional group (labelled X) at one end of the PEG4 linker. Step two reacts the 4-azido-L-phenylalanine (Faz) amino acid at position 577 on the surface of Hel308Mbu with the DBCO functional group (labelled Y) at the other end of the PEG4 linker using click chemistry.

Fig. 28 shows a 4-12% gel of samples from Example 12. The sample in each lane is as follows - lane a) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C), lane b) Hel308 Mbu-A577Faz (where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz), lane c) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Hel308 Mbu-A577Faz (where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz), lane d) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + 5kDa PEG, lane e) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + 5kDa PEG with an azide attached, lane f) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Azide Alexa Fluor® 555(Life Technologies, used to check for

non-specific interactions between the fluorophore and TrwC Cba-Q276C monomer), lane g) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO, lane h) TrwC Cba-Q276C-PEG4-DBCO (SEQ ID NO: 87 with the mutation Q276C which is attached to the PEG4-DBCO linker) + Hel308 Mbu (SEQ ID NO: 10), lane i) Hel308Mbu-A577Faz-PEG4 linker-TrwC Cba Q276C dimer (where the Hel308Mbu monomer unit comprises SEQ ID NO: 10 with the amino acid at position 577 mutated to a 4-azido-L-phenylalanine (Faz), which is attached by PEG4 linker to TrwC Cba monomer unit SEQ ID NO: 87 with the mutation Q276C, where the linker is attached to each monomer at position 577 on Hel 308 Mbu monomer and position 276 on TrwC Cba) plus unreacted TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO + Hel308 Mbu-A577Faz monomer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz), lane j) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO + 5kDa PEG with an azide attached, lane k) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO + Azide Alexa Fluor® 555(Life Technologies, used to check for non-specific interactions between the fluorophore and TrwC Cba-Q276C monomer). The band corresponding to the desired dimer product (Hel308Mbu-A577Faz-PEG4 linker-TrwC Cba Q276C dimer (where the Hel308Mbu monomer unit comprises SEQ ID NO: 10 with the amino acid at position 577 mutated to a 4-azido-L-phenylalanine (Faz), which is attached by PEG4 linker to TrwC Cba monomer unit SEQ ID NO: 87 with the mutation Q276C, where the linker is attached to each monomer at position 577 on Hel 308 Mbu monomer and position 276 on TrwC Cba)) is indicated by a grey arrow.

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 Hel308 motif.

SEQ ID NO: 9 shows the amino acid sequence of the extended Hel308 motif.

SEQ ID NO: 10 shows the amino acid sequence of Hel308 Mbu.

SEQ ID NO: 11 shows the Hel308 motif of Hel308 Mbu and Hel308 Mhu.

SEQ ID NO: 12 shows the extended Hel308 motif of Hel308 Mbu and Hel308 Mhu.

SEQ ID NO: 13 shows the amino acid sequence of Hel308 Csy.

SEQ ID NO: 14 shows the Hel308 motif of Hel308 Csy.

SEQ ID NO: 15 shows the extended Hel308 motif of Hel308 Csy.

SEQ ID NO: 16 shows the amino acid sequence of Hel308 Tga.

SEQ ID NO: 17 shows the Hel308 motif of Hel308 Tga.

SEQ ID NO: 18 shows the extended Hel308 motif of Hel308 Tga.

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

SEQ ID NO: 20 shows the RecD-like motif I.

SEQ ID NOs: 21, 22 and 23 show the extended RecD-like motif I.

SEQ ID NO: 24 shows the RecD motif I.

SEQ ID NO: 25 shows a preferred RecD motif I, namely G-G-P-G-T-G-K-T.

SEQ ID NOs: 26, 27 and 28 show the extended RecD motif I.

SEQ ID NO: 29 shows the RecD-like motif V.

SEQ ID NO: 30 shows the RecD motif V.

SEQ ID NOs: 31-38 show the MobF motif III.

SEQ ID NOs: 39-45 shows the MobQ motif III.

SEQ ID NO: 46 shows the amino acid sequence of TraI Eco.

SEQ ID NO: 47 shows the RecD-like motif I of TraI Eco.

SEQ ID NO: 48 shows the RecD-like motif V of TraI Eco.

SEQ ID NO: 49 shows the MobF motif III of TraI Eco.

SEQ ID NO: 50 shows the XPD motif V.

SEQ ID NO: 51 shows XPD motif VI.

SEQ ID NO: 52 shows the amino acid sequence of XPD Mbu.

SEQ ID NO: 53 shows the XPD motif V of XPD Mbu.

SEQ ID NO: 54 shows XPD motif VI of XPD Mbu.

SEQ ID NO: 55 shows the amino acid sequence of a preferred HhH domain.

SEQ ID NO: 56 shows the amino acid sequence of the SSB from the bacteriophage RB69, which is encoded by the gp32 gene.

SEQ ID NO: 57 shows the amino acid sequence of the SSB from the bacteriophage T7, which is encoded by the gp2.5 gene.

SEQ ID NO: 58 shows the amino acid sequence of the UL42 processivity factor from Herpes virus 1.

SEQ ID NO: 59 shows the amino acid sequence of subunit 1 of PCNA.

SEQ ID NO: 60 shows the amino acid sequence of subunit 2 of PCNA.

SEQ ID NO: 61 shows the amino acid sequence of subunit 3 of PCNA.

SEQ ID NO: 62 shows the amino acid sequence of Phi29 DNA polymerase.

SEQ ID NO: 63 shows the amino acid sequence (from 1 to 319) of the UL42 processivity factor from the Herpes virus 1.

SEQ ID NO: 64 shows the amino acid sequence of the SSB from the bacteriophage RB69, i.e. SEQ ID NO: 56, with its C terminus deleted (gp32RB69CD).

SEQ ID NO: 65 shows the amino acid sequence (from 1 to 210) of the SSB from the bacteriophage T7 (gp2.5T7-R211Del). The full length protein is shown in SEQ ID NO: 57.

SEQ ID NO: 66 shows the amino acid sequence of the 5th domain of Hel308 Hla.

SEQ ID NO: 67 shows the amino acid sequence of the 5th domain of Hel308 Hvo.

SEQ ID NO: 68 shows the polynucleotide sequence of a DNA strand used in helicase dimer production.

SEQ ID NO: 69 shows the polynucleotide sequence of a DNA strand used in a helicase fluorescent assay.

SEQ ID NO: 70 shows the polynucleotide sequence of a ssDNA strand used in Examples 5, 6 and 7. At the 5' end of SEQ ID NO: 70 there are four 2'-O-methyl uracil bases attached to a 50T leader sequence to aid capture by the nanopore.

SEQ ID NO: 71 shows the polynucleotide sequence of a ssDNA strand used in Examples 5, 6 and 7.

SEQ ID NOs: 72 and 73 show polynucleotide sequences of ssDNA strands used in Example 8.

SEQ ID NO: 74 shows the polynucleotide sequence of a ssDNA strand used in Example 8.

SEQ ID NO: 75 shows the amino acid sequence of the (HhH)2 domain.

SEQ ID NO: 76 shows the amino acid sequence of the (HhH)2-(HhH)2 domain.

SEQ ID NO: 77 shows the amino acid sequence of the human mitochondrial SSB (*H*smtSSB).

SEQ ID NO: 78 shows the amino acid sequence of the p5 protein from Phi29 DNA polymerase.

SEQ ID NO: 79 shows the amino acid sequence of the wild-type SSB from *E. coli*.

SEQ ID NO: 80 shows the amino acid sequence of the ssb from the bacteriophage T4, which is encoded by the gp32 gene.

SEQ ID NO: 81 shows the amino acid sequence of EcoSSB-CterAla.

SEQ ID NO: 82 shows the amino acid sequence of EcoSSB-CterNGGN.

SEQ ID NO: 83 shows the amino acid sequence of EcoSSB-Q152del.

SEQ ID NO: 84 shows the amino acid sequence of EcoSSB-G117del.

SEQ ID NO: 85 shows the GTGSGA linker.

SEQ ID NO: 86 shows the GTGSGT linker.

SEQ ID NO: 87 shows the amino acid sequence TrwC Cba.

SEQ ID NO: 88 shows part of the polynucleotide sequence used in Example 9. Attached to the 5' end of this sequence 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 this sequence is four iSpC3 spacer units which are attached to the 5' end of SEQ ID NO: 104.

SEQ ID NO: 89 shows the amino acid sequence of Topoisomerase V Mka (Methanopyrus Kandleri).

SEQ ID NO: 90 shows the amino acid sequence of TrwC Cba-TopoV Mka, where TrwC Cba is attached by the linker AYDVGA to domains H-L of Topoisomerase V Mka.

SEQ ID NOs: 91-93 show the polynucleotide sequences used in Example 10.

SEQ ID NO: 94 shows the amino acid sequence of domains H-L of Topoisomerase V Mka (Methanopyrus Kandleri).

SEQ ID NOs: 95 to 103 show some of the TraI sequences shown in Table 2.

SEQ ID NO: 104 shows part of 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: 88. Attached to the 5' end of SEQ ID NO: 88 is 28 iSpC3 spacer units the last of which has an additional two T's attached to the 5' end of the spacer group.

SEQ ID NO: 105 shows the amino acid sequence of Mutant S (*Escherichia coli*).

SEQ ID NO: 106 shows the amino acid sequence of Sso7d (*Sufolobus solfataricus*).

SEQ ID NO: 107 shows the amino acid sequence of Sso10b1 (*Sulfolobus solfataricus P2*).

SEQ ID NO: 108 shows the amino acid sequence of Sso10b2 (*Sulfolobus solfataricus P2*).

SEQ ID NO: 109 shows the amino acid sequence of Tryptophan repressor (*Escherichia coli*).

SEQ ID NO: 110 shows the amino acid sequence of Lambda repressor (*Enterobacteria phage lambda*).

SEQ ID NO: 111 shows the amino acid sequence of Cren7 (*Histone crenarchaea Cren7 Sso*).

SEQ ID NO: 112 shows the amino acid sequence of human histone (*Homo sapiens*).

SEQ ID NO: 113 shows the amino acid sequence of dsbA (*Enterobacteria phage T4*).

SEQ ID NO: 114 shows the amino acid sequence of Rad51 (*Homo sapiens*).

SEQ ID NO: 115 shows the amino acid sequence of PCNA sliding clamp (*Citromicrobium bathyomarinum JL354*).

Detailed description of the invention

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

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

Constructs for use in the invention

The present invention provides methods using a construct that is helpful for controlling the movement of a polynucleotide. The construct comprises a helicase and an additional polynucleotide binding moiety. The helicase is attached to the polynucleotide binding moiety. The construct has the ability to control the movement of a polynucleotide. The construct is artificial or non-natural.

As discussed in more detail below, the construct may comprise two or more helicases (i.e. the additional polynucleotide binding moiety is one or more additional helicases). In such embodiments, each helicase in the construct is capable of functioning on its own as a helicase. The construct itself is not a multimeric or oligomeric helicase, such as dimeric helicase. In other words, the construct itself is not a helicase that naturally exists as a multimer or an oligomer,

such as a dimer. The construct may comprise a multimeric, such as dimeric, helicase, but it must be attached to an additional polynucleotide binding moiety, such as another helicase. The helicase is preferably monomeric. The helicase is preferably not a helicase domain from a helicase enzyme. This is discussed in more detail below.

The constructs described herein are useful tools for controlling the movement of a polynucleotide during Strand Sequencing. A problem which occurs in sequencing polynucleotides, particularly those of 500 nucleotides or more, is that the molecular motor which is controlling translocation of the polynucleotide may disengage from the polynucleotide. This allows the polynucleotide to be pulled through the pore rapidly and in an uncontrolled manner in the direction of the applied field. The constructs described herein are less likely to disengage from the polynucleotide being sequenced. The construct can provide increased read lengths of the polynucleotide as it controls the translocation of the polynucleotide through a nanopore. The ability to translocate an entire polynucleotide through a nanopore under the control of a construct described herein allows characteristics of the polynucleotide, such as its sequence, to be estimated with improved accuracy and speed over known methods. This becomes more important as strand lengths increase and molecular motors are required with improved processivity. The constructs described herein are particularly effective in controlling the translocation of target polynucleotides of 500 nucleotides or more, for example 1000 nucleotides, 5000, 10000, 20000, 50000, 100000 or more.

A targeted construct that binds to a specific polynucleotide sequence can also be designed. As discussed in more detail below, the polynucleotide binding moiety may bind to a specific polynucleotide sequence and thereby target the helicase portion of the construct to the specific sequence.

The constructs described herein are also useful tools for isothermal polymerase chain reaction (PCR). In such methods, the strands of double stranded DNA are typically first separated by a construct described herein and coated by single stranded DNA (ssDNA)-binding proteins. In the second step, two sequence specific primers typically hybridise to each border of the DNA template. DNA polymerases may then be used to extend the primers annealed to the templates to produce a double stranded DNA and the two newly synthesized DNA products may then be used as substrates by the constructs described herein entering the next round of the reaction. Thus, a simultaneous chain reaction develops, resulting in exponential amplification of the selected target sequence.

The construct has the ability to control the movement of a polynucleotide. The ability of a construct to control the movement of a polynucleotide can be assayed using any method known

in the art. For instance, the construct may be contacted with a polynucleotide and the position of the polynucleotide may be determined using standard methods. The ability of a construct to control the movement of a polynucleotide is typically assayed as described in the Examples.

The construct may be isolated, substantially isolated, purified or substantially purified. A construct is isolated or purified if it is completely free of any other components, such as lipids, polynucleotides or pore monomers. A construct is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a construct is substantially isolated or substantially purified if it is present in a form that comprises less than 10%, less than 5%, less than 2% or less than 1% of other components, such as lipids, polynucleotides or pore monomers.

Attachment

The helicase is attached to the additional polynucleotide binding moiety. The helicase is preferably covalently attached to the additional polynucleotide binding moiety. The helicase may be attached to the moiety at more than one, such as two or three, points.

The helicase can be covalently attached to the moiety using any method known in the art. The helicase and moiety may be produced separately and then attached together. The two components may be attached in any configuration. For instance, they may be attached via their terminal (i.e. amino or carboxy terminal) amino acids. Suitable configurations include, but are not limited to, the amino terminus of the moiety being attached to the carboxy terminus of the helicase and *vice versa*. Alternatively, the two components may be attached via amino acids within their sequences. For instance, the moiety may be attached to one or more amino acids in a loop region of the helicase. In a preferred embodiment, terminal amino acids of the moiety are attached to one or more amino acids in the loop region of a helicase. Terminal amino acids and loop regions can be identified using methods known in the art (Edman P., Acta Chemica Scandinavia, (1950), 283-293). For instance, loop regions can be identified using protein modeling. This exploits the fact that protein structures are more conserved than protein sequences amongst homologues. Hence, producing atomic resolution models of proteins is dependent upon the identification of one or more protein structures that are likely to resemble the structure of the query sequence. In order to assess whether a suitable protein structure exists to use as a "template" to build a protein model, a search is performed on the protein data bank (PDB) database. A protein structure is considered a suitable template if it shares a reasonable level of sequence identity with the query sequence. If such a template exists, then the template sequence is "aligned" with the query sequence, i.e. residues in the query sequence are mapped

onto the template residues. The sequence alignment and template structure are then used to produce a structural model of the query sequence. Hence, the quality of a protein model is dependent upon the quality of the sequence alignment and the template structure.

The two components may be attached via their naturally occurring amino acids, such as cysteines, threonines, serines, aspartates, asparagines, glutamates and glutamines. Naturally occurring amino acids may be modified to facilitate attachment. For instance, the naturally occurring amino acids may be modified by acylation, phosphorylation, glycosylation or farnesylation. Other suitable modifications are known in the art. Modifications to naturally occurring amino acids may be post-translation modifications. The two components may be attached via amino acids that have been introduced into their sequences. Such amino acids are preferably introduced by substitution. The introduced amino acid may be cysteine or a non-natural amino acid that facilitates attachment. Suitable non-natural amino acids include, but are not limited to, 4-azido-L-phenylalanine (Faz), and any one of the amino acids numbered 1-71 included in figure 1 of Liu C. C. and Schultz P. G., Annu. Rev. Biochem., 2010, 79, 413-444. The introduced amino acids may be modified as discussed above.

In a preferred embodiment, the helicase is chemically attached to the moiety, for instance via a linker molecule. Linker molecules are discussed in more detail below. One suitable method of chemical attachment is cysteine linkage. This is discussed in more detail below.

The helicase may be transiently attached to the moiety by a hexa-his tag or Ni-NTA. The helicase and moiety may also be modified such that they transiently attach to each other.

In another preferred embodiment, the helicase is genetically fused to the moiety. A helicase is genetically fused to a moiety if the whole construct is expressed from a single polynucleotide sequence. The coding sequences of the helicase and moiety may be combined in any way to form a single polynucleotide sequence encoding the construct. Genetic fusion of a pore to a nucleic acid binding protein is discussed in International Application No. PCT/GB09/001679 (published as WO 2010/004265).

The helicase and moiety may be genetically fused in any configuration. The helicase and moiety may be fused via their terminal amino acids. For instance, the amino terminus of the moiety may be fused to the carboxy terminus of the helicase and *vice versa*. The amino acid sequence of the moiety is preferbly added in frame into the amino acid sequence of the helicase. In other words, the moiety is preferably inserted within the sequence of the helicase. In such embodiments, the helicase and moiety are typically attached at two points, i.e. via the amino and carboxy terminal amino acids of the moiety. If the moiety is inserted within the sequence of the helicase, it is preferred that the amino and carboxy terminal amino acids of the moiety are in

close proximity and are each attached to adjacent amino acids in the sequence of the helicase or variant thereof. In a preferred embodiment, the moiety is inserted into a loop region of the helicase.

The construct retains the ability of the helicase to control the movement of a polynucleotide. This ability of the helicase is typically provided by its three dimensional structure that is typically provided by its β-strands and α-helices. The α-helices and β-strands are typically connected by loop regions. In order to avoid affecting the ability of the helicase to control the movement of a polynucleotide, the moiety is preferably genetically fused to either end of the helicase or inserted into a surface-exposed loop region of the helicase. The loop regions of specific helicases can be identified using methods known in the art. For instance, the loop regions can be identified using protein modelling, x-ray diffraction measurement of the protein in a crystalline state (Rupp B (2009). Biomolecular Crystallography: Principles, Practice and Application to Structural Biology. New York: Garland Science.), nuclear magnetic resonance (NMR) spectroscopy of the protein in solution (Mark Rance; Cavanagh, John; Wayne J. Fairbrother; Arthur W. Hunt III; Skelton, Nicholas J. (2007). Protein NMR spectroscopy: principles and practice (2nd ed.). Boston: Academic Press.) or cryo-electron microscopy of the protein in a frozen-hydrated state (van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A (2000). "Single-particle electron cryo-microscopy: towards atomic resolution.". Q Rev Biophys. 33: 307–69. Structural information of proteins determined by above mentioned methods are publicly available from the protein bank (PDB) database.

For Hel308 helicases (SEQ ID NOs: 10, 13, 16 and 19), β-strands can only be found in the two RecA-like engine domains (domains 1 and 2). These domains are responsible for coupling the hydrolysis of the fuel nucleotide (normally ATP) with movement. The important domains for ratcheting along a polynucleotide are domains 3 and 4, but above all domain 4. Interestingly, both of domains 3 and 4 comprise only α-helices. There is an important α-helix in domain 4 called the ratchet helix. As a result, in the Hel308 embodiments of the invention, the moiety is preferably not genetically fused to any of the α-helixes.

In another embodiment, the helicase is attached to the moiety using intein-Tag sequences. Two proteins can be joined by genetically encoding compatible split intein-Tag sequences at the end of each protein. The inteins do not require catalysts or enzymes but self release while joining the two proteins. The join is traceless, leaving a single peptide chain. This method is generally for joining termini of proteins.

The helicase may be attached directly to the moiety. The helicase is preferably attached to the moiety using one or more, such as two or three, linkers. The one or more linkers may be designed to constrain the mobility of the moiety. The linkers may be attached to one or more reactive cysteine residues, reactive lysine residues or non-natural amino acids in the helicase and/or moiety. The non-natural amino acid may be any of those discussed above. The non-natural amino acid is preferably 4-azido-L-phenylalanine (Faz). Suitable linkers are well-known in the art.

The helicase is preferably attached to the moiety using one or more chemical crosslinkers or one or more peptide linkers. Suitable chemical crosslinkers are well-known in the art. Suitable chemical crosslinkers include, but are not limited to, those including the following functional groups: maleimide, active esters, succinimide, azide, alkyne (such as dibenzocyclooctynol (DIBO or DBCO), difluoro cycloalkynes and linear alkynes), phosphine (such as those used in traceless and non-traceless Staudinger ligations), haloacetyl (such as iodoacetamide), phosgene type reagents, sulphonyl chloride reagents, isothiocyanates, acyl halides, hydrazines, disulphides, vinyl sulfones, aziridines and photoreactive reagents (such as aryl azides, diaziridines). The crosslinker is preferably not bis(sulfosuccinimidyl) suberate (BS^3). The helicase and the moiety are preferably not crosslinked using formaldehyde.

Reactions between amino acids and functional groups may be spontaneous, such as cysteine/maleimide, or may require external reagents, such as Cu(I) for linking azide and linear alkynes.

Linkers can comprise any molecule that stretches across the distance required. Linkers can vary in length from one carbon (phosgene-type linkers) to many Angstroms. Examples of linear molecules, include but are not limited to, are polyethyleneglycols (PEGs), polypeptides, polysaccharides, deoxyribonucleic acid (DNA), peptide nucleic acid (PNA), threose nucleic acid (TNA), glycerol nucleic acid (GNA), saturated and unsaturated hydrocarbons, polyamides. These linkers may be inert or reactive, in particular they may be chemically cleavable at a defined position, or may be themselves modified with a fluorophore or ligand. The linker is preferably resistant to dithiothreitol (DTT).

Cleavable linkers can be used as an aid to separation of constructs from non-attached components and can be used to further control the synthesis reaction. For example, a hetero-bifunctional linker may react with the helicase, but not the moiety. If the free end of the linker can be used to bind the helicase protein to a surface, the unreacted helicases from the first reaction can be removed from the mixture. Subsequently, the linker can be cleaved to expose a group that reacts with the moiety. In addition, by following this sequence of linkage reactions,

conditions may be optimised first for the reaction to the helicase, then for the reaction to the moiety after cleavage of the linker. The second reaction would also be much more directed towards the correct site of reaction with the moiety because the linker would be confined to the region to which it is already attached.

Preferred crosslinkers include 2,5-dioxopyrrolidin-1-yl 3-(pyridin-2-yldisulfanyl)propanoate, 2,5-dioxopyrrolidin-1-yl 4-(pyridin-2-yldisulfanyl)butanoate, 2,5-dioxopyrrolidin-1-yl 8-(pyridin-2-yldisulfanyl)octananoate, di-maleimide PEG 1k, di-maleimide PEG 3.4k, di-maleimide PEG 5k, di-maleimide PEG 10k, bis(maleimido)ethane (BMOE), bis-maleimidohexane (BMH), 1,4-bis-maleimidobutane (BMB), 1,4 bis-maleimidyl-2,3-dihydroxybutane (BMDB), BM[PEO]2 (1,8-bis-maleimidodiethyleneglycol), BM[PEO]3 (1,11-bis-maleimidotriethylene glycol), tris[2-maleimidoethyl]amine (TMEA), DTME dithiobismaleimidoethane, bis-maleimide PEG3, bis-maleimide PEG11, DBCO-maleimide, DBCO-PEG4-maleimide, DBCO-PEG4-NH2, DBCO-PEG4-NHS, DBCO-NHS, DBCO-PEG-DBCO 2.8kDa, DBCO-PEG-DBCO 4.0kDa, DBCO-15 atoms-DBCO, DBCO-26 atoms-DBCO, DBCO-35 atoms-DBCO, DBCO-PEG4-S-S-PEG3-biotin, DBCO-S-S-PEG3-biotin and DBCO-S-S-PEG11-biotin. The most preferred crosslinkers are succinimidyl 3-(2-pyridyldithio)propionate (SPDP) and maleimide-PEG(2kDa)-maleimide (alpha,omega-bis-maleimido poly(ethylene glycol)).

The helicase may be covalently attached to the bifunctional crosslinker before the helicase/crosslinker complex is covalently attached to the moiety. Alternatively, the moiety may be covalently attached to the bifunctional crosslinker before the bifunctional crosslinker/moiety complex is attached to the helicase. The helicase and moiety may be covalently attached to the chemical crosslinker at the same time.

The helicase may be attached to the moiety using two different linkers that are specific for each other. One of the linkers is attached to the helicase and the other is attached to the moiety. Once mixed together, the linkers should react to form a construct described herein. The helicase may be attached to the moiety using the hybridization linkers described in International Application No. PCT/GB10/000132 (published as WO 2010/086602). In particular, the helicase may be attached to the moiety using two or more linkers each comprising a hybridizable region and a group capable of forming a covalent bond. The hybridizable regions in the linkers hybridize and link the moieties. The linked moieties are then coupled via the formation of covalent bonds between the groups. Any of the specific linkers disclosed in International Application No. PCT/GB10/000132 (published as WO 2010/086602) may be used in accordance with the invention.

The helicase and the moiety may be modified and then attached using a chemical crosslinker that is specific for the two modifications. Any of the crosslinkers discussed above may be used.

Alternatively, the linkers preferably comprise amino acid sequences. Such linkers are peptide linkers. The length, flexibility and hydrophilicity of the peptide linker are typically designed such that it does not to disturb the functions of the helicase and moiety. Preferred flexible peptide linkers are stretches of 2 to 20, such as 4, 6, 8, 10 or 16, serine and/or glycine amino acids. More preferred flexible linkers include $(SG)_1$, $(SG)_2$, $(SG)_3$, $(SG)_4$, $(SG)_5$, $(SG)_8$, $(SG)_{10}$, $(SG)_{15}$ or $(SG)_{20}$ wherein S is serine and G is glycine. Preferred rigid linkers are stretches of 2 to 30, such as 4, 6, 8, 16 or 24, proline amino acids. More preferred rigid linkers include $(P)_{12}$ wherein P is proline.

The linkers may be labeled. Suitable labels include, but are not limited to, fluorescent molecules (such as Cy3 or AlexaFluor®555), radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, antigens, polynucleotides and ligands such as biotin. Such labels allow the amount of linker to be quantified. The label could also be a cleavable purification tag, such as biotin, or a specific sequence to show up in an identification method, such as a peptide that is not present in the protein itself, but that is released by trypsin digestion.

A preferred method of attaching the helicase to the moiety is via cysteine linkage. This can be mediated by a bi-functional chemical linker or by a polypeptide linker with a terminal presented cysteine residue. Linkage can occur via natural cysteines in the helicase and/or moiety. Alternatively, cysteines can be introduced into the helicase and/or moiety. If the helicase is attached to the moiety via cysteine linkage, the one or more cysteines have preferably been introduced to the helicase and/or moiety by substitution.

The length, reactivity, specificity, rigidity and solubility of any bi-functional linker may be designed to ensure that the moiety is positioned correctly in relation to the helicase and the function of both the helicase and moiety is retained. Suitable linkers include bismaleimide crosslinkers, such as 1,4-bis(maleimido)butane (BMB) or bis(maleimido)hexane. One drawback of bi-functional linkers is the requirement of the helicase and moiety to contain no further surface accessible cysteine residues if attachment at specific sites is preferred, as binding of the bi-functional linker to surface accessible cysteine residues may be difficult to control and may affect substrate binding or activity. If the helicase and/or moiety does contain several accessible cysteine residues, modification of the helicase and/or moiety may be required to remove them while ensuring the modifications do not affect the folding or activity of the helicase and moiety. This is discussed in International Application No. PCT/GB10/000133 (published as WO

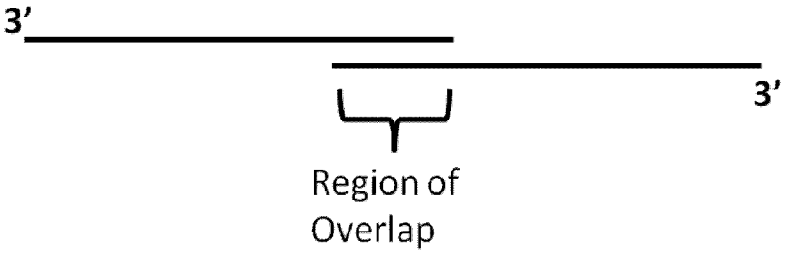
2010/086603). In a preferred embodiment, a reactive cysteine is presented on a peptide linker that is genetically attached to the moiety. This means that additional modifications will not necessarily be needed to remove other accessible cysteine residues from the moiety. The reactivity of cysteine residues may be enhanced by modification of the adjacent residues, for example on a peptide linker. 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 5,5'-dithiobis-(2-nitrobenzoic acid) (dTNB). These may be reacted with one or more cysteine residues of the moiety or helicase, either as a monomer or part of an oligomer, before a linker is attached. Selective deprotection of surface accessible cysteines may be possible using reducing reagents immobilized on beads (for example immobilized tris(2-carboxyethyl)phosphine, TCEP). Cysteine linkage of two or more helicases is discussed in more detail below.

Another preferred method of attaching the helicase to the moiety is via 4-azido-L-phenylalanine (Faz) linkage. This can be mediated by a bi-functional chemical linker or by a polypeptide linker with a terminal presented Faz residue. The one or more Faz residues have preferably been introduced to the helicase and/or moiety by substitution. Faz linkage of two or more helicases is discussed in more detail below.

Cross-linkage of helicases or moieties to themselves may be prevented by keeping the concentration of linker in a vast excess of the helicase and/or moiety. Alternatively, a "lock and key" arrangement may be used in which two linkers are used. Only one end of each linker may react together to form a longer linker and the other ends of the linker each react with a different part of the construct (i.e. helicase or moiety). This is discussed in more detail below.

The site of attachment is selected such that, when the construct is contacted with a polynucleotide, both the helicase and the moiety can bind to the polynucleotide and control its movement.

Attachment can be facilitated using the polynucleotide binding activities of the helicase and the moiety. For instance, complementary polynucleotides can be used to bring the helicase and moiety together as they hybridize. The helicase can be bound to one polynucleotide and the moiety can be bound to the complementary polynucleotide. The two polynucleotides can then be allowed to hybridise to each other. This will bring the helicase into close contact with the moiety, making the linking reaction more efficient. This is especially helpful for attaching two or more helicases in the correct orientation for controlling movement of a target polynucleotide. An example of complementary polynucleotides that may be used are shown below.



For helicase-Phi29 constructs the DNA below could be used.

Tags can be added to the construct to make purification of the construct easier. These tags can then be chemically or enzymatically cleaved off, if their removal is necessary. Fluorophores or chromophores can also be included, and these could also be cleavable.

A simple way to purify the construct is to include a different purification tag on each protein (i.e. the helicase and the moiety), such as a hexa-His-tag and a Strep-tag®. If the two proteins are different from one another, this method is particularly useful. The use of two tags enables only the species with both tags to be purified easily.

If the two proteins do not have two different tags, other methods may be used. For instance, proteins with free surface cysteines or proteins with linkers attached that have not reacted to form a construct could be removed, for instance using an iodoacetamide resin for maleimide linkers.

Constructs can also be purified from unreacted proteins on the basis of a different DNA processivity property. In particular, a construct can be purified from unreacted proteins on the basis of an increased affinity for a polynucleotide, a reduced likelihood of disengaging from a polynucleotide once bound and/or an increased read length of a polynucleotide as it controls the translocation of the polynucleotide through a nanopore

Helicase

Any helicase may be used in the contructs described herein. Helicases are often known as translocases and the two terms may be used interchangeably. Suitable helicases are well-known in the art (M. E. Fairman-Williams *et al.*, Curr. Opin. Struct Biol., 2010, 20 (3), 313-324, T. M. Lohman et al., Nature Reviews Molecular Cell Biology, 2008, 9, 391-401). The helicase is typically a member of one of superfamilies 1 to 6. The helicase is preferably a member of any of

the Moiety Classification (EC) groups 3.6.1.- and 2.7.7.-. The helicase is preferably an ATP-dependent DNA helicase (EC group 3.6.4.12), an ATP-dependent RNA helicase (EC group 3.6.4.13) or an ATP-independent RNA helicase.

The helicase may be a multimeric or oligomeric helicase. In other words, the helicase may need to form a multimer or an oligomer, such as a dimer, to function. However, as discussed above, the construct itself cannot be a multimeric or oligomeric helicase. The multimeric or oligomeric helicase must be attached to an additional polynucleotide binding moiety. The helicase is preferably monomeric. In other words, the helicase preferably does not need to form a multimer or an oligomer, such as a dimer, to function. Hel308, RecD, TraI and XPD helicases are all monomeric helicases. These are discussed in more detail below. The helicase is preferably not the hepatitis C virus NS3 helicase (also known as NS3h). NS3 helicase acts as an oligomer and so is not monomeric.

Monomeric helicases may comprise several domains attached together. For instance, TraI helicases and TraI subgroup helicases may contain two RecD helicase domains, a relaxase domain and a C-terminal domain. The domains typically form a monomeric helicase that is capable of functioning without forming oligomers.

The helicase is typically an entire helicase, such as an entire Hel308, RecD, TraI or XPD helicase. The helicase is preferably not a helicase domain from a helicase enzyme. For instance, the helicase is preferably not a RecD domain or the helicase domain from the Brome mosaic virus (BMV) viral replication protein 1a. The construct cannot itself be a helicase comprising two or more helicase domains attached together, such as a TraI helicase comprising two RecD domains attached together. The construct may comprise a helicase comprising two or more helicase domains, such as TraI helicase, but it must be attached to an additional polynucleotide binding moiety. The helicase is preferably capable of binding to the target polynucleotide at an internal nucleotide. An internal nucleotide is a nucleotide which is not a terminal nucleotide in the target polynucleotide. For example, it is not a 3' terminal nucleotide or a 5' terminal nucleotide. All nucleotides in a circular polynucleotide are internal nucleotides.

Generally, a helicase which is capable of binding at an internal nucleotide is also capable of binding at a terminal nucleotide, but the tendency for some helicases to bind at an internal nucleotide will be greater than others. For a helicase suitable for use in the invention, typically at least 10% of its binding to a polynucleotide will be at an internal nucleotide. Typically, at least 20%, at least 30%, at least 40% or at least 50% of its binding will be at an internal nucleotide. Binding at a terminal nucleotide may involve binding to both a terminal nucleotide and adjacent internal nucleotides at the same time. For the purposes of the invention, this is not binding to the

target polynucleotide at an internal nucleotide. In other words, the helicase used in the invention is not only capable of binding to a terminal nucleotide in combination with one or more adjacent internal nucleotides. The helicase must be capable of binding to an internal nucleotide without concurrent binding to a terminal nucleotide.

A helicase which is capable of binding at an internal nucleotide may bind to more than one internal nucleotide. Typically, the helicase binds to at least 2 internal nucleotides, for example at least 3, at least 4, at least 5, at least 10 or at least 15 internal nucleotides. Typically the helicase binds to at least 2 adjacent internal nucleotides, for example at least 3, at least 4, at least 5, at least 10 or at least 15 adjacent internal nucleotides. The at least 2 internal nucleotides may be adjacent or non-adjacent.

The ability of a helicase to bind to a polynucleotide at an internal nucleotide may be determined by carrying out a comparative assay. The ability of a motor to bind to a control polynucleotide A is compared to the ability to bind to the same polynucleotide but with a blocking group attached at the terminal nucleotide (polynucleotide B). The blocking group prevents any binding at the terminal nucleotide of strand B, and thus allows only internal binding of a helicase.

Examples of helicases which are capable of binding at an internal nucleotide include, but are not limited to, Hel308 Tga, Hel308 Mhu and Hel308 Csy. Hence, the molecular motor preferably comprises (a) the sequence of Hel308 Tga (i.e. SEQ ID NO: 16) or a variant thereof or (b) the sequence of Hel308 Csy (i.e. SEQ ID NO: 13) or a variant thereof or (c) the sequence of Hel308 Mhu (i.e. SEQ ID NO: 19) or a variant thereof. Variants of these sequences are discussed in more detail below. Variants preferably comprise one or more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment as discussed above.

The helicase is preferably a Hel308 helicase. Hel308 helicases are monomeric because they function without forming oligomers. Any Hel308 helicase may be used in accordance with the invention. Hel308 helicases are also known as ski2-like helicases and the two terms can be used interchangeably. Suitable Hel308 helicases are disclosed in Table 4 of US Patent Application Nos. 61,549,998 and 61/599,244 and International Application No. PCT/GB2012/052579 (published as WO 2013/057495).

The Hel308 helicase typically comprises the amino acid motif Q-X1-X2-G-R-A-G-R (hereinafter called the Hel308 motif; SEQ ID NO: 8). The Hel308 motif is typically part of the helicase motif VI (Tuteja and Tuteja, Eur. J. Biochem. 271, 1849–1863 (2004)). X1 may be C, M or L. X1 is preferably C. X2 may be any amino acid residue. X2 is typically a hydrophobic or

neutral residue. X2 may be A, F, M, C, V, L, I, S, T, P or R. X2 is preferably A, F, M, C, V, L, I, S, T or P. X2 is more preferably A, M or L. X2 is most preferably A or M.

The Hel308 helicase preferably comprises the motif Q-X1-X2-G-R-A-G-R-P (hereinafter called the extended Hel308 motif; SEQ ID NO: 9) wherein X1 and X2 are as described above.

The most preferred Hel308 motifs and extended Hel308 motifs are shown in the Table 1 below.

Table 1 – Preferred Hel308 helicases and their motifs

SEQ ID NO:	Helicase	Names	% Identity to Hel308 Mbu	Hel308 motif	Extended Hel308 motif
10	Hel308 Mbu	Methanococcoides burtonii	-	QMAGRAGR (SEQ ID NO: 11)	QMAGRAGRP (SEQ ID NO: 12)
13	Hel308 Csy	Cenarchaeum symbiosum	34%	QLCGRAGR (SEQ ID NO: 14)	QLCGRAGRP (SEQ ID NO: 15)
16	Hel308 Tga	Thermococcus gammatolerans EJ3	38%	QMMGRAGR (SEQ ID NO: 17)	QMMGRAGRP (SEQ ID NO: 18)
19	Hel308 Mhu	Methanospirillum hungatei JF-1	40%	QMAGRAGR (SEQ ID NO: 11)	QMAGRAGRP (SEQ ID NO: 12)

The most preferred Hel308 motif is shown in SEQ ID NO: 17. The most preferred extended Hel308 motif is shown in SEQ ID NO: 18. Other preferred Hel308 motifs and extended Hel308 motifs are found in Table 5 of US Patent Application Nos. 61,549,998 and 61/599,244 and International Application No. PCT/GB2012/052579 (published as WO 2013/057495).

The Hel308 helicase preferably comprises the sequence of Hel308 Mbu (i.e. SEQ ID NO: 10) or a variant thereof. The Hel308 helicase more preferably comprises (a) the sequence of Hel308 Tga (i.e. SEQ ID NO: 16) or a variant thereof, (b) the sequence of Hel308 Csy (i.e. SEQ ID NO: 13) or a variant thereof or (c) the sequence of Hel308 Mhu (i.e. SEQ ID NO: 19) or a variant thereof. The Hel308 helicase most preferably comprises the sequence shown in SEQ ID NO: 16 or a variant thereof.

A variant of a Hel308 helicase is an enzyme that has an amino acid sequence which varies from that of the wild-type helicase and which retains polynucleotide binding activity. In

particular, a variant of SEQ ID NO: 10, 13, 16 or 19 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 10, 13, 16 or 19 and which retains polynucleotide binding activity. Polynucleotide binding activity can be determined using methods known in the art. Suitable methods include, but are not limited to, fluorescence anisotropy, tryptophan fluorescence and electrophoretic mobility shift assay (EMSA). For instance, the ability of a variant to bind a single stranded polynucleotide can be determined as described in the Examples.

The variant retains helicase activity. This can be measured in various ways. For instance, the ability of the variant to translocate along a polynucleotide can be measured using electrophysiology, a fluorescence assay or ATP hydrolysis.

The variant may include modifications that facilitate handling of the polynucleotide encoding the helicase and/or facilitate its activity at high salt concentrations and/or room temperature. Variants typically differ from the wild-type helicase in regions outside of the Hel308 motif or extended Hel308 motif discussed above. However, variants may include modifications within these motif(s).

Over the entire length of the amino acid sequence of SEQ ID NO: 10, 13, 16 or 19, a variant will preferably be at least 30% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 40%, at least 45%, at least 50%, 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: 10, 13, 16 or 19 over the entire sequence. There may be at least 70%, for example at least 80%, at least 85%, at least 90% or at least 95%, amino acid identity over a stretch of 150 or more, for example 200, 300, 400, 500, 600, 700, 800, 900 or 1000 or more, contiguous amino acids ("hard homology"). Homology is determined as described below. The variant may differ from the wild-type sequence in any of the ways discussed below with reference to SEQ ID NOs: 2 and 4.

A variant of SEQ ID NO: 10, 13, 16 or 19 preferably comprises the Hel308 motif or extended Hel308 motif of the wild-type sequence as shown in Table 1 above. However, a variant may comprise the Hel308 motif or extended Hel308 motif from a different wild-type sequence. For instance, a variant of SEQ ID NO: 12 may comprise the Hel308 motif or extended Hel308 motif from SEQ ID NO: 13 (i.e. SEQ ID NO: 14 or 15). Variants of SEQ ID NO: 10, 13, 16 or 19 may also include modifications within the Hel308 motif or extended Hel308 motif of the relevant wild-type sequence. Suitable modifications at X1 and X2 are discussed above when defining the two motifs. A variant of SEQ ID NO: 10, 13, 16 or 19 preferably comprises one or

more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment as discussed above.

A variant of SEQ ID NO: 10 may lack the first 19 amino acids of SEQ ID NO: 10 and/or lack the last 33 amino acids of SEQ ID NO: 10. A variant of SEQ ID NO: 10 preferably comprises a sequence which is at least 70%, at least 75%, at least 80%, at least 85%, at least 90% or more preferably at least 95%, at least 97% or at least 99% homologous based on amino acid identity with amino acids 20 to 211 or 20 to 727 of SEQ ID NO: 10.

SEQ ID NO: 10 (Hel308 Mbu) contains five natural cysteine residues. However, all of these residues are located within or around the DNA binding grove of the enzyme. Once a DNA strand is bound within the enzyme, these natural cysteine residues become less accessible for external modifications. This allows specific cysteine mutants of SEQ ID NO: 10 to be designed and attached to the moiety using cysteine linkage as discussed above. Preferred variants of SEQ ID NO: 10 have one or more of the following substitutions: A29C, Q221C, Q442C, T569C, A577C, A700C and S708C. The introduction of a cysteine residue at one or more of these positions facilitates cysteine linkage as discussed above. Other preferred variants of SEQ ID NO: 10 have one or more of the following substitutions: M2Faz, R10Faz, F15Faz, A29Faz, R185Faz, A268Faz, E284Faz, Y387Faz, F400Faz, Y455Faz, E464Faz, E573Faz, A577Faz, E649Faz, A700Faz, Y720Faz, Q442Faz and S708Faz. The introduction of a Faz residue at one or more of these positions facilitates Faz linkage as discussed above.

The helicase is preferably a RecD helicase. RecD helicases are monomeric because they function without forming oligomers. Any RecD helicase may be used in accordance with the invention. The structures of RecD helicases are known in the art (FEBS J. 2008 Apr;275(8):1835-51. Epub 2008 Mar 9. ATPase activity of RecD is essential for growth of the Antarctic Pseudomonas syringae Lz4W at low temperature. Satapathy AK, Pavankumar TL, Bhattacharjya S, Sankaranarayanan R, Ray MK; EMS Microbiol Rev. 2009 May;33(3):657-87. The diversity of conjugative relaxases and its application in plasmid classification. Garcillán-Barcia MP, Francia MV, de la Cruz F; J Biol Chem. 2011 Apr 8;286(14):12670-82. Epub 2011 Feb 2. Functional characterization of the multidomain F plasmid TraI relaxase-helicase. Cheng Y, McNamara DE, Miley MJ, Nash RP, Redinbo MR).

The RecD helicase typically comprises the amino acid motif X1-X2-X3-G-X4-X5-X6-X7 (hereinafter called the RecD-like motif I; SEQ ID NO: 20), wherein X1 is G, S or A, X2 is any amino acid, X3 is P, A, S or G, X4 is T, A, V, S or C, X5 is G or A, X6 is K or R and X7 is T or S. X1 is preferably G. X2 is preferably G, I, Y or A. X2 is more preferably G. X3 is preferably P or A. X4 is preferably T, A, V or C. X4 is preferably T, V or C. X5 is preferably G.

X6 is preferably K. X7 is preferably T or S. The RecD helicase preferably comprises Q-$(X8)_{16-18}$-X1-X2-X3-G-X4-X5-X6-X7 (hereinafter called the extended RecD-like motif I; SEQ ID NOs: 21, 22 and 23), wherein X1 to X7 are as defined above and X8 is any amino acid. There are preferably 16 X8 residues (i.e. $(X8)_{16}$) in the extended RecD-like motif I. Suitable sequences for $(X8)_{16}$ can be identified in SEQ ID NOs: 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47 and 50 of US Patent Application No. 61/581,332 and SEQ ID NOs: 18, 21, 24, 25, 28, 30, 32, 35, 37, 39, 41, 42 and 44 of International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

The RecD helicase preferably comprises the amino acid motif G-G-P-G-Xa-G-K-Xb (hereinafter called the RecD motif I; SEQ ID NO: 24) wherein Xa is T, V or C and Xb is T or S. Xa is preferably T. Xb is preferably T. The Rec-D helicase preferably comprises the sequence G-G-P-G-T-G-K-T (SEQ ID NO: 25). The RecD helicase more preferably comprises the amino acid motif Q-$(X8)_{16-18}$-G-G-P-G-Xa-G-K-Xb (hereinafter called the extended RecD motif I; SEQ ID NOs: 26, 27 and 28), wherein Xa and Xb are as defined above and X8 is any amino acid. There are preferably 16 X8 residues (i.e. $(X8)_{16}$) in the extended RecD motif I. Suitable sequences for $(X8)_{16}$ can be identified in SEQ ID NOs: 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47 and 50 of US Patent Application No. 61/581,332 and SEQ ID NOs: 18, 21, 24, 25, 28, 30, 32, 35, 37, 39, 41, 42 and 44 of International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

The RecD helicase typically comprises the amino acid motif X1-X2-X3-X4-X5-$(X6)_3$-Q-X7 (hereinafter called the RecD-like motif V; SEQ ID NO: 29), wherein X1 is Y, W or F, X2 is A, T, S, M, C or V, X3 is any amino acid, X4 is T, N or S, X5 is A, T, G, S, V or I, X6 is any amino acid and X7 is G or S. X1 is preferably Y. X2 is preferably A, M, C or V. X2 is more preferably A. X3 is preferably I, M or L. X3 is more preferably I or L. X4 is preferably T or S. X4 is more preferably T. X5 is preferably A, V or I. X5 is more preferably V or I. X5 is most preferably V. $(X6)_3$ is preferably H-K-S, H-M-A, H-G-A or H-R-S. $(X6)_3$ is more preferably H-K-S. X7 is preferably G. The RecD helicase preferably comprises the amino acid motif Xa-Xb-Xc-Xd-Xe-H-K-S-Q-G (hereinafter called the RecD motif V; SEQ ID NO: 30), wherein Xa is Y, W or F, Xb is A, M, C or V, Xc is I, M or L, Xd is T or S and Xe is V or I. Xa is preferably Y. Xb is preferably A. Xd is preferably T. Xd is preferably V. Preferred RecD motifs I are shown in Table 5 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562). Preferred RecD-like motifs I are shown in Table 7 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562). Preferred RecD-like motifs V are

shown in Tables 5 and 7 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

The RecD helicase is preferably one of the helicases shown in Table 4 or 5 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562) or a variant thereof. Variants are described in US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

The RecD helicase is preferably a TraI helicase or a TraI subgroup helicase. TraI helicases and TraI subgroup helicases are monomeric because they function without forming oligomers. TraI helicases and TraI subgroup helicases may contain two RecD helicase domains, a relaxase domain and a C-terminal domain. The TraI subgroup helicase is preferably a TrwC helicase. The TraI helicase or TraI subgroup helicase is preferably one of the helicases shown in Table 6 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562) or a variant thereof. Variants are described in US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

The TraI helicase or a TraI subgroup helicase typically comprises a RecD-like motif I as defined above (SEQ ID NO: 20) and/or a RecD-like motif V as defined above (SEQ ID NO: 29). The TraI helicase or a TraI subgroup helicase preferably comprises both a RecD-like motif I (SEQ ID NO: 20) and a RecD-like motif V (SEQ ID NO: 29). The TraI helicase or a TraI subgroup helicase typically further comprises one of the following two motifs:

- The amino acid motif H-$(X1)_2$-X2-R-$(X3)_{5-12}$-H-X4-H (hereinafter called the MobF motif III; SEQ ID NOs: 31-38), wherein X1 and X2 are any amino acid and X2 and X4 are independently selected from any amino acid except D, E, K and R. $(X1)_2$ is of course X1a-X1b. X1a and X1b can be the same of different amino acid. X1a is preferably D or E. X1b is preferably T or D. $(X1)_2$ is preferably DT or ED. $(X1)_2$ is most preferably DT. The 5 to 12 amino acids in $(X3)_{5-12}$ can be the same or different. X2 and X4 are independently selected from G, P, A, V, L, I, M, C, F, Y, W, H, Q, N, S and T. X2 and X4 are preferably not charged. X2 and X4 are preferably not H. X2 is more preferably N, S or A. X2 is most preferably N. X4 is most preferably F or T. $(X3)_{5-12}$ is preferably 6 or 10 residues in length. Suitable embodiments of $(X3)_{5-12}$ can be derived from SEQ ID NOs: 58, 62, 66 and 70 shown in Table 7 of US Patent Application No. 61/581,332 and SEQ ID NOs: 61, 65, 69, 73, 74, 82, 86, 90, 94, 98, 102, 110, 112, 113, 114, 117, 121, 124, 125, 129, 133, 136, 140, 144, 147, 151, 152, 156, 160, 164 and 168 of International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

- The amino acid motif G-X1-X2-X3-X4-X5-X6-X7-H-$(X8)_{6-12}$-H-X9 (hereinafter called the MobQ motif III; SEQ ID NOs: 39-45), wherein X1, X2, X3, X5, X6, X7 and X9 are independently selected from any amino acid except D, E, K and R, X4 is D or E and X8 is any amino acid. X1, X2, X3, X5, X6, X7 and X9 are independently selected from G, P, A, V, L, I, M, C, F, Y, W, H, Q, N, S and T. X1, X2, X3, X5, X6, X7 and X9 are preferably not charged. X1, X2, X3, X5, X6, X7 and X9 are preferably not H. The 6 to 12 amino acids in $(X8)_{6-12}$ can be the same or different. Preferred MobF motifs III are shown in Table 7 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562).

The TraI helicase or TraI subgroup helicase is more preferably one of the helicases shown in Table 6 or 7 of US Patent Application No. 61/581,332 and International Application No. PCT/GB2012/053274 (published as WO 2012/098562) or a variant thereof. The TraI helicase most preferably comprises the sequence shown in SEQ ID NO: 46 or a variant thereof. SEQ ID NO: 46 is TraI Eco (NCBI Reference Sequence: NP_061483.1; Genbank AAQ98619.1; SEQ ID NO: 46). TraI Eco comprises the following motifs: RecD-like motif I (GYAGVGKT; SEQ ID NO: 47), RecD-like motif V (YAITAHGAQG; SEQ ID NO: 48) and Mob F motif III (HDTSRDQEPQLHTH; SEQ ID NO: 49).

The TraI helicase or TraI subgroup helicase more preferably comprises the sequence of one of the helicases shown in Table 2 below, i.e. one of SEQ ID NOs: 46, 87, 98 and 102, or a variant thereof.

Table 2 – More preferred TraI helicase and TraI subgroup helicases

SEQ ID NO	**Name**	**Strain**	**NCBI ref**	**% Identity to TraI Eco**	**RecD-like motif I (SEQ ID NO:)**	**RecD-like motif V (SEQ ID NO:)**	**Mob F motif III (SEQ ID NO:)**
46	TraI Eco	Escherichia coli	NCBI Reference Sequence: NP_061483.1 Genbank AAQ98619.1	-	GYAGVGKT (47)	YAITAHGAQG (48)	HDTSRDQEPQLHTH 49)
87	TrwC Cba	Citromicrobium bathyomarinum JL354	NCBI Reference Sequence: ZP_06861556.1	15%	GIAGAGKS (95)	YALNVHMAQG (96)	HDTNRNQEPNLHFH (97)
98	TrwC	Halothiob	NCBI	11.5%	GAAGA	YCITIH	HEDAR

	Hne	acillus neapolitanus c2	Reference Sequence: YP_003262832.1		GKT (99)	RSQG (100)	TVDDIADPQLHTH (101)
102	TrwC Eli	Erythrobacter litoralis HTCC2594	NCBI Reference Sequence: YP_457045.1	16%	GIAGAGKS (95)	YALNAHMAQG (103)	HDTNRNQEPNLHFH (97)

A variant of a RecD helicase, TraI helicase or TraI subgroup helicase is an enzyme that has an amino acid sequence which varies from that of the wild-type helicase and which retains polynucleotide binding activity. This can be measured as described above. In particular, a variant of SEQ ID NO: 46, 87, 98 or 102 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 46, 87, 98 or 102 and which retains polynucleotide binding activity. The variant retains helicase activity. The variant must work in at least one of the two modes discussed below. Preferably, the variant works in both modes. The variant may include modifications that facilitate handling of the polynucleotide encoding the helicase and/or facilitate its activity at high salt concentrations and/or room temperature. Variants typically differ from the wild-type helicase in regions outside of the motifs discussed above. However, variants may include modifications within these motif(s).

Over the entire length of the amino acid sequence of any one of SEQ ID NOs: 46, 87, 98 and 102, a variant will preferably be at least 10% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 20%, at least 25%, at least 30%, at least 40%, at least 45%, at least 50%, 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 any one of SEQ ID NOs: 46, 87, 98 or 102 over the entire sequence. There may be at least 70%, for example at least 80%, at least 85%, at least 90% or at least 95%, amino acid identity over a stretch of 150 or more, for example 200, 300, 400, 500, 600, 700, 800, 900 or 1000 or more, contiguous amino acids ("hard homology"). Homology is determined as described below. The variant may differ from the wild-type sequence in any of the ways discussed above with reference to SEQ ID NOs: 2 and 4.

A variant of any one of SEQ ID NOs: 46, 87, 98 and 102 preferably comprises the RecD-like motif I and/or RecD-like motif V of the wild-type sequence. However, a variant of SEQ ID NO: 46, 87, 98 or 102 may comprise the RecD-like motif I and/or extended RecD-like motif V from a different wild-type sequence. For instance, a variant may comprise any one of the preferred motifs shown in Tables 5 and 7 of US Patent Application No. 61/581,332 and

International Application No. PCT/GB2012/053274 (published as WO 2012/098562). Variants of SEQ ID NO: 46, 87, 98 or 102 may also include modifications within the RecD-like motifs I and V of the wild-type sequence. A variant of SEQ ID NO: 46, 87, 98 or 102 preferably comprises one or more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment as discussed above.

The helicase is preferably an XPD helicase. XPD helicases are monomeric because they function without forming oligomers. Any XPD helicase may be used in accordance with the invention. XPD helicases are also known as Rad3 helicases and the two terms can be used interchangeably.

The structures of XPD helicases are known in the art (Cell. 2008 May 30;133(5):801-12. Structure of the DNA repair helicase XPD. Liu H, Rudolf J, Johnson KA, McMahon SA, Oke M, Carter L, McRobbie AM, Brown SE, Naismith JH, White MF). The XPD helicase typically comprises the amino acid motif X1-X2-X3-G-X4-X5-X6-E-G (hereinafter called XPD motif V; SEQ ID NO: 50). X1, X2, X5 and X6 are independently selected from any amino acid except D, E, K and R. X1, X2, X5 and X6 are independently selected from G, P, A, V, L, I, M, C, F, Y, W, H, Q, N, S and T. X1, X2, X5 and X6 are preferably not charged. X1, X2, X5 and X6 are preferably not H. X1 is more preferably V, L, I, S or Y. X5 is more preferably V, L, I, N or F. X6 is more preferably S or A. X3 and X4 may be any amino acid residue. X4 is preferably K, R or T.

The XPD helicase typically comprises the amino acid motif Q-Xa-Xb-G-R-Xc-Xd-R-$(Xe)_3$-Xf-$(Xg)_7$-D-Xh-R (hereinafter called XPD motif VI; SEQ ID NO: 51). Xa, Xe and Xg may be any amino acid residue. Xb, Xc and Xd are independently selected from any amino acid except D, E, K and R. Xb, Xc and Xd are typically independently selected from G, P, A, V, L, I, M, C, F, Y, W, H, Q, N, S and T. Xb, Xc and Xd are preferably not charged. Xb, Xc and Xd are preferably not H. Xb is more preferably V, A, L, I or M. Xc is more preferably V, A, L, I, M or C. Xd is more preferably I, H, L, F, M or V. Xf may be D or E. $(Xg)_7$ is X_{g1}, X_{g2}, X_{g3}, X_{g4}, X_{g5}, X_{g6} and X_{g7}. X_{g2} is preferably G, A, S or C. X_{g5} is preferably F, V, L, I, M, A, W or Y. X_{g6} is preferably L, F, Y, M, I or V. X_{g7} is preferably A, C, V, L, I, M or S.

The XPD helicase preferably comprises XPD motifs V and VI. The most preferred XPD motifs V and VI are shown in Table 5 of US Patent Application No. 61/581,340 and International Application No. PCT/GB2012/053273 (published as WO 2012/098561).

The XPD helicase preferably further comprises an iron sulphide (FeS) core between two Walker A and B motifs (motifs I and II). An FeS core typically comprises an iron atom

coordinated between the sulphide groups of cysteine residues. The FeS core is typically tetrahedral.

The XPD helicase is preferably one of the helicases shown in Table 4 or 5 of US Patent Application No. 61/581,340 and International Application No. PCT/GB2012/053273 (published as WO 2012/098561) or a variant thereof. The XPD helicase most preferably comprises the sequence shown in SEQ ID NO: 52 or a variant thereof. SEQ ID NO: 52 is XPD Mbu (Methanococcoides burtonii; YP_566221.1; GI:91773529). XPD Mbu comprises YLWGTLSEG (Motif V; SEQ ID NO: 53) and QAMGRVVRSPTDYGARILLDGR (Motif VI; SEQ ID NO: 54).

A variant of a XPD helicase is an enzyme that has an amino acid sequence which varies from that of the wild-type helicase and which retains polynucleotide binding activity. This can be measured as described above. In particular, a variant of SEQ ID NO: 52 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 52 and which retains polynucleotide binding activity. The variant retains helicase activity. The variant must work in at least one of the two modes discussed below. Preferably, the variant works in both modes. The variant may include modifications that facilitate handling of the polynucleotide encoding the helicase and/or facilitate its activity at high salt concentrations and/or room temperature. Variants typically differ from the wild-type helicase in regions outside of XPD motifs V and VI discussed above. However, variants may include modifications within one or both of these motifs.

Over the entire length of the amino acid sequence of SEQ ID NO: 52, such as SEQ ID NO: 10, a variant will preferably be at least 10%, preferably 30% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 40%, at least 45%, at least 50%, 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: 52 over the entire sequence. There may be at least 70%, for example at least 80%, at least 85%, at least 90% or at least 95%, amino acid identity over a stretch of 150 or more, for example 200, 300, 400, 500, 600, 700, 800, 900 or 1000 or more, contiguous amino acids ("hard homology"). Homology is determined as described below. The variant may differ from the wild-type sequence in any of the ways discussed above with reference to SEQ ID NOs: 2 and 4.

A variant of SEQ ID NO: 52 preferably comprises the XPD motif V and/or the XPD motif VI of the wild-type sequence. A variant of SEQ ID NO: 52 more preferably comprises both XPD motifs V and VI of SEQ ID NO: 52. However, a variant of SEQ ID NO: 52 may

comprise XPD motifs V and/or VI from a different wild-type sequence. For instance, a variant of SEQ ID NO: 52 may comprise any one of the preferred motifs shown in Table 5 of US Patent Application No. 61/581,340 and International Application No. PCT/GB2012/053273 (published as WO 2012/098561). Variants of SEQ ID NO: 52 may also include modifications within XPD motif V and/or XPD motif VI of the wild-type sequence. Suitable modifications to these motifs are discussed above when defining the two motifs. A variant of SEQ ID NO: 52 preferably comprises one or more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment as discussed above.

The helicase may be any of the modified helicases described and claimed in US Provisional Application No. 61/673,452 (filed 19 July 2012), US Provisional Application No. 61/774,862 (filed 8 March 2013) and the International Application being filed concurrently with this application (Oxford Nanopore Ref: ONT IP 033), all of which correspond to WO 2014/013260.

The helicase is more preferably a Hel308 helicase in which one or more cysteine residues and/or one or more non-natural amino acids have been introduced at one or more of the positions which correspond to D272, N273, D274, G281, E284, E285, E287, S288, T289, G290, E291, D293, T294, N300, R303, K304, N314, S315, N316, H317, R318, K319, L320, E322, R326, N328, S615, K717, Y720, N721 and S724 in Hel308 Mbu (SEQ ID NO: 10), wherein the helicase retains its ability to control the movement of a polynucleotide.

The Hel308 helicase preferably comprises a variant of one of SEQ ID NOs: 10, 13, 16 or 19 which comprises one or more cysteine residues and/or one or more non-natural amino acids at one or more of the positions which correspond to D272, N273, D274, G281, E284, E285, E287, S288, T289, G290, E291, D293, T294, N300, R303, K304, N314, S315, N316, H317, R318, K319, L320, E322, R326, N328, S615, K717, Y720, N721 and S724 in Hel308 Mbu (SEQ ID NO: 10).

The Hel308 helicase preferably comprises a variant of one of SEQ ID NOs: 10, 13, 16 or 19 which comprises one or more cysteine residues and/or one or more non-natural amino acids at one or more of the positions which correspond to D274, E284, E285, S288, S615, K717, Y720, E287, T289, G290, E291, N316 and K319 in Hel308 Mbu (SEQ ID NO: 10).

Tables 3a and 3b below show the positions in other Hel308 helicases which correspond to D274, E284, E285, S288, S615, K717, Y720, E287, T289, G290, E291, N316 and K319 in Hel308 Mbu (SEQ ID NO: 10). The lack of a corresponding position in another Hel308 helicase is marked as a "-".

Table 3a – Positions which correspond to D274, E284, E285, S288, S615, K717 and Y720 in Hel308 Mbu (SEQ ID NO: 10).

SEQ ID NO:	Hel308 homologue	A	B	C	D	E	F	G
10	Mbu	D274	E284	E285	S288	S615	K717	Y720
13	Csy	D280	K290	I291	S294	P589	T694	N697
16	Tga	L266	S276	L277	Q280	P583	K689	D692
19	Mhu	S269	Q277	E278	R281	S583	G685	R688

Table 3b – Positions which correspond to E287, T289, G290, E291, N316 and K319 in Hel308 Mbu (SEQ ID NO: 10).

SEQ ID NO:	Hel308 homologue	H	I	J	K	L	M
10	Mbu	E287	T289	G290	E291	N316	K319
13	Csy	S293	G295	G296	E297	D322	S325
16	Tga	S279	L281	E282	D283	V308	T311
19	Mhu	R280	L282	R283	D284	Q309	T312

The Hel308 helicase preferably comprises a variant of one of SEQ ID NOs: 10, 13, 16 or 19 which comprises one or more cysteine residues and/or one or more non-natural amino acids at one or more of the positions which correspond to D274, E284, E285, S288, S615, K717 and Y720 in Hel308 Mbu (SEQ ID NO: 10). The helicase may comprise one or more cysteine residues and/or one or more non-natural amino acids at any of the following combinations of the positions labelled A to G in each row of Table 3a: {A}, {B}, {C}, {D}, {G}, {E}, {F}, {A and B}, {A and C}, {A and D}, {A and G}, {A and E}, {A and F}, {B and C}, {B and D}, {B and G}, {B and E}, {B and F}, {C and D}, {C and G}, {C and E}, {C and F}, {D and G}, {D and E}, {D and F}, {G and E}, {G and F}, {E and F}, {A, B and C}, {A, B and D}, {A, B and G}, {A, B and E}, {A, B and F}, {A, C and D}, {A, C and G}, {A, C and E}, {A, C and F}, {A, D and G}, {A, D and E}, {A, D and F}, {A, G and E}, {A, G and F}, {A, E and F}, {B, C and D}, {B, C and G}, {B, C and E}, {B, C and F}, {B, D and G}, {B, D and E}, {B, D and F}, {B, G and E}, {B, G and F}, {B, E and F}, {C, D and G}, {C, D and E}, {C, D and F}, {C, G and E}, {C, G and F}, {C, E and F}, {D, G and E}, {D, G and F}, {D, E and F}, {G, E and F}, {A, B, C and D}, {A, B, C and G}, {A, B, C and E}, {A, B, C and F}, {A, B, D and G}, {A, B, D and E},

{A, B, D and F}, {A, B, G and E}, {A, B, G and F}, {A, B, E and F}, {A, C, D and G}, {A, C, D and E}, {A, C, D and F}, {A, C, G and E}, {A, C, G and F}, {A, C, E and F}, {A, D, G and E}, {A, D, G and F}, {A, D, E and F}, {A, G, E and F}, {B, C, D and G}, {B, C, D and E}, {B, C, D and F}, {B, C, G and E}, {B, C, G and F}, {B, C, E and F}, {B, D, G and E}, {B, D, G and F}, {B, D, E and F}, {B, G, E and F}, {C, D, G and E}, {C, D, G and F}, {C, D, E and F}, {C, G, E and F}, {D, G, E and F}, {A, B, C, D and G}, {A, B, C, D and E}, {A, B, C, D and F}, {A, B, C, G and E}, {A, B, C, G and F}, {A, B, C, E and F}, {A, B, D, G and E}, {A, B, D, G and F}, {A, B, D, E and F}, {A, B, G, E and F}, {A, C, D, G and E}, {A, C, D, G and F}, {A, C, D, E and F}, {A, C, G, E and F}, {A, D, G, E and F}, {B, C, D, G and E}, {B, C, D, G and F}, {B, C, D, E and F}, {B, C, G, E and F}, {B, D, G, E and F}, {C, D, G, E and F}, {A, B, C, D, G and E}, {A, B, C, D, G and F}, {A, B, C, D, E and F}, {A, B, C, G, E and F}, {A, B, D, G, E and F}, {A, C, D, G, E and F}, {B, C, D, G, E and F}, or {A, B, C, D, G, E and F}.

The Hel308 helicase more preferably comprises a variant of one of SEQ ID NOs: 10, 13, 16 or 19 which comprises one or more cysteine residues and/or one or more non-natural amino acids at one or more of the positions which correspond to D274, E284, E285, S288 and S615 in Hel308 Mbu (SEQ ID NO: 10).

Polynucleotide binding moiety

The constructs described hereincomprise an additional polynucleotide binding moiety. A polynucleotide binding moiety is a polypeptide that is capable of binding to a polynucleotide. The moiety is preferably capable of specific binding to defined polynucleotide sequence. In other words, the moiety preferably binds to a specific polynucleotide sequence, but displays at least 10 fold less binding to different sequences or more preferably at least 100 fold less binding to different sequences or most preferably at least 1000 fold less binding to different sequences. The different sequence may be a random sequence. In some embodiments, the moiety binds to a specific polynucleotide sequence, but binding to different sequences cannot be measured. Moieties that bind to specific sequences can be used to design constructs that are targeted to such sequences.

The moiety typically interacts with and modifies at least one property of a polynucleotide. The moiety may modify the polynucleotide by cleaving it to form individual nucleotides or shorter chains of nucleotides, such as di- or trinucleotides. The moiety may modify the polynucleotide by orienting it or moving it to a specific position, i.e. controlling its movement.

A polynucleotide, such as a nucleic acid, is a macromolecule comprising two or more nucleotides. The polynucleotide or nucleic acid may comprise any combination of any

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

A nucleotide typically contains a nucleobase, a sugar and at least one phosphate group. The nucleobase is typically heterocyclic. Nucleobases include, but are not limited to, purines and pyrimidines and more specifically adenine, guanine, thymine, uracil and cytosine. The sugar is typically a pentose sugar. Nucleotide sugars include, but are not limited to, ribose and deoxyribose. The nucleotide is typically a ribonucleotide or deoxyribonucleotide. The nucleotide typically contains a monophosphate, diphosphate or triphosphate. Phosphates may be attached on the 5' or 3' side of a nucleotide.

Nucleotides include, but are not limited to, adenosine monophosphate (AMP), guanosine monophosphate (GMP), thymidine monophosphate (TMP), uridine monophosphate (UMP), 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.

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

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

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

The polynucleotide can be a nucleic acid, such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The target polynucleotide can comprise one strand of RNA hybridized to one strand of DNA. The polynucleotide may be any synthetic nucleic acid known in the art, such as peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or other synthetic polymers with nucleotide side chains.

It is preferred that the tertiary structure of the moiety is known. Knowledge of the three dimensional structure of the moiety allows modifications to be made to the moiety to facilitate its function in the construct.

The moiety may be any size and have any structure. For instance, the moiety may be an oligomer, such as a dimer or trimer. The moiety is preferably a small, globular polypeptide formed from one monomer. Such moieties are easy to handle and are less likely to interfere with the ability of the helicase to control the movement of the polynucleotide, particularly if fused to or inserted into the sequence of the helicase.

The amino and carboxy terminii of the moiety are preferably in close proximity. The amino and carboxy terminii of the moiety are more preferably presented on same face of the moiety. Such embodiments facilitate insertion of the moiety into the sequence of the helicase. For instance, if the amino and carboxy terminii of the moiety are in close proximity, each can be attached by genetic fusion to adjacent amino acids in the sequence of the helicase.

It is also preferred that the location and function of the active site of the moiety is known. This prevents modifications being made to the active site that abolish the activity of the moiety. It also allows the moiety to be attached to the helicase so that the moiety binds to the polynucleotide and controls its movement. Knowledge of the way in which a moiety may bind to and orient polynucleotides also allows an effective construct to be designed.

The constructs described herein are useful in Strand Sequencing. The moiety preferably binds the polynucleotide in a buffer background which is compatible with Strand Sequencing and the discrimination of the nucleotides. The moiety preferably has at least residual activity in a salt concentration well above the normal physiological level, such as from 100mM to 2M. The moiety is more preferably modified to increase its activity at high salt concentrations. The moiety may also be modified to improve its processivity, stability and shelf life.

Suitable modifications can be determined from the characterisation of polynucleotide binding moieties from extremphiles such as halophilic, moderately halophilic bacteria, thermophilic and moderately thermophilic organisms, as well as directed evolution approaches to altering the salt tolerance, stability and temperature dependence of mesophilic or thermophilic exonucleases.

The polynucleotide binding moiety preferably comprises one or more domains independently selected from helix-hairpin-helix (HhH) domains, eukaryotic single-stranded binding proteins (SSBs), bacterial SSBs, archaeal SSBs, viral SSBs, double-stranded binding proteins, sliding clamps, processivity factors, DNA binding loops, replication initiation proteins,

telomere binding proteins, repressors, zinc fingers and proliferating cell nuclear antigens (PCNAs).

The helix-hairpin-helix (HhH) domains are polypeptide motifs that bind DNA in a sequence non-specific manner. They have been shown to confer salt stability and processivity when fused to polymerases, as well as increasing their thermal stability. Suitable domains include domain H (residues 696-751) and domain HI (residues 696-802) from Topoisomerase V from *Methanopyrus kandleri* (SEQ ID NO: 89). As discussed below, the polynucleotide binding moiety may be domains H-L of SEQ ID NO: 89 as shown in SEQ ID NO: 94. Topoisomerase V from *Methanopyrus kandleri* is an example of a double-stranded binding protein as discussed below.

The HhH domain preferably comprises the sequence shown in SEQ ID NO: 55 or 75 or 76 or a variant thereof. This domain increases the processivity and the salt tolerance of a helicase when used in a construct described herein. A variant of SEQ ID NO: 55 or 75 or 76 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 55 or 75 or 76 and which retains polynucleotide binding activity. This can be measured as described above. A variant typically has at least 50% homology to SEQ ID NO: 55 or 75 or 76 based on amino acid identity over its entire sequence (or any of the % homologies discussed above in relation to helicases) and retains polynucleotide binding activity. A variant may differ from SEQ ID NO: 55 or 75 or 76 in any of the ways discussed above in relation to helicases or below in relation to pores. A variant preferably comprises one or more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment to the helicase as discussed above.

SSBs bind single stranded DNA with high affinity in a sequence non-specific manner. They exist in all domains of life in a variety of forms and bind DNA either as monomers or multimers. Using amino acid sequence alignment and logorithms (such as Hidden Markov models) SSBs can be classified according to their sequence homology. The Pfam family, PF00436, includes proteins that all show sequence similarity to known SSBs. This group of SSBs can then be further classified according to the Structural Classification of Proteins (SCOP). SSBs fall into the following lineage: Class; All beta proteins, Fold; OB-fold, Superfamily: Nucleic acid-binding proteins, Family; Single strand DNA-binding domain, SSB. Within this family SSBs can be classified according to subfamilies, with several type species often characterised within each subfamily.

The SSB may be from a eukaryote, such as from humans, mice, rats, fungi, protozoa or plants, from a prokaryote, such as bacteria and archaea, or from a virus.

Eukariotic SSBs are known as replication protein A (RPAs). In most cases, they are hetero-trimers formed of different size units. Some of the larger units (e.g. RPA70 of *Saccharomyces cerevisiae*) are stable and bind ssDNA in monomeric form.

Bacterial SSBs bind DNA as stable homo-tetramers (e.g. *E.coli, Mycobacterium smegmatis* and *Helicobacter pylori*) or homo-dimers (e.g. *Deinococcus radiodurans* and *Thermotoga maritima*). The SSBs from archaeal genomes are considered to be related with eukaryotic RPAs. Few of them, such as the SSB encoded by the crenarchaeote *Sulfolobus solfataricus*, are homo-tetramers. The SSBs from most other species are closer related to the replication proteins from eukaryotes and are referred to as RPAs. In some of these species they have been shown to be monomeric (*Methanococcus jannaschii* and *Methanothermobacter thermoautotrophicum*). Still, other species of Archaea, including *Archaeoglobus fulgidus* and *Methanococcoides burtonii*, appear to each contain two open reading frames with sequence similarity to RPAs. There is no evidence at protein level and no published data regarding their DNA binding capabilities or oligomeric state. However, the presence of two oligonucleotide/oligosaccharide (OB) folds in each of these genes (three OB folds in the case of one of the *M.burtonii* ORFs) suggests that they also bind single stranded DNA.

Viral SSBs bind DNA as monomers. This, as well as their relatively small size renders them amenable to genetic fusion to other proteins, for instance via a flexible peptide linker. Alternatively, the SSBs can be expressed separately and attached to other proteins by chemical methods (e.g. cysteines, unnatural amino-acids). This is discussed in more detail below.

The SSB is preferably either (i) an SSB comprising a carboxy-terminal (C-terminal) region which does not have a net negative charge or (ii) a modified SSB comprising one or more modifications in its C-terminal region which decreases the net negative charge of the C-terminal region. Such SSBs do not block the transmembrane pore and therefore allow characterization of the target polynucleotide.

Examples of SSBs comprising a C-terminal region which does not have a net negative charge include, but are not limited to, the human mitochondrial SSB (*Hs*mtSSB; SEQ ID NO: 77), the human replication protein A 70kDa subunit, the human replication protein A 14kDa subunit, the telomere end binding protein alpha subunit from *Oxytricha nova*, the core domain of telomere end binding protein beta subunit from *Oxytricha nova*, the protection of telomeres protein 1 (Pot1) from *Schizosaccharomyces pombe*, the human Pot1, the OB-fold domains of BRCA2 from mouse or rat, the p5 protein from phi29 (SEQ ID NO: 78) or a variant of any of those proteins. A variant is a protein that has an amino acid sequence which varies from that of the wild-type protein and which retains single stranded polynucleotide binding activity.

Polynucleotide binding activity can be determined using methods known in the art (and as described above). For instance, the ability of a variant to bind a single stranded polynucleotide can be determined as described in the Examples.

A variant of SEQ ID NO 77 or 78 typically has at least 50% homology to SEQ ID NO: 77 or 78 based on amino acid identity over its entire sequence (or any of the % homologies discussed above in relation to helicases) and retains single stranded polynucleotide binding activity. A variant may differ from SEQ ID NO: 77 or 78 in any of the ways discussed above in relation to helicases. In particular, a variant may have one or more conservative substitutions as shown in Tables 8 and 9.

Examples of SSBs which require one or more modifications in their C-terminal region to decrease the net negative charge include, but are not limited to, the SSB of *E. coli* (*Eco*SSB; SEQ ID NO: 79), the SSB of *Mycobacterium tuberculosis*, the SSB of *Deinococcus radiodurans*, the SSB of *Thermus thermophiles*, the SSB from *Sulfolobus solfataricus*, the human replication protein A 32kDa subunit (RPA32) fragment, the CDC13 SSB from *Saccharomyces cerevisiae*, the Primosomal replication protein N (PriB) from *E. coli*, the PriB from *Arabidopsis thaliana*, the hypothetical protein At4g28440, the SSB from T4 (gp32; SEQ ID NO: 80), the SSB from RB69 (gp32; SEQ ID NO: 56), the SSB from T7 (gp2.5; SEQ ID NO: 57) or a variant of any of these proteins. Hence, the SSB used in the method of the invention may be derived from any of these proteins.

In addition to the one or more modifications in the C-terminal region, the SSB used in the method may include additional modifications which are outside the C-terminal region or do not decrease the net negative charge of the C-terminal region. In other words, the SSB used in the method of the invention is derived from a variant of a wild-type protein. A variant is a protein that has an amino acid sequence which varies from that of the wild-type protein and which retains single stranded polynucleotide binding activity. Polynucleotide binding activity can be determined as discussed above.

The SSB used in the invention may be derived from a variant of SEQ ID NO: 56, 57, 79 or 80. In other words, a variant of SEQ ID NO: 56, 57, 79 or 80 may be used as the starting point for the SSB used in the invention, but the SSB actually used further includes one or more modifications in its C-terminal region which decreases the net negative charge of the C-terminal region. A variant of SEQ ID NO: 56, 57, 79 or 80 typically has at least 50% homology to SEQ ID NO: 56, 57, 79 or 80 based on amino acid identity over its entire sequence (or any of the % homologies discussed above in relation to helicases) and retains single stranded polynucleotide binding activity. A variant may differ from SEQ ID NO: 56, 57, 79 or 80 in any of the ways

discussed above in relation to helicases. In particular, a variant may have one or more conservative substitutions as shown in Tables 8 and 9.

It is straightforward to identify the C-terminal region of the SSB in accordance with normal protein N to C nomenclature. The C-terminal region of the SSB is preferably about the last third of the SSB at the C-terminal end, such as the last third of the SSB at the C-terminal end. The C-terminal region of the SSB is more preferably about the last quarter, fifth or eighth of the SSB at the C-terminal end, such as the last quarter, fifth or eighth of the SSB at the C-terminal end. The last third, quarter, fifth or eighth of the SSB may be measured in terms of numbers of amino acids or in terms of actual length of the primary structure of the SSB protein. The length of the various amino acids in the N to C direction are known in the art.

The C-terminal region is preferably from about the last 10 to about the last 60 amino acids of the C-terminal end of the SSB. The C-terminal region is more preferably about the last 15, about the last 20, about the last 25, about the last 30, about the last 35, about the last 40, about the last 45, about the last 50 or about the last 55 amino acids of the C-terminal end of the SSB.

The C-terminal region typically comprises a glycine and/or proline rich region. This proline/glycine rich region gives the C-terminal region flexibility and can be used to identify the C-terminal region.

Suitable modifications for decreasing the net negative charge are disclosed in US Provisional Application No. 61/673,457 (filed 19 July 2012), US Provisional Application No. 61/774,688 (filed 8 March 2013) and the International application being filed concurrently with this application (Oxford Nanopore Ref: ONT IP 035), all of which correspond to WO 2014/013259. The SSB may be any of the SSBs disclosed in the US Provisional Applications and the International application.

The modified SSB most preferably comprises a sequence selected from those shown in SEQ ID NOs: 64, 65 and 81 to 84.

Double-stranded binding proteins bind double stranded DNA with high affinity. Suitable double-stranded binding proteins include, but are not limited to Mutator S (MutS; NCBI Reference Sequence: NP_417213.1; SEQ ID NO: 105), Sso7d (*Sufolobus solfataricus P2;* NCBI Reference Sequence: NP_343889.1; SEQ ID NO: 106; Nucleic Acids Research, 2004, Vol 32, No. 3, 1197-1207), Sso10b1 (NCBI Reference Sequence: NP_342446.1; SEQ ID NO: 107), Sso10b2 (NCBI Reference Sequence: NP_342448.1; SEQ ID NO: 108), Tryptophan repressor (Trp repressor; NCBI Reference Sequence: NP_291006.1; SEQ ID NO: 109), Lambda repressor (NCBI Reference Sequence: NP_040628.1; SEQ ID NO: 110), Cren7 (NCBI Reference Sequence: NP_342459.1; SEQ ID NO: 111), major histone classes H1/H5, H2A, H2B, H3 and

H4 (NCBI Reference Sequence: NP_066403.2, SEQ ID NO: 112), dsbA (NCBI Reference Sequence: NP_049858.1; SEQ ID NO: 113), Rad51 (NCBI Reference Sequence: NP_002866.2; SEQ ID NO: 114), sliding clamps and Topoisomerase V Mka (SEQ ID NO: 89) or a variant of any of these proteins. A variant of SEQ ID NO: 89, 105, 106, 107, 108, 109, 110, 111, 112, 113 or 114 typically has at least 50% homology to SEQ ID NO: 89, 105, 106, 107, 108, 109, 110, 111, 112, 113 or 114 based on amino acid identity over its entire sequence (or any of the % homologies discussed above in relation to helicases) and retains single stranded polynucleotide binding activity. A variant may differ from SEQ ID NO: 89, 105, 106, 107, 108, 109, 110, 111, 112, 113 or 114 in any of the ways discussed above in relation to helicases. In particular, a variant may have one or more conservative substitutions as shown in Tables 8 and 9. Most polymerases achieve processivity by interacting with sliding clamps. In general, these are multimeric proteins (homo-dimers or homo-trimers) that encircle dsDNA. These sliding clamps require accessory proteins (clamp loaders) to assemble them around the DNA helix in an ATP-dependent process. They also do not contact DNA directly, acting as a topological tether. As sliding clamps interact with their cognate polymerases in a specific manner via a polymerase domain, this fragment could be fused to the helicase in order to incite recruitment of helicases onto the sliding clamp. This interaction could be further stabilized by the generation of a covalent bond (introduction of cysteines or unnatural amino-acids).

Related to DNA sliding clamps, processivity factors are viral proteins that anchor their cognate polymerases to DNA, leading to a dramatic increase the length of the fragments generated. They can be monomeric (as is the case for UL42 from *Herpes simplex virus 1*) or multimeric (UL44 from *Cytomegalovirus* is a dimer), they do not form closed rings around the DNA strand and they contact DNA directly. UL42 has been shown to increase processivity without reducing the rate of its corresponding polymerase, suggesting that it interacts with DNA in a different mode to SSBs. The UL42 preferably comprises the sequence shown in SEQ ID NO: 58 or SEQ ID NO: 63 or a variant thereof. A variant of SEQ ID NO: 58 or 63 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 58 or 63 and which retains polynucleotide binding activity. This can be measured as described above. A variant typically has at least 50% homology to SEQ ID NO: 58 or 63 based on amino acid identity over its entire sequence (or any of the % homologies discussed above in relation to helicases) and retains polynucleotide binding activity. A variant may differ from SEQ ID NO: 58 or SEQ ID NO: 63 in any of the ways discussed above in relation to helicases or below in relation to pores. A variant preferably comprises one or more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment to the helicase as discussed above.

Attaching UL42 to a helicase could be done via genetic fusion or chemical attachment (cysteines, unnatural amino-acids). As the polymerase polypeptide that binds UL42 is visible in the crystal structure, these 35 amino acids (residues 1200-1235) could be fused onto the C-terminus of the helicase and the natural affinity between this polypeptide and the processivity factor used to form a complex. The interaction could be stabilized by introducing a covalent interaction (cysteines or unnatural amino-acids). One option is to utilize a natural UL42 cysteine (C300) that is located close to the polypeptide interaction site and introduce a point mutation into the polymerase polypeptide (e.g. L1234C).

A reported method of increasing polymerase processivity is by exploiting the interaction between *E.coli* thioredoxin (Trx) and the thioredoxin binding domain (TBD) of bacteriophage T7 DNA polymerase (residues 258-333). The binding of Trx to TBD causes the polypeptide to change conformation to one that binds DNA. TBD is believed to clamp down onto a DNA strand and limit the polymerase off-rate, thus increasing processivity. Chimeric polymerases have been made by transferring TBD onto a non-processive polymerase, resulting in 1000 fold increase in polymerised fragment length. There were no attempts to attach TBD to any other class of proteins, but a covalent link between TBD and Trx was engineered and can be used to stabilise the interaction.

Some helicases use accessory proteins in-vivo to achieve processivity (e.g. cisA from phage Φx174 and geneII protein from phage M13 for *E.coli* Rep helicase). Some of these proteins have been shown to interact with more than one helicase (e.g. MutL acts on both UvrD and Rep, though not to the same extent). These proteins have intrinsic DNA binding capabilities, some of them recognizing a specific DNA sequence. The ability of some of these accessory proteins to covalently attach themselves to a specific DNA sequence could also be used to create a set starting point for the helicase activity.

The proteins that protect the ends of chromosomes bind to telomeric ssDNA sequences in a highly specific manner. This ability could either be exploited as is or by using point mutations to abolish the sequence specificity.

Small DNA binding motifs (such as helix-turn-helix) recognize specific DNA sequences. In the case of the bacteriophage 434 repressor, a 62 residue fragment was engineered and shown to retain DNA binding abilities and specificity.

An abundant motif in eukaryotic proteins, zinc fingers consist of around 30 amino-acids that bind DNA in a specific manner. Typically each zinc finger recognizes only three DNA bases, but multiple fingers can be linked to obtain recognition of a longer sequence.

Proliferating cell nuclear antigens (PCNAs) form a very tight clamp (doughnut) which slides up and down the dsDNA or ssDNA. The PCNA from *crenarchaeota* is unique in being a hetero-trimer so it is possible to functionalise one subunit and retain activity. Its subunits are shown in SEQ ID NOs: 59, 60 and 61. The PCNA is preferably a trimer comprising the sequences shown in SEQ ID NOs: 59, 60 and 61 or a variant thereof. PCNA sliding clamp (NCBI Reference Sequence: ZP_06863050.1; SEQ ID NO: 115) forms a dimer. The pCNA is preferably a dimer comprising SEQ ID NO: 115 or a variant thereof. A variant is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 59, 60, 61 and 115 and which retains polynucleotide binding activity. This can be measured as described above. A variant is typically a trimer comprising sequences that have at least 50% homology to SEQ ID NOs: 59, 60, 61 and 115 respectively based on amino acid identity over their entire sequences (or any of the % homologies discussed above in relation to helicases) and which retains polynucleotide binding activity. A variant may comprise sequences which differ from SEQ ID NO: 59, 60, 61 or 115 in any of the ways discussed above in relation to helicases or below in relation to pores. A variant preferably comprises one or more substituted cysteine residues and/or one or more substituted Faz residues to facilitate attachment to the helicase as discussed above. In a preferred embodiment, subunits 1 and 2 (i.e. SEQ ID NOs: 59 and 60 or variants thereof) are attached, such as genetically fused, and the resulting protein is attached to a helicase to form a construct. During use of the construct, subunit 3 (i.e. SEQ ID NO: 61 or a variant thereof) may be added to complete the PCNA clamp (or doughnut) once the construct has bound the polynucleotide. In a preferred embodiment, one monomer (i.e. SEQ ID NO: 111 or variants thereof) is attached, such as genetically fused, and the resulting protein is attached to a helicase to form a construct of the invention. During use of the construct, the second monomer (i.e. SEQ ID NO: 111 or a variant thereof) may be added to complete the the PCNA clamp (or doughnut) once the construct has bound the polynucleotide.

The polynucleotide binding motif may be selected from any of those shown in Table 4 below.

Table 4. Suitable polynucleotide binding motifs

No.	Name	Class	Organism	Structure	Sequence	Functional form	MW (Da)	Notes
1	SSBEco	ssb	Escherichia coli	1QVC, 1EYG	P0AGE0	homo-tetramer	18975	
2	SSBBhe	ssb	Bartonella henselae	3LGJ, 3PGZ	Q6G302	homo-tetramer	16737	structure only
3	SSBCb	ssb	Coxiella	3TQY	Q83EP4	homo-	17437	structure only

	u		burnetii			tetramer		
4	SSBTma	ssb	Thermathoga maritima	1Z9F	Q9WZ73	homo-dimer	16298	small, thermostable, salt independent DNA binding
5	SSBHpy	ssb	Helicobacter pylori	2VW9	O25841	homo-tetramer	20143	
6	SSBDra	ssb	Deinococcus radiodurans	1SE8	Q9RY51	homo-dimer	32722	
7	SSBTaq	ssb	Thermus aquaticus	2FXQ	Q9KH06	homo-dimer	30026	
8	SSBMsm	ssb	Mycobacterium smegmatis	3A5U,1X3E	Q9AFI5	homo-tetramer	17401	tetramer more stable than E.coli, binding less salt dependent
9	SSBSso	ssb/RPA	Sulfolobus solfataricus	1O7I	Q97W73	homo-tetramer	16138	similarities with RPA
10	SSBMHsmt	ssb	Homo sapiens	3ULL	Q04837	homo-tetramer	17260	
11	SSBMle	ssb	Mycobacterium leprae	3AFP	P46390	homo-tetramer	17701	
12	gp32T4	ssb	Bacteriohage T4	1GPC	P03695	monomer	33506	Homo-dimer in the absence of DNA, monomer when binding DNA.
13	gp32RB69	ssb	Bacteriophage RB69	2A1K	Q7Y265	monomer	33118	
14	gp2.5T7	ssb	Bacteriohage T7	1JE5	P03696	monomer	25694	
15	UL42	processivity factor	Herpes virus 1	1DML	P10226	monomer	51159	binds ssDNA dsDNA, structure shows link with polymerase
16	UL44	processivity factor	Herpes virus 5 (cytomegalovirus)	1YYP	P16790	homo-dimer	46233	forms C shaped clamp on DNA
17	pf8	processivity factor	KSHV	3I2M	Q77ZG5	homo-dimer	42378	
18	RPAMja	RPA	Methanococcus jannaschii	3DM3	Q58559	monomer	73842	contains 4 OB folds. Structure of fragment
19	RPAMma	RPA	Methanococcus maripaludis	3E0E, 2K5V	Q6LYF9	monomer	71388	Core domain structure
20	RPAMth	RPA	Methanothermobacter thermoautotrophicus			monomer	120000	Shown to interact directly with Hel308. Sequence from paper.
21	RPA70Sce	RPA	Saccharomyces cerevisiae	1YNX	P22336	hetero-trimer	70348	unit has two OB folds and binds DNA
22	RPAMbu1	RPA	Methanococcoides burtonii		Q12V72	?	41227	three OB folds identified
23	RPAMbu2	RPA	Methanococcoides burtonii		Q12W96	?	47082	two OB folds identified

24	RPA70 Hsa	RPA	Homo sapiens	1JMC	P27694	hetero-trimer	68138	
25	RPA14 Hsa	RPA	Homo sapiens	3KDF	P35244	hetero-trimer	13569	in complex with RPA32
26	gp45T4	sliding clamp	Bacteriophage T4	1CZD	P04525	homo-trimer	24858	ring shape threads DNA
27	BetaEco	sliding clamp	E.coli	3BEP	P0A988	homo-dimer	40587	ring shape threads DNA, may bind ssDNA in poket
28	PCNA Sce	sliding clamp	Saccharomyces cerevisiae	1PLQ,3K4X	P15873	homo-dimer	28916	ring shape threads DNA
29	PCNA Tko	sliding clamp	Thermococcus kodakaraensis	3LX1	Q5JF32	homo-dimer	28239	
30	PCNA Hvo	sliding clamp	Haloferax volcanii	3IFV	D0VWY8	homo-dimer	26672	
31	PCNA Pfu	sliding clamp	Pyrococcus furiosus	1GE8	O73947	homo-dimer	28005	
32	PCNA Mbu	sliding clamp	Methanococcoides burtonii		Q12U18	homo-dimer	27121	Inferred from homology
33	BetaMtu	sliding clamp	Mycobacterium tuberculosis	3P16	Q50790	homo-dimer	42113	
34	BetaTma	sliding clamp	Thermotoga maritima	1VPK	Q9WYA0	homo-dimer	40948	
35	BetaSpy	sliding clamp	Streptococcus pyrogenes	2AVT	Q9EVR1	homo-dimer	41867	
36	gp45RB69	sliding clamp	Bacteriophage RB69	1B77	O80164	homo-trimer	25111	Structure shows interaction with polypeptide fom polymerase
37	p55Hsa	DNA binding protein	Homo sapiens (mitochondrial)	2G4C, 3IKL , 3IKM	Q9UHN	monomer	54911	interacts with specific polymerase domain
38	p55Dme	DNA binding protein	Drosophylla melanogaster		Q9VJV8	monomer	41027	associates with polymerase Gamma conferring salt tolerance, processivity and increased activity
39	p55Xla	DNA binding protein	Xenopus laevis		Q9W6G7	monomer	52283	

40	RepDSau	replication initiation protein	Staphylococcus aureus		P08115	homo-dimer	37874	increases processivity of PcrA, covalently and specifically links DNA
41	G2P	replication initiation protein	Enterobacteria phage 1		P69546	monomer	46168	increases processivity of Rep, covalently and specifically links DNA
42	MutLEco	mismatch repair protein	Escherichia coli	1BKN, 1B62, 1B63	P23367	homo-dimer	67924	increases processivity of UvrD (and Rep)
43	KuMtu	DNA repair protein	Mycobacterium tuberculosis		O05866	homo-dimer	30904	increases processivity of UvrD1. Structure available for human Ku
44	OnTEBP	telomere binding protein	Oxytricha nova-Alpha	1OTC	P29549	hetero-dimer	56082	Specific biding to 3' end T4G4T4G4. Alpha subunit may be enough
			Oxytricha nova-Beta		P16458		41446	
45	EcrTEBP	telomere binding protein	Euplotes crassus		Q06183	monomer	53360	Homolog to OnTEBP with no Beta subunit in genome
46	TteTEBP	telomere binding protein	Tetrachymena termophila Alpha		Q23FB9	hetero-dimer	53073	Homolog to OnTEBP-Alpha
			Tetrachymena termophila Beta		Q23FH0		54757	May be homolog to OnTEBP Beta
47	pot1Spo	telomere binding proteins	Schizosaccharomyces pombe		O13988	monomer	64111	related to TEBP
48	Cdc13pSce	telomere binding	Saccharomyces cerevisiae		C7GSV7	monomer	104936	specific binding to telomeric DNA

		proteins						
49	C1	repressor	Bacteriophage 434		P16117	homo-dimer	10426	binds DNA specifically as homo-dimer
50	LexA	repressor	Escherichia coli	1LEB	P0A7C2	homo-dimer	22358	binds DNA specifically as homo-dimer

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

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

Preferred enzymes are exonucleases, polymerases, helicases and topoisomerases, such as gyrases. Suitable exonucleases include, but are not limited to, exonuclease I from *E. coli*, exonuclease III enzyme from *E. coli*, RecJ from *T. thermophilus* and bacteriophage lambda exonuclease and variants thereof.

The polymerase is preferably a member of any of the Moiety Classification (EC) groups 2.7.7.6, 2.7.7.7, 2.7.7.19, 2.7.7.48 and 2.7.7.49. The polymerase is preferably a DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase, a DNA-dependent RNA polymerase or an RNA-dependent RNA polymerase. The polynucleotide binding moiety is preferably derived from Phi29 DNA polymerase (SEQ ID NO: 62). The moiety may comprise the sequence shown in SEQ ID NO: 62 or a variant thereof. A variant of SEQ ID NO: 62 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 62 and which retains polynucleotide binding activity. The variant may include modifications that facilitate binding of the polynucleotide and/or facilitate its activity at high salt concentrations and/or room temperature.

Over the entire length of the amino acid sequence of SEQ ID NO: 62, 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: 62 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 is determined as described below. The variant may differ from the wild-type sequence in any of the ways discussed below with reference to SEQ ID NOs: 2 and 4.

The helicase may be any of those discussed above. Helicase dimers and multimers are discussed in detail below. The polynucleotide binding moiety may be a polynucleotide binding domain derived from a helicase. For instance, the polynucleotide binding moiety preferably comprises the sequence shown in SEQ ID NO: 66 or 67 or a variant thereof. A variant of SEQ ID NO: 66 or 67 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 66 or 67 and which retains polynucleotide binding activity. This can be measured as described above. The variant may include modifications that facilitate binding of the polynucleotide and/or facilitate its activity at high salt concentrations and/or room temperature.

Over the entire length of the amino acid sequence of SEQ ID NO: 66 or 67, 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: 66 or 67 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 40 or more, for example 50, 60, 70 or 80 or more, contiguous amino acids ("hard homology"). Homology is determined as described above. The variant may differ from the wild-type sequence in any of the ways discussed below with reference to SEQ ID NOs: 2 and 4.

The topoisomerase is preferably a member of any of the Moiety Classification (EC) groups 5.99.1.2 and 5.99.1.3.

The polynucleotide binding moiety may be any of the enzymes discussed above. The construct preferably does not comprise a helicase attached to a polymerase, a topoisomerase or a primase. The construct preferably does not comprise two or more NS3 helicases attached

together. The construct more preferably does not comprise two or more NS3 helicases attached together using bis(sulfosuccinimidyl) suberate (BS^3).

The moiety may be labelled with a revealing label. The label may be any of those described above.

The moiety may be isolated from any moiety-producing organism, such as *E. coli*, *T. thermophilus* or bacteriophage, or made synthetically or by recombinant means. For example, the moiety may be synthesized by *in vitro* translation and transcription as described below. The moiety may be produced in large scale following purification as described below.

Preferred constructs

As will be clear from the discussion above, the polynucleotide binding moiety is preferably derived from a helicase. For instance, it may be a polynucleotide domain from a helicase. The moiety more preferably comprises one or more helicases. The helicases may be any of those discussed above. In such embodiments, the constructs of course comprise two or more helicases attached together. The invention provides a construct comprising two or more helicases attached together. As discussed above, each helicase must be capable of functioning as a helicase on its own. Any of the embodiments discussed above with reference to the constructs that may be used in the method of the invention, and in particular the types of helicases that may be used and the methods of attachment, are equally applicable to the constructs of the invention. The two or more helicases are preferably monomeric helicases. The two or more helicases are preferably not two or more helicase domains from helicase enzymes.

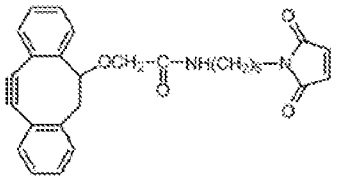
The construct of the invention preferably does not comprise two or more NS3 helicases attached together. The construct of the invention more preferably does not comprise two or more NS3 helicases attached together using bis(sulfosuccinimidyl) suberate (BS^3).

The construct may comprise two, three, four, five or more helicases. In other words, the constructs of the invention may comprise a helicase dimer, a helicase trimer, a helicase tetramer, a helicase pentamer and the like.

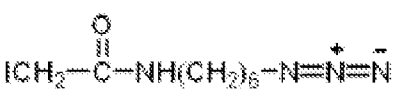
The two or more helicases can be attached together in any orientation. Identical or similar helicases may be attached via the same amino acid residue (i.e. same position) or spatially proximate amino acid residues (i.e. spatially proximate positions) in each helicase. This is termed the "head-to-head" formation. Alternatively, identical or similar helicases may be attached via amino acid residues (or positions) on opposite or different sides of each helicase. This is termed the "head-to-tail" formation. Helicase trimers comprising three identical or similar helicases may comprise both the head-to-head and head-to-tail formations.

The two or more helicases may be different from one another (i.e. the construct is a hetero-dimer, -trimer, -tetramer or -pentamer etc.). For instance, the constructs of the invention may comprise: (a) one or more Hel308 helicases and one or more XPD helicases; (b) one or more Hel308 helicases and one or more RecD helicases; (c) one or more Hel308 helicases and one or more TraI helicases; (d) one or more XPD helicases and one or more RecD helicases; (e) one or more XPD helicases and one or more TraI helicases; or (f) one or more RecD helicases and one or more TraI helicases. The construct may comprise two different variants of the same helicase. For instance, the construct may comprise two variants of one of the helicases discussed above with one or more cysteine residues or Faz residues introduced at different positions in each variant. In this instance, the helicases can be in a head-to-tail formation. In a preferred embodiment, a variant of SEQ ID NO: 10 comprising Q442C may be attached via cysteine linkage to a variant of SEQ ID NO: 10 comprising Q557C. Cys mutants of Hel308Mbu can also be made into hetero-dimers if necessary. In this approach, two different Cys mutant pairs such as Hel308Mbu-Q442C and Hel308Mbu-Q577C can be linked in head-to-tail fashion. Hetero-dimers can be formed in two possible ways. The first involves the use of a homo-bifunctional linker as discussed above. One of the helicase variants can be modified with a large excess of linker in such a way that one linker is attached to one molecule of the protein. This linker modified variant can then be purified away from unmodified proteins, possible homo-dimers and unreacted linkers to react with the other helicase variant. The resulting dimer can then be purified away from other species.

The second involves the use of hetero-bifunctional linkers. For example, one of the helicase variants can be modified with a first PEG linker containing maleimide or iodoacetamide functional group at one end and a cyclooctyne functional group (DIBO) at the other end. An example of this is shown below:



The second helicase variant can be modified with a second PEG linker containing maleimide or iodioacetamide functional group at one end and an azide functional group at the other end. An example is show below:

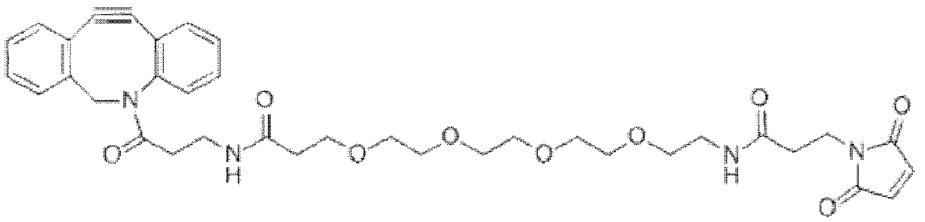


The two helicase variants with two different linkers can then be purified and clicked together (using Cu^{2+} free click chemistry) to make a dimer. Copper free click chemistry has been used in these applications because of its desirable properties. For example, it is fast, clean and not poisonous towards proteins. However, other suitable bio-orthogonal chemistries include, but are not limited to, Staudinger chemistry, hydrazine or hydrazide/aldehyde or ketone reagents (HyNic + 4FB chemistry, including all Solulink™ reagents), Diels-Alder reagent pairs and boronic acid/salicyhydroxamate reagents.

These two ways of linking two different variants of the same helicase are also valid for any of the constructs discussed above in which the helicase and the moiety are different from one another, such as dimers of two different helicases and a helicase-polymerase dimer.

Similar methodology may also be used for linking different Faz variants. One Faz variant (such as SEQ ID NO: 10 comprising Q442C) can be modified with a large excess of linker in such a way that one linker is attached to one molecule of the protein. This linker modified Faz variant can then be purified away from unmodified proteins, possible homo-dimers and unreacted linkers to react with the second Faz variant (such as SEQ ID NO: 10 comprising Q577Faz). The resulting dimer can then be purified away from other species.

Hetero-dimers can also be made by linking cysteine variants and Faz variants of the same helicase or different helicases. For example, any of the above cysteine variants (such as SEQ ID NO: 10 comprising Q442C) can be used to make dimers with any of the above Faz variants (such SEQ ID NO: 10 comprising Q577Faz). Hetero-bifunctional PEG linkers with maleimide or iodoacetamide functionalities at one end and DBCO functionality at the other end can be used in this combination of mutants. An example of such a linker is shown below (DBCO-PEG4-maleimide):

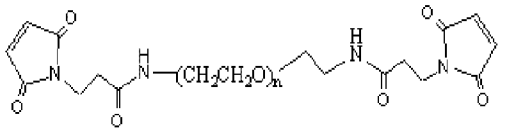


The length of the linker can be varied by changing the number of PEG units between the two functional groups.

Helicase hetero-trimers can comprise three different types of helicases selected from Hel308 helicases, XPD helicases, RecD helicases, TraI helicases and variants thereof. The same is true for oligomers comprising more than three helicases. The two or more helicases within a

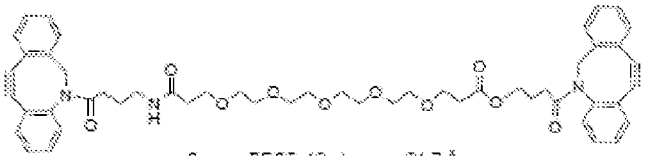
construct may be different variants of the same helicase, such as different variants of SEQ ID NO: 10, 13, 16 or 19. The different variants may be modified at different positions to facilitate attachment via the different positions. The hetero-trimers may therefore be in a head-to-tail and head-to-head formation.

The two or more helicases in the constructs of the invention may be the same as one another (i.e. the construct is a homo-dimer, -trimer, -tetramer or –pentamer etc.) Homo-oligomers can comprise two or more Hel308 helicases, two or more XPD helicases, two or more RecD helicases, two or more TraI helicases or two or more of any of the variants discussed above. In such embodiments, the helicases are preferably attached using the same amino acid residue (i.e. same position) in each helicase. The helicases are therefore attached head-to-head. The helicases may be linked using a cysteine residue or a Faz residue that has been substituted into the helicases at the same position. Cysteine residues in identical helicase variants can be linked using a homo-bifunctional linker containing thiol reactive groups such as maleimide or iodoacetamide. These functional groups can be at the end of a polyethyleneglycol (PEG) chain as in the following example:



The length of the linker can be varied to suit the required applications. For example, n can be 2, 3, 4, 8, 11, 12, 16 or more. PEG linkers are suitable because they have favourable properties such as water solubility. Other non PEG linkers can also be used in cysteine linkage.

By using similar approaches, identical Faz variants can also be made into homo-dimers. Homo-bifunctional linkers with DIBO functional groups can be used to link two molecules of the same Faz variant to make homo-dimers using Cu^{2+} free click chemistry. An example of a linker is given below:



The length of the PEG linker can vary to include 2, 4, 8, 12, 16 or more PEG units. Such linkers can also be made to incorporate a florescent tag to ease quantifications. Such fluorescence tags can also be incorporated into Maleimide linkers.

Preferred constructs of the invention are shown in the Table 5 below.

Preferred constructs of the invention

Hel308Mbu-A700C dimer 2kDa
Hel308Mbu-A700C dimer 3.4kDa
Hel308Mbu-Q442C 2kDa linker homodimer
Hel308Mbu-Q442C 3.4kDa linker homodimer
Hel308Mbu-A700C 2kDa linker homodimer
Hel308Mbu-A700C-strepII. 2kDa PEG homodimer
Hel308Mhu-WT 2kDa Dimer
Helicase 2k dimer (Hel308Mbu R681A, R687A, A700C - STrEP)
Helicase 2k dimer (Hel308Mbu R687A, A700C - STrEP)
Hel308Mhu-WT 2kDa Dimer
Hel308Tga N674C Dimer 2kDa
Hel308Tga N674C Dimer 2kDa tests for assay
Hel308 Tga-R657A-N674C-STrEP Dimer 2kDa

Preferred constructs of the invention are shown in the Table 6 below. Each row shows a preferred construct in which the helicase in the left-hand column is attached to additional polynucleotide binding moiety in the right-hand column in accordance with the invention.

5

Helicase	**Additional polynucleotide binding moiety**
Hel308 helicase as defined above (preferably SEQ ID NO: 10, 13, 16 or 19 or a variant thereof as defined above)	Polymerase (preferably SEQ ID NO: 62 or a variant thereof as defined above)
TraI helicase as defined above (preferably SEQ ID NO: 46, 87, 98 and 102 or a variant thereof as defined above)	Polymerase (preferably SEQ ID NO: 62 or a variant thereof as defined above)
Hel308 helicase as defined above (preferably SEQ ID NO: 10, 13, 16 or 19 or a variant thereof as defined above)	Hel308 helicase as defined above (preferably SEQ ID NO: 10, 13, 16 or 19 or a variant thereof as defined above)
TraI helicase as defined above (preferably SEQ ID NO: 46, 87, 98 and 102 or a variant thereof as defined above)	TraI helicase as defined above (preferably SEQ ID NO: 46, 87, 98 and 102 or a variant thereof as defined above)
Hel308 helicase as defined above (preferably SEQ ID NO: 10, 13, 16 or 19 or a variant thereof as defined above)	TraI helicase as defined above (preferably SEQ ID NO: 46, 87, 98 and 102 or a variant thereof as defined above)
TraI helicase as defined above (preferably SEQ ID NO: 46, 87, 98 and 102 or a variant thereof	Hel308 helicase as defined above (preferably SEQ ID NO: 10, 13, 16 or 19 or a variant

as defined above)	thereof as defined above)

The invention also provides a construct comprising a helicase and an amino acid sequence comprising SEQ ID NO: 94 (H-L domains from Topoisomerase V from *Methanopyrus kandleri;* SEQ ID NO: 89) or a variant thereof having at least 80% homology to SEQ ID NO: 94 based on amino acid identity over the entire sequence of SEQ ID NO: 94, wherein the helicase is attached to the amino acid sequence and the construct has the ability to control the movement of a polynucleotide. The helicase may be attached to the amino acid sequence in any of the ways discussed above.

The construct preferably comprises SEQ ID NO: 90 or a variant thereof having at least 80% homology to SEQ ID NO: 90 based on amino acid identity over the entire sequence of SEQ ID NO: 90.

Over the entire length of the amino acid sequence of SEQ ID NO: 94 or 90, a variant will preferably be at least 80% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be 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: 94 or 90 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 is determined as described below. The variant may differ from the wild-type sequence in any of the ways discussed below with reference to SEQ ID NOs: 2 and 4.

Polynucleotide sequences

The present invention also provides polynucleotide sequences which encode a construct in which the two or more helicases are genetically fused. It is straightforward to generate such polynucleotide sequences using standard techniques. A polynucleotide sequence encoding a helicase may be either fused to or inserted into a polynucleotide sequence encoding another helicase. The fusion or insertion is typically in frame. If a polynucleotide sequence encoding a helicase is inserted into a polynucleotide sequence encoding another helicase, the sequence encoding the moiety is typically flanked at both ends by restriction endonuclease sites, such as those recognized by *Bsp*E1. It may also be flanked at both ends by polynucleotide sequences encoding linkers, such as 5 to 10 codons each encoding serine or glycine.

Polynucleotide sequences may be isolated and replicated using standard methods in the art. Chromosomal DNA may be extracted from a helicase producing organism, such *as*

Methanococcoides burtonii, and/or a moiety producing organism, such as *E. coli, T. thermophilus* or bacteriophage. The gene encoding the helicase and moiety 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 encoding a helicase and/or moiety may be made by introducing a polynucleotide encoding a helicase and/or moiety 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 encoding a construct 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.

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 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 encoding a construct 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 C41 (DE3), BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, Origami and Origami B, can express a vector comprising the T7 promoter.

Methods of the invention

The invention provides a method of controlling the movement of a target polynucleotide using a construct of the invention, i.e. a construct comprising two or more helicases attached together. The method comprises contacting the target polynucleotide with a construct of the invention and thereby controlling the movement of the polynucleotide. The method is preferably carried out with a potential applied across the pore. As discussed in more detail below, the applied potential typically results in the formation of a complex between the pore and the construct. The applied potential may be a voltage potential. Alternatively, the applied potential may be a chemical potential. An example of this is using a salt gradient across an amphiphilic layer. A salt gradient is disclosed in Holden *et al.*, J Am Chem Soc. 2007 Jul 11;129(27):8650-5.

The invention also provides a method of characterising a target polynucleotide. The method comprises (a) contacting the target polynucleotide with a transmembrane pore and a construct described herein such that the construct controls the movement of the target polynucleotide through the pore. The method also comprises (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 target polynucleotide and thereby characterising the target polynucleotide.

Steps (a) and (b) are preferably carried out with a potential applied across the pore as discussed above. In some instances, the current passing through the pore as the polynucleotide moves with respect to the pore is used to determine the sequence of the target polynucleotide. This is Strand Sequencing.

The method of the invention is for characterising a target polynucleotide. A polynucleotide is defined above.

The whole or only part of the target polynucleotide may be characterised using this method. The target polynucleotide can be any length. For example, the polynucleotide can be at least 10, at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 400 or at least 500 nucleotide pairs in length. The polynucleotide can be 1000 or more nucleotide pairs, 5000 or more nucleotide pairs in length or 100000 or more nucleotide pairs in length.

The target polynucleotide is present in any suitable sample. The invention is typically carried out on a sample that is known to contain or suspected to contain the target polynucleotide. Alternatively, the invention may be carried out on a sample to confirm the

identity of one or more target polynucleotides whose presence in the sample is known or expected.

The sample may be a biological sample. The invention may be carried out *in vitro* on a sample obtained from or extracted from any organism or microorganism. The organism or microorganism is typically archaeal, prokaryotic or eukaryotic and typically belongs to one of the five kingdoms: plantae, animalia, fungi, monera and protista. The invention may be carried out *in vitro* on a sample obtained from or extracted from any virus. The sample is preferably a fluid sample. The sample typically comprises a body fluid of the patient. The sample may be urine, lymph, saliva, mucus or amniotic fluid but is preferably blood, plasma or serum. Typically, the sample is human in origin, but alternatively it may be from another mammal animal such as from commercially farmed animals such as horses, cattle, sheep 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, 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.

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

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

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

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

The 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 Example. Lipid bilayers are commonly formed by the method of Montal and Mueller (Proc. Natl. Acad. Sci. USA., 1972; 69: 3561-3566), in which a lipid monolayer is carried on aqueous solution/air interface past either side of an aperture which is perpendicular to that interface.

The 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, $A1_2O_3$, and SiO, organic and inorganic polymers such as polyamide, plastics such as Teflon® or elastomers such as two-component addition-cure silicone rubber, and glasses. The solid state layer may be formed from 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 polynucleotide may be 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 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 polynucleotide may be coupled directly to the membrane. The 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 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 helicase. If a linker is used, then the polynucleotide can be processed to completion. If a linker is used, the linker may be attached to the polynucleotide at any position. The linker is preferably attached to the 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 bilayer and the helicase's active site. If the coupling is transient, then when the coupled end randomly becomes free of the bilayer, 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 7 below.

Table 7

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." J Am Chem Soc **125**(13): 3696-7.
Biotin	Stable	Nikolov, V., R. Lipowsky, et al. (2007). "Behavior of giant vesicles with anchored DNA molecules." Biophys J **92**(12): 4356-68
Cholesterol	Transient	Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucletides to lipid membrane assemblies." J Am Chem Soc **126**(33): 10224-5
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." Langmuir **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 bilayer. 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 ATPγS (Grant, G. P. and P. Z. Qin (2007). "A facile method for attaching nitroxide spin labels at the 5' terminus of nucleic acids." Nucleic Acids Res **35**(10):

e77). 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 the addition of a short piece of DNA complementary to one already coupled to the bilayer, so that attachment can be achieved via hybridisation. Ligation of short pieces of ssDNA have been reported using T4 RNA ligase I (Troutt, A. B., M. G. McHeyzer-Williams, et al. (1992). "Ligation-anchored PCR: a simple amplification technique with single-sided specificity." Proc Natl Acad Sci U S A 89(20): 9823-5). Alternatively either ssDNA or dsDNA could be ligated to native dsDNA and then the two strands separated by thermal or chemical denaturation. To native dsDNA, it is possible to add either a piece of ssDNA to one or both of the ends of the duplex, or dsDNA to one or both ends. Then, when the duplex is melted, each single strand will have either a 5' or 3' modification if ssDNA was used for ligation or a modification at the 5' end, the 3' end or both if dsDNA was used for ligation. If the polynucleotide is a synthetic strand, the coupling chemistry can be incorporated during the chemical synthesis of the polynucleotide. For instance, the polynucleotide can be synthesized using a primer with a reactive group attached to it.

A common technique for the amplification of sections of genomic DNA is using polymerase chain reaction (PCR). Here, using two synthetic oligonucleotide primers, a number of copies of the same section of DNA can be generated, where for each copy the 5' of each strand in the duplex will be a synthetic polynucleotide. By using an antisense primer that has a reactive group, such as a cholesterol, thiol, biotin or lipid, each copy of the amplified target DNA will contain a reactive group for coupling.

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, 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, 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 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 and is called MS-(B2)8. The pore used in the invention is preferably MS-(B2)8. A variant of SEQ ID NO: 2 has the mutations G75S/G77S/L88N/Q126R in addition to all the mutations of MS-B1 and is called MS-B2C. The pore used in the invention is preferably MS-(B2)8 or MS-(B2C)8.

Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions replace amino acids with other amino acids of similar chemical structure, similar chemical properties or similar side-chain volume. The amino acids introduced may have similar polarity, hydrophilicity, hydrophobicity, basicity, acidity, neutrality or charge to the amino acids they replace. Alternatively, the conservative substitution may introduce another amino acid that is aromatic or aliphatic in the place of a pre-existing aromatic or aliphatic amino acid. 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 8 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 9.

Table 8 – 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 9 - Hydropathy scale

Side Chain	Hydropathy
Ile	4.5
Val	4.2

Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

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

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

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

As discussed above, a variant is a polypeptide that has an amino acid sequence which varies from that of SEQ ID NO: 2 and which retains its ability to form a pore. A variant typically contains the regions of SEQ ID NO: 2 that are responsible for pore formation. The pore forming ability of Msp, which contains a β-barrel, is provided by β-sheets in each subunit. A variant of SEQ ID NO: 2 typically comprises the regions in SEQ ID NO: 2 that form β-sheets.

One or more modifications can be made to the regions of SEQ ID NO: 2 that form β-sheets as long as the resulting variant retains its ability to form a pore. A variant of SEQ ID NO: 2 preferably includes one or more modifications, such as substitutions, additions or deletions, within its α-helices and/or loop regions.

The monomers derived from Msp may be modified to assist their identification or purification, for example by the addition of histidine residues (a hist 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 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 above.

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

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

The monomer derived from Msp can be produced using standard methods known in the art. The monomer derived from Msp may be made synthetically or by recombinant means. For example, the 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 construct. The variant preferably comprises one or more reactive cysteine residues that facilitate attachment to the construct. 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.

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

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

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

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

The variant 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.

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 construct may be covalently attached to the pore. The construct is preferably not covalently attached to the pore. The application of a voltage to the pore and construct typically results in the formation of a sensor that is capable of sequencing target polynucleotides. This is discussed in more detail below.

Any of the proteins described herein, i.e. the transmembrane protein pores or constructs, 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 pore or construct. 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 pore and/or construct may be labelled with a revealing label. The revealing label may be any suitable label which allows the pore to be detected. Suitable labels include, but are not limited to, fluorescent molecules, radioisotopes, e.g. ^{125}I, ^{35}S, enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

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

The pore and/or construct may also be produced using D-amino acids. For instance, the pore or construct 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 pore and/or construct may also contain other non-specific modifications as long as they do not interfere with pore formation or construct 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 $NaBH_4$, amidination with methylacetimidate or acylation with acetic anhydride.

The pore and construct can be produced using standard methods known in the art. Polynucleotide sequences encoding a pore or construct may be derived and replicated using standard methods in the art. Polynucleotide sequences encoding a pore or construct may be expressed in a bacterial host cell using standard techniques in the art. The pore and/or construct may be produced in a cell by *in situ* expression of the polypeptide from a recombinant expression vector. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide. These methods are described in Sambrook, J. and Russell, D. (2001). Molecular Cloning: A Laboratory Manual, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The pore and/or construct 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.

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 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.

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 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.

In a preferred embodiment, the method comprises:

(a) contacting the target polynucleotide with a transmembrane pore and a construct described herein such that the target polynucleotide moves through the pore and the construct controls the movement of the target polynucleotide through the pore; and

(b) measuring the current passing through the pore as the polynucleotide moves with respect to the pore wherein the current is indicative of one or more characteristics of the target polynucleotide and thereby characterising the target polynucleotide.

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

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

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

The methods of the invention may involve the measuring of a current passing through the pore as the polynucleotide moves with respect to the pore. Suitable conditions for measuring ionic currents through transmembrane protein pores are known in the art and disclosed in the Example. The method is typically carried out with a voltage applied across the membrane and pore. The voltage used is typically from +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 of 120 mV to 220 mV. It is possible to increase discrimination between different nucleotides by a pore by using an increased applied potential.

The methods are typically carried out in the presence of any charge carriers, such as metal salts, for example alkali metal salt, halide salts, for example chloride salts, such as alkali metal chloride salt. Charge carriers may include ionic liquids or organic salts, for example tetramethyl ammonium chloride, trimethylphenyl ammonium chloride, phenyltrimethyl ammonium chloride, or 1-ethyl-3-methyl imidazolium chloride. In the exemplary apparatus discussed above, the salt is present in the aqueous solution in the chamber. Potassium chloride (KCl), sodium chloride (NaCl), caesium chloride (CsCl) or a mixture of potassium ferrocyanide and potassium ferricyanide is typically used. KCl, NaCl and a mixture of potassium ferrocyanide and potassium ferricyanide are preferred. The 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 and TraI helicases surprisingly work under high salt concentrations. The method is preferably carried out using a salt concentration of at least 0.3 M, such as at least 0.4 M, at least 0.5 M, at least 0.6 M, at least 0.8 M, at least 1.0 M, at least 1.5 M, at least 2.0 M, at least 2.5 M or at least 3.0 M. High salt concentrations provide a high signal to noise ratio and allow for currents indicative of the presence of a nucleotide to be identified against the background of normal current fluctuations.

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

The methods may be carried out at from 0 °C to 100 °C, from 15 °C to 95 °C, from 16 °C to 90 °C, from 17 °C to 85 °C, from 18 °C to 80 °C, 19 °C to 70 °C, or from 20 °C to 60 °C. The methods are typically carried out at room temperature. The methods are optionally carried out at a temperature that supports enzyme function, such as about 37 °C.

The method may be carried out in the presence of free nucleotides or free nucleotide analogues and/or an enzyme cofactor that facilitates the action of the construct. The method may also be carried out in the absence of free nucleotides or free nucleotide analogues and in the absence of an enzyme 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 monophosphate (AMP), adenosine diphosphate (ADP), adenosine triphosphate (ATP), guanosine monophosphate (GMP), guanosine diphosphate (GDP), guanosine triphosphate (GTP), thymidine monophosphate (TMP), thymidine diphosphate (TDP), thymidine triphosphate (TTP), uridine monophosphate (UMP), uridine diphosphate (UDP), uridine triphosphate (UTP), cytidine monophosphate (CMP), cytidine diphosphate (CDP), cytidine triphosphate (CTP), cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), deoxyadenosine monophosphate (dAMP), deoxyadenosine diphosphate (dADP), deoxyadenosine triphosphate (dATP), deoxyguanosine monophosphate (dGMP), deoxyguanosine diphosphate (dGDP), deoxyguanosine triphosphate (dGTP), deoxythymidine monophosphate (dTMP), deoxythymidine diphosphate (dTDP), deoxythymidine triphosphate (dTTP), deoxyuridine monophosphate (dUMP), deoxyuridine diphosphate (dUDP), deoxyuridine triphosphate (dUTP), deoxycytidine monophosphate (dCMP), deoxycytidine diphosphate (dCDP) and deoxycytidine triphosphate (dCTP). The free nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP or dCMP. The free nucleotides are preferably adenosine triphosphate (ATP). The enzyme cofactor is a factor that allows the construct to function. The enzyme cofactor is preferably a divalent metal cation. The divalent metal cation is preferably Mg^{2+}, Mn^{2+}, Ca^{2+} or Co^{2+}. The enzyme cofactor is most preferably Mg^{2+}.

The target polynucleotide may be contacted with the construct and the pore in any order. In is preferred that, when the target polynucleotide is contacted with the construct and the pore,

the target polynucleotide firstly forms a complex with the construct. When the voltage is applied across the pore, the target polynucleotide/construct complex then forms a complex with the pore and controls the movement of the polynucleotide through the pore.

As discussed above, helicases may work in two modes with respect to the pore. The constructs described herein comprising such helicases can also work in two modes. First, the method is preferably carried out using the construct such that it moves the target sequence through the pore with the field resulting from the applied voltage. In this mode the 5' end of the DNA is first captured in the pore, and the construct moves the DNA into the pore such that the target sequence is passed through the pore with the field until it finally translocates through to the trans side of the bilayer. Alternatively, the method is preferably carried out such that the construct moves the target sequence through the pore against the field resulting from the applied voltage. In this mode the 3' end of the DNA is first captured in the pore, and the construct moves the DNA through the pore such that the target sequence is pulled out of the pore against the applied field until finally ejected back to the cis side of the bilayer.

Other methods

The invention also provides a method of forming a sensor for characterising a target polynucleotide. The method comprises forming a complex between a pore and a construct described herein. The complex may be formed by contacting the pore and the construct in the presence of the target polynucleotide and then applying a potential across the pore. The applied potential may be a chemical potential or a voltage potential as described above. Alternatively, the complex may be formed by covalently attaching the pore to the construct. Methods for covalent attachment are known in the art and disclosed, for example, in International Application Nos. PCT/GB09/001679 (published as WO 2010/004265) and PCT/GB10/000133 (published as WO 2010/086603). The complex is a sensor for characterising the target polynucleotide. The method preferably comprises forming a complex between a pore derived from Msp and a construct described herein. Any of the embodiments discussed above with reference to the methods of the invention equally apply to this method. The invention also provides a sensor produced using the method of the invention.

Kits

The present invention also provides a kit for characterising a target polynucleotide. The kit comprises (a) a pore and (b) a construct described herein. Any of the embodiments discussed above with reference to the method of the invention equally apply to the kits.

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 nucleotides.

Apparatus

The invention also provides an apparatus for characterising a target polynucleotide. The apparatus comprises a plurality of pores and a plurality of constructs described herein. The apparatus preferably further comprises instructions for carrying out the method of the invention. The apparatus may be any conventional apparatus for polynucleotide analysis, such as an array or a chip. Any of the embodiments discussed above with reference to the methods of the invention are equally applicable to the apparatus of the invention.

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

The apparatus preferably comprises:

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

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

The apparatus preferably comprises:

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

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

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

one or more containers for receiving respective samples, the fluidics system being configured to supply the samples selectively from the one or more containers to the sensor device. The apparatus may be any of those described in International Application No. No. PCT/GB08/004127 (published as WO 2009/077734), PCT/GB10/000789 (published as WO

2010/122293), International Application No. PCT/GB10/002206 (not yet published) or International Application No. PCT/US99/25679 (published as WO 00/28312).

Methods of producing constructs of the invention

The invention also provides a method of producing a construct of the invention. In one embodiment, the method comprises attaching, preferably covalently attaching, two or more helicases. In another embodiment, the method comprises attaching, preferably covalently attaching, a helicase to an amino acid sequence comprising SEQ ID NO: 94 or a variant thereof having at least 80% homology to SEQ ID NO: 94 based on amino acid identity over the entire sequence of SEQ ID NO: 94 and thereby producing the construct. Any of the helicases discussed above can be used in the methods. The site of and method of attachment are selected as discussed above.

The method may further comprise determining whether or not the construct is capable of controlling the movement of a polynucleotide. Assays for doing this are described above. If the movement of a polynucleotide can be controlled, the helicases have been attached correctly and a construct of the invention has been produced. If the movement of a polynucleotide cannot be controlled, a construct of the invention has not been produced.

The following Example illustrates the invention.

Example 1

In this Example and all of the following Examples, bismaleimide-functionalized PEG linkers are identified with reference to their molecular weight. For instance, "2kDa", "2kDA linker" or "2kDA PEG linker" refers to a bismaleimide-functionalized PEG linker having a molecular weight of 2kDa.

This Example describes the method of synthesising the Hel308 Mbu(R687A/A700C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutations R687A/A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker) and the Hel308 Mbu(R681A/R687A/A700C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutations R681A/R687A/A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker). In the case of these dimers the A700C is the added mutation that serves as the connection for the linker in the dimer proteins (Mbu/R687A **A700C** – 2kDa PEG linker – **A700C** Mbu/R687A and Mbu/R681A/R687A **A700C** – 2kDa PEG linker – **A700C** Mbu/R681A/R687A). 5 cysteines are naturally present in Hel308Mbu (SEQ ID NO: 10)

but these are not very reactive and, therefore, the reaction is almost completely centered on A700C.

DTT was added to Hel308Mbu R687A/A700C (2 mg/mL) (SEQ ID NO: 10 with the mutations R687A/A700C) to 5 mM and placed on a rotator for 30 min. The reduced protein was buffer exchanged into 100 mM potassium phosphate, 500 mM NaCl, 5 mM EDTA, 0.1% tween20™, pH7.2. 30 mer ssDNA (SEQ ID NO: 68) was added to the helicase (10 fold excess) to protect the internal cysteines and increase the likelihood of the protein remaining stable. The protein/DNA solution was diluted to 1.5 mg/mL with buffer, placed under an atmosphere of nitrogen and incubated at room temperature for 30 min. 0.016 mM bismaleimide-PEG (2kDa) was added and the reaction allowed to proceed at room temperature under an atmosphere of nitrogen for 2 h. 10 mM DTT was added to quench the reaction and break up any disulfide bridged species. The Hel308 Mbu(R687A/A700C)-2kDa mutant dimer was purified immediately using an initial Strep-tactin step to remove DNA and reagents, followed by an anion-exchange chromatography step to separate the dimer from all other species present in solution.

An AKTA purifier machine was used for the purifications. Strep purification was performed on a 1 mL StrepTactin Sepharose™ High Performance column. The protein solution was buffer exchanged into binding buffer (50 mM Tris, 500 mM NaCl, 2 mM EDTA, 2 mM DTT, 0.05% tween20, pH 8.0) before being loaded onto the column. After an initial wash step to remove all unbound material, the protein was washed with the same buffer containing 2 M salt to dissociate the DNA from the protein before eluting it with 10 mM desthiobiotin in buffer. The eluted protein was buffer exchanged into LOW buffer (50 mM Tris, 80 mM NaCl, 2 mM DTT, 0.05% tween20, pH 8.0) to prepare it for the separation between the dimer and all other species present in solution. The ion exchange step was performed on a GE Mini Q PC 3.2/3 column with a flow rate of 0.4 mL/min, on a gradient between LOW and HIGH buffer (50 mM Tris, 2 M NaCl, 2 mM DTT, 0.05% tween20, pH 8.0). The start, middle and end of the eluted dimer peak were pooled separately and given separate ID numbers. All three were activity assayed before the middle peak was used for tests in electrophysiology. Fig. 1 shows a gel of the Hel308 Mbu(R687A/A700C) monomer (SEQ ID NO: 10 with the mutations R687A/A700C, lane 2) and the Hel308 Mbu(R687A/A700C)-2kDa dimer (lane 4). The above procedure can be used in order to form the Hel308 Mbu(R681A/R687A/A700C)-2kDa dimer (Fig. 1, Lane 5) from the Hel308 Mbu(R681A/R687A/A700C) monomer (SEQ ID NO: 10 with the mutations R681A/R687A/A700C, Fig. 1, Lane 3).

Example 2

This Example describes the method of synthesising the Hel308 Mhu multimer (multiple units of SEQ ID NO: 19). There are no added mutations on Hel308Mhu (SEQ ID NO: 19), so the cysteines contained in the WT are used as the sites for linkage.

DTT was added to Hel308 Mhu (1.83mg/mL) (SEQ ID NO: 19) to 10 mM and placed on a rotator for 30 min. The reduced protein was buffer exchanged into 100 mM potassium phosphate, 500 mM NaCl, 5 mM EDTA, 0.1% tween20, pH7.2. 30mer ssDNA (SEQ ID NO: 68) was added to the helicase (6 fold excess) to protect the internal cysteines and increase the likelihood of the protein remaining stable. The protein/DNA solution was diluted to 1.6 mg/mL with buffer, placed under an atmosphere of nitrogen and incubated at room temperature for 30 min. 0.0095 mM bismaleimide-PEG (2kDa) was added and the reaction allowed to proceed at room temperature under an atmosphere of nitrogen for 2 h. 10 mM DTT was added to quench the reaction and break up any disulfide bridged species. The Hel308 Mhu multimer was purified immediately using an initial Strep-tactin step to remove DNA and reagents, followed by anion-exchange and gel filtration chromatography steps to remove the monomer from all other species present in solution.

An AKTA purifier machine was used for the purifications. Strep purification was performed on a 5 mL StrepTactin Sepharose High Performance column. The protein solution was buffer exchanged into binding buffer (50 mM Tris, 500 mM NaCl, 0.1% tween20, pH 8.0) before being loaded onto the column. After an initial wash step, to remove all unbound material, the protein was washed with the same buffer containing 2 M NaCl to dissociate the DNA from the protein, before eluting it with 10 mM desthiobiotin in buffer. The eluted protein was buffer exchanged into LOW buffer (50 mM Tris, 80 mM NaCl, 2 mM DTT, 0.05% tween20, pH 8.0) to prepare it for the separation between the dimer and all other species present in solution. The ion exchange step was performed on a GE Mono Q 5/50 GL column with a flow rate of 1 mL/min, on a gradient between LOW and HIGH buffer (50 mM Tris, 2 M NaCl, 2 mM DTT, 0.05% tween20, pH 8.0). The elution peak was further purified by anion exchange on the same column 3 times. The collected elution peak from the last purification step was collected, concentrated to 0.25 mL, buffer, exchanged into 50 mM Tris, 500 mM NaCl, 2 mM DTT, pH 8.0 and purified by gel filtration on a Superdex™ 10/300 GL column. The final protein had an approximate 1:1:1:1 ratio of monomer:dimer:trimer:multimers (ONT Ref - ONLP4454). Fig. 2 shows a gel of the Hel308 Mhu multimer sample (lane 2).

Example 3

This Example describes the method of synthesising the Hel308 Tga(R657A/N674C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 16 with the mutations R657A/N674C, with one monomer unit being linked to the other via position 674 of each monomer unit using a 2 kDa PEG linker). In the case of this dimer the N674C is the added mutation that serves as the connection for the linker in the dimer protein (Tga **N674C** – 2kDa PEG linker – **N674C** Tga).

DTT was added Hel308 Tga(R657A/N674C) (2mg/mL) (where each monomer unit comprises SEQ ID NO: 16 with the mutations R657A/N674C, with one monomer unit being linked to the other via position 674 of each monomer unit using a 2 kDa PEG linker) to 10 mM and placed on a rotator for 30 min. The reduced protein was buffer exchanged into 100 mM potassium phosphate, 500 mM NaCl, 5 mM EDTA, 0.1% tween20, pH7.2. 30mer ssDNA (SEQ ID NO: 68) was added to the helicase to increase the likelihood of the protein remaining stable. The protein/DNA solution was diluted to 1.6 mg/mL with buffer, placed under an atmosphere of nitrogen and incubated at room temperature for 30 min. 0.038 mM bismaleimide-PEG (2kDa) was added and the reaction allowed to proceed at 23 °C under an atmosphere of nitrogen for 2.5 h. 10 mM DTT was added to quench the reaction and break up any disulfide bridged species. The Hel308 Tga(R657A/N674C) mutant dimer was purified immediately using an initial Strep-tactin step to remove DNA and reagents, followed by an anion-exchange chromatography step to separate dimer from all other species present in solution.

An AKTA purifier machine was used for the purifications. Strep purification was performed on a 5 mL StrepTactin Sepharose High Performance column. The protein solution was buffer exchanged into binding buffer (50 mM Tris, 200 mM NaCl, 1 mM MgCl2, 2 mM DTT, 0.05% tween20, pH 8.0) before being loaded onto the column. After an initial wash step, to remove all unbound material, the protein was washed with the same buffer containing 4 mM ATP, to remove DNA from the protein, before eluting it with 10 mM desthiobiotin in buffer. The eluted protein was concentrated to 0.25 mL, buffer exchanged into 50 mM Tris, 250 mM NaCl, 2 mM DTT, 1mM MgCl2, 0.05% tween20, pH 8.0 and purified by gel filtration on a Superdex™ 10/300 GL column. Fig. 3 shows a gel of the Hel308 Tga(R657A/N674C)-2kDa dimer and monomer at various stages during formation and purification (Lane 1 = protein ladder, Lane 2 = Hel308 Tga(R657A/N674C)-2kDa dimer after heat treatment, Lane 3 = Hel308 Tga(R657A/N674C)-2kDa dimer (where each monomer unit comprises SEQ ID NO: 16 with the mutations R657A/N674C, with one monomer unit being linked to the other via position 674 of each monomer unit using a 2 kDa PEG linker), Lane 4 = strep elution, Lane 5 = initial reaction

mixture and Lane 6 = Hel308 Tga(R657A/N674C) monomer (SEQ ID NO: 16 with the mutations R657A/N674C)).

Example 4

This Example compares the DNA binding ability of various Hel308 Mbu helicase constructs with that of the Hel308 Mbu monomer (SEQ ID NO: 10) using a fluorescence based assay.

A custom fluorescent substrate was used to assay the ability of various helicases to bind to single-stranded DNA. The 88 nt single-stranded DNA substrate (1 nM final, SEQ ID NO: 69) has a carboxyfluorescein (FAM) base at its 5' end. As helicase binds to the oligonuclotide in a buffered solution (400 mM NaCl, 10 mM Hepes, pH8.0, 1 mM $MgCl_2$), the fluorescence anisotropy (a property relating to the rate of free rotation of the oligonucleotide in solution) increases. The lower the amount of helicase needed to affect an increase in anisotropy, the tighter the binding affinity between the DNA and helicase (Fig. 4). Fig.'s 5-8 show the change in anisotropy of the DNA oligonucleotide (SEQ ID NO: 69, which has a carboxyfluorescein base at its 5' end) with increasing amounts of various Hel308 (Mbu) constructs. All of the constructs tested show an increase in anisotropy at a lower concentration than the monomer. The constructs tested were:

1. Hel308 Mbu A700C 2kDa dimer (helicase dimer where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker),
2. Hel308 Mbu-GTGSGA-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a helix-hairpin-helix (HhH2) domain (SEQ ID NO: 75)),
3. Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a (HhH)2-(HhH)2 domain (SEQ ID NO: 76) where HhH is a helix-hairpin-helix domain),
4. Hel308 Mbu-GTGSGA-UL42HV1-I320Del (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to processivity factor UL42HV1-I320Del (SEQ ID NO: 63)),
5. Hel308 Mbu-GTGSGA-gp32RB69CD (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to SSB gp32RB69CD (SEQ ID NO: 64)),

6. Hel308 Mbu-GTGSGA-gp2.5T7-R211Del (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to SSB gp2.5T7-R211Del (SEQ ID NO: 65)) and
7. (gp32RB69CD)-Hel308 Mbu) (where the SSB gp32RB69CD (SEQ ID NO: 64) is attached by the linker sequence GTGSGT to the helicase monomer unit (SEQ ID NO: 10)).

Fig. 9 shows the relative equilibrium dissociation constants (K_d) (with respect to Hel308 Mbu monomer (SEQ ID NO: 10)) for various Hel308 (Mbu) constructs obtained through fitting *two phase dissociation* binding curves through the data shown in Fig.'s 5-8 using Graphpad Prism software. All of the helicase constructs with additional binding domains attached show a lower equilibrium dissociation constant than the Hel308 Mbu monomer (SEQ ID NO: 10) alone. Therefore, the Hel308 Mbu helicase with additional binding constructs all show stronger binding to DNA than the Hel308 Mbu monomer.

Example 5

This Example compares the ability of a Hel308 Mbu monomer (SEQ ID NO: 10), to control the movement of intact DNA strands (400 mer) through a nanopore, to that of the Hel308 Mbu A700C 2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker). The general method for controlled DNA translocation by the monomer is shown in Fig. 10 and by the dimer in Fig. 11.

Materials and Methods

400 mer DNA Sequences:

Primers were designed to amplify a ~400 bp fragment of PhiX174. Each of the 5'-ends of these primers included a 50 nucleotide non-complimentary region, either a homopolymeric stretch or repeating units of 10 nucleotide homopolymeric sections. These serve as identifiers for controlled translocation of the strand through a nanopore, as well as determining the directionality of translocation. In addition, the 5'-end of the forward primer was "capped" to include four 2'-O-methyl-uracil (mU) nucleotides and the 5'-end of the reverse primer was chemically phosphorylated. These primer modifications then allow for the controlled digestion of predominantly only the antisense strand, using lambda exonuclease. The mU capping protects the sense strand from nuclease digestion whilst the PO4 at the 5' of the antisense strand

promotes it. Therefore after incubation with lambda exonuclease only the sense strand of the duplex remains intact, now as single stranded DNA (ssDNA). The generated ssDNA was then PAGE purified as previously described.

The DNA substrate design used in all the experiments described here is shown in Fig. 12. The DNA substrate consists of a 400base section of ssDNA from PhiX, with, at the 5' end of the sequence, four 2'-O-methyl uracil bases attached to a 50T 5'-leader to aid capture by the nanopore (SEQ ID NO: 70, at the 5' end of SEQ ID NO: 70 there are four 2'-O-methyl uracil bases attached to a 50T leader sequence to aid capture by the nanopore). Annealed to this strand just after the 50T leader is a primer containing a 3' cholesterol TEG (SEQ ID NO: 71) to enrich the DNA on the surface of the bilayer, and thus improve capture efficiency.

Buffered solution: 1 M KCl, 10 mM Hepes pH8.0, 1 mM ATP, 10 mM $MgCl_2$, 1 mM DTT
Nanopore: E.coli MS(B1-L88N)8 MspA (SEQ ID NO: 2, with the mutation L88N)
Monomer Enzyme: Hel308 Mbu (SEQ ID NO: 10) added at 100 nM final.
Dimer Enzyme: Hel308 Mbu A700C 2kDa homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker) added at 10 nM final.

Electrical measurements were acquired from single MspA nanopores inserted in 1,2-diphytanoyl-glycero-3-phosphocholine lipid (Avanti Polar Lipids) bilayers. Bilayers were formed across ~100 um diameter apertures in 20 um thick PTFE films (in custom Delrin chambers) via the Montal-Mueller technique, separating two 1 mL buffered solutions. All experiments were carried out in the stated buffered solution. Single-channel currents were measured on Axopatch 200B amplifiers (Molecular Devices) equipped with 1440A digitizers. Ag/AgCl electrodes were connected to the buffered solutions so that the *cis* compartment (to which both nanopore and enzyme/DNA are added) is connected to the ground of the Axopatch headstage, and the *trans* compartment is connected to the active electrode of the headstage.

After achieving a single pore in the bilayer DTT (1 mM) and MgCl2 were added to the cis chamber and mixed well. DNA construct and helicase were then added to 100 uL of buffer and pre-incubated for 5 mins (DNA = 1.5 nM (SEQ ID NO's: 70 and 71 (which has a 3' cholesterol TEG)), monomer enzyme = 1 uM or dimer enzyme = 0.1 uM). This pre-incubation mix was added to 900 uL of buffer in the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore (to give final concentrations of DNA = 0.15 nM, monomer enzyme = 0.1 uM or dimer enzyme = 0.01 uM).

Helicase ATPase activity was initiated as required by the addition of dNTP (1 mM final ATP) to the *cis* compartment. Experiments were carried out at a constant potential of +120 mV.

Results and Discussion

The addition of helicase monomer-DNA substrate to MspA nanopores (as shown in Fig. 10) produces characteristic current blocks as shown in Fig. 13. For a given substrate, we observe a characteristic pattern of current transitions for each helicase controlled DNA movement that reflects the DNA sequence. DNA without a helicase bound to it interacts transiently with the nanopore producing short-lived blocks in current (<< 1 seconds). DNA with Hel308 Mbu monomer (SEQ ID NO: 10) bound and active (i.e. moving along the DNA strand under ATPase action) produces long characteristic blocks levels with stepwise changes in current as shown in Fig. 13. Different DNA motifs in the nanopore give rise to unique current block levels.

The addition of helicase dimer-DNA substrate to MspA nanopores (as shown in Fig. 11) produces characteristic current blocks as shown in Fig. 14. DNA with Hel308 Mbu A700C 2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker) bound and active (i.e. moving along the DNA strand under ATPase action) produces long characteristic blocks levels with stepwise changes in current as shown in Fig. 14. It is also possible to employ a ten-times lower concentration of dimer and still observe these characteristic current blocks.

The current blocks produced when the Hel308 Mbu monomer (SEQ ID NO: 10) controls translocation of the 400 mer DNA strand through the pore are similar to those produced by the Hel308 Mbu A700C 2kDa dimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A700C, with one monomer unit being linked to the other via position 700 of each monomer unit using a 2 kDa PEG linker). However, when you compare the overall length of the helicase controlled strand movements for two experiments (run under identical conditions with either monomer Hel308 Mbu or Hel308 Mbu A700C 2kDa dimer) then the helicase controlled DNA movements with the dimer are typically much longer (Fig. 16) than those observed by the monomer (Fig. 15). This indicates enzyme rebinding and therefore reduced enzyme dissociation. In addition, for the monomer run 37% of the helicase controlled DNA movements measured reached the polyT at the end of the DNA strand, whereas, for the dimer run 47% of the helicase controlled DNA movements measured reached the polyT, showing the reduced dissociation and improved processivity of the dimer. Fig. 8 and 9 each show six examples of the position in the known DNA sequence (vertical axis) of the state-fitted data for the strand movements as a

function of the state index (horizontal axis), when DNA motion is controlled by either the Hel308 Mbu monomer (Fig. 17) or the Hel308 Mbu A700C 2kDa dimer (Fig. 18). The monomer data shows processive linear movement through the sequence, with periodic dislocations back to previous parts of the sequence (highlighted with a dashed circle), which are the result of enzyme dissociation and the DNA slipping back under the applied field until it encounters a trailing enzyme. Many of the helicase controlled DNA movements do not make it to the end of the sequence due to enzyme dissociation. Whereas for the dimer data the enzymes proceed to control the movement of the DNA for much longer, and after dissociation the enzyme re-binds to the DNA.

Example 6

This Example shows that two different Hel308 Mbu homodimers, that are connected via different amino-acid positions and with different length linkers (2kb and 3.4kb linkers) in comparison to the Hel308 Mbu A700C 2kDa homodimer discussed in Example 5, were also capable of controlling the movement of intact DNA strands (400 mer) through a nanopore. The general method for controlled DNA translocation by the dimer is shown in Fig. 11.

Buffered solution: 400 mM KCl, 10 mM Hepes pH8.0, 1 mM ATP, 1 mM $MgCl_2$, 1 mM DTT
Nanopore: E.coli MS(B1-L88N)8 MspA (SEQ ID NO: 2 with the mutation L88N)
Dimer Enzymes: Hel308 Mbu Q442C 2 kDa linker homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation Q442C, with one monomer unit being linked to the other via position 442 of each monomer unit using a 2 kDa PEG linker) added at approximately 1 nM final and Hel308 Mbu Q442C 3.4 kDa linker homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation Q442C, with one monomer unit being linked to the other via position 442 of each monomer unit using a 3.4 kDa PEG linker) added at approximately 1 nM final.

Electrical experiments were set up as described in Example 5 in order to achieve a single pore inserted into a lipid bilayer. After achieving a single pore in the bilayer, DTT (1 mM) and $MgCl_2$ (1 mM) were added to the *cis* chamber and mixed well. DNA polynucleotide SEQ ID NO's: 70 and 71 (which has a 3' cholesterol TEG) (DNA = 0.15 nM), ATP (1 mM) and either Hel308 Mbu Q442C 2 kDa linker homodimer or Hel308 Mbu Q442C 3.4 kDa linker homodimer were then added to the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore. Experiments were carried out at a constant potential of +180 mV.

Results and Discussion

The use of the Hel308 Mbu Q442C 2 kDa linker homodimer (where each monomer unit comprises SEQ ID NO: 10 with the mutation Q442C, with one monomer unit being linked to the other via position 442 of each monomer unit using a 2 kDa PEG linker) helicase, to control DNA substrate translocation through MspA nanopores, produces characteristic current blocks as shown in Fig. 19A. For a given substrate, we observe a characteristic pattern of current transitions for each helicase controlled DNA movement that reflects the DNA sequence. DNA without a helicase bound to it interacts transiently with the nanopore producing short-lived blocks in current (<< 1 seconds). DNA with helicase dimer bound and active (i.e. moving along the DNA strand under ATPase action) produces long characteristic blocks levels with stepwise changes in current as shown in Fig. 19A. Different DNA motifs in the nanopore give rise to unique current block levels. The use of the Hel308 Mbu Q442C 3.4 kDa linker homodimer helicase (where each monomer unit comprises SEQ ID NO: 10 with the mutation Q442C, with one monomer unit being linked to the other via position 442 of each monomer unit using a 3.4 kDa PEG linker), to control DNA movement through MspA nanopores, also produces characteristic current blocks as shown in Fig. 19B. This illustrates that it is possible to attach two helicases together at different positions using two different linker lengths and still retain enzyme activity.

Example 7

This Example shows helicase-controlled DNA movement through nanopores using helicases with an additional binding domain fused to the C-terminus. The two Examples shown are Hel308 Mbu with an additional 5th domain from Hel308 Hla (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hla (SEQ ID NO: 66)) or Hel308 Hvo (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hvo (SEQ ID NO: 67)) genetically fused to the C-terminus.

Buffered solution: 400 mM NaCl, 10 mM Hepes pH8.0, 1 mM ATP, 1 mM $MgCl_2$, 1 mM DTT
Nanopore: E.coli MS(B1-L88N)8 MspA (SEQ ID NO: 2 with the mutation L88N)
Dimer Enzymes: Hel308 Mbu + 5th domain Hel308 Hla (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hla (SEQ ID NO: 66)) added at 100 nM final and Hel308 Mbu + 5th domain Hel308 Hvo (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hvo (SEQ ID NO: 67)) added at 100 nM final.

Electrical experiments were set up as described in Example 5 in order to achieve a single pore inserted into a lipid bilayer. After achieving a single pore in the bilayer DTT (1 mM) and$MgCl_2$ (1 mM) were added to the *cis* chamber and mixed well. A control recording at +140 mV was run for 5 minutes. DNA polynucleotide SEQ ID NO's: 70 and 71 (which has a 3' cholesterol TEG) (DNA = 0.6 nM) and either Hel308 Mbu + 5th domain Hel308 Hla (100 nM, where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hla (SEQ ID NO: 66)) or Hel308 Mbu + 5th domain Hel308 Hvo (100 nM, where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hvo (SEQ ID NO: 67)) were then added to the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore. A second control recording at +140 mV was run for 10 minutes. Finally helicase ATPase activity was initiated as required by the addition of ATP (1 mM) to the *cis* compartment. Experiments were carried out at a constant potential of +140 mV.

Results and Discussion

The use of the Hel308 Mbu + 5th domain Hel308 Hla helicase (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hla (SEQ ID NO: 66)), to control DNA substrate translocation through MspA nanopores, produces characteristic current blocks as shown in Fig. 20A. For a given substrate, we observe a characteristic pattern of current transitions for each helicase controlled DNA movement that reflects the DNA sequence. DNA without a helicase bound to it interacts transiently with the nanopore producing short-lived blocks in current (<< 1 seconds). DNA with Hel308 Mbu + 5th domain Hel308 Hla helicase bound and active (i.e. moving along the DNA strand under ATPase action) produces long characteristic blocks levels with stepwise changes in current as shown in Fig. 20A. Different DNA motifs in the nanopore give rise to unique current block levels. The use of the Hel308 Mbu + 5th domain Hel308 Hvo helicase (where a helicase monomer unit (SEQ ID NO: 10) is attached to the 5th domain of Hel308 Hvo (SEQ ID NO: 67)), to control DNA movement through MspA nanopores, also produces characteristic current blocks as shown in Fig. 20B. This illustrates that it is possible to attach an additional binding domain of a helicase to another helicase and still retain enzyme activity.

Example 8

This Example shows helicase-controlled DNA movement through nanopores using helicases with additional Helix-hairpin-Helix (HhH) domains attached. The two examples shown are Hel308 Mbu with either two or four helix-hairpin-helix domains attached at the C-terminus.

Materials and Methods

The DNA was formed by ligating a 50-polyT 5' leader to a ~900base fragment of PhiX dsDNA. The leader also contains a complementary section to which SEQ ID NO: 74 (which at the 3' end of the sequence has six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG) was hybridized to allow the DNA to be tethered to the bilayer. Finally the 3' end of the PhiX dsDNA was digested with AatII digestion enzyme to yield a 4nt 3'-overhang of ACGT (see Fig. 21 for diagram of the DNA substrate design).

Buffered solution: 400 mM NaCl, 100 mM Hepes pH8.0, 10 mM potassium ferrocyanide, 10 mM potassium ferricyanide, 1 mM ATP, 1 mM $MgCl_2$,
Nanopore: E.coli MS(B1-G75S-G77S-L88N-Q126R)8 MspA (SEQ ID NO:2 with the mutations G75S/G77S/L88N/Q126R)
Dimer Enzymes: Hel308 Mbu-GTGSGA-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a HhH2 domain (SEQ ID NO: 75)) added at 100 nM final and Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a (HhH)2-(HhH)2 domain (SEQ ID NO:76)) added at 100 nM final.

Electrical experiments were set up as described in Example 5 in order to achieve a single pore inserted into a lipid bilayer, however, platinum electrodes were employed instead of Ag/AgCl electrodes. After achieving a single pore in the bilayer, $MgCl_2$ (1 mM) and ATP (1 mM) were added to the chamber. A control recording at +140 mV was run for 5 minutes. DNA polynucleotide SEQ ID NO's 72, 73 and 74 (which at the 3' end of the sequence has six iSp18 spacers attached to two thymine residues and a 3' cholesterol TEG) (DNA = 0.1 nM) were added to the *cis* compartment of the electrophysiology chamber and DNA translocation events were observed. Finally, either Hel308 Mbu-GTGSGA-(HhH)2 (100 nM, where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a HhH2 domain (SEQ ID NO: 75)) or Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 (100 nM, where a helicase monomer unit

(SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a (HhH)2-(HhH)2 domain (SEQ ID NO: 76)) was then added to the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the MspA nanopore. Experiments were carried out at a constant potential of +140 mV.

Results and Discussion

The use of Hel308 Mbu-GTGSGA-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a HhH2 domain (SEQ ID NO: 75)), to control DNA substrate translocation through MspA nanopores, produces characteristic current blocks as shown in Fig. 22A. For a given substrate, we observe a characteristic pattern of current transitions for each helicase controlled DNA movement that reflects the DNA sequence. DNA without helicase bound interacts transiently with the nanopore producing short-lived blocks in current (<< 1 seconds). DNA with Hel308 Mbu-GTGSGA-(HhH)2 bound and active (i.e. moving along the DNA strand under ATPase action) produces long characteristic blocks levels with stepwise changes in current as shown in Fig. 22A. Different DNA motifs in the nanopore give rise to unique current block levels. The use of the Hel308 Mbu-GTGSGA-(HhH)2-(HhH)2 (where a helicase monomer unit (SEQ ID NO: 10) is attached by the linker sequence GTGSGA to a (HhH)2-(HhH)2 domain (SEQ ID NO: 76)), to control DNA movement through MspA nanopores, also produces characteristic current blocks as shown in Fig. 22B. This illustrates that it is possible to attach additional helix-hairpin-helix domains to a helicase and still retain enzyme activity.

Example 9

This Example compares the ability of a TrwC Cba monomer (SEQ ID NO: 87), to control the movement of intact DNA strands (attached to the 5' end of SEQ ID NO: 88 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: 88 is a further four iSpC3 spacers which are attached to the 5' end of SEQ ID NO. 104) through a nanopore, to that of the TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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 dimer results 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: 88 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: 88 is a further four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 104) and the enzyme (either a TrwC Cba monomer (1 nM, SEQ ID NO: 87) or TrwC Cba Q276C-3.4kDa dimer (0.3 nM, where each monomer unit comprises SEQ ID NO: 87 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/D90N/D91N/D93N/D118R/Q126R/D134R/E139K)8 MspA (SEQ ID NO: 2 with the mutations G75S/G77S/L88N/D90N/D91N/D93N/D118R/Q126R/D134R/E139K) inserted in block copolymer in buffer (625 mM KCl, 100 mM Hepes, 75 mM Potassium Ferrocyanide (II), 25 mM Potassium ferricyanide (III), pH 8). $MgCl_2$ (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 (attached to the 5' end of SEQ ID NO: 88 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: 88 is a further four iSpC3 spacers which are attached to the 5' end of SEQ ID NO: 104), enzyme pre-mix (either a TrwC Cba monomer (1 nM, SEQ ID NO: 87) or TrwC Cba Q276C-3.4kDa dimer (1 nM, where each monomer unit comprises SEQ ID NO: 87 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: 87) and TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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). 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, where each monomer unit comprises SEQ ID NO: 87 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) 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 10

This Example illustrates the salt tolerance of a TrwC Cba-TopoV Mka (where TrwC Cba is attached by the linker AYDVGA to domains H-L of Topoisomerase V Mka full sequence shown in SEQ ID NO: 90) using a fluorescence assay for testing enzyme activity.

Materials and Methods

A custom fluorescent substrate was used to assay the ability of the TrwC Cba-Topo V Mka (where TrwC Cba is attached by the linker AYDVGA to domains H-L of Topoisomerase V Mka full sequence shown in SEQ ID NO: 90) to displace hybridised dsDNA. As shown in 1) of Fig. 23, the fluorescent substrate strand (50 nM final) has both a 3'and 5' ssDNA overhang, and a 44 base section of hybridised dsDNA. The upper strand, containing the 3' ssDNA overhang, has a carboxyfluorescein base (labelled 5 in SEQ ID NO: 91) at the 5' end, and the hybrised complement has a black-hole quencher (BHQ-1) base (labelled 6 in SEQ ID NO: 92) at the 3' end. When hybridised the fluorescence from the fluorescein is quenched by the local BHQ-1, and the substrate is essentially non-fluorescent. 1 μM of a capture strand (SEQ ID NO: 93) that is part-complementary to the lower strand of the fluorescent substrate is included in the assay. As shown in 2), in the presence of ATP (1 mM) and $MgCl_2$ (1 mM), helicase (100 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 3), once the complementary strand with BHQ-1 is fully displaced the fluorescein on the major strand fluoresces. As shown in 4), the displaced strand preferentially anneals to an excess of capture strand to prevent re-annealing of initial substrate and loss of fluorescence.

Results and Discussion

The graph in Fig. 24 shows the initial rate of activity in buffer solution (10 mM Hepes pH 8.0, 1 mM ATP, 1 mM $MgCl_2$, 50 nM fluorescent substrate DNA (SEQ ID NOs: 91 and 92), 1 μM capture DNA (SEQ ID NO: 93)) for the TrwC Cba monomer (labeled A in Fig. 24; SEQ ID NO: 87) and the TrwC Cba-Topo V Mka (labeled B in Fig. 24; where TrwC Cba is attached by the linker AYDVGA to domains H-L of Topoisomerase V Mka full sequence shown in SEQ ID NO: 90) at 400 mM of NaCl. At the salt concentration investigated the TrwC Cba-Topo V Mka (where TrwC Cba is attached by the linker AYDVGA to domains H-L of Topoisomerase V Mka full sequence shown in SEQ ID NO: 90) exhibited a higher rate of dsDNA turnover than the TrwC Cba monomer (SEQ ID NO: 87) (see Fig. 24).

Example 11

This Example describes the method of synthesising the TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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).

Materials and Methods

DTT was added TrwC Cba Q276C monomer (2mg/mL, where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) to 10 mM and placed on a rotator for 30 min. The reduced protein was buffer exchanged into 100 mM potassium phosphate, 500 mM NaCl, 5 mM EDTA, 0.1% tween20, pH7.2. TrwC Cba Q276C monomer (2mg/mL) (where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) at 2mg/ml was added to bismaleimide-PEG (3.4kDa; 0.038 mM) was added and the reaction allowed to proceed at 23 °C under an atmosphere of nitrogen for 2.5 h. DTT (10 mM) was added to quench the reaction and break up any disulfide bridged species. The TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) was purified immediately using an initial Strep-tactin step to remove reagents, followed by GF-chromatography step to separate dimer from all other species present in solution.

An AKTA purifier machine was used for the purifications. Strep purification was performed on a 5 mL StrepTactin Sepharose High Performance column. The protein solution was buffer exchanged into binding buffer (50 mM Tris, 200 mM NaCl, 1 mM MgCl2, 2 mM

DTT, 0.05% tween20, pH 8.0) before being loaded onto the column. After an initial wash step to remove all unbound material it was eluted with 10 mM desthiobiotin in buffer. The eluted protein was concentrated to 0.25 mL, buffer exchanged into 50 mM Tris, 250 mM NaCl, 2 mM DTT, 1mM MgCl2, 0.05% tween20, pH 8.0 and purified by gel filtration on a Superdex™ 10/300 GL column. Figs. 25 and 26 show gels of the TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) and TrwC Cba Q276C monomer (where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) at various stages during formation and purification (Fig. 25 - Lane M = protein ladder, Lane 1 = E3-Q276C monomer starting material, Lane 2 = reaction mix, Lane 3 = reaction mix. Fig. 26 - Lane M = protein ladder, Lane X = reference lane for TrwC Cba Q276C-3.4kDa dimer, lanes 4-14 contain purified fractions from the elution of TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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 band which corresponds to TrwC Cba Q276C-3.4kDa dimer (where each monomer unit comprises SEQ ID NO: 87 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) is indicated by a grey arrow in both Figs. 25 and 26.

Using an analogous procedure to that described in this example, it was possible to make the following TrwC Cba Q276C-1kDa dimer (where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C, with one monomer unit being linked to the other via position 276 of each monomer unit using a 1kDa PEG linker).

Example 12

This Example describes the method of synthesising the Hel308Mbu-A577Faz-PEG4 linker-TrwC Cba Q276C dimer (where the Hel308Mbu monomer unit comprises SEQ ID NO: 10 with the amino acid at position 577 mutated to a 4-azido-L-phenylalanine (Faz), which is attached by PEG4 linker to TrwC Cba monomer unit SEQ ID NO: 87 with the mutation Q276C, where the linker is attached to each monomer at position 577 on Hel 308 Mbu monomer and position 276 on TrwC Cba). A cartoon representation of the method to attach the two monomer units is shown in Fig. 27.

Materials and Methods

DTT (10mM) was added to TrwC Cba Q276C monomer (0.9 mg/mL, 1 mL, where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) and the sample left at room temperature for 1 hour. Buffer exchange of the TrwC Cba Q276C monomer (0.9 mg/mL, where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) sample was performed twice using 40K Zeba columns in 100mM Tris 400mM NaCl pH 7.5. Mal-PEG4-DBCO (500μM) was added to the buffer exchanged TrwC Cba Q276C monomer (0.9 mg/mL, where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) and the sample left at room temperature for 3 hours. The TrwC Cba Q276C monomer (0.9 mg/mL, where each monomer unit comprises SEQ ID NO: 87 with the mutation Q276C) Mal-PEG4-DBCO was then buffer exchanged twice, into 100mM Tris 400mM NaCl pH 7.5, using 40K Zeba columns. Hel308 Mbu-A577Faz (1.1 mg/mL, 1 mL, where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz) was buffer exchanged into 100mM Tris 400mM NaCl pH 7.5 using a 40K Zeba column. The two buffer exchanged proteins were mixed together and left at room temperature for 3 hours. Finally, the following samples were then run on a 4-12% gel (shown in Fig. 28) lane a) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C), lane b) Hel308 Mbu-A577Faz (where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz), lane c) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Hel308 Mbu-A577Faz (where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz), lane d) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + 5kDa PEG, lane e) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + 5kDa PEG with an azide attached, lane f) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Azide Alexa Fluor® 555 (Life Technologies, used to check for non-specific interactions between the fluorophore and TrwC Cba-Q276C monomer), lane g) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO, lane h) TrwC Cba-Q276C-PEG4-DBCO (SEQ ID NO: 87 with the mutation Q276C which is attached to the PEG4-DBCO linker) + Hel308 Mbu (SEQ ID NO: 10), lane i) Hel308Mbu-A577Faz-PEG4 linker-TrwC Cba Q276C dimer (where the Hel308Mbu monomer unit comprises SEQ ID NO: 10 with the amino acid at position 577 mutated to a 4-azido-L-phenylalanine (Faz), which is attached by PEG4 linker to TrwC Cba monomer unit SEQ ID NO: 87 with the mutation Q276C, where the linker is attached to each monomer at position 577 on Hel 308 Mbu monomer and position 276 on TrwC Cba) plus unreacted TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO + Hel308 Mbu-A577Faz monomer (where each monomer unit comprises SEQ ID NO: 10 with the mutation A577Faz), lane j) TrwC Cba-

Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO + 5kDa PEG with an azide attached, lane k) TrwC Cba-Q276C monomer (SEQ ID NO: 87 with the mutation Q276C) + Mal-PEG4-DBCO + Azide Alexa Fluor® 555 (Life Technologies, used to check for non-specific interactions between the fluorophore and TrwC Cba-Q276C monomer).

5

WE CLAIM:

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

(a) contacting the target polynucleotide with a transmembrane pore and a construct comprising a helicase and an additional polynucleotide binding moiety, wherein the helicase is attached to the polynucleotide binding moiety and the construct has the ability to control the movement of a polynucleotide, such that the construct controls the movement of the target polynucleotide through the pore; and

(b) taking one or more electrical or optical measurements as the polynucleotide moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the target polynucleotide and thereby characterising the target polynucleotide; wherein the one or more characteristics of the target polynucleotide are selected from the length of the target polynucleotide, the identity of the target polynucleotide, the sequence of the target polynucleotide, the secondary structure of the target polynucleotide, and whether or not the target polynucleotide is modified;

wherein:

(i) the polynucleotide binding moiety comprises one or more helicases, wherein:

(ia) the two or more helicases in the construct are different from one another; or

(ib) the two or more helicases are the same or similar; or

(ic) the two or more helicases are the same or similar and are attached using the same amino acid residue in each helicase; or

(ii) the polynucleotide binding moiety:

(iia) comprises one or more domains independently selected from helix-hairpin-helix (HhH) domains, eukaryotic single-stranded binding proteins (SSBs), bacterial SSBs, archaeal SSBs, viral SSBs, double-stranded binding proteins, sliding clamps, processivity factors, DNA binding loops, replication initiation proteins, telomere binding proteins, repressors, zinc fingers and proliferating cell nuclear antigens (PCNAs); or

(iib) the polynucleotide binding moiety is selected from any one of: the SSBEco protein from *Escherichia coli* (sequence accession number P0AGE0); the SSBBhe protein from *Bartonella henselae* (sequence accession number Q6G302); the SSBCbu protein from *Coxiella burnetii* (sequence accession number Q83EP4); the SSBTma protein from *Thermathoga maritima* (sequence accession number

Q9WZ73); the SSBHpy protein from *Helicobacter pylori* (sequence accession number O25841); the SSBDra protein from *Deinococcus radiodurans* (sequence accession number Q9RY51); the SSBTaq protein from *Thermus aquaticus* (sequence accession number Q9KH06); the SSBMsm protein from *Mycobacterium smegmatis* (sequence accession number Q9AFI5); the SSBSso protein from *Sulfolobus solfataricus* (sequence accession number Q97W73); the SSBMHsmt protein from *Homo sapiens* (sequence accession number Q04837); the SSBMle protein from *Mycobacterium leprae* (sequence accession number P46390); the gp32T4 protein from Bacteriohage T4 (sequence accession number P03695); the gp32RB69 protein from Bacteriophage RB69 (sequence accession number Q7Y265); the gp2.5T7 protein from Bacteriohage T7 (sequence accession number P03696); the UL42 protein from Herpes virus 1 (sequence accession number P10226); the UL44 protein from Herpes virus 5 (cytomegalovirus) (sequence accession number P16790); the pf8 protein from KSHV (sequence accession number Q77ZG5); the RPAMja protein from *Methanococcus jannaschii* (sequence accession number Q58559); the RPAMma protein from *Methanococcus maripaludis* (sequence accession number Q6LYF9); the RPA70Sce protein from *Saccharomyces cerevisiae* (sequence accession number P22336); the RPAMbu1 protein from *Methanococcoides burtonii* (sequence accession number Q12V72); the RPAMbu2 protein from *Methanococcoides burtonii* (sequence accession number Q12W96); the RPA70Hsa protein from *Homo sapiens* (sequence accession number P27694); the RPA14Hsa protein from *Homo sapiens* (sequence accession number P35244); the gp45T4 protein from Bacteriophage T4 (sequence accession number P04525); the BetaEco protein from *E. coli* (sequence accession number P0A988); the PCNASce protein from *Saccharomyces cerevisiae* (sequence accession number P15873); the PCNATko protein from *Thermococcus kodakaraensis* (sequence accession number Q5JF32); the PCNAHvo protein from *Haloferax volcanii* (sequence accession number D0VWY8); the PCNAPfu protein from *Pyrococcus furiosus* (sequence accession number O73947); the PCNAMbu protein from *Methanococcoides burtonii* (sequence accession number Q12U18); the BetaMtu protein from *Mycobacterium tuberculosis* (sequence accession number Q50790); the BetaTma protein from *Thermotoga maritima* (sequence accession number Q9WYA0); the BetaSpy protein from

Date Reçue/Date Received 2020-09-29

Streptococcus pyrogenes (sequence accession number Q9EVR1); the gp45RB69 protein from Bacteriophage RB69 (sequence accession number O80164); the p55Hsa protein from *Homo sapiens* (mitochondrial) (sequence accession number Q9UHN); the p55Dme protein from *Drosophylla melanogaster* (sequence accession number Q9VJV8); the p55Xla protein from *Xenopus laevis* (sequence accession number Q9W6G7); the RepDSau protein from *Staphylococcus aureus* (sequence accession number P08115); the G2P protein from Enterobacteria phage 1 (sequence accession number P69546); the MutLEco protein from *Escherichia coli* (sequence accession number P23367); the KuMtu protein from *Mycobacterium tuberculosis* (sequence accession number O05866); the OnTEBP protein from *Oxytricha nova*-Alpha (sequence accession number P29549); the protein from *Oxytricha nova*-Beta (sequence accession number P16458); the EcrTEBP protein from *Euplotes crassus* (sequence accession number Q06183); the TteTEBP protein from *Tetrachymena termophila* Alpha (sequence accession number Q23FB9); the protein from *Tetrachymena termophila* Beta having sequence accession number Q23FH0; the pot1Spo protein from *Schizosaccharomyces pombe* (sequence accession number O13988); the Cdc13pSce protein from *Saccharomyces cerevisiae* (sequence accession number C7GSV7); the C1 protein from Bacteriophage 434 (sequence accession number P16117); and the LexA protein from *Escherichia coli* (sequence accession number P0A7C2); and variants thereof having at least 80% identity to the relevant sequence based on amino acid identity over the entire sequence; or

(iic) is an exonuclease, polymerase or topoisomerase; or

(iid) is Phi29 polymerase (SEQ ID NO: 62) or a variant thereof having at least 80% identity to SEQ ID NO: 62 based on amino acid identity over the entire sequence.

2. The method of claim 1, wherein the helicase and the polynucleotide binding moiety are covalently attached, are chemically attached or are genetically fused.

3. The method of claim 1 or claim 2 wherein the helicase and the polynucleotide binding moiety are attached by one or more linkers.

4. The method of claim 3 wherein the helicase and the polynucleotide binding moiety are attached by one or more linkers comprised of amino acids.

5. The method of any one of claims 1 to 4, wherein the one or more helicases are independently selected from superfamilies 1 to 6.

6. The method of any one of claims 1 to 5 wherein the one or more helicases are monomeric.

7. The method of any one of claims 1 to 6 wherein the one or more helicases are independently selected from Hel308 helicases, RecD helicases, Tral helicases, and XPD helicases.

8. The method of claim 7, wherein:

(a) the Hel308 helicase comprises the amino acid motif Q-X1-X2-G-R-A-G-R (SEQ ID NO: 8), wherein X1 is C, M or L and X2 is any amino acid residue; or

(b) the RecD helicase comprises the amino acid motif X1-X2-X3-G-X4-X5-X6-X7 (SEQ ID NO: 20), wherein X1 is G, S or A, X2 is any amino acid, X3 is P, A, S or G, X4 is T, A, V, S or C, X5 is G or A, X6 is K or R and X7 is T or S; or

(c) the TraI helicase comprises:

- the amino acid motif H-$(X1)_2$-X2-R-$(X3)_{5-12}$-H-X4-H (SEQ ID NO: 31-38), wherein X1 and X3 are any amino acid and X2 and X4 are independently selected from any amino acid except D, E, K and R; or

- the amino acid motif G-X1-X2-X3-X4-X5-X6-X7-H-$(X8)_{6-12}$-H-X9 (SEQ ID NO: 39-45), wherein X1, X2, X3, X5, X6, X7 and X9 are independently selected from any amino acid except D, E, K and R, X4 is D or E and X8 is any amino acid; or

(d) the XPD helicase comprises the amino acid motif X1-X2-X3-G-X4-X5-X6-E-G (SEQ ID NO: 8), wherein X1, X2, X5 and X6 are independently selected from any amino acid except D, E, K and R and wherein X3 and X4 may be any amino acid residue

9. The method of claim 7, wherein:

(a) the RecD helicase comprises the amino acid motif X1-X2-X3-X4-X5-$(X6)_3$-Q-X7 (SEQ ID NO: 29), wherein X1 is Y, W or F, X2 is A, T, S, M, C or V, X3 is any amino acid, X4 is T, N or S, X5 is A, T, G, S, V or I, X6 is any amino acid and X7 is G or S; or

(b) the XPD helicase comprises the amino acid motif Q-Xa-Xb-G-R-Xc-Xd-R-$(Xe)_3$-Xf-$(Xg)_7$-D-Xh-R (SEQ ID NO: 9), wherein Xa, Xe and Xg may be any amino acid residue and wherein Xb, Xc and Xd are independently selected from any amino acid except D, E, K and R.

10. The method of claim 8, wherein X2 in Hel308 helicase is A, F, M, C, V, L, I, S, T or P.

11. The method of claim 8 or claim 10 wherein the Hel308 helicase comprises (i) the sequence shown in any one of SEQ ID NOs: 10, 13, 16 or 19 or (ii) a variant thereof having at least 40% identity to the relevant sequence based on amino acid identity over the entire sequence and retains helicase activity.

12. The method of claim 8 wherein the TraI helicase comprises (i) the sequence shown in SEQ ID NO: 46, 87, 98 or 102 or (ii) a variant thereof having at least 40% identity to the relevant sequence based on amino acid identity over the entire sequence and retains helicase activity.

13. The method of claim 8 or claim 9 wherein:

(a) X1, X2, X5 and X6 and/or Xb, Xc and Xd in the XPD helicase are independently selected from G, P, A, V, L, I, M, C, F, Y, W, H, Q, N, S and T; or

(b) the XPD helicase comprises (i) the sequence shown in SEQ ID NO: 52 or (ii) a variant thereof having at least 40% identity to SEQ ID NO: 52 based on amino acid identity over the entire sequence and retains helicase activity.

14. The method of any one of claims 1 to 13, wherein the one or more helicases are modified to facilitate the attachment.

15. The method of any one of claims 1 to 14, wherein the one or more helicases are modified to facilitate the attachment by the introduction of one or more non-native cysteine residues and/or one or more 4-azido-L-phenylalanine (Faz) residues.

16. The method of any one of claims 1 to 15, wherein the one or more characteristics of the target polynucleotide are measured by electrical measurement which is a current measurement, an impedance measurement, a tunnelling measurement or a field effect transistor (FET) measurement.

Date Reçue/Date Received 2020-09-29

17. A construct comprising two or more helicases, wherein the helicases are attached together and the construct has the ability to control the movement of a polynucleotide, wherein the helicases are attached together by covalent attachment, chemical attachment or via one or more linkers, and wherein the helicases comprised with the construct are capable of attaching to the polynucleotide by binding to the polynucleotide at an internal nucleotide.

18. The construct of claim 17, wherein the two or more helicases are as defined in claim 1 or any one of claims 5 to 15.

19. A construct comprising (a) a helicase and a polypeptide having an amino acid sequence comprising SEQ ID NO: 94 or a variant thereof having at least 80% identity to SEQ ID NO: 94 based on amino acid identity over the entire sequence of SEQ ID NO: 94, wherein the helicase is attached to the polypeptide, or (b) a polypeptide having an amino acid sequence comprising SEQ ID NO: 90 or a variant thereof having at least 80% identity to SEQ ID NO: 90 based on amino acid identity over the entire sequence of SEQ ID NO: 90,
wherein the construct has the ability to control the movement of a polynucleotide.

20. A sensor for characterising a target polynucleotide, comprising a complex between a pore and a construct as defined in any one of claims 1 to 15 or a construct according to any one of claims 17 to 19.

21. Use of a construct as defined in any one of claims 1 to 15 or a construct according to any one of claims 17 to 19 to control the movement of a target polynucleotide through a pore.

Date Reçue/Date Received 2020-09-29

Fig. 1 1/28

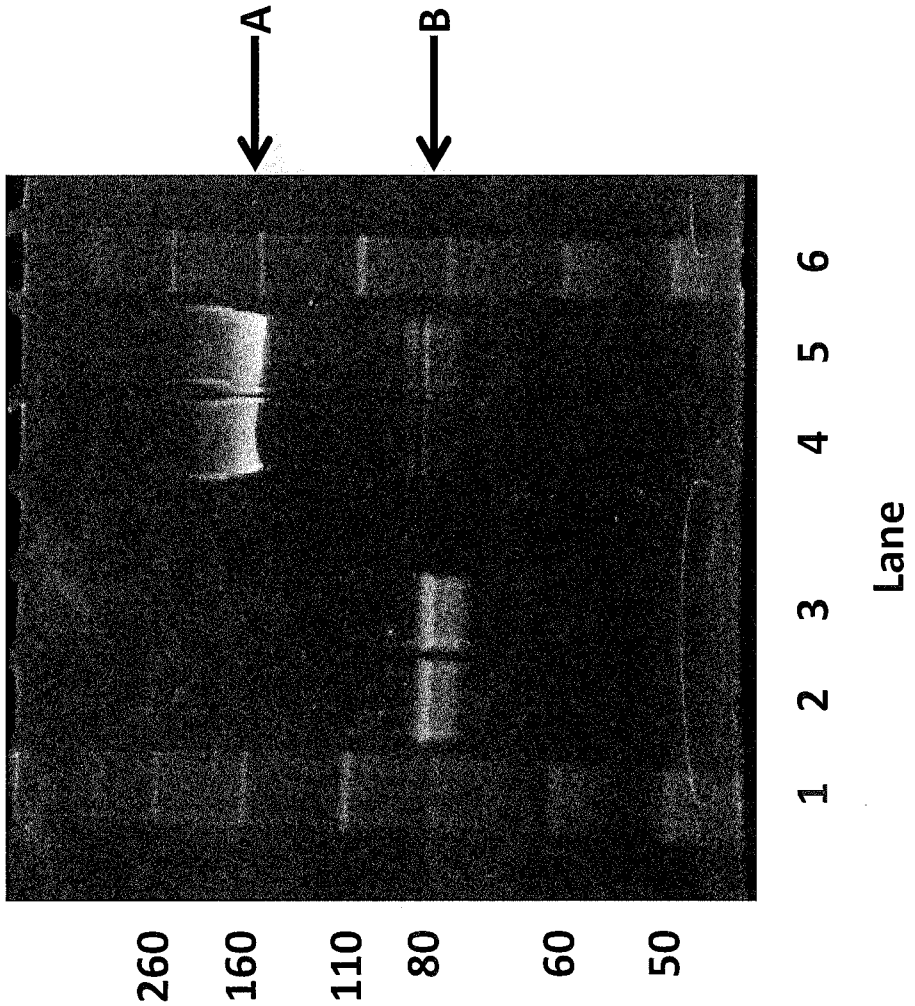


Fig. 2 2/28

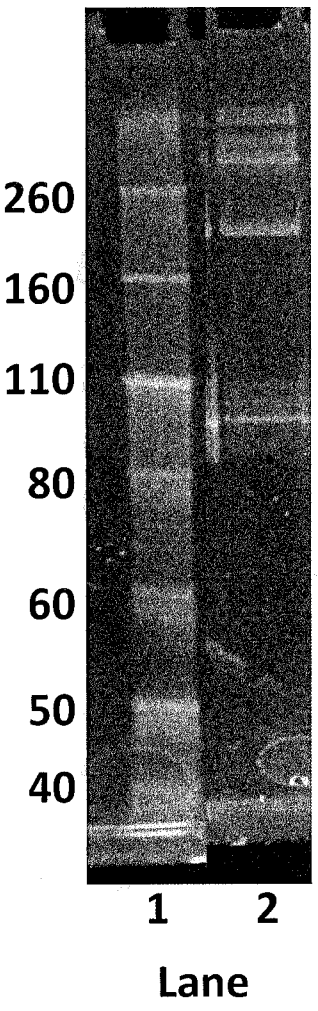


Fig. 3 3/28

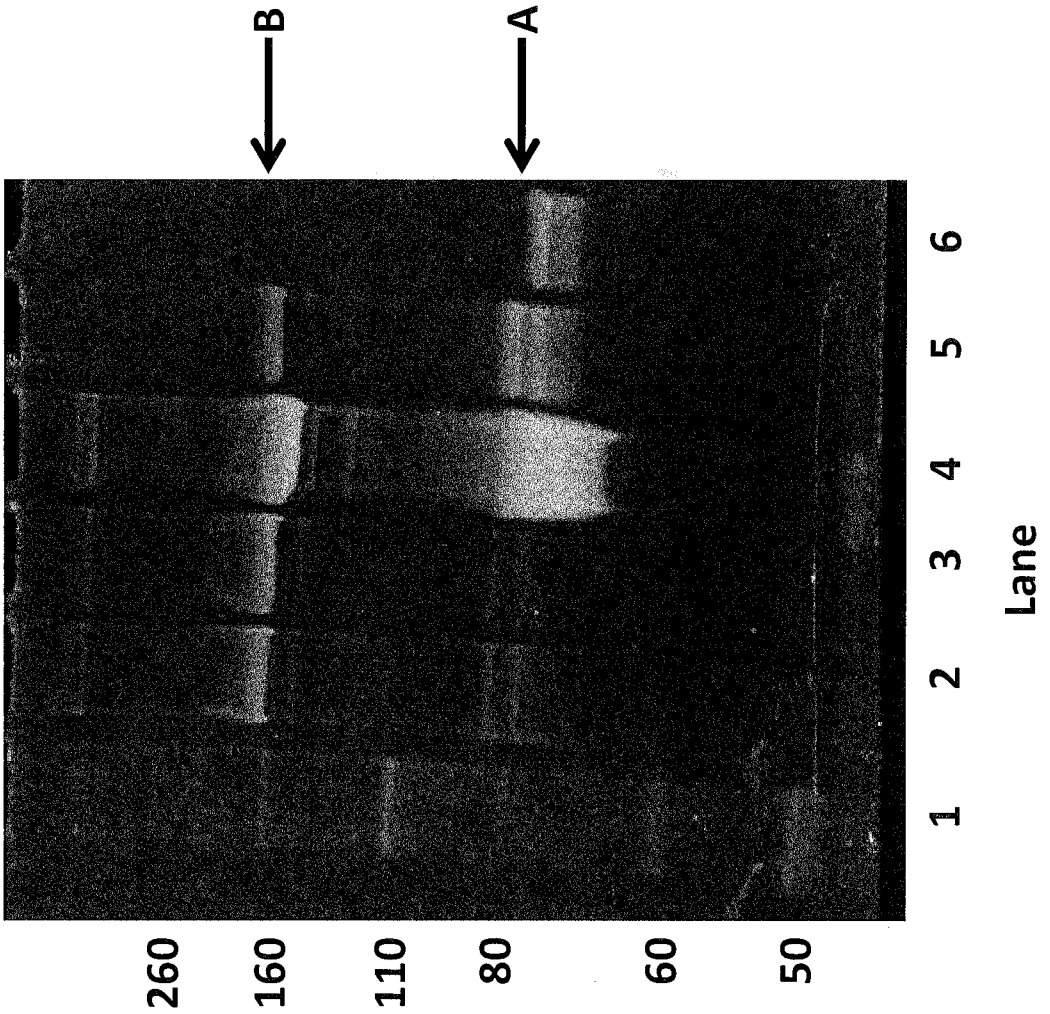


Fig. 4 4/28

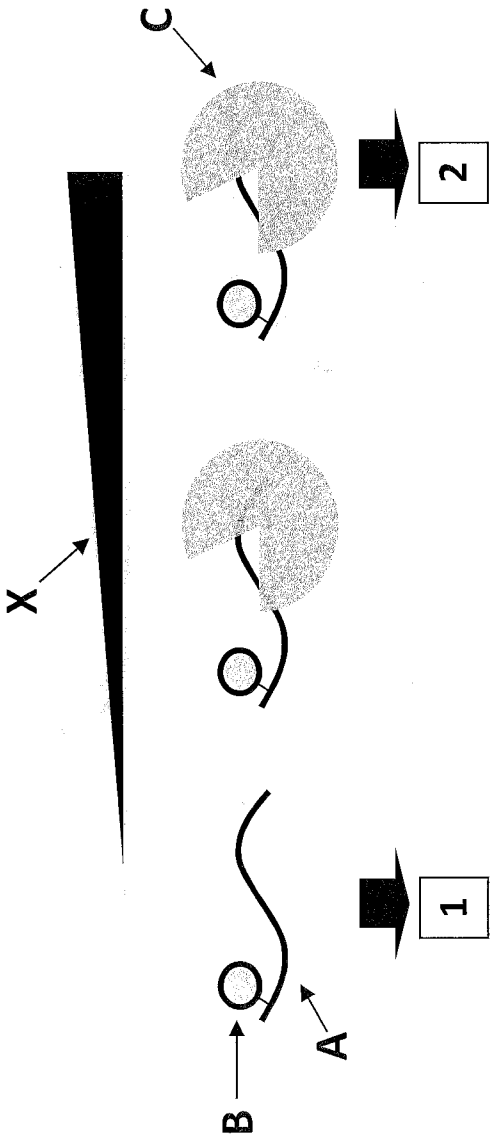


Fig. 5 5/28

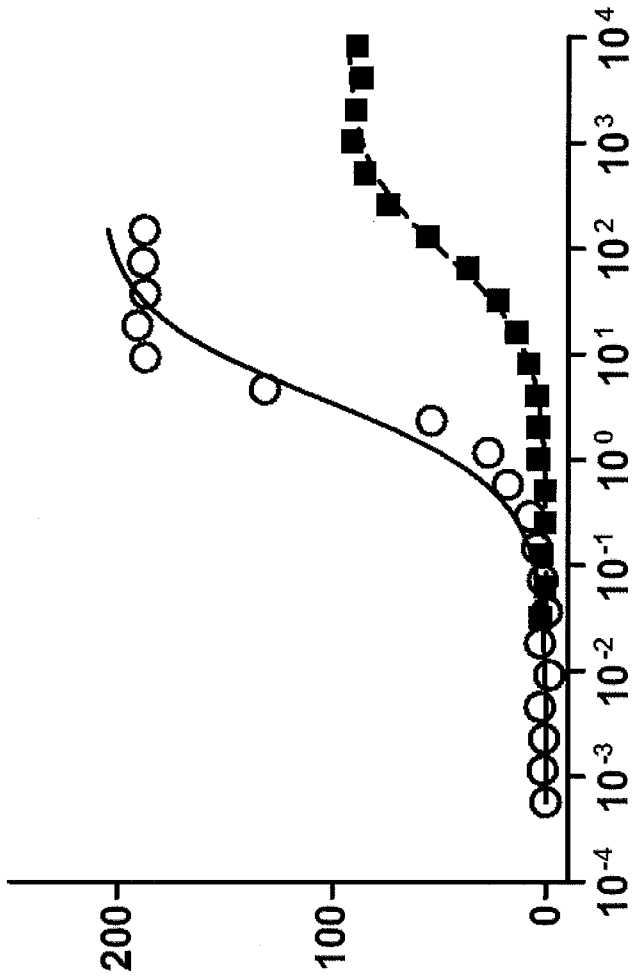


Fig. 6 6/28

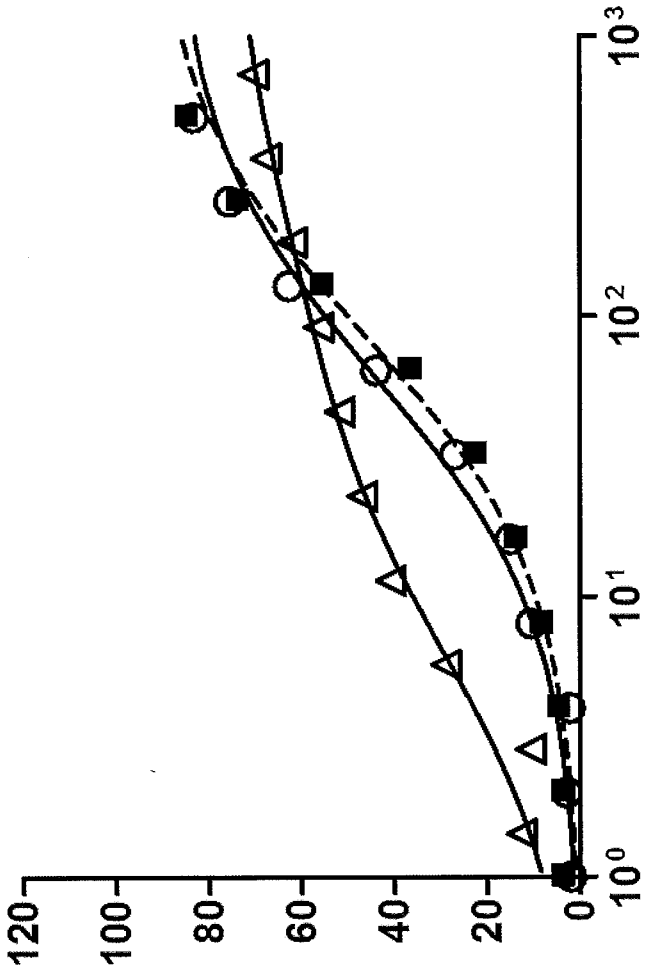


Fig. 7 7/28

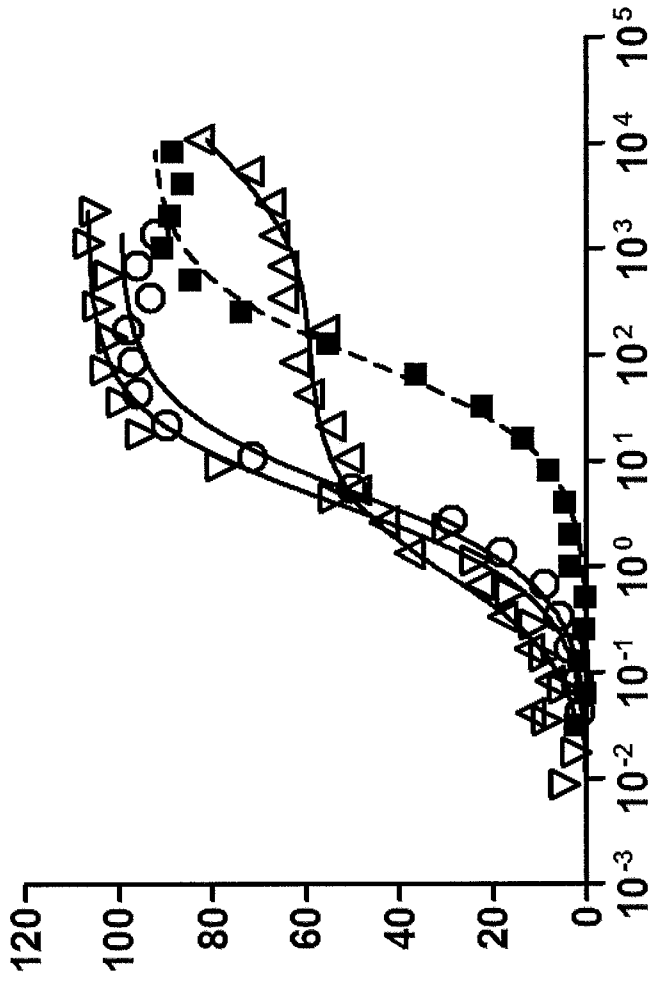


Fig. 8 8/28

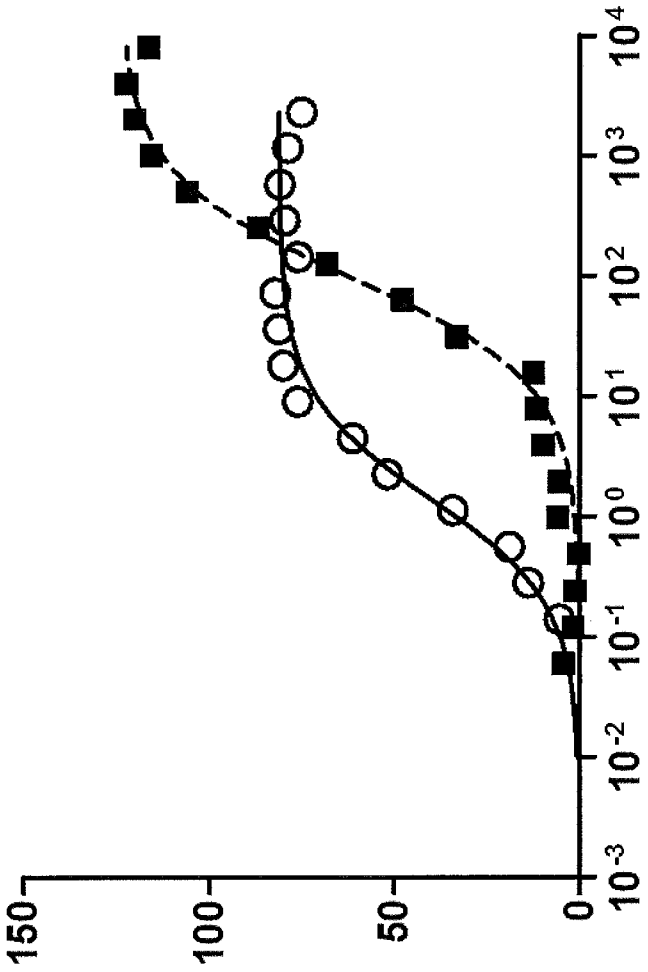


Fig. 9 9/28

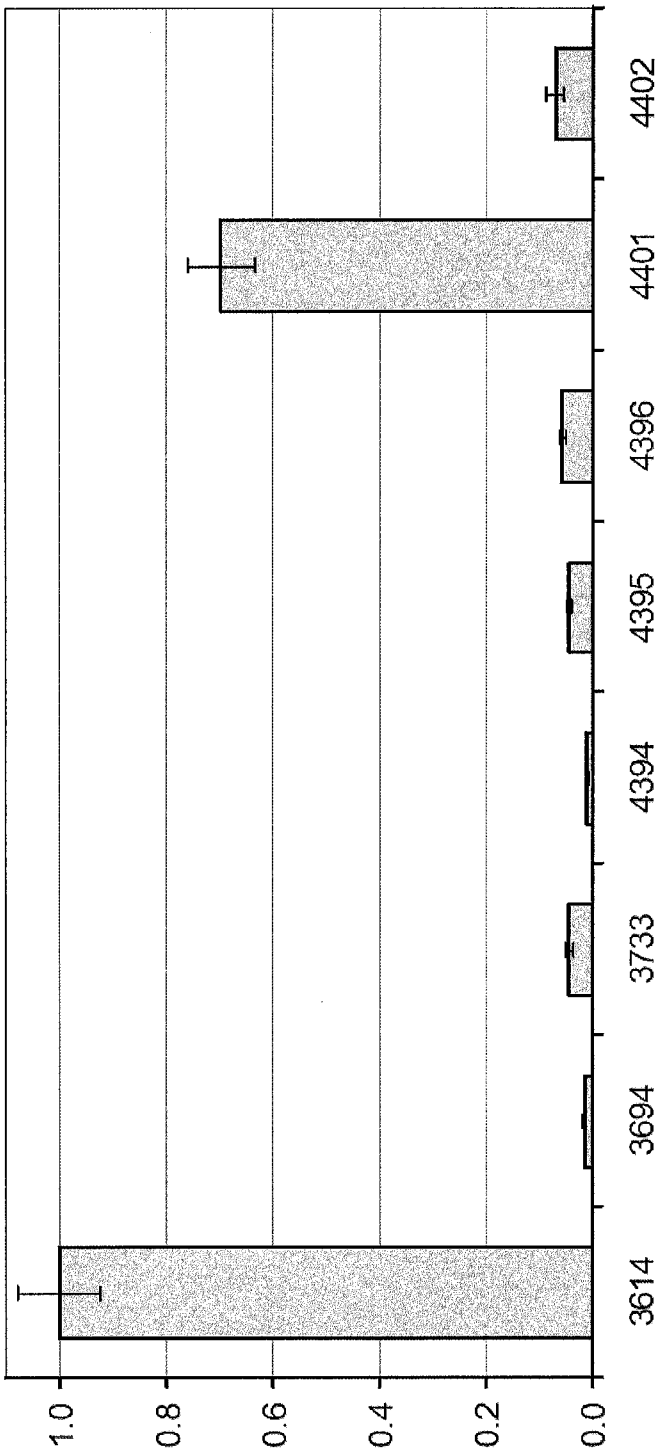


Fig. 10

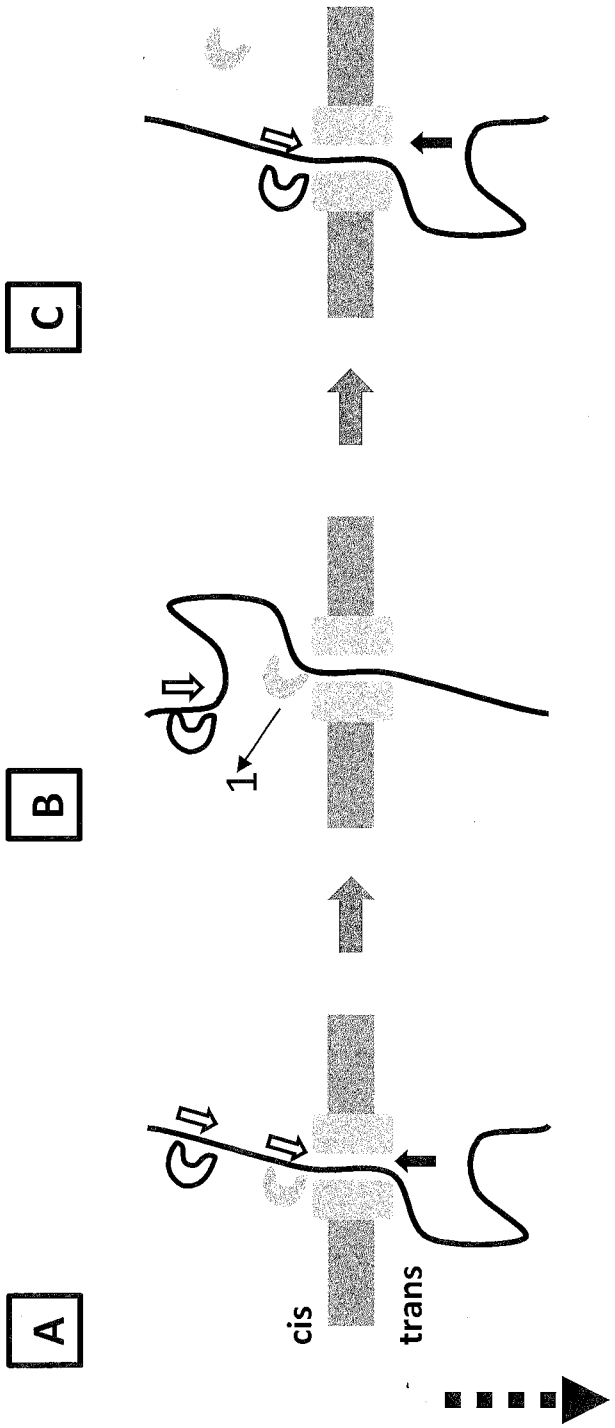


Fig. 11 11/28

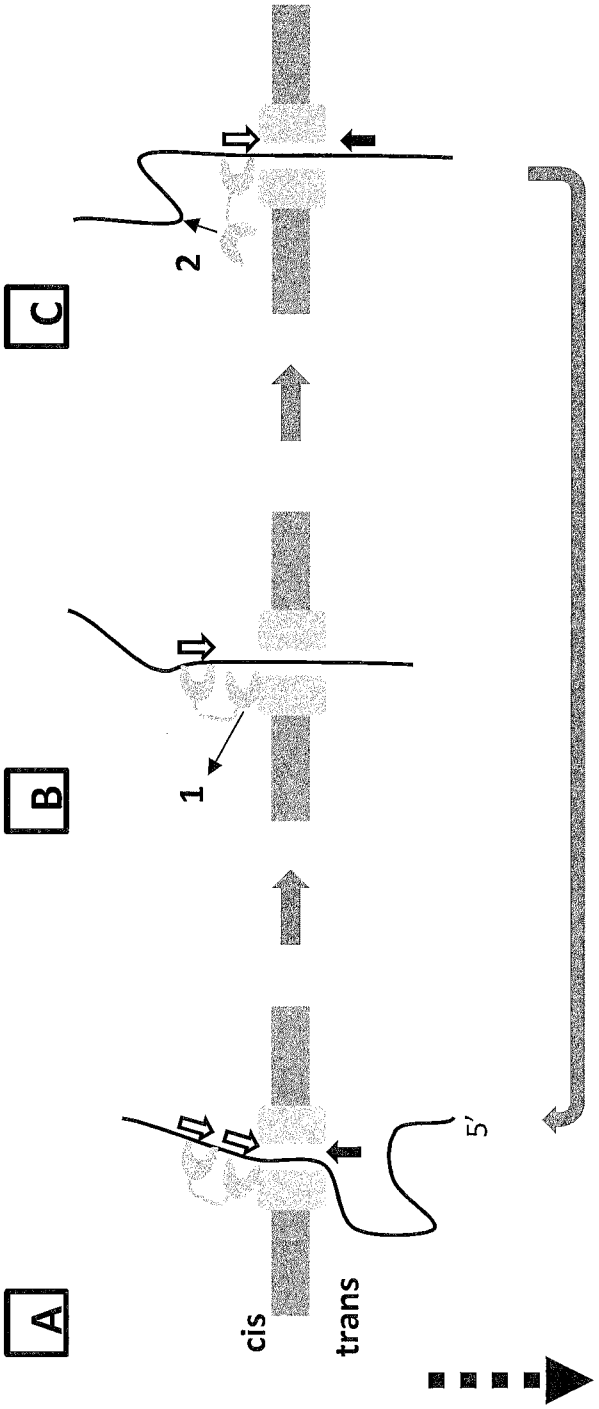


Fig. 12 12/28

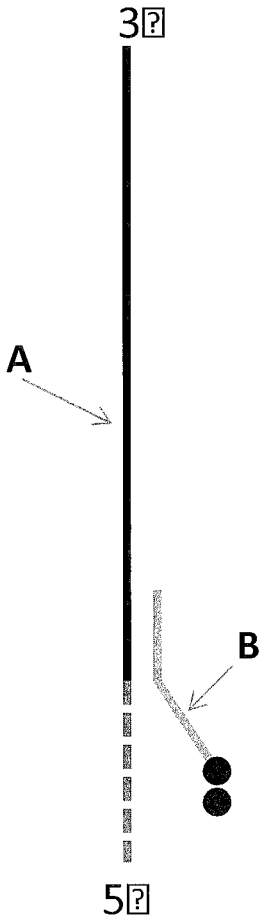


Fig. 13 13/28

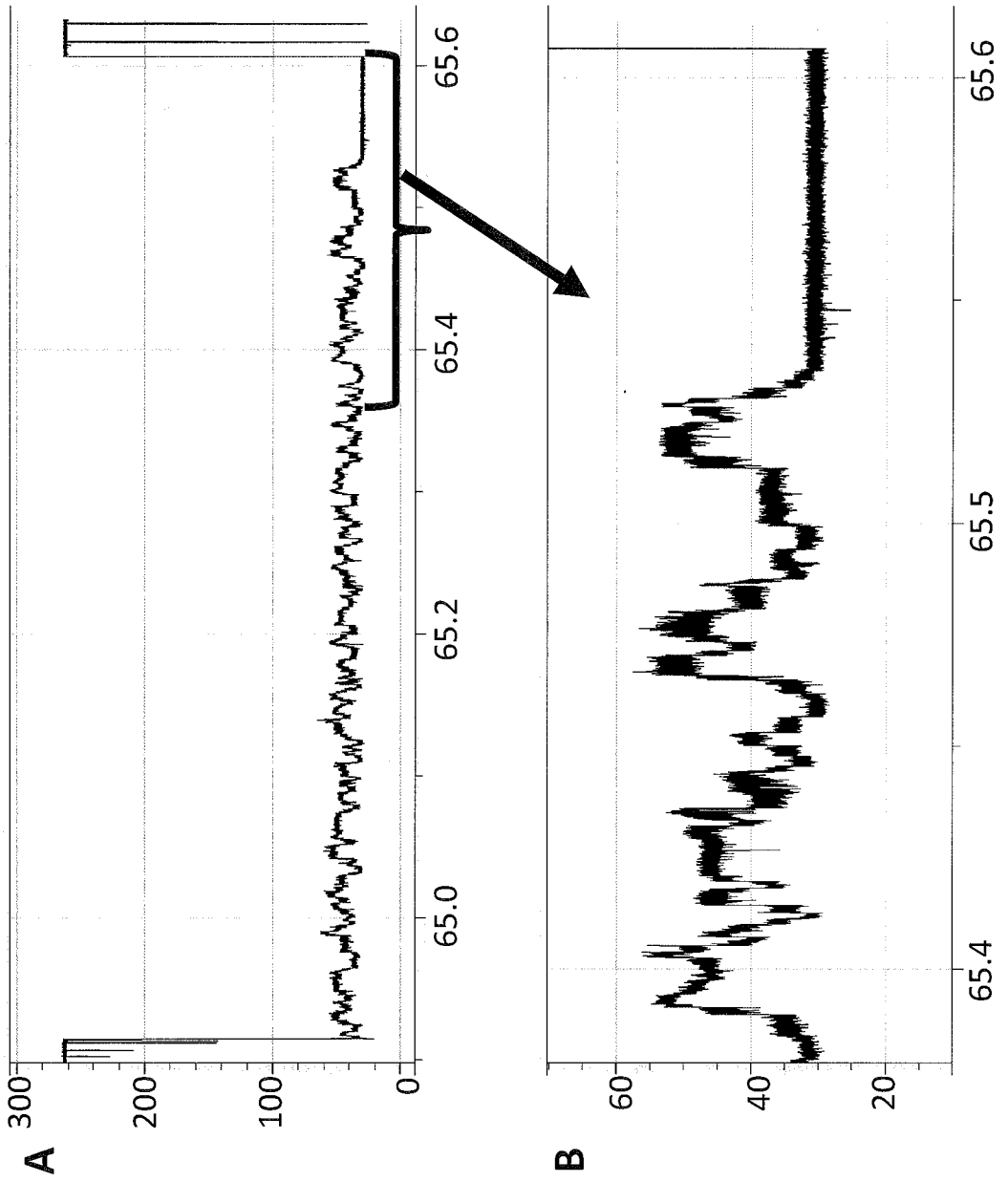


Fig. 14 14/28

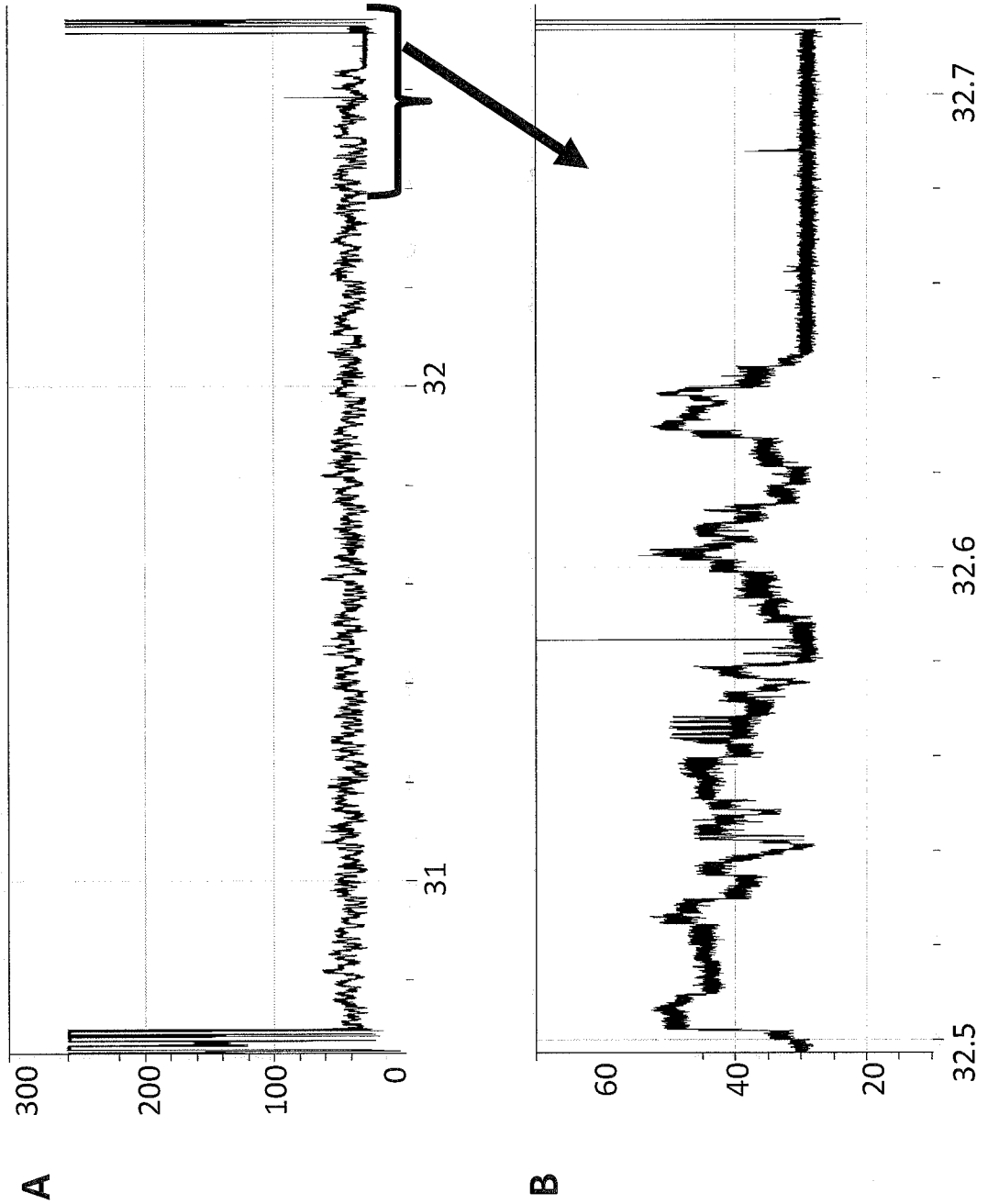


Fig. 15 15/28

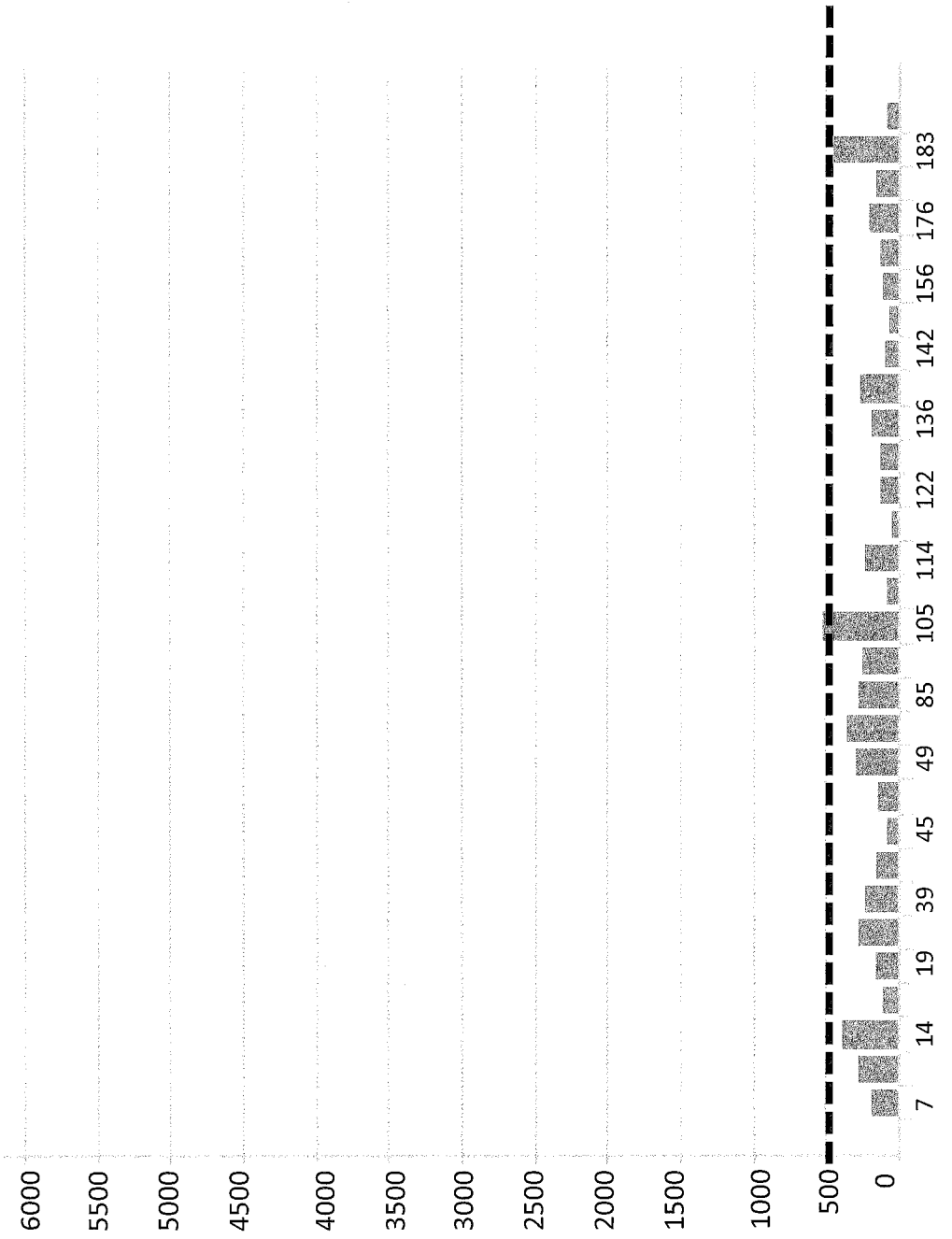

6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
7
14
19
39
45
49
85
105
114
122
136
142
156
176
183

Fig. 16 16/28

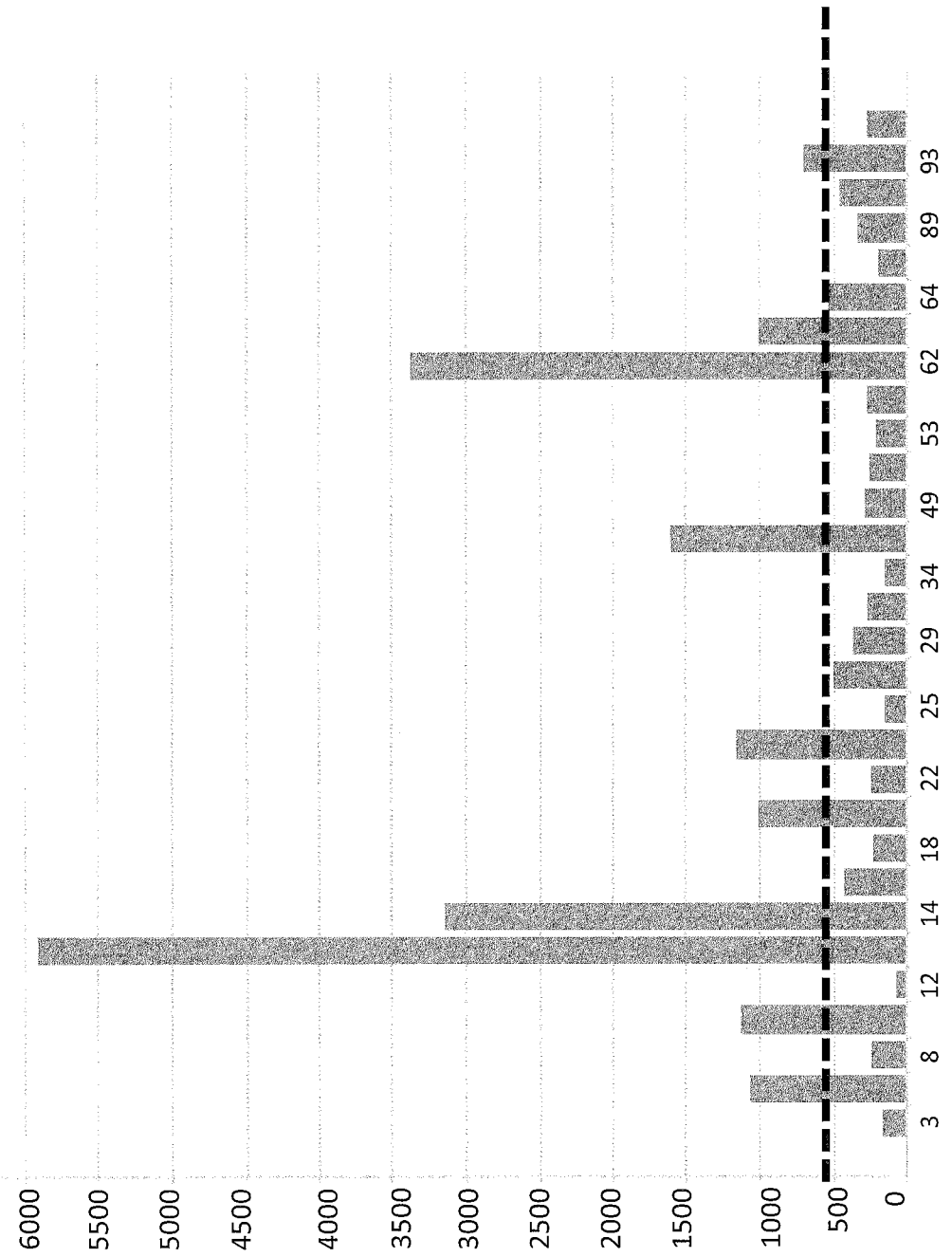

6000
5500
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
3
8
12
14
18
22
25
29
34
49
53
62
64
89
93

Fig. 17

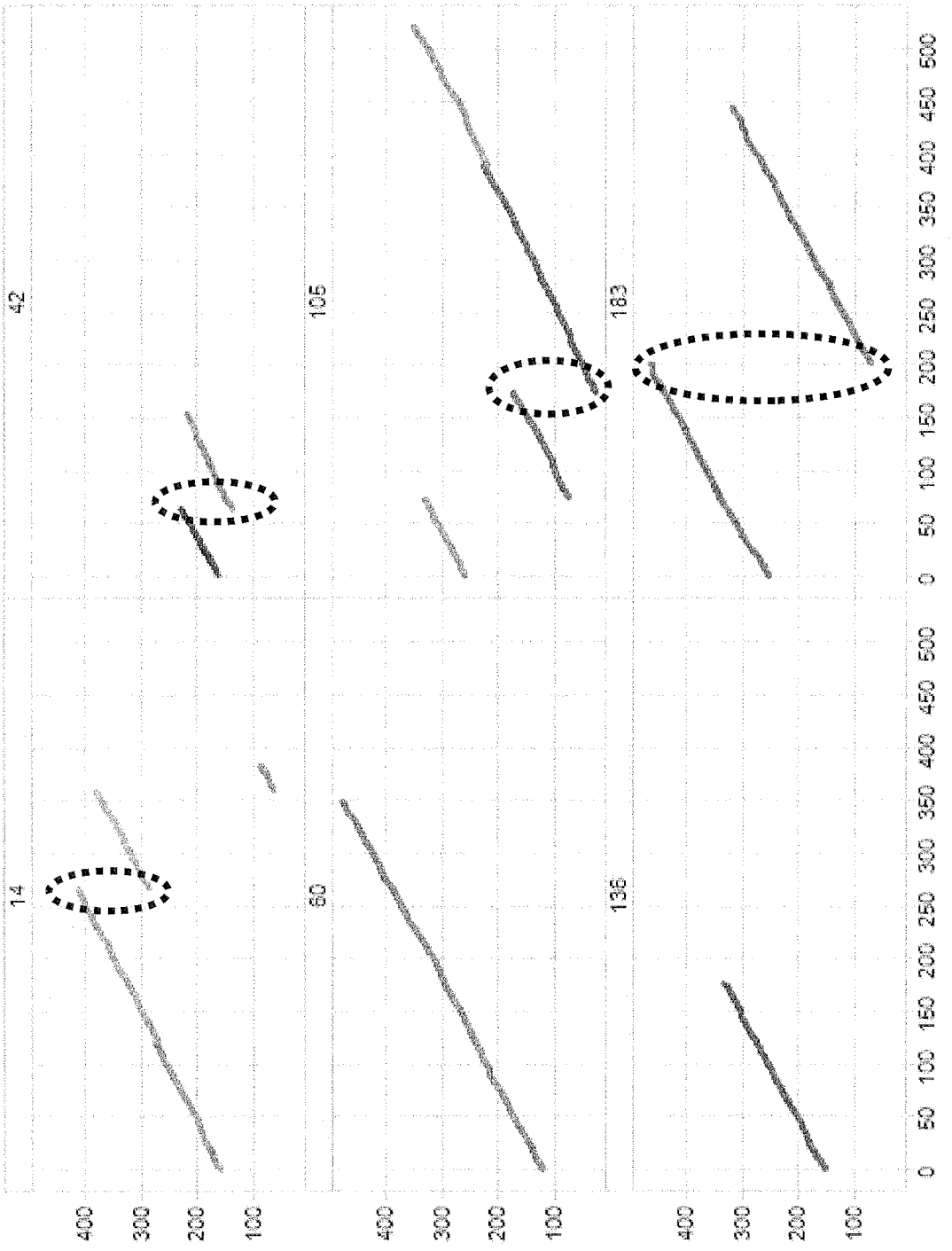

42
105
183
14
80
136
0 50 100 150 200 250 300 350 400 450 500
400 300 200 100

Fig. 18 18/28

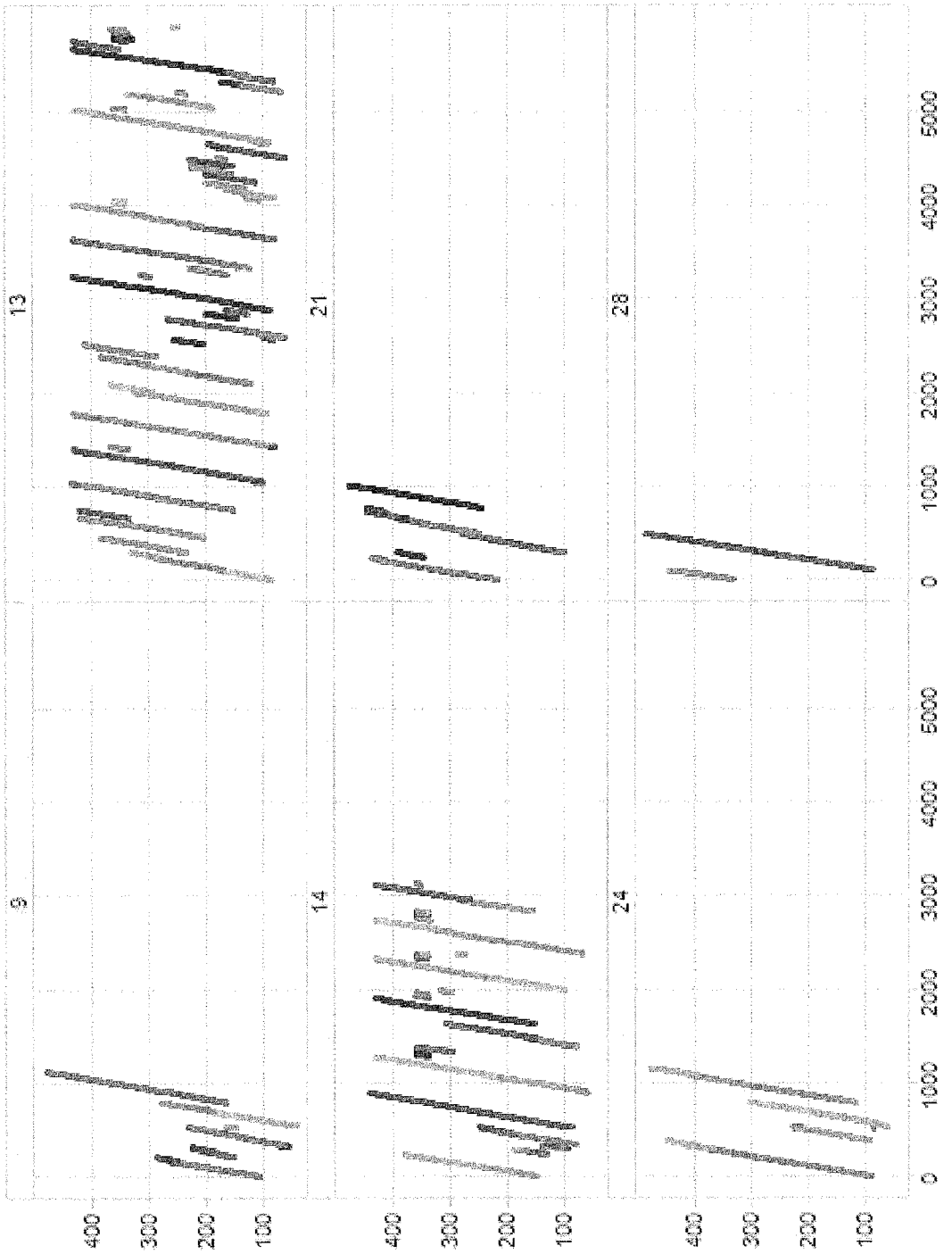


Fig. 19 19/28

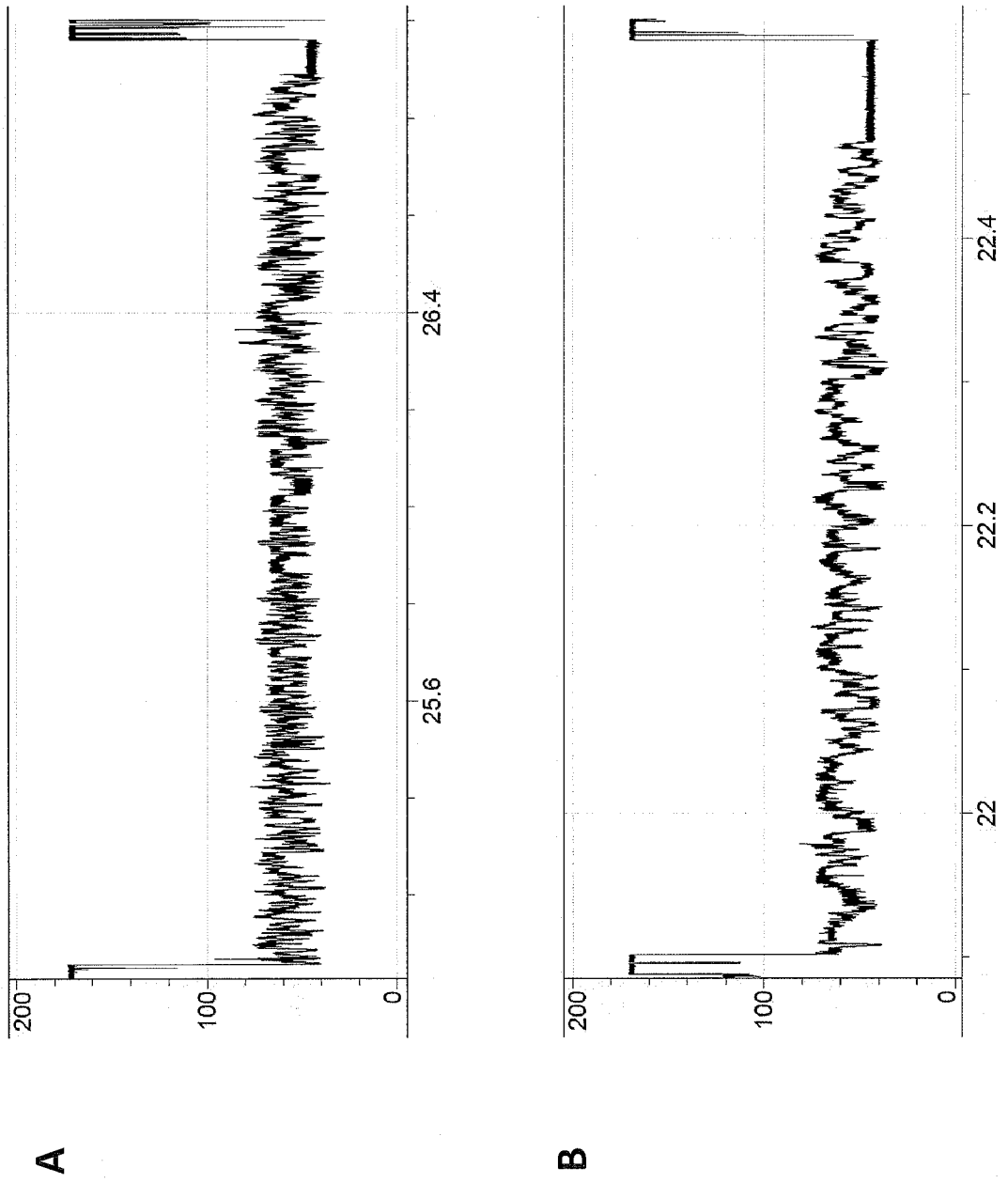


Fig. 20 20/28

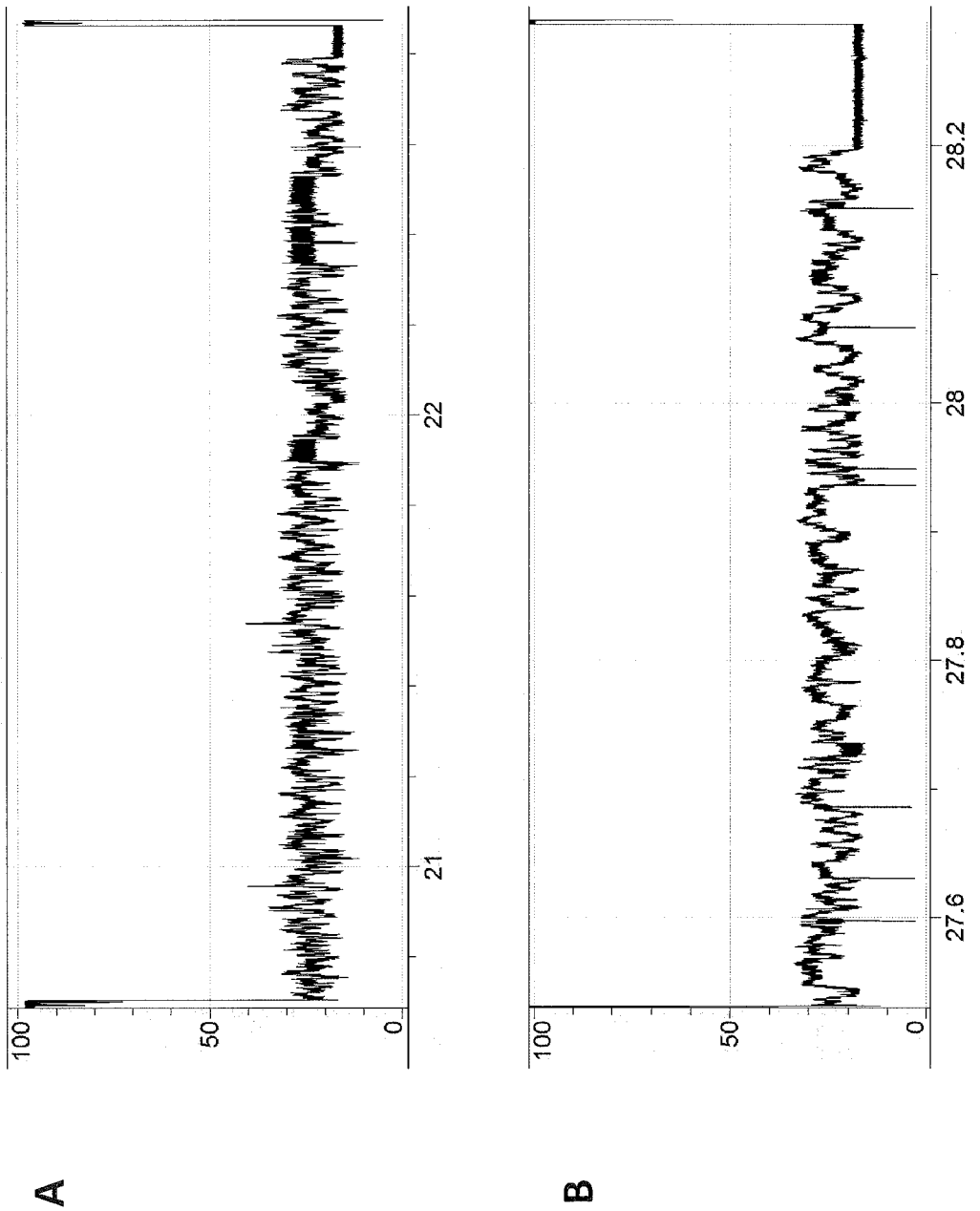


Fig. 21 21/28

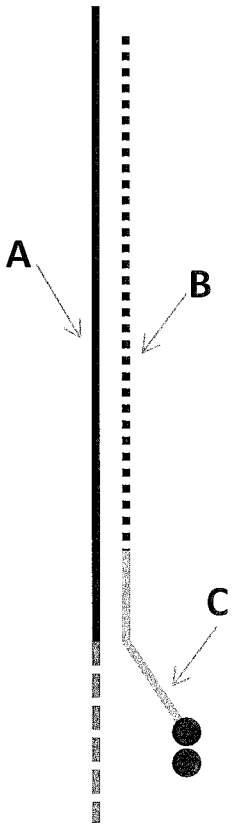


Fig. 22 22/28

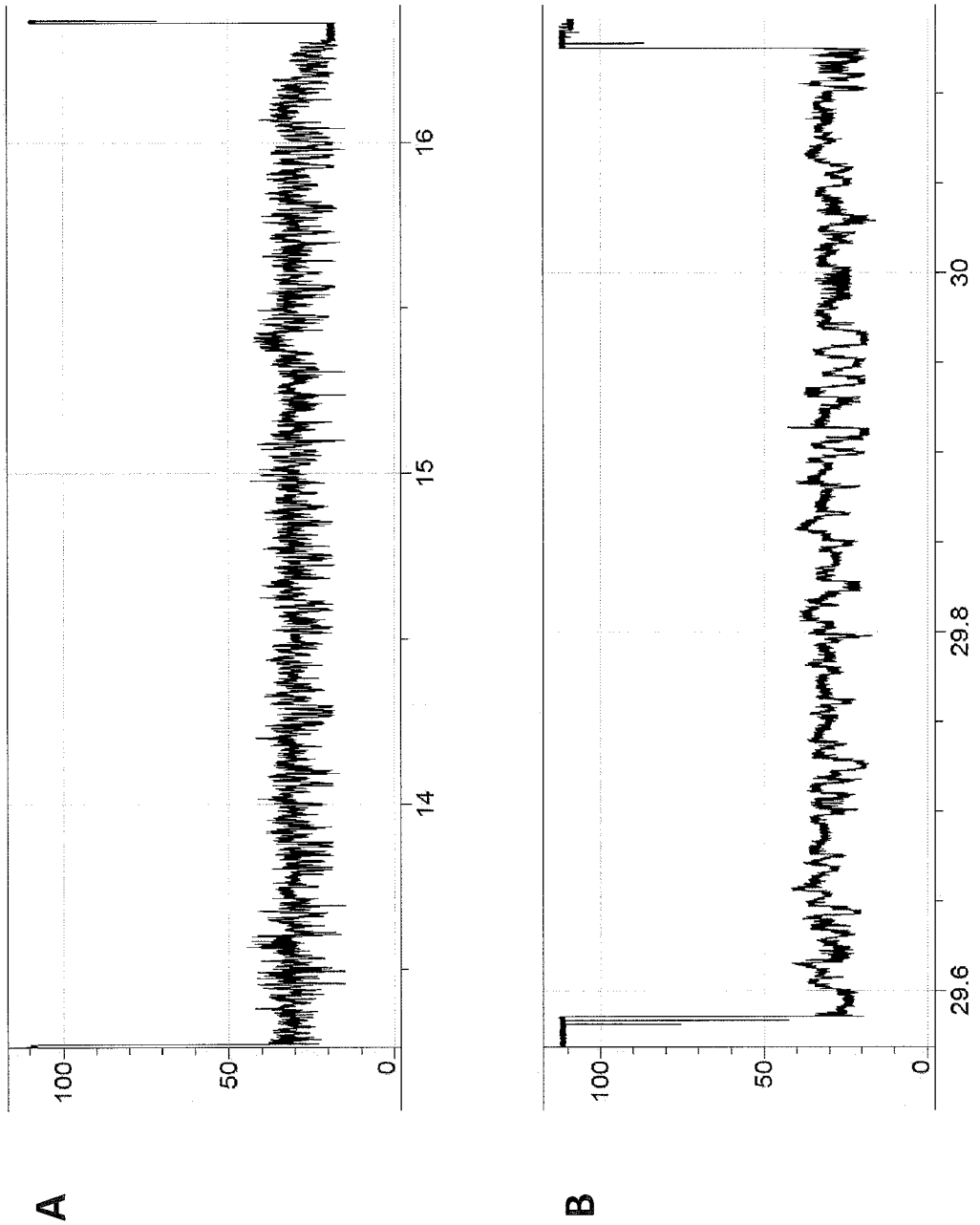


Fig. 23

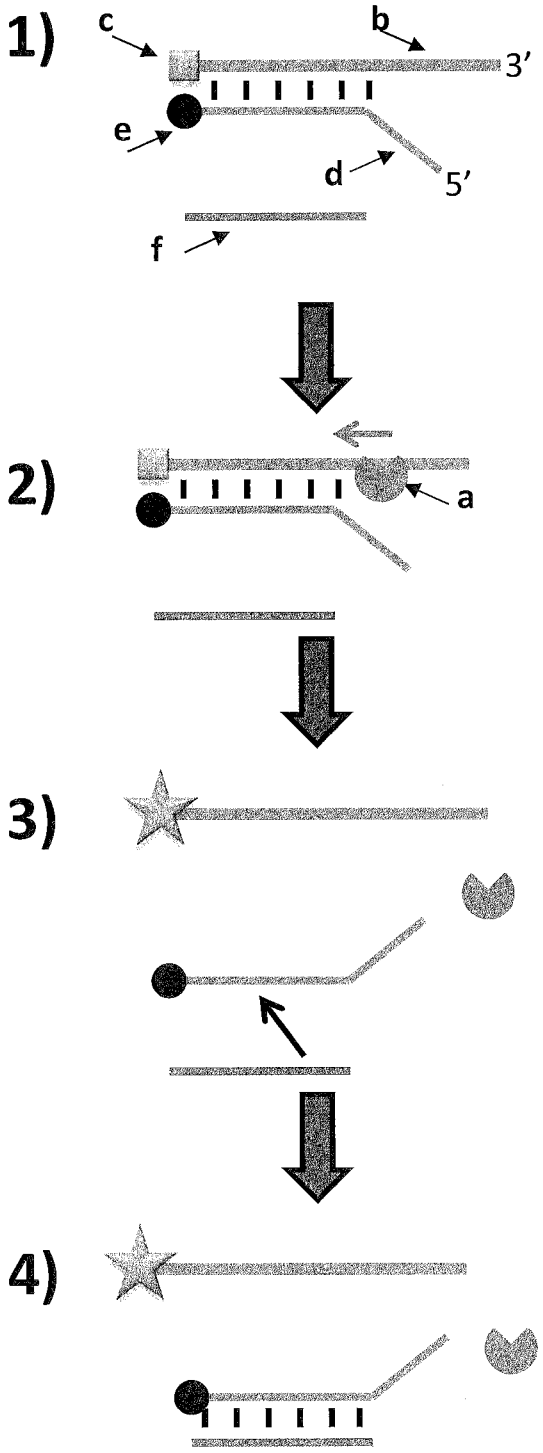


Fig. 24 24/28

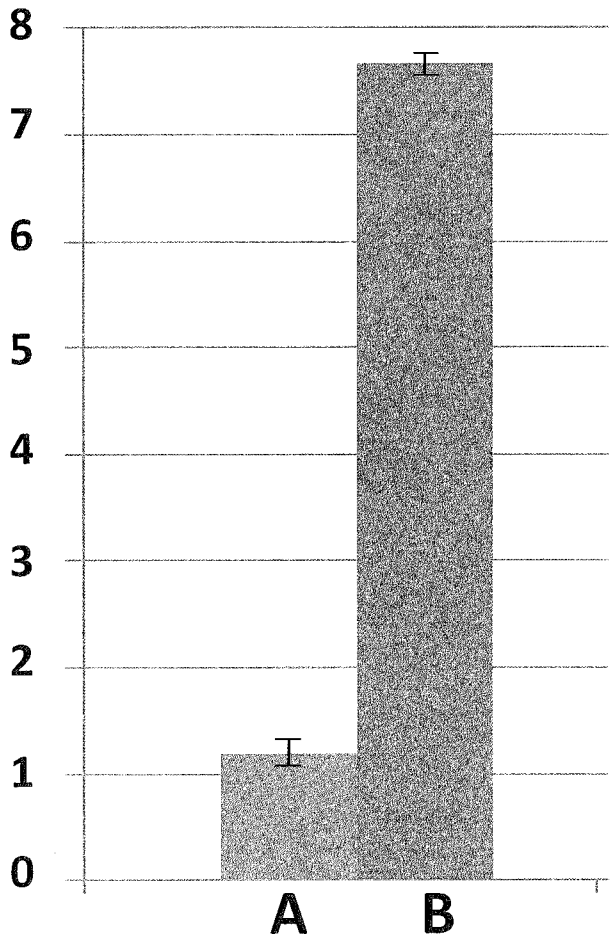


Fig. 25 25/28

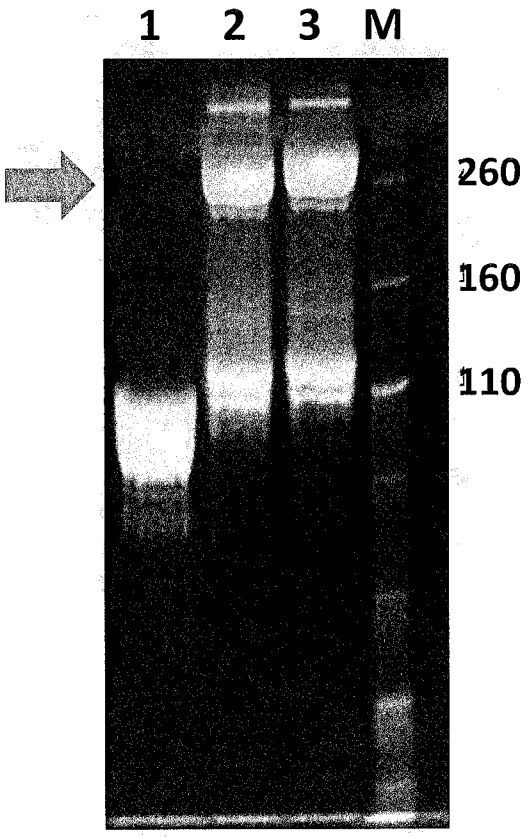


Fig. 26 26/28

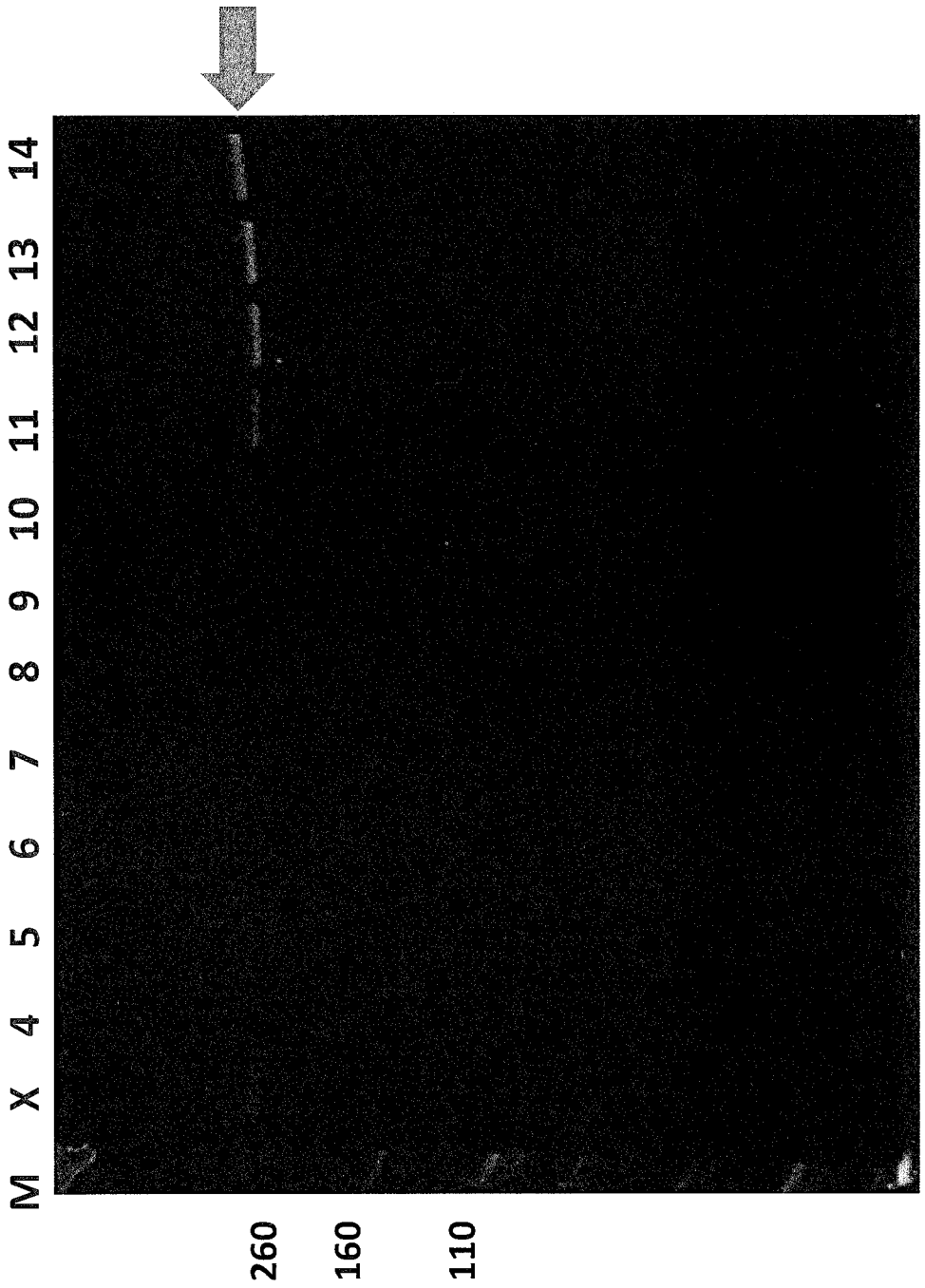


Fig. 27 27/28

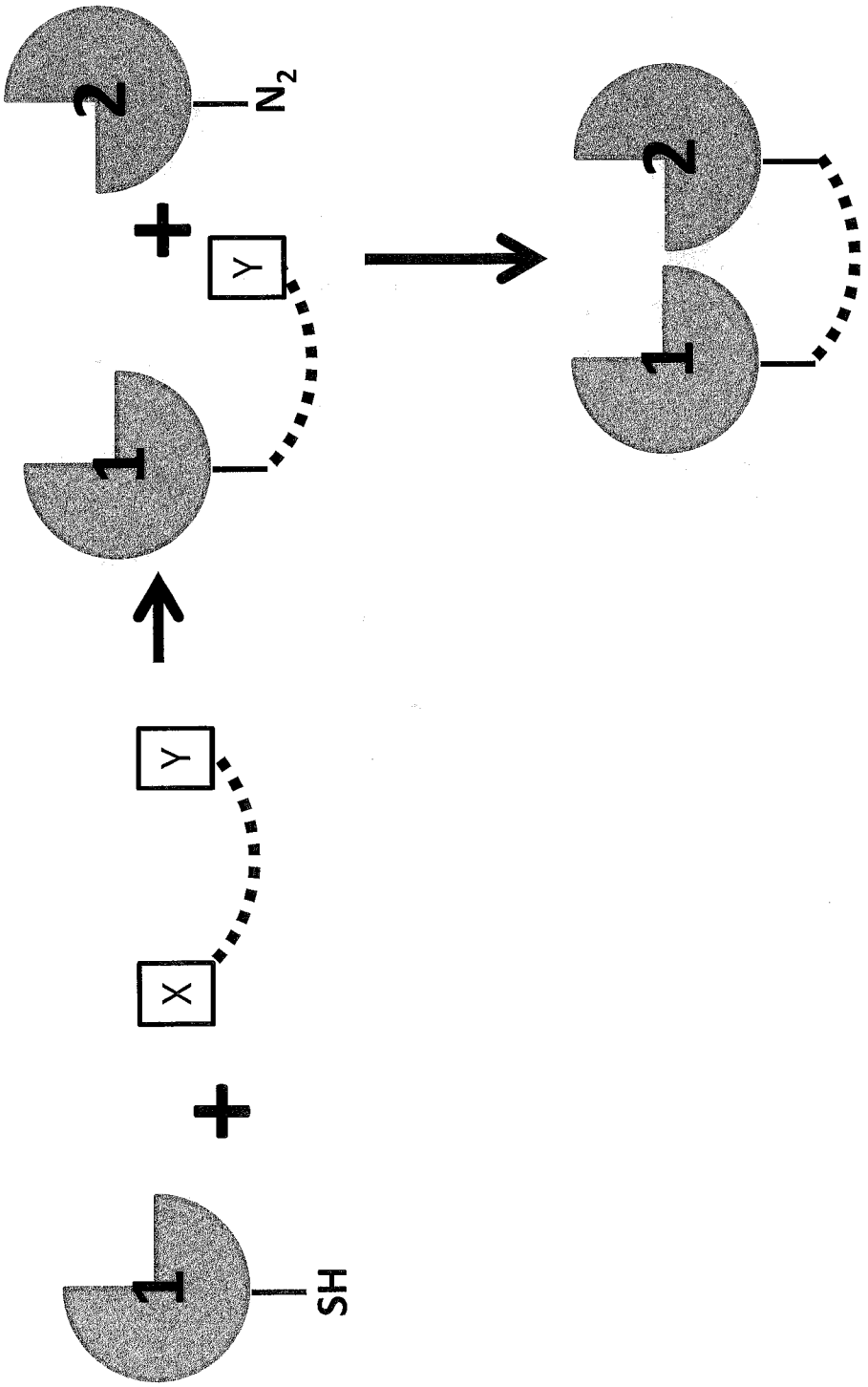
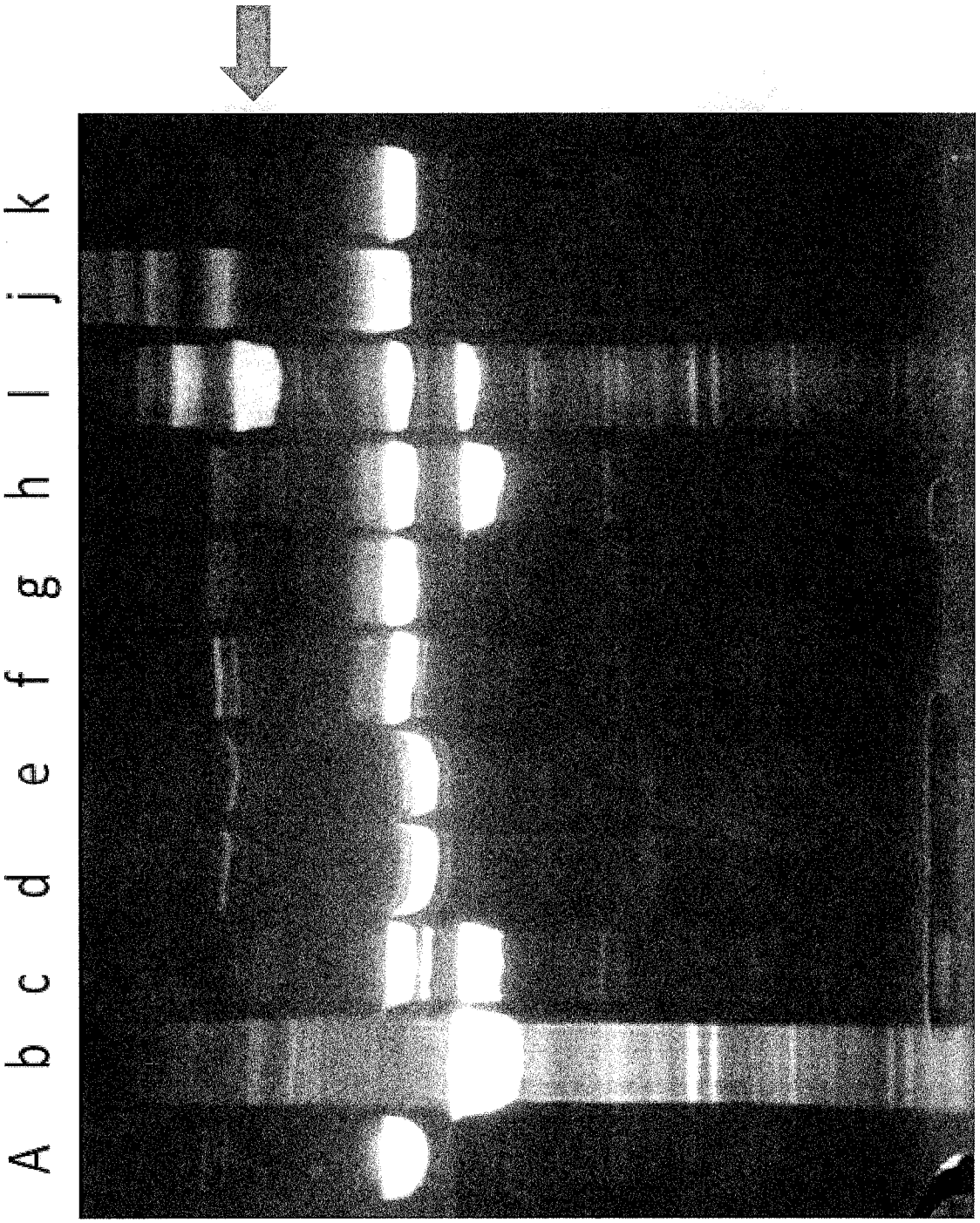
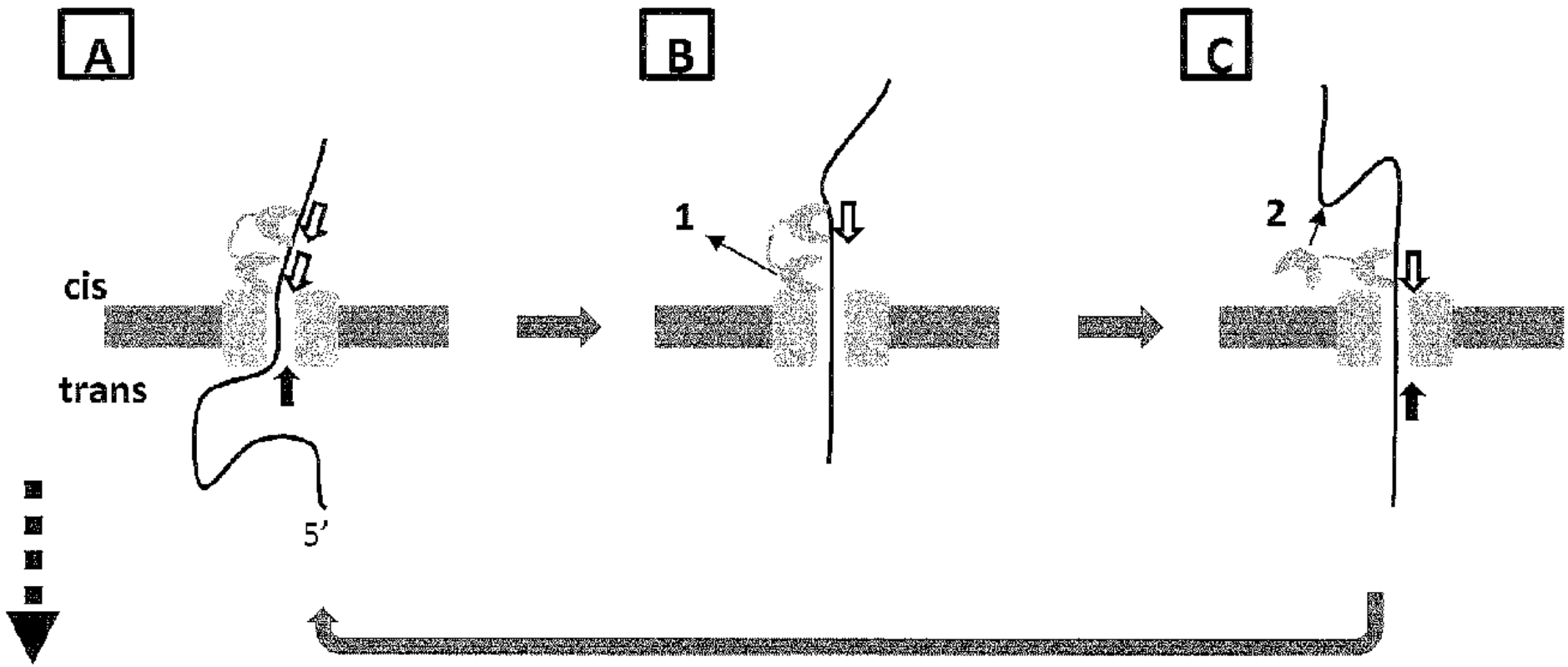


Fig. 28 28/28




A
B
C
cis
trans
5'
1
2