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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 April 2016 (21.04.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/059436 A1

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

(21) International Application Number:
PCT/GB2015/053097

(22) International Filing Date:
19 October 2015 (19.10.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

1418459.2	17 October 2014 (17.10.2014)	GB
1508270.4	14 May 2015 (14.05.2015)	GB
1517634.0	6 October 2015 (06.10.2015)	GB
PCT/GB2015/052916	6 October 2015 (06.10.2015)	GB

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

(72) Inventors: **GARALDE, Daniel Ryan**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). **HERON, Andrew John**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). **JAYASINGHE, Lakmal**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). **TURNER, Daniel John**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). **WHITE, James**; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB).

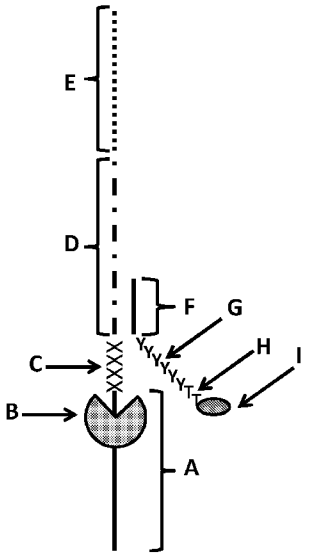
(74) Agent: **CHAPMAN, Lee Phillip**; 14 South Square, Gray's Inn, London Greater, London WC1R 5JJ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,

[Continued on next page]

(54) Title: METHOD FOR NANOPORE RNA CHARACTERISATION

Figure 3



(57) Abstract: The invention relates to a new method of characterising a target RNA polynucleotide by taking one or more measurements as the target RNA polynucleotide moves with respect to a transmembrane pore. The movement is controlled by a DNA helicase. The invention also relates to a modified RNA construct wherein the RNApolynucleotidehas been modified to increase DNA helicase binding thereto.

WO 2016/059436 A1

MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

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

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

METHOD FOR NANOPORE RNA CHARACTERISATION

Field of the invention

The invention relates to a new method of characterising a target RNA polynucleotide by taking one or more measurements as the target RNA polynucleotide moves with respect to a transmembrane pore. The movement of the target RNA polynucleotide with respect to the transmembrane pore is controlled by a DNA helicase enzyme and the target RNA polynucleotide is modified to increase DNA helicase binding thereto. The invention also relates to a modified RNA construct wherein the RNA polynucleotide has been modified to increase DNA helicase binding thereto.

Background of the invention

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

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

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

Messenger RNA provides a view of the dynamic state of an organism and the benefits and applications of direct RNA sequencing are vast, including use in health screening; for example metastasis progression in certain cancers and heart disease. Direct RNA sequencing also has applications in investigating disease resistance in crops, determining the response to stressses, for example; drought, UV and salinity and in cellular differentiation and determination during embryogenesis.

A problem which occurs in direct sequencing of RNA, particularly those of 500 nucleotides or more, is finding a suitable molecular motor which can control translocation of the RNA through a transmembrane pore. To date, a molecular motor that engages with RNA and provides consistent movement has not been shown. Consistent movement of the RNA polymer and the ability to read long segments of the polymer is desirable for characterising or sequencing the polynucleotide.

International Patent Application No. PCT/GB2014/053121 (WO 2015/056028) describes a method of characterising a target ribonucleic acid (RNA) involving forming a complementary polynucleotide and then characterising the complementary polynucleotide using a transmembrane pore. Such indirect RNA characterisation is prone to error and can result in the loss of vital information regarding, for example, methylation status of the RNA. Other important modifications can also be hidden in the conversion of RNA to cDNA.

Summary of the invention

The inventors have surprisingly demonstrated that it is possible to characterise a target RNA polynucleotide by taking one or more measurements as the target RNA polynucleotide moves with respect to a transmembrane pore under the control of a DNA helicase enzyme. Accordingly, in one embodiment, a method for increasing the ability or efficiency of an RNA to be sequenced through a pore is provided. In another embodiment, a method of generating a modified RNA capable of being sequenced through a pore with greater efficiency than the RNA in unmodified form is also provided. The invention therefore provides a method of characterising a target RNA polynucleotide comprising:

a) providing (i) an RNA polynucleotide wherein the RNA polynucleotide is modified to comprise a non-RNA polynucleotide and (ii) a DNA helicase enzyme;

b) contacting the RNA polynucleotide and DNA helicase enzyme provided in a) with a transmembrane pore such that the DNA helicase controls the movement of the RNA polynucleotide through the transmembrane pore;

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

The modification of the RNA polynucleotide to comprise a non-RNA polynucleotide (such as a polynucleotide region or sequence or construct) results in increased DNA helicase binding thereto. A "non-RNA polynucleotide" as defined herein is a polynucleotide wherein

at least one nucleotide of the polynucleotide is not a ribonucleotide, i.e. is not from RNA. The non-RNA polynucleotide may therefore comprise at least one ribonucleotide (or RNA nucleotide) but must also additionally comprise or include a non-RNA nucleotide or sequence i.e., a nucleotide or sequence of nucleotides that is not RNA. In a preferred embodiment of the invention the non-RNA polynucleotide (which may or may not comprise at least one ribonucleotide or

RNA nucleotide) comprises DNA or DNA analogue, preferably a DNA helicase binding site or a DNA adaptor. Preferably the non-RNA polynucleotide comprises a leader sequence which preferentially threads into the pore. Accordingly the non-RNA polynucletoide is read first, followed by the target RNA sequence to be characterised.

The invention provides a construct which is a "hybrid" polynucleotide comprising (i) an RNA polynucleotide and (ii) a non-RNA polynucleotide. Preferably the non-RNA polynucleotide comprises a DNA polynucleotide, wherein the DNA polynucleotide comprises or comprises only a DNA helicase binding site. Preferably the non-RNA polynucleotide further comprises a leader sequence which preferentially threads into a nanopore. More preferably the non-RNA polynucleotide further comprises a barcoding section on the polynucleotide strand. The barcoding section is preferably located between the leader sequence and the DNA helicase binding site. The barcoding section enables unambiguous identification of an analyte i.e., informing the user which of several samples is being sequenced.

The modification of a target RNA polynucleotide to comprise a non-RNA polynucleotide may involve the attachment of a non-RNA polynucleotide (possibly comprising an RNA sequence or at least one ribonucleotide) to the target RNA polynucleotide using any suitable attachment method, including one or more of the attachment methods described herein. As described herein, the attachment of a non-RNA polynucleotide to a target RNA is synonomous with the attachment of a target RNA to a non-RNA polynucleotide. Where the non-RNA polynucleotide comprises a ribonucleotide or an RNA sequence, the non-RNA polynucleotide may be attached to the target RNA polynucleotide via the ribonucleotide or RNA which is comprised within the non-RNA polynucleotide.

The non-RNA polynucleotide can be attached to the target RNA polynucleotide by a covalent bond formed between at least one reactive group on each of the target RNA polynucleotide and the non-RNA polynucleotide. The non-RNA polynucleotide can be chemically or enzymatically ligated to the RNA polynucleotide. The non-RNA

polynucleotide can additionally or alternatively be attached to the RNA polynucleotide by hybridisation and/or using one or more topoisomerases. Preferably the one or more characteristics to be determined by the method are selected from (i) length of the RNA polynucleotide, (ii) identity of the RNA polynucleotide, (iii) the sequence of the RNA polynucleotide, (iv) the secondary structure of the RNA polynucleotide and (v) whether or not the RNA polynucleotide is modified. The one or more characteristics of the RNA polynucleotide can be measured by electrical and/or optical measurement. Preferably step c) comprises measuring the current passing through the transmembrane pore as the RNA polynucleotide moves with respect to the transmembrane pore wherein the current is indicative of one or more characteristics of the RNA polynucleotide and thereby characterising the RNA polynucleotide.

The target RNA polynucleotide to be characterised may be additionally or further modified by methylation, by oxidation, by damage, with one or more proteins or with one or more labels, tags or spacers. The target RNA may contain base analogues. The RNA polynucleotide can be coupled to the membrane using one or more anchors.

Preferably the DNA helicase comprises a modification to reduce the size of an opening in the polynucleotide binding domain through which in at least one conformational state the RNA polynucleotide can unbind from the helicase. In one embodiment of the invention, the movement of the RNA polynucleotide is controlled by a series of one of more DNA helicases. The one or more helicases are a) Hel308 helicases, RecD helicases, XPD helicases or Dda helicases (b) helicases derived from any of the helicases in (a); or (c) a combination of any of the helicases in (a) and/or (b). The method may further comprise the use of one or more molecular brakes that are derived from helicases and are modified such that they bind the polynucleotide but do not function as a helicase.

The transmembrane pore can by a protein pore or a solid state pore. Preferably the transmembrane protein pore is a protein pore and is derived from a hemolysin, leukocidin, *Mycobacterium smegmatis* porin A (MspA), MspB, MspC, MspD, lysenin, CsgG, outer membrane porin F (OmpF), outer membrane porin G (OmpG), outer membrane phospholipase A, *Neisseria* autotransporter lipoprotein (NalP) and WZA.

The invention also provides a method of moving a target RNA polynucleotide with respect to a transmembrane pore when the movement is controlled by a DNA helicase enzyme, comprising:

a) providing (i) an RNA polynucleotide wherein the RNA polynucleotide is modified to comprise a non-RNA polynucleotide and a (ii)DNA helicase enzyme;

b) contacting the RNA polynucleotide and DNA helicase enzyme provided in a) with a transmembrane pore such that the DNA helicase controls the movement of the RNA polynucleotide with respect to the transmembrane pore.

The modification of the RNA polynucleotide to comprise a non-RNA polynucleotide results in increased DNA helicase binding thereto. In one embodiment of the invention the method comprises pre-binding the DNA helicase to the modified RNA polynucleotide before the contacting step.

The method of the invention provides more consistent movement of the RNA polynucleotide with respect to the transmembrane pore. The invention also provides a "hybrid" polynucleotide comprising (i) an RNA polynucleotide and (ii) a non-RNA polynucleotide. Preferably the non-RNA polynucleotide comprises a DNA polynucleotide, wherein the DNA polynucleotide comprises or comprises only a DNA helicase binding site. Preferably the non-RNA polynucleotide further comprises a leader sequence which preferentially threads into a nanopore. More preferably the non-RNA polynucleotide further comprises a barcoding section on the polynucleotide strand. The barcoding section is preferably located between the leader sequence and the DNA helicase binding site. The barcoding section enables unambiguous identification of an analyte i.e., informing the user which of several samples is being sequenced.

The invention also provides a combination of a target RNA polynucleotide and a DNA helicase in which a part of the RNA polynucleotide has been modified to interact with or bind to the DNA helicase. Preferably the RNA is modified to comprise a non-RNA polynucleotide, most preferably the non-RNA polynucleotide is DNA or a DNA analogue.

The invention also provides a kit for characterising a target RNA polynucleotide. Preferably the kit comprises a non-RNA polynucleotide, which is adapted for attachment to a target RNA to be characterised. The invention also provides an apparatus for characterising a target RNA polynucleotide in a sample.

The modified RNA constructs described herein provide a binding site for the DNA helicase enzyme. The DNA helicase is essentially "tricked" into reading the RNA polynucleotide: Once the DNA helicase has bound to the non-RNA polynucleotide it can transition along the RNA polynucleotide. The helicase may transition along the RNA in solution. The helicase may require the presence of the nanopore to facilitate the movement of the helicase along the RNA polynucleotide. The movement of the helicase along the RNA may be facilitated by a nanopore and via the application of an applied field across the nanopore.

The ability to translocate an entire RNA construct through a nanopore under the control of a DNA helicase allows characteristics of the RNA polynucleotide, such as its sequence, to be estimated with improved accuracy and speed over known methods: The methods of the present inventionare free from PCR bias, as compared with prior art methods. Modifications in RNA can be detected and non-coding isoforms and splice-variants can be correctly identified. RNA base analogues can be directely detected in the nanopore. The constructs described herein are particularly effective for achieving translocation of target RNAs of 500 nucleotides or more, for example 1000 nucleotides, 5000, 10000, 20000, 50000, 100000 or more.

In the methods of the invention the DNA helicase is essentially "tricked" into reading the target RNA sequence by virtue of the presence of a non-RNA leader sequence. Once movement of the DNA helicase is initiated by the non-RNA polynucleotide (which may comprise DNA or a DNA analogue), it can continue to move along the RNA. The methods of the invention also provide means to differentiate RNA and DNA from each other as a function of mean amplitude and range, even when the RNA and DNA sequences are the same.

Description of the Figures

Figure 1 shows a cartoon representation of methods of attaching a strand of eukaryotic RNA (shown as a dashed line) to a strand of DNA (shown as a solid line). Eukaryotic RNA has 7-methylguanosine cap at the 5' end (shown as a star shape and labelled A). Reaction Step 1 leaves the 7-methylguanosine cap in place and adds a reactive group (labelled B and shown as a square) to the 5' end of the eukaryotic RNA. The reactive group is added using a hypermethylase enzyme, for example Trimethyl guanosine synthase with modified SAdenosyl Methiamine . Step 2 shows a chemical reaction step where the DNA strand, which also has a reactive group labelled C attached and shown as a circle, is reacted with the other reactive group (B) to form a covalent bond. Step 3 removes the 7-methylguanosine cap using Tobacco acid phosphatase which results in the RNA strand labelled D. It is then possible to treat the RNA (with the 7-methylguanosine cap removed, labelled D) in a number of different ways in order to produce a strand of RNA attached to a strand of DNA (steps 4,6, 7, 8 or 11). In step 4 a reactive group (labelled B and shown as a square) is added to the 5' end of RNA strand D. Step 5 shows a chemical reaction step where the DNA strand which also has a reactive group labelled C attached and the reactive group is shown as a circle is reacted with the other reactive group (B) to form a covalent bond. Steps 6

and 7 show that a strand of DNA (labelled E) can be ligated directly onto the RNA using, for example, T4 RNA Polymerase 1, T4 RNA Polymerase 2 , Thermostable 5' App DNA/RNA ligase etc. In step 7 an enzyme (labelled G) is pre-bound to the DNA whereas in step 6 no enzyme is pre-bound. Steps 8 and 11 show the hybridisation of a DNA primer, with a leader (labelled F), to the RNA strand D. In step 11 an enzyme (labelled G) is pre-bound to the DNA primer whereas in step 8 no enzyme is pre-bound. Steps 9 and 12 show the reverse transcription of RNA strand D which results in a 3' overhang of three C's. Steps 10 and 13 show the ligation of a DNA hairpin (labelled H) to the double-stranded DNA/RNA (labelled I).

Figure 2 shows a cartoon representation of methods of attaching a strand of prokaryotic RNA (shown as a dashed line) to a strand of DNA (shown as a solid line). Reaction Step 1 uses, for example the PolyA Polymerase enzyme to add a poly(dA) tail (labelled A) onto the 3' end of the RNA strand resulting in strand B. Steps 2 and 5 show the hybridisation of a DNA primer with a leader (labelled C) to the poly(dA) region (labelled A). In step 5 an enzyme (labelled D) is pre-bound to the DNA primer whereas in Step 2 no enzyme is pre-bound. Steps 3 and 6 show the reverse transcription of RNA strand B which results in a 3' overhang of one to three C's. Steps 4 and 7 show the ligation of a DNA hairpin (labelled E) to the double-stranded DNA/RNA (labelled F). Reaction steps 8, 10 and 11 occur directly on prokaryotic RNA rather than by adding the poly(dA) region to the RNA. In step 8 a reactive group (labelled G and shown as a square) is added to the 5' end of RNA strand. Step 9 shows a chemical reaction step where the DNA strand which also has a reactive group attached (labelled H and the reactive group is shown as a circle) is reacted with the other reactive group (G) to form a covalent bond. Steps 10 and 11 show that a strand of DNA (labelled I) can be ligated directly onto the RNA using , for example, T4 RNA Polymerase 1, T4 RNA Polymerase 2 , Thermostable 5' App DNA/RNA ligase etc. In step 11 an enzyme (labelled D) is pre-bound to the DNA whereas in step 10 no enzyme is pre-bound.

Figure 3 shows a cartoon representation of the DNA/RNA strand which was translocated through an MspA nanopore using a DNA helicase (T4 Dda – E94C/A360C (SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1)) in Example 1. Region A corresponds to the 40 nucleotide poly(dT) leader (SEQ ID NO: 16) to which the DNA helicase (labelled B) binds. Region A is attached to four iSpC3 spacers (shown as X's and labelled C). Region D corresponds to the synthetic RNA region (SEQ ID NO: 17). Region E corresponds to the variable length poly(U). Region F is the DNA (SEQ ID NO: 18) which is hybridised to Region D. Attached to the DNA (SEQ ID NO: 18) is six iSp18 spacers (shown

as Y's and labelled G) two thymines (shown as T's and labeled H) and a 3' cholesterol TEG (labelled I).

Figure 4 shows a 5% PAGE TBE BioRad Criterion Gel (run at 140 V for 40 minutes) after the RNA poly(U) polymerase extension step (Example 1 step 1.1). Lane 1 shows a 100 bp TriDye ladder. Lane 2 shows synthetic DNA/RNA 1 (SEQ ID NO: 16 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 17) before polymerisation. Lane 3 shows the poly(U) polymerisation mix. Lane 4 shows the purified Sample 1 which contains poly(U) extended DNA/RNA 2. Arrow X corresponds to the non-extended DNA/RNA 1 and arrow Y corresponds to the poly(U) extended DNA/RNA 2.

Figure 5 shows an example trace of a helicase controlled DNA movement (y-axis = current (pA), x-axis = time (s)) where the DNA helicase (T4 Dda – E94C/A360C (SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1)) controlled the movement of DNA/RNA 2 (cartoon representation shown in Figure 3). Region 1 corresponds to the poly(dT) leader (SEQ ID NO: 16), region 2 corresponds to the iSpC3 spacers (the spacers allowed a larger amount of current to flow through the nanopore than the DNA or RNA regions), region 3 corresponds to the RNA sequence (SEQ ID NO: 17) and region 4 corresponds to the variable length poly(U) RNA region which was added in Example 1 step 1.1.

Figure 6 shows a possible method of attaching dsDNA onto a strand of RNA using a topoisomerase. In this cartoon figure, dsDNA is attached to a strand of RNA which has a free hydroxyl at its 5' end (the 7-methylguanosine cap and 5'phosphate present at the 5' end of eukaryotic RNA would need to be removed with, for example a phosphatase such as antarctic phosphatase or alkaline phosphatase). The vaccinia topoisomerase binds onto the dsDNA at the sequence shown (step 1). The red arrow highlights that there is a nick in the lower strand of the dsDNA opposite the second thymine. Once the topoisomerase has bound it cuts the upper strand of the DNA, after the second thymine, and remains bound to the dsDNA (step 2). The topoisomerase which is bound to the dsDNA was then incubated with RNA which has a free 5' hydroxyl (step 3). The topoisomerase then joins the RNA to the dsDNA.

Figure 7 shows another possible method of attaching dsDNA onto a strand of RNA using a topoisomerase. In this cartoon figure, the dsDNA is attached to a strand of RNA using a single-stranded region of DNA to hybridise to the RNA and assist in the RNA attachment. The vaccinia topoisomerase binds onto the dsDNA at the sequence shown (step 1). The lower strand of DNA does not have a nick in it. Once the topoisomerase has bound it cuts the upper

strand of the DNA only, after the second thymine, and remains bound to the dsDNA/ssDNA (step 2). The topoisomerase which is bound to the dsDNA was then incubated with RNA which has a free 5' hydroxyl (step 3). The topoisomerase then joins the RNA to the dsDNA. The ssDNA region assists in attracting the complementary RNA sequence to the point of attachment to the DNA.

Figure 8 shows a 10% PAGE TBE-Urea denaturing criterion gel run at 140 V for 60 minutes which includes samples from Example 2. Lane 1 shows a 100 bp TriDye ladder. Lane 2 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (0.5X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the absence of T4 DNA ligase. Lane 3 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (4X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the absence of T4 DNA ligase. Lane 4 shows unreacted DNA oligo 1 (SEQ ID NO: 21) and the DNA splint (SEQ ID NO: 20) mixed together as a control. Lane 5 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (0.5X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the presence of T4 DNA ligase. Lane 6 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (1X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the presence of T4 DNA ligase. Lane 7 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (2X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the presence of T4 DNA ligase. Lane 8 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (4X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the presence of T4 DNA ligase. Lane 9 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (0.5X concentration, SEQ ID NO: 20) after they have undergone the ligation reaction described in Example 2 in the presence of T4 DNA ligase and they have been incubated with further DNA splint (4.5X concentration of SEQ ID NO: 20) after the ligation step. Lane 10 shows the DNA oligo 1 (1X concentration, SEQ ID NO: 21), the RNA oligo 1 (1X concentration, SEQ ID NO: 19) and the DNA splint (1X concentration, SEQ ID NO: 20) after they have undergone the ligation

reaction described in Example 2 in the presence of T4 DNA ligase and they have been further heat treated and exposed to ExoI. Lane 11 shows the DNA oligo 1 (SEQ ID NO: 21) as a control. Lane 12 shows the RNA oligo 1 as a control. Lane 13 shows unreacted RNA oligo 1 (SEQ ID NO: 19) and the DNA splint (SEQ ID NO: 20) mixed together as a control. The band labelled 1 corresponds to RNA oligo 1 (SEQ ID NO: 19). The band labelled 2 corresponds to RNA oligo 1 (SEQ ID NO: 19) hybridised to the splint (SEQ ID NO: 20). The band labelled 3 corresponds to the splint (SEQ ID NO: 20). The band labelled 4 corresponds to the DNA oligo 1 (SEQ ID NO: 21). The band labelled 5 corresponds to DNA oligo 1 (SEQ ID NO: 21) hybridised to the splint (SEQ ID NO: 20). The region labelled A corresponds to the ligated substrate in the presence of the splint (DNA oligo 1 ligated to RNA oligo 1, hybridised to splint). The region labelled B corresponds to the ligated substrate in the absence of the splint (DNA oligo 1 ligated to RNA oligo 1).

Figure 9 shows a 5% PAGE TBE BioRad Criterion Gel run at 140 mV for 60 minutes, (A) before SYBR stain and (B) after SYBR stain, which includes samples from Example 3. Lanes 1 and 7 show a TriDye 1 kB ladder. Lanes 2 and 8 show the product produced from Example 3A (DNA X1 (SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end) reacted with firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail) in the presence of DNA splint X1 (SEQ ID NO: 24)). Lanes 3 and 9 show the product produced from Example 3B (DNA X2 (SEQ ID NO: 25 which had a CY3 attached to its 5' end and had a 3AzideN attached to the 3' end) reacted with firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail) in the presence of DNA splint X1 (SEQ ID NO: 24)). Lanes 4 and 10 show DNA X2 (SEQ ID NO: 25 which had a CY3 attached to its 5' end and had a 3AzideN attached to the 3' end) mixed with firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). Lanes 5 and 11 show DNA X1 (SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end). Lanes 6 and 12 show firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). The band labelled 1 corresponds to RNA X1 with and without DNA attached. Band 2 corresponds to unreacted

DNA X1 (SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end). Band 3 corresponds to unreacted DNA X2 (SEQ ID NO: 25 which had a CY3 attached to its 5' end and had a 3AzideN attached to the 3' end). Only DNA which contained the CY3 label was visible on the non-SYBR stained gel.

Figure 10 shows a cartoon representation of the DNA/RNA strand produced in Example 3A which is translocated through an MspA nanopore using a DNA helicase (T4 Dda – E94C/A360C (SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1)). Region A corresponds to the DNA leader (SEQ ID NO: 22) to which the DNA helicase (labelled B) binds. Region A is attached to four iSp18 spacers (shown as X's and labelled C). Region D corresponds to a second DNA sequence (SEQ ID NO: 23). Region E corresponds to the firefly luciferase mRNA with a 5'-hexynl-G region (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). Regions D and E are attached by click chemistry. The reacted azide and hexynl groups are represented by a box (labelled F). Region G is the DNA (SEQ ID NO: 18) which is hybridised to Region D. Attached to the DNA (SEQ ID NO: 18) is six iSp18 spacers (shown as Y's and labelled H) two thymines (shown as T's and labeled I) and a 3' cholesterol TEG (labelled J).

Figure 11 shows an example trace of a helicase controlled DNA movement (y-axis = current (pA), x-axis = time (s)) where the DNA helicase (T4 Dda – E94C/A360C (SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1)) controlled the movement of the DNA/RNA product produced by Example 3A (cartoon representation shown in Figure 3). Region 1 corresponds to the DNA leader that has not ligated onto the RNA (DNA X1), region 2 corresponds to SEQ ID NO: 22 attached to the four iSp18 spacers (the spacers allow a larger amount of current to flow through the nanopore than the DNA or RNA regions), region 3 corresponds to a DNA sequence (SEQ ID NO: 23) and region 4 corresponds to the Firefly luciferase mRNA region (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). The peak in current labelled with a * corresponds to the translocation of the click linkage which is made when the DNA and RNA are joined.

Figure 12 shows a cartoon representation of a method of attaching a strand of eukaryotic RNA (shown as a dashed line) to a strand of DNA (shown as a solid line). The eukaryotic RNA has a 7-methylguanosine cap which runs in the opposite orientation to the rest of the RNA strand (i.e., the 5' end comprises reversed bases, shown as a star shape and

labelled A). The eukaryotic RNA is ligated to a strand of DNA (labelled B) which comprises a region of reversed DNA bases (shown as a random sequence of n's and labelled C). In the region of reversed bases the bases also run in the opposite direction. This illustrates that the two regions of reversed bases on the RNA and DNA can be ligated together.

Figure 13 shows a cartoon representation of the RNA/cDNA construct which was produced in Example 4. Region B corresponds to SEQ ID NO: 30. Region C corresponded to the last three adenines in SEQ ID NO: 30. Region D corresponds to the SEQ ID NO: 29 in both the 3T and 10T hairpins. Region E corresponds to four iSpC3 spacers (shown as X's). Region F corresponds to SEQ ID NO 27 in the 3T hairpin and SEQ ID NO: 28 in the 10T hairpin. Region G corresponds to the last three thymines in SEQ ID NO: 27 or 28. Region H corresponds to the cDNA which was produced during the reverse transcription of RNA strand SEQ ID NO: 30.

Figure 14 shows a 10% PAGE TBE-Urea denaturing gel (run at 140 V for 60 minutes) showing various samples from Example 4 before and after hairpin ligation and reverse transcription. Lane 1 shows a TriDye ladder. Lane 2 shows the RNA strand used in Example 4 (SEQ ID NO: 30) before ligation or reverse transcription. Lane 3 shows the 10T hairpin control (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 28) before ligation or reverse transcription. Lane 4 shows the RNA strand (SEQ ID NO: 30) after reverse transcription using a primer. Lane 5 shows the RNA strand (SEQ ID NO: 30) after ligation to the 3T hairpin (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 27). Lane 6 shows the RNA strand (SEQ ID NO: 30) after ligation to the 3T hairpin (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 27) and subsequent reverse transcription. Lane 7 shows the RNA strand (SEQ ID NO: 30) after ligation to the 10T hairpin (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 28). Lane 8 shows the RNA strand (SEQ ID NO: 30) after ligation to the 10T hairpin (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 28) and subsequent reverse transcription. Lane 9 shows a control experiment where the RNA strand and the 10T hairpin were incubated

together in the absence of a ligase. Arrow A corresponds to the RNA strand used in Example 4 (SEQ ID NO: 30).

Figure 15 shows a cartoon representation of the DNA/RNA/cDNA strand produced in Example 5 which was translocated through an MspA nanopore using a DNA helicase (T4 Dda – E94C/C109A/C136A/A360C (SEQ ID NO: 14 with mutations E94C/C109A/C136A/A360C and then (ΔM1)G1)). Region A corresponds to the DNA leader (SEQ ID NO: 22) to which the DNA helicase (labelled B) binds. Region A is attached to four iSp18 spacers (shown as X's and labelled C). Region D corresponds to a second DNA sequence (SEQ ID NO: 23). Region E corresponds to the firefly luciferase mRNA with a 5'-hexynl-G region (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). Regions D and E are attached by click chemistry. The reacted azide and hexynl groups are represented by a box (labelled F). Region G is the DNA (SEQ ID NO: 18) which is hybridised to Region D. Attached to the DNA (SEQ ID NO: 18) is six iSp18 spacers (shown as Y's and labelled H) two thymines (shown as T's and labeled I) and a 3' cholesterol TEG (labelled J). Region K corresponds to the last three adenines in FLuc mRNA (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). Region L corresponds to the SEQ ID NO: 29 in the 10T hairpin. Region M corresponds to four iSpC3 spacers (shown as X's). Region N corresponds to SEQ ID NO: 28 in the 10T hairpin. Region O corresponds to the last three thymines in SEQ ID NO: 28. Region P corresponds to the cDNA which was produced during the reverse transcription of the RNA strand RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail.

Figure 16 shows an example trace of a helicase controlled DNA movement (y-axis = current (pA), x-axis = time (s)) where the DNA helicase (T4 Dda – E94C/C109A/C136A/A360C (SEQ ID NO: 14 with mutations E94C/C109A/C136A/A360C and then (ΔM1)G1)) controlled the movement of the DNA/RNA/cDNA product produced by Example 5 (cartoon representation shown in Figure 15). Region 1 corresponds to the DNA leader (DNA X1), region 2 corresponds to the Firefly luciferase mRNA region (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail), region 3 corresponds to the four iSpC3 spacers, region 4 corresponds to the polyT region of the 10T hairpin, region 5 corresponds to the cDNA which was produced by reverse transcription of the mRNA.

Figure 17 shows a 5% PAGE TBE-Urea denaturing gel (run at 140 V for 60 minutes) showing various samples from Example 5 before and after decapping and ligation. Lane 1 shows a TriDye ladder. Lane 2 shows the capped RNA strand (SEQ ID NO: 30 which has a 7-methylguanosine cap connected to the 5' end of the strand by a 5' to 5' triphosphate linkage). Lane 3 shows the non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG). Lane 4 shows the RNA strand (SEQ ID NO: 30) after ligation to the non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG). Arrow A corresponds to the RNA strand (SEQ ID NO: 30). Arrow B corresponds to the non-RNA polynucleotide. Arrow C corresponds to the ligated product where the RNA strand has been ligated to the non-RNA polynucleotide.

Figure 18 illustrates consensus current levels from RNA with and without modified bases. This figure shows that modified bases affect several consecutive current levels.

Figure 19 shows a DNA helicase controlled translocation of an RNA-DNA 2D strand (RNA-sense and DNA-antisense) through the nanopore. Figure 19 demonstrates the different mean amplitude and range observed for RNA *versus* DNA.

Figure 20 shows an example trace of a helicase controlled DNA movement (y-axis = current (pA), x-axis = time (s)) where the DNA helicase (Hel308Mbu-E284C/S615C (SEQ ID NO: 8 with mutations E284C/S615C)) controlled the movement of the DNA/RNA product produced by Example 8.

Figure 21 shows an example trace of a helicase controlled DNA movement (y-axis = current (pA), x-axis = time (s)) through a lysenin mutant where the DNA helicase (and T4 Dda – E94C/A360C (0.36 µl, 3.8 µM) , SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1) controlled the movement of the DNA/RNA product produced by Example 3.

Figure 22 shows an example trace of a helicase controlled DNA movement (y-axis = current (pA), x-axis = time (s)) through a CsgG mutant pore (CsgG-Eco-(Y51T/F56Q)-StrepII(C))9 (SEQ ID NO: 44 with mutations Y51T/F56Q where StepII(C) is SEQ ID NO: 45 and is attached at the C-terminus))where the DNA helicase (Hel308Mbu-E284C/S615C (SEQ

ID NO: 8 with mutations E284C/S615C)) controlled the movement of the DNA/RNA product produced by Example 5.

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 Hel308 Mbu.

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

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

SEQ ID NO: 11 shows the amino acid sequence of Hel308 Mhu.

SEQ ID NO: 12 shows the amino acid sequence of TraI Eco.

SEQ ID NO: 13 shows the amino acid sequence of XPD Mbu.

SEQ ID NO: 14 shows the amino acid sequence of Dda 1993.

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

SEQ ID NO: 16 shows a polynucleotide sequence used in Example 1.

SEQ ID NO: 17 shows a polynucleotide sequence used in Example 1.

SEQ ID NO: 18 shows a polynucleotide sequence used in Example 1 and 3.

SEQ ID NO: 19 shows a polynucleotide sequence used in Example 2. This sequence has a 5' phosphate group.

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

SEQ ID NO: 21 shows a polynucleotide sequence used in Example 2. This sequence has a 5' CY3 group.

SEQ ID NO: 22 shows a polynucleotide sequence used in Example 3.

SEQ ID NO: 23 shows a polynucleotide sequence used in Example 3. This sequence has a 3' AzideN group.

SEQ ID NO: 24 shows a polynucleotide sequence used in Example 3.

SEQ ID NO: 25 shows a polynucleotide sequence used in Example 3. This sequence has a 3' AzideN group and a 5' CY3 group.

SEQ ID NO: 26 shows the open reading frame of a polynucleotide sequence used in Examples 3 and 5. This sequence has 5'-hexynl group attached to the first G in the sequence.

SEQ ID NO: 27 shows a polynucleotide sequence used in Examples 4 and 5.

SEQ ID NO: 28 shows a polynucleotide sequence used in Example 4.

SEQ ID NO: 29 shows a polynucleotide sequence used in Examples 4 and 5.

SEQ ID NO: 30 shows a polynucleotide sequence used in Examples 4 and 6.

SEQ ID NO: 31 shows a polynucleotide sequence used in Example 6.

SEQ ID NO: 32 shows a polynucleotide sequence used in Example 6.

SEQ ID NO: 33 shows a sequence used to illustrate homopolymer read.

SEQ ID NO: 34 shows a sequence used to illustrate homopolymer read.

SEQ ID NO: 35 shows a polynucleotide sequence used in Example 8.

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

SEQ ID NO: 37 shows a polynucleotide sequence used in Example 8.

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

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

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

SEQ ID NO: 41 shows the polynucleotide sequence encoding the lysenin monomer.

SEQ ID NO: 42 shows the amino acid sequence of the the lysenin monomer.

SEQ ID NO: 43 shows the codon optimised polynucleotide sequence encoding the wild-type CsgG monomer from *Escherchia coli* Str. K-12 substr. MC4100. This monomer lacks the signal sequence.

SEQ ID NO: 44 shows the amino acid sequence of the mature form of the wild-type CsgG monomer from *Escherchia coli* Str. K-12 substr. MC4100. This monomer lacks the signal sequence. The abbreviation used for this CsgG = CsgG-Eco.

SEQ ID NO: 45 shows the amino acid sequence of StepII(C).

SEQ ID NOs: 46 to 64 are polynucleotide sequences described in the description.

SEQ ID NOs: 65 and 66 are polynucleotide sequences used in the Examples.

Detailed description of the invention

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

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

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

Characterising a target RNA polynucleotide

The method of the invention involves characterising a target RNA polynucleotide. The RNA polynucleotide is delivered to a transmembrane pore and the pore is used to characterise the RNA polynucleotide. The invention provides a method of characterising a target ribonucleic acid (RNA) polynucleotide by taking one or more measurements as the target RNA polynucleotide moves with respect to a transmembrane pore under the control of a DNA helicase enzyme.

Since the transmembrane pore is capable of detecting a single molecule of the target polynucleotide, there is no need for amplification of the target RNA polynucleotide. The method typically does not comprise polymerase chain reaction (PCR) or reverse transcription PCR (RT-PCR). This considerably reduces the amount of workflow needed to characterise a target RNA polynucleotide. It also avoids any biases and artifacts introduced by PCR.

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

{ii,iii}, {ii,iv}, {ii,v}, {iii,iv}, {iii,v}, {iv,v}, {i,ii,iii}, {i,ii,iv}, {i,ii,v}, {i,iii,iv}, {i,iii,v}, {i,iv,v}, {ii,iii,iv}, {ii,iii,v}, {ii,iv,v}, {iii,iv,v}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iv,v}, {i,iii,iv,v}, {ii,iii,iv,v} or {i,ii,iii,iv,v}. Different combinations of (i) to (v) may be measured including any of those combinations listed above. The method of the invention preferably comprises estimating the sequence of or sequencing the RNA polynucleotide.

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

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

For (iii), the sequence of the RNA 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 RNA polynucleotide to be distinguished.

For (v), the presence or absence of any modification may be measured. The method preferably comprises determining whether or not the polynucleotide is modified by methylation, by oxidation, by damage, with one or more proteins or with one or more labels, tags or spacers. Specific modifications will result in specific interactions with the pore which can be measured using the methods described below. For instance, methylcyotsine may be distinguished from cytosine on the basis of the current flowing through the pore during its interaction with each ribonucleotide. The methods of the invention can be used to distinguish between RNA and DNA even in a single sample: RNA and DNA can be differentiated from

each other as a function of mean amplitude and range even when the RNA and DNA sequences are the same.

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

The methods are typically carried out in the presence of any charge carriers, such as metal salts, for example alkali metal salt, halide salts, for example chloride salts, such as alkali metal chloride salt. Charge carriers may include ionic liquids or organic salts, for

example tetramethyl ammonium chloride, trimethylphenyl ammonium chloride, phenyltrimethyl ammonium chloride, or 1-ethyl-3-methyl imidazolium chloride. In the exemplary apparatus discussed above, the salt is present in the aqueous solution in the chamber. Potassium chloride (KCl), sodium chloride (NaCl), caesium chloride (CsCl) or a mixture of potassium ferrocyanide and potassium ferricyanide is typically used. KCl, NaCl and a mixture of potassium ferrocyanide and potassium ferricyanide are preferred. The charge carriers may be asymmetric across the membrane. For instance, the type and/or concentration of the charge carriers may be different on each side of the membrane.

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

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

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

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 helicase or 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 helicase or 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+}.

Target RNA

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

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

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

monophosphate (UMP), cytidine monophosphate (CMP), 5-methylcytidine monophosphate, 5-methylcytidine diphosphate, 5-methylcytidine triphosphate, 5-hydroxymethylcytidine monophosphate, 5-hydroxymethylcytidine diphosphate and 5-hydroxymethylcytidine triphosphate. The nucleotides are preferably selected from AMP, TMP, GMP, CMP and UMP.

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

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

RNA is an extremely diverse molecule. The target RNA polynucleotide may be any naturally occurring or synthetic ribonucleotide molecule, e.g., RNA, messenger RNA (mRNA), Ribosomal RNA (rRNA), Heterogenous nuclear RNA (hnRNA), Transfer RNA (tRNA), Transfer-messanger RNA (tmRNA), Micro RNA (miRNA), Small nuclear RNA (snRNA), Small nucleolar RNA (snoRNA), Signal recognition particle (SRP RNA), SmY RNA, Small Cajal body-speicifc RNA (scaRNA), Guide RNA (gRNA), Spliced Leader RNA (SL RNA), Antisense RNA (asRNA), Long noncoding RNA (lncRNA), Piwi-interacting RNA (piRNA), Small interfering RNA (siRNA), Trans-acting siRNA (tasiRNA), Repeat associated siRNA (rasiRNA), Y RNA, viral RNA or chromosomal RNA, all of which where appropriate may be single, double or triple stranded.

The target RNA polynucleotide is preferably messenger RNA (mRNA). The target mRNA may be an alternate splice variant. Altered amounts (or levels) of mRNA and/or alternate mRNA splice variants may be associated with diseases or conditions.

Alternatively the target RNA polynucleotide is a microRNA (or miRNA). One group of RNAs which are difficult to detect in low concentrations are micro-ribonucleic acids (micro-RNA or miRNAs). miRNAs are highly stable RNA oligomers, which can regulate protein production post-transcriptionally. They act by one of two mechanisms. In plants, miRNAs have been shown to act chiefly by directing the cleavage of messenger RNA, whereas in animals, gene regulation by miRNAs typically involves hybridisation of miRNAs to the 3' UTRs of messenger RNAs, which hinders translation (Lee *et al.*, Cell 75, 843-54 (1993); Wightman *et al.*, Cell 75, 855-62 (1993); and Esquela-Kerscher *et al.*, Cancer 6, 259-69 (2006)). miRNAs frequently bind to their targets with imperfect complementarity. They

have been predicted to bind to as many as 200 or more gene targets each and to regulate more than a third of all human genes (Lewis *et al.*, Cell 120, 15-20 (2005)).

Suitable miRNAs for use in the invention are well known in the art. For instance, suitable miRNAs are stored on publically available databases (Jiang Q., Wang Y., Hao Y., Juan L., Teng M., Zhang X., Li M., Wang G., Liu Y., (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res.). The expression level of certain microRNAs is known to change in tumours, giving different tumour types characteristic patterns of microRNA expression (Rosenfeld, N. *et al.*, Nature Biotechnology **26**, 462-9 (2008)). In addition, miRNA profiles have been shown to be able to reveal the stage of tumour development with greater accuracy than messenger RNA profiles (Lu *et al.*, Nature 435, 834-8 (2005) and Barshack *et al.*, The International Journal of Biochemistry & Cell Biology 42, 1355-62 (2010)). These findings, together with the high stability of miRNAs, and the ability to detect circulating miRNAs in serum and plasma (Wang *et al.*, Biochemical and Biophysical Research Communications 394, 184-8 (2010); Gilad *et al.*, PloS One 3, e3148 (2008); and Keller *et al.*, Nature Methods 8, 841-3 (2011)), have led to a considerable amount of interest in the potential use of microRNAs as cancer biomarkers. For treatment to be effective, cancers need to be classified accurately and treated differently, but the efficacy of tumour morphology evaluation as a means of classification is compromised by the fact that many different types of cancer share morphological features. miRNAs offer a potentially more reliable and less invasive solution.

The use of mRNAs and miRNAs to diagnose or prognose diseases or conditions are discussed in more detail below.

Any number of RNA's can be investigated. For instance, the method of the invention may concern determining the presence, absence or one or more characteristics of 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 50, 100 or more RNA molecules.

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

The target RNA polynucleotide can be any length. For example, the RNA 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 ribonucleotides in length. The target RNA can be 1000 or more ribonucleotides, 5000 or more ribonucleotides in length or 100000 or more ribonucleotides in length. The whole or only part of the target RNA may be characterised using this method. The part of the RNA to be sequenced preferably comprises all of the

target molecule, but may for example be less than the entire molecule, e.g., between 4 bases and 1kb, e.g., 4 to 100 bases.

The target RNA polynucleotide is typically present in or derived from any suitable sample. The invention is typically carried out on a sample that is known to contain or suspected to contain the target RNA polynucleotide. Alternatively, the invention may be carried out on a sample to confirm the identity of one or more target RNAs 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 target RNA polynucleotide may be derived from a eukaryotic cell or may be derived from a virus using a eukaryotic cell's transcription machinery. 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 or 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. The target RNA polynucleotide is typically extracted from the sample before it is used in the method of the invention. RNA extraction kits are commercially available from, for instance, New England Biolabs® and Invitrogen®.

Modification of the target RNA

The modification to the RNA polynucleotide may be any modification that facilitates DNA helicase binding and/or causes or has the effect of increased DNA helicase binding to the modified RNA polynucleotide. The term "binding" as used herein refers to affinity, or probability that the DNA helicase and substrate polynucleotide will be bound at any given time. Biochemically, this increase in affinity could be caused by an increase of the "on rate" or rate of binding, or a decrease of the "off rate" or rate of unbinding, or both an increase in "on rate" and a decrease in "off rate."

The modification of the RNA polynucleotide may increase the affinity of the DNA helicase for the modified RNA by at least 10%, preferably at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% Most preferably the modification of the RNA polynucleotide increases the affinity of the DNA helicase by 95% or more. Faciliated DNA helicase binding to the modified RNA is defined as a situation where DNA helicase binds more easily to the modified RNA polynucleotide as compared to non-modified or unmodified RNA polynucleotide. Increased DNA helicase binding to the modified RNA polynucleotide is defined as an amount or level of DNA helicase binding that is greater than, or more than, the amount or level of DNA helicase binding that is observed for non-modified or unmodified RNA polynucleotide i.e., an RNA that has not been modified in accordance with the modification methods of the invention. The level of binding of DNA helicase to a target RNA polynucleotide can be easily tested using routine methods which are known and routine to one of skill in the art.

The target RNA polynucleotide is modified to comprise a non-RNA polynucleotide, such as a polynucleotide region or sequence or construct. At least one nucleotide of the non-RNA polynucleotide is not RNA. The non-RNA polynucleotide may therefore comprise a ribonucleotide or RNA nucleotide but must also comprise or include a non-RNA nucleotide or sequence i.e., a nucleotide or sequence that is not RNA. The target RNA polynucleotide is modified by the addition or attachment of the target RNA to a non-RNA polynucleotide (which may or may not comprise a ribonucleotide or an RNA nucleotide) to form a construct of the invention. The addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that the interaction between the DNA helicase and the modified RNA construct is increased i.e., as compared with interaction that occurs between DNA helicase and RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Additionally or alternatively, the addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that the specificity of DNA helicase for the

modified RNA construct is increased i.e., as compared to the specificity of the DNA helicase for the RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Additionally or alternatively, the addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that DNA helicase binding to the modified RNA construct is facilitated and/or DNA helicase binding to the modified RNA construct is increased i.e., as compared with binding that occurs between DNA helicase and RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Additionally or alternatively, the addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that the DNA helicase binds more efficiently or more strongly to the modified RNA construct and is less likely to disengage from the modified construct i.e., as compared with binding that occurs between DNA helicase and RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Preferably the modification of the RNA polynucleotide decreases the unbinding of the DNA helicase from the RNA polynucleotide by at least 10%, preferably at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90%. Most preferably the modification of the RNA polynucleotide decreases the unbinding of the DNA helicase from the RNA polynucleotide by 95% or more.

Non-RNA polynucleotide

The target RNA polynucleotide sequence is modified to comprise a non-RNA polynucleotide. The target RNA is attached to a non-RNA polynucleotide. The target RNA is preferably covalently attached to a non-RNA polynucleotide. The non-RNA polynucleotide must comprise at least one nucleotide which is not a ribonucleotide, i.e. which is not from RNA. The non-RNA polynucleotide may additionally comprise a ribonucleotide or RNA but it must also comprise or include at least one non-RNA nucleotide i.e., a nucleotide that is not RNA. Typically the non-RNA polynucleotide which comprises RNA comprises less than 20 RNA nucleotides such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, or 19 RNA nucleotides. The non-RNA polynucleotide may therefore be a "hybrid" polynucleotide comprising, for example, RNA and another polynucleotide such as DNA or a DNA analogue. The non-RNA may also include DNA spacers etc. The skilled person will be aware that any of the attachment methods described as suitable for making a modified RNA construct of the invention are equally suitable for making the non-RNA polynucleotide, wherein two or more types of nucleic acid sequence may be combined.

Preferably the non-RNA polynucleotide comprises a leader sequence. The leader sequence preferentially threads into the pore.

Preferably the target RNA polynucleotide is modified by attachment of a non-RNA leader sequence to the RNA. The leader sequence facilitates the characterisation method of the invention. The leader sequence is designed to preferentially thread into the pore and thereby facilitate the movement of polynucleotide through the pore. The leader sequence can also be used to link the target RNA polynucleotide to the one or more anchors as discussed below. The leader sequence may be linked to the target RNA polynucleotide.

The leader sequence typically comprises a polymer region. The polymer region is preferably negatively charged. The polymer is preferably a polynucleotide, such as DNA, a modified polynucleotide (such as abasic DNA), PNA, LNA, polyethylene glycol (PEG) or a polypeptide. The leader sequence preferably comprises a single stranded polynucleotide.

The single stranded leader sequence may comprise a single strand of DNA, iSpC3, a poly dT section or a poly dC section. The leader sequence preferably comprises one or more spacers. The leader sequence is preferably part of a Y adaptor as defined below.
The leader sequence can be any length, but is typically 10 to 200 nucleotides in length., In one embodiment of the invention the non-RNA leader sequence comprises iSpC3. Most preferably the non-RNA leader sequence is an iSpC3 or repeating sequences of C and A (e.g., 4x(9C's and 1A)) of approx 40 nucleotides in length. The length of the leader typically depends on the transmembrane pore used in the method.

Preferably the non-RNA polynucleotide comprises a region to which a DNA helicase is capable of binding (a DNA helicase binding site) or a DNA adaptor. The target RNA polynucleotide may be modified to comprise a DNA binding site for the DNA helicase which controls movement of the target RNA through a transmembrane pore. As used herein the terms "DNA helicase binding site" includes a DNA or DNA analogue sequence of sufficient size/length to allow one or more DNA helicases to bind thereto. The length of the binding site depends on the number of helicases that should bind thereto. The region to which a DNA helicase is capable of binding is preferably a polynucleotide such as DNA, a modified polynucleotide (such as abasic DNA), PNA, LNA, polyethylene glycol (PEG) . Preferably the DNA helicase binding site is a single stranded, non hybridised region. The region may correspond to the leader sequence. Alternatively, the region may be distinct from the leader sequence. The DNA helicase may help to control the movement of the RNA polynucleotide through the pore as discussed in more detail below.

Preferably the non-RNA polynucleotide is further provided with a blocking site or blocking molecule, which may be located adjacent or in close proximity to the DNA binding site, at the end opposite to that which the one or more helicases are to be moved. The blocking molecule prevents backward movement of the helicase and prevents it slipping off the construct.

The non-RNA polynucleotide construct for attachment to the target RNA sequence preferably comprises: portion (i) a polymer of 5 or more charged units which preferably provides for capture of the target polynucleotide by a pore; and/or portion (ii) a blocking-strand hybridisation site of approximately 20 nucleotides in length whch ; and/or portion (iii) a DNA-helicase binding site of 1 or more non-RNA nucleotides, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, or 50 non-RNA nucleotides; and/or portion (iv) a stalling chemistry of 1 or more units e.g., 1, 2, 3, 4, 5, 6, 7, 8,9 ,10 or more units, such as Sp18, as described in WO2014/135838 which is incorporated by reference herein; and/or portion (v) a tether hybridisation site of approximately 30 nucleotides in length; and/or portion (vi) a sequence that facilitates ligation of the non-RNA polynucleotide to the RNA polynucleotide. The total length of the non-RNA polynucleotide for attachment to a target RNA sequence may therefore comprise approximately 50 to 200 nucleotides.

For example, in one embodiment, a non-RNA polynucleotide may comprise at least one of: (i) a polymer of 5 or more charged units; (ii) a blocking-strand hybridisation site of approximately 20 nucleotides in length; (iii) a DNA-helicase binding site of 1 or more non-RNA nucleotides, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, or 50 nucleotides; (iv) a stalling chemistry of 1 or more units e.g., 1, 2, 3, 4, 5, 6, 7, 8,9 ,10 or more units, such as Sp18, as described in WO2014/135838 which is incorporated by reference herein; (v) a tether hybridisation site of approximately 30 nucleotides in length; and/or (vi) a sequence that facilitates ligation of the non-RNA polynucleotide to the RNA polynucleotide, as described in the preceding sections.

Each of (i) to (vi) is discussed in more detail below:

Portion (i)

Portion (i) of the non-RNA polynucleotide is preferably a polymer with net negative charge. The polymer may be any of those discussed above for the leader sequence. Preferably the polymer lacks nucleobases or lacks nucleosides. The polymer may be any of the spacers discussed below. Representative examples of sequences that meet these criteria are:

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 16)
CCCCCCCCCACCCCCCCCCACCCCCCCCCACCCCCCCCCA (SEQ ID NO: 46)
6666666666666666666666666666666666666666
7777777777777777777777777777777777777777
8888888888888888888888888888888888888888

where 6 = 1,2-dedeoxy nucleotide monophosphate
where 7 = n-propylene phosphate (Spacer 3 groups)
where 8 = PEG3 phosphate (Spacer 9 groups)

Portion (ii)

Portion (ii) of the non-RNA polynucleotide is any polynucleotide sequence which allows hybridisation of a blocking strand and can include DNA or RNA or analogues like PNA, GNA, TNA, BNA, LNA or morpholino. Representative examples of sequences that meet these criteria are:

ACTCGCAGATCATTACGATC (SEQ ID NO: 47)
rArCrUrCrGrCrArGrArUrCrArUrUrArCrGArUrC (SEQ ID NO: 48)
PNA with the sequence of SEQ ID NO: 47
BNA with the sequence of SEQ ID NO: 47
CGATTGACTAAGCTATACGC (SEQ ID NO: 49)
rCrGrArUrUrGrArCrUrArArGrCrUrArUrArCrGrC (SEQ ID NO: 50)
PNA with the sequence of SEQ ID NO: 49
BNA with the sequence of SEQ ID NO: 49

Portion (iii)

Portion (iii) of the non-RNA polynucleotide should be of sufficient length to allow binding of the specific DNA helicase of use and should be composed of non-RNA nucleotides including analogues such as PNA, GNA, TNA, BNA, LNA or morpholino. In a preferred embodiment of the invention, portion (iii) is DNA. Representative examples of sequences that meet these criteria are:

TTTTTTTTT (SEQ ID NO: 51)
TTTTTTTTTTTTTTTTTTT (SEQ ID NO: 52)
CCCCCCCCCA(SEQ ID NO: 31)
CCCCCCCCCACCCCCCCCCA (SEQ ID NO: 53)
XXXXXXXXX (where X is independently selected from A, T, G or C).

Portion (iv)

Portion (iv) of the non-RNA polynucleotide should prevent or slow ATP-mediated translocation of the DNA helicase without the force of the pore. Representative examples of sequences that meet these criteria are:

rArArArArArArArArArA (SEQ ID NO: 54)
rUrCrCrArUrArCrGrArA (SEQ ID NO: 55)
9999

where 9 = PEG6 phosphate (Spacer 9 (iSp9) groups)

Portion (v)

Portion (v) of the non-RNA polynucleotide should allow hybridisation of a tethering oligo with a TM high enough to form a stable hybrid. Representative examples of sequences that meet these criteria are:

AACTACTAGGATCATCGATGTATCTGCTCA (SEQ ID NO: 56)
AGCTTAACATACGATACTCTTAGCTAACCA (SEQ ID NO: 57)
rArArCrUrArCrUrArGrGrArUrCrArUrCrGrArUrGrUrArUrCrUrGrCrUrCrA (SEQ ID NO: 58)
rArGrCrUrUrArArCrArUrArCrGrArUrArCrUrCrUrUrArGrCrUrArArCrCrA (SEQ ID NO: 59)
PNA with the sequence of SEQ ID NO: 56
PNA with the sequence of SEQ ID NO: 57

Portion (vi)

Portion (vi) of the non-RNA polynucleotide should facilitate ligation of the non-RNA polynucleotide to the RNA polynucleotide. Representative examples of sequences that meet these criteria are:

ACTCTGAACC (SEQ ID NO: 60)
ACTCTrGrArArCrC (SEQ ID NO: 61)
GCACAATGAT (SEQ ID NO: 62)
GCACArArTrGrArT (SEQ ID NO: 63)

Any combination of (i) to (vi) may be made in accordance with the invention: Preferably the non-RNA polynucleotide comprises (iii) and (vi) in combination with {i}, {ii}, {iv}, {v}, {i,ii}, {i,iv}, {i,v}, {ii,iv}, {ii,v}, {iv,v},{i,ii,iv}, {i,ii,v}, {i,iv,v}, {ii,iv,v},{i,ii,iv,v}.

Any of the different representative examples given above for each of portions (i) to (vi) can be used interchangeably to form a non-RNA polynucleotide for attachment to a target RNA sequence.

Seven different representative example of a non-RNA polynucleotide are as follows:

Construct 1: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)
(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
i.e., (40x SpC3)(SEQ ID NO: 47)(SEQ ID NO: 51)(4x Sp18)(SEQ ID NO: 56)(SEQ ID NO: 60)

Construct 2: (40x SpC3)(ACTCGCAGATCATTACGATC)(20x dT)(4x Sp18)
(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
i.e., (40x SpC3)(SEQ ID NO: 47)(SEQ ID NO: 52)(4x Sp18)(SEQ ID NO:56)(SEQ ID NO: 60)

Construct 3: (40x rU)(ACTCGCAGATCATTACGATC)(10x dT(4x Sp18)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
i.e., (40x rU)(SEQ ID NO: 47)(SEQ ID NO: 51)(4x Sp18)(SEQ ID NO: 56)(SEQ ID NO: 60)

Construct 4: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)
(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTrGrArArCrC)
i.e., (40x SpC3)(SEQ ID NO: 47)(SEQ ID NO: 51)(4x Sp18)(SEQ ID NO: 56) (SEQ ID NO: 61)

Construct 5: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(rArArArCrUrArCrGrCrU)(AACTACTAGGATCATCGATGTATCTGCTCA)
(ACTCTGAACC)
i.e., (40x SpC3)(SEQ ID NO: 47)(SEQ ID NO: 51)(SEQ ID NO: 64)(SEQ ID NO: 56) (SEQ ID NO: 60)

Construct 6: (40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)
(AGCTTAACATACGATACTCTTAGCTAACCA) (ACTCTGAACC)
i.e., (40x SpC3)(SEQ ID NO: 47)(SEQ ID NO: 51)(4x Sp18)(SEQ ID NO: 57)
(SEQ ID NO: 60)

Construct 7: (40x SpC3)(PNA with the sequence of SEQ ID NO: 47)(10x dT)(4x Sp18)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)
i..e, (40x SpC3)(PNA with the sequence of SEQ ID NO: 47)(SEQ ID NO: 51)(4x Sp18)(SEQ ID NO: 56)(SEQ ID NO: 60)

Attachment

The target RNA polynucleotide is attached to a non-RNA polynucleotide and/or the non-RNA polynucleotide is attached to the target RNA polynucleotide to form a modified RNA polynucleotide. Where the non-RNA polynucleotide comprises an ribonucletoide or an RNA sequence, the non-RNA polynucleotide may be attached to the target RNA polynucleotide via the ribonucleotide or RNA sequence which is comprised within the non-RNA polynucleotide.

The method comprises in step (a) providing the modified RNA polynucleotide and a DNA helicase enzyme. The method may further comprise, before step (a), attaching the RNA polynucleotide to the non-RNA polynucleotide and/or attaching the non-RNA polynucleotide to the RNA polynucleotide to form a modified RNA polynucleotide.

The target RNA polynucleotide and the non-RNA polynucleotide (which may be DNA or a DNA analogue) can be attached to each other in any manner using any method or methods known in the art. Preferably the target RNA polynucleotide and the non-RNA polynucleotide are attached using one or more of the various methods described below. The target RNA polynucleotide may be chemically attached to the non-RNA polynucleotide, for example by a covalent bond. The target RNA polynucleotide may be attached to the non-RNA polynucleotide by chemical or enzymatic ligation. The target RNA polynucleotide may be attached to the non-RNA polynucleotide by hybridisation and/or synthetic methods. The RNA polynucleotide may be attached to the non-RNA polynucleotide using a topoisomerase. The RNA polynucleotide may be attached to the non-RNA polynucleotide at more than one, such as two or three, points. The method of attachment may involve one, two, three, four, five or more different methods of attachment. Any combination of the attachment methods described below may be used in accordance with the invention.

The RNA polynucleotide and non-RNA polynucleotide (which may or may not comprise a ribonucleotide or RNA) 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. 5' or 3') ends. Suitable configurations include, but are not limited to, the 5' end of the RNA polynucleotide being attached to the 3' end of the non-RNA polynucleotide and *vice versa*. Alternatively, the two components may be attached via nucleotides within their sequences.

Polynucleotides may be attached via their naturally occurring nucleotides. Naturally occurring nucleotides may be modified to facilitate attachment. For instance, the naturally occurring nucleic acids may be modified by, for example, Trimethyl guanosine synthase for the mRNA cap. Other suitable modifications are known in the art. Modifications may be introduced by substitution. The RNA polynucleotide may be attached to the non-RNA polynucleotide via a linker molecule. The RNA polynucleotide may be attached to the non-RNA polynucleotide using one or more, such as two or three, linkers. 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 that are suitable for use as linkers, 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). 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 RNA polynucleotide may be attached to the non-RNA polynucleotide 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).

Reactions between the RNA polynucleotide and non-RNA polynucleotide may be spontaneous, such as cysteine/maleimide, or may require external reagents, such as Cu(I) for linking azide and linear alkynes.

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 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.

Cross-linkage of RNA polynucleotide or non-RNA polynucleotide to themselves may be prevented by keeping the concentration of linker in a vast excess of the RNA polynucleotide and/or non-RNA polynucleotide. 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. RNA polynucleotide or non-RNA polynucleotide).

The attachment of the RNA polynucleotide to the non-RNA polynucleotide may be permanent or stable (i.e. the RNA polynucleotide does not become detatched from the non-RNA polynucleotide in the method of the invention). A preferred permanent or stable attachment is covalent attachment.

Alternatively the attachment is transient, i.e. the RNA polynucleotide may detach from the non-RNA polynucleotide. It will be understood by one of skill in the art that any of the methods described herein are equally suitable for attaching a non-RNA polynucleotide to a target RNA and also in constructing the non-RNA polynucleotide itself which may, as described above, be a hybrid of two or more types of nucleic acid e.g., DNA and RNA.

Click Chemistry

The target RNA polynucleotide can be covalently attached to the non-RNA polynucleotide. The non-RNA polynucleotide may or may not comprise a pre-bound DNA helicase enzyme. In a preferred embodiment, the covalent bond between the RNA polynucleotide and non-RNA polynucleotide, e.g., a DNA leader sequence, can be made using copper free click chemistry or copper catalysed click chemistry. Click chemistry has been used in these applications because of its desirable properties and its scope for creating covalent links between diverse building blocks. For example, it is fast, clean and not poisonous, generating only inoffensive byproducts. Click chemistry is a term first introduced by Kolb *et al.*in 2001 to describe an expanding set of powerful, selective, and modular building blocks that work reliably in both small- and large-scale applications (Kolb HC, Finn, MG, Sharpless KB, Click chemistry: diverse chemical function from a few good reactions, Angew. Chem. Int. Ed. 40 (2001) 2004–2021). They have defined the set of stringent criteria for click chemistry as follows: "The reaction must be modular, wide in scope, give very high yields, generate only inoffensive byproducts that can be removed by nonchromatographic methods, and be stereospecific (but not necessarily enantioselective). The required process characteristics include simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation. Purification if required must be by nonchromatographic methods, such as crystallization or distillation, and the product must be stable under physiological conditions".

Suitable examples of click chemistry include, but are not limited to, the following:

(a) copper-free variant of the 1,3 dipolar cycloaddition reaction, where an azide reacts with an alkyne under strain, for example in a cyclooctane ring;

(b) the reaction of an oxygen nucleophile on one linker with an epoxide or aziridine reactive moiety on the other; and

(c) the Staudinger ligation, where the alkyne moiety can be replaced by an aryl phosphine, resulting in a specific reaction with the azide to give an amide bond.

Preferably the click chemistry reaction is the Cu (I) catalysed 1,3 dipolar cycloaddition reaction between an alkyne and an azide. In a preferred embodiment, the first group is an azide group and the second group is an alkyne group. Nucleic acid bases have already been synthesized incorporating azide and alkyne groups in preferred positions (for example Kocalka P, El-Sagheer AH, Brown T, Rapid and efficient DNA strand cross-linking by click chemistry, Chembiochem. 2008. 9(8):1280-5). Alkyne groups are available

commercially from Berry Associates (Michigan, USA) and azide groups are synthesized by ATDBio or IDT bio.

If nucleotides within the linkers' nucleic acid acid regions are modified to include groups that can form covalent bonds, the modified nucleotides are preferably offset from one another by one nucleotide in order to achieve the link. This follows the published work of Tom Brown (Kocalka *et al.* (2008) ChemBiochem 9 8 1280-1285).

A click-reactive base may be added to the target RNA polynucleotide when the RNA transcript is formed. Alternatively a click group may be added to a target RNA polynucleotide by hypermethylase enzyme (for capped mRNA).

Preferably the reactive groups are azide and hexynl groups such as 3AzideN and 5'-hexynl-G. Preferably the azide group is attached to the non-RNA polynucleotide which is preferably DNA (and may or may not comprise a ribonucleotide or RNA sequence) and the hexynl group is attached to the target RNA polynucleotide.

Examples 3 and 5 illustrate the use of a click reaction to join a non-RNA polynucleotide (comprising for example DNA) to a target RNA polynucleotide. Example 5 further describes the use of a bridging moiety to obtain a 2 dimensional (2D) RNA-cDNA construct. The construct of Example 5 which is illustrated in Figure 15 can be compared to the construct of Example 3 (1D construct illustrated in Figure 10) which does not comprise a bridging moiety.

Ligation

The target RNA polynucleotide may be ligated to the non-RNA polynucleotide (which RNA polynucleotide may or may not comprise a ribonucleotide or RNA sequence). Ligation is the joining of two nucleic acid fragments most commonly through the action of an enzyme or by chemical means. The ends of RNA and DNA fragments are joined together by the formation of phosphodiester bonds between the 3'-hydroxyl of one RNA or DNA terminus with the 5'-phosphoryl of another. A co-factor is generally involved in the reaction, and this is usually ATP or NAD^+. A splint of RNA or non-RNA polynucleotide, such as a DNA, PNA, glycerol nucleic acid (GNA), threose nucleic acid (TNA) or locked nucleic acid (LNA) may be used in the ligation reaction to facilitate ligation by holding the RNA polynucleotide and the non-RNA polynucleotide adjacent to each other by hybridisation to the splint. The non-RNA polynucleotide to be attached to the RNA polynucleotide may or may not comprise a pre-bound DNA helicase enzyme.

The non-RNA polynucleotide may be ligated to either end of the RNA polynucleotide, i.e. the 5' or the 3' end. The non-RNA polynucleotide may be ligated to both ends of the target RNA polynucleotide. Preferably the non-RNA polynucleotide is ligated to the 5' end of the target RNA polynucleotide. The non-RNA polynucleotide may be ligated to the RNA polynucleotide using any method known in the art. The one or more non-RNA polynucleotide may be ligated using a ligase, such as T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase, 9°N DNA ligase, T4 Polymerase I, T4 Polymerase 2, Thermostable 5' App DNA/RNA ligase, SplintR, circ Ligase, T4 RNA ligase 1 or T4 RNA ligase 2. The one or more non-RNA polynucleotides may be ligated to the RNA polynucleotide (or vice versa) in the absence of ATP or using gamma-S-ATP (ATPγS) instead of ATP.

The method preferably further comprises removing the ligase from the method conditions.

Example 2 illustrates the ligation of RNA polynucleotide to DNA using T4 DNA ligase.

Synthetic methods

An oligonucleotide or primer can be used to hybridise to any region of the target RNA polynucleotide and act as a starting point for DNA synthesis. The oligonucleotide or primer may or may not comprise a pre-bound DNA helicase enzyme.

Eukaryotic RNA typically comprises polyA tail, *i.e.* a stretch of consecutive adenosine monophosphates. The polyA tail is typically at the 3' end of the RNA. The Poly A Polymerase or Terminal Transferase enzyme can be utilised to add a poly(dA) tail onto the 3' end of a prokaryotic RNA strand if necessary. A primer can be hybridised to the polyA tail of the target RNA and used as a starting point for synthesis. The primer preferably comprises a polyT region, i.e. a region containing only nucleotides based on thymine, or a polyU region, i.e. a region containing only nucleotides based on uracil. The polyU region may contain UMP or dUMP. The polyU region may be any length, such as at least 10, at least 15, at least 20, at least 25 or more.

In one embodiment of the invention the non-RNA polynucleotide comprises a DNA primer with a leader sequence which is hybridised to an RNA strand at the poly(dA) region. One or more DNA helicase enzymes may be pre-bound to the DNA primer that is to be hybridised to the RNA strand. Alternatively the DNA primer that is to be hybridised to the RNA does not contain or comprise pre-bound DNA helicase enzyme. Reverse transcription

of the RNA strand from the DNA primer is allowed to occur and a bridging moiety such as a DNA hairpin can then be ligated to the double-stranded DNA/RNA. Such a bridging moiety, for example a hairpin loop adaptor, can be added to any double- stranded target RNA polynucleotide (RNA/RNA or RNA/DNA) or modified construct of the invention. A bridging moiety permits contiguous sequencing of both template and complement strands by connection into a single strand. Preferably an adaptor (e.g., a Y adaptor) containing a leader sequence is attached to one end of the RNA and a bridging moiety adaptor is attached to the other end. The leader sequence preferentially threads into the nanopore and the bridging moiety connecting the two strands (which may be RNA/RNA or RNA/DNA) allows both strands to be characterised as the polynucleotide unzips and both strands (connected via the bridging moiety) move through the pore. This is advantageous because it doubles the amount of information obtained from a double- stranded polynucleotide. Moreover, because the sequences in the two strands are complementary, the information from the two strands can be combined informatically. This mechanism provides a proof-reading capability that provides higher confidence observations.

Alternatively the bridging moiety can be added to a target RNA polynucleotide prior to synthesis of the complement and used as a primer for complement synthesis, as discussed in more detail below.

In one embodiment of the invention, the strands of a double stranded RNA polynucleotide or an RNA/DNA duplex (e.g., RNA and cDNA) are linked using a bridging moiety. The method of characterising a target RNA polynucleotide according to the invention then preferably comprises contacting (i) the linked construct comprising the target RNA polynucleotide, wherein the RNA polynucleotide is modified to comprise a non-RNA polynucleotide, and (ii) a DNA helicase enzyme with a transmembrane pore such that the target RNA moves through the pore. The method preferably comprises taking one or more measurements as the target RNA moves with respect to the pore wherein the measurements are indicative of one or more characteristics of the complementary polynucleotide (RNA or cDNA) and the target RNA and thereby characterising the target double stranded polynucleotide.

Linking and interrogating both strands of a target dsRNA or RNA/DNA duplex in this way increases the efficiency and accuracy of characterization.

The bridging moiety is capable of linking the two strands of a target dsRNA polynucleotide or RNA/DNA duplex. The bridging moiety typically covalently links the two strands of a target dsRNA polynucleotide or RNA/DNA duplex. The bridging moiety can be

anything that is capable of linking the two strands of a target dsRNA polynucleotide or RNA/DNA duplex , provided that the bridging moiety does not interfere with movement of the RNA polynucleotide through the transmembrane pore.

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

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

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

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

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

The complementary polynucleotide (RNA or cDNA) and the target RNA may be separated after or before the linked construct is contacted with the pore in accordance with the invention. They may be separated as the polynucleotide movement through the pore is controlled by a polynucleotide binding protein, such as a helicase, or molecular brake.

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

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

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

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

The linked construct preferably comprises a leader sequence at the opposite end from the bridging moiety, such as a hairpin loop or hairpin loop adaptor.

In one embodiment of the invention, a bridging moiety such as a hairpin-forming oligonucleotide is attached to the target RNA strand. Prior to the attachment of the bridging moiety, the target RNA strand may have been modified to comprise a non-RNA polynucleotide. For example the target RNA strand may have been attached to a non-RNA polynucleotide using chemical attachment, for example by a covalent bond, click chemistry,

chemical or enzymatic ligation, by hybridisation and/or synthetic methods. The RNA polynucleotide may have been attached to the non-RNA polynucleotide using a topoisomerase. Alternatively the briding moiety may be attached to the target RNA strand prior to the modification of the RNA strand to comprise a non-RNA polynucleotide. Similarly, the bridging moiety, such as a hairpin-forming oligonucleotide, may be attached to the target RNA strand by any of the attachment methods described herein. Preferably the bridging moiety is attached to the target RNA strand by ligation. Any suitable ligase described above may be used for the ligation of the bridging moiety to the target RNA strand e.g., T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase, 9°N DNA ligase, T4 Polymerase I, T4 Polymerase 2, Thermostable 5' App DNA/RNA ligase, SplintR, circ Ligase, T4 RNA ligase 1 or T4 RNA ligase 2.

Preferably the bridging moiety is attached to the 3' or 5'end of the target RNA strand, most preferably the bridging moiety is attached to the the 3' end of the target RNA strand. Reverse transcription from the bridging moiety, such as a hairpin-forming oligonucleotide, results in the formation of a RNA-cDNA construct that is joined by a hairpin. In this embodiment the bridging moiety acts as a primer for reverse transcription: The bridging moiety, such as hairpin forming oligonucleotide, is itself used as a primer for reverse-transcription to generate an RNA-cDNA construct and enable a 2D read. Preferably reverse transcription is initated at the 3' end of the bridging moiety. A poly-T overhang on the bridging moiety hybridizes to the poly A-tail of the RNA. The polyA-tail of the RNA is typically from 50 to 300 nucleotides in length. Preferably the poly-T overhang of the bridging moiety comprises less than 100 nucleotides, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57 ,58, 59, 60, 61, 62. 63. 64 .65 .66. 67.68, 69, 70, 71, 72, 73, 74 ,75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 nucleotides . Preferably the poly T overhang is on the 3' end of the bridging moiety. In this embodiment of the invention, where a bridging moiety is attached to a target RNA, (which target RNA has already been or is yet to be modified to comprise a non-RNA polynucleotide; preferably a DNA polynucleotide) and wherein the bridging moiety serves as a primer for reverse transcription, a DNA/RNA/cDNA construct is created (see Example 5 and Figure 15). This construct allows for a 2D read wherein both the target RNA and its complementary DNA sequence (formed by reverse transcription) can be sequenced by the transmembrane pore due to the presence of the bridging moiety. This method is advantageous because it doubles the amount of information

obtained from a single double stranded target polynucleotide construct. Moreover, because the sequence in the complementary (cDNA) strand is necessarily orthogonal to the sequence of the template RNA strand, the information from the two strands can be combined informatically. Thus, this mechanism provides an orthogonal proof-reading capability that provides higher confidence observations.

Furthermore, the other major advantages of the method of the invention are:

1) Coverage of missed nucleotides: the method substantially minimises issues of any missed nucleotides or groups of nucleotides (e.g. due to movement issues such as the RNA strand slipping through the pore), since any states that might be missed in one strand are likely to be covered by the orthogonal information obtained from its complement region.

2) Coverage of problematic sequence motifs: any difficult to sequence motifs are covered by the orthogonal and opposite information in the complementary strand, which having a different sequence will not have the same sequence dependent issues. For example, this is particularly relevant for sequence motifs that produce only small changes in current, or have similar current levels - i.e. consecutive base motifs that when moved through the nanopore produce the same current block, and are therefore not observed as there is no step change in current. Any similar current levels from one sequence motif will be covered by the entirely different current levels obtained from its orthogonal sequence in the complement strand.

In addition to the advantages discussed above there are a number of special cases where the concept of reading both strands of the double stranded polynucleotide can be utilized to provide further benefits:

1. Epigenetic information

Being able to identify epigenetic information (such as 5-methylcytosine or 5-hydroxymenthylcytosine nucleotides) or damaged bases within a natural RNA strand is desirable in a wide range of applications. Using the method of the present invention this information is obtained without chemical treatment or amplification, both of which can introduce errors. During nanopore sequencing, changes in current levels are measured as nucleotides pass through the nanopore. The current level is dictated by several bases rather than by a single nucleotide. When a modified nucleotide passes through the nanopore it affects several consecutive levels: This increases the confidence we can have in its detection (See Figure 18).

Nanopore sequencing is also a single molecule sequencing technology and therefore can be performed without the need for amplification. It has been shown that nanopores can

detect modifications to the standard four RNA nucleotides. Reading both strands of the polynucleotide can be useful in detecting RNA modifications in situations where a modified base behaves in a similar way (generates a similar current signal) to another base. For example if methylcytosine (mC) behaves in a similar way to uracil there is an error associated with assigning a mC to a U. In the template strand, there is a probability of the base being called a mC or a U. However, in the complement strand, the corresponding base may appear as a G with a high probability. Thus by "proof reading" the complement strand, it is highly likely that the base in the template strand was a mC rather than a U.

Reading the template and the complement strand can be performed without the need of amplification or replication. However, amplification or replication may be added as part of the sample preparation to aid the detection of epigenetic information.

The linked strands comprising the target RNA polynucleotide can be separated and duplicated at any stage before sequencing is carried out and as many times as necessary. For example, after separating the two linked strands of a first RNA/cDNA polynucleotide construct as described above, a complementary strand to the resulting single stranded polynucleotide can be generated to form another double stranded polynucleotide. The two strands of this double stranded polynucleotide can then be linked using a bridging moiety to form a second construct. This may be referred to herein as the "DUO" method. This construct may then be used in the invention. In such an embodiment, one strand of the double stranded polynucleotide in the resulting construct contains both strands of the original target double stranded polynucleotide (in the first construct) linked by a bridging moiety. The sequence of the original target double stranded polynucleotide or the complement strand can be estimated or determined. This process of replication can be repeated as many times as necessary and provides additional proof reading as the target polynucleotide is in effect being read multiple times.

A nucleotide strand may be constructed where the following information is read through the nanopore in the following order: template RNA (original), complement cDNA (original), –bridging moiety–, template RNA (complement), complement cDNA (complement)

In this scheme, information on the methylated base will be obtained four times. If the epigenetic base is in the original template strand (in this case, mC), the following information will be obtained with a high probability: template (original)-mC, complement (original)-G, template (complement)-C, and complement (complement)-G. It is clear that the original template read and the replicated template read will give different results, while the both

complementary reads will yield the same base call. This information can be used to indicate the position of the epigenetic base in the original template strand.

2. Homopolymer reads

Homopolymer reads may be a problem for single molecule nanopore sequencing. If the homopolymer region is longer than the reading section of the pore, the length of the homopolymer section will be difficult to determine.

To overcome the problem of homopolymer reads, the cDNA strand can be synthesised with the addition of a different/modified base in combination with the original dTTP, dGTP, dATP, dCTP. This could be a natural base analogue such as inosine (I). The base will have a random chance of incorporating compared to the correct natural base and the insertion rates can be controlled by varying the concentration of the triphosphate species.

Through the addition of the alternative base, there will be a probability of an alternative base being inserted into the reverse complement of a homopolymer region. The result of this is that the homopolymer run will be reduced in length to a point where it can be read by the reading section of the nanopore. For example, a homopolymer group of AAAAAAAAAAAA (SEQ ID NO: 33) will have random insertions of the alternative base and may give TTTITTIITTTI (SEQ ID NO: 34) (where I is inosine).

The homopolymer stretch is reduced to allow individual nucleotides or groups of nucleotides to be estimated or determined. The template strand will be a natural RNA strand, while the complementary/cDNA strand will contain a mixture of natural bases and base analogues. The combination of data from the template and the complementary reads can be used to estimate the length of the homopolymer run in the original RNA section.

Topoisomerases

Topoisomerases bind to either single-stranded or double-stranded DNA and cut the phosphate backbone of the DNA. This intermediate break allows the DNA to be untangled or unwound, and, at the end of these processes, the DNA backbone is resealed again. Suitable topoisomerase binding strategies are illustrated in Figures 6 and 7.

Other methods

An alternative method of attaching the RNA target polynucleotide to a non-RNA polynucleotide comprises exploiting the 5' ends of eukaryotic RNA which are modified by the addition of a 7 methyl guanosine cap in the opposite orientation. The 5' ends of eukaryotic

RNA therefore contains or comprises reversed bases i.e., at the 5' end of the RNA strand the individual base has its 3' end free, as opposed the 5' end. The target RNA polynucleotide can be ligated to a non-RNA polynucleotide (which may or may not comprise a ribonucleotide or RNA sequence) which also has a region with reversed bases such that it runs in the opposite direction, as illustrated by Figure 12 wherein the non-RNA polynucleotide is DNA. Alternatively non-RNA polynucleotide may have a section of RNA with reversed bases at one end. This section of reversed RNA bases can be attached to the reversed bases at the 5' end of eukaryotic RNA by ligation.

The non-RNA polynucleotide may or may not comprise a pre-bound DNA helicase enzyme.

Eukaryotic RNA

In eukaryotes the primary RNA transcript produced in the nucleus is processed in several ways before transport to the cytoplasm where it is used to program the translation machinery. First, a cap consisting of a 7-methyl guanosine residue linked to the 5' end of the transcript by a triphosphate bond is added during transcription (capping). Capping involves a 5'-5' triphosphate linkage. The caps are recognised by the translation machinery and protect the growing RNA chain from degradation by nucleases. Then stretches of adenosine residues are added at the 3' ends (polyadenylation). These polyA tails are 150 to 200 residues long. After these modifications, RNA splicing removes intervening sequences (i.e., the introns).

In one embodiment of the invention the 5' cap is left in place and at least one chemically reactive group is added to the 5’ end of the target RNA polynucleotide. Any methods of chemical attachment may be used in the methods of the invention so long as a) the chemical reaction does not damage the RNA or DNA structure and b) the reactive linkage that is produced is not so bulky that the DNA helicase enzyme cannot move along it or past it. The at least one reactive group may be added to the target RNA polynucleotide using a hypermethylase enzyme. In one embodiment of the invention the at least one reactive group added to the RNA polynucleotide is a click reactive group although this is not essential. The at least one reactive group added to the RNA polynucleotide may alternatively be any suitable reactive group such as Thiol. At least one reactive group is also attached to the end of a non-RNA polynucleotide (which may or may not comprise a ribonucleotide or RNA sequence). Preferably the at least one reactive group is attached to the 3' end of the non-RNA polynucleotide. Preferably the non-RNA polynucleotide is a DNA strand. The at least one

reactive group added to the non-RNA polynucleotide may be added using a hypermethylase enzyme. In one embodiment of the invention the at least one reactive group added to the non-RNA polynucleotide is a click reactive group although this is not essential. The one or more reactive groups on each of the RNA polynucleotide and non-RNA polynucleotide are then contacted under suitable conditions to form a covalent bond. ATP gamma S and enzyme can be used to add a thiophosphate to DNA which could then be attached to RNA which has a maleimide attached.

In an alternative embodiment of the invention the 7-methylguanosine cap is removed, preferably using Tobacco Acid Pyrophosphatase to form a de-capped RNA strand. The de-capped target RNA polynucleotide can then be treated in a number of different ways in order to produce a strand of RNA polynucleotide attached to a non-RNA polynucleotide. In one embodiment, at least one reactive group can be added to the 5' end of the de-capped RNA polynucleotide. The at least one reactive group may be added to the RNA polynucleotide using a hypermethylase enzyme. In one embodiment of the invention the at least one reactive group added to the RNA polynucleotide is a click reactive group although this is not essential. The reactive group added to the RNA polynucleotide may alternatively be any suitable reactive group such as Thiol. At least one reactive group is also attached to the end of a non-RNA polynucleotide. Preferably the at least one reactive group is attached to the 3' end of the non-RNA polynucleotide. Preferably the non-RNA polynucleotide is a DNA strand. The at least one reactive group added to the non-RNA polynucleotide may be added using a hypermethylase enzyme. In one embodiment of the invention the at least one reactive group added to the non-RNA polynucleotide is a click reactive group although this is not essential. The one or more reactive groups on each of the RNA polynucleotide and non-RNA polynucleotide are then contacted under suitable conditions to form a covalent bond.

Alternatively, a strand of non-RNA polynucleotide can be ligated directly onto the target RNA polynucleotide using a ligase, such as T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase, 9°N DNA ligase, T4 Polymerase I, T4 Polymerase 2, Thermostable 5' App DNA/RNA ligase, SplintR, circ Ligase, T4 RNA ligase 1, T4 RNA ligase 2. Preferably the non-RNA polynucleotide is a DNA strand. In one embodiment of the invention, one or more enzymes may be pre-bound to the non-RNA polynucleotide that is to be ligated to the RNA. Pre-loading the enzymes on the non-RNA polynucleotide speeds up the sample preparation process and means that fewer tubes are used. Alternatively the non-RNA polynucleotide that is to be ligated to the RNA polynucleotide does not contain or comprise pre-bound enzyme.

In an alternative embodiment of the invention, the non-RNA polynucleotide is a DNA primer with a leader sequence that is hybridised to an RNA polynucleotide. One or more enzymes may be pre-bound to the DNA primer that is to be hybridised to the RNA polynucleotide. Alternatively the DNA primer that is to be hybridised to the RNA polynucleotide does not contain or comprise pre-bound enzyme. Reverse transcription of the RNA polynucleotide from the DNA primer results in a 3' overhang of one to three C's, depending on the reverse transcriptase enzyme used. A DNA hairpin can then be ligated to the double-stranded DNA/RNA. Conditions that permit the hybridisation are well-known in the art (for example, Sambrook *et al.*, 2001, Molecular Cloning: a laboratory manual, 3rd edition, Cold Spring Harbour Laboratory Press; and Current Protocols in Molecular Biology, Chapter 2, Ausubel et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995)). Hybridisation can be carried out under low stringency conditions, for example in the presence of a buffered solution of 30 to 35% formamide, 1 M NaCl and 1 % SDS (sodium dodecyl sulfate) at 37°C followed by a wash in from 1X (0.1650 M Na+) to 2X (0.33 M Na+) SSC (standard sodium citrate) at 50°C. Hybridisation can be carried out under moderate stringency conditions, for example in the presence of a buffer solution of 40 to 45% formamide, 1 M NaCl, and 1 % SDS at 37°C, followed by a wash in from 0.5X (0.0825 M Na+) to 1X (0.1650 M Na+) SSC at 55°C. Hybridisation can be carried out under high stringency conditions, for example in the presence of a buffered solution of 50% formamide, 1 M NaCl, 1% SDS at 37°C, followed by a wash in 0.1X (0.0165 M Na+) SSC at 60°C. In particular, the conditions are preferably 10 uM oligomers in 10 mM Tris-HCl, 50 mM NaCl, pH 7 and heat to 98°C before cooling to 18°C at 2°C per minute.

Eukaryotic RNA typically comprises polyA tail, *i.e.* a stretch of consecutive adenosine monophosphates. The polyA tail is typically at the 3' end of the RNA. In such embodiments, a primer can be hybridised to the polyA tail of the target RNA. The primer preferably comprises a polyU region, i.e. region containing only nucleotides based on uracil. The polyU region may contain UMP or dUMP. The polyU region may be any length, such as at least 10, at least 15, at least 20, at least 25 or more. Alternatively the primer comprises a polyT region. The polyT region may be any length, such as at least 10, at least 15, at least 20, at least 25 or more. Methods of attaching a strand of eukaryotic RNA to a strand of DNA are illustrated in Figure 1.

In an alternative embodiment of the invention, a non-RNA polynucleotide can be attached to the target RNA polynucleotide using a topoisomerase-based strategy (Cheng and Shuman, 2000, Nucleic Acids Research, Vol. 28, No. 9 1893-1898, incorporated by reference

herein). DNA topoisomerase binds to duplex DNA and incises the phosphodiester backbone of one strand at a specific target site. The other strand of the duplex DNA comprises a nick or break in the sequence at a corresponding position. Once the topoisomerase has bound and cut the upper strand of the DNA it remains bound to the dsDNA. The topoisomerase which is bound to the dsDNA can transfer the DNA strand to a 5' OH terminated RNA strand to form a tandem DNA-RNA copolymer if it is then incubated with RNA which has a free 5' hydroxyl. The target RNA must have a free 5' hydroxyl: Eukaryotic RNA must be de-capped to produce RNA with a free 5' hydroxyl whereas microRNA has a free 5' hydroxyl group (See Figure 6).

In an alternative topoisomerase-based strategy, the dsDNA is attached to a strand of RNA using a single-stranded region of DNA to hybridise to the RNA and assist in the RNA attachment (Sekiguchi et al., 1997, The Journal of Biological Chemistry, Vol 272, No. 25, 15721-15728, incorporated by reference herein). The topoisomerase binds onto the dsDNA at the specific target sequence. The lower strand of DNA does not have a nick in it. Once the topoisomerase has bound it cuts the upper strand of the DNA only and remains bound to the dsDNA/ssDNA hybrid. The topoisomerase, which is bound to the dsDNA, is then incubated with RNA that has a free 5' hydroxyl and joins the RNA to the dsDNA. The ssDNA region assists in attracting the complementary RNA sequence to the point of attachment to the DNA (See Figure 7).

An alternative method of attaching an RNA polynucleotide to a non-RNA polynucleotide comprises exploiting the 5' ends of eukaryotic RNA which are modified by the addition of a 7-methylguanosine cap which runs in the opposite orientation. The 5' ends of eukaryotic RNA therefore contains or comprises a guanine base which has been reversed i.e., it is at the 5' end of the RNA strand but the individual base has its 3' end free, as opposed the 5' end. The target RNA can be ligated to a DNA (or RNA) sequence which also has a region with reversed bases such that it runs in the opposite direction (See Figure 12).

Prokaryotic RNA

The Poly A Polymerase enzyme can be utilised to add a poly(dA) tail onto the 3' end of the RNA polynucleotide. In one embodiment of the invention, the non-RNA polynucleotide is a DNA primer with a leader sequence which is hybridised to an RNA polynucleotide. A DNA primer with a leader sequence can be hybridised to the poly(dA) region. One or more enzymes may be pre-bound to the DNA primer that is to be hybridised to the RNA polynucleotide. Alternatively the DNA primer that is to be hybridised to the

RNA polynucleotide does not contain or comprise pre-bound enzyme. Reverse transcription of the RNA polynucleotide from the DNA primer results in a 3' overhang of one to three C's depending on the reverse transcriptase enzyme used (three to four for SuperScript II (MMLV) (see p1192 Biotechniques Vol 29 No 6 (2000) . A DNA hairpin can then be ligated to the double-stranded DNA/RNA.

In an alternative embodiment of the invention, a poly (dA) region is not added to the RNA polynucleotide and the prokaryotic RNA polynucleotide is used directly in the reaction steps. A reactive group can be added to the 5' or 3' end of the RNA polynucleotide. Preferably at least one reactive group is added to the 5' end of RNA polynucleotide. The at least one reactive group added to the RNA polynucleotide may be a click reactive group although this is not essential. The at least one reactive group added to the RNA polynucleotide may alternatively be any suitable reactive group such as Thiol. At least one reactive group is also attached to the end of a non-RNA polynucleotide. Preferably the at least one reactive group is attached to the 3' end of the non-RNA polynucleotide. Preferably the non-RNA polynucleotide comprises a DNA strand. The at least one reactive group added to the non-RNA polynucleotide may be added using a hypermethylase enzyme. In one embodiment of the invention the at least one reactive group added to the non-RNA polynucleotide is a click reactive group although this is not essential. The one or more reactive groups on each of the RNA and non-RNA polynucleotide are then contacted under suitable conditions to form a covalent bond.

In an alternative embodiment of the invention, a strand of non-RNA polynucleotide can be ligated directly onto the target RNA polynucleotide using a ligase, such as T4 DNA ligase, *E. coli* DNA ligase, Taq DNA ligase, Tma DNA ligase, 9°N DNA ligase, T4 Polymerase I, T4 Polymerase 2, Thermostable 5' App DNA/RNA ligase, SplintR, circ Ligase, T4 RNA ligase 1 and T4 RNA ligase 2. Preferably the non-RNA polynucleotide comprises a DNA strand. In one embodiment of the invention, one or more enzymes may be pre-bound to the non-RNA polynucleotide that is to be ligated to the RNA polynucleotide. Alternatively the non-RNA polynucleotide that is to be ligated to the RNA polynucleotide does not contain or comprise pre-bound enzyme. Methods of attaching prokaryotic RNA polynucleotide to a non-RNA (e.g., DNA) polynucleotide are illustrated in Figure 2.

In an alternative embodiment of the invention, a non-RNA polynucleotide can be attached to the target RNA polynucleotide using a topoisomerase-based strategy (Cheng and Shuman, 2000, Nucleic Acids Research, Vol. 28, No. 9 1893-1898, incorporated by reference herein). DNA topoisomerase binds to duplex DNA and incises the phosphodiester backbone

of one strand at a specific target site. The other strand of the duplex DNA comprises a nick or break in the sequence at a corresponding position. Once the topoisomerase has bound and cut the upper strand of the DNA it remains bound to the dsDNA. The topoisomerase which is bound to the dsDNA can transfer the DNA strand to a 5' OH terminated RNA strand to form a tandem DNA-RNA copolymer if it is then incubated with RNA which has a free 5' hydroxyl. The target RNA must have a free 5' hydroxyl. In an alternative topoisomerase-based strategy, the dsDNA is attached to a strand of RNA using a single-stranded region of DNA to hybridise to the RNA and assist in the RNA attachment (Sekiguchi et al., 1997, The Journal of Biological Chemistry, Vol 272, No. 25, 15721-15728, incorporated by reference herein). The topoisomerase binds onto the dsDNA at the specific target sequence. The lower strand of DNA does not have a nick in it. Once the topoisomerase has bound it cuts the upper strand of the DNA only and remains bound to the dsDNA/ssDNA hybrid. The topoisomerase which is bound to the dsDNA is then incubated with RNA which has a free 5' hydroxyl and joins the RNA to the dsDNA. The ssDNA region assists in attracting the complementary RNA sequence to the point of attachment to the DNA (See Figure 7).

No amplification

The target RNA polynucleotide is typically not amplified in the method of the invention. The method typically does not comprise making multiple copies of the target RNA.

The method preferably does not comprise polymerase chain reaction (PCR) or reverse transcription PCR (RT-PCR).

As discussed above, in one embodiment of the invention the RNA polynucleotide comprises a DNA complement that has been generated using a DNA primer and a reverse transcriptase. Linking of the RNA and DNA strands using a briding moiety enables a 2D read.

In an alternative embodiment of the invention, a hairpin forming oligonucleotide is itself used as a primer for reverse-transcription to generate an RNA-cDNA construct and enable a 2 dimensional read (i.e., RNA and cDNA strand).

DNA Helicase(s) and molecular brake(s)

DNA helicases are used to control the movement of RNA polynucleotide through the pore. The DNA helicase enzyme does not need to display enzymatic activity as long as it is capable of binding the target RNA polynucleotide and controlling its movement through the

pore. For instance, the enzyme may be modified to remove its enzymatic activity or may be used under conditions which prevent it from acting as an enzyme.

The helicase may be or be derived from a Hel308 helicase, a RecD helicase, such as TraI helicase or a TrwC helicase, a XPD helicase or a Dda helicase. The DNA helicase may be derived from Hel308 Mbu (SEQ ID NO: 8), Hel308 Csy (SEQ ID NO: 9), Hel308 Tga (SEQ ID NO: 10), Hel308 Mhu (SEQ ID NO: 11), TraI Eco (SEQ ID NO: 12), XPD Mbu (SEQ ID NO: 13), Dda 1993 (SEQ ID NO:14) or a variant thereof.

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

Helicases may work in two modes with respect to the pore. First, the method is preferably carried out using a DNA helicase such that it controls movement of the RNA polynucleotide through the pore with the field resulting from the applied voltage. In this mode the 5' end of the RNA polynucleotide is first captured in the pore, and the enzyme controls movement of the RNA polynucleotide into the pore such that the RNA polynucleotide is passed through the pore with the field until it finally translocates through to the trans side of the membrane. Alternatively, the method is preferably carried out such that a DNA helicase enzyme controls movement of the RNA polynucleotide through the pore against the field resulting from the applied voltage. In this mode the 3' end of the RNA polynucleotide is first captured in the pore, and the enzyme controls movement of the RNA polynucleotide through the pore such that the RNA polynucleotide is pulled out of the pore against the applied field until finally ejected back to the cis side of the membrane.

The DNA helicase preferably comprises the sequence shown in SEQ ID NO: 15 (Trwc Cba) or a variant thereof, the sequence shown in SEQ ID NO: 8 (Hel308 Mbu) or a variant thereof or the sequence shown in SEQ ID NO: 14 (Dda) or a variant thereof. Variants may differ from the native sequences in any of the ways discussed below for transmembrane pores. A preferred variant of SEQ ID NO: 14 comprises (a) E94C and A360C or (b) E94C, A360C, C109A and C136A and then optionally (ΔM1)G1 (i.e. deletion of M1 and then addition of G1).

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

numbers of helicases may be used. In one embodiment of the invention the DNA helicase is pre-bound to the non-RNA polynucleotide.

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

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

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

A variant of SEQ ID NO: 8, 9, 10, 11, 12, 13, 14 or 15 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 8, 9, 10, 11, 12, 13, 14 or 15 and which retains polynucleotide binding ability. This can be measured using any method known in the art. For instance, the variant can be contacted with a polynucleotide and its ability to bind to and move along the polynucleotide can be measured. The variant may include modifications that facilitate binding of the polynucleotide and/or facilitate its activity at high salt concentrations and/or room temperature. Variants may be modified such that they bind polynucleotides (i.e. retain polynucleotide binding ability) but do not function as a helicase (i.e. do not move along polynucleotides when provided with all the necessary components to facilitate movement, e.g. ATP and Mg^{2+}). Such modifications are known in the art. For instance, modification of the Mg^{2+} binding domain in helicases typically results in variants which do not function as helicases. These types of variants may act as molecular brakes (see below).

Over the entire length of the amino acid sequence of SEQ ID NO: 8, 9, 10, 11, 12, 13, 14 or 15, a variant will preferably be at least 50% homologous to that sequence based on amino acid similarity or 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 similarity or identity to the amino acid sequence of SEQ ID NO: 8, 9, 10, 11, 12, 13, 14 or 15

over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid similarity or identity over a stretch of 200 or more, for example 230, 250, 270, 280, 300, 400, 500, 600, 700, 800, 900 or 1000 or more, contiguous amino acids ("hard homology"). Homology is determined as described above. The variant may differ from the wild-type sequence in any of the ways discussed above with reference to SEQ ID NO: 2 and 4 above. The enzyme may be covalently attached to the pore. Any method may be used to covalently attach the enzyme to the pore.

In a preferred embodiment, the method comprises:

(a) providing RNA polynucleotide with one or more DNA helicases and one or more molecular brakes, wherein the RNA polynucleotide is modified to comprise a non-RNA polynucleotide and increase DNA helicase binding thereto;

(b) contacting the RNA polynucleotide with a transmembrane pore and applying a potential across the pore such that the one or more DNA helicases and the one or more molecular brakes are brought together and both control the movement of the RNA polynucleotide through the pore;

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

This type of method is discussed in detail in UK Application No. 1406151.9. A preferred molecular brake is TrwC Cba-Q594A (SEQ ID NO: 15 with the mutation Q594A). This variant does not function as a helicase (i.e. binds polynucleotides but does not move along them when provided with all the necessary components to facilitate movement, e.g. ATP and Mg^{2+}).

The one or more helicases may be any of those discussed above. The one or more molecular brakes may be any compound or molecule which binds to the RNA polynucleotide and slows the movement of the RNA polynucleotide through the pore. The one or more molecular brakes preferably comprise one or more compounds which bind to the RNA polynucleotide. The one or more compounds are preferably one or more macrocycles. Suitable macrocycles include, but are not limited to, cyclodextrins, calixarenes, cyclic peptides, crown ethers, cucurbiturils, pillararenes, derivatives thereof or a combination thereof. The cyclodextrin or derivative thereof may be any of those disclosed in Eliseev, A. V., and Schneider, H-J. (1994) *J. Am. Chem. Soc.* 116, 6081-6088. The agent is more preferably heptakis-6-amino-β-cyclodextrin (am_7-βCD), 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1-βCD) or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu_7-βCD).

The one or more molecular brakes are preferably not one or more single stranded binding proteins (SSB). The one or more molecular brakes are more preferably not a single-stranded binding protein (SSB) comprising a carboxy-terminal (C-terminal) region which does not have a net negative charge or (ii) a modified SSB comprising one or more modifications in its C-terminal region which decreases the net negative charge of the C-terminal region. The one or more molecular brakes are most preferably not any of the SSBs disclosed in International Application No. PCT/GB2013/051924 (published as WO 2014/013259).

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

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

Any number of molecular brakes derived from a helicase may be used. For instance, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more helicases may be used as molecular brakes. If two or more helicases are be used as molecular brakes, the two or more helicases are typically the same helicase. The two or more helicases may be different helicases.

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

The two or more helicases are preferably attached to one another. The two or more helicases are more preferably covalently attached to one another. The helicases may be attached in any order and using any method. The one or more molecular brakes derived from helicases are preferably modified to reduce the size of an opening in the polynucleotide

binding domain through which in at least one conformational state the polynucleotide can unbind from the helicase. This is disclosed in WO 2014/013260.

Preferred helicase constructs for use in the invention are described in International Application Nos. PCT/GB2013/051925 (published as WO 2014/013260); PCT/GB2013/051924 (published as WO 2014/013259) and PCT/GB2013/051928 (published as WO 2014/013262); and in UK Application No. 1318464.3 filed on 18 October 2013.

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

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

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

spacer. The one or more DNA helicases and the one or more molecular brakes are preferably brought together when the one or more DNA helicases reach the hairpin loop. The one or more DNA helicases may be attached to the Y adaptor before the Y adaptor is attached to the polynucleotide or after the Y adaptor is attached to the polynucleotide. The one or more molecular brakes may be attached to the hairpin loop adaptor before the hairpin loop adaptor is attached to the polynucleotide or after the hairpin loop adaptor is attached to the polynucleotide.

The one or more helicases and the one or more molecular brakes are preferably not attached to one another. The one or more helicases and the one or more molecular brakes are more preferably not covalently attached to one another. The one or more helicases and the one or more molecular brakes are preferably not attached as described in International Application Nos. PCT/GB2013/051925 (published as WO 2014/013260); PCT/GB2013/051924 (published as WO 2014/013259) and PCT/GB2013/051928 (published as WO 2014/013262); and in UK Application No. 1318464.3 filed on 18 October 2013.

Spacer(s)

One or more spacers can be included in the constructs of the invention. When a part of the RNA polynucleotide enters the pore and moves through the pore along the field resulting from the applied potential, the one or more helicases are moved past the spacer by the pore as the RNA polynucleotide moves through the pore. This is because the RNA polynucleotide (including the one or more spacers) moves through the pore and the one or more helicases remain on top of the pore. The one or more DNA helicases may be stalled at the one or more spacers as discussed in International Application No. PCT/GB2014/050175 (published as WO 2014/135838). Any configuration of one or more helicases and one or more spacers disclosed in the International Application may be used in this invention.

The one or more spacers can be part of the target RNA polynucleotide, for instance it/they interrupt(s) the polynucleotide sequence. The one or more spacers are preferably not part of one or more blocking molecules, such as speed bumps, hybridised to the target RNA. The one or more spacers can be part of the non-RNA polynucleotide (e.g., a DNA polynucleotide), for instance it/they interrupt(s) the polynucleotide sequence. The one or more spacers can be part of the RNA polynucleotide. The one or more spacers can be attached to the target RNA polynucleotide and/or the non-RNA polynucleotide. The one or more spacers may be positioned at the ends or the RNA polynucleotide or non-RNA

polynucleotide and/or the one or more spacers may be positioned within the RNA polynucleotideor non-RNA polynucleotide.

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

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

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

The one or more spacers typically comprise a linear molecule, such as a polymer. The one or more spacers typically have a different structure from the target RNA polynucleotide or non-RNA polynucleotide. For instance, the one or more spacers are typically not RNA. In particular, the one or more spacers preferably comprise peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or a synthetic polymer with nucleotide side chains. The one or more spacers may comprise one or more nucleotides in the opposite direction from the polynucleotide. For instance, the one or more spacers may comprise one or more nucleotides in the 3' to 5' direction when the polynucleotide is in the 5' to 3' direction. The nucleotides may be any of those discussed above.

The one or more spacers preferably comprises one or more nitroindoles, such as one or more 5-nitroindoles, one or more inosines, one or more acridines, one or more 2-aminopurines, one or more 2-6-diaminopurines, one or more 5-bromo-deoxyuridines, one or more inverted thymidines (inverted dTs), one or more inverted dideoxy-thymidines (ddTs),

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

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

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

The one or more spacers preferably comprise one or more abasic nucleotides (i.e. nucleotides lacking a nucleobase), such as 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more abasic nucleotides. The nucleobase can be replaced by –H (idSp) or –OH in the abasic nucleotide. Abasic spacers can be inserted into target polynucleotides by removing the nucleobases from one or more adjacent nucleotides. For instance, polunucleotides may be modified to include 3-methyladenine, 7-methylguanine, 1,N6-ethenoadenine inosine or hypoxanthine and the nucleobases may be removed from these nucleotides using Human Alkyladenine DNA Glycosylase (hAAG). Alternatively, polunucleotides may be modified to include uracil and the nucleobases removed with Uracil-DNA Glycosylase (UDG). In one embodiment, the one or more spacers do not comprise any abasic nucleotides.

The one or more DNA helicases may be stalled by (i.e. before) or on each linear molecule spacers. If linear molecule spacers are used, the construct is preferably provided with a double stranded region of polynucleotide adjacent to the end of each spacer past which the one or more helicases are to be moved. The double stranded region typically helps to stall the one or more helicases on the adjacent spacer. The presence of the double stranded

region(s) is particularly preferred if the method is carried out at at a salt concentration of about 100 mM or lower. Each double stranded region is typically at least 10, such as at least 12, nucleotides in length. If the target polynucleotide used in the invention is single stranded, a double stranded region may formed by hybridising a shorter polynucleotide to a region adjacent to a spacer. The shorter polynucleotide is typically formed from the same nucleotides as the target polynucleotide, but may be formed from different nucleotides. For instance, the shorter polynucleotide may be formed from LNA.

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

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

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

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

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

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

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

Preferred combinations of features are shown in the Table 1 below.

Table 1:

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

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

Transmembrane pore

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

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

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

The transmembrane protein pore may be a monomer or an oligomer. The pore is preferably made up of several repeating subunits, such as 6, 7, 8 or 9 subunits. The pore is preferably a hexameric, heptameric, octameric or nonameric pore.

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

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

Transmembrane protein pores for use in accordance with the invention can be derived from β-barrel pores or α-helix bundle pores. β-barrel pores comprise a barrel or channel that is formed from β-strands. Suitable β-barrel pores include, but are not limited to, β-toxins, such as α-hemolysin, anthrax toxin and leukocidins, and outer membrane proteins/porins of bacteria, such as *Mycobacterium smegmatis* porin (Msp), for example MspA, MspB, MspC or MspD, CsgG, 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 similarity or 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 similarity or 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 similarity or 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/).

Similarity can be measured using pairwise identity or by applying a scoring matrix such as BLOSUM62 and converting to an equivalent identity. Since they represent functional rather than evolved changes, deliberately mutated positions would be masked when determining homology. Similarity may be determined more sensitively by the application of position-specific scoring matrices using, for example, PSIBLAST on a comprehensive database of protein sequences. A different scoring matrix could be used that

reflect amino acid chemico-physical properties rather than frequency of substitution over evolutionary time scales (e.g. charge).

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

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 also discussed therein.

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).

The variant may include modifications that facilitate covalent attachment to or interaction with the helicase or construct. The variant preferably comprises one or more reactive cysteine residues that facilitate attachment to the helicase or 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.

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.

Membrane

Any membrane may be used in accordance with the invention. Suitable membranes are well-known in the art. The membrane is preferably an amphiphilic layer. An amphiphilic

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

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

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

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

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

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

In another preferred embodiment, the membrane is a solid state layer. A solid-state layer is not of biological origin. In other words, a solid state layer is not derived from or isolated from a biological environment such as an organism or cell, or a synthetically manufactured version of a biologically available structure. Solid state layers can be formed from both organic and inorganic materials including, but not limited to, microelectronic materials, insulating materials such as Si_3N_4, $A1_2O_3$, 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*.

Coupling

The target RNA polynucleotide is preferably coupled to the membrane comprising the transmembrane pore. This may be done using any known method. The method may comprise coupling the target RNA polynucleotide to the membrane comprising the transmembrane pore. The RNA polynucleotide is preferably coupled to the membrane using one or more anchors. The RNA polynucleotide may be coupled to the membrane using any known method.

Each anchor comprises a group which couples (or binds) to the RNA polynucleotide and a group that couples (or binds) to the membrane. Each anchor may covalently couple (or bind) to the RNA polynucleotide and/or the membrane. If a Y adaptor and/or a hairpin loop adaptors are used, the RNA is preferably coupled to the membrane using the adaptor(s).

The RNA polynucleotide may be coupled to the membrane using any number of anchors, such as 2, 3, 4 or more anchors. For instance, an RNA polynucleotide may be coupled to the membrane using two anchors each of which separately couples (or binds) to both the RNA polynucleotide and membrane.

The one or more anchors may comprise the one or more DNA helicases and/or the one or more molecular brakes discussed above.

If the membrane is an amphiphilic layer, such as a lipid bilayer (as discussed in detail above), the RNA 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 RNA polynucleotide may be coupled directly to the membrane. The RNA polynucleotide may be coupled to the membrane using any of the methods disclosed in International Application Number No. PCT/GB2012/051191 (published as WO 2012/164270). The RNA 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 an RNA is coupled directly to the membrane, then some data will be lost as the characterising run cannot continue to the end of the RNA due to the distance between the membrane and the pore and/or polynucleotide binding protein. If a linker is used, then the RNA can be processed to completion. If a linker is used, the linker may be attached to the RNA at any position. The linker is typically attached to the RNA 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 an RNA, then some data will be lost as the characterising run cannot continue to the end of the complementary polynucleotide due to the distance between the membrane and the pore and/or polynucleotide binding protein. If the coupling is transient, then when the coupled end randomly becomes free of the membrane, then the RNA polynucleotide can be processed to completion. Chemical groups that form stable or transient links with the membrane are discussed in more detail below. The RNA 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.

Coupling of polynucleotides to synthetic lipid bilayers has been carried out previously with various different tethering strategies. These are summarised in Table 2 below.

Table 2

Attachment group	Type of coupling	Reference
Thiol	Stable	Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." 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

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

Coupling of RNA polynucleotides can also be achieved by a number of other means provided that a reactive group can be added to the RNA polynucleotide.

Alternatively, the reactive group could be considered to be a short region in the RNA polynucleotide complementary to one already coupled to the membrane, so that attachment can be achieved via hybridisation. The region could be part of the RNA polynucleotide or ligated to it. Ligation of short pieces of ssDNA have been reported using T4 RNA ligase I (Troutt, A. B., M. G. McHeyzer-Williams, et al. (1992). "Ligation-anchored PCR: a simple amplification technique with single-sided specificity." Proc Natl Acad Sci U S A 89(20): 9823-5).

Most preferably, the RNA is coupled to the membrane using a cholesterol-tagged polynucleotide which hybridises to the RNA polynucleotide or non-RNA polynucleotide attached thereto.

Diagnosing or prognosing diseases or conditions

mRNA is preferably used in the invention to diagnose or prognose a disease or condition. Some diseases or conditions are associated with an altered amount (or level) of mRNA. The mRNA may be normal or wild-type mRNA, *i.e.* not alternately spliced. The amount (or level) of the mRNA may be increased or decreased in the disease or condition compared with the amount (or level) in a patient without the disease or condition. Such diseases or conditions may be diagnosed or prognosed by determining the amount of the mRNA in a sample from the patient using a method of the invention.

Many genetic diseases or conditions are caused by mutations that cause alternate mRNA splicing, such as mRNA splicing defects. A number of diseases or conditions are associated with alternate mRNA splicing which are not attributed to overt mutations. The presence or absence of alternate splicing can be identified by determining the presence or absence of an alternately spliced mRNA in a sample from the patient using the method of the invention. In some instances, alternate mRNA splicing may be the normal function of a cell. In such instances, an increased or decreased amount (or level) of the alternately spliced mRNA compared with the normal amount (*i.e.* the amount in a patient without the disease or condition) may be used to diagnose or prognose the disease or condition.

The invention provides a method of diagnosing or prognosing a disease or condition associated with an altered amount and/or alternate splicing of messenger RNA (mRNA) in a patient. The invention provides a method of determining whether or not a patient has or is at risk of developing a disease or condition associated with an altered amount and/or alternate splicing of messenger RNA (mRNA). In each instance, the method comprises determining the amount and/or identity of the mRNA in a sample from the patient using a method of the invention. The disease or condition may be any of those discussed below. The disease or condition is preferably cystic fibrosis, familial dysautonomia, frontotemporal lobar dementia, amyotrophic lateral sclerosis, Hutchinson–Gilford progeria syndrome, medium-chain acyl-CoA dehydrogenase (MCAD) deficiency, myotonic dystrophy, Prader–Willi syndrome, spinal muscular atrophy, tauopathy, hypercholesterolemia or cancer. These diseases, their causes and possible treatments are discussed in Tazi *et al.* (Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, Volume 1792, Issue 1, January 2009, Pages 14–26).

The presence of an altered (*i.e.* increased or decreased) amount (or level) of the mRNA in the sample from the patient typically diagnoses or prognoses the disease or condition, *i.e.* indicates that the patient has or is at risk of developing the disease or condition. The absence of an altered (*i.e.* increased or decreased) amount (or level) of the mRNA in the sample from the patient typically indicates that the patient does not have or is not at risk of developing the disease or condition. The amount of mRNA can be determined as discussed above.

The presence of the alternately spliced mRNA in the sample from the patient typically diagnoses or prognoses the disease or condition, *i.e.* indicates that the patient has or is at risk of developing the disease or condition. The absence of the alternately spliced mRNA in the sample from the patient typically indicates that the patient does not have or is not at risk of developing the disease or condition. The presence or absence of the alternately spliced mRNA can be determined by identifying RNA in the sample as discussed above.

An increased or decreased amount (or level) of the alternately spliced mRNA in the sample from the patient typically diagnoses or prognoses the disease or condition, *i.e.* indicates that the patient has or is at risk of developing the disease or condition. No change in the amount of the alternately spliced mRNA in the sample from the patient (compared with the amount or level in a patient without the disease or condition) typically indicates that the patient does not have or is not at risk of developing the disease or condition. The amount of the alternately spliced mRNA can be determined as discussed above.

miRNA is preferably used in the invention to diagnose or prognose a disease or condition. The invention provides a method of diagnosing or prognosing a disease or condition associated with a miRNA. The invention provides a method of determining whether or not a patient has or is at risk of developing a disease or condition associated with a miRNA. The method comprises determining the presence or absence of the miRNA in a sample from the patient using a method of the invention. The disease or condition may be any of those discussed below.

The presence of the miRNA in the sample from the patient typically indicates that the patient has or is at risk of developing the disease or condition. The absence of the miRNA in the sample from the patient typically indicates that the patient does not have or is not at risk of developing the disease or condition. The presence or absence of the miRNA can be determined by identifying any miRNAs in the sample as discussed above.

The disease or condition is preferably cancer, coronary heart disease, cardiovascular disease or sepsis. The disease or condition is more preferably abdominal aortic aneurysm,

acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), acute myocardial infarction, acute promyelocytic leukemia (APL), adenoma, adrenocortical carcinoma, alcoholic liver disease, Alzheimer's disease, anaplastic thyroid carcinoma (ATC), anxiety disorder, asthma, astrocytoma, atopic dermatitis, autism spectrum disorder (ASD), B-cell chronic lymphocytic leukemia, B-cell lymphoma, Becker muscular dystrophy (BMD), bladder cancer, brain neoplasm, breast cancer, Burkitt lymphoma, cardiac hypertrophy, cardiomyopathy, cardiovascular disease, cerebellar neurodegeneration, cervical cancer, cholangiocarcinoma, cholesteatoma, choriocarcinoma, chronic lymphocytic leukemia, chronic myeloid leukemia, chronic pancreatitis, colon carcinoma, colorectal cancer, congenital heart disease, coronary artery disease, cowden syndrome, dermatomyositis (DM), diabetic nephropathy, diarrhea predominant irritable bowel syndrome, diffuse large B-cell lymphoma, dilated cardiomyopathy, down syndrome (DS), duchenne muscular dystrophy (DMD), endometrial cancer, endometrial endometrioid adenocarcinoma, endometriosis, epithelial ovarian cancer, esophageal cancer, esophagus squamous cell carcinoma, essential thrombocythemia (ET), facioscapulohumeral muscular dystrophy (FSHD), follicular lymphoma (FL), follicular thyroid carcinoma (FTC), frontotemporal dementia, gastric cancer (stomach cancer), glioblastoma, glioblastoma multiforme (GBM), glioma, glomerular disease, glomerulosclerosis, hamartoma, HBV-related cirrhosis, HCV infection, head and neck cancer, head and neck squamous cell carcinoma (HNSCC), hearing loss, heart disease, heart failure, hepatitis B, hepatitis C , hepatocellular carcinoma (HCC), hilar cholangiocarcinoma, Hodgkin's lymphoma , homozygous sickle cell disease (HbSS), Huntington's disease (HD), hypertension, hypopharyngeal cancer, inclusion body myositis (IBM), insulinoma, intrahepatic cholangiocarcinoma (ICC), kidney cancer, kidney disease, laryngeal carcinoma, late insomnia (sleep disease), leiomyoma of lung, leukemia, limb-girdle muscular dystrophies types 2A (LGMD2A), lipoma, lung adenocarcinoma, lung cancer, lymphoproliferative disease, malignant lymphoma, malignant melanoma, malignant mesothelioma (MM), mantle cell lymphoma (MCL), medulloblastoma, melanoma, meningioma, metabolic disease, miyoshi myopathy (MM), multiple myeloma (MM), multiple sclerosis, MYC-rearranged lymphoma, myelodysplastic syndrome, myeloproliferative disorder, myocardial infarction, myocardial injury, myoma, nasopharyngeal carcinoma (NPC), nemaline myopathy (NM), nephritis, neuroblastoma (NB), neutrophilia, Niemann-Pick type C (NPC) disease, non-alcoholic fatty liver disease (NAFLD), non-small cell lung cancer (NSCLC), obesity, oral carcinomaosteosarcoma ovarian cancer (OC), pancreatic cancer, pancreatic ductal adenocarcinoma (PDAC), pancreatic neoplasia, panic disease,

papillary thyroid carcinoma (PTC), Parkinson's disease, PFV-1 infection, pharyngeal disease, pituitary adenoma, polycystic kidney disease , polycystic liver disease, polycythemia vera (PV), polymyositis (PM), primary biliary cirrhosis (PBC), primary myelofibrosis, prion disease, prostate cancer, psoriasic arthritis, psoriasis, pulmonary hypertension, recurrent ovarian cancer, renal cell carcinoma, renal clear cell carcinoma, retinitis pigmentosa (RP), retinoblastoma, rhabdomyosarcoma, rheumatic heart disease and atrial fibrillation, rheumatoid arthritis, sarcoma, schizophrenia, sepsis, serous ovarian cancer, Sezary syndrome, skin disease, small cell lung cancer, spinocerebellar ataxia, squamous carcinoma, T-cell leukemia, teratocarcinoma, testicular germ cell tumor , thalassemia, thyroid cancer, tongue squamous cell carcinoma, tourette's syndrome, type 2 diabetes, ulcerative colitis (UC), uterine leiomyoma (ULM), uveal melanoma, vascular disease, vesicular stomatitis or Waldenstrom macroglobulinemia (WM).

The patient may be any of the mammals discussed above. The patient is preferably human. The patient is an individual.

The sample may be any of those discussed above. The sample is typically from any tissue or bodily fluid. The sample typically comprises a body fluid and/or cells of the patient and may, for example, be obtained using a swab, such as a mouth swab. The sample may be, or be derived from, blood, urine, saliva, skin, cheek cell or hair root samples. The target RNA is typically extracted from the sample before it is used in the method of the invention.

The method may concern diagnosis of the disease or condition in the patient, *i.e.* determining whether or not the patient has the disease or condition. The patient may be symptomatic.

The method may concern prognosing the disease or condition in the patient, *i.e.* determining whether or not the patient is likely to develop the disease or condition. The patient can be asymptomatic. The patient can have a genetic predisposition to the disease or condition. The patient may have one or more family member(s) with the disease or condition.

Method of improving the movement of an RNA polynucleotide

The present invention also provides a method of moving a target RNA polynucleotide with respect to a transmembrane pore when the movement is controlled by a DNA helicase enzyme, comprising:

a) providing (i) an RNA polynucleotide wherein the RNA is modified to comprise a non-RNA polynucleotide and (ii) a DNA helicase enzyme;

b) contacting the RNA polynucleotide and DNA helicase enzyme provided in a) with a transmembrane pore such that the DNA helicase controls the movement of the RNA polynucleotide with respect to the transmembrane pore.

The modification of the RNA polynucleotide results in increased DNA helicase binding thereto. Increased DNA helicase binding to the modified RNA polynucleotide is defined as an amount or level of DNA helicase binding that is greater than, or more than, the amount or level of DNA helicase binding that is observed for non-modified or unmodified RNA polynucleotide i.e., an RNA that has not been modified in accordance with the modification methods of the invention. The level of binding of DNA helicase to a target RNA polynucleotide can be easily tested using routine methods which are known and routine to one of skill in the art.

Preferably the DNA helicase enzyme is pre-bound to the non-RNA polynucleotide. Any of the embodiments described above also apply to this method. For example, in one embodiment, a non-RNA polynucleotide may comprise at least one of-(i) a polymer of 5 or more charged units; (ii) a blocking-strand hybridisation site of approximately 20 nucleotides in length; (iii) a DNA-helicase binding site of 1 or more non-RNA nucleotides, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, or 50 nucleotides; (iv) a stalling chemistry of 1 or more units e.g., 1, 2, 3, 4, 5, 6, 7, 8,9 ,10 or more units, such as Sp18, as described in WO2014/135838 which is incorporated by reference herein; (v) a tether hybridisation site of approximately 30 nucleotides in length; and/or (vi) a sequence that facilitates ligation of the non-RNA polynucleotide to the RNA polynucleotide, as described in the preceding sections.

The Examples further illustrate the use of a DNA helicase to control the movement of a DNA/RNA strand through a nanopore. Accordingly, in one embodiment, provided herein is a method for increasing the ability or efficiency of an RNA to be sequenced through a pore.

Methods of producing constructs of the invention

The method of producing a construct comprises attaching a target RNA polynucleotide to a non-RNA polynucleotide. At least one nucleotide of the non-RNA polynucleotide is not a ribonucleotide, i.e. is not from RNA. The non-RNA polynucleotide may therefore comprise at least one ribonucleotide (or RNA nucleotide) but must also additionally comprise or include a non-RNA nucleotide or sequence i.e., a nucleotide or sequence of nucleotides that is not RNA. The non-RNA polynucleotide may comprise any of the embodiments as described above. Preferably the non-RNA polynucleotide comprises a

DNA helicase binding site or a DNA adaptor. More preferably the non-RNA polynucleotide comprises a leader sequence.

The site of and method of attachment are selected as discussed above. Preferably, the attachment method may be selected from chemical attachment, covalent attachment, enzymatic attachment, hybridization, synthetic methods, or using a topoisomerase. The RNA polynucleotide may be attached to the non-RNA polynucleotide at more than one, such as two or three, points. The method of attachment may involve one, two, three, four, five or more different methods of attachment. Any combination of the attachment methods described above may be used in accordance with the invention.

The method may further comprise determining whether or not the construct is capable of controlled movement through a nanopore under the control of a DNA helicase. Assays for testing this are known to those of skill in the art. If the movement of the RNA polynucleotide through a nanopore can be controlled, a construct of the invention has been produced. If the movement of the RNA polynucleotide cannot be controlled, a construct of the invention has not been produced.

In one embodiment of the invention the method of producing the construct comprising attaching a target RNA polynucleotide to a non-RNA polynucleotide takes place before the contacting step (b). In another embodiment, provided herein is a method of generating a modified RNA capable of being sequenced through a pore with greater efficiency than the RNA in unmodified form.

Modification method of the invention

The present invention provides a method of modifying a target RNA polynucleotide for characterisation, such as for sequencing. The modified RNA polynucleotide is characterised, or sequenced, in accordance with the invention. This is discussed in more detail above.

The method involves the formation of one or more modified RNA polynucleotides using one or more of the methods described. The one or more modified polynucleotides are easier to characterise than the unmodified polynucleotide, especially using strand sequencing.

Products/ Constructs of the invention

The invention also provides an RNA polynucleotide modified using a modification method of the invention. The target RNA polynucleotide is modified by attachment of the RNA polynucleotide to a non-RNA polynucleotide to form a construct of the invention. The

modification of the target RNA necessarily results in increased interaction between a DNA helicase and the modified RNA construct i.e., as compared with interaction that occurs between DNA helicase and RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Additionallly or alternatively, the addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that the specificity of DNA helicase for the modified RNA construct is increased i.e., as compared to the specificity of the DNA helicase for the RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Additionally or alternatively, the addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that DNA helicase binding to the modified RNA construct is facilitated and/or DNA helicase binding to the modified RNA construct is increased i.e., as compared with binding that occurs between DNA helicase and RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide. Additionally or alternatively, the addition or attachment of the non-RNA polynucleotide to the RNA polynucleotide means that the DNA helicase binds more efficiently or more strongly to the modified RNA construct and is less likely to disengage from the modified construct i.e., as compared with binding that occurs between DNA helicase and RNA polynucleotide in un-modified form, without the attached non-RNA polynucleotide.

The non-RNA polynucleotide may be any polynucleotide which is not RNA. The non-RNA polynucleotide may comprise at least one ribo nucleotide but must also additionally comprise or include a non-RNA nucleotide or sequence i.e., a nucleotide or sequence of nucleotides that is not RNA. The site of and method of attachment are selected as discussed above. The non-RNA polynucleotide may or may not comprise a pre-bound DNA helicase. Preferably the non-RNA polynucleotide comprises a region to which a DNA helicase is capable of binding (a DNA helicase binding site) or a DNA adaptor. More preferably the DNA helicase binding site or the DNA adaptor comprises a leader sequence which preferentially threads into a nanopore. The leader sequence can also be used to link the target RNA to the one or more anchors as discussed above. The leader sequence may be linked to the target RNA polynucleotide.

The construct of the invention is preferably a DNA/RNA hybrid strand which can be translocated through an MspA nanopore using a DNA helicase.

The construct may further comprise a barcoding section on the polynucleotide strand. Polynucleotide barcodes are well-known in the art (Kozarewa, I. *et al.*, (2011), *Methods Mol. Biol.* **733**, p279-298). A barcode is a specific sequence of polynucleotide that affects the current flowing through the pore in a specific and known manner. The barcoding section

enables unambiguous identification of an analyte. Preferably the barcoding section is located between the leader sequence and the DNA helicase binding site.

An anchor e.g., a DNA anchor may be hybridised to the RNA polynucleotide or the non-RNA polynucleotide, as described above. The DNA anchor may further comprise spacers and cholesterol.

The RNA polynucleotide may be further extended using Poly (U) Polymerase. This ensures that the full length of the RNA is read.

The modified RNA polynucleotide may come in a variety of forms depending on which modification method(s) of the invention is used. Possible forms, include, but are not limited to, one or more of the following:

- an RNA polynucleotide chemically attached to a non-RNA polynucleotide e.g., using click chemistry
- an RNA polynucleotide ligated to a non- RNA polynucleotide
- an RNA polynucleotide hybridised to a non RNA leader sequence, with or without a briding moiety
- an RNA polynucleotide hybridised to a cDNA sequence, with or without a briding moiety
- an RNA polynucleotide attached to a non-RNA polynucleotide using a topoisomerase
- an RNA polynucleotide ligated to a non-RNA polynucleotide which has a region with reversed bases such that it runs in the opposite direction to the remainder of the non-RNA polynucleotide. Preferably the non-RNA polynucleotide is a DNA polynucleotide.

Kits

The invention also provides a kit for characterising a target RNA polynucleotide. The kit comprises a non-RNA polynucleotide which is adapted to attach to any target RNA polynucleotide for characterisation.

Preferably the non-RNA polynucleotide which is adapted to attach to any target RNA polynucleotide for characterisation has a reactive group attached, e.g., a click reactive group. The non-RNA polynucleotide with reactive group attached can ultimately be used by the end-user to react with any target RNA polynucleotide of choice to form a covalent bond. Preferably the target RNA polynucleotide also has a reactive group attached, .e.g., a click

reactive group which reacts with the reactive group attached to the non-RNA polynucleotide to form a covalent bond.

Alternatively the non-RNA polynucleotide which is adapted to attach to any target RNA polynucleotide for characterisation is provided with a ligase or is an oligonucleotide or primer which can be used by the end user to hybridise to any region of the target RNA polynucleotide of choice and act as a starting point for cDNA synthesis.

Alternatively the non-RNA polynucleotide which is adapted to attach to any target RNA polynucleotide for characterisation comprises a topoisomerase bound to a specific DNA polynucleotide. The topoisomerase bound DNA can ultimately be used by the end-user to attach the non-RNA (DNA) polynucleotide to any target RNA polynucleotide of choice. The end user can incubate the topoisomerase bound DNA with RNA which has a free 5' hydroxyl. The topoisomerase then joins the RNA to the DNA.

Alternatively the non-RNA polynucleotide which is adapted to attach to any target RNA polynucleotide for characterisation comprises a region with reversed bases (e.g A DNA region of reversed bases). This reversed region can be attached to the 5' ends of eukaryotic RNA which are modified by the addition of a 7-methylguanosine cap which runs in the opposite orientation.

Any of the embodiments discussed above with reference to the method of the invention equally apply to the kits. The kit may further comprise a DNA helicase binding protein which may be pre-bound to the non-RNA polynucleotide. The kit may further comprise a pore and 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), 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 typically comprises nucleotides. The kit preferably comprises dAMP, dTMP, dGMP and dCMP. The kit preferably does not comprise means to amplify and/or express polynucleotides.

The following Examples illustrate the invention.

Example 1

This example shows the sample preparation procedure which 1) extended the RNA region of an RNA/DNA strand, 2) annealed an anchor, 3) bound an enzyme and then 4) tested the resultant strand in an electrophysiology experiment. This example illustrated that it was possible to use a DNA helicase (T4 Dda – E94C/A360C (SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1)) to control the movement of a synthetic DNA/RNA strand (DNA leader attached to an RNA strand, shown in Figure 3) through an MspA nanopore.

Materials and Methods

1.1 **Extension of the 3' End of an DNA/RNA Strand using Poly(U) Polymerase**

The reagents listed in Table 3 below were mixed and incubated at 37 °C for 10 minutes. The mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 μL SPRI beads per μL of sample. This sample was known as sample 1 (DNA/RNA 2). Figure 4 shows that the polymerase extension reaction was successful as a broad band labelled Y in the figure corresponded to variably extended DNA/RNA 1.

Reagent	***Volume***	***Concentration of Stock***	***Final Concentration***
Synthetic DNA/RNA 1 (SEQ ID NO: 16 attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 17)	0.4 μl	100 μM	1 μM
rUTP	0.4 μl	100 mM	1 mM
NEBuffer	4 μl	10 x	1 x
PolyU Pol (NEB)	2 μl	2 U/μl	4 U
NF H_2O	33.2 μl		
Total	40 μl		

Table 3

1.2 **Anchor Anneal**

The reagents listed in Table 4 below were mixed and incubated at 65 °C and then cooled to 4 °C at a rate of 0.1 °C per second. This sample was known as Sample 2.

Reagent	*Volume*	*Concentration of Stock*	*Final Concentration*
Sample 1	9 µl	~1 µM	942 µM
Anchor (SEQ ID NO: 18 attached at its 3' end to six iSp18 spacers, two thymines and a 3' cholesterol TEG)	0.36 µl	100 µM	3.77 µM
10 mM TRIS pH 7.5 50 mM NaCl	0.19 µl	50 x	1 x
Total	9.55 µl		

Table 4

1.3 **Bind DNA Helicase**

Sample 2 (0.28 µL) was incubated with T4 Dda – E94C/A360C (0.36 µL, 3.8 µM, SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1) in buffer (10 mM TRIS pH 7.5, 50 mM NaCl) at room temperature for one hour. This sample was known as Sample 3.

1.4 **Electrophysiology**

Sample 3 was diluted into buffer (1221 µL of 600 mM KCl, 50 mM HEPES pH 8.0, 463 mM glycerol). MgCl2 (13 µL, 1 M) and ATP (65µL, 100 mM) were added to the sample 3 buffer mixture giving a total volume of 1300 µL.

Electrical measurements were acquired from single MspA nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide ~ pH 8.0). After achieving a single pore inserted in the block co-polymer, buffer (2 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide, pH 8.0) was flowed through the system to remove any excess MspA nanopores.

An excess of KCl buffer (600 mM KCl, 50 mM HEPES pH ~8, 463 mM glycerol) was flowed through the system and this KCl buffer was separated from an electrode buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide, pH 8.0) by an agarose bridge.

The experiment was run at - 120 mV and helicase-controlled DNA movement monitored.

Results

Helicase-controlled DNA movement was observed when Sample 3 (cartoon representation of the DNA/RNA 2 is shown in Figure 3) was added to the nanopore system. An example of helicase controlled DNA movement for Sample 3 is shown in Figure 5. The various regions of the synthetic strand were identified as the strand translocated through the nanopore (region 1 = poly(dT) leader (SEQ ID NO: 16), region 2 = iSpC3 spacers, region 3 = RNA sequence (SEQ ID NO: 17) and region 4 = the variable length poly(U) RNA). This example showed that it was possible to use a DNA helicase to control the movement of a DNA/RNA strand (cartoon representation shown in figure 3) through an MspA nanopore. The Poly (U) Polymerase extension step ensured that the full length of the RNA was read.

Example 2

This example shows the ligation of a DNA strand (SEQ ID NO: 21) to an RNA strand (SEQ ID NO: 19) using T4 DNA ligase.

Materials and Methods

2.1 Ligation of a DNA strand to an RNA Strand using T4 DNA Ligase

The reagents listed in Table 5 below were mixed and placed on a thermocycler. The thermocycler was set to the program in Table 6 below. The samples were then analysed using a 10% PAGE TBE-Urea denaturing BioRad Criterion Gel which was run at 140 V for 60 minutes.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
5' phosphate RNA (SEQ ID NO: 19)	0.3 ul	50 uM	1 uM
T4 DNA ligase buffer	1.5	10 x	1 x
T4 DNA ligase	1 ul	10 U/ul	10 U
DNA splint (SEQ ID NO: 20)	0.3 ul	50 uM	1 uM
Cy3 DNA (SEQ ID NO: 21)	0.6 ul	50 uM	2 uM
ATP	0.3 ul	50 mM	1 mM
NF H_2O	11 ul		
Total	15 ul		

Table 5

Number of Cycles	Step	Temp (°C)	Time
1	Ligate	25	1:00:00
1	Denature	65	0:10:00

Table 6

Results

The TBE-Urea denaturing gel was used to analyse the ligation of CY3 DNA(SEQ ID NO: 21) to RNA (SEQ ID NO: 19). Figure 8, lanes 2-3 showed the control reactions of the ligation step with increasing concentration of the DNA splint (SEQ ID NO: 20) in the absence of T4 DNA ligase. For these control reactions no bands in regions A and B were observed indicating that no ligation reaction occurred under these conditions. Lanes 5-8 showed the ligation step with increasing concentrations of the DNA splint (SEQ ID NO: 20) in the presence of T4 DNA ligase. For all of lanes 5-8 a band was visible in both region A and region B which corresponded to ligated substrate with the hybridised DNA splint (A) and the ligated substrate without the hybridised DNA splint (B). As the concentration of splint was increased the intensity of band A also increased. Two further control reactions were carried out shown in lanes 9 and 10. Lane 9 corresponded to the same sample as shown in lane 5 which was further treated with the addition of extra DNA splint (4.5x) added after the ligation step. This showed the expected increase in intensity of the upper band (corresponding to ligated product with the splint hybridised) and decrease in intensity of the lower band B (corresponding to ligated product without the splint hybridised) when compared to lane 5. Lane 10 corresponded to the same sample as shown in lane 6 that was further treated with ExoI at 37ºC for 30 minutes. Heating during ExoI treatment resulted in the separation of the ligated strand and the DNA splint. The DNA splint was preferentially digested by ExoI because it had a 3' DNA end. This resulted in the disappearance of the band in region A which corresponded to the ligated product hybridised to the DNA splint, as the DNA splint was digested by the ExoI. The band in region B was still visible after digestion with ExoI. This meant that the ligation step had been successful because if the DNA had not been ligated to the RNA then the DNA that had not ligated would have been digested by the ExoI and no band would have been visible in region B. Therefore, this example showed that it was possible to ligate DNA to RNA using a T4 DNA ligase.

Example 3

This example shows a sample preparation procedure that chemically attached a DNA strand to an RNA strand using click chemistry. This was carried out on two different samples one of which had a fluorescent group attached to the DNA, in order for the chemical attachment step to be confirmed using gel electrophoresis. The DNA/RNA strand that did not have a fluorescent group attached was then tested in an electrophysiology experiment. This example illustrated that it was possible to use a DNA helicase (T4 Dda – E94C/A360C (SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1)) to control the movement of an RNA strand that was attached to a non-RNA polynucleotide by copper-mediated click-chemistry (cartoon representation of the construct is shown in Figure 10) through an MspA or Lysenin nanopore.

Materials and Methods

3A.1 Click reaction of DNA X1 to RNA X1

The RNA X1, DNA X1 and splint X1 (listed in Table 7 below) were mixed in buffer (TRIS-NaCl (500 mM-2.5M) pH 8). The DNA X1, RNA X1 and splint X1 were annealed in a PCR machine (protocol heat to 55 °C and cool to 4 °C at 0.1 °C/s). CuSO4, Tris(3-hydroxypropyltriazolylmethylamine) and sodium ascorbate (Sigma A4034) were then added to the DNA X1/RNA X1/splint X1 mixture and the sample was then placed on a thermocycler. The thermocycler was set to the program in Table 8 below. The sample was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8µl SPRI beads per µL of sample. This sample was known as sample 3A (DNA/RNA 3A). This sample was then analysed on a 5% PAGE TBE BioRad Criterion Gel and in electrophysiology as described in 3.3 below.

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
Firefly luciferase mRNA (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail) from Trilink Biotech,CA	12.9 ul	1.1 uM	0.94 uM
DNA with Azide (DNA X1, SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end)	0.3 ul	50 uM	1 uM

Splint X1 (SEQ ID NO: 24)	0.6 ul	50 uM	2 uM
$CuSO_4$	0.3 ul	50 mM	1 mM
Tris(3-hydroxypropyltriazolylmethylamine)	0.3 ul	100 mM	2 mM
Sodium Ascorbate (Sigma A4034)	0.3 ul	200 mM	4 mM
TRIS-NaCl (500mM-2.5M) pH 8	0.3 ul	500 mM	10 mM
Total	15 ul		

Table 7

Number of Cycles	Step	Temp (°C)	Time
1	Click	23	0:30:00
1	Click	37	0:30:00

Table 8

3B.1 Click reaction of DNA X2 to RNA X1

The RNA X1, DNA X2 and splint X1 (listed in Table 9 below) were mixed in buffer (TRIS or MOPS (500mM-2.5M) pH 6.8-7). The DNA X2, RNA X1 and splint X1 were annealed in a PCR machine (protocol heat to 55 °C and cool to 4 °C at 0.1 °C/s). CuSO4, Tris(3-hydroxypropyltriazolylmethylamine) and sodium ascorbate (Sigma A4034) were then added to the DNA X2/RNA X1/splint X1 mixture and the sample was then placed on a thermocycler. The thermocycler was set to the program in Table 10 below. The sample was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8µl SPRI beads per µL of sample. This sample was known as sample 3B (DNA/RNA 3B). This sample was then analysed on a 5% PAGE TBE BioRad Criterion Gel.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
Firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail	12.9 ul	1.1 uM	0.94 uM
Cy3 DNA with Azide (DNA X2, SEQ ID NO: 25 which had a Cy3 attached to its 5' end and had a 3AzideN attached to the 3' end)	0.3 ul	50 uM	1 uM
Splint X1 (SEQ ID NO: 24)	0.6 ul	50 uM	2 uM
$CuSO_4$	0.3 ul	50 mM	1 mM
Cu Ligand	0.3 ul	100 mM	2 mM
Sodium Ascorbate (Sigma A4034)	0.3 ul	200 mM	4 mM
TRIS-NaCl (500mM-2.5M) pH 8	0.3 ul	500 mM	10 mM

Total	15 ul		

Table 9

Number of Cycles	Step	Temp (°C)	Time
1	Click	23	0:30:00
1	Click	37	0:30:00

Table 10

3.3 Electrophysiology

Tether was annealed to Sample 3A as per Example 1 (see section 1.2). Sample 3A (0.28µl) and T4 Dda – E94C/A360C (0.36 µl, 3.8 µM) , SEQ ID NO: 14 with mutations E94C/A360C and then (ΔM1)G1) were diluted into buffer (1221 µL of 500 mM KCl, 25 mM potassium phosphate pH 8.0). MgCl2 (13 µL, 1 M) and ATP (65µL, 100 mM) were added to the sample 3A buffer mixture giving a total volume of 1300 µL.

Electrical measurements were acquired from single MspA or Lysenin nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide ~ pH 8.0). After achieving a single pore inserted in the block co-polymer, then buffer (2 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide, pH 8.0) was flowed through the system to remove any excess MspA or Lysenin nanopores.

An excess of buffer (500 mM KCl, 25 mM potassium phosphate pH 8.0) was flowed through the system prior to the addition of sample. Finally, T4 Dda – E94C/A360C bound to sample 3A was then added to the nanopore system, the experiment was run at - 120 mV and helicase-controlled DNA movement monitored.

Results

A TBE-Urea denaturing gel was used to analyse the click reaction of DNA (either DNA X1 (Example 3A, SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end) or DNA X2 (Example 3B, SEQ ID NO: 25 which had a CY3 attached to its 5' end and had a 3AzideN attached to the 3' end)) to RNA X1. The DNA that was used in Example 3B was visible before the SYBR stain owing to the Cy3 group which was attached. Figure 9, lanes 3 and 9 showed the sample produced after example 3B's click reaction. The fluorescent band at the top of lane 3 (highlighted by a white box in Figure 9) showed that the DNA with

the Cy3 label had been attached to the mRNA (the fluorescent band for the DNA-only was at arrow 3 and the DNA bound to mRNA was at arrow 1). Lanes 2 and 8 showed the sample produced after the click reaction in Example 3A (no fluorescent group on the DNA). The band in lane 11 corresponded to unreacted DNA (arrow 2) and the DNA bound to mRNA was arrow 1. Therefore, this gel showed that the click reaction was successful at joining DNA (with a fluorescent label (X1) or without a fluorescent label (X2)) to mRNA (RNA X1).

Helicase-controlled RNA movement was observed when Sample 3A (cartoon representation of the DNA/RNA 3A is shown in Figure 10) was added to the nanopore system. An example of helicase controlled RNA movement through MspA is shown in Figure 11 (regions 2, 3 and 4) and movement through a Lysenin mutant in Figure 21. The various regions of the strand in Figure 11 were identified as the strand translocated through the nanopore. Taken together, regions 2, 3 and 4 represent the DNA leader reacted to the target RNA by click reaction (Region 2 = SEQ ID NO: 22 joined to four iSp18 spacers; Region 3 = SEQ ID NO: 23; Region 4 = Firefly luciferase mRNA with open reading frame SEQ ID NO:26 and the arrow marked with * highlights the linkage formed by the click reaction). Region 1 shows a separate nanopore translocation of a DNA leader not reacted to RNA (SEQ ID NO: 22 joined to four iSp18 spacers joined to SEQ ID NO:23). Similar features can be identified for the Lysenin trace in figure 21. This example showed that it was possible to use a DNA helicase to control the movement of a long mRNA strand through a nanopore by linking a non-RNA polynucleotide (SEQ ID NO: 22 joined to four iSp18 spacers joined to SEQ ID NO:23) to the long mRNA strand, (cartoon representation shown in figure 10) through an MspA or Lysenin nanopore.

Example 4

This example shows ligation of hairpin-forming oligos (3T hairpin = SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 27, or the 10T hairpin = SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 28) to the 3' end of an RNA strand (SEQ ID NO: 30) using T4 DNA ligase and subsequent reverse-transcription from the 3' end of the hairpin-forming oligo. This ligation and reverse-transcription demonstrated a method of constructing a RNA/cDNA construct (see Figure 13).

A poly T overhang at the 3' end of the hairpin-forming oligos (either the 3T or 10T hairpins) hybridizes to the poly A-tail of the mRNA and acts as a splint for efficient DNA to RNA ligation. The hairpin can act as a primer for subsequent reverse-transcription.

Materials and Methods

The reagents listed in Table 11 below were mixed and placed on a thermocycler. The thermocycler was set to the program in Table 12 below. The mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample. After purification, reverse transcription was performed using Life Technologies Super Script II: reagents in Table 13 were mixed according to the manufacturer's protocol and placed on a thermocycler set to the program in Table 14. Samples were then analysed using a 10% PAGE TBE-Urea denaturing BioRad Criterion Gel which was run at 140 V for 60 minutes.

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
RNA strand (SEQ ID NO: 30)	0.72 ul	556ng/ul	0.2 uM
polyT hairpin (3T hairpin or 10T hairpin described above)	0.4 ul	50uM	1 uM
T4 DNA ligase buffer	4 ul	5 x	1 x
T4 DNA ligase	1 ul	2000U	2000 U
NF H_2O	13.88 ul		
Total	20 ul		

Table 11

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Ligate	16	2:00:00

Table 12

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
RNA after polyT hairpin ligation	7 ul	35.5ng/ul	248.5ng/reaction
dNTPs	1 ul	10uM each	0.5 uM
NF H_2O	5 ul		
First-Strand Buffer	4 ul	5x	1x
0.1M DTT	2 ul	0.1M	0.01M
Super Script II	1ul	200U	200U
Total	20 ul		

Table 13

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Reverse Transcription	42	0:50:00
2	Denaturation	70	0:15:00

Table 14

Results

A TBE-Urea denaturing gel (see Figure 14) was used to analyse the 3T and 10T hairpin ligation and reverse transcription. In lane 5, partial ligation of the 3T-hairpin was visible in the form of an upward shift of the RNA strand band (SEQ ID NO: 30, the RNA strand is labelled as band A in Figure 14). After reverse transcription (Lane 6) the band was shifted below the level of the RNA strand (SEQ ID NO: 30, band A in figure 14) due to the fact that the double stranded construct of RNA/cDNA (shown in Figure 13) migrated faster than the single stranded RNA (SEQ ID NO: 30 shown as band A in Figure 14). In the lane 7, ligation of the 10T hairpin occurred with nearly 100% efficiency and was visible as a single band above the level of single stranded RNA strand (SEQ ID NO: 30 shown as band A in Figure 14). Lane 8 shows the sample after reverse transcription, primed by the 10T hairpin, where the ligated and reverse transcribed product was shifted downwards below the single stranded RNA band (SEQ ID NO: 30 shown as band A in Figure 14). A hybridization control sample in lane 9 (reaction mixture without T4 DNA ligase) was visible at the level of single stranded RNA (SEQ ID NO: 30 shown as band A in figure 14) and showed that a shift was observed only for the ligated product.

Example 5

This example shows how an RNA/cDNA construct was analysed in an electrophysiology experiment. To obtain the RNA/cDNA construct the 10T hairpin (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 28) was ligated to the 3' end of firefly luciferase mRNA (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail). Prior to the ligation, click chemistry was used to attach DNA with Azide (DNA X1, SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end) to the 5' end of Firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail) as described in Example 3. After hairpin ligation reverse transcription was performed.

Materials and Methods

Click reaction

The RNA of Firefly luciferase mRNA with a 5'-hexynl-G (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail) was ligated to DNA with Azide (DNA X1, SEQ ID NO: 22 attached at its 3' end to four iSp18 spacers attached at their opposite end to the 5' end of SEQ ID NO: 23 which has a 3AzideN attached to the 3' end) as described in Example 3B click reaction. This sample was known as Sample 3B.

Hairpin ligation and RT

The sample 3B (DNA/RNA 3B) was then ligated to the 10T hairpin and reverse transcribed as described in Example 4. The volumes and quantities of reagents are shown in Tables 15 and 17 below and the thermocycle conditions in tables 16 and 18. The sample produced after ligation/revers transcription was known as sample 5B. A cartoon representation of the sample construct is shown in Figure 15.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
Sample 3B	10 ul	22ng/ul	0.017 uM
polyT hairpin (3T hairpin or 10T hairpin described above)	0.4 ul	50uM	1 uM
T4 DNA ligase buffer	4 ul	5 x	1 x

T4 DNA ligase	1 ul	2000U/ul	2000 U
NF H_2O	4.6 ul		
Total	20 ul		

Table 15

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Ligate	16	2:00:00

Table 16

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
Sample 3B after reverse transcription	12 ul	20ng/ul	240ng/reaction
dNTPs	1 ul	10uM each	0.5 uM
First-Strand Buffer	4 ul	5x	1x
0.1M DTT	2 ul	0.1M	0.01M
Super Script II	1ul	200U/ul	200U
Total	20 ul		

Table 17

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Reverse Transcription	42	0:50:00
2	Denaturation	70	0:15:00

Table 18

Electrophysiology

A tether was annealed to Sample 5B as described in Example 1 (see section 1.2). Sample 5B (4µl) and (T4 Dda – E94C/C109A/C136A/A360C (0.36 µl, 3.8 µM, SEQ ID NO: 14 with mutations E94C/C109A/C136A/A360C and then (ΔM1)G1) were diluted into buffer (1221 µL of 500 mM KCl, 25 mM potassium phosphate pH 8.0). MgCl2 (13 µL, 1 M) and ATP (65µL, 100 mM) were added to the sample 5B (DNA/RNA/cDNA 5B) buffer mixture giving a total volume of 1300 µL.

Electrical measurements were acquired from single MspA or CsgG-Eco-(Y51T/F56Q)-StrepII(C))9 (SEQ ID NO: 44 with mutations Y51T/F56Q where StepII(C) is SEQ ID NO: 45 and is attached at the C-terminus) nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide ~ pH 8.0). After achieving a single pore inserted in the block co-polymer, then buffer (2 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium

Ferricyanide, pH 8.0) was flowed through the system to remove any excess MspA or CsgG nanopores.

An excess of buffer (500 mM KCl, 25 mM potassium phosphate pH 8.0) was flowed through the system prior to the addition of sample. Finally, (T4 Dda – E94C/C109A/C136A/A360C bound to sample 5B (DNA/RNA/cDNA 5B) was then added to the nanopore system, the experiment was run at - 140 mV and helicase-controlled DNA movement monitored.

Results

Helicase-controlled DNA/RNA/cDNA movement was observed when Sample 5B was added to the nanopore system. An example of a helicase-controlled DNA/RNA/cDNA movement event is shown in Figure 16 (regions 1-5) for MspA or Figure 22 for CsgG. The various regions of the strand in figure 16 were identified as the strand translocated through the nanopore; Region 1 represented the non-RNA polynucleotide reacted to the target RNA by click reaction; Region 2 = Firefly luciferase mRNA; Region 3 = the iSpC3 spacers present in the 10T hairpin which was ligated to the mRNA; Region 4 = the polyT region of the 10T hairpin which was ligated to the mRNA; Region 5 = the cDNA which was produced by reverse transcription of the mRNA. Similar features can be identified in the CsgG trace in Figure 22. This example showed that it was possible to use a DNA helicase to control the movement of a strand DNA/RNA/cDNA (formed by linking a DNA strand to an mRNA strand, ligating a hairpin to that mRNA strand and using the hairpin to reverse transcribe the mRNA) (cartoon representation shown in figure 15) through an MspA or CsgG nanopore.

Example 6

This example shows how a non-RNA polynucleotide can be attached to an RNA strand that has a 5' methylguanosine cap (homologous to some native cellular mRNAs) by removing the cap before ligation. In this case, a capped RNA strand (SEQ ID NO: 30 which has a 7-methylguanosine cap connected to the 5' end of the strand by a 5' to 5' triphosphate linkage) was used, which was then decapped using RNA 5' pyrophosphohydrolase (RppH) and subsequently ligated to a non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG) using T4 RNA ligase 1.

Materials and Methods

In order to ligate a non-RNA polynucleotide to the 5' end of the capped RNA strand (SEQ ID NO: 30 which has a 7-methylguanosine cap connected to the 5' end of the strand by a 5' to 5' triphosphate linkage) the mRNA was first decapped using RppH as a decapping enzyme. The reagents listed in Table 19 were mixed and the reaction mixture was then placed on a thermocycler set to the program in Table 20. The resulting reaction mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample. Subsequently, a non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG) was ligated to the RNA by mixing the reagents listed in a Table 21 and placing the mixture on a thermocycler set to the program in Table 22. The reaction mixture was then analysed using a 5% PAGE TBE-Urea denaturing BioRad Criterion Gel which was run at 140 V for 60 minutes.

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
Capped RNA strand (SEQ ID NO: 30 which has a 7-methylguanosine cap connected to the 5' end of the strand by a 5' to 5' triphosphate linkage)	5.39 ul	556ng/ul	2996.8ng/reaction
NEB2 buffer	2 ul	10x	1x
RNA 5' Pyrophosphohydrolase	2 ul	5U/ul	10U
NF H_2O	10.61ul		
Total	20 ul		

Table 19

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Decapping	37	1:00:00

Table 20

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
Decapped RNA strand (SEQ ID NO: 30)	1.5 ul	166ng/ul	250ng/reaction
T4 RNA ligase 1 reaction buffer	2 ul	10x	1x

non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence	2.4 ul	50uM	8.33uM
ATP	0.4 ul	50mM	1mM
NF H_2O	0.8 ul		
T4 RNA ligase 1	2.9ul	10U/ul	29U
PEG 8k	10 ul	50%	25%
Total	20 ul		

Table 21

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Ligation	16	4:00:00

Table 22

Results

The TBE-Urea denaturing gel (See Figure 17) was used to analyse the success of the decapping of the RNA strand (SEQ ID NO: 30) and subsequent ligation with the non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG). Lane 4 of Figure 17 showed an additional band (labelled C) which was observed above the level of the RNA strand (SEQ ID NO: 30, control shown in lane 2 and labelled A) indicating that successful ligation of the non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG) occurred with ~40% efficiency.

Example 7

This example shows how a 2D (RNA-sense and DNA-antisense) library was prepared from cellular mRNA extracted from *Saccharomyces cerevisiae* to allow DNA helicase controlled movement of native *Saccharomyces cerevisiae* mRNA strands through a nanopore.

Materials and methods

A 2D (RNA-sense and DNA-antisense) library was prepared from cellular mRNA extracted from *Saccharomyces cerevisiae* by ligating a hairpin to the 3' of the mRNA, decapping the 5' end of the mRNA, reverse transcribing to create a DNA complement, and ligating a non-RNA polynucleotide to the 5' end of the mRNA.

Ligate 3' hairpin

PolyA+ mRNA from *Saccharomyces cerevisiae* was purchased from Clontech. A hairpin that hybridises to the polyA tail of the mRNA was ligated to the mRNA by mixing the reagents shown in Table 23 below and placing the mixture in a thermocycler with the program shown in Table 24 below. The mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample and eluted in 16 ul of NF H_2O.

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
polyA+ mRNA	0.5 ul	1 ug/ul	500 ng
polyT hairpin (SEQ ID NO: 29 is attached at its 5' end to a phosphate group and is attached at its 3' end to four iSpC3 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 28)	0.4 ul	25uM	1 uM
T4 DNA ligase buffer	4 ul	5 x	1 x
T4 DNA ligase	1 ul	2000U/ul	2000 U
NF H_2O	14.1 ul		
Total	20 ul		

Table 23

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Ligate	16	0:30:00

Table 24

Decap

Prior to ligating the non-RNA polynucleotide to the 5' end of the RNA strand (SEQ ID NO: 30), which has a 7-methylguanosine cap connected to its 5' end via a 5' to 5' triphosphate linkage, the cap was enzymatically removed by RNA-5'-pyrophosphohydrolase (RppH).

Decapping was achieved by mixing the reagents in Table 25 below and placing the mixture in a thermocycler set to the program in table 26 below. The resulting reaction mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample and eluted in 12 ul of NF H_2O.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
Yeast polyA+ mRNA (Clontech)	16 ul	28ng ug/ul	500 ng/reaction
NEB2 buffer	2 ul	10x	1x
RNA 5' Pyrophosphohydrolase	2 ul	5U/ul	10U
Total	20 ul		

Table 25

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Decapping	37	1:00:00

Table 26

RT

After decapping, reverse transcription was performed using Life Technologies Super Script II: reagents in Table 27 below were mixed according to the manufacturer's protocol and placed on a thermocycler set to the program in Table 28.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
3B (DNA/RNA 3B) after reverse	12 ul	~20ng/ul	240ng/reaction
dNTPs	1 ul	10uM each	0.5 uM
First-Strand Buffer	4 ul	5x	1x
0.1M DTT	2 ul	0.1M	0.01M
Super Script II	1ul	200U/ul	200U
Total	20 ul		

Table 27

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Reverse Transcription	42	0:50:00
2	Denaturation	70	0:15:00

Table 28

The mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample and eluted into 10 ul of NF H_2O.

Non-RNA polynucleotide ligation

The non-RNA polynucleotide (30 SpC3 spacers attached to the 5' end of SEQ ID NO: 31 which was attached at the 3' end to four iSp18 spacers which were attached at the opposite end to the 5' end of SEQ ID NO: 32 which was attached at the 3' end to four 5-nitroindoles which were attached at the opposite end to the RNA sequence CAAGGG) was ligated to the reverse-transcribed mRNA by mixing the reagents listed in a Table 29 and placing the mixture in a thermocycler set to the program in Table 30. The sample produced after non-RNA polynucleotide ligation was known as Sample 7A.

A 5' leader was ligated by

Reagent	***Volume***	***Stock Concentration***	***Final Concentration***
Reverse-transcribed mRNA	10.0 ul		
Non-RNA polynucleotide (described above)	1.2 ul	50 uM	2 ul
T4 RNA ligase buffer	3.0 ul	10x	1x
ATP	0.6 ul	50 mM	1 mM
8k PEG	12.0 ul	50%	
T4 RNA Ligase 1	3.2 ul	10 U/ul	
Total	30.0 ul		

Table 29

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Ligate	25	0:30:00

Table 30

hairpin purification

Sample EX was purified from ligase and unreacted components via a sequence specific tether and Lifetech MyOne C1 streptavidin beads.

Sample 7A was hybridized to a tether (/5desthiobiotin/TT/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/(SEQ ID NO: 65)/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/TT/3 CholTEG/) by mixing 1.25 ul of 100 uM

tether with 30 ul of Sample 7A and 8.75 ul of NF **H_2O and incubating the mixture for 15 min at room temperature.**

Using a magnetic rack, 20 ul of MyOne C1 streptavidin beads were washed with 200 ul of 1x Binding and washing buffer (B&W buffer) as specified by the manufacturer then resuspended in 40 ul of 2x B&W buffer.

2x B&W buffer specified by Lifetech:
10 mM TRIS-HCl pH 7.5
1 mM EDTA
2 M NaCl

40 ul of tethered Sample 7A was added to 40 ul of streptavidin beads and incubated on a roller for 15 minutes. The solution was then placed on a magnetic rack and washed twice with 1x B&W buffer per manufacturer's instructions.

The sample was eluted from the streptavidin beads by adding 15 uL of 133 uM biotin in H_2O to the beads and heating to 37C for 10 min. The tube was quickly placed on a magnetic rack and the supernatant was removed from the beads. This product is the purified sample and was known as Sample 7B.

Electrophysiology

Sample 7B (9 µl) was hybridised to a tethering oligo by mixing 3 ul of (1 uM tethering oligo (SEQ ID NO: 66/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18 /TT/3CholTEG/), 750 mM KCl, 125 mM KPhos Buffer pH 7, and 5 mM EDTA) and incubating the mixture at room temperature for 20 minutes.

This tethered Sample 7B was then incubated with 2 µl of 17.4 µM T4 Dda (E94C/F98W/C109A/C136A/K194L/A360C (SEQ ID NO: 8 with mutations E94C/F98W/C109A/C136A/K194L/A360C and then (ΔM1)G1)) for 15 minutes. 2.1 µl of 800 µM TMAD was then added to the incubated mixture and kept at room temperature for 10 min. This sample was then diluted into buffer (282 µL of 500 mM KCl, 25 mM potassium phosphate pH 8.0) and mixed with 2 ul of (70 uM ATP and 75 uM $MgCl_2$).

Electrical measurements were acquired from single MspA nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM

Potassium Ferricyanide ~ pH 8.0). After achieving a single pore inserted in the block co-polymer, then buffer (2 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide, pH 8.0) was flowed through the system to remove any excess MspA nanopores.

An excess of buffer (500 mM KCl, 25 mM potassium phosphate pH 8.0) was flowed through the system prior to the addition of sample. Finally, T4 Dda – E94C/A360C bound to sample 3A was then added to the nanopore system, the experiment was run at - 120 mV and helicase-controlled DNA movement monitored.

Results

Figure 19 shows an ionic current recording from a single nanopore when sample 7B was added. The electrophysiology experiment showed good throughput, with 2D strands of varying lengths (corresponding to different native cellular mRNA transcript lengths). Figure 19 also demonstrates that RNA has a different mean amplitude and range *versus* DNA. Thus, RNA and DNA can be differentiated from each other as a function of mean amplitude and range even when the RNA and DNA sequences are the same.

Example 8

This example shows how a DNA-containing leader was attached to messenger RNA (mRNA) to facilitate loading of a DNA helicase, Hel308Mbu-E284C/S615C (SEQ ID NO: 8 with mutations E284C/S615C), and subsequent helicase controlled movement of the RNA through a nanopore. The 1.9 kb mRNA was purchased from Trilink Biotech. The DNA-containing leader was ligated to the 3' end of the mRNA. Hel308 was then loaded onto a DNA binding site in the leader and the substrate was analysed by the nanopore.

Materials and Methods

A DNA-containing leader pre-annealed with several oligos (SEQ ID NO: 35 (which has a 5' phosphate) was attached at its 3' end to four iSp18 spacers which are attached at the opposite end to the 5' end of SEQ ID NO: 36, which is attached at its 5' end to thirty iSpC3 spacers was annealed with SEQ ID NO: 37, 38 and 39 (which has a 3' cholesterol TEG)) and this pre-annealed leader was ligated to the mRNA by mixing the reagents listed in a Table 31 and placing the mixture on a thermocycler set to the program in Table 32.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
FLuc mRNA (RNA X1, mRNA with an open reading frame of SEQ ID NO: 26 which has a 5'-hexynl-G as the 5' most nucleotide and has a 3' polyA tail)	2.5 ul	1 ug/ul	2500 ng/reaction
Leader pre-annealed 1:1:1:1	3.5 ul	2 uM ea	1x
NEB Quick Ligase Buffer (Lifetech)	4.0 ul	5 x	1x
NEB Quick Ligase	2.0 ul	2,000 U/ul	4,000 U
Splint (SEQ ID NO: 40)	0.2 ul	100 uM	1 uM
Water	8 ul		
Total	20.2 ul		

Table 31

Number of Cycles	**Step**	**Temp (°C)**	**Time**
1	Ligate	16	0:30:00

Table 32

The mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample.

Reverse Transcription

The sample was reverse transcribed using SuperScript II kit by mixing the reagents listed in a Table 33 and placing the mixture on a thermocycler set to the program in Table 34.

Reagent	*Volume*	*Stock Concentration*	*Final Concentration*
Sample from ligation	2.2 ul		500 ng
dNTPs	1 ul	10 mM Ea	
First Strand Buffer (Lifetech)	4 ul	5 x	1 x
DTT	2 ul	0.1 M	
SSII Reverse Transcriptase	1 ul		
NF H_2O	9.8 ul		
Total	20 ul		

Table 33

Number of Cycles	Step	Temp (°C)	Time
1	Reverse Transcription	42	0:50:00
2	Denaturation	70	0:15:00

Table 34

The mixture was then purified using Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample. Hel308Mbu-E284C/S615C (SEQ ID NO: 8 with mutations E284C/S615C) was buffer exchanged into 50 mM HEPES pH8, 100 mM KAc using a 7 kda Zeba column. The reverse transcribed RNA sample was mixed with an equal volume of 100 mM HEPES pH8, 200 mM KAc then mixed with Hel308Mbu-E284C/S615C at a mole ratio of 1:100 and buffer exchanged into 100 mM KPhos pH 8, 100 mM NaCl, 5 mM EDTA, 0.1% TWEEN. 20 mM BMOE dissolved in DMF was added to 100 mM KPhos pH 8, 100 mM NaCl, 5 mM EDTA, 0.1% TWEEN to a concentration of 5 mM. This 5 mM BMOE solution was added to the mixture of RNA and Hel308 to a final concentration of 40 uM BMOE. This solution was incubated at room temperature for 2 hours. The solution was then bound to Agencourt Ampure SPRI beads at a ratio of 1.8 µL SPRI beads per µL of sample. Instead of washing the SPRI beads with an EtOH mixture, a single wash with a modified wash buffer of 20% PEG, 2.5M Nacl, 50mM Tris was used. The sample was eluted in 30 ul of 500 mM KCl 25 mM KPhosphate pH 8. This was known as Sample 6.

Electrophysiology

Sample 6 (4 ul) was mixed with buffer (500 mM KCl 25 mM KPhosphat pH 8, 2 mM ATP, 2 mM $MgCl_2$, 295 ul)

Electrical measurements were acquired from single MspA nanopores inserted in block co-polymer in buffer (25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide ~ pH 8.0). After achieving a single pore inserted in the block co-polymer, then buffer (2 mL, 25mM K Phosphate buffer, 150mM Potassium Ferrocyanide, 150mM Potassium Ferricyanide, pH 8.0) was flowed through the system to remove any excess MspA nanopores.

An excess of buffer (500 mM KCl, 25 mM potassium phosphate pH 8.0) was flowed through the system prior to the addition of Sample 6. Finally, Sample 6 was then added to the nanopore system, the experiment was run at - 140 mV and helicase-controlled DNA movement was monitored.

Results

This example shows how a DNA-containing leader was attached to messenger RNA (mRNA) to facilitate loading of a DNA helicase, Hel308Mbu-E284C/S615C (SEQ ID NO: 8 with mutations E284C/S615C), and subsequent helicase controlled movement of the RNA was observed. An example of a helicase controlled RNA movement is shown in Figure 20.

CLAIMS:

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

a) providing (i) an RNA polynucleotide wherein the RNA polynucleotide is modified to comprise a non-RNA polynucleotide and (ii) a DNA helicase enzyme;

b) contacting the RNA polynucleotide and DNA helicase enzyme provided in a) with a transmembrane pore such that the DNA helicase controls the movement of the RNA polynucleotide through the transmembrane pore;

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

2. A method according to claim 1 wherein the non-RNA polynucleotide comprises a DNA helicase binding site or a DNA adaptor.

3. A method according to claim 2 wherein the DNA helicase binding site or the DNA adaptor comprises a leader sequence.

4. A method according to claim 3 wherein the leader sequence preferentially threads into the pore.

5. A method according to any one of claims 1 to 4 wherein the non-RNA polynucleotide is attached to the RNA polynucleotide by means of a covalent bond formed between at least one reactive group on each of the RNA polynucleotide and the non-RNA polynucleotide.

6. A method according to any one of claims 1 to 4 wherein the non-RNA polynucleotide is ligated to the RNA polynucleotide by chemical or enzymatic ligation.

7. A method according to any one of claims 1 to 4 wherein the non-RNA polynucleotide is hybridised to the RNA polynucleotide.

8. A method according to any one of claims 1 to 7, wherein the one or more characteristics are selected from (i) length of the RNA polynucleotide, (ii) identity of the RNA polynucleotide, (iii) the sequence of the RNA polynucleotide, (iv) the secondary structure of the RNA polynucleotide and (v) whether or not the RNA polynucleotide is modified.

9. A method according to any one of claims 1 to 8 wherein the one or more characteristics of the RNA polynucleotide are measured by electrical and/or optical measurement.

10. A method according to any one of claims 1 to 9, wherein step c) comprises measuring the current passing through the transmembrane pore as the RNA polynucleotide moves with respect to the transmembrane pore wherein the current is indicative of one or more characteristics of the RNA polynucleotide and thereby characterising the RNA polynucleotide.

11. A method according to any one of the preceding claims wherein the RNA polynucleotide comprises modification by methylation, by oxidation, by damage, with one or more proteins, base analogues or with one or more labels, tags or spacers.

12. A method according to any one of the preceding claims, wherein the RNA polynucleotide is coupled to the membrane using one or more anchors.

13. A method according to any one of the preceding claims wherein the DNA helicase comprises a modification to reduce the size of an opening in the polynucleotide binding domain through which in at least one conformational state the RNA polynucleotide can unbind from the helicase.

14. A method according to any one of the preceding claims wherein the movement is controlled by a series of one of more DNA helicases.

15. A method according to any one of the preceding claims, wherein the one or more helicases are a) Hel308 helicases, RecD helicases, XPD helicases or Dda helicases (b)

helicases derived from any of the helicases in (a); or (c) a combination of any of the helicases in (a) and/or (b).

16. A method according to any one of the preceding claims further comprising one or more molecular brakes that are derived from helicases and are modified such that they bind the polynucleotide but do not function as a helicase.

17. A method according to any one of the preceding claims wherein the transmembrane pore is a protein pore or a solid state pore.

18. A method according to claim 17, wherein the transmembrane protein pore is a protein pore and is derived from a hemolysin, leukocidin, *Mycobacterium smegmatis* porin A (MspA), MspB, MspC, MspD, CsgG, lysenin, outer membrane porin F (OmpF), outer membrane porin G (OmpG), outer membrane phospholipase A, *Neisseria* autotransporter lipoprotein (NalP) and WZA.

19. A method of moving a target RNA polynucleotide with respect to a transmembrane pore when the movement is controlled by a DNA helicase enzyme, comprising:

a) providing (i) an RNA polynucleotide wherein the RNA polynucleotide is modified to comprise a non-RNA polynucleotide and a (ii) a DNA helicase enzyme;

b) contacting the RNA polynucleotide and DNA helicase enzyme provided in a) with a transmembrane pore such that the DNA helicase controls the movement of the RNA polynucleotide with respect to the transmembrane pore.

20. A method according to claim 19 wherein the method comprises before step (b) binding the DNA helicase enzyme to the modified RNA polynucleotide.

21 A method according to claim 19 or claim 20, wherein the RNA polynucleotide is modified to comprise a DNA helicase binding site or a DNA adaptor.

22. A method according to any one of claims 19 to claim 21, wherein the method provides more consistent movement of the RNA polynucleotide with respect to the transmembrane pore.

23. A polynucleotide comprising an RNA polynucleotide and DNA polynucleotide wherein the DNA polynucleotide comprises or comprises only a DNA helicase binding site.

24. A polynucleotide according to claim 23 further comprising a leader sequence which preferentially threads into a nanopore.

25. A polynucleotide according to claim 23 or claim 24 further comprising a barcoding section on the polynucleotide strand.

26. A polynucleotide according to claim 25 wherein the barcoding section is located between the leader sequence and the DNA helicase binding site.

27. A combination of an RNA polynucleotide and a DNA helicase in which a part of the RNA polynucleotide has been modified to comprise a non-RNA polynucleotide and interact with the DNA helicase..

28. A kit for characterising a target RNA polynucleotide comprising a non-RNA polynucleotide which is adapted to attach to any target RNA polynucleotide for characterisation.

Figure 1 1/22

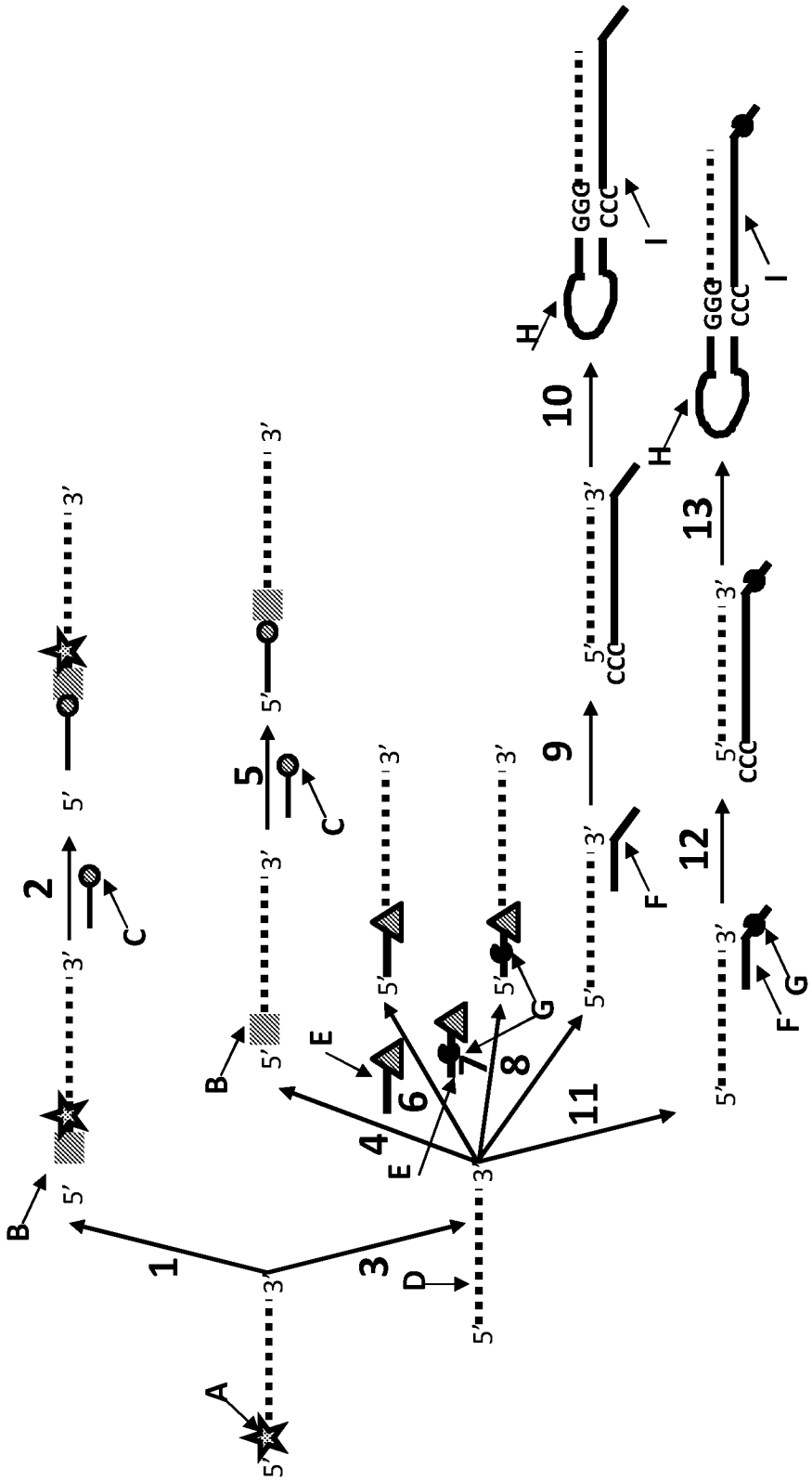


Figure 2

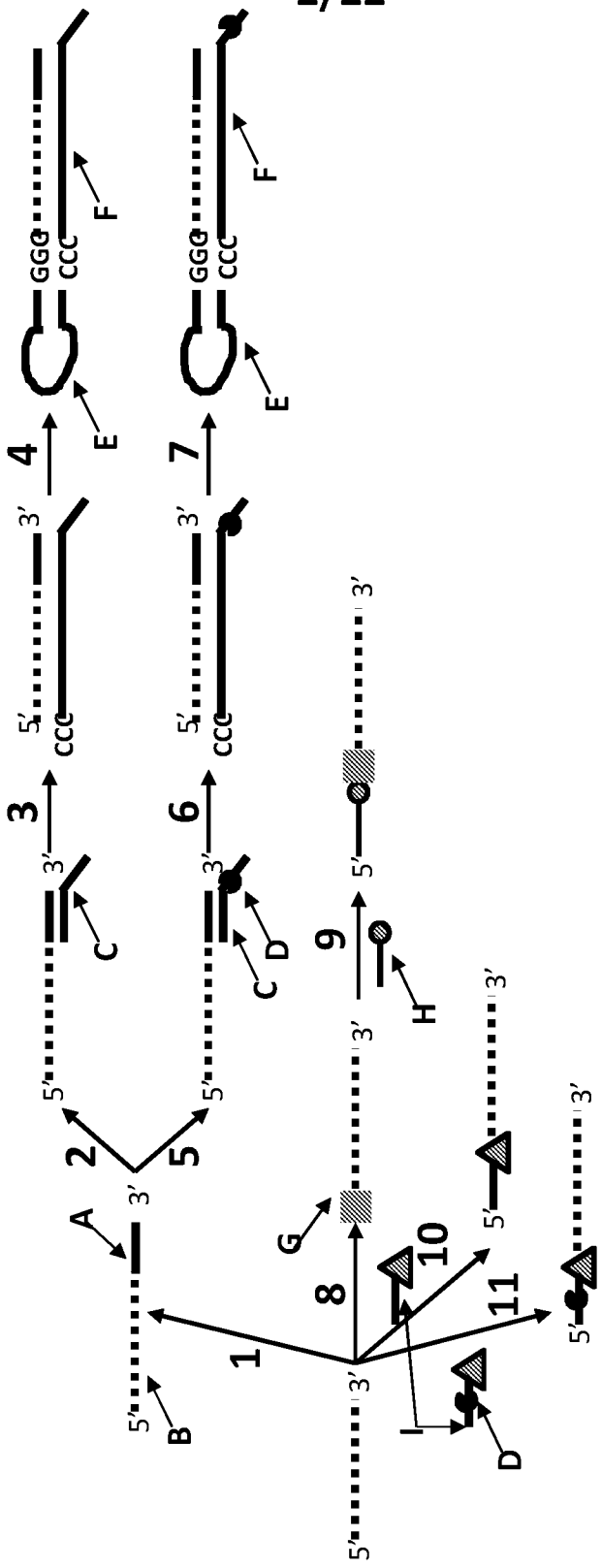


Figure 3 3/22

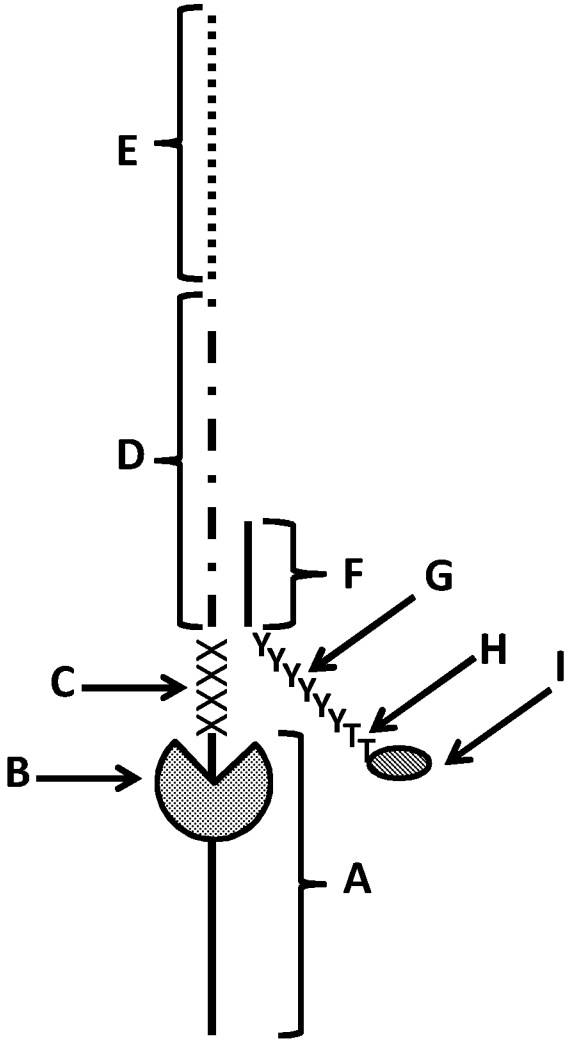


Figure 4

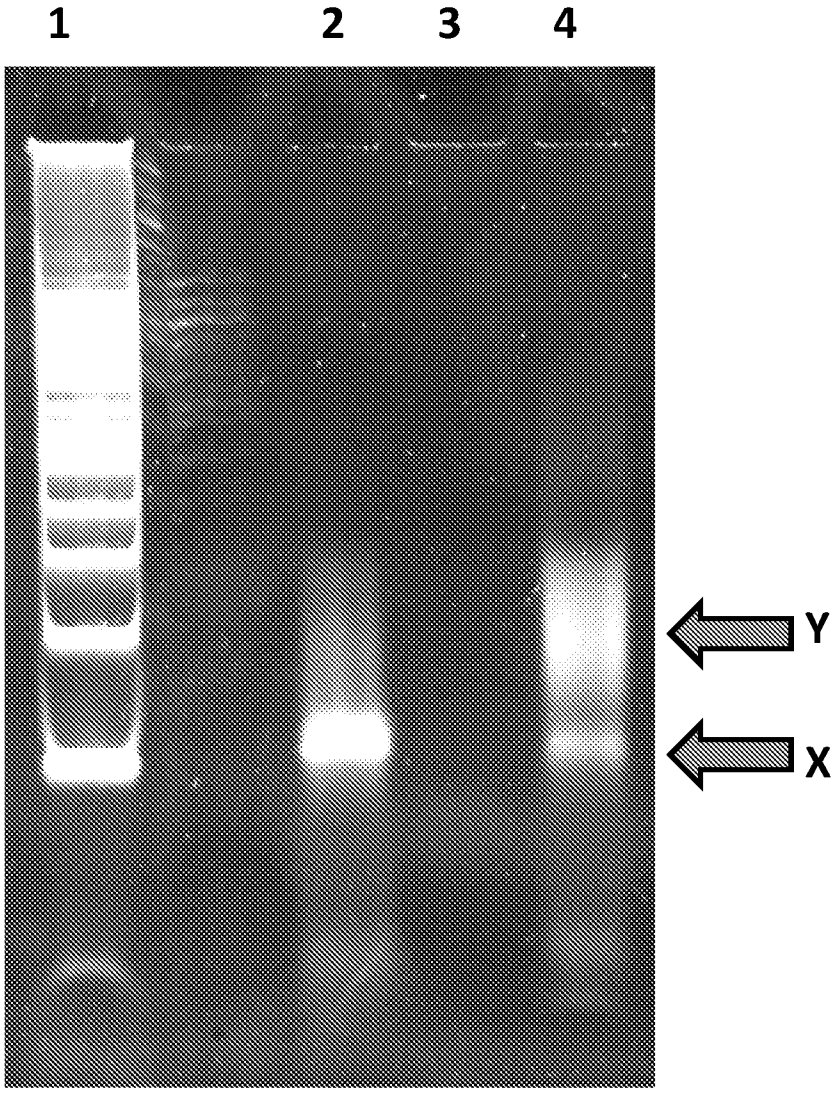


Figure 5 5/22

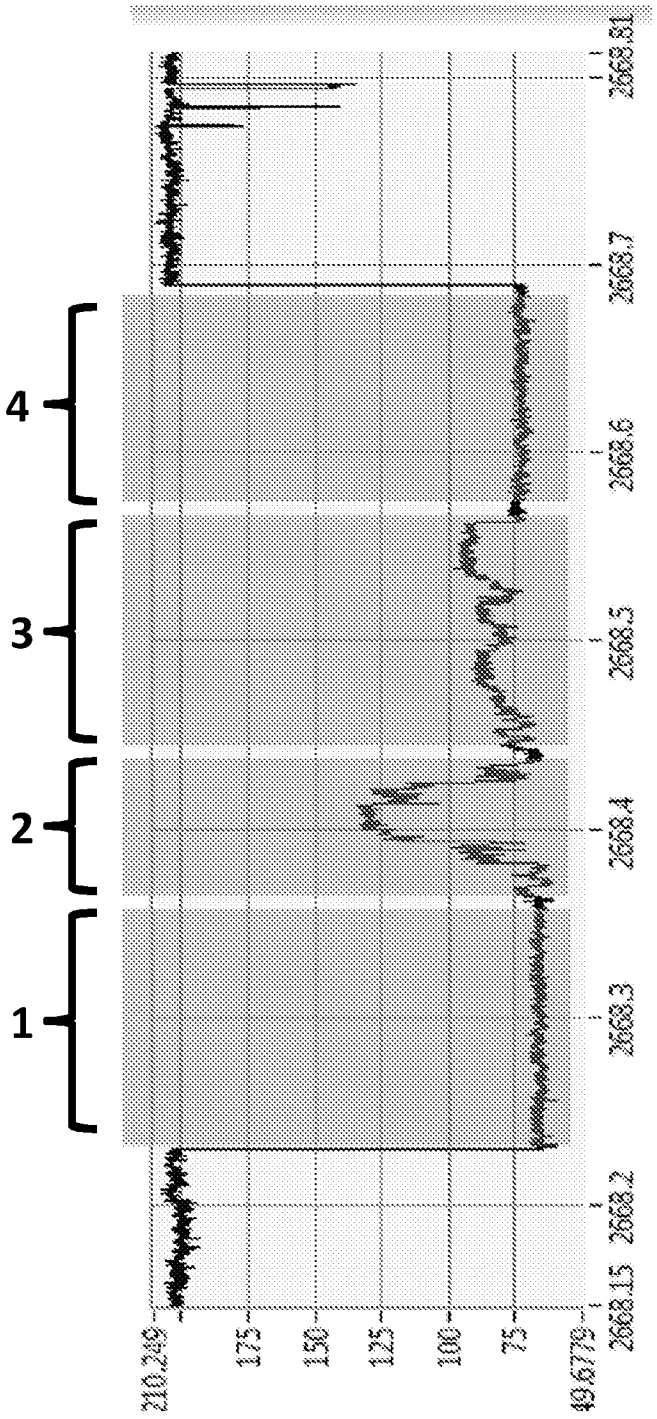

4
3
2
1
2668.81
2668.7
2668.6
2668.5
2668.4
2668.3
2668.2
2668.15
210.249
175
150
125
100
75
49.6779

Figure 6

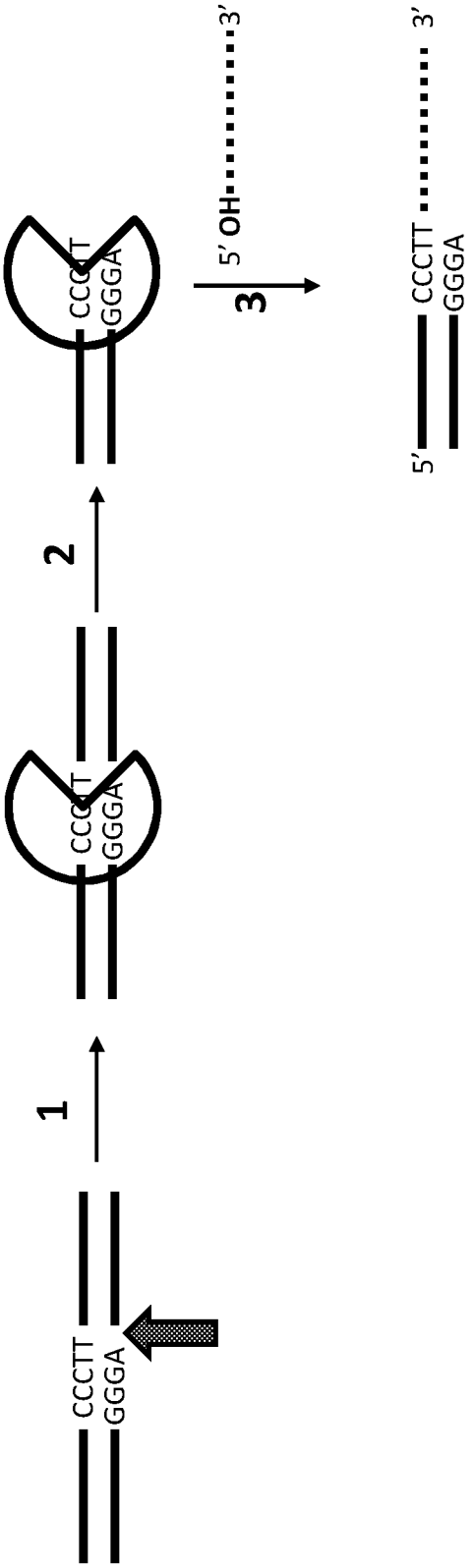


Figure 7

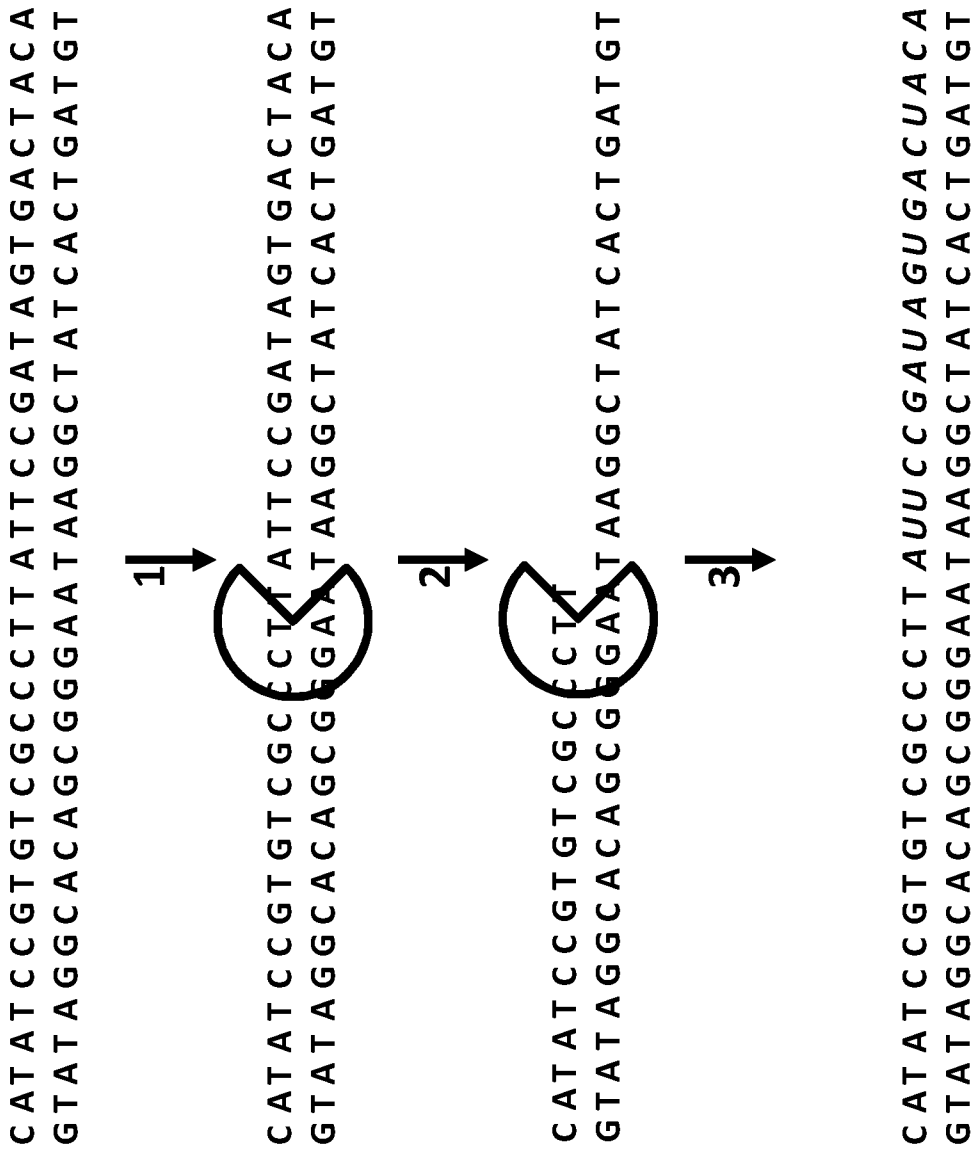

CATATCCGTGTCGCCCTTATTCCGATAGTGACTACA
GTATAGGCACAGCGGGAATAAGGCTATCACTGATGT
1
CATATCCGTGTCGCCCTTATTCCGATAGTGACTACA
GTATAGGCACAGCGGGAATAAGGCTATCACTGATGT
2
CATATCCGTGTCGCCCTTATTCCGATAGTGACTACA
GTATAGGCACAGCGGGAATAAGGCTATCACTGATGT
3
CATATCCGTGTCGCCCTTTAUUCCGAUAGUGACUACA
GTATAGGCACAGCGGGAATAAGGCTATCACTGATGT

Figure 8 8/22

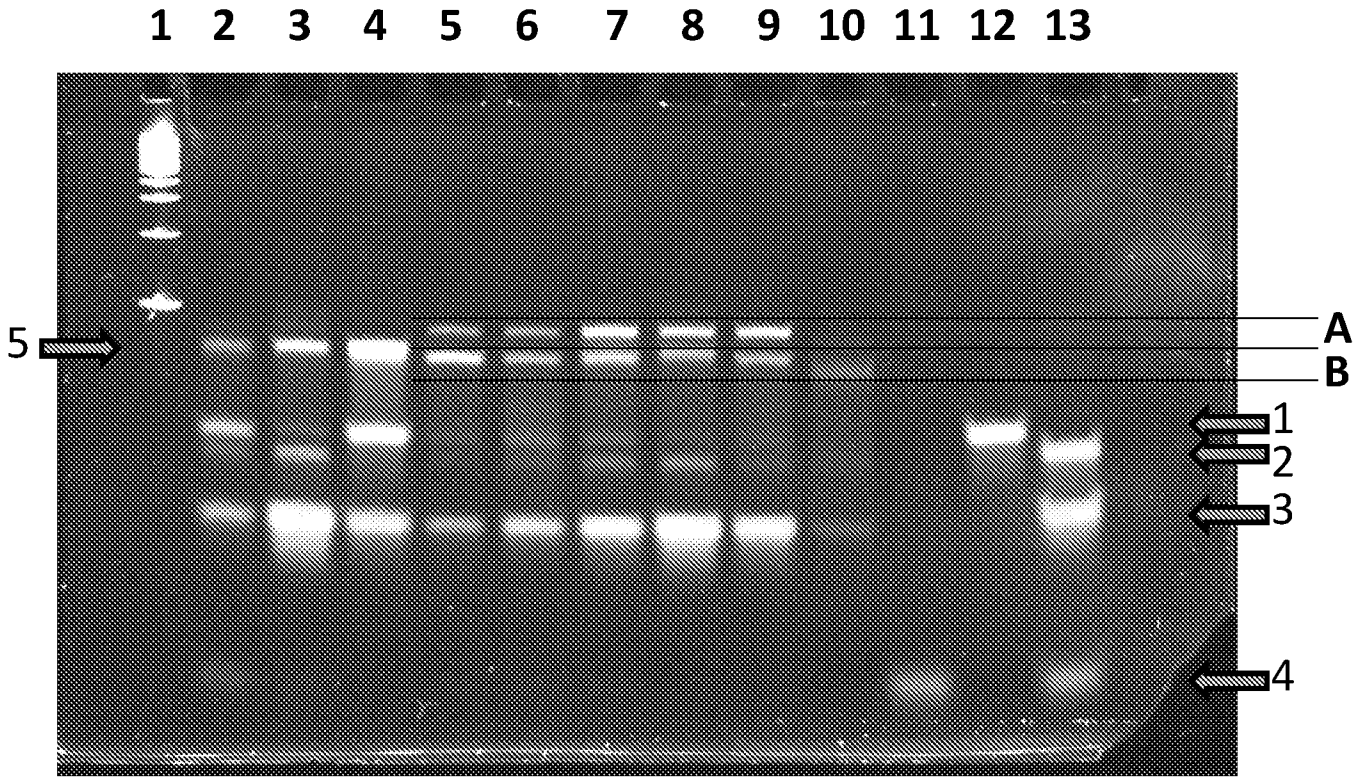


Figure 9 9/22

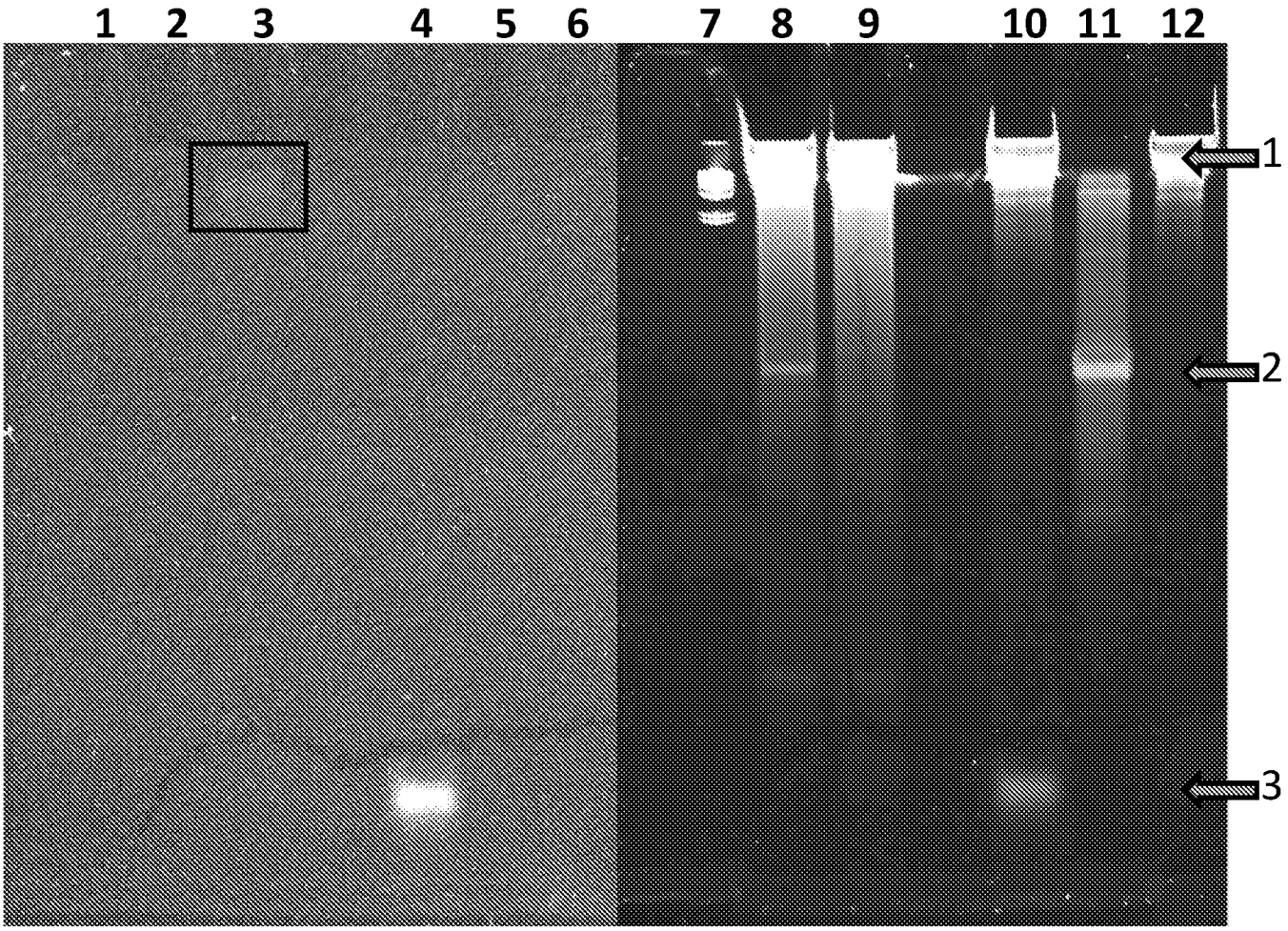


Figure 10

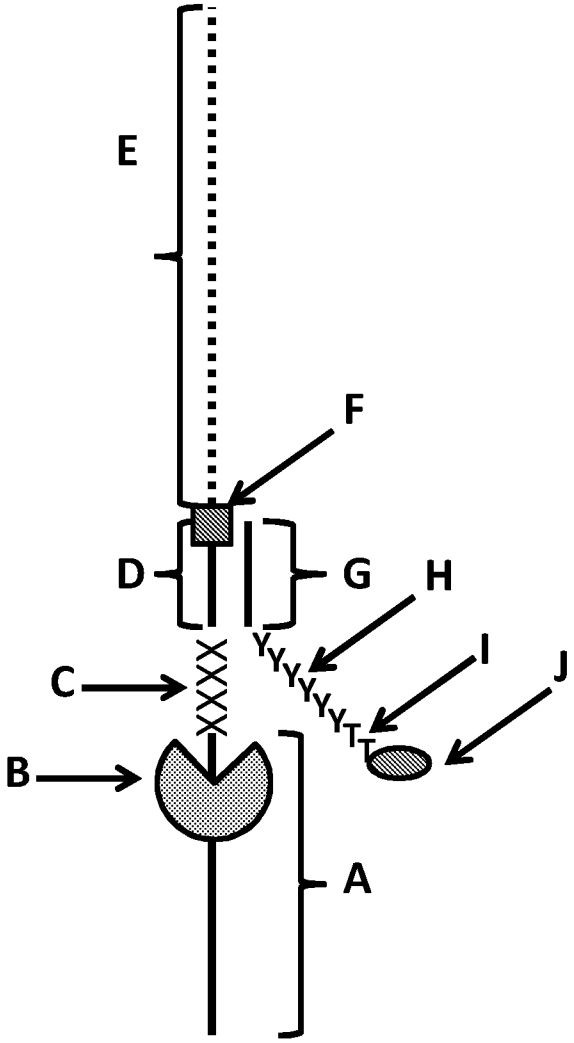


Figure 11 11/22

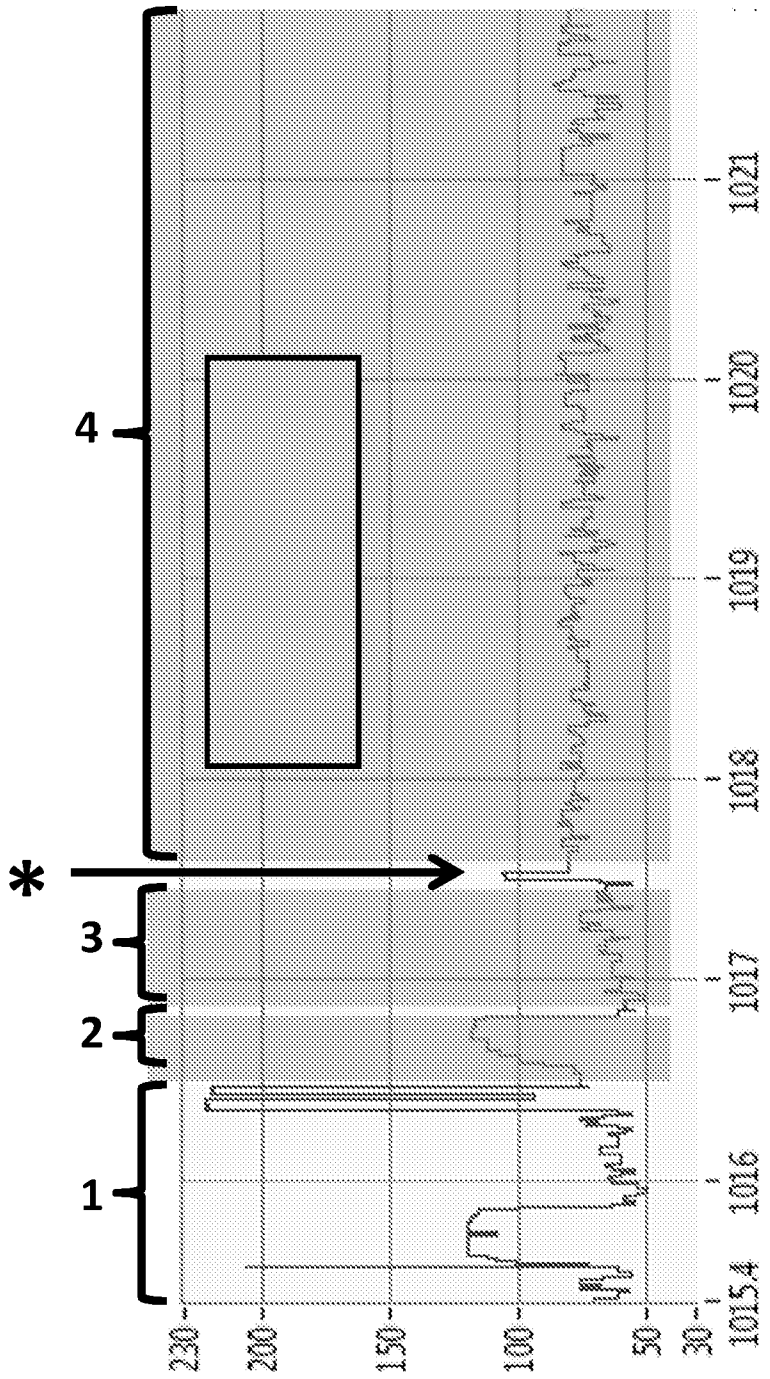


Figure 12

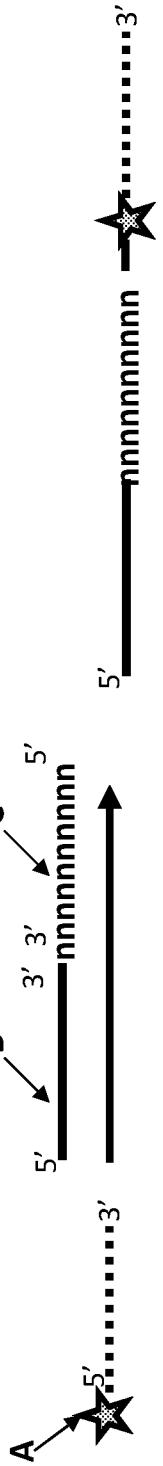


Figure 13 13/22

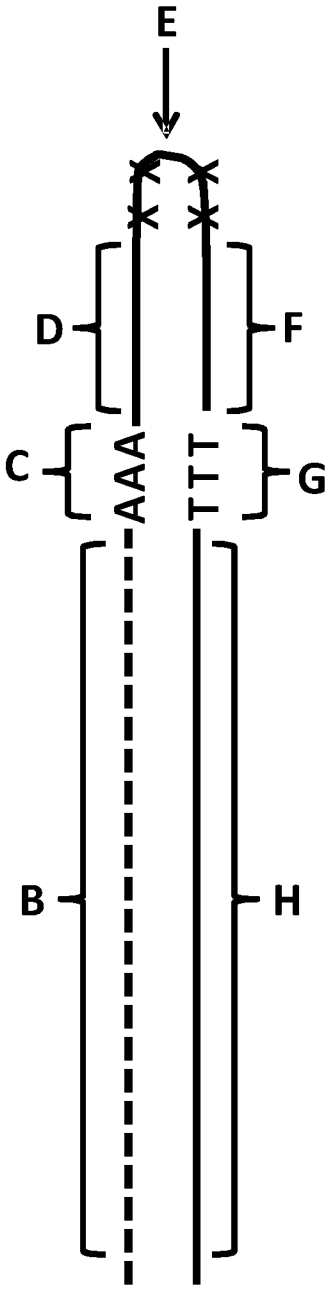


Figure 14

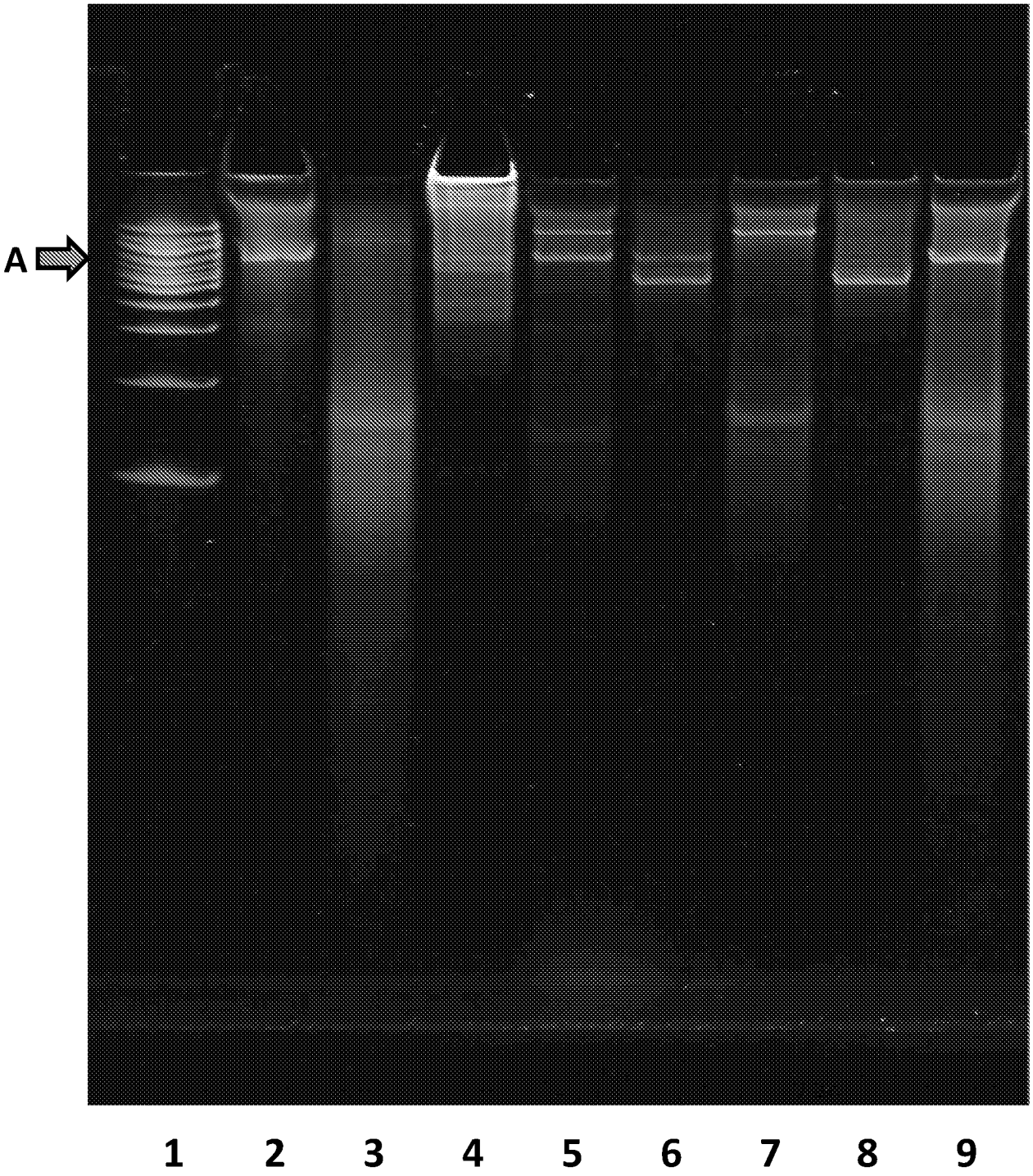


Figure 15

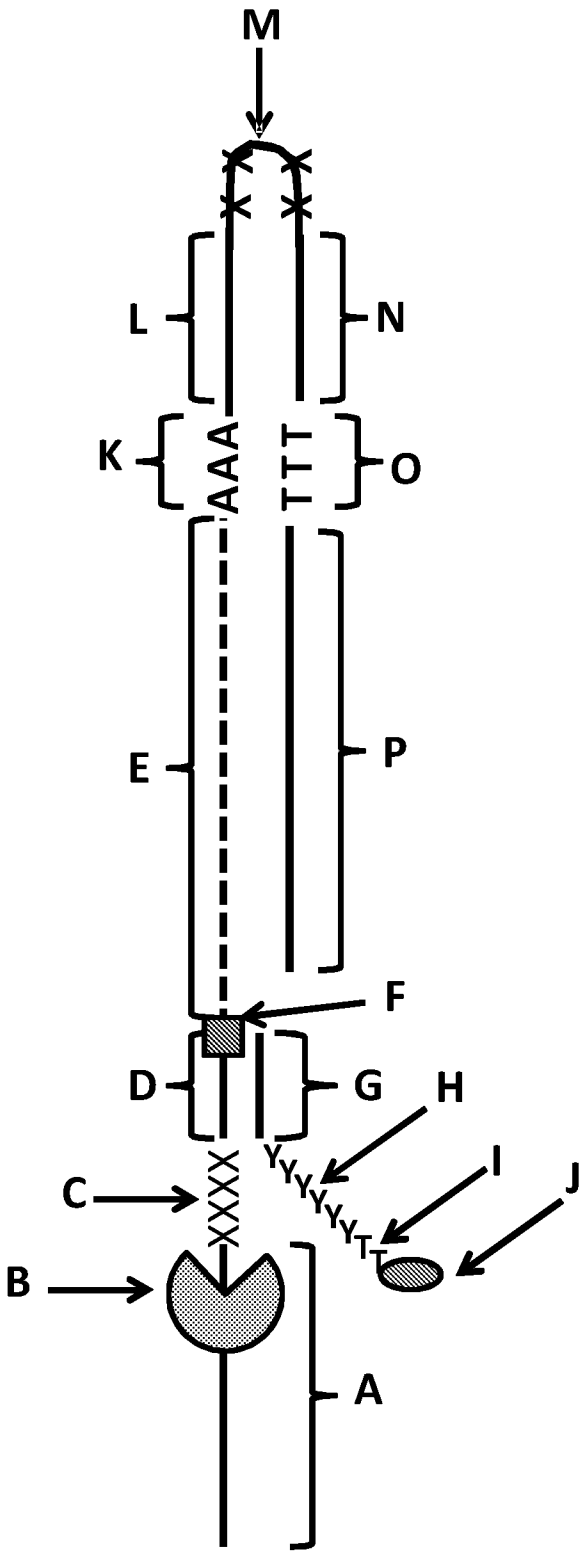


Figure 16

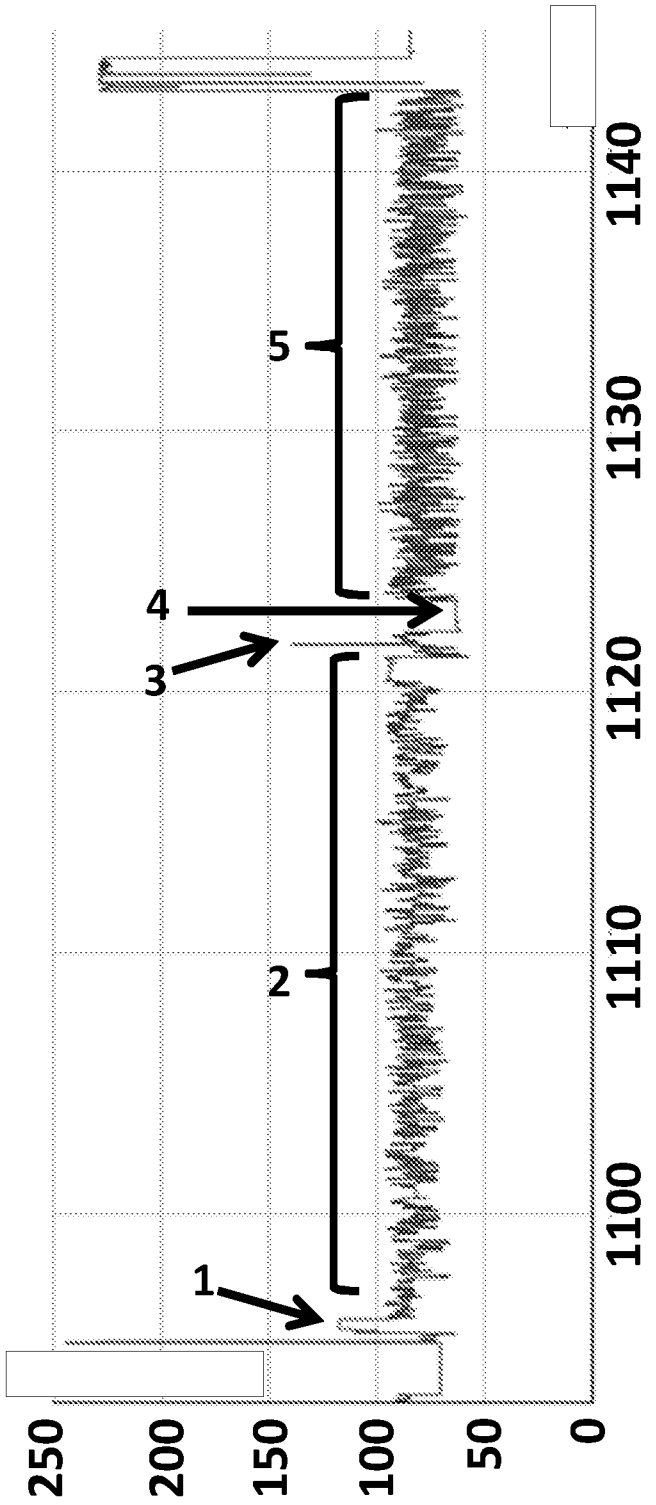


Figure 17

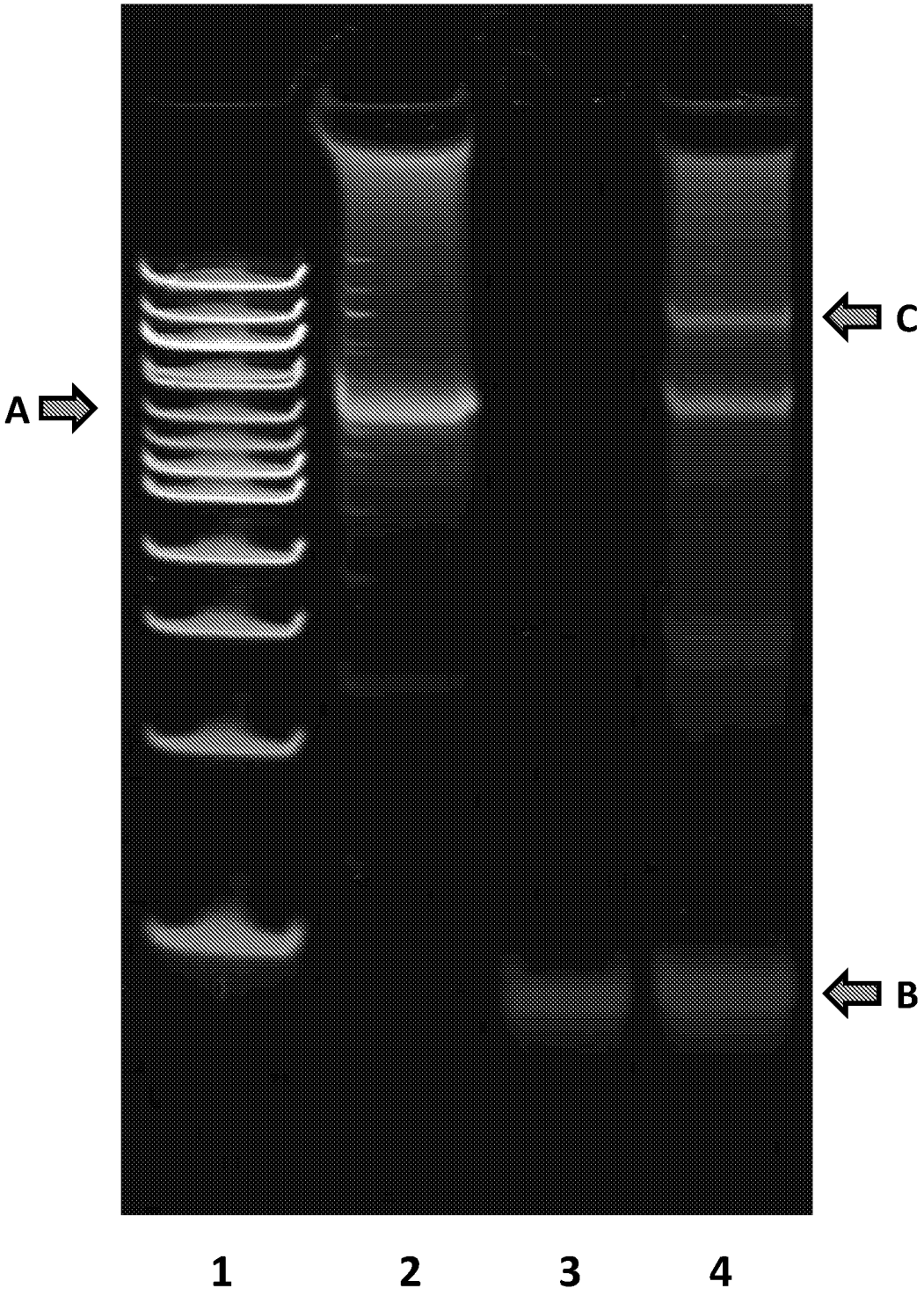


Figure 18 **18/22**

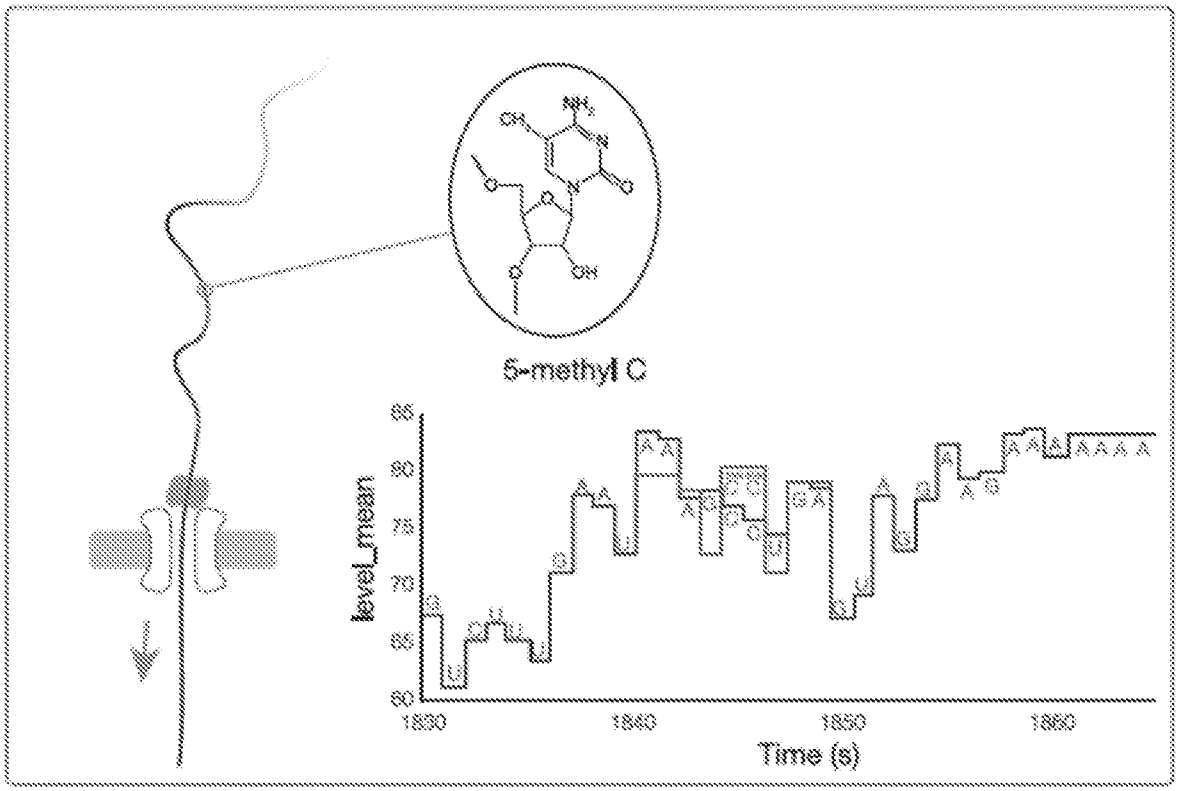


Figure 19 **19/22**

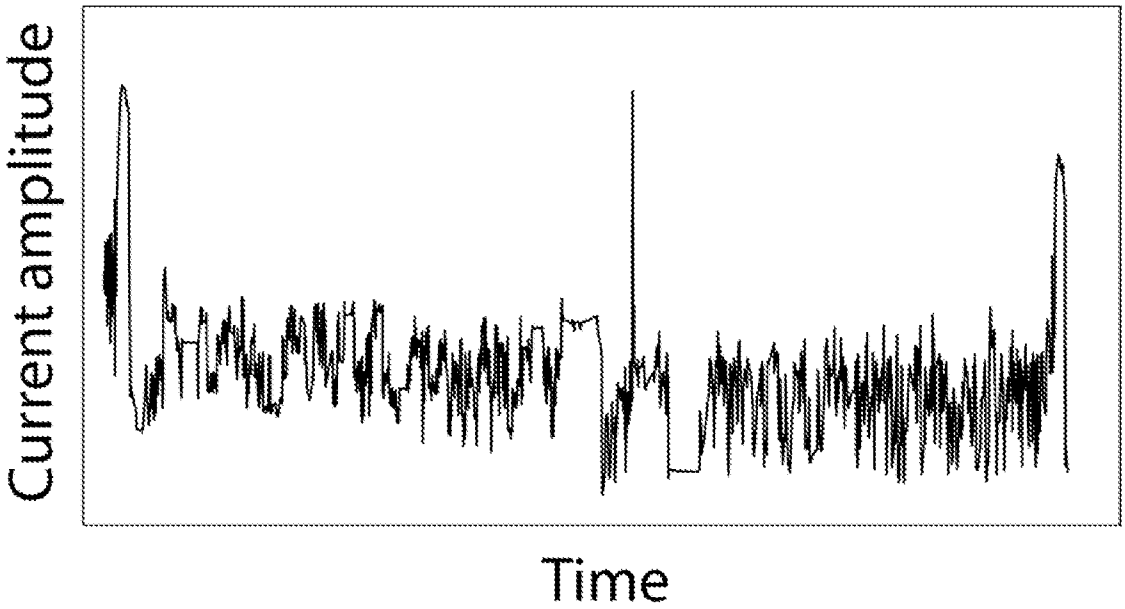


Figure 20

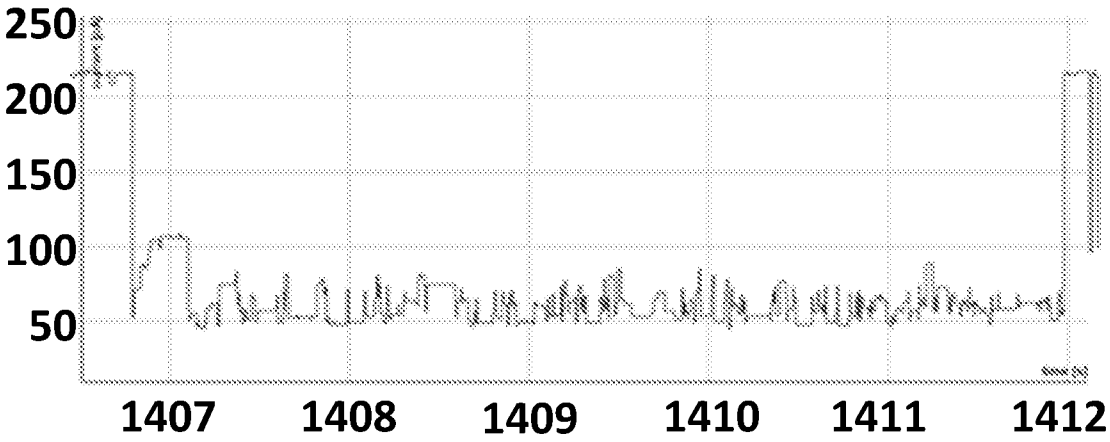


Figure 21 21/22

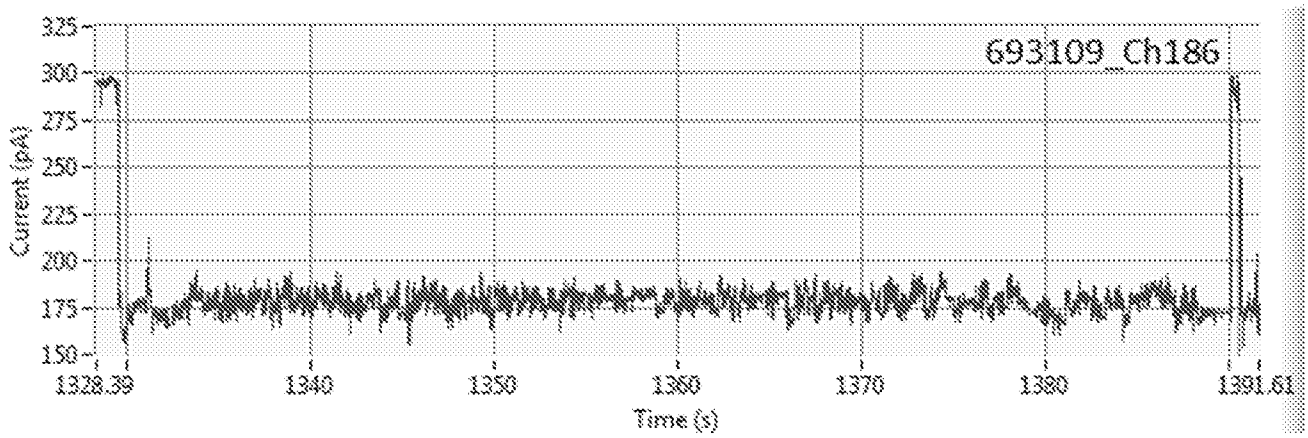
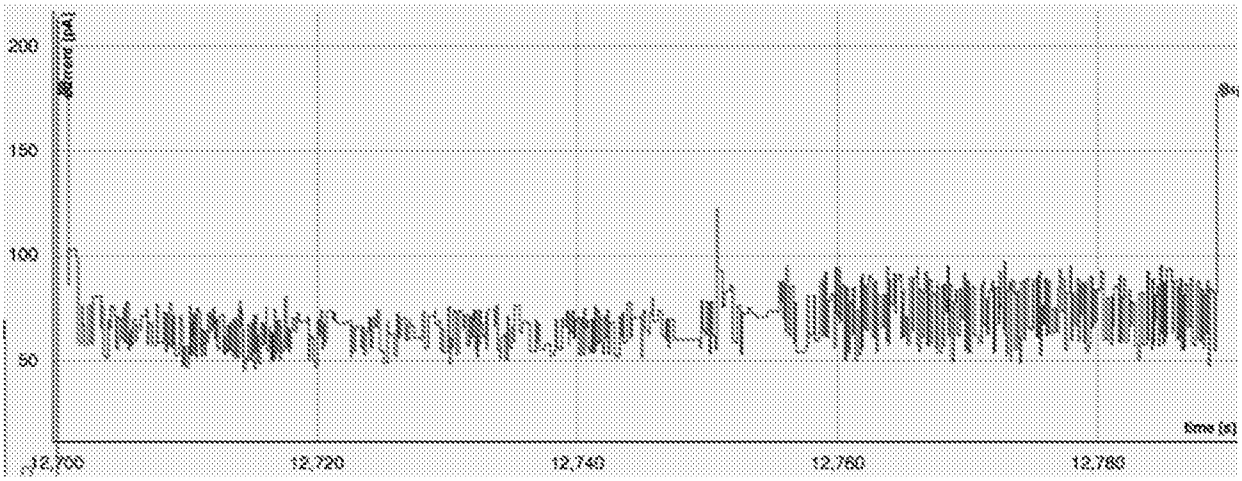


Figure 22 **22/22**



INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/053097

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

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

B. FIELDS SEARCHED

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013/014451 A1 (OXFORD NANOPORE TECH LTD [GB]; BROWN CLIVE [GB]; CLARKE JAMES [GB]; HA) 31 January 2013 (2013-01-31) page 1; claims 1-8; figure 10,	1-28
X	WO 2013/057495 A2 (OXFORD NANOPORE TECH LTD [GB]) 25 April 2013 (2013-04-25) page 1; figure 10	1-28
X	WO 2010/109197 A2 (ISIS INNOVATION [GB]; STODDART DAVID [GB]; HERON ANDREW JOHN [GB]; MAG) 30 September 2010 (2010-09-30) pages 10-11	1-28
X	WO 2013/098562 A2 (OXFORD NANOPORE TECH LTD [GB]) 4 July 2013 (2013-07-04) figure 1	1-28

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 December 2015	18/12/2015

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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2015/053097

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2015/053097

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2012/033524 A2 (UNIV CALIFORNIA [US]; CHERF GERALD MAXWELL [US]; LIBERMAN KATE R [US];) 15 March 2012 (2012-03-15) the whole document -----	1-28
T	WO 2015/056028 A1 (OXFORD NANOPORE TECH LTD [GB]) 23 April 2015 (2015-04-23) the whole document -----	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2015/053097

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013014451 A1	31-01-2013	AU 2012288629 A1	06-03-2014
		CA 2843136 A1	31-01-2013
		CN 103827320 A	28-05-2014
		EP 2737084 A1	04-06-2014
		JP 2014531196 A	27-11-2014
		KR 20140050067 A	28-04-2014
		US 2015152492 A1	04-06-2015
		WO 2013014451 A1	31-01-2013
WO 2013057495 A2	25-04-2013	AU 2012324639 A1	22-05-2014
		CA 2852812 A1	25-04-2013
		CN 104039979 A	10-09-2014
		EP 2768977 A2	27-08-2014
		JP 2014534812 A	25-12-2014
		KR 20140090633 A	17-07-2014
		US 2014255921 A1	11-09-2014
		WO 2013057495 A2	25-04-2013
WO 2010109197 A2	30-09-2010	CN 102365371 A	29-02-2012
		EP 2411538 A2	01-02-2012
		US 2012107802 A1	03-05-2012
		WO 2010109197 A2	30-09-2010
WO 2013098562 A2	04-07-2013	AU 2012360244 A1	24-07-2014
		CA 2861808 A1	04-07-2013
		CN 104126018 A	29-10-2014
		EP 2798084 A2	05-11-2014
		JP 2015503917 A	05-02-2015
		KR 20140108706 A	12-09-2014
		US 2014335512 A1	13-11-2014
		WO 2013098562 A2	04-07-2013
WO 2012033524 A2	15-03-2012	CN 103392008 A	13-11-2013
		EP 2614156 A2	17-07-2013
		US 2014051068 A1	20-02-2014
		WO 2012033524 A2	15-03-2012
WO 2015056028 A1	23-04-2015	NONE	