

PIF1-LIKE HELICASE AND USE THEREOF

TECHNICAL FIELD

[0001] The present application relates to the technical fields of gene sequencing, molecular detection and clinical detection, particularly relates to a modified Pif1-like helicase, a construct comprising the Pif1-like helicase, and its use in characterising a target polynucleotide or controlling the movement of a target polynucleotide through a pore.

BACKGROUND

[0002] Nanopore sequencing technology refers to a gene sequencing technology that uses a single nucleic acid molecule as a measurement unit and uses nanopores to read its sequence information in real time and continuously.

[0003] The nanopore sequencing technology uses nanopores that can provide channels for ion current. Electrophoresis drives polynucleotides through the nanopore. Since a polynucleotide passes through the nanopore, the current passing through the nanopore may be reduced. Each nucleotide or series of nucleotides passing through the nanopore produce a characteristic current, and the current level corresponds to the polynucleotide sequence. In “strand sequencing” (such as using a helicase to control the movement of a polynucleotide through a pore), a single polynucleotide chain passes through the pore and enables identification of nucleotides.

[0004] The sequencing technology has many advantages, such as the library may be easily constructed without amplification; the technology has a fast reading speed, and may reach a reading speed of tens of thousands of bases per hour for a single molecule; the technology has a long read length, usually thousands of bases; and it is possible to make a direct measurement of RNA and DNA methylation. These are beyond the reach of the existing second-generation sequencing technology.

[0005] However, nanopore sequencing technology also has difficult problems that need to be solved. For example, the translocation of polynucleotides through the nanopore is so fast that the current level of a single nucleotide is too short to be distinguished. Especially when the nucleotide sequence has a very long length, such as 500 or more nucleotides, the molecular motor which is controlling the movement of the polynucleotide may disengage from the polynucleotide. This allows the polynucleotide to be pulled through the pore rapidly and in an uncontrolled manner in the direction of the applied field.

[0006] There are existing technologies that can solve this problem. For example, patent application WO2013057495A3 discloses a new method of characterising a target polynucleotide, wherein the method comprises controlling the movement of the target polynucleotide through a pore by Hel308 helicase or molecular motors. Patent application US20150065354A1 discloses a method of characterising a target polynucleotide by using XPD helicase, wherein the method comprises controlling the movement of the target polynucleotide through pores by XPD helicase. Patent application CN107109380A discloses a modified enzyme, which is a modified Dda helicase that can control the movement of a target polynucleotide through a pore. All of the above methods may control the movement of a target polynucleotide through a pore to a certain extent. However, there is still a need to develop a new helicase that controls the movement of a target polynucleotide through a pore.

[0007] In addition, the ability of a helicase to continuously and uniformly control the movement of nucleic acids plays an important role in accurately identifying information of a nucleic acid sequence. In the movement of nucleic acids through a nanopore controlled by a helicase, a uniform speed at which the enzyme moves on different nucleic acid molecules, and a uniform speed at which the enzyme moves on different bases of the same nucleic acid molecule are very important for the accurate identification of base signals since the nanopore sequencing technology is a single-molecule sequencing technology, in addition to the need for continuously binding the enzyme to the nucleic acid. Specifically, when enzymes control the movement of different nucleic acid molecules, the speeds at which different enzymes control the nucleic acids to pass through the nanopore are usually different due to the heterogeneity of different enzyme molecules. In addition, some enzymes stall when controlling nucleic acids to pass through the nanopore due to their poor activity, which is not conducive to the analysis of base signals and the identities of base sequences, or is not conducive to the efficient use of pore channels, thereby reducing accuracy or throughput of sequencing. The translocation of enzymes on the same nucleic acid molecule may lead to some phenomena that are not conducive to sequencing, including but not limited to, “slippage” of individual or consecutive multiple bases from the enzyme due to the electric field force, or half or multiple bases per step length due to inconsistent step length of the enzyme, or shifting of the enzyme forward one or more step length and then backward relative to the nucleic acid.

[0008] Therefore, effectively improving the homogeneity among different enzyme molecules and effectively controlling the movement of nucleic acids through nanopores by a single enzyme molecule are very important for improving the performance of the nanopore strand sequencing technology.

SUMMARY

[0009] The present application provides a new modified Pif1-like helicase to solve the problem of too fast translocation of polynucleotides through a nanopore. The modified Pif1-like helicase may maintain the binding to the polynucleotide for a longer time and control the movement of the polynucleotide through the pore. The Pif1-like helicase of the present application is a useful tool for controlling the movement of polynucleotides in the strand sequencing, causing the polynucleotides to move in a controlled and stepwise manner along or against the electric field caused by the applied voltage, thereby controlling the speed of the polynucleotides passing through the nanopore, and obtaining a recognizable current level. Especially when the polynucleotide chain has an increased length, such as 500 or more nucleotides, and molecular motors are required with improved processivity, the Pif1-like helicase of the present application will still not disengage from the polynucleotide, that is, it is particularly effective in controlling the movement of polynucleotides of 500, 1000, 5000, 10000, 20000, 50000, 100000 or more nucleotides. The Pif1-like helicase of the present application has the advantages of controlling the smooth movement of polynucleotides from a pore and reducing slippage or irregular movement, thereby promoting more accurate reading of nucleotides and longer read lengths. The Pif1-like helicase of the present application can effectively improve the ability of the enzyme molecule to control the movement of nucleic acid molecules through a nanopore while keeping the enzyme being continuously bound to the nucleic acid, so as to improve the accuracy and throughput of nucleic acid detection or sequencing.

[0010] In the first aspect of the present application, provided is a Pif1-like helicase comprising at least one cysteine residue and/or at least one non-natural amino acid introduced into a tower domain, a pin domain and/or a 1A (RecA-like motor) domain of the Pif1-like helicase, wherein the helicase further comprises: (A) substitution or deletion of at least one or more native amino acids, wherein the substituted or deleted native amino acids comprise amino acids corresponding to C308 and/or C419 in SEQ ID NO: 6; and/or (B) substitution of at least one amino acid that interacts with one or more nucleotides in single-stranded DNA or double-stranded DNA, wherein the substituted amino acids comprise amino acids corresponding to H87 and/or V422 and/or I282 in SEQ ID NO: 6, wherein the Pif1-like helicase retains its ability to control the movement of a polynucleotide.

[0011] Specifically, provided is a Pif1-like helicase of SEQ ID NOs: 1-11 comprising at least one cysteine residue and/or at least one non-natural amino acid introduced into a tower domain, a pin domain and/or a 1A (RecA-like motor) domain of the Pif1-like helicase, wherein the

helicase further comprises: (A) substitution or deletion of at least one or more native amino acids, wherein the substituted or deleted native amino acids comprise C308 and/or C419 in SEQ ID NO: 6 or amino acids in SEQ ID NOs: 1 to 5 and 7 to 11 which correspond to C308 and/or C419 in SEQ ID NO: 6; and/or (B) substitution of at least one amino acid that interacts with one or more nucleotides in single-stranded DNA or double-stranded DNA, wherein the substituted amino acids comprise H87 and/or V422 and/or I282 in SEQ ID NO: 6 or amino acids in SEQ ID NOs: 1 to 5 and 7 to 11 which correspond to H87, V422 and/or I282 in SEQ ID NO: 6, wherein the Pif1-like helicase retains its ability to control the movement of a polynucleotide.

[0012] Preferably, in the Pif1-like helicase the at least one cysteine residue and/or at least one non-natural amino acid have been introduced into any one selected from the group consisting of the following:

[0013] (A) the tower domain;

[0014] (B) the pin domain;

[0015] (C) the 1A domain;

[0016] (D) the tower domain and the pin domain;

[0017] (E) the tower domain and the 1A domain;

[0018] (F) the 1A domain and the pin domain;

[0019] (G) the tower domain, the pin domain and the 1A domain.

[0020] Preferably, into the tower domain, the pin domain and/or the 1A domain, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or more cysteine residues may be introduced or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or more non-natural amino acids may be introduced, or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or more cysteine residues and non-natural amino acids may be introduced.

[0021] Preferably, the introduction of cysteine residues and/or at least one non-natural amino acids make the binding of the Pif1-like helicase to the polynucleotide more stable and enhances its ability to control the movement.

[0022] Preferably, the Pif1-like helicase is selected from Pba-PM2, Aph-Acj61, Aph-PX29, Avi-Aeh1, Sph-CBH8, Eph-Pei26, Aph-AM101, PphPspYZU05, Eph-EcS1, Eph- Cronus2 or Mph-MP1. Some wild-type Pif1-like helicases are shown in Table 1.

[0023] Table 1. Wild-type Pif1-like helicases

Wild-type Pif1-like helicase		UniProt	Length
Pba-PM2 (SEQ ID NO:1)	Pectobacterium bacteriophage	A0A0A0Q3A4	444

	PM2		
Aph-Acj61 (SEQ ID NO:2)	Acinetobacter phage Acj61	E5E403	441
Aph-PX29 (SEQ ID NO:3)	Aeromonas phage PX29	E5DQH4	452
Avi-Aeh1 (SEQ ID NO:4)	Aeromonas virus Aeh1	Q76YI7	453
Sph-CBH8 (SEQ ID NO:5)	Serratia phage CBH8	A0A1Z1LY03	441
Eph-Pei26 (SEQ ID NO:6)	Edwardsiella phage PEi26	A0A0B6VP37	446
Aph-AM101 (SEQ ID NO:7)	Acinetobacter phage AM101	A0A4Y1NMU3	452
PphPspYZU0 (SEQ ID NO:8)	Pseudomonas phage PspYZU05	A0A2U7NRX4	454
Eph-EcS1 (SEQ ID NO:9)	Escherichia phage EcS1	A0A2Z5ZC85	441
Eph-Cronus2 (SEQ ID NO:10)	Erwinia phage Cronus	A0A2S1GM41	443
Mph-MP1 (SEQ ID NO:11)	Morganella phage vB_MmoM_MP1	A0A192Y9K3	441

[0024] The amino acid residue positions of the tower domain, the pin domain, the hook domain, the 1A domain and the 2A domain of the Pif1-like helicase are shown in Table 2.

[0025] Table 2 Residues making up each domain in each Pif1-like helicase as identified.

Pif1-like helicase	SEQ ID NO	1A domain	2 A domain
Pba-PM2	1	M1-L88, M106-V181	E264-P278, N296-A394
Aph-Acj61	2	M1-L88, I106-M180	E265-P279, N297-A392
Aph-PX29	3	M1-L88, K110-V182	T266-P280, N298-S403
Avi-Aeh1	4	M1-L88, K110-V182	T266-P280, N298-S404
Sph-CBH8	5	M1-L84, M103-K177	E260-P274, N292-A391
Eph-Pei26	6	M1-L90, M108-M183	E266-P280, N298-A396
Aph-AM101	7	M1-L99, D117-M191	T276-P290, N308-P402
PphPspYZU05	8	M1-L95, I113-K187	D274-P288, N306-A404
Eph-EcS1	9	M1-L85, M103-K177	E260-P274, N292-A391
Eph-Cronus2	10	M1-L87, I105-K180	E265-P279, H297-A393
Mph-MP1	11	M1-L96, P114-K184	E264-P278, N296-P389

[0026]

Pif1-like helicase	SEQ ID NO	Pin domain	Tower domain	Hook domain
--------------------	-----------	------------	--------------	-------------

Pba-PM2	1	K89-E105	E264-P278, N296-A394	E277-F295
Aph-Acj61	2	K89-D105	E265-P279, N297-A392	E278-F296
Aph-PX29	3	K89-A109	T266-P280, N298-S403	E279-L297
Avi-Aeh1	4	K89-A109	T266-P280, N298-S404	E279-L297
Sph-CBH8	5	K86-E102	E260-P274, N292-A391	E273-F291
Eph-Pei26	6	K91-E107	E266-P280, N298-A396	E279-F297
Aph-AM101	7	K100-D116	T276-P290, N308-P402	E289-F307
PphPspYZU05	8	K95-E112	D274-P288, N306-A404	G287-Y305
Eph-EcS1	9	K86-E102	E260-P274, N292-A391	E273-F291
Eph-Cronus2	10	K88-E104	E265-P279, H297-A393	E278-F296
Mph-MP1	11	K97-A113	E264-P278, N296-P389	E277-F295

[0027] In a specific embodiment of the present application, the Pif1-like helicase is selected from Mph-MP1, Sph-CBH8, Eph-Pei26 or PphPspYZU05.

[0028] Preferably, the helicase comprises a variant of SEQ ID NO: 11 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E264-P278 and N296-P389), and/or the pin domain (residues K97-A113), and/or the 1A domain (residues M1-L96 and P114-K184).

[0029] Preferably, the helicase comprises a variant of SEQ ID NO: 1 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E264-P278 and N296-A394), and/or the pin domain (residues K89-E105), and/or the 1A domain (residues M1-L88 and M106-V181).

[0030] Preferably, the helicase comprises a variant of SEQ ID NO: 2 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E265-P279 and N297-A392), and/or the pin domain (residue K89-D105), and/or the 1A domain (residue M1-L88 and I106-M180).

[0031] Preferably, the helicase comprises a variant of SEQ ID NO: 3 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues T266-P280 and N298-S403), and/or the pin domain (residues K89-A109), and/or the 1A domain (residues M1-L88 and K110-V182).

[0032] Preferably, the helicase comprises a variant of SEQ ID NO: 4 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues T266-P280 and N298-S404), and/or the pin domain (residues K89-A109), and/or the 1A domain (residues M1-L88 and K110-V182).

[0033] Preferably, the helicase comprises a variant of SEQ ID NO: 5 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E260-P274 and N292-A391), and/or the pin domain (residues K86-E102), and/or the 1A domain (residues M1-L84 and M103-K177).

[0034] Preferably, the helicase comprises a variant of SEQ ID NO: 6 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E266-P280 and N298-A396), and/or the pin domain (residues K91-E107), and/or the 1A domain (residues M1-L90 and M108-M183).

[0035] Preferably, the helicase comprises a variant of SEQ ID NO: 7 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues T276-P290 and N308-P402), and/or the pin domain (residues K100-D116), and/or the 1A domain (residues M1-L99 and D117-M191).

[0036] Preferably, the helicase comprises a variant of SEQ ID NO: 8 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues D274-P288 and N306-A404), and/or the pin domain (residues K95-E112), and/or the 1A domain (residues M1-L95 and I113-K187).

[0037] Preferably, the helicase comprises a variant of SEQ ID NO: 9 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E260-P274 and N292-A391), and/or the pin domain (residues K86-E102), and/or the 1A domain (residues M1-L85 and M103-K177).

[0038] Preferably, the helicase comprises a variant of SEQ ID NO: 10 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into the tower domain (residues E265-P279 and H297-A393), and/or the pin domain (residues K88-E104), and/or the 1A domain (residues M1-L87 and I105-K180).

[0039] Preferably, the helicase comprises a variant of SEQ ID NO: 11, which comprises (i) E105C and/or A362C; (ii) E104C and/or K360C; (iii) E104C and/or A362C; (iv) E104C and/or Q363C; (v) E104C and/or K366C; (vi) E105C and/or M356C; (vii) E105C and/or K360C; (viii) E104C and/or M356C; (ix) E105C and/or Q363C; (x) E105C and/or K366C; (xi) F108C and/or M356C; (xii) F108C and/or K360C; (xiii) F108C and/or A362C; (xiv) F108C and/or Q363C; (xv) F108C and/or K366C; (xvi) K134C and/or M356C; (xvii) K134C and/or K360C; (xviii) K134C and/or A362C; (xix) K134C and/or Q363C; (xx) K134C and/or K366C; (xxi) any of (i) to (xx) and G359C; (xxii) any of (i) to (xx) and Q111C; (xxiii) any of (i) to (xx) and I138C; (xxiv) any of (i) to (xx) and Q111C and I138C; (xxv) E105C and/or F377C; (xxvi) Y103L, E105Y, N352N, A362C and Y365N; (xxvii) E105Y and A362C; (xxviii) A362C; (xxix)

Y103L, E105C, N352N, A362Y and Y365N; (xxx) Y103L, E105C and A362Y; (xxxi) E105C and/or A362C, and I280A; (xxxii) E105C and/or L358C; (xxxiii) E104C and/or G359C; (xxxiv) E104C and/or A362C; (Xxxv) K106C and/or W378C; (xxxvi) T102C and/or N382C; (xxxvii) T102C and/or W378C; (xxxviii) E104C and/or Y355C; (xxxix) E104C and/or N382C; (xl) E104C and/or K381C; (xli) E104C and/or K379C; (xlii) E104C and/or D376C; (xliii) E104C and/or W378C; (xliv) E104C and/or W374C; (xlv) E105C and/or Y355C; (xlvi) E105C and/or N382C; (xlvii) E105C and/or K381C; (xlviii) E105C and/or K379C; (xlix) E105C and/or D376C; (l) E105C and/or W378C; (li) E105C and/or W374C; (lii) E105C and A362Y; (liii) E105C, G359C and A362C; or (liv) I2C, E105C and A362C.

[0040] Preferably, the helicase comprises a variant of any one of SEQ ID NOs: 1 to 10, which comprises a cysteine residue at the positions which correspond to those in SEQ ID NO: 11 as defined in any of (i) to (liv).

[0041] Preferably, the amino acid sequence of the Pif1-like helicase is one of the amino acid sequences shown in SEQ ID NOs: 1 to 11 or has at least 30%, at least 40%, at least 50%, 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.9% homology with one of the amino acid sequences shown in SEQ ID NOs: 1 to 11, and has the ability to control the movement of a polynucleotide.

[0042] In a specific embodiment of the present application, a cysteine residue has been introduced at the amino acid position(s) corresponding to position(s) 10 and/or 362 of SEQ ID NO: 11, for example, position(s) 97 and/or 363 of SEQ ID NO: 2, position(s) 96 and/or 371 of SEQ ID NO: 3, position(s) 94 and/or 361 of SEQ ID NO: 5, position(s) 99 and/or 366 of SEQ ID NO: 6, and position(s) 104 and/or 375 of SEQ ID NO: 8, and the like.

[0043] In order to improve the stability of the Pif1-like helicase of the present application binding to a polynucleotide, and decrease the ability of the helicase to disengage from the polynucleotide, the introduced cysteines are connected to one another, the introduced non-natural amino acids are connected to one another, the introduced cysteine and the introduced non-natural amino acid are connected to one another, the introduced cysteine and a native amino acid of the helicase are connected to one another, or the introduced non-natural amino acid and a native amino acid of the helicase are connected to one another.

[0044] Any number and combination of two more of the introduced cysteines and/or non-natural amino acids may be connected to one another. For instance, 3, 4, 5, 6, 7, 8 or more cysteines and/or non-natural amino acids may be connected to one another. One or more cysteines may be connected to one or more cysteines. One or more cysteines may be connected

to one or more non-natural amino acids, such as Faz. One or more non-natural amino acids, such as Faz, may be connected to one or more non-natural amino acids, such as Faz. One or more cysteines may be connected to one or more native amino acids of the helicase. One or more non-natural amino acids, such as Faz may be connected to one or more native amino acids of the helicase.

[0045] The connection may be in any manner, including direct connection or indirect connection. Preferably, the connection may be a transient contact or a permanent connection. More preferably, the connection may be a non-covalent connection or a covalent connection.

[0046] In a specific embodiment of the present application, the covalent connection may be implemented by using chemical crosslinkers, linear molecules or catalysts. The chemical crosslinkers include, but are not limited to, maleimide, active esters, succinimide, azides, alkynes (such as dibenzocyclooctynol (DIBO or DBCO), difluoro cycloalkynes and linear alkynes), and the like. The chemical crosslinkers may vary in length from one carbon (phosgene-type linkers) to many Angstroms. Examples of linear molecules, include but are not limited to, are polyethyleneglycols (PEGs), polypeptides, polysaccharides, deoxyribonucleic acid (DNA), peptide nucleic acid (PNA), threose nucleic acid (TNA), glycerol nucleic acid (GNA), saturated and unsaturated hydrocarbons, polyamides. The catalysts include, but are not limited to, any catalysts, such as TMAD, that can produce covalent bonds between cysteine residues, between non-natural amino acids, between cysteine residues and non-natural amino acids, between non-natural amino acids and native amino acids or between cysteine residues and native amino acids.

[0047] Preferably, in (A), the native amino acid is substituted with a non-polar amino acid, a polar amino acid or a charged amino acid; the non-polar amino acid includes but not limited to glycine (G), alanine (A) or valine (V); the polar amino acids include but are not limited to serine (S), threonine (T), tyrosine (Y), asparagine (N) or glutamine (Q); the charged amino acids include but are not limited to Aspartic acid (D), glutamic acid (E) or histidine (H).

[0048] Preferably, the Pif1-like helicase is further modified by removal of one or more cysteine residues.

[0049] Preferably, the Pif1-like helicase further comprises a substitution of at least one or more native cysteine residues; more preferably the cysteine residues being substituted with alanine (A), glycine, serine (S), threonine (T), aspartic acid (D) or valine (V).

[0050] Preferably, the Pif1-like helicase comprises a variant of SEQ ID NO: 6, and the one or more substituted native cysteine residues are a combination of (i) C308 and/or C419 and (ii) one or more of C114, C119, and C141.

[0051] Preferably, the Pif1-like helicase comprises a variant of any one of SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11, and the one or more substituted native cysteine residues are shown in Table 3.

[0052] Table 3. Corresponding cysteine residue positions in SEQ ID NOs: 1-11

SEQ ID No:	Amino acid positions
1	C39/C112/C117/C417
2	C112/C117/C139/C202/C307/C327/C344/C395/C415
3	C15/C62/C142/C420/C422
4	C16/C63/C143/C422/C444
5	C109/C114/C136/C199/C302/C414/C394
6	C114/C119/C141/C308/C419
7	C23/C123/C128/C150/C318/C405/C425
8	C119/C124/C146/C209/C407/C424
9	C109/C114/C136/C199/C302/C414/C394
10	C14/C111/C138/C116/C243/C396/C410/C416/C347
11	C125/C128/C412/C315

[0053] Preferably, the helicase is further modified to reduce its surface negative charge.

[0054] Preferably, the Pif1-like helicase further comprises a substitution increasing the net positive charge. More preferably, the substitution increasing the net positive charge comprises a substitution or modification of a negatively charged amino acid, a polar or non-polar amino acid, or introduction a positively charged amino acid at positions near a surface negatively-charged amino acid, a polar or non-polar amino acid. More preferably, the substitution increasing the net positive charge comprises a substitution of negatively charged amino acids, uncharged amino acids, aromatic amino acids, polar or non-polar amino acids with positively charged amino acids. More preferably, the substitution increasing the net positive charge comprises a substitution of negatively charged amino acids, aromatic amino acids, polar or non-polar amino acids with uncharged amino acids. Suitable positively charged amino acids include, but are not limited to, histidine (H), lysine (K) and/or arginine (R). Uncharged amino acids have no net charge. Suitable uncharged amino acids include, but are not limited to, cysteine (C), serine (S), threonine (T), methionine (M), asparagine (N) or glutamine (Q). Non-polar amino acids have non-polar side chains. Non-polar amino acids include, but are not limited to, glycine (G), alanine (A), proline (P), isoleucine (I), leucine (L) or valine (V).

Aromatic amino acids have aromatic side chains. Suitable aromatic amino acids include, but are not limited to, histidine (H), phenylalanine (F), tryptophan (W) or tyrosine (Y).

[0055] The positively charged amino acids, uncharged amino acids, polar, non-polar amino acids, or aromatic amino acids may be natural or non-natural amino acids, which may be artificially synthesized or modified natural amino acids.

[0056] Preferred substitutions include, but are not limited to, a substitution of glutamic acid (E) with arginine (R), a substitution of glutamic acid (E) with lysine (K), and a substitution of glutamic acid (E) with asparagine (N), a substitution of aspartic acid (D) with lysine (K), and a substitution of aspartic acid (D) with arginine (R).

[0057] Preferably, the Pif1-like helicase comprises a variant of SEQ ID NO: 6, and the one or more negatively charged or uncharged amino acids are one or more of S9, S173, D208 or T218.

[0058] Preferably, the Pif1-like helicase comprises a variant of SEQ ID NO: 7, and the one or more negatively charged or uncharged amino acids are one or more of D8, E11, T26, S186, D216 or S226. Any number of these amino acids may be neutralised, such as 1, 2, 3, 4, 5, 6, 7 or 8 of them.

[0059] Preferably, the Pif1-like helicase comprises a variant of any one of SEQ ID NOs: 1 to 5 and 8 to 11, and the one or more negatively charged or uncharged amino acids correspond to one or more of S9, S173, D208 or T218 in SEQ ID NO: 6; or one or more of D8, E11, T26, S186, D216 or S226 in SEQ ID NO: 7. Amino acids in SEQ ID NOs: 1 to 5 and 8 to 11 which correspond to S9, S173, D208 or T218 in SEQ ID NO: 6; or one or more of D8, E11, T26, S186, D216 or S226 in SEQ ID NO: 7 may be determined using the alignment shown in Figures 11A and 11B.

[0060] Preferably, the non-natural amino acid is selected from 4-Azido-L-phenylalanine (Faz), 4-Acetyl-L-phenylalanine, 3-Acetyl-L-phenylalanine, 4-Acetoacetyl-L-phenylalanine, O-Allyl-L-tyrosine, 3-(Phenylselanyl)-L-alanine, O-2-Propyn-1-yl-L-tyrosine, 4-(Dihydroxyboryl)-L-phenylalanine, 4-[(Ethylsulfanyl)carbonyl]-L-phenylalanine, (2S)-2-amino-3-{4-[(propan-2-ylsulfanyl)carbonyl]phenyl}propanoic acid, (2S)-2-amino-3-{4-[(2-amino-3-sulfanylpropanoyl)amino]phenyl}propanoic acid, O-Methyl-L-tyrosine, 4-Amino-L-phenylalanine, 4-Cyano-L-phenylalanine, 3-Cyano-L-phenylalanine, 4-Fluoro-L-phenylalanine, 4-Iodo-L-phenylalanine, 4-Bromo-L-phenylalanine, O-(Trifluoromethyl)tyrosine, 4-Nitro-L-phenylalanine, 3-Hydroxy-L-tyrosine, 3-Amino-L-tyrosine, 3-Iodo-L-tyrosine, 4-Isopropyl-L-phenylalanine, 3-(2-Naphthyl)-L-alanine, 4-Phenyl-L-phenylalanine, (2S)-2-amino-3-(naphthalen-2-ylamino)propanoic acid, 6-

(Methylsulfanyl)norleucine, 6-Oxo-L-lysine, D-tyrosine, (2R)-2-Hydroxy-3-(4-hydroxyphenyl)propanoic acid, (2R)-2-Ammoniooctanoate-3-(2,2'-Bipyridin-5-yl)-D-alanine, 2-amino-3-(8-hydroxy-3-quinolyl)propanoic acid, 4-Benzoyl-L-phenylalanine, S-(2-Nitrobenzyl)cysteine, (2R)-2-amino-3-[(2-nitrobenzyl)sulfanyl]propanoic acid, (2S)-2-amino-3-[(2-nitrobenzyl)oxy]propanoic acid, O-(4,5-Dimethoxy-2-nitrobenzyl)-L-serine, (2S)-2-amino-6-({[(2-nitrobenzyl)oxy]carbonyl}amino)hexanoic acid, O-(2-Nitrobenzyl)-L-tyrosine, 2-Nitrophenylalanine, 4-[(E)-Phenyldiazenyl]-L-phenylalanine, 4-[3-(Trifluoromethyl)-3H-diaziren-3-yl]-D-phenylalanine, 2-amino-3-[[5-(dimethylamino)-1-naphthyl]sulfonylamino]propanoic acid, (2S)-2-amino-4-(7-hydroxy-2-oxo-2H-chromen-4-yl)butanoic acid, (2S)-3-[(6-acetylnaphthalen-2-yl)amino]-2-aminopropanoic acid, 4-(Carboxymethyl)phenylalanine, 3-Nitro-L-tyrosine, O-Sulfo-L-tyrosine, (2R)-6-Acetamido-2-ammoniohexanoate, 1-Methylhistidine, 2-Aminononanoic acid, 2-Aminodecanoic acid, L-Homocysteine, 5-Sulfanylnorvaline, 6-Sulfanyl-L-norleucine, 5-(Methyl sulfanyl)-L-norvaline, N⁶-{[(2R,3R)-3-Methyl-3,4-dihydro-2H-pyrrol-2-yl]carbonyl}-L-lysine, N⁶-[(Benzyloxy)carbonyl]lysine, (2S)-2-amino-6-[(cyclopentylcarbonyl)amino]hexanoic acid, N⁶-[(Cyclopentyl)oxy]carbonyl-L-lysine, (2S)-2-amino-6-{{(2R)-tetrahydrofuran-2-ylcarbonyl}amino}hexanoic acid, (2S)-2-amino-8-[(2R,3S)-3-ethynyltetrahydrofuran-2-yl]-8-oxooctanoic acid, N⁶-(tert-Butoxycarbonyl)-L-lysine, (2S)-2-Hydroxy-6-({[(2-methyl-2-propanyl)oxy]carbonyl}amino)hexanoic acid, N⁶-[(Allyloxy)carbonyl]lysine, (2S)-2-amino-6-({[(2-azidobenzyl)oxy]carbonyl}amino)hexanoic acid, N⁶-L-Prolyl-L-lysine, (2S)-2-amino-6-{{[(prop-2-yn-1-yloxy)carbonyl]amino}hexanoic acid and N⁶-[(2-Azidoethoxy)carbonyl]-L-lysine.

[0061] More preferably, in (B), at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is substituted with an amino acid containing a larger side chain (R group).

[0062] Preferably, the Pif1-like helicase comprises substitution of at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA with an amino acid containing a larger side chain (R group). Any number of amino acids may be substituted, such as, 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, or 6 or more amino acids. Each amino acid may interact with a base, a sugar, or a base and a sugar ring. Protein modeling may be used to identify amino acids that interact with a sugar ring and/or base of one or more nucleotides in single-stranded or double-stranded DNA.

[0063] Preferably, the Pif1-like helicase comprises (a) a variant of SEQ ID NO: 11, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more

nucleotides in single-stranded DNA or double-stranded DNA is at least one of P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 or P100. These numbers correspond to the relevant positions in SEQ ID NO: 11; or (b) a variant of SEQ ID NO: 6, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is a combination of (i) H87, V422 and/or I282 and (ii) at least one of P94, F103, V155, P67, M124, D126, E156, P157, E279, S293, N93, H403 or F246. Relative to SEQ ID NOs: 11 and 6, it may be necessary to make changes when one or more amino acids have been inserted or deleted in the variant. As mentioned above, those skilled in the art may determine the corresponding positions in the variant. More preferably, the Pif1-like helicase comprises a variant of SEQ ID NO: 11, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is F109 and one or more of P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 or P100.

[0064] Preferably, the Pif1-like helicase comprises a variant of any one of SEQ ID NOs: 1 to 5 and 7 to 10, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is at least one amino acid corresponding to P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 or P100 in SEQ ID NO: 11, or amino acid corresponding P94, F103, I282, V155, P67, M124, D126, E156, P157, E279, V422, S293, H87, N93, H403 or F246 in SEQ ID NO: 6. More preferably, the Pif1-like helicase comprises a variant of any one of SEQ ID NOs: 1 to 10, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is amino acid corresponding to F109 and one or more of amino acids corresponding to P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 or P100 in SEQ ID NO: 11.

[0065] Table 4 shows amino acids in SEQ ID NOs: 1 to 10 corresponding to P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 and P100 in SEQ ID NO: 11.

[0066] Table 4. Amino acid positions corresponding to SEQ ID NOs: 1 to 10

SEQ ID NO	Amino acid positions
-----------	----------------------

ID NO :																
11	P7 3	H9 3	N99	P10 0	F10 9	F130	D13 2	A16 1	D16 2	D16 3	Y24 4	E27 7	I280	Q29 1	H39 6	K41 5
1	P6 6	H8 5	N91	P92	F10 1	M12 2	G12 4	V15 3	E15 4	P15 5	F24 4	E27 7	I281	S29 1	H40 1	I420
2	P6 5	H8 5	N92	P92	F10 1	M12 2	D12 4	V15 3	A15 4	P15 5	F24 5	E27 8	M28 1	T29 2	H39 9	S41 8
3	P6 5	H8 5	H91	P92	F10 1	M12 6	D12 8	V15 7	R15 8	H15 9	F24 6	E27 9	M28 2	T29 3	H41 0	F42 9
4	P6 5	H8 5	H91	P92	F10 1	M12 6	D12 8	V15 7	R15 8	H15 9	F24 6	E27 9	M28 2	T29 3	H41 1	F43 0
5	P6 3	H8 2	S88	P89	F98	M11 9	D12 1	V15 0	S15 1	P15 2	F24 0	E27 3	M27 6	N28 7	H39 8	L41 7
6	P6 7	H8 7	N93	P94	F10 3	M12 4	D12 6	V15 5	E15 6	P15 7	F24 6	E27 9	I282	S29 3	H40 3	V42 2
7	P7 6	H9 6	N10 2	P10 3	F11 2	M13 3	D13 5	V16 4	A16 5	P16 6	F25 6	E29 0	M29 2	T30 3	H40 9	R42 8
8	P7 2	H9 2	S98	P99	F10 8	M12 9	D13 1	V16 0	S16 1	P16 2	Y25 4	G28 7	T29 1	K30 1	H41 1	S43 0
9	P6 3	H8 2	S88	P89	F98	M11 9	D12 1	V15 0	S15 1	P15 2	Y24 0	E27 3	M27 7	N28 7	H39 8	L41 7
10	P6 4	H8 4	N91	P91	F10 0	M12 1	D12 3	V15 2	E15 3	L15 4	Y24 5	E27 8	T28 1	S29 2	H40 0	K41 9

[0067] The larger side chain (R group) preferably (a) comprises an increased number of carbon atoms, (b) has an increased length, (c) has an increased molecular volume, and/or (d) has an increased van der Waals volume. The larger side chain (R group) is preferably (a); (b); (c); (d); (a) and (b); (a) and (c); (a) and (d); (b) and (c); (b) and (d); (c) and (d); (a), (b) and (c); (a), (b) and (d); (a), (c) and (d); (b), (c) and (d); or (a), (b), (c) and (d). Each of (a) to (d) may be measured by standard methods in the art.

[0068] More preferably, the larger side chain (R group) increases (i) electrostatic interaction, (ii) hydrogen bonding, (iii) cation- π (cation- π) interaction and/or (iv) π - π interaction between

the at least one amino acid and one or more nucleotides in the single-stranded DNA or double-stranded DNA. For example, in (i), positively charged amino acids such as arginine (R), histidine (H), and lysine (K) have R groups that increase electrostatic interaction. For example, in (ii), amino acids such as asparagine (N), serine (S), glutamine (Q), threonine (T), and histidine (H) have R groups that increase hydrogen bonding. For example, in (iii), aromatic amino acids such as phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H) have R groups that increase cation-pi (cation- π) interaction.

[0069] The amino acid containing the larger side chain (R) may be non-natural amino acid.

[0070] In a specific embodiment of the present application, the amino acid containing the larger side chain (R group) is not alanine (A), cysteine (C), glycine (G), selenocysteine (U), methionine (M), aspartic acid (D) or glutamic acid (E).

[0071] In a specific embodiment of the present application, the Pif1-like helicase comprises one or more of the following substitutions:

[0072] a) Histidine (H) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q) or asparagine (N); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); (iv) tyrosine (Y), arginine (R) or glutamine (Q); or (v) arginine (R), tyrosine (Y), asparagine (N) or glutamine (Q). Histidine (H) is more preferably substituted with iv) Y, R or Q or (v) R, Y, N or Q.

[0073] b) Asparagine (N) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); or (iv) glutamine (Q), arginine (R), histidine (H) or lysine (K). Asparagine (N) is more preferably substituted with Q, R, H or K.

[0074] c) Proline (P) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N), threonine (T) or histidine (H); (iii) tyrosine (Y), phenylalanine (F) or tryptophan (W); (iv) leucine (L), valine (V) or isoleucine (I); or (v) tryptophan (W), phenylalanine (F) or leucine (L). Proline (P) is more preferably substituted with W, For L.

[0075] d) Valine (V) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); (iv) isoleucine (I) or leucine (L); or (v) tyrosine (Y), arginine (R), histidine (H) or tryptophan (W). Valine (V) is more preferably substituted with Y, R, H or W.

[0076] e) Phenylalanine (F) is preferably substituted with (i) arginine (R) or lysine (K); (ii) histidine (H); (iii) tyrosine (Y) or tryptophan (W); or (iv) arginine (R), tyrosine (Y), glutamine (Q) or histidine (H). Phenylalanine (F) is more preferably substituted with R, Y, Q or H.

[0077] f) Glutamine (Q) is preferably substituted with (i) arginine (R) or lysine (K) or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W).

[0078] g) Alanine (A) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W) or (iv) isoleucine (I) or leucine (L).

[0079] h) Serine (S) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); (iv) isoleucine (I) or leucine (L); or (v) arginine (R), asparagine (N) or glutamine (Q). Serine (S) is preferably substituted with R, N or Q.

[0080] i) Lysine (K) is preferably substituted with (i) arginine (R) or (ii) tyrosine (Y) or tryptophan (W).

[0081] j) Arginine (R) is preferably substituted with (i) Tyrosine (Y) or Tryptophan (W).

[0082] k) Methionine (M) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H) or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W).

[0083] l) Leucine (L) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q) or asparagine (N) or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W).

[0084] m) Aspartic acid (D) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W). Aspartic acid (D) is more preferably substituted with H, Y or K.

[0085] n) Glutamic acid (E) is preferably substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H) or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W).

[0086] o) Isoleucine (I) is preferably substituted with (i) phenylalanine (F) or tryptophan (W); (ii) valine (V) or leucine (L); or (iii) histidine (H), lysine (K) or arginine (R).

[0087] p) Tyrosine (Y) is preferably substituted with (i) arginine (R) or lysine (K); or (ii) tryptophan (W). Tyrosine (Y) is more preferably substituted with W or R.

[0088] In a specific embodiment of the present application, the Pif1-like helicase more preferably comprises a variant of SEQ ID NO: 11 comprising:

[0089] P73L; P73V; P73I; P73E; P73T; P73F; H93N; H93Q; H93W; N99R; N99H; N99W; N99Y; P100L; P100V; P100I; P100E; P100T; P100F; F130W; F130Y; F130H; D132H; D132Y; D132K; A161I; A161L; A161N; A161W; A161H; D162H; D162Y; D162K; D163W; D163F; D163Y; D163H; D163I; D163L; D163V; Y244W; Y244H; E277G; I280H; I280K; I280W; Q291K; Q291R; Q291W; Q291F; H396N; H396Q; H396W; K415W; K415R; K415H;

K415Y; F109W/P73L; F109W/P73V; F109W/P73I; F109W/P73E; F109W/P73T; F109W/P73F; F109W/H93N; F109W/H93Q; F109W/H93W; F109W/N99R; F109W/N99H; F109W/N99W; F109W/N99Y; F109W/P100L; F109W/P100V; F109W/P100I; F109W/P100E; F109W/P100T; F109W/P100F; F109W/F130W; F109W/F130Y; F109W/F130H; F109W/D132H; F109W/D132Y; F109W/D132K; F109W/A161I; F109W/A161L; F109W/A161N; F109W/A161W; F109W/A161H; F109W/D162H; F109W/D162Y; F109W/D162K; F109W/D163W; F109W/D163F; F109W/D163Y; F109W/D163H; F109W/D163I; F109W/D163L; F109W/D163V; F109W/Y244W; F109W/Y244H; F109W/E277G; F109W/I280H; F109W/I280K; F109W/I280W; F109W/Q291K; F109W/Q291R; F109W/Q291W; F109W/Q291F; F109W/H396N; F109W/H396Q; F109W/H396W; F109W/K415W; F109W/K415R; F109W/K415H; or F109W/K415Y.

[0090] In a specific embodiment of the present application, the Pif1-like helicase more preferably comprises a variant of SEQ ID NO: 6 comprising:

[0091] P94W; P94F; F103W; I282F; V155L; V155I; D126H; D126N; D126Q; P157W; P157F; P157L; V422Y; V422R; V422H; V422W; S293R; S293N; S293Q; H87R; H87Y; H87N; H87Q; N93Q; N93R; N93K; N93H; H403Y; H403R; H403Q; F246R; F246Y; or F246Q.

[0092] The helicase of the present application is preferably a helicase in which at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is substituted. Any number of amino acids may be substituted, for example, 1 or more, 2 or more, 3 or more, 4 or more, 5 or more, or 6 or more amino acids. The nucleotides in ssDNA each contain three phosphate groups. Each amino acid that is substituted may interact with any number of phosphate groups at a time, for example, one, two, or three phosphate groups at a time. Protein modeling may be used to identify amino acids that interact with one or more phosphate groups.

[0093] The substitution preferably increases (i) electrostatic interaction; (ii) hydrogen bonding and/or (iii) cation- π (cation- π) interaction between the at least one amino acid and one or more phosphate groups in the ssDNA or dsDNA. The preferred substitutions of (i), (ii) and (iii) are discussed below using the labels (i), (ii) and (iii).

[0094] The substitution preferably increases the net positive charge of the position. Methods known in the art may be used to measure the net charge at any position. For example, the isoelectric point may be used to define the net charge of an amino acid. The net charge is usually measured at about 7.5. The substitution is preferably a substitution of a negatively charged amino acid with a positively charged, uncharged, non-polar or aromatic amino acid. A negatively charged amino acid is an amino acid with a net negative charge. Negatively charged

amino acids include, but are not limited to, aspartic acid (D) and glutamic acid (E). A positively charged amino acid is an amino acid with a net positive charge. The positively charged amino acids may be naturally occurring or non-naturally occurring. The positively charged amino acids may be synthetic or modified. For example, modified amino acids with a net positive charge may be specifically designed for use in the present application. Many different types of modifications to amino acids are well known in the art. Preferred naturally occurring positively charged amino acids include, but are not limited to, histidine (H), lysine (K), and arginine (R).

[0095] Uncharged amino acids, non-polar amino acids, or aromatic amino acids may be naturally occurring or non-naturally occurring. It may be synthetic or modified. Uncharged amino acids have no net charge. Suitable uncharged amino acids include, but are not limited to, cysteine (C), serine (S), threonine (T), methionine (M), asparagine (N) and glutamine (Q). Non-polar amino acids have non-polar side chains. Suitable non-polar amino acids include, but are not limited to, glycine (G), alanine (A), proline (P), isoleucine (I), leucine (L) and valine (V). Aromatic amino acids have aromatic side chains. Suitable aromatic amino acids include, but are not limited to, histidine (H), phenylalanine (F), tryptophan (W), and tyrosine (Y).

[0096] The Pif1-like helicase preferably comprises a variant of SEQ ID NO: 11, in which the at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is at least one of H75, T91, S94, K97, N246, N247, N284, K288, N297, T394 or K397. The Pif1-like helicase preferably comprises a variant of SEQ ID NO: 6, in which the at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is at least one of K91, T85, R88, H69, K404, T401, N299, N248, E280, K290 or K249. These numbers correspond to the relevant positions in SEQ ID NO: 11 or 6. Relative to SEQ ID NO: 11 or 6, it may be necessary to make changes when one or more amino acids have been inserted or deleted in the variant. Those skilled in the art may determine the corresponding positions in the variants described above.

[0097] The Pif1-like helicase preferably comprises a variant of any one of SEQ ID NOs: 1 to 5 and 7 to 10, in which the at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is at least one amino acid corresponding to H75, T91, S94, K97, N246, N247, N284, K288, N297, T394 or K397 in SEQ ID NO: 11, or at least one amino acid corresponding to K91, T85, R88, H69, K404, T401, N299, N248, E280, K290 or K249 in SEQ ID NO: 6.

[0098] Table 5 shows the amino acids in SEQ ID NO: 1 to 10 corresponding to H75, T91, S94, K97, N246, N247, N284, K288, N297, T394, and K397 in SEQ ID NO: 11.

[0099] Table 5. Amino acids positions in SEQ ID NOs: 1 to 10 corresponding to SEQ ID NO: 11.

SEQ ID NO:	Amino acids positions										
11	H75	T91	S94	K97	N24 6	N24 7	N28 4	K28 8	N29 7	T39 4	K39 7
1	H67	T83	R86	K89	N24 6	K24 7	E28 4	K28 8	N29 7	T39 9	K40 2
2	H67	T83	S86	K89	N24 7	K24 8	K28 5	K28 9	N29 8	T39 7	K40 0
3	H67	T83	S86	K89	N24 8	D24 9	E28 6	E29 0	N29 9	T40 8	K41 1
4	H67	T83	S86	K89	N24 8	D24 9	E28 6	E29 0	N29 9	T40 9	K41 2
5	H64	T80	S83	K86	N24 2	K24 3	E28 0	K28 4	N29 3	T39 6	K39 9
6	H69	T85	R88	K91	N24 8	K24 9	E28 6	K29 0	N29 9	T40 1	K40 4
7	H78	T94	S97	K10 0	N25 8	K25 9	T29 6	K30 0	N30 9	T40 7	K41 0
8	H74	T90	S93	K96	N25 6	K25 7	T29 4	K29 8	N30 7	T40 9	K41 2
9	H64	T80	S83	K86	N24 2	K24 3	E28 0	K28 4	N29 3	T39 6	K40 0
10	H66	T82	S85	K86	N24 7	E24 8	K28 5	K28 9	N29 8	T39 8	K40 1

[00100] Preferably, the Pif1-like helicase comprises any one or more of the following:

[00101] a) Histidine (H) is substituted with (i) arginine (R) or lysine (K); (ii) asparagine (N), serine (S), glutamine (Q) or threonine (T); (iii) phenylalanine (F), tryptophan (W) or tyrosine (Y); or (iv) asparagine (N), glutamine (Q) or arginine (R);

[00102] b) Threonine (T) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) asparagine (N), serine (S), glutamine (Q) or histidine (H); (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); (iv) asparagine (N), glutamine (Q) or arginine (R); or (v) asparagine (N), histidine (H), lysine (K) or arginine (R);

[00103] c) Serine (S) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) asparagine (N), glutamine (Q), threonine (T) or histidine (H); or (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H);

[00104] d) Asparagine (N) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) serine (S), glutamine (Q), threonine (T) or histidine (H); (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); or (iv) glutamine (Q), arginine (R), histidine (H) or lysine (K);

[00105] e) Lysine (K) is substituted with (i) arginine (R) or histidine (H); (ii) asparagine (N), serine (S), glutamine (Q), threonine (T) or histidine (H); (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); or (iv) arginine (R), glutamine (Q) or asparagine (N); and/or,

[00106] f) Arginine (R) is substituted with (i) asparagine (N), serine (S) or glutamine (Q).

[00107] In a specific embodiment of the present application, the Pif1-like helicase is a variant of SEQ ID NO: 11 comprising one or more of (a) to (k): (a) H75N, H75Q, H75K or H75F; (b) T91K, T91Q or T91N; (c) S94H, S94N, S94K, S94T, S94R or S94Q; (d) K97Q, K97H or K97Y; (e) N246H or N246Q; (f) N247Q or N247H; (g) N284H or N284Q; (h) K288Q or K288H; (i) N297Q, N297K or N297H; (j) T394K, T394H or T394N; or (k) K397R, K397H or K397Y.

[00108] In a specific embodiment of the present application, the Pif1-like helicase is a variant of SEQ ID NO: 6, which comprises one or more of (a) to (i): (a) K91R; (b) T85N, T85Q or T85R; (c) R88N or R88Q; (d) H69N, H69Q or H69R; (e) K404R; (f) T401N, T401H, T401K or T401R; (g) N299Q, N299R, N299H or N299K; (h) N248Q, N248R, N248H or N248K; or (i) K249R, K249Q or K249N.

[00109] More preferably, in (B), at least one amino acid that interacts with the double strand of one or more nucleotides in a double stranded DNA is substituted.

[00110] Preferably, the Pif1-like helicase comprises:

[00111] (a) a variant of SEQ ID NO: 6, in which the at least one amino acid that interacts with the double strand of one or more nucleotides in a double dsDNA is at least one of M81, V59, Q52, E286 or K290; or,

[00112] (b) a variant of any one of SEQ ID NOs: 1 to 5 and 7 to 11, in which the at least one amino acid that interacts with the double strand of one or more nucleotides in a double dsDNA is at least one amino acid corresponding to M81, V59, Q52, E286 or K290 in SEQ ID NO: 6.

[00113] Preferably, the Pif1-like helicase is a variant of SEQ ID NO: 6 comprising one or more of (a) to (e): (a) M81K, M81R or M81H; (b) V59K, V59R or V59H; (c) Q52K, Q52R or Q52H; (d) E286K, E286R or E286H; or (e) K290R.

[00114] Preferably, the Pif1-like helicase is a variant of SEQ ID NO: 6, which comprises one or more of the following amino acid substitutions: T85N, H87Y, H87Q, H87N, R88Q, R88N, V155I, V155L, K91R, F103W, S239N, F246R, F246Y, K249R, I282F, E286K or V422H. The helicase of the present application is also such a helicase, wherein a part of the helicase that interacts with a transmembrane pore contains one or more modifications, preferably one or more substitutions. Further preferably, the helicase comprises substitution of at least one amino acid that interacts with a transmembrane pore. The part of the helicase that interacts with a transmembrane pore is usually the part of the helicase that interacts with the transmembrane pore when the helicase is used to control the movement of polynucleotides through the pore, for example, as discussed in more details below. When helicases are used to control the movement of polynucleotides through a pore, the part usually contains amino acids that interact or contact with the pore. When an electric potential is applied, and the helicase is bound or linked to a polynucleotide that is moving through the pore, the part usually contains amino acids that interact or contact with the pore.

[00115] The part that interacts with the transmembrane pore preferably comprises a variant of SEQ ID NO: 11, wherein one or more, such as 2, 3, 4 or 5 amino acids on E196, W202, N199 or G201 are substituted.

[00116] The part that interacts with a transmembrane pore preferably comprises a variant of any one of SEQ ID NOs: 1 to 10, which comprises substitutions of at least one or more of the amino acids corresponding to SEQ ID NO: 11 at positions (a) E196; (b) W202; (c); N199; or (d) G201.

[00117] Table 6 shows amino acids in SEQ ID NOs: 1 to 10 corresponding to E196, W202, N199, and G201 in SEQ ID NO: 11.

[00118] Table 6. Amino acid positions in SEQ ID NOs: 1 to 10 corresponding to SEQ ID NO: 11

SEQ ID NO:	Amino acid positions			
11	E196	W202	N199	G201
1	E193	W199	M196	S198
2	E192	W198	H195	K197
3	K196	W202	N199	G201
4	K196	W202	N199	G201
5	E189	W195	N192	G194
6	D195	W201	T198	G200
7	E203	W209	K206	G208
8	E199	W205	K202	Q204
9	E189	W195	T192	G194
10	S192	W198	T195	G197

[00119] In a specific embodiment of the present application, the Pif1-like helicase is a variant of SEQ ID NO. 11 comprising substitutions at the following positions:

[00120] -F109/E196/H75, such as, F109W/E196L/H75N, F109W/E196L/H75Q, F109W/E196L/H75K or F109W/E196L/H75F;

[00121] -F109/E196/T91, such as, F109W/E196L/T91K, F109W/E196L/T91Q or F109W/E196L/T91N;

[00122] -F109/S94/E196, such as, F109W/S94H/E196L, F109W/S94T/E196L, F109W/S94R/E196L, F109W/S94Q/E196L, F109W/S94N/E196L, or F109W/S94K/E196L;

[00123] F109/N99/E196, such as, F109W/N99R/E196L, F109W/N99H/E196L, F109W/N99W/E196L or F109W/N99Y/E196L;

[00124] -F109/S94/E196/I280, such as, F109W/S94H/E196L/I280K;

[00125] -F109/P100/E196, such as, F109W/P100L/E196L, F109W/P100V/E196L, F109W/P100I/E196L or

[00126] F109W/P100T/E196L;

[00127] -F109/D132/E196, such as, F109W/D132H/E196L, F109W/D132Y/E196L or F109W/D132K/E196L;

[00128] -F109/A161/E196, such as, F109W/A161I/E196L, F109W/A161L/E196L, F109W/A161N/E196L, F109W/A161W/E196L or F109W/A161H/E196L;

[00129] -F109/D163/E196, such as, F109W/D163W/E196L, F109W/D163F/E196L, F109W/D163Y/E196L, F109W/D163H/E196L, F109W/D163I/E196L, F109W/D163L/E196L or F109W/D163V/D163L/E196L;

[00130] -F109/Y244/E196, such as, F109W/Y244W/E196L, F109W/Y244Y/E196L or F109W/Y244H/E196L;

[00131] -F109/N246/E196, such as, F109W/N246H/E196L or F109W/N246Q/E196L;

[00132] -F109/E196/I280, such as, F109W/E196L/I280K, F109W/E196L/I280H, F109W/E196L/I280W or F109W/E196L/I280R;

[00133] -F109/E196/Q291, such as, F109W/E196L/Q291K, F109W/E196L/Q291R, F109W/E196L/Q291W or F109W/E196L/Q291F;

[00134] -F109/N297/E196, such as, F109W/N297Q/E196L, F109W/N297K/E196L or F109W/N297H/E196L;

[00135] -F109/T394/E196, such as, F109W/T394K/E196L, F109W/T394H/E196L or F109W/T394N/E196L;

[00136] -F109/H396/E196, such as, F109W/H396Y/E196L, F109W/H396F/E196L, F109W/H396Q/E196L or F109W/H396K/E196L;

[00137] -F109/K397/E196, such as, F109W/K397R/E196L, F109W/K397H/E196L or F109W/K397Y/E196L; or,

[00138] -F109/Y416/E196, such as, F109W/Y416W/E196L or F109W/Y416R/E196L.

[00139] A variant of a Pif1-like helicase is an enzyme that has an amino acid sequence which varies from that of the wild-type helicase and which retains polynucleotide binding activity. In particular, a variant of any one of SEQ ID NOs: 1 to 11 is an enzyme that has an amino acid sequence which varies from that of any one of SEQ ID NOs: 1 to 11 and which retains polynucleotide binding activity. Polynucleotide binding activity can be determined using methods known in the art. Suitable methods include, but are not limited to, fluorescence anisotropy, tryptophan fluorescence and electrophoretic mobility shift assay (EMSA). For instance, the ability of a variant to bind a single stranded polynucleotide can be determined as described in the Examples.

[00140] Over the entire length of the amino acid sequence of any one of SEQ ID NOs: 1 to 11, a variant will preferably be at least 20% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 30%, at least 40%, at least 45%), at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of any one of SEQ ID NOs: 1 to 11 over the entire sequence. There may be at least 70%, for example at least 80%, at least 85%, at least 90% or at least 95%, amino acid identity over a stretch of 100 or more, for example 150, 200, 300, 400 or 500 or more, contiguous amino acids ("strict homology").

[00141] In the second aspect of the present application, provided is a construct, which comprises at least one or more Pif1-like helicases of the present application.

[00142] Preferably, the construct further comprises a polynucleotide binding moiety.

[00143] Preferably, the construct has the ability to control the movement of a polynucleotide.

[00144] Preferably, the polynucleotide binding moiety may be a moiety that binds to a base of the polynucleotide, and/or a moiety that binds to a sugar ring of the polynucleotide, and/or a moiety that binds to a phosphate group of the polynucleotide.

[00145] Preferably, the Pif1-like helicase and polynucleotide binding moiety constituting the construct can be prepared separately and then directly connected. The construct can also be directly prepared by genetic fusion. For example, the nucleotides encoding the Pif1-like helicase and the polynucleotide binding moiety are connected, then transferred into host cells for expression, and purified.

[00146] More preferably, the polynucleotide binding moiety is a polypeptide capable of binding to a polynucleotide, including but not limited to one or a combination of two or more of eukaryotic single-chain binding proteins, bacterial single-chain binding proteins, archaea single-chain binding proteins, and viral single-chain binding proteins or double-stranded binding proteins.

[00147] In a specific embodiment of the present application, the polynucleotide binding moiety includes but is not limited to any one of those shown in Table 7:

[00148] Table 7. Binding moiety bound to a polynucleotide

Name	Classification	Source	UniProt Database	Molecular Weight (Da)
SSBEco	ssb	Escherichia coli	P0AGE0	18975
SSBBhe	ssb	Bartonella henselae	Q6G302	19312
SSBCbu	ssb	Coxiella burnetii	Q83EP4	17437
SSBTma	ssb	Thermatogamaritima	Q9WZ73	16298
SSBHpy	ssb	Helicobacter pylori	O25841	20043
SSBDra	ssb	Deinococcus radiodurans	Q9RY51	32722
SSBTaq	ssb	Thermus aquaticus	Q9KH06	30029
SSBMsm	ssb	Mycobacterium	Q9AFI5	17401

		smegmatis		
SSBSso	ssb/RPA	Sulfolobussolfataricus	Q97W73	16138
SSBSso7 D	ssb	Sulfolobussolfataricus	P39476	7280
SSBMHs mt	ssb	Homo sapiens	Q04837	17260
SSBMle	ssb	Mycobacterium leprae	P46390	17701
gp32T4	ssb	Bacteriophage T4	P03695	33506
gp32RB6 9	ssb	Bacteriophage RB69	Q7Y265	33118
gp2.5T7	ssb	Bacteriophage T7	P03696	25694

[00149] In the third aspect of the present application, provided is a nucleic acid that encodes the Pif1-like helicase or the construct of the present application.

[00150] In the fourth aspect of the present application, provided is an expression vector comprising the nucleic acid of the present application.

[00151] Preferably, the nucleic acid is operably linked to a regulatory element in an expression vector, wherein the regulatory element is preferably a promoter.

[00152] In a specific embodiment of the present application, the promoter is selected from T7, trc, lac, ara or λ_L .

[00153] Preferably, the expression vector includes but is not limited to a plasmid, virus or phage.

[00154] In the fifth aspect of the present application, provided is a host cell, comprising the nucleic acid or the expression vector of the present application.

[00155] Preferably, the host cell includes but is not limited to *Escherichia coli*.

[00156] In a specific embodiment of the present application, the host cell is selected from BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, C41 (DE3), Rosetta2 (DE3), Origami, Origami B, and the like.

[00157] In the sixth aspect of the present application, provided is a method for preparing Pif1-like helicase. The method comprises providing a wild-type Pif1-like helicase, and then modifying the wild-type Pif1-like helicase to obtain the Pif1-like helicase of the present application.

[00158] In the seventh aspect of the present application, provided is a method for preparing Pif1-like helicase. The method comprises culturing the host cell of the present

application, and performing an inducible expression and purification to obtain the Pif1-like helicase.

[00159] In a specific embodiment of the present application, the method comprises obtaining a nucleic acid sequence encoding a Pif1-like helicase according to the amino acid sequence of the Pif1-like helicase of the present application, digesting the nucleic acid sequence and ligating to an expression vector, then transforming into *E. coli*, and performing an inducible expression and purification to obtain the Pif1-like helicase.

[00160] In the eighth aspect of the present application, provided is a method of controlling the movement of a polynucleotide, comprising contacting the Pif1-like helicase or the construct of the present application with the polynucleotide.

[00161] Preferably, said controlling the movement of the polynucleotide is controlling the movement of the polynucleotide through a pore. The pore is a nanopore, and the nanopore is a transmembrane pore. The pore may be a natural or artificial pore, including but not limited to a protein pore, a polynucleotide pore or a solid state pore.

[00162] In a specific embodiment of the present application, the transmembrane pore is selected from a biological pore, a solid state pore, or a biological and solid state hybrid pore.

[00163] In a specific embodiment of the present application, the pores include, but are not limited to, those derived from *Mycobacterium smegmatis* porin A, *Mycobacterium smegmatis* porin B, *Mycobacterium smegmatis* porin C, and *Mycobacterium smegmatis* porin D, hemolysin, lysin, interleukin, outer membrane porin F, outer membrane porin G, outer membrane phospholipase A, WZA or *Neisseria* autotransporter lipoprotein, etc.

[00164] Preferably, the method may include controlling the movement of the polynucleotide with one or more Pif1-like helicases.

[00165] In the ninth aspect of the present application, provided is a method for characterising a target polynucleotide, comprising:

[00166] I) contacting the Pif1-like helicase or the construct of the present application with the target polynucleotide and a pore, such that the Pif1-like helicase or the construct controls the movement of the target polynucleotide through the pore; and

[00167] II) taking one or more characteristics of the target polynucleotide when the nucleotide of the target polynucleotide interacts with the pore, and thereby characterising the target polynucleotide.

[00168] Any number of Pif1-like helicases of the present application may be used in these methods. More preferably, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more helicases may be used. If two or more Pif1-like helicases of the present application are used, they may be the same or

different. It may also comprise wild-type Pif1-like helicases or other types of helicases. Further, if two or more helicases are used, they may be attached to one another, or bound to the polynucleotide separately to perform the function of controlling the movement of the polynucleotide.

[00169] Preferably, when a force (such as a voltage) is applied to the pore, speed of the polynucleotide passing through the pore is controlled by the Pif1-like helicase or construct, so as to obtain a recognizable and stable current level for determining the characteristics of the target polynucleotide.

[00170] Preferably, steps I) and II) are repeated one or more times.

[00171] Preferably, the method further comprises the step of applying a potential difference across the pore contacting the helicase or the construct and the target polynucleotide.

[00172] Preferably, the pore is a structure that permits hydrated ions driven by an applied potential to flow from one side of the membrane to the other side of the membrane. More preferably, the pores are nanopores, and the nanopores are transmembrane pores. The transmembrane pores provide channels for the movement of the target polynucleotide.

[00173] The membrane may be any membranes well-known in the art, preferably an amphiphilic layer, i.e., 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. More preferably, the membrane is a lipid bilayer membrane.

[00174] The target polynucleotide may be coupled to the membrane using any known method. If the membrane is an amphiphilic layer, such as a lipid bilayer, the polynucleotide is preferably coupled to the membrane via a polypeptide present in the membrane or a hydrophobic anchor present in the membrane. The hydrophobic anchor is preferably a lipid, fatty acid, sterol, carbon nanotube or amino acid.

[00175] Preferably, the pore is selected from a biological pore, a solid state pore, or a biological and solid state hybrid pore.

[00176] In a specific embodiment of the present application, the pores include, but are not limited to, those derived from *Mycobacterium smegmatis* porin A, *Mycobacterium smegmatis* porin B, *Mycobacterium smegmatis* porin C, and *Mycobacterium smegmatis* porin D, hemolysin, lysin, interleukin, outer membrane porin F, outer membrane porin G, outer membrane phospholipase A, WZA or *Neisseria* autotransporter lipoprotein, etc.

[00177] When provided with all the necessary components to facilitate movement, the Pif1-like helicase moves along the DNA in the 5'-3' direction, but the orientation of the DNA

in the nanopore (dependent on which end of the DNA is captured) means that the enzyme can be used to either move the DNA out of the nanopore against the applied field, or move the DNA into the nanopore with the applied field.

[00178] Preferably, the target polynucleotide is single-stranded, double-stranded, or at least partially double-stranded.

[00179] More preferably, the target polynucleotide may be modified by tags, spacers, methylation, oxidation or damage.

[00180] In a specific embodiment of the present application, at least a part of the target polynucleotide is double-stranded. The double-stranded part constitutes a Y adapter structure comprising a leader sequence that is preferentially screwed into the pore.

[00181] More preferably, the length of the target polynucleotide may be 10 to 100000 or more.

[00182] In a specific embodiment of the present application, the length of the target polynucleotide may be at least 10, at least 50, at least 100, at least 200, at least 300, at least 400, at least 500, at least 1,000, at least 2,000, at least 5,000, at least 10,000, at least 50,000, or at least 100,000, etc.

[00183] Preferably, the helicase is incorporated into an internal nucleotide of a single-stranded polynucleotide.

[00184] Preferably, the one or more characteristics are selected from the source, length, identity, sequence, secondary structure of the target polynucleotide, or whether or not the target polynucleotide is modified.

[00185] Preferably, taking the one or more characteristics are carried out by electrical measurement and/or optical measurement.

[00186] More preferably, electrical and/or optical signals are generated by electrical measurement and/or optical measurement, wherein each nucleotide corresponds to a signal level, and then the electrical signals and/or optical signals are converted into characteristics of the nucleotide.

[00187] In a specific embodiment of the present application, the electrical measurement includes, but is not limited to, a current measurement, an impedance measurement, a tunnelling measurement, a wind tunnel measurement or a field effect transistor (FET) measurement, etc.

[00188] The electrical signal of the present application is selected from the measured values of current, voltage, tunneling, resistance, potential, conductivity or lateral electrical measurement.

[00189] In a specific embodiment of the present application, the electrical signal is a current passing through the pore.

[00190] In the tenth aspect of the present application, provided is a product for characterising a target polynucleotide, comprising the Pif1-like helicase, the construct, the nucleic acid, the expression vector or the host cell of the present application, and a pore.

[00191] Preferably, the product comprises multiple Pif1-like helicases and/or multiple constructs.

[00192] Preferably, the product comprises multiple pores. More preferably, the pores are nanopores, and the nanopores are transmembrane pores.

[00193] In a specific embodiment of the present application, the transmembrane pore is selected from a biological pore, a solid state pore, or a biological and solid state hybrid pore.

[00194] In a specific embodiment of the present application, the pores include, but are not limited to, those derived from Mycobacterium smegmatis porin A, Mycobacterium smegmatis porin B, Mycobacterium smegmatis porin C, and Mycobacterium smegmatis porin D, hemolysin, lysin, interleukin, outer membrane porin F, outer membrane porin G, outer membrane phospholipase A, WZA or Neisseria autotransporter lipoprotein, etc.

[00195] In a specific embodiment of the present application, the product comprises multiple Pif1-like helicases or multiple constructs, and multiple pores.

[00196] Preferably, the product is selected from a kit, a device or a sensor.

[00197] More preferably, the kit also comprises a chip containing a lipid bilayer. The pores span the lipid bilayer.

[00198] The kit of the present application comprises one or more lipid bilayers, and each lipid bilayer comprises one or more of the pores.

[00199] The kit of the present application also comprises reagents or devices for performing the characterising of a target polynucleotide. Preferably, the reagents include buffers and tools required for PCR amplification.

[00200] In the eleventh aspect of the present application, provided is use of the Pif1-like helicase, the construct, the nucleic acid, the expression vector, the host cell or the product of the present application in characterising a target polynucleotide or controlling the movement of a target polynucleotide through a pore.

[00201] In the twelfth aspect of the present application, provided is a kit for characterising a target polynucleotide, comprising the Pif1-like helicase, the construct or the nucleic acid, the expression vector or the host cell of the present application, and pores.

[00202] In the thirteenth aspect of the present application, provided is a device for characterising a target polynucleotide, comprising the Pif1-like helicase, the construct or the nucleic acid, the expression vector or the host cell of the present application, and the pore.

[00203] Preferably, the device comprises a sensor that is capable of supporting the plurality of pores and can transmit signals of the interaction of the pores with the polynucleotide, and at least one reservoir for storing the target polynucleotide, and a solution required to perform the characterising.

[00204] Preferably, the device comprises a plurality of Pif1-like helicases and/or a plurality of constructs, and a plurality of pores.

[00205] In the fourteenth aspect of the present application, provided is a sensor for characterising a target polynucleotide, comprising a complex formed between the pore and the Pif1-like helicase or the construct of the present application.

[00206] Preferably, the pore is brought into contact with the helicase or construct in the presence of the target polynucleotide, and a potential is applied across the pore. The potential is selected from voltage potential or chemical potential.

[00207] In the fifteenth aspect of the present application, provided is a method for forming a sensor for characterising a target polynucleotide, comprising forming a complex between the pore and the Pif1-like helicase or the construct of the present application, thereby forming a sensor that characterises the target polynucleotide.

[00208] In the sixteenth aspect of the present application, provided is a structure of two or more helicases attached to a polynucleotide, wherein at least one of the two or more helicases is a Pif1-like helicase of the present application.

[00209] In the seventeenth aspect of the present application, provided is a Pif1-like helicase oligomer. The Pif1-like helicase oligomer comprises one or more Pif1-like helicases of the present.

[00210] Preferably, the Pif1-like helicase oligomer may also comprise wild-type Pif1-like helicase or other types of helicase, wherein the other types of helicase may be Hel308 helicase, XPD helicase, Dda helicase, TraI helicase or TrwC helicase, etc.

[00211] Preferably, the Pif1-like helicase and the wild-type Pif1-like helicase, the Pif1-like helicase and the Pif1-like helicase, the wild-type Pif1-like helicase and the wild-type Pif1-like helicase, the Pif1-like helicase and other types of helicases, or the wild-type Pif1-like helicase and other types of helicases, may be connected or arranged in a head-to-head, tail-to-tail or head-to-tail manner.

[00212] Preferably, the Pif1-like helicase oligomer comprises more than two Pif1-like helicases of the present application, wherein the Pif1-like helicases may be different or the same.

[00213] The “Pif1-like helicase” of the present application is modified relative to the wild-type or natural helicase.

[00214] The “Pif1-like helicase”, “construct” or “pore” of the present application may be modified to assist their identification or purification, for example by the addition of histidine residues (a his tag), aspartic acid residues (an asp tag), a streptavidin tag, a Flag tag, a SUMO tag, a GST tag or a MBP tag, or by the addition of a signal sequence to promote their secretion from a cell where the polypeptide does not naturally contain such a sequence. An alternative to introducing a genetic tag is to chemically react a tag onto a native or engineered position on the helicase, pore or construct.

[00215] The “nucleotides” in the present application include, but are not limited to, adenosine monophosphate (AMP), guanosine monophosphate (GMP), thymidine monophosphate (TMP), uridine monophosphate (UMP), cytidine monophosphate (CMP), cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), deoxyadenosine monophosphate (dAMP), deoxyguanosine monophosphate (dGMP), deoxythymidine monophosphate (dTMP), deoxyuridine monophosphate (dUMP) and deoxycytidine monophosphate (dCMP). The nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP, dCMP and dUMP.

[00216] The “conservative amino acid substitutions” in the present application include, but are not limited to, a substitution between alanine and serine, glycine, threonine, valine, proline or glutamic acid; and/or a substitution between aspartame and glycine, asparagine or glutamic acid; and/or a substitution between serine and glycine, asparagine or threonine; and/or a substitution between leucine and isoleucine or valine; and/or a substitution between valine and leucine, isoleucine; and/or a substitution between tyrosine and phenylalanine; and/or, a substitution between lysine and arginine. The above-mentioned substitution basically does not change the activity of the amino acid sequence of the present application.

[00217] The term “two or more” in the present application includes two, three, four, five, six, seven, eight or more and so on.

[00218] The term “plurality” or “multiple” in the present application includes but is not limited to two or more, three or more, four or more, five or more, six or more, seven or more, eight or more or more, and so on.

[00219] The term “at least one” in the present application includes but is not limited to one or more, two or more, three or more, four or more, five or more, six or more, seven or more, eight or more or more, and so on.

[00220] The term “and/or” in the present application includes the selection of one and any number of combinations of the listed items.

[00221] The term “comprise/comprising” in the present application is an open-ended term, comprising the specified ingredients or steps described, as well as other ingredients or steps that have no substantial effect.

[00222] The term “non-natural amino acid” in the present application is an amino acid that is not naturally present in Pif1-like helicase. Preferably, it includes but is not limited to, 4-Azido-L-phenylalanine (Faz), 4-Acetyl-L-phenylalanine, 3-Acetyl-L-phenylalanine, 4-Acetoacetyl-L-phenylalanine, O-Allyl-L-tyrosine, 3-(Phenylselanyl)-L-alanine, O-2-Propyn-1-yl-L-tyrosine, 4-(Dihydroxyboryl)-L-phenylalanine, 4-[(Ethylsulfanyl)carbonyl]-L-phenylalanine, (2*S*)-2-amino-3-{4-[(propan-2-ylsulfanyl)carbonyl]phenyl}propanoic acid, (2*S*)-2-amino-3-{4-[(2-amino-3-sulfanylpropanoyl)amino]phenyl}propanoic acid, O-Methyl-L-tyrosine, 4-Amino-L-phenylalanine, 4-Cyano-L-phenylalanine, 3-Cyano-L-phenylalanine, 4-Fluoro-L-phenylalanine, 4-Iodo-L-phenylalanine, 4-Bromo-L-phenylalanine, O-(Trifluoromethyl)tyrosine, 4-Nitro-L-phenylalanine, 3-Hydroxy-L-tyrosine, 3-Amino-L-tyrosine, 3-Iodo-L-tyrosine, 4-Isopropyl-L-phenylalanine, 3-(2-Naphthyl)-L-alanine, 4-Phenyl-L-phenylalanine, (2*S*)-2-amino-3-(naphthalen-2-ylamino)propanoic acid, 6-(Methylsulfanyl)norleucine, 6-Oxo-L-lysine, D-tyrosine, (2*R*)-2-Hydroxy-3-(4-hydroxyphenyl)propanoic acid, (2*R*)-2-Ammoniooctanoate3-(2,2'-Bipyridin-5-yl)-D-alanine, 2-amino-3-(8-hydroxy-3-quinolyl)propanoic acid, 4-Benzoyl-L-phenylalanine, S-(2-Nitrobenzyl)cysteine, (2*R*)-2-amino-3-[(2-nitrobenzyl)sulfanyl]propanoic acid, (2*S*)-2-amino-3-[(2-nitrobenzyl)oxy]propanoic acid, O-(4,5-Dimethoxy-2-nitrobenzyl)-L-serine, (2*S*)-2-amino-6-({[(2-nitrobenzyl)oxy]carbonyl}amino)hexanoic acid, O-(2-Nitrobenzyl)-L-tyrosine, 2-Nitrophenylalanine, 4-[(*E*)-Phenyldiazenyl]-L-phenylalanine, 4-[3-(Trifluoromethyl)-3H-diaziren-3-yl]-D-phenylalanine, 2-amino-3-[[5-(dimethylamino)-1-naphthyl]sulfonylamino]propanoic acid, (2*S*)-2-amino-4-(7-hydroxy-2-oxo-2H-chromen-4-yl)butanoic acid, (2*S*)-3-[(6-acetylnaphthalen-2-yl)amino]-2-aminopropanoic acid, 4-(Carboxymethyl)phenylalanine, 3-Nitro-L-tyrosine, O-Sulfo-L-tyrosine, (2*R*)-6-Acetamido-2-ammoniohexanoate, 1-Methylhistidine, 2-Aminononanoic acid, 2-Aminodecanoic acid, L-Homocysteine, 5-Sulfanylnorvaline, 6-Sulfanyl-L-norleucine, 5-(Methyl sulfanyl)-L-norvaline, N⁶-{[(2*R*,3*R*)-3-Methyl-3,4-dihydro-2H-pyrrol-2-yl]carbonyl}-L-lysine, N⁶-

[(Benzyloxy)carbonyl]lysine, (2*S*)-2-amino-6-[(cyclopentylcarbonyl)amino]hexanoic acid, N⁶-[(Cyclopentylloxy)carbonyl]-L-lysine, (2*S*)-2-amino-6-{[(2*R*)-tetrahydrofuran-2-ylcarbonyl]amino}hexanoic acid, (2*S*)-2-amino-8-[(2*R*,3*S*)-3-ethynyltetrahydrofuran-2-yl]-8-oxooctanoic acid, N⁶-(tert-Butoxycarbonyl)-L-lysine, (2*S*)-2-Hydroxy-6-({[(2-methyl-2-propanyl)oxy]carbonyl}amino)hexanoic acid, N⁶-[(Allyloxy)carbonyl]lysine, (2*S*)-2-amino-6-({[(2-azidobenzyl)oxy]carbonyl}amino)hexanoic acid, N⁶-L-Prolyl-L-lysine, (2*S*)-2-amino-6-({[(prop-2-yn-1-yloxy)carbonyl]amino}hexanoic acid and N⁶-(2-Azidoethoxy)carbonyl]-L-lysine.

BRIEF DESCRIPTION OF THE DRAWINGS

[00223] Hereinafter, the embodiments of the present application will be described in detail with reference to the accompanying drawings, in which:

[00224] Figure 1 shows a schematic diagram of fluorescence assay for testing enzyme activity of Pif1-like helicase, wherein the fluorescent substrate strand has a 5' ssDNA overhang, and a 50 base section of hybridized dsDNA. As shown in a), the major upper strand (B) has a black-hole quencher (BHQ-1) base (E) at the 3' end, and the hybridized complement (D) has a carboxyfluorescein base (C) at the 5' end. 0.5 μ M of a capture strand (F) that is complementary to the shorter strand of the fluorescent substrate (D) is also included. As shown in b), in the presence of ATP (5 mM) and MgCl₂ (5 mM), helicase (200 nM) added to the substrate binds to the 5' part of the fluorescent substrate, moves along the major strand, and unwinds the complementary strand. After that, excess of capture strand preferentially anneals to the complementary DNA to prevent re-annealing of initial substrate and loss of fluorescence. As shown in c), after adding the excess of the capture strand (A) that is completely complementary to the major strand, part of the entangled dsDNA will have a strand unwinding effect due to the presence of excessive A, and finally all the dsDNA will be untwisted, and the fluorescence value will reach the highest.

[00225] Figure 2 shows a measurement of the hybridized dsDNA unwinding capabilities of Pif1-like helicase using a fluorescence assay, specifically a graph of the time-dependent dsDNA unwinding ratio in a buffer containing 400mM NaCl.

[00226] Figure 3 shows gel measurements of the DNA-binding capabilities of different PIF1-like helicases. Lane 1 represents the pre-constructed dsDNA (hybridization of SEQ ID NO: 18 with SEQ ID NO: 19 modified with 5'FAM). Lanes 2-6 comprise Aph Acj61, Aph PX29, Sph CBH8, PphPspYZU05, and Mph MP1 pre-linked to dsDNA, respectively. Lanes 7-

11 comprise Aph Acj61-D97C/A363C, Aph PX29-D96C/A371C, Sph CBH8-A94C/A361C/C136A, PphPspYZU05-D104C/A375C/C146A and Mph MP1-E105C/A362C pre-linked to dsDNA, respectively. Band A corresponds to SEQ ID NO:19 to be hybridized with SEQ ID NO:18. The regions labeled 1A, 2A, 3A, 4A, and 5A correspond to combination of 1, 2, 3, 4, and 5 helicases with hybridization of SEQ ID NO: 18 with SEQ ID NO: 19, respectively.

[00227] Figure 4 shows DNA construct which is used in examples, in which SEQ ID NO: 13 (with the nucleotide sequence “tttttt”, labelled B) is attached at its 5' end to twenty iSpC3 spacers (labelled A) and at its 3' end to four iSpC3 spacers (labelled C) which are attached to the 5' end of SEQ ID NO: 14 (labelled D) which is attached at its 3' end to SEQ ID NO: 17 or 24 (labelled E), and the SEQ ID NO: 15 region (labelled F) of this construct is hybridized to SEQ ID NO: 16 (labelled G, which has a 5' cholesterol tether).

[00228] Figure 5 shows an illustrative current trace (y-axis = Current (pA, 0 to 250), x-axis = Sampling frequency (hz, 0 to 3.5×10^5)) when Pif1-like helicase (Mph MP1-E105C/A362C (SEQ ID NO: 11 with mutations E105C/A362C)) controlled the translocation of DNA construct A through the nanopore 8MspA (SEQ ID NO: 12).

[00229] Figure 6 shows an enlarged current trace in regions of the Pif1-like helicase-controlled DNA movement shown in Figure 5 (y-axis = Current (pA, 30 to 100, x-axis = Sampling frequency (hz, 2.346 to 2.366×10^5)).

[00230] Figure 7 shows an illustrative current trace (y-axis = Current (pA), x-axis = Time (s)) when Pif1-like helicase (Sph CBH8-A94C/A361C/C136A (SEQ ID NO: 5 with A94C/A361C/C136A mutation)) controlled the translocation of DNA construct B through the MspA nanopore.

[00231] Figure 8 shows an illustrative current trace (y-axis = Current (pA), x-axis = Time (s)) when helicase (Eph Pei26-D99C/A366C/C141A (SEQ ID NO: 6 with D99C/A366C/C141A mutation)) controlled the translocation of DNA construct B through the MspA nanopore.

[00232] Figure 9 shows an illustrative current trace (y-axis = Current (pA), x-axis = Time (s)) when helicase (Pph PspYZU05-D104C/A375C/C146A (SEQ ID NO: 8 with D104C/A375C and C146A mutations)) controlled the translocation of DNA construct B through the MspA nanopore.

[00233] Figure 10 shows the SDS-PAGE gel electrophoresis image of purified Mph MP1 (SEQ ID NO: 11), in which M represents Marker (Kd), and lane 1 represents the electrophoresis result of the helicase Mph MP1.

[00234] Figures 11 A and 11B show the correspondence among the amino acid sequences of SEQ NOs: 1 to 11.

[00235] Figure 12 shows the speed distribution of a nucleic acid library passing through a nanopore controlled by the mutants Eph-Pei26-D99C/A366C/F103W/E286K (A) and Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/ F103W/E286K (B).

[00236] Figure 13 shows movements of a nucleic acid through a nanopore controlled by the mutants Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D (SEQ ID NO: 6 with mutation D99C/A366C/C114V/C119V/C141S/C308S/C419D) and Eph-Pei26-D99C/A366C/C114V /C119V/C141S/C308S/C419D/F103W (SEQ ID NO: 6 with mutation D99C/A366C/C114V/ C119V/C141S/C308S/C419D/F103W). Figure A is a simulated current signal characteristic of the nucleic acid sequence SEQ ID NO: 24 moving through the nanopore one by one; Figure B is a current signal generated by the movement of the nucleic acid sequence SEQ ID NO: 24 through the nanopore controlled by the mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W; and Figure C is a current signal generated by the movement of the nucleic acid sequence SEQ ID NO: 24 through the nanopore controlled by the mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D.

[00237] Figure 14 shows an accuracy rate of the validation set of randomly digested *E.coli* and human genomes library sequenced by using mutants Eph-Pei26-24 and Eph-Pei26-25.

DETAILED DESCRIPTION

[00238] The following examples further illustrate the content of the present application, but should not be construed as limiting the present application. Without departing from the spirit and essence of the present application, modifications or substitutions made to the methods, steps or conditions of the present application fall within the scope of claims.

[00239] Unless otherwise specified, the technical means used in the examples are conventional means well known to those skilled in the art. The equipment and reagents used in each example are conventionally commercially available. The sequencer used for sequencing is the QNome-9604 gene sequencer, and the sequencing algorithm is the open-source sequencing model bonito-etc of ONT (Oxford Nanopore Technologies), which may be found at <https://github.com/nanoporetech/bonito>. This is an open source CTC-based end-to-end seq2seq model, based on NVIDIA's open source speech recognition network QuartzNet that is described in paper titled "QUARTZNET: DEEP AUTOMATIC SPEECH RECOGNITION WITH 1D TIME-CHANNEL SEPARABLE CONVOLUTIONS".

[00240] **Example 1. Preparation of Pif1-like helicase**

[00241] 1. Materials and methods

[00242] A recombinant plasmid containing the Pif1-like helicase sequence (a variant of amino acid sequence SEQ ID NOs:1-11, corresponding to nucleotide sequence SEQ ID NOs:25-35) was transformed into BL21(DE3) competent cell by heat shock. The resuscitated bacteria solution was spread on an ampicillin-resistant solid LB plate, and then cultured overnight at 37°C. A single colony was picked and inoculated into 100ml of ampicillin-resistance liquid LB medium and cultivated at 37°C. 1% of the inoculum was transferred to ampicillin-resistant LB liquid medium for expansion culture at 37°C and 200 rpm, and its OD600 value was measured continuously. When OD600=0.6-0.8, the culture solution in LB medium was cooled to 18°C, and isopropyl thiogalactoside (Isopropyl β -D-Thiogalactoside, IPTG) was added to induce expression, so that the final concentration reached 1 mM. After 12-16h, the bacteria were collected at 18°C. The bacteria were crushed under high pressure, purified by FPLC, and samples were collected.

[00243] 2. Results

[00244] Figure 10 shows an SDS-PAGE gel electrophoresis image of purified Mph MP1 (variant of SEQ ID NO: 11).

[00245] **Example 2**

[00246] This Example illustrates the capabilities of Pif1-like helicase to unwind hybridized dsDNA by using a fluorescence assay to detect enzyme activity.

[00247] 1. Materials and methods

[00248] In Figure 1, as shown in a), the fluorescent substrate strand (100 nM final) has a 5' ssDNA overhang, and a 50 base section of hybridized dsDNA. The major upper strand has a black-hole quencher (BHQ-1) base (SEQ ID NO:20--BHQ-3') at the 3' end, and the hybridized complement has a carboxyfluorescein base (5'FAM-SEQ ID NO:21) at the 5' end. When the fluorescence from the fluorescein is quenched by the local BHQ-1, and the substrate is essentially non-fluorescent. 0.5 μ M of a capture strand (SEQ ID NO:22) that is complementary to the shorter strand of the fluorescent substrate is also included. As shown in b), in the presence of ATP (5 mM) and MgCl₂ (5 mM), helicase (200 nM) added to the substrate binds to the 5' part of the fluorescent substrate, moves along the major strand, and unwinds the complementary strand. After that, excess of capture strand preferentially anneals to the complementary DNA to prevent re-annealing of initial substrate and loss of fluorescence. A certain amount of hybrid dsDNA that is not unwound by Pif1-like helicase is present in the system. As shown in c), after adding excessive capture strand A (SEQ ID NO:23) that is completely complementary to the major strand, part of the entangled dsDNA will have a strand

unwinding effect due to the presence of excessive A, and finally all the dsDNA will be untwisted, and the fluorescence value will reach the highest.

[00249] 2. Results

[00250] Figure 2 shows a graph of the time-dependent dsDNA unwinding ratio in a buffer containing 400mM NaCl (10 mM Hepes pH 8.0, 5 mM ATP, 5 mM MgCl₂, 100 nM fluorescent substrate DNA, 0.5 μM capture DNA).

[00251] **Example 3**

[00252] This example shows an exemplary DNA binding capabilities of modified Pif1-like helicase of the present application by using gel measurements.

[00253] Specifically, DNA binding capabilities of Aph Acj61-D97C/A363C (SEQ ID NO: 2 with D97C/A363C mutation), Aph PX29-D96C/A371C (SEQ ID NO: 3 with D96C/A371C mutation), Sph CBH8-A94C/A361C/C136A (SEQ ID NO: 5 with A94C/A361C and C136A mutations) and Pph PspYZU05-D104C/A375C/C146A (SEQ ID NO: 8 with D104C/A375C and C146A mutations) and Mph MP1-E105C/A362C (SEQ ID NO: 11 with E105C/A362C mutant) were measured.

[00254] 1. Materials and methods

[00255] The annealed DNA complex (hybridization of SEQ ID NO: 18 with SEQ ID NO: 19 modified with 5'FAM) were mixed with Aph Acj61, Aph PX29, Sph CBH8, Pph PspYZU05, Mph MP1, Aph Acj61-D97C/A363C, Aph PX29-D96C/A371C, Sph CBH8-A94C/A361C/C136A, Pph PspYZU05-D104C/A375C/A375C/MPC/A375C E105C/A362C, respectively, in a ratio of (1:1, volume/volume) in 10mM HEPE, pH 8.0, and 400mM potassium chloride to reach a final concentration of Pif1-like helicase of 600nM and a final concentration of DNA of 30nM. The Pif1-like helicase was allowed to bind to DNA for 1 hour at room temperature. To each sample TMAD was added to a final concentration of 5 μM, and incubated at room temperature for 1 hour. The sample was loaded on a 4-20% TBE gel and run the gel at 160V for 1.5 hours. The gel was placed under blue fluorescence to observe the DNA bands.

[00256] 2. Results

[00257] Figure 3 shows the effect of modifications on the DNA-binding capabilities of Pif1-like helicases. Lanes 2 to 6 show that a large amount of DNA was not bound by the Pif1-like helicase during electrophoresis and no obvious binding bands are observed. Lanes 7-11 show the binding bands of different numbers of enzymes bound to DNA, and lanes 9 and 10 show that up to 5 Pif1-like helicases could bind to the single-stranded moiety of SEQ ID

NO:18. This indicates that the modified Pif1-like helicase significantly enhanced its binding strength to DNA.

[00258] Example 4

[00259] This example describes how a Mph MP1-E105C/A362C (SEQ ID NO: 11 with mutation E105C/A362C) controlled the movement of intact DNA strands through a single MspA nanopore (SEQ ID NO: 12).

[00260] 1. Materials and methods

[00261] DNA construct A as shown in Figure 4 was prepared. SEQ ID NO: 13 (labelled B) was attached at its 5' end to twenty iSpC3 spacers (labelled A) and at its 3' end to four iSpC3 spacers (labelled C) which were attached to the 5' end of SEQ ID NO: 14 (labelled D) which was attached at its 3' end to SEQ ID NO: 17 (labelled E), and the SEQ ID NO: 15 region (labelled F) of this construct was hybridized to SEQ ID NO: 16 (labelled G, having a 5' cholesterol tether).

[00262] The prepared DNA construct and Mph MP1-E105C/A362C were pre-incubated together for 30 minutes at 25°C in a buffer (10 mM HEPES, pH 8.0, 400 mM NaCl, 5% glycerol, 2 mM DTT).

[00263] Electrical measurements were acquired from MspA nanopores inserted in 1,2-diphytanoyl-glycero-3-phosphocholine lipid bilayers. Bilayers were formed across about 25 µm diameter apertures on a PTFE film via the Montal-Mueller technique, separating two 100 µL buffered solutions. All experiments were carried out in the stated buffered solution. Single-channel currents were measured on an amplifier equipped with digitizers. Ag/AgCl electrodes were connected to the buffered solutions so that the *cis* compartment is connected to the ground of the amplifier, and the *trans* compartment is connected to the active electrode of the amplifier.

[00264] After achieving a single pore in the bilayer, DNA polynucleotide and the Pif1-like helicase were added to 70 µL of buffer in the *cis* compartment of the electrophysiology chamber to initiate capture of the helicase-DNA complexes in the nanopore. Helicase ATPase activity was initiated as required by the addition of divalent metal (5 mM MgCl₂) and NTP (2.86µM ATP) to the *cis* compartment. Experiments were carried out at a constant potential of +180 mV.

[00265] Results and Discussion

[00266] DNA movement controlled by the Pif1-like helicase was observed for the DNA construct. Observation of Pif1-like helicase-controlled DNA movement is shown in Figure 5. The Pif1-like helicase-controlled DNA movement was 50 seconds long and corresponded to approximately 10 kB of the DNA construct's translocation through the nanopore. Figure 6

shows an enlarged graph of a partial region of the Pif1-like helicase-controlled DNA movement.

[00267] Example 5

[00268] This example describes how Sph CBH8-A94C/A361C/C136A (SEQ ID NO: 5 with A94C/A361C/C136A mutation), Eph Pei26-D99C/A366C/C141A (SEQ ID NO: 6 with D99C/A366C/C141A mutation) and Pph PspYZU05-D104C/A375C/C146A (SEQ ID NO: 8 with D104C/A375C and C146A mutations) controlled the movement of intact DNA strands through a single MspA nanopore.

[00269] 1. Materials and methods

[00270] DNA construct B as shown in Figure 4 was prepared. SEQ ID NO: 13 was attached at its 5' end to twenty iSpC3 spacers and at its 3' end to four iSpC3 spacers which were attached to the 5' end of SEQ ID NO: 14 which was attached at its 3' end to SEQ ID NO: 24, and the SEQ ID NO: 15 region of this construct was hybridized to SEQ ID NO: 16 (having a 5' cholesterol tether). This DNA construct B was similar to the construct used in Example 4, except that the region labeled E corresponded to SEQ ID NO: 24.

[00271] DNA construct B in a buffer (in 50 mM NaCl, 10 mM Tris pH 7.5) and Sph CBH8-A94C/A361C/C136A, Eph Pei26-D99C/A366C/C141A or Pph PspYZU05-D104C/A375C/C146A in a buffer (50 mM KCl, 10 mM HEPES, pH 8.0) were pre-incubated together for 30 minutes at room temperature. TMAD was then added to the DNA/enzyme pre-mix and incubated for a further 30 minutes. Finally, a buffer (10 mM HEPES, 600 mM KCl, pH 8.0, 3 mM MgCl₂) and ATP were added to the pre-mix.

[00272] Electrical measurements were acquired at room temperature from single MspA nanopore inserted in block co-polymer in a buffer (10 mM HEPES, 400 mM KCl, pH 8.0). After achieving a single pore inserted in the block co-polymer, the pre-mix of the Pif1-like helicase (Sph CBH8-A94C/A361C/C136A, Eph Pei26-D99C/A366C/C141A or Pph PspYZU05-D104C/A375C/C146A (1nM final)), DNA (0.3nM final), fuel (ATP, 3mM final) was then added to the single nanopore experimental system. Each experiment was carried out for 2 hours at a holding potential of 180mV and helicase-controlled DNA movement was monitored.

[00273] 2. Results

[00274] DNA movement controlled by Pif1-like helicase was observed for the DNA construct B. Observations of Sph CBH8-A94C/A361C/C136A, Eph Pei26-

D99C/A366C/C141A or Pph PspYZU05-D104C/A375C/C146A-controlled DNA movements are shown in Figures 7-9, respectively.

[00275] Example 6

[00276] In this example Eph-Pei26-D99C/A366C/F103W/E286K (SEQ ID NO: 6 with D99C/A366C/F103W/E286K mutations) and Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K (SEQ ID NO: 6 with D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K mutation) are taken as an example, to illustrate that by replacing or deleting certain types of natural amino acids of helicase, non-specific modifications of amino acids at one or more sites of this type can be avoided during chemical modification or treatment of helicase, the heterogeneity among enzymes can be reduced, and the speed uniformity of different enzyme molecules controlling different nucleic acid molecules passing through the MspA nanopore (SEQ ID NO: 12) can be improved.

[00277] 1. Materials and methods

[00278] DNA construct C as shown in Figure 4 was prepared. SEQ ID NO: 13 (labelled B) was attached at its 5' end to twenty iSpC3 spacers (labelled A) and at its 3' end to four iSpC3 spacers (labelled C) which were attached to the 5' end of SEQ ID NO: 14 (labelled D) which was attached at its 3' end to SEQ ID NO: 17 (labelled E), and the SEQ ID NO: 15 region (labelled F) of this construct was hybridized to SEQ ID NO: 16 (labelled G, having a 5' cholesterol tether).

[00279] A ligated segment was synthesized from the A, B, C, and D segments at a concentration of 10 μ M. The ligated segment, and the F segment and G segment in a ratio of 1:1:1 were added to an annealing buffer (10mM Tris, pH7.0, 50mM NaCl). The annealing process was carried out in accordance with 98°C 10min, -0.1°C/0.6s, 300 cycles, 65°C 5min, -0.1°C/0.6s, 400 cycles (wherein segments A, B, C, D, F, and G were provided by Sangon Biotech).

[00280] The prepared DNA construct C and Eph-Pei26-D99C/A366C/F103W/E286K or Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K were pre-incubated together for 30 minutes at 25°C in a buffer (10mM HEPES, 50mM KCl, pH 7.0). Then, TMAD with a final concentration of 1 mM was added thereto for catalytic treatment for 30 minutes. Finally, the DNA construct/enzyme mixture purified by magnetic beads and end-repaired nucleic acid library SEQ ID NO: 17 with a fixed-length of 10kb (E region) were added to a T4 rapid ligation reaction system (as shown in Table 8 below, the T4 rapid ligation kit was

provided by Enzymatics) at a molar concentration ratio of 1.5:1, the rapid TA ligation was performed at room temperature for 10 minutes. The ligation product was purified by magnetic beads, and then put into the subsequent sequencing system.

[00281] Table 8. T4 rapid ligation reaction system

Reaction component	Final concentration
2x rapid ligation Buffer	1x
DNA construct/enzyme complex	10uM
End-repaired nucleic acid library SEQ ID NO: 17 with a fixed-length of 10kb	10uM
Nuclease-free water	N/A
T4 DNA Ligase (600U/ul)	30U/ul
Total volume	20ul

[00282] Electrical measurements were acquired from MspA nanopores inserted in 1,2-diphytanoyl-glycero-3-phosphocholine lipid bilayers. Bilayers were formed across about 25 μm diameter apertures on a PTFE film via the Montal-Mueller technique, separating two 100 μL buffered solutions. All experiments were carried out in the stated buffered solution. Single-channel currents were measured on an amplifier equipped with digitizers. Ag/AgCl electrodes were connected to the buffered solutions so that the *cis* compartment was connected to the ground of the amplifier, and the *trans* compartment was connected to the active electrode of the amplifier. After achieving a single pore in the bilayer, the incubated DNA/enzyme mixture was added to 70 μL of buffer (10mM HEPES, 500mM KCl, 50mM MgCl_2 , 50mM ATP, pH 7.0) in the *cis* compartment of the electrophysiology chamber and a constant potential of +180 mV was applied at 35°C to initiate capture of the DNA construct/enzyme complex in the MspA nanopore (SEQ ID NO: 12) and the movement of nucleic acid through the nanopore controlled by the helicase, and the DNA movement controlled by helicase was monitored and recorded.

[00283] 2. Results

[00284] Certain types of native amino acids in the helicase sequence have an effect on the uniformity of the speeds at which the helicase controls the movement of nucleic acids through a nanopore. These types of native amino acids are present in or introduced into a nucleic acid binding pocket of the helicase, which was shrunk after chemical or biological modification. In this example, in order to identify and analyze the improvement of the speed uniformity of nucleic acid moving through nanopores controlled by the helicase by replacing or deleting these types of native amino acids in the helicase, the time of each original current signal obtained by sequencing was extracted, and the speed of each original current signal was calculated according to the base length of the fixed-length library. Current signals indicating

passing nucleic acid library through the nanopore controlled by Eph-Pei26-D99C/A366C/F103W/E286K and Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K were identified and processed, the speeds of complete passage of the different nucleic acid libraries through the nanopore was calculated and statistically analyzed.

[00285] As shown in Figure 12, it can be seen from figure 12A that the speed at which the mutant Eph-Pei26-D99C/A366C/F103W/E286K controlled nucleic acid passing through the nanopore shows multimodal distribution, while it can be seen from figure 12B that the speed at which the mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K controlled the nucleic acid passing through the nanopore shows unimodal distribution. In mutants Eph-Pei26-D99C/A366C/F103W/E286K and Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K, the mutation sites D99C/A366C are cysteine mutations introduced in domain 1B and domain 2B related to the nucleic acid binding pocket of Eph-Pei26 helicase, respectively. The mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W/E286K is a mutant with C114V/C119V/C141S/C308S/C419D mutations in the cysteine (C) at positions other than the cysteines introduced into the 1B domain and 2B domain of the Eph-Pei26 helicase. C114V/C119V/C141S/C308S/C419D mutation may also avoid the introduction of modifications at other cysteine positions such as C114, C119, C141, C308 or C419 during the formation of a disulfide bond between D99C/A366C catalyzed by TMAD, which would otherwise affect the function of the enzyme and cause the heterogeneity among enzyme molecules, leading to a multimodal distribution of the rate. The multimodal distribution of speed makes it more difficult for the algorithm to recognize signals or is not conducive to the identification of bases. By replacing some types of native amino acids in the helicase sequence, the uniformity of speed of helicase can be effectively improved, thus improving the throughput and accuracy of effective sequencing data.

[00286] Example 7

[00287] This example provides a comparative analysis of the current signal generated by the movement of a nucleic acid through the MspA nanopore (SEQ ID NO: 12) controlled by Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D (SEQ ID NO: 6 with D99C/A366C/C114V/C119V/C141S/C308S/C419D mutation) and Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W (SEQ ID NO: 6 with

D99C/A366C/ C114V/C119V/C141S/C308S/C419D/F103W mutations) respectively, to illustrate the influence of the amino acid side chain interacting with the phosphate backbone or sugar ring or base of the nucleic acid on the current signal, which in turn affects the accuracy of sequencing.

[00288] 1. Materials and methods

[00289] DNA construct D as shown in Figure 4 was prepared. SEQ ID NO: 13 (labelled B) was attached at its 5' end to twenty iSpC3 spacers (labelled A) and at its 3' end to four iSpC3 spacers (labelled C) which were attached to the 5' end of SEQ ID NO: 14 (labelled D) which was attached at its 3' end to SEQ ID NO: 17 (labelled E), and the SEQ ID NO: 15 region (labelled F) of this construct was hybridized to SEQ ID NO: 16 (labelled G, having a 5' cholesterol tether).

[00290] A ligated segment were synthesized from the A, B, C, and D segments at a concentration of 10 μ M. The ligated segment, and the F segment and G segment in a ratio of 1:1:1 were added to an annealing buffer (10mM Tris, pH7.0, 50mM NaCl). The annealing process was carried out in accordance with 98°C 10min, -0.1°C/0.6s, 300 cycles, 65°C 5min, -0.1°C/0.6s, 400 cycles (wherein segments A, B, C, D, F, and G were provided by Sangon Biotech).

[00291] The prepared DNA construct D and Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D or Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W were pre-incubated together for 30 minutes at 25°C in a buffer (10mM HEPES, 50mM KCl, pH 7.0). Then, TMAD with a final concentration of 1 mM was added thereto for catalytic treatment for 30 minutes. Finally, the DNA construct/enzyme mixture purified by magnetic beads and a double-stranded DNA target sequence formed by annealing of SEQ ID NO: 24 and its complementary sequence (E region) were added to the T4 rapid ligation reaction system (as shown in Table 9 below, the T4 rapid ligation kit iwas provided by Enzymatics) at a molar concentration ratio of 1.5:1, and the rapid ligation was performed at room temperature for 10 minutes. The ligation product was purified by magnetic beads, and then put into the subsequent sequencing system.

[00292] Table 9. T4 rapid ligation reaction system

Reaction component	Final concentration
2x rapid ligation Buffer	1x
DNA construct/enzyme complex	10uM
dsDNA formed by annealing of SEQ ID NO: 24 and its complementary sequence	10uM

Nuclease-free water	N/A
T4 DNA Ligase (600U/ul)	30U/ul
Total volume	20ul

[00293] Electrical measurements were acquired from MspA nanopores inserted in 1,2-diphytanoyl-glycero-3-phosphocholine lipid bilayers. Bilayers were formed across about 25 μm diameter apertures on a PTFE film via the Montal-Mueller technique, separating two 100 μL buffered solutions. All experiments were carried out in the stated buffered solution. Single-channel currents were measured on an amplifier equipped with digitizers. Ag/AgCl electrodes were connected to the buffered solutions so that the *cis* compartment was connected to the ground of the amplifier, and the *trans* compartment was connected to the active electrode of the amplifier. After achieving a single pore in the bilayer, the incubated DNA/enzyme mixture was added to 70 μL of buffer (10mM HEPES, 600mM KCl, 3mM MgCl_2 , 3mM ATP, pH 7.0) in the *cis* compartment of the electrophysiology chamber and a constant potential of +180 mV was applied at 30°C to initiate capture of the DNA construct/enzyme complex in the MspA nanopore (SEQ ID NO: 12) and the movement of nucleic acid through the nanopore controlled by the helicase, and the DNA movement controlled by helicase was monitored and recorded.

[00294] 2. Results

[00295] DNA movement controlled by the Eph-Pei26 helicase mutant was observed for the DNA construct. Figure 13 shows the current signal characteristics of the target nucleic acid sequence SEQ ID NO: 24 passing through the nanopore controlled by the Eph-Pei26 helicase mutant. Figure A is a simulated current signal characteristic of the target nucleic acid sequence SEQ ID NO: 24 moving through the nanopore one by one; Figure B is an actual current signal generated by the movement of the target nucleic acid sequence SEQ ID NO: 24 through the nanopore controlled by the mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W; and Figure C is an actual current signal generated by the movement of the target nucleic acid sequence SEQ ID NO: 24 through the nanopore controlled by the mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D. It can be seen from figure 13, there are more current step signals (labelled with arrows in Figure B) when the mutant Eph-Pei26-D99C/A366C/C114V/C119V/C141S/C308S/C419D/F103W is used to control the target nucleic acid SEQ ID NO: 24 to pass through the nanopore, resulting in an increase in errors of insertion. The results suggest that the accuracy of sequencing is affected by the amino acid side chain that interacts with the nucleic acid base sequence or the amino acid side chain that interacts with the nucleic acid base at a specific domain site.

[00296] Example 8

[00297] This example uses different Eph-Pei26 helicase mutants to control the movement of DNA constructs through a nanopore. The tested helicases have a substitution of at least one amino acid that interacts with one or more nucleotides of a template single-stranded DNA or complementary single-stranded DNA. By analyzing the different parameters of different Eph-Pei26 helicase mutants that control the movement of nucleic acid through the MspA nanopore (SEQ ID NO: 12), this example shows the effects of the substitution of amino acids in helicases that interact with one or more nucleotides of a template or complementary single-stranded DNA on the speed and accuracy of nucleic acid sequencing and provides several mutants that improve the accuracy and/or speed of sequencing.

[00298] 1. Materials and methods

[00299] DNA construct E as shown in Figure 4 was prepared. SEQ ID NO: 13 (labelled B) was attached at its 5' end to twenty iSpC3 spacers (labelled A) and at its 3' end to four iSpC3 spacers (labelled C) which were attached to the 5' end of SEQ ID NO: 14 (labelled D) which was attached at its 3' end to SEQ ID NO: 17 (labelled E), and the SEQ ID NO: 15 region (labelled F) of this construct was hybridized to SEQ ID NO: 16 (labelled G, having a 5' cholesterol tether).

[00300] A ligated segment were synthesized from the A, B, C, and D segments at a concentration of 10 μ M. The ligated segment, and the F segment and G segment in a ratio of 1:1:1 were added to an annealing buffer (10mM Tris, pH7.0, 50mM NaCl). The annealing process was carried out in accordance with 98°C 10min, -0.1°C/0.6s, 300 cycles, 65°C 5min, -0.1°C/0.6s, 400 cycles (wherein segments A, B, C, D, F, and G were provided by Sangon Biotech).

[00301] The prepared DNA construct and the different mutants of the helicase EPH-PEI26 shown in the table 10 below were pre-incubated together for 30 minutes at 25°C in a buffer (10mM HEPES, 50mM KCl, pH 7.0). Then, TMAD with a final concentration of 1 mM was added thereto for catalytic treatment for 30 minutes. Finally, the DNA construct/enzyme mixture purified by magnetic beads and the double-stranded DNA sequence of target nucleic acid of SEQ ID NO: 17 (E region) were added to a T4 rapid ligation reaction system (as shown in Table 11 below, the T4 rapid ligation kit was provided by Enzymatics) at a molar concentration ratio of 1.5:1, and the rapid ligation was performed at room temperature for 10 minutes. The ligation product was purified by magnetic beads, and then put into the subsequent sequencing system.

[00302] Table 10: Different mutants of helicase EPH-PEI26 used in this example

Mutant No.	Mutant region 1	Mutant region 2	Mutant region 3
Eph-Pei26-1	D99C/A366C		
Eph-Pei26-2	D99C/A366C		R88Q
Eph-Pei26-3	D99C/A366C		H87Y
Eph-Pei26-4	D99C/A366C		R88N
Eph-Pei26-5	D99C/A366C		K249R
Eph-Pei26-6	D99C/A366C		H87Q
Eph-Pei26-7	D99C/A366C		V422H
Eph-Pei26-8	D99C/A366C		S293R
Eph-Pei26-9	D99C/A366C		E286K
Eph-Pei26-10	D99C/A366C		V155I
Eph-Pei26-11	D99C/A366C		K91R
Eph-Pei26-12	D99C/A366C		V155L
Eph-Pei26-13	D99C/A366C		S293N
Eph-Pei26-14	D99C/A366C		F246Y
Eph-Pei26-15	D99C/A366C		F246R
Eph-Pei26-16	D99C/A366C	C308T/C419D	
Eph-Pei26-17	D99C/A366C	C308T/C419D	E286K
Eph-Pei26-18	D99C/A366C	C308T/C419D	E286K/V422H
Eph-Pei26-19	D99C/A366C	C308T/C419D	E286K/F246R
Eph-Pei26-20	D99C/A366C	C308T/C419D	E286K/F246R/V422H
Eph-Pei26-21	D99C/A366C	C308T/C419D	H87Q/E286K/F246R
Eph-Pei26-22	D99C/A366C	C308T/C419D	H87Q/E286K/F246R/V422H
Eph-Pei26-23	D99C/A366C	C308T/C419D	E286K/F246Y/S293N

Eph-Pei26-24	D99C/A366C	C308T/C419D	E286K/F246Y/S293N/V422H
Eph-Pei26-25	D99C/A366C	C114V/C119V/C141S/C308S/C419D	F103W/E286K
Eph-Pei26-26	D99C/A366C	C114V/C119V/C141S/C308S/C419D	
Eph-Pei26-27	D99C/A366C	C114V/C141S/C308T/C419D	E286K/F246Y
Eph-Pei26-28	D99C/A366C	C114V/C141S/C308T/C419D	I282F/E286K/F246Y

[00303] Table 11. T4 rapid ligation reaction system

Reaction component	Final concentration
2x rapid ligation Buffer	1x
DNA construct/enzyme complex	10uM
End-repaired nucleic acid library SEQ ID NO: 17 with a fixed-length of 10kb	10uM
Nuclease-free water	N/A
T4 DNA Ligase (600U/ul)	30U/ul
Total volume	20ul

[00304] Electrical measurements were acquired from MspA nanopores inserted in 1,2-diphytanoyl-glycero-3-phosphocholine lipid bilayers. Bilayers were formed across about 25 μm diameter apertures on a PTFE film via the Montal-Mueller technique, separating two 100 μL buffered solutions. All experiments were carried out in the stated buffered solution. Single-channel currents were measured on an amplifier equipped with digitizers. Ag/AgCl electrodes were connected to the buffered solutions so that the *cis* compartment was connected to the ground of the amplifier, and the *trans* compartment was connected to the active electrode of the amplifier. After achieving a single pore in the bilayer, the incubated DNA/enzyme mixture was added to 70 μL of buffer (10mM HEPES, 500mM KCl, 50mM MgCl_2 , 50mM ATP, pH 7.0) in the *cis* compartment of the electrophysiology chamber and a constant potential of +180 mV was applied at 35°C to initiate capture of the DNA construct/enzyme complex in the MspA nanopore (SEQ ID NO: 12) and the movement of nucleic acid through the nanopore controlled by the helicase, and the DNA movement controlled by helicase was monitored and recorded.

[00305] 2. Results

[00306] This example analyzes the characteristic parameters of the current signals generated by the movement of the nucleic acid sequence of SEQ ID NO: 17 through the nanopore controlled by different mutants of helicase Eph-Pei26, such as a median speed, a deletion ratio, an insertion ratio, an uneven ratio, and flick ratio, etc., to illustrate the influence of substitutions of amino acids that interact with the template strand or complementary single-stranded nucleic acid on the ability of the enzyme to control the movement of nucleic acids,

thereby screening out helicase mutants that effectively control the movement of nucleic acids through a nanopore and improve the accuracy and throughput of sequencing. The above parameters can be calculated by the method as follows. The sequencing current signal of an original fixed-length library was segmented and compared with that of a reference base sequence by a common sequence alignment algorithm (such as dynamic time warping), giving a processed current signal and corresponding base sequence information, and each parameter can be calculated separately based on the segmentation and comparison results. The above parameters are defined or calculated as follows: (1) The median speed is calculated in the same way as described in Example 6; (2) The deletion ratio is defined as the value obtained by dividing the number of missing bases by the total number of bases in the reference sequence, wherein the missing bases are bases missing in the original current signal compared to the reference sequence; (3) The insertion ratio is defined as the ratio of insertion steps in the original current signal, and one insertion is present if two or more steps correspond to one base according to the comparison result; (4) The uneven ratio is defined as the proportion of uneven steps. Ideally, each base may be expressed as a flat continuous signal on the original current signal. In fact, however the continuous signal is only relatively flat, and a signal block is considered as an uneven step if the standard deviation of the current value is greater than a certain threshold. The proportion of uneven steps is defined as the value obtained by dividing the number of uneven steps by the total number of steps; (5) The flick ratio is defined as the proportion of signal blocks with large signal disturbances. The signal is high-pass filtered to obtain a component with a relatively high frequency. If the high-pass filtered signal block has a standard deviation greater than a certain threshold, it is considered that this signal block has high interference. The number of high interference blocks divided by the total number is the flick ratio. The lower the value of parameters (2)-(5), the better the sequencing results.

[00307] The analysis results of various parameters of different mutants are shown in Table 12. It can be seen from the results of parameter analysis that different amino acid positions have different effects on the movement of nucleic acids controlled by enzyme, including the speed, especially the uneven ratio, deletion ratio and insertion ratio. This may be caused by interference of amino acid side chains on base shift or slippage. The smaller the above parameters, the better the quality of the current signals generated by the movement of nucleic acids controlled by the enzyme. In the evaluation of different mutants, the effects on various parameters were combined to investigate the effects of mutations. The mutants with large differences among the above parameters were selected for subsequent base recognition

model training in order to verify that the effect of modifications of amino acid sites on the difference in the parameters of movement of the nucleic acid controlled by the enzyme and the effect of modifications of amino acid sites on the accuracy of sequencing.

[00308] Table 12: Current signal characteristic parameters of different mutants

Mutant No.	Insertion ratio	Uneven ratio	Deletion ratio	Flick ratio	Median speed (bp/s)
Eph-Pei26-1 (control)	7.26%	57.10%	11.47%	14.39%	541
Eph-Pei26-2	4.38%	49.10%	13.19%	12.70%	640
Eph-Pei26-3	5.49%	51.48%	11.52%	10.71%	453
Eph-Pei26-4	5.55%	51.33%	11.89%	12.82%	590
Eph-Pei26-5	5.95%	55.25%	12.47%	14.53%	588
Eph-Pei26-6	6.58%	54.65%	13.28%	13.92%	566
Eph-Pei26-7	6.72%	56.13%	12.23%	14.60%	583
Eph-Pei26-8	6.94%	55.05%	10.44%	13.47%	513
Eph-Pei26-9	7.27%	59.38%	10.98%	15.26%	502
Eph-Pei26-10	7.46%	56.22%	8.23%	11.19%	416
Eph-Pei26-11	7.53%	54.77%	9.64%	12.12%	477
Eph-Pei26-12	7.57%	56.39%	9.01%	11.89%	477
Eph-Pei26-13	7.65%	55.39%	8.66%	12.57%	471
Eph-Pei26-14	9.16%	63.49%	6.16%	15.06%	405
Eph-Pei26-15	10.39%	67.01%	9.24%	17.07%	366
Eph-Pei26-16	5.39%	53.48%	12.97%	13.77%	587
Eph-Pei26-17	5.41%	45.64%	13.19%	10.45%	583
Eph-Pei26-18	5.94%	50.28%	13.23%	12.95%	608
Eph-Pei26-19	7.79%	52.13%	12.73%	12.78%	471
Eph-Pei26-20	7.00%	51.80%	13.10%	12.93%	505
Eph-Pei26-21	7.95%	63.02%	12.44%	19.58%	427
Eph-Pei26-22	6.80%	51.42%	14.07%	11.81%	471
Eph-Pei26-23	5.27%	47.07%	13.27%	10.50%	587
Eph-Pei26-24	4.56%	45.18%	12.52%	10.28%	598
Eph-Pei26-25	8.14%	60.40%	11.11%	16.65%	445
Eph-Pei26-26	6.26%	56.24%	10.50%	14.33%	443
Eph-Pei26-27	11.79%	51.93%	8.84%	15.2%	494
Eph-Pei26-28	12.9%	53.49%	9.1%	16.1%	497

[00309] Example 9

[00310] In this example, mutants Eph-Pei26-24 and Eph-Pei26-25 were selected because they showed significant differences in the comprehensive performance of the four parameters including deletion ratio, insertion ratio, uneven ratio and flick ratio, as shown in Table 12 of Example 8. The sequencing data of the movement of randomly digested library from Escherichia coli and the human genome through the MspA nanopore (SEQ ID NO: 12) controlled by the mutants were collected, and the same algorithm model was used to train base

recognition models for different mutants. By evaluating the performance of models of different mutants in the validation set of chip data and test chip data, the differences in the ability of different mutants to control nucleic acid movement were evaluated with the various parameters in Table 12 of Example 8, enabling selection of suitable mutants for nanopore nucleic acid sequencing.

[00311] 1. Methods

[00312] The *E. coli* and human genome nucleic acids were randomly digested, and then were used to replace the E fragment of the DNA construct shown in Figure 4. Fragment E and fragments A/B/C/D/F were annealed to form a complex. The ligation product was purified by magnetic beads and incubated with Eph-Pei26-24 and Eph-Pei26-25 helicase mutants and catalyzed with TMAD. The processed products were purified and added to the sequencer chip for sequencing and collecting signals.

[00313] 2. Results

[00314] Figure 14 shows the accuracy of the model after multiple rounds of training of the validation set data of randomly digested *E. coli* and human genome library sequenced by Eph-Pei26-24 and Eph-Pei26-25. Table 13 shows the test results of the base recognition models trained for mutants EPH-PEI26-24 and EPH-PEI26-25 on their respective chip data. The accuracy rate of the validation set refers to the ratio of the sequence aligned completely with the reference sequence in a single library signal after the model obtained after training performs base recognition on the chip data of the validation set. Correspondingly, the accuracy rate of the test chip refers to the ratio of the sequence aligned completely with the reference sequence in a single library signal when the trained model is used to identify the data of the test chip. The accuracy rates in the example are the median values after statistics. It can be seen that the accuracy of the mutant Eph-Pei26-24 on the *E. coli* and human genome validation set data is higher than that of Eph-Pei26-25, and this difference has also been verified on the test chip. The difference in accuracy between the test chips and between the test chip and the validation set, especially the difference between the chips is related to the repeatability of batches of experiments. The results demonstrate that the parameters in Example 8 regarding the control of the movement of nucleic acid by different mutants are effective for assessing the difference in the ability of enzymes to control the movement of the nucleic acid.

[00315] Table 13: Test results of mutants Eph-Pei26-24 and Eph-Pei26-25 on their respective test chip data

Library type	Test parameters	Eph-Pei26-25	Eph-Pei26-24
--------------	-----------------	--------------	--------------

E. coli genome	Number of signals	9676	47537
	Total throughput	43.07M	256.75M
	Aligned throughput	40.04M	247.71M
	Median accuracy	83.29%	86.50%
	Alignment rate	92.96%	96.48%
Human genome	Number of signals	10285	68274
	Total throughput	51.06M	390.97M
	Aligned throughput	48.23M	374.2M
	Median accuracy	86.72%	87.54%
	Alignment rate	94.46%	95.78%
Phage genome	Number of signals	5745	11939
	Total throughput	40.17M	88.84M
	Aligned throughput	38.25M	85.51M
	Median accuracy	82.88%	85.20%
	Alignment rate	95.22%	96.25%

[00316] The preferred embodiments of the present application are described in detail above. However, the present application is not limited to the specific details in the above embodiments. Within the scope of the technical concept of the present application, various simple variations could be made to the technical solution of the present application, and these simple variations belong to the protection scope of claims.

[00317] In addition, it should be noted that the various specific technical features described in the above specific embodiments can be combined in any suitable manner without contradiction. In order to avoid unnecessary repetition, the present application does not specify any possible combination mode separately.

WHAT IS CLAIMED IS:

1. A Pif1-like helicase in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into a tower domain, a pin domain and/or a 1A domain of the Pif1-like helicase, wherein the helicase further comprises:

(A) substitution or deletion of at least one or more native amino acids, wherein the substituted or deleted native amino acids comprise amino acids corresponding to C308 and/or C419 in SEQ ID NO: 6; and/or

(B) substitution of at least one amino acid that interacts with one or more nucleotides in single-stranded DNA or double-stranded DNA, wherein the substituted amino acids comprise amino acids corresponding to H87 and/or V422 and/or I282 in SEQ ID NO: 6, wherein the Pif1-like helicase retains its ability to control the movement of a polynucleotide.

2. The Pif1-like helicase according to claim 1, wherein the helicase comprises:

(a) a variant of SEQ ID NO: 1 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E264-P278 and N296-A394 of the tower domain, and/or residues K89-E105 of the pin domain, and/or residues M1-L88 and M106-V181 of the 1A domain;

(b) a variant of SEQ ID NO: 2 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E265-P279 and N297-A392 of the tower domain, and/or residue K89-D105 of the pin domain, and/or residue M1-L88 and I106-M180 of the 1A domain;

(c) a variant of SEQ ID NO: 3 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues T266-P280 and N298-S403 of the tower domain, and/or residues K89-A109 of the pin domain, and/or residues M1-L88 and K110-V182 of the 1A domain;

(d) a variant of SEQ ID NO: 4 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues T266-P280 and N298-S404 of the tower domain, and/or residues K89-A109 of the pin domain, and/or residues M1-L88 and K110-V182 of the 1A domain;

(e) a variant of SEQ ID NO: 5 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E260-P274 and N292-A391 of the tower domain, and/or residues K86-E102 of the pin domain, and/or residues M1-L84 and M103-K177 of the 1A domain;

- (f) a variant of SEQ ID NO: 6 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E266-P280 and N298-A396 of the tower domain, and/or residues K91-E107 of the pin domain, and/or residues M1-L90 and M108-M183 of the 1A domain;
- (g) a variant of SEQ ID NO: 7 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues T276-P290 and N308-P402 of the tower domain, and/or residues K100-D116 of the pin domain, and/or residues M1-L99 and D117-M191 of the 1A domain;
- (h) a variant of SEQ ID NO: 8 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues D274-P288 and N306-A404 of the tower domain, and/or residues K95-E112 of the pin domain, and/or residues M1-L95 and I113-K187 of the 1A domain;
- (i) a variant of SEQ ID NO: 9 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E260-P274 and N292-A391 of the tower domain, and/or residues K86-E102 of the pin domain, and/or residues M1-L85 and M103-K177 of the 1A domain;
- (j) a variant of SEQ ID NO: 10 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E265-P279 and H297-A393 of the tower domain, and/or residues K88-E104 of the pin domain, and/or residues M1-L87 and I105-K180 of the 1A domain; or
- (k) a variant of SEQ ID NO: 11 in which at least one cysteine residue and/or at least one non-natural amino acid have been introduced into residues E264-P278 and N296-P389 of the tower domain, and/or residues K97-A113 of the pin domain, and/or residues M1-L96 and P114-K184 of the 1A domain.

3. The Pif1-like helicase according to claim 2, wherein the helicase comprises:

- (a) a variant of SEQ ID NO: 11, which comprises (i) E105C and/or A362C; (ii) E104C and/or K360C; (iii) E104C and/or A362C; (iv) E104C and/or Q363C; (v) E104C and/or K366C; (vi) E105C and/or M356C; (vii) E105C and/or K360C; (viii) E104C and/or M356C; (ix) E105C and/or Q363C; (x) E105C and/or K366C; (xi) F108C and/or M356C; (xii) F108C and/or K360C; (xiii) F108C and/or A362C; (xiv) F108C and/or Q363C; (xv) F108C and/or K366C; (xvi) K134C and/or M356C; (xvii) K134C and/or K360C; (xviii) K134C and/or A362C; (xix) K134C and/or Q363C; (xx) K134C and/or K366C; (xxi) any of (i) to (xx) and G359C; (xxii) any of (i) to (xx) and Q111C; (xxiii) any of (i) to (xx) and I138C; (xxiv) any of (i) to (xx) and Q111C and I138C; (xxv) E105C and/or F377C; (xxvi) Y103L, E105Y, N352N, A362C and

Y365N; (xxvii) E105Y and A362C; (xxviii) A362C; (xxix) Y103L, E105C, N352N, A362Y and Y365N; (xxx) Y103L, E105C and A362Y; (xxxi) E105C and/or A362C, and I280A; (xxxii) E105C and/or L358C; (xxxiii) E104C and/or G359C; (xxxiv) E104C and/or A362C ; (Xxxv) K106C and/or W378C; (xxxvi) T102C and/or N382C; (xxxvii) T102C and/or W378C; (xxxviii) E104C and/or Y355C; (xxxix) E104C and/or N382C; (xl) E104C and/or K381C; (xli) E104C and/or K379C; (xlii) E104C and/or D376C; (xliii) E104C and/or W378C; (xliv) E104C and/or W374C; (xlv) E105C and/or Y355C; (xlvi) E105C and/or N382C; (xlvii) E105C and/or K381C; (xlviii) E105C and/or K379C; (xlix) E105C and/or D376C; (l) E105C and/or W378C; (li) E105C and/or W374C; (lii) E105C and A362Y; (liii) E105C, G359C and A362C; or (liv)I2C, E105C and A362C; or,

(b) a variant of any one of SEQ ID NOs: 1 to 10, which comprises a cysteine residue at the positions which correspond to those in SEQ ID NO: 11 as defined in any of (i) to (liv).

4. The Pif1-like helicase according to claim 1, wherein the amino acid sequence of the Pif1-like helicase is any of the amino acid sequences shown in SEQ ID NOs:1 to 11 or has at least 30%, at least 40%, at least 50%, 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.9% homology with any of the amino acid sequences shown in SEQ ID NOs: 1 to 11, and has the ability to control the movement of a polynucleotide.

5. The Pif1-like helicase according to claim 1 or 2, wherein the introduced cysteines are connected to one another, the introduced non-natural amino acids are connected to one another, the introduced cysteine and the introduced non-natural amino acid are connected to one another, the introduced cysteine and the native amino acid are connected to one another, or the introduced non-natural amino acid and the native amino acid are connected to one another.

6. The Pif1-like helicase according to claim 1, wherein in (A), the native amino acid is substituted with a non-polar amino acid, a polar amino acid or a charged amino acid; the non-polar amino acid includes but not limited to glycine (G), alanine (A) or valine (V); the polar amino acids include but are not limited to serine (S), threonine (T), tyrosine (Y), asparagine (N) or glutamine (Q); the charged amino acids include but are not limited to aspartic acid (D), glutamic acid (E) or histidine (H); and/or the one or more cysteines (C) are substituted with alanine (A), serine (S) , threonine (T), aspartic acid (D) or valine (V).

7. The Pif1-like helicase according to claim 6, wherein the Pif1-like helicase comprises:

(a) a variant of SEQ ID NO: 6, and the one or more substituted or deleted native amino acids are a combination of (i) C308 and/or C419 and (ii) one or more of C114, C119, and C141;

- (b) a variant of SEQ ID NO: 1, and the one or more substituted or deleted native amino acids are one or more of C39, C112, C117, and C417;
- (c) a variant of SEQ ID NO: 2, and the one or more substituted or deleted native amino acids are one or more of C112, C117, C139, C202, C307, C327, C344, C395, and C415; or,
- (d) a variant of SEQ ID NO: 3, and the one or more substituted or deleted native amino acids are one or more of C15, C62, C142, C420, and C422;
- (e) a variant of SEQ ID NO: 4, and the one or more substituted or deleted native amino acids are one or more of C16, C63, C143, C422, and C444;
- (f) a variant of SEQ ID NO: 5, and the one or more substituted or deleted native amino acids are one or more of C109, C114, C136, C199, C302, C414, and C394;
- (g) a variant of SEQ ID NO: 7, and the one or more substituted or deleted native amino acids are one or more of C23, C123, C128, C150, C318, C405, and C425;
- (h) a variant of SEQ ID NO: 8, and the one or more substituted or deleted native amino acids are one or more of C119, C124, C146, C209, C407, and C424;
- (i) a variant of SEQ ID NO: 9, and the one or more substituted or deleted native amino acids are one or more of C109, C114, C136, C199, C302, C414, and C394;
- (j) a variant of SEQ ID NO: 10, and the one or more substituted or deleted native amino acids are one or more of C14, C111, C138, C116, C243, C396, C410, C416, and C347; or,
- (k) a variant of SEQ ID NO: 11, and the one or more substituted or deleted native amino acids are one or more of C125, C128, C412, and C315.

8. The Pif1-like helicase according to claim 1, wherein the Pif1-like helicase further comprises substitution or modification of surface negatively-charged amino acids, polar or non-polar amino acids; the substitution comprises substitution of negatively charged amino acids, uncharged amino acids, aromatic amino acids, polar or non-polar amino acids with positively charged amino acids, or uncharged amino acids.

9. The Pif1-like helicase of claim 8, wherein the Pif1-like helicase comprises:

- (a) a variant of SEQ ID NO: 6 and the one or more negatively charged or uncharged amino acids are one or more of S9, S173, D208 or T218;
- (b) a variant of SEQ ID NO: 7 and the one or more negatively charged or uncharged amino acids are one or more of D8, E11, T26, S186, D216 or S226; or,
- (c) a variant of any one of SEQ ID NOs: 1 to 5 and 8 to 11, wherein the one or more negatively charged or uncharged amino acids correspond to one or more of S9, S173, D208 or T218 in SEQ ID NO: 6; or one or more of D8, E11, T26, S186, D216 or S226 in SEQ ID NO: 7.

10. The Pif1-like helicase according to claim 1, wherein the non-natural amino acid is selected from 4-Azido-L-phenylalanine (Faz), 4-Acetyl-L-phenylalanine, 3-Acetyl-L-phenylalanine, 4-Acetoacetyl-L-phenylalanine, O-Allyl-L-tyrosine, 3-(Phenylselanyl)-L-alanine, O-2-Propyn-1-yl-L-tyrosine, 4-(Dihydroxyboryl)-L-phenylalanine, 4-[(Ethylsulfanyl)carbonyl]-L-phenylalanine, (2*S*)-2-amino-3-{4-[(propan-2-ylsulfanyl)carbonyl]phenyl}propanoic acid, (2*S*)-2-amino-3-{4-[(2-amino-3-sulfanylpropanoyl)amino]phenyl}propanoic acid, O-Methyl-L-tyrosine, 4-Amino-L-phenylalanine, 4-Cyano-L-phenylalanine, 3-Cyano-L-phenylalanine, 4-Fluoro-L-phenylalanine, 4-Iodo-L-phenylalanine, 4-Bromo-L-phenylalanine, O-(Trifluoromethyl)tyrosine, 4-Nitro-L-phenylalanine, 3-Hydroxy-L-tyrosine, 3-Amino-L-tyrosine, 3-Iodo-L-tyrosine, 4-Isopropyl-L-phenylalanine, 3-(2-Naphthyl)-L-alanine, 4-Phenyl-L-phenylalanine, (2*S*)-2-amino-3-(naphthalen-2-ylamino)propanoic acid, 6-(Methylsulfanyl)norleucine, 6-Oxo-L-lysine, D-tyrosine, (2*R*)-2-Hydroxy-3-(4-hydroxyphenyl)propanoic acid, (2*R*)-2-Ammoniooctanoate-3-(2,2'-Bipyridin-5-yl)-D-alanine, 2-amino-3-(8-hydroxy-3-quinolyl)propanoic acid, 4-Benzoyl-L-phenylalanine, S-(2-Nitrobenzyl)cysteine, (2*R*)-2-amino-3-[(2-nitrobenzyl)sulfanyl]propanoic acid, (2*S*)-2-amino-3-[(2-nitrobenzyl)oxy]propanoic acid, O-(4,5-Dimethoxy-2-nitrobenzyl)-L-serine, (2*S*)-2-amino-6-({[(2-nitrobenzyl)oxy]carbonyl}amino)hexanoic acid, O-(2-Nitrobenzyl)-L-tyrosine, 2-Nitrophenylalanine, 4-[(*E*)-Phenyldiazenyl]-L-phenylalanine, 4-[3-(Trifluoromethyl)-3H-diaziren-3-yl]-D-phenylalanine, 2-amino-3-[[5-(dimethylamino)-1-naphthyl]sulfonylamino]propanoic acid, (2*S*)-2-amino-4-(7-hydroxy-2-oxo-2H-chromen-4-yl)butanoic acid, (2*S*)-3-[(6-acetylnaphthalen-2-yl)amino]-2-aminopropanoic acid, 4-(Carboxymethyl)phenylalanine, 3-Nitro-L-tyrosine, O-Sulfo-L-tyrosine, (2*R*)-6-Acetamido-2-ammoniohexanoate, 1-Methylhistidine, 2-Aminononanoic acid, 2-Aminodecanoic acid, L-Homocysteine, 5-Sulfanylnorvaline, 6-Sulfanyl-L-norleucine, 5-(Methyl sulfanyl)-L-norvaline, N⁶-{[(2*R*,3*R*)-3-Methyl-3,4-dihydro-2H-pyrrol-2-yl]carbonyl}-L-lysine, N⁶-[(Benzyloxy)carbonyl]lysine, (2*S*)-2-amino-6-[(cyclopentylcarbonyl)amino]hexanoic acid, N⁶-[(Cyclopentylloxy)carbonyl]-L-lysine, (2*S*)-2-amino-6-{[(2*R*)-tetrahydrofuran-2-ylcarbonyl]amino}hexanoic acid, (2*S*)-2-amino-8-[(2*R*,3*S*)-3-ethynyltetrahydrofuran-2-yl]-8-oxooctanoic acid, N⁶-(tert-Butoxycarbonyl)-L-lysine, (2*S*)-2-Hydroxy-6-({[(2-methyl-2-propanyl)oxy]carbonyl}amino)hexanoic acid, N⁶-[(Allyloxy)carbonyl]lysine, (2*S*)-2-amino-6-({[(2-azidobenzyl)oxy]carbonyl}amino)hexanoic acid, N⁶-L-Prolyl-L-lysine, (2*S*)-2-amino-6-({[(prop-2-yn-1-yloxy)carbonyl]amino}hexanoic acid and N⁶-(2-Azidoethoxy)carbonyl]-L-lysine.

11. The Pif1-like helicase according to claim 1, wherein in (B), at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is substituted with an amino acid containing a larger side chain.
12. The Pif1-like helicase according to claim 11, wherein the Pif1-like helicase comprises:
 - (a) a variant of SEQ ID NO: 11, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is at least one of P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 or P100;
 - (b) a variant of SEQ ID NO: 6, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA is a combination of (i) H87, V422 and/or I282 and (ii) at least one of P94, F103, V155, P67, M124, D126, E156, P157, E279, S293, N93, H403 or F246; or,
 - (c) a variant of any one of SEQ ID NOs: 1 to 5 and 7 to 10, wherein the at least one amino acid that interacts with a sugar ring and/or base of one or more nucleotides in single-stranded DNA or double-stranded DNA corresponds to at least one of P73, H93, N99, F109, I280, A161, F130, D132, D162, D163, E277, K415, Q291, H396, Y244 or P100 in SEQ ID NO: 11, or a combination of (i) H87, V422 and/or I282 and (ii) at least one of P94, F103, V155, P67, M124, D126, E156, P157, E279, S293, N93, H403 or F246 in SEQ ID NO: 6.
13. The Pif1-like helicase according to claim 11 or 12, wherein the larger side chain comprises an increased number of carbon atoms, has an increased length, an increased molecular volume, and/or has an increased van der Waals volume.
14. The Pif1-like helicase according to claim 11, wherein the larger side chain increases (i) electrostatic interaction, (ii) hydrogen bonding, (iii) cation- π interaction and/or (iv) π - π interaction between the at least one amino acid and one or more nucleotides in the single-stranded DNA or double-stranded DNA.
15. The Pif1-like helicase according to claim 11, wherein the amino acid containing the larger side chain is not alanine (A), cysteine (C), glycine (G), selenocysteine (U), methionine (M), aspartic acid (D) or glutamic acid (E).
16. The Pif1-like helicase according to claim 11, wherein:
 - a) histidine (H) is substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q) or asparagine (N); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); (iv) tyrosine (Y), arginine (R) or glutamine (Q); or (v) arginine (R), tyrosine (Y), asparagine (N) or glutamine (Q);

- b) asparagine (N) is substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); or (iv) glutamine (Q), arginine (R), histidine (H) or lysine (K);
- c) proline (P) is substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N), threonine (T) or histidine (H); (iii) tyrosine (Y), phenylalanine (F) or tryptophan (W); (iv) leucine (L), valine (V) or isoleucine (I); or (v) tryptophan (W), phenylalanine (F) or leucine (L);
- d) phenylalanine (F) is substituted with (i) arginine (R) or lysine (K); (ii) histidine (H); (iii) tyrosine (Y) or tryptophan (W); or (iv) arginine (R), tyrosine (Y), glutamine (Q) or histidine (H);
- e) aspartic acid (D) is substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W);
- f) valine (V) is substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); (iv) isoleucine (I) or leucine (L); or (v) tyrosine (Y), arginine (R), histidine (H) or tryptophan (W);
- g) serine (S) is substituted with (i) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W); (iv) isoleucine (I) or leucine (L); or (v) arginine (R), asparagine (N) or glutamine (Q);
- h) tyrosine (Y) is substituted with (i) arginine (R) or lysine (K); or (ii) tryptophan (W); and/or,
- i) isoleucine (I) is (i) phenylalanine (F) or tryptophan (W); (ii) valine (V) or leucine (L); or (iii) histidine (H), lysine (K) or arginine (R).

17. The Pif1-like helicase according to claim 11, wherein the Pif1-like helicase is a variant of SEQ ID NO: 11 comprising:

P73L; P73V; P73I; P73E; P73T; P73F; H93N; H93Q; H93W; N99R; N99H; N99W; N99Y; P100L; P100V; P100I; P100E; P100T; P100F; F130W; F130Y; F130H; D132H; D132Y; D132K; A161I; A161L; A161N; A161W; A161H; D162H; D162Y; D162K; D163W; D163F; D163Y; D163H; D163I; D163L; D163V; Y244W; Y244H; E277G; I280H; I280K; I280W; Q291K; Q291R; Q291W; Q291F; H396N; H396Q; H396W; K415W; K415R; K415H; K415Y; F109W/P73L; F109W/P73V; F109W/P73I; F109W/P73E; F109W/P73T; F109W/P73F; F109W/H93N; F109W/H93Q; F109W/H93W; F109W/N99R; F109W/N99H; F109W/N99W; F109W/N99Y; F109W/P100L; F109W/P100V; F109W/P100I; F109W/P100E; F109W/P100T; F109W/P100F; F109W/F130W; F109W/F130Y; F109W/F130H; F109W/D132H; F109W/D132Y; F109W/D132K; F109W/A161I; F109W/A161L;

F109W/A161N; F109W/A161W; F109W/A161H; F109W/D162H; F109W/D162Y; F109W/D162K; F109W/D163W; F109W/D163F; F109W/D163Y; F109W/D163H; F109W/D163I; F109W/D163L; F109W/D163V; F109W/Y244W; F109W/Y244H; F109W/E277G; F109W/I280H; F109W/I280K; F109W/I280W; F109W/Q291K; F109W/Q291R; F109W/Q291W; F109W/Q291F; F109W/H396N; F109W/H396Q; F109W/H396W; F109W/K415W; F109W/K415R; F109W/K415H; or F109W/K415Y; and/or the Pif1-like helicase is a variant of SEQ ID NO: 6 comprising:

P94W; P94F; F103W; I282F; V155L; V155I; D126H; D126N; D126Q; P157W; P157F; P157L; V422Y; V422R; V422H; V422W; S293R; S293N; S293Q; H87R; H87Y; H87N; H87Q; N93Q; N93R; N93K; N93H; H403Y; H403R; H403Q; F246R; F246Y; or F246Q.

18. The Pif1-like helicase according to claim 1, wherein, in (B), at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in single-stranded DNA or double-stranded DNA is substituted.

19. The Pif1-like helicase of claim 18, wherein the Pif1-like helicase comprises:

(a) a variant of SEQ ID NO: 11, in which the at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is at least one of H75, T91, S94, K97, N246, N247, N284, K288, N297, T394 or K397;

(b) a variant of SEQ ID NO: 6, in which the at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is at least one of K91, T85, R88, H69, K404, T401, N299, N248, E280, K290 or K249; or,

(c) a variant of any one of SEQ ID NOs: 1 to 5 and 7 to 10, in which the at least one amino acid that interacts with one or more phosphate groups of one or more nucleotides in ssDNA or dsDNA is at least one amino acid corresponding to H75, T91, S94, K97, N246, N247, N284, K288, N297, T394 or K397 in SEQ ID NO: 11, or at least one amino acid corresponding to K91, T85, R88, H69, K404, T401, N299, N248, E280, K290 or K249 in SEQ ID NO: 6.

20. The Pif1-like helicase according to claim 19, wherein:

a) histidine (H) is substituted with (i) arginine (R) or lysine (K); (ii) asparagine (N), serine (S), glutamine (Q) or threonine (T); (iii) phenylalanine (F), tryptophan (W) or tyrosine (Y); or (iv) asparagine (N), glutamine (Q) or arginine (R);

b) threonine (T) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) asparagine (N), serine (S), glutamine (Q) or histidine (H); (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); (iv) asparagine (N), glutamine (Q) or arginine (R); or (v) asparagine (N), histidine (H), lysine (K) or arginine (R);

c) serine (S) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) asparagine (N), glutamine (Q), threonine (T) or histidine (H); or (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H);

d) asparagine (N) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) serine (S), glutamine (Q), threonine (T) or histidine (H); (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); or (iv) glutamine (Q), arginine (R), histidine (H) or lysine (K);

e) lysine (K) is substituted with (i) arginine (R) or histidine (H); (ii) asparagine (N), serine (S), glutamine (Q), threonine (T) or histidine (H); (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); or (iv) arginine (R), glutamine (Q) or asparagine (N); and/or,

f) arginine (R) is substituted with (i) asparagine (N), serine (S) or glutamine (Q).

21. The Pif1-like helicase according to claim 18, wherein the Pif1-like helicase is a variant of SEQ ID NO: 11 comprising one or more of (a) to (k),

(a) H75N, H75Q, H75K or H75F; (b) T91K, T91Q or T91N; (c) S94H, S94N, S94K, S94T, S94R or S94Q; (d) K97Q, K97H or K97Y; (e) N246H or N246Q ; (f) N247Q or N247H; (g) N284H or N284Q; (h) K288Q or K288H; (i) N297Q, N297K or N297H; (j) T394K, T394H or T394N; or (k) K397R, K397H or K397Y; or

the Pif1-like helicase is a variant of SEQ ID NO: 6 comprising one or more of (a) to (i),

(a) K91R; (b) T85N, T85Q or T85R; (c) R88N or R88Q; (d) H69N, H69Q or H69R; (e) K404R; (f) T401N, T401H, T401K or T401R; (g) N299Q, N299R, N299H or N299K; (h) N248Q, N248R, N248H or N248K; or (i) K249R, K249Q or K249N.

22. The Pif1-like helicase according to claim 1, wherein, in (B), at least one amino acid that interacts with the double strand of one or more nucleotides in the double stranded DNA is substituted.

23. The Pif1-like helicase according to claim 22, wherein, the Pif1-like helicase comprises:

(a) a variant of SEQ ID NO: 6, in which the at least one amino acid that interacts with the double strand of one or more nucleotides in the double dsDNA is at least one of M81, V59, Q52, E286 or K290; or,

(b) a variant of any one of SEQ ID NOs: 1 to 5 and 7 to 11, in which the at least one amino acid that interacts with the double strand of one or more nucleotides in the double dsDNA is at least one amino acid corresponding to M81, V59, Q52, E286 or K290 in SEQ ID NO: 6.

24. The Pif1-like helicase according to claim 23, wherein, the Pif1-like helicase is a variant of SEQ ID NO: 6 comprising one or more of (a) to (e),

(a) M81K, M81R or M81H; (b) V59K, V59R or V59H; (c) Q52K, Q52R or Q52H; (d) E286K, E286R or E286H; or (e) K290R.

25. The Pif1-like helicase according to any one of claims 11 and 18-24, wherein the Pif1-like helicase is a variant of SEQ ID NO: 6, which comprises one or more of the following amino acid substitutions: T85N, H87Y, H87Q, H87N, R88Q, R88N, V155I, V155L, K91R, F103W, S239N, F246R, F246Y, K249R, I282F, E286K or V422H. 26. The Pif1-like helicase according to claim 1, wherein the helicase comprises substitution of at least one amino acid that interacts with the transmembrane pore:

(a) a variant of SEQ ID NO: 11, in which the Pif1-like helicase comprises one or more substitutions on (a) E196 (b) W202 (c) N199 or (d) G201; or ,

(b) a variant of any one of SEQ ID NOs: 1 to 10, in which the Pif1-like helicase comprises at least one substitutions on amino acid corresponding to (a) E196 (b) W202 (c) N199 or (d) G201.

27. The Pif1-like helicase according to claim 26, wherein the Pif1-like helicase is a variant of SEQ ID NO: 11 comprising substitutions at the following positions:

-F109/E196/H75, such as, F109W/E196L/H75N, F109W/E196L/H75Q, F109W/E196L/H75K or F109W/E196L/H75F;

-F109/E196/T91, such as, F109W/E196L/T91K, F109W/E196L/T91Q or F109W/E196L/T91N;

-F109/S94/E196, such as, F109W/S94H/E196L, F109W/S94T/E196L, F109W/S94R/E196L, F109W/S94Q/E196L, F109W/S94N/E196L or F109W/S94K/E196L;

-F109/N99/E196, such as, F109W/N99R/E196L, F109W/N99H/E196L, F109W/N99W/E196L or F109W/N99Y/E196L;

-F109/S94/E196/I280, such as, F109W/S94H/E196L/I280K;

-F109/P100/E196, such as, F109W/P100L/E196L, F109W/P100V/E196L, F109W/P100I/E196L or F109W/P100T/E196L;

-F109/D132/E196, such as, F109W/D132H/E196L, F109W/D132Y/E196L or F109W/D132K/E196L;

-F109/A161/E196, such as, F109W/A161I/E196L, F109W/A161L/E196L, F109W/A161N/E196L, F109W/A161W/E196L or F109W/A161H/E196L;

-F109/D163/E196, such as, F109W/D163W/E196L, F109W/D163F/E196L, F109W/D163Y/E196L, F109W/D163H/E196L, F109W/D163I/E196L, F109W/D163L/E196L or F109W/D163V/D163L/E196L;

-F109/Y244/E196, such as, F109W/Y244W/E196L, F109W/Y244Y/E196L or F109W/Y244H/E196L;

-F109/N246/E196, such as, F109W/N246H/E196L or F109W/N246Q/E196L;

-F109/E196/I280, such as, F109W/E196L/I280K, F109W/E196L/I280H, F109W/E196L/I280W or F109W/E196L/I280R;

-F109/E196/Q291, such as, F109W/E196L/Q291K, F109W/E196L/Q291R, F109W/E196L/Q291W or F109W/E196L/Q291F;

-F109/N297/E196, such as, F109W/N297Q/E196L, F109W/N297K/E196L or F109W/N297H/E196L;

-F109/T394/E196, such as, F109W/T394K/E196L, F109W/T394H/E196L or F109W/T394N/E196L;

-F109/H396/E196, such as, F109W/H396Y/E196L, F109W/H396F/E196L, F109W/H396Q/E196L or F109W/H396K/E196L;

-F109/K397/E196, such as, F109W/K397R/E196L, F109W/K397H/E196L or F109W/K397Y/E196L; or,

-F109/Y416/E196, such as, F109W/Y416W/E196L or F109W/Y416R/E196L.

28. A construct comprising at least one Pif1-like helicase according to any one of claims 1 to 27.

29. The construct according to claim 28, wherein the construct further comprises a polynucleotide binding moiety.

30. A nucleic acid encoding the Pif1-like helicase according to any one of claims 1 to 27 or the construct according to any one of claims 28 to 29.

31. An expression vector comprising the nucleic acid according to claim 30.

32. A host cell comprising the nucleic acid according to claim 30 or the expression vector according to claim 31.

33. A method of controlling the movement of a polynucleotide, comprising contacting the Pif1-like helicase according to any one of claims 1 to 27 or the construct according to any one of claims 28 to 29 with the polynucleotide.

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

I) contacting the Pif1-like helicase according to any one of claims 1 to 27 or the construct according to any one of claims 28 to 29 with the target polynucleotide and a pore, such that the Pif1-like helicase or the construct controls the movement of the target polynucleotide through the pore; and

II) taking one or more characteristics of the target polynucleotide when the nucleotide of the target polynucleotide interacts with the pore, and thereby characterising the target polynucleotide.

35. The method according to claim 34, wherein the method further comprises the step of applying a potential difference across the pore contacting the helicase or the construct and the target polynucleotide; the pore is selected from a biological pore, a solid state pore, or a biological and solid state hybrid pore.
36. The method according to claim 34 or 35, wherein the target polynucleotide is single-stranded, double-stranded, or at least partially double-stranded.
37. The method according to claim 34, wherein the one or more characteristics are selected from the source, length, identity, sequence, secondary structure of the target polynucleotide, or whether or not the target polynucleotide is modified; the one or more characteristics are measured by electrical measurement and/or optical measurement.
38. A product for characterising a target polynucleotide comprising the Pif1-like helicase according to any one of claims 1 to 27, the construct according to any one of claims 28 to 29, the nucleic acid according to claim 30, the expression vector according to claim 31 or the host cell according to claim 32, and a pore.
39. The product of claim 38, wherein the product is selected from a kit, a device or a sensor.
40. Use of the Pif1-like helicase according to any one of claims 1 to 27, the construct according to any one of claims 28 to 29, the nucleic acid according to claim 30, the expression vector according to claim 31, the host cell according to claim 32 or the product according to any one of claims 38 to 39 in characterising a target polynucleotide or controlling the movement of a target polynucleotide through a pore.