

PIF1-LIKE HELICASE AND APPLICATION 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] .

SUMMARY

[0008] 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.

[0009] 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 Pif1-like helicase retains its ability to control the movement of a polynucleotide.

[0010] 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:

[0011] (A) the tower domain;

[0012] (B) the pin domain;

[0013] (C) the 1A domain;

[0014] (D) the tower domain and the pin domain;

[0015] (E) the tower domain and the 1A domain;

[0016] (F) the 1A domain and the pin domain;

[0017] (G) the tower domain, the pin domain and the 1A domain.

[0018] 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.

[0019] 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.

[0020] 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.

[0021] Table 1. Wild-type Pif1-like helicases

Wild-type Pif1-like helicase		UniProt	Length
Pba-PM2 (SEQ ID NO:1)	Pectobacterium bacteriophage PM2	A0A0A0Q3A4	444
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
PphPspYZU05 (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

[0022] 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.

[0023] 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

[0024]

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

[0025] In a specific embodiment of the present application, the Pif1-like helicase is selected from Mph-MP1, Sph-CBH8, Eph-Pei26 or PphPspYZU05.

[0026] 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).

[0027] 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).

[0028] 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).

[0029] 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).

[0030] 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).

[0031] 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).

[0032] 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).

[0033] 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).

[0034] 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).

[0035] 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).

[0036] 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).

[0037] 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.

[0038] 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).

[0039] 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.

[0040] 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.

[0041] 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.

[0042] 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.

[0043] 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.

[0044] 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.

[0045] Preferably, the Pif1-like helicase is further modified by removal of one or more cysteine residues.

[0046] 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, serine or valine.

[0047] Preferably, the Pif1-like helicase comprises a variant of SEQ ID NO: 5, and the one or more substituted native cysteine residues are one or more of C109, C114, C136 or C414.

[0048] Preferably, the Pif1-like helicase comprises a variant of any one of SEQ ID NOs: 1, 2, 3, 4, 6, 7, 8, 9, 10, and 11, and the one or more substituted native cysteine residues correspond to one or more of C109, C114, C136 or C414 in SEQ ID NO:5.

[0049] Table 3 shows the amino acid positions in SEQ ID NOs: 1, 2, 3, 4, 6, 7, 8, 9, 10, and 11 corresponding to C109, C114, C136, and C414 in SEQ ID NO: 5.

[0050] Table 3. Corresponding amino acid positions in SEQ ID NOs: 1-4, and 6-11

SEQ ID NO:	Amino acid positions			
5	C109	C114	C136	C414
1	C112	C118	C139	C417
2	C113	C117	C139	C415
3	-	-	C143	-
4	-	-	C144	-
6	C114	C119	C141	C420
7	C124	C128	C150	C426
8	C119	C124	C146	
9	C109	C114	C136	C414
10	C111	C116	C138	C416
11	-	C125	-	C412

[0051] Preferably, the helicase is further modified to reduce its surface negative charge.

[0052] 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).

[0053] 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.

[0054] 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).

[0055] Preferably, the Pif1-like helicase comprises a variant of SEQ ID NO: 11, and the one or more negatively charged amino acids are one or more of D5, E9, E24, E87, I65, S58, D209 or D216. Any number of these amino acids may be neutralised, such as 1, 2, 3, 4, 5, 6, 7 or 8 of them. Any combination may be neutralised.

[0056] Preferably, the Pif1-like helicase comprises a variant of any one of SEQ ID NOs: 1 to 10, and the one or more negatively charged amino acids correspond to one or more of D5, E9, E24, E87, I65, S58, D209, or D216 in SEQ ID NO: 11. Amino acids in SEQ ID NOs: 1 to 10 which correspond to D5, E9, E24, E87, I65, S58, D209, or D216 in SEQ ID NO: 11 may be determined using the alignment shown in Figure 11.

[0057] Preferred substitution also comprises a variant of SEQ ID NO: 1, and the one or more negatively charged amino acids further comprise S171.

[0058] Preferred substitution also comprises a variant of SEQ ID NO: 9, and the one or more negatively charged amino acids further comprise S173.

[0059] 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, (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)-3*H*-diaziren-3-yl]-D-phenylalanine, 2-amino-3-[[5-(dimethylamino)-1-naphthyl]sulfonylamino]propanoic acid, (2*S*)-2-amino-4-(7-hydroxy-2-oxo-2*H*-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-Sulfamoylnorvaline, 6-Sulfamoyl-L-norleucine, 5-(Methyl sulfamoyl)-L-norvaline, N⁶-{[(2*R*,3*R*)-3-Methyl-3,4-dihydro-2*H*-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.

[0060] Preferably, the Pif1-like helicase further comprises: (a) substitution of at least one amino acid that interacts with one or more nucleotides in single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA); and/or, (b) substitution of at least one amino acid that interacts

with a transmembrane pore, wherein the Pif1-like helicase has the ability to control the movement of a polynucleotide.

[0061] More preferably, in (a), at least one amino acid that interacts with a sugar 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 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. Protein modeling may be used to identify amino acids that interact with a sugar and/or base of one or more nucleotides in single-stranded or double-stranded DNA.

[0063] 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 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. Relative to SEQ ID NO: 11, 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 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 10, wherein the at least one amino acid that interacts with a sugar 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. 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 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														
11	P73	H93	N99	P100	F109	F130	D132	A161	D162	D163	Y244	E277	I280	Q291	H396
1	P66	H85	N91	P92	F101	M122	G124	V153	E154	P155	F244	E277	I281	S291	H401
2	P65	H85	N92	P92	F101	M122	D124	V153	A154	P155	F245	E278	M281	T292	H399
3	P65	H85	H91	P92	F101	M126	D128	V157	R158	H159	F246	E279	M282	T293	H410
4	P65	H85	H91	P92	F101	M126	D128	V157	R158	H159	F246	E279	M282	T293	H411
5	P63	H82	S88	P89	F98	M119	D121	V150	S151	P152	F240	E273	M276	N287	H398
6	P67	H88	N93	P94	F103	M124	D126	V155	E156	P157	F246	E279	I283	S293	H403
7	P76	H96	N102	P103	F112	M133	D135	V164	A165	P166	F256	E290	M292	T303	H409
8	P72	H92	S98	P99	F108	M129	D131	V160	S161	P162	Y254	G287	T291	K301	H411
9	P63	H82	S88	P89	F98	M119	D121	V150	S151	P152	Y240	E273	M277	N287	H398
10	P64	H84	N91	P91	F100	M121	D123	V152	E153	L154	Y245	E278	T281	S292	H400

[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 and/or (iii) cation-pi (cation- 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. The non-natural amino acid may be any of those discussed below.

[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); or (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W). Histidine (H) is more preferably substituted with (a) N, Q or W or (b) Y, F, Q or K.

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

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

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

[0076] E) Phenylalanine (F) is preferably substituted with (i) arginine (R) or lysine (K); (ii) histidine (H); or (iii) tyrosine (Y) or tryptophan (W). Phenylalanine (F) is more preferably substituted with (a) W, (b) W, Y or H, (c) W, R or K or (d) K, H, W or R.

[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); or (iv) isoleucine (I) or leucine (L). Serine (S) is preferably substituted with K, R, W or F.

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

[0081] J) Arginine (R) is preferably substituted with (iii) 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) arginine (R) or lysine (K); (ii) glutamine (Q), asparagine (N) or histidine (H); (iii) phenylalanine (F), tyrosine (Y) or tryptophan (W) or (iv) leucine (L).

[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] 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.

[0091] 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).

[0092] 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).

[0093] 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).

[0094] 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. These numbers correspond to the relevant positions in SEQ ID NO: 11. Relative to SEQ ID NO: 11, 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.

[0095] The Pif1-like helicase preferably comprises a variant of any one of SEQ ID NOs: 1 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.

[0096] 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.

[0097] 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	N246	N247	N284	K288	N297	T394	K397
1	H67	T83	R86	K89	N246	K247	E284	K288	N297	T399	K402
2	H67	T83	S86	K89	N247	K248	K285	K289	N298	T397	K400
3	H67	T83	S86	K89	N248	D249	E286	E290	N299	T408	K411
4	H67	T83	S86	K89	N248	D249	E286	E290	N299	T409	K412
5	H64	T80	S83	K86	N242	K243	E280	K284	N293	T396	K399
6	H69	T85	R88	K91	N248	K249	E286	K290	N299	T401	K404
7	H78	T94	S97	K100	N258	K259	T296	K300	N309	T407	K410
8	H74	T90	S93	K96	N256	K257	T294	K298	N307	T409	K412
9	H64	T80	S83	K86	N242	K243	E280	K284	N293	T396	K400
10	H66	T82	S85	K86	N247	E248	K285	K289	N298	T398	K401

[0098] Preferably, the Pif1-like helicase comprises any one or more of the following:

[0099] a) Histidine (H) is substituted with (i) arginine (R) or lysine (K); (ii) asparagine (N), serine (S), glutamine (Q) or threonine (T); or (iii) phenylalanine (F), tryptophan (W) or tyrosine (Y);

[00100] b) Threonine (T) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) asparagine (N), serine (S), glutamine (Q) or histidine (H); or (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H);

[00101] c) Serine 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);

[00102] d) Asparagine (N) is substituted with (i) arginine (R), histidine (H) or lysine (K); (ii) serine (S), glutamine (Q), threonine (T) or histidine (H); or (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); and/or,

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

[00104] 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.

[00105] 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.

[00106] 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.

[00107] 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.

[00108] Table 6 shows amino acids in SEQ ID NOs: 1 to 10 corresponding to E196, W202, N199, and G201 in SEQ ID NO: 11.

[00109] Table 6. Amino acid positions in SEQ ID NOs: 1 to 10 corresponding to SEQ ID NO: 11

SEQIDNO:	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

[00110] 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:

[00111] -F109/E196/H75, such as, F109W/E196L/H75N, F109W/E196L/H75Q, F109W/E196L/H75K or F109W/E196L/H75F;

[00112] -F109/E196/T91, such as, F109W/E196L/T91K, F109W/E196L/T91Q or F109W/E196L/T91N;

[00113] -F109/S94/E196, such as, F109W/S94H/E196L, F109W/S94T/E196L, F109W/S94R/E196L, F109W/S94Q/E196L, F109W/S94N/E196L, or F109W/S94K/E196L;

[00114] F109/N99/E196, such as, F109W/N99R/E196L, F109W/N99H/E196L, F109W/N99W/E196L or F109W/N99Y/E196L;

[00115] -F109/S94/E196/I280, such as, F109W/S94H/E196L/I280K;

- [00116] -F109/P100/E196, such as, F109W/P100L/E196L, F109W/P100V/E196L, F109W/P100I/E196L or
- [00117] F109W/P100T/E196L;
- [00118] -F109/D132/E196, such as, F109W/D132H/E196L, F109W/D132Y/E196L or F109W/D132K/E196L;
- [00119] -F109/A161/E196, such as, F109W/A161I/E196L, F109W/A161L/E196L, F109W/A161N/E196L, F109W/A161W/E196L or F109W/A161H/E196L;
- [00120] -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;
- [00121] -F109/Y244/E196, such as, F109W/Y244W/E196L, F109W/Y244Y/E196L or F109W/Y244H/E196L;
- [00122] -F109/N246/E196, such as, F109W/N246H/E196L or F109W/N246Q/E196L;
- [00123] -F109/E196/I280, such as, F109W/E196L/I280K, F109W/E196L/I280H, F109W/E196L/I280W or F109W/E196L/I280R;
- [00124] -F109/E196/Q291, such as, F109W/E196L/Q291K, F109W/E196L/Q291R, F109W/E196L/Q291W or F109W/E196L/Q291F;
- [00125] -F109/N297/E196, such as, F109W/N297Q/E196L, F109W/N297K/E196L or F109W/N297H/E196L;
- [00126] -F109/T394/E196, such as, F109W/T394K/E196L, F109W/T394H/E196L or F109W/T394N/E196L;
- [00127] -F109/H396/E196, such as, F109W/H396Y/E196L, F109W/H396F/E196L, F109W/H396Q/E196L or F109W/H396K/E196L;
- [00128] -F109/K397/E196, such as, F109W/K397R/E196L, F109W/K397H/E196L or F109W/K397Y/E196L; or,
- [00129] -F109/Y416/E196, such as, F109W/Y416W/E196L or F109W/Y416R/E196L.
- [00130] 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.

[00131] 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").

[00132] 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.

[00133] Preferably, the construct further comprises a polynucleotide binding moiety.

[00134] Preferably, the construct has the ability to control the movement of a polynucleotide.

[00135] 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 of the polynucleotide, and/or a moiety that binds to a phosphate group of the polynucleotide.

[00136] 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.

[00137] 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.

[00138] 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:

[00139] 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 smegmatis	Q9AFI5	17401
SSBSso	ssb/RPA	Sulfolobus solfataricus	Q97W73	16138
SSBSso7D	ssb	Sulfolobus solfataricus	P39476	7280
SSBMHsmt	ssb	Homo sapiens	Q04837	17260
SSBMle	ssb	Mycobacterium leprae	P46390	17701
gp32T4	ssb	Bacteriophage T4	P03695	33506
gp32RB69	ssb	Bacteriophage RB69	Q7Y265	33118
gp2.5T7	ssb	Bacteriophage T7	P03696	25694

[00140] 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.

[00141] In the fourth aspect of the present application, provided is an expression vector comprising the nucleic acid of the present application.

[00142] Preferably, the nucleic acid is operably linked to a regulatory element in an expression vector, wherein the regulatory element is preferably a promoter.

[00143] In a specific embodiment of the present application, the promoter is selected from T7, trc, lac, ara or λ_L .

[00144] Preferably, the expression vector includes but is not limited to a plasmid, virus or phage.

[00145] 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.

[00146] Preferably, the host cell includes but is not limited to *Escherichia coli*.

[00147] 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.

[00148] 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.

[00149] 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.

[00150] 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 a expression vector, then transforming into *E. coli*, and performing an inducible expression and purification to obtain the Pif1-like helicase.

[00151] 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.

[00152] 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.

[00153] 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.

[00154] 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.

[00155] Preferably, the method may include controlling the movement of the polynucleotide with one or more Pif1-like helicases.

[00156] In the ninth aspect of the present application, provided is a method for characterising a target polynucleotide, comprising:

[00157] 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

[00158] 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.

[00159] 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.

[00160] 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.

[00161] Preferably, steps I) and II) are repeated one or more times.

[00162] 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.

[00163] 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.

[00164] 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.

[00165] 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.

[00166] Preferably, the pore is selected from a biological pore, a solid state pore, or a biological and solid state hybrid pore.

[00167] 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.

[00168] 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.

[00169] Preferably, the target polynucleotide is single-stranded, double-stranded, or at least partially double-stranded.

[00170] More preferably, the target polynucleotide may be modified by tags, spacers, methylation, oxidation or damage.

[00171] 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.

[00172] More preferably, the length of the target polynucleotide may be 10 to 100000 or more.

[00173] 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.

[00174] Preferably, the helicase is incorporated into an internal nucleotide of a single-stranded polynucleotide.

[00175] 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.

[00176] Preferably, taking the one or more characteristics are carried out by electrical measurement and/or optical measurement.

[00177] 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.

[00178] 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.

[00179] The electrical signal of the present application is selected from the measured values of current, voltage, tunneling, resistance, potential, conductivity or lateral electrical measurement.

[00180] In a specific embodiment of the present application, the electrical signal is a current passing through the pore.

[00181] Preferably, the characterising further comprises using an improved Viterbi algorithm.

[00182] 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.

[00183] Preferably, the product comprises multiple Pif1-like helicases and/or multiple constructs.

[00184] Preferably, the product comprises multiple pores. More preferably, the pores are nanopores, and the nanopores are transmembrane pores.

[00185] 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.

[00186] 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.

[00187] In a specific embodiment of the present application, the product comprises multiple Pif1-like helicases or multiple constructs, and multiple pores.

[00188] Preferably, the product is selected from a kit, a device or a sensor.

[00189] More preferably, the kit also comprises a chip containing a lipid bilayer. The pores span the lipid bilayer.

[00190] The kit of the present application comprises one or more lipid bilayers, and each lipid bilayer comprises one or more of the pores.

[00191] 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.

[00192] 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.

[00193] 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.

[00194] 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.

[00195] 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.

[00196] Preferably, the device comprises a plurality of Pif1-like helicases and/or a plurality of constructs, and a plurality of pores.

[00197] 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.

[00198] 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.

[00199] 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.

[00200] 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.

[00201] 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.

[00202] 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.

[00203] 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.

[00204] 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.

[00205] The “Pif1-like helicase” of the present application is modified relative to the wild-type or natural helicase.

[00206] 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.

[00207] 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.

[00208] 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.

[00209] The term “two or more” in the present application includes two, three, four, five, six, seven, eight or more and so on.

[00210] 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.

[00211] 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.

[00212] The term “and/or” in the present application includes the selection of one and any number of combinations of the listed items.

[00213] 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.

[00214] 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-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, (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-[(Cyclopentylloxy)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.

BRIEF DESCRIPTION OF THE DRAWINGS

[00215] Hereinafter, the embodiments of the present application will be described in detail with reference to the accompanying drawings, in which:

[00216] 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 μM 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.

[00217] 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.

[00218] 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.

[00219] Figure 4 shows DNA construct which is used in examples, in which SEQ ID NO: 13 (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 3' cholesterol tether).

[00220] 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).

[00221] 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)).

[00222] 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.

[00223] 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.

[00224] 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.

[00225] 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.

[00226] Figure 11 shows the correspondence among the amino acid sequences of SEQ NOs: 1 to 11.

DETAILED DESCRIPTION

[00227] 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.

[00228] 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.

[00229] Example 1. Preparation of Pif1-like helicase

[00230] 1. Materials and methods

[00231] 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.

[00232] 2. Results

[00233] Figure 10 shows an SDS-PAGE gel electrophoresis image of purified Mph MP1 (variant of SEQ ID NO: 11).

[00234] **Example 2**

[00235] This Example illustrates the capabilities of Pif1-like helicase to unwind hybridized dsDNA by using a fluorescence assay to detect enzyme activity.

[00236] 1. Materials and methods

[00237] 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.

[00238] 2. Results

[00239] 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 μ M capture DNA).

[00240] **Example 3**

[00241] This example shows an exemplary DNA binding capabilities of modified Pif1-like helicase of the present application by using gel measurements.

[00242] 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.

[00243] 1. Materials and methods

[00244] 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 μ M, 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.

[00245] 2. Results

[00246] 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.

[00247] Example 4

[00248] 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).

[00249] 1. Materials and methods

[00250] DNA construct A 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: 17, and the SEQ ID NO: 15 region of this construct was hybridized to SEQ ID NO: 16 (having a 3' cholesterol tether).

[00251] 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).

[00252] 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.

[00253] 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.

[00254] Results and Discussion

[00255] 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.

[00256] Example 5

[00257] 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.

[00258] 1. Materials and methods

[00259] 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

3' 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.

[00260] 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.

[00261] 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.

[00262] 2. Results

[00263] 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.

[00264] 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.

[00265] 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.

Sequence Listing

<110> Qitan Technology Ltd., Beijing

<120> A PIF1-LIKE HELICASE AND THE USE THEREOF

<160> 35

<170> SIPOSequenceListing 1.0

<210> 1

<211> 444

<212> PRT

<213> Artificial Sequence

<400> 1

```

Met Ser Leu Val Phe Glu Asp Leu Lys Gln Gly Gln Arg Glu Ala Phe
1          5          10          15
Asn Arg Ile Ile Glu Val Val Lys Lys Arg Ser Gly Gly Arg Ile Thr
20          25          30
Leu Asn Gly Pro Ala Gly Cys Gly Lys Thr Thr Leu Thr Lys Phe Ile
35          40          45
Ile Asp His Leu Val Arg Asn Gly Ile Leu Gly Val Val Leu Ala Ala
50          55          60
Pro Thr His Gln Ala Lys Lys Val Leu Ser Lys Leu Ser Gly Val Glu
65          70          75          80
Ala Asn Thr Ile His Arg Ile Leu Lys Ile Asn Pro Asn Thr Tyr Glu
85          90          95
Asp Gln Asp Ile Phe Glu Gln Arg Glu Met Pro Asp Leu Ser Lys Cys
100         105         110
Asn Val Leu Ile Cys Asp Glu Ala Ser Met Tyr Gly Asp Lys Leu Phe
115         120         125
Gly Ile Ile Leu Arg Ser Val Pro Ser Trp Cys Val Ile Ile Gly Ile
130         135         140
Gly Asp Arg Glu Gln Leu Pro Pro Val Glu Pro Gly Ser Asp Gly Gln
145         150         155         160
Thr Leu Ile Ser Pro Phe Phe Thr His Pro Ser Phe Glu Gln Leu Tyr
165         170         175
Leu Thr Glu Val Val Arg Ser Asn Thr Pro Ile Ile Asp Val Ala Thr
180         185         190
Glu Ile Arg Met Gly Ser Trp Leu Arg Glu Asn Ile Val Asp Gly His
195         200         205
Gly Val His Glu Phe Asn Ser Ser Thr Ala Leu Lys Asp Tyr Met Thr
210         215         220
Glu Tyr Phe Asn Val Val Lys Asp Ala Asp Asp Leu Ile Glu Thr Arg
225         230         235         240
Met Leu Ala Phe Thr Asn Lys Ser Val Asp Lys Leu Asn Ser Ile Ile
245         250         255
Arg Arg Arg Leu Tyr Glu Thr Glu Thr Ser Phe Ile Lys Asp Glu Ile
260         265         270
Ile Val Met Gln Glu Pro Met Ile Lys Glu Leu Glu Phe Asp Gly Lys
275         280         285
Lys Phe Ser Glu Thr Ile Phe Asn Asn Gly Gln Leu Val Arg Ile Lys

```

290		295		300											
Asp	Ala	Met	Leu	Thr	Ser	Gly	Phe	Leu	Ser	Ala	Arg	Asn	Val	Ser	Thr
305					310					315					320
Arg	Gln	Met	Ile	Asn	Tyr	Trp	Ser	Leu	Glu	Val	Glu	Thr	Ala	Glu	Asp
			325						330					335	
Asp	Glu	Glu	Tyr	Arg	Val	Asp	Val	Ile	Lys	Phe	Leu	Pro	Ala	Asp	Gln
			340					345					350		
Val	Glu	Lys	Phe	Asn	Tyr	Phe	Leu	Ala	Lys	Thr	Ala	Thr	Thr	Tyr	Arg
	355						360					365			
Glu	Met	Lys	Asn	Ala	Gly	Lys	Lys	Ala	Pro	Trp	Glu	Asp	Phe	Trp	Lys
370						375					380				
Ala	Lys	Arg	Thr	Phe	Leu	Lys	Val	Arg	Ala	Leu	Pro	Val	Ser	Thr	Ile
385					390					395					400
His	Lys	Ala	Gln	Gly	Val	Ser	Val	Asn	Arg	Ser	Phe	Leu	Tyr	Thr	Pro
			405					410						415	
Cys	Ile	His	Ile	Ala	Glu	Ala	Gln	Leu	Ala	Lys	Gln	Leu	Ala	Tyr	Val
			420					425					430		
Gly	Val	Thr	Arg	Ala	Arg	His	Asp	Val	Tyr	Tyr	Val				
	435						440								

<210> 2
 <211> 442
 <212> PRT
 <213> Artificial Sequence

<400>	2														
Met	Thr	Asp	Leu	Gln	Phe	Asp	Asp	Leu	Thr	Glu	Gly	Gln	Gln	Asn	Ala
1	5	10	15												
Phe	Asn	Ala	Ala	Leu	Glu	Ala	Met	Lys	Thr	Lys	Gly	Gln	His	Ile	Thr
	20	25	30												
Ile	Asn	Gly	Pro	Ala	Gly	Thr	Gly	Lys	Thr	Thr	Leu	Thr	Lys	Phe	Leu
	35	40	45												
Ile	Asn	His	Leu	Ile	Arg	Thr	Gly	Glu	Ser	Gly	Ile	Met	Leu	Ala	Ala
	50	55	60												
Pro	Thr	His	Gln	Ala	Lys	Lys	Val	Leu	Ser	Lys	Leu	Ala	Gly	Met	Glu
65	70	75	80												
Ala	Gln	Thr	Ile	His	Ser	Leu	Leu	Lys	Ile	Asn	Pro	Thr	Thr	Tyr	Glu
	85	90	95												
Asp	Ala	Thr	Leu	Phe	Glu	Gln	Ser	Asp	Val	Pro	Asp	Leu	Ser	Glu	Cys
	100	105	110												
Arg	Val	Leu	Ile	Cys	Asp	Glu	Val	Ser	Met	Tyr	Asp	Arg	Glu	Leu	Phe
	115	120	125												
Arg	Ile	Leu	Met	Ala	Ser	Val	Pro	Tyr	Trp	Cys	Thr	Ile	Ile	Gly	Leu
	130	135	140												
Gly	Asp	Ile	Ala	Gln	Ile	Arg	Pro	Val	Ala	Pro	Asn	Ser	Asn	Ile	Pro
145	150	155	160												
Glu	Val	Ser	Ala	Phe	Phe	Leu	Asn	Glu	Lys	Phe	Glu	Gln	Val	Ala	Leu
	165	170	175												
Thr	Glu	Val	Met	Arg	Ser	Asn	Ala	Pro	Ile	Ile	Glu	Val	Ala	Thr	Glu
	180	185	190												
Ile	Arg	His	Gly	Lys	Trp	Ile	Arg	Glu	Cys	Leu	Leu	Asn	Gly	Glu	Gly
	195	200	205												
Val	His	Asp	Met	Val	Leu	Pro	Thr	Gly	Gly	Ser	Val	Ala	Asn	Phe	Met
	210	215	220												

Tyr	Lys	Tyr	Phe	Asp	Ile	Val	Lys	Thr	Pro	Glu	Asp	Leu	Phe	Glu	Asn
225					230					235					240
Arg	Met	Leu	Ala	Phe	Thr	Asn	Lys	Ser	Val	Gly	Asn	Leu	Asn	Lys	Ile
				245					250					255	
Ile	Arg	Arg	Lys	Leu	Tyr	Gln	Thr	Glu	Val	Pro	Phe	Ile	Asn	Asp	Glu
			260					265					270		
Val	Leu	Val	Met	Gln	Glu	Pro	Leu	Met	Arg	Thr	His	Lys	Phe	Glu	Gly
		275					280					285			
Lys	Ser	Phe	Thr	Asp	Val	Arg	Phe	Asn	Asn	Gly	Glu	Leu	Val	Arg	Val
	290					295				300					
Leu	Ser	Cys	Gln	Pro	Ile	Thr	Lys	Arg	Leu	Ala	Ile	Arg	Gly	Ile	Asp
305					310					315					320
Gln	Glu	Asp	Val	Val	Lys	Cys	Trp	His	Leu	Glu	Leu	Arg	Ala	Ile	Glu
				325					330					335	
Thr	Asp	Val	Val	Asp	Ser	Ile	Cys	Val	Ile	Glu	Asp	Glu	Arg	Gln	Met
			340					345					350		
Lys	Ile	Phe	Gln	His	Tyr	Leu	Ser	Ala	Val	Ala	Tyr	Glu	Phe	Lys	Asn
		355					360					365			
Ser	Asn	Thr	Gly	Lys	Arg	Pro	Asn	Trp	Ser	Gly	Trp	Trp	Asp	Leu	Arg
	370					375				380					
Lys	Glu	Phe	His	Lys	Val	Lys	Ala	Leu	Pro	Cys	Gly	Thr	Ile	His	Lys
385					390					395					400
Ser	Gln	Gly	Thr	Ser	Val	Asp	Asn	Val	Phe	Leu	Tyr	Thr	Pro	Cys	Ile
				405					410					415	
His	Ser	Ala	Asp	Ala	Asp	Leu	Ala	Gln	Gln	Leu	Leu	Tyr	Val	Gly	Ala
			420					425					430		
Thr	Arg	Ala	Arg	Asn	Asn	Val	Tyr	Phe	Val						
		435					440								

<210> 3
 <211> 453
 <212> PRT
 <213> Artificial Sequence

<400> 3															
Met	Ser	Glu	Ala	Glu	Ser	Leu	Leu	Ala	Lys	Ile	Val	Leu	Thr	Asp	Cys
1				5					10					15	
Gln	Lys	Asn	Ala	Ile	Gly	Met	Val	Leu	Ser	Asp	Lys	Ser	His	Ile	Thr
			20					25					30		
Ile	Ser	Gly	Pro	Ala	Gly	Ser	Gly	Lys	Ser	Phe	Leu	Thr	Lys	Ile	Leu
		35				40						45			
Ile	Lys	Lys	Leu	Leu	Glu	Leu	Asn	Asn	Gly	Ser	Val	Val	Ala	Cys	Ala
	50					55				60					
Pro	Thr	His	Gln	Ala	Lys	Ile	Val	Leu	Ser	Lys	Met	Ser	Gly	Met	Thr
65				70						75				80	
Ala	Ala	Thr	Ile	His	Ser	Ile	Leu	Lys	Ile	His	Pro	Asp	Thr	Tyr	Glu
				85					90					95	
Asp	Val	Arg	Glu	Phe	Lys	Gln	Ser	Lys	Ser	Asp	Lys	Ala	Lys	Glu	Asp
			100					105					110		
Leu	Lys	Glu	Val	Arg	Tyr	Leu	Leu	Val	Asp	Glu	Gly	Ser	Met	Val	Asp
		115				120						125			
Asn	Asp	Leu	Phe	Glu	Ile	Leu	Leu	Lys	Ser	Val	His	Pro	Tyr	Cys	Gln
	130					135					140				
Ile	Ile	Ala	Ile	Gly	Asp	Lys	His	Gln	Ile	Gln	Pro	Val	Arg	His	Ala

145					150					155				160	
Pro	Gly	Glu	Ile	Ser	Pro	Phe	Phe	Thr	Asp	Lys	Arg	Phe	Arg	Leu	Ala
				165					170					175	
Glu	Leu	Lys	Thr	Ile	Val	Arg	Gln	Gln	Ala	Gly	Asn	Pro	Ile	Ile	Gln
			180					185					190		
Val	Ala	Thr	Lys	Ile	Arg	Asn	Gly	Gly	Trp	Phe	Glu	Thr	Asn	Trp	Asp
		195					200				205				
Lys	Glu	Ser	Gly	Thr	Gly	Val	Leu	Asp	Val	Lys	Ser	Val	Ala	Asn	Leu
	210					215				220					
Met	Lys	Ile	Tyr	Leu	Ser	Lys	Val	Lys	Thr	Pro	Asp	Asp	Leu	Leu	Asn
225					230					235				240	
Tyr	Arg	Met	Leu	Ala	Tyr	Thr	Asn	Asp	Val	Val	Asn	Arg	Phe	Asn	Lys
				245				250						255	
Ala	Ile	Arg	Lys	Gln	Val	Tyr	Asn	Thr	Thr	Glu	Pro	Phe	Val	Asp	Asn
		260						265					270		
Glu	Tyr	Leu	Val	Met	Gln	Glu	Pro	Val	Met	Lys	Glu	Ser	Glu	Ile	Gly
	275						280				285				
Gly	Glu	Thr	Phe	Thr	Glu	Thr	Leu	Leu	Asn	Asn	Gly	Glu	Thr	Val	Lys
	290					295				300					
Ile	Lys	Glu	Gly	Ser	Ile	Lys	Arg	Gln	Met	Lys	Tyr	Ile	Ser	Leu	Pro
305					310					315				320	
Tyr	Val	Asp	Pro	Ile	Gln	Ile	Glu	Ile	Ala	Thr	Met	Thr	Val	Ile	Arg
				325				330						335	
Asn	Glu	Val	Asp	Leu	Thr	Glu	Ile	Asp	Gly	Asp	Leu	Glu	Val	Glu	Leu
		340						345				350			
Ser	Val	Val	Trp	Asp	Ala	Asp	Gly	Gln	Val	Gln	Leu	Asp	Glu	Ala	Leu
	355						360				365				
Ser	Tyr	Ala	Ala	Ser	Gln	Tyr	Lys	Gln	Met	Gly	Ser	Gly	Lys	Ala	Thr
	370					375				380					
Ser	Arg	Leu	Trp	Glu	Ser	Phe	Trp	Gln	Val	Lys	Gly	Met	Phe	Thr	Asn
385					390					395				400	
Thr	Lys	Ser	Leu	Gly	Ala	Ser	Thr	Phe	His	Lys	Ser	Gln	Gly	Thr	Thr
				405				410						415	
Val	Ile	Gly	Val	Cys	Val	Tyr	Thr	Gly	Asp	Met	Asn	Phe	Ala	Gln	Phe
		420						425				430			
Glu	Ile	Gln	Thr	Gln	Leu	Gly	Tyr	Val	Gly	Cys	Thr	Arg	Ala	Gln	Lys
	435					440					445				
Trp	Val	Met	Tyr	Cys											
	450														

<210> 4
 <211> 454
 <212> PRT
 <213> Artificial Sequence

<400> 4
 Met Ser Glu Ala Ala Ser Leu Leu Ala Lys Ile Ile Leu Thr Asp Cys
 1 5 10 15
 Gln Lys Thr Ala Ile Asp Ala Val Leu Thr Asp Lys Lys His Ile Thr
 20 25 30
 Ile Ser Gly Pro Ala Gly Ser Gly Lys Ser Phe Leu Thr Lys Ile Leu
 35 40 45
 Ile Gln Lys Leu Leu Asp Leu Asn Ser Gly Ala Val Ile Thr Cys Ala
 50 55 60

Pro	Thr	His	Gln	Ala	Lys	Ile	Val	Leu	Ser	Lys	Met	Ser	Gly	Phe	Thr
65					70					75					80
Ala	Ser	Thr	Ile	His	Ser	Val	Leu	Lys	Ile	His	Pro	Asp	Thr	Tyr	Glu
			85						90					95	
Asp	Val	Arg	Glu	Phe	Lys	Gln	Ser	Lys	Ser	Asp	Lys	Ala	Lys	Glu	Asp
			100					105					110		
Leu	Lys	Ala	Val	Arg	Tyr	Leu	Ile	Val	Asp	Glu	Ala	Ser	Met	Val	Asp
		115					120					125			
Asn	Asp	Leu	Phe	Glu	Ile	Leu	Leu	Lys	Ser	Val	His	Pro	Phe	Cys	Gln
		130					135				140				
Ile	Ile	Ala	Ile	Gly	Asp	Lys	His	Gln	Ile	Gln	Pro	Val	Arg	His	Ala
145					150					155					160
Pro	Gly	Glu	Ile	Ser	Pro	Phe	Phe	Thr	Asp	Lys	Arg	Phe	Arg	Leu	Ala
				165					170					175	
Glu	Leu	Lys	Thr	Val	Val	Arg	Gln	Gln	Ala	Gly	Asn	Pro	Ile	Ile	Gln
			180					185					190		
Val	Ala	Thr	Lys	Ile	Arg	Asn	Gly	Gly	Trp	Phe	Glu	Thr	Asn	Trp	Asp
		195					200					205			
Lys	Ala	Thr	Gly	Thr	Gly	Val	Leu	Asp	Val	Lys	Thr	Ile	Ala	Lys	Met
		210				215					220				
Met	Gln	Ile	Tyr	Leu	Ser	Lys	Val	Lys	Thr	Pro	Glu	Asp	Leu	Leu	Asn
225					230					235					240
Tyr	Arg	Met	Leu	Ala	Tyr	Thr	Asn	Asp	Val	Val	Asn	Ser	Phe	Asn	Arg
				245					250					255	
Val	Ile	Arg	Lys	His	Val	Tyr	Lys	Thr	Thr	Glu	Pro	Phe	Val	Asp	Asn
			260					265					270		
Glu	Tyr	Leu	Val	Met	Gln	Glu	Pro	Val	Met	Arg	Glu	Glu	Glu	Ile	Gly
		275					280					285			
Gly	Glu	Thr	Phe	Thr	Glu	Thr	Leu	Leu	Asn	Asn	Gly	Glu	Thr	Val	Lys
		290				295					300				
Ile	Ile	Pro	Gly	Ser	Ile	Lys	Arg	Gln	Leu	Lys	Tyr	Ile	Ser	Leu	Pro
305					310					315					320
Tyr	Val	Glu	Pro	Ile	Gln	Ile	Glu	Val	Ala	Thr	Met	Leu	Val	Glu	Arg
				325					330					335	
Gln	Glu	Thr	Asp	Val	Thr	Asp	Asn	Val	Asp	Ser	Asp	Lys	Glu	Val	Glu
			340					345					350		
Ile	Ser	Val	Val	Trp	Asp	Ala	Ser	Ser	Gln	Val	Leu	Leu	Asp	Glu	Ala
		355					360					365			
Leu	Ser	Tyr	Ala	Ala	Ser	Gln	Tyr	Lys	Gln	Met	Gly	Ser	Gly	Lys	Ala
		370				375					380				
Thr	Ser	Arg	Leu	Trp	Glu	Ser	Phe	Trp	Gln	Val	Lys	Gly	Met	Phe	Val
385					390					395					400
Asn	Thr	Lys	Ser	Leu	Gly	Ala	Ser	Thr	Phe	His	Lys	Ser	Gln	Gly	Thr
				405					410					415	
Thr	Val	Ile	Gly	Val	Cys	Val	Tyr	Thr	Gly	Asp	Met	Asn	Phe	Ala	Gln
			420					425					430		
Phe	Glu	Ile	Gln	Thr	Gln	Leu	Gly	Tyr	Val	Gly	Cys	Thr	Arg	Ala	Gln
		435					440					445			
Lys	Trp	Val	Ile	Tyr	Cys										
		450													

<210> 5
 <211> 441
 <212> PRT

<213> Artificial Sequence

<400> 5

```

Met Asn Phe Asp Asp Leu Thr Val Gly Gln Lys Glu Ala Phe Glu Ile
1      5      10      15
Val Ile Glu Ala Ile Arg Thr Lys Lys His His Val Met Ile Asn Gly
      20      25      30
Pro Ala Gly Thr Gly Lys Thr Thr Met Thr Lys Phe Ile Leu Glu His
      35      40      45
Leu Val Arg Asn Gly Glu Leu Gly Ile Met Leu Ala Ala Pro Thr His
      50      55      60
Gln Ala Lys Lys Val Leu Ser Lys Leu Ser Gly His Gln Ala Ala Thr
      65      70      75      80
Ile His Ser Ile Leu Lys Ile Ser Pro Thr Thr Tyr Glu Ala Glu Ser
      85      90      95
Ile Phe Glu Gln Lys Glu Met Pro Asp Leu Ala Lys Cys Arg Val Leu
      100     105     110
Leu Cys Asp Glu Gly Ser Met Tyr Asp Gly Ala Leu Phe Lys Ile Leu
      115     120     125
Met Asn Thr Ile Pro Ser His Cys Thr Val Ile Gly Ile Gly Asp Glu
      130     135     140
Glu Gln Leu Arg Pro Val Ser Pro Gly Asp Ser Leu Pro Ser Lys Ser
      145     150     155     160
Pro Phe Phe Ser Asp His Arg Phe Lys Gln Val Thr Leu Thr Glu Val
      165     170     175
Lys Arg Ser Asn Gly Pro Ile Ile Lys Val Ala Thr Glu Ile Arg Asn
      180     185     190
Gly Gly Trp Phe Arg Glu Cys Ile Glu Asp Gly His Gly Phe His Gly
      195     200     205
Phe Asn Gly Asp Lys Pro Leu Gln Gln Tyr Met Met Lys Tyr Phe Asp
      210     215     220
Val Val Lys Ser Pro Glu Asp Leu Phe Glu Thr Arg Met Leu Ala Tyr
      225     230     235     240
Thr Asn Lys Ser Val Asp Lys Leu Asn Gly Ile Ile Arg Arg Lys Leu
      245     250     255
Tyr Glu Thr Glu Ser Pro Phe Ile Val Gly Glu Val Leu Val Met Gln
      260     265     270
Glu Pro Tyr Met Lys Ser Leu Glu Phe Asp Gly Lys Lys Phe Asn Glu
      275     280     285
Ile Ile Phe Asn Asn Gly Gln Met Val Arg Ile Leu Asp Cys Lys Leu
      290     295     300
Thr Ser Thr Phe Leu Lys Ala Arg Asp Val Ser Val Lys Gln Met Ile
      305     310     315     320
Ser Tyr Trp His Leu Glu Val Glu Thr Val Asp Glu Asp Asp Asp Tyr
      325     330     335
Gln Arg Glu Thr Ile Lys Val Leu Ala Asp Asp Asn Glu Lys Gln Lys
      340     345     350
Phe Asp Met Phe Leu Ala Lys Val Ala Thr Ser Tyr Arg Glu Leu Lys
      355     360     365
Ser Ala Gly Arg Arg Pro His Trp Ala Asp Phe Trp Asp Ala Lys Arg
      370     375     380
Thr Phe Leu Lys Val Lys Ala Leu Pro Cys Ser Thr Ile His Lys Ser
      385     390     395     400
Gln Gly Ile Ser Val Asp Asn Ala Phe Ile Tyr Thr Pro Cys Ile Thr

```

				405					410					415			
Leu	Ala	Asp	Ile	Asp	Leu	Ala	Lys	Gln	Leu	Ala	Tyr	Val	Ser	Ser	Thr		
			420					425					430				
Arg	Ala	Arg	His	Asp	Val	Tyr	Phe	Val									
		435					440										

<210> 6
 <211> 446
 <212> PRT
 <213> Artificial Sequence

<400>	6																
Met	Ser	Gln	Glu	Val	Thr	Phe	Glu	Ser	Leu	Asn	Lys	Gly	Gln	Arg	Glu		
1				5					10					15			
Ala	Phe	Asp	Ile	Ile	Thr	Ser	Ala	Ile	Gln	Arg	Arg	Asn	Gly	Glu	Arg		
		20						25					30				
Leu	Thr	Leu	Asn	Gly	Pro	Ala	Gly	Thr	Gly	Lys	Thr	Thr	Leu	Thr	Lys		
		35					40					45					
Phe	Ile	Ile	Gln	His	Ile	Val	Arg	Asn	Gly	Val	Leu	Gly	Val	Val	Leu		
	50					55				60							
Ala	Ala	Pro	Thr	His	Gln	Ala	Lys	Lys	Val	Leu	Ala	Lys	Met	Ser	Gly		
65					70				75						80		
Met	Glu	Ala	Asn	Thr	Ile	His	Arg	Val	Leu	Lys	Ile	Asn	Pro	Met	Thr		
			85					90						95			
Tyr	Glu	Asp	Gln	Asp	Val	Phe	Glu	Gln	Arg	Glu	Met	Pro	Asp	Met	Ser		
		100						105					110				
Lys	Cys	Asn	Val	Leu	Val	Cys	Asp	Glu	Ala	Ser	Met	Leu	Asp	Gly	Lys		
		115					120					125					
Ile	Phe	Lys	Ile	Ile	Leu	Asn	Ser	Ile	Pro	Pro	Trp	Ala	Val	Leu	Ile		
	130					135					140						
Gly	Ile	Gly	Asp	Arg	Glu	Gln	Ile	Gln	Pro	Val	Glu	Pro	Gly	Ser	Asp		
145					150				155						160		
Gly	Thr	Pro	Gln	Ile	Ser	Pro	Phe	Phe	Thr	His	Pro	Ser	Phe	Lys	Gln		
			165					170						175			
Val	His	Leu	Thr	Glu	Val	Met	Arg	Ser	Asn	Ala	Pro	Ile	Ile	Asp	Val		
		180					185						190				
Ala	Thr	Asp	Ile	Arg	Thr	Gly	Gly	Trp	Leu	Arg	His	His	Ile	Ile	Asp		
	195					200						205					
Gly	His	Gly	Val	His	Glu	Phe	Ala	Ser	Thr	Thr	Ala	Leu	Lys	Asp	Phe		
	210					215					220						
Met	Met	Gln	Tyr	Phe	Asp	Val	Val	Lys	Thr	Pro	Glu	Asp	Leu	Phe	Glu		
225					230				235						240		
Thr	Arg	Met	Leu	Ala	Phe	Thr	Asn	Lys	Ser	Val	Glu	Lys	Leu	Asn	Asn		
			245					250						255			
Ile	Ile	Arg	Arg	Lys	Leu	Tyr	Glu	Thr	Glu	Val	Pro	Phe	Ile	Asn	Glu		
		260					265						270				
Glu	Val	Ile	Val	Met	Gln	Glu	Pro	Phe	Ile	Lys	Glu	Leu	Glu	Phe	Asp		
	275					280					285						
Gly	Lys	Lys	Phe	Ser	Glu	Ile	Val	Phe	Asn	Asn	Gly	Glu	Met	Val	Arg		
	290					295					300						
Ile	Lys	Asp	Cys	Met	Leu	Thr	Ser	Met	Pro	Leu	Ile	Ala	Arg	Asn	Val		
305					310				315						320		
Ser	Thr	Lys	Gln	His	Ile	Asn	Tyr	Trp	Ala	Leu	Glu	Val	Glu	Thr	Ile		
			325					330						335			

Asp	Pro	Asp	Glu	Glu	Tyr	Lys	Ile	Glu	Val	Ile	Lys	Val	Leu	Pro	Leu
			340					345					350		
Asp	Gln	Tyr	Gln	Lys	Met	Asp	Met	Phe	Leu	Ala	Lys	Val	Ala	Thr	Thr
		355					360					365			
Tyr	Arg	Glu	Met	Lys	Ala	Ala	Gly	Lys	Arg	Pro	Pro	Trp	Asp	Asp	Phe
	370					375					380				
Trp	Lys	Ile	Lys	Arg	Thr	Phe	Leu	Lys	Val	Arg	Ala	Leu	Pro	Val	Ser
385					390					395					400
Thr	Ile	His	Lys	Ser	Gln	Gly	Ile	Ser	Val	Asn	Asn	Ser	Phe	Ile	Tyr
			405						410					415	
Thr	Pro	Cys	Ile	His	Val	Ala	Glu	Val	Gln	Leu	Ala	Arg	Gln	Leu	Ala
		420						425					430		
Tyr	Val	Gly	Leu	Thr	Arg	Ala	Arg	His	Asp	Ala	Tyr	Tyr	Val		
		435					440					445			

<210> 7
 <211> 452
 <212> PRT
 <213> Artificial Sequence

<400> 7															
Met	Asn	Asp	Leu	Lys	Leu	Glu	Asp	Leu	Asn	Glu	Gly	Gln	Arg	Ala	Ala
1				5					10					15	
Phe	Asp	Ala	Val	Ile	Lys	Val	Ile	Ser	Thr	Lys	Ser	Met	Met	Ala	Thr
		20						25				30			
Thr	Ser	Ser	Cys	Glu	Val	Lys	Pro	His	Ile	Thr	Ile	Asn	Gly	Pro	Ala
	35					40						45			
Gly	Thr	Gly	Lys	Thr	Thr	Leu	Thr	Arg	Phe	Leu	Ile	Asn	His	Leu	Ile
	50					55					60				
Ser	Ser	Gly	Glu	Lys	Gly	Val	Met	Leu	Ala	Ala	Pro	Thr	His	Gln	Ala
65				70					75					80	
Lys	Lys	Val	Leu	Ala	Lys	Leu	Ser	Gly	Met	Glu	Ala	Ser	Thr	Ile	His
			85					90						95	
Ser	Leu	Leu	Lys	Ile	Asn	Pro	Thr	Thr	Tyr	Glu	Asp	Ser	Thr	Val	Phe
			100					105					110		
Gln	Gln	Ser	Asp	Asp	Pro	Asp	Leu	Ser	Asp	Cys	Arg	Val	Leu	Ile	Cys
	115					120						125			
Asp	Glu	Val	Ser	Met	Tyr	Asp	Arg	Glu	Leu	Phe	Arg	Ile	Leu	Met	Ala
	130					135				140					
Ser	Ile	Pro	Pro	Trp	Ala	Thr	Ile	Ile	Gly	Leu	Gly	Asp	Ile	Ala	Gln
145				150					155					160	
Ile	Arg	Pro	Val	Ala	Pro	Asn	Ser	Thr	Thr	Pro	Glu	Leu	Ser	Ala	Phe
			165					170						175	
Phe	Phe	Asn	Asp	Lys	Phe	Gln	Gln	Val	Ser	Leu	Thr	Glu	Val	Met	Arg
		180						185					190		
Ser	Asn	Ala	Pro	Ile	Ile	Glu	Val	Ala	Thr	Glu	Ile	Arg	Lys	Gly	Gly
	195					200						205			
Trp	Ile	Arg	Glu	Asn	Leu	Val	Asp	Gly	Gln	Gly	Val	His	Ser	Met	Val
	210				215					220					
Arg	Ser	Asn	Gly	Gly	Ser	Val	Ala	Ala	Phe	Leu	Thr	Lys	Tyr	Phe	Glu
225				230					235					240	
Ile	Val	Lys	Asp	Pro	Asp	Asp	Leu	Phe	Asp	Asn	Arg	Met	Leu	Ala	Phe
			245					250					255		
Thr	Asn	Lys	Ser	Val	Asn	Asp	Leu	Asn	Asn	Ile	Ile	Arg	Lys	Lys	Leu

Lys Phe Lys Gln Ile Lys Leu Thr Glu Val Lys Arg Ser Asn Gly Pro
 180 185 190
 Ile Ile Thr Val Ala Arg Glu Ile Leu Lys Gly Gln Trp Leu Arg Glu
 195 200 205
 Cys Leu Asp Glu Asp Gly Gln Gly Val His Ala Tyr Asp Pro Glu Ser
 210 215 220
 Asp Ile Pro Ser Leu His Trp Phe Leu Lys Glu Tyr Phe Lys Val Val
 225 230 235 240
 Lys Thr Lys Glu Asp Phe Val Asn Thr Arg Val Met Ala Tyr Thr Asn
 245 250 255
 Lys Val Val Asn Thr Leu Asn Lys Ile Ile Arg Lys Arg Ile Phe Asn
 260 265 270
 Thr Asp Glu Pro Phe Ile Glu Asp Glu Ile Ile Val Met Gln Gly Pro
 275 280 285
 Leu Thr Glu Ser Leu Ile Val Asp Gly Lys Lys Val Lys Lys Leu Ile
 290 295 300
 Tyr Asn Asn Gly Gln Arg Val Arg Ile Val Arg Val Asn Lys Thr Val
 305 310 315 320
 His Thr Leu Arg Ala Arg Phe Val Glu Ser Thr Lys Glu Ile Asp Val
 325 330 335
 Trp Thr Leu Thr Val Glu Thr Ala Asp Lys Asn Ile Asp Glu Tyr His
 340 345 350
 Leu Lys Asp Leu His Ile Val Asp Glu Gly Ser Glu Leu Val Leu Lys
 355 360 365
 Glu Phe Leu Ser Glu Thr Ala Asn Thr Tyr Arg Tyr Trp Glu Leu Pro
 370 375 380
 Gly Lys Ala Pro Trp Gly Glu Phe Trp Thr Ile Lys Glu Arg Tyr Ser
 385 390 395 400
 Asn Val Lys Ala Glu Pro Cys Ser Thr Ile His Lys Ala Gln Gly Ile
 405 410 415
 Ser Val Asp Asn Ala Phe Leu Cys Thr Ser Gly Leu Ser Ser Met Asp
 420 425 430
 Pro Asp Leu Val Lys Glu Leu Ile Tyr Val Gly Ser Thr Arg Pro Lys
 435 440 445
 His Asn Leu Tyr Trp Ile
 450

<210> 9
 <211> 441
 <212> PRT
 <213> Artificial Sequence

<400> 9
 Met Lys Phe Glu Asp Leu Thr Val Gly Gln Lys Ser Ala Phe Asp Val
 1 5 10 15
 Val Ile Glu Ala Ile Lys Thr Lys Lys Phe His Val Met Ile Asn Gly
 20 25 30
 Pro Ala Gly Thr Gly Lys Thr Thr Met Thr Lys Phe Ile Leu Glu His
 35 40 45
 Leu Val Arg Asn Gly Glu Leu Gly Ile Met Leu Ala Ala Pro Thr His
 50 55 60
 Gln Ala Lys Lys Val Leu Ser Lys Leu Ser Gly His Glu Ala Ala Thr
 65 70 75 80
 Ile His Ser Val Leu Lys Ile Ser Pro Thr Thr Tyr Glu Ala Glu Ser

				85					90					95			
Ile	Phe	Glu	Gln	Lys	Glu	Met	Pro	Asp	Leu	Ala	Lys	Cys	Arg	Val	Leu		
			100					105					110				
Phe	Cys	Asp	Glu	Gly	Ser	Met	Tyr	Asp	Gly	Ala	Leu	Phe	Lys	Ile	Leu		
		115					120					125					
Met	Asn	Thr	Ile	Pro	Ser	His	Cys	Thr	Val	Ile	Gly	Ile	Gly	Asp	Glu		
		130				135					140						
Glu	Gln	Leu	Arg	Pro	Val	Ser	Pro	Gly	Asp	Ser	Leu	Pro	Ser	Lys	Ser		
145					150				155						160		
Pro	Phe	Phe	Ser	Asp	Ser	Arg	Phe	Lys	Gln	Val	Thr	Leu	Thr	Glu	Val		
				165				170						175			
Lys	Arg	Ser	Asn	Gly	Pro	Ile	Ile	Lys	Val	Ala	Thr	Glu	Ile	Arg	Thr		
			180				185						190				
Gly	Gly	Trp	Phe	Arg	Glu	Cys	Ile	Glu	Asp	Gly	His	Gly	Phe	His	Gly		
		195				200					205						
Phe	Gly	Gly	Asp	Lys	Pro	Leu	Gln	Gln	Tyr	Met	Met	Lys	Tyr	Phe	Asp		
		210				215				220							
Val	Val	Lys	Ser	Pro	Glu	Asp	Leu	Phe	Glu	Thr	Arg	Met	Leu	Ala	Tyr		
225					230				235						240		
Thr	Asn	Lys	Ser	Val	Asp	Lys	Leu	Asn	Gly	Ile	Ile	Arg	Arg	Lys	Leu		
				245				250						255			
Tyr	Glu	Thr	Glu	Asn	Pro	Phe	Ile	Val	Gly	Glu	Val	Leu	Val	Met	Gln		
			260				265						270				
Glu	Pro	Tyr	Met	Lys	Gln	Leu	Glu	Phe	Asp	Gly	Lys	Lys	Phe	Asn	Glu		
		275				280						285					
Ile	Ile	Phe	Asn	Asn	Gly	Gln	Met	Ile	Arg	Ile	Leu	Asp	Cys	Lys	Leu		
		290				295					300						
Thr	Ser	Thr	Phe	Leu	Lys	Ala	Arg	Asp	Val	Ser	Val	Lys	Gln	Met	Ile		
305					310				315					320			
Ser	Tyr	Trp	His	Leu	Glu	Val	Glu	Thr	Val	Asp	Glu	Asp	Asp	Asp	Tyr		
				325				330						335			
Gln	Arg	Glu	Thr	Ile	Lys	Val	Leu	Ala	Asp	Ala	Asn	Glu	Lys	Gln	Lys		
			340				345						350				
Phe	Asp	Met	Phe	Leu	Ala	Lys	Val	Ala	Thr	Thr	Tyr	Arg	Glu	Leu	Lys		
		355				360						365					
Ser	Ala	Gly	Arg	Arg	Pro	His	Trp	Ala	Asp	Phe	Trp	Asp	Ala	Lys	Arg		
		370				375					380						
Thr	Phe	Leu	Lys	Val	Lys	Ala	Leu	Pro	Cys	Ser	Thr	Ile	His	Lys	Ser		
385					390				395					400			
Gln	Gly	Ile	Ser	Val	Asp	Asn	Thr	Phe	Ile	Tyr	Thr	Pro	Cys	Ile	Thr		
				405				410						415			
Leu	Ala	Asp	Ile	Asp	Leu	Ala	Lys	Gln	Leu	Ala	Tyr	Val	Ser	Ala	Thr		
			420					425					430				
Arg	Ala	Arg	His	Asp	Val	Tyr	Phe	Val									
		435				440											

<210> 10
 <211> 443
 <212> PRT
 <213> Artificial Sequence

<400> 10
 Met Thr Ile Gly Phe Asp Asp Leu Thr Glu Gly Gln Lys Cys Ala Phe
 1 5 10 15

Glu	Thr	Ala	Val	Glu	Leu	Val	Asn	Ser	Lys	Arg	Lys	His	Met	Thr	Leu			
			20					25					30					
Asn	Gly	Pro	Ala	Gly	Ser	Gly	Lys	Thr	Thr	Trp	Thr	Arg	Phe	Phe	Ile			
		35					40					45						
Asp	His	Leu	Val	Arg	Ser	Gly	Glu	Ser	Gly	Val	Ile	Leu	Ala	Ala	Pro			
		50					55				60							
Thr	His	Gln	Ala	Lys	Lys	Val	Leu	Ser	Lys	Leu	Ser	Gly	Val	Glu	Ala			
65							70				75				80			
Ser	Thr	Ile	His	Ser	Ile	Leu	Lys	Ile	Asn	Pro	Thr	Thr	Tyr	Glu	Glu			
					85				90					95				
Asn	Val	Leu	Phe	Glu	Gln	Lys	Glu	Ile	Pro	Asp	Leu	Ala	Lys	Cys	Arg			
			100					105					110					
Val	Leu	Ile	Cys	Asp	Glu	Ala	Ser	Met	Tyr	Asp	Arg	Lys	Leu	Phe	Asp			
		115						120				125						
Ile	Leu	Met	Asn	Ser	Ile	Pro	Ser	Trp	Cys	Ile	Val	Ile	Ala	Leu	Gly			
		130					135				140							
Asp	Lys	Asp	Gln	Leu	Arg	Pro	Val	Glu	Leu	Asn	Ser	Glu	Gly	Lys	Gly			
145						150				155					160			
Gln	Ile	Ser	Ala	Phe	Phe	Tyr	Asp	Pro	Arg	Phe	Glu	Gln	Val	Phe	Leu			
					165				170					175				
Ser	Glu	Ile	Lys	Arg	Ser	Asn	Ser	Pro	Ile	Ile	Glu	Val	Ala	Thr	Ser			
					180			185					190					
Ile	Arg	Thr	Gly	Gly	Trp	Leu	Tyr	His	Asn	Leu	Gly	Asp	Asp	Gly	Thr			
		195					200					205						
Gly	Val	His	Gly	Tyr	Met	Asn	Lys	Gly	Ser	Ala	Leu	Lys	Asp	Phe	Phe			
		210					215				220							
Gly	Gln	Tyr	Phe	Asp	Thr	Val	Arg	Lys	Pro	Glu	Asp	Leu	Phe	Glu	Asn			
225						230				235					240			
Arg	Met	Cys	Ala	Tyr	Thr	Asn	Glu	Ser	Val	Asn	Lys	Leu	Asn	Ser	Ile			
					245				250					255				
Ile	Arg	Arg	Lys	Ile	Tyr	Asp	Thr	Glu	Asp	Pro	Phe	Val	Val	Asn	Glu			
			260					265					270					
Val	Leu	Val	Met	Gln	Glu	Pro	Leu	Thr	Lys	Glu	Ile	Lys	Phe	Glu	Gly			
		275					280					285						
Lys	Arg	Phe	Ser	Glu	Met	Ile	Phe	His	Asn	Gly	Gln	Met	Val	Arg	Val			
		290					295				300							
Val	Lys	Ala	Glu	Lys	Thr	Ser	Lys	Phe	Leu	Arg	Ala	Lys	Gly	Val	Ser			
						310				315					320			
Gly	Glu	Gln	Met	Ile	Arg	Tyr	Trp	Ser	Leu	Val	Val	Glu	Thr	Asn	Asp			
					325				330					335				
Ala	Glu	Asp	Glu	Tyr	Phe	Arg	Glu	Gln	Ile	Cys	Val	Leu	Ser	Asp	Glu			
					340			345					350					
Asn	Glu	Ile	Asn	Lys	Tyr	Tyr	Tyr	Phe	Leu	Ala	Lys	Val	Ala	Asp	Ala			
		355					360					365						
Tyr	Lys	Ser	Gly	Ala	Val	Lys	Ala	His	Trp	Ala	Asp	Phe	Trp	Ala	Ala			
		370					375				380							
Lys	Arg	Ala	Phe	Ile	Lys	Val	Lys	Ala	Leu	Pro	Cys	Ser	Thr	Ile	His			
						390				395				400				
Lys	Val	Gln	Gly	Ile	Ser	Val	Asp	Asn	Cys	Phe	Leu	Tyr	Thr	Pro	Cys			
					405				410					415				
Ile	His	Lys	Ala	Asp	Ala	Asp	Leu	Ala	Lys	Gln	Leu	Thr	Tyr	Val	Gly			
			420					425					430					
Ala	Thr	Arg	Pro	Arg	Phe	Asn	Leu	His	Tyr	Val								
		435					440											

<210> 11
 <211> 441
 <212> PRT
 <213> Artificial Sequence

<400> 11
 Met Ile Thr Ile Asp Gln Leu Thr Glu Gly Gln Phe Asp Ser Leu Gln
 1 5 10 15
 Arg Ala Lys Val Leu Ile Gln Glu Ala Thr Lys Asn Asp Gly Asn Trp
 20 25 30
 Asn His Arg Thr Lys His Leu Thr Ile Asn Gly Pro Ala Gly Thr Gly
 35 40 45
 Lys Thr Thr Met Met Lys Phe Leu Val Ser Trp Leu Arg Asp Glu Gly
 50 55 60
 Ile Thr Gly Val Ala Leu Ala Ala Pro Thr His Ala Ala Lys Lys Val
 65 70 75 80
 Leu Ala Asn Ala Val Gly Glu Glu Val Ser Thr Ile His Ser Ile Leu
 85 90 95
 Lys Ile Asn Pro Thr Thr Tyr Glu Glu Lys Gln Phe Phe Glu Gln Ser
 100 105 110
 Ala Pro Pro Asp Leu Ser Lys Ile Arg Ile Leu Ile Cys Glu Glu Cys
 115 120 125
 Ser Phe Tyr Asp Ile Lys Leu Phe Glu Ile Leu Met Asn Ser Ile Gln
 130 135 140
 Pro Trp Thr Ile Ile Ile Gly Ile Gly Asp Arg Ala Gln Leu Arg Pro
 145 150 155 160
 Ala Asp Asp Lys Gly Ile Ser Arg Phe Phe Thr Asp Gln Arg Phe Glu
 165 170 175
 Gln Thr Tyr Leu Thr Glu Ile Lys Arg Ser Asn Met Pro Ile Ile Glu
 180 185 190
 Val Ala Thr Glu Ile Arg Asn Gly Gly Trp Ile Arg Glu Asn Ile Ile
 195 200 205
 Asp Asp Leu Gly Val Lys Gln Asp Lys Ser Val Ser Glu Phe Met Thr
 210 215 220
 Asn Tyr Phe Lys Val Val Lys Ser Ile Asp Asp Leu Tyr Glu Thr Arg
 225 230 235 240
 Met Tyr Ala Tyr Thr Asn Asn Ser Val Asp Thr Leu Asn Lys Ile Ile
 245 250 255
 Arg Lys Lys Leu Tyr Glu Thr Glu Gln Asp Phe Ile Val Gly Glu Pro
 260 265 270
 Ile Val Met Gln Glu Pro Leu Ile Arg Asp Ile Asn Tyr Glu Gly Lys
 275 280 285
 Arg Phe Gln Glu Ile Val Phe Asn Asn Gly Glu Tyr Leu Glu Val Ser
 290 295 300
 Glu Ile Lys Pro Met Glu Ser Val Leu Lys Cys Arg Asn Ile Asp Tyr
 305 310 315 320
 Gln Leu Val Leu His Tyr Tyr Gln Leu Lys Val Lys Ser Ile Asp Thr
 325 330 335
 Gly Glu Ser Gly Leu Ile Asn Thr Ile Ser Asp Lys Asn Glu Leu Asn
 340 345 350
 Lys Phe Tyr Met Phe Leu Gly Lys Val Ala Gln Asp Tyr Lys Ser Gly
 355 360 365
 Thr Ile Lys Ala Phe Trp Asp Asp Phe Trp Lys Ile Lys Asn Asn Tyr

370		375		380
His Arg Val Lys Pro Leu	Pro Val Ser Thr	Ile His Lys Gly Gln Gly		
385		390		400
Ser Thr Val Asp Asn Ser Phe Leu Tyr	Thr Pro Cys Ile Thr Lys Tyr			
	405	410		415
Ala Glu Pro Asp Leu Ala Ser Gln Leu Leu Tyr Val Gly Val Thr Arg				
	420	425		430
Ala Arg His Asn Val Asn Phe Val Gly				
	435	440		

<210> 12
 <211> 185
 <212> PRT
 <213> Artificial Sequence

<400> 12
Met Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr
1 5 10 15
Leu Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu
20 25 30
Asp Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys
35 40 45
Tyr Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu
50 55 60
Leu Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn
65 70 75 80
Phe Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asn Asn Gly Asn Ile Thr
85 90 95
Ala Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro
100 105 110
Gly Val Ser Ile Ser Ala Arg Leu Gly Asn Gly Pro Gly Ile Gln Glu
115 120 125
Val Ala Thr Phe Ser Val Arg Val Ser Gly Ala Lys Gly Gly Val Ala
130 135 140
Val Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu
145 150 155 160
Leu Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr
165 170 175
Thr Tyr Gly Glu Pro Trp Asn Met Asn
180 185

<210> 13
 <211> 7
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 13
 ttttttt

7

<210> 14
 <211> 16
 <212> DNA/RNA

<213> Artificial Sequence

<400> 14

ccgttctcat tgggtgc

16

<210> 15

<211> 20

<212> DNA/RNA

<213> Artificial Sequence

<400> 15

tcactatcgc atttctcatga

20

<210> 16

<211> 20

<212> DNA/RNA

<213> Artificial Sequence

<400> 16

tcatgagaat gcgatagtgga

20

<210> 17

<211> 10563

<212> DNA/RNA

<213> Artificial Sequence

<400> 17

atcctgagac cccttttttt tgttttgcct tttagaattt tattcgccat tcaggctgcg

60

caactgttgg gaagggcgat cgggtgcgggc ctcttcgcta ttacgccagc tggcgaaagg

120

gggatgtgct gcaaggcgat taagttgggt aacgccaggg tttaccagcgt cacgacgttg

180

taaaacgacg gccagtgaat tcgagctcgg tacctcgcga atgcatctag atatcggtac

240

cttttttagaa tttttacctg ggcagacagg actgcgcggg cattcaaata catgtgggat

300

gcggtgctgg atattggctg tcctgatacc gcgcaggaga tgctgattaa ggcagaggct

360

gcgtataaga aagcagacga catctggaat ctgcgcaagg atgattattt tgttaacgat

420

gaagcgcggg cgcgttactg ggatgatcgt gaaaaggccc gtcttgcgct tgaagccgcc

480

cgaagaagg ctgagcagca gactcaacag gacaaaaatg cgcagcagca gagcgatacc

540

gaagcgtcac ggctgaaata taccgaagag gcgcagaagg cttacgaacg gctgcagacg

600

ccgctggaga aatataccgc ccgtcaggaa gaactgaaca aggcactgaa agacgggaaa

660

atcctgcagg cggattacaa cacgctgatg gcggcggcga aaaaggatta tgaagcgacg

720

ctgaaaaagc cgaacagtc cagcgtgaag gtgtctgcgg gcgatcgtca ggaagacagt	780
gctcatgctg ccctgctgac gcttcaggca gaactccgga cgctggagaa gcatgccgga	840
gcaaatgaga aaatcagcca gcagcgccgg gatttgtgga aggcggagag tcagttcgcg	900
gtactggagg aggcggcgca acgtcgccag ctgtctgcac aggagaaatc cctgctggcg	960
cataaagatg agacgctgga gtacaaacgc cagctggctg cacttggcga caaggttacg	1020
tatcaggagc gcctgaacgc gctggcgag caggcggata aattcgaca gcagcaacgg	1080
gcaaacggg ccgccattga tgcgaaaagc cgggggctga ctgaccggca ggcagaacgg	1140
gaagccacgg aacagcgcct gaaggaacag tatggcgata atccgctggc gctgaataac	1200
gtcatgtcag agcagaaaaa gacctgggcg gctgaagacc agcttcgcgg gaactggatg	1260
gcaggcctga agtccggctg gagtgagtgg gaagagagcg ccacggacag tatgtcgag	1320
gtaaaaagtg cagccacgca gacctttgat ggtattgcac agaatatggc ggcgatgctg	1380
accggcagtg agcagaactg gcgcagcttc acccgttccg tgctgtccat gatgacagaa	1440
attctgctta agcaggcaat ggtggggatt gtcgggagta tcggcagcgc cattggcggg	1500
gctgttggtg gcggcgcatc cgcgtcaggc ggtacagcca ttcaggccgc tgcggcga	1560
ttccattttg caaccggagg atttacggga accggcggca aatatgagcc agcggggatt	1620
gttcaccgtg gtgagtttgt cttcacgaag gaggaacca gccggattgg cgtggggaat	1680
ctttaccggc tgatgcgcgg ctatgccacc ggcggttatg tcggtacacc gggcagcatg	1740
gcagacagcc ggtcgcaggc gtccgggacg tttgagcaga ataaccatgt ggtgattaac	1800
aacgacggca cgaacgggca gataggtccg gctgctctga aggcgggtgta tgacatggcc	1860
cgcaagggtg cccgtgatga aattcagaca cagatgcgtg atggtggcct gttctccgga	1920
ggtggacgat gaagaccttc cgctggaaag tgaaaccgg tatggatgtg gcttcgggtc	1980
cttctgtaag aaaggtgcgc tttggtgatg gctatttctca gcgagcgctt gccgggctga	2040
atgccaacct gaaaacgtac agcgtgacgc tttctgtccc ccgtgaggag gccacggtac	2100
tggagtcgtt tctggaagag cacgggggct ggaaatcctt tctgtggacg ccgccttatg	2160
agtggcggca gataaaggtg acctgcgcaa aatggtcgtc gcgggtcagt atgctgcgtg	2220
ttgagttcag cgcagagttt gaacaggtgg tgaactgatg caggatatcc ggcaggaaac	2280
actgaatgaa tgcacccgtg cggagcagtc ggccagcgtg gtgctctggg aaatcgacct	2340

gacagaggtc ggtggagaac gttatTTTTT ctgtaatgag cagaacgaaa aaggtgagcc	2400
ggtcacctgg caggggcgac agtatcagcc gtatcccatt caggggagcg gttttgaact	2460
gaatggcaaa ggcaccagta cgcgccccac gctgacggtt tctaacctgt acggtatggt	2520
caccgggatg gcggaagata tgcagagtct ggtcggcgga acggtgggtcc ggcgtaaggt	2580
ttacgcccggt tttctggatg cgggtgaactt cgtcaacgga aacagttacg ccgatccgga	2640
gcaggaggtg atcagccgct ggcgcatgga gcagtgcagc gaactgagcg cgggtgagtc	2700
ctcctttgta ctgtccacgc cgacggaaac ggatggcgct gtttttccgg gacgtatcat	2760
gctggccaac acctgcacct ggacctatcg cggtgacgag tgcggttata gcggtccggc	2820
tgtcgcggat gaatatgacc agccaacgtc cgatatcacg aaggataaat gcagcaaatg	2880
cctgagcgggt tgtaagttcc gcaataacgt cggcaacttt ggcggttcc tttcattaa	2940
caaactttcg cagtaaattc catgacacag acagaatcag cgattctggc gcacgcccgg	3000
cgatgtgctc cagcggagtc gtgcggcttc gtggtaagca cgccggaggg ggaaagatat	3060
ttcccctgct tgaatatctc cggtgagccg gaggtatatt ccgtatgtcg ccggaagact	3120
ggctgcaggc agaaatgcag ggtgagattg tggcgctggt ccacagccac cccggtggtc	3180
tgccctggct gagtgaggcc gaccggcggc tgcaggtgca gagtgatttg ccgtggtggc	3240
tggctctgcc ggggacgatt cataagttcc gctgtgtgcc gcatctcacc gggcggcgct	3300
ttgagcacgg tgtgacggac tgttacacac tgttccggga tgcttatcat ctggcgggga	3360
ttgagatgcc ggactttcat cgtgaggatg actggtggcg taacggccag aatctctatc	3420
tggataatct ggaggcgacg gggctgtatc aggtgccgtt gtcagcggca cagccgggcg	3480
atgtgctgct gtgctgtttt ggttcacag tgccgaatca cgccgcaatt tactgcggcg	3540
acggcgagct gctgcacat attcctgaac aactgagcaa acgagagagg tacaccgaca	3600
aatggcagcg acgcacacac tccctctggc gtcaccgggc atggcgcgca tctgccttta	3660
cggggattta caacgatttg gtcgcccat cgaccttctg gtgaaaacgg gggctgaagc	3720
catccgggca ctggccacac agctcccggc gtttcgtcag aaactgagcg acggctggta	3780
tcaggtagcg attgccgggc gggacgtcag cacgtccggg ttaacggcg agttacatga	3840
gactctgcct gatggcgctg taattcatat tgttcccaga gtcgccgggg ccaagtcagg	3900
tggcgtattc cagattgtcc tgggggctgc cgccattgcc ggatcattct ttaccgccgg	3960

agccaccctt gcagcatggg gggcagccat tggggccggt ggtatgaccg gcatcctgtt	4020
ttctctcggg gccagtatgg tgctcggtag tgtggcgcag atgctggcac cgaaagccag	4080
aactccccgt atacagacaa cggataacgg taagcagaac acctatttct cctcactgga	4140
taacatgggt gccaggggca atgttctgcc tgttctgtac ggggaaatgc gcgtggggtc	4200
acgcgtgggt tctcaggaga tcagcacggc agacgaaggg gacggtggtc aggttgtggg	4260
gattggtcgc tgatgcaaaa tgttttatgt gaaaccgcct gcgggcgggt ttgtcattta	4320
tggagcgtga ggaatgggta aaggaagcag taaggggcat accccgcgcg aagcgaagga	4380
caacctgaag tccacgcagt tgctgagtgt gatcgatgcc atcagcgaag ggccgattga	4440
aggtccgggt gatggcttaa aaagcgtgct gctgaacagt acgccgggtgc tggacactga	4500
ggggaatacc aacatatccg gtgtcacggg ggtgttccgg gctggtgagc aggagcagac	4560
tccgccggag ggatttgaat cctccggctc cgagacgggt ctgggtacgg aagtgaata	4620
tgacacgccg atcacccgca ccattacgtc tgcaaacatc gaccgtctgc gctttacctt	4680
cgggtgtacag gcactgggtg aaaccacctc aaagggtgac aggaatccgt cggaagtccg	4740
cctgctgggt cagatacaac gtaacggtag ctgggtgacg gaaaaagaca tcaccattaa	4800
gggcaaaacc acctcgcagt atctggcctc ggtggtgatg ggtaacctgc cgccgcgccc	4860
gtttaatatc cggatgcgca ggatgacgcc ggacagcacc acagaccagc tgcagaacaa	4920
aacgctctgg tcgtcataca ctgaaatcat cgatgtgaaa cagtgtctacc cgaacacggc	4980
actggtcggc gtgcaggtag actcggagca gttcggcagc cagcaggtag gccgtaatta	5040
tcatctgcgc gggcgtattc tgcaggtagc gtcgaactat aaccgcgaga cgcggcaata	5100
cagcggtagc tgggacggaa cgtttaaacc ggcatcacgc aacaacatgg cctggtgtct	5160
gtgggatatg ctgaccatc cgcgctacgg catggggaaa cgtcttggtg cggcggatgt	5220
ggataaatgg gcgctgtatg tcatcgcca gtactgcgac cagtcagtgc cggacggctt	5280
tggcggcacg gagccgcgca tcacctgtaa tgcgtacctg accacacagc gtaaggcgtg	5340
ggatgtgctc agcgatttct gctcggcgat gcgctgtatg ccggtatgga acgggcagac	5400
gctgacgttc gtgcaggacc gaccgtcgga taagacgtgg acctataacc gcagtaatgt	5460
ggtgatgccg gatgatggcg cgccgttccg ctacagcttc agcgcctga aggaccgcca	5520
taatgccgtt gaggtgaact ggattgaccc gaacaacggc tgggagacgg cgacagagct	5580

tgttgaagat acgcaggcca ttgcccgtta cggtcgtaat gttacgaaga tggatgcctt	5640
tggctgtacc agccgggggc aggcacaccg cgccgggctg tggctgatta aaacagaact	5700
gctggaaacg cagaccgtgg atttcagcgt cggcgcagaa gggcttcgcc atgtaccggg	5760
cgatgttatt gaaatctgcg atgatgacta tgccgggtatc agcaccgggtg gtcgtgtgct	5820
ggcgggtgaac agccagaccc ggacgctgac gctcgaccgt gaaatcacgc tgccatcctc	5880
cgggtaccgcg ctgataagcc tggttgacgg aagtggcaat ccggtcagcg tggaggttca	5940
gtccgtcacc gacggcgtga aggtaaaagt gagccgtgtt cctgacgggtg ttgctgaata	6000
cagcgtatgg gagctgaagc tgccgacgct gcgccagcga ctgttccgct gcgtgagtat	6060
ccgtgagaac gacgacggca cgtatgccat caccgccgtg cagcatgtgc cggaaaaaga	6120
ggccatcgtg gataacgggg cgcactttga cggcgaacag agtggcacgg tgaatggtgt	6180
cacgccgcca gcggtgcagc acctgaccgc agaagtcact gcagacagcg gggaatatca	6240
ggtgctggcg cgatgggaca caccgaaggt ggtgaagggc gtgagtttcc tgctccgtct	6300
gaccgtaaca gcggacgacg gcagtgcgag gctggtcagc acggcccgga cgacggaaac	6360
cacataccgc ttcacgcaac tggcgctggg gaactacagg ctgacagtcc gggcggtaaa	6420
tgcgtggggg cagcagggcg atccggcgtc ggtatcgttc cggattgccg caccggcagc	6480
accgtcgagg attgagctga cgccgggcta ttttcagata accgccacgc cgcatcttgc	6540
cgtttatgac ccgacgggtac agtttgagtt ctggttctcg gaaaagcaga ttgcggatat	6600
cagacaggtt gaaaccagca cgcgttatct tggtagggcg ctgtactgga tagccgccag	6660
tatcaatatc aaaccgggcc atgattatta cttttatatc gcagtggtga acaccgttgg	6720
caaatcggca ttcgtggagg ccgtcggtcg ggcgagcgat gatgcggaag gttacctgga	6780
ttttttcaaa ggcaagataa ccgaatcca tctcggcaag gagctgctgg aaaaagtcga	6840
gctgacggag gataacgcca gcagactgga ggagttttcg aaagagtgga aggatgccag	6900
tgataagtgg aatgccatgt gggctgtcaa aattgagcag accaaagacg gcaaacatta	6960
tgtcgcgggt attggcctca gcatggagga cacggaggaa ggcaaactga gccagtttct	7020
ggttgccgcc aatcgtatcg catttattga cccggcaaac gggaatgaaa cgccgatgtt	7080
tgtggcgag ggcaaccaga tattcatgaa cgacgtgttc ctgaagcgcc tgacggcccc	7140
caccattacc agcggcggca atcctccggc cttttccctg acaccggacg gaaagctgac	7200

cgctaaaaat gcgatatca gtggcagtgt gaatgcgaac tccgggacgc tcagtaatgt	7260
gacgatagct gaaaactgta cgataaacgg tacgctgagg gcggaaaaaa tcgtcgggga	7320
cattgtaaaag gcggcgagcg cggcctttcc gcgccagcgt gaaagcagtg tggactggcc	7380
gtcaggtacc cgtactgtca ccgtgaccga tgaccatcct tttgatcgcc agatagtgg	7440
gcttccgctg acgtttcgcg gaagtaagcg tactgtcagc ggaggagcaa cgtattcgat	7500
gtgttatctg aaagtactga tgaacggcgc ggtgatttat gatggcgagg cgaacgaggc	7560
ggtagcagtg ttctcccgta ttgttgacat gccagcgggt cggggaaaacg tgatcctgac	7620
gttcacgctt acgtccacac ggcattcggc agatattccg ccgtatacgt ttgccagcga	7680
tgtgcaggtt atggtgatta agaaacaggc gctgggcacg agcgtggctt gagtgtgtta	7740
cagaggctcg tccgggaacg ggcgttttat tataaaacag tgagagggtga acgatgcgta	7800
atgtgtgtat tgccgttgct gtctttgccg cacttgccgt gacagtcact ccggcccgtg	7860
cgggaagggtg acatgggtacg ttacgggtgg gctattttca agtgaaaccg ggtacattgc	7920
cgctcgttgc gggcggggat accggtgtga gtcactgaa agggattaac gtgaagtacc	7980
gttatgagct gacggacagt gtgggggtga tggcttcctt ggggttcgcc gcgtcgaaaa	8040
agagcagcac agtgatTTTT ttctagattt ttagaatttt taagcttgct tggcgtaatc	8100
atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg	8160
agccggaagc ataaagtgta aagcctgggg tgcctaata gaagagctaac tcacattaat	8220
tgcgttgccg tcaactgccc ctttcagtc gggaaacctg tcgtgccagc tgcattaatg	8280
aatcggccaa cgcgcgggga gaggcgggtt gcgtattggg cgcactaccg cttcctcgct	8340
cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc	8400
ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg	8460
ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg	8520
ccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg	8580
actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac	8640
cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca	8700
tagctcacgc tgtaggtatc tcagttcggg ttaggtcgtt cgctccaagc tgggctgtgt	8760
gcacgaaccc ccgcttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc	8820

caacccggtgta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag	8880
agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac	8940
tagaagaaca gtatttggtg tctgctctct gctgaagcca gttaccttcg gaaaaagagt	9000
tggtagctct tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa	9060
gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat ctttgatct tttctacggg	9120
gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga gattatcaaa	9180
aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa tctaaagtat	9240
atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac ctatctcagc	9300
gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga taactacgat	9360
acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgtgaac cacgctcacc	9420
ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgggtcc	9480
tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta gagtaagtag	9540
ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgtcacg	9600
ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc gagttacatg	9660
atcccccatg ttgtgcaaaa aagcggttag ctcttcgggt cctccgatcg ttgtcagaag	9720
taagtgggcc gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt	9780
catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt cattctgaga	9840
atagtgtatg cggcgaccga gttgctcttg cccggcgta atacgggata ataccgcgcc	9900
acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc	9960
aaggatctta ccgctgttga gatccagttc gatgtaacct actcgtgcac ccaactgatc	10020
ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc	10080
cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct tcctttttca	10140
atattattga agcatttatc agggttattg tctcatgagc ggatacatat ttgaatgtat	10200
ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtc cacctgacgt	10260
ctaagaaacc attattatca tgacattaac ctataaaaat aggcgtatca cgaggccctt	10320
tcgtctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc tcccggagac	10380
ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg gcgcgtcagc	10440

gggtgttggc gggtgtcggg gctggcttaa ctatgcggca tcagagcaga ttgtactgag	10500
agtgcaccat atgcggtgtg aaataccgca caggtctcta agctctgtct cttatacaca	10560
tct	10563

<210> 18
 <211> 83
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 18	
tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tcgctgctcc	60
acaggtctca gcttgagcag cga	83

<210> 19
 <211> 33
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 19	
tcgctgctca agctgagacc tgtggagcag cga	33

<210> 20
 <211> 90
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 20	
tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt acagaatagg	60
gctaacaac aagaacata aacagaatag	90

<210> 21
 <211> 40
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 21	
ctattctgtt tatgtttctt gtttgtagc cctattctgt	40

<210> 22
 <211> 40
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 22	
acagaatagg gctaacaac aagaacata aacagaatag	40

<210> 23
 <211> 90
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 23
 ctattctgtt tatgtttctt gtttgtagc cctattctgt aaaaaaaaaa aaaaaaaaaa 60
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 90

<210> 24
 <211> 142
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 24
 atcctttttt ttcagagatt ttttttcaga gatttttttt cagagathtt ttttaatgta 60
 cttcgttcag ttacgtattg cttttttttt aatgtacttc gttcagttac gtattgcttt 120
 tttttttttt tttttttttt tt 142

<210> 25
 <211> 1335
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 25
 atgtctctgg ttttcgaaga cctgaaacag ggtagcgtg aagcgttcaa ccgtatcatc 60
 gaagtgttta aaaaacgttc tggtagcgt atcacctga acggtccggc tggtagcgtt 120
 aaaaccacc tgaccaaatt catcatcgac cacctggttc gtaacggtat cctgggtggt 180
 gttctggctg ctccgacca ccaggctaaa aaagtctgt ctaaactgtc tggtagtgaa 240
 gctaacacca tccaccgtat cctgaaaatc aaccgaaca cctacgaatg ccaggacatc 300
 ttcgaacagc gtgaaatgcc ggacctgtct aaatgcaac ttctgatctg cgacgaagct 360
 tctatgtacg gtgacaaact gttcggtatc atcctgcgtt ctgttccgtc ttgggcggtt 420
 atcatcggtg tcggtgaccg tgaacagctg ccgccggtt aaccgggttc tgacggtcag 480
 accctgatct ctccgttctt caccacccg tctttcgaac agctgtacct gaccgaagtt 540
 gttcgttcta acacccgat catcgacgtt gctaccgaaa tccgtatggg ttcttggtg 600
 cgtgaaaaca tcgttgacgg tcacggtggt cacgaattta actcttctac cgctctgaaa 660

gactacatga ccgaatactt caacgttggt aaagacgctg acgacctgat cgaaacccgt	720
atgctggctt tcaccaacaa atctgttgac aaactgaact ctatcatccg tcgtcgtctg	780
tacgaaaccg aaacctcttt catcaaagac gaaatcatcg ttatgcagga accgatgatc	840
aaagaactgg aatttgacgg taaaaaattc tctgaaacca tcttcaacaa cggtcagctg	900
gttcgtatca aagacgctat gctgacctct ggtttcctgt ctgctcgtaa cgtttctacc	960
cgtcagatga tcaactactg gtctctggaa gttgaaaccg ctgaagacga cgaagaatac	1020
cgtgttgacg ttatcaaatt cctgccggct gaccagggtg aaaaattcaa ctacttcctg	1080
gctaaaacct gcaccaccta ccgtgaaatg aaaaacgctg gtaaaaaagc tccgtgggaa	1140
gacttctgga aagctaaacg taccttcctg aaagttcgtg ctctgccggg ttctaccatc	1200
cacaaagctc aggggtgtttc tgtaaacgt tctttcctgt acaccccgct catccacatc	1260
gctgaagctc agctggctaa acagctggct tacgttggtg ttacccgtgc tcgtcatgac	1320
gtatactacg tataa	1335

<210> 26
 <211> 1329
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 26	
atgaccgacc tgcagtttga cgatctgacc gagggtcagc aaaacgcgtt caacgcggcg	60
ctggaagcga tgaagaccaa aggtcaacac atcaccatta acggtccggc gggcaccggc	120
aagaccacc tgaccaaatt tctgatcaac cacctgattc gtaccggcga gagcggatc	180
atgctggcgg cgccgaccca ccaggcgaag aaagtgtga gcaagctggc gggcatggaa	240
gcgcaaacca tccacagcct gctgaaaatt aaccgacca cctacgagtg cgcgaccctg	300
tttgaacaga gcgacgtgcc ggatctgagc gagtgccgtg tgctgatctg cgacgaagtt	360
agcatgtacg atcgtgaact gttccgtatt ctgatggcga gcgttccgta ttggtgcacc	420
atcattggtc tgggcgatat cgcgcagatt cgtccggtgg cgccgaacag caacatccc	480
gaagtgagcg cgttctttct gaacgagaag tttgaacaag tggcgtgac cgaagttatg	540
cgtagcaacg cgccgatcat tgaggtggcg accgaaatcc gtcacggtaa atggattcgt	600
gagtgccctgc tgaacggtga aggcgtgcac gacatggttc tgccgaccgg tggcagcgtg	660
gcgaacttca tgtacaagta cttcgacatc gttaagaccc cggaggacct gttcgaaaac	720

cgtatgctgg cgttcaccaa caagagcgtg ggcaacctga acaagatcat tcgtcgtaaa	780
ctgtaccaga ccgaggttcc gttcattaac gacgaggtgc tggttatgca agaaccgctg	840
atgcgtaccc acaagttcga gggcaagagc ttcaccgatg ttcgtttcaa caacggcgaa	900
ctggtgcgtg ttctgagctg ccagccgatc accaagcgtc tggcgatccg tggatttgac	960
caagaagatg tggttaaatg ctggcacctg gagctgcgtg cgatcgaaac cgacgtggtt	1020
gatagcatct gcgtgattga ggatgaacgt cagatgaaaa tttttcaaca ctacctgagc	1080
gcggtttgct atgagttcaa gaacagcaac accggtaaac gtccgaactg gagcggttgg	1140
tgggacctgc gtaaggaatt tcacaagggtg aaagcgtgc cgtgcggcac catccacaaa	1200
agccagggca ccagcgtgga taacgttttc ctgtataccc cgtgcattca cagcgcggac	1260
gcggatctgg cgcagcaact gctgtacgtt ggtgcgaccc gtgcgcgtaa caacgtgtat	1320
ttcgttttaa	1329

<210> 27
 <211> 1362
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 27	
atgagcgagg cggaaagcct gctggcgaag atcgttctga ccgactgcc aaaaaacgcg	60
attggtatgg tgctgagcga taagagccac atcaccatta gcggtccggc gggtagcggc	120
aagagcttcc tgacaaaaat cctgattaag aaactgctgg aactgaacaa cggtagcgtg	180
gttgctgctg cgccgaccca ccaggcgaag atcgttctga gaaaaatgag cggcatgacc	240
gcggcgacca ttcacagcat cctgaagatt caccgggaca cctacgagtg cgtgcgtgag	300
ttcaagcaaa gcaaaagcga caaggcgaag gaggatctga aagaagttcg ttacctgctg	360
gttgacgaag gtagcatggt tgacaacgac ctgttcgaga tcctgctgaa gagcgtgcac	420
ccgtactgcc agatcattgc gatcggtgac aaacaccaga ttcaaccggt tcgtcacgcg	480
ccgggcgaaa ttagcccggt ctttaccgat aagcgtttcc gtctggcgga gctgaaaacc	540
atcgttcgtc agcaagcggg caaccgatc attcaggtgg cgaccaagat tcgtaacggt	600
ggctggtttg agaccaactg ggacaaagaa agcggcaccg gtgtgctgga cgtaagagc	660
gtggcgaacc tgatgaaaat ctatctgagc aaggtgaaaa ccccgacga tctgctgaac	720

taccgtatgc tggcgtatac caacgacgtg gttaaccgtt tcaacaaggc gatccgtaaa	780
caggttttaca acaccaccga gccgtttgtg gataacgaat atctggttat gcaagagccg	840
gtgatgaaag agagcgaaat tggcggcgag accttcaccg aaaccctgct gaacaacggt	900
gaaaccgtta agatcaaaga gggcagcatt aagcgtcaga tgaaatacat cagcctgccg	960
tatgtggacc cgatccaaat tgaaatcgcg accatgaccg ttatccgtaa cgaggtggat	1020
ctgaccgaaa ttgacggtga cctggaagtg gaactgagcg tggtttgga cgcggtggc	1080
caggtgcaac tggacgaagc gctgagctac tgcgcgagcc agtataagca aatgggtagc	1140
ggcaaagcga ccagccgtct gtgggagagc ttctggcagg ttaaggcat gtttaccac	1200
acaaaaagcc tgggcgcgag cacctttcac aagagccaag gcaccaccgt gatcggcgtg	1260
tgcgtttaca ccggtgatat gaacttcgcg cagtttgaga ttcagacca actgggttac	1320
gttggctgca cccgtgcgca aaaatgggtg atgtattgct aa	1362

<210> 28
 <211> 1365
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 28	
atgagcgagg cggcgagcct gctggcgaaa atcattctga ccgactgccga gaagaccgcg	60
atcgacgcgg tgctgaccga taagaaacac atcaccatta gcggtccggc gggtagcggc	120
aagagcttcc tgacaaaaat cctgattcag aagctgctgg atctgaacag cggcgcggtg	180
atcacctgcg cgccgacca ccaagcgaaa attgttctga gcaagatgag cggcttcacc	240
gcgagcacca tccacagcgt gctgaaaatt caccgggaca cctacgaatg cgttcgtgag	300
tttaagcaga gcaaaagcga caaggcgaaa gaagatctga aggcgggtgcg ttatctgatc	360
gttgacgagg cgagcatggt ggacaacgac ctgttcgaaa ttctgctgaa aagcgttcac	420
ccgttttgcc aaatcattgc gatcggtgac aagcaccaga ttcagccggt gcgtcatgcg	480
ccgggtgaaa tcagcccgtt ctttaccgat aaacgtttcc gtctggcgga actgaagacc	540
gtggttcgtc agcaagcggg caaccgatc attcaggttg cgacaaaaat tcgtaacggt	600
ggctggtttg agaccaactg ggacaaggcg accggcaccg gtgtgctgga cgtaagacc	660
atcgcgaaaa tgatgcaaat ttacctgagc aaggtgaaaa ccccggaaga cctgctgaac	720
taccgtatgc tggcgtatac caacgatgtg gttaacagct tcaaccgtgt gatccgtaag	780

cacgtttaca aaaccaccga accgtttggt gataacgagt atctggtgat gcaggaaccg	840
gttatgcgtg aggaagagat tggtagcgaa accttcaccg agaccctgct gaacaacggc	900
gagaccgtga aaatcattcc gggcagcatc aagcgtcaac tgaaatacat tagcctgccg	960
tatgttgaac cgatccagat tgaggtggcg accatgctgg ttgaacgtca agagaccgac	1020
gtgaccgata acgttgacag cgataaggaa gtggagatca gcgtggtttg ggacgcgagc	1080
agccagggtc tgctggatga ggcgctgagc tactgcgcga gccagtataa acaaatgggt	1140
agcggcaagg cgaccagccg tctgtgggag agcttctggc aggtgaaagg catgtttgtt	1200
aacaccaaga gcctgggcgc gagcaccttt caaaaagcc aaggcaccac cgttatcggc	1260
gtgtgcgttt ataccggtga catgaacttc gcgcagtttg aaattcagac ccaactgggt	1320
tacgtgggct gcacccgtgc gcaaaagtgg gttatctatt gctaa	1365

<210> 29
 <211> 1326
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 29	
atgaacttcg acgacctgac cgttggtcag aaagaagcgt tcgaaatcgt tatcgaagct	60
atccgtacca aaaaacacca cgttatgata aacggctcgg ctggtaccgg taaaaccacc	120
atgaccaaatt tcatcctgga acacctggtt cgtaacggtg aactgggtat catgctggct	180
gctccgaccc accaggctaa aaaagtcttg tctaaactgt ctggtcacca ggctgctacc	240
atccactcta tcctgaaaat ctctccgacc acctacgaat gcgaatctat cttcgaacag	300
aaagaaatgc cggacctggc taaatgccgt gttctgctgt gcgacgaagg ttctatgtac	360
gacggtgctc tgttcaaaat cctgatgaac accatcccgt ctcacgcgac cgttatcggc	420
atcggtgacg aagaacagct gcgtccggtg agcccgggtg actctctgcc gtctaaatct	480
ccgttcttct ctgaccaccg tttcaaacag gttaccctga ccgaagttaa acgttctaac	540
ggtccgatca tcaaagttgc taccgaaatc cgtaacggtg gttggttccg tgaatgcatc	600
gaagacggtc acggtttcca cggtttcaac ggtgacaaac cgctgcagca gtacatgatg	660
aaatacttcg acgttggtta atctccggaa gacctgttcg aaaccgtat gctggcttac	720
accaacaaat ctgttgacaa actgaacggt atcatccgtc gtaaactgta cgaaaccgaa	780

tctccgttca tcgttggtga agttctgggt atgcaggaac cgtacatgaa atctctggaa	840
tttgacggta aaaaattcaa cgaaatcatc ttcaacaacg gtcagatggg tcgtatcctg	900
gactgcaaac tgacctctac cttcctgaaa gtcctgtgacg tttctgttaa acagatgatc	960
tcttactggc acctggaagt tgaaaccggt gacgaagacg acgactacca gcgtgaaacc	1020
atcaaagttc tggctgacga caacgaaaaa cagaaattcg acatgttcct ggctaaagtt	1080
tgacaccttt accgtgaact gaaatctgct ggctgtcgtc cgactggggc tgacttctgg	1140
gacgctaaac gtaccttcct gaaagttaaa gctctgccgt gctctaccat ccacaaatct	1200
cagggtatct ctgttgacaa cgctttcatc tacaccccggt gcatcaccct ggctgacatc	1260
gacctgggta aacagctggc ttacgtttct tctaccctg ctcgtcatga cgtgtacttc	1320
gtctaa	1326

<210> 30
 <211> 1341
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 30	
atgtctcagg aagttacctt cgaatctctg aacaaagggtc agcgtgaagc gttcgacatc	60
atcacctctg ctatccagcg tcgtaacggg gaacgtctga ccctgaacgg tccggctggg	120
accggtaaaa ccaccctgac caaattcatc atccagcaca tcgttcgtaa cgggtgttctg	180
gggtgtgttc tggctgctcc gaccaccag gctaaaaaag ttctggctaa aatgtctggg	240
atggaagcta acaccatcca ccgtgttctg aaaatcaacc cgatgaccta cgaatgccag	300
gacgttttcg aacagcgtga aatgccggac atgtctaaat gcaacgttct ggtttgcgac	360
gaagcgtcca tgctggacgg taaaatcttc aaaatcatcc tgaactctat cccgccgtgg	420
gcggttctga tcggtatcgg tgaccgtgaa cagatccagc cggttgaacc gggttctgac	480
ggtacccgc agatctctcc gttcttcacc caccgtctt tcaaacagggt tcacctgacc	540
gaagttatgc gttctaacgc tccgatcatc gacgttgcta ccgacatccg taccgggtgg	600
tggctgcgtc accacatcat cgacggtcac ggtgttcacg aatttgcttc taccaccgct	660
ctgaaagact tcatgatgca gtatttcgac gttgttaaaa ccccggaaga cctgttcgaa	720
accggtatgc tggctttcac caacaaatct gttgaaaaac tgaacaacat catccgtcgt	780
aaactgtacg aaaccgaagt tccgttcacg aacgaagaag ttatcggtat gcaggaaccg	840

ttcatcaaag aactggaatt tgacggtaaa aaattctctg aaatcgtttt caacaacggt	900
gaaatgggtc gtatcaaaga ctgcatgctg acctctatgc cgctgatcgc tcgtaacggt	960
tctaccaaac agcacatcaa ctactgggct ctggaagttg aaaccatcga cccggacgaa	1020
gaatacaaaa tcgaagttat caaagtcttg ccgctggacc agtaccagaa aatggacatg	1080
ttcctggcta aagtttgac cacctaccgt gaaatgaaag ctgctggtaa acgtccgccg	1140
tgggacgact tctggaaaat caaacgtacc ttcctgaaag ttcgtgctct gccggtttct	1200
accatccaca aatctcaggg tatctctggt aacaactctt tcattctacac cccgtgcatc	1260
cacgttgctg aagttcagct ggctcgtcag ctggcttacg ttggtctgac ccgtgctcgt	1320
cacgacgctt actacgttta a	1341

<210> 31
 <211> 1359
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 31	
atgaacgacc tgaaactgga agacctgaac gaaggtcagc gtgctgcttt cgacgctgtt	60
atcaaagtta tctctaccaa atctatgatg gctaccacct cttcttgca agttaaacg	120
cacatcacca tcaacgggtc ggctgggtacc ggtaaaacta ctctgacccg tttcctgatc	180
aaccatctga tctcttctgg tgaaaaaggt gttatgctgg ctgctccgac ccaccaggct	240
aaaaaagttc tggctaaact gtctggatat gaagcgtcta ccatccactc tctgctgaaa	300
atcaaccgga ccacctacga atgctctacc gttttccagc agtctgacga cccggacctg	360
tctgactgcc gtgttctgat ctgcgacgaa gtttctatgt acgaccgtga actgttccgt	420
atcctgatgg cttctatccc gccgtgggcg accatcatcg gtctgggtga catcgctcag	480
atccgtccgg ttgctccgaa ctctaccacc ccggaactgt ctgctttctt cttcaacgac	540
aaattccagc aggtttctct gaccgaagtt atgcgttcta acgctccgat catcgaagtt	600
gctaccgaaa tccgtaaagg tggttggatt cgtgaaaacc tggttgacgg tcagggtgtt	660
cactctatgg ttcgttctaa cggtggttct gttgctgctt tcctgaccaa atacttcgaa	720
atcgttaaag acccggacga cctgttcgac aaccgtatgc tggctttcac caacaaatct	780
gttaacgacc tgaacaacat catccgtaaa aaactgtacc agaccaccgt tccgtacatc	840

aaagacgaag ttctggttat gcaggaaccg ctgatgcgtt ctcacacctt cgaaggtaaa	900
accttcaccg aagttatctt caacaacggg gaactgggtc gtatcatcaa ctgccgtgaa	960
aaatacgtta acctgttcat caaagggttc aaagggttctg actctatcaa agtttgggaa	1020
ctggaaatcc gtggtgttga ctctgacgct gttgacatga tcaaagttat ccacgacgaa	1080
caggaaactga acaaattcca gtacttcatg tctaaatctt gctctgaatt taaaaacgct	1140
cgtgacaaac gtccgaactg gaaagggttg tgggacctga aagctcagtt ccacaaagtt	1200
aaaccgctgc cgtgcggtac catccacaaa tctcagggtt ctaccctgga caacgttttc	1260
ctgttcaccc cgtgcatcca ccgtgctgac ccggctctgg ctcagcagct gctgtacgtt	1320
ggtgctaccc gtgctaaaca caacgtttac ttcgtttaa	1359

<210> 32
 <211> 1365
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 32	
atggttatca ccttcgacga cctgaccgaa gggtcagaaac tggctttcaa cgaagttggt	60
gacgctatca aaaccagaa cctgaacctg cgtgctgaca caaaaccca catcaccatc	120
aacgggtgaag ctggtaccgg taaaaccacc ctgaccaaat tcctgatcga ctacatcatc	180
aaagaaggta tcaacggtgt tatcctggct gtcctgaccc acgtgctaa aaccgttctg	240
tctaaactgt ctggtatgga agctgaaacc atccactctg ttctgaaaat ctctccgacc	300
aactacgaat gccagaccgt tttcgaacag cgtgaaatcc cgaacctggc tgaatgccgt	360
atcctgatct gcgacgaagg ttctatgtac gaccgtaaac tggttcagct gatcctgaac	420
accgttccga aatgggcgct gggtatcgtt ctgggtgaca aagaacagat ccgtccgggt	480
tctccgggtg aaaccctgcc gggatatctt ccgttcttca cccacaaaaa attcaaacag	540
atcaaaactga ccgaagttaa acgttctaac ggtccgatca tcaccttgct tcgtgaaatc	600
ctgaaaggct agtggctgct tgaatgcctg gacgaagacg gtcagggtgt tcacgcttac	660
gacccggaat ctgacatccc gtctctgcac tggttcctga aagaatactt caaagttggt	720
aaaaccaaag aagacttcgt taacaccgtt gttatggctt acaccaacaa agttgttaac	780
accctgaaca aaatcatccg taaacgtatc ttcaacaccg acgaaccgtt catcgaagac	840
gaaatcatcg ttatgcaggg tccgctgacc gaatctctga tcgttgacgg taaaaaagtt	900

aaaaaactga tctacaacaa cggtcagcgt gttcgtatcg ttcgtgttaa caaaaccgtt	960
cacaccctgc gtgctcgttt cgttgaatct accaaagaaa tcgacgtttg gaccctgacc	1020
gttgaaaccg ctgacaaaaa catcgacgaa taccacctga aagacctgca catcgttgac	1080
gaaggttctg aactggttct gaaagaattc ctgtctgaaa cctgcaacac ctaccgttac	1140
tgggaaactgc cgggtaaagc tccgtggggg gaattctgga ccatcaaaga acgttactct	1200
aacgttaaag ctgaaccgtg ctctaccatc cacaagctc agggatatctc tgttgacaac	1260
gctttcctgt gcacctctgg tctgtcttct atggaccggg acctgggttaa agaactgatc	1320
tacgttggtt ctaccctgcc gaaacacaac ctgtactgga tctaa	1365

<210> 33
 <211> 1326
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 33	
atgaaattcg aagacctgac cgttggtcag aaatctgctt tcgacgttgt tatcgaagct	60
atcaaaacca aaaaattcca cgttatgatc aacgggccgg ctggtaccgg taaaaccacc	120
atgaccaaatt catcctgga acacctgggt cgtaacgggtg aactgggtat catgctggct	180
gctccgaccc accaggctaa aaaagtcttg tctaaactgt ctggtcacga agctgctacc	240
atccactctg ttctgaaaat ctctccgacc acctacgaat gcgaatctat cttcgaacag	300
aaagaaatgc cggacctggc taaatgccgt gttctgttct gcgacgaagg ttctatgtac	360
gacggtgctc tgttcaaaat cctgatgaac accatcccgt ctacgcgac cgttatcggg	420
atcggtgacg aagaacagct gcgtccggta agcccgggtg actctctgcc gtctaaatct	480
ccgttcttct ctgactctcg tttcaaacag gttaccctga ccgaagttaa acgttctaac	540
ggtccgatca tcaaagttgc taccgaaatc cgtaccgggtg gttggttccg tgaatgcac	600
gaagacggtc acggtttcca cggtttcggg ggtgacaaac cgctgcagca gtacatgatg	660
aaatacttctg acgttggttaa atctccggaa gacctgttcg aaaccgtat gctggcttac	720
accaacaaat ctgttgacaa actgaacggg atcatccgtc gtaaactgta cgaaaccgaa	780
aaccgttca tcgttggtga agttctgggt atgcaggaac cgtacatgaa acagctggaa	840
tttgacggta aaaaattcaa cgaaatcatc ttcaacaacg gtcagatgat ccgtatcctg	900

gactgcaaac tgacctctac cttcctgaaa gctcgtgacg tttctgttaa acagatgac	960
tcttactggc acctggaagt tgaaaccgtt gacgaagacg acgactacca gcgtgaaacc	1020
atcaaagttc tggctgacgc taacgaaaaa cagaaattcg acatgttcct ggctaaagtt	1080
tgcaccacct accgtgaact gaaatctgct ggctcgtcgtc cgcactgggc tgacttctgg	1140
gacgctaaac gtaccttcct gaaagttaaa gctctgccgt gctctaccat ccacaaatct	1200
cagggtatct ctgttgacaa caccttcac tacacccgt gcatcacct ggctgacatc	1260
gacctggcta aacagctggc ttacgtttct gctaccgtg ctcgtcatga cgtgtacttc	1320
gtctaa	1326

<210> 34
 <211> 1332
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 34	
atgaccatcg gtttcgacga cctgaccgaa ggtcagaaat gcgctttcga aaccgctgtt	60
gaactggtta actctaaacg taaacacatg accctgaacg gtccggctgg ttctggtaaa	120
accacttggga ccaggttctt catcgaccat ctggttcgtt ctggtgaatc tgggtgtatc	180
ctggctgctc cgaccaccca ggctaaaaaa gttctgtcta aactgtctgg tgttgaagct	240
tctaccatcc actctatcct gaaaatcaac ccgaccacct acgaatgcaa cgttctgttc	300
gaacagaaag aaatcccgga cctggctaaa tgccgtgttc tgatctgcga cgaagcttct	360
atgtacgacc gtaaactgtt cgacatcctg atgaactcta tccgtcttg ggcgatcgtt	420
atcgctctgg gtgacaaaga ccagctgcgt ccggttgaac tgaactctga aggtaaaggt	480
cagatctctg ctttcttcta cgaccgcgt ttcgaacagg tttcctgtc tgaaatcaaa	540
cgttctaact ctccgatcat cgaagtgtc acctctatcc gtaccggtgg ttggctgtac	600
cacaacctgg gtgacgacgg taccggtgtt cacggttaca tgaacaaagg ttctgctctg	660
aaagacttct tcggtcagta cttcgacacc gttcgtaaac cggaagacct gttcgaaaac	720
cgtatgtgcg cttacaccaa cgaatctgtt aacaaactga actctatcat ccgtcgtaaa	780
atctacgaca ccgaagacct gttcgttgtt aacgaagttc tggttatgca ggaaccgctg	840
accaaagaaa tcaaattcga aggtaaacgt ttctctgaaa tgatcttcca caacggtcag	900
atggttcgtg ttgttaaagc tgaaaaaacc tctaaattcc tgcgtgctaa aggtgtttct	960

ggtgaacaga tgatccgtta ctggtctctg gttgttgaaa ccaacgacgc tgaagacgaa	1020
tacttccgtg aacagatctg cgttctgtct gacgaaaacg aaatcaacaa atactactac	1080
ttcctggcta aagtttgca cgcttacaaa tctggtgctg ttaaagctca ctgggctgac	1140
ttctgggctg ctaaactgac tttcatcaaa gttaaagctc tgccgtgctc taccatccac	1200
aaagttcagg gtatctctgt tgacaactgc ttcctgtaca cccgtgcat ccacaaagct	1260
gacgctgacc tggctaaaca gctgacctac gttggtgcta cccgtccgag tttcaacctg	1320
cactacgttt aa	1332

<210> 35
 <211> 1326
 <212> DNA/RNA
 <213> Artificial Sequence

<400> 35	
atgatcacca tcgaccagct gaccgaaggt cagttcgact ctctgcagcg tgctaaagtt	60
ctgatccagg aagctaccaa aaacgacggt aactggaacc accgtaccaa acacctgacc	120
atcaacggctc cggtcggtac cggtaaaacc accatgatga aattcctggt ttcttggtg	180
cgtgacgaag gtatcacagg tggtgctctg gctgctccga cccacgctgc taaaaaagtt	240
ctggctaacg ctgttggtga agaagtttct accatccact ctatcctgaa aatcaacccg	300
accacctacg aatgcaaaca gttcttcgaa cagtctgctc cgccggacct gtctaaaatc	360
cgtatcctga tctgcgaaga atgctctttc tacgacatca aactgttcga aatcctgatg	420
aactctatcc agccgtggac catcatcatc ggtatcggtg accgtgctca gctgcgtccg	480
gctgacgaca aaggtatctc tcgtttcttc accgaccagc gtttcgaaca gacctacctg	540
accgaaatca aacgttctaa catgccgatc atcgaagttg ctaccgaaat ccgtaacggt	600
ggttggattc gtgaaaacat catcgacgac ctgggtgtta aacaggacaa atctgtttct	660
gaatttatga ccaactactt caaagttggt aaatctatcg acgacctgta cgaaaccgtg	720
atgtacgctt acaccaacaa ctctgttgac accctgaaca aaatcatccg taaaaaactg	780
tacgaaaccg aacaggactt catcgttggt gaaccgatcg ttatgcagga accgctgatc	840
cgtgacatca actacgaagg taaacgtttc caggaaatcg ttttcaacaa cggtgaatac	900
ctggaagttt ctgaaatcaa accgatggaa tctgtttctga aatgccgtaa catcgactac	960

cagctggttc tgcactacta ccagctgaaa gttaaatacta tcgacaccgg tgaatctggt	1020
ctgatcaaca ccatctctga caaaaacgaa ctgaacaaat tctacatggt cctgggtaaa	1080
gtttgccagg actacaaatc tggtagcatc aaagcggtct gggacgactt ctggaaaatc	1140
aaaaacaact accaccgtgt taaaccgctg ccggtttcta ccatccacaa aggtcagggt	1200
tctaccgttg acaactcttt cctgtacacc ccgtgcatca ccaaatacgc tgaaccggac	1260
ctggcttctc agctgctgta cgttggtggt acccgtgctc gtcacaacgt taacttcgtt	1320
ggttaa	1326

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 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; (xxxii) E105C and/or A362C, and I280A; (xxxiii) E105C and/or L358C; (xxxiv) E104C and/or G359C; (xxxv) E104C and/or A362C; (xxxvi) K106C and/or W378C; (xxxvii) T102C and/or N382C; (xxxviii) T102C and/or W378C; (xxxix) E104C and/or Y355C; (xl) E104C and/or N382C; (xli) E104C and/or K381C; (xlii) E104C and/or K379C; (xliii) E104C and/or D376C; (xliv) E104C and/or W378C; (xlv) 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 any one of claims 1 to 3, 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 any one of claims 1 to 4, 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 any one of claims 1 to 5, wherein the Pif1-like helicase further comprises substitution of at least one or more cysteine residues; preferably the cysteine residues are substituted with alanine, serine or valine.

7. The Pif1-like helicase according to any one of claims 1 to 6, wherein the Pif1-like helicase comprises:

(a) a variant of SEQ ID NO: 5, and the one or more substituted native cysteine residues are one or more of C109, C114, C136 or C414; or,

(b) a variant of any one of SEQ ID NOs: 1, 2, 3, 4, 6, 7, 8, 9, 10, and 11, and the one or more substituted native cysteine residues correspond to one or more of C109, C114, C136 or C414 in SEQ ID NO: 5.

8. The Pif1-like helicase according to any one of claims 1 to 7, wherein the Pif1-like helicase further comprises substitution or modification of surface negatively-charged amino acids, polar or non-polar amino acids; more preferably, 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: 11, and the one or more negatively charged amino acids are one or more of D5, E9, E24, E87, I65, S58, D209 or D216; or,

b) a variant of any one of SEQ ID NOs: 1 to 10, and the one or more negatively charged amino acids correspond to one or more of D5, E9, E24, E87, I65, S58, D209, or D216 in SEQ ID NO: 11.

10. The Pif1-like helicase according to claim 9, wherein the Pif1-like helicase further comprises:

a) a variant of SEQ ID NO: 1, and the one or more negatively charged amino acids further comprise S171; or,

b) a variant of SEQ ID NO: 9, and the one or more negatively charged amino acids further comprise S173.

11. The Pif1-like helicase according to any one of claims 1 to 10, 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, (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-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, (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-[(Cyclopentylloxy)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.

12. The Pif1-like helicase according to any one of claims 1 to 11, wherein the Pif1-like helicase further comprises:

(a) substitution of at least one amino acid that interacts with one or more nucleotides in single-stranded DNA or double-stranded DNA; and/or,

(b) substitution of at least one amino acid that interacts with a transmembrane pore, wherein the Pif1-like helicase has the ability to control the movement of a polynucleotide.

13. The Pif1-like helicase according to claim 12, wherein in (a), 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.

14. The Pif1-like helicase according to claim 13, 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; or,

(b) 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 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.

15. The Pif1-like helicase according to claim 13 or 14, 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.

16. The Pif1-like helicase according to any one of claims 13 to 15, wherein the larger side chain increases (i) electrostatic interaction, (ii) hydrogen bonding and/or (iii) cation-pi interaction

between the at least one amino acid and one or more nucleotides in the single-stranded DNA or double-stranded DNA.

17. The Pif1-like helicase according to any one of claims 13 to 16, 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).

18. The Pif1-like helicase according to any one of claims 12 to 17, wherein:

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

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

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); or (iv) leucine (L), valine (V) or isoleucine (I);

D) phenylalanine (F) is substituted with (i) arginine (R) or lysine (K); (ii) histidine (H); or (iii) tyrosine (Y) or tryptophan Acid (W);

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); or (iv) isoleucine (I) or leucine (L);

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); or (iv) isoleucine (I) or leucine (L); and/or,

H) tyrosine (Y) is substituted with (i) arginine (R) or lysine (K); or (ii) tryptophan (W).

19. The Pif1-like helicase according to any one of claims 12 to 18, 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.

20. The Pif1-like helicase according to any one of claims 12 to 19, wherein, in (a), 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.

21. The Pif1-like helicase of claim 20, 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; or,

(b) a variant of any one of SEQ ID NOs: 1 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.

22. The Pif1-like helicase according to claim 21, wherein:

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

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

c) serine 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); or (iii) phenylalanine (F), tryptophan (W), tyrosine (Y) or histidine (H); and/or,

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

23. The Pif1-like helicase according to any one of claims 20 to 22, 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.

24. The Pif1-like helicase according to any one of claims 12 to 23, 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.

25. The Pif1-like helicase according to any one of claims 1-24, 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, for example, 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.

26. A construct comprising at least one Pif1-like helicase according to any one of claims 1 to 25.

27. The construct according to claim 26, wherein the construct further comprises a polynucleotide binding moiety.

28. A nucleic acid encoding the Pif1-like helicase according to any one of claims 1 to 25 or the construct according to any one of claims 26 to 27.

29. An expression vector comprising the nucleic acid according to claim 28.

30. A host cell comprising the nucleic acid according to claim 28 or the expression vector according to claim 29.

31. A method of controlling the movement of a polynucleotide, comprising contacting the Pif1-like helicase according to any one of claims 1 to 25 or the construct according to any one of claims 26 to 27 with the polynucleotide.

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

I) contacting the Pif1-like helicase according to any one of claims 1 to 25 or the construct according to any one of claims 26 to 27 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.

33. The method according to claim 32, 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; preferably, the pore is selected from a biological pore, a solid state pore, or a biological and solid state hybrid pore.

34. The method according to claim 32 or 33, wherein the target polynucleotide is single-stranded, double-stranded, or at least partially double-stranded.

35. The method according to any one of claims 32 to 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; preferably, the one or more characteristics are measured by electrical measurement and/or optical measurement.

36. A product for characterising a target polynucleotide comprising the Pif1-like helicase according to any one of claims 1 to 25, the construct according to any one of claims 26 to 27, the nucleic acid according to claim 28, the expression vector according to claim 29 or the host cell according to claim 30, and a pore.

37. The product of claim 36, wherein the product is selected from a kit, a device or a sensor.

38. Use of the Pif1-like helicase according to any one of claims 1 to 25, the construct according to any one of claims 26 to 27, the nucleic acid according to claim 28, the expression vector according to claim 29, the host cell according to claim 30 or the product according to any one of claims 36 to 37 in characterising a target polynucleotide or controlling the movement of a target polynucleotide through a pore.