US 20120100530A1

(19) **United States**

(12) **Patent Application Publication**

Moysey et al.

(10) **Pub. No.: US 2012/0100530 A1**

(43) **Pub. Date: Apr. 26, 2012**

(54) ENZYME MUTANT

(75) Inventors: **Ruth Moysey**, Oxford (GB); **Michael Knaggs**, Oxford (GB); **Lakmal Jayasinghe**, Oxford (GB); **James White**, Oxford (GB); **Brian Mckeown**, Oxon (GB); **John Milton**, Oxford (GB)

(73) Assignee: **OXFORD NANOPORE TECHNOLOGIES LIMITED**, Oxford (GB)

(21) Appl. No.: **13/147,171**

(22) PCT Filed: **Jan. 29, 2010**

(86) PCT No.: **PCT/GB10/00133**

§ 371 (c)(1),
(2), (4) Date: **Nov. 10, 2011**

Related U.S. Application Data

(60) Provisional application No. 61/148,726, filed on Jan. 30, 2009.

Publication Classification

(51) **Int. Cl.**

C12Q 1/68	(2006.01)
C12N 9/16	(2006.01)
C07H 21/04	(2006.01)
C12N 11/00	(2006.01)

(52) **U.S. Cl.** **435/6.1**; 435/174; 435/196; 536/23.2

(57) ABSTRACT

The invention relates to constructs comprising a nucleic acid binding protein and a surface. At least one native accessible cysteine residue is removed from the binding protein. The binding protein is attached to the surface via one or more accessible cysteine residues. The removal of other accessible cysteine residues from the protein allows control attachment to the surface. The constructs can be used to generate transmembrane pores having a nucleic acid binding protein attached thereto. Such pores are particularly useful for sequencing nucleic acids. The enzyme handles the nucleic acid in such a way that the pore can detect each of its component nucleotides by stochastic sensing.

Figure 1

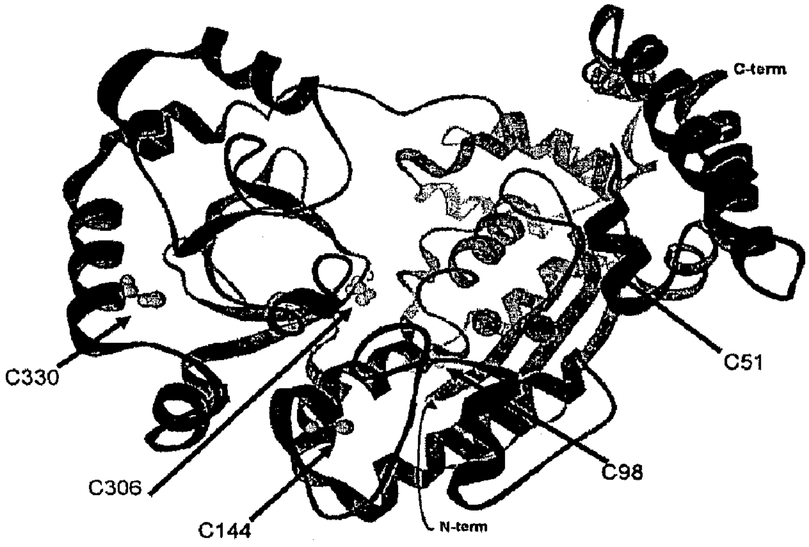


Figure 2

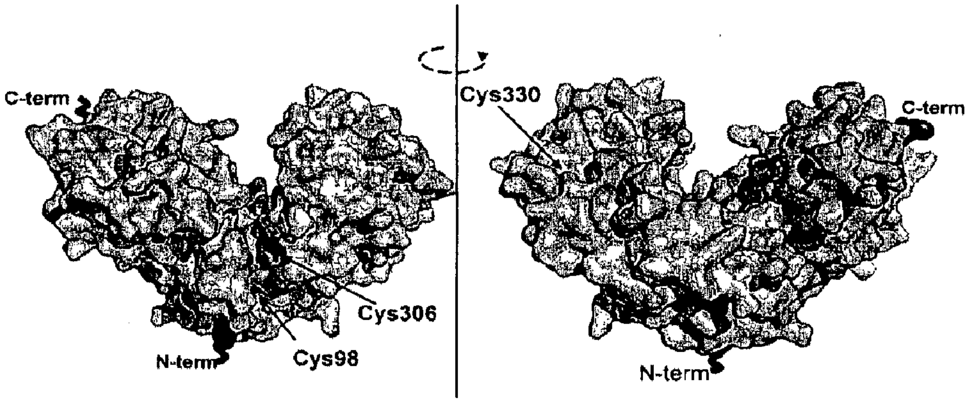


Figure 3

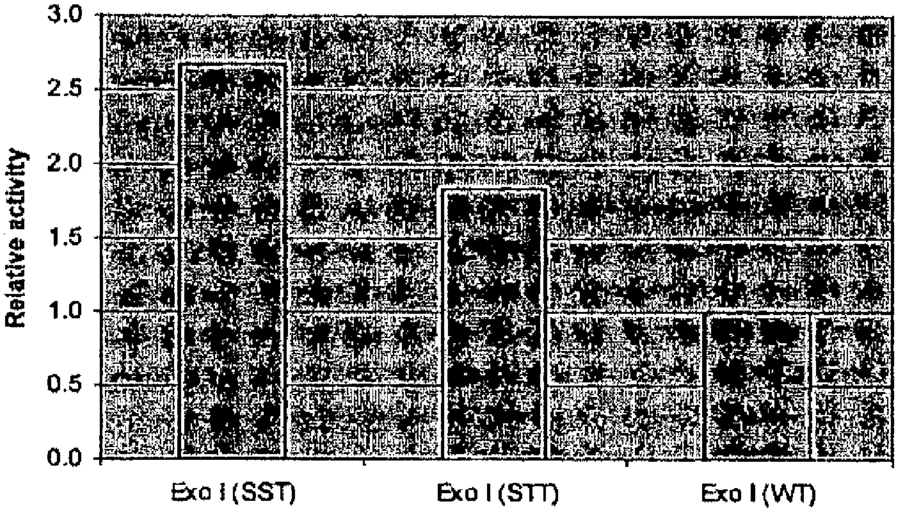


Figure 4

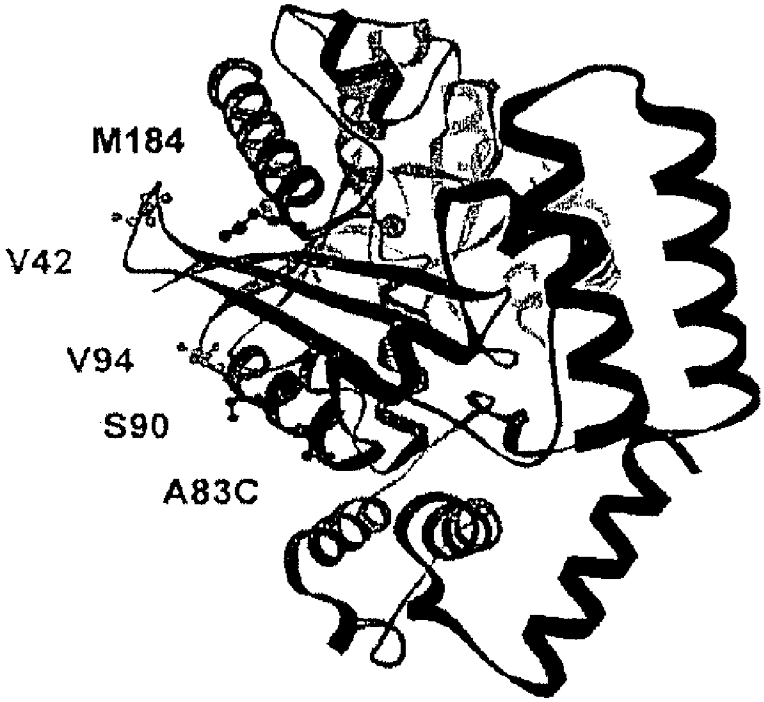


Figure 5a

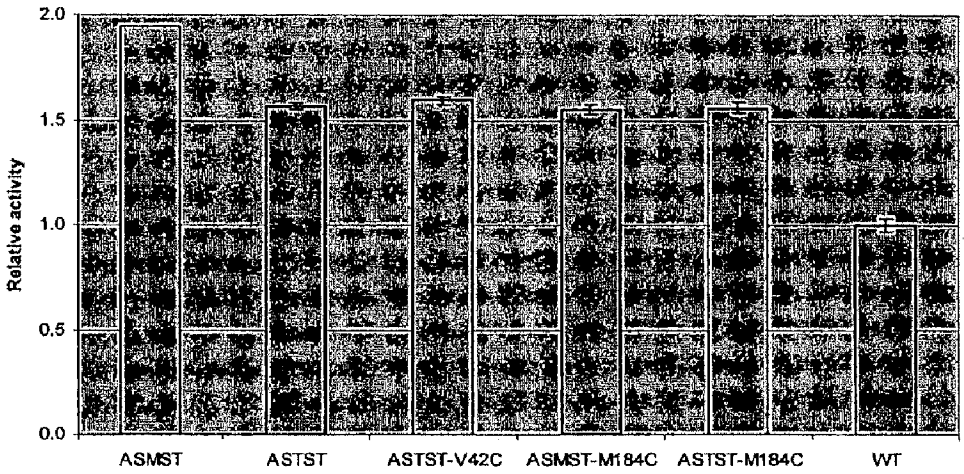


Figure 5b

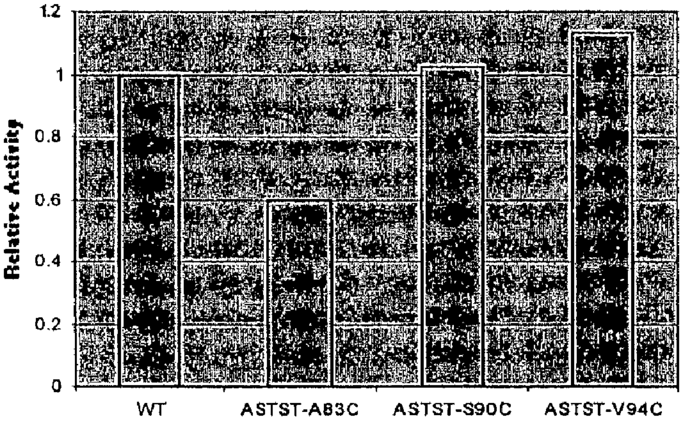


Figure 6

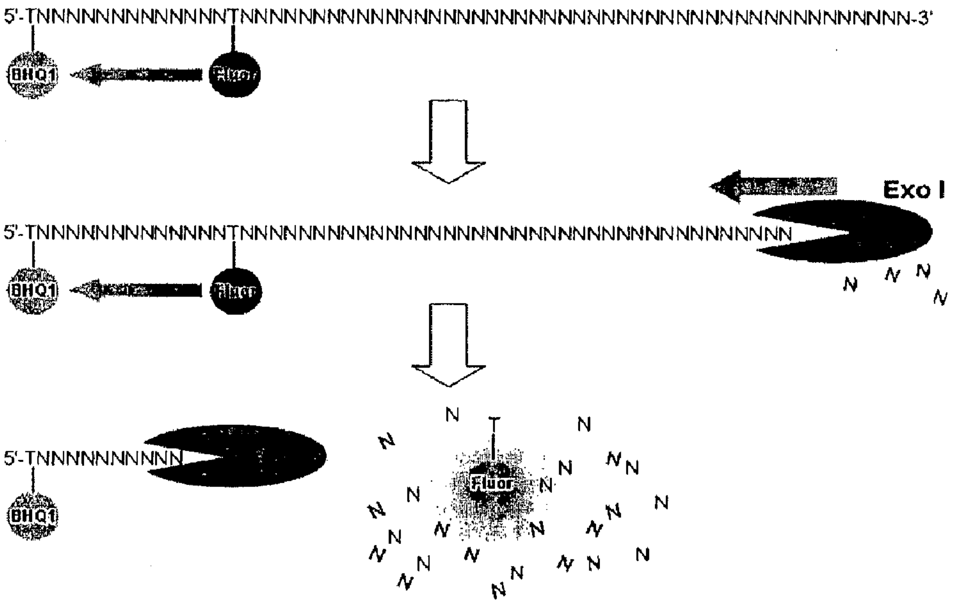


Figure 7

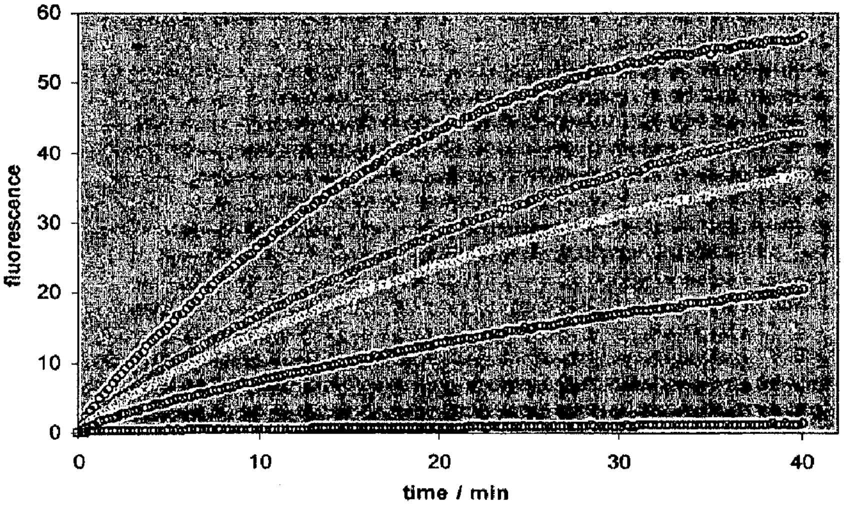


Figure 8

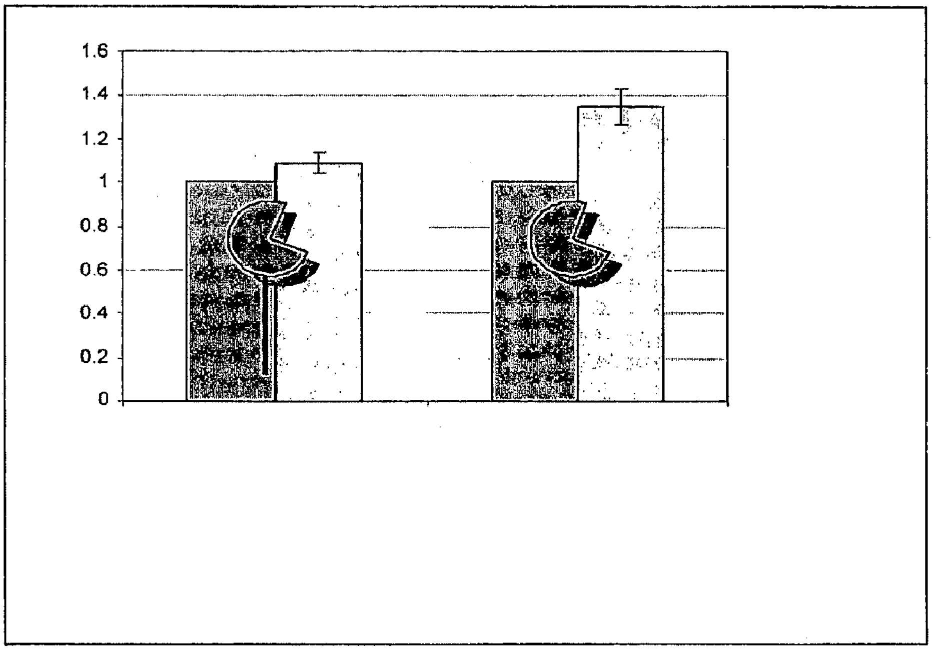


Figure 9

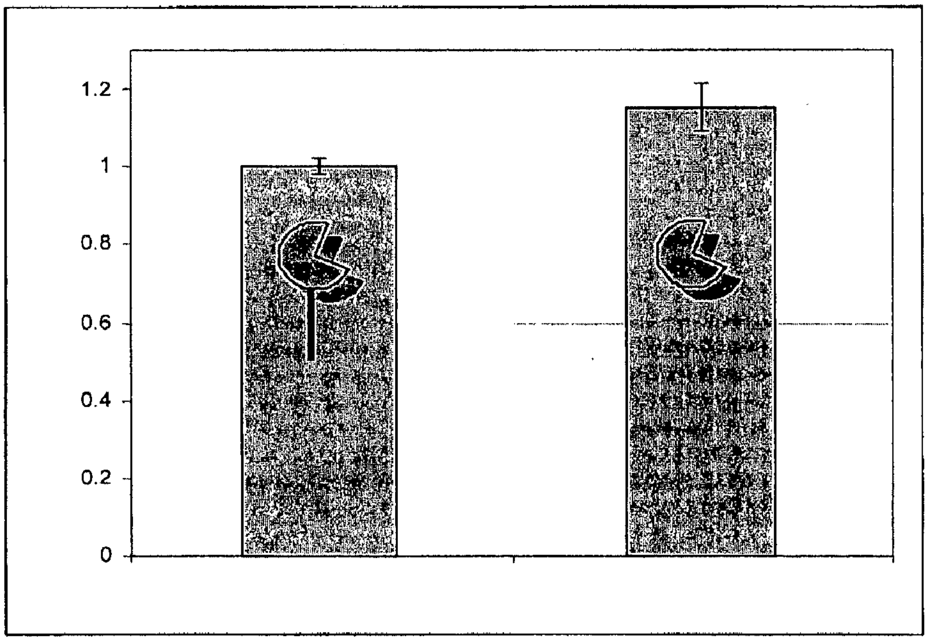
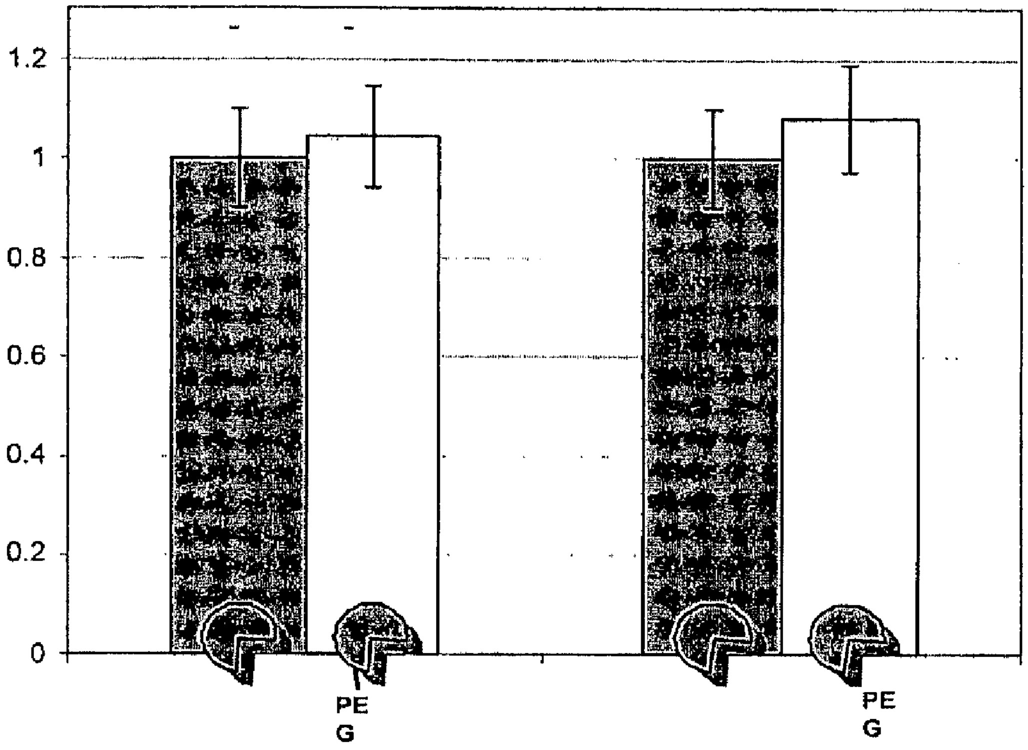


Figure 10



ENZYME MUTANT

FIELD OF THE INVENTION

[0001] The invention relates to constructs comprising a nucleic acid binding protein and a surface. At least one native accessible cysteine residue is removed from the binding protein. The binding protein is attached to the surface via one or more accessible cysteine residues. The removal of other accessible cysteine residues from the protein allows control attachment to the surface. The constructs can be used to generate transmembrane pores having a nucleic acid binding protein attached thereto. Such pores are particularly useful for sequencing nucleic acids. The enzyme handles the nucleic acid in such a way that the pore can detect each of its component nucleotides by stochastic sensing.

BACKGROUND OF THE INVENTION

[0002] Stochastic detection is an approach to sensing that relies on the observation of individual binding events between analyte molecules and a receptor. Stochastic sensors can be created by placing a single pore of nanometer dimensions in an insulating membrane and measuring voltage-driven ionic transport through the pore in the presence of analyte molecules. The frequency of occurrence of fluctuations in the current reveals the concentration of an analyte that binds within the pore. The identity of an analyte is revealed through its distinctive current signature, notably the duration and extent of current block (Braha, O., Walker, B., Cheley, S., Kasianowicz, J. J., Song, L., Gouaux, J. E., and Bayley, H. (1997) *Chem. Biol.* 4, 497-505; and Bayley, H., and Cremer, P. S. (2001) *Nature* 413, 226-230).

[0003] Engineered versions of the bacterial pore forming toxin α-hemolysin (α-HL) have been used for stochastic sensing of many classes of molecules (Bayley, H., and Cremer, P. S. (2001) *Nature* 413, 226-230; Shin, S., H., Luchian, T., Cheley, S., Braha, O., and Bayley, H. (2002) *Angew. Chem. Int. Ed.* 41, 3707-3709; and Guan, X., Gu, L.-Q., Cheley, S., Braha, O., and Bayley, H. (2005) *ChemBioChem* 6, 1875-1881). In the course of these studies, it was found that attempts to engineer α-HL to bind small organic analytes directly can prove taxing, with rare examples of success (Guan, X., Gu, L.-Q., Cheley, S., Braha, O., and Bayley, H. (2005) *ChemBioChem* 6, 1875-1881). Fortunately, a different strategy was discovered, which utilised non-covalently attached molecular adaptors, notably cyclodextrins (Gu, L.-Q., Braha, O., Conlan, S., Cheley, S., and Bayley, H. (1999) *Nature* 398, 686-690), but also cyclic peptides (Sanchez-Quesada, J., Ghadiri, M. R., Bayley, H., and Braha, O. (2000) *J Am. Chem. Soc.* 122, 11758-11766) and cucurbiturils (Braha, O., Webb, J., Gu, L.-Q., Kim, K., and Bayley, H. (2005) *ChemPhysChem* 6, 889-892). Cyclodextrins become transiently lodged in the α-HL pore and produce a substantial but incomplete channel block. Organic analytes, which bind within the hydrophobic interiors of cyclodextrins, augment this block allowing analyte detection (Gu, L.-Q., Braha, O., Conlan, S., Cheley, S., and Bayley, H. (1999) *Nature* 398, 686-690).

[0004] There is currently a need for rapid and cheap DNA or RNA sequencing technologies across a wide range of applications. Existing technologies are slow and expensive mainly because they rely on amplification techniques to produce large volumes of nucleic acid and require a high quantity of specialist fluorescent chemicals for signal detection. Stochastic sensing has the potential to provide rapid and cheap DNA sequencing by reducing the quantity of nucleotide and reagents required.

SUMMARY OF THE INVENTION

[0005] The inventors have surprisingly demonstrated that a nucleic acid binding protein remains functional if all or all but one its native accessible cysteine residues are removed by substitution. The inventors have also shown that generation of a nucleic acid binding protein with only one or more accessible cysteine residues allows the controlled attachment of the protein to a surface. The one or more accessible cysteine residues may be non-native residues introduced by mutagenesis. Alternatively, the one or more accessible residues may be native residues remaining once the other native accessible cysteine residues are removed. The invention therefore provides a construct comprising a nucleic acid binding protein and a surface, wherein at least one native accessible cysteine residue is removed from the binding protein, wherein the binding protein is attached to the surface via one or more accessible cysteine residues and wherein the binding protein retains its ability to bind nucleic acids. The nucleic acid binding protein is preferably a nucleic acid handling enzyme. The surface is preferably a pore or a pore subunit.

[0006] The inventors have also surprisingly demonstrated that the constructs of the invention can be used to generate transmembrane pores that are capable of both binding a nucleic acid and sequencing the nucleic acid via stochastic sensing. The fixed nature and close proximity of the nucleic acid binding protein to the pore means that a proportion of the nucleotides in a target nucleic acid will interact with the pore and affect the current flowing through the pore in a distinctive manner. As a result, transmembrane pores comprising such constructs are useful tools for stochastic sensing and especially for sequencing nucleic acids.

[0007] Accordingly, the invention also provides:

- **[0008]** a modified pore for use in sequencing nucleic acids, comprising at least one construct of the invention;
- **[0009]** a kit for producing a modified pore for use in sequencing nucleic acids, comprising:
  - **[0010]** (a) at least one construct of the invention in which the surface is a pore subunit; and
  - **[0011]** (b) the remaining subunits needed to form a pore;
- **[0012]** a method of producing a construct of the invention, comprising:
  - **[0013]** (a) attaching a nucleic acid binding protein comprising one or more accessible cysteine residues to a surface via the cysteine residues; and
  - **[0014]** (b) determining whether or not the resulting construct is capable of binding nucleic acids;
- **[0015]** a method of producing a modified pore of the invention, comprising:
  - **[0016]** (a) allowing at least one construct of the invention in which the surface is a pore subunit to form a pore with other suitable subunits; and
  - **[0017]** (b) determining whether or not the resulting pore is capable of binding nucleic acids and detecting nucleotides;

[0018] a method of sequencing a target nucleic acid sequence, comprising:

[0019] (a) contacting the target sequence with a pore of the invention, which comprises an exonuclease, such that the exonuclease digests an individual nucleotide from one end of the target sequence;

[0020] (b) contacting the nucleotide with the pore so that the nucleotide interacts with the adaptor;

[0021] (c) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide; and

[0022] (d) repeating steps (a) to (c) at the same end of the target sequence and thereby determining the sequence of the target sequence;

[0023] a method of sequencing a target nucleic acid sequence, comprising:

[0024] (a) contacting the target sequence with a pore of the invention comprising a nucleic acid handling enzyme so that the enzyme pushes or pulls the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore; and

[0025] (b) measuring the current passing through the pore during each interaction and thereby determining the sequence of the target sequence;

[0026] an exonuclease enzyme comprising the sequence shown in any one of SEQ ID NOs: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 or a variant thereof; and

[0027] a polynucleotide sequence which encodes an exonuclease enzyme of the invention.

DESCRIPTION OF THE FIGURES

[0028] FIG. **1** shows a cartoon of the structure of EcoExo I showing the position and residue number of naturally occurring cysteines (obtained from crystal structure).

[0029] FIG. **2** shows a surface diagram of EcoExo I showing the position of accessible cysteines (obtained from crystal structure).

[0030] FIG. **3** shows the relative activity of Exo I mutants compared to the wild type.

[0031] FIG. **4** shows a diagram of EcoExo I showing the positions at which a single cysteine could be introduced for coupling to a protein nanopore

[0032] FIGS. **5***a* and **5***b* show the relative activity of Exo I mutants compared to the wild type.

[0033] FIG. **6** shows a schematic representation of the Exonuclease I assay.

[0034] FIG. **7** shows an example of data from the Exonuclease I assay.

[0035] FIG. **8** shows that the attachment of a linker to Exo-ATTTT-A83C does not affect enzyme activity. The Y-axis is relative activity. The left-hand pair of columns is ONLP 1403 and the right-hand pair of columns is ONLP 1405. In each pair, the left-hand (darker) column is the enzyme with linker and the right-hand (lighter column) is enzyme without linker.

[0036] FIG. **9** shows that the attachment of PNA to Exo-ATTTT-A83C does not affect enzyme activity. The Y-axis is relative activity. The left-hand column is ONLP 1498 and the right-hand column is ONLP 1499. FIG. **10** shows that the attachment of a PEG linker to Exo-ATTTT or Exo-ATTTT-M184C does not affect enzyme activity. The Y-axis is relative activity. The left-hand pair of columns is Exo-ATTTT and the right-hand pair of columns is Exo-ATTTT-M184C. In each pair, the left-hand (darker) column is the enzyme with linker and the right-hand (lighter column) is enzyme without linker.

DESCRIPTION OF THE SEQUENCE LISTING

[0037] SEQ ID NO: 1 shows the polynucleotide sequence encoding one subunit of wild type α-hemolysin (α-HL).

[0038] SEQ ID NO: 2 shows the amino acid sequence of one subunit of wild type α-HL. Amino acids 2 to 6, 73 to 75, 207 to 209, 214 to 216 and 219 to 222 form α-helices. Amino acids 22 to 30, 35 to 44, 52 to 62, 67 to 71, 76 to 91, 98 to 103, 112 to 123, 137 to 148, 154 to 159, 165 to 172, 229 to 235, 243 to 261, 266 to 271,285 to 286 and 291 to 293 form β-strands. All the other non-terminal amino acids, namely 7 to 21, 31 to 34, 45 to 51, 63 to 66, 72, 92 to 97, 104 to 111, 124 to 136, 149 to 153, 160 to 164, 173 to 206, 210 to 213, 217, 218, 223 to 228, 236 to 242, 262 to 265, 272 to 274 and 287 to 290 form loop regions. Amino acids 1 and 294 are terminal amino acids.

[0039] SEQ ID NO: 3 shows the polynucleotide sequence encoding one subunit of α-HL L135C/N139Q (HL-CQ).

[0040] SEQ ID NO: 4 shows the amino acid sequence of one subunit of ═-HL L135C/N139Q (HL-CQ). The same amino acids that form α-helices, β-strands and loop regions in wild type α-HL form the corresponding regions in this subunit.

[0041] SEQ ID NO: 5 shows the codon optimised polynucleotide sequence derived from the sbcB gene from *E. coli*. It encodes the exonuclease I enzyme (EcoExo I) from *E. coli*.

[0042] SEQ ID NO: 6 shows the amino acid sequence of exonuclease I enzyme (EcoExo I) from *E. coli*. This enzyme performs processive digestion of 5' monophosphate nucleosides from single stranded DNA (ssDNA) in a 5' to 3' direction. Amino acids 60 to 68, 70 to 78, 80 to 93, 107 to 119, 124 to 128, 137 to 148, 165 to 172, 182 to 211, 213 to 221, 234 to 241, 268 to 286, 313 to 324, 326 to 352, 362 to 370, 373 to 391, 401 to 454 and 457 to 475 form α-helices. Amino acids 10 to 18, 28 to 26, 47 to 50, 97 to 101, 133 to 136, 229 to 232, 243 to 251, 258 to 263, 298 to 302 and 308 to 311 form β-strands. All the other non-terminal amino acids, 19 to 27, 37 to 46, 51 to 59, 69, 79, 94 to 96102 to 106, 120 to 123, 129 to 132, 149 to 164, 173 to 181, 212, 222 to 228 233, 242, 252 to 257, 264 to 267, 287 to 297, 303 to 307, 312, 325, 353 to 361, 371, 372, 392 to 400, 455 and 456, form loops. Amino acids 1 to 9 are terminal amino acids. The overall fold of the enzyme is such that three regions combine to form a molecule with the appearance of the letter C, although residues 355-358, disordered in the crystal structure, effectively convert this C into an O-like shape. The amino terminus (1-206) forms the exonuclease domain and has homology to the DnaQ superfamily, the following residues (202-354) form an SH3-like domain and the carboxyl domain (359-475) extends the exonuclease domain to form the C-like shape of the molecule. Four acidic residues of EcoExo I are conserved with the active site residues of the DnaQ superfamily (corresponding to D15, E17, D108 and D186). It is suggested a single metal ion is bound by residues D15 and 108. Hydrolysis of DNA is likely catalyzed by attack of the scissile phosphate with an activated water molecule, with H181 being the catalytic residue and aligning the nucleotide substrate.

[0043] SEQ ID NO: 7 shows the codon optimised polynucleotide sequence encoding EcoExo I C98S/C306S/C330T/C51A (ONLD0393).

[0044] SEQ ID NO: 8 shows the amino acid sequence of EcoExo I C98S/C306S/C330T/C51A (ONLD0393).

[0045] SEQ ID NO: 9 shows the codon optimised polynucleotide sequence encoding EcoExo I C98S/C306S/C330T/C144M (ONLD0403).

[0046] SEQ ID NO: 10 shows the amino acid sequence of EcoExo I C98S/C306S/C330T/C144M (ONLD0403).

[0047] SEQ ID NO: 11 shows the codon optimised polynucleotide sequence encoding EcoExo I C98S/C306S/C330T/C144T (ONLD0404).

[0048] SEQ ID NO: 12 shows the amino acid sequence of EcoExo I C98S/C306S/C330T/C144T (ONLD0404).

[0049] SEQ ID NO: 13 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144M/C306S/C330TN42C (ONLD0415).

[0050] SEQ ID NO: 14 shows the amino acid sequence of EcoExo I C51A/C98S/C144M/C306S/C330TN42C (ONLD0415).

[0051] SEQ ID NO: 15 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144T/C306S/C330TN42C (ONLD0416).

[0052] SEQ ID NO: 16 shows the amino acid sequence of EcoExo I C51A/C98S/C144T/C306S/C330TN42C (ONLD0416).

[0053] SEQ ID NO: 17 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144M/C306S/C330T/M184C (ONLD0417).

[0054] SEQ ID NO: 18 shows the amino acid sequence of EcoExo I C51A/C98S/C144M/C306S/C330T/M184C (ONLD0417).

[0055] SEQ ID NO: 19 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144T/C306S/C330T/M184C (ONLD0418).

[0056] SEQ ID NO: 20 shows the amino acid sequence of EcoExo I C51A/C98S/C144T/C306S/C330T/M184C (ONLD0418).

[0057] SEQ ID NO: 21 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411).

[0058] SEQ ID NO: 22 shows the amino acid sequence of EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411).

[0059] SEQ ID NO: 23 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432).

[0060] SEQ ID NO: 24 shows the amino acid sequence of EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432).

[0061] SEQ ID NO: 25 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433).

[0062] SEQ ID NO: 26 shows the amino acid sequence of EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433).

[0063] SEQ ID NO: 27 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451).

[0064] SEQ ID NO: 28 shows the amino acid sequence of EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451).

[0065] SEQ ID NO: 29 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452).

[0066] SEQ ID NO: 30 shows the amino acid sequence of EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452).

[0067] SEQ ID NO: 31 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453).

[0068] SEQ ID NO: 32 shows the amino acid sequence of EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453).

[0069] SEQ ID NO: 33 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98D/C144T/C306S/C330T (ONLD0491).

[0070] SEQ ID NO: 34 shows the amino acid sequence of EcoExo I C51A/C98D/C144T/C306S/C330T (ONLD0491).

[0071] SEQ ID NO: 35 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454).

[0072] SEQ ID NO: 36 shows the amino acid sequence of EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454).

[0073] SEQ ID NO: 37 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455).

[0074] SEQ ID NO: 38 shows the amino acid sequence of EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455).

[0075] SEQ ID NO: 39 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456).

[0076] SEQ ID NO: 40 shows the amino acid sequence of EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456).

[0077] SEQ ID NO: 41 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476).

[0078] SEQ ID NO: 42 shows the amino acid sequence of EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476).

[0079] SEQ ID NO: 43 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477).

[0080] SEQ ID NO: 44 shows the amino acid sequence of EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477).

[0081] SEQ ID NO: 45 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478).

[0082] SEQ ID NO: 46 shows the amino acid sequence of EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478).

[0083] SEQ ID NO: 47 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479).

[0084] SEQ ID NO: 48 shows the amino acid sequence of EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479).

[0085] SEQ ID NO: 49 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480).

[0086] SEQ ID NO: 50 shows the amino acid sequence of EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480).

[0087] In all of the mutants described in SEQ ID NOs: 7 to 50, the same amino acids that form α-helices, β-strands and loop regions in wild type EcoExo I form the corresponding regions in this mutant.

[0088] SEQ ID NO: 51 shows the codon optimised polynucleotide sequence derived from the xthA gene from *E. coli*. It encodes the exonuclease III enzyme from *E. coli*.

[0089] SEQ ID NO: 52 shows the amino acid sequence of the exonuclease III enzyme from *E. coli*. This enzyme performs distributive digestion of 5' monophosphate nucleosides from one strand of double stranded DNA (dsDNA) in a 3'-5' direction. Enzyme initiation on a strand requires a 5' overhang of approximately 4 nucleotides. Amino acids 11 to 13, 15 to 25, 39 to 41, 44 to 49, 85 to 89, 121 to 139, 158 to 160, 165 to 174, 181 to 194, 198 to 202, 219 to 222, 235 to 240 and 248 to 252 form α-helices. Amino acids 2 to 7, 29 to 33, 53 to 57, 65 to 70, 75 to 78, 91 to 98, 101 to 109, 146 to 151, 195 to 197, 229 to 234 and 241 to 246 form β-strands. All the other non-terminal amino acids, 8 to 10, 26 to 28, 34 to 38, 42, 43, 50 to 52, 58 to 64, 71 to 74, 79 to 84, 90, 99, 100, 110 to 120, 140 to 145, 152 to 157, 161 to 164, 175 to 180, 203 to 218, 223 to 228, 247 and 253 to 261, form loops. Amino acids 1, 267 and 268 are terminal amino acids. The enzyme active site is formed by loop regions connecting β_1-α_1, β_3-β_4, β_5-β_6, β_{III}-α_I, β_{IV}-α_{II} and β_V-β_{VI} (consisting of amino acids 8 -10, 58-64, 90, 110-120, 152 -164, 175-180, 223-228 and 253-261 respectively). A single divalent metal ion is bound at residue E34 and aids nucleophilic attack on the phosphodiester bond by the D229 and H259 histidine-aspartate catalytic pair.

[0090] SEQ ID NO: 53 shows the codon optimised polynucleotide sequence derived from the recJ gene from *T. thermophilus*. It encodes the RecJ enzyme from *T. thermophilus* (TthRecJ-cd).

[0091] SEQ ID NO: 54 shows the amino acid sequence of the RecJ enzyme from *T. thermophilus* (TthRecJ-cd). This enzyme performs processive digestion of 5' monophosphate nucleosides from ssDNA in a 5'-3' direction. Enzyme initiation on a strand requires at least 4 nucleotides. Amino acids 19 to 33, 44 to 61, 80 to 89, 103 to 111, 136 to 140, 148 to 163, 169 to 183, 189 to 202, 207 to 217, 223 to 240, 242 to 252, 254 to 287, 302 to 318, 338 to 350 and 365 to 382 form α-helices. Amino acids 36 to 40, 64 to 68, 93 to 96, 116 to 120, 133 to 135, 294 to 297, 321 to 325, 328 to 332, 352 to 355 and 359 to 363 form β-strands. All the other non-terminal amino acids, 34, 35, 41 to 43, 62, 63, 69 to 79, 90 to 92, 97 to 102, 112 to 115, 121 to 132, 141 to 147, 164 to 168, 184 to 188203 to 206, 218 to 222, 241, 253, 288 to 293, 298 to 301, 319, 320, 326, 327, 333 to 337, 351 to 358 and 364, form loops. Amino acids 1 to 18 and 383 to 425 are terminal amino acids. The crystal structure has only been resolved for the core domain of RecJ from *Thermus thermophilus* (residues 40-463). To ensure initiation of translation and in vivo expression of the RecJ core domain a methionine residue was added at its amino terminus, this is absent from the crystal structure information. The resolved structure shows two domains, an amino (2-253) and a carboxyl (288-463) region, connected by a long α-helix (254-287). The catalytic residues (D46, D98, H122, and D183) co-ordinate a single divalent metal ion for nucleophilic attack on the phosphodiester bond. D46 and H120 proposed to be the catalytic pair; however, mutation of any of these conserved residues in the *E. coli* RecJ was shown to abolish activity.

[0092] SEQ ID NO: 55 shows the codon optimised polynucleotide sequence derived from the bacteriophage lambda exo (redX) gene. It encodes the bacteriophage lambda exonuclease.

[0093] SEQ ID NO: 56 shows the amino acid sequence of the bacteriophage lambda exonuclease. The sequence is one of three identical subunits that assemble into a trimer. The enzyme performs highly processive digestion of nucleotides from one strand of dsDNA, in a 5'-3'direction (http://www.neb.com/nebecomm/products/productMO262.asp). Enzyme initiation on a strand preferentially requires a 5' overhang of approximately 4 nucleotides with a 5' phosphate. Amino acids 3 to 10, 14 to 16, 22 to 26, 34 to 40, 52 to 67, 75 to 95, 135 to 149, 152 to 165 and 193 to 216 form α-helices. Amino acids 100 to 101, 106 to 107, 114 to 116, 120 to 122, 127 to 131, 169 to 175 and 184 to 190 form β-strands. All the other non-terminal amino acids, 11 to 13, 17 to 21, 27 to 33, 41 to 51, 68 to 74, 96 to 99, 102 to 105, 108 to 113, 117 to 119, 123 to 126, 132 to 134, 150 to 151, 166 to 168, 176 to 183, 191 to 192, 217 to 222, form loops. Amino acids 1, 2 and 226 are terminal amino acids. Lambda exonuclease is a homo-trimer that forms a toroid with a tapered channel through the middle, apparently large enough for dsDNA to enter at one end and only ssDNA to exit at the other. The catalytic residues are undetermined but a single divalent metal ion appears bound at each subunit by residues D119, E129 and L130.

[0094] SEQ ID NO: 57 shows the nucleic sequence from which preferred nucleic acid linkers can be generated.

[0095] SEQ ID NO: 58 shows a preferred nucleic acid linker. MAL is maleimide. This linker is used in combination with SEQ ID NO: 61.

[0096] SEQ ID NO: 59 shows a preferred nucleic acid linker. MAL is maleimide. This linker is used in combination with SEQ ID NO: 62.

[0097] SEQ ID NO: 60 shows a preferred nucleic acid linker. MAL is maleimide. This linker is used in combination with SEQ ID NO: 63.

[0098] SEQ ID NO: 61 shows a preferred nucleic acid linker. MAL is maleimide. This linker is complementary to and used in combination with SEQ ID NO: 58.

[0099] SEQ ID NO: 62 shows a preferred nucleic acid linker. MAL is maleimide. This linker is complementary to and used in combination with SEQ ID NO: 59.

[0100] SEQ ID NO: 63 shows a preferred nucleic acid linker. MAL is maleimide. This linker is complementary to and used in combination with SEQ ID NO: 60.

[0101] SEQ ID NO: 64 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 65.

[0102] SEQ ID NO: 65 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 64.

[0103] SEQ ID NO: 66 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 68.

[0104] SEQ ID NO: 67 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 69.

[0105] SEQ ID NO: 68 shows a preferred nucleic acid linker. This linker is complementary to and used in combination with SEQ ID NO: 66.

[0106] SEQ ID NO: 69 shows a preferred nucleic acid linker. This linker is complementary to and used in combination with SEQ ID NO: 67.

[0107] SEQ ID NO: 70 shows a preferred nucleic acid linker. This linker is complementary to and used in combination with SEQ ID NO: 73.

[0108] SEQ ID NO: 71 shows a preferred nucleic acid linker. This linker is complementary to and used in combination with SEQ ID NO: 74.

[0109] SEQ ID NO: 72 shows a preferred nucleic acid linker This linker is used in combination with SEQ ID NO: 75.

[0110] SEQ ID NO: 73 shows a preferred nucleic acid linker This linker is used in combination with SEQ ID NO: 70.

[0111] SEQ ID NO: 74 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 71.

[0112] SEQ ID NO: 75 shows a preferred nucleic acid linker. This linker is complementary to and used in combination with SEQ ID NO: 72.

DETAILED DESCRIPTION OF THE INVENTION

[0113] It is to be understood that different applications of the disclosed products and methods may be tailored to the specific needs in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

[0114] In addition as used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a construct" includes "constructs", reference to "a transmembrane protein pore" includes two or more such pores, reference to "a molecular adaptor" includes two or more such adaptors, and the like.

[0115] All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

Constructs

[0116] The invention provides constructs that are useful for binding nucleic acids. In particular, the invention provides constructs that are useful for handling nucleic acids, such as when sequencing nucleic acids. The constructs comprise a nucleic acid binding protein and a surface. At least one native accessible cysteine residue is removed from the nucleic acid binding protein. However, the nucleic acid binding protein comprises one or more accessible cysteine residues and is attached to the surface via those residues. The presence of a limited number of accessible cysteine residues allows controlled attachment to the surface.

[0117] The ionisable side chain of cysteine residues is a potent nucleophile for engaging in addition reactions and so is a popular choice for use in bioconjugation techniques. However, a common barrier to the effective use of these techniques is the native cysteine residues present in wild type proteins. Modification of one or more of the native cysteine residues can cause both ambiguous enzyme activity and uncontrolled linkage.

[0118] The invention concerns site directed mutagenesis of native cysteine residues from nucleic acid binding protein. All but one or more of the accessible native cysteines can be removed. Alternatively, all the accessible cysteine residues can be removed and one or more cysteine residues can be introduced to the protein. The presence of specific accessible cysteine residues facilitates the attachment of the binding protein to a surface, such as another protein, a solid support or a chemical reagent.

[0119] Removal of native cysteine residues and targeted addition of single or multiple cysteine residue improves on previous methodologies as it conveys the ability to control attachment, while also minimising the possible interaction of the nucleic acid binding proteins with one another to form dimers, trimers and the like via undesired surface thiols. If the protein contains only a single cysteine residue at a designed position in its structure, then the thiol group of that residue can be used to form either a direct disulphide bond with other sulfhydryl groups or coupling can be mediated by a reactive linker. These targeted cysteine residues ensure only the desired number of enzyme molecules cross-link to the desired number of target molecules, whilst also conferring a degree of favourable conformation, so that for instance active sites can be optimally orientated.

[0120] The native cysteine residues present in the wild type enzyme do not readily permit the use of specific linkers containing thiol reactive groups, such as maleimide, iodoactemide or ortho-pyridyl disulphide (OPSS). The potential for reaction with any structurally or catalytically important cysteine residues could impact upon the binding protein's activity, which is of particular importance for use in single molecule applications. The use of linkers for bioconjugation is preferred to ensure some spatial separation so that binding protein activity is not affected by being in close proximity to another protein or a surface, such as solid support.

[0121] The surface is preferably a pore or a pore subunit. The constructs of the invention are therefore useful tools for forming pores that are capable of sequencing nucleic acids by stochastic sensing. The constructs of the invention are particularly useful for generating transmembrane pores that can both bind a target nucleic acid sequence and discriminate between the different nucleotides in the target sequence. As described in more detail below, the nucleic acid binding protein preferably handles a target nucleic acid in such a way that the pore can identify nucleotides in the target sequence and thereby sequence the target sequence. In particular, the constructs of the invention not only allow co-localisation of a nucleic acid binding protein, such as a nucleic acid handling enzyme, to a nanopore, but also the enzyme active site to be orientated to optimise base translocation once catalysis has occurred.

[0122] The surface does not have to be a pore or pore subunit. Other sequencing techniques commonly rely on nucleic acids being immobilised onto a solid support, such as a bead, before an enzyme binds from solution and metal ions are added to trigger catalytic activity. Liberated bases can then be detected downstream in a number of manners. However, one barrier to this method is the time consuming and expensive amplification and subsequent chemical modification required in order to couple the template nucleic acid to the bead surface. Having a nucleic acid binding protein pre-attached to a bead minimises the need for any nucleic acid modification. Existing methods of enzyme attachment would not be suitable as they rely on the undirected bulk attachment of nucleic binding proteins and so are highly variable in terms of the amount of attached protein and the activity of the attached protein due to the unknown orientation or spatial separation from the surface. The constructs of the invention allow controlled addition of a single nucleic acid binding protein to a surface and therefore vastly improve on this sequencing method.

[0123] One important class of nucleic acid binding proteins is the exonucleases. The common use for exonucleases in molecular biology is the removal of the amplification primers from PCR reactions to allow the direct use of the product for sequencing by addition of a specific sequencing primer. There is still a requirement however for the exonuclease to be either denatured or removed in order to prevent degradation of the sequencing primer. A method for the rapid and easy removal of the exonuclease at this stage is therefore desirable both to aid in the automation of this process, as well as reduce costly

DNA purification procedures. A construct of the invention comprising an exonuclease attached to a surface would be ideal for such applications. The problems associated with the existing methods of attaching exonucleases to a surface, such as a glass slide, the bottom of a well or a bead, have already been outlined above. The covalent coupling of an exonuclease to a surface, such as an agarose bead, through a single cysteine residue would vastly improve upon known methods, as well as help any potential enzyme solubility problems. In addition, as can be seen from the Example, removal of native cysteine residues is capable of producing proteins with an altered activity and stability. An exonuclease with a high activity, but lowered stability may also aid in the removal of the enzyme activity from the PCR reaction. Similar potential applications for many other commercially important enzymes should also be considered.

[0124] The constructs of the invention may be isolated, substantially isolated, purified or substantially purified. A construct is isolated or purified if it is completely free of any other components, such as lipids or other pore monomers. A construct is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a construct is substantially isolated or substantially purified if it present in a form that comprises less than 10%, less than 5%, less than 2% or less than 1% of other components, such as lipids or other pore monomers. A construct of the invention may be present in a lipid bilayer.

Mutagenesis of the Cysteine Residues

[0125] Constructs of the invention comprise a nucleic acid binding protein in which at least one accessible cysteine residue is removed. The protein also has one or more accessible cysteine residues. Accessible cysteine residues are residues that are available for reaction with a thiol specific group. Methods for determining whether cysteine residues are available for reaction with a thiol specific group are well known in the art. Accessible cysteine residues are typically present on the accessible surface of the protein. Accessible cysteine residues are cysteine residues that are not buried within the binding protein and/or do not form an intramolecular disulphide bond. Cysteine residues that are buried within the binding protein and/or form intramolecular disulphide bonds do not need to be removed from the binding protein because they will not interfere with attachment to the surface. Cysteine residues can of course form disulphide bonds with other cysteine residues in the protein or with other chemical conjugates containing a reactive thiol group.

[0126] Accessible cysteine residues can be mapped using sequence information and molecular modeling. Methods for such modeling are known in the art. One method is also described in the Example. Once the accessible cysteine residues have been mapped, at least one of them can be removed as described below.

[0127] Any number of native accessible cysteine residues, such as two, five, ten or more, can be removed from the binding protein. In some embodiments, all the native accessible cysteine residues are removed from the nucleic acid binding protein. This is discussed in more detail below.

[0128] Native accessible cysteine residues can be removed using any method known in the art. Native residues are residues that are present in the native or wild type protein. Native cysteine residues can be removed by deletion. Native cysteine residues are preferably removed by substitution. Native cysteine residues can be substituted with naturally occurring residues or non-naturally occurring residues. Native cysteine residues are typically substituted with residues lacking a reactive thiol group. Native cysteine residues are preferably substituted with residues having a similar structure. Suitable residues for replacing cysteine residues include, but are not limited to, alanine, serine, threonine, methionine and valine. Cysteine residues are most preferably substituted with threonine. The number of accessible cysteines removed from the nucleic acid binding protein will of course depend on the number present in the native protein.

[0129] In one embodiment, all the native accessible cysteine residues are removed from the binding protein and one or more one, such as two, three, four, five or more, non-native accessible cysteine residue are introduced into the binding protein. Non-native residues are residues that are not present in the native or wild type protein. Non-native cysteine residues can be introduced by addition. Non-native cysteine residues are preferably introduced by substitution. Non-native cysteine residues are typically introduced in place of residues lacking a reactive thiol group. Any residue, such as methionine, valine, serine or alanine, can be replaced with cysteine.

[0130] In another embodiment, all but one or more of the native accessible cysteine residues are removed from the binding protein. In this embodiment, the nucleic acid binding protein comprises one or more, such as two, three, four, five or more, native accessible cysteine residues. Preferably, all but one or all but two of the native accessible cysteine residues are removed from the binding protein. The remaining native accessible cysteine residues are used to attach the binding protein to the surface in a specific manner. In this embodiment, the nucleic acid binding protein may further comprise one or more, such as two, three, four, five or more, non-native accessible cysteine residues introduced as described above. The nucleic acid binding protein may be attached to the surface by the native residues, the non-native residues or both the native and non-native residues.

[0131] In preferred embodiments, the nucleic acid binding protein comprises only one or only two accessible cysteine residues. The cysteine residues may be native, non-native or a combination of the two as described above. The nucleic acid protein is attached to the surface via these cysteine residues. The presence of only one or only two accessible cysteine residues can be determined using any method known in the art. For instance, the nucleic acid binding protein can be reacted with conjugates containing a reactive thiol group and the number of conjugates attached to the protein can be determined. Methods for carrying out such reactions are discussed in more detail below. A nucleic acid binding protein comprises only one accessible cysteine residue if only one conjugate is attached to the protein following such a reaction. A nucleic acid binding protein comprises only two accessible cysteine residues if only two conjugates are attached to the protein following such a reaction.

[0132] The nucleic acid binding protein retains its ability to bind nucleic acids. This allows the construct to bind nucleic acids and preferably sequence nucleic acids as described below. The ability of a construct to bind nucleic acids can be assayed using any method known in the art. For instance, constructs or pores formed from the constructs can be tested for their ability to bind specific sequences of nucleic acids. The ability of a construct or a pore to bind nucleic acids is typically assayed as described in the Example.

[0133] The nucleic acid binding protein also retains its ability to be expressed using standard techniques. An inability to express the protein once one or more native cysteine residues are removed of course means that the protein cannot be attached to the surface. It is straightforward to determine whether a particular mutant can be expressed or not. The ability of a nucleic acid binding protein to be expressed is typically assayed as described in the Example.

Nucleic Acid Binding Protein

[0134] The constructs of the invention comprise a nucleic acid binding protein. Examples of such proteins include, but are not limited to, nucleic acid handling enzymes, such as nucleases, polymerases, topoisomerases, ligases and helicases, and non-catalytic binding proteins such as those classified by SCOP (Structural Classification of Proteins) under the Nucleic acid-binding protein superfamily (50249). The nucleic acid binding protein is modified to remove and/or replace cysteine residues as described above.

[0135] A nucleic acid is a macromolecule comprising two or more nucleotides. The nucleic acid bound by the protein may comprise any combination of any nucleotides. The nucleotides can be naturally occurring or artificial. The nucleotide can be oxidized or methylated. A nucleotide typically contains a nucleobase, a sugar and at least one phosphate group. The nucleobase is typically heterocyclic. Nucleobases include, but are not limited to, purines and pyrimidines and more specifically adenine, guanine, thymine, uracil and cytosine. The sugar is typically a pentose sugar. Nucleotide sugars include, but are not limited to, ribose and deoxyribose. The nucleotide is typically a ribonucleotide or deoxyribonucleotide. The nucleotide typically contains a monophosphate, diphosphate or triphosphate. Phosphates may be attached on the 5' or 3' side of a nucleotide.

[0136] Nucleotides include, but are not limited to, adenosine monophosphate (AMP), adenosine diphosphate (ADP), adenosine triphosphate (ATP), guanosine monophosphate (GMP), guanosine diphosphate (GDP), guanosine triphosphate (GTP), thymidine monophosphate (TMP), thymidine diphosphate (TDP), thymidine triphosphate (TTP), uridine monophosphate (UMP), uridine diphosphate (UDP), uridine triphosphate (UTP), cytidine monophosphate (CMP), cytidine diphosphate (CDP), cytidine triphosphate (CTP), cyclic adenosine monophosphate (cAMP), cyclic guanosine monophosphate (cGMP), deoxyadenosine monophosphate (dAMP), deoxyadenosine diphosphate (dADP), deoxyadenosine triphosphate (dATP), deoxyguanosine monophosphate (dGMP), deoxyguanosine diphosphate (dGDP), deoxyguanosine triphosphate (dGTP), deoxythymidine monophosphate (dTMP), deoxythymidine diphosphate (dTDP), deoxythymidine triphosphate (dTTP), deoxyuridine monophosphate (dUMP), deoxyuridine diphosphate (dUDP), deoxyuridine triphosphate (dUTP), deoxycytidine monophosphate (dCMP), deoxycytidine diphosphate (dCDP) and deoxycytidine triphosphate (dCTP). The nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP or dCMP.

[0137] The nucleic acid can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). The nucleic acid may be any synthetic nucleic acid known in the art, such as peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or other synthetic polymers with nucleotide side chains. The nucleic acid bound by the protein is preferably single stranded, such as cDNA, RNA, GNA, TNA or LNA. The nucleic acid bound by the protein is preferably double stranded, such as DNA. Proteins that bind single stranded nucleic acids may be used to sequence double stranded DNA as long as the double stranded DNA is dissociated into a single strand before it is bound by the protein.

[0138] It is preferred that the tertiary structure of the nucleic acid binding protein is known. Knowledge of the three dimensional structure of the binding protein allows modifications to be made to the protein to facilitate its function in the construct or pore of the invention.

[0139] The protein may be any size and have any structure. For instance, the protein may be an oligomer, such as a dimer or trimer. The protein is preferably a small, globular polypeptide formed from one monomer. Such proteins are easy to handle and are less likely to interfere with the surface. For instance, such proteins are less likely to interfere with the pore forming ability of a pore subunit.

[0140] It is also preferred that the location and function of the active site of the protein is known. This prevents modifications being made to the active site that abolish the activity of the protein. It also allows the protein to be attached to the surface so that the protein binds the target nucleic acid sequence in a particular way. For instance, if the protein is being attached to a transmembrane protein pore for sequencing purposes, it allows a proportion of the nucleotides in a target sequence to interact with the pore as described below. In such embodiments, it is beneficial to position the active site of the protein as close as possible to the opening of the barrel of channel of the pore, without the protein itself presenting a block to the flow of current. Knowledge of the way in which a protein may orient nucleic acids also allows an effective construct to be designed.

[0141] As discussed in more detail below, it may be necessary to purify the construct of the invention. It is preferred that the nucleic acid binding protein is capable of withstanding the conditions used to purify the construct.

[0142] The surface in a construct of the invention may comprise a pore. Such pores may be used to sequence nucleic acids. In order that most of the nucleotides in the target nucleic acid are correctly identified by stochastic sensing, the protein preferably binds the nucleic acid in a buffer background which is compatible with discrimination of the nucleotides. The protein preferably has at least residual activity in a salt concentration well above the normal physiological level, such as from 100 mM to 2000 mM. The protein is more preferably modified to increase its activity at high salt concentrations. The protein may also be modified to improve its processivity, stability and shelf life.

[0143] Suitable modifications can be determined from the characterisation of nucleic acid handling enzymes from extremphiles such as halophilic, moderately halophilic bacteria, thermophilic and moderately thermophilic organisms, as well as directed evolution approaches to altering the salt tolerance, stability and temperature dependence of mesophilic or thermophilic exonucleases.

[0144] The enzyme also preferably retains at least partial activity at temperatures from 10° C. to 60° C., such as at room temperature. This allows the construct to sequence nucleic acids at a variety of temperatures, including room temperature.

[0145] The nucleic acid binding protein is preferably a nucleic acid handling enzyme. A nucleic acid handling enzyme is a polypeptide that is capable of interacting with and modifying at least one property of a nucleic acid. The enzyme may modify the nucleic acid by cleaving it to form individual nucleotides or shorter chains of nucleotides, such as di- or trinucleotides. The enzyme may modify the nucleic acid by orienting it or moving it to a specific position.

[0146] The nucleic acid handling enzyme is preferably derived from a nucleolytic enzyme. The nucleic acid handling enzyme used in the construct of the enzyme is more preferably derived from a member of any of the Enzyme Classification (EC) groups 3.1.11, 3.1.13, 3.1.14, 3.1.15, 3.1.16, 3.1.21, 3.1.22, 3.1.25, 3.1.26, 3.1.27, 3.1.30 and 3.1.31. The nucleic acid handling enzyme is more preferably based on any one of the following enzymes:

- **[0147]** 3.1.11.-Exodeoxyribonucleases producing 5'-phosphomonoesters.
  - **[0148]** 3.1.11.1 Exodeoxyribonuclease I.
  - **[0149]** 3.1.11.2 Exodeoxyribonuclease III.
  - **[0150]** 3.1.11.3 Exodeoxyribonuclease (lambda-induced).
  - **[0151]** 3.1.11.4 Exodeoxyribonuclease (phage SP3-induced).
  - **[0152]** 3.1.11.5 Exodeoxyribonuclease V.
  - **[0153]** 3.1.11.6 Exodeoxyribonuclease VII.
- **[0154]** 3.1.13.-Exoribonucleases producing 5'-phosphomonoesters.
  - **[0155]** 3.1.13.1 Exoribonuclease II.
  - **[0156]** 3.1.13.2 Exoribonuclease H.
  - **[0157]** 3.1.13.3 Oligonucleotidase.
  - **[0158]** 3.1.13.4 Poly(A)-specific ribonuclease.
  - **[0159]** 3.1.13.5 Ribonuclease D.
- **[0160]** 3.1.14.-Exoribonucleases producing 3'-phosphomonoesters.
  - **[0161]** 3.1.14.1 Yeast ribonuclease.
- **[0162]** 3.1.15.-Exonucleases active with either ribo- or deoxyribonucleic acid producing 5' phosphomonoesters
  - **[0163]** a 3.1.15.1 Venom exonuclease.
- **[0164]** 3.1.16.-Exonucleases active with either ribo- or deoxyribonucleic acid producing 3' phosphomonoesters
  - **[0165]** 3.1.16.1 Spleen exonuclease.
- **[0166]** 3.1.21.-Endodeoxyribonucleases producing 5'-phosphomonoesters.
  - **[0167]** 3.1.21.1 Deoxyribonuclease I.
  - **[0168]** 3.1.21.2 Deoxyribonuclease IV (phage-T(4)-induced).
  - **[0169]** 3.1.21.3 Type I site-specific deoxyribonuclease.
  - **[0170]** 3.1.21.4 Type II site-specific deoxyribonuclease.
  - **[0171]** 3.1.21.5 Type III site-specific deoxyribonuclease.
  - **[0172]** 3.1.21.6 CC-preferring endodeoxyribonuclease.
  - **[0173]** 3.1.21.7 Deoxyribonuclease V.
- **[0174]** 3.1.22.-Endodeoxyribonucleases producing other than 5'-phosphomonoesters.
  - **[0175]** 3.1.22.1 Deoxyribonuclease II.
  - **[0176]** 3.1.22.2 Aspergillus deoxyribonuclease K(1).
  - **[0177]** 3.1.22.3 Transferred entry: 3.1.21.7.
  - **[0178]** 3.1.22.4 Crossover junction endodeoxyribonuclease.
  - **[0179]** 3.1.22.5 Deoxyribonuclease X.
- **[0180]** 3.1.25.-Site-specific endodeoxyribonucleases specific for altered bases.
  - **[0181]** 3.1.25.1 Deoxyribonuclease (pyrimidine dimer).
  - **[0182]** 3.1.25.2 Transferred entry: 4.2.99.18.
- **[0183]** 3.1.26.-Endoribonucleases producing 5'-phosphomonoesters.
  - **[0184]** 3.1.26.1 Physarum polycephalum ribonuclease.
  - **[0185]** 3.1.26.2 Ribonuclease alpha.
  - **[0186]** 3.1.26.3 Ribonuclease III.
  - **[0187]** 3.1.26.4 Ribonuclease H.
  - **[0188]** 3.1.26.5 Ribonuclease P.
  - **[0189]** 3.1.26.6 Ribonuclease IV.
  - **[0190]** 3.1.26.7 Ribonuclease P4.
  - **[0191]** 3.1.26.8 Ribonuclease M5.
  - **[0192]** 3.1.26.9 Ribonuclease (poly-(U)-specific).
  - **[0193]** 3.1.26.10 Ribonuclease IX.
  - **[0194]** 3.1.26.11 Ribonuclease Z.
- **[0195]** 3.1.27.-Endoribonucleases producing other than 5'-phosphomonoesters.
  - **[0196]** 3.1.27.1 Ribonuclease T(2).
  - **[0197]** 3.1.27.2 *Bacillus subtilis* ribonuclease.
  - **[0198]** 3.1.27.3 Ribonuclease T(1).
  - **[0199]** 3.1.27.4 Ribonuclease U(2).
  - **[0200]** 3.1.27.5 Pancreatic ribonuclease.
  - **[0201]** 3.1.27.6 *Enterobacter* ribonuclease.
  - **[0202]** 3.1.27.7 Ribonuclease F.
  - **[0203]** 3.1.27.8 Ribonuclease V.
  - **[0204]** 3.1.27.9 tRNA-intron endonuclease.
  - **[0205]** 3.1.27.10 rRNA endonuclease.
- **[0206]** 3.1.30.-Endoribonucleases active with either ribo- or deoxyribonucleic producing 5' phosphomonoesters
  - **[0207]** 3.1.30.1 *Aspergillus* nuclease S(1).
  - **[0208]** 3.1.30.2 *Serratia marcescens* nuclease.
- **[0209]** 3.1.31.-Endoribonucleases active with either ribo- or deoxyribonucleic producing 3' phosphomonoesters
  - **[0210]** 3.1.31.1 Micrococcal nuclease.

[0211] The enzyme is most preferably derived from an exonuclease, such as a deoxyribonuclease, which cleaves nucleic acids to form individual nucleotides. The advantages of exodeoxyribonucleases are that they are active on both single stranded and double stranded DNA and hydrolyse bases either in the 5'-3' or 3'-5' direction.

[0212] An individual nucleotide is a single nucleotide. An individual nucleotide is one which is not bound to another nucleotide or nucleic acid by any bond, such as a phosphodiester bond. A phosphodiester bond involves one of the phosphate groups of a nucleotide being bound to the sugar group of another nucleotide. An individual nucleotide is typically one which is not bound in any manner to another nucleic acid sequence of at least 5, at least 10, at least 20, at least 50, at least 100, at least 200, at least 500, at least 1000 or at least 5000 nucleotides.

[0213] Preferred enzymes for use in the invention include exonuclease I from *E. coli* (SEQ ID NO: 6), exonuclease III enzyme from *E. coli* (SEQ ID NO: 52), RecJ from *T. thermophilus* (SEQ ID NO: 54) and bacteriophage lambda exonuclease (SEQ ID NO: 56) and variants thereof. Three identical subunits of SEQ ID NO: 56 interact to form a trimer exonuclease.

[0214] The enzyme is most preferably based on exonuclease I from *E. coli* (SEQ ID NO: 6).

[0215] The nucleic acid handling enzyme is preferably derived from an exonuclease enzyme comprising any of the sequences shown in SEQ ID NOs: 6, 52, 54 and 56 or a variant thereof. In other words, the enzyme preferably comprises any of the sequences shown in SEQ ID NOs: 6, 52, 54 and 56 or a variant thereof before its cysteine residues are modified as described above.

[0216] A variant of SEQ ID NO: 6, 52, 54 or 56 is an enzyme that has an amino acid sequence which varies from that of SEQ ID NO: 6, 52, 54 or 56 and which retains nucleic acid handling ability. The ability of a variant to handle nucleic acids can be assayed using any method known in the art. For instance, the ability of a variant to handle nucleic acids can be assayed as described in the Example. The variant must also retain its ability to be expressed as described in the Example.

[0217] The variant may include modifications that facilitate handling of the nucleic acid and/or facilitate its activity at high salt concentrations and/or room temperature.

[0218] The enzyme, may be a naturally occurring variant which is expressed by an organism, for instance by an *E. coli* bacterium. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the amino acid sequence of SEQ ID NO: 6, 52, 54 or 56, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of SEQ ID NO: 6, 52, 54 or 56 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 200 or more, for example 230, 250, 270 or 280 or more, contiguous amino acids ("hard homology").

[0219] Standard methods in the art may be used to determine homology. For example the UWGCG Package provides the BESTFIT program which can be used to calculate homology, for example used on its default settings (Devereux et al (1984) *Nucleic Acids Research* 12, p387-395). The PILEUP and BLAST algorithms can be used to calculate homology or line up sequences (such as identifying equivalent residues or corresponding sequences (typically on their default settings)), for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S.F et al (1990) J Mol Biol 215:403-10.

[0220] Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pair (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighbourhood word score threshold (Altschul et al, supra). These initial neighbourhood word hits act as seeds for initiating searches to find HSP's containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extensions for the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a word length (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) *Proc. Natl. Acad. Sci.* USA 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

[0221] The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) *Proc. Natl. Acad. Sci.* USA 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0222] Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 6, 52, 54 or 56 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions may be made, for example, according to Table 1 below.

TABLE 1

Conservative substitutions
Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other.

NON-AROMATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		H K R
AROMATIC		H F W Y

[0223] One or more amino acid residues of the amino acid sequence of SEQ ID NO: 6, 52, 54 or 56 may additionally be deleted from the polypeptides described above. Up to 1, 2, 3, 4, 5, 10, 20 or 30 residues may be deleted, or more.

[0224] Variants may be fragments of SEQ ID NO: 6, 52, 54 or 56. Such fragments retain nucleic acid handling activity. Fragments may be at least 50, 100, 200 or 250 amino acids in length. A fragment preferably comprises the nucleic acid handling domain of SEQ ID NO: 6, 52, 54 or 56.

[0225] One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the amino terminus or carboxy terminus of the amino acid sequence of SEQ ID NO: 6, 52, 54 or 56 or a variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids. A carrier protein may be fused to a subunit or variant.

[0226] As discussed above, a variant of SEQ ID NO: 6, 52, 54 or 56 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 6, 52, 54 or 56 and which retains its ability to handle nucleic acids. A variant typically contains the regions of SEQ ID NO: 6, 52, 54 or 56 that are responsible for handling nucleic acids. The catalytic domains of SEQ ID NOs: 6, 52, 54 and 56 are discussed above in the description of the sequence listing. A variant of SEQ ID NO: 6, 52, 54 or 56 preferably comprises the relevant catalytic domain. A variant SEQ ID NO: 6, 52, 54 or 56 typically includes one or more modifications, such as substitutions, additions or deletions, outside the relevant catalytic domain. Specific variants of SEQ ID NO: 6 are discussed in more detail below.

[0227] The variant may be modified for example by the addition of histidine or aspartic acid residues to assist its identification or purification 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.

[0228] Preferred enzymes that are capable of pushing or pulling the target nucleic acid sequence through the pore include polymerases, nucleases, helicases and topoisomerases, such as gyrases. The nucleic acid handling enzyme can be derived from any of these types of enzymes. The polymerase is preferably a member of any of the Enzyme Classification (EC) groups 2.7.7.6, 2.7.7.7, 2.7.7.19, 2.7.7.48 and 2.7.7.49. The polymerase is preferably a DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase, a DNA-dependent RNA polymerase or an RNA-dependent RNA polymerase. The helicase is preferably based on a member of any of the Enzyme Classification (EC) groups 3.6.1.- and 2.7.7.-. The helicase is preferably an ATP-dependent DNA helicase (EC group 3.6.1.8), an ATP-dependent RNA helicase (EC group 3.6.1.8) or an ATP-independent RNA helicase. The topoisomerase is preferably a member of any of the Enzyme Classification (EC) groups 5.99.1.2 and 5.99.1.3.

[0229] The nucleic acid binding protein may be labelled with a revealing label. The revealing label may be any suitable label which allows the pore to be detected. Suitable labels include, but are not limited to, fluorescent molecules, radioisotopes, e.g. ^{125}I, ^{35}S, ^{14}C, enzymes, antibodies, antigens, polynucleotides and ligands such as biotin.

[0230] The nucleic acid binding protein may be isolated from a binding protein producing organism, such as *E. coli, T thermophilus* or bacteriophage, or made synthetically or by recombinant means. For example, the nucleic acid binding protein may be synthesised by in vitro translation and transcription. The amino acid sequence of the nucleic acid binding protein may be modified to include non-naturally occurring amino acids or to increase the stability of the protein. When the nucleic acid binding protein is produced by synthetic means, such amino acids may be introduced during production. The nucleic acid binding protein may also be altered following either synthetic or recombinant production.

[0231] The nucleic acid binding protein may also be produced using D-amino acids. For instance, the nucleic acid binding proteins may comprise a mixture of L-amino acids and D-amino acids. This is conventional in the art for producing such proteins or peptides.

[0232] The nucleic acid binding protein may also contain other non-specific chemical modifications as long as they do not interfere with its ability to handle nucleic acids or attach to the surface. A number of non-specific side chain modifications are known in the art and may be made to the side chains of the pores. Such modifications include, for example, reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with amidination with methylacetimidate or acylation with acetic anhydride. The modifications to the nucleic acid binding protein can be made after expression of the nucleic acid binding protein or after the nucleic acid binding protein has been used to form a construct of the invention.

[0233] The nucleic acid binding protein can be produced using standard methods known in the art. Polynucleotide sequences encoding a nucleic acid binding protein may be isolated and replicated using standard methods in the art. Such sequences are discussed in more detail below. Polynucleotide sequences encoding a nucleic acid binding protein may be expressed in a bacterial host cell using standard techniques in the art. The nucleic acid binding protein may be produced in a cell by in situ expression of the polypeptide from a recombinant expression vector. The expression vector optionally carries an inducible promoter to control the expression of the polypeptide.

[0234] A nucleic acid binding protein may be produced in large scale following purification by any protein liquid chromatography system from pore producing organisms or after recombinant expression as described below. Typical protein liquid chromatography systems include FPLC, AKTA systems, the Bio-Cad system, the Bio-Rad BioLogic system and the Gilson HPLC system.

Preferred Nucleic Acid Handling Enzymes

[0235] The nucleic acid handling enzyme is preferably derived from an exonuclease enzyme comprising the sequence shown in SEQ ID NO: 6 or a variant thereof. In other words, the enzyme preferably comprises the sequence shown in SEQ ID NO: 6 or a variant thereof before its cysteine residues are modified as described above. Variants of SEQ ID NO: 6 are discussed above.

[0236] SEQ ID NO: 6 has five native cysteine residues at positions 51, 98, 144, 306 and 330. All five cysteine residues at these positions are accessible. A variant of SEQ ID NO: 6 preferably comprises all five of these residues before it is modified in accordance with the invention.

[0237] In one embodiment, all five of the cysteines at positions 51, 98, 144, 306 and 330 are removed from SEQ ID NO: 6 and one or more non-native cysteine residues are introduced.

[0238] In another embodiment, all but one or more of the five cysteine residues at positions 51, 98, 144, 306 and 330 are removed from SEQ ID NO: 6. Any combination of one or more of the five cysteine residues at positions 51, 98, 144, 306 and 330 can remain in an enzyme derived from SEQ ID NO: 6 after mutagenesis. Preferred combinations include, but are not limited, to 144 and 330. In a preferred embodiment, only the native cysteine residue at position 144 or 330 of SEQ ID NO: 6 remains.

[0239] One or more of the five cysteine residues in SEQ ID NO: 6 are preferably substituted with alanine, serine, methionine or threonine. The cysteine residue at position 51 in SEQ ID NO: 6 is more preferably substituted with alanine. The cysteine residue at position 98 in SEQ ID NO: 6 is more preferably substituted with serine or threonine. The cysteine residue at position 144 in SEQ ID NO: 6 is more preferably substituted with methionine or threonine. The cysteine residue at position 306 in SEQ ID NO: 6 is more preferably substituted with serine or threonine. The cysteine residue at position 330 in SEQ ID NO: 6 is more preferably substituted with threonine. In the most preferred embodiment, the cysteine residue at position 51 in SEQ ID NO: 6 is substituted with alanine, the cysteine residue at position 98 in SEQ ID NO: 6 is substituted with threonine, the cysteine residue at position 144 in SEQ ID NO: 6 is substituted with threonine, the cysteine residue at position 306 in SEQ ID NO: 6 is substituted with threonine and the cysteine residue at position 330 in SEQ ID NO: 6 is substituted with threonine.

[0240] The constructs of the invention most preferably comprise a nucleic acid handling enzyme comprising the sequence shown in any one of SEQ ID NOs: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 or a variant thereof. A variant of SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 is a protein that has an amino acid sequence which varies from that of SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 and which retains its ability to handle nucleic acids. Variants may differ from SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 or 50 to the same extent as variants of SEQ ID NO: 6 differ from SEQ ID NO: 6 as discussed above. However, variants of SEQ ID NOs: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 must comprise the changes made compared with the wild type enzyme. For instance, variants of SEQ ID NOs: 8, 10 and 12 must comprise the same residues as SEQ ID NOs: 8, 10 and 12 at positions 98, 144, 306 and 330. Variants of SEQ ID NOs: 14 and 16 must comprise the same residues as SEQ ID NOs: 14 and 16 at positions 42, 51, 98, 144, 306 and 330. Variants of SEQ ID NOs: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 must comprise the same residues as SEQ ID NOs: 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 at positions 51, 98, 144, 184, 306 and 330.

[0241] The nucleic acid handling enzyme is typically attached to the surface via one or more of its surface accessible residues having side chains of a preferred orientation. The following residues in SEQ ID NO: 6 should not be used as attachment points because they are buried and inaccessible to solvent or attachment to a linker will result is severe disruption of the protein global structure: 13, 101, 104, 111, 114, 115, 117, 190, 194, 197, 208, 232, 233, 236, 246, 247, 248, 258, 259, 261, 262, 283, 298, 309, 310, 347, 406, and 267. The following residues in SEQ ID NO: 6 are close to DNA binding groove and would not be a good location to attach a linker even though they are surface accessible: 21, 22, 103, 105, 106, 107, 109, 110, 114, 115, 117, 118, 231, 242, 252, 256, 257, 285, 287, 288, 289, 300, 301, 302, 304, 305, 307, 355, 356, 357, 358, 359, 360 368, 369 and 371. All the remaining residues in SEQ ID NO: 6 could potentially be used as attachment points.

[0242] The following residues in SEQ ID NO: 6 are preferred attachment points because they provide a side chain, are exposed and are solvent accessible and mutation of the side chain would be predicted to lead to little disruption of the overall protein structure: 8, 9, 37, 38, 39, 41, 43, 44, 45, 47, 76, 77, 96, 150, 151, 153, 156, 159, 161, 171, 173, 176, 178, 179, 184, 195, 198, 199, 200, 203, 209, 218, 222, 225, 227, 256, 273, 275, 277, 278, 280, 281, 282, 285, 292, 293, 311, 313, 316, 318, 321, 326, 327, 328, 332, 335, 338, 339, 340, 342, 345, 353, 374, 381, 385, 387, 389, 390, 395, 397, 401, 417, 420, 423, 424, 429, 432, 437, 438, 441, 445, 448, 452, 456, 458, 459, 465, 458, 459, 462, 466 and 467. The following residues are most preferred as attachment points: 184, 83, 42, 94, 90, 188, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 215, 216, 217, 218, 219, 220, 221, 222 and 223. Any of the preferred attachment residues may be modified by substitution. Preferably one or more of the preferred attachment residues is substituted with cysteine. Preferred substitutions in SEQ ID NO: 6 or a variant thereof include, but are not limited to, M184C, A83C, V42C, V94C, S90C, Y188C, Y188C, A219C, A219C, M218C and M218C. As discussed in more detail below, the enzyme may be attached to the surface via more than one residue.

Surface

[0243] The constructs of the invention comprise a surface. Any surface that can be attached to the nucleic acid binding protein via the one or more accessible cysteine residues can be used in accordance with the invention. Suitable surfaces include, but are not limited to beads (such as agarose or magnetic beads), column supports, immobilised metal affinity matrices (such as

[0244] NiNTA, Streptavidin-tag or cobalt), glutathione sepharose, dextrin sepharose, affinity matrix, IgG sepharose, activated thiol-sepharose, metal films (such as gold solid supports), nanoparticles, glass, treated glass, plastics, resin surfaces, solid state nanopores, sensor surfaces, single molecular detectors and transmembrane protein pores.

[0245] Suitable solid state pores include, but are not limited to, silicon nitride pores, silicon dioxide pores and graphene pores. Other suitable solid state pores and methods of producing them are discussed in U.S. Pat. No. 6,464,842, WO 03/003446, WO 2005/061373, U.S. Pat. No. 7,258,838, U.S. Pat. No. 7,466,069, U.S. Pat. No. 7,468,271 and U.S. Pat. No. 7,253,434.

[0246] The surface is preferably derived from a transmembrane protein pore. A transmembrane protein pore is a polypeptide or a collection of polypeptides that permits ions driven by an applied potential to flow from one side of a membrane to the other side of the membrane. The pore preferably permits nucleotides to flow from one side of a membrane to the other along the applied potential. The pore preferably allows a nucleic acid, such as DNA or RNA, to be pushed or pulled through the pore.

[0247] The pore may be a monomer or an oligomer. The pore is preferably made up of several repeating subunits, such as 6, 7 or 8 subunits. The pore is more preferably a heptameric pore. The pore typically comprises a barrel or channel through which the ions may flow. The subunits of the pore typically surround a central axis and contribute strands to a transmembrane β barrel or channel or a transmembrane α-helix bundle or channel.

[0248] The barrel or channel of the pore typically comprises amino acids that facilitate interaction with nucleotides or nucleic acids. These amino acids are preferably located near a constriction of the barrel or channel. The pore typically comprises one or more positively charged amino acids, such as arginine, lysine or histidine. These amino acids typically facilitate the interaction between the pore and nucleotides or nucleic. The nucleotide detection can be facilitated with an adaptor. This is discussed in more detail below.

[0249] Pores for use in accordance with the invention can be β-barrel pores, α-helix bundle pores or solid state pores. β-barrel pores comprise a barrel or channel that is formed from β-strands. Suitable β-barrel pores include, but are not limited to, β-toxins, such as α-hemolysin, anthrax toxin and leukocidins, and outer membrane proteins/porins of bacteria, such as *Mycobacterium smegmatis* porin A (MspA), outer membrane porin F (OmpF), outer membrane porin G (OmpG), outer membrane phospholipase A and *Neisseria* autotransporter lipoprotein (NalP). α-helix bundle pores comprise a barrel or channel that is formed from α-helices. Suitable α-helix bundle pores include, but are not limited to, inner membrane proteins and α outer membrane proteins, such as WZA.

[0250] The surface may be a pore itself. Alternatively, if the pore is an oligomer, the surface may be a pore subunit. The constructs of the invention may be part of a pore. Alternatively, the construct may be isolated, substantially isolated, purified or substantially purified as described above.

[0251] The pore or subunit is preferably derived from α-hemolysin (α-HL). The wild type α-HL pore is formed of seven identical monomers or subunits (i.e. it is heptameric). The sequence of one wild type monomer or subunit of α-hemolysin is shown in SEQ ID NO: 2. The surface in the constructs of the invention preferably comprises the sequence shown in SEQ ID NO: 2 or a variant thereof. Amino acids 1, 7 to 21, 31 to 34, 45 to 51, 63 to 66, 72, 92 to 97, 104 to 111, 124 to 136, 149 to 153, 160 to 164, 173 to 206, 210 to 213, 217, 218, 223 to 228, 236 to 242, 262 to 265, 272 to 274, 287 to 290 and 294 of SEQ ID NO: 2 form loop regions. Residues 113 and 147 of SEQ ID NO: 2 form part of a constriction of the barrel or channel of α-HL. The nucleic acid binding protein is preferably attached to one or more of amino acids 8, 9, 17, 18, 19, 44, 45, 50 and 51 of SEQ ID NO: 2.

[0252] A variant of SEQ ID NO: 2 is a subunit that has an amino acid sequence which varies from that of SEQ ID NO: 2 and which retains its pore forming ability. The ability of a variant to form a pore can be assayed using any method known in the art. For instance, the variant may be inserted into a membrane along with other appropriate subunits and its ability to oligomerise to form a pore may be determined. Methods are known in the art for inserting subunits into membranes, such as lipid bilayers. For example, subunits may be suspended in a purified form in a solution containing a lipid bilayer such that it diffuses to the lipid bilayer and is inserted by binding to the lipid bilayer and assembling into a functional state. Alternatively, subunits may be directly inserted into the membrane using the "pick and place" method described in M. A. Holden, H. Bayley. J. Am. Chem. Soc. 2005, 127, 6502-6503 and International Application No. PCT/GB2006/001057 (published as WO 2006/100484).

[0253] The variant may include modifications that facilitate covalent attachment to or interaction with the nucleic acid binding protein. The variant preferably comprises one or more reactive cysteine residues that facilitate attachment to the nucleic acid binding protein. For instance, the variant may include a cysteine at one or more of positions 8, 9, 17, 18, 19, 44, 45, 50, 51, 237, 239 and 287 and/or on the amino or carboxy terminus of SEQ ID NO: 2. Preferred variants comprise a substitution of the residue at position 8, 9, 17, 237, 239 and 287 of SEQ ID NO: 2 with cysteine (K8C, T9C, N17C, K237C, S239C or E287C).

[0254] The variant may also include modifications that facilitate any interaction with nucleotides or facilitate orientation of a molecular adaptor as discussed below. The variant may also contain modifications that facilitate covalent attachment of a molecular adaptor.

[0255] In particular, the variant preferably has a glutamine at position 139 of SEQ ID NO: 2. The variant preferably has a cysteine at position 119, 121 or 135 of SEQ ID NO: 2. SEQ ID NO: 4 shows the sequence of SEQ ID NO: 2 except that it has an cysteine at position 135 (L135C) and a glutamine at position 139 (N139Q). SEQ ID NO: 4 or a variant thereof may be used to form a pore in accordance with the invention. The variant may have an arginine at position 113 of SEQ ID NO: 2.

[0256] The variant may be a naturally occurring variant which is expressed naturally by an organism, for instance by a *Staphylococcus* bacterium, or expressed recombinantly by a bacterium such as *Escherichia coli*. Variants also include non-naturally occurring variants produced by recombinant technology. Over the entire length of the amino acid sequence of SEQ ID NO: 2 or 4, a variant will preferably be at least 50% homologous to that sequence based on amino acid identity. More preferably, the variant polypeptide may be at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% and more preferably at least 95%, 97% or 99% homologous based on amino acid identity to the amino acid sequence of SEQ ID NO: 2 or 4 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95%, amino acid identity over a stretch of 200 or more, for example 230, 250, 270 or 280 or more, contiguous amino acids ("hard homology"). Homology can be measured as described above.

[0257] Amino acid substitutions may be made to the amino acid sequence of SEQ ID NO: 2 or 4 in addition to those discussed above, for example up to 1, 2, 3, 4, 5, 10, 20 or 30 substitutions. Conservative substitutions may be made, for example, according to Table 1 above.

[0258] One or more amino acid residues of the amino acid sequence of SEQ ID NO: 2 may additionally be deleted from the polypeptides described above. Up to 1, 2, 3, 4, 5, 10, 20 or 30 residues may be deleted, or more.

[0259] Variants may fragments of SEQ ID NO: 2 or 4. Such fragments retain pore forming activity. Fragments may be at least 50, 100, 200 or 250 amino acids in length. A fragment preferably comprises the pore forming domain of SEQ ID NO: 2 or 4. Fragments typically include residues 119, 121, 135. 113 and 139 of SEQ ID NO: 2 or 4.

[0260] One or more amino acids may be alternatively or additionally added to the polypeptides described above. An extension may be provided at the amino terminus or carboxy terminus of the amino acid sequence of SEQ ID NO: 2 or 4 or a variant or fragment thereof. The extension may be quite short, for example from 1 to 10 amino acids in length. Alternatively, the extension may be longer, for example up to 50 or 100 amino acids. A carrier protein may be fused to a pore or variant.

[0261] As discussed above, a variant of SEQ ID NO: 2 or 4 is a subunit that has an amino acid sequence which varies from that of SEQ ID NO: 2 or 4 and which retains its ability to form a pore. A variant typically contains the regions of SEQ ID NO: 2 or 4 that are responsible for pore formation. The pore forming ability of α-HL, which contains a β-barrel, is provided by β-strands in each subunit. A variant of SEQ ID NO: 2 or 4 typically comprises the regions in SEQ ID NO: 2 that form β-strands. The amino acids of SEQ ID NO: 2 or 4 that form β-strands are discussed above. One or more modifications can be made to the regions of SEQ ID NO: 2 or 4 that form β-strands as long as the resulting variant retains its ability to form a pore. Specific modifications that can be made to the β-strand regions of SEQ ID NO: 2 or 4 are discussed above.

[0262] A variant of SEQ ID NO: 2 or 4 preferably includes one or more modifications, such as substitutions, additions or deletions, within its α-helices and/or loop regions. Amino acids that form α-helices and loops are discussed above.

[0263] The variant may be modified for example by the addition of histidine or aspartic acid residues to assist its identification or purification 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.

[0264] Variants may also comprise any of the non-specific modifications discussed above for the nucleic acid binding protein. Subunits or pores can be made as discussed above.

Attachment

[0265] The nucleic acid binding protein is attached to the surface via one or more accessible cysteine residues. The nucleic acid binding protein may be attached to the surface at more than one, such as two or three, points. Attaching the nucleic acid binding protein to the surface at more than one point can be used to constrain the mobility of the protein. For instance, multiple attachments may be used to constrain the freedom of the protein to rotate or its ability to move away from the surface.

[0266] If the surface is the subunit of an oligomeric pore, the subunit may be in a monomeric form when it is attached to the nucleic acid binding protein (post expression modification). Alternatively, the subunit may be part of an oligomeric pore when it is attached to the nucleic acid binding protein (post oligomerisation modification).

[0267] The nucleic acid binding protein can be attached to the surface using any method known in the art. The nucleic acid binding protein and surface may be produced separately and then attached together. If the surface is itself a protein, such as a pore subunit, the two components may be attached in any configuration. For instance, they may be attached via their terminal (i.e. amino or carboxy terminal) amino acids. Suitable configurations include, but are not limited to, the amino terminus of the nucleic acid binding protein being attached to the carboxy terminus of the surface and vice versa. Alternatively, the two components may be attached via amino acids within their sequences. For instance, the nucleic acid binding protein may be attached to one or more amino acids in a loop region of the surface. In a preferred embodiment, terminal amino acids of the nucleic acid binding protein are attached to one or more amino acids in the loop region of the surface. Terminal amino acids and loop regions are discussed above.

[0268] The nucleic acid binding protein is preferably chemically fused to the surface. A nucleic acid binding protein is chemically fused to a surface if the two parts are chemically attached, for instance via a linker molecule. Any method of chemical fusion or attachment can be used. Suitable methods include, but are not limited to, histidine tag binding to a metal affinity matrix, Ni-NTA, biotin binding to streptavidin, antibody binding to an antigen, primary amine coupling, GST tags binding to glutathione, MBP tags binding to dextrin, Protein A binding to IgG, reaction between thiols, nucleic acid hybridization linkers and cysteine linkage. DNA hybridization linkers and cysteine linkage are discussed in more detail below. The nucleic acid binding protein is preferably covalently attached to the surface.

[0269] If the surface is a protein, the nucleic acid binding protein may be genetically fused to the surface. A nucleic acid binding protein is genetically fused to a protein surface if the whole construct is expressed from a single polynucleotide sequence. The coding sequences of the nucleic acid binding protein and surface may be combined in any way to form a single polynucleotide sequence encoding the construct.

[0270] The nucleic acid binding protein and surface may be genetically fused in any configuration, such as via their terminal amino acids. The amino acid sequence of the nucleic acid binding protein is typically added in frame into the amino acid sequence of the surface. In a preferred embodiment, the nucleic acid binding protein is inserted into a loop region of a transmembrane protein pore or pore subunit. In an especially preferred embodiment, the nucleic acid binding protein is inserted between amino acids, 18 and 19, 44 and 45 or 50 and 51 of SEQ ID NO: 2.

[0271] The nucleic acid binding protein retains its ability to bind nucleic acids. This ability is typically provided by its secondary structural elements (α-helices and β-strands) and tertiary structural elements. In order to avoid adversely affecting the nucleic acid binding ability of the protein, it is preferably attached to the surface in a manner that does not affect its secondary or tertiary structure.

[0272] If the surface is a pore or pore subunit, the pore or pore subunit retains it ability to form pores. The pore forming ability of subunits is typically provided by their α-helices and β-strands. β-barrel pores comprise a barrel or channel that is formed from n-strands, whereas α-helix bundle pores comprise a barrel or channel that is formed from α-helices. The α-helices and β-strands are typically connected by loop regions. In order to avoid affecting the pore forming ability of the subunit, the nucleic acid binding protein is preferably attached to a loop region of the subunit. The loop regions of specific subunits are discussed in more detail above.

[0273] The nucleic acid binding protein may be attached directly to the surface. For instance, native and/or non-native accessible cysteine residues can be attached directly to activated thiol-sepharose.

[0274] The nucleic acid binding protein can be attached to the surface at one or more positions, such as at one, two, three or four positions. The nucleic acid binding protein is preferably attached to the surface at one or two positions. After removal of native cysteine residues from the nucleic acid binding protein, one or more cysteine residuess can be incorporated into the protein at specific positions for the attachments. Attachments can be done either by direct cross linking of the cytseine residues in the nucleic acid binding protein to cysteines in the surface (i.e. via a disulphide bond) or by using cross linkers. Attachment at two positions can reduce the flexibility of the complex and can fix the nucleic acid binding protein on the surface in a chosen specific orientation.

[0275] In a preferred embodiment, the construct comprises (1) exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof with the substitutions M184C and V94C and (2) α-HL (SEQ ID NO: 2) or a variant thereof having the substitutions N17C and E287C. The M184C position in SEQ ID NO: 6 or a variant thereof is cross linked to the N17C position in SEQ ID NO: 2 or a variant thereof either by direct cysteine cross linking or by short linkers. The V94C position in SEQ ID NO: 6 or a variant therof is cross linked to the E287C position in SEQ ID NO: 2 or a variant thereof by a different linker. In this example, the nucleic acid binding protein is attached to the surface via at two distant positions.

[0276] Table 2 below shows the positions of other preferred first and second attachment points.

	First A SEQ ID NO: 6 or variant thereof	First B (attached to First A) SEQ ID NO: 2 or variant thereof	Second A SEQ ID NO: 6 or variant thereof	Second B (attached to Second A) SEQ ID NO: 2 or variant thereof
1	M184C	N17C	A83C	K237C
2	M184C	N17C	S90C	S239C
3	M184C	N17C	V94C	E287C
4	V42C	T19C	A83C	K237C
5	V42C	T19C	S90C	S239C
6	V42C	T19C	V94C	E287C

[0277] The nucleic acid binding protein is preferably attached to the surface using one or more, such as two or three, linkers. The one or more linkers may be designed to constrain the mobility of the nucleic acid binding protein. The linkers are typically attached to the one or more accessible cysteine residues in the nucleic acid binding protein. The linkers may be attached to one or more reactive groups, such as cysteine residues, reactive lysine residues or non-natural amino acids, in the surface. Suitable linkers are well known in the art. Suitable linkers include, but are not limited to, chemical crosslinkers and peptide linkers. Preferred chemical crosslinkers are nucleic acid hybridization linkers. The length, flexibility and hydrophilicity of the nucleic acid hybridization linkers are typically designed such that they do not to disturb the functions of the nucleic acid binding protein and surface. An advantage of using hybridization linkers is that the formation of unwanted dimers (surface-surface or protein-protein) is minimized. The nucleic acid hybridization linkers can comprise any of the nucleic acids discussed above. For instance, they may comprise deoxyribonucleic acid (DNA), ribonucleic acid (RNA) or any synthetic nucleic acid known in the art, such as peptide nucleic acid (PNA), glycerol nucleic acid (GNA), threose nucleic acid (TNA), locked nucleic acid (LNA) or other synthetic polymers with nucleotide side chains. The linkers can also be modified such they react with one another once they have hybridised. Alternatively, agents may be used to crosslink the linkers once they have hybridised to one another.

[0278] Preferred nucleic acid hybridization linkers correspond to the first 15, 25 or 35 nucleotides from the 5' end of SEQ ID NO: 57. The linker preferably also has TT at the 3' end to provide extra flexibility. At the 3' end, the linkers have a group, such as maleimide or Thiol, that allows the linker to be attached to the nucleic acid binding protein or surface. Maleimide or Thiol modified oliognucleotides can be obtained commercially, for instance from ATDBio. More preferred linkers are shown in SEQ ID NOs: 58, 59 and 60. Complementary linkers are shown in SEQ ID NOs: 61, 62 and 63. SEQ ID NO: 58, 59 or 60 may be attached to one of the nucleic acid binding protein and surface and the complementary linker (SEQ ID NO: 61, 62 or 63 respectively) is attached to the other of the nucleic acid binding protein and surface. The nucleic acid binding protein and surface can then be attached together by hybridizing the linkers.

[0279] The stability of the hybridization depends on the melting temperature of the hybridizing linkers. Depending on the application and the required stability, this can be optimized by changing the sequences of the linkers (e.g. changing the linkers to more GC rich will increase their melting temperature and hence the stability), the length of the linkers (i.e. increasing their length will increase the stability) or the reaction conditions (e.g. increasing their concentration will increase the stability).

[0280] For maximum stability of hybridization, it is desirable to have long hybridizing linkers with high melting temperatures, for example linkers more than 15 nucleotides in length, particularly 15 to 45 nucleotides in length, such as 15, 20, 25, 30, 35, 40 or 45 nucleotides in length. However, the use of long linkers increases the distance between the moieties. This may be disadvantageous because interaction between the moieties is disrupted or because proximity is required for the surface to detect a substrate which has been released from the nucleic acid binding protein. Increased distance may be advantageous as it may prevent aggregation or electrostatic interactions and may permit flexing. The disadvantages of the increased distance can be overcome by changing the orientation of the nucleic acid attachment. Most preferably, the linkers comprise a nucleic acid that is from 6 to 15 nucleotides in lenght, such as 6, 8 or 10 nucleotides long.

[0281] The hybridization linkers preferably have an affinity of from 1 fM to 1 uM at concentrations of from 1 pM to 1 mM. The linkers more preferably have an affinity of from 1 fM to 10nM at concnetrations of from 1 pM to 1 uM. The linkers most preferably have an affinity of from 1 pM to 100 pM at concentrations of from 100 pM to 10 nM.

[0282] A preferred linker is shown in the SEQ ID NO: 64. The 3' end of this linker can be attached to a cysteine residue on the surface. The linker preferably also has TTTTT at the 3' end to provide extra flexibility. The 5' end of SEQ ID NO: 65 can then be attached to the nucleic acid binding protein. In this example, the 3' end of the SEQ ID NO 64 is complementary to a stretch of sequence at the 5' end of SEQ ID NO: 65.

[0283] Once the linkers are hybridized to each other, they melt (fall apart) under certain conditions (for example, at high temperatures or lower salt conditions) unless there is a permanent bond between the two linkers. To form a permanent bond, the linkers are preferably modified such they react with one another once they have hybridized. Each linker preferably contains a group capable of forming a covalent bond with a group in another linker. A pair of linkers can be linked by one or more covalent bonds, for example one, two or three, covalent bonds.

[0284] Typically the bond will be a simple disulfide bond between the two linkers. The linkers can also be modified to incorporate thiol groups at one or more, such as two, positions. Depending on the application and preferences, thiols groups can be internal or terminal.

[0285] Linkers can also be modified either internally or terminally to include one or more, such as two, iodoacetamide groups. A hybridizing linker with one or more iodoactamide groups can be covalently linked to thiols on the complementary hybridizing linker.

[0286] Linkers can also be modified with alkene groups. One or more internal or terminal alkene groups in preferred positions can be subjected to olefin metathesis to make a covalent bond between the alkenes in the hybridization linkers.

[0287] If necessary, a small linker can be added between the linker and the reactive groups, such as thiol groups, iodoacetamide groups and alkene groups, to obtain the proper distances necessary to make an efficient covalent bond between the linkers.

[0288] In a preferred embodiment, the covalent bond between the linkers can be made using the click chemistry. Click chemistry is a term first introduced by Kolb et al. in 2001 to describe an expanding set of powerful, selective, and modular building blocks that work reliably in both small- and large-scale applications (Kolb H C, Finn, M G, Sharpless K B, Click chemistry: diverse chemical function from a few good reactions, Angew. Chem. Int. Ed. 40 (2001) 2004-2021). They have defined the set of stringent criteria for click chemistry as follows: "The reaction must be modular, wide in scope, give very high yields, generate only inoffensive byproducts that can be removed by nonchromatographic methods, and be stereospecific (but not necessarily enantioselective). The required process characteristics include simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation. Purification if required must be by nonchromatographic methods, such as crystallization or distillation, and the product must be stable under physiological conditions".

[0289] Suitable example of click chemistry include, but are not limited to, the following:

[0290] (a) copper-free variant of the 1,3 dipolar cycloaddition reaction, where an azide reacts with an alkyne under strain, for example in a cyclooctane ring;

[0291] (b) the reaction of an oxygen nucleophile on one linker with an epoxide or aziridine reactive moiety on the other; and

[0292] (c) the Staudinger ligation, where the alkyne moiety can be replaced by an aryl phosphine, resulting in a specific reaction with the azide to give an amide bond.

[0293] Preferably the click chemistry reaction is the Cu (I) catalysed 1,3 dipolar cycloaddition reaction between an alkyne and an azide. Nucleic acid bases have already been synthesized incorporating azide and alkyne groups in preferred positions (for example Kocalka P, El-Sagheer A H, Brown T, Rapid and efficient DNA strand cross-linking by click chemistry, Chembiochem. 2008. 9(8):1280-5).

[0294] If nucleotides within the linkers' nucleic acid acid regions are modified to include groups that can form covalent bonds, the modified nucleotides are preferably offset from one another by one nucleotide in order to achieve the link. This follows the published work of Tom Brown (Kocalka et al. (2008) ChemBiochem 9 8 1280-1285).

[0295] In a preferred embodiment, a single azide group (SEQ ID NO: 66) or more such as two (SEQ ID NO: 67) can be incorporated into uracil bases at specific places in a 15 base deoxyribonucleic acid sequence. The cysteine residues on the surface can then be modified with these azide hybridization linkers using the thiol group at the 5'end (SEQ ID NOs: 66 and 67). Alkyne groups can also be incorporated into uracil bases at preferred positions in sequences complementary to the SEQ ID NOs: 66 and 67 (SEQ ID NOs: 68 and 69 respectively). These sequences can be used to modify the cysteines on the nucleic acid binding protein. Using DNA hybridization followed by 'click chemistry' between the azide and alkyne groups, hybridization linkers can be covalently cross linked.

[0296] The distance between the surface and the nucleic acid binding protein can be modulated by changing the length of the hybridization linkers. The position of the azide and alkyne modified bases then needs to be changed accordingly.

[0297] In a preferred embodiment 6 mer (SEQ ID NO 70), 8 mer (SEQ ID NO 71) or 10 mer (SEQ ID NO 72) DNA in which two uracil bases are modified with azide groups can be attach to the cysteines of the moiety. Complementary sequences of 6 mer (SEQ ID NO 73), 8 mer (SEQ ID NO 74) or 10 mer (SEQ ID NO 75) DNA in which two uracil bases are modified with alkyne groups can be attached to the cysteines of a moiety such as a DNA binding protein. Covalent cross linking between these hybridization linkers will bring the moieties closer to each other than with the hybridization linkers (SEQ ID NO 67 and 69). Incorporation of azide and alkyne groups into uracil base units of DNA has been developed by ATDBio.

[0298] Other preferred chemical crosslinkers are shown in the following Table 3.

TABLE 3

Some preferred linkers

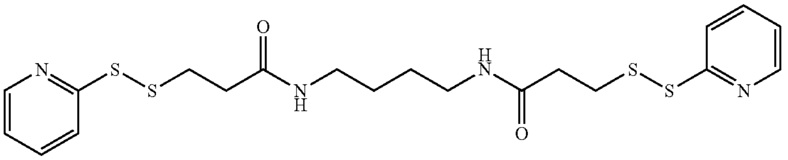
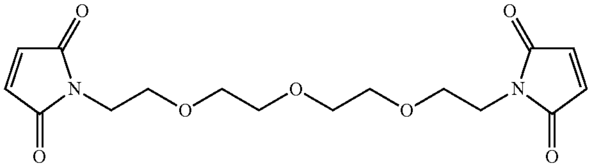
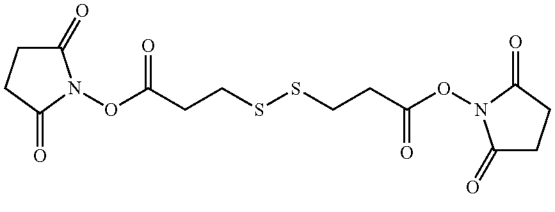
Name	Reacts with	Structure
1,4-Bis[3-(2-pyridyldithio) propionamido]butane	Thiols	
1,11-bis-Maleimidotri-ethyleneglycol	Thiols	
3,3'-Dithiodipropionic acid di(N-hydroxysuccinimide ester)	Primary amines	

TABLE 3-continued

Some preferred linkers

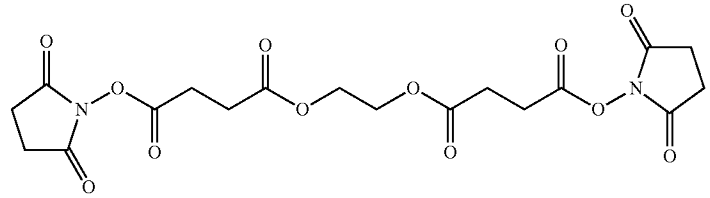
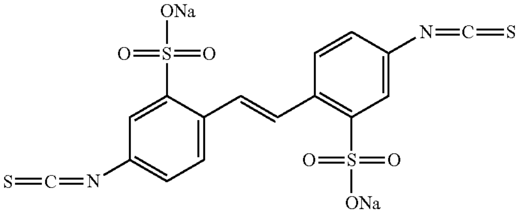
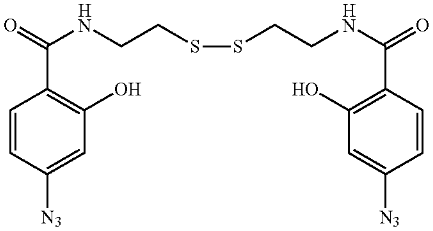
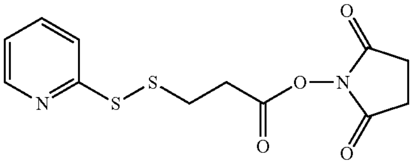
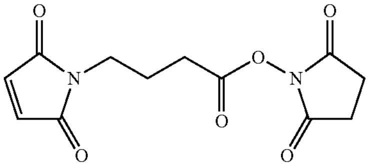
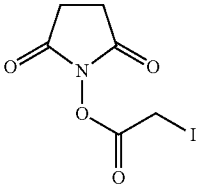
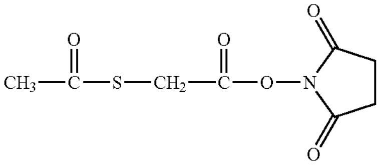
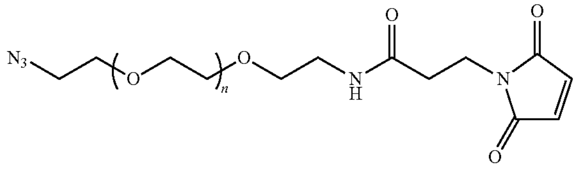
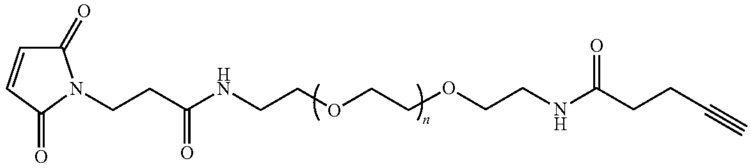
Name	Reacts with	Structure
Ethylene glycol-bis(succinic acid N-hydroxysuccinimide ester)	Primary amines	
4,4'-Diisothiocyanatostilbene-2,2'-disulfonic acid disodium salt	Primary amines	
Bis[2-(4-azidosalicylamido)ethyl] disulfide	Photo-activated, non-specific	
3-(2-Pyridyldithio)propionic acid N-hydroxysuccinimide ester	Thiols, primary amines	
4-Maleimidobutyric acid N-hydroxysuccinimide ester	Thiols, primary amines	
Iodoacetic acid N-hydroxysuccinimide ester	Thiols, primary amines	
S-Acetylthioglycolic acid N-hydroxysuccinimide ester	Thiols, primary amines	

TABLE 3-continued

Some preferred linkers		
Name	Reacts with	Structure
Azide-PEG-maleimide	Thiols, alkyne	 n = 5, 10
Alkyne-PEG-maleimide	Thiols, azide	 n = 6, 10

[0299] Linkers may be attached to the nucleic acid binding protein first and then the surface, the surface first and then the nucleic acid binding protein or the surface and nucleic acid binding protein at the same time. When the linker is attached to a pore subunit (as the surface), it may be a monomeric subunit, part of an oligomer of two or more monomers or part of complete oligomeric pore. It is preferred that the linker is reacted before any purification step to remove any unbound linker.

[0300] The preferred method of attaching the nucleic acid binding protein to the surface is via cysteine linkage. This can be mediated by a bi-functional chemical linker or by a polypeptide linker with a terminal presented cysteine residue. α-HL (SEQ ID NO: 2) lacks native cysteine residues so the introduction of a cysteine into the sequence of SEQ ID NO: 2 enables the controlled covalent attachment of the nucleic acid binding protein to the subunit. Cysteines can be introduced at various positions, such as position K8, T9, N17 or E287 of SEQ ID NO: 2 or at the carboxy terminus of SEQ ID NO: 2. The length, reactivity, specificity, rigidity and solubility of any bi-functional linker may be designed to ensure that the enzyme is positioned correctly in relation to the subunit and the function of both the subunit and enzyme is retained. Suitable linkers include those described above.

[0301] Cross-linkage of subunits or enzymes to themselves may be prevented by keeping the concentration of linker in a vast excess of the nucleic acid binding protein and/or the surface. Alternatively, a "lock and key" arrangement may be used in which two linkers are used. For instance, click chemistry, such as azide alkyne Huisgen cycloaddition, may be used to ensure that the nucleic acid binding protein only binds to the surface and not to itself and vice versa. In a preferred embodiment, the azide-PEG-maleimide and alkyne-PEG-maleimide linkers shown in Table 3 above are used. One is attached to the nucleic acid binding protein and the other is attached to the surface. This ensures that binding only occurs between the nucleic acid binding protein and the surface.

[0302] Only one end of each linker may react together to form a longer linker and the other ends of the linker each react with a different part of the surface (i.e. subunit or monomer).

[0303] In a preferred embodiment, the site of covalent attachment is selected such that, when the construct is used to form a pore, the nucleic acid binding protein handles a target nucleic acid sequence in such a way that a proportion of the nucleotides in the target sequence interacts with the pore. Nucleotides are then distinguished on the basis of the different ways in which they affect the current flowing through the pore during the interaction.

[0304] There are a number of ways that pores can be used to sequence nucleic acid molecules. One way involves the use of an exonuclease enzyme, such as a deoxyribonuclease. In this approach, the exonuclease enzyme is used to sequentially detach the nucleotides from a target nucleic strand. The nucleotides are then detected and discriminated by the pore in order of their release, thus reading the sequence of the original strand. For such an embodiment, the exonuclease enzyme is preferably attached to a pore subunit such that a proportion of the nucleotides released from the target nucleic acid is capable of entering and interacting with the barrel or channel of a pore comprising the subunit. The exonuclease is preferably attached to the subunit at a site in close proximity to the part of the subunit that forms the opening of the barrel of channel of the pore. The exonuclease enzyme is more preferably attached to the subunit such that its nucleotide exit trajectory site is orientated towards the part of the subunit that forms part of the opening of the pore.

[0305] Another way of sequencing nucleic acids involves the use of an enzyme that pushes or pulls the target nucleic acid strand through the pore in combination with an applied potential. In this approach, the ionic current fluctuates as a nucleotide in the target strand passes through the pore. The fluctuations in the current are indicative of the sequence of the strand. For such an embodiment, the enzyme is preferably attached to a pore subunit such that it is capable of pushing or pulling the target nucleic acid through the barrel or channel of a pore comprising the subunit and does not interfere with the flow of ionic current through the pore. The enzyme is preferably attached to the subunit at a site in close proximity to the part of the subunit that forms part of the opening of the barrel of channel of the pore. The enzyme is more preferably attached to the subunit such that its active site is orientated towards the part of the subunit that forms part of the opening of the pore.

[0306] A third way of sequencing a nucleic acid strand is to detect the byproducts of a polymerase in close proximity to a pore detector. In this approach, nucleoside phosphates (nucleotides) are labelled so that a phosphate labelled species is released upon the addition of a polymerase to the nucleotide strand and the phosphate labelled species is detected by the pore. The phosphate species contains a specific label for each nucleotide. As nucleotides are sequentially added to the nucleic acid strand, the bi-products of the base addition are detected. The order that the phosphate labelled species are detected can be used to determine the sequence of the nucleic acid strand.

[0307] The nucleic acid binding protein is preferably attached to the part of a pore or a subunit thereof that forms part of the cis side of a pore. In electrophysiology, the cis side is the grounded side by convention. If a hemolysin pore is inserted correctly into an electrophysiology apparatus, the Cap region is on the cis side. It is well known that, under a positive potential, nucleotides will migrate from the cis to the trans side of pores used for stochastic sensing. Positioning the nucleic acid binding protein at the cis side of a pore allows it to handle the target nucleic acid such that a proportion of the nucleotides in the sequence enters the barrel or channel of the pore and interacts with it. Preferably, at least 20%, at least 40%, at least 50%, at least 80% or at least 90% of the nucleotides in the sequence enters the barrel or channel of the pore and interacts with it.

[0308] The site and method of covalent attachment is preferably selected such that mobility of the nucleic acid binding protein is constrained. This helps to ensure that the protein handles the target nucleic acid sequence in such a way that a proportion of the nucleotides in the target sequence interacts with the pore. For instance, constraining the ability of nucleic acid binding protein to move means that its active site can be permanently orientated towards the part of the subunit that forms part of the opening of the barrel of channel of the pore. The mobility of the nucleic acid binding protein may be constrained by increasing the number of points at which the protein is attached to the surface and/or the use of specific linkers.

Preferred Conjugates

[0309] In a preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having at least one native accessible cysteine residue removed attached to α-HL (SEQ ID NO: 2) or a variant thereof via one or more accessible cysteine residues. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residue removed attached to α-HL (SEQ ID NO: 2) or a variant thereof via one or more accessible cysteine residues that have been substituted into the exonuclease. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having only the native cysteine residues at position 144 and/or 330 attached to α-HL (SEQ ID NO: 2) or a variant thereof via one or more of the native cysteine residues. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having no native cysteine residues and A83C attached to α-HL (SEQ ID NO: 2) or a variant thereof having E287C via the non-native cysteine residues.

[0310] In a preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having at least one native accessible cysteine residue removed attached to α-HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4) via one or more accessible cysteine residues. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residue removed attached to α-HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4) via one or more accessible cysteine residues that have been substituted into the exonuclease. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6)or a variant thereof having only the native cysteine residues at position 144 and/or 330 attached to α-HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4) via one or more of the native cysteine residues. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having no native cysteine residues and A83C attached to α-HL or a variant therof having L135C, N139Q and E287C via the non-native cysteine residues.

[0311] In a preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having at least one native accessible cysteine residue removed attached to α-HL (SEQ ID NO: 2) or a variant thereof via one or more hybridization linkers attached to accessible cysteine residues in the exonuclease. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residue removed attached to α-HL (SEQ ID NO: 2) or a variant thereof via one or more hybridization linkers attached to accessible cysteine residues that have beens substituted into the exonuclease. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having only the native cysteine residues at position 144 and/or 330 attached to α-HL (SEQ ID NO: 2) or a variant thereof via hybridization linkers attached to one or more of the native cysteine residue. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having no native cysteine residues and A83C attached to α-HL (SEQ ID NO: 2) or a variant thereof having E287C via a hybridization linker attached to the non-native cysteine residues.

[0312] In another preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having at least one native accessible cysteine residue removed attached to α-HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4). via one or more hybridization linkers attached to accessible cysteine residues in the exonuclease. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residue removed attached to α-HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4) via one or more hybridization linkers attached to accessible cysteine residues that have beens substituted into the exonuclease. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having only the native cysteine residues at position 144 and/or 330 attached to α-HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4) via hybridization linkers attached to the one or more native cysteine residue. In a more preferred embodiment, the construct comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having no native cysteine residues and A83C attached to α-HL or a variant thereof having L135C, N139Q and E287C via a hybridization linker attached to the non-native cysteine residues.

Modified Pores

[0313] The invention also provides modified pores for use in sequencing nucleic acids. The pores comprise at least one construct of the invention in which the surface is a transmembrane protein pore or a subunit thereof. The pores may comprise more than one, such as 2, 3 or 4, constructs of the invention.

[0314] A pore of the invention may be isolated, substantially isolated, purified or substantially purified. A pore of the invention is isolated or purified if it is completely free of any other components, such as lipids or other pores. A pore is substantially isolated if it is mixed with carriers or diluents which will not interfere with its intended use. For instance, a pore is substantially isolated or substantially purified if it present in a form that comprises less than 10%, less than 5%, less than 2% or less than 1% of other components, such as lipids or other pores. Alternatively, a pore of the invention may be present in a lipid bilayer or in a surfactant micelle.

[0315] The nucleic acid binding protein, which is preferably a nucleic acid handling enzyme, attached to the construct handles a target nucleic acid sequence in such a way that a proportion of the nucleotide in the target sequence interacts with the pore, preferably the barrel or channel of the pore. Nucleotides are then distinguished on the basis of the different ways in which they affect the current flowing through the pore during the interaction.

[0316] The fixed nature of the nucleic acid binding protein means that a target nucleic acid sequence is handled by the pore in a specific manner. For instance, each nucleotide may be digested from one of the target sequence in a processive manner or the target sequence may be pushed or pulled through the pore. This ensures that a proportion of the nucleotides in the target nucleic acid sequence interacts with the pore and is identified. The lack of any interruption in the signal is important when sequencing nucleic acids. In addition, the fixed nature of the enzyme and the pore means they can be stored together, thereby allowing the production of a ready-to-use sensor.

[0317] In a preferred embodiment, an exonuclease enzyme, such as a deoxyribonuclease, is attached to the pore such that a proportion of the nucleotides is released from the target nucleic acid and interacts with the barrel or channel of the pore. In another preferred embodiment, an enzyme that is capable of pushing or pulling the target nucleic acid sequence through the pore is attached to the pore such that the target nucleic acid sequence is pushed or pulled through the barrel or channel of the pore and a proportion of the nucleotides in the target sequence interacts with the barrel or channel. In this embodiment, the nucleotides may interact with the pore in blocks or groups of more than one, such as 2, 3 or 4. Suitable enzymes include, but are not limited to, polymerases, nucleases, helicases and topoisomerases, such as gyrases. In each embodiment, the enzyme is preferably attached to the pore at a site in close proximity to the opening of the barrel of channel of the pore. The enzyme is more preferably attached to the pore such that its active site is orientated towards the opening of the barrel of channel of the pore. This means that a proportion of the nucleotides of the target nucleic acid sequence is fed in the barrel or channel. The enzyme is preferably attached to the cis side of the pore.

[0318] The modified pore may be derived from any of the transmembrane protein pores discussed above, including the β-barrel pores and α-helix bundle pores.

[0319] For constructs comprising the sequence shown in SEQ ID NO: 2 or a variant thereof, the pore typically comprises an appropriate number of additional subunits comprising the sequence shown in SEQ ID NO: 2 or a variant thereof. A preferred pore of the invention comprises one construct comprising the sequence shown in SEQ ID NO: 2 or a variant thereof and six subunits comprising the sequence shown in SEQ ID NO: 2 or a variant thereof. The pore may comprise one or more subunits comprising the sequence shown in SEQ ID NO: 4 or a variant thereof. SEQ ID NO: 4 shows the sequence of SEQ ID NO: 2 except that it has an cysteine at position 135 (L135C) and a glutamine at position 139 (N139Q). A variant of SEQ ID NO: 4 may differ from SEQ ID NO: 4 in the same way and to the same extent as discussed for SEQ ID NO: 2 above. A preferred pore of the invention comprises one construct comprising the sequence shown in SEQ ID NO: 2 or a variant thereof and six subunits comprising the sequence shown in SEQ ID NO: 4 or a variant thereof.

[0320] The pores may comprise a molecular adaptor that facilitates the interaction between the pore and the nucleotides or the target nucleic acid sequence. The presence of the adaptor improves the host-guest chemistry of the pore and nucleotides released from or present in the target nucleic acid sequence. The principles of host-guest chemistry are well-known in the art. The adaptor has an effect on the physical or chemical properties of the pore that improves its interaction with nucleotides. The adaptor typically alters the charge of the barrel or channel of the pore or specifically interacts with or binds to nucleotides thereby facilitating their interaction with the pore.

[0321] The adaptor mediates the interaction between nucleotides released from or present in the target nucleic acid sequence and the pore. The nucleotides preferably reversibly bind to the pore via or in conjunction with the adaptor. The nucleotides most preferably reversibly bind to the pore via or in conjunction with the adaptor as they pass through the pore across the membrane. The nucleotides can also reversibly bind to the barrel or channel of the pore via or in conjunction with the adaptor as they pass through the pore across the membrane. The adaptor preferably constricts the barrel or channel so that it may interact with the nucleotides.

[0322] The adaptor is typically cyclic. The adaptor preferably has the same symmetry as the pore. An adaptor having seven-fold symmetry is typically used if the pore is heptameric (e.g. has seven subunits around a central axis that contribute 14 strands to a transmembrane β barrel). Likewise, an adaptor having six-fold symmetry is typically used if the pore is hexameric (e.g. has six subunits around a central axis that contribute 12 strands to a transmembrane β barrel, or is a 12-stranded β barrel). Any adaptor that facilitates the interaction between the pore and the nucleotide can be used. Suitable adaptors include, but are not limited to, cyclodextrins, cyclic peptides and cucurbiturils. The adaptor is preferably a cyclodextrin or a derivative thereof. The adaptor is more preferably heptakis-6-amino-β-cyclodextrin (am_7-βCD), 6-monodeoxy-6-monoamino-β-cyclodextrin (am_1-βCD) or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu_7-βCD). Table 4 below shows preferred combinations of pores and adaptors.

TABLE 3

Suitable combinations of pores and adaptors

Pore	Number of strands in the transmembrane β-barrel	Adaptor
Leukocidin	16	γ-cyclodextrin (γ-CD)
OmpF	16	γ-cyclodextrin (γ-CD)
α-hemolysin (or a variant thereof discussed above)	14	β-cyclodextrin (β-CD) 6-monodeoxy-6-monoamino-β-cyclodextrin ($am_1\beta$-CD) heptakis-6-amino-β-cyclodextrin (am_7-β-CD) heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu_7-β-CD)
OmpG	14	β-cyclodextrin (β-CD) 6-monodeoxy-6-monoamino-β-cyclodextrin ($am_1\beta$-CD) heptakis-6-amino-β-cyclodextrin (am_7-β-CD) heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu_7-β-CD)
NalP	12	α-cyclodextrin (α-CD)
OMPLA	12	α-cyclodextrin (α-CD)

[0323] The adaptor is preferably covalently attached to the pore. The adaptor can be covalently attached to the pore using any method known in the art. The adaptor may be attached directly to the pore. The adaptor is preferably attached to the pore using a bifunctional crosslinker. Suitable crosslinkers are well-known in the art. Preferred crosslinkers include 2,5-dioxopyrrolidin-1-yl 3-(pyridin-2-yldisulfanyl)propanoate, 2,5-dioxopyrrolidin-1-yl 4-(pyridin-2-yldisulfanyl)butanoate and 2,5-dioxopyrrolidin-1-yl 8-(pyridin-2-yldisulfanyl)octananoate. The most preferred crosslinker is succinimidyl 3-(2-pyridyldithio)propionate (SPDP). Typically, the adaptor is covalently attached to the bifunctional crosslinker before the adaptor/crosslinker complex is covalently attached to the pore but it is also possible to covalently attach the bifunctional crosslinker to the pore before the bifunctional crosslinker/pore complex is attached to the adaptor.

[0324] The site of covalent attachment is selected such that the adaptor facilitates interaction of nucleotides released from or present in the target nucleic acid sequence with the pore and thereby allows detection of nucleotides. For pores based on α-HL, the correct orientation of the adaptor within the barrel or channel of the pore and the covalent attachment of adaptor to the pore can be facilitated using specific modifications to the pore. In particular, every subunit of the pore, including the construct(s), preferably has a glutamine at position 139 of SEQ ID NO: 2. One or more of the subunits of the pore, including the construct(s), may have an arginine at position 113 of SEQ ID NO: 2. One or more of the subunits of the pore, including the construct(s), may have a cysteine at position 119, 121 or 135 of SEQ ID NO: 2 to facilitate attachment of the molecular adaptor to the pore.

Methods of Producing Constructs of the Invention

[0325] The invention also provides methods of producing a construct of the invention. The methods comprise attaching a nucleic acid binding protein comprising one or more accessible cysteines to a surface via the cysteine residues. Any of the nucleic acid binding proteins and surfaces discussed above can be used in the methods. The site of and method of covalent attachment are selected as discussed above.

[0326] As described above, the methods preferably comprise removing all the native accessible cysteine residues from the binding protein and introducing one or more non-native accessible cysteine residues into the binding protein. Alternatively, the method comprises removing all but one or more of the native accessible cysteine residues from the binding protein.

[0327] The methods also comprise determining whether or not the construct is capable of binding nucleic acids. Assays for doing this are described above. If nucleic acids can be bound, the protein and surface have been attached correctly and a construct of the invention has been produced. If nucleic acids cannot be bound, a construct of the invention has not been produced.

Methods of Producing Modified Pores

[0328] The invention also provides methods of producing modified pores of the invention. The modified pore may be formed by allowing at least one construct of the invention in which the surface is a pore subunit to form a pore with other suitable subunits. Any of the constructs, binding proteins, surfaces or pores discussed above can be used in the methods. The site of and method of covalent attachment are selected as discussed above.

[0329] The methods also comprise determining whether or not the pore is capable of binding nucleic acids and detecting nucleotides. The pore may be assessed for its ability to detect individual nucleotides or short chains of nucleotides, such as di- or trinucleotides. Assays for doing this are described above and below. If the pore is capable of binding nucleic acids and detecting nucleotides, the subunit and enzyme have been attached correctly and a pore of the invention has been produced. If a pore cannot bind nucleic acids and detect nucleotides, a pore of the invention has not been produced.

[0330] The pores can be purified by inserting the construct(s) and remaining subunits into synthetic lipid vesicles and allowed them to oligomerise. Methods for inserting the construct(s) and remaining subunits into synthetic vesicles are well known in the art.

[0331] The synthetic vesicles should have similar properties to rabbit cell membranes, but should lack the rabbit cell membrane proteins. The vesicles may comprise any components and are typically made of a blend of lipids. Suitable lipids are well-known in the art. The synthetic vesicles preferably comprise 30% cholesterol, 30% phosphatidylcholine (PC), 20% phosphatidylethanolamine (PE), 10% sphingomyelin (SM) and 10% phosphatidylserine (PS).

[0332] The pore is then purified using an anionic surfactants, such as sodium dodecyl sulphate (SDS), before it is attached to the nucleic acid binding protein.

Methods of Sequencing Nucleic Acids

[0333] The invention also provides methods of sequencing a target nucleic acid sequence. In one embodiment, the method comprises (a) contacting the target sequence with a pore of the invention, which comprises an exonuclease, such that the exonuclease digests an individual nucleotide from one end of the target sequence; (b) contacting the nucleotide with the pore so that the nucleotide interacts with the adaptor; (c) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide; and (d) repeating steps (a) to (c) at the same end of the target sequence and thereby determining the sequence of the target sequence. Hence, the method involves stochastic sensing of a proportion of the nucleotides in a target nucleic acid sequence in a successive manner in order to sequence the target sequence. Individual nucleotides are described above.

[0334] In another embodiment, the method comprises (a) contacting the target sequence with a pore of the invention comprising a nucleic acid handling enzyme so that the target sequence is pushed or pulled through the pore and a proportion of the nucleotides in the target sequence interacts with the pore and (b) measuring the current passing through the pore during each interaction and thereby determining the sequence of the target sequence. Hence, the method involves stochastic sensing of a proportion of the nucleotides in a target nucleic acid sequence as the nucleotides pass through the barrel or channel in a successive manner in order to sequence the target sequence.

[0335] Pores comprising a construct of the invention are particularly suited to these methods. In order to effectively sequence the nucleic acid, it is important to ensure that a proportion of the nucleotides in the nucleic acid is identified in a successive manner. The fixed nature of the enzyme means that a proportion of the nucleotides in the target sequence affects the current flowing through the pore.

[0336] The whole or only part of the target nucleic acid sequence may be sequenced using this method. The nucleic acid sequence can be any length. For example, the nucleic acid sequence can be at least 10, at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 400 or at least 500 nucleotides in length. The nucleic acid sequence can be naturally occurring or artificial. For instance, the method may be used to verify the sequence of a manufactured oligonucleotide. The methods are typically carried out in vitro.

[0337] The methods may be carried out using any suitable membrane/pore system in which a pore comprising a construct of the invention is inserted into a membrane. The methods are typically carried out using (i) an artificial membrane comprising a pore comprising a construct of the invention, (ii) an isolated, naturally occurring membrane comprising a pore comprising a construct of the invention, or (iii) a cell expressing a pore comprising a construct of the invention. The methods are preferably carried out using an artificial membrane. The membrane may comprise other transmembrane and/or intramembrane proteins as well as other molecules in addition to the pore of the invention.

[0338] The membrane forms a barrier to the flow of ions, nucleotides and nucleic acids. The membrane is preferably a lipid bilayer. Lipid bilayers suitable for use in accordance with the invention can be made using methods known in the art. For example, lipid bilayer membranes can be formed using the method of Montal and Mueller (1972). Lipid bilayers can also be formed using the method described in International Application No. PCT/GB08/000563 and PCT/GB07/002856.

[0339] The methods of the invention may be carried out using lipid bilayers formed from any membrane lipid including, but not limited to, phospholipids, glycolipids, cholesterol and mixtures thereof. Any of the lipids described in International Application No. PCT/GB08/000563 may be used.

[0340] Methods are known in the art for inserting pores into membranes, such as lipid bilayers. Some of those methods are discussed above.

Interaction Between the Pore and Nucleotides

[0341] The nucleotide or nucleic acid may be contacted with the pore on either side of the membrane. The nucleotide or nucleic acid may be introduced to the pore on either side of the membrane. The nucleotide or nucleic acid is typically contacted with the side of the membrane on which the enzyme is attached to the pore. This allows the enzyme to handle the nucleic acid during the method.

[0342] A proportion of the nucleotides of the target nucleic acid sequence interacts with the pore and/or adaptor as it passes across the membrane through the barrel or channel of the pore. Alternatively, if the target sequence is digested by an exonuclease, the nucleotide may interact with the pore via or in conjunction with the adaptor, dissociate from the pore and remain on the same side of the membrane. The methods may involve the use of pores in which the orientation of the adaptor is fixed. In such embodiments, the nucleotide is preferably contacted with the end of the pore towards which the adaptor is oriented. Most preferably, the nucleotide is contacted with the end of the pore towards which the portion of the adaptor that interacts with the nucleotide is orientated.

[0343] The nucleotides may interact with the pore in any manner and at any site. As discussed above, the nucleotides preferably reversibly bind to the pore via or in conjunction with the adaptor. The nucleotides most preferably reversibly bind to the pore via or in conjunction with the adaptor as they pass through the pore across the membrane. The nucleotides can also reversibly bind to the barrel or channel of the pore via or in conjunction with the adaptor as they pass through the pore across the membrane.

[0344] During the interaction between a nucleotides and the pore, the nucleotide affects the current flowing through the pore in a manner specific for that nucleotide. For example, a particular nucleotide will reduce the current flowing through the pore for a particular mean time period and to a particular extent. In other words, the current flowing through the pore is distinctive for a particular nucleotide. Control experiments may be carried out to determine the effect a particular nucleotide has on the current flowing through the pore. Results from carrying out the method of the invention on a test sample can then be compared with those derived from such a control experiment in order to identify a particular nucleotide.

Apparatus

[0345] The methods may be carried out using any apparatus that is suitable for investigating a membrane/pore system in which a pore comprising a construct of the invention is inserted into a membrane. The methods may be carried out using any apparatus that is suitable for stochastic sensing. For example, the apparatus comprises a chamber comprising an aqueous solution and a barrier that separates the chamber into two sections. The barrier has an aperture in which the membrane containing the pore is formed. The nucleotide or nucleic acid may be contacted with the pore by introducing the nucleic acid into the chamber. The nucleic acid may be introduced into either of the two sections of the chamber, but is preferably introduced into the section of the chamber containing the enzyme.

[0346] The methods may be carried out using the apparatus described in International Application No. PCT/GB08/000562.

[0347] The methods involve measuring the current passing through the pore during interaction with the nucleotides. Therefore the apparatus also comprises an electrical circuit capable of applying a potential and measuring an electrical signal across the membrane and pore. The methods may be carried out using a patch clamp or a voltage clamp. The methods preferably involve the use of a voltage clamp.

Conditions

[0348] The methods of the invention involve the measuring of a current passing through the pore during interaction with nucleotides of a target nucleic acid sequence. Suitable conditions for measuring ionic currents through transmembrane pores are known in the art and disclosed in the Examples. The method is carried out with a voltage applied across the membrane and pore. The voltage used is typically from −400mV to +400mV. The voltage used is preferably in a range having a lower limit selected from −400 mV, −300mV, −200 mV, −150 mV, −100 mV, −50 mV, −20mV and 0 mV and an upper limit independently selected from +10 mV, +20 mV, +50 mV, +100 mV, +150 mV, +200 mV, +300 mV and +400 mV. The voltage used is more preferably in the range 120 mV to 170 mV. It is possible to increase discrimination between different nucleotides by a pore of the invention by varying the applied potential.

[0349] The methods are carried out in the presence of any alkali metal chloride salt. In the exemplary apparatus discussed above, the salt is present in the aqueous solution in the chamber. Potassium chloride (KCl), sodium chloride (NaCl) or caesium chloride (CsCl) is typically used. KCl is preferred. The salt concentration is typically from 0.1 to 2.5M, from 0.3 to 1.9M, from 0.5 to 1.8M, from 0.7 to 1.7M, from 0.9 to 1.6M or from 1M to 1.4M. High salt concentrations provide a high signal to noise ratio and allow for currents indicative of the presence of a nucleotide to be identified against the background of normal current fluctuations. However, lower salt concentrations may have to be used so that the enzyme is capable of functioning.

[0350] The methods are typically carried out in the presence of a buffer. In the exemplary apparatus discussed above, the buffer is present in the aqueous solution in the chamber. Any buffer may be used in the methods. One suitable buffer is Tris-HCl buffer. The methods are typically carried out at a pH of from 4.0 to 10.0, from 4.5 to 9.5, from 5.0 to 9.0, from 5.5 to 8.8, from 6.0 to 8.7 or from 7.0 to 8.8 or 7.5 to 8.5. The pH used is preferably about 7.5.

[0351] The methods are typically carried out at from 0° C. to 100° C., from 15° C. to 95° C., from 16° C. to 90° C., from 17° C. to 85 ° C., from 18° C. to 80° C., 19° C. to 70° C., or from 20° C. to 60° C. The methods may be carried out at room temperature. The methods are preferably carried out at a temperature that supports enzyme function, such as about 37° C. Good nucleotide discrimination can be achieved at low salt concentrations if the temperature is increased. However, lower temperatures, particularly those below room temperature, result in longer dwell times and can therefore be used to obtain a higher degree of accuracy.

[0352] In addition to increasing the solution temperature, there are a number of other strategies that can be employed to increase the conductance of the solution, while maintaining conditions that are suitable for enzyme activity. One such strategy is to use the lipid bilayer to divide two different concentrations of salt solution, a low salt concentration of salt on the enzyme side and a higher concentration on the opposite side. One example of this approach is to use 200 mM of KCl on the cis side of the membrane and 500 mM KCl in the trans chamber. At these conditions, the conductance through the pore is expected to be roughly equivalent to 400 mM KCl under normal conditions, and the enzyme only experiences 200 mM if placed on the cis side. Another possible benefit of using asymmetric salt conditions is the osmotic gradient induced across the pore. This net flow of water could be used to pull nucleotides into the pore for detection. A similar effect can be achieved using a neutral osmolyte, such as sucrose, glycerol or PEG. Another possibility is to use a solution with relatively low levels of KCl and rely on an additional charge carrying species that is less disruptive to enzyme activity.

Exonuclease-Based Methods

[0353] In one embodiment, the method of sequencing a target nucleic acid sequence involves contacting the target sequence with a pore having an exonuclease enzyme, such as deoxyribonuclease, attached thereto. The constructs needed to make such pores are discussed above. Any of the exonuclease enzymes discussed above may be used in the method. The exonuclease releases individual nucleotides from one end of the target sequence. Exonucleases are enzymes that typically latch onto one end of a nucleic acid sequence and digest the sequence one nucleotide at a time from that end. The exonuclease can digest the nucleic acid in the 5' to 3' direction or 3' to 5' direction. The end of the nucleic acid to which the exonuclease binds is typically determined through the choice of enzyme used and/or using methods known in the art. Hydroxyl groups or cap structures at either end of the nucleic acid sequence may typically be used to prevent or facilitate the binding of the exonuclease to a particular end of the nucleic acid sequence.

[0354] The method involves contacting the nucleic acid sequence with the exonuclease so that the nucleotides are digested from the end of the nucleic acid at a rate that allows identification of a proportion of nucleotides as discussed above. Methods for doing this are well known in the art. For example, Edman degradation is used to successively digest single amino acids from the end of polypeptide such that they may be identified using High Performance Liquid Chromatography (HPLC). A homologous method may be used in the invention.

[0355] The rate at which the exonuclease functions can be altered by mutation compared to the wild type enzyme. A suitable rate of activity of the exonuclease in the method of sequencing involves digestion of from 0.5 to 1000 nucleotides per second, from 0.6 to 500 nucleotides per second, 0.7 to 200 nucleotides per second, from 0.8 to 100 nucleotides per second, from 0.9 to 50 nucleotides per second or 1 to 20 or 10 nucleotides per second. The rate is preferably 1, 10, 100, 500 or 1000 nucleotides per second. A suitable rate of exonuclease activity can be achieved in various ways. For example, variant exonucleases with a reduced or improved optimal rate of activity may be used in accordance with the invention.

Pushing or Pulling DNA through the Pore

[0356] Strand sequencing involves the controlled and stepwise translocation of nucleic acid polymers through a pore. The majority of DNA handling enzymes are suitable for use in this application provided they hydrolyse, polymerise or process single stranded DNA or RNA. Preferred enzymes are polymerases, nucleases, helicases and topoisomerases, such as gyrases. The enzyme moiety is not required to be in as close a proximity to the pore lumen as for individual nucleotide sequencing as there is no potential for disorder in the series in which nucleotides reach the sensing moiety of the pore.

[0357] The two strategies for single strand DNA sequencing are the translocation of the DNA through the nanopore, both cis to trans and trans to cis, either with or against an applied potential. The most advantageous mechanism for strand sequencing is the controlled translocation of single strand DNA through the nanopore with an applied potential. Exonucleases that act progressively or processively on double stranded DNA can be used on the cis side of the pore to feed the remaining single strand through under an applied potential or the trans side under a reverse potential. Likewise, a helicase that unwinds the double stranded DNA can also be used in a similar manner There are also possibilities for sequencing applications that require strand translocation against an applied potential, but the DNA must be first "caught" by the enzyme under a reverse or no potential. With the potential then switched back following binding the strand will pass cis to trans through the pore and be held in an extended conformation by the current flow. The single strand DNA exonucleases or single strand DNA dependent polymerases can act as molecular motors to pull the recently translocated single strand back through the pore in a controlled stepwise manner, trans to cis, against the applied potential.

Kits

[0358] The invention also provides kits for producing a modified pore for use in sequencing nucleic acids. In one embodiment, the kits comprise at least one construct of the invention in which the surface is a pore subunit and any remaining subunits need to form a pore. The kits may comprise enough constructs of the invention to form a complete pore (i.e. a homo-oligomer). The kits may comprise any of the constructs and subunits discussed above. A preferred kit comprises (i) a construct comprising a subunit comprising the sequence shown in SEQ ID NO: 2 or a variant thereof and (ii) six subunits comprising the sequence shown in SEQ ID NO: 2 or a variant thereof.

[0359] The kits of the invention may additionally comprise one or more other reagents or instruments which enable any of the embodiments mentioned above to be carried out. Such reagents or instruments include one or more of the following: suitable buffer(s) (aqueous solutions), means to obtain a sample from a subject (such as a vessel or an instrument comprising a needle), means to amplify and/or express polynucleotide sequences, a membrane as defined above or voltage or patch clamp apparatus. Reagents may be present in the kit in a dry state such that a fluid sample resuspends the reagents. The kit may also, optionally, comprise instructions to enable the kit to be used in the method of the invention or details regarding which patients the method may be used for. The kit may, optionally, comprise nucleotides.

Enzymes and Polynucleotide Sequences

[0360] The invention also provides an exonuclease enzyme comprising the sequence shown in any one of SEQ ID NOs: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 or a variant thereof. SEQ ID NOs: 8, 10 and 12 each have all but one native accessible cysteine residues removed. SEQ ID NOs: 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 each have all native accessible cysteine residues removed and one non-native accessible cysteine residue introduced. Variants include any of those discussed above with reference to the constructs of the invention.

[0361] The invention also provides polynucleotide sequences which encode an exonuclease enzyme of the invention. The polynucleotide sequence preferably comprises SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47 or 49 or a variant thereof. Variants of SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47 or 49 are sequences that are at least 50%, 60%, 70%, 80%, 90% or 95% homologous based on nucleotide identity to sequence of SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47 or 49 over the entire sequence. There may be at least 80%, for example at least 85%, 90% or 95% nucleotide identity over a stretch of 600 or more, for example 700, 750, 850 or 900 or more, contiguous nucleotides ("hard homology"). Homology may be calculated as described above. The polynucleotide sequence may comprise a sequence that differs from SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47 or 49 on the basis of the degeneracy of the genetic code.

[0362] Polynucleotide sequences may be isolated and replicated using standard methods in the art. Chromosomal DNA may be extracted from an enzyme producing organism, such as *E. coli, T. thermophilus* or bacteriophage. The gene encoding the enzyme may be amplified using PCR involving specific primers. The amplified sequences may then be incorporated into a recombinant replicable vector such as a cloning vector. The vector may be used to replicate the polynucleotide in a compatible host cell. Thus polynucleotide sequences encoding the enzyme may be made by introducing a polynucleotide encoding the enzyme into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells for cloning of polynucleotides are known in the art and described in more detail below.

[0363] The polynucleotide sequence may be cloned into suitable expression vector. In an expression vector, the polynucleotide sequence encoding a construct is typically operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell. Such expression vectors can be used to express a construct.

[0364] The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. Multiple copies of the same or different polynucleotide may be introduced into the vector.

[0365] The expression vector may then be introduced into a suitable host cell. Thus, a construct can be produced by inserting a polynucleotide sequence encoding a construct into an expression vector, introducing the vector into a compatible bacterial host cell, and growing the host cell under conditions which bring about expression of the polynucleotide sequence. The recombinantly-expressed construct may self-assemble into a pore in the host cell membrane. Alternatively, the recombinant construct produced in this manner may be isolated from the host cell and inserted into another membrane. When producing an oligomeric pore comprising a construct of the invention and at least one different subunit, the construct and different subunits may be expressed separately in different host cells as described above, removed from the host cells and assembled into a pore in a separate membrane, such as a rabbit cell membrane.

[0366] The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide sequence and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene. Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. A T7, trc, lac, ara or λ_L promoter is typically used. The host cell typically expresses the construct at a high level. Host cells transformed with a polynucleotide sequence encoding a construct will be chosen to be compatible with the expression vector used to transform the cell. The host cell is typically bacterial and preferably *E. coli*. Any cell with a λ DE3 lysogen, for example C41 (DE3), BL21 (DE3), JM109 (DE3), B834 (DE3), TUNER, Origami and Origami B, can express a vector comprising the T7 promoter.

[0367] The following Example illustrates the invention:

EXAMPLE

1 Materials and Methods

1.1 Overview

[0368] One method of sequencing nucleic acids using stochastic sensing involves the use of an exonuclease attached to the α-HL pore. This can be done by introducing a chemical linkage between the two proteins. The chemical linkage joins one cysteine on the exonuclease to one cysteine on the hemolysin, either directly (a disulphide bond), or via a linker molecule between the two cysteine residues.

[0369] In order to introduce the chemical linkage in a controlled and directed approach it was necessary to replace 5 naturally occurring cysteine residues found within Exonuclease I with other residues. An additional mutation would then be made in order to introduce the cysteine residue attachment point. Computer modeling was used to identify suitable attachment positions (see below).

1.2 Modelling Exo I CYS Knockout Mutations

[0370] Molecular modelling has been used in an attempt to predict which mutations of the 5 wild type cysteine residues of *E. coli* Exo I are compatible with protein stability. To this end alchemical free energy perturbation simulations where performed for these mutations, allowing predictions to be made about their relative stability. These predictions were used to guide experimental work.

[0371] VMD in combination with NAMD was used in conjunction with the CHARMM27 forcefield to perform alchemical free energy perturbation calculations upon wild type

[0372] CYS to XXX mutations of the *E. coli* Exo I protein structure (PDB accession code: 1 fxx [14]) were determined. The resulting lambda vs. deltaG curves were used to determine the free energy change upon mutation and if the calculations had converged.

1.3 Expressions and Purification

[0373] A codon optimised version of the Exonuclease I gene, scsB, from *Escherichia coli* (SEQ ID NOs: 5 and 6) was obtained from GenScript Corporation. Site directed mutagenesis of cysteine residues was performed by an adapted form of in vivo homologous recombination of PCR products (Jones, D. H. PCR mutagenesis and recombination in vivo. (1994) Genome Research, 3. pp141-148). Plasmids containing the gene of interest were transformed into BL21 (DE3) pLysS (stratagene). Cultures containing 500 ml Terrific Broth in 2 litre baffled shake flasks supplemented with 100 μg/ml ampicillin and 20 μg/ml chloramphenical were inoculated with colonies from agar plates and grown to OD600=2 (approx 6 hours) before induction with 0.2 mM IPTG at reduced temperature (18° C.). Cells were harvested by centrifugation at 4K followed by lysis with bugbuster (Novagen) and benzonase (Merck). Cell debris was removed by centrifugation at 10K and the supernatant filtered before loading onto a HisTrap column (GE Healthcare). Eluted protein was further purified by gel filtration.

1.4 Enzyme Assay

[0374] The Exonuclease I (Exo I) assay employs an oligodeoxynucleotide (DNA template) labelled with a fluorescein dye ("Fluor") on a dT at position 15, and a fluorescein quencher ("BHQ1") on dT at position 1. The template initially exists in a quenched state, with the fluorescein and quencher in close proximity. Upon the introduction of Exo I, the DNA template is digested in a processive manner from the 3'-end, eventually releasing dT-Fluor into solution and leading to a subsequent increase in fluorescence as the dye diffuses away from the quencher (FIG. **6**).

[0375] To conduct the assay, DNA template is transferred into a cuvette and placed in a Cary Eclipse fluorimeter to equilibrate at a desired temperature. Meanwhile, Exo I is equilibrated separately using a water bath and then added to the template-containing cuvette. The subsequent increase in fluorescence signal is monitored over time (see FIG. **7** for example data) with the initial rate of increase extracted. The end-point-fluorescence after the template has been fully digested is correlated with the initial amount of DNA template in order to convert the fluorescence-based rate to a molecular one.

2 Results

2.1 Modelling Exo I CYS Knockout Mutations

[0376] The results of modelling of mutations is summarised in the Tables below.

Mutation	deltaG (vacuo)	deltaG (water)	Converged (vacuo)	Converged (water)
C51A	+3.4	+3.2	YES	YES
C51R	−239.2	−251.2	NO	NO
C51N	−69.5	−66.8	NO	NO

-continued

Mutation	deltaG (vacuo)	deltaG (water)	Converged (vacuo)	Converged (water)
C51D	−67.7	−82.6	NO	NO
C51C	+0.9	+0.9	YES	YES
C51Q	−49.5	−50.7	NO	NO
C51E	−55.2	−68.9	NO	NO
C51G	−0.8	−0.5	YES	YES
C51H	−24.9	−18.5	YES	YES
C51I	+8.6	+11.0	YES	YES
C51L	−1.8	−11.7	YES	YES
C51K	−37.9	−12.1	YES	YES
C51M	−1.8	−0.4	YES	YES
C51F	+15.6	+9.4	YES	YES
C51S	+4.6	+11.9	YES	NO
C51T	−11.3	−10.7	YES	YES
C51W	+15.6	+11.6	YES	YES
C51Y	−9.3	−12.3	YES	YES
C51V	+3.4	+4.2	YES	YES

Mutation	deltaG (vacuo)	deltaG (water)	Converged (vacuo)	Converged (water)
C98A	+0.4	+4.2	YES	YES
C98R	−203.4	−231.0	NO	NO
C98N	−79.3	−76.1	NO	NO
C98D	−111.5	−113.7	NO	NO
C98C	+0.0	−0.3	YES	YES
C98Q	−48.2	−54.3	NO	NO
C98E	−55.7	−99.4	NO	NO
C98G	+1.3	−2.0	YES	YES
C98H	−14.9	−16.8	YES	YES
C98I	+6.5	+10.0	YES	YES
C98L	−2.7	−11.3	YES	YES
C98K	+11.7	−25.5	YES	YES
C98M	+0.7	+2.0	YES	YES
C98F	+10.1	+23.9	YES	YES
C98S	+3.0	+9.3	YES	NO
C98T	−13.3	−11.8	YES	YES
C98W	+17.7	+23.6	YES	YES
C98Y	+9.9	+2.2	YES	YES
C98V	+2.0	−0.3	YES	YES

Mutation	deltaG (vacuo)	deltaG (water)	Converged (vacuo)	Converged (water)
C144A	+1.2	−0.9	YES	YES
C144R	−218.0	−243.8	NO	NO
C144N	−64.9	−73.5	NO	NO
C144D	−70.6	−111.9	NO	NO
C144C	+0.4	+0.0	YES	YES
C144Q	−45.9	−52.4	NO	NO
C144E	−60.2	−71.4	NO	NO
C144G	−3.2	−1.5	YES	YES
C144H	−14.0	−16.0	YES	YES
C144I	+9.9	+7.3	YES	YES
C144L	−12.9	−11.4	YES	YES
C144K	−11.9	−25.6	YES	YES
C144M	−2.6	−2.5	YES	YES
C144F	+16.0	+8.5	YES	YES
C144S	+7.1	+7.6	YES	YES
C144T	−12.9	−9.2	YES	YES
C144W	+14.2	+9.2	YES	YES
C144Y	+5.8	−7.4	YES	YES
C144V	+0.4	+0.5	YES	YES

Mutation	deltaG (vacuo)	deltaG (water)	Converged (vacuo)	Converged (water)
C306A	+3.4	+1.1	YES	YES
C306R	−207.6	−245.2	NO	NO
C306N	−67.8	−77.1	NO	NO
C306D	−125.2	−106.6	NO	NO
C306C	+2.5	+0.0	YES	YES
C306Q	−49.8	−51.7	NO	NO
C306E	−66.6	−108.9	NO	NO
C306G	−5.1	−3.3	YES	YES
C306H	−12.6	−18.0	YES	NO
C306I	+11.0	+7.4	YES	YES
C306L	−7.3	−13.5	YES	YES
C306K	−25.1	−21.4	YES	YES
C306M	+0.5	−4.4	YES	YES
C306F	+10.7	+2.5	YES	YES
C306S	+2.7	+5.8	YES	YES
C306T	−15.5	−12.9	NO	YES
C306W	+15.5	+13.6	YES	YES
C306Y	−6.8	−15.7	YES	YES
C306V	+1.9	+4.0	YES	YES

Mutation	deltaG (vacuo)	deltaG (water)	Converged (vacuo)	Converged (water)
C330A	+2.5	+1.6	YES	YES
C330R	−219.2	−241.1	NO	NO
C330N	−77.1	−65.8	NO	NO
C330D	−62.2	−101.9	NO	NO
C330C	+2.1	−0.1	YES	YES
C330Q	−47.0	−49.8	NO	NO
C330E	−36.9	−66.2	NO	NO
C330G	−1.1	−0.2	YES	YES
C330H	−13.1	−19.0	YES	YES
C330I	+7.0	+2.6	YES	YES
C330L	−10.5	−10.9	YES	YES
C330K	−31.8	−26.4	YES	NO
C330M	−2.9	−4.6	YES	YES
C330F	+8.3	+12.2	YES	YES
C330S	+5.1	+4.7	YES	YES
C330T	−12.5	−14.8	YES	YES
C330W	+5.8	+14.6	YES	YES
C330Y	−1.8	−2.2	YES	YES
C330V	+5.3	+1.3	YES	YES

[0377] It is clear from the tables above that the obvious mutation (Cys to Ser) at each of these sites is not predicted to be energetically favourable (positive predicted deltaG). The mutations which are predicted to stabilise the protein vary significantly, as a consequence of their different local environments.

[0378] It should be noted that, although we have both activity and expression data, this does not directly relate to the delta G values in these tables. The delta G values can be related to thermodynamic data (such as that obtained from DSC), providing it is assumed that the mutations do not change the protein folding pathway significantly (which in many applications of alchemical FEP in the literature, has been found to be applicable). There is no direct relationship between expression and delta G of mutation or activity.

[0379] Comparison of the crystal structures of the Klenow fragment of DNA polymerase I and EcoExo I [14] suggest indicate that the residues involved in protein catalysis and metal binding are Asp15, Glu17, Asp108, Asp186 and His 181.

[0380] Residue types other than Cys can also be chemically modified, including Asp [1,2], Glu [1,2], Lys[3,4,5], Arg[6, 7,8], His[9,10], Tyr[11,12], Trp[11,12] and Met[13]. Therefore mutating these residues, in order to allow directed chemical modifications of Exo I allows us to remove competing chemical reactions.

[0381] Experimental work (see below) shows that many of these mutations result in proteins that retain their activity and in some cases appear to enhance activity, either as single mutations or as combinations of mutants. Comparison of the *E. coli* Exo I sequence with other related Exo I protein sequences, and the construction of homology models of these sequences suggests that many of these Cys residues are conserved both in sequence and in structure. This suggests that the properties of the Cys mutants of *E. coli* Exo I may be applicable to similar mutants of related Exo I enzymes.

[0382] A sequence alignment of all the Exo I enzymes in the Swissprot database shows that C51, C98, C144 and C306 are conserved.

2.2 Identification of 5 Cysteine Residues

[0383] Exonuclease I (SEQ ID N06) contains 5 naturally occurring cysteine residues (FIG. **1**). Some of the cysteine residues are conserved within other homologues of *E. coli* Exonuclease I (EcoExo I) (see below). None of the cysteine residues within Exonuclease I form internal disulphide bonds.

2.3 Removal of all 5 Cysteines by Mutation to Serine

[0384] The serine (Ser) amino acid is the most similar in structure to cysteine (Cys), with the S atom being replaced by an O. The most obvious mutation to make is therefore a Cys (C) to Ser (S). To test the activity of this mutant, all 5 cysteine residues were mutated to serine and protein expression was attempted.

[0385] Expression of ONLD0217 (Exo I C51S/C98S/C144S/C306S/C330S) resulted in reduced yields, poor expression and unstable protein. The protein was not able to be purified. This demonstrates that when all the Cys are replaced with Ser, the protein is not stable. Further investigation of the protein structure was required to engineer a cysteine free EcoExo I. This was guided in part by computer modeling.

2.4 Replacement of Individual Cysteines by Mutation to Serine

[0386] To identify cysteine positions that would tolerate substitution individual cysteine residues were replaced in turn with a serine residue.

[0387] Constructs ONLD0233 (C51S), ONLD0234 (C98S), ONLD0235 (C144S), ONLD0236 (C3065), ONLD0237 (C330S) were expressed. The mutant C98S expressed to sufficient levels, after a few attempts sufficient levels of C306S were also achieved. However expression levels of all the other position were much reduced.

[0388] Alternative mutations were therefore considered to produce sufficient expression levels so that the mutants could be tested for exonuclease activity. The cysteines were grouped into highly accessible positions (C98, C306, C330; FIG. **2**) and more buried residues (C51, C 144).

2.5 Replacement of C306 and C330 with Alternatives to Serine

[0389] Replacement of cysteine residues at positions 306 and 330 were investigated by examination of the following mutants.

[0390] Exo I C306→D, M, N, S, T

[0391] Exo I C330→D, M, N, T, Y, H, L, M, Q

[0392] Expression levels of ONLD0335 (C306T) were highest. Other mutations that were tolerated in this position were S, D and N.

[0393] Expression levels of ONLD0342 (C330T) were highest. Other mutations that were tolerated in this position were H, Q and M and to a lesser extent D, N, L and Y.

2.6 Combining C98S with C306S and C330T

[0394] In case double and triple mutations adversely affected each other the following combinations were investigated.

ONLD Number	EcoExo I Mutant	Expression
ONLD0366	C98S/C306S/C330T (SST)	High
ONLD0368	C98S/C306T/C330T (STT)	High
ONLD0367	C98S/C306T/C330H (STH)	Low
ONLD0364	C98S/C306D/C330T (SDT)	Low
ONLD0365	C98S/C306S/C330H (SSH)	Failed
ONLD0363	C98S/C306D/C330H (SDH)	Failed

[0395] The combination ONLD0366 gave highest expression followed by ONLD0368. The combinations ONLD0365 and ONLD0363 failed to express. Lower yields were obtained from ONLD0367, ONLD0364.

[0396] Activity assays (see Methods) showed that the relative activity of the Exo I mutants was higher than the wild type (FIG. **3**).

2.7 Introduction of Cysteine Attachment Points

[0397] Attachment points were designed to form a covalent bond from α-HL direct to the corresponding cysteine of Exonuclease I or via a linker. The attachment positions (V42C, A83C, S90C, V94C, M184C) predicted via modeling are shown in FIG. **4**.

[0398] Introduction of the above cysteines to the mutant Exo I C98S, C306S, C330T gave the following constructs:

ONLD Number	EcoExo I Mutant	Cysteine Position
ONLD0379	C98S/C306S/C330T/M184C	184
ONLD0372	C98S/C306S/C330T/V42C	42
ONLD0373	C98S/C306S/C330T/S90C	90
ONLD0378	C98S/C306S/C330T/A83C	83
ONLD0374	C98S/C306S/C330T/V94C	84

[0399] The activity of the cysteines was assessed by reacting with malemide-PEG which acts as a gel-shift reagent. The resultant gels showed that each protein had three reactive cysteines. This data showed that the cysteines at C144 and C51 were more accessible than previously assumed and were also reacting with the malemide-PEG. Further rounds of mutagenesis were carried out to replace C144 and C51 to ensure that the protein only react at one position.

2.8 Replacement of Residue C144 with Alternatives to Serine

[0400] Possible mutations were assessed from a total of 18 using molecular modeling (A, E, G, I, L, N, R, T, W, D, F, H, K, M, Q, S, V, Y). The six most stable mutants at residue 144 predicted from the modeling were made:

Exo I C98S, C306S, C330S, C144→A, G, L, T, H, M

[0401] From the 6 mutations tested (ONLD0403) Exo I C98S/C306S/C330T/C144M and (ONLD0404) Exo I C98S/C306S/C330T/C144T expressed well. Expression levels were reduced in the other mutants.

2.9 Replacement of Residue C51 with Alternatives to Serine

[0402] Possible mutations were assessed from a total of 17 using molecular modeling (A, D, E, F, G, H, I, K, L, M, N, R, S, T, V, W, Y). The eight most stable mutants at residue 51 predicted from the modeling were made:

Exo I C98S, C306S, C330S, C51→A, G, H, K, L, M, T, W

[0403] From the 8 mutations tested only (ONLD0393) Exo I C98S/C306S/C330T/C51A expressed.

2.10 Combination of 5 Cysteine Replacements and Addition of Attachement Point

[0404] The following combinations were assessed. These included all 5 cysteine replacements and addition of a cysteine attachment point. Attachment could also be made to a cysteine added to the N or C-termini.

ONLD Number	EcoExo I Mutant	Expression
ONLD0416	C51A/C98S/C144T/C306S/C330T/V42C	Excellent
ONLD0418	C51A/C98S/C144T/C306S/C330T/M184C	Excellent
ONLD0415	C51A/C98S/C144M/C306S/C330T/V42C	Good
ONLD0417	C51A/C98S/C144M/C306S/C330T/M184C	Good
ONLD0413	C51A/C98S/C144M/C306S/C330T/n-term C	Failed
ONLD0414	C51A/C98S/C144T/C306S/C330T/n-term C	Failed
ONLD0425	C51A/C98S/C144T/C306S/C330T/V94C	v. Poor/failed *
ONLD0422	C51A/C98S/C144T/C306S/C330T/S90C	Poor
ONLD0421	C51A/C98S/C144T/C306S/C330T/A83C	Good

* ONLD0425 did not express well - not enough to measure the concentration but enough to just confirm activity.

[0405] The two mutants with N-terminal cysteine residues (OND0413, ONLD0414) failed to express. All other combinations expressed and were active exonucleases. Combinations ONLD0415 and ONLD0417 expressed less well than ONLD0416 and ONLD0418. The activity results are shown in FIG. **5**.

[0406] The following combinations were also assessed.

Position 98 and 306

[0407]

ONLD Number	EcoExo I Mutant	Expression
ONLD0411	C51A/C98S/C144T/C306S/C330T	High
ONLD0432	C51A/C98T/C144T/C306S/C330T	High
ONLD0433	C51A/C98T/C144T/C306T/C330T	High

Position 98

[0408]

ONLD Number	EcoExo I Mutant	Expression
ONLD0411	C51A/C98S/C144T/C306S/C330T	Excellent
ONLD0432	C51A/C98T/C144T/C306S/C330T	Excellent
ONLD0453	C51A/C98G/C144T/C306S/C330T	Good
ONLD0454	C51A/C98K/C144T/C306S/C330T	Good
ONLD0455	C51A/C98L/C144T/C306S/C330T	Failed
ONLD0456	C51A/C98V/C144T/C306S/C330T	Excellent

Position 306

[0409] **[text missing or illegible when filed]**

ONLD Number	EcoExo I Mutant	Expression
ONLD0433	C51A/C98T/C144T/C306T/C330T	Good
ONLD0477	C51A/C98T/C144T/C306M/C330T	Good
ONLD0478	C51A/C98T/C144T/C306N/C330T	Good
ONLD0479	C51A/C98T/C144T/C306D/C330T	Poor
ONLD0480	C51A/C98T/C144T/C306A/C330T	Poor

[0410] 2.11 DSC Analysis of Proteins

[0411] Differential scanning calorimetry (DSC) is a technique that allows measurement of thermal transitions within a sample. Its primary use is measuring the thermal stability of a bio-molecule as a transition mid-point. The transition mid-point is used as a marker of thermal stability.

[0412] The earliest mutant analysed with all 5 cysteines replaced (Exo-ASTST) was found to be quite unstable with a high loss of activity over time. The Tm was determined to be 35° C., compared to wildtype at 49° C.

[0413] Following further rounds of mutant screening the enzyme Exo-ATTTT was identified with a Tm of 40° C. and greatly improved stability over time. Other combinations eg Exo-ATTMT and Exo-AVTAT had a similar but not improved Tm of 40° C.

[0414] The mutant AVTMT was found to have slightly higher Tm, however was found more prone to aggregation under some conditions.

[0415] Addition of the preferred non-native cysteine (A83C) added for attachment purposes does not appear to have a detrimental effect on stability.

Sample	ONLD	Tm (melting point) C.									
ASTST	ONLD0411	35.5 C.	35.8 C.								
ATTTT	ONLD0433	40.5 C.	40.3 C.	39.8 C.	39.8 C.	39.8	39.6	39.6	39.3	39.7	39.9
WT	ONLD0267	49	48.9								
ATTST	ONLD0432	39.8 C.	39.9	40	40.1						
CSCST	ONLD0366	41.7	41.8								
AVTST	ONLD0456			39.7 C.	39.9 C.						
AGTST	ONLD0453			36 C.*	36 C.*						
AKTST	ONLD0454			nd	nd						
ATTMT	ONLD0477					40.8	40.7				
ATTNT	ONLD0478					39.6	39.5				
ATTDT	ONLD0479					37.5*	37.4*				
ATTAT	ONLD0480					38.6*	38.5*				
AVTTT	ONLD0476							40.3	40.2		
AVTMT	ONLD0492							42.13	42.3		
AVTNT	ONLD0493							40.15	40.18		
AVTDT	ONLD0494							38.4	38.3		
AVTAT	ONLD0495							40.1	40.1		
NTTTT	ONLD0497									x	x
NVTMT	ONLD0510									40.0	39.9
NVTNT	ONLD0511									37.7	37.8
NVTAD	ONLD0514									nd	nd
NVTTT	ONLD0496									37.7	37.7
ADTST	ONLD0491									x	x

DSC results

Sample	ONLD	Tm Exp 1a	Tm Exp 1b
Exo-ATTTT-A83C-Strep-Strep	ONLD0540	40.68	40.94

2.12 Other Mutations of Interest: C1441 and C330I

[0416] C144 is adjacent to both the DNA binding site and the catalytic site of Exo-I. Residue C144 and C330 have not been mutated as much as other cysteine residues to find the best combination. Based upon the analysis of protein sequence and previous modelling work, models of the following 2 proteins (ATTTT and ATITI) were constructed.

[0417] Modeling work comparing ATTTT with ATITI suggest that T330 side chain causes little disruption of the protein secondary structure. T144 however is packed with the hydrophobic core which appears to be disrupted when the T144 side chain flips, causing significant disruption to the core. The ATITI model simulation does not show the same degree of disruption around T144.

2.13 Enzyme Activity

[0418] The construct examined was Exo-ATTTT-A83C. The linker used in these examples is a ssDNA linker with the 5'end free and the 3' end attached to the enzyme. The activity of Exo-ATTTT-A83C was measured with ssDNA linker attached and then the linker was removed by addition of DTT and activity remeasured. No change in activity was observed confirming the presence of the linker does not have a detrimental impact on activity (FIG. **8**). In this examples activity of Exo-ATTTT-A83C was compared of free enzyme and PNA modified enzyme. No loss of activity was observed by addition of the PNA linker (FIG. **9**). Exo ATTTT-M184C (with PEG). Controls were Exo ATTTT in the presence of PEG. No loss of activity was observed by addition of the PEG linker (FIG. **10**).

3 Conclusion

[0419] We have created a range of mutants of EcoExo I where all the native cysteines have been removed. In addition, a single cysteine has been introduced to control the attachment chemistry of the ExoEco I. Somewhat surprisingly, the activity of some of the mutants is higher than the wild type.

```
                              SEQUENCE LISTING


<160> NUMBER OF SEQ ID NOS: 75

<210> SEQ ID NO 1
<211> LENGTH: 882
<212> TYPE: DNA
<213> ORGANISM: Staphylococcus aureus

<400> SEQUENCE: 1

atggcagatt ctgatattaa tattaaaacc ggtactacag atattggaag caatactaca      60
```

-continued

```
gtaaaaacag gtgatttagt cacttatgat aaagaaaatg gcatgcacaa aaaagtattt       120

tatagtttta tcgatgataa aaatcacaat aaaaaactgc tagttattag aacaaaaggt       180

accattgctg gtcaatatag agtttatagc gaagaaggtg ctaacaaaag tggtttagcc       240

tggccttcag cctttaaggt acagttgcaa ctacctgata atgaagtagc tcaaatatct       300

gattactatc caagaaattc gattgataca aaagagtata tgagtacttt aacttatgga       360

ttcaacggta atgttactgg tgatgataca ggaaaaattg gcggccttat tggtgcaaat       420

gtttcgattg gtcatacact gaaatatgtt caacctgatt tcaaaacaat tttagagagc       480

ccaactgata aaaaagtagg ctggaaagtg atatttaaca atatggtgaa tcaaaattgg       540

ggaccatacg atcgagattc ttggaacccg gtatatggca atcaactttt catgaaaact       600

agaaatggtt ctatgaaagc agcagataac ttccttgatc ctaacaaagc aagttctcta       660

ttatcttcag ggttttcacc agacttcgct acagttatta ctatggatag aaaagcatcc       720

aaacaacaaa caaatataga tgtaatatac gaacgagttc gtgatgatta ccaattgcat       780

tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgttcttca       840

gaaagatata aaatcgattg ggaaaaagaa gaaatgacaa at                          882


<210> SEQ ID NO 2
<211> LENGTH: 293
<212> TYPE: PRT
<213> ORGANISM: Staphylococcus aureus

<400> SEQUENCE: 2

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1               5                   10                  15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
            20                  25                  30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
        35                  40                  45

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
    50                  55                  60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65                  70                  75                  80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
                85                  90                  95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
            100                 105                 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp
        115                 120                 125

Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His
    130                 135                 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
145                 150                 155                 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
                165                 170                 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly
            180                 185                 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
        195                 200                 205
```

-continued

```
Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
    210                 215                 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys
225                 230                 235                 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr
                245                 250                 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp
            260                 265                 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys
        275                 280                 285

Glu Glu Met Thr Asn
    290


<210> SEQ ID NO 3
<211> LENGTH: 882
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: a-HL L135C/N139Q

<400> SEQUENCE: 3

atggcagatt ctgatattaa tattaaaacc ggtactacag atattggaag caatactaca     60

gtaaaaacag gtgatttagt cacttatgat aaagaaaatg gcatgcacaa aaaagtattt    120

tatagtttta tcgatgataa aaatcacaat aaaaaactgc tagttattag aacaaaaggt    180

accattgctg gtcaatatag agtttatagc gaagaaggtg ctaacaaaag tggtttagcc    240

tggccttcag cctttaaggt acagttgcaa ctacctgata atgaagtagc tcaaatatct    300

gattactatc caagaaattc gattgataca aaagagtata tgagtacttt aacttatgga    360

ttcaacggta atgttactgg tgatgataca ggaaaaattg gcggctgtat tggtgcacaa    420

gtttcgattg gtcatacact gaaatatgtt caacctgatt tcaaaacaat tttagagagc    480

ccaactgata aaaaagtagg ctggaaagtg atatttaaca atatggtgaa tcaaaattgg    540

ggaccatacg atcgagattc ttggaacccg gtatatggca atcaactttt catgaaaact    600

agaaatggtt ctatgaaagc agcagataac ttccttgatc ctaacaaagc aagttctcta    660

ttatcttcag ggttttcacc agacttcgct acagttatta ctatggatag aaaagcatcc    720

aaacaacaaa caaatataga tgtaatatac gaacgagttc gtgatgatta ccaattgcat    780

tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgttcttca    840

gaaagatata aaatcgattg ggaaaaagaa gaaatgacaa at                       882


<210> SEQ ID NO 4
<211> LENGTH: 293
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: a-HL L135C/N139Q

<400> SEQUENCE: 4

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
1               5                   10                  15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
            20                  25                  30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
        35                  40                  45
```

-continued

```
Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
    50                  55                  60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
65                  70                  75                  80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
                85                  90                  95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Glu Tyr
            100                 105                 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp
        115                 120                 125

Thr Gly Lys Ile Gly Gly Cys Ile Gly Ala Gln Val Ser Ile Gly His
    130                 135                 140

Thr Leu Lys Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
145                 150                 155                 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
                165                 170                 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly
            180                 185                 190

Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
        195                 200                 205

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
    210                 215                 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys
225                 230                 235                 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr
                245                 250                 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp
            260                 265                 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys
        275                 280                 285

Glu Glu Met Thr Asn
    290


<210> SEQ ID NO 5
<211> LENGTH: 1390
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 5

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat tgcaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac ctgcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcgt gctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600
```

-continued

```
cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaatgccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcattgt ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380

gtggcgctgc                                                          1390
```

```
<210> SEQ ID NO 6
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 6

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Cys Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Cys Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Cys
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205
```

-continued

```
Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Cys Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Cys Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 7
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C98S/C306S/C330T/C51A (ONLD0393)

<400> SEQUENCE: 7

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360
```

-continued

```
tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcgt gctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 8
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C98S/C306S/C330T/C51A (ONLD0393)

<400> SEQUENCE: 8

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Cys
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
```

-continued

```
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 9
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C98S/C306S/C330T/C144M (ONLD0403)

<400> SEQUENCE: 9

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60
```

-continued

```
acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat tgcaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga tgtatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac    1440

catcatcatc accac                                                     1455
```

```
<210> SEQ ID NO 10
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: 10 EcoExo I C98S/C306S/C330T/C144M (ONLD0403)

<400> SEQUENCE: 10

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
      35                  40                      45

Phe Tyr Cys Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95
```

-continued

```
Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
    115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Met
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
            165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
    195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
            245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
    275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
            325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
    355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
            405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
    435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

-continued

```
<210> SEQ ID NO 11
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: 11 EcoExo I C98S/C306S/C330T/C144T (ONLD0404)

<400> SEQUENCE: 11

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc    120

aatgtgattg gcgaaccgga agtgttttat tgcaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 12
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C98S/C306S/C330T/C144T (ONLD0404)

<400> SEQUENCE: 12

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30
```

-continued

```
Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Cys Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
```

-continued

435 440 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
450 455 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465 470 475 480

His His His His His
485

<210> SEQ ID NO 13
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415)

<400> SEQUENCE: 13

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt 60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc 120

aattgcattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag 180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac 240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg 300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt 360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg 420

atgcgcgcga tgtatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc 480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc 540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt 600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg 660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc 720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt 780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt 840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg 900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg 960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac 1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc 1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg 1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat 1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat 1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg 1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa 1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac 1440

catcatcatc accac 1455

<210> SEQ ID NO 14
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence

-continued

```
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144M/C306S/C330T/V42C
      (ONLD0415)

<400> SEQUENCE: 14

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Cys Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Met
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
```

-continued

```
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
            405                 410                     415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 15
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144T/C306S/C330T/V42C
      (ONLD0416)

<400> SEQUENCE: 15

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aattgcattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc ccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt      780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260
```

-continued

```
tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 16
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416)

<400> SEQUENCE: 16

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Cys Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Met
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
```

-continued

```
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 17
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144M/C306S/C330T/M184C
      (ONLD0417)

<400> SEQUENCE: 17

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga tgtatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcgt gcgccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900
```

-continued

```
gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac    1440

catcatcatc accac                                                     1455


<210> SEQ ID NO 18
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144M/C306S/C330T/M184C
      (ONLD0417)

<400> SEQUENCE: 18

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Met
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Cys Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240
```

-continued

```
Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

<210> SEQ ID NO 19
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144T/C306S/C330T/M184C
(ONLD0418)

<400> SEQUENCE: 19

```
atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540
```

-continued

```
catgatgcgt gcgccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac    1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 20
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144T/C306S/C330T/M184C
      (ONLD0418)

<400> SEQUENCE: 20

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                      40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175
```

-continued

```
His Ser Asn Ala His Asp Ala Met Ala Asp Cys Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 21
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411)

<400> SEQUENCE: 21

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180
```

-continued

```
ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cagcattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455
```

```
<210> SEQ ID NO 22
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411)

<400> SEQUENCE: 22

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Ser Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110
```

-continued

```
Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485

<210> SEQ ID NO 23
<211> LENGTH: 1455
<212> TYPE: DNA
```

```
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432)

<400> SEQUENCE: 23

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc    120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 24
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432)

<400> SEQUENCE: 24

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
```

-continued

```
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460
```

-continued

```
Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

<210> SEQ ID NO 25
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433)

<400> SEQUENCE: 25

```
atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaacccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac    1440

catcatcatc accac                                                    1455
```

<210> SEQ ID NO 26
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433)

<400> SEQUENCE: 26

-continued

```
Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Thr Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400
```

-continued

```
Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 27
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451)

<400> SEQUENCE: 27

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc    120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga tgtatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440
```

```
catcatcatc accac                                                    1455


<210> SEQ ID NO 28
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451)

<400> SEQUENCE: 28

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Met
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
```

-continued

```
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 29
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452)

<400> SEQUENCE: 29

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc    120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga tgtatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaacccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140
```

-continued

```
aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455
```

```
<210> SEQ ID NO 30
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452)

<400> SEQUENCE: 30

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Met
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285
```

-continued

```
Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Thr Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

<210> SEQ ID NO 31
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453)

<400> SEQUENCE: 31

```
atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cggcattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780
```

-continued

```
atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455
```

```
<210> SEQ ID NO 32
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453)

<400> SEQUENCE: 32

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Gly Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220
```

-continued

```
Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

<210> SEQ ID NO 33
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0491)

<400> SEQUENCE: 33

```
atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cgatattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480
```

-continued

```
ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc ccgctggcc  tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455
```

```
<210> SEQ ID NO 34
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0491)

<400> SEQUENCE: 34

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Asp Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
```

-continued

```
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 35
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454)

<400> SEQUENCE: 35

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120
```

-continued

```
aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caaaattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc ccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                     1455


<210> SEQ ID NO 36
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454)

<400> SEQUENCE: 36

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Lys Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110
```

-continued

```
Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485

<210> SEQ ID NO 37
<211> LENGTH: 1455
```

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455)

<400> SEQUENCE: 37

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc    120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cctgattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                     1455


<210> SEQ ID NO 38
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455)

<400> SEQUENCE: 38

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45
```

-continued

```
Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Leu Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
```

-continued

```
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 39
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456)

<400> SEQUENCE: 39

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cgtgattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaagccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac    1440

catcatcatc accac                                                     1455


<210> SEQ ID NO 40
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456)
```

-continued

```
<400> SEQUENCE: 40

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Val Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ser Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400
```

-continued

```
Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 41
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98V/C144T/C306T/C330T
      (ONLD0476)

<400> SEQUENCE: 41

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac cgtgattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaacccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg    1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat    1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat    1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg    1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa    1380
```

-continued

```
gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 42
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476)

<400> SEQUENCE: 42

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Val Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Thr Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335
```

-continued

```
Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 43
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477)

<400> SEQUENCE: 43

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc     480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc     540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt     600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg     660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc     720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt     780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt     840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg     900

gttcacatta acaaaatgcc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg     960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac    1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc    1080
```

-continued

```
gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455


<210> SEQ ID NO 44
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477)

<400> SEQUENCE: 44

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270
```

-continued

```
Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Met Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

<210> SEQ ID NO 45
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478)

<400> SEQUENCE: 45

```
atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc    120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag    180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac    240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg    300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt    360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg    420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720
```

-continued

```
ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaaaaccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                               1455
```

<210> SEQ ID NO 46
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478)

<400> SEQUENCE: 46

```
Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
```

-continued

```
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Asn Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 47
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479)

<400> SEQUENCE: 47

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt      60

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc     120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag     180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac     240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg     300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt     360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg     420
```

-continued

```
atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc    480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc    540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt    600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg    660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc    720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt    780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt    840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg    900

gttcacatta acaaagatcc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg    960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac   1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc   1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg   1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat   1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat   1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg   1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa   1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac   1440

catcatcatc accac                                                    1455
```

<210> SEQ ID NO 48
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479)

<400> SEQUENCE: 48

```
Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1               5                   10                  15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
            20                  25                  30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
        35                  40                  45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
    50                  55                  60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65                  70                  75                  80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
                85                  90                  95

Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160
```

-continued

```
Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Asp Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485


<210> SEQ ID NO 49
<211> LENGTH: 1455
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480)

<400> SEQUENCE: 49

atgatgaacg atggcaaaca gcagagcacc ttcctgtttc atgattatga aaccttcggt     60
```

-continued

acccatccgg ccctggatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc 120

aatgtgattg gcgaaccgga agtgttttat gcgaaaccgg ccgatgatta tctgccgcag 180

ccgggtgcgg tgctgattac cggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac 240

gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac caccattctg 300

ggctataaca atgtgcgctt cgatgatgaa gttacccgta atatctttta tcgtaacttt 360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg 420

atgcgcgcga cctatgcgct gcgcccggaa ggcattaatt ggccggaaaa cgatgatggc 480

ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc 540

catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaacccgt 600

cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg 660

attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc 720

ggcaacacca gctgggtggc cccgctggcc tggcacccgg aaaatcgtaa cgccgtgatt 780

atggttgatc tggccggtga tattagcccg ctgctggaac tggatagcga taccctgcgt 840

gaacgcctgt ataccgccaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg 900

gttcacatta acaaagcgcc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg 960

gatcgtctgg gtattaatcg ccagcatacc ctggataatc tgaaaatcct gcgtgaaaac 1020

ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccccgagc 1080

gataacgtgg atgcgcagct gtataacggc ttctttagcg atgccgatcg cgcggcgatg 1140

aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat 1200

aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat 1260

tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg 1320

cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa 1380

gtggcgctgc tgaaagcgct gtggcagtat gcggaagaaa tcgtttctgg ctctggtcac 1440

catcatcatc accac 1455

<210> SEQ ID NO 50
<211> LENGTH: 485
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480)

<400> SEQUENCE: 50

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
1 5 10 15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
20 25 30

Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
35 40 45

Phe Tyr Ala Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
50 55 60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
65 70 75 80

Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
85 90 95

-continued

```
Thr Thr Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
            100                 105                 110

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
        115                 120                 125

Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Thr
    130                 135                 140

Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
145                 150                 155                 160

Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
                165                 170                 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
            180                 185                 190

Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
        195                 200                 205

Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro
    210                 215                 220

Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg
225                 230                 235                 240

Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg
                245                 250                 255

Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu
            260                 265                 270

Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr
        275                 280                 285

Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn
    290                 295                 300

Lys Ala Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala
305                 310                 315                 320

Asp Arg Leu Gly Ile Asn Arg Gln His Thr Leu Asp Asn Leu Lys Ile
                325                 330                 335

Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala
            340                 345                 350

Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr
        355                 360                 365

Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu
    370                 375                 380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp
385                 390                 395                 400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro
                405                 410                 415

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg
            420                 425                 430

Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln
        435                 440                 445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu
    450                 455                 460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His
465                 470                 475                 480

His His His His His
                485
```

-continued

```
<210> SEQ ID NO 51
<211> LENGTH: 804
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 51

atgaaatttg tctcttttaa tatcaacggc ctgcgcgcca gacctcacca gcttgaagcc      60

atcgtcgaaa agcaccaacc ggatgtgatt ggcctgcagg agacaaaagt tcatgacgat     120

atgtttccgc tcgaagaggt ggcgaagctc ggctacaacg tgttttatca cgggcagaaa     180

ggccattatg gcgtggcgct gctgaccaaa gagacgccga ttgccgtgcg tcgcggcttt     240

cccggtgacg acgaagaggc gcagcggcgg attattatgg cggaaatccc ctcactgctg     300

ggtaatgtca ccgtgatcaa cggttacttc ccgcagggtg aaagccgcga ccatccgata     360

aaattcccgg caaaagcgca gttttatcag aatctgcaaa actacctgga aaccgaactc     420

aaacgtgata atccggtact gattatgggc gatatgaata tcagccctac agatctggat     480

atcggcattg gcgaagaaaa ccgtaagcgc tggctgcgta ccggtaaatg ctctttcctg     540

ccggaagagc gcgaatggat ggacaggctg atgagctggg ggttggtcga taccttccgc     600

catgcgaatc cgcaaacagc agatcgtttc tcatggtttg attaccgctc aaaaggtttt     660

gacgataacc gtggtctgcg catcgacctg ctgctcgcca gccaaccgct ggcagaatgt     720

tgcgtagaaa ccggcatcga ctatgaaatc cgcagcatgg aaaaaccgtc cgatcacgcc     780

cccgtctggg cgaccttccg ccgc                                            804


<210> SEQ ID NO 52
<211> LENGTH: 268
<212> TYPE: PRT
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 52

Met Lys Phe Val Ser Phe Asn Ile Asn Gly Leu Arg Ala Arg Pro His
1               5                   10                  15

Gln Leu Glu Ala Ile Val Glu Lys His Gln Pro Asp Val Ile Gly Leu
            20                  25                  30

Gln Glu Thr Lys Val His Asp Asp Met Phe Pro Leu Glu Glu Val Ala
        35                  40                  45

Lys Leu Gly Tyr Asn Val Phe Tyr His Gly Gln Lys Gly His Tyr Gly
    50                  55                  60

Val Ala Leu Leu Thr Lys Glu Thr Pro Ile Ala Val Arg Arg Gly Phe
65                  70                  75                  80

Pro Gly Asp Asp Glu Glu Ala Gln Arg Arg Ile Ile Met Ala Glu Ile
                85                  90                  95

Pro Ser Leu Leu Gly Asn Val Thr Val Ile Asn Gly Tyr Phe Pro Gln
            100                 105                 110

Gly Glu Ser Arg Asp His Pro Ile Lys Phe Pro Ala Lys Ala Gln Phe
        115                 120                 125

Tyr Gln Asn Leu Gln Asn Tyr Leu Glu Thr Glu Leu Lys Arg Asp Asn
    130                 135                 140

Pro Val Leu Ile Met Gly Asp Met Asn Ile Ser Pro Thr Asp Leu Asp
145                 150                 155                 160

Ile Gly Ile Gly Glu Glu Asn Arg Lys Arg Trp Leu Arg Thr Gly Lys
                165                 170                 175
```

-continued

```
Cys Ser Phe Leu Pro Glu Glu Arg Glu Trp Met Asp Arg Leu Met Ser
            180                 185                 190

Trp Gly Leu Val Asp Thr Phe Arg His Ala Asn Pro Gln Thr Ala Asp
        195                 200                 205

Arg Phe Ser Trp Phe Asp Tyr Arg Ser Lys Gly Phe Asp Asp Asn Arg
    210                 215                 220

Gly Leu Arg Ile Asp Leu Leu Leu Ala Ser Gln Pro Leu Ala Glu Cys
225                 230                 235                 240

Cys Val Glu Thr Gly Ile Asp Tyr Glu Ile Arg Ser Met Glu Lys Pro
                245                 250                 255

Ser Asp His Ala Pro Val Trp Ala Thr Phe Arg Arg
            260                 265


<210> SEQ ID NO 53
<211> LENGTH: 1275
<212> TYPE: DNA
<213> ORGANISM: Thermus thermophilus

<400> SEQUENCE: 53

atgtttcgtc gtaaagaaga tctggatccg ccgctggcac tgctgccgct gaaaggcctg     60

cgcgaagccg ccgcactgct ggaagaagcg ctgcgtcaag gtaaacgcat tcgtgttcac    120

ggcgactatg atgcggatgg cctgaccggc accgcgatcc tggttcgtgg tctggccgcc    180

ctgggtgcgg atgttcatcc gtttatcccg caccgcctgg aagaaggcta tggtgtcctg    240

atggaacgcg tcccggaaca tctggaagcc tcggacctgt ttctgaccgt tgactgcggc    300

attaccaacc atgcggaact gcgcgaactg ctggaaaatg gcgtggaagt cattgttacc    360

gatcatcata cgccgggcaa aacgccgccg ccgggtctgg tcgtgcatcc ggcgctgacg    420

ccggatctga aagaaaaacc gaccggcgca ggcgtggcgt ttctgctgct gtgggcactg    480

catgaacgcc tgggcctgcc gccgccgctg gaatacgcgg acctggcagc cgttggcacc    540

attgccgacg ttgccccgct gtggggttgg aatcgtgcac tggtgaaaga aggtctggca    600

cgcatcccgg cttcatcttg ggtgggcctg cgtctgctgg ctgaagccgt gggctatacc    660

ggcaaagcgg tcgaagtcgc tttccgcatc gcgccgcgca tcaatgcggc ttcccgcctg    720

ggcgaagcgg aaaaagccct gcgcctgctg ctgacggatg atgcggcaga agctcaggcg    780

ctggtcggcg aactgcaccg tctgaacgcc cgtcgtcaga ccctggaaga agcgatgctg    840

cgcaaactgc tgccgcaggc cgacccggaa gcgaaagcca tcgttctgct ggacccggaa    900

ggccatccgg gtgttatggg tattgtggcc tctcgcatcc tggaagcgac cctgcgcccg    960

gtctttctgg tggcccaggg caaaggcacc gtgcgttcgc tggctccgat ttccgccgtc   1020

gaagcactgc gcagcgcgga agatctgctg ctgcgttatg gtggtcataa agaagcggcg   1080

ggtttcgcaa tggatgaagc gctgtttccg gcgttcaaag cacgcgttga agcgtatgcc   1140

gcacgtttcc cggatccggt tcgtgaagtg gcactgctgg atctgctgcc ggaaccgggc   1200

ctgctgccgc aggtgttccg tgaactggca ctgctggaac cgtatggtga aggtaacccg   1260

gaaccgctgt tcctg                                                    1275


<210> SEQ ID NO 54
<211> LENGTH: 425
<212> TYPE: PRT
<213> ORGANISM: Thermus thermophilus
```

-continued

```
<400> SEQUENCE: 54

Met Phe Arg Arg Lys Glu Asp Leu Asp Pro Pro Leu Ala Leu Leu Pro
1               5                   10                  15

Leu Lys Gly Leu Arg Glu Ala Ala Ala Leu Leu Glu Glu Ala Leu Arg
            20                  25                  30

Gln Gly Lys Arg Ile Arg Val His Gly Asp Tyr Asp Ala Asp Gly Leu
        35                  40                  45

Thr Gly Thr Ala Ile Leu Val Arg Gly Leu Ala Ala Leu Gly Ala Asp
    50                  55                  60

Val His Pro Phe Ile Pro His Arg Leu Glu Glu Gly Tyr Gly Val Leu
65                  70                  75                  80

Met Glu Arg Val Pro Glu His Leu Glu Ala Ser Asp Leu Phe Leu Thr
                85                  90                  95

Val Asp Cys Gly Ile Thr Asn His Ala Glu Leu Arg Glu Leu Leu Glu
            100                 105                 110

Asn Gly Val Glu Val Ile Val Thr Asp His His Thr Pro Gly Lys Thr
        115                 120                 125

Pro Pro Pro Gly Leu Val Val His Pro Ala Leu Thr Pro Asp Leu Lys
    130                 135                 140

Glu Lys Pro Thr Gly Ala Gly Val Ala Phe Leu Leu Leu Trp Ala Leu
145                 150                 155                 160

His Glu Arg Leu Gly Leu Pro Pro Pro Leu Glu Tyr Ala Asp Leu Ala
                165                 170                 175

Ala Val Gly Thr Ile Ala Asp Val Ala Pro Leu Trp Gly Trp Asn Arg
            180                 185                 190

Ala Leu Val Lys Glu Gly Leu Ala Arg Ile Pro Ala Ser Ser Trp Val
        195                 200                 205

Gly Leu Arg Leu Leu Ala Glu Ala Val Gly Tyr Thr Gly Lys Ala Val
    210                 215                 220

Glu Val Ala Phe Arg Ile Ala Pro Arg Ile Asn Ala Ala Ser Arg Leu
225                 230                 235                 240

Gly Glu Ala Glu Lys Ala Leu Arg Leu Leu Leu Thr Asp Asp Ala Ala
                245                 250                 255

Glu Ala Gln Ala Leu Val Gly Glu Leu His Arg Leu Asn Ala Arg Arg
            260                 265                 270

Gln Thr Leu Glu Glu Ala Met Leu Arg Lys Leu Leu Pro Gln Ala Asp
        275                 280                 285

Pro Glu Ala Lys Ala Ile Val Leu Leu Asp Pro Glu Gly His Pro Gly
    290                 295                 300

Val Met Gly Ile Val Ala Ser Arg Ile Leu Glu Ala Thr Leu Arg Pro
305                 310                 315                 320

Val Phe Leu Val Ala Gln Gly Lys Gly Thr Val Arg Ser Leu Ala Pro
                325                 330                 335

Ile Ser Ala Val Glu Ala Leu Arg Ser Ala Glu Asp Leu Leu Leu Arg
            340                 345                 350

Tyr Gly Gly His Lys Glu Ala Ala Gly Phe Ala Met Asp Glu Ala Leu
        355                 360                 365

Phe Pro Ala Phe Lys Ala Arg Val Glu Ala Tyr Ala Ala Arg Phe Pro
    370                 375                 380

Asp Pro Val Arg Glu Val Ala Leu Leu Asp Leu Leu Pro Glu Pro Gly
385                 390                 395                 400
```

-continued

```
Leu Leu Pro Gln Val Phe Arg Glu Leu Ala Leu Leu Glu Pro Tyr Gly
                405                 410                 415

Glu Gly Asn Pro Glu Pro Leu Phe Leu
            420                 425


<210> SEQ ID NO 55
<211> LENGTH: 738
<212> TYPE: DNA
<213> ORGANISM: Bacteriophage

<400> SEQUENCE: 55

tccggaagcg gctctggtag tggttctggc atgacaccgg acattatcct gcagcgtacc     60

gggatcgatg tgagagctgt cgaacagggg gatgatgcgt ggcacaaatt acggctcggc    120

gtcatcaccg cttcagaagt tcacaacgtg atagcaaaac cccgctccgg aaagaagtgg    180

cctgacatga aaatgtccta cttccacacc ctgcttgctg aggtttgcac cggtgtggct    240

ccggaagtta acgctaaagc actggcctgg ggaaaacagt acgagaacga cgccagaacc    300

ctgtttgaat tcacttccgg cgtgaatgtt actgaatccc cgatcatcta tcgcgacgaa    360

agtatgcgta ccgcctgctc tcccgatggt ttatgcagtg acggcaacgg ccttgaactg    420

aaatgcccgt ttacctcccg ggatttcatg aagttccggc tcggtggttt cgaggccata    480

aagtcagctt acatggccca ggtgcagtac agcatgtggg tgacgcgaaa aaatgcctgg    540

tactttgcca actatgaccc gcgtatgaag cgtgaaggcc tgcattatgt cgtgattgag    600

cgggatgaaa agtacatggc gagttttgac gagatcgtgc cggagttcat cgaaaaaatg    660

gacgaggcac tggctgaaat tggttttgta tttggggagc aatggcgatc tggctctggt    720

tccggcagcg gttccgga                                                  738


<210> SEQ ID NO 56
<211> LENGTH: 226
<212> TYPE: PRT
<213> ORGANISM: Bacteriophage

<400> SEQUENCE: 56

Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala
1               5                   10                  15

Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile
            20                  25                  30

Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys
        35                  40                  45

Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu
    50                  55                  60

Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu Ala Trp
65                  70                  75                  80

Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe Thr Ser
                85                  90                  95

Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu Ser Met
            100                 105                 110

Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn Gly Leu
        115                 120                 125

Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe Arg Leu
    130                 135                 140
```

-continued

```
Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val Gln Tyr
145                 150                 155                 160

Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn Tyr Asp
                165                 170                 175

Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu Arg Asp
            180                 185                 190

Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe Ile Glu
        195                 200                 205

Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly Glu Gln
    210                 215                 220

Trp Arg
225
```

<210> SEQ ID NO 57
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic sequence from which preferred nucleic
      acid linkers can be generated

<400> SEQUENCE: 57

```
tgtgttctat gtcttattct tacttcgtta ttcttgtctc tattctgttt atgtttcttg     60

tttgtta                                                               67
```

<210> SEQ ID NO 58
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 58

```
tgtgttctat gtctttt                                                    17
```

<210> SEQ ID NO 59
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 59

```
tgtgttctat gtcttattct tactttt                                         27
```

<210> SEQ ID NO 60
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 60

```
tgtgttctat gtcttattct tacttcgtta ttctttt                              37
```

<210> SEQ ID NO 61
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

-continued

```
<400> SEQUENCE: 61

aagacataga acacatt                                                17


<210> SEQ ID NO 62
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 62

aagtaagaat aagacataga acacatt                                     27


<210> SEQ ID NO 63
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 63

aagaataacg aagtaagaat aagacataga acacatt                          37


<210> SEQ ID NO 64
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 64

aagaataacg aagtaagaat aagacataga acacattttt                       40


<210> SEQ ID NO 65
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker

<400> SEQUENCE: 65

tgtgttctat gtcttattct tacttcgtta ttctt                            35


<210> SEQ ID NO 66
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: U base with azide group

<400> SEQUENCE: 66

cctagtctcc gnagc                                                  15


<210> SEQ ID NO 67
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with azide group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: U base with azide group

<400> SEQUENCE: 67

ccnagtctcc gnagc                                                      15


<210> SEQ ID NO 68
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with alkyne group

<400> SEQUENCE: 68

gcnacggaga ctagg                                                      15


<210> SEQ ID NO 69
<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with alkyne group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: U base with alkyne group

<400> SEQUENCE: 69

gcnacggaga cnagg                                                      15


<210> SEQ ID NO 70
<211> LENGTH: 6
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (2)..(2)
<223> OTHER INFORMATION: U base with azide group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(5)
<223> OTHER INFORMATION: U base with azide group

<400> SEQUENCE: 70

cnaang                                                                 6


<210> SEQ ID NO 71
<211> LENGTH: 8
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
```

-continued

```
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (2)..(2)
<223> OTHER INFORMATION: U base with azide group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(5)
<223> OTHER INFORMATION: U base with azide group

<400> SEQUENCE: 71

cnacnagc                                                                  8


<210> SEQ ID NO 72
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with azide group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: U base with azide group

<400> SEQUENCE: 72

ccnagcnagc                                                               10


<210> SEQ ID NO 73
<211> LENGTH: 6
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with alkyne group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (4)..(4)
<223> OTHER INFORMATION: U base with alkyne group

<400> SEQUENCE: 73

cannag                                                                    6


<210> SEQ ID NO 74
<211> LENGTH: 8
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with alkyne group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6)..(6)
<223> OTHER INFORMATION: U base with alkyne group

<400> SEQUENCE: 74

gcnagnag                                                                  8


<210> SEQ ID NO 75
<211> LENGTH: 10
```

-continued

```
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: nucleic acid linker
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: U base with alkyne group
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: U base with alkyne group

<400> SEQUENCE: 75

gcnagcnagg                                                    10
```

1-39. (canceled)

40. A construct comprising a nucleic acid binding protein and a surface, wherein at least one native accessible cysteine residue is removed from the binding protein, wherein the binding protein is attached to the surface via one or more accessible cysteine residues and wherein the binding protein retains its ability to bind nucleic acids.

41. A construct according to claim **40**, wherein:
- (a) all the native accessible cysteine residues are removed from the binding protein and one or more non-native accessible cysteine residues are introduced into the binding protein;
- (b) all the native accessible cysteine residues are removed from the binding protein and one or more non-native accessible cysteine residue(s) are introduced into the binding protein by substitution;
- (c) all but one or more of the native accessible cysteine residues are removed from the binding protein;
- (d) all or all but one of the native accessible cysteine residues are removed from the binding protein by substitution; or
- (e) all or all but one of the native accessible native cysteine residues are removed from the binding protein and substituted with threonine.

42. A construct according to claim **40**, wherein the accessible cysteine(s) are available for reaction with a thiol specific group.

43. A construct according to claim **40**, wherein the nucleic acid binding protein is a nucleic acid handling enzyme and retains its ability to handle nucleic acids.

44. A construct according to claim **43**, wherein:
- (a) the nucleic acid handling enzyme is derived from a nuclease;
- (b) the nucleic acid handling enzyme is derived from a nuclease and the nuclease is member of any of the Enzyme Classification (EC) groups 3.1.11, 3.1.13, 3.1.14, 3.1.15, 3.1.16, 3.1.21, 3.1.22, 3.1.25, 3.1.26, 3.1.27, 3.1.30 and 3.1.31;
- (c) the nucleic acid handling enzyme is derived from an exonuclease; or
- (d) the nucleic handling enzyme is derived from an exonuclease and the exonuclease comprises the sequence shown in any one of SEQ ID NOs: 6, 52, 54 and 56 or a variant thereof.

45. A construct according to claim **40**, wherein the nucleic acid binding protein comprises the sequence shown in any one of SEQ ID NOs: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 or a variant thereof.

46. A construct according to claim **43**, wherein:
- (a) the nucleic acid handling enzyme is derived from a polymerase, helicase or topoisomerase;
- (b) the nucleic acid handling enzyme is derived from a polymerase helicase or topoisomerase and:
- (i) the polymerase is member of any of the Enzyme Classification (EC) groups 2.7.7.6, 2.7.7.7, 2.7.7.19, 2.7.7.48 and 2.7.7.49;
- (ii) the helicase is member of any of the Enzyme Classification (EC) groups 3.6.1.- and 2.7.7.-; or
- (iii) the topoisomerase is member of any of the Enzyme Classification (EC) groups 5.99.1.2 and 5.99.1.3;
- (c) the nucleic acid handling enzyme is derived from a polymerase and the polymerase is a DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase, a DNA-dependent RNA polymerase or an RNA-dependent RNA polymerase;
- (d) the nucleic acid handling enzyme is derived from a helicase and the helicase an ATP-dependent DNA helicase, an ATP-dependent RNA helicase or an ATP-independent RNA helicase; or
- (e) the nucleic acid handling enzyme is derived from a topoisomerase and the topoisomerase is a gyrase.

47. A construct according to claim **40**, wherein the surface is a transmembrane protein pore or a subunit thereof.

48. A construct according to claim **47**, wherein the transmembrane protein pore is α-hemolysin (α-HL) or MspA.

49. A construct according to claim **48**, wherein the surface comprises the sequence shown in SEQ ID NO: 2 or a variant thereof.

50. A construct according claim **40**, wherein the enzyme is attached to the surface by one or more linkers, or the enzyme is attached to the surface by one or more linkers which are nucleic acid hybridization linkers.

51. A modified pore for use in sequencing nucleic acids, comprising at least one construct according to claim **49**.

52. A pore according to claim **51**, wherein:
- (a) the pore comprises the construct and six subunits comprising the sequence shown SEQ ID NO: 2 or a variant thereof; or
- (b) the pore comprises the construct and six subunits comprising the sequence shown in SEQ ID NO: 2 and all seven subunits have a glutamine at position 139 of SEQ

ID NO: 2 and one of the subunits has a cysteine at position 135 of SEQ ID NO: 2; or

(c) the pore comprises the construct and six subunits comprising the sequence shown in SEQ ID NO: 2 and all seven subunits have a glutamine at position 139 of SEQ ID NO: 2 and one of the subunits has a cysteine at position 135 of SEQ ID NO: 2 andall seven subunits have an arginine at position 113 of SEQ ID NO: 2.

53. A pore according to claim **51**, wherein:

(a) the pore comprises a molecular adaptor that facilitates an interaction between the pore and one or more nucleotide(s);

(b) the pore comprises a molecular adaptor that facilitates an interaction between the pore and one or more nucleotide and the molecular adaptor is a cyclodextrin or a derivative thereof; or

(c) the pore comprises a molecular adaptor that is cyclodextrin and the cyclodextrin is heptakis-6-amino-β-cyclodextrin (am7-βCD), 6-monodeoxy-6-monoamino-β-cyclodextrin (aml-bCD) or heptakis-(6-deoxy-6-guanidino)-cyclodextrin (gu7-βCD).

54. A kit for producing a modified pore for use in sequencing nucleic acids, comprising:

(a) (i) at least one construct according to of claim **47**; and
(ii) the remaining subunits needed to form a pore; or

(b) (i) a construct wherein the pore is α-hemolysin; and
(ii) six subunits each comprising the sequence shown in SEQ ID NO: 2 or a variant thereof.

55. A method of producing a construct according to claim **40**, comprising:

(a) attaching a nucleic acid binding protein comprising one or more accessible cysteine residues to a surface via the cysteine residues; and

(b) determining whether or not the resulting construct is capable of binding nucleic acids.

56. A method according to claim **55**, further comprising before step (a):

(i) removing all the native accessible cysteine residues from the binding protein and introducing one or more non-native accessible cysteine residues into the binding protein; or

(ii) removing all but one or more of the native accessible cysteine residues from the binding protein.

57. A method of producing a modified pore according to claim **51**, comprising:

(a) allowing at least one construct wherein the surface is a transmembrane protein pore or a subunit thereof to form a pore with other suitable subunits; and

(b) determining whether or not the resulting pore is capable of binding nucleic acids and detecting nucleotides.

58. A method of sequencing a target nucleic acid sequence, comprising:

(a) contacting the target sequence with a pore according to claim **51**, which comprises an exonuclease, such that the exonuclease digests an individual nucleotide from one end of the target sequence;

(b) contacting the nucleotide with the pore so that the nucleotide interacts with the adaptor;

(c) measuring the current passing through the pore during the interaction and thereby determining the identity of the nucleotide; and

(d) repeating steps (a) to (c) at the same end of the target sequence and thereby determining the sequence of the target sequence.

59. A method of sequencing a target nucleic acid sequence, comprising:

(a) contacting the target sequence with a pore according to claim **51** so that the enzyme provides controlled and stepwise translocation of the target sequence through the pore and a proportion of the nucleotides in the target sequence interacts with the pore; and

(b) measuring the current passing through the pore during each interaction and thereby determining the sequence of the target sequence.

60. An exonuclease enzyme comprising the sequence shown in any one of SEQ ID NOs: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48 and 50 or a variant thereof.

61. A polynucleotide sequence which encodes an exonuclease enzyme according to claim **60**.

* * * * *