

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 August 2010 (05.08.2010)

PCT

(10) International Publication Number
WO 2010/086602 A1

- (51) **International Patent Classification:**
C07K 19/00 (2006.01) C12N 15/11 (2006.01)
- (21) **International Application Number:**
PCT/GB2010/000132
- (22) **International Filing Date:**
29 January 2010 (29.01.2010)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/148,726 30 January 2009 (30.01.2009) US
- (71) **Applicant (for all designated States except US):** OXFORD NANOPORE TECHNOLOGIES LIMITED [GB/GB]; Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** JAYASINGHE, Lakmal [LK/GB]; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). MILTON, John; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). MCNEILL, Luke; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). CLARKE, James; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). WHITE, James; Oxford Nanopore Technologies Limited, Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB). MOYSEY, Ruth; Oxford Nanopore Technologies Limited,

Edmund Cartwright House, 4 Robert Robinson Avenue, Oxford Science Park, Oxford OX4 4GA (GB).

(74) **Agents:** CHAPMAN, Lee, Phillip et al.; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

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

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** HYBRIDIZATION LINKERS

(57) **Abstract:** The invention provides method of covalently coupling two or more moieties, the method comprising: (a) providing a first moiety having covalently attached thereto (i) at least one first linker comprising a first hybridizable region and (ii) at least one first group capable of forming a covalent bond; (b) providing a second moiety having covalently attached thereto (i) at least one second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and (ii) at least a second group capable of forming a covalent bond with the first group; (c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and link the moieties; and (d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups.



WO 2010/086602 A1

HYBRIDIZATION LINKERS

Field of the invention

The invention relates to a method of covalently coupling two or more moieties. The invention involves the use of two or more linkers each comprising a hybridizable region and the use of two or more groups capable of forming a covalent bond. The hybridizable regions in the linkers hybridize and link the moieties. The linked moieties are then coupled via the formation of covalent bonds between the groups.

Background of the invention

The covalent coupling of two or more moieties has many applications within chemistry and biochemistry. When chemical coupling methods are employed, the reaction rate between moieties is dependent upon the concentrations of the moieties available to react. If the concentration of moieties is low, then the reaction rate may be rate limited by the moiety concentration. For many applications the concentration of moieties may be low, it is therefore desirable to produce methods that enhance the rate or reaction (or alternatively, methods that lower the concentration of moieties required).

It is well known that certain molecules, or collections of molecules, can be made to self assemble to form larger, more complex structures, examples of these include: DNA, proteins, lipids, liquid crystals, polymers etc. The molecular components of these species often have a very high affinity for each other and self assemble in low concentration. A drawback to self-assembling systems is that they often aggregate through non-covalent interactions and therefore have limited stability.

Summary of the invention

The inventors have surprisingly demonstrated that two or more moieties can be covalently coupled at low concentrations utilising the high affinity of self-assembling systems. More specifically, the inventors have surprisingly demonstrated that two or more high affinity hybridization linkers can be used to link two or more moieties before they are coupled using chemical groups that are capable of covalently reacting together. The resulting complex has a greater stability than the equivalent non-covalent complex formed using just the high affinity hybridization linkers.

The invention therefore provides a method of covalently coupling two or more moieties, the method comprising:

- (a) providing a first moiety having covalently attached thereto (i) at least one first linker comprising a first hybridizable region and (ii) at least one first group capable of forming a covalent bond;
- 5 (b) providing a second moiety having covalently attached thereto (i) at least one second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and (ii) at least one second group capable of forming a covalent bond with the first group;
- (c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and link the moieties; and
- 10 (d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups.

The invention also provides:

- first and second moieties coupled using a method of the invention;
- a first moiety coupled to a second moiety using first and second linkers of the invention;
- 15 - a pair of first and second linkers of the invention; and
- use of a pair of linkers of the invention to couple a first moiety to a second moiety.

20 Description of the Figures

Figure 1 shows a cartoon of the structure of EcoExo I showing the position and residue number of naturally occurring cysteines (obtained from crystal structure).

Figure 2 shows a surface diagram of EcoExo I showing the position of accessible cysteines (obtained from crystal structure).

25 Figure 3 shows the relative activity of Exo I mutants compared to the wild type.

Figure 4 shows a diagram of EcoExo I showing the positions at which a single cysteine could be introduced for coupling to a protein nanopore

Figures 5a and 5b show the relative activity of Exo I mutants compared to the wild type.

30 Figure 6 shows a schematic representation of the Exonuclease I assay.

Figure 7 shows an example of data from the Exonuclease I assay.

Figure 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 ONLP1403 and the right-hand pair of columns is ONLP1405. In each pair, the left-hand (darker) column

is the enzyme with linker and the right-hand (lighter column) is enzyme without linker.

Figure 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 ONLP1498 and the right-hand column is ONLP1499.

5 Figure 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.

10 Figure 11 shows the copper (I) binding ligand used in the Examples.

Description of the Sequence Listing

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

15 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
20 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.

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

25 SEQ ID NO: 4 shows the amino acid sequence of one subunit of α -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.

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*.

30 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 96, 102 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.

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

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

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

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

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

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

SEQ ID NO: 13 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415).

SEQ ID NO: 14 shows the amino acid sequence of EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415).

SEQ ID NO: 15 shows the codon optimised polynucleotide sequence encoding EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416).

SEQ ID NO: 16 shows the amino acid sequence of EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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*.

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 $\beta_1 - \alpha_1$, $\beta_3 - \beta_4$, $\beta_5 - \beta_6$, $\beta_{III} - \alpha_I$, $\beta_{IV} - \alpha_{II}$ and $\beta_V - \beta_{VI}$ (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.

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).

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 188, 203 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.

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

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 3' – 5' direction. 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.

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

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

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

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

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.

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.

5 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.

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

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

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

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

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

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

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

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

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

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

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

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

SEQ ID NO: 76 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 77.

SEQ ID NO: 77 shows a preferred nucleic acid linker. This linker is used in combination with SEQ ID NO: 76.

SEQ ID NO: 78 shows a preferred nucleic acid linker. This linker is complementary to and used in combination with SEQ ID NO: 67.

Detailed description of the invention

5 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.

10 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 linker” includes “linkers”, reference to “a transmembrane protein pore” includes two or more such pores, reference to “a moiety” includes two or more such moieties, and the like.

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

Methods

20 The invention provides a method for coupling two or more moieties. The invention involves the use of (1) two or more linkers each comprising a hybridizable region and (2) two or more groups capable of forming a covalent bond. The hybridizable regions in the linkers hybridize and link the moieties. The linked moieties are then coupled via the formation of covalent bonds between the groups.

25 The method of the invention covalently couples two or more moieties. The method comprises providing a first moiety having covalently attached thereto (i) at least one first linker comprising a first hybridizable region and (ii) at least one first group capable of forming a covalent bond. The method also comprises providing a second moiety having covalently attached thereto (i) at least one second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and (ii) at least a second group capable of forming a covalent bond with the first group. The first and second moieties are contacted
30 under conditions that allow the first and second hybridizable regions to hybridize and link the moieties. The high affinity of the hybridization reaction allows the moieties to be linked at low concentrations. This is discussed in more detail below. The linked moieties are then exposed to conditions that allow the formation of a covalent bond between the first and second

groups. The formation of one or more covalent bonds strengthens the link between the two moieties and results in a more stable complex than would be achieved using just hybridization.

It is preferred that none of the groups used in the invention react with themselves. This improves the yield of the method when two or more different moieties are being coupled
5 because it prevents a moiety from coupling with another version of itself rather than to the different moiety to which it should be coupled.

In one embodiment, the at least one first group is covalently attached directly to the first moiety or is attached to the first moiety via a separate linker (i.e. a different linker from the first linker) and/or the at least one second group is covalently attached directly to the
10 second moiety or is attached to the second moiety via a separate linker (i.e. a different linker from the second linker). The separate linker may be any type of linker. Suitable linkers are known in the art and may be any of the linkers discussed in more detail below.

In another, more preferred, embodiment, the at least one first group is covalently attached to the first moiety via the first linker and/or the at least one second group is
15 covalently attached to the second moiety via the second linker. In this embodiment, the first moiety preferably has covalently attached thereto at least one first linker comprising both a first hybridizable region and a first group capable of forming a covalent bond and/or the second moiety preferably has covalently attached thereto at least one second linker comprising
20 a second hybridizable region capable of hybridizing to the first hybridizable region and a second group capable of forming a covalent bond with the first group. In the most preferred embodiment, the first and second groups are covalently attached to the first and second moieties via the first and second linkers respectively. Hence, in a preferred embodiment, the method comprises (a) providing a first moiety having covalently attached thereto at least one
25 first linker comprising a first hybridizable region and a first group capable of forming a covalent bond; (b) providing a second moiety having covalently attached thereto at least one second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and a second group capable of forming a covalent bond with the first
30 group; (c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and link the moieties; and (d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups.

Each moiety has covalently attached thereto at least one linker. Methods for covalently attaching linkers to moieties are well-known in the art. For instance, the linker may be attached a cysteine residue is a protein moiety via a disulphide bond. This is discussed in

more detail below. Each moiety may have more than one linker, such as 2, 3, 4 or more linkers, covalently attached thereto. In such embodiments, the two or more moieties being coupled typically have the same number of linkers covalently attached thereto. However, the two or more moieties being coupled may each have different numbers of linkers covalently attached thereto. This is also discussed in more detail below.

If more than one linker is used per moiety, it is preferred that linkers on each moiety form pairs that attach together. In other words, the hybridizable regions of two linkers on respective moieties are preferably capable of specifically hybridizing to each other. The most straightforward way of achieving this is to design the hybridizable regions of two linkers such that they only hybridize to each other and not to any of the other hybridization regions being used in the invention. If the first and second groups are attached to the first and second linkers, pairs of linkers may also be formed using click chemistry as described below.

In a preferred embodiment, the first moiety has 2, 3 or 4 first linkers covalently attached thereto and the second moiety has the corresponding number of second linkers attached thereto and the linkers form one or more pairs of first and second linkers in which the hybridizable regions in each pair specifically hybridize to each other but do not hybridize to any of the hybridizable regions in the other the pairs. In this embodiment, the moieties will be form a stable complex because they will be coupled by 2, 3 or 4 covalent bonds. This embodiment can also be used to fix the two moieties in a desired orientation.

In a more preferred embodiment, the moieties are coupled by two linkers and the method comprises (a) providing a first moiety having covalently attached thereto (i) a first linker comprising a first hybridizable region, (ii) a first group capable of forming a covalent bond, (iii) a third linker comprising a third hybridizable region and (iv) a third group capable of forming a covalent bond; (b) providing a second moiety having covalently attached thereto (i) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region, (ii) a second group capable of forming a covalent bond with the first group, (iii) a fourth linker comprising a fourth hybridizable region capable of hybridizing to the third hybridizable region and (iv) a fourth group capable of forming a covalent bond with the third group; (c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and allow the third and fourth hybridizable regions to hybridize and link the moieties; and (d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups and the third and fourth groups. In a more preferred embodiment, the method comprises (a) providing a first moiety having covalently attached thereto (1) a first linker comprising a first hybridizable

region and a first group capable of forming a covalent bond and (2) a third linker comprising a third hybridizable region and a third group capable of forming a covalent bond; (b) providing a second moiety having covalently attached thereto (1) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and a second group capable of forming a covalent bond with the first group and (2) a fourth linker comprising a fourth hybridizable region capable of hybridizing to the third hybridizable region and a fourth group capable of forming a covalent bond with the third group; (c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and allow the third and fourth hybridizable regions to hybridize and link the moieties; and (d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups and the third and fourth groups.

It is preferred that the first hybridizable region does not hybridize to the third and fourth hybridizable regions, that the second hybridizable region does not hybridize to the third and fourth hybridizable regions, that the third hybridizable region does not hybridize to the first and second hybridizable region and that the fourth hybridizable region does not hybridize to the first and second hybridizable regions.

There are also advantages to coupling the two or more moieties using different numbers of linkers on each moiety. For instance, the relative orientations of the moieties can be controlled. In one embodiment, one of the moieties has 2, 3 or 4 linkers and the other moiety has only one linker. The single or individual linker will of course need to comprise 2, 3 or 4 hybridizable regions that are capable of hybridizing to the hybridizable regions of the multiple linkers on the corresponding moiety. Hybridization to the multiple linkers may cause the single or individual linker to bend or flex. In a preferred embodiment, the first moiety has covalently attached thereto (i) a first linker comprising a first hybridizable region, (ii) a first group capable of forming a covalent bond, (iii) a third linker comprising a third hybridizable region and (iv) a third group capable of forming a covalent bond and the second moiety has covalently attached thereto (i) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and a fourth hybridizable region capable of hybridizing to the third hybridizable region, (ii) a second group capable of forming a covalent bond with the first group and (iii) a fourth group capable of forming a covalent bond with the third group. In a more preferred embodiment, the first moiety has a first linker and a third linker covalently attached thereto and the second moiety has a second linker covalently attached thereto, wherein the first linker comprises a first hybridizable region and a first group capable of forming a covalent bond, wherein the third linker comprises a third hybridizable

region and a third group capable of forming a covalent bond and wherein the second linker comprises a second hybridizable region capable of hybridizing to the first hybridizable region, a fourth hybridizable region capable of hybridizing to the third hybridizable region and second and fourth groups capable of forming covalent bonds with the first and third groups.

5 Each moiety has covalently attached thereto at least one group. Methods for covalently attaching groups to the moieties or the linkers are well-known in the art. This is discussed in more detail below. Each moiety may have more than one group, such as 2, 3, 4 or more groups, covalently attached thereto. This allows the formation of a strong complex containing the moieties. The linkers and groups can also be designed to fix the moieties in a
10 desired spatial orientation relative to each other. Multiple groups may also be used to constrain the mobility of the moieties relative to each other. For instance, multiple groups may be used to constrain the freedom of one moiety to rotate or its ability to move away from the other moiety.

More than two moieties, such as 3, 4, 5, 6 or more moieties, may be coupled in
15 accordance with the invention. This allows complexes comprising several moieties to be produced. It also allows complexes comprising several moieties in a particular arrangement to be produced. For instance, the hybridizing regions or the groups of the linkers can be designed such that they couple the moieties in a particular arrangement. For instance, if three moieties are used (A, B and C), they can be coupled in each of the specific arrangements,
20 namely A-B-C, A-C-B, B-A-C, B-C-A, C-A-B and C-B-A. The terminal moieties of any of these arrangements can also be coupled together to form a "loop" of moieties.

A person skilled in the art will be able to design suitable moiety and linker combinations to allow the moieties to be linked. For instance, if three moieties are used (A, B and C) as described above, A-B-C can be formed by providing B with two linkers covalently
25 attached thereto, one of which allows coupling via a single linker on A and the other of which allows coupling via a single linker on B. Alternatively, each moiety may have two linkers covalently attached thereto. If one of the linkers on A allows coupling via one of the linkers on B, the other linker on B allows coupling via one of the linkers on C and the other linker on C allows coupling via the remaining linker on A, the three moieties may be coupled in a "loop".
30 Multiple linkers may also be used to couple each moiety as discussed above.

In yet a further embodiment, the linkers may be designed to self-assemble into a structure to which the moieties are covalently attached. Nucleic acid self-assembly is well-known in the art. The interactions between synthetic oligonucleotides, and thus the structures that they form by hybridization, can be controlled through design of their nucleotide sequences

using simple rules (Watson-Crick base pairing, supplemented by rules prescribing the design of structural motifs such as junctions, G-quadruplexes, i-motifs etc.). For instance, Seeman *et al.* have prepared various three-dimensional DNA lattices (for example NC Seeman *et al.*, *J. Theor. Biol.* **99**, 237 (1982)). The self-assembled structure is coupled by the covalent bonds
5 formed between the groups on the linkers.

In a preferred embodiment, the method comprises coupling 3, 4 or 5 moieties and wherein each moiety has covalently attached thereto (i) at least one linker which comprises a hybridizable region capable of hybridizing to the hybridizable region of a linker on a different moiety and (ii) at least one covalent group capable of forming a covalent bond with the group
10 on a different moiety. In a more preferred embodiment, the method comprises coupling 3, 4 or 5 moieties and wherein each moiety has at least one linker covalently attached thereto which comprises hybridizable regions capable of hybridizing to the hybridizable region of another linker and a covalent group capable of forming a covalent bond with the group on another linker.

In an even more preferred embodiment, the method comprises a) providing a first moiety having covalently attached thereto (i) a first linker comprising a first hybridizable region and (ii) a first group capable of forming a covalent bond; b) providing a second moiety having covalently attached thereto (i) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and (ii) a second group capable of
20 forming a covalent bond with the first group, (iii) a third linker comprising a third hybridizable region and (iv) a third group capable of forming a covalent bond; c) providing a third moiety having covalently attached thereto (i) a fourth linker comprising a fourth hybridizable region capable of hybridizing to the third hybridizable region, (ii) a fourth group capable of forming a covalent bond with the third group; d) contacting the first, second and third moieties under
25 conditions that allow the hybridizable regions to hybridize and link the moieties; and e) exposing the linked moieties to conditions that allow the formation of a covalent bond between the groups.

In an even more preferred embodiment, the method comprises (a) providing a first moiety having covalently attached thereto a first linker comprising a first hybridizable region
30 and a first group capable of forming a covalent bond; (b) providing a second moiety having covalently attached thereto (1) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and a second group capable of forming a covalent bond with the first group and (2) a third linker comprising a third hybridizable region and a third group capable of forming a covalent bond; (c) providing a third moiety having

covalently attached thereto a fourth linker comprising a fourth hybridizable region capable of hybridizing to the third hybridizable region and a fourth group capable of forming a covalent bond with the third group; (d) contacting the first, second and third moieties under conditions that allow the hybridizable regions to hybridize and link the moieties; and (e) exposing the
5 linked moieties to conditions that allow the formation of a covalent bond between the groups.

Moieties

The two or more moieties may be any substance or surface that may be covalently coupled using linkers. Preferred moieties include, but are not limited to, polypeptides,
10 proteins, nucleic acids, chemical groups, dyes, affinity tags, polymers, beads, antibodies, supports, immobilised metal affinity matrices, glutathione sepharose, dextrin sepharose, affinity matrices, IgG sepharoses, activated thiol-sepharoses, metal films, nanoparticles, glass, treated glass, plastics, resin surfaces, solid state nanopores, sensor surfaces and single molecular detectors.

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

20 The two or more moieties may be the same or different.

It is preferred that the first moiety is a protein and the second moiety is a protein, a nucleic acid, a reporter group, an affinity tag or a support. The two or more moieties may be attached in any configuration. For instance, if they are proteins, they may be attached via their terminal (i.e. amino or carboxy terminal) amino acids.

25 It is more preferred that the first moiety is a nucleic acid binding protein and the second moiety is a transmembrane protein pore. 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 transmembrane protein pore and *vice versa*. Alternatively, the two components may be attached via amino acids within their sequences. For instance, the
30 nucleic acid binding protein may be attached to one or more amino acids in a loop region of the moiety. 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 moiety. Terminal amino acids and loop regions are discussed below.

If the first moiety is a nucleic acid binding protein and the second moiety is a transmembrane protein pore, the first and second groups are preferably from 0.05 nm to 10 nm from their respective moieties, more preferably from 0.05 nm to 5 nm and most preferably about 0.1 nm, or 3 nucleotides, from their respective moieties. If the first moiety is a nucleic acid binding protein and the second moiety is a transmembrane protein pore, the distance between the coupled first and second moieties is preferably from 0.1 nm to 100 nm, more preferably from 0.1 nm to 25 nm, more preferably from 0.1 nm to 5 nm and most preferably 0.5 nm. To achieve distances of around 0.5 nm between moieties it will be necessary to have the attached linkers in a "head to head" arrangement. If the linkers are used in a "head to tail" arrangement, a minimum of 5 or 6 nucleotides in each linker is needed to provide sufficient hybridization. Shorter linkers will not hybridize adequately, if at all, and thus shorter moiety-moiety distances can only be achieved using a different conformation.

In a preferred embodiment, the site of covalent attachment is selected such that, when the moiety is 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.

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 or 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.

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 or 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.

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.

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.

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 or 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 moiety and/or the use of specific linkers.

Biologically active moieties are typically coupled in such a way that they retain their biological activity. The biological activity of any of the moieties described herein can be

determined using methods well-known in the art. If a nucleic acid binding protein and a transmembrane protein pore are coupled together, both proteins preferably retain their binding and pore properties respectively.

If one or more of the moieties is a protein, at least one native accessible cysteine residue may be removed from the protein as described in a co-pending International application claiming priority from US Application No. 61/148,726 and being filed simultaneously with this application [J A Kemp & Co Ref: N.106566A; Oxford Nanopore Technologies Ref: ONL IP 007]. However, the protein may comprise one or more accessible cysteine residues and may be attached to the other moieties via those residues. The presence of a limited number of accessible cysteine residues allows controlled coupling between moieties.

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.

The invention may involve site directed mutagenesis of native cysteine residues from protein moieties. 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 another moiety, such as another protein, a solid support or a chemical reagent.

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 coupling, while also minimising the possible interaction of the 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.

The native cysteine residues present in wild type proteins do not readily permit the use of specific linkers containing thiol reactive groups, such as maleimide, iodoacetamide or ortho-

pyridyl disulphide (OPSS). The potential for reaction with any structurally or catalytically important cysteine residues could impact upon the 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 protein activity is not affected by being in close proximity to another moiety, such as a protein or a surface.

Nucleic acid binding protein

One of the moieties is preferably 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 handling enzyme is modified to remove and/or replace cysteine residues as described above and in a co-pending International application claiming priority from US Application No. 61/148,726 and being filed simultaneously with this application [J A Kemp & Co Ref: N.106566A; Oxford Nanopore Technologies Ref: ONL IP 007]. At least one native accessible cysteine residue is preferably removed from the nucleic acid binding protein. However, the nucleic acid binding protein comprises one or more accessible cysteine residues and is attached to another moiety via those residues. The presence of a limited number of accessible cysteine residues allows controlled attachment to another moiety.

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.

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),
5 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
10 diphosphate (dCDP) and deoxycytidine triphosphate (dCTP). The nucleotides are preferably selected from AMP, TMP, GMP, CMP, UMP, dAMP, dTMP, dGMP or dCMP.

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)
15 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.

20 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 coupled structure.

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
25 formed from one monomer. Such proteins are easy to handle and are less likely to interfere with the other moiety. For instance, such proteins are less likely to interfere with the pore forming ability of a pore subunit.

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
30 the protein. It also allows the protein to be attached to the other moiety 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 coupled structure to be designed.

It may be necessary to purify the structure produced by the method of the invention. It is preferred that the nucleic acid binding protein is capable of withstanding the conditions used to purify the structure.

The other moiety 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.

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

The enzyme also preferably retains at least partial activity at temperatures from 10 °C to 60 °C, such as at room temperature. This allows nucleic acids to be sequenced at a variety of temperatures, including room temperature.

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.

The nucleic acid handling enzyme is preferably derived from a nucleolytic enzyme. The nucleic acid handling 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:

- **3. 1.11.- Exodeoxyribonucleases producing 5'-phosphomonoesters.**
 - 3.1.11.1 Exodeoxyribonuclease I.
 - 3.1.11.2 Exodeoxyribonuclease III.
 - 3.1.11.3 Exodeoxyribonuclease (lambda-induced).
 - 5 ○ 3.1.11.4 Exodeoxyribonuclease (phage SP3-induced).
 - 3.1.11.5 Exodeoxyribonuclease V.
 - 3.1.11.6 Exodeoxyribonuclease VII.

- **3. 1.13.- Exoribonucleases producing 5'-phosphomonoesters.**
 - 10 ○ 3.1.13.1 Exoribonuclease II.
 - 3.1.13.2 Exoribonuclease H.
 - 3.1.13.3 Oligonucleotidase.
 - 3.1.13.4 Poly(A)-specific ribonuclease.
 - 3.1.13.5 Ribonuclease D.

- 15
- **3. 1.14.- Exoribonucleases producing 3'-phosphomonoesters.**
 - 3.1.14.1 Yeast ribonuclease.

- **3. 1.15.- Exonucleases active with either ribo- or deoxyribonucleic acid producing 5' phosphomonoesters**
 - 20 ○ 3.1.15.1 Venom exonuclease.

- **3. 1.16.- Exonucleases active with either ribo- or deoxyribonucleic acid producing 3' phosphomonoesters**
 - 25 ○ 3.1.16.1 Spleen exonuclease.

- **3. 1.21.- Endodeoxyribonucleases producing 5'-phosphomonoesters.**
 - 3.1.21.1 Deoxyribonuclease I.
 - 3.1.21.2 Deoxyribonuclease IV (phage-T(4)-induced).
 - 30 ○ 3.1.21.3 Type I site-specific deoxyribonuclease.
 - 3.1.21.4 Type II site-specific deoxyribonuclease.
 - 3.1.21.5 Type III site-specific deoxyribonuclease.
 - 3.1.21.6 CC-preferring endodeoxyribonuclease.
 - 3.1.21.7 Deoxyribonuclease V.

- **3. 1.22.- Endodeoxyribonucleases producing other than 5'-phosphomonoesters.**
 - 3.1.22.1 Deoxyribonuclease II.
 - 3.1.22.2 Aspergillus deoxyribonuclease K(1).
 - 5 ○ 3.1.22.3 Transferred entry: 3.1.21.7.
 - 3.1.22.4 Crossover junction endodeoxyribonuclease.
 - 3.1.22.5 Deoxyribonuclease X.

- **3. 1.25.- Site-specific endodeoxyribonucleases specific for altered bases.**
 - 10 ○ 3.1.25.1 Deoxyribonuclease (pyrimidine dimer).
 - 3.1.25.2 Transferred entry: 4.2.99.18.

- **3. 1.26.- Endoribonucleases producing 5'-phosphomonoesters.**
 - 3.1.26.1 Physarum polycephalum ribonuclease.
 - 15 ○ 3.1.26.2 Ribonuclease alpha.
 - 3.1.26.3 Ribonuclease III.
 - 3.1.26.4 Ribonuclease H.
 - 3.1.26.5 Ribonuclease P.
 - 3.1.26.6 Ribonuclease IV.
 - 20 ○ 3.1.26.7 Ribonuclease P4.
 - 3.1.26.8 Ribonuclease M5.
 - 3.1.26.9 Ribonuclease (poly-(U)-specific).
 - 3.1.26.10 Ribonuclease IX.
 - 3.1.26.11 Ribonuclease Z.
 - 25

- **3. 1.27.- Endoribonucleases producing other than 5'-phosphomonoesters.**
 - 3.1.27.1 Ribonuclease T(2).
 - 3.1.27.2 Bacillus subtilis ribonuclease.
 - 3.1.27.3 Ribonuclease T(1).
 - 30 ○ 3.1.27.4 Ribonuclease U(2).
 - 3.1.27.5 Pancreatic ribonuclease.
 - 3.1.27.6 Enterobacter ribonuclease.
 - 3.1.27.7 Ribonuclease F.
 - 3.1.27.8 Ribonuclease V.

- 3.1.27.9 tRNA-intron endonuclease.
- 3.1.27.10 rRNA endonuclease.

- **3.1.30.- Endoribonucleases active with either ribo- or deoxyribonucleic producing 5' phosphomonoesters**

- 3.1.30.1 *Aspergillus* nuclease S(1).
- 3.1.30.2 *Serratia marcescens* nuclease.

- **3.1.31.- Endoribonucleases active with either ribo- or deoxyribonucleic producing 3' phosphomonoesters**

- 3.1.31.1 Micrococcal nuclease.

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.

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.

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

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.

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.

5 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.

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
10 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.
15 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").

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,
20 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.

25 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
30 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.

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.

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

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.

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.

5 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.
10 A carrier protein may be fused to a subunit or variant.

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
15 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.

20 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.

Preferred enzymes that are capable of pushing or pulling the target nucleic acid
25 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
30 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.

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.

5 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
10 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.

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-
15 amino acids. This is conventional in the art for producing such proteins or peptides.

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 other moiety. 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,
20 reductive alkylation of amino acids by reaction with an aldehyde followed by reduction with NaBH_4 , amidination with methylacetimidate or acylation with acetic anhydride. The 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 in the method of the invention.

25 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
30 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.

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.

5 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.

10 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.

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.

15 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
20 position 144 or 330 of SEQ ID NO: 6 remains.

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
25 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
30 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.

The nucleic acid handling enzyme preferably 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. 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.

The nucleic acid handling enzyme is typically attached to the moiety 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 in 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.

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 moiety via more than one residue.

The nucleic acid binding protein can be attached to another moiety at one or more positions, such as at one, two, three or four positions. The nucleic acid binding protein is preferably attached at one or two positions. After removal of native cysteine residues from the nucleic acid binding protein, one or more cysteine residues can be incorporated into the protein at specific positions for the attachments. Attachment at two positions can reduce the flexibility of the complex and can fix the nucleic acid binding protein on the moiety in a chosen specific orientation.

In a preferred embodiment, the method comprises coupling (1) exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof with the substitutions M184C and V94C to (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 coupled to the N17C position in SEQ ID NO: 2 or a variant thereof. The V94C position in SEQ ID NO: 6 or a variant thereof is coupled to the E287C position in SEQ ID NO: 2 or a variant thereof by a different linker.

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

	First A	First B (attached to First A)	Second A	Second B (attached to Second A)
	SEQ ID NO: 6 or variant thereof	SEQ ID NO: 2 or variant thereof	SEQ ID NO: 6 or variant thereof	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

Transmembrane protein pore

One of the moieties is preferably 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.

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.

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.

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.

The moiety may be a pore itself. Alternatively, if the pore is an oligomer, the moiety may be a pore subunit. The moiety may be isolated, substantially isolated, purified or substantially purified as described above.

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 moiety 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.

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).

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).

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.

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 a 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.

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.

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.

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

Variants may 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.

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.

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.

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.

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.

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.

Linkers and groups

The two or more moieties have at least one linker and at least one group as defined above covalently attached thereto. Methods for attaching linkers and groups to moieties are well-known in the art. The linkers and groups are preferably chemically fused to the moiety. Suitable methods include, but are not limited to, primary amine/carboxylic acid coupling, reaction between thiols, reaction between non-natural amino acids and cysteine linkage.

The method of the invention may further comprise the step of attaching the at least one linker and/or the at least one group to each moiety.

The first moiety is preferably attached to the second moiety using linkers at one or more positions, for example at two, three or four positions. The position of attachment of each linker to its relevant moiety may be designed to constrain the relative mobility of the moieties. The number of positions at which the moieties are linked will have an effect on the rigidity of the final complex.

The linkers and/or groups may be attached to one or more reactive groups on the moiety, such as cysteine residues, reactive lysine residues or non-natural amino acids. Preferably, the linkers and/or groups are attached to one or more accessible cysteine residues in the moiety. Typically the linkers and/or groups are attached to the moiety via a disulphide bridge from one cysteine residue to another.

The linkers are high affinity hybridization linkers. They comprise at least one hybridizable region and preferably comprise at least one group capable of forming a covalent bond. Particularly preferred linkers which are effective at low concentrations, for example at nanomolar affinity. The length, flexibility and hydrophilicity of linkers are typically designed such that they do not disturb the functions of the moieties to be linked. An advantage of using hybridization linkers is that the formation of unwanted homodimers of the moieties is minimized.

A hybridizable region is a region that is capable of undergoing hybridization with a hybridizable region in another linker. The hybridizable region preferably comprises a nucleic acid. 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 nucleic acids that allow two linkers to hybridize may be the same or different. Most preferably the linkers comprise DNA. Typically, when respective linkers comprise DNA, at least one linker must comprise single-stranded DNA. Preferably, both linkers comprise single-stranded DNA.

Hybridization may be between two or more linkers, preferably between two or three linkers. In the case where three linkers are used, they form a triplex. The linkers may hybridize in a variety of conformations, for example in a "head to head" conformation or a "head to tail" conformation. In the "head to head" conformation, the linkers align so that the moieties are adjacent at the same end of the hybridized linkers. In the "head to tail" conformation, the linkers align so the moieties are at opposite ends of the hybridized linkers. The linkers may have any suitable three-dimensional conformation. Typically, a linker is linear, but it can also be triangular or shaped in another geometrical confirmation.

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, which allows the linker to be attached to the moiety. Maleimide or Thiol modified oligonucleotides 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 moieties and the complementary linker (SEQ ID NO: 61, 62 or 63 respectively) is attached to the other moiety. The moieties can then be linked by hybridizing the linkers.

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).

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 one moiety to detect a substrate which has been released from the other moiety. 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 length, such as 6, 8 or 10 nucleotides long.

The first and second hybridizable regions preferably have an affinity of from 1 fM to 1 uM at concentrations of from 1 pM to 1 mM. The first and second hybridizable regions more preferably have an affinity of from 1 fM to 10nM at concentrations of from 1 pM to 1 uM. The first and second hybridizable regions most preferably have an affinity of from 1 pM to 100 pM at concentrations of from 100 pM to 10 nM.

A preferred sequence 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 TTTT at the 3' end to provide extra flexibility. 5' end of SEQ ID NO 65 can then be attached to the nucleic acid binding protein. In this example, 3' end of the SEQ ID NO 64 is complementary to a stretch of sequence at the 5' end of the SEQ ID NO 65.

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 moieties are modified such they react with one another once they have hybridized. Each moiety has covalently attached thereto a group capable of forming a covalent bond with a group in another moiety. As discussed above, the group is preferably attached to the moiety via the

hybridization linker. A pair of moieties can be linked by one or more covalent bonds, for example one, two or three, covalent bonds.

Typically the bond will be a simple disulfide bond between the two moieties. The moieties or hybridization 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 incorporated into the linkers either internally or terminally.

The moieties or hybridization linkers can also be modified to include one or more, such as two, iodoacetamide groups. Again, the iodoacetamide groups can be incorporated into the linkers either internally or terminally. A moiety or a hybridization linker with one or more iodoacetamide groups can be covalently linked to thiols on the complementary moiety or hybridization linker.

The moieties or hybridization linkers can also be modified with alkene groups. If incorporated into the hybridization linkers, the alkene groups can be internal or terminal. One or more alkene groups in preferred positions can be subjected to olefin metathesis to make a covalent bond between the alkenes in the moieties or hybridization linkers.

If necessary, a small linker can be added between the moiety or hybridization 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 moieties or hybridization linkers.

In a preferred embodiment, the covalent bond between the moieties or hybridization 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 HC, Finn, MG, Sharpless KB, Click chemistry: diverse chemical function from a few good reactions, *Angew. Chem. Int. Ed.* 40 (2001) 2004–2021). They have defined the set of stringent criteria for click chemistry as follows: “The reaction must be modular, wide in scope, give very high yields, generate only inoffensive byproducts that can be removed by nonchromatographic methods, and be stereospecific (but not necessarily enantioselective). The required process characteristics include simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation. Purification if required must be by nonchromatographic methods, such as crystallization or distillation, and the product must be stable under physiological conditions”.

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

- (a) copper-free variant of the 1,3 dipolar cycloaddition reaction, where an azide reacts with an alkyne under strain, for example in a cyclooctane ring;
- (b) the reaction of an oxygen nucleophile on one linker with an epoxide or aziridine reactive moiety on the other; and
- 5 (c) the Staudinger ligation, where the alkyne moiety can be replaced by an aryl phosphine, resulting in a specific reaction with the azide to give an amide bond.

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

10

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).

15

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 a moiety can then be modified with these azide hybridization linkers using the thiol group at the 5' end (SEQ ID NO 66 and 67). Alkyne groups can also be incorporated into uracil bases at preferred positions in sequences complementary to the SEQ ID NO 66 and 67 (SEQ ID NO 68 and 69 respectively). These sequences can be attached to the cysteines in one of the moieties. Using DNA hybridization followed by 'click chemistry' between the azide and alkyne groups, hybridization linkers can be covalently cross-linked. If the distance between the moieties has been modulated by changing the length of the hybridization linkers, the position of the azide and alkyne modified bases then needs to be changed accordingly.

20

25

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 attached 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. Using these hybridization linkers will bring the moieties closer to each other than

30

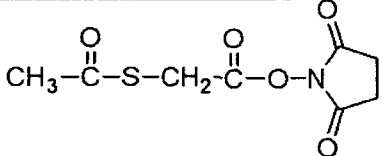
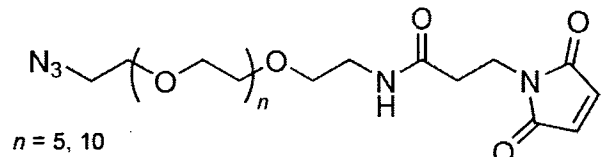
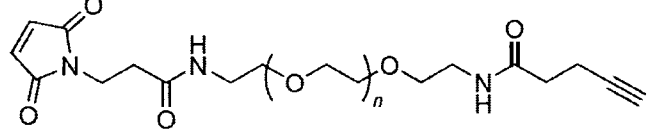
with other preferred hybridization linkers of the invention, such as SEQ ID NO 67 and 69.

Incorporation of azide and alkyne groups into uracil base units of DNA has been developed by ATDBio.

Other preferred groups for use in the invention are shown in the following Table 3.

5 Table 3 – Some preferred groups capable of forming covalent bonds

Name	Reacts with	Structure
1,4-Bis[3-(2-pyridyldithio)propionamid o]butane	Thiols	
1,11-bis-Maleimidotriethyleneglycol	Thiols	
3,3'-Dithiodipropionic acid di(N-hydroxysuccinimide ester)	Primary amines	
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-activate d, 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	
Azide-PEG-maleimide	Thiols, alkyne	
Alkyne-PEG-maleimide	Thiols, azide	

Conditions

The moieties are first contacted under conditions that allow the first and second hybridizable regions to hybridize and link the moieties. Conditions that permit the hybridization are well-known in the art (for example, Sambrook *et al.*, 2001, Molecular Cloning: a laboratory manual, 3rd edition, Cold Spring Harbour Laboratory Press; and Current Protocols in Molecular Biology, Chapter 2, Ausubel *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995)). Hybridization can be carried out under low stringency conditions, for example in the presence of a buffered solution of 30 to 35% formamide, 1 M NaCl and 1 % SDS (sodium dodecyl sulfate) at 37°C followed by a wash in from 1X (0.1650 M Na⁺) to 2X (0.33 M Na⁺) SSC (standard sodium citrate) at 50°C. Hybridization can be carried out under moderate stringency conditions, for example in the presence of a buffer solution of 40 to 45% formamide, 1 M NaCl, and 1 % SDS at 37°C, followed by a wash in from 0.5X (0.0825 M Na⁺) to 1X (0.1650 M Na⁺) SSC at 55°C. Hybridization can be carried out under high stringency conditions, for example in the presence of a buffered solution of 50% formamide, 1 M NaCl, 1% SDS at 37°C, followed by a wash in 0.1X (0.0165 M Na⁺) SSC at 60°C. Preferred conditions are those described in Example 1.

The linked moieties are then exposed to conditions that allow the formation of a covalent bond between the first and second groups. These conditions are of course dependent on the groups used. The reaction is preferably carried out at less than pH 7.0. Preferred conditions are those described in Example 1.

If one or more of the moieties are enzymes, such as nucleic acid binding proteins, the activity of the enzyme is preferably inhibited during the coupling and linking reactions. Magnesium-dependent enzymes, such as Exonuclease I, may be inhibited by EDTA or by CaCl₂ (Lehman (1960) *J. Biol. Chem.* **235** 5 1479-1487 and internal data).

Preferred combinations of moieties

In a preferred embodiment, the first moiety is a protein and the second moiety is a reporter group. Suitable reporter groups include, but are not limited to, optical or fluorescent labels, steric tags, increases to mass, changes to charge, spin labels, radiolabels, magnetic groups and any combination thereof.

In another preferred embodiment, the first moiety is a protein and the second moiety is an affinity tag. Examples of affinity tags are well-known in the art. Suitable affinity tags include, but are not limited to, His-NTA, FLAG-Antibody-M2, myc, STREP- streptavidin, NANO- streptavidin and biotin-avidin.

In another preferred embodiment, the first moiety is a protein and the second moiety is a support. Examples of supports are well-known in the art. Suitable supports include, but are not limited to, The support may be, but is not limited to metal surfaces, polymer beads, magnetic beads, glass, modified glass and silica gels.

In a preferred embodiment, the first moiety comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having at least one native accessible cysteine residue and the second moiety comprises α -HL (SEQ ID NO: 2) or a variant thereof. In a more preferred embodiment, the first moiety comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residues removed and the second moiety comprises α -HL (SEQ ID NO: 2) or a variant thereof. In a more preferred embodiment, the first moiety 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 and the second moiety comprises α -HL (SEQ ID NO: 2) or a variant thereof. In a more preferred embodiment, the first moiety comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residues removed and A83C and the second moiety comprises α -HL (SEQ ID NO: 2) or a variant thereof having E287C.

In a preferred embodiment, the first moiety comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having at least one native accessible cysteine residue removed and the second moiety comprises α -HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4). In a more preferred embodiment, the first moiety comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residues removed and the second moiety comprises α -HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4). In a more preferred embodiment, the first moiety 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 and the second moiety comprises α -HL or a variant thereof having L135C and N139Q (e.g. SEQ ID NO: 4). In a more preferred embodiment, the first moiety comprises exonuclease I from *E. coli* (SEQ ID NO: 6) or a variant thereof having all its native accessible cysteine residues removed and A83C and the second moiety comprises α -HL or a variant thereof having L135C, N139Q and E287C.

Products

The invention also provides first and second moieties coupled using a method of the invention and a first moiety coupled to a second moiety using first and second linkers as defined above. The invention further provides a pair of first and second linkers as defined above. The linkers preferably comprise the groups capable of forming covalent bonds. Any of the embodiments described above with reference to the methods of the invention are equally applicable to the products of the invention.

The following Examples illustrate the invention:

15

1. Example 1

1.1 Materials and Methods

In this manifestation, an azide functional group and an alkyne functional group react with one another in the presence of a copper (I) catalyst to form a triazole ring, in the copper-catalysed alkyne azide cycloaddition reaction (CuAAC).

Copper (II) sulphate, sodium ascorbate, MOPS free acid, sodium phosphate, sodium chloride, calcium chloride, and EDTA were obtained at the highest grade available from standard laboratory suppliers, eg Sigma and Fisher. The copper (I) binding ligand (shown in Figure 11).

Oligonucleotides with thiol end groups, or with internal azide or alkyne groups were from AtdBio (Southampton, UK); unmodified oligonucleotides were from AtdBio, or IDTDNA (Leuven, Belgium). The sequences of the oligonucleotides are shown in SEQ ID NOs: 64 to 78. In all cases of complementary oligonucleotides, the modified bases are offset from one another by one base in order to achieve the link..

30

1.2 Hybridization

When hybridizing oligonucleotides, the hybridization was carried out by incubating the oligonucleotides in buffer at 4°C for at least 1 hour. For oligonucleotides 15 bases long, this

buffer could be 100mM MOPS pH6.8 or 100mM sodium phosphate pH 7.0. For oligos shorter than this, 1000mM NaCl was included. When enzyme inhibitors (10mM EDTA or 20mM CaCl₂) were added, this was in 100mM phosphate pH 6.8 (for EDTA) or 100mM MOPS pH6.8 (for either EDTA or CaCl₂).

5

1.3 Linking reaction

The linking reaction was carried out after an overnight hybridization (for 15bp duplexes and longer) or a 1 hour incubation (for 6-10bp duplexes) at 4°C. The linking reaction was achieved by first pre-mixing copper (II) ion, ascorbate reducing agent, and a copper (I) binding ligand. The three components were mixed at the appropriate concentrations to achieve the final conditions. For instance, if the final required conditions were 1mM copper, 2mM ascorbate, 1mM ligand in 100µl reaction volume, equal volumes of 100mM copper sulphate (in water), 200mM sodium ascorbate (in a buffer at pH <7.0), and 100mM ligand (in water) were mixed in the order specified, then 3µl of the resulting mixture were added to the 100µl reaction. After this addition, the reaction was left at 4°C in the dark for at least 60 minutes.

10

15

Oligos linked	Length of duplex	Link conditions / [Cu ^I], [ascorbate], [ligand], buffer	Link?	Confirmation
76+77	15bp	5µM, 10µM, 7µM, 1	No	Gel
76+77	15bp	0.5mM, 1mM, 0.5mM, 1	Yes	Gel
66+68 (1 link)	15bp	1mM, 2mM, 1mM, 2	Yes	Gel, T _m
67+69 (2 links)	15bp	1mM, 2mM, 1mM, 2	Yes	Gel, T _m
70+73	6bp	1mM, 2mM, 1mM, 3	No	
71+74	8bp	1mM, 2mM, 1mM, 3	Yes	Gel, T _m
72+75	10bp	1mM, 2mM, 1mM, 3	Yes	Gel, T _m

Buffer 1 = 100mM phosphate, 500mM NaCl pH7.0

Buffer 2 = 100mM phosphate pH6.8

20 Buffer 3 = 100mM MOPS, 1000mM NaCl, 20mM CaCl₂ pH6.8

1.4 Attachment reaction

To attach an oligonucleotide to a protein, the oligonucleotides must first be activated. This procedure changes the chemical nature of the supplied oligonucleotides from one that is

inert to the protein to one that can react with the protein. This process makes no difference to the linking reaction.

When required, oligonucleotides were attached to proteins through disulphide bonds. This was achieved by incubating a 5x excess of activated oligonucleotide with freshly reduced protein for 1 hour at 37°C (for exonuclease), or overnight at room temperature (for nanopore). In all cases, oligonucleotides containing the azide base(s) were attached to the nanopore (through E287C on one of the monomers), and the oligonucleotides containing the alkyne base(s) were attached to the exonuclease (through A83C). Unreacted DNA was removed by purification on Streptavidin affinity beads.

10

1.5 Coupling of proteins

Purified modified proteins were incubated together under conditions designed to promote hybridization between the oligonucleotides. For 15 base oligonucleotides and longer this buffer could be 100mM MOPS pH6.8 or 100mM sodium phosphate pH7.0. When enzyme inhibitors were required, 20mM CaCl₂ was supplemented in to 100mM MOPS pH6.8. For 6-10 base oligos, the buffer was 100mM MOPS, 1000mM NaCl, 20mM CaCl₂ pH6.8.

15

The linking reaction was carried out as described above, as these conditions have been shown to be compatible with enzyme activity.

20

Protein 1	Protein 2	Duplex length	Link conditions / [Cu ^I], [ascorbate], [ligand], buffer	ONT colloquial name
Exo-SEQ 78	Pore-SEQ 67	15bp	1mM, 2mM, 1mM, 4	PK V3
Exo-SEQ 73	Pore-SEQ 70	6bp	1mM, 2mM, 1mM, 3	TL

Buffer 4 = 100mM MOPS, 20mM CaCl₂ pH6.8

2. Example 2

2.1 Materials and Methods

25

2.1.1 Overview

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

30

cysteine on the hemolysin, either directly (a disulphide bond), or via a linker molecule between the two cysteine residues.

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).

2.1.2 Modelling Exo I CYS Knockout Mutations

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 were performed for these mutations, allowing predictions to be made about their relative stability. These predictions were used to guide experimental work.

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

CYS to XXX mutations of the *E. coli* Exo I protein structure (PDB accession code: 1fx [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.

20

2.1.3 Expressions and Purification

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 500ml Terrific Broth in 2 litre baffled shake flasks supplemented with 100µg/ml ampicillin and 20µg/ml chloramphenicol were inoculated with colonies from agar plates and grown to OD₆₀₀=2 (approx 6hours) before induction with 0.2mM 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.

2.1.4 Enzyme Assay

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 (Figure 6).

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 Figure 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.

2.2 Results

2.2.1 Modelling Exo I CYS Knockout Mutations

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
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

50

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

5 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.

10 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.

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
5 binding are Asp15, Glu17, Asp108, Asp186 and His181.

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.

10 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.
15 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.

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

20 2.2.2 Identification of 5 Cysteine Residues

Exonuclease I (SEQ ID NO: 6) contains 5 naturally occurring cysteine residues (Figure 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.

25

2.2.3 Removal of all 5 Cysteines by Mutation to Serine

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
30 protein expression was attempted.

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.

2.2.4 Replacement of Individual Cysteines by Mutation to Serine

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

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

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; Figure 2) and more buried residues (C51, C144).
15

2.2.5 Replacement of C306 and C330 with Alternatives to Serine

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

Exo I C306 → D, M, N, S, T

20 Exo I C330 → D, M, N, T, Y, H, L, M, Q

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

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.
25

2.2.6 Combining C98S with C306S and C330T

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

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

5 Activity assays (see Methods) showed that the relative activity of the Exo I mutants was higher than the wild type (Figure 3).

2.2.7 Introduction of Cysteine Attachment Points

10 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 Figure 4.

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

15 The activity of the cysteines was assessed by reacting with maleimide-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 maleimide-PEG. Further rounds of mutagenesis were
20 carried out to replace C144 and C51 to ensure that the protein only react at one position.

2.2.8 Replacement of Residue C144 with Alternatives to Serine

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
25 predicted from the modeling were made:

Exo I C98S, C306S, C330S, C144 \rightarrow A, G, L, **T**, H, **M**

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.

5 2.2.9 Replacement of Residue C51 with Alternatives to Serine

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:

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

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

15 2.2.10 Combination of 5 Cysteine Replacements and Addition of Attachment Point

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

20

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

The two mutants with N-terminal cysteine residues (ONLD0413, ONLD0414) failed to express. All other combinations expressed and were active exonucleases. Combinations
 25 ONLD0415 and ONLD0417 expressed less well than ONLD0416 and ONLD0418. The activity results are shown in Figure 5.

The following combinations were also assessed.

Position 98 and 306

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

5

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

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

10 2.2.11 DSC analysis of proteins

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 biomolecule as a transition mid-point. The transition mid-point is used as a marker of thermal stability.

15 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 T_m was determined to be 35°C, compared to wildtype at 49°C.

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

The mutant AVTMT was found to have slightly higher T_m , however was found more prone to aggregation under some conditions.

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

Sample	ONLD	T _m (melting point) C									
ASTST	ONLD0411	35.5C	35.8C								
ATTTT	ONLD0433	40.5C	40.3C	39.8C	39.8C	39.8	39.6	39.6	39.3	39.7	39.9
WT	ONLD0267	49	48.9								
ATTST	ONLD0432	39.8C	39.9	40	40.1						
CSCST	ONLD0366	41.7	41.8								
AVTST	ONLD0456			39.7C	39.9C						
AGTST	ONLD0453			36C*	36C*						
AKTST	ONLD0454			nd	nd						
ATTMT	ONLD0477					40.8	40.7				
ATTNT	ONLD0478					39.6	39.5				
ATTDI	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		
AVTDI	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	T _m	T _m
		Exp 1a	Exp 1b
Exo-ATTTT-A83C-Strep-Strep	ONLD0540	40.68	40.94

2.2.12 Other mutations of interest: C144I and C330I

5 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.

10 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.

15

2.2.13 Enzyme activity

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 (Figure 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 (Figure 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 (Figure 10).

10

2.3 Conclusion

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.

15

CLAIMS

1. A method of covalently coupling two or more moieties, the method comprising:
 - a) providing a first moiety having covalently attached thereto (i) at least one first linker comprising a first hybridizable region and (ii) at least one first group capable of forming a covalent bond;
 - b) providing a second moiety having covalently attached thereto (i) at least one second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and (ii) at least one second group capable of forming a covalent bond with the first group;
 - c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and link the moieties; and
 - d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups.

2. A method according to claim 1, wherein:
 - a) the at least one first group is covalently attached directly to the first moiety or is attached to the first moiety via a separate linker; and/or
 - b) the at least one second group is covalently attached directly to the second moiety or is attached to the second moiety via a separate linker.

3. A method according to claim 1, wherein the at least one first group is covalently attached to the first moiety via the first linker and/or the at least one second group is covalently attached to the second moiety via the second linker.

4. A method according to any one of claims 1 to 3, wherein the first moiety has 2, 3 or 4 first linkers covalently attached thereto and the second moiety has the corresponding number of second linkers attached thereto and wherein the linkers form one or more pairs of first and second linkers in which the hybridizable regions

in each pair specifically hybridize to each other but do not hybridize to any of the hybridizable regions in the other the pairs.

5. A method according to claim 4, wherein the method comprises:
- (a) providing a first moiety having covalently attached thereto (i) a first linker comprising a first hybridizable region, (ii) a first group capable of forming a covalent bond, (iii) a third linker comprising a third hybridizable region and (iv) a third group capable of forming a covalent bond;
 - (b) providing a second moiety having covalently attached thereto (i) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region, (ii) a second group capable of forming a covalent bond with the first group, (iii) a fourth linker comprising a fourth hybridizable region capable of hybridizing to the third hybridizable region and (iv) a fourth group capable of forming a covalent bond with the third group;
 - (c) contacting the first and second moieties under conditions that allow the first and second hybridizable regions to hybridize and allow the third and fourth hybridizable regions to hybridize and link the moieties; and
 - (d) exposing the linked moieties to conditions that allow the formation of a covalent bond between the first and second groups and the third and fourth groups.
6. A method according to any one of claims 1 to 3, wherein the first moiety has covalently attached thereto (i) a first linker comprising a first hybridizable region, (ii) a first group capable of forming a covalent bond, (iii) a third linker comprising a third hybridizable region and (iv) a third group capable of forming a covalent bond and the second moiety has covalently attached thereto (i) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and a fourth hybridizable region capable of hybridizing to the third hybridizable

region, (ii) a second group capable of forming a covalent bond with the first group and (iii) a fourth group capable of forming a covalent bond with the third group.

7. A method according to claim 1, wherein the method comprises coupling 3, 4 or 5 moieties and wherein each moiety has covalently attached thereto (i) at least one linker which comprises a hybridizable region capable of hybridizing to the hybridizable region of a linker on a different moiety and (ii) at least one covalent group capable of forming a covalent bond with the group on a different moiety.

8. A method according to claim 7, wherein the method comprises:

- a) providing a first moiety having covalently attached thereto (i) a first linker comprising a first hybridizable region and (ii) a first group capable of forming a covalent bond;
- b) providing a second moiety having covalently attached thereto (i) a second linker comprising a second hybridizable region capable of hybridizing to the first hybridizable region and (ii) a second group capable of forming a covalent bond with the first group, (iii) a third linker comprising a third hybridizable region and (iv) a third group capable of forming a covalent bond;
- c) providing a third moiety having covalently attached thereto (i) a fourth linker comprising a fourth hybridizable region capable of hybridizing to the third hybridizable region, (ii) a fourth group capable of forming a covalent bond with the third group;
- d) contacting the first, second and third moieties under conditions that allow the hybridizable regions to hybridize and link the moieties; and
- e) exposing the linked moieties to conditions that allow the formation of a covalent bond between the groups.

9. A method according to any one of claims 5 to 8, wherein the groups are covalently attached to their respective moieties via the linkers comprising one or more hybridizable regions.

10. A method according to any one of the preceding claims, wherein none of the groups react with themselves.
11. A method according to any one of the preceding claims, wherein the first moiety and/or the second moiety is a polypeptide, a protein, a nucleic acid, a chemical group, a dye, an affinity tag, a polymer, a bead, an antibody, a support, an immobilised metal affinity matrix, glutathione sepharose, dextrin sepharose, affinity matrix, IgG sepharose, activated thiol-sepharose, a metal film, a nanoparticle, glass, treated glass, a plastic, a resin surface, a solid state nanopore, a sensor surface and a single molecular detector.
12. A method according to claim 11, wherein the first moiety is a protein and the second moiety is a protein, a nucleic acid, a reporter group, an affinity tag or a support.
13. A method according to claim 12, wherein the first moiety is a nucleic acid binding protein and the second moiety is a transmembrane protein pore.
14. A method according to claim 13, wherein the nucleic acid binding protein is derived from an exonuclease, a nuclease, a polymerase, a helicase or a topoisomerase.
15. A method according to claim 14, wherein the exonuclease comprises the sequence shown in any one of SEQ ID NOs: 6 or a variant thereof.
16. A method according to claim 15, wherein the exonuclease comprises the sequence shown in any one of SEQ ID NOs: 6, 52, 54 and 56 or a variant thereof.
17. A method according to any one of claims 13 to 16, wherein the transmembrane protein pore is α -hemolysin (α -HL).

18. A method according to claim 17, wherein the pore comprises the sequence shown SEQ ID NO: 2 or 4 or a variant thereof.
19. A method according to any one claims 13 to 18, wherein the first and second groups is from 0.05 nm and 10 nm from their respective moieties.
20. A method according to any one of the claims 13 to 19, wherein the distance between the coupled first and second moieties is from 0.1 nm and 500 nm.
21. A method according to any one of the preceding claims, wherein the first and second hybridizable regions comprise a nucleic acid.
22. A method according to claim 21, wherein the nucleic acid is 6 to 15 nucleotides in length.
23. A method according to any one of the preceding claims, wherein the first and second hybridizable regions have an affinity of from 0.1 fM to 0.1 μ M at a concentration from 1 pM to 1 mM.
24. A method according to any one of the preceding claims, wherein the first hybridizable region corresponds to the first 15, 25 or 35 nucleotides from the 5' end of SEQ ID NO: 57.
25. A method according to any one of the preceding claims, wherein the first and second hybridizable regions comprise the following sequences respectively:
 - (a) the nucleotide sequences shown in SEQ ID NOs: 58 and 61;
 - (b) the nucleotide sequences shown in SEQ ID NOs: 59 and 62;
 - (c) the nucleotide sequences shown in SEQ ID NOs: 60 and 63;
 - (d) the nucleotide sequences shown in SEQ ID NOs: 64 and 65;
 - (e) the nucleotide sequences shown in SEQ ID NOs: 66 and 68;
 - (f) the nucleotide sequences shown in SEQ ID NOs: 67 and 69;
 - (g) the nucleotide sequences shown in SEQ ID NOs: 70 and 73;

- (h) the nucleotide sequences shown in SEQ ID NOs: 71 and 74;
- (i) the nucleotide sequences shown in SEQ ID NOs: 72 and 75;
- (j) the nucleotide sequences shown in SEQ ID NOs: 76 and 77; or
- (k) the nucleotide sequences shown in SEQ ID NOs: 78 and 67.

26. A method according to any one of the preceding claims, wherein the first and second groups form a covalent bond by click chemistry.
27. A method according to claim 26, wherein the first group is an azide group and the second group is an alkyne group.
28. A method according to any one of claims 1 to 25, wherein the first and second groups form a disulphide bond, an iodo acetamide bond or a bond via a metathesis reaction.
29. First and second moieties coupled using a method according to any one of the preceding claims.
30. A first moiety coupled to a second moiety using first and second linkers as defined in any one of claims 1 to 28.
31. A pair of first and second linkers as defined in any one of claims 1 to 28.
32. Use of a pair of linkers according to claim 31 to couple a first moiety to a second moiety.

Figure 1

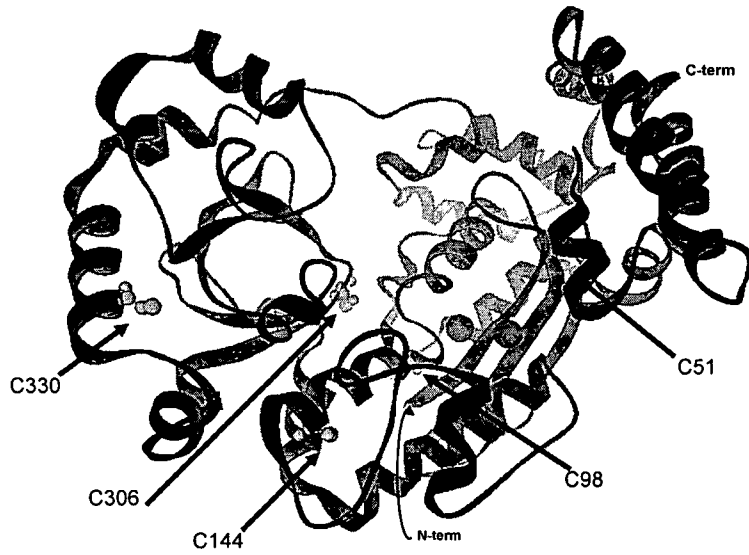


Figure 2

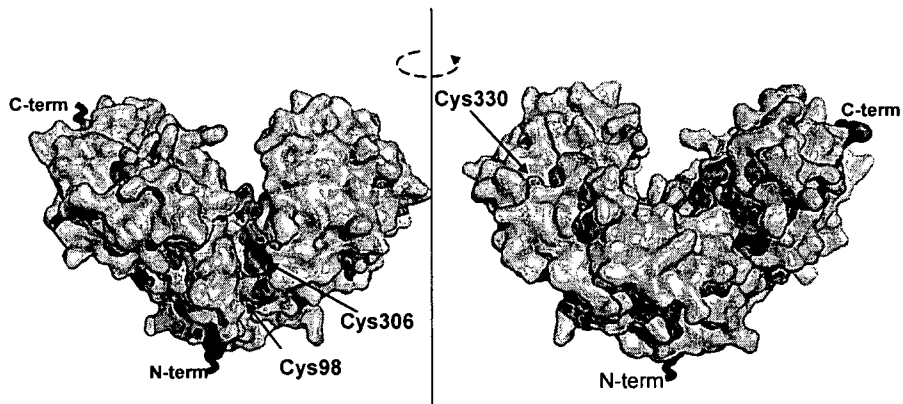


Figure 3

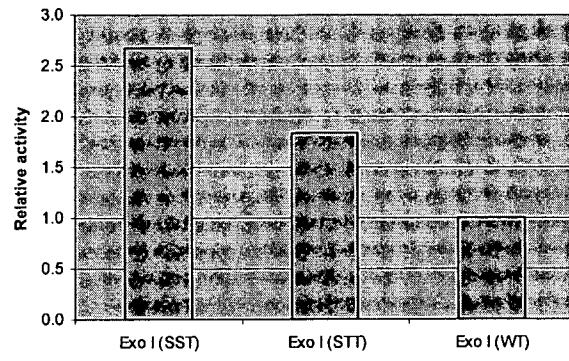


Figure 4

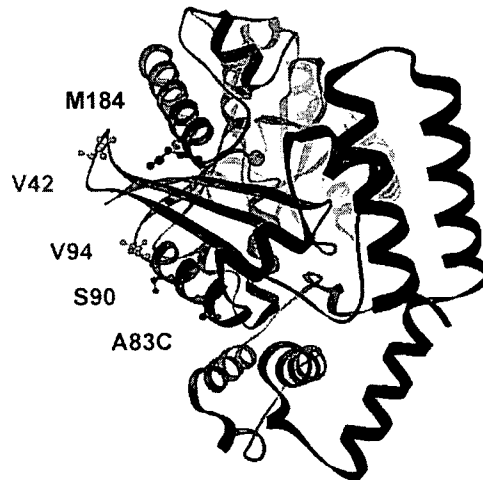


Figure 5a

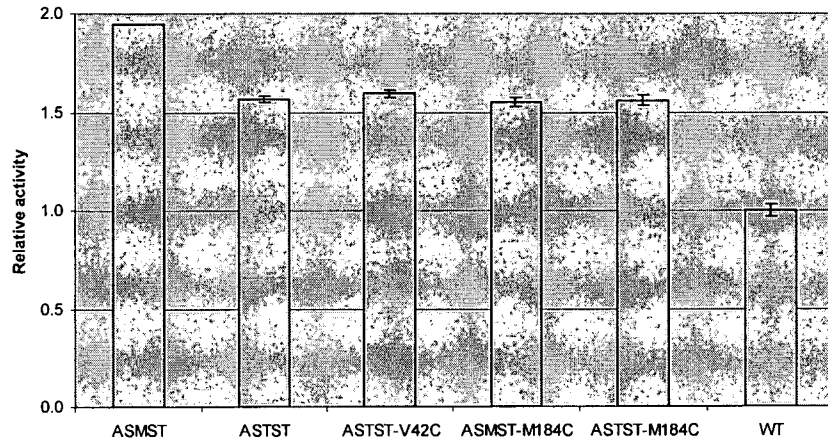


Figure 5b

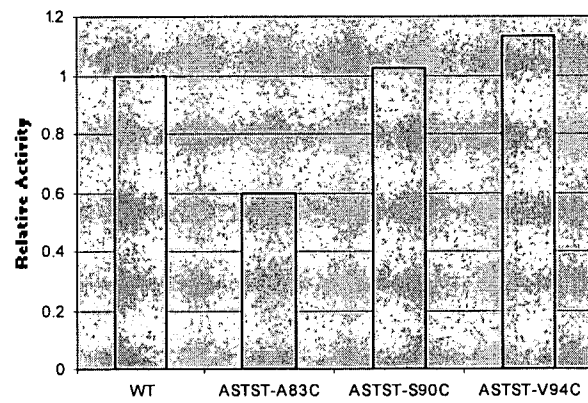


Figure 8

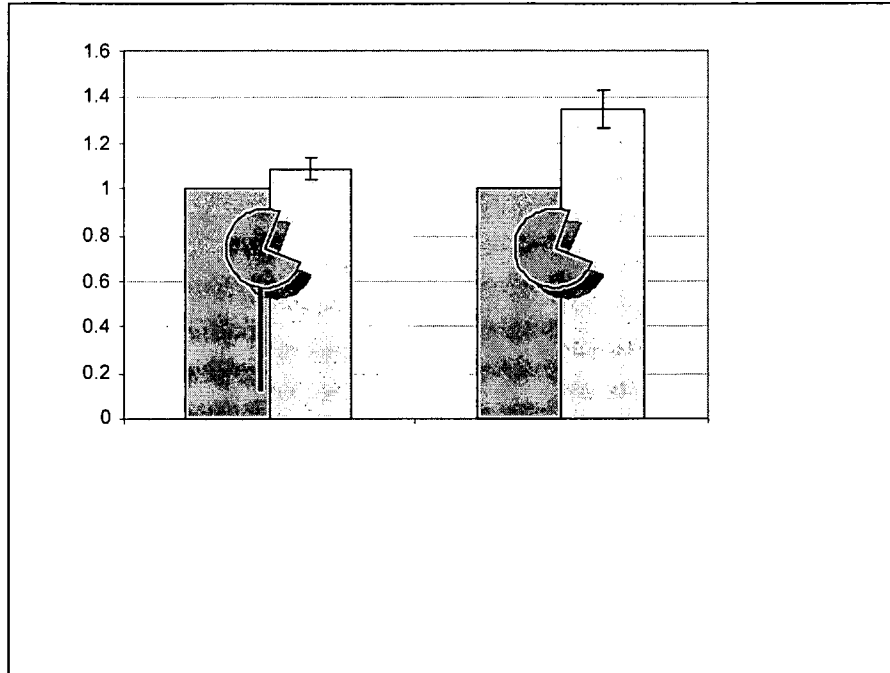


Figure 9

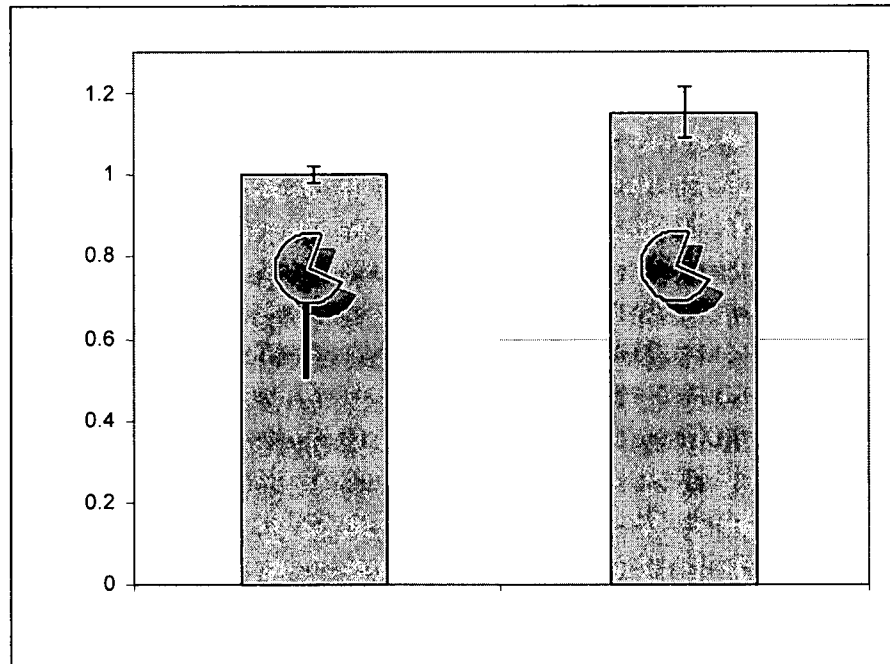


Figure 10

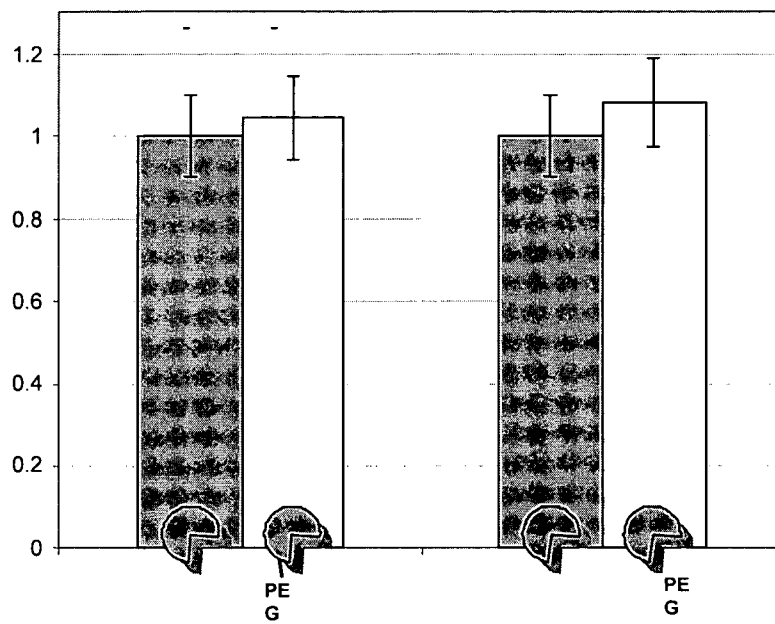
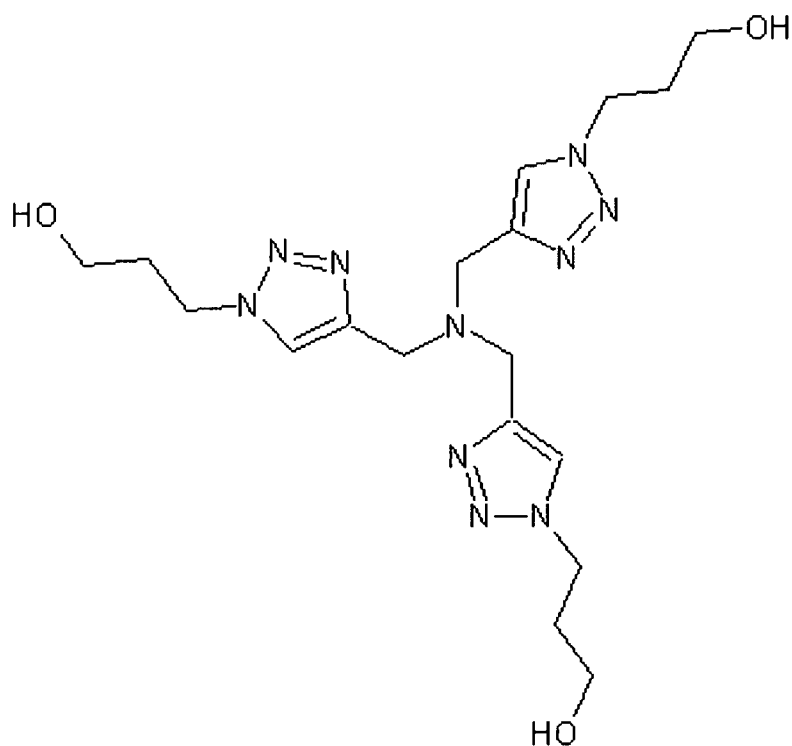


Figure 11



INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2010/000132

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K19/00 C12N15/11
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BRANTON DANIEL ET AL: "The potential and challenges of nanopore sequencing" NATURE BIOTECHNOLOGY, NATURE PUBLISHING GROUP, NEW YORK, NY, US LNKD- DOI:10.1038/NBT.1495, vol. 26, no. 10, 1 October 2008 (2008-10-01), pages 1146-1153, XP002570701 ISSN: 1087-0156 [retrieved on 2008-10-09] figure 1b	1-32
A	WO 2008/124107 A1 (UNIV CALIFORNIA [US]; AKESON MARK A [US]; DEAMER DAVID W [US]; DUNBAR) 16 October 2008 (2008-10-16) page 23, lines 1-3	1-32
	----- -/-- -----	

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

10 May 2010

Date of mailing of the international search report

02/06/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Herrmann, Klaus

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2010/000132

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2005/053961 A1 (AKESON MARK [US] ET AL AKESON MARK [US] ET AL) 10 March 2005 (2005-03-10) paragraph [0076]	1-32
A	----- TUNG C-H ET AL: "PREPARATION AND APPLICATIONS OF PEPTIDE-OLIGONUCLEOTIDE CONJUGATES" BIOCONJUGATE CHEMISTRY, ACS, WASHINGTON, DC, US LNKD- DOI:10.1021/BC0000334, vol. 11, no. 5, 1 September 2000 (2000-09-01), pages 605-618, XP001155364 ISSN: 1043-1802 the whole document	1-32
A	----- LOVRINOVIC ET AL: "Rapid synthesis of DNA-cysteine conjugates for expressed protein ligation" BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, ACADEMIC PRESS INC. ORLANDO, FL, US LNKD- DOI:10.1016/J.BBRC.2005.08.001, vol. 335, no. 3, 30 September 2005 (2005-09-30), pages 943-948, XP005100411 ISSN: 0006-291X abstract page 947, right-hand column	1-32
A	----- WANG Q ET AL: "Bioconjugation by copper(I)-catalyzed azide-alkyne [3+2] cycloaddition" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, AMERICAN CHEMICAL SOCIETY, NEW YORK, US LNKD- DOI:10.1021/JA021381E, vol. 125, no. 11, 1 January 2003 (2003-01-01), pages 3192-3193, XP002450640 ISSN: 0002-7863 the whole document	1-32
A	----- SEO T S ET AL: "Click chemistry to construct fluorescent oligonucleotides for DNA sequencing" JOURNAL OF ORGANIC CHEMISTRY, AMERICAN CHEMICAL SOCIETY, EASTON.; US LNKD- DOI:10.1021/J0026615R, vol. 68, 1 January 2003 (2003-01-01), pages 609-612, XP002977877 ISSN: 0022-3263 the whole document	1-32

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210.

Continuation of Box II.2

Claims Nos.: 1-32(partially)

Independent claim 1 covers methods of covalently coupling two or more moieties comprising the use of any "moieties" having covalently attached thereto any "linker" comprising a hybridizable region and a group capable of forming a covalent bond. However, the application provides support (Art. 6 PCT) and disclosure (Art. 5 PCT) for the use of only a limited number of such moieties and linkers. Thus, a meaningful search over the whole of the scope claimed is impossible. Besides an exonuclease having covalently attached thereto an oligonucleotide via an azide group forming a disulphide bond with a cysteine residue of the exonuclease and alpha-hemolysin having covalently attached thereto an oligonucleotide via an alkyne group forming a disulphide bond with a cysteine residue of the alpha-hemolysin, wherein the two oligonucleotides hybridize with each other and eventually the azide group and the alkyne group form a disulphide bond, no other moieties or linkers, respectively, have been shown to lead to the desired effect of covalently coupling two moieties (cf. p. 46 of present description). Since the independent claims do not contain the technical feature of claims 14, 17, 27 and 28 they do not meet the requirements of Art. 5/6 PCT. Consequently, the search has been limited to a method of covalently coupling two moieties comprising the use of moieties and linkers mentioned above, and subject-matter relating thereto (cf. Art. 17(2)(a)(ii) PCT).

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2) declaration be overcome.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2010/000132

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: **1-32(partially)**
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2010/000132

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008124107	A1	16-10-2008	AU 2008236694 A1 16-10-2008
			CA 2684801 A1 16-10-2008
			CN 101680873 A 24-03-2010
			EP 2156179 A1 24-02-2010
			US 2010035260 A1 11-02-2010

US 2005053961	A1	10-03-2005	NONE
