



(12) 发明专利申请

(10) 申请公布号 CN 102439043 A

(43) 申请公布日 2012. 05. 02

(21) 申请号 201080014472. 2

(22) 申请日 2010. 01. 29

(30) 优先权数据

61/148, 726 2009. 01. 30 US

(85) PCT申请进入国家阶段日

2011. 09. 30

(86) PCT申请的申请数据

PCT/GB2010/000132 2010. 01. 29

(87) PCT申请的公布数据

W02010/086602 EN 2010. 08. 05

(71) 申请人 牛津纳米孔技术有限公司

地址 英国牛津

(72) 发明人 L·贾雅辛格 J·米尔顿

L·麦克尼尔 J·克拉克 J·怀特

R·莫伊希

(74) 专利代理机构 北京北翔知识产权代理有限公司 11285

代理人 张广育 姜建成

(51) Int. Cl.

C07K 19/00 (2006. 01)

C12N 15/11 (2006. 01)

权利要求书 3 页 说明书 45 页

序列表 15 页 附图 5 页

(54) 发明名称

杂交连接物

(57) 摘要

本发明提供一种共价连接两个或多个部分的方法,所述方法包括:(a) 提供第一部分,其上共价连接有:(i) 至少一个第一连接物,其包含第一可杂交区域;和(ii) 至少一个能够形成共价键的第一基团;(b) 提供第二部分,其上共价连接有:(i) 至少一个第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域;和(ii) 至少一个能够与所述第一基团形成共价键的第二基团;(c) 在允许所述第一和第二可杂交区域杂交并连接所述部分的条件下使所述第一和第二部分接触;并且(d) 将所述连接的部分暴露于允许在所述第一和第二基团之间形成共价键的条件。

1. 一种共价偶联两个或多个部分的方法,所述方法包括:

(a) 提供第一部分,其上共价连接有:(i)至少一个第一连接物,其包含第一可杂交区域;和(ii)至少一个能够形成共价键的第一基团;

(b) 提供第二部分,其上共价连接有:(i)至少一个第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域;和(ii)至少一个能够与所述第一基团形成共价键的第二基团;

(c) 在允许所述第一和第二可杂交区域杂交并连接所述部分的条件下使所述第一和第二部分接触;并且

(d) 将所述连接的部分暴露于允许所述第一和第二基团之间形成共价键的条件。

2. 权利要求1的方法,其中:

a) 所述至少一个第一基团直接共价连接至所述第一部分,或者通过单独的连接物连接至所述第一部分,并且/或者

b) 所述至少一个第二基团直接共价连接至所述第二部分,或者通过单独的连接物连接至所述第二部分。

3. 权利要求1的方法,其中所述至少一个第一基团通过所述第一连接物共价连接至所述第一部分,并且/或者所述至少一个第二基团通过所述第二连接物共价连接至所述第二部分。

4. 权利要求1至3任一项的方法,其中所述第一部分共价连接有2、3或4个第一连接物,而所述第二部分连接有相应数目的第二连接物,并且其中所述连接物形成一个或多个第一和第二连接物的对,在所述连接物对中,每对中的可杂交区域相互特异性杂交,但不与其他连接物对中的任何可杂交区域杂交。

5. 权利要求4的方法,其中所述方法包括:

(a) 提供第一部分,其上共价连接有:(i)第一连接物,其包含第一可杂交区域;(ii)能够形成共价键的第一基团;(iii)包含第三可杂交区域的第三连接物;和(iv)能够形成共价键的第三基团;

(b) 提供第二部分,其上共价连接有:(i)第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域;(ii)能够与所述第一基团形成共价键的第二基团;(iii)第四连接物,其包含能够与所述第三可杂交区域杂交的第四可杂交区域;和(iv)能够与所述第三基团形成共价键的第四基团;

(c) 在允许所述第一和第二可杂交区域杂交并允许所述第三和第四可杂交区域杂交并连接所述部分的条件下使所述第一和第二部分接触;并且

(d) 将所述连接的部分暴露于允许所述第一和第二基团之间和所述第三和第四基团之间形成共价键的条件。

6. 权利要求1至3任一项的方法,其中所述第一部分共价连接有:(i)包含第一可杂交区域的第一连接物,(ii)能够形成共价键的第一基团,(iii)包含第三可杂交区域的第三连接物,和(iv)能够形成共价键的第三基团;并且所述第二部分共价连接有:(i)第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域和能够与所述第三可杂交区域杂交的第四可杂交区域,(ii)能够与所述第一基团形成共价键的第二基团,和(iii)能够与所述第三基团形成共价键的第四基团。

7. 权利要求 1 的方法,其中所述方法包括偶联 3、4 或 5 个部分,并且其中每个部分均共价连接有:(i) 至少一个连接物,其包含能够与不同部分上的连接物的可杂交区域杂交的可杂交区域,和(ii) 至少一个能够与不同部分上的基团形成共价键的共价基团。

8. 权利要求 7 的方法,其中所述方法包括:

a) 提供第一部分,其共价连接有:(i) 包含第一可杂交区域的第一连接物,和(ii) 能够形成共价键的第一基团;

b) 提供第二部分,其共价连接有:(i) 第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域,和(ii) 能够与所述第一基团形成共价键的第二基团,(iii) 包含第三可杂交区域的第三连接物,和(iv) 能够形成共价键的第三基团;

c) 提供第三部分,其共价连接有:(i) 第四连接物,其包含能够与所述第三可杂交区域杂交的第四可杂交区域,(ii) 能够与所述第三基团形成共价键的第四基团;

d) 在允许所述可杂交区域杂交并连接所述部分的条件下使所述第一、第二和第三部分接触;并且

e) 将所述连接的部分暴露于允许所述基团之间形成共价键的条件。

9. 权利要求 5 至 8 任一项的方法,其中所述基团通过包含一个或多个可杂交区域的所述连接物共价连接至它们各自的部分。

10. 前述权利要求任一项的方法,其中全部所述基团都不与自身反应。

11. 前述权利要求任一项的方法,其中所述第一部分和/或所述第二部分是多肽、蛋白质、核酸、化学基团、染料、亲和标签、聚合物、小珠、抗体、载体、固定的金属亲和基质、谷胱甘肽琼脂糖凝胶、糊精琼脂糖凝胶、亲和基质、IgG 琼脂糖凝胶、活化的巯基-琼脂糖凝胶、金属膜、纳米颗粒、玻璃、处理的玻璃、塑料、树脂表面、固态纳米孔、传感器表面和单分子探测器。

12. 权利要求 11 的方法,其中所述第一部分是蛋白质,所述第二部分是蛋白质、核酸、报告基团、亲和标签或载体。

13. 权利要求 12 的方法,其中所述第一部分是核酸结合蛋白,所述第二部分是跨膜蛋白质孔。

14. 权利要求 13 的方法,其中所述核酸结合蛋白源自于核酸外切酶、核酸酶、聚合酶、螺旋酶或拓扑异构酶。

15. 权利要求 14 的方法,其中所述核酸外切酶包含 SEQ ID NO:6 或其变体中任一所示的序列。

16. 权利要求 15 的方法,其中所述核酸外切酶包含 SEQ ID NO:6、52、54 和 56 或其变体中任一所示的序列。

17. 权利要求 13 至 16 任一项的方法,其中所述跨膜蛋白质孔是 α -溶血素(α -HL)。

18. 权利要求 17 的方法,其中所述孔包含 SEQ ID NO:2 或 4 或其变体所示的序列。

19. 权利要求 13 至 18 任一项的方法,其中所述第一和第二基团距离它们各自的部分 0.05nm 至 10nm。

20. 权利要求 13 至 19 任一项的方法,其中所述偶联的第一和第二部分之间的距离为 0.1nm 至 500nm。

21. 前述权利要求任一项的方法,其中所述第一和第二可杂交区域包含核酸。

22. 权利要求 21 的方法,其中所述核酸长 6 至 15 个核苷酸。
23. 前述权利要求任一项的方法,其中所述第一和第二可杂交区域在 1pM 至 1mM 的浓度下具有 0.1fM 至 0.1 μ M 的亲和力。
24. 前述权利要求任一项的方法,其中所述第一可杂交区域对应于来自 SEQ ID NO:57 的 5' 末端的前 15、25 或 35 个核苷酸。
25. 前述权利要求任一项的方法,其中所述第一和第二可杂交区域分别包含以下序列:
- (a) SEQ ID NO:58 和 61 所示的核苷酸序列;
 - (b) SEQ ID NO:59 和 62 所示的核苷酸序列;
 - (c) SEQ ID NO:60 和 63 所示的核苷酸序列;
 - (d) SEQ ID NO:64 和 65 所示的核苷酸序列;
 - (e) SEQ ID NO:66 和 68 所示的核苷酸序列;
 - (f) SEQ ID NO:67 和 69 所示的核苷酸序列;
 - (g) SEQ ID NO:70 和 73 所示的核苷酸序列;
 - (h) SEQ ID NO:71 和 74 所示的核苷酸序列;
 - (i) SEQ ID NO:72 和 75 所示的核苷酸序列;
 - (j) SEQ ID NO:76 和 77 所示的核苷酸序列;或
 - (k) SEQ ID NO:78 和 67 所示的核苷酸序列。
26. 前述权利要求任一项的方法,其中所述第一和第二基团通过链接化学形成共价键。
27. 权利要求 26 的方法,其中所述第一基团是叠氮基团,第二基团是炔基团。
28. 权利要求 1 至 25 任一项的方法,其中所述第一和第二基团形成二硫键、碘乙酰胺键或通过复分解反应形成的键。
29. 使用前述权利要求任一项的方法偶联的第一和第二部分。
30. 使用权利要求 1 至 28 任一项中所定义的第一和第二连接物偶联至第二部分的第一部分。
31. 权利要求 1 至 28 任一项中所定义的第一和第二连接物的对。
32. 权利要求 31 的连接物对用于将第一部分偶联至第二部分的用途。

杂交连接物

技术领域

[0001] 本发明涉及一种共价偶联 (couple) 两个或多个部分的方法。本发明涉及两个或多个各自包含可杂交区域的连接物的用途以及两个或多个能够形成共价键的基团的用途。所述连接物上的可杂交区域杂交并连接所述部分。然后所述连接的部分通过所述基团之间形成共价键而偶联。

背景技术

[0002] 两个或多个部分的共价偶联在化学和生物化学领域中有很多应用。当使用化学偶联方法时,部分之间的反应速率依赖于能够反应的所述部分的浓度。如果所述部分的浓度低,那么所述反应速率可能会由所述部分的浓度限速。对于很多应用来说,所述部分的浓度可能较低,因此需要提高反应速率的方法(或者,降低所需的部分的浓度的方法)。

[0003] 已知可以制备一些分子或分子集合来进行自组装以形成更大的、更复杂的结构,这些的实例包括:DNA、蛋白质、脂质、液晶、聚合物等。这些物质的分子组分通常相互之间有很高的亲和力,并且在低浓度下可以自组装。自组装系统的一个缺点是它们经常通过非共价相互作用聚集,因此稳定性有限。

发明内容

[0004] 发明人出乎意料地证明利用自组装系统的高亲和性,可在低浓度下使两个或多个部分共价偶联。更具体地,发明人出乎意料地证明两个或多个高亲和力杂交连接物可被用于连接两个或多个部分,然后再将所述部分使用能够一起共价反应的化学基团偶联。所得到的复合物的稳定性比仅使用高亲和力杂交连接物形成的等价非共价复合物的稳定性高。

[0005] 因此,本发明提供共价偶联两个或多个部分的方法,所述方法包括:

[0006] (a) 提供第一部分,其上共价连接有:(i) 至少一个第一连接物,其包含第一可杂交区域;和(ii) 至少一个能够形成共价键的第一基团;

[0007] (b) 提供第二部分,其上共价连接有:(i) 至少一个第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域;和(ii) 至少一个能够与所述第一基团形成共价键的第二基团;

[0008] (c) 在允许所述第一和第二可杂交区域杂交并连接所述部分的条件下使所述第一和第二部分接触;并且

[0009] (d) 将所述连接的部分暴露于允许所述第一和第二基团之间形成共价键的条件。

[0010] 本发明还提供:

[0011] - 使用本发明的方法偶联的第一和第二部分;

[0012] - 使用本发明的第一和第二连接物偶联至第二部分的第一部分;

[0013] - 一对本发明的第一和第二连接物;和

[0014] - 本发明的一对连接物用于将第一部分偶联至第二部分的用途。

附图说明

[0015] 图 1 显示 EcoExo I 结构的卡通图,其显示天然存在的半胱氨酸的位置和残基数目(从晶体结构得到)。

[0016] 图 2 显示 EcoExo I 的表面图,显示可及的半胱氨酸的位置(从晶体结构得到)。

[0017] 图 3 显示与野生型比较的 Exo I 突变体的相对活性。

[0018] 图 4 显示 EcoExo I 的图,显示可引入用于与蛋白质纳米孔偶联的单个半胱氨酸的位置。

[0019] 图 5a 和 5b 显示与野生型比较的 Exo I 突变体的相对活性。

[0020] 图 6 显示核酸外切酶 I 测定法的图示。

[0021] 图 7 显示来自核酸外切酶 I 测定法的数据的实例。

[0022] 图 8 显示连接物与 Exo-ATTTT-A83C 的连接不影响酶活性。Y-轴是相对活性。左边的一对柱是 ONLP1403,右边的一对柱是 ONLP1405。在每对柱中,左边柱(深色柱)是带有连接物的酶,右边柱(浅色柱)是没有连接物的酶。

[0023] 图 9 显示 PNA 与 Exo-ATTTT-A83C 的连接不影响酶活性。Y-轴是相对活性。右边柱是 ONLP1498,左边柱是 ONLP1499。

[0024] 图 10 显示 PEG 连接物与 Exo-ATTTT 或 Exo-ATTTT-M184C 的连接不影响酶活性。Y-轴是相对活性。左边的一对柱是 Exo-ATTTT,右边的一对柱是 Exo-ATTTT-M184C。在每对柱中,左边柱(深色柱)是有连接物的酶,右边的(浅色柱)是没有连接物的酶。

[0025] 图 11 显示在实施例中所使用的铜(I)结合配体。

[0026] 序列表的说明

[0027] SEQ ID NO :1 显示编码野生型 α -溶血素(α -HL)的一个亚基的多核苷酸序列。

[0028] SEQ ID NO :2 显示野生型 α -HL 的一个亚基的氨基酸序列。氨基酸 2 至 6、73 至 75、207 至 209、214 至 216 和 219 至 222 形成 α -螺旋。氨基酸 22 至 30、35 至 44、52 至 62、67 至 71、76 至 91、98 至 103、112 至 123、137 至 148、154 至 159、165 至 172、229 至 235、243 至 261、266 至 271、285 至 286 和 291 至 293 形成 β -折叠。所有其他非末端氨基酸,也就是 7 至 21、31 至 34、45 至 51、63 至 66、72、92 至 97、104 至 111、124 至 136、149 至 153、160 至 164、173 至 206、210 至 213、217、218、223 至 228、236 至 242、262 至 265、272 至 274 和 287 至 290 形成环区域。氨基酸 1 和 294 是末端氨基酸。

[0029] SEQ ID NO :3 显示编码 α -HL L135C/N139Q(HL-CQ)的一个亚基的多核苷酸序列。

[0030] SEQ ID NO :4 显示 α -HL L135C/N139Q(HL-CQ)的一个亚基的氨基酸序列。在野生型 α -HL 中形成 α -螺旋、 β -折叠和环区域的相同氨基酸在该亚基中形成相应的区域。

[0031] SEQ ID NO :5 显示衍生自大肠杆菌(*E. coli*)*sbcB* 基因的密码子优化的多核苷酸序列。它编码来自大肠杆菌的核酸外切酶 I(EcoExo I)。

[0032] SEQ ID NO :6 显示来自大肠杆菌的核酸外切酶 I(EcoExo I)的氨基酸序列。该酶从 5' 至 3' 方向从单链 DNA(ssDNA)持续消化 5' 单磷酸核苷。氨基酸 60 至 68、70 至 78、80 至 93、107 至 119、124 至 128、137 至 148、165 至 172、182 至 211、213 至 221、234 至 241、268 至 286、313 至 324、326 至 352、362 至 370、373 至 391、401 至 454 和 457 至 475 形成 α -螺旋。氨基酸 10 至 18、28 至 26、47 至 50、97 至 101、133 至 136、229 至 232、243 至 251、258 至 263、298 至 302 和 308 至 311 形成 β -折叠。所有其他非末端氨基酸,19 至 27、37 至 46、51

至 59、69、79、94 至 96、102 至 106、120 至 123、129 至 132、149 至 164、173 至 181、212、222 至 228、233、242、252 至 257、264 至 267、287 至 297、303 至 307、312、325、353 至 361、371、372、392 至 400、455 和 456 形成环。氨基酸 1 至 9 是末端氨基酸。酶整体折叠为三个区域结合形成外形为字母 C 的分子,但是在晶体结构中无序的残基 355-358 可有效地将该 C 转变为 O-样形状。氨基末端 (1-206) 形成核酸外切酶结构域,与 DnaQ 超家族有同源性,后面的残基 (202-354) 形成 SH3-样结构域,羧基结构域 (359-475) 伸出所述核酸外切酶结构域以形成所述分子的 C-样形状。EcoExo I 的 4 个酸性残基与所述 DnaQ 超家族的活性位点残基是保守的 (对应于 D15、E17、D108 和 D186)。据称残基 D15 和 108 结合有单个金属离子。DNA 的水解可能是通过活化的水分子攻击易被切断的磷酸基来催化的, H181 是催化残基并对准所述核苷酸底物。

[0033] SEQ ID NO :7 显示编码 EcoExo I C98S/C306S/C330T/C51A (ONLD0393) 的密码子优化的多核苷酸序列。

[0034] SEQ ID NO :8 显示 EcoExo I C98S/C306S/C330T/C51A (ONLD0393) 的氨基酸序列。

[0035] SEQ ID NO :9 显示编码 EcoExo I C98S/C306S/C330T/C144M (ONLD0403) 的密码子优化的多核苷酸序列。

[0036] SEQ ID NO :10 显示 EcoExo I C98S/C306S/C330T/C144M (ONLD0403) 的氨基酸序列。

[0037] SEQ ID NO :11 显示编码 EcoExo I C98S/C306S/C330T/C144T (ONLD0404) 的密码子优化多核苷酸序列。

[0038] SEQ ID NO :12 显示 EcoExo I C98S/C306S/C330T/C144T (ONLD0404) 的氨基酸序列。

[0039] SEQ ID NO :13 显示编码 EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415) 的密码子优化的多核苷酸序列。

[0040] SEQ ID NO :14 显示 EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415) 的氨基酸序列。

[0041] SEQ ID NO :15 显示编码 EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416) 的密码子优化的多核苷酸序列。

[0042] SEQ ID NO :16 显示 EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416) 的氨基酸序列。

[0043] SEQ ID NO :17 显示编码 EcoExo I C51A/C98S/C144M/C306S/C330T/M184C (ONLD0417) 的密码子优化的多核苷酸序列。

[0044] SEQ ID NO :18 EcoExo I C51A/C98S/C144M/C306S/C330T/M184C (ONLD0417) 的氨基酸序列。

[0045] SEQ ID NO :19 显示编码 EcoExo I C51A/C98S/C144T/C306S/C330T/M184C (ONLD0418) 的密码子优化的多核苷酸序列。

[0046] SEQ ID NO :20 显示 EcoExo I C51A/C98S/C144T/C306S/C330T/M184C (ONLD0418) 的氨基酸序列。

[0047] SEQ ID NO :21 显示编码 EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411) 的密码子优化的多核苷酸序列。

[0048] SEQ ID NO :22 显示 EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411) 的氨基酸序列。

[0049] SEQ ID NO :23 显示编码 EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432) 的密码子优化的多核苷酸序列。

[0050] SEQ ID NO :24 :显示 EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432) 的氨基酸序列。

[0051] SEQ ID NO :25 显示编码 EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433) 的密码子优化的多核苷酸序列。

[0052] SEQ ID NO :26 显示 EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433) 的氨基酸序列。

[0053] SEQ ID NO :27 显示编码 EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451) 的密码子优化的多核苷酸序列。

[0054] SEQ ID NO :28 显示 EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451) 的氨基酸序列。

[0055] SEQ ID NO :29 显示编码 EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452) 的密码子优化的多核苷酸序列。

[0056] SEQ ID NO :30 显示 EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452) 的氨基酸序列。

[0057] SEQ ID NO :31 显示编码 EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453) 的密码子优化的多核苷酸序列。

[0058] SEQ ID NO :32 显示 EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453) 的氨基酸序列。

[0059] SEQ ID NO :33 显示编码 EcoExo I C51A/C98D/C144T/C306S/C330T (ONLD0491) 的密码子优化的多核苷酸序列。

[0060] SEQ ID NO :34 显示 EcoExo I C51A/C98D/C144T/C306S/C330T (ONLD0491) 的氨基酸序列。

[0061] SEQ ID NO :35 显示编码 EcoExo IC51A/C98K/C144T/C306S/C330T (ONLD0454) 的密码子优化的多核苷酸序列。

[0062] SEQ ID NO :36 显示 EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454) 的氨基酸序列。

[0063] SEQ ID NO :37 显示编码 EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455) 的密码子优化的多核苷酸序列。

[0064] SEQ ID NO :38 显示 EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455) 的氨基酸序列。

[0065] SEQ ID NO :39 显示编码 EcoExo IC51A/C98V/C144T/C306S/C330T (ONLD0456) 的密码子优化的多核苷酸序列。

[0066] SEQ ID NO :40 显示 EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456) 的氨基酸序列。

[0067] SEQ ID NO :41 显示编码 EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476) 的

密码子优化的多核苷酸序列。

[0068] SEQ ID NO :42 显示 EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476) 的氨基酸序列。

[0069] SEQ ID NO :43 显示编码 EcoExo IC51A/C98T/C144T/C306M/C330T (ONLD0477) 的密码子优化的多核苷酸序列。

[0070] SEQ ID NO :44 显示 EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477) 的氨基酸序列。

[0071] SEQ ID NO :45 显示编码 EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478) 的密码子优化的多核苷酸序列。

[0072] SEQ ID NO :46 显示 EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478) 的氨基酸序列。

[0073] SEQ ID NO :47 显示编码 EcoExo IC51A/C98T/C144T/C306D/C330T (ONLD0479) 的密码子优化的多核苷酸序列。

[0074] SEQ ID NO :48 显示 EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479) 的氨基酸序列。

[0075] SEQ ID NO :49 显示编码 EcoExo IC51A/C98T/C144T/C306A/C330T (ONLD0480) 的密码子优化的多核苷酸序列。

[0076] SEQ ID NO :50 显示 EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480) 的氨基酸序列。

[0077] 在 SEQ ID NO :7 至 50 所描述的全部突变体中,在野生型 EcoExo I 中形成 α -螺旋、 β -折叠和环区域的相同氨基酸在该突变体中形成相应区域。

[0078] SEQ ID NO :51 显示衍生自大肠杆菌 xthA 基因的密码子优化的多核苷酸序列。它编码大肠杆菌的核酸外切酶 III。

[0079] SEQ ID NO :52 显示来自大肠杆菌的核酸外切酶 III 的氨基酸序列。该酶从 3'-5' 方向从双链 DNA (dsDNA) 的一条链上分配性消化 5' 单磷酸核苷。在链上的酶启动需要大约 4 个核苷酸的 5' 突出。氨基酸 11 至 13、15 至 25、39 至 41、44 至 49、85 至 89、121 至 139、158 至 160、165 至 174、181 至 194、198 至 202、219 至 222、235 至 240 和 248 至 252 形成 α -螺旋。氨基酸 2 至 7、29 至 33、53 至 57、65 至 70、75 至 78、91 至 98、101 至 109、146 至 151、195 至 197、229 至 234 和 241 至 246 形成 β -折叠。所有其他非末端氨基酸,8 至 10、26 至 28、34 至 38、42、43、50 至 52、58 至 64、71 至 74、79 至 84、90、99、100、110 至 120、140 至 145、152 至 157、161 至 164、175 至 180、203 至 218、223 至 228、247 和 253 至 261 形成环。氨基酸 1、267 和 268 是末端氨基酸。酶活性位点是通过连接 $\beta_1-\alpha_1$ 、 $\beta_3-\beta_4$ 、 $\beta_5-\beta_6$ 、 $\beta_{III}-\alpha_I$ 、 $\beta_{IV}-\alpha_{II}$ 和 $\beta_V-\beta_{VI}$ 的环区域 (分别由氨基酸 8-10、58-64、90、110-120、152-164、175-180、223-228 和 253-261 构成) 形成的。单个二价金属离子结合在残基 E34 上,帮助 D229 和 H259 组氨酸-天冬氨酸催化对亲核攻击磷酸二酯键。

[0080] SEQ ID NO :53 显示衍生自嗜热栖热菌 (*T. thermophilus*) 的 recJ 基因的密码子优化的多核苷酸序列。它编码来自嗜热栖热菌的 RecJ 酶 (TthRecJ-cd)。

[0081] SEQ ID NO :54 显示嗜热栖热菌的 RecJ 酶 (TthRecJ-cd) 的氨基酸序列。该酶从 5'-3' 方向从 ssDNA 持续消化 5' 单磷酸核苷。链上的酶启动需要至少 4 个核苷酸。氨基

酸 19 至 33、44 至 61、80 至 89、103 至 111、136 至 140、148 至 163、169 至 183、189 至 202、207 至 217、223 至 240、242 至 252、254 至 287、302 至 318、338 至 350 和 365 至 382 形成 α -螺旋。氨基酸 36 至 40、64 至 68、93 至 96、116 至 120、133 至 135、294 至 297、321 至 325、328 至 332、352 至 355 和 359 至 363 形成 β -折叠。所有其他非末端氨基酸,34、35、41 至 43、62、63、69 至 79、90 至 92、97 至 102、112 至 115、121 至 132、141 至 147、164 至 168、184 至 188、203 至 206、218 至 222、241、253、288 至 293、298 至 301、319、320、326、327、333 至 337、351 至 358 和 364 形成环。氨基酸 1 至 18 和 383 至 425 是末端氨基酸。仅对嗜热栖热菌 (*Thermus thermophilus*) 的 RecJ 的核心结构域 (残基 40-463) 的晶体结构进行了解析。为了确保 RecJ 核心结构域的翻译起始和体内表达,在其氨基末端加入了甲硫氨酸残基,而这在晶体结构信息中不存在。所解析的结构显示两个结构域 -- 氨基 (2-253) 和羧基 (288-463) 区域,通过长的 α -螺旋 (254-287) 连接。催化残基 (D46、D98、H122 和 D183) 与单个二价金属离子配位以亲核攻击所述磷酸二酯键。D46 和 H120 被认为是所述催化对;但是,对大肠杆菌 RecJ 中的这些保守残基任意一个的突变都被证明可消除活性。

[0082] SEQ ID NO :55 显示衍生自噬菌体 λ exo (redX) 基因的密码子优化的多核苷酸序列。它编码噬菌体 λ 核酸外切酶。

[0083] SEQ ID NO :56 显示噬菌体 λ 核酸外切酶的氨基酸序列。所述序列是组装成三聚体的三个相同亚基之一的序列。所述酶高度持续地从 dsDNA 的一条链上消化核苷酸,方向是 3' -5'。在链上的酶启动优先地需要大约 4 个具有 5' 磷酸的核苷酸的 5' 突出端。氨基酸 3 至 10、14 至 16、22 至 26、34 至 40、52 至 67、75 至 95、135 至 149、152 至 165 和 193 至 216 形成 α -螺旋。氨基酸 100 至 101、106 至 107、114 至 116、120 至 122、127 至 131、169 至 175 和 184 至 190 形成 β -折叠。所有其他非末端氨基酸,11 至 13、17 至 21、27 至 33、41 至 51、68 至 74、96 至 99、102 至 105、108 至 113、117 至 119、123 至 126、132 至 134、150 至 151、166 至 168、176 至 183、191 至 192、217 至 222 形成环。氨基酸 1、2 和 226 是末端氨基酸。 λ 核酸外切酶是同源三聚体,所述同源三聚体形成圆环结构 (toroid),其具有穿过中间的锥形通道,该通道在一端显然大到足以使 dsDNA 进入,而另一端仅能使 ssDNA 出去。未确定催化残基,但是单个二价金属离子看起来通过残基 D119、E129 和 L130 与各个亚基结合。

[0084] SEQ ID NO :57 显示核酸序列,从其可产生优选的核酸连接物。

[0085] SEQ ID NO :58 显示优选的核酸连接物。MAL 是马来酰亚胺。该连接物与 SEQ ID NO :61 结合使用。

[0086] SEQ ID NO :59 显示优选的核酸连接物。MAL 是马来酰亚胺。该连接物与 SEQ ID NO :62 结合使用。

[0087] SEQ ID NO :60 显示优选的核酸连接物。MAL 是马来酰亚胺。该连接物与 SEQ ID NO :63 结合使用。

[0088] SEQ ID NO :61 显示优选的核酸连接物。MAL 是马来酰亚胺。该连接物与 SEQ ID NO :58 互补并且用于与 SEQ ID NO :58 结合使用。

[0089] SEQ ID NO :62 显示优选的核酸连接物。MAL 是马来酰亚胺。该连接物与 SEQ ID NO :59 互补并且与 SEQ ID NO :59 结合使用。

[0090] SEQ ID NO :63 显示优选的核酸连接物。MAL 是马来酰亚胺。该连接物与 SEQ ID

NO :60 互补并且与 SEQ ID NO :60 结合使用。

[0091] SEQ ID NO :64 显示优选的核酸连接物。该连接物与 SEQ ID NO :65 结合使用。

[0092] SEQ ID NO :65 显示优选的核酸连接物。该连接物与 SEQ ID NO :64 结合使用。

[0093] SEQ ID NO :66 显示优选的核酸连接物。该连接物与 SEQ ID NO :68 结合使用。

[0094] SEQ ID NO :67 显示优选的核酸连接物。该连接物与 SEQ ID NO :69 结合使用。

[0095] SEQ ID NO :68 显示优选的核酸连接物。该连接物与 SEQ ID NO :66 互补并且与 SEQ ID NO :66 结合使用。

[0096] SEQ ID NO :69 显示优选的核酸连接物。该连接物与 SEQ ID NO :67 互补并且与 SEQ ID NO :67 结合使用。

[0097] SEQ ID NO :70 显示优选的核酸连接物。该连接物与 SEQ ID NO :73 互补并且与 SEQ ID NO :73 结合使用。

[0098] SEQ ID NO :71 显示优选的核酸连接物。该连接物与 SEQ ID NO :74 互补并且与 SEQ ID NO :74 结合使用。

[0099] SEQ ID NO :72 显示优选的核酸连接物。该连接物与 SEQ ID NO :75 结合使用。

[0100] SEQ ID NO :73 显示优选的核酸连接物。该连接物与 SEQ ID NO :70 结合使用。

[0101] SEQ ID NO :74 显示优选的核酸连接物。该连接物与 SEQ ID NO :71 结合使用。

[0102] SEQ ID NO :75 显示优选的核酸连接物。该连接物与 SEQ ID NO :72 互补并且与 SEQ ID NO :72 结合使用。

[0103] SEQ ID NO :76 显示优选的核酸连接物。该连接物与 SEQ ID NO :77 结合使用。

[0104] SEQ ID NO :77 显示优选的核酸连接物。该连接物与 SEQ ID NO :76 结合使用。

[0105] SEQ ID NO :78 显示优选的核酸连接物。该连接物与 SEQ ID NO :67 互补并且与 SEQ ID NO :67 结合使用。

具体实施方式

[0106] 应理解所公开的产物和方法的不同应用可根据本领域具体需要而调整。还应理解本文使用的术语仅用于描述本发明的具体实施方案的目的,并且无意于限制。

[0107] 此外,除非上下文明确说明相反,否则本说明书和所附权利要求书所使用的单数形式“一种”、“一个”和“所述”包括复数指示物。因此,例如,提及“一种连接物”包括“多种连接物”,提及“一种跨膜蛋白质孔”包括两种或多种这样的孔,提及“一种部分”包括两种或多种这样的部分,等。

[0108] 本文引用的所有出版物、专利和专利申请,无论是在上文还是在下文中,都以引用的方式全文纳入本说明书。

[0109] 方法

[0110] 本发明提供用于偶联两个或多个部分的方法。本发明涉及 (1) 两个或多个各自包含可杂交区域的连接物和 (2) 两个或多个能够形成共价键的基团的用途。所述连接物中的可杂交区域杂交并连接所述部分。然后所述连接的部分通过所述基团之间形成的共价键偶联。

[0111] 本发明的方法共价地偶联两个或多个部分。所述方法包括提供第一部分,其上共价连接有:(i) 至少一个第一连接物,其包含第一可杂交区域;和(ii) 至少一个能够形成共

价键的第一基团。所述方法还包括提供第二部分,其上共价连接有:(i)至少一个第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域;和(ii)至少一个能够与所述第一基团形成共价键的第二基团。在允许所述第一和第二可杂交区域杂交并连接所述部分的条件下使所述第一和第二部分接触。所述杂交反应的高亲和性使得所述部分可以在低浓度下连接。这将在下文中进一步详细讨论。然后将所述连接的部分暴露于允许所述第一和第二基团之间形成共价键的条件。一个或多个共价键的形成可加强所述两个部分之间的连接,并且形成比仅用杂交实现时的更稳定的复合物。

[0112] 本发明中使用的所有基团优选地不与本身反应。这会在偶联两个或多个不同部分时提高所述方法的产率,因为它会防止一个部分与其自身的另一个形式偶联,而不与其应该偶联的不同部分偶联。

[0113] 在一个实施方案中,所述至少一个第一基团直接共价连接至所述第一部分,或者通过单独的连接物(例如,与所述第一连接物不同的连接物)连接至所述第一部分,并且/或者所述至少一个第二基团直接共价连接至第二部分,或者通过单独的连接物(即,与所述第二连接物不同的连接物)连接至第二部分。所述单独的连接物可以是任意类型的连接物。合适的连接物是本领域中已知的,并且可以是下文更详细描述的任何连接物。

[0114] 在另一个更优选的实施方案中,所述至少一个第一基团通过所述第一连接物共价连接至所述第一部分,并且/或者至少一个所述第二基团通过所述第二连接物共价连接至第二部分。在该实施方案中,所述第一部分优选地共价连接有至少一个第一连接物,其包含第一可杂交区域和能够形成共价键的第一基团;并且/或者第二部分优选地共价连接有第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域和能够与第一基团形成共价键的第二基团。在最优选的实施方案中,所述第一和第二基团分别通过所述第一和第二连接物共价连接至所述第一和第二部分。因此,在一个优选的实施方案中,所述方法包括(a)提供第一部分,其上共价连接有至少一个第一连接物,其包含第一可杂交区域和能够形成共价键的第一基团;(b)提供第二部分,其上共价连接有至少一个第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域和能够与所述第一基团形成共价键的第二基团;(c)在允许所述第一和第二可杂交区域杂交并连接所述部分的条件下使所述第一和第二部分接触;并且(d)将所述连接的部分暴露于允许所述第一和第二基团之间形成共价键的条件。

[0115] 每个部分都共价连接有至少一个连接物。用于将连接物共价连接至部分的方法是本领域中公知的。例如,所述连接物可通过二硫键连接至蛋白质部分的半胱氨酸残基。这将在下文更详细地描述。每个部分均可共价连接有超过一个连接物,例如2、3、4或更多个连接物。在这样的实施方案中,被偶联的所述两个或多个部分一般具有相同数目的连接物共价连接于其上。然而,被偶联的所述两个或多个部分可各自具有不同数目的连接物共价连接其上。这还将在下文更详细地描述。

[0116] 如果每个部分使用超过一个连接物,则优选每个部分上的连接物形成连接在一起的连接物对。换言之,各个部分上的两个连接物的可杂交区域优选地能够彼此特异性地杂交。实现上述内容的最简单的方法是将两个连接物的可杂交区域设计为使得它们仅彼此杂交而不与本发明使用的任何其他可杂交区域杂交。如果所述第一和第二基团连接至所述第一和第二连接物,则还可使用下文描述的链接化学(click chemistry)形成连接物对。

[0117] 在一个优选的实施方案中,所述第一部分共价连接有 2、3 或 4 个第一连接物,而第二部分共价连接有相应数目的第二连接物,并且所述连接物形成一个或多个第一和第二连接物对,在所述连接物对中,每对中的可杂交区域相互特异性地杂交,但不与其他连接物对的任何可杂交区域杂交。在该实施方案中,所述部分会形成稳定的复合物,因为它们会由 2、3 或 4 个共价键连接。该实施方案还可被用于以需要的取向固定所述两个部分。

[0118] 在一个更优选的实施方案中,所述部分由两个连接物连接,所述方法包括 (a) 提供第一部分,其上共价连接有:(i) 第一连接物,其包含第一可杂交区域,(ii) 能够形成共价键的第一基团,(iii) 包含第三可杂交区域的第三连接物,和 (iv) 能够形成共价键的第三基团;(b) 提供第二部分,其上共价连接有:(i) 第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域,(ii) 能够与所述第一基团形成共价键的第二基团,(iii) 包含能够与所述第三可杂交区域杂交的第四可杂交区域的第四连接物,和 (iv) 能够与所述第三基团形成共价键的第四基团;(c) 在允许所述第一和第二可杂交区域杂交并允许所述第三和第四可杂交区域杂交并且连接所述部分的条件下使所述第一和第二部分接触;并且 (d) 将所述连接的部分暴露于以下条件:即允许所述第一和第二基团之间和所述第三和第四基团之间形成共价键。在一个更优选的实施方案中,所述方法包括 (a) 提供第一部分,其共价连接有 (1) 第一连接物,其包含第一可杂交区域和能够形成共价键的第一基团,和 (2) 第三连接物,其包含第三可杂交区域和能够形成共价键的第三基团;(b) 提供第二部分,其共价连接有 (1) 第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域和能够与所述第一基团形成共价键的第二基团,和 (2) 第四连接物,其包含能够与所述第三可杂交区域杂交的第四可杂交区域和能够与所述第三基团形成共价键的第四基团;(c) 在允许所述第一和第二可杂交区域杂交并允许所述第三和第四可杂交区域杂交并且连接所述部分的条件下使所述第一和第二部分接触;并且 (d) 将所述连接的部分暴露于以下条件:即允许所述第一和第二基团之间和第三和第四基团之间形成共价键。

[0119] 优选地,所述第一可杂交区域不与所述第三和第四可杂交区域杂交,所述第二可杂交区域不与所述第三和第四可杂交区域杂交,所述第三可杂交区域不与所述第一和第二可杂交区域杂交,并且所述第四可杂交区域不与所述第一和第二可杂交区域杂交。

[0120] 使用每个部分上不同数目的连接物来偶联所述两个或多个部分也有益处。例如,可控制所述部分的相对取向。在一个实施方案中,所述部分之一有 2、3 或 4 个连接物,而另一个部分仅有 1 个连接物。所述单个或单独的连接物当然需要包含能够与相应部分上多个连接物的可杂交区域杂交的 2、3 或 4 个可杂交区域。与多个连接物的杂交可引起所述单个或单独的连接物弯曲或折曲。在一个优选的实施方案中,所述第一部分共价连接有 (i) 包含第一可杂交区域的第一连接物,(ii) 能够形成共价键的第一基团,(iii) 包含第三可杂交区域的第三连接物,和 (iv) 能够形成共价键的第三基团;并且所述第二部分共价连接有 (i) 第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域和能够与所述第三可杂交区域杂交的第四可杂交区域,(ii) 能够与所述第一基团形成共价键的第二基团,和 (iii) 能够与所述第三基团形成共价键的第四基团。在一个更优选的实施方案中,所述第一部分共价连接有第一连接物和第三连接物,而所述第二部分共价连接有第二连接物,其中所述第一连接物包含第一可杂交区域和能够形成共价键的第一基团,其中所述第三连接物包含第三可杂交区域和能够形成共价键的第三基团,并且其中所述第二连接物包

含能够与所述第一可杂交区域杂交的第二可杂交区域、能够与所述第三可杂交区域杂交的第四可杂交区域和能够与所述第一和第三基团形成共价键的第二和第四基团。

[0121] 每个部分均共价连接有至少一个基团。用于将基团共价连接至所述部分或所述连接物的方法是本领域中公知的。这将在下文中更详细地描述。每个部分均共价连接有超过一个基团,例如 2、3、4 或更多个基团。这使得可形成包含所述部分的牢固复合物。所述连接物和基团还可被设计来以所需的所述部分相互之间的空间取向固定所述部分。还可使用多个基团来限制所述部分相互之间的活动性。例如,可使用多个基团来限制一个部分旋转的自由度或其移动远离另一个部分的能力。

[0122] 根据本发明,可偶联超过两个部分,例如 3、4、5、6 或更多个部分。这使得可以形成包含数个部分的复合物。它还使得可以形成包含具有特定排布方式的数个部分的复合物。例如,可将所述连接物的可杂交区域或基团设计为使它们以特定排布方式偶联所述部分。例如,如果使用三个部分(A、B 或 C),则可以各个特定的排布方式来偶联它们,也就是 A-B-C、A-C-B、B-A-C、B-C-A、C-A-B 和 C-B-A。任意这些排布的末端部分还可被偶联起来形成部分的“环”。

[0123] 本领域技术人员能够设计出合适的部分和连接物组合来使所述部分可以连接起来。例如,如果如上所述使用三个部分(A、B 和 C),则可通过提供共价连接有两个连接物的 B 来形成 A-B-C,所述连接物之一允许通过 A 上的单个连接物进行偶联,而所述连接物的另一个允许通过 B 上的单个连接物进行偶联。或者,每个部分都可以共价连接有两个连接物。如果 A 上的一个连接物允许通过 B 上的一个连接物进行偶联,B 上的另一个连接物允许通过 C 上的一个连接物进行偶联,C 上的另一个连接物允许通过 A 上的余下连接物进行偶联,则所述三个部分可以偶联成“环”。还可如上文所述使用多个连接物来偶联每个部分。

[0124] 在另一个实施方案中,可将所述连接物设计为可自组装为共价连接有所述部分的结构。核酸自组装是本领域中公知的。合成的寡核苷酸之间的相互作用以及因此它们通过杂交形成的结构可通过使用简单规则(Watson-Crick 碱基配对,加上规定结构基序(例如连接点、G-四链体、i-基序等)设计的规则)设计它们的核苷酸序列来加以控制。例如,Seeman 等人已制备了多种三维 DNA 晶体(例如 NC Seeman et al., J. Theor. Biol. 99, 237 (1982))。所述自组装结构可通过所述连接物上的所述基团之间形成的共价键来偶联。

[0125] 在一个优选的实施方案中,所述方法包括偶联 3、4 或 5 个部分,并且其中每个部分共价连接有 (i) 至少一个连接物,其包含能够与不同部分上的连接物的可杂交区域杂交的可杂交区域,和 (ii) 至少一个能够与不同部分上的基团形成共价键的共价基团。在一个更优选的实施方案中,所述方法包括连接 3、4 或 5 个部分,并且其中每个部分共价连接有至少一个连接物,其包含能够与另一个连接物的可杂交区域杂交的可杂交区域和能够与另一个连接物上的基团形成共价键的共价基团。

[0126] 在一个甚至更优选的实施方案中,所述方法包括 a) 提供第一部分,其共价连接有:(i) 包含第一可杂交区域的第一连接物,和 (ii) 能够形成共价键的第一基团;b) 提供第二部分,其共价连接有:(i) 第二连接物,其包含能够与所述第一可杂交区域杂交的第二可杂交区域,和 (ii) 能够与所述第一基团形成共价键的第二基团,(iii) 包含第三可杂交区域的第三连接物,和 (iv) 能够形成共价键的第三基团;c) 提供第三部分,其共价连接有:(i) 第四连接物,其包含能够与所述第三可杂交区域杂交的第四可杂交区域,(ii) 能够与

所述第三基团形成共价键的第四基团；d) 在允许所述可杂交区域杂交并连接所述部分的条件下使所述第一、第二和第三部分接触；并且 e) 将所述连接的部分暴露于允许所述基团之间形成共价键的条件。

[0127] 在一个甚至更优选的实施方案中，所述方法包括 (a) 提供第一部分，其共价连接有第一连接物，所述第一连接物包含第一可杂交区域和能够形成共价键的第一基团；(b) 提供第二部分，其共价连接有 (1) 第二连接物，其包含能够与所述第一可杂交区域杂交的第二可杂交区域和能够与所述第一基团形成共价键的第二基团，和 (2) 第三连接物，其包含第三可杂交区域和能够形成共价键的第三基团；(c) 提供第三部分，其共价连接有第四连接物，所述第四连接物包含能够与所述第三可杂交区域杂交的第四可杂交区域和能够与所述第三基团形成共价键的第四基团；(d) 在允许所述可杂交区域杂交并且连接所述部分的条件下使所述第一、第二和第三部分接触；并且 (e) 将所述连接的部分暴露于允许所述基团之间形成共价键的条件。

[0128] 部分

[0129] 所述两个或多个部分可以是可使用连接物共价偶联的任意物质或表面。优选的部分包括但不局限于多肽、蛋白质、核酸、化学基团、染料、亲和标签、聚合物、小珠、抗体、载体、固定的金属亲和基质、谷胱甘肽琼脂糖凝胶、糊精琼脂糖凝胶、亲和基质、IgG 琼脂糖凝胶、活化的巯基-琼脂糖凝胶、金属薄膜、纳米颗粒、玻璃、经处理的玻璃、塑料、树脂表面、固态纳米孔、传感器表面和单分子探测器。

[0130] 合适的固态孔包括但不局限于氮化硅孔、二氧化硅孔和石墨烯 (graphene) 孔。美国专利 No. 6, 464, 842、WO 03/003446、WO2005/061373、美国专利 No. 7, 258, 838、美国专利 No. 7, 466, 069、美国专利 No. 7, 468, 271 和美国专利 No. 7, 253, 434 中描述了其他合适的固态孔和生产它们的方法。

[0131] 所述两个或多个部分可以是相同的或者是不同的。

[0132] 优选地，所述第一部分是蛋白质，所述第二部分是蛋白质、核酸、报告基团、亲和标签或载体。所述两个或多个部分可以任意构型连接。例如，如果它们是蛋白质，它们可通过其末端（即氨基或羧基末端）氨基酸连接。

[0133] 更优选地，所述第一部分是核酸结合蛋白，第二部分是跨膜蛋白质孔。合适的构型包括但不局限于，所述核酸结合蛋白的氨基末端连接至所述跨膜蛋白质孔的羧基末端，反之亦然。或者，所述两个组分可通过其序列中的氨基酸连接。例如，所述核酸结合蛋白可连接至所述部分的环区域中的一个或多个氨基酸。在一个优选的实施方案中，所述核酸结合蛋白的末端氨基酸连接至所述部分的环区域中的一个或多个氨基酸。下文将描述末端氨基酸和环区域。

[0134] 如果所述第一部分是核酸结合蛋白，所述第二部分是跨膜蛋白质孔，则所述第一和第二基团优选地与它们各自的部分相距 0.05nm 至 10nm，更优选地相距 0.05nm 至 5nm，最优选地与它们各自的部分相距约 0.1nm 或 3 个核苷酸。如果所述第一部分是核酸结合蛋白，所述第二部分是跨膜蛋白质孔，则所述偶联的第一和第二部分之间的距离优选地是 0.1nm 至 100nm，更优选地是 0.1nm 至 25nm，更优选地是 0.1nm 至 5nm，最优选地是 0.5nm。为了达到部分之间约 0.5nm 的距离，需要使所述连接物以“头对头”方式排布。如果所述连接物以“头对尾”的排布方式使用，则每个连接物中需要最少有 5 或 6 个核苷酸以提供充分

的杂交。更短的连接物即使能杂交,也不能够充分地杂交,因此只有使用不同的构型才能达到更短的部分-部分距离。

[0135] 在一个优选的实施方案中,选择所述共价连接的位点以使得当所述部分为孔时,所述核酸结合蛋白会以这样的方式处理靶核酸序列,即使得所述靶序列中的一部分核苷酸与所述孔相互作用。然后根据核苷酸影响所述相互作用期间流经所述孔的电流的不同方式,区分所述核苷酸。

[0136] 有很多可将所述孔用于对核酸分子测序的方法。一种方法涉及核酸外切酶例如脱氧核糖核酸酶的使用。在该方法中,所述核酸外切酶被用于从靶核酸链中连续分离核苷酸。然后所述孔可按照所述核苷酸释放的顺序对其进行检测和分辨,从而读出原始链的序列。对于这样的实施方案,优选将所述核酸外切酶连接至孔亚基,这样使得从所述靶核酸上释放的一部分核苷酸就能够进入包含所述亚基的孔的桶状体或通道,并且与其相互作用。优选将所述核酸外切酶连接至所述亚基中紧密靠近形成所述孔的桶状体或通道开口的亚基部分的位置。所述核酸外切酶更优选地被连接至所述亚基,这样使得其核苷酸出口轨道位置的取向就朝向形成所述孔的开口部分的亚基部分。

[0137] 对核酸测序的另一种方法包括将所述靶核酸推过或者拉过所述孔的酶与外加电势的结合使用。在该方法中,离子电流随着所述靶链中的核苷酸通过所述孔而波动。所述电流波动指示所述链的序列。对于这样的实施方案,优选将所述酶连接至孔亚基,这样使得它能够所述靶核酸推过或拉过包含所述亚基的孔的桶状体或通道而不影响通过所述孔的离子电流。优选地将所述酶连接至所述亚基中紧密靠近形成所述孔的桶状体或通道的开口部分的亚基部分的位置。更优选将所述酶连接至所述亚基使得其活性位点的取向朝向形成所述孔的开口部分的亚基部分。

[0138] 对核酸链测序的第三种方法是检测紧密靠近孔检测器的聚合酶的副产物。在该方法中,将核苷磷酸(核苷酸)标记,这样使得向所述核苷酸链加入聚合酶后,磷酸标记的物质(phosphate labeled species)就会释放出来,并且所述磷酸标记的物质会被所述孔检测到。所述磷酸物质包含针对每种核苷酸特异性的标记。随着核苷酸被顺序地加入所述核酸链,可检测到碱基添加的副产物。所述磷酸标记的物质被检测到的顺序可被用于确定所述核酸链的序列。

[0139] 所述核酸结合蛋白优选地连接至孔或其亚基中形成所述孔顺侧(cis side)部分的部分。在电生理学中,所述顺侧按照惯例是接地的一侧。如果溶血素孔被正确地插入电生理装置中,则所述Cap区域位于顺侧。已知在正电势下,核苷酸将从用于随机传感的孔的顺侧迁移至反侧(trans side)。将所述核酸结合蛋白放置在孔的顺侧会使其能够处理所述靶核酸,这样使得所述序列中的一部分核苷酸能够进入所述孔的桶状体或通道并与其相互作用。优选地,所述序列中至少20%、至少40%、至少50%、至少80%或至少90%的核苷酸会进入所述孔的桶状体或通道并与其相互作用。

[0140] 优选地选择共价连接的位点和方法,从而使得限制所述核酸结合蛋白的移动性。这有助于确保所述蛋白质以这样的方式处理所述靶核酸序列,即使所述靶序列中的一部分核苷酸与所述孔相互作用。例如,限制核酸结合蛋白移动的能力的含义是它的活性位点的取向可永久地朝向形成所述孔的桶状体或通道的开口部分的亚基部分。可通过增加所述蛋白连接至所述部分的位点的数量并且/或者使用特异性的连接物来限制所述核酸结合蛋

白的移动性。

[0141] 生物学活性部分一般以保持它们的生物活性的方式偶联。可使用本领域中公知的方法来确定本文描述的任何部分的生物学活性。如果核酸结合蛋白和跨膜蛋白质孔被偶联在一起,则这两种蛋白质优选地分别保持它们的结合特性和孔特性。

[0142] 如果一个或多个所述部分是蛋白质,可如要求美国申请 No. 61/148,726 的优先权并与该申请同时提交的共同未决国际申请 [J A Kemp & Co Ref :N.106566A ;Oxford Nanopore Technologies Ref :ONL IP 007] 中所描述的,可从所述蛋白质中除去至少一个天然可及的半胱氨酸残基。但是,所述蛋白质可包含一个或多个可及的半胱氨酸残基,并且可通过这些残基连接至其他部分。存在有限数量的可及半胱氨酸残基使得可以控制部分之间的偶联。

[0143] 半胱氨酸残基的可离子化侧链是用于进行加成反应的强效亲核体,因此广泛地被选用于生物缀合 (bioconjugation) 技术。但是,对于这些技术有效使用的一个常见障碍是存在于野生型蛋白质中的天然半胱氨酸残基。对一个或多个所述天然半胱氨酸残基的修饰可导致不明确的酶活性和不受控制的连接。

[0144] 本发明可涉及蛋白质部分的天然半胱氨酸残基的定点诱变。可保留一个或多个所述可及的天然半胱氨酸,除去其他可及的天然半胱氨酸。或者,可除去全部可及的半胱氨酸残基,并向所述蛋白质引入一个或多个半胱氨酸残基。特定的可及半胱氨酸残基的存在便于所述结合蛋白与另一部分的连接,所述另一部分例如另一蛋白质、固体载体或化学试剂。

[0145] 天然半胱氨酸残基的移除以及单个或多个半胱氨酸的靶向添加改进了之前的方法,因为它赋予控制偶联的能力,同时还使得所述蛋白质彼此之间可能的相互作用最小化,所述相互作用通过不需要的表面巯基形成二聚体、三聚体等。如果所述蛋白质仅在其结构中的预定位点处含有单个半胱氨酸残基,那么该残基的巯基基团能够被用于与其他巯基基团形成直接的二硫键或者可通过反应活性连接物介导偶联。这些被靶向的半胱氨酸残基可确保只有需要数量的酶分子与需要数量的靶分子交联,同时还赋予一定程度的有利构象,这样使得例如活性位点就能够最优地取向。

[0146] 存在于野生型蛋白质中的天然半胱氨酸残基使得不能容易地使用包含巯基反应活性基团的特定连接物,例如马来酰亚胺、碘乙酰胺或邻-吡啶基-二硫化物 (ortho-pyridyl disulphide (OPSS))。与任意在结构和催化方面重要的半胱氨酸残基反应的可能性能够影响所述蛋白的活性,这对于用于单分子应用很重要。优选使用用于生物缀合的连接物以确保一定的空间分离,这样使得蛋白活性不因与另一个部分 (例如蛋白质或表面) 紧密接近而受到影响。

[0147] 核酸结合蛋白

[0148] 一种所述部分优选地是核酸结合蛋白。这类蛋白的实例包括但不局限于核酸处理酶,例如核酸酶、聚合酶、拓扑异构酶、连接酶和解旋酶;和非催化的结合蛋白,例如由 SCOP (蛋白质的结构分类) 分类在核酸结合蛋白超家族 (50249) 下的那些。如上文所述和如要求美国申请 No. 61/148,726 的优先权并与本申请同时提交的共同未决国际申请 [J A Kemp & Co Ref :N.106566A ;Oxford Nanopore Technologies Ref :ONL IP 007] 中所述,将所述核酸处理酶修饰以去除并且 / 或者代替半胱氨酸残基。优选地从所述核酸结合蛋白中去除至少一个天然可及的半胱氨酸残基。然而,所述核酸结合蛋白包含一个或多个可及的

半胱氨酸残基,并且通过这些残基连接至另一部分。有限数量可及半胱氨酸残基的存在使得可以控制与另一部分的连接。

[0149] 核酸是包含两个或多个核苷酸的大分子。由蛋白质结合的核酸可包含任意核苷酸的任意组合。所述核苷酸可以是天然存在的或者是人造的。所述核苷酸可被氧化或者甲基化。核苷酸一般包含核碱基、糖和至少一个磷酸基团。所述核碱基一般是杂环的。核碱基包括但不局限于嘌呤和嘧啶,更具体地是腺嘌呤、鸟嘌呤、胸腺嘧啶、尿嘧啶和胞嘧啶。所述糖一般是戊糖。核苷酸糖包括但不局限于核糖和脱氧核糖。所述核苷酸一般是核糖核苷酸或脱氧核糖核苷酸。所述核苷酸一般包含单磷酸、二磷酸或三磷酸。磷酸可连接在核苷酸的 5' 或 3' 侧。

[0150] 核苷酸包括但不限于单磷酸腺苷 (AMP)、二磷酸腺苷 (ADP)、三磷酸腺苷 (ATP)、单磷酸鸟苷 (GMP)、二磷酸鸟苷 (GDP)、三磷酸鸟苷 (GTP)、单磷酸胸苷 (TMP)、二磷酸胸苷 (TDP)、三磷酸胸苷 (TTP)、单磷酸尿苷 (UMP)、二磷酸尿苷 (UDP)、三磷酸尿苷 (UTP)、单磷酸胞苷 (CMP)、二磷酸胞苷 (CDP)、三磷酸胞苷 (CTP)、环单磷酸腺苷 (cAMP)、环单磷酸鸟苷 (cGMP)、单磷酸脱氧腺苷 (dAMP)、二磷酸脱氧腺苷 (dADP)、三磷酸脱氧腺苷 (dATP)、单磷酸脱氧鸟苷 (dGMP)、二磷酸脱氧鸟苷 (dGDP)、三磷酸脱氧鸟苷 (dGTP)、单磷酸脱氧胸苷 (dTMP)、二磷酸脱氧胸苷 (dTDP)、三磷酸脱氧胸苷 (dTTP)、单磷酸脱氧尿苷 (dUMP)、二磷酸脱氧尿苷 (dUDP)、三磷酸脱氧尿苷 (dUTP)、单磷酸脱氧胞苷 (dCMP)、二磷酸脱氧胞苷 (dCDP) 和三磷酸脱氧胞苷 (dCTP)。所述核苷酸优选选自 AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP 或 dCMP。

[0151] 所述核酸可以是脱氧核糖核酸 (DNA) 或核糖核酸 (RNA)。所述核酸可以是本领域中已知的任意合成的核酸,例如肽核酸 (PNA)、甘油核酸 (GNA)、苏糖核酸 (TNA)、锁核酸 (LNA) 或具有核苷酸侧链的其他合成的聚合物。由蛋白质结合的核酸优选地为单链的,例如 cDNA、RNA、GNA、TNA 或 LNA。由蛋白质结合的核酸优选地是双链的,例如 DNA。结合单链核酸的蛋白质可被用于对双链 DNA 测序,条件是所述双链 DNA 在其被所述蛋白质结合之前解离为单链。

[0152] 优选地,所述核酸结合蛋白的三级结构是已知的。所述结合蛋白的三维结构的知识可使得对所述蛋白质做出修饰以促进其在所述偶联结构中的功能。

[0153] 所述蛋白质可具有任意大小和任意结构。例如,所述蛋白质可以是低聚体,例如二聚体或三聚体。所述蛋白质优选地是由一个单体形成的小的球状多肽。这样的蛋白质容易处理并且干扰其他部分的可能性较小。例如,这样的蛋白质干扰孔亚基的孔形成能力的可能性较小。

[0154] 还优选地,所述蛋白质的活性位点的位置和功能是已知的。这可以防止对所述活性位点做出修饰可消除所述蛋白质活性的修饰。还允许所述蛋白质连接至其他部分,这样使得所述蛋白质以特定的方式结合所述靶核酸序列。例如,如果所述蛋白质被连接于用于测序的跨膜蛋白质孔,则它使得靶序列中的一部分核苷酸与所述孔相互作用,如下文中所描述。在这样的实施方案中,将所述蛋白质的活性位点设置在尽可能靠近所述孔通道的桶状体的开口处并且使得所述蛋白质本身不会阻断电流是有益的。关于其中蛋白质对核酸取向的方式的知识也使得可设计有效的偶联结构。

[0155] 可能必须纯化由本发明所述方法产生的所述结构。优选所述核酸结合蛋白能够承

受用于纯化所述结构的条件。

[0156] 所述另一个部分可包含孔。这样的孔可被用于对核酸测序。为了使所述靶核酸中的大部分核苷酸可通过随机传感被正确地鉴定,所述蛋白质优选地在缓冲液背景下结合所述核酸,所述缓冲液背景适于所述核苷酸的辨别。所述蛋白质优选地在远超于正常生理水平的盐浓度(例如 100mM 至 2000mM)下至少具有残余活性。更优选地,将所述蛋白质修饰以增加其在高盐浓度下的活性。还可对所述蛋白质修饰以提高其持续处理能力、稳定性和保存期限。

[0157] 可根据对来自嗜极微生物(extremophile)的核酸处理酶的特征以及改变嗜温或嗜热核酸外切酶的耐盐性、稳定性和温度依赖性的定向进化方法来确定合适的修饰,所述嗜极微生物例如嗜盐细菌、中度嗜盐细菌,嗜热生物和中度嗜热生物。

[0158] 所述酶还优选地在 10°C 至 60°C 的温度(例如室温)下保持至少部分活性。这使得可在多种温度(包括室温)下对核酸测序。

[0159] 所述核酸结合蛋白优选地是核酸处理酶。核酸处理酶是能够与核酸相互作用并且改变核酸的至少一个性质的多肽。所述酶可通过切割所述核酸以形成单个核苷酸或较短的核苷酸链(例如二核苷酸或三核苷酸)来修饰所述核酸。所述酶可通过对所述核酸取向或将其移动至特定位置来修饰所述核酸。

[0160] 所述核酸处理酶优选地来自溶核酶。所述核酸处理酶更优选地来自酶分类(EC)组 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 和 3.1.31 中任一组的成员。所述核酸处理酶更优选地基于以下酶中的任一种:

[0161] ● 3.1.11. - 产生 5' - 磷酸单酯的脱氧核糖核酸外切酶。

[0162] ○ 3.1.11.1 脱氧核糖核酸外切酶 I。

[0163] ○ 3.1.11.2 脱氧核糖核酸外切酶 III。

[0164] ○ 3.1.11.3 脱氧核糖核酸外切酶(λ 诱导的)。

[0165] ○ 3.1.11.4 脱氧核糖核酸外切酶(噬菌体 SP3 诱导的)。

[0166] ○ 3.1.11.5 脱氧核糖核酸外切酶 V。

[0167] ○ 3.1.11.6 脱氧核糖核酸外切酶 VII。

[0168] ● 3.1.13. - 产生 5' - 磷酸单酯的核糖核酸外切酶。

[0169] ○ 3.1.13.1 核糖核酸外切酶 II。

[0170] ○ 3.1.13.2 核糖核酸外切酶 H。

[0171] ○ 3.1.13.3 寡核苷酸酶。

[0172] ○ 3.1.13.4 Poly(A) - 特异的核糖核酸酶。

[0173] ○ 3.1.13.5 核糖核酸酶 D。

[0174] ● 3.1.14. - 产生 3' - 磷酸单酯的核糖核酸外切酶。

[0175] ○ 3.1.14.1 酵母核糖核酸酶。

[0176] ● 3.1.15. - 产生 5' - 磷酸单酯的可作用于核糖核酸或脱氧核糖核酸的核酸外切酶。

[0177] ○ 3.1.15.1 毒液核酸外切酶(Venom exonuclease)。

[0178] ● 3.1.16. - 产生 3' - 磷酸单酯的可作用于核糖核酸或脱氧核糖核酸的核酸外切酶。

- [0179] ○ 3.1.16.1 脾脏核酸外切酶 (Spleen exonuclease)。
- [0180] ● 3.1.21. - 产生 5' - 磷酸单酯的脱氧核糖核酸内切酶。
- [0181] ○ 3.1.21.1 脱氧核糖核酸酶 I。
- [0182] ○ 3.1.21.2 脱氧核糖核酸酶 IV (噬菌体 -T(4)- 诱导的)。
- [0183] ○ 3.1.21.3I 型位点特异的脱氧核糖核酸酶。
- [0184] ○ 3.1.21.4II 型位点特异的脱氧核糖核酸酶。
- [0185] ○ 3.1.21.5III 型位点特异的脱氧核糖核酸酶。
- [0186] ○ 3.1.21.6CC- 嗜性脱氧核糖核酸内切酶 (CC-preferring endodeoxyribonuclease)。
- [0187] ○ 3.1.21.7 脱氧核糖核酸酶 V。
- [0188] ● 3.1.22. - 不产生 5' - 磷酸单酯的脱氧核糖核酸内切酶。
- [0189] ○ 3.1.22.1 脱氧核糖核酸酶 II。
- [0190] ○ 3.1.22.2 曲霉属 (*Aspergillus*) 脱氧核糖核酸酶 K(1)。
- [0191] ○ 3.1.22.3 转向条目 :3.1.21.7。
- [0192] ○ 3.1.22.4 交换型连接脱氧核糖核酸内切酶 (Crossover junction endodeoxyribonuclease)。
- [0193] ○ 3.1.22.5 脱氧核糖核酸酶 X。
- [0194] ● 3.1.25. - 特异性用于改变的碱基的位点特异的脱氧核糖核酸内切酶。
- [0195] ○ 3.1.25.1 脱氧核糖核酸酶 (嘧啶二聚体)。
- [0196] ○ 3.1.25.2 转向条目 :4.2.99.18。
- [0197] ● 3.1.26. - 产生 5' - 磷酸单酯的核糖核酸内切酶。
- [0198] ○ 3.1.26.1 绒胞菌多聚脑磷脂核糖核酸酶 (*Physarum polycephalum* ribonuclease)。
- [0199] ○ 3.1.26.2 核糖核酸酶 α 。
- [0200] ○ 3.1.26.3 核糖核酸酶 III。
- [0201] ○ 3.1.26.4 核糖核酸酶 H。
- [0202] ○ 3.1.26.5 核糖核酸酶 P。
- [0203] ○ 3.1.26.6 核糖核酸酶 IV。
- [0204] ○ 3.1.26.7 核糖核酸酶 P4。
- [0205] ○ 3.1.26.8 核糖核酸酶 M5。
- [0206] ○ 3.1.26.9 核糖核酸酶 (poly-(U)- 特异的)。
- [0207] ○ 3.1.26.10 核糖核酸酶 IX。
- [0208] ○ 3.1.26.11 核糖核酸酶 Z。
- [0209] ● 3.1.27. - 不产生 5' - 磷酸单酯的核糖核酸内切酶。
- [0210] ○ 3.1.27.1 核糖核酸酶 T(2)。
- [0211] ○ 3.1.27.2 枯草芽孢杆菌 (*Bacillus subtilis*) 核糖核酸酶。
- [0212] ○ 3.1.27.3 核糖核酸酶 T(1)。
- [0213] ○ 3.1.27.4 核糖核酸酶 U(2)。
- [0214] ○ 3.1.27.5 胰核糖核酸酶。

- [0215] ○ 3.1.27.6 肠杆菌属核糖核酸酶 (Enterobacter ribonuclease)。
- [0216] ○ 3.1.27.7 核糖核酸酶 F。
- [0217] ○ 3.1.27.8 核糖核酸酶 V。
- [0218] ○ 3.1.27.9 tRNA- 内含子核酸内切酶。
- [0219] ○ 3.1.27.10 rRNA 核酸内切酶。
- [0220] ● 3.1.30.- 产生 5' - 磷酸单酯的可作用于核糖核酸或脱氧核糖核酸的核糖核酸内切酶。
- [0221] ○ 3.1.30.1 曲霉属核酸酶 S(1)。
- [0222] ○ 3.1.30.2 粘质沙雷氏菌核酸酶 (*Serratia marcescens* nuclease)。
- [0223] ● 3.1.31.- 产生 3' - 磷酸单酯的可作用于核糖或脱氧核糖的核糖核酸内切酶。
- [0224] 3.1.31.1 微球菌核酸酶 (*Micrococcal* nuclease)。
- [0225] 所述酶最优选地衍生自核酸外切酶,例如脱氧核糖核酸酶,其切断核酸以形成个体核苷酸。脱氧核糖核酸外切酶的优点在于它们可作用于单链和双链 DNA 并且可从 5'-3' 或 3'-5' 方向水解碱基。
- [0226] 个体核苷酸是单个核苷酸。个体核苷酸是不与另一个核苷酸或核酸以任意键(例如磷酸二酯键)结合的核苷酸。磷酸二酯键包含结合在另一个核苷酸糖基团上的核苷酸磷酸基团之一。个体核苷酸一般是不与另一核酸序列以任意方式结合的核苷酸,所述另一核酸序列有至少 5 个、至少 10 个、至少 20 个、至少 50 个、至少 100 个、至少 200 个、至少 500 个、至少 1000 个或至少 5000 个核苷酸。
- [0227] 用于本发明的优选的酶包括来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6), 来自大肠杆菌的核酸外切酶 III (SEQ ID NO :52)、来自嗜热栖热菌的 RecJ (SEQ ID NO :54) 和噬菌体 λ 核酸外切酶 (SEQ ID NO :56) 以及它们的变体。SEQ ID NO :56 的三个相同亚基相互作用以形成三聚体核酸外切酶。所述酶最优选地以来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 为基础。
- [0228] 所述核酸处理酶优选地衍生自包含 SEQ ID NO :6、52、54 和 56 所示序列中任一个或其变体的核酸外切酶。换言之,所述酶优选地在如上文所述修饰其半胱氨酸前包含 SEQ ID NO :6、52、54 和 56 所示序列中任一个或其变体。
- [0229] SEQ ID NO :6、52、54 或 56 的变体是具有与 SEQ ID NO :6、52、54 或 56 的氨基酸序列不同的氨基酸序列但是保持核酸处理能力的酶。所述变体处理核酸的能力可使用本领域中已知的任意方法来分析。例如,可如实施例中所描述的来测定变体处理核酸的能力。如实施例中所述,所述变体还必须保持其被表达的能力。
- [0230] 所述变体可包括有助于处理所述核酸并且 / 或者促进其在高盐浓度和 / 或室温下的活性的修饰。
- [0231] 所述酶可以是由生物体(例如由大肠杆菌)表达的天然存在的变体。变体还包括由重组技术产生的非天然存在的变体。在 SEQ ID NO :6、52、54 或 56 的氨基酸序列全长中,基于氨基酸同一性,变体优选地与该序列有至少 50% 的同源性。更优选地,基于氨基酸同一性,所述变体多肽可与 SEQ ID NO :6、52、54 或 56 的氨基酸序列在全长序列上具有至少 55%、至少 60%、至少 65%、至少 70%、至少 75%、至少 80%、至少 85%、至少 90%、并且更优选至少 95%、97% 或 99% 的同源性。在 200 或更多(例如 230、250、270、280 或者更多)

个连续氨基酸的片段上有至少 80% 例（如至少 85%、90% 或 95%）的氨基酸同一性（“硬同源性 (hard homology)”）。

[0232] 可以用本领域中的标准方法来确定同源性。例如，UWCGC 软件包提供的 BESTFIT 程序可以用于计算同源性（例如使用它的默认设置）(Devereux et al(1984)Nucleic Acids Research 12,387-395)。例如，如 Altschul S.F. (1993)J Mol Evol 36 :290-300 ; Altschul, S, F et al(1990)JMol Biol 215 :403-10 中所述，可以使用 PILEUP 和 BLAST 算法计算同源性或者对序列进行比对（例如鉴定等价残基或对应序列（一般使用它们的默认设置））。

[0233] 进行 BLAST 分析的软件公众可以从美国国家生物技术信息中心 (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/>) 公开获得。所述算法涉及到首先鉴别高打分序列对 (HSP)，所述鉴别步骤通过如下方式实现：鉴别查询序列中长度为 W 的短字，所述短字就是在与数据库序列中相同长度的字比较时，能匹配或者满足某些正值阈值分值 T 的字。T 是指邻近字分值阈值 (Altschul et al, 见上文)。将这些初始邻近字匹配作为种子启动检索，以发现包含它们的 HSP。在累积比对分值能够增加期间，在沿每条序列的两个方向上进行字匹配延伸。字匹配在每个方向上的延伸停止的条件是：所述累积比对分值从其所达到的最大值下降 X 量；由于一个或多个负打分残基比对的累积，所述累积分值降到 0 或小于 0；或者达到任何一个序列的端点。所述 BLAST 算法的参数 W、T 和 X 决定了比对的敏感度和速度。所述 BLAST 程序使用的默认字段 (W) 为 11, BLOSUM62 打分矩阵（见 Henikoff and Henikoff(1992)Proc. Natl. Acad. Sci. USA 89 :10915-10919）比对 (B) 为 50, 期望值 (E) 为 10, M = 5, N = 4, 并且是进行双链比较。

[0234] 所述 BLAST 算法可进行两个序列之间的相似性的统计分析；参见例如 Karlin and Altschul(1993)Proc. Natl. Acad. Sci. USA 90 :5873-5787。所述 BLAST 算法提供的一种相似性的量度是最小和概率 (P(N))，所述最小和概率表示两个氨基酸序列之间随机出现匹配的概率。例如，如果一个序列与另一个序列比较中的最小和概率小于约 1, 优选小于约 0.1, 更优选小于约 0.01, 并且最优选小于约 0.001, 那么所述第一序列被认为与所述第二序列相似。

[0235] 除了上文所描述的那些之外，还可以对 SEQ ID NO :6、52、54 或 56 的氨基酸序列做出氨基酸置换，例如，最多达 1、2、3、4、5、10、20 或 30 个置换。例如，可根据下表 1 做出保守性置换。

[0236] 表 1- 保守性置换

[0237] 在第二列同一格，优选在第三列同一行中的氨基酸可相互置换。

[0238]

非芳香族	非-极性	GAP
		ILV
	极性- 不荷电	CSTM
		NQ
	极性 - 荷电	DE
		HKR
芳香族		HFVY

[0239] 还可从上文描述的多肽中额外缺失 SEQ ID NO :6、52、54 或 56 的氨基酸序列的一个或多个氨基酸残基。可以缺失最长达 1、2、3、4、5、10、20 或 30 个残基, 或者更多个残基。

[0240] 变体可以是 SEQ ID NO :6、52、54 或 56 的片段。这类片段保留了核酸处理活性。片段长度可以为至少 50、100、200 或 250 个氨基酸。片段优选包含 SEQ ID NO :6、52、54 或 56 的核酸处理结构域。

[0241] 可将一个或多个氨基酸替代地或额外地添加至上述的多肽上。在 SEQ ID NO :6、52、54 或 56 或其变体或片段的氨基酸序列的氨基末端或羧基末端, 可提供一段延长片段。所述延长片段可以非常短, 例如长度为 1-10 个氨基酸。或者, 所述延长片段可以较长, 例如最长达 50 或 100 个氨基酸。可将载体蛋白融合于亚基或变体。

[0242] 如上文所述, SEQ ID NO :6、52、54 或 56 的变体是这样的蛋白质, 即该蛋白质具有不同于 SEQ ID NO :6、52、54 或 56 的氨基酸序列且保留了其处理核酸的能力。变体一般包含 SEQ ID NO :6、52、54 或 56 中负责处理核酸的区域。在上文序列列表说明中描述了 SEQ ID NO :6、52、54 和 56 的催化结构域。SEQ ID NO :6、52、54 或 56 的变体优选地包含相关的催化结构域。SEQ ID NO :6、52、54 或 56 的变体一般在相关催化结构域外边包含一个或多个修饰, 例如置换、添加或缺失。SEQ ID NO :6 的特定变体将在下文进一步描述。

[0243] 所述变体可通过例如添加组氨酸或天冬氨酸残基以帮助它们的鉴定或纯化, 或者通过添加信号序列以促进它们从其中所述多肽天然不含这类序列的细胞中分泌而被修饰。

[0244] 能够将所述靶核酸序列拉过或推过所述孔的优选酶包括聚合酶、核酸酶、解旋酶和拓扑异构酶, 例如促旋酶。所述核酸处理酶可以衍生自任意这些类型的酶。所述聚合酶优选地是酶分类 (EC) 组 2.7.7.6、2.7.7.7、2.7.7.19、2.7.7.48 和 2.7.7.49 中任一组的成员。所述聚合酶优选地是 DNA 依赖的 DNA 聚合酶、RNA 依赖的 DNA 聚合酶、DNA 依赖的 RNA 聚合酶或 RNA 依赖的 RNA 聚合酶。所述解旋酶优选地是基于酶分类 (EC) 组 3.6.1.- 和 2.7.7.- 中任一组的成员。所述解旋酶优选地是 ATP 依赖的 DNA 解旋酶 (EC 组 3.6.1.8)、ATP- 依赖的 RNA 解旋酶 (EC 组 3.6.1.8) 或 ATP- 依赖的 RNA 解旋酶。所述拓扑异构酶优选地是酶分类 (EC) 组 5.99.1.2 和 5.99.1.3 中任一组的成员。

[0245] 所述核酸结合蛋白可由显示标签标记。所述显示标签可以是使得所述孔可被检测的任意合适标签。合适的标签包括但不限于荧光分子、反射性同位素 (例如 ^{125}I 、 ^{35}S 、 ^{14}C)、酶、抗体、抗原、多核苷酸和配体 (例如生物素)。

[0246] 所述核酸结合蛋白可从产生结合蛋白的生物体中分离, 或者通过合成或通过重组

方法制造,所述产生结合蛋白的生物体例如大肠杆菌、嗜热栖热菌或噬菌体。例如,所述核酸结合蛋白可通过体外翻译和转录来合成。所述核酸结合蛋白的氨基酸序列可被修饰以引入非天然存在的氨基酸或增加所述蛋白质的稳定性。当所述核酸结合蛋白是通过合成方法来生产时,这样的氨基酸可在生产期间被引入。所述核酸结合蛋白还可在合成或重组生产之后被改变。

[0247] 还可使用 D-氨基酸来生产所述核酸结合蛋白。例如,所述核酸结合蛋白可包含 L-氨基酸和 D-氨基酸的混合物。这类蛋白或肽的生产是本领域中常规的。

[0248] 所述核酸结合蛋白还可以包含其他非特异性化学修饰,条件是它们不干扰其处理核酸或连接至其他部分的能力。许多非特异性侧链修饰是本领域中已知的,并且可对所述侧链进行这些修饰。例如,这类修饰包括对氨基酸的还原性烷基化(按照如下方式进行:首先与醛反应,然后以 NaBH₄ 进行还原)、以甲基乙酰亚胺酯(methylacetimidate)进行脒基化(amidination)或者以乙酸酐进行酰化。可在所述核酸结合蛋白表达之后或者所述核酸结合蛋白已被用于本发明的方法之后做出对所述核酸结合蛋白的修饰。

[0249] 可以使用本领域中已知的标准方法产生所述核酸结合蛋白。可以使用本领域中的标准方法分离并复制编码核酸结合蛋白的多核苷酸序列。这类序列在下文有更详细的叙述。可以使用本领域中的标准技术将编码核酸结合蛋白的多核苷酸序列在细菌宿主细胞中表达。所述核酸结合蛋白可以通过在细胞中以重组表达载体原位表达所述多肽来产生。所述表达载体任选地携带诱导型启动子以控制所述多肽的表达。

[0250] 核酸结合蛋白可通过如下方式大规模产生:从产生孔的生物体或者在如下文所述的重组表达后通过任一种蛋白质液相色谱体系进行纯化。一般的蛋白质液相色谱体系包括 FPLC、AKTA 系统、Bio-Cad 系统、Bio-Rad BioLogic 系统和 Gilson HPLC 系统。

[0251] 所述核酸处理酶优选地衍生自包含 SEQ ID NO:6 所示的序列或其变体的核酸外切酶。换言之,所述酶在其半胱氨酸残基被如上文所述修饰之前,优选地包含 SEQ ID NO:6 所示的序列或其变体。SEQ ID NO:6 的变体上文描述过。

[0252] SEQ ID NO:6 在位置 51、98、144、306 和 330 有 5 个天然半胱氨酸残基。这些位置的所有 5 个半胱氨酸残基都是可及的。SEQ ID NO:6 的一种变体优选地在其按照本发明被修饰之前包含所有 5 个这些残基。

[0253] 在一个实施方案中,从 SEQ ID NO:6 中去除位置 51、98、144、306 和 330 处的所有 5 个半胱氨酸,并且引入一个或多个非天然半胱氨酸残基。

[0254] 在另一个实施方案中,将位置 51、98、144、306 和 330 处的所有 5 个半胱氨酸除保留 1 个或多个外,都从 SEQ ID NO:6 中去除。衍生自 SEQ ID NO:6 的酶在诱变后可保留位置 51、98、144、306 和 330 处的所述 5 个半胱氨酸残基中的一个或者多个的任意结合。优选的结合包括但不局限于 144 和 330。在一个优选的实施方案中,仅保留 SEQ ID NO:6 的位置 144 或 330 处的天然半胱氨酸残基。

[0255] SEQ ID NO:6 中所述 5 个半胱氨酸残基中的一个或多个优选地由丙氨酸、丝氨酸、甲硫氨酸或苏氨酸置换。SEQ ID NO:6 中位置 51 处的半胱氨酸残基更优选地由丙氨酸置换。SEQ ID NO:6 中位置 98 处的半胱氨酸残基更优选地由丝氨酸或苏氨酸置换。SEQ ID NO:6 中位置 144 处的半胱氨酸残基更优选地由甲硫氨酸或苏氨酸置换。SEQ ID NO:6 中位置 306 处的半胱氨酸残基更优选地由丝氨酸或苏氨酸置换。SEQ ID NO:6 中位置 330 处的

半胱氨酸残基更优选地由苏氨酸置换。在最优选的实施方案中,SEQ ID NO :6 中位置 51 处的半胱氨酸残基由丙氨酸置换,SEQ ID NO :6 中位置 98 处的半胱氨酸残基由苏氨酸置换,SEQ ID NO :6 中位置 144 处的半胱氨酸残基由苏氨酸置换,SEQ ID NO :6 中位置 306 处的半胱氨酸残基由苏氨酸置换,而 SEQ ID NO :6 中位置 330 处的半胱氨酸残基由苏氨酸置换。

[0256] 所述核酸处理酶优选地包含 SEQ ID NO :8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 和 50 中任意一个显示的序列或其变体。SEQ ID NO :8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 和 50 的变体是具有与 SEQ ID NO :8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 和 50 的氨基酸序列不同的氨基酸序列并且保持其处理核酸的能力的蛋白质。各个变体与 SEQ ID NO :8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 或 50 不同的程度可与上文描述的 SEQ ID NO :6 的变体与 SEQ ID NO :6 不同的程度相同。但是,SEQ ID NO :8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 和 50 的变体必须包含与野生型酶比较做出的改变。例如,SEQ ID NO :8、10 和 12 的变体必须在位置 98、144、306 和 330 处包含与 SEQ ID NO :8、10 和 12 相同的残基。SEQ ID NO :14 和 16 的变体必须在位置 42、51、98、144、306 和 330 处包含与 SEQ ID NO :14 和 16 相同的残基。SEQ ID NO :18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 和 50 的变体必须在位置 51、98、144、184、306 和 330 处包含与 SEQ ID NO :18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 和 50 相同的残基。

[0257] 所述核酸处理酶一般通过一个或多个它的表面可及残基来连接至所述部分,所述表面可及残基具有优选取向的侧链。SEQ ID NO :6 中的以下残基不应该被用作连接点,因为它们被埋没并且不能触及溶剂,或者与连接物的连接将导致所述蛋白质整体结构的严重破坏: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 和 267。SEQ ID NO :6 中的以下残基接近 DNA 结合沟,尽管它们是表面可及的,但也不是连接连接物的良好位点: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 和 371。SEQ ID NO :6 中的所有剩余残基都可能被用作连接点。

[0258] SEQ ID NO :6 中的以下残基是优选的连接点,因为它们提供侧链,暴露在外并且是溶剂可及的,并且可预测对所述侧链的突变几乎不会导致对整个蛋白质结构的破坏: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 和 467。以下的残基是最优选的连接点: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 和 223。任意所述优选的连接残基都可通过置换来修饰。所述优选的连接残基中的一个或者多个优选地是由半胱氨酸置换。SEQ ID NO :6 或其变体中的优选置换包括但不限于 M184C、A83C、V42C、V94C、S90C、Y188C、Y188C、A219C、A219C、M218C 和 M218C。如下文所更详细描述,所述酶可通过超过 1 个残基连接至所述部

分。

[0259] 所述核酸结合蛋白可在一个或多个位置（例如 1、2、3 或 4 个位置）处连接于另一个部分。所述核酸结合蛋白优选地在 1 个或 2 个位置处连接。在从所述核酸结合蛋白中去除天然半胱氨酸残基后，可在所述蛋白质的特定位置处纳入一个或多个半胱氨酸残基用于连接。在两个位置处的连接可降低所述复合物的柔性，并且能够将所述核酸结合蛋白以所选择的特定取向固定在所述部分上。

[0260] 在一个优选的实施方案中，所述方法包括将 (1) 具有 M184C 和 V94C 置换的来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体偶联至 (2) 具有 N17C 和 E287C 置换的 α -HL (SEQ ID NO :2) 或其变体。将 SEQ ID NO :6 或其变体中的 M184C 位置偶联至 SEQ ID NO :2 或其变体中的 N17C 位置。将 SEQ ID NO :6 或其变体中的 V94C 位置由不同的连接物偶联至 SEQ ID NO :2 或其变体中的 E287C 位置。

[0261] 下文表 2 显示其他优选的第一和第二连接点的位置。

[0262]

	第一个 A	第一个 B (连接至第一个 A)	第二个 A	第二个 B (连接至第二个 A)
	SEQ ID NO: 6 或其变体	SEQ ID NO: 2 或其变体	SEQ ID NO: 6 或其变体	SEQ ID NO: 2 或其变体
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

[0263] 跨膜蛋白质孔

[0264] 一种所述部分优选地是跨膜蛋白质孔。跨膜蛋白质孔是多肽或多肽的集合体，其允许由外加电势驱动的离子从膜的一侧流到所述膜的另一侧。所述孔优选地允许核苷酸顺着外加电势从膜的一侧流到所述膜的另一侧。所述孔优选地允许核酸（例如 DNA 或 RNA）被推过或拉过所述孔。

[0265] 所述孔可以是单体或低聚体。所述孔优选地由一些重复亚基例如 6、7 或 8 个亚基构成。所述孔更优选地是七聚体孔。所述孔一般包含离子可流过的桶状体或通道。所述孔的亚基一般围绕中心轴并且为跨膜 β 桶状体或通道或者跨膜 α -螺旋束状体或通道提供链。

[0266] 所述孔的所述桶状体或通道一般包含便于与核苷酸或核酸的相互作用的氨基酸。这些氨基酸的位置优选地靠近所述桶状体或通道的缩窄处。所述孔一般包含一个或多个带正电的氨基酸，例如精氨酸、赖氨酸或组氨酸。这些氨基酸一般便于所述孔和核苷酸或核酸之间的相互作用。衔接体可促进所述核苷酸检测。这将在下文更详细描述。

[0267] 可用于本发明的孔可以是 β -桶状体孔、 α -螺旋束状体孔或固态孔。 β -桶状体孔包含由 β -折叠形成的桶状体或通道。合适的 β -桶状体孔包括但不限于 β -毒素, 例如 α -溶血素、炭疽毒素和杀白细胞素; 和细菌的外膜蛋白/通道蛋白 (porin), 例如包皮垢分支杆菌 (*Mycobacterium smegmatis*) 通道蛋白 A (MspA)、外膜通道蛋白 F (OmpF)、外膜通道蛋白 G (OmpG)、外膜磷脂酶 A 和奈瑟氏菌 (*Neisseria*) 自转运脂蛋白 (Nalp)。 α -螺旋束状体孔包含由 α -螺旋形成的桶状体或通道。合适的 α -螺旋束状体孔包括但不限于内膜蛋白和 α -外膜蛋白, 例如 WZA。

[0268] 所述部分可以是孔本身。或者, 如果所述孔是低聚体, 则所述部分可以是孔亚基。所述部分可如上所述被分离、基本上分离、纯化或基本上纯化。

[0269] 所述孔或亚基优选地衍生自 α -溶血素 (α -HL)。野生型 α -HL 孔是由 7 个相同的单体或亚基形成 (即, 它是七聚体)。 α -溶血素的一种野生型单体或亚基的序列在 SEQ ID NO :2 中显示。所述部分优选地包含 SEQ ID NO :2 所示的序列或其变体。SEQ ID NO :2 的氨基酸 1、7 至 21、31 至 34、45 至 51、63 至 66、72、92 至 97、104 至 111、124 至 136、149 至 153、160 至 164、173 至 206、210 至 213、217、218、223 至 228、236 至 242、262 至 265、272 至 274、287 至 290 和 294 形成环区域。SEQ ID NO :2 的残基 113 和 147 形成 α -HL 的桶状体或通道的缩窄处的一部分。所述核酸结合蛋白优选地连接 SEQ ID NO :2 的氨基酸 8、9、17、18、19、44、45、50 和 51 中的一个或者多个。

[0270] SEQ ID NO :2 的一个变体是具有不同于 SEQ ID NO :2 的氨基酸序列并且保持其孔形成能力的亚基。可使用本领域中已知的任意方法来测定变体形成孔的能力。例如, 可将所述变体与其他合适的亚基一起插入膜中并且测定其低聚化以形成孔的能力。用于将亚基插入膜 (例如脂双层) 中的方法是本领域中已知的。例如, 可将亚基以纯化的形式悬浮于含有脂双层的溶液中, 这样使得它会扩散到所述脂双层中, 并且通过结合至所述脂双层并且组装为有功能状态而插入。或者, 可使用 M. A. Holden, H. Bayley. *J. Am. Chem. Soc.* 2005, 127, 6502-6503 和国际申请 PCT/GB2006/001057 (公开号 WO 2006/100484) 所述的“拾取放置 (pick and place)”方法, 将亚基直接插入至所述膜中。

[0271] 所述变体可包括便于与所述核酸结合蛋白共价连接或相互作用的修饰。所述变体优选地包含便于连接至所述核酸结合蛋白的一个或多个反应活性半胱氨酸残基。例如, 所述变体可在 SEQ ID NO :2 的位置 8、9、17、18、19、44、45、50、51、237、239 和 287 中的一处或多处和/或氨基或羧基末端包含半胱氨酸。优选的变体包含用半胱氨酸对 SEQ ID NO :2 的位置 8、9、17、237、239 和 287 处残基进行的置换 (K8C、T9C、N17C、K237C、S239C 或 E287C)。

[0272] 所述变体还可以包含便于与核苷酸的任意相互作用或者便于分子衔接体取向的修饰, 这将在下文描述。所述变体还可包含便于分子衔接体的共价连接的修饰。

[0273] 具体地, 所述变体优选地在 SEQ ID NO :2 的位置 139 处具有谷氨酰胺。所述变体优选地在 SEQ ID NO :2 的位置 119、121 或 135 处具有半胱氨酸。SEQ ID NO :4 示出了 SEQ ID NO :2 的序列, 但其在位置 135 处具有半胱氨酸 (L135C), 并在位置 139 处有谷氨酰胺 (N139Q)。SEQ ID NO :4 或其变体可被用于形成本发明的孔。所述变体在 SEQ ID NO :2 的位置 113 处可具有精氨酸。

[0274] 所述变体可以是天然存在的变体, 其由例如葡萄球菌属 (*Staphylococcus*) 细菌的生物体天然地表达, 或者由例如大肠杆菌 (*Escherichia coli*) 的细菌以重组方式表达。

变体还包括通过重组技术产生的非天然存在的变体。在 SEQ ID NO :2 或 4 的氨基酸序列全长中,基于氨基酸同一性,变体优选地与所述序列具有至少 50% 的同源性。基于氨基酸同一性,所述变体多肽更优选地在整个序列上可以与 SEQ ID NO :2 或 4 的氨基酸序列具有至少 55%、至少 60%、至少 65%、至少 70%、至少 75%、至少 80%、至少 85%、至少 90%、更优选至少 95%、97% 或 99% 的同源性。在 200 个或更多个(例如 230、250、270 或 280 个或更多个)连续氨基酸的片段中,可以有至少 80%,例如至少 85%、90% 或 95% 的氨基酸同一性(“硬同源性”)。可如上所述测量同源性。

[0275] 除上文所描述的那些之外,可对 SEQ ID NO :2 或 4 的氨基酸序列进行氨基酸置换,例如最多达 1、2、3、4、5、10、20 或 30 个置换。例如,可根据上表 1 进行保守性置换。

[0276] 可额外地将 SEQ ID NO :2 的氨基酸序列中的一个或多个氨基酸残基从上述多肽中缺失。最多达 1、2、3、4、5、10、20 或 30 或更多个残基可被缺失。

[0277] 变体可以是 SEQ ID NO :2 或 4 的片段。这样的片段保持孔形成活性。片段长度可为至少 50、100、200 或 250 个氨基酸。片段优选地包含 SEQ ID NO :2 或 4 的孔形成结构域。片段一般包括 SEQ ID NO :2 或 4 的残基 119、121、135、113 和 139。

[0278] 替代地或额外地,一个或多个氨基酸可被加入上述多肽中。可在 SEQ ID NO :2 或 4 或者其变体或片段的氨基酸序列的氨基末端或羧基末端提供延长片段。所述延长片段可以很短,例如长度为 1 至 10 个氨基酸。或者,所述延长片段可以较长,例如最长达 50 或 100 个氨基酸。载体蛋白可被融合至孔或变体中。

[0279] 如上所述,SEQ ID NO :2 或 4 的变体是具有不同于 SEQ ID NO :2 或 4 的氨基酸序列并且保持了其孔形成能力的亚基。变体一般包含 SEQ ID NO :2 或 4 中负责孔形成的区域。含有 β -桶状体的 α -HL 的孔形成能力由每个亚基中的 β -折叠提供。SEQ ID NO :2 或 4 的变体一般包含 SEQ ID NO :2 或 4 中形成 β -折叠的区域。上文描述了 SEQ ID NO :2 或 4 中形成 β -折叠的氨基酸。可对 SEQ ID NO :2 或 4 中形成 β -折叠的区域进行一种或多种修饰,只要所得到的变体保持其形成孔的能力。上文描述了可对 SEQ ID NO :2 或 4 的 β -折叠区域进行的具体修饰。

[0280] SEQ ID NO :2 或 4 的变体优选包含在其 α -螺旋和 / 或环区域内的一种或多种修饰,如置换、添加或缺失。上文描述了形成 α -螺旋和环的氨基酸。

[0281] 可通过以下方式对所述变体进行修饰,例如可通过添加组氨酸或天冬氨酸残基以帮助它的鉴定或纯化,或者通过添加信号序列以促进它们从其中所述多肽天然不含这类序列的细胞中分泌。

[0282] 变体还可包含上文针对所述核酸结合蛋白描述的任意非特异修饰。可如上文所述制造亚基或孔。

[0283] 连接物和基团

[0284] 所述两个或多个部分共价连接有上文定义的至少一个连接物和至少一个基团。用于将连接物和基团连接至部分的方法是本领域中公知的。优选地将所述连接物和基团以化学方式融合至所述部分。合适的方法包括但不限于伯胺 / 羧酸偶联、巯基之间的反应、非天然氨基酸之间的反应和半胱氨酸连接。

[0285] 本发明的方法还可包括将所述至少一个连接物和 / 或至少一个基团连接至每个部分的步骤。

[0286] 优选地使用连接物在一个或多个位置（例如 2 个、3 个或 4 个位置）将所述第一部分连接至所述第二部分。每个连接物与其相关部分的连接位置可被设计为限制所述部分的相对活动性。所述部分被连接的位置的数目对最终的复合物的刚性有影响。

[0287] 所述连接物和 / 或基团可被连接至所述部分上的一个或多个反应活性基团，例如半胱氨酸残基、反应活性赖氨酸残基或非天然氨基酸。所述连接物和 / 或基团优选地连接至所述部分中的一个或多个可及的半胱氨酸残基。所述连接物和 / 或基团一般通过由一个半胱氨酸残基和另一个半胱氨酸残基形成的二硫键连接至所述部分。

[0288] 所述连接物是高亲和力杂交连接物。它们包含至少一个可杂交区域，并且优选地包含至少一个能够形成共价键的基团。特别优选的连接物是在低浓度下（例如在纳摩尔亲和力和力下）有效的连接物。连接物的长度、柔性和亲水性一般被设计为使得它们不会破坏待连接的部份的功能。使用杂交连接物的一个益处是所述部分的不需要的同二聚体形成被最小化。

[0289] 可杂交区域是能够与另一个连接物中的可杂交区域进行杂交的区域。所述可杂交区域优选地包含核酸。所述核酸杂交连接物可包含任意上文描述的核酸。例如，它们可包含脱氧核糖核酸 (DNA)、核糖核酸 (RNA) 和本领域中已知的任意合成的核酸，例如肽核酸 (PNA)、甘油核酸 (GNA)、苏糖核酸 (TNA)、锁核酸 (LNA) 或其他有核苷酸侧链的合成的聚合物。允许两个连接物杂交的核酸可以相同也可以不相同。最优选地，所述连接物包含 DNA。一般地，当各个连接物均包含 DNA 时，至少一个连接物必须包含单链 DNA。优选地，两个连接物都包含单链 DNA。

[0290] 杂交可以在两个或多个连接物之间发生，优选地在两个或三个连接物之间发生。在使用三个连接物的情况下，它们形成三聚体。所述连接物可以多种构象杂交，例如以“头对头”构象或“头对尾”的构象杂交。在所述“头对头”构象中，所述连接物排列为使得所述部分在所述杂交的连接物的相同末端相邻。在所述“头对尾”构象中，所述连接物排列为使得所述部分处于所述杂交的连接物的相对末端。所述连接物可具有任意合适的三维构象。一般地，连接物是直链的，但它也可能是三角形或形成其他几何构象。

[0291] 优选的核酸杂交连接物对应于来自 SEQ ID NO :57 的 5' 末端的前 15、25 或 35 个核苷酸。所述连接物优选地还在 3' 末端具有 TT 以提供额外的柔性。在所述 3' 末端，所述连接物具有使得所述连接物可连接至所述部分的基团，例如马来酰亚胺或巯基。马来酰亚胺或巯基修饰的寡核苷酸可购买获得，例如从 ATDBio 购买。更优选的连接物在 SEQ ID NO : 58、59 和 60 中显示。互补连接物在 SEQ ID NO :61、62 和 63 中显示。SEQ ID NO :58、59 或 60 可被连接至所述部分之一，而互补连接物（分别为 SEQ ID NO :61、62 或 63）被连接至所述另一个部分。然后可通过杂交所述连接物来将所述部分连接在一起。

[0292] 杂交的稳定性依赖于所述杂交连接物的解链温度。根据应用和所需要的稳定性，可通过改变所述连接物的序列（例如，将所述连接物变为更富含 GC 将增加它们的解链温度，并进而增强稳定性）、所述连接物的长度（即增加它们的长度会增强稳定性）或反应条件（例如，增加它们的浓度将增强稳定性）来对此进行优化。

[0293] 为了获得杂交的最大稳定性，需要具有高解链温度的长杂交连接物，例如长度超过 15 个核苷酸的连接物，尤其是长度为 15 至 45 个核苷酸的连接物，例如长度为 15、20、25、30、35、40 或 45 个核苷酸的连接物。但是，使用长连接物会增加所述部分之间的距离。这可

能是不利的,因为所述部分之间的相互作用会遭到破坏,或者因为需要近距离以使得一个部分检测已从另一个部分上释放的底物。增加的距离可能是有利的,因为它可以阻止聚集或者静电相互作用,并且可允许弯曲。可通过改变所述核酸连接的取向来克服所述距离增加的不利之处。所述连接物最优选地包含具有长度为 6 至 15 个核苷酸的核酸,例如长度为 6、8 或 10 个核苷酸的核酸。

[0294] 所述第一和第二可杂交区域优选地在 1pM 至 1mM 的浓度下具有 1fM 至 1uM 的亲合力。所述第一和第二可杂交区域更优选地在 1pM 至 1uM 的浓度下具有 1fM 至 10nM 的亲合力。所述第一和第二可杂交区域最优选地在 100pM 至 10nM 的浓度下具有 1pM 至 100pM 的亲合力。

[0295] SEQ ID NO 64 示出了优选的序列。该连接物的 3' 端可连接至所述表面的半胱氨酸残基上。所述连接物优选地在所述 3' 端还具有 TTTTT 以提供额外的柔性。然后,SEQ ID NO 65 的 5' 端能够连接在所述核酸结合蛋白上。在该实例中,SEQ ID NO 64 的 3' 末端与 SEQ ID NO 65 的 5' 末端处的一段序列互补。

[0296] 一旦所述连接物相互杂交,它们在一定条件下(例如,高温或低盐条件)解链(分开),除非所述两个连接物之间有固定的键。为了形成固定的键,对所述部分进行修饰从而使得一旦它们已杂交即相互反应。每个部分共价连接有能够与另一个部分中的基团形成共价键的基团。如上所述,所述基团优选地通过所述杂交连接物与所述部分连接。一对部分可由一个或多个共价键(例如 1 个、2 个或 3 个共价键)连接。

[0297] 一般地,所述键是所述两个部分之间的单个二硫键。所述部分或杂交连接物还可被修饰以在一个或多个(例如 2 个)位置处纳入巯基基团。根据应用和偏好,巯基基团可以被纳入所述连接物的内部或者末端。

[0298] 还可修饰所述部分或杂交连接物以包括一个或多个(例如 2 个)碘乙酰胺基团。此外,所述碘乙酰胺基团可被纳入所述连接物的内部或者末端。具有一个或多个碘乙酰胺基团的部分或杂交连接物可与所述互补部分或杂交连接物上的巯基共价连接。

[0299] 还可用烯炔基团修饰所述部分或杂交连接物。如果被纳入所述杂交连接物,则所述烯炔基团可以在内部或者末端。可对在优选位置中的一个或多个烯基基团进行烯炔复分解(olefin metathesis)以使所述部分或杂交连接物上的烯基之间形成共价键。

[0300] 如果需要,可在所述部分或杂交连接物与所述反应活性基团之间加上小的连接物(例如巯基基团、碘乙酰胺基团和烯炔基团)以得到使所述部分或杂交连接物之间形成有效的共价键所必需的合适距离。

[0301] 在一个优选的实施方案中,可使用链接化学(click chemistry)来形成所述部分或杂交连接物之间的共价键。链接化学是由 Kolb 等在 2001 年首先提出的术语,其用于描述在小规模和大规模应用中均可靠地起作用的一大组强有力、有选择性并且是模块化的构件块(Kolb HC, Finn, MG, Sharpless KB, Click chemistry: diverse chemical function from a few good reactions, Angew. Chem. Int. Ed. 40 (2001) 2004-2021)。他们为链接化学定义了如下的严格标准:“所述反应必须是模块化的,范围广泛,产量很高,仅产生可通过非色谱方法除去的无害的副产物,并且是立体定向的(但不必是对映选择性的)。所需要的方法特征包括:反应条件简单(理想地,所述方法应该对氧气和水不敏感)、容易得到起始材料和试剂、不使用溶剂或使用温和的(例如水)或者容易除去的溶剂,以及产物分离简单。

如需纯化,则纯化必须是非色谱方法,例如结晶或蒸馏,并且所述产物在生理条件下必需是稳定的”。

[0302] 链接化学的合适实例包括但不限于以下:

[0303] (a) 1,3 偶极环加成反应的无铜变化方案,其中叠氮化物与炔在应变下反应,例如在环辛烷环中;

[0304] (b) 一个连接物上的氧亲核试剂与另一个连接物上的环氧化物或氮丙啶反应活性部分的反应;和

[0305] (c) Staudinger 连接,其中所述炔部分可由芳基膦代替,导致与叠氮化物的特异反应,得到酰胺键。

[0306] 所述链接化学反应优选地是 Cu(I) 催化的炔和叠氮化物之间的 1,3 偶极环加成反应。在一个优选的实施方案中,所述第一个基团是叠氮基团,所述第二基团是炔基基团。已经合成了在优选位置上纳入叠氮化物和炔基基团的核酸碱基(例如 Kocalka P, El-Sagheer AH, Brown T, Rapid and efficient DNA strand cross-linking by click chemistry, *Chembiochem.* 2008. 9(8):1280-5)。炔基基团可从 Berry Associates (Michigan, USA) 市购,叠氮基团可由 ATDBio 合成。

[0307] 如果所述连接物核酸区域中的核苷酸被修饰以包括可形成共价键的基团,则所述经修饰的核苷酸优选地相互偏离一个核苷酸以实现所述连接。这依照 Tom Brown (Kocalka et al. (2008) *ChemBiochem* 9:1280-1285) 公开的研究进行。

[0308] 在一个优选的实施方案中,可将单个叠氮基团 (SEQ ID NO 66) 或多个例如两个 (SEQ ID NO 67) 叠氮基团在 15 个碱基的脱氧核糖核酸序列的特定位点处纳入尿嘧啶碱基。然后可用这些在 5' 末端使用巯基基团的叠氮化物杂交连接物 (SEQ ID NO 66 和 67) 来修饰部分上的半胱氨酸残基。还可在与 SEQ ID NO 66 和 67 互补的序列(分别为 SEQ ID NO 68 和 69) 中优选位置处的尿嘧啶中纳入炔基基团。这些序列可被连接至一个所述部分中的半胱氨酸上。通过使用 DNA 杂交,然后进行所述叠氮化物和炔基基团之间的“链接化学”,杂交连接物能被共价交联。如果所述部分之间的距离已经通过改变所述杂交连接物的长度作了调整,则所述叠氮化物和炔修饰的碱基的位置就需要相应地改变。

[0309] 在一个优选的实施方案中,可将其中两个尿嘧啶碱基被叠氮基团修饰的 6mer (SEQ ID NO 70)、8mer (SEQ ID NO 71) 或 10mer (SEQ ID NO 72) DNA 与所述部分的半胱氨酸连接。可将其中两个尿嘧啶碱基被炔基团修饰的 6mer (SEQ ID NO 73)、8mer (SEQ ID NO 74) 或 10mer (SEQ ID NO 75) DNA 的互补序列连接至例如 DNA 结合蛋白的部分的半胱氨酸。与使用本发明其他优选的杂交连接物(例如, SEQ ID NO 67 和 69) 相比,使用这些杂交连接物将使得所述部分相互之间更接近。将叠氮基团和炔基团纳入 DNA 的尿嘧啶碱基单元中的方法已经被 ATDBio 开发。

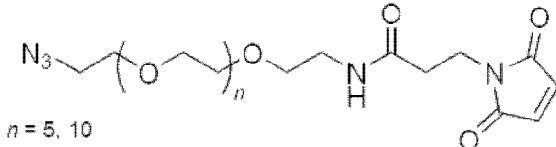
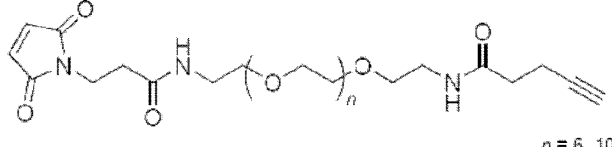
[0310] 可用于本发明的其他优选基团在下表 3 中显示。

[0311] 表 3- 能够形成共价键的一些优选基团。

[0312]

名称	与以下物质反应	结构
1,4-二[3-(2-吡啶基二硫)丙酰氨基]丁烷	巯基	
1,11-二-马来酰亚胺三甘醇	巯基	
3,3'-二硫二丙酸二(N-羟基琥珀酰亚胺酯)	伯胺	
乙二醇-二(琥珀酸 N-羟基琥珀酰亚胺酯)	伯胺	
4,4'-二异硫氰酸根合芪-2,2'-二磺酸二钠盐	伯胺	
二[2-(4-叠氮水杨酰氨基)乙基]二硫化物	光活化的, 非特异的	
3-(2-吡啶基二硫)丙酸 N-羟基琥珀酰亚胺酯	巯基, 伯胺	
4-马来酰亚胺丁酸 N-羟基琥珀酰亚胺酯	巯基, 伯胺	
碘乙酸 N-羟基琥珀酰亚胺酯	巯基, 伯胺	
S-乙酰基硫代乙醇酸 N-羟基琥珀酰亚胺酯	巯基, 伯胺	

[0313]

叠氮-PEG-马来酰亚胺	巯基, 炔	
炔-PEG-马来酰亚胺	巯基, 叠氮化物	

[0314] 条件

[0315] 首先将所述部分在允许所述第一和第二可杂交区域杂交并连接所述部分的条件下接触。允许杂交的条件是本领域中公知的(例如, Sambrook et al., 2001, Molecular Cloning: a laboratory manual, 3rd edition, Cold Spring Harbour Laboratory Press; 和 Current Protocols in Molecular Biology, Chapter 2, Ausubel et al., Eds., Greene Publishing and Wiley-Interscience, New York (1995))。可在低严格性条件下进行杂交, 例如如下条件: 存在 37°C 的 30 至 35% 甲酰胺、1M NaCl 和 1% SDS (十二烷基硫酸钠) 的缓冲溶液, 之后在 50°C 的 1X (0.1650M Na+) 至 2X (0.33M Na+) SSC (标准柠檬酸钠) 中洗涤。可在中度严格性条件下进行杂交, 例如如下条件: 存在 37°C 的 40 至 45% 甲酰胺、1M NaCl 和 1% SDS 的缓冲溶液, 之后在 55°C 的 0.5X (0.0825M Na+) 至 1X (0.1650M Na+) SSC 中洗涤。可在高严格性条件下进行杂交, 例如如下条件: 存在 37°C 的 50% 甲酰胺、1M NaCl 和 1% SDS 的缓冲溶液, 之后在 60°C 的 0.1X (0.0165M Na+) SSC 中洗涤。优选的条件是在实施例 1 中所述描述的那些条件。

[0316] 然后将所述连接的部分暴露于允许所述第一和第二基团形成共价键的条件。这些条件当然依赖于所使用的基团。所述反应优选地在小于 pH 7.0 的 pH 下进行。优选的条件是实施例 1 中所描述的那些。

[0317] 如果一个或多个所述部分是酶, 例如核酸结合蛋白, 则所述酶的活性优选地在所述偶联和连接反应中被抑制。依赖镁的酶(例如核酸外切酶 I) 可被 EDTA 或被 CaCl₂ 抑制 (Lehman (1960) J. Biol. Chem. 235:1479-1487 和内部数据)。

[0318] 部分的优选结合

[0319] 在一个优选的实施方案中, 所述第一部分是蛋白质, 所述第二部分是报告基团。合适的报告基团包括但不限于, 光学或荧光标记物、立体化学标签 (steric tag)、质量增加物、电荷改变物、自旋标记物、放射性标签、磁性基团 (magnetic group) 和它们的任一组合。

[0320] 在另一个优选的实施方案中, 所述第一部分是蛋白质而第二部分是亲和标签。亲和标签的实例是本领域中公知的。合适的亲和标签包括但不限于 His-NTA、FLAG-抗体-M2、myc、STrEP-链亲和素、NANO-链亲和素和生物素-抗生物素蛋白。

[0321] 在另一个优选的实施方案中, 所述第一部分是蛋白质而第二部分是载体。载体的实例是本领域中公知的。合适的载体包括但不限于, 所述载体可以是但不限于金属表面、聚合物小珠、磁珠、玻璃、改性的玻璃和硅胶。

[0322] 在一个优选的实施方案中, 所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体, 其具有至少一个天然可及半胱氨酸残基, 所述第二部分包含 α-HL (SEQ

ID NO :2) 或其变体。在一个更优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,全部天然可及的半胱氨酸残基都被去除,所述第二部分包含 α -HL (SEQ ID NO :2) 或其变体。在一个更优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,仅在位置 144 和 / 或 330 处具有天然半胱氨酸残基,所述第二部分包含 α -HL (SEQ ID NO :2) 或其变体。在一个更优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,所有天然半胱氨酸残基都被去除并且具有 A83C 置换,所述第二部分包含具有 E287C 的 α -HL (SEQ ID NO :2) 或其变体。

[0323] 在一个优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,其中至少一个天然可及的半胱氨酸残基被去除,并且所述第二部分包含具有 L135C 和 N139Q 的 α -HL (SEQ ID NO :2) 或其变体 (例如, SEQ ID NO :4)。在一个更优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,其全部天然可及的半胱氨酸残基都被去除,所述第二部分包含具有 L135C 和 N139Q 的 α -HL 或其变体 (例如, SEQ ID NO :4)。在一个更优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,其仅在位置 144 和 / 或 330 处具有天然半胱氨酸残基,所述第二部分包含具有 L135C 和 N139Q 的 α -HL 或其变体 (例如, SEQ ID NO :4)。在一个更优选的实施方案中,所述第一部分包含来自大肠杆菌的核酸外切酶 I (SEQ ID NO :6) 或其变体,其全部天然可及的半胱氨酸都被去除并具有 A83C 置换,所述第二部分包含具有 L135C、N139Q 和 E287C 的 α -HL 或其变体。

[0324] 产品

[0325] 本发明还提供使用本发明的方法偶联的第一和第二部分,和使用上文描述的第一和第二连接物偶联至第二部分的第一部分。本发明还提供如上所述的第一和第二连接物对。所述连接物优选地包含能够形成共价键的基团。上文描述的述及本发明方法的任意实施方案可同等地适用于本发明的产品。

[0326] 以下实施例举例说明本发明:

[0327] 1. 实施例 1

[0328] 1.1 材料和方法

[0329] 在该实例中,在铜催化的炔叠氮环加成反应 (CuAAC) 中,叠氮官能团和炔官能团在铜 (I) 催化剂的存在下相互反应形成三唑环。

[0330] 从标准实验室供应商 (例如 Sigma 和 Fisher) 获得最高等级的硫酸铜 (II)、抗坏血酸钠、MOPS 游离酸、磷酸钠、氯化钠、氯化钙和 EDTA。铜 (I) 结合配体 (图 11 中显示)。

[0331] 具有巯基末端基团或者具有内部叠氮或炔基基团的寡核苷酸来自 AtdBio (Southampton, UK); 未修饰的寡核苷酸来自 AtdBio 或 IDTDNA (Leuven, 比利时)。所述寡核苷酸的序列显示在 SEQ ID NO :64 至 78 中。在所有互补寡核苷酸的情况下,所述修饰的碱基相互之间偏离一个碱基以实现所述连接。

[0332] 1.2 杂交

[0333] 当杂交寡核苷酸时,通过在 4°C 下将所述寡核苷酸在缓冲液中孵育至少 1 小时来进行所述杂交。对于长 15 个碱基的寡核苷酸,该缓冲液可以是 100mM MOPS (pH 6.8) 或 100mM 磷酸钠 (pH 7.0)。对于比 15 个碱基的寡核苷酸短的寡聚体,包含 1000mM NaCl。当加入酶

抑制剂 (10mM EDTA 或 20mM CaCl₂) 时,这就在 100mM 磷酸盐 (pH6.8) (对于 EDTA) 或 100mM MOPS (pH6.8) (对于 EDTA 或 CaCl₂) 中。

[0334] 1.3 连接反应

[0335] 在 4°C 下过夜杂交 (对于 15bp 和更长的双链体) 或孵育 1 小时 (对于 6-10bp 的双链体) 之后进行所述连接反应。所述连接反应是通过首先预混合铜 (II) 离子、抗坏血酸盐还原剂和铜 (I) 结合配体来实现。将所述三个组分在适合的浓度下混合以达到最终的条件。例如,如果最终需要的条件是 100 μl 反应体积中有 1mM 铜、2mM 抗坏血酸盐、1mM 配体,则以指定的顺序混合等体积的 100mM 硫酸铜 (溶于水)、200mM 抗坏血酸钠 (溶于 pH < 7.0 的缓冲液) 和 100mM 配体 (溶于水),然后将 3 μl 的所得混合物加至所述 100 μl 反应体系中。在该项加入之后,在 4°C 下和黑暗中进行所述反应至少 60 分钟。

[0336]

连接的寡聚体	双链体的长度	连接条件/ [Cu ^I], [抗坏血酸盐], [配体], 缓冲液	连接?	确认
76+77	15 bp	5 μM, 10 μM, 7 μM, 1	否	凝胶
76+77	15 bp	0.5 mM, 1 mM, 0.5 mM, 1	是	凝胶
66+68 (1 连接)	15 bp	1 mM, 2 mM, 1 mM, 2	是	凝胶, T _m
67+69 (2 连接)	15 bp	1 mM, 2 mM, 1 mM, 2	是	凝胶, T _m
70+73	6 bp	1 mM, 2 mM, 1 mM, 3	否	
71+74	8 bp	1 mM, 2 mM, 1 mM, 3	是	凝胶, T _m
72+75	10 bp	1 mM, 2 mM, 1 mM, 3	是	凝胶, T _m

[0337] 缓冲液 1 = 100mM 磷酸盐, 500mM NaCl pH7.0

[0338] 缓冲液 2 = 100mM 磷酸盐 pH6.8

[0339] 缓冲液 3 = 100mM MOPS, 1000mM NaCl, 20mM CaCl₂ pH6.8

[0340] 1.4 连接反应

[0341] 为了将寡核苷酸连接至蛋白质,必须首先将所述寡核苷酸活化。该步骤将所提供的寡核苷酸的化学特性从对所述蛋白质惰性改变为能够与所述蛋白质反应。该过程与所述连接反应没有差异。

[0342] 当需要时,将寡核苷酸通过二硫键连接至蛋白质。这是通过将 5× 过量的活化寡核苷酸与新还原的蛋白质在 37°C 下孵育 1 小时 (对于核酸外切酶) 或在室温下过夜孵育 (对于纳米孔) 实现的。在所有的情况下,包含所述叠氮碱基的寡核苷酸被连接至所述纳米孔 (通过所述单体之一上的 E287C), 包含所述炔碱基的寡核苷酸被连接至所述核酸外切酶 (通过 A83C)。通过在链亲和素亲和小珠上纯化来除去未反应的 DNA。

[0343] 1.5 蛋白质的偶联

[0344] 在设计为促进所述寡核苷酸之间杂交的条件下一起孵育纯化的经修饰蛋白质。对于 15 个碱基和更长的寡核苷酸,该缓冲液可以是 100mM MOPS (pH6.8) 或 100mM 磷酸钠

(pH7.0)。当需要酶抑制剂时,将 20mM CaCl₂ 补充至 100mM MOPS (pH6.8) 中。对于 6-10 个碱基的寡聚体,所述缓冲液是 100mM MOPS, 1000mM NaCl, 20mM CaCl₂ (pH6.8)。

[0345] 如上所述进行所述连接反应,因为这些条件已被证明与酶活性相容。

[0346]

蛋白质 1	蛋白质 2	双链体长度	连接条件/ [Cu ^I], [抗坏血酸盐], [配 体], 缓冲液	ONT 俗名
Exo-SEQ 78	Pore-SEQ 67	15 bp	1 mM, 2 mM, 1 mM, 4	PK V3
Exo-SEQ 73	Pore-SEQ 70	6 bp	1 mM, 2 mM, 1 mM, 3	TL

[0347] 缓冲液 4 = 100mM MOPS, 20mM CaCl₂ pH6.8

[0348] 2. 实施例 2

[0349] 2.1 材料和方法

[0350] 2.1.1 概述

[0351] 使用随机传感对核酸测序的一种方法涉及使用连接至 α -HL 孔的核酸外切酶。这可以通过在所述两个蛋白质之间引入化学键来进行。所述化学键将所述核酸外切酶上的一个半胱氨酸残基连接至所述溶血素的一个半胱氨酸上,连接方式为直接的(二硫键)或通过所述两个半胱氨酸残基之间的连接物分子。

[0352] 为了以受控并且定向的方法将所述化学键引入,需要用其他残基来代替存在于核酸外切酶 I 中的 5 个天然存在的半胱氨酸残基。然后将作出另外的突变以引入所述半胱氨酸残基连接点。使用了计算机建模来鉴定合适的连接位置(参见下文)。

[0353] 2.1.2 对 Exo I CYS 敲除突变建模

[0354] 已经使用了分子建模来尝试预测对大肠杆菌 Exo I 的 5 个野生型半胱氨酸残基的哪个突变与蛋白质稳定性相容。为此目的对这些突变进行了 alchemical 自由能微扰模拟(alchemical free energy perturbation simulation),使得可以预测它们的相对稳定性。这些预测被用于指导实验工作。

[0355] 将 VMD 和 NAMD 与 CHARMM27 力场结合使用以对野生型进行 alchemical 自由能微扰计算。

[0356] 确定了对大肠杆菌 Exo I 蛋白质结构(PDB 登录号:1fxX[14])的 CYS 至 XXX 突变。将所得到的 λ 与 ΔG 曲线用于确定突变时的自由能变化以及计算结果是否趋同(converged)。

[0357] 2.1.3 表达和纯化

[0358] 从 GenScript 公司得到来自大肠杆菌的核酸外切酶 I 基因 scsB 的一个密码子优化版本(SEQ ID NO:5 和 6)。通过 PCR 产物的体内同源重组的经修改形式来进行半胱氨酸残基的定点诱变(Jones, D. H. PCR mutagenesis and recombination in vivo. (1994) Genome Research, 3. pp141-148)。将含有目的基因的质粒转化进入 BL21 (DE3) pLysS (stratagene)。

用来自琼脂板的菌落接种培养基（装于 2 升带盖摇瓶中，补充有 100 μ g/ml 氨苄青霉素和 20 μ g/ml 氯霉素，含有 500ml Terrific Broth）并培养至 OD600 = 2（大约 6 个小时），之后在降低的温度（18°C）下用 0.2mM IPTG 诱导。通过在 4K 下离心收集细胞，之后将其用 bugbuster (Novagen) 和 benzonase (Merck) 裂解。通过在 10K 下离心除去细胞碎片，并过滤上清液，之后将其上样至 HisTrap 柱 (GE Healthcare)。通过凝胶过滤进一步纯化洗脱的蛋白。

[0359] 2.1.4 酶测定法

[0360] 所述核酸外切酶 I (Exo I) 测定法使用寡脱氧核苷酸 (DNA 模板)，其在位置 15 处的 dT 上标记有荧光素染料 (“Fluor”)，在位置 1 处的 dT 上标记有荧光素猝灭剂 (“BHQ1”)。所述模板开始处于猝灭状态，其中所述荧光素和猝灭剂紧密靠近。加入 Exo I 后，所述 DNA 模板被以连续的方式从 3' - 末端消化，最终将 dT-Fluor 释放进入溶液，并且因为染料扩散远离了所述猝灭剂而导致随后荧光的增强（图 6）。

[0361] 为进行所述测定法，将 DNA 模板转移至比色皿中，并将其放置在 Cary Eclipse 荧光计中以在需要的温度下平衡。同时，单独地使用水浴来平衡 Exo I，然后将其加入含有模板的比色皿中。监测随后的荧光信号随时间的增加（参见图 7 的实例数据），并推出初始的增加速率。在所述模板已被完全消化之后的终点荧光与 DNA 的初始量联系起来，以将基于荧光的速率转化为分子速率。

[0362] 2.2 结果

[0363] 2.2.1 对 Exo I CYS 敲除突变的建模

[0364] 下表总结了突变建模的结果。

[0365]

突变	ΔG (真空)	ΔG (水)	趋同(真空)	趋同(水)
C51A	+3.4	+3.2	是	是
C51R	-239.2	-251.2	否	否
C51N	-69.5	-66.8	否	否
C51D	-67.7	-82.6	否	否
C51C	+0.9	+0.9	是	是
C51Q	-49.5	-50.7	否	否
C51E	-55.2	-68.9	否	否
C51G	-0.8	-0.5	是	是
C51H	-24.9	-18.5	是	是
C51I	+8.6	+11.0	是	是
C51L	-1.8	-11.7	是	是
C51K	-37.9	-12.1	是	是
C51M	-1.8	-0.4	是	是
C51F	+15.6	+9.4	是	是
C51S	+4.6	+11.9	是	否
C51T	-11.3	-10.7	是	是
C51W	+15.6	+11.6	是	是
C51Y	-9.3	-12.3	是	是
C51V	+3.4	+4.2	是	是

[0366]

[0367]

突变	ΔG (真空)	ΔG (水)	趋同(真空)	趋同(水)
C98A	+0.4	+4.2	是	是
C98R	-203.4	-231.0	否	否
C98N	-79.3	-76.1	否	否
C98D	-111.5	-113.7	否	否
C98C	+0.0	-0.3	是	是
C98Q	-48.2	-54.3	否	否
C98E	-55.7	-99.4	否	否
C98G	+1.3	-2.0	是	是
C98H	-14.9	-16.8	是	是
C98I	+6.5	+10.0	是	是
C98L	-2.7	-11.3	是	是
C98K	+11.7	-25.5	是	是
C98M	+0.7	+2.0	是	是
C98F	+10.1	+23.9	是	是
C98S	+3.0	+9.3	是	否
C98T	-13.3	-11.8	是	是
C98W	+17.7	+23.6	是	是
C98Y	+9.9	+2.2	是	是
C98V	+2.0	-0.3	是	是

[0368]

[0369]

突变	ΔG (真空)	ΔG (水)	趋同(真空)	趋同(水)
C144A	+1.2	-0.9	是	是
C144R	-218.0	-243.8	否	否
C144N	-64.9	-73.5	否	否
C144D	-70.6	-111.9	否	否
C144C	+0.4	+0.0	是	是
C144Q	-45.9	-52.4	否	否
C144E	-60.2	-71.4	否	否
C144G	-3.2	-1.5	是	是
C144H	-14.0	-16.0	是	是
C144I	+9.9	+7.3	是	是
C144L	-12.9	-11.4	是	是
C144K	-11.9	-25.6	是	是
C144M	-2.6	-2.5	是	是
C144F	+16.0	+8.5	是	是
C144S	+7.1	+7.6	是	是
C144T	-12.9	-9.2	是	是
C144W	+14.2	+9.2	是	是
C144Y	+5.8	-7.4	是	是
C144V	+0.4	+0.5	是	是

[0370]

[0371]

突变	ΔG (真空)	ΔG (水)	趋同(真空)	趋同(水)
C306A	+3.4	+1.1	是	是
C306R	-207.6	-245.2	否	否
C306N	-67.8	-77.1	否	否
C306D	-125.2	-106.6	否	否
C306C	+2.5	+0.0	是	是
C306Q	-49.8	-51.7	否	否
C306E	-66.6	-108.9	否	否
C306G	-5.1	-3.3	是	是
C306H	-12.6	-18.0	是	否
C306I	+11.0	+7.4	是	是
C306L	-7.3	-13.5	是	是
C306K	-25.1	-21.4	是	是
C306M	+0.5	-4.4	是	是
C306F	+10.7	+2.5	是	是
C306S	+2.7	+5.8	是	是
C306T	-15.5	-12.9	否	是
C306W	+15.5	+13.6	是	是
C306Y	-6.8	-15.7	是	是
C306V	+1.9	+4.0	是	是

[0372]

突变	ΔG (真空)	ΔG (水)	趋同(真空)	趋同(水)
C330A	+2.5	+1.6	是	是
C330R	-219.2	-241.1	否	否
C330N	-77.1	-65.8	否	否
C330D	-62.2	-101.9	否	否
C330C	+2.1	-0.1	是	是
C330Q	-47.0	-49.8	否	否
C330E	-36.9	-66.2	否	否
C330G	-1.1	-0.2	是	是
C330H	-13.1	-19.0	是	是
C330I	+7.0	+2.6	是	是
C330L	-10.5	-10.9	是	是
C330K	-31.8	-26.4	是	否
C330M	-2.9	-4.6	是	是
C330F	+8.3	+12.2	是	是
C330S	+5.1	+4.7	是	是
C330T	-12.5	-14.8	是	是
C330W	+5.8	+14.6	是	是
C330Y	-1.8	-2.2	是	是
C330V	+5.3	+1.3	是	是

[0373]

[0374] 从上文表中可明显看出在每个这些位点处的明显突变 (Cys 至 Ser) 不被预测是能量有利的 (正的预测 ΔG)。被预测能稳定所述蛋白质的突变变化很大,这是因为它们的不同局部环境。

[0375] 应该注意,尽管发明人有活性和表达数据,但是这并不直接与这些表中的 ΔG 值相关。所述 ΔG 值可与热力学数据 (例如从 DSC 得到的数据) 联系起来,条件是假设所述突变不会显著改变蛋白质折叠路径 (已发现其适用于文献中的许多 alchemical FEP 应用)。

在表达与突变的 ΔG 或活性之间没有直接关系。

[0376] 对 DNA 聚合酶 I 和 EcoExo I [14] 的 Klenow 片段的晶体结构比较表明参与蛋白质催化和金属结合的残基是 Asp15、Glu17、Asp108、Asp186 和 His181。

[0377] 还可以通过化学方法修饰非半胱氨酸的残基,包括 Asp[1,2]、Glu[1,2]、Lys[3,4,5]、Arg[6,7,8]、His[9,10]、Tyr[11,12]、Trp[11,12] 和 Met[13]。因此将这些残基突变(以定向化学修饰 Exo I)使得发明人可以消除竞争性化学反应。

[0378] 实验研究(参见下文)显示很多这些突变均使得这些蛋白质保留了它们的活性,并且在一些情况下似乎还会增强活性,所述突变可以是单突变或者突变的组合。对大肠杆菌 Exo I 序列与其他相关的 Exo I 蛋白质序列的比较,和这些序列的同源模型的构建表明很多这些 Cys 残基在序列和结构上是保守的。这表明大肠杆菌 Exo I 的 Cys 突变体的特性可适用于相关 Exo I 酶的类型突变体。

[0379] 对 Swissprot 数据库中的所有 Exo I 酶的序列对比显示 C51、C98、C144 和 C306 是保守的。

[0380] 2.2.25 个半胱氨酸残基的鉴定

[0381] 核酸外切酶 I (SEQ ID NO :6) 包含 5 个天然存在的半胱氨酸残基(图 1)。一些所述半胱氨酸残基在大肠杆菌核酸外切酶 I (EcoExo I) 的其他同源物中是保守的(参见下文)。核酸外切酶 I 中的半胱氨酸不形成内部二硫键。

[0382] 2.2.3 通过突变为丝氨酸来去除所有 5 个半胱氨酸

[0383] 丝氨酸 (Ser) 氨基酸是与半胱氨酸 (Cys) 在结构上最相似的氨基酸,半胱氨酸中的 S 原子被丝氨酸中的 O 代替。因此可做出的最显而易见的突变是将 Cys (C) 变为 Ser (S)。为了测试该突变体的活性,将所有 5 个半胱氨酸都突变为丝氨酸,并尝试蛋白质表达。

[0384] ONLD0217 (Exo I C51S/C98S/C144S/C306S/C330S) 的表达导致产量下降、表达减少并且蛋白质不稳定。不能纯化所述蛋白质。这证明当所有 Cys 被 Ser 代替后,所述蛋白质不稳定。为了用基因工程方法制造无半胱氨酸的 EcoExo I,需要进一步研究所述蛋白质的结构。这部分地是在计算机建模的指导下进行的。

[0385] 2.2.4 通过突变为丝氨酸来代替各个半胱氨酸

[0386] 为了鉴别能够容许置换的半胱氨酸位置,用丝氨酸残基轮流置换各个半胱氨酸残基。

[0387] 表达构建体 ONLD0233 (C51S)、ONLD0234 (C98S)、ONLD0235 (C144S)、ONLD0236 (C306S)、ONLD0237 (C330S)。突变体 C98S 表达至充分的水平,数次尝试后,实现了充分水平的 C306S。然而,所有其他位置的表达水平大大下降。

[0388] 因此考虑备选突变来产生充分的表达水平,这样使得可测试所述突变体的核酸外切酶活性。将所述半胱氨酸分为高度可及的位置 (C98, C306, C330 ;图 2) 和深度埋藏的残基 (C51, C144)。

[0389] 2.2.5 用丝氨酸的备选氨基酸代替 C306 和 C330

[0390] 通过检测以下突变体来研究对位置 306 和 330 处的半胱氨酸残基的代替。

[0391] Exo I C306 → D, M, N, S, T

[0392] Exo I C330 → D, M, N, T, Y, H, L, M, Q

[0393] ONLD0335 (C306T) 的表达水平最高。该位置可容许的其他突变为 S、D 和 N。

[0394] ONLD0342(C330T) 的表达水平最高。该位置可容许的其他突变为 H、Q 和 M, 容许程度较低的是 D、N、L 和 Y。

[0395] 2.2.6 将 C98S 与 C306S 和 C330T 组合

[0396] 为避免双突变和三突变相互不利地影响, 研究以下组合。

[0397]

ONLD 编号	EcoExo I 突变体	表达
ONLD0366	C98S/C306S/C330T (SST)	高
ONLD0368	C98S/C306T/C330T (STT)	高
ONLD0367	C98S/C306T/C330H (STH)	低
ONLD0364	C98S/C306D/C330T (SDT)	低
ONLD0365	C98S/C306S/C330H (SSH)	失败
ONLD0363	C98S/C306D/C330H (SDH)	失败

[0398] 组合 ONLD0366 给出了最高的表达, ONLD0368 次之。组合 ONLD0365 和 ONLD0363 不能表达。ONLD0367 和 ONLD0364 的产率较低。

[0399] 活性分析 (参见方法部分) 显示所述 Exo I 突变体的相对活性比野生型的高 (图 3)。

[0400] 2.2.7 半胱氨酸连接点的引入

[0401] 连接点被设计为由 α -HL 直接与核酸外切酶 I 的相应半胱氨酸形成共价键或者通过连接物形成共价键。通过建模预测的连接点 (V42C、A83C、S90C、V94C、M184C) 在图 4 中显示。

[0402] 将上述半胱氨酸引入突变体 Exo I C98S、C306S、C330T 得到以下构建体:

[0403]

ONLD 编号	EcoExo I 突变体	半胱氨酸位置
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

[0404] 通过与作为凝胶迁移试剂起作用的马来酰亚胺 (maleimide)-PEG 反应来评估所述半胱氨酸的活性。所得的凝胶显示每种蛋白质均有 3 个反应活性半胱氨酸。该数据表明 C144 和 C51 处的半胱氨酸的可及程度比之前推测的高, 并且也能与马来酰亚胺 (maleimide)-PEG 反应。进行更多轮突变来代替 C144 和 C51 以确保所述蛋白质仅在一个位置处反应。

[0405] 2.2.8 用丝氨酸的备选氨基酸代替残基 C144

[0406] 使用分子建模从总共 18 个氨基酸中 (A、E、G、I、L、N、R、T、W、D、F、H、K、M、Q、S、V、Y) 评估可能的突变。制备从所述建模中预测的位置 144 处的 6 个最稳定的突变体:

[0407] Exo I C98S, C306S, C330S, C144 → A, G, L, T, H, M

[0408] 在 6 个经测试的突变体中, (ONLD0403)Exo I C98S/C306S/C330T/C144M 和 (ONLD0404)Exo I C98S/C306S/C330T/C144T 表达良好。其他突变体的表达水平下降。

[0409] 2.2.9 用丝氨酸的备选氨基酸置换残基 C51

[0410] 使用分子建模从总共 17 个氨基酸中 (A、D、E、F、G、H、I、K、L、M、N、R、S、T、V、W、Y) 评估可能的突变。制备从所述建模中预测的位置 51 处的 8 个最稳定的突变体:

[0411] Exo I C98S, C306S, C330S, C51 → A, G, H, K, L, M, T, W

[0412] 在 8 个经测试的突变体中, 仅 (ONLD0393)Exo I C98S/C306S/C330T/C51A 表达。

[0413] 2.2.105 个半胱氨酸替换和加入连接位点的组合

[0414] 对以下组合进行评估。这些包括所有 5 个半胱氨酸替换和半胱氨酸连接点的加入。还可对添加至 N 或 C 末端的半胱氨酸进行连接。

[0415]

ONLD 编号	EcoExo I 突变体	表达
ONLD0416	C51A/C98S/C144T/C306S/C330T/V42C	很好
ONLD0418	C51A/C98S/C144T/C306S/C330T/M184C	很好
ONLD0415	C51A/C98S/C144M/C306S/C330T/V42C	好
ONLD0417	C51A/C98S/C144M/C306S/C330T/M184C	好
ONLD0413	C51A/C98S/C144M/C306S/C330T/n-端 C	失败
ONLD0414	C51A/C98S/C144T/C306S/C330T/n-端 C	失败
ONLD0425	C51A/C98S/C144T/C306S/C330T/V94C	非常差/失败 *
ONLD0422	C51A/C98S/C144T/C306S/C330T/S90C	差
ONLD0421	C51A/C98S/C144T/C306S/C330T/A83C	好

[0416] * ONLD0425 没有很好地表达 – 不足以确定浓度但正好足以确认活性。

[0417] 带有 N- 末端半胱氨酸残基的两个突变体 (OND0413, ONLD0414) 没有表达。所有其他组合都表达, 并且均是活性核酸外切酶。组合 ONLD0415 和 ONLD0417 的表达程度没有 ONLD0416 和 ONLD0418 的表达程度好。

[0418] 图 5 示出了活性结果。

[0419] 还评估了以下组合。

[0420] 位置 98 和 306

[0421]

ONLD 编号	EcoExo I 突变体	表达
ONLD0411	C51A/C98S/C144T/C306S/C330T	高
ONLD0432	C51A/C98T/C144T/C306S/C330T	高
ONLD0433	C51A/C98T/C144T/C306T/C330T	高

[0422] 位置 98

[0423]

ONLD 编号	EcoExo I 突变体	表达
ONLD0411	C51A/C98S/C144T/C306S/C330T	很好
ONLD0432	C51A/C98T/C144T/C306S/C330T	很好
ONLD0453	C51A/C98G/C144T/C306S/C330T	好
ONLD0454	C51A/C98K/C144T/C306S/C330T	好
ONLD0455	C51A/C98L/C144T/C306S/C330T	失败
ONLD0456	C51A/C98V/C144T/C306S/C330T	很好

[0424] 位置 306

[0425]

ONLD 编号	EcoExo I 突变体	表达
ONLD0433	C51A/C98T/C 144T/C306T/C330T	好
ONLD0477	C51A/C98T/C 144T/C306M/C330T	好
ONLD0478	C51A/C98T/C 144T/C306N/C330T	好
ONLD0479	C51A/C98T/C 144T/C306D/C330T	差
ONLD0480	C51A/C98T/C 144T/C306A/C330T	差

[0426] 2.2.11 蛋白质的 DSC 分析

[0427] 示差扫描量热法 (DSC) 是可测量样本内的热转变的技术。它的主要用途是测量生物分子的热稳定性 (表示为转变中点)。所述转变中点可被用作热稳定性的标记。

[0428] 发现最早分析的所有 5 个半胱氨酸均被代替的突变体 (Exo-ASTST) 非常不稳定, 活性随着时间流逝大大丧失。测得 T_m 为 35°C , 而野生型为 49°C 。

[0429] 在更多轮突变体筛选之后, 鉴定到了酶 Exo-ATTTT, 其 T_m 为 40°C , 稳定性随时间流逝大大提高。其他组合 (例如 Exo-ATTMT 和 Exo-AVTAT) 具有类似的但是未提高的 T_m , 40°C 。

[0430] 发现突变体 AVTMT 有略微高的 T_m , 但是在一些条件下更容易聚集。

[0431] 为连接目的而加入的优选的非天然半胱氨酸 (A83C) 的加入看起来对稳定性没有不利影响。

[0432]

样本	ONLD	T _m (解链温度) 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

[0433]	DSC 结果	T _m		
	样本	ONLD	Exp1a	Exp1b
	Exo-ATTTT-A83C-Strep-Strep	ONLD0540	40.68	40.94

[0434] 2.2.12 其他目的突变体 :C144I 和 C330I

[0435] C144 邻近 Exo-I 的 DNA 结合位点和催化位点。没有对残基 C144 和 C330 进行像其他半胱氨酸那样多的突变来寻找最好的组合。基于对蛋白质序列的分析和之前的建模工作,构建了以下两个蛋白质 (ATTTT 和 ATITI) 的模型。

[0436] 将 ATTTT 与 ATITI 进行比较的建模工作表明 T330 侧链几乎不会导致对所述蛋白质二级结构的破坏。但是 T144 被疏水核心包住,当 T144 侧链翻转时所述疏水核心看起来遭到了破坏,导致所述核心的显著破坏。ATITI 模型模拟没有显示出相同程度的 T144 周围的破坏。

[0437] 2.2.13 酶活性

[0438] 所检查的所述构建体是 Exo-ATTTT-A83C。用在这些实施例中的连接物是一种 ssDNA 连接物,其 5' 末端游离,而 3' 末端连接至所述酶。测量了连接有 ssDNA 连接物的 Exo-ATTTT-A83C 的活性,然后通过加入 DTT 除去所述连接物并测量活性。没有观察到活性变化,证实所述连接物的存在不会对活性有不利影响 (图 8)。在该实施例中,针对 Exo-ATTTT-A83C 比较了游离酶和经 PNA 修饰的酶的活性。没有观察到加入 PNA 连接物后会有活性丧失 (图 9)。Exo ATTTT-M184C (带有 PEG)。对照是 PEG 存在的情况下的 Exo ATTTT。

没有观察到加入 PEG 连接物后会有活性丧失（图 10）。

[0439] 2.3 结论

[0440] 发明人制造了一系列 EcoExo I 的突变体，其所有天然半胱氨酸都已被去除。此外，单个半胱氨酸已被引入以控制 EcoExo I 的连接化学。有点出乎意料的是，所述突变体中的一些的活性比野生型的高。

[0001]

序列表

SEQ ID NO: 1 (WT α -HL)

1	ATGGCAGATT	CTGATATTA	TATTAACC	GGTACTACAG	ATATTGGAAG	CAATACTACA	GTAACACAG
71	GTGATTTAGT	CACTTATGAT	AAAGAAAATG	GCATGCACAA	AAAAGTATTT	TATAGTTTTA	TCGATGATAA
141	AAATCACAAAT	AAAAAACTGC	TAGTTATTAG	AACAAAAGGT	ACCATTGCTG	GTCATAATAG	AGTTTATAGC
211	GAAGAAGGTG	CTAACAAAAG	TGGTTTAGCC	TGGCCTTCAG	CCTTTAAGGT	ACAGTTGCAA	CTACCTGATA
281	ATGAAGTAGC	TCAAAATATCT	GATTAATATC	CAAGAAATTC	GATTGATACA	AAAGAGTATA	TGAGTACTTT
351	AACTTATGGA	TTCACCGTA	ATGTTACTGG	TGATGATACA	GGAAAAATTG	GCGGCCTTAT	TGGTGCAAAAT
421	GTTTCGATTG	GTCATACACT	GAAATATGTT	CAACCTGATT	TCAAAAACAA	TTTAGAGAGC	CCAACTGATA
491	AAAAAGTAGG	CTGAAAAGTG	ATATTTAACA	ATATGGTGAA	TCAAAAATTG	GGACCATACG	ATCGAGATTC
561	TGGAACCCG	GTATATGGCA	ATCAACTTTT	CATGAAAAC	AGAAATGGTT	CTATGAAAAG	AGCAGATAAC
631	TTCCTTGATC	CTAACAAAAG	AAGTTCTCTA	TTATCTTCAG	GGTTTTCCAC	AGACTTCGCT	ACAGTTATTA
701	CTATGGATAG	AAAAGCATCC	AAACAACAAA	CAAATATAGA	TGTAATATAC	GAACGAGTTC	GTGATGATTA
771	CCAATTGCAT	TGGACTTCAA	CAAATTGGAA	AGGTACCAAT	ACTAAAGATA	AATGGACAGA	TCGTTCTTCA
841	GAAAGATATA	AAATCGATTG	GGAAAAAGAA	GAAATGACAA	AT		

SEQ ID NO: 2 (WT α -HL)

1	ADSDINIKTG	TTDIGSNTTV	KTGDLVITYDK	ENGMHKKVYF	SFIDDKNHNK	KLLVIRTKGT	IAGQYRVYSE
71	EGANKSGLAW	PSAFKVQLQL	PDNEVAQISD	YYPNRSIDTK	EYMSLTLYGF	NGNVTDGDTG	KIGGLIGANV
141	SIGHTLKYYQ	PDFKTIIESP	TDKKGWVKVI	FNNMNVQNWG	PYDRDSWNPV	YGNQLFMKTR	NGSMKAADNF
211	LDPNKASSLL	SSGFSPDFAT	VITMDRKASK	QQTNIIDVIYE	RVRDDYQLHW	TSTNWKGTNT	KDKWTDRESSE
281	RYKIDWEKEE	MTN					

SEQ ID NO: 3 (α -HL L135C/N139Q)

1	ATGGCAGATT	CTGATATTA	TATTAACC	GGTACTACAG	ATATTGGAAG	CAATACTACA	GTAACACAG
71	GTGATTTAGT	CACTTATGAT	AAAGAAAATG	GCATGCACAA	AAAAGTATTT	TATAGTTTTA	TCGATGATAA
141	AAATCACAAAT	AAAAAACTGC	TAGTTATTAG	AACAAAAGGT	ACCATTGCTG	GTCATAATAG	AGTTTATAGC
211	GAAGAAGGTG	CTAACAAAAG	TGGTTTAGCC	TGGCCTTCAG	CCTTTAAGGT	ACAGTTGCAA	CTACCTGATA
281	ATGAAGTAGC	TCAAAATATCT	GATTAATATC	CAAGAAATTC	GATTGATACA	AAAGAGTATA	TGAGTACTTT
351	AACTTATGGA	TTCACCGTA	ATGTTACTGG	TGATGATACA	GGAAAAATTG	GCGGCCTTAT	TGGTGCAAAAT
421	GTTTCGATTG	GTCATACACT	GAAATATGTT	CAACCTGATT	TCAAAAACAA	TTTAGAGAGC	CCAACTGATA
491	AAAAAGTAGG	CTGAAAAGTG	ATATTTAACA	ATATGGTGAA	TCAAAAATTG	GGACCATACG	ATCGAGATTC
561	TGGAACCCG	GTATATGGCA	ATCAACTTTT	CATGAAAAC	AGAAATGGTT	CTATGAAAAG	AGCAGATAAC
631	TTCCTTGATC	CTAACAAAAG	AAGTTCTCTA	TTATCTTCAG	GGTTTTCCAC	AGACTTCGCT	ACAGTTATTA
701	CTATGGATAG	AAAAGCATCC	AAACAACAAA	CAAATATAGA	TGTAATATAC	GAACGAGTTC	GTGATGATTA
771	CCAATTGCAT	TGGACTTCAA	CAAATTGGAA	AGGTACCAAT	ACTAAAGATA	AATGGACAGA	TCGTTCTTCA
841	GAAAGATATA	AAATCGATTG	GGAAAAAGAA	GAAATGACAA	AT		

SEQ ID NO: 4 (α -HL L135C/N139Q)

1	ADSDINIKTG	TTDIGSNTTV	KTGDLVITYDK	ENGMHKKVYF	SFIDDKNHNK	KLLVIRTKGT	IAGQYRVYSE
71	EGANKSGLAW	PSAFKVQLQL	PDNEVAQISD	YYPNRSIDTK	EYMSLTLYGF	NGNVTDGDTG	KIGGCIQAQV
141	SIGHTLKYYQ	PDFKTIIESP	TDKKGWVKVI	FNNMNVQNWG	PYDRDSWNPV	YGNQLFMKTR	NGSMKAADNF
211	LDPNKASSLL	SSGFSPDFAT	VITMDRKASK	QQTNIIDVIYE	RVRDDYQLHW	TSTNWKGTNT	KDKWTDRESSE
281	RYKIDWEKEE	MTN					

SEQ ID NO: 5 (WT EcoExo I; ONLD0267)

1	ATGATGAACG	ATGGCAACA	CGAGAGCACC	TTCTGTTC	ATGATTATGA	AACCTTCGGT	ACCCATCCGG
71	CCCTGGATCG	TCCGGCGCAG	TTTGGCGCCA	TTCGCACCGA	TAGCGAATTC	AATGTGATTC	GCGAACCCGA
141	AGTGTFTTAT	TGCAAACCCG	CCGATGATTA	TCTGCCGCAG	CCGGGTGCGG	TGCTGATTAC	CGGTATTACC
211	CCGCAGGAAG	CGCGCGCGAA	AGGTGAAAAC	GAAGCGGCGT	TTGCCGCGCG	CATTCATAGC	CTGTTTACC
281	TGCCGAAAAC	CTGCATTCTG	GGCTATAACA	ATGTGCGCTT	CGATGATGAA	GTTACCCGTA	ATATCTTTTA
351	TGCTAACTTT	TATGATCCGT	ATGCGTGGAG	CTGGCAGCAT	GATAACAGCC	GTTGGGATCT	GCTGGATGTC
421	ATGCGCGCGT	GCTATGCGCT	GCGCCCGGAA	GGCATTAAAT	GGCCGAAAA	CGATGATGGC	CTGCCGAGCT
491	TTCGCTGGA	ACATCTGACC	AAAGCCAACG	GCATTGAACA	TAGCAATGCC	CATGATGCGA	TGGCCGATGT
561	TTATGCGACC	ATTGCGATGG	CGAAACTGGT	TAAAACCCGT	CAGCCGCGCC	TGTTTGAATTA	TCTGTTTACC
631	CACCGTAAACA	AACACAAACT	GATGGCGCTG	ATTGATGTTT	CGCAGATGAA	ACCCTGGTGG	CATGTTAGCC
701	GCATGTTTGG	CGCCTGGCGC	GGCAACACCA	GCTGGGTGGC	CCCCTGGCC	TGGCACCCGG	AAAATCGTAA
771	CGCCGTGATT	ATGGTTGATC	TGGCCGGTGA	TATTAGCCCG	CTGCTGGAAC	TGGATAGCGA	TACCTGCGT
841	GAACGCGTGT	ATACCGCCAA	AACCGATCTG	GGCGATAATG	CCGCGGTGCC	GGTGAAACTG	GTTTACATTA
911	ACAAATGCCC	GGTGTGGGCC	CAGGCGAACA	CCCTGCGCCC	CGAAGATGCG	GGAAGATGCG	GTATTAATCG
981	CCAGCATTGT	CTGGATAATC	TGAAAATCCT	CGGTGAAAAC	CCGCAGGTGC	GTGAAAAAGT	GGTGGCGATC
1051	TTCGGGAAAG	CGGAACCGTT	CACCCCGAGC	GATAACGTGG	ATGCGCAGCT	GTATAACGGC	TTCCTTAGCG
1121	ATGCCGATCG	CGCGCGGATG	AAAATCGTTC	TGAAAACCGA	ACCGCGCAAT	CTGCCGCGCG	TGGATATTAC
1191	CTTTGTTGAT	AAACGTATTG	AAAAACTGCT	GTTTAATTTA	CGTGCAGCGA	ATTTCCGGGG	TACCCGTGAT
1261	TATGCCGAAC	AGCAGCGTTG	GCTGGAACAT	CGTGTGTCAG	TTTTCACCCC	GGAAATTTCTG	CAGGGTTATG
1331	CGGATGAACT	GCAGATGCTG	GTTTCAGCAT	ATGCCGATGA	TAAAGAAAAA	GTGGCGCTGC	

[0002]

SEQ ID NO: 6 (WT EcoExo I; ONLD0267)

```

1      MMNDGKQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVfy
51     CKPADDYLPQ PGAVLITGTT PQEARAKGEN EAAFAARIHS LFTVPKTCIL
101    GYNVRFDDDE VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRACYALRPE
151    GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201    QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251    WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYTAKTDL GDNAAVPVKL
301    VHINKCPVLA QANTLRPEDA DRLGINRQHC LDNLKILREN PQVREKVVAI
351    FAEAEPTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401    KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451    VQYADDKEK VALLKALWQY AEEIV SGSGH HHHHH

```

SEQ ID NO: 7 EcoExo I C98S/C306S/C330T/C51A (ONLD0393)

```

1      ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71     CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141    AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGC GG TGCTGATTAC CGGTATTACC
211    CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCGCGCG CATTATAGC CTGTTTACC
281    TGCCGAAAAC CAGCATTCTG GGTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351    TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421    ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAA CGATGATGGC CTGCCGAGCT
491    TTCGTCGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561    TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGCC TGTTTGATTA TCTGTTTACC
631    CACCGTAACA AACACAAACT GATGCGCGTG ATPGATGTTT CCGAGATGAA ACCGCTGGTG CATGTGAGCG
701    GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771    CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841    GAACGCCTGT ATACCGCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911    ACAAAGCCC GGTGCTGGCC CAGCGGAACA CCTGCGCCC CCGAGATGCG GATCGTCTGG GTATTAATCG
981    CCAGCATACC CTGGATAATC TGAATACTCT GCGTGAATAA CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051   TTCGCGGAAG CGGAACCGTT CACCCGAGC GATAACGTGG ATGCGCAGCT GTATAACCGC TTCTTTAGCG
1121   ATGCGGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGAATTTAC
1191   CTTTGTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGCGCGCA ATTTCCGGG TACCCTGGAT
1261   TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTACCCC GGAATTTCTG CAGGGTTATG
1331   CGGATGAAC GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAA GTGGCGCTGC TGAAGCGCT
1401   GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 8 EcoExo I C98S/C306S/C330T/C51A (ONLD0393)

```

1      MMNDGKQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVfy
51     AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101    GYNVRFDDDE VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRACYALRPE
151    GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201    QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251    WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYTAKTDL GDNAAVPVKL
301    VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351    FAEAEPTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401    KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451    VQYADDKEK VALLKALWQY AEEIV SGSGH HHHHH

```

SEQ ID NO: 9 EcoExo I C98S/C306S/C330T/C144M (ONLD0403)

```

1      ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71     CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141    AGTGTTTTAT TGCAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGC GG TGCTGATTAC CGGTATTACC
211    CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCGCGCG CATTATAGC CTGTTTACC
281    TGCCGAAAAC CAGCATTCTG GGTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351    TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421    ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAA CGATGATGGC CTGCCGAGCT
491    TTCGTCGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561    TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGCC TGTTTGATTA TCTGTTTACC
631    CACCGTAACA AACACAAACT GATGCGCGTG ATPGATGTTT CCGAGATGAA ACCGCTGGTG CATGTGAGCG
701    GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771    CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGATAGCGA TACCCTGCGT
841    GAACGCCTGT ATACCGCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911    ACAAAGCCC GGTGCTGGCC CAGCGGAACA CCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981    CCAGCATACC CTGGATAATC TGAATACTCT GCGTGAATAA CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051   TTCGCGGAAG CGGAACCGTT CACCCGAGC GATAACGTGG ATGCGCAGCT GTATAACCGC TTCTTTAGCG
1121   ATGCGGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGAATTTAC
1191   CTTTGTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGCGCGCA ATTTCCGGG TACCCTGGAT
1261   TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTACCCC GGAATTTCTG CAGGGTTATG
1331   CGGATGAAC GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAA GTGGCGCTGC TGAAGCGCT

```

[0003]

1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

SEQ ID NO: 10 EcoExo I C98S/C306S/C330T/C144M (ONLD0403)

1 MMNDGKQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVFY
51 CKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRAMYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IAMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVS GMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPF T PS DNVDAQLYNG FFSADADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIV SGS GH HHHHH

SEQ ID NO: 11 EcoExo I C98S/C306S/C330T/C144T (ONLD0404)

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141 AGTGT TTTAT TGCAAACCGG CCGATGATTA TCTGCCGCGAG CCGGGTGCGG TGCTGATTAC CGGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGG CATTCTAGC CTGTTTACCG
281 TGCCGAAAAC CAGCATCTCG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTGTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGCGCGTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA CCGTGGTGCC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAAGCCC GGTGCTGGCC CAGGCGAACA CCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATAAC CTGGATAAAT TGAATACTCT GCGTGA AAAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTGCGGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACCGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGT TGGAT AAACGTATTG AAAAATGCT GTTTAATFAT CGTGCGCGCA ATTTCCCGG TACCCTGGAT
1261 TATGCCG AAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCTG
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

SEQ ID NO: 12 EcoExo I C98S/C306S/C330T/C144T (ONLD0404)

1 MMNDGKQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVFY
51 CKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IAMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVS GMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPF T PS DNVDAQLYNG FFSADADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIV SGS GH HHHHH

SEQ ID NO: 13 EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415)

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141 AGTGT TTTAT GCGAAAACCG CCGATGATTA TCTGCCGCGAG CCGGGTGCGG TGCTGATTAC CGGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGG CATTCTAGC CTGTTTACCG
281 TGCCGAAAAC CAGCATCTCG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTGTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGCGCGTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA CCGTGGTGCC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAAGCCC GGTGCTGGCC CAGGCGAACA CCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATAAC CTGGATAAAT TGAATACTCT GCGTGA AAAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTGCGGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCCGAGCT GTATAACCGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGT TGGAT AAACGTATTG AAAAATGCT GTTTAATFAT CGTGCGCGCA ATTTCCCGG TACCCTGGAT
1261 TATGCCG AAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCTG

[0004]

1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

SEQ ID NO: 14 EcoExo I C51A/C98S/C144M/C306S/C330T/V42C (ONLD0415)

1 MMNDGKQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NCIGEPEVFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRAMYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IAMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVS GMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPFPTS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIV SGS GH HHHHH

SEQ ID NO: 15 EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416)

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGCGG TGTGATTAC CGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGG CATTCTAGC CTGTTTACC
281 TGCCGAAAAC CAGCATCTCG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTGTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGT CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA CCGTGGTGCC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATAAC CTGGATAAAT TGAATACTCT GCGTGA AAAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTGCGCGAAG CCGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACCGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGTGTGAT AAACGTATTG AAAAATCTGCT GTTTAATFAT CGTGCGCGCA ATTTCCGCGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCTG
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

SEQ ID NO: 16 EcoExo I C51A/C98S/C144T/C306S/C330T/V42C (ONLD0416)

1 MMNDGKQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NCIGEPEVFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRAMYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IAMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVS GMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPFPTS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIV SGS GH HHHHH

SEQ ID NO: 17 EcoExo I C51A/C98S/C144M/C306S/C330T/M184C (ONLD0417)

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGCGG TGTGATTAC CGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGG CATTCTAGC CTGTTTACC
281 TGCCGAAAAC CAGCATCTCG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGT GCGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTGTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGT CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA CCGTGGTGCC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATAAC CTGGATAAAT TGAATACTCT GCGTGA AAAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTGCGCGAAG CCGAACCGTT CACCCCGAGC GATAACGTGG ATGCCGAGCT GTATAACCGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGTGTGAT AAACGTATTG AAAAATCTGCT GTTTAATFAT CGTGCGCGCA ATTTCCGCGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCTG

[0005]

1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

SEQ ID NO: 18 EcoExo I C51A/C98S/C144M/C306S/C330T/M184C (ONLD0417)

1 MMNDGKQQST FLFHDIYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRAMYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADCYAT IAMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVS GMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIV SGS GH HHHHH

SEQ ID NO: 19 EcoExo I C51A/C98S/C144T/C306S/C330T/M184C (ONLD0418)

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG CCGAACCGGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCAG CCGGGTGGCG TGCTGATTAC CCGTATTACC
211 CCGCAGGAAG CGCGCGCAA AGGTGAAAAC GAAGCGCGT TGGCCGCGC CATTCATAGC CTGTTTACC
281 TGCCGAAAAC CAGCATTCG GGCATAACA ATGTGCGCT CGATGATGAA GTTACCCGTA ATATCTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGAATG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GCGCGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGT GCGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGC TGTGTTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCG GCAACACCA GCTGGGTGGC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGATAGCGA TACCCTGCGT
841 GAACGCTGT ATACCGCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTACATTA
911 AAAAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGATAATC TGAATTCCT GCGTAAAAAC CCGCAGGTGC GTAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CCGAACCGTT CACCCCGAG GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGTTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGCGCGCA ATTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CCGCTCAGG TTTACACCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAAC TGCAGTGTG GTTCAGCAG ATGCCGATGA TAAAGAAAA GTGCCGCTGC TGAAGCGCT
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

SEQ ID NO: 20 EcoExo I C51A/C98S/C144T/C306S/C330T/M184C (ONLD0418)

1 MMNDGKQQST FLFIYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRAMYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADCYAT IAMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVS GMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIV SGS GH HHHHH

SEQ ID NO: 21 EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411)

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG CCGAACCGGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCAG CCGGGTGGCG TGCTGATTAC CCGTATTACC
211 CCGCAGGAAG CGCGCGCAA AGGTGAAAAC GAAGCGCGT TGGCCGCGC CATTCATAGC CTGTTTACC
281 TGCCGAAAAC CAGCATTCG GGCATAACA ATGTGCGCT CGATGATGAA GTTACCCGTA ATATCTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGAATG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GCGCGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGC TGTGTTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCG GCAACACCA GCTGGGTGGC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGATAGCGA TACCCTGCGT
841 GAACGCTGT ATACCGCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTACATTA
911 AAAAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGATAATC TGAATTCCT GCGTAAAAAC CCGCAGGTGC GTAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CCGAACCGTT CACCCCGAG GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGTTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGCGCGCA ATTTCCGGG TACCCTGGAT

[0006]

```

1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CCGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCTG
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 22 EcoExo I C51A/C98S/C144T/C306S/C330T (ONLD0411)

```

1 MMNDGKQST FLFHIDYETFG THPALDRPAQ FFAIRTDSEF NVIGEPEVIFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTSIL
101 GYNVRFDDDE VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201 QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELDSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPFPTS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 23 EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432)

```

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCAG CCGGGTGGCG TGCTGATTAC CCGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGC CATTCATAGC CTGTTTACCG
281 TGCCGAAAAC CACCATTCTG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAAA CGATGATGGC CTGCCGAGCT
491 TTCGTCGTGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAACTGGT TAAAACCCGT CAGCCGCGCC TGTGTTGATTA TCTGTTTACC
631 CACCGTAAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGGATAATC TGAATACTCT GCGTGA AAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTTGTGAT AAACGTATTG AAAAAGTCTT GTTAAATTAT CGTGGCGCA ATTTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACA TCGTGCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CCGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCTG
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 24 EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0432)

```

1 MMNDGKQST FLFHIDYETFG THPALDRPAQ FFAIRTDSEF NVIGEPEVIFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTTIL
101 GYNVRFDDDE VTRNIFYRNF YDPYAWSWQH DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201 QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELDSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPFPTS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 25 EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433)

```

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCAG CCGGGTGGCG TGCTGATTAC CCGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGC CATTCATAGC CTGTTTACCG
281 TGCCGAAAAC CACCATTCTG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAAA CGATGATGGC CTGCCGAGCT
491 TTCGTCGTGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAACTGGT TAAAACCCGT CAGCCGCGCC TGTGTTGATTA TCTGTTTACC
631 CACCGTAAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGGATAATC TGAATACTCT GCGTGA AAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG

```

[0007]

```

1121   ATGCCGATCG CGCGGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGGCGC TGGATATTAC
1191   CTTGTGTGAT AACCGTATTG AAAAATCGCT GTTAAATFAT CGTGCCGCGCA ATTTTCGGGG TACCCTGGAT
1261   TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTATG
1331   CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401   GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 26 EcoExo I C51A/C98T/C144T/C306T/C330T (ONLD0433)

```

1      MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEV FY
51     AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAFAARIHS LFTVPKTTIL
101    GYNNVRFDD E VTRNIFYRNF YDPYAWSQWQ DNSRWDL LDV MRATYALRPE
151    GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IAMAKLVKTR
201    QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251    WHPENRNAV I MVDLAGDISP LLELDSDTLR ERLYTAKTDL GDNAAVPVKL
301    VHINKTPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351    FAEAEPFPT S DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401    KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGVADELQML
451    VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 27 EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451)

```

1      ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71     CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCGGA
141    AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGGCG TGCTGATTAC CGGTATTACC
211    CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGC CATTCATAGC CTGTTTACCG
281    TGCCGAAAA CACCATTCTG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351    TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421    ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAA CGATGATGGC CTGCCGAGCT
491    TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561    TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTTTGATTA TCTGTTTACC
631    CACCGTAACA AACACAAACT GATGCGCGTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701    GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771    CGCCGTGATT ATGTTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841    GAACGCCTGT ATACCGCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911    ACAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981    CCAGCATACC CTGGATAATC TGAATATCCT CGGTGAAAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGGATC
1051   TTCGCGGAAG CGGAACCGTT CACCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGCG TTCTTTAGCG
1121   ATGCCGATCG CGCGCGGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGGCGC TGGATATTAC
1191   CTTGTGTGAT AACCGTATTG AAAAATCGCT GTTAAATFAT CGTGCCGCGCA ATTTTCGGGG TACCCTGGAT
1261   TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTATG
1331   CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401   GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 28 EcoExo I C51A/C98T/C144M/C306S/C330T (ONLD0451)

```

1      MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEV FY
51     AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAFAARIHS LFTVPKTTIL
101    GYNNVRFDD E VTRNIFYRNF YDPYAWSQWQ DNSRWDL LDV MRAMYALRPE
151    GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IAMAKLVKTR
201    QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251    WHPENRNAV I MVDLAGDISP LLELDSDTLR ERLYTAKTDL GDNAAVPVKL
301    VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351    FAEAEPFPT S DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401    KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGVADELQML
451    VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 29 EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452)

```

1      ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71     CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCGGA
141    AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGGCG TGCTGATTAC CGGTATTACC
211    CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCCGCGC CATTCATAGC CTGTTTACCG
281    TGCCGAAAA CACCATTCTG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351    TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421    ATGCGCGCGA TGTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAA CGATGATGGC CTGCCGAGCT
491    TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561    TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTTTGATTA TCTGTTTACC
631    CACCGTAACA AACACAAACT GATGCGCGTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701    GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771    CGCCGTGATT ATGTTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841    GAACGCCTGT ATACCGCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911    ACAAAGCCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG

```

[0008]

981	CCAGCATACC	CTGGATAATC	TGAAAAATCCT	CGGTGAAAAC	CCGCAGGTGC	GTGAAAAAGT	GGTGGCGATC
1051	TTGCGGGAAG	CGGAACCGTT	CACCCCGAGC	GATAACGTGG	ATGCGCAGCT	GTATAACGGC	TTCTTTAGCC
1121	ATGCCGATCG	CGCGCGATG	AAAAATCGTTC	TGAAAACCGA	ACCGCGCAAT	CTGCCGGCGC	TGGATAATTAC
1191	CTTTGTTGAT	AAACGTATTG	AAAAACTGCT	GTTTAATFAT	CGTGCGCGCA	ATTTTCCGGG	TACCCCTGGAT
1261	TATGCCGAAC	AGCAGCGTTG	GCTGGAACAT	CGTCGTCAGG	TTTTACCCCC	GGAAATTTCTG	CAGGGTTATG
1331	CGGATGAACT	GCAGATGCTG	GTTCAGCAGT	ATGCCGATGA	TAAAGAAAAA	GTGGCGCTGC	TGAAAGCGCT
1401	GTGGCAGTAT	CGGGAAGAAA	TCGTTTCTGG	CTCTGGTCAC	CATCATCATC	ACCAC	

SEQ ID NO: 30 EcoExo I C51A/C98T/C144M/C306T/C330T (ONLD0452)

1	MMNDGKQST	FLFHDYETFG	THPALDRPAQ	FAAIRTDSEF	NVIGEPEVIFY		
51	AKPADDYLPQ	PGAVLITGIT	PQEARAKGEN	EAAFAARIHS	LFTVPKTTIL		
101	GYNVRFDDDE	VTRNIFYRNF	YDPYAWSQWH	DNSRWDLDDV	MRAMYALRPE		
151	GINWPENDDG	LPSFRLEHLT	KANGIEHSNA	HDMADVYAT	IAMAKLVKTR		
201	QPRLFDYLFY	HRNKHKLML	IDVPQMKPLV	HVSGMFGAWR	GNTSWVAPLA		
251	WHPENRNAVY	MVDLAGDISP	LLELSDTLR	ERLYTAKTDL	GDNAAVPVKL		
301	VHINKTPVLA	QANTLRPEDA	DRLGINRQHT	LDNLKILREN	PQVREKVVAI		
351	FAEAEPTFPS	DNVDAQLYNG	FFSDADRAAM	KIVLETEPRN	LPALDITFVD		
401	KRIEKLFFNY	RARNFPGTLD	YAEQQRWLEH	RRQVFTPEFL	QGYADELQML		
451	VQYADDKEK	VALLKALWQY	AEEIVSGSGH	HHHHH			

SEQ ID NO: 31 EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453)

1	ATGATGAACG	ATGGCAAACA	GCAGAGCACC	TTCTGTFTTC	ATGATTATGA	AACCTTCGGT	ACCCATCCGG
71	CCCTGGATCG	TCCGGCGCAG	TTTGGCGCCA	TTCGCACCGA	TAGCGAATTC	AATGTGATTG	GCGAACCCGGA
141	AGTGTTTTAT	CGGAAACCGG	CCGATGATTA	TCTGCCGCGCAG	CCGGGTGCGG	TGCTGATTAC	CGGTATTACC
211	CCGCAGGAAG	CGCGCGCGAA	AGGTGAAAAC	GAAGCGGCGT	TTGCCCGCGC	CATTCATAGC	CTGTTTACCC
281	TGCCGAAAAC	CGGATTTCTG	GGCTATAACA	ATGTGCGCTT	CGATGATGAA	GTTACCCGTA	ATATCTTTTA
351	TCGTAACCTT	TATGATCCGT	ATGCGTGGAG	CTGGCAGCAT	GATAACAGCC	GTTGGGATCT	GCTGGATGTG
421	ATGCGCGCGA	CCTATGCGCT	CGCCCCGAAA	GGCATTAAAT	GGCCGGAAAA	CGATGATGCC	CTGCCGAGCT
491	TTGCTCTGGA	ACATCTGACC	AAAGCCAACG	GCATTGAACA	TAGCAATGCC	CATGATGCGA	TGGCCGATGT
561	TTATGCGACC	ATTGCGATGG	CGAAACTGGT	TAAAACCCGT	CAGCCGCGCC	TGTTTGATTA	TCTGTTTACC
631	CACCGTAACA	AACACAAACT	GATGCGCGTG	ATTGATGTTT	CGCAGATGAA	ACCGCTGGTG	CATGTGAGCC
701	GCAATGTTGG	CGCCTGGCGC	GGCAACACCA	GCTGGGTGGC	CCCGCTGGCC	TGGCACCCGG	AAAAATCGTAA
771	CGCCGTGATT	ATGTTTGATC	TGGCCGGTGA	TATTAGCCCG	CTGCTGGAAC	TGGATAGCGA	TACCCTGCGT
841	GAACGCCTGT	ATACCCGCAA	AACCGATCTG	GCGGATAATG	CCGCCGTGCC	GGTGAAGTCT	GTTACACATTA
911	ACAAAAGCCC	GGTGTGGGCC	CAGGCGAACA	CCCTGCGGCC	GGAAAGATGCG	GATCGTCTGG	GTATTAATCC
981	CCAGCATACC	CTGGATAAATC	TGAAAATCCT	CGGTGAAAAC	CCGCAGGTGC	GTGAAAAAGT	GGTGGCGATC
1051	TTGCGGGAAG	CGGAACCGTT	CACCCCGAGC	GATAACGTGG	ATGCGCAGCT	GTATAACGGC	TTCTTTAGCC
1121	ATGCCGATCG	CGCGCGATG	AAAAATCGTTC	TGAAAACCGA	ACCGCGCAAT	CTGCCGGCGC	TGGATAATTAC
1191	CTTTGTTGAT	AAACGTATTG	AAAAACTGCT	GTTTAATFAT	CGTGCGCGCA	ATTTTCCGGG	TACCCCTGGAT
1261	TATGCCGAAC	AGCAGCGTTG	GCTGGAACAT	CGTCGTCAGG	TTTTACCCCC	GGAAATTTCTG	CAGGGTTATG
1331	CGGATGAACT	GCAGATGCTG	GTTCAGCAGT	ATGCCGATGA	TAAAGAAAAA	GTGGCGCTGC	TGAAAGCGCT
1401	GTGGCAGTAT	CGGGAAGAAA	TCGTTTCTGG	CTCTGGTCAC	CATCATCATC	ACCAC	

SEQ ID NO: 32 EcoExo I C51A/C98G/C144T/C306S/C330T (ONLD0453)

1	MMNDGKQST	FLFHDYETFG	THPALDRPAQ	FAAIRTDSEF	NVIGEPEVIFY		
51	AKPADDYLPQ	PGAVLITGIT	PQEARAKGEN	EAAFAARIHS	LFTVPKTGIL		
101	GYNVRFDDDE	VTRNIFYRNF	YDPYAWSQWH	DNSRWDLDDV	MRATYALRPE		
151	GINWPENDDG	LPSFRLEHLT	KANGIEHSNA	HDMADVYAT	IAMAKLVKTR		
201	QPRLFDYLFY	HRNKHKLML	IDVPQMKPLV	HVSGMFGAWR	GNTSWVAPLA		
251	WHPENRNAVY	MVDLAGDISP	LLELSDTLR	ERLYTAKTDL	GDNAAVPVKL		
301	VHINKTPVLA	QANTLRPEDA	DRLGINRQHT	LDNLKILREN	PQVREKVVAI		
351	FAEAEPTFPS	DNVDAQLYNG	FFSDADRAAM	KIVLETEPRN	LPALDITFVD		
401	KRIEKLFFNY	RARNFPGTLD	YAEQQRWLEH	RRQVFTPEFL	QGYADELQML		
451	VQYADDKEK	VALLKALWQY	AEEIVSGSGH	HHHHH			

SEQ ID NO: 33 EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0491)

1	ATGATGAACG	ATGGCAAACA	GCAGAGCACC	TTCTGTFTTC	ATGATTATGA	AACCTTCGGT	ACCCATCCGG
71	CCCTGGATCG	TCCGGCGCAG	TTTGGCGCCA	TTCGCACCGA	TAGCGAATTC	AATGTGATTG	GCGAACCCGGA
141	AGTGTTTTAT	CGGAAACCGG	CCGATGATTA	TCTGCCGCGCAG	CCGGGTGCGG	TGCTGATTAC	CGGTATTACC
211	CCGCAGGAAG	CGCGCGCGAA	AGGTGAAAAC	GAAGCGGCGT	TTGCCCGCGC	CATTCATAGC	CTGTTTACCC
281	TGCCGAAAAC	CGATATTCTG	GGCTATAACA	ATGTGCGCTT	CGATGATGAA	GTTACCCGTA	ATATCTTTTA
351	TCGTAACCTT	TATGATCCGT	ATGCGTGGAG	CTGGCAGCAT	GATAACAGCC	GTTGGGATCT	GCTGGATGTG
421	ATGCGCGCGA	CCTATGCGCT	CGCCCCGAAA	GGCATTAAAT	GGCCGGAAAA	CGATGATGCC	CTGCCGAGCT
491	TTGCTCTGGA	ACATCTGACC	AAAGCCAACG	GCATTGAACA	TAGCAATGCC	CATGATGCGA	TGGCCGATGT
561	TTATGCGACC	ATTGCGATGG	CGAAACTGGT	TAAAACCCGT	CAGCCGCGCC	TGTTTGATTA	TCTGTTTACC
631	CACCGTAACA	AACACAAACT	GATGCGCGTG	ATTGATGTTT	CGCAGATGAA	ACCGCTGGTG	CATGTGAGCC
701	GCAATGTTGG	CGCCTGGCGC	GGCAACACCA	GCTGGGTGGC	CCCGCTGGCC	TGGCACCCGG	AAAAATCGTAA
771	CGCCGTGATT	ATGTTTGATC	TGGCCGGTGA	TATTAGCCCG	CTGCTGGAAC	TGGATAGCGA	TACCCTGCGT

[0009]

841	GAACGCCTGT	ATACCGCCAA	AACCGATCTG	GGCGATAATG	CCGCCGTGCC	GGTGAAACTG	GTTACACATTA
911	ACAAAAGCCC	GGTGCCTGGC	CAGGCGAACA	CCCTGCGCCC	GGAAGATGCG	GATCGTCTGG	GTATTAATCG
981	CCAGCATAAC	CTGGATAATC	TGAAAATCCT	CGGTGAAAAC	CCGCAGGTGC	GTGAAAAAGT	GGTGGCGATC
1051	TTCCGCGAAG	CGGAACCGTT	CACCCGAGC	GATAACGTGG	ATGCGCAGCT	GTATAACGGC	TTCTTTAGCG
1121	ATGCCGATCG	CGCGCGGATG	AAAATCGTTC	TGGAAACCGA	ACCGCGCAAT	CTGCCGGCGC	TGGATATTAC
1191	CTTTGTTGAT	AAACGTATTG	AAAAATGCT	GTTTAATTAT	CGTGCGCGCA	ATTTCCGGG	TACCCCTGGAT
1261	TATGCCGAAC	AGCAGCGTTG	GCTGGAACAT	CGTCGTCAGG	TTTTACCCC	GGAAATTTCTG	CAGGGTTATG
1331	CGGATGAACT	GCAGATGCTG	GTTCAGCAGT	ATGCCGATGA	TAAAGAAAAA	GTGGCGCTGC	TGAAAGCGCT
1401	GTGGCAGTAT	CGGAAGAAA	TCGTTTCTGG	CTCTGGTCAC	CATCATCATC	ACCAC	

SEQ ID NO: 34 EcoExo I C51A/C98T/C144T/C306S/C330T (ONLD0491)

1	MMNDGKQST	FLFHDYETFG	THPALDRPAQ	FAAIRTDSEF	NVIGEPEVIFY		
51	AKPADDYLPQ	PGAVLITGIT	PQEARAKGEN	EAAFAARIHS	LFTVPKTDIL		
101	GYNVRFDDDE	VTRNIFYRNF	YDPYAWSQWH	DNSRWDLDDV	MRATYALRPE		
151	GINWPENDDG	LPSFRLEHLT	KANGIEHSNA	HDMADVYAT	IAMAKLVKTR		
201	QPRLFYDLFT	HRNKHKLML	IDVPQMKPLV	HVSGMFGAWR	GNTSWVAPLA		
251	WHPENRNAV	MVDLAGDISP	LLELSDTLR	ERLYTAKTDL	GDNAAVPVKL		
301	VHINKSPVLA	QANTLRPEDA	DRLGINRQHT	LDNLKILREN	PQVREKVVAI		
351	FAEAEPFPTS	DNVDAQLYNG	FFSDADRAAM	KIVLETEPRN	LPALDITFVD		
401	KRIEKLIFNY	RARNFPGTLD	YAEQQRWLEH	RRQVFTPEFL	QGYADELQML		
451	VQYADDKEK	VALLKALWQY	AEEIVSGSGH	HHHHH			

SEQ ID NO: 35 EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454)

1	ATGATGAACG	ATGGCAAACA	GCAGAGCACC	TTCCTGTTTC	ATGATTATGA	AACCTTCGGT	ACCCATCCGG
71	CCCTGGATCG	TCCGGCGCAG	TTTGGCGCCA	TTCGCACCGA	TAGCGAATTC	AATGTGATTG	GCGAACCGGA
141	AGTGTTTTAT	GCGAAACCGG	CCGATGATTA	TCTGCCGCG	CCGGGTGCGG	TGCTGATTAC	CGGTATTACC
211	CCGCAGGAAG	CGCGCGCGAA	AGGTGAAAAC	GAAGCGGCGT	TTGCCCGCGC	CATTCATAGC	CTGTTTACCG
281	TGCCGAAAAC	CAAAATTCG	GGCTATAACA	ATGTGCGCTT	CGATGATGAA	GTTACCCGTA	ATATCTTTTA
351	TCGTAACCTT	TATGATCCGT	ATGCGTGGAG	CTGGCAGCAT	GATAACAGCC	GTTGGGATCT	GCTGGATGTG
421	ATGCGCGCGA	CCTATGCGCT	CGCCCCGAA	GGCATTAAAT	GGCCGGAAAA	CGATGATGGC	CTGCCGAGCT
491	TTGCTCTGGA	ACATCTGACC	AAAGCCAACG	GCATTGAACA	TAGCAATGCC	CATGATGCGA	TGGCCGATGT
561	TTATGCGACC	ATTGCGATGG	CGAAACTGGT	TAAAACCCGT	CAGCCGCGCC	TGTTTGATTA	TCTGTTTACC
631	CACCGTAACA	AACACAAACT	GATGGCGCTG	ATTGATGTTT	CGCAGATGAA	ACCGCTGGTG	CATGTGAGCG
701	GCATGTTTGG	CGCCTGGCGC	GGCAACACCA	GCTGGGTGGC	CCCGCTGGCC	TGGCACCCTGG	AAAAATCGTAA
771	CGCCGTGATT	ATGGTTGATC	TGGCCGGTGA	TATTAGCCCG	CTGCTGGAAC	TGGATAGCGA	TACCCTCGCT
841	GAACGCCCTG	ATACCGCCAA	AACCGATCTG	GGCGATAATG	CCGCCGTGCC	GGTGAAACTG	GTTACACATTA
911	ACAAAAGCCC	GGTGCCTGGC	CAGGCGAACA	CCCTGCGCCC	GGAAGATGCG	GATCGTCTGG	GTATTAATCG
981	CCAGCATAAC	CTGGATAATC	TGAAAATCCT	CGGTGAAAAC	CCGCAGGTGC	GTGAAAAAGT	GGTGGCGATC
1051	TTCCGCGAAG	CGGAACCGTT	CACCCGAGC	GATAACGTGG	ATGCGCAGCT	GTATAACGGC	TTCTTTAGCG
1121	ATGCCGATCG	CGCGCGGATG	AAAATCGTTC	TGGAAACCGA	ACCGCGCAAT	CTGCCGGCGC	TGGATATTAC
1191	CTTTGTTGAT	AAACGTATTG	AAAAATGCT	GTTTAATTAT	CGTGCGCGCA	ATTTCCGGG	TACCCCTGGAT
1261	TATGCCGAAC	AGCAGCGTTG	GCTGGAACAT	CGTCGTCAGG	TTTTACCCC	GGAAATTTCTG	CAGGGTTATG
1331	CGGATGAACT	GCAGATGCTG	GTTCAGCAGT	ATGCCGATGA	TAAAGAAAAA	GTGGCGCTGC	TGAAAGCGCT
1401	GTGGCAGTAT	CGGAAGAAA	TCGTTTCTGG	CTCTGGTCAC	CATCATCATC	ACCAC	

SEQ ID NO: 36 EcoExo I C51A/C98K/C144T/C306S/C330T (ONLD0454)

1	MMNDGKQST	FLFHDYETFG	THPALDRPAQ	FAAIRTDSEF	NVIGEPEVIFY		
51	AKPADDYLPQ	PGAVLITGIT	PQEARAKGEN	EAAFAARIHS	LFTVPKTKIL		
101	GYNVRFDDDE	VTRNIFYRNF	YDPYAWSQWH	DNSRWDLDDV	MRATYALRPE		
151	GINWPENDDG	LPSFRLEHLT	KANGIEHSNA	HDMADVYAT	IAMAKLVKTR		
201	QPRLFYDLFT	HRNKHKLML	IDVPQMKPLV	HVSGMFGAWR	GNTSWVAPLA		
251	WHPENRNAV	MVDLAGDISP	LLELSDTLR	ERLYTAKTDL	GDNAAVPVKL		
301	VHINKSPVLA	QANTLRPEDA	DRLGINRQHT	LDNLKILREN	PQVREKVVAI		
351	FAEAEPFPTS	DNVDAQLYNG	FFSDADRAAM	KIVLETEPRN	LPALDITFVD		
401	KRIEKLIFNY	RARNFPGTLD	YAEQQRWLEH	RRQVFTPEFL	QGYADELQML		
451	VQYADDKEK	VALLKALWQY	AEEIVSGSGH	HHHHH			

Seq ID NO: 37 EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455)

1	ATGATGAACG	ATGGCAAACA	GCAGAGCACC	TTCCTGTTTC	ATGATTATGA	AACCTTCGGT	ACCCATCCGG
71	CCCTGGATCG	TCCGGCGCAG	TTTGGCGCCA	TTCGCACCGA	TAGCGAATTC	AATGTGATTG	GCGAACCGGA
141	AGTGTTTTAT	GCGAAACCGG	CCGATGATTA	TCTGCCGCG	CCGGGTGCGG	TGCTGATTAC	CGGTATTACC
211	CCGCAGGAAG	CGCGCGCGAA	AGGTGAAAAC	GAAGCGGCGT	TTGCCCGCGC	CATTCATAGC	CTGTTTACCG
281	TGCCGAAAAC	CCTGATTCG	GGCTATAACA	ATGTGCGCTT	CGATGATGAA	GTTACCCGTA	ATATCTTTTA
351	TCGTAACCTT	TATGATCCGT	ATGCGTGGAG	CTGGCAGCAT	GATAACAGCC	GTTGGGATCT	GCTGGATGTG
421	ATGCGCGCGA	CCTATGCGCT	CGCCCCGAA	GGCATTAAAT	GGCCGGAAAA	CGATGATGGC	CTGCCGAGCT
491	TTGCTCTGGA	ACATCTGACC	AAAGCCAACG	GCATTGAACA	TAGCAATGCC	CATGATGCGA	TGGCCGATGT
561	TTATGCGACC	ATTGCGATGG	CGAAACTGGT	TAAAACCCGT	CAGCCGCGCC	TGTTTGATTA	TCTGTTTACC
631	CACCGTAACA	AACACAAACT	GATGGCGCTG	ATTGATGTTT	CGCAGATGAA	ACCGCTGGTG	CATGTGAGCG

[0010]

```

701 GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GGCATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAAAGCCC GGTGCTGGCC CAGGCCAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATAACC CTGGATAATC TGAATACTCT GCGTGAATAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGGCGC TGGATATTAC
1191 CTTTGTGAT AAACGTATTG AAAAATCGCT GTTTAATFAT CGTGGCGCA ATTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

Seq ID NO: 38 EcoExo I C51A/C98L/C144T/C306S/C330T (ONLD0455)

```

1 MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVYF
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTLIL
101 GYNNVRFDE VTRNIFYRNF YDPYAWSQWH DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201 QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAT
351 FAEAEPFTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 39 EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456)

```

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCTCGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCGGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCG CCGGGTGGCG TGCTGATTAC CGGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCGCGCG CATTATAGC CTGTTTACCG
281 TGCCGAAAAC CGTGATTCTG GGTATAACA ATGTGCGCTT CGATGATGAA GTTACCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCCGCGCA CCTATGCGCT GCGCCCGGAA GGCAATTAAT GGCCGGAAAA CGATGATGCG CTGCCAGACT
491 TTCGTCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGTT TAAAACCCGT CAGCCCGGCC TGTGTTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTAGCCG
701 GCAIGTITGG CGCCTGGCGC GGCAACACCA CCGGGTGGC CCGCCGTGCC TGGCAACCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GGCATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAAAGCCC GGTGCTGGCC CAGGCCAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATAACC CTGGATAATC TGAATACTCT GCGTGAATAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGGCGC TGGATATTAC
1191 CTTTGTGAT AAACGTATTG AAAAATCGCT GTTTAATFAT CGTGGCGCA ATTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 40 EcoExo I C51A/C98V/C144T/C306S/C330T (ONLD0456)

```

1 MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVYF
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTVIL
101 GYNNVRFDE VTRNIFYRNF YDPYAWSQWH DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201 QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVPVKL
301 VHINKSPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAT
351 FAEAEPFTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 41 EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476)

```

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCTCGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCGGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCG CCGGGTGGCG TGCTGATTAC CGGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCGCGCG CATTATAGC CTGTTTACCG
281 TGCCGAAAAC CGTGATTCTG GGTATAACA ATGTGCGCTT CGATGATGAA GTTACCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCCGCGCA CCTATGCGCT GCGCCCGGAA GGCAATTAAT GGCCGGAAAA CGATGATGCG CTGCCAGACT
491 TTCGTCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT

```

[0011]

```

561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGCC TGTGTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGCGCGTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTTGG CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTACACATTA
911 ACAAAAACCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGGATAATC TGAATACTCT CCGTGAATAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGGATC
1051 TTCGCGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGTTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGGCGCA ATTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAA GTGGCGCTGC TGAAGCGCT
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 42 EcoExo I C51A/C98V/C144T/C306T/C330T (ONLD0476)

```

1 MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVIFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTVIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSQW DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELDSOTLR ERLYAKTDL GDNAAVPVKL
301 VHINKPVL A QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEFP T PS DNVDAQLYNG FFSADADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQQYADDKE K VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 43 EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477)

```

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGGCG TGCTGATTAC CGGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAA GAAGCGGCGT TTGCCGCGCG CATTCAATAGC CTGTTTACCG
281 TGCCGAAAA C CACCATCTCG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GGCCGAAAAA CGATGATGCG CTGCCGAGT
491 TTCGCTGGA ACATCTGACC AAAGCCCAAG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561 TTAGCGGACC ATTGGCATGG CGAAACTGGT TAAAACCCGT CAGCCCGGCC TGTGTGATTA TCTGTTTACC
631 CACCGTAACA AACACAAACT GATGCGCGTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTTGG CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTACACATTA
911 ACAAAAAGCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGGATAATC TGAATACTCT CCGTGAATAAC CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTCGCGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CGCGGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTGTTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGGCGCA ATTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAACT GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAA GTGGCGCTGC TGAAGCGCT
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 44 EcoExo I C51A/C98T/C144T/C306M/C330T (ONLD0477)

```

1 MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRTDSEF NVIGEPEVIFY
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTTIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSQW DNSRWDLDDV MRATYALRPE
151 GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELDSOTLR ERLYAKTDL GDNAAVPVKL
301 VHINKMPVL A QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEFP T PS DNVDAQLYNG FFSADADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQQYADDKE K VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 45 EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478)

```

1 ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71 CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG GCGAACCCGA
141 AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGGCG TGCTGATTAC CGGTATTACC
211 CCGCAGGAAG CGCGCGCGAA AGGTGAAAAA GAAGCGGCGT TTGCCGCGCG CATTCAATAGC CTGTTTACCG
281 TGCCGAAAA C CACCATCTCG GGCTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG

```

[0012]

```

421   ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GCGCGGAAAA CGATGATGGC CTGCCGAGCT
491   TTGCTCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561   TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGCC TGTTTGATTA TCTGTTTACC
631   CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701   GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771   CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841   GAAACGCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911   AAAAAAACC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981   CCAGCATACC CTGATAATC TAAAAATCCT GCGTAAAAAC CCGCAGGTGC GTAAAAAAGT GGTGGCGATC
1051  TTGCGGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121  ATGCCGATCG CGCGCGCATG AAAATCGTTC TGGAAACCGA ACCCGCAAT CTGCCGCGC TGGATATTAC
1191  CTTGTTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGGCGCA ATTTCCGGG TACCCTGGAT
1261  TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTACCCC GGAATTTCTG CAGGGTTATG
1331  CGGATGAAC GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401  GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 46 EcoExo I C51A/C98T/C144T/C306N/C330T (ONLD0478)

```

1     MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRDSEF NVIGEPEVfy
51    AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTTIL
101   GYNNVRFDEE VTRNIFYRNF YDPYAWSWQH DNSRWDL.LDV MRATYALRPE
151   GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201   QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251   WHPENRNAV I MVDLAGDISP LLELDSDTLR ERLYTAKTDI GDNAAVPVKL
301   VHINKNPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351   FAEAEPTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401   KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451   VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 47 EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479)

```

1     ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71    CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG CGGAACCGGA
141   AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGGCG TCGTATTAC CCGTATTACC
211   CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCGCGCG CATTCAATAGC CTGTTTACCG
281   TGCCGAAAAC CACCATTCTG GGTATAAACA ATGTGCGCTT CGATGATGAA GTTACCCTGA ATATCTTTTA
351   TCGTAACCTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT CTGGGATGTG
421   ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAATT GCGCGGAAAA CGATGATGCG CATGCCGAGCT
491   TTGCTCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCGA TGGCCGATGT
561   TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGCC TGTTTGATTA TCTGTTTACC
631   CACCGTAACA AACACAAACT GATGGCGCTG ATTGATGTTT CGCAGATGAA ACCGCTGGTG CATGTGAGCG
701   GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCCGCTGGCC TGGCACCCGG AAAATCGTAA
771   CGCCGTGATT ATGGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841   GAACGCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911   ACAAAGATCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981   CCAGCATACC CTGATAATC TAAAAATCCT GCGTAAAAAC CCGCAGGTGC GTAAAAAAGT GGTGGCGATC
1051  TTGCGGGAAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121  ATGCCGATCG CGCGCGCATG AAAATCGTTC TGGAAACCGA ACCCGCAAT CTGCCGCGC TGGATATTAC
1191  CTTGTTGAT AAACGTATTG AAAAATGCT GTTTAATTAT CGTGGCGCA ATTTCCGGG TACCCTGGAT
1261  TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTACCCC GGAATTTCTG CAGGGTTATG
1331  CGGATGAAC GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401  GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGGTCAC CATCATCATC ACCAC

```

SEQ ID NO: 48 EcoExo I C51A/C98T/C144T/C306D/C330T (ONLD0479)

```

1     MMNDGKQST FLFHDYETFG THPALDRPAQ FAAIRDSEF NVIGEPEVfy
51    AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTTIL
101   GYNNVRFDEE VTRNIFYRNF YDPYAWSWQH DNSRWDL.LDV MRATYALRPE
151   GINWPENDDG LPSFRLEHLT KANGIEHSNA HDAMADVYAT IMAKLVKTR
201   QPRLFDYLFY HRNKHKLML IDVPQMKPLV HVSGMFGAWR GNTSWVAPLA
251   WHPENRNAV I MVDLAGDISP LLELDSDTLR ERLYTAKTDI GDNAAVPVKL
301   VHINKNPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351   FAEAEPTPS DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401   KRIEKLLFNY RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451   VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 49 EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480)

```

1     ATGATGAACG ATGGCAAACA GCAGAGCACC TTCCTGTTTC ATGATTATGA AACCTTCGGT ACCCATCCGG
71    CCCTGGATCG TCCGGCGCAG TTTGCGGCCA TTCGCACCGA TAGCGAATTC AATGTGATTG CGGAACCGGA
141   AGTGTTTTAT GCGAAACCGG CCGATGATTA TCTGCCGCGC CCGGGTGGCG TCGTATTAC CCGTATTACC
211   CCGCAGGAAG CGCGCGCGAA AGGTGAAAAC GAAGCGGCGT TTGCCGCGCG CATTCAATAGC CTGTTTACCG

```

[0013]

```

281 TGCCGAAAAC CACCATTCTG GGTATAACA ATGTGCGCTT CGATGATGAA GTTACCCGTA ATATCTTTTA
351 TCGTAACTTT TATGATCCGT ATGCGTGGAG CTGGCAGCAT GATAACAGCC GTTGGGATCT GCTGGATGTG
421 ATGCGCGCGA CCTATGCGCT GCGCCCGGAA GGCATTAAAT GGCCGGAAAA CGATGATGGC CTGCCGAGCT
491 TTCGCTGGA ACATCTGACC AAAGCCAACG GCATTGAACA TAGCAATGCC CATGATGCCA TGGCCGATGT
561 TTATGCGACC ATTGCGATGG CGAAACTGGT TAAAACCCGT CAGCCGCGCC TGTITGATTA TCTGTTTACC
631 CACCGTAAAC AACACAAACT GATGCGCGTG ATTGATGTTT CCGAGATGAA ACCGCTGGTG CATGTGAGCG
701 GCATGTTTGG CGCCTGGCGC GGCAACACCA GCTGGGTGGC CCGCTGGCC TGGCACCCGG AAAATCGTAA
771 GCGCGTGATT ATGTTGATC TGGCCGGTGA TATTAGCCCG CTGCTGGAAC TGGATAGCGA TACCCTGCGT
841 GAACGCCTGT ATACCGCCAA AACCGATCTG GCGGATAATG CCGCCGTGCC GGTGAAACTG GTTCACATTA
911 ACAAAGCGCC GGTGCTGGCC CAGGCGAACA CCCTGCGCCC GGAAGATGCG GATCGTCTGG GTATTAATCG
981 CCAGCATACC CTGGATAATC TGAATACTCT GCGTGAATAA CCGCAGGTGC GTGAAAAAGT GGTGGCGATC
1051 TTCGCGGAG CGGAACCGTT CACCCCGAGC GATAACGTGG ATGCGCAGCT GTATAACGGC TTCTTTAGCG
1121 ATGCCGATCG CCGCGCGATG AAAATCGTTC TGGAAACCGA ACCGCGCAAT CTGCCGCGC TGGATATTAC
1191 CTTTGTGAT AAACGTATTG AAAAAGTCTT GTTTAAATAT CGTGGCGCGA ATTTCCGGG TACCCTGGAT
1261 TATGCCGAAC AGCAGCGTTG GCTGGAACAT CGTCGTCAGG TTTTCACCCC GGAATTTCTG CAGGGTTATG
1331 CGGATGAAC GCAGATGCTG GTTCAGCAGT ATGCCGATGA TAAAGAAAAA GTGGCGCTGC TGAAGCGCT
1401 GTGGCAGTAT GCGGAAGAAA TCGTTTCTGG CTCTGCTCAC CATCATCATC ACCAC

```

SEQ ID NO: 50 EcoExo I C51A/C98T/C144T/C306A/C330T (ONLD0480)

```

1 MMNDGKQST FLFHDYFTFG THPALDRPAQ FAARTDSEF NVIGEPEVYF
51 AKPADDYLPQ PGAVLITGIT PQEARAKGEN EAAFAARIHS LFTVPKTTIL
101 GYNNVRFDD E VTRNIFYRNF YDPYAWSQW DNSRWDLDDV MRATYALRPE
151 GYNWPFDDG LPSFRLFHLI KANGIEHSNA HDMADVYAT TAMAKL VKTR
201 QPRLFDYLF T HRNKHKLML IDVPQMKPLV HVGSMFGAWR GNTSWVAPLA
251 WHPENRNAV I MVDLAGDISP LLELSDTLR ERLYAKTDL GDNAAVVPKL
301 VHINKAPVLA QANTLRPEDA DRLGINRQHT LDNLKILREN PQVREKVVAI
351 FAEAEPTFP S DNVDAQLYNG FFSADRAAM KIVLETEPRN LPALDITFVD
401 KRIEKLLFN Y RARNFPGTLD YAEQQRWLEH RRQVFTPEFL QGYADELQML
451 VQYADDKEK VALLKALWQY AEEIVSGSGH HHHHH

```

SEQ ID NO: 51

```

1 ATGAAATTTG TCTCTTTAA TATCAACGGC CTGCGCGCCA GACCTACCA GCTTGAAGCC ATCGTCGAAA
71 AGCACCAACC GGATGTGATT GGCCTGCAGG AGACAAAAGT TCATGACGAT ATGTTTCCGC TCGAAGAGGT
141 GCGAAGCTC GGTACAACG TGTTTATCA CCGGCAGAAA GGCATTATG CCGTGGCGT GCTGACAAA
211 GAGACCGCGA TTGCGTGGC TCGCGCTTT CCGGTGACG ACGAAGAGG CCGACGCGG ATTATTATG
281 CGGAAATCCC CTCACCTGCT GGTAAATGCA CCGTGAACA CGGTTACTIC CCGCAGGGTG AAAAGCCGCA
351 CCATCCGATA AAATCCCGG CAAAAGCGCA GTTTTATCAG AATCTGCAA ACTACCTGGA AACCGAATC
421 AAACGTGATA ATCCGGTACT GATTATGGG GATATGAATA TCAGCCCTAC AGATCTGGAT ATCGGCATTG
491 GCGAAGAAAA CCGTAAGCGC TGGCTGCGTA CCGGTAATG CTCTTTCCTG CCGGAAGAGC GCGAATGGAT
561 GGACAGGCTG ATGAGCTGG GGTGGTCCA TACCTTCCGC CATGCGAATC CGCAAACAGC AGATCGTTTC
631 TCATGGTTG ATTACCGCTC AAAAGTTTT GACGATAACC GTGGTCTGG CATCGACCTG CTGCTGCGCA
701 GCCAACCGCT GGCAGAAATG TCGTAGAAA CCGGCATCGA CTATGAAATC CGCAGCATGG AAAAACCGTC
771 CGATCACGCC CCCGTCTGGG CGACCTTCCG CCGC

```

SEQ ID NO: 52

```

1 MKFVSFNIG LRARPHQLEA IVEKHQPDVI GLQETKVHDD MFPLEEVAKL GYNVFIHGQK GHYGVALLTK
71 ETPIAVRRGF PGDDEEAQRR IIMAEIPSLN GNVTVINGYF PQGESRDHPI KFPKAQFYQ NLQNYLETEL
141 KRDNPVLMG DMNISPTDLD IGIGEENRRK WLRTGKCSFL PEEREWMDRL MSWGLVDFR HANPQTADRF
211 SWFDYRSKGF DDNRGLRIDL LLASQPLAEC CVETGIDYEI RSMKPSDHA PVWATFRR

```

SEQ ID NO: 53

```

1 ATGTTTCGTC GTAAAGAAGA TCTGGATCCG CCGCTGGCAC TGCTGCCGCT GAAAGGCTG CCGGAAGCCG
71 CCGCACTGCT GGAAGAAGCG CTGCGTCAAG GTAAACGCAT TCGTGTTCAC GCGCACTATG ATGCGGATGG
141 CCTGACCGGC ACCCGCATCC TGGTTCGTGG TCTGGCCGCC CTGGGTGCGG ATGTTTATCCG GTTTATCCCG
211 CACCGCCTGG AAGAAGGCTA TGGTGTCTG ATGGAACGCG TCCCGGAAAC TCTGGAAGCC TCGGACCTGT
281 TTCTGACCGT TGA CTGCGGC ATTACCAACC ATGCGGAATC CCGCGAACTG CTGGAATAAT CCGTGGAAAT
351 CATTGTTACC GATCATCATA CGCCGGGCAA AACGCGCGCG CCGGGTCTGG TCGTGCATCC GCGCTGACG
421 CCGGATCTGA AAAAAAACC GACCGCGCA GCGTGGCGT TCTGCTGCT GTGGGCACTG CATGAACGCC
491 TGGCCCTGCC GCGCCGCTG GAATACGCG ACCTGGCAGC CGTTGGCACC ATTTCCGACG TTGCCCCGCT
561 GTGGGTTGG AATCGTGCAC TGGTGAAGA AGGTCTGGCA CGCATCCCGG CTTCATCTTG GGTGGGCTG
631 CGTCTGCTGG CTGAAGCCGT GGGCTATACC GGCAAAAGCG TCGAAGTCG TTTCCGCATC GCGCCGCGCA
701 TCAATGCGGC TTCCCGCTG GCGGAAGCG AAAAAGCCCT GCGCCTGCTG CTGACGGATG ATGCGGCGA
771 AGCTCAGGCG CTGGTCGGC AACTGCACC TCTGAACGCC CGTCTGTCAG CCCTGGAAGA AGCGATGCTG
841 CGCAAACCTG TCCCGCAGGC CGAACCGGAA CCGAAAAGCCA TCGTTCTGCT GGACCCGAAA GGCATCCCG
911 GTGTTATGG TATGTGGCC TCTGCATCC TGAAGCGAC CCTGCGCCCG GTCTTTCTG TGGCCAGGG
981 CAAAGGCACC GTGCGTTCG TGGCTCCGAT TTCGCGCTC GAAGCACTG CAGCGCGGA AGATCTGCTG
1051 CTGCGTTATG GTGGTCATAA AGAAGCGCG GGTTCGCAA TGGATGAAG CCGTGTTCGG CCGTTCAAAG
1121 CACGCGTTGA AGCGTATGCC GCACGTTTCC CCGATCCGGT TCGTGAAGT GCACTGCTGG ATCTGCTGCC

```

[0014]

1191 GGAACCGGGC CTGCTGCCG AGGTGTTCCG TGAAGTGGCA CTGCTGGAAC CGTATGGTGA AGGTAACCCG
1261 GAACCGCTGT TCCTG

SEQ ID NO: 54

1 MFRKEDLDP PLALLPLKGL REAAALLEE LRGKRIRVH GDYDADGLTG TAILVRGLAA LGADVHPFIP
71 HRLEEGYGV L MERVPEHLEA SDLFLTVDCC ITNHAELREL LENGVEVIVT DHHTPGKTPP PGLVVHPALT
141 PDLKEKPTGA GVAFLLWAL HERLGLPPPL EYADLAAVGT IADVAPLWGW NRALVKEGLA RIPASSWVGL
211 RLLAEAVGYT GKAVEVAFRI APRINAASRL GEAEKALRLL LTDDAAEAQA LVGELHRLNA RRQTLEEAML
281 RKLLPQADPE AKAIIVLLDPE GHPGVMGIVA SRILEATLRP VFLVAQGGKT VRSLAPISAV EALRSAEDLL
351 LRYGGHKEAA GFAMDEALFP AFKARVEAYA ARFPDPVREV ALLDLLPEPG LLPQVRELA LLEPYGEGNP
421 EPLFL

SEQ ID NO: 55

1 TCCGGAAGCG GCTCTGGTAG TGGTTCGGC ATGACACCGG ACATTATCCT GCAGCGTACC GGGATCGATG
71 TGAGAGCTGT CGAACAGGGG GATGATGCGT GGCACAAATT ACGGCTCGGC GTCATCACCG CTTCAGAAGT
141 TCACAACGTG ATAGCAAAAC CCCGCTCCGG AAAGAAGTGG CCTGACATGA AAATGTCCTA CTCCACACC
211 CTGCTTGGT AGGTTTGCAC CGGTGTGGCT CCGGAAGTTA ACGCTAAAGC ACTGGCCTGG GAAAAACAGT
281 ACGAGAACGA CGCCAGAACC CTGTTTGAAT TCACTCCGG CGTGAATGTT ACTGAATCCC CGATCATCTA
351 TCGCGACGAA AGTATGCGTA CCGCCTGCTC TCCCGATGGT TTATGCAGTG ACGGCAACGG CCTTGAACCT
421 AAATGCCCGT TTACCTCCCG GGATTTATG AAGTTCGGC TCGGTGGTTT CGAGGCCATA AAGTCAGCTT
491 ACATGGCCCA GGTGCAGTAC AGCATGTGGG TGACGCGAAA AAATGCCTGG TACTTTGCCA ACTATGACCC
561 GCGTATGAAG CGTGAAGGCC TGCATTATGT CGTGATTGAG CGGGATGAAA AGTACATGGC GAGTTTGGC
631 GAGATCGTGC CCGAGTTCAT CGAAAAATG GACGAGGCAC TGGCTGAAAT TGGTTTGTGTA TTTGGGGAGC
701 AATGGCGATC TGGCTCTGGT TCCGGCAGCG GTTCCGGA

SEQ ID NO: 56

1 MTPDIILQRT GIDVRAVEQG DDAWHKRLG VITASEVHV IAKPRSGKKW PDMKMSYFHT LLAEVCTGVA
71 PEVNAKALAW GKQYENDART LFEFTSGVNV TESPIIYRDE SMRTACSPDG LCSDBGLEL KCPFTSRDFM
141 KFRLGGFEAI KSAYMAQVQY SMWVTRKNAW YFANYDPRMK REGLHYVVIE RDEKYMASF D EIVPEFIEKM
211 DEALAEIGFV FGEQWR

SEQ ID NO: 57

TGTGTTCTATGTCTTATTCTTACTTCGTTATTCTTGTCTCTATTCTGTTTATGTTTCTGTTTGTIA

SEQ ID NO: 58

TGTGTTCTATGTCTTTT - (CH2)4 - MAL

SEQ ID NO: 59

TGTGTTCTATGTCTTATTCTTACTTTT - (CH2)4

SEQ ID NO: 60

TGTGTTCTATGTCTTATTCTTACTTCGTTATTCTTTT - (CH2)4 - MAL

SEQ ID NO: 61

AAGACATAGAACACATT - (CH2)4 - MAL

SEQ ID NO: 62

AAGTAAGAATAAGACATAGAACACATT - (CH2)4 - MAL

SEQ ID NO: 63

AAGAATAACGAAGTAAGAATAAGACATAGAACACATT - (CH2)4 - MAL

SEQ ID NO: 64

AAGAATAACGAAGTAAGAATAAGACATAGAACACATTTT - (CH2)3-THIOL - 3'

SEQ ID NO: 65

THIOL - (CH2)6 - TGTGTTCTATGTCTTATTCTTACTTCGTTATTCTT

SEQ ID NO: 66

THIOL - CCTAGTCTCCGXAGC where X is U base modified with azide group

[0015]

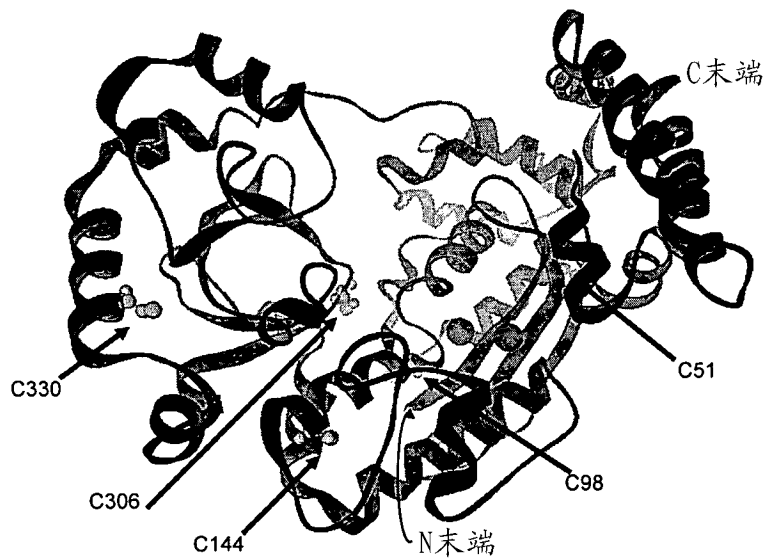


图 1

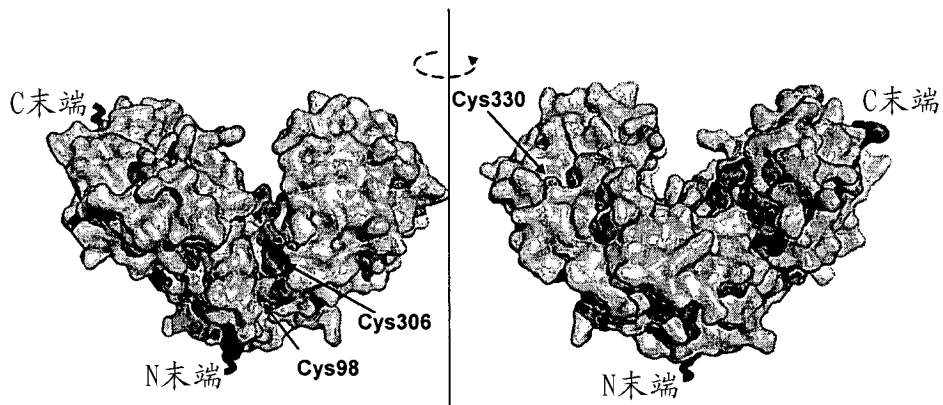


图 2

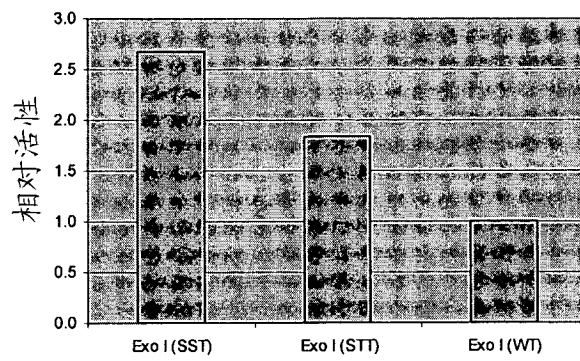


图 3

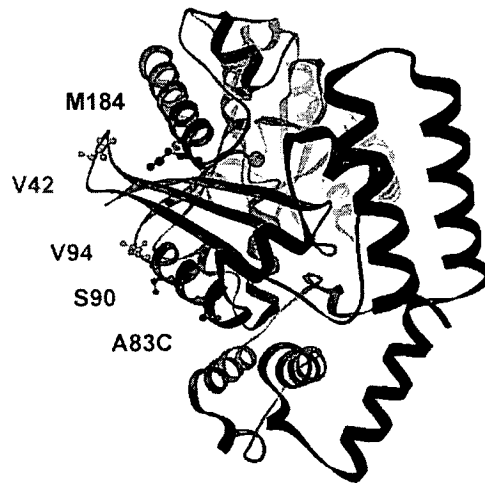


图 4

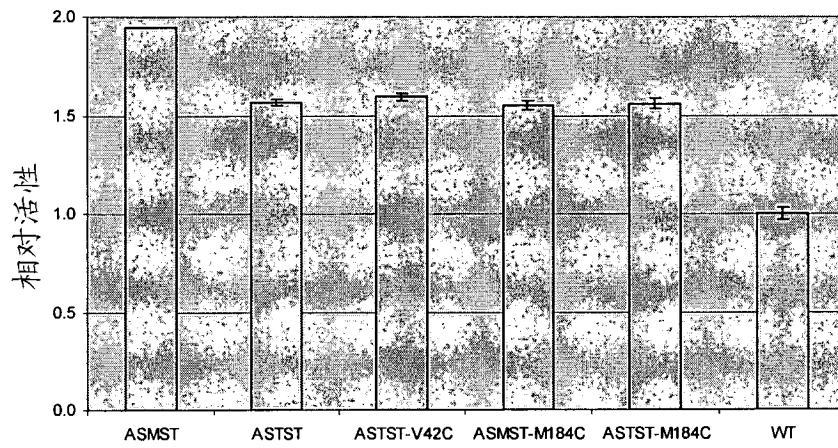


图 5a

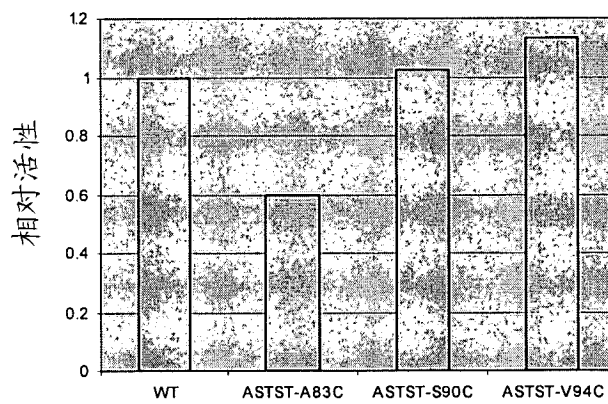


图 5b

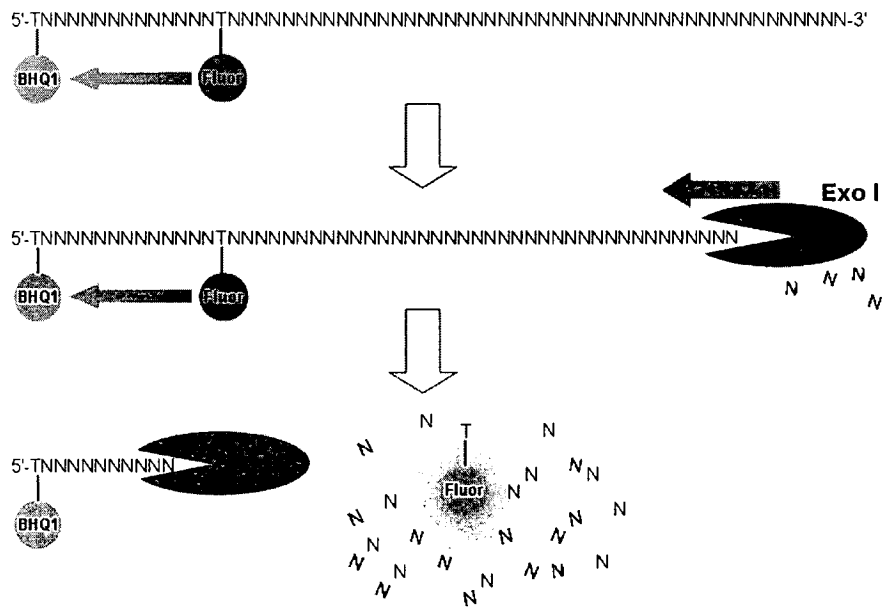


图 6

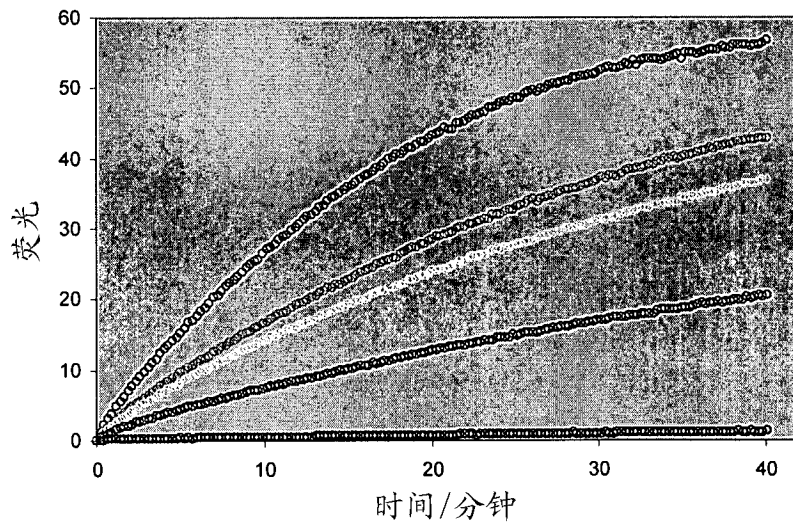


图 7

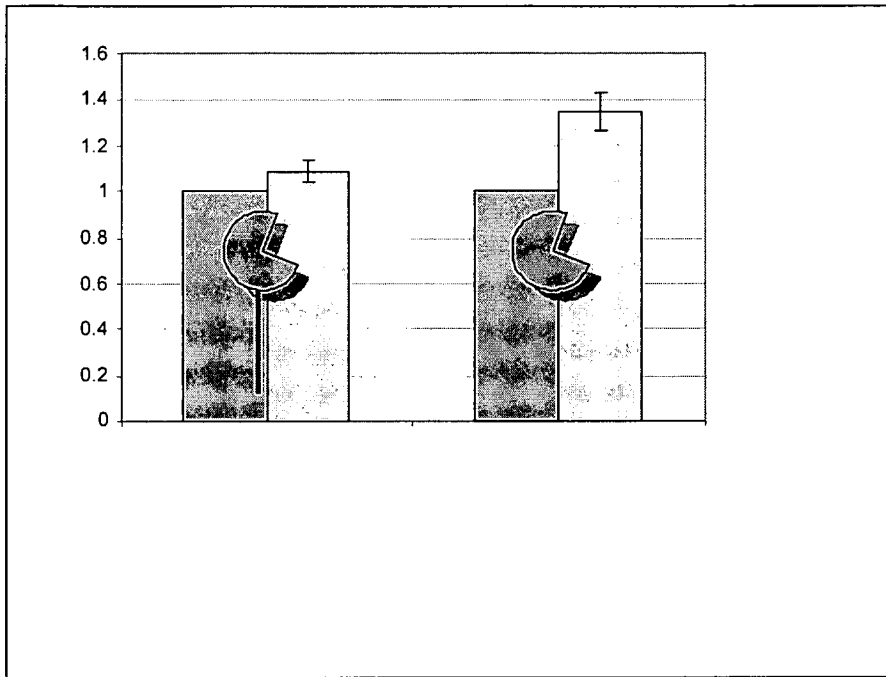


图 8

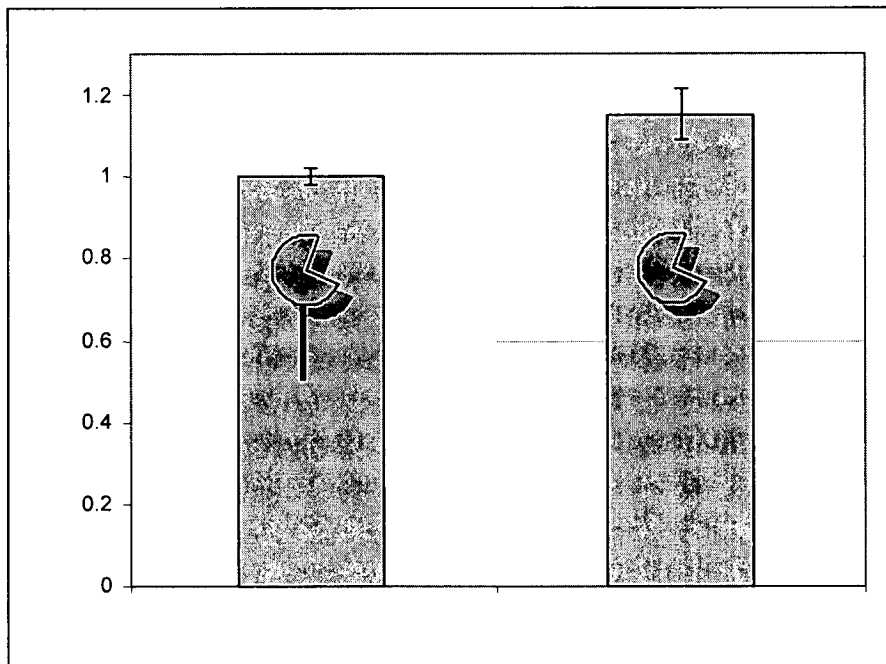


图 9

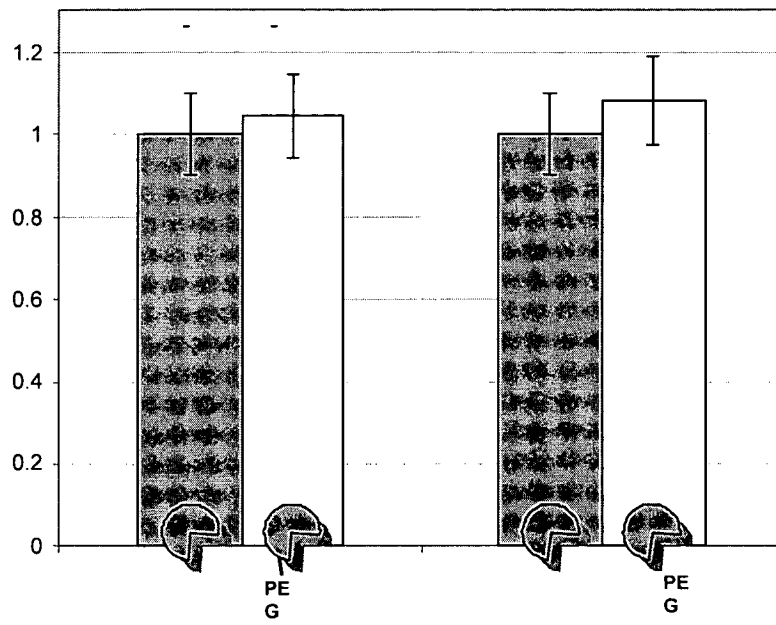


图 10

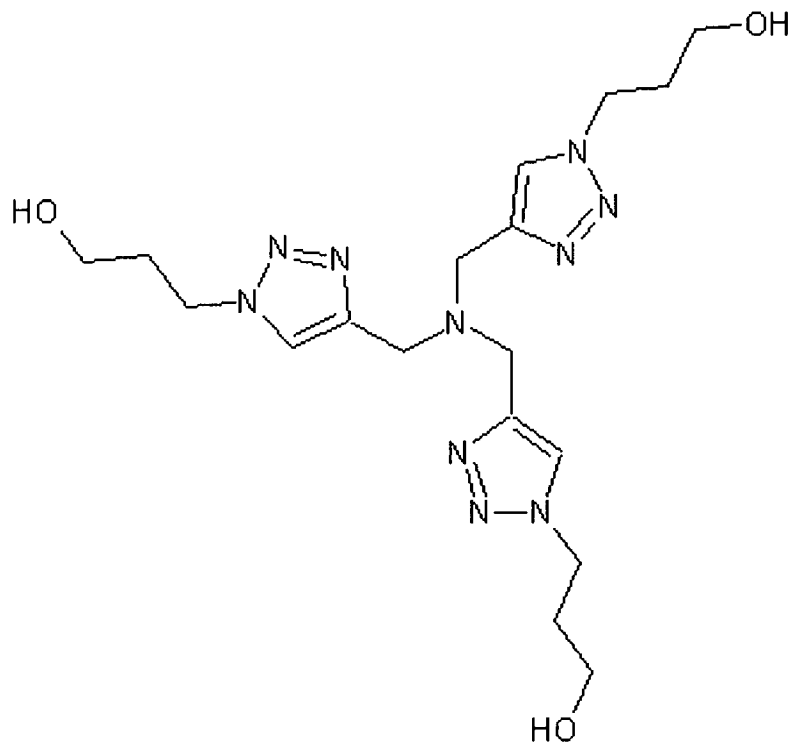


图 11