



(12) 发明专利

(10) 授权公告号 CN 107109483 B

(45) 授权公告日 2021. 07. 20

(21) 申请号 201580062575.9

(22) 申请日 2015.02.18

(65) 同一申请的已公布的文献号
申请公布号 CN 107109483 A

(43) 申请公布日 2017.08.29

(30) 优先权数据
1418159.8 2014.10.14 GB

(85) PCT国际申请进入国家阶段日
2017.05.18

(86) PCT国际申请的申请数据
PCT/GB2015/050461 2015.02.18

(87) PCT国际申请的公布数据
W02016/059363 EN 2016.04.21

(73) 专利权人 牛津纳米孔技术公司
地址 英国牛津郡

(72) 发明人 詹姆斯·怀特

(74) 专利代理机构 北京同立钧成知识产权代理有限公司 11205

代理人 李艳 臧建明

(51) Int.Cl.
C12Q 1/6806 (2018.01)
C12Q 1/6869 (2018.01)

(56) 对比文件
WO 2013185137 A1, 2013.12.12
WO 2013014451 A1, 2013.01.31
CN 102369298 A, 2012.03.07
CN 103827320 A, 2014.05.28
CN 102245760 A, 2011.11.16
CN 105705656 A, 2016.06.22
Nicholas J. Heredia et al..In vitro
double transposition for DNA
identification.《Analytical Biochemistry》
.2009,第399卷

审查员 李煜

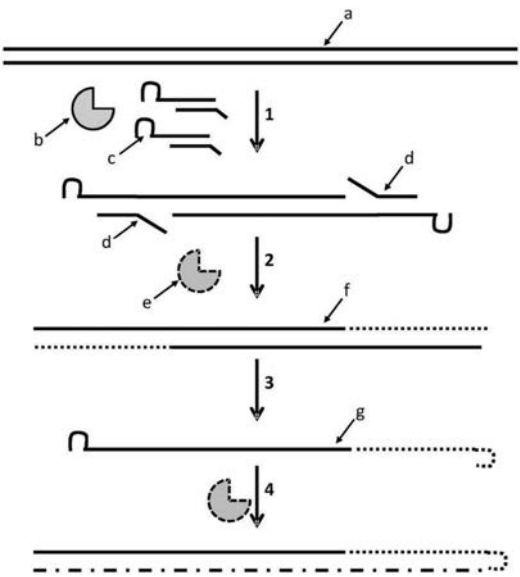
权利要求书2页 说明书29页
序列表63页 附图8页

(54) 发明名称

一种修饰模板双链多核苷酸的方法

(57) 摘要

本发明涉及一种修饰模板双链多核苷酸的方法,特别是使用纳米孔测序进行表征的方法。该方法从模板产生多个经修饰的双链多核苷酸。然后可以表征这些经修饰的多核苷酸。



1. 一种修饰模板双链多核苷酸的方法,包括:

(a) 使模板多核苷酸与MuA转座酶以及一群双链MuA底物接触,每个双链MuA底物包括(i) 至少一个突出端和(ii) 与包括所述至少一个突出端的链相对的链中的至少一个发夹环,使得转座酶将所述模板多核苷酸片段化并将底物连接到双链片段的一端或两端,从而产生多个片段/底物构建体;

(b) 使所述片段/底物构建体与聚合酶接触,使得所述聚合酶置换包括所述突出端的链,并用与包括发夹环的链互补的链替换它们,从而产生多个双链构建体,每个所述双链构建体包括模板多核苷酸的双链片段;和

(c) 分离所述双链构建体的两条链,并使用所述链作为模板以形成多个经修饰的双链多核苷酸,每个所述经修饰的双链多核苷酸包括由至少一个发夹环连接的两条互补链。

2. 根据权利要求1所述的方法,其中步骤(c) 包括通过增加pH、温度和离子强度中的一个或多个来分离所述双链构建体的两条链。

3. 根据权利要求1所述的方法,其中步骤(c) 包括使分离的链与聚合酶接触,使得所述聚合酶使用所述链作为模板以形成所述多个经修饰的双链多核苷酸。

4. 根据权利要求1所述的方法,其中步骤(c) 包括(i) 使多个分离的链与一群核苷酸寡聚体接触,所述一群核苷酸寡聚体包括在所述寡聚体能够与所述链杂交的条件下与所述链中所有核苷酸互补的核苷酸的每种可能组合,和(ii) 将那些与所述链杂交的寡聚体连接在一起以形成所述多个经修饰的双链多核苷酸。

5. 根据权利要求1所述的方法,其中步骤(c) 包括使所述多个双链构建体与聚合酶接触,使得所述聚合酶同时分离所述双链构建体的两条链,并将所述链用作形成所述多个经修饰的双链多核苷酸的模板。

6. 根据权利要求1所述的方法,其中所述至少一个发夹环不连接每个底物的两条链。

7. 根据权利要求1所述的方法,其中所述至少一个突出端长度为4、5或6个核苷酸。

8. 根据权利要求1所述的方法,其中每个底物包括设计为有助于多核苷酸通过孔的运动的前导序列。

9. 根据权利要求1所述的方法,其中所述方法还包括(d) 在发夹环的相对端将Y适配器连接到所述多个经修饰的双链多核苷酸。

10. 根据权利要求9所述的方法,其中所述Y适配器包括设计为有助于多核苷酸通过孔的运动的前导序列。

11. 根据权利要求1所述的方法,其中所述方法还包括将一个或多个多核苷酸结合蛋白与所述多个经修饰的双链多核苷酸结合。

12. 一群用于修饰模板多核苷酸的双链多核苷酸MuA底物,其中所述底物如权利要求1, 6和7中任一项所定义,并且其中所述至少一个发夹环不连接每个底物的两条链。

13. 一种表征模板多核苷酸的方法,包括:

a) 使用根据权利要求1所述的方法修饰模板多核苷酸,以产生多个经修饰的多核苷酸;

b) 使每个经修饰的多核苷酸与跨膜孔接触,使得每个多核苷酸的至少一条链移动通过所述孔;和

c) 当每个多核苷酸相对于孔移动时获取一个或多个测量值,其中所述测量值指示每个多核苷酸的一个或多个特征,从而表征所述模板多核苷酸。

14. 根据权利要求13所述的方法, 其中所述经修饰的多核苷酸/每个经修饰的多核苷酸的两条链移动通过所述孔。

15. 根据权利要求13所述的方法, 其中所述一个或多个特征选自 (i) 多核苷酸的长度, (ii) 多核苷酸的同源性, (iii) 多核苷酸的序列, (iv) 多核苷酸的二级结构和 (v) 多核苷酸是否被修饰。

16. 根据权利要求13所述的方法, 其中接触步骤 (a) 或接触步骤 (b) 还包括使所述经修饰的多核苷酸/每个经修饰的多核苷酸与多核苷酸结合蛋白接触, 使得所述蛋白控制所述多核苷酸/每个多核苷酸通过所述孔的移动。

17. 根据权利要求16所述的方法, 其中所述方法包括

(i) 使所述经修饰的多核苷酸/每个经修饰的多核苷酸与跨膜孔以及一个或多个多核苷酸结合蛋白接触, 使得所述多核苷酸/每个多核苷酸的至少一条链移动通过所述孔, 并且所述一个或多个蛋白控制所述多核苷酸/每个多核苷酸通过孔的移动; 和

(ii) 当所述多核苷酸/每个多核苷酸相对于所述孔移动时, 测量通过所述孔的电流, 其中所述电流指示所述多核苷酸/每个多核苷酸的一个或多个特征, 从而表征所述多核苷酸。

18. 根据权利要求17所述的方法, 其中所述一个或多个多核苷酸结合蛋白在其与所述跨膜孔接触之前与所述经修饰的多核苷酸/每个经修饰的多核苷酸结合。

19. 根据权利要求17或18所述的方法, 其中所述一个或多个多核苷酸结合蛋白衍生自解旋酶。

20. 一种用于修饰模板双链多核苷酸的试剂盒, 其包括 (a) 如权利要求1、6和7中任一项所定义的一群MuA底物, 其中所述至少一个发夹环不连接每个底物的两条链, (b) MuA转座酶和 (c) 聚合酶。

一种修饰模板双链多核苷酸的方法

技术领域

[0001] 本发明涉及一种修饰模板双链多核苷酸的方法,特别是使用纳米孔测序进行表征的方法。该方法从模板产生多个经修饰的双链多核苷酸。然后可以表征这些经修饰的多核苷酸。

背景技术

[0002] 存在许多需要制备核酸库的商业情况。这通常使用转座酶来实现。根据用于制备所述库的转座酶,在能够使用所述库之前,例如在测序中,可能需要在体外修复转座事件。

[0003] 目前需要在广泛的应用中快速且便宜的多核苷酸(例如DNA或RNA)测序和鉴定技术。现有技术是缓慢且昂贵的,主要是因为它们依赖于扩增技术来生产大量的多核苷酸,并且需要大量专门的荧光化学品用于信号检测。

[0004] 跨膜孔(纳米孔)具有作为聚合物和各种小分子的直接电生物传感器的巨大潜力。特别是,最近已经重点将纳米孔作为潜在的DNA测序技术。

[0005] 当跨纳米孔施加电位时,当诸如核苷酸等的分析物瞬间驻留在桶中达一段时间时,存在电流的变化。核苷酸的纳米孔检测给出了已知识别标识(signature)和持续时间的当前的变化。在链测序方法中,单个多核苷酸链通过孔,得到核苷酸的身份。链测序可涉及使用多核苷酸结合蛋白质来控制多核苷酸通过孔的运动。

发明内容

[0006] 发明人惊奇地证明可以修饰模板双链多核苷酸以产生多个更短的经修饰的双链多核苷酸。经修饰的双链多核苷酸可以包括例如发夹环或单链前导序列。这些修饰可以被设计成使得经修饰的双链多核苷酸各自比原始模板多核苷酸更容易表征,例如通过链测序。经修饰的多核苷酸的随后的表征允许更容易确定模板多核苷酸的特征。

[0007] 本发明的修饰方法使用MuA转座酶、MuA底物群和聚合酶,并总结在图1中。MuA底物包含相对链上的突出端和发夹环。MuA转座酶能够使模板多核苷酸片段化,并在两端产生具有突出端的片段。MuA转座酶还能够将底物在双链片段的一端或两端连接到突出端。没有突出端的底物的链通常用突出端连接到片段的链上。这在所得双链构建体中留下单链间隙。双链构建体还具有在与间隙相对的链上的发夹环。

[0008] 该聚合酶能够使用包含发夹环的链作为模板并且移位含有单链间隙的链。所得的双链构建体含有包含模板多核苷酸的片段的两条互补链。该构建体中的两条链可以分离,并优选同时用作模板,以产生包含模板多核苷酸的片段的一个双链构建体,其中两条链通过发夹环连接。

[0009] 因此,本发明提供了一种修饰模板双链多核苷酸的方法,其包括:

[0010] (a) 使模板多核苷酸与MuA转座酶和双链MuA底物群接触,双链MuA底物群中的每个包括(i)至少一个突出端和(ii)与包括至少一个突出端的链相对的链中的至少一个发夹环,使得转座酶将所述模板多核苷酸片段化并将底物连接到双链片段的一端或两端,并且

从而产生多个片段/底物构建体；

[0011] (b) 使所述片段/底物构建体与聚合酶接触，使得所述聚合酶置换包括突出端的链，并用与包括发夹环的链互补的链替换它们，从而产生多个双链构建体，每个所述双链构建体包括模板多核苷酸的双链片段；和

[0012] (c) 分离双链构建体的两条链，并使用所述链作为模板以形成多个经修饰的双链多核苷酸，每个所述经修饰的双链多核苷酸包括由至少一个发夹环连接的两条互补链。

[0013] 本发明还提供：

[0014] -使用本发明的方法制备的多个经修饰的双链多核苷酸；

[0015] -一种用于修饰模板多核苷酸的双链多核苷酸MuA底物群，其中所述底物如以上定义；

[0016] -一种表征使用本发明的方法修饰的至少一种多核苷酸的方法，其包括：

[0017] a) 使经修饰的多核苷酸与跨膜孔接触，使得多核苷酸的至少一条链移动通过孔；和

[0018] b) 当至少一条链相对于孔移动时获取一个或多个测量值，其中所述测量值指示所述至少一条链的一个或多个特征，从而表征所述经修饰的多核苷酸；

[0019] -一种表征模板多核苷酸的方法，包括：

[0020] a) 使用本发明的方法修饰模板多核苷酸，以产生多个经修饰的多核苷酸；

[0021] b) 使每个经修饰的多核苷酸与跨膜孔接触，使得每个多核苷酸的至少一条链移动通过孔；和

[0022] c) 当每个多核苷酸相对于孔移动时获取一个或多个测量值，其中所述测量值指示每个多核苷酸的一个或多个特征，从而表征模板多核苷酸；以及

[0023] -一种用于修饰模板双链多核苷酸的试剂盒，其包括(a) 如以上定义的MuA底物群，(b) MuA转座酶和(c) 聚合酶。

附图说明

[0024] 图1示出了修饰模板双链多核苷酸(标记为a)的方法的卡通表示。步骤1涉及使模板双链多核苷酸与MuA转座酶(标记为b)以及双链MuA底物群(标记为c，双链MuA底物均含有5'发夹环)接触，使得MuA转座酶将模板双链多核苷酸片段化，并将MuA底物在分裂点的每一侧插入。步骤2涉及用聚合酶(标记为e)和dNTP处理模板链，dNTP替换标记为d的DNA片段并产生与DNA 5'发夹环互补的链。步骤3涉及对标记为f的双链DNA构建体的热处理，使得所述链变性成单链DNA(标记为g)。最后，步骤4涉及形成互补链的DNA聚合酶。

[0025] 图2示出了修饰实施例1中概述的模板双链多核苷酸(标记为a)的方法的卡通表示。步骤1涉及使模板双链多核苷酸与MuA转座酶(标记为b)以及双链MuA底物群(标记为c，双链MuA底物均含有5'发夹环)接触，使得MuA转座酶将模板双链多核苷酸片段化，并将MuA底物在分裂点的每一侧插入。步骤2涉及用聚合酶(标记为e)和dNTP处理模板链，dNTP替换标记为d的DNA片段并产生对于DNA 5'发夹环的互补链。步骤3涉及标记为f的双链DNA构建体的热处理，使得所述链变性成单链DNA(标记为g)。步骤4涉及用形成互补链的DNA聚合酶进行的第二次处理。最后，步骤5涉及步骤4中产生的双链DNA构建体的dA-拖尾(dA-tailing)、具有酶(标记为h)预先结合的适配器的连接，以及含有胆固醇系链(标记为j)的

DNA链(标记为i)的杂交。这产生了在实施例1中描述的纳米孔系统中被测试的最终DNA构建体。

[0026] 图3示出了当解旋酶(T4Dda-E94C/C109A/C136A/A360C(SEQ ID NO:24具有突变E94C/C109A/C136A/A360C))控制DNA样品6通过MspA纳米孔易位时的示例性电流迹线(y轴坐标=电流(pA),x轴坐标=时间(s))。

[0027] 图4示出了Agilent 12,000DNA芯片迹线。标记为1的线是未处理的MuA片段化DNA输入材料,标记为2的线是具有68°C培育步骤(在实施例1的1.2中)的分析物,并且随后经历了实施例1的步骤1.3的全部,标记为3的线在实施例1的步骤1.2中没有68°C培育但是经历了实施例1的步骤1.3的全部。区域X对应于双链DNA库,区域Y对应于Agilent 12,000的上标记并且区域Z对应于Agilent 12,000芯片的下标记。

[0028] 图5示出了修饰模板双链多核苷酸(标记为a)的优选方法的卡通表示。除了每个底物包含通过间隔基(spacer)(xxx;标记为h)与发夹环分离的前导序列(标记为i)之外,图5与图1相同。引导序列不用作模板,因为聚合酶(标记为e)不能移动通过间隔基。

[0029] 图6示出了当解旋酶(T4Dda-E94C/C109A/C136A/A360C)控制DNA样品7通过MspA纳米孔易位时的示例性电流迹线(y轴坐标=电流(pA),x轴坐标=时间(s))。

[0030] 图7示出了修饰模板双链多核苷酸(标记为a)的方法的卡通表示。步骤1涉及使模板双链多核苷酸与MuA转座酶(标记为b)以及双链MuA底物群(标记为C,其中双链MuA底物均含有5'发夹环,其在替换G/C's的发夹(标记为h并且显示为黑色圆圈)中含有I/Z's)接触,使得MuA转座酶将模板双链多核苷酸片段化,并将MuA底物在分裂点的每一侧插入。步骤2涉及用聚合酶(标记为e)和dNTP处理模板链,dNTP替换标记为d的DNA片段并产生对于DNA 5'发夹环的互补链(产生的dsDNA标记为f)。由聚合酶形成的双链区域(标记为1X)由均能够形成发夹环的两条链组成。由链F2形成的发夹环具有比双链区域1X的 T_m 更高的 T_m ,这是因为链F2的发夹环由C/T/A/G组成,并且双链区域1X是链f2杂交至链f1,其中链F1由Z/T/A/I组成(并且Z和I仅形成两个氢键,而C/G形成3个氢键)。因此,F2形成发夹环(标记为f2h),并且F1形成发夹环(标记为f1h),由链F1形成的发夹环具有比由链F2形成的发夹环更高的 T_m 。然后DNA聚合酶能够产生以虚线/点线显示的互补链(标记为i1和i2的整个dsDNA构建体)。因此,聚合酶能够形成互补链(如虚线/点线所示),而不需要加热在步骤2中产生的dsDNA(并且标记为f1,其与f2杂交)。

[0031] 图8示出了本发明的优选方法的卡通表示。步骤1至4与图1中相同。步骤5涉及向图1中形成的构建体添加发夹环。步骤6涉及修饰的双链多核苷酸的热处理,使得所述链变性能成单链构建体。最后,步骤7涉及形成互补链的DNA聚合酶。

[0032] 序列表说明

[0033] SEQ ID NO:1示出了对MS-B1突变体MspA单体进行编码的密码子优化的多核苷酸序列。该突变体缺乏信号序列并且包括以下突变:D90N,D91N,D93N,D118R,D134R和E139K。

[0034] SEQ ID NO:2示出了MspA单体的MS-B1突变体的成熟形式的氨基酸序列。该突变体缺乏信号序列并且包括以下突变:D90N,D91N,D93N,D118R,D134R和E139K。

[0035] SEQ ID NO:3示出了对 α -溶血素-E111N/K147N(α -HL-NN;Stoddart等人,PNAS,2009;106(19):7702-7707)的1种单体进行编码的多核苷酸序列。

[0036] SEQ ID NO:4示出了 α -HL-NN的1种单体的氨基酸序列。

- [0037] SEQ ID NO:5至7示出了MspB,C和D的氨基酸序列。
- [0038] SEQ ID NO:8示出了对Phi29DNA聚合酶进行编码的多核苷酸序列。
- [0039] SEQ ID NO:9示出了Phi29DNA聚合酶的氨基酸序列。
- [0040] SEQ ID NO:10示出了由来自大肠杆菌的sbcB基因得到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酶 (EcoExo I)。
- [0041] SEQ ID NO:11示出了来自大肠杆菌的核酸外切酶I酶 (EcoExo I) 的氨基酸序列。
- [0042] SEQ ID NO:12示出了由来自大肠杆菌的xthA基因得到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III酶。
- [0043] SEQ ID NO:13示出了来自大肠杆菌的核酸外切酶III酶的氨基酸序列。该酶在3'-5'方向上从双链DNA(dsDNA)的一条链进行5'单磷酸核苷的分布消化。链上的酶引发需要约4个核苷酸的5'突出端。
- [0044] SEQ ID NO:14示出了由来自嗜热栖热菌(*T.thermophilus*)的recJ基因衍生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)。
- [0045] SEQ ID NO:15示出了来自嗜热栖热菌的RecJ酶(TthReeJ-cd)的氨基酸序列。该酶在5'-3'方向上从ssDNA进行5'单磷酸核苷的进行性消化(processive digestion)。链上的酶引发需要至少4个核苷酸。
- [0046] SEQ ID NO:16示出了由噬菌体λ_{exo}(redX)基因衍生的密码子优化的多核苷酸序列。它编码细菌噬菌体λ核酸外切酶。
- [0047] SEQ ID NO:17示出了细菌噬菌体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的亚基之一。该酶在5'-3'方向从dsDNA的一条链进行核苷酸的高度进行性消化(<http://www.neb.com/nebecomm/products/productM0262.asp>)。链上的酶引发优选需要约4个具有5'磷酸的核苷酸的5'突出端。
- [0048] SEQ ID NO:18示出了He1308Mbu的氨基酸序列。
- [0049] SEQ ID NO:19示出了He1308 Csy的氨基酸序列。
- [0050] SEQ ID NO:20示出了He1308Tga的氨基酸序列。
- [0051] SEQ ID NO:21示出了He1308Mhu的氨基酸序列。
- [0052] SEQ ID NO:22示出了TraI Eco的氨基酸序列。
- [0053] SEQ ID NO:23示出了XPD Mbu的氨基酸序列。
- [0054] SEQ ID NO:24示出了Dda 1993的氨基酸序列。
- [0055] SEQ ID NO:25示出了Trwc Cba的氨基酸序列。
- [0056] SEQ ID NO:26至28示出了本发明优选的MuA底物的序列。
- [0057] SEQ ID NO:29示出了实施例1中使用的多核苷酸序列。
- [0058] SEQ ID NO:30示出了实施例1中使用的多核苷酸序列。该序列在其5'末端连接有如下列多核苷酸序列——GATCU。
- [0059] SEQ ID NO:31示出了实施例1中使用的肠杆菌噬菌体λ的多核苷酸序列。该序列包含连接在模板链5'末端的另外的12碱基突出端。此处示出的序列仅为模板链的序列(模板互补体未示出)。
- [0060] SEQ ID NO:32示出了实施例1中使用的多核苷酸序列。
- [0061] SEQ ID NO:33示出了实施例1中使用的多核苷酸序列。

- [0062] SEQ ID NO:34示出了实施例1中使用的多核苷酸序列。
- [0063] SEQ ID NO:35示出了实施例1中使用的多核苷酸序列。
- [0064] SEQ ID NO:36示出了实施例2中使用的多核苷酸序列。
- [0065] SEQ ID NO:37示出了实施例2中使用的多核苷酸序列。

具体实施方式

[0066] 应当理解,所公开的产品和方法的不同应用可以根据本领域的具体需要进行调整。还应当理解,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,而不是限制性的。

[0067] 此外,除非另有明确规定,在本说明书和所附权利要求书中使用的单数形式“一个”,“一种”和“该”包括复数指示物。因此,例如,提及“一个多核苷酸”时包括“多个多核苷酸”,提及“一个底物”时包括两个或更多个这样的底物,提及“一个跨膜蛋白孔”时包括两个或更多个这样的孔,等。

[0068] 本文引用的所有出版物、专利和专利申请,无论是上文还是下文中,均通过整体引用并入本文。

[0069] 本发明的修饰方法

[0070] 本发明提供了修饰模板多核苷酸的方法。可以为任何目的修饰模板。该方法优选用于修饰用于表征例如用于链测序的模板多核苷酸。模板多核苷酸通常是根据本发明最终将被表征或测序的多核苷酸。这将在下面更详细地讨论。

[0071] 该方法涉及形成多个修饰的双链多核苷酸。这些修饰的双链多核苷酸通常比模板多核苷酸更容易表征,特别是使用链测序进行表征。所述多个修饰的双链多核苷酸自身可以被表征,以便于表征模板多核苷酸。例如,可以通过测序每个修饰的双链多核苷酸来确定模板多核苷酸的序列。

[0072] 修饰的双链多核苷酸通常比模板多核苷酸短,因此使用链测序来表征它们是更容易的。修饰的双链多核苷酸还包括如下所述的信息量的两倍。

[0073] 可以通过将标记包含在MuA底物中来选择性地标记修饰的双链多核苷酸。合适的标记包括但不限于校准序列,偶联部分和适配器结合的酶。

[0074] 在一些实施方案中,该方法引入双链多核苷酸修饰中,这有助于使用链测序对其进行表征。众所周知,将多核苷酸偶联到含有纳米孔的膜中使得多核苷酸的量——允许其表征或测序所需的量——降低了几个数量级。这在申请号为PCT/GB2012/051191(公开为W0 2012/164270)的国际申请中进行了讨论。本发明的方法允许生产多个双链多核苷酸,每个双链多核苷酸包括将多核苷酸偶联到膜上的装置。这将在下面更详细地讨论。

[0075] 使用纳米孔表征双链多核苷酸通常需要存在设计成优先螺旋进入纳米孔的前导序列。本发明的方法允许生产多个双链多核苷酸,每个双链多核苷酸包括单链前导序列。这将在下面更详细地讨论。

[0076] 还已经确定的是,通过诸如发夹环的桥接部分连接双链多核苷酸的两条链使得所述多核苷酸的两条链通过纳米孔进行表征或测序。这是有利的,因为它使得从单个双链多核苷酸获得的信息量加倍。此外,由于模板互补链中的序列必然与模板链的序列正交,所以来自两条链的信息可以进行信息组合。因此,该机制提供了正交校验能力,提供更高置信度

观察结果。这在申请号为PCT/GB2012/051786(公开为W0 2013/014451)的国际申请中进行了讨论。本发明的方法允许产生多个修饰的双链多核苷酸,其中每个多核苷酸的两条链使用发夹环连接。

[0077] 模板多核苷酸

[0078] 本发明的方法修饰模板双链多核苷酸,优选用于表征。模板多核苷酸通常是根据本发明最终将被表征或测序的多核苷酸。它也可以称为目标双链多核苷酸或感兴趣的双链多核苷酸。

[0079] 多核苷酸,例如核酸,是包含两个或多个核苷酸的大分子。多核苷酸或核酸可以包含任何核苷酸的任何组合。核苷酸可以是天然存在的或人造的。多核苷酸中的一个或多个核苷酸可被氧化或甲基化。多核苷酸中的一个或多个核苷酸可被损坏。例如,多核苷酸可以包含嘧啶二聚体。这种二聚体通常与紫外线的损伤相关,并且是皮肤黑色素瘤的主要原因。可以修饰多核苷酸中的一个或多个核苷酸,例如用标记或标签进行修饰。合适的标记如下所述。多核苷酸可以包含一个或多个间隔基。

[0080] 核苷酸通常含有核碱基、糖和至少一个磷酸基团。核碱基和糖形成核苷。

[0081] 核碱基通常是杂环的。核碱基包括但不限于嘌呤和嘧啶,并且更具体地包括腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、尿嘧啶(U)和胞嘧啶(C)。

[0082] 糖通常是戊糖。核苷酸糖包括但不限于核糖和脱氧核糖。糖优选为脱氧核糖。

[0083] 多核苷酸优选包含以下核苷:脱氧腺苷(dA)、脱氧尿苷(dU)和/或胸苷(dT),脱氧鸟苷(dG)和脱氧胞苷(dC)。

[0084] 核苷酸通常是核糖核苷酸或脱氧核糖核苷酸。核苷酸通常含有一磷酸盐、二磷酸盐或三磷酸盐。核苷酸可以包含多于三个的磷酸盐,例如4或5个磷酸盐。磷酸盐可以连接在核苷酸的5'或3'侧上。核苷酸包括但不限于腺苷单磷酸(AMP)、鸟苷单磷酸(GMP)、胸苷单磷酸(TMP)、尿苷单磷酸(UMP)、5-甲基胞苷单磷酸,5-羟甲基胞苷单磷酸、胞苷单磷酸(CMP)、环状腺苷单磷酸(cAMP)、环状鸟苷单磷酸(cGMP)、脱氧腺苷单磷酸(dAMP)、脱氧鸟苷单磷酸(dGMP)、脱氧胸苷单磷酸(dTMP)、脱氧尿苷单磷酸(dUMP)、脱氧胞苷单磷酸(dCMP)和脱氧甲基胞苷单磷酸。核苷酸优选地选自AMP,TMP,GMP,CMP,UMP,dAMP,dTMP,dGMP,dCMP和dUMP。

[0085] 核苷酸可以是脱碱基的(即缺少核碱基)。核苷酸也可以缺少核碱基和糖(即C3间隔基)。

[0086] 多核苷酸中的核苷酸可以以任何方式彼此连接。核苷酸通常通过其糖和磷酸酯基团连接,如核酸中一样。核苷酸可以通过其核碱基连接,如嘧啶二聚体中一样。

[0087] 多核苷酸是双链的。多核苷酸的至少一部分优选是双链的。

[0088] 多核苷酸可以是核酸,例如脱氧核糖核酸(DNA)或核糖核酸(RNA)。多核苷酸可以包含与DNA的一条链杂交的一条RNA链。多核苷酸可以是本领域已知的任何合成核酸,例如肽核酸(PNA)、甘油核酸(GNA)、苏糖核酸(TNA)、锁定核酸(LNA)或具有核苷酸侧链的其他合成聚合物。PNA骨架由通过肽键连接的重复的N-(2-氨基乙基)-甘氨酸单元组成。GNA骨架由通过磷酸二酯键连接的重复的二醇单元组成。TNA骨架由通过磷酸二酯键连接在一起的重复的苏糖组成。LNA由如上所述的核糖核苷酸形成,所述核糖核苷酸具有连接核糖部分中的2'氧和4'碳的额外的桥。

[0089] 多核苷酸最优选为核糖核酸(RNA)或脱氧核糖核酸(DNA)。

[0090] 多核苷酸可以是任何长度。例如,多核苷酸的长度可以是至少10个、至少50个、至少100个、至少150个、至少200个、至少250个、至少300个、至少400个或至少500个核苷酸或核苷酸对。多核苷酸的长度可以是1000个或更多个核苷酸或核苷酸对、5000个或更多个核苷酸或核苷酸对或者100000个或更多个核苷酸或核苷酸对。

[0091] 可以使用本发明研究任何数目的多核苷酸。例如,本发明可能涉及表征 2,3,4,5,6,7,8,9,10,20,30,51,100或更多个多核苷酸。如果两个或更多个多核苷酸被表征,它们可以是不同的多核苷酸或同一种多核苷酸的两个实例。

[0092] 多核苷酸可以是天然存在的或人造的。例如,该方法可以用于验证制造的寡核苷酸的序列。该方法通常在体外进行。

[0093] 模板多核苷酸通常存在于任何合适的样品中。本发明通常针对已知含有或怀疑含有模板多核苷酸的样品进行。或者,本发明可以针对样品进行,以确认一种或多种已知或预期在样品中存在的模板多核苷酸的身份。

[0094] 样品可以是生物样品。本发明可以针对从任何生物体或微生物获得或提取的样品在体外进行。生物体或微生物通常是古菌、原核或真核的,并且通常属于五界之一:植物界,动物界,真菌界,原核生物界和原生生物界。本发明可以对从任何病毒获得或从提取的样品在体外进行。样品优选为流体样品。样品通常包括患者的体液。样品可以是尿液、淋巴、唾液、粘液或羊水,但优选为血液、血浆或血清。通常,样品是来源于人的,但是也可以来自另一种哺乳动物,例如来自商业养殖的动物如马、牛、绵羊或猪,或者还可以是宠物如猫或狗。或者,植物来源的样品通常从商业作物获得,例如谷物、豆类、水果或蔬菜,如小麦、大麦、燕麦、油菜、玉米、大豆、稻、香蕉、苹果、西红柿、土豆、葡萄、烟草、菜豆、扁豆、甘蔗、可可、棉花。

[0095] 样品可以是非生物样品。非生物样品优选为流体样品。非生物样品的实例包括手术液,水如饮用水、海水或河水,以及实验室试验用试剂。

[0096] 样品通常在用于本发明之前进行处理,例如通过离心,或通过穿过膜过滤掉不需要的分子或细胞例如红细胞。取样后可以立即测量样品。样品通常也可以在测定之前储存,优选低于-70℃。

[0097] MuA和条件

[0098] 模板多核苷酸与MuA转座酶接触。这种接触在允许转座酶起作用的条件下进行,所述起作用即,将模板多核苷酸片段化并将MuA底物连接到片段的一端或两端。MuA转座酶可从例如Thermo Scientific(目录号F-750C,20μL(1.1μg/μL))购得。MuA转座酶将起作用的条件是本领域已知的。合适的条件在实施例中描述。

[0099] 底物群

[0100] 使模板多核苷酸与双链MuA底物群接触。双链底物是多核苷酸底物,并且可以由上述任何核苷酸或核酸形成。底物通常由与模板多核苷酸相同的核苷酸形成。

[0101] 底物群通常是同源的(即通常含有多个相同的底物)。底物群可能是异源的(即可能包含多个不同的底物)。

[0102] 用于MuA转座酶的合适底物是本领域已知的(Saariaho和Savilahti,Nucleic Acids Research,2006;34(10):3139-3149和Lee和Harshey,J.Mol.Biol.,2001;314:433-444)。

[0103] 每个底物通常包含提供其作为MuA转座酶底物的活性的双链部分。双链部分在每个底物中通常是相同的。底物群可以包括不同的双链部分。

[0104] 每个底物中的双链部分通常长度为至少50个核苷酸对,例如至少55个,至少60个或至少65个核苷酸对。每个底物中的双链部分优选包含在每条链的3'端含有脱氧胞苷(dC)和脱氧腺苷(dA)的二核苷酸。dC和dA通常在双链部分的两条链中具有不同的取向,即在从5'到3'进行读取时,在3'端,一条链具有dC/dA,而另一条链具有dA/dC。

[0105] 双链部分的一条链优选包含SEQ ID NO:26中所示的序列,双链部分的另一条链优选包含SEQ ID NO:27中所示的序列。

[0106] 5'-GTTTTTCGCATTTATCGTGAAACGCTTTTCGCGTTTTTCGTGCGCCGCTTCA-3'(SEQ 26)

[0107] 3'-CAAAAGCGTAAATAGCACTTTGCGAAAGCGCAAAAAGCACGCGGCGAAGT-5'(SEQ 27)

[0108] 每个底物包括至少一个突出端。突出端通常是核苷酸突出端。在每个底物的一端或两端可能有突出端。如果每个底物中的双链部分包含与SEQ ID NO:27所示的序列杂交的SEQ ID NO:26所示的序列,则至少一个突出端优选在SEQ ID NO:27所示的序列的5'端。

[0109] 每个底物可以包括两个突出端,即每个底物两端各一个。如果在底物的两端都存在突出端,则每个突出端通常在双链多核苷酸部分的不同链上。优选地,突出端位于双链部分的一条链的5'端。

[0110] 每个底物优选包括仅一个突出端。该仅一个突出端优选在双链部分的一条链的5'端。

[0111] 突出端长度可以是至少3个、至少4个、至少5个、至少6个或至少7个核苷酸。突出端长度优选为5个核苷酸。

[0112] 在优选的实施例中,底物的一条链包含SEQ ID NO:26中所示的序列,底物的另一条链包含SEQ ID NO:28中所示的序列(见下文)。

[0113] 5'-GTTTTTCGCATTTATCGTGAAACGCTTTTCGCGTTTTTCGTGCGCCGCTTCA-3'(SEQ26)

[0114] 3'-CAAAAGCGTAAATAGCACTTTGCGAAAGCGCAAAAAGCACGCGGCGAAGTCTAG-5'(SEQ 28)

[0115] 所述底物群中的底物可以具有申请号为PCT/GB2014/052505的国际申请中公开的任何结构。

[0116] 每个底物在与包含至少一个突出端的链相对的链中包含至少一个发夹环。发夹环通常不连接底物的两条链。发夹环可以是内部发夹环,即不在与包含至少一个突出端的链相对的链的末端。内部发夹环优选与间隔基相邻,在本发明的方法中使用的任何聚合酶不能移动经过该间隔基。间隔基可以位于发夹环的任一侧。可以使用以下讨论的任何间隔基,例如一个或多个iSpC3基团(即缺少糖和碱基的核苷酸)、一个或多个间隔基9(iSp9)基团或一个或多个间隔基18(iSp18)基团。内部发夹环优选与诸如硝基吡啶等的非天然核苷酸相邻,本发明方法中使用的任何聚合酶不能移动经过该核苷酸。可以使用以下讨论的任何不同的核苷酸种类。

[0117] 发夹环优选位于与包含至少一个突出端的链的相对的链的端部处或该端部附近。如果距离与包含至少一个突出端的链相对的链的末端20个核苷酸或更少、15个核苷酸或更少、10个核苷酸或更少或5个核苷酸或更少,则发夹环在与含至少一个突出端的链相对的链的端部附近。如果在形成链末端的干部分(杂交部分)的最后核苷酸之间存在20个或更少的核苷酸,则发夹环距离该链末端20个核苷酸或更少。发夹环优选在与包含至少一个突出

端的链相对的链的末端。在每个底物的一端或两端可能存在发夹环。发夹环优选位于与至少一个突出端相对的底物末端。

[0118] 发夹环通常是核苷酸发夹环。如果每个底物中的双链部分包含与SEQ ID NO:27中所示的序列杂交的SEQ ID NO:26中所示的序列,则至少一个发夹环优选在SEQ ID NO:26中所示序列的5'端。

[0119] 每个底物可以包含两个发夹环,即每个底物的两条链中各一个或在每个底物两端各一个。如果在底物的两端存在发夹环,则每个发夹环通常位于双链多核苷酸部分的不同链上。发夹环优选位于双链部分的链的5'端。

[0120] 每个底物优选包含仅一个发夹环。该仅一个发夹环优选在与包含至少一个突出端的链相对的链中。该仅一个发夹环优选位于与至少一个突出端相对的底物末端,并且在与包含至少一个突出端的链相对的链中。该仅一个发夹环优选在双链部分的一条链的5'端并且在与包含至少一个突出端的链相对的链中。

[0121] 在一个优选的实施例中,每个底物包含在双链部分的一条链的5'端处的一个突出端和在双链部分的另一条链的5'端处的发夹环。在最优的实施例中,底物的一条链包含SEQ ID NO:26中所示的序列,底物的另一条链包含SEQ ID NO:28中所示的序列(见上文),并且发夹环位于SEQ ID NO:26中所示序列的5'端。

[0122] 可以使用本领域已知的方法设计合适的发夹环。发夹环可以是任何长度。发夹环的长度通常为110个或更少的核苷酸,例如100个或更少的核苷酸、90个或更少的核苷酸、80个或更少的核苷酸、70个或更少的核苷酸、60个或更少的核苷酸、50个或更少的核苷酸、40个或更少的核苷酸、30个或更少的核苷酸、20个或更少的核苷酸或10个或更少的核苷酸。发夹环的长度优选为约1至110,2至100,5至80或6至50个核苷酸。

[0123] 发夹环可以由任何上述的核苷酸形成。发夹环可以由与双链部分相同的核苷酸形成。发夹环优选由导致发夹环具有比双链部分更低的熔化温度(T_m)的核苷酸形成。可以使用常规技术测量熔化温度。如果双链部分包含RNA,则发夹优选由含有腺苷(A)、尿苷(U)、肌苷(I)和zebularine(Z)的核苷酸形成。如果双链部分包含DNA,则发夹优选由含有脱氧腺苷(dA)、胸苷(dT)、脱氧肌苷(dI)和脱氧zebularine(dZ)的核苷酸形成。用肌苷(I)/脱氧肌苷(dI)代替鸟苷(G)/脱氧鸟苷(dG)以及用zebularine(Z)/脱氧zebularine(dZ)取代胞苷(C)/脱氧胞苷(dC)相比于双链部分降低了发夹的 T_m 。I/dI和Z/dZ仅形成两个氢键,而G/dG和C/dC形成三个氢键。在本发明的方法中,聚合物用与包含发夹环的链互补的新链取代突出链。具有较低 T_m 的发夹环可用于形成具有较高 T_m 的互补的发夹,即由具有较高 T_m 的核苷酸形成的发夹。聚合酶可以用与包含发夹环的链互补的新链取代突出链,其中新链包含具有比模板链中发夹环更高的 T_m 的发夹环。例如,由含有腺苷(A)/脱氧腺苷(dA)、尿苷(U)/胸苷(dT)、肌苷(I)/脱氧肌苷(dI)和zebularine(Z)/脱氧zebularine(dZ)的核苷酸形成的发夹环可以用于形成互补的RNA或DNA发夹环。两个发夹之间的 T_m 差异意味着它们作为个体发夹比杂交在一起时更稳定。这意味着两个发夹环形成它们各自的环,而不是杂交在一起。这有助于所述方法的最后一步,其中双链构建体的两条链分离并用作模板以形成多个修饰的双链多核苷酸,每个修饰的双链多核苷酸包含由至少一个发夹环连接的两条互补链。例如,分离可以在室温下进行。

[0124] 每个底物可以包含可选择的结合部分。如果存在,可选择的结合部分优选在发夹

环中。可选择的结合部分是可以基于其结合性质而选择的部分。因此,可选择的结合部分优选是特异性结合至表面的部分。如果可选择的结合部分以比在本发明中使用的任何其它部分以大得多的程度结合到表面,则可选择的结合部分特异性地结合到表面。在优选的实施方案中,该部分结合到没有其它本发明中使用的部分结合的表面。

[0125] 合适的选择性结合部分是本领域已知的。优选的选择性结合部分包括但不限于生物素、核酸序列、抗体、抗体片段,例如Fab和ScSv、抗原、核酸结合蛋白、多组氨酸尾部和GST标签。最优选的选择性结合部分是生物素和可选择的核酸序列。生物素特异性结合到包被有抗生物素蛋白的表面。可选择的核酸序列特异性结合(即杂交)到包被有同源序列的表面。或者,可选择的核酸序列特异性结合包被有核酸结合蛋白的表面。

[0126] 每个底物可以包含前导序列。前导序列通常与至少一个发夹环在同一条链上。前导序列通常位于与发夹环相同的底物末端。前导序列通常位于包含至少一个发夹环的链末端(即,发夹环位于末端前导序列和底物的其余部分之间)。前导序列通常通过间隔基与发夹环分离,在本发明的方法中使用的任何聚合酶不能移动通过该间隔基。可以使用以下任何间隔基,例如一个或多个iSpC3基团(即缺少糖和碱基的核苷酸)、一个或多个间隔基9(iSp9)基团或一个或多个间隔基18(iSp18)基团。间隔基意味着前导序列在步骤(b)和(c)中不用作模板,因此在方法结束时保持单链。这允许前导序列执行其功能。这样的例子在图5中示出。

[0127] 前导序列优先旋入孔中。前导序列有助于本发明的表征方法。前导序列被设计成优先旋入孔中,从而有助于多核苷酸通过孔的运动。前导序列也可以用于将多核苷酸连接到一个或多个锚,如下所述。前导序列通常包含聚合物。聚合物优选是带负电荷的。聚合物优选为多核苷酸,例如DNA或RNA,经修饰的多核苷酸(例如脱碱基DNA),PNA,LNA,聚乙二醇(PEG)或多肽。前导序列优选包含多核苷酸,更优选包含单链多核苷酸。前导序列可以包含任何上述的多核苷酸。单链前导序列最优选包含单链DNA,例如聚dT区段。前导序列优选包含一个或多个间隔基。

[0128] 前导序列可以是任何长度,但是长度通常为10至150个核苷酸,例如长度为20至150个核苷酸。前导序列的长度通常取决于方法中使用的跨膜孔。

[0129] 片段化

[0130] 转座酶将模板双链多核苷酸片段化,形成多个双链片段。转座酶还将底物连接到双链片段的一端或两端,从而产生多个片段/底物构建体。转座酶优选将底物连接到双链片段的末端,从而产生多个片段/底物构建体,其各自在末端具有发夹环。这样的例子可以在图1中看到。

[0131] 聚合酶

[0132] 由转座酶产生的片段/底物构建体与聚合酶接触。可以使用以下讨论的任何聚合酶。聚合酶优选为Klenow或9°North。聚合酶更优选为LongAmp®Taq DNA聚合酶(可以从New England Biolabs®Inc.购得),Phusion®High-Fidelity DNA聚合酶(可以从New England Biolabs Inc.购得)或KAPAHiFi(可以从KAPABiosystems购得)。

[0133] 在聚合酶可以置换突出的链并形成互补的多核苷酸的条件下,使构建体与聚合酶接触。这些条件是本领域已知的。例如,通常将这些构建体与可购得的聚合酶缓冲液(例如来自New England Biolabs或KAPA Biosystems的缓冲液)中的聚合酶接触。对于Klenow,

温度优选为20至37℃,对于9°North,LongAmp®Taq DNA聚合酶,Phusion®High-Fidelity DNA 聚合酶或KAPA HiFi,温度优选为60至75℃。

[0134] 聚合酶置换包含来自片段/底物构建体的突出端的链。聚合酶用与包含发夹环的链互补的新链取代突出链。这产生多个双链构建体,每个双链构建体包含模板多核苷酸的双链片段。由聚合酶形成的部分新链通常与发夹环互补。这意味着发夹环通常在构建体中形成双链多核苷酸的一部分。这样的例子可以在图1中看到。

[0135] 聚合酶可以形成包含上文和下文所讨论的任何核苷酸的新链。聚合酶具有与包含发夹环的链中的核苷酸互补的游离核苷酸群。聚合酶可以使用游离核苷酸形成新的链。

[0136] 分离/复制

[0137] 分离双链构建体的两条链,并且将该链用作模板以形成多个修饰的双链多核苷酸,每个双链多核苷酸包含由至少一个发夹环连接的两条互补链。一个例子如图1所示。

[0138] 在将两条链用作模板之前,可以将所述两条链完全分离。两条链可以分离并在同一时间被用作模板(即同时)。换句话说,两条链不需要完全分离,或者两条链在用作模板之前可被部分分离。

[0139] 两条链可以以任何方式分离。该方法优选包括通过增加pH、温度和离子强度中的一种或多种来分离双链构建体的两条链。优选升高的温度。该方法优选包括将温度升高至95℃。该方法优选包括将温度升高至95℃,然后将温度降至55℃。该方法优选包括将温度升高至95℃,将温度降低至55℃,然后将温度升高至68℃。该方法最优选包括在95℃下培育双链构建体2分钟,在55℃下培育30秒,然后在68℃下培育30分钟。可以使用甲酰胺或氢氧化钠(NaOH)来实现pH的增加。酶,例如解旋酶或消化模板链的酶(例如,如果该链具有dU而不是dT,则为USER)也可用于分离链。下面讨论的任何解旋酶都可以使用。

[0140] 如下面更详细讨论的,可以使用聚合酶分离两条链。聚合酶可以是上面或下面讨论的那些。

[0141] 可以使用任何方法来形成使用分离的链作为模板的新的多核苷酸。该方法优选包括使链与聚合酶接触,使得聚合酶使用链作为模板形成多个修饰的双链多核苷酸。可以使用上述或下面讨论的任何聚合酶。

[0142] 或者,该方法可以包括(i)使多个链与一群核苷酸寡聚体接触,核苷酸寡聚体包括在寡聚体能够杂交到链的条件下与链中所有核苷酸互补的核苷酸的每一种可能组合,和(ii)将那些与链杂交的寡聚体连接在一起以形成多个经修饰的双链多核苷酸。允许杂交的条件是本领域已知的(例如,Sambrook等人,2001,Molecular Cloning:a laboratory manual,第3版,Cold Spring Harbor Laboratory Press;和Current Protocols in Molecular Biology,第2章,Ausubel et al.,Eds.,Greene Publishing and Wiley-Interscience,New York(1995))。杂交可以在低严格条件下进行,例如在37℃下在30-35%甲酰胺、1M NaCl和1%SDS(十二烷基硫酸钠)的缓冲溶液存在下进行,然后在50℃下在从1X(0.1650MNa+)到2X(0.33MNa+)的SSC(标准柠檬酸钠)中洗涤。杂交可以在中度严格条件下进行,例如在37℃下在40至45%甲酰胺、1M NaCl和1%SDS的缓冲溶液的存在下进行,然后在55℃下在从0.5X(0.0825M Na+)到1X(0.1650MNa+)的SSC中洗涤。杂交可以在高严格条件下进行,例如在37℃下在50%甲酰胺、1M NaCl、1%SDS的缓冲溶液存在下进行,然后在60℃下在0.1X(0.0165M Na+)的SSC中洗涤。优选的条件优选为10mM Tris-HCl,50mM NaCl,

pH7中的10uM寡聚体,并加热至98℃,然后以2℃/分钟冷却至18℃。

[0143] 群中的寡聚体通常具有2至16个核苷酸。群中的所有寡聚体可具有 2,3,4,5,6,7,8,9,10,11,12,13,14,15或16个核苷酸。群中的寡聚体可以具有不同的长度。群中的所有寡聚体优选具有相同的长度。寡聚体可以包含任何上述核苷酸。核苷酸与和寡聚体杂交的链中的核苷酸互补。对于本领域技术人员来说,识别与那些核苷酸互补的核苷酸是较容易的。如果核苷酸通过碱基配对(优选Watson和Crick碱基配对)与另一个核苷酸杂交,则所述核苷酸与所述另一个核苷酸互补。互补的核苷酸可以与不与其互补的其他核苷酸杂交,但是比其杂交到与其互补的核苷酸的程度小。N优选包含核碱基腺嘌呤(A)、尿嘧啶(U)、鸟嘌呤(G)或胞嘧啶(C)。或者,N优选包含核碱基A、胸腺嘧啶(T)、G或C。A与 T或U互补,反之亦然。G与C互补,反之亦然。

[0144] 所述群包括与链中所有核苷酸互补的核苷酸的每种可能组合。这意味着寡聚体将与绝大多数(如果不是全部)的链杂交,无论其序列如何。例如,如果N包含核碱基腺嘌呤(A)、尿嘧啶(U)、鸟嘌呤(G)或胞嘧啶(C),则所述群包括A、U、G和C的每一种可能的组合。类似地,如果N包含核碱基A、胸腺嘧啶(T)、G或C,则所述群包含A、T、G和 C的每一种可能的组合。

[0145] 设计和获得具有所需组合的寡聚体群是较为简单的。例如,如果群中所有的寡聚体都包含或由NN组成,并且N是A、T、G或C,则所述群包含AT,AG,AC,TA,TG,TC, GA,GT,GC,CA,CT和CG。类似地,如果群中所有寡聚体包含或由NNN组成,并且 N是A、T、G或C,则所述群包含ATG,ATC,AGT,AGC,ACT,ACG,TAG,TAC, TGA,TGC,TCA,TCG,GAT,GAC,GTA,GTC,GCA,GCT,CAT,CAG,CTA, CTG,CGA和CGT。一旦设计了通用公式,例如NN或NNN,则包含N的所有可能组合的群可购得,例如购自Intergrated DNA Technologies (IDT),Sigma和Invitrogen。

[0146] 寡聚体能够根据本发明连接在一起。群中的所有寡聚体优选在5'端具有磷酸基团或腺苷酸基团。

[0147] 杂交的寡聚体可以使用本领域已知的任何方法连接在一起。寡聚体优选使用连接酶连接,例如T4DNA连接酶,大肠杆菌DNA连接酶,Taq DNA连接酶,Tma DNA连接酶和 9°N DNA连接酶。

[0148] 如果反应性基团存在于寡聚体的末端,则寡聚体也可以化学连接。在这样的实施方案中,需要采取步骤以防止寡聚体在溶液中彼此结合。通常使用链上的发夹作为引物来引发连接反应。

[0149] 在优选的实施方案中,该方法优选包括使多个双链构建体与聚合酶接触,使得聚合酶同时分离双链构建体的两条链,并将该链用作模板以形成多个修饰的双链多核苷酸。可以使用上面或下面讨论的任何聚合酶。聚合酶可以形成包含上面或下面讨论的任何核苷酸的新链。聚合酶具有与模板链中的核苷酸互补的游离核苷酸群。聚合酶可以使用游离核苷酸形成新的链。

[0150] 经修饰的多核苷酸

[0151] 如果聚合酶使用链作为模板以形成多个修饰的双链多核苷酸,则该方法可以包括在聚合酶使用链作为模板以形成多个修饰的双链多核苷酸的条件使链与聚合酶以及游离核苷酸群接触,其中当形成修饰的双链多核苷酸时,所述聚合酶用不同的核苷酸种类代替链中的一个或多个核苷酸种类。如上所述,聚合酶可以用于同时分离各链。在申请号为

1403096.9的英国申请中描述了这种类型的修饰。可以使用上面或下面讨论的任何聚合酶。聚合酶优选为 Klenow或9°North。以上讨论了合适的条件。

[0152] 使用跨膜孔进行的多核苷酸的表征例如测序通常涉及分析由k个核苷酸组成的聚合物单元,其中k是正整数(即'k聚体')。这在申请号为PCT/GB2012/052343(公开为W0 2013/041878)的国际申请中进行了讨论。尽管希望在不同k聚体的电流测量值之间有明确的分离,但是这些测量值中的一些通常是重叠的。特别是在k聚体中具有高数量的聚合物单元,即高的k值时,可能变得难以分辨由不同的k聚体产生的测量值,从而不利于得到关于多核苷酸的信息,例如对多核苷酸的潜在序列的估计。

[0153] 通过用修饰的双链多核苷酸的新链(即使用聚合酶产生的链)中的不同核苷酸种类替换链中的一个或多个核苷酸种类,新链含有不同于那些在模板链中的k聚体。新链中的不同的k聚体能够从模板链中的k聚体产生不同的电流测量值,因此新的链提供与模板链不同的信息。来自新链的额外信息可以使得更容易表征修饰的双链多核苷酸以及模板多核苷酸。在一些情况下,修饰的双链多核苷酸本身可能更易于表征。例如,修饰的双链多核苷酸可以被设计成包括具有在其电流测量值之间的增加的分离或清晰的分离的k聚体或具有降低的噪声的k聚体。

[0154] 当形成修饰的双链多核苷酸时,聚合酶优选不同核苷酸种类取代模板链中的两个或更多个核苷酸种类。聚合酶可以用不同的核苷酸种类代替模板链中的两个或更多个核苷酸种类中的每一个。聚合酶可以用相同的核苷酸种类代替模板链中的两个或更多个核苷酸种类中的每一个。

[0155] 如果模板链是DNA,则不同的核苷酸种类通常包含不同于腺嘌呤,鸟嘌呤,胸腺嘧啶,胞嘧啶或甲基胞嘧啶的核碱基和/或包含与脱氧腺苷,脱氧鸟苷,胸苷,脱氧胞苷或脱氧甲基胞苷不同的核苷。如果模板链是RNA,则经修饰的多核苷酸中的不同核苷酸种类通常包含不同于腺嘌呤,鸟嘌呤,尿嘧啶,胞嘧啶或甲基胞嘧啶的核碱基和/或包含不同于腺苷,鸟苷,尿苷,胞苷或甲基胞苷的核苷。

[0156] 不同的核苷酸种类可以是通用核苷酸。通用核苷酸是在某种程度上与模板链中的所有核苷酸杂交或结合的核苷酸。通用核苷酸优选是一定程度上与包含核苷腺苷(A),胸腺嘧啶(T),尿嘧啶(U),鸟嘌呤(G)和胞嘧啶(C)的核苷酸杂交或结合的核苷酸。通用核苷酸可以比其他核苷酸更强地杂交或结合某些核苷酸。例如,包含核苷,2'-脱氧肌苷的通用核苷酸(I)将显示 $I-C > I-A > I-G \approx I-T$ 的配对的优先顺序。如果通用核苷酸替代了群中的核苷酸种类,则聚合酶将用通用核苷酸替代核苷酸种类。例如,如果与一群游离dAMP,dTMP,dCMP和通用核苷酸接触,聚合酶将以通用核苷酸取代dGMP。

[0157] 通用核苷酸优选包含以下核碱基之一:次黄嘌呤,4-硝基吡啶,5-硝基吡啶,6-硝基吡啶,甲酰基吡啶,3-硝基吡咯,硝基咪唑,4-硝基吡唑,4-硝基苯并咪唑,5-硝基吡唑,4-氨基苯并咪唑或苯基(C6-芳环)。通用核苷酸更优选包含以下核苷之一:2'-脱氧肌苷,肌苷,7-脱氮-2'-脱氧肌苷,7-脱氮肌苷,2-氮杂-脱氧肌苷,2-氮杂-肌苷,2-O'-甲基肌苷,4-硝基吡啶2'-脱氧核糖核苷,4-硝基吡啶核糖核苷,5-硝基吡啶2'-脱氧核糖核苷,5-硝基吡啶核糖核苷,6-硝基吡啶2'-脱氧核糖核苷,6-硝基吡啶核糖核苷,3-硝基吡咯2'-脱氧核糖核苷,3-硝基吡咯核糖核苷,次黄嘌呤的无环糖类似物,硝基咪唑2'-脱氧核糖核苷,硝基咪唑核糖核苷,4-硝基吡唑2'-脱氧核糖核苷,4-硝基吡唑核糖核苷,4-硝基苯并咪唑2'-脱

氧核糖核苷, 4-硝基苯并咪唑核糖核苷, 5-硝基咪唑2'-脱氧核糖核苷, 5-硝基咪唑核糖核苷, 4-氨基苯并咪唑2'-脱氧核糖核苷, 4-氨基苯并咪唑核糖核苷, 苯基C-核糖核苷, 苯基C-2'-脱氧核糖核苷, 2'-脱氧水粉草素, 2'-脱氧异鸟苷, K-2'-脱氧核糖, P-2'-脱氧核糖和吡咯烷。通用核苷酸更优选包含2'-脱氧肌苷。通用核苷酸更优选为IMP或dIMP。通用核苷酸最优选为dPMP (2'-脱氧-P-核苷单磷酸) 或dKMP (N6-甲氧基-2,6-二氨基嘌呤单磷酸酯)。

[0158] 不同的核苷酸种类优选包含其所取代的核苷酸种类中不存在的化学原子或基团。化学基团优选为丙炔基, 硫基, 氧代基, 甲基, 羟甲基, 甲酰基, 羧基, 羰基, 苄基, 炔丙基或炔丙胺基。化学基团或原子可以是或可以包含荧光分子, 生物素, 地高辛, DNP (二硝基苯酚), 光不稳定基团, 炔炔, DBCO, 叠氮化物, 游离氨基, 氧化还原染料, 汞原子或硒原子。

[0159] 包含不存在于天然存在的核苷中的化学基团的市售核苷包括但不限于6-硫代-2'-脱氧鸟苷, 7-脱氮-2'-脱氧腺苷, 7-脱氮-2'-脱氧鸟苷, 7-脱氮-2'-脱氧腺苷, 7-脱氮-8-氮杂-2'-脱氧腺苷, 8-5' (5'S) -环-2'-脱氧腺苷, 8-氨基-2'-脱氧腺苷, 8-氨基-2'-脱氧鸟苷, 8-氘代-2'-脱氧鸟苷, 8-氧-2'-脱氧腺苷, 8-氧-2'-脱氧鸟苷, 亚乙烯基-2'-脱氧腺苷, N6-甲基-2'-脱氧腺苷, O6-甲基-2'-脱氧鸟苷, O6-苯基-2'-脱氧肌苷, 2'-脱氧假尿苷, 2-硫代胸苷, 4-硫代-2'-脱氧尿苷, 4-硫代胸苷, 5'氨基胸苷, 5-(1-苊基乙炔基)-2'-脱氧尿苷, 5-(C2-EDTA)-2'-脱氧尿苷, 5-(羧基) 乙烯基-2'-脱氧尿苷, 5,6-二氢-2'-脱氧尿苷, 5,6-二氢胸苷, 5-溴-2'-脱氧胞苷, 5-溴-2'-脱氧尿苷, 5-羧基-2'-脱氧胞苷, 5-氟-2'-脱氧尿苷, 5-甲酰基-2'-脱氧胞苷, 5-羟基-2'-脱氧胞苷, 5-羟基-2'-脱氧尿苷, 5-羟甲基-2'-脱氧胞苷, 5-羟甲基-2'-脱氧尿苷, 5-碘-2'-脱氧胞苷, 5-碘-2'-脱氧尿苷, 5-甲基-2'-脱氧胞苷, 5-甲基-2'-脱氧异胞苷, 5-丙炔基-2'-脱氧胞苷, 5-丙炔基-2'-脱氧尿苷, 6-O-(TMP)-5-F-2'-脱氧尿苷, C4-(1,2,4-三唑-1-基)-2'-脱氧尿苷, C8-炔-胸苷, dT-二茂铁, N4-乙基-2'-脱氧胞苷, O4-甲基胸苷, 吡咯-2'-脱氧胞苷, 胸苷乙二醇, 4-硫尿苷, 5-甲基胞苷, 5-甲基尿苷, 吡咯基胞苷, 3-脱氮-5-氮杂-2'-0-甲基胞苷, 5-氟-2'-0-甲基尿苷, 5-氟-4-0-TMP-2'-0-甲基尿苷, 5-甲基-2'-0-甲基胞苷, 5-甲基-2'-0-甲基胸苷, 2', 3'-脱氧腺苷, 2', 3'-二脱氧胞苷, 2', 3'-二脱氧鸟苷, 2', 3'-二脱氧胸苷, 3'-脱氧腺苷, 3'-脱氧胞苷, 3'-脱氧鸟苷, 3'-脱氧胸苷和5'-0-甲基胸苷。不同的核苷酸种类可以包含任何这些核苷。

[0160] 或者, 不同的核苷酸种类优选缺乏其所替代的核苷酸种类中存在的化学基团或原子。

[0161] 与被替换的一个或多个核苷酸相比, 所述不同的核苷酸种类优选具有改变的电负性。具有改变的电负性的不同核苷酸种类优选包含卤素原子。卤素原子可以连接到不同核苷酸种类上的任何位置, 例如核碱基和/或糖。卤素原子优选为氟(F), 氯(Cl), 溴(Br) 或碘(I)。卤素原子最优选为F或I。

[0162] 包含卤素的市售核苷包括但不限于8-溴-2'-脱氧腺苷, 8-溴-2'-脱氧鸟苷, 5-溴尿苷, 5-碘尿苷, 5'-碘胸苷和5-溴-2'-0-甲基尿苷。不同的核苷酸种类可以包含任何这些核苷。

[0163] 该方法优选还包括选择性地从修饰的双链多核苷酸中的一个或多个不同核苷酸种类中去除核碱基。这导致修饰的双链多核苷酸中的脱碱基核苷酸。脱碱基核苷酸是缺乏核碱基的核苷酸。脱碱基核苷酸通常含有糖和至少一个磷酸基。糖通常是戊糖, 如核糖和脱

氧核糖。脱碱基核苷酸通常是脱碱基核糖核苷酸或脱碱基脱氧核糖核苷酸。脱碱基核苷酸通常含有一磷酸盐,二磷酸盐或三磷酸盐。磷酸盐可以连接在无碱基核苷酸的5'或3'侧。

[0164] 可以使用本领域已知的任何方法选择性地除去核碱基。例如,某些DNA修复蛋白(例如人烷基腺嘌呤DNA糖基化酶(hAAG))能够从核苷酸中选择性地除去3-甲基腺嘌呤,7-甲基鸟嘌呤,1,N6-乙烯基腺嘌呤和次黄嘌呤。此外,可以使用尿嘧啶DNA糖基化酶选择性除去dUMP。

[0165] 附加聚合酶步骤

[0166] 在另一个优选的实施方案中,将修饰的双链多核苷酸中的信息量加倍以利于模板多核苷酸的表征。该方法的一个示例在图8中示出。该方法优选包含(d)将修饰的双链多核苷酸的两条链分离并使用该链作为模板以形成多个适应的双链多核苷酸,每个双链多核苷酸包含由至少一个发夹环连接的两个互补链,其中每个互补链包含两个互补序列。每个互补链中的两个互补序列之一衍生自模板双链多核苷酸。步骤(d)通常包括,在分离之前将发夹环在修饰的双链多核苷酸的远离连接互补链的所述至少一个发夹环的另一端连接到修饰的双链多核苷酸。该发夹环优选不连接经修饰的双链多核苷酸的链。发夹可以形成聚合酶的成核点。当修饰的双链多核苷酸的分离的链用作模板时,连接的发夹环也用作模板并连接适应的双链多核苷酸的两个互补链,即,连接来自修饰的双链多核苷酸的模板链与由模板形成的新链。

[0167] 步骤(d)可以以上述任何方式进行。例如,步骤(d)可以包括通过增加pH,温度和离子强度中的一个或多个来分离修饰的双链多核苷酸的两条链。步骤(d)可以包括使分离的链与聚合酶接触,使得聚合酶使用链作为模板来形成多个适应的双链多核苷酸。步骤(d)可以包括(i)使多个分离的链与一群核苷酸寡聚体接触,所述一群核苷酸寡聚体包括在寡聚体能够杂交到链的条件下与链中所有核苷酸互补的核苷酸的每种可能组合,和(ii)将那些与所述链杂交的寡聚体连接在一起以形成多个适应的双链多核苷酸。步骤(d)可以包括使多个修饰的双链多核苷酸与聚合酶接触,使得聚合酶同时分离修饰的双链多核苷酸的两条链,并将该链用作多个适应的双链多核苷酸的模板。上述任何实施例可以应用于步骤(d)。例如,步骤(d)可以包括用新链中的不同核苷酸种类替换模板链中的一个或多个核苷酸种类。

[0168] Y适配器

[0169] 如果每个底物不包含前导序列,则该方法优选还包括在发夹环的相对端将Y适配器连接到多个修饰的双链多核苷酸。Y适配器通常是多核苷酸适配器。它们可以由任何上述的多核苷酸形成。Y适配器通常包含(a)双链区域和(b)单链区域或在另一端不互补的区域。如果Y适配器包括单链区域,则Y适配器可以被描述为具有突出端。Y适配器中非互补区域的存在使其具有Y形状,因为与双链部分不同,这两条链通常不彼此杂交。Y适配器可以包括一个或多个锚,如下面更详细地讨论的。

[0170] Y适配器可以连接到修饰的双链多核苷酸。可以使用本领域已知的任何方法进行连接。例如,可以使用连接酶例如T4 DNA连接酶,大肠杆菌DNA连接酶,Taq DNA连接酶,Tma DNA连接酶和9°N DNA连接酶连接Y适配器。

[0171] 本发明的产品

[0172] 本发明还提供了用于修饰模板多核苷酸的双链MuA底物群,其中每个底物包含通

用核苷酸的至少一个突出端。本发明还提供了用于修饰模板多核苷酸的双链MuA底物群,其中每个底物包含(i)至少一个突出端和(ii)在与包含所述至少一个突出端的链相对的链中的至少一个发夹环。底物可以是上述那些中的任一种。底物优选包含如上所定义的双链部分。双链部分优选包含如上所述的SEQ ID NO:26和27。双链部分更优选包含如上所述的SEQ ID NO:26和28。本发明的优选的群是其中每个底物包含在一端的突出端和在另一端的发夹环的那些群。

[0173] 本发明还提供了使用本发明的方法修饰的多个多核苷酸。所述多个多核苷酸可以是上述任何形式。修饰的双链多核苷酸包含两个互补链,其包含通过发夹环连接的模板多核苷酸的双链片段。

[0174] 群或多个多核苷酸可以是分离的,基本上分离的,纯化的或基本上纯化的。如果群或多个多核苷酸完全不含任何其它组分,例如模板多核苷酸,脂质或孔,则群或多个多核苷酸是分离的或纯化的。如果与不妨碍其预期用途的载体或稀释剂混合,则群或多个多核苷酸基本上分离的。例如,如果群或多个多核苷酸以包含小于10%,小于5%,小于2%或小于1%的其它成分(例如脂质或孔)的形式存在,则群或多个多核苷酸基本上是分离的或基本上是纯化的。

[0175] 表征方法

[0176] 本发明还提供了表征使用本发明的方法修饰的至少一种多核苷酸的方法。经修饰的多核苷酸与跨膜孔接触,使得多核苷酸的至少一条链移动通过所述孔。随着至少一条链相对于孔移动,获取一个或多个测量值。所述测量值指示所述至少一条链的一个或多个特征,并且这允许表征经修饰的多核苷酸。

[0177] 本发明还提供了表征模板多核苷酸的方法。使用本发明修饰模板多核苷酸以产生多个经修饰的多核苷酸。每个经修饰的多核苷酸与跨膜孔接触,使得每个多核苷酸的至少一条链移动通过孔。随着每个多核苷酸相对于孔移动,获取一个或多个测量值。所述测量值指示每个多核苷酸的一个或多个特征,这允许对模板多核苷酸进行表征。

[0178] 在优选的实施方案中,经修饰的多核苷酸的两条链/每个经修饰的多核苷酸移动通过孔。如果两条链都移动穿过孔,则两条链通常是分离的。可以使用本领域已知的任何方法分离两条链。例如,它们可以被多核苷酸结合蛋白分离,或者使用有利于脱杂交的条件(有利于脱杂交的条件的实例包括但不限于高温、高pH和添加可以破坏氢键或碱基配对的试剂,如甲酰胺和脲)而分离。

[0179] 跨膜孔

[0180] 跨膜孔是在一定程度上穿过膜的结构。它允许由施加电势驱动的水合离子流过膜或在膜内流动。跨膜孔通常穿过整个膜,使得水合离子可以从膜的一侧流到膜的另一侧。然而,跨膜孔不必跨过膜。它可能在一端被封闭。例如,孔可以是水合离子可以沿其流动或流入其中的膜内的阱、间隙、通道、沟槽或狭缝。

[0181] 一个或多个选择性扩增探针或一个或多个扩增产物优选通过下述进行表征:(i)使探针或扩增产物与跨膜孔接触,使得探针或扩增产物移动通过孔,并且(ii)随着探针或扩增产物相对于孔移动获取一个或多个测量值,其中所述测量值指示探针或扩增产物的一个或多个特征,从而表征探针或扩增产物。

[0182] 任何跨膜孔可用于本发明。孔可以是生物的或人造的。合适的孔包括但不限于蛋

白质孔,多核苷酸孔和固态孔。孔可以是DNA折纸孔(Langecker等人,Science,2012;338:932-936)。

[0183] 跨膜孔优选为跨膜蛋白孔。

[0184] 根据本发明使用的跨膜蛋白孔可以衍生自 β -桶状孔或 α -螺旋束孔。 β -桶状孔包括由 β -链形成的桶状体或通道。合适的 β -桶状孔包括但不限于,形成毒素的 β 孔,例如 α -溶血素,炭疽毒素和杀白细胞素,和细菌的外膜蛋白/孔蛋白,例如耻垢分枝杆菌(*Mycobacterium smegmatis*)孔蛋白(Msp),如MspA,MspB,MspC或MspD,外膜孔蛋白F(OmpF),外膜孔蛋白G(OmpG),外膜磷脂酶A和奈瑟氏菌(*Neisseria*)自体转运脂蛋白(NalP)以及其他孔,如lysenin。 α -螺旋束孔包括由 α -螺旋形成的桶状体或通道。合适的 α -螺旋束孔包括但不限于内膜蛋白和 α 外膜蛋白,例如WZA和ClyA毒素。跨膜孔可以衍生自lysenin。衍生自lysenin的合适的孔在申请号为PCT/GB2013/050667(公开为WO 2013/153359)的国际申请中公开。跨膜孔可以衍生自Msp,如MspA,或衍生自 α -溶血素(α -HL)。野生型 α -HL孔由七个相同的单体或亚基形成(即其为七聚体)。 α -溶血素-NN的一个单体或亚基的序列在SEQ ID NO:4中示出。

[0185] 跨膜蛋白孔优选衍生自Msp,优选来自MspA。这样的孔将是低聚的,并且通常包含衍生自Msp的7,8,9或10个单体。孔可以是衍生自包含相同单体的Msp的同源寡聚孔。或者,孔可以是衍生自包含至少一种与其它单体不同的单体的Msp的异源寡聚孔。优选地,孔衍生自MspA或其同源物或旁系同源物。

[0186] 衍生自Msp的单体通常包含SEQ ID NO:2或其变体中所示的序列。SEQ ID NO:2是MspA单体的MS-(B1)8突变体。它包括以下突变:D90N,D91N,D93N,D118R,D134R和E139K。SEQ ID NO:2的变体是具有下述氨基酸序列的多肽:从SEQ ID NO:2的氨基酸序列变化而来并保留其形成孔的能力。合适的变体在申请号为PCT/GB2012/050301(公开为WO 2012/107778)的国际申请和申请号为1407809.1(ONT IP 057)的英国申请中公开。SEQ ID NO:2的优选变体包含N93D。可以使用本领域已知的任何方法测定变体形成孔的能力。例如,可以将变体与其他合适的亚基一起插入两亲层,并且可以确定其低聚以形成孔的能力。本领域已知用于将亚基插入诸如两亲层的膜中的方法。例如,亚基可以在含有三嵌段共聚物膜的溶液中以纯化的形式悬浮,使得其扩散到膜并通过结合到膜上并组装成功能状态而插入。或者,亚基可以使用M.A.Holden,H.Bayley,J.Am.Chem.Soc.2005,127,6502-6503和申请号为PCT/GB2006/001057(公开为WO 2006/100484)的国际申请中所述的“摘取和放置(pick and place)”方法直接插入膜中。

[0187] 在SEQ ID NO:2的氨基酸序列的整个长度上,基于氨基酸同一性,变体将优选与该序列至少50%同源。更优选地,基于氨基酸同一性,变体可以与SEQ ID NO:2的氨基酸序列在整个长度上至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,和更优选地至少95%,97%或99%同源。在100个或更多个,例如125,150,175或200个或更多个连续氨基酸的片段上可以具有至少80%,例如至少85%,90%或95%的氨基酸同一性(“严格同源性(hard homology)”)。

[0188] 本文所述的任何蛋白质,例如跨膜蛋白孔,可以通过合成而制备或通过重组方法制备。例如,孔可以通过体外翻译和转录(IVTT)合成。孔的氨基酸序列可以被修饰为包括非天然存在的氨基酸或增加蛋白质的稳定性。当通过合成方法产生蛋白质时,可以在生产过

程中引入这些氨基酸。在合成或重组生产之后,孔也可以改变。

[0189] 表征

[0190] 该方法可以涉及测量经修饰多核苷酸或模板多核苷酸的两个,三个,四个或五个或更多个特征。一个或多个特征优选选自 (i) 多核苷酸的长度, (ii) 多核苷酸的同一性, (iii) 多核苷酸的序列, (iv) 多核苷酸的二级结构, 以及 (v) 多核苷酸是否被修饰。(i) 至 (v) 的任何组合可以根据本发明进行测量, 例如 {i}, {ii}, {iii}, {iv}, {v}, {i, ii}, {i, iii}, {i, iv}, {i, v}, {ii, iii}, {ii, iv}, {ii, v}, {iii, iv}, {iii, v}, {iv, v}, {i, ii, iii}, {i, ii, iv}, {i, ii, v}, {i, iii, iv}, {i, iii, v}, {i, iv, v}, {ii, iii, iv}, {ii, iii, v}, {ii, iv, v}, {iii, iv, v}, {i, ii, iii, iv}, {i, ii, iii, v}, {i, ii, iv, v}, {i, iii, iv, v}, {ii, iii, iv, v} 或 {i, ii, iii, iv, v}。相比于第二多核苷酸, 可以对第一多核苷酸测量 (i) 至 (v) 的不同组合, 包括上述任何一种组合。

[0191] 对于 (i), 可以例如通过确定多核苷酸和孔之间的相互作用的次数或多核苷酸和孔之间的相互作用持续时间来测量多核苷酸的长度。

[0192] 对于 (ii), 可以以多种方式测量多核苷酸的同一性。多核苷酸的同一性可以结合多核苷酸序列的测量或不结合多核苷酸的序列的测量来测量。前者较简单; 对多核苷酸进行测序并鉴定。后者可以通过几种方式完成。例如, 可以测量多核苷酸中特定基序的存在 (不测量多核苷酸的剩余序列)。或者, 该方法中特定电和/或光信号的测量可以鉴定来自特定来源的多核苷酸。

[0193] 对于 (iii), 可以如先前所述测定多核苷酸的序列。在 Stoddart D 等人, Proc Natl Acad Sci, 12;106 (19):7702-7, Lieberman KR 等人, J Am Chem Soc. 2010;132 (50):17961-72 和申请号为 WO 2000/28312 的国际申请中描述了合适的测序方法, 特别是使用电测量的测序方法。

[0194] 对于 (iv), 二级结构可以以多种方式测量。例如, 如果该方法涉及电测量, 则可以使用驻留时间的变化或流过孔的电流的变化来测量二级结构。这允许区分单链和双链多核苷酸的区域。

[0195] 对于 (v), 可以测量是否存在任何修饰。该方法优选包括, 确定多核苷酸是否通过甲基化, 氧化, 损伤、用一种或多种蛋白质或用一种或多种标记、标签或间隔基, 进行了修饰。特定的修饰将导致与孔的特异性相互作用, 其可以使用下述方法测量。例如, 可基于在与每个核苷酸相互作用期间流过孔的电流, 将甲基胞嘧啶与胞嘧啶进行区分。

[0196] 多核苷酸与跨膜孔接触。孔通常存在于膜中。下面讨论合适的膜。该方法可以使用任何适合于研究膜中存在孔的膜/孔系统的设备进行。该方法可以使用适用于跨膜孔感测的任何设备来进行。例如, 该设备包括一个包含水性溶液的腔室和一个将腔室分成两个部分的屏障。屏障通常具有形成包含孔的膜的孔洞。或者, 屏障形成其中存在孔的膜。

[0197] 该方法可以使用申请号为 PCT/GB08/000562 (WO 2008/102120) 的国际申请中描述的设备进行。

[0198] 可以进行多种不同类型的测量。这包括但不限于: 电测量和光测量。可能的电测量包括: 电流测量, 阻抗测量, 隧道测量 (tunnelling measurement) (Ivanov AP 等人, Nano Lett. 2011 Jan 12;11 (1):279-85) 和 FET 测量 (申请号为 WO 2005/124888 的国际申请)。光测量可以与电测量结合 (Soni GV 等人, Rev Sci Instrum. 2010 Jan;81 (1):014301)。测量可

以是跨膜电流测量,例如流过孔的离子电流的测量。

[0199] 可以使用Stoddart D等人,Proc Natl Acad Sci,12;106(19):7702-7,Lieberman KR等人,J Am Chem Soc.2010;132(50):17961-72,和申请号为W0 2000/28312的国际申请中描述的标准单通道记录装置进行电测量。或者,可以使用如国际申请W0 2009/077734和国际申请W0 2011/067559中所述的多通道系统进行电测量。

[0200] 该方法优选以跨膜施加的电势进行。所施加的电势可以是电压电势。或者,施加的电势可以是化学电势。其一个例子是使用跨膜的盐梯度,所述膜例如两亲层。在Holden等人,J Am Chem Soc.2007Jul 11;129(27):8650-5中公开了盐梯度。在一些情况下,随着多核苷酸相对于孔的移动,使用通过孔的电流来估计或确定多核苷酸的序列。这就是链测序。

[0201] 该方法可以涉及测量随着多核苷酸相对于孔的移动通过孔的电流。因此,该方法中使用的设备还可以包括能够在膜和孔上施加电势并测量电信号的电路。该方法可以使用膜片钳或电压钳进行。所述方法优选地涉及使用电压钳。

[0202] 本发明的方法可以涉及测量随着多核苷酸相对于孔的移动通过孔的电流。测量通过跨膜蛋白孔的离子电流的合适条件是本领域已知的并且在实施例公开。该方法通常通过施加在膜和孔上的电压进行。所使用的电压通常为+5V至-5V,例如从+4V至-4V,+3V至-3V或+2V至-2V。所使用的电压通常为-600mV至+600mV或-400mV至+400mV。所使用的电压优选在具有选自-400mV,-300mV,-200mV,-150mV,-100mV,-50mV,-20mV和0mV的下限和独立地选自+10mV,+20mV,+50mV,+100mV,+150mV,+200mV,+300mV和+400mV的上限的范围内。所使用的电压更优选在100mV至240mV的范围内,最优选在120mV至220mV的范围内。通过使用增加的施加电势,可以由孔增加不同核苷酸之间的区别。

[0203] 该方法通常在任何电荷载体存在下进行,例如金属盐,例如碱金属盐,卤化物盐,例如氯化物盐,如碱金属氯化物盐。电荷载体可以包括离子液体或有机盐,例如四甲基氯化铵,三甲基苯基氯化铵,苯基三甲基氯化铵或1-乙基-3-甲基咪唑鎓氯化物。在上述示例性设备中,盐存在于腔室中的水性溶液中。通常使用氯化钾(KCl),氯化钠(NaCl),氯化铯(CsCl)或亚铁氰化钾和铁氰化钾的混合物。KCl,NaCl和亚铁氰化钾和铁氰化钾的混合物是优选的。电荷载体在膜上可能是不对称的。例如,电荷载体的类型和/或浓度在膜的每一侧可以是不同的。

[0204] 盐浓度可以是饱和的。盐浓度可以是3M或更低,通常为0.1至2.5M,0.3至1.9M,0.5至1.8M,0.7至1.7M,0.9至1.6M或1M至1.4M。盐浓度优选为150mM至1M。该方法优选使用至少为0.3M,例如至少0.4M,至少0.5M,至少0.6M,至少0.8M,在至少1.0M,至少1.5M,至少2.0M,至少2.5M或至少3.0M的盐浓度进行。高盐浓度提供高信噪比,并允许电流指示正常电流波动背景下待鉴定的核苷酸的存在。

[0205] 该方法通常在存在缓冲液的情况下进行。在上述示例性设备中,缓冲液存在于腔室中的水性溶液中。任何缓冲液可用于本发明的方法。通常,缓冲液是磷酸盐缓冲液。其他合适的缓冲液是HEPES和三羟甲基氨基甲烷盐酸盐(Tris-HCl)缓冲液。该方法通常在4.0至12.0,4.5至10.0,5.0至9.0,5.5至8.8,6.0至8.7或7.0至8.8或7.5至8.5的pH下进行。所用的pH优选为约7.5。

[0206] 该方法可以在0℃至100℃,15℃至95℃,16℃至90℃,17℃至85℃,18℃至80℃,19℃至70℃或在20℃至60℃进行。所述方法通常在室温下进行。所述方法任选地在支持酶

功能的温度下进行,例如约37℃。

[0207] 多核苷酸结合蛋白

[0208] 该方法优选包括使所述/每个多核苷酸与多核苷酸结合蛋白接触,使得所述蛋白控制所述/每个多核苷酸的至少一条链通过孔的运动。

[0209] 更优选地,该方法包括(a)使所述/每个多核苷酸与孔以及多核苷酸结合蛋白接触,使得所述蛋白控制所述/每个多核苷酸的至少一条链通过孔的运动,和(b)随着所述/每个多核苷酸相对于孔移动,获取一个或多个测量值,其中测量值指示所述/每个多核苷酸的一个或多个特征,从而表征修饰的多核苷酸或模板多核苷酸。

[0210] 多核苷酸结合蛋白可以是能够结合到多核苷酸并控制其通过孔的运动的任何蛋白质。确定蛋白质是否与多核苷酸结合在本领域是较简单的。蛋白质通常与多核苷酸相互作用并修饰多核苷酸的至少一种特性。蛋白质可以通过将多核苷酸裂解以形成单个核苷酸或核苷酸短链,例如二核苷酸或三核苷酸,来修饰多核苷酸。蛋白质可以通过使其定向到特定位置或将其移动到特定位置,即控制其运动,来修饰多核苷酸。

[0211] 多核苷酸结合蛋白优选衍生自多核苷酸处理酶。多核苷酸处理酶是能够与多核苷酸相互作用并修饰多核苷酸的至少一种性质的多肽。酶可以通过将多核苷酸裂解以形成单个核苷酸或核苷酸短链,例如二核苷酸或三核苷酸,来修饰多核苷酸。酶可以通过使其定向到特定位置或将其移动到特定位置来修饰多核苷酸。多核苷酸处理酶不需要显示酶活性,只要其能够结合多核苷酸并控制其通过孔的移动即可。例如,酶可以被修饰以去除其酶活性,或者可以在防止其用作酶的条件下使用。以下更详细地讨论这些条件。

[0212] 多核苷酸处理酶优选衍生自溶核酶(nucleolytic enzyme)。酶的构建体中使用多核苷酸处理酶更优选衍生自酶分类(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中任一成员。酶可以是申请号为PCT/GB10/000133(公开为W0 2010/086603)的国际申请中公开的那些酶中的任一种。

[0213] 优选的酶是聚合酶、核酸外切酶、解旋酶和拓扑异构酶,如促旋酶。合适的酶包括但不限于来自大肠杆菌的核酸外切酶I(SEQ ID NO:11),来自大肠杆菌的核酸外切酶III(SEQ ID NO:13),来自嗜热栖热菌的RecJ(SEQ ID NO:15),以及细菌噬菌体λ核酸外切酶(SEQ ID NO:17),TatD核酸外切酶及其变体。包含SEQ ID NO:15或其变体中所示序列的三个亚基相互作用以形成三聚体核酸外切酶。聚合酶可以是PyroPhage®3173 DNA聚合酶(可从Lucigen®Corporation购得),SD聚合酶(可从Bioron®购得)或其变体。酶优选为Phi29 DNA聚合酶(SEQ ID NO:9)或其变体。拓扑异构酶优选为酶分类(EC)组5.99.1.2和5.99.1.3中的任何成员。

[0214] 该酶最优选衍生自解旋酶,例如He1308Mbu(SEQ ID NO:18),He1308Csy(SEQ ID NO:19),He1308Tga(SEQ ID NO:20),He1308Mhu(SEQ ID NO:21),TraI Eco(SEQ ID NO:22),XPD Mbu(SEQ ID NO:23)或其变体。在本发明中可以使用任何解旋酶。解旋酶可以是或来自He1308解旋酶,RecD解旋酶,例如TraI解旋酶或TrwC解旋酶,XPD解旋酶或Dda解旋酶。解旋酶可以是申请号为PCT/GB2012/052579(公开为W0 2013/057495),PCT/GB2012/053274(公开为W0 2013/098562),PCT/GB2012/053273(公开为W02013098561),PCT/GB2013/051925(公开为W0 2014/013260),PCT/GB2013/051924(公开为W0 2014/013259),PCT/GB2013/051928(公开为W0 2014/013262)和PCT/GB2014/052736的国际申请中公开的

任何解旋酶,修饰的解旋酶或解旋酶构建体。

[0215] 解旋酶优选包含SEQ ID NO:25 (Trwc Cba) 或其变体中所示的序列,SEQ ID NO:18 (He1308Mbu) 或其变体中所示的序列,或者SEQ ID NO:24 (Dda) 或其变体中所示的序列。变体可不同于以下以任何方式讨论的跨膜孔的天然序列。SEQ ID NO:24的优选变体包括(a) E94C和A360C或(b) E94C,A360C,C109A和C136A,然后可选地(Δ M1) G1G2 (即M1缺失,然后添加G1和G2)。

[0216] 在链测序中,多核苷酸顺着或逆着施加的电势被移位穿过孔。在双链多核苷酸上逐渐或进行性地起作用的核酸外切酶可以用在孔的顺式侧以使剩余的单链在施加电势下通过或在反式侧在反向电势下使其通过。同样地,解开双链DNA的解旋酶也可以以类似的方式使用。也可以使用聚合酶。对于测序应用还有可能需要逆着施加的电势的链移位,但DNA必须首先在反向电势或无电势的条件下被酶“捕获”。然后随着电势在结合后切换回来,链将沿顺式到反式通过所述孔并通过电流保持为扩展的构象。单链DNA核酸外切酶或单链DNA依赖性聚合酶可以作为分子马达,以将最近移位的单链逆着施加的电势按反式到顺式以受控的逐步方式穿过细孔拉回。

[0217] 任何解旋酶可用于该方法中。解旋酶可以相对于孔以两种模式工作。首先,该方法优选使用解旋酶进行,使得其利用由施加的电压产生的场使多核苷酸移动通过孔。在该模式下,首先将多核苷酸的5'端捕获在孔中,然后解旋酶将多核苷酸移动到孔中,使得其利用场通过孔,直到其最终移位到膜的反式侧。或者,优选进行该方法,使得解旋酶逆着由施加的电压产生的场将多核苷酸移动通过孔。在该模式中,首先将多核苷酸的3'端捕获在孔中,然后解旋酶将多核苷酸移动通过孔,使得其逆着施加的场被从孔中拉出,直到最终被驱逐回膜的顺式侧。

[0218] (一种或多种)解旋酶和(一种或多种)分子制动器

[0219] 在优选的实施例中,该方法包括:

[0220] (i) 向所述/每个多核苷酸提供一个或多个解旋酶和连接到所述/每个多核苷酸的一个或多个分子制动器;

[0221] (b) 使所述/每个多核苷酸与跨膜孔接触并且跨孔施加电势,使得一个或多个解旋酶和一个或多个分子制动器被聚集在一起,并且两者都控制所述/每个多核苷酸的至少一条链通过孔的运动;

[0222] (c) 随着所述/每个多核苷酸相对于孔移动,获取一个或多个测量值,其中所述测量值指示多核苷酸的一个或多个特征,从而表征经修饰的多核苷酸或模板多核苷酸。

[0223] 这种类型的方法在国际申请PCT/GB2014/052737中有详细讨论。

[0224] 间隔基

[0225] 如申请号为PCT/GB2014/050175 (公开为W02014/135838) 的国际申请中所述,一个或多个解旋酶可以停滞在一个或多个间隔基处。在本发明中可以使用该国际申请中公开的一个或多个解旋酶和一个或多个间隔基的任何构象。

[0226] 当多核苷酸的一部分进入孔并沿着由施加的电势产生的场移动通过孔时,随着多核苷酸移动通过孔,一个或多个解旋酶通过孔被移动通过间隔基。这是因为多核苷酸(包括一个或多个间隔基) 移动通过孔并且一个或多个解旋酶保留在孔的顶部。

[0227] 一个或多个间隔基优选是多核苷酸的一部分,例如它们中断所述多核苷酸的序

列。一个或多个间隔基优选不是与多核苷酸杂交的一个或多个阻断分子(例如减速带(speed bump))的一部分。

[0228] 多核苷酸中可以存在任何数目的间隔基,例如1,2,3,4,5,6,7,8,9,10或更多个间隔基。多核苷酸中优选有两个,四个或六个间隔基。一个或多个间隔基优选地在Y适配器或前导序列中。多核苷酸的不同区域中可以有一个或多个间隔基,例如Y适配器和/或发夹环适配器中的一个或多个间隔基。

[0229] 一个或多个间隔基各自提供一个或多个解旋酶即使在活动模式下也不能克服的能量屏障。一个或多个间隔基可以通过减少解旋酶的牵引力(例如通过从多核苷酸中的核苷酸中去除碱基)或物理阻断一个或多个解旋酶的运动(例如使用庞大的化学基团)来使一个或多个解旋酶停滞。

[0230] 一个或多个间隔基可以包含使一个或多个解旋酶停滞的任何分子或分子的组合。一个或多个间隔基可以包含阻止一个或多个解旋酶沿多核苷酸移动的任何分子或分子的组合。在不存在跨膜孔和施加的电势的情况下,确定一个或多个解旋酶是否停滞在一个或多个间隔基处是较为容易的。例如,通过PAGE可以测量解旋酶移动通过间隔基并置换DNA的互补链的能力。

[0231] 一个或多个间隔基通常包含直链分子,例如聚合物。一个或多个间隔基通常具有与多核苷酸不同的结构。例如,如果多核苷酸是DNA,则一个或多个间隔基通常不是DNA。特别地,如果多核苷酸是脱氧核糖核酸(DNA)或核糖核酸(RNA),则一个或多个间隔基优选包含肽核酸(PNA),甘油核酸(GNA),苏糖核酸(TNA),锁定核酸(LNA)或具有核苷酸侧链的合成聚合物。一个或多个间隔基可以包含与多核苷酸方向相反的一个或多个核苷酸。例如,当多核苷酸处于5'至3'方向时,一个或多个间隔基可以包含3'至5'方向的一个或多个核苷酸。核苷酸可以是上面讨论的任何核苷酸。

[0232] 一个或多个解旋酶可被每个线性分子间隔基停滞(即停滞在间隔基前面)或被停滞在每个线性分子间隔基上。如果使用直链分子间隔基,则多核苷酸优选具有与每个间隔基的末端相邻的多核苷酸的双链区域,其中一个或多个解旋酶将被移动通过每个间隔基。如果使用线性分子间隔基,则多核苷酸优选在每个间隔基的与一个或多个解旋酶将被移动通过的末端相对末端设置有阻断分子。这可以帮助确保一个或多个解旋酶保持停滞在每个间隔基上。在它/它们在溶液中扩散开的情况下,还可以帮助将一个或多个解旋酶保持在多核苷酸上。阻断分子可以是物理上导致一个或多个解旋酶停滞的下述任何化学基团。阻断分子可以是多核苷酸的双链区域。阻断分子可以是BNA。

[0233] 该方法可能涉及将两个或更多个解旋酶移动通过间隔基。在这种情况下,间隔基的长度通常被增加,以防止尾部解旋酶在没有孔和施加电势的情况下推动先导解旋酶经过间隔基。如果该方法涉及移动两个或更多个解旋酶经过一个或多个间隔基,则上述间隔基长度可以增加至少1.5倍,例如2倍,2.5倍或3倍。例如,如果该方法涉及移动两个或多个解旋酶经过一个或多个间隔基,则间隔基长度可以增加1.5倍,2倍,2.5倍或3倍。

[0234] 膜

[0235] 本发明中使用的孔可以存在于膜中。在本发明的方法中,多核苷酸通常与膜中的孔接触。根据本发明,可以使用任何膜。合适的膜是本领域公知的。膜优选为两亲层。两亲层是由具有亲水性和亲油性的两亲分子形成的层,例如磷脂。两亲分子可以是合成的或天然

存在的。非天然存在的两亲物和形成单层的两亲物是本领域已知的,并且包含例如嵌段共聚物 (Gonzalez-Perez等人,Langmuir,2009,25,10447-10450)。嵌段共聚物是两个或更多个单体子单元聚合在一起形成单个聚合物链的聚合材料。嵌段共聚物通常具有由每个单体子单元贡献的性质。然而,嵌段共聚物可以具有从各个子单元形成的聚合物不具有的独特性质。嵌段共聚物可以被改造成使得单体子单元之一是疏水的(即亲油的),而其他(一个或多个)子单元在水性介质中是亲水的。在这种情况下,嵌段共聚物可以具有两亲性质并且可以形成能模拟生物膜的结构。嵌段共聚物可以是二嵌段(由两个单体子单元组成),但也可以由多于两个的单体子单元构成,以形成表现为两亲体的更为复杂的布置。共聚物可以是三嵌段,四嵌段或五嵌段共聚物。膜优选为三嵌段共聚物膜。

[0236] 膜最优选是国际申请PCT/GB2013/052766或PCT/GB2013/052767中公开的膜之一。

[0237] 两亲性分子可以被化学修饰或功能化以促进多核苷酸的偶联。

[0238] 偶联

[0239] 所述/每个经修饰的多核苷酸优选与包含孔的膜偶联。该方法可以包括将所述/每个多核苷酸偶联到包含孔的膜。优选使用一个或多个锚将多核苷酸与膜偶联。多核苷酸可以使用任何已知的方法与膜偶联。

[0240] 每个锚包含与多核苷酸偶联(或结合)的基团和与膜偶联(或结合)的基团。每个锚可以与多核苷酸和/或膜共价偶联(或结合)。优选使用Y适配器或前导序列和/或发夹环将多核苷酸偶联到膜上。

[0241] 多核苷酸可以使用任何数量的锚,例如2,3,4或更多个锚偶联到膜上。例如,多核苷酸可以使用两个锚偶联到膜上,两个锚中的每个单独地既与多核苷酸偶联(或结合)也与膜偶联(或结合)。

[0242] 一个或多个锚可以包含一个或多个解旋酶和/或一个或多个分子制动器。

[0243] 如果膜是两亲层,例如共聚物膜或脂双层,则一个或多个锚优选包含存在于膜中的多肽锚和/或存在于膜中的疏水性锚。疏水性锚优选为脂质,脂肪酸,固醇,碳纳米管,多肽,蛋白质或氨基酸,例如胆固醇,棕榈酸酯或生育酚。在优选实施例中,一个或多个锚不是孔。

[0244] 膜的组分,例如两亲分子、共聚物或脂质,可以被化学修饰或功能化以形成一个或多个锚。以下将更详细地讨论合适的化学修饰和将膜的组分功能化的合适方式的示例。任何比例的膜组分可以被功能化,例如至少0.01%,至少0.1%,至少1%,至少10%,至少25%,至少50%或100%。

[0245] 多核苷酸可以被直接偶联到膜。用于将多核苷酸偶联到膜上的一个或多个锚优选包含连接体。一个或多个锚可以包括一个或多个,例如2,3,4或更多个连接体。一个连接体可用于将多于一个,例如2,3,4或更多个多核苷酸偶联到膜上。

[0246] 优选的连接体包括但不限于聚合物,例如多核苷酸,聚乙二醇(PEG),多糖和多肽。这些连接体可以是直链,支链或环状的。例如,连接体可以是环状多核苷酸。多核苷酸可以与环状多核苷酸连接体上的互补序列杂交。

[0247] 在以下讨论的测序实施例中优选使用连接体。如果多核苷酸直接永久地偶联到膜上,在这种意义上,当与孔相互作用时不进行解偶联(即在步骤(b)或(e))中不进行解偶联),则某些序列数据将丢失,因为由于膜与孔之间的距离测序不能继续到多核苷酸的末

端。如果使用连接体,则多多核苷酸可以进行到完成为止。

[0248] 偶联可以是永久的或稳定的。换句话说,偶联可以是这样的:当多核苷酸与孔相互作用时,多核苷酸保持与膜偶联。

[0249] 偶联可能是瞬态的。换句话说,偶联是这样的:当多核苷酸与孔相互作用时,多核苷酸可以与膜解偶联。

[0250] 在申请号为PCT/GB12/051191(公开为W0 2012/164270)的国际申请和申请号为1406155.0的英国申请中公开了合适的偶联方法。

[0251] 解偶联

[0252] 本发明的方法可以包括表征多个修饰的双链多核苷酸和解偶联至少第一修饰的双链多核苷酸。

[0253] 在优选的实施方案中,本发明涉及表征两个或更多个修饰的双链多核苷酸。该方法包括:

[0254] (a) 在第一样品中提供第一经修饰的双链多核苷酸;

[0255] (b) 在第二样品中提供第二经修饰的双链多核苷酸;

[0256] (c) 使用一个或更多锚将第一样品中的第一多核苷酸与膜偶联;

[0257] (d) 使第一多核苷酸与跨膜孔接触,使得第一多核苷酸的至少一条链移动穿过孔;

[0258] (e) 随着第一多核苷酸相对于孔移动获取一个或多个测量值,其中所述测量值指示第一多核苷酸的一个或多个特征,从而表征第一个多核苷酸;

[0259] (f) 将第一个多核苷酸从膜上解偶联;

[0260] (g) 使用一个或多个锚将第二样品中的第二多核苷酸偶联到膜上;

[0261] (h) 使第二多核苷酸与孔接触,使得第二多核苷酸的至少一条链移动穿过孔;和

[0262] (i) 随着第二多核苷酸相对于孔移动而获取一个或多个测量值,其中所述测量值指示第二多核苷酸的一个或多个特征,从而表征第二多核苷酸。

[0263] 这种类型的方法在申请号为1406155.0的英国申请中有详细讨论。

[0264] 其他表征方法

[0265] 在另一个实施例中,所述/每个经修饰的双链多核苷酸通过下述进行表征:检测当聚合酶将核苷酸并入多核苷酸中时释放的标记物种。聚合酶使用多核苷酸作为模板。每个标记物种对于每个核苷酸是特异性的。所述/每个多核苷酸与跨膜孔、聚合酶和标记的核苷酸接触,使得当通过聚合酶将核苷酸加入到多核苷酸时,磷酸盐标记的物种依次释放,其中磷酸盐物种含有对每个核苷酸特异的标记。聚合酶可以是上面讨论的任何聚合酶。使用孔检测磷酸盐标记的物种,从而表征多核苷酸。这种方法在申请号为13187149.3(公布为EP 2682460)的欧洲申请中公开。以上讨论的任何实施例同样适用于该方法。

[0266] 试剂盒

[0267] 本发明还提供了用于修饰模板多核苷酸的试剂盒。试剂盒包含(a)本发明的MuA底物群和(b)MuA转座酶以及(c)聚合酶。上文参照本发明的方法和产品讨论的任何实施方案同样适用于试剂盒。

[0268] 试剂盒还可以包含膜的组分,例如两亲层或脂质双层的组分。试剂盒还可以包含跨膜孔或跨膜孔的组分。试剂盒还可以包含多核苷酸结合蛋白。上面讨论了合适的膜、孔和多核苷酸结合蛋白。

[0269] 本发明的试剂盒可以另外包含一种或多种使上述任何实施例能够进行的其它试剂或仪器。这样的试剂或仪器包括以下一种或多种：合适的缓冲液（一种或多种）（水性溶液），从受试者获得样品的设备（例如容器或包含针的仪器），用于扩增和/或表达多核苷酸的装置，如上所定义的膜，或者电压或膜片钳装置。试剂可能以干燥状态存在于试剂盒中，使得流体样品使该试剂重悬。试剂盒还可以任选地包括使试剂盒能够用于本发明方法或关于该方法可用于哪些患者的细节的说明。试剂盒可以任选地包含核苷酸。

[0270] 以下实施例说明了本发明。

[0271] 实施例1

[0272] 该实施例描述了用于修饰模板双链多核苷酸的方法，特别是使用纳米孔测序进行表征的方法。该实施例示出了MuA转座酶能够含有发夹环的MuA底物插入。然后使用聚合酶填充构建体中的间隙，然后加热双链构建体以熔化双链DNA。这得到具有发夹的单链DNA，聚合酶从该发夹产生互补体。然后将该构建体连接到具有预结合的酶的适配器，并最终与系链杂交。然后，该DNA构建体显示出，解旋酶控制DNA运动通过纳米孔。

[0273] 材料和方法

[0274] 1.1-使用MuA转座酶将DNA模板片段化

[0275] 在该实施例中使用的MuA适配器X具有5'21bp发夹（在图2中适配器标记为c，上链= SEQ ID NO:29，下链=SEQ ID NO:30，下链的5'端连接到序列GATCU的3'端）。适配器的上链和下链在10uM下，从95℃，以2℃min⁻¹，在10mM Tris pH 7.5, 50mM NaCl中退火。

[0276] MuA片段化反应（10μL）如下表1中所述进行设置，并在30℃温育1小时。然后通过75℃加热15分钟将MuA酶进行热灭活。最后将得到的DNA进行1.5×SPRI纯化并在无核酸酶的水（42uL，样品1）中洗脱。

试剂	组分
λ DNA (SEQ ID NO: 31)	50 ng μL ⁻¹
MuA 适配器 X	200 nM
MuA (Thermo)	50 nM
缓冲液	25 mM Tris-HCl pH 8.0, 10 mM MgCl ₂ , 110 mM NaCl, 0.05 % Triton X-100, 10 %甘油

[0277] 表1

[0278] 1.2-用DNA聚合酶培育DNA模板

[0279] 在MuA片段化过程之后，然后将纯化的DNA用DNA聚合酶培育以复制上链发夹。

[0280] DNA聚合酶反应（50uL）如下表2所述设置，并在68℃下培育10分钟。最后，将所得的DNA进行1.5×SPRI纯化并在无核酸酶的水（42uL，样品2）中洗脱。

[0282]

试剂	组分
样品 1	39 μ l
10x ThermoPol 缓冲液	5 μ l 的 200 mM Tris-HCl pH 8.8, 100 mM (NH ₄) ₂ SO ₄ , 100 mM KCl, 20 mM MgSO ₄ , 1.0 % Triton® X-100
9°N 聚合酶 (NEB)	5 μ L
dNTP	1 μ L 10 mM (最终浓度 0.2 mM)

[0283] 表2

[0284] 1.3-热变性和聚合酶填充

[0285] 在发夹复制阶段后,将样品2处理,以进行单一变性步骤和聚合酶填充。对于聚合酶填充反应,对聚合酶提供有dCTP/dGTP/dATP,但用不同的核苷酸种类5-丙炔基-dU代替标准 dTTP。反应(50 μ)如下表3所述进行设置,并在95℃培育2分钟,55℃培育30秒,68℃培育30分钟。最后,将所得DNA进行1.5×SPRI纯化并在无核酸酶的水(45 μ L,样品3)中洗脱。

[0286]

试剂	组分
样品 2	39 μ l
10x ThermoPol 缓冲剂	5 μ l 的 200 mM Tris-HCl pH 8.8, 100 mM (NH ₄) ₂ SO ₄ , 100 mM KCl, 20 mM MgSO ₄ , 1.0 % Triton® X-100
9°N 聚合酶(NEB)	5 μ L
dNTPs- (5-丙炔基-dU/dCTP/dGTP/dATP)	1 μ L 10 mM (最终浓度 0.2 mM)

[0287] 表3

[0288] 1.4-dA拖尾反应

[0289] 然后将样品3如下表4所述进行dA拖尾,并在37℃下培育30分钟。将所得DNA进行1.5×SPRI纯化并在无核酸酶的水(20 μ L,样品4)中洗脱。

[0290]

试剂	组分
样品3	42 μ l
10x NEB dA-拖尾缓冲液	5 μ l
Klenow exo ⁻	3 μ L

[0291] 表4

[0292] 1.5-用预加载的酶连接适配器

[0293] 然后将样品4用预加载的酶(T4Dda-E94C/A360C/C109A/C136A(具有突变E94C/A360C/C114A/C171A/C421D以及(Δ M1)G1G2)的SEQ ID NO:24)连接到Y适配器1(上链=在3'端连接到SEQ ID NO:32的20个iSpC3间隔基,SEQ ID NO:32的3'端连接到4个iSp18间隔基,iSp18间隔基3'端连接到SEQ ID NO:33,底部链=连接有5'磷酸盐的SEQ ID NO:34),如下表5所述,并在室温下培育20分钟。将所得的DNA进行0.4×SPRI 纯化,并用缓冲液(200 μ L

的750mM NaCl,10%PEG 8000,50mM Tris.HCl pH8)洗涤,并在缓冲液(20uL的40mMCAPS pH10,40mMKCl样品5)中洗脱。

[0294]	试剂	组分
	样品4	20uL
	Y-适配器1	5μL
	NEB Blunt TA MM (2x)	25μL

[0295] 表5

[0296] 1.6-系链的退火

[0297] 然后将存在于样品5中的DNA分析物退火至系链。将样品5用DNA系链(AACAACCT序列其5'端连接到三个iSp18间隔基,两个胸腺嘧啶和5'胆固醇TEG,并且序列AACAACCT其3'端连接到三个3'端连接到SEQ ID NO:35的iSp18间隔基),500nM,5μL)在室温下培育10分钟。所得样品称为样品6。

[0298] 1.7-电生理检测

[0299] 在设置实验之前,将DNA样品6(样品6的总体积的四分之一)加入到缓冲液(25mM磷酸钾缓冲液(pH7.5),500mM KCl),MgCl₂(1mM)和ATP(2mM),所得总体积为150μL。

[0300] 在缓冲液(25mM K磷酸盐缓冲液,150mM亚铁氰化钾(II))和150mM铁氰化钾(III),pH8.0)中从插入嵌段共聚物中的单个MspA纳米孔获得电测量值。在实现单孔插入嵌段共聚物中之后,将缓冲液(2mL,25mM K磷酸盐缓冲液,150mM亚铁氰化钾(II),150mM 铁氰化钾(III),pH 8.0)流过系统以去除任何过量MspA纳米孔。然后将酶(T4 Dda-E94C/C109A/C136A/A360C,10nM最终浓度),DNA样品6和燃料(MgCl₂ 2m最终浓度,ATP 2mM最终浓度)预混合物(总共150μL)流入单纳米孔实验系统中。实验在120mV 下进行,并且监测解旋酶控制的DNA运动6小时。

[0301] 结果

[0302] 观察到,样品制备过程(样品6)结束时产生的DNA的解旋酶控制的DNA运动。图3 显示了解旋酶控制DNA运动的示例。

[0303] 还使用Agilent 12,000DNA芯片迹线分析样品制备过程。如果没有在68℃下的预培育,则在步骤1.2(其中5'发夹被转录)之前,随后在链离解(热变性步骤1.3)之后,没有制得合成的互补体(在图2的步骤4之后以虚线/点线示出),因为链缺少聚合酶引发所必需的3'发夹。这在图4所示的Agilent 12,000DNA芯片迹线中被观察到,其中标记为1的线是未处理的MuA片段化的DNA输入材料,标记为2的线是具有68℃培育步骤(在以上的1.2中)且随后经历了所有1.3步骤的分析物,并且标记为3的线在步骤1.2中没有68℃培育,但经历了所有1.3步骤。因此,对于线3,没有制得dsDNA,因此在Agilent迹线上观察到平坦的线(区域标记为X),因为在95℃下链离解之前没有被复制的发夹。然而,对于线2,发夹被转录并且因此在链离解时,聚合酶从新的3'发夹引发填充。这意味着线2示出了在区域X 中的峰,其对应于由复制的发夹制成的dsDNA产物。

[0304] 然而,如上所述,在步骤1.3中重复上述过程,向聚合酶提供标准DNA dNTP's-dCTP/dATP/dGTP/dTTP,而不是产生DNA样品7的5-丙炔基-dU。图6示出了DNA 样品7(其在步骤1.3中使用标准DNA dNTP's制备)的示例性的解旋酶控制的DNA移动。该样品制备过程是成功的,并且观察到该样品的解旋酶控制的DNA运动。

[0305] 实施例2

[0306] 该实施例描述了用于修饰模板双链多核苷酸的方法,特别是使用纳米孔测序进行表征的方法。图7示出了下面步骤2.1和2.2中描述的样品制备步骤的卡通表示。该实施例示出了 MuA转座酶能够将含有发夹环(该发夹环含有dG和dC的类似物)的MuA底物插入MuA 适配器的发夹环(dG被脱氧肌苷取代并且dC被脱氧zebularine取代)。然后使用聚合酶填充构建体中的间隙,所述聚合酶用与包含发夹环的链互补的新链取代突出链。与包含发夹环的链互补的新链也能够形成发夹环。新链的发夹环具有比双链区更高的 T_m ,该双链区在互补链和由A/T/Z/I(在图7中标记为1X)制成的发夹环之间形成。因此,形成了新链中的发夹(图7中标记为f2h)并且也形成了由A/T/Z/I制成的发夹环(在图7中标记为f1h)。然后聚合酶将发夹环用作引物以制备互补链。因此,不需要额外的加热步骤来分离在图7的步骤2之后产生的dsDNA构建体。

[0307] 2.1-使用MuA转座酶片段化DNA模板

[0308] 在该实施例中使用的MuA适配器具有5'7bp发夹,dG被替换为dInosine,dC被替换为 dZebularine。适配器P的上链(经修饰的多核苷酸序列IZITAZ(其中I是脱氧肌苷,Z是脱氧zebularine)被连接到未经修饰的多核苷酸序列TTTTTA的5'端,TTTTTA的3'端与连接到SEQ ID NO:39的5'端的经修饰的多核苷酸序列ITAZIZ(其中I是脱氧肌苷,Z是脱氧zebularine)连接)和下链(SEQ ID NO:38)在10uM,95°C,以2°Cmin⁻¹,在10mM Tris pH 7.5, 50mM NaCl中退火。

[0309] 如上表1所述,MuA片段化反应(10μL)使用适配器P而不是适配器X设置,并在30°C下培育1小时。然后通过75°C加热15分钟将MuA酶进行热灭活。最后,将所得的DNA 进行1.5×SPRI纯化并在无核酸酶的水(42μL)中洗脱。

[0310] 2.2-用DNA聚合酶培育DNA模板

[0311] 在MuA片段化过程之后,将纯化的DNA用DNA聚合酶培育以复制上链发夹(其G/C 用I/Z替换)。

[0312] 在此步骤期间,与包含发夹环的链互补的新链形成发夹环。这是由于由新链形成的发夹环比在互补链和含有dZ和dI的类似物的发夹环之间形成的双链区具有更高的 T_m 。因此,不需要加热双链DNA,来将其分离成ssDNA,因为具有较高 T_m 的发夹环优先形成,然后聚合酶使用该发夹环作为引物来制得互补链。

[0313] DNA聚合酶反应(50uL)如下表所述设置,并在37°C培育30分钟。最后,将所得的DNA进行1.5×SPRI纯化并在无核酸酶的水(42uL)中洗脱。

[0314]

试剂	组分
样品 1	39 μ l
10x NE 缓冲液	5 μ l 的 100 mM Tris-HCl (pH 7.9), 500 mM NaCl ₂ , 100 mM MgCl ₂ , 10 mM DTT
Klenow 片段(NEB)	2.5 μ L
SSB (Promega)	2.5 μ L
dNTPs	1 μ L 的 10 mM (最终浓度 0.2 mM)

[0315] 表6

[0316] 该链可以通过dA拖尾进一步修饰,将适配器与预加载的酶连接并与系链杂交(如实施例1.4-1.6中所述),从而产生可以使用纳米孔系统表征的链(如实施例1.7中所述)。

[0001]	序列表
[0002]	<110> 牛津纳米孔技术公司
[0003]	<120> 方法
[0004]	<130> N404112W0
[0005]	<140> GB1418159.8
[0006]	<141> 2014-10-14
[0007]	<160> 37
[0008]	<170> PatentIn version 3.5
[0009]	<210> 1
[0010]	<211> 558
[0011]	<212> DNA
[0012]	<213> 人工序列
[0013]	<220>
[0014]	<223> 耻垢分枝杆菌(Mycobacterium smegmatis)孔蛋白A突变体
[0015]	(D90N/D91N/D93N/D118R/D134R/E193K)
[0016]	<400> 1
[0017]	atgggtctgg ataatgaact gagcctgggtg gacggtcaag atcgtaccct gacggtgcaa 60
[0018]	caatgggata cctttctgaa tggcgttttt cgcctggatc gtaatcgctt gacccgtgaa 120
[0019]	tggtttcatt ccggtcgcgc aaaatatatc gtcgcaggcc cgggtgctga cgaattcgaa 180
[0020]	ggcacgctgg aactgggtta tcagattggc tttccgtggt cactgggcgt tggtatcaac 240
[0021]	ttctcgtaca ccacgccgaa tattctgata aacaatggta acattaccgc accgccgttt 300
[0022]	ggcctgaaca gcgtgattac gccgaacctg tttccgggtg ttagcatctc tgcccgtctg 360
[0023]	ggcaatggtc cgggcattca agaagtggca accttttagtg tgcgcgtttc cggcgctaaa 420
[0024]	ggcgggtgtcg cgggtgtctaa cgcccacggt accgttacgg gcgcggccgg cgggtgtcctg 480
[0025]	ctgcgtccgt tcgcgcgcct gattgcctct accggcgaca gcgttacgac ctatggcgaa 540
[0026]	ccgtggaata tgaactaa 558
[0027]	<210> 2
[0028]	<211> 184
[0029]	<212> PRT
[0030]	<213> 人工序列
[0031]	<220>
[0032]	<223> 耻垢分枝杆菌孔蛋白A突变体
[0033]	(D90N/D91N/D93N/D118R/D134R/E139K)
[0034]	<400> 2
[0035]	Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
[0036]	1 5 10 15
[0037]	Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
[0038]	20 25 30

[0039]	Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr	
[0040]	35	40 45
[0041]	Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu	
[0042]	50	55 60
[0043]	Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe	
[0044]	65	70 75 80
[0045]	Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asn Asn Gly Asn Ile Thr Ala	
[0046]	85	90 95
[0047]	Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly	
[0048]	100	105 110
[0049]	Val Ser Ile Ser Ala Arg Leu Gly Asn Gly Pro Gly Ile Gln Glu Val	
[0050]	115	120 125
[0051]	Ala Thr Phe Ser Val Arg Val Ser Gly Ala Lys Gly Gly Val Ala Val	
[0052]	130	135 140
[0053]	Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu	
[0054]	145	150 155 160
[0055]	Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr	
[0056]	165	170 175
[0057]	Tyr Gly Glu Pro Trp Asn Met Asn	
[0058]	180	
[0059]	<210> 3	
[0060]	<211> 885	
[0061]	<212> DNA	
[0062]	<213> 人工序列	
[0063]	<220>	
[0064]	<223> α-溶血素突变体 (E111N/K147N)	
[0065]	<400> 3	
[0066]	atggcagatt ctgatattaa tattaaaacc ggtactacag atattggaag caatactaca	60
[0067]	gtaaaaacag gtgatttagt cacttatgat aaagaaaatg gcatgcacaa aaaagtat	120
[0068]	tatagtttta tcgatgataa aaatcacaat aaaaaactgc tagttattag aacaaaaggt	180
[0069]	accattgctg gtcaatatag agtttatagc gaagaagggtg ctaacaaaag tggtttagcc	240
[0070]	tggccttcag cctttaaggt acagttgcaa ctacctgata atgaagtagc tcaaatatct	300
[0071]	gattactatc caagaaattc gattgataca aaaaactata tgagtacttt aacttatgga	360
[0072]	ttcaacggta atgttactgg tgatgataca ggaaaaattg gcggccttat tggtgcaaat	420
[0073]	gtttcgattg gtcatacact gaactatgtt caacctgatt tcaaaacaat tttagagagc	480
[0074]	ccaactgata aaaaagtagg ctggaaagtg atatttaaca atatggtgaa tcaaaattgg	540
[0075]	ggaccatacg atcgagattc ttggaacccg gtatatggca atcaactttt catgaaaact	600
[0076]	agaaatgggt ctatgaaagc agcagataac ttccttgatc ctaacaaagc aagttctcta	660
[0077]	ttatcttcag ggttttcacc agacttcgct acagttatta ctatggatag aaaagcatcc	720

[0078] aaacaacaaa caaatataga tgtaatat ac gaacgagttc gtgatgatta ccaattgcat 780
 [0079] tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgttcttca 840
 [0080] gaaagatata aaatcgattg ggaaaaagaa gaaatgacaa attaa 885
 [0081] <210> 4
 [0082] <211> 293
 [0083] <212> PRT
 [0084] <213> 人工序列
 [0085] <220>
 [0086] <223> α -溶血素突变体(E111N/K147N)
 [0087] <400> 4
 [0088] Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser
 [0089] 1 5 10 15
 [0090] Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn
 [0091] 20 25 30
 [0092] Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His
 [0093] 35 40 45
 [0094] Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln
 [0095] 50 55 60
 [0096] Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp
 [0097] 65 70 75 80
 [0098] Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala
 [0099] 85 90 95
 [0100] Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Asn Tyr
 [0101] 100 105 110
 [0102] Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp
 [0103] 115 120 125
 [0104] Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His
 [0105] 130 135 140
 [0106] Thr Leu Asn Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro
 [0107] 145 150 155 160
 [0108] Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn
 [0109] 165 170 175
 [0110] Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly
 [0111] 180 185 190
 [0112] Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp
 [0113] 195 200 205
 [0114] Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe
 [0115] 210 215 220
 [0116] Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys

[0117]	225	230	235	240
[0118]	Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr			
[0119]	245	250	255	
[0120]	Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp			
[0121]	260	265	270	
[0122]	Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys			
[0123]	275	280	285	
[0124]	Glu Glu Met Thr Asn			
[0125]	290			
[0126]	<210> 5			
[0127]	<211> 184			
[0128]	<212> PRT			
[0129]	<213> 耻垢分枝杆菌			
[0130]	<400> 5			
[0131]	Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu			
[0132]	1	5	10	15
[0133]	Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp			
[0134]	20	25	30	
[0135]	Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr			
[0136]	35	40	45	
[0137]	Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu			
[0138]	50	55	60	
[0139]	Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe			
[0140]	65	70	75	80
[0141]	Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala			
[0142]	85	90	95	
[0143]	Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly			
[0144]	100	105	110	
[0145]	Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val			
[0146]	115	120	125	
[0147]	Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val			
[0148]	130	135	140	
[0149]	Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu			
[0150]	145	150	155	160
[0151]	Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr			
[0152]	165	170	175	
[0153]	Tyr Gly Glu Pro Trp Asn Met Asn			
[0154]	180			
[0155]	<210> 6			

[0156]	<211>	184
[0157]	<212>	PRT
[0158]	<213>	耻垢分枝杆菌
[0159]	<400>	6
[0160]	Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu	
[0161]	1	5 10 15
[0162]	Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp	
[0163]	20 25 30	
[0164]	Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr	
[0165]	35 40 45	
[0166]	Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu	
[0167]	50 55 60	
[0168]	Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe	
[0169]	65 70 75 80	
[0170]	Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly	
[0171]	85 90 95	
[0172]	Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly	
[0173]	100 105 110	
[0174]	Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val	
[0175]	115 120 125	
[0176]	Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val	
[0177]	130 135 140	
[0178]	Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu	
[0179]	145 150 155 160	
[0180]	Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr	
[0181]	165 170 175	
[0182]	Tyr Gly Glu Pro Trp Asn Met Asn	
[0183]	180	
[0184]	<210>	7
[0185]	<211>	183
[0186]	<212>	PRT
[0187]	<213>	耻垢分枝杆菌
[0188]	<400>	7
[0189]	Val Asp Asn Gln Leu Ser Val Val Asp Gly Gln Gly Arg Thr Leu Thr	
[0190]	1	5 10 15
[0191]	Val Gln Gln Ala Glu Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg	
[0192]	20 25 30	
[0193]	Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Thr Tyr His	
[0194]	35 40 45	

[0195]	Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly	
[0196]	50	60
[0197]	Tyr Gln Val Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser	
[0198]	65	80
[0199]	Tyr Thr Thr Pro Asn Ile Leu Ile Asp Gly Gly Asp Ile Thr Gln Pro	
[0200]	85	95
[0201]	Pro Phe Gly Leu Asp Thr Ile Ile Thr Pro Asn Leu Phe Pro Gly Val	
[0202]	100	110
[0203]	Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala	
[0204]	115	125
[0205]	Thr Phe Ser Val Asp Val Lys Gly Ala Lys Gly Ala Val Ala Val Ser	
[0206]	130	140
[0207]	Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg	
[0208]	145	160
[0209]	Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr	
[0210]	165	175
[0211]	Gly Glu Pro Trp Asn Met Asn	
[0212]	180	
[0213]	<210> 8	
[0214]	<211> 1830	
[0215]	<212> DNA	
[0216]	<213> 枯草芽孢杆菌噬菌体(Bacillus subtilis phage)phi29	
[0217]	<400> 8	
[0218]	atgaaacaca tgccgcgtaa aatgtatagc tgccggtttg aaaccacgac caaagtggaa	60
[0219]	gattgtcgcg tttgggccta tggctacatg aacatcgaag atcattctga atacaaaatc	120
[0220]	ggtaacagtc tggatgaatt tatggcatgg gtgctgaaag ttcaggcgga tctgtacttc	180
[0221]	cacaacctga aatttgatgg cgcattcatt atcaactggc tggaacgtaa tggcttttaa	240
[0222]	tggagcgcgg atggtctgcc gaacacgtat aataccatta tctctcgtat gggccagtgg	300
[0223]	tatatgattg atatctgcct gggctacaaa ggtaaacgca aaattcatac cgtgatctat	360
[0224]	gatagcctga aaaaactgcc gtttccggtg aagaaaattg cgaaagattt caaactgacg	420
[0225]	gttctgaaag gcgatattga ttatcacaaa gaacgtccgg ttggttacaa aatcaccccg	480
[0226]	gaagaatacg catacatcaa aaacgatatc cagatcatcg cagaagcgct gctgattcag	540
[0227]	tttaaacagg gcctggatcg catgaccgcg ggcagtgata gcctgaaagg tttcaaagat	600
[0228]	atcatcacga ccaaaaaatt caaaaaagtg ttcccgcgc tgagcctggg tctggataaa	660
[0229]	gaagttcggt atgcctaccg cggcggtttt acctggctga acgatcggtt caaagaaaaa	720
[0230]	gaaattggcg agggatatgg gtttgatgtt aatagtctgt atccggcaca gatgtacagc	780
[0231]	cgctgctgc cgtatggcga accgatcgtg ttcgaggta aatatgtttg ggatgaagat	840
[0232]	taccgctgc atattcagca catccgttgt gaatttgaac tgaaagaagg ctatatccg	900
[0233]	accattcaga tcaaacgtag tcgcttctat aagggtaac aatacctgaa aagctctggc	960

[0234]	ggtgaaatcg cggatctgtg gctgagtaac gtggatctgg aactgatgaa agaacactac	1020
[0235]	gatctgtaca acgttgaata catcagcggc ctgaaattta aagccacgac cggctctgttc	1080
[0236]	aaagatttca tcgataaatg gacctacatc aaaacgacct ctgaaggcgc gattaaacag	1140
[0237]	ctggccaaac tgatgctgaa cagcctgtat ggcaaattcg cctctaattcc ggatgtgacc	1200
[0238]	ggtaaagtgc cgtacctgaa agaaaatggc gcaactgggtt ttgcctggg cgaagaagaa	1260
[0239]	acgaaagatc cgggtgtatac cccgatgggt gttttcatta cggcctgggc acgttacacg	1320
[0240]	accatcaccg cggcccaggc atgctatgat cgcattatct actgtgatac cgattctatt	1380
[0241]	catctgacgg gcaccgaaat cccgatgtg attaaagata tcgttgatcc gaaaaaactg	1440
[0242]	ggttattggg cccacgaaag tacgtttaaa cgtgcaaaat acctgcgcca gaaaacctac	1500
[0243]	atccaggata tctacatgaa agaagtggat ggcaaactgg ttgaaggttc tccggatgat	1560
[0244]	tacaccgata tcaaattcag tgtgaaatgc gccggcatga cggataaaat caaaaaagaa	1620
[0245]	gtgaccttcg aaaacttcaa agttggtttc agccgcaaaa tgaaccgaa accggtgcag	1680
[0246]	gttccgggcg gtgtggttct ggtggatgat acgtttacca ttaaatctgg cggtagtgcg	1740
[0247]	tggagccatc cgcagttcga aaaaggcggg ggctctggtg gcggttctgg cggtagtgcc	1800
[0248]	tggagccacc cgcagtttga aaaataataa	1830
[0249]	<210> 9	
[0250]	<211> 608	
[0251]	<212> PRT	
[0252]	<213> 枯草芽孢杆菌噬菌体phi29	
[0253]	<400> 9	
[0254]	Met Lys His Met Pro Arg Lys Met Tyr Ser Cys Ala Phe Glu Thr Thr	
[0255]	1 5 10 15	
[0256]	Thr Lys Val Glu Asp Cys Arg Val Trp Ala Tyr Gly Tyr Met Asn Ile	
[0257]	20 25 30	
[0258]	Glu Asp His Ser Glu Tyr Lys Ile Gly Asn Ser Leu Asp Glu Phe Met	
[0259]	35 40 45	
[0260]	Ala Trp Val Leu Lys Val Gln Ala Asp Leu Tyr Phe His Asn Leu Lys	
[0261]	50 55 60	
[0262]	Phe Asp Gly Ala Phe Ile Ile Asn Trp Leu Glu Arg Asn Gly Phe Lys	
[0263]	65 70 75 80	
[0264]	Trp Ser Ala Asp Gly Leu Pro Asn Thr Tyr Asn Thr Ile Ile Ser Arg	
[0265]	85 90 95	
[0266]	Met Gly Gln Trp Tyr Met Ile Asp Ile Cys Leu Gly Tyr Lys Gly Lys	
[0267]	100 105 110	
[0268]	Arg Lys Ile His Thr Val Ile Tyr Asp Ser Leu Lys Lys Leu Pro Phe	
[0269]	115 120 125	
[0270]	Pro Val Lys Lys Ile Ala Lys Asp Phe Lys Leu Thr Val Leu Lys Gly	
[0271]	130 135 140	
[0272]	Asp Ile Asp Tyr His Lys Glu Arg Pro Val Gly Tyr Lys Ile Thr Pro	

[0273]	145	150	155	160
[0274]	Glu Glu Tyr Ala Tyr Ile Lys Asn Asp Ile Gln Ile Ile Ala Glu Ala			
[0275]		165	170	175
[0276]	Leu Leu Ile Gln Phe Lys Gln Gly Leu Asp Arg Met Thr Ala Gly Ser			
[0277]		180	185	190
[0278]	Asp Ser Leu Lys Gly Phe Lys Asp Ile Ile Thr Thr Lys Lys Phe Lys			
[0279]		195	200	205
[0280]	Lys Val Phe Pro Thr Leu Ser Leu Gly Leu Asp Lys Glu Val Arg Tyr			
[0281]		210	215	220
[0282]	Ala Tyr Arg Gly Gly Phe Thr Trp Leu Asn Asp Arg Phe Lys Glu Lys			
[0283]		225	230	235
[0284]	Glu Ile Gly Glu Gly Met Val Phe Asp Val Asn Ser Leu Tyr Pro Ala			
[0285]		245	250	255
[0286]	Gln Met Tyr Ser Arg Leu Leu Pro Tyr Gly Glu Pro Ile Val Phe Glu			
[0287]		260	265	270
[0288]	Gly Lys Tyr Val Trp Asp Glu Asp Tyr Pro Leu His Ile Gln His Ile			
[0289]		275	280	285
[0290]	Arg Cys Glu Phe Glu Leu Lys Glu Gly Tyr Ile Pro Thr Ile Gln Ile			
[0291]		290	295	300
[0292]	Lys Arg Ser Arg Phe Tyr Lys Gly Asn Glu Tyr Leu Lys Ser Ser Gly			
[0293]		305	310	315
[0294]	Gly Glu Ile Ala Asp Leu Trp Leu Ser Asn Val Asp Leu Glu Leu Met			
[0295]		325	330	335
[0296]	Lys Glu His Tyr Asp Leu Tyr Asn Val Glu Tyr Ile Ser Gly Leu Lys			
[0297]		340	345	350
[0298]	Phe Lys Ala Thr Thr Gly Leu Phe Lys Asp Phe Ile Asp Lys Trp Thr			
[0299]		355	360	365
[0300]	Tyr Ile Lys Thr Thr Ser Glu Gly Ala Ile Lys Gln Leu Ala Lys Leu			
[0301]		370	375	380
[0302]	Met Leu Asn Ser Leu Tyr Gly Lys Phe Ala Ser Asn Pro Asp Val Thr			
[0303]		385	390	395
[0304]	Gly Lys Val Pro Tyr Leu Lys Glu Asn Gly Ala Leu Gly Phe Arg Leu			
[0305]		405	410	415
[0306]	Gly Glu Glu Glu Thr Lys Asp Pro Val Tyr Thr Pro Met Gly Val Phe			
[0307]		420	425	430
[0308]	Ile Thr Ala Trp Ala Arg Tyr Thr Thr Ile Thr Ala Ala Gln Ala Cys			
[0309]		435	440	445
[0310]	Tyr Asp Arg Ile Ile Tyr Cys Asp Thr Asp Ser Ile His Leu Thr Gly			
[0311]		450	455	460

[0312]	Thr Glu Ile Pro Asp Val Ile Lys Asp Ile Val Asp Pro Lys Lys Leu		
[0313]	465	470	475 480
[0314]	Gly Tyr Trp Ala His Glu Ser Thr Phe Lys Arg Ala Lys Tyr Leu Arg		
[0315]		485	490 495
[0316]	Gln Lys Thr Tyr Ile Gln Asp Ile Tyr Met Lys Glu Val Asp Gly Lys		
[0317]		500	505 510
[0318]	Leu Val Glu Gly Ser Pro Asp Asp Tyr Thr Asp Ile Lys Phe Ser Val		
[0319]		515	520 525
[0320]	Lys Cys Ala Gly Met Thr Asp Lys Ile Lys Lys Glu Val Thr Phe Glu		
[0321]		530	535 540
[0322]	Asn Phe Lys Val Gly Phe Ser Arg Lys Met Lys Pro Lys Pro Val Gln		
[0323]	545	550	555 560
[0324]	Val Pro Gly Gly Val Val Leu Val Asp Asp Thr Phe Thr Ile Lys Ser		
[0325]		565	570 575
[0326]	Gly Gly Ser Ala Trp Ser His Pro Gln Phe Glu Lys Gly Gly Gly Ser		
[0327]		580	585 590
[0328]	Gly Gly Gly Ser Gly Gly Ser Ala Trp Ser His Pro Gln Phe Glu Lys		
[0329]		595	600 605
[0330]	<210> 10		
[0331]	<211> 1390		
[0332]	<212> DNA		
[0333]	<213> 大肠杆菌(Escherichia coli)		
[0334]	<400> 10		
[0335]	atgatgaacg atggcaaaca gcagagcacc ttctgtttc atgattatga aaccttcggt	60	
[0336]	accatccgg ccttgatcg tccggcgcag tttgcggcca ttcgcaccga tagcgaattc	120	
[0337]	aatgtgattg gcgaaccgga agtgttttat tgcaaacgg ccgatgatta tctgccgcag	180	
[0338]	ccgggtgcgg tgctgattac cgggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac	240	
[0339]	gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac ctgcattctg	300	
[0340]	ggctataaca atgtgcgctt cgatgatgaa gttaccgta atatctttta tcgtaacttt	360	
[0341]	tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg	420	
[0342]	atgcgcgcgt gctatgcgct gcgcccggaa ggcattaatt ggccgaaaa cgatgatggc	480	
[0343]	ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc	540	
[0344]	catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactgg taaaaccctg	600	
[0345]	cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg	660	
[0346]	attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgcctggcgc	720	
[0347]	ggcaacacca gctgggtggc cccgctggcc tggcaccgg aaaatcgtaa cgccgtgatt	780	
[0348]	atggttgatc tggccgtgga tattagcccg ctgctggaac tggatagcga taccctgcgt	840	
[0349]	gaacgcctgt ataccgcaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg	900	
[0350]	gttcacatta acaaatgccc ggtgctggcc caggcgaaca ccctgcgccc ggaagatgcg	960	

[0351] gatcgtctgg gtattaatcg ccagcattgt ctggataatc tgaaaatcct gcgtgaaaac 1020
 [0352] ccgcaggtgc gtgaaaaagt ggtggcgatc ttccgcgaag cggaaccgtt caccgccgagc 1080
 [0353] gataacgtgg atgcgcagct gtataacggc ttcttttagcg atgccgatcg cgcggcgatg 1140
 [0354] aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat 1200
 [0355] aaacgtattg aaaaactgct gtttaattat cgtgcgcgca attttccggg taccctggat 1260
 [0356] tatgccgaac agcagcgttg gctggaacat cgtcgtcagg ttttcacccc ggaatttctg 1320
 [0357] cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa 1380
 [0358] gtggcgctgc 1390
 [0359] <210> 11
 [0360] <211> 485
 [0361] <212> PRT
 [0362] <213> 大肠杆菌
 [0363] <400> 11
 [0364] Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr
 [0365] 1 5 10 15
 [0366] Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala
 [0367] 20 25 30
 [0368] Ala Ile Arg Thr Asp Ser Glu Phe Asn Val Ile Gly Glu Pro Glu Val
 [0369] 35 40 45
 [0370] Phe Tyr Cys Lys Pro Ala Asp Asp Tyr Leu Pro Gln Pro Gly Ala Val
 [0371] 50 55 60
 [0372] Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn
 [0373] 65 70 75 80
 [0374] Glu Ala Ala Phe Ala Ala Arg Ile His Ser Leu Phe Thr Val Pro Lys
 [0375] 85 90 95
 [0376] Thr Cys Ile Leu Gly Tyr Asn Asn Val Arg Phe Asp Asp Glu Val Thr
 [0377] 100 105 110
 [0378] Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp
 [0379] 115 120 125
 [0380] Gln His Asp Asn Ser Arg Trp Asp Leu Leu Asp Val Met Arg Ala Cys
 [0381] 130 135 140
 [0382] Tyr Ala Leu Arg Pro Glu Gly Ile Asn Trp Pro Glu Asn Asp Asp Gly
 [0383] 145 150 155 160
 [0384] Leu Pro Ser Phe Arg Leu Glu His Leu Thr Lys Ala Asn Gly Ile Glu
 [0385] 165 170 175
 [0386] His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala
 [0387] 180 185 190
 [0388] Met Ala Lys Leu Val Lys Thr Arg Gln Pro Arg Leu Phe Asp Tyr Leu
 [0389] 195 200 205

[0390]	Phe Thr His Arg Asn Lys His Lys Leu Met Ala Leu Ile Asp Val Pro		
[0391]	210	215	220
[0392]	Gln Met Lys Pro Leu Val His Val Ser Gly Met Phe Gly Ala Trp Arg		
[0393]	225	230	235 240
[0394]	Gly Asn Thr Ser Trp Val Ala Pro Leu Ala Trp His Pro Glu Asn Arg		
[0395]	245	250	255
[0396]	Asn Ala Val Ile Met Val Asp Leu Ala Gly Asp Ile Ser Pro Leu Leu		
[0397]	260	265	270
[0398]	Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr		
[0399]	275	280	285
[0400]	Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn		
[0401]	290	295	300
[0402]	Lys Cys Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala		
[0403]	305	310	315 320
[0404]	Asp Arg Leu Gly Ile Asn Arg Gln His Cys Leu Asp Asn Leu Lys Ile		
[0405]	325	330	335
[0406]	Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala		
[0407]	340	345	350
[0408]	Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr		
[0409]	355	360	365
[0410]	Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu		
[0411]	370	375	380
[0412]	Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp		
[0413]	385	390	395 400
[0414]	Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro		
[0415]	405	410	415
[0416]	Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg		
[0417]	420	425	430
[0418]	Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln		
[0419]	435	440	445
[0420]	Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu		
[0421]	450	455	460
[0422]	Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His		
[0423]	465	470	475 480
[0424]	His His His His His		
[0425]	485		
[0426]	<210> 12		
[0427]	<211> 804		
[0428]	<212> DNA		

[0429]	<213> 大肠杆菌
[0430]	<400> 12
[0431]	atgaaatttg tctcttttaa tatcaacggc ctgcgcgcca gacctcacca gcttgaagcc 60
[0432]	atcgctgaaa agcaccaacc ggatgtgatt ggcctgcagg agacaaaagt tcatgacgat 120
[0433]	atgtttccgc tcgaagaggt ggcgaagctc ggctacaacg tgttttatca cgggcagaaa 180
[0434]	ggccattatg gcggtggcgt gctgaccaa gagacgccga ttgccgtgcg tcgcggcttt 240
[0435]	cccggtgacg acgaagaggc gcagcggcgg attattatgg cggaaatccc ctactgctg 300
[0436]	ggtaatgtca ccgtgatcaa cggttacttc ccgcagggtg aaagccgcga ccatccgata 360
[0437]	aaattcccg caaaagcgca gttttatcag aatctgcaa actacctgga aaccgaactc 420
[0438]	aaacgtgata atccggtact gattatgggc gatatgaata tcagccctac agatctggat 480
[0439]	atcggcattg gcgaagaaaa ccgtaagcgc tggctgcgta ccggtaaatg ctctttcctg 540
[0440]	ccggaagagc gcgaatggat ggacaggctg atgagctggg ggttggtcga taccttccgc 600
[0441]	catgcgaatc cgcaaacagc agatcgtttc tcatggtttg attaccgctc aaaaggtttt 660
[0442]	gacgataacc gtggtctgcg catcgacctg ctgctcgcca gccaacgct ggacgaatgt 720
[0443]	tgcgtagaaa ccggcatcga ctatgaaatc cgcagcatgg aaaaaccgtc cgatcacgcc 780
[0444]	cccgtctggg cgaccttccg ccgc 804
[0445]	<210> 13
[0446]	<211> 268
[0447]	<212> PRT
[0448]	<213> 大肠杆菌
[0449]	<400> 13
[0450]	Met Lys Phe Val Ser Phe Asn Ile Asn Gly Leu Arg Ala Arg Pro His
[0451]	1 5 10 15
[0452]	Gln Leu Glu Ala Ile Val Glu Lys His Gln Pro Asp Val Ile Gly Leu
[0453]	20 25 30
[0454]	Gln Glu Thr Lys Val His Asp Asp Met Phe Pro Leu Glu Glu Val Ala
[0455]	35 40 45
[0456]	Lys Leu Gly Tyr Asn Val Phe Tyr His Gly Gln Lys Gly His Tyr Gly
[0457]	50 55 60
[0458]	Val Ala Leu Leu Thr Lys Glu Thr Pro Ile Ala Val Arg Arg Gly Phe
[0459]	65 70 75 80
[0460]	Pro Gly Asp Asp Glu Glu Ala Gln Arg Arg Ile Ile Met Ala Glu Ile
[0461]	85 90 95
[0462]	Pro Ser Leu Leu Gly Asn Val Thr Val Ile Asn Gly Tyr Phe Pro Gln
[0463]	100 105 110
[0464]	Gly Glu Ser Arg Asp His Pro Ile Lys Phe Pro Ala Lys Ala Gln Phe
[0465]	115 120 125
[0466]	Tyr Gln Asn Leu Gln Asn Tyr Leu Glu Thr Glu Leu Lys Arg Asp Asn
[0467]	130 135 140

[0468]	Pro Val Leu Ile Met Gly Asp Met Asn Ile Ser Pro Thr Asp Leu Asp				
[0469]	145	150	155	160	
[0470]	Ile Gly Ile Gly Glu Glu Asn Arg Lys Arg Trp Leu Arg Thr Gly Lys				
[0471]		165	170	175	
[0472]	Cys Ser Phe Leu Pro Glu Glu Arg Glu Trp Met Asp Arg Leu Met Ser				
[0473]		180	185	190	
[0474]	Trp Gly Leu Val Asp Thr Phe Arg His Ala Asn Pro Gln Thr Ala Asp				
[0475]		195	200	205	
[0476]	Arg Phe Ser Trp Phe Asp Tyr Arg Ser Lys Gly Phe Asp Asp Asn Arg				
[0477]		210	215	220	
[0478]	Gly Leu Arg Ile Asp Leu Leu Leu Ala Ser Gln Pro Leu Ala Glu Cys				
[0479]		225	230	235	240
[0480]	Cys Val Glu Thr Gly Ile Asp Tyr Glu Ile Arg Ser Met Glu Lys Pro				
[0481]		245	250	255	
[0482]	Ser Asp His Ala Pro Val Trp Ala Thr Phe Arg Arg				
[0483]		260	265		
[0484]	<210> 14				
[0485]	<211> 1275				
[0486]	<212> DNA				
[0487]	<213> 嗜热栖热菌(Thermus thermophilus)				
[0488]	<400> 14				
[0489]	atgtttcgtc gtaaagaaga tctggatccg ccgctggcac tgctgccgct gaaaggcctg	60			
[0490]	cgcgaaagccg ccgcaactgct ggaagaagcg ctgcgtcaag gtaaacgcat tcgtgttcac	120			
[0491]	ggcgactatg atgcggatgg cctgaccggc accgcgatcc tggttcgtgg tctggccgcc	180			
[0492]	ctgggtgcgg atgttcatcc gtttatcccg caccgcctgg aagaaggcta tgggtgtcctg	240			
[0493]	atggaacgcg tcccgaaca tctggaagcc tcggacctgt ttctgaccgt tgactgcggc	300			
[0494]	attaccaacc atgcggaact gcgcgaactg ctggaaaatg gcgtggaagt cattgttacc	360			
[0495]	gatcatcata cgccgggcaa aacgccggcg ccgggtctgg tcgtgcatcc ggcgctgacg	420			
[0496]	ccggatctga aagaaaaacc gaccggcgca ggctggcgct ttctgctgct gtgggcactg	480			
[0497]	catgaacgcc tgggcctgcc gccgccgtg gaatacgcg accctggcage cgttggcacc	540			
[0498]	attgccgacg ttgccccgct gtgggggttg aatcgtgcac tggtgaaaga aggtctggca	600			
[0499]	cgcatecccg cttcatcttg ggtgggcctg cgtctgctgg ctgaagccgt gggctatacc	660			
[0500]	ggcaaagcgg tcgaagtgc tttccgcata gcgccgcga tcaatgcggc ttcccgccctg	720			
[0501]	ggcgaagcgg aaaaagccct gcgcctgctg ctgacggatg atgcggcaga agctcaggcg	780			
[0502]	ctggtcggcg aactgcaccg tctgaacgcc cgtcgtcaga ccctggaaga agcgatgctg	840			
[0503]	cgcaaactgc tgccgcagge cgaccgggaa gcgaaagcca tcgttctgct ggaccgggaa	900			
[0504]	ggccatccgg gtgttatggg tattgtggcc tctcgcatac tggaagcgac cctgcgcccc	960			
[0505]	gtctttcttg tggcccaggg caaaggcacc gtgcgttcgc tggtccgat ttccgccgtc	1020			
[0506]	gaagcactgc gcagcgcgga agatctgctg ctgcgttatg gtggtcataa agaagcggcg	1080			

[0507] ggttttcgcaa tggatgaagc gctgtttccg gcgttcaaag cacgcgttga agcgtatgcc 1140
 [0508] gcacgtttcc cggatccggt tcgtgaagtg gcactgctgg atctgctgcc ggaaccgggc 1200
 [0509] ctgctgccgc aggtgttccg tgaactggca ctgctggaac cgtatggtga aggtaaccgc 1260
 [0510] gaaccgctgt tcctg 1275
 [0511] <210> 15
 [0512] <211> 425
 [0513] <212> PRT
 [0514] <213> 嗜热栖热菌
 [0515] <400> 15
 [0516] Met Phe Arg Arg Lys Glu Asp Leu Asp Pro Pro Leu Ala Leu Leu Pro
 [0517] 1 5 10 15
 [0518] Leu Lys Gly Leu Arg Glu Ala Ala Ala Leu Leu Glu Glu Ala Leu Arg
 [0519] 20 25 30
 [0520] Gln Gly Lys Arg Ile Arg Val His Gly Asp Tyr Asp Ala Asp Gly Leu
 [0521] 35 40 45
 [0522] Thr Gly Thr Ala Ile Leu Val Arg Gly Leu Ala Ala Leu Gly Ala Asp
 [0523] 50 55 60
 [0524] Val His Pro Phe Ile Pro His Arg Leu Glu Glu Gly Tyr Gly Val Leu
 [0525] 65 70 75 80
 [0526] Met Glu Arg Val Pro Glu His Leu Glu Ala Ser Asp Leu Phe Leu Thr
 [0527] 85 90 95
 [0528] Val Asp Cys Gly Ile Thr Asn His Ala Glu Leu Arg Glu Leu Leu Glu
 [0529] 100 105 110
 [0530] Asn Gly Val Glu Val Ile Val Thr Asp His His Thr Pro Gly Lys Thr
 [0531] 115 120 125
 [0532] Pro Pro Pro Gly Leu Val Val His Pro Ala Leu Thr Pro Asp Leu Lys
 [0533] 130 135 140
 [0534] Glu Lys Pro Thr Gly Ala Gly Val Ala Phe Leu Leu Leu Trp Ala Leu
 [0535] 145 150 155 160
 [0536] His Glu Arg Leu Gly Leu Pro Pro Pro Leu Glu Tyr Ala Asp Leu Ala
 [0537] 165 170 175
 [0538] Ala Val Gly Thr Ile Ala Asp Val Ala Pro Leu Trp Gly Trp Asn Arg
 [0539] 180 185 190
 [0540] Ala Leu Val Lys Glu Gly Leu Ala Arg Ile Pro Ala Ser Ser Trp Val
 [0541] 195 200 205
 [0542] Gly Leu Arg Leu Leu Ala Glu Ala Val Gly Tyr Thr Gly Lys Ala Val
 [0543] 210 215 220
 [0544] Glu Val Ala Phe Arg Ile Ala Pro Arg Ile Asn Ala Ala Ser Arg Leu
 [0545] 225 230 235 240

[0546]	Gly Glu Ala Glu Lys Ala Leu Arg Leu Leu Leu Thr Asp Asp Ala Ala		
[0547]		245	250 255
[0548]	Glu Ala Gln Ala Leu Val Gly Glu Leu His Arg Leu Asn Ala Arg Arg		
[0549]		260	265 270
[0550]	Gln Thr Leu Glu Glu Ala Met Leu Arg Lys Leu Leu Pro Gln Ala Asp		
[0551]		275	280 285
[0552]	Pro Glu Ala Lys Ala Ile Val Leu Leu Asp Pro Glu Gly His Pro Gly		
[0553]		290	295 300
[0554]	Val Met Gly Ile Val Ala Ser Arg Ile Leu Glu Ala Thr Leu Arg Pro		
[0555]	305	310	315 320
[0556]	Val Phe Leu Val Ala Gln Gly Lys Gly Thr Val Arg Ser Leu Ala Pro		
[0557]		325	330 335
[0558]	Ile Ser Ala Val Glu Ala Leu Arg Ser Ala Glu Asp Leu Leu Leu Arg		
[0559]		340	345 350
[0560]	Tyr Gly Gly His Lys Glu Ala Ala Gly Phe Ala Met Asp Glu Ala Leu		
[0561]		355	360 365
[0562]	Phe Pro Ala Phe Lys Ala Arg Val Glu Ala Tyr Ala Ala Arg Phe Pro		
[0563]		370	375 380
[0564]	Asp Pro Val Arg Glu Val Ala Leu Leu Asp Leu Leu Pro Glu Pro Gly		
[0565]	385	390	395 400
[0566]	Leu Leu Pro Gln Val Phe Arg Glu Leu Ala Leu Leu Glu Pro Tyr Gly		
[0567]		405	410 415
[0568]	Glu Gly Asn Pro Glu Pro Leu Phe Leu		
[0569]		420	425
[0570]	<210> 16		
[0571]	<211> 738		
[0572]	<212> DNA		
[0573]	<213> 细菌噬菌体λ (Bacteriophage lambda)		
[0574]	<400> 16		
[0575]	tccggaagcg gctctggtag tggttctggc atgacaccgg acattatcct gcagcgtacc	60	
[0576]	gggatcgatg tgagagctgt cgaacagggg gatgatgcgt ggcacaaatt acggctcggc	120	
[0577]	gtcatcaccg cttcagaagt tcacaacgtg atagcaaac cccgctccgg aaagaagtgg	180	
[0578]	cctgacatga aaatgtccta cttccacacc ctgcttgcgt aggtttgcac cgggtgtggct	240	
[0579]	ccggaagtta acgctaaagc actggcctgg ggaaaacagt acgagaacga cgccagaacc	300	
[0580]	ctgtttgaat tcaattccgg cgtgaatgtt actgaatccc cgatcatcta tcgcgacgaa	360	
[0581]	agtatgcgta ccgcctgctc tcccgatggt ttatgcagtg acggcaacgg ccttgaactg	420	
[0582]	aaatgcccgt ttacctcccg ggatttcatg aagttccggc tcggtggttt cgaggccata	480	
[0583]	aagtcagctt acatggccca ggtgcagtac agcatgtggg tgacgcgaaa aaatgcctgg	540	
[0584]	tactttgcc aactatgacc gcgtatgaag cgtgaaggcc tgcattatgt cgtgattgag	600	

[0585] cgggatgaaa agtacatggc gagttttgac gagatcgtgc cggagtcat cgaaaaaatg 660
 [0586] gacgaggcac tggctgaaat tggttttgta tttggggagc aatggcgatc tggctctggt 720
 [0587] tccggcagcg gttccgga 738
 [0588] <210> 17
 [0589] <211> 226
 [0590] <212> PRT
 [0591] <213> 细菌噬菌体λ
 [0592] <400> 17
 [0593] Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala
 [0594] 1 5 10 15
 [0595] Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile
 [0596] 20 25 30
 [0597] Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys
 [0598] 35 40 45
 [0599] Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu
 [0600] 50 55 60
 [0601] Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu Ala Trp
 [0602] 65 70 75 80
 [0603] Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe Thr Ser
 [0604] 85 90 95
 [0605] Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu Ser Met
 [0606] 100 105 110
 [0607] Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn Gly Leu
 [0608] 115 120 125
 [0609] Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe Arg Leu
 [0610] 130 135 140
 [0611] Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val Gln Tyr
 [0612] 145 150 155 160
 [0613] Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn Tyr Asp
 [0614] 165 170 175
 [0615] Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu Arg Asp
 [0616] 180 185 190
 [0617] Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe Ile Glu
 [0618] 195 200 205
 [0619] Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly Glu Gln
 [0620] 210 215 220
 [0621] Trp Arg
 [0622] 225
 [0623] <210> 18

[0624] <211> 760
 [0625] <212> PRT
 [0626] <213> 伯顿拟甲烷球菌(Methanococcoides burtonii)
 [0627] <400> 18
 [0628] Met Met Ile Arg Glu Leu Asp Ile Pro Arg Asp Ile Ile Gly Phe Tyr
 [0629] 1 5 10 15
 [0630] Glu Asp Ser Gly Ile Lys Glu Leu Tyr Pro Pro Gln Ala Glu Ala Ile
 [0631] 20 25 30
 [0632] Glu Met Gly Leu Leu Glu Lys Lys Asn Leu Leu Ala Ala Ile Pro Thr
 [0633] 35 40 45
 [0634] Ala Ser Gly Lys Thr Leu Leu Ala Glu Leu Ala Met Ile Lys Ala Ile
 [0635] 50 55 60
 [0636] Arg Glu Gly Gly Lys Ala Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala
 [0637] 65 70 75 80
 [0638] Ser Glu Lys Phe Glu Arg Phe Lys Glu Leu Ala Pro Phe Gly Ile Lys
 [0639] 85 90 95
 [0640] Val Gly Ile Ser Thr Gly Asp Leu Asp Ser Arg Ala Asp Trp Leu Gly
 [0641] 100 105 110
 [0642] Val Asn Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu
 [0643] 115 120 125
 [0644] Arg Asn Gly Thr Ser Trp Met Asp Glu Ile Thr Thr Val Val Val Asp
 [0645] 130 135 140
 [0646] Glu Ile His Leu Leu Asp Ser Lys Asn Arg Gly Pro Thr Leu Glu Val
 [0647] 145 150 155 160
 [0648] Thr Ile Thr Lys Leu Met Arg Leu Asn Pro Asp Val Gln Val Val Ala
 [0649] 165 170 175
 [0650] Leu Ser Ala Thr Val Gly Asn Ala Arg Glu Met Ala Asp Trp Leu Gly
 [0651] 180 185 190
 [0652] Ala Ala Leu Val Leu Ser Glu Trp Arg Pro Thr Asp Leu His Glu Gly
 [0653] 195 200 205
 [0654] Val Leu Phe Gly Asp Ala Ile Asn Phe Pro Gly Ser Gln Lys Lys Ile
 [0655] 210 215 220
 [0656] Asp Arg Leu Glu Lys Asp Asp Ala Val Asn Leu Val Leu Asp Thr Ile
 [0657] 225 230 235 240
 [0658] Lys Ala Glu Gly Gln Cys Leu Val Phe Glu Ser Ser Arg Arg Asn Cys
 [0659] 245 250 255
 [0660] Ala Gly Phe Ala Lys Thr Ala Ser Ser Lys Val Ala Lys Ile Leu Asp
 [0661] 260 265 270
 [0662] Asn Asp Ile Met Ile Lys Leu Ala Gly Ile Ala Glu Glu Val Glu Ser

[0663]	275	280	285
[0664]	Thr Gly Glu Thr Asp Thr Ala Ile Val Leu Ala Asn Cys Ile Arg Lys		
[0665]	290	295	300
[0666]	Gly Val Ala Phe His His Ala Gly Leu Asn Ser Asn His Arg Lys Leu		
[0667]	305	310	315
[0668]	Val Glu Asn Gly Phe Arg Gln Asn Leu Ile Lys Val Ile Ser Ser Thr		
[0669]	325	330	335
[0670]	Pro Thr Leu Ala Ala Gly Leu Asn Leu Pro Ala Arg Arg Val Ile Ile		
[0671]	340	345	350
[0672]	Arg Ser Tyr Arg Arg Phe Asp Ser Asn Phe Gly Met Gln Pro Ile Pro		
[0673]	355	360	365
[0674]	Val Leu Glu Tyr Lys Gln Met Ala Gly Arg Ala Gly Arg Pro His Leu		
[0675]	370	375	380
[0676]	Asp Pro Tyr Gly Glu Ser Val Leu Leu Ala Lys Thr Tyr Asp Glu Phe		
[0677]	385	390	395
[0678]	Ala Gln Leu Met Glu Asn Tyr Val Glu Ala Asp Ala Glu Asp Ile Trp		
[0679]	405	410	415
[0680]	Ser Lys Leu Gly Thr Glu Asn Ala Leu Arg Thr His Val Leu Ser Thr		
[0681]	420	425	430
[0682]	Ile Val Asn Gly Phe Ala Ser Thr Arg Gln Glu Leu Phe Asp Phe Phe		
[0683]	435	440	445
[0684]	Gly Ala Thr Phe Phe Ala Tyr Gln Gln Asp Lys Trp Met Leu Glu Glu		
[0685]	450	455	460
[0686]	Val Ile Asn Asp Cys Leu Glu Phe Leu Ile Asp Lys Ala Met Val Ser		
[0687]	465	470	475
[0688]	Glu Thr Glu Asp Ile Glu Asp Ala Ser Lys Leu Phe Leu Arg Gly Thr		
[0689]	485	490	495
[0690]	Arg Leu Gly Ser Leu Val Ser Met Leu Tyr Ile Asp Pro Leu Ser Gly		
[0691]	500	505	510
[0692]	Ser Lys Ile Val Asp Gly Phe Lys Asp Ile Gly Lys Ser Thr Gly Gly		
[0693]	515	520	525
[0694]	Asn Met Gly Ser Leu Glu Asp Asp Lys Gly Asp Asp Ile Thr Val Thr		
[0695]	530	535	540
[0696]	Asp Met Thr Leu Leu His Leu Val Cys Ser Thr Pro Asp Met Arg Gln		
[0697]	545	550	555
[0698]	Leu Tyr Leu Arg Asn Thr Asp Tyr Thr Ile Val Asn Glu Tyr Ile Val		
[0699]	565	570	575
[0700]	Ala His Ser Asp Glu Phe His Glu Ile Pro Asp Lys Leu Lys Glu Thr		
[0701]	580	585	590

[0702]	Asp	Tyr	Glu	Trp	Phe	Met	Gly	Glu	Val	Lys	Thr	Ala	Met	Leu	Leu	Glu
[0703]						595										600
[0704]	Glu	Trp	Val	Thr	Glu	Val	Ser	Ala	Glu	Asp	Ile	Thr	Arg	His	Phe	Asn
[0705]						610										615
[0706]	Val	Gly	Glu	Gly	Asp	Ile	His	Ala	Leu	Ala	Asp	Thr	Ser	Glu	Trp	Leu
[0707]						625										630
[0708]	Met	His	Ala	Ala	Ala	Lys	Leu	Ala	Glu	Leu	Leu	Gly	Val	Glu	Tyr	Ser
[0709]						645										650
[0710]	Ser	His	Ala	Tyr	Ser	Leu	Glu	Lys	Arg	Ile	Arg	Tyr	Gly	Ser	Gly	Leu
[0711]						660										665
[0712]	Asp	Leu	Met	Glu	Leu	Val	Gly	Ile	Arg	Gly	Val	Gly	Arg	Val	Arg	Ala
[0713]						675										680
[0714]	Arg	Lys	Leu	Tyr	Asn	Ala	Gly	Phe	Val	Ser	Val	Ala	Lys	Leu	Lys	Gly
[0715]						690										695
[0716]	Ala	Asp	Ile	Ser	Val	Leu	Ser	Lys	Leu	Val	Gly	Pro	Lys	Val	Ala	Tyr
[0717]						705										710
[0718]	Asn	Ile	Leu	Ser	Gly	Ile	Gly	Val	Arg	Val	Asn	Asp	Lys	His	Phe	Asn
[0719]						725										730
[0720]	Ser	Ala	Pro	Ile	Ser	Ser	Asn	Thr	Leu	Asp	Thr	Leu	Leu	Asp	Lys	Asn
[0721]						740										745
[0722]	Gln	Lys	Thr	Phe	Asn	Asp	Phe	Gln								
[0723]						755										760
[0724]	<210> 19															
[0725]	<211> 707															
[0726]	<212> PRT															
[0727]	<213> 共生餐古菌(Cenarchaeum symbiosum)															
[0728]	<400> 19															
[0729]	Met	Arg	Ile	Ser	Glu	Leu	Asp	Ile	Pro	Arg	Pro	Ala	Ile	Glu	Phe	Leu
[0730]						1										5
[0731]	Glu	Gly	Glu	Gly	Tyr	Lys	Lys	Leu	Tyr	Pro	Pro	Gln	Ala	Ala	Ala	Ala
[0732]						20										25
[0733]	Lys	Ala	Gly	Leu	Thr	Asp	Gly	Lys	Ser	Val	Leu	Val	Ser	Ala	Pro	Thr
[0734]						35										40
[0735]	Ala	Ser	Gly	Lys	Thr	Leu	Ile	Ala	Ala	Ile	Ala	Met	Ile	Ser	His	Leu
[0736]						50										55
[0737]	Ser	Arg	Asn	Arg	Gly	Lys	Ala	Val	Tyr	Leu	Ser	Pro	Leu	Arg	Ala	Leu
[0738]						65										70
[0739]	Ala	Ala	Glu	Lys	Phe	Ala	Glu	Phe	Gly	Lys	Ile	Gly	Gly	Ile	Pro	Leu
[0740]						85										90

[0741]	Gly	Arg	Pro	Val	Arg	Val	Gly	Val	Ser	Thr	Gly	Asp	Phe	Glu	Lys	Ala
[0742]				100					105					110		
[0743]	Gly	Arg	Ser	Leu	Gly	Asn	Asn	Asp	Ile	Leu	Val	Leu	Thr	Asn	Glu	Arg
[0744]				115					120					125		
[0745]	Met	Asp	Ser	Leu	Ile	Arg	Arg	Arg	Pro	Asp	Trp	Met	Asp	Glu	Val	Gly
[0746]				130					135					140		
[0747]	Leu	Val	Ile	Ala	Asp	Glu	Ile	His	Leu	Ile	Gly	Asp	Arg	Ser	Arg	Gly
[0748]	145								150					155		160
[0749]	Pro	Thr	Leu	Glu	Met	Val	Leu	Thr	Lys	Leu	Arg	Gly	Leu	Arg	Ser	Ser
[0750]									165					170		175
[0751]	Pro	Gln	Val	Val	Ala	Leu	Ser	Ala	Thr	Ile	Ser	Asn	Ala	Asp	Glu	Ile
[0752]														180		185
[0753]	Ala	Gly	Trp	Leu	Asp	Cys	Thr	Leu	Val	His	Ser	Thr	Trp	Arg	Pro	Val
[0754]														195		200
[0755]	Pro	Leu	Ser	Glu	Gly	Val	Tyr	Gln	Asp	Gly	Glu	Val	Ala	Met	Gly	Asp
[0756]														210		215
[0757]	Gly	Ser	Arg	His	Glu	Val	Ala	Ala	Thr	Gly	Gly	Gly	Pro	Ala	Val	Asp
[0758]	225													230		235
[0759]	Leu	Ala	Ala	Glu	Ser	Val	Ala	Glu	Gly	Gly	Gln	Ser	Leu	Ile	Phe	Ala
[0760]														245		250
[0761]	Asp	Thr	Arg	Ala	Arg	Ser	Ala	Ser	Leu	Ala	Ala	Lys	Ala	Ser	Ala	Val
[0762]														260		265
[0763]	Ile	Pro	Glu	Ala	Lys	Gly	Ala	Asp	Ala	Ala	Lys	Leu	Ala	Ala	Ala	Ala
[0764]														275		280
[0765]	Lys	Lys	Ile	Ile	Ser	Ser	Gly	Gly	Glu	Thr	Lys	Leu	Ala	Lys	Thr	Leu
[0766]														290		295
[0767]	Ala	Glu	Leu	Val	Glu	Lys	Gly	Ala	Ala	Phe	His	His	Ala	Gly	Leu	Asn
[0768]	305													310		315
[0769]	Gln	Asp	Cys	Arg	Ser	Val	Val	Glu	Glu	Glu	Phe	Arg	Ser	Gly	Arg	Ile
[0770]														325		330
[0771]	Arg	Leu	Leu	Ala	Ser	Thr	Pro	Thr	Leu	Ala	Ala	Gly	Val	Asn	Leu	Pro
[0772]														340		345
[0773]	Ala	Arg	Arg	Val	Val	Ile	Ser	Ser	Val	Met	Arg	Tyr	Asn	Ser	Ser	Ser
[0774]														355		360
[0775]	Gly	Met	Ser	Glu	Pro	Ile	Ser	Ile	Leu	Glu	Tyr	Lys	Gln	Leu	Cys	Gly
[0776]														370		375
[0777]	Arg	Ala	Gly	Arg	Pro	Gln	Tyr	Asp	Lys	Ser	Gly	Glu	Ala	Ile	Val	Val
[0778]	385													390		395
[0779]	Gly	Gly	Val	Asn	Ala	Asp	Glu	Ile	Phe	Asp	Arg	Tyr	Ile	Gly	Gly	Glu

[0780]	405							410							415							
[0781]	Pro	Glu	Pro	Ile	Arg	Ser	Ala	Met	Val	Asp	Asp	Arg	Ala	Leu	Arg	Ile						
[0782]	420							425							430							
[0783]	His	Val	Leu	Ser	Leu	Val	Thr	Thr	Ser	Pro	Gly	Ile	Lys	Glu	Asp	Asp						
[0784]	435							440							445							
[0785]	Val	Thr	Glu	Phe	Phe	Leu	Gly	Thr	Leu	Gly	Gly	Gln	Gln	Ser	Gly	Glu						
[0786]	450							455							460							
[0787]	Ser	Thr	Val	Lys	Phe	Ser	Val	Ala	Val	Ala	Leu	Arg	Phe	Leu	Gln	Glu						
[0788]	465	470							475							480						
[0789]	Glu	Gly	Met	Leu	Gly	Arg	Arg	Gly	Gly	Arg	Leu	Ala	Ala	Thr	Lys	Met						
[0790]	485							490							495							
[0791]	Gly	Arg	Leu	Val	Ser	Arg	Leu	Tyr	Met	Asp	Pro	Met	Thr	Ala	Val	Thr						
[0792]	500							505							510							
[0793]	Leu	Arg	Asp	Ala	Val	Gly	Glu	Ala	Ser	Pro	Gly	Arg	Met	His	Thr	Leu						
[0794]	515							520							525							
[0795]	Gly	Phe	Leu	His	Leu	Val	Ser	Glu	Cys	Ser	Glu	Phe	Met	Pro	Arg	Phe						
[0796]	530							535							540							
[0797]	Ala	Leu	Arg	Gln	Lys	Asp	His	Glu	Val	Ala	Glu	Met	Met	Leu	Glu	Ala						
[0798]	545	550							555							560						
[0799]	Gly	Arg	Gly	Glu	Leu	Leu	Arg	Pro	Val	Tyr	Ser	Tyr	Glu	Cys	Gly	Arg						
[0800]	565							570							575							
[0801]	Gly	Leu	Leu	Ala	Leu	His	Arg	Trp	Ile	Gly	Glu	Ser	Pro	Glu	Ala	Lys						
[0802]	580							585							590							
[0803]	Leu	Ala	Glu	Asp	Leu	Lys	Phe	Glu	Ser	Gly	Asp	Val	His	Arg	Met	Val						
[0804]	595							600							605							
[0805]	Glu	Ser	Ser	Gly	Trp	Leu	Leu	Arg	Cys	Ile	Trp	Glu	Ile	Ser	Lys	His						
[0806]	610							615							620							
[0807]	Gln	Glu	Arg	Pro	Asp	Leu	Leu	Gly	Glu	Leu	Asp	Val	Leu	Arg	Ser	Arg						
[0808]	625	630							635							640						
[0809]	Val	Ala	Tyr	Gly	Ile	Lys	Ala	Glu	Leu	Val	Pro	Leu	Val	Ser	Ile	Lys						
[0810]	645							650							655							
[0811]	Gly	Ile	Gly	Arg	Val	Arg	Ser	Arg	Arg	Leu	Phe	Arg	Gly	Gly	Ile	Lys						
[0812]	660							665							670							
[0813]	Gly	Pro	Gly	Asp	Leu	Ala	Ala	Val	Pro	Val	Glu	Arg	Leu	Ser	Arg	Val						
[0814]	675							680							685							
[0815]	Glu	Gly	Ile	Gly	Ala	Thr	Leu	Ala	Asn	Asn	Ile	Lys	Ser	Gln	Leu	Arg						
[0816]	690							695							700							
[0817]	Lys	Gly	Gly																			
[0818]	705																					

[0819]	<210> 20															
[0820]	<211> 720															
[0821]	<212> PRT															
[0822]	<213> 抗辐射热球菌(Thermococcus gammatolerans)															
[0823]	<400> 20															
[0824]	Met	Lys	Val	Asp	Glu	Leu	Pro	Val	Asp	Glu	Arg	Leu	Lys	Ala	Val	Leu
[0825]	1				5					10					15	
[0826]	Lys	Glu	Arg	Gly	Ile	Glu	Glu	Leu	Tyr	Pro	Pro	Gln	Ala	Glu	Ala	Leu
[0827]					20					25					30	
[0828]	Lys	Ser	Gly	Ala	Leu	Glu	Gly	Arg	Asn	Leu	Val	Leu	Ala	Ile	Pro	Thr
[0829]					35					40				45		
[0830]	Ala	Ser	Gly	Lys	Thr	Leu	Val	Ser	Glu	Ile	Val	Met	Val	Asn	Lys	Leu
[0831]					50					55				60		
[0832]	Ile	Gln	Glu	Gly	Gly	Lys	Ala	Val	Tyr	Leu	Val	Pro	Leu	Lys	Ala	Leu
[0833]	65									70				75		80
[0834]	Ala	Glu	Glu	Lys	Tyr	Arg	Glu	Phe	Lys	Glu	Trp	Glu	Lys	Leu	Gly	Leu
[0835]										85				90		95
[0836]	Lys	Val	Ala	Ala	Thr	Thr	Gly	Asp	Tyr	Asp	Ser	Thr	Asp	Asp	Trp	Leu
[0837]										100				105		110
[0838]	Gly	Arg	Tyr	Asp	Ile	Ile	Val	Ala	Thr	Ala	Glu	Lys	Phe	Asp	Ser	Leu
[0839]										115				120		125
[0840]	Leu	Arg	His	Gly	Ala	Arg	Trp	Ile	Asn	Asp	Val	Lys	Leu	Val	Val	Ala
[0841]										130				135		140
[0842]	Asp	Glu	Val	His	Leu	Ile	Gly	Ser	Tyr	Asp	Arg	Gly	Ala	Thr	Leu	Glu
[0843]	145									150				155		160
[0844]	Met	Ile	Leu	Thr	His	Met	Leu	Gly	Arg	Ala	Gln	Ile	Leu	Ala	Leu	Ser
[0845]										165				170		175
[0846]	Ala	Thr	Val	Gly	Asn	Ala	Glu	Glu	Leu	Ala	Glu	Trp	Leu	Asp	Ala	Ser
[0847]										180				185		190
[0848]	Leu	Val	Val	Ser	Asp	Trp	Arg	Pro	Val	Gln	Leu	Arg	Arg	Gly	Val	Phe
[0849]										195				200		205
[0850]	His	Leu	Gly	Thr	Leu	Ile	Trp	Glu	Asp	Gly	Lys	Val	Glu	Ser	Tyr	Pro
[0851]										210				215		220
[0852]	Glu	Asn	Trp	Tyr	Ser	Leu	Val	Val	Asp	Ala	Val	Lys	Arg	Gly	Lys	Gly
[0853]	225									230				235		240
[0854]	Ala	Leu	Val	Phe	Val	Asn	Thr	Arg	Arg	Ser	Ala	Glu	Lys	Glu	Ala	Leu
[0855]										245				250		255
[0856]	Ala	Leu	Ser	Lys	Leu	Val	Ser	Ser	His	Leu	Thr	Lys	Pro	Glu	Lys	Arg
[0857]										260				265		270

[0858]	Ala	Leu	Glu	Ser	Leu	Ala	Ser	Gln	Leu	Glu	Asp	Asn	Pro	Thr	Ser	Glu
[0859]		275						280					285			
[0860]	Lys	Leu	Lys	Arg	Ala	Leu	Arg	Gly	Gly	Val	Ala	Phe	His	His	Ala	Gly
[0861]		290						295					300			
[0862]	Leu	Ser	Arg	Val	Glu	Arg	Thr	Leu	Ile	Glu	Asp	Ala	Phe	Arg	Glu	Gly
[0863]	305						310				315					320
[0864]	Leu	Ile	Lys	Val	Ile	Thr	Ala	Thr	Pro	Thr	Leu	Ser	Ala	Gly	Val	Asn
[0865]					325					330					335	
[0866]	Leu	Pro	Ser	Phe	Arg	Val	Ile	Ile	Arg	Asp	Thr	Lys	Arg	Tyr	Ala	Gly
[0867]				340					345					350		
[0868]	Phe	Gly	Trp	Thr	Asp	Ile	Pro	Val	Leu	Glu	Ile	Gln	Gln	Met	Met	Gly
[0869]			355					360					365			
[0870]	Arg	Ala	Gly	Arg	Pro	Arg	Tyr	Asp	Lys	Tyr	Gly	Glu	Ala	Ile	Ile	Val
[0871]		370						375				380				
[0872]	Ala	Arg	Thr	Asp	Glu	Pro	Gly	Lys	Leu	Met	Glu	Arg	Tyr	Ile	Arg	Gly
[0873]	385					390					395					400
[0874]	Lys	Pro	Glu	Lys	Leu	Phe	Ser	Met	Leu	Ala	Asn	Glu	Gln	Ala	Phe	Arg
[0875]				405						410					415	
[0876]	Ser	Gln	Val	Leu	Ala	Leu	Ile	Thr	Asn	Phe	Gly	Ile	Arg	Ser	Phe	Pro
[0877]			420						425				430			
[0878]	Glu	Leu	Val	Arg	Phe	Leu	Glu	Arg	Thr	Phe	Tyr	Ala	His	Gln	Arg	Lys
[0879]		435						440					445			
[0880]	Asp	Leu	Ser	Ser	Leu	Glu	Tyr	Lys	Ala	Lys	Glu	Val	Val	Tyr	Phe	Leu
[0881]		450						455				460				
[0882]	Ile	Glu	Asn	Glu	Phe	Ile	Asp	Leu	Asp	Leu	Glu	Asp	Arg	Phe	Ile	Pro
[0883]	465					470					475					480
[0884]	Leu	Pro	Phe	Gly	Lys	Arg	Thr	Ser	Gln	Leu	Tyr	Ile	Asp	Pro	Leu	Thr
[0885]			485						490				495			
[0886]	Ala	Lys	Lys	Phe	Lys	Asp	Ala	Phe	Pro	Ala	Ile	Glu	Arg	Asn	Pro	Asn
[0887]			500						505				510			
[0888]	Pro	Phe	Gly	Ile	Phe	Gln	Leu	Ile	Ala	Ser	Thr	Pro	Asp	Met	Ala	Thr
[0889]			515						520				525			
[0890]	Leu	Thr	Ala	Arg	Arg	Arg	Glu	Met	Glu	Asp	Tyr	Leu	Asp	Leu	Ala	Tyr
[0891]		530						535				540				
[0892]	Glu	Leu	Glu	Asp	Lys	Leu	Tyr	Ala	Ser	Ile	Pro	Tyr	Tyr	Glu	Asp	Ser
[0893]	545					550					555					560
[0894]	Arg	Phe	Gln	Gly	Phe	Leu	Gly	Gln	Val	Lys	Thr	Ala	Lys	Val	Leu	Leu
[0895]				565					570				575			
[0896]	Asp	Trp	Ile	Asn	Glu	Val	Pro	Glu	Ala	Arg	Ile	Tyr	Glu	Thr	Tyr	Ser

[0897]	580	585	590
[0898]	Ile Asp Pro Gly Asp Leu Tyr Arg Leu Leu Glu Leu Ala Asp Trp Leu		
[0899]	595	600	605
[0900]	Met Tyr Ser Leu Ile Glu Leu Tyr Lys Leu Phe Glu Pro Lys Glu Glu		
[0901]	610	615	620
[0902]	Ile Leu Asn Tyr Leu Arg Asp Leu His Leu Arg Leu Arg His Gly Val		
[0903]	625	630	635
[0904]	Arg Glu Glu Leu Leu Glu Leu Val Arg Leu Pro Asn Ile Gly Arg Lys		
[0905]	645	650	655
[0906]	Arg Ala Arg Ala Leu Tyr Asn Ala Gly Phe Arg Ser Val Glu Ala Ile		
[0907]	660	665	670
[0908]	Ala Asn Ala Lys Pro Ala Glu Leu Leu Ala Val Glu Gly Ile Gly Ala		
[0909]	675	680	685
[0910]	Lys Ile Leu Asp Gly Ile Tyr Arg His Leu Gly Ile Glu Lys Arg Val		
[0911]	690	695	700
[0912]	Thr Glu Glu Lys Pro Lys Arg Lys Gly Thr Leu Glu Asp Phe Leu Arg		
[0913]	705	710	715
[0914]	<210> 21		
[0915]	<211> 799		
[0916]	<212> PRT		
[0917]	<213> 亨氏甲烷螺旋菌(Methanospirillum hungatei)		
[0918]	<400> 21		
[0919]	Met Glu Ile Ala Ser Leu Pro Leu Pro Asp Ser Phe Ile Arg Ala Cys		
[0920]	1	5	10
[0921]	His Ala Lys Gly Ile Arg Ser Leu Tyr Pro Pro Gln Ala Glu Cys Ile		
[0922]	20	25	30
[0923]	Glu Lys Gly Leu Leu Glu Gly Lys Asn Leu Leu Ile Ser Ile Pro Thr		
[0924]	35	40	45
[0925]	Ala Ser Gly Lys Thr Leu Leu Ala Glu Met Ala Met Trp Ser Arg Ile		
[0926]	50	55	60
[0927]	Ala Ala Gly Gly Lys Cys Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala		
[0928]	65	70	75
[0929]	Ser Glu Lys Tyr Asp Glu Phe Ser Lys Lys Gly Val Ile Arg Val Gly		
[0930]	85	90	95
[0931]	Ile Ala Thr Gly Asp Leu Asp Arg Thr Asp Ala Tyr Leu Gly Glu Asn		
[0932]	100	105	110
[0933]	Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu Arg Asn		
[0934]	115	120	125
[0935]	Arg Thr Pro Trp Leu Ser Gln Ile Thr Cys Ile Val Leu Asp Glu Val		

[0936]	130	135	140
[0937]	His Leu Ile Gly Ser Glu Asn Arg Gly Ala Thr Leu Glu Met Val Ile		
[0938]	145	150	155
[0939]	Thr Lys Leu Arg Tyr Thr Asn Pro Val Met Gln Ile Ile Gly Leu Ser		
[0940]	165	170	175
[0941]	Ala Thr Ile Gly Asn Pro Ala Gln Leu Ala Glu Trp Leu Asp Ala Thr		
[0942]	180	185	190
[0943]	Leu Ile Thr Ser Thr Trp Arg Pro Val Asp Leu Arg Gln Gly Val Tyr		
[0944]	195	200	205
[0945]	Tyr Asn Gly Lys Ile Arg Phe Ser Asp Ser Glu Arg Pro Ile Gln Gly		
[0946]	210	215	220
[0947]	Lys Thr Lys His Asp Asp Leu Asn Leu Cys Leu Asp Thr Ile Glu Glu		
[0948]	225	230	235
[0949]	Gly Gly Gln Cys Leu Val Phe Val Ser Ser Arg Arg Asn Ala Glu Gly		
[0950]	245	250	255
[0951]	Phe Ala Lys Lys Ala Ala Gly Ala Leu Lys Ala Gly Ser Pro Asp Ser		
[0952]	260	265	270
[0953]	Lys Ala Leu Ala Gln Glu Leu Arg Arg Leu Arg Asp Arg Asp Glu Gly		
[0954]	275	280	285
[0955]	Asn Val Leu Ala Asp Cys Val Glu Arg Gly Ala Ala Phe His His Ala		
[0956]	290	295	300
[0957]	Gly Leu Ile Arg Gln Glu Arg Thr Ile Ile Glu Glu Gly Phe Arg Asn		
[0958]	305	310	315
[0959]	Gly Tyr Ile Glu Val Ile Ala Ala Thr Pro Thr Leu Ala Ala Gly Leu		
[0960]	325	330	335
[0961]	Asn Leu Pro Ala Arg Arg Val Ile Ile Arg Asp Tyr Asn Arg Phe Ala		
[0962]	340	345	350
[0963]	Ser Gly Leu Gly Met Val Pro Ile Pro Val Gly Glu Tyr His Gln Met		
[0964]	355	360	365
[0965]	Ala Gly Arg Ala Gly Arg Pro His Leu Asp Pro Tyr Gly Glu Ala Val		
[0966]	370	375	380
[0967]	Leu Leu Ala Lys Asp Ala Pro Ser Val Glu Arg Leu Phe Glu Thr Phe		
[0968]	385	390	395
[0969]	Ile Asp Ala Glu Ala Glu Arg Val Asp Ser Gln Cys Val Asp Asp Ala		
[0970]	405	410	415
[0971]	Ser Leu Cys Ala His Ile Leu Ser Leu Ile Ala Thr Gly Phe Ala His		
[0972]	420	425	430
[0973]	Asp Gln Glu Ala Leu Ser Ser Phe Met Glu Arg Thr Phe Tyr Phe Phe		
[0974]	435	440	445

[0975]	Gln His Pro Lys Thr Arg Ser Leu Pro Arg Leu Val Ala Asp Ala Ile
[0976]	450 455 460
[0977]	Arg Phe Leu Thr Thr Ala Gly Met Val Glu Glu Arg Glu Asn Thr Leu
[0978]	465 470 475 480
[0979]	Ser Ala Thr Arg Leu Gly Ser Leu Val Ser Arg Leu Tyr Leu Asn Pro
[0980]	485 490 495
[0981]	Cys Thr Ala Arg Leu Ile Leu Asp Ser Leu Lys Ser Cys Lys Thr Pro
[0982]	500 505 510
[0983]	Thr Leu Ile Gly Leu Leu His Val Ile Cys Val Ser Pro Asp Met Gln
[0984]	515 520 525
[0985]	Arg Leu Tyr Leu Lys Ala Ala Asp Thr Gln Leu Leu Arg Thr Phe Leu
[0986]	530 535 540
[0987]	Phe Lys His Lys Asp Asp Leu Ile Leu Pro Leu Pro Phe Glu Gln Glu
[0988]	545 550 555 560
[0989]	Glu Glu Glu Leu Trp Leu Ser Gly Leu Lys Thr Ala Leu Val Leu Thr
[0990]	565 570 575
[0991]	Asp Trp Ala Asp Glu Phe Ser Glu Gly Met Ile Glu Glu Arg Tyr Gly
[0992]	580 585 590
[0993]	Ile Gly Ala Gly Asp Leu Tyr Asn Ile Val Asp Ser Gly Lys Trp Leu
[0994]	595 600 605
[0995]	Leu His Gly Thr Glu Arg Leu Val Ser Val Glu Met Pro Glu Met Ser
[0996]	610 615 620
[0997]	Gln Val Val Lys Thr Leu Ser Val Arg Val His His Gly Val Lys Ser
[0998]	625 630 635 640
[0999]	Glu Leu Leu Pro Leu Val Ala Leu Arg Asn Ile Gly Arg Val Arg Ala
[1000]	645 650 655
[1001]	Arg Thr Leu Tyr Asn Ala Gly Tyr Pro Asp Pro Glu Ala Val Ala Arg
[1002]	660 665 670
[1003]	Ala Gly Leu Ser Thr Ile Ala Arg Ile Ile Gly Glu Gly Ile Ala Arg
[1004]	675 680 685
[1005]	Gln Val Ile Asp Glu Ile Thr Gly Val Lys Arg Ser Gly Ile His Ser
[1006]	690 695 700
[1007]	Ser Asp Asp Asp Tyr Gln Gln Lys Thr Pro Glu Leu Leu Thr Asp Ile
[1008]	705 710 715 720
[1009]	Pro Gly Ile Gly Lys Lys Met Ala Glu Lys Leu Gln Asn Ala Gly Ile
[1010]	725 730 735
[1011]	Ile Thr Val Ser Asp Leu Leu Thr Ala Asp Glu Val Leu Leu Ser Asp
[1012]	740 745 750
[1013]	Val Leu Gly Ala Ala Arg Ala Arg Lys Val Leu Ala Phe Leu Ser Asn

[1014]	755	760	765
[1015]	Ser Glu Lys Glu Asn Ser Ser Ser Asp Lys Thr Glu Glu Ile Pro Asp		
[1016]	770	775	780
[1017]	Thr Gln Lys Ile Arg Gly Gln Ser Ser Trp Glu Asp Phe Gly Cys		
[1018]	785	790	795
[1019]	<210> 22		
[1020]	<211> 1756		
[1021]	<212> PRT		
[1022]	<213> 大肠杆菌		
[1023]	<400> 22		
[1024]	Met Met Ser Ile Ala Gln Val Arg Ser Ala Gly Ser Ala Gly Asn Tyr		
[1025]	1	5	10
[1026]	Tyr Thr Asp Lys Asp Asn Tyr Tyr Val Leu Gly Ser Met Gly Glu Arg		
[1027]	20	25	30
[1028]	Trp Ala Gly Lys Gly Ala Glu Gln Leu Gly Leu Gln Gly Ser Val Asp		
[1029]	35	40	45
[1030]	Lys Asp Val Phe Thr Arg Leu Leu Glu Gly Arg Leu Pro Asp Gly Ala		
[1031]	50	55	60
[1032]	Asp Leu Ser Arg Met Gln Asp Gly Ser Asn Lys His Arg Pro Gly Tyr		
[1033]	65	70	75
[1034]	Asp Leu Thr Phe Ser Ala Pro Lys Ser Val Ser Met Met Ala Met Leu		
[1035]	85	90	95
[1036]	Gly Gly Asp Lys Arg Leu Ile Asp Ala His Asn Gln Ala Val Asp Phe		
[1037]	100	105	110
[1038]	Ala Val Arg Gln Val Glu Ala Leu Ala Ser Thr Arg Val Met Thr Asp		
[1039]	115	120	125
[1040]	Gly Gln Ser Glu Thr Val Leu Thr Gly Asn Leu Val Met Ala Leu Phe		
[1041]	130	135	140
[1042]	Asn His Asp Thr Ser Arg Asp Gln Glu Pro Gln Leu His Thr His Ala		
[1043]	145	150	155
[1044]	Val Val Ala Asn Val Thr Gln His Asn Gly Glu Trp Lys Thr Leu Ser		
[1045]	165	170	175
[1046]	Ser Asp Lys Val Gly Lys Thr Gly Phe Ile Glu Asn Val Tyr Ala Asn		
[1047]	180	185	190
[1048]	Gln Ile Ala Phe Gly Arg Leu Tyr Arg Glu Lys Leu Lys Glu Gln Val		
[1049]	195	200	205
[1050]	Glu Ala Leu Gly Tyr Glu Thr Glu Val Val Gly Lys His Gly Met Trp		
[1051]	210	215	220
[1052]	Glu Met Pro Gly Val Pro Val Glu Ala Phe Ser Gly Arg Ser Gln Ala		

[1053]	225	230	235	240
[1054]	Ile Arg Glu Ala Val Gly Glu Asp Ala Ser Leu Lys Ser Arg Asp Val			
[1055]		245	250	255
[1056]	Ala Ala Leu Asp Thr Arg Lys Ser Lys Gln His Val Asp Pro Glu Ile			
[1057]		260	265	270
[1058]	Arg Met Ala Glu Trp Met Gln Thr Leu Lys Glu Thr Gly Phe Asp Ile			
[1059]		275	280	285
[1060]	Arg Ala Tyr Arg Asp Ala Ala Asp Gln Arg Thr Glu Ile Arg Thr Gln			
[1061]		290	295	300
[1062]	Ala Pro Gly Pro Ala Ser Gln Asp Gly Pro Asp Val Gln Gln Ala Val			
[1063]	305	310	315	320
[1064]	Thr Gln Ala Ile Ala Gly Leu Ser Glu Arg Lys Val Gln Phe Thr Tyr			
[1065]		325	330	335
[1066]	Thr Asp Val Leu Ala Arg Thr Val Gly Ile Leu Pro Pro Glu Asn Gly			
[1067]		340	345	350
[1068]	Val Ile Glu Arg Ala Arg Ala Gly Ile Asp Glu Ala Ile Ser Arg Glu			
[1069]		355	360	365
[1070]	Gln Leu Ile Pro Leu Asp Arg Glu Lys Gly Leu Phe Thr Ser Gly Ile			
[1071]		370	375	380
[1072]	His Val Leu Asp Glu Leu Ser Val Arg Ala Leu Ser Arg Asp Ile Met			
[1073]	385	390	395	400
[1074]	Lys Gln Asn Arg Val Thr Val His Pro Glu Lys Ser Val Pro Arg Thr			
[1075]		405	410	415
[1076]	Ala Gly Tyr Ser Asp Ala Val Ser Val Leu Ala Gln Asp Arg Pro Ser			
[1077]		420	425	430
[1078]	Leu Ala Ile Val Ser Gly Gln Gly Gly Ala Ala Gly Gln Arg Glu Arg			
[1079]		435	440	445
[1080]	Val Ala Glu Leu Val Met Met Ala Arg Glu Gln Gly Arg Glu Val Gln			
[1081]		450	455	460
[1082]	Ile Ile Ala Ala Asp Arg Arg Ser Gln Met Asn Leu Lys Gln Asp Glu			
[1083]	465	470	475	480
[1084]	Arg Leu Ser Gly Glu Leu Ile Thr Gly Arg Arg Gln Leu Leu Glu Gly			
[1085]		485	490	495
[1086]	Met Ala Phe Thr Pro Gly Ser Thr Val Ile Val Asp Gln Gly Glu Lys			
[1087]		500	505	510
[1088]	Leu Ser Leu Lys Glu Thr Leu Thr Leu Leu Asp Gly Ala Ala Arg His			
[1089]		515	520	525
[1090]	Asn Val Gln Val Leu Ile Thr Asp Ser Gly Gln Arg Thr Gly Thr Gly			
[1091]		530	535	540

[1092]	Ser	Ala	Leu	Met	Ala	Met	Lys	Asp	Ala	Gly	Val	Asn	Thr	Tyr	Arg	Trp
[1093]	545						550				555					560
[1094]	Gln	Gly	Gly	Glu	Gln	Arg	Pro	Ala	Thr	Ile	Ile	Ser	Glu	Pro	Asp	Arg
[1095]						565					570					575
[1096]	Asn	Val	Arg	Tyr	Ala	Arg	Leu	Ala	Gly	Asp	Phe	Ala	Ala	Ser	Val	Lys
[1097]						580					585					590
[1098]	Ala	Gly	Glu	Glu	Ser	Val	Ala	Gln	Val	Ser	Gly	Val	Arg	Glu	Gln	Ala
[1099]						595					600					605
[1100]	Ile	Leu	Thr	Gln	Ala	Ile	Arg	Ser	Glu	Leu	Lys	Thr	Gln	Gly	Val	Leu
[1101]						610										620
[1102]	Gly	His	Pro	Glu	Val	Thr	Met	Thr	Ala	Leu	Ser	Pro	Val	Trp	Leu	Asp
[1103]	625						630					635				640
[1104]	Ser	Arg	Ser	Arg	Tyr	Leu	Arg	Asp	Met	Tyr	Arg	Pro	Gly	Met	Val	Met
[1105]						645						650				655
[1106]	Glu	Gln	Trp	Asn	Pro	Glu	Thr	Arg	Ser	His	Asp	Arg	Tyr	Val	Ile	Asp
[1107]						660						665				670
[1108]	Arg	Val	Thr	Ala	Gln	Ser	His	Ser	Leu	Thr	Leu	Arg	Asp	Ala	Gln	Gly
[1109]						675										685
[1110]	Glu	Thr	Gln	Val	Val	Arg	Ile	Ser	Ser	Leu	Asp	Ser	Ser	Trp	Ser	Leu
[1111]						690										700
[1112]	Phe	Arg	Pro	Glu	Lys	Met	Pro	Val	Ala	Asp	Gly	Glu	Arg	Leu	Arg	Val
[1113]	705						710					715				720
[1114]	Thr	Gly	Lys	Ile	Pro	Gly	Leu	Arg	Val	Ser	Gly	Gly	Asp	Arg	Leu	Gln
[1115]						725						730				735
[1116]	Val	Ala	Ser	Val	Ser	Glu	Asp	Ala	Met	Thr	Val	Val	Val	Pro	Gly	Arg
[1117]						740										750
[1118]	Ala	Glu	Pro	Ala	Ser	Leu	Pro	Val	Ser	Asp	Ser	Pro	Phe	Thr	Ala	Leu
[1119]						755										765
[1120]	Lys	Leu	Glu	Asn	Gly	Trp	Val	Glu	Thr	Pro	Gly	His	Ser	Val	Ser	Asp
[1121]						770										780
[1122]	Ser	Ala	Thr	Val	Phe	Ala	Ser	Val	Thr	Gln	Met	Ala	Met	Asp	Asn	Ala
[1123]	785						790					795				800
[1124]	Thr	Leu	Asn	Gly	Leu	Ala	Arg	Ser	Gly	Arg	Asp	Val	Arg	Leu	Tyr	Ser
[1125]						805						810				815
[1126]	Ser	Leu	Asp	Glu	Thr	Arg	Thr	Ala	Glu	Lys	Leu	Ala	Arg	His	Pro	Ser
[1127]						820						825				830
[1128]	Phe	Thr	Val	Val	Ser	Glu	Gln	Ile	Lys	Ala	Arg	Ala	Gly	Glu	Thr	Leu
[1129]						835										845
[1130]	Leu	Glu	Thr	Ala	Ile	Ser	Leu	Gln	Lys	Ala	Gly	Leu	His	Thr	Pro	Ala

[1131]	850	855	860
[1132]	Gln Gln Ala Ile His Leu Ala Leu Pro Val Leu Glu Ser Lys Asn Leu		
[1133]	865	870	875 880
[1134]	Ala Phe Ser Met Val Asp Leu Leu Thr Glu Ala Lys Ser Phe Ala Ala		
[1135]	885	890	895
[1136]	Glu Gly Thr Gly Phe Thr Glu Leu Gly Gly Glu Ile Asn Ala Gln Ile		
[1137]	900	905	910
[1138]	Lys Arg Gly Asp Leu Leu Tyr Val Asp Val Ala Lys Gly Tyr Gly Thr		
[1139]	915	920	925
[1140]	Gly Leu Leu Val Ser Arg Ala Ser Tyr Glu Ala Glu Lys Ser Ile Leu		
[1141]	930	935	940
[1142]	Arg His Ile Leu Glu Gly Lys Glu Ala Val Thr Pro Leu Met Glu Arg		
[1143]	945	950	955 960
[1144]	Val Pro Gly Glu Leu Met Glu Thr Leu Thr Ser Gly Gln Arg Ala Ala		
[1145]	965	970	975
[1146]	Thr Arg Met Ile Leu Glu Thr Ser Asp Arg Phe Thr Val Val Gln Gly		
[1147]	980	985	990
[1148]	Tyr Ala Gly Val Gly Lys Thr Thr Gln Phe Arg Ala Val Met Ser Ala		
[1149]	995	1000	1005
[1150]	Val Asn Met Leu Pro Ala Ser Glu Arg Pro Arg Val Val Gly Leu		
[1151]	1010	1015	1020
[1152]	Gly Pro Thr His Arg Ala Val Gly Glu Met Arg Ser Ala Gly Val		
[1153]	1025	1030	1035
[1154]	Asp Ala Gln Thr Leu Ala Ser Phe Leu His Asp Thr Gln Leu Gln		
[1155]	1040	1045	1050
[1156]	Gln Arg Ser Gly Glu Thr Pro Asp Phe Ser Asn Thr Leu Phe Leu		
[1157]	1055	1060	1065
[1158]	Leu Asp Glu Ser Ser Met Val Gly Asn Thr Glu Met Ala Arg Ala		
[1159]	1070	1075	1080
[1160]	Tyr Ala Leu Ile Ala Ala Gly Gly Gly Arg Ala Val Ala Ser Gly		
[1161]	1085	1090	1095
[1162]	Asp Thr Asp Gln Leu Gln Ala Ile Ala Pro Gly Gln Ser Phe Arg		
[1163]	1100	1105	1110
[1164]	Leu Gln Gln Thr Arg Ser Ala Ala Asp Val Val Ile Met Lys Glu		
[1165]	1115	1120	1125
[1166]	Ile Val Arg Gln Thr Pro Glu Leu Arg Glu Ala Val Tyr Ser Leu		
[1167]	1130	1135	1140
[1168]	Ile Asn Arg Asp Val Glu Arg Ala Leu Ser Gly Leu Glu Ser Val		
[1169]	1145	1150	1155

[1170]	Lys Pro Ser Gln Val Pro Arg Leu Glu Gly Ala Trp Ala Pro Glu
[1171]	1160 1165 1170
[1172]	His Ser Val Thr Glu Phe Ser His Ser Gln Glu Ala Lys Leu Ala
[1173]	1175 1180 1185
[1174]	Glu Ala Gln Gln Lys Ala Met Leu Lys Gly Glu Ala Phe Pro Asp
[1175]	1190 1195 1200
[1176]	Ile Pro Met Thr Leu Tyr Glu Ala Ile Val Arg Asp Tyr Thr Gly
[1177]	1205 1210 1215
[1178]	Arg Thr Pro Glu Ala Arg Glu Gln Thr Leu Ile Val Thr His Leu
[1179]	1220 1225 1230
[1180]	Asn Glu Asp Arg Arg Val Leu Asn Ser Met Ile His Asp Ala Arg
[1181]	1235 1240 1245
[1182]	Glu Lys Ala Gly Glu Leu Gly Lys Glu Gln Val Met Val Pro Val
[1183]	1250 1255 1260
[1184]	Leu Asn Thr Ala Asn Ile Arg Asp Gly Glu Leu Arg Arg Leu Ser
[1185]	1265 1270 1275
[1186]	Thr Trp Glu Lys Asn Pro Asp Ala Leu Ala Leu Val Asp Asn Val
[1187]	1280 1285 1290
[1188]	Tyr His Arg Ile Ala Gly Ile Ser Lys Asp Asp Gly Leu Ile Thr
[1189]	1295 1300 1305
[1190]	Leu Gln Asp Ala Glu Gly Asn Thr Arg Leu Ile Ser Pro Arg Glu
[1191]	1310 1315 1320
[1192]	Ala Val Ala Glu Gly Val Thr Leu Tyr Thr Pro Asp Lys Ile Arg
[1193]	1325 1330 1335
[1194]	Val Gly Thr Gly Asp Arg Met Arg Phe Thr Lys Ser Asp Arg Glu
[1195]	1340 1345 1350
[1196]	Arg Gly Tyr Val Ala Asn Ser Val Trp Thr Val Thr Ala Val Ser
[1197]	1355 1360 1365
[1198]	Gly Asp Ser Val Thr Leu Ser Asp Gly Gln Gln Thr Arg Val Ile
[1199]	1370 1375 1380
[1200]	Arg Pro Gly Gln Glu Arg Ala Glu Gln His Ile Asp Leu Ala Tyr
[1201]	1385 1390 1395
[1202]	Ala Ile Thr Ala His Gly Ala Gln Gly Ala Ser Glu Thr Phe Ala
[1203]	1400 1405 1410
[1204]	Ile Ala Leu Glu Gly Thr Glu Gly Asn Arg Lys Leu Met Ala Gly
[1205]	1415 1420 1425
[1206]	Phe Glu Ser Ala Tyr Val Ala Leu Ser Arg Met Lys Gln His Val
[1207]	1430 1435 1440
[1208]	Gln Val Tyr Thr Asp Asn Arg Gln Gly Trp Thr Asp Ala Ile Asn

[1209]	1445	1450	1455
[1210]	Asn Ala Val Gln Lys Gly Thr Ala His Asp Val Leu Glu Pro Lys		
[1211]	1460	1465	1470
[1212]	Pro Asp Arg Glu Val Met Asn Ala Gln Arg Leu Phe Ser Thr Ala		
[1213]	1475	1480	1485
[1214]	Arg Glu Leu Arg Asp Val Ala Ala Gly Arg Ala Val Leu Arg Gln		
[1215]	1490	1495	1500
[1216]	Ala Gly Leu Ala Gly Gly Asp Ser Pro Ala Arg Phe Ile Ala Pro		
[1217]	1505	1510	1515
[1218]	Gly Arg Lys Tyr Pro Gln Pro Tyr Val Ala Leu Pro Ala Phe Asp		
[1219]	1520	1525	1530
[1220]	Arg Asn Gly Lys Ser Ala Gly Ile Trp Leu Asn Pro Leu Thr Thr		
[1221]	1535	1540	1545
[1222]	Asp Asp Gly Asn Gly Leu Arg Gly Phe Ser Gly Glu Gly Arg Val		
[1223]	1550	1555	1560
[1224]	Lys Gly Ser Gly Asp Ala Gln Phe Val Ala Leu Gln Gly Ser Arg		
[1225]	1565	1570	1575
[1226]	Asn Gly Glu Ser Leu Leu Ala Asp Asn Met Gln Asp Gly Val Arg		
[1227]	1580	1585	1590
[1228]	Ile Ala Arg Asp Asn Pro Asp Ser Gly Val Val Val Arg Ile Ala		
[1229]	1595	1600	1605
[1230]	Gly Glu Gly Arg Pro Trp Asn Pro Gly Ala Ile Thr Gly Gly Arg		
[1231]	1610	1615	1620
[1232]	Val Trp Gly Asp Ile Pro Asp Asn Ser Val Gln Pro Gly Ala Gly		
[1233]	1625	1630	1635
[1234]	Asn Gly Glu Pro Val Thr Ala Glu Val Leu Ala Gln Arg Gln Ala		
[1235]	1640	1645	1650
[1236]	Glu Glu Ala Ile Arg Arg Glu Thr Glu Arg Arg Ala Asp Glu Ile		
[1237]	1655	1660	1665
[1238]	Val Arg Lys Met Ala Glu Asn Lys Pro Asp Leu Pro Asp Gly Lys		
[1239]	1670	1675	1680
[1240]	Thr Glu Leu Ala Val Arg Asp Ile Ala Gly Gln Glu Arg Asp Arg		
[1241]	1685	1690	1695
[1242]	Ser Ala Ile Ser Glu Arg Glu Thr Ala Leu Pro Glu Ser Val Leu		
[1243]	1700	1705	1710
[1244]	Arg Glu Ser Gln Arg Glu Arg Glu Ala Val Arg Glu Val Ala Arg		
[1245]	1715	1720	1725
[1246]	Glu Asn Leu Leu Gln Glu Arg Leu Gln Gln Met Glu Arg Asp Met		
[1247]	1730	1735	1740

[1248]	Val Arg Asp Leu Gln Lys Glu Lys Thr Leu Gly Gly Asp
[1249]	1745 1750 1755
[1250]	<210> 23
[1251]	<211> 726
[1252]	<212> PRT
[1253]	<213> 伯顿拟甲烷球菌
[1254]	<400> 23
[1255]	Met Ser Asp Lys Pro Ala Phe Met Lys Tyr Phe Thr Gln Ser Ser Cys
[1256]	1 5 10 15
[1257]	Tyr Pro Asn Gln Gln Glu Ala Met Asp Arg Ile His Ser Ala Leu Met
[1258]	20 25 30
[1259]	Gln Gln Gln Leu Val Leu Phe Glu Gly Ala Cys Gly Thr Gly Lys Thr
[1260]	35 40 45
[1261]	Leu Ser Ala Leu Val Pro Ala Leu His Val Gly Lys Met Leu Gly Lys
[1262]	50 55 60
[1263]	Thr Val Ile Ile Ala Thr Asn Val His Gln Gln Met Val Gln Phe Ile
[1264]	65 70 75 80
[1265]	Asn Glu Ala Arg Asp Ile Lys Lys Val Gln Asp Val Lys Val Ala Val
[1266]	85 90 95
[1267]	Ile Lys Gly Lys Thr Ala Met Cys Pro Gln Glu Ala Asp Tyr Glu Glu
[1268]	100 105 110
[1269]	Cys Ser Val Lys Arg Glu Asn Thr Phe Glu Leu Met Glu Thr Glu Arg
[1270]	115 120 125
[1271]	Glu Ile Tyr Leu Lys Arg Gln Glu Leu Asn Ser Ala Arg Asp Ser Tyr
[1272]	130 135 140
[1273]	Lys Lys Ser His Asp Pro Ala Phe Val Thr Leu Arg Asp Glu Leu Ser
[1274]	145 150 155 160
[1275]	Lys Glu Ile Asp Ala Val Glu Glu Lys Ala Arg Gly Leu Arg Asp Arg
[1276]	165 170 175
[1277]	Ala Cys Asn Asp Leu Tyr Glu Val Leu Arg Ser Asp Ser Glu Lys Phe
[1278]	180 185 190
[1279]	Arg Glu Trp Leu Tyr Lys Glu Val Arg Ser Pro Glu Glu Ile Asn Asp
[1280]	195 200 205
[1281]	His Ala Ile Lys Asp Gly Met Cys Gly Tyr Glu Leu Val Lys Arg Glu
[1282]	210 215 220
[1283]	Leu Lys His Ala Asp Leu Leu Ile Cys Asn Tyr His His Val Leu Asn
[1284]	225 230 235 240
[1285]	Pro Asp Ile Phe Ser Thr Val Leu Gly Trp Ile Glu Lys Glu Pro Gln
[1286]	245 250 255

[1287]	Glu Thr Ile Val Ile Phe Asp Glu Ala His Asn Leu Glu Ser Ala Ala		
[1288]		260	265
[1289]	Arg Ser His Ser Ser Leu Ser Leu Thr Glu His Ser Ile Glu Lys Ala		270
[1290]		275	280
[1291]	Ile Thr Glu Leu Glu Ala Asn Leu Asp Leu Leu Ala Asp Asp Asn Ile		285
[1292]		290	295
[1293]	His Asn Leu Phe Asn Ile Phe Leu Glu Val Ile Ser Asp Thr Tyr Asn		300
[1294]	305	310	315
[1295]	Ser Arg Phe Lys Phe Gly Glu Arg Glu Arg Val Arg Lys Asn Trp Tyr		320
[1296]		325	330
[1297]	Asp Ile Arg Ile Ser Asp Pro Tyr Glu Arg Asn Asp Ile Val Arg Gly		335
[1298]		340	345
[1299]	Lys Phe Leu Arg Gln Ala Lys Gly Asp Phe Gly Glu Lys Asp Asp Ile		350
[1300]		355	360
[1301]	Gln Ile Leu Leu Ser Glu Ala Ser Glu Leu Gly Ala Lys Leu Asp Glu		365
[1302]		370	375
[1303]	Thr Tyr Arg Asp Gln Tyr Lys Lys Gly Leu Ser Ser Val Met Lys Arg		380
[1304]	385	390	395
[1305]	Ser His Ile Arg Tyr Val Ala Asp Phe Met Ser Ala Tyr Ile Glu Leu		400
[1306]		405	410
[1307]	Ser His Asn Leu Asn Tyr Tyr Pro Ile Leu Asn Val Arg Arg Asp Met		415
[1308]		420	425
[1309]	Asn Asp Glu Ile Tyr Gly Arg Val Glu Leu Phe Thr Cys Ile Pro Lys		430
[1310]		435	440
[1311]	Asn Val Thr Glu Pro Leu Phe Asn Ser Leu Phe Ser Val Ile Leu Met		445
[1312]		450	455
[1313]	Ser Ala Thr Leu His Pro Phe Glu Met Val Lys Lys Thr Leu Gly Ile		460
[1314]	465	470	475
[1315]	Thr Arg Asp Thr Cys Glu Met Ser Tyr Gly Thr Ser Phe Pro Glu Glu		480
[1316]		485	490
[1317]	Lys Arg Leu Ser Ile Ala Val Ser Ile Pro Pro Leu Phe Ala Lys Asn		495
[1318]		500	505
[1319]	Arg Asp Asp Arg His Val Thr Glu Leu Leu Glu Gln Val Leu Leu Asp		510
[1320]		515	520
[1321]	Ser Ile Glu Asn Ser Lys Gly Asn Val Ile Leu Phe Phe Gln Ser Ala		525
[1322]		530	535
[1323]	Phe Glu Ala Lys Arg Tyr Tyr Ser Lys Ile Glu Pro Leu Val Asn Val		540
[1324]	545	550	555
[1325]	Pro Val Phe Leu Asp Glu Val Gly Ile Ser Ser Gln Asp Val Arg Glu		560

[1326]		565		570		575
[1327]	Glu Phe Phe Ser Ile Gly Glu Glu Asn Gly Lys Ala Val Leu Leu Ser					
[1328]		580		585		590
[1329]	Tyr Leu Trp Gly Thr Leu Ser Glu Gly Ile Asp Tyr Arg Asp Gly Arg					
[1330]		595		600		605
[1331]	Gly Arg Thr Val Ile Ile Ile Gly Val Gly Tyr Pro Ala Leu Asn Asp					
[1332]		610		615		620
[1333]	Arg Met Asn Ala Val Glu Ser Ala Tyr Asp His Val Phe Gly Tyr Gly					
[1334]	625		630		635	640
[1335]	Ala Gly Trp Glu Phe Ala Ile Gln Val Pro Thr Ile Arg Lys Ile Arg					
[1336]		645		650		655
[1337]	Gln Ala Met Gly Arg Val Val Arg Ser Pro Thr Asp Tyr Gly Ala Arg					
[1338]		660		665		670
[1339]	Ile Leu Leu Asp Gly Arg Phe Leu Thr Asp Ser Lys Lys Arg Phe Gly					
[1340]		675		680		685
[1341]	Lys Phe Ser Val Phe Glu Val Phe Pro Pro Ala Glu Arg Ser Glu Phe					
[1342]		690		695		700
[1343]	Val Asp Val Asp Pro Glu Lys Val Lys Tyr Ser Leu Met Asn Phe Phe					
[1344]	705		710		715	720
[1345]	Met Asp Asn Asp Glu Gln					
[1346]		725				
[1347]	<210> 24					
[1348]	<211> 439					
[1349]	<212> PRT					
[1350]	<213> 肠杆菌噬菌体(Enterobacteria phage)T4					
[1351]	<400> 24					
[1352]	Met Thr Phe Asp Asp Leu Thr Glu Gly Gln Lys Asn Ala Phe Asn Ile					
[1353]	1	5		10		15
[1354]	Val Met Lys Ala Ile Lys Glu Lys Lys His His Val Thr Ile Asn Gly					
[1355]		20		25		30
[1356]	Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr Lys Phe Ile Ile Glu Ala					
[1357]		35		40		45
[1358]	Leu Ile Ser Thr Gly Glu Thr Gly Ile Ile Leu Ala Ala Pro Thr His					
[1359]		50		55		60
[1360]	Ala Ala Lys Lys Ile Leu Ser Lys Leu Ser Gly Lys Glu Ala Ser Thr					
[1361]	65		70		75	80
[1362]	Ile His Ser Ile Leu Lys Ile Asn Pro Val Thr Tyr Glu Glu Asn Val					
[1363]		85		90		95
[1364]	Leu Phe Glu Gln Lys Glu Val Pro Asp Leu Ala Lys Cys Arg Val Leu					

[1365]	100	105	110
[1366]	Ile Cys Asp Glu Val Ser Met Tyr Asp Arg Lys Leu Phe Lys Ile Leu		
[1367]	115	120	125
[1368]	Leu Ser Thr Ile Pro Pro Trp Cys Thr Ile Ile Gly Ile Gly Asp Asn		
[1369]	130	135	140
[1370]	Lys Gln Ile Arg Pro Val Asp Pro Gly Glu Asn Thr Ala Tyr Ile Ser		
[1371]	145	150	155
[1372]	Pro Phe Phe Thr His Lys Asp Phe Tyr Gln Cys Glu Leu Thr Glu Val		
[1373]	165	170	175
[1374]	Lys Arg Ser Asn Ala Pro Ile Ile Asp Val Ala Thr Asp Val Arg Asn		
[1375]	180	185	190
[1376]	Gly Lys Trp Ile Tyr Asp Lys Val Val Asp Gly His Gly Val Arg Gly		
[1377]	195	200	205
[1378]	Phe Thr Gly Asp Thr Ala Leu Arg Asp Phe Met Val Asn Tyr Phe Ser		
[1379]	210	215	220
[1380]	Ile Val Lys Ser Leu Asp Asp Leu Phe Glu Asn Arg Val Met Ala Phe		
[1381]	225	230	235
[1382]	Thr Asn Lys Ser Val Asp Lys Leu Asn Ser Ile Ile Arg Lys Lys Ile		
[1383]	245	250	255
[1384]	Phe Glu Thr Asp Lys Asp Phe Ile Val Gly Glu Ile Ile Val Met Gln		
[1385]	260	265	270
[1386]	Glu Pro Leu Phe Lys Thr Tyr Lys Ile Asp Gly Lys Pro Val Ser Glu		
[1387]	275	280	285
[1388]	Ile Ile Phe Asn Asn Gly Gln Leu Val Arg Ile Ile Glu Ala Glu Tyr		
[1389]	290	295	300
[1390]	Thr Ser Thr Phe Val Lys Ala Arg Gly Val Pro Gly Glu Tyr Leu Ile		
[1391]	305	310	315
[1392]	Arg His Trp Asp Leu Thr Val Glu Thr Tyr Gly Asp Asp Glu Tyr Tyr		
[1393]	325	330	335
[1394]	Arg Glu Lys Ile Lys Ile Ile Ser Ser Asp Glu Glu Leu Tyr Lys Phe		
[1395]	340	345	350
[1396]	Asn Leu Phe Leu Gly Lys Thr Ala Glu Thr Tyr Lys Asn Trp Asn Lys		
[1397]	355	360	365
[1398]	Gly Gly Lys Ala Pro Trp Ser Asp Phe Trp Asp Ala Lys Ser Gln Phe		
[1399]	370	375	380
[1400]	Ser Lys Val Lys Ala Leu Pro Ala Ser Thr Phe His Lys Ala Gln Gly		
[1401]	385	390	395
[1402]	Met Ser Val Asp Arg Ala Phe Ile Tyr Thr Pro Cys Ile His Tyr Ala		
[1403]	405	410	415

[1404]	Asp Val Glu Leu Ala Gln Gln Leu Leu Tyr Val Gly Val Thr Arg Gly
[1405]	420 425 430
[1406]	Arg Tyr Asp Val Phe Tyr Val
[1407]	435
[1408]	<210> 25
[1409]	<211> 970
[1410]	<212> PRT
[1411]	<213> 肉毒杆菌(Clostridium botulinum)
[1412]	<400> 25
[1413]	Met Leu Ser Val Ala Asn Val Arg Ser Pro Ser Ala Ala Ala Ser Tyr
[1414]	1 5 10 15
[1415]	Phe Ala Ser Asp Asn Tyr Tyr Ala Ser Ala Asp Ala Asp Arg Ser Gly
[1416]	20 25 30
[1417]	Gln Trp Ile Gly Asp Gly Ala Lys Arg Leu Gly Leu Glu Gly Lys Val
[1418]	35 40 45
[1419]	Glu Ala Arg Ala Phe Asp Ala Leu Leu Arg Gly Glu Leu Pro Asp Gly
[1420]	50 55 60
[1421]	Ser Ser Val Gly Asn Pro Gly Gln Ala His Arg Pro Gly Thr Asp Leu
[1422]	65 70 75 80
[1423]	Thr Phe Ser Val Pro Lys Ser Trp Ser Leu Leu Ala Leu Val Gly Lys
[1424]	85 90 95
[1425]	Asp Glu Arg Ile Ile Ala Ala Tyr Arg Glu Ala Val Val Glu Ala Leu
[1426]	100 105 110
[1427]	His Trp Ala Glu Lys Asn Ala Ala Glu Thr Arg Val Val Glu Lys Gly
[1428]	115 120 125
[1429]	Met Val Val Thr Gln Ala Thr Gly Asn Leu Ala Ile Gly Leu Phe Gln
[1430]	130 135 140
[1431]	His Asp Thr Asn Arg Asn Gln Glu Pro Asn Leu His Phe His Ala Val
[1432]	145 150 155 160
[1433]	Ile Ala Asn Val Thr Gln Gly Lys Asp Gly Lys Trp Arg Thr Leu Lys
[1434]	165 170 175
[1435]	Asn Asp Arg Leu Trp Gln Leu Asn Thr Thr Leu Asn Ser Ile Ala Met
[1436]	180 185 190
[1437]	Ala Arg Phe Arg Val Ala Val Glu Lys Leu Gly Tyr Glu Pro Gly Pro
[1438]	195 200 205
[1439]	Val Leu Lys His Gly Asn Phe Glu Ala Arg Gly Ile Ser Arg Glu Gln
[1440]	210 215 220
[1441]	Val Met Ala Phe Ser Thr Arg Arg Lys Glu Val Leu Glu Ala Arg Arg
[1442]	225 230 235 240

[1443]	Gly	Pro	Gly	Leu	Asp	Ala	Gly	Arg	Ile	Ala	Ala	Leu	Asp	Thr	Arg	Ala
[1444]					245				250						255	
[1445]	Ser	Lys	Glu	Gly	Ile	Glu	Asp	Arg	Ala	Thr	Leu	Ser	Lys	Gln	Trp	Ser
[1446]					260				265						270	
[1447]	Glu	Ala	Ala	Gln	Ser	Ile	Gly	Leu	Asp	Leu	Lys	Pro	Leu	Val	Asp	Arg
[1448]					275				280						285	
[1449]	Ala	Arg	Thr	Lys	Ala	Leu	Gly	Gln	Gly	Met	Glu	Ala	Thr	Arg	Ile	Gly
[1450]					290				295						300	
[1451]	Ser	Leu	Val	Glu	Arg	Gly	Arg	Ala	Trp	Leu	Ser	Arg	Phe	Ala	Ala	His
[1452]	305					310						315				320
[1453]	Val	Arg	Gly	Asp	Pro	Ala	Asp	Pro	Leu	Val	Pro	Pro	Ser	Val	Leu	Lys
[1454]					325					330						335
[1455]	Gln	Asp	Arg	Gln	Thr	Ile	Ala	Ala	Ala	Gln	Ala	Val	Ala	Ser	Ala	Val
[1456]					340					345						350
[1457]	Arg	His	Leu	Ser	Gln	Arg	Glu	Ala	Ala	Phe	Glu	Arg	Thr	Ala	Leu	Tyr
[1458]					355					360						365
[1459]	Lys	Ala	Ala	Leu	Asp	Phe	Gly	Leu	Pro	Thr	Thr	Ile	Ala	Asp	Val	Glu
[1460]					370					375						380
[1461]	Lys	Arg	Thr	Arg	Ala	Leu	Val	Arg	Ser	Gly	Asp	Leu	Ile	Ala	Gly	Lys
[1462]	385					390						395				400
[1463]	Gly	Glu	His	Lys	Gly	Trp	Leu	Ala	Ser	Arg	Asp	Ala	Val	Val	Thr	Glu
[1464]					405						410					415
[1465]	Gln	Arg	Ile	Leu	Ser	Glu	Val	Ala	Ala	Gly	Lys	Gly	Asp	Ser	Ser	Pro
[1466]					420						425					430
[1467]	Ala	Ile	Thr	Pro	Gln	Lys	Ala	Ala	Ala	Ser	Val	Gln	Ala	Ala	Ala	Leu
[1468]					435						440					445
[1469]	Thr	Gly	Gln	Gly	Phe	Arg	Leu	Asn	Glu	Gly	Gln	Leu	Ala	Ala	Ala	Arg
[1470]					450							455				460
[1471]	Leu	Ile	Leu	Ile	Ser	Lys	Asp	Arg	Thr	Ile	Ala	Val	Gln	Gly	Ile	Ala
[1472]	465					470						475				480
[1473]	Gly	Ala	Gly	Lys	Ser	Ser	Val	Leu	Lys	Pro	Val	Ala	Glu	Val	Leu	Arg
[1474]					485							490				495
[1475]	Asp	Glu	Gly	His	Pro	Val	Ile	Gly	Leu	Ala	Ile	Gln	Asn	Thr	Leu	Val
[1476]					500							505				510
[1477]	Gln	Met	Leu	Glu	Arg	Asp	Thr	Gly	Ile	Gly	Ser	Gln	Thr	Leu	Ala	Arg
[1478]					515							520				525
[1479]	Phe	Leu	Gly	Gly	Trp	Asn	Lys	Leu	Leu	Asp	Asp	Pro	Gly	Asn	Val	Ala
[1480]					530							535				540
[1481]	Leu	Arg	Ala	Glu	Ala	Gln	Ala	Ser	Leu	Lys	Asp	His	Val	Leu	Val	Leu

[1482]	545	550	555	560
[1483]	Asp Glu Ala Ser Met Val Ser Asn Glu Asp Lys Glu Lys Leu Val Arg			
[1484]		565	570	575
[1485]	Leu Ala Asn Leu Ala Gly Val His Arg Leu Val Leu Ile Gly Asp Arg			
[1486]		580	585	590
[1487]	Lys Gln Leu Gly Ala Val Asp Ala Gly Lys Pro Phe Ala Leu Leu Gln			
[1488]		595	600	605
[1489]	Arg Ala Gly Ile Ala Arg Ala Glu Met Ala Thr Asn Leu Arg Ala Arg			
[1490]		610	615	620
[1491]	Asp Pro Val Val Arg Glu Ala Gln Ala Ala Ala Gln Ala Gly Asp Val			
[1492]	625	630	635	640
[1493]	Arg Lys Ala Leu Arg His Leu Lys Ser His Thr Val Glu Ala Arg Gly			
[1494]		645	650	655
[1495]	Asp Gly Ala Gln Val Ala Ala Glu Thr Trp Leu Ala Leu Asp Lys Glu			
[1496]		660	665	670
[1497]	Thr Arg Ala Arg Thr Ser Ile Tyr Ala Ser Gly Arg Ala Ile Arg Ser			
[1498]		675	680	685
[1499]	Ala Val Asn Ala Ala Val Gln Gln Gly Leu Leu Ala Ser Arg Glu Ile			
[1500]		690	695	700
[1501]	Gly Pro Ala Lys Met Lys Leu Glu Val Leu Asp Arg Val Asn Thr Thr			
[1502]	705	710	715	720
[1503]	Arg Glu Glu Leu Arg His Leu Pro Ala Tyr Arg Ala Gly Arg Val Leu			
[1504]		725	730	735
[1505]	Glu Val Ser Arg Lys Gln Gln Ala Leu Gly Leu Phe Ile Gly Glu Tyr			
[1506]		740	745	750
[1507]	Arg Val Ile Gly Gln Asp Arg Lys Gly Lys Leu Val Glu Val Glu Asp			
[1508]		755	760	765
[1509]	Lys Arg Gly Lys Arg Phe Arg Phe Asp Pro Ala Arg Ile Arg Ala Gly			
[1510]		770	775	780
[1511]	Lys Gly Asp Asp Asn Leu Thr Leu Leu Glu Pro Arg Lys Leu Glu Ile			
[1512]	785	790	795	800
[1513]	His Glu Gly Asp Arg Ile Arg Trp Thr Arg Asn Asp His Arg Arg Gly			
[1514]		805	810	815
[1515]	Leu Phe Asn Ala Asp Gln Ala Arg Val Val Glu Ile Ala Asn Gly Lys			
[1516]		820	825	830
[1517]	Val Thr Phe Glu Thr Ser Lys Gly Asp Leu Val Glu Leu Lys Lys Asp			
[1518]		835	840	845
[1519]	Asp Pro Met Leu Lys Arg Ile Asp Leu Ala Tyr Ala Leu Asn Val His			
[1520]		850	855	860

[1521]	Met Ala Gln Gly Leu Thr Ser Asp Arg Gly Ile Ala Val Met Asp Ser
[1522]	865 870 875 880
[1523]	Arg Glu Arg Asn Leu Ser Asn Gln Lys Thr Phe Leu Val Thr Val Thr
[1524]	885 890 895
[1525]	Arg Leu Arg Asp His Leu Thr Leu Val Val Asp Ser Ala Asp Lys Leu
[1526]	900 905 910
[1527]	Gly Ala Ala Val Ala Arg Asn Lys Gly Glu Lys Ala Ser Ala Ile Glu
[1528]	915 920 925
[1529]	Val Thr Gly Ser Val Lys Pro Thr Ala Thr Lys Gly Ser Gly Val Asp
[1530]	930 935 940
[1531]	Gln Pro Lys Ser Val Glu Ala Asn Lys Ala Glu Lys Glu Leu Thr Arg
[1532]	945 950 955 960
[1533]	Ser Lys Ser Lys Thr Leu Asp Phe Gly Ile
[1534]	965 970
[1535]	<210> 26
[1536]	<211> 50
[1537]	<212> DNA
[1538]	<213> 人工序列
[1539]	<220>
[1540]	<223> 本发明的实施例MuA底物。
[1541]	<400> 26
[1542]	gttttcgcat ttatcgtgaa acgctttcgc gtttttcgtg cgccgcttca 50
[1543]	<210> 27
[1544]	<211> 50
[1545]	<212> DNA
[1546]	<213> 人工序列
[1547]	<220>
[1548]	<223> 本发明的实施例MuA底物。
[1549]	<400> 27
[1550]	caaaagcgta aatagcactt tgcgaaagcg caaaaagcac gcggcgaagt 50
[1551]	<210> 28
[1552]	<211> 54
[1553]	<212> DNA
[1554]	<213> 人工序列
[1555]	<220>
[1556]	<223> 本发明的实施例MuA底物。
[1557]	<400> 28
[1558]	caaaagcgta aatagcactt tgcgaaagcg caaaaagcac gcggcgaagt ctag 54
[1559]	<210> 29

[1560] <211> 100
 [1561] <212> DNA
 [1562] <213> 人工序列
 [1563] <220>
 [1564] <223> 实施例1中使用的序列。
 [1565] <400> 29
 [1566] gcgttctgtt tcgatgtat gttttcatac atccgaaaca gaacgctttt gttttcgc 60
 [1567] ttatcgtgaa acgctttcgc gtttttcgtg cgccgcttca 100
 [1568] <210> 30
 [1569] <211> 48
 [1570] <212> DNA
 [1571] <213> 人工序列
 [1572] <220>
 [1573] <223> 实施例1中使用的序列。
 [1574] <400> 30
 [1575] gaagcggcgc acgaaaaacg cgaaagcgtt tcacgataat gcgaaaac 48
 [1576] <210> 31
 [1577] <211> 48502
 [1578] <212> DNA
 [1579] <213> 人工序列
 [1580] <220>
 [1581] <223> 实施例1中使用的序列。
 [1582] <400> 31
 [1583] gggcggcgac ctgcgggtt ttcgctatth atgaaaattt tccggtttta ggcgtttccg 60
 [1584] ttcttcttcg tcataactta atgtttttat ttaaaatacc ctctgaaaag aaaggaaacg 120
 [1585] acaggtgctg aaagcgaggc tttttggcct ctgtcggttc ctttctctgt ttttgtccgt 180
 [1586] ggaatgaaca atggaagtca acaaaaagca gctggctgac attttcggtg cgagtatccg 240
 [1587] taccattcag aactggcagg aacagggaat gcccgttctg cgaggcggtg gcaagggtta 300
 [1588] tgaggtgctt tatgactctg ccgccgtcat aaaatggtat gccgaaaggg atgctgaaat 360
 [1589] tgagaacgaa aagctgcgcc gggagggtga agaactgcgg caggccagcg aggcagatct 420
 [1590] ccagccagga actattgagt acgaacgcca tcgacttacg cgtgcgcagg ccgacgcaca 480
 [1591] ggaactgaag aatgccagag actccgctga agtggtggaa accgcattct gtactttcgt 540
 [1592] gctgtcgcgg atcgcagggt aaattgccag tattctcgac gggctcccc tgctcggtgca 600
 [1593] gcggcgtttt ccggaactgg aaaaccgaca tgttgatttc ctgaaacggg atatcatcaa 660
 [1594] agccatgaac aaagcagccg cgctggatga actgataccg gggttgctga gtgaatatat 720
 [1595] cgaacagtca ggttaacagg ctgcggcatt ttgtccgcgc cgggcttcgc tcaactgttca 780
 [1596] ggccggagcc acagaccgcc gttgaatggg cggtatgctaa ttactatctc ccgaaagaat 840
 [1597] ccgcatacca ggaagggcgc tgggaaacac tgccctttca gcgggccatc atgaatgcga 900
 [1598] tgggcagcga ctacatccgt gaggtgaatg tggatgaagtc tgcccgtgtc gggttattcca 960

[1599]	aaatgctgct	gggtgtttat	gcctacttta	tagagcataa	gcagcgcaac	acccttatct	1020
[1600]	ggttgccgac	ggatggatgat	gccgagaact	ttatgaaaac	ccacgttgag	ccgactattc	1080
[1601]	gtgatattcc	gtcgtctgctg	gcgctggccc	cgtggatg	caaaaagcac	cgggataaca	1140
[1602]	cgctcaccat	gaagcgtttc	actaatgggc	gtggcttctg	gtgcctgggc	ggtaaagcgg	1200
[1603]	caaaaaacta	ccgtgaaaag	tcggtggatg	tggcgggtta	tgatgaactt	gctgcttttg	1260
[1604]	atgatgatat	tgaacaggaa	ggctctccga	cgttcctggg	tgacaagcgt	attgaaggct	1320
[1605]	cggctctggc	aaagtccatc	cgtggctcca	cgccaaaagt	gagaggcacc	tgtcagattg	1380
[1606]	agcgtgcagc	cagtgaatcc	ccgcatttta	tgcttttca	tgttgccctg	ccgcattgcg	1440
[1607]	gggaggagca	gtatcttaaa	tttggcgaca	aagagacgcc	gtttggcctc	aaatggacgc	1500
[1608]	cggatgacct	ctccagcgtg	ttttatctct	gcgagcataa	tgctgcgtc	atccgccagc	1560
[1609]	aggagctgga	ctttactgat	gcccgttata	tctgcgaaaa	gaccgggac	tggaccctg	1620
[1610]	atggcattct	ctggttttcg	tcacccggtg	aagagattga	gccacctgac	agtgtgacct	1680
[1611]	ttcacatctg	gacagcgtac	agcccggttc	ccacctgggt	gcagattgtc	aaagactgga	1740
[1612]	tgaaaacgaa	aggggatacg	ggaaaacgta	aaaccttcgt	aaacaccacg	ctcggtgaga	1800
[1613]	cgtgggagcg	gaaaattggc	gaacgtccgg	atgctgaagt	gatggcagag	cggaaagagc	1860
[1614]	attattcagc	gcccgttcct	gaccgtgtgg	cttacctgac	cgccggtatc	gactcccagc	1920
[1615]	tggaccgcta	cgaatgcgc	gtatggggat	gggggccggg	tgaggaaagc	tggctgattg	1980
[1616]	accggcagat	tattatgggc	cgccacgacg	atgaacagac	gctgctgcgt	gtggatgagg	2040
[1617]	ccatcaataa	aacctatacc	cgccggaatg	gtgcagaaat	gtcgatatcc	cgtatctgct	2100
[1618]	gggatactgg	cgggattgac	ccgaccattg	tgtatgaacg	ctcgaaaaaa	catgggctgt	2160
[1619]	tccgggtgat	ccccattaaa	ggggcatccg	tctacggaaa	gccggtggcc	agcatgccac	2220
[1620]	gtaagcgaaa	caaaaacggg	gtttacctta	ccgaaatcgg	tacggatacc	gcgaaagagc	2280
[1621]	agatttataa	ccgcttcaca	ctgacgccgg	aaggggatga	accgcttccc	ggtgccgttc	2340
[1622]	acttcccga	taaccgggat	atttttgatc	tgaccgaagc	gcagcagctg	actgctgaag	2400
[1623]	agcaggtcga	aaaatgggtg	gatggcagga	aaaaaatact	gtgggacagc	aaaaagcgac	2460
[1624]	gcaatgaggc	actcgactgc	ttcgtttatg	cgctggcggc	gctgcgcac	agtatttccc	2520
[1625]	gctggcagct	ggatctcagt	gcgctgctgg	cgagcctgca	ggaagaggat	ggtgcagcaa	2580
[1626]	ccaacaagaa	aacactggca	gattacgccc	gtgccttatc	cggagaggat	gaatgacgcg	2640
[1627]	acaggaagaa	cttgccgctg	cccgtgcggc	actgcatgac	ctgatgacag	gtaaaccggg	2700
[1628]	ggcaacagta	cagaaagacg	gacgaagggt	ggagtttacg	gccacttccg	tgtctgacct	2760
[1629]	gaaaaaatat	attgcagagc	tggaagtgca	gaccggcatg	acacagcgac	gcaggggacc	2820
[1630]	tgcaggattt	tatgtatgaa	aacgcccacc	attcccacc	ttctggggcc	ggacggcatg	2880
[1631]	acatcgctgc	gcgaatatgc	cggttatcac	ggcgggtggca	gcggatttgg	agggcagttg	2940
[1632]	cggctcgtgga	accaccgag	tgaaagtgtg	gatgcagccc	tgttgcccaa	ctttaccctg	3000
[1633]	ggcaatgccc	gcgcagacga	tctggtacgc	aataacggct	atgccgccaa	cgccatccag	3060
[1634]	ctgcatcagg	atcatatcgt	cgggtctttt	ttccggctca	gtcatcgccc	aagctggcgc	3120
[1635]	tatctgggca	tcggggagga	agaagcccgt	gccttttccc	gcgaggttga	agcggcatgg	3180
[1636]	aaagagtttg	ccgaggatga	ctgctgctgc	attgacgttg	agcgaaaacg	cacgtttacc	3240
[1637]	atgatgattc	gggaagggtg	ggccatgcac	gcctttaacg	gtgaactgtt	cgttcaggcc	3300

[1638]	acctgggata	ccagttcgtc	gcggtttttc	cggacacagt	tccgcatggt	cagcccgaag	3360
[1639]	cgcatcagca	acccgaacaa	taccggcgac	agccggaact	gccgtgccgg	tgtgcagatt	3420
[1640]	aatgacagcg	gtgcggcgct	gggatattac	gtcagcgagg	acgggtatcc	tggctggatg	3480
[1641]	ccgcagaaat	ggacatggat	accccgtgag	ttaccggcg	ggcgcgcctc	gttcattcac	3540
[1642]	gtttttgaac	ccgtggagga	cgggcagact	cgcggtgcaa	atgtgtttta	cagcgtgatg	3600
[1643]	gagcagatga	agatgctcga	cacgctgcag	aacacgcagc	tgcagagcgc	cattgtgaag	3660
[1644]	gcgatgtatg	ccgccaccat	tgagagttag	ctggatacgc	agtcagcgat	ggattttatt	3720
[1645]	ctgggcgcga	acagtcagga	gcagcgggaa	aggctgaccg	gctggattgg	tgaattgcc	3780
[1646]	gcgtattacg	ccgcagcgcc	ggtccggctg	ggaggcgcaa	aagtaccgca	cctgatgccg	3840
[1647]	ggtgactcac	tgaacctgca	gacggctcag	gatacggata	acggctactc	cgtgtttgag	3900
[1648]	cagtcactgc	tgcggtatat	cgctgccggg	ctgggtgtct	cgtatgagca	gctttcccg	3960
[1649]	aattacgccc	agatgagcta	ctccacggca	cgggccagt	cgaacgagtc	gtgggcgtac	4020
[1650]	tttatggggc	ggcgaatt	cgctgcaccc	cgtcaggcga	gccagatgtt	tctgtgctgg	4080
[1651]	ctggaagagg	ccatcgcttcg	ccgcgtggtg	acgttacctt	caaaagcgcg	cttcagtttt	4140
[1652]	caggaagccc	gcagtgcctg	ggggaactgc	gactggatag	gctccggtcg	tatggccatc	4200
[1653]	gatggtctga	aagaagttca	ggaagcggtg	atgctgatag	aagccggact	gagtacctac	4260
[1654]	gagaaagagt	gcgcaaacg	cgggtgacgac	tatcaggaaa	tttttgccca	gcaggtccgt	4320
[1655]	gaaacgatgg	agcgccgtgc	agccggtctt	aaaccgccc	cctgggcggc	tgcagcattt	4380
[1656]	gaatccgggc	tgcgacaatc	aacagaggag	gagaagagt	acagcagagc	tgcgtaatct	4440
[1657]	cccgcataat	gccagcatgg	cctttaatga	gccgctgatg	cttgaaccgc	cctatgcgcg	4500
[1658]	ggttttcttt	tgtgcgcttg	caggccagct	tgggatcagc	agcctgacgg	atgcggtgtc	4560
[1659]	cggcgacagc	ctgactgccc	aggaggcact	cgcgacgctg	gcattatccg	gtgatgatga	4620
[1660]	cggaccacga	caggcccgcga	gttatcaggt	catgaacggc	atcgccgtgc	tgccggtgtc	4680
[1661]	cggcacgctg	gtcagccgga	cgcgggcgct	gcagccgtac	tcggggatga	ccggttacia	4740
[1662]	cggcattatc	gcccgctctg	aacaggctgc	cagcgatccg	atggtggacg	gcattctgct	4800
[1663]	cgatatggac	acgcccggcg	ggatggtggc	gggggcattt	gactgcgctg	acatcatcgc	4860
[1664]	ccgtgtgcgt	gacataaaac	cggatatggc	gcttgccaac	gacatgaact	gcagtgcagg	4920
[1665]	tcagttgctt	gccagtgccg	cctcccggcg	tctggtcacg	cagaccgccc	ggacaggctc	4980
[1666]	catcggcgtc	atgatggctc	acagtaatta	cgggtgctgc	ctggagaaac	agggtgtgga	5040
[1667]	aatcacgctg	atttacagcg	gcagccataa	ggtggatggc	aaccctaca	gccatcttcc	5100
[1668]	ggatgacgtc	cgggagacac	tgcagtcccg	gatggacgca	acccgccaga	tgtttgcgca	5160
[1669]	gaaggtgtcg	gcataataccg	gcctgtccgt	gcaggttgtg	ctggataccg	aggctgcagt	5220
[1670]	gtacagcggt	caggaggcca	ttgatgccgg	actggctgat	gaacttgta	acagcaccga	5280
[1671]	tgcgatcacc	gtcatgcgtg	atgcaactgga	tgcacgtaaa	tcccgctctt	caggaggcg	5340
[1672]	aatgacaaa	gagactcaat	caacaactgt	ttcagccact	gcttcgcagg	ctgacgttac	5400
[1673]	tgacgtggtg	ccagcgacgg	aggcgagaa	cgcagcgcg	gcgcagccgg	acgtgaacgc	5460
[1674]	gcagatcacc	gcagcggttg	cggcagaaaa	cagccgcatt	atggggatcc	tcaactgtga	5520
[1675]	ggaggctcac	ggacgcgaag	aacaggcacg	cgtgctggca	gaaacccccg	gtatgaccgt	5580
[1676]	gaaaacggcc	cgccgcattc	tggccgcagc	accacagagt	gcacaggcgc	gcagtgcac	5640

[1677]	tgcgctggat	cgtctgatgc	agggggcacc	ggcaccgctg	gctgcaggta	acccggcatc	5700
[1678]	tgatgccgtt	aacgatttgc	tgaacacacc	agtgttaagg	atgtttatga	cgagcaaaga	5760
[1679]	aacctttacc	cattaccagc	cgcagggcaa	cagtgaccgc	gctcataccg	caaccgcgcc	5820
[1680]	cggcggattg	agtgcgaaag	cgcctgcaat	gaccccgctg	atgctggaca	cctccagccg	5880
[1681]	taagctgggt	gcgtgggatg	gcaccaccga	cgggtgctgc	gttggcattc	ttgcggttgc	5940
[1682]	tgctgaccag	accagcacca	cgctgacgtt	ctacaagtcc	ggcacgttcc	gttatgagga	6000
[1683]	tgtgctctgg	ccggaggctg	ccagcgacga	gacgaaaaaa	cggaccgcgt	ttgccggaac	6060
[1684]	ggcaatcagc	atcgtttaac	tttacccttc	atcactaaag	gccgcctgtg	cggctttttt	6120
[1685]	tacgggattt	ttttatgtcg	atgtacacaa	ccgcccaact	gctggcggca	aatgagcaga	6180
[1686]	aattttaagtt	tgatccgctg	tttctgcgtc	tctttttccg	tgagagctat	cccttcacca	6240
[1687]	cggagaaagt	ctatctctca	caaattccgg	gactggtaaa	catggcgctg	tacgttttcg	6300
[1688]	cgattgtttc	cggtgaggtt	atccgttccc	gtggcggtc	cacctctgaa	tttacgcccg	6360
[1689]	gatattgtcaa	gccgaagcat	gaagtgaatc	cgcagatgac	cctgcgtcgc	ctgccggatg	6420
[1690]	aagatccgca	gaatctggcg	gacccggctt	accgccgccg	tcgcatcatc	atgcagaaca	6480
[1691]	tgcgtgacga	agagctggcc	attgctcagg	tcgaagagat	gcaggcagtt	tctgccgtgc	6540
[1692]	ttaagggcaa	atacaccatg	accggtgaag	ccttcgatcc	ggttgaggtg	gatatgggcc	6600
[1693]	gcagtgagga	gaataacatc	acgcagtccg	gcggcacgga	gtggagcaag	cgtgacaagt	6660
[1694]	ccacgtatga	cccgaccgac	gatatcgaag	cctacgcgct	gaacgccagc	ggtgtggtga	6720
[1695]	atatcatcgt	gttcgatccg	aaaggctggg	cgctgttccg	ttccttcaaa	gccgtcaagg	6780
[1696]	agaagctgga	taccgcgtgt	ggctctaatt	ccgagctgga	gacagcggtg	aaagacctgg	6840
[1697]	gcaaagcgg	gtcctataag	gggatgtatg	gcgatgtggc	catcgtcgtg	tattccggac	6900
[1698]	agtacgtgga	aaacggcgct	aaaaagaact	tcctgccgga	caacacgatg	gtgctgggga	6960
[1699]	acactcaggc	acgcggtctg	cgcacctatg	gctgcattca	ggatgcggac	gcacagcgcg	7020
[1700]	aaggcattaa	cgcctctgcc	cgttaccgca	aaaactgggt	gaccaccggc	gatccggcgc	7080
[1701]	gtgagttcac	catgattcag	tcagcaccgc	tgatgctgct	ggctgaccct	gatgagttcg	7140
[1702]	tgtccgtaca	actggcgtaa	tcattggcct	tcggggccat	tgtttctctg	tggaggagtc	7200
[1703]	catgacgaaa	gatgaactga	ttgccgtct	ccgctcgtg	ggtgaacaac	tgaaccgtga	7260
[1704]	tgtagcctg	acggggacga	aagaagaact	ggcgtccgt	gtggcagagc	tgaagagga	7320
[1705]	gcttgatgac	acggatgaaa	ctgccggtca	ggacaccct	ctcagccggg	aaaatgtgct	7380
[1706]	gaccggacat	gaaaatgagg	tgggatcagc	gcagccggat	accgtgattc	tggatacgtc	7440
[1707]	tgaactggtc	acggctcgtg	cactggtgaa	gctgcatact	gatgcacttc	acgccacgcg	7500
[1708]	ggatgaacct	gtggcatttg	tgctgccggg	aacggcgttt	cgtgtctctg	ccggtgtggc	7560
[1709]	agccgaaatg	acagagcgcg	gcctggccag	aatgcaataa	cgggagcgcg	tgtggctgat	7620
[1710]	ttcgataacc	tgttcgatgc	tgccattgcc	cgcgccgatg	aaacgatacg	cgggtacatg	7680
[1711]	ggaacgtcag	ccaccattac	atccggtgag	cagtcagggt	cggtgatacg	tgggtgtttt	7740
[1712]	gatgaccctg	aaaatatcag	ctatgccgga	cagggcgctg	gcgttgaaag	ctccagcccg	7800
[1713]	tccctgtttg	tccggactga	tgaggtgcgg	cagctgcggc	gtggagacac	gctgaccatc	7860
[1714]	ggtgagggaa	atttctgggt	agatcgggtt	tcgccggatg	atggcggaag	ttgtcatctc	7920
[1715]	tggcttgac	ggggcggtacc	gcctgccgtt	aaccgtcgcc	gctgaaaggg	ggatgtatgg	7980

[1716]	ccataaaagg	tcttgagcag	gccgttgaaa	acctcagccg	tatcagcaaa	acggcgggtgc	8040
[1717]	ctggtgccgc	cgcaatggcc	attaaccgcg	ttgcttcac	cgcgatatcg	cagtcggcgt	8100
[1718]	cacaggttgc	ccgtgagaca	aaggtacgcc	ggaaactggt	aaaggaaagg	gccaggctga	8160
[1719]	aaagggccac	ggtcaaaaat	ccgcaggcca	gaatcaaagt	taaccggggg	gatttgcccc	8220
[1720]	taatcaagct	gggtaatgcg	cgggttggtc	tttcgcgccg	caggcgctcg	aaaaaggggc	8280
[1721]	agcggttcac	cctgaaaggt	ggcggcagcg	tgcttggtgt	gggtaaccgt	cgtattcccc	8340
[1722]	gcgcgtttat	tcagcaactg	aaaaatggcc	ggtggcatgt	catgcagcgt	gtggctggga	8400
[1723]	aaaaccgtta	ccccattgat	gtggtgaaaa	tcccgatggc	ggtgccgctg	accacggcgt	8460
[1724]	ttaaacaaaa	tattgagcgg	atacggcgtg	aacgtcttcc	gaaagagctg	ggctatgcgc	8520
[1725]	tgcagcatca	actgaggatg	gtaataaagc	gatgaaacat	actgaactcc	gtgcagccgt	8580
[1726]	actggatgca	ctggagaagc	atgacaccgg	ggcgacgttt	tttgatggtc	gccccgctgt	8640
[1727]	ttttgatgag	gcggattttc	cggcagttgc	cgtttatctc	accggcgtg	aatacacggg	8700
[1728]	cgaagagctg	gacagcgata	cctggcaggc	ggagctgcat	atcgaagttt	tcttgccctgc	8760
[1729]	tcaggtgccg	gattcagagc	tggatgcgtg	gatggagtcc	cggattttac	cggatgatgag	8820
[1730]	cgatatcccc	gcaactgtcag	atttgatcac	cagtatgggt	gccagcggct	atgactaccg	8880
[1731]	gcgcgacgat	gatgcgggct	tgtggagttc	agccgatctg	acttatgtca	ttacctatga	8940
[1732]	aatgtgagga	cgctatgcct	gtaccaaata	ctacaatgcc	ggtgaaaggt	gccggggacca	9000
[1733]	ccctgtgggt	ttataagggg	agcggtgacc	cttacgcgaa	tccgctttca	gacgttgact	9060
[1734]	ggtcgcgtct	ggcaaaagtt	aaagacctga	cgccccgcga	actgaccgct	gagtcctatg	9120
[1735]	acgacagcta	tctcgatgat	gaagatgcag	actggactgc	gaccgggcag	gggcagaaat	9180
[1736]	ctgccggaga	taccagcttc	acgctggcgt	ggatgcccgg	agagcagggg	cagcaggcgc	9240
[1737]	tgctggcgtg	gtttaatgaa	ggcgataccc	gtgcctataa	aatccgcttc	ccgaacggca	9300
[1738]	cggtcgatgt	gttccgtggc	tgggtcagca	gtatcggtaa	ggcggtgacg	gcgaaggaag	9360
[1739]	tgatcacccg	cacggtgaaa	gtcaccaatg	tgggacgtcc	gtcgatggca	gaagatcgca	9420
[1740]	gcacggtaac	agcggcaacc	ggcatgaccg	tgacgcctgc	cagcacctcg	gtggtgaaag	9480
[1741]	ggcagagcac	cacgctgacc	gtggccttcc	agccggaggg	cgtaaccgac	aagagctttc	9540
[1742]	gtgcgggtgc	tgcggataaa	acaaaagcca	ccgtgtcggt	cagtggatatg	accatcacccg	9600
[1743]	tgaacggcgt	tgctgcaggc	aaggtcaaca	ttccggttgt	atccggtaat	ggtgagtttg	9660
[1744]	ctgcggttgc	agaaattacc	gtcacgcgca	gttaatccgg	agagtcagcg	atgttcctga	9720
[1745]	aaaccgaatc	atttgaacat	aacggtgtga	ccgtcacgct	ttctgaactg	tcagccctgc	9780
[1746]	agcgcattga	gcatctcgcc	ctgatgaaac	ggcaggcaga	acaggcggag	tcagacagca	9840
[1747]	accggaagtt	tactgtggaa	gacgccatca	gaaccggcgc	gtttctgggtg	gcgatgtccc	9900
[1748]	tgtggcataa	ccatccgcag	aagacgcaga	tgccgtccat	gaatgaagcc	gttaaacaga	9960
[1749]	ttgagcagga	agtgcctacc	acctggccca	cggaggcaat	ttctcatgct	gaaaacgtgg	10020
[1750]	tgtaccggct	gtctggatatg	tatgagtttg	tggtgaataa	tgcccctgaa	cagacagagg	10080
[1751]	acgccgggcc	cgagagcct	gtttctgcgg	gaaagtgttc	gacggtgagc	tgagttttgc	10140
[1752]	cctgaaactg	gcgcgtgaga	tggggcgacc	cgactggcgt	gccatgcttg	ccgggatgtc	10200
[1753]	atccacggag	tatgccgact	ggcaccgctt	ttacagtacc	cattattttc	atgatgttct	10260
[1754]	gctggatatg	cacttttccg	ggctgacgta	caccgtgctc	agcctgtttt	tcagcgatcc	10320

[1755]	ggatatgcat	ccgctggatt	tcagtctgct	gaaccggcgc	gaggctgacg	aagagcctga	10380
[1756]	agatgatgtg	ctgatgcaga	aagcggcagg	gcttgccgga	ggtgtccgct	ttggcccgga	10440
[1757]	cgggaatgaa	gttatccccg	cttccccgga	tgtggcggac	atgacggagg	atgacgtaat	10500
[1758]	gctgatgaca	gtatcagaag	ggatcgcagg	aggagtccgg	tatggctgaa	ccggtaggcg	10560
[1759]	atctggtcgt	tgatttgagt	ctggatgcgg	ccagatttga	cgagcagatg	gccagagtca	10620
[1760]	ggcgtcattt	ttctggtacg	gaaagtgatg	cgaaaaaac	agcggcagtc	gttgaacagt	10680
[1761]	cgctgagccg	acaggcgctg	gctgcacaga	aagcggggat	ttccgtcggg	cagtataaag	10740
[1762]	ccgccatgcg	tatgctgcct	gcacagttca	ccgacgtggc	cacgcagctt	gcaggcgggc	10800
[1763]	aaagtccgtg	gctgatcctg	ctgcaacagg	gggggcagg	gaaggactcc	ttcggcggga	10860
[1764]	tgatccccat	gttcaggggg	cttgccgggtg	cgatcacctt	gccgatgggtg	ggggccacct	10920
[1765]	cgctggcggt	ggcgaccggt	gcgctggcgt	atgcctggta	tcagggaac	tcaacctgt	10980
[1766]	ccgatttcaa	caaaacgctg	gtcctttccg	gcaatcaggc	gggactgacg	gcagatcgta	11040
[1767]	tgctggctct	gtccagagcc	gggcaggcgg	cagggtgac	gtttaaccag	accagcgagt	11100
[1768]	cactcagcgc	actggttaag	gcgggggtaa	gcggtgaggc	tcagattgcg	tccatcagcc	11160
[1769]	agagtgtggc	gcgtttctcc	tctgcatccg	gcgtggagg	ggacaaggtc	gctgaagcct	11220
[1770]	tcgggaagct	gaccacagac	ccgacgtcgg	ggctgacggc	gatggctcgc	cagttccata	11280
[1771]	acgtgtcggc	ggagcagatt	gcgtatgttg	ctcagttgca	gcgttccggc	gatgaagccg	11340
[1772]	gggcattgca	ggcggcgaac	gaggccgcaa	cgaaagggtt	tgatgaccag	acccgccgcc	11400
[1773]	tgaaagagaa	catgggcacg	ctggagacct	gggcagacag	gactgcgcgg	gcattcaa	11460
[1774]	ccatgtggga	tgcggtgctg	gatattggtc	gtcctgatac	cgcgccaggag	atgctgatta	11520
[1775]	aggcagaggc	tcgtataaag	aaagcagacg	acatctggaa	tctgcgcaag	gatgattatt	11580
[1776]	ttgttaacga	tgaagcgcgg	gcgcgttact	gggatgatcg	tgaaaaggcc	cgtcttgccg	11640
[1777]	ttgaagccgc	ccgaaagaag	gctgagcagc	agactcaaca	ggacaaaaat	gcgcagcagc	11700
[1778]	agagcgatac	cgaagcgta	cggctgaaat	ataccgaaga	ggcgccagaag	gcttacgaac	11760
[1779]	ggctgcagac	gccgctggag	aaatataccg	cccgtcagga	agaactgaac	aaggcactga	11820
[1780]	aagacgggaa	aatcctgcag	gcggattaca	acacgctgat	ggcggcggcg	aaaaaggatt	11880
[1781]	atgaagcgac	gctgaaaaag	ccgaaacagt	ccagcgtgaa	ggtgtctgcg	ggcgatcgtc	11940
[1782]	aggaagacag	tgctcatgct	gccctgctga	cgcttcaggc	agaactccgg	acgctggaga	12000
[1783]	agcatgccgg	agcaaatgag	aaaatcagcc	agcagcggcg	ggatttgtgg	aaggcggaga	12060
[1784]	gtcagttcgc	ggtactggag	gaggcggcgc	aacgtcgcca	gctgtctgca	caggagaaat	12120
[1785]	ccctgctggc	gcataaagat	gagacgctgg	agtacaaacg	ccagctggct	gcacttggcg	12180
[1786]	acaaggttac	gtatcaggag	gcctgaacg	cgctggcgca	gcaggcgat	aaattcgac	12240
[1787]	agcagcaacg	ggcaaaacgg	gccgccattg	atgcgaaaag	ccgggggctg	actgaccggc	12300
[1788]	aggcagaacg	ggaagccacg	gaacagcgcc	tgaaggaaca	gtatggcgat	aatccgctgg	12360
[1789]	cgctgaataa	cgtcatgtca	gagcagaaaa	agacctgggc	ggctgaagac	cagcttcgcg	12420
[1790]	ggaactggat	ggcaggcctg	aagtccggct	ggagtgaagt	ggaagagagc	gccacggaca	12480
[1791]	gtatgtcgca	ggtaaaaagt	gcagccacgc	agaccttga	tggtattgca	cagaatatgg	12540
[1792]	cggcgatgct	gaccggcagt	gagcagaact	ggcgagctt	caccggttcc	gtgctgtcca	12600
[1793]	tgatgacaga	aattctgctt	aagcaggcaa	tggtggggat	gtcgggag	atcggcagcg	12660

[1794]	ccattggcgg	ggctgttgg	ggcggcgcat	ccgcgtcagg	cggtacagcc	attcaggccg	12720
[1795]	ctgcggcgaa	attccatttt	gcaaccggag	gatttacggg	aaccggcggc	aaatatgagc	12780
[1796]	cagcggggat	tgttcacctg	ggtgagtttg	tcttcacgaa	ggaggcaacc	agccggattg	12840
[1797]	gcgtggggaa	tctttaccgg	ctgatgcgcg	gctatgccac	cggcggttat	gtcgggtacac	12900
[1798]	cgggcagcat	ggcagacagc	cggtcgcagg	cgtccgggac	gtttgagcag	aataaccatg	12960
[1799]	tggtgattaa	caacgacggc	acgaacgggc	agataggtcc	ggctgctctg	aaggcgggtg	13020
[1800]	atgacatggc	ccgcaagggt	gcccgtgatg	aaattcagac	acagatgcgt	gatggtggcc	13080
[1801]	tgttctccgg	aggtggacga	tgaagacctt	ccgctggaaa	gtgaaacccg	gtatggatgt	13140
[1802]	ggcttcggtc	ccttctgtaa	gaaaggtgcg	ctttgggtgat	ggctattctc	agcgagcgcc	13200
[1803]	tgccgggctg	aatgccaacc	tgaaaacgta	cagcgtgacg	ctttctgtcc	cccgtgagga	13260
[1804]	ggccacggta	ctggagtcgt	ttctggaaga	gcacgggggc	tggaaatcct	ttctgtggac	13320
[1805]	gccgccttat	gagtggcggc	agataaaggt	gacctgcgca	aaatggctgt	cgcgggtcag	13380
[1806]	tatgctgcgt	gttgagttca	gcgcagagtt	tgaacaggtg	gtgaactgat	gcaggatata	13440
[1807]	cggcagga	caactgaatga	atgcacccgt	gcggagcagt	cggccagcgt	ggtgctcttg	13500
[1808]	gaaatcgacc	tgacagaggt	cgggtggagaa	cgttattttt	tctgtaatga	gcagaacgaa	13560
[1809]	aaaggtgagc	cggtcacctg	gcaggggcca	cagtatcagc	cgtatcccat	tcaggggagc	13620
[1810]	ggttttgaac	tgaatggcaa	aggcaccagt	acgcgcccc	cgtgacgggt	ttctaacctg	13680
[1811]	tacggtatgg	tcaccgggat	ggcgggaagat	atgcagagtc	tggtcggcgg	aacggtggtc	13740
[1812]	cggcgtaagg	tttacgccc	ttttctggat	gcggtgaact	tcgtcaacgg	aaacagttac	13800
[1813]	gccgatccgg	agcaggaggt	gatcagccgc	tggcgcattg	agcagtcag	cgaactgagc	13860
[1814]	gcggtgagtg	cctcctttgt	actgtccacg	ccgacggaaa	cggatggcgc	tgtttttccg	13920
[1815]	ggacgtatca	tgctggccaa	cacctgcacc	tggacctatc	gcggtgacga	gtgcggttat	13980
[1816]	agcggctccg	ctgtcgcgga	tgaatatgac	cagccaacgt	ccgatatac	gaaggataaa	14040
[1817]	tgcagcaaat	gcctgagcgg	ttgtaagtgc	cgcaataacg	tcggcaactt	tggcggcttc	14100
[1818]	ctttccatta	acaaactttc	gcagtaaata	ccatgacaca	gacagaatca	gcgattcttg	14160
[1819]	cgcacgccc	gcgatgtgcg	ccagcggagt	cgtgcggctt	cgtggtaagc	acgccggagg	14220
[1820]	gggaaagata	tttcccctgc	gtgaatatct	ccggtgagcc	ggaggctatt	tccgtatgtc	14280
[1821]	gccggaagac	tggtcgcagg	cagaaatgca	gggtgagatt	gtggcgtg	tccacagcca	14340
[1822]	ccccgggtgt	ctgccctggc	tgagtgaggg	cgaccggcgg	ctgcaggtgc	agagtgattt	14400
[1823]	gccgtggtgg	ctggtctgcc	gggggacgat	tcataagtgc	cgtgtgtg	cgcattctac	14460
[1824]	cgggcggcgc	tttgagcacg	gtgtgacgga	ctgttacaca	ctgttccggg	atgcttatca	14520
[1825]	tctggcgggg	attgagatgc	cggactttca	tcgtgaggat	gactgggtgg	gtaacggcca	14580
[1826]	gaatctctat	ctggataatc	tggaggcgac	ggggctgtat	cagggtccgt	tgtcagcggc	14640
[1827]	acagccgggc	gatgtgtg	tgtgtgtgtt	tggttcatca	gtgccgaatc	acgccgcaat	14700
[1828]	ttactgcggc	gacggcgagc	tgctgcacca	tattcctgaa	caactgagca	aacgagagag	14760
[1829]	gtacaccgac	aaatggcagc	gacgcacaca	ctccctctgg	cgtcaccggg	catggcgcg	14820
[1830]	atctgccttt	acggggattt	acaacgattt	ggtcgccgca	tcgaccttcg	tgtgaaaacg	14880
[1831]	ggggctgaag	ccatccgggc	actggccaca	cagctcccgg	cgtttcgtca	gaaactgagc	14940
[1832]	gacggctggt	atcagggtacg	gattgccggg	cgggacgtca	gcacgtccgg	gttaacggcg	15000

[1833]	cagttacatg agactctgcc tgatggcgct gtaattcata ttgttcccag agtcgccggg	15060
[1834]	gccaaagtcag gtggcggtatt ccagattgtc ctgggggctg ccgccattgc cggatcattc	15120
[1835]	tttaccgccg gagccaccct tgcagcatgg ggggcagcca ttggggccgg tggatatgacc	15180
[1836]	ggcatcctgt tttctctcgg tgccagtatg gtgctcgggt gtgtggcgca gatgctggca	15240
[1837]	ccgaaagcca gaactccccg tatacagaca acggataacg gtaagcagaa cacctatttc	15300
[1838]	tcctcactgg ataacatggg tggccagggc aatgtttctgc ctgttctgta cggggaaatg	15360
[1839]	cgcgtgggggt cacgcgtggg ttctcaggag atcagcacgg cagacgaagg ggacggtggg	15420
[1840]	caggttgtgg tgattggtcg ctgatgcaaa atgttttatg tgaaaccgcc tgcgggcggg	15480
[1841]	tttgtcattt atggagcgtg aggaatgggt aaaggaagca gtaaggggca taccgccgcg	15540
[1842]	gaagcgaagg acaacctgaa gtccacgcag ttgctgagtg tgatcgatgc catcagcgaa	15600
[1843]	gggccgattg aaggtccggg ggatggctta aaaagcgtgc tgctgaacag tacgccggtg	15660
[1844]	ctggacactg aggggaatac caacatatcc ggtgtcacgg tgggtttccg ggctggtgag	15720
[1845]	caggagcaga ctccgccgga gggatttgaa tcctccggct ccgagacggg gctgggtacg	15780
[1846]	gaagtgaat atgacacgcc gatcacccgc accattacgt ctgcaaacat cgaccgtctg	15840
[1847]	cgttttacct tcggtgtaca ggcactggtg gaaaccacct caaaggtga caggaatccg	15900
[1848]	tcggaagtcc gcctgctggg tcagatacaa cgtaacgggt gctgggtgac ggaaaaagac	15960
[1849]	atcaccatta agggcaaaac cacctcgcag tatctggcct cgggtggtgat gggtaacctg	16020
[1850]	ccgccgcgcc cgtttaatat ccgatgcgc aggatgacgc cggacagcac cacagaccag	16080
[1851]	ctgcagaaca aaacgctctg gtcgtcatac actgaaatca tcgatgtgaa acagtgttac	16140
[1852]	ccgaacacgg cactggtcgg cgtgcagggt gactcggagc agttcggcag ccagcagggtg	16200
[1853]	agccgtaatt atcatctcgc cgggcgtatt ctgcagggtc cgctcgaacta taaccgcag	16260
[1854]	acgcggcaat acagcgggtat ctgggacgga acgtttaaac cggcatacag caacaacatg	16320
[1855]	gcctggtgtc tgtgggatat gctgacccat ccgcgctacg gcatggggaa acgtcttggg	16380
[1856]	gcggcggatg tggataaatg ggcgctgtat gtcacggcc agtactgcga ccagtcagtg	16440
[1857]	ccggacggct ttggcggcac ggagccgcgc atcacctgta atgcgtacct gaccacacag	16500
[1858]	cgtaaggcgt gggatgtgct cagcgatttc tgctcggcga tgcgctgtat gccggtatgg	16560
[1859]	aacgggcaga cgctgacgtt cgtgcaggac cgaccgtcgg ataagacgtg gacctataac	16620
[1860]	cgcagtaatg tggatgatgcc ggatgatggc gcgccgttcc gctacagctt cagcgccctg	16680
[1861]	aaggaccgcc ataatgccgt tgaggtgaac tggattgacc cgaacaacgg ctgggagacg	16740
[1862]	gcgacagagc ttgttgaaaga tacgcaggcc attgccggtt acggctgtaa tgttacgaag	16800
[1863]	atggatgcct ttggctgtac cagccggggg caggcacacc gcgccgggct gtggctgatt	16860
[1864]	aaaacagaac tgctggaac gcagaccgtg gatttcagcg tcggcgcaga agggcttcgc	16920
[1865]	catgtaccgg gcgatgttat tgaaatctgc gatgatgact atgccggtat cagcaccggg	16980
[1866]	ggctgtgtgc tggcggtgaa cagccagacc cggacgtga cgctcgaccg tgaaatcacg	17040
[1867]	ctgccatcct ccggtaccgc gctgataagc ctggttgacg gaagtggcaa tccggtcagc	17100
[1868]	gtggaggttc agtccgtcac cgacggcgtg aaggtaaaag tgagccgtgt tcctgacggg	17160
[1869]	gttctgtaat acagcgtatg ggagctgaag ctgccgacgc tgcgccagcg actgttcgcg	17220
[1870]	tgcgtgagta tccgtgagaa cgacgacggc acgtatgcca tcaccgccgt gcagcatgtg	17280
[1871]	ccggaaaaag aggccatcgt ggataacggg gcgcactttg acggcgaaca gagtggcacg	17340

[1872]	gtgaatggtg	tcacgccgcc	agcgggtgcag	cacctgaccg	cagaagtcac	tgacagacagc	17400
[1873]	ggggaatatc	aggtgctggc	gcgatgggac	acaccgaagg	tggtgaaggg	cgtgagtttc	17460
[1874]	ctgctccgtc	tgaccgtaac	agcggacgac	ggcagtgagc	ggctggtcag	cacggcccgg	17520
[1875]	acgacggaaa	ccacataccg	cttcacgcaa	ctggcgctgg	ggaactacag	gctgacagtc	17580
[1876]	cgggcggtaa	atgcgtgggg	gcagcagggc	gatccggcgt	cggtatcggt	ccggattgcc	17640
[1877]	gcaccggcag	caccgtcgag	gattgagctg	acgccgggct	attttcagat	aaccgccacg	17700
[1878]	ccgcattctg	ccgtttatga	cccgcaggta	cagtttgagt	tctggttctc	ggaaaagcag	17760
[1879]	attgcggata	tcagacaggt	tgaaccagc	acgcgttatc	ttggtacggc	gctgtactgg	17820
[1880]	atagccgcca	gtatcaatat	caaaccgggc	catgattatt	acttttatat	ccgcagtgtg	17880
[1881]	aacaccgttg	gcaaatcggc	attcgtggag	gccgtcggtc	gggcgagcga	tgatgcggaa	17940
[1882]	ggttacctgg	attttttcaa	aggcaagata	accgaatccc	atctcggcaa	ggagctgctg	18000
[1883]	gaaaaagtcg	agctgacgga	ggataacgcc	agcagactgg	aggagttttc	gaaagagtgg	18060
[1884]	aaggatgcca	gtgataagt	gaatgccatg	tggtgtgtca	aaattgagca	gaccaaagac	18120
[1885]	ggcaaacatt	atgtcgcggg	tattggcctc	agcatggagg	acacggagga	aggcaaactg	18180
[1886]	agccagtttc	tggttgccgc	caatcgtatc	gcattttattg	acccggcaaa	cgggaatgaa	18240
[1887]	acgccgatgt	ttgtggcgca	gggcaaccag	atattcatga	acgacgtgtt	cctgaagcgc	18300
[1888]	ctgacggccc	ccaccattac	cagcggcggc	aatcctccgg	ccttttccct	gacaccggac	18360
[1889]	ggaaagctga	ccgctaaaaa	tgcggatata	agtggcagt	tgaatgcgaa	ctccgggacg	18420
[1890]	ctcagtaatg	tgacgatagc	tgaaaactgt	acgataaacg	gtacgtcgag	ggcggaaaaa	18480
[1891]	atcgtcgggg	acattgtaaa	ggcggcgagc	gcggcttttc	cgccgacgag	tgaaagcagt	18540
[1892]	gtggactggc	cgctcaggtac	ccgtactgtc	accgtgaccg	atgaccatcc	ttttgatcgc	18600
[1893]	cagatagtgg	tgcttccgct	gacgtttcgc	ggaagtaagc	gtactgtcag	cggcaggaca	18660
[1894]	acgtattcga	tgtgttatct	gaaagtactg	atgaacgggt	cggtgattta	tgatggcgcg	18720
[1895]	gcgaacgagg	cggtacaggt	gttctcccgt	attgttgaca	tgccagcggg	tcggggaaac	18780
[1896]	gtgatectga	cgttcacgct	tacgtccaca	cggcattcgg	cagatattcc	gccgtatacg	18840
[1897]	tttgccagcg	atgtgcaggt	tatggtgatt	aagaaacagg	cgctgggcat	cagcgtggtc	18900
[1898]	tgagtgtgtt	acagaggttc	gtccgggaac	gggcgtttta	ttataaaaaca	gtgagaggtg	18960
[1899]	aacgatgcgt	aatgtgtgta	ttgccgttgc	tgtctttgcc	gcacttgccg	tgacagtcac	19020
[1900]	tccggcccgt	gcggaaggtg	gacatggtac	gtttacgggt	ggctattttc	aagtgaacc	19080
[1901]	gggtacattg	ccgtcgttgt	cgggcgggga	taccgggtgtg	agtcactctga	aagggattaa	19140
[1902]	cgtgaagtac	cgttatgagc	tgacggacag	tgtgggggtg	atggcttccc	tggggttcgc	19200
[1903]	cgcgtcga	aaagacgagc	cagtgatgac	cggggaggat	acgtttcact	atgagagcct	19260
[1904]	gcgtggacgt	tatgtgagcg	tgatggccgg	accggtttta	caaatcagta	agcaggtcag	19320
[1905]	tgcttacgcc	atggccggag	tggttcacag	tcggtggtcc	ggcagtacaa	tggattaccg	19380
[1906]	taagacggaa	atcaactccc	ggtatatgaa	agagacgacc	actgccaggg	acgaaagtgc	19440
[1907]	aatgcggcat	acctcagtg	cgtggagtgc	aggtatacag	attaatccgg	cagcgtccgt	19500
[1908]	cgttggtgat	attgcttatg	aaggctccgg	cagtggcgac	tggcgtactg	acggattcat	19560
[1909]	cgttggggtc	ggttataaat	tctgattagc	caggtaacac	agtgttatga	cagcccgcg	19620
[1910]	gaaccggtgg	gcttttttgt	ggggtgaata	tggcagtaaa	gatttcagga	gtcctgaaag	19680

[1911]	acggcacagg	aaaaccggta	cagaactgca	ccattcagct	gaaagccaga	cgtaacagca	19740
[1912]	ccacggtggt	ggtgaacacg	gtgggctcag	agaatccgga	tgaagccggg	cgttacagca	19800
[1913]	tggatgtgga	gtacggtcag	tacagtgtca	tcctgcaggt	tgacggtttt	ccaccatcgc	19860
[1914]	acgccgggac	catcaccggtg	tatgaagatt	cacaaccggg	gacgctgaat	gattttctct	19920
[1915]	gtgccatgac	ggaggatgat	gcccggccgg	aggtgctgcg	tcgtcttgaa	ctgatggtgg	19980
[1916]	aagagggtggc	gcgtaacgcg	tccgtggtgg	cacagagtac	ggcagacgcg	aagaaatcag	20040
[1917]	ccggcgatgc	cagtgcataca	gctgctcagg	tcgcgccct	tgtgactgat	gcaactgact	20100
[1918]	cagcacgcgc	cgccagcacg	tccgccggac	aggtgcatac	gtcagctcag	gaagcgctct	20160
[1919]	ccggcgacga	agcggcatca	gcaaaggcca	ctgaagcgga	aaaaagtgcc	gcagccgcag	20220
[1920]	agtctcaaaa	aaacgcggcg	gccaccagtg	ccggtgcggc	gaaaacgtca	gaaacgaatg	20280
[1921]	ctgcagcgtc	acaacaatca	gccgccacgt	ctgctccac	cgcggccacg	aaagcgtcag	20340
[1922]	aggccgccac	ttcagcacga	gatgcggtgg	cctcaaaaaga	ggcagcaaaa	tcatcagaaa	20400
[1923]	cgaacgcata	atcaagtgcc	ggtcgtgcag	cttctcggc	aacggcgga	gaaaattctg	20460
[1924]	ccagggcggc	aaaaacgtcc	gagacgaatg	ccaggtcatc	tgaacagca	gcggaacgga	20520
[1925]	gcgcctctgc	cgcggcagac	gcaaaaacag	cggcggggg	gagtgcgtca	acggcatcca	20580
[1926]	cgaaggcgac	agaggctgcg	ggaagtgcgg	tatcagcatc	gcagagcaaa	agtgcggcag	20640
[1927]	aagcggcggc	aatacgtgca	aaaaattcgg	caaacgtgc	agaagatata	gcttcagctg	20700
[1928]	tcgcgcttga	ggatgcggac	acaacgagaa	aggggatagt	gcagctcagc	agtgcaacca	20760
[1929]	acagcacgtc	tgaacgctt	gctgcaacgc	caaaggcggt	taagggtgta	atggatgaaa	20820
[1930]	cgaacagaaa	agcccactgg	acagtccggc	actgaccgga	acgccaacag	caccaaccgc	20880
[1931]	gctcagggga	acaacaata	cccagattgc	gaacaccgct	tttgtactgg	ccgcgattgc	20940
[1932]	agatgttatc	gacgcgtcac	ctgacgcact	gaatacgtg	aatgaactgg	ccgcagcgct	21000
[1933]	cgggaatgat	ccagattttg	ctaccacat	gactaacgcg	cttgccggta	aacaaccgaa	21060
[1934]	gaatgcgaca	ctgacggcgc	tggcagggct	ttccacggcg	aaaaataaat	taccgtattt	21120
[1935]	tgcggaaaat	gatgccgcca	gcctgactga	actgactcag	gttggcaggg	atattctggc	21180
[1936]	aaaaaattcc	gttgacagatg	ttcttgaata	ccttggggcc	ggtgagaatt	cggcctttcc	21240
[1937]	ggcaggtgcg	ccgatcccg	ggccatcaga	tatcgttccg	tctggctacg	tcctgatgca	21300
[1938]	ggggcaggcg	tttgacaaat	cagcctaccc	aaaacttgct	gtcgcgtatc	catcggtgt	21360
[1939]	gcttctgat	atgcgaggct	ggacaatcaa	ggggaaaccc	gccagcggtc	gtgctgtatt	21420
[1940]	gtctcaggaa	caggatggaa	ttaagtcgca	caccacaggt	gccagtgcac	ccggtacgga	21480
[1941]	tttggggacg	aaaaccacat	cgctgcttga	ttacgggacg	aaaacaacag	gcagtttcga	21540
[1942]	ttacggcacc	aaatcgacga	ataacacggg	ggctcatgct	cacagtctga	gcggttcaac	21600
[1943]	aggggccgcg	ggtgctcatg	cccacacaag	tggtttaagg	atgaacagtt	ctggctggag	21660
[1944]	tcagtatgga	acagcaacca	ttacaggaag	tttatccaca	gttaaaggaa	ccagcacaca	21720
[1945]	gggtattgct	tatttatcga	aaacggacag	tcagggcagc	cacagtcact	cattgtccgg	21780
[1946]	tacagccgtg	agtccgggtg	cacatgcgca	tacagttggt	attggtgcgc	accagcatcc	21840
[1947]	ggttggttatc	ggtgctcatg	cccattcttt	cagtattggt	tcacacggac	acaccatcac	21900
[1948]	cgttaacgct	gcgggtaacg	cggaaaacac	cgtaaaaaac	attgcattta	actatattgt	21960
[1949]	gaggcttgca	taatggcatt	cagaatgagt	gaacaaccac	ggaccataaa	aatttataat	22020

[1950]	ctgctggccg	gaactaatga	atattattggt	gaaggtgacg	catatattcc	gcctcatacc	22080
[1951]	ggtctgcctg	caaacagtac	cgatattgca	ccgccagata	ttccggctgg	ctttgtggct	22140
[1952]	gttttcaaca	gtgatgaggc	atcgtggcat	ctcgttgaag	accatcgggg	taaaaccgtc	22200
[1953]	tatgacgtgg	cttccggcga	cgcgttattt	atctctgaac	tcgggtccgtt	accggaaaat	22260
[1954]	tttacctggt	tatcgccggg	aggggaatat	cagaagtgga	acggcacagc	ctgggtgaag	22320
[1955]	gatacggaag	cagaaaaact	gttccggatc	cgggaggcgg	aagaaacaaa	aaaaagcctg	22380
[1956]	atgcaggtag	ccagtgaagc	tattgcgccg	cttcaggatg	ctgcagatct	ggaaattgca	22440
[1957]	acgaaggaag	aaacctcgtt	gctggaagcc	tggaagaagt	atcgggtggt	gctgaaccgt	22500
[1958]	gttgatacat	caactgcacc	tgatattgag	tgccctgctg	tccctgttat	ggagtaatcg	22560
[1959]	ttttgtgata	tgccgcagaa	acgttgtatg	aaataacgtt	ctgcggttag	ttagtatatt	22620
[1960]	gtaaagctga	gtattggttt	atctggcgat	tattatcttc	aggagaataa	tggaagttct	22680
[1961]	atgactcaat	tgttcatagt	gtttacatca	ccgccaatg	cttttaagac	tgaacgcatt	22740
[1962]	aaatatggtt	tttcgtcatg	ttttgagctt	gctgttgata	tttctaaagt	cggttttttt	22800
[1963]	tcttcgtttt	ctctaactat	tttccatgaa	atacattttt	gattattatt	tgaatcaatt	22860
[1964]	ccaattacct	gaagtctttc	atctataatt	ggcattgtat	gtattggttt	attggagtag	22920
[1965]	atgcttgctt	ttctgagcca	tagctctgat	atccaaatga	agccataggc	atttgttatt	22980
[1966]	ttggctctgt	cagctgcata	acgccaaaaa	atatattttt	ctgcttgatc	ttcaaatgtt	23040
[1967]	gtattgatta	aatcaattgg	atggaattgt	ttatcataaa	aaattaatgt	ttgaatgtga	23100
[1968]	taaccgtcct	ttaaaaaagt	cgtttctgca	agcttggctg	tatagtcaac	taactcttct	23160
[1969]	gtcgaagtga	tatttttagg	cttatctacc	agtttttagc	gctctttaat	atcttcagga	23220
[1970]	attattttat	tgatcatatt	tatcatgcta	aatgacaatt	tgcttatgga	gtaatctttt	23280
[1971]	aattttaaat	aagttattct	cctggcttca	tcaaataaag	agtcgaatga	tgttggcgaa	23340
[1972]	atcacatcgt	cacccattgg	attgtttatt	tgtatgcaa	gagagttaca	gcagttatac	23400
[1973]	attctgccat	agattatagc	taaggcatgt	aataattcgt	aatcttttag	cgtattagcg	23460
[1974]	acccatcgtc	tttctgattt	aataatagat	gattcagtta	aatatgaagg	taatttcttt	23520
[1975]	tgtgcaagtc	tgactaactt	ttttatacca	atgtttaaca	tactttcatt	tgtaataaac	23580
[1976]	tcaatgtcat	tttcttcaat	gtaagatgaa	ataagagtag	cctttgcctc	gctatacatt	23640
[1977]	tctaaatcgc	cttggttttc	tatcgtattg	cgagaatttt	tagcccaagc	cattaatgga	23700
[1978]	tcatttttcc	atttttcaat	aacattattg	ttataccaaa	tgatcatatc	tataatctgg	23760
[1979]	tttttggttt	tttgaataat	aaatgttact	gttcttgctg	tttgaggaa	ttgattcaaa	23820
[1980]	ttcaagcgaa	ataattcagg	gtcaaaaatat	gtatcaatgc	agcatttgag	caagtgcgat	23880
[1981]	aaatctttaa	gtcttctttc	ccatggtttt	ttagtcataa	aactctccat	tttgataggt	23940
[1982]	tgcatgctag	atgctgatat	attttagagg	tgataaaatt	aactgcttaa	ctgtcaatgt	24000
[1983]	aatacaagtt	gtttgatctt	tgcaatgatt	cttatcagaa	accatatagt	aaattagtta	24060
[1984]	cacaggaaat	ttttaatatt	attattatca	ttcattatgt	attaaaatta	gagttgtggc	24120
[1985]	ttggctctgc	taacacgttg	ctcataggag	atatggtaga	gccgcagaca	cgtcgtatgc	24180
[1986]	aggaacgtgc	tgccgctggc	tggtgaactt	ccgatagtgc	gggtgttgaa	tgatttccag	24240
[1987]	ttgctaccga	ttttacatat	tttttgcatg	agagaatttg	taccacctcc	caccgacct	24300
[1988]	ctatgactgt	acgccactgt	ccctaggact	gctatgtgcc	ggagcggaca	ttacaaacgt	24360

[1989]	ccttctcggg	gcatgccact	gttgccaatg	acctgcctag	gaattgggta	gcaagttact	24420
[1990]	accggatttt	gtaaaaacag	ccctcctcat	ataaaaagta	ttcgttcact	tccgataagc	24480
[1991]	gtcgtaat	tctatctttc	atcatattct	agatccctct	gaaaaaatct	tccgagtttg	24540
[1992]	ctaggcactg	atacataact	cttttccaat	aattggggaa	gtcattcaaa	tctataatag	24600
[1993]	gtttcagatt	tgcttcaata	aattctgact	gtagctgctg	aaacgttgcg	gttgaactat	24660
[1994]	atttccttat	aactttttacg	aaagagtttc	tttgagtaat	cacttcactc	aagtgccttc	24720
[1995]	ctgcctccaa	acgataacctg	ttagcaatat	ttaatagctt	gaaatgatga	agagctctgt	24780
[1996]	gtttgtcttc	ctgcctccag	ttcgccgggc	attcaacata	aaaactgata	gcacccggag	24840
[1997]	ttccggaaac	gaaatttgca	tataccatt	gtcacgaaa	aaaaatgtcc	ttgtcgatat	24900
[1998]	agggatgaat	cgcttggtgt	acctcatcta	ctgcgaaaac	ttgacctttc	tctcccatat	24960
[1999]	tgtagtcg	gcacgatgga	actaaattaa	taggcacac	cgaaaattca	ggataatgtg	25020
[2000]	caataggaag	aaaatgatct	atattttttg	tctgtcctat	atcaccacaa	aatggacatt	25080
[2001]	tttcacctga	tgaacaagc	atgtcatcgt	aatatgttct	agcgggttg	tttttatctc	25140
[2002]	ggagattatt	ttcataaagc	ttttctaatt	taacctttgt	caggttacca	actactaagg	25200
[2003]	ttgtaggctc	aagagggtgt	gtcctgtcgt	aggtaaataa	ctgacctgtc	gagcttaata	25260
[2004]	ttctatattg	ttgttctttc	tgcaaaaaag	tggggaagtg	agtaatgaaa	ttattttctaa	25320
[2005]	catttatctg	catcatacct	tccgagcatt	tattaagcat	ttcgtataaa	gttctcgtg	25380
[2006]	gaagaggtag	ttttttcatt	gtactttacc	ttcatctctg	ttcattatca	tcgcttttaa	25440
[2007]	aacggttcga	ccttctaate	ctatctgacc	attataat	tttagaatgg	tttcataaga	25500
[2008]	aagctctgaa	tcaacggact	gcgataataa	gtggtggtat	ccagaatttg	tcacttcaag	25560
[2009]	taaaaacacc	tcacgagtta	aaacacctaa	gttctcaccg	aatgtctcaa	tatccggacg	25620
[2010]	gataatattt	attgcttctc	ttgaccgtag	gactttccac	atgcaggatt	ttggaacctc	25680
[2011]	ttgcagtact	actggggaat	gagttgcaat	tattgctaca	ccattgcgtg	catcgagtaa	25740
[2012]	gtcgtttaat	gttcgtaaaa	aagcagagag	caaagggtga	tgcatatgaa	cctctggttc	25800
[2013]	atcgaataaa	actaatgact	tttcgccaac	gacatctact	aatcttgtga	tagtaaataa	25860
[2014]	aacaattgca	tgtccagagc	tcattcgaag	cagatatttc	tgatatttgt	cataaaacaa	25920
[2015]	tttagtgaat	ttatcatcgt	ccacttgaat	ctgtggttca	ttacgtctta	actcttcata	25980
[2016]	tttagaaatg	aggctgatga	gttccatatt	tgaaaagttt	tcactactac	ttagtttttt	26040
[2017]	gatagettca	agccagagtt	gtctttttct	atctactctc	atacaaccaa	taaatgctga	26100
[2018]	aatgaattct	aagcggagat	cgcctagtga	ttttaaacta	ttgctggcag	cattcttgag	26160
[2019]	tccaatataa	aagtattgtg	taccttttgc	tggttcaggt	tggtcttttag	gaggagtaaa	26220
[2020]	aggatcaaat	gcactaaacg	aaactgaaac	aagcgatcga	aaatatccct	ttgggattct	26280
[2021]	tgactcgata	agtctattat	tttcagagaa	aaaatattca	ttgttttctg	ggttggtgat	26340
[2022]	tgaccaatc	attccattca	aaattgttgt	tttaccacac	ccattccgcc	cgataaaagc	26400
[2023]	atgaatgttc	gtgctgggca	tagaattaac	cgtcacctca	aaaggtatag	ttaaatcact	26460
[2024]	gaatccggga	gcactttttc	tattaaatga	aaagtggaaa	tctgacaatt	ctggcaaac	26520
[2025]	atttaacaca	cgtgcgaact	gtccatgaat	ttctgaaaga	gttaccctc	taagtaatga	26580
[2026]	ggtgttaagg	acgttttcat	tttcaatgtc	ggctaatacga	tttgccata	ctactaaatc	26640
[2027]	ctgaatagct	ttaagaaggt	tatgttttaa	accatcgctt	aatttgctga	gattaacata	26700

[2028]	gtagtcaatg	ctttcaccta	aggaaaaaaa	catttcaggg	agttgactga	atTTTTtattc	26760
[2029]	tattaatgaa	taagtgetta	cttcttcttt	ttgacctaca	aaaccaatTT	taacattttcc	26820
[2030]	gatatcgcat	ttttcaccat	gctcatcaaa	gacagtaaga	taaaacattg	taacaaagga	26880
[2031]	atagtcattc	caaccatctg	ctcgtaggaa	tgccttattt	ttttctactg	caggaatata	26940
[2032]	cccgctctt	tcaataacac	taaactccaa	catatagtaa	cccttaattt	tattaaaata	27000
[2033]	accgcaatTT	atttggcggc	aacacaggat	ctctctttta	agttactctc	tattacatac	27060
[2034]	gttttccatc	taaaaattag	tagtattgaa	cttaacgggg	catcgtattg	tagttttcca	27120
[2035]	tatttagctt	tctgcttcct	tttgataac	ccactgttat	tcatgttgca	tgggtgactg	27180
[2036]	tttataccaa	cgatatagtc	tattaatgca	tatatagtat	cgccgaacga	ttagctcttc	27240
[2037]	aggcttctga	agaagcgTTT	caagtactaa	taagccgata	gatagccacg	gacttcgtag	27300
[2038]	ccatTTTTca	taagtgttaa	cttccgctcc	tcgctcataa	cagacattca	ctacagttat	27360
[2039]	ggcggaaagg	tatgcatgct	gggtgtgggg	aagtcgtgaa	agaaaagaag	tcagctgcgt	27420
[2040]	cgTTTgacat	cactgctatc	ttcttactgg	ttatgcaggt	cgtagtgggt	ggcacacaaa	27480
[2041]	gctttgcact	ggattgcgag	gctttgtgct	tctctggagt	gcgacaggTT	tgatgacaaa	27540
[2042]	aaattagcgc	aagaagacaa	aaatcacctt	gcgctaattc	tctgttacag	gtcactaata	27600
[2043]	ccatctaagt	agttgattca	tagtgactgc	atatgttgTg	ttttacagta	ttatgtagtc	27660
[2044]	tgTTTTttat	gcaaaatcta	atttaatat	ttgatattta	tatcatttta	cgtttctcgt	27720
[2045]	tcagctTTTT	tatactaagt	tggcattata	aaaaagcatt	gcttatcaat	ttgttgcaac	27780
[2046]	gaacaggtca	ctatcagtca	aaataaaatc	attatttgat	ttcaattttg	tcccactccc	27840
[2047]	tgcctctgtc	atcacgatac	tgtgatgcca	tgggtgccga	cttatgcccc	agaagatgtt	27900
[2048]	gagcaaaact	atcgcttatac	tgcttctcat	agagtcttgc	agacaaactg	cgcaactcgt	27960
[2049]	gaaaggtagg	cggatccccT	tcgaaggaaa	gacctgatgc	ttttcgtgcg	cgcataaaat	28020
[2050]	accttgatac	tgtgccggat	gaaagcggtt	cgcgacgagt	agatgcaatt	atggtttctc	28080
[2051]	cgccaagaat	ctctttgcat	ttatcaagtg	tttcttcat	tgatattccg	agagcatcaa	28140
[2052]	tatgcaatgc	tgTTgggatg	gcaattttta	cgctgtttt	gctttgctcg	acataaagat	28200
[2053]	atccatctac	gatatcagac	cacttcattt	cgcataaatc	accaactcgt	tgcccggtaa	28260
[2054]	caacagccag	ttccattgca	agtctgagcc	aacatggTga	tgattctgct	gcttgataaa	28320
[2055]	ttttcaggta	ttcgtcagcc	gtaagtcttg	atctccttac	ctctgatttt	gctgcgcgag	28380
[2056]	tggcagcgac	atggtttgTt	gttatatggc	cttcagctat	tgcctctcgg	aatgcatcgc	28440
[2057]	tcagtgttga	tctgattaac	ttggctgacg	ccgcttgcc	ctcgtctatg	tatccattga	28500
[2058]	gcattgccgc	aattttcttt	gtggTgatgt	cttcaagtgg	agcatcagge	agaccctcc	28560
[2059]	ttattgcttt	aattttgctc	atgtaattta	tgagtgtctt	ctgcttgatt	cctctgctgg	28620
[2060]	ccaggatttt	ttcgtagcga	tcaagccatg	aatgtaacgt	aacggaatta	tactgttga	28680
[2061]	ttctcgtgt	cagagcttg	tgtttgtgtc	ctgaaaataa	ctcaatgttg	gcctgtatag	28740
[2062]	cttcagtgat	tgcgattcgc	ctgtctctgc	ctaatccaaa	ctctttacc	gtccttgggt	28800
[2063]	ccctgtagca	gtaatatcca	ttgtttctta	tataaaggtt	agggggtaaa	tcccgcgct	28860
[2064]	catgacttcg	ccttcttccc	atttctgatac	ctcttcaaaa	ggccacctgt	tactggTcga	28920
[2065]	tttaagtcaa	cctttaccgc	tgattcgtgg	aacagatact	ctcttccatc	cttaaccgga	28980
[2066]	ggtgggaata	tcctgcattc	ccgaacccat	cgacgaactg	tttcaaggct	tcttggacgt	29040

[2067]	cgctggcgtg	cgttccactc	ctgaagtgtc	aagtacatcg	caaagtctcc	gcaattacac	29100
[2068]	gcaagaaaaa	accgccatca	ggcggttg	gtttctttca	gttcttcaat	tcgaatattg	29160
[2069]	gttacgtctg	catgtgctat	ctgcgcccat	atcatccagt	ggtcgtagca	gtcgttgatg	29220
[2070]	ttctccgctt	cgataactct	gttgaatggc	tctccattcc	attctcctgt	gactcggaag	29280
[2071]	tgcatttatc	atctccataa	aacaaaaccc	gccgtagcga	gttcagataa	aataaatccc	29340
[2072]	cgcgagtgcg	aggattgtta	tgtaatattg	ggtttaatca	tctatatgtt	ttgtacagag	29400
[2073]	agggaagta	tcgtttccac	cgtactcgtg	ataataattt	tgcacggtat	cagtcatttc	29460
[2074]	tcgcacattg	cagaatgggg	atttgtcttc	attagactta	taaaccttca	tggaatattt	29520
[2075]	gtatgccgac	tctatatcta	taccttcata	tacataaaca	ccttcgtgat	gtctgcatgg	29580
[2076]	agacaagaca	ccggatctgc	acaacattga	taacgcccac	tctttttgct	cagactctaa	29640
[2077]	ctcattgata	ctcatttata	aactccttgc	aatgtatgtc	gtttcagcta	aacgggtatca	29700
[2078]	gcaatgttta	tgtaaagaaa	cagtaagata	atactcaacc	cgatgtttga	gtacggtcac	29760
[2079]	catctgacac	tacagactct	ggcctcgtg	tgaagacgac	gcgaaattca	gcattttcac	29820
[2080]	aagcgttatc	ttttacaaaa	ccgatctcac	tctcctttga	tgcgaaatgcc	agcgtcagac	29880
[2081]	atcatatgca	gatactcacc	tgcactctga	accattgac	ctccaacccc	gtaatagcga	29940
[2082]	tgcgtaatga	gtcgatagt	tactaacggg	tcttgttcga	ttaactgccg	cagaaactct	30000
[2083]	tccaggtcac	cagtgcagtg	cttgataaca	ggagtcttcc	caggatggcg	aacaacaaga	30060
[2084]	aactggtttc	cgtcttcacg	gacttcgttg	ctttccagtt	tagcaatacg	cttactccca	30120
[2085]	tccagataaa	caccttcgta	atactcacgc	tgctcgttga	gttttgattt	tgctgtttca	30180
[2086]	agctcaacac	gcagtttccc	tactgttagc	gcaatatcct	cgttctcctg	gtcgcggcgt	30240
[2087]	ttgatgtatt	gctggtttct	ttcccgttca	tccagcagtt	ccagcacaat	cgatggtgtt	30300
[2088]	accaattcat	ggaaaaggtc	tgcgtcaa	ccccagtcgt	catgcattgc	ctgctctgcc	30360
[2089]	gcttcacgca	gtgcctgaga	gttaatttcg	ctcacttcga	acctctctgt	ttactgataa	30420
[2090]	gttccagatc	ctcctggcaa	cttgacacaag	tccgacaacc	ctgaacgacc	aggcgtcttc	30480
[2091]	gttcatctat	cggatcgcca	cactcacaac	aatgagtggc	agatatagcc	tggtggttca	30540
[2092]	ggcggcgcat	ttttattgct	gtgttgcgct	gtaattcttc	tatttctgat	gctgaatcaa	30600
[2093]	tgatgtctgc	catctttcat	taatccctga	actgttggtt	aatacgcttg	agggtgaatg	30660
[2094]	cgaataataa	aaaaggagcc	tgtagctccc	tgatgatttt	gcttttcatg	ttcatcggtc	30720
[2095]	cttaaagacg	ccgtttaaca	tgccgattgc	caggetttaa	tgagtcggtg	tgaatcccat	30780
[2096]	cagcgttacc	gtttcgcggt	gcttcttcag	tacgtacgg	caaattgcat	cgacgttttt	30840
[2097]	atccggaaac	tgctgtctgg	ctttttttga	tttcagaatt	agcctgacgg	gcaatgctgc	30900
[2098]	gaagggcggt	ttcctgctga	ggtgtcattg	aacaagtccc	atgtcggcaa	gcataagcac	30960
[2099]	acagaatatg	aagcccgtg	ccagaaaaat	gcattccgtg	gttgtcatac	ctggtttctc	31020
[2100]	tcacttgctt	ctgctttcgc	caccatcatt	tccagctttt	gtgaaaggga	tgcggctaac	31080
[2101]	gtatgaaatt	cttcgtctgt	ttctactggt	attggcacia	acctgattcc	aatttgagca	31140
[2102]	aggctatgtg	ccatctcgat	actcgttctt	aactcaacag	aagatgcttt	gtgcatacag	31200
[2103]	cccctcgttt	attattttatc	tcctcagcca	gccgctgtgc	tttcagtggg	tttcggataa	31260
[2104]	cagaaaggcc	gggaaatacc	cagcctcgct	ttgtaacgga	gtagacgaaa	gtgattgcgc	31320
[2105]	ctacccggat	attatcggtg	ggatgcgtca	tcgccattgc	tcccaaata	caaaaccaat	31380

[2106]	ttcagccagt	gcctcgtcca	ttttttcgat	gaactccggc	acgatctcgt	caaaactcgc	31440
[2107]	catgtacttt	tcatcccgt	caatcacgac	ataatgcagg	ccttcacgct	tcatacgcgg	31500
[2108]	gtcatagttg	gcaaagtacc	aggcattttt	tcgcgtcacc	cacatgctgt	actgcacctg	31560
[2109]	ggccatgtaa	gctgacttta	tggcctcgaa	accaccgagc	cggaacttca	tgaaatcccg	31620
[2110]	ggaggtaaac	gggcatttca	gttcaaggcc	gttgccgtca	ctgcataaac	catcgggaga	31680
[2111]	gcaggcggta	cgcatacttt	cgtcgcgata	gatgatcggg	gattcagtaa	cattcacgcc	31740
[2112]	ggaagtgaat	tcaaacaggg	ttctggcgtc	gttctcgtac	tgttttcccc	aggccagtgc	31800
[2113]	tttagcgtta	acttccggag	ccacaccggt	gcaaacctca	gcaagcaggg	tgtggaagta	31860
[2114]	ggacattttc	atgtcaggcc	acttcttttc	ggagcgggggt	tttgctatca	cgttgtgaac	31920
[2115]	ttctgaagcg	gtgatgacgc	cgagccgtaa	tttgtgccac	gcatcatccc	cctgttcgac	31980
[2116]	agctctcaca	tcgatcccg	tacgtcgcag	gataatgtcc	ggtgtcatgc	tgccaccttc	32040
[2117]	tgctctgcgg	ctttctgttt	caggaatcca	agagctttta	ctgcttcggc	ctgtgtcagt	32100
[2118]	tctgacgatg	cacgaatgtc	gcggcgaaat	atctgggaac	agagcggcaa	taagtcgtca	32160
[2119]	tcccatgttt	tatccagggc	gatcagcaga	gtgttaatct	cctgcatggt	ttcatcgtta	32220
[2120]	accggagtga	tgctcgttcc	cggctgacgt	tctgcagtgt	atgcagtatt	ttcgacaatg	32280
[2121]	cgctcggctt	catccttgtc	atagatacca	gcaaatccga	aggccagacg	ggcacactga	32340
[2122]	atcatggctt	tatgacgtaa	catccgtttg	ggatgcgact	gccacggccc	cgtgatttct	32400
[2123]	ctgccttcgc	gagttttgaa	tggttcgcgg	cggcattcat	ccatccattc	ggtaacgcag	32460
[2124]	atcggatgat	tacggtcctt	gcggtaaatc	cggcatgtac	aggattcatt	gtcctgtctca	32520
[2125]	aagtccatgc	catcaaactg	ctggttttca	ttgatgatgc	gggaccagcc	atcaacgccc	32580
[2126]	accaccgga	cgatgccatt	ctgcttatca	ggaaaggcgt	aaattttctt	cgtccacgga	32640
[2127]	ttaaggccgt	actggttggc	aacgatcagt	aatgcgatga	actgcgcata	gctggcatca	32700
[2128]	cctttaaatg	ccgtctggcg	aagagtgggtg	atcagttcct	gtgggtcgac	agaatccatg	32760
[2129]	ccgacacgtt	cagccagctt	cccagccagc	gttgcgagt	cagtactcat	tcgttttata	32820
[2130]	cctctgaatc	aatatcaacc	tgggtggtgag	caatggtttc	aacctatgtac	cggatgtgtt	32880
[2131]	ctgccatgcg	ctcctgaaac	tcaacatcgt	catcaaacgc	acgggtaatg	gattttttgc	32940
[2132]	tggccccgtg	gcgttgcaaa	tgatcgatgc	atagcgattc	aaacaggtgc	tggggcaggc	33000
[2133]	ctttttccat	gtcgtctgcc	agttctgcct	ctttctcttc	acgggcgagc	tgctggtagt	33060
[2134]	gacgcgcccc	gctctgagcc	tcaagacgat	cctgaatgta	ataagcgttc	atggctgaac	33120
[2135]	tcctgaaata	gctgtgaaaa	tatgcgccgc	gaaatgccgg	gctgattagg	aaaacaggaa	33180
[2136]	aggggggttag	tgaatgcttt	tgcttgatct	cagtttcagt	attaatatcc	attttttata	33240
[2137]	agcgtcgacg	gcttcacgaa	acatcttttc	atcgccaata	aaagtggcga	tagtgaattt	33300
[2138]	agtctggata	gccataagtg	tttgatccat	tctttgggac	tcctggctga	ttaagtatgt	33360
[2139]	cgataaggcg	tttccatccg	tcacgtaatt	tacgggtgat	tcgttcaagt	aaagattcgg	33420
[2140]	aagggcagcc	agcaacaggc	caccctgcaa	tggcatattg	catggtgtgc	tccttattta	33480
[2141]	tacataacga	aaaacgcctc	gagtgaagcg	ttattggtat	gcggtaaaac	cgcactcagg	33540
[2142]	cggccttgat	agtcataatca	tctgaatcaa	atattcctga	tgtatcgata	tcggtaattc	33600
[2143]	ttattccttc	gctaccatcc	attggaggcc	atccttcctg	accatttcca	tcattccagt	33660
[2144]	cgaactcaca	cacaacacca	tatgcattta	agtcgcttga	aattgctata	agcagagcat	33720

[2145]	gttgccgag	catgattaat	acagcattta	atacagagcc	gtgtttattg	agtcggtatt	33780
[2146]	cagagtctga	ccagaaatta	ttaatctggt	gaagtttttc	ctctgtcatt	acgtcatggt	33840
[2147]	cgatttcaat	ttctattgat	gctttccagt	cgtaatcaat	gatgtatttt	ttgatgtttg	33900
[2148]	acatctgttc	atatcctcac	agataaaaaa	tcgccctcac	actggagggc	aaagaagatt	33960
[2149]	tccaataatc	agaacaagtc	ggctcctgtt	tagttacgag	cgacattgct	ccgtgtattc	34020
[2150]	actcgttggg	atgaatacac	agtgcagtgt	ttattctgtt	atttatgcc	aaaataaagg	34080
[2151]	ccactatcag	gcagctttgt	tgttctgttt	accaagtctt	ctggcaatca	ttgccgtcgt	34140
[2152]	tcgtattgcc	catttatcga	catatttccc	atcttccatt	acaggaaaca	tttcttcagg	34200
[2153]	cttaaccatg	cattccgatt	gcagcttgca	tccattgcat	cgcttgaatt	gtccacacca	34260
[2154]	ttgattttta	tcaatagtcg	tagtcatacg	gatagtctcg	gtattgttcc	atcacatcct	34320
[2155]	gaggatgctc	ttcgaactct	tcaaattctt	cttccatata	tcaccttaaa	tagtggattg	34380
[2156]	cggtagtaaa	gattgtgcct	gtcttttaac	cacatcaggc	tcggtggttc	tcgtgtaccc	34440
[2157]	ctacagcgag	aaatcggata	aactattaca	accctacag	tttgatgagt	atagaaatgg	34500
[2158]	atccactcgt	tattctcgga	cgagtgttca	gtaatgaacc	tctggagaga	accatgtata	34560
[2159]	tgatcgttat	ctgggttggg	cttctgcttt	taagcccaga	taactggcct	gaatatgtta	34620
[2160]	atgagagaat	cggtattcct	catgtgtggc	atgttttcgt	ctttgctctt	gcattttcgc	34680
[2161]	tagcaattaa	tgtgcatcga	ttatcagcta	ttgccagcgc	cagatataag	cgatttaagc	34740
[2162]	taagaaaacg	cattaagatg	caaaacgata	aagtgcgatc	agtaattcaa	aaccttacag	34800
[2163]	aagagcaatc	tatggttttg	tgcgcagccc	ttaatgaagg	caggaaagtat	gtgggttacat	34860
[2164]	caaaacaatt	cccatacatt	agtgagttga	ttgagcttgg	tgtgttgaac	aaaacttttt	34920
[2165]	cccgatggaa	tggaaagcat	atattattcc	ctattgagga	tatttactgg	actgaattag	34980
[2166]	ttgccagcta	tgatccatat	aattattgaga	taaagccaag	gccaatatct	aagtaactag	35040
[2167]	ataagaggaa	tcgattttcc	cttaattttc	tggcgtccac	tgcatgttat	gccgcgttcg	35100
[2168]	ccaggcttgc	tgtacatgtg	gcgctgattc	ttgcgctcaa	tacgttgcag	gttgctttca	35160
[2169]	atctgtttgt	ggtattcagc	cagcactgta	aggtctatcg	gatttagtgc	gctttctact	35220
[2170]	cgtgatttcg	gtttgcgatt	cagcgagaga	atagggcggt	taactggttt	tgcgcttacc	35280
[2171]	ccaaccaaca	ggggatttgc	tgctttccat	tgagcctggt	tctctgcgcg	acgttcgcgg	35340
[2172]	cggcgtgttt	gtgcatccat	ctggattctc	ctgtcagtta	gctttggtgg	tgtgtggcag	35400
[2173]	ttgtagtcct	gaacgaaaac	ccccgcgat	tggcacattg	gcagctaate	cggaatcgca	35460
[2174]	cttacggcca	atgcttcgtt	tcgtatcaca	caccccaaag	ccttctgctt	tgaatgctgc	35520
[2175]	ccttcttcag	ggcttaattt	ttaagagcgt	caccttcattg	gtggtcagtgc	cgctctgctg	35580
[2176]	atgtgctcag	tatcaccgcc	agtggatttt	atgtcaacac	cgccagagat	aatttatcac	35640
[2177]	cgcagatggt	tatctgtatg	ttttttatat	gaattttatt	tttgcagggg	ggcattgttt	35700
[2178]	ggtaggtgag	agatctgaat	tgctatgttt	agtgagttgt	atctatttat	ttttcaataa	35760
[2179]	atacaattgg	ttatgtgttt	tgggggcgat	cgtgaggcaa	agaaaacccg	gcgctgaggc	35820
[2180]	cgggttattc	ttgttctctg	gtcaaatatt	atagttggaa	aacaaggatg	catatatgaa	35880
[2181]	tgaacgatgc	agaggcaatg	ccgatggcga	tagtgggtat	catgtagccg	cttatgctgg	35940
[2182]	aaagaagcaa	taaccgcgag	aaaaacaaag	ctccaagctc	aacaaaacta	agggcataga	36000
[2183]	caataactac	cgatgtcata	taccataact	ctctaattct	ggccagtcgg	cgcgttctgc	36060

[2184]	ttccgattag	aaacgtcaag	gcagcaatca	ggattgcaat	catggttcct	gcatatgatg	36120
[2185]	acaatgtcgc	cccaagacca	tctctatgag	ctgaaaaaga	aacaccagga	atgtagtggc	36180
[2186]	ggaaaaggag	atagcaaattg	cttacgataa	cgtaaggaat	tattactatg	taaacaccag	36240
[2187]	gcatgattct	gttccgcata	attactcctg	ataattaatc	cttaactttg	cccacctgcc	36300
[2188]	ttttaaaaca	ttccagtata	tcacttttca	ttcttgcgta	gcaatatgcc	atctcttcag	36360
[2189]	ctatctcagc	attggtgacc	ttgttcagag	gcgctgagag	atggcctttt	tctgatagat	36420
[2190]	aatgtttctgt	taaaatatct	ccggcctcat	cttttgcccg	caggctaattg	tctgaaaatt	36480
[2191]	gaggtgacgg	gttaaaaaata	atatccttgg	caaccttttt	tatatccctt	ttaaattttg	36540
[2192]	gcttaatgac	tatatccaat	gagtcaaaaa	gctccccttc	aatatctggt	gcccctaaga	36600
[2193]	cctttaatat	atcgccaaat	acaggtagct	tggtctctac	cttcaccgtt	gttcggccga	36660
[2194]	tgaatgcat	atgcataaca	tcgtcttttg	tggttccctt	catcagtggc	tctatctgaa	36720
[2195]	cgcgctctcc	actgcttaat	gacattcctt	tcccgattaa	aaaatctgtc	agatcggatg	36780
[2196]	tggtcggccc	gaaaacagtt	ctggcaaaaac	caatggtgtc	gccttcaaca	aacaaaaaag	36840
[2197]	atgggaatcc	caatgattcg	tcactctgca	ggctgttctt	aatatcttca	actgaagctt	36900
[2198]	tagagcgatt	tatcttctga	accagactct	tgtcatttgt	tttggtaaag	agaaaagttt	36960
[2199]	ttccatcgat	tttatgaata	tacaaataat	tggagccaac	ctgcaggatga	tgattatcag	37020
[2200]	ccagcagaga	attaaggaaa	acagacaggt	ttattgagcg	cttatctttc	cctttatttt	37080
[2201]	tgctgcggta	agtcgcataa	aaaccattct	tcataattca	atccatttac	tatgttatgt	37140
[2202]	tctgagggga	gtgaaaattc	ccctaattcg	atgaagattc	ttgctcaatt	gttatcagct	37200
[2203]	atgcgccgac	cagaacacct	tgccgatcag	ccaaacgtct	cttcaggcca	ctgactagcg	37260
[2204]	ataactttcc	ccacaacgga	acaactctca	ttgcatggga	tcattgggta	ctgtgggttt	37320
[2205]	agtggttgta	aaaacacctg	accgctatcc	ctgatcagtt	tcttgaaggt	aaactcatca	37380
[2206]	cccccaagtc	tggctatgca	gaaatcacct	ggctcaacag	cctgctcagg	gtcaacgaga	37440
[2207]	attaacattc	cgtcaggaaa	gcttggcttg	gagcctgttg	gtcgggtcat	ggaattacct	37500
[2208]	tcaacctcaa	gccagaatgc	agaatcactg	gcttttttgg	ttgtgcttac	ccatctctcc	37560
[2209]	gcatcacctt	tggtaaaggt	tctaagctta	ggtgagaaca	tccctgcctg	aacatgagaa	37620
[2210]	aaaacagggt	actcatactc	acttctaagt	gacggctgca	tactaaccgc	ttcatacatc	37680
[2211]	tcgtagattt	ctctggcgat	tgaagggcta	aattcttcaa	cgctaacttt	gagaattttt	37740
[2212]	gtaagcaatg	cggcgttata	agcattttaat	gcattgatgc	cattaaataa	agcaccaacg	37800
[2213]	cctgactgcc	ccatcccat	cttgtctgcg	acagattcct	gggataagcc	aagttcattt	37860
[2214]	ttcttttttt	cataaattgc	tttaaggcga	cgtgcgtcct	caagctgctc	ttgtgttaat	37920
[2215]	ggtttctttt	ttgtgctcat	acgtttaaate	tatcacgcga	agggataaat	atctaaccac	37980
[2216]	gtgcgtgttg	actattttac	ctctggcggt	gataatggtt	gcatgtacta	aggaggttgt	38040
[2217]	atggaacaac	gcataaccct	gaaagattat	gcaatgcgct	ttgggcaaac	caagacagct	38100
[2218]	aaagatctcg	gcgtatatca	aagcgcgatc	aacaaggcca	ttcatgcagg	ccgaaagatt	38160
[2219]	tttttaacta	taaacgtga	tgggaagcgtt	tatgcggaag	aggtaaagcc	cttcccagat	38220
[2220]	aacaaaaaaa	caacagcata	aataaccccg	ctcttacaca	ttccagccct	gaaaaagggc	38280
[2221]	atcaaattaa	accacaccta	tgggtgtatgc	atttatttgc	atacattcaa	tcaattgtta	38340
[2222]	tctaaggaaa	tacttacata	tggttcgtgc	aaacaaacgc	aacgaggctc	tacgaatcga	38400

[2223]	gagtgcgttg	cttaacaaaa	tcgcaatgct	tggaactgag	aagacagcgg	aagctgtggg	38460
[2224]	cgttgataag	tcgcagatca	gcaggtggaa	gagggactgg	attccaaagt	tctcaatgct	38520
[2225]	gcttgctggt	cttgaatggg	gggtcggtga	cgacgacatg	gctcgattgg	cgcgacaagt	38580
[2226]	tgctgcgatt	ctcaccaata	aaaaacgccc	ggcggcaacc	gagcgttctg	aacaaatcca	38640
[2227]	gatggagttc	tgaggtcatt	actggatcta	tcaacaggag	tcattatgac	aaatacagca	38700
[2228]	aaaatactca	acttcggcag	aggtaacttt	gccggacagg	agcgtaatgt	ggcagatctc	38760
[2229]	gatgatgggt	acgccagact	atcaaatatg	ctgcttgagg	cttattcggg	cgcagatctg	38820
[2230]	accaagcgac	agtttaaagt	gctgcttgcc	attctgcgta	aaacctatgg	gtggaataaa	38880
[2231]	ccaatggaca	gaatcaccga	ttctcaactt	agcgagatta	caaagttacc	tgtcaaacgg	38940
[2232]	tgcaatgaag	ccaagttaga	actcgtcaga	atgaatatta	tcaagcagca	aggcggcatg	39000
[2233]	tttgaccac	ataaaaacat	ctcagaatgg	tgcacccctc	aaaacgaggg	aaaatcccct	39060
[2234]	aaaacgaggg	ataaaacatc	cctcaaattg	ggggattgct	atccctcaaa	acagggggac	39120
[2235]	acaaaagaca	ctattacaaa	agaaaaaaga	aaagattatt	cgtcagagaa	ttctggcgaa	39180
[2236]	tcctctgacc	agccagaaaa	cgacctttct	gtggtgaaac	cggatgctgc	aattcagagc	39240
[2237]	ggcagcaagt	gggggacagc	agaagacctg	accgccgcag	agtggatggt	tgacatggtg	39300
[2238]	aagactatcg	caccatcagc	cagaaaaccg	aattttgctg	ggtgggctaa	cgatatccgc	39360
[2239]	ctgatgcgtg	aacgtgacgg	acgtaaccac	cgcgacatgt	gtgtgctggt	ccgctgggca	39420
[2240]	tgccaggaca	acttctggtc	cggtaacgtg	ctgagcccgg	ccaaactccg	cgataagtgg	39480
[2241]	acccaactcg	aaatcaaccg	taacaagcaa	caggcaggcg	tgacagccag	caaaccacaaa	39540
[2242]	ctcgacctga	caaacacaga	ctggattttac	ggggtggatc	tatgaaaaac	atcgccgcac	39600
[2243]	agatggttaa	ctttgaccgt	gagcagatgc	gtcggatcgc	caacaacatg	ccggaacagt	39660
[2244]	acgacgaaaa	gccgcaggta	cagcaggtag	cgcagatcat	caacggtgtg	ttcagccagt	39720
[2245]	tactggcaac	tttcccggcg	agcctggcta	accgtgacca	gaacgaagtg	aacgaaatcc	39780
[2246]	gtcgccagtg	ggttctggct	tttcgggaaa	acgggatcac	cacgatggaa	caggttaacg	39840
[2247]	caggaatgcg	cgtagcccg	cggcagaatc	gaccatttct	gccatcacc	gggcagtttg	39900
[2248]	ttgcatgggt	ccgggaagaa	gcatccgtta	ccgccggact	gccaaacgtc	agcgagctgg	39960
[2249]	ttgatatggt	ttacgagtat	tgccggaagc	gaggcctgta	tccgatgctg	gagtcttatc	40020
[2250]	cgtggaaatc	aaacgcgcac	tactggctgg	ttaccaacct	gtatcagaac	atgcgggcca	40080
[2251]	atgcgcttac	tgatgcggaa	ttacgccgta	aggccgcaga	tgagcttgct	catatgactg	40140
[2252]	cgagaattaa	ccgtgggtgag	gcgatccctg	aaccagtaaa	acaacttcct	gtcatgggcg	40200
[2253]	gtagacctct	aaatcgtgca	caggctctgg	cgaagatcgc	agaaatcaaa	gctaagttcg	40260
[2254]	gactgaaagg	agcaagtgtg	tgacgggcaa	agaggcaatt	attcattacc	tggggacgca	40320
[2255]	taatagcttc	tgtgcgccgg	acgttgccgc	gctaacaggc	gcaacagtaa	ccagcataaa	40380
[2256]	tcaggccgcg	gctaaaatgg	cacgggcagg	tcttctgggt	atcgaaggta	aggtctggcg	40440
[2257]	aacggtgtat	taccggtttg	ctaccaggga	agaacgggaa	ggaaagatga	gcacgaacct	40500
[2258]	ggttttttaag	gagtgtcgcc	agagtgcgcg	gatgaaacgg	gtattggcgg	tatatggagt	40560
[2259]	taaaagatga	ccatctacat	tactgagcta	ataacaggcc	tgctggtaat	cgcaggcctt	40620
[2260]	tttatttggg	ggagagggaa	gtcatgaaaa	aactaacctt	tgaattcga	tctccagcac	40680
[2261]	atcagcaaaa	cgctattcac	gcagtacagc	aatccttcc	agaccaacc	aaaccaatcg	40740

[2262]	tagtaacat	tcaggaacgc	aaccgcagct	tagacaaaa	caggaagcta	tgggcctgct	40800
[2263]	taggtgacgt	ctctcgtcag	gttgaatggc	atggtcgctg	gctggatgca	gaaagctgga	40860
[2264]	agtgtgtgtt	taccgcagca	ttaaagcagc	aggatgttgt	tcctaaccctt	gccgggaatg	40920
[2265]	gctttgtggt	aataggccag	tcaaccagca	ggatgcgtgt	aggcgaatctt	gcgagctat	40980
[2266]	tagagcttat	acaggcattc	ggtacagagc	gtggcgtaa	gtggtcagac	gaagcgagac	41040
[2267]	tggctctgga	gtggaaagcg	agatggggag	acagggctgc	atgataaatg	tcgttagttt	41100
[2268]	ctccggtggc	aggacgtcag	catatttgct	ctggctaata	gagcaaaagc	gacgggcagg	41160
[2269]	taaagacgtg	cattacgttt	tcatggatac	aggttgtgaa	catccaatga	catatcggtt	41220
[2270]	tgtcaggga	gttgtgaagt	tctgggatat	accgctcacc	gtattgcagg	ttgatataca	41280
[2271]	cccggagctt	ggacagccaa	atgggtatac	ggtatgggaa	ccaaaggata	ttcagacgcg	41340
[2272]	aatgctgtt	ctgaagccat	ttatcgatat	ggtaaagaaa	tatggcactc	catacgtcgg	41400
[2273]	cggcgcttc	tgcactgaca	gattaaaact	cgttccttc	accaaatact	gtgatgacca	41460
[2274]	tttcgggcga	gggaattaca	ccacgtggat	tggcatcaga	gctgatgaac	cgaagcggct	41520
[2275]	aaagccaaag	cctggaatca	gatatcttgc	tgaactgtca	gactttgaga	aggaagatat	41580
[2276]	cctcgcatgg	tggaagcaac	aaccattcga	tttgcaaata	ccggaacatc	tcggttaactg	41640
[2277]	catattctgc	attaaaaaat	caacgcaaaa	aatcggactt	gcctgcaaag	atgaggaggg	41700
[2278]	attgcagcgt	gtttttaatg	aggtcatcac	gggatcccat	gtgcgtgacg	gacatcgga	41760
[2279]	aacgccaag	gagattatgt	accgaggaag	aatgtcgtg	gacggtatcg	cgaaaatgta	41820
[2280]	ttcagaaaat	gattatcaag	ccctgtatca	ggacatggta	cgagctaaaa	gattcgatac	41880
[2281]	cggctcttgt	tctgagtcac	gcgaaatatt	tggaggcgag	cttgatttcg	acttcgggag	41940
[2282]	ggaagctgca	tgatgcgatg	ttatcggtgc	ggtgaatgca	aagaagataa	ccgcttcgga	42000
[2283]	ccaaatcaac	cttactggaa	tcgatggtgt	ctccggtgtg	aaagaacacc	aacaggggtg	42060
[2284]	ttaccactac	cgcaggaaaa	ggaggacgtg	tggcgagaca	gcgacgaagt	atcaccgaca	42120
[2285]	taatctgcga	aaactgcaaa	taccttccaa	cgaacgcac	cagaaataaa	cccaagccaa	42180
[2286]	tcccaaaaga	atctgacgta	aaaaccttca	actacacggc	tcacctgtgg	gatataccggt	42240
[2287]	ggctaagacg	tcgtgcgagg	aaaacaaggt	gattgaccaa	aatcgaagtt	acgaacaaga	42300
[2288]	aagcgtcgag	cgagctttta	cgtgcgctaa	ctgcggtcag	aagctgcatg	tgctggaagt	42360
[2289]	tcacgtgtgt	gagcactgct	gcgcagaact	gatgagcgat	ccgaatagct	cgatgcacga	42420
[2290]	ggaagaagat	gatggctaaa	ccagcgcgaa	gacgatgtaa	aaacgatgaa	tgccgggaat	42480
[2291]	ggtttcaccc	tgcattcgct	aatcagtggt	ggtgctctcc	agagtgtgga	accaagatag	42540
[2292]	cactcgaacg	acgaagtaaa	gaacgcgaaa	aagcggaaaa	agcagcagag	aagaaacgac	42600
[2293]	gacgagagga	gcagaaacag	aaagataaac	ttaagattcg	aaaactcgcc	ttaaagcccc	42660
[2294]	gcagttactg	gattaaacaa	gcccacaag	ccgtaaacgc	cttcatcaga	gaaagagacc	42720
[2295]	gcgacttacc	atgtatctcg	tgcggaacgc	tcacgtctgc	tcagtgggat	gccggacatt	42780
[2296]	accggacaac	tgctgcggca	cctcaactcc	gatttaatga	acgcaatatt	cacaagcaat	42840
[2297]	gcgtggtgtg	caaccagcac	aaaagcggaa	atctcgttcc	gtatcgctgc	gaactgatta	42900
[2298]	gccgcacgg	gcaggaagca	gtagacgaaa	tcgaatcaaa	ccataaccgc	catcgctgga	42960
[2299]	ctatcgaaga	gtgcaaggcg	atcaaggcag	agtaccaaca	gaaactcaaa	gacctgcgaa	43020
[2300]	atagcagaag	tgaggccgca	tgacgttctc	agtaaaaacc	attccagaca	tgctcgttga	43080

[2301]	aacatacggg	aatcagacag	aagtagcacg	cagactgaaa	tgtagtcgcg	gtacggtcag	43140
[2302]	aaaatacgtt	gatgataaag	acgggaaaat	gcacgccatc	gtcaacgacg	ttctcatggt	43200
[2303]	tcatcgcgga	tggagtgaag	gagatgcgct	attacgaaaa	aattgatggc	agcaaatacc	43260
[2304]	gaaatatttg	ggtagttggc	gatctgcacg	gatgctacac	gaacctgatg	aacaaactgg	43320
[2305]	atacgattgg	attcgacaac	aaaaaagacc	tgcttatctc	ggtgggcat	ttggttgatc	43380
[2306]	gtggtgcaga	gaacgttgaa	tgccctggaat	taatcacatt	cccctggttc	agagctgtac	43440
[2307]	gtggaaacca	tgagcaaatg	atgattgatg	gcttatcaga	gcgtggaaac	gttaatcact	43500
[2308]	ggctgcttaa	tggcgggtgg	tggttcttta	atctcgatta	cgacaaagaa	attctggcta	43560
[2309]	aagctcttgc	ccataaagca	gatgaacttc	cgtaaatcat	cgaactgggtg	agcaaagata	43620
[2310]	aaaaatatgt	tatctgccac	gccgattatc	cctttgacga	atacgagttt	ggaaagccag	43680
[2311]	ttgatcatca	gcaggtaatc	tggaaaccg	aacgaatcag	caactcaca	aacgggatcg	43740
[2312]	tgaagaaat	caaaggcgcg	gacacgttca	tctttggtca	tacgccagca	gtgaaaccac	43800
[2313]	tcaagtttgc	caaccaaag	tatatcgata	ccggcgagct	gttctgcgga	aacctaacat	43860
[2314]	tgattcaggt	acaggagaga	ggcgcatgag	actcgaaagc	gtagctaaat	ttcattcgcc	43920
[2315]	aaaaagcccc	atgatgagcg	actcaccacg	ggccacggct	tctgactctc	tttcgggtac	43980
[2316]	tgatgtgatg	gctgctatgg	ggatggcgca	atcacaagcc	ggattcggtg	tggctgcatt	44040
[2317]	ctgcggtaag	cacgaactca	gccagaacga	caaacaaaag	gctatcaact	atctgatgca	44100
[2318]	atttgcacac	aaggatatcg	ggaaataacc	tgggtgtggca	aagcttgaag	gaaataactaa	44160
[2319]	ggcaaaggtg	ctgcaagtgc	tcgcaacatt	cgcttatgcg	gattattgcc	gtagtgcgcg	44220
[2320]	gacgccgggg	gcaagatgca	gagattgcca	tggtagagcg	cgtgcggttg	atattgcca	44280
[2321]	aacagagctg	tgggggagag	ttgtcgagaa	agagtgcgga	agatgcaaag	gcgtcggtta	44340
[2322]	ttcaaggatg	ccagcaagcg	cagcatatcg	cgctgtgacg	atgctaatac	caaaccttac	44400
[2323]	ccaaccacc	tggtcacgca	ctgttaagcc	gctgtatgac	gctctggtgg	tgcaatgcca	44460
[2324]	caaagaagag	tcaatcgag	acaacatttt	gaatgcggtc	acacgttagc	agcatgattg	44520
[2325]	ccacggatgg	caacatatta	acggcatgat	attgacttat	tgaataaaat	tgggtaaatt	44580
[2326]	tgactcaacg	atgggttaat	tcgctcggtt	tggtagtgag	atgaaaagag	gcggcgctta	44640
[2327]	ctaccgattc	cgcctagttg	gtcacttcga	cgtatcgtct	ggaactccaa	ccatcgagg	44700
[2328]	cagagaggtc	tgcaaaatgc	aatcccgaag	cagttcgag	gtaatagtta	gagcctgcat	44760
[2329]	aacggtttcg	ggatttttta	tatctgcaca	acaggtaaga	gcattgagtc	gataatcggt	44820
[2330]	aagagtcggc	gagcctgggt	agccagtgc	ctttccgttg	tgctgaatta	agcgaatacc	44880
[2331]	ggaagcagaa	ccggatcacc	aatgcgtac	aggcgtcatc	gccgccagc	aacagcaca	44940
[2332]	cccaaactga	gccgtagcca	ctgtctgtcc	tgaattcatt	agtaatagtt	acgtcgggc	45000
[2333]	cttttacaca	tgaccttcgt	gaaagcgggt	ggcaggaggt	cgcgctaaca	acctcctgcc	45060
[2334]	gttttgcccc	tgcatatcgg	tcacgaacaa	atctgattac	taaacacagt	agcctggatt	45120
[2335]	tgttctatca	gtaatcgacc	ttattcctaa	ttaaataagag	caaataccct	tattgggggt	45180
[2336]	aagacatgaa	gatgccagaa	aaacatgacc	tgttgccgcg	cattctcgcg	gcaaaggaac	45240
[2337]	aaggcatcgg	ggcaatcctt	gcgtttgcaa	tggcgtacct	tcgcggcaga	tataatggcg	45300
[2338]	gtgcgtttac	aaaaacagta	atcgacgcaa	cgatgtgcgc	cattatcgcc	tagttcattc	45360
[2339]	gtgaccttct	cgacttcgcc	ggactaagta	gcaatctcgc	ttatataacg	agcgtgttta	45420

[2340]	tcggtacat	cgtactgac	tcgattggtt	cgcttatcaa	acgcttcgct	gctaaaaaag	45480
[2341]	ccggagtaga	agatggtaga	aatcaataat	caacgtaagg	cgttcctcga	tatgtctggcg	45540
[2342]	tggtcggagg	gaactgataa	cggacgtcag	aaaaccagaa	atcatggtta	tgacgtcatt	45600
[2343]	gtaggcggag	agctatttac	tgattactcc	gatcacctc	gcaaacttgt	cacgctaaac	45660
[2344]	ccaaaactca	aatcaacagg	cgccggacgc	taccagcttc	tttcccgttg	gtgggatgcc	45720
[2345]	taccgcaagc	agcttggcct	gaaagacttc	tctccgaaaa	gtcaggacgc	tgtggcattg	45780
[2346]	cagcagatta	aggagcgtgg	cgctttacct	atgattgata	gtggtgatat	ccgtcaggca	45840
[2347]	atcgaccgtt	gcagcaatat	ctgggcttca	ctgccgggcg	ctggttatgg	tcagttcgag	45900
[2348]	cataaggctg	acagcctgat	tgcaaaattc	aaagaagcgg	gcggaacggt	cagagagatt	45960
[2349]	gatgtatgag	cagagtcacc	gcgattatct	ccgctctggt	tatctgatac	atcgtctgcc	46020
[2350]	tgtcatgggc	tgtaaatcat	taccgtgata	acgccattac	ctacaaagcc	cagcgcgaca	46080
[2351]	aaaatgccag	agaactgaag	ctggcgaaacg	cggcaattac	tgacatgcag	atgcgtcagc	46140
[2352]	gtgatgttgc	tgcgctcgat	gcaaaataca	cgaaggagtt	agctgatgct	aaagctgaaa	46200
[2353]	atgatgctct	gcgtgatgat	gttgccgctg	gtcgctcgcg	gttgacatac	aaagcagtct	46260
[2354]	gtcagtcagt	gcgtgaagcc	accaccgcct	ccggcgtgga	taatgcagcc	tcccccgac	46320
[2355]	tggcagacac	cgtgaacgg	gattatttca	ccctcagaga	gaggtgatc	actatgcaaa	46380
[2356]	aacaactgga	aggaaccag	aagtatatata	atgagcagtg	cagatagagt	tgcccatatc	46440
[2357]	gatgggcaac	tcatgcaatt	attgtgagca	atacacacgc	gcttcacg	gagtataaat	46500
[2358]	gcctaaagta	ataaaaccga	gcaatccatt	tacgaatgtt	tgctgggttt	ctgttttaac	46560
[2359]	aacattttct	gcgccgccac	aaattttggc	tgcatcgaca	gttttcttct	gcccattcc	46620
[2360]	agaaacgaag	aaatgatggg	tgatggtttc	ctttggtgct	actgctgccg	gtttgttttg	46680
[2361]	aacagtaaac	gtctgttgag	cacatcctgt	aataagcagg	gccagcgcag	tagcgagtag	46740
[2362]	catttttttc	atgggtgttat	tcccgatgct	ttttgaagtt	cgcagaatcg	tatgtgtaga	46800
[2363]	aaattaaaca	aaccctaaac	aatgagttga	aatttcatat	tgtaatatatt	tattaatgta	46860
[2364]	tgtcagggtc	gatgaatcgt	cattgtattc	ccggattaac	tatgtccaca	gccctgacgg	46920
[2365]	ggaacttctc	tgcgggagtg	tccgggaata	attaaaacga	tgcacacagg	gttttagcgcg	46980
[2366]	tacacgtatt	gcattatgcc	aacgccccgg	tgctgacacg	gaagaaaccg	gacgttatga	47040
[2367]	tttagcgtgg	aaagatttgt	gtagtgttct	gaatgctctc	agtaaatagt	aatgaattat	47100
[2368]	caaaggtata	gtaatatctt	ttatgttcat	ggatatttgt	aacccatcgg	aaaactcctg	47160
[2369]	cttttagcaag	attttccctg	tattgctgaa	atgtgatttc	tcttgatttc	aacctatcat	47220
[2370]	aggacgtttc	tataagatgc	gtgtttcttg	agaatttaac	atttacaacc	tttttaagtc	47280
[2371]	cttttattaa	cacggtgtta	tcgttttcta	acacgatgtg	aatattatct	gtggctagat	47340
[2372]	agtaaataata	atgtgagacg	ttgtgacgtt	ttagttcaga	ataaaacaat	tcacagtcta	47400
[2373]	aatcttttcg	cacttgatcg	aatatcttct	taaaaatggc	aacctgagcc	attggtaaaa	47460
[2374]	ccttccatgt	gatacagagg	cgcgtagttt	gcattatcgt	ttttatcggt	tcaatctggt	47520
[2375]	ctgacctcct	tgtgttttgt	tgatgattta	tgtaaaatat	taggaatgtt	ttcacttaat	47580
[2376]	agtattgggt	gcgtaacaaa	gtgcggtcct	gctggcattc	tgaggggaaa	tacaaccgac	47640
[2377]	agatgtatgt	aaggccaacg	tgctcaaatc	ttcatacaga	aagatttgaa	gtaatatattt	47700
[2378]	aaccgctaga	tgaagagcaa	gcgcatggag	cgacaaaatg	aataaagaac	aatctgctga	47760

[2379]	tgatccctcc gtggatctga ttcgtgtaaa aaatatgctt aatagcacca tttctatgag	47820
[2380]	ttaccctgat gttgtaattg catgtataga acataagtg tctctggaag cattcagagc	47880
[2381]	aattgaggca gcgttggtga agcacgataa taatatgaag gattattccc tgggtggtga	47940
[2382]	ctgatcacca taactgctaa tcattcaaac tatttagtct gtgacagagc caacacgcag	48000
[2383]	tctgtcactg tcaggaaagt ggtaaaactg caactcaatt actgcaatgc cctcgtaatt	48060
[2384]	aagtgaattt acaatatcgt cctgttcgga gggaagaacg cgggatgttc attcttcac	48120
[2385]	acttttaatt gatgtatatg ctctcttttc tgacgttagt ctccgacggc aggcttcaat	48180
[2386]	gaccaggtc gagaaattcc cggacccttt ttgctcaaga gcgatgttaa tttgttcaat	48240
[2387]	catttggtta ggaaagcgga tgttgccggg tgttgttctg cgggttctgt tcttcgttga	48300
[2388]	catgaggttg ccccgatttc agtgtcgtg atttgtattg tctgaagttg tttttacgtt	48360
[2389]	aagttgatgc agatcaatta atacgatacc tgcgtcataa ttgattattt gacgtggttt	48420
[2390]	gatggcctcc acgcacgttg tgatatgtag atgataatca ttatcacttt acgggtcctt	48480
[2391]	tccggtgatc cgacaggtta cg	48502
[2392]	<210> 32	
[2393]	<211> 12	
[2394]	<212> DNA	
[2395]	<213> 人工序列	
[2396]	<220>	
[2397]	<223> 实施例1中使用的序列。	
[2398]	<400> 32	
[2399]	tttttttttt tt	12
[2400]	<210> 33	
[2401]	<211> 51	
[2402]	<212> DNA	
[2403]	<213> 人工序列	
[2404]	<220>	
[2405]	<223> 实施例1中使用的序列。	
[2406]	<400> 33	
[2407]	ggttgtttct gttggtgctg atattgcggc gtctgcttgg gtgtttaacc t	51
[2408]	<210> 34	
[2409]	<211> 68	
[2410]	<212> DNA	
[2411]	<213> 人工序列	
[2412]	<220>	
[2413]	<223> 实施例1中使用的序列。	
[2414]	<400> 34	
[2415]	ggttaaacac ccaagcagac gccgcaatat cagcaccaac agaaacaacc tttgaggcga	60
[2416]	gcggtcaa	68
[2417]	<210> 35	

- [2418] <211> 15
- [2419] <212> DNA
- [2420] <213> 人工序列
- [2421] <220>
- [2422] <223> 实施例1中使用的序列。
- [2423] <400> 35
- [2424] ttgaccgctc gcctc 15
- [2425] <210> 36
- [2426] <211> 53
- [2427] <212> DNA
- [2428] <213> 人工序列
- [2429] <220>
- [2430] <223> 实施例2中使用的序列。
- [2431] <400> 36
- [2432] gatctgaagc ggcgcacgaa aaacgcgaaa gcgtttcacg ataatgcgaa aac 53
- [2433] <210> 37
- [2434] <211> 54
- [2435] <212> DNA
- [2436] <213> 人工序列
- [2437] <220>
- [2438] <223> 实施例2中使用的序列。
- [2439] <400> 37
- [2440] ttttgttttc gcatttatcg tgaaacgctt tcgcgttttt cgtgcgccgc ttca 54

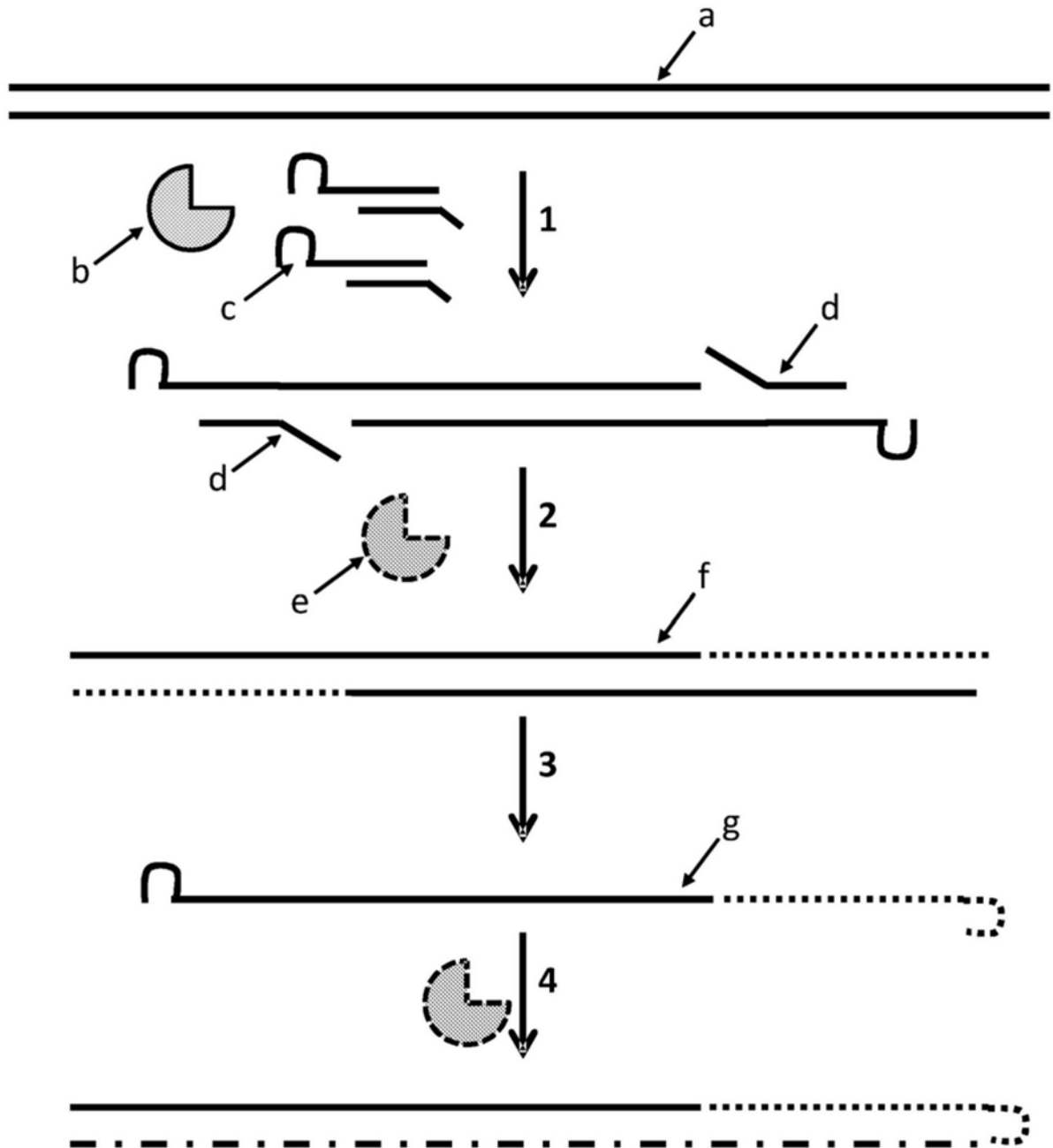


图1

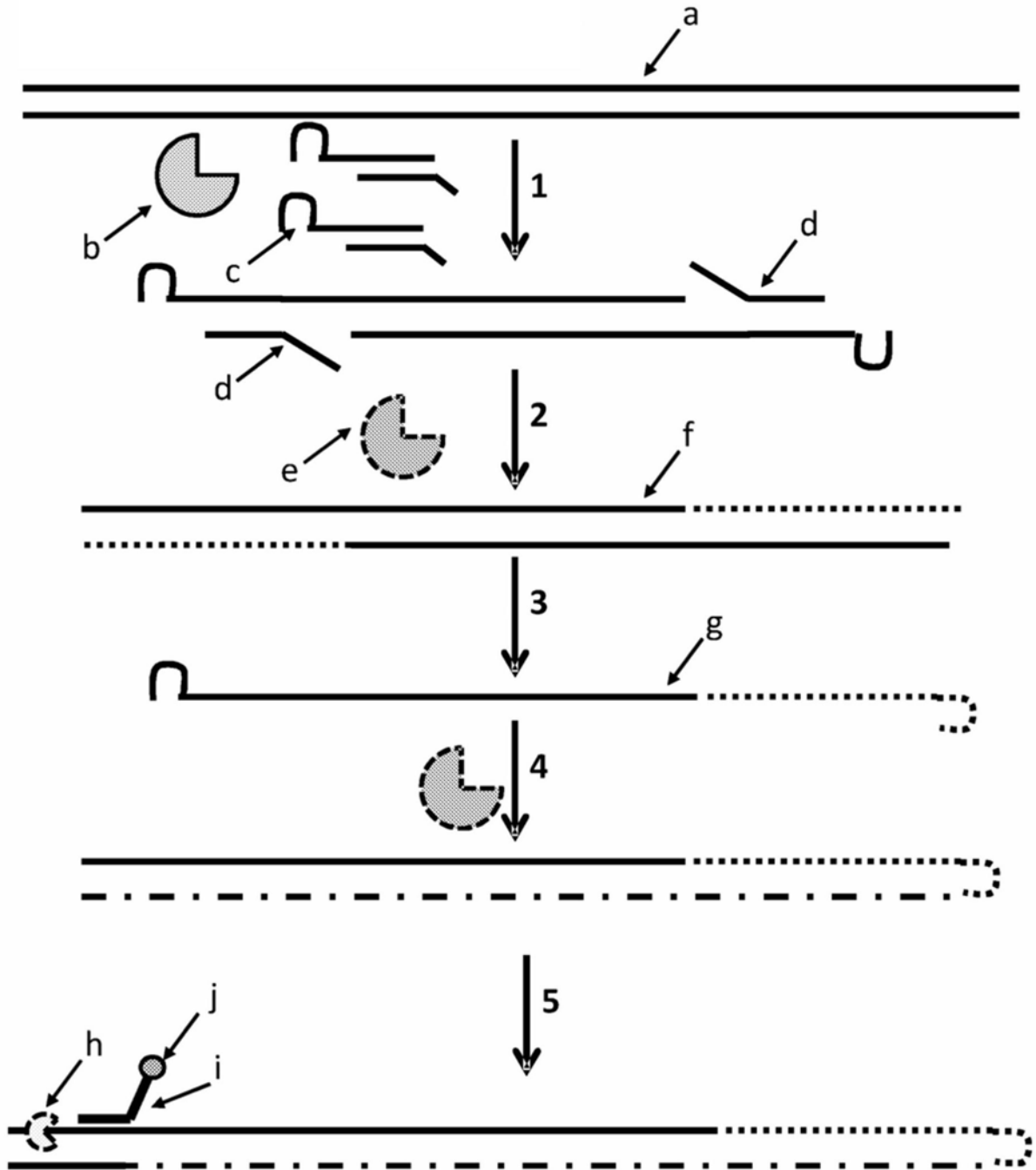


图2

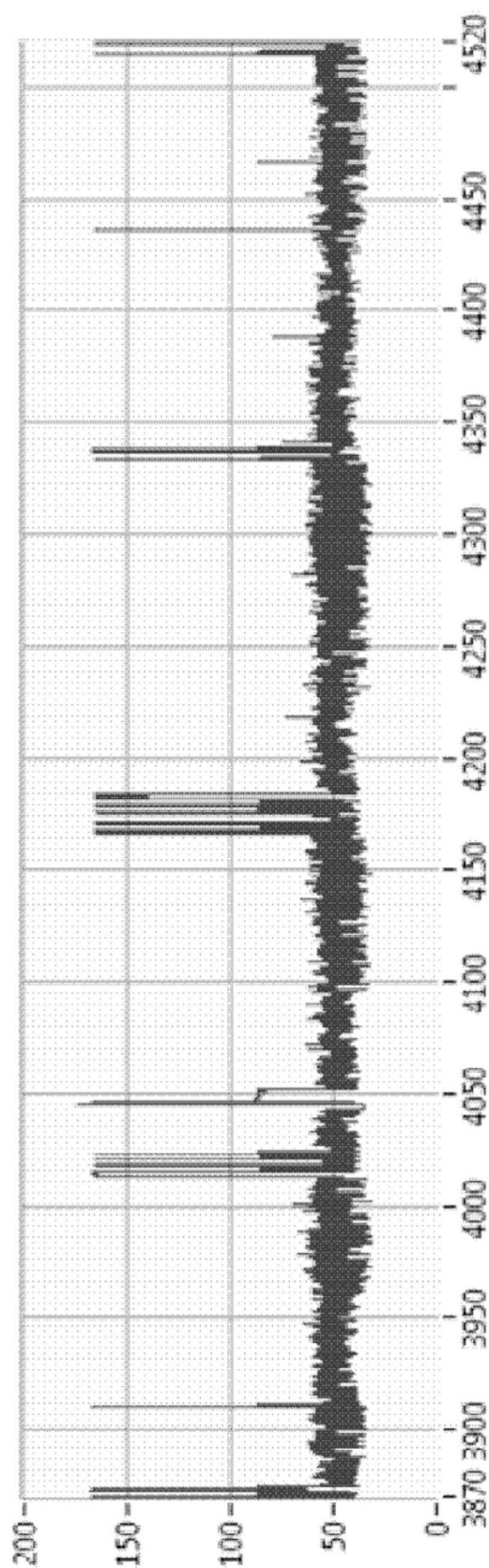


图3

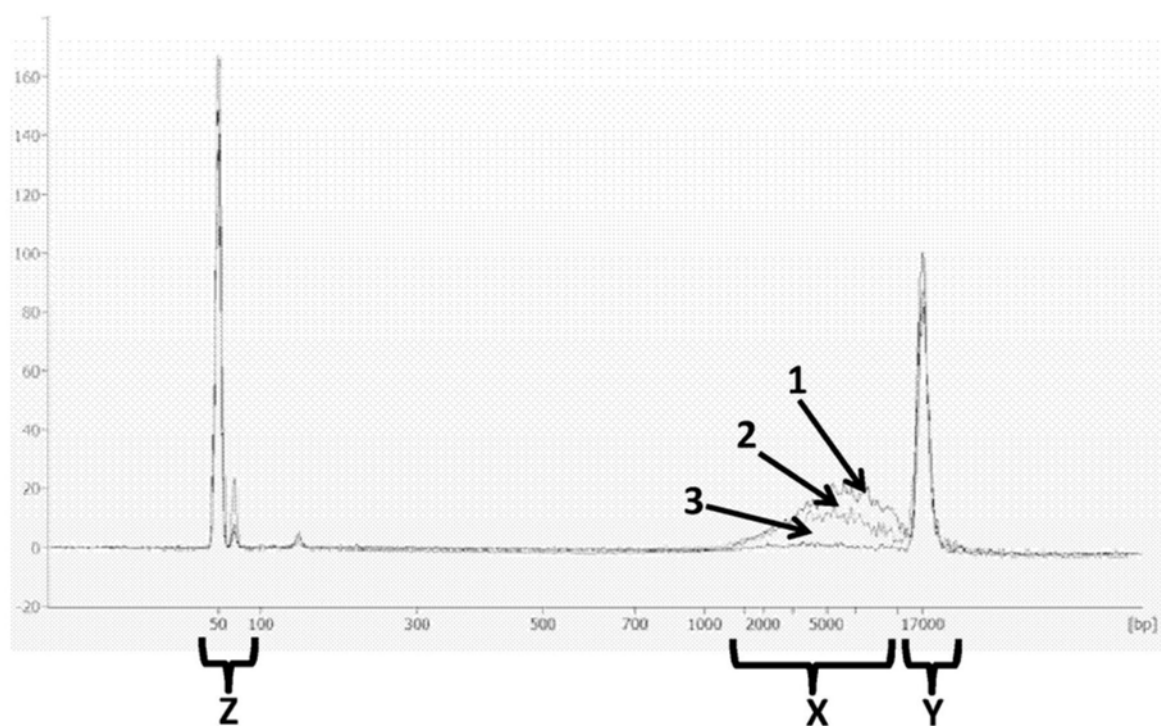


图4

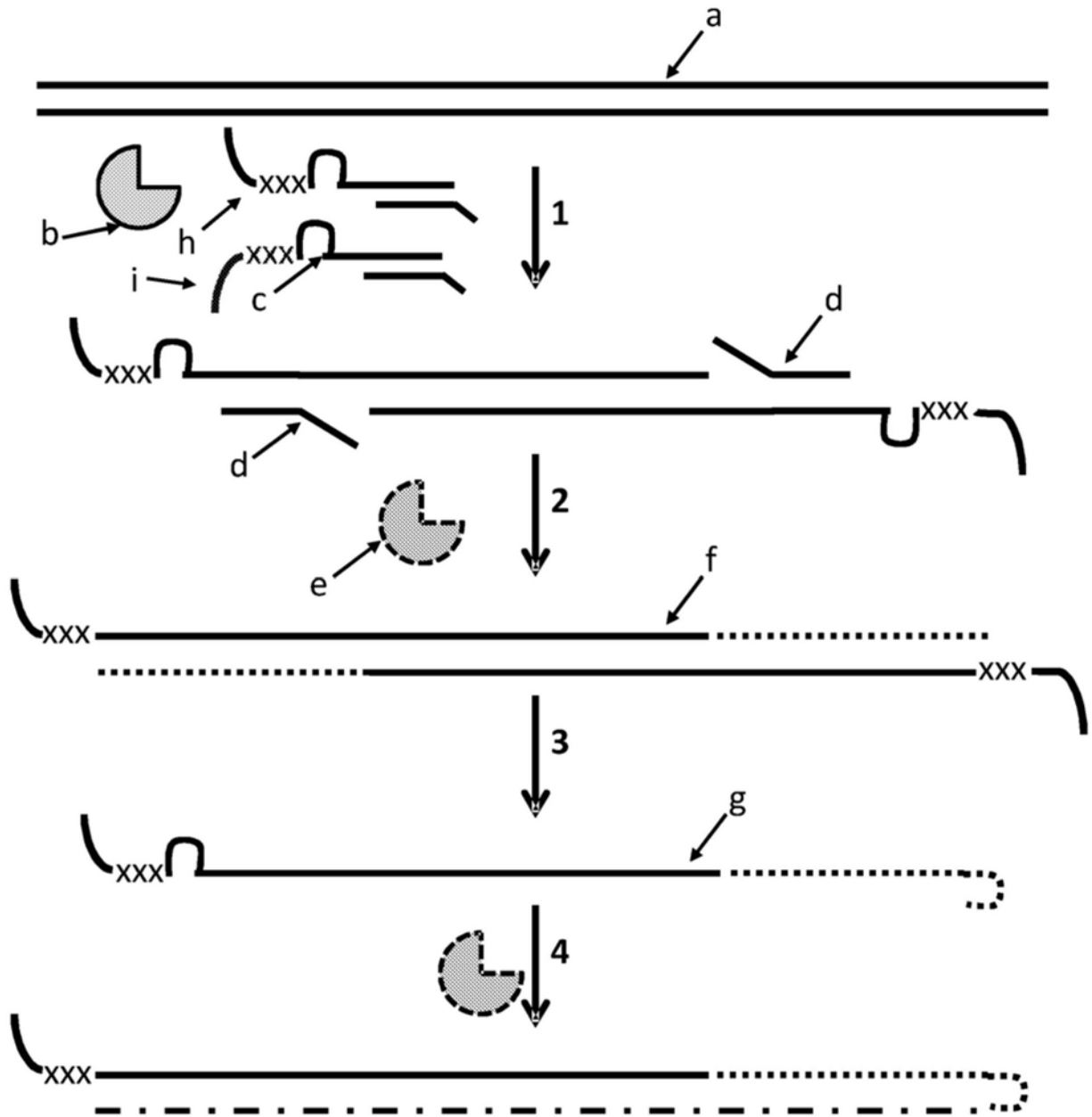


图5

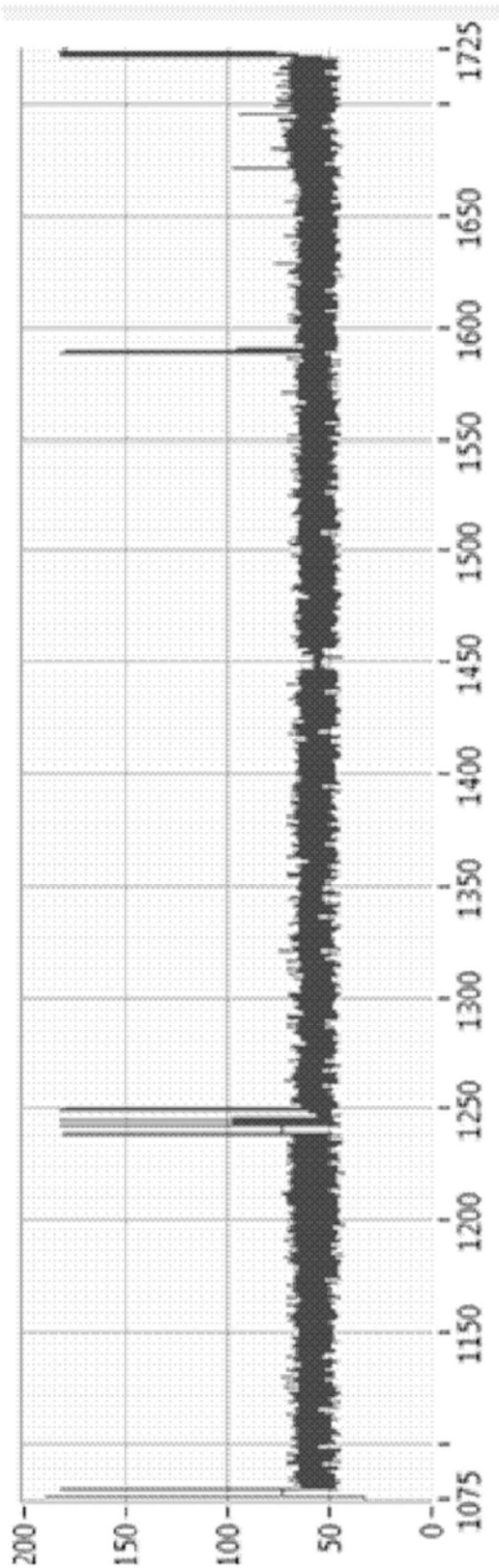


图6

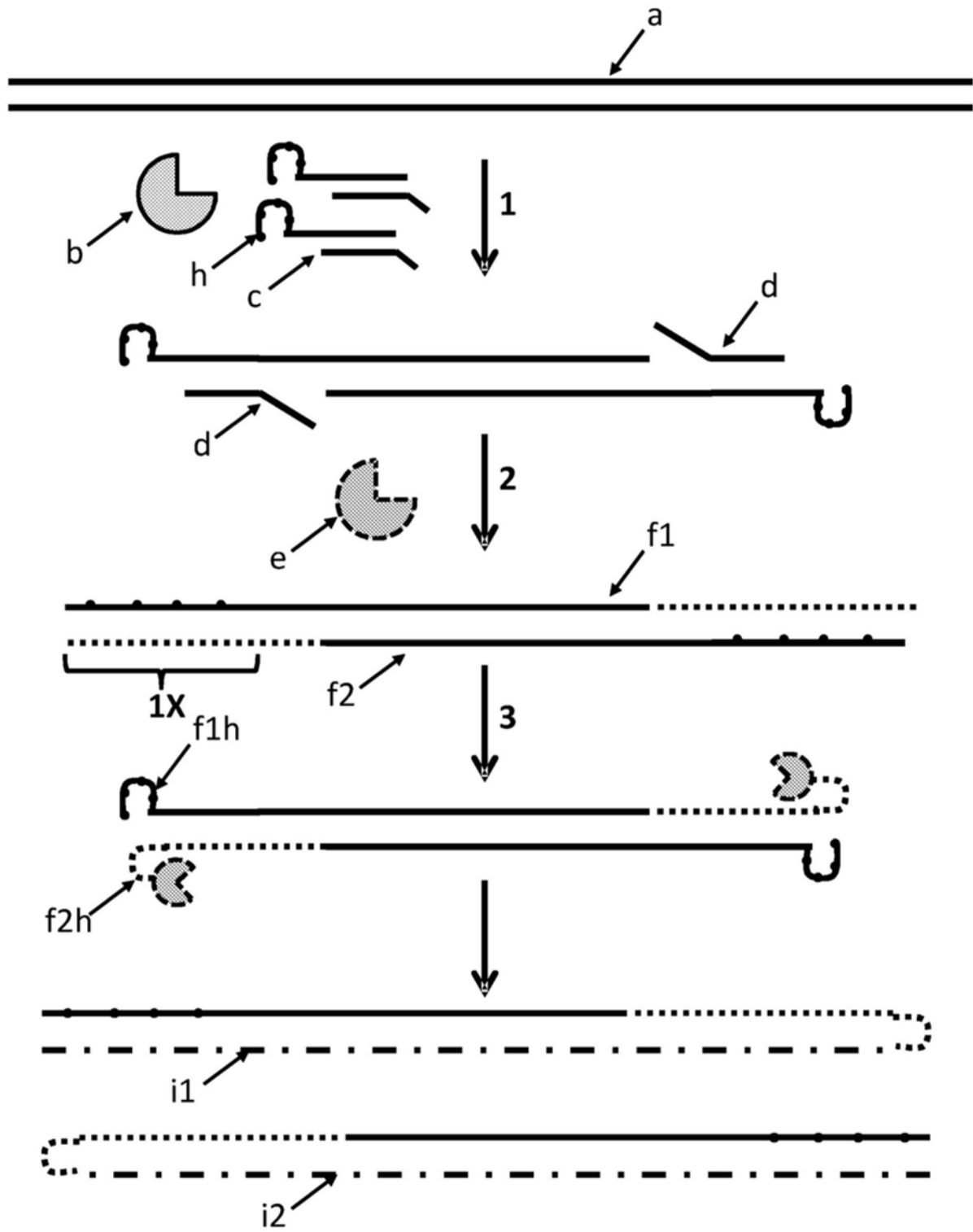


图7

如图 1 中

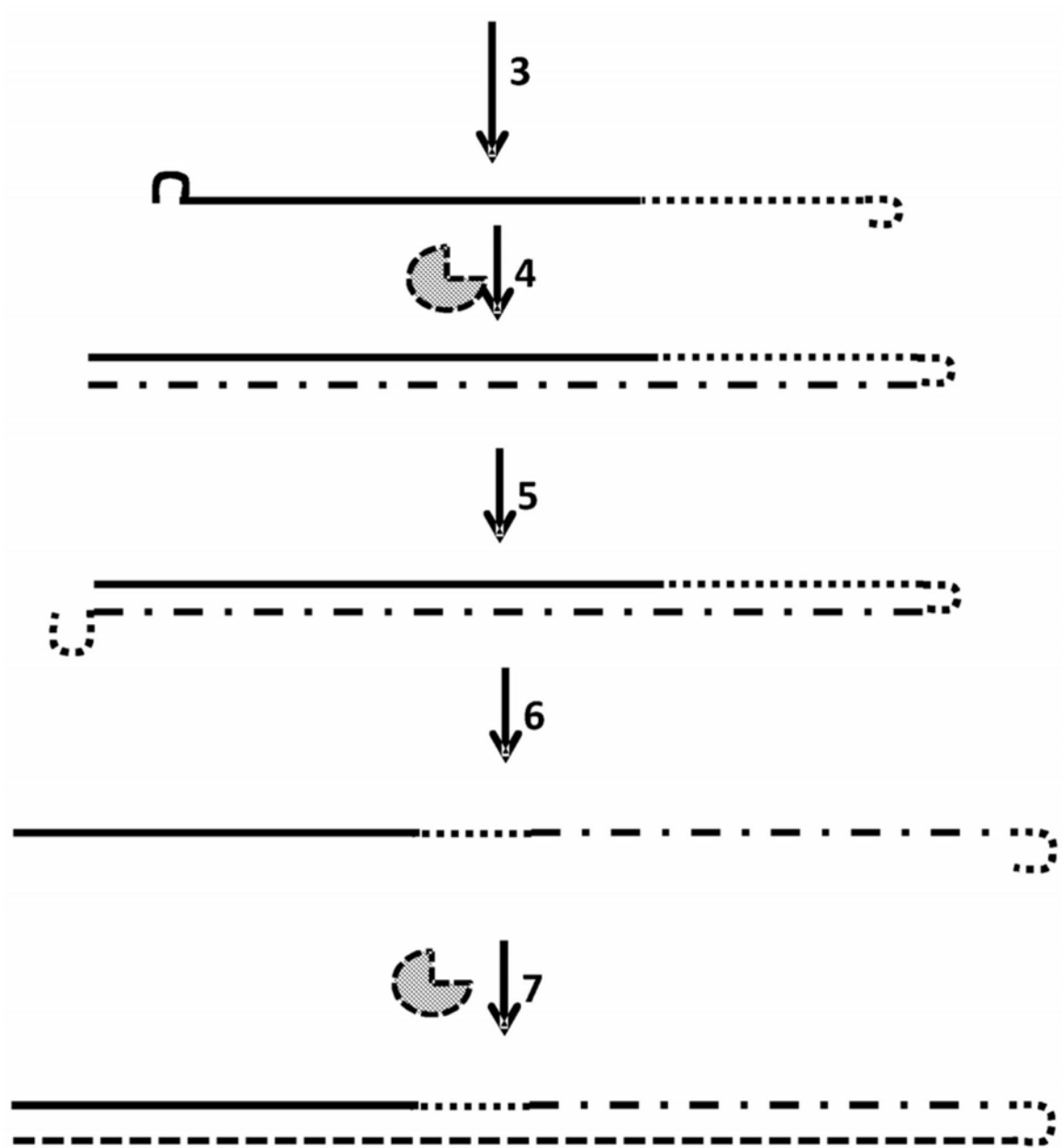


图8