



(12)发明专利

(10)授权公告号 CN 106103741 B

(45)授权公告日 2020.03.13

(21)申请号 201580005650.8

(72)发明人 安迪·赫伦 克莱夫·布朗

(22)申请日 2015.01.22

瑞贝卡·鲍恩 詹姆斯·怀特
丹尼尔·约翰·特纳
约瑟夫·哈格里夫·劳埃德
克里斯多夫·彼得·约德

(65)同一申请的已公布的文献号
申请公布号 CN 106103741 A

(43)申请公布日 2016.11.09

(74)专利代理机构 北京同立钧成知识产权代理
有限公司 11205

(30)优先权数据

代理人 李艳 臧建明

PCT/GB2014/050175 2014.01.22 GB
1406151.9 2014.04.04 GB
1406155.0 2014.04.04 GB
1416197.0 2014.09.12 GB

(51)Int.Cl.

C12Q 1/6806(2018.01)
C12Q 1/6869(2018.01)

(85)PCT国际申请进入国家阶段日

(56)对比文件

2016.07.22

Anna-Helena Saariaho等.

(86)PCT国际申请的申请数据

“Characteristics of MuA transposase-catalyzed processing of model transposon end DNA hairpin substrates”.《Nucleic Acids Research》.2006,第3139-3149页.

PCT/GB2015/050140 2015.01.22

(87)PCT国际申请的公布数据

审查员 郭玉洁

W02015/110813 EN 2015.07.30

权利要求书3页 说明书51页
序列表93页 附图16页

(73)专利权人 牛津纳米孔技术公司

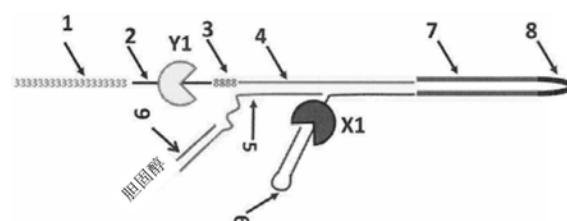
地址 英国牛津郡

(54)发明名称

将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的方法

(57)摘要

本发明涉及将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的新方法。本发明还涉及表征靶多核苷酸的新方法。



1. 一种用于将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的方法,包括:
 - (a) 提供结合到一个或多个装载部分的所述一个或多个多核苷酸结合蛋白,所述一个或多个装载部分包含装载多核苷酸,其中所述一个或多个多核苷酸结合蛋白结合并停滞在所述装载多核苷酸上的一个或多个间隔器处;以及
 - (b) 将所述一个或多个装载部分连接到所述靶多核苷酸,其中所述一个或多个多核苷酸结合蛋白保持结合到所述一个或多个装载部分。
2. 根据权利要求1所述的方法,其中所述方法包括在步骤(a)之前将所述多核苷酸结合蛋白结合到所述一个或多个装载部分。
3. 根据权利要求1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一个或多个多核苷酸处理酶。
4. 根据权利要求3所述的方法,其中所述一个或多个多核苷酸处理酶是一个或多个聚合酶、核酸外切酶、解旋酶、拓扑异构酶,或其组合。
5. 根据权利要求4所述的方法,其中所述一个或多个解旋酶为(a) Hsl308解旋酶、RecD解旋酶、XPD解旋酶或Dda解旋酶;(b) 衍生自(a)中所述任一解旋酶的解旋酶;或(c) (a)和/(b)中所述解旋酶的任意组合。
6. 根据权利要求5所述的方法,其中对所述一个或多个解旋酶进行修饰,以减小所述解旋酶的多核苷酸结合域中开口的大小,所述多核苷酸在至少一个构象下穿过所述开口从解旋酶上解绑。
7. 根据权利要求1所述的方法,其中所述方法涉及连接两个或更多个多核苷酸结合蛋白。
8. 根据权利要求7所述的方法,其中所述两个或更多个多核苷酸结合蛋白除非由所述一个或多个装载部分连接,否则彼此不连接。
9. 根据权利要求7所述的方法,其中所述两个或更多个多核苷酸结合蛋白彼此不同。
10. 根据权利要求1所述的方法,其中所述一个或多个装载部分是合成的。
11. 根据权利要求1所述的方法,其中所述一个或多个装载部分包含所述一个或多个多核苷酸结合蛋白结合的单链多核苷酸。
12. 根据权利要求11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,并且停滞在所述一个或多个装载多核苷酸上的一个或多个间隔器处。
13. 根据权利要求1所述的方法,其中所述靶多核苷酸是双链多核苷酸。
14. 根据权利要求13所述的方法,其中所述一个或多个装载部分中的至少一个是Y衔接体。
15. 根据权利要求14所述的方法,其中所述Y衔接体包含优先螺入孔中的前导序列。
16. 根据权利要求13所述的方法,其中所述一个或多个装载部分中的至少一个是桥连部分。
17. 根据权利要求16所述的方法,其中所述桥连部分是发夹环。
18. 根据权利要求17所述的方法,其中所述发夹环的茎干为少于200个核苷酸对的长度。
19. 根据权利要求1-18中任一项所述的方法,其中所述一个或多个装载部分包含一个或多个能够耦合到膜的锚。

20. 根据权利要求1-18中任一项所述的方法,其中步骤(b)包含使用连接酶将所述一个或多个装载部分连接到所述靶多核苷酸。

21. 根据权利要求20所述的方法,其中所述方法进一步包括(c)从方法条件中去除所述连接酶。

22. 根据权利要求20所述的方法,其中步骤(b)是在ATP不存在或使用 γ -S-ATP(ATP γ S)代替ATP的情况下进行的。

23. 一种表征靶多核苷酸的方法,包括:

(i) 实施前述权利要求中任一项所述的方法;

(ii) 将在步骤(i)中提供的具有所述一个或多个多核苷酸结合蛋白连接的所述靶多核苷酸与跨膜孔接触,使得所述一个或多个多核苷酸结合蛋白控制所述多核苷酸相对于所述孔的移动;以及

(iii) 随着所述多核苷酸相对于所述孔移动,获取一个或多个测量值,其中所述测量值代表所述多核苷酸的一个或多个特征,且由此表征所述靶多核苷酸。

24. 一种制备用于表征的靶多核苷酸的方法,包括:

(i) 实施权利要求1-22中任一项所述的方法,其中所述一个或多个多核苷酸结合蛋白包含一个或多个聚合酶;以及

(ii) 使用步骤(i)中提供的所述靶多核苷酸作为所述一个或多个聚合酶的模板,形成与所述靶多核苷酸互补的一个或多个多核苷酸,且由此制备用于表征的靶多核苷酸。

25. 一种表征靶多核苷酸的方法,包括:

(1) 实施权利要求24所述的方法;

(2) 将所述靶多核苷酸和/或在步骤(1)中产生的与所述靶多核苷酸互补的所述一个或多个多核苷酸与跨膜孔接触,使得所述靶多核苷酸和/或在步骤(1)中产生的与所述靶多核苷酸互补的所述一个或多个多核苷酸相对于所述孔移动;以及

(3) 在所述靶多核苷酸和/或步骤(1)中产生的与所述靶多核苷酸互补的所述一个或多个多核苷酸相对于所述孔移动时,获取一个或多个测量值,其中所述测量值代表所述靶多核苷酸和/或在步骤(1)中产生的与所述靶多核苷酸互补的所述一个或多个多核苷酸的一个或多个特征,并由此表征所述靶多核苷酸。

26. 根据权利要求23或25所述的方法,其中所述孔是跨膜蛋白孔或固态孔。

27. 根据权利要求26所述的方法,其中所述跨膜蛋白孔衍生自溶血素、杀白细胞素、耻垢分枝杆菌孔蛋白A(MspA)、MspB、MspC、MspD、胞溶素、外膜孔蛋白F(OmpF)、外膜孔蛋白G(OmpG)、外膜磷脂酶A、奈瑟球菌属自转运脂蛋白(NalP)和WZA。

28. 根据权利要求27中所述的方法,其中所述跨膜蛋白孔为:

(a) 由8个如SEQ ID NO:2中所示的相同亚基形成的MspA,或(b)MspA的变体,其中基于氨基酸同一性,所述8个亚基中的一个或多个与SEQ ID NO:2在整个序列上具有至少50%同源性且保留了孔活性;或

(c) 由7个如SEQ ID NO:4中所示的相同亚基形成的 α -溶血素,或(d) α -溶血素的变体,其中基于氨基酸同一性,所述7个亚基中的一个或多个与SEQ ID NO:4在整个序列上具有至少50%同源性且保留了孔活性。

29. 根据权利要求28所述的方法,其中所述一个或多个特征选自(i)所述靶多核苷酸的

长度, (ii) 所述靶多核苷酸的同一性, (iii) 所述靶多核苷酸的序列, (iv) 所述靶多核苷酸的二级结构, 以及 (v) 所述靶多核苷酸是否被修饰。

30. 根据权利要求29所述的方法, 其中所述靶多核苷酸通过甲基化、氧化、损伤、用一个或多个蛋白或用一个或多个标记物、标签或间隔器进行修饰。

31. 根据权利要求23-25中任一项所述的方法, 其中所述靶多核苷酸的一个或多个特征通过电测量和/或光学测量来测量。

32. 根据权利要求31所述的方法, 其中所述电测量是电流测量、阻抗测量、风洞测量或场效应晶体管 (FET) 测量。

33. 一种靶多核苷酸, 使用根据权利要求1-22和24中任一项所述的方法进行修饰。

34. 一种装载部分, 包含装载多核苷酸并且具有一个或多个结合的解旋酶, 所述解旋酶停滞在所述装载多核苷酸上的一个或多个间隔器处。

35. 根据权利要求34所述的装载部分, 其中所述装载部分如下述任一项所限定:

所述装载部分是合成的,

所述装载部分包含所述一个或多个解旋酶结合的单链多核苷酸,

所述装载部分是Y衔体,

所述装载部分是桥连部分,

所述装载部分包含一个或多个能够耦合到膜的锚;

和/或其中所述一个或多个解旋酶如权利要求5或6所限定。

36. 根据权利要求35所述的装载部分, 其中所述装载部分是Y衔体, 所述Y衔体包含优先螺入孔中的前导序列。

37. 根据权利要求35所述的装载部分, 其中所述装载部分是桥连部分, 所述桥连部分是发夹环。

38. 根据权利要求37所述的装载部分, 其中所述发夹环的茎干为少于200个核苷酸对的长度。

将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的方法

技术领域

[0001] 本发明涉及将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的新方法。本发明还涉及表征靶多核苷酸的新方法。

背景技术

[0002] 目前需要一种具有广泛的应用范围的快速且廉价的多核苷酸(如DNA或RNA)测序和鉴定技术。现有的技术是缓慢的且昂贵的,这主要由于它们依赖于扩增技术来产生大量的多核苷酸,且需要大量特定的用于信号检测的荧光化学物质。

[0003] 跨膜孔(纳米孔)作为用于聚合物和各种小分子的直接的、电生物传感器,具有很大的潜力。特别是,目前作为一种有潜力的DNA测序技术的纳米孔得到了许多关注。

[0004] 当跨纳米孔施加电势时,在,当分析物如核苷酸在桶(barrel)中短暂停留一段时间时,会产生电流的变化。所述核苷酸的纳米孔检测能产生已知特征和持续时间的电流变化。在链测序方法中,单个多核苷酸链穿过所述孔并实现对核苷酸的鉴定。链测序可包括使用多核苷酸结合蛋白来控制所述多核苷酸穿过所述孔的移动。

发明内容

[0005] 令人惊讶地,发明人已经证明可以预装载一个或多个多核苷酸结合蛋白到一个或多个装载部分,然后连接所述一个或多个装载部分到所述靶多核苷酸。令人惊讶的是,所述一个或多个多核苷酸结合蛋白对所述一个或多个装载部分到多核苷酸的连接没有空间位阻。还令人惊讶的是,连接过程不影响所述一个或多个多核苷酸结合蛋白,且所述一个或多个多核苷酸结合蛋白在连接到靶多核苷酸后仍保留其结合到所述一个或多个装载部分的功能和能力。

[0006] 因此,本发明提供了一种用于将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的方法,包括:

[0007] (a) 提供结合到一个或多个装载部分的一个或多个多核苷酸结合蛋白;和

[0008] (b) 将所述一个或多个装载部分连接到所述靶多核苷酸。

[0009] 本发明还提供了表征靶多核苷酸的方法,包括:

[0010] (a) 实施本发明所述的方法;

[0011] (b) 将在步骤(a)中提供的连接有一个或多个多核苷酸结合蛋白的靶多核苷酸与跨膜孔接触,以使得所述一个或多个多核苷酸结合蛋白控制所述多核苷酸相对于所述孔的移动;

[0012] 和

[0013] (c) 随着所述多核苷酸相对于所述孔移动,获取一个或多个测量值,其中所述测量值代表所述多核苷酸的一个或多个特征,并由此表征所述靶多核苷酸。

[0014] 本发明还提供了一种制备用于表征的靶多核苷酸的方法,包括:

[0015] (a) 实施本发明的方法,其中所述一个或多个多核苷酸结合蛋白包含一个或多个

聚合酶;和

[0016] (a) 使用所述靶多核苷酸作为模板,使得在步骤(a)中提供的连接到所述靶多核苷酸的所述一个或多个聚合酶形成一个或多个多核苷酸,并由此制备用于表征的所述靶多核苷酸。

[0017] 本发明还提供一种表征靶多核苷酸的方法,包括:

[0018] (a) 实施本发明所述的基于聚合酶的方法;

[0019] (b) 将在步骤(a)中产生的所述靶多核苷酸及所述一个或多个多核苷酸与跨膜孔接触,使得所述靶多核苷酸及所述一个或多个多核苷酸相对于所述孔移动;和

[0020] (c) 随着所述靶多核苷酸和所述一个或多个多核苷酸相对于所述孔移动,获取一个或多个测量值,其中所述测量值代表所述多核苷酸的一个或多个特征,并由此表征所述靶多核苷酸。

[0021] 本发明还提供:

[0022] -使用本发明所述的方法修饰的靶多核苷酸;

[0023] -具有一个或多个所结合的多核苷酸结合蛋白的装载部分;和

[0024] -用于将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的试剂盒,包含(a)结合到一个或多个装载部分的所述一个或多个多核苷酸结合蛋白;和(b)连接酶。

附图说明

[0025] 图1示出了使用8mL POROS HQ-10柱将预结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C纯化后的示例性FPLC轨迹。

[0026] 图2示出了使用5mL Histrap HP柱将预结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C纯化后的示例性FPLC轨迹。

[0027] 图3示出了TBE(天然的)PAGE凝胶分析(DNA校准带对应在凝胶的顶部示出的DNA浓度),柱1示出了在只有缓冲液中结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C(显示为条带X),柱2示出了添加TRIS缓冲液后结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C,柱3示出了添加双马来酰亚胺基乙烷后结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C,柱4-6示出了经两次FPLC纯化后结合到DNA发夹衔接体的不同稀释浓度的TrwC Cba-L376C/Q594A/K762C。条带X对应结合到DNA的酶。条带Y对应没有酶结合的DNA。

[0028] 图4示出了SDS PAGE凝胶分析(DNA校准带对应在凝胶的顶部示出的DNA浓度),柱1示出了在只有缓冲液中结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C(显示为条带X),柱2示出了添加TRIS缓冲液后结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C,柱3示出了添加双马来酰亚胺基乙烷后结合到DNA发夹衔接体的TrwC Cba-L376C/Q594A/K762C,柱4-6示出了经两次FPLC纯化后结合到DNA发夹衔接体的不同稀释浓度的TrwC Cba-L376C/Q594A/K762C。条带Y对应DNA上没有闭合的酶。条带X对应由双马来酰亚胺基乙烷连接体结合到DNA上的酶。

[0029] 图5示出了TBE(天然的)PAGE凝胶分析(DNA校准带对应在凝胶的顶部显示的DNA浓度),柱1示出了在添加TMAD前结合到DNA发夹衔接体的T4 Dda-E94C/C109A/C136A/A360C。柱2示出了添加KC1和ATP之后结合到DNA发夹衔接体的T4 Dda-E94C/C109A/C136A/A360C,柱3示出了SPRI纯化后结合到DNA发夹衔接体的T4 Dda-E94C/C109A/C136A/A360C。条带2:1示出了

两个结合的酶,条带1:1示出了一个结合的酶。DNA条带对应单独的DNA。

[0030] 图6示出了当解旋酶T4 Dd-E94C/C109A/C136A/A360C和TrwC Cba-L376C/Q594A/K762C控制DNA构建体(参见上文电流轨迹)移位穿过MspA纳米孔时的示例电流轨迹。x轴对应时间(秒)以及y轴对应电流(pA)。所述轨迹示出了在所述两种解旋酶控制下单独DNA链移动穿过纳米孔,标记的区域1和2对应DNA构建体的区域1和2的移位。所述轨迹示出了在解旋酶T4 Dda-E94C/C109A/C136A/A360C和TrwC Cba-L376C/Q594A/K762C二者控制下构建体Y移位穿过纳米孔时观察到的电流轨迹。箭头标记的3示出了当发夹衔体中的间隔器移位穿过纳米孔时电流的尖峰。所述发夹衔体和Y衔体间隔器在电流轨迹上方的DNA构建体的图片中以x示出。

[0031] 图7示出了Agilent生物分析仪轨迹,表明酶预结合到MuA Y衔体。A)未预结合衔体。B)预结合衔体。

[0032] 图8示出了Agilent生物分析仪轨迹,表明当酶预结合到MuA Y衔体时,没有见到对靶DNA的副作用。A)未预结合衔体。B)预结合衔体。

[0033] 图9示出了当由解旋酶控制DNA移动穿过使用由MuA标签技术(tagmentation)产生的DNA制备的纳米孔时的示例电流轨迹。

[0034] 图10示出了具有预结合的E1(A片段)及E2(END片段)的A片段(见分图A)和END片段(见分图B)的装载部分。这两个装载部分接合到实施例6中的基因组DNA。所述A片段具有标记为1的39个SpC3间隔器的区域。SEQ ID NO:32对应标记为2的区域,其为E1结合的区域。标记为3的区域对应4个iSp18间隔器。标记为4的区域对应SEQ ID NO:33。标记为5的区域对应SEQ ID NO:34。标记为6的区域对应一个iBNA-meC、两个iBNA-A和两个iBNA-meC。标记为7的区域对应SEQ ID NO:35。END片段中的标记为8的区域对应SEQ ID NO:36。标记为9的区域对应SEQ ID NO:37。标记为11区域对应SEQ ID NO:38。

[0035] 图11示出了当解旋酶T4 Dda-(H82R/E94C/A360C)(具有突变H82R/E94C/A360C的SEQ ID NO:24,E1)和T4 Dda-E94C/C109A/C136A/A360C(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:24,E2)控制样本4中的DNA构建体图(图顶部示出)移位穿过MspA纳米孔时的示例性图。x轴对应移动指数,而y轴对应电流(pA)。对于移动穿过所述孔的每条DNA链,测量作为时间的函数的电流。移动的DNA导致测得的电流水平的逐步变化。将观察到的电流水平拟合以得到每一步的平均电流并为所述电流水平分配递增的移动指数点数。相对于移动指数的平均电流因此近似于原有的电流信号并且用于表征移位的DNA。该图示出了在所述两种酶的控制下单独DNA链移动穿过纳米孔,标记的区域1和2对应DNA构建体的区域1和2的移位。该轨迹示出了在解旋酶T4 Dda-(H82R/E94C/A360C)和T4 Dda-E94C/C109A/C136A/A360C二者控制下当构建体移位穿过孔时所观察到的移动指数。箭头标记的3示出了在发夹衔体中的间隔器移位穿过纳米孔时的电流中的尖峰。所述的发夹衔体和Y衔体间隔器在所述轨迹上面的DNA构建体图片中以x示出。

[0036] 图12示出了解旋酶前导复合体(见分图A)、聚合酶链复合体(见分图B)和具有预结合的聚合酶(标记为X1,Phi29-A411C/Q560C(具有突变A411C/Q560C的SEQ ID NO:9))和解旋酶(标记为Y1,T4 Dda-E94C/C109A/C136A/A360C(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:24))二者的解旋酶/聚合酶前导复合体的最终装载部分。在实施例7中,该最终装载部分被接合到3.6kb DNA链(SEQ ID NO:46)。标记为1的区域对应30个SpC3间隔器。SEQ

ID NO:27对应标记为2的区域,其为T4 Dda-E94C/C109A/C136A/A360C所结合的区域。标记为3的区域对应4个iSp18间隔器。标记为4的区域对应SEQ ID NO:28。标记为5的区域对应SEQ ID NO:43。标记为6的区域对应SEQ ID NO:44,其3'末端连接到4个iSpC3间隔器,所述间隔器在相对端连接到SEQ ID NO:45的5'末端。Phi29-A411C/Q560C结合到区域6。

[0037] 图13示出了在实施例7中在连接步骤之后并且聚合步骤之前产生的DNA构建体。标记为1的区域对应30个SpC3间隔器。SEQ ID NO:27对应标记为2的区域,其为T4 Dda-E94C/C109A/C136A/A360C所结合的区域。标记为3的区域对应4个iSp18间隔器。标记为4的区域对应SEQ ID NO:28。标记为5的区域对应SEQ ID NO:43。标记为6的区域对应SEQ ID NO:44,其3'末端连接到4个iSpC3间隔器,所述间隔器在相对端连接到SEQ ID NO:45的5'末端。Phi29-A411C/Q560C结合到区域6。标记为7的区域对应SEQ ID NO:46。标记为8的区域对应SEQ ID NO:47。标记为9的区域对应SEQ ID NO:31。

[0038] 图14A是包含标记区域1和2、含间隔器x(3)的发夹、与区域2和1互补的标记区域4和5以及含间隔器(x)和解旋酶T4 Dda-E94C/C109A/C136A/A360C(Y1)的Y衔接体的构建体的图。图14B示出了当解旋酶T4 Dda-E94C/C109A/C136A/A360C(在图14A中标记为Y1)控制样本5中的DNA构建体(在图14A中示出)移位穿过MspA纳米孔时的示例性图。x轴对应移动指数,而y轴对应电流(pA)。对于移动穿过所述孔的每条DNA链,测量作为时间函数的电流。移动的DNA导致测得的电流水平的逐步变化。将观察到的电流水平拟合以得到每一步的平均电流并为所述电流水平分配递增的移动指数点数。相对于移动指数的平均电流因此近似于原有电流信号并且用于表征转移的DNA。该图示出了在解旋酶的控制下DNA单链移动穿过纳米孔,标记区域1和2对应原始的3.6kb DNA构建体(SEQ ID NO:46)的区域1和2的移位。标记区域4和5对应使用Phi29-A411C/Q560C经聚合产生的区域1和2的互补链。轨迹示出了在T4 Dda-E94C/C109A/C136A/A360C的控制下当构建体移位穿过孔时观察到的移动指数。箭头标记的3示出了在最终构建体的发夹中的间隔器(在图14A中示为x和标记为3)移位穿过纳米孔时的电流中的尖峰。所述的发夹衔接体和Y衔接体间隔器在图14A中以x示出。

[0039] 图15示出了在200V运行60分钟然后使用SYBR染色的4-20%的TBE PAGE。在凝胶上运行的各个样品为400nM(5μL)。列1对应100bp梯子(条带对应的碱基对的数目沿凝胶的侧部示出)。列2对应在图12A中示出的没有结合解旋酶的解旋酶前导复合体。列3对应在图12A中示出的结合了解旋酶的解旋酶前导复合体。列4对应在图12B中示出的没有结合聚合酶的聚合酶链复合体。列5对应在图12B中示出的结合了聚合酶的聚合酶链复合体。条带A对应SEQ ID NO:43。条带B对应在3'端连接到SEQ ID NO:27的5'末端的DNA链X=30个iSpC3间隔器,SEQ ID NO:27在3'末端连接到4个iSp18间隔器,所述4个iSp18间隔器在相对端连接到SEQ ID NO:28的5'末端。条带C对应结合到DNA链X的解旋酶(T4 Dda-E94C/C109A/C136A/A360C)。条带D对应没有结合酶的聚合酶链复合体(SEQ ID NO:44,其3'末端连接到4个iSpC3间隔器,所述iSpC3间隔器在相对端连接到SEQ ID NO:45的5'末端)。条带E对应结合到所述聚合酶链复合体的Phi29-A411C/Q560C。

[0040] 图16的A)中示出了在连接前没有将聚合酶(标记为x)预结合到装载部分的双链靶多核苷酸的衔接体连接方法和聚合方法的实例,B)中示出了用预结合的解旋酶(标记为y)和预结合的聚合酶(标记为x)两者连接装载部分,然后聚合所述双链靶多核苷酸的本发明方法的实例。步骤1A显示了将衔接体连接到双链靶多核苷酸的任一端。步骤2A示出了聚合酶(标

记为x)的结合。步骤3A示出了双链靶多核苷酸的聚合,使用所述靶作为模板形成的多核苷酸用虚线示出。步骤4A示出了新的双链构建体的末端修复、构建体的A尾以及用预结合的解旋酶与装载部分的连接(标记为y)。步骤1B示出了含有预结合的解旋酶(标记为y)和预结合的聚合酶(标记为x)二者的装载部分的连接。步骤2B示出了双链靶多核苷酸的聚合。B中不需要进一步的步骤,因此该方法包括比A中示出的方法明显更少的步骤。

[0041] 图17示出了包含预结合的解旋酶(标记为y)和预结合的聚合酶(标记为x)的装载部分,该装载部分随后被连接(步骤1)到双链靶多核苷酸的两条链(一条链标记为T1用于模板,另一条标记为C1用于互补),该双链靶多核苷酸在一端由桥连部分衔接体(标记为z)连接。在装载部分中,解旋酶从与聚合酶结合的链结合到相对的链。所述与聚合酶结合的链含有3'发夹环(标记为v)。在本实施方案中,聚合酶将产生双链构建体(步骤2),其中所述构建体的两条链在一端由桥连部分(标记为v)连接,而且所述构建体的每条链包含靶多核苷酸的一条链(示为实线,且标记为T1和C1)和由聚合酶形成的互补多核苷酸(示为虚线,且标记为T2和C2)。

[0042] 图18示出了包含预结合的聚合酶(标记为x)和预结合的解旋酶(标记为y)的装载部分,其随后连接(步骤1)到双链靶多核苷酸的两条链的每一端,所述两条链包括靶链(T1)和互补链(C1)。所述装载部分不包括桥连部分,所以T1和C1的任一端没有由桥连部分连接。在装载部分中,解旋酶从与聚合酶结合的链结合到相对的链。所述与聚合酶结合的链含有3'发夹环(标记为v)。在该图中,双链靶多核苷酸末端的预结合的聚合酶、预结合的解旋酶和两个装载部分的发夹环分别标记为x1和x2,y1和y2,v1和v2。在本实施方案中,聚合酶将产生两个双链构建体(步骤2),其中每个构建体的两条链在一端由桥连部分连接(标记为v1或v2),并且每个构建体包含靶多核苷酸的一条链(示为实线)和由聚合酶形成的互补多核苷酸(示为虚线)。

[0043] 图19示出了包含预结合的聚合酶(标记为x)和预结合的解旋酶(标记为y)的装载部分(标记为v的桥连部分衔接体和标记为w的杂合前导体),其随后连接(步骤1)到包括靶链(T1)和互补链(C1)的双链靶多核苷酸的每一端,以使T1和C1在两端由装载部分(在一个装载部分中,桥连部分、杂合前导体、预结合的聚合酶和预结合的解旋酶分别标记为v1、w1、x1和y1。在另一个装载部分中,桥连部分、杂合前导体、预结合的聚合酶和预结合的解旋酶分别标记为v2、w2、x2和y2)连接以形成环形构建体。由每个多核苷酸前导体(标记为w1和w2)杂合到所述桥连部分衔接体而形成的双链段形成了用于结合聚合酶的两个引物位点。当聚合酶扩展启动时(步骤2)(例如,通过添加核苷酸和辅因子),用为原始靶DNA的互补拷贝的DNA生成两个构建体,并且取决于聚合酶围绕该环形构建体进行的程度,所述两个构建体将含有T1和C1的多个拷贝段。

[0044] 序列表的说明

[0045] SEQ ID NO:1示出了编码MS-B1突变体MspA单体的密码子优化的多核苷酸序列。该突变缺少信号序列,并包括下列突变:D90N、D91N、D93N、D118R、D134R和E139K。

[0046] 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。该突变缺少信号序列,并包括下列突变:D90N、D91N、D93N、D118R、D134R和E139K。

[0047] SEQ ID NO:3示出了编码α-溶血素-E111N/K147N(α-HL-NN; Stoddart等人, PNAS, 2009; 106 (19) : 7702-7707)的一个单体的多核苷酸序列。

- [0048] SEQ ID NO:4显示了 α -HL-NN的一个单体的氨基酸序列。
- [0049] SEQ ID NO:5至SEQ ID NO:7显示了MspB,C和D的氨基酸序列。
- [0050] SEQ ID NO:8显示了编码Phi29DNA聚合酶的多核苷酸序列。
- [0051] SEQ ID NO:9显示了Phi29DNA聚合酶的氨基酸序列。
- [0052] SEQ ID NO:10显示了由大肠杆菌的sbcB基因衍生的密码子优化的多核苷酸序列。它编码来自大肠杆菌的核酸外切酶I (EcoExo I)。
- [0053] SEQ ID NO:11显示了来自大肠杆菌的核酸外切酶I酶 (EcoExo I) 的氨基酸序列。
- [0054] SEQ ID NO:12显示了从大肠杆菌的xthA基因衍生的密码子优化的多核苷酸序列。其编码来自大肠杆菌的核酸外切酶III。
- [0055] SEQ ID NO:13显示了来自大肠杆菌的核酸外切酶III的氨基酸序列。这种酶在3'-5'的方向对双链DNA (dsDNA) 的一条链的5'单磷酸核苷进行分配消化。链上酶的启动需要约4个核苷酸的5'突出。
- [0056] SEQ ID NO:14显示了由嗜热栖热菌 (*T. thermophilus*) 的recJ基因衍生的密码子优化的多核苷酸序列。其编码来自嗜热栖热菌的RecJ酶 (TthRecJ-cd)。
- [0057] SEQ ID NO:15显示了来自嗜热栖热菌的RecJ酶 (TthRecJ-cd) 的氨基酸序列。该酶在5'-3'的方向对ssDNA的5'单磷酸核苷进行进行性 (processive) 消化。链上酶的启动需要至少4个核苷酸。
- [0058] SEQ ID NO:16显示了由细菌噬菌体 λ (redX) 核酸外切酶基因衍生的密码子优化的多核苷酸序列。其编码细菌噬菌体 λ 核酸外切酶。
- [0059] SEQ ID NO:17显示了细菌噬菌体 λ 核酸外切酶的氨基酸序列。该序列是装配成三聚体的三个相同的亚基之一。该酶在5'-3'方向上对dsDNA的一条链的核苷酸进行高度的进行性消化, (<http://www.neb.com/nebcomm/products/productM0262.asp>)。链上酶的启动优选需要约4个具有5'磷酸的核苷酸的5'突出。
- [0060] SEQ ID NO:18显示了He1308 Mbu的氨基酸序列。
- [0061] SEQ ID NO:19显示了He1308 Csy的氨基酸序列。
- [0062] SEQ ID NO:20显示了He1308 Tga的氨基酸序列。
- [0063] SEQ ID NO:21显示了He1308 Mhu的氨基酸序列。
- [0064] SEQ ID NO:22显示了Tral Eco的氨基酸序列。
- [0065] SEQ ID NO:23显示了XPD Mbu的氨基酸序列。
- [0066] SEQ ID NO:24显示了Dda 1993的氨基酸序列。
- [0067] SEQ ID NO:25显示了Trwc Cba的氨基酸序列。
- [0068] SEQ ID NO:26显示了在实施例1中使用的多核苷酸序列。此序列具有5'磷酸。
- [0069] SEQ ID NO:27显示了在实施例2和7中使用的多核苷酸序列。
- [0070] SEQ ID NO:28显示了在实施例2和7中使用的多核苷酸序列。
- [0071] SEQ ID NO:29显示了在实施例1中使用的多核苷酸序列。此序列具有5'磷酸。
- [0072] SEQ ID NO:30显示了在实施例1中使用的多核苷酸序列。
- [0073] SEQ ID NO:31显示了在实施例7中使用的多核苷酸序列。此序列具有5'胆固醇TEG。
- [0074] SEQ ID NO:32显示了在实施例6中使用的多核苷酸序列。连接到SEQ ID NO:32的5'末端的是39个SpC3间隔器。连接到SEQ ID NO:32的3'末端的是在相对端连接到SEQ ID

NO:33的5'末端的4个iSP18间隔器。

[0075] SEQ ID NO:33显示了在实施例6中使用的多核苷酸序列。连接到SEQ ID NO:33的5'末端的是在相对端连接到SEQ ID NO:32的3'末端的4个iSP18间隔器。

[0076] SEQ ID NO:34显示了在实施例6中使用的多核苷酸序列。连接到3'末端的是一个iBNA-meC、两个iBNA-A和两个iBNA-meC，所述两个iBNA-meC在相对端连接到3个iSp18间隔器，所述3个iSp18间隔器连接到SEQ ID NO:35的5'末端。

[0077] SEQ ID NO:35显示了在实施例6中使用的多核苷酸序列。连接到5'末端的是3个iSp18间隔器，所述间隔器在相对端连接两个iBNA-meC、两个iBNA-A和一个iBNA-meC，所述iBNA-meC在相对端连接到连接SEQ ID NO:34的3'末端。

[0078] SEQ ID NO:36显示了在实施例6中使用的多核苷酸序列。连接到SEQ ID NO:36的5'末端的是磷酸。连接到SEQ ID NO:36的3'末端的是四个在相对端连接到SEQ ID NO:37的5'末端的iSp18间隔器。

[0079] SEQ ID NO:37显示了在实施例6中使用的多核苷酸序列。连接到SEQ ID NO:37的5'末端的是在相对端连接到SEQ ID NO:36的3'末端的4个iSp18间隔器。

[0080] SEQ ID NO:38显示了在实施例6中使用的多核苷酸序列。

[0081] SEQ ID NO:39显示了在实施例6中使用的多核苷酸序列。连接于SEQ ID NO:39的5'末端的是磷酸。连接到SEQ ID NO:39的3'末端的是在相对端连接到SEQ ID NO:40的5'末端的四个iSpC3间隔器。

[0082] SEQ ID NO:40显示了在实施例6中使用的多核苷酸序列。连接到SEQ ID NO:40的5'末端的是在相对端连接到SEQ ID NO:39的3'末端的四个iSpC3间隔器。

[0083] SEQ ID NO:41显示了在实施例6中使用的多核苷酸序列。SEQ ID NO:41是与SEQ ID NO:42互补的。

[0084] SEQ ID NO:42显示了在实施例6中使用的多核苷酸序列。SEQ ID NO:42是与SEQ ID NO:41互补的。

[0085] SEQ ID NO:43显示了在实施例7中使用的多核苷酸序列。

[0086] SEQ ID NO:44显示了在实施例7中使用的多核苷酸序列。该序列具有5'磷酸。SEQ ID NO:44的3'末端连接到4个iSpC3间隔器，所述4个iSpC3间隔器在相对端连接到SEQ ID NO:45的5'末端。

[0087] SEQ ID NO:45显示了在实施例7中使用的多核苷酸序列。该序列在其5'末端连接到四个iSpC3间隔器，所述四个iSpC3间隔器在相对端连接到SEQ ID NO:44的3'末端。

[0088] SEQ ID NO:46显示了在实施例7中使用的多核苷酸序列。

[0089] SEQ ID NO:47显示了在实施例7中使用的多核苷酸序列。该序列具有5'磷酸和在该序列的3'末端的基于胸腺嘧啶的硫代磷酸酯碱基。

具体实施方式

[0090] 应该理解的是，所公开的产品和方法的不同应用可以适用于本领域中的特定需要。也应理解的是，这里使用的术语仅用于描述本发明的具体实施方案，并且不意在限制。

[0091] 此外，除非内容另外明确指出，否则用于本说明书和所附权利要求书中的单数形式“一”，“一个”，和“所述”包括复数指代。因此，例如，涉及“多核苷酸”时包括两个或多个多

核苷酸,涉及“酶”时包括两个或多个酶,涉及“解旋酶”时包括两个或多个解旋酶,涉及“分子制动器 (molecular brake)”时指两个或多个分子制动器,涉及“跨膜孔”时包括两个或多个孔,等。

[0092] 文本无论在上文还是下文中引用的所有出版物、专利和专利申请,以全文参考的方式引入本文。

[0093] 本发明的方法

[0094] 本发明提供一种将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的方法。提供了结合到(或连接到)一个或多个装载部分的所述一个或多个多核苷酸结合蛋白。所述一个或多个装载部分被连接到所述的靶多核苷酸。这将所述一个或多个多核苷酸结合蛋白连接到所述靶多核苷酸。一旦所述一个或多个多核苷酸结合蛋白以此种方式连接,它们可被用于控制所述靶多核苷酸穿过跨膜孔的移动或使用所述靶多核苷酸作为模板形成一个或多个多核苷酸(例如修饰所述靶多核苷酸)。这允许对待表征的靶多核苷酸在下文做更具体的讨述。

[0095] 本发明具有多种优点。

[0096] i. 将多核苷酸结合蛋白预装载到装载部分上加速了样品制备过程,并且意味着使用更少的试管。

[0097] ii. 尽可能接近100%的所得靶多核苷酸的实例具有连接的多核苷酸结合蛋白。这可以改善后续的必需过程的产率。

[0098] iii. 如果客户不将一个或多个多核苷酸结合蛋白装载到一个或多个装载部分上,客户错误减少。

[0099] iv. 连接后无过量的多核苷酸结合蛋白(其可阻塞所述孔或具有一些其他不需要的活动)残留。

[0100] v. 提高了所述一个或多个多核苷酸结合蛋白的稳定性。结合到多核苷酸的蛋白可能更稳定。

[0101] vi. 过量的装载部分可作为正确地建立系统的对照(例如在缓冲液中的ATP等)。

[0102] vii. 用户可控制一个或多个多核苷酸结合蛋白的顺序,以使它们以正确的序列装载在一个装载部分上。

[0103] viii. 使用者可控制哪个多核苷酸结合蛋白连接到哪个装载部分,例如Y衔接对桥连部分。

[0104] ix. 所述一个或多个多核苷酸结合蛋白可用于纯化所述一个或多个装载部分。

[0105] x. 用户可以对连接到不同的装载部分的多核苷酸结合蛋白做出不同的修饰。

[0106] xi. 不同的多核苷酸结合蛋白可以优选不同的结合条件,并因此它可有助于能够预装载它们到不同的装载部分上(而不是一起装载在靶多核苷酸上)。

[0107] xii. 如果一个或多个多核苷酸结合蛋白被预装载,连接后应该没有任何游离的多核苷酸结合蛋白存在,并因此所述蛋白可用于纯化构建体。

[0108] xiii. 本发明最小化的使用所述的一个或多个多核苷酸结合蛋白,因此存在较少浪费。

[0109] xiv. 用户可以控制所述一个或多个多核苷酸结合蛋白相对于靶多核苷酸结合的位置或不相对于所述靶多核苷酸结合。通过预装载所述一个或多个多核苷酸结合蛋白到一个或多个装载部分上,所述一个或多个多核苷酸结合蛋白不直接结合到所述靶多核苷酸。

[0110] xv. 本发明还提高了序列信息的产率, 例如当聚合酶复制靶链时。产率可通过低效的蛋白结合限制, 且还缺乏持续合成能力。预装载克服了装载低效性, 且产生了克服持续合成问题的闭合复合体。

[0111] 本发明还提供一种表征靶多核苷酸的方法。一旦所述一个或多个多核苷酸结合蛋白已经被装载到靶多核苷酸上, 它们可与跨膜孔接触, 以使得所述一个或多个多核苷酸结合蛋白控制多核苷酸相对于所述孔的移动, 例如穿过所述孔。该方法还包括当所述多核苷酸相对于所述孔移动时, 获取一个或多个测量值。所述测量值表示多核苷酸的一个或多个特征, 如序列。

[0112] 已经表明了可以使用跨膜孔有效地表征双链多核苷酸, 如果所述双链多核苷酸被修饰为包括含有前导序列的Y衔接体(双链茎干和两个非互补臂)和桥连部分衔接体, 如发夹环衔接体(WO2013/014451)。优选为, 含有前导序列的Y衔接体连接到多核苷酸的一端, 而桥连部分衔接体连接到另一端。前导序列优选螺入纳米孔, 且连接所述多核苷酸的两条链的桥连部分(如发夹环)使得待测试的作为多核苷酸的两条链解开且使得两条链(经桥连部分连接)相对于所述孔移动, 如穿过所述孔。这是有利的, 因为其使从单个双链多核苷酸获得的信息量加倍。另外, 因为两条链中的序列是互补的, 因此来自两条链中的信息可以进行信息合并。这种机制提供了提供更高的可信观察结果的正交校阅能力。根据本发明使用的一个或多个装载部分可以是Y衔接体和/或桥连部分衔接体。这在下文更详细地讨论。

[0113] 本发明还提供了一种制备用于表征的靶多核苷酸的方法。该方法可用于改进用于表征的靶多核苷酸。该方法可以用于修饰或扩增靶多核苷酸。一旦一个或多个聚合酶已被装载到靶多核苷酸上, 它们可以被允许使用所述靶多核苷酸作为模板以形成一个或多个多核苷酸。如果靶多核苷酸是单链的, 则形成另一个互补的多核苷酸。如果所述靶多核苷酸是双链的, 则优选通过所述一个或多个聚合酶将两条链用作模板。因为来自靶多核苷酸的链和由一个或多个聚合酶产生的新的多核苷酸是互补的, 来自它们的信息可以如上所述进行信息合并。这种类型的方法也在WO2013/014451中公开。由聚合酶形成的多核苷酸可包含与靶多核苷酸相同种类的多核苷酸或不同种类的多核苷酸, 如下详述。靶多核苷酸可以被修饰, 如下文所详述。所述的一个或多个聚合酶可以使用Y型衔接体和/或桥连部分衔接体被装载到靶多核苷酸上, 如下文所详述。

[0114] 多核苷酸

[0115] 所述靶多核苷酸可以是任何多核苷酸。多核苷酸如核酸是含有两个或更多个核苷酸的大分子。所述多核苷酸或核酸可包括任何核苷酸的任意组合。核苷酸可以是天然存在的或人工合成的。多核苷酸中的一个或多个核苷酸可以被氧化或甲基化。多核苷酸中的一个或多个核苷酸的可被损坏。例如, 多核苷酸可包含嘧啶二聚体。此类二聚体通常与紫外线导致的损坏相关, 且是皮肤黑素瘤的首要原因。多核苷酸中的一个或多个核苷酸可被修饰, 例如用标记物或标签。合适的标记物如下所述。所述多核苷酸可包含一个或多个间隔器。

[0116] 核苷酸通常含有核碱基、糖和至少一个磷酸基团。所述核碱基和糖形成核苷。核苷酸可以是天然核苷酸或非天然核苷酸。

[0117] 核苷碱基通常为杂环的。核碱基包括但不限于: 嘌呤和嘧啶, 更具体地, 腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、尿嘧啶(U)和胞嘧啶(C)。

[0118] 糖通常为戊糖。核苷酸糖包括但不限于, 核糖和脱氧核糖。所述糖优选为脱氧核糖。

[0119] 所述多核苷酸中的核苷酸通常是核糖核苷酸或脱氧核糖核苷酸。所述多核苷酸可包含以下核苷：腺苷，尿苷，鸟苷和胞苷。核苷酸优选为脱氧核糖核苷酸。所述多核苷酸优选包含下列核苷：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)。

[0120] 所述核苷酸通常含有单磷酸、二磷酸或三磷酸。磷酸酶可以被连接在核苷酸的5'或3'侧。

[0121] 合适的核苷酸包括但不限于，单磷酸腺苷(AMP)、单磷酸鸟苷(GMP)、单磷酸胸苷(TMP)、单磷酸尿苷(UMP)、单磷酸胞苷(CMP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、单磷酸脱氧鸟苷(dGMP)、单磷酸脱氧胸苷(dTMP)、单磷酸脱氧尿苷(dUMP)和单磷酸脱氧胞苷(dCMP)。所述核苷酸优选选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP、dCMP和dUMP。核苷酸最优先选自dAMP、dTMP、dGMP、dCMP和dUMP。多核苷酸优选包含下列核苷酸：dAMP、dUMP和/或dTMP、dGMP和dCMP。

[0122] 多核苷酸中的核苷酸可以以任何方式彼此连接。如在核酸中一样，核苷酸通常通过它们的糖和磷酸基团连接。如嘧啶二聚体中一样，所述核苷酸可通过它们的核碱基连接。

[0123] 所述多核苷酸可以是单链或双链的。至少所述多核苷酸的一部分优选为双链的。

[0124] 所述多核苷酸可以是核酸。多核苷酸可以是本领域中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)、锁核酸(LNA)、桥连核酸(BNA)或其它具有核苷酸侧链的合成聚合物。PNA骨架是由通过肽键连接的N-(2-氨基乙基)甘氨酸重复单元组成。GNA骨架是由通过磷酸二酯键连接的甘油重复单元组成。TNA骨架是由通过磷酸二酯键连接在一起的苏糖重复单元组成。LNA是由在核糖部分中具有连接2'氧和4'碳的额外的桥的如上所述的核糖核苷酸形成。桥连核酸(BNA)为经修饰的RNA核苷酸。它们也可以被称为被约束的或不可接近的RNA。BNA单体可含有具有“固定的”C3'-内糖起褶(C3'-endo sugar pucker)的五元的、六元的或甚至七元的桥连结构。所述桥连是在核糖的2',4'-位置通过合成而引入的，以产生2',4'-BNA单体。

[0125] 所述多核苷酸最优先选为核糖核酸(RNA)或脱氧核糖核酸(DNA)。

[0126] 所述多核苷酸可以是任何长度。例如，多核苷酸可以是至少10、至少50、至少100、至少150、至少200、至少250、至少300、至少400或至少500个核苷酸的长度。所述多核苷酸可以是1000或更多个核苷酸、5000或更多个核苷酸，或100000或更多个核苷酸长度。

[0127] 在本发明的方法中解旋酶可以沿整个多核苷酸或仅多核苷酸的一部分移动。可以使用本发明的方法来表征所述整个靶多核苷酸或仅靶多核苷酸的一部分。

[0128] 所述多核苷酸可以是单链的。至少多核苷酸的一部分优选为双链的。解旋酶通常结合到单链多核苷酸。如果所述多核苷酸的至少一部分为双链的，所述多核苷酸优选包含单链区域或非杂合区域。所述一个或多个解旋酶能够结合到该单链区域或所述的非杂合区域的一条链。所述多核苷酸优选包含一个或多个单链区域或一个或多个非杂合区域。

[0129] 所述一个或多个间隔器优选包括在多核苷酸的单链区域或非杂合区域中。所述多核苷酸可包含多于一个的单链区域或多于一个的非杂合区域。所述多核苷酸可在其序列内和/或在其一端或两端包含单链区域或非杂合区域。所述一个或多个间隔器可以被包括在多核苷酸的双链区域中。

[0130] 如果在所述方法中使用的一个或多个解旋酶以5'至3'的方向移动，则多核苷酸优先地在其5'末端包含单链区域或非杂合区域。如果在所述方法中使用的一个或多个解旋酶

以3'至5'的方向移动,则所述多核苷酸优选地在其3'末端包含单链区域或非杂合区域。如果所述一个或多个解旋酶在非激活模式下使用(即作为制动器),则其中单链区域或非杂合区域定位的位置是不重要的。

[0131] 单链区域优选包含优选螺入所述孔的前导序列。这将在下文详述。

[0132] 如果所述多核苷酸的至少一部分为双链,则所述双链部分的两条链优选使用桥连部分,如发夹或发夹环,进行连接。这有助于本发明的表征方法,并在下文详述。

[0133] 所述多核苷酸存在于任何合适的样本中。本发明通常针对已知含有或怀疑含有所述多核苷酸的样品实施。可对样品实施本发明,以确认在样品中的存在是已知的或期望的一个或多个多核苷酸的同一性。

[0134] 所述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中获得或提取的样品在体外实施。所述生物体或微生物通常是古细菌的(archaeal),原核的或真核的,并且通常属于以下五界中的一个:植物界,动物界,真菌,原核生物和原生生物。本发明针对从任何病毒中获得或提取的样品在体外实施。所述样品优选是液体样品。样品通常包括患者的体液。所述样品可以是尿液,淋巴液,唾液,粘液或羊水,但优选血液,血浆或血清。通常,所述样品来源于人,但替代地可以是来自其他哺乳动物,如自商业上养殖的动物如马,牛,绵羊,鱼,鸡或猪,或者替代地可以是宠物如猫或狗。或者,所述样品可以来源于植物,例如从商业作物获得的样品,如谷类,豆类,水果或蔬菜,例如小麦,大麦,燕麦,芸苔,玉米,大豆,水稻,大黄,香蕉,苹果,番茄,土豆,葡萄,烟草,菜豆,小扁豆,甘蔗,可可,棉花。

[0135] 所述样品可以是非生物样品。所述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,以及用于实验室试验的试剂。

[0136] 所述样品通常是在本发明中使用前处理,例如通过离心,或通过膜过滤掉不需要的分子或细胞,例如红细胞。所述样品可在采集后立即测量。样品也可通常在测定前被存储,优选低于-70°C存储。

[0137] 多核苷酸结合蛋白

[0138] 多核苷酸结合蛋白可以是能够结合到多核苷酸且控制其相对于孔移动例如穿过所述孔的任何蛋白质。现有技术中可直接确定蛋白质是否结合到多核苷酸。所述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个特性。所述蛋白质可通过裂解多核苷酸以形成单个核苷酸或较短链的核苷酸,如二核苷酸或三核苷酸,来修饰所述多核苷酸。所述部分可以通过定向多核苷酸或移动多核苷酸到特定的位置例如控制其移动来修饰该多核苷酸。

[0139] 任何数目的多核苷酸蛋白可以被连接到靶多核苷酸。例如,1、2、3、4、5、6、7、8、9、10或更多个蛋白质可以被连接。

[0140] 所述一个或多个多核苷酸结合蛋白可以是一个或多个单链结合蛋白(SSB)。所述一个或多个单链结合蛋白(SSB)可以包含不具有净负电荷的羧基末端(C-末端)的区域或(ii)在其C-末端区包含一个或多个修饰且可降低C-末端区域的净负电荷的被修饰的SSB。所述一个或多个多核苷酸结合蛋白可以是在国际申请No.PCT/GB2013/051924中(公开为W02014/013259)公开的任何SSB。

[0141] 所述一个或多个多核苷酸结合蛋白优选衍生自多核苷酸处理酶。所述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个特性的多肽。所述酶可以通过裂解多核苷酸以形成单个核苷酸或较短链的核苷酸,如二核苷酸或三核苷酸,来修饰多核

昔酸。所述酶可以通过定向多核昔酸或将多核昔酸移动到特定的位置而修饰该多核昔酸。只要多核昔酸处理酶能够结合多核昔酸并控制其相对于孔的移动,例如穿过所述孔,所述多核昔酸处理酶并不需要显示酶活性。例如,所述酶可被修饰以去除其酶活性或者可在防止其作为酶的条件下使用。这样的条件将在下文更详细的讨论。

[0142] 所述一个或多个多核昔酸结合蛋白优选衍生自核酸溶解酶。所述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中的任何成员。所述酶可以是在国际申请No.PCT/GB10/000133(公开为WO2010/086603)中公开的任何酶。

[0143] 优选的酶是聚合酶、核酸外切酶、解旋酶和拓扑异构酶如旋转酶和逆转录酶。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酶 (SEQ ID NO:13)、来自嗜热栖热菌的RecJ (SEQ ID NO:15) 和细菌噬菌体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中示出的序列的三个亚基或其变体相互作用以形成三聚体核酸外切酶。所述聚合酶可以是 PyroPhage®3173DNA聚合酶(其可商购自 Lucigen®公司)、SD聚合酶(可商购自 Bioron®)、来自NEB的Klenow或其变体。所述酶优选为Phi29DNA聚合酶 (SEQ ID NO:9) 或其变体。Phi29的优选版本将在下文详述。可以在本发明中使用的Phi29聚合酶的经修饰版本将在下文论述且在美国专利No.5576204中公开。所述拓扑异构酶优选为部分分类(EC)组5.99.1.2和5.99.1.3中的任何成员。逆转录酶是能够催化从RNA模板形成cDNA的酶。它们可商购自例如New EnglandBiolabs®和Invitrogen®。

[0144] 所述一个或多个多核昔酸结合蛋白优选衍生自解旋酶。所述解旋酶可以在至少两个激活操纵模式(当为解旋酶提供有助于移动的所有必要组分例如ATP和镁离子时)和一个未激活操纵模式(当未为解旋酶提供有助于移动的必要组分时,或解旋酶被修饰成能防止或阻碍移动时)下控制多核昔酸的移动。当向解旋酶提供了促进移动的所有必要组分时,解旋酶沿多核昔酸以5'到3'或3'至5'的方向移动(取决于解旋酶),但多核昔酸在孔中的取向(取决于多核昔酸的哪一端被孔捕获)意味着解旋酶可用于逆着施加电场将多核昔酸移出孔,或顺着施加电场将多核昔酸移入孔中。当解旋酶朝其移动的多核昔酸的一端被孔捕获时,所述解旋酶可以逆着施加电势产生的电场的方向工作,并将螺旋的多核昔酸拉出孔,并拉入顺室中。然而,当所述解旋酶远离其而移动的一端在孔中被捕获时,所述解旋酶可以顺着所施加电势而产生的电场的方向工作,并将螺旋的多核昔酸推入孔并进入反式室中。

[0145] 当未为解旋酶提供有助于移动的必要组分时,在所述解旋酶通过施加的电势产生的场被拉入孔时,所述解旋酶可以结合到所述多核昔酸,且作为制动器减缓多核昔酸的移动。在非激活模式中,多核昔酸的哪一端被捕获并不重要,是施加的电场将多核昔酸朝着反式侧拉入孔中,所述解旋酶起制动器的作用。当在非激活模式中,解旋酶对多核昔酸的移动控制可以包括棘轮(ratcheting)、滑动和制动在内的多种方式进行描述。

[0146] 在本发明的表征方法中,利用施加的电势所产生的电场,所述一个或多个解旋酶优选控制所述靶多核昔酸相对于所述孔的移动,如穿过所述孔。在一个优选的实施方案中,所述一个或多个解旋酶是以激活模式使用,且所述一个或多个解旋酶远离其而移动的一端被所述孔捕获,使得所述一个或多个解旋酶顺着所施加的电势所产生的电场工作,并且相对于孔推动多核昔酸,例如穿过所述孔。如果所述一个或多个解旋酶以5'到3'方向移动,所

述多核苷酸的5'末端优选被孔捕获。在本实施方案中,所述一个或多个解旋酶以5'到3'方向沿多核苷酸移动。如果所述一个或多个解旋酶以3'到5'方向移动,所述的多核苷酸的3'末端优选被孔捕获。在这类实施方案中,所述一个或多个解旋酶以3'到5'方向沿多核苷酸移动。

[0147] 在表征方法的另一个优选实施方案中,所述一个或多个解旋酶以非激活模式使用,使得所施加的场相对于孔牵拉多核苷酸,如穿过所述孔,并且所述一个或多个解旋酶充当制动器。在另一个优选的实施方案中,所述一个或多个解旋酶被修饰,使得它们保持它们的多核苷酸结合能力,但缺乏解旋酶活性(即主动地沿多核苷酸移动的能力),使得所施加的场相对于孔牵拉多核苷酸,如穿过所述孔,并且所述一个或多个解旋酶充当制动器。在本发明的方法中,优选所述一个或多个解旋酶利用由所施加的电势产生的场,减缓或制动多核苷酸相对于孔的移动,如穿过所述孔。在任一情况下,所述的一个或多个解旋酶通常太大而不能相对于孔移动,如穿过所述孔,并且在所施加的电势产生的场下当多核苷酸相对于孔移动,如穿过所述孔时,所述孔沿所述多核苷酸推动所述一个或多个解旋酶。

[0148] 使用一个或多个解旋酶进行的表征方法的任何步骤通常在游离核苷酸或游离核苷酸类似物存在下,以及有助于一个或多个解旋酶的作用的酶辅因子的存在下进行。所述游离核苷酸可以是一个或多个上述的任意单个核苷酸。所述游离核苷酸包括但不限于,单磷酸腺苷(AMP)、二磷酸腺苷(ADP)、三磷酸腺苷(ATP)、单磷酸鸟苷(GMP)、二磷酸鸟苷(GDP)、三磷酸鸟苷(GTP)、单磷酸胸苷(TMP)、二磷酸胸苷(TDP)、三磷酸胸苷(TTP)、单磷酸尿苷(UMP)、二磷酸尿苷(UDP)、三磷酸尿苷(UTP)、单磷酸胞苷(CMP)、二磷酸胞苷(CDP)、三磷酸胞苷(CTP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、二磷酸脱氧腺苷(dADP)、三磷酸脱氧腺苷(dATP)、单磷酸脱氧鸟苷(dGMP)、二磷酸脱氧鸟苷(dGDP)、三磷酸脱氧鸟苷(dGTP)、单磷酸脱氧胸苷(dTMP)、二磷酸脱氧胸苷(dTDP)、三磷酸脱氧胸苷(dTTP)、单磷酸脱氧尿苷(dUMP)、二磷酸脱氧尿苷(dUDP)、三磷酸脱氧尿苷(dUTP)、单磷酸脱氧胞苷(dCMP)、二磷酸脱氧胞苷(dCDP)和三磷酸脱氧胞苷(dCTP)。所述游离核苷酸优选选自AMP、TMP、GMP、CMP、UMP、dAMP、dTDP、dGMP或dCMP。所述游离核苷酸优选为三磷酸腺苷(ATP)。所述酶辅因子是允许构建体起作用的因子。所述酶辅因子优选为二价金属阳离子。所述二价金属阳离子优选为Mg²⁺、Mn²⁺、Ca²⁺或Co²⁺。所述酶辅因子最优选Mg²⁺。

[0149] 在本发明中可以使用任何解旋酶。解旋酶可以衍生自He1308解旋酶、RecD解旋酶,如TraI解旋酶或TrwC解旋酶, XPD解旋酶或Dda解旋酶。所述的解旋酶可以是在国际申请No.PCT/GB2012/052579(公开为W02013/057495)、PCT/GB2012/053274(公开为W02013/098562)、PCT/GB2012/053273(公开号为W02013098561)、PCT/GB2013/051925(公开为W02014/013260)、PCT/GB2013/051924(公开为W02014/013259)和PCT/GB2013/051928(公开为W02014/013262)和在国际申请No.PCT/GB2014/052736中公开的任何的解旋酶、被修饰的解旋酶或解旋酶构建体。

[0150] 所述一个或多个多核苷酸结合蛋白可衍生自解旋酶,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。所述一个或多个多核苷酸结合蛋白优选包含在SEQ ID NO:25(Trwc Cba)中示出的序列或其变体、在SEQ ID NO:18(He1308Mbu)中示出的序列(He1308Mbu)或其变体,或在SEQ ID NO:24(Dda)中示出序列或其变体。对解旋酶或跨膜孔,变体可以下面所论述的任何方式不同于天然序列。

[0151] SEQ ID NO:24的优选的变体包含(或仅包含) (a) E94C/A360C, (b) E94C/A360C, 然后(Δ M1) G1G2(即M1缺失,然后添加G1和G2), (c) E94C/A360C/C109A/C136A或(d) E94C/A360C/C109A/C136A,然后(Δ M1) G1G2(即M1缺失,然后添加G1和G2)。

[0152] SEQ ID NO:24的其它优选的变体包含W378A。SEQ ID NO:24的优选变体包含(或仅包括) (a) E94C/A360C/W378A, (b) E94C/A360C/W378A, 然后(Δ M1) G1G2(即M1缺失,然后添加G1和G2), (c) E94C/A360C/C109A/C136A/W378A或(d) E94C/A360C/C109A/C136A/W378A,然后(Δ M1) G1G2(即M1缺失,然后添加G1和G2)。

[0153] SEQ ID NO:25的优选的变体包含(或仅包括) (a) Q594A, (b) L376C/Q594A/K762C, (c) L376C/Q594A/A779C, (d) Q346C/Q594A/A779C, (e) Q346C/Q594A/A783C, (f) D411/Q594A/A783C, (g) Q594A/R353C/E722C, (h) Q594A/Q357C/T720C, (i) Q594A/R358C/T720C, (j) Q594A/H354C/T720C, (k) Q594A/F374C/E722C或(l) Q594A/S350C/E722C。(a)至(l)中的任一项还可以进一步包含,然后(Δ M1) G1G2(即M1缺失,然后添加G1和G2)。其它优选的变体如上所述。

[0154] 所述一个或多个解旋酶优选修饰为减少多核苷酸结合结构域的开口的大小,所述多核苷酸可以在至少一个构象状态下穿过所述开口从解旋酶上解绑。所述一个或多个解旋酶优选修饰为闭合该多核苷酸结合域的开口,所述多核苷酸可以在至少一个构象状态下穿过所述开口从解旋酶上解绑。这在W02014/013260和PCT/GB2014/052736中被公开。可以在本发明中使用在W02014/013260和PCT/GB2014/052736中所公开的任何修饰。

[0155] 解旋酶结合到多核苷酸和从多核苷酸解绑的能力可以使用本领域中已知的任何方法来确定。合适的结合/解绑测定包括但不限于,天然聚丙烯酰胺凝胶电泳(PAGE),荧光各向异性、量热法和表面等离子共振(SPR,如BiacoreTM)。解旋酶从多核苷酸解绑的能力当然可以通过测量解旋酶可以控制多核苷酸移动的时间来确定。这也可使用本领域中已知的任何方法来确定。解旋酶控制多核苷酸移动的能力通常在纳米孔系统中测定,例如如下所述的那些。解旋酶控制多核苷酸移动的能力可以如实施例中所描述的来确定。

[0156] 如PCT/GB2014/052736中公开的,本发明中使用的解旋酶可以是Dda解旋酶,所述Dda中至少一个半胱氨酸残基和/或至少一个非天然的氨基酸已经被引入到钩结构域和/或2A(RecA状电机)结构域,其中所述解旋酶保留其控制多核苷酸移动的能力。任何数量的半胱氨酸残基和/或非天然氨基酸可以被引入到每个结构域。例如,1、2、3、4、5、6、7、8、9、10个或更多个半胱氨酸残基可被引入,和/或1、2、3、4、5、6、7、8、9、10个或更多个非天然氨基酸可被引入。可以仅一个或多个半胱氨酸残基被引入。可以仅一个或多个非天然氨基酸被引入。一个或多个半胱氨酸残基和一个或多个非天然氨基酸的组合可被引入。优选通过取代引入至少一个半胱氨酸残基和/或至少一个非天然氨基酸。这样做的方法是本领域已知的。这些Dda修饰不阻止解旋酶结合到多核苷酸。这些修饰降低了多核苷酸从解旋酶上解绑或分离的能力。换言之,所述一个或多个修饰通过防止Dda解旋酶从多核苷酸链的解离来增加Dda解旋酶的持续合成能力(processivity)。酶的热稳定性通常也通过一个或多个修饰给予酶在链测序中有利的提高的结构稳定性来增加。非天然氨基酸为不是在Dda解旋酶中天然发现的氨基酸。非天然氨基酸优选不是组氨酸、丙氨酸、异亮氨酸、精氨酸、亮氨酸、天冬酰胺、赖氨酸、天冬氨酸、甲硫氨酸、半胱氨酸、苯丙氨酸、谷氨酸、苏氨酸、谷氨酰胺、色氨酸、甘氨酸、缬氨酸、脯氨酸、丝氨酸或酪氨酸。非天然氨基酸更优选不是前句中二十种氨基酸中的任一种氨基酸或硒代半胱氨酸。

[0157] 在本发明中使用的优选的非天然氨基酸包括但不限于,4-叠氮基-L-苯丙氨酸(Faz)、4-乙酰基-L-苯丙氨酸、3-乙酰基-L-苯丙氨酸、4-乙酰乙酰基-L-苯丙氨酸、0-烯丙基-L-酪氨酸、3-(苯硒基)-L-丙氨酸、0-2-丙炔-1-基-L-酪氨酸、4-(二羟基硼基)-L-苯丙氨酸、4-[(乙基硫烷基) 羰基]-L-苯丙氨酸、(2S)-2-氨基-3-{4-[(丙烷-2-基硫烷基) 羰基] 苯基} 丙酸、(2S)-2-氨基-3-{4-[(2-氨基-3-硫烷基丙酰基) 氨基] 苯基} 丙酸、0-甲基-L-酪氨酸、4-氨基-L-苯丙氨酸、4-氰基-L-苯丙氨酸、3-氰基-L-苯丙氨酸、4-氟-L-苯丙氨酸、4-碘-L-苯丙氨酸、4-溴-L-苯丙氨酸、0-(三氟甲基) 酪氨酸、4-硝基-L-苯丙氨酸、3-羟基-L-酪氨酸、3-氨基-L-酪氨酸、3-碘-L-酪氨酸、4-异丙基-L-苯丙氨酸、3-(2-萘基)-L-丙氨酸、4-苯基-L-苯丙氨酸、(2S)-2-氨基-3-(萘-2-基氨基) 丙酸、6-(甲基硫烷基) 正亮氨酸、6-氧化-L-赖氨酸、D-酪氨酸、(2R)-2-羟基-3-(4-羟基苯基) 丙酸、(2R)-2-铵基辛酸酯3-(2,2'-双吡啶-5-基)-D-丙氨酸、2-氨基-3-(8-羟基-3-喹啉基) 丙酸、4-苯甲酰基-L-苯丙氨酸、S-(2-硝基苄基) 半胱氨酸、(2R)-2-氨基-3-[(2-硝基苄基) 硫烷基] 丙酸、(2S)-2-氨基-3-[(2-硝基苄基) 氧基] 丙酸、0-(4,5-二甲氧基-2-硝基苄基)-L-丝氨酸、(2S)-2-氨基-6-({[(2-硝基苄基) 氧基] 羰基} 氨基) 己酸、0-(2-硝基苄基)-L-酪氨酸、2-硝基苯丙氨酸、4-[(E)-苯二氮烯基]-L-苯丙氨酸、4-[3-(三氟甲基)-3H-双吖丙啶-3-基]-D-苯丙氨酸(4-[3-(Trifluoromethyl)-3H-diaziren-3-yl]-D-phenylalanine)、2-氨基3-[[5-(二甲基氨基)-1-萘基] 硼酰基氨基] 丙酸、(2S)-2-氨基-4-(7-羟基-2-氧化-2H-色烯-4-基) 丁酸、(2S)-3-[(6-乙酰萘-2-基) 氨基]-2-氨基丙酸、4-(羧甲基) 苯丙氨酸、3-硝基-L-酪氨酸、0-磺基-L-酪氨酸、(2R)-6-乙酰胺基-2-铵基己酸酯、1-甲基组氨酸、2-氨基壬酸、2-氨基癸酸、L-高半胱氨酸、5-硫烷基正缬氨酸、6-硫烷基-L-正亮氨酸、5-(甲基硫烷基)-L-正缬氨酸、N⁶-{[(2R,3R)-3-甲基-3,4-二氢-2H-吡咯-2-基] 羰基}-L-赖氨酸、N⁶-[(苄氧基) 羰基] 赖氨酸、(2S)-2-氨基-6-[(环戊基羰基) 氨基] 己酸、N⁶-[(环戊氧基) 羰基]-L-赖氨酸、(2S)-2-氨基-6-{[(2R)-四氢呋喃-2-基羰基] 氨基} 己酸、(2S)-2-氨基-8-{(2R,3S)-3-乙炔基四氢呋喃-2-基}-8-氧化辛酸、N⁶-(叔丁氧基羰基)-L-赖氨酸、(2S)-2-羟基-6-{[(2-甲基-2-丙基) 氧基] 羰基} 氨基) 己酸、N⁶-[(烯丙氧基) 羰基] 赖氨酸、(2S)-2-氨基-6-{[(2-叠氮基苄基) 氧基] 羰基} 氨基) 己酸、N⁶-L-脯氨酰-L-赖氨酸、(2S)-2-氨基-6-{[(丙-2-炔-1-基氧基) 羰基] 氨基} 己酸和N⁶-[(2-叠氮基乙氧基) 羰基]-L-赖氨酸。最优秀的非天然氨基酸是4-叠氮基-L-苯丙氨酸(Faz)。

[0158] 本发明中使用的解旋酶优选包含SEQ ID NO:24的变体,其中至少一个半胱氨酸残基和/或至少一个非天然氨基酸已经被引入(i)塔域(残基D260-P274和N292-A389),和/或(ii)销域(残基K86-E102),和/或(iii)1A域(残基M1-L85和V103-K177)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N292-A389中。

[0159] 如上所述,将至少两个半胱氨酸引入到SEQ ID NO:24和SEQ ID NO25中减小了解旋酶的多核苷酸结合域中的开口的大小或闭合了所述开口。

[0160] 在本发明中使用的优选的解旋酶构建体在国际申请No.PCT/GB2013/051925(公开为W02014/013260)、PCT/GB2013/051924(公开为W02014/013259)和PCT/GB2013/051928(公开为W02014/013262)和于2013年10月18日在英国提交的申请号1318464.3中描述。

[0161] SEQ ID NO:9、SEQ ID N011、SEQ ID N013、SEQ ID N015、SEQ ID N017、SEQ ID N018、SEQ ID N019、SEQ ID N020、SEQ ID N021、SEQ ID N022、SEQ ID N023、SEQ ID N024

或SEQ ID N025的变体是具有从SEQ ID N0:9、SEQ ID N011、SEQ ID N013、SEQ ID N015、SEQ ID N017、SEQ ID N018、SEQ ID N019、SEQ ID N020、SEQ ID N021、SEQ ID N022、SEQ ID N023、SEQ ID N024或SEQ ID N025变化而来且保持多核苷酸结合能力的氨基酸序列的酶。这可以使用任何本领域中已知的方法来测定。例如,可使变体与多核苷酸接触并且测量其结合到多核苷酸并沿着多核苷酸移动的能力。所述变体可以包括有助于结合多核苷酸的修饰,和/或有助于其在高盐浓度和/或室温下活性的修饰。变体可被修饰,以使得其结合多核苷酸(即保留多核苷酸结合能力),但没有酶的功能。例如,解旋酶的变体可被修饰以使其结合多核苷酸(即保留多核苷酸结合能力),但没有解旋酶的功能(即当提供所有有助于移动的必要组分,例如ATP和Mg²⁺时,解旋酶不沿多核苷酸移动)。这样的修饰在本领域是已知的。例如,解旋酶中的Mg²⁺结合域的修饰通常导致不具有解旋酶功能的变体。这些类型的变体可以作为分子制动器。优选的分子制动器为TrwC CBA-Q594A(具有突变Q594A的SEQ ID N0:25)。其他的参考SEQ ID N0:25在上文进行了论述。该变体不具有解旋酶的功能(即当提供所有有助于移动的必要组分,例如ATP和Mg²⁺时,结合到多核苷酸,但是并不沿着多核苷酸移动)。一个或多个分子制动器解旋酶可以上文所述的任何方向和/或模式使用。

[0162] 在SEQ ID N0:9、SEQ ID N0:11、SEQ ID N0:13、SEQ ID N0:15、SEQ ID N0:17、SEQ ID N0:18、SEQ ID N0:19、SEQ ID N0:20、SEQ ID N0:21、SEQ ID N0:22、SEQ ID N0:23、SEQ ID N0:24或SEQ ID N0:25的氨基酸序列的整个长度上,基于氨基酸的同一性,变体将优选与所述序列为至少50%的同源性。更优选地,变体多肽可以与SEQ ID N0:9、SEQ ID N0:11、SEQ ID N0:13、SEQ ID N0:15、SEQ ID N0:17、SEQ ID N0:18、SEQ ID N0:19、SEQ ID N0:20、SEQ ID N0:21、SEQ ID N0:22、SEQ ID N0:23、SEQ ID N0:24或SEQ ID N0:25的氨基酸序列在整个序列上基于氨基酸的同一性具有至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,更优选至少95%,97%或99%的同源性。在200或更多,例如230、250、270、280、300、400、500、600、700、800、900或1000个或更多个连续氨基酸的长度上,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”)。同源性如上文所描述的确定。关于下文的SEQ ID N0:2和SEQ ID N0:4,所述的变体可以上文所述的任何方式不同于野生型序列。

[0163] 如果两个或多个多核苷酸结合蛋白被使用,它们可以是相同的或不同的。可使用上述的蛋白质的任意组合。例如,两个或多个蛋白可以是相同蛋白例如解旋酶的不同变体。所述两个或多个多核苷酸结合蛋白优选包含一个或多个解旋酶和一个或多个聚合酶。

[0164] 如果两个或多个多核苷酸结合蛋白被使用,它们可以彼此连接。所述两个或多个多核苷酸结合蛋白可以共价地彼此连接。多核苷酸结合蛋白可以以任何顺序和使用任何方法连接。本发明中使用的优选的解旋酶构建体在国际申请No.PCT/GB2013/051925(公开为W02014/013260)、PCT/GB2013/051924(公开为W02014/013259)和PCT/GB2013/051928(公开为W02014/013262)和于2013年10月18日在英国提交的申请号1318464.3中描述。

[0165] 如果使用了两个或多个多核苷酸结合蛋白,则除了经由多核苷酸外,它们优选不彼此连接。所述两个或多个多核苷酸结合蛋白更优选彼此不共价连接。

[0166] 一个或多个解旋酶和一个或多个分子制动器

[0167] 在某些情况下,本发明的表征方法可以涉及使用一个或多个解旋酶和一个或多个分子制动器。当靶多核苷酸与所述孔相接触时,所述一个或多个解旋酶和一个或多个分子

制动器聚集在一起，并且这两者控制多核苷酸相对于所述孔的移动，如穿过所述孔。

[0168] 所述一个或多个解旋酶可以是任何上述的那些解旋酶。所述一个或多个分子制动器可以是结合到多核苷酸并减缓多核苷酸相对于孔的移动如穿过所述孔的任何化合物或分子。所述一个或多个分子制动器优选包括结合到多核苷酸的一个或多个化合物。所述一个或多个化合物优选是一个或多个大环化合物。合适的大环化合物包括，但不限于，环糊精、杯芳烃、环肽、冠醚、瓜环、柱芳烃、其衍生物或其组合。环糊精或其衍生物可以是任何在 Eliseev, A.V., 与 Schneider, H-J. (1994) J. Am. Chem. Soc. 116, 6081-6088 中公开的那些。所述的试剂更优选是七-6-氨基- β -环糊精 (am₇- β CD)、6-单脱氧-6-单氨基- β -环糊精 (am₁- β CD) 或七-(6-脱氧-6-氨基)-环糊精 (gu₇- β CD)。

[0169] 所述一个或多个分子制动器优选为一个或多个单链结合蛋白 (SSB)。所述一个或多个分子制动器更优选为包含不具有净负电荷的羧基末端 (C-末端) 区域的单链结合蛋白 (SSB)，或为 (ii) C-末端区域中包含能减少 C-末端区域的净负电荷的一个或多个修饰的经修饰SSB。所述一个或多个分子制动器最优选为在国际申请No.PCT/GB2013/051924 (公开为 WO2014/013259) 中公开的任何SSB。

[0170] 所述一个或多个分子制动器优选为一个或多个多核苷酸结合蛋白。所述多核苷酸结合蛋白可以是能够结合到多核苷酸并控制其相对于孔的移动，如穿过所述孔的任何蛋白。现有技术中可直接确定蛋白质是否结合到多核苷酸。所述蛋白通常与多核苷酸相互作用并修饰所述多核苷酸的至少一个特性。所述蛋白可通过裂解以形成单个核苷酸或较短链的核苷酸，如二核苷酸或三核苷酸，来修饰多核苷酸。所述部分可以通过将多核苷酸定向或通过移动多核苷酸到特定的位置，即控制多核苷酸的移动，修饰多核苷酸。

[0171] 所述多核苷酸结合蛋白优选衍生自多核苷酸处理酶。所述一个或多个分子制动器可以衍生自上述的任何多核苷酸处理酶。充当分子制动器的Phi29聚合酶的经修饰的版本 (SEQ ID NO:9) 在美国专利No. 5576204 中公开。充当分子制动器的Phi29聚合酶的经修饰的版本 (SEQ ID NO:9) 在下文中公开。所述一个或多个分子制动器优选衍生自解旋酶。

[0172] 任何数量的衍生自解旋酶的分子制动器可以被使用。例如，1、2、3、4、5、6、7、8、9、10 或更多个解旋酶可被用作分子制动器。如果两个或更多个解旋酶被用作分子制动器，所述两个或更多个解旋酶通常是相同的解旋酶。所述两个或更多个解旋酶可以是不同的解旋酶。

[0173] 所述两个或更多个解旋酶可以是上文提到的解旋酶的任意组合。所述两个或更多个解旋酶可以是两个或更多个Dda解旋酶。所述两个或更多个解旋酶可以是一个或多个Dda解旋酶和一个或多个TrwC解旋酶。所述两个或更多个解旋酶可以是相同解旋酶的不同变体。

[0174] 所述两个或更多个解螺旋酶优选彼此连接。所述两个或更多个解旋酶更优选共价地彼此连接。所述解旋酶可以以任何顺序和使用任何方法连接。从解旋酶衍生的一个或多个分子制动器优选被修饰以减少多核苷酸结合域中开口的大小，所述多核苷酸可在至少一个构象状态下通过所述开口从解旋酶解绑。这在 WO2014/013260 中公开。

[0175] 用于本发明的优选的解旋酶构建体在国际申请No.PCT/GB2013/051925 (公开为 WO2014/013260)、PCT/GB2013/051924 (公开为 WO2014/013259)、PCT/GB2013/051928 (公开为 WO2014/013262) 和 PCT/GB2014/052736 中描述。

[0176] 如果所述一个或多个解旋酶以激活模式使用 (即当为一个或多个解旋酶提供有助于移动的所有必要组分，例如ATP和Mg²⁺时)，所述一个或多个分子制动器优选 (a) 在非激活

模式下使用(即在不存在有助于移动的必要组分或没有能力主动移动时),(b)在激活模式下使用,其中所述一个或多个分子制动器以与所述一个或多个解旋酶相反的方向移动,或(c)在激活模式下使用,其中所述一个或多个分子制动器以与所述一个或多个解旋酶相同的方向移动,且比所述一个或多个解旋酶移动得更慢。

[0177] 如果在激活模式下使用所述一个或多个解旋酶(即当未为所述一个或多个解旋酶提供有助于移动的所有必要组分,例如ATP和Mg²⁺或不能主动移动时),所述一个或多个分子制动器优选(a)在非激活模式下使用(即在不存在有助于移动所需的组分时或不能主动移动时)或(b)在激活模式下使用,其中所述一个或多个分子制动器以与多核苷酸相对于孔的相同方向沿多核苷酸移动,如穿过所述孔。

[0178] 所述一个或多个解旋酶和一个或多个分子制动器可以在任何位置被连接到多核苷酸,以使它们聚集在一起并且这两者控制多核苷酸相对于所述孔的移动,如穿过所述孔。所述一个或多个解旋酶和一个或多个分子制动器为至少一个核苷酸间隔,例如至少5个、至少10个、至少50个、至少100个、至少500个、至少1000、至少5000、至少10,000、至少50,000个核苷酸间隔或更多个核苷酸间隔。如果该方法涉及用在一端设置的Y衔接体和在另一端设置的桥连部分衔接体如发夹环衔接体表征双链多核苷酸,则所述一个或多个解旋酶被优选地连接到Y衔接体,而所述一个或多种分子制动器优选被连接到桥连部分衔接体。在本实施方案中,所述一个或多个分子制动器优选是经修饰的一个或多个解旋酶,使得它们结合多核苷酸但不具有解旋酶的功能。连接到Y衔接体的一个或多个解旋酶如下文中详述的优选被停滞在间隔器处。连接到桥连部分衔接体的一个或多个分子制动器优选不停滞在间隔器处。当所述一个或多个解旋酶到达桥连部分时,所述一个或多个解旋酶和一个或多个分子制动器优选聚集在一起。所述一个或多个解旋酶可以在Y衔接体被连接到多核苷酸之前或Y衔接体被连接到多核苷酸之后,连接到Y衔接体。一个或多个分子制动器可以在所述桥连部分衔接体被连接到多核苷酸之前或桥连部分衔接体被连接到多核苷酸之后,连接到桥连部分衔接体。

[0179] 所述一个或多个解旋酶和一个或多个分子制动器优选不彼此连接。所述一个或多个解旋酶和一个或多个分子制动器更优选不共价地彼此连接。所述一个或多个解旋酶和一个或多个分子制动器优选不连接,如在国际申请No.PCT/GB2013/051925(公开为W02014/013260)、PCT/GB2013/051924(公开为W02014/013259)和PCT/GB2013/051928(公开为W02014/013262),以及于2013年10月18日在英国提交的申请号为1318464.3中所描述的。

[0180] 一个或多个聚合酶

[0181] 本发明的方法优选涉及将一个或多个聚合酶连接到靶多核苷酸。可以使用上文或下文所述的任何聚合酶。例如,聚合酶可包含上文所述的SEQ ID NO:9或SEQ ID NO:31中所示的序列或其变体。

[0182] SEQ ID NO:9的优选的变体包括,但不限于,包含以下取代物的变体:(a)G410C和P562C、(b)A411C和Q560C、(c)K402C和I94C、(d)A406C和E75C、以及(e)L412C和Q560C。将这些半胱氨酸引入到SEQ ID NO:9中减小了上文针对解旋酶所述的聚合酶的多核苷酸结合域中的开口的大小,或闭合了该开口。(a)至(e)中的变体可以被用作如上所述的分子制动器。

[0183] 一旦一个或多个聚合酶被装载到靶多核苷酸上,它/它们可以使用靶多核苷酸作为模板形成一个或多个多核苷酸。这通常包括在聚合酶使用靶多核苷酸作为模板形成一个或多个多核苷酸的条件下,将靶多核苷酸与一个或多个具有一组游离核苷酸的聚合酶接

触。合适的条件将在下文论述。任何数量的多核苷酸可以由一个或多个聚合酶形成。优选形成一个或两个多核苷酸。

[0184] 所述一组游离核苷酸可以包含上文和下文所述的任何核苷酸。由所述一个或多个聚合酶形成的一个或多个多核苷酸的性质将取决于该组中的游离核苷酸。所述一个或多个聚合酶可以形成与靶多核苷酸相同类型的一个或多个多核苷酸。例如,如果靶多核苷酸为DNA,本发明可使用一组游离的DNA核苷酸(即包括脱氧腺苷、脱氧鸟苷、胸苷、脱氧胞苷和脱氧甲基胞苷的核苷酸),使得所述一个或多个聚合酶使用靶多核苷酸作为模板形成一个或多个DNA多核苷酸。所述一个或多个聚合酶可以形成与靶多核苷酸不同类型的一个或多个多核苷酸。例如,如果靶多核苷酸是RNA,本发明可使用一组游离的DNA核苷酸,使得所述一个或多个聚合酶使用靶多核苷酸作为模板形成一个或多个DNA多核苷酸。所述一个或多个聚合酶可以如下文详述来修饰靶多核苷酸。

[0185] 引物或3'发夹通常用作用于聚合酶扩展的成核点。所述一个或多个聚合酶可以本文所述的任何方式连接到靶多核苷酸。所述一个或多个聚合酶通常被提供结合到一个或多个装载部分,如一个或多个Y衔接体和/或一个或多个桥连部分衔接体(如一个或多个发夹环衔接体)。

[0186] 本发明的方法优选包括根据本发明将一个或多个聚合酶连接到单链靶多核苷酸,并使一个或多个聚合酶使用所述靶多核苷酸作为模板形成多核苷酸,并由此产生双链多核苷酸。所述双链多核苷酸包含靶多核苷酸(模板)和由所述一个或多个聚合酶形成的互补多核苷酸。以这种方式形成的双链多核苷酸的两条链可以与桥连部分衔接体如发夹环衔接体相连,且该双链多核苷酸的两条链可如下文所述被表征。

[0187] 本发明提供了制备用于表征的靶多核苷酸的方法。该方法包括使用本发明将一个或多个聚合酶连接到靶多核苷酸,然后使用所述靶多核苷酸作为模板,用所述一个或多个聚合酶产生一个或多个额外的多核苷酸。产生与靶多核苷酸互补的多核苷酸有助于其表征,如上所述。

[0188] 所述方法优选包括根据本发明将一个或多个聚合酶连接到双链靶多核苷酸,并使所述一个或多个聚合酶使用靶多核苷酸的每条链作为模板形成一个或多个多核苷酸。

[0189] 在一个优选的实施方案中,双链靶多核苷酸的两条链在一端与桥连部分衔接体(如发夹环衔接体)连接,且使用包含另一桥连部分的装载部分将一个或多个聚合酶连接到所述双链靶多核苷酸的另一端。所述装载部分可以是Y衔接体。所述另一桥连部分通常在一个或更多个聚合酶结合的Y衔接体的链的末端由发夹环形成。所述装载部分可以包括一个或多个解旋酶,优选结合到一个或多个聚合酶结合的链的相对链。在本实施方案中,一个或多个聚合酶将产生双链构建体,其中所述构建体的两条链在一端与桥连部分连接,而且所述构建体的每条链包含靶多核苷酸的一条链和由一个或多个聚合酶形成的互补的多核苷酸。这在图17中示出。双链构建体的两条链可以如下文所述被表征。在本实施方案中,原始靶多核苷酸的每条链被表征两次。所述一个或多个聚合酶可以用于控制双链构建体相对于孔的移动,如穿过所述孔。所述一个或多个聚合酶可以是分子制动器。如果装载部分还包括一个或多个解旋酶,所述一个或多个解旋酶可以用于控制双链构建体相对于所述孔的移动,如穿过所述孔。在一些实施方案中,一个或多个聚合酶和一个或多个解旋酶可以都被用于控制双链构建体相对于所述孔的移动,如穿过所述孔。

[0190] 在另一个实施方案中,双链靶多核苷酸的两条链的任一端不与桥连部分衔接体(如

发夹环衔接体)连接。一个或多个聚合酶使用包含桥连部分的装载部分在每一端被连接到双链靶多核苷酸。在每一端的装载部分可以是相同的或不同的。每个装载部分可以是Y衔接体。桥连部分通常在一个或多个聚合酶结合的Y衔接体的链的末端由发夹环形成。所述装载部分可包含一个或多个解旋酶,优选结合到与一个或多个聚合酶结合的链的相对链。在本实施方案中,一个或多个聚合酶将产生两个双链构建体,其中每个构建体的两条链一端与桥连部分连接,且每个构建体包含靶多核苷酸的一条链和由所述一个或多个聚合酶形成的互补的多核苷酸。这在图18中示出。所述两个构建体可以如下文论述被表征。在本实施方案中,原始靶多核苷酸的每条链表征两次,每个构建体表征一次。所述一个或多个聚合酶可以控制每个双链构建体相对于所述孔的移动,如穿过所述孔。所述一个或多个聚合酶可以是分子制动器。如果每个装载部分还包括一个或多个解旋酶,则所述一个或多个解旋酶可被用于控制每个双链构建体相对于所述孔的移动,如穿过所述孔。在一些实施方案中,所述一个或多个聚合酶和一个或多个解旋酶可都被用于控制每个双链构建体相对于孔的移动,如穿过所述孔。

[0191] 在另一个实施方案中,双链靶多核苷酸的两条链的两端与桥连部分衔接体(如发夹环衔接体)连接,以形成环形构建体,并且一个或多个聚合酶被连接到每个桥连部分衔接体。所述的一个或所述的两个桥连部分衔接体可以包含一个或多个解旋酶,优选结合到杂合到桥连部分衔接体的多核苷酸。通过杂合该多核苷酸到桥接衔接体部分形成的双链多核苷酸可由所述一个或多个聚合酶来成核,用于聚合酶延伸。在本实施方案中,在环形构建体的每一端的一个或多个聚合酶将使用该环形构建体作为模板,产生包含所述靶多核苷酸的每条链的多个拷贝的构建体。由于有两组一个或多个聚合酶,两个构建体将从每个靶多核苷酸产生。这在图19中示出。每个构建体可以如下所述被表征。每个构建体将通常包含靶多核苷酸的每条链的多个拷贝。构建体中的拷贝可以杂合在一起形成双链区域。在本实施方案中,原始靶多核苷酸的每条链将进行与一个或多个聚合酶复制它们时一样多次数的表征。所述一个或多个聚合酶可控制构建体相对于所述孔的移动,如穿过所述孔。所述一个或多个聚合酶可以是分子制动器。如果每个桥连部分还包括一个或多个解旋酶,所述一个或多个解旋酶可以被用于控制构建体相对于所述孔的移动,如穿过所述孔。在一些实施方案中,所述一个或多个聚合酶和一个或多个解旋酶可以都被用于控制构建体相对于所述孔的移动,如穿过所述孔。

[0192] 使用一个或多个聚合酶产生一个或多个多核苷酸之后,该方法优选包括将靶多核苷酸和一个或多个多核苷酸与跨膜孔接触,使得靶多核苷酸和所述一个或多个多核苷酸相对于孔的移动,如穿过所述孔。该方法优选包括将所述靶多核苷酸和一个或多个多核苷酸与跨膜孔接触,使得所述一个或多个聚合酶控制靶多核苷酸和一个或多个多核苷酸相对于孔的移动,如穿过所述孔。一个或多个解旋酶也可以用于控制靶多核苷酸和一个或多个多核苷酸相对于孔的移动,如穿过所述孔。所述一个或多个解旋酶可以连接到所述靶多核苷酸和所述一个或多个多核苷酸,如上所述。

[0193] 该方法还包含当所述靶多核苷酸和一个或多个多核苷酸相对于所述孔移动时,获取一个或多个测量值,其中测量值代表多核苷酸的一个或多个特征,并由此表征所述靶多核苷酸。下文所述的任何实施方案适用于本方法。

[0194] 装载部分

[0195] 一个或多个多核苷酸结合蛋白被提供结合到(或连接到)一个或多个装载部分。在

一个优选的实施方案中,该方法还包括结合(或连接)所述一个或多个多核苷酸结合蛋白到所述一个或多个装载部分。

[0196] 每个装载部分可以是能够被连接到靶多核苷酸的任何部分。只要多核苷酸结合蛋白可结合且其能被连接到靶多核苷酸,每个装载部分可以是任何长度。

[0197] 所述一个或多个装载部分优选是合成的或人工的。所述一个或多个装载部分优选是非天然的。

[0198] 合适的装载部分包括,但不限于,聚合连接体、化学连接体、多核苷酸或多肽。所述一个或多个装载部分优选包含多核苷酸或装载多核苷酸。在此类实施方案中,一个或多个多核苷酸结合蛋白优选结合到(或连接到)多核苷酸。也可以使用以上所述的任何多核苷酸。优选地,所述一个或多个装载部分包含DNA、RNA、经修饰的DNA(如脱碱基的DNA)、RNA、PNA、LNA、BNA或PEG。所述一个或多个装载部分更优选包含单链或双链DNA或RNA。

[0199] 所述一个或多个装载部分优选包括所述一个或多个多核苷酸结合蛋白结合的(或连接的)的单链多核苷酸。

[0200] 至少一个所述一个或多个装载部分优选为Y衔接体。Y衔接体在关于双偶联的章节中定义,并且可以包括前导序列。

[0201] 至少一个所述一个或多个装载部分优选为桥连部分。所述桥连部分最优选为发夹环或发夹环衔接体。合适的发夹环衔接体可以使用本领域中已知的方法来设计。发夹环可以是任何长度。如果发夹环用作装载部分,其通常为400个或更少个核苷酸,如350个或更少个核苷酸、300个或更少个核苷酸、250个或更少个核苷酸、200个或更少个核苷酸、150个或更少个核苷酸、100个或更少个核苷酸、90或更少个核苷酸、80个或更少个核苷酸、70个或更少个核苷酸、60个或更少个核苷酸、50个或更少个核苷酸、40个或更少个核苷酸、30个或更少个核苷酸、20个或更少个核苷酸、或10个或更少个核苷酸的长度。发夹环优选为从约1个至400个、2个至300个、5个至200、6个至100个核苷酸的长度。当多核苷酸的两个互补部分杂合以形成双链序列(称为茎干)时,形成发夹环。如果用作装载部分,发夹环的茎干优选200个或更少个核苷酸对、例如150个或更少个核苷酸对、100个或更少个核苷酸对、90个或更少个核苷酸对、80个或更少个核苷酸对、70个或更少个核苷酸对、60个或更少个核苷酸对、50个或更少个核苷酸对、40个或更少个核苷酸对、30个或更少个核苷酸对、20个或更少个核苷酸对或10个或更少个核苷酸对的长度。所述一个或多个多核苷酸结合蛋白通常结合到发夹的环,即不是茎干。

[0202] 如果靶多核苷酸是双链的,所述一个或多个装载部分优选包含Y衔接体和可选地桥连部分,如发夹环衔接体。如果至少一个或多个所述装载部分是Y衔接体,它可与不具有任何多核苷酸结合蛋白结合或连接的桥接衔接体组合使用。

[0203] 如果所述一个或多个多核苷酸结合蛋白衍生自解旋酶,它们可停滞在所述一个或多个装载部分上的一个或多个间隔器处。下文进行更详细的讨论。

[0204] 可以使用任何数量的一个或多个装载部分。本方法可以包括连接两个或多个装载部分,其中每个装载部分具有与其结合(连接)的一个或多个多核苷酸结合蛋白。例如,装载部分可连接到靶多核苷酸的每一端。在此类实施方案中,一个装载部分优选为Y衔接体,且另一装载部分可以是桥连部分,如发夹环衔接体。这些将在下文更详细地讨论。

[0205] 所述一个或多个装载部分可以以任何方式连接到靶多核苷酸。所述一个或多个装

载部分优选共价地连接到靶多核苷酸。

[0206] 所述一个或多个装载部分最优选接合到靶多核苷酸。所述一个或多个装载部分可以接合到多核苷酸的任一端,即5'或3'末端。装载部分可被接合到靶多核苷酸的两端。所述一个或多个装载部分可使用本领域中已知的任何方法接合到多核苷酸。所述一个或多个装载部分可在没有ATP存在下或使用 γ -S-ATP(ATP γ S)代替ATP接合到多核苷酸。

[0207] 所述一个或多个装载部分可使用连接酶接合,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。连接酶优选在实施例3中设定的条件下使用。

[0208] 该方法优选进一步包括从本方法的条件中去除连接酶。

[0209] 一旦装载部分已经连接到靶多核苷酸,所述一个或多个多核苷酸结合蛋白优选保持结合(连接)到所述装载部分。根据本发明所述一个或多个多核苷酸结合蛋白被连接后,可以将它们从所述一个或多个装载部分解绑。

[0210] 膜

[0211] 根据本发明可使用任何膜。合适的膜是本领域公知的。所述膜优选为两亲层。两亲层是由具有亲水性和亲脂性的两亲分子形成的层,如磷脂。所述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在本领域中是已知的,且包括,例如,嵌段共聚物(Gonzalez-Perez等人,Langmuir,2009,25,10447-10450)。嵌段共聚物是其中两个或多个单体亚基一起聚合以生成单个聚合物链的聚合物材料。嵌段共聚物通常具有由每个单体亚基贡献的特性。然而,嵌段共聚物可具有由各个亚基形成的聚合物不具有的独特的特性。嵌段共聚物可以被加工,使得单体亚基中的一个是疏水性的(即亲脂的),而其他亚基在含水介质中是亲水性的。在这种情况下,该嵌段共聚物可具有两亲特性,并且可以形成模仿生物膜的结构。嵌段共聚物可以是二嵌段(由两个单体亚基组成),但也可以由多于两个的单体亚基构成以形成充当两亲物的更复杂的排列。所述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。

[0212] 古细菌双极四醚脂质是天然存在的脂质,其被构造为使得所述脂质形成单层膜。这些脂质通常在恶劣的生物环境、嗜热菌、嗜盐菌和嗜酸菌中生存的嗜极菌(extremophiles)中被发现。它们的稳定性被认为是从最终双层的融合性质中衍生的。比较简单的是,通过生成具有亲水性-疏水性-亲水性一般基序的三嵌段聚合物构造模仿这些生物实体的嵌段共聚物材料。所述材料可以形成行为类似于脂质双层且包含从囊泡到层状膜的一系列相行为的单体膜。由这些三嵌段共聚物形成的膜与生物脂质膜相比持有一些优点。因为三嵌段共聚物是合成的,因此可以仔细控制其确切构造以提供用于形成膜和与孔及其它蛋白相互作用所需的正确链长度和特性。

[0213] 嵌段共聚物也可以从不被分类为脂质材料的亚基构建;例如疏水性聚合物可从硅氧烷或其它非基于烃的单体制成。嵌段共聚物的亲水性子段也可具有低的蛋白质结合特性,这使得当暴露于未处理的生物样品时,其可生成高抗性的膜。该头部基团单元也可以从非经典的脂质头部基团衍生。

[0214] 三嵌段共聚物膜与生物脂质膜相比,还具有增强的机械稳定性和环境稳定性,例如高得多的操作温度或pH范围。嵌段共聚物的合成性质为定制用于广泛应用范围的基于聚合物的膜提供了平台(platform)。

[0215] 所述膜最优选为在国际申请No.PCT/GB2013/052766或PCT/GB2013/052767中公开的膜之一。

[0216] 所述两亲分子可以是经化学修饰的或官能化的以有助于多核苷酸的耦合。

[0217] 该两亲层可以是单层或双层。两亲层通常是平面的。两亲层可以是弯曲的。两亲层可以被支撑。

[0218] 两亲膜通常自然是移动的,本质上充当具有约 10^{-8}cm s^{-1} 的脂质扩散速率的二维流体。这意味着检测器和耦合的多核苷酸通常在两亲膜内移动。

[0219] 膜可以是脂质双层。脂质双层是细胞膜的模型,并充当用于一系列实验研究的优秀的平台。例如,脂质双层可通过单通道记录用于膜蛋白的体外研究。或者,脂质双层可用作生物传感器以检测一系列物质的存在。脂质双层可以是任何脂质双层。合适的脂质双层包括,但不限于,平面脂双层、支撑双层或脂质体。脂质双层优选平面脂质双层。合适的脂质双层在国际申请No.PCT/GB08/000563(公开为WO 2008/102121)、国际申请No.PCT/GB08/004127(公开为WO2009/077734)和国际申请No.PCT/GB2006/001057(公开为WO 2006/100484)中公开。

[0220] 用于形成脂质双层的方法在本领域中是已知的。合适的方法在实施例中公开。脂质双层通常由Montal和Mueller的方法(Proc.Natl.Acad.Sci.USA.,1972;69:3561-3566)共同来形成,其中脂质单层是在水溶液/空气界面上进行的,所述水溶液/空气界面穿过垂直于该界面的孔的任一侧。脂质通常通过首先将其在有机溶剂中溶解,然后使得一滴所述溶剂在孔的任一侧上的水溶液的表面上蒸发,而添加到电解质水溶液的表面。一旦有机溶剂蒸发了,在孔的任一侧上的溶液/空气界面物理地上下移动穿过所述孔,直到双层形成。平面脂质双层可跨越膜中的孔或跨越开口而形成凹部。

[0221] Montal和Mueller的方法是流行的,因为该方法是形成适于蛋白孔插入的良好品质脂质双层的成本合算和相对简单的方法。双层形成的其它常用方法包括脂质体双层的尖部浸渍,喷涂双层和膜片钳。

[0222] 尖部浸渍形成双层必需使孔表面(例如,移液管尖部)接触到承载脂质单层的测试溶液的表面。另外,首先通过将一滴溶解在有机溶剂中的脂质在溶液表面蒸发而在溶液/空气界面处产生脂质单层。然后由Langmuir-Schaefer工艺形成所述双层,并且所述双层需要机械自动化以相对于溶液表面移动所述孔。

[0223] 对于喷涂双层,将溶解在有机溶剂中的脂质的液滴直接施加到所述孔,该孔浸没在水性测试溶液中。使用画笔或等效物将脂质溶液薄薄地涂在孔上。溶剂的稀释(thinning)导致脂质双层的形成。然而,从双层完全去除溶剂是困难的,并因此由该方法形成的双层是较不稳定的且在电化学测量过程中更容易有噪声。

[0224] 膜片钳通常在生物细胞膜的研究中使用。细胞膜通过抽吸被夹到移液管的端部,且膜的小块贴附到孔上。该方法已被改为用于通过夹住脂质体,然后脂质体爆裂而离开移液管的孔上的脂质双层密封,而产生脂质双层。该方法需要稳定的、巨大的和单层的脂质体,以及需要在具有玻璃表面的材料中制造小孔。

[0225] 脂质体可通过超声、挤出或Mozafari法(Colas等人(2007)Micron 38:841-847)来形成。

[0226] 在优选的实施方案中,脂质双层如在国际申请No.PCT/GB08/004127(公开为

W02009/077734) 中描述的来形成。本方法中有利的是, 脂质双层是由干燥的脂质形成的。在最优选的实施方案中, 脂质双层跨越开口而形成, 如在W02009/077734 (PCT/GB08/004127) 中描述的。

[0227] 脂质双层是由脂质的两个相对的层形成的。脂质的两层被排列为使得它们的疏水性尾部基团面向彼此以形成疏水性内部。脂质的亲水性头部基团面向外朝向该双层的每一侧的含水环境。该双层可存在于许多脂质相中, 所述脂质相包括但不限于, 液态无序相(流体层状)、液态有序相、固态有序相(层状凝胶相、相互交叉的凝胶相)和平面双层晶体(层状子凝胶相、层状结晶相)。

[0228] 可使用形成脂质双层的任何脂质组合物。脂质组合物被选择为, 使得形成的脂双层具有所需的特性, 例如表面电荷、支撑膜蛋白的能力、填充密度或机械特性。脂质组合物可包含一个或多个不同的脂质。例如, 脂质组合物可含有多达100个脂质。脂质组合物优选含有1至10个脂质。脂质组合物可以包括天然存在的脂质和/或人工脂质。

[0229] 脂质通常包括头部基团、界面部分和可以相同或不同的两个疏水尾基。合适的头部基团包括, 但不限于, 中性头部基团, 如二酰基甘油酯(DG)和神经酰胺(CM); 两性离子头部基团, 如磷脂酰胆碱(PC)、磷脂酰乙醇胺(PE)和鞘磷脂(SM); 带负电荷的头部基团, 如磷脂酰甘油(PG); 磷脂酰丝氨酸(PS)、磷脂酰肌醇(PI)、磷酸(PA)和心磷脂(CA); 和带正电荷的头部基团, 如三甲基铵-丙烷(TAP)。合适的界面部分包括, 但不限于, 天然存在的界面部分, 例如基于甘油或基于神经酰胺的部分。合适的疏水性尾部基团包括, 但不限于, 饱和烃链, 如月桂酸(正十二烷酸)、肉豆蔻酸(正十四烷酸)、棕榈酸(正十六烷酸)、硬脂酸(正十八烷酸)和花生酸(正二十烷酸); 不饱和烃链, 如油酸(顺-9-十八烷酸); 和支链烃链, 如植烷酰(phytanoyl)。链的长度和不饱和烃链中双键的位置和数量可以改变。链的长度和在支链烃链中的支链如甲基基团的位置和数目可以改变。疏水性尾部基团可以被连接到作为醚或酯的界面部分。脂质可以是霉菌酸。

[0230] 脂质也可以是被化学修饰的。脂质的头部基团或尾部基团可以是化学修饰的。头部基团已被化学修饰的合适的脂质包括, 但不限于, PEG修饰的脂质, 例如1,2-二酰基-sn-甘油基-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000]; 官能化的PEG脂质, 如1,2-二硬脂酰-sn-甘油基-3-磷酸乙醇胺-N-[生物素基(聚乙二醇)2000]; 为共轭而修饰的脂质, 如1,2-二油酰-sn-甘油基-3-磷酸乙醇胺-N-(琥珀酰基)和1,2-二棕榈酰基-sn-甘油基-3-磷酸乙醇胺-N-(生物素基)。尾部基团已被化学修饰的合适的脂质包括, 但不限于, 可聚合的脂质, 如1,2-双(10,12-二十三二炔酰)-sn甘油基-3-磷酸胆碱; 氟化脂质, 例如1-棕榈酰基-2-(16-氟代棕榈酰基)-sn甘油基-3-磷酸胆碱; 氯化脂质, 如1,2-二棕榈酰基-D62--sn-甘油基-3-磷酸胆碱; 和醚连接的脂质, 如1,2-二-0-植烷基-sn-甘油基-3-磷酸胆碱。脂质可以是化学修饰的或官能化的, 以有助于多核苷酸的耦合。

[0231] 两亲层例如, 脂质组合物, 通常包含将影响该层的性能的一个或多个添加剂。合适的添加剂包括, 但不限于, 脂肪酸, 如棕榈酸、肉豆蔻酸和油酸; 脂肪醇, 如棕榈醇、肉豆蔻醇和油醇; 甾醇, 如胆固醇、麦角固醇、羊毛甾醇, 谷甾醇和豆甾醇; 溶血磷脂, 如1-酰基-2-羟基-sn-甘油基-3-磷酸胆碱; 和神经酰胺。

[0232] 在另一个优选的实施方案中, 所述膜为固态层。固态层可以由有机和无机材料形成, 该有机和无机材料包括但不限于, 微电子材料、绝缘材料, 如Si₃N₄、Al₂O₃和SiO, 有机和

无机聚合物,例如聚酰胺、塑料如Teflon®,或弹性体如双组分加成固化硅橡胶,和玻璃。固态层可以由石墨烯形成。合适的石墨烯层在国际申请No.PCT/US2008/010637(公开为W02009/035647)中公开。

[0233] 该方法通常使用(i)包含孔的人工两亲层, (ii)分离的、包含孔的天然存在的脂质双层,或(iii)具有嵌入其中的孔的细胞进行。该方法通常使用人工两亲层进行,如人工三嵌段共聚物层。除了孔以外,所述层可包含其它跨膜和/或膜内蛋白以及其他分子。合适的设备和条件在下面讨论。本发明的方法通常在体外实施。

[0234] **耦合**

[0235] 所述一个或多个装载部分优选包含一个或多个能够耦合到膜的锚。根据本发明,一旦被连接,所述一个或多个锚能够将靶多核苷酸耦合到膜。

[0236] 在本发明的表征方法中,靶多核苷酸优选通过使用一个或多个锚耦合到膜。靶多核苷酸可以使用任何已知的方法耦合到膜。

[0237] 每个锚包含耦合(或结合)到所述一个或多个装载部分的基团和耦合(或结合)到膜上的基团。每个锚可以共价耦合(或结合)到所述部分和/或所述膜。

[0238] 每个部分可包含任何数量的锚,如2、3、4个或更多个锚。例如,一个靶多核苷酸可以使用两个锚耦合到所述膜,所述两个锚各自经部分和膜单独耦合(或结合)到两个多核苷酸。

[0239] 如果膜是两亲层,如共聚物膜或脂质双层,则所述一个或多个锚优选地包括在所述膜中存在的多肽锚和/或在所述膜中存在的疏水锚。所述疏水锚优选是脂质、脂肪酸、甾醇、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。在优选的实施方案中,所述一个或多个锚不是检测器。

[0240] 膜的组分,如两亲分子、共聚物或脂质,可以是化学修饰的或官能化的,以形成一个或多个锚。合适的化学修饰和合适的官能化膜的组分的方式的实例在下面更详细地讨论。任何比例的膜组分可以被官能化,例如至少0.01%、至少0.1%、至少1%、至少10%、至少25%、至少50%或100%。

[0241] 用于耦合多核苷酸到膜的一个或多个锚优选包含连接体。所述一个或多个锚可包含一个或多个连接体,例如2、3、4或更多个。一个连接体可用于耦合多于1个,例如2、3、4或更多个多核苷酸到膜上。

[0242] 优选的连接体包括,但不限于,聚合物,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的、支链的或环状的。例如,所述连接体可以是环形的多核苷酸。所述多核苷酸可杂合到环状多核苷酸连接体上的互补序列。

[0243] 所述一个或多个锚或一个或多个连接体可包含可切割或可分解的组分,如限制性位点或光不稳定基团。

[0244] 官能化的连接体以及它们耦合分子的方式在本领域中是已知的。例如,用马来酰亚胺基团官能化的连接体将与蛋白质中的半胱氨酸残基反应并连接到其上。在本发明的上下文中,该蛋白质可以存在于该膜中,或者可以用来耦合(或结合)到所述一个或多个装载部分。这在下面更详细地讨论。

[0245] 可使用“锁和钥匙”排列避免多核苷酸的交联。各连接体仅一端可一起反应以形成更长的连接体,且连接体的其他端各自分别与装载部分或膜反应。此类连接体在国际申请No.PCT/GB10/000132(公开为W0 2010/086602)中描述。

[0246] 连接体优选用在下文讨论的测序实施方案中。如果多核苷酸被永久地直接耦合到所述膜, 意义在于当多核苷酸与检测器相互作用时, 多核苷酸不解耦(即在步骤(b)或(e)中不解耦), 则由于所述膜和检测器之间的距离, 当测序运行不能继续至多核苷酸的末端时, 一些序列数据将丢失。如果使用连接体, 则多核苷酸可以进行到完成为止。

[0247] 所述耦合可以是永久的或稳定的。换言之, 耦合可以是, 使得当多核苷酸与所述孔相互作用时, 多核苷酸保持耦合到膜上。

[0248] 该耦合可以是短暂的。换言之, 耦合可以是, 使得当多核苷酸与所述孔相互作用时, 多核苷酸可以去耦。

[0249] 对于某些应用, 如适体(aptamer)检测, 耦合的短暂性质是优选的。如果永久的或稳定的连接体直接连接到多核苷酸的5'或3'末端, 并且连接体比膜和跨膜孔的通道之间的距离更短, 则当测序运行不能继续至多核苷酸的末端, 一些序列数据将丢失。如果耦合是短暂的, 那么当耦合的末端随机地变得没有膜时, 该多核苷酸可以被加工至完成为止。形成永久的/稳定的或短暂的链接的化学基团在下面更详细地讨论。所述多核苷酸可使用胆固醇或脂肪酰链短暂地耦合到两亲层或三嵌段共聚物膜。可以使用具有6至30个碳原子的长度的任何脂肪酰基链, 例如十六烷酸。

[0250] 在优选的实施方案中, 一个或多个锚能够被耦合到两亲层, 如三嵌段共聚物膜或脂质双层。将核酸耦合到合成脂双层已经先前用多种不同的拴系策略实施。这些总结于下表1中。

[0251] 表1

锚包含	耦合类型	参考文件
[0252]	硫醇	Yoshina-Ishii, C. and S. G. Boxer (2003). "Arrays of mobile tethered vesicles on supported lipid bilayers." <i>J Am Chem Soc</i> 125 (13): 3696-7.
	生物素	Nikolov, V., R. Lipowsky, et al. (2007). "Behavior of giant vesicles with anchored DNA molecules." <i>Biophys J</i> 92 (12): 4356-68
[0253]	胆固醇	Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies." <i>J Am Chem Soc</i> 126 (33): 10224-5
	表面活性剂(例如脂质, 棕榈酸酯等)	van Lengerich, B., R. J. Rawle, et al. "Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." <i>Langmuir</i> 26 (11): 8666-72

[0254] 合成的多核苷酸和/或连接体可使用在合成反应中经修饰的亚磷酰胺官能化, 所述经修饰的亚磷酰胺与用于直接加成合适的锚固基团, 如胆固醇、生育酚、棕榈酸、硫醇、脂质和生物素基团是容易兼容的。这些不同的连接化学物给出了用于连接装载部分的一套选择。每个不同的修饰基团以略微不同的方式耦合装载部分, 并且耦合并不总是永久的, 所以给予了连接到膜的多核苷酸以不同的停留时间(dwell time)。短暂的耦合的优点如上所述。

[0255] 将多核苷酸耦合到连接体, 或将多核苷酸耦合到官能化的膜还可以通过许多其他手段实现, 条件是互补的反应性基团或锚固基团被加成到所述多核苷酸装载部分。将反应

性基团加成到多核苷酸的任一端在以前已经有报道。硫醇基团可以使用T4多核苷酸激酶和ATP γ S加成到ssDNA或dsDNA的5'末端(Grant, G.P. and P.Z. Qin (2007). "在核酸的5'末端连接硝基氧自旋标记的简便方法", *Nucleic Acids Res* 35 (10) :e77)。叠氮基可以用T4多核苷酸激酶和 γ -[2-叠氮基乙基]-ATP或 γ -[6-叠氮己基]-ATP加成到ssDNA或dsDNA的5'-磷酸酯。使用硫醇或点击化学(Click chemistry), 将含有硫醇、碘乙酰胺OPSS或马来酰亚胺基团(对硫醇具有反应性)或DIBO(二苯并环辛炔(dibenzocyclooctyne))或炔基(对叠氮化物具有反应性)的系链共价连接到多核苷酸装载部分。更多样化选择的化学基团, 如生物素、硫醇和荧光团, 可以使用末端转移酶进行加成, 以结合修饰的寡核苷酸到ssDNA的3'末端(Kumar, A., P.Tchen, et al. (1988). "以末端脱氧核苷酸转移酶非放射性标记合成的寡核苷酸探针" *Anal Biochem* 169 (2) :376-82)。链霉亲和素/生物素和/或链霉亲和素/脱硫生物素耦合可用于任何其它的多核苷酸装载部分。也可以将锚使用具有经合适修饰的核苷酸(例如, 胆固醇或棕榈酸酯)的末端转移酶直接加成到多核苷酸装载部分。

[0256] 所述一个或多个锚优选能够将装载部分/多核苷酸经杂合偶联到膜。一个或多个锚中的杂合可使以如上所述的短暂停方式耦合。杂合可以存在于一个或多个锚的任何部分中, 例如一个或多个锚与装载部分之间、所述一个或多个锚内或在一个或多个锚和膜之间。例如, 连接体可以包含杂合在一起的两个或更多个多核苷酸, 如3、4或5个多核苷酸。所述一个或多个锚可以杂合到多核苷酸装载部分。所述一个或多个锚可以直接地杂合到多核苷酸装载部分, 如直接杂合到Y衔接体和/或前导序列, 或直接杂合到桥连部分衔接体, 如发夹环衔接体(如下面讨论的)。或者, 一个或多个锚可以杂合到一个或多个, 例如2或3个中间多核苷酸(或“夹板体(splint)”), 所述中间多核苷酸杂合到多核苷酸装载部分, 或杂合到Y衔接体和/或前导序列, 或杂合到连接到多核苷酸的桥连部分衔接体(如下面讨论)。

[0257] 所述一个或多个锚可包括单链多核苷酸或双链多核苷酸。锚的一部分可以接合到单链或双链多核苷酸装载部分。使用T4RNA连接酶I接合ssDNA的短片段已经被报道(Troutt, A.B., M.G. McHeyzer-Williams等人(1992). "接合-锚定PCR技术: 具有单侧特异性的简单扩增技术." *Proc Natl Acad Sci U S A* 89 (20) :9823-5)。或者, 单链或双链多核苷酸可以接合到双链多核苷酸, 然后通过热变性或化学变性分离两条链。对于双链靶多核苷酸, 可以将单链多核苷酸的小段添加到该双链体的一端或两端, 或将一双链多核苷酸添加到其一端或两端。对于将单链多核苷酸添加到双链多核苷酸, 可以使用用于连接到单链多核苷酸的其他区域的T4RNA连接酶I来实现。对于将双链多核苷酸添加到双链多核苷酸随后接合可以用分别在多核苷酸和添加的多核苷酸上的互补的3'dA/dT尾部(如同常规地对许多样品制备应用所做的, 以防止多联体或者二聚体形成)或使用由多核苷酸的限制消化和相容衔接体的接合而产生的“粘性末端”, 进行“钝化终止(blunt-ended)”。然后, 当双链体熔化时, 每条单链将具有或5'或3'修饰, 如果使用单链多核苷酸进行连接或修饰, 则在5'末端或3'末端进行修饰, 或如果使用双链多核苷酸进行接合, 则在这两端进行修饰。

[0258] 如果装载部分是合成的, 所述一个或多个锚可以在其化学合成过程中并入。例如, 多核苷酸装载部分可以使用具有连接到它的反应性基团的引物来合成。

[0259] 腺苷酰化的多核苷酸是接合反应中的中间体, 其中单磷酸腺苷连接到多核苷酸的5'-磷酸。多种试剂盒可用于产生该中间体, 如来自NEB的5'-DNA腺苷酸化试剂盒。通过在反应中用ATP替代修饰的核苷酸三磷酸, 可以将反应性基团(例如硫醇、胺、生物素、叠氮化物

等)加成到多核苷酸的5'。也可以使用具有适当被修饰的核苷酸(例如,胆固醇或棕榈酸酯)的5'DNA腺苷酰化试剂盒将锚直接加成到多核苷酸装载部分。

[0260] 用于基因组DNA的片段的扩增的常用技术是使用聚合酶链反应(PCR)。本文中,使用两个合成寡核苷酸引物可以生成许多DNA的相同片段的拷贝,其中对于每个拷贝,双链体中每条链的5'将是合成的多核苷酸。单个或多个核苷酸可以通过使用聚合酶被添加到单链或双链的DNA的3'末端。可使用的聚合酶的实例包括,但不限于,末端转移酶、Klenow和大肠杆菌多(A)聚合酶。通过在反应中用ATP替代被修饰的核苷酸三磷酸,则可将锚,例如胆固醇、硫醇、胺、叠氮化物、生物素或脂质,结合至双链多核苷酸装载部分中。因此,扩增的多核苷酸装载部分的每个拷贝将包含锚。

[0261] 理想化的,装载部分/多核苷酸耦合到膜,而无需必须将所述装载部分/多核苷酸官能化。这可以通过耦合一个或多个锚,如多核苷酸结合蛋白质或化学基团到膜,并使得所述一个或多个锚与装载部分/多核苷酸相互作用或通过官能化所述膜来实现。所述一个或多个锚可以通过本文所述的任何方法被耦合到膜上。特别地,所述一个或多个锚可以包含一个或多个连接体,如马来酰亚胺官能化的连接体。

[0262] 所述一个或多个锚可包含耦合到、结合到下述物质或与之相互作用:单链或双链多核苷酸、装载部分内特定的核苷酸序列、或装载部分内经修饰核苷酸的模式,或在装载部分上存在的任何其它配体。

[0263] 在锚中使用的合适的结合蛋白包括但不限于,大肠杆菌单链结合蛋白、P5单链结合蛋白、T4gp32单链结合蛋白、TOP0 V dsDNA结合区、人类组蛋白、大肠杆菌HU DNA结合蛋白和其他古细菌、原核或真核的单链或双链多核苷酸(或核酸)结合蛋白,包括下文列出的那些。

[0264] 装载部分中的特定的核苷酸序列可以是由转录因子、核糖体、核酸内切酶、拓扑异构酶或复制起始因子识别的序列。修饰的核苷酸的模式可以是甲基化模式或损伤模式。

[0265] 所述一个或多个锚可包含耦合到、结合到、嵌入多核苷酸装载部分或与多核苷酸装载部分相互反应的任何基团。该基团可以通过静电结合、氢键结合或范德华相互作用而嵌入或与多核苷酸相互作用。这类基团包括赖氨酸单体,聚赖氨酸(其将与ssDNA或dsDNA相互作用),溴化乙锭(其将嵌入dsDNA),通用碱基或通用核苷酸(可以与任何多核苷酸杂合)和锇络合物(可以对甲基化碱基反应)。因此多核苷酸装载部分可使用连接到膜的一个或多个通用核苷酸耦合到膜。每个通用核苷酸可以使用一个或多个连接体耦合到膜。通用核苷酸优选包括以下的核碱基中的一个:次黄嘌呤、4-硝基吲哚、5-硝基吲哚、6-硝基吲哚、甲酰基吲哚、3-硝基吡咯、硝基咪唑、4-硝基吡唑、4-硝基苯并咪唑、5-硝基吲唑、4-氨基苯并咪唑或苯基(C 6-芳族环)。通用核苷酸更优选包括下列核苷中的一个:2'-脱氧肌苷,肌苷、7-脱氮杂-2'-脱氧肌苷、7-脱氮杂-肌苷、2-氮杂-脱氧肌苷、2-氮杂-肌苷、2-0'-甲基肌苷、4-硝基吲哚2'-脱氧核糖核苷、4-硝基吲哚核糖核苷、5-硝基吲哚-2'-脱氧核糖核苷、5-硝基吲哚核糖核苷、6-硝基吲哚-2'-脱氧核糖核苷、6-硝基吲哚核糖核苷、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-硝基吲唑2'-脱氧核糖核苷、5-硝基吲唑核糖核苷、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'-脱氧核糖基核苷、2'-脱氧水粉蕈素、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。通用核苷酸最优先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)。

[0266] 所述一个或多个锚可以通过Hoogsteen氢键(其中两个核碱基通过氢键保持在一起)或反向Hoogsteen氢键(其中一个核碱基相对于另一个核碱基旋转180°)耦合到(或结合到)多核苷酸装载部分。例如,一个或多个锚可以包含一个或多个核苷酸、一个或多个寡核苷酸,或与多核苷酸装载部分形成Hoogsteen氢键或反向Hoogsteen氢键的一个或多个多核苷酸。这种类型的氢键允许第三个核苷酸链绕双链螺旋卷绕并形成三链体。所述一个或多个锚可以通过用双链二链体形成三链体而耦合到(或结合到)双链多核苷酸装载部分。

[0267] 在本实施方案中,至少1%、至少10%、至少25%、至少50%或100%的膜组分可以被官能化。

[0268] 如果所述一个或多个锚包含蛋白,它们无需进一步官能化,能够直接锚固到膜中,例如,如果所述蛋白已经具有一个与膜相容的外部疏水区时。这类蛋白的实例包括,但不限于,跨膜蛋白、膜内蛋白和膜蛋白。或者,所述蛋白可以用与膜相容的遗传融合的疏水区来表达。这种疏水蛋白区在本领域中是公知的。

[0269] 所述一个或多个锚优选在与膜接触之前,与一个或多个装载部分混合,但是所述一个或多个锚也可以与膜接触,然后与一个或多个装载部分接触。

[0270] 在另一个方面,装载部分可使用上述方法官能化,以使其可以被特定的结合基团识别。具体地,所述装载部分也可以用配体如生物素(用于结合到链霉亲和素)、直链淀粉(用于结合到麦芽糖结合蛋白或融合蛋白)、Ni-NTA(用于结合到聚组氨酸或聚组氨酸标记的蛋白)或肽(如抗原)官能化。

[0271] 根据优选的实施方案,当多核苷酸连接到包含优先螺入孔中的前导序列的装载部分上时,所述一个或多个锚可用于将多核苷酸耦合到膜上。前导序列是在下面更详细地讨论。优选地,该多核苷酸被连接到(例如接合到)优先螺入孔中的前导序列。这种前导序列可以包含均聚多核苷酸或脱碱基区。前导序列通常设计为,被直接杂合到或由一个或多个中间体多核苷酸(或夹板体)杂合到所述一个或多个锚。在这种情况下,所述一个或多个锚通常包含与前导序列中的序列或一个或多个中间体多核苷酸(或夹板体)中的序列互补的多核苷酸序列。在这种情况下,所述一个或多个夹板体通常包含与前导序列中的序列互补的多核苷酸序列。

[0272] 在化学连接中使用的分子的实例是EDC (1-乙基-3-[3-二甲基氨基丙基]碳二亚胺盐酸盐)。反应性基团也可以使用商购的试剂盒(Thermo Pierce, 货号No. 22980)被加成到多核苷酸的5'末端。合适的方法包括,但不限于,使用组氨酸残基和Ni-NTA进行的短暂亲和连接,以及由反应性半胱氨酸、赖氨酸或非天然氨基酸进行的更牢固的共价连接。

[0273] 跨膜孔

[0274] 本发明的表征方法包括当靶多核苷酸相对于孔移动时,获取一个或多个测量值。可使用孔得到多种不同类型的测量值。这包括但不限于:电测量和光学测量。可能的电测量包括:电流测量、阻抗测量、风洞测量(Ivanov AP et al., Nano Lett. 2011 Jan; 11 (1): 279-85)以及FET测量(国际申请号WO 2005/124888)。光学测量可以与电测量组合(Soni GV et al., Rev Sci Instrum. 2010 Jan; 81 (1): 014301)。所述测量值可以是跨膜电流测量值,如流经孔的离子电流的测量值。

[0275] 电测量值可以使用在Stoddart D et al., Proc Natl Acad Sci, 12;106 (19) : 7702-7, Lieberman KR et al, J Am Chem Soc. 2010; 132 (50) : 17961-72, 和国际申请WO 2000/28312中所描述的标准的单通道记录设备获得。或者,电测量可以使用多通道系统进行,例如国际申请WO 2009/077734和国际申请号WO 2011/067559中描述的。

[0276] 本方法优选用跨膜施加发电势实施。所施加的电势可以是电压电势。或者,所施加的电势可以是化学势。这方面的实例是跨膜,如两亲层,使用盐梯度。盐梯度在Holden et al., J Am Chem Soc. 2007 Jul 11; 129 (27) : 8650-5中公开。在一些情况下,当多核苷酸相对于孔移动时,使用流经检测器(或孔)的电流来评估或确定多核苷酸的序列。这就是链测序。

[0277] 该方法包括将靶多核苷酸与跨膜孔接触。跨膜孔是在一定程度上穿过所述膜的结构。其允许由施加的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,使得水合离子可从膜的一侧流到膜的另一侧。然而,跨膜孔不是必须穿过膜。它可以是一端闭合的。例如,所述孔可以是位于水合离子沿其流动或流入其中的膜中的孔道、间隙、通道、沟槽或狭缝。

[0278] 任何跨膜孔可以在本发明中使用。所述孔可以是生物的或人工的。合适的孔包括但不限于,蛋白孔、多核苷酸孔和固态孔。所述孔可以是DNA折纸(origami)孔(Langecker et al., Science, 2012; 338:932-936)。

[0279] 跨膜孔优选为跨膜蛋白孔。跨膜蛋白孔是允许水合离子如分析物从膜的一侧流向膜的另一侧的多肽或多肽的集合。在本发明中,跨膜蛋白孔能够形成孔,所述孔允许由施加的电势驱动的水合离子从膜的一侧流到膜的另一侧。跨膜蛋白质孔优选允许分析物如核苷酸从膜如三嵌段共聚物膜的一侧流动到另一侧。跨膜蛋白孔允许多核苷酸如DNA或RNA,相对于孔移动,如穿过所述孔。

[0280] 跨膜蛋白质孔可以是单体或低聚物。孔优选由几个重复的亚基,例如至少6个、至少7、至少8或至少9个亚基组成。孔优选是六聚体的、七聚体的、八聚体的或九聚体的孔。所述孔可以是同源低聚物或异源低聚物。

[0281] 跨膜蛋白质孔通常包含该离子可流动通过的桶状体或通道。孔的亚基通常围绕中心轴并且将链贡献到跨膜 β 桶状体或通道或跨膜 α 螺旋束或通道。

[0282] 跨膜蛋白孔的桶状体或通道通常包含有助于与分析物如核苷酸、多核苷酸或核酸相互作用的氨基酸。这些氨基酸优选定位于靠近桶状体或通道的缢痕(constriiction)。跨膜蛋白孔通常包括一个或多个带正电荷的氨基酸,例如精氨酸、赖氨酸或组氨酸,或芳族氨基酸,例如酪氨酸或色氨酸。这些氨基酸通常有助于孔和核苷酸、多核苷酸或核酸之间的相互作用。

[0283] 本发明使用的跨膜蛋白孔可衍生自 β -桶状体孔或 α -螺旋束孔。 β -桶状体孔包含由 β -链形成的桶状体或通道。合适的 β -桶状体孔包括,但不限于, β -毒素、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,如耻垢分枝杆菌(*Mycobacterium smegmatis*)孔蛋白(Msp),例如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和奈瑟球菌属(*Neisseria*)自转运脂蛋白(NalP)以及其他孔,如胞溶素(lysenin)。 α -螺旋束孔包含由 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,如WZA和ClyA毒素。跨膜孔可以从lysenin衍生。从胞溶素衍生的合适的孔在国际申请No. PCT/GB2013/050667(公开为WWO 2013/153359)中公开。跨膜孔

可以衍生自Msp或 α -溶血素(α -HL)。

[0284] 跨膜蛋白孔优选衍生自Msp, 优选MspA。该孔为低聚的且通常包含衍生自Msp的7个, 8个, 9个或10个单体。所述孔可以是衍生自含相同单体的Msp的同源寡聚孔。或者, 所述孔可以是从含至少一个与其他单体不同的单体的Msp衍生的异源寡聚孔。优选地, 所述孔衍生自MspA或其同源物或并系同源物(paralog)。

[0285] 衍生自Msp的单体通常包含SEQ ID NO:2所示的序列或其变体。SEQ ID NO:2为MspA单体的MS-(B1)8突变体。其包括以下突变:D90N, D91N, D93N, D118R, D134R和E139K。SEQ ID NO:2的变体是多肽, 所述多肽具有从SEQ ID NO:2的氨基酸序列变化而来且保留了其形成孔的能力的氨基酸序列。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, 变体可以连同其他合适的亚基被插入到两亲层中, 且可以确定其寡聚形成孔的能力。将亚基插入膜例如两亲膜中的方法是本领域已知的。例如, 亚基可以以纯化的形式悬浮在含有三嵌段共聚体膜的溶液中, 从而使其扩散到膜并通过结合而插入膜中, 并组装成功能状态。或者, 可以使用如M.A.Holden, H.Bayley. J. Am. Chem. Soc. 2005, 127, 6502-6503和国际申请No.PCT/GB2006/001057(公开为W0 2006/100484)中所描述的“拾取和放置”方法将亚基直接插入膜中。

[0286] 对于SEQ ID NO:2的氨基酸序列的整个长度, 基于氨基酸同一性, 变体优选与该序列具有至少50%的同源性。更优选地, 基于氨基酸同一性, 变体可与SEQ ID NO:2的氨基酸序列的整个序列具有至少55%, 至少60%, 至少65%, 至少70%, 至少75%, 至少80%, 至少85%, 至少90%, 更优选至少95%, 97%或99%的同源性。在100或更多, 例如125, 150, 175或200或更多的连续氨基酸长度上, 以具有至少为80%, 例如至少85%, 90%或95%的氨基酸同一性(严格同源性)。

[0287] 可使用本领域的标准方法确定同源性。例如, UWGCG软件包提供BESTFIT程序, 该程序可以用来计算同源性, 例如用于其默认设置(Devereux et al (1984) Nucleic Acids Research 12, p387-395)。PILEUP和BLAST算法可以用来计算同源性或比对序列(例如鉴定当量残基或相应的序列(通常在它们的默认设置下)), 如Altschul S.F. (1993) J Mol Evol 36:290-300; Altschul, S.F et al (1990) J Mol Biol 215:403-10中所描述的。公众可通过国家生物技术信息中心获得用于进行BLAST分析的软件(<http://www.ncbi.nlm.nih.gov/>)。

[0288] SEQ ID NO:2是MspA单体的MS-(B1)8突变体。该变体与MspA比较, 可包含MspB, MspC或MspD单体中的任何突变。MspB, MspC和MspD的成熟形式在SEQ ID NO:5至SEQ ID NO:7中示出。特别地, 所述变体可以包含存在于MspB中的下列取代:A138P。所述变体可以包含存在于MspC中的下列取代中的一个或多个:A96G, N102E和A138P。所述变体可包含存在于MspD中的下列突变中的一个或多个:G1缺失, L2V, E5Q, L8V, D13G, W21A, D22E, K47T, I49H, I68V, D91G, A96Q, N102D, S103T, V104I, S136K和G141A。所述变体可以包含来自Msp B, Msp C和Msp D的突变体和取代中的一个或多个的组合。所述变体优选包含突变L88N。SEQ ID NO:2的变体除了具有MS-B1的所有突变外还具有突变L88N, 并被称为MS-(B2)8。在本发明中使用的孔优选为MS-(B2)8。SEQ ID NO:2的变体除了具有MS-B1的所有突变外, 还具有突变G75S/G77S/L88N/Q126R, 并被称为MS-B2C。在本发明使用的孔优选MS-(B2)8或MS-(B2C)8。

[0289] 除了以上讨论的那些, 还可对SEQ ID NO:2的氨基酸序列进行氨基酸取代, 例如高

达1,2,3,4,5,10,20或30个取代。保守取代用相似化学结构,相似化学性质或相似侧链体积的其他氨基酸替代氨基酸。引入的氨基酸可以与它们要替代的氨基酸具有相似的极性,亲水性,疏水性,碱性,酸性,中性或带电性。或者,所述保守取代可在预先存在的芳香族或脂肪族氨基酸的位置引入其他芳香族或脂肪族氨基酸。保守氨基酸的改变是本领域公知的,并且可以根据下表2中所定义的20种主要氨基酸的特性进行选择。如果氨基酸具有相似极性,这也可以通过参考表3中的氨基酸侧链的亲水性大小而确定。

[0290] 表2-氨基酸的化学性质

[0291]

Ala	脂肪族,疏水,中性	Met	疏水,中性
Cys	极性,疏水,中性	Asn	极性,亲水,中性
Asp	极性,亲水,带电荷 (-)	Pro	疏水,中性
Glu	极性,亲水,带电荷 (-)	Gln	极性,亲水,中性
Phe	芳香族,疏水,中性	Arg	极性,亲水,带电荷 (+)
Gly	脂肪族,中性	Ser	极性,亲水性,中性
His	芳香族,极性,亲水性,带电荷 (+)	Thr	极性,亲水性,中性
Ile	脂肪族,疏水,中性	Val	脂肪族,疏水,中性
Lys	极性,亲水,带电荷 (+)	Trp	芳香族,疏水,中性
Leu	脂肪族,疏水,中性	Tyr	芳香族,极性,疏水

[0292]

表3-亲水性大小

侧链	亲水性值
----	------

[0293]

Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Trp	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

[0294]

[0295] SEQ ID NO:2的氨基酸序列的一个或多个氨基酸残基可另外从上述多肽中缺失。可缺失高达1,2,3,4,5,10,20或30个或更多个氨基酸残基。

[0296] 变体可包括SEQ ID NO:2的片段。这些片段保持孔形成活性。片段可以至少为50, 100, 150或200个氨基酸长度。这类片段可用于产生孔。片段优选包含SEQ ID NO:2的孔形成域。片段必须包含SEQ ID NO:2的残基88、90、91、105、118和134之一。通常, 片段包括SEQ ID NO:2的所有残基88、90、91、105、118和134。

[0297] 一个或多个氨基酸可以替代地或另外地添加到上述多肽。可在SEQ ID NO:2或多肽变体或其片段的氨基酸序列的氨基末端或羧基末端提供延伸。所述延伸可以很短, 例如, 1到10个氨基酸长度。或者, 延伸可以更长, 例如长达50或100个氨基酸。载体蛋白可以根据本发明被融合到氨基酸序列。其他融合蛋白在下文中更详细地讨论。

[0298] 如上所讨论的, 变体为多肽, 所述多肽具有从SEQ ID NO:2的氨基酸序列变化而来的氨基酸序列且保留其形成孔能力。变体通常包含负责孔形成的SEQ ID NO:2的区域。包含 β -桶状体的Msp的孔形成能力由每个亚基中的 β -片体(β -sheet)提供。SEQ ID NO:2的变体通常包含SEQ ID NO:2中形成 β -片体的区域。可对SEQ ID NO:2中形成 β -片体的区域进行一个或多个修饰, 只要得到的变体保留了其形成孔的能力即可。SEQ ID NO:2的变体优选地包括在 α -螺旋和/或环状区域内的一个或多个修饰, 如取代, 添加或缺失。

[0299] 衍生自Msp的单体可以被修饰以辅助它们的鉴定或纯化, 例如通过添加组氨酸残基(hist标签), 天冬氨酸残基(asp标签), 链霉亲和素标签或Flag标签, 或通过添加信号序列以促进它们从细胞中分泌, 该细胞中的多肽不天然地含有该信号序列。引入遗传标签的一种替代方式是通过化学反应将标签连到孔上天然或工程位点。其例子是将凝胶迁移试剂与工程连接在孔外的半胱氨酸进行反应。这已被证明是用于分离溶血素异源寡聚物的方法(Chem Biol. 1997 Jul; 4(7):497-505)。

[0300] 衍生自Msp的单体可以用可显示标记物(revealing label)进行标记。所述显示标记物可为任何使孔可被检测的合适的标记物。合适的标记物如下所述。

[0301] 衍生自Msp的单体也可以使用D-氨基酸生产。例如, 衍生自Msp的单体可包含L-氨基酸和D-氨基酸的混合物。这是本领域中用于生产这类蛋白质或肽常规的。

[0302] 衍生自Msp的单体含有一个或多个特异性修饰以促进对核苷酸的鉴别。衍生自Msp的单体也可以含有其它非特异性修饰, 只要它们不干扰孔的形成即可。许多非特异性侧链修饰是本领域已知的, 并可以针对衍生自Msp的单体的侧链进行。这种修饰包括, 例如, 通过与醛反应然后用NaBH₄还原进行氨基酸的还原性烷基化, 用甲基乙酰亚胺酯(methylacetimidate)进行脒基化, 或用乙酸酐进行酰化。

[0303] 衍生自Msp的单体可以使用本领域中已知的标准方法来制备。衍生自Msp的单体可以通过合成或重组方式制成。例如, 所述孔可通过体外翻译和转录(IVTT)合成。制造孔的合适的方法在国际申请No.PCT/GB09/001690(公开为WO 2010/004273), PCT/GB09/001679(公开为WO 2010/004265)或PCT/GB10/000133(公开为WO 2010/086603)中描述。论述了将孔插入膜的方法。

[0304] 跨膜蛋白质孔还优选衍生自 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。

[0305] 在一些实施例中, 跨膜蛋白孔是经化学修饰的。所述孔可以任何方式在任何位点

进行化学修饰。所述跨膜蛋白孔优选通过下述进行化学修饰：将分子连接到一个或多个半胱氨酸(半胱氨酸连接)，将分子连接到一个或多个赖氨酸，将分子连接到一个或多个非天然氨基酸，表位的酶修饰或末端的修饰。适用于进行这些修饰的方法是本领域已知的。跨膜蛋白孔可通过连接任何分子被化学修饰。例如，所述孔可以通过连接染料或荧光团被化学修饰。

[0306] 可对孔中任何数量的单体进行化学修饰。一个或多个，例如2,3,4,5,6,7,8,9或10个单体优选如上所述进行化学修饰。

[0307] 半胱氨酸残基的反应性可以通过相邻残基的修饰而增强。例如，侧翼精氨酸、组氨酸或赖氨酸残基的碱基将使半胱氨酸硫醇基团的pKa变为更大反应性的S⁻基团的pKa。半胱氨酸残基的反应性可通过硫醇保护基如dTNB进行保护。所述保护基可在连接体连接之前与孔的一个或多个半胱氨酸残基反应。

[0308] 所述分子(使用其对孔进行化学修饰)可以直接地连接到孔或通过连接体连接到孔，如国际申请No.PCT/GB09/001690(公开为WO 2010/004273), PCT/GB09/001679(公布为WO 2010/004265)或PCT/GB10/000133(公开为WO 2010/086603)所公开的。

[0309] 本文所描述的任何蛋白质，例如跨膜蛋白孔，可以被修饰以有助于它们的鉴定或纯化，所述修饰为例如通过添加组氨酸残基(his标签)，天冬氨酸残基(asp标签)，链霉亲和素标签，Flag标签，SUMO标签，GST标签或MBP标签，或通过添加信号序列以促进它们从细胞中分泌，该细胞中的多肽不天然地含有该信号序列。引入遗传标签的另一方式是通过化学反应将标签连到孔或构建体上的天然或工程位点。其例子是将凝胶迁移试剂与工程连接在所述孔外的半胱氨酸反应。这已被证明是用于分离溶血素异源寡聚物的方法(Chem Biol. 1997 Jul; 4(7):497-505)。

[0310] 所述孔可以用显示标记物标记。所述显示标记物可以是使得所述孔被检测的任何合适的标记物。合适的标记物包括，但不限于，荧光分子，放射性同位素，例如¹²⁵I, ³⁵S, 酶，抗体，抗原，多核苷酸和配体例如生物素。

[0311] 本文所描述的任何蛋白质，例如跨膜蛋白孔可以通过人工合成或重组手段制备。例如，所述孔可通过体外翻译和转录(IVTT)合成。该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被修饰为用以提高蛋白质的稳定性。当通过合成手段制备蛋白质时，可以在制备过程中引入这种氨基酸。所述孔也可在合成或重组生产后被改变。

[0312] 所述孔还可以使用D-氨基酸生产。例如，所述孔或构建体可包含L-氨基酸和D-氨基酸的混合物。这是本领域中用于生产这类蛋白质或肽常规的。

[0313] 所述孔还可以包含其他非特异性修饰，只要它们不干扰孔形成或构建体功能即可。许多非特异性侧链修饰是本领域中已知的，并且可以针对蛋白质的侧链进行。这样的修饰包括，例如，通过与醛反应然后用NaBH₄还原进行氨基酸的还原性烷基化，用甲基乙酰亚胺酯进行的肽基化，或用乙酸酐进行的酰化。

[0314] 本文所描述的任何蛋白质例如跨膜蛋白孔可以使用本领域已知的标准方法制备。可以使用本领域的标准方法衍生和复制编码孔或构建体的多核苷酸序列。可以使用本领域的标准技术在细菌宿主细胞中表达编码孔或构建体的多核苷酸序列。所述孔可在细胞中由重组表达载体原位表达多肽而产生。所述表达载体可选地携带诱导型启动子以控制所述多肽的表达。这些方法在Sambrook, J. 和 Russell, D. (2001). 分子克隆: 实验手册, 第3版. 冷泉

港实验室出版社,冷泉港,纽约 (Molecular Cloning: A Laboratory Manual, 3rd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) 中描述。

[0315] 所述孔可以从制备生物体的蛋白质通过任何蛋白质液相色谱系统进行纯化后,或在重组表达后,而大规模制备。通常的蛋白质液相色谱系统包括FPLC、AKTA系统、Bio-Cad系统、Bio-Rad BioLogic系统和Gilson HPLC的系统。

[0316] 间隔器

[0317] 如果所述一个或多个多核苷酸结合蛋白是解旋酶,且所述一个或多个装载部分包含多核苷酸,则所述一个或多个解旋酶可以停滞在一个或多个间隔器处,如在国际申请 No. PCT/GB2014/050175 (公开为 WO 2014/135838) 中描述。在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器的任何构造均可以在本发明中使用。

[0318] 当部分靶多核苷酸进入所述孔,并相对于所述孔沿着由所施加的电势而产生的场移动如穿过所述孔时,随着多核苷酸相对于所述孔移动,如穿过所述孔,所述一个或多个解旋酶被所述孔移动穿过间隔器。这是因为,多核苷酸(包括所述一个或多个间隔器)相对于所述孔移动,如穿过所述孔,并且所述一个或多个解旋酶保留在所述孔的顶部。

[0319] 所述一个或多个间隔器优选为装载部分多核苷酸的一部分,例如它/它们中断多核苷酸序列。所述一个或多个间隔器优选不是杂合到多核苷酸的一个或多个阻断分子如减速障 (speed bump) 的一部分。

[0320] 装载部分多核苷酸中可以有任何数量的间隔器,如1、2、3、4、5、6、7、8、9、10或更多个间隔器。多核苷酸中优选有2个、4个或6个间隔器。不同的装载部分多核苷酸中可以有间隔器,如在前导序列中的间隔器,和在桥连部分或发夹环中的间隔器。

[0321] 所述一个或多个间隔器其各自提供使一个或多个解旋酶即使在激活模式下也无法克服的能量屏障。所述一个或多个间隔器可通过减弱解旋酶的牵引(例如通过从多核苷酸中的核苷酸去除碱基)或物理地阻断所述一个或多个解旋酶的移动(例如使用庞大的化学基团)而使所述一个或多个解旋酶停滞 (stall)。

[0322] 所述一个或多个间隔器可包含使一个或多个解旋酶停滞的任何分子或分子组合。所述一个或多个间隔器可包含防止所述一个或多个解旋酶沿着多核苷酸移动的任何分子或分子组合。这可以在不存在跨膜孔和施加的电势的情况下,直接确定所述一个或多个解旋酶是否被停滞在一个或多个间隔器处。例如,其可如实施例中所示被检测,例如可以用 PAGE 测定解旋酶移动经过间隔器且置换DNA的互补链的能力。

[0323] 所述一个或多个间隔器通常包含线性分子,如聚合物。所述一个或多个间隔器通常具有与多核苷酸不同的结构。例如,如果该多核苷酸是DNA时,则所述一个或多个间隔器通常不是DNA。特别是,如果所述多核苷酸是脱氧核糖核酸 (DNA) 或核糖核酸 (RNA),则所述一个或多个间隔器优选包含肽核酸 (PNA)、甘油核酸 (GNA)、苏糖核酸 (TNA)、锁核酸 (LNA)、桥连核酸 (BNA) 或具有核苷酸侧链的合成聚合物。所述一个或多个间隔器可以沿与所述多核苷酸相反的方向包含一个或多个核苷酸。例如,当多核苷酸在5'到3'的方向时,所述一个或多个间隔器可在3'至5'方向包含一个或多个核苷酸。核苷酸可以是上文论述的任何核苷酸。

[0324] 所述一个或多个间隔器优选包含一个或多个硝基吲哚,例如一个或多个5-硝基吲哚、一个或多个肌苷、一个或多个吖啶、一个或多个2-氨基嘌呤、一个或多个2-6-二氨基嘌

呤、一个或多个5-溴-脱氧尿苷、一个或多个倒立胸苷(倒立dT)、一个或多个倒立二脱氧胸苷(ddT)、一个或多个二脱氧胞苷(ddC)、一个或多个5-甲基胞苷、一个或多个5-羟甲基胞苷、一个或多个2'-0-甲基RNA碱基、一个或多个异脱氧胞苷(Iso-dC)、一个或多个异脱氧鸟苷(Iso-dG)、一个或多个iSpC3基团(即缺乏糖和碱基的核苷酸)、一个或多个光可切割(PC)基团、一个或多个己二醇基团、一个或多个间隔器9(iSp9)基团、一个或多个间隔器18(iSp18)基团、聚合物或一个或多个硫醇的连接。所述一个或多个间隔器可包含这些基团的任意组合。这些基团中的许多都是商购自IDT®(Integrated DNA Technologies®)。

[0325] 所述一个或多个间隔器可包含任何数量的这些基团。例如,对于2-氨基嘌呤、2-6-二氨基嘌呤、5-溴-脱氧尿苷、倒立dT、ddT、ddC、5-甲基胞苷、5-羟甲基胞苷、2'-0-甲基RNA碱基、异-dC、异-dG、iSpC3基团、PC基团、己二醇基团和硫醇的连接,所述一个或多个间隔器优选包含2、3、4、5、6、7、8、9、10、11、12或更多个。所述一个或多个间隔器优选包含2、3、4、5、6、7、8个或更多个iSp9基团。所述一个或多个间隔器优选包含2、3、4、5或6个或更多个iSp18基团。最优选的间隔器为四个iSp18基团。

[0326] 所述聚合物优选为多肽或聚乙二醇(PEG)。所述多肽优选地包含2、3、4、5、6、7、8、9、10、11、12或更多个氨基酸。所述PEG优选包含2、3、4、5、6、7、8、9、10、11、12或更多个单体单元。

[0327] 所述一个或多个间隔器优选包含一个或多个脱碱基核苷酸(即缺乏核碱基的核苷酸),如2、3、4、5、6、7、8、9、10、11、12或更多个脱碱基核苷酸。所述核碱基可以被脱碱基核苷酸中的-H(idSP)或-OH替代。脱碱基间隔器可以通过从一个或多个相邻的核苷酸中去除核碱基而被嵌入到多核苷酸中。例如,多核苷酸可以被修饰为包括3-甲基腺嘌呤、7-甲基鸟嘌呤、1,N 6-亚乙烯基腺嘌呤肌苷或次黄嘌呤,并且所述核碱基可使用人类烷基腺嘌呤DNA转葡萄糖基酶(hAAG)从这些核苷酸中去除。或者,多核苷酸可以被修饰为包括尿嘧啶和用尿嘧啶-DNA转葡萄糖基酶(UDG)去除的核碱基。在一个实施方案中,所述一个或多个间隔器不包含任何脱碱基核苷酸。

[0328] 所述一个或多个解旋酶可通过每个线性分子间隔器停滞(即停滞在其前面)或在每个线性分子间隔器上停滞。如果使用线性分子间隔器,则优选所述装载部分多核苷酸优选提供有与一个或多个解旋酶待移动穿过的每个间隔器的末端相邻近的多核苷酸的双链区。所述双链区通常有助于将一个或多个解旋酶停滞到相邻的间隔器上。如果该方法是在约100mM或更低的盐浓度下进行的,则存在双链区是特别优选的。每个双链区通常为至少10个,如至少12个核苷酸的长度。如果本发明中使用的装载部分多核苷酸是单链的,则双链区可以通过将较短的多核苷酸杂合到与间隔器邻近的区域而形成。所述较短的多核苷酸通常由与装载部分多核苷酸相同的核苷酸形成,但也可以由不同的核苷酸形成。例如,所述较短的多核苷酸可由LNA或BNA形成。

[0329] 如果使用线性分子间隔器,所述装载部分多核苷酸优选提供有阻断分子,该阻断分子在每个间隔器的与所述一个或多个解旋酶待移动穿过的末端相对的末端。这有助于确保所述一个或多个解旋酶保持停滞在每个间隔器上。在它/它们在溶液中扩散的情况下,也有助于在多核苷酸上保持一个或多个解旋酶。阻断分子可以是下文所述的能物理地引起一个或多个解旋酶停滞的任何化学基团。阻断分子可以是多核苷酸的双链区。

[0330] 所述一个或多个间隔器优选包含物理地引起一个或多个解旋酶停滞的一个或多

个化学基团。所述一个或多个化学基团优选为一个或多个侧挂的化学基团。所述一个或多个化学基团可以被连接到所述多核苷酸中的一个或多个核碱基。所述一个或多个化学基团可以被连接到多核苷酸骨架。可以存在任意数量的这些化学基团,如2、3、4、5、6、7、8、9、10、11、12或更多个。合适的基团包括,但不限于,荧光团、链霉亲和素和/或生物素、胆固醇、亚甲蓝、二硝基苯酚(DNP)、地高辛配基和/或抗地高辛配基和二苯甲基环辛炔基团。

[0331] 装载部分多核苷酸中的不同间隔器可以包含不同的停滞分子。例如,一个间隔器可以包含上述的线性分子中的一个,而另一个间隔器可以包含物理地引起一个或多个解旋酶停滞的一个或多个化学基团。间隔器可以包含任何上述的线性分子和一个或多个物理地引起一个或多个解旋酶停滞的化学基团,如一个或多个脱碱基体和荧光团。

[0332] 合适的间隔器可以根据装载部分多核苷酸的类型和本发明的方法实施的条件进行设计。大多数的解旋酶结合DNA并沿DNA移动,且由此可以使用不为DNA的任何物质使其停滞。合适的分子如上所述。

[0333] 本发明的表征方法优选在游离核苷酸存在下和/或在解旋酶辅因子存在下进行。这在下文更详细地讨论。在跨膜孔和施加电势不存在时,所述一个或多个间隔器优选能够在游离核苷酸存在下和/或在解旋酶辅因子存在下停滞所述一个或多个解旋酶。

[0334] 如果本发明的表征方法如下文所述在游离核苷酸和解旋酶辅因子存在下进行(使得一个或更多的解旋酶是在激活模式下),则通常使用一个或多个较长间隔器来确保所述一个或多个解旋酶是在它们与跨膜孔接触前且电势施加前被停滞在多核苷酸上。在不存在游离核苷酸和解旋酶辅因子情况下(使得所述一个或多个解旋酶是在非激活模式下),可以使用一个或多个较短间隔器。

[0335] 所述盐浓度也影响所述一个或多个间隔器停滞一个或多个解旋酶的能力。在跨膜孔和施加的电势不存在时,所述一个或多个间隔器优选能够在约100mM或更低的盐浓度下停滞所述一个或多个解旋酶。本发明的方法中使用的盐浓度越高,通常使用的所述一个或多个间隔器越短,反之亦然。

[0336] 特征的优选组合在下表4中示出。

多核苷酸	间隔器组成*	间隔器长度(即其的数量*)	盐[1]	游离核苷酸?	解旋酶辅因子?
[0337]	DNA	iSpC3	4	1M	有
	DNA	iSp18	4	100-1000 mM	有
	DNA	iSp18	6	<100-1000 mM	有
	DNA	iSp18	2	1M	有
	DNA	iSpC3	12	<100-1000 mM	有
	DNA	iSpC3	20	<100-1000 mM	有
	DNA	iSp9	6	100-1000 mM	有
	DNA	idSp	4	1 M	有

[0338] 所述方法可以涉及,移动两个或更多个解旋酶穿过间隔器。在这种情况下,间隔器的长度通常是增加的以防止在不存在孔和施加的电势下尾部解旋酶推动前导解旋酶穿过

间隔器。如果该方法涉及移动两个或更多个解螺酶穿过一个或多个间隔器，则上文所述的间隔器的长度可以增加至少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更多个解螺酶穿过一个或多个间隔器，则在上表4的第三列中的间隔器的长度可以增加1.5倍、2倍、2.5倍或3倍。

[0339] 所述两个或更多个解旋酶也可以被分开，使得每个解旋酶具有其自身的一个或多个间隔器。这在下面更详细地讨论。

[0340] 双链多核苷酸

[0341] 靶多核苷酸可以是双链的。如果多核苷酸是双链的，本发明优选包括在多核苷酸的一端连接桥连部分衔体，如发夹环衔体，以及分离所述多核苷酸的两条链，以形成单链多核苷酸构建体。根据本发明，单链多核苷酸构建体可以随后相对于所述孔移动，如穿过所述孔。以这种方式连接或内接合 (interrogating) 双链构建体的两条链提高了表征的效率和准确性。

[0342] 所述桥连部分能够连接多核苷酸的两条链。桥连部分通常共价连接多核苷酸的两条链。只要桥连部分不干扰所述多核苷酸穿过跨膜孔的移动，桥连部分可以是能够连接多核苷酸的两条链的任何物质。

[0343] 桥连部分可通过本领域中已知的任何合适的手段被连接到多核苷酸。桥连部分可以被分别合成并化学连接或通过酶接合到多核苷酸。或者，桥连部分可以在多核苷酸的加工过程中生成。

[0344] 桥连部分在多核苷酸的一端或接近多核苷酸的一端被连接到多核苷酸。桥连部分优选连接到所述多核苷酸末端的10个核苷酸中的多核苷酸。

[0345] 尽管桥连部分是优选的装载部分，但是所述桥连部分不是必须具有一个或多个结合的(或连接的)多核苷酸结合蛋白，只要所述装载部分如Y衔体具有连接到靶多核苷酸的另一端的一个或多个结合的(或连接的)多核苷酸结合蛋白即可。在一些实施方案中，所述装载部分可以被连接到本发明的多核苷酸的两端，优选其中一个是Y衔体，而另一个是桥连部分，如发夹环衔体。合适的桥连部分在上文论述，参考装载部分。如果发夹环不被用作装载部分，发夹环通常是110个或更少个核苷酸，例如100个或更少个核苷酸、90个或更少个核苷酸、80个或更少个核苷酸、70个或更少个核苷酸、60个或更少个核苷酸、50个或更少个核苷酸、40个或更少个核苷酸、30个或更少个核苷酸、20个或更少个核苷酸，或10个或更少个核苷酸的长度。发夹环优选为约1至110、2到100、5至80、或6到50个核苷酸的长度。如果发夹环被包含在衔体的差分可选的结合 (differential selectable binding) 中，优选所述发夹环具有较长的长度，如50到110个核苷酸。类似地，如果发夹环如下文所述不包含在差分可选的结合中，优选所述发夹环具有较短的长度，如1至5个核苷酸。

[0346] 所述桥连部分衔体或发夹环衔体可以被连接到或接合到靶多核苷酸，如上文所述。

[0347] 多核苷酸的两条链可以使用本领域中已知的任何方法分离。例如，它们可以由一个或多个多核苷酸结合蛋白或使用有利于去杂合化的条件进行分离(有利于去杂合化的条件的实例包括，但不限于高温、高pH值以及添加可以破坏氢键或碱基配对的试剂，如甲酰胺和脲)。如果使用一个或多个多核苷酸结合蛋白来分离链，则所述一个或多个多核苷酸结合蛋白通常在桥连部分的另一端被连接到靶多核苷酸，例如使用Y衔体进行连接。

[0348] 所述一个或多个装载部分,优选Y衔接体和/或桥连部分衔接体(如发夹环衔接体),优选包含可选结合部分。这使得多核苷酸被纯化或分离。可选结合部分是在其结合特性的基础上选择出的部分。因此,可选结合部分优选为特异性结合到表面的部分。如果可选结合部分,与本发明中使用的任何其他部分相比,以大得多的程度结合到表面,则可选结合部分特异性结合到该表面。在优选的实施方案中,所述部分结合到未结合有本发明使用的其他部分的表面。

[0349] 合适的可选结合部分是本领域已知的。优选的可选结合部分包括,但不限于,生物素、多核苷酸序列、抗体、抗体片段,如Fab和ScFv,抗原、多核苷酸结合蛋白、聚组氨酸尾部和GST标签。最优选的可选结合部分是生物素和可选的多核苷酸序列。生物素特异性结合到涂覆有抗生物素蛋白的表面。可选的多核苷酸序列特异性结合(即杂合)到涂覆有同源序列的表面。或者,可选的多核苷酸序列特异性结合到涂覆有多聚核苷酸结合蛋白的表面。

[0350] 所述一个或多个装载部分,优选Y衔接体和/或桥连部分衔接体(如发夹环衔接体)和/或可选的结合部分,可包含可以被切割的、带缺口的、裂缝的或水解的区域。这种区域可以被设计为允许第一和/或第二多核苷酸从其所结合的表面移除,随后纯化或分离。合适的区域在本领域中是公知的。合适的区域包括,但不限于, RNA区域、包含脱硫生物素和链霉亲和素的区域、二硫键以及可光解的区域。

[0351] 前导序列

[0352] 所述一个或多个装载部分可提供有优先螺入孔中的前导序列。前导序列有助于本发明的方法。前导序列被设计为优先螺入跨膜孔中,并由此有助于靶多核苷酸相对于所述孔的移动,如穿过所述孔。前导序列还可以用于将多核苷酸连接到一个或多个锚,如上文所述。

[0353] 前导序列通常包含聚合物。所述聚合物优选带负电荷。聚合物优选为多核苷酸,如DNA或RNA,修饰的核苷酸(如脱碱基的DNA)、PNA、LNA、BNA、聚乙二醇(PEG)或多肽。所述前导序列优选包含多核苷酸,并且更优选包含单链多核苷酸。前导序列可以包含上述的任何多核苷酸。单链前导序列最优选包含DNA的单链,如聚dT段。前导序列优选地包含所述一个或多个间隔器。

[0354] 前导序列可以是任意长度,但通常为10至150个核苷酸长度,例如20至150个核苷酸长度。前导序列的长度通常取决于在本方法中使用的跨膜孔。

[0355] 双耦合

[0356] 在优选的实施方案中,本发明包括将一个或多个多核苷酸结合蛋白连接到靶双链多核苷酸的方法,包括:

[0357] (a) 提供结合(或连接)有一个或多个多核苷酸结合蛋白并且具有用于将多核苷酸耦合到膜的一个或多个第一锚的Y衔接体,且提供具有一个或多个第二锚的桥连部分衔接体,如发夹环衔接体,其中所述桥连部分衔接体耦合到膜的强度比Y衔接体耦合到膜的强度更大;以及

[0358] (b) 将所述Y衔接体连接到所述靶多核苷酸的一端,并且将所述桥连部分连接到所述靶多核苷酸的另一端。桥连部分优选具有与其连接的一个或多个多核苷酸结合蛋白,优选一个或多个分子制动器。

[0359] 本发明还提供表征靶双链多核苷酸的方法,包括

[0360] (c) 将在上面步骤(b)中提供的多核苷酸与跨膜孔接触,使得所述一个或多个多核

昔酸结合蛋白控制靶多核昔酸相对于所述孔的移动,如穿过所述孔;和

[0361] (d) 随着所述多核昔酸相对于所述孔移动,获取一个或多个测量值,其中测量值代表多核昔酸的一个或多个特征,并由此表征所述靶多核昔酸。

[0362] 这种类型的方法在英国申请No. 1406147.7中被详细讨论。

[0363] Y衔体和/或桥连部分衔体通常为多核昔酸衔体。它们可以由上文所述的任何多核昔酸形成。

[0364] Y衔体通常包含(a) 双链区域和(b) 单链区域或另一端不互补的区域。如果Y衔体包括一个单链区域,则Y衔体可被描述为具有悬突(overhang)。与双链部分不同的是,由于两条链通常不彼此杂合,因此Y衔体中非互补区域的存在给予该衔体Y形。Y衔体包括一个或多个第一锚。锚在上文更详细地论述。

[0365] Y衔体优选包括优先螺入孔中的前导序列。这是上文进行了论述。

[0366] 所述桥连部分衔体优选包含如上所述的可选结合部分。所述桥连部分衔体和/或可选结合部分可包含如上所述的可切割的、带缺口的、裂缝的或水解的区域。

[0367] Y衔体和/或桥连部分衔体可以使用本领域中已知的任何方法接合到多核昔酸。所述的一个或两个衔体可使用连接酶,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连接酶进行接合。

[0368] 桥连部分衔体耦合(或结合)到膜的强度比Y衔体耦合(或结合)到膜的强度更大。这可以用任何方式来测定。用于测定耦合(或结合)的强度的合适的方法在英国申请No. 1406147.7的实施例中公开。

[0369] 桥连部分衔体耦合(或结合)的强度优选至少是Y衔体耦合(或结合)的强度的1.5倍,如是Y衔体耦合(或结合)的强度的至少二倍,至少三倍,至少四倍,至少五倍或至少十倍。用于膜的桥连部分衔体的亲和常数(Kd)优选为Y衔体的亲和常数的至少1.5倍,如至少二倍,至少三倍,至少四倍,至少五倍或至少十倍。

[0370] 使桥连部分衔体比Y衔体更牢固地耦合(或结合)到膜的方式有多种。例如,所述桥连部分衔体可以包含比Y衔体更多的锚。例如,桥连部分衔体可包含2、3或更多个第二锚,而Y衔体可包含一个第一锚。

[0371] 一个或多个第二锚耦合(或结合)到膜的强度可以比一个或多个第一锚耦合(或结合)到膜的强度更大。将一个或多个第二锚耦合(或结合)到桥连部分衔体的强度可以比将一个或多个第一锚耦合(或结合)到Y衔体的强度更大。所述一个或多个第一锚和所述一个或多个第二锚定可通过杂合被连接到其各自的衔体,且在一个或多个第二锚中杂合的强度比在一个或多个第一锚中更大。这些实施方案的任意组合也可以用于本发明中。耦合(或结合)的强度可以使用本领域的已知技术来测定。

[0372] 所述一个或多个第二锚优选包含耦合(或结合)到膜的一个或多个基团,所述一个或多个基团在一个或多个第一锚中耦合(或结合)到膜的一个或多个基团具有更高的强度。在优选的实施方案中,使用胆固醇将桥连部分衔体/一个或多个第二锚耦合(或结合)到膜,且使用棕榈酸酯将Y衔体/一个或多个第一锚耦合(或结合)到膜。胆固醇比棕榈酸酯更牢固地结合到三嵌段共聚物膜和脂质膜。在可替代的实施方案中,使用单酰基物种,如棕榈酸膜,将桥连部分衔体/一个或多个第二锚耦合(或结合)到膜,并且使用二酰基物种如二棕榈酰磷脂酰胆碱将Y衔体/一个或多个第一锚耦合(或结合)到膜。

[0373] 添加发夹环及前导序列

[0374] 根据本发明，双链多核苷酸可以与MuA转座酶和一组双链MuA底物接触，其中所述组中的底物的一部分是结合到一个或多个多核苷酸结合蛋白且包含前导序列的Y衔接体，且其中所述组中的底物的一部分是桥连部分衔接体，例如结合到一个或多个多核苷酸结合蛋白的发夹环衔接体。所述Y衔接体和/或桥连部分衔接体有所述装载部分的功能。转座酶分裂双链多核苷酸并且将MuA底物接合到分裂碎片的一端或两端。这产生多个经修饰的双链多核苷酸，所述经修饰的双链多核苷酸一端含有具有一个或多个多核苷酸结合蛋白和前导序列的Y衔接体，且在另一端含有具有一个或多个多核苷酸结合蛋白的桥连部分(或发夹环)。所述经修饰的双链多核苷酸随后可以使用本发明的方法进行表征。

[0375] 所述组中的每个底物优选包含至少一个通用核苷酸的悬突，使得转座酶将模板多核苷酸分裂且将底物接合到双链分裂碎片的一端或两端，并由此产生多个碎片/底物构建体，并且其中所述方法进一步包括将悬突接合到构建体中的碎片，并由此产生多个经修饰的双链多核苷酸。合适的通用核苷酸如上所述。悬突优选为5个核苷酸的长度。

[0376] 或者，所述组中的每个底物优选包含(i)至少一个悬突，以及(ii)在与至少一个悬突相同的链中的至少一个核苷酸，所述悬突包含在模板多核苷酸中不存在的核苷，使得转座酶分裂所述模板多核苷酸，且将底物接合到双链碎片的一端或两端，并由此产生多个碎片/底物构建体，且其中所述方法进一步包括(a)通过选择性地去除所述至少一个核苷酸而从所述构建体去除所述悬突，且由此产生多个含有单链间隙的双链构建体，以及(b)修复所述构建体中的单链间隙，并由此产生多个经修饰的双链多核苷酸。多核苷酸分析物通常包含核苷脱氧腺苷(dA)、脱氧尿苷(dU)和/或脱氧胸苷(dT)、脱氧鸟苷(dG)和脱氧胞苷(dC)。不存在于多核苷酸中的核苷优选是脱碱基的腺苷(A)、尿苷(U)、5-甲基尿苷(m^5U)、胞苷(C)或鸟苷(G)或包含脲、5,6-二羟基胸腺嘧啶、胸腺嘧啶二醇、5-羟基-5-甲基乙内酰脲(5-hydroxy-5 methylhydantoin)、尿嘧啶二醇、6-羟基-5,6-二氢硫胺、甲基丙醇二酰脲(tartronylurea)、7,8-二氢-8-氧鸟嘌呤(8-氧鸟嘌呤)、8-氧代腺嘌呤、fapy-鸟嘌呤、甲基-fapy-鸟嘌呤、fapy-腺嘌呤、黄曲霉毒素B1-fapy-鸟嘌呤、5-羟基胞嘧啶、5-羟基尿嘧啶、3-甲基腺嘌呤、7-甲基鸟嘌呤、1,N 6-亚乙烯基腺嘌呤、次黄嘌呤、5-羟基尿嘧啶、5-羟基甲基尿嘧啶、5-甲酰基尿嘧啶或顺式-环丁烷嘧啶二聚体。所述至少一个核苷酸优选为来自悬突的10个或更少的核苷酸。所述至少一个核苷酸优选为悬突中的第一个核苷酸。悬突中的所有核苷酸优选包含在模板多核苷酸中不存在的核苷。

[0377] 这些基于MuA的方法在英国申请No. 1314695.6中公开。它们还在英国申请No. 1406147.7中有详细论述。

[0378] 一个或多个解旋酶可在与双链多核苷酸和MuA转座酶接触之前，连接到MuA底物Y衔接体(即装载部分)。或者，一个或多个解旋酶可在与双链多核苷酸和MuA转座酶接触之后，连接到MuA底物Y衔接体(即装载部分)。

[0379] 一个或多个分子制动器可在与双链多核苷酸和MuA转座接触之前，连接到MuA底物桥连部分(或发夹环)衔接体。或者，一个或多个分子制动器可在与双链多核苷酸和MuA转座酶接触后，连接到MuA底物桥连部分(或发夹环)衔接体。

[0380] 多核苷酸表征

[0381] 本发明提供了表征靶多核苷酸的方法。靶多核苷酸也可称为模板多核苷酸或感兴

趣的多核苷酸。

[0382] 本发明的方法涉及测量所述多核苷酸的一个或多个特征。特别地,上述用于控制多核苷酸移动穿过跨膜孔的方法中之一按步骤(a)实施,然后在步骤(b)中,当所述多核苷酸相对于所述孔移动时,获取一个或多个测量值,其中测量值代表多核苷酸的一个或多个特征。合适的测量值如上文所述。

[0383] 可以测查任意数量的靶多核苷酸。例如,本发明的方法可以涉及表征2、3、4、5、6、7、8、9、10、20、30、50、100个或更多个靶多核苷酸。所述靶多核苷酸可以是天然存在的或人工的。例如,该方法可以被用于检验所制造的寡核苷酸的序列。该方法一般在体外进行。

[0384] 该方法可以涉及测量一个,两个,三个,四个或五个或更多个靶多核苷酸的特征。所述一个或多个特征优选选自(i)所述多核苷酸的长度,(ii)所述多核苷酸的同一性,(iii)所述多核苷酸的序列,(iv)所述多核苷酸的二级结构,和(v)所述多核苷酸是否被修饰。(i)至(v)的任意组合可以根据本发明进行测量,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}。

[0385] 对于(i),多核苷酸的长度例如可以通过确定所述多核苷酸和所述孔之间的相互作用次数,或所述多核苷酸与所述孔之间相互作用的持续时间来测定。

[0386] 对于(ii),所述多核苷酸的同一性可以通过多种方式测定。所述多核苷酸的同一性可以联合所述多核苷酸的序列的测定来测定,或不联合多核苷酸的序列的测定进行测定。前者是直接进行的;对所述多核苷酸进行测序,并由此进行鉴定。后者可以以几种方式来完成。例如,可以测定多核苷酸中特定模序的存在(而无需测定该多核苷酸的其余序列)。或者,所述方法中测定的特定的电和/或光信号可鉴定来自特定来源的多核苷酸。

[0387] 对于(iii),多核苷酸的序列可以如以前所述确定。合适的测序方法,特别是那些使用电测量的方法,描述于Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7,Lieberman KR et al., J Am Chem Soc. 2010;132(50):17961-72,和国际申请WO 2000/28312中。

[0388] 对于(iv),所述二级结构可以以多种方式测量。例如,如果该方法包含电测量,二级结构可以利用相对于孔流动例如穿过孔时的停留时间的改变或电流变化进行测量。这使得单链和双链多核苷酸的区域能够得到识别。

[0389] 对于(v),可以测定任何修饰的存在或不存在。该方法优选包括确定所述多核苷酸是否通过甲基化、氧化、损伤、用一个或多个蛋白质或一个或多个标记物、标签或间隔器进行了修饰。特异性修饰将导致与孔的特异性相互作用,这可以使用下面描述的方法来测定。例如,甲基胞嘧啶可以基于其与每个核苷酸的相互作用过程中相对于孔流动例如穿过孔时的电流,与胞嘧啶区别开。

[0390] 所述方法可以使用任何适于研究膜/孔系统(其中孔存在于膜中)的设备进行实施,该方法可以使用任何适合于感测跨膜孔的设备实施。例如,所述设备包括一个室,所述室包括水溶液和将该室分割为两部分的屏障(barrier)。所述屏障通常具有开口,其中在开口中形成包括所述孔的膜。或者该屏障形成其中存在所述孔的膜。

[0391] 该方法可以使用在国际申请No.PCT/GB08/000562 (WO 2008/102120) 中描述的设备实施。

[0392] 该方法可以包括随着多核苷酸相对于所述孔移动, 测量经过所述孔的电流。因此该装置也可以包括能够跨所述膜和孔而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实施。所述方法优选包含使用电压钳。

[0393] 本发明的方法包括测量当所述多核苷酸相对于所述孔移动时经过所述孔的电流。用于测量通过跨膜蛋白孔的离子电流的合适的条件是本领域已知的, 并且在实施例中公开。该方法通常用跨所述膜和孔施加的电压进行实施。使用的电压通常为+5V到-5V, 例如+4V到-4V, +3V到-3V或+2V到-2V。使用的电压通常为-600mV到+600mV或-400mV到+400mV。所使用的电压优选在具有以下下限和上限的范围内: 所述下限选自-400mV, -300mV, -200mV, -150mV, -100mV, -50mV, -20mV和0mV的范围内, 所述上限独立地选自+10mV, +20mV, +50mV, +100mV, +150mV, +200mV, +300mV, 和+400mV。所用的电压更优选在100mV到240mV的范围内并最优选在120mV到220mV的范围内。可通过施加提高的电势通过孔来提高对不同核苷酸的鉴别力。

[0394] 该方法通常在任何载荷子, 如金属盐, 如碱金属盐, 卤盐, 如盐酸盐, 如碱金属盐酸盐的存在下实施。载荷子可包括离子型液体或有机盐, 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。在上面论述的示例性设备中, 所述盐存在于所述室中的水溶液中。通常使用氯化钾(KCl), 氯化钠(氯化钠)、氯化铯(CsCl)或亚铁氰化钾和铁氰化钾的混合物。优选KCl、NaCl, 和亚铁氰化钾与铁氰化钾的混合物。所述载荷子可以不对称的跨所述膜。例如在膜的每侧, 所述载荷子的种类和/或浓度可以是不同的

[0395] 盐浓度可为饱和的。盐浓度可以是3M或更低, 通常为0.1M至2.5M, 0.3M至1.9M, 0.5M至1.8M, 0.7M至1.7M, 0.9M至1.6M或1M至1.4M。优选盐浓度为150mM到1M。所述方法优选使用至少为0.3M, 例如至少为0.4M, 至少为0.5M, 至少为0.6M, 至少为0.8M, 至少为1.0M, 至少为1.5M, 至少为2.0M, 至少为2.5M, 或者至少为3.0M的盐浓度进行实施。高盐浓度提供了高信噪比, 并使得在正常电流波动的背景下, 代表核苷酸存在的电流能被识别。

[0396] 该方法通常在缓冲剂存在下实施。在上面讨论的示例性设备中, 缓冲剂存在于所述室的水溶液中。本发明的方法可使用任何缓冲剂。通常地, 缓冲剂是磷酸缓冲液。另一个合适的缓冲剂是HEPES和Tris-HCl(三羟甲基氨基甲烷-HCl)缓冲剂。该方法通常在pH值为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, 7.0至8.8, 或7.5至8.5下实施。所使用的pH优选约为7.5。

[0397] 该方法可以通常在0°C到100°C, 15°C到95°C, 16°C至90°C, 17°C至85°C, 18°C至80°C, 19°C至70°C, 或者从20°C至60°C实施。该方法通常在室温下进行。该方法可选地在支持酶功能的温度下如约37°C实施。

[0398] 桥连部分测序

[0399] 在优选的实施方案中, 靶双链多核苷酸在一端设置有桥连部分(或发夹环)衔接体, 并且该方法包括将多核苷酸与跨膜孔接触, 使得该多核苷酸的两条链相对于孔移动, 如穿过所述孔, 并且随着多核苷酸的两条链相对于所述孔移动, 获取一个或多个测量值, 其中所述测量值代表多核苷酸的链的一个或多个特征, 并由此表征所述靶双链多核苷酸。上述的

任何实施方案同样适用于本实施例。

[0400] 经修饰的多核苷酸

[0401] 在使用本发明的方法之前,在聚合酶使用多核苷酸作为模板形成经修饰的多核苷酸的条件下,通过将所述多核苷酸与聚合酶和一组游离核苷酸接触,来修饰所述多核苷酸,其中当形成经修饰的多核苷酸时,所述聚合酶用不同的核苷酸物种来代替所述多核苷酸中的一个或多个核苷酸物种。所述经修饰的多核苷酸随后可以在本发明的方法中使用。这种类型的修饰在英国申请No.1403096.9中描述。可以使用任何上述的聚合酶。所述聚合酶优选Klenow or 9°North。

[0402] 在聚合酶使用该多核苷酸作为模板形成经修饰的多核苷酸的条件下,使所述多核苷酸与聚合酶接触。这种条件在本领域中是已知的。例如,通常在商购的聚合酶缓冲液中如来自New England Biolabs®的缓冲液中,使所述多核苷酸与聚合酶接触。对于Klenow,温度优选20到37°C,或对于9° North,温度优选60到75°C。通常使用引物或3'发夹作为聚合酶扩展的成核点。根据本发明聚合酶可以与靶多核苷酸接触,即通过下述进行接触: (a) 提供结合到一个或多个装载部分的一个或多个聚合酶,及 (b) 将所述一个或多个装载部分连接到所述靶多核苷酸。

[0403] 使用跨膜孔对多核苷酸的表征,如测序,通常包括,分析由k个核苷酸组成的聚合物单元(即‘k聚体’),其中k是正整数。这在国际申请No.PCT/GB2012/052343(公开为W02013/041878)中有论述。虽然理想的是在不同k-聚体的电流测量值之间能清楚的分离开,但是通常这些测量中的一些存在重叠。特别是k-聚体中具有大量的聚合物单元时,即具有非常高的k值时,难以分辨由不同的k-聚体产生的测量值,不利于导出关于多核苷酸的信息,例如估计多核苷酸的基础序列。

[0404] 通过用所述经修饰的多核苷酸中的不同的核苷酸物种代替所述多核苷酸中的一个或多个核苷酸物种,所述经修饰的多核苷酸含有的k-聚体不同于所述多核苷酸中的k-聚体。所述经修饰的多核苷酸中的不同k-聚体能够产生与来自所述多核苷酸中的k聚体不同的电流测量值,所以所述经修饰的多核苷酸提供了与所述多核苷酸不同的信息。来自所述经修饰的多核苷酸的额外信息使得可以更容易表征所述多核苷酸。在一些情况下,所述经修饰的多核苷酸其自身可以更容易表征。例如,所述经修饰的多核苷酸可以被设计为,包括在它们的电流测量值之间具有增强的分离或清楚的分离的k-聚体或具有降低的噪声的k-聚体。

[0405] 所述聚合酶优选在形成经修饰的多核苷酸时的不同核苷酸物种代替所述多核苷酸中的两个或多个核苷酸物种。所述聚合酶可以用有区别的核苷酸物种代替所述多核苷酸中的两个或多个核苷酸物种中的每一个。所述聚合酶可以用相同的核苷酸物种代替所述多核苷酸中的两个或多个核苷酸物种中的每一个。

[0406] 如果所述多核苷酸为DNA,则所述经修饰的多核苷酸中的不同的核苷酸物种通常包含与腺嘌呤,鸟嘌呤,胸腺嘧啶,胞嘧啶或甲基胞嘧啶不同的核碱基和/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。如果所述多核苷酸是RNA,则所述经修饰的多核苷酸中不同的核苷酸物种通常包含与腺嘌呤,鸟嘌呤,尿嘧啶,胞嘧啶或甲基胞嘧啶不同的核碱基,和/或包含与腺苷,鸟苷,尿苷,胞苷或甲基胞苷不同的核苷。所述不同的核苷酸物种可以是上述的任何通用核苷酸。

[0407] 所述聚合酶可以用含有在一个或多个核苷酸物种中不存在的化学基团或原子的不同核苷酸物种代替一个或多个核苷酸物种。所述化学基团可以是丙炔基,硫代基,氧化基,甲基,羟甲基,甲酰基,羧基,羰基,苄基,炔丙基或炔丙胺基。

[0408] 所述聚合酶可以用缺少存在于一个或多个核苷酸物种中的化学基团或原子的不同核苷酸种类来代替一个或多个核苷酸物种。聚合酶可以用具有改变的电负性的不同核苷酸物种来代替一个或多个核苷酸物种。具有改变的电负性的不同核苷酸物种优选包含卤素原子。

[0409] 该方法优选进一步包括选择性地从所述经修饰的多核苷酸中的一个或多个不同核苷酸物种中去除核碱基。

[0410] 产物

[0411] 本发明还提供使用本发明修饰的靶多核苷酸。本发明还提供具有一个或多个多核苷酸结合蛋白结合其上的装载部分。上文针对本发明方法所论述的任何实施方案均适用于本发明的多核苷酸和部分。

[0412] 试剂盒

[0413] 本发明还提供了一种用于将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的试剂盒,包括:(a)结合到一个或多个装载部分的一个或多个多核苷酸结合蛋白和(b)连接酶。上文所述的任何实施方案均适用于试剂盒。

[0414] 所述试剂盒优选用于通过跨膜孔的双链多核苷酸,且该试剂盒优选地包括具有一个或多个解旋酶附连的Y衔接体和具有一个或多个分子制动器附连的桥连部分(或发夹环)衔接体。所述Y衔接体优选包含一个或多个用于耦合所述多核苷酸到膜的第一锚,所述桥连部分(或发夹环)衔接体优选包含一个或多个用于耦合所述多核苷酸到膜的第二锚,并且将桥连部分(或发夹环)衔接体耦合到所述膜的强度优选比将所述Y衔接体耦合到所述膜的强度更大。

[0415] 所述试剂盒优选还包含跨膜孔。任何以上所述的膜和孔均可以在试剂盒中。

[0416] 上文针对本发明方法所述的任何实施方案同样适用于所述试剂盒。所述试剂盒还可以包括膜的组分,如两亲层或三嵌段共聚物膜的组分。

[0417] 本发明的试剂盒可以额外包括一个或多个能使上述任何实施方案实施的其它试剂或仪器。这类试剂或仪器包括以下中的一个或多个:合适的缓冲液(水性溶液),从受试者获得样品的工具(例如包含针的容器或仪器),扩增和/或表达多核苷酸的工具,或上文定义的膜,或电压或膜片钳装置。试剂可以干燥状态存在于试剂盒中,使得流体样品能重悬所述试剂。所述试剂盒还可以任选包括能使试剂盒用于本发明方法的介绍,或关于本发明方法可用于哪种生物的详情。

[0418] 下列实施例阐明了本发明。

[0419] 实施例1

[0420] 本实施例描述了用于将多核苷酸结合蛋白结合到装载部分的样品制备过程。

[0421] 材料与方法

[0422] TrwC Cba-L376C/Q594A/K762C (6.5 μ M,具有突变L376C/Q594A/K762C的SEQ ID NO:25) 和DNA发夹衔接体(4 μ M,SEQ ID NO:29在3'末端连接到四个iSpC3间隔器,所述四个iSpC3间隔器在相对端连接到SEQ ID NO:30的5'末端) 在缓冲液(100mM KC1,100mM CAPS (pH为10) 和1mM EDTA) 中混合,并培养30分钟。将TRIS(0.1体积的1M TRIS pH7.0) 添加到该

DNA/酶混合物中,并将该混合物充分混合。最后,添加0.025体积的DMF中的1.29mM双马来酰亚胺基乙烷,且将混合物进一步培养15分钟。各成分的最终浓度如下:TrwC Cba-L376C/Q594A/K762C (5.8 μ M),DNA发夹衔接体 (3.56 μ M),缓冲液 (89mM KC1,89mM CAPS,pH 10,0.89mM EDTA,89mM Tris (pH 7)) 和双马来酰亚胺基乙烷 (28.7 μ M),该混合物被称为样品1。

[0423] 然后使用8mL POROS HQ-10柱 (FPLC) 和随后的洗脱缓冲液 (缓冲液A——50mM乙醇胺,300mM NaCl,0.1% β -OTG,1mM TCEP,pH 10.0,缓冲液B——50mM乙醇胺,700mM NaCl,0.1% β -OTG,1mM TCEP,pH 10.0) 使用所述的步骤将预结合到DNA发夹衔接体的TrwC CBA-L376C/Q594A/K762C纯化。将样品1装载到柱上,用5柱体积的缓冲液A将没有TrwC CBA-L376C/Q594A/K762C结合的任何DNA或没有结合到DNA的酶洗掉。用11.8柱体积的0-100%缓冲液B对预结合到DNA发夹衔接体的TrwC CBA-L376C/Q594A/K762C进行洗脱,该样品被称为样品2。示例性FPLC轨迹在图1中示出,标有P1的峰对应预结合到DNA发夹衔接体的TrwC CBA-L376C/Q594A/K762C。

[0424] 然后将样品2使用5mL的Histrap HP柱 (FPLC),使用下列洗脱缓冲液 (缓冲液C——20mM的Na-CAPS,100mM的NaCl,0.1% β -OTG,1mM TCEP,pH 10.0,缓冲液D——20mM的Na-CAPS,2M NaCl,10% (w/v) 甘油,0.1% β -OTG,1mM的TCEP,pH 10.0,和缓冲剂E——20mM的Na-CAPS,100mM的NaCl,300mM咪唑,0.1% β -OTG,1mM的TCEP,pH 10.0) 进一步纯化。将样品2装载到所述柱上,用缓冲液C将没有TrwC CBA-L376C/Q594A/K762C结合的任何DNA或没有结合到DNA的酶从柱上洗脱掉。然后用缓冲液D洗涤该柱,然后用缓冲液C再次洗涤。最后用10柱体积的0-100%缓冲液E将预结合到DNA发夹衔接体的TrwC CBA-L376C/Q594A/K762C进行洗脱。然后汇集(pool) 主洗脱峰(实例在图2中示出,用P2标记主峰),测量其浓度,并使用TBE(天然的) PAGE和SDS PAGE凝胶分析DNA样品。TBE(天然的) PAGE为4-20%TBE凝胶,在150V下运行25分钟,然后使用SYBR金染料染色。这种染料使得样品中的任何DNA(有或没有酶结合)均是可见的。图3显示了这种凝胶。4-6列表明纯化后,DNA仍结合到酶。SDS PAGE为10%Bis-Tris凝胶,XT MOPS,在200V下运行60分钟,然后使用SYPRO红宝石染料染色。这种染料使得样品中的任何酶(结合到或未结合到DNA)均是可见的。图4的4-6列中显示了纯化后,使用双马来酰亚胺基乙烷连接体将酶闭合到DNA上。

[0425] 实施例2

[0426] 本实施例描述了用于将多核苷酸结合蛋白结合到装载部分的样品制备过程。

[0427] 材料与方法

[0428] 将T4 Dda-E94C/C109A/C136A/A360C (3 μ M,具有突变E94C/C109A/C136A/A360C的SEQ ID NO:24) 和DNA Y衔接体 (500nM,SEQ ID NO:26杂合到DNA链X=30个iSpC3间隔器,所述间隔器在其3'末端连接到SEQ ID NO:27的5'末端,SEQ ID NO:27在其3'末端连接到4个iSp18间隔器,所述4个iSp18间隔器在相对端连接到SEQ ID NO:28的5'末端) 在缓冲液 (50mM HEPES,100mM KOAc (pH 8)) 中混合,并添加EDTA (1mM) 和TMAD (100 μ M),并在室温下培养60分钟。然后将混合物在KC1/ATP溶液中 (500mM KC1,1mM ATP最终浓度) 以1:1稀释,并在环境温度下培育25分钟,该混合物被称为样品3。

[0429] 然后使用SPRI珠粒纯化该样品。使用下面的过程制备SPRI珠粒——

[0430] 1. 使用血清移液管将洗涤缓冲液 (30mL,50mMTris pH7.5,2.5M NaCl) 转移到50mL Falcon试管中。

- [0431] 2. 将速度珠粒羧酸酯修饰的磁性颗粒 (Thermo) 的存储罐 (stock pot) 涡旋以重悬所述珠粒。
- [0432] 3. 将珠粒悬浮液 (600μL) 转移到30mL的洗涤缓冲液中 (步骤1) 并涡旋以重悬所述珠粒。
- [0433] 4. 然后将Falcon试管放入DynaMag-50磁道中,使得珠粒在所述试管的邻近该磁体的一侧上积聚,并且使用血清移液管去除上清液。
- [0434] 5. 将另外的30mL的洗涤缓冲液添加到所述试管中并将所述试管充分涡旋以重悬所述珠粒。
- [0435] 6. 然后将Falcon试管放回DynaMag-50磁道中,使得所述珠粒在所述试管的邻近该磁体的一侧上积聚。
- [0436] 7. 一旦溶液已经澄清,将上清液从所述试管中去除。
- [0437] 8. 依次重复步骤5、6和7三次。
- [0438] 9. 将结合缓冲液 (30mL, 50mM Tris pH7.5, 2.5M NaCl, 20% PEG8000) 添加到来自步骤8的Falcon试管中,然后将该试管涡旋以重悬所述珠粒。
- [0439] 然后使用前将所述珠粒在4℃下储存,但是所述珠粒在纯化过程中在室温条件下使用。
- [0440] 使用如下所述的方法纯化样品3——
- [0441] 1. 将来自上述步骤9的试管涡旋以悬浮珠粒,并将18.5mL (样品3) 转移到干净的50mL的Falcon试管中。
- [0442] 2. 然后将Falcon试管放入DynaMag-50磁道中,使得所述珠粒在该试管的邻近该磁体的一侧上积聚。注意,由于珠粒悬浮于结合缓冲液 (50mM的Tris pH7.5, 2.5M NaCl, 20% PEG8000) 中,所述积聚进行超过10分钟。
- [0443] 3. 一旦溶液已经澄清,将上清液去除,添加18.5mL的结合缓冲液并将样品彻底涡旋以重悬所述珠粒。
- [0444] 4. 然后将Falcon试管放入DynaMag-50磁道中,使得珠粒在所述试管的邻近该磁体的一侧上积聚。注意,由于珠粒悬浮于结合缓冲液 (50mM Tris pH7.5, 2.5M NaCl, 20% PEG8000),珠粒积聚进行超过10分钟。
- [0445] 5. 然后重复步骤3。
- [0446] 从该时间点起,不对样品进行移液器混合或进行涡旋。
- [0447] 6. 将Falcon试管放回DynaMag-50磁道中,使得珠粒在所述试管的邻近该磁体的一侧上积聚。注意,由于所述缓冲液含有PEG8000,珠粒积聚进行超过10分钟。
- [0448] 7. 使用血清吸液管去除上清液,并用足够的结合缓冲液替换所述上清液以覆盖珠粒 (~25mL)。
- [0449] 8. 将Falcon试管留在DynaMag-50磁道中,并将所述试管沿其轴线顺时针旋转90°,使得积聚的珠粒围绕试管的侧部移动直到它们在邻近该磁体的新位置处沉降。
- [0450] 9. 将步骤8重复三次,直至所述的Falcon试管返回到其原始的位置。
- [0451] 10. 使用血清移液管去除上清液,然后除去尽可能多的残留上清液。
- [0452] 11. 将Falcon试管从DynaMag-50磁道移除,并添加洗脱缓冲液 (2.5mL, 20mM NaCl, 50mM Tris, pH7.5)。

[0453] 12. 通过轻弹将试管轻轻搅动以重悬所述珠粒, 然后将样品在环境条件下培育5分钟。

[0454] 13. 然后将所述的Falcon试管放回至DynaMag-50磁道, 使得珠粒在该试管的邻近该磁体的一侧上积聚, 将上清液转移到5mL的蛋白质LoBind微量离心管。

[0455] 14. 通过添加洗脱缓冲液将来自步骤13的洗脱物质稀释两倍。

[0456] 15. 在将所述材料分装到试管中之前, 将所述材料在4°C储存。

[0457] 然后, 如前文在实施例1中所述, 使用TBE PAGE凝胶来分析所得结合到DNA的酶, 不同在于将凝胶运行45分钟。凝胶在图5中示出, 列3示出了SPRI纯化法纯化结合到DNA Y衔接体的单个酶。

[0458] 实施例3

[0459] 本实施例描述了用于将预结合到Y衔接体的酶和结合到发夹衔接体的酶接合到基因组双链DNA的随机链的样品的制备过程。

[0460] 材料与方法

[0461] 使用NEBNext dA-尾部模块 (NEB) 将随机基因组的双链DNA序列进行dA结尾。无需进一步纯化。将dA结尾的基因组DNA (30μL) 与具有T4 Dda-E94C/C109A/C136A/A360C预结合的Y衔接体DNA (10μL 200nM) 和具有TrwC CBA-L376C/Q594A/K762C预结合的发夹衔接体 (10μL, 1 μM) 以及Blunt/TA连接酶主要混合物 (master mix) (50μL) 混合, 将样品倒置10次以混合。然后短暂地将样品在微量离心机中向下旋转。然后, 将样品在室温下培育10分钟。按照制造商的实验手册, 使用0.4倍体积的Agencourt AMPure XP珠粒纯化改造的DNA, 但使用下列洗涤缓冲液 (750mM NaCl, 10% PEG8000, 50mM Tris.HCl pH 8) 和洗脱缓冲液 (40mM CAPS pH 10, 40mM KCl和适当的DNA系链), 如下详述。

[0462] 在按照制造商的实验手册之后, 将丸状珠粒在150μL的洗涤缓冲液中浸浴30秒。然后小心地吸出洗涤缓冲液, 当心不要干扰到丸粒。然后将所述样品在微量离心机中短暂地旋转, 以充分地从珠粒排干多余的洗涤缓冲液。然后将试管重置于磁体上将珠粒制成丸状, 然后放置约1-2分钟。最后, 将剩余的洗涤缓冲液吸出, 并将丸状珠粒重悬浮于25μL的洗脱缓冲液中, 并充分混合。然后在丸状化以及除去洗脱液之前将样品放置10分钟。最后, 将在一端具有胆固醇连接的DNA链杂合到样品。这即是随后在实施例4中测试的样品DNA。

[0463] 实施例4

[0464] 本实施例说明了上述样品制备过程之后, 制备的DNA构建体具有功能预结合酶 (T4 Dda-E94C/C109A/C136A/A360C and TrwC Cba-L376C/Q594A/K762C), 该酶能够控制DNA移动穿过MspA纳米孔。

[0465] 材料与方法

[0466] 将在实施例3中制备的具有两种预结合酶 (见图6数据) 的DNA构建体 (添加到纳米孔系统中的终浓度为0.1nM) 添加到缓冲液中 (添加到纳米孔系统中的终浓度是500mM KCl, 25mM的磷酸钾pH 8.0), ATP (添加到纳米孔系统中的终浓度是2mM) 和MgCl₂ (添加到纳米孔系统的终浓度是2mM) 中。这是预混合物, 然后将其添加到纳米孔系统中 (总体积150μL)。

[0467] 电测量是在缓冲液中从嵌入到嵌段共聚物的单个MspA纳米孔获得的 (25mM磷酸钾, 150mM的亚铁氰化钾(II), 150mM铁氰化钾(III), pH 8.0)。在实现单个孔嵌入到嵌段共聚物后, 则将缓冲液 (2mL, 25mM磷酸钾, 150mM的亚铁氰化钾(II), 150mM铁氰化钾(III), pH

8.0) 流经该系统以去除任何过量的MspA纳米孔。然后将预先结合到构建体Y(有T4 Dda-E94C/C109A/C136A/A360C和TrwC Cba-L376C/Q594A/K762C两种预结合酶)的酶和燃料(MgCl₂和ATP)的预混合物(总量150μL)流入所述的单个纳米孔实验系统中,并使实验在-120mV保持电位下运行6小时(电势翻转至+60mV达2秒)并监测解旋酶控制的DNA移动。

[0468] 结果

[0469] 对于使用串联的T4 Dda-E94C/C109A/C136A/A360C和TrwC Cba-L376C/Q594A/K762C的DNA构建体Y,观察解旋酶控制的DNA移动(参照图6)。该图显示了图中标记的区域1和2的控制移位,所述区域1和2对应随机基因组双链DNA。当在发夹衔接体中存在的间隔器移位穿过纳米孔时,可观察到增加的电流,见标记3。因此该样品解旋酶控制的DNA移动表明,样品的制备过程是成功的,区域1和2在酶的控制下移位穿过纳米孔,并且可使用所述增加的电流的尖峰清楚地识别所述区域之间的过渡。

[0470] 实施例5:MuA预装载的酶衔接体

[0471] 在本实施例中,我们发现酶可以预结合到MuA衔接体,并且这不影响MuA的功能,即MuA仍可将衔接体连接到DNA。

[0472] 将T4 Dda-E94C/C109A/C136A/A360C(具有突变4C/C109A/C136A/A360C的SEQ ID NO:24)预结合到MuA Y衔接体:如图7所示,衔接体峰偏移(A是未预结合的,B是预结合的)。

[0473] 标签技术使用转座酶产生片段,并连接衔接体到基因组DNA。标签技术的条件如下:用酶预装载衔接体。使用以下来实施标签技术:100nM具有400nM T4 Dda-E94C/C109A/C136A/A360C(具有4C/C109A/C136A/A360C突变的SEQ ID NO:24)预结合衔接体的MuA四聚体。25ng ul-1靶λDNA。25mM Tris-HCl pH 8.0,20℃,10mM MgCl₂,110mM NaCl,0.05%的Triton X-100,10%甘油,30℃下1小时。

[0474] 为了在Agilent生物分析仪上可视,将T4 Dda-E94C/C109A/C136A/A360C(具有突变4C/C109A/C136A/A360C的SEQ ID NO:24)和MuA在75℃热灭活10分钟。如图8所示,没有看到有关靶DNA的标签技术的不利效应。这可通过靶DNA涂迹(smear)这一事实看到(A是未预结合的,B是预结合的)。结果显示,酶连接与否并不重要,碎片化的基因组DNA具有相同的碎片尺寸分布。

[0475] 使用与实施例4中所述仅使用T4 Dda-E94C/C109A/C136A/A360C作为解旋酶类似的过程,测试之前描述的使用标签技术方案制备的DNA样品。如图9中所示,所得到的样品库可直接添加到芯片(chip)上,并观察解旋酶控制的DNA移动。

[0476] 实施例6

[0477] 实施例描述了将两个不同的酶装载到两个DNA组分上,然后将所述两个DNA组分接合到一起并连接到基因组DNA。下文的实施例描述了两个不同的酶的装载,然而,该过程同样适用于连接两个相同的酶。

[0478] 材料与方法

[0479] 进行与实施例2中所述相同的过程,且将两个酶(E1=T4 Dda-(H82R/E94C/A360C)(具有突变H82R/E94C/A360C的SEQ ID NO:24)和E2=T4 Dda-E94C/C109A/C136A/A360C(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:24),图10中)分别装载到两个DNA构建体(图10中的A片段和END片段)上,随后按实施例2中所述纯化。A片段包含两个杂合在一起的DNA寡核苷酸(参见图10的图解,完整说明了该寡核苷酸)。END片段包含三个杂合在一起的DNA

寡核苷酸(参见图10的图解,完整说明了该寡核苷酸)。

[0480] 然后将纯化的酶装载的A片段和END片段DNA衍体使用下文描述的过程与发夹衍体3(具有5'磷酸的SEQ ID NO:39,其在3'末端连接到四个iSpC3间隔器,所述iSpC3间隔器在相对端连接到SEQ ID NO:40的5'末端)一起接合到基因组λdsDNA的3.6kb区段(模板和互补的SEQID)。

[0481] 将3.6kb基因组DNA(0.34μL,5nM的终浓度,SEQ ID NO:41杂合到SEQ ID NO:42)与A片段E1DNA(6.25uL,50nM的终浓度)、END片段E2-DNA(3.125uL,25nM的终浓度)、发夹3(1.25uL,25nM的终浓度)用接合缓冲液(5uL,2mMATP γ S,4mM MgCl₂,10mM的Hepes pH8.0,6%PEG 8000,10mM NaCl终浓度)、快速T4 DNA连接酶NEB(2.5uL,200U/uL的终浓度)和无水核酸酶(8uL)混合。将所述样品在室温下培育10分钟。最后,将在一端连接胆固醇的DNA链杂合到所述样品。样品4可不用进一步纯化而如实施例4中所述在纳米孔电生理学中使用。

[0482] 结果

[0483] 对于样品4中使用串联的T4 Dda-(H82R/E94C/A360C)和T4 Dda-E94C/C109A/C136A/A360C(参见图11)的DNA构建体(参见图11中示出的DNA构建体),观察解旋酶控制的DNA移动。该图显示了在该图中标记的区域1和2的可控移位,所述区域1和2对应基因组双链DNA。当存在于发夹衍体中的间隔器移位穿过纳米孔时,可观察到增加的电流,见箭标3。因此该样品解旋酶控制的DNA移动表明,该样品制备过程是成功的,区域1和2在酶的控制下移位穿过纳米孔,并且可使用所述增加的电流的尖峰来清楚地识别所述区域之间的过渡。

[0484] 实施例7

[0485] 本实施例描述了将解旋酶和聚合酶预装载到装载部分上,然后将所述装载部分接合到DNA的到3.6kB链。然后聚合酶被用于制造DNA的3.6kB链的模板链和互补链二者的拷贝。接合和聚合步骤的卡通示意图在图17中示出。

[0486] 材料与方法

[0487] 如实施例2中所述制备解旋酶前导复合物(在图12的A部分中示出的图),不同在于,用SEQ ID NO:43代替SEQ ID NO:26。图15泳道3显示了样品在TBE PAGE上的运行,表明酶已经结合到DNA(标记为C的条带)。

[0488] 通过将Phi29-A411C/Q560C(具有突变A411C/Q560C的SEQ ID NO:9,标记为X1)预结合到DNA发夹链(SEQ ID NO:44在其3'末端连接到四个iSpC3间隔器,所述iSpC3间隔器在相对端连接到SEQ ID NO:45的5'末端)来制备聚合酶链复合物(图12的B部分中示出的图)。Phi29-A411C/Q560C是按照实验手册(<https://www.piercenet.com/instructions/2161515.pdf>)使用Zeba 0.5mL脱盐柱(89882,Piercenet)交换成下列缓冲液(50mM的Tris pH8,20mM (NH₄)₂SO₄,10mM的MgCl₂,4%甘油)并稀释到400nM的缓冲液。将DNA发夹链(SEQ ID NO:44在其3'末端连接到4个iSpC3间隔器,所述4个iSpC3间隔器在其相对端连接到SEQ ID NO:45的5'末端)添加到phi29-A411C/Q560C,制得phi29-A411C/Q560C:DNA发夹链比例为1:1的样品(400nM)(本样品对应图15中的条带E)。将其在室温下培育15分钟。然后添加TMAD(125uM),并将样品在室温下进一步培育15分钟。然后,如之前所描述的,将样品再次进行缓冲液交换,不同在于,使用的缓冲液为(10mM HEPES pH8,10mM MgCl₂)。然后将样品与解旋酶-前导复合体(该解旋酶-前导复合体对应图15的条带C)以1:1混合,以产生图12的C部分中所示的解旋酶/聚合酶前导复合体。

[0489] 用2:1过量的发夹 (SEQ ID NO:47) ,以2:1过量,将解旋酶/聚合酶前导复合体接合到3.6kb DNA链 (SEQ ID NO:46) 。所述接合在10%的T4快速连接酶存在下,在5x ATP接合缓冲液 (5x:150mM Tris pH8,50mM MgCl₂,5mM ATP,30%PEG 8000) 中实施。将样品在室温下培育15分钟。接合步骤之后产生的构建体在图13中示出。对于聚合酶填充步骤,添加dNTP (0.5mM) 。然后在30℃将样品培育1小时。最后,SEQ ID NO:31以5倍过量添加到样品中,并将样品在室温下培育最少15分钟。

[0490] 最终样品5可不用进一步纯化而如实施例4中所述在纳米孔电生理学中使用。

[0491] 结果

[0492] 对于样品5中使用T4 Dda-E94C/C109A/C136A/A360C (参见图14) 的DNA构建体 (参见图13中示出的DNA构建体) ,观察解旋酶控制的DNA移动。该图显示了该图中标记的区域1、2、4和5的可控移位,所述区域对应原始的3.6kB DNA链 (区段1和2) 和通过聚合酶制备的互补链 (区段4和5) 。当最终构建体 (表示为x,在图14的顶部构件体图中标记为3) 的发夹中存在的间隔器移位穿过纳米孔时,可观察到增加的电流,见箭头标记3。因此该样品由解旋酶控制的DNA移动表明,样品的接合和聚合制备过程是成功的,区域1,2和聚合区域4和5在酶的控制下移位穿过纳米孔,并且可使用所述增加的电流的尖峰清楚地识别原始链区和聚合链区之间的过渡。

<110> 牛津纳米孔技术公司

<120> 方法

<130> N403885WO

<140> PCT/GB2014/050175

<141> 2014-01-22

<140> 1406151.9

<141> 2014-04-04

<140> 1406155.0

<141> 2014-04-04

<140> 1416197.0

<141> 2014-09-12

[0001] <160> 47

<170> PatentIn 3.5 版

<210> 1

<211> 558

<212> DNA

<213> 人工序列

<220>

<223> 耻垢分枝杆菌孔蛋白 A 突变体

(D90N/D91N/D93N/D118R/D134R/E193K)

<400> 1

atgggtctgg ataatgaact gagcctggtg gacggtaag atcgtaccct gacggtgcaa 60

caatggata ctttctgaa tggcggtttt cccgtggatc gtaatcgctt gaccgtgaa 120

tgtttcatt cccgtcgccgca aaaatatacgtcgcaggcc cgggtgtga cgaattcgaa 180

ggcacgctgg aactgggtta tcagattggc ttccgttgt cactggcggt tggtatcaac	240
ttctcgtaa ccacgcccgaatttctgtt aacaatggta acattaccgc accggcgttt	300
ggcctgaaca gctgttac gccgaacctg ttccgggtt ttagcatctc tgcccgcttg	360
ggcaatggtc cgggcattca agaagtggca accttagtg tgcgcgttccggcgctaa	420
ggcgggttcg cgggtctaa cggccacggt accgttacgg gggggccgg cgggtcttg	480
ctgcgtccgt tcgcgcgcct gattgcctt accggcgaca gcttacgac ctatggcgaa	540
ccgttggaaata tgaactaa	558

[0002] <210> 2

<211> 184

<212> PRT

<213> 人工序列

<220>

<223> 耻垢分枝杆菌孔蛋白 A 突变体

(D90N/D91N/D93N/D118R/D134R/E139K)

<400> 2

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu

1 5 10 15

Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp

20 25 30

Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr

35 40 45

Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu

50 55 60

Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe

65 70 75 80

Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asn Asn Gly Asn Ile Thr Ala

85 90 95

Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly

100 105 110

Val Ser Ile Ser Ala Arg Leu Gly Asn Gly Pro Gly Ile Gln Glu Val

115 120 125

[0003]

Ala Thr Phe Ser Val Arg Val Ser Gly Ala Lys Gly Gly Val Ala Val

130 135 140

Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu

145 150 155 160

Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr

165 170 175

Tyr Gly Glu Pro Trp Asn Met Asn

180

<210> 3

<211> 885

<212> DNA

<213> 人工序列

<220>

<223> α -溶血素突变体 (E111N/K147N)

<400> 3

atggcagatt ctgatattaa tattaaaacc ggtactacag atattggaag caataactaca	60
gtaaaaacag gtgatttagt cacttatgt aaagaaaaatg gcatgcacaa aaaagtattt	120
tatagttta tcgatgataa aaatcacaat aaaaaactgc tagtattag aacaaaaggt	180
accattgctg gtcaatatacg agtttatagc gaagaaggtg ctaacaaaag tggtttagcc	240
tggccttcag ccttaaggt acagttgcaa ctacctgata atgaagttagc tcaaataatct	300
gattactatc caagaaattc gattgataca aaaaactata tgagtacttt aacttatgga	360
ttcaacggta atgttactgg tgatgataca gaaaaatttgc gggccttgc tggtgcaat	420
[0004]	
gtttagattt gtcataactt gaactatgtt caacctgatt tcaaaaacaat tttagagagc	480
ccaactgata aaaaagtagg ctggaaagtgc atatttaca atatggtaa tcaaaaattgg	540
ggaccatacg atcgagattc ttggaaacccg gtatatggca atcaactttt catgaaaact	600
agaaaatggtt ctatgaaagc agcagataac ttccctgatc ctaacaagc aagttctta	660
ttatcttcag ggtttcacc agacttcgtt acagtttataat cttatggatag aaaagcatcc	720
aaacaacaaa caaatataga tggatataac gaacgaggatc gtatggatca ccaattgcatt	780
tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgttctca	840
gaaagatata aaatcgattt ggaaaaagaa gaaatgacaa attaa	885

<210> 4

<211> 293

<212> PRT

<213> 人工序列

<220>

<223> α -溶血素突变体 (E111N/K147N)

<400> 4

Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser

1 5 10 15

Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn

20 25 30

Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His

35 40 45

[0005]

Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln

50 55 60

Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp

65 70 75 80

Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala

85 90 95

Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Asn Tyr

100 105 110

Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp

115 120 125

Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His

130 135 140

Thr Leu Asn Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro

145 150 155 160

Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn

165 170 175

Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly

180 185 190

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

195 200 205

[0006]

Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe

210 215 220

Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys

225 230 235 240

Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr

245 250 255

Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp

260 265 270

Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys

275 280 285

Glu Glu Met Thr Asn

290

<210> 5

<211> 184

<212> PRT

<213> 耻垢分枝杆菌

<400> 5

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu

1

5

10

15

Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp

20

25

30

[0007]

Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr

35

40

45

Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu

50

55

60

Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe

65

70

75

80

Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala

85

90

95

Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly

100

105

110

Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
 115 120 125

Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val
 130 135 140

Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
 145 150 155 160

Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
 165 170 175

Tyr Gly Glu Pro Trp Asn Met Asn
 180

[0008]

<210> 6
 <211> 184
 <212> PRT
 <213> 耻垢分枝杆菌

<400> 6

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
 1 5 10 15

Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
 20 25 30

Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
 35 40 45

Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu

50 55 60

Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe

65 70 75 80

Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly

85 90 95

Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly

100 105 110

Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val

115 120 125

[0009]

Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val

130 135 140

Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu

145 150 155 160

Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr

165 170 175

Tyr Gly Glu Pro Trp Asn Met Asn

180

<210> 7

<211> 183

<212> PRT

<213> 耻垢分枝杆菌

<400> 7

Val Asp Asn Gln Leu Ser Val Val Asp Gly Gln Gly Arg Thr Leu Thr

1 5 10 15

Val Gln Gln Ala Glu Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg

20 25 30

Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Thr Tyr His

35 40 45

Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly

[0010] 50 55 60

Tyr Gln Val Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser

65 70 75 80

Tyr Thr Thr Pro Asn Ile Leu Ile Asp Gly Gly Asp Ile Thr Gln Pro

85 90 95

Pro Phe Gly Leu Asp Thr Ile Ile Thr Pro Asn Leu Phe Pro Gly Val

100 105 110

Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala

115 120 125

Thr Phe Ser Val Asp Val Lys Gly Ala Lys Gly Ala Val Ala Val Ser

130

135

140

Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg

145

150

155

160

Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr

165

170

175

Gly Glu Pro Trp Asn Met Asn

180

<210> 8

<211> 1830

<212> DNA

<213> 枯草芽孢杆菌噬菌体 (Bacillus subtilis phage) phi29
[0011]

<400> 8

atgaaaacaca tgccgcgtaa aatgtatagc tgccgcgttg aaaccacgac caaagtggaa 60

gattgtcgcg tttggcccta tggctacatg aacatcgaaatc atcattctga atacaaaatc 120

ggtaacagtc tggatgaatt tatggcatgg gtgctgaaatg ttccggcgaa tctgtacttc 180

cacaacctga aatttgcgtt cgcattcatt atcaactggc tggaaacgtaa tggctttaaa 240

tggagcgcgg atggctgcgc gaacacgtat aataccatta tctctgtat gggccagtgg 300

tatatgattt atatctgcct gggctacaaa ggttaaacgca aaattcatac cgtgtatcat 360

gatagcctga aaaaactgcgc gttccgggtg aagaaaattt cgaaagattt caaactgcgc 420

gttctgaaat gcgatatttttttgcataaaa gaacgtccgg ttggttacaa aatcaccccg 480

gaagaatacg catacatcaa aaacgatac cagatcatcg cagaagcgct gctgattcag 540

tttaaacagg gcctggatcg catgaccgcg ggcagtgata gcctgaaagg tttcaaagat	600
atcatcacga caaaaaaatt caaaaaagtgc ttcccgacgc tgagcctggg tctggataaa	660
gaagttcgatgtt atgcctaccg cggcggtttt acctggctga aegatcgaaa caaagaaaaa	720
gaaattggcg agggatgggt gtttgatgtt aatagtcgtt atccggcaca gatgtacagc	780
cgccctgcgc cgtatggcga accgatcggtt ttcgagggtt aatatgttg ggtatgaagat	840
tacccgcgtgc atattcagca catccgttgtt gaatttgaac taaaagaagg ctatattccg	900
accatcaga tcaaacgttag tcgttctat aagggttaacg aataacctgaa aagetctggc	960
ggtaaatcg cggatctgtt gctgatgtt aactgtatggaa agaacactac	1020
gatctgtaca acgttataa catcagggc ctgaaatttta aagccacgac cggatctgttca	1080
aaagatttca tcgataatgc gacccatc aaaaacgaccc ctgaaaggcgc gattaaacag	1140
[0012] ctggccaaac tggatgttca cagccgttat ggcataattcg cctataatcc ggtatgttca	1200
ggtaaatcg cgtacccgttca agaaaatggc gcaatgggtt ttegectggg cgaagaagaa	1260
acgaaagatc cggatgtatac cccgatgggtt gtttcatttta cggccgtggc acgttataacg	1320
accatcacgg cggcccgaggc atgtatgtatc cgcattatct actgtatgttca cgttattttt	1380
catctgacgg gcacccgaaat cccggatgtt attaaagata tcgttataatcc gaaaaaaactg	1440
ggttattttttt cccacgaaatg tacgtttttttt cgtccaaat acctgcgcgc gaaaacccatc	1500
atccaggata tctacatgaa agaagtggat ggcataactgg ttgttgcgttccggatgtat	1560
tacacccgata tcaaatttcag ttttttttttttccggatgtatccggatggatcaaaaaaagaa	1620
gtgacccatcg aaaacttcaaaatggatgttccggatgtatccggatggatcaaaaaaagaa	1680
gttccggggcg gtgtggatgtatccggatgtatccggatggatcaaaaaaagaa	1740

tggagccatc cgcaaggatcgaaaaggcggt ggctctggtg gcggttctgg cggtagtgcc 1800

tggagccacc cgcagttgaaaataataa 1830

<210> 9

<211> 608

<212> PRT

<213> 枯草芽孢杆菌噬菌体 phi29

<400> 9

Met Lys His Met Pro Arg Lys Met Tyr Ser Cys Ala Phe Glu Thr Thr

1 5 10 15

Thr Lys Val Glu Asp Cys Arg Val Trp Ala Tyr Gly Tyr Met Asn Ile

20 25 30

[0013]

Glu Asp His Ser Glu Tyr Lys Ile Gly Asn Ser Leu Asp Glu Phe Met

35 40 45

Ala Trp Val Leu Lys Val Gln Ala Asp Leu Tyr Phe His Asn Leu Lys

50 55 60

Phe Asp Gly Ala Phe Ile Ile Asn Trp Leu Glu Arg Asn Gly Phe Lys

65 70 75 80

Trp Ser Ala Asp Gly Leu Pro Asn Thr Tyr Asn Thr Ile Ile Ser Arg

85 90 95

Met Gly Gln Trp Tyr Met Ile Asp Ile Cys Leu Gly Tyr Lys Gly Lys

100 105 110

Arg Lys Ile His Thr Val Ile Tyr Asp Ser Leu Lys Lys Leu Pro Phe

115

120

125

Pro Val Lys Lys Ile Ala Lys Asp Phe Lys Leu Thr Val Leu Lys Gly

130

135

140

Asp Ile Asp Tyr His Lys Glu Arg Pro Val Gly Tyr Lys Ile Thr Pro

145

150

155

160

Glu Glu Tyr Ala Tyr Ile Lys Asn Asp Ile Gln Ile Ile Ala Glu Ala

165

170

175

Leu Leu Ile Gln Phe Lys Gln Gly Leu Asp Arg Met Thr Ala Gly Ser

180

185

190

[0014]

Asp Ser Leu Lys Gly Phe Lys Asp Ile Ile Thr Thr Lys Phe Lys

195

200

205

Lys Val Phe Pro Thr Leu Ser Leu Gly Leu Asp Lys Glu Val Arg Tyr

210

215

220

Ala Tyr Arg Gly Gly Phe Thr Trp Leu Asn Asp Arg Phe Lys Glu Lys

225

230

235

240

Glu Ile Gly Glu Gly Met Val Phe Asp Val Asn Ser Leu Tyr Pro Ala

245

250

255

Gln Met Tyr Ser Arg Leu Leu Pro Tyr Gly Glu Pro Ile Val Phe Glu

260

265

270

Gly Lys Tyr Val Trp Asp Glu Asp Tyr Pro Leu His Ile Gln His Ile

275

280

285

Arg Cys Glu Phe Glu Leu Lys Glu Gly Tyr Ile Pro Thr Ile Gln Ile

290

295

300

Lys Arg Ser Arg Phe Tyr Lys Gly Asn Glu Tyr Leu Lys Ser Ser Gly

305

310

315

320

Gly Glu Ile Ala Asp Leu Trp Leu Ser Asn Val Asp Leu Glu Leu Met

325

330

335

[0015]

Lys Glu His Tyr Asp Leu Tyr Asn Val Glu Tyr Ile Ser Gly Leu Lys

340

345

350

Phe Lys Ala Thr Thr Gly Leu Phe Lys Asp Phe Ile Asp Lys Trp Thr

355

360

365

Tyr Ile Lys Thr Thr Ser Glu Gly Ala Ile Lys Gln Leu Ala Lys Leu

370

375

380

Met Leu Asn Ser Leu Tyr Gly Lys Phe Ala Ser Asn Pro Asp Val Thr

385

390

395

400

Gly Lys Val Pro Tyr Leu Lys Glu Asn Gly Ala Leu Gly Phe Arg Leu

405

410

415

Gly Glu Glu Glu Thr Lys Asp Pro Val Tyr Thr Pro Met Gly Val Phe
 420 425 430

Ile Thr Ala Trp Ala Arg Tyr Thr Thr Ile Thr Ala Ala Gln Ala Cys
 435 440 445

Tyr Asp Arg Ile Ile Tyr Cys Asp Thr Asp Ser Ile His Leu Thr Gly
 450 455 460

Thr Glu Ile Pro Asp Val Ile Lys Asp Ile Val Asp Pro Lys Lys Leu
 465 470 475 480

Gly Tyr Trp Ala His Glu Ser Thr Phe Lys Arg Ala Lys Tyr Leu Arg
 485 490 495

[0016]

Gln Lys Thr Tyr Ile Gln Asp Ile Tyr Met Lys Glu Val Asp Gly Lys
 500 505 510

Leu Val Glu Gly Ser Pro Asp Asp Tyr Thr Asp Ile Lys Phe Ser Val
 515 520 525

Lys Cys Ala Gly Met Thr Asp Lys Ile Lys Lys Glu Val Thr Phe Glu
 530 535 540

Asn Phe Lys Val Gly Phe Ser Arg Lys Met Lys Pro Lys Pro Val Gln
 545 550 555 560

Val Pro Gly Gly Val Val Leu Val Asp Asp Thr Phe Thr Ile Lys Ser
 565 570 575

Gly Gly Ser Ala Trp Ser His Pro Gln Phe Glu Lys Gly Gly Ser
 580 585 590

Gly Gly Gly Ser Gly Gly Ser Ala Trp Ser His Pro Gln Phe Glu Lys
 595 600 605

<210> 10

<211> 1390

<212> DNA

<213> 大肠杆菌

<400> 10

atgatgaacg atggcaaaca gcagagcacc ttctgttc atgattatga aacctcggt 60

accatccgg ccctggatcg tccggcgcag ttgcggcca ttgcaccga tagcgaattc 120

[0017] aatgtgattg gcgaaacggg agtgtttat tgcacaaacgg ccgatgatta tctgcccgcag 180

ccgggtgcgg tgctgattac cggattacc cgcgcggaaag cgcgcgcgaa aggtgaaaac 240

gaagcggcgt ttgcgcgcg cattcatagc ctgttaccg tgccgaaaac ctgcattctg 300

ggctataaca atgtgcgcgtt cgatgtgaa gttaccgtt atatcttta tcttaacttt 360

tatgatccgt atgcgtggag ctggcagcat gataacagcc gttggatct gctggatgtg 420

atgcgcgcgtt gctatgcgcgtt ggcgcggaa ggcattaaatt ggccggaaaaa cgatgtggc 480

ctgcgcgcgtt ttcgtctggaa acatctgacc aaagccaaacg gcatgttgcgcgtt 540

catgtgcgtt tggccgcgtt ttatgcgcacc attgcgcgtt cggaaactggtaaaaacccgt 600

cagccgcgcgtt tgcttgcgtt tctgtttacc caccgttgcgtt aacacaaact gatggcgctg 660

attgtatgttc cgcagatgaa accgcgtgggtt catgtgcgtt gcatgtttgg cgcctggcgcgtt 720

gcaaacacca gctgggtgac cccgctggcc tggcacccgg aaaatcgaa cgcgtgatt	780
atggttgatc tggccggta tattagcccg ctgctggaa ac tggatagcga taccctgcgt	840
gaacgcgtgt ataccgccaa aaccgatctg ggccataatg ccgcgtgac ggtgaaactg	900
gttcacattt acaaatgccc ggtgctggcc caggcgaaca ccctgcgcac ggaagatgcg	960
gategtctgg gtattaaatcg ccagcattgt ctggataatc taaaatctt ggtgaaaac	1020
ccgcagggtgc gtgaaaaagt ggtggcgatc ttgcggaa ag cggaccgtt caccccgagc	1080
gataacgtgg atgcgcagct gtataacggc ttcttagcg atgcgcgtcg cgccgcgtat	1140
aaaatcggtt tggaaaccga accgcgcaat ctgcggcgcc tggatattac ctgttgcgt	1200
aaacgtattt gaaaactgtt gtttaattat cgtgcgcgca attttccggg taccctggat	1260
[0018] tatgcggaaac agcagcggtt gctggaaacat cgtgcgcagg ttgcaccc ggaatttcgt	1320
cagggtttagt cggatgaact gcagatgtt gttcagcgtt atgcgcgttga taaagaaaa	1380
gtggcgctgc	1390

<210> 11

<211> 485

<212> PRT

<213> 大肠杆菌

<400> 11

Met Met Asn Asp Gly Lys Gln Gln Ser Thr Phe Leu Phe His Asp Tyr

1 5 10 15

Glu Thr Phe Gly Thr His Pro Ala Leu Asp Arg Pro Ala Gln Phe Ala

20 25 30

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

35 40 45

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

50 55 60

Leu Ile Thr Gly Ile Thr Pro Gln Glu Ala Arg Ala Lys Gly Glu Asn

65 70 75 80

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

85 90 95

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

100 105 110

[0019]

Arg Asn Ile Phe Tyr Arg Asn Phe Tyr Asp Pro Tyr Ala Trp Ser Trp

115 120 125

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

130 135 140

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

145 150 155 160

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

165 170 175

His Ser Asn Ala His Asp Ala Met Ala Asp Val Tyr Ala Thr Ile Ala

180 185 190

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

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

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

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

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

[0020]

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

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

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

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

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

340

345

350

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

355

360

365

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

370

375

380

Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp

385

390

395

400

Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro

405

410

415

[0021]

Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg

420

425

430

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

435

440

445

Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu

450

455

460

Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His

465

470

475

480

His His His His His

485

<210>	12
<211>	804
<212>	DNA
<213>	大肠杆菌
<400>	12
atgaaatttg tcttttaa tatcaacggc ctgcgcgcca gacctcacca gcttgaagcc	60
atcgtcgaaa agcaccaacc ggatgtgatt ggcctgcagg agacaaaagt tcatgacgat	120
atgttccgc tgaagaggt ggcgaagctc ggctacaacg ttttatca cggcagaaa	180
ggccattatg gctggcgct gctgaccaaa gagacgcgga ttgcgtgcg tcggcgctt	240
cccggtgacg acgaagaggg gcagcggcgg attattatgg cggaaatccc ctcactgctg	300
gttaatgtca ccgtatcaa cggtaatcc cgcagggtg aaagccgcga ccatccgata	360
aaattcccgaa caaaagcgcgca gttttatcag aatctgcaaa actacctgga aaccgaactc	420
[0022]	
aaacgtgata atccgtact gattatggc gatataata tcagccctac agatctggat	480
atccggcattg gcaagaaaa ccgtaaagcgc tggctgcgtc ccggtaatgt ctcttctg	540
ccggaagagc gcaatggat ggacaggctg atgagctggg gtttgcgtac tacctccgc	600
catgcataatc cgcaaacagc agatgttcc tcatggtttgcattaccgcgtc aaaaggttt	660
gacgataacc gtggctgcg catgcacccgt ctgcgcgcgc gccaaccgtt ggcagaatgt	720
tgcgttagaaa ccggcatacgatc atgaaatc cgcacatgg aaaaaccgtc cgatcagcc	780
cccgctggg cgaccctccg ccgc	804
<210>	13
<211>	268
<212>	PRT
<213>	大肠杆菌

<400> 13

Met Lys Phe Val Ser Phe Asn Ile Asn Gly Leu Arg Ala Arg Pro His
1 5 10 15

Gln Leu Glu Ala Ile Val Glu Lys His Gln Pro Asp Val Ile Gly Leu
20 25 30

Gln Glu Thr Lys Val His Asp Asp Met Phe Pro Leu Glu Glu Val Ala
35 40 45

Lys Leu Gly Tyr Asn Val Phe Tyr His Gly Gln Lys Gly His Tyr Gly
50 55 60

[0023]

Val Ala Leu Leu Thr Lys Glu Thr Pro Ile Ala Val Arg Arg Gly Phe
65 70 75 80

Pro Gly Asp Asp Glu Glu Ala Gln Arg Arg Ile Ile Met Ala Glu Ile
85 90 95

Pro Ser Leu Leu Gly Asn Val Thr Val Ile Asn Gly Tyr Phe Pro Gln
100 105 110

Gly Glu Ser Arg Asp His Pro Ile Lys Phe Pro Ala Lys Ala Gln Phe
115 120 125

Tyr Gln Asn Leu Gln Asn Tyr Leu Glu Thr Glu Leu Lys Arg Asp Asn
130 135 140

Pro Val Leu Ile Met Gly Asp Met Asn Ile Ser Pro Thr Asp Leu Asp
 145 150 155 160

Ile Gly Ile Gly Glu Glu Asn Arg Lys Arg Trp Leu Arg Thr Gly Lys
 165 170 175

Cys Ser Phe Leu Pro Glu Glu Arg Glu Trp Met Asp Arg Leu Met Ser
 180 185 190

Trp Gly Leu Val Asp Thr Phe Arg His Ala Asn Pro Gln Thr Ala Asp
 195 200 205

Arg Phe Ser Trp Phe Asp Tyr Arg Ser Lys Gly Phe Asp Asp Asn Arg
 210 215 220

[0024]

Gly Leu Arg Ile Asp Leu Leu Ala Ser Gln Pro Leu Ala Glu Cys
 225 230 235 240

Cys Val Glu Thr Gly Ile Asp Tyr Glu Ile Arg Ser Met Glu Lys Pro
 245 250 255

Ser Asp His Ala Pro Val Trp Ala Thr Phe Arg Arg
 260 265

<210> 14
 <211> 1275
 <212> DNA
 <213> 嗜热栖热菌 (Thermus thermophilus)

<400> 14
 atgtttcgtc gtaaagaaga tctggatccg ccgcgtggcac tgctgccgt gaaaggcctg 60

cgcgaagccg ccgcactgctt ggaagaagcg ctgcgtcaag gtaaacgcac tcgtgttac	120
ggcgactatg atgcggatgg cctgaccggc accgcgcattc tggttcgtgg tctggccgccc	180
ctgggtgcgg atgttcatcc gtttatcccg caccgcctgg aagaaggcta tgggtgcctg	240
atggaacgcg tcccgaaaca tctggaagcc tcggacctgt ttctgaccgt tgactgcggc	300
attaccaacc atgcggaaact gcgcgaactg ctggaaaatg gcgttggaaatg cattgttacc	360
gatcatcata cgccgggcaaa aacgcgcgcg cgggtctgg tcgtgcattc ggcgcgtgacg	420
ccggatctga aagaaaaacc gaccggcgca ggctggcggt ttctgtgtgt gtgggcactg	480
catgaacgcc tgggcctgcc gcccgcgtg gaatacgcgg acctggcagc cggtggcacc	540
attgccgacg ttgcggcgat gtgggggtgg aatcgacac tggtaaaaga aggtctggca	600
[0025] cgcattccgg cttcatcttgc ggtgggcctg cgtctgttgg ctgaaggccgt gggctataacc	660
ggcaaaaggccg tcgaagtgcgc ttccgcatttgc gcgcgcgcgc tcaatgcggc ttccgcctg	720
ggcgaaggccg aaaaaggccct gcgcctgttgc ctgacggatg atgcggcaga agctcaggcg	780
ctggtcggcg aactgcacccg tctgaacgcc cgtcgatcaga ccctggaaaga agcgatgttgc	840
cgcaaaactgc tgccgcaggc cgaccggaa gcaaaaggcc tggatgttgc ttccggggaa	900
ggccatccgg gtgttatggg tattgtggcc tctgcattcc tggaaaggccac cctgcgcggc	960
gtcttcgttgc tggcccgaggc caaaggccacc gtgcgttcgc tggttcgttgc ttccggccgtc	1020
gaagcactgc gcagcgcggaa agatctgttgc ctgcgttgc tggttcgttgc agaaggccggc	1080
ggttcgccaa tggatgttgc gctgtttccg gctgttgc gcaaggccacc cacgcgttgc agcgatgttgc	1140
gcacgttcc cggatccggt tcgttgcgttgc gcaactgttgc atctgttgc gcaaccggggc	1200

ctgctgccgc aggtgttccg tgaactggca ctgctggAAC cgtatggta aggtAACCCG 1260

gaaccgctgt tcctg 1275

<210> 15

<211> 425

<212> PRT

<213> 嗜热栖热菌

<400> 15

Met Phe Arg Arg Lys Glu Asp Leu Asp Pro Pro Leu Ala Leu Leu Pro

1 5 10 15

Leu Lys Gly Leu Arg Glu Ala Ala Ala Leu Leu Glu Ala Leu Arg

20 25 30

[0026]

Gln Gly Lys Arg Ile Arg Val His Gly Asp Tyr Asp Ala Asp Gly Leu

35 40 45

Thr Gly Thr Ala Ile Leu Val Arg Gly Leu Ala Ala Leu Gly Ala Asp

50 55 60

Val His Pro Phe Ile Pro His Arg Leu Glu Glu Gly Tyr Gly Val Leu

65 70 75 80

Met Glu Arg Val Pro Glu His Leu Glu Ala Ser Asp Leu Phe Leu Thr

85 90 95

Val Asp Cys Gly Ile Thr Asn His Ala Glu Leu Arg Glu Leu Leu Glu

100 105 110

Asn Gly Val Val Ile Val Thr Asp His His Thr Pro Gly Lys Thr		
115	120	125
Pro Pro Pro Gly Leu Val Val His Pro Ala Leu Thr Pro Asp Leu Lys		
130	135	140
Glu Lys Pro Thr Gly Ala Gly Val Ala Phe Leu Leu Leu Trp Ala Leu		
145	150	155
His Glu Arg Leu Gly Leu Pro Pro Leu Glu Tyr Ala Asp Leu Ala		
165	170	175
Ala Val Gly Thr Ile Ala Asp Val Ala Pro Leu Trp Gly Trp Asn Arg		
180	185	190
[0027]		
Ala Leu Val Lys Glu Gly Leu Ala Arg Ile Pro Ala Ser Ser Trp Val		
195	200	205
Gly Leu Arg Leu Leu Ala Glu Ala Val Gly Tyr Thr Gly Lys Ala Val		
210	215	220
Glu Val Ala Phe Arg Ile Ala Pro Arg Ile Asn Ala Ala Ser Arg Leu		
225	230	235
240		
Gly Glu Ala Glu Lys Ala Leu Arg Leu Leu Leu Thr Asp Asp Ala Ala		
245	250	255
Glu Ala Gln Ala Leu Val Gly Glu Leu His Arg Leu Asn Ala Arg Arg		
260	265	270

Gln Thr Leu Glu Glu Ala Met Leu Arg Lys Leu Leu Pro Gln Ala Asp
 275 280 285

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

Val Met Gly Ile Val Ala Ser Arg Ile Leu Glu Ala Thr Leu Arg Pro
 305 310 315 320

Val Phe Leu Val Ala Gln Gly Lys Gly Thr Val Arg Ser Leu Ala Pro
 325 330 335

Ile Ser Ala Val Glu Ala Leu Arg Ser Ala Glu Asp Leu Leu Arg
 340 345 350

[0028]

Tyr Gly Gly His Lys Glu Ala Ala Gly Phe Ala Met Asp Glu Ala Leu
 355 360 365

Phe Pro Ala Phe Lys Ala Arg Val Glu Ala Tyr Ala Ala Arg Phe Pro
 370 375 380

Asp Pro Val Arg Glu Val Ala Leu Leu Asp Leu Leu Pro Glu Pro Gly
 385 390 395 400

Leu Leu Pro Gln Val Phe Arg Glu Leu Ala Leu Leu Glu Pro Tyr Gly
 405 410 415

Glu Gly Asn Pro Glu Pro Leu Phe Leu

420

425

<210> 16
<211> 738
<212> DNA
<213> 细菌噬菌体 λ

<400> 16

tccggaagcg gctctggtag tggttctggc atgacacccgg acattatcct gcagcgtacc 60

gggatcgatg tgagagctgt cgaacagggg gatgatgcgt ggcacaaatt acggctcgcc 120

gtcatcacccg cttcagaagt tcacaacgtg atagcaaaaac cccgctccgg aaagaagtgg 180

cctgacatga aaatgtccta ctccacaccc ctgcttgctg aggttgcac cggtgtggct 240

ccggaagttt acgctaaagc actggctgg ggaaaacagt acgagaacga cgccagaacc 300

[0029] ctgtttgaat tcacttccgg cgtgaatgtt actgaatccc cgatcatcta tcggcgacgaa 360

agtatgcgta ccgcctgctc tcccgatggt ttatgcagtg acggcaacgg ccttgaactg 420

aatgccccgt ttacccctcccg ggatttcatg aagtccggc tcgggtggtt cgaggccata 480

aagtca gttt acatggccca ggtgcagtac agcatgtggg tgacgcgaaa aaatgcctgg 540

cgggataaa agtacatggc gagtttgac gagatcgac cgagttcat cgaaaaaatg 660

gacgaggcac tggctgaaat tggtttgta ttggggagc aatggcgatc tggctctgg 720

tccggcagcg gttccgga

<210> 17

<211> 226

<212> PRT

<213> 细菌噬菌体λ

<400> 17

Met Thr Pro Asp Ile Ile Leu Gln Arg Thr Gly Ile Asp Val Arg Ala

1 5 10 15

Val Glu Gln Gly Asp Asp Ala Trp His Lys Leu Arg Leu Gly Val Ile

20 25 30

Thr Ala Ser Glu Val His Asn Val Ile Ala Lys Pro Arg Ser Gly Lys

35 40 45

Lys Trp Pro Asp Met Lys Met Ser Tyr Phe His Thr Leu Leu Ala Glu

50 55 60

[0030]

Val Cys Thr Gly Val Ala Pro Glu Val Asn Ala Lys Ala Leu Ala Trp

65 70 75 80

Gly Lys Gln Tyr Glu Asn Asp Ala Arg Thr Leu Phe Glu Phe Thr Ser

85 90 95

Gly Val Asn Val Thr Glu Ser Pro Ile Ile Tyr Arg Asp Glu Ser Met

100 105 110

Arg Thr Ala Cys Ser Pro Asp Gly Leu Cys Ser Asp Gly Asn Gly Leu

115 120 125

Glu Leu Lys Cys Pro Phe Thr Ser Arg Asp Phe Met Lys Phe Arg Leu

130 135 140

Gly Gly Phe Glu Ala Ile Lys Ser Ala Tyr Met Ala Gln Val Gln Tyr
 145 150 155 160

Ser Met Trp Val Thr Arg Lys Asn Ala Trp Tyr Phe Ala Asn Tyr Asp
 165 170 175

Pro Arg Met Lys Arg Glu Gly Leu His Tyr Val Val Ile Glu Arg Asp
 180 185 190

Glu Lys Tyr Met Ala Ser Phe Asp Glu Ile Val Pro Glu Phe Ile Glu
 195 200 205

Lys Met Asp Glu Ala Leu Ala Glu Ile Gly Phe Val Phe Gly Glu Gln
 210 215 220

[0031]

Trp Arg
 225

<210> 18
 <211> 760
 <212> PRT
 <213> 拟甲烷球菌属 (Methanococcoides burtonii)

<400> 18

Met Met Ile Arg Glu Leu Asp Ile Pro Arg Asp Ile Ile Gly Phe Tyr
 1 5 10 15

Glu Asp Ser Gly Ile Lys Glu Leu Tyr Pro Pro Gln Ala Glu Ala Ile
 20 25 30

Glu Met Gly Leu Leu Glu Lys Lys Asn Leu Leu Ala Ala Ile Pro Thr

35 40 45

Ala Ser Gly Lys Thr Leu Leu Ala Glu Leu Ala Met Ile Lys Ala Ile

50 55 60

Arg Glu Gly Gly Lys Ala Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala

65 70 75 80

Ser Glu Lys Phe Glu Arg Phe Lys Glu Leu Ala Pro Phe Gly Ile Lys

85 90 95

Val Gly Ile Ser Thr Gly Asp Leu Asp Ser Arg Ala Asp Trp Leu Gly

100 105 110

[0032]

Val Asn Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu

115 120 125

Arg Asn Gly Thr Ser Trp Met Asp Glu Ile Thr Thr Val Val Val Asp

130 135 140

Glu Ile His Leu Leu Asp Ser Lys Asn Arg Gly Pro Thr Leu Glu Val

145 150 155 160

Thr Ile Thr Lys Leu Met Arg Leu Asn Pro Asp Val Gln Val Val Ala

165 170 175

Leu Ser Ala Thr Val Gly Asn Ala Arg Glu Met Ala Asp Trp Leu Gly

180 185 190

Ala Ala Leu Val Leu Ser Glu Trp Arg Pro Thr Asp Leu His Glu Gly
195 200 205

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

Asp Arg Leu Glu Lys Asp Asp Ala Val Asn Leu Val Leu Asp Thr Ile
225 230 235 240

Lys Ala Glu Gly Gln Cys Leu Val Phe Glu Ser Ser Arg Arg Asn Cys
245 250 255

Ala Gly Phe Ala Lys Thr Ala Ser Ser Lys Val Ala Lys Ile Leu Asp
260 265 270

[0033]

Asn Asp Ile Met Ile Lys Leu Ala Gly Ile Ala Glu Glu Val Glu Ser
275 280 285

Thr Gly Glu Thr Asp Thr Ala Ile Val Leu Ala Asn Cys Ile Arg Lys
290 295 300

Gly Val Ala Phe His His Ala Gly Leu Asn Ser Asn His Arg Lys Leu
305 310 315 320

Val Glu Asn Gly Phe Arg Gln Asn Leu Ile Lys Val Ile Ser Ser Thr
325 330 335

Pro Thr Leu Ala Ala Gly Leu Asn Leu Pro Ala Arg Arg Val Ile Ile

340	345	350
-----	-----	-----

Arg Ser Tyr Arg Arg Phe Asp Ser Asn Phe Gly Met Gln Pro Ile Pro		
355	360	365

Val Leu Glu Tyr Lys Gln Met Ala Gly Arg Ala Gly Arg Pro His Leu		
370	375	380

Asp Pro Tyr Gly Glu Ser Val Leu Leu Ala Lys Thr Tyr Asp Glu Phe		
385	390	395
		400

Ala Gln Leu Met Glu Asn Tyr Val Glu Ala Asp Ala Glu Asp Ile Trp		
405	410	415

[0034]

Ser Lys Leu Gly Thr Glu Asn Ala Leu Arg Thr His Val Leu Ser Thr		
420	425	430

Ile Val Asn Gly Phe Ala Ser Thr Arg Gln Glu Leu Phe Asp Phe Phe		
435	440	445

Gly Ala Thr Phe Phe Ala Tyr Gln Gln Asp Lys Trp Met Leu Glu Glu		
450	455	460

Val Ile Asn Asp Cys Leu Glu Phe Leu Ile Asp Lys Ala Met Val Ser		
465	470	475
		480

Glu Thr Glu Asp Ile Glu Asp Ala Ser Lys Leu Phe Leu Arg Gly Thr		
485	490	495

Arg Leu Gly Ser Leu Val Ser Met Leu Tyr Ile Asp Pro Leu Ser Gly
 500 505 510

Ser Lys Ile Val Asp Gly Phe Lys Asp Ile Gly Lys Ser Thr Gly Gly
 515 520 525

Asn Met Gly Ser Leu Glu Asp Asp Lys Gly Asp Asp Ile Thr Val Thr
 530 535 540

Asp Met Thr Leu Leu His Leu Val Cys Ser Thr Pro Asp Met Arg Gln
 545 550 555 560

Leu Tyr Leu Arg Asn Thr Asp Tyr Thr Ile Val Asn Glu Tyr Ile Val
 565 570 575

[0035]

Ala His Ser Asp Glu Phe His Glu Ile Pro Asp Lys Leu Lys Glu Thr
 580 585 590

Asp Tyr Glu Trp Phe Met Gly Glu Val Lys Thr Ala Met Leu Leu Glu
 595 600 605

Glu Trp Val Thr Glu Val Ser Ala Glu Asp Ile Thr Arg His Phe Asn
 610 615 620

Val Gly Glu Gly Asp Ile His Ala Leu Ala Asp Thr Ser Glu Trp Leu
 625 630 635 640

Met His Ala Ala Ala Lys Leu Ala Glu Leu Leu Gly Val Glu Tyr Ser
 645 650 655

Ser His Ala Tyr Ser Leu Glu Lys Arg Ile Arg Tyr Gly Ser Gly Leu
 660 665 670

Asp Leu Met Glu Leu Val Gly Ile Arg Gly Val Gly Arg Val Arg Ala
 675 680 685

Arg Lys Leu Tyr Asn Ala Gly Phe Val Ser Val Ala Lys Leu Lys Gly
 690 695 700

Ala Asp Ile Ser Val Leu Ser Lys Leu Val Gly Pro Lys Val Ala Tyr
 705 710 715 720

Asn Ile Leu Ser Gly Ile Gly Val Arg Val Asn Asp Lys His Phe Asn
 725 730 735

[0036]

Ser Ala Pro Ile Ser Ser Asn Thr Leu Asp Thr Leu Leu Asp Lys Asn
 740 745 750

Gln Lys Thr Phe Asn Asp Phe Gln
 755 760

<210> 19
 <211> 707
 <212> PRT
 <213> 共生餐古菌 (Cenarchaeum symbiosum)

<400> 19

Met Arg Ile Ser Glu Leu Asp Ile Pro Arg Pro Ala Ile Glu Phe Leu
 1 5 10 15

Glu Gly Glu Gly Tyr Lys Lys Leu Tyr Pro Pro Gln Ala Ala Ala Ala
 20 25 30

Lys Ala Gly Leu Thr Asp Gly Lys Ser Val Leu Val Ser Ala Pro Thr
 35 40 45

Ala Ser Gly Lys Thr Leu Ile Ala Ala Ile Ala Met Ile Ser His Leu
 50 55 60

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

Ala Ala Glu Lys Phe Ala Glu Phe Gly Lys Ile Gly Gly Ile Pro Leu
 85 90 95

[0037]

Gly Arg Pro Val Arg Val Gly Val Ser Thr Gly Asp Phe Glu Lys Ala
 100 105 110

Gly Arg Ser Leu Gly Asn Asn Asp Ile Leu Val Leu Thr Asn Glu Arg
 115 120 125

Met Asp Ser Leu Ile Arg Arg Pro Asp Trp Met Asp Glu Val Gly
 130 135 140

Leu Val Ile Ala Asp Glu Ile His Leu Ile Gly Asp Arg Ser Arg Gly
 145 150 155 160

Pro Thr Leu Glu Met Val Leu Thr Lys Leu Arg Gly Leu Arg Ser Ser
 165 170 175

Pro Gln Val Val Ala Leu Ser Ala Thr Ile Ser Asn Ala Asp Glu Ile
 180 185 190

Ala Gly Trp Leu Asp Cys Thr Leu Val His Ser Thr Trp Arg Pro Val
 195 200 205

Pro Leu Ser Glu Gly Val Tyr Gln Asp Gly Glu Val Ala Met Gly Asp
 210 215 220

Gly Ser Arg His Glu Val Ala Ala Thr Gly Gly Pro Ala Val Asp
 225 230 235 240

Leu Ala Ala Glu Ser Val Ala Glu Gly Gly Gln Ser Leu Ile Phe Ala
 245 250 255

[0038]

Asp Thr Arg Ala Arg Ser Ala Ser Leu Ala Ala Lys Ala Ser Ala Val
 260 265 270

Ile Pro Glu Ala Lys Gly Ala Asp Ala Ala Lys Leu Ala Ala Ala
 275 280 285

Lys Lys Ile Ile Ser Ser Gly Gly Glu Thr Lys Leu Ala Lys Thr Leu
 290 295 300

Ala Glu Leu Val Glu Lys Gly Ala Ala Phe His His Ala Gly Leu Asn
 305 310 315 320

Gln Asp Cys Arg Ser Val Val Glu Glu Glu Phe Arg Ser Gly Arg Ile

325

330

335

Arg Leu Leu Ala Ser Thr Pro Thr Leu Ala Ala Gly Val Asn Leu Pro

340

345

350

Ala Arg Arg Val Val Ile Ser Ser Val Met Arg Tyr Asn Ser Ser Ser

355

360

365

Gly Met Ser Glu Pro Ile Ser Ile Leu Glu Tyr Lys Gln Leu Cys Gly

370

375

380

Arg Ala Gly Arg Pro Gln Tyr Asp Lys Ser Gly Glu Ala Ile Val Val

385

390

395

400

[0039]

Gly Gly Val Asn Ala Asp Glu Ile Phe Asp Arg Tyr Ile Gly Gly Glu

405

410

415

Pro Glu Pro Ile Arg Ser Ala Met Val Asp Asp Arg Ala Leu Arg Ile

420

425

430

His Val Leu Ser Leu Val Thr Thr Ser Pro Gly Ile Lys Glu Asp Asp

435

440

445

Val Thr Glu Phe Phe Leu Gly Thr Leu Gly Gly Gln Gln Ser Gly Glu

450

455

460

Ser Thr Val Lys Phe Ser Val Ala Val Ala Leu Arg Phe Leu Gln Glu

465

470

475

480

Glu Gly Met Leu Gly Arg Arg Gly Arg Gly Leu Ala Ala Thr Lys Met
 485 490 495

Gly Arg Leu Val Ser Arg Leu Tyr Met Asp Pro Met Thr Ala Val Thr
 500 505 510

Leu Arg Asp Ala Val Gly Glu Ala Ser Pro Gly Arg Met His Thr Leu
 515 520 525

Gly Phe Leu His Leu Val Ser Glu Cys Ser Glu Phe Met Pro Arg Phe
 530 535 540

Ala Leu Arg Gln Lys Asp His Glu Val Ala Glu Met Met Leu Glu Ala
 545 550 555 560

[0040]

Gly Arg Gly Glu Leu Leu Arg Pro Val Tyr Ser Tyr Glu Cys Gly Arg
 565 570 575

Gly Leu Leu Ala Leu His Arg Trp Ile Gly Glu Ser Pro Glu Ala Lys
 580 585 590

Leu Ala Glu Asp Leu Lys Phe Glu Ser Gly Asp Val His Arg Met Val
 595 600 605

Glu Ser Ser Gly Trp Leu Leu Arg Cys Ile Trp Glu Ile Ser Lys His
 610 615 620

Gln Glu Arg Pro Asp Leu Leu Gly Glu Leu Asp Val Leu Arg Ser Arg
 625 630 635 640

Val Ala Tyr Gly Ile Lys Ala Glu Leu Val Pro Leu Val Ser Ile Lys
 645 650 655

Gly Ile Gly Arg Val Arg Ser Arg Arg Leu Phe Arg Gly Gly Ile Lys
 660 665 670

Gly Pro Gly Asp Leu Ala Ala Val Pro Val Glu Arg Leu Ser Arg Val
 675 680 685

Glu Gly Ile Gly Ala Thr Leu Ala Asn Asn Ile Lys Ser Gln Leu Arg
 690 695 700

Lys Gly Gly
 705

[0041]

<210> 20
 <211> 720
 <212> PRT
 <213> 抗辐射热球菌 (Thermococcus gammatolerans)

<400> 20

Met Lys Val Asp Glu Leu Pro Val Asp Glu Arg Leu Lys Ala Val Leu
 1 5 10 15

Lys Glu Arg Gly Ile Glu Glu Leu Tyr Pro Pro Gln Ala Glu Ala Leu
 20 25 30

Lys Ser Gly Ala Leu Glu Gly Arg Asn Leu Val Leu Ala Ile Pro Thr
 35 40 45

Ala Ser Gly Lys Thr Leu Val Ser Glu Ile Val Met Val Asn Lys Leu

50 55 60

Ile Gln Glu Gly Gly Lys Ala Val Tyr Leu Val Pro Leu Lys Ala Leu

65 70 75 80

Ala Glu Glu Lys Tyr Arg Glu Phe Lys Glu Trp Glu Lys Leu Gly Leu

85 90 95

Lys Val Ala Ala Thr Thr Gly Asp Tyr Asp Ser Thr Asp Asp Trp Leu

100 105 110

Gly Arg Tyr Asp Ile Ile Val Ala Thr Ala Glu Lys Phe Asp Ser Leu

115 120 125

[0042]

Leu Arg His Gly Ala Arg Trp Ile Asn Asp Val Lys Leu Val Val Ala

130 135 140

Asp Glu Val His Leu Ile Gly Ser Tyr Asp Arg Gly Ala Thr Leu Glu

145 150 155 160

Met Ile Leu Thr His Met Leu Gly Arg Ala Gln Ile Leu Ala Leu Ser

165 170 175

Ala Thr Val Gly Asn Ala Glu Glu Leu Ala Glu Trp Leu Asp Ala Ser

180 185 190

Leu Val Val Ser Asp Trp Arg Pro Val Gln Leu Arg Arg Gly Val Phe

195 200 205

His Leu Gly Thr Leu Ile Trp Glu Asp Gly Lys Val Glu Ser Tyr Pro

210 215 220

Glu Asn Trp Tyr Ser Leu Val Val Asp Ala Val Lys Arg Gly Lys Gly

225 230 235 240

Ala Leu Val Phe Val Asn Thr Arg Arg Ser Ala Glu Lys Glu Ala Leu

245 250 255

Ala Leu Ser Lys Leu Val Ser Ser His Leu Thr Lys Pro Glu Lys Arg

260 265 270

Ala Leu Glu Ser Leu Ala Ser Gln Leu Glu Asp Asn Pro Thr Ser Glu

275 280 285

[0043]

Lys Leu Lys Arg Ala Leu Arg Gly Gly Val Ala Phe His His Ala Gly

290 295 300

Leu Ser Arg Val Glu Arg Thr Leu Ile Glu Asp Ala Phe Arg Glu Gly

305 310 315 320

Leu Ile Lys Val Ile Thr Ala Thr Pro Thr Leu Ser Ala Gly Val Asn

325 330 335

Leu Pro Ser Phe Arg Val Ile Ile Arg Asp Thr Lys Arg Tyr Ala Gly

340 345 350

Phe Gly Trp Thr Asp Ile Pro Val Leu Glu Ile Gln Gln Met Met Gly

355	360	365
-----	-----	-----

Arg Ala Gly Arg Pro Arg Tyr Asp Lys Tyr Gly Glu Ala Ile Ile Val			
---	--	--	--

370	375	380	
-----	-----	-----	--

Ala Arg Thr Asp Glu Pro Gly Lys Leu Met Glu Arg Tyr Ile Arg Gly			
---	--	--	--

385	390	395	400
-----	-----	-----	-----

Lys Pro Glu Lys Leu Phe Ser Met Leu Ala Asn Glu Gln Ala Phe Arg			
---	--	--	--

405	410	415	
-----	-----	-----	--

Ser Gln Val Leu Ala Leu Ile Thr Asn Phe Gly Ile Arg Ser Phe Pro			
---	--	--	--

420	425	430	
-----	-----	-----	--

[0044]

Glu Leu Val Arg Phe Leu Glu Arg Thr Phe Tyr Ala His Gln Arg Lys			
---	--	--	--

435	440	445	
-----	-----	-----	--

Asp Leu Ser Ser Leu Glu Tyr Lys Ala Lys Glu Val Val Tyr Phe Leu			
---	--	--	--

450	455	460	
-----	-----	-----	--

Ile Glu Asn Glu Phe Ile Asp Leu Asp Leu Glu Asp Arg Phe Ile Pro			
---	--	--	--

465	470	475	480
-----	-----	-----	-----

Leu Pro Phe Gly Lys Arg Thr Ser Gln Leu Tyr Ile Asp Pro Leu Thr			
---	--	--	--

485	490	495	
-----	-----	-----	--

Ala Lys Lys Phe Lys Asp Ala Phe Pro Ala Ile Glu Arg Asn Pro Asn			
---	--	--	--

500	505	510	
-----	-----	-----	--

Pro Phe Gly Ile Phe Gln Leu Ile Ala Ser Thr Pro Asp Met Ala Thr

515 520 525

Leu Thr Ala Arg Arg Arg Glu Met Glu Asp Tyr Leu Asp Leu Ala Tyr

530 535 540

Glu Leu Glu Asp Lys Leu Tyr Ala Ser Ile Pro Tyr Tyr Glu Asp Ser

545 550 555 560

Arg Phe Gln Gly Phe Leu Gly Gln Val Lys Thr Ala Lys Val Leu Leu

565 570 575

Asp Trp Ile Asn Glu Val Pro Glu Ala Arg Ile Tyr Glu Thr Tyr Ser

580 585 590

[0045]

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

595 600 605

Met Tyr Ser Leu Ile Glu Leu Tyr Lys Leu Phe Glu Pro Lys Glu Glu

610 615 620

Ile Leu Asn Tyr Leu Arg Asp Leu His Leu Arg Leu Arg His Gly Val

625 630 635 640

Arg Glu Glu Leu Leu Glu Leu Val Arg Leu Pro Asn Ile Gly Arg Lys

645 650 655

Arg Ala Arg Ala Leu Tyr Asn Ala Gly Phe Arg Ser Val Glu Ala Ile

660 665 670

Ala Asn Ala Lys Pro Ala Glu Leu Leu Ala Val Glu Gly Ile Gly Ala
 675 680 685

Lys Ile Leu Asp Gly Ile Tyr Arg His Leu Gly Ile Glu Lys Arg Val
 690 695 700

Thr Glu Glu Lys Pro Lys Arg Lys Gly Thr Leu Glu Asp Phe Leu Arg
 705 710 715 720

<210> 21

<211> 799

<212> PRT

<213> 亨氏甲烷螺旋菌 (Methanospirillum hungatei)

<400> 21

[0046]

Met Glu Ile Ala Ser Leu Pro Leu Pro Asp Ser Phe Ile Arg Ala Cys
 1 5 10 15

His Ala Lys Gly Ile Arg Ser Leu Tyr Pro Pro Gln Ala Glu Cys Ile
 20 25 30

Glu Lys Gly Leu Leu Glu Gly Lys Asn Leu Leu Ile Ser Ile Pro Thr
 35 40 45

Ala Ser Gly Lys Thr Leu Leu Ala Glu Met Ala Met Trp Ser Arg Ile
 50 55 60

Ala Ala Gly Gly Lys Cys Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala
 65 70 75 80

Ser	Glu	Lys	Tyr	Asp	Glu	Phe	Ser	Lys	Lys	Gly	Val	Ile	Arg	Val	Gly
85							90						95		
Ile Ala Thr Gly Asp Leu Asp Arg Thr Asp Ala Tyr Leu Gly Glu Asn															
100							105						110		
Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu Arg Asn															
115							120						125		
Arg Thr Pro Trp Leu Ser Gln Ile Thr Cys Ile Val Leu Asp Glu Val															
130							135						140		
His Leu Ile Gly Ser Glu Asn Arg Gly Ala Thr Leu Glu Met Val Ile															
145							150						155		160
[0047]															
Thr Lys Leu Arg Tyr Thr Asn Pro Val Met Gln Ile Ile Gly Leu Ser															
165							170						175		
Ala Thr Ile Gly Asn Pro Ala Gln Leu Ala Glu Trp Leu Asp Ala Thr															
180							185						190		
Leu Ile Thr Ser Thr Trp Arg Pro Val Asp Leu Arg Gln Gly Val Tyr															
195							200						205		
Tyr Asn Gly Lys Ile Arg Phe Ser Asp Ser Glu Arg Pro Ile Gln Gly															
210							215						220		
Lys Thr Lys His Asp Asp Leu Asn Leu Cys Leu Asp Thr Ile Glu Glu															
225							230						235		240

Gly	Gly	Gln	Cys	Leu	Val	Phe	Val	Ser	Ser	Arg	Arg	Asn	Ala	Glu	Gly
245															255
Phe	Ala	Lys	Lys	Ala	Ala	Gly	Ala	Leu	Lys	Ala	Gly	Ser	Pro	Asp	Ser
260															270
Lys	Ala	Leu	Ala	Gln	Glu	Leu	Arg	Arg	Leu	Arg	Asp	Arg	Asp	Glu	Gly
275															285
Asn	Val	Leu	Ala	Asp	Cys	Val	Glu	Arg	Gly	Ala	Ala	Phe	His	His	Ala
290															300
Gly	Leu	Ile	Arg	Gln	Glu	Arg	Thr	Ile	Ile	Glu	Glu	Gly	Phe	Arg	Asn
305															320
[0048]															
Gly	Tyr	Ile	Glu	Val	Ile	Ala	Ala	Thr	Pro	Thr	Leu	Ala	Ala	Gly	Leu
325															335
Asn	Leu	Pro	Ala	Arg	Arg	Val	Ile	Ile	Arg	Asp	Tyr	Asn	Arg	Phe	Ala
340															350
Ser	Gly	Leu	Gly	Met	Val	Pro	Ile	Pro	Val	Gly	Glu	Tyr	His	Gln	Met
355															365
Ala	Gly	Arg	Ala	Gly	Arg	Pro	His	Leu	Asp	Pro	Tyr	Gly	Glu	Ala	Val
370															380
Leu	Leu	Ala	Lys	Asp	Ala	Pro	Ser	Val	Glu	Arg	Leu	Phe	Glu	Thr	Phe

385

390

395

400

Ile Asp Ala Glu Ala Glu Arg Val Asp Ser Gln Cys Val Asp Asp Ala

405

410

415

Ser Leu Cys Ala His Ile Leu Ser Leu Ile Ala Thr Gly Phe Ala His

420

425

430

Asp Gln Glu Ala Leu Ser Ser Phe Met Glu Arg Thr Phe Tyr Phe Phe

435

440

445

Gln His Pro Lys Thr Arg Ser Leu Pro Arg Leu Val Ala Asp Ala Ile

450

455

460

[0049]

Arg Phe Leu Thr Thr Ala Gly Met Val Glu Glu Arg Glu Asn Thr Leu

465

470

475

480

Ser Ala Thr Arg Leu Gly Ser Leu Val Ser Arg Leu Tyr Leu Asn Pro

485

490

495

Cys Thr Ala Arg Leu Ile Leu Asp Ser Leu Lys Ser Cys Lys Thr Pro

500

505

510

Thr Leu Ile Gly Leu Leu His Val Ile Cys Val Ser Pro Asp Met Gln

515

520

525

Arg Leu Tyr Leu Lys Ala Ala Asp Thr Gln Leu Leu Arg Thr Phe Leu

530

535

540

Phe Lys His Lys Asp Asp Leu Ile Leu Pro Leu Pro Phe Glu Gln Glu			
545	550	555	560
Glu Glu Glu Leu Trp Leu Ser Gly Leu Lys Thr Ala Leu Val Leu Thr			
565	570	575	
Asp Trp Ala Asp Glu Phe Ser Glu Gly Met Ile Glu Glu Arg Tyr Gly			
580	585	590	
Ile Gly Ala Gly Asp Leu Tyr Asn Ile Val Asp Ser Gly Lys Trp Leu			
595	600	605	
Leu His Gly Thr Glu Arg Leu Val Ser Val Glu Met Pro Glu Met Ser			
610	615	620	
[0050]			
Gln Val Val Lys Thr Leu Ser Val Arg Val His His Gly Val Lys Ser			
625	630	635	640
Glu Leu Leu Pro Leu Val Ala Leu Arg Asn Ile Gly Arg Val Arg Ala			
645	650	655	
Arg Thr Leu Tyr Asn Ala Gly Tyr Pro Asp Pro Glu Ala Val Ala Arg			
660	665	670	
Ala Gly Leu Ser Thr Ile Ala Arg Ile Ile Gly Glu Gly Ile Ala Arg			
675	680	685	
Gln Val Ile Asp Glu Ile Thr Gly Val Lys Arg Ser Gly Ile His Ser			
690	695	700	

Ser Asp Asp Asp Tyr Gln Gln Lys Thr Pro Glu Leu Leu Thr Asp Ile
 705 710 715 720

Pro Gly Ile Gly Lys Lys Met Ala Glu Lys Leu Gln Asn Ala Gly Ile
 725 730 735

Ile Thr Val Ser Asp Leu Leu Thr Ala Asp Glu Val Leu Leu Ser Asp
 740 745 750

Val Leu Gly Ala Ala Arg Ala Arg Lys Val Leu Ala Phe Leu Ser Asn
 755 760 765

Ser Glu Lys Glu Asn Ser Ser Asp Lys Thr Glu Glu Ile Pro Asp
 770 775 780

[0051]

Thr Gln Lys Ile Arg Gly Gln Ser Ser Trp Glu Asp Phe Gly Cys
 785 790 795

<210> 22
 <211> 1756
 <212> PRT
 <213> 大肠杆菌

<400> 22

Met Met Ser Ile Ala Gln Val Arg Ser Ala Gly Ser Ala Gly Asn Tyr
 1 5 10 15

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

Trp Ala Gly Lys Gly Ala Glu Gln Leu Gly Leu Gln Gly Ser Val Asp

35

40

45

Lys Asp Val Phe Thr Arg Leu Leu Glu Gly Arg Leu Pro Asp Gly Ala

50

55

60

Asp Leu Ser Arg Met Gln Asp Gly Ser Asn Lys His Arg Pro Gly Tyr

65

70

75

80

Asp Leu Thr Phe Ser Ala Pro Lys Ser Val Ser Met Met Ala Met Leu

85

90

95

Gly Gly Asp Lys Arg Leu Ile Asp Ala His Asn Gln Ala Val Asp Phe

100

105

110

[0052]

Ala Val Arg Gln Val Glu Ala Leu Ala Ser Thr Arg Val Met Thr Asp

115

120

125

Gly Gln Ser Glu Thr Val Leu Thr Gly Asn Leu Val Met Ala Leu Phe

130

135

140

Asn His Asp Thr Ser Arg Asp Gln Glu Pro Gln Leu His Thr His Ala

145

150

155

160

Val Val Ala Asn Val Thr Gln His Asn Gly Glu Trp Lys Thr Leu Ser

165

170

175

Ser Asp Lys Val Gly Lys Thr Gly Phe Ile Glu Asn Val Tyr Ala Asn

180

185

190

Gln Ile Ala Phe Gly Arg Leu Tyr Arg Glu Lys Leu Lys Glu Gln Val
 195 200 205

Glu Ala Leu Gly Tyr Glu Thr Glu Val Val Gly Lys His Gly Met Trp
 210 215 220

Glu Met Pro Gly Val Pro Val Glu Ala Phe Ser Gly Arg Ser Gln Ala
 225 230 235 240

Ile Arg Glu Ala Val Gly Glu Asp Ala Ser Leu Lys Ser Arg Asp Val
 245 250 255

Ala Ala Leu Asp Thr Arg Lys Ser Lys Gln His Val Asp Pro Glu Ile
 260 265 270

[0053]

Arg Met Ala Glu Trp Met Gln Thr Leu Lys Glu Thr Gly Phe Asp Ile
 275 280 285

Arg Ala Tyr Arg Asp Ala Ala Asp Gln Arg Thr Glu Ile Arg Thr Gln
 290 295 300

Ala Pro Gly Pro Ala Ser Gln Asp Gly Pro Asp Val Gln Gln Ala Val
 305 310 315 320

Thr Gln Ala Ile Ala Gly Leu Ser Glu Arg Lys Val Gln Phe Thr Tyr
 325 330 335

Thr Asp Val Leu Ala Arg Thr Val Gly Ile Leu Pro Pro Glu Asn Gly

340	345	350
-----	-----	-----

Val Ile Glu Arg Ala Arg Ala Gly Ile Asp Glu Ala Ile Ser Arg Glu		
---	--	--

355	360	365
-----	-----	-----

Gln Leu Ile Pro Leu Asp Arg Glu Lys Gly Leu Phe Thr Ser Gly Ile		
---	--	--

370	375	380
-----	-----	-----

His Val Leu Asp Glu Leu Ser Val Arg Ala Leu Ser Arg Asp Ile Met		
---	--	--

385	390	395	400
-----	-----	-----	-----

Lys Gln Asn Arg Val Thr Val His Pro Glu Lys Ser Val Pro Arg Thr		
---	--	--

405	410	415
-----	-----	-----

[0054]

Ala Gly Tyr Ser Asp Ala Val Ser Val Leu Ala Gln Asp Arg Pro Ser		
---	--	--

420	425	430
-----	-----	-----

Leu Ala Ile Val Ser Gly Gln Gly Ala Ala Gly Gln Arg Glu Arg		
---	--	--

435	440	445
-----	-----	-----

Val Ala Glu Leu Val Met Met Ala Arg Glu Gln Gly Arg Glu Val Gln		
---	--	--

450	455	460
-----	-----	-----

Ile Ile Ala Ala Asp Arg Arg Ser Gln Met Asn Leu Lys Gln Asp Glu		
---	--	--

465	470	475	480
-----	-----	-----	-----

Arg Leu Ser Gly Glu Leu Ile Thr Gly Arg Arg Gln Leu Leu Glu Gly		
---	--	--

485	490	495
-----	-----	-----

Met Ala Phe Thr Pro Gly Ser Thr Val Ile Val Asp Gln Gly Glu Lys

500

505

510

Leu Ser Leu Lys Glu Thr Leu Thr Leu Leu Asp Gly Ala Ala Arg His

515

520

525

Asn Val Gln Val Leu Ile Thr Asp Ser Gly Gln Arg Thr Gly Thr Gly

530

535

540

Ser Ala Leu Met Ala Met Lys Asp Ala Gly Val Asn Thr Tyr Arg Trp

545

550

555

560

Gln Gly Gly Glu Gln Arg Pro Ala Thr Ile Ile Ser Glu Pro Asp Arg

565

570

575

[0055]

Asn Val Arg Tyr Ala Arg Leu Ala Gly Asp Phe Ala Ala Ser Val Lys

580

585

590

Ala Gly Glu Glu Ser Val Ala Gln Val Ser Gly Val Arg Glu Gln Ala

595

600

605

Ile Leu Thr Gln Ala Ile Arg Ser Glu Leu Lys Thr Gln Gly Val Leu

610

615

620

Gly His Pro Glu Val Thr Met Thr Ala Leu Ser Pro Val Trp Leu Asp

625

630

635

640

Ser Arg Ser Arg Tyr Leu Arg Asp Met Tyr Arg Pro Gly Met Val Met

645

650

655

Glu Gln Trp Asn Pro Glu Thr Arg Ser His Asp Arg Tyr Val Ile Asp
 660 665 670

Arg Val Thr Ala Gln Ser His Ser Leu Thr Leu Arg Asp Ala Gln Gly
 675 680 685

Glu Thr Gln Val Val Arg Ile Ser Ser Leu Asp Ser Ser Trp Ser Leu
 690 695 700

Phe Arg Pro Glu Lys Met Pro Val Ala Asp Gly Glu Arg Leu Arg Val
 705 710 715 720

Thr Gly Lys Ile Pro Gly Leu Arg Val Ser Gly Gly Asp Arg Leu Gln
 725 730 735

[0056]

Val Ala Ser Val Ser Glu Asp Ala Met Thr Val Val Val Pro Gly Arg
 740 745 750

Ala Glu Pro Ala Ser Leu Pro Val Ser Asp Ser Pro Phe Thr Ala Leu
 755 760 765

Lys Leu Glu Asn Gly Trp Val Glu Thr Pro Gly His Ser Val Ser Asp
 770 775 780

Ser Ala Thr Val Phe Ala Ser Val Thr Gln Met Ala Met Asp Asn Ala
 785 790 795 800

Thr Leu Asn Gly Leu Ala Arg Ser Gly Arg Asp Val Arg Leu Tyr Ser
 805 810 815

Ser Leu Asp Glu Thr Arg Thr Ala Glu Lys Leu Ala Arg His Pro Ser
 820 825 830

Phe Thr Val Val Ser Glu Gln Ile Lys Ala Arg Ala Gly Glu Thr Leu
 835 840 845

Leu Glu Thr Ala Ile Ser Leu Gln Lys Ala Gly Leu His Thr Pro Ala
 850 855 860

Gln Gln Ala Ile His Leu Ala Leu Pro Val Leu Glu Ser Lys Asn Leu
 865 870 875 880

Ala Phe Ser Met Val Asp Leu Leu Thr Glu Ala Lys Ser Phe Ala Ala
 885 890 895

[0057]

Glu Gly Thr Gly Phe Thr Glu Leu Gly Gly Glu Ile Asn Ala Gln Ile
 900 905 910

Lys Arg Gly Asp Leu Leu Tyr Val Asp Val Ala Lys Gly Tyr Gly Thr
 915 920 925

Gly Leu Leu Val Ser Arg Ala Ser Tyr Glu Ala Glu Lys Ser Ile Leu
 930 935 940

Arg His Ile Leu Glu Gly Lys Glu Ala Val Thr Pro Leu Met Glu Arg
 945 950 955 960

Val Pro Gly Glu Leu Met Glu Thr Leu Thr Ser Gly Gln Arg Ala Ala

965

970

975

Thr Arg Met Ile Leu Glu Thr Ser Asp Arg Phe Thr Val Val Gln Gly
 980 985 990

Tyr Ala Gly Val Gly Lys Thr Thr Gln Phe Arg Ala Val Met Ser Ala
 995 1000 1005

Val Asn Met Leu Pro Ala Ser Glu Arg Pro Arg Val Val Gly Leu
 1010 1015 1020

Gly Pro Thr His Arg Ala Val Gly Glu Met Arg Ser Ala Gly Val
 1025 1030 1035

[0058]

Asp Ala Gln Thr Leu Ala Ser Phe Leu His Asp Thr Gln Leu Gln
 1040 1045 1050

Gln Arg Ser Gly Glu Thr Pro Asp Phe Ser Asn Thr Leu Phe Leu
 1055 1060 1065

Leu Asp Glu Ser Ser Met Val Gly Asn Thr Glu Met Ala Arg Ala
 1070 1075 1080

Tyr Ala Leu Ile Ala Ala Gly Gly Gly Arg Ala Val Ala Ser Gly
 1085 1090 1095

Asp Thr Asp Gln Leu Gln Ala Ile Ala Pro Gly Gln Ser Phe Arg
 1100 1105 1110

Leu Gln Gln Thr Arg Ser Ala Ala Asp Val Val Ile Met Lys Glu

1115

1120

1125

Ile Val Arg Gln Thr Pro Glu Leu Arg Glu Ala Val Tyr Ser Leu

1130

1135

1140

Ile Asn Arg Asp Val Glu Arg Ala Leu Ser Gly Leu Glu Ser Val

1145

1150

1155

Lys Pro Ser Gln Val Pro Arg Leu Glu Gly Ala Trp Ala Pro Glu

1160

1165

1170

His Ser Val Thr Glu Phe Ser His Ser Gln Glu Ala Lys Leu Ala

1175

1180

1185

[0059]

Glu Ala Gln Gln Lys Ala Met Leu Lys Gly Glu Ala Phe Pro Asp

1190

1195

1200

Ile Pro Met Thr Leu Tyr Glu Ala Ile Val Arg Asp Tyr Thr Gly

1205

1210

1215

Arg Thr Pro Glu Ala Arg Glu Gln Thr Leu Ile Val Thr His Leu

1220

1225

1230

Asn Glu Asp Arg Arg Val Leu Asn Ser Met Ile His Asp Ala Arg

1235

1240

1245

Glu Lys Ala Gly Glu Leu Gly Lys Glu Gln Val Met Val Pro Val

1250

1255

1260

Leu Asn Thr Ala Asn Ile Arg Asp Gly Glu Leu Arg Arg Leu Ser
1265 1270 1275

Thr Trp Glu Lys Asn Pro Asp Ala Leu Ala Leu Val Asp Asn Val
1280 1285 1290

Tyr His Arg Ile Ala Gly Ile Ser Lys Asp Asp Gly Leu Ile Thr
1295 1300 1305

Leu Gln Asp Ala Glu Gly Asn Thr Arg Leu Ile Ser Pro Arg Glu
1310 1315 1320

Ala Val Ala Glu Gly Val Thr Leu Tyr Thr Pro Asp Lys Ile Arg
1325 1330 1335

[0060]

Val Gly Thr Gly Asp Arg Met Arg Phe Thr Lys Ser Asp Arg Glu
1340 1345 1350

Arg Gly Tyr Val Ala Asn Ser Val Trp Thr Val Thr Ala Val Ser
1355 1360 1365

Gly Asp Ser Val Thr Leu Ser Asp Gly Gln Gln Thr Arg Val Ile
1370 1375 1380

Arg Pro Gly Gln Glu Arg Ala Glu Gln His Ile Asp Leu Ala Tyr
1385 1390 1395

Ala Ile Thr Ala His Gly Ala Gln Gly Ala Ser Glu Thr Phe Ala
1400 1405 1410

Ile Ala Leu Glu Gly Thr Glu Gly Asn Arg Lys Leu Met Ala Gly
 1415 1420 1425

Phe Glu Ser Ala Tyr Val Ala Leu Ser Arg Met Lys Gln His Val
 1430 1435 1440

Gln Val Tyr Thr Asp Asn Arg Gln Gly Trp Thr Asp Ala Ile Asn
 1445 1450 1455

Asn Ala Val Gln Lys Gly Thr Ala His Asp Val Leu Glu Pro Lys
 1460 1465 1470

Pro Asp Arg Glu Val Met Asn Ala Gln Arg Leu Phe Ser Thr Ala
 1475 1480 1485

[0061]

Arg Glu Leu Arg Asp Val Ala Ala Gly Arg Ala Val Leu Arg Gln
 1490 1495 1500

Ala Gly Leu Ala Gly Gly Asp Ser Pro Ala Arg Phe Ile Ala Pro
 1505 1510 1515

Gly Arg Lys Tyr Pro Gln Pro Tyr Val Ala Leu Pro Ala Phe Asp
 1520 1525 1530

Arg Asn Gly Lys Ser Ala Gly Ile Trp Leu Asn Pro Leu Thr Thr
 1535 1540 1545

Asp Asp Gly Asn Gly Leu Arg Gly Phe Ser Gly Glu Gly Arg Val

1550 1555 1560

Lys Gly Ser Gly Asp Ala Gln Phe Val Ala Leu Gln Gly Ser Arg
1565 1570 1575

Asn Gly Glu Ser Leu Leu Ala Asp Asn Met Gln Asp Gly Val Arg
1580 1585 1590

Ile Ala Arg Asp Asn Pro Asp Ser Gly Val Val Val Arg Ile Ala
1595 1600 1605

Gly Glu Gly Arg Pro Trp Asn Pro Gly Ala Ile Thr Gly Gly Arg
1610 1615 1620

[0062]

Val Trp Gly Asp Ile Pro Asp Asn Ser Val Gln Pro Gly Ala Gly
1625 1630 1635

Asn Gly Glu Pro Val Thr Ala Glu Val Leu Ala Gln Arg Gln Ala
1640 1645 1650

Glu Glu Ala Ile Arg Arg Glu Thr Glu Arg Arg Ala Asp Glu Ile
1655 1660 1665

Val Arg Lys Met Ala Glu Asn Lys Pro Asp Leu Pro Asp Gly Lys
1670 1675 1680

Thr Glu Leu Ala Val Arg Asp Ile Ala Gly Gln Glu Arg Asp Arg
1685 1690 1695

Ser Ala Ile Ser Glu Arg Glu Thr Ala Leu Pro Glu Ser Val Leu
 1700 1705 1710

Arg Glu Ser Gln Arg Glu Arg Glu Ala Val Arg Glu Val Ala Arg
 1715 1720 1725

Glu Asn Leu Leu Gln Glu Arg Leu Gln Gln Met Glu Arg Asp Met
 1730 1735 1740

Val Arg Asp Leu Gln Lys Glu Lys Thr Leu Gly Gly Asp
 1745 1750 1755

<210> 23

<211> 726

[0063] <212> PRT

<213> 拟甲烷球菌属

<400> 23

Met Ser Asp Lys Pro Ala Phe Met Lys Tyr Phe Thr Gln Ser Ser Cys
 1 5 10 15

Tyr Pro Asn Gln Gln Glu Ala Met Asp Arg Ile His Ser Ala Leu Met
 20 25 30

Gln Gln Gln Leu Val Leu Phe Glu Gly Ala Cys Gly Thr Gly Lys Thr
 35 40 45

Leu Ser Ala Leu Val Pro Ala Leu His Val Gly Lys Met Leu Gly Lys
 50 55 60

Thr Val Ile Ile Ala Thr Asn Val His Gln Gln Met Val Gln Phe Ile

65 70 75 80

Asn Glu Ala Arg Asp Ile Lys Lys Val Gln Asp Val Lys Val Ala Val

85 90 95

Ile Lys Gly Lys Thr Ala Met Cys Pro Gln Glu Ala Asp Tyr Glu Glu

100 105 110

Cys Ser Val Lys Arg Glu Asn Thr Phe Glu Leu Met Glu Thr Glu Arg

115 120 125

Glu Ile Tyr Leu Lys Arg Gln Glu Leu Asn Ser Ala Arg Asp Ser Tyr

130 135 140

[0064]

Lys Lys Ser His Asp Pro Ala Phe Val Thr Leu Arg Asp Glu Leu Ser

145 150 155 160

Lys Glu Ile Asp Ala Val Glu Glu Lys Ala Arg Gly Leu Arg Asp Arg

165 170 175

Ala Cys Asn Asp Leu Tyr Glu Val Leu Arg Ser Asp Ser Glu Lys Phe

180 185 190

Arg Glu Trp Leu Tyr Lys Glu Val Arg Ser Pro Glu Glu Ile Asn Asp

195 200 205

His Ala Ile Lys Asp Gly Met Cys Gly Tyr Glu Leu Val Lys Arg Glu

210 215 220

Leu Lys His Ala Asp Leu Leu Ile Cys Asn Tyr His His Val Leu Asn
 225 230 235 240

Pro Asp Ile Phe Ser Thr Val Leu Gly Trp Ile Glu Lys Glu Pro Gln
 245 250 255

Glu Thr Ile Val Ile Phe Asp Glu Ala His Asn Leu Glu Ser Ala Ala
 260 265 270

Arg Ser His Ser Ser Leu Ser Leu Thr Glu His Ser Ile Glu Lys Ala
 275 280 285

Ile Thr Glu Leu Glu Ala Asn Leu Asp Leu Leu Ala Asp Asp Asn Ile
 290 295 300

[0065]

His Asn Leu Phe Asn Ile Phe Leu Glu Val Ile Ser Asp Thr Tyr Asn
 305 310 315 320

Ser Arg Phe Lys Phe Gly Glu Arg Glu Arg Val Arg Lys Asn Trp Tyr
 325 330 335

Asp Ile Arg Ile Ser Asp Pro Tyr Glu Arg Asn Asp Ile Val Arg Gly
 340 345 350

Lys Phe Leu Arg Gln Ala Lys Gly Asp Phe Gly Glu Lys Asp Asp Ile
 355 360 365

Gln Ile Leu Leu Ser Glu Ala Ser Glu Leu Gly Ala Lys Leu Asp Glu
 370 375 380

Thr Tyr Arg Asp Gln Tyr Lys Lys Gly Leu Ser Ser Val Met Lys Arg
 385 390 395 400

Ser His Ile Arg Tyr Val Ala Asp Phe Met Ser Ala Tyr Ile Glu Leu
 405 410 415

Ser His Asn Leu Asn Tyr Tyr Pro Ile Leu Asn Val Arg Arg Asp Met
 420 425 430

Asn Asp Glu Ile Tyr Gly Arg Val Glu Leu Phe Thr Cys Ile Pro Lys
 435 440 445

Asn Val Thr Glu Pro Leu Phe Asn Ser Leu Phe Ser Val Ile Leu Met
 450 455 460

[0066]

Ser Ala Thr Leu His Pro Phe Glu Met Val Lys Lys Thr Leu Gly Ile
 465 470 475 480

Thr Arg Asp Thr Cys Glu Met Ser Tyr Gly Thr Ser Phe Pro Glu Glu
 485 490 495

Lys Arg Leu Ser Ile Ala Val Ser Ile Pro Pro Leu Phe Ala Lys Asn
 500 505 510

Arg Asp Asp Arg His Val Thr Glu Leu Leu Glu Gln Val Leu Leu Asp
 515 520 525

Ser Ile Glu Asn Ser Lys Gly Asn Val Ile Leu Phe Phe Gln Ser Ala

530

535

540

Phe Glu Ala Lys Arg Tyr Tyr Ser Lys Ile Glu Pro Leu Val Asn Val

545

550

555

560

Pro Val Phe Leu Asp Glu Val Gly Ile Ser Ser Gln Asp Val Arg Glu

565

570

575

Glu Phe Phe Ser Ile Gly Glu Glu Asn Gly Lys Ala Val Leu Leu Ser

580

585

590

Tyr Leu Trp Gly Thr Leu Ser Glu Gly Ile Asp Tyr Arg Asp Gly Arg

595

600

605

[0067]

Gly Arg Thr Val Ile Ile Ile Gly Val Gly Tyr Pro Ala Leu Asn Asp

610

615

620

Arg Met Asn Ala Val Glu Ser Ala Tyr Asp His Val Phe Gly Tyr Gly

625

630

635

640

Ala Gly Trp Glu Phe Ala Ile Gln Val Pro Thr Ile Arg Lys Ile Arg

645

650

655

Gln Ala Met Gly Arg Val Val Arg Ser Pro Thr Asp Tyr Gly Ala Arg

660

665

670

Ile Leu Leu Asp Gly Arg Phe Leu Thr Asp Ser Lys Lys Arg Phe Gly

675

680

685

Lys Phe Ser Val Phe Glu Val Phe Pro Pro Ala Glu Arg Ser Glu Phe

690

695

700

Val Asp Val Asp Pro Glu Lys Val Lys Tyr Ser Leu Met Asn Phe Phe

705

710

715

720

Met Asp Asn Asp Glu Gln

725

<210> 24

<211> 439

<212> PRT

<213> 未知

<220>

<223> Dda 1993

[0068]

<400> 24

Met Thr Phe Asp Asp Leu Thr Glu Gly Gln Lys Asn Ala Phe Asn Ile

1

5

10

15

Val Met Lys Ala Ile Lys Glu Lys Lys His His Val Thr Ile Asn Gly

20

25

30

Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr Lys Phe Ile Ile Glu Ala

35

40

45

Leu Ile Ser Thr Gly Glu Thr Gly Ile Ile Leu Ala Ala Pro Thr His

50

55

60

Ala Ala Lys Lys Ile Leu Ser Lys Leu Ser Gly Lys Glu Ala Ser Thr

65

70

75

80

Ile His Ser Ile Leu Lys Ile Asn Pro Val Thr Tyr Glu Glu Asn Val

85

90

95

Leu Phe Glu Gln Lys Glu Val Pro Asp Leu Ala Lys Cys Arg Val Leu

100

105

110

Ile Cys Asp Glu Val Ser Met Tyr Asp Arg Lys Leu Phe Lys Ile Leu

115

120

125

Leu Ser Thr Ile Pro Pro Trp Cys Thr Ile Ile Gly Ile Gly Asp Asn

130

135

140

[0069]

Lys Gln Ile Arg Pro Val Asp Pro Gly Glu Asn Thr Ala Tyr Ile Ser

145

150

155

160

Pro Phe Phe Thr His Lys Asp Phe Tyr Gln Cys Glu Leu Thr Glu Val

165

170

175

Lys Arg Ser Asn Ala Pro Ile Ile Asp Val Ala Thr Asp Val Arg Asn

180

185

190

Gly Lys Trp Ile Tyr Asp Lys Val Val Asp Gly His Gly Val Arg Gly

195

200

205

Phe Thr Gly Asp Thr Ala Leu Arg Asp Phe Met Val Asn Tyr Phe Ser

210

215

220

Ile Val Lys Ser Leu Asp Asp Leu Phe Glu Asn Arg Val Met Ala Phe
 225 230 235 240

Thr Asn Lys Ser Val Asp Lys Leu Asn Ser Ile Ile Arg Lys Lys Ile
 245 250 255

Phe Glu Thr Asp Lys Asp Phe Ile Val Gly Glu Ile Ile Val Met Gln
 260 265 270

Glu Pro Leu Phe Lys Thr Tyr Lys Ile Asp Gly Lys Pro Val Ser Glu
 275 280 285

Ile Ile Phe Asn Asn Gly Gln Leu Val Arg Ile Ile Glu Ala Glu Tyr
 290 295 300

[0070]

Thr Ser Thr Phe Val Lys Ala Arg Gly Val Pro Gly Glu Tyr Leu Ile
 305 310 315 320

Arg His Trp Asp Leu Thr Val Glu Thr Tyr Gly Asp Asp Glu Tyr Tyr
 325 330 335

Arg Glu Lys Ile Lys Ile Ile Ser Ser Asp Glu Glu Leu Tyr Lys Phe
 340 345 350

Asn Leu Phe Leu Gly Lys Thr Ala Glu Thr Tyr Lys Asn Trp Asn Lys
 355 360 365

Gly Gly Lys Ala Pro Trp Ser Asp Phe Trp Asp Ala Lys Ser Gln Phe
 370 375 380

Ser Lys Val Lys Ala Leu Pro Ala Ser Thr Phe His Lys Ala Gln Gly
 385 390 395 400

Met Ser Val Asp Arg Ala Phe Ile Tyr Thr Pro Cys Ile His Tyr Ala
 405 410 415

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

Arg Tyr Asp Val Phe Tyr Val
 435

<210> 25
 <211> 970
 [0071] <212> PRT
 <213> 肉毒梭菌 (Clostridium botulinum)
 <400> 25

Met Leu Ser Val Ala Asn Val Arg Ser Pro Ser Ala Ala Ser Tyr
 1 5 10 15

Phe Ala Ser Asp Asn Tyr Tyr Ala Ser Ala Asp Ala Asp Arg Ser Gly
 20 25 30

Gln Trp Ile Gly Asp Gly Ala Lys Arg Leu Gly Leu Glu Gly Lys Val
 35 40 45

Glu Ala Arg Ala Phe Asp Ala Leu Leu Arg Gly Glu Leu Pro Asp Gly
 50 55 60

Ser	Ser	Val	Gly	Asn	Pro	Gly	Gln	Ala	His	Arg	Pro	Gly	Thr	Asp	Leu
65															80
Thr	Phe	Ser	Val	Pro	Lys	Ser	Trp	Ser	Leu	Leu	Ala	Leu	Val	Gly	Lys
															95
Asp	Glu	Arg	Ile	Ile	Ala	Ala	Tyr	Arg	Glu	Ala	Val	Val	Glu	Ala	Leu
															110
His	Trp	Ala	Glu	Lys	Asn	Ala	Ala	Glu	Thr	Arg	Val	Val	Glu	Lys	Gly
															125
Met	Val	Val	Thr	Gln	Ala	Thr	Gly	Asn	Leu	Ala	Ile	Gly	Leu	Phe	Gln
															140
[0072]															
His	Asp	Thr	Asn	Arg	Asn	Gln	Glu	Pro	Asn	Leu	His	Phe	His	Ala	Val
															160
Ile	Ala	Asn	Val	Thr	Gln	Gly	Lys	Asp	Gly	Lys	Trp	Arg	Thr	Leu	Lys
															175
Asn	Asp	Arg	Leu	Trp	Gln	Leu	Asn	Thr	Thr	Leu	Asn	Ser	Ile	Ala	Met
															190
Ala	Arg	Phe	Arg	Val	Ala	Val	Glu	Lys	Leu	Gly	Tyr	Glu	Pro	Gly	Pro
															205
Val	Leu	Lys	His	Gly	Asn	Phe	Glu	Ala	Arg	Gly	Ile	Ser	Arg	Glu	Gln
															220

Val Met Ala Phe Ser Thr Arg Arg Lys Glu Val Leu Glu Ala Arg Arg
 225 230 235 240

Gly Pro Gly Leu Asp Ala Gly Arg Ile Ala Ala Leu Asp Thr Arg Ala
 245 250 255

Ser Lys Glu Gly Ile Glu Asp Arg Ala Thr Leu Ser Lys Gln Trp Ser
 260 265 270

Glu Ala Ala Gln Ser Ile Gly Leu Asp Leu Lys Pro Leu Val Asp Arg
 275 280 285

Ala Arg Thr Lys Ala Leu Gly Gln Gly Met Glu Ala Thr Arg Ile Gly
 290 295 300

[0073]

Ser Leu Val Glu Arg Gly Arg Ala Trp Leu Ser Arg Phe Ala Ala His
 305 310 315 320

Val Arg Gly Asp Pro Ala Asp Pro Leu Val Pro Pro Ser Val Leu Lys
 325 330 335

Gln Asp Arg Gln Thr Ile Ala Ala Gln Ala Val Ala Ser Ala Val
 340 345 350

Arg His Leu Ser Gln Arg Glu Ala Ala Phe Glu Arg Thr Ala Leu Tyr
 355 360 365

Lys Ala Ala Leu Asp Phe Gly Leu Pro Thr Thr Ile Ala Asp Val Glu

370 375 380

Lys Arg Thr Arg Ala Leu Val Arg Ser Gly Asp Leu Ile Ala Gly Lys

385 390 395 400

Gly Glu His Lys Gly Trp Leu Ala Ser Arg Asp Ala Val Val Thr Glu

405 410 415

Gln Arg Ile Leu Ser Glu Val Ala Ala Gly Lys Gly Asp Ser Ser Pro

420 425 430

Ala Ile Thr Pro Gln Lys Ala Ala Ala Ser Val Gln Ala Ala Ala Leu

435 440 445

[0074]

Thr Gly Gln Gly Phe Arg Leu Asn Glu Gly Gln Leu Ala Ala Ala Arg

450 455 460

Leu Ile Leu Ile Ser Lys Asp Arg Thr Ile Ala Val Gln Gly Ile Ala

465 470 475 480

Gly Ala Gly Lys Ser Ser Val Leu Lys Pro Val Ala Glu Val Leu Arg

485 490 495

Asp Glu Gly His Pro Val Ile Gly Leu Ala Ile Gln Asn Thr Leu Val

500 505 510

Gln Met Leu Glu Arg Asp Thr Gly Ile Gly Ser Gln Thr Leu Ala Arg

515 520 525

Phe Leu Gly Gly Trp Asn Lys Leu Leu Asp Asp Pro Gly Asn Val Ala

530 535 540

Leu Arg Ala Glu Ala Gln Ala Ser Leu Lys Asp His Val Leu Val Leu

545 550 555 560

Asp Glu Ala Ser Met Val Ser Asn Glu Asp Lys Glu Lys Leu Val Arg

565 570 575

Leu Ala Asn Leu Ala Gly Val His Arg Leu Val Leu Ile Gly Asp Arg

580 585 590

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

595 600 605

[0075]

Arg Ala Gly Ile Ala Arg Ala Glu Met Ala Thr Asn Leu Arg Ala Arg

610 615 620

Asp Pro Val Val Arg Glu Ala Gln Ala Ala Ala Gln Ala Gly Asp Val

625 630 635 640

Arg Lys Ala Leu Arg His Leu Lys Ser His Thr Val Glu Ala Arg Gly

645 650 655

Asp Gly Ala Gln Val Ala Ala Glu Thr Trp Leu Ala Leu Asp Lys Glu

660 665 670

Thr Arg Ala Arg Thr Ser Ile Tyr Ala Ser Gly Arg Ala Ile Arg Ser

675 680 685

Ala Val Asn Ala Ala Val Gln Gln Gly Leu Leu Ala Ser Arg Glu Ile			
690	695	700	
Gly Pro Ala Lys Met Lys Leu Glu Val Leu Asp Arg Val Asn Thr Thr			
705	710	715	720
Arg Glu Glu Leu Arg His Leu Pro Ala Tyr Arg Ala Gly Arg Val Leu			
725	730	735	
Glu Val Ser Arg Lys Gln Gln Ala Leu Gly Leu Phe Ile Gly Glu Tyr			
740	745	750	
Arg Val Ile Gly Gln Asp Arg Lys Gly Lys Leu Val Glu Val Glu Asp			
755	760	765	
[0076]			
Lys Arg Gly Lys Arg Phe Arg Phe Asp Pro Ala Arg Ile Arg Ala Gly			
770	775	780	
Lys Gly Asp Asp Asn Leu Thr Leu Leu Glu Pro Arg Lys Leu Glu Ile			
785	790	795	800
His Glu Gly Asp Arg Ile Arg Trp Thr Arg Asn Asp His Arg Arg Gly			
805	810	815	
Leu Phe Asn Ala Asp Gln Ala Arg Val Val Glu Ile Ala Asn Gly Lys			
820	825	830	
Val Thr Phe Glu Thr Ser Lys Gly Asp Leu Val Glu Leu Lys Lys Asp			
835	840	845	

Asp Pro Met Leu Lys Arg Ile Asp Leu Ala Tyr Ala Leu Asn Val His

850 855 860

Met Ala Gln Gly Leu Thr Ser Asp Arg Gly Ile Ala Val Met Asp Ser

865 870 875 880

Arg Glu Arg Asn Leu Ser Asn Gln Lys Thr Phe Leu Val Thr Val Thr

885 890 895

Arg Leu Arg Asp His Leu Thr Leu Val Val Asp Ser Ala Asp Lys Leu

900 905 910

Gly Ala Ala Val Ala Arg Asn Lys Gly Glu Lys Ala Ser Ala Ile Glu

915 920 925

[0077]

Val Thr Gly Ser Val Lys Pro Thr Ala Thr Lys Gly Ser Gly Val Asp

930 935 940

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

945 950 955 960

Ser Lys Ser Lys Thr Leu Asp Phe Gly Ile

965 970

<210> 26

<211> 68

<212> DNA

<213> 人工序列

<220>

<223> 实施例 1 中使用的序列

<400> 26

ggtaaacac ccaagcagac gccgcaatat cagcaccaac agaaacaacc tttgaggcga 60

gcggtaaa

68

<210> 27

<211> 12

<212> DNA

<213> 人工序列

<220>

<223> 实施例 2 和 7 中使用的序列

<400> 27

tttttttt tt

12

[0078]

<210> 28

<211> 51

<212> DNA

<213> 人工序列

<220>

<223> 实施例 2 和 7 中使用的序列

<400> 28

ggttgttct gtttgtgctg atattgcggc gtctgcttgg gtgtttaacc t

51

<210> 29

<211> 48

<212> DNA

<213> 人工序列

<220>

<223> 实施例 1 中使用的序列

<400> 29

cgttctgttt atgtttcttg gacactgatt gacacggttt agtagaacac 48

<210> 30

<211> 49

<212> DNA

<213> 人工序列

<220>

<223> 实施例 1 中使用的序列

<400> 30

ttttttttt tttttttttt ttttttttca agaaacataa acagaacgt 49

<210> 31

[0079] <211> 15

<212> PRT

<213> 人工序列

<220>

<223> 实施例 7 中使用的序列

<400> 31

Thr Thr Gly Ala Cys Cys Gly Cys Thr Cys Gly Cys Cys Thr Cys

1

5

10

15

<210> 32

<211> 10

<212> DNA

<213> 人工序列

<220>

<223> 实施例 6 中使用的序列

<400> 32	
tttttttttt	10
<210> 33	
<211> 29	
<212> DNA	
<213> 人工序列	
<220>	
<223> 实施例 6 中使用的序列	
<400> 33	
ggttgtttct gttgggtgctg atattgcac	29
[0080]	<210> 34
	<211> 22
	<212> DNA
	<213> 人工序列
<220>	
<223> 实施例 6 中使用的序列	
<400> 34	
gcaatatcag caccaacaga aa	22
[0080]	<210> 35
	<211> 15
	<212> DNA
	<213> 人工序列
<220>	
<223> 实施例 6 中使用的序列	
<400> 35	
gaggcgagcg gtcaa	15

<210> 36
<211> 32
<212> DNA
<213> 人工序列

<220>
<223> 实施例 6 中使用的序列

<400> 36
tgagtgacca atcagctacg tttttttttt tt

32

<210> 37
<211> 28
<212> DNA
<213> 人工序列

<220>
[0081] <223> 实施例 6 中使用的序列

<400> 37
ggttgtttct gttggtgctg atattgtt

28

<210> 38
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 实施例 6 中使用的序列

<400> 38
cgttagctgat tggtaactca gt

22

<210> 39

<211> 48

<212> DNA

<213> 人工序列

<220>

<223> 实施例 6 中使用的序列

<400> 39

cgtcctgtcg ctgtgtctcg gacactgatt gacacggttt agtagagc

48

<210> 40

<211> 52

<212> DNA

<213> 人工序列

<220>

<223> 实施例 6 中使用的序列

[0082] <400> 40

ttttttttt tttttttttt ttttttttctg agacacagcg acaggacgtc ct

52

<210> 41

<211> 3568

<212> DNA

<213> 人工序列

<220>

<223> 实施例 6 中使用的序列

<400> 41

gccccatcgat tgtgtttgtt agtcgtttt tttttttgga atttttttt tggaattttt

60

tttttgcgtt aacaacccccc tgccgttttgc cccgtgcata tcggcacgca acaaatctga

120

ttactaaaca cagtagccgtg gatttttttct atcagtaatc gacccattttc ctaattaaat

180

agagcaaatc cccttattgg gggtaagaca tgaagatgcc agaaaaacat gacccgttgg

240

ccggcattct cgcggcaaaag gaacaaggca tcggggcaat cttcgctt gcaatggcgt	300
accttcgccc cagatataat ggccgtgcgt ttacaaaaac agtaatcgac gcaacgatgt	360
gcgcattat cgcctagtcc attcgtaacc ttctcgactt cgccggacta agtagcaatc	420
tcgttatata aacgagcgtg tttatcggtt acatcggtac tgactcgatt ggttcgctt	480
tcaaacgctt cgctgtaaa aaagccggag tagaagatgg tagaaatcaa taatcaacgt	540
aaggcggtcc tcgatatgtt ggcgtggcgg gagggactg ataacggacg tcagaaaacc	600
agaaatcatg gttatgacgt cattgttaggc ggagagatatttactgatta ctcggatcac	660
cctcgcaaac ttgtcacgtt aaacccaaaaa ctcaaatcaa caggcgccgg acgttaccag	720
cctttttccc gttgggtggga tgcctaccgc aagcagcttgcgtt gctgaaaga cttctctccg	780
[0083] aaaagtcaagg acgtgtggc attgcagcag attaaggagc gtggcgctt acctatgatt	840
gatcggtggat atatccgtca ggcaatcgac cgttgcggca atatctggc ttcaactgcgg	900
ggcgctgggtt atggtcagtt cgagcataag gctgacagcc tgattgcaaa attcaaagaa	960
ggggcggaa cggtcagaga gattgtatgtt tgagcagagt caccggcattt atctccgcgc	1020
tgttatctg catcatcgcc tgcctgtcat gggctgttaa tcattaccgt gataacgcca	1080
ttacctacaa agcccgccgc gacaaaaatgc ccagagaact gaagctggcg aacggggcaaa	1140
ttactgacat gcagatgcgtt cagcgtgtatg ttgtcgctt cgttgcggatc tacacgaagg	1200
agtttagctgtt gctaaagct gaaaatgtt ctctcggttga tgatgttgcg gctggcg	1260
gtgggttgcacatcaaaggca gtcgtcaagt cagtgcgttga agccaccacc gcttccggcg	1320
tggataatgc agccctcccccc cggactggcag acaccgttga acgggatttttacccctca	1380

gagagaggct gatcaactatg caaaaacaac tggaaggaac ccagaagtat attaatgagc	1440
agtgcagata gagttgccta tatcgatgg caactcatgc aattattgtg agcaatacac	1500
acgcgettcc agcggagtagt aaatgcctaa agtaataaaa ccgagcaatc cattacgaa	1560
tgttgctgg gttctgttt taacaacatt ttctgcgcgg ccacaaattt tggctgcata	1620
gacagtttc ttctgccaa ttccagaaac gaagaaatga tgggtgatgg tttccttgg	1680
tgtactgtc gcgggttgtt tttgaacagt aaacgtctgt tgagcacatc ctgtataag	1740
cagggccagc gcagtagcga gtagcattt ttcatggtg ttattccaga tgcgtttga	1800
atgtcgcaga atcgatgtg tagaaaatta aacaaccct aaacaatgag ttgaaattc	1860
atattgttaa tatttattaa tgtatgtcag gtgcgtgaa tgcattgtt attccggat	1920
taactatgtc cacagccctg acggggaaact ttctgcggg agtgcggg aataattaa	1980
[0084]	
acatgcaca cagggttag cgcgtacacg tattgcatta tgccaacgcc cgggtgtga	2040
cacggaagaa accggacgtt atgatttagc gtggaaagat ttgtgtatgt ttctgaatgc	2100
tctcgtaaa tagtaatgaa ttatcaaagg tatagtataa tctttatgt tcatggat	2160
ttgttaacca tcggaaaact cctgcattttt caagatttc cctgtattgc tgaatgtga	2220
tttcttgc ttcaaccta tcataggacg ttctataag atgcgtgtt ctggagaatt	2280
taacatttac aacctttta agtccttta ttaacacgggt gttatcgatgtcaacacga	2340
tgtaatatt atctgtggct agatagtaaa tataatgtga gacgtgtga cgtttttgtt	2400
cagaataaaa caattcacag tctaaatctt ttgcacttg atcgaatatt tctttaaaaa	2460
tgcaacctg agccattggt aaaacccctcc atgtatacg agggcgcgtt gtttgcattt	2520
tgcgtttat cgttcaatc tggctgtacc ttctgtgtt ttgtgtatgtaa tttatgtcaa	2580

atattaggaa tggttcaact taatagtatt ggttgcgtaa caaagtgcgg tctgtgtggc	2640
attctggagg gaaataacaac cgacagatgt atgtaaggcc aacgtgtca aatcttcata	2700
cagaaagatt tgaagtaata tttaaccgc tagatgaaga gcaaggcat ggagcgacaa	2760
aatgaataaa gaacaatctg ctgatgatec ctccgtggat ctgattcgtg taaaaatat	2820
gcttaatagc accatttcta tggatgttacc tgatgttgta attgcatgtg tagaacataa	2880
ggtgtctctg gaagcattca gagcaattga ggcagcggtg gtgaaggcag ataataat	2940
gaaggattat tccctggtgg ttgactgate accataactg ctaatcatc aaactatttta	3000
gtctgtgaca gagccaacac gcagtcgtc actgtcagga aagtggtaaa actgcaactc	3060
aattactgca atgcctcgt aattaagtga atttacaata tgcgttgcgtt cggagggaaag	3120
aacgcgggat gtcattttt catcaatttta aatgtatgtat tatgtctctt ttctgtact	3180
[0085]	
tagtctcga cggcaggcattt caatgaccca ggctgagaaaa ttccggacc cttttgctc	3240
aagagcgatg ttaattttttt caatcatttgc gtttagaaag cggatgtgc gggttgtgt	3300
tctgcgggtt ctgttctcg ttgacatgag gttggcccgat attcgtgtc gtcgtttgt	3360
attgtctgaa gttgtttttt cgttaagtttgc atgcagatca attaatacgtt tacctgcgtc	3420
ataatttgatt atttgacgtg gtttgatggc ctccacgcac gttgtgtat gtagatgata	3480
atcattatca ctttacgggt ctttccggat gaaaaaaaaaaag gtaccaaaaaaa aaacatcgat	3540
gtgagtagtg aaccgttacgc atgttagga	3568
<210> 42	
<211> 3565	
<212> DNA	

<213> 人工序列

<220>

<223> 实施例 6 中使用的序列

<400> 42

acatgcctac ggftcaactac tcacgacgt gtttttttg gtacctttt ttcacccgga 60

aaggacccgt aaagtgataa tgattatcat ctacatatca caacgtgegt ggaggccatc 120

aaaccacgta aaataatcaa ttatgacgca ggtatcgat taattgatct gcatcaactt 180

aacgtaaaaa caacttcaga caatacaaat cagcgacact gaatacgggg caacctcatg 240

tcaacgaaga acagaacccg cagaacaaca acccgcaaca tccgcttcc taaccaaatg 300

attgaacaaa ttaacatcgc tcttgagcaa aaagggtccg ggaatttctc agcctgggtc 360

attgaagcct gccgtcgag actaacgtca gaaaagagag catatacatc aattaaaagt 420

[0086]

gatgaagaat gaacatcccg egttttccc tccgaacagg acgtattgt aaattcactt 480

aattacgagg gcattgcagt aattgagttg cagtttacc actttctga cagtgcacaga 540

ctgcgtgttgc tctgtcac agactaaata gtttgaatga ttagcagtta tggtgatcag 600

tcaaccacca gggaaataatc cttcatatta ttatcgatc tcaccaacgc tgcctcaatt 660

gtctgaatg cttccagaga caccttatgt tctatacatg caattacaac atcaggtaa 720

ctcatagaaa tggcttatt aagcatattt ttacacgaa tcagatccac ggagggatca 780

tcagcagatt gtttttatt cattttgtcg ctccatgcgc tgcgtttca tctagcggtt 840

aaaatattac ttcaaatctt tctgtatgaa gatttgagca cgttggcatt acatacatct 900

gtcggttgta ttccctcca gaatgccagc aggaccgcac ttgttacgc aaccaatact 960

attaagtgaa aacattcta atattgaca taaatcatca acaaaacaca aggaggcag 1020

accagattga aacgataaaaa acgataatgc aaactacgctg ccctcgatc acatggagg	1080
tttaccaat ggctcagggtt gccattttta aagaatattt cgtcaagtgc gaaaagatt	1140
tagactgtga attgtttat tctgaactaa aacgtcacaa cgtctcacat tatatttact	1200
atctagccac agataatatt cacatcggt tagaaaacga taacaccgtg ttaataaaaag	1260
gacttaaaaaa ggtttaaat gttaaattct caagaaacac gcatcttata gaaacgtct	1320
atgatagggtt gaaatcaaga gaaatcacat ttccagcaata caggaaaaat ctgtctaaag	1380
caggagttt ccgatgggtt acaaataatcc atgaacataa aagatattac tataccttg	1440
ataattcatt actatttact gagagcattc agaacactac acaaatctt ccacgttaaa	1500
tcataacgtc cgggttcttc cgtgtcagca cggggcggtt ggcataatgc aatacgtgt	1560
[0087] cgcgttaaac cctgtgtca tcgttttaat tattccggta cactccgca gagaagttcc	1620
ccgtcagggc tgtggacata gtaatccgg gaatacaatg acgattcatc gcacctgaca	1680
tacattaata aatattaaca atatgaaatt tcaactcatt gtttagggtt tgtttaattt	1740
tctacacata cgattctgctg aacttcggaa agcatcggtt ataacaccat gaaaaaaatg	1800
ctactcgcta ctgcgtggc cctgttttattt acaggatgtg ctcaacagac gtttactgtt	1860
caaaacaaac cggcagcagt agcaccaaaag gaaaccatca cccatcattt ctgcgttct	1920
ggaattgggc agaagaaaaac tgtcgatca gccaaaattt gtggcgccgc agaaaatgtt	1980
gtttaaaacag aaacccagca aacattcgta aatggattgc tcggttttat tacttttaggc	2040
atttataactc cgctggaaagc gctgtgtat tgcacaaat aattgcatga gttggccatc	2100
gatatgggca actctatctg cactgtctat taatatactt ctgggttctt tccagtttt	2160

ttgcatagt gatcagecctc tctctgaggg taaaataatc cgggtcagcg gtgtctgc	2220	
gtggggggga ggctgcatta tccacgcccgg aggccggtggt ggcttcacgc actgactgac	2280	
agactgcgtt gatgtgeaac cgacgacgac cagcggcaac atcatcacgc agagcatcat	2340	
ttcagcttt agcatcagct aactccctcg tgcatttgc atcgagcgca gcaacatcac	2400	
gtgacgcat ctgcattgtca gtaattgcgg cgttgcgcag cttcagttct ctggcatttt	2460	
tgcgcgtg ggctttgttag gtaatggcgt tatacggta atgattaaca gccccatgaca	2520	
ggcagacgat gatgcagata accagagcgg agataatgc ggtgactctg ctcatacatc	2580	
aatctctctg accgttccgc cccgttcttt gaatttgca atcaggtgt cagccttatg	2640	
ctgcactga ccataaccag cgcgcggcag tgaagcccg atattgcgc aacggcgtat	2700	
tgcgtacgg atatcaccac gatcaatcat aggtaaagcg ccacgtctt taatctgtc	2760	
[0088]	caatgccaca ggcgttgcac tttcggaga gaagtcttc aggccaagct gcttgcggta	2820
	ggcatcccac caacgggaaa gaagctggta ggcgttgcgc cctgttgcatt tgagtttgg	2880
	gttagcgtg acaagttgc gagggtgatc ggagtaatca gtaaatagct ctccgcctac	2940
	aatgacgtca taaccatgtat ttctggttt ctgacgtccg ttatcgttc ctccgcacca	3000
	cgccagcata tcgaggaacg cttacgttg attattgatt tctaccatct tctactccgg	3060
	cttttttagc agcgaaggcg ttgataagcg aaccaatcga gtcagtaccg atgtagccga	3120
	taaacacgct cgttatataa gggagattgc tacttagtcc ggcgaagtcg agaagggtcac	3180
	gaatgaacta ggcgataatg ggcacatcg ttgcgtcgat tactgtttt gtaaacgcac	3240
	cgccattata tctgcgcga aggtacgcca ttgcaaacgc aaggatgcc cccatgcctt	3300
	gttccttgc cgcgagaatg gggccaaaca ggtcatgttt ttctggcata ttcatgttt	3360

acccccaata aggggatttg ctctatttaa ttaggaataa ggtcgattac tgatagaaca	3420
aatccaggct actgtgtta gtaatcagat ttgtcgtga ccgatatgca cgggcaaaac	3480
ggcaggaggt tgtagcgc aaaaaaaaaat tccaaaaaaaaaa aaattccaaa aaaaaaaaaagc	3540
gactaacaaa cacaatctga tggca	3565

<210> 43

<211> 45

<212> DNA

<213> 人工序列

<220>

<223> 实施例 7 中使用的序列

<400> 43

[0089] gcaatatcag caccaacaga aacaacctt gaggcgagcg gtcaa	45
---	----

<210> 44

<211> 54

<212> DNA

<213> 人工序列

<220>

<223> 实施例 7 中使用的序列

<400> 44

ggtaaacac ccaagcagac gcctttttt ttctgtctg tcgtgtgtc tcgg	54
---	----

<210> 45

<211> 20

<212> DNA

<213> 人工序列

<220>

<223> 实施例 7 中使用的序列

<400> 45

cgagacacag cgacaggacg

20

<210> 46

<211> 3560

<212> DNA

<213> 人工序列

<220>

<223> 实施例 7 中使用的序列

<400> 46

gccatcagat tggtttgtt agtgcgtttt ttttttttggaaatttttt 60

ttttgcgtt aacaacctcc tgcgttttgc cccgtgcata tcggcacga acaaatctga 120

[0090]

ttactaaaca cagtagcctg gattgttct atcagtaatc gaccttatttc ctaatataat 180

agagcaaatc cccttattgg gggtaagaca tgaagatgcc agaaaaacat gacctgttgg 240

ccgcattctt cggggcaaaag gaacaaggca tcggggcaat cttgcgtttt gcaatggcgt 300

accttcggg cagatataat ggcgggtcg ttacaaaaac agtaatcgac gcaacgtatgt 360

gcgcattat cgcctagttt attcgtgacc ttctcgactt cgcggacta agtagcaatc 420

tcgttatat aacgagcggtt ttatcggtt acatcggtac tgactcgatt ggtcgctta 480

tcaaacgctt cgctgttaaa aaagccggag tagaagatgg tagaaatcaa taatcaacgt 540

aaggcggtcc tcgatatgtt ggcgtggcgtt gagggaactg ataacggacg tcagaaaacc 600

agaaaatcatg gttatgacgt cattgttagcc ggagagctat ttactgtttaa ctccgatcac 660

ctcgcaaac ttgtcacgtt aaacccaaaa ctcaaataa cagggcgccgg acgttaccag 720

cttccttccc gttgggtggga tgcctaccgc aagcagcttg gcctgaaaga cttctctccg	780
aaaagtgcagg acgcgtgtggc attgcagcag attaaggagc gtggcgctt acctatgatt	840
gatcgtggtg atatccgtca ggcaatcgac cggtgcagca atatctggc ttcactgccg	900
ggcgcgtgggtt atggtcagtt cgagcataag gctgacagcc tgattgcaaa attcaaagaa	960
gccccggaa cggtcagaga gattgatgta tgagcagagt caccgcgatt atctccgtc	1020
tggttatctg catcatcgtc tgcctgtcat gggctgttaa tcattaccgt gataacgcca	1080
ttacctacaa agcccagcgc gacaaaaatg ccagagaact gaagctggcg aacgcggcaa	1140
ttactgacat gcagatgcgt cagcgtgatg ttgcgtcgct cgtgcacaaa tacacgaagg	1200
agtttagctga tgctaaagct gaaaatgatg ctctgcgtga tgaatgtgcc gctggtcgtc	1260
[0091] gtcgggttgc a catcaaagca gtcgtcagt cagtgcgtga agccaccacc gcctccggcg	1320
tggataatgc agectcccc c gactggcag acaccgctga acgggattat ttcaccctca	1380
gagagaggt gatcaactatg caaaaacaac tggaaaggaaac ccagaagtat attaatgagc	1440
agtgcagata gagttgc cca tatcgatggg caactcatgc aattattgtg agcaatacac	1500
acgcgcgttcc agcggaggtat aaatgcctaa agtaataaaa cccggcaatc cattacgaa	1560
tgtttgttgg gtttctgttt taacaacatt ttctgcgeeg ccacaaattt tggctgcate	1620
gacagtttc ttctgc cca ttccagaaac gaagaaatga tgggtgttgg ttcccttgg	1680
tgcgtactgtc gcccgggttgtt ttgaacagt aaacgtctgt tgagcacatc ctgtataag	1740
caggccca gca gtcgtacgca gtagcatttt ttcatgggtt ttattcccgaa tgcgtttgtt	1800
aggcgcaga atcgatgtg tagaaaat aacaaaccct aaacaatgag ttgaaatttc	1860

atattgttaa tatttattaa tgtatgtcag gtgcgatgaa tcgtcattgt attccggat	1920
taactatgtc cacagccctg acgggaaact tcctgcggg agtgtccggg aataattaa	1980
acgatgcaca cagggttag cgegtacacg tattgcatta tgccaacgac ccgggtgtga	2040
cacggaagaa accggacgtt atgatttagc gtggaaagat ttgtgttagt ttctgaatgc	2100
tctcagtaaa tagtaatgaa ttatcaaagg tatagtaata tctttatgt tcatggatat	2160
ttgtAACCCa tcggaaaact cctgttttag caagatttc cctgtattgc tgaaatgtga	2220
tttctttga ttcaaccta tcataggacg ttctataag atgcgtgtt ctggagaatt	2280
taacatttac aaccctttta agtcctttta ttaacacggt gttatcgtt tctaacacga	2340
tgtgaatatt atctgtggct agatagtaaa tataatgtga gacgttgta cgtttttagt	2400
cagaataaaa caattcacag tctaaatctt ttgcacttg atcgaatatt tctttaaaaa	2460
[0092]	
tgccaacctg agccatttgtt aaaaccttcc atgtgatacg agggcgctgtt gtttgattt	2520
tctttttat cgttcaatc tggtctgacc tcctgtgtt ttgtgtatga tttatgtcaa	2580
atattaggaa tgtttcact taatagtatt ggtgcgtaa caaagtgcgg tcctgtggc	2640
attctggagg gaaataacaac cgacagatgt atgtaaaggcc aacgtgcata aatcttcata	2700
cagaaagatt tgaagtaata tttaaccgc tagatgaaga gcaagcgtat ggaggcaca	2760
aatgaataaaa gaacaatctg ctgtatgtcc ctccgtggat ctgattctgt taaaaatata	2820
gttttaatgc accatttcata tgagttaccc tggatgttgc attgcgtatgc tagaaacataa	2880
gggtctctg gaagcattca gagcaattga ggcagcgttg gtgaaggcactg ataataatata	2940
gaaggattat tccctgggg ttgactgtac accataactg ctaatcatc aaactattta	3000
gtctgtgaca gagccaaacac gcagtcgtc actgtcagga aagtggtaaa actgcaact	3060

aattactgca atgcctctgt aattaagtga atttacaata tgcgtctgtt cggagggaag	3120
aacgcggat gttcattctt catcacttt aattgtatgtat tatgtctctt ttctgacgt	3180
tagtctccga cggcaggtt caatgaccca ggctgagaaa ttccggacc cttttgctc	3240
aagagcgatg ttaatttgtt caatcatttgc gtttagaaag cggatgtgc gggttgtgt	3300
tctgcgggtt ctgttctcg ttgacatgag gttgccccgtt attcagtgtc gctgatttgt	3360
attgtctgaa gttgtttta cgttaagtttgc atgcagatca attaatacga tacctgcgtc	3420
ataatttgcattt atttgcgtt gtttgcgtt ctccacgcac gttgtgtat gtagatgata	3480
atcattatca ctttacgggtt ctttccgggtt gaaaaaaaaag gtacccaaaaaa aaacatgtc	3540
[0093]	
gtgagtagtg aaccgttaagc	3560

<210> 47

<211> 64

<212> DNA

<213> 人工序列

<220>

<223> 实施例 7 中使用的序列

<400> 47

cgttctgttt atgtttcttg ttgttagcc ttggctaa caaacaagaa acataaacag 60

aacg

64

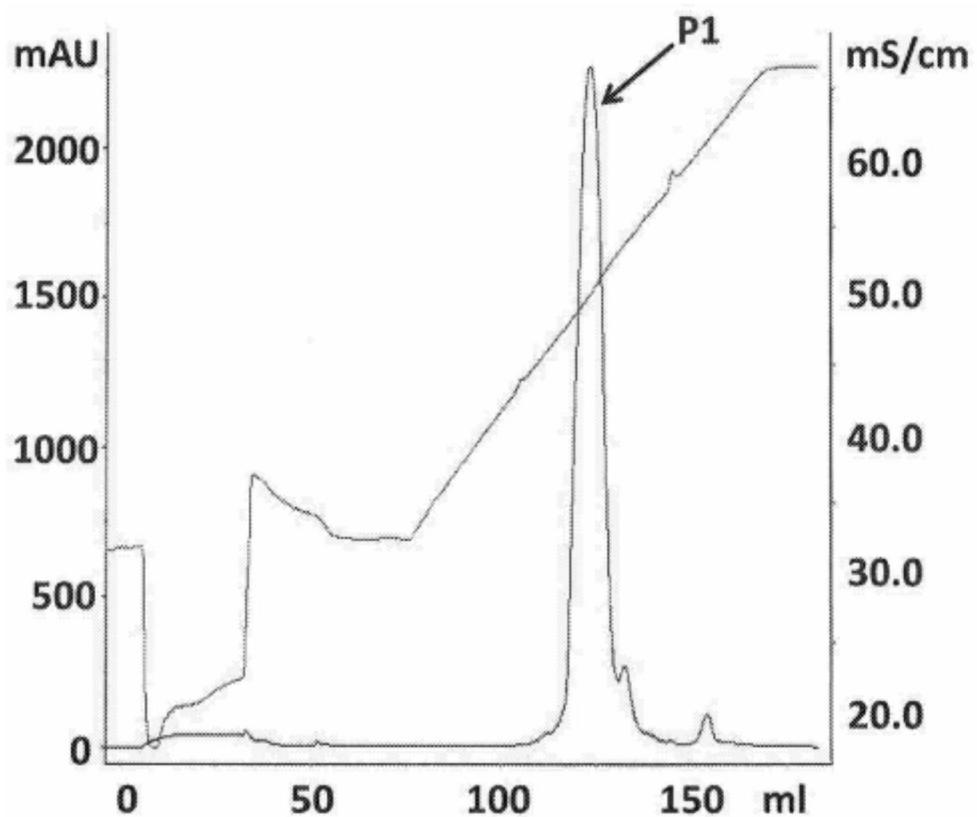


图1

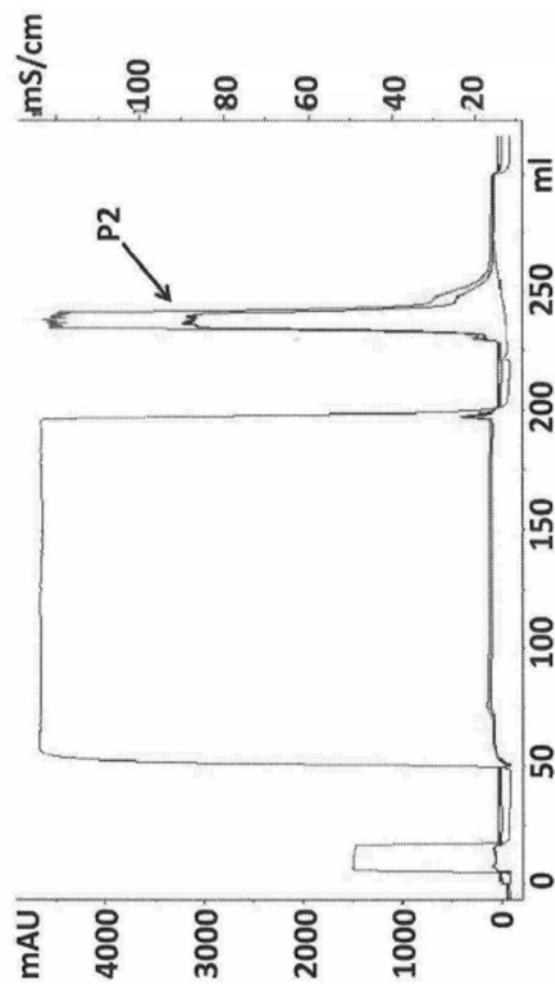


图2

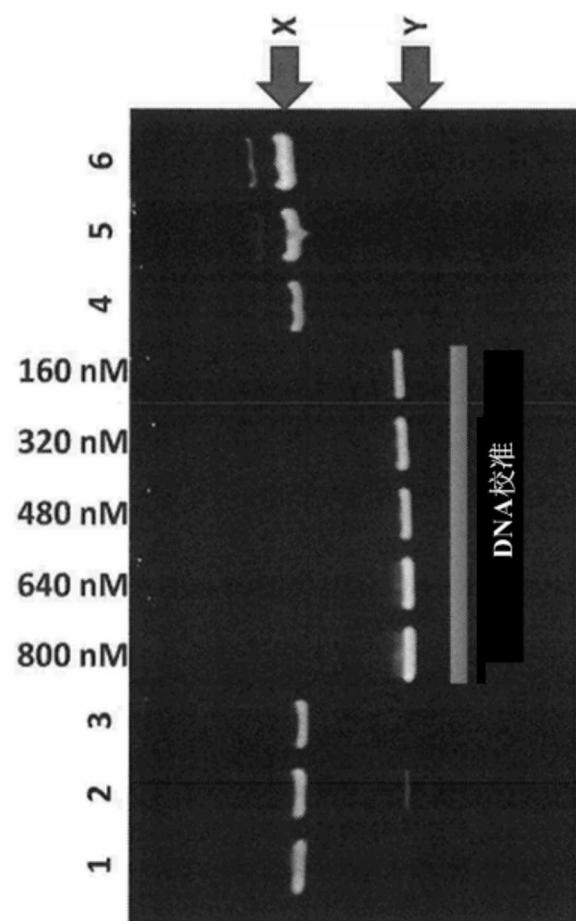


图3

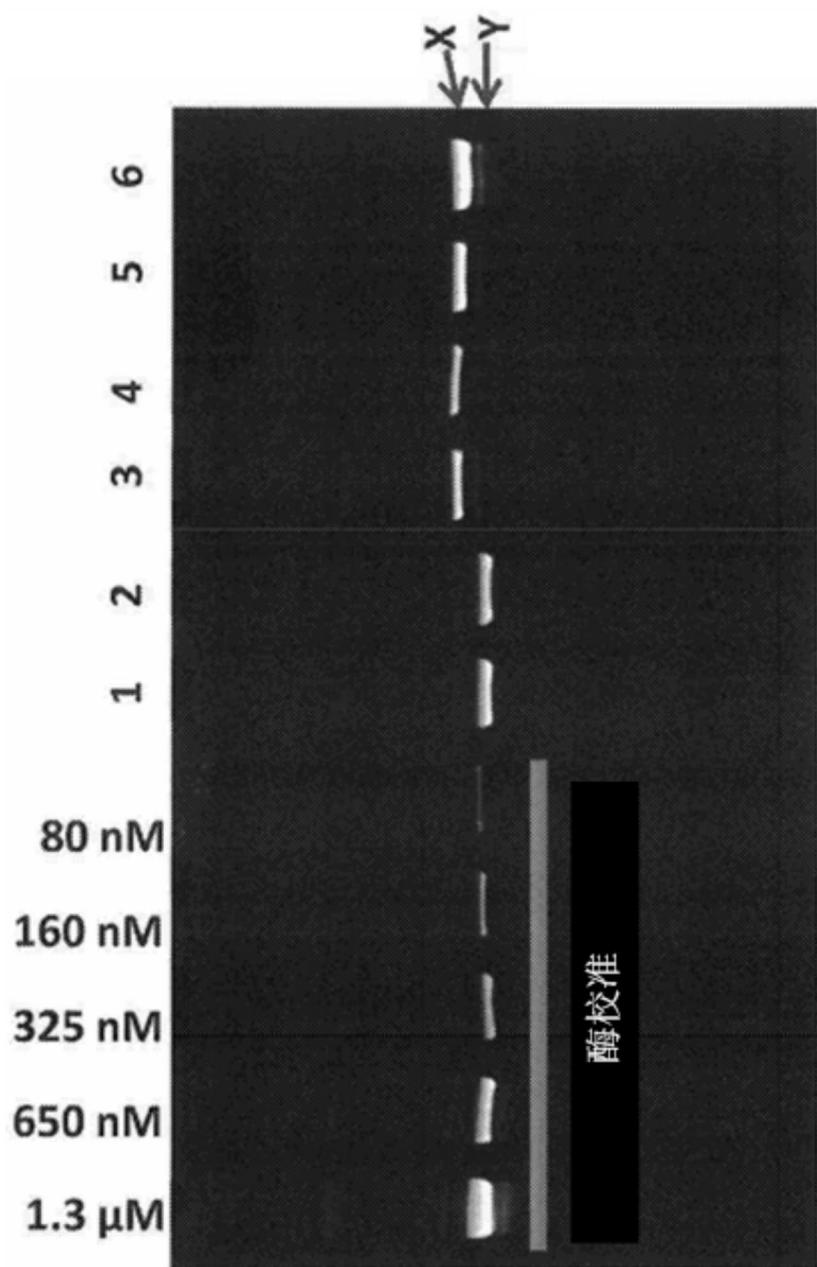


图4

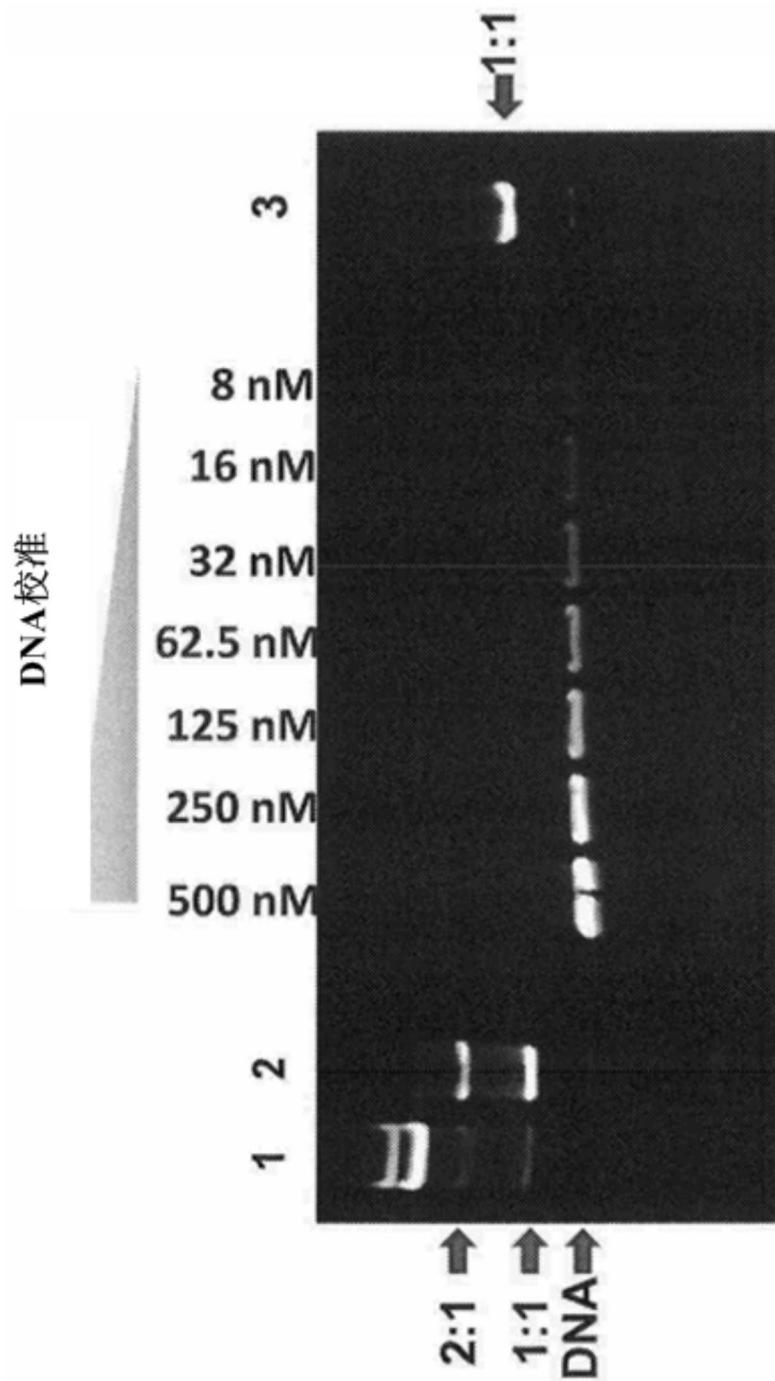


图5

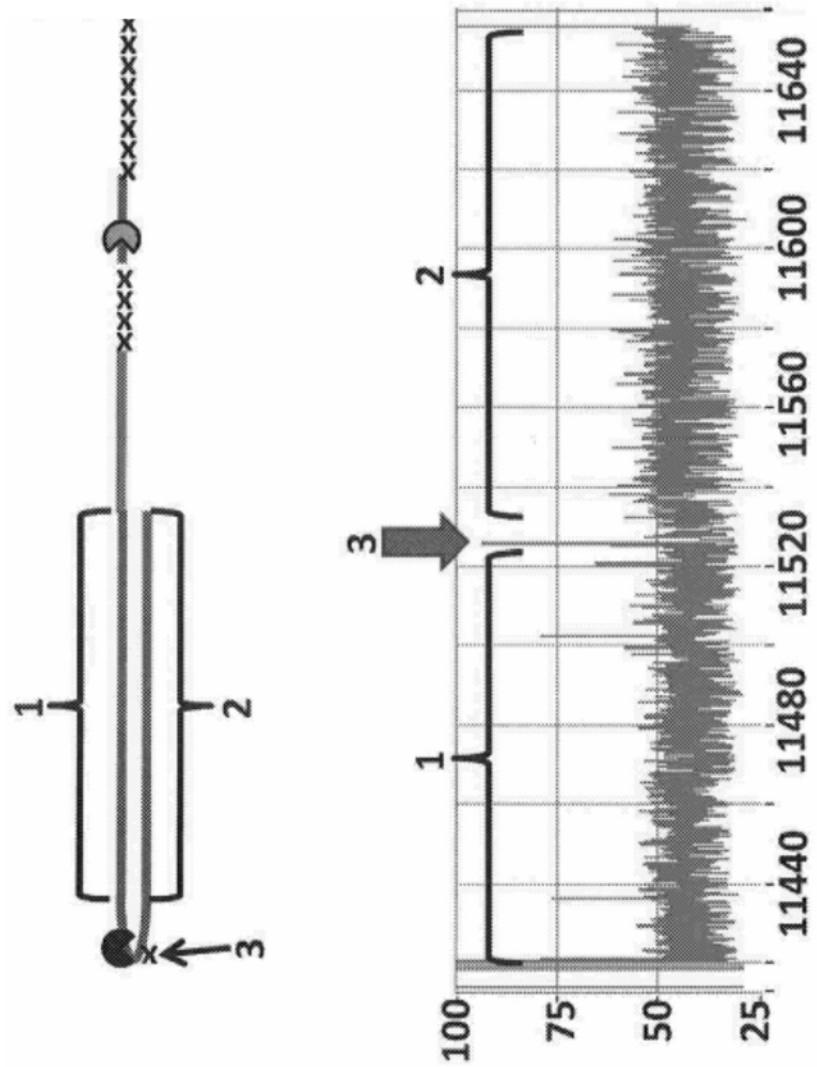


图6

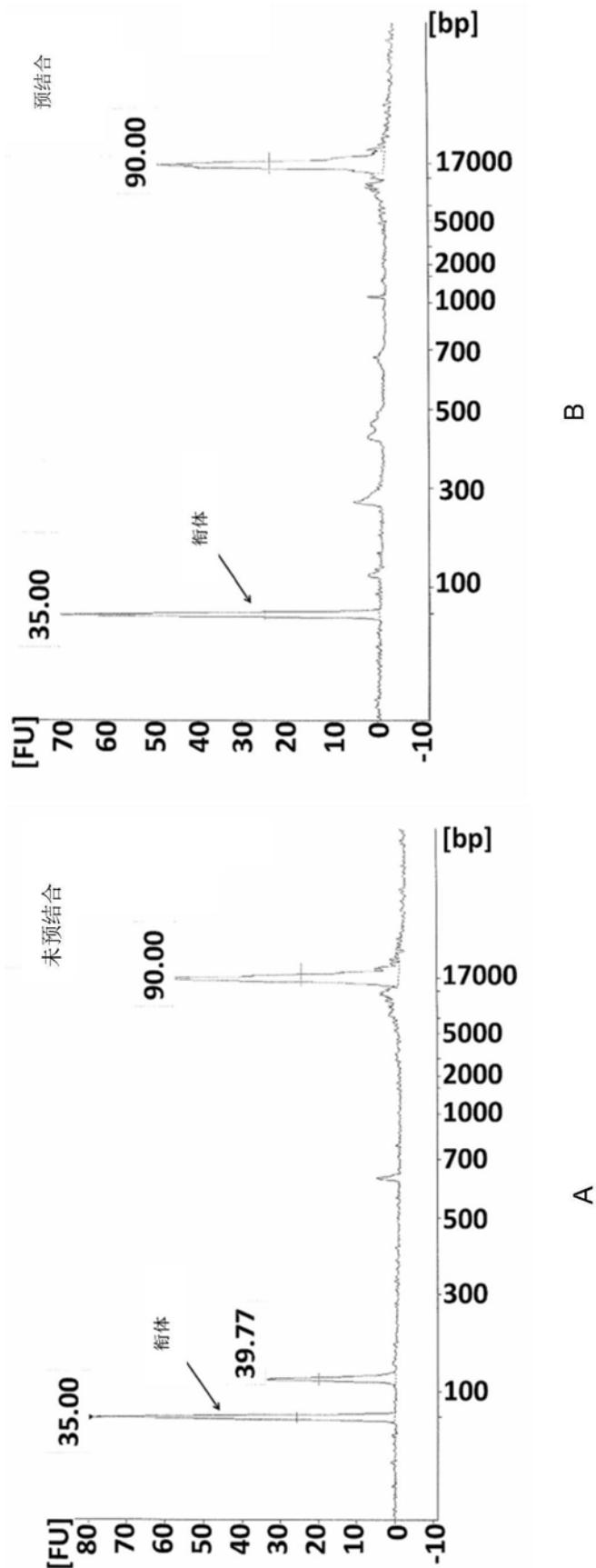


图7

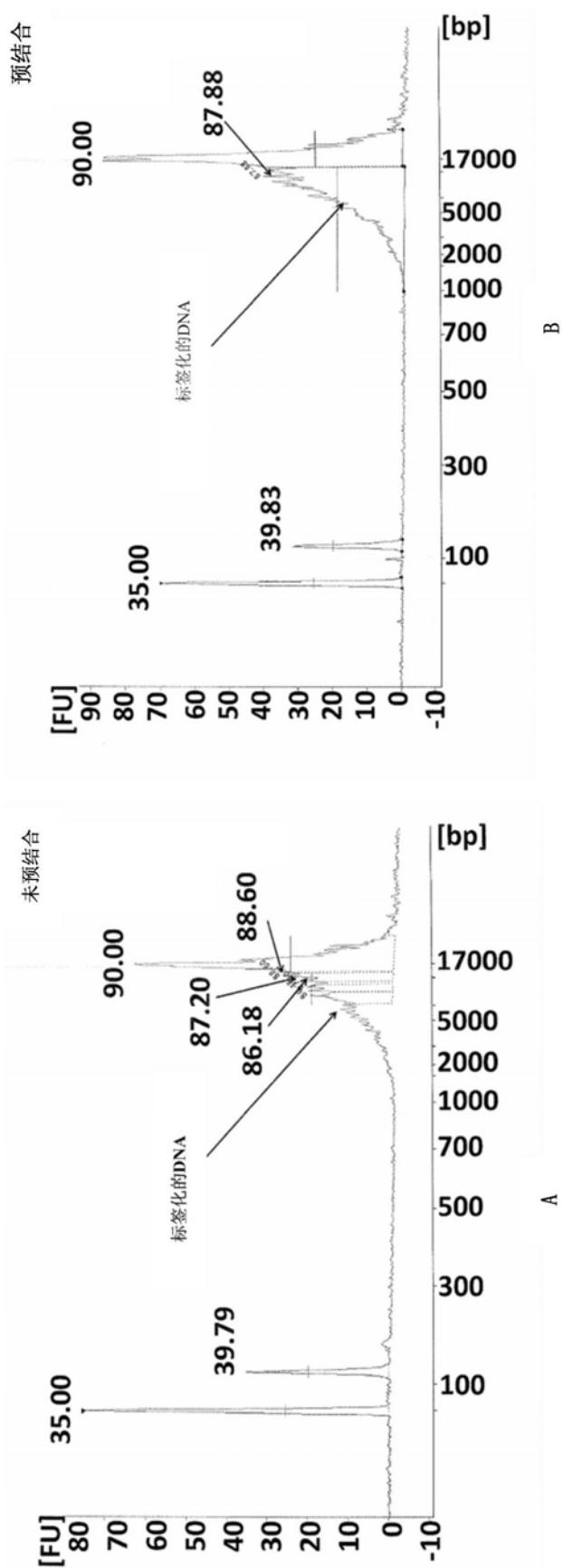


图8

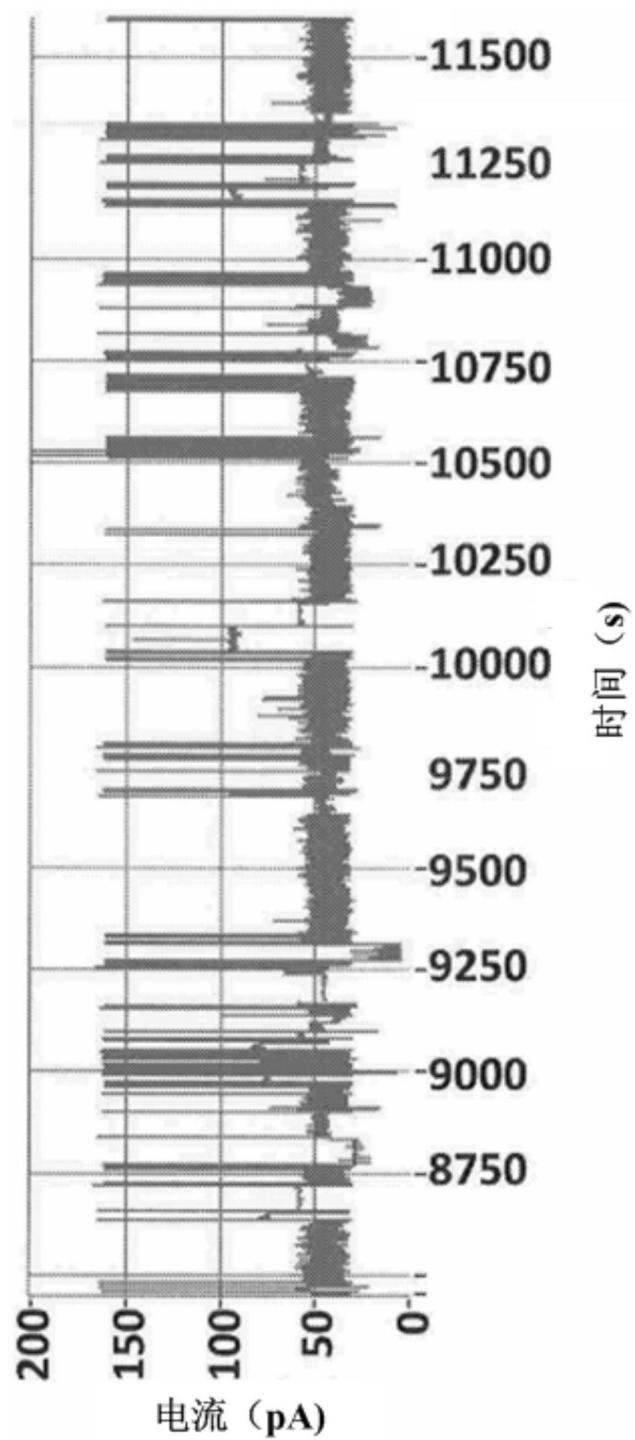


图9

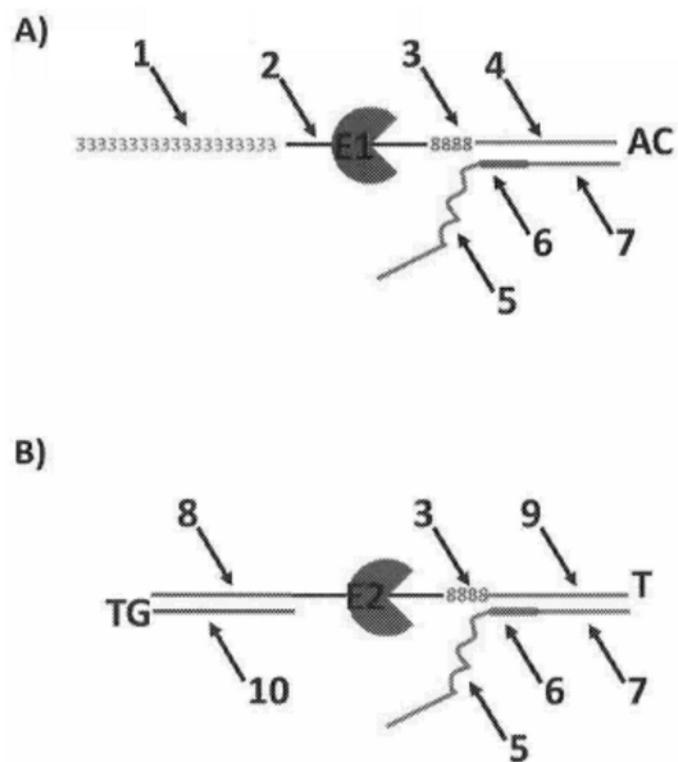


图10

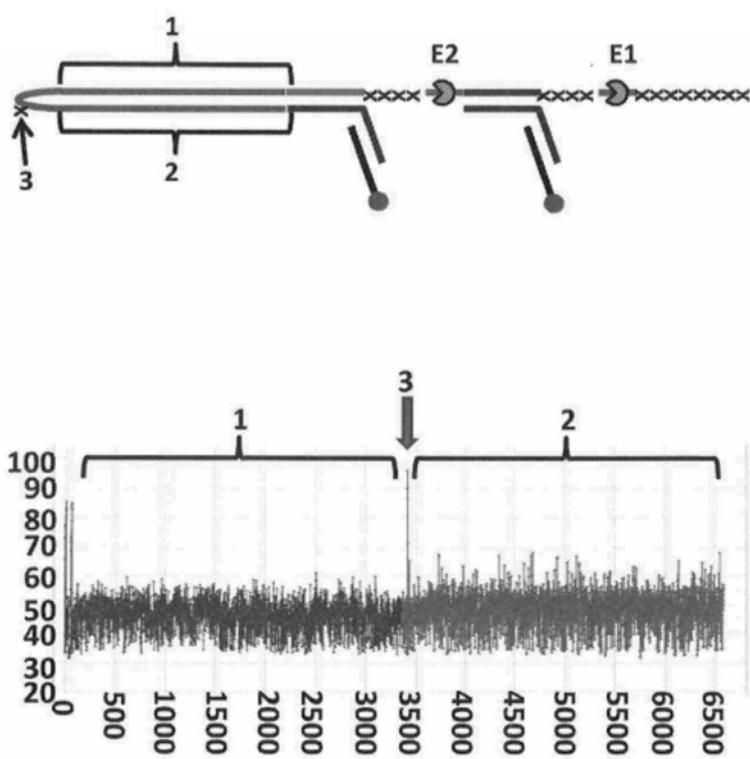


图11

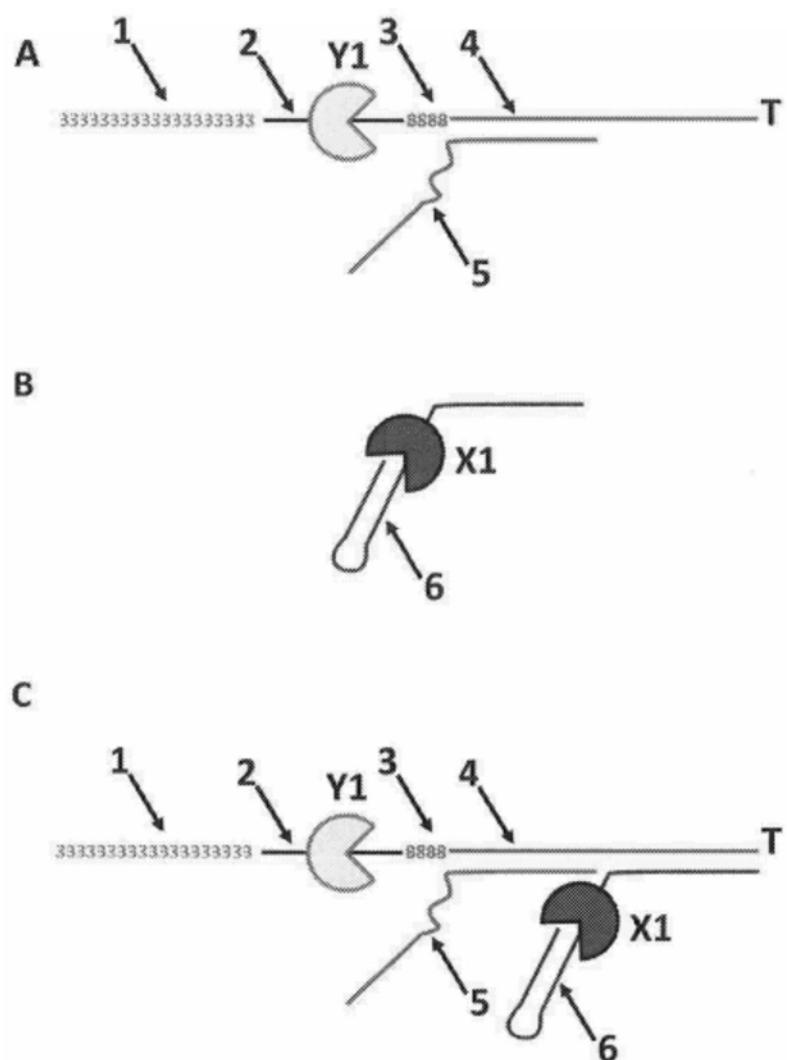


图12

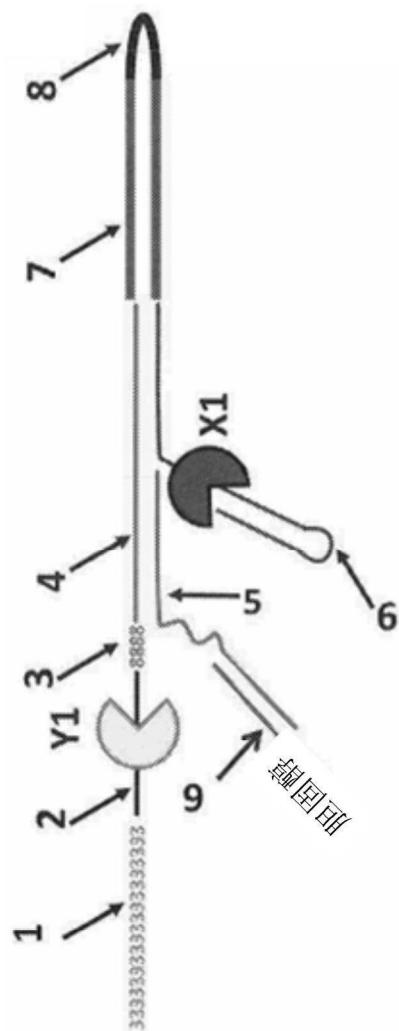


图13

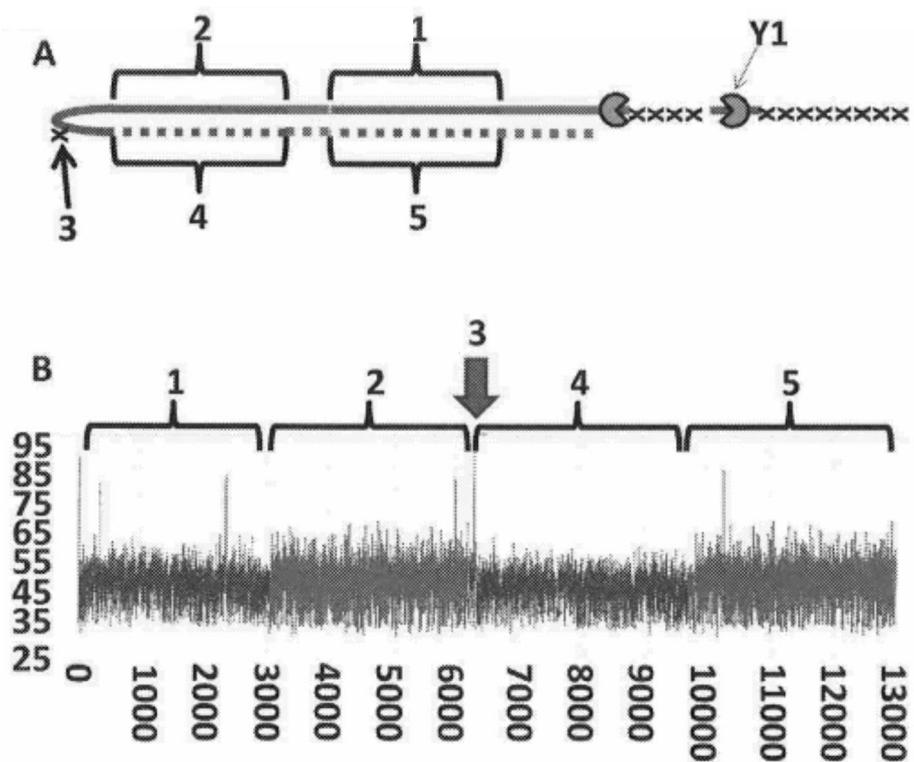


图14

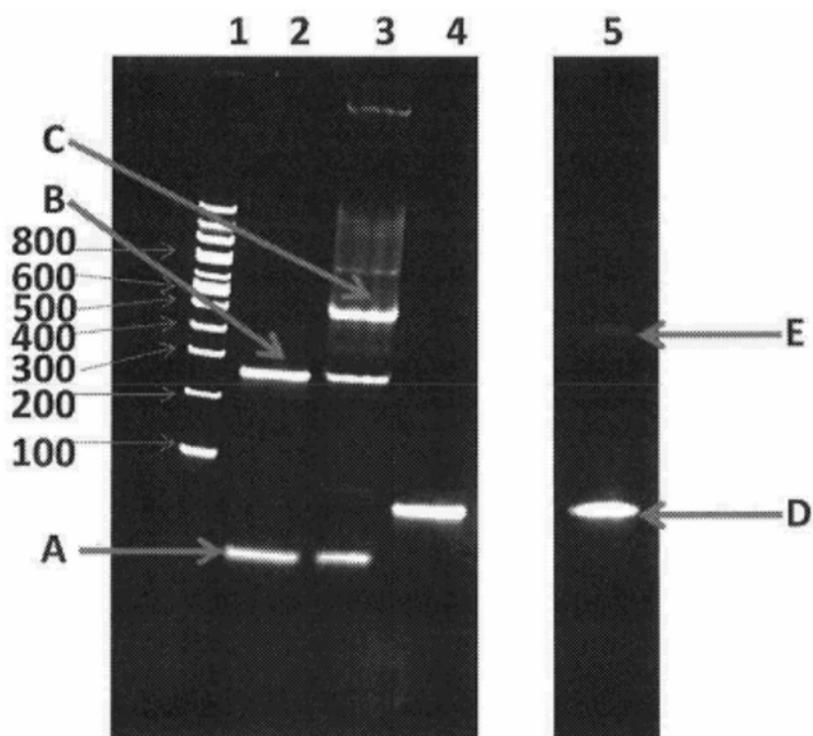


图15

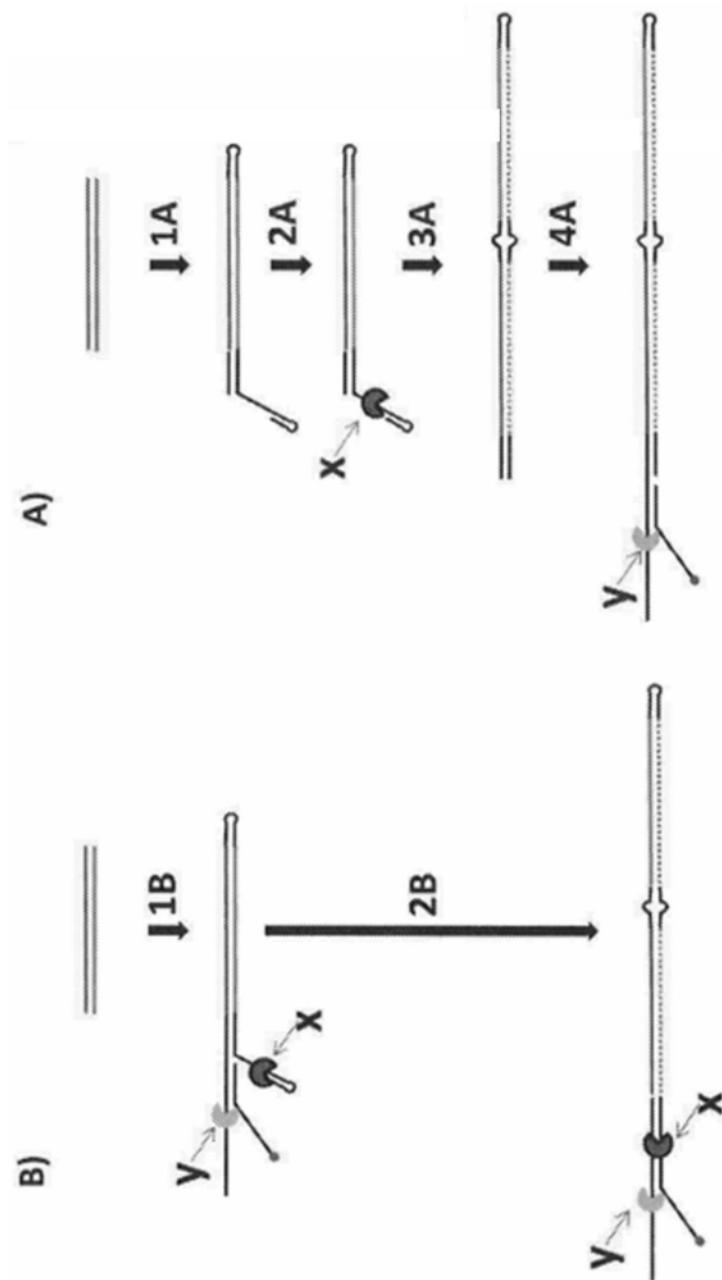


图16

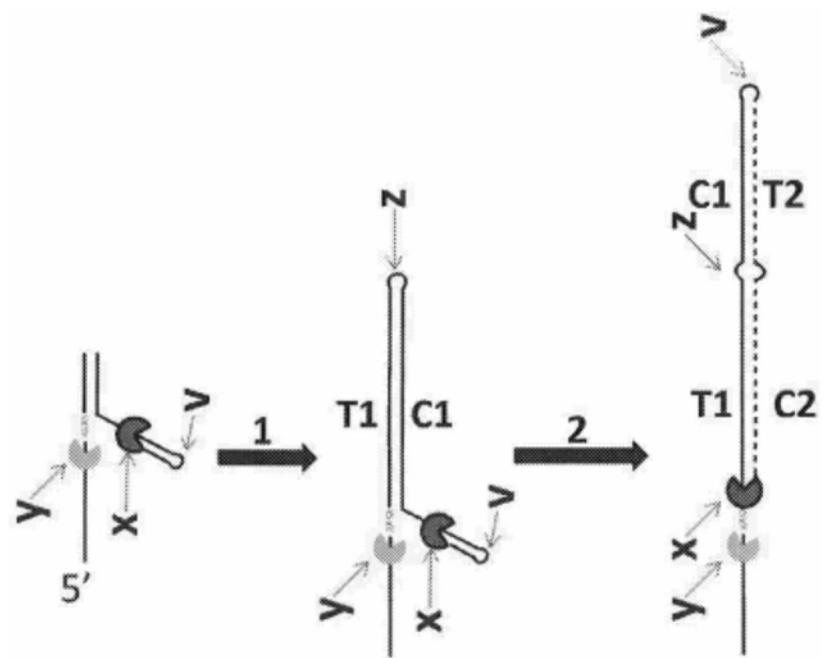


图17

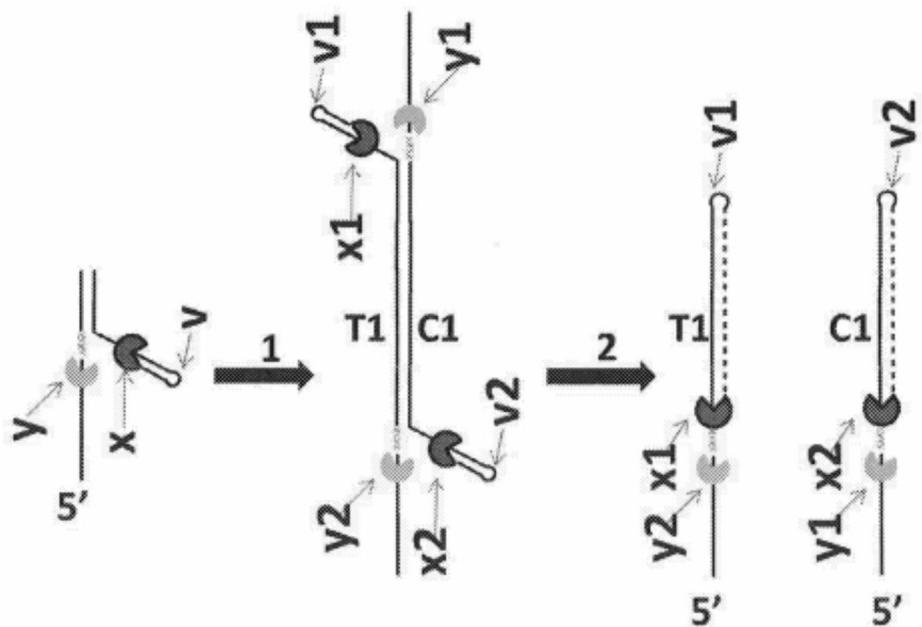


图18

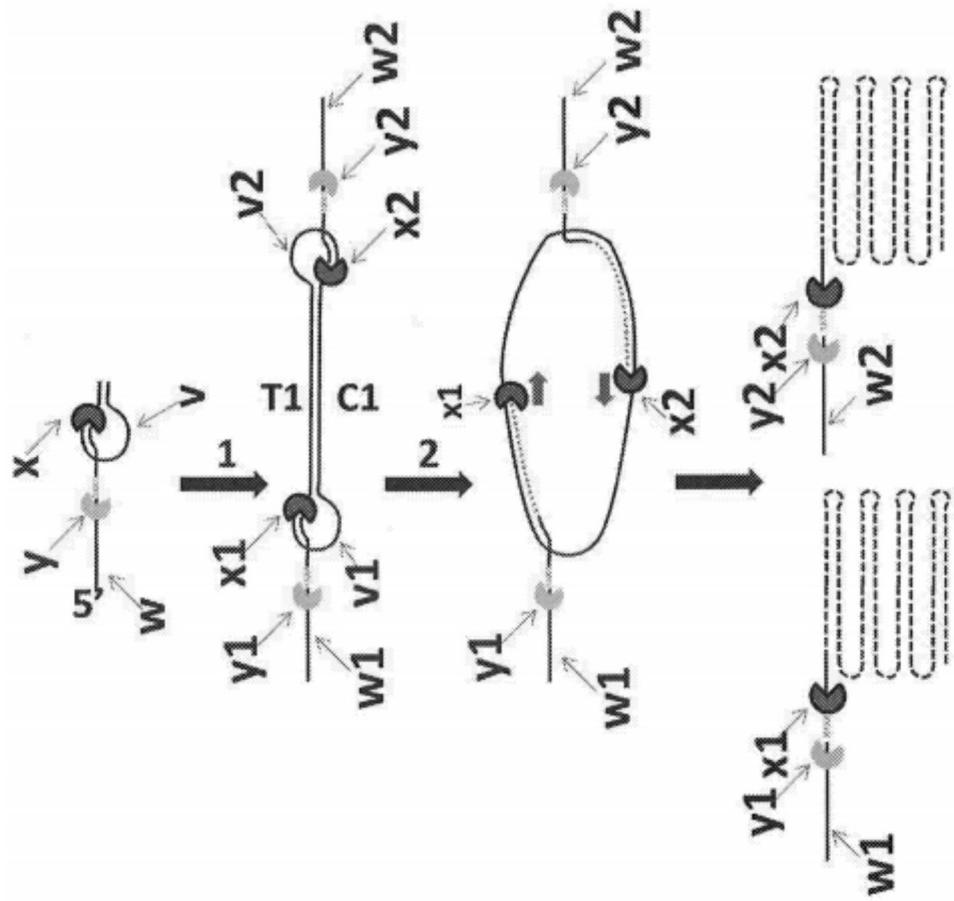


图19