



(12)发明专利申请

(10)申请公布号 CN 106460061 A

(43)申请公布日 2017.02.22

(21)申请号 201580029863.4

詹姆斯·怀特

(22)申请日 2015.03.31

(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

1406147.7 2014.04.04 GB

1407815.8 2014.05.02 GB

PCT/GB2014/052737 2014.09.10 GB

(51)Int.Cl.

C12Q 1/68(2006.01)

C12N 15/115(2010.01)

(85)PCT国际申请进入国家阶段日

2016.12.05

(86)PCT国际申请的申请数据

PCT/GB2015/050991 2015.03.31

(87)PCT国际申请的公布数据

W02015/150786 EN 2015.10.08

(71)申请人 牛津纳米孔技术公司

地址 英国牛津郡

权利要求书4页 说明书43页

序列表73页 附图16页

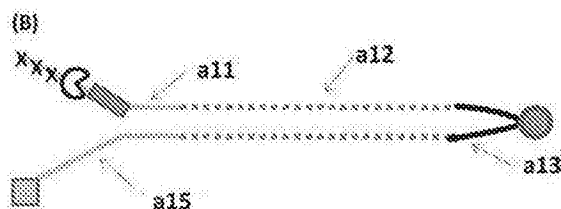
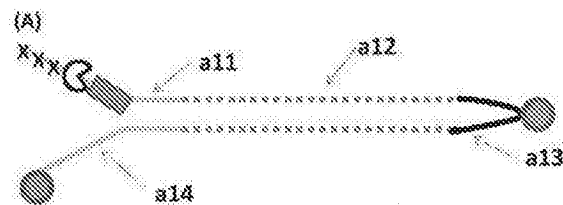
(72)发明人 马里恩·路易丝·克劳福德

(54)发明名称

在双链核酸分子的两端使用纳米孔和锚分子来表征所述双链核酸的方法

(57)摘要

本发明涉及一种用于表征例如测序目标双链多核苷酸的方法。使用具有不同的到膜的耦合强度的至少两个适配体将所述多核苷酸耦合到膜上。



1. 一种使用膜中的跨膜孔表征目标双链多核苷酸的方法,包括:

a) 提供一端具有Y适配体和另一端具有发夹环适配体的所述目标双链多核苷酸,其中所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹环适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,并且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;

b) 将在步骤a)中提供的所述多核苷酸与所述跨膜孔接触,使得所述多核苷酸的至少一条链移动穿过所述孔;和

c) 随着所述多核苷酸的至少一条链相对于所述孔移动,获取一个或多个测量值,其中所述测量值表示所述多核苷酸的至少一条链的一个或多个特征且由此表征所述目标双链多核苷酸。

2. 根据权利要求1所述的方法,其中(a)所述发夹环适配体比所述Y适配体包括更多的锚,(b)将所述一个或多个第二锚耦合到所述膜的强度比将所述一个或多个第一锚耦合到所述膜的强度大,(c)将所述一个或多个第二锚耦合到所述发夹环适配体的强度比将所述一个或多个第一锚耦合到所述Y适配体的强度大,(d)所述一个或多个第一锚和一个或多个第二锚通过杂交将它们各自的适配体耦合到所述膜,且所述一个或多个第二锚中的杂交强度大于所述一个或多个第一锚中的杂交强度,(e)所述一个或多个第一锚包括一个或多个刚性连接体,且所述一个或多个第二锚包括一个或多个柔性连接体,或(f) (a)至(e)的任何组合。

3. 根据权利要求2所述的方法,其中(i)使用胆固醇将所述一个或多个第二锚耦合到所述膜,且使用棕榈酸酯将所述一个或多个第一锚耦合到所述膜,(ii)使用单酰基物质将所述一个或多个第二锚耦合到所述膜,且使用二酰基物质将所述一个或多个第一锚耦合到所述膜,(iii)所述一个或多个第一锚包括一个或多个多核苷酸连接体,且所述一个或多个第二锚包括一个或多个柔性连接体,所述柔性连接体包括一个或多个间隔区9(iSp9)基团或一个或多个间隔区18(iSp18),(iv) (i)和(iii)的组合,或(v) (ii)和(iii)的组合。

4. 根据前述权利要求中任一项所述的方法,其中步骤a)包括,修饰所述目标双链多核苷酸,使得其一端包括Y适配体和另一端包括发夹环适配体。

5. 根据前述权利要求中任一项所述的方法,其中步骤b)包括将在步骤a)中提供的所述多核苷酸与跨膜孔接触,使得所述多核苷酸的两条链移动穿过所述孔,且步骤c)包括随着所述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,其中所述测量值表示所述多核苷酸的链的一个或多个特征,且由此表征所述目标多核苷酸。

6. 根据前述权利要求中任一项所述的方法,其中所述一个或多个特征选自:(i)所述多核苷酸的长度,(ii)所述多核苷酸的同源性,(iii)所述多核苷酸的序列,(iv)所述多核苷酸的二级结构,以及(v)所述多核苷酸是否被修饰。

7. 根据前述权利要求中任一项所述的方法,其中所述多核苷酸的所述一个或多个特征通过电测量和/或光学测量进行测量。

8. 根据权利要求7所述的方法,其中所述电测量是电流测量、阻抗测量、隧道测量或场效应晶体管(FET)测量。

9. 根据前述权利要求中任一项所述的方法,其中步骤b)进一步包括将在步骤a)中提供的所述多核苷酸与多核苷酸结合蛋白接触,使得所述蛋白控制所述多核苷酸的至少一条链

移动穿过所述孔。

10. 根据权利要求9所述的方法,其中所述方法包括:

b) 将在步骤a)中提供的所述多核苷酸与跨膜孔和多核苷酸结合蛋白接触,使得所述多核苷酸的至少一条链移动穿过所述孔,且所述蛋白控制所述多核苷酸的至少一条链移动穿过所述孔;和

c) 随着所述多核苷酸的至少一条链相对于所述孔移动,测量穿过所述孔的电流,其中所述电流表示所述多核苷酸的至少一条链的一个或多个特征,且由此表征所述目标双链多核苷酸。

11. 根据权利要求9或10所述的方法,其中所述多核苷酸结合蛋白衍生自解旋酶。

12. 根据前述权利要求中任一项所述的方法,其中所述膜是两亲性层或固态层。

13. 根据前述权利要求中任一项所述的方法,其中所述跨膜孔是跨膜蛋白孔。

14. 根据权利要求13所述的方法,其中所述跨膜蛋白孔衍生自耻垢分枝杆菌孔蛋白(Msp), $\alpha$ -溶血素( $\alpha$ -HL)或胞溶素。

15. 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括将Y适配体接合到所述多核苷酸的一端,且将发夹环适配体接合到所述多核苷酸的另一端,其中所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹环适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大,且由此提供经修饰的目标双链多核苷酸。

16. 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括

(a) 将Y适配体接合到所述多核苷酸的一端,且将发夹环适配体接合到所述多核苷酸的另一端;和

(b) 向所述Y适配体上连接用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,向所述发夹环适配体上连接用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且由此提供经修饰的目标双链多核苷酸;

其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大。

17. 根据权利要求15或16所述的方法,其中所述适配体和/或锚如权利要求2或3中所定义。

18. 一种目标双链多核苷酸,使用权利要求15至17中任一项所述的方法进行修饰。

19. 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括将所述目标多核苷酸与MuA转座酶和一组双链MuA底物接触,其中所述组中底物的一部分是含有将所述多核苷酸耦合到所述膜的一个或多个第一锚的Y适配体,其中所述组中底物的一部分是含有将所述多核苷酸耦合到所述膜的一个或多个第二锚的发夹环适配体,并且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大,且由此产生多个经修饰的双链多核苷酸。

20. 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括:

(a) 将所述目标多核苷酸与MuA转座酶和一组双链MuA底物接触,其中所述组中底物的一部分是Y适配体,且其中所述组中底物的一部分是发夹环适配体,使得所述转座酶将所述

目标多核苷酸片段化,且将底物接合到所述双链片段的一端或两端,且由此产生多个片段/底物构建体;和

(b) 向所述多个片段/底物构建体中的所述Y适配体上连接一个或多个第一锚,且向所述多个片段/底物构建体中的发夹环适配体上连接一个或多个第二锚,且由此产生多个修饰的双链多核苷酸;

其中将每个发夹环适配体耦合到所述膜的强度比将每个Y适配体耦合到所述膜的强度大。

21. 根据权利要求19或20所述的方法,其中所述组中的每个底物包括通用核苷酸的至少一个突出端,使得所述转座酶将所述目标多核苷酸片段化,且将底物接合到所述双链片段的一端或两端,并由此产生多个片段/底物构建体,且其中所述方法进一步包括将所述突出端接合到所述构建体中的片段,并由此产生多个修饰的双链多核苷酸。

22. 权利要求19或20所述的方法,其中所述组中的每个底物包括(i)至少一个突出端,和(ii)与所述至少一个突出端在同一链中的至少一个核苷酸,其包括不存在于所述目标多核苷酸中的核苷,使得所述转座酶将所述目标多核苷酸片段化,且将底物接合到所述双链片段的一端或两端,且由此产生多个片段/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从构建体去除突出端,且由此产生包含单链空位的多个双链构建体;和

(b) 修复所述构建体中的所述单链空位,并由此产生多个修饰的双链多核苷酸。

23. 使用根据权利要求19至22中任一项所述的方法修饰的多个多核苷酸。

24. 一种使用膜中的跨膜孔表征目标双链多核苷酸的方法,包括:

a) 提供在一端具有Y适配体和另一端具有发夹环适配体的所述目标双链多核苷酸,其中所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹环适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;

b) 将在步骤a)中提供的所述多核苷酸与所述跨膜孔、聚合酶和标记的核苷酸接触,使得当通过所述聚合酶将核苷酸加入到所述多核苷酸时,磷酸盐标记的物质依次释放,其中所述磷酸盐物质含有对每个核苷酸具有特异性的标记;和

c) 使用所述孔检测所述磷酸盐标记的物质,且由此表征所述目标双链多核苷酸。

25. 一个适配体对,用于修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征,其中一个适配体是Y适配体,所述Y适配体包含用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中另一个适配体是发夹环适配体,所述发夹环适配体包含用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,并且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大。

26. 一个适配体组,用于修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征,其中所述适配体中的一部分是Y适配体,其包含用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述适配体中的一部分是发夹环适配体,其包含用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大。

27. 根据权利要求25所述的适配体对,或根据权利要求26所述的适配体组,其中每个适

配体包括双链MuA底物。

28. 根据权利要求27所述的适配体对或适配体组,其中每个适配体如权利要求21或22中所定义。

29. 一个用于修饰目标多核苷酸的试剂盒,包括 (a) 根据权利要求28所述的适配体对或适配体组,和 (b) MuA转座酶。

## 在双链核酸分子的两端使用纳米孔和锚分子来表征所述双链核酸的方法

### 技术领域

[0001] 本发明涉及一种用于表征如测序目标双链多核苷酸的方法。使用具有耦合到所述膜的不同强度的至少两个适配体(adaptor),将所述多核苷酸耦合到膜上。

### 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。

[0005] 以前已经证明,通过将所述分析物耦合到其中存在相关检测器的膜,能够实现超低浓度分析物递送。这使检测需要的分析物的量降低了若干数量级(WO 2012/164270)。

[0006] 还已经表明,如果双链多核苷酸被修饰为包括含有前导序列和发夹适配体的Y适配体(双链茎和两个非互补臂),还可以使用链测序有效地表征双链多核苷酸(WO 2013/014451)。优选的为,含有前导序列的Y适配体被连接到所述多核苷酸的一端,且发夹适配体被连接到另一端。所述前导序列优选旋入纳米孔,且连接所述多核苷酸的两条链的发夹环使得待作为所述多核苷酸研究的两条链解开且移动穿过所述孔。这是有利的,因为它将从单个双链多核苷酸获得的信息量加倍。此外,因为两条链中的序列是互补的,因此从这两条链中获得的信息可以信息化组合。这种机制提供了正交校验功能,从而提供更高可信度的观察值。当Y适配体和发夹环以这种方式一起使用时,Y适配体通常包含将所述多核苷酸耦合到包含纳米孔的膜的锚。在一些情况下,在两端具有Y适配体的双链多核苷酸在样品制备中产生。所述前导序列和和Y适配体中的锚的存在意味着系统通常偏向于表征这些多核苷酸。然而,连接这种多核苷酸中的两条链的发夹环的缺乏意味着只有一条链被研究。

### 发明内容

[0007] 本发明人已令人惊讶地证明,通过在系统中包括两个锚,可以避免偏向在两端包含Y适配体的双链多核苷酸。本发明人已经使用包括用于将多核苷酸耦合到所述膜的一个或多个第一锚的Y适配体和包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚的发夹适配体。发夹适配体以比Y适配体更大的强度将所述多核苷酸耦合到所述膜。

[0008] 因此,本发明提供了一种使用膜中的跨膜孔表征目标双链多核苷酸的方法,包括:

[0009] a) 提供在一端具有Y适配体并在另一端具有发夹适配体所述目标双链多核苷酸,其中所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,并且其中将所述发夹适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;

[0010] b) 将在步骤a)中提供的所述多核苷酸与所述跨膜孔接触,使得所述多核苷酸的至少一条链穿过所述孔;和

[0011] c) 随着所述多核苷酸的至少一条链相对于所述孔移动,获取一个或多个测量值,其中所述测量值表示所述多核苷酸的至少一条链的一个或多个特性且由此表征所述目标双链多核苷酸。

[0012] 本发明还提供:

[0013] 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括将Y适配体接合到所述多核苷酸的一端,且将发夹适配体连接到所述多核苷酸的另一端,其中所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且其中将所述发夹适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大,且由此提供修饰的目标双链多核苷酸;

[0014] 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括(a)将Y适配体接合到多核苷酸的一端,且将发夹适配体接合到所述多核苷酸的另一端;和(b)将用于耦合所述多核苷酸到所述膜的一个或多个第一锚连接到所述Y适配体,将用于耦合所述多核苷酸到所述膜的一个或多个第二锚连接到所述发夹适配体,且由此提供修饰的目标双链多核苷酸;其中将所述发夹适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;

[0015] 一种使用本发明的方法修饰的目标双链多核苷酸;

[0016] 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括使所述目标多核苷酸与MuA转座酶以及一组双链MuA底物接触,其中所述组中底物的一部分是含有用于将所述多核苷酸与所述膜耦合的一个或多个第一锚的Y适配体,其中所述组中底物的一部分是包括将所述多核苷酸耦合到所述膜的一个或多个第二锚的发夹适配体,并且其中所述发夹适配体耦合到所述膜的强度比所述Y适配体耦合到所述膜的强度大,且由此产生多个修饰的双链多核苷酸;

[0017] 一种修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括:(a)将所述目标多核苷酸与MuA转座酶和一组双链MuA底物接触,其中所述组中底物的一部分是Y适配体,且其中所述组中底物的一部分是发夹适配体,使得所述转座酶将所述目标多核苷酸片段化,且将所述底物接合到所述双链片段的一端或两端,且由此产生多个片段/底物构建体;和(b)向多个片段/底物构建体中的所述Y适配体连接一个或多个第一锚,且向所述多个片段/底物构建体中的发夹适配体连接一个或多个第二锚,且由此产生多个修饰的双链多核苷酸;其中每个发夹适配体耦合到所述膜的强度比每个Y适配体耦合到所述膜的强度大;

[0018] 一种使用本发明方法修饰的多个多核苷酸;

[0019] 一种使用膜中的跨膜孔表征目标双链多核苷酸的方法,包括:

[0020] a) 提供在一端具有Y适配体且在另一端具有发夹适配体的所述目标双链多核苷酸,其中所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且其中所述发夹适配体耦合到所述膜的强度比所述Y适配体耦合到所述膜的强度大;

[0021] b) 使在步骤a)中提供的所述多核苷酸与所述跨膜孔、聚合酶和标记的核苷酸接触,使得当通过聚合酶将核苷酸添加到所述多核苷酸时,磷酸盐标记的物质(species)依次释放,其中所述磷酸盐物质含有对每个核苷酸具有特异性的标记;和

[0022] c) 使用所述孔检测所述磷酸盐标记的物种,且由此表征所述目标双链多核苷酸;

[0023] 一种用于修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的适配体对,其中一个适配体组是包含用于将所述多核苷酸耦合到所述膜的一个或多个第一锚的Y适配体,其中另一个适配体是包含用于将所述多核苷酸耦合到所述膜的一个或多个第二锚的发夹适配体,并且其中将所述发夹适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;

[0024] 一组用于修饰目标多核苷酸以使用膜中的跨膜孔进行表征的适配体,其中所述适配体的一部分是用于将所述多核苷酸耦合到所述膜包含一个或多个第一锚的Y适配体,其中所述适配体的一部分是包含用于将所述多核苷酸耦合到所述膜的一个或多个第二锚的发夹适配体,且其中将所述发夹适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;以及

[0025] 一种用于修饰目标多核苷酸的试剂盒,包括(a)本发明的一对适配体或一组适配体,和(b)MuA转座酶。

## 附图说明

[0026] 图1的(A)部分和(B)部分示出了在实施例2中使用的DNA构建体的卡通示意图——四个iSpC3间隔区示为十字,且4个5-硝基吡啶为黑色0,胆固醇耦合剂为灰色椭圆,且棕榈酸酯耦合剂为灰色正方形;标签a1-10在实施例1中全面描述。

[0027] 图2的(A)部分和(B)部分示出了在实施例3中使用的DNA构建体的卡通示意图——四个iSpC3间隔区示为十字,且4个5-硝基吡啶为黑色0,胆固醇耦合剂为灰色椭圆,且棕榈酸酯锚为灰色方形;标签a1-9在实施例1中全面描述。

[0028] 图3示出了针对构建体1检测的解旋酶控制的DNA移动的比例,其中链a1移位穿过纳米孔和构建体2,其中链a2移位穿过纳米孔。x轴示出了沿 $\lambda$ DNA序列,解旋酶控制的DNA移动所映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$ 区域的模板和互补模板对应的链a1和a2相同的区域)。Y轴示出了链a1(构建体1)解旋酶控制的DNA移动(标记为a1)与链a2(构建体2)解旋酶控制的DNA移动(标记为a2)的比例。当使用单个胆固醇将构建体耦合到所述膜上时,观察到该比例为大约50:50。本对照显示,如果使用相同的耦合剂将不同的构建体耦合到所述膜上,则解旋酶控制的DNA移动的数目对每个构建体而言大概相等,例如没有观察到偏向。

[0029] 图4示出了针对其中链a2移位穿过纳米孔的构建体2和其中链a1移动穿过纳米孔的构建体3检测的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA序列,解旋酶控制的DNA移动映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$ 区域的模板和互补模板对

应的链a1和a2相同的区域)。Y轴示出了链a2(构建体2)解旋酶控制的DNA移动(标记为a2)与链a1(构建体3)解旋酶控制的DNA移动(标记为a1)的比例。当使用单个胆固醇将构建体2耦合到所述膜上时,与使用胆固醇和棕榈酸酯将构建体3耦合到所述膜上时比较,观察到该比例为大约5:95。该实验表明,与单个耦合剂相比(胆固醇),当使用两种耦合剂将构建体耦合到所述膜(其中一个耦合剂比另一个耦合剂强,如胆固醇比棕榈酸酯强)时,解旋酶控制的DNA移动的数目比单耦合的构建体更强地偏向双重耦合的构建体。

[0030] 图5示出了针对其中链a1移位穿过纳米孔的构建体4和其中链a1或a2移动穿过纳米孔的构建体5检测的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA序列,解旋酶控制的DNA移动映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$ 区域的模板和互补模板对应的链a1和a2相同的区域)。Y轴示出了链a1(构建体4或5)解旋酶控制的DNA移动(标记为a1)与链a2(仅构建体5)解旋酶控制的DNA移动(标记为a2)的比例。当使用两个胆固醇将这两个构建体耦合到所述膜上时,观察到该比例为约33:66(a2:a1解旋酶控制的DNA移动)。由于如果构建体4被纳米孔捕获,仅能检测到a2移动,因此可以预测到2:1偏向a1解旋酶控制的DNA移动,而a1移动将从构建体4和5检测到,因此可以预计到2倍的a1移动。

[0031] 图6示出了针对其中链a1移位穿过纳米孔的构建体3和其中链a1或a2移动穿过纳米孔的构建体6检测的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA序列,解旋酶控制的DNA移动映射的位置(它们全部与模板对应的链a1和a2集中在相同的区域,且 $\lambda$ 的区域模板互补在45,042bp和48,487bp之间)。Y轴示出了链a1(构建体3或6)解旋酶控制的DNA移动(标记为a1)与链a2(仅构建体6)解旋酶控制的DNA移动(标记为a2)的比例。当使用胆固醇和棕榈酸酯的组合将构建体3耦合到所述膜上和使用两种棕榈酸酯将构建体6耦合到所述膜上时,观察到该比例为约5:95(a2:a1解旋酶控制的DNA移动)。与图5中示出的对照实验相比,偏向解旋酶控制的DNA移动表明,使用不同强度的两种不同耦合剂的组合时,可以选择这种构建体,而不是含有两种类似强度的耦合剂的构建体。

[0032] 图7的(A)部分和(B)部分示出了中在实施例4中使用的DNA构建体的卡通示意图——四个iSpC3间隔区示为十字,且4个5-硝基吡啶为黑色0,胆固醇耦合剂为灰色椭圆,且在本实施例中研究的另外的耦合剂为黑色三角形;标签a1-10(a-c)在实施例1和4中全面描述。

[0033] 图8示出了针对其中链a1移位穿过纳米孔的构建体1(使用单个胆固醇耦合到所述膜)和其中链a2移动穿过纳米孔的构建体2(也使用单个胆固醇耦合到所述膜)检测的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA,序列解旋酶控制的DNA移动映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$ 区域的模板和互补模板对应的链a1和a2相同的区域)。Y轴示出了链a1(构建体1)解旋酶控制的DNA移动(标记为a1)和链a2(构建体2)解旋酶控制的DNA移动(标记为a2)的比例。当使用单个胆固醇将每个构建体耦合到所述膜上时,观察到该比例为大约50:50。该对照表明,如果使用相同的耦合剂将不同的构建体耦合到所述膜上,解旋酶控制的DNA移动的数目对每个构建体大约相等,例如没有观察到偏向。

[0034] 图9示出了针对其中链a1移位穿过纳米孔的构建体1(使用单个胆固醇耦合到所述膜)与针对其中链a2移动穿过纳米孔的构建体7a(使用连接有两个胆固醇TEG's的DNA片段耦合到所述膜,每个以黑色三角形显示)检测的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA序列解旋酶控制的DNA移动映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$

区域的模板和互补模板对应的链a1和a2相同的区域)。Y轴示出了链a1(构建体1)解旋酶控制的DNA移动(标记为a1)与链a2(构建体7a)解旋酶控制的DNA移动(标记为a2)的比例。当使用单个胆固醇将构建体1耦合到所述膜上时,与使用两个胆固醇将构建体7a耦合到所述膜上时相比,观察到该比例为大约50:50。该实验表明,当使用连接到相同的DNA片段的两个胆固醇将构建体7a耦合到所述膜上时,与使用单个胆固醇时相比,检测到的解旋酶控制的DNA移动的数目显示,没有偏向任何一个构建体。因此,在相同的DNA片段中两个胆固醇时观察到的耦合强度与单个胆固醇时相似。

[0035] 图10示出了针对其中链a1移位穿过纳米孔的构建体1(使用单个胆固醇耦合到所述膜)和针对其中链a2移动穿过纳米孔的构建体7b(使用生育酚耦合到所述膜,示为黑色三角形)检测到的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA序列,解旋酶控制的DNA移动映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$ 区域的模板和互补模板对应的链a1和a2相同的区域)。Y轴示出了链a1(构建体1)解旋酶控制的DNA移动(标记为a1)与链a2(构建体7b)解旋酶控制的DNA移动(标记为a2)的比例。当使用单个胆固醇将构建体1耦合到所述膜上时,与使用单个生育酚将构建体7b耦合到所述膜上时相比,观察到该比例为大约35:65。该实验表明,当使用生育酚将构建体7b耦合到所述膜上时,与使用单个胆固醇时相比,检测到的解旋酶控制的DNA移动的数目显示偏向于使用生育酚耦合的构建体。因此,生育酚时所观察到的耦合强度比胆固醇时观察到的耦合强度略强。

[0036] 图11示出了针对其中链a1移位穿过纳米孔的构建体1(使用单个胆固醇耦合到所述膜)和针对中链a2移动穿过纳米孔的构建体7c(使用棕榈酸酯耦合到所述膜,示为黑色三角形)检测的解旋酶控制的DNA移动的比例。x轴示出了沿 $\lambda$ DNA序列,解旋酶控制的DNA移动映射的位置(它们全部集中在与45,042bp和48,487bp之间 $\lambda$ 区域的模板和互补模板对应的链a1和a2相同的区域)。Y轴示出了链a1(构建体1)解旋酶控制的DNA移动(标记为a1)与链a2(构建体7c)解旋酶控制的DNA移动(标记为a2)的比例。当使用单个胆固醇将构建体1耦合到所述膜上时,与使用单个棕榈酸酯将构建体7c耦合到所述膜上时相比,观察到该比例为大约70:30。该实验表明,当使用棕榈酸酯将构建体7c耦合到所述膜上时,与使用单个胆固醇时相比,检测到的解旋酶控制的DNA移动的数目显示偏向于使用胆固醇时耦合的构建体。因此,棕榈酸酯时所观察到的耦合强度弱于胆固醇时观察到的耦合强度。

[0037] 图12的(A)部分和(B)部分示出了在实施例6中使用的DNA构建体的卡通示意图——25个iSpC3间隔区表示为三个黑色十字,4个iSp18'表示为灰色矩形,胆固醇耦合剂表示为灰色圆,且棕榈酸酯耦合剂表示为灰色方块;标签a11-15在实施例5中全面描述。

[0038] 图13示出片段化和连接适配体后产生的可能的构建体(如实施例5中描述的)。构建体10具有两个发夹适配体且不能被纳米孔捕获。构建体8是所需的构建体,其具有1个Y适配体和1个发夹,链X和链Y在被捕获后移位穿过纳米孔。构建体11具有两个Y适配体,因此,仅可以捕获和移位链X或链Y。

[0039] 图14示出片段化和连接适配体后产生的可能的构建体(如实施例5中描述的)。构建体10具有两个发夹适配体且不能被纳米孔捕获。构建体9是所需的构建体,其具有1个Y适配体和1个发夹环,链X和链Y被捕获后移位穿过纳米孔。构建体12具有两个Y适配体,因此,仅可以捕获和移位链X或链Y。

[0040] 图15的(A)部分和(B)部分示出了在实施例7中使用的两个DNA构建体(构建体13和

14)的卡通示意图——30个iSpC3间隔区表示为4个黑色十字,4个iSp18'表示为灰色矩形,胆固醇耦合剂表示为灰色圆,且棕榈酸酯耦合剂表示为灰色方块;标签a12,a16-21在实施例7中全面描述。DNA构建体的两个区域标记为R1和R2。

[0041] 图16的(A)部分和(B)部分示出了在实施例7中使用的两个DNA构建体(构建体15和16)的卡通示意图——30个iSpC3间隔区表示为4个黑色十字,4个iSp18'表示为灰色矩形,胆固醇耦合剂表示为灰色圆形,标记a12,a16-a18,a20,a22-23在实施例7中全面描述。DNA构建体的两个区域标记为R1和R2。

[0042] 图17示出了检测到的解旋酶控制的DNA移动的%的图,其对应于区域R1和R2穿过纳米孔的受控移位,以观察到的所有的解旋酶控制的DNA移动的比例表示。y轴标记=百分比(%),x轴标记=DNA构建体数目(13-16)。图上的每个点对应单独的实验(n=3)。

[0043] 图18示出了对于每个DNA构建体,每个纳米孔中检测到的解旋酶控制的DNA移动的数目的图。Y轴标记=每个纳米孔中解旋酶控制的DNA移动,x轴标记=DNA构建体数目(13-16)。图上的每个点对应单独的实验(n=3)。

#### [0044] 序列列表说明

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

[0047] SEQ ID NO:3示出了编码 $\alpha$ -溶血素-E111N/K147N( $\alpha$ -HL-NN;Stoddart等人,PNAS,2009;106(19):7702-7707)的1个单体的多核苷酸序列。

[0048] SEQ ID NO:4显示了 $\alpha$ -HL-NN的1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Phi29 DNA聚合酶的多核苷酸序列。

[0051] SEQ ID NO:9显示了Phi29 DNA聚合酶的氨基酸序列。

[0052] SEQ ID NO:10显示了衍生自大肠杆菌的sbcB基因的密码子优化的多核苷酸序列。它编码来自大肠杆菌的核酸外切酶I(EcoExo I)。

[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显示了衍生自大肠杆菌的xthA基因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。这种酶沿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'突出端(overhang)。

[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 $\lambda$ (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 $\lambda$ 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/nebecomm/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′突出端。

[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TraI 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示出了本发明的MuA底物的双链部分的多核苷酸序列。

[0069] SEQ ID NO:27示出了本发明的MuA底物的双链部分的多核苷酸序列。该序列与SEQ ID NO:26互补,不同之处在于它在3′端含有U。

[0070] SEQ ID NO:28示出了本发明的双链MuA底物的突出端链的多核苷酸序列。

[0071] SEQ ID NOs:29-32示出了在实施例1中使用的多核苷酸序列。

[0072] SEQ ID NO:33-41示出了在实施例1、2和3中使用的多核苷酸序列。

[0073] SEQ ID NO:42示出了在实施例4中使用的多核苷酸序列。

[0074] SEQ ID NO:43示出了肠杆菌噬菌体λ的多核苷酸序列。该序列包含连接在模板链的5′端的额外的12碱基突出端。这里示出的序列仅为模板链(未示出模板互补链)。该序列在实施例5中使用。

[0075] SEQ ID NO:44-48示出了在实施例5和6中使用的多核苷酸序列。

[0076] SEQ ID NO:49示出了在实施例7中使用的多核苷酸序列。在5′端,该序列含有磷酸基团。

[0077] SEQ ID NO:50示出了在实施例7中使用的多核苷酸序列。在第3′端,该序列中的最后的胸腺嘧啶具有硫代磷酸基团。

[0078] SEQ ID NO:51示出了在实施例7中使用的多核苷酸序列。

[0079] SEQ ID NO:52示出了在实施例7中使用的多核苷酸序列。

[0080] SEQ ID NO:53示出了在实施例7中使用的多核苷酸序列。在5′端,所述序列含有磷酸基。

[0081] SEQ ID NO:54示出在实施例7中使用的多核苷酸序列。

## 具体实施方式

[0082] 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。

[0083] 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”时包括两个或更多个多核苷酸,涉及“锚”时包括两个或更多个锚,涉及“解旋酶”时包括两个或更多个解旋酶,涉

及“跨膜孔”时包括两个或更多个孔,等。

[0084] 本文中,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。

#### [0085] 本发明的表征方法

[0086] 当Y适配体和发夹环适配体一起使用以表征双链多核苷酸时,所述Y适配体通常包括用于将多核苷酸耦合到所述膜的锚,以及优先旋入所述孔的前导序列。所述锚和更重要地所述前导序列通常会导致偏向于表征在两端具有Y适配体的双链多核苷酸。

[0087] 当使用大致等量的Y适配体和发夹环适配体修饰一组双链多核苷酸时,产生三组不同的修饰的多核苷酸。近似地,50%的修饰的多核苷酸在一端具有Y适配体,和在另一端具有发夹环适配体(即是“拐角(round the corner)”或RTC多核苷酸)。优选的是,这些RTC多核苷酸使用在WO 2013/014451中公开的方法表征。近似地,25%的修饰的多核苷酸在每端具有Y适配体(即双Y多核苷酸)和约25%的修饰的多核苷酸在每端具有发夹环适配体(即双链发夹多核苷酸)。如果Y适配体包括所述锚,则偏向于表征双Y多核苷酸,因为它们具有两个锚并因此能更强地耦合到所述膜。如果Y适配体包括前导,则偏向于表征双Y多核苷酸,因为它们具有两个前导且因此优选旋入所述孔中。与锚的存在相比,前导的存在通常导致更大的偏向。如果Y适配体既包括锚也包括前导,则偏向于表征双Y多核苷酸的程度最大。本发明人已经发现,具有约50倍的偏向于双Y多核苷酸。双链发夹多核苷酸通常不能被特征,因为它们不具有能够旋入用于表征的纳米孔中的单链区域。

[0088] 本发明人已表明,在发夹环适配体上使用一个或多个锚,不仅降低了对双Y多核苷酸的表征,同时也显著降低了观察到的双Y多核苷酸的吞吐量(throughput)。本发明人由此得出结论:该吞吐量取决于在旋入所述纳米孔中的修饰的多核苷酸的一端的锚的存在(即RTC多核苷酸的Y适配体末端)。

[0089] 本发明人还令人惊奇地表明,偏向于双Y多核苷酸可以通过使用对膜具有不同耦合强度的Y适配体和发夹环适配体来克服。如果发夹环适配体比Y适配体具有更大的耦合强度,则可实现偏向于RTC多核苷酸。“强”的发夹环适配体有效地将RTC多核苷酸耦合到所述膜,这使得它们能够竞争含有两个“较弱”系链的双Y适配体。由于“较弱”的Y适配体,维持了RTC多核苷酸的高吞吐量。

[0090] 本发明涉及使用在膜中的跨膜孔表征目标双链多核苷酸。所述目标双链多核苷酸配置为在一端具有Y适配体,而在另一端具有发夹环适配体。这样的方法在下面详细论述。Y适配体包括将多核苷酸耦合到膜的一个或多个第一锚,且发夹环适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚。将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大。

[0091] 配置有两个适配体的多核苷酸与跨膜孔接触,使得所述多核苷酸的至少一条链移动穿过所述孔。随着所述多核苷酸的至少一条链相对于所述孔的移动,获取一个或多个测量值。所述测量值表示所述多核苷酸的至少一条链的一个或多个特征。

[0092] 本发明优选涉及测序或估计目标双链多核苷酸的序列。如下文所详细描述,只有目标双链多核苷酸的一部分可以被表征或测序。

#### [0093] 目标双链多核苷酸

[0094] 本发明的方法表征目标双链多核苷酸。所述目标双链多核苷酸也可被称为模板双

链多核苷酸或感兴趣的双链多核苷酸。

[0095] 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。目标多核苷酸中的一个或多个核苷酸可以被氧化或甲基化。目标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,且是皮肤黑色素瘤的首要原因。目标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。所述目标多核苷酸可包含一个或多个间隔区。

[0096] 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。

[0097] 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)。

[0098] 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。

[0099] 所述目标双链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)。

[0100] 核苷酸通常为核糖核苷酸或脱氧核糖核苷酸。核苷酸优选是脱氧核糖核苷酸。所述核苷酸通常含有单磷酸、二磷酸或三磷酸。磷酸可以被连接在核苷酸的5'或3'侧。

[0101] 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)、5-甲基胞苷单磷酸、5-羟甲基胞苷单磷酸、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。核苷酸最优选选自dAMP、dTMP、dGMP、dCMP和dUMP。目标双链多核苷酸优选包含下列核苷酸:dAMP、dUMP和/或dTMP、dGMP和dCMP。

[0102] 所述目标双链多核苷酸优选包含下列核苷酸:dAMP、dUMP和/或dTMP、dGMP和dCMP。

[0103] 核苷酸可以是无碱基的(即缺乏核碱基)。核苷酸也可以是缺乏核碱基和糖(即是C3间隔区)。

[0104] 目标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。

[0105] 目标多核苷酸是双链的。目标多核苷酸可以含有一些单链区,但至少目标多核苷酸的一部分是双链的。

[0106] 所述目标多核苷酸可以是核酸,例如脱氧核糖核酸(DNA)或核糖核酸(RNA)。目标多核苷酸可以包含与DNA的一条链杂交的一条RNA链。多核苷酸可以是本领域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醇(glycol)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'-内糖褶皱(C3'-endo sugar pucker)的五元的、六元的或甚至七元的桥连结构。所述桥

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。

[0107] 所述多核苷酸最优选为核糖核酸(RNA)或脱氧核糖核酸(DNA)。

[0108] 所述目标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个核苷酸对的长度。所述多核苷酸可以是1000或更多个核苷酸对、5000或更多个核苷酸对长度,或100000或更多个核苷酸对长度。

[0109] 所述目标多核苷酸通常存在于任何合适的样品中。本发明通常针对已知含有或怀疑含有所述目标多核苷酸的样品实施。或者,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目标多核苷酸的同源性。

[0110] 所述样品可以是生物样品。本发明可以针对从任何有机体或微生物中获得或提取的样品在体外实施。所述有机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之一: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。

[0111] 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。

[0112] 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。

[0113] 提供具有两个适配体的多核苷酸

[0114] 目标双链多核苷酸上配置为在一端具有Y适配体,而在另一端具有发夹环适配体。Y适配体和/或发夹环适配体是典型的多核苷酸适配体。它们可以由上面论述的任何多核苷酸形成。

[0115] Y适配体通常包括双链区域和单链区域或在另一端不互补的区域。如果Y适配体包括单链区域,则可被描述为具有突出端。Y适配体中非互补区的存在给予适配体以Y形,因为这两条链与双链部分不同,通常彼此不杂交。Y适配体包括一个或多个第一锚。锚在下面更详细地论述。

[0116] Y适配体优选包括优先旋入所述孔中的前导序列。所述前导序列有助于本发明的方法。前导序列被设计为优先地旋入跨膜孔,且由此有助于多核苷酸穿过所述孔的移动。所述前导序列也可以用于将所述多核苷酸连接到锚(一个或多个),如下面所论述的。

[0117] 所述前导序列通常包括聚合物。所述聚合物优选带负电荷。所述聚合物优选多核苷酸,如DNA或RNA、修饰的核苷酸(如脱碱基的DNA)、PNA、LNA、聚乙二醇(PEG)或多肽。前导优选包含多核苷酸,更优选包含单链多核苷酸。前导序列可以包括上面论述的任何多核苷酸。单链前导序列最优选包括DNA的单链,如聚胸苷部分。所述前导序列优选包括所述一个或多个间隔区。

[0118] 所述前导序列可以是任意长度,但通常长度为10至150个核苷酸长度,例如20至150个核苷酸长度。所述前导的长度通常取决于本方法中使用的跨膜孔。

[0119] 合适的发夹环适配体可以使用本领域已知的方法来设计。发夹环可以是任何长度。发夹环一般为110或更少个核苷酸,例如100或更少个核苷酸,90或更少个核苷酸,80或更少个核苷酸,70或更少个核苷酸,60或更少个核苷酸,50或更少个核苷酸,40或更少个核苷酸,30个或更少核苷酸,20或更少个核苷酸,或10或更少个核苷酸长度。发夹环优选约1至110,2到100,5至80,或6到50个核苷酸长度。如果发夹环包含在适配体的不同选择性中,较长长度的发夹环,如50到110个核苷酸,是优选的。类似地,如果所述发夹环不包含在如下论述的可选结合中,较短长度的发夹环,如1至5个核苷酸,是优选的。

[0120] 所述发夹环适配体优选包括可选结合部分。这允许所述第一和/或第二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可以被选择的部分。因此,可选结合部分优选为特异性结合到表面的部分。可选结合部分特异性地结合到表面上——如果其与本发明中使用的任何部分相比,能以更大的程度结合到表面。在优选的实施方案中,所述部分结合到本发明中使用的何其他部分未结合的表面。

[0121] 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ScSv,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(即杂交)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。

[0122] 所述发夹环适配体和/或所述可选结合部分可包括可被切割、缺口、劈裂或水解的区域。这种区域可以被设计为允许所述第一和/或第二多核苷酸在纯化或分离之后从其所结合的表面去除。合适的区域是本领域中公知的。合适的区域包括,但不限于,RNA区域、包括脱硫生物素和链霉亲和素、二硫键的区域,和可光裂解的区域。

[0123] 可以使用本领域中已知的任何方法,将Y适配体和/或发夹环适配体接合到所述多核苷酸。适配体中的一个或两者可使用连接酶,如T4 DNA连接酶、大肠杆菌DNA连接酶、Taq酶DNA连接酶、Tma DNA连接酶和9°N DNA连接酶进行接合。或者,所述适配体可以使用以下论述的本发明的方法加入到所述多核苷酸。

[0124] 在优选的实施方案中,本方法的步骤a)包括修饰目标双链多核苷酸,使得其一端包括Y适配体且在另一端包括发夹环适配体。任何修饰方式均可以使用。本方法优选包括根据本发明修饰目标双链多核苷酸。这在下面更详细地论述。修饰和表征的方法可以以任何方式进行组合。

[0125] 膜

[0126] 根据本发明可以使用任何膜。合适的膜是本领域公知的。膜优选是两亲性层。两亲性层是由两亲性分子形成的层,例如磷脂,其具有亲水性和亲脂性。两亲分子可以是合成的或天然存在的。形成单层的非天然存在的两亲物(一个或多个)是本领域已知的,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更多个单体亚单元聚合在一起产生单一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单元贡献的性质。然而,嵌段共聚物可具有由各个亚单元形成的聚合物不具有的独特性质。可以改造嵌段共聚物,使得单体亚单元之一是疏水的(即亲脂的),而其它

亚单元在水性介质中是亲水的。在这种情况下,嵌段共聚物可以具有两亲性质并且可以形成模拟生物膜的结构。嵌段共聚物可以是二嵌段(由两个单体亚单元组成),但也可以由多于两个单体亚单元构成,以形成表现为两亲物的更复杂的构型。共聚物可以是三嵌段,四嵌段或五嵌段共聚物。膜优选是三嵌段共聚物膜。

[0127] 古细菌双极性性四醚脂质(bipolar tetraether lipids)是天然存在的脂质,其被构建为使得该脂质形成单层膜。这些脂质通常存在于在恶劣的生物环境存活的极端微生物,嗜热菌,嗜盐菌和嗜酸菌中。它们的稳定性被认为源于最终双层的融合性质。通过产生具有亲水-疏水-亲水的通用基序的三嵌段聚合物,可以直接构建模拟这些生物实体的嵌段共聚物材料。这种材料可以形成单体膜,其类似于脂质双层而发挥作用,并且包括从囊泡到层状膜的一系列相行为。由这些三嵌段共聚物形成的膜与生物脂质膜相比具有一些优点。因为所述三嵌段共聚物是合成的,可仔细控制精确的构造以提供形成膜和与孔及其它蛋白质相互作用所需的正确的链长度和性质。

[0128] 嵌段共聚物也可以由不归类为脂质亚材料的亚单元构造而成;例如疏水性聚合物可以由硅氧烷或其它非烃基单体制成。嵌段共聚物的亲水性子部也可以具有低的蛋白质结合性质,这允许产生当暴露于未加工的生物样品时高度耐受的膜。该头部基团单元还可以衍生自非经典脂质头部基团。

[0129] 与生物脂质膜相比,三嵌段共聚物膜还具有增加的机械和环境稳定性,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广泛的应用提供了定制基于聚合物的膜的平台。

[0130] 在优选的实施方案中,本发明提供了在包含三嵌段共聚物的膜中使用跨膜孔来表征目标双链多核苷酸的方法,可选地,其中所述膜被修饰为有助于耦合。

[0131] 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。

[0132] 所述两亲分子可以是经化学修饰的或官能化的以有助于对多核苷酸的耦合。

[0133] 该两亲性层可以是单层或双层。两亲性层通常是平面的。两亲性层可以是弯曲的。两亲性层可以被支撑。

[0134] 两亲膜通常是天然可移动的,本质上充当具有约 $10^{-8}\text{cm s}^{-1}$ 的脂质扩散速率的二维流体。这意味着所述孔和耦合的多核苷酸通常在两亲膜内移动。

[0135] 膜可以是脂质双层。脂质双层是细胞膜的模型,并且作为一系列实验研究的优良平台。例如,脂质双层可以用于通过单通道记录进行的膜蛋白的体外研究。或者,脂质双层可用作生物传感器以检测多种物质的存在。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质双层、受支撑双层,或脂质体。脂质双层优选为平面脂质双层。合适的脂质双层在国际申请号PCT/GB08/000563(公开为WO 2008/102121),国际申请号PCT/GB08/004127(公开为WO 2009/077734)和国际申请号PCT/GB2006/001057(公开为WO 2006/100484)中公开。

[0136] 形成脂质双层的方法是本领域已知的。合适的方法公开在实施例。脂质双层通常通过Montal和Mueller(Proc.Natl.Acad.Sci.USA.,1972;69:3561-3566)的方法形成,其中脂质单层被携带在水溶液/空气界面上,穿过垂直于该界面的孔的任一侧。通常通过下述将脂质添加到电解质水溶液的表面:首先将脂质溶解在有机溶剂中,然后允许溶剂滴在孔

的任一侧上的水溶液的表面上蒸发。一旦有机溶剂蒸发,孔的任一侧上的溶液/空气界面物理地上下移动,穿过所述孔直到形成双层。平面脂质双层可以形成为,穿过膜中的孔或穿过开口而进入凹部。

[0137] Montal&Mueller的方法是受流行的,因为它是形成适合于蛋白质孔插入的优质脂质双层的成本划算和相对直接的方法。形成双层的其他常见方法包括脂质体双层的尖端浸渍、喷涂双层和膜片钳。

[0138] 尖端浸渍双层的形成需要使孔表面(例如,移液管尖端)接触到携带脂质单层的测试溶液的表面上。另外,首先通过允许溶解在有机溶剂中的脂质滴在溶液表面蒸发而在溶液/空气界面处产生脂质单层。然后通过Langmuir-Schaefer方法形成双层,并且需要机械自动化以相对于溶液表面移动所述孔。

[0139] 对于喷涂双层,将溶解在有机溶剂中的脂质滴直接施加到孔中,将该孔浸没在水性测试溶液中。使用油漆刷或等效物将脂质溶液薄薄地铺在孔上。溶剂的稀释(thinning)导致脂质双层的形成。然而,从双层中完全去除溶剂是困难的,因此由该方法形成的双层在电化学测量期间不太稳定且更容易产生噪声。

[0140] 膜片钳通常用于生物细胞膜的研究。通过抽吸将细胞膜夹到移液管的末端,并且膜的小块贴附在孔上。该方法已经被调整为通过夹住脂质体,然后脂质体破裂而在移液管的孔口上形成脂质双层密封,而产生脂质双层。该方法需要稳定的巨大且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。

[0141] 脂质体可以通过超声,挤出或Mozafari方法(Colas等人(2007) Micron 38:841-847)形成。

[0142] 在优选的实施方案中,脂质双层如国际申请号PCT/GB08/004127(公开为W0 2009/077734)中所述形成。有利地,在该方法中,脂质双层由干燥的脂质形成。在最优选的实施方案中,脂质双层穿越开口形成,如W02009/077734(PCT/GB08/004127)中所述。

[0143] 脂质双层由两个相对的脂质层形成。两层脂质排列成使得它们的疏水尾部基团彼此面对以形成疏水内部。脂质的亲水性头部基团在双层的每一侧上面面向外朝向水性环境。双层可存在于许多脂质相中,包括但不限于液体无序相(流体层状相),液态有序相,固态有序相(层状凝胶相,交叉凝胶相(interdigitated gel phase))和平面双层晶体(层状亚凝胶相,层状结晶相)。

[0144] 可以使用形成脂质双层的任何脂质组合物。选择脂质组合物,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(packing density)或机械特性。脂质组合物可以包含一个或多个不同的脂质。例如,脂质组合物可以含有多达100个脂质。脂质组合物优选含有1至10个脂质。脂质组合物可以包含天然存在的脂质和/或人工脂质。

[0145] 脂质通常包含头部基团、界面部分和两个可以相同或不同的疏水尾部基团。合适的头部基团包括但不限于中性头部基团,例如二酰基甘油酯(DG)和神经酰胺(CM);两性离子头部基团,例如磷脂酰胆碱(PC)、磷脂酰乙醇胺(PE)和鞘磷脂(SM);带负电荷的头部基团,例如磷脂酰甘油(PG);磷脂酰丝氨酸(PS),磷脂酰肌醇(PI),磷酸(PA)和心磷脂(CA);以及带正电荷的头部基团,例如三甲基铵-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,例如月桂酸(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,例如油酸(顺-9-十八烷酸);和支链烃链,例如植烷酰基(phytanoyl)。链的长度和不饱和烃链中双键的位置和数目可以变化。链的长度和支链烃链中支链(例如甲基)的位置和数目可以变化。疏水性尾部基团可以作为醚或酯连接到界面部分。脂质可以是分枝菌酸。

[0146] 脂质也可以是化学修饰的。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,例如1,2-二硬脂酰基-sn-甘油基-3-磷酸乙醇胺-N-[生物素基(聚乙二醇)2000];和修饰为用于轭合的脂质,例如1,2-二油酰基-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)。尾部基团已经化学修饰的合适的脂质包括但不限于可聚合脂质,例如1,2-双(10,12-二十三碳二炔酰基)-sn-甘油基-3-磷酸胆碱;氟化脂质,例如1-棕榈酰基-2-(16-氟棕榈酰基)-sn-甘油基-3-磷酸胆碱;氘代脂质,例如1,2-二棕榈酰基-D62-sn-甘油-3-磷酸胆碱;和醚连接的脂质,例如1,2-二-O-植烷基-sn-甘油基-3-磷酸胆碱。脂质可以被化学修饰或官能化以促进多核苷酸的耦合。

[0147] 两亲性层,例如脂质组合物,通常包含一个或多个将影响该层的性质的添加剂。合适的添加剂包括但不限于脂肪酸,例如棕榈酸,肉豆蔻酸和油酸;脂肪醇,例如棕榈醇,肉豆蔻醇和油醇;甾醇,例如胆固醇,麦角固醇,羊毛甾醇,谷甾醇和豆甾醇;溶血磷脂,例如1-酰基-2-羟基-sn-甘油基-3-磷酸胆碱;和神经酰胺。

[0148] 在另一个优选的实施方案中,膜是固态层。固态层可以由有机和无机材料形成,包括但不限于微电子材料,绝缘材料例如 $\text{Si}_3\text{N}_4$ ,  $\text{Al}_2\text{O}_3$ 和 $\text{SiO}_2$ ,有机和无机聚合物例如聚酰胺,塑料例如Teflon<sup>®</sup>或者弹性体例如双组分加成固化硅橡胶,以及玻璃。固态层可以由石墨烯形成。合适的石墨烯层公开在国际申请号PCT/US2008/010637(公开为WO 2009/035647)中。

[0149] 该方法通常使用(i)包含孔的人工两亲性层,(ii)包含孔的分离的天然存在的脂质双层,或(iii)具有插入其中的孔的细胞进行实施。该方法通常使用人工两亲性层,如人工三嵌段共聚物层进行实施。该层除了孔之外还可以包含其它跨膜蛋白和/或膜内蛋白以及其他分子。下面论述合适的装置和条件。本发明的方法通常在体外进行。

[0150] 耦合

[0151] 所述多核苷酸可以使用任何已知的方法耦合到所述膜。使用至少两个锚将所述多核苷酸耦合到所述膜。

[0152] 所述Y适配体包括一个或多个第一锚。发夹环适配体包括一个或多个第二锚。每个锚包括耦合到(或结合到)所述适配体的基团,和耦合到(或结合到)所述膜的基团。每个锚可以共价地耦合到(或结合到)所述适配体和/或所述膜。

[0153] 所述Y适配体可以包含任何数量的第一锚,如2,3,4个或更多的锚。发夹环适配体可以包含任意数量的第二锚,例如2,3,4个或更多的锚。例如,一个或这两个适配体可包括两个锚,其中每个锚独立耦合到(或结合到)适配体(一个或多个)和膜。

[0154] 所述一个或多个第一锚和/或所述一个或多个第二锚可以包括一个或多个多核苷酸结合蛋白。每个锚可以包含一个或多个多核苷酸结合蛋白。所述多核苷酸结合蛋白(一个或多个)可以是上文论述的任何多核苷酸结合蛋白。

[0155] 如果膜是两亲性层,如共聚物膜或脂质双层,则所述一个或多个锚优选地包括存在于所述膜中的多肽锚和/或存在于所述膜中的疏水锚。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。

[0156] 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所述膜的组分的方式的实例在下面更详细地论述。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%。

[0157] 所述Y适配体和/或发夹环适配体可以直接耦合到膜上。所述一个或多个第一锚和/或所述一个或多个第二锚优选包括连接体。所述一个或多个第一锚和/或所述一个或多个第二锚可以包括一个或多个,例如2,3,4个或更多个连接体。1个连接体可用于将多于一个的,例如2,3,4个或更多个多核苷酸或适配体耦合到所述膜上。

[0158] 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。表征的所述多核苷酸可杂交到环状多核苷酸连接体上的互补序列。

[0159] 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断开的组分,如限制性位点或光不稳定基团。

[0160] 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,或者可以用于耦合(或结合)到多核苷酸。这在下面更详细地论述。

[0161] 可使用“锁和钥匙”排列避免多核苷酸的交联。每个连接体仅一端可一起反应以形成更长的连接体,连接体的其他端各自分别与多核苷酸或膜反应。此类连接体在国际申请No.PCT/GB10/000132(公开为WO 2010/086602)中有描述。

[0162] 连接体优选在下文论述的测序实施方案中。如果多核苷酸被永久地直接耦合到所述膜,其意义在于当多核苷酸与孔相互作用时不解耦合,则一些序列数据将由于膜和检测器之间的距离导致测序不能继续进行到多核苷酸的末端而丢失。如果使用连接体,则多核苷酸可以进行到完成为止。

[0163] 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。

[0164] 所述耦合可以是短暂的。换言之,耦合可以是,使得当多核苷酸与所述孔相互作用时,多核苷酸可以从所述膜上解耦合。

[0165] 形成永久的/稳定的或短暂的连接的化学基团在下面更详细地论述。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性层或三嵌段共聚物膜。可以使用具有6至30个碳原子长度的任何脂肪酰链,例如十六烷酸。

[0166] 在优选的实施方案中,多核苷酸,例如核酸,被耦合到两亲性层,如三嵌段共聚物膜或脂质双层。将核酸耦合到合成脂双层在以前曾使用多种不同的拴系策略(tethering strategies)实施。这些总结于下表1中。

[0167] 表1

[0168]

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

[0169] 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易于与直接添加合适的锚固基团,如胆固醇、生育酚、棕榈酸酯、硫醇、脂质和生物素基团兼容。这些不同的连接化学物给出了用于连接多核苷酸的一组选择。每个不同的修饰基团以略微不同的方式耦合多核苷酸,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(dwelling time)。短暂的耦合的优点如上所述。

[0170] 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添加到所述多核苷酸。将反应性基团添加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  $\gamma$  S添加到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物的简伸方法", *Nucleic Acids Res* **35**(10): e77)。叠氮基可以用T4多核苷酸激酶和 $\gamma$ -[2-叠氮基乙基]-ATP或 $\gamma$ -[6-叠氮己基]-ATP添加到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。更多样化的选择的化学基团,如生物素、硫醇和荧光团,可以使用末端转移酶进行添加,以将修饰的寡核苷酸纳入到ssDNA的3'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。也可以使用具有经合适修饰的核苷酸(例如,胆固醇或棕榈酸酯)的末端转移酶将锚直接添加到多核苷酸。

[0171] 所述一个或多个第一锚和/或所述一个或多个第二锚优选将目标双链多核苷酸经杂交耦合到所述膜。所述杂交可以存在于一个或多个锚中的任何部分,例如一个或多个锚和多核苷酸之间,所述一个或多个锚内,或所述一个或多个锚和膜之间。在一个或多个锚中的杂交使得以如上所论述的瞬时方式耦合。例如,连接体可以包括两个或更多个杂交在一起的多核苷酸,如3,4或5个多核苷酸。所述一个或多个第一锚和/或所述一个或多个第二锚可以杂交到所述目标双链多核苷酸。所述一个或多个第一锚可以直接杂交到目标双链多核苷酸,或直接杂交到Y适配体和/或前导序列。所述一个或多个第二锚可以直接杂交到所述

目标双链多核苷酸,或直接杂交到发夹环适配体。替代地,所述一个或多个第一锚可以被杂交到一个或多个,例如2或3个,杂交到目标双链目标多核苷酸或Y适配体和/或前导序列的中间多核苷酸(或“夹板”)。替代地,所述一个或多个第二锚可以杂交到一个或多个,例如2或3个,杂交到目标目标双链多核苷酸或发夹环适配体的中间多核苷酸(或“夹板”)。如下面更详细地论述的,所述一个或多个锚与一个或多个夹板的杂交可形成一个或多个刚性双链多核苷酸连接体。

[0172] 所述一个或多个第一锚和/或一个或多个第二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双链多核苷酸。使用T4 RNA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)。替代地,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。对于双链目标多核苷酸,可以将单链多核苷酸的片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T4RNA连接酶I来实现,用于连接到单链多核苷酸的其他区域。对于将双链多核苷酸添加到双链多核苷酸,接合可以用分别在目标多核苷酸和添加的多核苷酸上互补的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适配体的接合而产生的“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5'或3'修饰——如果使用单链多核苷酸在5'端进行连接或修饰,每天单链具有3'端或两端修饰——如果使用双链多核苷酸进行接合。

[0173] 如果多核苷酸是合成链,可以在多核苷酸的化学合成过程中引入适配体和锚。例如,多核苷酸可以使用具有连接到它的反应性基团的引物来合成。

[0174] 腺苷化(Adenylated)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,例如来自NEB的5'-DNA腺苷化试剂盒。通过在反应中用ATP取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)添加到多核苷酸的5'端。也可以使用具有适当修饰的核苷酸(例如,胆固醇或棕榈酸酯)的5'DNA腺苷酰化试剂盒将锚直接添加到多核苷酸。

[0175] 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的DNA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'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) (Poly(A))聚合酶。通过在反应中用dATP取代被修饰的核苷酸三磷酸,可将锚,例如胆固醇、硫醇、胺、叠氮化物、生物素或脂质,纳入到双链多核苷酸中。因此,扩增的多核苷酸的每个拷贝将包含锚。

[0176] 所述锚可包含耦合到、结合到下述物质或与之相互作用:单链或双链多核苷酸、多核苷酸内特定的核苷酸序列,或多核苷酸内经修饰核苷酸的模式,或多核苷酸上存在的任何其它配体。

[0177] 在锚中使用的合适的多核苷酸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、TOPO V dsDNA结合区域、人类组蛋白、大肠杆菌HU DNA结合蛋白和其他古细菌的、原核的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。

[0178] 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。

[0179] 所述一个或多个第一锚和/或所述一个或多个第二锚可包含结合到、嵌入多核苷酸或与多核苷酸相互反应的任何基团。该基团可以通过静电结合、氢键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交)和钼络合物(可以对甲基化碱基反应)。因此多核苷酸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-硝基吡啶、5-硝基吡啶、6-硝基吡啶、甲酰基吡啶、3-硝基吡咯、硝基咪唑、4-硝基吡唑、4-硝基苯并咪唑、5-硝基吡唑、4-氨基苯并咪唑或苯基(C6-芳族环)。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-脱氧杂-2'-脱氧肌苷,7-脱氧杂-肌苷,2-氮杂-脱氧肌苷,2-氮杂-肌苷,2-0'-甲基肌苷,4-硝基吡啶-2'-脱氧核糖核苷、4-硝基吡啶核糖核苷、5-硝基吡啶-2'-脱氧核糖核苷、5-硝基吡啶核糖核苷、6-硝基吡啶-2'-脱氧核糖核苷、6-硝基吡啶核糖核苷、3-硝基吡咯-2'-脱氧核糖核苷、3-硝基吡咯核糖核苷、次黄嘌呤的无环糖类似物、硝基咪唑-2'-脱氧核糖核苷、硝基咪唑核糖核苷、4-硝基吡唑-2'-脱氧核糖核苷、4-硝基吡唑核糖核苷、4-硝基苯并咪唑-2'-脱氧核糖核苷、4-硝基苯并咪唑核糖核苷、5-硝基吡唑-2'-脱氧核糖核苷、5-硝基吡唑核糖核苷、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'-deoxynebularine)、2'-脱氧异鸟苷、K-2'-脱氧核糖、P-2'-脱氧核糖和吡咯烷。通用核苷酸更优选包含2'-脱氧肌苷。通用核苷酸更优选为IMP或dIMP。通用核苷酸最优选为dPMP(2'-脱氧-P-核苷单磷酸)或dKMP(N6-甲氧基-2,6-二氨基嘌呤单磷酸)。

[0180] 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。例如,所述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,或一个或多个多核苷酸,其与多核苷酸形成Hoogsteen氢键或反向Hoogsteen氢键。这种类型的氢键允许第三个多核苷酸链缠绕到双链螺旋并形成三链体。所述一个或多个锚可以通过与双链二链体形成三链体而耦合到(或结合到)双链多核苷酸。

[0181] 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。

[0182] 如果所述一个或多个第一锚和/或一个或多个第二锚包含蛋白,其能够无需进一步官能化,直接锚固到膜中,例如,如果其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已知的。

[0183] 在另一个方面,所述Y适配体和/或所述发夹环适配体可使用上述方法官能化,从而使其可以被特定的结合基团识别。具体地,所述适配体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。

[0184] 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) 被添加到多核苷酸的5' 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。

[0185] 耦合强度

[0186] 将发夹环适配体耦合 (或结合) 到膜的强度比将Y适配体耦合 (或结合) 到膜的强度大。这可以用任何方式来测量。用于测量耦合 (或结合) 的强度的合适方法在实施例中公开。

[0187] 发夹环适配体的耦合 (或结合) 强度优选为Y适配体的耦合 (或结合) 强度的至少1.5倍, 例如至少2倍, 至少3倍, 至少4倍, 至少5倍或至少10倍于Y适配体的耦合 (或结合) 强度。发夹环适配体对膜的亲和常数 (Kd) 优选为Y适配体对膜的亲和常数的至少1.5倍, 例如至少2倍, 至少3倍, 至少4倍, 至少5倍或至少10倍于Y适配体的耦合强度。

[0188] 有几种将所述发夹环适配体比所述Y适配体更强地耦合 (或结合) 到所述膜的方法。例如, 发夹环适配体可以比Y适配体包括更多锚。例如, 所述发夹环适配体可以包括2个、3个或更多个第二锚, 而Y适配体可以包括1个第一锚。

[0189] 所述一个或多个第二锚耦合 (或结合) 到所述膜的强度可以比所述一个或多个第一锚耦合 (或结合) 到所述膜的强度大。所述一个或多个第二锚耦合 (或结合) 到所述发夹环适配体的强度可以比所述一个或多个第一锚耦合 (或结合) 到所述Y适配体的强度大。所述一个或多个第一锚和所述一个或多个第二锚可通过杂交连接到各自的适配体, 且杂交的强度在所述一个或多个第一锚中比在所述一个或多个第二锚中更大。这些实施方案的任何组合也可以用于本发明。耦合 (或结合) 的强度可以使用本领域已知的技术测量。

[0190] 所述一个或多个第二锚优选包括耦合 (或结合) 到所述膜的一个或多个基团, 其比耦合 (或结合) 到所述膜的所述一个或多个第一锚中的一个或多个基团具有更大的强度。在优选的实施方案中, 所述发夹环适配体/一个或多个第二锚使用胆固醇耦合 (或结合) 到所述膜, Y适配体/一个或多个第一锚使用棕榈酸酯耦合 (或结合) 到所述膜。胆固醇比棕榈酸酯更强地结合到三嵌段共聚物膜和脂质膜。在一个替代性实施方案中, 所述发夹环适配体/一个或多个第二锚使用单酰基物质例如棕榈酸酯耦合 (或结合) 到所述膜, 且Y适配体/一个或多个第一锚使用二酰基物质例如二棕榈酰磷脂酰胆碱耦合 (或结合) 到所述膜。可能会影响所述一个或多个基团耦合 (或结合) 到所述膜的能力的其它因素包括, 但不限于, 所述一个或多个锚或所述一个或多个连接体的电荷或疏水性, 或者熵值 (entropic cost)。例如, 所述一个或多个第一锚和一个或多个第二锚可包括胆固醇, 但所述一个或多个第二锚的电荷可以增加其胆固醇耦合 (或结合) 到膜的能力, 或所述一个或多个第一锚的电荷可以降低其胆固醇耦合 (或结合) 到膜的能力。

[0191] 所述一个或多个第一锚可包括一个或多个刚性连接体且所述一个或多个第二锚可包括一个或多个柔性连接体。所述柔性连接体可以使得耦合 (或结合) 到膜一个或多个基团能相对于目标双链多核苷酸移动, 且由此增加耦合的可能性。在本实施例中, 在第一锚和第二锚中, 耦合 (或结合) 到所述膜的所述一个或多个基团可以相同。例如, 所述一个或多个第一锚可包括一个或多个刚性连接体和胆固醇, 且所述一个或多个第二锚可包括一个或多个柔性连接体和胆固醇。合适的连接体如上所述。所述刚性连接体是基本上不弯曲或挠曲, 且不允许沿着它们的长度或在它们的连接点大量旋转的那些连接体。刚性连接体是不允许

耦合(或结合)到所述膜的所述一个或多个基团相对于目标双链多核苷酸大量移动的那些连接体。在刚性连接体包括聚合部分或低聚部分的情况下,在刚性测量时,所述聚合部分或低聚部分比对应聚合物或低聚物的持续长度(persistence length)短。持续长度可以使用本领域中已知的方法测量或计算。例如,双链DNA的持续长度为大约50nm或500埃或147个核苷酸对。小于147个核苷酸对长度的DNA的双链DNA的链段被认为是刚性的。刚性连接体包括,但不限于,包含双链多核苷酸的那些刚性连接体,包括DNA,共轭有机部分或具有限制构象自由度的庞大侧链的有机部分。所述一个或多个刚性连接体优选包含一个或多个双链多核苷酸。也可以使用上面论述的任何多核苷酸。所述一个或多个刚性连接体的优选长度为5,10,15,20,25,27,30,35,40或更多个核苷酸对。

[0192] 柔性连接体是基本上弯曲或挠曲或旋转的那些连接体。柔性连接体可以沿着它们的长度或在一个或两个其连接点处弯曲或挠曲。柔性连接体允许耦合(或结合)到膜的一个或多个基团相对于目标双链多核苷酸的大量移动。柔性连接体是允许目标双链多核苷酸的取向进行实质变化的那些连接体。柔性连接体可通过一维或二维挠曲允许所述一个或多个基团移动。替代地或另外地,柔性连接体可允许一个或多个基团绕轴旋转而移动。在柔性连接体包括聚合部分的或低聚部分的情况下,如果聚合部分或低聚部分与相应的聚合物或低聚物的持续长度相似或比其更长,则聚合部分或低聚部分贡献显著的柔性。柔性连接体包括,但不限于,那些含有单链寡核苷酸或多核苷酸、短碳间隔区(例如烷烃和烯烃)、多肽、多组氨酸标签、核苷衍生物、短多糖或它们的组合的连接体。柔性连接体通常还允许耦合(或结合)到膜的一个或多个基团的旋转。柔性连接体可以允许在至少一个轴中的转动。所述轴优选为连接体的纵向轴。所述一个或多个柔性连接体优选包括一个或多个间隔区(iSp9)基团或一个或多个间隔区18(iSp18)基团。所述一个或多个柔性连接体优选包括1,2,3,4,5,6,7,8,9,10或更多个间隔区9(iSp9)基团和/或1,2,3,4,5,6,7,8,9,10或更多个间隔区18(iSp18)基团。

[0193] 优选地,(a)所述发夹环适配体比Y适配体包括更多锚,(b)所述一个或多个第二锚耦合到膜的强度比所述一个或多个第一锚耦合到膜的强度大,(c)将所述一个或多个第二锚耦合到所述发夹环适配体的强度比将所述一个或多个第一锚耦合到Y适配体的强度大,(d)所述一个或多个第一锚和所述一个或多个第二锚经杂交将它们各自的适配体耦合到膜,且杂交的强度在所述一个或多个第二锚中比在所述一个或多个第一锚中大,(e)所述一个或多个第一锚包括一个或多个刚性连接体,且所述一个或多个第二锚包括一个或多个柔性连接体,或(f)(a)到(e)的任何组合。换句话说,本发明优选包含(a)和(b)的组合,(b)和(c)的组合,(c)和(d)的组合,(a)和(c)的组合,(b)和(d)的组合,(a)和(d)的组合,(a)和(e)的组合,(b)和(e)的组合,(c)及(e)的组合,(d)和(e)的组合,(a)、(b)及(c)的组合,(b)、(c)及(d)的组合,(a)、(b)和(d)的组合,(a)、(c)和(d)的组合,(a)、(b)和(e)的组合,(a)、(c)和(e)的组合,(a)、(d)和(e)的组合,(b)、(c)和(e)的组合,(b)、(d)和(e)的组合,(c)、(d)和(e)的组合,(a)、(b)、(c)及(d)的组合,(a)、(b)、(c)和(e)的组合,(a)、(b)、(d)和(e)的组合,(a)、(c)、(D)和(e)的组合,(b)、(c)、(d)和(e)的组合,或(a)、(b)、(c)、(d)和(e)的组合。

[0194] 跨膜孔

[0195] 所述方法包括当多核苷酸的至少一条链相对于跨膜孔移动时,获取一个或多个测

量值。可使用所述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、隧道测量(tunnelling measurement) (Ivanov AP et al., Nano Lett. 2011 Jan 12; 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)。所述测量可以是跨膜电流测量,如流经孔的离子电流的测量值。

[0196] 电测量值可以使用在Stoddart D等人, Proc Natl Acad Sci, 12; 106(19): 7702-7, Lieberman KR等人, J Am Chem Soc. 2010; 132(50): 17961-72, 和国际申请WO 2000/28312中所描述的标准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。

[0197] 本方法优选跨膜施加的电势实施。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, 如两亲性层, 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。

[0198] 该方法包括将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阱、间隙、通道、沟槽或狭缝。

[0199] 步骤(c)优选包括: 随着至少一条链相对于所述孔的移动, 获取流经所述孔的电流的一个或多个测量值, 其中所述一个或多个测量值表示所述至少一条链的一个或多个特征。

[0200] 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等人, Science, 2012; 338: 932-936)。

[0201] 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, 移动穿过所述孔。

[0202] 跨膜蛋白质孔可以是单体或低聚物。所述孔优选由几个重复的亚基, 例如至少3个, 至少4个, 至少5个, 至少6个、至少7个、至少8个或至少9个, 至少10个, 至少11个, 至少12个, 至少13个, 或至少14个亚基组成。例如3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13或14个亚基。所述孔优选是六聚体的、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。

[0203] 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 $\beta$ 桶状体或通道或跨膜 $\alpha$ -螺旋束或通道。

[0204] 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收缩部(constriction)。

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。

[0205] 本发明使用的跨膜蛋白孔可衍生自 $\beta$ 桶状体孔或 $\alpha$ -螺旋束孔。 $\beta$ 桶状体孔包含由 $\beta$ 链形成的桶状体或通道。合适的 $\beta$ 桶状体孔包括,但不限于, $\beta$ -毒素、如 $\alpha$ -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、胞溶素、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)。 $\alpha$ -螺旋束孔包含由 $\alpha$ -螺旋形成的桶状体或通道。合适的 $\alpha$ -螺旋束孔包括但不限于内膜蛋白和 $\alpha$ -外膜蛋白,如WZA和ClyA毒素。跨膜孔可以衍生自胞溶素。衍生自胞溶素的合适的孔在国际申请No. PCT/GB2013/050667(公开为W/O 2013/153359)中公开。跨膜孔可以衍生自Msp或 $\alpha$ -溶血素( $\alpha$ -HL)。

[0206] 跨膜蛋白孔优选衍生自Msp,优选MspA。这类孔为寡聚的且通常包含7个,8个,9个或10个衍生自Msp的单体。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或其同源物或旁系同源物(paralog)。

[0207] 衍生自Msp的单体通常包含SEQ ID NO:2或其变体所示的序列。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,所述多肽具有这样的氨基酸序列: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。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,变体可以与其他合适的亚基一起插入到两亲性层中,且可以确定其寡聚形成孔的能力。将亚基插入膜例如两亲性层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到膜并组装成功能状态而插入。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W/O 2006/100484)中所描述的“拾取和放置”方法将亚基直接插入膜中。

[0208] 在SEQ ID NO:2的氨基酸序列的整个长度上,基于氨基酸同一性,变体将优选与该序列至少50%同源。更优选地,基于氨基酸同一性,所述变体可以是至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,且更优选至少95%,97%或99%在整个序列上与SEQ ID NO:2的氨基酸序列同源。在100个或更多个,例如125,150,175或200个或更多个连续氨基酸的片段(stretch)上,可以具有至少80%,例如至少85%,90%或95%的氨基酸同一性(“严格同源性(hard homology)”)。

[0209] 可用本领域的标准方法测定同源性。例如,UWGCG包提供了用于计算同源性的BESTFIT程序,例如使用其默认设置(Devereux等人(1984) *Nucleic Acids Research* 12, p387-395)。可用PILEUP和BLAST算法计算同源性或比对(line up)序列(例如鉴定等效残基或相应序列(通常基于其默认设置)),例如Altschul S.F. (1993) *J Mol Evol* 36:290-300; Altschul, S.F等人(1990) *J Mol Biol* 215:403-10中所述。用于执行BLAST分析的软件可通过国家生物技术信息中心(<http://www.ncbi.nlm.nih.gov/>)公开获得。

[0210] SEQ ID NO:2是MspA单体的MS-(B1)8突变体。与MspA相比,该变体可以包含MspB、C

或D单体中的任何突变。MspB、C和D的成熟形式示于SEQ ID NO:5至7中。特别地,所述变体可以包含存在于MspB:A138P中的以下取代。所述变体可以包含存在于MspC中的一个或多个下列取代:A96G,N102E和A138P。所述变体可包含存在于MspD中的一个或多个以下突变:G1,L2V,E5Q,L8V,D13G,W21A,D22E,K47T,I49H,I68V,D91G,A96Q,N102D,S103T,V104I,S136K和G141A的缺失。所述变体可以包含来自Msp B、C和D中的一个或多个突变和取代的组合。变体优选包含突变L88N。SEQ ID NO:2的变体除了MS-B1的所有突变之外还具有突变L88N,并称为MS-(B2)8。本发明中使用的孔优选为MS-(B2)8。除了MS-B1的所有突变之外,SEQ ID NO:2的变体具有突变G75S/G77S/L88N/Q126R,并称为MS-B2C。本发明中使用的孔优选为MS-(B2)8或MS-(B2C)8。

[0211] 除了上文论述的那些之外,可以对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。保守取代用具有相似化学结构、相似化学性质或相似侧链体积的其他氨基酸替换氨基酸。引入的氨基酸可以具有与它们替代的氨基酸相似的极性、亲水性、疏水性、碱性、酸性、中性或电荷。或者,保守取代可以引入另一个芳香族或脂肪族氨基酸代替之前存在的芳香族或脂肪族氨基酸。

[0212] SEQ ID NO:2的氨基酸序列的一个或多个氨基酸残基可以另外从上述多肽中缺失。可以缺失多至1,2,3,4,5,10,20或30个残基或更多个残基。

[0213] 变体可以包括SEQ ID NO:2的片段。这类片段保留成孔活性。片段的长度可以为至少50,100,150或200个氨基酸。这样的片段可以用于产生孔。片段优选包含SEQ ID NO:2的孔形成结构域。片段必须包括SEQ ID NO:2的残基88,90,91,105,118和134中的一个。通常,片段包括SEQ ID NO:2的所有残基88,90,91,105,118和134。

[0214] 一个或多个氨基酸可以替代地或另外地添加到上述多肽。可以在SEQ ID NO:2或其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。延伸可以非常短,例如1至10个氨基酸长度。或者,延伸可以较长,例如高达50或100个氨基酸。载体蛋白可以与本发明的氨基酸序列融合。其它融合蛋白在下面更详细地论述。

[0215] 如上所述,变体是具有下述氨基酸序列的多肽:从SEQ ID NO:2的氨基酸序列变化而来并且保留其形成孔的能力。变体通常包含负责形成孔的SEQ ID NO:2的区域。含有 $\beta$ 桶状体的Msp的孔形成能力由每个亚基中的 $\beta$ 片体提供。SEQ ID NO:2的变体通常包含SEQ ID NO:2中形成 $\beta$ 片体的区域。可以对SEQ ID NO:2的形成 $\beta$ 片体的区域进行一个或多个修饰,只要所得变体保留其形成孔的能力即可。SEQ ID NO:2的变体优选包括在其 $\alpha$ 螺旋和/或环区域内的一个或多个修饰,例如取代、添加或缺失。

[0216] 衍生自Msp的单体含有一个或多个特异性修饰以便于对核苷酸的鉴别。衍生自Msp的单体也可以含有其它非特异性修饰,只要它们不干扰孔形成即可。许多非特异性侧链修饰是本领域已知的,并且可以对衍生自Msp的单体的侧链进行。这样的修饰包括例如通过与醛反应、然后用NaBH<sub>4</sub>还原而进行的氨基酸的还原性烷基化,用甲基乙酰亚胺酯(methylacetimidate)进行的脒基化,或用乙酸酐进行的酰化。

[0217] 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(公开号为W0 2010/004273),PCT/GB09/001679(公开号为W0 2010/004265)或PCT/GB10/000133(公开号为W0 2010/086603)中有论述。论

述了将孔插入膜的方法。

[0218] 跨膜蛋白孔也优选衍生自 $\alpha$ -溶血素( $\alpha$ -HL)。野生型 $\alpha$ -HL孔由七个相同的单体或亚基形成(即它是七聚体)。 $\alpha$ -溶血素-NN的一个单体或亚基的序列显示在SEQ ID NO:4中。

[0219] 在一些实施方案中,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。所述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。

[0220] 孔中的任何数量的单体可以是化学修饰的。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。

[0221] 半胱氨酸残基的反应性可以通过修饰相邻残基来增强。例如,侧翼精氨酸、组氨酸或赖氨酸残基的碱基将半胱氨酸硫醇基团的pKa变为更具反应性的S-基团的pKa。半胱氨酸残基的反应性可以用硫醇保护基团如dTNB保护。所述保护基团可以在接头连接之前与孔的一个或多个半胱氨酸残基反应。

[0222] 所述分子(使用其对孔进行化学修饰)可以直接连接到所述孔或如国际申请No. PCT/GB09/001690(公开号WO 2010/004273),PCT/GB09/001679(公开号WO 2010/004265)或PCT/GB10/000133(公开号WO 2010/086603)中所公开通过接头连接到所述孔。

[0223] 本文所述的任何蛋白质,例如跨膜蛋白孔可以使用本领域已知的标准方法制备。编码孔或构建体的多核苷酸序列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标准技术在细菌宿主细胞中表达。孔可以通过在细胞中由重组表达载体的多肽的原位表达而制备。表达载体任选地携带诱导型启动子以控制多肽的表达。这些方法在Sambrook, J. 和Russell, D. (2001). *Molecular Cloning: A Laboratory Manual*, 第3版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY中有描述。

[0224] 孔可以从产生蛋白质的生物体通过任何蛋白质液相色谱系统纯化后,或在重组表达后大规模生产。典型的蛋白质液相色谱系统包括FPLC系统、AKTA系统、Bio-Cad系统、Bio-Rad BioLogic系统和Gilson HPLC系统。。

[0225] 多核苷酸表征

[0226] 本发明的方法包括测量目标双链多核苷酸的一个或多个特征。

[0227] 可以研究任意数量的多核苷酸。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个或更多个多核苷酸。

[0228] 所述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两个或更多个寡核苷酸的序列。该方法一般在体外进行。

[0229] 该方法可以包括测量多核苷酸的两个、三个、四个或五个或更多个特征。所述一个或多个特征优选选自(i)所述多核苷酸的长度,(ii)所述多核苷酸的同源性,(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。根据本发明可以测量(i)至(v)的任意组合,例如{i},{ii},{iii},{iv},{v},{i,ii},{i,iii},{i,iv},{i,v},{ii,iii},{ii,iv},{ii,v},{iii,iv},{iii,v},{iv,v},{i,ii,iii},{i,ii,

iv}, {i,ii,v}, {i,iii,iv}, {i,iii,v}, {i,iv,v}, {ii,iii,iv}, {ii,iii,v}, {ii,iv,v}, {iii,iv,v}, {i,ii,iii,iv}, {i,ii,iii,v}, {i,ii,iv,v}, {i,iii,iv,v}, {ii,iii,iv,v} 或 {i,ii,iii,iv,v}。可以测量第一多核苷酸的与第二多核苷酸相比不同的(i)到(v)的组合,包括上面所列的任意组合。

[0230] 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。

[0231] 对于(ii),对多核苷酸的鉴别可以通过多种方式进行测定。对多核苷酸的识别可以联合多核苷酸序列的测定或不联合多核苷酸序列的测定而进行测定。前者是直接测定;对所述多核苷酸进行测序,并由此进行识别。后者可以多种方式完成。例如,可以测定多核苷酸中特定模序的存在(不需要测定多核苷酸的其余序列)。替代地,在所述方法中特定电和/或光信号的测定可以鉴别来自特定来源的多核苷酸。

[0232] 对于(iii),多核苷酸的序列可以如前文所述进行测定。适合的测序方法,特别是那些使用电测量的方法,在Stoddart D等人,Proc Natl Acad Sci,12;106(19):7702-7, Lieberman KR等人,J Am Chem Soc.2010;132(50):17961-72,以及国际申请WO 2000/28312中有描述。

[0233] 对于(iv),二级结构可以多种方式测定。例如,如果所述方法涉及电测量,则二级结构可以利用流经孔的停留时间的变化或电流的变化进行测定。这使得能够对单链和双链多核苷酸区域进行辨别。

[0234] 对于(v),可以测定任何修饰的存在与否。所述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、标签或间隔区进行了修饰。特定的修饰将导致与孔的特定的相互作用,这可使用下述方法测定。例如,可以基于孔与每个核苷酸的相互作用过程中流经所述孔的电流,来区分甲基胞嘧啶和胞嘧啶。

[0235] 所述方法可以使用适于研究膜/孔系统——其中孔存在于膜中——的任何设备进行。该方法可以使用适合于跨膜孔传感的任何设备进行。例如,所述设备包括含水溶液的腔室,以及将所述腔室分成两个区段的屏障。所述屏障典型地具有开孔,含所述孔的膜形成于所述开孔中。替代地,所述屏障形成其中存在孔的膜。

[0236] 所述方法可使用国际申请No.PCT/GB08/000562(WO 2008/102120)中描述的设备进行。

[0237] 所述方法可以包括测量所述多核苷酸相对于所述孔移动时穿过所述孔的电流。因此,所述设备也可以包括能够施加电位并测量穿过所述膜和孔的电信号的电路。所述方法可以使用膜片钳或电压钳进行。所述方法优选涉及使用电压钳。

[0238] 本发明方法可包括,当多核苷酸相对于孔移动时,测量穿过该孔的电流。用于测量穿过跨膜蛋白孔的离子电流的适宜条件是本领域已知的并在实施例中公开。所述方法通常通过跨膜和孔施加电压来实施。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。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,所述上限独立地选自+10mV,+20mV,+50mV,+100mV,+150mV,+200mV,+300mV和+400mV。使用的电压更优选在100mV到240mV范围内,最优选在120mV到220mV范围内。可通过对孔施加提高的电位来提高对不同核苷酸的鉴别力。

[0239] 所述方法通常在任何载荷子诸如金属盐(如碱金属盐)、卤盐(如氯化物盐(如碱金属氯化物盐))的存在下实施。载荷子可包括离子液体或有机盐,例如四甲基氯化铵、三甲基苯基氯化铵、苯基三甲基氯化铵或1-乙基-3-甲基氯化咪唑鎓。在上述示例性设备中,所述盐以水溶液存在于室中。通常使用氯化钾(KCl)、氯化钠(NaCl)、氯化铯(CsCl),或亚铁氰化钾和铁氰化钾的混合物。优选KCl、NaCl,和亚铁氰化钾和铁氰化钾的混合物。载荷子可以是跨所述膜不对称的。所述载荷子的类型和/或浓度在膜的每一侧上可以不同。

[0240] 所述盐浓度可以是饱和的。所述盐浓度可以是3M或更低并且通常为0.1-2.5M, 0.3-1.9M, 0.5-1.8M, 0.7-1.7M, 0.9-1.6M或1M-1.4M。所述盐浓度优选为150mM-1M。所述方法优选使用至少0.3M,例如至少0.4M,至少0.5M,至少0.6M,至少0.8M,至少1.0M,至少1.5M,至少2.0M,至少2.5M或至少3.0M的盐浓度而实施。高的盐浓度提供高的信噪比并使得电流能在正常电流波动的背景下识别代表核苷酸存在。

[0241] 所述方法通常在缓冲剂的存在下实施。在上述示例性设备中,所述缓冲剂以水溶液存在于所述室中。任何缓冲剂均可用在本发明方法中。通常,所述缓冲剂为磷酸盐缓冲液。另一个适宜的缓冲剂为HEPES和三羟甲基氨基甲烷-HCl(Tris-HCl)缓冲液。所述方法通常在4.0-12.0, 4.5-10.0, 5.0-9.0, 5.5-8.8, 6.0-8.7或7.0-8.8或7.5-8.5的pH下实施。所用pH优选约为7.5。

[0242] 所述方法可在0°C-100°C、15°C-95°C、16°C-90°C、17°C-85°C、18°C-80°C、19°C-70°C或20°C-60°C实施。所述方法通常在室温实施。所述方法可选地在能支持酶功能的温度下,例如约37°C实施。

[0243] 步骤(b)优选包含将在步骤a)中提供的多核苷酸与多核苷酸结合蛋白接触,使得蛋白控制所述多核苷酸的至少一条链穿过所述孔的移动。

[0244] 更优选地,所述方法包括b)将在步骤a)中提供的多核苷酸与跨膜孔和多核苷酸结合蛋白接触,使得所述多核苷酸的至少一条链移动穿过所述孔,且蛋白质控制所述多核苷酸的至少一条链穿过所述孔的移动;以及c)随着所述多核苷酸的至少一条链相对于所述孔的移动,测量穿过所述孔的电流,其中所述电流代表所述多核苷酸的至少一条链的一个或多个特征,且由此表征目标双链多核苷酸。

[0245] 所述多核苷酸结合蛋白可以是能够结合到多核苷酸且控制其移动穿过所述孔的任何蛋白。现有技术中可直接确定所述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各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即控制其移动来修饰该多核苷酸。

[0246] 所述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形成各单个核苷酸或较短链的核苷酸,如二核苷酸或三核苷酸,来修饰该多核苷酸。所述酶可以通过定向多核苷酸或将多核苷酸移动到特定的位置而修饰该多核苷酸。所述多核苷酸处理酶不需要显示酶活性,只要其能够结合多核苷酸并控制其移动穿过所述孔即可。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论述。

[0247] 所述多核苷酸处理酶优选衍生自溶核酶。所使用的多核苷酸处理酶更优选衍生自

酶分类(EC)组3.1.11,3.1.13,3.1.14,3.1.15,3.1.16,3.1.21,3.1.22,3.1.25,3.1.26,3.1.27,3.1.30和3.1.31中任何组的成员。所述酶可为在国际申请No.PCT/GB10/000133(公开为WO 2010/086603)中公开的任何酶。

[0248] 优选的酶为聚合酶、核酸外切酶、解旋酶和拓扑异构酶,例如促旋酶。合适的酶包括,但不限于,来自大肠杆菌的核酸外切酶I(SEQ ID NO:11)、来自大肠杆菌的核酸外切酶III(SEQ ID NO:13)、来自嗜热栖热菌的RecJ(SEQ ID NO:15),和噬菌体 $\lambda$ 核酸外切酶(SEQ ID NO:17)及其变体。包含SEQ ID NO:15或其变体所示序列的三个亚基相互作用形成三聚体核酸外切酶。所述酶优选是Phi29DNA聚合酶(SEQ ID NO:9)或其变体。拓扑异构酶优选是部分分类(EC)组5.99.1.2和5.99.1.3中的任意成员。

[0249] 酶最优选衍生自解旋酶,例如He1308Mbu(SEQ ID NO:18),He1308Csy(SEQ ID NO:19),He1308Mhu(SEQ ID NO:21),TRAI Eco(SEQ ID NO:22),XPD Mbu(SEQ ID NO:23)或其变体。本发明可以使用任何解旋酶。所述解旋酶可以是或者衍生自He1308解旋酶、RecD解旋酶,例如TraI解旋酶或TrwC解旋酶、XPD解旋酶或Dda解旋酶。所述解旋酶可以是国际申请No.PCT/GB2012/052579(公开号WO 2013/057495);PCT/GB2012/053274(公开号WO 2013/098562);PCT/GB2012/053273(公开号WO2013098561);PCT/GB2013/051925(公开号WO 2014/013260);PCT/GB2013/051924(公开号WO 2014/013259);PCT/GB2013/051928(公开号WO 2014/013262);以及PCT/GB2014/052736中公开的解旋酶、经修饰的解旋酶或解旋酶构建体中的任意成员。

[0250] 所述解旋酶优选包含SEQ ID NO:25(Trwc Cba)或其变体中所示序列,SEQ ID NO:18(He1308 Mbu)或其变体中所示序列,或者SEQ ID NO:24(Dda)或其变体中所示序列。各变体可以以下文针对跨膜孔所述任意方式不同于天然序列。SEQ ID NO:24的优选变体包含E94C/A360C和之后的( $\Delta$ M1)G1G2(即M1缺失,然后添加G1和G2)或E94C/A360C/C109A/C136A和之后的( $\Delta$ M1)G1G2。

[0251] 本发明可以使用任何数目的解旋酶。例如,可以使用1,2,3,4,5,6,7,8,9,10或更多个解旋酶。在一些实施方案中,可使用其他数目的解旋酶。

[0252] 本发明的方法优选包括使目标双链多核苷酸与两个或更多个解旋酶接触。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。

[0253] 所述两个或更多个解旋酶可以是上文提及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。

[0254] 所述两个或更多个解旋酶优选彼此连接。所述两个或更多个解旋酶更优选彼此共价连接。所述解旋酶可以以任何顺序使用任何方法连接。用于本发明的优选的解旋酶构建体在国际申请PCT/GB2013/051925(公开号WO 2014/013260);PCT/GB2013/051924(公开号WO 2014/013259);PCT/GB2013/051928(公开号WO 2014/013262);以及PCT/GB2014/052736中有描述。

[0255] SEQ ID NO:9,11,13,15,17,18,19,20,21,22,23,24或25的变体是具有下述氨基酸序列的酶:所述氨基酸序列从SEQ ID NO:9,11,13,15,17,18,19,20,21,22,23,24或25变化而来并保留多核苷酸结合能力。这可以使用本领域已知的任何方法测定。例如,使所述变

体与多核苷酸接触,然后测量其结合到多核苷酸并沿多核苷酸移动的能力。所述变体可以包括能促进对多核苷酸的结合和/或促进其在高盐浓度和/或室温下活性的修饰。各变体可以被修饰为,使其结合多核苷酸(即保留多核苷酸结合能力),但不具有解旋酶功能(即当提供了有助于移动的所有必要组分例如ATP和 $Mg^{2+}$ 时仍不能沿多核苷酸移动)。这种修饰是本领域公知的。例如,对解旋酶中 $Mg^{2+}$ 结合结构域的修饰通常导致不具有解旋酶功能的变体。这些类型的变体可用作分子制动器。

[0256] 基于氨基酸同一性,在SEQ ID NO:9,11,13,15,17,18,19,20,21,22,23,24或25的氨基酸序列的整个长度上,变体将优选与该序列至少50%同源。更优选,基于氨基酸同一性,所述变体多肽在整个长度上可以与SEQ ID NO:9,11,13,15,17,18,19,20,21,22,23,24或25的氨基酸序列,至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%且更优选至少95%,97%或99%同源。在200或更多,例如230,250,270,280,300,400,500,600,700,800,900或1000或更多个连续氨基酸片段上,可以具有至少为80%,例如至少85%,90%或95%的氨基酸同源性(“严格同源性”)。同源性可如上文所述进行确定。所述变体可以上文针对SEQ ID NO:2和4论述的任意方式不同于野生型序列。所述酶可以共价连接到孔。可使用任何方法将所述酶共价连接到孔。

[0257] 优选的分子制动器是TrwC Cba-Q594A (SEQ ID NO:25,具有突变Q594A)。该变异不具有解旋酶的功能(即能结合多核苷酸,但当提供了有助于移动的所有必要组分例如ATP和 $Mg^{2+}$ 时仍不能沿多核苷酸移动)。

[0258] 在链测序中,所述多核苷酸顺着或逆着所施加电位移位穿过所述孔。逐渐地或进行性地作用于双链多核苷酸上的核酸外切酶可用于在施加的电位下在所述孔的顺侧上使剩余的单链穿过,或者在反向电位下在所述孔的反侧上使剩余的单链穿过。同样地,解开双链DNA的解旋酶也可以类似的方式使用。也可以使用聚合酶。还有可能有这样的测序应用:其需要使链逆着施加电位移位,但是DNA必须在反向电位或没有电位的情况下首先被所述酶“捕获”。然后随着在结合之后电位被翻转,所述链将从顺侧到反侧而穿过所述孔并通过电流保持延伸的构造。单链DNA核酸外切酶或单链DNA依赖性聚合酶可用作分子马达,用于以受控且逐步的方式将最近移位的单链逆着施加的电位从反侧到顺侧穿过所述孔而拉回。

[0259] 解旋酶(一个或多个)和分子制动器(一个或多个)

[0260] 在优选的实施方案中,本发明提供一种控制双链多核苷酸穿过跨膜孔移动的方法,包括:

[0261] (a) 提供在一端具有Y适配体和另一端具有发夹环适配体的双链多核苷酸,其中所述Y适配体包括一个或多个解旋酶和用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,其中所述发夹环适配体包括一个或多个分子制动器和用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且其中将所述发夹环适配体耦合到所述膜的强度比将所述Y适配体耦合到所述膜的强度大;

[0262] (b) 将在步骤a)中提供的目标多核苷酸与所述孔接触;和

[0263] (c) 跨所述孔施加电势,使得所述一个或多个解旋酶和所述一个或多个分子制动器聚集到一起,且二者一起控制目标多核苷酸穿过所述孔的移动。

[0264] 这类方法在国际申请PCT/GB2014/052737中有详细论述。

[0265] 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所述多核苷酸并控制其穿过所述孔的运动的任何蛋白。在本领域中可以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且修饰所述多核苷酸的至少一个特性。该蛋白可通过使多核苷酸裂解形成各单个核苷酸或短链核苷酸，例如二核苷酸或三核苷酸，而修饰所述多核苷酸。所述部分可以通过将多核苷酸定向或移动到特定位置，即控制其移动，而修饰所述多核苷酸。

[0266] 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衍生自上面论述的任何多核苷酸处理酶。充当分子制动器的phi29聚合酶的经修饰的变体 (SEQ ID NO:8) 在美国专利No.5576204中公开。所述一个或多个分子制动器优选衍生自解旋酶。

[0267] Y适配体中的间隔区

[0268] 如在国际申请No.PCT/GB2014/050175 (公开为WO 2014/135838) 中所论述的, 连接到Y适配体的所述一个或多个解旋酶可以在一个或多个间隔区处停滞。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区的任何结构可以在本发明中使用。

[0269] RTC测序

[0270] 在优选的实施方案中, 步骤b) 包括, 将在步骤a) 中提供的多核苷酸与跨膜孔接触, 使得所述多核苷酸的两条链移动穿过所述孔, 且步骤c) 包括, 随着所述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, 其中所述测量值表示所述多核苷酸链的一个或多个特征, 并由此表征所述目标多核苷酸。上面论述的任何实施方式也同样适用于本实施方式。

[0271] 解耦合

[0272] 本发明的方法可以包括表征多个目标双链多核苷酸和将第一目标双链多核苷酸解耦合。

[0273] 在优选的实施方案中, 本发明涉及表征两个或更多个目标双链多核苷酸。所述方法优选包括:

[0274] (a) 在第一样品中提供第一目标双链多核苷酸, 其一端具有第一Y适配体且另一端具有第一发夹环适配体, 其中所述第一Y适配体包含用于将所述多核苷酸耦合至膜的一个或多个第一锚, 其中所述第一发夹环适配体包含用于将所述多核苷酸耦合至所述膜的一个或多个第二锚, 并且其中所述第一发夹环适配体耦合到所述膜的强度大于所述第一Y适配体耦合到所述膜的强度;

[0275] (b) 在第二样品中提供第二目标双链多核苷酸, 其一端具有第二Y适配体且另一端具有第二发夹环适配体, 其中所述第二Y适配体包含用于将所述多核苷酸耦合至膜的一个或多个第三锚, 其中所述第二发夹环适配体包含用于将所述多核苷酸耦合至所述膜的一个或多个第四锚, 并且其中所述第二发夹环适配体耦合到所述膜的强度大于所述第二Y适配体耦合到所述膜的强度;

[0276] (c) 将步骤a) 中提供的所述第一多核苷酸耦合到膜;

[0277] (d) 将步骤(c) 中耦合的所述第一多核苷酸与跨膜孔接触, 使得所述第一多核苷酸的至少一条链移动穿过所述孔;

[0278] (e) 随着所述第一多核苷酸的至少一条链相对于所述孔移动, 获取一个或多个测量值, 其中所述测量值表示所述第一多核苷酸的至少一条链的一个或多个特征, 且由此表

征所述第一多核苷酸；

[0279] (f) 将所述第一多核苷酸从所述膜解耦合；

[0280] (g) 将在步骤(b)中提供的所述第二多核苷酸耦合到所述膜；

[0281] (h) 将步骤(g)中耦合的所述第二多核苷酸与跨膜孔接触,使得所述第二多核苷酸的至少一条链移动穿过所述孔;和

[0282] (i) 随着所述第二多核苷酸的至少一条链相对于所述孔的移动,获取一个或多个测量值,其中所述测量值表示所述第二多核苷酸的至少一条链的一个或多个特征,并由此表征所述第一多核苷酸。

[0283] 这类方法在英国申请1406155.0中和与本申请同时申请的国际申请(ONT IP 055)中有详细论述。其中论述的任何实施方式均适用于本方法。

[0284] 步骤(f)(即第一多核苷酸的解耦合)可以在步骤(g)之前(即在将所述第二多核苷酸耦合到膜之前)进行。步骤(g)可以在步骤(f)之前进行。如果所述第二多核苷酸在所述第一多核苷酸解耦合之前被耦合到所述膜,则步骤(f)优选包括选择性地将所述第一多核苷酸从所述膜解耦合(即将所述第一多核苷酸从所述膜解耦合,而不是将所述第二多核苷酸从所述膜解耦合)。本领域技术人员可以设计能实现选择性解耦合的系统。步骤(f)和步骤(g)可以同时进行。这在下面更详细地论述。

[0285] 去除或洗涤

[0286] 尽管在步骤(f)中所述第一多核苷酸从所述膜解耦合,但是不需要将所述第一多核苷酸去除或洗去。如果所述第二多核苷酸可以容易地与所述第一多核苷酸区分开,则没有必要去除所述第一多核苷酸。

[0287] 步骤(f)和步骤(g)之间,所述方法优选进一步包括从所述膜去除至少一些所述第一样品。至少10%的所述第一样品可以被去除,例如至少20%,至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,或至少90%的所述第一样品可以被去除。

[0288] 所述方法更优选进一步包括,从所述膜去除所有的第一样品。这可以用任何方式来完成。例如,可以在所述第一多核苷酸已经解耦合之后,用缓冲液洗涤所述膜。合适的缓冲液在下面论述。

[0289] 经修饰的目标双链多核苷酸

[0290] 在步骤a)之前,可以在使用目标双链多核苷酸作为模板使所述聚合酶形成经修饰的目标双链多核苷酸的情况下,将目标双链多核苷酸与聚合酶和一组游离核苷酸接触,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质来代替所述目标多核苷酸中的一个或多个核苷酸物质。然后,所述经修饰的目标双链多核苷酸可以像在步骤a)中一样进行提供。这种类型的修饰在国际申请No. PCT/GB2015/050483中有描述。可以使用任何上述的聚合酶。所述聚合酶优选为Klenow或9o North。

[0291] 在聚合酶使用该模板多核苷酸作为模板形成经修饰的多核苷酸的情况下,使模板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,多核苷酸通常与可商购的聚合酶缓冲液中的聚合酶接触,例如来自New England Biolabs®的缓冲液。对于Klenow,温度优选为20至37°C,或者,对于9o North,温度优选为60至75°C。引物或3'发夹通常用作聚合酶延伸时的成核点。

[0292] 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。虽然期望在不同k聚体的电流测量值之间具有清楚的分离,但是通常这些测量中的一些存在重叠。特别是k聚体中具有很高数量的聚合物单元时,即具有非常高的k值时,变得难以分辨由不同的k聚体产生的用于确定关于多核苷酸的导出信息的测量值,所述导出信息例如对多核苷酸基本序列(underlying sequence)的估计。

[0293] 通过用经修饰的多核苷酸中的不同的核苷酸物质代替目标多核苷酸中的一个或多个核苷酸物质,所述经修饰的多核苷酸含有的k聚体不同于所述目标多核苷酸中的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目标多核苷酸中的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目标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目标多核苷酸。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,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。

#### [0294] 其他表征方法

[0295] 在另一个实施方案中,所述多核苷酸通过检测当聚合酶将核苷酸结合到多核苷酸中时所释放的标记物质进行表征。所述聚合酶使用所述多核苷酸为目标。每个标记物质对于每个核苷酸具有特异性。在步骤a)中,目标双链多核苷酸如以上所述而提供。在步骤b)中,将步骤a)中提供的多核苷酸与所述跨膜孔、聚合酶和标记的核苷酸接触,使得当核苷酸通过聚合酶被添加到多核苷酸时,磷酸盐标记的物质被依次释放,其中所述磷酸盐物质含有对每个核苷酸具有特异性的标记。所述聚合酶可以是任何上述论述的聚合酶。在步骤c)中,使用所述孔检测磷酸盐标记的物质,并由此表征所述目标双链多核苷酸。步骤b)和c)在欧洲专利申请No.13187149.3(公开为EP 2682460)中公开。上面论述的任何实施例同样适用于本方法。

#### [0296] 本发明的修饰方法

[0297] 本发明还提供了制备在表征方法的步骤a)中提供的经修饰的目标多核苷酸的方法。

[0298] 本发明提供了用于修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法。所述方法包括,将Y适配体接合到所述多核苷酸的一端,并将发夹环适配体接合到所述多核苷酸的另一端。如上所述,所述Y适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第一锚,所述发夹环适配体包括用于将所述多核苷酸耦合到所述膜的一个或多个第二锚,且将所述发夹环适配体耦合到膜的强度比将所述Y适配体耦合到膜的强度大。

[0299] 本发明提供了另一种替代方法,其中在所述适配体已接合到目标双链多核苷酸后,将所述锚连接到所述适配体。所述方法包括,将Y适配体接合到所述多核苷酸的一端,且将发夹环适配体接合到所述多核苷酸的另一端。然后将一个或多个第一锚连接到所述Y适配体,且将一个或多个第二锚连接到所述发夹环适配体。如上所述,发夹环适配体耦合到膜的强度比Y适配体耦合到膜的强度大。所述的一个或多个锚可以任何方式连接到适配体,包括上文所述的那些方式。所述的一个或多个锚优选通过杂交连接到所述适配体。

[0300] 可以使用任何接合方法。合适的方法在上文已公开。

[0301] Y适配体和发夹环适配体可以是任何上述论述的那些。

[0302] 基于MuA的方法

[0303] 本发明还提供了用于修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征的方法,包括将所述目标多核苷酸与MuA转座酶和一组双链MuA底物接触,其中所述组中底物的一部分是Y适配体,所述Y适配体包括将所述多核苷酸耦合到膜的一个或多个第一锚,其中所述组中底物的一部分是发夹环适配体,所述发夹环适配体包括将所述多核苷酸连接到所述膜的一个或多个第二锚,且其中所述发夹环适配体耦合到所述膜的强度比所述Y适配体耦合到所述膜的强度大,且由此产生多个经修饰的双链多核苷酸。

[0304] 本发明提供了另一种替代方法,其中在目标双链多核苷酸已被MuA转座酶片段化且所述底物已被接合到所述片段之后,所述锚被连接到所述底物。所述方法包括将所述目标多核苷酸与MuA转座酶和一组双链MuA底物接触,其中所述组中底物的一部分为Y适配体且其中所述组中底物的一部分是发夹环适配体。所述转座酶将目标多核苷酸片段化且将底物接合到该双链片段的一端或两端,并由此产生多个片段/底物构建体。转座酶优选将目标多核苷酸片段化且将底物接合到该双链片段的两端,并由此产生多个片段/底物构建体。转座酶优选产生多个片段/底物构建体,其各自包括位于一端的Y适配体和位于另一端的发夹环适配体。所述方法还包括向多个片段/底物构建体中的Y适配体连接一个或多个第一锚,和向多个片段/底物构建体中的发夹环适配体连接一个或多个第二锚,并由此产生多个经修饰的双链多核苷酸。每个发夹环适配体耦合到膜的强度比每个Y适配体耦合到膜的强度大。

[0305] 所述Y适配体和发夹环适配体可以是任何上述论述的那些。

[0306] 基于MuA使目标双链多核苷酸片段化在国际申请No. PCT/GB2014/052505 (公开为W02015/022544) 中公开。

[0307] 所述目标多核苷酸与MuA转座酶接触。这种接触发生在下述条件下:使转座酶能够发挥功能,即将目标多核苷酸片段化且将MuA底物接合到所述片段的一端或两端。MuA转座酶是可商购的,例如从Thermo Scientific商购(目录号F-750°C, 20 $\mu$ L (1.1 $\mu$ g/ $\mu$ L))。MuA转座酶起作用的条件在本领域中是已知的。

[0308] 使所述目标多核苷酸与一组双链MuA底物接触。所述双链底物是多核苷酸底物且可以由任何上文论述的核苷酸、多核苷酸或核酸形成。

[0309] 每个底物通常包括,能提供其作为MuA转座酶的底物的活性的双链部分。所述双链部分通常在每个底物中都是相同的。底物组可以包括不同的双链部分。每个底物优选包含双链部分,所述双链部分包含SEQ ID NO:26中示出的序列,所述序列杂交到与SEQ ID NO:26中示出的序列互补的序列。至少一个突出端优选位于与SEQ ID NO:26中所示的序列互补的序列的5'端。

[0310] 在一个优选的实施方案中,底物组中的每个底物包括通用核苷酸的至少一个突出端,使得所述转座酶将所述目标多核苷酸片段化且将底物接合到该双链片段的一端或两端,优选两端,且由此产生多个片段/底物构建体,且其中所述方法进一步包括将所述突出端接合到所述构建体中的片段并由此产生多个经修饰的双链多核苷酸。转座酶优选产生多个片段/底物构建体,其中每一个包括在一端的Y适配体和在一端的发夹环适配体。

[0311] 每个底物优选包括仅一个突出端。所述仅一个突出端是优选位于所述双链部分中的一条链的5'端。

[0312] 所述突出端可以是至少3个,至少4个,至少5个,至少6个或至少7个核苷酸长度。所述突出端优选为5个核苷酸长度。

[0313] 通用核苷酸是以某种程度杂交到目标多核苷酸中的所有核苷酸的核苷酸。合适的通用核苷酸在国际申请No.PCT/GB2014/052505(公开为W0 2015/022544)中描述。

[0314] 通用核苷酸的突出端(一个或多个)可以进一步包括反应性基团,其优选位于5'端。所述反应性基团可用于将所述突出端接合到构建体中的片段,如下所述。所述反应性基团可以使用点击化学(click chemistry)将所述片段接合到所述突出端。合适的反应性基团在国际申请No.PCT/GB2014/052505(公开为W0 2015/022544)中公开。

[0315] 在进一步的实施方案中,所述修饰方法使用MuA转座酶和一组MuA底物,其各自包括具有反应性基团的至少一个突出端。所述突出端(一个或多个)可以是任何长度并且可以包括任意个核苷酸(一个或多个)的任何组合。合适的长度和核苷酸在上面有公开。合适的反应性基团在上文有描述。

[0316] 在另一个实施方案中,所述方法包括将目标多核苷酸与一组双链MuA底物接触,其各自包含(i)至少一个突出端,和(ii)与所述至少一个突出端在相同的链中的至少一个核苷酸,其包括在目标多核苷酸中不存在的核苷,以使得所述转座酶将所述目标多核苷酸片段化且将底物接合到该双链片段的一端或两端,且由此产生多个片段/底物构建体。本领域技术人员可以如国际申请No.PCT/GB2014/052505(公开为W0 2015/022544)中所述来识别所述目标多核苷酸中不存在的核苷酸。转座酶优选将目标多核苷酸片段化且将底物接合到该双链片段的两端,并由此产生多个片段/底物构建体。所述转座酶优选产生多个片段/底物构建体,其各自包括在一端的Y适配体和在一端的发夹环适配体。所述突出端可以从所述构建体通过选择性地去除所述至少一个核苷酸而去除,以产生包括单链空位(gap)的多个双链构建体。构建体中的单链空位可以被修复以产生多个经修饰的双链多核苷酸。

[0317] 双链部分的一条链优选包含SEQ ID NO:26中所示的序列,且所述双链部分的另一条链优选包含这样的序列:其与SEQ ID NO:26中所示的序列互补,且其被修饰为,包括至少一个在目标多核苷酸中不存在的核苷酸。这种“另一条链”还包括突出端。在一个更优选的实施方案中,所述双链部分的一条链包含SEQ ID NO:26中示出的序列,且所述双链部分的另一条链包含SEQ ID NO:27中示出的序列(见下文)。在SEQ ID NO:27中,dC中的dA和在3'端的dA二核苷酸已经被dU取代。当目标多核苷酸包含脱氧腺苷(dA),胸苷(dT),脱氧鸟苷(dG)和脱氧胞苷(dC)而不包含脱氧尿苷(dU)时,可以使用该双链部分(如下所示)。

[0318] 5'-GTTTTTCGCATTTATCGTGAAACGCTTTTCGCGTTTTTCGTGCGCCGCTTCA-3'(SEQ 26)

[0319] 3'-CAAAAGCGTAAATAGCACTTTGCGAAAGCGCAAAAAGCACGCGGCGAAGU-5'(SEQ 27)

[0320] 在最优选的实施方案中,所述底物中的一条链包含SEQ ID NO:26中示出的序列,所述底物中的另一条链包含SEQ ID NO:28中所示的序列(见下文)。当目标多核苷酸包含脱氧腺苷(dA),胸苷(dT),脱氧鸟苷(dG)和脱氧胞苷(dC)而不包含脱氧尿苷(dU)时,可以使用该底物(下文示出)。

[0321] 5'-GTTTTTCGCATTTATCGTGAAACGCTTTTCGCGTTTTTCGTGCGCCGCTTCA-3'(SEQ 26)

[0322] 3'-CAAAAGCGTAAATAGCACTTTGCGAAAGCGCAAAAAGCACGCGGCGAAGUCTA G-5'(SEQ 28)

[0323] 在所有上述实施方式中,一种类型的底物的比例可以是任意比例,例如至少约5%,至少约10%,至少约20%,至少约30%,至少约40%,至少约50%,至少约60%,至少约

70%，至少约80%，至少约90%或至少约95%。底物组中底物的剩余比例通常是另一类型的底物的。例如，所述组可以包括40%的包含发夹环的底物和包括约60%的Y底物。这两种类型的底物的比例优选约为50%。

[0324] 用于将所述突出端接合到所述片段、将核苷酸(一个或多个)——其包含在目标多核苷酸中不存在的核苷酸——选择性地从接合的构建体中去除，和用于修复在双链构建体中的单链空位的方法在国际申请No. PCT/GB2014/052505 (公开为W02015/022544) 中公开。

#### [0325] 本发明的产品

[0326] 本发明还提供使用本发明的方法修饰的目标双链多核苷酸。经修饰的多核苷酸包括连接在多核苷酸一端的Y适配体和连接在另一端的发夹环适配体，其中所述Y适配体包括用于将所述多核苷酸耦合到膜的一个或多个第一锚，其中所述发夹环适配体包括用于将多核苷酸连接到膜的一个或多个第二膜，并且其中所述发夹环适配体耦合到所述膜的强度比所述Y适配体耦合到所述膜的强度大，且由此提供经修饰的目标双链多核苷酸。

[0327] 本发明还提供使用本发明的基于MuA的方法修饰的多个多核苷酸。这些经修饰的多核苷酸优选包括连接在一端的Y适配体和连接在另一端的发夹环适配体，其中所述Y适配体包括用于将多核苷酸耦合到膜的一个或多个第一锚，其中所述发夹环适配体包括用于将多核苷酸耦合到膜的一个或多个第二锚，且其中所述发夹环适配体耦合到膜的强度比所述Y适配体耦合到膜的强度大，且由此提供经修饰的目标双链多核苷酸。一些多核苷酸可以如上所述在两端具有Y适配体或如上所述在两端具有发夹环适配体。

[0328] 本发明还提供了一对适配体，用于修饰目标双链多核苷酸以使用膜中的跨膜孔进行表征，其中，一个适配体是Y适配体，该Y适配体包括用于将多核苷酸耦合到隔膜的一个或多个第一锚，其中另一个适配体是发夹环适配体，该发夹环适配体包括用于将多核苷酸耦合到膜的一个或多个第二锚，并且其中发夹环适配体耦合到膜的强度比Y适配体耦合到膜的强度大。

[0329] 本发明还提供了适配体组，用于修饰目标多核苷酸以使用膜中的跨膜孔进行表征，其中一部分适配体是Y适配体，其包含用于将多核苷酸耦合到膜的一个或多个第一锚，其中一部分适配体是发夹环适配体，其包含将多核苷酸耦合到膜的一个或多个第二锚，且其中所述发夹环适配体耦合到所述膜的强度比所述Y适配体耦合到所述膜的强度大。

[0330] 本发明适配体对或组中的每个适配体优选包括双链MuA底物。所述底物可以是任何上文描述的底物。所述底物优选包含如上文所述的双链部分。所述双链部分优选包括如上所述的SEQ ID NO:26和SEQ ID NO:27。所述双链部分更优选包括如上所述的SEQ ID NO:26和SEQ ID NO:28。

[0331] 上文本发明方法中给出的比例值同样适用于本发明的所述组。

[0332] 上文参照本发明的方法所述的任何实施方式同样适用于本发明的多核苷酸、对和组。

[0333] 所述组或多个可以被分离、基本上分离、纯化或基本上纯化。组或多个可以被分离或纯化，如果它完全不含任何其它组分，例如目标多核苷酸、脂质或孔。组或多个被基本上分离，如果它与不干扰其预期用途的载体或稀释剂混合。例如，组或多个被基本上分离或基本上纯化，如果它以包含小于10%，小于5%，小于2%或小于1%的其它组分如脂质或孔的形式存在。

### [0334] 试剂盒

[0335] 本发明还提供了修饰目标多核苷酸的试剂盒,所述试剂盒包括(a)本发明的一对适配体或本发明的一组适配体,和(b) MuA转座酶。所述对或组中的每个适配体优选包括双链MuA底物。

[0336] 上文参照本发明的方法和产品所述的任何实施方式同样适用于所述试剂盒。

[0337] 所述试剂盒可以进一步包括膜的组件,如两亲性层或三嵌段共聚物膜的组件。所述试剂盒可以进一步包括跨膜孔或跨膜孔的组件。所述试剂盒可以进一步包括多核苷酸结合蛋白。合适的膜、孔和多核苷酸结合蛋白质在上文已述。

[0338] 本发明的试剂盒可以额外包括一个或多个能实施上述任何实施方案的其它试剂或仪器。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,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地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患者的详细说明。所述试剂盒可以可选地包括核苷酸。

[0339] 下列实施例用于说明本发明。

### [0340] 实施例1

[0341] 本实施例描述了在图1和图2中示出的且在实施例2和3中使用的DNA构建体1-6的样品制备过程。

### [0342] 材料和方法

[0343] 在这项研究中使用的链来自 $\lambda$ 基因组的45042bp和48487bp之间的区域。分析物通过聚合酶PCR法制备以按需要在模板和互补模板中每一个的限定端包括杂交位点。PCR从 $\lambda$ 基因组DNA实施。

[0344] 本模板(SEQ ID NO:29,杂交到SEQ ID NO:30)使用KAPA HiFi 2x主混合物、 $\lambda$ DNA(NEB)和引物SEQ ID NO:31和SEQ ID NO:32制备。使反应循环20次,将正确尺寸的产品通过在Sephacryl S1000柱上凝胶过滤而纯化,并使用Millipore Ultracel 1550kDa浓缩器浓缩至0.25mg/ml。

[0345] 用于电生理学实验的DNA构建体(1,2,3,4,5和6)均是根据相同的反应混合物制备:2 $\times$ LongAmp Taq主混合物,300nM的引物1和2或3和4(在图1和图2中标记为a1的链使用引物1和引物2制备,其中引物1=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,所述间隔区在相对端连接到SEQ ID NO:34的5'端,SEQ ID NO:34的3'端连接到4个iSpC3间隔区,所述间隔区连接到4个5-硝基吡啶,5-硝基吡啶在相对端连接到SEQ ID NO:36的5'端,引物2=SEQ ID NO:37,其3'端连接到4个5-硝基吡啶,5-硝基吡啶在相对端连接到SEQ ID NO:38的5'端;图1和图2中标记为a2的链使用引物3和引物4制备,其中引物3=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,所述间隔区在相对端连接到SEQ ID NO:37的5'端,SEQ ID NO:37的3'端连接到4个iSpC3间隔区,所述间隔区连接到4个5-硝基吡啶,5-硝基吡啶在相对端连接到SEQ ID NO:38的5'端,引物4=SEQ ID NO:34,其3'端连接到4个5-硝基吡啶,5-硝基吡啶在相对端连接到SEQ ID NO:36的5'端)和1.2ng  $\mu$ l<sup>-1</sup> DNA模板(SEQ ID NO:29,杂交到SEQ ID NO:30)。DNA构建体全部根据相同的循环程序扩增:94 $^{\circ}$ C 2分钟,[94 $^{\circ}$ C 15秒,58 $^{\circ}$ C 30秒,65 $^{\circ}$ C 2分钟]<sub>12</sub>和65 $^{\circ}$ C 5分钟。DNA构建体全部根据制造商的说明(Qiagen凝胶提取试

剂盒)从0.8%琼脂糖凝胶中纯化,然后根据制造商的说明进行SPRI纯化(Agencourt AMPure小珠)。

[0346] 最后,所产生的DNA链被杂交到DNA的互补链,其中一些含有锚例如胆固醇或棕榈酸酯。将互补链SEQ ID NO:39和SEQ ID NO:40(与锚连接和与不锚连接)在25mM磷酸钾缓冲液、151mM氯化钾、pH 8.0中在超过室温5倍的温度下退火十分钟。

[0347] 如1(A)中所示,DNA构建体1是由4个不同链杂交在一起构成的——a1=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:34的5'端;SEQ ID NO:34在3'端连接到4个iSpC3间隔区的3'端,该iSpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:35的5'端,a3=SEQ ID NO:40,其3'端连接到6个iSp18间隔区,该iSp18间隔区的相对端连接到2个胸腺嘧啶,胸腺嘧啶的相对端连接到3'胆固醇TEG,a4=SEQ ID NO:39,a5=与a3互补且与a1的一部分互补的序列。

[0348] 如图1(A)和(B)中所示,DNA构建体2是由4个不同链杂交在一起构成的——a2=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区相对端连接到SEQ ID NO:37的5'端;SEQ ID NO:37的3'端连接到4个iSpC3间隔区,iSpC3间隔区相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:41的5'端,a6=SEQ ID NO:39,其3'端连接到6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'胆固醇TEG,a7=SEQ ID NO:40,a8=与a6互补且与a2的一部分互补的序列。

[0349] 如图1(B)和2(B)中所示,DNA构建体3是由4个不同链杂交在一起构成的——a1=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:34的5'端;SEQ ID NO:34的3'端连接到4个iSpC3间隔区,iSpC3间隔区的相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:35的5'端,a3=SEQ ID NO:40,其3'端连接到6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'胆固醇TEG,a9=SEQ ID NO:39,其3'端连接到6个iSp18间隔区,iSp18间隔区相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'棕榈酸酯,a5=与a3互补且与a1的一部分互补的序列。

[0350] 如图2(A)中所示,DNA构建体4是由4个不同链杂交在一起构成的——a1=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:34的5'端;SEQ ID NO:34的3'端连接到4个iSpC3间隔区,iSpC3间隔区相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:35的5'端,a3=SEQ ID NO:40,其3'端连接有6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'胆固醇TEG,a6=SEQ ID NO:39,其3'端连接到6个iSp18间隔区,iSp18间隔区相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'胆固醇TEG,a5=与a3互补且与a1的一部分互补的序列。

[0351] 如图2(A)中所示,DNA构建体5是由4个不同链杂交在一起构成的——a1=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:34的5'端;SEQ ID NO:34在3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:35的5'端,a3=SEQ ID NO:40,其3'端连接到6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶在相对端连接到3'胆固醇TEG,a2=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:37的5'端;SEQ ID NO:37在3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:41的5'端,a6=SEQ ID NO:39,其3'端连接到

6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶在相对端连接到3'胆固醇TEG。

[0352] 如图2(B)中所示,DNA构建体6是由4个不同链杂交在一起构成的——a1=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:34的5'端;SEQ ID NO:34在3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:35的5'端,a10=SEQ ID NO:40,其3'端连接到6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶在相对端连接到3'棕榈酸酯,a2=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:37的5'端;SEQ ID NO:37在3'端连接到4个iSpC3间隔区,SpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:41的5'端,a9=SEQ ID NO:39,其3'端连接到6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶,胸腺嘧啶在相对端连接到3'胆固醇TEG。

### [0353] 实施例2

[0354] 本实施例将使用单个锚将构建体耦合到膜和使用不同强度的两个锚将构建体耦合到膜进行比较。为了使解旋酶控制的DNA(通过纳米孔系统检测)移动偏向双耦合到膜的构建体,使用了两个锚。

### [0355] 材料和方法

[0356] 在设置实验之前,将DNA构建体1和2或构建体2和3(原料浓度为20nM,加入到纳米孔系统中的最终浓度0.1nM)分别用在缓冲液(151mM KCl,25mM磷酸,2mM EDTA,pH8.0)中的T4 Dda-E94C/A360C(原料浓度250nM,加入到纳米孔系统中的最终浓度1nM,具有突变E94C/A360C的SEQ ID NO:24)在室温下预培育5分钟。五分钟后,将TMAD(500 $\mu$ M)加入到预混合物中且再培育混合物5分钟。最后,将MgCl<sub>2</sub>(10mM最终浓度),ATP(2.5mM最终浓度)和缓冲液(150mM的亚铁氰化钾(II),150mM铁氰化钾和25mM磷酸钾pH 8.0)加入到预混合物中。

[0357] 从在缓冲液(25mM磷酸钾,150mM亚铁氰化钾(II),150mM铁氰化钾(III),pH 8.0)中插入嵌段共聚物中的单个MspA纳米孔(MspA-B2C)获得电测量值。在实现将单个孔插入嵌段共聚物中之后,使缓冲液(2mL,25mM磷酸钾pH 8.0,150mM的亚铁氰化钾(II)和150mM铁氰化钾(III))流经系统以去除任何多余的MspA纳米孔。然后将酶(T4 Dda-E94C/A360C,1nM的最终浓度)、DNA构建体1和2或2和3中(0.1nM的最终浓度)、燃料(MgCl<sub>2</sub> 10mM最终浓度,ATP 2.5mM最终浓度)的预混合物(总计150 $\mu$ L)加入到单个纳米孔实验系统,并在120mV的保持电势下进行实验2小时,且监测解旋酶控制的DNA移动。

### [0358] 结果与讨论

[0359] 对于对照实验,监测构建体1和构建体2的解旋酶控制的DNA移动。图3上部的卡通显示,其中解旋酶可以结合到构建体1和构建体2。对应于每个构建体的解旋酶控制的DNA移动被识别,且将对应于链a1(构建体1)和链a2(构建体2)的移动的比例进行比较(见图3的下半部分)。当使用相同的锚(胆固醇锚)将每个构建体耦合到膜,然后观察解旋酶控制的DNA移动时,大约50%对应于链a1和50%对应于链a2。这表明,使用相同的单个锚时,不存在相比于一个构建体,偏向于另一个构建体的解旋酶控制的移位运动。

[0360] 实验还对使用单个锚(胆固醇)将构建体2耦合到膜与使用两个不同的锚(不同强度——胆固醇和棕榈酸酯)将构建体3耦合到所述膜进行了比较。图4上部的卡通表明了其

中解旋酶可以结合到链a1(构建体3)和链a2(构建体2)。识别对应于每个构建体的解旋酶控制的DNA的移动,并将对应于a1和a2(分别为构建体3和构建体2)的动的比例进行了比较(参见图4的下半部分)。在这种情况下,检测到,多于95%的解旋酶控制的DNA移动对应于使用胆固醇和棕榈酸酯耦合到膜的链a1(构建体3)。小于5%的解旋酶控制的DNA移动对应于链a2(构建体2)。该实验表明,由纳米孔系统检测到,使用两个不同的锚时,与单一耦合的构建体相比,解旋酶控制的DNA移动强烈的偏向双重耦合的构建体。

#### [0361] 实施例3

[0362] 本实施例比较了使用相等强度的两个锚(双胆固醇构建体4和构建体5或双棕榈酸酯构建体6)与使用不同强度的两个锚(棕榈酸酯和胆固醇构建体3)。

#### [0363] 材料和方法

[0364] 将DNA构建体4和5或构建体3和6如上文实施例2中所述使用T4 Dda-E94C/A360C解旋酶预培育。

[0365] 如实施例2中所述从在缓冲液中插入嵌段共聚物中的单个MspA纳米孔(MspA-B2C),获得电测量值,并监测对于构建体4和构5或构建体3和6的解旋酶控制的DNA移动。

#### [0366] 结果与讨论

[0367] 对于对照实验,监测构建体4和构建体5的解旋酶控制DNA的移动。图5上部的卡通显示,其中解旋酶可以结合到构建体4(仅在a1上)和构建体5(在a1和a2上)。构建体4的链a1能够被纳米孔捕获,而构建体5的链a1或链a2能够被纳米孔捕获。对应于每个构建体的解旋酶控制的DNA移动被识别,且将对应于链a1(构建体4和5)和链a2(仅构建体5)的动的比例进行比较(见图5的下半部分)。当使用相同的锚(两个胆固醇锚)将每个构建体耦合到膜,然后观察解旋酶控制的DNA移动时,大约66%对应于链a1(构建体4和5),33%对应于链a2(仅构建体5)。可以预期到,2:1偏向于对a1解旋酶控制的DNA移动的检测,因为a2移动仅在构建体4被纳米孔捕获时被检测到,而a1移动能从构建体4和构建体5的捕获中被检测到,因此,可以预期到两倍的a1移动(在当量浓度的构建体4和5被加入到纳米孔实验系统中时)。

[0368] 该实验还比较了使用不同强度的两种耦合剂(胆固醇和棕榈酸酯)将构建体3耦合到膜上,与使用两种相同的耦合剂(均为棕榈酸酯)将构建体6耦合到膜上。图6上部的卡通表明,其中解旋酶可以结合到链a1(构建体3和6)和链a2(仅构建体6)。构建体3的链a1能够被纳米孔捕获,而构建体6的链a1或a2能够被纳米孔捕获。对应于每个构建体的解旋酶控制的DNA移动被识别,且将对应于链a1(构建体2和3)和链a2(仅构建体2)的动的比例进行比较(见图6的下半部分)。在这种情况下检测到,多于95%的解旋酶控制的DNA移动对应于使用胆固醇和棕榈酸酯或两个棕榈酸酯耦合到所述膜的链a1(构建体3和6)。小于5%的解旋酶控制的DNA移动对应于链a2(仅构建体6)。该实验说明,可以由纳米孔系统检测到,通过使用两种不同的耦合剂,有可能使解旋酶控制的DNA移动强烈偏向由不同强度的耦合剂耦合的构建体而不是由相同强度的两种相同耦合剂耦合的构建体。这也表明,在仅一端被纳米孔捕获的构建体上使用较强和较弱锚的组合时,可能偏向于该构建体,而不是偏向于在两端被纳米孔捕获的构建体上具有两个较弱锚的构建体。

#### [0369] 实施例4

[0370] 本实施例说明了如何比较大量耦合剂的相对强度。这使得能选择合适强度的耦合剂将构建体耦合到膜上。

### [0371] 材料和方法

[0372] 使用实施例1中所述的方法制备DNA构建体1、构建体2和构建体7(a-c)。构建体1和构建体2在上文有详细描述,构建体7(a-c)在下文进行描述。

[0373] 如图7(B)中所示,DNA构建体7a是由4个不同链杂交在一起构成的——a2=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:37的5'端;SEQ ID NO:37在3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:41的5'端,a10(a)=SEQ ID NO:42,其5'端连接到6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶和5'胆固醇TEG,SEQ ID NO:42具有连接到其3'端的另外6个iSp18间隔区,iSp18间隔区在相对端连接到2个胸腺嘧啶和3'胆固醇TEG,a7=SEQ ID NO:40,a8=与a2的一部分和A10a互补的序列。

[0374] 如图7(B)中所示,DNA构建体7b是由4个不同链杂交在一起构成的——a2=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到SEQ ID NO:37的5'端;SEQ ID NO:37在3'端连接到4个iSpC3间隔区,iSpC3间隔区在相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:41的5'端,a10(b)=SEQ ID NO:39,其3'端连接到6个iSp18间隔区,iSp18间隔区相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'生育酚,a7=SEQ ID NO:40,a8=与a2的一部分和A10b互补的序列。

[0375] 如图7(B)中所示,DNA构建体7c是由4个不同链杂交在一起构成的——a2=SEQ ID NO:33,其3'端连接到4个iSpC3间隔区,iSpC3间隔区相对端连接到SEQ ID NO:37的5'端;SEQ ID NO:37的3'端连接到4个iSpC3间隔区,iSpC3间隔区相对端连接到4个5-硝基吡啶,5-硝基吡啶连接到SEQ ID NO:41的5'端,a10(c)=SEQ ID NO:39,其3'端连接到6个iSp18间隔区,iSp18间隔区相对端连接到2个胸腺嘧啶,胸腺嘧啶相对端连接到3'棕榈酸酯,a7=SEQ ID NO:40,a8=与a2的一部分和A10c互补的序列。

[0376] DNA构建体1、构建体2和构建体7(a-c)如上文实施例2中所述使用T4 Dda-E94C/A360C解旋酶预培育。

[0377] 如实施例2中所述从在缓冲液中插入嵌段共聚物中的单个MspA纳米孔(MspA-B2C)获取电测量值,且检测对于构建体1、2和7(a-c)的解旋酶控制的DNA移动。

### [0378] 结果与讨论

[0379] 对于对照实验,监测构建体1和构建体2的解旋酶控制的DNA的移动。图8上部的卡通显示,其中解旋酶可以结合到构建体1和构建体2。对应于每个构建体的解旋酶控制的DNA移动被识别,且将对应于链a1(构建体1)和链a2(构建体2)的移动的比例进行比较(见图8的下半部分)。当使用相同的耦合剂(胆固醇)将每个构建体耦合到膜时,观察到,解旋酶控制的DNA移动中,大约50%对应于链a1,50%对应于链a2。这表明,使用相同的单个耦合剂时,不存在相对于一个构建体,偏向另一个构建体的解旋酶控制的移位运动。

[0380] 该实验还对使用多种不同耦合剂将构建体7(a-c)耦合到膜与使用胆固醇将构建体1耦合到膜进行了比较。图9-11上部的卡通显示,其中解旋酶可以结合到链a1(构建体1)和链a2(构建体7(a-c))。对应于每个构建体的解旋酶控制DNA的移动被识别,并将对应于a1和a2(分别为构建体1和构建体7(a-c))的移动的比例进行比较(参见图9-11的下半部分)。在相同的DNA链(A10a)中使用两个胆固醇时观察到,其具有与使用单个胆固醇时相同的耦合强度(见图9)。而使用生育酚时观察到,所述生育酚是比胆固醇强的耦合剂(见图10),因

为通过生育酚锚定构建体时比通过胆固醇锚定构建体时观察到较大量的DNA控制的解旋酶移动。此外,观察到棕榈酸酯是比胆固醇弱的耦合剂,因为通过棕榈酸酯锚定构建体时比通过胆固醇锚定构建体时观察到较少的解旋酶控制的DNA移动(见图11)。因此,通过使用本方法比较各种耦合剂的耦合强度,可以选择两种不同强度的耦合剂和使系统偏向于选择所希望的构建体(进一步详细内容见实施例6)。

#### [0381] 实施例5

[0382] 本实施例描述了图12中所示的和实施例6中所用的DNA构建体8和构建体9的样品制备过程。

#### [0383] 材料和方法

[0384] 将λDNA (1μg, SEQ ID NO:43) 用5U的SnaBI和20U的BamHI-HF在1×NEBuffer 4在37℃下进行限制性酶消化1小时。消化以后,根据制造商的使用说明,使用1.5x Agencourt AMPure小珠将DNA纯化。然后将样品在50uL无核酸酶的水中洗脱。然后遵循制造商的使用说明,使用NEB的NEBNext端部修复系统对DNA进行末端修复。将样品使用1.5x Agencourt AMPure小珠再次进行纯化,然后在42uL无核酸酶的水中进行DNA洗脱。末端修复步骤以后,根据制造商的使用说明,将纯化的DNA使用NEB的NEBNext dA-拖尾系统进行dA拖尾(dA-tailed)。将DNA使用1.5x Agencourt AMPure小珠再次纯化,在20uL无核酸酶的水中洗脱。遵循制造商的使用说明,使用NEB的NEBNext快速连接酶模块,将所述DNA在A260nm进行定量然后以对每个5'端过量25倍实施适配体接合。构建体8通过将以下适配体接合到DNA的片段(图12中标记为a12)而制备(适配体1(图12中标记为a11)=25个iSpC3间隔区,其连接到SEQ ID NO:44的5'端,SEQ ID NO:44的3'端连接4个iSp18间隔区,所述间隔区相对端连接到SEQ ID NO 45的5'端,SEQ ID NO 45的3'端连接到序列AACCT(其通过硫代磷酸酯(phosphorothioate)键连接在一起),其中所述序列中的最后的T被连接到另外的未修饰的T;适配体2(图12中标记为a14)=在序列GGTT(其通过硫代磷酸酯键连接在一起)的5'端具有磷酸基团,该序列3'端被连接到SEQ ID NO:46的5'端,SEQ ID NO:46的3'端连接到6个iSp18间隔区,iSp18间隔区相对端连接到两个胸腺嘧啶,胸腺嘧啶相对端连接到3'胆固醇TEG,适配体3(图12中标记为a13)=SEQ ID NO:47的5'端具有磷酸基团,且其具有内部胆固醇TEG连接到SEQ ID NO:47的3'端,所述内部胆固醇还连接到SEQ ID NO:48的5'端,SEQ ID NO:48的3'端用硫代磷酸酯键连接到G和未修饰的T)。构建体9通过将以下适配体接合到DNA的片段(图12中标记为a12)而制备(适配体1(图12中标记为a11)=25 iSpC3间隔区,其5'端连接到SEQ ID NO:44,SEQ ID NO:44的3'端连接到4个iSp18间隔区,iSp18间隔区相对端连接到SEQ ID NO:45的5'端,SEQ ID NO:45的3'端连接到序列AACCT(其通过硫代磷酸酯键连接在一起),其中,所述序列中的最后的T被连接到另一个未修饰的T;适配体4(图12中标记为a15)=在序列GGTT(其通过硫代磷酸酯键连接在一起)的5'端具有磷酸基团,该序列的3'端连接到SEQ ID NO:46的5'端,SEQ ID NO:46的3'端连接到6个iSp18间隔区,所述间隔区相对端连接到两个胸腺嘧啶,胸腺嘧啶相对端连接到3'棕榈酸酯,适配体3(图12中标记为a13)=SEQ ID NO:47的5'端具有磷酸基团,且其具有内部胆固醇TEG连接到SEQ ID NO:47的3'端,所述内部胆固醇也连接到SEQ ID NO:48的5'端,SEQ ID NO:48的3'端用硫代磷酸酯键连接到G和未修饰的T)。消化以后,将接合的DNA使用0.4x Agencourt AMPure小珠纯化,在16uL无核酸酶的水中洗脱。所述DNA在A260nm下进行定量然后添加757.5mM KCl,

125mM磷酸钾缓冲液(pH 7)和5mM EDTA。

#### [0385] 实施例6

[0386] 本实施例比较了针对构建体8和构建体9检测到的解旋酶控制的DNA移动。将构建体8使用两个胆固醇(二者为强耦合剂)耦合到膜。将构建体9使用发夹环中的胆固醇和Y适配体中的棕榈酸酯耦合到膜。所述构建体均含有由上文所述的方法片段化的 $\lambda$ DNA。

#### [0387] 材料和方法

[0388] 将DNA构建体8和9分别用T4 Dda-E94C/A360C/C109A/C136A解旋酶按与上述实施例2中所述类似的方法预温育。

[0389] 如实施例2中所述从在缓冲液中插入嵌段共聚物中的单个MspA纳米孔(MspA-B2C)获取电测量值,且监测对于构建体8和构建体9的解旋酶控制的DNA移动。

#### [0390] 结果与讨论

[0391] 在片段化和适配体连接过程中(实施例5中所述),制备图12中所示的所需构建体。然而,连接两个Y适配体(构建体11和构建体12)或连接两个发夹环适配体(构建体10)的过程中,还制备了其它构建体(见图13和图14)。为了使纳米孔能够捕获DNA,所述DNA必须具有至少一个Y适配体,使得有游离的末端被纳米孔捕获。因此,连接有两个发夹环的DNA构建体(构建体10)在纳米孔实验中不能被观察到。在 $\lambda$ DNA片段的两端上具有Y适配体的构建体能够被纳米孔通过DNA(构建体11和构建体12)的两端而捕获。但是,因为没有发夹环连接到所述DNA,因此仅链x或链y移位穿过纳米孔。然而,如果构建体8或构建体9被纳米孔(经Y适配体端)捕获,则链X和链Y两者由于连接至相对端的发夹环而移位穿过纳米孔。因此,希望的是,含有Y适配体和发夹的构建体优选被纳米孔捕获。为了仅分离所需的构建体8和构建体9而分离出各个构建体(10、8和11,或10、9和12),这是不可能。

[0392] 对含有Y适配体和发夹适配体的构建体是否可以通过使用耦合剂的不同组合被纳米孔优先选择和捕获进行了研究。在本实施例中,比较了两个胆固醇的耦合(构建体8)与胆固醇和棕榈酸酯的耦合(构建体9)。测试的这两种构建体展出发夹/Y适配体构建体(8和9)或双Y适配体构建体(11和12)的大数量的解旋酶控制的DNA移动。当在Y适配体和发夹二者(构建体8)中使用相同的耦合剂(胆固醇)时,检测到25%的解旋酶控制的DNA移动对应于构建体8且75%的解旋酶控制的DNA移动对应于构建体11。因此,检测到,只有四分之一的移动对应于所希望的构建体。当在发夹中使用强耦合剂(在这种情况下为胆固醇)且在Y适配体中使用较弱的耦合剂(在这种情况下为棕榈酸酯)时,检测到,46%的解旋酶控制的DNA移动对应于构建体9且56%的解旋酶控制的DNA移动对应于构建体12。这是对选择所需构建体9的显著改进。在检测所需构建体中的这一改进是由于使用了两种不同强度的耦合剂。将使用两种较弱耦合剂耦合双Y适配体构建体12到膜与使用一个弱锚和一个强锚耦合Y适配体/发夹构建体9到膜进行比较。由于所需构建体采用了一个比双Y适配体构建体中使用的那些锚强的锚,因此其优先耦合到膜上,且因此检测到,更大比例的解旋酶控制的DNA移动对应于所需的构建体9。

#### [0393] 实施例7

[0394] 本实施例比较了通过若干手段耦合到膜的许多不同DNA构建体13、14、15和16的耦合效率。构建体13使用发夹区域中的胆固醇和Y适配体区域中的由刚性DNA连接体杂交的棕榈酸酯耦合到膜。构建体14使用发夹区域中的胆固醇和Y适配体区域中的由柔性DNA连接体

杂交的棕榈酸酯耦合到膜。构建体15使用发夹区域中的胆固醇和Y适配体区域中的由刚性DNA连接体杂交的胆固醇耦合到膜。构建体16使用发夹区域中的胆固醇和Y适配体区域中的由柔性DNA连接体杂交的胆固醇耦合到所述膜。

#### [0395] 材料和方法

[0396] 所有构建体包含已通过上文实施例5中所述的方法片段化的 $\lambda$ DNA片段(图15和图16中标记为a12),然而,不同的适配体然后被接合到各DNA构建体。所述适配体接合而形成DNA构建体13-16(在图15和16中示出)将在下面详细描述。

[0397] 构建体13、14、15和16通过将下列各适配体接合到 $\lambda$ DNA的片段(在图15和图16中标记为a12)而制备(适配体1(图15中标记为a16)=30个iSpC3间隔区,其连接到SEQ ID NO:51的5'端,SEQ ID NO:51的3'端连接到4个iSp18间隔区,所述iSp18间隔区的相对端连接到SEQ ID NO:52的5'端;适配体2(图15中标记为a17)=在SEQ ID NO:49的5'端具有磷酸基团,该序列的3'端连接到4个iSpC3间隔区,所述iSpC3间隔区的相对端连接到SEQ ID NO:50的5'端,所述SEQ ID NO:50的3'端经硫代磷酸酯键连接到胸腺嘧啶;适配体3(图15中标记为a18)=SEQ ID NO:53的5'端具有磷酸基团。

[0398] 在按实施例5中所述纯化以上构建体后,将5'端连接有棕榈酸酯或胆固醇的刚性或柔性DNA连接体在以下条件下杂交到所述构建体——将0.2 $\mu$ l的每个DNA连接体(10 $\mu$ M)加入到13 $\mu$ l接合的DNA(原液浓度约为20-30nM)中,并在室温下培养10分钟。

[0399] 构建体13杂交到适配体4(在图15中标记为a20)和适配体5(在图15中标记为a19),其中适配体4=SEQ ID NO:40,其5'端连接到6个iSp18间隔区,所述iSp18间隔区的相对端连接到两个胸腺嘧啶和5'胆固醇TEG,适配体5=SEQ ID NO:54,其5'端连接到棕榈酸酯。

[0400] 构建体14杂交到适配体4(在图15中标记为a20)和适配体6(在图15中标记为a21),其中适配体4=SEQ ID NO:40,其5'端连接到6个iSp18间隔区,所述iSp18间隔区的相对端连接到两个胸腺嘧啶和5'胆固醇TEG,适配体6=SEQ ID NO:54,其5'端连接到6个iSp18间隔区,所述iSp18间隔区的相对端连接两个胸腺嘧啶和棕榈酸酯。

[0401] 构建体15杂交到适配体4(图15中标记为a20)和适配体5(在图15中标记为a22),其中适配体4=SEQ ID NO:40,其5'端连接到6个iSp18间隔区,所述iSp18间隔区的相对端连接到两个胸腺嘧啶和5'胆固醇TEG,适配体5=SEQ ID NO:54,其5'端连接到胆固醇TEG。

[0402] 构建体16杂交到适配体4(在图15中标记为a20)和适配体6(在图15中标记为a23),其中适配体4=SEQ ID NO:40,其5'端连接到6个iSp18间隔区,所述iSp18间隔区的相对端连接到两个胸腺嘧啶和5'胆固醇TEG,适配体6=SEQ ID NO:54,其5'端连接到6个iSp18间隔区,所述iSp18间隔区的相对端连接两个胸腺嘧啶和胆固醇TEG。

[0403] 在设置该实验之前,将DNA构建体13-16(加入到纳米孔系统的最终浓度0.1nM)使用在缓冲液(151mM KCl,25mM磷酸盐,5%甘油,pH7.0)中的T4Dda-E94C/A360C/C109A/C136A(加入到纳米孔系统的最终浓度为10nM)在室温下独立地预培育5分钟。五分钟后,将TMAD(500 $\mu$ M)加入到预混合物中,且再培育混合物5分钟。最后,MgCl<sub>2</sub>(2mM最终浓度),ATP(2.5mM最终浓度)和缓冲液(500mM KCl和25mM磷酸钾pH 8.0)加入到预混合物中。

[0404] 如实施例2中所述从在缓冲液中插入嵌段共聚物中的单个MspA纳米孔(MspA-B2C)获取电测量值,并监测构建体13-16的解旋酶控制的DNA移动。

#### [0405] 结果与讨论

[0406] 本实施例通过下述比较了多个不同DNA构建体13、14、15和16的耦合效率:观察研究在试验过程中每个纳米孔的解旋酶控制的DNA移动的数目,以及与检测到的所有其他解旋酶控制的DNA移动相比时,这些解旋酶控制的移动——对应于区域R1和R2的移位——的百分比。

[0407] 观测了对于所测试的所有这四个不同构建体的解旋酶控制的DNA移动。在这个实验中,我们比较了两种不同类型的锚——棕榈酸酯和胆固醇的强度。还可以比较,两种不同类型的双链多核苷酸连接体的柔性是如何影响耦合的强度的。在所有4个构建体中,在发夹a20中的耦合剂(柔性胆固醇系链)保持不变,Y适配体中使用的耦合剂在棕榈酸酯/胆固醇和柔性/刚性连接体之间变化。

[0408] 本实验中使用的构建体是通过如实施例5中所述的相同的片段化和适配体连接方法制备的。这制得在图15和16中示出的所需构建体以及在图13和14中所示构建体的构建体类似物,它们具有两个连接的Y适配体或两个连接的发夹适配体。如在上文实施例6详细描述,为了使纳米孔能够捕获DNA,DNA必须已具有至少一个Y适配体以使其具有游离端能够被所述孔捕获。因此,在纳米孔实验中观察不到连接有两个发夹的DNA构建体。在DNA片段的两端上具有Y适配体的构建体通过所述DNA的两端被纳米孔捕获。但是,当没有发夹连接到DNA时,仅R1或R2被移位穿过纳米孔。然而,如果所需的构建体13-16被纳米孔捕获(通过Y适配体端),则区域R1和R2将都移位通过纳米孔。因此,所需的是含有Y适配体和发夹的构建体优先地被纳米孔捕获。

[0409] 图17所示的数据说明,所检测到的解旋酶控制的DNA移动总数的%,其对应于R1和R2区域穿过纳米孔的移位(从含有Y适配体和发夹适配体的所需构建体制备的)。观察到,其他解旋酶控制的DNA移动对应于下述构建体的移位:其具有两个Y适配体而不是Y适配体和发夹环适配体,连接到λDNA的片段。当棕榈酸酯与刚性或柔性连接体(构建体13和构建体14)一起使用时,R1和R2区域的解旋酶控制DNA移动穿过纳米孔(并对应于所需构建体)的最高百分比被观察到。具有刚性系链的胆固醇比具有柔性系链的胆固醇具有更高的解旋酶控制的DNA移动的百分比,其对应于R1和R2穿过纳米孔的移位。

[0410] 从图18所示数据也可清楚,对于棕榈酸酯(构建体13和构建体14)和胆固醇(构建体15和构建体16),刚性排列是比柔性排列较弱的耦合结构,因为构建体13和构建体15导致对每个纳米孔检测到较少的解旋酶控制的DNA移动。测试到的最弱的耦合排列是在其中使用刚性连接体连接到棕榈酸酯的构建体15中。

[0411] 优选的耦合系统是这样的耦合系统:其引起每个纳米孔具有较高数量的解旋酶控制的DNA移动,其也具有最高百分比的解旋酶控制的DNA移动——对应于区域R1和区域R2的移位。实验表明,对于R1和R2的最高%的解旋酶控制的DNA移动导致每个纳米孔的最低数量的解旋酶控制的DNA移动,反之亦然。因此,对于所需的解旋酶控制的DNA移动的最佳整体吞吐量,优选的构建体设计是这样的构建体15:其每个纳米孔具有大量的解旋酶控制的DNA移动,其中解旋酶控制的DNA移动——对应于R1和R2的移位——的百分比刚好在50%以下。

## 序列表

<110> 牛津纳米孔技术公司

<120> 方法

<130> N402116W0

<150> GB1406147.7

<151> 2014-04-04

<150> GB1407815.8

<151> 2014-05-02

<150> PCT/GB2014/052737

<151> 2014-09-10

<160> 54

<170> PatentIn version 3.5

<210> 1

<211> 558

<212> DNA

<213> 人工序列

<220>

<223> 耻垢分枝杆菌 (*Mycobacterium smegmatis*) 孔蛋白A突变体  
(D90N/D91N/D93N/D118R/D134R/E193K)

<400> 1

```

atgggtctgg ataatgaact gagcctggtg gacggtcaag atcgtaccet gacggtgcaa 60
caatgggata cctttctgaa tggcgTTTTT ccgctggate gtaatecct gaccctgaa 120
tggtttcatt ccggtcgcgc aaaatatac gtcgcagcc cgggtgctga cgaattcgaa 180
ggcacgctgg aactgggtta tcagattggc tttccgtggt cactgggcgt tggtatcaac 240
ttctcgtaca ccacgccgaa tattctgac aacaatggtta acattaccgc accgccgttt 300
ggcctgaaca gcgtgattac gccgaacctg tttccgggtg ttagcatctc tgcccgtctg 360
ggcaatggtc cgggcattca agaagtggca accttagtg tgcgcgtttc cggcgctaaa 420
ggcgggtgctg cgggtgtctaa cgcccacggt accgttacgg gcgcggccgg cgggtgtcctg 480
ctgcgtccgt tcgcgcgct gattgcctet accggcgaca gcgttacgac ctatggcgaa 540
ccgtggaata tgaactaa                                     558

```

<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  
 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>  $\alpha$ -溶血素突变体 (E111N/K147N)

<400> 3

atggcagatt ctgatattaa tattaaaacc ggtactacag atattggaag caactactaca 60  
 gtaaaaacag gtgatttagt cacttatgat aaagaaatg gcatgcacaa aaaagtattt 120  
 tatagtttta tcgatgataa aaatcacaat aaaaaactgc tagttattag aacaaaaggt 180  
 accattgctg gtcaatatag agtttatagc gaagaaggtg ctaacaaaag tggtttagcc 240  
 tggccttcag cctttaaggt acagttgcaa ctacctgata atgaagtage tcaaatatct 300  
 gattactatc caagaaatc gattgataca aaaaactata tgagtacttt aacttatgga 360  
 ttcaacggta atgttactgg tgatgataca ggaaaaattg gcggccttat tggtgcaaat 420  
 gtttcgattg gtcatacaact gaactatggt caacctgatt tcaaaacaat tttagagagc 480

ccaactgata aaaaagtagg ctggaaagtg atatttaaca atatggtgaa tcaaaattgg 540  
 ggaccatacg atcgagattc ttggaacccg gtatatggca atcaactttt catgaaaact 600  
 agaaatgggtt ctatgaaagc agcagataac ttccttgatc ctaacaaagc aagttctcta 660  
 ttatcttcag ggttttcacc agacttcgct acagttatta ctatggatag aaaagcatcc 720  
 aaacaacaaa caaatataga tgtaataac gaacgagttc gtgatgatta ccaattgcat 780  
 tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgttcttca 840  
 gaaagatata aaatcgattg ggaaaaagaa gaaatgacaa attaa 885

<210> 4

<211> 293

<212> PRT

<213> 人工序列

<220>

<223>  $\alpha$ -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			
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
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
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
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
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
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> 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
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  
 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(Bacillus subtilis)噬菌体phi29

<400> 8

atgaaacaca tgccgcgtaa aatgtatagc tgcgcgtttg aaaccacgac caaagtggaa 60  
 gattgtcgcg tttgggccta tggtacatg aacatcgaag atcattctga atacaaaatc 120  
 ggtaacagtc tggatgaatt tatggcatgg gtgctgaaag ttcaggcgga tctgtacttc 180  
 cacaacctga aatttgatgg cgcattcatt atcaactggc tggaacgtaa tggctttaa 240  
 tggagcgcgg atggtctgcc gaacacgtat aataccatta tctctcgtat gggccagtgg 300  
 tatatgattg atatctgcct gggtacaaa ggtaaacgca aaattcatac cgtgatctat 360  
 gatagcctga aaaaactgcc gtttccggtg aagaaaattg cgaaagattt caaactgacg 420  
 gttctgaaag gcgatattga ttatcacaaa gaacgtccgg ttggttaca aatcaccccg 480  
 gaagaatacg catacatcaa aaacgatata cagatcatcg cagaagcgt gctgattcag 540  
 tttaaacagg gcctggatcg catgaccgcg ggcagtgata gcctgaaagg tttcaaagat 600  
 atcatcacga ccaaaaaatt caaaaaagtg ttcccgcgc tgagcctggg tctggataaa 660  
 gaagttcggt atgcctaccg cggcggtttt acctggetga acgatcgttt caaagaaaa 720

gaaattggcg agggatatggt gtttgatggt aatagtctgt atccggcaca gatgtacagc 780  
 cgctctgtgc cgtatggcga accgatcgtg ttcgaggta aatatgtttg ggatgaagat 840  
 taccgcgtgc atattcagca catccgttgt gaatttgaac tgaaagaagg ctatattccg 900  
 accattcaga tcaaacgtag tcgcttctat aagggtaac aatacctgaa aagctctggc 960  
 ggtgaaatcg cggatctgtg gctgagtaac gtggatctgg aactgatgaa agaacactac 1020  
 gatctgtaca acgttgaata catcagcggc ctgaaattta aagccacgac cggctctgttc 1080  
 aaagatttca tcgataaatg gacctacatc aaaacgacct ctgaaggcgc gattaaacag 1140  
 ctggccaaac tgatgctgaa cagcctgtat ggcaaattcg cctctaatcc ggatgtgacc 1200  
 ggtaaagtgc cgtacctgaa agaaaatggc gcaactgggt ttcgcctggg cgaagaagaa 1260  
 acgaaagatc cgggtgtatac cccgatgggt gttttcatta cggcctgggc acgttacacg 1320  
 accatcaccg cgccccagge atgctatgat cgcattatct actgtgatac cgattctatt 1380  
 catctgacgg gcaccgaaat cccggatgtg attaaagata tcgttgatcc gaaaaaactg 1440  
 ggttattggg cccacgaaaag tacgtttaa cgtgcaaaat acctgcgcca gaaacactac 1500  
 atccaggata tctacatgaa agaagtggat ggcaaactgg ttgaaggttc tccgatgat 1560  
 tacaccgata tcaaattcag tgtgaaatgc gccggcatga cggataaaat caaaaaagaa 1620  
 gtgaccttcg aaaacttcaa agttggtttc agccgcaaaa tgaaaccgaa accggtgcag 1680  
 gttccgggcg gtgtggttct ggtggatgat acgtttacca ttaaactctg cggtagtgcg 1740  
 tggagccatc cgcagttcga aaaaggcgggt ggctctggtg gcggttctgg cggtagtgcc 1800  
 tggagccacc cgcagtttga aaaataataa 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  
 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
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
Asp Ser Leu Lys Gly Phe Lys Asp Ile Ile Thr Thr Lys 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
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
Gly Glu Ile Ala Asp Leu Trp Leu Ser Asn Val Asp Leu Glu Leu Met		
325	330	335
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
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  
 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 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(Escherichia coli)

<400> 10

atgatgaacg atggcaaaca gcagagcacc ttctgtttc atgattatga aaccttcggt 60  
 acccatccgg ccctggatcg tccggcgcag tttgeggcca ttcgcaccga tagcgaattc 120  
 aatgtgattg gcgaaccgga agtgttttat tgcaaaccgg ccgatgatta tctgccgcag 180  
 ccgggtgcgg tgctgattac cgggtattacc ccgcaggaag cgcgcgcgaa aggtgaaaac 240  
 gaagcggcgt ttgccgcgcg cattcatagc ctgtttaccg tgccgaaaac ctgcattctg 300  
 ggctataaca atgtgcgctt cgatgatgaa gttaccgta atatctttta tcgtaacttt 360  
 tatgatccgt atgcgtggag ctggcagcat gataacagcc gttgggatct gctggatgtg 420  
 atgcgcgcgt gctatgcgct gcgcccggaa ggcattaatt ggccgaaaa cgatgatggc 480  
 ctgccgagct ttcgtctgga acatctgacc aaagccaacg gcattgaaca tagcaatgcc 540  
 catgatgcga tggccgatgt ttatgcgacc attgcgatgg cgaaactggt taaaaccgct 600  
 cagccgcgcc tgtttgatta tctgtttacc caccgtaaca aacacaaact gatggcgctg 660  
 attgatgttc cgcagatgaa accgctggtg catgtgagcg gcatgtttgg cgctggcgc 720

ggcaacacca gctgggtggc cccgctggcc tggcaccceg aaaatcgtaa cgccgtgatt 780  
 atggttgatc tggccggtga tattagcccc ctgctggaac tggatagcga taccctgcgt 840  
 gaacgcctgt ataccgcaa aaccgatctg ggcgataatg ccgccgtgcc ggtgaaactg 900  
 gttcacatta acaaatgccc ggtgctggcc caggcgaaca cctgcgccc ggaagatgcg 960  
 gatcgtctgg gtattaatcg ccagcattgt ctggataatc tgaaaatcct gcgtgaaaac 1020  
 ccgcaggtgc gtgaaaaagt ggtggcgatc ttcgcggaag cggaaccgtt caccgccgagc 1080  
 gataacgtgg atgcgcagct gtataacggc ttcttttagcg atgccgatcg cgcggcgatg 1140  
 aaaatcgttc tggaaaccga accgcgcaat ctgccggcgc tggatattac ctttgttgat 1200  
 aaacgtattg aaaaactgct gttaattat cgtgcgcgca atttccggg taccctggat 1260  
 tatgccgaac agcagcgttg gctggaacat cgtcgteagg ttttaccgcc ggaatttctg 1320  
 cagggttatg cggatgaact gcagatgctg gttcagcagt atgccgatga taaagaaaaa 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
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  
 Glu Leu Asp Ser Asp Thr Leu Arg Glu Arg Leu Tyr Thr Ala Lys Thr  
 275 280 285  
 Asp Leu Gly Asp Asn Ala Ala Val Pro Val Lys Leu Val His Ile Asn  
 290 295 300  
 Lys Cys Pro Val Leu Ala Gln Ala Asn Thr Leu Arg Pro Glu Asp Ala  
 305 310 315 320  
 Asp Arg Leu Gly Ile Asn Arg Gln His Cys Leu Asp Asn Leu Lys Ile  
 325 330 335  
 Leu Arg Glu Asn Pro Gln Val Arg Glu Lys Val Val Ala Ile Phe Ala  
 340 345 350  
 Glu Ala Glu Pro Phe Thr Pro Ser Asp Asn Val Asp Ala Gln Leu Tyr  
 355 360 365  
 Asn Gly Phe Phe Ser Asp Ala Asp Arg Ala Ala Met Lys Ile Val Leu  
 370 375 380  
 Glu Thr Glu Pro Arg Asn Leu Pro Ala Leu Asp Ile Thr Phe Val Asp  
 385 390 395 400  
 Lys Arg Ile Glu Lys Leu Leu Phe Asn Tyr Arg Ala Arg Asn Phe Pro  
 405 410 415  
 Gly Thr Leu Asp Tyr Ala Glu Gln Gln Arg Trp Leu Glu His Arg Arg  
 420 425 430  
 Gln Val Phe Thr Pro Glu Phe Leu Gln Gly Tyr Ala Asp Glu Leu Gln  
 435 440 445  
 Met Leu Val Gln Gln Tyr Ala Asp Asp Lys Glu Lys Val Ala Leu Leu  
 450 455 460  
 Lys Ala Leu Trp Gln Tyr Ala Glu Glu Ile Val Ser Gly Ser Gly His  
 465 470 475 480  
 His His His His His

485

&lt;210&gt; 12

&lt;211&gt; 804

&lt;212&gt; DNA

&lt;213&gt; 大肠杆菌

&lt;400&gt; 12

```

atgaaatttg tctcttttaa tatcaacggc ctgcgcgcca gacctcacca gcttgaagcc 60
atcgtcgaaa agcaccaacc ggatgtgatt ggctgcagg agacaaaagt tcatgacgat 120
atgtttccgc tcgaagaggt ggcgaagctc ggctacaacg tgttttatca cgggcagaaa 180
ggccattatg gcgtaggctt gctgacaaa gagacgccga ttgccgtgcg tcgaggcttt 240
cccggtagcg acgaagagge gcagcggcgg attattatgg cggaaatccc ctcactgctg 300
ggtaatgtca ccgtgatcaa cggttacttc ccgcagggtg aaagccgcca ccatccgata 360
aaattcccgg caaaaagcga gttttatcag aatctgcaa actacctgga aaccgaacte 420
aaacgtgata atccggtact gattatgggc gatatgaata tcagccctac agatctggat 480
atcggcattg gcgaagaaaa ccgtaagcgc tggetgcgta ccggtaaatg ctctttcctg 540
ccggaagagc gcgaatggat ggacaggetg atgagctggg ggttggtcga taccttccgc 600
catgcgaatc cgcaaacage agatcgtttc tcatggtttg attaccgctc aaaaggtttt 660
gacgataacc gtggtctgcg catcgacctg ctgctcgcca gccaacgct ggcagaatgt 720
tgcgtagaaa ccggcatcga ctatgaaatc cgcagcatgg aaaaaccgct cgatcacgcc 780
cccgtctggg cgaccttccg ccgc                                     804

```

&lt;210&gt; 13

&lt;211&gt; 268

&lt;212&gt; PRT

&lt;213&gt; 大肠杆菌

&lt;400&gt; 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
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  
 Gly Leu Arg Ile Asp Leu Leu Leu Ala Ser Gln Pro Leu Ala Glu Cys  
 225 230 235 240  
 Cys Val Glu Thr Gly Ile Asp Tyr Glu Ile Arg Ser Met Glu Lys Pro  
 245 250 255  
 Ser Asp His Ala Pro Val Trp Ala Thr Phe Arg Arg  
 260 265

<210> 14

<211> 1275

<212> DNA

<213> 嗜热栖热菌(Thermus thermophilus)

<400> 14

atgtttcgtc gtaaagaaga tctggatccg ccgctggcac tgctgccgct gaaaggcctg 60  
 cgcgaaagccg ccgcactgct ggaagaagcg ctgcgtcaag gtaaacgcat tcgtgttcac 120  
 ggcgactatg atgcggatgg cctgaccggc accgcgatcc tggttcgtgg tctggcccgc 180  
 ctgggtgcgg atgttcatcc gtttatcccg caccgcctgg aagaaggcta tgggtgcctg 240  
 atggaacgcg tcccgaaca tctggaagcc teggacctgt ttctgacctg tgactgcggc 300  
 attaccaacc atgcggaact gcgcgaactg ctggaaaatg gcgtggaagt cattgttacc 360  
 gatcatcata cgccgggcaa aacgccgccc ccgggtctgg tcgtgcatcc ggcgctgacg 420  
 ccgatctga aagaaaaacc gaccggcgca ggcgtggcgt ttctgctgct gtgggactg 480  
 catgaacgcc tgggctgcc gccgccgctg gaatacgcgg acctggcagc cgttggcacc 540  
 attgccgacg ttgccccgct gtggggttgg aatcgtgcac tggtgaaaga aggtctggca 600  
 cgcateccgg ctcatcttg ggtgggctg cgtctgctgg ctgaagcctg ggcctatacc 660  
 ggcaaagcgg tcgaagtgc tttccgcata gcgccgca tcaatgcggc ttcccgcctg 720  
 ggcgaagcgg aaaaagccct gcgcctgctg ctgacggatg atgcggcaga agctcaggcg 780  
 ctggtcggcg aactgcaccg tctgaacgcc cgtcgtcaga ccctggaaga agcgatgctg 840

cgcaaactgc tgccgcagge cgacccggaa gcgaaagcca tcgttctgct ggacccggaa 900  
 ggccatccgg gtgttatggg tattgtggcc tctcgcatcc tggaagcgac cctgcgcccc 960  
 gtctttctgg tggcccaggg caaaggcacc gtgcgttcgc tggctccgat ttccgccgctc 1020  
 gaagcactgc gcagcgcgga agatctgctg ctgcgttatg gtggtcataa agaagcggcg 1080  
 ggtttcgcaa tggatgaagc gctgtttccg gcgttcaaag cacgcgttga agcgtatgcc 1140  
 gcacgtttcc cggatccggt tctggaagtg gcaactgctgg atctgctgcc ggaaccgggc 1200  
 ctgctgccgc aggtgttccg tgaactggca ctgctggaac cgtatggtga aggtaaccgg 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	Glu	Ala	Leu	Arg
			20					25					30		
Gln	Gly	Lys	Arg	Ile	Arg	Val	His	Gly	Asp	Tyr	Asp	Ala	Asp	Gly	Leu
			35				40						45		
Thr	Gly	Thr	Ala	Ile	Leu	Val	Arg	Gly	Leu	Ala	Ala	Leu	Gly	Ala	Asp
			50				55					60			
Val	His	Pro	Phe	Ile	Pro	His	Arg	Leu	Glu	Glu	Gly	Tyr	Gly	Val	Leu
65						70					75			80	
Met	Glu	Arg	Val	Pro	Glu	His	Leu	Glu	Ala	Ser	Asp	Leu	Phe	Leu	Thr
						85					90			95	
Val	Asp	Cys	Gly	Ile	Thr	Asn	His	Ala	Glu	Leu	Arg	Glu	Leu	Leu	Glu
			100								105			110	
Asn	Gly	Val	Glu	Val	Ile	Val	Thr	Asp	His	His	Thr	Pro	Gly	Lys	Thr
			115								120			125	
Pro	Pro	Pro	Gly	Leu	Val	Val	His	Pro	Ala	Leu	Thr	Pro	Asp	Leu	Lys
			130								135			140	
Glu	Lys	Pro	Thr	Gly	Ala	Gly	Val	Ala	Phe	Leu	Leu	Leu	Trp	Ala	Leu
145						150					155			160	
His	Glu	Arg	Leu	Gly	Leu	Pro	Pro	Pro	Leu	Glu	Tyr	Ala	Asp	Leu	Ala
						165					170			175	
Ala	Val	Gly	Thr	Ile	Ala	Asp	Val	Ala	Pro	Leu	Trp	Gly	Trp	Asn	Arg
						180					185			190	
Ala	Leu	Val	Lys	Glu	Gly	Leu	Ala	Arg	Ile	Pro	Ala	Ser	Ser	Trp	Val
			195								200			205	

Gly Leu Arg Leu Leu Ala Glu Ala Val Gly Tyr Thr Gly Lys Ala Val  
 210 215 220  
 Glu Val Ala Phe Arg Ile Ala Pro Arg Ile Asn Ala Ala Ser Arg Leu  
 225 230 235 240  
 Gly Glu Ala Glu Lys Ala Leu Arg Leu Leu Leu Thr Asp Asp Ala Ala  
 245 250 255  
 Glu Ala Gln Ala Leu Val Gly Glu Leu His Arg Leu Asn Ala Arg Arg  
 260 265 270  
 Gln Thr Leu Glu Glu Ala Met Leu Arg Lys Leu Leu Pro Gln Ala Asp  
 275 280 285  
 Pro Glu Ala Lys Ala Ile Val Leu Leu Asp Pro Glu Gly His Pro Gly  
 290 295 300  
 Val Met Gly Ile Val Ala Ser Arg Ile Leu Glu Ala Thr Leu Arg Pro  
 305 310 315 320  
 Val Phe Leu Val Ala Gln Gly Lys Gly Thr Val Arg Ser Leu Ala Pro  
 325 330 335  
 Ile Ser Ala Val Glu Ala Leu Arg Ser Ala Glu Asp Leu Leu Leu Arg  
 340 345 350  
 Tyr Gly Gly His Lys Glu Ala Ala Gly Phe Ala Met Asp Glu Ala Leu  
 355 360 365  
 Phe Pro Ala Phe Lys Ala Arg Val Glu Ala Tyr Ala Ala Arg Phe Pro  
 370 375 380  
 Asp Pro Val Arg Glu Val Ala Leu Leu Asp Leu Leu Pro Glu Pro Gly  
 385 390 395 400  
 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 atgacaccgg acattatcct gcagcgtacc 60  
 gggatcgatg tgagagctgt cgaacagggg gatgatgcgt ggcacaaatt acggctcggc 120  
 gtcataccg cttcagaagt tcacaacgtg atageaaaac cccgctccgg aaagaagtgg 180  
 cctgacatga aaatgtccta ctccacacc ctgcttgcg aggtttgcac cgggtgtggct 240  
 ccggaagtta acgctaaage actggcctgg ggaaaacagt acgagaacga cgccagaacc 300  
 ctgtttgaat tcaactccgg cgtgaatggt actgaatccc cgatcateta tcgacgacgaa 360

agtatgcgta ccgcctgctc tcccgatggt ttatgcagtg acggcaacgg cettgaactg 420  
 aaatgcccggt ttacctcccg ggatttcatg aagttccggc tcggtggttt cgaggccata 480  
 aagtcagctt acatggccca ggtgcagtac agcatgtggg tgacgcgaaa aaatgcctgg 540  
 tactttgcca actatgaccc gcgtatgaag cgtgaaggcc tgcattatgt cgtgattgag 600  
 cgggatgaaa agtacatggc gagttttgac gagatcgtgc cggagtcat cgaaaaaatg 660  
 gacgaggcac tggctgaaat tggttttgta tttggggagc aatggcgatc tggctctggt 720  
 tccggcagcg gttccgga 738

<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				
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
Trp Arg		
225		
<210> 18		
<211> 760		
<212> PRT		
<213> 伯顿拟甲烷球菌 (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		
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
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
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
Ala Gln Leu	Met Glu Asn	Tyr Val Glu	Ala Asp Ala Glu Asp Ile Trp
	405	410	415
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
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





370	375	380
Arg Ala Gly Arg Pro Gln Tyr Asp Lys Ser Gly Glu Ala Ile Val Val		
385	390	395
Gly Gly Val Asn Ala Asp Glu Ile Phe Asp Arg Tyr Ile Gly Gly Glu		400
	405	410
Pro Glu Pro Ile Arg Ser Ala Met Val Asp Asp Arg Ala Leu Arg Ile		415
	420	425
His Val Leu Ser Leu Val Thr Thr Ser Pro Gly Ile Lys Glu Asp Asp		430
	435	440
Val Thr Glu Phe Phe Leu Gly Thr Leu Gly Gly Gln Gln Ser Gly Glu		445
	450	455
Ser Thr Val Lys Phe Ser Val Ala Val Ala Leu Arg Phe Leu Gln Glu		460
465	470	475
Glu Gly Met Leu Gly Arg Arg Gly Gly Arg Leu Ala Ala Thr Lys Met		480
	485	490
Gly Arg Leu Val Ser Arg Leu Tyr Met Asp Pro Met Thr Ala Val Thr		495
	500	505
Leu Arg Asp Ala Val Gly Glu Ala Ser Pro Gly Arg Met His Thr Leu		510
	515	520
Gly Phe Leu His Leu Val Ser Glu Cys Ser Glu Phe Met Pro Arg Phe		525
	530	535
Ala Leu Arg Gln Lys Asp His Glu Val Ala Glu Met Met Leu Glu Ala		540
545	550	555
Gly Arg Gly Glu Leu Leu Arg Pro Val Tyr Ser Tyr Glu Cys Gly Arg		560
	565	570
Gly Leu Leu Ala Leu His Arg Trp Ile Gly Glu Ser Pro Glu Ala Lys		575
	580	585
Leu Ala Glu Asp Leu Lys Phe Glu Ser Gly Asp Val His Arg Met Val		590
	595	600
Glu Ser Ser Gly Trp Leu Leu Arg Cys Ile Trp Glu Ile Ser Lys His		605
	610	615
Gln Glu Arg Pro Asp Leu Leu Gly Glu Leu Asp Val Leu Arg Ser Arg		620
625	630	635
Val Ala Tyr Gly Ile Lys Ala Glu Leu Val Pro Leu Val Ser Ile Lys		640
	645	650
Gly Ile Gly Arg Val Arg Ser Arg Arg Leu Phe Arg Gly Gly Ile Lys		655
	660	665
Gly Pro Gly Asp Leu Ala Ala Val Pro Val Glu Arg Leu Ser Arg Val		670
	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  
 <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  
 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  
 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  
 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
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			
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
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
Gly Gly Gln Cys Leu Val Phe Val Ser Ser Arg Arg Asn Ala Glu Gly		
245	250	255
Phe Ala Lys Lys Ala Ala Gly Ala Leu Lys Ala Gly Ser Pro Asp Ser		
260	265	270
Lys Ala Leu Ala Gln Glu Leu Arg Arg Leu Arg Asp Arg Asp Glu Gly		
275	280	285
Asn Val Leu Ala Asp Cys Val Glu Arg Gly Ala Ala Phe His His Ala		
290	295	300
Gly Leu Ile Arg Gln Glu Arg Thr Ile Ile Glu Glu Gly Phe Arg Asn		
305	310	315
Gly Tyr Ile Glu Val Ile Ala Ala Thr Pro Thr Leu Ala Ala Gly Leu		
325	330	335
Asn Leu Pro Ala Arg Arg Val Ile Ile Arg Asp Tyr Asn Arg Phe Ala		
340	345	350
Ser Gly Leu Gly Met Val Pro Ile Pro Val Gly Glu Tyr His Gln Met		
355	360	365
Ala Gly Arg Ala Gly Arg Pro His Leu Asp Pro Tyr Gly Glu Ala Val		
370	375	380
Leu Leu Ala Lys Asp Ala Pro Ser Val Glu Arg Leu Phe Glu Thr Phe		
385	390	395
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  
 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  
 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 Ser Asp Lys Thr Glu Glu Ile Pro Asp  
    770                      775                      780  
 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  
 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
Ile Arg Glu Ala Val Gly Glu Asp Ala	Ser Leu Lys Ser Arg Asp Val	
	245	250
Ala Ala Leu Asp Thr Arg Lys Ser Lys	Gln His Val Asp Pro Glu Ile	
	260	265
Arg Met Ala Glu Trp Met Gln Thr	Leu Lys Glu Thr Gly Phe Asp Ile	
	275	280
Arg Ala Tyr Arg Asp Ala Ala Asp	Gln Arg Thr Glu Ile Arg Thr Gln	
	290	295
Ala Pro Gly Pro Ala Ser Gln Asp	Gly Pro Asp Val Gln Gln Ala Val	
305	310	315
Thr Gln Ala Ile Ala Gly Leu Ser	Glu Arg Lys Val Gln Phe Thr Tyr	
	325	330
Thr Asp Val Leu Ala Arg Thr Val	Gly Ile Leu Pro Pro Glu Asn Gly	
	340	345
Val Ile Glu Arg Ala Arg Ala Gly	Ile Asp Glu Ala Ile Ser Arg Glu	
	355	360
Gln Leu Ile Pro Leu Asp Arg Glu	Lys Gly Leu Phe Thr Ser Gly Ile	
	370	375
His Val Leu Asp Glu Leu Ser Val	Arg Ala Leu Ser Arg Asp Ile Met	
385	390	395
Lys Gln Asn Arg Val Thr Val His	Pro Glu Lys Ser Val Pro Arg Thr	
	405	410
Ala Gly Tyr Ser Asp Ala Val Ser	Val Leu Ala Gln Asp Arg Pro Ser	
	420	425
Leu Ala Ile Val Ser Gly Gln Gly	Gly Ala Ala Gly Gln Arg Glu Arg	
	435	440
Val Ala Glu Leu Val Met Met Ala	Arg Glu Gln Gly Arg Glu Val Gln	
	450	455
Ile Ile Ala Ala Asp Arg Arg Ser	Gln Met Asn Leu Lys Gln Asp Glu	
465	470	475
Arg Leu Ser Gly Glu Leu Ile Thr	Gly Arg Arg Gln Leu Leu Glu Gly	
	485	490
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  
 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  
 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
Ala Phe Ser Met Val Asp Leu Leu Thr Glu Ala Lys Ser Phe Ala Ala		
885	890	895
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
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
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
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
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
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
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  
 <212> PRT  
 <213> 伯顿拟甲烷球菌  
 <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  
 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  
 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  
 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



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
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
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



Val Leu Lys His Gly Asn Phe Glu Ala Arg Gly Ile Ser Arg Glu Gln  
 210 215 220  
 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  
 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 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  
 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
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
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
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
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
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
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

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> 50

<212> DNA

<213> 人工序列

<220>

<223> MuA底物的双链部分的多核苷酸序列

<400> 26

gttttcgcat ttatcgtgaa acgctttcgc gtttttcgtg cgccgcttca 50

<210> 27

<211> 50

<212> DNA

<213> 人工序列

<220>

<223> MuA底物的双链部分的多核苷酸序列

<400> 27

caaaaagcgt aatagcactt tgcgaaagcg caaaaagcac gcggcgaagu 50

<210> 28

<211> 54

<212> DNA

<213> 人工序列

<220>

<223> 双链MuA底物的突出端的多核苷酸序列

<400> 28

caaaagcgta aatagcactt tgcgaaagcg caaaaagcac gcggcgaagu cttag 54

<210> 29

<211> 3560

<212> DNA

<213> 人工序列

<220>

<223> 实施例1中使用的序列

<400> 29

gccatcagat tgtgtttggt agtcgctttt tttttttgga attttttttt tggaattttt 60  
 ttttttgcgct aacaacctcc tgcggttttg cccgtgcata tcggtcacga acaaatctga 120  
 ttactaaaca cagtagcctg gatttgttct atcagtaate gaccttatte ctaattaaat 180  
 agagcaaatc cccttattgg gggtaagaca tgaagatgcc agaaaaacat gacctgttgg 240  
 ccgccattct cgcggcaaaag gaacaaggca tcggggcaat ccttgcgttt gcaatggcgt 300  
 accttcgcgg cagatataat ggcggtgcgt ttacaaaaac agtaatcgac gcaacgatgt 360  
 gcgccattat cgctagttc attcgtgacc ttctcgactt cgccggacta agtagcaatc 420  
 tcgcttatat aacgagcgtg tttatcggtc acatcggtac tgactcgatt ggttcgctta 480  
 tcaaacgctt cgctgctaaa aaagccggag tagaagatgg tagaaatcaa taatcaacgt 540  
 aaggcgttcc tcgatatgct ggcgtggtcg gagggaactg ataacggacg tcagaaaacc 600  
 agaaatcatg gttatgacgt cattgtaggc ggagagctat ttactgatta ctccgatcac 660  
 cctcgcaaac ttgtcacgct aaacccaaaa ctcaaatcaa caggcgccgg acgctaccag 720  
 cttctttccc gttggtggga tgcctaccgc aagcagcttg gcctgaaaga cttctctccc 780  
 aaaagtcagg acgctgtggc attgcagcag attaaggagc gtggcgcttt acctatgatt 840  
 gatcgtgggtg atatccgtca ggcaatcgac cgttcagca atatctgggc ttcactgccg 900  
 ggcgctgggt atggtcagtt cgagcataag gctgacagcc tgattgcaa attcaaagaa 960  
 gcgggcgga cggtcagaga gattgatgta tgagcagagt caccgcgatt atctccgctc 1020  
 tggttatctg catcatcgtc tgcctgtcat gggctgttaa tcattaccgt gataacgcca 1080  
 ttacctaaa agcccagcgc gacaaaaatg ccagagaact gaagctggcg aacgcgga 1140  
 ttactgacat gcagatgcgt cagcgtgatg ttgctgcgct cgatgcaaaa tacacgaagg 1200  
 agttagctga tgctaaagct gaaaaatgat ctctgcgtga tgatgttcc gctggtcgctc 1260  
 gtcggttgca catcaaagca gtctgtcagt cagtgcgtga agccaccacc gcctccggcg 1320  
 tggataatgc agcctcccc cgaactggcag acaccgctga acgggattat ttaccctca 1380  
 gagagaggct gatcactatg caaaaacaac tggaaggaac ccagaagtat attaatgagc 1440  
 agtgcagata gaggttgcca ttcgatggg caactcatgc aattattgtg agcaatacac 1500  
 acgcgcttcc agcggagtat aaatgcctaa agtaataaaa ccgagcaate catttacgaa 1560  
 tgtttgctgg gtttctgttt taacaacatt ttctgcgccg ccacaaattt tggctgcatc 1620  
 gacagttttc ttctgcecaa ttccagaaac gaagaaatga tgggtgatgg tttcctttgg 1680  
 tgctactgct gccggtttgt tttgaacagt aaacgtctgt tgagcacatc ctgtaataag 1740

cagggccagc gcagtagcga gtagcatttt tttcatgggtg ttattcccga tgctttttga 1800  
 agttcgcaga atcgtatgtg tagaaaatta acaaacctt aaacaatgag ttgaaatttc 1860  
 atattgttaa tatttattaa tgtatgtcag gtgcgatgaa tcgtcattgt attccccgat 1920  
 taactatgtc cacagccctg acggggaact tctctgcggg agtgtccggg aataattaa 1980  
 acgatgcaca cagggtttag cgcgtacacg tattgcatta tgccaacgcc ccggtgctga 2040  
 cacggaagaa accggacgtt atgatttagc gtggaaagat ttgtgtagtg ttctgaatgc 2100  
 tctcagtaaa tagtaatgaa ttatcaaagg tataagtaata tcttttatgt tcatggatat 2160  
 ttgtaaccca tcggaact cctgcttag caagatttc cctgtattgc tgaaatgtga 2220  
 tttctcttga tttcaacct tcataggacg tttctataag atgcgtgtt cttgagaatt 2280  
 taacatttac aacctttta agtctttta ttaacacggt gttatcgtt tctaacacga 2340  
 tgtgaatatt atctgtggct agatagtaaa tataatgtga gacgttgtga cgttttagtt 2400  
 cagaataaaa caattcacag tctaaatctt ttcgcacttg atcgaatatt tctttaaaaa 2460  
 tggcaacctg agccattggt aaaaccttc atgtgatacg agggcgcgta gtttgcatta 2520  
 tcgtttttat cgtttcaatc tggctgacc tcttgtgtt ttgttgatga tttatgtcaa 2580  
 atattaggaa tgtttctact taatagtatt ggttgcgtaa caaagtgcgg tctgtctggc 2640  
 attctggagg gaaatacaac cgacagatgt atgtaaggcc aacgtgctca aatcttcata 2700  
 cagaaagatt tgaagtaata ttttaaccgc tagatgaaga gcaagcgcac ggagcgacaa 2760  
 aatgaataaa gaacaatctg ctgatgatcc ctccgtggat ctgattcgtg taaaaaatat 2820  
 gcttaatagc accatttcta tgagttacc tgatgttga attgcatgta tagaacataa 2880  
 ggtgtctctg gaagcattca gagcaattga ggcagcgtt gtgaagcacg ataataatat 2940  
 gaaggattat tccctgggtg ttgactgatc accataactg ctaatcattc aaactattta 3000  
 gtctgtgaca gagccaacac gcagtctgtc actgtcagga aagtggtaaa actgcaactc 3060  
 aattactgca atgccctcgt aattaagtga attacaata tcgtcctgtt cggaggaag 3120  
 aacgcgggat gttcattctt catcactttt aattgatgta tatgctctct tttctgacgt 3180  
 tagtctccga cggcaggctt caatgacca ggctgagaaa tcccggacc ctttttgctc 3240  
 aagagcgatg ttaatttgtt caatcatttg gttaggaaag cggatgttgc ggttgttgt 3300  
 tctgcgggtt ctgttcttcg ttgacatgag gttgccccgt attcagtgtc gctgatttgt 3360  
 attgtctgaa gttgttttta cgtaagtgt atgcagatca attaatacga tacctgcgtc 3420  
 ataattgatt atttgacgtg gtttgatggc ctccacgcac gttgtgatat gtagatgata 3480  
 atcattatca ctttacgggt ctttccgggt gaaaaaaaaag gtaccaaaaa aaacatcgtc 3540  
 gtgagtagtg aaccgtaagc 3560  
 <210> 30  
 <211> 3560  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 实施例1中使用的序列  
 <400> 30  
 gcttacggtt cactactcac gacgatgtt tttttgtac ctttttttc accgaaagg 60

acccgtaaag tgataatgat tatcatctac atatcacaac gtgcgtggag gccatcaaac 120  
 cacgtcaaat aatcaattat gacgcaggta tcgtattaat tgatctgcat caacttaacg 180  
 taaaaacaac ttcagacaat acaaatcagc gacactgaat acggggcaac ctcatgtcaa 240  
 cgaagaacag aaccgcaga acaacaacc gcaacatccg ctttcctaac caaatgattg 300  
 aacaaattaa catcgtcttt gagcaaaaag ggtccgggaa tttctcagcc tgggtcattg 360  
 aagcctgccg tcggagacta acgtcagaaa agagagcata tacatcaatt aaaagtgatg 420  
 aagaatgaac atccccgctt cttccctccg aacaggacga tattgtaaatt tcaacttaatt 480  
 acgagggcat tgcagtaatt gagttgcagt tttaccactt tctgacagt gacagactgc 540  
 gtgttggtc tgcacagac taaatagttt gaatgattag cagttatggt gatcagtcaa 600  
 ccaccaggga ataatectt atattattat cgtgetteac caacgtgcc tcaattgctc 660  
 tgaatgctt cagagacacc ttatgttcta tacatgcaat tacaacatca gggtaactca 720  
 tagaaatggt gctattaage atatTTTTTta cacgaatcag atccacggag ggatcatcag 780  
 cagattgtt tttattcatt ttgtcctcc atgccttgc tcttcatcta gcggttaaaa 840  
 tattacttca aatctttctg tatgaagatt tgagcacgtt ggccttacet acatctgtcg 900  
 gttgtattt cctccagaat gccagcagga ccgcacttg ttacgcaacc aatactatta 960  
 agtgaaaaca ttcttaatat ttgacataaa tcatcaaaa aacacaagga ggtcagacca 1020  
 gattgaaacg ataaaaacga taatgcaaac tacgcgcct cgtatcacat ggaaggtttt 1080  
 accaatggct caggttgcca tttttaaaga aatattcgat caagtgcgaa aagatttaga 1140  
 ctgtgaattg ttttattctg aactaaaacg tcacaacgtc tcacattata tttactatct 1200  
 agccacagat aatattcaca tcgtgttaga aaacgataac accgtgttaa taaaaggact 1260  
 taaaaagggt gtaaagtta aattctcaag aaacacgcat cttatagaaa cgtcctatga 1320  
 taggttgaaa tcaagagaaa tcacatttca gcaatacagg gaaaatcttg ctaaagcagg 1380  
 agttttccga tgggttacia atatccatga acataaaaga tattactata ctttgataa 1440  
 ttactacta tttactgaga gcattcagaa cactacacia atctttccac gctaaatcat 1500  
 aacgtccggt ttcttccgt tcagcaccgg ggcgttgga taatgcaata cgtgtacgcg 1560  
 ctaaaccctg tgtgcatcgt ttaattatt cccggacact cccgcagaga agttccccgt 1620  
 cagggctgtg gacatagtta atccgggaat acaatgacga ttcacgcac ctgacataca 1680  
 ttaataaata ttaacaatat gaaatttcaa ctattgttt agggtttgtt taattttcta 1740  
 cacatacgat tctgcgaact tcaaaaagca tcgggaataa caccatgaaa aaaatgctac 1800  
 tcgctactgc gctggccctg cttattacag gatgtgetca acagacgttt actgttcaaa 1860  
 acaaacggc agcagtagca ccaaaggaaa ccatcecca tcatttctt gtttctggaa 1920  
 ttgggcagaa gaaaactgtc gatgcagcca aaatttgtgg cggcgcagaa aatgttgta 1980  
 aaacagaaac ccagcaaaac ttcgtaaatg gattgctcgg ttttattact ttaggcattt 2040  
 atactccgct ggaagcgcgt gtgtattget cacaataatt gcatgagttg cccatcgata 2100  
 tgggcaactc tatctgcaact getcattaat ataettctgg gttccttcca gttgtttttg 2160  
 catagtgate agcctctctc tgagggtgaa ataateccgt tcagcgggtg ctgccagtcg 2220  
 gggggaggct gcattatcca cgccggaggc ggtggtggct tcacgactg actgacagac 2280  
 tgctttgatg tgcaaccgac gacgaccagc ggcaacatca tcacgcagag catcattttc 2340  
 agcttttagca tcagctaact ccttctgtga ttttgcctc agcgcagcaa catcagctg 2400

acgcatctgc atgtcagtaa ttgccgcggt cgccagette agttctctgg catttttgtc 2460  
 gcgctgggct ttgtaggtaa tggcgttatc acggtaatga ttaacagccc atgacaggca 2520  
 gacgatgatg cagataacca gagcggagat aatcgcggtg actctgctca tacatcaatc 2580  
 tctctgaccg ttccgcccgc ttctttgaat tttgcaatca ggctgtcagc cttatgctcg 2640  
 aactgaccat aaccagcgcg cggcagtgaa gccagatat tgctgcaacg gtcgattgcc 2700  
 tgacggatat caccacgata aatcataggt aaagcgccac gctcctaat ctgctgcaat 2760  
 gccacagcgt cctgactttt cggagagaag tctttcagge caagctgctt gcggtaggca 2820  
 tcccaccaac gggaaagaag ctggtagcgt ccggcgectg ttgatttgag ttttgggttt 2880  
 agcgtgacaa gtttgcgagg gtgatcggag taatcagtaa atagctctcc gcctacaatg 2940  
 acgtcataac catgatttct ggttttctga cgtecgttat cagtccctc cgaccacgcc 3000  
 agcatatcga ggaacgcctt acgttgatta ttgattteta ccatctteta ctccggcttt 3060  
 tttagcagcg aagcgtttga taagcgaacc aatcgagtea gtaccgatgt agccgataaa 3120  
 cacgctcgtt atataagcga gattgtaact tagtccggcg aagtcgagaa ggtcacgaat 3180  
 gaactaggcg ataatggcgc acatcgttgc gtcgattact gtttttgtaa acgcaccgcc 3240  
 attatatctg ccgcgaaggt acgccattgc aaacgcaagg attgccccga tgccttgttc 3300  
 ctttgccgcg agaatggcgg ccaacaggtc atgtttttct ggcattctca tgtcttacc 3360  
 ccaataaggg gatttgctct atttaattag gaataagtc gattactgat agaacaaatc 3420  
 caggctactg tgtttagtaa tcagatttgt tcgtgaccga tatgcacggg caaaacggca 3480  
 ggaggttggt agcgcaaaaa aaaaattcca aaaaaaaaaat tccaaaaaaa aaaagcgact 3540  
 aacaaacaca atctgatggc 3560

<210> 31

<211> 85

<212> DNA

<213> 人工序列

<220>

<223> 实施例1中使用的序列

<400> 31

gccatcagat tggtttgggt agtcgctttt ttttttggga attttttttt tggaattttt 60  
 tttttgcgct aacaacctcc tgccg 85

<210> 32

<211> 72

<212> DNA

<213> 人工序列

<220>

<223> 实施例1中使用的序列

<400> 32

gcttacgggt cactactcac gacgatgttt tttttgttac cttttttttc accggaaagg 60  
 acccgtaaag tg 72

<210> 33

<211> 46  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 实施例1、2和3中使用的序列  
 <400> 33  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttt 46  
 <210> 34  
 <211> 27  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 实施例1、2和3中使用的序列  
 <400> 34  
 ggttgtttct gttggtgctg atattgc 27  
 <210> 35  
 <211> 3587  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 实施例1、2和3中使用的序列  
 <400> 35  
 gccatcagat tgtgtttggt agtcgctgcc atcagattgt gtttgtagt cgcttttttt 60  
 ttttggaaatt ttttttttgg aatttttttt ttgcgetaac aacctcctgc cgttttgccc 120  
 gtgcatatcg gtcacgaaca aatctgatta ctaaacacag tagcctggat ttgttctatc 180  
 agtaatcgac cttattccta attaaataga gcaaattccc ttattggggg taagacatga 240  
 agatgccaga aaaacatgac ctgttgccg ccattctcgc ggcaaaggaa caaggcatcg 300  
 gggcaatcct tgcgtttgca atggcgtacc ttcgcggcag atataatggc ggtgcgttta 360  
 caaaaacagt aatcgacgca acgatgtgag ccattatcgc ctagttcatt cgtgaccttc 420  
 tcgacttcgc cggactaagt agcaatctcg cttatataac gagegtgttt atcggctaca 480  
 tcggtactga ctcgattggt tcgcttatac aacgcttcgc tgctaaaaaa gccggagtag 540  
 aagatgtagt aatcaataa tcaacgtaag gcgttctcgc atatgctggc gtggtcggag 600  
 ggaactgata acggacgtca gaaaaccaga aatcatggtt atgacgtcat tgtaggcgga 660  
 gagctattta ctgattactc cgatcaccct cgcaaacttg tcacgctaaa cccaaaactc 720  
 aatcaacag gcgccggacg ctaccagett ctttcccgtt ggtgggatgc ctaccgcaag 780  
 cagcttggcc tgaagaactt ctctccgaaa agtcaggacg ctgtggcatt gcagcagatt 840  
 aaggagcgtg gcgctttacc tatgattgat cgtggtgata tccgtcaggc aatcgaccgt 900  
 tgcagcaata tctgggcttc actgccgggc gctggttatg gtcagttcga gcataagget 960  
 gacagcctga ttgcaaaatt caaagaagcg ggcggaacgg tcagagagat tgatgtatga 1020

gcagagtcac cgcgattatc tccgctctgg ttatctgcat catcgtctgc ctgtcatggg 1080  
ctgttaatca ttaccgtgat aacgccatta cctacaaage ccagecgcac aaaaatgcc 1140  
gagaactgaa gctggcgaac gcggcaatta ctgacatgca gatgcgtcag cgtgatgttg 1200  
ctgcgctcga tgcaaaatac acgaaggagt tagctgatgc taaagctgaa aatgatgctc 1260  
tgcgtgatga tgttgccgct ggtcgtcgtc ggttgacat caaagcagtc tgtcagtcag 1320  
tgcgtgaagc caccaccgcc tccggcgtgg ataatgcagc ctcccccca ctggcagaca 1380  
ccgctgaacg ggattatttc accctcagag agaggctgat cactatgcaa aaacaactgg 1440  
aaggaaccca gaagtatatt aatgagcagt gcagatagag ttgcccatat cgatgggcaa 1500  
ctcatgcaat tattgtgagc aatacacacg cgcttcagc ggagtataaa tgcctaaagt 1560  
aataaaaccg agcaatccat ttacgaatgt ttgctgggtt tctgttttaa caacattttc 1620  
tgcgccgcca caaattttgg ctgcatcgac agttttcttc tgcccaattc cagaaacgaa 1680  
gaaatgatgg gtgatggttt cctttgggtc tactgctgcc ggtttgtttt gaacagtaaa 1740  
cgtctgttga gcacatcctg taataagcag ggccagcgca gtagcagta gcattttttt 1800  
catggtgta tccccgatgc tttttgaagt tcgcagaatc gtatgtgtag aaaattaac 1860  
aaaccctaaa caatgagttg aaatttcata ttgttaatat ttattaatgt atgtcaggtg 1920  
cgatgaatcg tcattgtatt cccggattaa ctatgtccac agccctgacg gggaacttct 1980  
ctgcgggagt gtccgggaat aattaaacg atgcacacag ggttttagcgc gtacacgtat 2040  
tgcattatgc caacgccccg gtgctgacac ggaagaaacc ggacgttatg atttagcgtg 2100  
gaaagatttg ttagtggtc tgaatgctc cagtaaatag taatgaatta tcaaaggtat 2160  
agtaatatct tttatgttca tggatatttg taaccatcg gaaaactcct gctttagcaa 2220  
gattttccct gtattgctga aatgtgattt ctcttgattt caacctatca taggacgttt 2280  
ctataagatg cgtgtttctt gagaatttaa cattacaac ctttttaagt ctttttatta 2340  
acacgggtgt atcgttttct aacacgatgt gaatattatc tgtggctaga tagtaaata 2400  
aatgtgagac gttgtgacgt tttagttcag aataaaaca ttcacagtct aaatcttttc 2460  
gcaactgac gaatatttct ttaaaaatgg caacctgagc cattggtaaa acctccatg 2520  
tgatacagg gcgcgtagtt tgcaattatc tttttatcgt ttcaatctgg tctgacctcc 2580  
ttgtgttttg ttgatgattt atgtcaata ttaggaatgt tttacttaa tagtattggt 2640  
tgcgtaacaa agtgcggtcc tgctggcatt ctggaggaa atacaaccga cagatgtatg 2700  
taaggccaac gtgctcaaat cttcatacag aaagatttga agtaatattt taaccgctag 2760  
atgaagagca agcgcgatga gcgacaaaat gaataaagaa caatctgctg atgatccctc 2820  
cgtggatctg attcgtgtaa aaaatatgct taatagcacc atttctatga gttaccctga 2880  
tgttgtaatt gcatgtatag aacataaggt gtctctggaa gcattcagag caattgaggc 2940  
agcgttggtg aagcacgata ataatatgaa ggattattec ctggtggttg actgatcacc 3000  
ataactgcta atcattcaaa ctatttagtc tgtgacagag ccaacacgca gtctgtcact 3060  
gtcaggaaag tggtaaaaact gcaactcaat tactgcaatg cctcgtaat taagtgaatt 3120  
tacaatatcg tcctgttcgg agggaagaac gcgggatgtt cattcttcat cactttta 3180  
tgatgtatat gctctctttt ctgacgttag tctccgacgg caggcttcaa tgaccaggc 3240  
tgagaaattc ccggaccctt tttgetcaag agcgtatgta atttgttcaa tcaattggtt 3300  
aggaaagcgg atgttgccgg ttgtgttct gcgggttctg ttcttcgttg acatgaggtt 3360

gccccgtatt cagtgtcgt gatttgtatt gtctgaagtt gtttttacgt taagttgatg 3420  
 cagatcaatt aatacgatac ctgcgcata attgattatt tgacgtggtt tgatggcctc 3480  
 cacgcacgtt gtgatatgta gatgataatc attatcactt tacgggtcct ttccggtgaa 3540  
 aaaaaaggta ccaaaaaaaaa catcgtcgtg agtagtgaac cgtaagc 3587

<210> 36

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 实施例1、2和3中使用的序列

<400> 36

gccatcagat tgtgtttggt agtcgct 27

<210> 37

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 实施例1、2和3中使用的序列

<400> 37

acactgattg acacggttta gtagaac 27

<210> 38

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 实施例1、2和3中使用的序列

<400> 38

gcttacggtt cactactcac gacgatg 27

<210> 39

<211> 29

<212> DNA

<213> 人工序列

<220>

<223> 实施例1、2和3中使用的序列

<400> 39

gcaatatcag caccaacaga aacaacctt 29

<210> 40

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 实施例1、2和3中使用的序列

<400> 40

gttctactaa accgtgtcaa tcagtgtc 28

<210> 41

<211> 3560

<212> DNA

<213> 人工序列

<220>

<223> 实施例1、2和3中使用的序列

<400> 41

gcttacggtt cactactcac gacgatgttt tttttgttac ctttttttc accggaaagg 60  
 acccgtaaag tgataatgat tatcatctac atatcacaac gtgcgtggag gccatcaaac 120  
 cacgtcaaat aatcaattat gacgcaggta tcgtattaat tgatctgcat caacttaacg 180  
 taaaaacaac ttcagacaat acaaatcagc gacactgaat acggggcaac ctcatgtcaa 240  
 cgaagaacag aaccgcaga acaacaacc gcaacatccg ctttcctaac caaatgattg 300  
 aacaaattaa catcgtcttt gagcaaaaag ggtccgggaa tttctcagcc tgggtcattg 360  
 aagcctgccg tcggagacta acgtcagaaa agagagcata tacatcaatt aaaagtgatg 420  
 aagaatgaac atcccgcgtt cttccctccg aacaggacga tattgtaaatt tcaacttaatt 480  
 acgagggcat tgcagtaatt gagttgcagt ttaccactt tcttgacagt gacagactgc 540  
 gtgttggtc tgtcacagac taaatagttt gaatgattag cagttatggt gatcagtcaa 600  
 ccaccagga ataatccttc atattattat cgtgettcaac caacgtgcc tcaattgctc 660  
 tgaatgcttc cagagacacc ttatgttcta tacatgcaat tacaacatca gggtaactca 720  
 tagaaatggg gctattaagc atatttttta cacgaatcag atccacggag ggatcatcag 780  
 cagattgttc tttattcatt ttgtcgtcc atgcgcttgc tcttcatcta gcggttaaaa 840  
 tattacttca aatctttctg tatgaagatt tgagcacgtt ggccttacet acatctgtcg 900  
 gttgtatttc cctccagaat gccagcagga ccgcacttg ttacgcaacc aatactatta 960  
 agtgaaaaca ttcctaatat ttgacataaa tcatcaacaa aacacaagga ggtcagacca 1020  
 gattgaaacg ataaaaacga taatgcaaac tacgegcct cgtatcacet ggaaggtttt 1080  
 accaatggct caggttgcca tttttaaaga aatattcgat caagtgcgaa aagattttaga 1140  
 ctgtgaattg ttttattctg aactaaaacg tcacaacgtc tcacattata tttactatct 1200  
 agccacagat aatattcaca tcgtgttaga aaacgataac accgtgttaa taaaaggact 1260  
 taaaaagggt gtaaatgtta aattctcaag aaacacgcat cttatagaaa cgtcctatga 1320  
 taggttgaat tcaagagaaa tcacatttca gcaatacagg gaaaatcttg ctaaagcagg 1380  
 agttttccga tgggttacia atatccatga acataaaaga tattactata cttttgataa 1440  
 ttcattacta tttactgaga gcattcagaa cactacacaa atctttccac gctaaatcat 1500  
 aacgtccggt ttcttccgtg tcagcaccgg ggcgttgga taatgcaata cgtgtacgag 1560  
 ctaaaccctg tgtgcatcgt ttttaattatt cccggacact cccgcagaga agttccccgt 1620

cagggctgtg gacatagtta atccgggaat acaatgacga ttcatcgcac ctgacataca 1680  
 ttaataaata ttaacaatat gaaatttcaa ctcatgtttt agggtttggt taattttcta 1740  
 cacatacgat tctgcgaact tcaaaaagca tcgggaataa caccatgaaa aaaatgctac 1800  
 tcgctactgc gctggccctg cttattacag gatgtgctca acagacgttt actgttcaaa 1860  
 acaaaccggc agcagtagca ccaaaggaaa ccatcaccca tcatttcttc gtttctggaa 1920  
 ttgggcagaa gaaaactgtc gatgcagcca aaatttgtgg cggcgcagaa aatgttgtta 1980  
 aaacagaaac ccagcaaaca ttcgtaaagt gattgctcgg ttttattact ttaggcattt 2040  
 atactccgct ggaagcgcgt gtgtattgct cacaataatt gcatgagttg cccatcgata 2100  
 tgggcaactc tatctgcaact gctcattaat ataactctgg gttccttcca gttgtttttg 2160  
 catagtgate agcctctctc tgagggtgaa ataateccgt tcagcgggtg ctgccagtcg 2220  
 gggggaggct gcattatcca cgccggaggc ggtggtggct tcacgcactg actgacagac 2280  
 tgctttgatg tgcaaccgac gacgaccagc ggcaacatca tcacgcagag catcatttct 2340  
 agcttttagc tcagctaaact ccttcgtgta ttttgcatcg agcgcagcaa catcacgctg 2400  
 acgcatctgc atgtcagtaa ttgccgcgtt cgccagcttc agttctctgg catttttctc 2460  
 gcgctgggct ttgtaggtaa tgccgttate acggtaatga ttaacagccc atgacaggca 2520  
 gacgatgatg cagataacca gagcggagat aatcgcgggt actctgctca tacatcaatc 2580  
 tctctgaccg ttccgcccgc ttctttgaat tttgcaatca ggctgtcagc cttatgctcg 2640  
 aactgaccat aaccagcgc cggcagtgaa gccagatat tgctgcaacg gtcgattgcc 2700  
 tgacggatat caccacgate aatcataggt aaagcggcac gctcettaat ctgctgcaat 2760  
 gccacagcgt cctgactttt cggagagaag tctttcaggc caagctgctt gcggtaggca 2820  
 tcccaccaac gggaaagaag ctggtagcgt ccggcgccctg ttgatttgag ttttgggttt 2880  
 agcgtgacaa gtttgcgagg gtgatcggag taatcagtaa atagctctcc gcctacaatg 2940  
 acgtcataac catgatttct ggttttctga cgccgcttat cagtccctc cgaccacgcc 3000  
 agcatatcga ggaacgcctt acgttgatta ttgattteta ccatctteta ctccggcttt 3060  
 tttagcagcg aagcgtttga taagcgaacc aatcgagtca gtaccgatgt agccgataaa 3120  
 cacgctcgtt atataagcga gattgctact tagtccggcg aagtcgagaa ggtcacgaat 3180  
 gaactaggcg ataatggcgc acatcgttgc gtcgattact gtttttgtaa acgcaccgcc 3240  
 attatatctg ccgcgaaggt acgccattgc aaacgcaagg attgccccga tgccttgttc 3300  
 ctttgccgcg agaatggcgg ccaacaggtc atgttttctt ggcattctca tgtcttacc 3360  
 ccaataaggg gatttgctct atttaattag gaataaggtc gattactgat agaacaaatc 3420  
 caggctactg tgtttagtaa tcagatttgt tcgtgaccga tatgcacggg caaaacggca 3480  
 ggaggttgtt agcgcaaaaa aaaaattcca aaaaaaaaaat tcaaaaaaaa aaaagcgact 3540  
 aacaaacaca atctgatggc 3560

<210> 42

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> 实施例4中使用的序列

<400> 42

ttgcaatatac agcaccaaca gaaacaacct t 31

<210> 43

<211> 48502

<212> DNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列

<400> 43

gggcggcgac ctcgcgggtt ttcgctatth atgaaaatth tccggtthaa ggcgthtccg 60  
 thtthtthtcg tcataactta atgthththt thaaaatacc ctctgaaaag aaaggaaacg 120  
 acaggthctg aaagcggagc thththggcct ctgthctthc ththctctgt ththgtccgt 180  
 ggaatgaaca atggaagtca acaaaaagca gthggctgac atththcggth cgagthtccg 240  
 taccatthcag aactggcagg aacagggaat gccctthctg cggagcggth gcaaggghaa 300  
 thaggthctt thtactctg ccgcccctat aaaatgghat gccgaaaggg atgctgaaat 360  
 thgaaacgaa aagctgcgcc gggaggthga agaactgcgg caggccagcg aggcagatct 420  
 ccagccagga actatthagt acgaaaccca thgactthac cgtgcgcagg ccgacgcaca 480  
 ggaactgaaag aatgccagag actcccgtga agthgghgaa accgcattct gthctthctg 540  
 gctgthcggg atcgcaggth aaatthccag ththctgcac gggctcccc thctggthga 600  
 gcggcgttht ccggaactgg aaaaccgaca ththgathth ctgaaacggg atathcatca 660  
 agccatgaaac aaagcagccc cgtgghatga actgataacc gggthgctga gthgaaatath 720  
 cgaacagthca gghthaacagg ctgcggcatt thgtcccgcg cgggctthcg thactgthca 780  
 ggcccggagcc acagaccgcc gthgaaatggg cggatgghaa thactathct ccgaaagaat 840  
 ccgcatacca ggaagggcgc thggaaacac thcccctthca gcgggcccac atgaaatgcga 900  
 thggcagcga ctacatcccgt gaggthgaaat thgthgaaat thcccctgth gththattcca 960  
 aaatgctgct gggthgthth gcctactthta thagacataa gcagcghaac accctthctt 1020  
 gghthcccagc gghatgghat gccgagaact thatgaaaac ccagththgag ccgactathc 1080  
 gthgaththcc gthcgtgctg gcgctggccc cthgghatthg caaaaagcac cggghataaca 1140  
 cgtcaccat gaagcgtthc actaatgggc thgghctctg thgctgggc ghthaaagcgg 1200  
 caaaaaacta ccgthgaaaag thgghgghat thggcggghth thgthgaaact gctgctththg 1260  
 atgathgath thgaaaggaa gghctctcca cghthctggg thgaaagcgt atthgaggct 1320  
 ccgthctggcc aaagthccac cthgghctca cghcaaaagth gagaggcacc thctgagathg 1380  
 agcgtgcagc cagthgaaat ccgcaththta thcghththca ththgctgth ccgcaththg 1440  
 gggaggagca gthctththaa ththggcgaca aagagcgcg gththggcctc aaatggagcgc 1500  
 ccgathgacc ctccagcgtg thththctct gcgagcataa thcctgthgth atcccagcgc 1560  
 aggagctgga ctthactgath gccctthata thctgcaaaa gaccgggath thgaccctgth 1620  
 atggcathct ctgghththc thctccgghg aagagathga gccacctgac agthgthacct 1680  
 thcacathct gacagcghac agcccghthc ccacctgggh gcgaththgth aaagactgga 1740  
 thgaaacgaa agggghathc ghgaaacgha aaacctctgth aacaccagc ctgghthgaa 1800

cgtgggaggc gaaaattggc gaacgtccgg atgctgaagt gatggcagag cggaaagagc 1860  
 attattcagc gcccgttccct gaccgtgtgg cttacctgac cgcccggtatc gactcccagc 1920  
 tggaccgcta cgaaatgcgc gtatggggat gggggccggg tgaggaaagc tggctgattg 1980  
 accggcagat tattatgggc cgccacgacg atgaacagac gctgctgcgt gtggatgagg 2040  
 ccatcaataa aacctatacc cgccggaatg gtgcagaaat gtcgatatcc cgtatctgct 2100  
 gggatactgg cgggattgac ccgaccattg tgtatgaacg ctcgaaaaaa catgggctgt 2160  
 tccgggtgat ccccatataa ggggcatccg tctacggaaa gccggtggcc agcatgccac 2220  
 gtaagcgaaa caaaaacggg gtttacctta ccgaaatcgg tacggatacc gcgaaagagc 2280  
 agatttataa ccgcttcaca ctgacgccgg aaggggatga accgcttccc ggtgccgttc 2340  
 acttcccga taaccgggat atttttgatc tgaccgaagc gcagcagctg actgctgaag 2400  
 agcaggctga aaaaatgggtg gatggcagga aaaaaatact gtgggacagc aaaaagcgac 2460  
 gcaatgaggc actcgactgc ttcgtttatg cgctggcgge gctgcgcate agtatttccc 2520  
 gctggcagct ggatctcagt gcgctgctgg cgagcctgca ggaagaggat ggtgcagcaa 2580  
 ccaacaagaa aacactggca gattacgcc gtgccttate cggagaggat gaatgacgcg 2640  
 acaggaagaa cttgccgctg cccgtgcggc actgcatgac ctgatgacag gtaaaccgggt 2700  
 ggcaacagta cagaaagacg gacgaagggt ggagtttacg gccacttccg tgtctgacct 2760  
 gaaaaaata attgacagagc tggaaagtgc gaccggcatg acacagcgac gcaggggacc 2820  
 tgcaggattt tatgtatgaa aacgcccacc attcccacc ttctggggcc ggacggcatg 2880  
 acatcgctgc gcgaatatgc cggttatcac ggcggtggca gcggatttgg agggcagttg 2940  
 cggtcgtgga acccaccgag tgaaagtgtg gatgcagccc tgttgcccaa ctttaccctg 3000  
 ggcaatgcc gcgcagacga tctggtacgc aataacggct atgccgcaa cgccatccag 3060  
 ctgcatcagg atcatatcgt cgggtctttt ttccggetca gtcategcc aagctggcgc 3120  
 tatctgggca tcggggagga agaagcccgt gccttttccc gcgaggttga agcggcatgg 3180  
 aaagagtttg ccgaggatga ctgctgctgc attgacgttg agcgaaaacg cacgtttacc 3240  
 atgatgattc gggaaaggtg ggccatgcac gcctttaacg gtgaactgtt cgttcaggcc 3300  
 acctgggata ccagttcgtc gcggcttttc cggacacagt tccggatggt cagcccgaag 3360  
 cgcatcagca acccgaacaa taccggcgac agccggaact gccgtgccgg tgtgcagatt 3420  
 aatgacagcg gtgcggcgtc gggatattac gtcagcgagg acgggtatcc tggctggatg 3480  
 ccgcagaaat ggacatggat acccctgag ttaccggcgg gcgcgcctc gttcattcac 3540  
 gtttttgaac ccgtggagga cgggcagact cgcggtgcaa atgtgtttta cagcgtgatg 3600  
 gagcagatga agatgctcga cacgctgcag aacacgcagc tgcagagcgc cattgtgaag 3660  
 gcgatgtatg ccgccacat tgagagtgag ctggatacgc agtcagcagat ggattttatt 3720  
 ctgggcgcga acagtcagga gcagcgggaa aggetgaccg gctggattgg tgaattgcc 3780  
 gcgtattacg ccgcagcgc ggtccggetg ggagcgcaa aagtaccgca cctgatgccg 3840  
 ggtgactcac tgaacctgca gacggetcag gatacggata acggctactc cgtgtttgag 3900  
 cagtcactgc tgcggtatat cgetgcccgg ctgggtgtct cgtatgagca gctttcccgg 3960  
 aattacgcc agatgagcta ctccacggca cgggccagtg cgaacgagtc gtgggcgtac 4020  
 tttatggggc ggcgaaaatt cgtcgcaccc cgtcagcgca gccagatgtt tctgtgctgg 4080  
 ctggaagagg ccatcgttcg ccgctggtg acgttacctt caaaagcgcg cttcagtttt 4140

caggaagccc gcagtgcctg ggggaactgc gactggatag gctccggtcg tatggccatc 4200  
 gatggtctga aagaagttca ggaagcggtg atgctgatag aagccggact gactacctac 4260  
 gagaaagagt gcgcaaacg cggtagcagc tatcaggaat tttttgccc gtaggtccgt 4320  
 gaaacgatgg agcgcctg agccggtctt aaaccgcccc cctgggcggc tgcagcattt 4380  
 gaatccgggc tgcgacaatc aacagaggag gagaagagt acagcagagc tgcgtaatct 4440  
 cccgcatatt gccagcatgg cctttaatga gccgctgatg cttgaacctg cctatgctgc 4500  
 ggttttcttt tgtgctgctg caggccagct tgggatcagc agcctgacgg atgcggtgct 4560  
 cggcgacagc ctgactgccc aggaggcact cgcgacgctg gcattatccg gtgatgatga 4620  
 cggaccacga caggcccgca gttatcaggt catgaacggc atcgcctgctg tgcggtgctc 4680  
 cggcacgctg gtcagccgga cgcgggcgct gcagccgtac tcgggatga ccggttaca 4740  
 cggcattatc gcccgtctgc aacaggctgc cagcgatecc atggtggacg gcattctgct 4800  
 cgatatggac acgcccggcg ggatggtggc gggggcattt gactgctgctg acatcctgc 4860  
 ccgtgtgctg gacataaac cggatggtggc gcttgccaac gacatgaact gcagtgcagg 4920  
 tcagttgctt gccagtgccg cctcccggcg tctggtcacg cagaccgcc ggacaggctc 4980  
 catcggcgtc atgatggctc acagtaatta cggtagctgc ctggagaaac aggggtgtga 5040  
 aatcacgctg atttacagcg gcagccataa ggtggatggc aaccctaca gccatcttc 5100  
 ggatgacgct cgggagacac tgcagtcctg gatggacgca acccgccaga tgtttgcga 5160  
 gaagggtgct gcataaccg gcctgtccgt gcaggttgtg ctggataccg aggtgcagc 5220  
 gtacagcggc caggaggcca ttgatgccg actggtgatg gaacttgta acagcaccga 5280  
 tgcgatcacc gtcatgctg atgactgga tgcagtaaa tcccgtctct caggaggcg 5340  
 aatgacaaa gagactcaat caacaactgt ttcagccact gcttcgagg ctgacgttac 5400  
 tgacgtggtg ccagcagcgg agggcgagaa cgcagcgcg gcgcagccg acgtgaacgc 5460  
 gcagatcacc gcagcggctg cggcagaaaa cagccgcatc atgggatcc tcaactgtga 5520  
 ggaggctcac ggacgcgaag aacaggcacg cgtgctggca gaaaccccc gtatgaccgt 5580  
 gaaaacggcc ccccgcattc tggccgcagc accacagagt gcacagcgc gcagtgcac 5640  
 tgcgctggat cgtctgatg agggggcacc ggcaccgctg gctgcaggta acccgcatc 5700  
 tgatgccgtt aacgatttgc tgaacacacc agtgaaggg atgtttatga cgagcaaaga 5760  
 aacctttacc cattaccagc cgcagggcaa cagtaccgg gctcaccg caaccgccc 5820  
 cggcggattg agtgcgaaag cgctgcaat gaccccgctg atgctggaca cctccagccg 5880  
 taagctgggt gcgtgggatg gcaccaccga cggtagctgc gttggcattc ttgcggttgc 5940  
 tgctgaccag accagcacca cgtgacgct ctacaagtc ggcacgttcc gttatgagga 6000  
 tgtgctctgg ccggaggctg ccagcagcga gacgaaaaa cggaccgctg ttcccgaac 6060  
 ggcaatcagc atcgtttaac tttaccctc atcaactaa gcccctgtg cggtttttt 6120  
 tacgggattt tttatgtcg atgtacaca ccgcccaact gctggcggca aatgagcaga 6180  
 aatttaagt tgatccgctg tttctgctc ttttttccg tgagagctat cccttacc 6240  
 cggagaaagt ctatctctca caaatccgg gactggtaaa catggcctg tacgttctgc 6300  
 cgattgttcc cggtagggtt atccgttccc gtggcggctc cacctctgaa tttaccggc 6360  
 gatatgtcaa gccgaagcat gaagtgaatc cgcagatgac cctgcctgctg ctccggatg 6420  
 aagatccgca gaatctggcg gaccggctt accgcccgcg tcgcatcctc atgcagaaca 6480

tgcgtgacga agagctggcc attgctcagg tcgaagagat gcaggcagtt tctgccgtgc 6540  
ttaagggcaa atacaccatg accggtgaag ctttcgatcc ggttgaggat gatatgggcc 6600  
gcagtgagga gaataacatc acgcagtccg gcggcacgga gtggagcaag cgtgacaagt 6660  
ccacgtatga cccgaccgac gatatcgaag cctacgcgct gaacgccagc ggtgtggtga 6720  
atatcatcgt gttcgatccg aaaggctggg cgctgttccg ttccttcaa gccgtcaagg 6780  
agaagctgga taccgctcgt ggctctaatt ccgagctgga gacagcggat aaagacctgg 6840  
gcaaagcggg gtcctataag gggatgtatg gcgatgtggc catcgtcgtg tattccggac 6900  
agtacgtgga aaacggcgtc aaaaagaact tctgcccga caacacgatg gtgctgggga 6960  
acactcaggc acgcggtctg cgcacctatg gctgcattca ggatgcggac gcacagcgcg 7020  
aaggcattaa cgctctgcc cgttaccga aaaactgggt gaccaccggc gatccggcgc 7080  
gtgagttcac catgattcag tcagcaccgc tgatgctgct ggctgaccct gatgagttcg 7140  
tgtccgtaca actggcgtaa tcatggcctc tcggggccat tgtttctctg tggaggagtc 7200  
catgacgaaa gatgaactga ttgcccgtct ccgctcgtg ggtgaacaac tgaaccgtga 7260  
tgtcagcctg acggggacga aagaagaact ggcgctccgt gtggcagagc tgaagagga 7320  
gcttgatgac acggatgaaa ctgcccgtca ggacaccctc ctcagccggg aaaatgtgct 7380  
gaccggacat gaaaatgagg tgggatcagc gcagccggat accgtgattc tggatacgtc 7440  
tgaactggtc acggtcgtgg cactggtgaa gctgcatact gatgcacttc acgccacgcg 7500  
ggatgaacct gtggcatttg tgctgccggg aacggcgttt cgtgtctctg ccggtgtggc 7560  
agccgaaatg acagagcgcg gcctggccag aatgcaataa cgggagcgcg tgtggctgat 7620  
ttcgataacc tgttcgatgc tgccattgcc cgcgccgatg aaacgatac cgggtacatg 7680  
ggaacgtcag ccaccattac atccggtgag cagtcaggat cgggtgatac tgggtgtttt 7740  
gatgaccctg aaaatatcag ctatgccgga caggcgtgct gcgttgaagg ctccagcccc 7800  
tcctgtttg tccggactga tgaggtgcgg cagctgcggc gtggagacac gctgaccatc 7860  
ggtgaggaaa atttctgggt agatcgggtt tcgccgatg atggcggag ttgtcatctc 7920  
tggcttgac ggggcgtacc gcctgccgtt aaccgtgcc gctgaaagg ggtatgatgg 7980  
ccataaaagg tcttgagcag gccgttgaaa acctcagccg tatcagcaa acggcgggtc 8040  
ctgggtccgc cgcaatggcc attaaccgct ttgcttcatc cgcgatatc cagtcggcgt 8100  
cacaggttgc ccgtgagaca aaggtacgcc ggaaactggt aaaggaaagg gccaggctga 8160  
aaagggccac ggtcaaaaaat ccgcaggcca gaatcaaagt taaccggggg gatttgcccc 8220  
taatcaagct gggtaatgct cgggttgctc tttcgcgcc caggcgtcgt aaaaaggggc 8280  
agcgttcatc cctgaaagggt ggcggcagcg tgcttgggt gggtaacctg cgtattcccc 8340  
gcgcgtttat tcagcaactg aaaaatggcc ggtggcatgt catgcagcgt gtggctggga 8400  
aaaaccgtta cccattgat gtggtgaaaa tcccgatggt ggtgccgctg accacggcgt 8460  
ttaaacaana tattgagcgg atacggcgtg aacgtcttc gaaagagctg ggctatgcgc 8520  
tgcagcatca actgaggatg gtaataaagc gatgaaacat actgaactcc gtgcagccgt 8580  
actggatgca ctggagaagc atgacaccgg ggcgacgttt tttgatggc gccccgctgt 8640  
ttttgatgag gcggatttcc cggcagttgc cgtttatctc accggcgtg aatacacggg 8700  
cgaagagctg gacagcgata cctggcaggc ggagctgcat atcgaagttt tctgctgc 8760  
tcaggtgccg gattcagagc tggatgcgtg gatggagtcc cggatttate cggatgatgag 8820

cgatatcccg gcactgtcag atttgatcac cagtatgggt gccagecggt atgactaccg 8880  
gcgcgacgat gatgcgggct tgtggagttc agccgatctg acttatgtca ttacctatga 8940  
aatgtgagga cgctatgcct gtaccaaate ctacaatgcc ggtgaaaggt gccgggacca 9000  
ccctgtgggt ttataagggg agcggtgacc cttacgcgaa tccgcttca gacgttgact 9060  
ggtcgcgtct ggcaaaagtt aaagacctga cgcccggcga actgaccgct gagtcctatg 9120  
acgacagcta tctcgatgat gaagatgcag actggactgc gaccgggcag gggcagaaat 9180  
ctgccggaga taccagcttc acgctggcgt ggatgcccgg agagcagggg cagcagggcgc 9240  
tgctggcgtg gtttaatgaa ggcgataccc gtgcctataa aatccgcttc ccgaacggca 9300  
cggtcgatgt gttccgtggc tgggtcagca gtatcggtaa ggcggtgacg gcgaaggaag 9360  
tgatcacccg cacggtgaaa gtcaccaatg tgggacgtcc gtcgatggca gaagatcgca 9420  
gcacggtaac agcggcaacc ggcatgaccg tgacgcctgc cagcacctcg gtggtgaaag 9480  
ggcagagcac cacgctgacc gtggccttc agccggaggg cgtaaccgac aagagcttcc 9540  
gtgcgggtgc tgcggataaa acaaaagcca ccgtgtcgggt cagtggatg accatcaccg 9600  
tgaacggcgt tgctgcagge aaggtcaaca ttccggttgt atccggtaat ggtgagtttg 9660  
ctgcggttgc agaaattacc gtcaccgcca gttaatccgg agagtcagcg atgttctctga 9720  
aaaccgaatc atttgaacat aacgggtgta ccgtcacgct ttctgaactg tcagccctgc 9780  
agcgcattga gcatctcgcc ctgatgaaac ggcaggcaga acaggcggag tcagacagca 9840  
accggaagtt tactgtggaa gacgccatca gaaccggcgc gtttctggtg gcgatgtccc 9900  
tgtggcataa ccatccgcag aagacgcaga tgccgtccat gaatgaagcc gttaaacaga 9960  
ttgagcagga agtgcttacc acctggccca cggaggcaat ttctcatgct gaaaacgtgg 10020  
tgtaccggct gtctggatg tatgagtttg tggtgataa tgcccctgaa cagacagagg 10080  
acgccgggcc cgacagacct gtttctgcgg gaaagtgttc gacggtgagc tgagttttgc 10140  
cctgaaactg gcgcgtgaga tggggcgacc cgactggcgt gccatgcttg ccgggatgtc 10200  
atccacggag tatgccgact ggcaccgctt ttacagtacc cattattttc atgatgttct 10260  
gctggatatg cacttttccg ggctgacgta caccgtgctc agcctgtttt tcagcgatcc 10320  
ggatatgcat ccgctggatt tcagtctgct gaaccggcgc gaggctgacg aagagcctga 10380  
agatgatgtg ctgatgcaga aagcggcagg gcttgccgga ggtgtccgct ttggcccgga 10440  
cgggaatgaa gttatccccg cttccccgga tgtggcggac atgacggagg atgacgtaat 10500  
gctgatgaca gtatcagaag ggatcgcagg aggagtccgg tatggctgaa ccggtaggcg 10560  
atctggctgt tgatttgagt ctggatgcgg ccagatttga cgagcagatg gccagagtca 10620  
ggcgtcattt ttctggtacg gaaagtgat cgaaaaaac agcggcagtc gttgaacagt 10680  
cgctgagccg acaggcgtg gctgcacaga aagcggggat ttccgtcggg cagtataaag 10740  
ccgccatgcg tatgctgcct gcacagttea ccgacgtggc cacgcagctt gcaggcgggc 10800  
aaagtccgtg gctgatcctg ctgcaacagg gggggcaggt gaaggactcc ttcggcggga 10860  
tgatcccat gttcaggggg cttgcccgtg cgatcacct gccgatggtg ggggccacct 10920  
cgctggcgggt ggcgaccggt gcgctggcgt atgcctggta tcagggaac tcaacctgt 10980  
ccgatttcaa caaacgctg gtccttccg gcaatcagge gggactgac gcagatcgta 11040  
tgctggctct gtccagagcc gggcaggcgg cagggtgac gtttaaccag accagcagat 11100  
cactcagcgc actggttaag gcgggggtaa gcggtgagge tcagattgcg tccatcagcc 11160

agagtgtggc gcgtttctcc tctgcatccg gcgtggaggt ggacaaggte gctgaagcct 11220  
 tcgggaagct gaccacagac ccgacgtcgg ggctgacggc gatggctcgc cagttccata 11280  
 acgtgtcggc ggagcagatt gcgtatgttg ctcaattgca gcgttccggc gatgaagccg 11340  
 gggcattgca ggccggcgaac gaggccgcaa cgaaagggtt tgatgaccag acccggccgc 11400  
 tgaaagagaa catgggcacg ctggagacct gggcagacag gactgcgcgg gcattcaaat 11460  
 ccatgtggga tgcggtgctg gatattggtc gtctgatac cgcgcaggag atgctgatta 11520  
 aggcagaggc tgcgtataag aaagcagacg acatctggaa tctgcgcaag gatgattatt 11580  
 ttgttaacga tgaagcgcgg gcgcgttact gggatgatcg tgaaaaggcc cgtcttgcgc 11640  
 ttgaagccgc ccgaaagaag gctgagcagc agactcaaca ggacaaaaat gcgcagcagc 11700  
 agagcgatac cgaagcgtca cggctgaaat ataccgaaga ggcgcagaag gcttacgaac 11760  
 ggctgcagac gccgctggag aaatataaccg cccgtcagga agaactgaac aaggcactga 11820  
 aagacgggaa aatcctgcag gcggattaca acacgctgat ggcggcggcg aaaaaggatt 11880  
 atgaagcgac gctgaaaaag ccgaaacagt ccagcgtgaa ggtgtctgcg ggcgatcgtc 11940  
 aggaagacag tgctcatgct gccctgetga cgettcaggc agaactccgg acgctggaga 12000  
 agcatgccgg agcaaatgag aaaatcagcc agcagcggcg ggatttgtgg aaggcggaga 12060  
 gtcagttcgc ggtactggag gaggcggcgc aacgtcgcca gctgtctgca caggagaaat 12120  
 ccctgctggc gcataaagat gagacgctgg agtaciaaac ccagctggct gcacttggcg 12180  
 acaaggttac gtatcaggag gcctgaacg cgctggcgca gcaggcggat aaattcgcac 12240  
 agcagcaacg ggcaaacgg gccgccattg atgcgaaaag ccgggggctg actgaccggc 12300  
 aggcagaacg ggaagccacg gaacagcgcc tgaaggaaca gtatggcgat aatccgctgg 12360  
 cgctgaataa cgtcatgca gagcagaaaa agacctgggc ggctgaagac cagcttcgcg 12420  
 ggaactggat ggcaggcctg aagtccggct ggagtgagtg ggaagagagc gccacggaca 12480  
 gtatgtcgca ggtaaaaagt gcagccacgc agaccttga tggattgca cagaatatgg 12540  
 cggcgatgct gaccggcagt gagcagaact ggcgcagett cacccttcc gtgctgtcca 12600  
 tgatgacaga aattctgctt aagcaggcaa tggtagggat tgcggggagt atcggcagcg 12660  
 ccattggcgg ggctgttggg ggcggcgcac ccgcgtcagg cgttacagcc attcaggccg 12720  
 ctgcggcgaa attccatttt gcaaccggag gatttacggg aaccggcggc aaatatgagc 12780  
 cagcggggat tgttcaccgt ggtgagttg tcttcacgaa ggaggcaacc agccgattg 12840  
 gcgtggggaa tctttaccgg ctgatgcgcg gctatgccac cggcggttat gtcggtacac 12900  
 cgggcagcat ggcagacagc cggtcgcagg cgtccgggac gtttgagcag aataaccatg 12960  
 tggtgattaa caacgacggc acgaacgggc agataggtec ggctgctctg aaggcgggtg 13020  
 atgacatggc ccgcaagggt gccctgatg aaattcagac acagatgcgt gatggtggcc 13080  
 tgttctccgg aggtggacga tgaagacctt ccgctggaaa gtgaaaccgg gtatggatgt 13140  
 ggcttcggtc cttctgtaa gaaaggtgcg ctttggatg ggctattctc agcgagcgcc 13200  
 tgccgggctg aatgccaacc tgaaaacgta cagcgtgacg ctttctgtcc cccgtgagga 13260  
 ggccacggta ctggagtcgt ttctggaaga gcacgggggc tggaatcct ttctgtggac 13320  
 gccgccttat gagtggcggc agataaaggt gacctgcgca aaatggctgt cgcgggtcag 13380  
 tatgctcgt gttgagtca gcgcagagtt tgaacaggtg gtgaactgat gcaggatata 13440  
 cggcaggaaa cactgaatga atgcaccctg gcggagcagt cggccagcgt ggtgctctgg 13500

gaaatcgacc tgacagaggt cggtaggaaa cgttattttt tctgtaatga gcagaacgaa 13560  
 aaaggtgagc cggtcacctg gcaggggaga cagtatcagc cgtatcccat tcaggggagc 13620  
 ggttttgaac tgaatggcaa aggcaccagt acgcgcecca cgctgacggt ttctaacctg 13680  
 tacggtatgg tcaccgggat ggcggaagat atgcagagtc tggtcggcgg aacggtggtc 13740  
 cggcgtaagg ttacgcccg tttctggat gcggtgaact tcgtcaacgg aaacagttac 13800  
 gccgatccgg agcaggaggt gatcagccgc tggcgcattg agcagtgcag cgaactgagc 13860  
 gcggtgagtg cctcctttgt actgtccacg ccgacggaaa cggatggcgc tgtttttccg 13920  
 ggacgtatca tgctggccaa cacctgcacc tggacctatc gcggtgacga gtgcggttat 13980  
 agcggtcagg ctgtcgcgga tgaatatgac cagccaacgt ccgatatac gaaggataaa 14040  
 tgcagcaaat gcctgagcgg ttgtaagttc cgcaataacg tcggcaactt tggcggcttc 14100  
 ctttccatta acaaaccttc gcagtaaate ccatgacaca gacagaatca gcgattctgg 14160  
 cgcacgcccg gcgatgtgag ccagcggagt cgtgcggctt cgtggtaagc acgccggagg 14220  
 gggaaagata tttcccctgc gtgaatatct ccggtgagcc ggaggctatt tccgtatgtc 14280  
 gccggaagac tggctgcagg cagaaatgca gggtagatt gtggcgtgg tccacagcca 14340  
 ccccggtggc ctgccctggc tgagtgagcc cgaccggcgg ctgcaggtgc agagtgattt 14400  
 gccgtggtag ctggtctgcc gggggacgat tcataagttc cgctgtgtgc cgcattctac 14460  
 cgggcggcgc tttgagcacg gtgtgacgga ctgttacaca ctgttccggg atgcttatca 14520  
 tctggcgggg attgagatgc cggactttca tcgtgaggat gactggtggc gtaacggcca 14580  
 gaatctctat ctggataatc tggaggcagc ggggctgtat caggtgccgt tgtcagcggc 14640  
 acagccgggc gatgtgctgc tgtgctgttt tggttcatca gtgccgaatc acgccgaat 14700  
 ttactgcggc gacggcgagc tgctgcacca tattctgaa caactgagca aacgagagag 14760  
 gtacaccgac aaatggcagc gacgcacaca ctccctctgg cgtcaccggg catggcgcgc 14820  
 atctgccttt acggggattt acaacgattt ggtcgcgca tcgacctcg tgtgaaaacg 14880  
 ggggctgaag ccatccgggc actggccaca cagctcccgg cgtttcgtca gaaactgagc 14940  
 gacggctggc atcaggtacg gattgccggg cgggacgtca gcacgtccgg gttaacggcg 15000  
 cagttacatg agactctgcc tgatggcgtc gtaattcata ttgttcccag agtcgccggg 15060  
 gccaaagcag gtggcgtatt ccagattgtc ctgggggctg ccgccattgc cggatcattc 15120  
 tttaccgccg gagccaccct tgacagatgg ggggcagcca ttggggccgg tggatgacc 15180  
 ggcatcctgt tttctctcgg tgccagtatg gtgctcgggt gtgtggcga gatgctggca 15240  
 ccgaaagcca gaactccccg tatacagaca acgataacg gtaagcagaa cacctatttc 15300  
 tcctcactgg ataacatggt tgcccagggc aatgtttctg ctgttctgta cggggaaatg 15360  
 cgcgtggggc cacgcgtggt ttctcaggag atcagcacgg cagacgaagg ggacgggtgg 15420  
 caggttggg tgattggctg ctgatgcaaa atgttttatg tgaaccgcc tcggggcgg 15480  
 tttgtcattt atggagcgtg aggaatgggt aaaggaagca gtaaggggca tccccgcgc 15540  
 gaagcgaagg acaacctgaa gtccacgcag ttgctgagtg tgatcgatgc catcagcga 15600  
 gggccgattg aaggtccggg ggatggetta aaaagcgtgc tgctgaacag tacgccggtg 15660  
 ctggacactg aggggaatac caacatatcc ggtgtcacgg tgggtttccg ggctggtgag 15720  
 caggagcaga ctccgccgga gggatttgaa tcctccgget ccgagacggt gctgggtacg 15780  
 gaagtgaat atgacacgcc gatcaccgc accattacgt ctgcaaacat cgaccgtctg 15840

cgctttacct tcggtgtaca ggcactggtg gaaaccacct caaagggtga caggaatccg 15900  
 tcggaagtcc gcctgctggt tcagatacaa cgtaacggtg gctgggtgac ggaaaaagac 15960  
 atcaccatta agggcaaaac cacctcgcag tatctggcct cggtggtgat gggtaacctg 16020  
 ccgccgcgcc cgtttaatat ccggatgcgc aggatgacgc cggacagcac cacagaccag 16080  
 ctgcagaaca aaacgctctg gtcgtcatac actgaaatca tcgatgtgaa acagtgtctac 16140  
 ccgaacacgg cactggtcgg cgtgcaggtg gactcggagc agttcggcag ccagcaggtg 16200  
 agccgtaatt atcatctgcg cgggcgtatt ctgcaggtgc cgtcgaacta taaccgcag 16260  
 acgcggcaat acagcggat ctgggacgga acgtttaaac cggcatacag caacaacatg 16320  
 gcctggtgtc tgtgggatat gctgacccat ccgcgctacg gcatggggaa acgtcttgg 16380  
 gcggcggatg tggataaatg ggcgctgtat gtcateggcc agtactgca ccagtcagtg 16440  
 ccggacggct ttggcggcac ggagccgcgc atcaactgta atgcgtacct gaccacacag 16500  
 cgtaaggcgt gggatgtgct cagcgatttc tgctcggcga tgcgctgtat gccggtatgg 16560  
 aacgggcaga cgctgacgtt cgtgcaggac cgaccgtcgg ataagacgtg gacctataac 16620  
 cgcagtaatg tggatgatgcc ggatgatgga gcgccgttc gctacagctt cagcgccctg 16680  
 aaggaccgcc ataatgccgt tgaggtgaa tggattgacc cgaacaacgg ctgggagacg 16740  
 gcgacagagc ttgttgaaga tacgcaggcc attgccgtt acggtcgtaa tgttacgaag 16800  
 atggatgcct ttggctgtac cagccggggg caggcacacc gcgccgggct gtggctgatt 16860  
 aaaacagaac tgctggaaac gcagaccgtg gatttcagcg tcggcgcaga agggcttcgc 16920  
 catgtaccgg gcgatgttat tgaatctgc gatgatgact atgccggtat cagcaccggt 16980  
 ggtcgtgtgc tggcggtgaa cagccagacc cggacgctga cgctcgaccg tgaatcacg 17040  
 ctgccatcct ccgtaccgc gctgataagc ctggttgacg gaagtggcaa tccggtcagc 17100  
 gtggaggttc agtccgtcac cgacggcgtg aaggtaaaag tgagccgtgt tctgacggt 17160  
 gttgctgaat acagcgtatg ggagctgaag ctgccgacgc tgcgccagcg actgttccgc 17220  
 tgcgtgagta tccgtgagaa cgacgacggc acgtatgcca tcaccgccgt gcagcatgtg 17280  
 ccggaaaaag aggccatcgt ggataacggg gcgcactttg acggcgaaca gactggcacg 17340  
 gtgaatggtg tcacgccgc agcgggtcag cacctgaccg cagaagtcac tgcagacagc 17400  
 ggggaatatc aggtgctggc gcgatgggac acaccgaagg tggatgaagg cgtgagtttc 17460  
 ctgctccgtc tgaccgtaac agcggacgac ggcagtgagc ggctggtcag cacggcccgg 17520  
 acgacggaaa ccacataccg cttcacgcaa ctggcgtgga ggaactacag gctgacagtc 17580  
 cggcgggtaa atgcgtgggg gcagcagggc gatccggcgt cggtatcgtt ccgattgcc 17640  
 gcaccggcag caccgtcag gattgagctg acgccgggct attttcagat aaccgccacg 17700  
 ccgcattctg ccgttatga cccgacggta cagtttgagt tctggttctc ggaaaagcag 17760  
 attgcggata tcagacaggt tgaaccagc acgcgttate ttggtacggc gctgtactgg 17820  
 atagccgcca gtatcaatat caaacgggc catgattatt acttttatat ccgcagtgtg 17880  
 aacaccgttg gcaaatcggc attcgtggag gccgtcggtc gggcgcgcga tgatgcggaa 17940  
 ggttacctgg atttttcaa aggcaagata accgaatccc atctcggcaa ggagctgctg 18000  
 gaaaaagtgc agctgacgga ggataacgcc agcagactgg aggagtttc gaaagagtgg 18060  
 aaggatgcca gtgataagt gaatgccatg tgggctgtca aaattgagca gaccaaagac 18120  
 ggcaaacatt atgtcgcggg tattggcctc agcatggagg acacggagga aggcaactg 18180

agccagtttc tggttgccgc caatcgtatc gcatttattg acceggcaaa cgggaatgaa 18240  
 acgccgatgt ttgtggcgca gggcaaccag atattcatga acgacgtggt cctgaagcgc 18300  
 ctgacggccc ccaccattac cagcggcggc aatcctccgg ccttttcctt gacaccggac 18360  
 ggaaagctga ccgctaaaaa tgcggatatac agtggcagtg tgaatgcgaa ctccgggacg 18420  
 ctcagtaatg tgacgatagc tgaaaactgt acgataaacg gtacgctgag ggcggaaaaa 18480  
 atcgtcgggg acattgtaaa ggcggcgagc gcggcttttc cgcgccagcg tgaagcagtg 18540  
 gtggactggc cgtcaggtag ccgtactgtc accgtgaccg atgaccatcc ttttgatcgc 18600  
 cagatagtagg tgcttccgct gacgtttcgc ggaagtaagc gtactgtcag cggcaggaca 18660  
 acgtattcga tgtgttatct gaaagtactg atgaacgggt cgggtgattt tgatggcgcg 18720  
 gcgaacgagg cggtagcagg gttctcccgt attgttgaca tgccagcggg tcggggaaac 18780  
 gtgatcctga cgttcacgct tacgtccaca cggcattcgg cagatattcc gccgtatacg 18840  
 tttgccagcg atgtgcaggc tatgggtgatt aagaaacagg cgctgggcat cagcgtggtc 18900  
 tgagtgtgtt acagaggctt gtccgggaac gggcgtttta ttataaaaca gtgagagggtg 18960  
 aacgatgcgt aatgtgtgta ttgccgttgc tgtctttgcc gcacttgccg tgacagtcac 19020  
 tccggcccgt gcggaagggt gacatggtac gtttacgggt ggctattttc aagtgaace 19080  
 gggtagattg ccgtcgttgt cgggcgggga taccggtgtg agtcatctga aagggattaa 19140  
 cgtgaagtac cgttatgagc tgacggacag tgtgggggtg atggcttccc tggggttcgc 19200  
 cgcgtcgaaa aagagcagca cagtgatgac cggggaggat acgtttcact atgagagcct 19260  
 gcgtggacgt tatgtgagcg tgatggccgg accggtttta caaatcagta agcaggtcag 19320  
 tgcgtacgcc atggccggag tggctcacag tcggtggtcc ggcagtaca tggattaccg 19380  
 taagacggaa atcactccc ggtatatgaa agagacgacc actgccaggg acgaaagtgc 19440  
 aatgcggcat acctcagtag cgtggagtgc aggtatacag attaatccgg cagcgtccgt 19500  
 cgttgttgat attgcttatg aaggctccgg cagtggcgac tggcgtactg acggattcat 19560  
 cgttggggtc ggtataaat tctgattagc caggtaacac agtgttatga cagcccgccg 19620  
 gaaccggtgg gcttttttgt ggggtgaata tggcagtaaa gatttcagga gtcctgaaag 19680  
 acggcacagg aaaaccggtg cagaactgca ccattcagct gaaagccaga cgtaacagca 19740  
 ccacgggtgg ggtgaacacg gtgggctcag agaatccgga tgaagccggg cgttacagca 19800  
 tggatgtgga gtacggtagc tacagtgtca tctgcagggt tgacggtttt ccaccatcgc 19860  
 acgccgggac catcaccgtg tatgaagatt cacaaccggg gacgctgaat gattttctct 19920  
 gtgccatgac ggaggatgat gcccggccgg aggtgctgag tcgtcttgaa ctgatggtgg 19980  
 aagagggtgg gcgtaacgcg tccgtggtgg cacagagtac ggcagacgcg aagaaatcag 20040  
 ccggcagatc cagtgcacga gctgctcagg tcgcggccct tgtgactgat gcaactgact 20100  
 cagcacgcgc cgccagcacg tccgccggac aggetgcate gtcagctcag gaagcgtcct 20160  
 ccggcgcaga agcggcatca gcaaaggcca ctgaagcgga aaaaagtgcc gcagccgcag 20220  
 agtcctcaaa aaacgcggcg gccaccagtg ccggtgcggc gaaaacgtca gaaacgaatg 20280  
 ctgcagcgtc acaacaatca gccgccacgt ctgectccac cgcggccacg aaagcgtcag 20340  
 aggccgccac ttcagcacga gatgcggtgg cctcaaaaga ggcagcaaaa tcatcagaaa 20400  
 cgaacgcate atcaagtgcc ggtcgtgcag ctctctcggc aacggcggca gaaaattctg 20460  
 ccagggcggc aaaaacgtcc gagacgaatg ccaggtcate tgaacagca gcggaacgga 20520

gcgcctctgc cgcggcagac gcaaaaacag cggcggcggg gagtgcgtca acggcatcca 20580  
 cgaaggcgac agaggctgcg ggaagtgcgg tatcagcatc gcagagcaaa agtgcggcag 20640  
 aagcggcggc aatacgtgca aaaaattcgg caaacgtgc agaagatata gcttcagctg 20700  
 tcgcgcttga ggatgcggac acaacgagaa aggggatagt gcagctcagc agtgcaacca 20760  
 acagcacgtc tgaaacgctt gctgcaacgc caaaggcggg taaggtggta atggatgaaa 20820  
 cgaacagaaa agcccactgg acagtccggc actgaccgga acgccaacag caccaaccgc 20880  
 gctcagggga acaaacataa cccagattgc gaacaccgct tttgtactgg ccgcgattgc 20940  
 agatgttatac gacgcgtcac ctgacgcaact gaatacgtc aatgaactgg ccgcagcgtc 21000  
 cggaatgat ccagattttg ctaccacat gactaacgc cttgcgggta aacaaccgaa 21060  
 gaatgcgaca ctgacggcgc tggcagggct ttccacggcg aaaaataaat taccgtattt 21120  
 tgcggaaaat gatgccgcca gctgactga actgactcag gttggcaggg atattctggc 21180  
 aaaaaattcc gttgcagatg ttcttgaata cttggggcc ggtgagaatt cggcctttcc 21240  
 ggcaggtgcg ccgatcccgt ggccatcaga tategttccg tctggctacg tctgatgca 21300  
 ggggcaggcg tttgacaaat cagcctacc aaaacttget gtcgcgtatc catcgggtgt 21360  
 gcttctgat atgcgaggt ggacaatcaa ggggaaacc gccagcggtc gtgctgtatt 21420  
 gtctcaggaa caggatggaa ttaagtcgca cacccacagt gccagtgcac ccggtacgga 21480  
 tttggggacg aaaaccacat cgtcgtttga ttacgggacg aaaacaacag gcagtttcga 21540  
 ttacggcacc aaatcgacga ataacacggg ggctcatgct cacagtctga gcggttcaac 21600  
 aggggcccg ggtgctcatg cccacacaag tggtttaagg atgaacagtt ctggctggag 21660  
 tcagtatgga acagcaacca ttacaggaag tttatccaca gttaaaggaa ccagcacaca 21720  
 gggatttgct tatttatcga aaacggacag tcagggcagc cacagtcact cattgtccgg 21780  
 tacagccgtg agtgccggg cacaatgcgca tacagttggt attggtgcgc accagcatcc 21840  
 ggttgcttatac ggtgctcatg cccattcttt cagtattggt tcacacggac acaccatcac 21900  
 cgtaaacgct gcgggtaacg cggaaaacac cgtcaaaaac attgcattta actatattgt 21960  
 gaggcttgca taatggcatt cagaatgagt gaacaaccac ggaccataaa aatttataat 22020  
 ctgctggccg gaactaatga atttattggt gaaggtgacg catatattcc gcctcatacc 22080  
 ggtctgcctg caaacagtac cgatattgca ccgccagata ttccggctgg ctttgtggct 22140  
 gttttcaaca gtgatgaggc atcgtggcat ctcttgaag accatcgggg taaaaccgtc 22200  
 tatgacgtgg cttccggcga cgcgttattt atttctgaac tcggtccgtt accgaaaat 22260  
 tttacctggt tatcgcggg aggggaatat cagaagtgga acggcacagc ctgggtgaag 22320  
 gatacggag cagaaaaact gttccggatc cgggaggcgg aagaaacaaa aaaaagcctg 22380  
 atgcaggtag ccagtgcgca tattgcgcc cttcaggatg ctgcagatct ggaaattgca 22440  
 acgaaggaag aaacctcgtt gctggaagcc tggaagaagt atcgggtggt gctgaaccgt 22500  
 gttgatacat caactgcacc tgatattgag tgacctgctg tcctgttat ggagtaatcg 22560  
 ttttgtgata tgccgcagaa acgttgtatg aaataacgtt ctgcggttag ttagtatatt 22620  
 gtaaagctga gtattggttt atttggcgtt tattatcttc aggagaataa tggaagtctt 22680  
 atgactcaat tgttcatagt gtttacaatc ccgccaattg cttttaagac tgaacgatg 22740  
 aaatatgggt tttcgtcatg ttttgagtct gctgttgata tttctaaagt cggttttttt 22800  
 tcttcgtttt ctctaactat tttccatgaa atacattttt gattattatt tgaatcaatt 22860

ccaattacct gaagctcttc atctataatt ggcattgtat gtattggttt attggagtag 22920  
 atgcttgctt ttctgagcca tagctctgat atccaaatga agccataggc atttgttatt 22980  
 ttggctctgt cagctgcata acgccccaaa atatatttat ctgcttgatc ttcaaagtgt 23040  
 gtattgatta aatcaattgg atggaattgt ttatcataaa aaattaatgt ttgaatgtga 23100  
 taaccgtcct ttaaaaaagt cgtttctgca agcttggctg tatagtcaac taactcttct 23160  
 gtcgaagtga ttttttagg cttatctacc agttttagac gctctttaat atcttcagga 23220  
 attattttat tgcataattg tatcatgcta aatgacaatt tgcttatgga gtaatctttt 23280  
 aattttaaat aagttattct cctggcttca tcaaataaag agtcgaatga tgttggcgaa 23340  
 atcacatcgt cacccattgg attgtttatt tgtatgcca gagagttaca gcagttatac 23400  
 attctgccat agattatagc taaggcatgt aataatcgt aatcttttag cgtatttagc 23460  
 acccatcgtc tttctgattt aataatagat gattcagtta aatatgaagg taatttcttt 23520  
 tgtgcaagtc tgactaaact ttttatacca atgtttaaca tactttcatt tgtaataaac 23580  
 tcaatgtcat tttcttcaat gtaagatgaa ataagagtag cctttgctc gctatacatt 23640  
 tctaaatcgc ctgttttctc tctcgtattg cgagaatttt tagcccaagc cattaatgga 23700  
 tcatttttcc atttttcaat aacattattg ttataccaaa tgcataatcc tataatctgg 23760  
 tttttgtttt ttgaaataat aaatgttact gttcttgctg tttggaggaa ttgattcaaa 23820  
 ttcaagcgaa ataattcagg gtcaaaaat gatacaatgc agcatttgag caagtgcgat 23880  
 aaatctttaa gtcttcttcc ccatggtttt ttagtcataa aactctccat tttgataggt 23940  
 tgcattgctg atgctgatat attttagagg tgataaaatt aactgcttaa ctgtcaatgt 24000  
 aatacaagtt gtttgatctt tgcaatgatt cttatcagaa accatatagt aaattagtta 24060  
 cacaggaaat ttttaaatatt attattatca ttcatatgt attaaaatta gagttgtggc 24120  
 ttggctctgc taacacgttg ctcataggag atatggtaga gccgcagaca cgtcgtatgc 24180  
 aggaacgtgc tgcggctggc tggngaactt ccgatagtgc ggggtgtgaa tgatttccag 24240  
 ttgctaccga ttttacatat tttttgcatg agagaatttg taccacctcc caccgacct 24300  
 ctatgactgt acgccactgt ccctaggact gctatgtgcc ggagcggaca ttacaaactg 24360  
 ccttctcggg catgccact gttgccaatg acctgcctag gaattggta gcaagttact 24420  
 accggatttt gtaaaaacag cctcctcat ataaaaagta ttcgttact tccgataagc 24480  
 gtcgtaattt tctatcttcc atcatattct agatcctct gaaaaaatct tccgatttg 24540  
 ctaggcaactg atacataact cttttccaat aattggggaa gtcattcaaa tctataatag 24600  
 gtttcagatt tgcttcaata aattctgact gtagctgctg aaacgttgcg gttgaactat 24660  
 atttccttat aacttttacg aaagagtctt tttgagtaat cacttctc aagtgcctcc 24720  
 ctgcctccaa acgataactg ttagcaatat ttaatagctt gaaatgatga agagctctgt 24780  
 gtttgtcttc ctgcctccag ttcgcccggc attcaacata aaaactgata gcaccggag 24840  
 ttccggaaac gaaattgca tataaccatt gtcacgaaa aaaaatgtcc ttgtcgatat 24900  
 agggatgaat cgcttgggtg acctcateta ctgcgaaaac ttgaccttc tctccatat 24960  
 tgcagtcgcg gcacgatgga actaaattaa taggcatcac cgaaaattca ggataatgtg 25020  
 caataggaag aaaaatgatc atattttttg tctgtctat atcaccaca aatggacatt 25080  
 tttcacctga tgaacaagc atgtcatcgt aatatgtct agcgggtttg tttttatctc 25140  
 ggagattatt ttcataaagc ttttctaatt taaccttgt caggttacca actactaagg 25200

ttgtaggctc aagaggggtgt gtcctgtcgt aggtaaataa ctgacctgctc gagcttaata 25260  
 ttctatattg ttgttcttcc tgcaaaaaag tggggaagtg agtaatgaaa ttatttctaa 25320  
 catttatctg catcatacct tccgagcatt tattaagcat ttcgctataa gttctcgcctg 25380  
 gaagaggtag ttttttcatt gtactttacc ttcatctctg ttattatca tcgcttttaa 25440  
 aacggttcga ccttctaate ctatctgacc attataattt tttagaatgg tttcataaga 25500  
 aagctctgaa tcaacggact gcgataataa gtgggtggtat ccagaatttg tcacttcaag 25560  
 taaaaacacc tcacgagtta aaacacctaa gttctcaccg aatgtctcaa tatccggacg 25620  
 gataatattt attgcttctc ttgaccgtag gactttccac atgcaggatt ttggaacctc 25680  
 ttgcagtact actggggaat gagttgcaat tattgctaca ccattgcctg catcgagtaa 25740  
 gtcgcttaat gttcgtaaaa aagcagagag caaaggtgga tgcagatgaa cctctggttc 25800  
 atcgaataaaa actaatgact tttcgccaac gacatctact aatcttgtga tagtaaataa 25860  
 aacaattgca tgtccagagc tcattcgaag cagatatttc tggatattgt cataaaaca 25920  
 tttagtgaat ttatcatcgt ccaettgaaat ctgtggttca ttacgtctta actcttcata 25980  
 tttagaaatg aggcctgatga gttccatatt tgaaaagttt tcatcactac ttagtttttt 26040  
 gatagcttca agccagagtt gtctttttct atctactctc atacaacca taaatgctga 26100  
 aatgaattct aagcggagat cgcctagtga ttttaacta ttgctggcag cattcttgag 26160  
 tccaataaaa aagtattgtg taccttttgc tgggtcaggt tgttctttag gaggagtaaa 26220  
 aggatcaaat gcactaaacg aaactgaaac aagcagatga aaatatccct ttgggattct 26280  
 tgactcgata agtctattat tttcagagaa aaaatattca ttgttttctg ggttggtgat 26340  
 tgcaccaate attccattca aaattgttgt tttaccacac ccattccgcc cgataaaagc 26400  
 atgaatgttc gtgctgggca tagaattaac cgtcacctca aaaggtatag ttaaactact 26460  
 gaatccggga gcacttttcc tattaaatga aaagtggaaa tctgacaatt ctggcaaacc 26520  
 atttaacaca cgtgcgaact gtccatgaat ttctgaaaga gttaccctc taagtaatga 26580  
 ggtgttaagg acgctttcat tttcaatgtc ggctaatcga tttggccata ctactaaatc 26640  
 ctgaatagct ttaagaaggt tatgtttaaa accatcgctt aatttgctga gattaacata 26700  
 gtagtcaatg ctttcaccta aggaaaaaaaa catttcaggg agttgactga attttttatc 26760  
 tattaatgaa taagtgtcta cttcttcttt ttgacctaca aaaccaattt taacatttcc 26820  
 gatatcgcat ttttcacat gctcatcaaa gacagtaaga taaaacattg taacaaagga 26880  
 atagtcatc caaccatctg ctctagtagaa tgccttattt tttctactg caggaatata 26940  
 cccgcctctt tcaataaac taaactccaa catatagtaa cccttaattt tattaaaata 27000  
 accgcaattt atttgcggcg aacacaggat ctctctttta agttactctc tattacatac 27060  
 gttttccatc taaaaattag tagtattgaa cttaacgggg catcgtattg tagttttcca 27120  
 tatttagctt tctgcttctt tttggataac ccactgttat tcatgttgca tgggtgactg 27180  
 tttataccaa cgatatagtc tattaatgca tatatagtat cgccgaacga ttagctcttc 27240  
 aggttctgga agaagcgttt caagtactaa taagccgata gatagccacg gacttcgtag 27300  
 ccatttttca taagtgttaa ctccctctc tctctcctaa cagacattca ctacagttat 27360  
 ggcggaaagg tatgcatgct ggggtgtgggg aagtcgtgaa agaaaagaag tcagctgcgt 27420  
 cgtttgacat cactgctatc ttcttactgg ttatgcaggt cgtagtgggt ggcacacaaa 27480  
 gctttgcaact ggattgcgag gctttgtgct tctctggagt gcgacaggtt tgatgacaaa 27540

aaattagcgc aagaagacaa aaatcacctt gcgetaatgc tctgttacag gtcactaata 27600  
 ccatctaagt agttgattca tagtgactgc atatgttgtg ttttacagta ttatgtagtc 27660  
 tgttttttat gcaaaatcta atttaataata ttgatattta tatcatttta cgtttctcgt 27720  
 tcagcttttt tatactaagt tggcattata aaaaagcatt gcttatcaat ttgttgcaac 27780  
 gaacaggcca ctatcagtca aaataaaatc attatttggat ttcaattttg tcccactccc 27840  
 tgcctctgtc atcacgatac tgtgatgcca tgggtgccga cttatgcccc agaagatggt 27900  
 gagcaaaact atcgcttata tgccttctcat agagtcttgc agacaaactg cgcaactcgt 27960  
 gaaaggtagg cggatcccct tcgaaggaaa gacctgatgc ttttcgtgcg cgcataaaat 28020  
 accttgatac tgtgccggat gaaagcgggt cgcgacgagt agatgcaatt atggtttctc 28080  
 cgccaagaat ctctttgcat ttatcaagtg tttcttctcat tgatattccg agagcatcaa 28140  
 tatgcaatgc tgttgggatg gcaattttta cgctgtttt gctttgctcg acataaagat 28200  
 atccatctac gatatacagac cacttcattt cgcataaatc accaactcgt tgccccgtaa 28260  
 caacagccag ttccattgca agtctgagcc aacatggtga tgattctgct gcttgataaa 28320  
 ttttcaggta ttcgtcagcc gtaagtcttg atctccttac ctctgatttt gctgcgcgag 28380  
 tggcagcgac atggtttgtt gttatatggc ctccagctat tgcctctcgg aatgcatcgc 28440  
 tcagtgttga tctgattaac ttggetgacg ccgccttgc ctcgtctatg tatccattga 28500  
 gcattgccgc aatttctttt gtgggtgatgt cttcaagtgg agcatcaggc agaccctcc 28560  
 ttattgcttt aattttgctc atgtaattta tgagtgtctt ctgcttgatt cctctgctgg 28620  
 ccaggatttt ttcgtagcga tcaagccatg aatgtaacgt aacggaatta tcaactgtga 28680  
 ttctcgtgtg cagaggcttg tgtttgtgct ctgaaaataa ctcaatgttg gcctgtatag 28740  
 cttcagtgat tgcgattcgc ctgtctctgc ctaatccaaa ctctttacce gtccttgggt 28800  
 cctgtagca gtaatatcca ttgtttctta tataaagggt agggggtaaa tcccggcgct 28860  
 catgacttgc ctttcttccc atttctgatc ctcttcaaaa ggccacctgt tactggctga 28920  
 ttttaagtcaa cctttaccgc tgattcgtgg aacagatact ctcttccatc cttaccgga 28980  
 ggtgggaata tcctgcattc ccgaaccatc cgacgaactg tttcaaggct tcttgacgt 29040  
 cgctggcgtg cgttccactc ctgaagtgtc aagtacatcg caaagtctcc gcaattacac 29100  
 gcaagaaaaa accgccatca ggcggcttgg tgttctttca gttcttcaat tcgaatattg 29160  
 gttacgtctg catgtgctat ctgcgcccac atcatccagt ggtcgtagca gtcgttgatg 29220  
 ttctccgctt cgataactct gttgaatggc tctccattcc attctctgt gactcggaag 29280  
 tgcatttatc atctccataa aacaaaacc gcegtagega gttcagataa aataaatccc 29340  
 cgcgagtgcg aggattgta tgtaaatatt ggttttaata tctatatggt ttgtacagag 29400  
 aggcaagta tcgtttccac cgtactcgtg ataataattt tgcacggatc cagtcatttc 29460  
 tcgcacattg cagaatgggg atttgtcttc attagactta taaacctca tggaatattt 29520  
 gtatgccgac tctatatcta taccttctac tacataaaca ccttcgtgat gtctgcatgg 29580  
 agacaagaca ccgatctgc acaacattga taacgcccac tctttttgct cagactctaa 29640  
 ctcatgata ctcatttata aactccttgc aatgtatgtc gtttcagcta aacggatca 29700  
 gcaatgttta tgtaaagaaa cagtaagata atactcaacc cgatgtttga gtacggctca 29760  
 catctgacac tacagactct ggcactcgtg tgaagacgac gcgaaattca gcattttcac 29820  
 aagcgttata ttttcaaaaa ccgatctcac tctccttga tgcgaatgcc agcgtcagac 29880

atcatatgca gatactcacc tgcacctga acccattgac ctccaacccc gtaatagcga 29940  
tgcgtaatga tgtcgaatgt tactaacggg tcttgttcga ttaactgccg cagaaactct 30000  
tccaggtcac cagtgcagtg cttgataaca ggagtcttcc caggatggcg aacaacaaga 30060  
aactggtttc cgtcttcacg gacttcggtg ctttccagtt tagcaatacg cttactccca 30120  
tccgagataa caccttcgta atactcacgc tgctcgttga gttttgattt tgctgtttca 30180  
agctcaacac gcagtttccc tactgttagc gcaatacctt cgttctcctg gtcgcggcgt 30240  
ttgatgtatt gctggtttct ttcccgttca tccagcagtt ccagcacaat cgatgggtgt 30300  
accaattcat ggaaaaggtc tgcgtcaaat ccccagtcgt catgcattgc ctgctctgcc 30360  
gcttcacgca gtgcctgaga gttaatttcg ctacttcga acctctctgt ttactgataa 30420  
gttccagatc ctctggcaa cttgcacaag tccgacaacc ctgaacgacc aggcgtcttc 30480  
gttcatctat cggatcgcca cactcacaac aatgagtggc agatatagcc tgggtggttca 30540  
ggcggcgcac ttttattgct gtgttgcgct gtaattcttc tatttctgat gctgaatcaa 30600  
tgatgtctgc catctttcat taatccctga actgttggtt aatacgttg aggggtgaatg 30660  
cgaataataa aaaaggagcc tgtagctccc tgatgatttt gcttttcatg ttcacgttc 30720  
cttaaagacg ccgtttaaca tgccgattgc caggettaaa tgagtcggtg tgaatcccat 30780  
cagcgttacc gtttcgcggt gcttctcag tacgtacgg caaatgtcat cgacgttttt 30840  
atccggaaac tgctgtctgg ctttttttga tttcagaatt agcctgacgg gcaatgctgc 30900  
gaagggcggt ttctgtctga ggtgtcattg aacaagtccc atgtcggcaa gcataagcac 30960  
acagaatatg aagcccgtg ccagaaaaat gcattccgtg gttgtcatac ctggtttctc 31020  
tcatctgctt ctgctttcgc caccatcatt tccagctttt gtgaaaggga tgcggctaac 31080  
gtatgaaatt ctctgtctgt ttctactggt attggcacia acctgattcc aatttgagca 31140  
aggctatgtg ccatctcgat actcgttctt aactcaacag aagatgcttt gtgcatacag 31200  
cccctgttt attatttacc tctcagcca gccgctgtgc tttcagtgga tttcggataa 31260  
cagaaaggcc gggaaatacc cagcctcgtt ttgtaacgga gtagacgaaa gtgattgcgc 31320  
ctaccgggat attatcgtga ggatgcgtca tcgccattgc tcccaaata caaaaccaat 31380  
ttcagccagt gcctcgtcca tttttctgat gaactccggc acgatctcgt caaaactcgc 31440  
catgtacttt tcatcccgtt caatcacgac ataatgcagg ctttcacgct tcatcgcgg 31500  
gtcatagtgt gcaaagtacc aggcattttt tcgcgtcacc cacatgctgt actgcacctg 31560  
ggccatgtaa gctgacttta tggcctcga accaccgagc cggaacttca tgaatccccg 31620  
ggaggtaaac gggcatttca gttcaaggcc gttgcccgtc ctgcataaac catcgggaga 31680  
gcaggcggta cgcatacttt cgtcgcgata gatgatcggg gattcagtaa cattcacgcc 31740  
ggaagtgaat tcaaacaggg ttctggcgtc gttctcgtac tgttttcccc aggccagtgc 31800  
tttagcgta acttccggag ccacaccggt gcaaacctca gcaagcaggg tgtggaagta 31860  
ggacattttc atgtcaggcc acttcttctc ggagcgggggt tttgctatca cgttgtgaac 31920  
ttctgaagcg gtgatgacgc cgagccgtaa tttgtgccac gcatactccc cctgttcgac 31980  
agctctcaca tcgatecccg tacgtctgag gataatgtcc ggtgtcatgc tgcaccttc 32040  
tgctctcgg ctttctgttt caggaatcca agagctttta ctgcttcggc ctgtgtcagt 32100  
tctgacgatg cacgaatgtc gcggcgaaat atctgggaac agagcggcaa taagtcgtca 32160  
tcccatgttt tatccagggc gatcagcaga gtgttaatct cctgcatggt ttcacgttta 32220

accggagtga tgtcgcgttc cggctgacgt tctgcagtgt atgcagtatt ttcgacaatg 32280  
 cgctcggcctt catccttgtc atagatacca gcaaatccga aggccagacg ggcacactga 32340  
 atcatggcctt tatgacgtaa catccgtttg ggatgcgact gccacggccc cgtgatttct 32400  
 ctgccttcgc gagttttgaa tggttcgcgg cggcattcat ccatccattc ggtaacgcag 32460  
 atcggatgat tacggtcctt gcggtaaadc cggcatgtac aggattcatt gtccctgtca 32520  
 aagtccatgc catcaaactg ctggttttca ttgatgatgc gggaccagcc atcaacgccc 32580  
 accaccggaa cgatgccatt ctgcttatca ggaaaggcgt aaatttcttt cgtccacgga 32640  
 ttaaggccgt actggttggc aacgatcagt aatgcgatga actgcgcatac gctggcatca 32700  
 cctttaaatg ccgtctggcg aagagtggcg atcagttcct gtgggtcgac agaatccatg 32760  
 ccgacacgtt cagccagctt ccagccagc gttgcgagtg cagtactcat tcgttttata 32820  
 cctctgaatc aatatcaacc tgggtggtgag caatggttcc aaccatgtac cggatgtgtt 32880  
 ctgccatgcg ctctgaaac tcaacatcgt catcaaacgc acgggtaatg gattttttgc 32940  
 tggccccgtg gcgttgcaaa tgatcgatgc atagegatcc aacaggtgc tggggcaggc 33000  
 ctttttccat gtcgtctgcc agttctgect ctttctcttc acgggcgagc tgctggtagt 33060  
 gacgcgcca gctctgagcc tcaagacgat cctgaatgta ataagcgttc atggctgaac 33120  
 tctgaaata gctgtgaaaa tctcgcgccg gaaatgccgg gctgattagg aaaacaggaa 33180  
 agggggtag tgaatgcttt tgettgatct cagtttcagt attaatatcc attttttata 33240  
 agcgtcgacg gcttcacgaa acatcttttc atcgccaata aaagtggcga tagtgaattt 33300  
 agtctggata gccataagt tttgatccat tctttgggac tcttggtgta ttaagtatgt 33360  
 cgataaggcg tttccatccg tcacgtaatt tacgggtgat tcgttcaagt aaagattcgg 33420  
 aagggcagcc agcaacaggc caccctgcaa tggcatattg catggtgtgc tccttattta 33480  
 tacataacga aaaacgcctc gagtgaagcg ttattggtat gcggtaaac cgcactcagg 33540  
 cggccttgat agtcatatca tctgaatcaa atattcctga tgtatcgata tcggtaattc 33600  
 ttattccttc gctaccatcc attggaggcc atccttctg accatttcca tcattccagt 33660  
 cgaactcaca cacaacacca tatgcattta agtcgcttga aattgctata agcagagcat 33720  
 gttgcgccag catgattaat acagcattta atacagagcc gtgtttattg agtcggtatt 33780  
 cagagtctga ccagaaatta ttaatctggt gaagtttttc ctctgtcatt acgtcatggt 33840  
 cgatttcaat ttctattgat gtttccagt cgtaatcaat gatgtatttt ttgatgtttg 33900  
 acatctgttc atatcctcac agataaaaaa tcgccctcac actggagggc aaagaagatt 33960  
 tccaataatc agaacaagtc ggtcctgtt tagttacgag cgacattgct ccgtgtattc 34020  
 actcgttggga atgaatacac agtgcagtgt ttattctggt atttatgcca aaaataaagg 34080  
 ccactatcag gcagctttgt tgttctgttt accaagttct ctggcaatca ttgccgtcgt 34140  
 tcgtattgcc catttatega catatttccc atcttccatt acaggaaaca tttcttcagg 34200  
 cttaacatg cattccgatt gcagettgca tccattgcat cgcttgaatt gtccacacca 34260  
 ttgattttta tcaatagtcg tagtcatacg gatagtcctg gtattgttcc atcacatcct 34320  
 gaggatgctc ttcgaactct tcaaattett ctccatata tcaccttaa tagtggattg 34380  
 cggtagtaaa gattgtgect gtcttttaac cacatcaggc tcggtggttc tcgtgtacc 34440  
 ctacagcgag aaatcggata aactattaca acccctacag tttgatgagt atagaaatgg 34500  
 atccactcgt tattctcgga cgagtgttca gtaatgaacc tctggagaga accatgtata 34560

tgatcgttat ctgggttga cttctgcttt taagcccaga taactggcct gaatatgtta 34620  
atgagagaat cggatttctt catgtgtggc atgttttctt ctttgccttt gcattttctgc 34680  
tagcaattaa tgtgcatcga ttatcagcta ttgccagcgc cagatataag cgatttaagc 34740  
taagaaaacg cattaagatg caaaacgata aagtgcgatc agtaattcaa aaccttacag 34800  
aagagcaatc tatggttttg tgcgcagccc ttaatgaagg caggaagtat gtggttacat 34860  
caaaacaatt cccatacatt agtgagtga ttgagcttgg tgtgttgaac aaaacttttt 34920  
cccgatggaa tggaaagcat atattattcc ctattgagga tatttactgg actgaattag 34980  
ttgccagcta tgatccatat aatattgaga taaagccaag gccaatatct aagtaactag 35040  
ataagaggaa tcgattttcc ctttaattttc tggegtccac tgcattgtat gccgcgttcg 35100  
ccaggcttgc tgtaccatgt gcgctgatte ttgcgctcaa tacgttgcag gttgctttca 35160  
atctgtttgt ggtattcage cagcaactgta aggtctatcg gatttagtgc gctttctact 35220  
cgtgatttctg gtttgcgatt cagcgagaga atagggcggg taactggttt tgcgcttacc 35280  
ccaaccaaca ggggatttgc tgettccat tgagcctgtt tctctgcgcg acgttcgceg 35340  
cggcgtgttt gtgcatccat ctggattctc ctgtcagtta gctttggtgg tgtgtggcag 35400  
ttgtagtctt gaacgaaaac cccccgcat tggeacattg gcagctaate cggaatcgca 35460  
cttacggcca atgcttcggt tcgtatcaca caccccaag cttctgctt tgaatgctgc 35520  
ccttcttcag ggcttaattt ttaagagcgt caccttcagt gtggtcagtg cgtcctgctg 35580  
atgtgctcag tatcaccgcc agtggatttt atgtcaacac cgccagagat aatttatcac 35640  
cgcagatggt tatctgtatg tttttatat gaattttttt tttgcagggg ggcattgttt 35700  
ggtaggtgag agatctgaat tgctatgttt agtgagtgtt atctatttat tttcaataa 35760  
atacaattgg ttatgtgttt tggggcgat cgtgaggcaa agaaaaccg gcgctgaggc 35820  
cgggttatte ttgttctctg gtcaaattat atagttggaa aacaaggatg catatatgaa 35880  
tgaacgatgc agaggcaatg ccgatggcga tagtgggtat catgtagccg cttatgctgg 35940  
aaagaagcaa taaccgcag aaaaacaaag ctccaagctc acaaaaacta agggcataga 36000  
caataactac cgatgtcata taccatact ctctaactt ggccagtcgg cgcgttctgc 36060  
ttccgattag aaacgtcaag gcagcaatca ggattgcaat catggttctt gcatatgatg 36120  
acaatgtcgc cccaagacca tctctatgag ctgaaaaga aacaccagga atgtagtggc 36180  
ggaaaaggag atagcaaatg cttacgataa cgtaaggaat tattactatg taaacaccag 36240  
gcatgattct gttccgata attactctg ataattaatc cttactttg cccacctgcc 36300  
ttttaaaaca ttccagtata tcaactttca ttcttgegta geaatatgcc atctcttcag 36360  
ctatctcagc attggtgacc ttgttcagag gcgctgagag atggcctttt tctgatagat 36420  
aatgttctgt taaaatatct ccggcctcat cttttgccg caggctaatg tctgaaaatt 36480  
gaggtgacgg gttaaaaata atatccttgg caacctttt tatatecctt ttaaattttg 36540  
gcttaatgac tatatccaat gagtcaaaaa getcecttc aatatctgtt gccctaaga 36600  
cctttaatat atgcctcaat acaggtaget tggettctac cttaccgctt gttcggccga 36660  
tgaatgcat atgcataaca tcgtttttgg tggttccct catcagtggc tctatctgaa 36720  
cgcgctctcc actgcttaat gacattcctt tcccgattaa aaaatctgtc agatcggatg 36780  
tggtcggccc gaaaacagtt ctggcaaac caatggtgtc gccttcaaca acaaaaaag 36840  
atgggaatcc caatgattcg tcactctcga ggctgttctt aatatctca actgaagctt 36900

tagagcgatt tatcttctga accagactct tgcatttgt tttggtaaag agaaaagttt 36960  
ttccatcgat tttatgaata taciaaataat tggagccaac ctgcaggtga tgattatcag 37020  
ccagcagaga attaaggaaa acagacaggt ttattgagcg cttatctttc cttttatttt 37080  
tgctgcggta agtcgcataa aaaccattct tcataattca atccatttac tatgttatgt 37140  
tctgagggga gtgaaaattc ccctaattcg atgaagattc ttgctcaatt gttatcagct 37200  
atgcgccgac cagaacacct tgccgatcag ccaaactctt cttcaggcca ctgactagcg 37260  
ataactttcc ccacaacgga acaactctca ttgcatggga tcattgggta ctgtgggttt 37320  
agtggttgta aaaacacctg accgctatcc ctgatcagtt tcttgaaggt aaactcatca 37380  
cccccaagtc tggctatgca gaaatcacct ggctcaacag cctgctcagg gtcaacgaga 37440  
attaacattc cgtcaggaaa gcttggcttg gagcctgttg gtgcggtcac ggaattacct 37500  
tcaacctcaa gccagaatgc agaactactg gcttttttgg ttgtgcttac ccatctctcc 37560  
gcatcacctt tggtaaaggt tctaagetta ggtgagaaca tcctgcttg aacatgagaa 37620  
aaaacagggt actcatactc acttetaagt gacggctgca tactaacgc ttatacatc 37680  
tcgtagattt ctctggcgat tgaagggeta aattcttcaa cgtaacttt gagaattttt 37740  
gtaagcaatg cggcggtata agcatttaat gcattgatgc cattaataa agcaccaacg 37800  
cctgactgcc ccatcccat cttgtctgcg acagattcct gggataagcc aagttcattt 37860  
ttcttttttt cataaattgc ttttaaggcga cgtgcgtcct caagctgctc ttgtgttaat 37920  
ggtttctttt ttgtgctcat acgttaaate tatcaccgca agggataaat atctaacacc 37980  
gtgcgtgttg actattttac ctctggcggt gataatggtt gcattgacta aggaggttgt 38040  
atggaacaac gcataacct gaaagattat gcaatgcgct ttgggcaaac caagacagct 38100  
aaagatctcg gcgtatatca aagcgcgatc aacaaggcca ttcattgcagg ccgaaagatt 38160  
tttttaacta taaacgctga tggaaagcgtt tatgcggaag aggtaaagcc cttcccagat 38220  
aacaacaaaa caacagcata aataaccccg ctcttacaca ttccagccct gaaaaagggc 38280  
atcaaattaa accacaccta tgggtgatgc atttatttgc atacattcaa tcaattgtta 38340  
tctaaggaaa tacttacata tggttcgtgc aaacaaacgc aacgaggctc tacgaatcga 38400  
gagtgcgttg cttaacaaaa tcgcaatgct tggaaactgag aagacagcgg aagctgtggg 38460  
cgttgataag tcgcagatca gcaggtggaa gagggactgg attccaaagt tctcaatgct 38520  
gcttgctggt cttgaatggg gggtcgttga cgacgacatg gctcgattgg cgcgacaagt 38580  
tgctgcgatt ctaccaata aaaaacgcc ggccggcaacc gagcgttctg aacaaatcca 38640  
gatggagttc tgaggtcatt actggatcta tcaacaggag tcattatgac aaatacagca 38700  
aaaatactca acttcggcag aggttaacttt gccggacagg agcgtaatgt ggcagatctc 38760  
gatgatgggt acgccagact atcaaatatg ctgcttgagg cttattcggg cgcagatctg 38820  
accaagcgac agtttaaagt getgettgc attctgcgta aaacctatgg gtggaataaa 38880  
ccaatggaca gaatcaccga ttctcaact agcgagatta caaagttacc tgtcaaacgg 38940  
tgcaatgaag ccaagttaga actcgtcaga atgaatatta tcaagcagca aggcggcatg 39000  
tttgaccaca ataaaaacat ctcagaatgg tgcatecctc aaaacgaggg aaaatccctt 39060  
aaaacgaggg ataaaaacat cctcaaattg ggggattgct atccctcaa acagggggac 39120  
acaaaagaca ctattacaaa agaaaaaaga aaagattatt cgtcagagaa ttctggcgaa 39180  
tcctctgacc agccagaaaa cgacctttct gtggtgaaac cggatgctgc aattcagagc 39240

ggcagcaagt gggggacagc agaagacctg accgccgcag agtggatggt tgacatggtg 39300  
 aagactatcg caccatcagc cagaaaaccg aatthtctg ggtgggctaa cgatatccgc 39360  
 ctgatgcgtg aacgtgacgg acgtaaccac cgcgacatgt gtgtgctggt ccgctgggca 39420  
 tgccaggaca acttctggtc cggtaacgtg ctgagccccg ccaaactccg cgataagtgg 39480  
 acccaactcg aaatcaaccg taacaagcaa caggcaggcg tgacagccag caaaccaaaa 39540  
 ctgcacctga caaacacaga ctggatttac ggggtggatc tatgaaaaac atcgccgcac 39600  
 agatgggttaa ctttgaccgt gagcagatgc gtcggatcgc caacaacatg ccggaacagt 39660  
 acgacgaaaa gccgcaggta cagcaggtag cgcagatcat caacggtgtg ttcagccagt 39720  
 tactggcaac tttcccggcg agcctggcta accgtgacca gaacgaagtg aacgaaatcc 39780  
 gtcgccagtg ggttctggct tttcgggaaa acgggateac cacgatggaa caggttaacg 39840  
 caggaatgcg cgtagcccgt cggcagaatc gaccatttct gccatcccc gggcagtttg 39900  
 ttgcatggtg ccgggaagaa gcatccgtta ccgccggact gccaaacgtc agcgagctgg 39960  
 ttgatatggt ttacgagtat tgccggaagc gaggectgta tccggatgcg gagtcttata 40020  
 cgtggaaatc aaacgcgcac tactggetgg ttaccaacct gtatcagaac atgcgggcca 40080  
 atgcgcttac tgatgcggaa ttacgccgta agcccgcaga tgagcttgtc catatgactg 40140  
 cgagaattaa ccgtggtgag gcgatccctg aaccagtaaa acaacttct gtcatgggcg 40200  
 gtagacctct aaatcgtgca caggetctgg cgaagatcgc agaaatcaaa gctaagttcg 40260  
 gactgaaagg agcaagtgta tgacgggcaa agaggcaatt attcattacc tggggacgca 40320  
 taatagcttc tgtgcgccgg acgttgccgc gctaacaggc gcaacagtaa ccagcataaa 40380  
 tcaggcccg gctaaaatgg cacgggcagg tcttctggtt atcgaaggta aggtctggcg 40440  
 aacggtgtat taccggtttg ctaccaggga agaacgggaa ggaaagatga gcacgaacct 40500  
 ggtttttaag gagtgtcgc agagtccgc gatgaaacgg gtattggcgg tatatggagt 40560  
 taaaagatga ccatctacat tactgagcta ataacaggcc tgctggtaat cgcaggcctt 40620  
 tttatttggg ggagagggaa gtcataaaaa aactaacctt tgaaattcga tctccagcac 40680  
 atcagcaaaa cgctattcac gcagtacagc aaatccttcc agaccaacc aaaccaatcg 40740  
 tagtaacat tcaggaacgc aaccgcagct tagacaaaa caggaagcta tgggcctgct 40800  
 taggtgacgt ctctcgtcag gttgaatggc atggtcgtg gctggatgca gaaagctgga 40860  
 agtgtgtggt taccgcagca ttaaagcagc aggatgtgt tccaacctt gccgggaatg 40920  
 gctttgtggt aataggccag tcaaccagca ggatgcgtgt aggcaattt gcggagctat 40980  
 tagagcttat acaggcattc ggtacagagc gtggcgtaa gtggtcagac gaagcgagac 41040  
 tggctctgga gtggaaagcg agatggggag acagggtgc atgataaatg tcgttagttt 41100  
 ctccggtggc aggacgtcag catatttctg ctggctaata gagcaaaagc gacgggcagg 41160  
 taaagacgtg cattacgttt tcatggatac aggttgtgaa catccaatga catatcggtt 41220  
 tgtcagggaa gttgtgaagt tctgggatat accgctcacc gtattgcagg ttgatataca 41280  
 cccggagctt ggacagccaa atggttatac ggtatgggaa ccaaaggata ttcagacgcg 41340  
 aatgcctgtt ctgaagccat ttatcgatat ggtaaagaaa tatggcactc catacgtcgg 41400  
 cggcgcgttc tgcaactgaca gattaaaact cgttccctc accaaatact gtgatgacca 41460  
 tttcgggcca gggaattaca ccacgtggat tggcatcaga gctgatgaac cgaagcggt 41520  
 aaagccaaag cctggaatca gatatcttgc tgaactgtca gactttgaga aggaagatat 41580

cctcgcatgg tggaaagcaac aaccattcga tttgcaaata ccggaacatc tcggttaactg 41640  
catattctgc attaaaaaat caacgcaaaa aatcggaactt gcctgcaaag atgaggaggg 41700  
attgcagcgt gtttttaatg aggtcatcac gggatcccat gtgcgtgacg gacatcggga 41760  
aacgccaag gagattatgt accgaggaag aatgtcgctg gacggtatcg cgaaaatgta 41820  
ttcagaaaaat gattatcaag ccctgtatca ggacatggta cgagctaaaa gattcgatac 41880  
cggctcttgt tctgagtcac gcgaaatatt tggagggcag cttgatttcg acttcgggag 41940  
ggaagctgca tgatgcatg ttatcgggtc ggtgaatgca aagaagataa ccgcttccga 42000  
ccaaatcaac cttactggaa tcgatgggtg ctccgggtg aaagaacacc aacaggggtg 42060  
ttaccactac cgcaggaaaa ggaggacgtg tggcgagaca gcgacgaagt atcaccgaca 42120  
taatctgcga aaactgcaaa taccttccaa cgaaacgcac cagaaataaa cccaagccaa 42180  
tcccaaaaaga atctgacgta aaaacettca actacacggc tcacctgtgg gatatccggt 42240  
ggctaagacg tcgtgcgagg aaaacaaggt gattgacca aatcgaagt acgaacaaga 42300  
aagcgtcgag cgagctttaa cgtgcgctaa ctgcggctag aagctgcatg tgctggaagt 42360  
tcacgtgtgt gagcaactgt gcgcagaact gatgagcgt ccgaatagct cgatgcacga 42420  
ggaagaagat gatggctaaa ccagcgcgaa gacgatgtaa aaacgatgaa tgccgggaat 42480  
ggtttcaccc tgcattcgt aatcagtggg ggtgctctcc agagtgtgga accaagatag 42540  
cactcgaacg acgaagtaaa gaacgcgaaa aagcggaaaa agcagcagag aagaaacgac 42600  
gacgagagga gcagaaacag aaagataaac ttaagattcg aaaactcgc ttaaagcccc 42660  
gcagttactg gattaaacia gcccaacaag ccgtaaacgc cttcatcaga gaaagagacc 42720  
gcgacttacc atgtatctcg tgcggaacgc tcacgtctgc tcagtgggat gccggacatt 42780  
accggacaac tgctgcggca cctcaactcc gatttaatga acgcaatatt cacaagcaat 42840  
gcgtgggtgt caaccagcac aaaagcggaa atctcgttcc gtatcgcgtc gaactgatta 42900  
gccgcatcgg gcaggaagca gtagacgaaa tcgaatcaaa ccataaccgc catcgtctgga 42960  
ctatcgaaga gtgcaaggcg atcaaggcag agtaccaca gaaactcaaa gacctgcgaa 43020  
atagcagaag tgaggccgca tgacgttctc agtaaaaacc attccagaca tgctcgttga 43080  
aacatacgg aatcagacag aagtagcacg cagactgaaa thtagtcgcg gtacggctcag 43140  
aaaatacgtt gatgataaag acgggaaaat gcacgccatc gtcaacgacg ttctcatggt 43200  
tcacgcgga tggagtgaag gagatgcgct attacgaaa aattgatggc agcaaatacc 43260  
gaaatatttg ggtagttggc gatctgcacg gatgctacac gaacctgatg aacaaactgg 43320  
atacgattgg attcgacaac aaaaaagacc tgcttatctc ggtggcgat ttggttgatc 43380  
gtggtgcaga gaacgttgaa tgcttggaat taatcacatt ccctggttc agagctgtac 43440  
gtggaacca tgagcaaatg atgattgatg gcttatcaga gcgtggaaac gttaatcact 43500  
ggctgcttaa tggcgggtggc tggttcttta atctcgatta cgacaaagaa attctggcta 43560  
aagctcttgc ccataaagca gatgaaette cgttaatcat cgaactggtg agcaaagata 43620  
aaaaatatgt tatctgccac gccgattatc cctttgacga atacgagttt ggaaagccag 43680  
ttgatcatca gcaggtaatc tggaaaccgg aacgaatcag caactacaa aacgggatcg 43740  
tgaagaaaat caaaggcgcg gacacgttca tctttggtca tacgccagca gtgaaaccac 43800  
tcaagtttgc caaccaaag tatatcgata ccggcgcagt gttctgcgga aacctaact 43860  
tgattcaggt acagggagaa ggcgcatgag actcgaaagc gtagctaaat tcattcgc 43920

aaaaagcccc atgatgagcg actcaccacg ggccacggct tetgactetc tttccggtag 43980  
 tgatgtgatg gctgctatgg ggatggcgca atcacaagcc ggattcggta tggctgcatt 44040  
 ctgcbgtaag cacgaactca gccagaacga caaacaaaag gctatcaact atctgatgca 44100  
 atttgcacac aaggtatcgg ggaaataccg tgggtggca aagcttgaag gaaatactaa 44160  
 ggcaaaggta ctgcaagtgc tcgcaacatt cgcttatgcg gattattgcc gtagtgccgc 44220  
 gacgccgggg gcaagatgca gagattgcca tggtagaggc cgtgcbggtg atattgcca 44280  
 aacagagctg tgggggagag ttgtcgagaa agagtgcgga agatgcaaag gcgtcggcta 44340  
 ttcaaggatg ccagcaagcg cagcatatcg cgctgtgacg atgctaacc caaaccttac 44400  
 ccaaccacc tggtcacgca ctgtaagcc gctgtatgac gctctggtg tgcaatgcca 44460  
 caaagaagag tcaatcgag acaacatddd gaatgcggtc acacgttagc agcatgattg 44520  
 ccacggatgg caacatatta acggcatgat attgacttat tgaataaaat tgggtaaatt 44580  
 tgactcaacg atgggttaat tcgctcggtg tggtagtgag atgaaaagag gcggcgctta 44640  
 ctaccgattc cgctagttg gtaacttcca cgtatcgtct ggaactcaa ccatcgagg 44700  
 cagagaggtc tgcaaaatgc aatcccgaag cagttcgag gtaatagtta gagcctgcat 44760  
 aacggtttcg ggatttttta tatctgcaca acaggtaaga gcattgagtc gataatcgtg 44820  
 aagagtcggc gagcctggtt agccagtget ctttccggtg tgctgaatta agcgaatacc 44880  
 ggaagcagaa ccggatcacc aaatgcgtac aggcgtcacc gccgccagc aacagcacia 44940  
 cccaaactga gccgtagcca ctgtctgtcc tgaattcatt agtaatagtt acgctgcggc 45000  
 cttttacaca tgaccttctg gaaagcgggt ggcaggaggt cgcgtaaca acctctgcc 45060  
 gttttgcccg tgcatatcgg tcacgaacaa atctgattac taaacacagt agcctggatt 45120  
 tgtttatca gtaatcgacc ttattcctaa ttaaataagag caaatccct tattgggggt 45180  
 aagacatgaa gatgccagaa aaacatgacc tgttgccgc cattctcgc gcaaaggaac 45240  
 aaggcatcgg ggcaatcctt gcgtttgcaa tggcgtacct tcgcbgca gataatggcg 45300  
 gtgcbgttac aaaaacagta atcgacgcaa cgatgtcgc cattatgcc tagttcattc 45360  
 gtgaccttct cgacttccc ggactaagta gcaatctcgc ttatataacg agcgtgttta 45420  
 tcggctacat cggtagtagc tcgattggtt cgcttatcaa acgcttcgct gctaaaaag 45480  
 ccggagtaga agatggtaga aatcaataat caacgtaagg cgttcctcga tatgctggcg 45540  
 tggtaggagg gaactgataa cggacgtcag aaaaccagaa atcatggta tgacgtcatt 45600  
 gtaggcggag agctatttac tgattactcc gatcaccctc gcaaacttgt cacgctaaac 45660  
 ccaaaactca aatcaacagg cgccggagc taccagcttc tttccggtg gtgggatgcc 45720  
 taccgcaagc agcttggcct gaaagacttc tctccgaaa gtcaggagc tgtggcattg 45780  
 cagcagatta aggagcgtgg cgctttacct atgattgac gtggtgatat ccgtcaggca 45840  
 atcgaccgtt gcagcaatat ctgggettea ctgcbggbg ctggttatgg tcagttcgag 45900  
 cataaggctg acagcctgat tgcaaaatc aaagaagcgg gcggaacggt cagagagatt 45960  
 gatgtatgag cagagtcacc gcgattatct ccgctctggt tatctgcacc atcgtctgcc 46020  
 tgtcatgggc tgtaatcat taccgtgata acgcaatc ctacaaagcc cagcgcgaca 46080  
 aaaatgccag agaactgaag ctggcgaac cggaattac tgacatgcag atcgtcagc 46140  
 gtgatgttgc tgcgctgat gcaaaataca cgaaggagtt agctgatgct aaagctgaaa 46200  
 atgatgctct gcgtgatgat gttgccgctg gtcgctcgc gttgcacatc aaagcagctc 46260

```

gtcagtcagt gcgtgaagcc accaccgcct ccggcgtgga taatgcagcc tccccccgac 46320
tggcagacac cgctgaacgg gattatttca ccctcagaga gaggctgata actatgcaaa 46380
aacaactgga aggaaccag aagtatatta atgagcagtg cagatagagt tgcccatatc 46440
gatgggcaac tcatgcaatt attgtgagca atacacacgc gcttcagcg gagtataaat 46500
gcctaaagta ataaaaccga gcaatccatt tacgaatggt tgctggggtt ctgttttaac 46560
aacattttct gcgccgccac aaattttggc tgcacgaca gttttcttct gcccaattcc 46620
agaaacgaag aatgatggg tgatggtttc ctttgggtct actgctgccg gtttgttttg 46680
aacagtaaac gtctgttgag cacatcctgt aataagcagg gccagcgcag tagcgagtag 46740
catttttttc atgggtttat tcccgatgct ttttgaagtt cgcagaatcg tatgtgtaga 46800
aaattaaaca aaccctaaac aatgagttga aatttcatat tgtaaatatt tattaatgta 46860
tgtcaggtgc gatgaatcgt cattgtatc ccggattaac tatgtccaca gccctgacgg 46920
ggaacttctc tgcgggagtg tccgggaata attaaaacga tgcacacagg gtttagcgcg 46980
tacacgtatt gcattatgcc aacgccccgg tgcagacag gaagaaaccg gacgttatga 47040
tttagcgtgg aaagatttgt gtagtgttct gaatgctctc agtaaatagt aatgaattat 47100
caaaggata gtaatatctt ttatgttcat ggatatttgt aaccatcgg aaaactcctg 47160
ctttagcaag attttccctg tattgctgaa atgtgatttc tcttgatttc aacctatcat 47220
aggacgttcc tataagatgc gtgtttcttg agaatttaac atttacaacc tttttaagtc 47280
cttttattaa cacggtgta tcgttttcta acacgatgtg aatattatct gtggctagat 47340
agtaaataata atgtgagacg ttgtgacggt ttagttcaga ataaaacaat tcacagtcta 47400
aatcttttcg cacttgatcg aatatttctt taaaaatggc aacctgagcc attggtaaaa 47460
ccttccatgt gatacagagg cgcgtagttt gcattatcgt ttttatcgtt tcaatctggt 47520
ctgacctcct tgtgttttgt tgatgattta tgcataat taggaatggt ttcacttaat 47580
agtattgggt gcgtaacaaa gtgcggtcct gctggcattc tggagggaaa tacaaccgac 47640
agatgtatgt aaggccaacg tgcataatc ttcatcaga aagatttgaa gtaatatttt 47700
aaccgctaga tgaagagcaa gcgcatggag cgacaaaatg aataaagaac aatctgctga 47760
tgatccctcc gtggatctga ttcgtgtaaa aaatatgctt aatagcacca tttctatgag 47820
ttaccctgat gttgtaattg catgtataga acataagggt tctctggaag cattcagagc 47880
aattgaggca gcgttggtga agcacgataa taatatgaag gattattccc tgggtggtga 47940
ctgatcacca taactgctaa tcattcaaac tatttagtct gtgacagagc caacacgcag 48000
tctgtcactg tcaggaaagt ggtaaaactg caactcaatt actgcaatgc cctcgtaatt 48060
aagtgaattt acaatategt cctgttcgga gggaagaacg cgggatgttc attcttcatc 48120
acttttaatt gatgtatatg ctctcttttc tgacgttagt ctccgacggc aggttcaat 48180
gaccaggct gagaaattcc cggaccctt ttgctcaaga gcgatgttaa tttgttcaat 48240
catttggtta ggaaagcgga tgttgccgggt tgttgctctg cgggttctgt tcttcgttga 48300
catgaggttg ccccgatc agtgctcctg attgtattg tctgaagttg tttttacgtt 48360
aagttgatgc agatcaatta atacgatacc tgcgtcataa ttgattattt gacgtggttt 48420
gatggcctcc acgcacgttg tgatatgtag atgataatca ttatcacttt acgggtcctt 48480
tccggtgatc cgacaggta cg
48502

```

<210> 44

<211> 25  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例5和6中使用的序列  
<400> 44  
tttttttttt tttttttttt ttttt 25  
<210> 45  
<211> 23  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例5和6中使用的序列  
<400> 45  
gcaatatcag caccaacaga aac 23  
<210> 46  
<211> 25  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例5和6中使用的序列  
<400> 46  
gtttctgttg gtgctgatat tgctt 25  
<210> 47  
<211> 14  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例5和6中使用的序列  
<400> 47  
ccttctgtta cgtt 14  
<210> 48  
<211> 13  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例5和6中使用的序列  
<400> 48  
ttcgtaacag aag 13

- <210> 49  
<211> 48  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例7中使用的序列  
<400> 49  
cgttctgttt atgtttcttg gacactgatt gacacggttt agtagaac 48
- <210> 50  
<211> 48  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例7中使用的序列  
<400> 50  
tttttttttt tttttttttt ttttttttca agaaacataa acagaacg 48
- <210> 51  
<211> 12  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例7中使用的序列  
<400> 51  
tttttttttt tt 12
- <210> 52  
<211> 51  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例7中使用的序列  
<400> 52  
ggttgtttct gttggtgctg atattgcggc gtctgcttgg gtgtttaacc t 51
- <210> 53  
<211> 68  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例7中使用的序列  
<400> 53

---

ggttaaacac ccaagcagac gccgcaatat cagcaccaac agaaacaacc ttgaggcga 60  
gcggtcaa 68  
<210> 54  
<211> 15  
<212> DNA  
<213> 人工序列  
<220>  
<223> 实施例7中使用的序列  
<400> 54  
ttgaccgctc gcctc 15

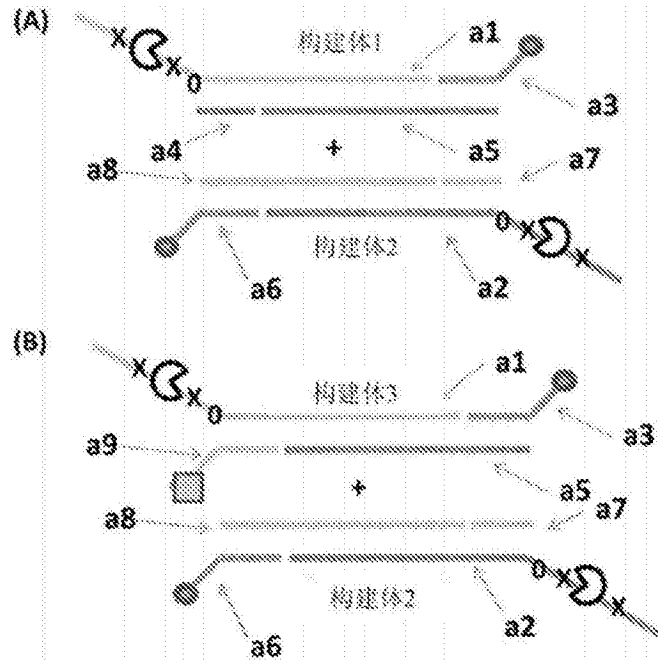


图1

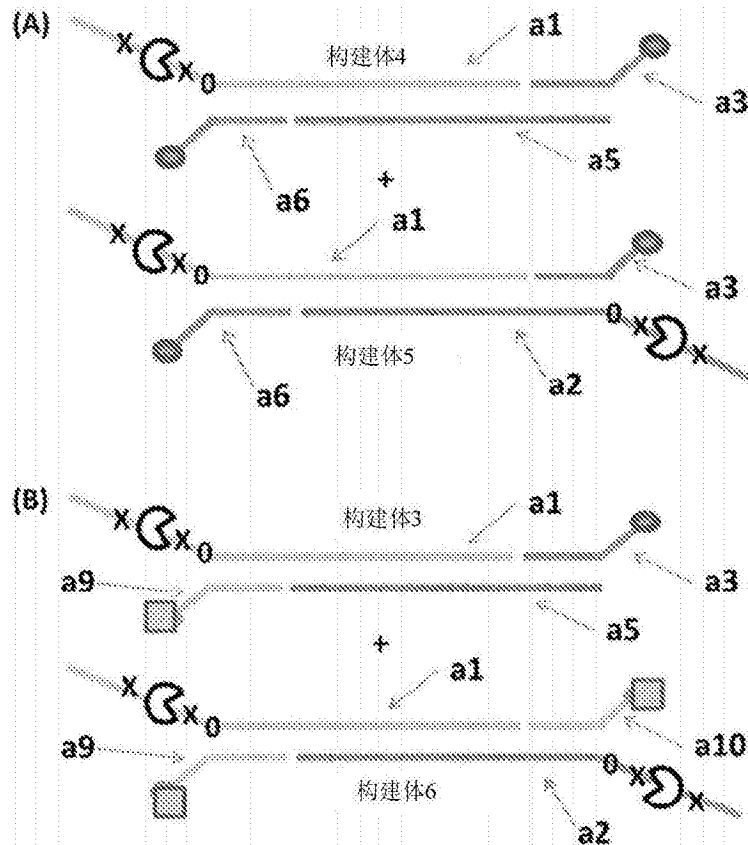


图2

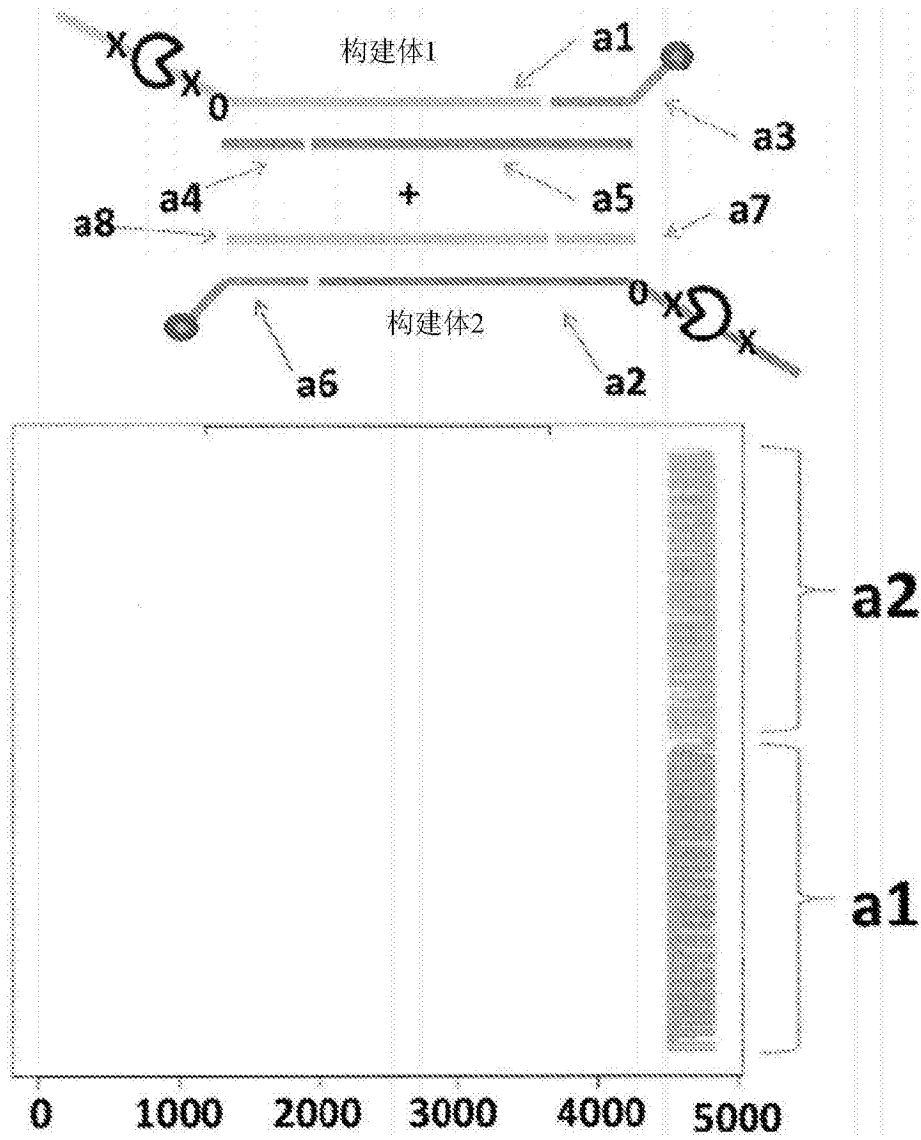


图3

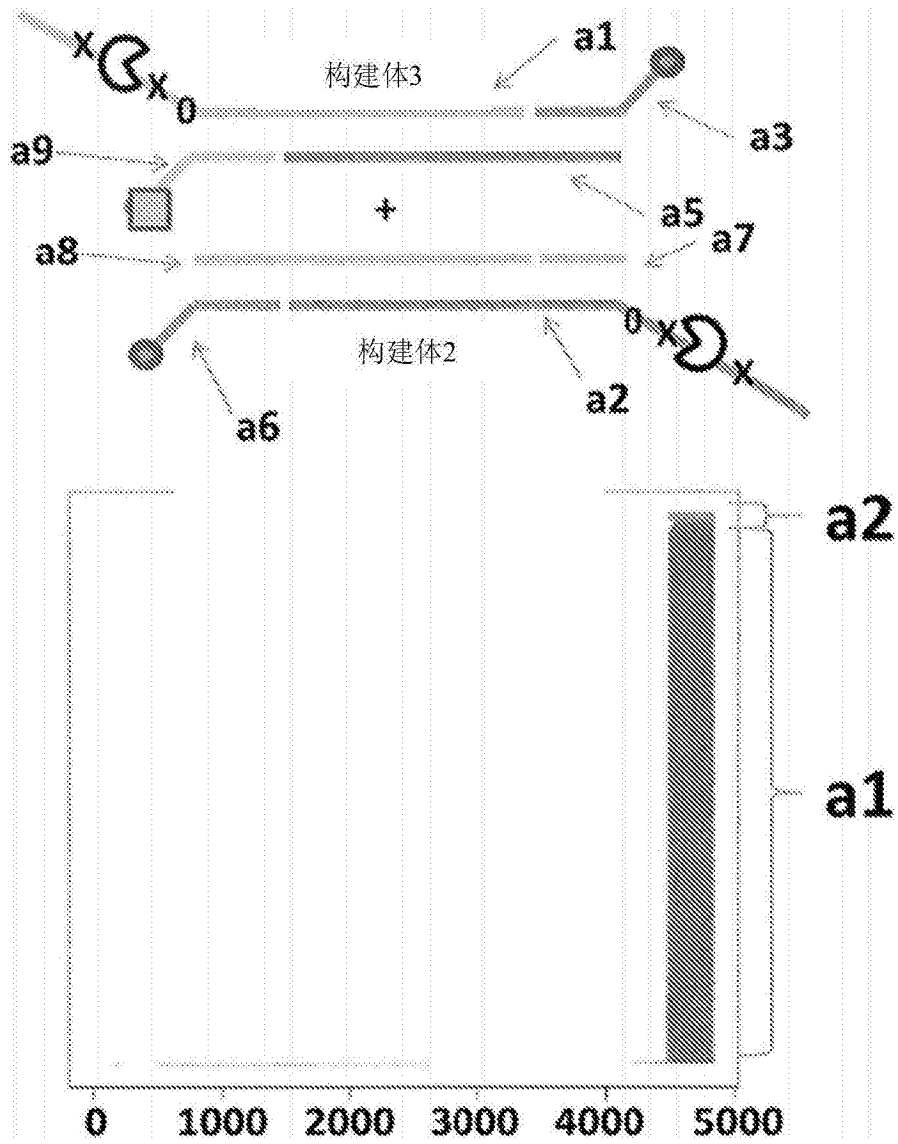


图4

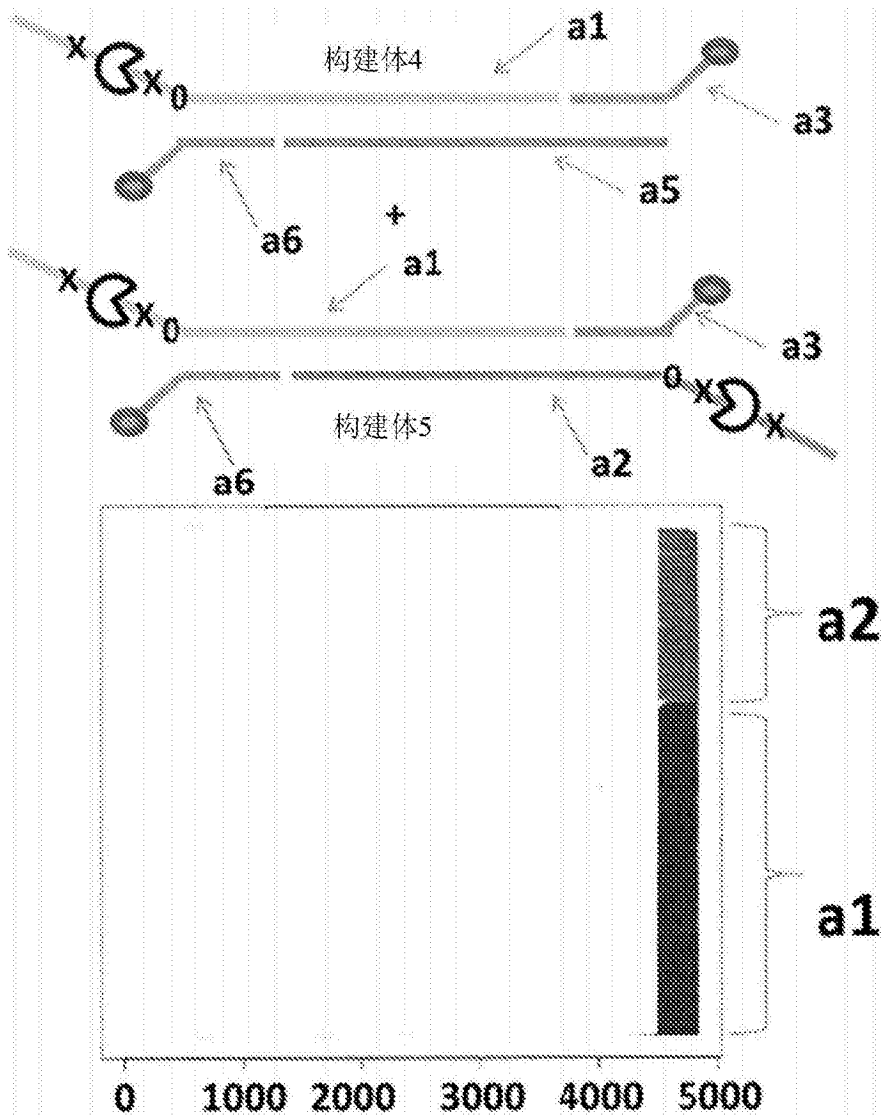


图5



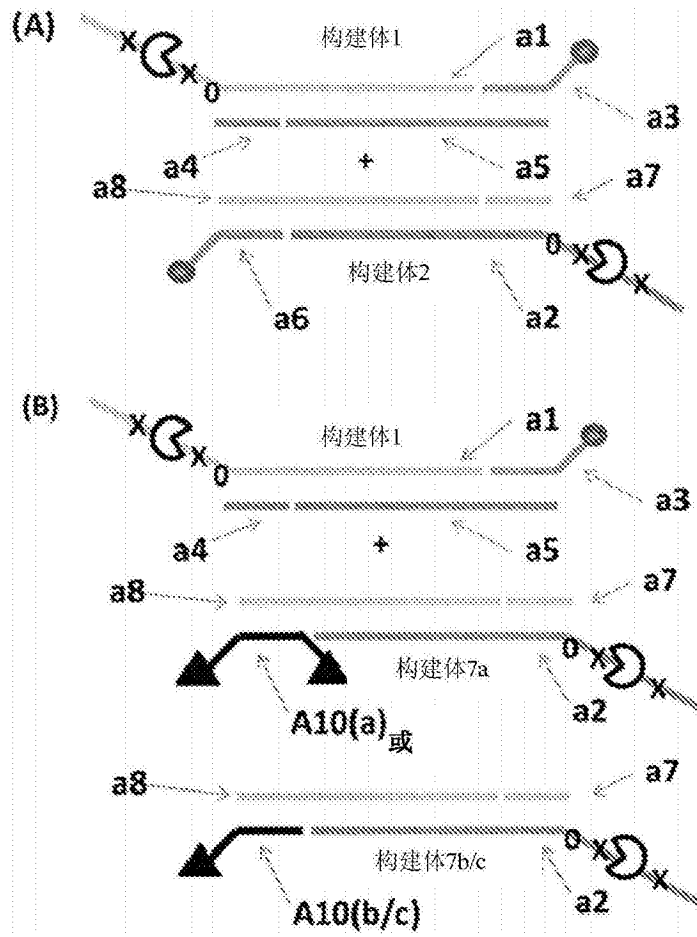


图7

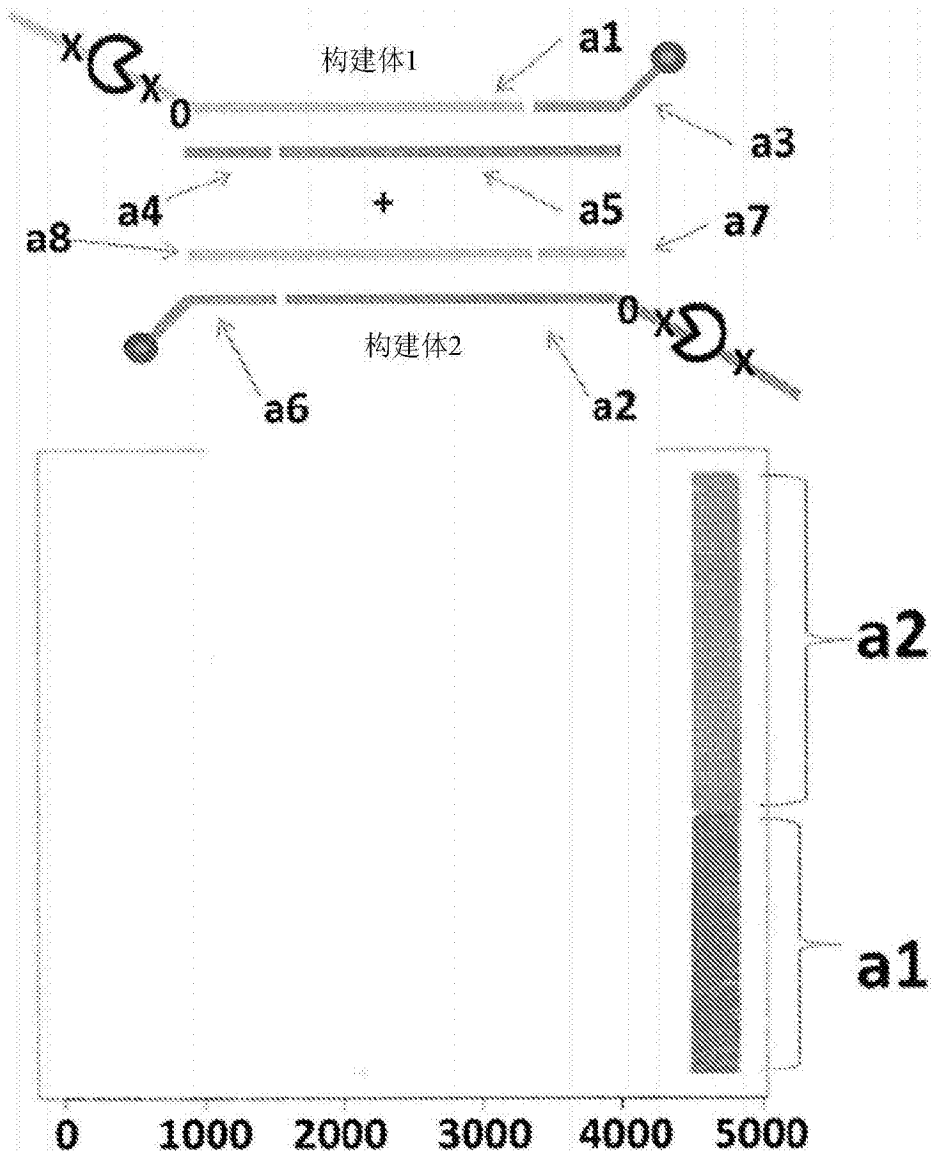


图8

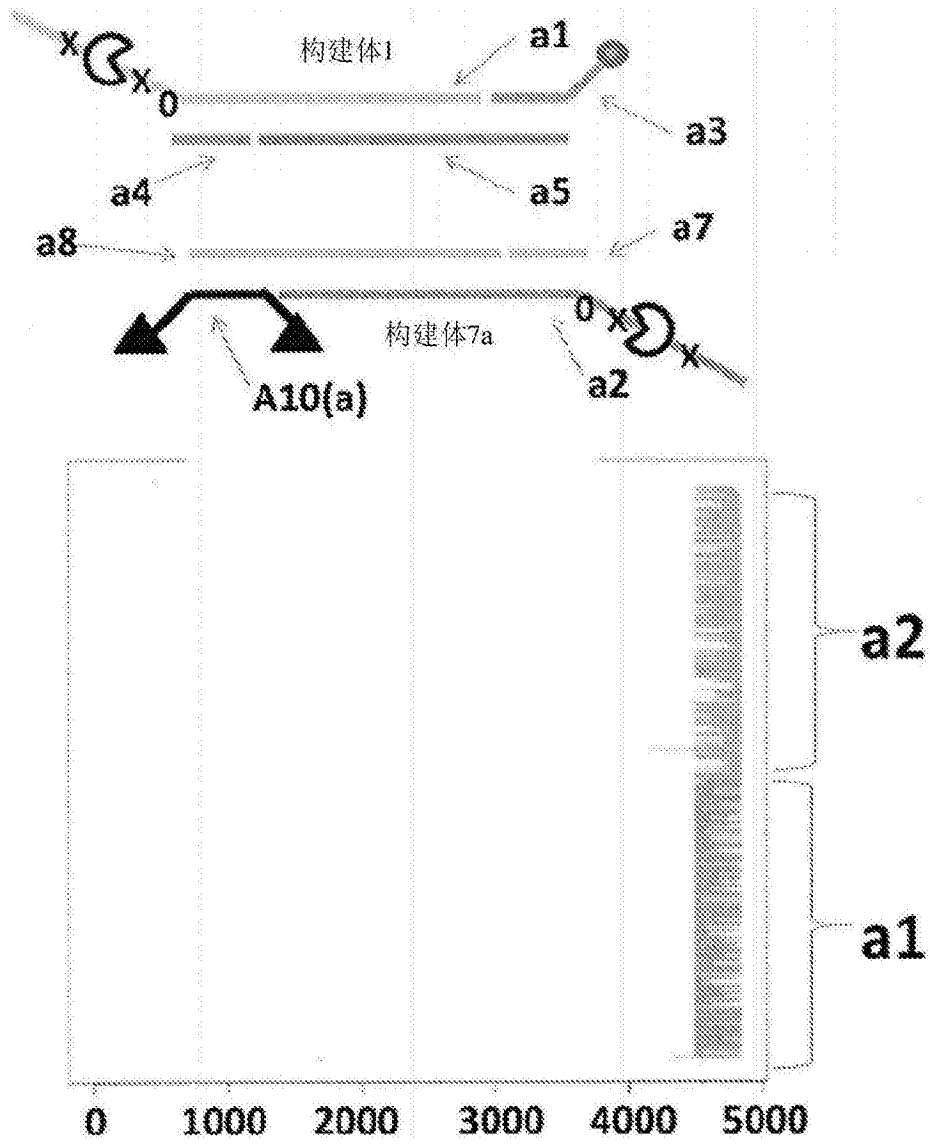


图9

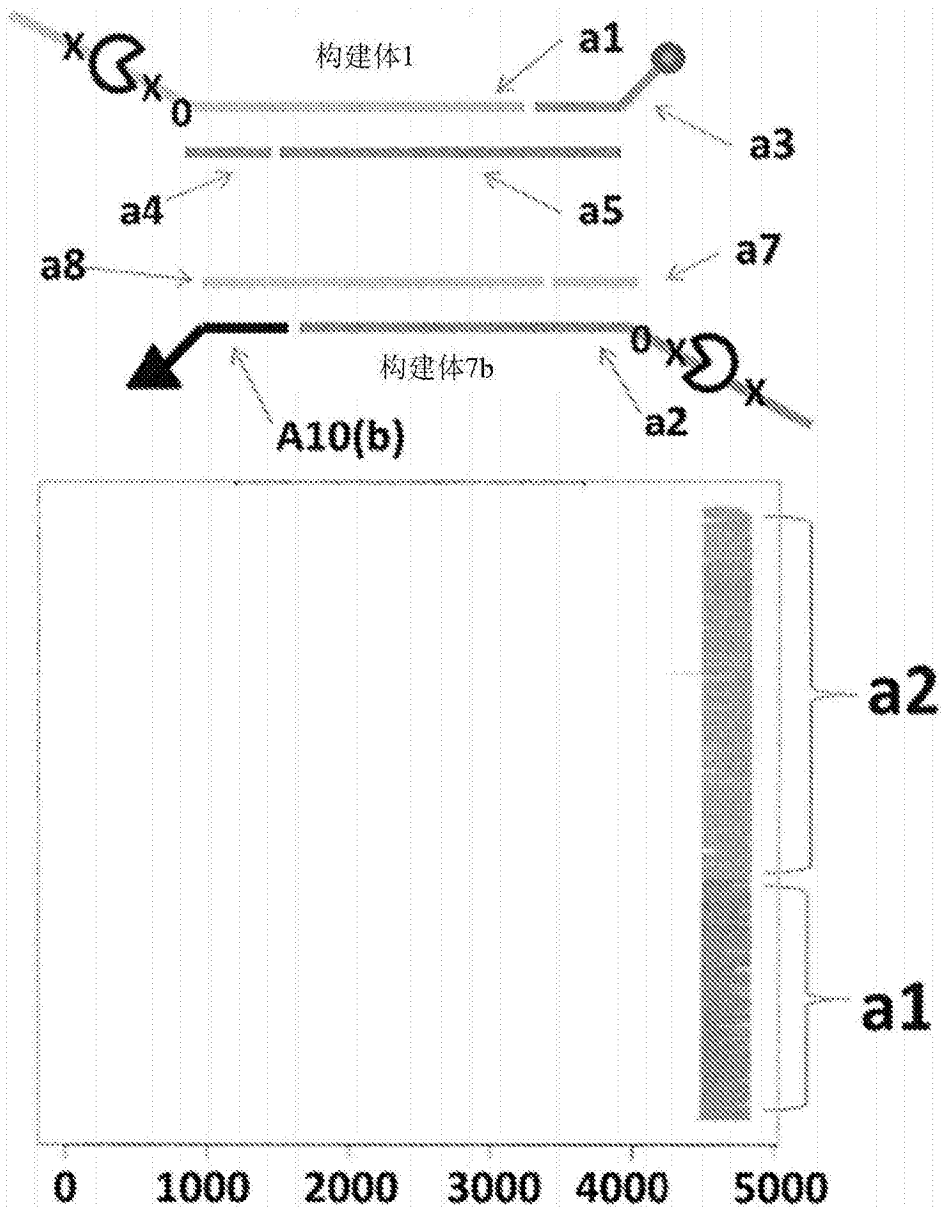


图10

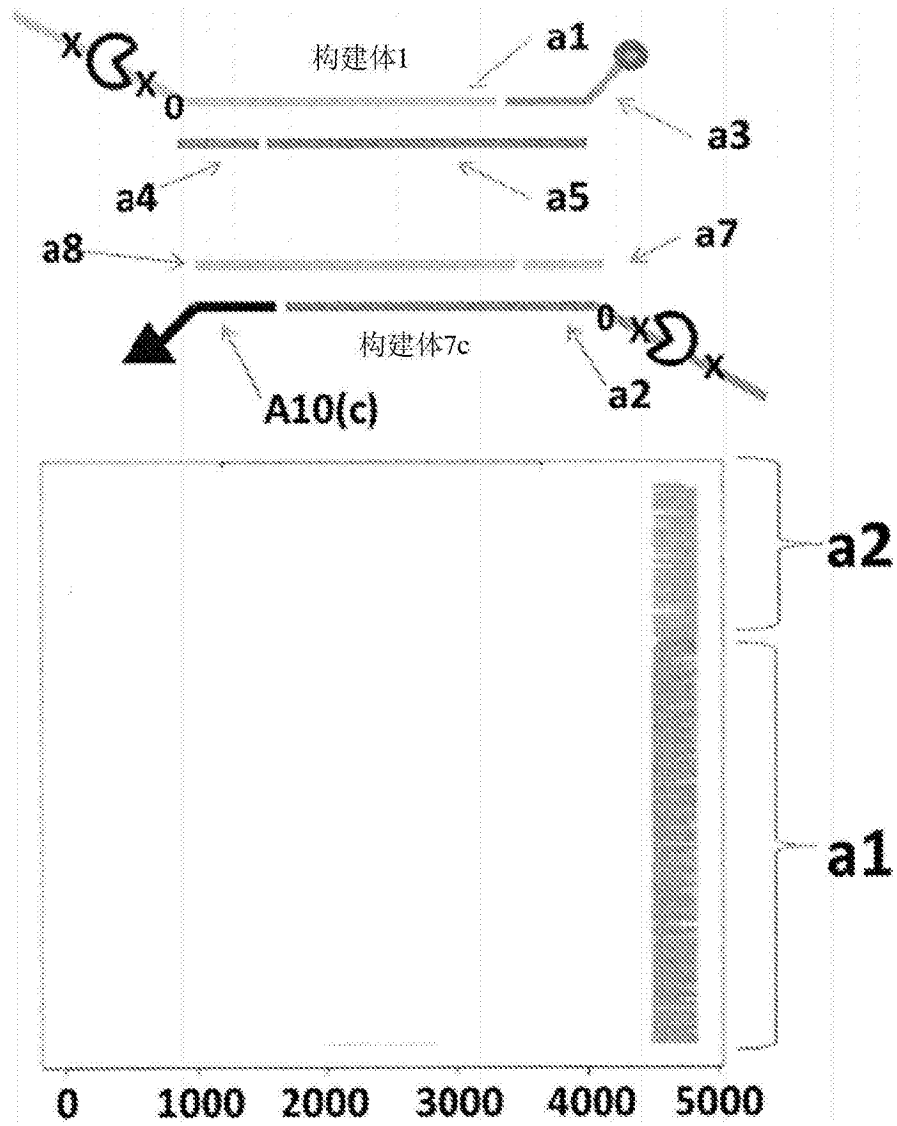


图11



图12

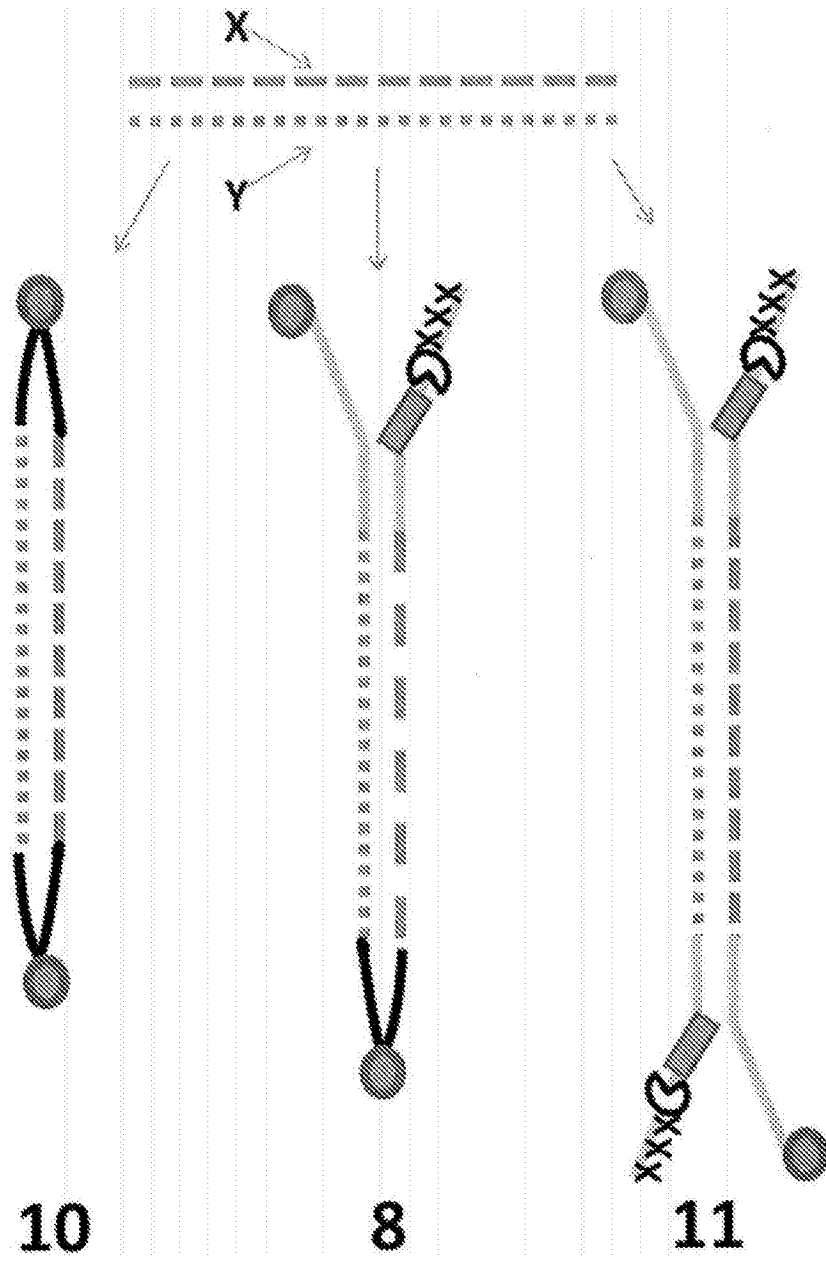


图13

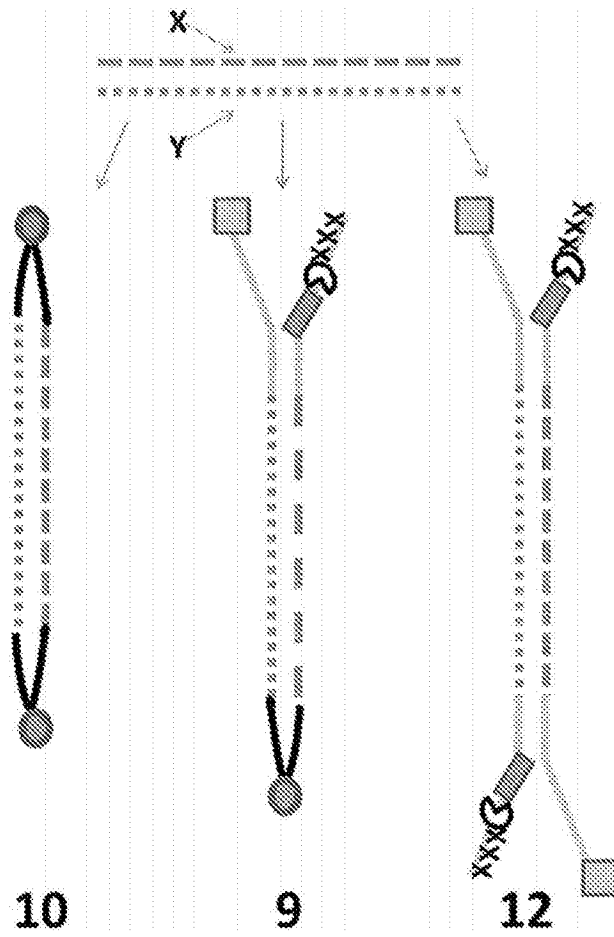


图14

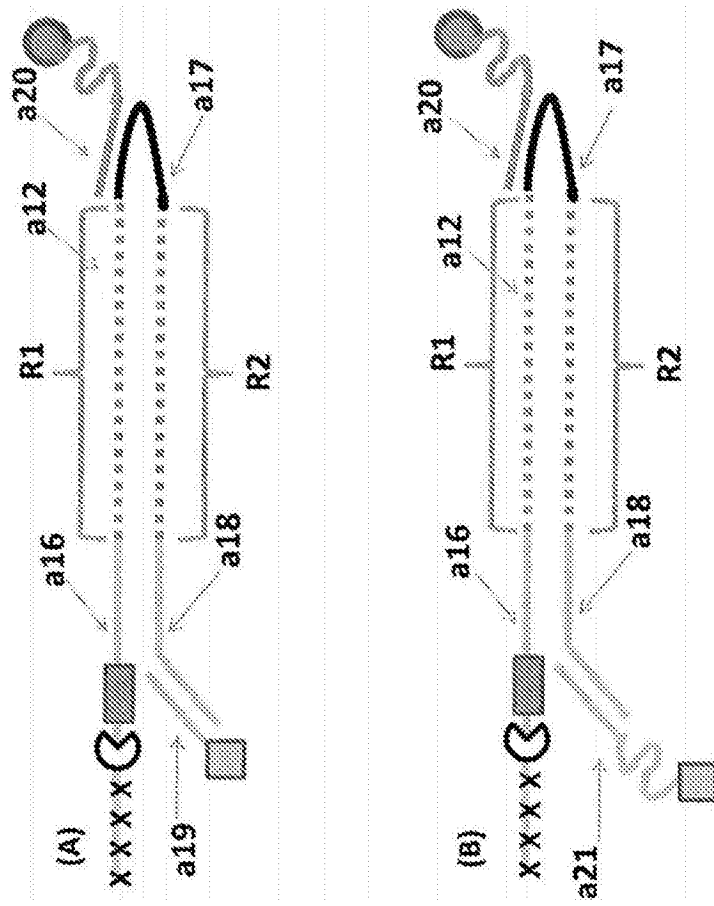


图15

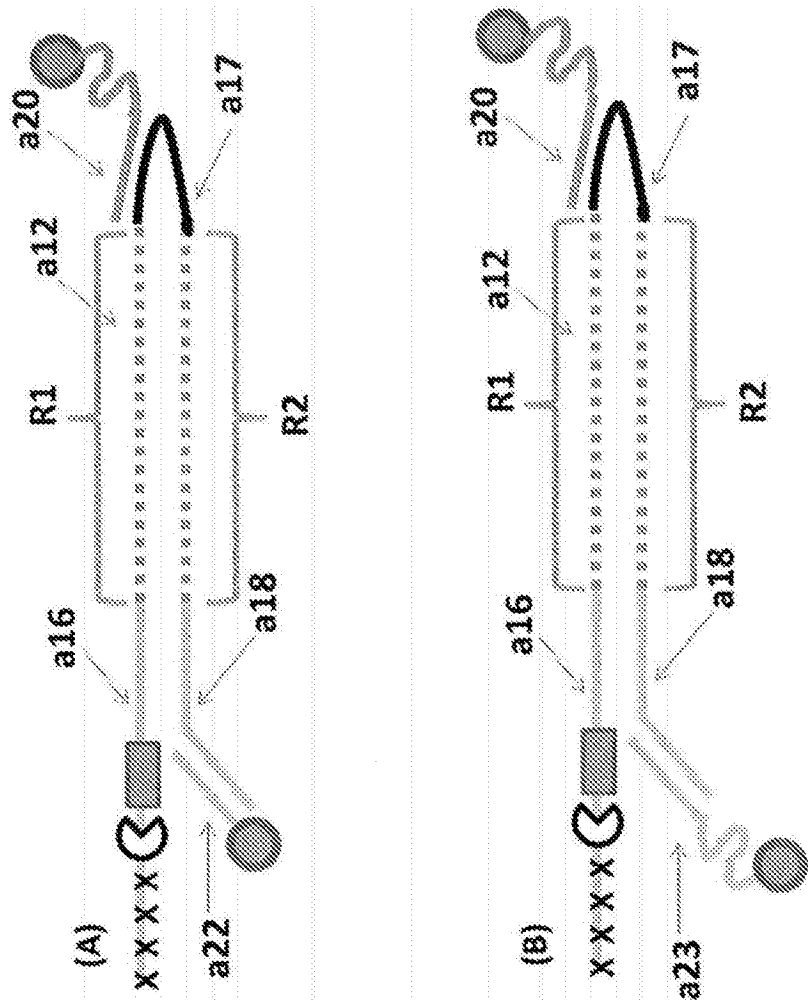


图16

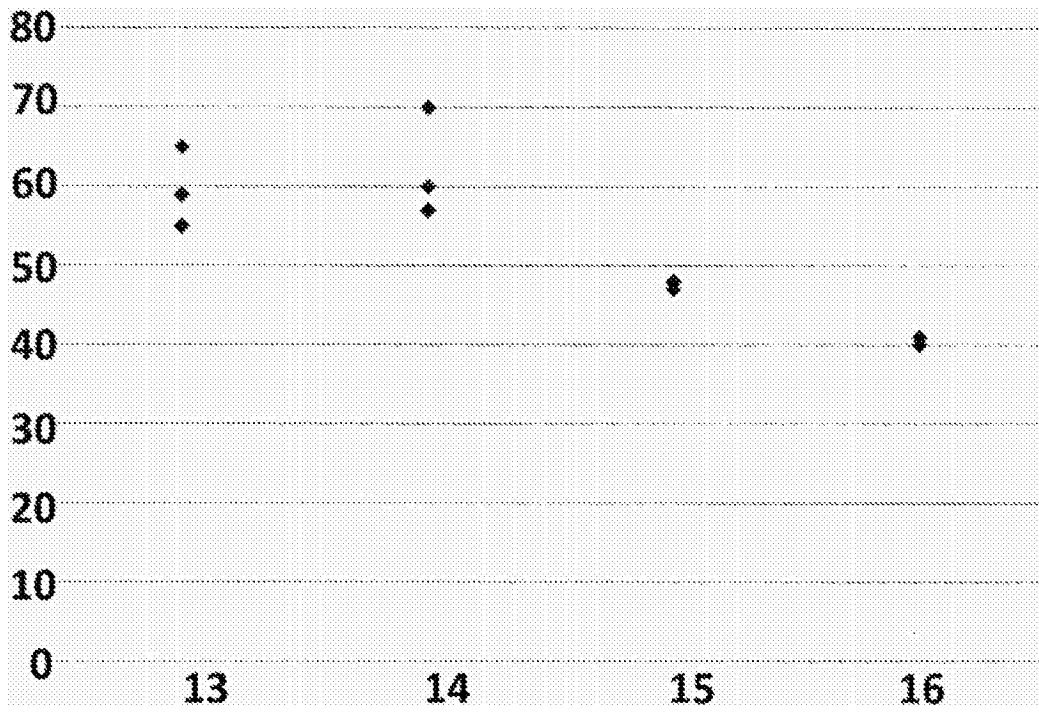


图17

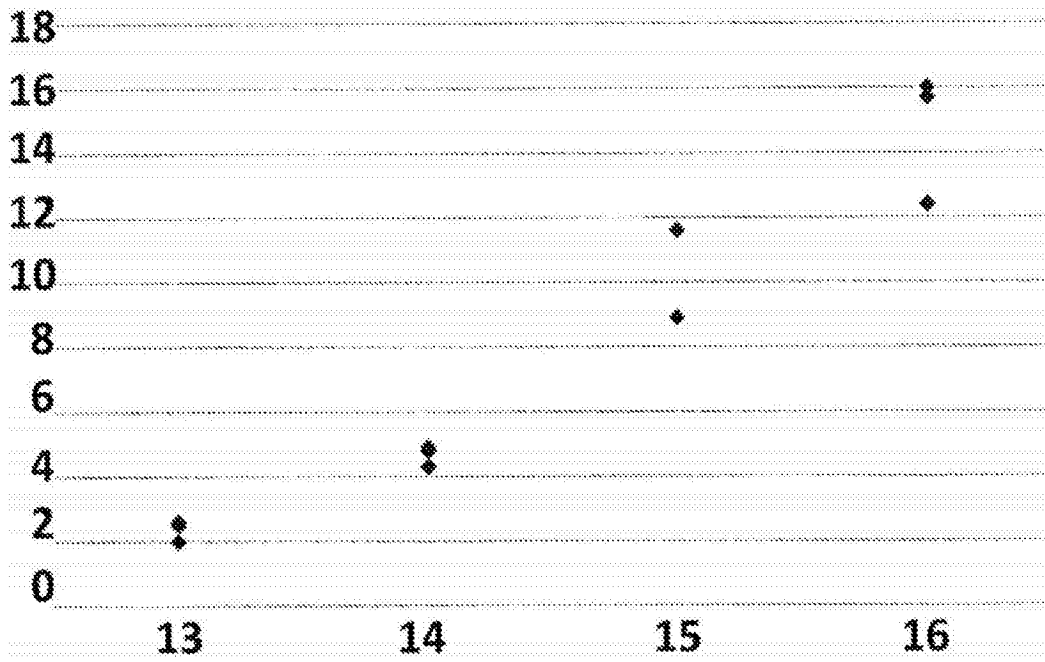


图18