



## (12) 发明专利

(10) 授权公告号 CN 107109489 B

(45) 授权公告日 2021.11.30

(21) 申请号 201580069027.9

(22) 申请日 2015.10.19

(65) 同一申请的已公布的文献号  
申请公布号 CN 107109489 A

(43) 申请公布日 2017.08.29

(30) 优先权数据  
1418459.2 2014.10.17 GB  
1508270.4 2015.05.14 GB  
1517634.0 2015.10.06 GB  
PCT/GB2015/052916 2015.10.06 GB

(85) PCT国际申请进入国家阶段日  
2017.06.16

(86) PCT国际申请的申请数据  
PCT/GB2015/053097 2015.10.19

(87) PCT国际申请的公布数据  
W02016/059436 EN 2016.04.21

(73) 专利权人 牛津纳米孔技术公司

地址 英国牛津郡

(72) 发明人 丹尼尔·赖安·加拉尔达  
安德鲁·约翰·赫伦  
拉科马·贾亚辛格  
丹尼尔·约翰·特纳  
詹姆斯·怀特

(74) 专利代理机构 北京同立钧成知识产权代理有限公司 11205

代理人 彭雪瑞 臧建明

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

审查员 何宇

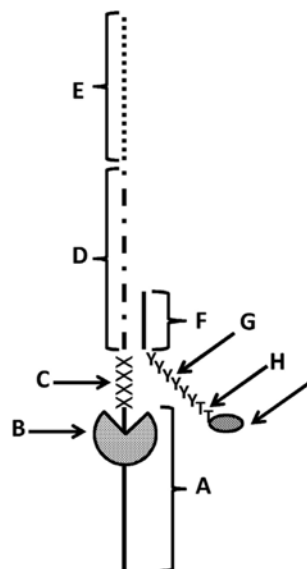
权利要求书1页 说明书65页  
序列表41页 附图18页

### (54) 发明名称

纳米孔RNA表征方法

### (57) 摘要

本发明涉及通过随着目标RNA多核苷酸相对于跨膜孔移动,获取一个或多个测量值来表征目标RNA多核苷酸的新方法。通过DNA解旋酶控制该移动。本发明也涉及修饰的RNA构建体,其中所述RNA多核苷酸已被修饰以增强DNA解旋酶对其的结合。



1. 一种表征目标RNA多核苷酸的方法,包括:
  - a) 提供 (i) RNA多核苷酸和 (ii) DNA解旋酶,其中所述RNA多核苷酸被修饰以包含DNA前导序列,所述DNA前导序列包括DNA解旋酶结合位点;
  - b) 将a) 中提供的所述RNA多核苷酸和DNA解旋酶与跨膜孔接触,使得所述DNA解旋酶控制所述RNA多核苷酸穿过所述跨膜孔的移动;
  - c) 随着所述RNA多核苷酸相对于所述跨膜孔移动,获取一个或多个测量值,其中所述测量值代表所述RNA多核苷酸的一个或多个特征,并由此表征所述目标RNA多核苷酸。
2. 根据权利要求1所述的方法,其中所述前导序列优先地穿入所述跨膜孔。
3. 根据权利要求1所述的方法,其中通过共价键将所述DNA前导序列连接到所述RNA多核苷酸,所述共价键形成在所述RNA多核苷酸和所述DNA前导序列的各自至少一个反应基团之间。
4. 根据权利要求1所述的方法,其中通过化学或酶促连接将所述DNA前导序列连接到所述RNA多核苷酸。
5. 根据权利要求1所述的方法,其中所述DNA前导序列杂交到所述RNA多核苷酸。
6. 根据权利要求1所述的方法,其中所述一个或多个特征选自 (i) 所述RNA多核苷酸的长度, (ii) 所述RNA多核苷酸的同一性, (iii) 所述RNA多核苷酸的序列, (iv) 所述RNA多核苷酸的二级结构和 (v) 所述RNA多核苷酸是否是修饰的。
7. 根据权利要求1所述的方法,其中所述RNA多核苷酸的一个或多个特征通过电测量和/或光测量来测定。
8. 根据权利要求1所述的方法,其中步骤c) 包括随着所述RNA多核苷酸相对于所述跨膜孔移动,测量流过所述跨膜孔的电流,其中所述电流代表所述RNA多核苷酸的一个或多个特征,并由此表征所述RNA多核苷酸。
9. 根据权利要求1所述的方法,其中所述RNA多核苷酸包括通过甲基化、氧化、损伤、用一个或多个蛋白,碱基类似物或用一个或多个标记物、标签或间隔器进行修饰。
10. 根据权利要求1所述的方法,其中所述RNA多核苷酸使用一个或多个锚耦合到所述膜。
11. 根据前述任一项权利要求所述的方法,其中所述DNA解旋酶包含修饰,以减小所述多核苷酸结合域中开口的大小,所述RNA多核苷酸可以在至少一个构象状态下穿过所述开口从所述解旋酶上解绑。
12. 根据前述任一项权利要求所述的方法,其中所述DNA解旋酶为 (a): Hel308 解旋酶、RecD 解旋酶、XPD 解旋酶或Dda 解旋酶; (b): 衍生自 (a) 中所述任何解旋酶的解旋酶; 或 (c): (a) 和/或 (b) 中所述解旋酶的任意组合。
13. 根据权利要求1所述的方法,进一步包含一个或多个衍生自解旋酶的分子制动器,所述分子制动器被修饰使得其结合所述多核苷酸但不发挥解旋酶的功能。
14. 根据权利要求1所述的方法,其中所述跨膜孔是蛋白孔或固态孔。
15. 根据权利要求14所述的方法,其中所述跨膜蛋白质孔是蛋白孔,并衍生自溶血素、杀白细胞素,耻垢分枝杆菌孔蛋白A (MspA)、MspB、MspC、MspD、CsgG、胞溶素、外膜孔蛋白F (OmpF)、外膜孔蛋白G (OmpG), 外膜磷脂酶A、奈瑟球菌属自转运脂蛋白 (NalP) 和WZA。

## 纳米孔RNA表征方法

### 技术领域

[0001] 本发明涉及通过随着目标RNA多核苷酸相对于跨膜孔移动,获取一个或多个测量值来表征所述目标RNA多核苷酸的新方法。所述目标RNA多核苷酸相对于所述跨膜孔的移动通过DNA解旋酶控制,并且修饰所述目标RNA多核苷酸以增强DNA解旋酶对其的结合。本发明还涉及被修饰的RNA的构建体,其中所述RNA多核苷酸已被修饰以增强DNA解旋酶对其的结合。

### 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,当分析物如核苷酸在桶(桶状体)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] 信使RNA提供了对生物体动态的观察,并且直接RNA测序的益处和应用是巨大的,包括用于健康筛查;例如某些癌症的转移过程和心脏病。直接RNA测序能在调查农作物的抗病性中应用,确定农作物对应激因素,例如干旱、紫外线和盐度的反应,以及在胚胎发育过程的细胞分化和决定中应用。

[0006] 存在于RNA,特别是500个或更多的核苷酸的RNA的直接测序中的问题是寻找合适的能够控制RNA穿过跨膜孔的移位的分子马达。至今,用于RNA并提供持续移动分子马达还没有出现。对于表征或测序多核苷酸,需要RNA聚合物的持续移动和读取长片段聚合物的能力。

[0007] 国际专利申请No.PCT/GB2014/053121(WO 2015/056028)公开了表征目标核糖核酸(RNA)的方法,包括形成互补多核苷酸,然后使用跨膜孔表征所述互补多核苷酸。这种间接RNA表征易于出错并可能导致关于例如,RNA的甲基化状态的重要信息的丢失。RNA到cDNA的转换过程中其他重要的修饰也可能被隐藏。

### 发明内容

[0008] 本发明人出乎意料地证明了在DNA解旋酶的控制下,可以通过随着目标RNA多核苷酸相对于跨膜孔移动,获取一个或多个测量值来表征所述目标RNA多核苷酸。相应地,在一个具体实施方式中,提出了用于提高要被测序的RNA穿过孔的能力和效率的方法。在另一具体实施方式中,也提出制备被修饰的RNA的方法,所述被修饰的RNA比未修饰形式的RNA能以更高效率穿过孔而被测序。本发明提供的一种表征目标RNA多核苷酸的方法,包括:

[0009] a) 提供 (i) RNA多核苷酸和 (ii) DNA解旋酶, 其中所述RNA多核苷酸被修饰以包含非-RNA多核苷酸;

[0010] b) 将a) 中提供的所述RNA多核苷酸和DNA解旋酶与跨膜孔接触, 使得所述DNA解旋酶控制所述RNA多核苷酸穿过所述跨膜孔的移动;

[0011] c) 随着所述RNA多核苷酸相对于所述跨膜孔移动, 获取一个或多个测量值, 其中所述测量值代表所述RNA多核苷酸的一个或多个特征, 并由此表征所述目标RNA多核苷酸。

[0012] 修饰所述RNA多核苷酸以包含非-RNA多核苷酸(例如多核苷酸区域或序列或构建体)导致了增强的DNA解旋酶对其的结合。在此定义的“非-RNA多核苷酸”是多核苷酸, 其中所述多核苷酸的至少一个核苷酸不是核糖核苷酸, 即不是来自RNA。所述非-RNA多核苷酸因此可以包含至少一个核糖核苷酸(或RNA核苷酸), 但必须额外包含或包括非-RNA核苷酸或序列, 即不是RNA的核苷酸或核苷酸序列。在本发明的优选的实施方式中, 所述非-RNA多核苷酸(其可能或可能不包含至少一个核糖核苷酸或RNA核苷酸)包含DNA或DNA类似物, 优选DNA解旋酶结合位点或DNA适配器。优选地所述非-RNA多核苷酸包含优先地穿入所述孔的前导序列。因此首先读取所述非-RNA多核苷酸, 然后读取要被表征的目标RNA序列。

[0013] 本发明提供的“杂交的”多核苷酸的构建体, 包含 (i) RNA多核苷酸和 (ii) 非-RNA多核苷酸。优选地所述非-RNA多核苷酸包含DNA多核苷酸, 其中所述DNA多核苷酸包含或仅包含DNA解旋酶结合位点。优选地所述非-RNA多核苷酸进一步包含前导序列, 其优先地穿入纳米孔。更优选地所述非-RNA多核苷酸进一步包含在多核苷酸链上的条形码部分。所述条形码部分优选地位于所述前导序列和所述DNA解旋酶结合位点之间。所述条形码部分使得能够清楚地识别分析物, 即通知使用者多个样品中哪个被测序。

[0014] 修饰目标RNA多核苷酸以包含非-RNA多核苷酸可包括使用任何合适的连接方法, 包括在此描述的一个或多个的连接方法, 将非-RNA多核苷酸(可能包含RNA序列或至少一个核糖核苷酸)连接到所述目标RNA多核苷酸。正如本文所述, 将非-RNA多核苷酸连接到目标RNA与将目标RNA连接到非-RNA多核苷酸同义。其中所述非-RNA多核苷酸包含核糖核苷酸或RNA序列, 所述非-RNA多核苷酸可以通过包含在所述非-RNA多核苷酸内的核糖核苷酸或RNA连接到所述目标RNA多核苷酸。

[0015] 所述非-RNA多核苷酸可以通过共价键连接到所述目标RNA多核苷酸, 所述共价键形成在所述目标RNA多核苷酸和所述非-RNA多核苷酸的各自至少一个反应基团之间。所述非-RNA多核苷酸可以用化学方法地或酶促地连接到所述RNA多核苷酸。通过杂交和/或使用一个或多个拓扑异构酶, 所述非-RNA多核苷酸可以额外地或替换地连接到所述RNA多核苷酸。优选地通过所述方法确定的一个或多个特征选自 (i) 所述RNA多核苷酸的长度, (ii) 所述RNA多核苷酸的同一性, (iii) 所述RNA多核苷酸的序列, (iv) 所述RNA多核苷酸的二级结构和 (v) 所述RNA多核苷酸是否是修饰的。所述RNA多核苷酸的一个或多个特征可以通过电测量和/或光测量来测量。优选地步骤c) 包括随着所述RNA多核苷酸相对于所述跨膜孔移动, 测量流过所述跨膜孔的电流, 其中所述电流代表所述RNA多核苷酸的一个或多个特征, 并由此表征所述RNA多核苷酸。

[0016] 要被表征的目标RNA多核苷酸可以额外地或进一步通过甲基化、氧化、损伤、用一个或多个蛋白, 或用一个或多个标记物、标签或间隔器进行修饰。所述目标RNA可以包含碱基类似物。所述RNA多核苷酸可以使用一个或多个锚耦合到所述膜。

[0017] 优选地所述DNA解旋酶包含减小多核苷酸结合域中开口的大小的修饰,所述RNA多核苷酸可以在至少一个构象状态下穿过所述开口从所述解旋酶上解绑。在本发明的一个实施方式中,通过一系列的一个或多个DNA解旋酶控制所述RNA多核苷酸的移动。所述一个或多个解旋酶为a) Hel308解旋酶、RecD解旋酶、XPD解旋酶或Dda解旋酶; (b) 衍生自(a)中所述任何解旋酶的解旋酶;或(c) (a)和/或(b)中所述解旋酶的任意组合。所述方法可以进一步包含使用一个或多个衍生自解旋酶的分子制动器,所述分子制动器被修饰使得其结合多核苷酸但不发挥解旋酶的功能。

[0018] 所述跨膜孔可以是蛋白孔或固态孔。优选地所述跨膜蛋白质孔是蛋白孔,并衍生自溶血素、杀白细胞素,耻垢分枝杆菌(*Mycobacterium smegmatis*)孔蛋白A(MspA)、MspB、MspC、MspD、胞溶素(lysenin)、CsgG、外膜孔蛋白F(OmpF)、外膜孔蛋白G(OmpG),外膜磷脂酶A、奈瑟球菌属(*Neisseria*)自转运脂蛋白(NalP)和WZA。

[0019] 本发明也提供了目标RNA多核苷酸相对于跨膜孔移动的方法,所述移动被DNA解旋酶控制,所述方法包括:

[0020] a) 提供(i) RNA多核苷酸和(ii) DNA解旋酶,所述RNA多核苷酸被修饰以包含非-RNA多核苷酸;

[0021] b) 将a)中提供的所述RNA多核苷酸和DNA解旋酶与跨膜孔接触,使得所述DNA解旋酶控制所述RNA多核苷酸相对于所述跨膜孔的移动。

[0022] 修饰所述RNA多核苷酸以包含非-RNA多核苷酸导致增强的DNA解旋酶对其的结合。在本发明的一个实施方式中,所述方法包含在接触步骤之前将所述DNA解旋酶预先结合到所述修饰的RNA多核苷酸。

[0023] 本发明的方法提供了所述RNA多核苷酸相对于所述跨膜孔的更持续的移动。本发明也提供了“杂交的”多核苷酸,包括(i) RNA多核苷酸和(ii) 非-RNA多核苷酸。优选地所述非-RNA多核苷酸包含DNA多核苷酸,其中所述DNA多核苷酸包含或仅包含DNA解旋酶结合位点。优选地所述非-RNA多核苷酸进一步包含前导序列,其优先地穿入纳米孔。更优选地所述非-RNA多核苷酸进一步包含在所述多核苷酸链上的条形码部分。所述条形码部分优选地位于所述前导序列和所述DNA解旋酶结合位点之间。所述条形码部分使得能够清楚地识别分析物,即通知使用者多个样品中的哪个被测序。

[0024] 本发明也提供了目标RNA多核苷酸和DNA解旋酶的结合体,其中部分的所述RNA多核苷酸被修饰以与所述DNA解旋酶相互作用或结合到所述DNA解旋酶。优选地所述RNA被修饰以包含非-RNA多核苷酸,最优选地所述非-RNA多核苷酸是DNA或DNA类似物。

[0025] 本发明也提供了一种用于表征目标RNA多核苷酸的试剂盒。优选地所述试剂盒包含非-RNA多核苷酸,所述非-RNA多核苷酸适于连接到要被表征的目标RNA。本发明也提供了一种用于表征样品中的目标RNA多核苷酸的设备。

[0026] 本文所述的修饰的RNA构建体提供用于DNA解旋酶的结合位点。所述DNA解旋酶本质上“被欺骗地”(“tricked”)读取所述RNA多核苷酸:一旦所述DNA解旋酶结合到所述非-RNA多核苷酸,其会沿着所述RNA多核苷酸移动(transition)。所述解旋酶可以在溶液中沿着RNA移动。所述解旋酶可能需要纳米孔的存在以促进所述解旋酶沿着所述RNA多核苷酸的移动。所述解旋酶沿着所述RNA的移动可被纳米孔和通过穿过所述纳米孔的施加的场而促进。

[0027] 整个RNA构建体在DNA解旋酶的控制下穿过纳米孔的移位的能力允许以与已知方法相比改善的精度和速度表征所述RNA多核苷酸,例如评估其序列。与现有技术方法相比,本发明的方法不存在PCR偏好性。RNA中的修饰可被检测到并且非编码亚型和剪接变体可被正确地鉴定。在纳米孔中可直接地检测RNA碱基类似物。本文所述的构建体对实现500个或更多个核苷酸,例如1000,5000,10000,20000,50000,100000或更多个核苷酸的目标RNA的移位特别有效。

[0028] 在本发明的方法中,所述DNA解旋酶通过借助非-RNA前导序列的存在本质上“被欺骗地”(“tricked”)读取所述目标RNA序列。一旦通过所述非-RNA多核苷酸(其可以包含DNA或DNA类似物)引发所述DNA解旋酶的移动,它可以继续沿着所述RNA移动。本发明的方法也提供了区分RNA和DNA的方法,作为平均振幅和范围的函数,即使当RNA和DNA序列相同时。

## 附图说明

[0029] 图1显示了将真核RNA的链(如虚线所示)连接到DNA的链(如实线所示)的方法的卡通示意图。真核RNA在其5'末端具有7-甲基鸟苷帽(如星形所示并标记为A)。反应步骤1中将7-甲基鸟苷帽留在原处,并且添加一反应基团(标记为B并如方形所示)到该真核RNA的5'末端。使用超甲基化酶例如具有修饰的S腺苷甲基胺(SAdenosyl Methiamine)的三甲基鸟苷合成酶来添加该反应基团。步骤2显示一化学反应步骤,其中也具有连接的标记为C并显示为圆形的反应基团的DNA链与另一反应基团(B)反应以形成共价键。步骤3使用烟草酸性磷酸酶去除7-甲基鸟苷帽,其导致标记为D的RNA链。然后有可能以多种不同方式处理该RNA(去除了7-甲基鸟苷帽,标记为D),从而制备连接到DNA链的RNA链(步骤4,6,7,8或11)。在步骤4中,一反应基团(标记为B并如方形所示)添加到RNA链D的5'末端。步骤5显示一化学反应步骤,其中DNA链也具有连接的标记为C并显示为圆形的反应基团,其与另一反应基团(B)反应以形成共价键。步骤6和7显示使用例如,T4RNA聚合酶1,T4RNA聚合酶2,热稳定的5'App DNA/RNA连接酶等可将DNA链(标记为E)直接地连接到RNA。在步骤7中,一酶(标记为G)预先结合到该DNA,而在步骤6中没有预先结合酶。步骤8和11显示具有前导区(标记为F)的DNA引物杂交到RNA链D。在步骤11中,一酶(标记为G)预先结合到该DNA引物,而在步骤8中没有预先结合酶。步骤9和12显示RNA链D的反转录,其导致3个C的3'突出。步骤10和13显示DNA发夹(标记为H)连接到双链DNA/RNA(标记为I)。

[0030] 图2显示了代表将原核RNA的链(如虚线所示)连接到DNA的链(如实线所示)的方法的卡通示意图。反应步骤1使用,例如PolyA聚合酶将poly(dA)尾(标记为A)添加到RNA链的3'末端获得链B。步骤2和5显示具有前导区(标记为C)的DNA引物杂交到poly(dA)区域(标记为A)。在步骤5中,一酶(标记为D)预先结合到该DNA引物,而在步骤2中没有预先结合酶。步骤3和6显示RNA链B的反转录,其导致1至3个C 3'突出。步骤4和7显示DNA发夹(标记为E)连接到双链DNA/RNA(标记为F)。反应步骤8,10和11直接发生在原核RNA上,而不是通过将poly(dA)区域添加到RNA。在步骤8中,一反应基团(标记为G并如方形所示)添加到RNA链的5'末端。步骤9显示一化学反应步骤,其中DNA链也具有连接的反应基团(标记为H并且该反应基团如圆形所示),其与另一反应基团(G)反应以形成共价键。步骤10和11显示使用例如,T4RNA聚合酶1,T4RNA聚合酶2,热稳定的5'App DNA/RNA连接酶等可将DNA链(标记为I)直接连接到RNA。在步骤11中,一酶(标记为D)预先结合与DNA结合,而在步骤10中没有预先结合

酶。

[0031] 图3示出了实施例1中的DNA/RNA链的卡通示意图,其使用DNA解旋酶(T4Dda-E94C/A360C (SEQ ID NO:14,具有突变E94C/A360C,然后( $\Delta$ M1)G1))移位穿过MspA纳米孔。区域A对应于40个核苷酸聚(dT)前导区(SEQ ID NO:16),DNA解旋酶(标记为B)与其结合。区域A与4个iSpC3间隔器(显示为X并标记为C)连接。区域D对应于合成的RNA区域(SEQ ID NO:17)。区域E对应于可变长度的聚(U)。区域F是杂交到区域D的DNA(SEQ ID NO:18)。与DNA(SEQ ID NO:18)连接的是6个iSp18间隔器(显示为Y并标记为G),2个胸腺嘧啶(显示为T并标记为H)和3'胆固醇TEG(标记为I)。

[0032] 图4显示在RNA poly(U)聚合酶延长步骤(实施例1步骤1.1)之后的5%PAGE TBE BioRad标准凝胶(在140V运行40分钟)。泳道1显示100bp TriDye梯。泳道2显示在聚合作用前的合成的DNA/RNA 1(SEQ ID NO:16在其3'末端连接到4个iSpC3间隔器,所述4个iSpC3间隔器在相对端连接到SEQ ID NO:17的5'末端)。泳道3显示poly(U)聚合混合物。泳道4显示纯化的样品1,其包含延长的DNA/RNA 2的poly(U)。箭头X对应于非延长的DNA/RNA 1并且箭头Y对应于poly(U)延长的DNA/RNA 2。

[0033] 图5显示解旋酶控制的DNA移动的迹线实例(y轴=电流(pA),x轴=时间(s)),其中DNA解旋酶(T4Dda-E94C/A360C (SEQ ID NO:14具有突变E94C/A360C,然后( $\Delta$ M1)G1))控制DNA/RNA 2(卡通示意图如图3所示)的移动。区域1对应于poly(dT)前导区(SEQ ID NO:16),区域2对应于iSpC3间隔器(该间隔器允许更大量的电流流过纳米孔,与DNA或RNA区域相比),区域3对应于RNA序列(SEQ ID NO:17),区域4对应于在实施例1步骤1.1中添加的可变长度的poly(U) RNA区域。

[0034] 图6显示使用拓扑异构酶将dsDNA连接到RNA的链上的可行方法。在该卡通图中,dsDNA与RNA的链连接,该RNA在其5'末端具有游离羟基(真核RNA的5'末端出现的7-甲基鸟苷帽和5'磷酸需要用例如磷酸酶诸如南极磷酸酶或碱性磷酸酶去除)。该牛痘拓扑异构酶结合到dsDNA的所示序列上(步骤1)。该红色箭头突出显示在dsDNA的下游链中第二个胸腺嘧啶对面有一缺口。一旦该拓扑异构酶已结合,其切割第二个胸腺嘧啶后的DNA的上游链,并保持与dsDNA结合(步骤2)。该与dsDNA结合的拓扑异构酶之后与具有游离5'羟基的RNA孵育(步骤3)。该拓扑异构酶然后将该RNA连接到dsDNA。

[0035] 图7显示使用拓扑异构酶将dsDNA连接到RNA的链上的另一个可行方法。在该卡通图中,使用DNA的单链区域将dsDNA连接到RNA的链从而杂交到该RNA并帮助RNA的连接。牛痘拓扑异构酶结合到dsDNA的所示序列上(步骤1)。DNA的下游链中没有缺口。一旦该拓扑异构酶已结合,其只切割第二个胸腺嘧啶后的DNA的上游链,并保持与dsDNA/ssDNA结合(步骤2)。结合到dsDNA的拓扑异构酶之后与具有游离5'羟基的RNA孵育(步骤3)。该拓扑异构酶然后将该RNA连接到dsDNA。ssDNA区域帮助将互补的RNA序列吸引到至DNA的连接位点。

[0036] 图8显示包含来自实施例2的样品的10%PAGE TBE-Urea变性标准凝胶在140V运行60分钟。泳道1显示100bp TriDye梯。泳道2显示了在T4DNA连接酶不存在下经过实施例2描述的连接反应之后的DNA寡核苷酸1(1X溶度,SEQ ID NO:21),RNA寡核苷酸1(1X浓度,SEQ ID NO:19)和DNA夹板(0.5X浓度,SEQ ID NO:20)。泳道3显示了在T4DNA连接酶存在下经过实施例2描述的连接反应之后的DNA寡核苷酸1(1X浓度,SEQ ID NO:21),RNA寡核苷酸1(1X浓度,SEQ ID NO:19)和DNA夹板(4X浓度,SEQ ID NO:20)。泳道4显示了未反应的

DNA寡核苷酸 1 (SEQ ID NO:21) 和DNA夹板 (SEQ ID NO:20) 混合在一起作为对照。泳道5显示了在T4DNA连接酶存在下经过实施例 2描述的连接反应之后的DNA寡核苷酸 1 (1X浓度, SEQ ID NO:21), RNA寡核苷酸 1 (1X浓度, SEQ ID NO:19) 和DNA夹板 (0.5X浓度, SEQ ID NO:20)。泳道6显示了在T4DNA连接酶存在下经过实施例 2描述的连接反应之后的DNA寡核苷酸 1 (1X浓度, SEQ ID NO:21), RNA寡核苷酸 1 (1X浓度, SEQ ID NO:19) 和DNA夹板 (1X浓度, SEQ ID NO:20)。泳道7显示了在T4DNA连接酶存在下经过实施例 2描述的连接反应之后的DNA寡核苷酸 1 (1X浓度, SEQ ID NO:21), RNA寡核苷酸 1 (1X浓度, SEQ ID NO:19) 和DNA夹板 (2X浓度, SEQ ID NO:20)。泳道8显示了在T4DNA连接酶存在下经过实施例2描述的连接反应之后的DNA寡核苷酸 1 (1X浓度, SEQ ID NO:21), RNA寡核苷酸 1 (1X浓度, SEQ ID NO:19) 和DNA夹板 (4X浓度, SEQ ID NO:20)。泳道9显示了在T4DNA连接酶存在下经过实施例 2描述的连接反应, 并且其在连接步骤之后进一步与DNA夹板 (4.5X浓度的SEQ ID NO:20) 孵育之后的DNA寡核苷酸 1 (1X浓度, SEQ ID NO:21), RNA寡核苷酸 1 (1X浓度, SEQ ID NO:19) 和DNA夹板 (0.5X浓度, SEQ ID NO:20)。泳道10显示了在T4DNA连接酶存在下经过实施例 2描述的连接反应, 并进一步热处理并暴露于ExoI之后的DNA寡核苷酸 1 (1X浓度, SEQ ID NO:21), RNA寡核苷酸 1 (1X浓度, SEQ ID NO:19) 和DNA夹板 (1X浓度, SEQ ID NO:20)。泳道11显示了作为对照的DNA寡核苷酸 1 (SEQ ID NO:21)。泳道12显示了作为对照的RNA寡核苷酸 1。泳道13显示了未反应的RNA寡核苷酸 1 (SEQ ID NO:19) 和DNA夹板 (SEQ ID NO:20) 混合在一起作为对照。标记为1的条带对应于RNA寡核苷酸 1 (SEQ ID NO:19)。标记为2的条带对应于杂交到夹板 (SEQ ID NO:20) 的RNA寡核苷酸 1 (SEQ ID NO:19)。标记为3的条带对应于夹板 (SEQ ID NO:20)。标记为4的条带对应于DNA寡核苷酸 1 (SEQ ID NO:21)。标记为5的条带对应于杂交到夹板 (SEQ ID NO:20) 的DNA寡核苷酸 1 (SEQ ID NO:21)。标记为A的区域对应于在夹板存在下连接的底物 (连接到RNA寡核苷酸1的DNA寡核苷酸1, 杂交到夹板)。标记为B的区域对应于在夹板不存在下连接的底物 (与RNA寡核苷酸1结合的DNA寡核苷酸1)。

[0037] 图9显示5%PAGE TBE BioRad标准凝胶在140mV下运行60分钟, (A) SYBR染色之前和 (B) SYBR染色之后, 其包括来自实施例3的样品。泳道1和7显示TriDye 1kD梯。泳道2和8显示产自实施例 3A的产物 (在DNA夹板X1 (SEQ ID NO:24) 的存在下, DNA X1 (SEQ ID NO:22在其3' 末端连接4个iSp18间隔器, 所述4个iSp18间隔器在其相对末端连接SEQ ID NO:23的5' 末端, SEQ ID NO:23的具有连接在其3' 末端的3叠氮化物N) 与具有 5' -己基-G的萤火虫荧光素酶mRNA反应 (RNA X1, 具有SEQ ID NO:26的开放阅读框的mRNA, 所述SEQ ID NO:26具有 5' -己基-G作为大多数5' 核苷酸并具有 3' 聚A尾))。泳道3和9显示产自实施例 3B的产物 (在DNA夹板X1 (SEQ ID NO:24) 的存在下, DNA X2 (SEQ ID NO:25在其5' 末端连接CY3, 并在其3' 末端连接3叠氮化物N) 与具有5' -己基-G的萤火虫荧光素酶mRNA (RNA X1, 具有SEQ ID NO:26的开放阅读框的mRNA, 所述SEQ ID NO:26具有5' -己基-G作为大多数5' 核苷酸并具有 3' 聚A尾) 反应)。泳道4和10显示DNA X2 (SEQ ID NO:25, 具有连接到其5' 末端的CY3, 和具有连接到其3' 末端的3叠氮化物N) 与具有5' -己基-G的萤火虫荧光素酶mRNA (RNA X1, 具有SEQ ID NO:26的开放阅读框的mRNA, 所述SEQ ID NO:26具有5' -己基-G作为大多数5' 核苷酸并具有 3' 聚A尾) 混合。泳道5和11显示DNA X1 (SEQ ID NO:22在其3' 末端连接4个iSp18间隔器, 所述4个iSp18间隔器在其相对末端连接SEQ ID NO:23的5' 末端, SEQ ID NO:23的具有连接在其3' 末端的3叠氮化物N)。泳道6和12显示具有5' -己基-G的萤火虫荧光素酶mRNA



(RNA X1, 具有SEQ ID NO:26的开放阅读框的mRNA, 所述SEQ ID NO:26具有5'-己基-G作为大多数5' 核苷酸并具有3' 聚A尾)。标记为1的条带对应于具有和不具有连接的DNA的RNA X1。条带2对应于未反应的DNA X1 (SEQ ID NO:22在其3' 末端连接4个iSp18间隔器, 所述4个iSp18间隔器在其相对末端连接SEQ ID NO:23的5' 末端, SEQ ID NO:23的具有连接在其3' 末端的3叠氮化物N)。条带3对应于未反应的DNA X2 (SEQ ID NO:25, 具有连接到其5' 末端的CY3和具有连接到其3' 末端的3叠氮化物N)。在非SYBR染色的凝胶上只有包含CY3标签的DNA是可见的。

[0038] 图10显示实施例3A中制备的DNA/RNA链的卡通示意图, 其使用DNA解旋酶 (T4Dda-E94C/A360C (SEQ ID NO:14, 具有突变E94C/A360C, 并然后 ( $\Delta$ M1)G1)) 移位穿过MspA纳米孔。区域A对应于与所述DNA解旋酶 (标记为B) 结合的DNA前导区 (SEQ ID NO:22)。区域A连接到4个iSp18间隔器 (显示为X并标记为C)。区域D对应于第二DNA序列 (SEQ ID NO:23)。区域E对应于具有5'-己基-G的萤火虫荧光素酶mRNA (RNA X1, 具有SEQ ID NO:26的开放阅读框的mRNA, 所述SEQ ID NO:26具有5'-己基-G作为大多数5' 核苷酸并具有3' 聚A尾)。区域D和E通过点击化学连接。反应的叠氮化物和己基基团由盒子 (标记为F) 表示。区域G是杂交到区域D的DNA (SEQ ID NO:18)。6个iSp18间隔器 (显示为Y并标记为H), 2个胸腺嘧啶 (显示为T并标记为I) 和3' 胆固醇TEG (标记为J) 连接到DNA (SEQ ID NO:18)。

[0039] 图11显示解旋酶控制的DNA移动的迹线实例 (y轴=电流 (pA), x轴=时间 (s)), 其中所述DNA解旋酶 (T4Dda-E94C/A360C (SEQ ID NO:14具有突变E94C/A360C然后 ( $\Delta$ M1)G1)) 控制由实施例3A制备的DNA/RNA产物 (卡通示意图如图3所示) 的移动。区域1对应于没有连接到RNA的DNA前导区 (DNA X1), 区域2对应于与4个iSp18间隔器 (该间隔器允许更大量的电流流过纳米孔, 与DNA或RNA区域相比) 连接的SEQ ID NO:22, 区域3对应于DNA序列 (SEQ ID NO:23) 和区域4对应于萤火虫荧光素酶mRNA区域 (RNA X1, 具有SEQ ID NO:26的开放阅读框的mRNA, 所述SEQ ID NO:26具有5'-己基-G作为5' 大多数核苷酸并具有3' 聚A尾)。标记为\*的电流峰值对应于当DNA和RNA连接时产生的点击连接的移位。

[0040] 图12显示将真核RNA的链 (如虚线所示) 连接到DNA的链 (如实线所示) 的方法的卡通示意图。该真核RNA具有7-甲基鸟苷帽, 其沿相反方向运动到RNA链的末端 (即5' 末端包含反向碱基, 显示为星形并标记为A)。该真核RNA与包含反向DNA碱基 (显示为n的随机序列并标记为C) 的DNA (标记为B) 的链连接。在反向碱基的区域中, 该碱基也沿相反方向运动。这说明了在RNA和DNA上的反向碱基的两个区域可连接在一起。

[0041] 图13显示实施例4中制备的RNA/cDNA构建体的卡通示意图。区域B对应于SEQ ID NO:30。区域C对应于SEQ ID NO:30中的最后3个腺嘌呤。区域D对应于3T和10T发夹中的SEQ ID NO:29。区域E对应于4个iSpC3间隔器 (显示为X)。区域F对应于3T发夹中的SEQ ID NO:27和10T发夹中的SEQ ID NO:28。区域G对应于SEQ ID NO:27或28中的最后3个胸腺嘧啶。区域H对应于在RNA链SEQ ID NO:30的反转录期间制备的cDNA。

[0042] 图14显示10%PAGE TBE-尿素变性凝胶 (在140V运行60分钟), 其显示在发夹连接和反转录之前和之后的来自实施例4的多个样品。泳道1显示TriDye梯。泳道2显示在连接或反转录之前使用在实施例4中的RNA链 (SEQ ID NO:30)。泳道3显示在连接或反转录之前的10T发夹控制 (SEQ ID NO:29在其5' 末端连接磷酸基团并在其3' 末端连接4个iSpC3间隔器, 所述4个iSpC3间隔器在其相对末端连接SEQ ID NO:28的5' 末端)。泳道4显示在使用引物的

反转录之后的RNA链(SEQ ID NO:30)。泳道5显示与3T发夹(SEQ ID NO:29在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接SEQ ID NO:27的5'末端)连接后的RNA链(SEQ ID NO:30)。泳道6显示与3T发夹(SEQ ID NO:29在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接SEQ ID NO:27的5'末端)连接后的RNA链(SEQ ID NO:30)并接着进行反转录。泳道7显示与10T发夹(SEQ ID NO:29在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接SEQ ID NO:28的5'末端)连接后的RNA链(SEQ ID NO:30)。泳道8显示与10T发夹(SEQ ID NO:29在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接SEQ ID NO:28的5'末端)连接后的RNA链(SEQ ID NO:30)并接着进行反转录。泳道9显示一对照实验,其中RNA链和10T发夹在缺少连接酶的情况下一起孵育。箭头A对应于实施例4中使用的RNA链(SEQ ID NO:30)。

[0043] 图15显示代表实施例5中制备的DNA/RNA/cDNA链的卡通示意图,其使用DNA解旋酶(T4Dda-E94C/C109A/C136A/A360C(SEQ ID NO:14具有突变E94C/C109A/C136A/A360C,然后( $\Delta$ M1)G1))移位穿过MspA纳米孔。区域A对应于与所述DNA解旋酶(标记为B)结合的DNA前导区(SEQ ID NO:22)。区域A与4个iSp18间隔器(显示为X并标记为C)连接。区域D对应于第二DNA序列(SEQ ID NO:23)。区域E对应于具有5'-己基-G区域的萤火虫荧光素酶mRNA(RNA X1,具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为5'大多数核苷酸并具有3'聚A尾)。区域D和E通过点击化学连接。反应的叠氮化物和己基基团由盒子(标记为F)表示。区域G是杂交到区域D的DNA(SEQ ID NO:18)。连接到DNA(SEQ ID NO:18)的是6个iSp18间隔器(显示为Y并标记为H),2个胸腺嘧啶(显示为T并标记为I)和3'胆固醇TEG(标记为J)。区域K对应于FLuc mRNA(RNA X1,具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为大多数5'核苷酸并具有3'聚A尾)的最后3个腺嘌呤。区域L对应于10T发夹中的SEQ ID NO:29。区域M对应于4个iSpC3间隔器(显示为X)。区域N对应于10T发夹中的SEQ ID NO:28。区域O对应于SEQ ID NO:28中的最后3个胸腺嘧啶。区域P对应于cDNA,其是在RNA链RNA X1的反转录期间制备的,RNA X1即具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为大多数5'核苷酸并具有3'聚A尾,。

[0044] 图16显示解旋酶控制的DNA移动的迹线实例(y轴=电流(pA),x轴=时间(s)),该DNA解旋酶(T4Dda-E94C/C109A/C136A/A360C(SEQ ID NO:14具有突变E94C/C109A/C136A/A360C,然后( $\Delta$ M1)G1))控制由实施例5制备的DNA/RNA/cDNA产物(卡通示意图如图15所示)的移动。区域1对应于DNA前导区(DNA X1),区域2对应于萤火虫荧光素酶mRNA(RNA X1,具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为大多数5'核苷酸并具有3'聚A尾),区域3对应于4个iSpC3间隔器,区域4对应于10T的聚T区域,区域5对应于由mRNA反转录制备的cDNA。

[0045] 图17显示5%PAGE TBE-尿素变性凝胶(在140V运行60分钟),其显示在脱帽和连接之前和之后来自实施例5的多个样品。泳道1显示TriDye梯。泳道2显示加帽的RNA链(SEQ ID NO:30,其具有通过5'至5'三磷酸连接在链5'末端连接的7-甲基鸟苷帽)。泳道3显示非-RNA多核苷酸(30个SpC3间隔器连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接

4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)。泳道4显示与所述非-RNA多核苷酸(30个SpC3间隔器连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)结合之后的RNA链(SEQ ID NO:30)。箭头A对应于RNA链(SEQ ID NO:30)。箭头B对应于所述非-RNA多核苷酸。箭头C对应于连接的产物,其中RNA链已与所述非-RNA多核苷酸连接。

[0046] 图18说明了来自具有修饰或不具有修饰的碱基的RNA的平均(consensus)电流水平。本图显示修饰的碱基影响几个连续电流水平。

[0047] 图19显示DNA解旋酶控制的RNA-DNA 2D链(RNA-正义和DNA-反义)穿过纳米孔的移位。图19说明观察RNA相对于DNA的不同平均振幅和范围。

[0048] 图20显示解旋酶控制的DNA移动的迹线实例(y轴=电流(pA),x轴=时间(s)),其中所述DNA解旋酶(He1308Mbu-E284C/S615C(具有突变E284C/S615C的SEQ ID NO:8))控制由实施例 8制备的DNA/RNA产物的移动。

[0049] 图21显示解旋酶控制的DNA移动(y轴=电流(pA),x轴=时间(s))穿过胞溶素突变体的迹线的实例,其中所述DNA解旋酶(T4Dda-E94C/A360C(0.36 $\mu$ l,3.8 $\mu$ M),SEQ ID NO:14具有突变E94C/A360C然后( $\Delta$ M1)G1)控制由实施例 3制备的DNA/RNA产物的移动。

[0050] 图22显示解旋酶控制的DNA移动(y轴=电流(pA),x轴=时间(s))穿过CsgG突变孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C))9(具有突变Y51T/F56Q的SEQ ID NO:44,其中StepII(C)是SEQ ID NO:45并连接在其C-末端)),其中所述DNA解旋酶(He1308Mbu-E284C/S615C(具有突变E284C/S615C的SEQ ID NO:8))控制由实施例 5制备的DNA/RNA产物的移动。

[0051] 序列列表的说明

[0052] SEQ ID NO:1显示了密码子优化的多核苷酸序列,其编码MS-B1突变体MspA单体。该突变体缺少信号序列,并包含下列突变:D90N,D91N,D93N,D118R,D134R和E139K。

[0053] SEQ ID NO:2显示了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,并包含下列突变:D90N,D91N,D93N,D118R,D134R和E139K。

[0054] SEQ ID NO:3显示了编码 $\alpha$ -溶血素-E111N/K147N( $\alpha$ -HL-NN;Stoddart等人PNAS,2009;106(19):7702-7707)的一个单体的多核苷酸序列。

[0055] SEQ ID NO:4显示了 $\alpha$ -HL-NN的一个单体的氨基酸序列。

[0056] SEQ ID NO:5至7显示了MspB,C和D的氨基酸序列。

[0057] SEQ ID NO:8显示了He1308Mbu的氨基酸序列。

[0058] SEQ ID NO:9显示了He1308Csy的氨基酸序列。

[0059] SEQ ID NO:10显示了He1308Tga的氨基酸序列。

[0060] SEQ ID NO:11显示了He1308Mhu的氨基酸序列。

[0061] SEQ ID NO:12显示了TraI Eco的氨基酸序列。

[0062] SEQ ID NO:13显示了XPD Mbu的氨基酸序列。

[0063] SEQ ID NO:14显示了Dda 1993的氨基酸序列。

[0064] SEQ ID NO:15显示了Trwc Cba的氨基酸序列。

- [0065] SEQ ID NO:16显示了实施例 1中使用的多核苷酸序列。
- [0066] SEQ ID NO:17显示了实施例 1中使用的多核苷酸序列。
- [0067] SEQ ID NO:18显示了实施例 1和3中使用的多核苷酸序列。
- [0068] SEQ ID NO:19显示了实施例 2中使用的多核苷酸序列。所述序列具有5' 磷酸基团。
- [0069] SEQ ID NO:20显示了实施例 2中使用的多核苷酸序列。
- [0070] SEQ ID NO:21显示了实施例 2中使用的多核苷酸序列。所述序列具有5' CY3基团。
- [0071] SEQ ID NO:22显示了实施例 3中使用的多核苷酸序列。
- [0072] SEQ ID NO:23显示了实施例 3中使用的多核苷酸序列。所述序列具有3' 叠氮(AzideN)基团。
- [0073] SEQ ID NO:24显示了实施例 3中使用的多核苷酸序列。
- [0074] SEQ ID NO:25显示了实施例 3中使用的多核苷酸序列。所述序列具有3' 叠氮(AzideN)基团和5' CY3基团。
- [0075] SEQ ID NO:26显示了实施例3和5中使用的多核苷酸序列的开放阅读框。所述序列具有连接到该序列中首个G的5' -己基基团(5' -hexynyl group)。
- [0076] SEQ ID NO:27显示了实施例4和5中使用的多核苷酸序列。
- [0077] SEQ ID NO:28显示了实施例4中使用的多核苷酸序列。
- [0078] SEQ ID NO:29显示了实施例4和5中使用的多核苷酸序列。
- [0079] SEQ ID NO:30显示了实施例4和6中使用的多核苷酸序列。
- [0080] SEQ ID NO:31显示了实施例6中使用的多核苷酸序列。
- [0081] SEQ ID NO:32显示了实施例6中使用的多核苷酸序列。
- [0082] SEQ ID NO:33显示了用于说明均聚物阅读(homopolymer read)的序列。
- [0083] SEQ ID NO:34显示了用于说明均聚物阅读的序列。
- [0084] SEQ ID NO:35显示了实施例8中使用的多核苷酸序列。
- [0085] SEQ ID NO:36显示了实施例8中使用的多核苷酸序列。
- [0086] SEQ ID NO:37显示了实施例8中使用的多核苷酸序列。
- [0087] SEQ ID NO:38显示了实施例8中使用的多核苷酸序列。
- [0088] SEQ ID NO:39显示了实施例8中使用的多核苷酸序列。
- [0089] SEQ ID NO:40显示了实施例8中使用的多核苷酸序列。
- [0090] SEQ ID NO:41显示了编码胞溶素单体的多核苷酸序列。
- [0091] SEQ ID NO:42显示了所述胞溶素单体的氨基酸序列。
- [0092] SEQ ID NO:43显示了编码来自大肠杆菌Str.K-12substr.MC4100的野生型CsgG单体的密码子优化的多核苷酸序列。这个单体缺少信号序列。
- [0093] SEQ ID NO:44显示了来自大肠杆菌Str.K-12substr.MC4100的野生型CsgG单体的成熟形式的氨基酸序列。这个单体缺少信号序列。用于该CsgG的缩写=CsgG-Eco。
- [0094] SEQ ID NO:45显示了步骤II (C)的氨基酸序列。
- [0095] SEQ ID NO:46至 64是说明书中描述的多核苷酸序列。
- [0096] SEQ ID NO:65和 66是实施例中使用的多核苷酸序列。

## 具体实施方式

[0097] 应理解,公开的产品和方法的不同应用可以根据本领域的具体需求而调整。可以理解本文中使用的术语仅是为了描述本发明的具体实施方式的目的,而不意为对本发明的限制。

[0098] 另外,除非本文另有明确规定,否则本说明书和随附的权利要求中所使用的单数形式的“一”、“一个”和“所述”包括复数指代。因此,例如,涉及“多核苷酸”时包括两个或多核苷酸,涉及“多核苷酸结合蛋白质”包括两个或多个这样的蛋白,涉及“解旋酶”时包括两个或多个解旋酶,涉及“单体”指的是两个或多个单体,涉及“孔”时包括两个或多个孔,等。

[0099] 本文所引用的所有公开物、专利和专利申请,无论在前文或后文,均以引用的方式全文引入。

### [0100] 表征目标RNA多核苷酸

[0101] 本发明的方法包括表征目标RNA多核苷酸。所述RNA多核苷酸被运送到跨膜孔,并且用所述孔表征所述RNA多核苷酸。本发明提供了通过在DNA解旋酶的控制下,随着所述RNA多核苷酸相对于所述跨膜孔移动,获取一个或多个测量值来表征目标核糖核酸(RNA)多核苷酸的方法。

[0102] 由于跨膜孔能够检测目标多核苷酸的单个分子,从而不需要放大(扩增)所述目标RNA多核苷酸。所述方法通常不包括聚合酶链反应(PCR)或反转录PCR(RT-PCR)。这极大地减少了表征目标RNA多核苷酸需要的工作量。这也避免了由PCR引起的任何偏差和人为假象。

[0103] 本发明的方法可涉及确定或测量RNA多核苷酸的一个或多个特征。所述方法可包括确定或测量所述RNA多核苷酸的一个、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地选自(i)所述RNA多核苷酸的长度,(ii)所述RNA多核苷酸的同一性,(iii)所述RNA多核苷酸的序列,(iv)所述RNA多核苷酸的二级结构和(v)所述RNA多核苷酸是否是修饰的。根据本发明可以测量(i)至(v)的任何组合,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(i)至(v)的不同组合,包括以上列举的任何组合。本发明的方法优选地包含估计所述RNA多核苷酸的序列或对所述RNA多核苷酸测序。

[0104] 对于(i),所述RNA多核苷酸的长度例如可以通过确定所述RNA多核苷酸和所述孔之间的相互作用的次数,或所述RNA多核苷酸和所述孔之间的相互作用的持续时间来测定。

[0105] 对于(ii),所述RNA多核苷酸的同一性可以通过多种方式测定。所述RNA多核苷酸的同一性可联合所述RNA多核苷酸的序列测定来测定或不联合所述RNA多核苷酸的序列测定来测定。前者是直接的;对所述多核苷酸进行测序,并由此鉴定RNA多核苷酸的同一性。后者可以以几种方式来完成。例如,可以测定所述RNA多核苷酸中特定模序的存在(而无需测定所述RNA多核苷酸的其余序列)。或者,所述方法中测定的特定的电和/或光信号的测量值可鉴定来自特定来源的RNA多核苷酸。

[0106] 对于(iii),所述RNA多核苷酸的序列可以如前所述确定。合适的测序方法,特别是那些使用电测量值的方法,在Stoddart D et al.,Proc Natl Acad Sci,12:106(19):7702-7,Lieberman KR et al.,J Am Chem Soc.2010;132(50):17961-72,和国际申请WO

2000/28312中描述。

[0107] 对于(iv),所述二级结构可以多种方法测量。例如,如果所述方法包括电测量,所述二级结构可以利用穿过孔的停留时间的变化或电流变化来测量。这使得单链和双链RNA多核苷酸的区域被区别。

[0108] 对于(v),可以测定任何修饰的存在或不存在。所述方法优选地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,或用一个或多个标记物、标签或间隔器进行了修饰。特异性修饰将导致与孔的特异性相互作用,这可以使用下面描述的方法来测定。例如,可以基于孔与每个核苷酸的相互作用过程中穿过孔的电流,区别胞嘧啶与甲基胞嘧啶。本发明的方法可用于RNA和DNA的之间的区别,甚至在单一样品中:作为平均振幅和范围的函数,甚至RNA和DNA序列相同时, RNA和DNA可被相互区别。

[0109] 所述方法可使用任何适于研究膜/孔系统的设备来实施,其中在所述膜/孔系统中,孔存在于膜中。可使用任何适于跨膜孔传感的设备来实施所述方法。例如,所述设备包括一个室,所述室包括水溶液和将该室分割为两部分的屏障(barrier)。所述屏障通常具有缝隙,其中在缝隙中形成包括孔的膜。或者该屏障形成其中存在孔的膜。

[0110] 该方法可以使用在国际申请No. PCT/GB08/000562 (WO 2008/102120) 中描述的设备实施。

[0111] 该方法可以包括随着所述RNA多核苷酸相对于所述孔移动,测量通过所述孔的电流。因此该装置也可以包括能够跨膜和孔施加电势并测量电流信号的电路。该方法可以使用膜片钳或电压钳实施。所述方法优选包含使用电压钳。

[0112] 本发明的方法可包括随着所述RNA多核苷酸相对于所述孔移动来测量流过所述孔的电流。随着所述多核苷酸相对于所述孔移动流经所述孔的电流被用于确定所述目标RNA多核苷酸的序列。这是链测序。用于测量通过跨膜蛋白质孔的离子电流的合适的条件是本领域已知的,并且在实施例中公开。所述方法通过跨膜和孔施加的电压进行实施。使用的电压通常为+5V到-5V,例如从+4V到-4V,+3V到-3V或+2V到-2V。通常使用的电压通常为-600mV到+600mV,或-400mV到+400mV。使用的电压优选地在具有以下下限和上限的范围内,所述下限选自在-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。所用的电压更优选在100mV到240mV的范围内并最优选在120mV到220mV的范围内。可通过对孔施加提高的电势来提高对不同核糖核苷酸的分辨力。

[0113] 该方法通常在任何载荷子,如金属盐,例如碱金属盐,卤盐,例如氯盐,如碱金属氯盐的存在下实施。载荷子可包括离子型液体或有机盐,例如四甲基氯化铵,三甲苯基氯化铵,苯基三甲基氯化苯,或1-乙基-3-甲基咪唑鎓氯化物。在上面讨论的示例性设备中,所述盐存在于所述室中的水溶液中。通常使用氯化钾(KCl),氯化钠(氯化钠)或氯化铯(CsCl)或亚铁氰化钾和铁氰化钾的混合物。优选氯化钾,氯化钠和亚铁氰化钾和铁氰化钾的混合物。所述载荷子可以非对称地穿过所述膜。例如,在所述膜的各侧上载荷子的类型和/或浓度可能不同。

[0114] 盐浓度可为饱和的。盐浓度可以是3M或更低,通常为0.1M至2.5M,0.3M至1.9M,0.5M至1.8M,0.7M至1.7M,0.9M至1.6M或1M至1.4M。优选盐浓度为150mM到1M。所述方法优选使用至少为0.3M,例如至少为0.4M,至少为0.5M,至少为0.6M,至少为0.8M,至少为1.0M,至

少为1.5M,至少为2.0M,至少为2.5M,或者至少为3.0M的盐浓度进行实施。高盐浓度提供了高信噪比,并使得在正常电流波动的背景下,代表核糖核苷酸存在的电流能被识别。

[0115] 所述方法通常在缓冲剂存在下实施。在上面讨论的示例性设备中,缓冲剂在所述室中的水溶液中存在。本发明的方法可使用任何缓冲剂。通常地,缓冲剂是磷酸盐缓冲液。其他合适的缓冲剂为HEPES和Tris-HCl缓冲剂。该方法通常在pH值为4.0至12.0,4.5至10.0,5.0至9.0,5.5至8.8,6.0至8.7,7.0至8.8,或7.5至8.5下实施。所使用的pH优选约为7.5。

[0116] 所述方法可在0℃至100℃,15℃至95℃,16℃至90℃,17℃至85℃,18℃至80℃,19℃至70℃,或20℃至60℃下实施。所述方法通常在室温下进行。该方法可选地在支持酶功能的温度下,例如约37℃实施。

[0117] 所述方法可在游离核苷酸或游离核苷酸类似物和/或有利于发挥解旋酶或构建体的功能的酶辅因子的存在下实施。所述方法也可在游离核苷酸或游离核苷酸类似物不存在和酶辅因子不存在下实施。所述游离核苷酸可以是任何以上讨论的单个核苷酸的一种或多种。所述游离核苷酸包括,但不限于,单磷酸腺苷(AMP),二磷酸腺苷(ADP),三磷酸腺苷(ATP),单磷酸鸟苷(GMP),二磷酸鸟苷(GDP),三磷酸鸟苷(GTP),单磷酸胸苷(TMP),二磷酸胸苷(TDP),三磷酸胸苷(TTP),单磷酸尿苷(UMP),二磷酸尿苷(UDP),三磷酸尿苷(UTP),单磷酸胞苷(CMP),二磷酸胞苷(CDP),三磷酸胞苷(CTP),单磷酸环腺苷(cAMP),单磷酸环鸟苷(cGMP),单磷酸脱氧腺苷(dAMP),二磷酸脱氧腺苷(dADP),三磷酸脱氧腺苷(dATP),单磷酸脱氧鸟苷(dGMP),二磷酸脱氧鸟苷(dGDP),三磷酸脱氧鸟苷(dGTP),单磷酸脱氧胸苷(dTMP),二磷酸脱氧胸苷(dTDP),三磷酸脱氧胸苷(dTTP),单磷酸脱氧尿苷(dUMP),二磷酸脱氧尿苷(dUDP),三磷酸脱氧尿苷(dUTP),单磷酸脱氧胞苷(dCMP),二磷酸脱氧胞苷(dCDP)和三磷酸脱氧胞苷(dCTP)。所述游离核苷酸优选选自AMP,TMP,GMP,CMP,UMP,dAMP,dTMP,dGMP或dCMP。所述游离核苷酸优选腺苷三磷酸(ATP)。所述酶辅因子是使解旋酶或构建体发挥功能的因子。所述酶辅因子优选是二价金属阳离子。所述二价金属阳离子优选为 $Mg^{2+}$ , $Mn^{2+}$ , $Ca^{2+}$ 或 $Co^{2+}$ 。所述酶辅因子最优选为 $Mg^{2+}$ 。

[0118] 目标RNA

[0119] RNA是包含两个或多个核糖核苷酸的大分子。所述目标RNA多核苷酸可以是真核或原核RNA。所述目标RNA多核苷酸可以包含任何核糖核苷酸的任何组合。所述核糖核苷酸可以是天然存在的或人造的。所述目标RNA多核苷酸中的一个或多个核糖核苷酸可被氧化或甲基化。所述目标RNA中的一个或多个核糖核苷酸可被损伤。例如,所述目标RNA可以包含嘧啶二聚体,例如尿嘧啶二聚体。此类二聚体通常与紫外线导致的损伤相关,并是皮肤黑色素瘤的首要原因。所述目标RNA多核苷酸中的一个或多个核糖核苷酸可被修饰,例如用标记物或标签修饰。合适的标记物如下所述。所述目标RNA可包含一个或多个间隔器。

[0120] 核糖核苷酸含有碱基、核糖和至少一个磷酸基团。所述碱基通常为杂环的。碱基包括但不限于:嘌呤和嘧啶,更具体地,腺嘌呤、鸟嘌呤、胸腺嘧啶、尿嘧啶和胞嘧啶。所述核苷酸通常含有单磷酸,二磷酸或三磷酸。磷酸可被连接在核苷酸的5'或3'侧。

[0121] 核糖核苷酸包括但不限于,单磷酸腺苷(AMP),单磷酸鸟苷(GMP),单磷酸胸苷(TMP),单磷酸尿苷(UMP),单磷酸胞苷(CMP),单磷酸5-甲基胞苷,二磷酸5-甲基胞苷,三磷酸5-甲基胞苷,单磷酸5-羟甲基胞苷,二磷酸5-羟甲基胞苷和三磷酸5-羟甲基胞苷。所述核

苷酸优选选自AMP,TMP,GMP,CMP和UMP。

[0122] 核糖核苷酸可以是脱碱基的(即缺少碱基)。核糖核苷酸也可缺少碱基和糖(即是C3间隔区)。

[0123] 所述目标RNA多核苷酸的核糖核苷酸可以任何方式彼此连接。如在核酸中一样,所述核糖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,所述核糖核苷酸可通过它们的碱基连接。

[0124] RNA是非常多样的分子。所述目标RNA多核苷酸可以是任何天然产生的或合成的核糖核苷酸分子,例如,RNA,信使RNA(mRNA),核糖体RNA(rRNA),核不均一RNA(hnRNA),转移RNA(tRNA),转移信使RNA(tmRNA),微小RNA(miRNA),小核RNA(snRNA),小核仁RNA(snoRNA),信号识别颗粒(SRP RNA),SmY RNA,小卡侯氏体(Small Cajal body-specific) RNA(scaRNA),向导RNA(gRNA),剪接前导RNA(SL RNA),反义RNA(asRNA),长非编码RNA(lncRNA),Piwi-相互作用(Piwi-interacting) RNA(piRNA),小分子干扰RNA(siRNA),反式作用siRNA(tasiRNA),重复联系siRNA(rasiRNA),Y RNA,病毒性RNA或染色体的RNA,在适当情况下所有的RNA可以是单链的,双链的或三链的。

[0125] 所述目标RNA多核苷酸优选是信使RNA(mRNA)。所述目标mRNA可以是交替剪接变体(alternate splice variant)。mRNA和/或交替mRNA剪接变体的变异量(或等级)可能与疾病或健康状况有关。

[0126] 或者所述目标RNA多核苷酸是微小RNA(或miRNA)。很难在低浓度下检测到的一组RNAs是微小核糖核酸(micro-RNA或miRNAs)。miRNAs是高度稳定的RNA寡聚体,其能转录后调节蛋白质产物。它们通过两个机理中的一个起作用。在植物中,miRNAs已被证明主要通过引导信使RNA的分裂来起作用,然而在动物中,通过miRNAs的基因调控通常包括miRNAs到信使RNAs的3' UTRs的杂交,这阻碍了翻译(Lee et al.,Cell 75,843-54(1993);Wightman等,Cell 75,855-62(1993);和Esquela-Kerscher等,Cancer 6,259-69(2006))。miRNAs经常以有缺陷的互补与它们的目标结合。它们已被预测能与多达200个或更多的基因目标分别结合,并调节全人类超过三分之一的基因(Lewis等,Cell 120,15-20(2005))。

[0127] 用于本发明的合适的miRNAs是本领域已知的。例如,在公众可获得的数据库上存储的合适的miRNAs(Jiang Q.,Wang Y.,Hao Y.,Juan L.,Teng M.,Zhang X.,Li M.,Wang G.,Liu Y.,(2009) miR2Disease:a manually curated database for microRNA deregulation in human disease.Nucleic Acids Res.)。已知在肿瘤中某些microRNAs的表达水平会改变,产生了不同肿瘤类型特征模式的microRNA表达(Rosenfeld,N.等,Nature Biotechnology 26,462-9(2008))。另外,已知miRNA表达谱能够比信使RNA表达谱更精确地揭示肿瘤发展的阶段(Lu等,Nature 435,834-8(2005)和Barshack等,The International Journal of Biochemistry&Cell Biology 42,1355-62(2010))。这些发现,结合miRNAs的高稳定性,和在血清和血浆中检测循环的miRNAs的能力(Wang等,Biochemical and Biophysical Research Communications 394,184-8(2010);Gilad等,PloS One 3,e3148(2008);和Keller等,Nature Methods 8,841-3(2011)),引起了将microRNAs作为癌症生物标记物的潜在应用的大量兴趣。为了有效地治疗,癌症需要被精确地分类并不同地治疗,但是由于许多不同类型的癌症共享形态学特征这一事实,作为分类的方法的肿瘤形态评估的功效被削弱了。miRNAs提供了潜在地更可靠并低创伤性的解决方案。



[0128] mRNAs和miRNAs对诊断或预测疾病或身体状况的用途更详细地讨论如下。

[0129] 可研究任何数量的RNA。例如,本发明的方法可关注于确定3,4,5,6,7,8,9,10,20,30,50,100或更多个RNA分子的存在,缺失或一个或多个特征。

[0130] 所述多核苷酸可以是天然存在的或人工合成的。例如,所述方法可用于核实两个或多个人工制造的寡核苷酸的序列。所述方法通常在体外实施。

[0131] 所述目标RNA多核苷酸可以是任意长度。例如,所述RNA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个核糖核苷酸长度。所述目标RNA可以是1000或更多个核糖核苷酸,5000或更多个核苷酸或在100000或更多核糖核苷酸长度。全部或只有部分的所述目标RNA可以使用这种方法表征。要被测序的RNA部分优选包含全部目标分子,但可以例如比全部分子少,例如,4个碱基到1kb,例如,4到100个碱基。

[0132] 所述目标RNA多核苷酸通常存在于或来源于任何合适样本中。本发明通常在已知含有或怀疑含有所述目标RNA多核苷酸的样品中实施。替代地,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RNAs身份。

[0133] 所述样品可以是生物样品。本发明可以针对从任何生物体或微生物中获得或提取的样品在体外实施。所述生物体或微生物通常是古细菌的 (archaeal), 原核的或真核的, 并且通常属于以下五界中的一个: 植物界, 动物界, 真菌, 原核生物和原生生物。所述目标RNA多核苷酸可来源于真核细胞或可来源于使用真核细胞的转录机制的病毒。本发明可以针对从任何病毒中获得或提取的样品在体外实施。

[0134] 所述样品优选是液体样品。样品通常包括病人的体液。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, 羊或猪或替代地可以是宠物如猫或狗。或者, 来源于植物的样品通常从商业作物, 如谷类, 豆类, 水果或蔬菜, 例如小麦, 大麦, 燕麦, 菜籽油 (canola), 玉米, 大豆, 水稻, 香蕉, 苹果, 番茄, 土豆, 葡萄, 烟草, 菜豆 (beans), 小扁豆, 甘蔗, 可可或棉花。

[0135] 所述样品可以是非生物样品。所述非生物样品优选为流体样品。非生物样品的实例包括手术液, 水如饮用水、海水或河水, 以及用于实验室试验的试剂。

[0136] 所述样品通常是在被分析前处理, 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。样品也可通常在分析前被存储, 优选低于-70℃存储。所述目标RNA多核苷酸在用于本发明的方法之前通常从所述样品中提取。RNA提取试剂盒是可从例如, New England Biolabs®和Invitrogen®s商业获得的。

[0137] 目标RNA的修饰

[0138] 对所述RNA多核苷酸的修饰可以是任何促进DNA解旋酶与修饰的RNA多核苷酸结合的修饰和/或导致或具有增大的DNA解旋酶对修饰的RNA多核苷酸结合的效果的修饰。本文使用的术语“结合”指的是亲和力, 或所述DNA解旋酶和底物多核苷酸将在任何给定时间结合的可能性。生物化学上, 这种对亲和力的增加是由“连接速率” (“on rate”) 或结合速率的提高或“解开速率” (“off rate”) 或解绑速率的降低, 或同时“连接速率”提高和“解开速率”降低引起的。

[0139] RNA多核苷酸的修饰可以增加至少10%, 优选至少20%, 至少30%, 至少40%, 至少

50%，至少60%，至少70%，至少80%或至少90%的DNA解旋酶对修饰的RNA的亲合力，最优选所述RNA多核苷酸的修饰可以增加95%或更多的所述DNA解旋酶的亲合力。促进的DNA解旋酶对修饰的RNA的结合被定义为这样的情况，即与非修饰的或无修饰的RNA多核苷酸相比，DNA解旋酶更容易地与修饰的RNA多核苷酸结合。增加的DNA解旋酶对修饰的RNA的结合被定义为DNA解旋酶结合的数量或水平大于，或多于，从非修饰的或无修饰的RNA多核苷酸观察到的DNA解旋酶结合的数量或水平，所述非修饰的或无修饰的RNA多核苷酸即根据本发明的修饰方法没有被修饰的RNA。DNA解旋酶对目标RNA多核苷酸结合的水平可容易地通过使用本领域技术人员已知和惯常的例行方法检测。

[0140] 所述目标RNA多核苷酸被修饰以包含非-RNA多核苷酸，例如多核苷酸区域或序列或构建体。所述非-RNA多核苷酸的至少一个核苷酸不是RNA。所述非-RNA多核苷酸可以因此包含核糖核苷酸或RNA核苷酸但必须也包含或含有非-RNA核苷酸或序列，即不是RNA的核苷酸或序列。通过对目标RNA多核苷酸添加或连接非-RNA多核苷酸(其可以包含或不包含核糖核苷酸或RNA核苷酸)来修饰所述目标RNA多核苷酸以形成本发明的构建体。所述非-RNA多核苷酸添加或连接到所述RNA多核苷酸意味着增加了所述DNA解旋酶和修饰的RNA构建体之间的相互作用，即，与发生在DNA解旋酶和未修饰形式的没有连接非-RNA多核苷酸的RNA多核苷酸之间的相互作用相比。额外地或替换地，所述非-RNA多核苷酸到所述RNA多核苷酸的添加或连接意味着增加了DNA解旋酶对于修饰的RNA构建体的特异性，即与DNA解旋酶对于未修饰形式的没有连接非-RNA多核苷酸的RNA多核苷酸的特异性相比。额外地或替换地，所述非-RNA多核苷酸添加或连接到所述RNA多核苷酸意味着促进了DNA解旋酶与修饰的RNA构建体的结合和/或增加了DNA解旋酶与修饰的RNA构建体的结合，即与发生在DNA解旋酶和未修饰形式的没有连接非-RNA多核苷酸的RNA多核苷酸之间的结合相比。额外地或替换地，所述非-RNA多核苷酸添加或连接到所述RNA多核苷酸意味着所述DNA解旋酶更有效或更强地与修饰的RNA构建体结合并不太可能从修饰的构建体中脱离，即与发生在DNA解旋酶和未修饰形式的没有连接非-RNA多核苷酸的RNA多核苷酸之间的结合相比。优选所述RNA多核苷酸的修饰减少了至少10%，优选至少20%，至少30%，至少40%，至少50%，至少60%，至少70%，至少80%或至少90%DNA解旋酶从所述RNA多核苷酸的释放。最优选所述RNA多核苷酸的修饰减少了至少95%或更多所述DNA解旋酶从所述RNA多核苷酸的释放。

[0141] 非-RNA多核苷酸

[0142] 目标RNA多核苷酸序列被修饰以包含非-RNA多核苷酸。所述目标RNA连接到非-RNA多核苷酸。所述目标RNA优选共价地连接到非-RNA多核苷酸。所述非-RNA多核苷酸必须包含至少一个不是核糖核苷酸，即不是来自RNA的核苷酸。所述非-RNA多核苷酸可另外包含核糖核苷酸或RNA，但必须也包含或包括至少一个非-RNA核苷酸，即不是RNA的核苷酸。通常包含RNA的非-RNA多核苷酸含有小于20个的RNA核苷酸，例如1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,或19个RNA核苷酸。所述非-RNA多核苷酸可以因此是“杂交的”多核苷酸，包括例如，RNA和另外的多核苷酸诸如DNA或DNA类似物。非-RNA也可以包括DNA间隔器等。本领域技术人员将会知道任何描述为适合制备本发明的修饰的RNA构建体的连接方法同样地适合制备所述非-RNA多核苷酸，其中可以结合两种或多种类型的核酸序列。

[0143] 优选所述非-RNA多核苷酸包含前导序列。所述前导序列优先地旋入所述孔。

[0144] 优选地通过将非-RNA前导序列连接到RNA来修饰所述目标RNA多核苷酸。所述前导

序列有助于本发明的表征方法。所述前导序列被设计为优先地穿入所述孔,并因此促进多核苷酸穿过孔的移动。所述前导序列也可以用于将所述目标RNA多核苷酸连接到如下所述的一个或多个锚。所述前导序列可连接到所述目标RNA多核苷酸。

[0145] 所述前导序列通常包含聚合物区域。所述聚合物区域优选带负电荷。所述聚合物优选为多核苷酸,例如DNA,修饰的多核苷酸(例如无碱基DNA),PNA,LNA,聚乙二醇(PEG)或多肽。所述前导序列优选包含单链或双链多核苷酸。

[0146] 所述单链前导序列可包含单链的DNA,iSpC3,聚dT部分或聚dC部分。所述前导序列优选包含一个或多个间隔器。所述前导序列优选是以下描述的Y适配器的一部分。所述前导序列可以是任意长度,但通常是10至200个核苷酸的长度。在本发明的一个实施例中所示非-RNA前导序列包含iSpC3。最优选地所述非-RNA前导序列是iSpC3或大约40个核苷酸长度的C和A(即4x(9C's和1A))的重复序列。所述前导序列的长度通常取决于本方法中使用的跨膜孔。

[0147] 所述非-RNA多核苷酸优选包含DNA解旋酶能够结合(DNA解旋酶结合位点)的区域或包含DNA适配器。所述目标RNA多核苷酸可被修饰以包含用于所述DNA解旋酶的DNA结合位点,所述DNA解旋酶控制所述目标RNA穿过跨膜孔的移动。本文中使用的术语“DNA解旋酶结合位点”包括足够尺寸/长度的DNA序列或DNA类似物序列以使得一个或多个DNA解旋酶结合至此。所述结合位点的长度依赖于要连接至此的解旋酶的数量。DNA解旋酶能够结合的区域优选为多核苷酸例如DNA,修饰的多核苷酸(例如无碱基DNA),PNA,LNA,聚乙二醇(PEG)。优选所述DNA解旋酶结合位点是单链的,没有杂交区域。所述区域可对应于所述前导序列。或者,所述区域可不同于所述前导序列。所述DNA解旋酶可帮助控制如下更详细讨论的RNA多核苷酸穿过孔的移动。

[0148] 所述非-RNA多核苷酸优选进一步具有阻塞位点(blocking site)或阻塞分子,其可邻近或非常靠近DNA结合位点,位于一个或多个解旋酶要移动到DNA位点的相对末端。所述阻塞分子阻止了所述解旋酶的向后移动并阻止了其脱离构建体。

[0149] 用于连接至目标RNA序列的非-RNA多核苷酸构建体优选包含:部分(i) 5个或更多个带电单元的聚合物,其优选用于通过孔捕获目标多核苷酸;和/或部分(ii) 大约20个核苷酸长度的阻塞链杂交位点;和/或部分(iii) 1个或多个非-RNA核苷酸,例如1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,30,40,或50个非-RNA核苷酸的DNA-解旋酶结合位点;和/或部分(iv) 抑制化学(stalling chemistry)的1个或多个单元,例如,1,2,3,4,5,6,7,8,9,10或更多个单元,例如W02014/135838中描述的Sp18,其通过引用并入本文;和/或部分(v) 大约30个核苷酸长度的系链杂交位点;和/或部分(vi) 促进所述非-RNA多核苷酸结合到所述RNA多核苷酸的序列。用于连接至目标RNA序列的所述非-RNA多核苷酸的全长可以因此包含大约50到200个核苷酸。

[0150] 例如,在一个实施例中,非-RNA多核苷酸可包含以下至少一个:(i) 5个或更多个带电单元的聚合物;(ii) 大约20个核苷酸长度的阻塞链杂交位点;(iii) 1个或多个非-RNA核苷酸,例如1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,30,40,或50个核苷酸的DNA-解旋酶结合位点;(iv) 抑制化学(stalling chemistry)的1个或多个单元,例如,1,2,3,4,5,6,7,8,9,10或更多个单元,例如W02014/135838中描述的Sp18,其通过引用并入本文;(v) 大约30个核苷酸长度的系链杂交位点;和/或(vi) 促进所述非-RNA多核苷酸结合



[0183] 9999

[0184] 其中9=PEG6磷酸(间隔器9(iSp9)基团)

[0185] 部分(v)

[0186] 部分(v)的所述非-RNA多核苷酸应允许系链寡核苷酸(tethering oligo)的杂交,所述系链寡核苷酸具有足够高的TM以形成稳定的杂交体。满足这些条件的序列的典型例子为:

[0187] AACTACTAGGATCATCGATGTATCTGCTCA (SEQ ID NO:56)

[0188] AGCTTAACATACGATACTCTTAGCTAACCA (SEQ ID NO:57)

[0189] rArArCrUrArCrUrArGrGrArUrCrArUrCrGrArUrGrUrArUrCrUrGrCrUrCrA (SEQ ID NO:58)

[0190] rArGrCrUrUrArArCrArUrArCrGrArUrArCrUrCrUrUrArGrCrUrArArCrCrA (SEQ ID NO:59)

[0191] 具有SEQ ID NO:56的序列的PNA

[0192] 具有SEQ ID NO:57的序列的PNA

[0193] 部分(vi)

[0194] 部分(vi)的所述非-RNA多核苷酸应促进所述非-RNA多核苷酸结合到所述RNA多核苷酸。满足这些条件的序列的典型例子为:

[0195] ACTCTGAACC (SEQ ID NO:60)

[0196] ACTCTrGrArArCrC (SEQ ID NO:61)

[0197] GCACAATGAT (SEQ ID NO:62)

[0198] GCACArArTrGrArT (SEQ ID NO:63)

[0199] 根据本发明可制备(i)到(vi)的任何组合:优选所述非-RNA多核苷酸包含(iii)和(vi)与{i},{ii},{iv},{v},{i,ii},{i,iv},{i,v},{ii,iv},{ii,v},{iv,v},{i,ii,iv},{i,ii,v},{i,iv,v},{ii,iv,v},{i,ii,iv,v}组合。

[0200] 任何以上给出的用于各(i)至(vi)部分的不同典型例子可交换地使用以形成用于连接到目标RNA序列的非-RNA多核苷酸。

[0201] 七个非-RNA多核苷酸的典型例子如下:

[0202] 构建体1:(40x SpC3)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)

[0203] (AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)

[0204] 即(40x SpC3)(SEQ ID NO:47)(SEQ ID NO:51)(4x Sp18)(SEQ ID NO:56)(SEQ ID NO:60)

[0205] 构建体2:(40x SpC3)(ACTCGCAGATCATTACGATC)(20x dT)(4x Sp18)

[0206] (AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)

[0207] 即(40x SpC3)(SEQ ID NO:47)(SEQ ID NO:52)(4x Sp18)(SEQ ID NO:56)(SEQ ID NO:60)

[0208] 构建体3:(40x rU)(ACTCGCAGATCATTACGATC)(10x dT)(4x Sp18)(AACTACTAGGATCATCGATGTATCTGCTCA)(ACTCTGAACC)

[0209] 即(40x rU)(SEQ ID NO:47)(SEQ ID NO:51)(4x Sp18)(SEQ ID NO:56)(SEQ ID NO:60)

- [0210] 构建体4: (40x SpC3) (ACTCGCAGATCATTACGATC) (10x dT) (4x Sp18)
- [0211] (AACTACTAGGATCATCGATGTATCTGCTCA) (ACTCTrGrArArCrC)
- [0212] 即 (40x SpC3) (SEQ ID NO:47) (SEQ ID NO:51) (4x Sp18) (SEQ ID NO:56) (SEQ ID NO:61)
- [0213] 构建体5: (40x SpC3) (ACTCGCAGATCATTACGATC) (10x dT) (rArArArCrUrArCrGrCrU) (AACTACTAGGATCATCGATGTATCTGCTCA)
- [0214] (ACTCTGAACC)
- [0215] 即 (40x SpC3) (SEQ ID NO:47) (SEQ ID NO:51) (SEQ ID NO:64) (SEQ ID NO:56) (SEQ ID NO:60)
- [0216] 构建体6: (40x SpC3) (ACTCGCAGATCATTACGATC) (10x dT) (4x Sp18)
- [0217] (AGCTTAACATACGATACTCTTAGCTAACCA) (ACTCTGAACC)
- [0218] 即 (40x SpC3) (SEQ ID NO:47) (SEQ ID NO:51) (4x Sp18) (SEQ ID NO:57) (SEQ ID NO:60)
- [0219] 构建体7: (40x SpC3) (具有SEQ ID NO:47的序列的PNA) (10x dT) (4x Sp18) (AACTACTAGGATCATCGATGTATCTGCTCA) (ACTCTGAACC)
- [0220] 即 (40x SpC3) (具有SEQ ID NO:47的序列的PNA) (SEQ ID NO:51) (4x Sp18) (SEQ ID NO:56) (SEQ ID NO:60)
- [0221] 连接
- [0222] 所述目标RNA多核苷酸连接到非-RNA多核苷酸和/或所述非-RNA多核苷酸连接到所述目标RNA多核苷酸,以形成修饰的RNA多核苷酸。其中所述非-RNA多核苷酸包含核糖核苷酸或RNA序列。通过包含在所述非-RNA多核苷酸中的所述核糖核苷酸或RNA序列,所述非-RNA多核苷酸可连接到所述目标RNA多核苷酸。
- [0223] 所述方法包含:在步骤(a)中提供修饰的RNA多核苷酸和DNA解旋酶。所述方法可进一步包含,在步骤(a)之前,将RNA多核苷酸连接到非-RNA多核苷酸和/或将非-RNA多核苷酸连接到RNA多核苷酸以形成修饰的RNA多核苷酸。
- [0224] 可使用任何方法的任何方式或本领域已知的方法将所述目标RNA多核苷酸和所述非-RNA多核苷酸(其可以是DNA或DNA类似物)彼此连接。优选地使用以下描述的多种方法中的一个或多个将所述目标RNA多核苷酸和所述非-RNA多核苷酸连接。所述目标RNA多核苷酸可化学地连接到所述非-RNA多核苷酸,例如通过共价键。所述目标RNA多核苷酸可通过化学或酶结合连接到所述非-RNA多核苷酸。所述目标RNA多核苷酸可通过杂交和/或合成方法连接到所述非-RNA多核苷酸。可使用拓扑异构酶将所述RNA多核苷酸连接到所述非-RNA多核苷酸。所述RNA多核苷酸可在不止一个,例如两个或三个位点连接到所述非-RNA多核苷酸。连接方法可包括一个,两个,三个,四个,五个或更多不同的连接方法。根据本发明可使用任何以下描述的连接方法的组合。
- [0225] 所述RNA多核苷酸和非-RNA多核苷酸(其可能或可能不包含核糖核苷酸或RNA)可分别制备然后再连接到一起。这两个组分可以以任何构造连接。例如,它们可以通过它们的末端(即5'或3')连接。合适的构造包括,但不限于,RNA多核苷酸的5'末端连接到非-RNA多核苷酸的3'末端和反之亦然。或者,这两组分可通过它们序列内部的核苷酸连接。
- [0226] 多核苷酸可通过它们天然存在的核苷酸连接。天然存在的核苷酸可被修饰以促进

连接。例如,天然存在的核酸可通过例如,用于mRNA帽的三甲基鸟苷合成酶来修饰。其他合适的修饰是本领域已知的。可通过取代引入修饰。所述RNA多核苷酸可通过连接体分子连接到所述非-RNA多核苷酸。所述RNA多核苷酸可使用一个或多个,例如两个或三个连接体连接到所述非-RNA多核苷酸。连接体可包含任何片段长度跨所需距离的分子。连接体在长度上可从一个碳(光气型连接体)到许多埃变化。适合用作连接体的线性分子的例子包括但不限于,聚乙二醇(PEGs),多肽,多糖,脱氧核糖核酸(DNA),肽核酸(PNA),苏糖核酸(TNA),甘油核酸(GNA),饱和的和饱和的烃类,聚酰胺。这些连接体可以是惰性的或活性的,特别地它们可在限定的位置为化学可断裂的(cleavable),或它们自身可用荧光团或配体修饰。所述连接体优选地抗二硫苏糖醇(DTT)。优选的柔性肽连接体为2至20个,例如4,6,8,10或16个丝氨酸和/或甘氨酸片段。更优选柔性连接体包括(SG)<sub>1</sub>, (SG)<sub>2</sub>, (SG)<sub>3</sub>, (SG)<sub>4</sub>, (SG)<sub>5</sub>, (SG)<sub>8</sub>, (SG)<sub>10</sub>, (SG)<sub>15</sub>或(SG)<sub>20</sub>,其中S是丝氨酸并且G是甘氨酸。优选的刚性连接体为2至30个,例如4,6,8,16或24个脯氨酸片段。更优选的刚性连接体包括(P)<sub>12</sub>,其中P是脯氨酸。

[0227] 所述RNA多核苷酸可使用一个或多个化学交联剂或一个或多个肽连接体连接到所述非-RNA多核苷酸。合适的化学交联剂是本领域众所周知的。合适的化学交联剂包括但不限于,包括以下功能性基团的化学交联剂:马来酰亚胺,活性酯,琥珀酰亚胺,叠氮化物,炔烃(例如二苯并环辛醇(DIBO或DBCO),二氟脂环烃和线性炔烃),磷化氢(例如用在无痕和非无痕施陶丁格连接中的磷化氢),卤代乙酰(例如碘乙酰胺),光气型试剂,磺酰氯试剂,异硫氰酸酯,酰基卤类,肼,二硫化物,乙烯砜类,氮杂环丙烷和光敏试剂(例如芳香叠氮化物,双氮杂环丙烷)。

[0228] 所述RNA多核苷酸和非-RNA多核苷酸之间的反应可为自发的,例如半胱氨酸/马来酰亚胺,或可能需要外部试剂,例如用于连接叠氮化物和线性炔烃的Cu(I)。

[0229] 优选的交联剂包括2,5-二氧化吡咯烷-1-基3-(吡啶-2-基二硫烷基)丙酸酯、2,5-二氧化吡咯烷-1-基4-(吡啶-2-基二硫烷基)丁酸酯和2,5-二氧化吡咯烷-1-基8-(吡啶-2-基二硫烷基)辛酸酯、二马来酰亚胺PEG 1k、二马来酰亚胺PEG 3.4k、二马来酰亚胺PEG 5k、二马来酰亚胺PEG 10k、双(马来酰亚胺基)乙烷(BMOE)、双马来酰亚胺己烷(BMH)、1,4-双马来酰亚胺丁烷(BMB)、1,4-双马来酰亚胺基-2,3-二羟基丁烷(BMDB)、BM[PEO]2(1,8-双马来酰亚胺二乙二醇)、BM[PEO]3(1,11-双马来酰亚胺三乙二醇)、三[2-马来酰亚胺氨基乙基]胺(TMEA)、DTME二硫代二马来酰亚胺乙烷、双马来酰亚胺PEG3、双马来酰亚胺PEG11、DBCO-马来酰亚胺、DBCO-PEG4-马来酰亚胺、DBCO-PEG4-NH<sub>2</sub>、DBCO-PEG4-NHS、DBCO-NHS、DBCO-PEG-DBCO 2.8kDa、DBCO-PEG-DBCO 4.0kDa、DBCO-15原子-DBCO、DBCO-26原子-DBCO、DBCO-35原子-DBCO、DBCO-PEG4-S-S-PEG3-生物素、DBCO-S-S-PEG3-生物素和DBCO-S-S-PEG11-生物素。最优的交联剂是3-(2-吡啶基二硫代)丙酸琥珀酰亚胺酯(SPDP)和马来酰亚胺-PEG(2kDa)-马来酰亚胺( $\alpha$ ,  $\omega$ -双马来酰亚胺基聚乙二醇)。

[0230] 所述连接体可被标记。合适的标签包括但不限于,荧光分子(例如Cy3或AlexaFluor®555),放射性同位素,如<sup>125</sup>I, <sup>35</sup>S,酶,抗体,抗原,多核苷酸和配体例如生物素。这种标签使得可确定连接体的量。所述标签也可是能断裂的纯化标签,例如生物素,或鉴定方法中出现的具体序列。

[0231] 通过保持连接体的溶度大量过剩于RNA多核苷酸和/或非-RNA多核苷酸可防止RNA多核苷酸或非-RNA多核苷酸自身的交联。或者,在其中使用两个连接体的情况下,可使用

“锁和钥匙”的设置。各连接体仅一个末端可一起反应以形成更长的连接体,所述各连接体的另一端与构建体(即RNA多核苷酸或非-RNA多核苷酸)的不同部分反应。

[0232] 所述RNA多核苷酸到所述非-RNA多核苷酸的连接可以是永久或稳定的(即在本发明的方法中,所述RNA多核苷酸不会从所述非-RNA多核苷酸脱离)。优选永久或稳定的连接是共价连接。

[0233] 或者,所述连接是暂时的,即所述RNA多核苷酸可从所述非-RNA多核苷酸脱离。本领域技术人员可以理解的是本文描述的任何方法都同样适合于将非-RNA多核苷酸连接至目标RNA,并且在构建所述非-RNA多核苷酸自身过程中,如上描述的所述非-RNA多核苷酸可以是两种或多种类型的核酸如DNA和RNA的杂交体。

[0234] 点击化学

[0235] 所述目标RNA多核苷酸可共价地连接到所述非-RNA多核苷酸。所述非-RNA多核苷酸可能包含或可能不包含预先结合的DNA解旋酶。在优选实施例中,可使用游离铜点击化学或铜催化的点击化学来制得所述RNA多核苷酸和非-RNA多核苷酸例如DNA的前导序列之间的共价键。由于点击化学令人满意的性质和其对于在多种构建块(building blocks)之间生成共价连接的范围,使得在这些应用中使用点击化学。例如,它是快速的,清洁的并且无毒的,只产生无害的副产物。点击化学是由Kolb等在2001首次介绍的术语,为描述更广泛的一系列强大,有选择性的和模块化的构建块,所述构建块可靠地用于小规模和大规模应用(Kolb HC, Finn, MG, Sharpless KB, 点击chemistry: diverse chemical function from a few good reactions, Angew. Chem. Int. Ed. 40 (2001) 2004-2021)。他们定义了如下一系列严格标准用于点击化学:“反应必须是模块化的,宽的范围,给出非常高的产量,只产生无害的可通过非色谱法去除的副产物,并且是立体定向的(但不必然是对映选择性)。所要求的方法特征包括简单的反应条件(理想地所述方法应对氧气和水不敏感),容易获得的起始物质和试剂,无溶剂或溶剂的使用,所述溶剂是温和的(例如水)或容易去除的,和简单的产物分离。纯化如果需要必须是通过非色谱法,例如结晶或蒸馏,并且所述产物在生理状态下必须是稳定的”。

[0236] 点击化学的合适例子包括但不限于以下:

[0237] (a) 游离铜的变体的1,3偶极环加成反应,其中叠氮化物与炔烃在应力例如在环辛烷环中反应;

[0238] (b) 在一个连接体上的氧亲核试剂与在其他连接体上的环氧化合物或氮杂环丙烷反应性部分的反应;和

[0239] (c) 施陶丁格连接,其中炔烃部分可被芳基膦取代,导致和叠氮化物的特异性反应,从而得到酰胺键。

[0240] 优选所述点击化学反应是在炔烃和叠氮化物之间Cu(I)催化的1,3偶极环加成反应。在优选实施例中,第一基团是叠氮化物基团,第二基团是炔烃基团。核酸碱基已被合成,在优选位置中插入叠氮化物和炔烃基团(例如Kocalka P, El-Sagheer AH, Brown T, Rapid and efficient DNA strand cross-linking by click chemistry, Chembiochem. 2008.9 (8):1280-5)。炔烃基团是从Berry Associates (Michigan, USA) 商业可获得的,并且叠氮化物基团是通过ATDBio或IDT bio合成的。

[0241] 如果连接体的核酸酸区域内的核苷酸被修饰以包含能形成共价键的基团,为了实



现连接,多个该修饰的核苷酸优选彼此偏移一个核苷酸。这遵循Tom Brown公布的工作(Kocalka等.(2008) ChemBiochem 9 8 1280-1285)。

[0242] 当RNA转录物形成时,可将点击反应性碱基添加到目标RNA多核苷酸。或者点击基团可通过超甲基化酶(用于给mRNA带帽)添加到目标RNA多核苷酸。

[0243] 优选地反应基团是叠氮化物 and 己基基团,例如3叠氮化物N和5'-己基-G。优选地所述叠氮化物基团连接到非-RNA多核苷酸,该非-RNA多核苷酸优选为DNA(和可能或可能不包含核糖核苷酸或RNA的序列),并且所述己基基团连接到目标RNA多核苷酸。

[0244] 实施例3和5说明了使用点击反应对将非-RNA多核苷酸(包含例如DNA)加入进目标RNA多核苷酸。实施例5进一步描述了使用桥接部分以获得2维(2D) RNA-cDNA构建体。如图15所示的实施例5的构建体可与实施例3的不包含桥接部分的构建体(如图10所示的1D构建体)比较。

[0245] 连接(ligation)

[0246] 目标RNA多核苷酸可被连接到非-RNA多核苷酸(其中RNA多核苷酸可能或可能不包含核糖核苷酸或RNA序列)。连接是两个核酸片段通常通过酶的行为或化学方法的结合。通过在一个RNA或DNA末端的3'-羟基和另一个的5'-磷酸基之间形成磷酸二酯键将RNA和DNA片段的末端结合在一起。在反应中通常包括辅因子,并且其通常为ATP或NAD<sup>+</sup>。在连接反应中可使用RNA或非-RNA多核苷酸的夹板(splint),例如DNA,PNA,甘油核酸(GNA),苏糖核酸(TNA)或锁定核酸(LNA),使得通过杂交到所述夹板来使得所述RNA多核苷酸和所述非-RNA多核苷酸保持彼此相邻从而促进连接。连接到所述RNA多核苷酸的所述非-RNA多核苷酸可能或可能不包含预先结合的DNA解旋酶。

[0247] 所述非-RNA多核苷酸可连接到所述RNA多核苷酸的任一末端,即5'或3'末端。所述非-RNA多核苷酸可连接到所述目标RNA多核苷酸的两末端。优选地所述非-RNA多核苷酸连接到所述目标RNA多核苷酸的5'末端。可使用本领域已知的任何方法将所述非-RNA多核苷酸连接到所述RNA多核苷酸。一个或多个非-RNA多核苷酸可使用连接酶进行连接,所述连接酶诸如T4DNA连接酶,大肠杆菌DNA连接酶,Taq DNA连接酶,Tma DNA连接酶,9°N DNA连接酶,T4聚合酶I,T4聚合酶2,热稳定5'App DNA/RNA连接酶,夹板R,环状连接酶,T4RNA连接酶1或T4RNA连接酶2。在缺少ATP或使用伽玛-S-ATP(ATP  $\gamma$  S)代替ATP,可将所述一个或多个非-RNA多核苷酸连接到所述RNA多核苷酸(或反之亦然)。

[0248] 所述方法优选进一步包含从所述方法条件中去除连接酶。

[0249] 实施例2说明了使用T4DNA连接酶的RNA多核苷酸到DNA的连接。

[0250] 合成方法

[0251] 寡核苷酸或引物可用于杂交到目标RNA多核苷酸的任何区域并作为DNA合成的起始点。所述寡核苷酸或引物可能或可能不包含预先结合的DNA解旋酶。

[0252] 真核RNA通常包含聚A尾,即连续的单磷酸胸苷片段。该聚A尾通常位于RNA的3'末端。如果需要的话,可使用Poly A聚合酶或末端转移酶将poly(dA)尾添加到原核RNA链的3'末端上。引物可杂交到目标RNA的polyA尾,并且所述引物可用作合成的起始点。所述引物优选包含聚T区域,即只包含基于胸腺嘧啶的核苷酸区域,或polyU区域,即只包含基于尿嘧啶的核苷酸的区域。所述polyU区域可包含UMP或dUMP。所述polyU区域可以是任意长度,例如至少10,至少15,至少20,至少25个或更多核苷酸长度。

[0253] 在本发明的一个实施例中,非-RNA多核苷酸包含具有前导序列的DNA引物,所述前导序列杂交到RNA链的poly (dA) 区域。一个或多个DNA解旋酶可预先结合到DNA引物,所述DNA引物将杂交到RNA链。或者该将要杂交到RNA链的DNA引物不包括或包含预先结合的DNA解旋酶。由DNA引物的RNA链的反转录被允许发生,并且桥接部例如DNA发夹可之后被连接到双链的DNA/RNA。这样的桥接部例如发夹环适配器,可添加到本发明任何双链目标RNA多核苷酸(RNA/RNA或RNA/DNA) 或修饰的构建体。桥接部通过接入单链中来允许模板链和互补链二者的连续测序。优选地包含适配器(例如Y适配器)的前导序列连接到RNA的一个末端,并且桥接部适配器连接到另一端。所述前导序列优先地穿入纳米孔,并且连接两条链(其可以是RNA/RNA或RNA/DNA) 的桥接部使得随着多核苷酸被拉开(unzips) 以及两条链(通过桥接部连接) 移动穿过所述孔对两条链都进行表征。这是有利的,因为它加倍了从双链多核苷酸获得的信息量。而且,因为两条链的序列是互补的,来自两条链的信息可有益地(informatively) 结合起来。这种机制提供了具有更高置信观察值的校对阅读能力。

[0254] 或者,在互补序列合成之前,将所述桥接部添加至目标RNA多核苷酸,并且所述桥接部用作互补合成的引物,其更详细讨论如下。

[0255] 在本发明的一个实施例中,使用桥接部连接双链RNA多核苷酸或RNA/DNA双链体(如RNA和cDNA) 的链。根据本发明的表征目标RNA多核苷酸的方法然后优选包含将(i) 连接的构建体,所述构建体包含所述目标RNA多核苷酸,其中所述RNA多核苷酸被修饰以包含非-RNA多核苷酸,和(ii) DNA解旋酶与跨膜孔接触,使得所述目标RNA移动穿过所述孔。所述方法优选包含随着所述目标RNA相对于所述孔移动,获取一个或多个测量值,其中所述测量值代表互补多核苷酸(RNA或cDNA) 和所述目标RNA的一个或多个特征,并由此表征目标双链多核苷酸。

[0256] 以这种方式连接并查询目标dsRNA或RNA/DNA双链体的两条链提高了表征的效率和精度。

[0257] 所述桥接部能够连接目标dsRNA多核苷酸或RNA/DNA双链体的两条链。所述桥接部通常共价地连接目标dsRNA多核苷酸或RNA/DNA双链体的两条链。所述桥接部可以为任何能够连接目标dsRNA多核苷酸或RNA/DNA双链体的两条链的物质,条件是所述桥接部不干涉RNA多核苷酸穿过跨膜孔的移动。

[0258] 通过本领域已知的任何合适的方法可将所述桥接部连接至目标多核苷酸。所述桥接部可为单独合成的并且化学连接或酶促连接至RNA目标多核苷酸。或者,在处理目标多核苷酸过程中生成所述桥接部。

[0259] 所述桥接部连接到目标多核苷酸的一个末端或接近所述末端。所述桥接部优选连接到目标多核苷酸末端的10个核苷酸内。

[0260] 合适的桥接部包括但不限于聚合物连接体,化学连接体,多核苷酸或多肽。优选地,所述桥接部包含DNA,RNA,修饰的DNA(例如无碱基DNA),RNA,PNA,LNA或PEG。所述桥接部更优选为DNA或RNA。

[0261] 所述桥接部最优选为发夹环或发夹环适配器。合适的发夹适配器可使用本领域已知的方法设计。所述发夹环可以是任何长度。所述发夹环通常为110个或更少的核苷酸长度,例如100个或更少的核苷酸,90个或更少的核苷酸,80个或更少的核苷酸,70个或更少的核苷酸,60个或更少的核苷酸,50个或更少的核苷酸,40个或更少的核苷酸,30个或更少的

核苷酸,20个或更少的核苷酸或10个或更少的核苷酸长度。所述发夹环优选为大约1至110个,2至100个,5至80个或6至50个核苷酸长度。如果环包涉及适配器的差异选择性,优选更长长度的发夹环,例如50至110个核苷酸长度。相似地,如果环不涉及选择性结合,优选更短长度的发夹环,例如从1至5个核苷酸长度。

[0262] 所述发夹适配器可连接到目标多核苷酸的任何一端,即5' 或3' 末端。可使用任何本领域已知的方法连接所述发夹适配器。所述发夹适配器可使用连接酶,例如T4DNA连接酶,大肠杆菌DNA连接酶,Taq DNA连接酶,Tma DNA连接酶和9°N DNA连接酶,进行连接。

[0263] 根据本发明在连接的构建体与孔接触之前或之后,可将互补的多核苷酸(RNA或cDNA)和目标RNA分开。可以随着多核苷酸结合蛋白例如解旋酶,或分子制动器控制的多核苷酸移动穿过孔,将它们分开。

[0264] 所述互补多核苷酸(RNA或cDNA)和所述目标RNA可使用任何本领域已知的方法分离。例如,它们可以通过多核苷酸结合蛋白或使用有助于去杂交(dehybridisation)的条件(有助于去杂交的条件的实例包括,但不限于,高温,高pH和添加能破坏氢键或碱基对的试剂,例如甲酰胺和尿素)分离。

[0265] 所述发夹适配器优选包含可选择性结合部分。这允许连接的构建体被纯化或分离。可选择性结合部分是根据其结合性质被挑选的部分。因此,可选择性结合部分优选是特异性地结合到表面的部分。如果可选择性结合部分比本发明中使用的任何其他部分更大程度地结合到表面,那么所述可选择性结合部分特异性地结合到该表面。在优选实施例中,所述部分结合到本发明使用的其他部分没有结合到的表面。

[0266] 合适的选择性结合部分是本领域已知的。优选的选择性结合部分包括但不限于,生物素,多核苷酸序列,抗体,抗体片段,例如Fab和ScSv,抗原,多核苷酸结合蛋白,多组氨酸尾和GST标签。最优选的选择性结合部分是生物素和可选择性多核苷酸序列。生物素特异性地结合到用抗生物素蛋白包覆的表面。可选择性多核苷酸序列特异性地结合(即杂交)到用同源序列包覆的表面。或者,可选择性多核苷酸序列特异性地结合到用多核苷酸结合蛋白包覆的表面。

[0267] 所述发夹适配器和/或所述可选择性结合部分可以包含能被切开,割开,劈开或水解的区域。可以设计这样的区域以允许互补多核苷酸和/或目标RNA在纯化或分离后从其结合的表面被去除。合适的区域是领域已知的。合适的区域包括但不限于,RNA区域,包含脱硫生物素和链霉亲和素的区域,二硫键和光可断裂(photocleavable)区域。

[0268] 所述连接的构建体优选地在桥接部的相对末端包含前导序列,所述桥接部例如为发夹环或发夹环适配器。

[0269] 在本发明的一个实施例中,桥接部例如发夹型寡核苷酸连接到目标RNA链。在连接所述桥接部之前,所述目标RNA链可被修饰以包含非-RNA多核苷酸。例如所述目标RNA链可使用化学连接,例如通过共价键,点击化学,化学连接或酶促连接,使用杂交和/或合成方法连接到非-RNA多核苷酸。所述RNA多核苷酸可使用拓扑异构酶连接到所述非-RNA多核苷酸。或者所述桥接部可在修饰RNA链以包含非-RNA多核苷酸之前连接到所述目标RNA链。相似地,所述桥接部,例如发夹型寡核苷酸,可通过本文所述的任何连接方法连接到所述目标RNA链。优选地所述桥接部通过连接(ligation)结合到所述目标RNA链。上文所述的任何合适的连接酶可用于桥接部到所述目标RNA链的连接,即T4DNA连接酶,大肠杆菌DNA连接酶,

Taq DNA连接酶, Tma DNA连接酶, 9°N DNA连接酶, T4聚合酶I, T4聚合酶2, 热稳定的5' App DNA/RNA连接酶, 夹板R, 环形连接酶, T4RNA连接酶1或T4RNA连接酶2。

[0270] 优选地所述桥接部连接到目标RNA链的3'或5'末端, 最优选所述桥接部连接到所述目标RNA链的3'末端。从桥接部例如发夹型寡核苷酸的反转录, 导致通过发夹连接的RNA-cDNA构建体的形成。在这一实施例中, 所述桥接部充当引物用于反转录: 所述桥接部, 例如发夹型寡核苷酸, 自身用作引物用于反转录以生成RNA-cDNA构建体并进行2D读取。优选反转录在所述桥接部的3'末端开始。在所述桥接部的poly-T突出杂交到RNA的poly A-尾。所述RNA的poly A-尾通常为50至300个核苷酸长度。优选在桥接部的poly-T突出含有少于100个的核苷酸, 例如1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 90, 91, 92, 93, 94, 95, 96, 97, 98或99个核苷酸。优选地所述poly T突出位于所述桥接部的3'末端上。在本发明的这一实施例中, 其中桥接部连接到目标RNA, (目标RNA已被修饰或有待被修饰以包含非-RNA多核苷酸; 优选为DNA多核苷酸) 并且其中所述桥接部充当引物用于反转录, 产生了DNA/RNA/cDNA构建体 (参见实施例5和图15)。这种构建体允许2D读取, 其中由于所述桥接部的存在, 所述目标RNA和其互补DNA序列 (通过反转录形成) 都可通过跨膜孔测序。这种方法是有利的, 因为它加倍了从单个双链的目标多核苷酸构建体获得的信息量。而且, 因为在互补(cDNA)链中的序列是必定正交于模板RNA链的序列, 来自两条链的信息可有益地结合起来。因此, 这种机制提供了更高置信观察的正交校对阅读能力。

[0271] 进一步, 本发明的方法的另外主要优势是:

[0272] 1) 丢失的核苷酸的覆盖: 所述方法本质上使任何丢失的核苷酸或核苷酸的基团 (即由于随RNA链滑过孔的移动问题) 的问题最小化, 因为在一条链上任何可能丢失的区域都很可能通过从其互补区域获得的正交信息进行涵盖。

[0273] 2) 不确定的序列基序的覆盖: 通过互补链中的正交或相反信息来涵盖任何有困难的序列基序, 该互补链具有不同序列不会有相同的依赖于序列的问题。例如, 这与只在电流中产生小变化的序列基序或具有相似电流水平的序列基序特别相关, 即连续的碱基基序, 当其移动穿过纳米孔产生相同电流块 (电流block) 时, 由于在电流中没有阶跃变化, 其因此不能观察到。来自一条序列基序的任何相似电流水平将会被从其互补链的正交序列获得的完全不同电流水平涵盖。

[0274] 除以上讨论的优势以外, 这也有许多具体例子, 其可利用读取双链多核苷酸的全部两条链的概念来提供进一步益处:

#### [0275] 1. 表观遗传信息

[0276] 在广泛的应用中, 能够鉴定天然RNA链中的表观遗传信息 (例如5-甲基胞嘧啶或5-羟甲基胞嘧啶核苷酸) 或受损碱基是令人满意的。使用本发明的方法, 获得这一信息不必化学处理或扩增, 所述化学处理或扩增都会引入误差。在纳米孔测序期间, 随着核苷酸穿过纳米孔检测电流水平的变化。所述电流水平被数个碱基支配而不是被单个核苷酸支配。当修饰的核苷酸穿过纳米孔时, 它影响了几个连续水平: 这增加了我们在检测中的置信度 (参见图18)。

[0277] 纳米孔测序也是单分子测序技术,因此其可在不需要扩增的情况下被实施。已显示纳米孔可检测对标准的4个RNA核苷酸的修饰。读取多核苷酸的双链可对检测在以下情况中的RNA修饰有用,所述情况为修饰的碱基以与其他碱基相似的方式(产生相似的电流信号)表现。例如如果甲基胞嘧啶(mC)以与尿嘧啶相似的方式表现,这会有与将mC分配给U相关的错误。在模板链中,这会有碱基被称为mC或U的可能性。然而,在互补链中,相应的碱基可能有很高可能性表现为G。因此通过“校正阅读”(“proof reading”)互补链,很有可能在模板链中的碱基是mC而不是U。

[0278] 读取模板和互补链可以在不需要扩增的情况下实施。然而,可增加扩增或复制作为样品制备的一部分,从而帮助表观遗传信息的检测。

[0279] 在测序之前的任何阶段可实施包含目标RNA多核苷酸的连接链的分离和复制,并根据需要进行多次。例如,在分离如以上所述的第一RNA/cDNA多核苷酸构建体的两条连接链之后,可生成最终单链多核苷酸的互补链,从而形成另一双链多核苷酸。这双链多核苷酸的两条链可之后使用桥接部连接起来,从而形成第二构建体。这在本文中被称作“DUO”方法。这种构建体可随后用于本发明。在这样的实施例中,最终构建体中的双链多核苷酸的一条链包含通过桥接部连接的原始目标双链多核苷酸(在第一构建体中)的两条链。原始目标双链多核苷酸或互补链的序列可被估算或测定。可按需要重复多次复制过程,并且由于目标多核苷酸实际上被多次读取,提供了额外的校正阅读。

[0280] 可构建核苷酸链,其中以以下顺序穿过纳米孔读取以下信息:模板RNA(原始),互补cDNA(原始),-桥接部-,模板RNA(互补),互补cDNA(互补)。

[0281] 在这方案中,将四次获得甲基化碱基上的信息。如果表观遗传碱基在原始模板链(在此例中是mC)中,将以高可能性获得以下信息:模板(原始)-mC,互补(原始)-G,模板(互补)-C,和互补(互补)-G。清楚的是原始模板读取和复制模板读取将给出不同结果,而两互补读取会产生相同的碱基响应。此信息可被用于显示原始模板链中的表观遗传碱基的位置。

## [0282] 2. 均聚物读取

[0283] 对于单分子纳米孔测序,均聚物读取可能是问题。如果均聚物区域比孔的读取部分更长,所述均聚物部分的长度将很难确定。

[0284] 为了克服均聚物读取的问题,可合成cDNA链,其中该cDNA链添加了结合原始dTTP, dGTP, dATP, dCTP的不同/修饰碱基。这可以是天然碱基类似物例如肌苷(I)。所述碱基与正确的天然碱基相比有随机的插入机会,并且插入率可通过变化三磷酸物种的浓度来控制。

[0285] 通过添加替换的碱基,将有将替换的碱基插入均聚物区域的反向互补链的可能性。这一结果是所述均聚物移动(homopolymer run)在长度上将减小至一点,该点可以被纳米孔的读取部分读取。例如,AAAAAAAAAAAAA (SEQ ID NO:33)的均聚物基团将会具有替换的碱基的随机插入,并产生TTTITTTIITTTI (SEQ ID NO:34) (其中I是肌苷)。

[0286] 所述均聚物片段被减小以允许各核苷酸或核苷酸的基团被估计或确定。模板链将是天然RNA链,而互补/cDNA链将包含天然碱基和碱基类似物的混合物。从模板和互补链读取的数据的结合可用于估计在原始RNA部分中均聚物移动(homopolymer run)的长度。

## [0287] 拓扑异构酶

[0288] 拓扑异构酶连接到单链DNA或连接到双链DNA,并切除该DNA的磷酸骨架。此中间物

破坏使得DNA为解缠绕的或未卷绕的,并且在此过程的最后,DNA骨架再次封装。合适的拓扑异构酶结合策略将在图6和7中说明。

#### [0289] 其他方法

[0290] 将RNA目标多核苷酸连接到非-RNA多核苷酸的替换性方法包含利用真核RNA的5'末端,所述真核RNA的5'末端通过在其相反方向添加7-甲基鸟苷帽来进行修饰。所述真核RNA的5'末端因此含有或包含反向碱基(reversed bases),即在RNA链的5'末端,各碱基具有其自由的3'末端,与5'末端相对。所述目标RNA多核苷酸可结合至非-RNA多核苷酸(其可能或可能不包含核糖核苷酸或RNA序列),所述非-RNA多核苷酸也具有含有反向碱基的区域,从而其在相反方向移动,正如图12所示,其中所述非-RNA多核苷酸是DNA。或者非-RNA多核苷酸可具有在其一端具有反向碱基的RNA片段。这反向RNA碱基的片段可通过连接结合至真核RNA的5'末端的反向碱基。

[0291] 所述非-RNA多核苷酸可能或可能不包含预先结合的DNA解旋酶。

#### [0292] 真核RNA

[0293] 在真核细胞中,核中产生的主要RNA转录物在其运输至细胞质之前以一些方式进行处理,在所述细胞质中编制翻译机制。首先,在转录(加帽)期间,添加由7-甲基鸟苷残基组成的帽,所述帽通过三磷酸键连接至转录物的5'末端。加帽包含5'-5'三磷酸连接。通过翻译机制认出所述帽,并且所述帽保护RNA链的生长以防由于核酸酶的分解。之后胸苷残基片段被添加至3'末端(聚腺苷酸化)。这些聚A尾为150至200个残基长度。在这些修饰后,RNA剪接去除插入序列(即内含子)。

[0294] 在本发明的一个实施例中,5'帽保留在原处,至少一个化学性反应基团添加至所述目标RNA多核苷酸的5'末端。在本发明的方法中可使用任何化学连接方法,只要a)化学反应不损害RNA或DNA的结构,和b)产生的反应性连接不是如此庞大以至于DNA解旋酶不能沿其移动或穿过它。可使用超甲基化酶将至少一个反应基团连接至所述目标RNA多核苷酸。在本发明的一个实施例中,添加至所述RNA多核苷酸的至少一个反应基团是点击反应基团,虽然其不是必要的。添加至所述RNA多核苷酸的至少一个反应基团可替换的是任何合适的反应基团例如巯基(thiol)。至少一个反应基团也添加至非-RNA多核苷酸(其可能包含或可能不包含核糖核苷酸或RNA序列)的末端。优选地所述至少一个反应基团连接至所述非-RNA多核苷酸的3'末端。优选地所述非-RNA多核苷酸是DNA链。可使用超甲基化酶添加至少一个反应基团至所述非-RNA多核苷酸。在本发明的一个实施例中,添加至所述非-RNA多核苷酸的至少一个反应基团是点击反应基团,虽然其不是必要的。然后各RNA多核苷酸和非-RNA多核苷酸上的一个或多个反应基团在合适的条件下进行接触以形成共价键。ATP gamma S和酶可用于将硫代磷酸盐添加到DNA,然后所述DNA可连接至连接有马来酰亚胺的RNA。

[0295] 在本发明的选择性实施例中,去除了7-甲基鸟苷帽,优选地使用烟草酸性焦磷酸酶来形成去帽的RNA链。然后可以多种不同方法处理去帽的目标RNA多核苷酸以制备连接有非-RNA多核苷酸的RNA多核苷酸的链。在一个实施例中,至少一个反应基团可添加至去帽的RNA多核苷酸的5'末端。可使用超甲基化酶将至少一个反应基团添加到所述RNA多核苷酸。在本发明的一个实施例中,添加至所述RNA多核苷酸的至少一个反应基团是点击反应基团,虽然其不是必要的。添加至所述RNA多核苷酸的反应基团可替换的是任何合适的反应基团例如巯基。至少一个反应基团也添加至非-RNA多核苷酸的末端。所述至少一个反应基团连

接至所述非-RNA多核苷酸的3'末端。优选地所述非-RNA多核苷酸是DNA链。可使用超甲基化酶添加至少一个反应基团至所述非-RNA多核苷酸。在本发明的一个实施例中,添加至所述非-RNA多核苷酸的至少一个反应基团是点击反应基团,虽然其不是必要的。然后各RNA多核苷酸和非-RNA多核苷酸上的一个或多个反应基团在合适的条件下进行接触以形成共价键。

[0296] 或者,使用连接酶,例如T4DNA连接酶,大肠杆菌DNA连接酶,Taq DNA连接酶,Tma DNA连接酶,9°N DNA连接酶,T4聚合酶I,T4聚合酶2,热稳定的5'App DNA/RNA连接酶,夹板R,环连接酶,T4RNA连接酶1,T4RNA连接酶2可将非-RNA多核苷酸的链直接结合到目标RNA多核苷酸上。优选地所述非-RNA多核苷酸是DNA链。在本发明的一个实施例中,一种或多种酶可预先结合到要与RNA连接的非-RNA多核苷酸上。在所述非-RNA多核苷酸上预先加载的酶加速了样品制备过程,意味着使用更少的管。或者,将要与所述RNA多核苷酸结合的非-RNA多核苷酸不包含或包含预先结合的酶。

[0297] 在本发明的替换性实施例中,所述非-RNA多核苷酸是具有前导序列的DNA引物,所述前导序列杂交到RNA多核苷酸。一种或多种酶可预先结合到DNA引物,所述DNA引物将杂交到所述RNA多核苷酸。或者杂交到RNA多核苷酸的DNA引物不包含或包含预先结合的酶。根据使用的转录酶,从DNA引物的RNA多核苷酸的反转录导致了1至3个C的3'突出。然后DNA发夹可结合至双链的DNA/RNA。允许杂交的条件是本领域已知的(例如,Sambrook等,2001, *Molecular Cloning: a laboratory manual*, 3rd edition, Cold Spring Harbour Laboratory Press; and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel等, Eds., Greene Publishing and Wiley-Interscience, New York (1995))。可在低严格条件下例如在37℃的30至35%甲酰胺, 1M NaCl和1% SDS (十二烷基硫酸钠) 缓冲溶液存在下实施杂交,然后在50℃进行从1X (0.1650M Na<sup>+</sup>) 至2X (0.33M Na<sup>+</sup>) SSC (标准的柠檬酸钠) 的冲洗。可在中等严格条件下例如37℃的40至45%甲酰胺, 1M NaCl和1% SDS缓冲溶液存在下实施杂交,然后在55℃进行从0.5X (0.0825M Na<sup>+</sup>) 至1X (0.1650M Na<sup>+</sup>) SSC的冲洗。可在高严格条件下例如37℃的50%甲酰胺, 1M NaCl, 1% SDS缓冲溶液存在下实施杂交,然后在60℃, 0.1X (0.0165M Na<sup>+</sup>) SSC中进行冲洗。特别地,所述条件优选为10uM寡聚体在10mM Tris-HCl, 50mM NaCl, pH 7中,加热至98℃然后以每分钟冷却2℃冷却到18℃。

[0298] 真核RNA通常包含聚A尾,即连续的单磷酸胸苷片段。所述聚A尾通常在RNA的3'末端。在这样的实施例中,可将引物杂交到目标RNA的聚A尾。所述引物优选包含polyU区域,即包含仅基于尿嘧啶的核苷酸的区域。所述polyU区域可包含UMP或dUMP。所述polyU区域可以是任何长度,例如至少10,至少15,至少20,至少25或更多个。或者所述引物包含聚T区域。所述聚T区域可以是任何长度,例如至少10,至少15,至少20,至少25或更多个。将真核RNA的链连接到DNA的链的方法如图1所示。

[0299] 在本发明替换的实施例中,可使用基于拓扑异构酶的策略将非-RNA多核苷酸连接到所述目标RNA多核苷酸(Cheng和Shuman, 2000, *Nucleic Acids Research*, Vol. 28, No. 9, 1893-1898, 通过并入的方式引入本文)。DNA拓扑异构酶结合到双链DNA,并在特定的目标位点切割一条链的磷酸二酯骨架。所述双链DNA的另一条链在其序列上相应的位置包含缺口或破口。一旦拓扑异构酶结合并切割所述DNA的上部的链,所述酶保持与dsDNA结合。结合到dsDNA的所述拓扑异构酶可将DNA链转移到具有5' OH末端的RNA链,从而形成串联的DNA-RNA共聚物,如果它之后与具有自由5' 羟基的RNA一起孵育。所述目标RNA必须具有自由



5'羟基:真核RNA必须去帽以得到具有自由5'羟基的RNA,而microRNA具有自由5'羟基基团(参见图6)。

[0300] 在替换的基于拓扑异构酶的策略中,使用DNA的单链区域将所述dsDNA连接到RNA的链,从而与RNA杂交并帮助RNA连接(Sekiguchi等,1997,The Journal of Biological Chemistry,Vol 272,No.25,15721-15728,通过并入引入本文)。所述拓扑异构酶结合到dsDNA的特定目标序列上。DNA下部的链在其中没有缺口。一旦结合拓扑异构酶,其只切割DNA的上部的链,并保持与dsDNA/ssDNA杂交体结合。结合到dsDNA的所述拓扑异构酶,然后与具有自由5'羟基的RNA孵育,并将RNA连接至dsDNA。所述ssDNA区域帮助将互补的RNA序列吸引至连接DNA的点(参见图7)。

[0301] 将RNA多核苷酸连接至非-RNA多核苷酸的替换性方法包括利用真核RNA的5'末端,其通过添加在相反方向上移动的7-甲基鸟苷帽来进行修饰。真核RNA的5'末端因此包括或包含反向鸟嘌呤碱基,即其在RNA链的5'末端,但该碱基具有自己的自由3'末端,与5'末端相对。所述目标RNA可连接到也具有反向碱基区域的DNA(或RNA)序列,从而其在相反方向移动(参见图12)。

[0302] 原核RNA

[0303] 可利用Poly A聚合酶将poly(dA)尾添加到RNA多核苷酸的3'末端上。在本发明的一个实施例中,所述非-RNA多核苷酸是具有前导序列的DNA引物,所述前导序列杂交到RNA多核苷酸。具有前导序列的DNA引物可杂交到poly(dA)区域。一个或多个酶可预先结合到DNA引物,所述DNA引物将杂交到所述RNA多核苷酸。或者要杂交到RNA多核苷酸的DNA引物不包含或包含预先结合的酶。根据使用的反转录酶,从DNA引物的RNA多核苷酸的反转录导致了1至3个C的3'突出。然后DNA发夹可结合至双链的DNA/RNA。允许杂交的条件是本领域已知的(3至4个用于上标II(MMLV)(参见p1192Biotechniques Vol 29No 6(2000)。DNA发夹可然后结合双链的DNA/RNA。

[0304] 在本发明的替换性实施例中,poly(dA)区域不添加到所述RNA多核苷酸,并且在反应步骤中直接使用原核RNA多核苷酸。可将反应基团添加到RNA多核苷酸的5'或3'末端。优选地至少一个反应基团被添加到RNA多核苷酸的5'末端。添加至所述RNA多核苷酸的至少一个反应基团是点击反应基团,虽然其不是必要的。添加至所述RNA多核苷酸的至少一个反应基团可替换的是任何合适的反应基团例如巯基。至少一个反应基团也添加至非-RNA多核苷酸的末端。优选地所述至少一个反应基团连接至所述非-RNA多核苷酸的3'末端。优选地所述非-RNA多核苷酸包含DNA链。可使用超甲基化酶添加至少一个反应基团至所述非-RNA多核苷酸。添加至所述非-RNA多核苷酸的至少一个反应基团是点击反应基团,虽然其不是必要的。然后各RNA和非-RNA多核苷酸上的一个或多个反应基团在合适的条件下进行接触以形成共价键。

[0305] 在本发明的替换性实施例中,使用连接酶,例如T4DNA连接酶,大肠杆菌DNA连接酶,Taq DNA连接酶,Tma DNA连接酶,9°N DNA连接酶,T4聚合酶I,T4聚合酶2,热稳定的5' App DNA/RNA连接酶,夹板R,环连接酶,T4RNA连接酶1和T4RNA连接酶2可将非-RNA多核苷酸的链直接连接到目标RNA多核苷酸上。优选地所述非-RNA多核苷酸包含DNA链。在本发明的一个实施例中,一个或多个酶可预先结合到非-RNA多核苷酸,所述非-RNA多核苷酸将与RNA多核苷酸结合。或者将要与所述RNA多核苷酸结合的非-RNA多核苷酸不包含或包含预先结



合的酶。将原核RNA多核苷酸连接到非-RNA(即DNA)多核苷酸的方法在图2中示出。

[0306] 在本发明的替换性实施例中,可使用基于拓扑异构酶的策略将非-RNA多核苷酸连接到所述目标RNA多核苷酸(Cheng和Shuman,2000,Nucleic Acids Research,Vol.28, No.9:1893-1898,通过并入的方式引入本文)。DNA拓扑异构酶结合到双链DNA,并在特定的目标位点切割一条链的磷酸二酯骨架。所述双链DNA的另一条链在其序列上相应的位置包含缺口或破口。一旦拓扑异构酶结合并切割所述DNA的上部的链,所述酶保持与dsDNA结合。结合到dsDNA的所述拓扑异构酶可将DNA链转移到具有5'OH末端的RNA链,从而形成串联的DNA-RNA共聚物,如果它之后与具有自由5'羟基的RNA一起孵育。所述目标RNA必须具有自由5'羟基。在替换的基于拓扑异构酶的策略中,使用DNA的单链区域将所述dsDNA连接到RNA的链,从而与RNA杂交并帮助RNA连接(Sekiguchi等,1997,The Journal of Biological Chemistry,Vol 272,No.25,15721-15728,通过并入引入本文)。所述拓扑异构酶结合到dsDNA的特定目标序列上。DNA下部的链在其中没有缺口。一旦结合拓扑异构酶,其只切割DNA的上部的链,并保持与dsDNA/ssDNA杂交体结合。结合到dsDNA的所述拓扑异构酶,然后与具有自由5'羟基的RNA孵育,并将RNA连接至dsDNA。所述ssDNA区域帮助将互补的RNA序列吸引至连接DNA的点(参见图7)。

[0307] 无扩增

[0308] 本发明的方法中,所述目标RNA多核苷酸通常不扩增。所述方法通常不包含制备多个所述目标RNA的副本。

[0309] 所述方法优选不包含聚合酶链反应(PCR)或反转录PCR(RT-PCR)。

[0310] 正如以上讨论,在本发明的一个实施例中,所述RNA多核苷酸包含DNA互补链,其通过使用DNA引物和反转录酶制备。使用桥接部将RNA和DNA链连接能够进行2D读取。

[0311] 在本发明的替换性实施例中,发夹型寡核苷酸其自身用作引物,来用于反转录从而制备RNA-cDNA构建体,并能够进行2维读取(即RNA和cDNA链)。

[0312] DNA解旋酶和分子制动器

[0313] DNA解旋酶被用于控制RNA多核苷酸穿过孔的移动。所述DNA解旋酶不需要展示酶活性,只要其能够结合目标RNA多核苷酸并控制所述目标RNA多核苷酸穿过孔的移动。例如,所述酶可被修饰以去除其酶活性或可在阻止其发挥酶功能的条件下使用。

[0314] 所述解旋酶可以是或衍生自Hel308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DNA解旋酶可源自Hel308Mbu(SEQ ID NO:8), Hel308Csy(SEQ ID NO:9),Hel308Tga(SEQ ID NO:10),Hel308Mhu(SEQ ID NO:11),TraI Eco(SEQ ID NO:12),XPD Mbu(SEQ ID NO:13),Dda 1993(SEQ ID NO:14)或其变体。

[0315] 所述解旋酶可以是国际申请No.PCT/GB2012/052579(公布为W0 2013/057495); PCT/GB2012/053274(公布为W0 2013/098562);PCT/GB2012/053273(公布为W02013098561);PCT/GB2013/051925(公布为W0 2014/013260);PCT/GB2013/051924(公布为W0 2014/013259);PCT/GB2013/051928(公布为W0 2014/013262)和PCT/GB2014/052736中公开的任何解旋酶,修饰的解旋酶或解旋酶构建体。

[0316] 解旋酶可以两种模式相对于孔工作。第一,优选以以下方式使用DNA解旋酶实施本方法,使得所述DNA解旋酶顺着由施压电压产生的场控制RNA多核苷酸穿过孔的移动。在此模式中,所述RNA多核苷酸的5'末端首先被捕获在孔中,然后所述酶控制所述RNA多核苷酸

移进所述孔,使得所述RNA多核苷酸顺着所述场穿过所述孔直到它最终移位穿到膜的反侧。或者,所述方法优选以以下方式实施,使得所述DNA解旋酶逆着由施压电压产生的场,控制所述RNA多核苷酸穿过孔的移动。在此模式中,所述RNA多核苷酸的3'末端首先首先被捕获在孔中,然后所述酶控制所述RNA多核苷酸移动穿过所述孔,使得所述RNA多核苷酸逆着施加的场被拉出所述孔直到最终返回到所述膜的顺侧。

[0317] 所述DNA解旋酶优选包含如SEQ ID NO:15 (TrwC Cba)所示的序列或其变体,所述序列如SEQ ID NO:8 (He1308Mbu)所示的序列或其变体,或所述序列如SEQ ID NO:14 (Dda)所示的序列或其变体。变体可以以下针对跨膜孔讨论的任意方式不同于天然序列。SEQ ID NO:14的优选变体包含 (a) E94C和A360C或 (b) E94C,A360C,C109A和C136A,然后可选地包含 ( $\Delta$  M1) G1 (即M1缺失并之后添加G1)。

[0318] 根据本发明可使用任何数量的解旋酶。例如,可使用1,2,3,4,5,6,7,8,9,10或更多个解旋酶。在一些实施例中,可使用不同数量的解旋酶。在本发明的一个实施例中,所述DNA解旋酶预先结合到非-RNA多核苷酸。

[0319] 本发明的方法优选包含将多核苷酸与两个或多个解旋酶接触。所述两个或多个解旋酶通常是相同的解旋酶。所述两个或多个解旋酶可为不同的解旋酶。

[0320] 所述两个或多个解旋酶可为以上提及的解旋酶的任何组合。所述两个或多个解旋酶可为两个或多个的Dda解旋酶。所述两个或多个解旋酶可为一个或多个的Dda解旋酶和一个或多个的TrwC解旋酶。所述两个或多个解旋酶可为相同解旋酶的不同变体。

[0321] 所述两个或多个解旋酶优选彼此连接。所述两个或多个解旋酶更优选共价地彼此连接。所述解旋酶可以任意次序或使用任何方法连接。用于本发明的优选的解旋酶构建体描述在国际申请No. PCT/GB2013/051925 (公开为WO 2014/013260); PCT/GB2013/051924 (公开为WO 2014/013259); PCT/GB2013/051928 (公开为WO 2014/013262) 和PCT/GB2014/052736中。

[0322] SEQ ID NO:8,9,10,11,12,13,14或15的变体是具有以下氨基酸序列的酶,所述氨基酸由SEQ ID NO:8,9,10,11,12,13,14或15的氨基酸序列变化而来并保持多核苷酸结合能力的酶。这可以使用本领域已知的任何方法测定。例如所述变体可与多核苷酸接触,并且它与多核苷酸结合的能力和沿着所述多核苷酸移动的能力可以被测定。所述变体可包括促进多核苷酸结合和/或提高它的高盐浓度和/或室温下活性的能力的修饰。变体可被修饰,使其结合多核苷酸(即保持多核苷酸结合能力),但不作为解旋酶发挥功能(即当其被提供促进移动的所有必要成分即ATP和 $Mg^{2+}$ 时,不能沿着多核苷酸移动)。这些修饰是本领域已知的。例如,解旋酶中的 $Mg^{2+}$ 结合域的修饰通常导致不发挥解旋酶功能的变体。这种类型的变体可表现为分子制动器(参见以下)。

[0323] 基于氨基酸的相似性或同一性,对于SEQ ID NO:8,9,10,11,12,13,14或15的氨基酸序列的整个长度,变体优选与该序列具有至少50%的同源性。更优选,基于氨基酸的相似性或同一性,对于SEQ ID NO:8,9,10,11,12,13,14或15的氨基酸序列的整个长度,变体多肽与该序列具有至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%和更优选至少95%,97%或99%的同源性。在200或更多,例如230,250,270,280,300,400,500,600,700,800,900或1000或更多的连续氨基酸长度上,可以具有至少80%,例如至少85%,90%或95%的氨基酸相似性或同一性(“严格同源性”)。同源性如上所述进行

确定。所述变体可以按上述针对SEQ ID NO:2和4所描述的多种方式中的任意一种不同于野生型序列。酶可以共价地连接到孔。可以使用任何方法将酶共价连接到孔。

[0324] 在优选实施例中,所述方法包含:

[0325] (a) 提供具有一个或多个DNA解旋酶的RNA多核苷酸和一个或多个分子制动器,其中所述RNA多核苷酸被修饰以包含非-RNA多核苷酸并增强了DNA解旋酶对其的结合;

[0326] (b) 将所述RNA多核苷酸和跨膜孔接触,并提供跨所述孔的电势,使得所述一个或多个DNA解旋酶和一个或多个分子制动器连接在一起并控制所述RNA多核苷酸穿过所述孔的移动;

[0327] (c) 随着所述RNA多核苷酸相对于所述孔移动,获取一个或多个测量值,其中所述测量值代表所述RNA多核苷酸的一个或多个特征,并由此表征所述目标RNA多核苷酸。

[0328] 此类方法更详细地在英国申请No.1406151.9中讨论。优选分子制动器是TrwC Cba-Q594A(具有突变Q594A的SEQ ID NO:15)。此变体不作为解旋酶发挥功能(即与多核苷酸结合,但当其被提供有促进移动的所有必要成分即ATP和 $Mg^{2+}$ 时,不沿着多核苷酸移动)。

[0329] 所述一个或多个解旋酶可以是任何上述的那些解旋酶。所述一个或多个分子制动器可以是结合到RNA多核苷酸并减缓所述RNA多核苷酸穿过所述孔的任何化合物或分子。所述一个或多个分子制动器优选包括结合到所述RNA多核苷酸的一个或多个化合物。所述一个或多个化合物优选是一个或多个大环化合物。合适的大环化合物包括,但不限于,环糊精、杯芳烃、环肽、冠醚、瓜环、柱芳烃、其衍生物或其组合。环糊精或其衍生物可以是任何在Eliseev, A.V., 与Schneider, H-J. (1994) J. Am. Chem. Soc. 116, 6081-6088中公开的那些。所述的试剂更优选是七-6-氨基- $\beta$ -环糊精( $am_7$ -BCD)、6-单脱氧-6-单氨基- $\beta$ -环糊精( $am_1$ - $\beta$ CD)或七-(6-脱氧-6-胍基)-环糊精( $gu_7$ -BCD)。

[0330] 所述一个或多个分子制动器优选不是一个或多个单链结合蛋白(SSB)。所述一个或多个分子制动器更优选不是包含不具有净负电荷的羧基末端(C-末端)区域的单链结合蛋白(SSB),或(ii)C-末端区域中包含能减少C-末端区域的净负电荷的一个或多个修饰的经修饰SSB。所述一个或多个分子制动器最优选不是在国际申请No.PCT/GB2013/051924(公开为W02014/013259)中公开的任何SSB。

[0331] 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RNA多核苷酸并控制其相对于孔的移動的任何蛋白。现有技术中可直接确定蛋白质是否结合到多核苷酸。所述蛋白通常与多核苷酸相互作用并修饰所述多核苷酸的至少一个特性。所述蛋白可通过切开多核苷酸以形成单个核苷酸或较短链的核苷酸,如二核苷酸或三核苷酸,来修饰多核苷酸。所述部分可以通过将多核苷酸定向或通过移动多核苷酸到特定的位置,即控制多核苷酸的移动,来修饰多核苷酸。

[0332] 所述多核苷酸结合蛋白优选衍生自多核苷酸处理酶。所述一个或多个分子制动器可以衍生自上述的任何多核苷酸处理酶。充当分子制动器的Phi29聚合酶(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。

[0333] 任何数量的衍生自解旋酶的分子制动器可以被使用。例如,1、2、3、4、5、6、7、8、9、10或更多个解旋酶可被用作分子制动器。如果两个或更多个解旋酶被用作分子制动器,所述两个或更多个解旋酶通常是相同的解旋酶。所述两个或更多个解旋酶可以是不同的解旋酶。

[0334] 所述两个或更多个解旋酶可以是上文提到的解旋酶的任意组合。所述两个或更多个解旋酶可以是两个或更多个Dda解旋酶。所述两个或更多个解旋酶可以是一个或多个Dda解旋酶和一个或多个TrwC解旋酶。所述两个或更多个解旋酶可以是相同解旋酶的不同变体。

[0335] 所述两个或更多个解螺旋酶优选彼此连接。所述两个或更多个解旋酶更优选共价地彼此连接。所述解旋酶可以以任何顺序和使用任何方法连接。从解旋酶衍生的一个或多个分子制动器优选被修饰以减少多核苷酸结合域中开口的大小,所述多核苷酸可在至少一个构象状态下通过所述开口从解旋酶解绑。这在W02014/013260中公开。

[0336] 用于本发明的优选的解旋酶构建体在国际申请No. PCT/GB2013/051925 (公开为W02014/013260); PCT/GB2013/051924 (公开为W0 2014/013259) 和PCT/GB2013/051928 (公开为W0 2014/013262); 和2013年10月18日提交的英国专利申请No. 1318464.3中所描述的。

[0337] 如果所述一个或多个解旋酶以活性模式使用(即当为一个或多个解旋酶提供有助于移动的所有必要组分,例如ATP和 $Mg^{2+}$ 时),所述一个或多个分子制动器优选(a)在非活性模式下使用(即在不存在有助于移动的必需组分或没有能力主动移动时), (b) 在活性模式下使用,其中所述一个或多个分子制动器以与所述一个或多个解旋酶相反的方向移动,或(c) 在活性模式下使用,其中所述一个或多个分子制动器以与所述一个或多个解旋酶相同的方向移动,且比所述一个或多个解旋酶移动得更慢。

[0338] 如果在非活性模式下使用所述一个或多个解旋酶(即当未为所述一个或多个解旋酶提供有助于移动的所有必要组分,例如ATP和 $Mg^{2+}$ 或不能主动移动时),所述一个或多个分子制动器优选(a) 在非活性模式下使用(即在不存在有助于移动所需的组分时或不能主动移动时)或(b) 在活性模式下使用,其中所述一个或多个分子制动器以与多核苷酸相对于孔的相同方向沿多核苷酸移动穿过所述孔。

[0339] 所述一个或多个解旋酶和一个或多个分子制动器可以在任何位置被连接到RNA,以使它们聚集在一起并且这两者控制RNA穿过孔的移动。所述一个或多个解旋酶和一个或多个分子制动器至少间隔一个核苷酸,例如间隔至少5个、至少10个、至少50个、至少100个、至少500个、至少1000、至少5000、至少10,000、至少50,000个核苷酸或更多个核苷酸。如果该方法涉及表征具有在一端设置的Y适配器和在另一端设置发夹环适配器的双链RNA多核苷酸,则所述一个或多个解旋酶优选连接到Y适配器,而所述一个或多个分子制动器优选连接到所述发夹环适配器。在本实施方案中,所述一个或多个分子制动器优选是经修饰的一个或多个解旋酶,该修饰使得它们结合RNA多核苷酸但不发挥解旋酶的功能。连接到Y适配器的一个或多个解旋酶如下文中详述的优选被停滞在间隔器处。连接到发夹环适配器的一个或多个分子制动器优选不停滞在间隔器处。当所述一个或多个解旋酶到达所述发夹环,所述一个或多个解旋酶和一个或多个分子制动器优选聚集在一起。所述一个或多个解旋酶可以在Y适配器被连接到多核苷酸之前或Y适配器被连接到多核苷酸之后,连接到Y适配器。一个或多个分子制动器可以在所述发夹环适配器被连接到多核苷酸之前或所述发夹环适配器被连接到多核苷酸之后,连接到所述发夹环适配器。

[0340] 所述一个或多个解旋酶和一个或多个分子制动器优选不彼此连接。所述一个或多个解旋酶和一个或多个分子制动器更优选不共价地彼此连接。所述一个或多个解旋酶和一个或多个分子制动器优选不连接,如在国际申请No. PCT/GB2013/051925 (公开为W0 2014/

013260);PCT/GB2013/051924(公开为W0 2014/013259)和PCT/GB2013/051928(公开为W02014/013262);和2013年10月18日提交的英国专利申请No.1318464.3中所描述的。

[0341] 间隔器

[0342] 在本发明的构建体中包括一个或多个间隔器。当部分的RNA多核苷酸进入所述孔并沿着由所施加的电势而产生的场移动穿过所述孔时,随着所述RNA多核苷酸移动穿过所述孔,所述一个或多个解旋酶被所述孔移动穿过间隔器。这是因为,所述RNA多核苷酸(包括一个或多个间隔器)移动穿过所述孔,并且所述一个或多个解旋酶保留在所述孔的顶部。正如国际申请No.PCT/GB2014/050175(公开为W0 2014/135838)中讨论的,一个或多个DNA解旋酶可停滞(stall)在一个或多个间隔器处。在该国际申请中公开的一个或多个解旋酶和一个或多个间隔器的任何构造均可以在本发明中使用。

[0343] 所述一个或多个间隔器可以是部分的目标RNA多核苷酸,例如它/它们中断所述多核苷酸序列。所述一个或多个间隔器优选不是杂交到所述目标RNA的一个或多个阻断分子如减速障(speed bump)的一部分。所述一个或多个间隔器可以是部分的非-RNA多核苷酸(即DNA多核苷酸),例如它/它们中断所述多核苷酸序列。所述一个或多个间隔器可以是部分的RNA多核苷酸。所述一个或多个间隔器可连接至目标RNA多核苷酸和/或非-RNA多核苷酸。所述一个或多个间隔器可设置在所述RNA多核苷酸或非-RNA多核苷酸的末端,和/或所述一个或多个间隔器可设置在所述RNA多核苷酸或非-RNA多核苷酸中。

[0344] 在目标RNA多核苷酸或非-RNA多核苷酸中可以有任何数量的间隔器,如1、2、3、4、5、6、7、8、9、10或更多个间隔器。在本发明的构建体中优选有2个、4个或6个间隔器。在构建体的不同区域中可以有一个或多个间隔器,例如在前导序列中的间隔器和在发夹环中的间隔器。

[0345] 所述一个或多个间隔器各自提供了使一个或多个解旋酶即使在活性模式下也无法克服的能量屏障。所述一个或多个间隔器可通过减弱解旋酶的牵引(例如将目标RNA多核苷酸或非-RNA多核苷酸中的核苷酸去除碱基)或物理地阻断所述一个或多个解旋酶的移动(例如使用庞大的化学基团)而使所述一个或多个解旋酶停滞(stall)。

[0346] 所述一个或多个间隔器可包含使一个或多个解旋酶停滞的任何分子或分子组合。所述一个或多个间隔器可包含防止所述一个或多个解旋酶沿着目标RNA多核苷酸移动的任何分子或分子组合。这可以在不存在跨膜孔和施加的电势的情况下,直接确定所述一个或多个解旋酶是否被停滞在一个或多个间隔器处。例如,可以使用PAGE测定解旋酶移动经过间隔器的能力。

[0347] 所述一个或多个间隔器通常包含线性分子,如聚合物。所述一个或多个间隔器通常具有与目标RNA多核苷酸或非-RNA多核苷酸不同的结构。例如,所述一个或多个间隔器通常不是RNA。特别地,所述一个或多个间隔器优选包含肽核酸(PNA)、甘油核酸(GNA)、苏糖核酸(TNA)、锁核酸(LNA)或具有核苷酸侧链的合成聚合物。所述一个或多个间隔器可以沿与所述多核苷酸相反的方向包含一个或多个核苷酸。例如,当多核苷酸在5'到3'的方向时,所述一个或多个间隔器可在3'至5'方向包含一个或多个核苷酸。核苷酸可以是上文论述的任何核苷酸。

[0348] 所述一个或多个间隔器优选包含一个或多个硝基吡啶,例如一个或多个5-硝基吡啶、一个或多个肌苷、一个或多个嘧啶、一个或多个2-氨基嘌呤、一个或多个2-6-二氨基嘌呤

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

[0349] 所述一个或多个间隔器可包含任何数量的这些基团。例如,对于2-氨基嘌呤、2-6-二氨基嘌呤、5-溴-脱氧尿苷、倒立dT、ddT、ddC、5-甲基胞苷、5-羟甲基胞苷、2'-O-甲基RNA碱基、异-dC、异-dG、iSpC3基团、PC基团、己二醇基团和硫醇的连接,所述一个或多个间隔器优选包含2、3、4、5、6、7、8、9、10、11、12或更多个。所述一个或多个间隔器优选包含2、3、4、5、6、7、8个或更多个iSp9基团。所述一个或多个间隔器优选包含2、3、4、5或6个或更多个iSp18基团。最优选的间隔器为四个iSpC3基团。

[0350] 所述聚合物优选为多肽或聚乙二醇 (PEG)。所述多肽优选地包含2、3、4、5、6、7、8、9、10、11、12或更多个氨基酸。所述PEG优选包含2、3、4、5、6、7、8、9、10、11、12或更多个单体单元。

[0351] 所述一个或多个间隔器优选包含一个或多个无碱基核苷酸 (即缺乏核碱基的核苷酸),如2、3、4、5、6、7、8、9、10、11、12或更多个无碱基核苷酸。所述核碱基可以被无碱基核苷酸中的-H (idSP) 或-OH替代。无碱基间隔器可以通过从一个或多个相邻的核苷酸中去除核碱基而被插入到目标多核苷酸中。例如,多核苷酸可以被修饰为包括3-甲基腺嘌呤、7-甲基鸟嘌呤、1,N 6-亚乙烯基腺嘌呤肌苷或次黄嘌呤,并且所述核碱基可使用人类烷基腺嘌呤DNA转葡萄糖基酶 (hAAG) 从这些核苷酸中去除。或者,多核苷酸可以被修饰为包括尿嘧啶和用尿嘧啶-DNA转葡萄糖基酶 (UDG) 去除的核碱基。在一个实施方案中,所述一个或多个间隔器不包含任何无碱基核苷酸。

[0352] 所述一个或多个DNA解旋酶可通过每个线性分子间隔器停滞 (即停滞在其前面) 或在每个线性分子间隔器上停滞。如果使用线性分子间隔器,构建体优选提供有与一个或多个解旋酶待移动穿过的每个间隔器的末端相邻的多核苷酸的双链区。所述双链区通常有助于将一个或多个解旋酶停滞到相邻的间隔器上。如果该方法是在约100mM或更低的盐浓度下进行的,则存在双链区是特别优选的。每个双链区通常为至少10个,如至少12个核苷酸的长度。如果本发明中使用的目标多核苷酸是单链,则双链区可以通过将较短的多核苷酸杂合到与间隔器邻近的区域而形成。所述较短的多核苷酸通常由与目标多核苷酸相同的核苷酸形成,但也可以由不同的核苷酸形成。例如,所述较短的多核苷酸可由LNA形成。

[0353] 如果使用线性分子间隔器,构建体优选提供有阻塞分子,该阻塞分子在每个间隔器的与所述一个或多个解旋酶待移动穿过的末端相对的末端。这有助于确保所述一个或多个解旋酶保持停滞在每个间隔器上。在它/它们在溶液中扩散的情况下,也有助于在构建体上保持一个或多个解旋酶。阻塞分子可以是下文所述的能物理地引起一个或多个解旋酶停滞的任何化学基团。阻塞分子可以是多核苷酸的双链区。

[0354] 所述一个或多个间隔器优选包含物理地引起一个或多个解旋酶停滞的一个或多个化学基团。所述一个或多个化学基团优选为一个或多个侧挂的化学基团。所述一个或多

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

[0355] 目标多核苷酸中的不同间隔器可以包含不同的停滞分子。例如,一个间隔器可以包含上述的线性分子中的一个,而另一个间隔器可以包含物理地引起一个或多个解旋酶停滞的一个或多个化学基团。间隔器可以包含任何上述的线性分子和一个或多个物理地引起一个或多个解旋酶停滞的化学基团,如一个或多个无碱基体和荧光团。

[0356] 合适的间隔器可以根据目标多核苷酸的类型和本发明的方法实施的条件进行设计。大多数的解旋酶结合DNA并沿DNA移动,且由此可以使用不为DNA的任何物质使其停滞。合适的分子如上所述。

[0357] 本发明的方法优选在游离核苷酸存在下和/或在解旋酶辅因子存在下进行。这在下文更详细地讨论。在跨膜孔和施加电势不存在时,所述一个或多个间隔器优选能够在游离核苷酸存在下和/或在解旋酶辅因子存在下停滞所述一个或多个解旋酶。

[0358] 如果本发明的方法如下文所述在游离核苷酸和解旋酶辅因子存在下进行(使得一个或更多的解旋酶是在活性模式下),则通常使用一个或多个较长间隔器来确保所述一个或多个解旋酶是在它们与跨膜孔接触前且电势施加前被停滞在多核苷酸上。在不存在游离核苷酸和解旋酶辅因子情况下(使得所述一个或多个解旋酶是在非活性模式下),可以使用一个或多个较短间隔器。

[0359] 所述盐浓度也影响所述一个或多个间隔器停滞一个或多个解旋酶的能力。在跨膜孔和施加的电势不存在时,所述一个或多个间隔器优选能够在约100mM或更低的盐浓度下停滞所述一个或多个解旋酶。本发明的方法中使用的盐浓度越高,通常使用的所述一个或多个间隔器越短,反之亦然。

[0360] 特征的优选组合在下表1中示出。

[0361] 表1:

| [0362] | 多核苷酸 | 间隔器组成* | 间隔器长度<br>(即其的数量*) | 盐[ ]         | 游离核苷酸? | 解旋酶辅因子? |
|--------|------|--------|-------------------|--------------|--------|---------|
|        | DNA  | iSpC3  | 4                 | 1 M          | 有      | 有       |
|        | DNA  | iSp18  | 4                 | 100-1000 mM  | 有      | 有       |
|        | DNA  | iSp18  | 6                 | <100-1000 mM | 有      | 有       |
|        | DNA  | iSp18  | 2                 | 1 M          | 有      | 有       |
|        | DNA  | iSpC3  | 12                | <100-1000 mM | 有      | 有       |
|        | DNA  | iSpC3  | 20                | <100-1000 mM | 有      | 有       |
|        | DNA  | iSp9   | 6                 | 100-1000 mM  | 有      | 有       |
| [0363] | DNA  | idSp   | 4                 | 1 M          | 有      | 有       |

[0364] 所述方法可以涉及,移动两个或更多个解旋酶穿过间隔器。在这种情况下,间隔器的长度通常是增加的以防止在不存在孔和施加的电势下尾部解旋酶推动前导解旋酶穿过间隔器。如果该方法涉及移动两个或更多个解旋酶穿过一个或多个间隔器,则上文所述的

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

#### [0365] 跨膜孔

[0366] 跨膜孔是在一定程度上穿过膜的结构。其允许由施加的电势驱动的水合离子流过膜或在膜内流动。跨膜孔通常穿过整个膜,使得水合离子可从膜的一侧流到膜的另一侧。然而,跨膜孔不是必须穿过膜。它可以是一端闭合的。例如,所述孔可以是位于水合离子沿其流动或流入其中的膜中的孔道、间隙、通道、沟槽或狭缝。

[0367] 任何跨膜孔可以在本发明中使用。所述孔可以是生物的或人工的。合适的孔包括但不限于,蛋白孔、多核苷酸孔和固态孔。所述孔可以是DNA折纸(origami)孔(Langecker et al., Science, 2012; 338: 932-936)。

[0368] 跨膜孔优选为跨膜蛋白质孔。跨膜蛋白质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,跨膜蛋白质孔能够形成孔,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脂质双分子层的一侧流动到另一侧。跨膜蛋白质孔允许多核苷酸或核酸,例如DNA或RNA,移动穿过所述孔。

[0369] 跨膜蛋白质孔可以是单体或低聚物。所述孔优选地由几个重复的亚基,例如至少6个、至少7、至少8或至少9个亚基组成。所述孔优选是六聚体的、七聚体的、八聚体的或九聚体的孔。

[0370] 跨膜蛋白质孔通常包含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。

[0371] 跨膜蛋白孔的桶状体或通道通常包含有助于与分析物如核苷酸、多核苷酸或核酸相互作用的氨基酸。这些氨基酸优选定位于靠近桶状体或通道的缢痕(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。

[0372] 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, CsgG, 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Msp或 $\alpha$ -溶血素( $\alpha$ -HL)。

[0373] 跨膜蛋白孔优选衍生自Msp, 优选MspA。该孔为低聚的且通常包含衍生自Msp的7个, 8个, 9个或10个单体。所述孔可以是衍生自含相同单体的Msp的同源寡聚孔。或者, 所述孔可以是含至少一个与其他单体不同的单体的Msp衍生的异源寡聚孔。优选地, 所述孔衍生自MspA或其同源物或并系同源物(paralog)。

[0374] 衍生自Msp的单体通常包含SEQ ID NO:2所示的序列或其变体。SEQ ID NO:2为MspA单体的MS-(B1)8突变体。其包括以下突变:D90N, D91N, D93N, D118R, D134R和E139K。SEQ



ID NO:2的变体是多肽,所述多肽具有从SEQ ID NO:2的氨基酸序列变化而来且保留了其形成孔的能力的氨基酸序列。可以使用本领域已知的任何方法测定变体形成孔的能力。例如,变体可以连同其他合适的亚基被插入到两亲层中,且可以确定其寡聚形成孔的能力。将亚基插入膜例如两亲膜中的方法是本领域已知的。例如,亚基可以以纯化的形式悬浮在含有脂质双分子层的溶液中,从而使其扩散到脂质双分子层并通过结合到脂质双分子层而插入膜中,并组装成功能状态。或者,可以使用如M.A.Holden,H.Bayley.J.Am.Chem.Soc.2005,127,6502-6503和国际申请No.PCT/GB2006/001057(公开为W0 2006/100484)中所描述的“拾取和放置”方法将亚基直接插入膜中。

[0375] 对于SEQ ID NO:2的氨基酸序列的整个长度,基于氨基酸相似性或同一性,变体优选与该序列具有至少50%的同源性。更优选地,基于氨基酸相似性或同一性,变体可与SEQ ID NO:2的氨基酸序列的整个序列具有至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,更优选至少95%,97%或99%的同源性。在100或更多,例如125,150,175或200或更多的连续氨基酸长度上,具有至少为80%,例如至少85%,90%或95%的氨基酸同一性(“严格同源性”)。

[0376] 可使用本领域的标准方法确定同源性。例如,UWCGC软件包提供BESTFIT程序,该程序可以用来计算同源性,例如用于其默认设置(Devereux et al(1984)Nucleic Acids Research 12,p387-395)。PILEUP和BLAST算法可以用来计算同源性或比对序列(例如鉴定等价残基或相应的序列(通常在它们的默认设置下)),如Altschul S.F.(1993)J Mol Evol 36:290-300;Altschul,S.F et al(1990)J Mol Biol 215:403-10中所描述的。公众可通过国家生物技术信息中心获得用于进行BLAST分析的软件(<http://www.ncbi.nlm.nih.gov/>)。

[0377] 相似性可使用成对同一性来测定或通过应用打分矩阵例如BLOSUM62并转为等值同一性来测定。由于它们代表功能变化而非进化变化,当确定同源性时,故意突变位置将被遮掩。通过在蛋白质序列的综合数据库上使用,例如PSIBLAST的位置特异性得分矩阵的应用,可以更敏感地确定相似性。可使用不同的打分矩阵,其在整个进化时间刻度上(即电荷)反映氨基酸化学-物理性质而不是替换的频率。

[0378] SEQ ID NO:2是MspA单体的MS-(B1)8突变体。该变体与MspA比较,可包含MspB,MspC或MspD单体中的任何突变。MspB,MspC和MspD的成熟形式在SEQ ID NO:5至SEQ ID NO:7中示出。特别地,所述变体可以包含存在于MspB中的下列取代:A138P。所述变体可以包含存在于MspC中的下列取代中的一个或多个:A96G,N102E和A138P。所述变体可包含存在于MspD中的下列突变中的一个或多个:G1缺失,L2V,E5Q,L8V,D13G,W21A,D22E,K47T,I49H,I68V,D91G,A96Q,N102D,S103T,V104I,S136K和G141A。所述变体可以包含来自Msp B,Msp C和Msp D的突变体和取代中的一个或多个的组合。所述变体优选包含突变L88N。SEQ ID NO:2的变体除了具有MS-(B1)8的所有突变外还具有突变L88N,并被称为MS-(B2)8。在本发明中使用的孔优选为MS-(B2)8。进一步优选的变体包括突变G75S/G77S/L88N/Q126R。SEQ ID NO:2的变体除了具有MS-(B1)8的所有突变外,还具有突变G75S/G77S/L88N/Q126R,并被称为MS-(B2C)8。在本发明使用的孔优选MS-(B2)8或MS-(B2C)8。

[0379] 衍生自Msp的单体可以使用本领域中已知的标准方法来制备。衍生自Msp的单体可以通过合成或重组方式制成。例如,所述孔可通过体外翻译和转录(IVTT)合成。制造孔的合

适的方法在国际申请No.PCT/GB09/001690 (公开为W0 2010/004273), PCT/GB09/001679 (公开为W0 2010/004265) 或PCT/GB10/000133 (公开为W0 2010/086603) 中描述。本文也论述了将孔插入膜的方法。

[0380] 跨膜蛋白质孔还优选衍生自 $\alpha$ -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,其为七聚体)。 $\alpha$ -溶血素-NN的一个单体或亚基的序列在SEQ ID NO:4中示出。所述跨膜蛋白质孔优选地包含七个单体,各单体包含如SEQ ID NO:4或其变体所示的序列。SEQ ID NO:4的氨基酸1,7至21,31至34,45至51,63至66,72,92至97,104至111,124至136,149至153,160至164,173至206,210至213,217,218,223至228,236至242,262至265,272至274,287至290和294形成了环区域。SEQ ID NO:4的残基113和147形成了部分收缩的 $\alpha$ -HL的桶状体或通道。

[0381] 在此实施例中,本发明的方法优选使用包含七个蛋白质或单体的孔,所述七个蛋白质或单体各自包含如SEQ ID NO:4或其变体所示的序列。该七个蛋白质可为相同的(同源七聚体)或不同的(异源七聚体)。

[0382] 变体可包含能促进与解旋酶或构建体的共价连接或相互作用的修饰。所述变体优选包含一个或多个能促进与解旋酶或构建体连接的反应性半胱氨酸残基。例如,变体可在SEQ ID NO:4的位置8,9,17,18,19,44,45,50,51,237,239和287中一个或多个和/或氨基或羧基末端包含半胱氨酸。优选的变体包含用半胱氨酸取代在SEQ ID NO:4的位置8,9,17,237,239和287的残基(A8C,T9C,N17C,K237C,S239C或E287C)。所述变体优选是任何一个国际专利No.PCT/GB09/001690 (公布为W0 2010/004273), PCT/GB09/001679 (公布为W0 2010/004265) 或PCT/GB10/000133 (公布为W0 2010/086603) 中描述的变体。

[0383] 变体也可包含能促进任何与核苷酸的相互作用的修饰。

[0384] 变体可以是天然存在的变体,其通过生物体,例如通过金黄色葡萄菌(*Staphylococcus bacterium*)天然地表达。或者,变体可在体外表达或通过细菌例如大肠杆菌重组表达。变体也包含通过重组技术制备的非天然存在的变体。基于氨基酸同一性,对于SEQ ID NO:4的氨基酸序列的整个长度,变体优选与该序列具有至少50%的同源性。更优选地,基于氨基酸同一性,对于SEQ ID NO:4的氨基酸序列的整个长度,变体多肽与该序列具有至少55%,至少60%,至少65%,至少70%,至少75%,至少80%,至少85%,至少90%和更优选至少95%,97%或99%的同源性。在200或更多,例如230,250,270或280或更多的连续氨基酸长度上,可以具有至少80%,例如至少85%,90%或95%的氨基酸同一性(“严格同源性”)。同源性如上所述进行确定。

[0385] 除了以上讨论的以外,可以对SEQ ID NO:4的氨基酸序列进行取代,例如可进行多达1,2,3,4,5,10,20或30个取代。可进行保守性取代。

[0386] SEQ ID NO:4的氨基酸序列的一个或多个氨基酸残基可额外地从以上描述的多肽中缺失。可以缺失多达1,2,3,4,5,10,20或30个残基或更多。

[0387] 变体可以是SEQ ID NO:4的片段。这种片段保留了形成孔的活性。片段至少为50,100,200或250个氨基酸长度。片段优选包含SEQ ID NO:4的形成孔的域。片段通常包括SEQ ID NO:4的残基119,121,135,113和139。

[0388] 一个或多个氨基酸可替换地或额外地添加到以上描述的多肽。在SEQ ID NO:4或其变体或其片段的氨基酸序列的氨基末端或羧基末端可提供延长。此延长可以非常短,例

如1至10个氨基酸长度。或者,此延长可以更长,例如多达到50或100个氨基酸长度。载体蛋白可以融合至孔或变体。

[0389] 正如以上讨论,SEQ ID NO:4是具有从SEQ ID NO:4的氨基酸序列变化而来且保留了其形成孔的能力的氨基酸序列的变体。变体通常包含SEQ ID NO:4的负责孔形成的区域。包含 $\beta$ -桶状体的 $\alpha$ -HL的孔形成能力是由在各亚基中的 $\beta$ -链提供的。SEQ ID NO:4的变体通常包含SEQ ID NO:4中的形成 $\beta$ -链的区域。SEQ ID NO:4的形成 $\beta$ -链的氨基酸如上讨论。可对SEQ ID NO:4的形成 $\beta$ -链的所述区域进行一个或多个修饰,只要最终的变体保留其形成孔的能力。以上讨论了可对SEQ ID NO:4的 $\beta$ -链区域进行的特定修饰。

[0390] SEQ ID NO:4的变体优选包含在 $\alpha$ -螺旋体和/或环区域中的一个或多个修饰,例如取代,添加或缺失。形成 $\alpha$ -螺旋体和/或环的氨基酸如上讨论。

[0391] 可以如上所述对变体进行修饰以帮助其鉴定或提纯。

[0392] 可以如上所述参考源自Msp的孔来制备源自 $\alpha$ -HL的孔。

[0393] 膜

[0394] 根据本发明可使用任何膜。合适的膜是现有技术中已知的。该膜优选为两性分子层。两性分子层是一种由具有至少一个亲水性部分和至少一个亲脂性或疏水性部分的两性分子诸如磷脂质形成的层。两性分子可以是合成的或天然存在的。形成单分子层的非天然存在的两亲物质和两亲物质是本领域已知的,并且包括例如嵌段共聚物(Gonzalez-Perez et al., Langmuir, 2009, 25, 10447-10450)。嵌段共聚物是其中两个或多个单体亚单元聚合在一起而形成单一聚合物链的聚合物材料。嵌段共聚物具有的性质通常由每个单体亚单元贡献。然而,嵌段共聚物可具有从各亚单元形成的聚合物不具备的独特性质。嵌段共聚物可以被改造,使得该单体亚单元之一是疏水性的(即亲脂性),而其它亚单元在水介质中是亲水性的。在这种情况下,该嵌段共聚物可具有两亲特性,并且可以形成模仿生物膜的结构。嵌段共聚物可以是二嵌段共聚物(由两个单体亚单元组成),但也可以由两个以上的单体亚单元构成,以形成相当于两亲物质的更复杂的结构。所述共聚物可以为三嵌段,四嵌段或五嵌段共聚物。

[0395] 两性分子层通常为单层或双层。两性分子层通常为平面脂质双分子层或支撑双层。

[0396] 两性分子层通常是脂质双分子层。脂质双分子层为细胞膜的模型,且作为一系列实验研究的优秀的平台。例如,脂质双分子层可通过单通道记录器用于膜蛋白的体外研究。或者,脂质双分子层可作为生物传感器来检测一系列物质的存在。脂质双分子层可以为任何脂质双分子层。合适的脂质双分子层包括但不限于,平面脂质双分子层,支撑双层或脂质体。脂质双分子层优选为平面脂质双分子层。合适的脂质双分子层在国际申请No. PCT/GB08/000563(公布为WO 2008/102121),国际申请No. PCT/GB08/004127(公布为WO 2009/077734)和国际申请No. PCT/GB2006/001057(公布为WO 2006/100484)中公开。

[0397] 用于形成脂质双分子层的方法是本领域已知的。在实施例中公开了合适的方法。脂质双分子层通常由Montal和Mueller的方法形成(Proc. Natl. Acad. Sci. USA., 1972; 69: 3561-3566),其中的脂质单分子层在穿过孔的任一侧负载在水溶液/空气界面上,所述孔垂直于所述界面。

[0398] Montal和Mueller的方法非常受欢迎,因为其成本经济并是一种相对迅速的形

具有良好质量的适合蛋白质孔嵌入的脂质双分子层的方法。形成脂质双分子层的其它常规方法包括尖浸渍 (tip dipping), 涂覆双分子层 (painting bilayers) 和膜片钳。

[0399] 在优选的实施例中, 如国际申请No. PCT/GB08/004127 (公开为W0 2009/077734) 中所描述形成脂质双分子层。

[0400] 在另一个优选的实施方案中, 膜是固态层。固态层不是生物来源的。换句话说, 固态层不是由生物环境获得或分离出的, 所述生物环境例如生物体或细胞, 或生物学上可获得结构的合成制造形式。固态层可以由有机材料和无机材料形成, 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  $\text{SiO}_2$ , 有机和无机聚合物如聚酰胺, 塑料如特氟隆 (Teflon®) 或柔性体如双组分加成固化硅橡胶, 以及玻璃。固态层可由单原子层例如石墨烯, 或只有几个原子厚度的层形成。合适的石墨层在国际申请No. PCT/US2008/010637 (公开为W0 2009/035647) 中公开。

[0401] 实施该方法通常利用 (i) 包含孔的人工两性分子层, (ii) 分离的, 天然存在的包含孔的脂质双分子层, 或 (iii) 含有孔嵌入其中的细胞。该方法通常使用人工两性分子层, 如人工脂质双分子层实施。除了孔, 该分子层可包含其他跨膜和/或膜内蛋白以及其他分子。合适的设备和条件在下面讨论。本发明的方法通常在体外实施。

[0402] 偶联

[0403] 目标RNA多核苷酸优选偶联到包含跨膜孔的膜。这可以使用任何已知方法实施。所述方法可包含将所述目标RNA多核苷酸偶联到包含跨膜孔的膜。所述RNA多核苷酸优选使用一个或多个锚偶联到膜。所述RNA多核苷酸可使用任何已知方法偶联到膜。

[0404] 每个锚包含偶联 (或结合) 到RNA多核苷酸的基团和偶联 (或结合) 到膜的基团。每个锚可共价地偶联 (或结合) 到RNA多核苷酸和/或膜。如果使用Y适配器和/或发夹环适配器, RNA优选使用适配器偶联到膜。

[0405] 可使用任何数量的锚, 例如2, 3, 4或更多个锚将RNA多核苷酸偶联到膜。例如, 可使用两个锚将RNA多核苷酸偶联到膜, 其中各个锚独立地偶联 (或结合) 到RNA多核苷酸和膜。

[0406] 一个或多个锚可包含如上讨论的一个或多个DNA解旋酶和/或一个或多个分子制动器。

[0407] 如果膜是两性分子层, 例如脂质双分子层 (详细讨论如上所示), RNA优选通过膜中存在的多肽或膜中存在的疏水性锚偶联到膜。所述疏水性锚优选是脂质, 脂肪酸, 甾醇, 碳纳米管或氨基酸。

[0408] RNA多核苷酸可直接地偶联到膜。可使用国际申请No. PCT/GB2012/051191 (公开为W02012/164270) 中描述的任何方法将RNA多核苷酸偶联到膜。RNA多核苷酸优选通过连接体偶联到膜。优选连接体包括但不限于, 聚合物, 例如多核苷酸, 聚乙二醇 (PEGs) 和多肽。如果RNA直接地偶联到膜, 然后由于膜和孔和/或多核苷酸结合蛋白之间的距离, 则当表征运行不能继续至RNA的末端时, 一些数据将丢失。如果使用连接体, 则RNA可以进行到完成为止。如果使用连接体, 连接体可连接到RNA的任何位置。连接体通常连接到RNA的尾聚合物。

[0409] 偶联可以是稳定的或暂时的。对于某些应用, 偶联的暂时性质是优选的。如果稳定的偶联分子直接连接到RNA任一5' 或3' 末端, 然后由于膜和孔和/或多核苷酸结合蛋白之间的距离, 则当表征运行不能继续至互补多核苷酸的末端, 一些序列数据将丢失。如果偶联是暂时的, 那么当偶联的末端随机地变得没有膜时, 所述RNA多核苷酸可以被加工至完成为

止。与膜形成稳定或短暂连接的化学基团在下面更详细地讨论。RNA多核苷酸可使用胆固醇或脂肪酰链暂时地偶联到两性分子层,例如脂质双分子层。可以使用具有6至30个碳原子的长度的任何脂肪酰链,例如十六烷酸。

[0410] 多核苷酸到合成脂质双分子层的偶联之前已经使用多种不同的系链策略进行实施。这些总结于下表2中。

[0411] 表2

| [0412] | 连接基团 | 偶联类型 | 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  | 暂时的  | Pfeiffer, I. and F. Hook (2004). "Bivalent cholesterol-based coupling of oligonucleotides nucleotides to lipid membrane assemblies." <i>J Am Chem Soc</i> <b>126</b> (33): 10224-5                           |
|        | 脂质   | 稳定的  | van Lengerich, B., R. J. Rawle, et al. "Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." <i>Langmuir</i> <b>26</b> (11): 8666-72 |

[0413] 合成的多核苷酸可在合成反应中使用修饰的亚磷酰胺进行官能化,所述经修饰的亚磷酰胺与用于直接添加合适的锚定基团,如硫醇、胆固醇、脂质和生物素基团是容易兼容的。这些不同的连接化学物给出了用于连接多核苷酸的一套选择。每个不同的修饰基团以略微不同的方式偶联多核苷酸,并且偶联并不总是永久的,所以给予了连接到膜的多核苷酸不同的停留时间(dwelling time)。暂时的偶联的优点如上所述。

[0414] RNA多核苷酸的偶联还可以通过许多其他手段实现,条件是反应基团被添加到RNA多核苷酸。

[0415] 或者,反应基团也可被认为是RNA多核苷酸中的短区域,其与已和膜偶联的短区域互补,这样通过杂交可以实现连接。该区域可为部分的RNA多核苷酸或与连接到RNA多核苷酸的区域。使用T4RNA连接酶I连接ssDNA的短片段已经被报道(Troutt, A.B., M.G.McHeyzer-Williams等人(1992). "Ligation-anchored PCR: a simple amplification technique with single-sided specificity (连接-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)。

[0416] 最优选地,使用胆固醇-标记的多核苷酸将RNA偶联到膜,所述胆固醇-标记的多核苷酸与RNA多核苷酸或连接至此的非-RNA多核苷酸杂交。

[0417] 诊断或预测疾病或状况 (conditions)

[0418] 本发明优选使用mRNA来诊断或预测疾病或状况。一些疾病或状况与mRNA的改变量(或水平)相关。mRNA可为正常或野生型mRNA,即不是交替拼接的。有疾病或状况的mRNA的量

(或水平)与没有疾病或状况的mRNA的量(或水平)相比可以增加或减少。通过使用本发明的方法测定来自病人的样品中的mRNA的量以诊断或预测这样的疾病或状况。

[0419] 许多遗传疾病或状况是通过由交替mRNA拼接,例如mRNA拼接缺陷引起的突变导致的。许多疾病或状况与并不归因为明显突变的交替mRNA拼接相关。交替拼的存在或不存在可通过使用本发明的方法测定来自病人的样品中的交替拼接的mRNA的存在或不存在来确定。在某些情况下,交替mRNA拼接可以是正常功能的细胞。在此情况下,交替拼接的mRNA的量(或水平)的增加或减少与正常量(即没有疾病或状况的病入的量)相比可用来诊断或预测疾病或状况。

[0420] 本发明提供了诊断或预测与病人的信使RNA(mRNA)的改变量和/或交替拼接相关的疾病或状况的方法。本发明提供了确定病人是否具有疾病或状况发展风险或处于疾病或状况发展风险中的方法,疾病或状况与信使RNA(mRNA)的改变量和/或交替拼接相关。每种情况中,所述方法包含使用本发明的方法来测定来自病人的样品中的mRNA的量和/或同一性。所述疾病或状况可为任何以下讨论的疾病或状况。所述疾病或状况优选是囊胞性纤维症,家族性自主神经功能异常,额颞叶痴呆大叶性,肌萎缩性脊髓侧索硬化症,早年衰老症候群,中链酰CoA脱氢酶(MCAD)缺乏症,肌强直性营养不良,普瑞德威利症候群,脊髓性肌萎缩,滔蛋白病变,高胆固醇血症或癌症。这些疾病,它们的起因或可能的治疗在Tazi et al. (Biochimica et Biophysica Acta(BBA)-Molecular Basis of Disease,Volume 1792, Issue 1,January 2009,Pages14-26)中讨论。

[0421] 来自病人的样品中的mRNA的改变(即增加的或减少的)量(或水平)的存在通常诊断预测疾病或状况,即表明病人具有疾病或状况发展风险或处于疾病或状况发展风险中。不存在来自病人的样品中的mRNA的改变(即增加的或减少的)量(或水平)通常表明病人不具有疾病或状况发展风险或不处于疾病或状况发展风险中。mRNA的量可如上所述进行确定。

[0422] 来自病人的样品中的交替拼接mRNA的存在通常诊断或预测疾病或状况,即表明病人具有疾病或状况发展风险或处于疾病或状况发展风险。来自病人的样品中不存在交替拼接mRNA通常表明病人不具有疾病或状况发展风险或不处于疾病或状况发展风险。交替拼接mRNA的存在或不存在可以如上所述通过鉴定样品中的RNA来确定。

[0423] 来自病人的样品中的交替拼接mRNA的增加量或降低量(或水平)通常诊断或预测疾病或状况,即表明病人具有疾病或状况发展风险或处于疾病或状况发展风险中。来自病人的样品中的交替拼接mRNA的量的无变化(与没有疾病或状况的病入的量或水平对比)通常表明病人不具有疾病或状况发展风险或不处于疾病或状况发展风险中。交替拼接mRNA的量可如上所述进行确定。

[0424] 本发明中优选使用miRNA来诊断或预测疾病或状况。本发明提供了诊断或预测与miRNA相关联的疾病或状况的方法。本发明提供了确定是否病人具有或处于与miRNA相关联的疾病或状况发展风险的方法。该方法包括使用本发明的方法确定来自病人的样品中的miRNA的存在或不存在。所述疾病或状况可为任何以下讨论的疾病或状况。

[0425] 来自病人的样品中的miRNA的存在通常表明病人具有疾病或状况发展风险或处于疾病或状况发展风险中。来自病人的样品中不存在miRNA通常表明病人不具有疾病或状况发展风险或不处于疾病或状况发展风险中。miRNA的存在或不存在可以如上所述通过鉴

定样品中的任何miRNAs来确定。

[0426] 疾病或状况优选是癌症,冠心病,心血管疾病或败血症。疾病或状况更优选是腹主动脉瘤,急性淋巴细胞白血病(ALL),急性髓性白血病(AML),急性心肌梗塞,急性早幼粒细胞白血病(APL),腺瘤,肾上腺皮质癌,酒精肝,阿兹海默症,未分化型甲状腺癌(ATC),焦虑症,哮喘,星形细胞瘤,过敏性皮炎,自闭症谱系障碍(ASD),B细胞慢性骨髓细胞性白血病,B细胞淋巴瘤,贝克肌营养不良(BMD),膀胱癌,脑肿瘤,乳腺癌,伯基特淋巴瘤,心脏肥厚,心肌症,心血管疾病,小脑的神经退化,宫颈癌,胆管癌,胆脂瘤,绒毛膜癌,慢性淋巴细胞性白血病,慢性骨髓白血病,慢性胰腺炎,结肠癌,大肠癌,先天性心脏病,冠状动脉疾病,多发性错构瘤综合征,皮炎(DM),糖尿病性肾病,腹泻型肠易激综合征,弥漫性大B细胞淋巴瘤,扩张型心肌病,唐氏综合症(DS),杜氏肌营养不良(DMD),子宫内膜癌,子宫内膜样腺癌,子宫内膜异位,上皮性卵巢癌,食管癌,食管鳞状细胞癌,原发性血小板增多症(ET),面肩胛肱型肌营养不良症(FSHD),滤泡性淋巴瘤(FL),滤泡性甲状腺癌(FTC),额颞叶痴呆,胃癌,恶性胶质瘤,多形性成胶质细胞瘤(GBM),胶质瘤,血管小球病,肾小球硬化症,错构瘤,HBV-相关的肝硬化,HCV感染,头颈癌,头颈部鳞状细胞癌(HNSCC),听觉损失,心脏病,心力衰竭,肝炎B,肝炎C,肝细胞癌(HCC),肝门部胆管癌,霍奇金氏淋巴瘤,纯合子的镰状细胞疾病(HbSS),亨廷顿氏疾病(HD),高血压,下咽癌,包涵体肌炎(IBM),胰岛瘤,肝内胆管细胞癌(ICC),肾肿瘤,肾病,喉癌,失眠(睡眠疾病),肺的平滑肌瘤,白血病,类型2A肢带肌营养不良(LGMD2A),脂肪瘤,肺腺癌,肺癌,淋巴组织增生症,恶性淋巴瘤,恶性黑色素瘤,恶性间皮细胞瘤(MM),套细胞淋巴瘤(MCL),成神经管细胞瘤,黑素瘤,脑膜瘤,新陈代谢疾病,三好氏远端肌肉病变(MM),多发性骨髓瘤(MM),多发性硬化,MYC-重排的淋巴瘤,骨髓增生异常综合征,骨髓增生性疾病,心肌梗塞,心肌损伤,肌瘤,鼻咽癌(NPC),线状体肌病(NM),肾炎,成神经细胞瘤(NB),中性白细胞增多症,c型尼曼-皮克病(NPC)疾病,非酒精性脂肪肝病(NAFLD),非小细胞性肺癌(NSCLC),肥胖症,口腔癌骨肉瘤卵巢癌(oral carcinomasteosarcoma ovarian cancer)(OC),胰腺癌,胰腺导管腺癌(PDAC),胰腺肿瘤,恐慌症,乳头状甲状腺癌(PTC),帕金森氏综合症,PFV-1感染,咽部疾病,垂体腺瘤,多囊肾病,多囊肝病,真性红细胞增多(PV),多肌炎(PM),原发性胆汁性肝硬化(PBC),原发性骨髓纤维化,朊病毒疾病,前列腺癌,牛皮癣的关节炎,牛皮癣,肺动脉高血压,复发性卵巢癌,肾细胞癌,肾透明细胞癌,色素性视网膜炎(RP),成视网膜细胞瘤,横纹肌肉瘤,风湿性心脏病和心房颤动,类风湿性关节炎,肉瘤,精神分裂症,败血症,浆液性卵巢癌,塞扎里综合征,皮肤病,小细胞肺癌,小脑萎缩症,鳞状细胞癌,T细胞白血病,畸胎瘤,睾丸生殖细胞肿瘤,地中海贫血,甲状腺癌,舌鳞状细胞癌,图雷特综合征,二型糖尿病,溃疡性结肠炎(UC),子宫肌瘤(ULM),葡萄膜黑色素瘤,血管疾病,水泡性口炎或华氏巨球蛋白血症(WM)。

[0427] 病人可以是任何以上讨论的哺乳动物。病人优选是人类。病人是个体。

[0428] 样品可以是任何以上讨论的。样品通常来自任何组织或体液。样品通常包含病人的体液和/或细胞,并可以例如通过使用棉签,如口腔拭子来获得。样品可以是,或来源于,血液,尿,唾液,皮肤,颊上皮细胞或发根样品。本发明方法中目标RNA通常在使用前从样品中提取。

[0429] 该方法可涉及诊断病人的疾病或状况,即确定是否病人患有疾病或状况。病人可以是有症状的。

[0430] 该方法可涉及预测病人的疾病或状况,即确定是否病人有可能发展疾病或状况。病人可以是无症状的。病人可对疾病或状况具有遗传易感性。病人可以具有一个或多个具有疾病或状况的家族成员。

[0431] 改进RNA核苷酸的移动的方法

[0432] 本发明也提供了一种目标RNA多核苷酸相对于所述跨膜孔移动的方法,所述移动被DNA解旋酶控制,所述方法包括:

[0433] a) 提供 (i) RNA多核苷酸和 (ii) DNA解旋酶,其中所述RNA多核苷酸被修饰以包含非-RNA多核苷酸;

[0434] b) 将a) 中提供的所述RNA多核苷酸和DNA解旋酶与跨膜孔接触,使得所述DNA解旋酶控制所述RNA多核苷酸相对于所述跨膜孔的移动。

[0435] RNA多核苷酸的修饰导致结合到其的DNA解旋酶增加。结合到修饰的RNA多核苷酸的DNA解旋酶的增加被定义为结合的DNA解旋酶的量或水平大于,或多于与观察到的非修饰或未修饰的RNA多核苷酸结合的DNA解旋酶的量或水平,所述非修饰或未修饰的RNA多核苷酸即根据本发明的修饰方法没有被修饰的RNA。DNA解旋酶结合到目标RNA多核苷酸的水平可使用常规方法容易地测试,所述常规方法是本领域技术人员已知或常规的。

[0436] 优选DNA解旋酶预先结合到非-RNA多核苷酸。以上描述的任何实施例也采用此方法。例如,在一实施例中,非-RNA多核苷酸可包含至少一个 (i) 具有5个或更多个带电单元的聚合物; (ii) 大约20个核苷酸长度的阻塞-链杂交位点; (iii) 1个或多个非-RNA核苷酸,例如,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,30,40,或50个核苷酸的DNA-解旋酶结合位点; (iv) 1个或多个单元,例如1,2,3,4,5,6,7,8,9,10或更多个单元的停滞化学体,例如Sp18,如W02014/135838描述的,并通过并入引入本文; (v) 大约30个核苷酸长度的系链杂交位点; 和/或 (vi) 促进非-RNA多核苷酸连接到RNA多核苷酸的序列,如前述部分所描述的。

[0437] 实施例进一步说明了DNA解旋酶对于控制DNA/RNA链穿过纳米孔的移动的用途。因此,在一实施例中,提供了用于提高RNA穿过孔被测序的能力或效率的方法。

[0438] 制备本发明构建体的方法

[0439] 制备构建体的方法包含将目标RNA多核苷酸连接到非-RNA多核苷酸。非-RNA多核苷酸的至少一个核苷酸不是核糖核苷酸,即不来自RNA。所述非-RNA多核苷酸可因此包含至少一个核糖核苷酸(或RNA核苷酸)但必须也额外地包含或包括非-RNA核苷酸或序列,即不是RNA的核苷酸或核苷酸的序列。所述非-RNA多核苷酸可包含以上描述的任何实例。优选所述非-RNA多核苷酸包含DNA解旋酶结合位点或DNA适配器。更优选地所述非-RNA多核苷酸包含前导序列。

[0440] 如上讨论地挑选连接方法的位点。优选地,连接方法可选自化学连接,共价连接,酶促连接,杂交,合成方法连接,或使用拓扑异构酶连接。RNA多核苷酸可连接到非-RNA多核苷酸不止一个点,例如两个或三个点。所述连接方法可包含一个,两个,三个,四个,五个或更多个不同的连接方法。根据本发明可使用任何以上描述的连接方法的组合。

[0441] 所述方法可进一步包含确定在DNA解旋酶的控制下构建体是否能够受控移动穿过纳米孔。对于此测试的分析是本领域技术人员已知的。如果RNA多核苷酸穿过纳米孔的移动是受控的,那么本发明的构建体已制成。如果RNA多核苷酸穿过纳米孔的移动是不受控的,



那么本发明的构建体没有制成。

[0442] 在本发明的一个实施例中,制备构建体的方法包含在接触步骤(b)之前将目标RNA多核苷酸连接到非-RNA多核苷酸。在另一实施例中,提供了产生修饰的RNA的方法,其中所述修饰的RNA比未修饰形式的RNA更有效率地穿过孔测序。

[0443] 本发明的方法的修饰

[0444] 本发明提供了修饰目标RNA多核苷酸的方法,所述修饰目标RNA多核苷酸用于表征,例如用于测序。根据本发明,所述修饰的RNA多核苷酸被表征,或被测序。这在上面更详细地讨论了。

[0445] 所述方法包括使用描述的一个或多个方法形成一个或多个修饰的RNA多核苷酸。一个或多个修饰的多核苷酸比未修饰的多核苷酸更容易表征,特别使用链测序。

[0446] 本发明的产品/构建体

[0447] 本发明也提供了使用本发明的修饰方法修饰的RNA多核苷酸。目标RNA多核苷酸通过将RNA多核苷酸连接到非-RNA多核苷酸来修饰以形成本发明的构建体。目标RNA的修饰必然导致DNA解旋酶和修饰的RNA构建体之间增加的相互作用,即与发生在DNA解旋酶和未修饰形式的没有连接的非-RNA多核苷酸的RNA多核苷酸之间的相互作用相比。额外或替换地,非-RNA多核苷酸添加或连接到RNA多核苷酸意味着增加了DNA解旋酶对于修饰的RNA构建体的特异性,即,与DNA解旋酶对于未修饰形式的没有连接的非-RNA多核苷酸的RNA多核苷酸的特异性相比。额外或替换地,非-RNA多核苷酸添加或连接到RNA多核苷酸意味着促进了DNA解旋酶结合到修饰的RNA构建体和/或增加了DNA解旋酶结合到修饰的RNA构建体,即与发生在DNA解旋酶和未修饰形式的没有连接的非-RNA多核苷酸的RNA多核苷酸之间的结合相比。额外或替换地,非-RNA多核苷酸到RNA多核苷酸的添加或连接意味着DNA解旋酶更有效地或更强地结合到修饰的RNA构建体,并且其不太可能从修饰的构建体脱离,即与发生在DNA解旋酶和未修饰形式的没有连接的非-RNA多核苷酸的RNA多核苷酸之间的结合相比。

[0448] 非-RNA多核苷酸可以是任何不是RNA的多核苷酸。非-RNA多核苷酸可以包含至少一个核糖核苷酸,但必须也额外地包含或包括非-RNA核苷酸或序列,即不是RNA的核苷酸或核苷酸的序列。连接的位点和方式如以上讨论进行选择。非-RNA多核苷酸可能或可能不包含预先结合的DNA解旋酶。优选地非-RNA多核苷酸包含DNA解旋酶能够结合的区域(DNA解旋酶结合位点)或DNA适配器。更优选地DNA解旋酶结合位点或DNA适配器包含前导序列,其优先地穿过纳米孔。所述前导序列也可被用于将目标RNA连接到如上讨论的一个或多个锚。所述前导序列也可被连接到目标RNA多核苷酸。

[0449] 本发明的构建体优选是DNA/RNA杂交链,其能使用DNA解旋酶移位穿过MspA纳米孔。

[0450] 所述构建体可进一步包含在多核苷酸链上的条形码部分。多核苷酸条形码是本领域已知的(Kozarewa, I. et al., (2011), Methods Mol. Biol. 733, p279-298)。条形码是核苷酸的特定序列,其能以特定和已知的方式影响流过所述孔的电流。所述条形码部分能够明确地鉴定分析物。优选地所述条形码部分位于前导序列和DNA解旋酶结合位点之间。

[0451] 锚,例如DNA锚,如上所述可杂交到RNA多核苷酸或非-RNA多核苷酸。所述DNA锚可进一步包含间隔器和胆固醇。

[0452] 所述RNA多核苷酸可进一步使用Poly (U) 聚合酶延伸。这确保了读取RNA的全长。

[0453] 根据本发明使用的修饰方法,修饰的RNA多核苷酸可为多种形式。可能的形式,包括但不限于,以下的一个或多个:

[0454] -例如使用点击化学,化学连接到非-RNA多核苷酸的RNA多核苷酸;

[0455] -连接至非-RNA多核苷酸的RNA多核苷酸;

[0456] -杂交到非-RNA前导序列的RNA多核苷酸,具有或不具有桥接部分;

[0457] -杂交到cDNA序列的RNA多核苷酸,具有或不具有桥接部分;

[0458] -使用拓扑异构酶连接到非-RNA多核苷酸的RNA多核苷酸;

[0459] -连接至具有反向碱基区域的非-RNA多核苷酸的RNA多核苷酸,从而该RNA多核苷酸在相反方向运行到剩余的非-RNA多核苷酸。优选地所述非-RNA多核苷酸是DNA多核苷酸。

#### [0460] 试剂盒

[0461] 本发明还提供了用于表征目标RNA多核苷酸的试剂盒。所述试剂盒包含非-RNA多核苷酸,所述非-RNA多核苷酸适于连接到任何用于表征的目标RNA多核苷酸。

[0462] 优选地该适于连接到用于表征的任何目标RNA多核苷酸的非-RNA多核苷酸具有连接的反应基团,例如点击反应基团。连接有反应基团的非-RNA多核苷酸可最终被末端-使用者(end-user)使用来与选择的任何目标RNA多核苷酸进行反应以形成共价键。优选地所述目标RNA多核苷酸也具有连接的反应基团。例如,点击反应基团,其与连接到非-RNA多核苷酸的反应基团进行反应以形成共价键。

[0463] 或者,该适于连接到用于表征的任何目标RNA多核苷酸的非-RNA多核苷酸具有连接酶,或是寡核苷酸或引物,能够通过末端使用者来杂交到选择的目标RNA多核苷酸的任意区域,并作为用于cDNA合成的起始点。

[0464] 或者,该适于连接到用于表征的任何目标RNA多核苷酸的非-RNA多核苷酸包含结合至特定DNA多核苷酸的拓扑异构酶。该结合DNA的拓扑异构酶可通过末端-使用者最终被用来将非-RNA (DNA) 多核苷酸连接到选择的任何目标RNA多核苷酸。该末端使用者可将具有自由5' 羟基的RNA与结合了DNA的拓扑异构酶孵育。该拓扑异构酶然后将RNA连接到DNA。

[0465] 或者该适于连接到用于表征的任何目标RNA多核苷酸的非-RNA多核苷酸包含具有反向碱基的区域(例如反向碱基的DNA区域)。该反向区域能够被连接到真核RNA的5' 末端,该真核RNA通过添加在相反方向运行的7-甲基鸟苷来进行修饰。

[0466] 关于本发明方法的以上讨论的任何实施方式同样适用于所述试剂盒。该试剂盒可进一步包含能够预先结合到非-RNA多核苷酸的DNA解旋酶结合蛋白。该试剂盒可进一步包含孔和膜的组件,例如形成两性分子层例如脂质双分子层需要的磷脂。

[0467] 本发明的试剂盒可以另外包括能使上面描述的任何实施方式实施的一种或多种其他试剂或仪器。所述试剂或仪器包括以下的一种或多种:合适的缓冲剂(水溶液),用于从受体获得样本的工具(诸如导管或含有针的仪器),如上所定义的膜或者电压钳或膜片钳装置。存在于所述试剂盒中的试剂可以是干态的,使得液体样本可使所述试剂重悬。所述试剂盒还可以可选的包括能使试剂盒用于本发明方法的说明书,或关于本发明方法可用于哪种病人的细节。所述试剂盒通常包含核苷酸。所述试剂盒优选包含dAMP, dTMP, dGMP和dCMP。所述试剂盒优选不包含用以扩增和/或表达多核苷酸的工具。

[0468] 下列实施例阐明了本发明。

#### [0469] 实施例1

[0470] 本实施例显示了样品制备过程,其1) 延长RNA/DNA链的RNA区域,2) 使锚退火,3) 与酶结合,和然后4) 在电生理学实验中检测最终的链。本实施例说明可以使用DNA解旋酶(T4Dda-E94C/A360C (SEQ ID NO:14,具有突变E94C/A360C和然后( $\Delta$ M1)G1))来控制合成的DNA/RNA链(连接到RNA链的DNA前导序列,如图3所示)穿过MspA纳米孔的移动。

#### [0471] 材料与方法

[0472] 1.1使用Poly (U) 聚合酶的DNA/RNA链的3' 末端的延长

[0473] 将下表3列出的试剂进行混合,在37℃孵育10分钟。然后使用Agencourt Ampure SPRI珠以每 $\mu$ L样品1.8 $\mu$ L SPRI珠的比例纯化该混合物。该样品被称为样品1 (DNA/RNA 2)。图4显示聚合酶延长反应是成功的,由于在图中标注Y的宽条带对应于可变延长的DNA/RNA1。

| [0474] | 试剂  | 体积           | 储液的浓度        | 最终浓度      |
|--------|---|--------------|--------------|-----------|
|        | 合成的 DNA/RNA 1 (SEQ ID NO: 16 在其 3' 末端连接到 4 个 iSpC3 间隔器,所述 4 个 iSpC3 间隔器在相对端连接到 SEQ ID NO: 17 的 5' 末端) | 0.4 $\mu$ l  | 100 $\mu$ M  | 1 $\mu$ M |
|        | rUTP  | 0.4 $\mu$ l  | 100 mM       | 1 mM      |
|        | 缓冲液   | 4 $\mu$ l    | 10 x         | 1 x       |
| [0475] | PolyU Pol (NEB)   | 2 $\mu$ l    | 2 U/ $\mu$ l | 4 U       |
|        | NF H <sub>2</sub> O   | 33.2 $\mu$ l |              |           |
|        |   |              |              |           |
|        | 总计  | 40 $\mu$ l   |              |           |

[0476] 表3

[0477] 1.2锚退火

[0478] 将以下表4列出的试剂进行混合,并在65℃孵育,然后以每秒0.1℃的速度冷却至4℃。该样品被称为样品2。

| [0479] | 试剂  | 体积           | 储液浓度        | 最终浓度         |
|--------|---|--------------|-------------|--------------|
|        | 样品 1  | 9 $\mu$ l    | ~1 $\mu$ M  | 942 $\mu$ M  |
|        | 锚 (SEQ ID NO: 18 在其 3'末端连接到 6 个 iSp18 间隔器, 2 个胸腺嘧啶和 3' 胆固醇 TEG) | 0.36 $\mu$ l | 100 $\mu$ M | 3.77 $\mu$ M |
|        | 10 mM TRIS pH 7.5 50 mM NaCl                                    | 0.19 $\mu$ l | 50 x        | 1 x          |
|        |   |              |             |              |
|        | 总计  | 9.55 $\mu$ l |             |              |

[0480] 表4

[0481] 1.3结合DNA解旋酶

[0482] 样品2 (0.28 $\mu$ L) 和T4Dda-E94C/A360C (0.36 $\mu$ L, 3.8 $\mu$ M, SEQ ID NO:14, 具有突变E94C/A360C和然后 ( $\Delta$  M1) G1) 在缓冲液 (10mM TRIS pH 7.5, 50mM NaCl) 中在室温下孵育1小时。该样品被称为样品3。

[0483] 1.4电生理现象

[0484] 将样品3稀释到缓冲液 (1221 $\mu$ L 600mM KCl, 50mM HEPES pH 8.0, 463mM甘油) 中。将MgCl<sub>2</sub> (13 $\mu$ L, 1M) 和ATP (65 $\mu$ L, 100mM) 加入到样品3缓冲液混合物中达到总体积1300 $\mu$ L。

[0485] 在缓冲液 (25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾~pH 8.0) 中, 从插入到嵌段共聚物的单MspA纳米孔获得电测量值。在获得插入到嵌段共聚物的单孔后, 让缓冲液 (2mL, 25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾, pH 8.0) 流过系统以去除任何过量的MspA纳米孔。

[0486] 过量的KCl缓冲液 (600mM KCl, 50mM HEPES pH~8, 463mM甘油) 流过系统, 该KCl缓冲液与电极缓冲液 (25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾, pH8.0) 通过琼脂糖桥分离。

[0487] 在-120mV电压下运行本实验, 并且监控解旋酶控制的DNA移动。

[0488] 结果

[0489] 当将样品3 (DNA/RNA 2的卡通示意图在图3中示出) 加入到纳米孔系统时, 观察解旋酶控制的DNA移动。对于样品3的解旋酶控制的DNA移动的实例在图5中示出。随着链移位穿过纳米孔, 合成链的多个区域被识别 (区域1=poly (dT) 前导 (SEQ ID NO:16), 区域2=iSpC3间隔器, 区域3=RNA序列 (SEQ ID NO:17) 和区域4=可变长度的poly (U) RNA)。本实施例显示可以使用DNA解旋酶来控制DNA/RNA链 (卡通示意图在图3中示出) 穿过MspA纳米孔的移动。该Poly (U) 聚合酶延长步骤确保读取全长的RNA。

[0490] 实施例2

[0491] 本实施例显示使用T4DNA连接酶将DNA链 (SEQ ID NO:21) 连接到RNA链 (SEQ ID NO:19)。

[0492] 材料与方法

[0493] 2.1使用T4DNA连接酶将DNA链连接到RNA链

[0494] 将下表5列出的试剂混合并放置在热循环仪上。该热循环仪设置为如下表6中的程序。然后使用10%PAGE TBE-尿素变性BioRad标准凝胶, 在140V运行60分钟分析该样品。

[0495]

| 试剂                       | 体积    | 储液的浓度  | 最终浓度 |
|--------------------------|-------|--------|------|
| 5' 磷酸盐RNA (SEQ ID NO:19) | 0.3u1 | 50uM   | 1uM  |
| T4DNA连接酶缓冲液              | 1.5   | 10x    | 1x   |
| T4DNA连接酶                 | 1u1   | 10U/u1 | 10U  |
| DNA夹板 (SEQ ID NO:20)     | 0.3u1 | 50uM   | 1uM  |
| Cy3DNA (SEQ ID NO:21)    | 0.6u1 | 50uM   | 2uM  |
| ATP                      | 0.3u1 | 50mM   | 1mM  |
| NF H <sub>2</sub> O      | 11u1  |        |      |
|                          |       |        |      |

|        |      |      |         |         |
|--------|------|------|---------|---------|
|        | 总计   | 15ul |         |         |
| [0496] | 表5   |      |         |         |
| [0497] | 循环序号 | 步骤   | 温度 (°C) | 时间      |
| [0498] | 1    | 连接   | 25      | 1:00:00 |
|        | 1    | 变性   | 65      | 0:10:00 |

[0499] 表6

[0500] Results

[0501] 使用该TBE-尿素变性凝胶分析CY3DNA (SEQ ID NO:21) 到RNA (SEQ ID NO:19) 的连接。图8,泳道2-3显示了在没有T4DNA连接酶的情况下,随着DNA夹板 (SEQ ID NO:20) 浓度增加控制连接步骤的反应。对于这些控制反应,区域A和B没有观察到条带,表明在这些情况下没有发生连接反应。泳道5-8显示了在T4DNA连接酶存在下,随着DNA夹板 (SEQ ID NO:20) 浓度增加的连接步骤。对于所有的泳道5-8,在区域A和区域B中都可见条带,其对应于具有杂交的DNA夹板 (A) 的连接底物和不具有杂交的DNA夹板 (B) 的连接底物。随着夹板浓度的增加,条带A的强度也增加。在泳道9和10中示出实施的两个进一步的控制反应。泳道9对应于泳道5所示的相同样品,其进一步用添加额外的DNA夹板 (4.5x) 来处理,所述DNA夹板 (4.5x) 在连接步骤之后加入。显示了,当与泳道5相比时,期望的上部的条带 (对应于具有杂交夹板的连接产物) 强度的增加和下部条带B (对应于不具有杂交夹板的连接产物) 强度的减小。泳道10对应于泳道6所示的相同样品,其进一步用ExoI在37°C处理30分钟。在ExoI处理期间,加热导致了连接链和DNA夹板的分离。所述DNA夹板优先被ExoI消化,因为其具有3' DNA末端。这导致了区域A中的条带的消失,其对应于杂交到DNA夹板的连接产物,,因为DNA夹板被ExoI消化。在用ExoI进行消化后,区域B中的条带仍然可见。这意味着连接步骤是成功的,因为如果没有DNA连接到RNA,那么没有连接的DNA将会被ExoI消化掉,从而在区域B中不会看到任何条带。因此,本实施例显示了可以使用T4DNA连接酶将DNA连接到RNA。

[0502] 实施例3

[0503] 本实施例显示了使用点击化学的化学连接DNA链与RNA链的样品制备方法。这在两个不同的样品上实施,其中一个样品具有连接到DNA的荧光基团,以为了通过使用凝胶电泳来确定化学连接步骤。没有连接荧光基团的DNA/RNA链然后在电生理学实验中进行测试。本实施例表明可以使用DNA解旋酶 (T4Dda-E94C/A360C (SEQ ID NO:14,具有突变E94C/A360C并然后 (Δ M1) G1)) 来控制RNA链穿过MspA或胞溶素纳米孔的移动,所述RNA链通过铜-介导的 (copper-mediated) 的点击化学来连接到非-RNA多核苷酸 (该构建体的卡通示意图如图10所示)。

[0504] 材料与方法

[0505] 3A.1 DNA X1至RNA X1的点击反应

[0506] 将RNA X1,DNA X1和夹板X1 (下表7中所列) 在缓冲液 (TRIS-NaCl (500mM-2.5M) pH8) 中混合。该DNA X1,RNA X1和夹板X1在PCR反应器 (程序升温至55°C并以0.1°C/s冷却至4°C) 中退火。然后将CuSO<sub>4</sub>,三羟甲基氨基甲烷 (3-羟丙基三唑基甲基胺) 和抗坏血酸钠 (Sigma A4034) 加入到DNA X1/RNA X1/夹板X1混合物中,该样品然后放置在热循环仪上。该热循环仪设置为如下表8中的程序。然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL

SPRI珠的比例纯化该样品。该样品被称为样品3A (DNA/RNA 3A)。然后在5%PAGE TBE BioRad标准凝胶上分析该样品,其电生理学如以下3.3所述。

| [0507] | 试剂   | 体积      | 储液的浓度  | 最终浓度    |
|--------|--|---------|--------|---------|
|        | 来自 Trilink Biotech, CA 的萤火虫荧光素酶 mRNA (RNA X1, 具有 SEQ ID NO: 26 的开放阅读框的 mRNA, 所述 SEQ ID NO: 26 具有 5'-己基-G 作为大多数 5'核苷酸并具有 3'聚 A 尾)                 | 12.9 ul | 1.1 uM | 0.94 uM |
|        | 具有叠氮化物的 DNA (DNA X1, SEQ ID NO: 22, 在其 3'末端连接 4 个 iSp18 间隔器, 该 4 个 iSp18 间隔器在其相对末端连接到 SEQ ID NO: 23 的 5'末端, 所述 SEQ ID NO: 23 在其 3'末端连接 3 叠氮化物 N) | 0.3 ul  | 50 uM  | 1 uM    |
|        | 夹板 X1 (SEQ ID NO: 24)  | 0.6 ul  | 50 uM  | 2 uM    |
|        | CuSO <sub>4</sub>  | 0.3 ul  | 50 mM  | 1 mM    |
|        | 三羟甲基氨基甲烷(3-羟丙基三唑基甲基胺)  | 0.3 ul  | 100 mM | 2 mM    |
|        | 抗坏血酸钠 (Sigma A4034)  | 0.3 ul  | 200 mM | 4 mM    |
|        | TRIS-NaCl (500mM-2.5M) pH 8  | 0.3 ul  | 500 mM | 10 mM   |
|        |  |         |        |         |
|        | 总计   | 15 ul   |        |         |

[0508] 表7

| [0509] | 循环序号 | 步骤 | 温度 (°C) | 时间      |
|--------|------|----|---------|---------|
|        | 1    | 点击 | 23      | 0:30:00 |
|        | 1    | 点击 | 37      | 0:30:00 |

[0510] 表8

#### [0511] 3B.1 DNA X2至RNA X1的点击反应

[0512] 将RNA X1, DNA X2和夹板X1 (下表9中所列) 在缓冲液 (TRIS或MOPS (500mM-2.5M) pH 6.8-7) 中混合。该DNA X2, RNA X1和夹板X1在PCR反应器 (程序升温至55℃并以0.1℃/s冷却至4℃) 中退火。然后将CuSO<sub>4</sub>, 三羟甲基氨基甲烷 (3-羟丙基三唑基甲基胺) 和抗坏血酸钠 (Sigma A4034) 加入到DNA X2/RNA X1/夹板X1混合物中, 该样品然后放置在热循环仪上。该热循环仪设置为如下表10中的程序。然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI磁珠的比率纯化该样品。该样品被称为样品3B (DNA/RNA 3B)。然后在5%PAGE TBE BioRad标准凝胶上分析该样品。

| [0513] | 试剂   | 体积      | 储液的浓度  | 最终浓度    |
|--------|--|---------|--------|---------|
|        | 具有 5'-己基-G 的萤火虫荧光素酶 mRNA (RNA X1, 具有 SEQ ID NO: 26 的开放阅读框的 mRNA, 所述 SEQ ID NO: 26 具有 5'-己基-G 作为大多数 5'核苷酸 并具有 3' 聚 A 尾) | 12.9 ul | 1.1 uM | 0.94 uM |
|        | 具有叠氮化物的 Cy3 DNA (DNA X2, SEQ ID NO: 25, 在其 5'末端连接 Cy3 并且在其 3'末端连接 3 叠氮化物 N)  | 0.3 ul  | 50 uM  | 1 uM    |
|        | 夹板 X1 (SEQ ID NO: 24)  | 0.6 ul  | 50 uM  | 2 uM    |
|        | CuSO <sub>4</sub>  | 0.3 ul  | 50 mM  | 1 mM    |
|        | Cu 配体  | 0.3 ul  | 100 mM | 2 mM    |
|        | 抗坏血酸钠 (Sigma A4034)  | 0.3 ul  | 200 mM | 4 mM    |
|        | TRIS-NaCl (500mM-2.5M) pH 8  | 0.3 ul  | 500 mM | 10 mM   |
|        | 总计   | 15 ul   |        |         |

[0514] 表9

| [0515] | 循环序号 | 步骤 | 温度 (°C) | 时间      |
|--------|------|----|---------|---------|
|        | 1    | 点击 | 23      | 0:30:00 |
|        | 1    | 点击 | 37      | 0:30:00 |

[0516] 表10

## [0517] 3.3电生理学

[0518] 如实施例1 (参见1.2节) 系链退火到样品3A。样品3A (0.28μl) 和T4Dda-E94C/A360C (0.36μL, 3.8μM, SEQ ID NO:14, 具有突变E94C/A360C, 然后 (Δ M1) G1) 稀释到缓冲液 (1221μL 的500mM KCl, 25mM磷酸钾, pH 8.0) 中。将MgCl<sub>2</sub> (13μL, 1M) 和ATP (65μL, 100mM) 加入到样品3A缓冲液混合物中以达到总体积1300μL。

[0519] 在缓冲液 (25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾~pH 8.0) 中, 从插入到嵌段共聚物的单MspA或胞溶素纳米孔获得电测量值。获得插入到嵌段共聚物的单孔后, 然后让缓冲液 (2mL, 25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾, pH 8.0) 流过系统以去除任何过量的MspA或胞溶素纳米孔。

[0520] 在添加样品之前, 让过量的缓冲液 (500mM KCl, 25mM磷酸钾, pH 8.0) 流过系统。最后, 之后将与样品3A结合的T4Dda-E94C/A360C添加到该纳米孔系统, 在-120mV电压下运行本实验, 并且监控解旋酶控制的DNA移动。

## [0521] 结果

[0522] 使用TBE-尿素变性凝胶来分析DNA (或者DNA X1 (实施例3A, SEQ ID NO:22, 在其3'末端连接4个iSp18间隔器, 该4个iSp18间隔器在其相对末端连接到SEQ ID NO:23的5'末

端,所述SEQ ID NO:23在其3'末端连接有3叠氮化物N)或DNA X2(实施例3B,SEQ ID NO:25,在其5'末端连接CY3并且在其3'末端连接3叠氮化物N))到RNA X1的点击反应。实施例3B中使用的DNA由于连接的Cy3基团在SYBR染色之前是可见的。图9,泳道3和9显示了在实施例3B的点击反应后制备的样品。泳道3的顶部的荧光带(在图9中以白色框高亮表示)显示具有Cy3标签的DNA已连接到mRNA(只有DNA的荧光带在箭头3处,结合mRNA的DNA在箭头1处)。泳道2和8显示了在实施例3A的点击反应后制备的样品(在DNA上没有荧光基团)。泳道11中的条带对应未反应的DNA(箭头2)并且结合mRNA的DNA是箭头1。因此,该凝胶显示了用于将DNA(具有荧光标签(X1)或不具有荧光标签(X2))连接到mRNA(RNA X1)的点击反应是成功的。

[0523] 当将样品3A(代表DNA/RNA 3A的卡通示意图如图10所示)加入到纳米孔系统时,观察解旋酶控制的RNA移动。在图11(区域2,3和4)中显示解旋酶控制的RNA移动穿过MspA的实施例,在图21中显示解旋酶控制的RNA移动穿过胞溶素突变体的实施例。图11中的链的多个区域随着该链移位穿过纳米孔被鉴定。区域2,3和4合起来代表通过点击反应与目标RNA反应的DNA前导区(区域2=连接4个iSp18间隔器的SEQ ID NO:22;区域3=SEQ ID NO:23;区域4=萤火虫荧光素酶,具有SEQ ID NO:26的具有开放阅读框的mRNA,并且具有\*的箭头突出表示了通过点击反应形成的连接)。区域1显示没有与RNA(连接4个iSp18间隔器的SEQ ID NO:22,该4个iSp18间隔器连接到SEQ ID NO:23)反应的单独的DNA前导区的纳米孔移位。在图21中对于胞溶素轨迹也可识别到相似特征。该实施例显示通过将非-RNA多核苷酸(连接4个iSp18间隔器的SEQ ID NO:22,该4个iSp18间隔器连接到SEQ ID NO:23)连接到长mRNA链(卡通示意图如图10所示),可以使用DNA解旋酶控制长mRNA链穿过MspA或胞溶素纳米孔的移动。

#### [0524] 实施例4

[0525] 本实施例显示使用T4DNA连接酶将发夹型寡核苷酸(3T发夹=SEQ ID NO:29,在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接到SEQ ID NO:27的5'末端,或10T发夹=SEQ ID NO:29,在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接到SEQ ID NO:28的5'末端)连接到RNA链(SEQ ID NO:30)的3'末端,并随后从发夹型寡核苷酸的3'末端进行反转录。该结合和反转录演示了构建RNA/cDNA构建体的方法(参见图13)。

[0526] 发夹型寡核苷酸(或3T或10T发夹)的3'末端的聚T突出杂交到mRNA的聚A-尾,并作为夹板用于高效将DNA连接到RAN。所述发夹作为用于序列反转录的引物。

#### [0527] 材料与方法

[0528] 将下表11列出的试剂进行混合并放置在热循环仪上。该热循环仪设置为如下表12中的程序。然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化该混合物。纯化后,使用生命技术超级脚本(Life Technologies Super Script)II实施反转录:根据制造商的程序混合下表13中的试剂并放置在按照表14中设置的程序的热循环仪上。然后使用10%PAGE TBE-尿素变性BioRad标准凝胶分析该样品,在140V运行60分钟。



[0529]

| 试剂                           | 体积       | 储液的浓度    | 最终浓度   |
|------------------------------|----------|----------|--------|
| RNA 链 (SEQ ID NO: 30)        | 0.72 ul  | 556ng/ul | 0.2 uM |
| 聚 T 发夹 (以上描述的 3T 发夹或 10T 发夹) | 0.4 ul   | 50uM     | 1 uM   |
| T4 DNA 连接酶缓冲液                | 4 ul     | 5 x      | 1 x    |
| T4 DNA 连接酶                   | 1 ul     | 2000U    | 2000 U |
| NF 水                         | 13.88 ul |          |        |
|                              |          |          |        |
| 总计                           | 20 ul    |          |        |

[0530] 表11

[0531]

| 循环序号 | 步骤 | 温度 (°C) | 时间      |
|------|----|---------|---------|
| 1    | 连接 | 16      | 2:00:00 |

[0532] 表12

[0533]

| 试剂                     | 体积   | 储液的浓度     | 最终浓度       |
|------------------------|------|-----------|------------|
| 聚T发夹连接后的RNA            | 7ul  | 35.5ng/ul | 248.5ng/反应 |
| 脱氧核糖核苷三磷酸 (dNTP)       | 1ul  | 各10uM     | 0.5uM      |
| NF水                    | 5ul  |           |            |
| 第一链缓冲液                 | 4ul  | 5x        | 1x         |
| 0.1M二硫苏糖醇 (DTT)        | 2ul  | 0.1M      | 0.01M      |
| 超级脚本 (Super Script) II | 1ul  | 200U      | 200U       |
| 总计                     | 20ul |           |            |

[0534] 表13

[0535]

| 循环序号 | 步骤  | 温度 (°C) | 时间      |
|------|-----|---------|---------|
| 1    | 反转录 | 42      | 0:50:00 |
| 2    | 变性  | 70      | 0:15:00 |

[0536] 表14

[0537] 结果

[0538] 使用TBE-尿素变性凝胶(看图14)分析3T和10T发夹连接和反转录。在泳道5中,3T-发夹的局部连接是以向上偏移的RNA链(SEQ ID NO:30,图14中该RNA链被标记为条带A)的条带的形式可见。在反转录(泳道6)后,由于RNA/cDNA的双链构建体(在图13中示出)比单链RNA(SEQ ID NO:30,在图14中显示为条带A)迁移更快,该条带移动到RNA链(SEQ ID NO:30,在图14中为条带A)的水平之下。在泳道7中,10T发夹的连接以接近100%效率发生,并以在单链RNA链(SEQ ID NO:30,在图14中显示为条带A)的水平之上的单条带可见。泳道8显示通过10T发夹作为引物反转录后的样品,其中该连接并反转录的产物向下移动到单链RNA条带(SEQ ID NO:30,在图14中显示为条带A)之下。在泳道9中的杂交控制样品(不具有T4DNA连接酶的反应混合物)在单链RNA条带(SEQ ID NO:30,在图14中显示为条带A)水平处可见,并

只观察到对于连接产物的移动。

#### [0539] 实施例5

[0540] 本实施例显示在电生理学实验中如何分析RNA/cDNA构建体。为了获得RNA/cDNA构建体,将10T发夹(SEQ ID NO:29,在其5'末端连接磷酸基团并在其3'末端连接4个iSpC3间隔器,所述4个iSpC3间隔器在其相对末端连接到SEQ ID NO:28的5'末端)连接到萤火虫荧光素酶mRNA(RNA X1,具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为大多数5'核苷酸并具有3'聚A尾)的3'末端。在连接之前,使用点击化学将具有叠氮化物的DNA(DNA X1,SEQ ID NO:22,在其3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:23的5'末端,所述SEQ ID NO:23在其3'末端连接3叠氮化物N)连接到具有5'-己基-G的萤火虫荧光素酶mRNA(RNA X1,具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为大多数5'核苷酸并具有3'聚A尾)的5'末端,如实施例3中描述的。之后进行发夹连接反转录。

#### [0541] 材料与方法

#### [0542] 点击反应

[0543] 如实施例3B点击反应描述的,将具有5'-己基-G的萤火虫荧光素酶mRNA的RNA(RNA X1,具有SEQ ID NO:26的开放阅读框的mRNA,所述SEQ ID NO:26具有5'-己基-G作为大多数5'核苷酸并具有3'聚A尾)连接至具有叠氮化物的DNA((DNA X1,SEQ ID NO:22,在其3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:23的5'末端,所述SEQ ID NO:23在其3'末端连接3叠氮化物N))。该样品被称为样品3B。

#### [0544] 发夹连接和反转录

[0545] 如实施例4中描述的,之后将样品3B(DNA/RNA3B)连接到10T发夹并反转录。试剂的体积和用量在以下表15和17中示出,热循环条件在表16和18中示出。在连接/反转录后制备的样品被称为样品5B。该样品构建体的卡通示意图如图15所示。

| [0546] | 试剂                           | 体积     | 储液的浓度    | 最终浓度     |
|--------|------------------------------|--------|----------|----------|
|        | 样品 3B                        | 10 ul  | 22ng/ul  | 0.017 uM |
|        | 聚 T 发夹 (以上描述的 3T 发夹或 10T 发夹) | 0.4 ul | 50uM     | 1 uM     |
|        | T4 DNA 连接酶缓冲液                | 4 ul   | 5 x      | 1 x      |
|        | T4 DNA 连接酶                   | 1 ul   | 2000U/ul | 2000 U   |
|        | NF 水                         | 4.6 ul |          |          |
|        |                              |        |          |          |
|        | 总计                           | 20 ul  |          |          |

#### [0547] 表15

| [0548] | 循环序号 | 步骤 | 温度(°C) | 时间      |
|--------|------|----|--------|---------|
|        | 1    | 连接 | 16     | 2:00:00 |

#### [0549] 表16

|        |                        |      |         |          |
|--------|------------------------|------|---------|----------|
| [0550] | 试剂                     | 体积   | 储液的浓度   | 最终浓度     |
|        | 反转录后的样品3B              | 12ul | 20ng/ul | 240ng/反应 |
|        | dNTPs                  | 1ul  | 各10uM   | 0.5uM    |
|        | 第一链缓冲液                 | 4ul  | 5x      | 1x       |
|        | 0.1M二硫苏糖醇              | 2ul  | 0.1M    | 0.01M    |
|        | 超级脚本 (Super Script) II | 1ul  | 200U/ul | 200U     |
|        | 总计                     | 20ul |         |          |

[0551] 表17

|        |      |     |         |         |
|--------|------|-----|---------|---------|
| [0552] | 循环序号 | 步骤  | 温度 (°C) | 时间      |
|        | 1    | 反转录 | 42      | 0:50:00 |
|        | 2    | 变性  | 70      | 0:15:00 |

[0553] 表18

[0554] 电生理学

[0555] 如实施例1 (参见1.2节), 将系链退火到样品5B。样品5B (4μl) 和T4Dda-E94C/C109A/C136A/A360C (0.36μL, 3.8μM, SEQ ID NO:14, 具有突变E94C/C109A/C136A/A360C然后 (Δ M1) G1) 稀释到缓冲液 (1221μL的500mM KCl, 25mM磷酸钾, pH 8.0) 中。将MgCl<sub>2</sub> (13μL, 1M) 和ATP (65μL, 100mM) 加入到样品5B (DNA/RNA/cDNA 5B) 缓冲液混合物中以达到总体积1300μL。

[0556] 在缓冲液 (25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾~pH 8.0) 中, 从插入到嵌段共聚物的单MspA或CsgG-Eco- (Y51T/F56Q) -StrepII (C)) 9 (具有突变Y51T/F56Q的SEQ ID NO:44, 其中StepII (C) 是SEQ ID NO:45, 并被连接在C-末端) 纳米孔获得电测量值。获得插入到嵌段共聚物的单孔后, 然后让缓冲液 (2mL, 25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾, pH 8.0) 流过系统以去除任何过量的MspA或CsgG纳米孔。

[0557] 在添加样品之前, 让过量的缓冲液 (500mM KCl, 25mM磷酸钾, pH 8.0) 流过系统。最后, 连接到样品5B (DNA/RNA/cDNA 5B) 的T4Dda-E94C/C109A/C136A/A360C之后添加到该纳米孔系统, 在-140mV电压下运行本实验, 并且监控解旋酶控制的DNA移动。

[0558] 结果

[0559] 当将样品5B添加到纳米孔系统, 观察解旋酶控制的DNA/RNA/cDNA移动。对于MspA的解旋酶控制的DNA/RNA/cDNA移动事件的实施例如图16 (区域1-5) 所示, 对于CsgG则如图22所示。图16的链的多个区域随着该链移位穿过纳米孔被鉴定; 区域1代表通过点击反应与目标RNA反应的非-RNA多核苷酸; 区域2=萤火虫荧光素酶mRNA; 区域3=连接到mRNA的10T发夹中存在的iSpC3间隔器; 区域4=连接至mRNA的10T发夹的聚T区域; 区域5=通过mRNA的反转录生成的cDNA。在图22中的CsgG迹线中可以鉴定相似的特征。该实施例显示可以使用DNA解旋酶来控制链DNA/RNA/cDNA (通过将DNA链连接至mRNA链形成, 将发夹连接至该mRNA链, 并使用该发夹对mRNA进行反转录) (卡通示意图如图15所示) 穿过MspA或CsgG纳米孔的移动。

[0560] 实施例6

[0561] 本实施例显示通过在连接前移除帽, 如何将非-RNA多核苷酸连接到具有5' 甲基鸟苷帽的RNA链 (与一些天然细胞mRNA同源)。在这种情况下, 使用加帽的RNA链 (SEQ ID NO:

30,其具有通过5'至5'三磷酸连接在该链的5'末端连接的7-甲基鸟苷帽),然后使用RNA 5'焦磷酸水解酶(RppH)使其脱帽,接着使用T4RNA连接酶1使其连接至非-RNA多核苷酸(30个SpC3间隔器连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)。

#### [0562] 材料与amp;方法

[0563] 为了将非-RNA多核苷酸连接至加帽的RNA链(SEQ ID NO:30,其具有通过5'至5'三磷酸连接在该链的5'末端连接的7-甲基鸟苷帽)的5'末端,首先使用作为脱帽酶的RppH让mRNA脱帽。将下表19列出的试剂进行混合并放置在热循环仪上,该热循环仪设置为如下表20中的程序。然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化该反应混合物。接着,通过混合表21中的试剂将非-RNA多核苷酸(30个SpC3间隔器连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)连接至RNA,并将该混合物放置在热循环仪上,该热循环仪设置为如下表22中的程序。然后使用5%PAGE TBE-尿素变性BioRad标准凝胶分析该样品,在140V运行60分钟。

|        |   |         |          |             |
|--------|---|---------|----------|-------------|
| [0564] | 试剂  | 体积      | 储液的浓度    | 最终浓度        |
|        | 加帽的 RNA 链 (SEQ ID NO: 30, 其具有通过 5' 至 5' 三磷酸连接在该链的 5' 末端连接的 7-甲基鸟苷帽) | 5.39 ul | 556ng/ul | 2996.8ng/反应 |
|        | NEB2 缓冲液  | 2 ul    | 10x      | 1x          |
|        | RNA 5' 焦磷酸水解酶   | 2 ul    | 5U/ul    | 10U         |
|        | NF H <sub>2</sub> O   | 10.61ul |          |             |
|        | 总计  | 20 ul   |          |             |

#### [0565] 表19

|        |      |    |       |         |
|--------|------|----|-------|---------|
| [0566] | 循环序号 | 步骤 | 温度(℃) | 时间      |
|        | 1    | 脱帽 | 37    | 1:00:00 |

#### [0567] 表20

|        |   |        |          |          |
|--------|---|--------|----------|----------|
|        | 试剂  | 体积     | 储液的浓度    | 最终浓度     |
|        | 脱帽的 RNA 链 (SEQ ID NO: 30)   | 1.5 ul | 166ng/ul | 250ng/反应 |
|        | T4 RNA 连接酶 1 反应缓冲液  | 2 ul   | 10x      | 1x       |
| [0568] | 非-RNA 多核苷酸 (30 个 SpC3 间隔器连接连接至 SEQ ID NO: 31 的 5'末端, SEQ ID NO: 31 的 3'末端连接 4 个 iSp18 间隔器, 所述 4 个 iSp18 间隔器在其相对末端连接 SEQ ID NO: 32 的 5'末端, SEQ ID NO: 32 的 3'末端连接到 4 个 5-硝基吡啶, 该 4 个 5-硝基吡啶在其相对末端连接 RNA 序列 CAAGGG) | 2.4 ul | 50uM     | 8.33uM   |
|        | ATP   | 0.4 ul | 50mM     | 1mM      |
|        | NF H <sub>2</sub> O   | 0.8 ul |          |          |
|        | T4 RNA 连接酶 1  | 2.9ul  | 10U/ul   | 29U      |
| [0569] | PEG 8k  | 10 ul  | 50%      | 25%      |
|        | 总计  | 20 ul  |          |          |

[0570] 表21

|        |      |    |         |         |
|--------|------|----|---------|---------|
| [0571] | 循环序号 | 步骤 | 温度 (°C) | 时间      |
|        | 1    | 连接 | 16      | 4:00:00 |

[0572] 表22

[0573] 结果

[0574] 使用TBE-尿素变性凝胶(参见图17)来分析成功脱帽的RNA链(SEQ ID NO:30)并随后分析与非-RNA多核苷酸(30个SpC3间隔器连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)的连接。图17的泳道4显示额外的条带(以C标记),其在RNA链(SEQ ID NO:30,在泳道2中显示并标记为A的对照)的水平之上观察到,表明发生具有~40%效率的非-RNA多核苷酸(30个SpC3间隔器连接连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)的成功连接。

[0575] 实施例7

[0576] 本实施例显示如何从酿酒酵母提取的细胞mRNA制备2D(正义RNA和反义DNA)文库,以允许DNA解旋酶控制天然酿酒酵母mRNA链穿过纳米孔的移动。

[0577] 材料与方法

[0578] 通过将一发夹连接到mRNA的3'末端,将该mRNA的5'末端脱帽,反转录以产生DNA互补链,以及将非-RNA多核苷酸连接到该mRNA的5'末端,来从酿酒酵母提取的细胞mRNA制备

2D (正义RNA和反义DNA)。

[0579] 连接3' 发夹

[0580] 来自酿酒酵母的聚A+mRNA从Clontech中购买。通过将下表23列出的试剂进行混合并将该混合物放置在如下表24所设程序的热循环仪上,使得杂交到mRNA的聚A尾的A发夹连接到mRNA。然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化该混合物并在16ul的NF水中洗脱。

| [0581] | 试剂        | 体积     | 储液的浓度   | 最终浓度   |
|--------|-----------|--------|---------|--------|
|        | 聚 A+ mRNA | 0.5 ul | 1 ug/ul | 500 ng |

|        |  |         |          |        |
|--------|--|---------|----------|--------|
| [0582] | 聚 T 发夹 (SEQ ID NO: 29, 在其 5'末端连接磷酸基团, 在其 3'末端连接四个 iSpC3 间隔器, 所述四个 iSpC3 间隔器在其相对末端连接到 SEQ ID NO: 28 的 5'末端) | 0.4 ul  | 25uM     | 1 uM   |
|        | T4 DNA 连接酶缓冲液  | 4 ul    | 5 x      | 1 x    |
|        | T4 DNA 连接酶   | 1 ul    | 2000U/ul | 2000 U |
|        | NF 水   | 14.1 ul |          |        |
|        | 总计   | 20 ul   |          |        |

[0583] 表23

| [0584] | 循环序号 | 步骤 | 温度 (°C) | 时间      |
|--------|------|----|---------|---------|
|        | 1    | 连接 | 16      | 0:30:00 |

[0585] 表24

[0586] 脱帽

[0587] 在非-RNA多核苷酸连接到RNA链 (SEQ ID NO:30) 的5' 末端之前,所述RNA链具有通过5' 至5' 三磷酸连接在其5' 末端连接的7-甲基鸟苷帽,该帽通过RNA-5'-焦磷酸水解酶 (RppH) 酶促地去除。通过通过将下表25列出的试剂进行混合并将该混合物放置在如下表26所设程序的热循环仪上来完成脱帽。然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化获得的反应混合物并在12ul的NF水中洗脱。

| [0588] | 试剂                   | 体积   | 储液的浓度      | 最终浓度     |
|--------|----------------------|------|------------|----------|
|        | 酵母聚A+mRNA (Clontech) | 16ul | 28ng ug/ul | 500ng/反应 |
|        | NEB2缓冲液              | 2ul  | 10x        | 1x       |
|        | RNA 5'焦磷酸水解酶         | 2ul  | 5U/ul      | 10U      |
|        |                      |      |            |          |
|        | 总计                   | 20ul |            |          |

[0589] 表25

| [0590] | 循环序号 | 步骤 | 温度 (°C) | 时间 |
|--------|------|----|---------|----|
|        |      |    |         |    |

|   |    |    |         |
|---|----|----|---------|
| 1 | 脱帽 | 37 | 1:00:00 |
|---|----|----|---------|

[0591] 表26

[0592] 反转录

[0593] 脱帽之后,使用使用生命技术超级脚本 (Life Technologies Super Script) II实施反转录:根据制造商的程序混合下表27中的试剂并放在按照表28中设置的程序的热循环仪上。

[0594]

| 试剂                   | 体积   | 储液的浓度    | 最终浓度     |
|----------------------|------|----------|----------|
| 反转录后的3B (DNA/RNA 3B) | 12ul | ~20ng/ul | 240ng/反应 |
| 脱氧核糖核苷三磷酸 (dNTP)     | 1ul  | 各10uM    | 0.5uM    |
| 第一链缓冲液               | 4ul  | 5x       | 1x       |
| 0.1M二硫苏糖醇 (DTT)      | 2ul  | 0.1M     | 0.01M    |
| 超级脚本II               | 1ul  | 200U/ul  | 200U     |
| 总计                   | 20ul |          |          |

[0595] 表27

[0596]

| 循环序号 | 步骤  | 温度 (°C) | 时间      |
|------|-----|---------|---------|
| 1    | 反转录 | 42      | 0:50:00 |
| 2    | 变性  | 70      | 0:15:00 |

[0597] 表28

[0598] 然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化所述混合物并在10ul的NF水中洗脱。

[0599] 非-RNA多核苷酸连接

[0600] 通过将表29列出的试剂进行混合并将该混合物放置在如表30所设程序的热循环仪上来将所述非-RNA多核苷酸(30个SpC3间隔器连接至SEQ ID NO:31的5'末端,SEQ ID NO:31的3'末端连接4个iSp18间隔器,所述4个iSp18间隔器在其相对末端连接SEQ ID NO:32的5'末端,SEQ ID NO:32的3'末端连接到4个5-硝基吡啶,该4个5-硝基吡啶在其相对末端连接RNA序列CAAGGG)连接到反转录的mRNA。在连接非-RNA多核苷酸后制备的样品被称为样品7A。

[0601] A 5' 前导区用以下试剂进行连接

[0602]

| 试剂                 | 体积      | 储液的浓度 | 最终浓度 |
|--------------------|---------|-------|------|
| 反转录的 mRNA          | 10.0 ul |       |      |
| 非-RNA 多核苷酸 (以上描述的) | 1.2 ul  | 50 uM | 2 ul |

|        |               |         |         |      |
|--------|---------------|---------|---------|------|
| [0603] | T4 RNA 连接酶缓冲液 | 3.0 ul  | 10x     | 1x   |
|        | 三磷酸腺苷 (ATP)   | 0.6 ul  | 50 mM   | 1 mM |
|        | 8k 聚乙二醇 (PEG) | 12.0 ul | 50%     |      |
|        | T4 RNA 连接酶 1  | 3.2 ul  | 10 U/ul |      |
|        |               |         |         |      |
|        |               |         |         |      |
|        | 总计            | 30.0 ul |         |      |

[0604] 表29

|        |      |    |         |         |
|--------|------|----|---------|---------|
| [0605] | 循环序号 | 步骤 | 温度 (°C) | 时间      |
|        | 1    | 连接 | 25      | 0:30:00 |

[0606] 表30

[0607] 发夹纯化

[0608] 通过序列特异性系链和Lifetech MyOne C1链霉亲和素珠从连接酶和未反应的组分中纯化样品EX。

[0609] 通过将1.25ul, 100uM系链 (/5脱硫生物素/TT/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/(SEQ ID NO:65)/iSp18//iSp18//iSp18//iSp18//iSp18//iSp18/TT/3CholTEG/) 和30ul样品7A和8.75ul NF水混合并在室温下将该混合物孵育15分钟来将样品7A杂交到系链。

[0610] 使用磁道, 用200ul制造商指定的1x连接和清洗缓冲液 (B&W缓冲液) 清洗20ul MyOne C1链霉亲和素珠, 然后在40ul 2x B&W缓冲液中重悬。

[0611] Lifetech指定的2x B&W缓冲液:

[0612] 10mM三羟甲基氨基甲烷盐酸盐 (TRIS-HCl) pH 7.5

[0613] 1mM乙二胺四乙酸 (EDTA)

[0614] 2M NaCl

[0615] 将40ul系链的 (tethered Sample) 样品7A添加到40ul链霉亲和素珠, 并在混合仪上孵育15分钟。该溶液然后放置在磁道上并用制造商说明的1x B&W缓冲液清洗两次。

[0616] 通过将15uL的133uM生物素水溶液添加到磁珠并加热至37°C保持10分钟, 使样品从该链霉亲和素珠中洗脱。快速地将管子放置在磁道上, 将上层清液从磁珠中去除。此产物是纯化的样品, 并被称为样品7B。

[0617] 电生理学

[0618] 通过混合3ul的 (1uM系链寡核苷酸 (tethering oligo) (SEQ ID NO:66/iSp18//iSp18//iSp18//iSp18//iSp18/TT/3CholTEG/), 750mM KCl, 125mM KPhos缓冲液pH 7和5mM EDTA), 并在室温下培养混合物20分钟, 从而将样品7B (9ul) 杂交到该系链寡核苷酸。

[0619] 然后该系链的样品7B用2ul 17.4uM T4Dda (E94C/F98W/C109A/C136A/K194L/A360C (SEQ ID NO: 8具有突变E94C/F98W/C109A/C136A/K194L/A360C然后 (ΔM1) G1)) 孵育15分钟。然后将2.1ul 800uM TMAD添加到该孵育的混合物中, 并在室温下保持10分钟。该样品稀释到缓冲液 (282uL 500mM KCl, 25mM磷酸钾pH 8.0) 中, 并与2ul的 (70uM ATP和75uM



MgCl<sub>2</sub>) 混合。

[0620] 在缓冲液 (25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾~pH 8.0) 中, 从插入到嵌段共聚物的单MspA纳米孔获得电测量值。获得插入到嵌段共聚物的单孔后, 然后让缓冲液 (2mL, 25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾, pH 8.0) 流过系统以去除任何过量的MspA纳米孔。

[0621] 在添加样品之前, 让过量的缓冲液 (500mM KCl, 25mM磷酸钾pH 8.0) 流过系统。将与样品3A连接的T4Dda-E94C/A360C添加到该纳米孔系统, 在-120mV电压下运行本实验, 并且监控解旋酶控制的DNA移动。

[0622] 结果

[0623] 图19显示当添加样品7B时, 记录的来自单纳米孔的离子电流。该电生理学实验显示对于具有不同长度 (对应于不同天然细胞mRNA转录物长度) 的2D链的良好通量。图19也说明了RNA与DNA相比具有不同的平均振幅和幅度。因此, 作为平均振幅和幅度的函数, 甚至RNA和DNA序列相同时, 也可彼此区分RNA和DNA。

[0624] 实施例8

[0625] 本实施例显示包含DNA的前导区如何连接到信使RNA (mRNA) 以促进加载DNA解旋酶, 即Hel308Mbu-E284C/S615C (具有突变E284C/S615C的SEQ ID NO:8), 并随后观察解旋酶控制RNA穿过纳米孔的移动。从Trilink Biotech购买1.9kb mRNA。该包含DNA的前导区与该mRNA的3' 末端连接。然后将Hel308加载到前导区中的DNA连接位点, 通过纳米孔分析底物。

[0626] 材料与方法

[0627] 与几个寡核苷酸 (SEQ ID NO:35 (其具有5' 磷酸) 预先退火的包含DNA的前导区在其3' 末端连接到4个iSp18间隔器, 该4个iSp18间隔器在其相对末端连接SEQ ID NO:36的5' 末端, SEQ ID NO:36在其5' 末端连接到30个iSpC3间隔器, 其用SEQ ID NO:37, 38和39 (其具有3' 胆固醇TEG) 进行退火, 通过混合表31列出的试剂并将该混合物放置在如表32所设程序的热循环仪上来将该预先退火的前导区与mRNA连接。

| [0628] | 试剂  | 体积      | 储液的浓度       | 最终浓度       |
|--------|---|---------|-------------|------------|
|        | FLuc mRNA (RNA X1, 具有 SEQ ID NO: 26 的开放阅读框的 mRNA, 所述 SEQ ID NO: 26 具有 5'-己基-G 作为大多数 5'核苷酸并具有 3'聚 A 尾) | 2.5 ul  | 1 ug/ul     | 2500 ng/反应 |
|        | 预先退火 1:1:1:1 的前导区   | 3.5 ul  | 各 2 uM      | 1x         |
|        | NEB Quick 连接酶缓冲液 (Lifetech)   | 4.0 ul  | 5 x         | 1x         |
|        | NEB Quick 连接酶   | 2.0 ul  | 2, 000 U/ul | 4, 000 U   |
|        | 夹板 (SEQ ID NO: 40)  | 0.2 ul  | 100 uM      | 1 uM       |
|        | 水   | 8 ul    |             |            |
|        | 总计  | 20.2 ul |             |            |

[0629] 表31

| 循环序号 | 步骤 | 温度(℃) | 时间      |
|------|----|-------|---------|
| 1    | 连接 | 16    | 0:30:00 |

[0631] 表32

[0632] 然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化该混合物。

[0633] 反转录

[0634] 通过混合表33中的试剂并将该混合物放置在如表34所设程序的热循环仪上,使用超级脚本II试剂盒(SuperScript II kit)将该样品反转录。

| 试剂                       | 体积     | 储液的浓度   | 最终浓度   |
|--------------------------|--------|---------|--------|
| 来自连接的样品                  | 2.2 ul |         | 500 ng |
| 脱氧核糖核苷三磷酸                | 1 ul   | 各 10 mM |        |
| [0635] 第一链缓冲液 (Lifetech) | 4 ul   | 5 x     | 1 x    |
| 二硫苏糖醇                    | 2 ul   | 0.1 M   |        |
| SSII 反转录酶                | 1 ul   |         |        |
| NF 水                     | 9.8 ul |         |        |
|                          |        |         |        |
| [0636] 总计                | 20 ul  |         |        |

[0637] 表33

| 循环序号 | 步骤  | 温度(℃) | 时间      |
|------|-----|-------|---------|
| 1    | 反转录 | 42    | 0:50:00 |
| 2    | 变性  | 70    | 0:15:00 |

[0639] 表34

[0640] 然后使用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例纯化该混合物。使用7kda Zeba柱将Hel308Mbu-E284C/S615C(具有突变E284C/S615C的SEQ ID NO:8)缓冲交换进入50mM HEPES pH8,100mM KAc。该反转录的RNA样品与等体积的100mM HEPES pH8,200mM KAc混合,然后以1:100的摩尔比与Hel308Mbu-E284C/S615C混合,并缓冲交换进入100mM KPhos pH 8,100mM NaCl,5mM EDTA,0.1%TWEEN。将20mM BMOE的DMF溶液加入到100mM KPhos pH 8,100mM NaCl,5mM EDTA,0.1%TWEEN以达到5mM的浓度。该5mM BMOE溶液加入到RNA和Hel308的混合物中以达到最终浓度为40uM BMOE。该溶液在室温下孵育2小时。然后该溶液与Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例连接。代替用EtOH混合物清洗该SPRI珠,使用具有20%PEG,2.5M NaCl,50mM Tris的修饰的清洗缓冲液的清洗单次。该样品在30ul 500mM KCl 25mM,pH 8下被洗脱。这被称为样品6。

[0641] 电生理学

[0642] 将样品6(4ul)与缓冲液(500mM KCl 25mM KPhosphat pH 8,2mM ATP,2mM MgCl<sub>2</sub>,295ul)混合。

[0643] 在缓冲液 (25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾~pH 8.0) 中, 从插入到嵌段共聚物的单MspA纳米孔获得电测量值。获得插入到嵌段共聚物的单孔后, 然后让缓冲液 (2mL, 25mM K磷酸盐缓冲液, 150mM亚铁氰化钾, 150mM铁氰化钾, pH 8.0) 流过系统以去除任何过量的MspA纳米孔。

[0644] 在添加样品6之前, 让过量的缓冲液 (500mM KCl, 25mM磷酸钾pH 8.0) 流过系统。最后, 然后将样品6添加到该纳米孔系统, 在-140mV电压下运行本实验, 并且监控解旋酶控制的DNA移动。

[0645] 结果

[0646] 本实施例显示包含DNA的前导区如何连接到信使RNA (mRNA) 以促进加载DNA解旋酶, 即He1308Mbu-E284C/S615C (具有突变E284C/S615C的SEQ ID NO:8), 并随后观察解旋酶控制RNA的移动。解旋酶控制的RNA的移动的实例如图20所示。

- [0001] 序列表
- [0002] <110> 牛津纳米孔技术公司
- [0003] <120> 纳米孔RNA表征方法
- [0004] <130> N404137W0
- [0005] <150> GB1418459.2
- [0006] <151> 2014-10-17
- [0007] <150> GB1508270.4
- [0008] <151> 2015-05-14
- [0009] <150> GB1517634.0
- [0010] <151> 2015-10-06
- [0011] <160> 66
- [0012] <170> 专利版本 3.5
- [0013] <210> 1
- [0014] <211> 558
- [0015] <212> DNA
- [0016] <213> 人工序列
- [0017] <220>
- [0018] <223> 耻垢分枝杆菌孔蛋白A突变体
- [0019] (D90N/D91N/D93N/D118R/D134R/E193K)
- [0020] <400> 1
- [0021] atgggtctgg ataatgaact gagcctggtg gacggtcaag atcgtagcct gacggtgcaa 60
- [0022] caatgggata cctttctgaa tggcgttttt ccgctggatc gtaatgcct gacccgtgaa 120
- [0023] tggtttcatt ccggtcgcgc aaaatatatc gtcgcaggcc cgggtgctga cgaattcgaa 180
- [0024] ggcacgctgg aactgggtta tcagattggc tttccgtggt cactgggcgt tggatatcaac 240
- [0025] ttctcgta caacgccgaa tattctgac aacaatggta acattaccgc accgccgttt 300
- [0026] ggctgaaca gcgtgattac gccgaacctg tttccgggtg ttagcatctc tgcccgctctg 360
- [0027] ggcaatggtc cgggcattca agaagtggca acctttagtg tgcgcgtttc cggcgctaaa 420
- [0028] ggcggtgtcg cgggtgtctaa cgcccacggt accgttacgg gcgcggccgg cgggtgtcctg 480
- [0029] ctgcgtccgt tcgcgcgcct gattgcctct accggcgaca gcgttacgac ctatggcgaa 540
- [0030] ccgtggaata tgaactaa 558
- [0031] <210> 2
- [0032] <211> 184
- [0033] <212> PRT
- [0034] <213> 人工序列
- [0035] <220>
- [0036] <223> 耻垢分枝杆菌孔蛋白A突变体
- [0037] (D90N/D91N/D93N/D118R/D134R/E139K)
- [0038] <400> 2
- [0039] Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
- [0040] 1 5 10 15
- [0041] Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp

|        |  |     |     |
|--------|--|-----|-----|
| [0042] | 20   | 25  | 30  |
| [0043] | Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr        |     |     |
| [0044] | 35   | 40  | 45  |
| [0045] | Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu        |     |     |
| [0046] | 50   | 55  | 60  |
| [0047] | Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe        |     |     |
| [0048] | 65   | 70  | 75  |
| [0049] | Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asn Asn Gly Asn Ile Thr Ala        |     |     |
| [0050] | 85   | 90  | 95  |
| [0051] | Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly        |     |     |
| [0052] | 100  | 105 | 110 |
| [0053] | Val Ser Ile Ser Ala Arg Leu Gly Asn Gly Pro Gly Ile Gln Glu Val        |     |     |
| [0054] | 115  | 120 | 125 |
| [0055] | Ala Thr Phe Ser Val Arg Val Ser Gly Ala Lys Gly Gly Val Ala Val        |     |     |
| [0056] | 130  | 135 | 140 |
| [0057] | Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu        |     |     |
| [0058] | 145  | 150 | 155 |
| [0059] | Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr        |     |     |
| [0060] | 165  | 170 | 175 |
| [0061] | Tyr Gly Glu Pro Trp Asn Met Asn  |     |     |
| [0062] | 180  |     |     |
| [0063] | <210> 3  |     |     |
| [0064] | <211> 885  |     |     |
| [0065] | <212> DNA  |     |     |
| [0066] | <213> 人工序列   |     |     |
| [0067] | <220>  |     |     |
| [0068] | <223> $\alpha$ -溶血素突变体(alpha-hemolysin mutant) (E111N/K147N)           |     |     |
| [0069] | <400> 3  |     |     |
| [0070] | atggcagatt ctgatattaa tattaacc ggtactacag atattggaag caatactaca 60     |     |     |
| [0071] | gtaaaacag gtgatttagt cacttatgat aaagaaaatg gcatgcacaa aaaagtattt 120   |     |     |
| [0072] | tatagtttta tcgatgataa aaatcacaat aaaaaactgc tagttattag aacaaaaggt 180  |     |     |
| [0073] | accattgctg gtcaatatag agtttatagc gaagaaggtg ctaacaaaag tggtttagcc 240  |     |     |
| [0074] | tggccttcag cctttaaggt acagttgcaa ctacctgata atgaagtagc tcaaatatct 300  |     |     |
| [0075] | gattactatc caagaaattc gattgataca aaaaactata tgagtacttt aacttatgga 360  |     |     |
| [0076] | ttcaacggta atgttactgg tgatgataca ggaaaaattg gcggccttat tgggtgcaaat 420 |     |     |
| [0077] | gtttcgattg gtcatacact gaactatggt caacctgatt tcaaaacaat ttagagagc 480   |     |     |
| [0078] | ccaactgata aaaaagtagg ctggaaagtg atatttaaca atatggtgaa tcaaaattgg 540  |     |     |
| [0079] | ggaccatacg atcgagattc ttggaaccgg gtatatggca atcaactttt catgaaaact 600  |     |     |
| [0080] | agaaatgggt ctatgaaagc agcagataac ttccttgatc ctaacaaagc aagttctcta 660  |     |     |
| [0081] | ttatcttcag ggttttcacc agacttcgct acagttatta ctatggatag aaaagcatcc 720  |     |     |
| [0082] | aaacaacaaa caaatataga tgtaatatag gaacgagttc gtgatgatta ccaattgcat 780  |     |     |
| [0083] | tggacttcaa caaattggaa aggtaccaat actaaagata aatggacaga tcgttcttca 840  |     |     |

[0084] gaaagatata aaatcgattg ggaaaaagaa gaaatgacaa attaa 885  
 [0085] <210> 4  
 [0086] <211> 293  
 [0087] <212> PRT  
 [0088] <213> 人工序列  
 [0089] <220>  
 [0090] <223>  $\alpha$ -溶血素突变体 (E111N/K147N)  
 [0091] <400> 4  
 [0092] Ala Asp Ser Asp Ile Asn Ile Lys Thr Gly Thr Thr Asp Ile Gly Ser  
 [0093] 1 5 10 15  
 [0094] Asn Thr Thr Val Lys Thr Gly Asp Leu Val Thr Tyr Asp Lys Glu Asn  
 [0095] 20 25 30  
 [0096] Gly Met His Lys Lys Val Phe Tyr Ser Phe Ile Asp Asp Lys Asn His  
 [0097] 35 40 45  
 [0098] Asn Lys Lys Leu Leu Val Ile Arg Thr Lys Gly Thr Ile Ala Gly Gln  
 [0099] 50 55 60  
 [0100] Tyr Arg Val Tyr Ser Glu Glu Gly Ala Asn Lys Ser Gly Leu Ala Trp  
 [0101] 65 70 75 80  
 [0102] Pro Ser Ala Phe Lys Val Gln Leu Gln Leu Pro Asp Asn Glu Val Ala  
 [0103] 85 90 95  
 [0104] Gln Ile Ser Asp Tyr Tyr Pro Arg Asn Ser Ile Asp Thr Lys Asn Tyr  
 [0105] 100 105 110  
 [0106] Met Ser Thr Leu Thr Tyr Gly Phe Asn Gly Asn Val Thr Gly Asp Asp  
 [0107] 115 120 125  
 [0108] Thr Gly Lys Ile Gly Gly Leu Ile Gly Ala Asn Val Ser Ile Gly His  
 [0109] 130 135 140  
 [0110] Thr Leu Asn Tyr Val Gln Pro Asp Phe Lys Thr Ile Leu Glu Ser Pro  
 [0111] 145 150 155 160  
 [0112] Thr Asp Lys Lys Val Gly Trp Lys Val Ile Phe Asn Asn Met Val Asn  
 [0113] 165 170 175  
 [0114] Gln Asn Trp Gly Pro Tyr Asp Arg Asp Ser Trp Asn Pro Val Tyr Gly  
 [0115] 180 185 190  
 [0116] Asn Gln Leu Phe Met Lys Thr Arg Asn Gly Ser Met Lys Ala Ala Asp  
 [0117] 195 200 205  
 [0118] Asn Phe Leu Asp Pro Asn Lys Ala Ser Ser Leu Leu Ser Ser Gly Phe  
 [0119] 210 215 220  
 [0120] Ser Pro Asp Phe Ala Thr Val Ile Thr Met Asp Arg Lys Ala Ser Lys  
 [0121] 225 230 235 240  
 [0122] Gln Gln Thr Asn Ile Asp Val Ile Tyr Glu Arg Val Arg Asp Asp Tyr  
 [0123] 245 250 255  
 [0124] Gln Leu His Trp Thr Ser Thr Asn Trp Lys Gly Thr Asn Thr Lys Asp  
 [0125] 260 265 270

|        |   |
|--------|---|
| [0126] | Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys |
| [0127] | 275 280 285   |
| [0128] | Glu Glu Met Thr Asn   |
| [0129] | 290   |
| [0130] | <210> 5   |
| [0131] | <211> 184   |
| [0132] | <212> PRT   |
| [0133] | <213> 耻垢分枝杆菌  |
| [0134] | <400> 5   |
| [0135] | Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu |
| [0136] | 1 5 10 15   |
| [0137] | Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp |
| [0138] | 20 25 30  |
| [0139] | Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr |
| [0140] | 35 40 45  |
| [0141] | Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu |
| [0142] | 50 55 60  |
| [0143] | Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe |
| [0144] | 65 70 75 80   |
| [0145] | Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala |
| [0146] | 85 90 95  |
| [0147] | Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly |
| [0148] | 100 105 110   |
| [0149] | Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val |
| [0150] | 115 120 125   |
| [0151] | Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val |
| [0152] | 130 135 140   |
| [0153] | Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu |
| [0154] | 145 150 155 160   |
| [0155] | Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr |
| [0156] | 165 170 175   |
| [0157] | Tyr Gly Glu Pro Trp Asn Met Asn                                 |
| [0158] | 180   |
| [0159] | <210> 6   |
| [0160] | <211> 184   |
| [0161] | <212> PRT   |
| [0162] | <213> 耻垢分枝杆菌  |
| [0163] | <400> 6   |
| [0164] | Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu |
| [0165] | 1 5 10 15   |
| [0166] | Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp |
| [0167] | 20 25 30  |

|        |   |
|--------|---|
| [0168] | Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr |
| [0169] | 35 40 45  |
| [0170] | Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu |
| [0171] | 50 55 60  |
| [0172] | Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe |
| [0173] | 65 70 75 80   |
| [0174] | Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly |
| [0175] | 85 90 95  |
| [0176] | Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly |
| [0177] | 100 105 110   |
| [0178] | Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val |
| [0179] | 115 120 125   |
| [0180] | Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val |
| [0181] | 130 135 140   |
| [0182] | Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu |
| [0183] | 145 150 155 160   |
| [0184] | Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr |
| [0185] | 165 170 175   |
| [0186] | Tyr Gly Glu Pro Trp Asn Met Asn                                 |
| [0187] | 180   |
| [0188] | <210> 7   |
| [0189] | <211> 183   |
| [0190] | <212> PRT   |
| [0191] | <213> 耻垢分枝杆菌  |
| [0192] | <400> 7   |
| [0193] | Val Asp Asn Gln Leu Ser Val Val Asp Gly Gln Gly Arg Thr Leu Thr |
| [0194] | 1 5 10 15   |
| [0195] | Val Gln Gln Ala Glu Thr Phe Leu Asn Gly Val Phe Pro Leu Asp Arg |
| [0196] | 20 25 30  |
| [0197] | Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Thr Tyr His |
| [0198] | 35 40 45  |
| [0199] | Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu Gly |
| [0200] | 50 55 60  |
| [0201] | Tyr Gln Val Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe Ser |
| [0202] | 65 70 75 80   |
| [0203] | Tyr Thr Thr Pro Asn Ile Leu Ile Asp Gly Gly Asp Ile Thr Gln Pro |
| [0204] | 85 90 95  |
| [0205] | Pro Phe Gly Leu Asp Thr Ile Ile Thr Pro Asn Leu Phe Pro Gly Val |
| [0206] | 100 105 110   |
| [0207] | Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala |
| [0208] | 115 120 125   |
| [0209] | Thr Phe Ser Val Asp Val Lys Gly Ala Lys Gly Ala Val Ala Val Ser |



|        |   |     |     |
|--------|---|-----|-----|
| [0210] | 130   | 135 | 140 |
| [0211] | Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg |     |     |
| [0212] | 145   | 150 | 155 |
| [0213] | Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr |     |     |
| [0214] |   | 165 | 170 |
| [0215] | Gly Glu Pro Trp Asn Met Asn                                     |     | 175 |
| [0216] |   | 180 |     |
| [0217] | <210> 8   |     |     |
| [0218] | <211> 760   |     |     |
| [0219] | <212> PRT   |     |     |
| [0220] | <213> 伯顿拟甲烷球菌(Methanococcoides burtonii)                        |     |     |
| [0221] | <400> 8   |     |     |
| [0222] | Met Met Ile Arg Glu Leu Asp Ile Pro Arg Asp Ile Ile Gly Phe Tyr |     |     |
| [0223] | 1   | 5   | 10  |
| [0224] | Glu Asp Ser Gly Ile Lys Glu Leu Tyr Pro Pro Gln Ala Glu Ala Ile |     | 15  |
| [0225] |   | 20  | 25  |
| [0226] | Glu Met Gly Leu Leu Glu Lys Lys Asn Leu Leu Ala Ala Ile Pro Thr |     | 30  |
| [0227] |   | 35  | 40  |
| [0228] | Ala Ser Gly Lys Thr Leu Leu Ala Glu Leu Ala Met Ile Lys Ala Ile |     | 45  |
| [0229] |   | 50  | 55  |
| [0230] | Arg Glu Gly Gly Lys Ala Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala |     | 60  |
| [0231] | 65  | 70  | 75  |
| [0232] | Ser Glu Lys Phe Glu Arg Phe Lys Glu Leu Ala Pro Phe Gly Ile Lys |     | 80  |
| [0233] |   | 85  | 90  |
| [0234] | Val Gly Ile Ser Thr Gly Asp Leu Asp Ser Arg Ala Asp Trp Leu Gly |     | 95  |
| [0235] |   | 100 | 105 |
| [0236] | Val Asn Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu |     | 110 |
| [0237] |   | 115 | 120 |
| [0238] | Arg Asn Gly Thr Ser Trp Met Asp Glu Ile Thr Thr Val Val Val Asp |     | 125 |
| [0239] |   | 130 | 135 |
| [0240] | Glu Ile His Leu Leu Asp Ser Lys Asn Arg Gly Pro Thr Leu Glu Val |     | 140 |
| [0241] | 145   | 150 | 155 |
| [0242] | Thr Ile Thr Lys Leu Met Arg Leu Asn Pro Asp Val Gln Val Val Ala |     | 160 |
| [0243] |   | 165 | 170 |
| [0244] | Leu Ser Ala Thr Val Gly Asn Ala Arg Glu Met Ala Asp Trp Leu Gly |     | 175 |
| [0245] |   | 180 | 185 |
| [0246] | Ala Ala Leu Val Leu Ser Glu Trp Arg Pro Thr Asp Leu His Glu Gly |     | 190 |
| [0247] |   | 195 | 200 |
| [0248] | Val Leu Phe Gly Asp Ala Ile Asn Phe Pro Gly Ser Gln Lys Lys Ile |     | 205 |
| [0249] |   | 210 | 215 |
| [0250] | Asp Arg Leu Glu Lys Asp Asp Ala Val Asn Leu Val Leu Asp Thr Ile |     | 220 |
| [0251] | 225   | 230 | 235 |
|        |   |     | 240 |

|        |   |     |             |
|--------|---|-----|-------------|
| [0252] | Lys Ala Glu Gly Gln Cys Leu Val Phe Glu Ser Ser Arg Arg Asn Cys |     |             |
| [0253] |   | 245 | 250 255     |
| [0254] | Ala Gly Phe Ala Lys Thr Ala Ser Ser Lys Val Ala Lys Ile Leu Asp |     |             |
| [0255] |   | 260 | 265 270     |
| [0256] | Asn Asp Ile Met Ile Lys Leu Ala Gly Ile Ala Glu Glu Val Glu Ser |     |             |
| [0257] |   | 275 | 280 285     |
| [0258] | Thr Gly Glu Thr Asp Thr Ala Ile Val Leu Ala Asn Cys Ile Arg Lys |     |             |
| [0259] |   | 290 | 295 300     |
| [0260] | Gly Val Ala Phe His His Ala Gly Leu Asn Ser Asn His Arg Lys Leu |     |             |
| [0261] |   | 305 | 310 315 320 |
| [0262] | Val Glu Asn Gly Phe Arg Gln Asn Leu Ile Lys Val Ile Ser Ser Thr |     |             |
| [0263] |   | 325 | 330 335     |
| [0264] | Pro Thr Leu Ala Ala Gly Leu Asn Leu Pro Ala Arg Arg Val Ile Ile |     |             |
| [0265] |   | 340 | 345 350     |
| [0266] | Arg Ser Tyr Arg Arg Phe Asp Ser Asn Phe Gly Met Gln Pro Ile Pro |     |             |
| [0267] |   | 355 | 360 365     |
| [0268] | Val Leu Glu Tyr Lys Gln Met Ala Gly Arg Ala Gly Arg Pro His Leu |     |             |
| [0269] |   | 370 | 375 380     |
| [0270] | Asp Pro Tyr Gly Glu Ser Val Leu Leu Ala Lys Thr Tyr Asp Glu Phe |     |             |
| [0271] |   | 385 | 390 395 400 |
| [0272] | Ala Gln Leu Met Glu Asn Tyr Val Glu Ala Asp Ala Glu Asp Ile Trp |     |             |
| [0273] |   | 405 | 410 415     |
| [0274] | Ser Lys Leu Gly Thr Glu Asn Ala Leu Arg Thr His Val Leu Ser Thr |     |             |
| [0275] |   | 420 | 425 430     |
| [0276] | Ile Val Asn Gly Phe Ala Ser Thr Arg Gln Glu Leu Phe Asp Phe Phe |     |             |
| [0277] |   | 435 | 440 445     |
| [0278] | Gly Ala Thr Phe Phe Ala Tyr Gln Gln Asp Lys Trp Met Leu Glu Glu |     |             |
| [0279] |   | 450 | 455 460     |
| [0280] | Val Ile Asn Asp Cys Leu Glu Phe Leu Ile Asp Lys Ala Met Val Ser |     |             |
| [0281] |   | 465 | 470 475 480 |
| [0282] | Glu Thr Glu Asp Ile Glu Asp Ala Ser Lys Leu Phe Leu Arg Gly Thr |     |             |
| [0283] |   | 485 | 490 495     |
| [0284] | Arg Leu Gly Ser Leu Val Ser Met Leu Tyr Ile Asp Pro Leu Ser Gly |     |             |
| [0285] |   | 500 | 505 510     |
| [0286] | Ser Lys Ile Val Asp Gly Phe Lys Asp Ile Gly Lys Ser Thr Gly Gly |     |             |
| [0287] |   | 515 | 520 525     |
| [0288] | Asn Met Gly Ser Leu Glu Asp Asp Lys Gly Asp Asp Ile Thr Val Thr |     |             |
| [0289] |   | 530 | 535 540     |
| [0290] | Asp Met Thr Leu Leu His Leu Val Cys Ser Thr Pro Asp Met Arg Gln |     |             |
| [0291] |   | 545 | 550 555 560 |
| [0292] | Leu Tyr Leu Arg Asn Thr Asp Tyr Thr Ile Val Asn Glu Tyr Ile Val |     |             |
| [0293] |   | 565 | 570 575     |

|        |   |
|--------|---|
| [0294] | Ala His Ser Asp Glu Phe His Glu Ile Pro Asp Lys Leu Lys Glu Thr |
| [0295] | 580 585 590   |
| [0296] | Asp Tyr Glu Trp Phe Met Gly Glu Val Lys Thr Ala Met Leu Leu Glu |
| [0297] | 595 600 605   |
| [0298] | Glu Trp Val Thr Glu Val Ser Ala Glu Asp Ile Thr Arg His Phe Asn |
| [0299] | 610 615 620   |
| [0300] | Val Gly Glu Gly Asp Ile His Ala Leu Ala Asp Thr Ser Glu Trp Leu |
| [0301] | 625 630 635 640   |
| [0302] | Met His Ala Ala Ala Lys Leu Ala Glu Leu Leu Gly Val Glu Tyr Ser |
| [0303] | 645 650 655   |
| [0304] | Ser His Ala Tyr Ser Leu Glu Lys Arg Ile Arg Tyr Gly Ser Gly Leu |
| [0305] | 660 665 670   |
| [0306] | Asp Leu Met Glu Leu Val Gly Ile Arg Gly Val Gly Arg Val Arg Ala |
| [0307] | 675 680 685   |
| [0308] | Arg Lys Leu Tyr Asn Ala Gly Phe Val Ser Val Ala Lys Leu Lys Gly |
| [0309] | 690 695 700   |
| [0310] | Ala Asp Ile Ser Val Leu Ser Lys Leu Val Gly Pro Lys Val Ala Tyr |
| [0311] | 705 710 715 720   |
| [0312] | Asn Ile Leu Ser Gly Ile Gly Val Arg Val Asn Asp Lys His Phe Asn |
| [0313] | 725 730 735   |
| [0314] | Ser Ala Pro Ile Ser Ser Asn Thr Leu Asp Thr Leu Leu Asp Lys Asn |
| [0315] | 740 745 750   |
| [0316] | Gln Lys Thr Phe Asn Asp Phe Gln                                 |
| [0317] | 755 760   |
| [0318] | <210> 9   |
| [0319] | <211> 707   |
| [0320] | <212> PRT   |
| [0321] | <213> 共生餐古菌(Cenarchaeum symbiosum)                              |
| [0322] | <400> 9   |
| [0323] | Met Arg Ile Ser Glu Leu Asp Ile Pro Arg Pro Ala Ile Glu Phe Leu |
| [0324] | 1 5 10 15   |
| [0325] | Glu Gly Glu Gly Tyr Lys Lys Leu Tyr Pro Pro Gln Ala Ala Ala Ala |
| [0326] | 20 25 30  |
| [0327] | Lys Ala Gly Leu Thr Asp Gly Lys Ser Val Leu Val Ser Ala Pro Thr |
| [0328] | 35 40 45  |
| [0329] | Ala Ser Gly Lys Thr Leu Ile Ala Ala Ile Ala Met Ile Ser His Leu |
| [0330] | 50 55 60  |
| [0331] | Ser Arg Asn Arg Gly Lys Ala Val Tyr Leu Ser Pro Leu Arg Ala Leu |
| [0332] | 65 70 75 80   |
| [0333] | Ala Ala Glu Lys Phe Ala Glu Phe Gly Lys Ile Gly Gly Ile Pro Leu |
| [0334] | 85 90 95  |
| [0335] | Gly Arg Pro Val Arg Val Gly Val Ser Thr Gly Asp Phe Glu Lys Ala |

|        |   |     |     |
|--------|---|-----|-----|
| [0336] | 100   | 105 | 110 |
| [0337] | Gly Arg Ser Leu Gly Asn Asn Asp Ile Leu Val Leu Thr Asn Glu Arg |     |     |
| [0338] | 115   | 120 | 125 |
| [0339] | Met Asp Ser Leu Ile Arg Arg Arg Pro Asp Trp Met Asp Glu Val Gly |     |     |
| [0340] | 130   | 135 | 140 |
| [0341] | Leu Val Ile Ala Asp Glu Ile His Leu Ile Gly Asp Arg Ser Arg Gly |     |     |
| [0342] | 145   | 150 | 155 |
| [0343] | Pro Thr Leu Glu Met Val Leu Thr Lys Leu Arg Gly Leu Arg Ser Ser |     |     |
| [0344] | 165   | 170 | 175 |
| [0345] | Pro Gln Val Val Ala Leu Ser Ala Thr Ile Ser Asn Ala Asp Glu Ile |     |     |
| [0346] | 180   | 185 | 190 |
| [0347] | Ala Gly Trp Leu Asp Cys Thr Leu Val His Ser Thr Trp Arg Pro Val |     |     |
| [0348] | 195   | 200 | 205 |
| [0349] | Pro Leu Ser Glu Gly Val Tyr Gln Asp Gly Glu Val Ala Met Gly Asp |     |     |
| [0350] | 210   | 215 | 220 |
| [0351] | Gly Ser Arg His Glu Val Ala Ala Thr Gly Gly Gly Pro Ala Val Asp |     |     |
| [0352] | 225   | 230 | 235 |
| [0353] | Leu Ala Ala Glu Ser Val Ala Glu Gly Gly Gln Ser Leu Ile Phe Ala |     |     |
| [0354] | 245   | 250 | 255 |
| [0355] | Asp Thr Arg Ala Arg Ser Ala Ser Leu Ala Ala Lys Ala Ser Ala Val |     |     |
| [0356] | 260   | 265 | 270 |
| [0357] | Ile Pro Glu Ala Lys Gly Ala Asp Ala Ala Lys Leu Ala Ala Ala Ala |     |     |
| [0358] | 275   | 280 | 285 |
| [0359] | Lys Lys Ile Ile Ser Ser Gly Gly Glu Thr Lys Leu Ala Lys Thr Leu |     |     |
| [0360] | 290   | 295 | 300 |
| [0361] | Ala Glu Leu Val Glu Lys Gly Ala Ala Phe His His Ala Gly Leu Asn |     |     |
| [0362] | 305   | 310 | 315 |
| [0363] | Gln Asp Cys Arg Ser Val Val Glu Glu Glu Phe Arg Ser Gly Arg Ile |     |     |
| [0364] | 325   | 330 | 335 |
| [0365] | Arg Leu Leu Ala Ser Thr Pro Thr Leu Ala Ala Gly Val Asn Leu Pro |     |     |
| [0366] | 340   | 345 | 350 |
| [0367] | Ala Arg Arg Val Val Ile Ser Ser Val Met Arg Tyr Asn Ser Ser Ser |     |     |
| [0368] | 355   | 360 | 365 |
| [0369] | Gly Met Ser Glu Pro Ile Ser Ile Leu Glu Tyr Lys Gln Leu Cys Gly |     |     |
| [0370] | 370   | 375 | 380 |
| [0371] | Arg Ala Gly Arg Pro Gln Tyr Asp Lys Ser Gly Glu Ala Ile Val Val |     |     |
| [0372] | 385   | 390 | 395 |
| [0373] | Gly Gly Val Asn Ala Asp Glu Ile Phe Asp Arg Tyr Ile Gly Gly Glu |     |     |
| [0374] | 405   | 410 | 415 |
| [0375] | Pro Glu Pro Ile Arg Ser Ala Met Val Asp Asp Arg Ala Leu Arg Ile |     |     |
| [0376] | 420   | 425 | 430 |
| [0377] | His Val Leu Ser Leu Val Thr Thr Ser Pro Gly Ile Lys Glu Asp Asp |     |     |

|        |   |     |     |
|--------|---|-----|-----|
| [0378] | 435   | 440 | 445 |
| [0379] | Val Thr Glu Phe Phe Leu Gly Thr Leu Gly Gly Gln Gln Ser Gly Glu |     |     |
| [0380] | 450   | 455 | 460 |
| [0381] | Ser Thr Val Lys Phe Ser Val Ala Val Ala Leu Arg Phe Leu Gln Glu |     |     |
| [0382] | 465   | 470 | 475 |
| [0383] | Glu Gly Met Leu Gly Arg Arg Gly Gly Arg Leu Ala Ala Thr Lys Met |     |     |
| [0384] | 485   | 490 | 495 |
| [0385] | Gly Arg Leu Val Ser Arg Leu Tyr Met Asp Pro Met Thr Ala Val Thr |     |     |
| [0386] | 500   | 505 | 510 |
| [0387] | Leu Arg Asp Ala Val Gly Glu Ala Ser Pro Gly Arg Met His Thr Leu |     |     |
| [0388] | 515   | 520 | 525 |
| [0389] | Gly Phe Leu His Leu Val Ser Glu Cys Ser Glu Phe Met Pro Arg Phe |     |     |
| [0390] | 530   | 535 | 540 |
| [0391] | Ala Leu Arg Gln Lys Asp His Glu Val Ala Glu Met Met Leu Glu Ala |     |     |
| [0392] | 545   | 550 | 555 |
| [0393] | Gly Arg Gly Glu Leu Leu Arg Pro Val Tyr Ser Tyr Glu Cys Gly Arg |     |     |
| [0394] | 565   | 570 | 575 |
| [0395] | Gly Leu Leu Ala Leu His Arg Trp Ile Gly Glu Ser Pro Glu Ala Lys |     |     |
| [0396] | 580   | 585 | 590 |
| [0397] | Leu Ala Glu Asp Leu Lys Phe Glu Ser Gly Asp Val His Arg Met Val |     |     |
| [0398] | 595   | 600 | 605 |
| [0399] | Glu Ser Ser Gly Trp Leu Leu Arg Cys Ile Trp Glu Ile Ser Lys His |     |     |
| [0400] | 610   | 615 | 620 |
| [0401] | Gln Glu Arg Pro Asp Leu Leu Gly Glu Leu Asp Val Leu Arg Ser Arg |     |     |
| [0402] | 625   | 630 | 635 |
| [0403] | Val Ala Tyr Gly Ile Lys Ala Glu Leu Val Pro Leu Val Ser Ile Lys |     |     |
| [0404] | 645   | 650 | 655 |
| [0405] | Gly Ile Gly Arg Val Arg Ser Arg Arg Leu Phe Arg Gly Gly Ile Lys |     |     |
| [0406] | 660   | 665 | 670 |
| [0407] | Gly Pro Gly Asp Leu Ala Ala Val Pro Val Glu Arg Leu Ser Arg Val |     |     |
| [0408] | 675   | 680 | 685 |
| [0409] | Glu Gly Ile Gly Ala Thr Leu Ala Asn Asn Ile Lys Ser Gln Leu Arg |     |     |
| [0410] | 690   | 695 | 700 |
| [0411] | Lys Gly Gly   |     |     |
| [0412] | 705   |     |     |
| [0413] | <210> 10  |     |     |
| [0414] | <211> 720   |     |     |
| [0415] | <212> PRT   |     |     |
| [0416] | <213> 抗辐射热球菌(Thermococcus gammatolerans)                        |     |     |
| [0417] | <400> 10  |     |     |
| [0418] | Met Lys Val Asp Glu Leu Pro Val Asp Glu Arg Leu Lys Ala Val Leu |     |     |
| [0419] | 1   | 5   | 10  |
|        |   |     | 15  |

|        |   |
|--------|---|
| [0420] | Lys Glu Arg Gly Ile Glu Glu Leu Tyr Pro Pro Gln Ala Glu Ala Leu |
| [0421] | 20 25 30  |
| [0422] | Lys Ser Gly Ala Leu Glu Gly Arg Asn Leu Val Leu Ala Ile Pro Thr |
| [0423] | 35 40 45  |
| [0424] | Ala Ser Gly Lys Thr Leu Val Ser Glu Ile Val Met Val Asn Lys Leu |
| [0425] | 50 55 60  |
| [0426] | Ile Gln Glu Gly Gly Lys Ala Val Tyr Leu Val Pro Leu Lys Ala Leu |
| [0427] | 65 70 75 80   |
| [0428] | Ala Glu Glu Lys Tyr Arg Glu Phe Lys Glu Trp Glu Lys Leu Gly Leu |
| [0429] | 85 90 95  |
| [0430] | Lys Val Ala Ala Thr Thr Gly Asp Tyr Asp Ser Thr Asp Asp Trp Leu |
| [0431] | 100 105 110   |
| [0432] | Gly Arg Tyr Asp Ile Ile Val Ala Thr Ala Glu Lys Phe Asp Ser Leu |
| [0433] | 115 120 125   |
| [0434] | Leu Arg His Gly Ala Arg Trp Ile Asn Asp Val Lys Leu Val Val Ala |
| [0435] | 130 135 140   |
| [0436] | Asp Glu Val His Leu Ile Gly Ser Tyr Asp Arg Gly Ala Thr Leu Glu |
| [0437] | 145 150 155 160   |
| [0438] | Met Ile Leu Thr His Met Leu Gly Arg Ala Gln Ile Leu Ala Leu Ser |
| [0439] | 165 170 175   |
| [0440] | Ala Thr Val Gly Asn Ala Glu Glu Leu Ala Glu Trp Leu Asp Ala Ser |
| [0441] | 180 185 190   |
| [0442] | Leu Val Val Ser Asp Trp Arg Pro Val Gln Leu Arg Arg Gly Val Phe |
| [0443] | 195 200 205   |
| [0444] | His Leu Gly Thr Leu Ile Trp Glu Asp Gly Lys Val Glu Ser Tyr Pro |
| [0445] | 210 215 220   |
| [0446] | Glu Asn Trp Tyr Ser Leu Val Val Asp Ala Val Lys Arg Gly Lys Gly |
| [0447] | 225 230 235 240   |
| [0448] | Ala Leu Val Phe Val Asn Thr Arg Arg Ser Ala Glu Lys Glu Ala Leu |
| [0449] | 245 250 255   |
| [0450] | Ala Leu Ser Lys Leu Val Ser Ser His Leu Thr Lys Pro Glu Lys Arg |
| [0451] | 260 265 270   |
| [0452] | Ala Leu Glu Ser Leu Ala Ser Gln Leu Glu Asp Asn Pro Thr Ser Glu |
| [0453] | 275 280 285   |
| [0454] | Lys Leu Lys Arg Ala Leu Arg Gly Gly Val Ala Phe His His Ala Gly |
| [0455] | 290 295 300   |
| [0456] | Leu Ser Arg Val Glu Arg Thr Leu Ile Glu Asp Ala Phe Arg Glu Gly |
| [0457] | 305 310 315 320   |
| [0458] | Leu Ile Lys Val Ile Thr Ala Thr Pro Thr Leu Ser Ala Gly Val Asn |
| [0459] | 325 330 335   |
| [0460] | Leu Pro Ser Phe Arg Val Ile Ile Arg Asp Thr Lys Arg Tyr Ala Gly |
| [0461] | 340 345 350   |

|        |   |     |     |
|--------|---|-----|-----|
| [0462] | Phe Gly Trp Thr Asp Ile Pro Val Leu Glu Ile Gln Gln Met Met Gly |     |     |
| [0463] | 355   | 360 | 365 |
| [0464] | Arg Ala Gly Arg Pro Arg Tyr Asp Lys Tyr Gly Glu Ala Ile Ile Val |     |     |
| [0465] | 370   | 375 | 380 |
| [0466] | Ala Arg Thr Asp Glu Pro Gly Lys Leu Met Glu Arg Tyr Ile Arg Gly |     |     |
| [0467] | 385   | 390 | 395 |
| [0468] | Lys Pro Glu Lys Leu Phe Ser Met Leu Ala Asn Glu Gln Ala Phe Arg |     |     |
| [0469] | 405   | 410 | 415 |
| [0470] | Ser Gln Val Leu Ala Leu Ile Thr Asn Phe Gly Ile Arg Ser Phe Pro |     |     |
| [0471] | 420   | 425 | 430 |
| [0472] | Glu Leu Val Arg Phe Leu Glu Arg Thr Phe Tyr Ala His Gln Arg Lys |     |     |
| [0473] | 435   | 440 | 445 |
| [0474] | Asp Leu Ser Ser Leu Glu Tyr Lys Ala Lys Glu Val Val Tyr Phe Leu |     |     |
| [0475] | 450   | 455 | 460 |
| [0476] | Ile Glu Asn Glu Phe Ile Asp Leu Asp Leu Glu Asp Arg Phe Ile Pro |     |     |
| [0477] | 465   | 470 | 475 |
| [0478] | Leu Pro Phe Gly Lys Arg Thr Ser Gln Leu Tyr Ile Asp Pro Leu Thr |     |     |
| [0479] | 485   | 490 | 495 |
| [0480] | Ala Lys Lys Phe Lys Asp Ala Phe Pro Ala Ile Glu Arg Asn Pro Asn |     |     |
| [0481] | 500   | 505 | 510 |
| [0482] | Pro Phe Gly Ile Phe Gln Leu Ile Ala Ser Thr Pro Asp Met Ala Thr |     |     |
| [0483] | 515   | 520 | 525 |
| [0484] | Leu Thr Ala Arg Arg Arg Glu Met Glu Asp Tyr Leu Asp Leu Ala Tyr |     |     |
| [0485] | 530   | 535 | 540 |
| [0486] | Glu Leu Glu Asp Lys Leu Tyr Ala Ser Ile Pro Tyr Tyr Glu Asp Ser |     |     |
| [0487] | 545   | 550 | 555 |
| [0488] | Arg Phe Gln Gly Phe Leu Gly Gln Val Lys Thr Ala Lys Val Leu Leu |     |     |
| [0489] | 565   | 570 | 575 |
| [0490] | Asp Trp Ile Asn Glu Val Pro Glu Ala Arg Ile Tyr Glu Thr Tyr Ser |     |     |
| [0491] | 580   | 585 | 590 |
| [0492] | Ile Asp Pro Gly Asp Leu Tyr Arg Leu Leu Glu Leu Ala Asp Trp Leu |     |     |
| [0493] | 595   | 600 | 605 |
| [0494] | Met Tyr Ser Leu Ile Glu Leu Tyr Lys Leu Phe Glu Pro Lys Glu Glu |     |     |
| [0495] | 610   | 615 | 620 |
| [0496] | Ile Leu Asn Tyr Leu Arg Asp Leu His Leu Arg Leu Arg His Gly Val |     |     |
| [0497] | 625   | 630 | 635 |
| [0498] | Arg Glu Glu Leu Leu Glu Leu Val Arg Leu Pro Asn Ile Gly Arg Lys |     |     |
| [0499] | 645   | 650 | 655 |
| [0500] | Arg Ala Arg Ala Leu Tyr Asn Ala Gly Phe Arg Ser Val Glu Ala Ile |     |     |
| [0501] | 660   | 665 | 670 |
| [0502] | Ala Asn Ala Lys Pro Ala Glu Leu Leu Ala Val Glu Gly Ile Gly Ala |     |     |
| [0503] | 675   | 680 | 685 |

|        |   |
|--------|---|
| [0504] | Lys Ile Leu Asp Gly Ile Tyr Arg His Leu Gly Ile Glu Lys Arg Val |
| [0505] | 690 695 700   |
| [0506] | Thr Glu Glu Lys Pro Lys Arg Lys Gly Thr Leu Glu Asp Phe Leu Arg |
| [0507] | 705 710 715 720   |
| [0508] | <210> 11  |
| [0509] | <211> 799   |
| [0510] | <212> PRT   |
| [0511] | <213> 亨氏甲烷螺菌(Methanospirillum hungatei)                         |
| [0512] | <400> 11  |
| [0513] | Met Glu Ile Ala Ser Leu Pro Leu Pro Asp Ser Phe Ile Arg Ala Cys |
| [0514] | 1 5 10 15   |
| [0515] | His Ala Lys Gly Ile Arg Ser Leu Tyr Pro Pro Gln Ala Glu Cys Ile |
| [0516] | 20 25 30  |
| [0517] | Glu Lys Gly Leu Leu Glu Gly Lys Asn Leu Leu Ile Ser Ile Pro Thr |
| [0518] | 35 40 45  |
| [0519] | Ala Ser Gly Lys Thr Leu Leu Ala Glu Met Ala Met Trp Ser Arg Ile |
| [0520] | 50 55 60  |
| [0521] | Ala Ala Gly Gly Lys Cys Leu Tyr Ile Val Pro Leu Arg Ala Leu Ala |
| [0522] | 65 70 75 80   |
| [0523] | Ser Glu Lys Tyr Asp Glu Phe Ser Lys Lys Gly Val Ile Arg Val Gly |
| [0524] | 85 90 95  |
| [0525] | Ile Ala Thr Gly Asp Leu Asp Arg Thr Asp Ala Tyr Leu Gly Glu Asn |
| [0526] | 100 105 110   |
| [0527] | Asp Ile Ile Val Ala Thr Ser Glu Lys Thr Asp Ser Leu Leu Arg Asn |
| [0528] | 115 120 125   |
| [0529] | Arg Thr Pro Trp Leu Ser Gln Ile Thr Cys Ile Val Leu Asp Glu Val |
| [0530] | 130 135 140   |
| [0531] | His Leu Ile Gly Ser Glu Asn Arg Gly Ala Thr Leu Glu Met Val Ile |
| [0532] | 145 150 155 160   |
| [0533] | Thr Lys Leu Arg Tyr Thr Asn Pro Val Met Gln Ile Ile Gly Leu Ser |
| [0534] | 165 170 175   |
| [0535] | Ala Thr Ile Gly Asn Pro Ala Gln Leu Ala Glu Trp Leu Asp Ala Thr |
| [0536] | 180 185 190   |
| [0537] | Leu Ile Thr Ser Thr Trp Arg Pro Val Asp Leu Arg Gln Gly Val Tyr |
| [0538] | 195 200 205   |
| [0539] | Tyr Asn Gly Lys Ile Arg Phe Ser Asp Ser Glu Arg Pro Ile Gln Gly |
| [0540] | 210 215 220   |
| [0541] | Lys Thr Lys His Asp Asp Leu Asn Leu Cys Leu Asp Thr Ile Glu Glu |
| [0542] | 225 230 235 240   |
| [0543] | Gly Gly Gln Cys Leu Val Phe Val Ser Ser Arg Arg Asn Ala Glu Gly |
| [0544] | 245 250 255   |
| [0545] | Phe Ala Lys Lys Ala Ala Gly Ala Leu Lys Ala Gly Ser Pro Asp Ser |



|        |   |     |     |
|--------|---|-----|-----|
| [0546] | 260   | 265 | 270 |
| [0547] | Lys Ala Leu Ala Gln Glu Leu Arg Arg Leu Arg Asp Arg Asp Glu Gly |     |     |
| [0548] | 275   | 280 | 285 |
| [0549] | Asn Val Leu Ala Asp Cys Val Glu Arg Gly Ala Ala Phe His His Ala |     |     |
| [0550] | 290   | 295 | 300 |
| [0551] | Gly Leu Ile Arg Gln Glu Arg Thr Ile Ile Glu Glu Gly Phe Arg Asn |     |     |
| [0552] | 305   | 310 | 315 |
| [0553] | Gly Tyr Ile Glu Val Ile Ala Ala Thr Pro Thr Leu Ala Ala Gly Leu |     |     |
| [0554] | 325   | 330 | 335 |
| [0555] | Asn Leu Pro Ala Arg Arg Val Ile Ile Arg Asp Tyr Asn Arg Phe Ala |     |     |
| [0556] | 340   | 345 | 350 |
| [0557] | Ser Gly Leu Gly Met Val Pro Ile Pro Val Gly Glu Tyr His Gln Met |     |     |
| [0558] | 355   | 360 | 365 |
| [0559] | Ala Gly Arg Ala Gly Arg Pro His Leu Asp Pro Tyr Gly Glu Ala Val |     |     |
| [0560] | 370   | 375 | 380 |
| [0561] | Leu Leu Ala Lys Asp Ala Pro Ser Val Glu Arg Leu Phe Glu Thr Phe |     |     |
| [0562] | 385   | 390 | 395 |
| [0563] | Ile Asp Ala Glu Ala Glu Arg Val Asp Ser Gln Cys Val Asp Asp Ala |     |     |
| [0564] | 405   | 410 | 415 |
| [0565] | Ser Leu Cys Ala His Ile Leu Ser Leu Ile Ala Thr Gly Phe Ala His |     |     |
| [0566] | 420   | 425 | 430 |
| [0567] | Asp Gln Glu Ala Leu Ser Ser Phe Met Glu Arg Thr Phe Tyr Phe Phe |     |     |
| [0568] | 435   | 440 | 445 |
| [0569] | Gln His Pro Lys Thr Arg Ser Leu Pro Arg Leu Val Ala Asp Ala Ile |     |     |
| [0570] | 450   | 455 | 460 |
| [0571] | Arg Phe Leu Thr Thr Ala Gly Met Val Glu Glu Arg Glu Asn Thr Leu |     |     |
| [0572] | 465   | 470 | 475 |
| [0573] | Ser Ala Thr Arg Leu Gly Ser Leu Val Ser Arg Leu Tyr Leu Asn Pro |     |     |
| [0574] | 485   | 490 | 495 |
| [0575] | Cys Thr Ala Arg Leu Ile Leu Asp Ser Leu Lys Ser Cys Lys Thr Pro |     |     |
| [0576] | 500   | 505 | 510 |
| [0577] | Thr Leu Ile Gly Leu Leu His Val Ile Cys Val Ser Pro Asp Met Gln |     |     |
| [0578] | 515   | 520 | 525 |
| [0579] | Arg Leu Tyr Leu Lys Ala Ala Asp Thr Gln Leu Leu Arg Thr Phe Leu |     |     |
| [0580] | 530   | 535 | 540 |
| [0581] | Phe Lys His Lys Asp Asp Leu Ile Leu Pro Leu Pro Phe Glu Gln Glu |     |     |
| [0582] | 545   | 550 | 555 |
| [0583] | Glu Glu Glu Leu Trp Leu Ser Gly Leu Lys Thr Ala Leu Val Leu Thr |     |     |
| [0584] | 565   | 570 | 575 |
| [0585] | Asp Trp Ala Asp Glu Phe Ser Glu Gly Met Ile Glu Glu Arg Tyr Gly |     |     |
| [0586] | 580   | 585 | 590 |
| [0587] | Ile Gly Ala Gly Asp Leu Tyr Asn Ile Val Asp Ser Gly Lys Trp Leu |     |     |

|        |   |     |     |
|--------|---|-----|-----|
| [0588] | 595   | 600 | 605 |
| [0589] | Leu His Gly Thr Glu Arg Leu Val Ser Val Glu Met Pro Glu Met Ser |     |     |
| [0590] | 610   | 615 | 620 |
| [0591] | Gln Val Val Lys Thr Leu Ser Val Arg Val His His Gly Val Lys Ser |     |     |
| [0592] | 625   | 630 | 635 |
| [0593] | Glu Leu Leu Pro Leu Val Ala Leu Arg Asn Ile Gly Arg Val Arg Ala |     |     |
| [0594] | 645   | 650 | 655 |
| [0595] | Arg Thr Leu Tyr Asn Ala Gly Tyr Pro Asp Pro Glu Ala Val Ala Arg |     |     |
| [0596] | 660   | 665 | 670 |
| [0597] | Ala Gly Leu Ser Thr Ile Ala Arg Ile Ile Gly Glu Gly Ile Ala Arg |     |     |
| [0598] | 675   | 680 | 685 |
| [0599] | Gln Val Ile Asp Glu Ile Thr Gly Val Lys Arg Ser Gly Ile His Ser |     |     |
| [0600] | 690   | 695 | 700 |
| [0601] | Ser Asp Asp Asp Tyr Gln Gln Lys Thr Pro Glu Leu Leu Thr Asp Ile |     |     |
| [0602] | 705   | 710 | 715 |
| [0603] | Pro Gly Ile Gly Lys Lys Met Ala Glu Lys Leu Gln Asn Ala Gly Ile |     |     |
| [0604] | 725   | 730 | 735 |
| [0605] | Ile Thr Val Ser Asp Leu Leu Thr Ala Asp Glu Val Leu Leu Ser Asp |     |     |
| [0606] | 740   | 745 | 750 |
| [0607] | Val Leu Gly Ala Ala Arg Ala Arg Lys Val Leu Ala Phe Leu Ser Asn |     |     |
| [0608] | 755   | 760 | 765 |
| [0609] | Ser Glu Lys Glu Asn Ser Ser Ser Asp Lys Thr Glu Glu Ile Pro Asp |     |     |
| [0610] | 770   | 775 | 780 |
| [0611] | Thr Gln Lys Ile Arg Gly Gln Ser Ser Trp Glu Asp Phe Gly Cys     |     |     |
| [0612] | 785   | 790 | 795 |
| [0613] | <210> 12  |     |     |
| [0614] | <211> 1756  |     |     |
| [0615] | <212> PRT   |     |     |
| [0616] | <213> 大肠杆菌(Escherichia coli)                                    |     |     |
| [0617] | <400> 12  |     |     |
| [0618] | Met Met Ser Ile Ala Gln Val Arg Ser Ala Gly Ser Ala Gly Asn Tyr |     |     |
| [0619] | 1   | 5   | 10  |
| [0620] | Tyr Thr Asp Lys Asp Asn Tyr Tyr Val Leu Gly Ser Met Gly Glu Arg |     |     |
| [0621] | 20  | 25  | 30  |
| [0622] | Trp Ala Gly Lys Gly Ala Glu Gln Leu Gly Leu Gln Gly Ser Val Asp |     |     |
| [0623] | 35  | 40  | 45  |
| [0624] | Lys Asp Val Phe Thr Arg Leu Leu Glu Gly Arg Leu Pro Asp Gly Ala |     |     |
| [0625] | 50  | 55  | 60  |
| [0626] | Asp Leu Ser Arg Met Gln Asp Gly Ser Asn Lys His Arg Pro Gly Tyr |     |     |
| [0627] | 65  | 70  | 75  |
| [0628] | Asp Leu Thr Phe Ser Ala Pro Lys Ser Val Ser Met Met Ala Met Leu |     |     |
| [0629] | 85  | 90  | 95  |

|        |   |     |     |
|--------|---|-----|-----|
| [0630] | Gly Gly Asp Lys Arg Leu Ile Asp Ala His Asn Gln Ala Val Asp Phe |     |     |
| [0631] | 100   | 105 | 110 |
| [0632] | Ala Val Arg Gln Val Glu Ala Leu Ala Ser Thr Arg Val Met Thr Asp |     |     |
| [0633] | 115   | 120 | 125 |
| [0634] | Gly Gln Ser Glu Thr Val Leu Thr Gly Asn Leu Val Met Ala Leu Phe |     |     |
| [0635] | 130   | 135 | 140 |
| [0636] | Asn His Asp Thr Ser Arg Asp Gln Glu Pro Gln Leu His Thr His Ala |     |     |
| [0637] | 145   | 150 | 155 |
| [0638] | Val Val Ala Asn Val Thr Gln His Asn Gly Glu Trp Lys Thr Leu Ser |     |     |
| [0639] | 165   | 170 | 175 |
| [0640] | Ser Asp Lys Val Gly Lys Thr Gly Phe Ile Glu Asn Val Tyr Ala Asn |     |     |
| [0641] | 180   | 185 | 190 |
| [0642] | Gln Ile Ala Phe Gly Arg Leu Tyr Arg Glu Lys Leu Lys Glu Gln Val |     |     |
| [0643] | 195   | 200 | 205 |
| [0644] | Glu Ala Leu Gly Tyr Glu Thr Glu Val Val Gly Lys His Gly Met Trp |     |     |
| [0645] | 210   | 215 | 220 |
| [0646] | Glu Met Pro Gly Val Pro Val Glu Ala Phe Ser Gly Arg Ser Gln Ala |     |     |
| [0647] | 225   | 230 | 235 |
| [0648] | Ile Arg Glu Ala Val Gly Glu Asp Ala Ser Leu Lys Ser Arg Asp Val |     |     |
| [0649] | 245   | 250 | 255 |
| [0650] | Ala Ala Leu Asp Thr Arg Lys Ser Lys Gln His Val Asp Pro Glu Ile |     |     |
| [0651] | 260   | 265 | 270 |
| [0652] | Arg Met Ala Glu Trp Met Gln Thr Leu Lys Glu Thr Gly Phe Asp Ile |     |     |
| [0653] | 275   | 280 | 285 |
| [0654] | Arg Ala Tyr Arg Asp Ala Ala Asp Gln Arg Thr Glu Ile Arg Thr Gln |     |     |
| [0655] | 290   | 295 | 300 |
| [0656] | Ala Pro Gly Pro Ala Ser Gln Asp Gly Pro Asp Val Gln Gln Ala Val |     |     |
| [0657] | 305   | 310 | 315 |
| [0658] | Thr Gln Ala Ile Ala Gly Leu Ser Glu Arg Lys Val Gln Phe Thr Tyr |     |     |
| [0659] | 325   | 330 | 335 |
| [0660] | Thr Asp Val Leu Ala Arg Thr Val Gly Ile Leu Pro Pro Glu Asn Gly |     |     |
| [0661] | 340   | 345 | 350 |
| [0662] | Val Ile Glu Arg Ala Arg Ala Gly Ile Asp Glu Ala Ile Ser Arg Glu |     |     |
| [0663] | 355   | 360 | 365 |
| [0664] | Gln Leu Ile Pro Leu Asp Arg Glu Lys Gly Leu Phe Thr Ser Gly Ile |     |     |
| [0665] | 370   | 375 | 380 |
| [0666] | His Val Leu Asp Glu Leu Ser Val Arg Ala Leu Ser Arg Asp Ile Met |     |     |
| [0667] | 385   | 390 | 395 |
| [0668] | Lys Gln Asn Arg Val Thr Val His Pro Glu Lys Ser Val Pro Arg Thr |     |     |
| [0669] | 405   | 410 | 415 |
| [0670] | Ala Gly Tyr Ser Asp Ala Val Ser Val Leu Ala Gln Asp Arg Pro Ser |     |     |
| [0671] | 420   | 425 | 430 |

|        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| [0672] | Leu | Ala | Ile | Val | Ser | Gly | Gln | Gly | Gly | Ala | Ala | Gly | Gln | Arg | Glu | Arg |
| [0673] |     |     |     | 435 |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
| [0674] | Val | Ala | Glu | Leu | Val | Met | Met | Ala | Arg | Glu | Gln | Gly | Arg | Glu | Val | Gln |
| [0675] |     |     | 450 |     |     |     | 455 |     |     |     |     |     | 460 |     |     |     |
| [0676] | Ile | Ile | Ala | Ala | Asp | Arg | Arg | Ser | Gln | Met | Asn | Leu | Lys | Gln | Asp | Glu |
| [0677] | 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     | 480 |     |
| [0678] | Arg | Leu | Ser | Gly | Glu | Leu | Ile | Thr | Gly | Arg | Arg | Gln | Leu | Leu | Glu | Gly |
| [0679] |     |     |     |     | 485 |     |     |     | 490 |     |     |     |     |     | 495 |     |
| [0680] | Met | Ala | Phe | Thr | Pro | Gly | Ser | Thr | Val | Ile | Val | Asp | Gln | Gly | Glu | Lys |
| [0681] |     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |
| [0682] | Leu | Ser | Leu | Lys | Glu | Thr | Leu | Thr | Leu | Leu | Asp | Gly | Ala | Ala | Arg | His |
| [0683] |     |     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |
| [0684] | Asn | Val | Gln | Val | Leu | Ile | Thr | Asp | Ser | Gly | Gln | Arg | Thr | Gly | Thr | Gly |
| [0685] |     |     | 530 |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
| [0686] | Ser | Ala | Leu | Met | Ala | Met | Lys | Asp | Ala | Gly | Val | Asn | Thr | Tyr | Arg | Trp |
| [0687] | 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     | 560 |     |
| [0688] | Gln | Gly | Gly | Glu | Gln | Arg | Pro | Ala | Thr | Ile | Ile | Ser | Glu | Pro | Asp | Arg |
| [0689] |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |     |
| [0690] | Asn | Val | Arg | Tyr | Ala | Arg | Leu | Ala | Gly | Asp | Phe | Ala | Ala | Ser | Val | Lys |
| [0691] |     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| [0692] | Ala | Gly | Glu | Glu | Ser | Val | Ala | Gln | Val | Ser | Gly | Val | Arg | Glu | Gln | Ala |
| [0693] |     |     | 595 |     |     |     |     | 600 |     |     |     |     |     | 605 |     |     |
| [0694] | Ile | Leu | Thr | Gln | Ala | Ile | Arg | Ser | Glu | Leu | Lys | Thr | Gln | Gly | Val | Leu |
| [0695] |     | 610 |     |     |     |     | 615 |     |     |     |     |     | 620 |     |     |     |
| [0696] | Gly | His | Pro | Glu | Val | Thr | Met | Thr | Ala | Leu | Ser | Pro | Val | Trp | Leu | Asp |
| [0697] | 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     | 640 |     |
| [0698] | Ser | Arg | Ser | Arg | Tyr | Leu | Arg | Asp | Met | Tyr | Arg | Pro | Gly | Met | Val | Met |
| [0699] |     |     |     | 645 |     |     |     |     |     | 650 |     |     |     | 655 |     |     |
| [0700] | Glu | Gln | Trp | Asn | Pro | Glu | Thr | Arg | Ser | His | Asp | Arg | Tyr | Val | Ile | Asp |
| [0701] |     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |
| [0702] | Arg | Val | Thr | Ala | Gln | Ser | His | Ser | Leu | Thr | Leu | Arg | Asp | Ala | Gln | Gly |
| [0703] |     |     | 675 |     |     |     |     |     | 680 |     |     |     |     | 685 |     |     |
| [0704] | Glu | Thr | Gln | Val | Val | Arg | Ile | Ser | Ser | Leu | Asp | Ser | Ser | Trp | Ser | Leu |
| [0705] |     | 690 |     |     |     |     | 695 |     |     |     |     |     | 700 |     |     |     |
| [0706] | Phe | Arg | Pro | Glu | Lys | Met | Pro | Val | Ala | Asp | Gly | Glu | Arg | Leu | Arg | Val |
| [0707] | 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     | 720 |     |
| [0708] | Thr | Gly | Lys | Ile | Pro | Gly | Leu | Arg | Val | Ser | Gly | Gly | Asp | Arg | Leu | Gln |
| [0709] |     |     |     | 725 |     |     |     |     |     | 730 |     |     |     | 735 |     |     |
| [0710] | Val | Ala | Ser | Val | Ser | Glu | Asp | Ala | Met | Thr | Val | Val | Val | Pro | Gly | Arg |
| [0711] |     |     |     | 740 |     |     |     |     |     | 745 |     |     |     | 750 |     |     |
| [0712] | Ala | Glu | Pro | Ala | Ser | Leu | Pro | Val | Ser | Asp | Ser | Pro | Phe | Thr | Ala | Leu |
| [0713] |     |     | 755 |     |     |     |     |     | 760 |     |     |     |     | 765 |     |     |

|        |      |     |     |     |     |     |     |     |     |      |     |     |     |      |      |     |
|--------|------|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|------|------|-----|
| [0714] | Lys  | Leu | Glu | Asn | Gly | Trp | Val | Glu | Thr | Pro  | Gly | His | Ser | Val  | Ser  | Asp |
| [0715] | 770  |     |     |     |     |     | 775 |     |     |      |     | 780 |     |      |      |     |
| [0716] | Ser  | Ala | Thr | Val | Phe | Ala | Ser | Val | Thr | Gln  | Met | Ala | Met | Asp  | Asn  | Ala |
| [0717] | 785  |     |     |     |     | 790 |     |     |     |      | 795 |     |     |      |      | 800 |
| [0718] | Thr  | Leu | Asn | Gly | Leu | Ala | Arg | Ser | Gly | Arg  | Asp | Val | Arg | Leu  | Tyr  | Ser |
| [0719] |      |     |     |     | 805 |     |     |     |     | 810  |     |     |     |      | 815  |     |
| [0720] | Ser  | Leu | Asp | Glu | Thr | Arg | Thr | Ala | Glu | Lys  | Leu | Ala | Arg | His  | Pro  | Ser |
| [0721] |      |     |     |     | 820 |     |     |     |     | 825  |     |     |     |      | 830  |     |
| [0722] | Phe  | Thr | Val | Val | Ser | Glu | Gln | Ile | Lys | Ala  | Arg | Ala | Gly | Glu  | Thr  | Leu |
| [0723] |      |     | 835 |     |     |     |     |     |     | 840  |     |     |     | 845  |      |     |
| [0724] | Leu  | Glu | Thr | Ala | Ile | Ser | Leu | Gln | Lys | Ala  | Gly | Leu | His | Thr  | Pro  | Ala |
| [0725] | 850  |     |     |     |     |     |     | 855 |     |      |     |     |     | 860  |      |     |
| [0726] | Gln  | Gln | Ala | Ile | His | Leu | Ala | Leu | Pro | Val  | Leu | Glu | Ser | Lys  | Asn  | Leu |
| [0727] | 865  |     |     |     |     | 870 |     |     |     |      |     | 875 |     |      |      | 880 |
| [0728] | Ala  | Phe | Ser | Met | Val | Asp | Leu | Leu | Thr | Glu  | Ala | Lys | Ser | Phe  | Ala  | Ala |
| [0729] |      |     |     |     | 885 |     |     |     |     |      | 890 |     |     |      |      | 895 |
| [0730] | Glu  | Gly | Thr | Gly | Phe | Thr | Glu | Leu | Gly | Gly  | Glu | Ile | Asn | Ala  | Gln  | Ile |
| [0731] |      |     |     | 900 |     |     |     |     |     | 905  |     |     |     |      | 910  |     |
| [0732] | Lys  | Arg | Gly | Asp | Leu | Leu | Tyr | Val | Asp | Val  | Ala | Lys | Gly | Tyr  | Gly  | Thr |
| [0733] |      |     | 915 |     |     |     |     |     |     | 920  |     |     |     | 925  |      |     |
| [0734] | Gly  | Leu | Leu | Val | Ser | Arg | Ala | Ser | Tyr | Glu  | Ala | Glu | Lys | Ser  | Ile  | Leu |
| [0735] | 930  |     |     |     |     |     |     | 935 |     |      |     |     |     | 940  |      |     |
| [0736] | Arg  | His | Ile | Leu | Glu | Gly | Lys | Glu | Ala | Val  | Thr | Pro | Leu | Met  | Glu  | Arg |
| [0737] | 945  |     |     |     |     | 950 |     |     |     |      |     | 955 |     |      |      | 960 |
| [0738] | Val  | Pro | Gly | Glu | Leu | Met | Glu | Thr | Leu | Thr  | Ser | Gly | Gln | Arg  | Ala  | Ala |
| [0739] |      |     |     |     | 965 |     |     |     |     |      | 970 |     |     |      |      | 975 |
| [0740] | Thr  | Arg | Met | Ile | Leu | Glu | Thr | Ser | Asp | Arg  | Phe | Thr | Val | Val  | Gln  | Gly |
| [0741] |      |     |     | 980 |     |     |     |     |     | 985  |     |     |     |      | 990  |     |
| [0742] | Tyr  | Ala | Gly | Val | Gly | Lys | Thr | Thr | Gln | Phe  | Arg | Ala | Val | Met  | Ser  | Ala |
| [0743] |      |     | 995 |     |     |     |     |     |     | 1000 |     |     |     |      | 1005 |     |
| [0744] | Val  | Asn | Met | Leu | Pro | Ala | Ser | Glu | Arg | Pro  | Arg | Val | Val | Gly  | Leu  |     |
| [0745] | 1010 |     |     |     |     |     |     |     |     | 1015 |     |     |     | 1020 |      |     |
| [0746] | Gly  | Pro | Thr | His | Arg | Ala | Val | Gly | Glu | Met  | Arg | Ser | Ala | Gly  | Val  |     |
| [0747] | 1025 |     |     |     |     |     |     |     |     | 1030 |     |     |     | 1035 |      |     |
| [0748] | Asp  | Ala | Gln | Thr | Leu | Ala | Ser | Phe | Leu | His  | Asp | Thr | Gln | Leu  | Gln  |     |
| [0749] | 1040 |     |     |     |     |     |     |     |     | 1045 |     |     |     | 1050 |      |     |
| [0750] | Gln  | Arg | Ser | Gly | Glu | Thr | Pro | Asp | Phe | Ser  | Asn | Thr | Leu | Phe  | Leu  |     |
| [0751] | 1055 |     |     |     |     |     |     |     |     | 1060 |     |     |     | 1065 |      |     |
| [0752] | Leu  | Asp | Glu | Ser | Ser | Met | Val | Gly | Asn | Thr  | Glu | Met | Ala | Arg  | Ala  |     |
| [0753] | 1070 |     |     |     |     |     |     |     |     | 1075 |     |     |     | 1080 |      |     |
| [0754] | Tyr  | Ala | Leu | Ile | Ala | Ala | Gly | Gly | Gly | Arg  | Ala | Val | Ala | Ser  | Gly  |     |
| [0755] | 1085 |     |     |     |     |     |     |     |     | 1090 |     |     |     | 1095 |      |     |

|        |   |
|--------|---|
| [0756] | Asp Thr Asp Gln Leu Gln Ala Ile Ala Pro Gly Gln Ser Phe Arg |
| [0757] | 1100 1105 1110  |
| [0758] | Leu Gln Gln Thr Arg Ser Ala Ala Asp Val Val Ile Met Lys Glu |
| [0759] | 1115 1120 1125  |
| [0760] | Ile Val Arg Gln Thr Pro Glu Leu Arg Glu Ala Val Tyr Ser Leu |
| [0761] | 1130 1135 1140  |
| [0762] | Ile Asn Arg Asp Val Glu Arg Ala Leu Ser Gly Leu Glu Ser Val |
| [0763] | 1145 1150 1155  |
| [0764] | Lys Pro Ser Gln Val Pro Arg Leu Glu Gly Ala Trp Ala Pro Glu |
| [0765] | 1160 1165 1170  |
| [0766] | His Ser Val Thr Glu Phe Ser His Ser Gln Glu Ala Lys Leu Ala |
| [0767] | 1175 1180 1185  |
| [0768] | Glu Ala Gln Gln Lys Ala Met Leu Lys Gly Glu Ala Phe Pro Asp |
| [0769] | 1190 1195 1200  |
| [0770] | Ile Pro Met Thr Leu Tyr Glu Ala Ile Val Arg Asp Tyr Thr Gly |
| [0771] | 1205 1210 1215  |
| [0772] | Arg Thr Pro Glu Ala Arg Glu Gln Thr Leu Ile Val Thr His Leu |
| [0773] | 1220 1225 1230  |
| [0774] | Asn Glu Asp Arg Arg Val Leu Asn Ser Met Ile His Asp Ala Arg |
| [0775] | 1235 1240 1245  |
| [0776] | Glu Lys Ala Gly Glu Leu Gly Lys Glu Gln Val Met Val Pro Val |
| [0777] | 1250 1255 1260  |
| [0778] | Leu Asn Thr Ala Asn Ile Arg Asp Gly Glu Leu Arg Arg Leu Ser |
| [0779] | 1265 1270 1275  |
| [0780] | Thr Trp Glu Lys Asn Pro Asp Ala Leu Ala Leu Val Asp Asn Val |
| [0781] | 1280 1285 1290  |
| [0782] | Tyr His Arg Ile Ala Gly Ile Ser Lys Asp Asp Gly Leu Ile Thr |
| [0783] | 1295 1300 1305  |
| [0784] | Leu Gln Asp Ala Glu Gly Asn Thr Arg Leu Ile Ser Pro Arg Glu |
| [0785] | 1310 1315 1320  |
| [0786] | Ala Val Ala Glu Gly Val Thr Leu Tyr Thr Pro Asp Lys Ile Arg |
| [0787] | 1325 1330 1335  |
| [0788] | Val Gly Thr Gly Asp Arg Met Arg Phe Thr Lys Ser Asp Arg Glu |
| [0789] | 1340 1345 1350  |
| [0790] | Arg Gly Tyr Val Ala Asn Ser Val Trp Thr Val Thr Ala Val Ser |
| [0791] | 1355 1360 1365  |
| [0792] | Gly Asp Ser Val Thr Leu Ser Asp Gly Gln Gln Thr Arg Val Ile |
| [0793] | 1370 1375 1380  |
| [0794] | Arg Pro Gly Gln Glu Arg Ala Glu Gln His Ile Asp Leu Ala Tyr |
| [0795] | 1385 1390 1395  |
| [0796] | Ala Ile Thr Ala His Gly Ala Gln Gly Ala Ser Glu Thr Phe Ala |
| [0797] | 1400 1405 1410  |

|        |   |
|--------|---|
| [0798] | Ile Ala Leu Glu Gly Thr Glu Gly Asn Arg Lys Leu Met Ala Gly |
| [0799] | 1415 1420 1425  |
| [0800] | Phe Glu Ser Ala Tyr Val Ala Leu Ser Arg Met Lys Gln His Val |
| [0801] | 1430 1435 1440  |
| [0802] | Gln Val Tyr Thr Asp Asn Arg Gln Gly Trp Thr Asp Ala Ile Asn |
| [0803] | 1445 1450 1455  |
| [0804] | Asn Ala Val Gln Lys Gly Thr Ala His Asp Val Leu Glu Pro Lys |
| [0805] | 1460 1465 1470  |
| [0806] | Pro Asp Arg Glu Val Met Asn Ala Gln Arg Leu Phe Ser Thr Ala |
| [0807] | 1475 1480 1485  |
| [0808] | Arg Glu Leu Arg Asp Val Ala Ala Gly Arg Ala Val Leu Arg Gln |
| [0809] | 1490 1495 1500  |
| [0810] | Ala Gly Leu Ala Gly Gly Asp Ser Pro Ala Arg Phe Ile Ala Pro |
| [0811] | 1505 1510 1515  |
| [0812] | Gly Arg Lys Tyr Pro Gln Pro Tyr Val Ala Leu Pro Ala Phe Asp |
| [0813] | 1520 1525 1530  |
| [0814] | Arg Asn Gly Lys Ser Ala Gly Ile Trp Leu Asn Pro Leu Thr Thr |
| [0815] | 1535 1540 1545  |
| [0816] | Asp Asp Gly Asn Gly Leu Arg Gly Phe Ser Gly Glu Gly Arg Val |
| [0817] | 1550 1555 1560  |
| [0818] | Lys Gly Ser Gly Asp Ala Gln Phe Val Ala Leu Gln Gly Ser Arg |
| [0819] | 1565 1570 1575  |
| [0820] | Asn Gly Glu Ser Leu Leu Ala Asp Asn Met Gln Asp Gly Val Arg |
| [0821] | 1580 1585 1590  |
| [0822] | Ile Ala Arg Asp Asn Pro Asp Ser Gly Val Val Val Arg Ile Ala |
| [0823] | 1595 1600 1605  |
| [0824] | Gly Glu Gly Arg Pro Trp Asn Pro Gly Ala Ile Thr Gly Gly Arg |
| [0825] | 1610 1615 1620  |
| [0826] | Val Trp Gly Asp Ile Pro Asp Asn Ser Val Gln Pro Gly Ala Gly |
| [0827] | 1625 1630 1635  |
| [0828] | Asn Gly Glu Pro Val Thr Ala Glu Val Leu Ala Gln Arg Gln Ala |
| [0829] | 1640 1645 1650  |
| [0830] | Glu Glu Ala Ile Arg Arg Glu Thr Glu Arg Arg Ala Asp Glu Ile |
| [0831] | 1655 1660 1665  |
| [0832] | Val Arg Lys Met Ala Glu Asn Lys Pro Asp Leu Pro Asp Gly Lys |
| [0833] | 1670 1675 1680  |
| [0834] | Thr Glu Leu Ala Val Arg Asp Ile Ala Gly Gln Glu Arg Asp Arg |
| [0835] | 1685 1690 1695  |
| [0836] | Ser Ala Ile Ser Glu Arg Glu Thr Ala Leu Pro Glu Ser Val Leu |
| [0837] | 1700 1705 1710  |
| [0838] | Arg Glu Ser Gln Arg Glu Arg Glu Ala Val Arg Glu Val Ala Arg |
| [0839] | 1715 1720 1725  |

|        |   |
|--------|---|
| [0840] | Glu Asn Leu Leu Gln Glu Arg Leu Gln Gln Met Glu Arg Asp Met     |
| [0841] | 1730 1735 1740  |
| [0842] | Val Arg Asp Leu Gln Lys Glu Lys Thr Leu Gly Gly Asp             |
| [0843] | 1745 1750 1755  |
| [0844] | <210> 13  |
| [0845] | <211> 726   |
| [0846] | <212> PRT   |
| [0847] | <213> 伯顿拟甲烷球菌(Methanococcoides burtonii)                        |
| [0848] | <400> 13  |
| [0849] | Met Ser Asp Lys Pro Ala Phe Met Lys Tyr Phe Thr Gln Ser Ser Cys |
| [0850] | 1 5 10 15   |
| [0851] | Tyr Pro Asn Gln Gln Glu Ala Met Asp Arg Ile His Ser Ala Leu Met |
| [0852] | 20 25 30  |
| [0853] | Gln Gln Gln Leu Val Leu Phe Glu Gly Ala Cys Gly Thr Gly Lys Thr |
| [0854] | 35 40 45  |
| [0855] | Leu Ser Ala Leu Val Pro Ala Leu His Val Gly Lys Met Leu Gly Lys |
| [0856] | 50 55 60  |
| [0857] | Thr Val Ile Ile Ala Thr Asn Val His Gln Gln Met Val Gln Phe Ile |
| [0858] | 65 70 75 80   |
| [0859] | Asn Glu Ala Arg Asp Ile Lys Lys Val Gln Asp Val Lys Val Ala Val |
| [0860] | 85 90 95  |
| [0861] | Ile Lys Gly Lys Thr Ala Met Cys Pro Gln Glu Ala Asp Tyr Glu Glu |
| [0862] | 100 105 110   |
| [0863] | Cys Ser Val Lys Arg Glu Asn Thr Phe Glu Leu Met Glu Thr Glu Arg |
| [0864] | 115 120 125   |
| [0865] | Glu Ile Tyr Leu Lys Arg Gln Glu Leu Asn Ser Ala Arg Asp Ser Tyr |
| [0866] | 130 135 140   |
| [0867] | Lys Lys Ser His Asp Pro Ala Phe Val Thr Leu Arg Asp Glu Leu Ser |
| [0868] | 145 150 155 160   |
| [0869] | Lys Glu Ile Asp Ala Val Glu Glu Lys Ala Arg Gly Leu Arg Asp Arg |
| [0870] | 165 170 175   |
| [0871] | Ala Cys Asn Asp Leu Tyr Glu Val Leu Arg Ser Asp Ser Glu Lys Phe |
| [0872] | 180 185 190   |
| [0873] | Arg Glu Trp Leu Tyr Lys Glu Val Arg Ser Pro Glu Glu Ile Asn Asp |
| [0874] | 195 200 205   |
| [0875] | His Ala Ile Lys Asp Gly Met Cys Gly Tyr Glu Leu Val Lys Arg Glu |
| [0876] | 210 215 220   |
| [0877] | Leu Lys His Ala Asp Leu Leu Ile Cys Asn Tyr His His Val Leu Asn |
| [0878] | 225 230 235 240   |
| [0879] | Pro Asp Ile Phe Ser Thr Val Leu Gly Trp Ile Glu Lys Glu Pro Gln |
| [0880] | 245 250 255   |
| [0881] | Glu Thr Ile Val Ile Phe Asp Glu Ala His Asn Leu Glu Ser Ala Ala |



|        |   |     |     |
|--------|---|-----|-----|
| [0882] | 260   | 265 | 270 |
| [0883] | Arg Ser His Ser Ser Leu Ser Leu Thr Glu His Ser Ile Glu Lys Ala |     |     |
| [0884] | 275   | 280 | 285 |
| [0885] | Ile Thr Glu Leu Glu Ala Asn Leu Asp Leu Leu Ala Asp Asp Asn Ile |     |     |
| [0886] | 290   | 295 | 300 |
| [0887] | His Asn Leu Phe Asn Ile Phe Leu Glu Val Ile Ser Asp Thr Tyr Asn |     |     |
| [0888] | 305   | 310 | 315 |
| [0889] | Ser Arg Phe Lys Phe Gly Glu Arg Glu Arg Val Arg Lys Asn Trp Tyr |     |     |
| [0890] | 325   | 330 | 335 |
| [0891] | Asp Ile Arg Ile Ser Asp Pro Tyr Glu Arg Asn Asp Ile Val Arg Gly |     |     |
| [0892] | 340   | 345 | 350 |
| [0893] | Lys Phe Leu Arg Gln Ala Lys Gly Asp Phe Gly Glu Lys Asp Asp Ile |     |     |
| [0894] | 355   | 360 | 365 |
| [0895] | Gln Ile Leu Leu Ser Glu Ala Ser Glu Leu Gly Ala Lys Leu Asp Glu |     |     |
| [0896] | 370   | 375 | 380 |
| [0897] | Thr Tyr Arg Asp Gln Tyr Lys Lys Gly Leu Ser Ser Val Met Lys Arg |     |     |
| [0898] | 385   | 390 | 395 |
| [0899] | Ser His Ile Arg Tyr Val Ala Asp Phe Met Ser Ala Tyr Ile Glu Leu |     |     |
| [0900] | 405   | 410 | 415 |
| [0901] | Ser His Asn Leu Asn Tyr Tyr Pro Ile Leu Asn Val Arg Arg Asp Met |     |     |
| [0902] | 420   | 425 | 430 |
| [0903] | Asn Asp Glu Ile Tyr Gly Arg Val Glu Leu Phe Thr Cys Ile Pro Lys |     |     |
| [0904] | 435   | 440 | 445 |
| [0905] | Asn Val Thr Glu Pro Leu Phe Asn Ser Leu Phe Ser Val Ile Leu Met |     |     |
| [0906] | 450   | 455 | 460 |
| [0907] | Ser Ala Thr Leu His Pro Phe Glu Met Val Lys Lys Thr Leu Gly Ile |     |     |
| [0908] | 465   | 470 | 475 |
| [0909] | Thr Arg Asp Thr Cys Glu Met Ser Tyr Gly Thr Ser Phe Pro Glu Glu |     |     |
| [0910] | 485   | 490 | 495 |
| [0911] | Lys Arg Leu Ser Ile Ala Val Ser Ile Pro Pro Leu Phe Ala Lys Asn |     |     |
| [0912] | 500   | 505 | 510 |
| [0913] | Arg Asp Asp Arg His Val Thr Glu Leu Leu Glu Gln Val Leu Leu Asp |     |     |
| [0914] | 515   | 520 | 525 |
| [0915] | Ser Ile Glu Asn Ser Lys Gly Asn Val Ile Leu Phe Phe Gln Ser Ala |     |     |
| [0916] | 530   | 535 | 540 |
| [0917] | Phe Glu Ala Lys Arg Tyr Tyr Ser Lys Ile Glu Pro Leu Val Asn Val |     |     |
| [0918] | 545   | 550 | 555 |
| [0919] | Pro Val Phe Leu Asp Glu Val Gly Ile Ser Ser Gln Asp Val Arg Glu |     |     |
| [0920] | 565   | 570 | 575 |
| [0921] | Glu Phe Phe Ser Ile Gly Glu Glu Asn Gly Lys Ala Val Leu Leu Ser |     |     |
| [0922] | 580   | 585 | 590 |
| [0923] | Tyr Leu Trp Gly Thr Leu Ser Glu Gly Ile Asp Tyr Arg Asp Gly Arg |     |     |

|        |   |     |     |
|--------|---|-----|-----|
| [0924] | 595   | 600 | 605 |
| [0925] | Gly Arg Thr Val Ile Ile Ile Gly Val Gly Tyr Pro Ala Leu Asn Asp |     |     |
| [0926] | 610   | 615 | 620 |
| [0927] | Arg Met Asn Ala Val Glu Ser Ala Tyr Asp His Val Phe Gly Tyr Gly |     |     |
| [0928] | 625   | 630 | 635 |
| [0929] | Ala Gly Trp Glu Phe Ala Ile Gln Val Pro Thr Ile Arg Lys Ile Arg |     |     |
| [0930] | 645   | 650 | 655 |
| [0931] | Gln Ala Met Gly Arg Val Val Arg Ser Pro Thr Asp Tyr Gly Ala Arg |     |     |
| [0932] | 660   | 665 | 670 |
| [0933] | Ile Leu Leu Asp Gly Arg Phe Leu Thr Asp Ser Lys Lys Arg Phe Gly |     |     |
| [0934] | 675   | 680 | 685 |
| [0935] | Lys Phe Ser Val Phe Glu Val Phe Pro Pro Ala Glu Arg Ser Glu Phe |     |     |
| [0936] | 690   | 695 | 700 |
| [0937] | Val Asp Val Asp Pro Glu Lys Val Lys Tyr Ser Leu Met Asn Phe Phe |     |     |
| [0938] | 705   | 710 | 715 |
| [0939] | Met Asp Asn Asp Glu Gln   |     |     |
| [0940] | 725   |     |     |
| [0941] | <210> 14  |     |     |
| [0942] | <211> 439   |     |     |
| [0943] | <212> PRT   |     |     |
| [0944] | <213> 肠杆菌噬菌体T4(Enterobacteria phage T4)                         |     |     |
| [0945] | <400> 14  |     |     |
| [0946] | Met Thr Phe Asp Asp Leu Thr Glu Gly Gln Lys Asn Ala Phe Asn Ile |     |     |
| [0947] | 1   | 5   | 10  |
| [0948] | Val Met Lys Ala Ile Lys Glu Lys Lys His His Val Thr Ile Asn Gly |     |     |
| [0949] | 20  | 25  | 30  |
| [0950] | Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr Lys Phe Ile Ile Glu Ala |     |     |
| [0951] | 35  | 40  | 45  |
| [0952] | Leu Ile Ser Thr Gly Glu Thr Gly Ile Ile Leu Ala Ala Pro Thr His |     |     |
| [0953] | 50  | 55  | 60  |
| [0954] | Ala Ala Lys Lys Ile Leu Ser Lys Leu Ser Gly Lys Glu Ala Ser Thr |     |     |
| [0955] | 65  | 70  | 75  |
| [0956] | Ile His Ser Ile Leu Lys Ile Asn Pro Val Thr Tyr Glu Glu Asn Val |     |     |
| [0957] | 85  | 90  | 95  |
| [0958] | Leu Phe Glu Gln Lys Glu Val Pro Asp Leu Ala Lys Cys Arg Val Leu |     |     |
| [0959] | 100   | 105 | 110 |
| [0960] | Ile Cys Asp Glu Val Ser Met Tyr Asp Arg Lys Leu Phe Lys Ile Leu |     |     |
| [0961] | 115   | 120 | 125 |
| [0962] | Leu Ser Thr Ile Pro Pro Trp Cys Thr Ile Ile Gly Ile Gly Asp Asn |     |     |
| [0963] | 130   | 135 | 140 |
| [0964] | Lys Gln Ile Arg Pro Val Asp Pro Gly Glu Asn Thr Ala Tyr Ile Ser |     |     |
| [0965] | 145   | 150 | 155 |

|        |   |
|--------|---|
| [0966] | Pro Phe Phe Thr His Lys Asp Phe Tyr Gln Cys Glu Leu Thr Glu Val |
| [0967] | 165 170 175   |
| [0968] | Lys Arg Ser Asn Ala Pro Ile Ile Asp Val Ala Thr Asp Val Arg Asn |
| [0969] | 180 185 190   |
| [0970] | Gly Lys Trp Ile Tyr Asp Lys Val Val Asp Gly His Gly Val Arg Gly |
| [0971] | 195 200 205   |
| [0972] | Phe Thr Gly Asp Thr Ala Leu Arg Asp Phe Met Val Asn Tyr Phe Ser |
| [0973] | 210 215 220   |
| [0974] | Ile Val Lys Ser Leu Asp Asp Leu Phe Glu Asn Arg Val Met Ala Phe |
| [0975] | 225 230 235 240   |
| [0976] | Thr Asn Lys Ser Val Asp Lys Leu Asn Ser Ile Ile Arg Lys Lys Ile |
| [0977] | 245 250 255   |
| [0978] | Phe Glu Thr Asp Lys Asp Phe Ile Val Gly Glu Ile Ile Val Met Gln |
| [0979] | 260 265 270   |
| [0980] | Glu Pro Leu Phe Lys Thr Tyr Lys Ile Asp Gly Lys Pro Val Ser Glu |
| [0981] | 275 280 285   |
| [0982] | Ile Ile Phe Asn Asn Gly Gln Leu Val Arg Ile Ile Glu Ala Glu Tyr |
| [0983] | 290 295 300   |
| [0984] | Thr Ser Thr Phe Val Lys Ala Arg Gly Val Pro Gly Glu Tyr Leu Ile |
| [0985] | 305 310 315 320   |
| [0986] | Arg His Trp Asp Leu Thr Val Glu Thr Tyr Gly Asp Asp Glu Tyr Tyr |
| [0987] | 325 330 335   |
| [0988] | Arg Glu Lys Ile Lys Ile Ile Ser Ser Asp Glu Glu Leu Tyr Lys Phe |
| [0989] | 340 345 350   |
| [0990] | Asn Leu Phe Leu Gly Lys Thr Ala Glu Thr Tyr Lys Asn Trp Asn Lys |
| [0991] | 355 360 365   |
| [0992] | Gly Gly Lys Ala Pro Trp Ser Asp Phe Trp Asp Ala Lys Ser Gln Phe |
| [0993] | 370 375 380   |
| [0994] | Ser Lys Val Lys Ala Leu Pro Ala Ser Thr Phe His Lys Ala Gln Gly |
| [0995] | 385 390 395 400   |
| [0996] | Met Ser Val Asp Arg Ala Phe Ile Tyr Thr Pro Cys Ile His Tyr Ala |
| [0997] | 405 410 415   |
| [0998] | Asp Val Glu Leu Ala Gln Gln Leu Leu Tyr Val Gly Val Thr Arg Gly |
| [0999] | 420 425 430   |
| [1000] | Arg Tyr Asp Val Phe Tyr Val                                     |
| [1001] | 435   |
| [1002] | <210> 15  |
| [1003] | <211> 970   |
| [1004] | <212> PRT   |
| [1005] | <213> 肉毒杆菌(Clostridium botulinum)                               |
| [1006] | <400> 15  |
| [1007] | Met Leu Ser Val Ala Asn Val Arg Ser Pro Ser Ala Ala Ala Ser Tyr |

|        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| [1008] | 1   |     |     | 5   |     |     |     | 10  |     |     |     | 15  |     |     |     |     |
| [1009] | Phe | Ala | Ser | Asp | Asn | Tyr | Tyr | Ala | Ser | Ala | Asp | Ala | Asp | Arg | Ser | Gly |
| [1010] | 20  |     |     | 25  |     |     |     | 30  |     |     |     |     |     |     |     |     |
| [1011] | Gln | Trp | Ile | Gly | Asp | Gly | Ala | Lys | Arg | Leu | Gly | Leu | Glu | Gly | Lys | Val |
| [1012] | 35  |     |     | 40  |     |     |     | 45  |     |     |     |     |     |     |     |     |
| [1013] | Glu | Ala | Arg | Ala | Phe | Asp | Ala | Leu | Leu | Arg | Gly | Glu | Leu | Pro | Asp | Gly |
| [1014] | 50  |     |     | 55  |     |     |     | 60  |     |     |     |     |     |     |     |     |
| [1015] | Ser | Ser | Val | Gly | Asn | Pro | Gly | Gln | Ala | His | Arg | Pro | Gly | Thr | Asp | Leu |
| [1016] | 65  |     |     | 70  |     |     |     | 75  |     |     |     | 80  |     |     |     |     |
| [1017] | Thr | Phe | Ser | Val | Pro | Lys | Ser | Trp | Ser | Leu | Leu | Ala | Leu | Val | Gly | Lys |
| [1018] | 85  |     |     | 90  |     |     |     | 95  |     |     |     |     |     |     |     |     |
| [1019] | Asp | Glu | Arg | Ile | Ile | Ala | Ala | Tyr | Arg | Glu | Ala | Val | Val | Glu | Ala | Leu |
| [1020] | 100 |     |     | 105 |     |     |     | 110 |     |     |     |     |     |     |     |     |
| [1021] | His | Trp | Ala | Glu | Lys | Asn | Ala | Ala | Glu | Thr | Arg | Val | Val | Glu | Lys | Gly |
| [1022] | 115 |     |     | 120 |     |     |     | 125 |     |     |     |     |     |     |     |     |
| [1023] | Met | Val | Val | Thr | Gln | Ala | Thr | Gly | Asn | Leu | Ala | Ile | Gly | Leu | Phe | Gln |
| [1024] | 130 |     |     | 135 |     |     |     | 140 |     |     |     |     |     |     |     |     |
| [1025] | His | Asp | Thr | Asn | Arg | Asn | Gln | Glu | Pro | Asn | Leu | His | Phe | His | Ala | Val |
| [1026] | 145 |     |     | 150 |     |     |     | 155 |     |     |     | 160 |     |     |     |     |
| [1027] | Ile | Ala | Asn | Val | Thr | Gln | Gly | Lys | Asp | Gly | Lys | Trp | Arg | Thr | Leu | Lys |
| [1028] | 165 |     |     | 170 |     |     |     | 175 |     |     |     |     |     |     |     |     |
| [1029] | Asn | Asp | Arg | Leu | Trp | Gln | Leu | Asn | Thr | Thr | Leu | Asn | Ser | Ile | Ala | Met |
| [1030] | 180 |     |     | 185 |     |     |     | 190 |     |     |     |     |     |     |     |     |
| [1031] | Ala | Arg | Phe | Arg | Val | Ala | Val | Glu | Lys | Leu | Gly | Tyr | Glu | Pro | Gly | Pro |
| [1032] | 195 |     |     | 200 |     |     |     | 205 |     |     |     |     |     |     |     |     |
| [1033] | Val | Leu | Lys | His | Gly | Asn | Phe | Glu | Ala | Arg | Gly | Ile | Ser | Arg | Glu | Gln |
| [1034] | 210 |     |     | 215 |     |     |     | 220 |     |     |     |     |     |     |     |     |
| [1035] | Val | Met | Ala | Phe | Ser | Thr | Arg | Arg | Lys | Glu | Val | Leu | Glu | Ala | Arg | Arg |
| [1036] | 225 |     |     | 230 |     |     |     | 235 |     |     |     | 240 |     |     |     |     |
| [1037] | Gly | Pro | Gly | Leu | Asp | Ala | Gly | Arg | Ile | Ala | Ala | Leu | Asp | Thr | Arg | Ala |
| [1038] | 245 |     |     | 250 |     |     |     | 255 |     |     |     |     |     |     |     |     |
| [1039] | Ser | Lys | Glu | Gly | Ile | Glu | Asp | Arg | Ala | Thr | Leu | Ser | Lys | Gln | Trp | Ser |
| [1040] | 260 |     |     | 265 |     |     |     | 270 |     |     |     |     |     |     |     |     |
| [1041] | Glu | Ala | Ala | Gln | Ser | Ile | Gly | Leu | Asp | Leu | Lys | Pro | Leu | Val | Asp | Arg |
| [1042] | 275 |     |     | 280 |     |     |     | 285 |     |     |     |     |     |     |     |     |
| [1043] | Ala | Arg | Thr | Lys | Ala | Leu | Gly | Gln | Gly | Met | Glu | Ala | Thr | Arg | Ile | Gly |
| [1044] | 290 |     |     | 295 |     |     |     | 300 |     |     |     |     |     |     |     |     |
| [1045] | Ser | Leu | Val | Glu | Arg | Gly | Arg | Ala | Trp | Leu | Ser | Arg | Phe | Ala | Ala | His |
| [1046] | 305 |     |     | 310 |     |     |     | 315 |     |     |     | 320 |     |     |     |     |
| [1047] | Val | Arg | Gly | Asp | Pro | Ala | Asp | Pro | Leu | Val | Pro | Pro | Ser | Val | Leu | Lys |
| [1048] | 325 |     |     | 330 |     |     |     | 335 |     |     |     |     |     |     |     |     |
| [1049] | Gln | Asp | Arg | Gln | Thr | Ile | Ala | Ala | Ala | Gln | Ala | Val | Ala | Ser | Ala | Val |

|        |   |     |     |
|--------|---|-----|-----|
| [1050] | 340   | 345 | 350 |
| [1051] | Arg His Leu Ser Gln Arg Glu Ala Ala Phe Glu Arg Thr Ala Leu Tyr |     |     |
| [1052] | 355   | 360 | 365 |
| [1053] | Lys Ala Ala Leu Asp Phe Gly Leu Pro Thr Thr Ile Ala Asp Val Glu |     |     |
| [1054] | 370   | 375 | 380 |
| [1055] | Lys Arg Thr Arg Ala Leu Val Arg Ser Gly Asp Leu Ile Ala Gly Lys |     |     |
| [1056] | 385   | 390 | 395 |
| [1057] | Gly Glu His Lys Gly Trp Leu Ala Ser Arg Asp Ala Val Val Thr Glu |     |     |
| [1058] | 405   | 410 | 415 |
| [1059] | Gln Arg Ile Leu Ser Glu Val Ala Ala Gly Lys Gly Asp Ser Ser Pro |     |     |
| [1060] | 420   | 425 | 430 |
| [1061] | Ala Ile Thr Pro Gln Lys Ala Ala Ala Ser Val Gln Ala Ala Ala Leu |     |     |
| [1062] | 435   | 440 | 445 |
| [1063] | Thr Gly Gln Gly Phe Arg Leu Asn Glu Gly Gln Leu Ala Ala Ala Arg |     |     |
| [1064] | 450   | 455 | 460 |
| [1065] | Leu Ile Leu Ile Ser Lys Asp Arg Thr Ile Ala Val Gln Gly Ile Ala |     |     |
| [1066] | 465   | 470 | 475 |
| [1067] | Gly Ala Gly Lys Ser Ser Val Leu Lys Pro Val Ala Glu Val Leu Arg |     |     |
| [1068] | 485   | 490 | 495 |
| [1069] | Asp Glu Gly His Pro Val Ile Gly Leu Ala Ile Gln Asn Thr Leu Val |     |     |
| [1070] | 500   | 505 | 510 |
| [1071] | Gln Met Leu Glu Arg Asp Thr Gly Ile Gly Ser Gln Thr Leu Ala Arg |     |     |
| [1072] | 515   | 520 | 525 |
| [1073] | Phe Leu Gly Gly Trp Asn Lys Leu Leu Asp Asp Pro Gly Asn Val Ala |     |     |
| [1074] | 530   | 535 | 540 |
| [1075] | Leu Arg Ala Glu Ala Gln Ala Ser Leu Lys Asp His Val Leu Val Leu |     |     |
| [1076] | 545   | 550 | 555 |
| [1077] | Asp Glu Ala Ser Met Val Ser Asn Glu Asp Lys Glu Lys Leu Val Arg |     |     |
| [1078] | 565   | 570 | 575 |
| [1079] | Leu Ala Asn Leu Ala Gly Val His Arg Leu Val Leu Ile Gly Asp Arg |     |     |
| [1080] | 580   | 585 | 590 |
| [1081] | Lys Gln Leu Gly Ala Val Asp Ala Gly Lys Pro Phe Ala Leu Leu Gln |     |     |
| [1082] | 595   | 600 | 605 |
| [1083] | Arg Ala Gly Ile Ala Arg Ala Glu Met Ala Thr Asn Leu Arg Ala Arg |     |     |
| [1084] | 610   | 615 | 620 |
| [1085] | Asp Pro Val Val Arg Glu Ala Gln Ala Ala Ala Gln Ala Gly Asp Val |     |     |
| [1086] | 625   | 630 | 635 |
| [1087] | Arg Lys Ala Leu Arg His Leu Lys Ser His Thr Val Glu Ala Arg Gly |     |     |
| [1088] | 645   | 650 | 655 |
| [1089] | Asp Gly Ala Gln Val Ala Ala Glu Thr Trp Leu Ala Leu Asp Lys Glu |     |     |
| [1090] | 660   | 665 | 670 |
| [1091] | Thr Arg Ala Arg Thr Ser Ile Tyr Ala Ser Gly Arg Ala Ile Arg Ser |     |     |

|        |   |     |     |
|--------|---|-----|-----|
| [1092] | 675   | 680 | 685 |
| [1093] | Ala Val Asn Ala Ala Val Gln Gln Gly Leu Leu Ala Ser Arg Glu Ile |     |     |
| [1094] | 690   | 695 | 700 |
| [1095] | Gly Pro Ala Lys Met Lys Leu Glu Val Leu Asp Arg Val Asn Thr Thr |     |     |
| [1096] | 705   | 710 | 715 |
| [1097] | Arg Glu Glu Leu Arg His Leu Pro Ala Tyr Arg Ala Gly Arg Val Leu |     |     |
| [1098] | 725   | 730 | 735 |
| [1099] | Glu Val Ser Arg Lys Gln Gln Ala Leu Gly Leu Phe Ile Gly Glu Tyr |     |     |
| [1100] | 740   | 745 | 750 |
| [1101] | Arg Val Ile Gly Gln Asp Arg Lys Gly Lys Leu Val Glu Val Glu Asp |     |     |
| [1102] | 755   | 760 | 765 |
| [1103] | Lys Arg Gly Lys Arg Phe Arg Phe Asp Pro Ala Arg Ile Arg Ala Gly |     |     |
| [1104] | 770   | 775 | 780 |
| [1105] | Lys Gly Asp Asp Asn Leu Thr Leu Leu Glu Pro Arg Lys Leu Glu Ile |     |     |
| [1106] | 785   | 790 | 795 |
| [1107] | His Glu Gly Asp Arg Ile Arg Trp Thr Arg Asn Asp His Arg Arg Gly |     |     |
| [1108] | 805   | 810 | 815 |
| [1109] | Leu Phe Asn Ala Asp Gln Ala Arg Val Val Glu Ile Ala Asn Gly Lys |     |     |
| [1110] | 820   | 825 | 830 |
| [1111] | Val Thr Phe Glu Thr Ser Lys Gly Asp Leu Val Glu Leu Lys Lys Asp |     |     |
| [1112] | 835   | 840 | 845 |
| [1113] | Asp Pro Met Leu Lys Arg Ile Asp Leu Ala Tyr Ala Leu Asn Val His |     |     |
| [1114] | 850   | 855 | 860 |
| [1115] | Met Ala Gln Gly Leu Thr Ser Asp Arg Gly Ile Ala Val Met Asp Ser |     |     |
| [1116] | 865   | 870 | 875 |
| [1117] | Arg Glu Arg Asn Leu Ser Asn Gln Lys Thr Phe Leu Val Thr Val Thr |     |     |
| [1118] | 885   | 890 | 895 |
| [1119] | Arg Leu Arg Asp His Leu Thr Leu Val Val Asp Ser Ala Asp Lys Leu |     |     |
| [1120] | 900   | 905 | 910 |
| [1121] | Gly Ala Ala Val Ala Arg Asn Lys Gly Glu Lys Ala Ser Ala Ile Glu |     |     |
| [1122] | 915   | 920 | 925 |
| [1123] | Val Thr Gly Ser Val Lys Pro Thr Ala Thr Lys Gly Ser Gly Val Asp |     |     |
| [1124] | 930   | 935 | 940 |
| [1125] | Gln Pro Lys Ser Val Glu Ala Asn Lys Ala Glu Lys Glu Leu Thr Arg |     |     |
| [1126] | 945   | 950 | 955 |
| [1127] | Ser Lys Ser Lys Thr Leu Asp Phe Gly Ile                         |     |     |
| [1128] | 965   | 970 |     |
| [1129] | <210> 16  |     |     |
| [1130] | <211> 40  |     |     |
| [1131] | <212> DNA   |     |     |
| [1132] | <213> 人工序列  |     |     |
| [1133] | <220>   |     |     |

- [1134] <223> 实施例1中使用的序列.  
[1135] <400> 16  
[1136] tttttttttt tttttttttt tttttttttt tttttttttt 40  
[1137] <210> 17  
[1138] <211> 40  
[1139] <212> RNA  
[1140] <213> 人工序列  
[1141] <220>  
[1142] <223> 实施例1中使用的序列.  
[1143] <400> 17  
[1144] gguuguuucu guuggugcug auauugcauu uaagagcauu 40  
[1145] <210> 18  
[1146] <211> 28  
[1147] <212> DNA  
[1148] <213> 人工序列  
[1149] <220>  
[1150] <223> 实施例1和3中使用的序列.  
[1151] <400> 18  
[1152] gcaatatcag caccaacaga aacaacct 28  
[1153] <210> 19  
[1154] <211> 55  
[1155] <212> RNA  
[1156] <213> 人工序列  
[1157] <220>  
[1158] <223> 实施例2中使用的序列.  
[1159] <400> 19  
[1160] gggaaauaag agagaccaca ccaccacac caccacacc acccacacca cacca 55  
[1161] <210> 20  
[1162] <211> 39  
[1163] <212> DNA  
[1164] <213> 人工序列  
[1165] <220>  
[1166] <223> 实施例2中使用的序列.  
[1167] <400> 20  
[1168] tctctcttat ttcccaggtt aaacacccaa gcagacgcc 39  
[1169] <210> 21  
[1170] <211> 24  
[1171] <212> DNA  
[1172] <213> 人工序列  
[1173] <220>  
[1174] <223> 实施例2中使用的序列.  
[1175] <400> 21

- [1176] ggcgtctgct tgggtgttta acct 24
- [1177] <210> 22
- [1178] <211> 40
- [1179] <212> DNA
- [1180] <213> 人工序列
- [1181] <220>
- [1182] <223> 实施例3中使用的序列.
- [1183] <400> 22
- [1184] ccccccccca ccccccccca ccccccccca ccccccccca 40
- [1185] <210> 23
- [1186] <211> 51
- [1187] <212> DNA
- [1188] <213> 人工序列
- [1189] <220>
- [1190] <223> 实施例3中使用的序列.
- [1191] <400> 23
- [1192] ggttggttct gttggtgctg atattgcggc gtctgcttgg gtgtttaacc t 51
- [1193] <210> 24
- [1194] <211> 32
- [1195] <212> DNA
- [1196] <213> 人工序列
- [1197] <220>
- [1198] <223> 实施例3中使用的序列.
- [1199] <400> 24
- [1200] tctctcttat cccaaacacc caagcagacg cc 32
- [1201] <210> 25
- [1202] <211> 24
- [1203] <212> DNA
- [1204] <213> 人工序列
- [1205] <220>
- [1206] <223> 实施例3中使用的序列.
- [1207] <400> 25
- [1208] ggcgtctgct tgggtgttta acct 24
- [1209] <210> 26
- [1210] <211> 1653
- [1211] <212> RNA
- [1212] <213> 人工序列
- [1213] <220>
- [1214] <223> 实施例3和5中使用的序列.
- [1215] <400> 26
- [1216] auggaagacg ccaaaaacau aaagaaaggc ccggcgccau ucuaucgcu ggaagaugga 60
- [1217] accgcuggag agcaacugca uaaggcuaug aagagauacg ccugguucc uggaacaauu 120



|        |  |               |
|--------|--|---------------|
| [1218] | gcuuuuacag augcacauau cgagguggac aucacuuacg cugaguacuu cgaaaugucc  | 180           |
| [1219] | guucgguugg cagaagcuau gaaacgauau gggcugaaua caaaucacag aaucgucgua  | 240           |
| [1220] | ugcagugaaa acucucuua auucuuuauug ccgguguugg gcgcguuuuu uaucggaguu  | 300           |
| [1221] | gcaguugcgc ccgcgaacga cauuuauauu gaacgugaau ugcucaacag uaugggcauu  | 360           |
| [1222] | ucgcagccua ccgugguguu cguuuccaaa aagggguugc aaaaauuuu gaacgugcaa   | 420           |
| [1223] | aaaaagcucc caaucaucca aaaaauuuu aucauggauu cuaaaacgga uuaccaggga   | 480           |
| [1224] | uuucagucga uguacacguu cgucacauau caucuaccuc ccgguuuuaa ugaauacgau  | 540           |
| [1225] | uuugugccag aguccuucga uagggacaag acaauugcac ugaucaugaa cuccucugga  | 600           |
| [1226] | ucuacugguc ugccuaaagg ugucgcucug ccucauagaa cugccugcgu gagauucug   | 660           |
| [1227] | caugccagag auccuauuuu uggcaaucaa aucauuccgg auacugcgau uuuuaguguu  | 720           |
| [1228] | guuccauucc aucacgguuu uggaauuuu acuacacucg gauauuugau auguggauuu   | 780           |
| [1229] | cgagucgucu uaauguauag auuugaagaa gagcuguuuc ugaggagccu ucaggauuac  | 840           |
| [1230] | aagauucaa gugcgcugcu ggugccaacc cuauucuccu ucuucgcaa aagcacucug    | 900           |
| [1231] | auugacaaau acgauuuuac uauuuuacac gaaauugcuu cugguggcgc ucccucucu   | 960           |
| [1232] | aaggaagucg gggaagcggu ugccaagagg uuccaucugc caggaucag gcaaggauau   | 1020          |
| [1233] | gggcucacug agacuacauc agcuauucug auuacacccg agggggauga uaaaccgggc  | 1080          |
| [1234] | gcggucggua aaguuguucc auuuuuugaa gcgaagguug uggaucugga uaccgggaaa  | 1140          |
| [1235] | acgcugggcg uuaaucaaag aggcgaacug ugugugagag guccuau gau uauguccggu | 1200          |
| [1236] | uauguuaaca auccggaagc gaccaacgcc uugauugaca aggauggaug gcuacauuc   | 1260          |
| [1237] | ggagacauag cuuacuggga cgaagacgaa cacuucua ucguugaccg ccugaagucu    | 1320          |
| [1238] | cugauuaagu acaaaggcu ucagguggcu cccgcugaau uggaauccau cuugcuccaa   | 1380          |
| [1239] | cacccaaca ucuucgacgc aggugucgca ggucuucccg acgaugacgc cggugaacuu   | 1440          |
| [1240] | cccgccgcg uuguuuuuu ggagcacgga aagacgauga cggaaaaaga gaucguggau    | 1500          |
| [1241] | uacgucgcca gucaaguaac aaccgcgaaa aaguugcgcg gaggaguugu guuugggac   | 1560          |
| [1242] | gaaguaccga aaggucuua cggaaaacuc gacgcaagaa aaaucaagaga gauccucau   | 1620          |
| [1243] | aaggccaaga agggcggaau gaucgccgug uaa                               | 1653          |
| [1244] | <210>  | 27            |
| [1245] | <211>  | 58            |
| [1246] | <212>  | DNA           |
| [1247] | <213>  | 人工序列          |
| [1248] | <220>  |               |
| [1249] | <223>  | 实施例4和5中使用的序列。 |
| [1250] | <400>  | 27            |
| [1251] | ccccccccca cccccccca cccccccca agaaacataa acagaacgtt tttttttt      | 58            |
| [1252] | <210>  | 28            |
| [1253] | <211>  | 51            |
| [1254] | <212>  | DNA           |
| [1255] | <213>  | 人工序列          |
| [1256] | <220>  |               |
| [1257] | <223>  | 实施例4中使用的序列。   |
| [1258] | <400>  | 28            |
| [1259] | ccccccccca cccccccca cccccccca agaaacataa acagaacgtt t             | 51            |

- [1260] <210> 29  
[1261] <211> 48  
[1262] <212> DNA  
[1263] <213> 人工序列  
[1264] <220>  
[1265] <223> 实施例4和5中使用的序列.  
[1266] <400> 29  
[1267] cgttctgttt atgtttcttg gacactgatt gacacggttt agtagaac 48  
[1268] <210> 30  
[1269] <211> 288  
[1270] <212> RNA  
[1271] <213> 人工序列  
[1272] <220>  
[1273] <223> 实施例4和6中使用的序列.  
[1274] <400> 30  
[1275] gggaaauaag agagaaaaga agaguaagaa gaaauauaag agccaccaug guuuauauug 60  
[1276] cggccgcuaa auuaagcugc cuucugcggg gcuugccuuc uggccaugcc cuucucucu 120  
[1277] cccuugcacc uguaccucu ggucuuugaa uaaagccuga guaggaagaa aaaaaaaaaa 180  
[1278] aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 240  
[1279] aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa 288  
[1280] <210> 31  
[1281] <211> 10  
[1282] <212> DNA  
[1283] <213> 人工序列  
[1284] <220>  
[1285] <223> 实施例6中使用的序列.  
[1286] <400> 31  
[1287] ccccccccca 10  
[1288] <210> 32  
[1289] <211> 27  
[1290] <212> DNA  
[1291] <213> 人工序列  
[1292] <220>  
[1293] <223> 实施例6中使用的序列.  
[1294] <400> 32  
[1295] ggttgtttct gttggtgctg atattgc 27  
[1296] <210> 33  
[1297] <211> 12  
[1298] <212> DNA  
[1299] <213> 人工序列  
[1300] <220>  
[1301] <223> 用于说明均聚物读取的序列.

|        |  |
|--------|--|
| [1302] | <400> 33                               |
| [1303] | aaaaaaaaaa aa 12                       |
| [1304] | <210> 34                               |
| [1305] | <211> 12                               |
| [1306] | <212> DNA                              |
| [1307] | <213> 人工序列                             |
| [1308] | <220>                                  |
| [1309] | <223> 用于说明均聚物读取的序列.                    |
| [1310] | <220>                                  |
| [1311] | <221> n                                |
| [1312] | <222> (4) .. (4)                       |
| [1313] | <223> 肌苷                               |
| [1314] | <220>                                  |
| [1315] | <221> misc_feature                     |
| [1316] | <222> (4) .. (4)                       |
| [1317] | <223> n是a,c,g,或t                       |
| [1318] | <220>                                  |
| [1319] | <221> n                                |
| [1320] | <222> (7) .. (8)                       |
| [1321] | <223> 肌苷                               |
| [1322] | <220>                                  |
| [1323] | <221> misc_feature                     |
| [1324] | <222> (7) .. (8)                       |
| [1325] | <223> n是a,c,g,或t                       |
| [1326] | <220>                                  |
| [1327] | <221> n                                |
| [1328] | <222> (12) .. (12)                     |
| [1329] | <223> 肌苷                               |
| [1330] | <220>                                  |
| [1331] | <221> misc_feature                     |
| [1332] | <222> (12) .. (12)                     |
| [1333] | <223> n是a,c,g,或t                       |
| [1334] | <400> 34                               |
| [1335] | tttnttnntt tn 12                       |
| [1336] | <210> 35                               |
| [1337] | <211> 32                               |
| [1338] | <212> DNA                              |
| [1339] | <213> 人工序列                             |
| [1340] | <220>                                  |
| [1341] | <223> 实施例7中使用的序列.                      |
| [1342] | <400> 35                               |
| [1343] | gcaatatcag caccaacaga aarcrararc rc 32 |

- [1344] <210> 36  
[1345] <211> 48  
[1346] <212> DNA  
[1347] <213> 人工序列  
[1348] <220>  
[1349] <223> 实施例7中使用的序列.  
[1350] <400> 36  
[1351] ttctgaatta ggaaccattt tttttttttt atgatgcaag atacgcac 48  
[1352] <210> 37  
[1353] <211> 36  
[1354] <212> DNA  
[1355] <213> 人工序列  
[1356] <220>  
[1357] <223> 实施例7中使用的序列.  
[1358] <400> 37  
[1359] gaggcgagcg gtcaatttgt gcgtatcttg catcat 36  
[1360] <210> 38  
[1361] <211> 18  
[1362] <212> DNA  
[1363] <213> 人工序列  
[1364] <220>  
[1365] <223> 39  
[1366] <400> 38  
[1367] atggttccta attcagaa 18  
[1368] <210> 39  
[1369] <211> 15  
[1370] <212> DNA  
[1371] <213> 人工序列  
[1372] <220>  
[1373] <223> 实施例7中使用的序列.  
[1374] <400> 39  
[1375] ttgaccgctc gcctc 15  
[1376] <210> 40  
[1377] <211> 37  
[1378] <212> DNA  
[1379] <213> 人工序列  
[1380] <220>  
[1381] <223> 实施例7中使用的序列.  
[1382] <400> 40  
[1383] ggttgtttct gttggtgctg atattgcttt ttttttt 37  
[1384] <210> 41  
[1385] <211> 897

|        |  |
|--------|--|
| [1386] | <212> DNA  |
| [1387] | <213> 赤子爱胜蚓 (Eisenia fetida)   |
| [1388] | <400> 41   |
| [1389] | atgagtgcga aggctgctga aggttatgaa caaatcgaag ttgatgtggt tgctgtgtgg 60   |
| [1390] | aaggaagggtt atgtgtatga aaatcgtggt agtacctccg tggatcaaaa aattaccatc 120 |
| [1391] | acgaaaggca tgaagaacgt taatagcgaa acccgtagcg tcaccgcgac gcattctatt 180  |
| [1392] | ggcagtacca tctccacggg tgacgccttt gaaatcggt ccgtaggaagt ttcattattcg 240 |
| [1393] | catagccacg aagaatcaca agtttcgatg accgaaacgg aagtctacga atcaaaagtg 300  |
| [1394] | attgaacaca ccattacgat cccgccgacc tcgaagttca cgcgctggca gctgaacgca 360  |
| [1395] | gatgtcggcg gtgctgacat tgaatatatg tacctgatcg atgaagttac cccgattggc 420  |
| [1396] | ggtacgcaga gtattccgca agtgatcacc tcccgtagca aaattatcgt tggctgccag 480  |
| [1397] | attatcctgg gcaagaccga aattcgtatc aaacatgctg aacgcaagga atatatgacc 540  |
| [1398] | gtggttagcc gtaaactctg gccggcgccc acgctgggtc acagtaaact gtttaagttc 600  |
| [1399] | gtgctgtacg aagattgggg cggttttcgc atcaaaaccc tgaatacgat gtattctggt 660  |
| [1400] | tatgaatacg cgtatagctc tgaccagggc ggtatctact tcgatcaagg caccgacaac 720  |
| [1401] | ccgaaacagc gttgggccat taataagagc ctgccgctgc gccatggtga tgctgtgacc 780  |
| [1402] | tttatgaaca aatacttcac gcgttctggt ctgtgctatg atgacggccc ggcgaccaat 840  |
| [1403] | gtgtattgtc tggataaacg cgaagacaag tggattctgg aagttgtcgg ctaatga 897     |
| [1404] | <210> 42   |
| [1405] | <211> 297  |
| [1406] | <212> PRT  |
| [1407] | <213> 赤子爱胜蚓  |
| [1408] | <400> 42   |
| [1409] | Met Ser Ala Lys Ala Ala Glu Gly Tyr Glu Gln Ile Glu Val Asp Val        |
| [1410] | 1 5 10 15  |
| [1411] | Val Ala Val Trp Lys Glu Gly Tyr Val Tyr Glu Asn Arg Gly Ser Thr        |
| [1412] | 20 25 30   |
| [1413] | Ser Val Asp Gln Lys Ile Thr Ile Thr Lys Gly Met Lys Asn Val Asn        |
| [1414] | 35 40 45   |
| [1415] | Ser Glu Thr Arg Thr Val Thr Ala Thr His Ser Ile Gly Ser Thr Ile        |
| [1416] | 50 55 60   |
| [1417] | Ser Thr Gly Asp Ala Phe Glu Ile Gly Ser Val Glu Val Ser Tyr Ser        |
| [1418] | 65 70 75 80  |
| [1419] | His Ser His Glu Glu Ser Gln Val Ser Met Thr Glu Thr Glu Val Tyr        |
| [1420] | 85 90 95   |
| [1421] | Glu Ser Lys Val Ile Glu His Thr Ile Thr Ile Pro Pro Thr Ser Lys        |
| [1422] | 100 105 110  |
| [1423] | Phe Thr Arg Trp Gln Leu Asn Ala Asp Val Gly Gly Ala Asp Ile Glu        |
| [1424] | 115 120 125  |
| [1425] | Tyr Met Tyr Leu Ile Asp Glu Val Thr Pro Ile Gly Gly Thr Gln Ser        |
| [1426] | 130 135 140  |
| [1427] | Ile Pro Gln Val Ile Thr Ser Arg Ala Lys Ile Ile Val Gly Arg Gln        |

|        |   |     |     |     |
|--------|---|-----|-----|-----|
| [1428] | 145   | 150 | 155 | 160 |
| [1429] | Ile Ile Leu Gly Lys Thr Glu Ile Arg Ile Lys His Ala Glu Arg Lys     |     |     |     |
| [1430] |   | 165 | 170 | 175 |
| [1431] | Glu Tyr Met Thr Val Val Ser Arg Lys Ser Trp Pro Ala Ala Thr Leu     |     |     |     |
| [1432] |   | 180 | 185 | 190 |
| [1433] | Gly His Ser Lys Leu Phe Lys Phe Val Leu Tyr Glu Asp Trp Gly Gly     |     |     |     |
| [1434] |   | 195 | 200 | 205 |
| [1435] | Phe Arg Ile Lys Thr Leu Asn Thr Met Tyr Ser Gly Tyr Glu Tyr Ala     |     |     |     |
| [1436] |   | 210 | 215 | 220 |
| [1437] | Tyr Ser Ser Asp Gln Gly Gly Ile Tyr Phe Asp Gln Gly Thr Asp Asn     |     |     |     |
| [1438] | 225   | 230 | 235 | 240 |
| [1439] | Pro Lys Gln Arg Trp Ala Ile Asn Lys Ser Leu Pro Leu Arg His Gly     |     |     |     |
| [1440] |   | 245 | 250 | 255 |
| [1441] | Asp Val Val Thr Phe Met Asn Lys Tyr Phe Thr Arg Ser Gly Leu Cys     |     |     |     |
| [1442] |   | 260 | 265 | 270 |
| [1443] | Tyr Asp Asp Gly Pro Ala Thr Asn Val Tyr Cys Leu Asp Lys Arg Glu     |     |     |     |
| [1444] |   | 275 | 280 | 285 |
| [1445] | Asp Lys Trp Ile Leu Glu Val Val Gly                                 |     |     |     |
| [1446] |   | 290 | 295 |     |
| [1447] | <210> 43  |     |     |     |
| [1448] | <211> 786   |     |     |     |
| [1449] | <212> DNA   |     |     |     |
| [1450] | <213> 大肠杆菌  |     |     |     |
| [1451] | <400> 43  |     |     |     |
| [1452] | tgtctgaccg caccgccgaa agaagcggca cgtccgaccc tgatgccgcg tgcacagtct   | 60  |     |     |
| [1453] | tataaagatc tgaccatct gccggtccg acgggcaaaa tttttgtag cgtctataac      | 120 |     |     |
| [1454] | atccaggacg aaaccggtca atttaaaccg taccggcgca gtaatttctc cacggccgtt   | 180 |     |     |
| [1455] | ccgcagagtg caaccgctat gctggctcac gcaactgaaag attcccgttg gttcattccg  | 240 |     |     |
| [1456] | ctggaacgcc agggcctgca aaacctgctg aatgaacgta aaattatccg cgcagctcag   | 300 |     |     |
| [1457] | gaaaacggta ccgtggccat taacaatcgt attccgctgc aaagcctgac cgccgcaaac   | 360 |     |     |
| [1458] | atcatggttg aaggctctat catcggttac gaatcaaacg tcaaactcggg cgggtgtgggc | 420 |     |     |
| [1459] | gcacgttatt ttggcattgg tgctgatacc cagtaccaac tggaccagat cgcagttaac   | 480 |     |     |
| [1460] | ctgcgcgtgg ttaatgtcag caccggcgaa attctgagct ctgtgaatac cagcaaaacg   | 540 |     |     |
| [1461] | atcctgtctt acgaagtgca ggctgggtgt tttcgtttca ttgattatca acgcctgctg   | 600 |     |     |
| [1462] | gaaggcgaag tcggttacac ctcaaacgaa ccggtgatgc tgtgtctgat gtcggcgatt   | 660 |     |     |
| [1463] | gaaacgggtg ttattttcct gatcaatgat ggcacgacc gtggtctgtg ggatctgcag    | 720 |     |     |
| [1464] | aacaaagccg aacgtcaaaa tgacattctg gtgaaatacc gccacatgag tgttccgccg   | 780 |     |     |
| [1465] | gaatcc  | 786 |     |     |
| [1466] | <210> 44  |     |     |     |
| [1467] | <211> 262   |     |     |     |
| [1468] | <212> PRT   |     |     |     |
| [1469] | <213> 大肠杆菌  |     |     |     |

|        |   |
|--------|---|
| [1470] | <400> 44  |
| [1471] | Cys Leu Thr Ala Pro Pro Lys Glu Ala Ala Arg Pro Thr Leu Met Pro |
| [1472] | 1 5 10 15   |
| [1473] | Arg Ala Gln Ser Tyr Lys Asp Leu Thr His Leu Pro Ala Pro Thr Gly |
| [1474] | 20 25 30  |
| [1475] | Lys Ile Phe Val Ser Val Tyr Asn Ile Gln Asp Glu Thr Gly Gln Phe |
| [1476] | 35 40 45  |
| [1477] | Lys Pro Tyr Pro Ala Ser Asn Phe Ser Thr Ala Val Pro Gln Ser Ala |
| [1478] | 50 55 60  |
| [1479] | Thr Ala Met Leu Val Thr Ala Leu Lys Asp Ser Arg Trp Phe Ile Pro |
| [1480] | 65 70 75 80   |
| [1481] | Leu Glu Arg Gln Gly Leu Gln Asn Leu Leu Asn Glu Arg Lys Ile Ile |
| [1482] | 85 90 95  |
| [1483] | Arg Ala Ala Gln Glu Asn Gly Thr Val Ala Ile Asn Asn Arg Ile Pro |
| [1484] | 100 105 110   |
| [1485] | Leu Gln Ser Leu Thr Ala Ala Asn Ile Met Val Glu Gly Ser Ile Ile |
| [1486] | 115 120 125   |
| [1487] | Gly Tyr Glu Ser Asn Val Lys Ser Gly Gly Val Gly Ala Arg Tyr Phe |
| [1488] | 130 135 140   |
| [1489] | Gly Ile Gly Ala Asp Thr Gln Tyr Gln Leu Asp Gln Ile Ala Val Asn |
| [1490] | 145 150 155 160   |
| [1491] | Leu Arg Val Val Asn Val Ser Thr Gly Glu Ile Leu Ser Ser Val Asn |
| [1492] | 165 170 175   |
| [1493] | Thr Ser Lys Thr Ile Leu Ser Tyr Glu Val Gln Ala Gly Val Phe Arg |
| [1494] | 180 185 190   |
| [1495] | Phe Ile Asp Tyr Gln Arg Leu Leu Glu Gly Glu Val Gly Tyr Thr Ser |
| [1496] | 195 200 205   |
| [1497] | Asn Glu Pro Val Met Leu Cys Leu Met Ser Ala Ile Glu Thr Gly Val |
| [1498] | 210 215 220   |
| [1499] | Ile Phe Leu Ile Asn Asp Gly Ile Asp Arg Gly Leu Trp Asp Leu Gln |
| [1500] | 225 230 235 240   |
| [1501] | Asn Lys Ala Glu Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg His Met |
| [1502] | 245 250 255   |
| [1503] | Ser Val Pro Pro Glu Ser   |
| [1504] | 260   |
| [1505] | <210> 45  |
| [1506] | <211> 10  |
| [1507] | <212> PRT   |
| [1508] | <213> 人工序列  |
| [1509] | <220>   |
| [1510] | <223> StepII (C) 的氨基酸序列.  |
| [1511] | <400> 45  |

[1512] Ser Ala Trp Ser His Pro Gln Phe Glu Lys  
[1513] 1 5 10  
[1514] <210> 46  
[1515] <211> 40  
[1516] <212> DNA  
[1517] <213> 人工序列  
[1518] <220>  
[1519] <223> 说明书中描述的序列.  
[1520] <400> 46  
[1521] ccccccccca ccccccccca ccccccccca ccccccccca 40  
[1522] <210> 47  
[1523] <211> 20  
[1524] <212> DNA  
[1525] <213> 人工序列  
[1526] <220>  
[1527] <223> 说明书中描述的序列.  
[1528] <400> 47  
[1529] actcgcagat cattacgatc 20  
[1530] <210> 48  
[1531] <211> 39  
[1532] <212> RNA  
[1533] <213> 人工序列  
[1534] <220>  
[1535] <223> 说明书中描述的序列.  
[1536] <400> 48  
[1537] rarcrcrurg rcrcrgraru rcrcrurura rcrgarurc 39  
[1538] <210> 49  
[1539] <211> 20  
[1540] <212> DNA  
[1541] <213> 人工序列  
[1542] <220>  
[1543] <223> 说明书中描述的序列.  
[1544] <400> 49  
[1545] cgattgacta agctatacgc 20  
[1546] <210> 50  
[1547] <211> 40  
[1548] <212> RNA  
[1549] <213> 人工序列  
[1550] <220>  
[1551] <223> 说明书中描述的序列.  
[1552] <400> 50  
[1553] rcrcrgruru rgrarcrcrura rargrcrcrura rurarcrgrc 40



[1554] <210> 51  
[1555] <211> 10  
[1556] <212> DNA  
[1557] <213> 人工序列  
[1558] <220>  
[1559] <223> 说明书中描述的序列.  
[1560] <400> 51  
[1561] tttttttttt 10  
[1562] <210> 52  
[1563] <211> 20  
[1564] <212> DNA  
[1565] <213> 人工序列  
[1566] <220>  
[1567] <223> 说明书中描述的序列.  
[1568] <400> 52  
[1569] tttttttttt tttttttttt 20  
[1570] <210> 53  
[1571] <211> 20  
[1572] <212> DNA  
[1573] <213> 人工序列  
[1574] <220>  
[1575] <223> 说明书中描述的序列.  
[1576] <400> 53  
[1577] ccccccccca ccccccccca 20  
[1578] <210> 54  
[1579] <211> 20  
[1580] <212> RNA  
[1581] <213> 人工序列  
[1582] <220>  
[1583] <223> 说明书中描述的序列.  
[1584] <400> 54  
[1585] rarararara rarararara 20  
[1586] <210> 55  
[1587] <211> 20  
[1588] <212> RNA  
[1589] <213> 人工序列  
[1590] <220>  
[1591] <223> 说明书中描述的序列.  
[1592] <400> 55  
[1593] rurrcraru rarcrgrara 20  
[1594] <210> 56  
[1595] <211> 30

- [1596] <212> DNA  
[1597] <213> 人工序列  
[1598] <220>  
[1599] <223> 说明书中描述的序列.  
[1600] <400> 56  
[1601] aactactagg atcatcgatg tatctgctca 30  
[1602] <210> 57  
[1603] <211> 30  
[1604] <212> DNA  
[1605] <213> 人工序列  
[1606] <220>  
[1607] <223> 说明书中描述的序列.  
[1608] <400> 57  
[1609] agcttaacat acgatactct tagctaacca 30  
[1610] <210> 58  
[1611] <211> 60  
[1612] <212> RNA  
[1613] <213> 人工序列  
[1614] <220>  
[1615] <223> 说明书中描述的序列.  
[1616] <400> 58  
[1617] rararcrrura rcrurargrg rarurcraru rcrgrarurg rurarurcru rgcrurcra 60  
[1618] <210> 59  
[1619] <211> 60  
[1620] <212> RNA  
[1621] <213> 人工序列  
[1622] <220>  
[1623] <223> 说明书中描述的序列.  
[1624] <400> 59  
[1625] rargrcruru rararcrraru rarcgraru rarcrrurcu rurargrcru rararcrcra 60  
[1626] <210> 60  
[1627] <211> 10  
[1628] <212> DNA  
[1629] <213> 人工序列  
[1630] <220>  
[1631] <223> 说明书中描述的序列.  
[1632] <400> 60  
[1633] actctgaacc 10  
[1634] <210> 61  
[1635] <211> 15  
[1636] <212> DNA  
[1637] <213> 人工序列

- [1638] <220>  
[1639] <223> 说明书中描述的序列.  
[1640] <400> 61  
[1641] actctrgrar arcrc 15  
[1642] <210> 62  
[1643] <211> 10  
[1644] <212> DNA  
[1645] <213> 人工序列  
[1646] <220>  
[1647] <223> 说明书中描述的序列.  
[1648] <400> 62  
[1649] gcacaatgat 10  
[1650] <210> 63  
[1651] <211> 15  
[1652] <212> DNA  
[1653] <213> 人工序列  
[1654] <220>  
[1655] <223> 说明书中描述的序列.  
[1656] <400> 63  
[1657] gcacarartr grart 15  
[1658] <210> 64  
[1659] <211> 20  
[1660] <212> RNA  
[1661] <213> 人工序列  
[1662] <220>  
[1663] <223> 说明书中描述的序列.  
[1664] <400> 64  
[1665] rarararcru rarcgrcru 20  
[1666] <210> 65  
[1667] <211> 32  
[1668] <212> DNA  
[1669] <213> 人工序列  
[1670] <220>  
[1671] <223> 实施例7中使用的序列.  
[1672] <400> 65  
[1673] ttgttctact aaaccgtgtc aatcagtgtc tt 32  
[1674] <210> 66  
[1675] <211> 29  
[1676] <212> DNA  
[1677] <213> 人工序列  
[1678] <220>  
[1679] <223> 实施例7中使用的序列.

[1680] <400> 66

[1681] gcaatatcag caccaacaga aacaacctt 29

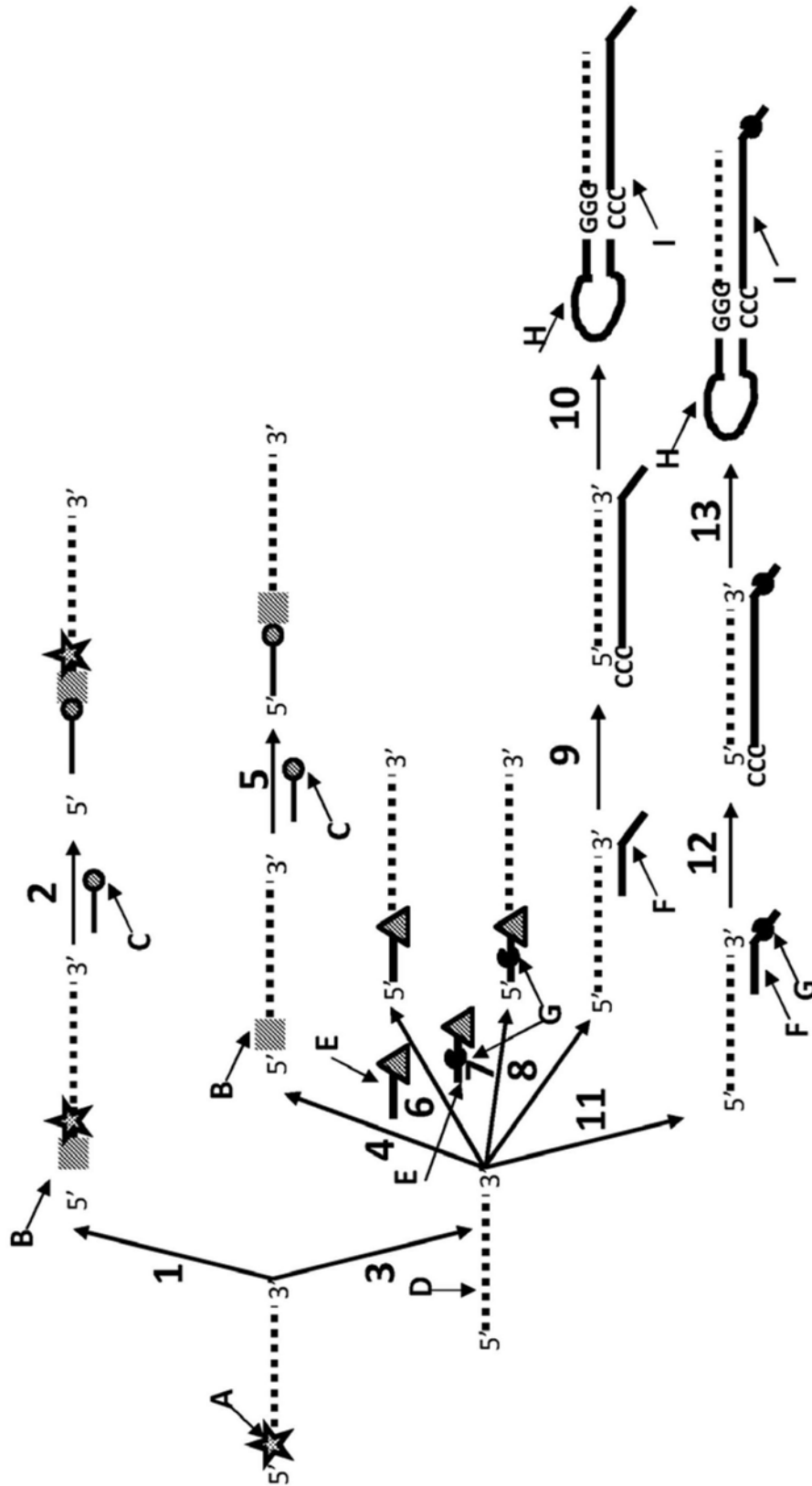


图1

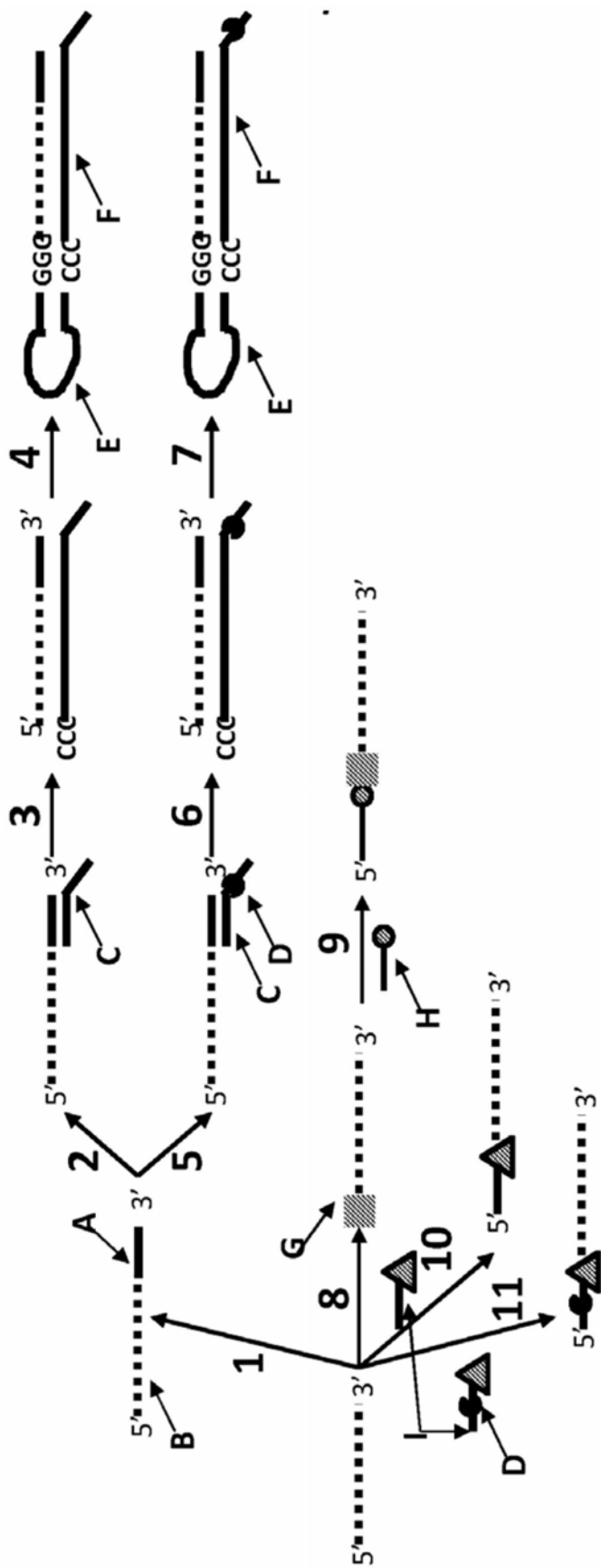


图2

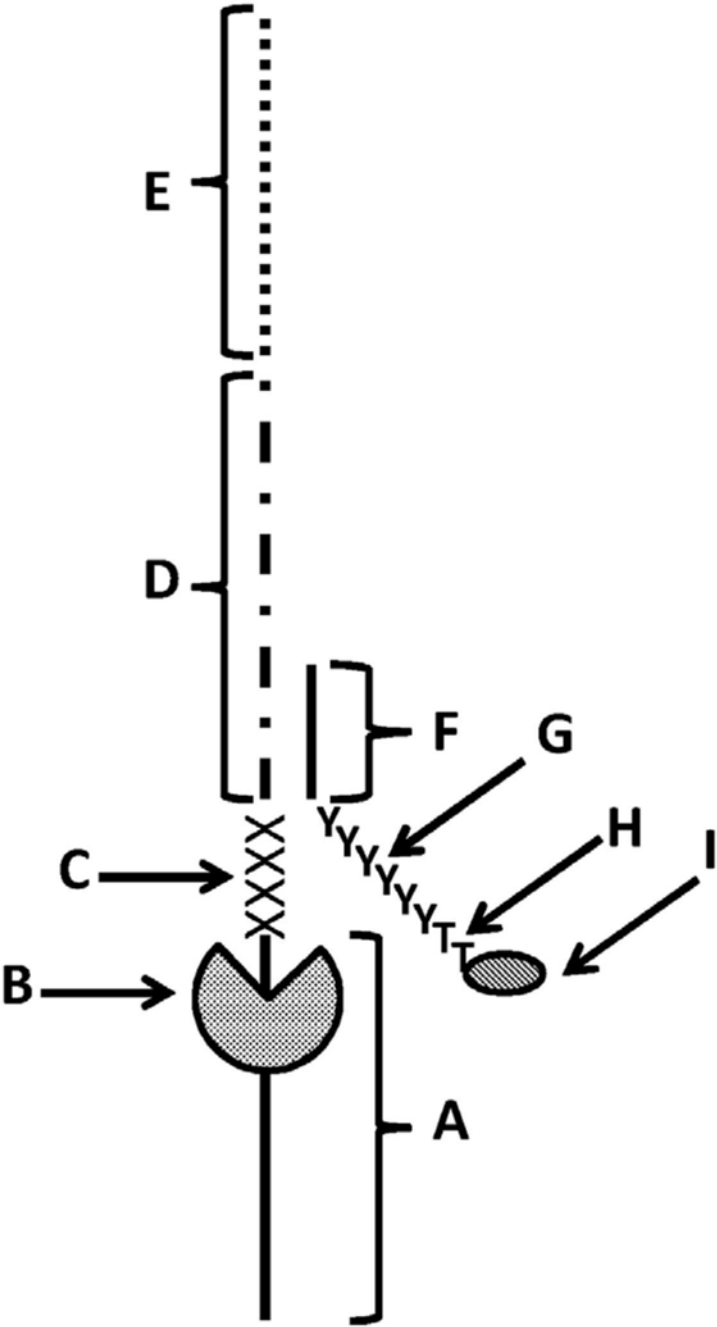


图3

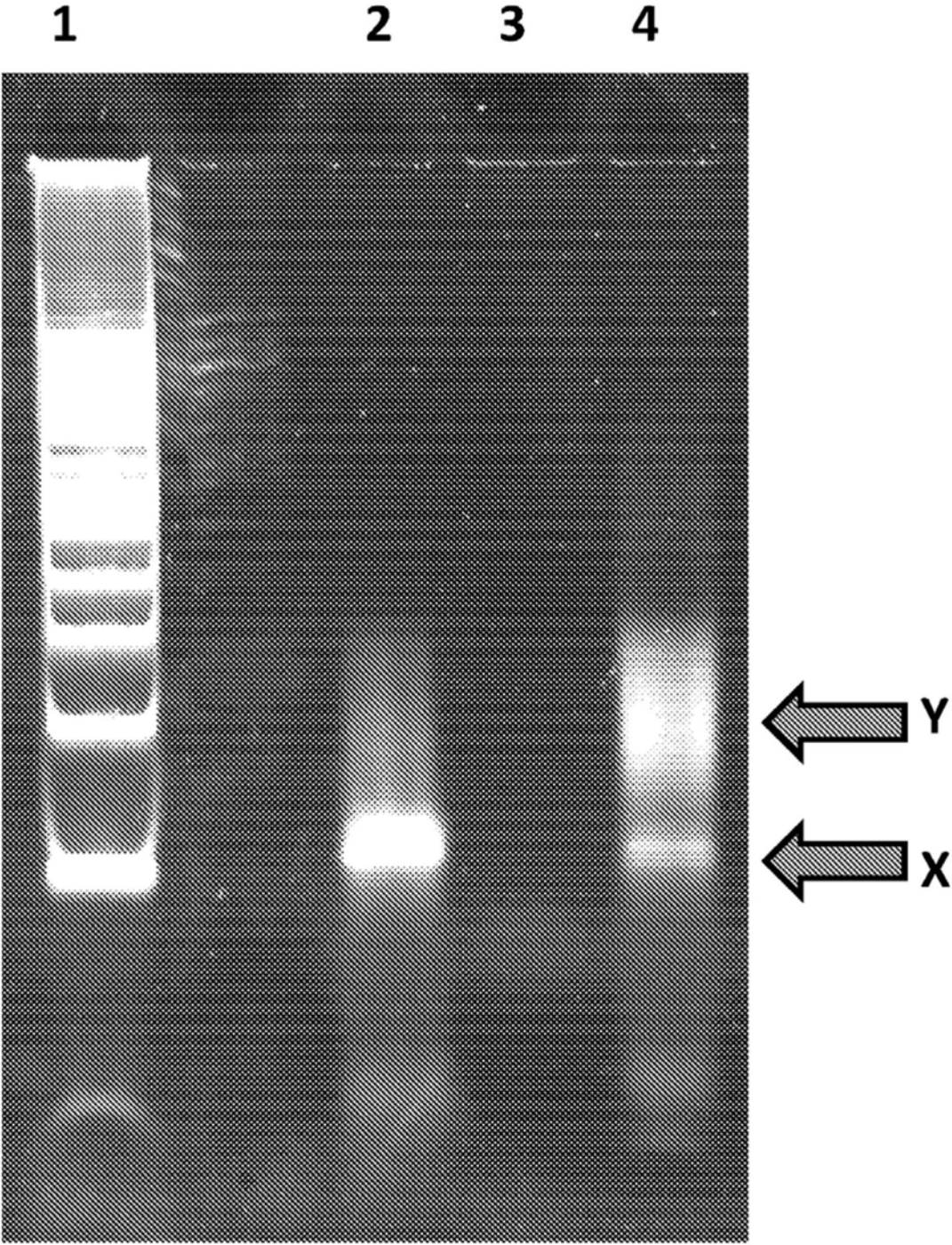


图4



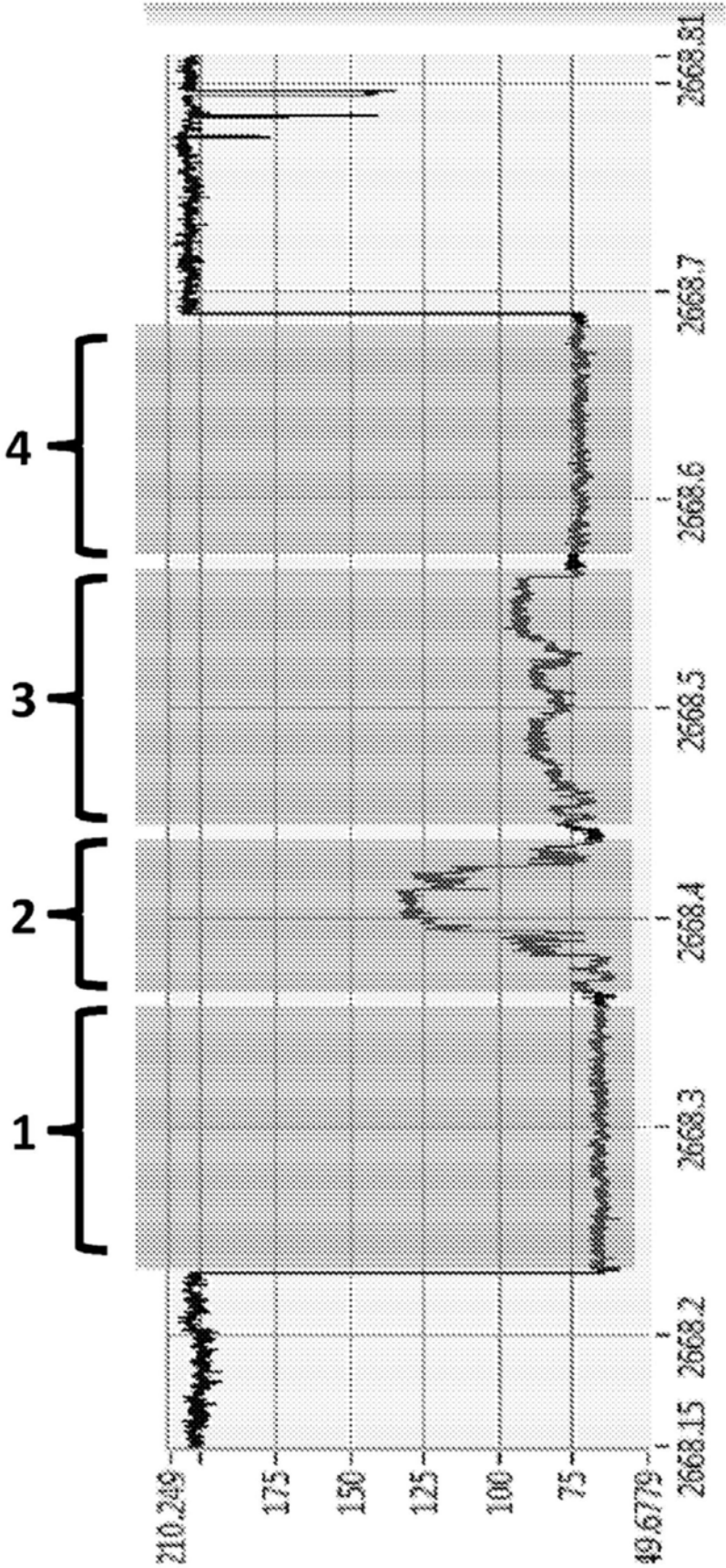


图5

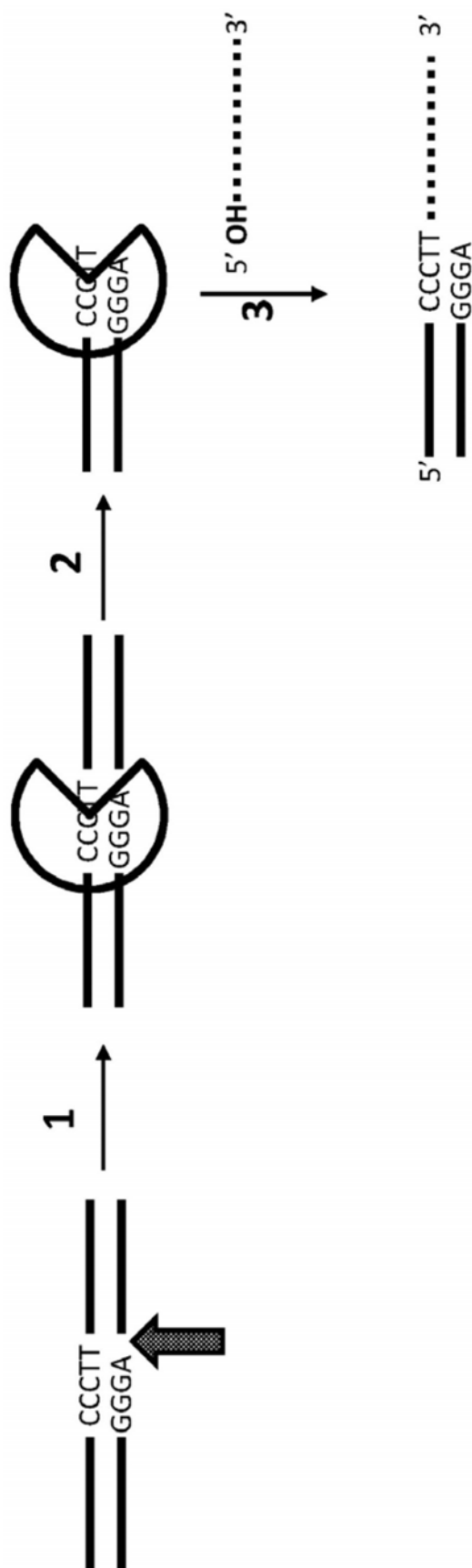


图6

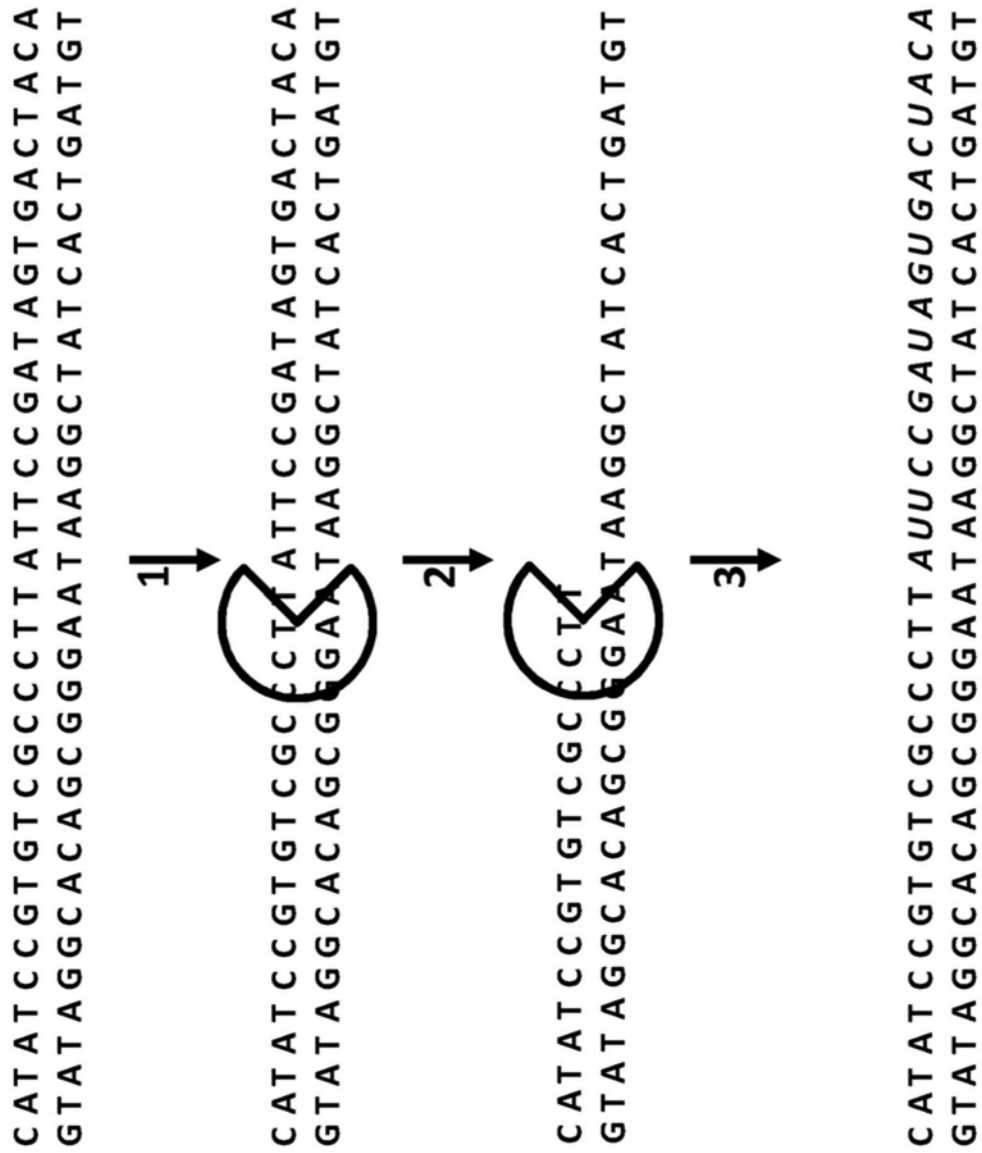


图7

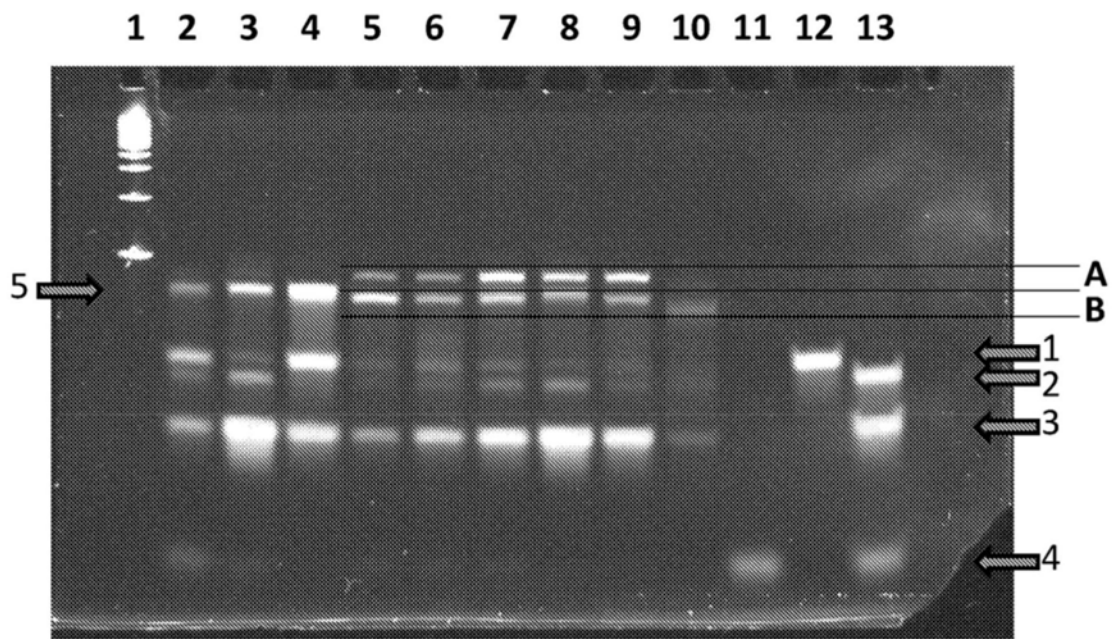


图8

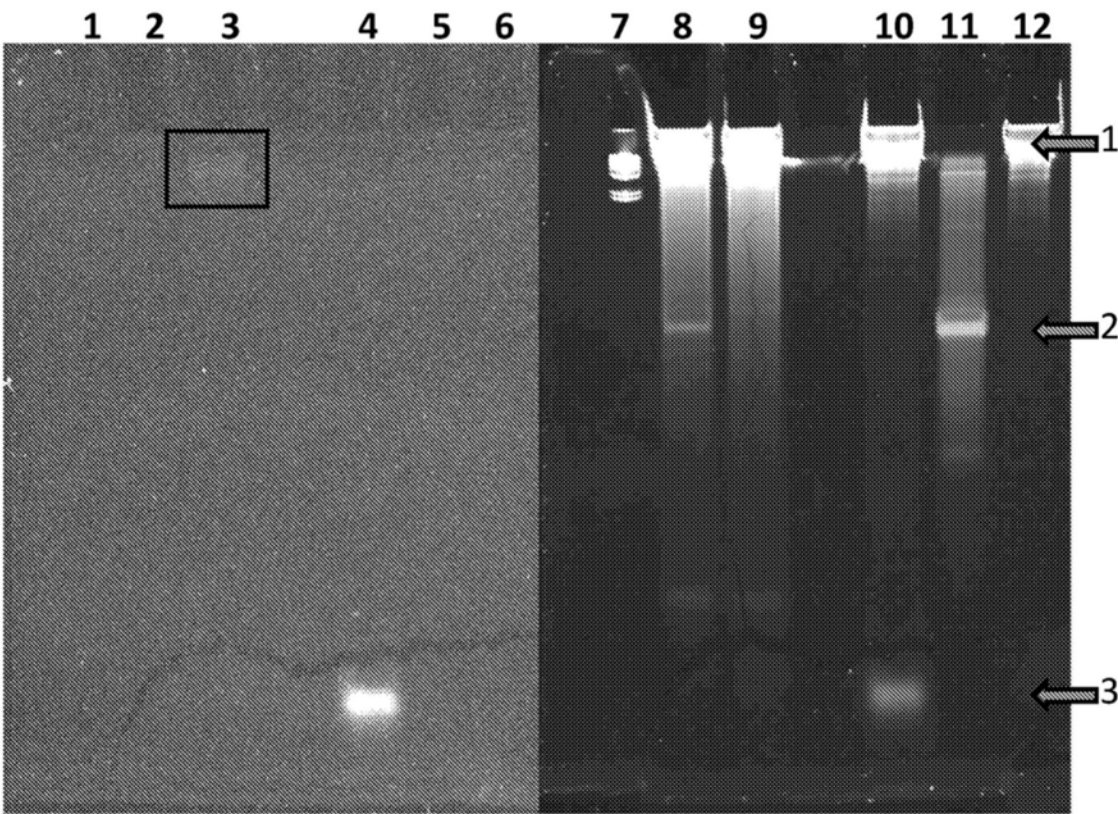


图9

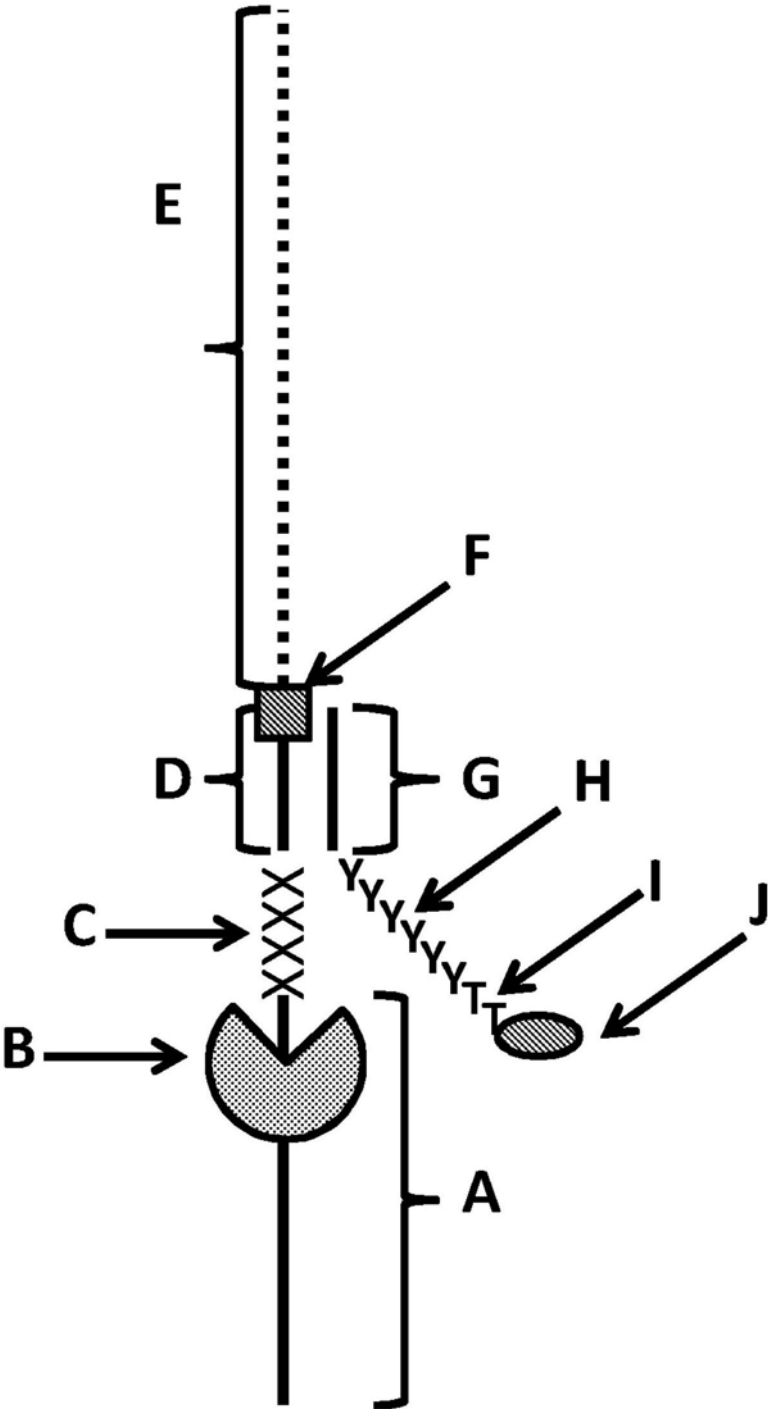


图10

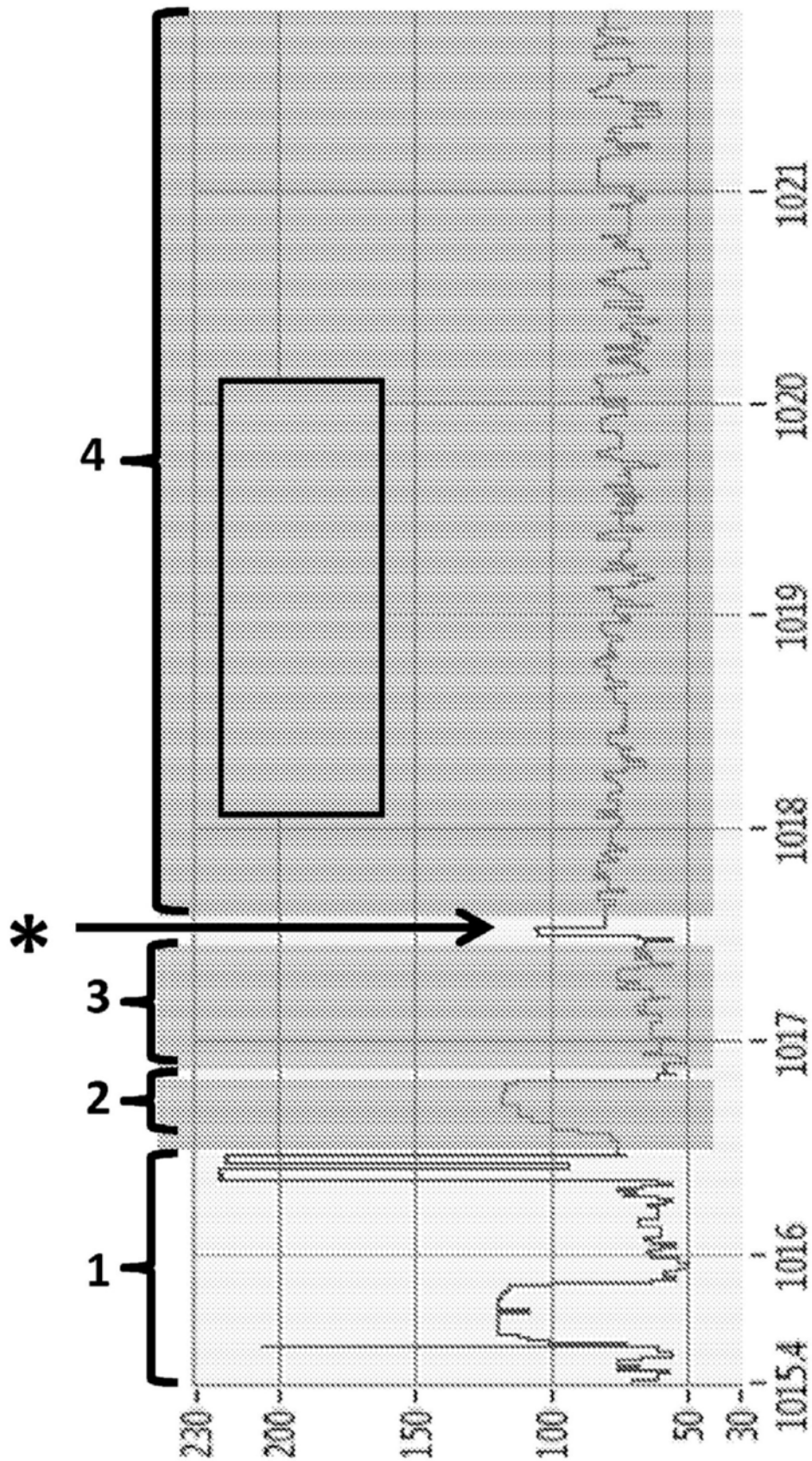


图11



图12

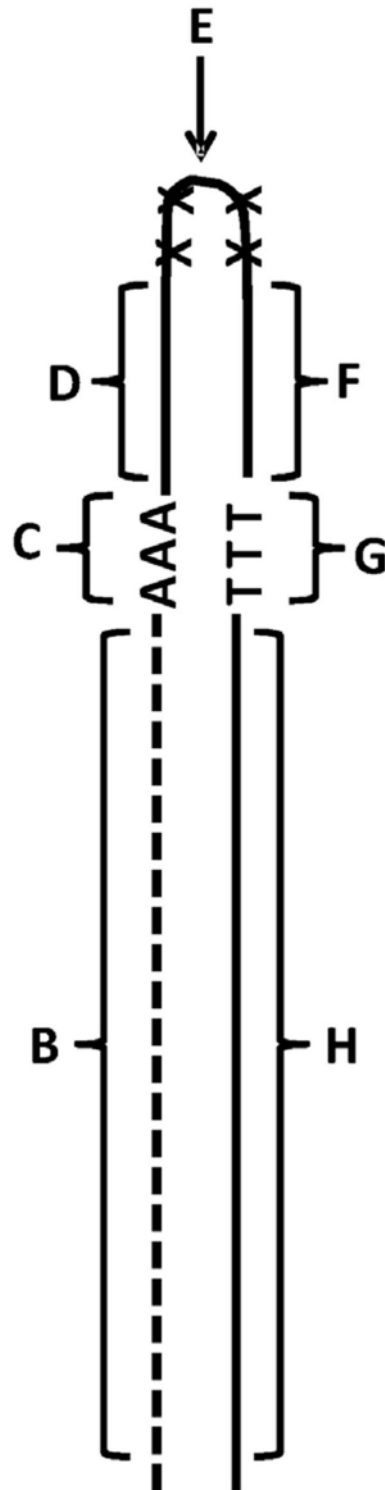


图13



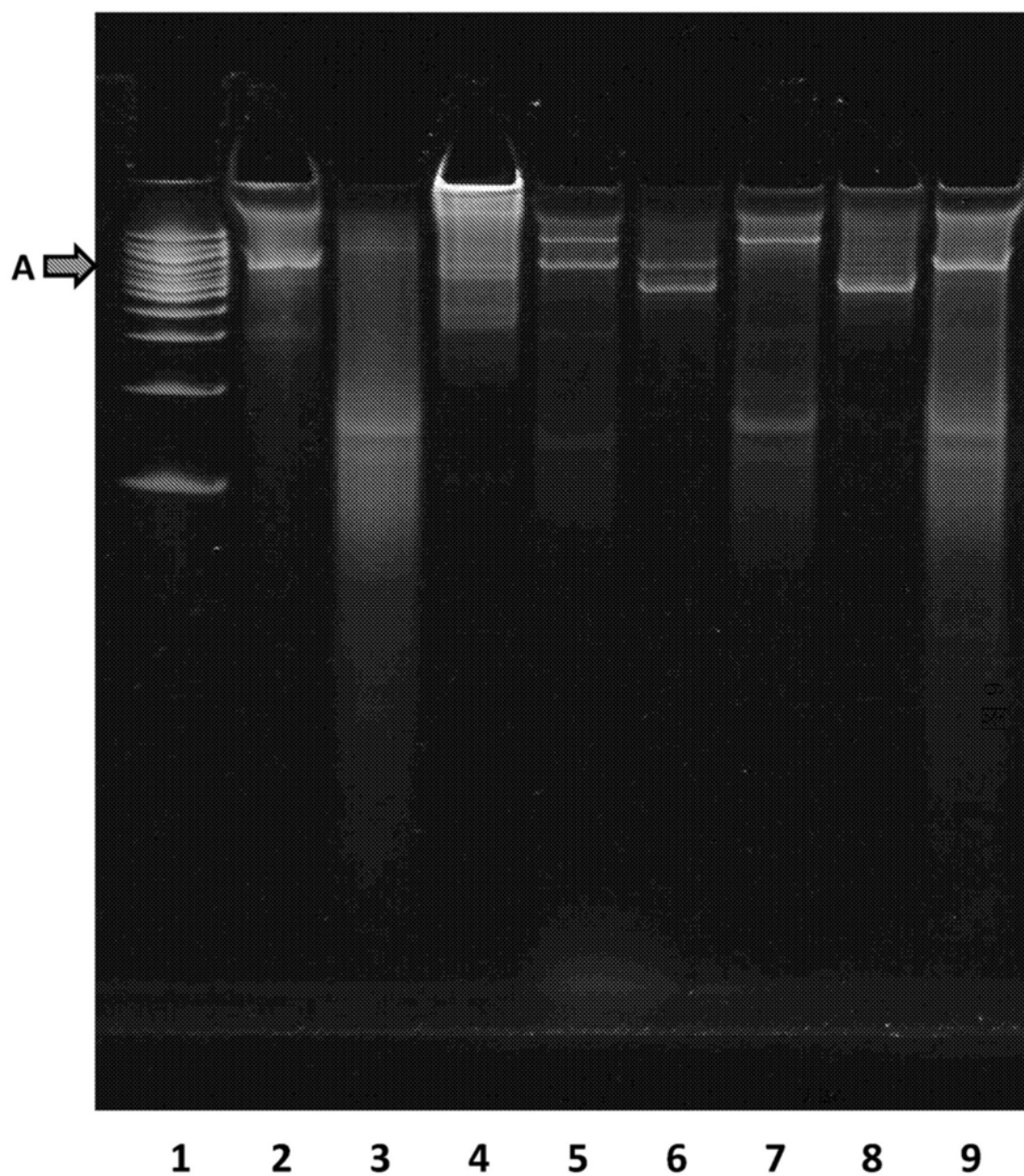


图14

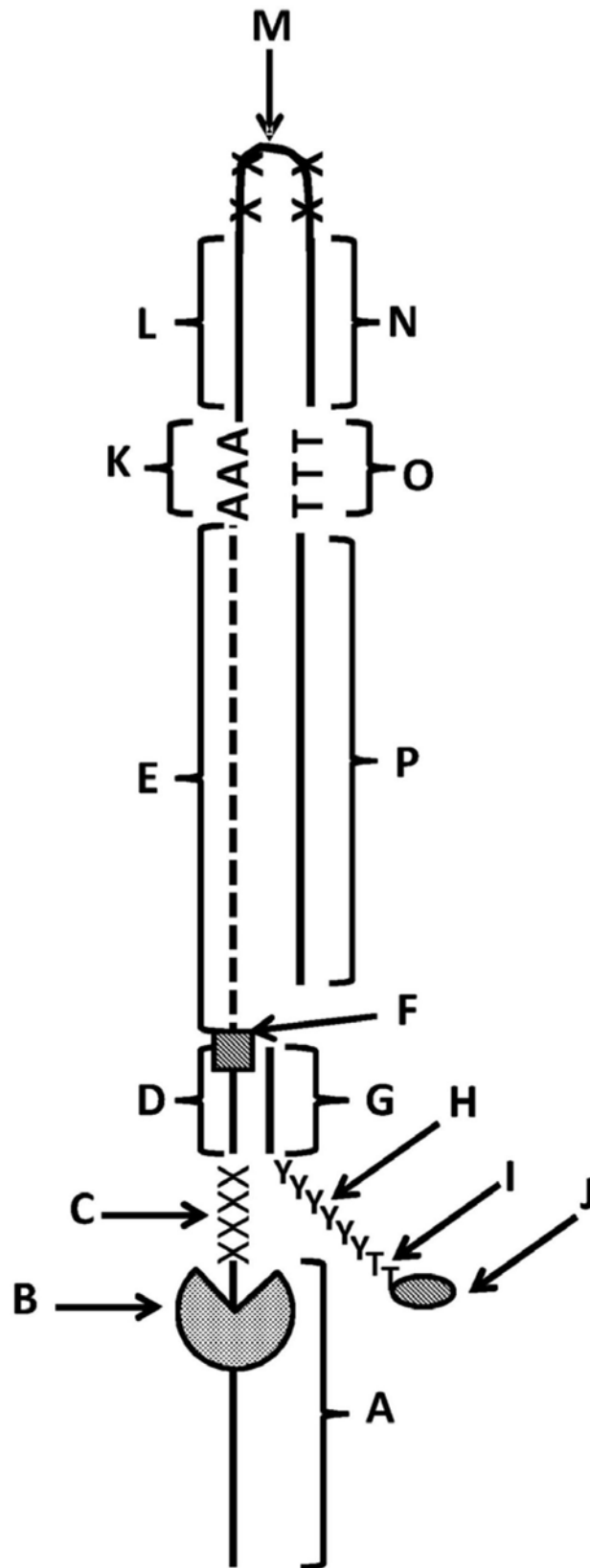


图15

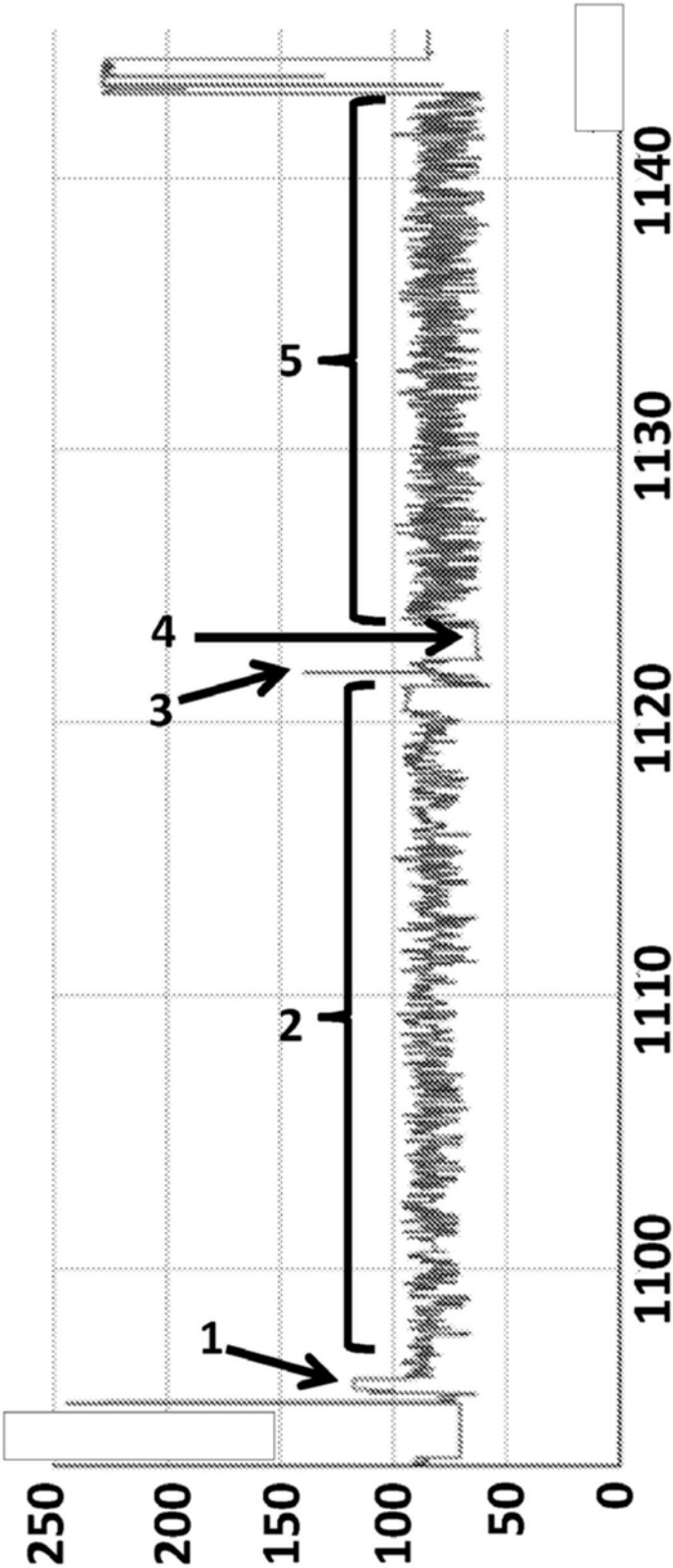


图16

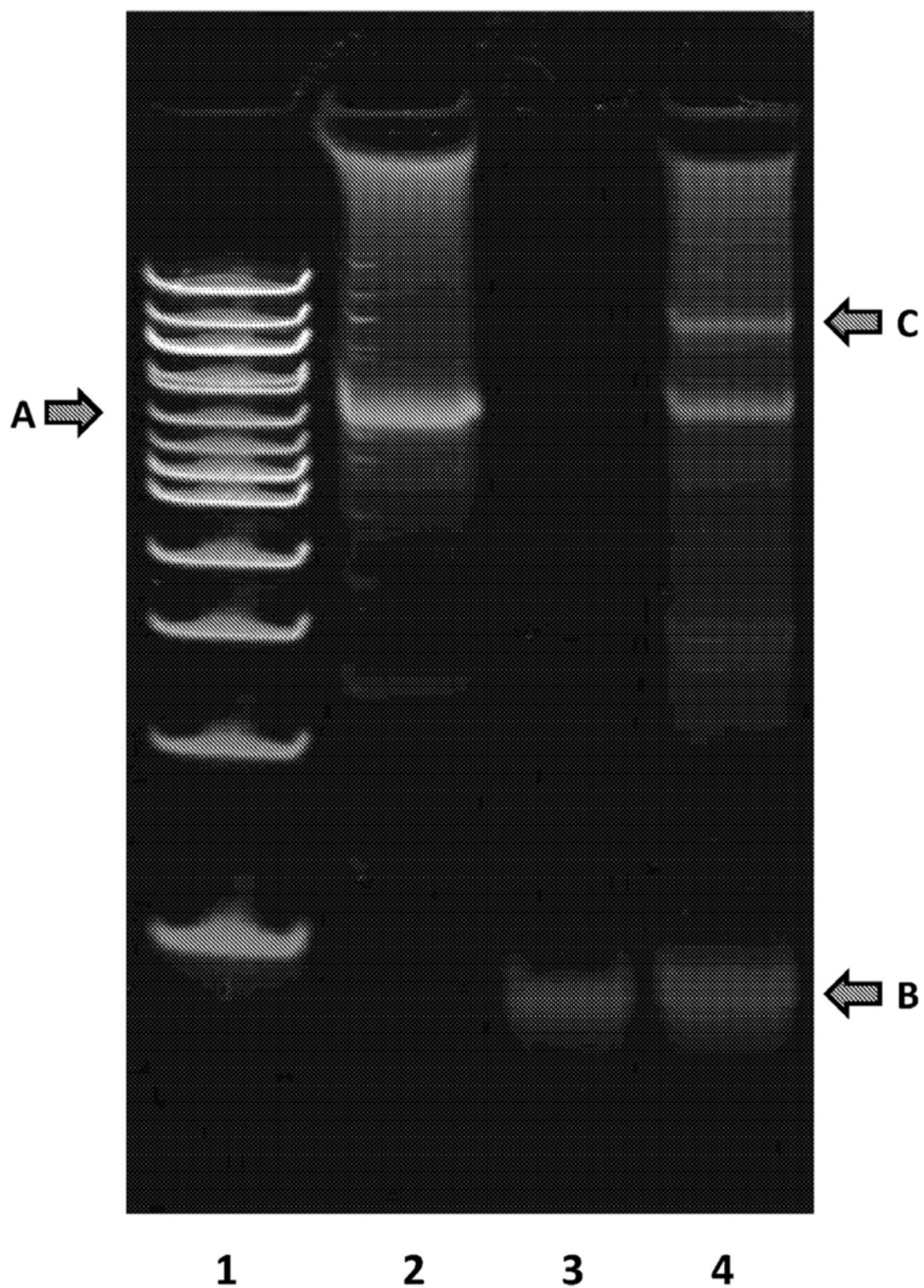


图17

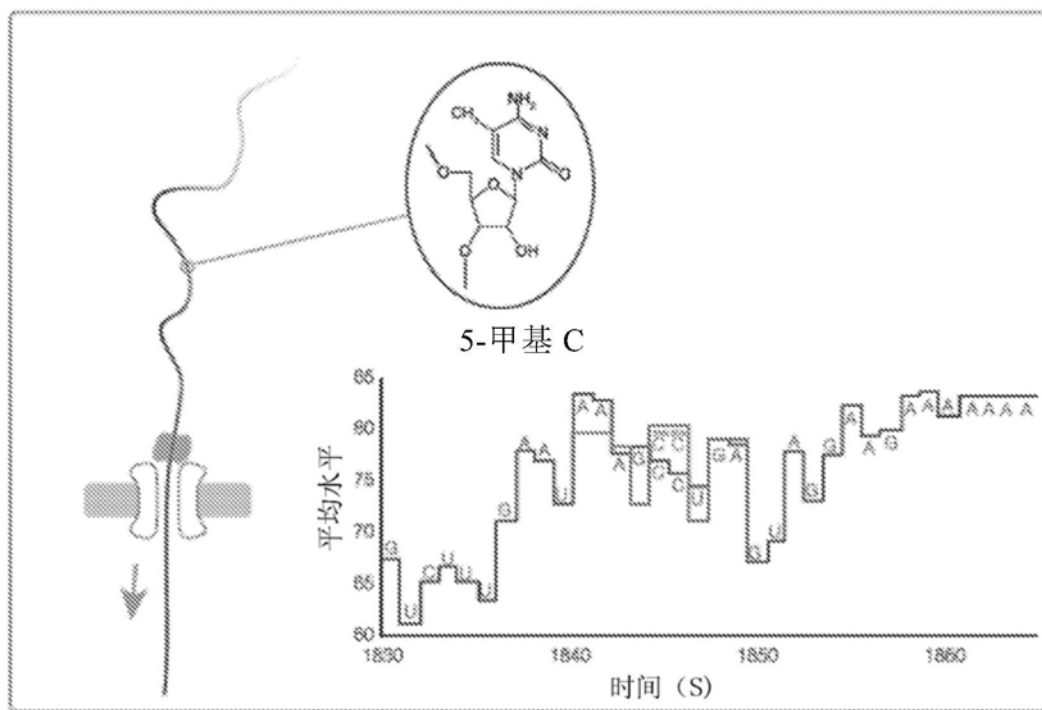


图18

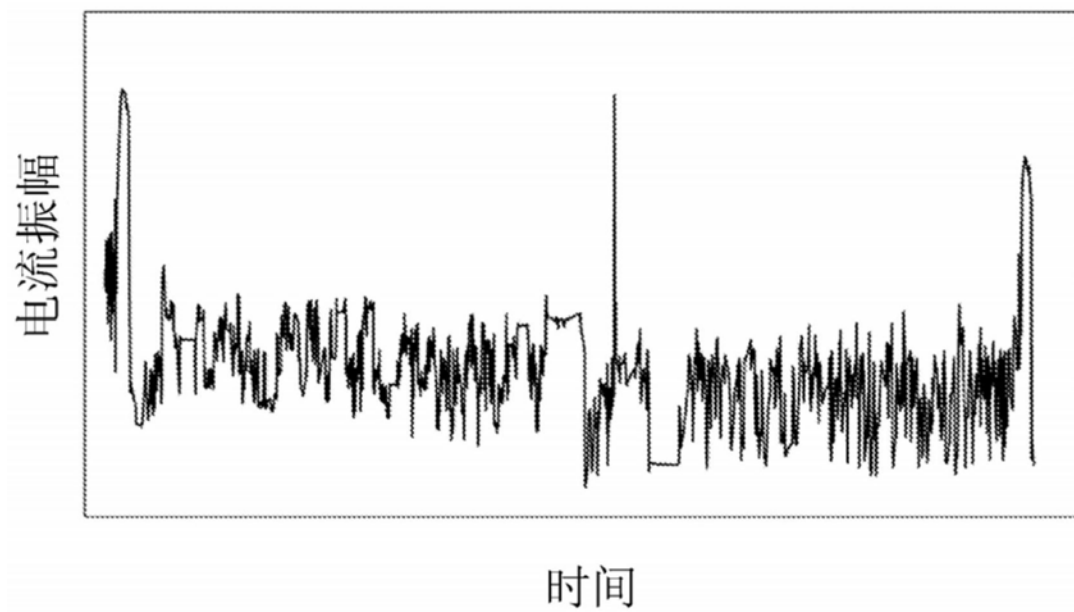


图19

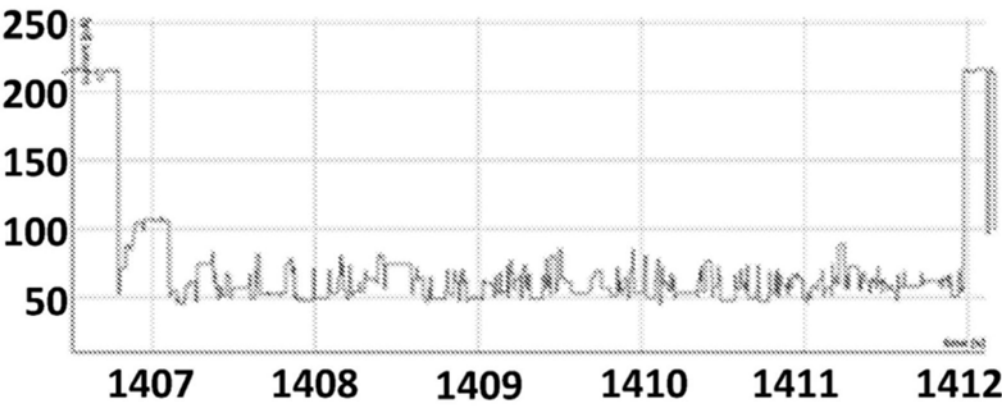


图20

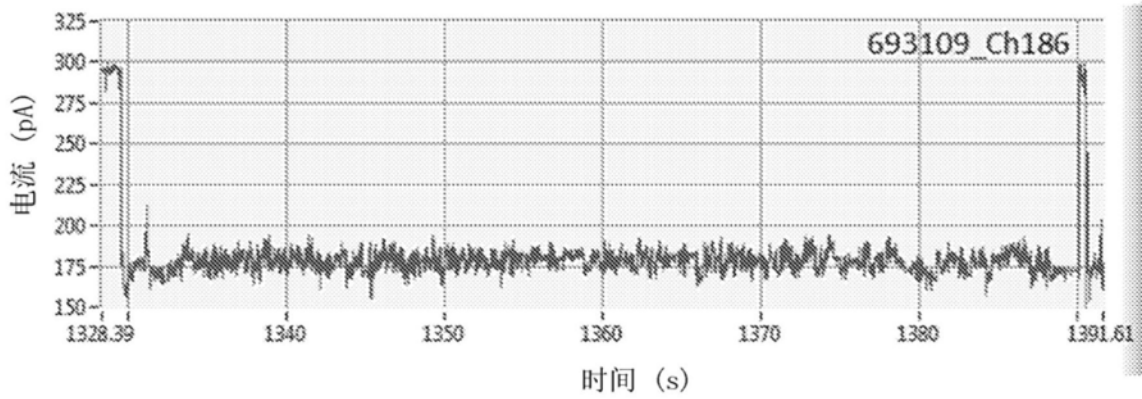


图21

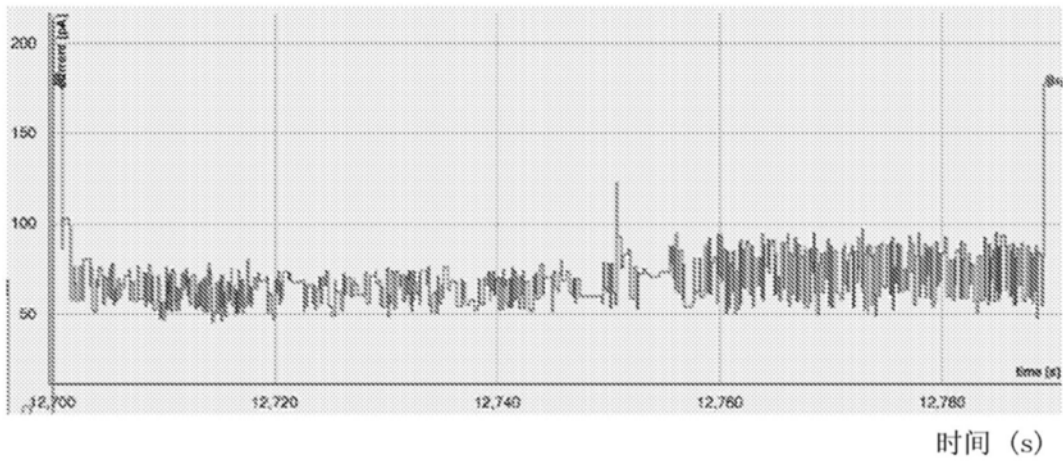


图22