



(12) 发明专利申请

(10) 申请公布号 CN 116514944 A

(43) 申请公布日 2023.08.01

(21) 申请号 202310422087.0

(22) 申请日 2017.04.06

(30) 优先权数据

1605899.2 2016.04.06 GB

1608274.5 2016.05.11 GB

(62) 分案原申请数据

201780022553.9 2017.04.06

(71) 申请人 牛津纳米孔科技公开有限公司

地址 英国牛津

(72) 发明人 拉科马·贾亚辛格 马克·布鲁斯

卢克·麦克尼尔

拉米兹·伊克巴尔·纳撒尼

普拉提克·拉吉·辛格

尼尔·罗杰·伍德

斯蒂芬·罗伯特·杨

(74) 专利代理机构 北京同立钧成知识产权代理有限公司 11205

专利代理师 严罗一 臧建明

(51) Int.Cl.

G07K 14/435 (2006.01)

C12Q 1/6869 (2018.01)

G01N 27/26 (2006.01)

权利要求书4页 说明书65页
序列表(电子公布) 附图7页

(54) 发明名称

突变体孔

(57) 摘要

本发明涉及胞溶素的突变形式。本发明还涉及使用胞溶素的所述突变形式进行分析物表征。

1. 一种突变胞溶素单体,其包括与SEQ ID NO:2中所示的氨基酸序列具有至少70%同一性的变异体,其中,所述变异体能够形成包含九个单体的孔,并且其中,所述变异体包括在位置T106的修饰,其增加了在位置T106的正电荷。

2. 根据权利要求1所述的突变单体,其中,所述变异体还包括在位置T104的修饰。

3. 根据权利要求1或2所述的突变单体,其中,所述变异体还包括在位置E94的修饰。

4. 根据前述权利要求中任一项所述的突变单体,其中,所述变异体还包括在一个或多个位置的修饰,所述位置选自K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、V95、Y96、S98、K99、V100、I101、P108、P109、T110、S111、K112以及T114。

5. 根据权利要求1到3中任一项所述的突变单体,其中,所述变异体还包括以下取代中的一个或多个:

D35N/S;	T93K;
K37N/W/S/Q;	E94D/Q/G/A/K/R/S/N, 优选 E94D/Q/K/N;
G43K	V95S;
K45D/R/N/Q/T/Y;	Y96D 或 Y96S;
V47K/S/N;	S98K 或 S98Q;
S49K/L;	K99Q/L 或 K99S;
T51K;	V100S;
S74K/R;	I101S
E76D/N;	E102N/Q/D/S;
S78R/K/N/Q;	T104R/K/Q;
S80K/R/N/Q;	T106R/K/Q;
S82K/R/N/Q;	P108K/R;
H83S/K;	P109K;
E84R/K/N/A;	T110K/R;
E85N;	S111K;
S86K/Q, 优选 S86K;	K112S;
V88I/T;	T114K;
S89K;	R115S;
M90K/I/A;	Q117S; 以及
T91K 或 T91S;	N119S。
E92D/S, 优选 E92D;	

6. 根据权利要求1到5中任一项所述的突变单体,其中,所述变异体还包括以下位置组合中的一个或多个处的修饰:

E94/K99/T106;	E94/Y96/T106;
E94/T93/T106;	K45/E94/T106;
E94/T91/T106;	E94/S98/K99/T106;
H83/E94/T106;	K37/E94/T106;

K45/E94/T106;	K45/V47/E92/E94/T106;
K37/E94/E102/T106;	E94/K99/T106;
K37/E94/T104/T106;	D35/E94/T106;
K45/E94/T106;	K37/E94/E102/T106;
K45/V47/E94/T106;	K45/S49/E92/E94/T106;
V47/E94/T106;	T51/E94D/T106;
T51/E94/T106;	E85/E94/T106;
K45/S49/E94/T106;	M90/E94/T106;
S49/E94/T106;	T91/E94/T106;
K45/T106;	E92/E94/T106;
G43/E94/T106;	T93/E94/T106;
V88/M90/E94/T106;	E94/T104; 以及
V47/V88/E94/T106;	E94/T106。
K45/S49/E94/E92/T106;	

7. 根据权利要求6所述的突变单体,其中,所述变异体还包括以下取代组合中的一个或多个:

E94D/K99Q/T106K;	G43K/E94D/T106K;
E94D/T93K/T106K;	V88I/M90A/E94D/T106K;
E94D/T91K/T106K;	V47N/V88T/E94D/T106K;
H83K/E94D/T106K;	K45N/S49K/E94N/E92D/T106K;
E94Q/Y96D/T106K;	K45N/V47K/E92D/E94N/T106K;
K45D/E94K/T106K;	E94D/K99Q/T106K;
K45R/E94D/T106K;	E94D/T106R;
E94D/S98K/K99L/T106K;	E94D/T106K;
K37N/E94D/T106K;	E94D/T104R;
K37W/E94D/T106K;	E94D/T104K;
K37S/E94D/T106K;	E94Q/T104Q;
K45N/E94N/T106K;	E94Q/T106Q;
K37Q/E94D/E102N/T106K;	E94D/M90K/T106K;
K37S/E94D/E102S/T106K;	E85N/E94D/T106K;
K37S/E94D/T104K/T106K;	D35N/E94D/T106K;
K45Q/E94Q/T106K;	D35S/E94D/T106K;
K45T/V47K/E94D/T106K;	E92D/E94Q/T106K;
V47S/E94D/T106K;	E94Q/T106K;
T51K/E94D/T106K;	M90I/E94D/T106K;
K45Y/S49K/E94D/T106K;	S82K/E94D/T106K;
S49L/E94D/T106K;	M90K/E94D/T106K;
K45R/T106K;	T91K/E94D/T106K;
V47K/E94D/T106K;	T93K/E94D/T106K;
E94Q/T106Q; 以及	E94Q/T106K。

8. 根据前述权利要求中任一项所述的突变胞溶素单体, 其中, 所述单体包括任何数目个在前述权利要求中任一项所定义的修饰和/或取代以及其任何组合。

9. 根据前述权利要求中任一项所述的突变胞溶素单体, 其中, 在所述变异体中, (a) 与SEQ ID NO:2的位置34到70处的氨基酸相对应的氨基酸中的2个、4个、6个、8个、10个、12个、14个、16个、18个或20个已经缺失, 并且(b) 与SEQ ID NO:2的位置71到107处的氨基酸相对应的氨基酸中的2个、4个、6个、8个、10个、12个、14个、16个、18个或20个已经缺失。

10. 根据权利要求9所述的突变单体, 其中, 与SEQ ID NO:2中的以下氨基酸相对应的氨基酸已经缺失:

- (i) N46/V47/T91/T92; 或
- (ii) N48/S49/T91/T92。

11. 根据前述权利要求中任一项所述的突变单体, 其中, 所述变异体进一步包括

SEQ ID NO:2的以下位置中的一个或多个处的取代: (a) E84、E85、E92、E97和D126; (b) E85、E97和D126; 或(c) E84和E92; 或

E84Q/E85K/E92Q/E97S/D126G中的一个或多个处或者在适当的情况下全部E84Q/E85K/

E92Q/E97S/D126G处的取代。

12. 根据权利要求11所述的突变单体,其中,所述变异体包括:E84Q/E85K/E92Q/E94D/E97S/T106K/D126G。

13. 根据前述权利要求中任一项所述的突变单体,其中,所述突变体被化学修饰。

14. 一种构建体,其包括两个或更多个衍生自胞溶素的共价连接单体,其中,所述单体中的至少一个是根据前述权利要求中任一项所述的突变胞溶素单体。

15. 一种多核苷酸,其对根据权利要求1到13中任一项所述的突变胞溶素单体或根据权利要求14所述的构建体进行编码。

16. 一种孔,其包括至少一个根据权利要求1到13中任一项所述的突变胞溶素单体和/或至少一个根据权利要求14所述的构建体。

17. 一种表征靶分析物的方法,其包括:

(a) 使所述靶分析物与根据权利要求16所述的孔接触,使得所述靶分析物移动穿过所述孔;以及

(b) 当所述分析物相对于所述孔移动时取得一个或多个测量结果,其中所述测量结果指示所述靶分析物的一种或多种特性,并且由此表征所述靶分析物。

突变体孔

[0001] 本申请为2017年4月6日递交的申请号为201780022553.9,发明名称为《突变体孔》的分案申请。

技术领域

[0002] 本发明涉及胞溶素(lysenin)的突变形式。本发明还涉及使用胞溶素的所述突变形式进行分析物表征。

背景技术

[0003] 纳米孔感测是一种依赖于对分析物分子与受体之间的单独结合或相互作用事件的观察的感测方法。可以通过在绝缘膜中放置纳米尺寸的单孔和测量在存在分析物分子的情况下通过所述孔的电压驱动离子转运来产生纳米孔传感器。分析物的同一性通过其独特的电流特征揭露,尤其是电流块的持续时间和程度以及电流水平的变化。这种纳米孔传感器是可商购的,例如,由牛津纳米孔科技有限公司(Oxford Nanopore Technologies Ltd)销售的MinION™装置,其包括集成在电子芯片内的纳米孔阵列。

[0004] 当前在广泛的应用范围内需要快速且便宜的核酸(例如,DNA或RNA)测序技术。现有技术缓慢且昂贵,这主要是因为其依赖于扩增技术以产生大量核酸并且需要大量专业荧光化学品以进行信号检测。纳米孔感测有可能通过降低所需核苷酸和试剂的量来提供快速且便宜的核酸测序。

[0005] 使用纳米孔感测对核酸进行测序的基本要素之一是控制核酸移动穿过孔。另一个要素是在核酸聚合物移动穿过孔时区分核苷酸。在过去,为了实现核苷酸区分,已经使核酸穿过溶血素的突变体。这已经提供了已经显示具有序列依赖性的电流特征。还已经显示,在使用溶血素孔时,大量核苷酸贡献于所观察到的电流,从而使得所观察到的电流与多核苷酸之间的直接关系具有挑战性。

[0006] 虽然已经通过溶血素孔突变提高了用于核苷酸区分的电流范围,但是如果可以进一步提高核苷酸之间的电流差异,则测序系统将具有更高的性能。另外,已经观察到,当核酸移动穿过孔时,一些电流状态显示出很高的变化。还已经显示,一些突变溶血素孔展现出比其它突变溶血素更高的变化。虽然这些状态的变化可能含有序列特异性信息,但是期望产生具有低变化的孔以简化系统。还期望降低贡献于所观察到的电流的核苷酸的数目。

[0007] 胞溶素(也称为efL1)是一种从蚯蚓赤子爱胜蚓的体腔液中纯化而来的成孔毒素。其特异性地结合到鞘磷脂,所述鞘磷脂抑制胞溶素诱导的溶血(Yamaji等人,《生物化学杂志(J.Biol.Chem.)》,1998年,第273卷,第9期,第5300到5306页)。在De Colbis等人,《结构(Structure)》,2012年,第20卷,第1498到1507页中公开了胞溶素单体的晶体结构。

发明内容

[0008] 诸位发明人惊奇地发现了新的突变胞溶素单体,在所述单体中已经进行了一个或多个修饰以提高所述单体与多核苷酸相互作用的能力。诸位发明人还惊奇地证实,包括新

颖突变单体的孔具有增强的与多核苷酸相互作用的能力并且因此显示出提高的用于估计多核苷酸的特性如其序列的性质。突变孔出人意料地显示出提高的核苷酸区分。具体地说，突变孔出人意料地显示出增大的电流范围和减少的状态变化，所述增大的电流范围使得更容易区分不同核苷酸，并且所述减少的状态变化增加信噪比。另外，在多核苷酸移动穿过所述孔时贡献于电流的核苷酸的数目减小。这使得更容易识别在多核苷酸移动穿过孔时的所观察到的电流与多核苷酸序列之间的直接关系。

[0009] 除非相反地陈述，否则本文所公开的所有氨基酸取代、缺失和/或添加均参考包括SEQ ID NO:2中所示的序列的变异体的突变胞溶素单体。

[0010] 提及包括SEQ ID NO:2中所示的序列的变异体的突变胞溶素单体涵盖包括如SEQ ID NO:14到16中所述的序列的变异体的突变胞溶素单体。可以对包括SEQ ID NO:2中所示的序列的变异体的胞溶素单体进行等同于本文中参考SEQ ID NO:2公开的取代、缺失和/或添加的氨基酸取代、缺失和/或添加。

[0011] 突变单体可以被视为分离单体。

[0012] 因此，本发明提供了一种突变胞溶素单体，其包括SEQ ID NO:2中所示的序列的变异体，其中所述单体能够形成孔，并且其中所述变异体包括以下位置中的一个或多个处的修饰：K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、V95、Y96、S98、K99、V100、I101、P108、P109、T110、S111、K112以及T114。

[0013] 本发明还提供了一种突变胞溶素单体，其包括SEQ ID NO:2中所示的序列的变异体，其中所述单体能够形成孔，并且其中所述变异体包括以下取代中的一个或多个：

[0014] D35N/S;

[0015] S74K/R;

[0016] E76D/N;

[0017] S78R/K/N/Q;

[0018] S80K/R/N/Q;

[0019] S82K/R/N/Q;

[0020] E84R/K/N/A;

[0021] E85N;

[0022] S86K/Q;

[0023] S89K;

[0024] M90K/I/A;

[0025] E92D/S;

[0026] E94D/Q/G/A/K/R/S/N;

[0027] E102N/Q/D/S;

[0028] T104R/K/Q;

[0029] T106R/K/Q;

[0030] R115S;

[0031] Q117S;以及

[0032] N119S。

[0033] 本发明还提供了一种突变胞溶素单体，其包括SEQ ID NO:2中所示的序列的变异

体,其中所述单体能够形成孔,并且其中所述变异体包括以下中的一个或多个处的突变:

- [0034] D35/E94/T106;
- [0035] K37/E94/E102/T106;
- [0036] K37/E94/T104/T106;
- [0037] K37/E94/T106;
- [0038] K37/E94/E102/T106;
- [0039] G43/E94/T106;
- [0040] K45/V47/E92/E94/T106;
- [0041] K45/V47/E94/T106;
- [0042] K45/S49/E92/E94/T106;
- [0043] K45/S49/E94/T106;
- [0044] K45/E94/T106;
- [0045] K45/T106;
- [0046] V47/E94/T106;
- [0047] V47/V88/E94/T106;
- [0048] S49/E94/T106;
- [0049] T51/E94D/T106;
- [0050] S74/E94;
- [0051] E76/E94;
- [0052] S78/E94;
- [0053] Y79/E94;
- [0054] S80/E94;
- [0055] S82/E94;
- [0056] S82/E94/T106;
- [0057] H83/E94;
- [0058] H83/E94/T106;
- [0059] E85/E94/T106;
- [0060] S86/E94;
- [0061] V88/M90/E94/T106;
- [0062] S89/E94;
- [0063] M90/E94/T106;
- [0064] T91/E94/T106;
- [0065] E92/E94/T106;
- [0066] T93/E94/T106;
- [0067] E94/Y96/T106;
- [0068] E94/S98/K99/T106;
- [0069] E94/K99/T106;
- [0070] E94/E102;
- [0071] E94/T104;

[0072] E94/T106;

[0073] E94/P108;

[0074] E94/P109;

[0075] E94/T110;

[0076] E94/S111;

[0077] E94/T114;

[0078] E94/R115;

[0079] E94/Q117;以及

[0080] E94/E119。

[0081] 本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中所述单体能够形成孔,并且其中所述变异体包括以下取代中的一个或多个:

[0082] E84R/E94D;

[0083] E84K/E94D;

[0084] E84N/E94D;

[0085] E84A/E94Q;

[0086] E84K/E94Q以及

[0087] E94Q/D121S。

[0088] 本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中所述变异体包括以下取代组合之一:

[0089] -E84Q/E85K/E92Q/E94D/E97S/D126G;

[0090] -E84Q/E85K/E92Q/E94Q/E97S/D126G;或

[0091] -E84Q/E85K/E92Q/E94D/E97S/T106K/D126G。

[0092] 本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中在所述变异体中,(a) SEQ ID NO:2的位置34到70处或与那些位置相对应的氨基酸中的2个、4个、6个、8个、10个、12个、14个、16个、18个或20个已经缺失,并且(b) SEQ ID NO:2的位置71到107处或与那些位置相对应的氨基酸中的2个、4个、6个、8个、10个、12个、14个、16个、18个或20个已经缺失。

[0093] 本发明还提供了:

[0094] -一种构建体,其包括衍生自胞溶素的两个或更多个共价连接单体,其中所述单体中的至少一个是本发明的突变胞溶素单体;

[0095] -一种多核苷酸,其对本发明的突变胞溶素单体或本发明的基因融合构建体进行编码;

[0096] -一种衍生自胞溶素的同源寡聚孔,其包括足够数目的本发明的突变胞溶素单体;

[0097] -一种衍生自胞溶素的异源寡聚孔,其包括至少一个本发明的突变胞溶素单体;

[0098] -一种孔,其包括至少一个本发明的构建体;

[0099] -一种表征靶分析物的方法,其包括:(a)使所述靶分析物与本发明的孔接触,使得所述靶分析物移动穿过所述孔;(b)当所述分析物相对于所述孔移动时取得一个或多个测量结果,其中所述测量结果指示所述靶分析物的一个或多个特性,并且由此表征所述靶分析物;

- [0100] -一种形成用于表征靶多核苷酸的传感器的方法,其包括在本发明的孔与多核苷酸结合蛋白之间形成复合体以及由此形成用于表征所述靶多核苷酸的传感器;
- [0101] -一种用于表征靶多核苷酸的传感器,其包括本发明的孔与多核苷酸结合蛋白之间的复合体;
- [0102] -本发明的孔用于表征靶分析物的用途;
- [0103] -一种用于表征靶多核苷酸的试剂盒,其包括(a)本发明的孔以及(b)膜;
- [0104] -一种用于表征样品中的靶多核苷酸的设备,其包括(a)多个本发明的孔和(b)多个多核苷酸结合蛋白;
- [0105] -一种提高包括SEQ ID NO:2中所示的序列的胞溶素单体表征多核苷酸的能力的方法,其包括:进行本发明的一个或多个修饰和/或取代;
- [0106] -一种产生本发明的构建体的方法,其包括:将至少一个本发明的突变胞溶素单体共价连接到衍生自胞溶素的一个或多个单体;以及
- [0107] -一种形成本发明的孔的方法,其包括:允许至少一个本发明的突变单体或至少一个本发明的构建体与足够数目的本发明的单体、本发明的构建体或衍生自胞溶素的单体寡聚以形成孔。

附图说明

- [0108] 图1示出了胞溶素突变体1的中值图。
- [0109] 图2示出了胞溶素突变体10的中值图。
- [0110] 图3示出了胞溶素突变体-胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85K/E92Q/E94D/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)的中值图。
- [0111] 图4示出了胞溶素突变体-胞溶素-具有通过E94C附接的2-碘-N-(2,2,2-三氟乙基)乙酰胺的(E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)的中值图。
- [0112] 图5示出了在实例中使用的衔接子。A对应于30个iSpC3。B对应于SEQ IN NO:19。C对应于4个iSp18。D对应于SEQ ID NO:20。E对应于SEQ ID NO:21,其具有附接到其5'末端的5BNA-G//iBNA-G//iBNA-T//iBNA-T//i-BNA-A。F对应于SEQ ID NO:22,其具有5'磷酸酯。G对应于SEQ ID NO:24。H对应于胆固醇。
- [0113] 图6示出了胞溶素单体的3D结构。在与含有鞘磷脂的膜相互作用时,胞溶素单体通过中间前孔组装在一起以形成九聚体孔。在组装过程期间,以黑色示出的多肽区段(对应于SEQ ID NO:2的氨基酸65到74)转化成图7所示的β桶的底部环。以黑色示出的多肽区段的任一侧的两个β折叠以及将那些β折叠链接到以黑色示出的多肽区段的多肽区段(对应于SEQ ID NO:2的氨基酸34到64和75到107)延伸以形成孔的β桶,如图7所示。这种大的结构变化使得难以通过研究单体结构来预测胞溶素孔的β桶区域。
- [0114] 图7描绘了胞溶素孔的区域。图7A示出了胞溶素的九聚体孔的3D结构,并且图7B示出了取自胞溶素孔的单体的结构。每个单体为胞溶素孔的桶贡献两个β折叠。β折叠(含有与SEQ ID NO:2的氨基酸34到64和75到107相对应的氨基酸)通过孔底部的非结构化环(与SEQ ID NO:2的位置65到74相对应的氨基酸)链接。

[0115] 图8是胞溶素的氨基酸序列(SEQ ID NO:2)与三个胞溶素相关蛋白的氨基酸序列(SEQ ID NO:14到16)的比对。通过使用非冗余蛋白序列的数据库执行BLAST搜索来识别了具有与胞溶素密切相关的序列的三个胞溶素同系物。将胞溶素相关蛋白1(LRP1)、胞溶素相关蛋白2(LRP2)和胞溶素相关蛋白3(LRP3)的蛋白序列与胞溶素的序列比对以显示所述四种蛋白的相似性。深灰色阴影表示一致的氨基酸存在于所有四个序列中的位置。LRP1与胞溶素的一致程度为大约75%，LRP2与胞溶素的一致程度为大约88%，并且LRP3与胞溶素的一致程度为大约79%。

[0116] 序列表说明

[0117] SEQ ID NO:1示出了编码胞溶素单体的多核苷酸序列。

[0118] SEQ ID NO:2示出了胞溶素单体的氨基酸序列。

[0119] SEQ ID NO:3示出了编码Phi29 DNA聚合酶的多核苷酸序列。

[0120] SEQ ID NO:4示出了Phi29 DNA聚合酶的氨基酸序列。

[0121] SEQ ID NO:5示出了衍生自来自大肠杆菌的sbcB基因的密码子优化多核苷酸序列。其编码来自大肠杆菌的核酸外切酶I酶(EcoExo I)。

[0122] SEQ ID NO:6示出了来自大肠杆菌的核酸外切酶I酶(EcoExo I)的氨基酸序列。

[0123] SEQ ID NO:7示出了衍生自来自大肠杆菌的xthA基因的密码子优化多核苷酸序列。其编码来自大肠杆菌的核酸外切酶III酶。

[0124] SEQ ID NO:8示出了来自大肠杆菌的核酸外切酶III酶的氨基酸序列。这种酶在3'到5'方向上对来自双链DNA(dsDNA)中的一条链的5'单磷酸酯核苷执行分配消化。链上的酶启动需要大约4个核苷酸的5'突出。

[0125] SEQ ID NO:9示出了衍生自来自嗜热菌的recJ基因的密码子优化多核苷酸序列。其编码来自嗜热菌的RecJ酶(TthRecJ-cd)。

[0126] SEQ ID NO:10示出了来自嗜热菌的RecJ酶(TthRecJ-cd)的氨基酸序列。这种酶在5'到3'方向上对来自ssDNA的5'单磷酸酯核苷执行进行性消化。链上的酶启动需要至少4个核苷酸。

[0127] SEQ ID NO:11示出了衍生自噬菌体λ_{exo}(redX)基因的密码子优化多核苷酸序列。其编码噬菌体λ核酸外切酶。

[0128] SEQ ID NO:12示出了噬菌体λ核酸外切酶的氨基酸序列。所述序列是组装成三聚体的三个一致的亚基之一。所述酶在5'到3'方向上对来自dsDNA的一个链的核苷酸执行高度进行性消化(<http://www.neb.com/nebecomm/products/productM0262.asp>)。链上的酶启动优先需要具有5'磷酸酯的大约4个核苷酸的5'突出。

[0129] SEQ ID NO:13示出了He1308 Mbu的氨基酸序列。

[0130] SEQ ID NO:14示出了胞溶素相关蛋白(LRP)1的氨基酸序列。

[0131] SEQ ID NO:15示出了胞溶素相关蛋白(LRP)2的氨基酸序列。

[0132] SEQ ID NO:16示出了胞溶素相关蛋白(LRP)3的氨基酸序列。

[0133] SEQ ID NO:17示出了伴孢晶体蛋白-2(parasporin-2)的活化版本的氨基酸序列。全长蛋白在其氨基和羧基末端处被切割以形成能够形成孔的活化版本。

[0134] SEQ ID NO:18示出了Dda 1993的氨基酸序列。

[0135] SEQ ID NO:19到24示出了在实例中使用的多核苷酸序列。

具体实施方式

[0136] 应理解,所公开产品和方法的不同应用可以根据所属领域的特定需要而定制。还应理解,本文中所使用的术语仅出于对本发明的特定实施例进行描述的目的而并不旨在是限制性的。

[0137] 另外,除非内容另外明确指明,否则如本说明书和所附权利要求书中所使用的,单数形式“一个/一种(a/an)”和“所述(the)”包含复数指示物。因此,例如,提及“一个突变单体”包含“多个突变单体”,提及“一个取代”包含两个或更多个这种取代,提及“一个孔”包含两个或更多个这种孔,提及“一个多核苷酸”包含两个或更多个这种多核苷酸,等等。

[0138] 在本说明书中,在特定位置处通过符号“/”将不同氨基酸分开的情况下,/符号“/”意指“或”。例如,P108R/K意指P108R或P108K。在本说明书中,在不同位置或不同取代通过符号“/”分开的情况下,/符号意指“和”。例如,E94/P108意指E94和P108,或者E94D/P108K意指E94D和P108K。

[0139] 本文中(无论是上文还是下文)所引用的所有公开、专利和专利申请特此通过引用以其全部内容并入。

[0140] 突变胞溶素单体

[0141] 在一方面,本发明提供了突变胞溶素单体。突变胞溶素单体可以用于形成本发明的孔。突变胞溶素单体是其序列与野生型胞溶素单体的序列不同的单体(例如SEQ ID NO: 2、SEQ ID NO:14、SEQ ID NO:15或SEQ ID NO:16)。突变胞溶素单体通常保留在存在本发明的其它单体或来自胞溶素或衍生自胞溶素的其它单体的情况下形成孔的能力。因此,突变单体通常能够形成孔。用于确认突变单体形成孔的能力的方法在所属领域中是众所周知的且在实例中进行描述。例如,可以通过电生理学确定孔的形成。孔通常插入膜中,所述膜可以是例如脂质膜或嵌段共聚物膜。可以从插入膜中的单胞溶素孔、如包括一个或多个本发明的单体的孔获得电气测量或光学测量。可以跨膜施加电势差,并且可以检测通过膜的电流。可以通过任何适当的方法、如通过电气手段或光学手段检测电流。孔转移多核苷酸、优选地单链多核苷酸的能力可以通过如下方式确定:添加多核苷酸结合蛋白、DNA、燃料(例如,MgCl₂、ATP)预混合物;施加电势差(例如,180mV);以及监测电流通过孔以检测多核苷酸结合蛋白控制的DNA移动。

[0142] 当存在于孔中时,突变单体具有改变的与多核苷酸相互作用的能力。因此,包括所述突变单体中的一个或多个的孔具有提高的核苷酸读取性质,例如,显示(1)提高的多核苷酸捕获和(2)提高的多核苷酸识别或区分。具体地说,由突变单体构造的孔比野生型更容易地捕获核苷酸和多核苷酸。另外,由突变单体构造的孔显示增大的电流范围和减少的状态变化,所述增大的电流范围使得更容易区分不同的核苷酸,并且所述减少的状态变化增加信噪比。另外,在多核苷酸移动穿过由突变体构造的孔时贡献于电流的核苷酸的数目减小。这使得更容易识别在多核苷酸移动穿过孔时的所观察到的电流与多核苷酸序列之间的直接关系。突变体的提高的核苷酸读取性质通过五种主要机制实现,即通过以下各项的变化:

- [0143] • 位阻(增加或减小氨基酸残基的大小);
- [0144] • 电荷(例如,引入或移除-ve电荷和/或引入或移除+ve电荷);
- [0145] • 氢结合(例如,引入可以氢结合到碱基对的氨基酸);
- [0146] • π 堆积(例如,引入通过离域电子 π 系统相互作用的氨基酸);和/或

[0147] • 孔结构改变(例如,引入增加桶或通道的大小的氨基酸)。

[0148] 这五种机制中的任何一种或多种可能是由本发明的突变单体形成的孔的性质提高的原因。例如,由于改变的位阻、改变的氢结合和改变的结构,包括本发明的突变单体的孔可以显示出提高的核苷酸读取性质。

[0149] 本发明的突变单体包括SEQ ID NO:2中所示的序列的变异体。SEQ ID NO:2是胞溶素单体的野生型序列。SEQ ID NO:2的变异体是氨基酸序列不同于SEQ ID NO:2的氨基酸序列的多肽。通常,变异体保留其形成孔的能力。

[0150] 包括包含在S80、T106、T104的取代的突变单体中的一个或多个的孔显示出提高的多核苷酸捕获。这种取代的具体实例包含S80K/R、T104R/K和T106R/K。在这些位置处增加了这些位置中的任何一个或多个、如2个、3个、4个或5个处的氨基酸侧链的正电荷的其它取代可以用于提高包括突变单体的孔的性质,即,与野生型孔或包括其它捕获增强突变的突变单体、如E84Q/E85K/E92Q/E97S/D126G的孔相比,提高对多核苷酸的捕获,所述孔例如包括仅包括那些突变的突变单体或包括以下突变E84Q/E85K/E92Q/E94D/E97S/D126G的突变单体的孔。通常,在相对于包括其它突变、如E84Q/E85K/E92Q/E97S/D126G或E84Q/E85K/E92Q/E94D/E97S/D126G的孔确定了提高的情况下,那些突变也存在于被测试的突变单体中,即,突变的一种或多种作用或突变的组合是相对于与被测试而不是处于一个或多个测试位置处的单体/孔一致的基线单体/孔确定的。包括突变单体或对照单体的孔的性质可以使用异源寡聚孔或更优选地同源寡聚孔来确定。在整个说明书中、例如在表9中描述了优选的突变组合的实例。

[0151] 包括包含D35、K37、K45、V47、S49、E76、S78、S82、V88、S89、M90、T91、E92、E94、Y96、S98、V100、T104处的取代的突变单体中的一个或多个的孔显示出提高的多核苷酸识别或区分。这种取代的具体实例包含D35N、K37N/S、K45R/K/D/T/Y/N、V47K/R、S49K/R/L、T51KE76S/N、S78N、S82N、V88I、S89Q、M90I/A、T91S、E92D/E、E94D/Q/N、Y96D、S98Q、V100S和T104K。如表9中所描述的,这些突变各自可以降低噪声、增大电流范围和/或减少通道门控。对SEQ ID NO:2中的指定位置或SEQ ID NO:2的变异体中的对应位置进行了以与对所进行的这些示例性突变中的任何一个或多个相同的方式增大或减小氨基酸侧链的大小、增加或减少电荷、导致相同的氢键形成和/或影响 π 堆积的其它突变。突变可以单独或组合引入。例如,可使这些位置中的2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个或18个突变以提高包括突变单体的孔的性质,即提高信噪比,增大范围和/或减少通道门控,使得相比于野生型孔、包括包含突变E84Q/E85K/E92QE97S/D126G的突变单体如仅包括那些突变的单体、包含突变E84Q/E85K/E92Q/E94D/E97S/D126G、包含突变E84Q/E85K/E92Q/E94Q/E97S/D126G和/或包含突变E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的突变单体的孔,多核苷酸识别和区分得到改进。通常,在相对于包括其它突变、如E84Q/E85K/E92Q/E97S/D126G、E84Q/E85K/E92Q/E94D/E97S/D126G、E84Q/E85K/E92Q/E94Q/E97S/D126G或E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的孔确定了提高的情况下,那些突变也存在于被测试的突变单体中,即,突变的一种或多种作用或突变的组合是相对于与被测试而不是处于一个或多个测试位置处的单体/孔一致的基线单体/孔确定的。包括突变单体或对照单体的孔的性质可以使用异源寡聚孔或更优选地同源寡聚孔来确定。在整个说明书中、例如在表9中描述了优选的突变组合的实例。

[0152] 相比于野生型孔或包括包含突变E84Q/E85K/E92QE97S/D126G的突变单体的孔,包括包含E94和/或Y96处的取代的突变单体中的一个或多个的孔可以在多核苷酸移动穿过孔时减少贡献于电流的核苷酸的数目。例如,可以进行取代Y96D/E,优选地与E94Q/D组合以减小读头(read head)的大小。相比于野生型孔或包括包含突变E84Q/E85K/E92QE97S/D126G的突变单体的孔,在多核苷酸移动穿过孔时贡献于电流的核苷酸的数目的减少还可通过以下方式实现:从单体的形成孔的桶的一部分的两个β链中的每一个缺失偶数数目的氨基酸(通常,将存在于孔的内腔中的氨基酸和将背对孔的内腔的相邻氨基酸),即与SEQ ID NO:2的氨基酸34到65和74到107相对应的位置,如本文所述。

[0153] 本发明的修饰

[0154] 本发明提供了一种突变胞溶素单体,其中相比于野生型胞溶素并且相比于在所属领域中、例如在WO 2013//153359中公开的胞溶素突变体而修饰了对胞溶素孔中的桶的结构有贡献的β折叠的氨基酸序列。本发明的修饰在与SEQ ID NO:2的氨基34到107、具体地说SEQ ID NO:2的氨基酸34到65和74到107相对应的胞溶素单体的区域中进行。LR1、LR2和LR3单体的对应区域显示于图8的比对中。

[0155] 因此,本发明提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中所述单体能够形成孔,并且其中所述变异体包括以下位置中的一个或多个如2个到22个、3个到20个、4个到15个、5个到10个、6个、7个、8个或9个处的修饰:K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、V95、Y96、S98、K99、V100、I101、P108、P109、T110、S111、K112以及T114。所述变异体可以包括任何数目的所述位置处以及所述位置的任何组合处的修饰。在一方面,所述修饰可以是氨基酸的取代、缺失或添加,并且优选地是取代或缺失突变。优选的修饰在下文中在标题“进一步修饰”下讨论。突变胞溶素单体可以包括SEQ ID NO:2的其它位置处的修饰。例如,除了本发明的一个或多个、例如2个到20个、3个到15个、4个到10个或6个到8个修饰之外,突变胞溶素单体可以具有SEQ ID NO:2的序列中在所属领域中、例如在WO 2013/153359中描述的一个或多个、如2个到20个、3个到15个、4个到10个或6个到8个氨基酸取代或缺失。

[0156] 所述变异体优选地包括以下位置T91、V95、Y96、S98、K99、V100、I101和K112中的一个或多个处的修饰。所述变异体可以具有任何数目的所述位置处以及所述位置的任何组合处的修饰。所述修饰优选地是使用丝氨酸(S)或谷氨酰胺(Q)进行的取代。所述变异体优选地包括取代T91S、V95S、Y96S、S98Q、K99S、V100S、I101S和K112S中的一个或多个。所述变异体可以包括任何数目的这些取代以及这些取代的任何组合。

[0157] 所述变异体优选地包括以下位置中的一个或多个处的修饰:K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、Y96、S98、K99、P108、P109、T110、S111以及T114。所述变异体可以包括任何数目的所述位置处以及所述位置的任何组合处的修饰。所述修饰优选地是使用天冬酰胺(N)、色氨酸(W)、丝氨酸(S)、谷氨酰胺(Q)、赖氨酸(K)、天冬氨酸(D)、精氨酸(R)、苏氨酸(T)、酪氨酸(Y)、亮氨酸(L)或异亮氨酸(I)进行的取代。所述变异体优选地包括以下取代中的一个或多个:K37N/W/S/Q、G43K、K45D/R/N/Q/T/Y、V47K/S/N、S49K/L、T51K、H83S/K、V88I/T、T91K、T93K、Y96D、S98K、K99Q/L、P108K/R、P109K、T110K/R、S111K以及T114K。所述变异体优选地包括以下位置中的一个或多个处的修饰:

E94/P108;	E94/S98/K99/T106;
E94/P109;	K37/E94/T106;
E94/T110;	K37/E94/T106;
E94/P108;	K37/E94/T106;
E94D/T110R;	K45/E94/T106;
E94D/S111K;	K37/E94/E102/T106;
E94D/T114K;	K37/E94/E102/T106;
H83S/E94Q;	K37/E94/T104/T106;
E94/K99/T106;	K45/E94/T106;
E94/T93/T106;	K45/V47/E94/T106;
E94/T91/T106;	V47/E94/T106;
H83/E94/T106;	T51/E94/T106;
E94/Y96/T106;	K45/S49/E94/T106;
K45/E94/T106;	S49/E94/T106;
K45/E94/T106;	K45/T106;
V47/E94/T106;	K45/S49/E94/E92/T106;
G43/E94/T106;	K45/V47/E92/E94/T106; 以及
V88/M90/E94/T106;	E94/K99/T106。
V47/V88/E94/T106;	

[0158] 所述变异体优选地包括以下取代中的一个或多个：

	E94D/P108K;	K45N/E94N/T106K;
	E94D/P109K;	K37Q/E94D/E102N/T106K;
	E94D/T110K;	K37S/E94D/E102S/T106K;
	E94D/P108R;	K37S/E94D/T104K/T106K;
	E94D/T110R;	K45Q/E94Q/T106K;
	E94D/S111K;	K45T/V47K/E94D/T106K;
	E94D/T114K;	V47S/E94D/T106K;
	H83S/E94Q;	T51K/E94D/T106K;
	E94D/K99Q/T106K;	K45Y/S49K/E94D/T106K;
[0159]	E94D/T93K/T106K;	S49L/E94D/T106K;
	E94D/T91K/T106K;	K45R/T106K;
	H83K/E94D/T106K;	V47K/E94D/T106K;
	E94Q/Y96D/T106K;	G43K/E94D/T106K;
	K45D/E94K/T106K;	V88I/M90A/E94D/T106K;
	K45R/E94D/T106K;	V47N/V88T/E94D/T106K;
	E94D/S98K/K99L/T106K;	K45N/S49K/E94N/E92D/T106K;
	K37N/E94D/T106K;	K45N/V47K/E92D/E94N/T106K; 以及
	K37W/E94D/T106K;	E94D/K99Q/T106K。
	K37S/E94D/T106K;	

[0160] 本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中所述单体能够形成孔,并且其中所述变异体包括以下取代中的一个或多个:

- [0161] D35N/S;
- [0162] S74K/R;
- [0163] E76D/N;
- [0164] S78R/K/N/Q;
- [0165] S80K/R/N/Q;
- [0166] S82K/R/N/Q;
- [0167] E84R/K/N/A;
- [0168] E85N;
- [0169] S86K/Q;
- [0170] S89K;
- [0171] M90K/I/A;
- [0172] E92D/S;
- [0173] E94D/Q/G/A/K/R/S/N;
- [0174] E102N/Q/D/S;
- [0175] T104R/K/Q;
- [0176] T106R/K/Q;

[0177] R115S;

[0178] Q117S;以及

[0179] N119S。

[0180] 所述变异体可以包括任何数目的这些取代以及这些取代的任何组合。所述变异体优选地包括以下取代中的一个或多个: E94D/Q/G/A/K/R/S、S86Q以及E92S, 如E94D/Q/G/A/K/R/S; S86Q; E92S; E94D/Q/G/A/K/R/S和S86Q; E94D/Q/G/A/K/R/S和E92S; S86Q和E92S; 或E94D/Q/G/A/K/R/S、S86Q和E92S。

[0181] 所述变异体优选地包括以下取代中的一个或多个:

[0182] D35N/S;

[0183] S74K/R;

[0184] E76D/N;

[0185] S78R/K/N/Q;

[0186] S80K/R/N/Q;

[0187] S82K/R/N/Q;

[0188] E84R/K/N/A;

[0189] E85N;

[0190] S86K;

[0191] S89K;

[0192] M90K/I/A;

[0193] E92D;

[0194] E94D/Q/K/N;

[0195] E102N/Q/D/S;

[0196] T104R/K/Q;

[0197] T106R/K/Q;

[0198] R115S;

[0199] Q117S;以及

[0200] N119S。

[0201] 所述变异体可以包括任何数目的这些取代以及这些取代的组合。

[0202] 所述变异体优选地包括以下取代中的一个或多个:

E94D/E102N;	S78Q/E94Q;
E94D/E102Q;	S80Q/E94Q;
E94D/S80K;	S82Q/E94Q;
S82K/E94D;	H83S/E94Q;
E94D/T106R;	E84A/E94Q;
E94D/T106K;	E84K/E94Q;
E94D/T104R;	E94Q/T104Q;
E94D/T104K;	E94Q/T106Q;
S78R/E94D;	E94Q/R115S;
S78K/E94D;	E94Q/Q117S;
S80R/E94D;	E94Q/N119S;
S82R/E94D;	E94Q/D121S;
E76D/E94D;	E76S/E94Q;
E76N/E94D;	E94D/K99Q/T106K;
E94D/E102D;	E94D/T93K/T106K;
[0203] E84R/E94D;	E94D/T91K/T106K;
E84K/E94D;	E94D/M90K/T106K;
E84N/E94D;	E85N/E94D/T106K;
S78N/E94D;	H83K/E94D/T106K;
S80N/E94D;	E94Q/Y96D/T106K;
S82N/E94D;	K45D/E94K/T106K;
E94D/P108K;	K45R/E94D/T106K;
E94D/P109K;	E94D/S98K/K99L/T106K;
S74K/E94D;	D35N/E94D/T106K;
E94D/T110K;	D35S/E94D/T106K;
S74R/E94D;	K37N/E94D/T106K;
E94D/P108R;	K37W/E94D/T106K;
E94D/T110R;	K37S/E94D/T106K;
S86K/E94D;	K45N/E94N/T106K;
S89K/E94D;	E92D/E94Q/T106K;
E94D/S111K;	K37Q/E94D/E102N/T106K;

	E94D/T114K;	E94Q/T106K;
	E76N/E94Q;	K37S/E94D/E102S/T106K;
	K37S/E94D/T104K/T106K;	V47K/E94D/T106K;
	K45Q/E94Q/T106K;	G43K/E94D/T106K;
	M90I/E94D/T106K;	V88I/M90A/E94D/T106K;
[0204]	K45T/V47K/E94D/T106K;	V47N/V88T/E94D/T106K;
	V47S/E94D/T106K;	K45N/S49K/E94N/E92D/T106K;
	T51K/E94D/T106K;	K45N/V47K/E92D/E94N/T106K;
	K45Y/S49K/E94D/T106K;	E94D/K99Q/T106K;
	S49L/E94D/T106K;	S82K/E94D/T106K; 以及
	K45R/T106K;	Y79S/E94Q。

[0205] 所述变异体可以包括任何数目的这些取代以及这些取代的任何组合。

[0206] 本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中所述单体能够形成孔,并且其中所述变异体包括以下中的一个或多个处的突变:

- [0207] D35/E94/T106;
- [0208] K37/E94/E102/T106;
- [0209] K37/E94/T104/T106;
- [0210] K37/E94/T106;
- [0211] K37/E94/E102/T106;
- [0212] G43/E94/T106;
- [0213] K45/V47/E92/E94/T106;
- [0214] K45/V47/E94/T106;
- [0215] K45/S49/E92/E94/T106;
- [0216] K45/S49/E94/T106;
- [0217] K45/E94/T106;
- [0218] K45/T106;
- [0219] V47/E94/T106;
- [0220] V47/V88/E94/T106;
- [0221] S49/E94/T106;
- [0222] T51/E94D/T106;
- [0223] S74/E94;
- [0224] E76/E94;
- [0225] S78/E94;
- [0226] Y79/E94;
- [0227] S80/E94;
- [0228] S82/E94;
- [0229] S82/E94/T106;

- [0230] H83/E94;
- [0231] H83/E94/T106;
- [0232] E85/E94/T106;
- [0233] S86/E94;
- [0234] V88/M90/E94/T106;
- [0235] S89/E94;
- [0236] M90/E94/T106;
- [0237] T91/E94/T106;
- [0238] E92/E94/T106;
- [0239] T93/E94/T106;
- [0240] E94/Y96/T106;
- [0241] E94/S98/K99/T106;
- [0242] E94/K99/T106;
- [0243] E94/E102;
- [0244] E94/T104;
- [0245] E94/T106;
- [0246] E94/P108;
- [0247] E94/P109;
- [0248] E94/T110;
- [0249] E94/S111;
- [0250] E94/T114;
- [0251] E94/R115;
- [0252] E94/Q117;以及
- [0253] E94/E119。
- [0254] 所述变异体优选地包括以下取代中的一个或多个:D35N/E94D/T106K;
- [0255] D35S/E94D/T106K;
- [0256] K37Q/E94D/E102N/T106K;
- [0257] K37S/E94D/E102S/T106K;
- [0258] K37S/E94D/T104K/T106K;
- [0259] K37N/E94D/T106K;
- [0260] K37W/E94D/T106K;
- [0261] K37S/E94D/T106K;
- [0262] G43K/E94D/T106K;
- [0263] K45N/V47K/E92D/E94N/T106K;
- [0264] K45T/V47K/E94D/T106K;
- [0265] K45N/S49K/E94N/E92D/T106K;
- [0266] K45Y/S49K/E94D/T106K;
- [0267] K45D/E94K/T106K;
- [0268] K45R/E94D/T106K;

- [0269] K45N/E94N/T106K;
- [0270] K45Q/E94Q/T106K;
- [0271] K45R/T106K;
- [0272] V47S/E94D/T106K;
- [0273] V47K/E94D/T106K;
- [0274] V47N/V88T/E94D/T106K;
- [0275] S49L/E94D/T106K;
- [0276] T51K/E94D/T106K;
- [0277] S74K/E94D;
- [0278] S74R/E94D;
- [0279] E76D/E94D;
- [0280] E76N/E94D;
- [0281] E76S/E94Q;
- [0282] E76N/E94Q;
- [0283] S78R/E94D;
- [0284] S78K/E94D;
- [0285] S78N/E94D;
- [0286] S78Q/E94Q;
- [0287] Y79S/E94Q;
- [0288] S80K/E94D;
- [0289] S80R/E94D;
- [0290] S80N/E94D;
- [0291] S80Q/E94Q;
- [0292] S82K/E94D;
- [0293] S82R/E94D;
- [0294] S82N/E94D;
- [0295] S82Q/E94Q;
- [0296] S82K/E94D/T106K;
- [0297] H83S/E94Q;
- [0298] H83K/E94D/T106K;
- [0299] E85N/E94D/T106K;
- [0300] S86K/E94D;
- [0301] V88I/M90A/E94D/T106K;
- [0302] S89K/E94D;
- [0303] M90K/E94D/T106K;
- [0304] M90I/E94D/T106K;
- [0305] T91K/E94D/T106K;
- [0306] E92D/E94Q/T106K;
- [0307] T93K/E94D/T106K;

[0308] E94Q/Y96D/T106K;

[0309] E94D/S98K/K99L/T106K;

[0310] E94D/K99Q/T106K;

[0311] E94D/E102N;

[0312] E94D/E102Q;

[0313] E94D/E102D;

[0314] E94D/T104R;

[0315] E94D/T104K;

[0316] E94Q/T104Q;

[0317] E94D/T106R;

[0318] E94D/T106K;

[0319] E94Q/T106Q;

[0320] E94Q/T106K;

[0321] E94D/P108K;

[0322] E94D/P108R;

[0323] E94D/P109K;

[0324] E94D/T110K;

[0325] E94D/T110R;

[0326] E94D/S111K;

[0327] E94D/T114K;

[0328] E94Q/R115S;

[0329] E94Q/Q117S;以及

[0330] E94Q/N119S。

[0331] 所述变异体可以包括任何数目的这些取代以及这些取代的任何组合。

[0332] 本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中所述单体能够形成孔,并且其中所述变异体包括以下取代中的一个或多个:

[0333] E84R/E94D;

[0334] E84K/E94D;

[0335] E84N/E94D;

[0336] E84A/E94Q;

[0337] E84K/E94Q以及

[0338] E94Q/D121S。

[0339] 所述变异体可以包括任何数目的这些取代以及这些取代的任何组合。

[0340] 本发明的突变单体优选地包括以上所述的修饰和/或取代的任何组合。示例性组合在实例中予以公开。

[0341] 桶缺失

[0342] 在另一实施例中,本发明还提供了一种突变胞溶素单体,其包括SEQ ID NO:2中所示的序列的变异体,其中在所述变异体中,(a)SEQ ID NO:2的位置34到70处的氨基酸中的2个、4个、6个、8个、10个、12个、14个、16个、18个或20个已经缺失,或者其中与SEQ ID NO:2的

位置34到70相对应的氨基酸已经缺失,并且(b) SEQ ID NO:2的位置71到107处的氨基酸中的2个、4个、6个、8个、10个、12个、14个、16个、18个或20个已经缺失,或者其中与SEQ ID NO:2的位置71到位置107相对应的位置处的氨基酸残基已经缺失,。

[0343] 从位置34到70缺失的氨基酸的数目可以不同于从位置71到107缺失的氨基酸的数目。从位置34到70缺失的氨基酸的数目优选地与从位置71到107缺失的氨基酸的数目相同。

[0344] 可以缺失来自位置34到70的氨基酸以及来自位置71到107的氨基酸。已经缺失的氨基酸的位置优选地在表1或表2的一行中或表1和/或表2的多于一行中示出。例如,如果从位置34到70缺失D35和V34,则可以从位置71到107缺失T104和I105。类似地,可以从位置34到70缺失D35、V34、K37和I38,并且可以从位置71到107缺失E102、H103、T104和I105。这确保了维持孔的桶的 β 折叠结构内衬。

[0345] 表1

V34 与 F70 之间面向桶的供缺失残基	V34 与 F70 之间面向外朝着膜的对应供缺失残基	I107 与 E71 之间面向桶的供缺失残基	面向外朝着膜 I107 与 E71 的对应供缺失残基
D35	V34	T104	I105
D35	V34	T104	H103
D35	Q36	T104	I105
D35	Q36	T104	H103
K37	Q36	E102	H103
K37	Q36	E102	I101
K37	I38	E102	H103
K37	I38	E102	I101
T39	I38	V100	I101
T39	I38	V100	K99
T39	I40	V100	I101
T39	I40	V100	K99
T41	I40	S98	K99
T41	I40	S98	E97
T41	K42	S98	K99
T41	K42	S98	E97
G43	K42	Y96	E97
G43	K42	Y96	V95
G43	M44	Y96	E97
G43	M44	Y96	V95
K45	M44	E94	V95
K45	M44	E94	T93
K45	N46	E94	V95
K45	N46	E94	T93
V47	N46	E92	T93
V47	N46	E92	T91
V47	N48	E92	T93
V47	N48	E92	T91
S49	N48	M90	T91
S49	N48	M90	S89
S49	E50	M90	T91

[0346]

[0347]

S49	E50	M90	S89
T51	E50	V88	S89
T51	E50	V88	Q87
T51	R52	V88	S89
T51	R52	V88	Q87
T53	R52	S86	Q87
T53	R52	S86	E85
T53	V54	S86	Q87
T53	V54	S86	E85
T55	V54	E84	E85
T55	V54	E84	H83
T55	A56	E84	E85
T55	A56	E84	H83
T57	A56	S82	H83
T57	A56	S82	H81
T57	H58	S82	H83
T57	H58	S82	H81
S59	H58	S80	H81
S59	H58	S80	Y79
S59	I60	S80	H81
S59	I60	S80	Y79
G61	I60	S78	Y79
G61	I60	S78	V77
G61	S62	S78	Y79
G61	S62	S78	V77
T63	S62	E76	V77
T63	S62	E76	V75
T63	I64	E76	V77
T63	I64	E76	V75
S65	I64	S74	V75
S65	I64	S74	G73
S65	T66	S74	V75
S65	T66	S74	G73
G67	T66	I72	G73
G67	T66	I72	E71
G67	D68	I72	G73
G67	D68	I72	E71
A69	D68	I72	G73
A69	D68	I72	E71
A69	D70	I72	G73
A69	D70	I72	E71

[0348] 表2

[0349]

I107 与 E71 之间面向桶的供缺失残基	面向外朝着膜 I107 与 E71 的对应供缺失残基	V34 与 F70 之间面向桶的供缺失残基	V34 与 F70 之间面向外朝着膜的对应供缺失残基
T106	I107	D35	V34
T106	I107	D35	Q36

[0350]

T106	I105	D35	V34
T106	I105	D35	Q36
T104	I105	D35	V34
T104	I105	D35	Q36
T104	H103	D35	V34
T104	H103	D35	Q36
E102	H103	K37	Q36
E102	H103	K37	I38
E102	I101	K37	Q36
E102	I101	K37	I38
V100	I101	T39	I38
V100	I101	T39	I40
V100	K99	T39	I38
V100	K99	T39	I40
S98	K99	T41	I40
S98	K99	T41	K52
S98	E97	T41	I40
S98	E97	T41	K52
Y96	E97	G43	K52
Y96	E97	G43	M44
Y96	V95	G43	K52
Y96	V95	G43	M44
E94	V95	K45	M44
E94	V95	K45	N46
E94	T93	K45	M44
E94	T93	K45	N46
E92	T93	V47	N46
E92	T93	V47	N48
E92	T91	V47	N46
E92	T91	V47	N48
M90	T91	S49	N48
M90	T91	S49	E50
M90	S89	S49	N48
M90	S89	S49	E50
V88	S89	T51	E50
V88	S89	T51	R52
V88	Q87	T51	E50
V88	Q87	T51	R52
S86	Q87	T53	R52
S86	Q87	T53	V54
S86	E85	T53	R52
S86	E85	T53	V54
E84	E85	T55	V54
E84	E85	T55	A56
E84	H83	T55	V54
E84	H83	T55	A56
S82	H83	T57	A56
S82	H83	T57	H58
S82	H81	T57	A56

	S82	H81	T57	H58
	S80	H81	S59	H58
	S80	H81	S59	I60
	S80	Y79	S59	H58
	S80	Y79	S59	I60
	S78	Y79	G61	I60
	S78	Y79	G61	S62
	S78	V77	G61	I60
	S78	V77	G61	S62
	E76	V77	T63	S62
	E76	V77	T63	I64
	E76	V75	T63	S62
[0351]	E76	V75	T63	I64
	S74	V75	S65	I64
	S74	V75	S65	T66
	S74	G73	S65	I64
	S74	G73	S65	T66
	I72	G73	G67	T66
	I72	G73	G67	D68
	I72	E71	G67	T66
	I72	E71	G67	D68
	I72	G73	A69	D68
	I72	G73	A69	F70
	I72	E71	A69	D68
	I72	E71	A69	F70

[0352] 从位置34到70和位置71到107缺失的氨基酸不必处于表1或2的一行中。例如,如果从位置34到70缺失D35和V34,则可以从位置71到107缺失I72和E71。

[0353] 从位置34到70缺失的氨基酸优选地是连续的。从位置71到107缺失的氨基酸优选地是连续的。从位置34到70缺失的氨基酸和从位置71到107缺失的氨基酸优选地是连续的。

[0354] 本发明优选地提供了突变单体,其中以下已经缺失:

[0355] (i)N46/V47/T91/T92;或

[0356] (ii)N48/S49/T91/T92。

[0357] 技术人员可以根据本发明识别可以缺失的氨基酸的其它组合。以下讨论使用SEQ ID NO:2中的残基的编号(即,在任何氨基酸如上文所述那样缺失之前)。

[0358] 桶缺失变异体进一步优选地包括在适当的情况下上文或下文中讨论的修饰和/或取代中的任何修饰和/或取代。“在适当的情况下”是指在桶缺失之后位置是否仍然存在于突变单体中。

[0359] 化学修饰

[0360] 在另一方面,本发明提供了经化学修饰的突变胞溶素单体。突变单体可以是上文或下文所讨论的突变单体中的任何突变单体。因此,可以根据本发明、如下文所讨论的那样对本发明的突变单体,如包括以下位置中的一个或多个处的修饰的SEQ ID NO:2的变异体:K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、V95、Y96、S98、K99、V100、I101、P108、P109、T110、S111、K112和T114或包括上述桶缺失的变异体进行化学修饰。

[0361] 可以对包括下文讨论的进一步修饰中的任何进一步修饰,即包括SEQ ID NO:2的约位置44位到约位置126的区域内、改变单体或优选地所述区域与多核苷酸相互作用的能力的一个或多个修饰的突变单体进行化学修饰。这些经化学修饰的单体不需要包括本发明

的修饰,即不需要包括以下位置中的一个或多个处的修饰:K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、V95、Y96、S98、K99、V100、I101、P108、P109、T110、S111、K112以及T114。经化学修饰的突变单体优选地包括SEQ ID NO:2的变异体,所述变异体包括SEQ ID NO:2的以下位置中的一个或多个处的取代:(a)E84、E85、E92、E97和D126;(b)E85、E97和D126或者(c)E84和E92。可以进行任何数目的下文讨论的取代或其任何组合。

[0362] 可以以任何方式对突变单体进行化学修饰,使得由单体形成的孔的桶或通道的直径减小或缩小。在下文中更详细地讨论这一点。

[0363] 化学修饰为使得化学分子优选地共价连接到突变单体。可以使用所属领域中已知的任何方法将化学分子共价连接到突变单体。化学分子通常通过化学连接附接。

[0364] 优选地通过将分子附接到一个或多个半胱氨酸(半胱氨酸连接)、将分子附接到一个或多个赖氨酸、将分子附接到一个或多个非天然氨基酸、表位的酶修饰对突变单体进行化学修饰。如果通过半胱氨酸连接附接化学修饰剂,则所述一个或多个半胱氨酸已经优选地通过取代引入到突变单体。用于执行这种修饰的适当方法在所属领域中是众所周知的。适当的非天然氨基酸包含但不限于4-叠氮基-L-苯丙氨酸(Faz)和Liu C.C.和Schultz P.G.,《生物化学年评(Annu.Rev.Biochem.)》,2010年,第79卷,第413到444页的图1中编号为1到71的氨基酸中的任何氨基酸。

[0365] 可以通过将具有减少或缩小由单体形成的孔的桶的直径的作用的任何分子附接在任何位置或位点处对突变单体进行化学修饰。可以通过附接如下各项来对突变单体进行化学修饰:(i) 马来酰亚胺,如:4-苯氮霉素、1.N-(2-羟乙基)马来酰亚胺、N-环己基马来酰亚胺、1.3-马来酰亚胺基丙酸、1.1-4-氨基苯基-1H-吡咯,2,5,二酮、1.1-4-羟基苯基-1H-吡咯,2,5,二酮、N-乙基马来酰亚胺、N-甲氧基羰基马来酰亚胺、N-叔丁基马来酰亚胺、N-(2-氨基乙基)马来酰亚胺、3-马来酰亚胺基-PROXYL、N-(4-氯苯基)马来酰亚胺、1-[4-(二甲氨基)-3,5-二硝基苯基]-1H-吡咯-2,5-二酮、N-[4-(2-苯并咪唑基)苯基]马来酰亚胺、N-[4-(2-苯并恶唑基)苯基]马来酰亚胺、N-(1-萘基)马来酰亚胺、N-(2,4-二甲苯基)马来酰亚胺、N-(2,4-二氟苯基)马来酰亚胺、N-(3-氯-对-甲苯基)-马来酰亚胺、1-(2-氨基-乙基)-吡咯-2,5-二酮盐酸盐、1-环戊基-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、1-(3-氨基丙基)-2,5-二氢-1H-吡咯-2,5-二酮盐酸盐、3-甲基-1-[2-氧代-2-(哌嗪-1-基)乙基]-2,5-二氢-1H-吡咯-2,5-二酮盐酸盐、1-苄基-2,5-二氢-1H-吡咯-2,5-二酮、3-甲基-1-(3,3,3-三氟丙基)-2,5-二氢-1H-吡咯-2,5-二酮、1-[4-(甲基氨基)环己基]-2,5-二氢-1H-吡咯-2,5-二酮三氟乙酸、SMILES O=C1C=CC(=O)N1CC=2C=CN=CC2、SMILES O=C1C=CC(=O)N1CN2CCNCC2、1-苄基-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、1-(2-氟苯基)-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、N-(4-苯氧基苯基)马来酰亚胺、N-(4-硝基苯基)马来酰亚胺;(ii) 碘代乙酰胺,如3-(2-碘乙酰氨基)-PROXYL、N-(环丙基甲基)-2-碘乙酰胺、2-碘-N-(2-苯乙基)乙酰胺、2-碘-N-(2,2,2-三氟乙基)乙酰胺、N-(4-乙酰基苯基)-2-碘代乙酰胺、N-(4-(氨基磺酰基)苯基)-2-碘代乙酰胺、N-(1,3-苯并噻唑-2-基)-2-碘代乙酰胺、N-(2,6-二乙基苯基)-2-碘代乙酰胺、N-(2-苯甲酰基-4-氯苯基)-2-碘代乙酰胺;(iii) 溴代乙酰胺:如N-(4-(乙酰氨基)苯基)-2-溴代乙酰胺、N-(2-乙酰基苯基)-2-溴代乙酰胺、2-溴-N-(2-氰基苯基)乙酰胺、2-溴-N-(3-(三氟甲基)苯基)乙酰胺、N-(2-苯甲酰基苯基)-2-溴代乙酰胺、2-溴-N-(4-氟苯基)-3-甲基丁酰胺、N-苄基-2-溴-N-苯基丙酰胺、N-(2-溴-丁酰

基)-4-氯-苯磺酰胺、2-溴-N-甲基-N-苯基乙酰胺、2-溴-N-苯乙基-乙酰胺、2-金刚烷-1-基-2-溴-N-环己基-乙酰胺、2-溴-N-(2-甲基苯基)丁酰胺、乙酰替对溴苯胺；(iv)二硫化物，如：ALDRITHIOL-2、ALDRITHIOL-4、异丙基二硫化物、1-(异丁基二硫烷基)-2-甲基丙烷、二苄基二硫化物、4-氨基苯基二硫化物、3-(2-吡啶基二硫代)丙酸、3-(2-吡啶基二硫代)丙酸酰肼、3-(2-吡啶基二硫代)丙酸N-琥珀酰亚胺酯、am6amPDP1-βCD；以及(v)硫醇，如：4-苯基噻唑-2-硫醇、PulpaId、5,6,7,8-四氢-喹唑啉-2-硫醇。

[0366] 可以通过附接聚乙二醇(PEG)、如DNA等核酸、染料、荧光团或发色团对突变单体进行化学修饰。在一些实施例中，使用促进包括单体的孔与靶分析物、靶核苷酸或靶多核苷酸序列之间的相互作用的分子衔接子对突变单体进行化学修饰。衔接子的存在改进了孔和核苷酸或多核苷酸的主客体化学，并且由此提高了由突变单体形成的孔的测序能力。

[0367] 化学修饰的突变单体优选地包括SEQ ID NO:2中所示的序列的变异体。变异体定义如下。所述变异体通常包括一个或多个取代，其中一个或多个残基被半胱氨酸、赖氨酸或非天然氨基酸取代。非天然氨基酸包含但不限于：4-叠氮基-L-苯丙氨酸(Faz)、4-乙酰基-L-苯丙氨酸、3-乙酰基-L-苯丙氨酸、4-乙酰乙酰基-L-苯丙氨酸、0-烯丙基-L-酪氨酸、3-(苯基乙烯基)-L-丙氨酸、0-2-丙炔-1-基-L-酪氨酸、4-(二羟基硼基)-L-苯丙氨酸、4-[(乙基硫烷基)羰基]-L-苯丙氨酸、(2S)-2-氨基-3-4-[(丙-2-基硫烷基)羰基]苯基；丙酸、(2S)-2-氨基-3-4-[(2-氨基-3-硫烷基丙酰基)氨基]苯基；丙酸、0-甲基-L-酪氨酸、4-氨基-L-苯丙氨酸、4-氰基-L-苯丙氨酸、3-氰基-L-苯丙氨酸、4-氟-L-苯丙氨酸、4-碘-L-苯丙氨酸、4-溴-L-苯丙氨酸、0-(三氟甲基)酪氨酸、4-硝基-L-苯丙氨酸、3-羟基-L-酪氨酸、3-氨基-L-酪氨酸、3-碘-L-酪氨酸、4-异丙基-L-苯丙氨酸、3-(2-萘基)-L-丙氨酸、4-苯基-L-苯丙氨酸、(2S)-2-氨基-3-(萘-2-基氨基)丙酸、6-(甲基硫烷基)正亮氨酸、6-氧代-L-赖氨酸、D-酪氨酸、(2R)-2-羟基-3-(4-羟基苯基)丙酸、(2R)-2-氨基辛酸酯3-(2,2'-联吡啶-5-基)-D-丙氨酸、2-氨基-3-(8-羟基-3-喹啉基)丙酸、4-苯甲酰基-L-苯丙氨酸、S-(2-硝基苄基)半胱氨酸、(2R)-2-氨基-3-[(2-硝基苄基)硫烷基]丙酸、(2S)-2-氨基-3-[(2-硝基苄基)氧基]丙酸、0-(4,5-二甲氧基-2-硝基苄基)-L-丝氨酸、(2S)-2-氨基-6-[(2-硝基苄基)氧基]羰基；氨基)己酸、0-(2-硝基苄基)-L-酪氨酸、2-硝基苯丙氨酸、4-[(E)-苯基二氮烯基]-L-苯丙氨酸、4-[3-(三氟甲基)-3H-二氮杂环庚烷-3-基]-D-苯丙氨酸、2-氨基-3-[[5-(二甲基氨基)-1-萘基]磺酰基氨基]丙酸、(2S)-2-氨基-4-(7-羟基-2-氧代-2H-苯并吡喃-4-基)丁酸、(2S)-3-[(6-乙酰基萘乙酰胺-2-基)氨基]-2-氨基丙酸、4-(羧甲基)苯丙氨酸、3-硝基-L-酪氨酸、0-磺基-L-酪氨酸、(2R)-6-乙酰氨基-2-氨基己酸、1-甲基组氨酸、2-氨基壬酸、2-氨基癸酸、-L-高半胱氨酸、5-硫烷基正缬氨酸、6-硫烷基-L-正亮氨酸、5-(甲硫基)-L-正缬氨酸、N⁶-[(2R,3R)-3-甲基-3,4-二氢-2H-吡咯-2-基]羰基；-L-赖氨酸、N⁶-[(苄氧基)羰基]赖氨酸、(2S)-2-氨基-6-[(环戊基羰基)氨基]己酸、N⁶-[(环戊基氧基)羰基]-L-赖氨酸、(2S)-2-氨基-6-[(2R)-四氢呋喃-2-基羰基]氨基；己酸、(2S)-2-氨基-8-[(2R,3S)-3-乙炔基四氢呋喃-2-基]-8-氧代辛酸、N⁶-(叔丁氧羰基)-L-赖氨酸、(2S)-2-羟基-6-[(2-甲基-2-丙基)氧基]羰基；氨基)己酸、N⁶-[(烯丙氧基)羰基]赖氨酸、(2S)-2-氨基-6-[(2-叠氮基苄基)氧基]羰基；氨基)己酸、N⁶-L-脯氨酰基-L-赖氨酸、(2S)-2-氨基-6-[(丙-2-炔-1-基氧基)羰基]氨基；己酸以及N⁶-[(2-叠氮基乙氧基)羰基]-L-赖氨酸。最优选的非天然氨基酸是4-叠氮基-L-苯丙氨酸(Faz)。

[0368] 可以通过SEQ ID NO:2的以下位置中的任何位置处附接任何分子对突变单体进行化学修饰:K37、V47、S49、T55、S86、E92以及E94。更优选地,可以通过在位置E92和/或E94处附接任何分子对突变单体进行化学修饰。在一个实施例中,通过这些位置将分子附接到一个或多个半胱氨酸(半胱氨酸连接)、一个或多个赖氨酸或一个或多个非天然氨基酸对突变单体进行化学修饰。突变单体优选地包括SEQ ID NO:2中所示序列的变异体,所述变异体包括K37C、V47C、S49C、T55C、S86C、E92C和E94C中的一个或多个,其中一个或多个分子附接到所述一个或多个引入的半胱氨酸。突变单体更优选地包括SEQ ID NO:2中所示序列的变异体,所述变异体其包括E92C和/或E94C,其中一个或多个分子附接到所引入的一个或多个半胱氨酸。在这两个优选实施例中的每一个中,所述一个或多个半胱氨酸(Cs)可以被一个或多个赖氨酸或如一个或多个Faz等一个或多个非天然氨基酸取代。

[0369] 可以通过修饰相邻残基来增强半胱氨酸残基的反应性。例如,侧接精氨酸、组氨酸或赖氨酸残基的碱性基团将半胱氨酸巯基的pKa改变更具反应性的S⁻基的pKa。可以通过如dTNB等巯基保护基来保护半胱氨酸残基的反应性。在附接接头之前,这些可以与突变单体的一个或多个半胱氨酸残基反应。

[0370] 可以将所述分子直接附接到突变单体。优选地,使用如化学交联剂或肽接头等接头将所述分子附接到突变单体。适当的化学交联剂在所属领域中是众所周知的。优选的交联剂包含3-(吡啶-2-基二磺酰基)丙酸2,5-二氧化吡咯烷-1-基酯、4-(吡啶-2-基二磺酰基)丁酸2,5-二氧化吡咯烷-1-基酯以及8-(吡啶-2-基二磺酰基)辛酸2,5-二氧化吡咯烷-1-基酯。最优的交联剂是3-(2-吡啶二硫代)丙酸丁二酰亚胺酯(SPDP)。通常,在分子/交联剂复合体共价连接到突变单体之前,分子共价连接到双功能交联剂,但也有可能在双功能交联剂/单体复合体附接到分子之前将双功能交联剂共价连接到单体。

[0371] 优选地,接头对二硫苏糖醇(DTT)具有抗性。适当的接头包含但不限于基于碘乙酰胺和基于马来酰亚胺的接头。

[0372] 以下更详细地讨论了包括本发明的经化学修饰的突变单体的孔的优点。

[0373] 下文讨论了可以根据本发明进行的进一步化学修饰。

[0374] 进一步修饰

[0375] 在适当的情况下,以上讨论的突变单体中的任何单体可以包括SEQ ID NO:2的约位置44到约位置126的区域内的进一步修饰(即,在相关的氨基位置保持处于突变单体中或未被另一个氨基酸修饰/取代的情况下)。此区域的至少一部分通常贡献于胞溶素的跨膜区域。此区域的至少一部分通常贡献于胞溶素的桶或通道。此区域的至少一部分通常贡献于胞溶素的内壁或内衬。

[0376] 已经将胞溶素的跨膜区域确定为SEQ ID NO:2的位置44到67(De Colbis等人,《结构(Structure)》,2012年,第20卷,第1498到1507页)。

[0377] 所述变异体优选地包括SEQ ID NO:2的约位置44到约位置126的区域内的一个或多个修饰,所述修饰改变单体或优选地所述区域与多核苷酸相互作用的能力。可以增加或减少单体与多核苷酸之间的相互作用。单体与多核苷酸之间的相互作用增加将例如促进通过包括突变单体的孔捕获多核苷酸。所述区域与多核苷酸之间相互作用减少将例如提高对多核苷酸的识别或区分。通过减少包括突变单体的孔的状态的变化(这增加信噪比)和/或减少多核苷酸中在多核苷酸移动穿过包括突变单体的孔时贡献于电流的核苷酸的数目,可

以提高对多核苷酸的识别或区分。

[0378] 单体与多核苷酸相互作用的能力可以使用所属领域中熟知的方法来确定。单体可以以任何方式与多核苷酸相互作用,例如通过非共价相互作用,如疏水相互作用、氢结合、范德瓦耳力, π (π) -阳离子相互作用或静电力。例如,可以使用常规的结合测定来测量所述区域结合到多核苷酸的能力。适当的测定包含但不限于基于荧光的结合测定、核磁共振(NMR)、等温滴定量热法(ITC)或电子自旋共振(ESR)光谱。可替代地,可以使用上文或下文讨论的方法中的任何方法确定包含所述突变单体中的一个或多个的孔与多核苷酸相互作用的能力。在实例中描述了优选的测定。

[0379] 可以在SEQ ID NO:2的约位置44到约位置126的区域内进一步进行一个或多个修饰。优选地,所述一个或多个修饰在以下区域中的任何区域内进行:约位置40到约位置125、约位置50到约位置120、约位置60到约位置110以及约位置70到约位置100。如果进行所述一个或多个修饰以提高多核苷酸捕获,则更优选地在以下区域中的任何区域内进行所述修饰:约位置44到约位置103、约位置68到约位置103、约位置84到约位置103、约位置44到约位置97、约位置68到约位置97或约位置84到约位置97。如果进行所述一个或多个修饰以提高多核苷酸识别或区分,则更优选地在以下区域中的任何区域内进行所述修饰:约位置44到约位置109、约位置44到约位置97或约位置48到约位置88。优选地,所述区域为SEQ ID NO:2的约位置44到约位置67。

[0380] 如果所述一个或多个修饰旨在提高多核苷酸识别或区分,则优选地,除了用于提高多核苷酸捕获之外的一个或多个修饰之外,还进行所述修饰。这允许由突变单体形成的孔有效捕获多核苷酸,并且然后表征多核苷酸,如估计其序列,如下文所讨论的。

[0381] 改变蛋白纳米孔与多核苷酸相互作用的能力,特别是提高其捕获和/或识别或区分多核苷酸的能力的蛋白纳米孔修饰在所属领域中有很好的记载。例如,在WO 2010/034018和WO 2010/055307中公开了这种修饰。可以对根据本发明的胞溶素单体进行类似的修饰。

[0382] 可以进行任何数目的修饰,如1个、2个、5个、10个、15个、20个、30个或更多个修饰。只要单体与多核苷酸相互作用的能力发生改变,就可以进行任何一个或多个修饰。适当的修饰包含但不限于氨基酸取代、氨基酸添加以及氨基酸缺失。优选地,所述一个或多个修饰是一个或多个取代。在下文中更详细地讨论这一点。

[0383] 所述一个或多个修饰优选地(a)改变单体的位阻效应或优选地改变所述区域的位阻效应;(b)改变所述单体的净电荷或优选地改变所述区域的净电荷;(c)改变所述单体或优选地所述区域与多核苷酸氢结合的能力;(d)引入或移除通过离域电子 π 系统相互作用的化学基团和/或(e)改变所述单体的结构或优选地改变所述区域的结构。所述一个或多个修饰更优选地产生(a)到(e)的任何组合,如(a)和(b);(a)和(c);(a)和(d);(a)和(e);(b)和(c);(b)和(d);(b)和(e);(c)和(d);(c)和(e);(d)和(e)、(a)、(b)和(c);(a)、(b)和(d);(a)、(b)和(e);(a)、(c)和(d);(a)、(c)和(e);(a)、(d)和(e);(b)、(c)和(d);(b)、(c)和(e);(b)、(d)和(e);(c)、(d)和(e);(a)、(b)、(c)和(d);(a)、(b)、(c)和(e);(a)、(b)、(d)和(e);(a)、(c)、(d)和(e);(b)、(c)、(d)和(e);以及(a)、(b)、(c)和(d)。

[0384] 对于(a),可以增加或减少单体的位阻效应。根据本发明可以使用任何改变位阻效应的方法。引入如苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)等大体积残基增加单体

的位阻。所述一个或多个修饰优选地是引入F、W、Y和H中的一个或多个。可以引入F、W、Y和H的任何组合。可以通过添加引入F、W、Y和H中的所述一个或多个。F、W、Y和H中的所述一个或多个优选地通过取代引入。以下更详细地讨论用于引入这种残基的适当位置。

[0385] 移除如苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)等大体积残基相反降低单体的位阻。所述一个或多个修饰优选地是移除F、W、Y和H中的一个或多个。可以移除F、W、Y和H的任何组合。可以通过缺失移除F、W、Y和H中的所述一个或多个。F、W、Y和H中的所述一个或多个优选地通过使用如丝氨酸(S)、苏氨酸(T)、丙氨酸(A)和缬氨酸(V)等具有较小侧基的残基进行取代而移除。

[0386] 对于(b),可以以任何方式改变净电荷。优选地增加或减少净正电荷。可以以任何方式增加净正电荷。优选地通过引入、优选地通过取代一个或多个带正电氨基酸和/或中和、优选通过取代一个或多个负电荷来增加净正电荷。

[0387] 优选地通过引入一个或多个带正电氨基酸来增加净正电荷。可以通过添加引入所述一个或多个带正电氨基酸。优选地通过取代引入所述一个或多个带正电氨基酸。带正电氨基酸是具有净正电氨基酸。所述一个或多个带正电氨基酸可以是天然存在的或非天然存在的。带正电氨基酸可以是合成的或修饰的。例如,具有净正电的经修饰氨基酸可以被专门设计用于本发明。对氨基酸的多种不同类型修饰在所属领域中是众所周知的。

[0388] 优选的天然存在的带正电氨基酸包含但不限于组氨酸(H)、赖氨酸(K)和精氨酸(R)。所述一个或多个修饰优选地是引入H、K和R中的一个或多个。可以引入任何数目的H、K和R和其任何组合。可以通过添加引入H、K和R中的所述一个或多个。H、K和R中的所述一个或多个优选地通过取代引入。以下更详细地讨论用于引入这种残基的适当位置。

[0389] 用于添加或取代天然存在的氨基酸的方法在所属领域中是众所周知的。例如,可以通过在对突变单体进行编码的多核苷酸中的相关位置处使用精氨酸的密码子(CGT)代替甲硫氨酸的密码子(ATG)而使用精氨酸(R)取代甲硫氨酸(M)。然后,可以如下文所讨论那样表达多核苷酸。

[0390] 用于添加或取代非天然存在的氨基酸的方法在所属领域中也是众所周知的。例如,可以通过在用于表达孔的IVTT系统中包括合成氨基酰基-tRNA来引入非天然存在的氨基酸。可替代地,可以通过在存在特定氨基酸的合成(即非天然存在的)类似物存在的情况下表达对于那些特定氨基酸来说是营养缺陷的大肠杆菌中表达突变单体来引入非天然存在的氨基酸。如果孔是使用部分肽合成来产生的,则非天然存在的氨基酸还可通过裸连接来产生。

[0391] 任何氨基酸可以被带正电氨基酸取代。一个或多个不带电氨基酸、非极性氨基酸和/或芳香族氨基酸可以被一个或多个带正电氨基酸取代。不带电氨基酸不具有净电荷。适当的不带电氨基酸包含但不限于半胱氨酸(C)、丝氨酸(S)、苏氨酸(T)、甲硫氨酸(M)、天冬酰胺(N)和谷氨酰胺(Q)。非极性氨基酸具有非极性侧链。适当的非极性氨基酸包含但不限于甘氨酸(G)、丙氨酸(A)、脯氨酸(P)、异亮氨酸(I)、亮氨酸(L)和缬氨酸(V)。芳香族氨基酸具有芳香族侧链。适当的芳香族氨基酸包含但不限于组氨酸(H)、苯丙氨酸(F)、色氨酸(W)和酪氨酸(Y)。优选地,一个或多个带负电氨基酸被一个或多个带正电氨基酸取代。适当的带负电氨基酸包含但不限于天冬氨酸(D)和谷氨酸(E)。

[0392] 优选的引入包含但不限于:使用K取代E、使用R取代M、使用H取代M、使用K取代M、使

用R取代D、使用H取代D、使用K取代D、使用R取代E、使用H取代E、使用R取代N、使用R取代T和使用R取代G。最优地,使用K取代E。

[0393] 可以引入或取代任何数目的带正电氨基酸。例如,可以引入或取代1个、2个、5个、10个、15个、20个、25个、30个或更多个带正电氨基酸。

[0394] 更优选地,通过中和一个或多个负电荷来增加净正电荷。可以通过使用一个或多个不带电氨基酸、非极性氨基酸和/或芳香族氨基酸取代一个或多个带负电氨基酸来中和所述一个或多个负电荷。移除负电荷会增加净正电荷。不带电氨基酸、非极性氨基酸和/或芳香族氨基酸可以是天然存在的或非天然存在的。它们可以是合成的或修饰的。以上讨论了适当的不带电氨基酸、非极性氨基酸和芳香族氨基酸。优选的取代包括但不限于:使用Q取代E、使用S取代E、使用A取代E、使用Q取代D、使用N取代E、使用N取代D、使用G取代D以及使用S取代D。

[0395] 可以取代任何数目的不带电氨基酸、非极性氨基酸和/或芳香族氨基酸和其任何组合。例如,可以取代1个、2个、5个、10个、15个、20个、25个或30个或更多个不带电氨基酸、非极性氨基酸和/或芳香族氨基酸。带负电氨基酸可以使用以下氨基酸取代:(1) 不带电氨基酸;(2) 非极性氨基酸;(3) 芳香族氨基酸;(4) 不带电氨基酸和非极性氨基酸;(5) 不带电氨基酸和芳香族氨基酸;以及(5) 非极性氨基酸和芳香族氨基酸;或(6) 不带电氨基酸、非极性氨基酸和芳香族氨基酸。

[0396] 可以通过在一个或多个带负电氨基酸附近如在1个、2个、3个或4个氨基酸内或在与一个或多个带负电氨基酸相邻的地方引入一个或多个带正电氨基酸来中和所述一个或多个负电荷。以上讨论了带正电氨基酸和带负电氨基酸的实例。带正电氨基酸可以以上讨论的任何方式引入,例如通过取代。

[0397] 优选地,通过引入一个或多个带负电氨基酸和/或中和一个或多个正电荷来减少净正电荷。根据以上参考增加净正电荷进行的讨论,可以实现这一点的方式将变得清楚。以上参考增加净正电荷所讨论的所有实施例同样适用于减少净正电荷,除了电荷以相反方式改变之外。具体地说,优选地通过使用一个或多个不带电氨基酸、非极性氨基酸和/或芳香族氨基酸取代一个或多个带正电氨基酸和/或通过在一个或多个带正电氨基酸附近如在其中的1个、2个、3个或4个氨基酸内或在与一个或多个带正电氨基酸相邻的地方引入一个或多个带负电氨基酸来中和所述一个或多个正电荷。

[0398] 优选地增加或减少净负电荷。以上参考增加或减少净正电荷所讨论的所有上述实施例同样适用于分别减少或增加净负电荷。

[0399] 对于(c),可以以任何方式改变单体进行氢结合的能力。丝氨酸(S)、苏氨酸(T)、天冬酰胺(N)、谷氨酰胺(Q)、酪氨酸(Y)或组氨酸(H)的引入提高了单体的氢结合能力。所述一个或多个修饰优选地是引入S、T、N、Q、Y和H中的一个或多个。可以引入S、T、N、Q、Y和H的任何组合。可以通过添加引入S、T、N、Q、Y和H中的所述一个或多个。优选地通过取代引入S、T、N、Q、Y和H中的所述一个或多个。以下更详细地讨论用于引入这种残基的适当位置。

[0400] 丝氨酸(S)、苏氨酸(T)、天冬酰胺(N)、谷氨酰胺(Q)、酪氨酸(Y)或组氨酸(H)的移除降低了单体的氢结合能力。所述一个或多个修饰优选地是移除S、T、N、Q、Y和H中的一个或多个。可以移除S、T、N、Q、Y和H的任何组合。可以通过缺失移除S、T、N、Q、Y和H中的所述一个或多个。优选地通过使用如丙氨酸(A)、缬氨酸(V)、异亮氨酸(I)和亮氨酸(L)等氢结合不太

好的其它氨基酸进行取代来移除S、T、N、Q、Y和H中的所述一个或多个。

[0401] 对于(d),引入如苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)等芳香族残基还增加了单体中的 π 堆积。移除如苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)等芳香族残基还减少了单体中的 π 堆积。可以如上文参考(a)所讨论的那样引入或移除这种氨基酸。

[0402] 对于(e),可以根据本发明进行改变单体的结构的一个或多个修饰。例如,可以移除、缩短或扩增一个或多个环区域。这通常有助于多核苷酸进入或离开孔。所述一个或多个环区域可以是孔的顺侧、孔的反侧或在孔的两侧上。可替代地,可以扩增或缺失孔的氨基末端和/或羧基末端的一个或多个区域。这通常改变孔的大小和/或电荷。

[0403] 根据以上讨论将清楚的是,引入某些氨基酸将增强单体通过多于一种机制与多核苷酸相互作用的能力。例如,使用H取代E将不仅根据(b)增加净正电荷(通过中和负电荷),而且将根据(c)氢键提高单体进行氢结合的能力。

[0404] 所述变异体优选地包括SEQ ID NO:2的以下位置中的一个或多个处的取代:M44、N46、N48、E50、R52、H58、D68、F70、E71、S74、E76、S78、Y79、S80、H81、S82、E84、E85、S86、Q87、S89、M90、E92、E94、E97、E102、H103、T104、T106、R115、Q117、N119、D121和D126。所述变异体优选地包括那些位置中的1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个或34个处的取代。所述变异体优选地包括SEQ ID NO:2的以下位置中的一个或多个处的取代:D68、E71、S74、E76、S78、S80、S82、E84、E85、S86、Q87、S89、E92、E102、T104、T106、R115、Q117、N119和D121。所述变异体优选地包括那些位置中的1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个或20个处的取代。

[0405] 所述变异体优选地包括SEQ ID NO:2的以下位置中的一个或多个处的取代:(a) E84、E85、E92、E97和D126;(b) E85、E97和D126或者(c) E84和E92。取代到变异体中的氨基酸可以是其天然存在的衍生物或非天然存在的衍生物。取代到变异体中的氨基酸可以是D-氨基酸。以上列出的每个位置可以使用以下各项取代:天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)、精氨酸(R)、甘氨酸(G)、酪氨酸(Y)、天冬氨酸(D)、亮氨酸(L)、赖氨酸(K)或丙氨酸(A)。

[0406] 所述变异体优选地包括SEQ ID NO:2的以下突变中的至少一个:

[0407] (a) 位置44处的丝氨酸(S);

[0408] (b) 位置46处的丝氨酸(S);

[0409] (c) 位置48处的丝氨酸(S);

[0410] (d) 位置52处的丝氨酸(S);

[0411] (e) 位置58处的丝氨酸(S);

[0412] (f) 位置68处的丝氨酸(S);

[0413] (g) 位置70处的丝氨酸(S);

[0414] (h) 位置71处的丝氨酸(S);

[0415] (i) 位置76处的丝氨酸(S);

[0416] (j) 位置79处的丝氨酸(S);

[0417] (k) 位置81处的丝氨酸(S);

- [0418] (l) 位置84处的丝氨酸(S)、天冬氨酸(D)或谷氨酰胺(Q)；
- [0419] (m) 位置85处的丝氨酸(S)或赖氨酸(K)；
- [0420] (n) 位置87处的丝氨酸(S)；
- [0421] (o) 位置90处的丝氨酸(S)；
- [0422] (p) 位置92处的天冬酰胺(N)或谷氨酰胺(Q)；
- [0423] (q) 位置94处的丝氨酸(S)或天冬酰胺(N)；
- [0424] (r) 位置97处的丝氨酸(S)或天冬酰胺(N)；
- [0425] (s) 位置102处的丝氨酸(S)；
- [0426] (t) 位置103处的丝氨酸(S)；
- [0427] (u) 位置121处的天冬酰胺(N)或丝氨酸(S)；
- [0428] (v) 位置50处的丝氨酸(S)；
- [0429] (w) 位置94处的天冬酰胺(N)或丝氨酸(S)；
- [0430] (x) 位置97处的天冬酰胺(N)或丝氨酸(S)；
- [0431] (y) 位置121处的丝氨酸(S)或天冬酰胺(N)；
- [0432] (z) 位置126处的天冬酰胺(N)或谷氨酰胺(Q)；以及
- [0433] (aa) 位置128处的丝氨酸(S)或天冬酰胺(N)。
- [0434] 所述变异体可以包含任何数目的突变(a)到(aa)，如所述突变中的1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26或27个。以下讨论了优选的突变组合。引入到变异体中的氨基酸可以是其天然存在的衍生物或非天然存在的衍生物。引入到变异体中的氨基酸可以是D-氨基酸。
- [0435] 所述变异体优选地包括SEQ ID NO:2的以下突变中的至少一个：
- [0436] (a) 位置68处的丝氨酸(S)；
- [0437] (b) 位置71处的丝氨酸(S)；
- [0438] (c) 位置76处的丝氨酸(S)；
- [0439] (d) 位置84处的天冬氨酸(D)或谷氨酰胺(Q)；
- [0440] (e) 位置85处的赖氨酸(K)；
- [0441] (f) 位置92处的天冬酰胺(N)或谷氨酰胺(Q)；
- [0442] (g) 位置102处的丝氨酸(S)；
- [0443] (h) 位置121处的天冬酰胺(N)或丝氨酸(S)；
- [0444] (i) 位置50处的丝氨酸(S)；
- [0445] (j) 位置94处的天冬酰胺(N)或丝氨酸(S)；
- [0446] (k) 位置97处的天冬酰胺(N)或丝氨酸(S)；以及
- [0447] (l) 位置126处的天冬酰胺(N)或谷氨酰胺(Q)。
- [0448] 所述变异体可包含任何数目的突变(a)到(l)，如所述突变中的1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个或12个。以下讨论了优选的突变组合。引入到变异体中的氨基酸可以是其天然存在的衍生物或非天然存在的衍生物。引入到变异体中的氨基酸可以是D-氨基酸。
- [0449] 所述变异体可以包含SEQ ID NO:2的约位置44到约位置126的区域之外的一个或

多个额外修饰,所述额外修饰与以上讨论的区域中的修饰组合提高多核苷酸捕获和/或提高多核苷酸识别或区分。适当的修饰包含但不限于D35、E128、E135、E134和E167中的一个或多个处的取代。具体地说,通过在位置128、135、134和167中的一个或多个处取代E来移除负电荷提高了多核苷酸捕获。可以以上讨论的方式中的任何方式取代这些位置中的一个或多个处的E。优选地,如以上所讨论的那样取代全部E128、E135、E134和E167。优选地使用A取代E。换言之,所述变异体优选地包括E128A、E135A、E134A和E167A中的一个或多个或全部。另一个优选的取代是D35Q。

[0450] 在优选的实施例中,所述变异体包括SEQ ID NO:2中的以下取代:

[0451] i. E84D和E85K中的一个或多个,如两个;

[0452] ii. E84Q、E85K、E92Q、E97S、D126G和E167A中的一个或多个,如2个、3个、4个、5个或6个;

[0453] iii. E92N、E94N、E97N、D121N和D126N中的一个或多个,如2个、3个、4个或5个;

[0454] iv. E92N、E94N、E97N、D121N、D126N和E128N中的一个或多个,如2个、3个、4个、5个或6个;

[0455] v. E76S、E84Q、E85K、E92Q、E97S、D126G和E167A中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0456] vi. E84Q、E85K、E92Q、E97S、D126G、E167A和E50S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0457] vii. E84Q、E85K、E92Q、E97S、D126G、E167A和E71S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0458] viii. E84Q、E85K、E92Q、E97S、D126G、E167A和E94S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0459] ix. E84Q、E85K、E92Q、E97S、D126G、E167A和E102S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0460] x. E84Q、E85K、E92Q、E97S、D126G、E167A和E128S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0461] xi. E84Q、E85K、E92Q、E97S、D126G、E167A和E135S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0462] xii. E84Q、E85K、E92Q、E97S、D126G、E167A和D68S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0463] xiii. E84Q、E85K、E92Q、E97S、D126G、E167A和D121S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0464] xiv. E84Q、E85K、E92Q、E97S、D126G、E167A和D134S中的一个或多个,如2个、3个、4个、5个、6个或7个;

[0465] xv. E84D、E85K和E92Q中的一个或多个,如2个或3个;

[0466] xvi. E84Q、E85K、E92Q、E97S、D126G和E135S中的一个或多个,如1个、2个、3个、4个、5个或6个;

[0467] xvii. E85K、E92Q、E94S、E97S和D126G中的一个或多个,如1个、2个、3个、4个或5个;

[0468] xviii. E76S、E85K、E92Q、E97S和D126G中的一个或多个,如1个、2个、3个、4个或5

个；

[0469] xix.E71S、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个或5个；

[0470] xx.D68S、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个或5个；

[0471] xxi.E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个或4个；

[0472] xxii.E84Q、E85K、E92Q、E97S、H103S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0473] xxiii.E84Q、E85K、M90S、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0474] xxiv.E84Q、Q87S、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0475] xxv.E84Q、E85S、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个或5个；

[0476] xxvi.E84S、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个或5个；

[0477] xxvii.H81S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0478] xxviii.Y79S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0479] xxix.F70S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0480] xxx.H58S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0481] xxxi.R52S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0482] xxxii.N48S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0483] xxxiii.N46S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0484] xxxiv.M44S、E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0485] xxxv.E92Q和E97S中的一个或多个，如两个；

[0486] xxxvi.E84Q、E85K、E92Q和E97S中的一个或多个，如1个、2个、3个或4个；

[0487] xxxvii.E84Q和E85K中的一个或多个，如两个；

[0488] xxxviii.E84Q、E85K和D126G中的一个或多个，如1个、2个或3个；

[0489] xxxix.E84Q、E85K、D126G和E167A中的一个或多个，如1个、2个、3个或4个；

[0490] xl.E92Q、E97S和D126G中的一个或多个，如1个、2个或3个；

[0491] xli.E84Q、E85K、E92Q、E97S和D126G中的一个或多个，如1个、2个、3个、4个或5个；

[0492] xlii.E84Q、E85K、E92Q、E97S和E167A中的一个或多个，如1个、2个、3个、4个或5个；

[0493] xliii.E84Q、E85K、E92Q、D126G和E167A中的一个或多个，如1个、2个、3个、4个或5个；

[0494] xliv.E84Q、E85K、E97S、D126G和E167A中的一个或多个，如1个、2个、3个、4个或5

个；

[0495] xlv.E84Q、E92Q、E97S、D126G和E167A中的一个或多个，如1个、2个、3个、4个或5个；

[0496] xlvi.E85K、E92Q、E97S、D126G和E167A中的一个或多个，如1个、2个、3个、4个或5个；

[0497] xlvii.E84D、E85K和E92Q中的一个或多个，如1个、2个或3个；

[0498] xlviii.E84Q、E85K、E92Q、E97S、D126G、E167A和D121S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0499] xlix.E84Q、E85K、E92Q、E97S、D126G、E167A和D68S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0500] l.E84Q、E85K、E92Q、E97S、D126G、E167A和E135S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0501] li.E84Q、E85K、E92Q、E97S、D126G、E167A和E128S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0502] lii.E84Q、E85K、E92Q、E97S、D126G、E167A和E102S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0503] liii.E84Q、E85K、E92Q、E97S、D126G、E167A和E94S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0504] liv.E84Q、E85K、E92Q、E97S、D126G、E167A和E71S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0505] lv.E84Q、E85K、E92Q、E97S、D126G、E167A和E50S中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0506] lvi.E76S、E84Q、E85K、E92Q、E97S、D126G和E167A中的一个或多个，如1个、2个、3个、4个、5个、6个或7个；

[0507] lvii.E92N、E94N、E97N、D121N、D126N和E128N中的一个或多个，如1个、2个、3个、4个、5个或6个；

[0508] lviii.E92N、E94N、E97N、D121N和D126N中的一个或多个，如1个、2个、3个、4个或5个；或

[0509] lix.E84Q、E85K、E92Q、E97S、D126G和E167A中的一个或多个，如1个、2个、3个、4个、5个或6个

[0510] 在上文中，第一个字母是指SEQ ID NO:2中被取代的氨基酸，数字是在SEQ ID NO:2中的位置，并且第二个字母是指将用于取代第一个的氨基酸。因此，E84D是指使用天冬氨酸(D)取代位置84处的谷氨酸(E)。

[0511] 变体可以包含i到lix中任何一个中的任何数目的取代，如1个、2个、3个、4个、5个、6个或7个。变体优选地包含上述i到lix中的任何一个中所示的所有取代。

[0512] 在优选的实施例中，变体包括上述i到xv中的任何一个中的取代。变体可以包含i到xv中任何一个中的任何数目的取代，如1个、2个、3个、4个、5个、6个或7个。变体优选地包含上述i到xv中的任何一个中所示的所有取代。

[0513] 如果所述一个或多个修饰旨在提高单体识别或区分多核苷酸的能力，则优选地，除了以上讨论的如E84Q、E85K、E92Q、E97S、D126G和E167A等提高多核苷酸捕获的修饰之外，

还可以进行所述一个或多个修饰。

[0514] 对所识别的区域进行的所述一个或多个修饰可以涉及使用存在于胞溶素的同源物或旁系同源物中的一个或多个对应位置处的氨基酸取代所述区域中的一个或多个氨基酸。在SEQ ID NO:14到17中示出了胞溶素的同源物的四个实例。这种取代的优点是它们可能导致形成孔的突变单体,因为同系物单体也形成孔。例如,可以在SEQ ID NO:2中的在SEQ ID NO:2与SEQ ID NO:14到17中的任何一个之间是不同的位置中的任何一个或多个处进行突变。这种突变可以是使用来自SEQ ID NO:14到17、优选地SEQ ID NO:14到16中的任何一个中的对应位置的氨基酸取代SEQ ID NO:2中的氨基酸。可替代地,在这些位置中的任何位置处的突变可以是使用任何氨基酸进行的取代,或者可以是缺失或插入突变,如1个到30个氨基酸如2个到20个、3个到10个或4个到8个氨基酸的取代、缺失或插入。除了本文公开的突变和现有技术中、例如在2013/153359中公开的突变之外,SEQ ID NO:2与全部SEQ ID NO:14到17、更优选地全部SEQ ID NO:14到16之间保守或一致的氨基酸优选地是保守的或存在于本发明的变异体中。然而,可以在SEQ ID NO:2与所有SEQ ID NO:14到17或更优选地SEQ ID NO:14到16之间是保守或一致的这些位置中的任何一个或多个处进行保守突变。

[0515] 本发明提供了一种胞溶素突变单体,其包括本文中被描述为在胞溶素单体的结构中 SEQ ID NO:2 中的特定位置对应的位置处被取代到 SEQ ID NO:2 的所述特定位置中的氨基酸中的任何一个或多个。可以通过所属领域中的标准技术来确定对应位置。例如,上文提到的 PILEUP 和 BLAST 算法可以用于将胞溶素单体的序列与 SEQ ID NO:2 进行比对并且因此识别对应残基。

[0516] 突变单体通常保留形成与野生型胞溶素单体相同的3D结构的能力,如与具有SEQ ID NO:2的序列的胞溶素单体相同的3D结构。胞溶素单体的3D结构在所属领域中是已知的并且在例如De Colbis等人,《结构(Structure)》,2012年,第20卷,第1498到1507页中得以公开。突变单体通常保留与其它胞溶素单体形成同源寡聚孔和/或异源寡聚孔的能力。当存在于孔中时,突变单体通常保留重折叠以形成与野生型胞溶素单体相同的3D结构的能力。在本文中的图7中示出了胞溶素孔中的胞溶素单体的3D结构。除了本文所述的突变之外,可以在野生型胞溶素序列中进行任何数目的突变,如2个到100个、3个到80个、4个到70个、5个到60个、10个到50个或20个到40个,条件是胞溶素突变单体保留了本发明的突变对其赋予提高性质中的一种或多种。

[0517] 通常,当胞溶素单体与其它一致的突变单体组装或与不同的胞溶素突变单体组装形成孔时,所述胞溶素单体将保留向胞溶素孔的桶贡献两个 β 折叠的能力。

[0518] 所述变异体进一步优选地包括E84Q/E85K/E92Q/E97S/D126G中的一个或多个或者在适当的情况下全部E84Q/E85K/E92Q/E97S/D126G中。“在适当的情况下”是指这些位置是仍存在于突变单体中还是未被不同的氨基酸修饰。

[0519] 除上文所讨论的特定突变之外,变异体可以包含其它突变。这些突变不一定增强单体与多核苷酸相互作用的能力。突变可以促进例如表达和/或纯化。在SEQ ID NO:2的氨基酸序列的整个长度内,基于氨基酸相似性或同一性,变异体将优选地与所述序列至少50%同源。更优选地,基于氨基酸相似性或同一性,变异体可以在整个序列内与SEQ ID NO:2的氨基酸序列至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,并且更优选地至少95%、97%或99%同源。在100个或更多个、例如125个、150个、

175个或200个或更多个连续氨基酸的延伸段内,可能存在至少80%、例如至少85%、90%或95%的氨基酸相似性或同一性(“硬同源性(hard homology)”)。

[0520] 所属领域中的标准方法可以用于确定同源性。例如,UWCGC程序包提供了BESTFIT程序,其可以用于计算同源性,例如根据其默认设置使用(Devereux等人(1984年)《核酸研究(Nucleic Acids Research)》第12卷,第387到395页)。PILEUP和BLAST算法可以用于计算同源性或对序列进行排序(如识别等同残基或对应序列(通常根据其默认设置),例如如Altschul S.F.(1993年),《分子进化杂志(J Mol Evol)》,第36卷,第290到300页;Altschul,S.F等人(1990年)《分子生物学杂志(J Mol Biol)》,第215卷,第403到410页中所描述。用于执行BLAST分析的软件可通过美国国家生物技术信息中心(National Center for Biotechnology Information)(<http://www.ncbi.nlm.nih.gov/>)公开获得。相似性可以使用成对同一性或通过应用如BLOSUM62等评分矩阵并转换为等效同一性来测量。由于它们表示功能变化而非演化变化,所以在确定同源性时将掩盖故意突变的位置。可以通过在蛋白序列的综合数据库上使用例如PSIBLAST来应用位置特异性评分矩阵从而更灵敏地确定相似性。可以使用反映氨基酸化学-物理性质而不是演化时间尺度(例如,电荷)内的取代频率的不同评分矩阵。

[0521] 可以对SEQ ID NO:2的氨基酸序列进行除了上文所讨论的取代之外的氨基酸取代,例如高达1个、2个、3个、4个、5个、10个、20个或30个取代。保守取代使用具有类似化学结构、类似化学性质或类似侧链体积的其它氨基酸来代替氨基酸。所引入的氨基酸可以具有与其所代替的氨基酸类似的极性、亲水性、疏水性、碱性、酸性、电中性或电荷。可替代地,保守取代可以引入另一芳香族或脂肪族氨基酸代替预先存在的芳香族或脂肪族氨基酸。保守氨基酸改变在所属领域中是众所周知的,且可以根据如下表3中限定的20种主要氨基酸的特性来进行选择。在氨基酸具有类似极性的情况下,还可以参考表4中的氨基酸侧链的亲水性尺度来确定这一点。

[0522] 表3-氨基酸的化学性质

Ala	脂肪族的、疏水性的、中性的	Met	疏水性的、中性的
Cys	极性的、疏水性的、中性的	Asn	极性的、亲水性的、中性的
Asp	极性的、亲水性的、带电的(-)	Pro	疏水性的、中性的
Glu	极性的、亲水性的、带电的(-)	Gln	极性的、亲水性的、中性的
Phe	芳香族的、疏水性的、中性的	Arg	极性的、亲水性的、带电的(+)
[0523] Gly	脂肪族的、中性的	Ser	极性的、亲水性的、中性的
His	芳香族的、极性的、亲水性的、带电的(+)	Thr	极性的、亲水性的、中性的
Ile	脂肪族的、疏水性的、中性的	Val	脂肪族的、疏水性的、中性的
Lys	极性的、亲水性的、带电的(+)	Trp	芳香族的、疏水性的、中性的
Leu	脂肪族的、疏水性的、中性的	Tyr	芳香族的、极性的、疏水性的

[0524] 表4-亲水性尺度

侧链	亲水性
[0525] Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8

	Cys	2.5
	Met	1.9
	Ala	1.8
	Gly	-0.4
	Thr	-0.7
	Ser	-0.8
	Trp	-0.9
[0526]	Tyr	-1.3
	Pro	-1.6
	His	-3.2
	Glu	-3.5
	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
	Arg	-4.5

[0527] 变体可以包括在上文指定的区域外部的一个或多个取代,其中氨基酸被代替为在胞溶素的同源物和旁系同源物中的一个或多个对应位置处的氨基酸。在SEQ ID NO:14到17中示出了胞溶素的同源物的四个实例。

[0528] 另外,可以从上文所述的变体缺失SEQ ID NO:2的氨基酸序列的一个或多个氨基酸残基。可缺失高达1个、2个、3个、4个、5个、10个、20个或30个或更多个残基。

[0529] 变体可以包含SEQ ID NO:2的片段。这种片段保留成孔活性。可以如上所述那样测定这一点。片段的长度可以为至少50个、100个、150个、200个或250个氨基酸。这种片段可以用于产生本发明的孔。由于可以通过根据本发明的一个或多个缺失修饰SEQ ID NO:2的约位置44到约位置126的区域,所以片段不必含有整个区域。因此,本发明设想了长度比未修饰区域的长度更短的片段。片段优选地包括SEQ ID NO:2的成孔结构域。片段更优选地包括根据本发明进行修饰的从SEQ ID NO:2的约位置44到约位置126的区域。

[0530] 可替代地或另外地,可以向上文所描述的变体添加一个或多个氨基酸。可以在SEQ ID NO:2、包含其片段的变体的氨基酸序列的氨基端或羧基端处提供延长物。延长物的长度可以非常短,例如1到10个氨基酸。可替代地,延长物可以较长,例如高达50或100个氨基酸。可以将载体蛋白与根据本发明的氨基酸序列融合。下文更详细地讨论了其它融合蛋白。

[0531] 如上文所讨论的,变体是具有不同于SEQ ID NO:2的氨基酸序列的氨基酸序列的且保留其形成孔的能力的多肽。变体通常含有SEQ ID NO:2的负责孔形成的区域,即约位置44到约位置126,并且此区域根据本发明如上文讨论的那样进行修饰。其可以含有此区域的片段,如上所讨论的。除了本发明的修饰之外,SEQ ID NO:2的变体可以包含一个或多个额外修饰,如取代、添加或缺失。这些修饰优选地位于变体中与SEQ ID NO:2的约位置1到约位置43和约位置127到约位置297相对应的延伸段(即,根据本发明修饰的区域外部)中。

[0532] 可以例如通过添加组氨酸残基(his标签)、天冬氨酸残基(asp标签)、抗生蛋白链霉素标签或flag标签或者通过添加用于促进突变单体从其中多肽并不天然地含有这种序列的细胞分泌的信号序列来对突变单体进行修饰以辅助对其进行识别或纯化。引入基因标签的替代性方案是将标签化学反应到孔上的原生或工程化位置上。这种操作的实例将是使

凝胶迁移试剂与孔外部上的工程化半胱氨酸反应。这已经被显示为一种用于分离溶血素同源寡聚物的方法(《生物化学(Chem Biol)》,1997年7月,第4卷,第7期,第497到505页)。

[0533] 可以使用揭露标签来标记突变单体。揭露标签可以是允许孔被检测到的任何适当标签。适当的标记包含但不限于荧光分子;放射性同位素,例如,¹²⁵I、³⁵S、酶、抗体、抗原、多核苷酸、聚乙二醇(PEG)、肽和配体,如生物素。

[0534] 还可以使用D-氨基酸来产生突变单体。例如,突变单体可以包括L-氨基酸与D-氨基酸的混合物。这在用于产生这种蛋白或肽的所属领域中是常规的。

[0535] 突变单体含有用于促进与多核苷酸的相互作用的一种或多种特异性修饰。突变单体还可以含有其它非特异性修饰,只要所述修饰不干扰孔形成即可。多种非特异性侧链修饰在所属领域中是已知的并且可以对突变单体的侧链进行所述多种非特异性侧链修饰。这种修饰包含例如通过与醛反应、随后使用NaBH₄进行还原、使用甲基乙酰亚氨酸进行脎化或使用乙酸酐进行酰化来对氨基酸进行的还原烷基化。

[0536] 可以使用所属领域中已知的标准方法产生突变单体。可以通过合成方式或通过重组手段来制备单体。例如,可以通过体外转译和转录(IVTT)来合成单体。在国际申请号PCT/GB09/001690(公开为WO 2010/004273)、PCT/GB09/001679(公开为WO 2010/004265)或PCT/GB10/000133(公开为WO 2010/086603)中讨论了用于产生孔单体的适当方法。以下讨论了用于将孔插入到膜中的方法。

[0537] 可以使用所属领域中的标准方法来衍生或复制对突变单体进行编码的多核苷酸序列。下文更详细地讨论了这种序列。可以使用所属领域中的标准技术在细菌宿主细胞中表达对突变单体进行编码的多核苷酸序列。可以在细胞中通过多肽的原位表达从重组表达载体产生突变单体。表达载体任选地携带用于控制多肽的表达的诱导型启动子。

[0538] 可以在通过任何蛋白质液相色谱系统从孔产生生物体进行纯化之后或在重组表达之后大规模产生突变单体,如下文所述。典型的蛋白质液相色谱系统包含FPLC、AKTA系统、Bio-Cad系统、Bio-Rad生物系统和Gilson HPLC系统。然后将突变单体插入天然存在的或人造的膜中以供根据本发明使用。以下讨论了用于将孔插入到膜中的方法。

[0539] 在一些实施例中,对突变单体进行化学修饰。可以以任何方式且在任何位点处对突变单体进行化学修饰。优选地通过将分子附接到一个或多个半胱氨酸(半胱氨酸连接)、将分子附接到一个或多个赖氨酸、将分子附接到一个或多个非天然氨基酸、表位的酶修饰或末端的修饰对突变单体进行化学修饰。用于执行这种修饰的适当方法在所属领域中是众所周知的。适当的非天然氨基酸包含但不限于4-叠氮基-L-苯丙氨酸(Faz)和Liu C.C.和Schultz P.G.,《生物化学年评(Annu.Rev.Biochem.)》,2010年,第79卷,第413到444页的图1中编号为1到71的氨基酸中的任何氨基酸。可以通过附接任何分子对突变单体进行化学修饰。例如,可以通过附接聚乙二醇(PEG)、如DNA等核酸、染料、荧光团或发色团对突变单体进行化学修饰。

[0540] 在一些实施例中,使用促进包括单体的孔与靶分析物、靶核苷酸或靶多核苷酸序列之间的相互作用的分子衔接子对突变单体进行化学修饰。衔接子的存在改进了孔和核苷酸或多核苷酸的主客体化学,并且由此提高了由突变单体形成的孔的测序能力。主客体化学的原理在所属领域中是众所周知的。衔接子对孔的物理或化学性质有影响,所述影响提高其与核苷酸或多核苷酸序列的相互作用。衔接子可改变孔的桶或通道的电荷,或与核苷

酸或多核苷酸特异性地相互作用或结合到核苷酸或多核苷酸,由此促进其与孔的相互作用。

[0541] 分子衔接子优选地是环状分子例如环糊精、能够杂化的物种、DNA结合剂或嵌入剂、肽或肽类似物、合成聚合物、芳香族平面分子、带正电的小分子或能够进行氢结合的小分子。

[0542] 衔接子可以是环状的。环状衔接子优选地具有与孔相同的对称性。

[0543] 衔接子通常通过主客体化学与分析物、核苷酸或多核苷酸相互作用。衔接子通常能够与核苷酸或多核苷酸相互作用。衔接子包括能够与核苷酸或多核苷酸相互作用的一个或多个化学基团。所述一个或多个化学基团优选地通过非共价相互作用、如疏水性相互作用、氢结合、范德华力、 π -阳离子相互作用和/或静电力与核苷酸或多核苷酸相互作用。能够与核苷酸或多核苷酸相互作用的所述一个或多个化学基团优选地带正电。能够与核苷酸或多核苷酸相互作用的所述一个或多个化学基团更优选地包括氨基。氨基可以附接到伯、仲或叔碳原子。衔接子甚至更优选地包括氨基环,如由6个、7个、8个或9个氨基构成的环。衔接子最优选地包括由6个或9个氨基构成的环。质子化氨基环可以与核苷酸或多核苷酸中的带负电磷酸基相互作用。

[0544] 可以通过衔接子与包括突变单体的孔之间的主客体化学来促进衔接子在孔内的正确定位。衔接子优选地包括能够与孔中的一个或多个氨基酸相互作用的一个或多个化学基团。衔接子更优选地包括能够通过非共价相互作用与孔中的一个或多个氨基酸相互作用的一个或多个化学基团,所述非共价相互作用如疏水性相互作用、氢结合、范德华力、 π -阳离子相互作用和/或静电力。能够与孔中的一个或多个氨基酸相互作用的化学基团通常是羟基或胺。羟基可以附接到伯、仲或叔碳原子。羟基可以与孔中的不带电氨基酸形成氢键。可以使用促进孔与核苷酸或多核苷酸之间的相互作用的任何衔接子。

[0545] 适当的衔接子包含但不限于环糊精、环肽和葫芦脲。衔接子优选地是环糊精或其衍生物。环糊精或其衍生物可以是在Eliseev, A.V. 和Schneider, H-J. (1994年),《美国化学会志(J. Am. Chem. Soc.)》,第116卷,第6081到6088页中公开的环糊精或其衍生物中的任何环糊精或其衍生物。衔接子更优选地是七-6-氨基- β -环糊精(am_7 - β CD)、6-单脱氧-6-单氨基- β -环糊精(am_1 - β CD)或七-(6-脱氧-6-胍基)-环糊精(gu_7 - β CD)。 gu_7 - β CD中的胍基具有比 am_7 - β CD中的伯胺高得多的pKa,且因此其带更多正电。这种 gu_7 - β CD衔接子可以用于增加核苷酸在孔中的停留时间、增加所测量的剩余电流的准确度、以及增加在高温或低数据获取速率下的碱基检测速率。

[0546] 如果如在下文中更详细地讨论的那样使用3-(2-吡啶二硫代)丙酸丁二酰亚胺酯(SPDP)交联剂,则衔接子优选地是七(6-脱氧-6-胺基)-6-N-单(2-吡啶基)二硫代丙酰基- β -环糊精(am_6amPDP_1 - β CD)。

[0547] 更多适当的衔接子包含 γ -环糊精,其包括8个糖单元(并且因此具有八重对称性)。 γ -环糊精可以含有接头分子,或可进行被修饰成包括所有或更多个在上文所讨论的 β -环糊精实例中使用的经修饰糖单元。

[0548] 分子衔接子优选地共价连接到突变单体。可以使用所属领域中已知的任何方法将衔接子共价连接到孔。通常通过化学连接附接衔接子。如果通过半胱氨酸连接附接分子衔接子,则已经优选地通过取代将所述一个或多个半胱氨酸引入到突变体中。本发明的突变

单体当然可以在位置272和283中的一个或两个处包括半胱氨酸残基。可以通过将分子衔接子附接到这些半胱氨酸中的一个或两个对突变单体进行化学修饰。可替代地,可以通过将分子附接到在其它位置处引入的一个或多个半胱氨酸或如FAz等非天然氨基酸对突变单体进行化学修饰。

[0549] 可以通过修饰相邻残基来增强半胱氨酸残基的反应性。例如,侧接精氨酸、组氨酸或赖氨酸残基的碱性基团将半胱氨酸巯基的pKa改变更具反应性的S⁻基的pKa。可以通过如dTNB等巯基保护基来保护半胱氨酸残基的反应性。在附接接头之前,这些可以与突变单体的一个或多个半胱氨酸残基反应。可以将所述分子直接附接到突变单体。优选地,使用如化学交联剂或肽接头等接头将所述分子附接到突变单体。

[0550] 适当的化学交联剂在所属领域中是众所周知的。优选的交联剂包含3-(吡啶-2-基二磺酰基)丙酸2,5-二氧化吡咯烷-1-基酯、4-(吡啶-2-基二磺酰基)丁酸2,5-二氧化吡咯烷-1-基酯以及8-(吡啶-2-基二磺酰基)辛酸2,5-二氧化吡咯烷-1-基酯。最优选的交联剂是3-(2-吡啶二硫代)丙酸丁二酰亚胺酯(SPDP)。通常,在分子/交联剂复合体共价连接到突变单体之前,分子共价连接到双功能交联剂,但也有可能在双功能交联剂/单体复合体附接到分子之前将双功能交联剂共价连接到单体。

[0551] 优选地,接头对二硫苏糖醇(DTT)具有抗性。适当的接头包含但不限于基于碘乙酰胺和基于马来酰亚胺的接头。

[0552] 在其它实施例中,单体可以附接到多核苷酸结合蛋白。这形成可以在本发明的方法中使用的模块化测序系统。下文讨论了多核苷酸结合蛋白。

[0553] 多核苷酸结合蛋白可以共价连接到突变单体。可以使用所属领域中已知的任何方法将蛋白质共价连接到孔。单体和蛋白质可以是化学融合的或基因融合的。如果从单个多核苷酸序列表达整个构建体,则单体和蛋白质是基因融合的。在国际申请号PCT/GB09/001679(公开为WO 2010/004265)中讨论了单体与多核苷酸结合蛋白的基因融合。

[0554] 如果通过半胱氨酸连接附接多核苷酸结合蛋白,则已经优选地通过取代将所述一个或多个半胱氨酸引入到突变体中。这种取代通常在环区域中进行,所述环区域在同源物中具有低保守性,表明可以耐受突变或插入。因此,其适合于附接多核苷酸结合蛋白。这种取代通常在SEQ ID NO:2的残基1到43和127到297中进行。可以通过如上文所描述的修饰来增强半胱氨酸残基的反应性。

[0555] 可以将多核苷酸结合蛋白直接附接到突变单体或通过一个或多个接头将其附接。可以使用在国际申请号PCT/GB10/000132(公开为WO 2010/086602)中描述的杂化接头将多核苷酸结合蛋白附接到突变单体。可替代地,可以使用肽接头。肽接头是氨基酸序列。肽接头的长度、柔性和亲水性通常被设计为使得其不干扰单体和分子的功能。优选的柔性肽接头是2个到20个,如4个、6个、8个、10个或16个丝氨酸和/或甘氨酸的延伸段。更优选的柔性接头包含(SG)₁、(SG)₂、(SG)₃、(SG)₄、(SG)₅和(SG)₈,其中S是丝氨酸且G是甘氨酸。优选的刚性接头是2个到30个,如4个、6个、8个、16个或24个脯氨酸的延伸段。更优选的刚性接头包含(P)₁₂,其中P是脯氨酸。

[0556] 可以使用分子接头和多核苷酸结合蛋白对突变单体进行化学修饰。

[0557] 制备突变胞溶素单体

[0558] 本发明还提供了一种提高包括SEQ ID NO:2中所示序列的胞溶素单体表征多核苷

酸的能力的方法。所述方法包括在SEQ ID NO:2中进行本发明的一个或多个修饰和/或取代。以上参考突变胞溶素单体并且以下参考表征多核苷酸讨论的实施例中的任何实施例同样适用于本发明的这种方法。

[0559] 构建体

[0560] 本发明还提供了一种构建体,其包括衍生自胞溶素的两个或更多个共价连接单体,其中所述单体中的至少一个是本发明的突变胞溶素单体。本发明的构建体保留其形成孔的能力。一个或多个本发明的构建体可以用于形成用于表征靶分析物的孔。一个或多个本发明的构建体可以用于形成用于表征靶多核苷酸、如对靶核苷酸进行测序的孔。构建体可以包括2个、3个、4个、5个、6个、7个、8个、9个或10个或更多个单体。所述两个或更多个单体可以相同或不同。

[0561] 构建体中的至少一个单体是本发明的突变单体。构建体中的2个或更多个、3个或更多个、4个或更多个、5个或更多个、6个或更多个、7个或更多个、8个或更多个、9个或更多个或10个或更多个单体可以是本发明的突变单体。构建体中的所有单体优选地是本发明的突变单体。突变单体可以相同或不同。在优选实施例中,构建体包括两个本发明的突变单体。

[0562] 构建体中的本发明的突变单体的长度优选地大致相同或相同。构建体中的本发明的突变单体的桶的长度优选地大致相同或相同。长度可以以氨基酸数目和/或长度单位的形式度量。构建体中的本发明的突变单体的氨基酸数目优选地与从如上所述的位置34到位置70和/或位置71到位置107缺失的氨基酸数目相同。

[0563] 构建体中的其它单体不必是本发明的突变单体。例如,至少一个单体可以包括SEQ ID NO:2中所示的序列。构建体中的至少一个单体可以是SEQ ID NO:2的旁系同源物或同源物。在SEQ ID NO:14到17中示出了适当的同源物。

[0564] 可替代地,至少一个单体可以包括SEQ ID NO:2的变异体,所述变异体基于氨基酸同一性在其整个序列内与SEQ ID NO:2至少50%同源,但不包含本发明的突变单体所需的特定突变中的任何突变,或者在所述变异体中,还未缺失氨基酸,如上所述。更优选地,基于氨基酸同一性,变异体可以在整个序列内与SEQ ID NO:2的氨基酸序列至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,并且更优选地至少95%、97%或99%同源。变异体可以是片段或上文讨论的任何其它变异体。本发明的构建体还可以包括SEQ ID NO:14、15、16或17的变异体,所述变异体基于氨基酸同一性在其整个序列内与SEQ ID NO:14、15、16或17至少50%同源或以上提及的其它同源性水平中的至少任何同源性水平。

[0565] 构建体中的所有单体可以是本发明的突变单体。突变单体可以相同或不同。在更优选的实施例中,构建体包括两个单体,并且所述单体中的至少一个是本发明的突变单体。

[0566] 单体可以是基因融合的。如果从单个多核苷酸序列表达整个构建体,则单体是基因融合的。可以以任何方式组合单体的编码序列以形成对构建体进行编码的单个多核苷酸序列。在国际申请号PCT/GB09/001679号(公开为WO 2010/004265)中讨论了基因融合。

[0567] 可以以任何配置来将单体基因融合。可以通过单体的末端氨基酸融合单体。例如,可将一个单体的氨基端与另一单体的羧基端融合。

[0568] 可将两个或更多个单体直接基因融合在一起。优选地使用接头对单体进行基因融

合。接头可以被设计成限制单体的移动性。优选的接头是氨基酸序列(即,肽接头)。可以使用上文所讨论的肽接头中的任何肽接头。

[0569] 肽接头的长度、柔性和亲水性各自通常被设计为使得其不干扰单体和分子的功能。优选的柔性肽接头是2个到20个,如4个、6个、8个、10个或16个丝氨酸和/或甘氨酸的延伸段。更优选的柔性接头包含(SG)₁、(SG)₂、(SG)₃、(SG)₄、(SG)₅和(SG)₈,其中S是丝氨酸且G是甘氨酸。优选的刚性接头是2个到30个,如4个、6个、8个、16个或24个脯氨酸的延伸段。更优选的刚性接头包含(P)₁₂,其中P是脯氨酸。

[0570] 在另一优选实施例中,单体是化学融合的。如果例如通过化学交联剂对单体进行化学附接,则单体是化学融合的。可以使用上文所讨论的化学交联剂中的任何化学交联剂。接头可以附接到引入到突变单体中的一个或多个半胱氨酸残基或如Faz等非天然氨基酸。可替代地,接头可以附接到构建体中的单体之一的末端。单体通常通过SEQ ID NO:2的残基1到43和127到297中的一个或多个链接。

[0571] 如果构建体含有不同的单体,则可以通过使接头的浓度大大超过单体来防止单体自身的交联。可替代地,可以使用其中两个接头被使用的“锁和钥”布置。每个接头的仅一端可以反应在一起以形成较长接头,并且接头的其它端各自与不同单体反应。在国际申请号PCT/GB10/000132(公开为W0 2010/086602)中描述了这种接头。

[0572] 本发明还提供了一种产生本发明的构建体的方法。所述方法包括:将至少一个本发明的突变胞溶素单体共价连接到衍生自胞溶素的一个或多个单体。以上参考本发明的构建体讨论的实施例中的任何实施例同样适用于产生构建体的方法。

[0573] 多核苷酸

[0574] 本发明还提供了对本发明的突变单体进行编码的多核苷酸。突变单体可以是上文所讨论的突变单体中的任何突变单体。基于核苷酸同一性,多核苷酸序列优选地包括在整个序列内与SEQ ID NO:1的序列至少50%、60%、70%、80%、90%或95%同源的序列。在300个或更多个、例如375个、450个、525个或600个或更多个的连续核苷酸的延伸段内,可能存在至少80%、例如至少85%、90%或95%的核苷酸同一性(“硬同源性”)。可以如上文所描述的那样计算同源性。在遗传密码的简并的基础上,多核苷酸序列可以包括不同于SEQ ID NO:1的序列。

[0575] 本发明还提供了对本发明的基因融合构建体中的任何构建体进行编码的多核苷酸序列。多核苷酸优选地包括两个或更多个如SEQ ID NO:1所示的序列或如上所述的其变异体。

[0576] 可以使用所属领域中的标准方法来衍生或复制多核苷酸序列。可以从如赤子爱胜蚓等孔产生生物体中提取对野生型胞溶素进行编码的染色体DNA。可以使用涉及特异性引物的PCR来扩增对孔单体进行编码的基因。然后可以使所扩增的序列经历定点诱变。适当的定点诱变方法在所属领域中是已知的并且包含例如组合链式反应。可以使用熟知的技术制备对本发明的构建体进行编码的多核苷酸,所述技术如在Sambrook, J. 和Russell, D. (2001年),《分子克隆:实验室手册(Molecular Cloning: A Laboratory Manual)》第3版,纽约冷泉港冷泉港实验室出版社(Cold Spring Harbor Laboratory Press)中描述的技术。

[0577] 然后将所得多核苷酸序列并入到如克隆载体等重组可复制载体中。可以将载体用于在相容的宿主细胞中复制多核苷酸。因此,可以通过将多核苷酸引入到可复制载体

中、将所述载体引入到相容的宿主细胞中以及在引起载体复制的条件下使宿主细胞生长来制备多核苷酸序列。可以从宿主细胞中回收载体。用于克隆多核苷酸的适当宿主细胞在所属领域中是已知的且在下文更详细地描述。

[0578] 可以将多核苷酸序列克隆到适当表达载体中。在表达载体中，多核苷酸序列通常可操作地连接到能够通过宿主细胞提供对编码序列的表达的控制序列。这种表达载体可以用于表达孔亚基。

[0579] 术语“可操作地连接”是指并置，其中所描述的组分处于允许其以其预期方式起作用的关系。“可操作地连接”到编码序列的控制序列是以使得在与控制序列相容的条件下实现对编码序列的表达的方式进行连接。可以将相同或不同的多核苷酸序列的多个拷贝引入到载体中。

[0580] 然后将表达载体引入到适当的宿主细胞中。因此，可以通过将多核苷酸序列插入到表达载体中、将载体引入到相容的细菌宿主细胞中以及在引起对多核苷酸序列的表达的条件下使宿主细胞生长来产生本发明的突变单体或构建体。可以将以重组方式表达的单体或构建体自组装到宿主细胞膜中的孔中。可替代地，以此方式产生的重组孔可以从宿主细胞移除并且插入到另一膜中。当产生包括至少两个不同亚基的孔时，不同亚基可以在如上文所描述的不同宿主细胞中单独表达、从宿主细胞中移除并且组装到如羊红细胞膜或含有鞘磷脂的脂质体等单独的膜中的孔中。

[0581] 例如，可以通过添加包括鞘磷脂和以下脂质中的一种或多种的脂质混合物并且例如在30°C下温育所述混合物60分钟来使胞溶素单体寡聚：磷脂酰丝氨酸；POPE；胆固醇；以及Soy PC。寡聚单体可以通过任何适当的方法纯化，例如通过如W02013/153359中描述的SDS-PAGE和凝胶纯化。

[0582] 载体可以是例如具有复制原点、任选地用于表达所述多核苷酸序列的启动子以及任选地启动子的调节因子的质粒、病毒或噬菌体载体。载体可以含有一个或多个可选标记基因，例如四环素抗性基因。可以将启动子和其它表达调控信号选择为与表达载体被设计用于的宿主细胞相容。通常使用T7、trc、lac、ara或 λ 启动子。

[0583] 宿主细胞通常以高水平表达孔亚基。使用多核苷酸序列转化的宿主细胞可以被选择为与用于转化细胞的表达载体相容。宿主细胞通常是细菌并优选地是大肠杆菌。具有 λ DE3溶源体，例如C41 (DE3)、BL21 (DE3)、JM109 (DE3)、B834 (DE3)、TUNER、Origami和Origami B的任何细胞可以表达包括T7启动子的载体。除了以上列出的条件之外，还可以使用《美国国家科学院院刊(Proc Natl Acad Sci USA)》，2008年12月30日，第105卷，第52期，第20647到20642页中所述的方法中的任何方法表达胞溶素蛋白。

[0584] 孔

[0585] 本发明还提供了各种孔。本发明的孔对于表征分析物来说是理想的。本发明的孔对于表征多核苷酸序列如对多核苷酸进行测序来说是特别理想的，这是因为其可以以较高灵敏度区分不同核苷酸。所述孔可以用于表征如DNA和RNA等核酸，包括对核酸进行测序和识别单碱基变化。本发明的孔甚至可以区别甲基化和未甲基化的核苷酸。本发明的孔的基础分辨率非常高。所述孔显示所有四种DNA核苷酸的几乎完全分离。所述孔可以进一步用于基于在孔中的停留时间和流过孔的电流区分脱氧胞苷单磷酸酯(dCMP)和甲基-dCMP。

[0586] 本发明的孔还可以在一系列条件下区分不同核苷酸。具体地说，孔将在有利于表

征多核苷酸如对其进行测序的条件下区分核苷酸。本发明的孔可以区分不同核苷酸的程度可以通过改变所施加的电势、盐浓度、缓冲液、温度和如脲、甜菜碱和DTT等添加剂的存在进行控制。这允许对孔的功能进行微调,尤其在测序时。在下文中更详细地讨论这一点。本发明的孔还可以用于根据与一个或多个单体的相互作用而不是在逐核苷酸的基础上识别多核苷酸聚合物。

[0587] 本发明的孔可以是分离的、基本上分离的、纯化的或基本上纯化的。如果本发明的孔完全不含任何其它组分、如脂质或其它孔,则所述孔是分离的或纯化的。如果孔与将不干扰其既定用途的载体或稀释剂混合,则其是基本上分离的。例如,如果孔以包括小于10%、小于5%、小于2%或小于1%的其它组分如脂质或其它孔的形式存在,则其是基本上分离的或基本上纯化的。可替代地,本发明的孔可以存在于脂质双层中。

[0588] 本发明的孔可以作为单独的孔或单个孔存在。可替代地,本发明的孔可以以两个或更多个孔的同源群体或异源群体或多个两个或更多个孔存在。

[0589] 同源寡聚孔

[0590] 本发明还提供了一种衍生自胞溶素的同源寡聚孔,其包括一致的本发明的突变单体。就单体的氨基酸序列而言,所述单体一致。本发明的同源寡聚孔对于表征多核苷酸、如对其进行测序来说是理想的。本发明的同源寡聚孔可以具有上文所讨论的优点中的任何优点。在实施例中说明了本发明的特定同源寡聚孔的优点。

[0591] 同源寡聚孔可以含有任意数目的突变单体。所述孔通常包括两个或更多个突变单体。同源寡聚孔可以含有任意数目的突变单体。所述孔通常包括至少6个、至少7个、至少8个、至少9个或至少10个一致的突变单体,如6个、7个、8个、9个或10个突变单体。所述孔优选地包括八个或九个一致的突变单体。所述孔最优选地包括九个一致的突变单体。此数目的单体在本文中被称为“足够数目”。

[0592] 所述突变单体中的一个或多个,如2个、3个、4个、5个、6个、7个、8个、9个或10个优选地如上文或下文所讨论的那样进行化学修饰。

[0593] 所述突变单体中的一个或多个优选地如上文或下文所讨论的那样进行化学修饰。换言之,只要所述单体中的每一个的氨基酸序列一致,所述单体中经化学修饰的一个或多个(以及其它未经化学修饰的单体)不会阻止孔成为同源寡聚物。

[0594] 在Yamaji等人,《生物化学杂志(J. Biol. Chem.)》,1998年,第273卷,第9期,第5300到5306页中描述了用于制备胞溶素孔的方法。

[0595] 异源寡聚孔

[0596] 本发明还提供了一种衍生自胞溶素的异源寡聚孔,其包括至少一个本发明的突变单体,其中所述单体中的至少一个不同于其它单体。就单体的氨基酸序列而言,所述单体不同于其它单体。本发明的异源寡聚孔对于表征多核苷酸、如对其进行测序来说是理想的。可以使用所属领域中已知的方法(例如,《蛋白质科学(Protein Sci)》,2002年7月,第11卷,第7期,第1813到1824页)制备异源寡聚孔。

[0597] 异源寡聚孔含有足以形成孔的单体。单体可以属于任何类型,包含但不限于野生型。所述孔通常包括两个或更多个单体。所述孔通常包括至少6个、至少7个、至少8个、至少9个或至少10个单体,如6个、7个、8个、9个或10个单体。所述孔优选地包括八个或九个单体。所述孔最优选地包括九个单体。此数目的单体在本文中被称为“足够数目”。

[0598] 孔以包括至少一个包括SEQ ID NO:2所示序列的单体、其旁系同源物、其同源物或其变异体,所述变异体不具有本发明的突变单体所需的突变或在所述变异体中,还未缺失氨基酸,如上所述。适当的变异体是以上参考本发明的构建体讨论的变异体中的任何变异体,包含SEQ ID NO:2、14、15、16和17及其变异体。在此实施例中,剩余单体优选地是本发明的突变单体。

[0599] 在优选的实施例中,孔包括(a)一个本发明的突变单体和(b)足以形成孔的数目的一致单体,其中(a)中的突变单体不同于(b)中的一致单体。(b)中的一致单体优选地包括SEQ ID NO:2中所示的序列、其旁系同源物、其同源物或其变异体,所述变异体不具有本发明的突变单体所需的突变。

[0600] 本发明的异源寡聚孔优选地仅包括一个本发明的突变胞溶素单体。

[0601] 在另一个优选实施例中,异源寡聚孔中的所有单体都是本发明的突变单体,并且它们中的至少一个与其它单体不同。

[0602] 孔中的本发明的突变单体的长度优选地大致相同或相同。孔中的本发明的突变单体的桶的长度优选地大致相同或相同。长度可以以氨基酸数目和/或长度单位的形式度量。孔中的本发明的突变单体的氨基酸数目优选地与从位置34到位置70和/或位置71到位置107缺失的氨基酸数目相同。

[0603] 在以上所讨论的全部实施例中,所述突变单体中的一个或多个优选地如上文或下文所讨论的那样进行化学修饰。在一个单体上存在化学修饰不会导致孔为异源寡聚物。至少一个单体的氨基酸序列必须不同于其它单体的一个或多个序列。在下文更详细地讨论了用于制备孔的方法。

[0604] 含构建体的孔

[0605] 本发明还提供了一种孔,其包括至少一个本发明的构建体。本发明的构建体包括衍生自胞溶素的两个或更多个共价连接单体,其中所述单体中的至少一个是本发明的突变胞溶素单体。换言之,构建体必须含有多于一个单体。孔中的单体中的至少两个呈本发明的构建体的形式。单体可以属于任何类型。

[0606] 孔通常含有(a)包括两个单体的一个构建体和(b)足以形成孔的数目的单体。构建体可以是上文所讨论的构建体中的任何构建体。单体可以是以上讨论的单体中的任何单体,包含本发明的突变单体。

[0607] 另一典型的孔包括多于一个本发明的构建体,如两个、三个或四个本发明的构建体。这种孔进一步包括足以形成孔的数目的单体。单体可以是上文所讨论的突变单体中的任何单体。本发明的进一步孔仅包括包含2个单体的构建体。根据本发明的特定孔包括各自包括两个单体的若干构建体。构建体可以寡聚成孔,所述孔具有使得来自每个构建体的仅一个单体贡献于孔的结构。通常,构建体的其它单体(即不形成孔的单体)将位于孔的外部。

[0608] 可以如以上所讨论的那样将突变引入到构建体中。突变可以是交替的,即突变对于双单体构建体内的每个单体来说是不同的,且构建体组装为同源寡聚物,从而产生交替的修饰。换言之,包括MutA和MutB的单体被融合和组装以形成A-B:A-B:A-B:A-B孔。可替代地,突变可以是相邻的,即一致的突变引入到构建体中的两个单体中,并且这然后与不同突变单体寡聚。换言之,包括MutA的单体被融合,随后是与含MutB的单体寡聚以形成A-A:B:B:B:B:B:B。

[0609] 可以如上文或下文所讨论的那样对含构建体的孔中的本发明的单体中的一个或多个进行化学修饰。

[0610] 本发明的经化学修饰的孔

[0611] 在另一方面,本发明提供了一种经化学修饰的胞溶素孔,其包括一个或多个突变单体,所述突变单体被化学修饰成使得组装孔的桶/通道的开口直径沿着桶的长度在一个位点或更多个位点、如两个、三个、四个或五个位点处减小、变窄或收缩。孔可以包括任何数目的上文参考本发明的同源寡聚孔和异源寡聚孔讨论的单体。孔优选地包括九个经化学修饰的单体。经化学修饰的孔可以是同源寡聚物,如上所述。换言之,经化学修饰的孔中的所有单体可以具有相同的氨基酸序列并且可以以相同的方式进行化学修饰。经化学修饰的孔可以是异源寡聚物,如上所述。换言之,孔可以包括(a) 仅一个经化学修饰的单体,(b) 多于一个,如两个、三个、四个、五个、六个、七个或八个经化学修饰的单体,其中所述经化学修饰的单体中的至少两个,如三个、四个、五个、六个或七个彼此不同或者(c) 仅经化学修饰的单体(即,所有单体都是经化学修饰的),其中所述经化学修饰的单体中的至少两个,如三个、四个、五个、六个、七个、八个或九个彼此不同。单体可以就其氨基酸序列、其化学修饰或其氨基酸序列和其化学修饰两者而言彼此不同。一个或多个经化学修饰的单体可以是上文和/或下文讨论的经化学修饰的单体中的任何经化学修饰的单体。

[0612] 本发明还提供了一种以下文所讨论的方式中的任何方式化学修饰的突变胞溶素单体。突变单体可以是上文或下文所讨论的突变单体中的任何突变单体。因此,可以根据本发明、如下文所讨论的那样对本发明的突变单体,如包括以下位置中的一个或多个处的修饰的SEQ ID NO:2的变异体:K37、G43、K45、V47、S49、T51、H83、V88、T91、T93、V95、Y96、S98、K99、V100、I101、P108、P109、T110、S111、K112和T114或包括上述桶缺失的变异体进行化学修饰。

[0613] 可以将突变单体化学修饰为使得组装孔的桶的直径通过取决于要通过孔的分析物的大小的任何减小因子减小或变窄。收缩区的宽度将通常决定在分析物转移期间的测量信号由于例如分析物减少了通过孔的离子流而破坏的程度。信号破坏越大,通常测量灵敏度越高。因此,可以将收缩区选择成略宽于待转移的分析物。对于例如ssDNA的转移,收缩区的宽度可以选自0.8nm到3.0nm范围内的值。

[0614] 化学修饰还可以决定收缩区的长度,所述收缩区的长度反过来将决定贡献于测量信号的聚合物单元,例如核苷酸的数目。在任何特定时间贡献于电流信号的核苷酸可以被称为k聚体,其中k是整数,并且可以是整数或分数。在测量具有4种类型的核碱基的多核苷酸的情况下,3聚体将产生 4^3 个潜在的信号电平。较大的k值会产生更大数目的信号电平。通常期望提供短的收缩区域,因为这简化了对测量信号数据的分析。

[0615] 化学修饰为使得化学分子优选地共价连接到突变单体或所述一个或多个突变单体。可以使用所属领域中已知的任何方法将化学分子共价连接到孔、突变单体或一个或多个突变单体。化学分子通常通过化学连接附接。

[0616] 优选地通过将分子附接到一个或多个半胱氨酸(半胱氨酸连接)、将分子附接到一个或多个赖氨酸、将分子附接到一个或多个非天然氨基酸、表位的酶修饰对突变单体或一个或多个突变单体进行化学修饰。如果通过半胱氨酸连接来附接化学修饰剂,则所述一个或多个半胱氨酸已经优选地通过取代引入到突变体。用于执行这种修饰的适当方法在所属

领域中是众所周知的。适当的非天然氨基酸包含但不限于4-叠氮基-L-苯丙氨酸(Faz)和Liu C.C.和Schultz P.G.,《生物化学年评(Annu.Rev.Biochem.)》,2010年,第79卷,第413到444页的图1中编号为1到71的氨基酸中的任何氨基酸。

[0617] 可以通过将具有减少或缩小组装孔的桶的直径的作用的任何分子附接在任何位置或位点处对突变单体或一个或多个突变单体进行化学修饰。可以通过附接如下各项来对突变单体进行化学修饰:(i) 马来酰亚胺,如:4-苯氮霉素(4-phenylazomaleinani)、1.N-(2-羟乙基)马来酰亚胺、N-环己基马来酰亚胺、1.3-马来酰亚胺基丙酸、1.1-4-氨基苯基-1H-吡咯,2,5,二酮、1.1-4-羟基苯基-1H-吡咯,2,5,二酮、N-乙基马来酰亚胺、N-甲氧基羰基马来酰亚胺、N-叔丁基马来酰亚胺、N-(2-氨基乙基)马来酰亚胺、3-马来酰亚胺基-PROXYL、N-(4-氯苯基)马来酰亚胺、1-[4-(二甲基氨基)-3,5-二硝基苯基]-1H-吡咯-2,5-二酮、N-[4-(2-苯并咪唑基)苯基]马来酰亚胺、N-[4-(2-苯并恶唑基)苯基]马来酰亚胺、N-(1-萘基)马来酰亚胺、N-(2,4-二甲苯基)马来酰亚胺、N-(2,4-二氟苯基)马来酰亚胺、N-(3-氯-对-甲苯基)-马来酰亚胺、1-(2-氨基-乙基)-吡咯-2,5-二酮盐酸盐、1-环戊基-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、1-(3-氨基丙基)-2,5-二氢-1H-吡咯-2,5-二酮盐酸盐、3-甲基-1-[2-氧代-2-(哌嗪-1-基)乙基]-2,5-二氢-1H-吡咯-2,5-二酮盐酸盐、1-苄基-2,5-二氢-1H-吡咯-2,5-二酮、3-甲基-1-(3,3,3-三氟丙基)-2,5-二氢-1H-吡咯-2,5-二酮、1-[4-(甲基氨基)环己基]-2,5-二氢-1H-吡咯-2,5-二酮三氟乙酸、SMILES O=C1C=CC(=O)N1CC=2C=CN=CC2、SMILES O=C1C=CC(=O)N1CN2CCNCC2、1-苄基-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、1-(2-氟苯基)-3-甲基-2,5-二氢-1H-吡咯-2,5-二酮、N-(4-苯氧基苯基)马来酰亚胺、N-(4-硝基苯基)马来酰亚胺;(ii) 碘代乙酰胺,如3-(2-碘乙酰氨基)-PROXYL、N-(环丙基甲基)-2-碘乙酰胺、2-碘-N-(2-苄乙基)乙酰胺、2-碘-N-(2,2,2-三氟乙基)乙酰胺、N-(4-乙酰基苯基)-2-碘代乙酰胺、N-(4-(氨基磺酰基)苯基)-2-碘代乙酰胺、N-(1,3-苯并噻唑-2-基)-2-碘代乙酰胺、N-(2,6-二乙基苯基)-2-碘代乙酰胺、N-(2-苯甲酰基-4-氯苯基)-2-碘代乙酰胺;(iii) 溴代乙酰胺:如N-(4-(乙酰氨基)苯基)-2-溴代乙酰胺、N-(2-乙酰基苯基)-2-溴代乙酰胺、2-溴-N-(2-氰基苯基)乙酰胺、2-溴-N-(3-(三氟甲基)苯基)乙酰胺、N-(2-苯甲酰基苯基)-2-溴代乙酰胺、2-溴-N-(4-氟苯基)-3-甲基丁酰胺、N-苄基-2-溴-N-苯基丙酰胺、N-(2-溴-丁酰基)-4-氯-苯磺酰胺、2-溴-N-甲基-N-苯基乙酰胺、2-溴-N-苄乙基-乙酰胺、2-金刚烷-1-基-2-溴-N-环己基-乙酰胺、2-溴-N-(2-甲基苯基)丁酰胺、乙酰替对溴苯胺;(iv) 二硫化物,如:ALDRITHIOL-2、ALDRITHIOL-4、异丙基二硫化物、1-(异丁基二硫烷基)-2-甲基丙烷、二苄基二硫化物、4-氨基苯基二硫化物、3-(2-吡啶基二硫代)丙酸、3-(2-吡啶基二硫代)丙酸酐、3-(2-吡啶基二硫代)丙酸N-琥珀酰亚胺酯、am6amPDP1-βCD;

[0618] 以及(v) 硫醇,如:4-苯基噻唑-2-硫醇、PulpaId、5,6,7,8-四氢-噻唑啉-2-硫醇。

[0619] 突变单体或一个或多个突变单体可以通过附接聚乙二醇(PEG)、如DNA等核酸、染料、荧光团或发色团进行化学修饰。在一些实施例中,使用促进包括单体的孔与靶分析物、靶核苷酸或靶多核苷酸序列之间的相互作用的分子衔接子对突变单体或一个或多个突变单体进行化学修饰。衔接子的存在改进了孔和核苷酸或多核苷酸的主客体化学,并且由此提高了由突变单体形成的孔的测序能力。

[0620] 可以通过将具有减少或缩小组装孔的桶的开口直径的作用的任何分子附接在任

何位置处对突变单体或一个或多个突变单体进行化学修饰。K37、V47、S49、T55、S86、E92、E94。更优选地，可以通过将具有减少或缩小组装孔的桶的开口直径的作用的任何分子附接在位置E92和E94处对突变单体进行化学修饰。在一个实施例中，通过这些位置处将分子附接到一个或多个半胱氨酸(半胱氨酸连接)对突变单体或一个或多个突变单体进行化学修饰。

[0621] 可以通过修饰相邻残基来增强半胱氨酸残基的反应性。例如，侧接精氨酸、组氨酸或赖氨酸残基的碱性基团将半胱氨酸巯基的pKa改变更具反应性的S⁻基的pKa。可以通过如dTNB等巯基保护基来保护半胱氨酸残基的反应性。在附接接头之前，这些可以与突变单体的一个或多个半胱氨酸残基反应。

[0622] 可以将所述分子直接附接到突变单体或一个或多个突变单体。优选地，使用如化学交联剂或肽接头等接头将所述分子附接到突变单体。适当的化学交联剂在所属领域中是众所周知的。优选的交联剂包含3-(吡啶-2-基二磺酰基)丙酸2,5-二氧化吡咯烷-1-基酯、4-(吡啶-2-基二磺酰基)丁酸2,5-二氧化吡咯烷-1-基酯以及8-(吡啶-2-基二磺酰基)辛酸2,5-二氧化吡咯烷-1-基酯。最优选的交联剂是3-(2-吡啶二硫代)丙酸丁二酰亚胺酯(SPDP)。通常，在分子/交联剂复合体共价连接到突变单体之前，分子共价连接到双功能交联剂，但也有可能在双功能交联剂/单体复合体附接到分子之前将双功能交联剂共价连接到单体。

[0623] 优选地，接头对二硫苏糖醇(DTT)具有抗性。适当的接头包含但不限于基于碘乙酰胺和基于马来酰亚胺的接头。

[0624] 以这种方式化学修饰的孔显示出以下特定优点：(i) 提高读头的清晰度(ii) 提高对碱基的区分以及(iii) 提高范围，即提高信噪比。

[0625] 通过使用化学分子来修饰桶内的特定位置，可以引入新的读头或者可以修饰旧的读头。由于经修饰分子的大小，可以显著改变读头的物理大小。类似地，由于经修饰分子的化学性质，可以改变读头的性质。已经证明两种效果的组合使得读头具有提高的分辨率和更好的碱基区分。不仅对不同位置处的不同碱基的信号相对贡献被改变，极端处的读头位置显示出少得多的区分，从而意味着它们对信号的贡献大大减少，并且因此在给定时刻测定的K聚体的长度较短。这种更清晰的读头使得从原始信号中解卷积K聚体的过程更简单。

[0626] 产生本发明的孔

[0627] 本发明还提供了一种产生本发明的孔的方法。所述方法包括：允许至少一个本发明的突变单体或至少一个本发明的构建体与足够数目的本发明的突变胞溶素单体、本发明的构建体、胞溶素单体或衍生自胞溶素的单体寡聚以形成孔。如果所述方法涉及制备本发明的同源寡聚孔，则所述方法中使用的所有单体都是具有相同氨基酸序列的本发明的突变胞溶素单体。如果所述方法涉及制备本发明的异源寡聚孔，则所述单体中的至少一个与其它单体不同。

[0628] 通常，单体在如上文所描述的宿主细胞中表达、从宿主细胞中移除并且组装到如羊红细胞膜或含有鞘磷脂的脂质体等单独的膜中的孔中。

[0629] 例如，可以通过添加包括鞘磷脂和以下脂质中的一种或多种的脂质混合物并且例如在30℃下温育所述混合物60分钟来使胞溶素单体寡聚：磷脂酰丝氨酸；POPE；胆固醇；以

及Soy PC。寡聚单体可以通过任何适当的方法纯化,例如通过如W02013/153359中描述的SDS-PAGE和凝胶纯化。

[0630] 以上参考本发明的孔讨论的实施例中的任何实施例同样适用于产生孔的方法。

[0631] 表征分析物的方法

[0632] 本发明提供了一种表征靶分析物的方法。所述方法包括:使所述靶分析物与本发明的孔接触,使得所述靶分析物移动穿过所述孔。所述孔可以是上文所讨论的孔中的任何孔。然后,当分析物相对于孔移动时,使用所属领域中已知的标准方法测量靶分析物的一个或多个特性。优选地在分析物移动穿过孔时测量靶分析物的一个或多个特性。步骤(a)和(b)优选地在跨孔施加电势的情况下进行。如下文更详细地讨论的,所施加的电势通常引起在所述孔与多核苷酸结合蛋白之间形成复合体。所施加的电势可以是电压电势。可替代地,所施加的电势可以是化学电势。这种操作的实例是跨两亲层使用盐梯度。在Holden等人,《美国化学会志(J Am Chem Soc.)》,2007年7月11日,第129卷,第27期,第8650到8655页中公开了盐梯度。

[0633] 本发明的方法用于表征靶分析物。所述方法用于表征至少一个分析物。所述方法可以涉及表征两个或更多个分析物。所述方法可以包括表征任何数目的分析物,如2个、5个、10个、15个、20个、30个、40个、50个、100个或更多个分析物。

[0634] 所述靶分析物优选地是金属离子、无机盐、聚合物、氨基酸、肽、多肽、蛋白质、核苷酸、寡核苷酸、多核苷酸、染料、漂白剂、药物、诊断剂、娱乐性药物、爆炸品或环境污染物。所述方法可以涉及表征相同类型的两个或更多个分析物,如两个或更多个蛋白质、两个或更多个核苷酸或两个或更多个药物。可替代地,所述方法可以涉及表征不同类型的两个或更多个分析物,如一个或多个蛋白质、一个或多个核苷酸和一个或多个药物。

[0635] 可以从细胞中分泌靶分析物。可替代地,靶分析物可以是存在于细胞内部的分析物,使得在可以执行本发明之前必须从细胞中提取分析物。

[0636] 分析物优选地是氨基酸、肽、多肽和/或蛋白质。氨基酸、肽、多肽或蛋白质可以是天然存在的或非天然存在的。多肽或蛋白质可以在其内包含合成或经修饰氨基酸。对氨基酸的多种不同类型修饰在所属领域中是已知的。适当的氨基酸和其修饰如上文。出于本发明的目的,应理解,可以通过所属领域中可用的任何方法修饰靶分析物。

[0637] 蛋白质可以是酶、抗体、激素、生长因子或生长调控蛋白,如细胞因子。细胞因子可以选自:白介素,优选地IFN-1、IL-1、IL-2、IL-4、IL-5、IL-6、IL-10、IL-12和IL-13;干扰素,优选地IL- γ ;以及其它细胞因子,如TNF- α 。蛋白质可以是细菌蛋白质、真菌蛋白质、病毒蛋白质或寄生虫衍生蛋白质。

[0638] 靶分析物优选地是核苷酸、寡核苷酸或多核苷酸。核苷酸通常含有核碱基、糖和至少一个磷酸基。核碱基通常是杂环的。核碱基包含但不限于嘌呤和嘧啶,以及更具体地,腺嘌呤、鸟嘌呤、胸腺嘧啶、尿嘧啶和胞嘧啶。糖通常是戊糖。核苷酸糖包含但不限于核糖和脱氧核糖。核苷酸通常是核糖核苷酸或脱氧核糖核苷酸。核苷酸通常含有单磷酸、二磷酸或三磷酸。磷酸可以附接在核苷酸的5'或3'侧上。

[0639] 核苷酸包含但不限于: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)、5-甲基胞苷单磷酸、5-甲基胞苷二磷酸、5-甲基胞苷三磷酸、5-羟基甲基胞苷单磷酸、5-羟基甲基胞苷二磷酸、5-羟基甲基胞苷三磷酸、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、二磷酸脱氧腺苷(dADP)、三磷酸脱氧腺苷(dATP)、单磷酸脱氧鸟苷(dGMP)、二磷酸脱氧鸟苷(dGDP)、三磷酸脱氧鸟苷(dGTP)、单磷酸脱氧胸苷(dTMP)、二磷酸脱氧胸苷(dTDP)、三磷酸脱氧胸苷(dTTP)、单磷酸脱氧尿苷(dUMP)、二磷酸脱氧尿苷(dUDP)、三磷酸脱氧尿苷(dUTP)、单磷酸脱氧胞苷(dCMP)、二磷酸脱氧胞苷(dCDP)和三磷酸脱氧胞苷(dCTP)、5-甲基-2'-脱氧胞苷单磷酸、5-甲基-2'-脱氧胞苷二磷酸、5-甲基-2'-脱氧胞苷三磷酸、5-羟基甲基-2'-脱氧胞苷单磷酸、5-羟基甲基-2'-脱氧胞苷二磷酸以及5-羟基甲基-2'-脱氧胞苷三磷酸。核苷酸优选地选自AMP、TMP、GMP、UMP、dAMP、dTMP、dGMP或dCMP。核苷酸可以无碱基的(即缺乏核碱基)。核苷酸可以含有额外修饰。具体地说,适当的经修饰核苷酸包含但不限于2'-氨基嘧啶(如2'-氨基胞苷和2'-氨基尿苷)、2'-羟基嘌呤(如2'-氟嘧啶(2'-氟胞苷和2'-氟尿苷)、羟基嘧啶(如5'- α -P-硼烷尿苷)、2'-O-甲基核苷酸(如2'-O-甲基腺苷、2'-O-甲基鸟苷、2'-O-甲基胞苷和2'-O-甲基尿苷)、4'-硫代嘧啶(如4'-硫代尿苷和4'-硫代胞苷),并且核苷酸具有对核碱基的修饰(如5-戊炔基-2'-脱氧尿苷、5-(3-氨基丙基)-尿苷和1,6-二氨基己基-N-5-氨基甲酰基甲基尿苷)。

[0640] 寡核苷酸是短核苷酸聚合物,其通常具有50个或更少核苷酸,如40个或更少、30个或更少、20个或更少、10个或更少或5个或更少核苷酸。寡核苷酸可以包括下文讨论的核苷酸中的任何核苷酸,包含无碱基和经修饰核苷酸。本发明的方法优选地用于表征靶多核苷酸。如核酸等多核苷酸是包括两个或更多个核苷酸的大分子。多核苷酸或核酸可以包括任何核苷酸的任何组合。核苷酸可以是天然存在的或人工的。靶多核苷酸中的一个或多个核苷酸可以是氧化或甲基化的。靶多核苷酸中的一个或多个核苷酸可以是受损的。例如,多核苷酸可以包括嘧啶二聚体。这种二聚体通常与紫外光引起的损坏相关联并且是皮肤黑色素瘤的主要病因。靶多核苷酸中的一个或多个核苷酸可以例如用标记或标签修饰。在下文描述了适当的标记。靶多核苷酸可以包括一个或多个间隔子。

[0641] 上文定义了核苷酸。多核苷酸中存在的核苷酸包含但不限于:单磷酸腺苷(AMP)、单磷酸鸟苷(GMP)、单磷酸胸苷(TMP)、单磷酸尿苷(UMP)、单磷酸胞嘧啶核苷(CMP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、单磷酸脱氧鸟苷(dGMP)、单磷酸脱氧胸苷(dTMP)、单磷酸脱氧尿苷(dUMP)和单磷酸脱氧胞苷(dCMP)。核苷酸优选地选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP、dCMP和dUMP。

[0642] 核苷酸可以无碱基的(即缺乏核碱基)。

[0643] 多核苷酸中的核苷酸可以以任何方式彼此附接。核苷酸通常通过其糖和磷酸基附接,如在核酸中那样。核苷酸可以通过其核碱基连接,如在嘧啶二聚体中那样。

[0644] 多核苷酸可以是单链的或双链的。多核苷酸的至少一部分优选地是双链的。单链多核苷酸可以具有与其杂交的一个或多个引物,并且因此包括双链多核苷酸的一个或多个短区。引物可以是与靶多核苷酸相同类型的多核苷酸或者可以是不同类型的多核苷酸。

[0645] 多核苷酸可以是核酸,如脱氧核糖核酸(DNA)或核糖核酸(RNA)。多核苷酸可以包括与一条DNA链杂交的一条RNA链。多核苷酸可以是所属领域中已知的任何合成核酸,如肽核酸(PNA)、甘油核酸(GNA)、苏糖核酸(TNA)、锁核酸(LNA)或具有核苷酸侧链的其它合成聚

合物。

[0646] 靶多核苷酸的整体或仅一部分可以使用此方法来表征。靶多核苷酸可以具有任何长度。例如,多核苷酸的长度可以是至少10个、至少50个、至少100个、至少150个、至少200个、至少250个、至少300个、至少400个或至少500个核苷酸对。多核苷酸的长度可以是1000个或更多个核苷酸对、5000个或更多个核苷酸对、或100000个或更多个核苷酸对。

[0647] 如靶多核苷酸等靶分析物存在于任何适当的样品中。通常对已知含有或疑似含有如靶多核苷酸等靶分析物的样品执行本发明。可替代地,可以对样品执行本发明以确认如一个或多个靶多核苷酸等其在样品中的存在是已知或预期的一个或多个靶分析物的同一性。

[0648] 样品可以是生物样品。可以对从任何生物体或微生物中获得或提取的样品体外执行本发明。生物体或微生物通常是古细菌、原核或真核微生物,并且通常属于以下五界之一:植物界、动物界、真菌界、原核生物界和原生生物界。可以对从任何病毒中获得或提取的样品体外执行本发明。样品优选地是流体样品。样品通常包括患者的体液。样品可以是尿液、淋巴液、唾液、粘液或羊膜液,但优选地是血液、血浆或血清。通常,样品来源于人,但可替代地其可以来自另一哺乳动物,如来自商业养殖动物,如马、牛、绵羊或猪,或可替代地可以是宠物,如猫或狗。可替代地,来源于植物的样品通常可以从经济作物获得,如谷物、豆科植物、果实或蔬菜,例如小麦、大麦、燕麦、油菜、玉米、大豆、稻谷、大黄、香蕉、苹果、蕃茄、马铃薯、葡萄、烟草、菜豆、小扁豆、甘蔗、可可、棉花。

[0649] 样品可以是非生物样品。非生物样品优选地是流体样品。非生物样品的实例包含手术液、如饮用水、海水或河水等水以及实验室测试用试剂。

[0650] 样品通常在测定之前处理,例如通过离心或穿过滤出不需要的分子或如红细胞等细胞的膜。可以在紧接着取得样品之后对其进行测量。通常还可以在分析之前优选地在低于-70℃下储存样品。

[0651] 孔通常存在于膜中。根据本发明可以使用任何膜。适当的膜在所属领域中是众所周知的。膜优选地包括鞘磷脂。膜优选地是两亲性层。两亲性层是由如磷脂等具有至少一个亲水部分和至少一个亲脂或疏水部分的两亲性分子形成的层。两亲分子可以是合成的或天然存在的。非天然存在的两亲物和形成单层的两亲物在所属领域中是已知的并且包括例如嵌段共聚物(Gonzalez-Perez等人,《朗缪尔(Langmuir)》,2009年,第25卷,第10447到10450页)。嵌段共聚物是两个或更多个单体亚基聚合在一起以产生单个聚合物链的聚合材料。嵌段共聚物通常具有通过每个单体亚基贡献的性质。然而,嵌段共聚物可以具有由单独的亚基形成的聚合物不拥有的独特性质。嵌段共聚物可以被工程化为使得单体亚基之一在水性介质中是疏水性的(即亲脂性),而一个或多个其它亚基是亲水性的。在这种情况下,嵌段共聚物可以拥有两亲性质,并且可以形成模拟生物膜的结构。嵌段共聚物可以是二嵌段(其由两个单体亚基组成),但也可以由多于两个单体亚基构造以形成表现为两亲物的更复杂布置。共聚物可以是三嵌段、四嵌段或五嵌段共聚物。

[0652] 两亲性层可以是单层或双层。两亲性层通常是平面脂质双层或支撑双层。

[0653] 两亲性层通常是脂质双层。脂质双层是细胞膜的模型,并且充当一系列实验研究的极佳平台。例如,脂质双层可以用于通过单通道记录对膜蛋白进行体外研究。可替代地,脂质双层可以用作检测一系列物质的存在的生物传感器。脂质双层可以是任何脂质双层。

适当的脂质双层包含但不限于平坦脂质双层、支撑双层或脂质体。脂质双层优选地是平坦脂质双层。在以下文献中公开了适当的脂质双层：国际申请号PCT/GB08/000563（公开为W0 2008/102121）、国际申请号PCT/GB08/004127（公开为W0 2009/077734）以及国际申请号PCT/GB2006/001057（公开为W0 2006/100484）。

[0654] 用于形成脂质双层的方法在所属领域中是已知的。在实例中公开了适当的方法。脂质双层通常通过Montal和Mueller的方法（《美国国家科学院院刊(Proc Natl Acad Sci USA), 1972年, 第69卷, 第3561页3566）形成, 在所述方法中, 脂质单层携带于通过开孔任一侧的水溶液/空气界面上, 所述开孔垂直于所述界面。

[0655] Montal和Mueller的方法是流行的, 这是因为其是具有成本效益的并且是形成适合于蛋白孔插入的良好质量脂质双层的相对直接了当的方法。双层形成的其它常见方法包含脂质体双层的尖端浸没、双层涂刷和贴片夹持。

[0656] 在优选实施例中, 如国际申请号PCT/GB08/004127（公开为W0 2009/077734）中公开的那样形成脂质双层。在另一优选实施例中, 膜是固态层。固态层不是生物来源的。换言之, 固态层不是来源于生物环境或与其隔离, 如生物体或细胞或生物学上可获得的结构的合成制造版本。固态层可以由有机材料和无机材料两者形成, 所述材料包含但不限于: 微电子材料; 绝缘材料, 如 Si_3N_4 、 Al_2O_3 和 SiO_2 ; 有机和无机聚合物, 如聚酰胺; 塑料, 如Teflon®, 或弹性体, 如二组分加成固化的硅橡胶; 以及玻璃。固态层可以由如石墨烯等单原子层或仅几个原子厚的层形成。在国际申请号PCT/US2008/010637（公开为W0 2009/035647）中公开了适当的石墨烯层。

[0657] 通常使用以下来执行所述方法: (i) 包括孔的人工两亲形层; (ii) 包括孔的分离的天然存在的脂质双层; 或 (iii) 具有插入其中的孔的细胞。通常使用如人工脂质双层等人工两亲性层执行所述方法。所述层可以包括其它跨膜和/或膜内蛋白质以及除孔以外的其它分子。下文讨论了适当的设备和条件。通常在体外执行本发明的方法。

[0658] 如靶多核苷酸等分析物可以偶联到膜。可以使用任何已知的方法完成这一点。如果膜是如脂质双层等两亲性层(如上文详细讨论的), 则如靶多核苷酸等分析物优选通过存在于膜中的多肽或存在于膜中的疏水锚偶联到膜。疏水性锚优选地是脂质、脂肪酸、固醇、碳纳米管或氨基酸。

[0659] 如靶多核苷酸等分析物可以直接偶联到膜。如靶多核苷酸等分析物优选地通过接头偶联到膜。优选的接头包含但不限于聚合物, 如多核苷酸、聚乙二醇(PEG)以及多肽。如果多核苷酸直接偶联到膜, 则一些数据将丢失, 因为运行的表征由于膜与孔内部之间的距离而无法继续到多核苷酸的末端。如果使用接头, 则多核苷酸可得到完全处理。如果使用接头, 则接头可以在任何位置附接到多核苷酸。接头优选地在尾部聚合物处附接到多核苷酸。

[0660] 偶联可以是稳定的或暂时性的。对于某些应用来说, 偶联的暂时性质是优选的。如果稳定的偶联分子直接附接到多核苷酸的5'或3'末端, 则一些数据将会丢失, 因为运行的表征由于双层与孔内部之间的距离而无法继续到多核苷酸的末端。如果偶联是暂时性的, 则当偶联的末端随机变为不含双层时, 那么多核苷酸可以得到完全处理。下面更详细地讨论了与膜形成稳定或瞬时连接的化学基团。如靶多核苷酸等分析物可以使用胆固醇或脂肪酰链暂时性地偶联到如脂质双层等两亲性层。可以使用长度为6个到30个碳原子的任何脂肪酰基链, 如十六烷酸。

[0661] 在优选的实施例中,如靶多核苷酸等分析物偶联到两亲性层。之前已经使用各种不同拴系策略执行如靶多核苷酸等分析物与合成脂质双层的偶联。在下表5中总结了这些信息:

[0662] 表5

衔接基团	偶联类型	参考文献
硫醇	稳定	Yoshina-Ishii, C.和 S.G.Boxer (2003 年), “《支撑脂质双层上的可移动拴系囊泡阵列 (Arrays of mobile tethered vesicles on supported lipid bilayers)》”,《美国化学会志 (J Am Chem Soc)》, 第 125 卷, 第 13 期, 第 3696 和 3697 页。
生物素	稳定	Nikolov, V.、R.Lipowsky 等人 (2007 年), “《具有锚定 DNA 分子的巨型囊泡的行为 (Behavior of giant vesicles with anchored DNA molecules)》”,《生物物理学杂志 (Biophys J)》, 第 92 卷, 第 12 期, 第 4356 到 4368 页
胆固醇	暂时性的	Pfeiffer, I.和 F.Hook (2004 年), “《寡核苷酸与脂膜组装件的基于二价胆固醇的偶联 (Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies)》”,《美国化学会志 (J Am Chem Soc)》, 第 126 卷, 第 33 期, 第 10224 和 10225 页
脂质	稳定	van Lengerich, B.、R.J.Rawle 等人, “《脂质囊泡与液体支撑双层的共价连接允许观测 DNA 介导的囊泡相互作用 (Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions)》”,《朗缪尔 (Langmuir)》, 第 26 卷, 第 11 期, 第 8666 到 8672 页

[0663] 可以在合成反应中使用经修饰的亚磷酰胺对多核苷酸进行官能化,所述经修饰的亚磷酰胺对于添加如硫醇、胆固醇、脂质和生物素基团等反应性基团来说容易相容。这些不同的衔接化学为多核苷酸产生一系列衔接选项。每个不同的修饰基团以稍微不同的方式拴系多核苷酸,并且偶联未必总是永久性的,因此给出不同的停留时间以供多核苷酸偶联到双层。上文讨论了暂时性偶联的优点。

[0664] 多核苷酸的偶联还可以通过多种其它手段来实现,条件是以向多核苷酸添加互异性基团。之前已经报告过向DNA的任一末端添加反应性基团。可以使用多核苷酸激酶和ATP γ S向ssDNA的5'添加巯基(Grant, G.P.和P.Z.Qin(2007年), “《一种用于在核酸的5'末端处衔接氮氧化物自旋标记的简易方法 (A facile method for attaching nitroxide spin labels at the 5' terminus of nucleic acids)》”,《核酸研究 (Nucleic Acids Res)》, 第 35 期, 第 10 卷, 第 e77 页)。可以使用末端转移酶将经修饰寡核苷酸并入到ssDNA的3'中来添加如生物素、硫醇和荧光团等更多样的化学基团集合(Kumar, A.、P.Tchen等人(1988年), “《使用末端脱氧核苷酸转移酶对合成寡核苷酸探针进行非放射性标记 (Nonradioactive labelling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase)》”,《分析生物化学 (Anal Biochem)》, 第 169 卷, 第 2 期, 第 376 到 382 页)。

[0665] 可替代地,可以将反应性基团视为是对与已经偶联到双层的DNA互补的短片DNA的添加,使得可以通过杂交实现衔接。已报告了使用T4 RNA连接酶I对短片ssDNA进行连接(Troutt, A.B.、M.G.McHeyzer-Williams等人(1992年), “《连接锚定的PCR:具有单侧特异性的简单扩增技术 (Ligation-anchored PCR: a simple amplification technique with single-sided specificity)》”,《美国国家科学院院刊 (Proc Natl Acad Sci U S A)》, 第 89 卷, 第 20 期, 第 9823 到 9825 页)。可替代地,可以将ssDNA或dsDNA连接到原生dsDNA,并且然

后通过热或化学变性分离两条链。对于原生dsDNA,可以将一片ssDNA添加到双链体的末端中一个或两个,或者将dsDNA添加到一个或两个末端。然后,当双链体熔化时,如果ssDNA用于在5'末端、3'末端处进行连接或修饰,则每个单链将具有5'修饰或3'修饰,或者如果dsDNA用于连接,则每个单链将具有5'和3'修饰两者。如果多核苷酸是合成链,则可以在化学合成多核苷酸期间并入偶联化学。例如,可以使用反应性基团附接到其上的引物来合成多核苷酸。

[0667] 用于扩增基因组DNA区段的常见技术是使用聚合酶链式反应(PCR)。此处,使用两个合成的寡核苷酸引物,可以产生相同DNA区段的大量拷贝,其中对于每个拷贝,双链体中的每个链的5'将是合成多核苷酸。通过使用如胆固醇、硫醇、生物素或脂质等具有反应性基团的反义引物,所扩增的靶DNA的每个拷贝将含有用于偶联的反应性基团。

[0668] 本发明的方法中使用的孔是本发明的孔(即,包括至少一个本发明的突变单体或至少一个本发明的构建体的孔)。可以以上讨论的方式中的任何方式对孔进行化学修饰。优选地使用共价衔接子能够与靶分析物相互作用的共价衔接子对孔进行修饰,如上所讨论的。

[0669] 所述方法优选地用于表征靶多核苷酸并且步骤(a)包括:使所述靶多核苷酸与所述孔和多核苷酸结合蛋白接触,并且所述多核苷酸结合蛋白控制所述靶多核苷酸移动穿过所述孔。多核苷酸结合蛋白可以是能够结合到多核苷酸并且控制其移动穿过孔的任何蛋白质。在所属领域中确定多核苷酸结合蛋白是否结合到多核苷酸是直接了当的。多核苷酸结合蛋白通常与多核苷酸相互作用并且修改多核苷酸的至少一个性质。多核苷酸结合蛋白可以通过切割多核苷酸以形成单独的核苷酸或如二或三核苷酸等较短核苷酸链来修饰多核苷酸。所述部分可以通过对多核苷酸进行定向或将其移动到特定位置、即控制其移动来修饰多核苷酸。

[0670] 多核苷酸结合蛋白优选地是多核苷酸处理酶。多核苷酸处理酶是能够与多核苷酸相互作用并且修改其至少一个性质的多肽。所述酶可以通过切割多核苷酸以形成单独的核苷酸或如二或三核苷酸等较短核苷酸链来修饰多核苷酸。所述酶可以通过对多核苷酸进行定向或将其移动到特定位置来修饰多核苷酸。多核苷酸结合蛋白通常包括多核苷酸结合结构域和催化结构域。多核苷酸处理酶不需要显示酶活性,只要其能够结合靶序列并且控制其移动穿过孔即可。例如,酶可以被修饰成移除其酶活性,或可以在防止其充当酶的条件下使用。下文更详细地讨论了这种条件。

[0671] 多核苷酸处理酶优选地衍生自溶核酶。在酶的构建体中使用的多核苷酸处理酶更优选地衍生自以下酶分类(EC)组中的任何组的成员:3.1.11、3.1.13、3.1.14、3.1.15、3.1.16、3.1.21、3.1.22、3.1.25、3.1.26、3.1.27、3.1.30和3.1.31。所述酶可以是国际申请号PCT/GB10/000133(公开为W0 2010/086603)中公开的酶中的任何酶。

[0672] 优选的酶是聚合酶、核酸外切酶、解旋酶和拓扑异构酶,如旋转酶。适当的酶包含但不限于来自大肠杆菌的核酸外切酶I(SEQ ID NO:6)、来自大肠杆菌的核酸外切酶III(SEQ ID NO:8)、来自嗜热菌的RecJ(SEQ ID NO:10)和噬菌体λ核酸外切酶(SEQ ID NO:12)和其变体。包括SEQ ID NO:10中所示序列的三个亚基或其变体相互作用以形成三聚体核酸外切酶。酶可以是Phi29 DNA聚合酶(SEQ ID NO:4)或其变体。酶可以是解旋酶或衍生自解旋酶。典型的解旋酶是Hel308、RecD或XPD,例如,Hel308 Mbu(SEQ ID NO:13)或其变

异体。

[0673] 酶最优选地衍生自解旋酶,如Hel308解旋酶、如TraI解旋酶或TrwC解旋酶等RecD解旋酶、XPD解旋酶或Dda解旋酶。解旋酶可以是在以下国际申请号中公开的解旋酶、经修饰解旋酶或解旋酶构建体中的任一者:PCT/GB2012/052579(公开为W0 2013/057495);PCT/GB2012/053274(公开为W0 2013/098562);PCT/GB2012/053273(公开为W02013098561);PCT/GB2013/051925(公开为W0 2014/013260);PCT/GB2013/051924(公开为W0 2014/013259);PCT/GB2013/051928(公开为W0 2014/013262)以及PCT/GB2014/052736。

[0674] 解旋酶优选地包括SEQ ID NO:18(Dda)中所示序列或其变异体。变异体可以以下文关于跨膜孔所讨论的方式中的任何方式不同于原生序列。SEQ ID NO:18的优选变异体包括:(a)E94C和A360C;或(b)E94C、A360C、C109A和C136A,并且然后任选地(Δ M1)G1G2(即缺失M1并且然后添加G1和G2)。

[0675] SEQ ID NO:4、6、8、10、12、13或18的变异体是氨基酸序列不同于SEQ ID NO:4、6、8、10、12、13或18的氨基酸序列且保留多核苷酸结合能力的酶。变异体可以包含促进多核苷酸的结合和/或促进多核苷酸在高盐浓度和/或室温下的活性的修饰。

[0676] 在SEQ ID NO:4、6、8、10、12、13或18的氨基酸序列的整个长度内,基于氨基酸同一性,变异体将优选地与所述序列至少50%同源。更优选地,基于氨基酸同一性,变异体多肽可以在整个序列内与SEQ ID NO:4、6、8、10、12、13或18的氨基酸序列至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,并且更优选地至少95%、97%或99%同源。在200个或更多个、例如230个、250个、270个或280个或更多个的连续氨基酸的延伸段内,可能存在至少80%、例如至少85%、90%或95%的氨基酸同一性(“硬同源性”)。如上文所述那样确定同源性。变异体可以以上文参考SEQ ID NO:2所讨论的方式中的任何方式不同于野生型序列。酶可以共价连接到孔,如上文所讨论的。

[0677] 存在两种使用纳米孔对多核苷酸进行测序的主要策略,即链测序和核酸外切酶测序。本发明的方法可以涉及链测序或核酸外切酶测序。

[0678] 在链测序中,通过所施加的电势或抵抗所施加的电势,DNA转移通过纳米孔。可以在孔的顺侧上使用渐进地或逐渐地对双链DNA起作用的核酸外切酶以在所施加的电势下馈送剩余单链或在反侧上在相反电势下馈送。同样,使双链DNA解旋的解旋酶也可以以类似的方式使用。还可以使用聚合酶。还存在需要抵抗所施加的电势的链转移的测序应用的可能性,但是DNA必须首先在相反或无电势下由酶“捕获”。在电势然后在结合之后切换回来的情况下,链将以顺式到反式的方式穿过孔并且通过电流保持处于延长的构形。单链DNA核酸外切酶或单链DNA依赖性聚合酶可充当分子马达以抵抗所施加的电势以逐步受控方式从反式到顺式将最近转移的单链拉回通过孔。

[0679] 在一个实施例中,表征靶多核苷酸的方法涉及使靶序列与孔和解旋酶接触。所述方法中可以使用任何解旋酶。解旋酶可以相对于孔在两种模式下起作用。首先,优选地使用解旋酶执行所述方法,使得所述解旋酶使用由所施加电压产生的场控制靶序列移动穿过孔。在这种模式下,DNA的5'末端首先在孔中被捕获,并且酶控制DNA移动到孔中,使得使用所述场使靶序列穿过孔,直到其最终转移通过到达双层的反侧。可替代地,所述方法优选地被执行为使得解旋酶抵抗由所施加电压产生的场而控制靶序列移动穿过孔。在这种模式下,DNA的3'末端首先在孔中被捕获,并且酶控制DNA移动穿过孔,使得靶序列抵抗所施加的

场而被拉出孔,直到最终推回到双层的顺侧。

[0680] 在核酸外切酶测序中,核酸外切酶从靶多核苷酸的一个末端释放单独的核苷酸,并且这些单独的核苷酸如下文所讨论的那样识别。在另一个实施例中,表征靶多核苷酸的方法涉及使靶序列与孔和核酸外切酶接触。在所述方法中可以使用以上讨论的核酸外切酶中的任何核酸外切酶。酶可以共价连接到孔,如上文所讨论的。

[0681] 核酸外切酶是通常锁到多核苷酸的一个末端且从那个末端一次一个核苷酸地消化序列的酶。核酸外切酶可以在5'到3'的方向或3'到5'的方向上消化多核苷酸。通常通过选择所属领域中使用的酶和/或使用所属领域中已知的方法来确定与核酸外切酶结合到的多核苷酸末端。多核苷酸的任一末端处的羟基或帽结构通常可以用于防止或促进核酸外切酶结合到多核苷酸的特定末端。

[0682] 所述方法涉及使多核苷酸与核酸外切酶接触,使得核苷酸以允许表征或识别一定比例的核苷酸的速率从多核苷酸的末端被消化,如上文所讨论的。用于进行这种操作的方法在所属领域中是众所周知的。例如,埃德曼降解法(Edman degradation)用于从多肽的末端连续地消化单一氨基酸,使得可使用高效液相色谱法(HPLC)识别所述氨基酸。在本发明中可以使用同源方法。

[0683] 核酸外切酶起作用的速率通常比野生型核酸外切酶的最优速率慢。在本发明的方法中的核酸外切酶的活性的适当速率包括以下消化速率:0.5个到1000个核苷酸每秒、0.6个到500个核苷酸每秒、0.7个到200个核苷酸每秒、0.8个到100个核苷酸每秒、0.9个到50个核苷酸每秒或1个到20个或10个核苷酸每秒。所述速率优选地是1个、10个、100个、500个或1000个核苷酸每秒。核酸外切酶活性的适当速率可以以不同方式实现。例如,可以根据本发明使用具有降低的最优活性速率的变异体核酸外切酶。

[0684] 本发明的方法涉及测量如靶多核苷酸等靶分析物的一个或多个特性。所述方法可以涉及测量如靶多核苷酸等靶分析物的两个、三个、四个或五个或更多个特性。对于靶多核苷酸,所述一个或多个特性优选地选自:(i)靶多核苷酸的长度;(ii)靶多核苷酸的一性;(iii)靶多核苷酸的序列;(iv)靶多核苷酸的二级结构;以及(v)靶多核苷酸是否被修饰。可以根据本发明测量(i)到(v)的任何组合。

[0685] 对于(i),可以使用靶多核苷酸与孔之间的相互作用的数目来测量多核苷酸的长度。

[0686] 对于(ii),可以以多种方式测量多核苷酸的一性。可以结合测量靶多核苷酸的序列或在不测量靶核苷酸的序列的情况下来测量多核苷酸的一性。前者是直接了当的;多核苷酸被测序并且由此被识别。可以以若干方式进行后者。例如,可以测量在多核苷酸中存在特定基序(而不测量多核苷酸的剩余序列)。可替代地,所述方法中的特定电信号和/或光信号的测量可以将靶多核苷酸识别为来自特定来源。

[0687] 对于(iii),可以如先前所描述的那样确定多核苷酸的序列。在Stoddart D等人,《美国国家科学院院刊(Proc Natl Acad Sci)》,2012年,第106卷,第19期,第7702到7707页;Lieberman KR等人,《美国化学会志(J Am Chem Soc)》,2010年,第132卷,第50期,第17961到17972页;以及国际申请W0 2000/28312中描述了适当的测序方法,尤其是使用电测量的测序方法。

[0688] 对于(iv),可以以多种方式测量二级结构。例如,如果所述方法涉及电测量,则可

以使用停留时间的变化或流过孔的电流的变化来测量二级结构。这允许区别单链多核苷酸和双链多核苷酸的区域。

[0689] 对于(v),可以测量是否存在任何修饰。所述方法优选地包括:使用一个或多个蛋白质或使用一个或多个标记、标签或间隔子确定靶多核苷酸是否通过甲基化、通过氧化、通过损坏修饰。特定修饰将引起与孔的特定相互作用,所述特定相互作用可以使用以下所描述的方法进行测量。例如,可以在孔与每个核苷酸相互作用期间流过孔的电流的基础上区别甲基胞嘧啶和胞嘧啶。

[0690] 本发明还提供了一种估计靶多核苷酸的序列的方法。本发明还提供了一种对靶多核苷酸进行测序的方法。

[0691] 可以进行各种不同类型的测量。这包括但不限于:电测量和光学测量。可能的电测量包含:电流测量、阻抗测量、隧道测量(Ivanov AP等人,《纳米快报(Nano Lett.)》,2011年1月12日,第11卷,第1期,第279到285页)以及FET测量(国际

[0692] 申请W0 2005/124888)。《美国化学会志(J. Am. Chem. Soc.)》,2009年,第131卷,第1652和1653页公开了涉及荧光测量的适当光学方法。可以将光学测量与点测量组合(Soni GV等人,《科学仪器综述(Rev Sci Instrum)》,2010年1月,第81卷,第1期,第014301页)。测量可以是跨膜电流测量,如对流过孔的离子电流的测量。

[0693] 可以使用如以下文献中描述的标准单通道记录设备来进行电测量:

[0694] Stoddart D等人,《美国国家科学院院刊(Proc Natl Acad Sci)》,2012年,第106卷,第19期,第7702到7707页;Lieberman KR等人,《美国化学会志(J Am Chem Soc)》,2010年,第132卷,第50期,第17961到17972页;以及国际申请

[0695] W0-2000/28312。可替代地,可使用例如如以下文献中描述的多通道系统来进行电测量:

[0696] 国际申请W0-2009/077734和国际申请W0-2011/067559。

[0697] 在优选实施例中,所述方法包括:

[0698] (a)使靶多核苷酸与本发明的孔和多核苷酸结合蛋白接触,使得所述靶多核苷酸移动穿过所述孔,并且所述结合蛋白控制所述靶多核苷酸移动穿过所述孔;以及

[0699] (b)测量所述多核苷酸相对于所述孔移动时通过所述孔的电流,其中所述电流指示所述靶多核苷酸的一个或多个特性,并且由此表征所述靶多核苷酸。

[0700] 可以使用适合于研究孔插入到膜中的膜/孔系统的任何设备来执行所述方法。可以使用适合于跨膜孔感测的任何设备来执行所述方法。例如,所述设备包括包含水溶液的室以及将所述室分成两个区段的屏障。所述屏障具有开孔,在开孔中形成含有孔的膜。

[0701] 可使用在国际申请号PCT/GB08/000562(W0 2008/102120)中描述的设备执行所述方法。

[0702] 所述方法可以涉及:测量如靶多核苷酸等分析物相对于孔移动时通过孔的电流。因此,所述设备还可以包括能够施加电势并且测量跨膜和孔的电信号的电路。可以使用贴片钳或电压钳来执行所述方法。所述方法优选地涉及使用电压钳。

[0703] 本发明的方法可以涉及:测量如靶多核苷酸等分析物相对于孔移动时通过孔的电流。用于测量通过跨膜蛋白孔的离子电流的适当条件在所属领域中是已知的且公开于实例中。通常使用跨膜和孔施加的电压来执行所述方法。所使用的电压通常为+2V到-2V,通常

为-400mV到+400mV。所使用的电压优选地处于具有下限和上限的范围内,所述下限选自-400mV、-300mV、-200mV、-150mV、-100mV、-50mV、-20mV和0mV,并且所述上限独立地选自+10mV、+20mV、+50mV、+100mV、+150mV、+200mV、+300mV和+400mV。所使用的电压更优选地处于100mV到240mV的范围内,并且最优选处于120mV到220mV的范围内。可以通过使用增大的所施加电势来增加通过孔区别不同的核苷酸。

[0704] 通常在存在任何电荷载流子的情况下执行所述方法,所述电荷载流子如金属盐,例如碱金属盐;卤盐,例如如碱金属氯化物盐等氯化物盐。电荷载流子可以包含离子液体或有机盐,例如四甲基氯化铵、三甲基苯基氯化铵、苯基三甲基氯化铵或1-乙基-3-甲基氯化咪唑鎓。在以上讨论的示例性设备中,盐存在于腔室中的水溶液中。通常使用氯化钾(KCl)、氯化钠(NaCl)或氯化铯(CsCl)。KCl是优选的。盐浓度可以是饱和的。盐浓度可以是3M或更低,并且通常为0.1M到2.5M、0.3M到1.9M、0.5M到1.8M、0.7M到1.7M、0.9M到1.6M或1M到1.4M。盐浓度优选地是150mM到1M。优选地使用至少0.3M的盐浓度来执行所述方法,如至少0.4M、至少0.5M、至少0.6M、至少0.8M、至少1.0M、至少1.5M、至少2.0M、至少2.5M或至少3.0M。高盐浓度提供高信噪比并且允许在正常电流波动的背景下识别指示核苷酸存在的电流。

[0705] 通常在存在缓冲液的情况下执行所述方法。在以上讨论的示例性设备中,缓冲液存在于腔室中的水溶液中。在本发明的方法中可以使用任何缓冲液。通常,缓冲液是HEPES。另一种适当的缓冲液是Tris-HCl缓冲液。通常在以下的pH下执行所述方法:4.0到12.0、4.5到10.0、5.0到9.0、5.5到8.8、6.0到8.7或7.0到8.8或7.5到8.5。所使用的pH优选地为约7.5。

[0706] 可以在以下温度下执行所述方法:0°C到100°C、15°C到95°C、16°C到90°C、17°C到85°C、18°C到80°C、19°C到70°C或20°C到60°C。通常在室温下执行所述方法。任选地在支持酶功能的温度、如约37°C下执行所述方法。

[0707] 通常在存在游离核苷酸或游离核苷酸类似物以及促进如解旋酶或核酸外切酶等多核苷酸结合蛋白的作用的酶辅因子的情况下执行所述方法。游离核苷酸可以是上文所讨论的单独核苷酸中的任何核苷酸中的一个或多个。游离核苷酸包含但不限于:单磷酸腺苷(AMP)、二磷酸腺苷(ADP)、三磷酸腺苷(ATP)、单磷酸鸟苷(GMP)、二磷酸鸟苷(GDP)、三磷酸鸟苷(GTP)、单磷酸胸苷(TMP)、二磷酸胸苷(TDP)、三磷酸胸苷(TTP)、单磷酸尿苷(UMP)、二磷酸尿苷(UDP)、三磷酸尿苷(UTP)、单磷酸胞苷(CMP)、二磷酸胞苷(CDP)、三磷酸胞苷(CTP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、单磷酸脱氧腺苷(dAMP)、二磷酸脱氧腺苷(dADP)、三磷酸脱氧腺苷(dATP)、单磷酸脱氧鸟苷(dGMP)、二磷酸脱氧鸟苷(dGDP)、三磷酸脱氧鸟苷(dGTP)、单磷酸脱氧胸苷(dTMP)、二磷酸脱氧胸苷(dTDP)、三磷酸脱氧胸苷(dTTP)、单磷酸脱氧尿苷(dUMP)、二磷酸脱氧尿苷(dUDP)、三磷酸脱氧尿苷(dUTP)、单磷酸脱氧胞苷(dCMP)、二磷酸脱氧胞苷(dCDP)和三磷酸脱氧胞苷(dCTP)。游离核苷酸优选地选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP或dCMP。游离核苷酸优选地是三磷酸腺苷(ATP)。酶辅因子是允许解旋酶起作用的因子。酶辅因子优选地是二价金属阳离子。二价金属阳离子优选地是 Mg^{2+} 、 Mn^{2+} 、 Ca^{2+} 或 Co^{2+} 。酶辅因子最优选地是 Mg^{2+} 。

[0708] 可按任何顺序使靶多核苷酸与孔和多核苷酸结合蛋白接触。优选的是,在使靶多核苷酸与多核苷酸结合蛋白和孔接触时,靶多核苷酸首先与多核苷酸结合蛋白形成复合

体。当跨孔施加电压时,靶多核苷酸/蛋白质复合体则与孔形成复合体并且控制多核苷酸移动穿过孔。

[0709] 识别单独核苷酸的方法

[0710] 本发明还提供了一种表征单独核苷酸的方法。换言之,靶分析物是单独核苷酸。所述方法包括:使核苷酸与本发明的孔接触,使得核苷酸与孔相互作用;以及测量在相互作用期间通过孔的电流并且由此表征核苷酸。因此,本发明涉及对单独核苷酸的纳米孔感测。本发明还提供了一种用于识别单独核苷酸的方法,其包括:测量在相互作用期间通过孔的电流并且由此确定核苷酸的同源性。可以使用上文讨论的孔中的任何孔。优选地使用分子衔接子对所述孔进行化学修饰,如以上所讨论的。

[0711] 如果电流以对核苷酸具有特异性的方式流过孔(即,如果检测到与分析物相关的独特电流流过孔),则存在核苷酸。如果电流并不以对核苷酸具有特异性的方式流过孔,则不存在核苷酸。

[0712] 本发明可以用于在相同结构的核苷酸对通过孔的电流具有的不同影响的基础上区分所述核苷酸。可以根据核苷酸与孔相互作用时的核苷酸电流振幅在单分子水平下识别单独的核苷酸。本发明还可以用于确定样品中是否存在特定核苷酸。本发明还可以用于测量样品中的特定核苷酸的浓度。

[0713] 孔通常存在于膜中。可以使用上述任何适当的膜/孔系统执行所述方法。

[0714] 单独的核苷酸是单个核苷酸。单独的核苷酸是不通过核苷酸键结合到另一核苷酸或多核苷酸的核苷酸。核苷酸键涉及结合到另一核苷酸的糖基团的核苷酸的磷酸基之一。单独的核苷酸通常是通过核苷酸键结合到由至少5个、至少10个、至少20个、至少50个、至少100个、至少200个、至少500个、至少1000个或至少5000个核苷酸构成的另一多核苷酸的核苷酸。例如,已从如DNA或RNA链等靶分析物多核苷酸序列消化单独的核苷酸。本发明的方法可以用于识别任何核苷酸。核苷酸可以是上文讨论的核苷酸中的任何核苷酸。

[0715] 核苷酸可以衍生自对如核糖核酸(RNA)或脱氧核糖核酸(DNA)等核酸序列的消化。可以使用所属领域中已知的任何方法消化核酸序列。适当的方法包含但不限于使用酶或催化剂的方法。在Deck等人,《无机化学(Inorg.Chem.)》,2002年,第41卷,第669到677页中公开了核酸的催化消化。

[0716] 可以使来自单个多核苷酸的单独核苷酸以顺序方式与孔接触,以便对多核苷酸的全部或部分进行测序。上文更详细地讨论了对多核苷酸进行测序。

[0717] 可以使核苷酸在膜的两侧与孔接触。可以在膜的两侧将核苷酸引入到孔。可以使核苷酸可以与膜的侧面接触,所述侧面允许核苷酸穿过孔到达膜的另一侧。例如,使核苷酸与孔的末端接触,所述末端在其原生环境中允许离子或如核苷酸等小分子进入孔的桶或通道中,使得核苷酸可以穿过孔。在这种情况下,核苷酸在跨膜穿过孔的桶或通道时与孔和/或衔接子相互作用。可替代地,核苷酸可以与膜的侧面接触,所述侧面允许核苷酸通过或结合衔接子与孔相互作用、与孔中解离并保持处于膜的同侧。本发明提供了孔,其中衔接子的位置是固定的。因此,核苷酸优选与孔的末端接触,所述末端允许衔接子与核苷酸相互作用。

[0718] 核苷酸可以任何方式且在任何位点处与孔相互作用。如以上所讨论的,核苷酸优选地通过或结合衔接子可逆地结合到孔。核苷酸最优选地在其跨膜穿过孔时通过或结合衔

接子可逆地结合到孔。核苷酸还可以在其跨膜穿过孔时通过或结合衔接子可逆地结合到孔的桶或通道。

[0719] 在核苷酸与孔之间相互作用期间,核苷酸以对所述核苷酸具有特异性的方式影响流过孔的电流。例如,特定核苷酸将在特定平均时间段内且在特定程度上减少流过孔的电流。换言之,流过孔的电流对于特定核苷酸来说是独特的。可以执行对照实验以确定特定核苷酸对流过孔的电流具有的影响。然后可以将对测试样品执行本发明的方法产生的结果与源自于这种对照实验的结果进行比较以识别样品中的特定核苷酸或确定样品中是否存在特定核苷酸。流过孔的电流以指示特定核苷酸的方式被影响的频率可以用于确定所述核苷酸在样品中的浓度。还可以计算样品内不同核苷酸的比率。例如,可以计算dCMP与甲基-dCMP的比率。

[0720] 所述方法可以涉及使用上文讨论的任何设备、样品或条件。

[0721] 形成传感器的方法

[0722] 本发明还提供一种形成用于表征靶多核苷酸的传感器的方法。所述方法包括:在本发明的孔与如解旋酶或核酸外切酶等多核苷酸结合蛋白之间形成复合体。可以通过在存在靶多核苷酸的情况下使孔和蛋白质接触并且然后跨孔施加电势来形成复合体。所施加的电势可以是如上所述的化学电势或电压电势。可替代地,可以通过将孔共价连接到蛋白质来形成复合体。用于共价连接的方法在所属领域中是已知的且例如公开于国际申请号PCT/GB09/001679(公开为WO 2010/004265)和PCT/GB10/000133(公开为WO 2010/086603)中。复合体是用于表征靶多核苷酸的传感器。所述方法优选地包括:在本发明的孔与解旋酶之间形成复合体。上文所讨论的实施例中的任何实施例同样适用于这种方法。

[0723] 本发明还提供了一种用于表征靶多核苷酸的传感器。传感器包括本发明的孔与多核苷酸结合蛋白之间的复合体。上文所讨论的实施例中的任何实施例同样适用于本发明的传感器。

[0724] 试剂盒

[0725] 本发明还提供了一种用于表征靶多核苷酸、如对其进行测序的试剂盒。试剂盒包括(a)本发明的孔以及(b)膜。试剂盒优选地进一步包括如解旋酶或核酸外切酶等多核苷酸结合蛋白。上文所讨论的实施例中的任何实施例同样适用于本发明的试剂盒。

[0726] 本发明的试剂盒可以另外包括使上文提到的实施例中的任何实施例能够被执行的一种或多种其它试剂或仪器。这种试剂或仪器包含以下各项中的一项或多项:一种或多种适当的缓冲液(水溶液)、用于从受试者获得样品的装置(如包括针的容器或仪器)、用于扩增和/或表达多核苷酸序列的装置、如上文定义的膜或者电压钳或贴片钳设备。试剂可以以干态存在于试剂盒中,使得流体样品使试剂再悬浮。试剂盒还可以任选地包括使试剂盒能够在本发明的方法使用的说明书或关于所述方法可以用于哪些患者的详情。试剂盒可以任选地包括核苷酸。

[0727] 设备

[0728] 本发明还提供了一种用于表征样品中的靶多核苷酸、如对其进行测序的装置。所述设备可以包括(a)多个本发明的孔以及(b)多个多核苷酸结合蛋白,如解旋酶或核酸外切酶。所述设备可以是用于分析物分析的任何常规设备,如阵列或芯片。

[0729] 所述阵列或芯片通常含有如嵌段共聚物膜等膜的多个井,所述井各自插入有单个

纳米孔。所述阵列可以集成在电子芯片内。

[0730] 所述设备优选地包括：

[0731] 传感器装置，其能够支撑多个孔并且能够操作以使用孔和蛋白质来执行多核苷酸表征或测序；

[0732] -至少一个储槽，其用于保持用于执行表征或测序的材料；

[0733] -流体学系统，其被配置成可控地将材料从所述至少一个储槽供应到所述传感器装置；以及

[0734] -多个容器，其用于容纳相应样品，所述流体学系统被配置成选择性地将所述样品从所述容器供应到所述传感器装置。

[0735] 所述设备可以是在国际申请号PCT/GB10/000789(公开为W0 2010/122293)、国际申请号PCT/GB10/002206(公开为W0 2011/067559)或国际申请号PCT/US99/25679(公开为W0 00/28312)中描述的设备中的任何设备。

[0736] 以下实例说明本发明。

[0737] 实例1

[0738] 此实例描述了如何使用解旋酶-T4Dda-E94C/C109A/C136A/A360C(具有突变E94C/C109A/C136A/A360C的SEQ ID NO:18)来控制DNA移动穿过多个不同的突变胞溶素纳米孔。在DNA转移穿过纳米孔时，所有测试的纳米孔都展现出电流变化。所测试的突变纳米孔展现出：1)增加的范围；2)降低的噪声；3)提高的信噪比；4)与突变对照纳米孔相比的增加的捕获；或5)与基线相比的改变的读头大小。

[0739] 材料和方法

[0740] DNA构建体制备

[0741] • 将70uL的T4 Dda-E94C/C109A/C136A/A360C缓冲液交换(使用Zeba柱)到具有2mM EDTA的70uL 1x KOAc缓冲液中。

[0742] • 将70uL的T4 Dda-E94C/C109A/C136A/A360C缓冲液交换混合物加入到70uL的2uM DNA衔接子中(有关序列的详细信息，请参见图5)。然后将样品混合并在室温下温育5分钟。

[0743] • 加入1uL的140mM TMAD，并且将样品混合并在室温下温育60分钟。此样品被称为样品A。然后取出2ul等分试样用于安捷伦(Agilent)分析。

[0744] HS/ATP步骤

[0745] • 将下表中的试剂混合并在室温下温育25分钟。此样品被称为样品B。

[0746]

试剂	体积	最终
样品A(500nM)	139	220nM
2x Hs缓冲液(100mM Hepes, 2M KCl, pH8)	150	1x
600mM MgCl ₂	7	14mM
100mM rATP	4.2	14mM
最终	300.2	

[0747] SPRI纯化

[0748] • 向样品B中加入1.1mL SPRI珠粒，并且然后将样品混合并温育5分钟。

[0749] • 沉淀珠粒并移除上清液。然后使用50mM Tris.HCl, 2.5M NaCl, 20% PEG8000洗

涤珠粒。

[0750] • 在70uL 10mM Tris.HCl,20mM NaCl中将样品C洗脱。

[0751] 使用酶将10kbλC连接到衔接子

[0752] • 在热循环仪中在20℃下将下表中的试剂温育10分钟。

		体积 (μL)	
[0753]	nH ₂ O	310.2	
	10 kb λ C DNA (SEQ ID NO: 23, 168.7 nM)	14.82	5 nM
	样品 C (500 nM)	25	25 nM
	连接缓冲液 (5x)	100	1x
	NEBNext 快速 T4 DNA 连接酶 (NEBNext Quick T4 DNA Ligase) (2000 U.μl ⁻¹)	50	5%

[0754]	总计	500	
--------	----	-----	--

[0755] • 然后对反应混合物(1x 500μl等分试样)进行如下处理:使用200μl的20% SPRI珠粒进行SPRI纯化;在750μl洗涤缓冲液1中洗涤;并且在125μl洗脱缓冲液1中洗脱。将最终的DNA序列(SEQ ID NO:24)与DNA杂交。此样品被称为样品D。

[0756] 连接缓冲液(5x)的组分

[0757]	试剂	体积	最终
	1M Tris.HCl pH8	15	150mM
	1M MgCl ₂	5	50mM
	100mM ATP	5	5mM
	40%PEG 8000	75	30%
	总计	100uL	

[0758] 洗涤缓冲液1的组分

[0759]	试剂	体积	最终
	水	1100	
	1M Tris.HCl pH8	100	50mM
	5M NaCl	300	750mM
	40%PEG 8000	500	10%
	总计	2000uL	

[0760] 洗脱缓冲液1的组分

[0761]	试剂	体积	最终
	水	906.7	高达1000uL
	0.5M CAPS pH10	80	40mM
	3M KCl	13.3	40mM
	总计	1000uL	

[0762] 电生理学实验

[0763] 从插入在缓冲液(25mM磷酸钾缓冲液,150mM亚铁氰化钾(II),150mM铁氰化钾(III),pH 8.0)中的含嵌段共聚物中的的单一胞溶素纳米孔获取电测量。在达到插入在嵌

段共聚物中的单个孔之后,然后使缓冲液(2mL,25mM磷酸钾缓冲液,150mM亚铁氰化钾(II),150mM铁氰化钾(III),pH 8.0)流过系统以移除任何过量的胞溶素纳米孔。然后使150 μ L的500mM KCl、25mM磷酸钾、pH8.0流过系统。10分钟后,使150 μ L的500mM KCl,25mM磷酸钾,pH8.0流过系统,并且然后使T4 Dda-E94C/C109A/C136A/A360C、DNA、燃料(MgCl₂,ATP)预混合物(总计150 μ L,样品D)流入单纳米孔实验系统中。在180mV下进行实验,并且监测解旋酶控制的DNA移动。

[0764] 结果

[0765] 研究了多个不同的纳米孔以确定突变对跨膜孔区域的影响。以下列出了所研究的突变孔连同与它们进行比较的基线纳米孔(基线孔1到4)。研究了多个不同的参数以识别改进的纳米孔:1)信号的平均噪声(其中噪声等于链中所有事件的标准偏差,在所有链上计算),在改进的纳米孔中,所述平均噪声将低于基线;2)平均电流范围,其是信号内电流电平的范围的度量,并且在改进的纳米孔中,其将高于基线;3)表中引用的平均信噪比是所有链中的信噪比(平均电流范围除以信号的平均噪声),并且在改进的纳米孔中,其将高于基线;4)DNA的捕获率,在改进的纳米孔中,其将高于基线;以及5)读头大小,在改进的纳米孔中,其可以根据基线的读头的大小而增大或减小。

[0766] 以下每个表包含对应基线纳米孔的相关数据。表6=突变体1,表7=突变体2,表8=突变体3,并且表9=突变体10,然后将其与突变孔进行比较。

[0767] 胞溶素突变体1=胞溶素-(E84Q/E85K/E92Q/E97S/D126G)9(具有突变E84Q/E85K/E92Q/E97S/D126G的SEQ ID NO:2)。(基线1)

[0768] 胞溶素突变体2=胞溶素-(E84Q/E85K/E92Q/E94D/E97S/D126G)9(具有突变E84Q/E85K/E92Q/E94D/E97S/D126G的SEQ ID NO:2)。(基线2)

[0769] 胞溶素突变体3=胞溶素-(E84Q/E85K/E92Q/E94Q/E97S/D126G)9(具有突变E84Q/E85K/E92Q/E94Q/E97S/D126G的SEQ ID NO:2)。(基线3)

[0770] 胞溶素突变体4=胞溶素-(E84Q/E85K/S89Q/E92Q/E97S/D126G)9(具有突变E84Q/E85K/S89Q/E92Q/E97S/D126G的SEQ ID NO:2)。

[0771] 胞溶素突变体5=胞溶素-(E84Q/E85K/T91S/E92Q/E97S/D126G)9(具有突变E84Q/E85K/T91S/E92Q/E97S/D126G的SEQ ID NO:2)。

[0772] 胞溶素突变体6=胞溶素-(E84Q/E85K/E92Q/E97S/S98Q/D126G)9(具有突变E84Q/E85K/E92Q/E97S/S98Q/D126G的SEQ ID NO:2)。

[0773] 胞溶素突变体7=胞溶素-(E84Q/E85K/E92Q/E97S/V100S/D126G)9(具有突变E84Q/E85K/E92Q/E97S/V100S/D126G的SEQ ID NO:2)。

[0774] 胞溶素突变体8=胞溶素-(E84Q/E85K/E92Q/E94D/E97S/S80K/D126G)9(具有突变E84Q/E85K/E92Q/E94D/E97S/S80K/D126G的SEQ ID NO:2)。

[0775] 胞溶素突变体9=胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T106R/D126G)9(具有突变E84Q/E85K/E92Q/E94D/E97S/T106R/D126G的SEQ ID NO:2)。

[0776] 胞溶素突变体10=胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T106K/D126G)9(具有突变E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。(基线4)

[0777] 胞溶素突变体11=胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T104R/D126G)9(具有突变E84Q/E85K/E92Q/E94D/E97S/T104R/D126G的SEQ ID NO:2)。

- [0778] 胞溶素突变体12=胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T104K/D126G)9(具有突变E84Q/E85K/E92Q/E94D/E97S/T104K/D126G的SEQ ID NO:2)。
- [0779] 胞溶素突变体13=胞溶素-(S78N/E84Q/E85K/E92Q/E94D/E97S/D126G)9(具有突变S78N/E84Q/E85K/E92Q/E94D/E97S/D126G的SEQ ID NO:2)。
- [0780] 胞溶素突变体14=胞溶素-(S82N/E84Q/E85K/E92Q/E94D/E97S/D126G)9(具有突变S82N/E84Q/E85K/E92Q/E94D/E97S/D126G的SEQ ID NO:2)。
- [0781] 胞溶素突变体15=胞溶素-(E76N/E84Q/E85K/E92Q/E94Q/E97S/D126G)9(具有突变E76N/E84Q/E85K/E92Q/E94Q/E97S/D126G的SEQ ID NO:2)。
- [0782] 胞溶素突变体16=胞溶素-(E76S/E84Q/E85K/E92Q/E94Q/E97S/D126G)9(具有突变E76S/E84Q/E85K/E92Q/E94Q/E97S/D126G的SEQ ID NO:2)。
- [0783] 胞溶素突变体17=胞溶素-(E84Q/E85K/E92Q/E94Q/Y96D/D97S/T106K/D126G)9(具有突变E84Q/E85K/E92Q/E94Q/Y96D/D97S/T106K/D126G的SEQ ID NO:2)。
- [0784] 胞溶素突变体18=胞溶素-(K45D/E84Q/E85K/E92Q/E94K/D97S/T106K/D126G)9(具有突变K45D/E84Q/E85K/E92Q/E94K/D97S/T106K/D126G的SEQ ID NO:2)。
- [0785] 胞溶素突变体19=胞溶素-(K45R/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G)9(具有突变K45R/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G的SEQ ID NO:2)。
- [0786] 胞溶素突变体20=胞溶素-(D35N/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G)9(具有突变D35N/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G的SEQ ID NO:2)。
- [0787] 胞溶素突变体21=胞溶素-(K37N/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G)9(具有突变K37N/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G的SEQ ID NO:2)。
- [0788] 胞溶素突变体22=胞溶素-(K37S/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G)9(具有突变K37S/E84Q/E85K/E92Q/E94D/D97S/T106K/D126G的SEQ ID NO:2)。
- [0789] 胞溶素突变体23=胞溶素-(E84Q/E85K/E92D/E94Q/D97S/T106K/D126G)9(具有突变E84Q/E85K/E92D/E94Q/D97S/T106K/D126G的SEQ ID NO:2)。
- [0790] 胞溶素突变体24=胞溶素-(E84Q/E85K/E92E/E94Q/D97S/T106K/D126G)9(具有突变E84Q/E85K/E92E/E94Q/D97S/T106K/D126G的SEQ ID NO:2)。
- [0791] 胞溶素突变体25=胞溶素-(K37S/E84Q/E85K/E92Q/E94D/D97S/T104K/T106K/D126G)9(具有突变K37S/E84Q/E85K/E92Q/E94D/D97S/T104K/T106K/D126G的SEQ ID NO:2)。
- [0792] 胞溶素突变体26=胞溶素-(E84Q/E85K/M90I/E92Q/E94D/E97S/T106K/D126G)9(具有突变E84Q/E85K/M90I/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。
- [0793] 胞溶素突变体27=胞溶素-(K45T/V47K/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G)9(具有突变K45T/V47K/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。
- [0794] 胞溶素突变体28=胞溶素-(T51K/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G)9(具有突变T51K/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。
- [0795] 胞溶素突变体29=胞溶素-(K45Y/S49K/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G)9(具有突变K45Y/S49K/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。
- [0796] 胞溶素突变体30=胞溶素-(S49L/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G)9(具有突变S49L/E84Q/E85K/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。

[0797] 胞溶素突变体31=胞溶素-(E84Q/E85K/V88I/M90A/E92Q/E94D/E97S/T106K/D126G)9(具有突变E84Q/E85K/V88I/M90A/E92Q/E94D/E97S/T106K/D126G的SEQ ID NO:2)。

[0798] 胞溶素突变体32=胞溶素-(K45N/S49K/E84Q/E85K/E92D/E94N/E97S/T106K/D126G)9(具有突变K45N/S49K/E84Q/E85K/E92D/E94N/E97S/T106K/D126G的SEQ ID NO:2)。

[0799] 胞溶素突变体33=胞溶素-(K45N/V47K/E84Q/E85K/E92D/E94N/E97S/T106K/D126G)9(具有突变K45N/V47K/E84Q/E85K/E92D/E94N/E97S/T106K/D126G的SEQ ID NO:2)。

突变号	与基线纳米孔的差异	范围(pA)	噪声 (pA)	信噪比	优点和观察
1	基线 1	11.1	1.56	7.12	
2	E94D	15.7	2.03	7.73	在孔中具有或不具有 DNA 的情况下,信噪比增加且通道门控减少
3	E94Q	23.6	3.55	6.65	在孔中具有或不具有 DNA 的情况下,范围加倍、通道门控减少
4	S89Q	11.96	1.38	8.67	较低噪声
5	T91S	12.21	1.31	9.32	较低噪声
6	S98Q	10.63	1.27	8.37	较低噪声
7	V100S	12.58	1.5	8.39	范围略有增大

[0801] 表6

突变号	与基线纳米孔的差异	范围(pA)	噪声 (pA)	信噪比	优点和观察
2	基线 2	15.7	2.03	7.73	
8	S80K	15.29	2.9	5.27	捕获率略微提高
9	T106R	15.83	2.55	6.21	捕获率大幅提高
10	T106K	15.73	1.99	7.90	捕获率大幅提高
11	T104R	17.36	3.59	4.84	捕获率大幅提高
12	T104K	15.55	2.57	6.05	捕获率大幅提高
13	S78N	14.54	1.77	8.21	噪声减少
14	S82N	15.03	1.81	8.30	噪声减少

[0803] 表7

突变号	与基线纳米孔的差异	范围(pA)	噪声 (pA)	信噪比	优点和观察
3	基线 3	23.6	3.55	6.65	
15	E76N	16.99	2.3	7.39	噪声大幅降低
16	E76S	18.35	2.38	7.71	噪声大幅降低

[0805] 表8

突变号	与基线纳米孔的差异	范围 (pA)	噪声 (pA)	信噪比	优点和观察
10	基线 4	13.48	1.35	9.99	
17	E94Q/Y96D	16.53	1.07	15.45	范围增大且噪声降低、读

					头大小减小	
18	K45D/E94K	14.11	1.05	13.44	噪声降低	
19	K45R	11.09	1.09	10.17	噪声降低	
20	D35N	15.91	1.68	9.47	范围增大	
21	K37N	14.39	1.28	11.24	范围增大且噪声降低	
22	K37S	14.47	1.28	11.30	范围增大且噪声降低	
23	E92D/E94Q	20.59	2.05	10.04	范围增大	
24	E92E/E94Q	21.45	1.83	11.72	范围增大	
[0807]	25	K37S/T104K	13.96	1.24	11.26	噪声降低
	26	M90I	14.08	1.14	12.35	噪声降低
	27	K45T/V47K	15.4	1.46	10.55	范围增大
	28	T51K	19.33	1.89	10.23	范围增大
	29	K45Y/S49K	20.69	1.71	12.10	范围增大
	30	S49L	12.51	1.1	11.37	噪声降低
	31	V88I/M90A	13.08	1.17	11.18	噪声降低
	32	K45N/S49K/E92D/E94N	15.84	1.44	11.00	范围增大
	33	K45N/V47K/E92D/E94N	12.31	1.13	10.89	噪声降低

[0808] 表9

[0809] 读头分析

[0810] 对于胞溶素突变体1和10,我们获得了所有可能的9聚体多核苷酸的预期离子电流分布的模型。所述模型可以包括每个9聚体的当前分布的平均和标准偏差。

[0811] 我们检查并比较了针对胞溶素突变体1和10获得的模型的结构。附图(见图1和图2)提供了这种比较的实例。在每个模型(即胞溶素1或10)的情况下,我们将形式为A, $x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$ 的所有9聚体的分布的均值组合,此处 x_i 表示选自{A,C,G,T}的任意多核苷酸,所述组合应用于取中位数的平均值。针对位置1中的所有核苷酸{A,C,G,T}并且针对所有位置重复这种中值平均,使得当核苷酸存在于9聚体的9个位置中的任何位置中时,我们获得36个对每个核苷酸的中值效应进行编码的中值。

[0812] 图1(胞溶素突变体1)和图2(胞溶素突变体2)标绘了两个不同孔的这些中值。图1和图2中的绘图示出了读头中的每个位置处的所有碱基之间的区分程度。区分越大,所述特定位置处的电流贡献水平之间的差异越大。如果某个位置不是读头的一部分,则所述位置处的电流贡献对于所有四个碱基来说将是相似的。图2(胞溶素突变体10)示出了读头的位置6到8处的所有四个碱基的类似电流贡献。图1(胞溶素突变体1)未示出读头中的任何位置处的所有四个碱基的类似电流贡献。因此,胞溶素突变体10的读头比比胞溶素突变体1的读头更短。读头较短可能是有利的,因为在任何一个时间较少的碱基对信号有贡献,这可以导致碱基调用准确度提高。

[0813] 实例2

[0814] 此实例描述了用于产生具有减小直径的桶/通道的经化学修饰的组装孔的方案。

[0815] 首先将单体胞溶素样品(约10 μ mol)还原以确保半胱氨酸残基的最大反应性,并且因此确保高效偶联反应。将单体胞溶素样品(约10 μ mol)与1mM二硫苏糖醇(DTT)一起温育5到15分钟。然后通过20,000rpm下离心10分钟沉淀细胞碎片和悬浮的聚集体。然后回收可溶性部分,并且使用截留分子量为7Kd的Zeba旋转柱(ThermoFisher)将其缓冲液交换为1mM Tris,1mM EDTA,pH8.0。

[0816] 将待附接的分子(例如:2-碘-N-(2,2,2-三氟乙基)乙酰胺)在适当的溶剂、通常为DMSO中溶解至100mM的浓度。将其加入缓冲液交换的胞溶素单体样品中至最终浓度为1mM。将所得溶液在30℃下温育2小时。然后通过加入来自Encapsula Nanosciences的20uL 5脂质混合物(磷脂酰丝氨酸(0.325mg/ml):POPE(0.55mg/ml):胆固醇(0.45mg/ml):Soy PC(0.9mg/ml):鞘磷脂(0.275mg/ml))使经修饰样品(100uL)寡聚。将样品在30℃下温育60分钟。然后使样品经受SDS-PAGE并如国际申请号PCT/GB2013/050667(公开为W02013/153359)中所述那样从凝胶中纯化。

[0817] 实例3

[0818] 此实例比较了具有减小直径的桶/通道的经化学修饰的组装胞溶素孔(胞溶素-具有通过E94C附接的2-碘-N-(2,2,2-三氟乙基)乙酰胺的(E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)与胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85K/E92Q/E94D/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)。

[0819] 材料和方法

[0820] 如实例1中所述那样制备的DNA构建体。如实例1中所述那样执行电生理学实验。

[0821] 结果

[0822] 电生理学实验表明,经化学修饰的组装孔(胞溶素-具有通过E94C附接的2-碘-N-(2,2,2-三氟乙基)乙酰胺的(E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)显示出21pA的中值范围,其大于显示出12pA的中值范围的胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T106K/D126G/C272A/C283A)9。这种中值范围增加为k聚体的分辨提供了更大的电流空间。

[0823] 图3(胞溶素-(E84Q/E85K/E92Q/E94D/E97S/T106K/D126G/C272A/C283A)9)和图4((胞溶素-具有通过E94C附接的2-碘-N-(2,2,2-三氟乙基)乙酰胺的(E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85K/E92Q/E94C/E97S/T106K/D126G/C272A/C283A)的SEQ ID NO:2)示出了如实例1中所述的中位数的绘图。在将图4与图3进行比较时,对不同位置处的不同碱基的信号相对贡献已经改变,图4中的极端读头位置(位置7到8)显示出少得多的区分,这意味着它们对信号的贡献大大减少,并且因此在给定时刻测定的K聚体的长度更短。这种较短的读头可能是有利的,因为在任何一个时间较少的碱基对信号有贡献,这可以导致碱基调用准确度提高。

[0824] 对胞溶素-具有通过E92C附接的2-碘-N-(2-苄基乙基)乙酰胺的(E84Q/E85S/E92C/E94D/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85S/E92C/E94D/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)和胞溶素-具有通过E92C附接的1-苄基-2,5-二氢-1H-吡咯-2,5-二酮的(E84Q/E85S/E92C/E94D/E97S/T106K/D126G/C272A/C283A)9(具有突变E84Q/E85S/E92C/E94D/E97S/T106K/D126G/C272A/C283A的SEQ ID NO:2)执行与实例3中所述的实验类似的实验。

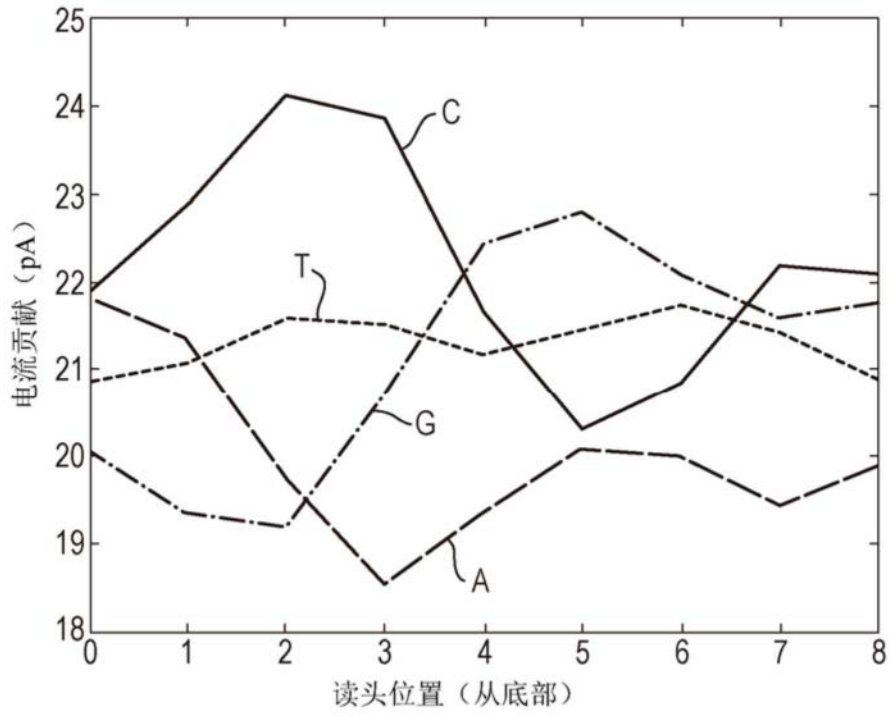


图1

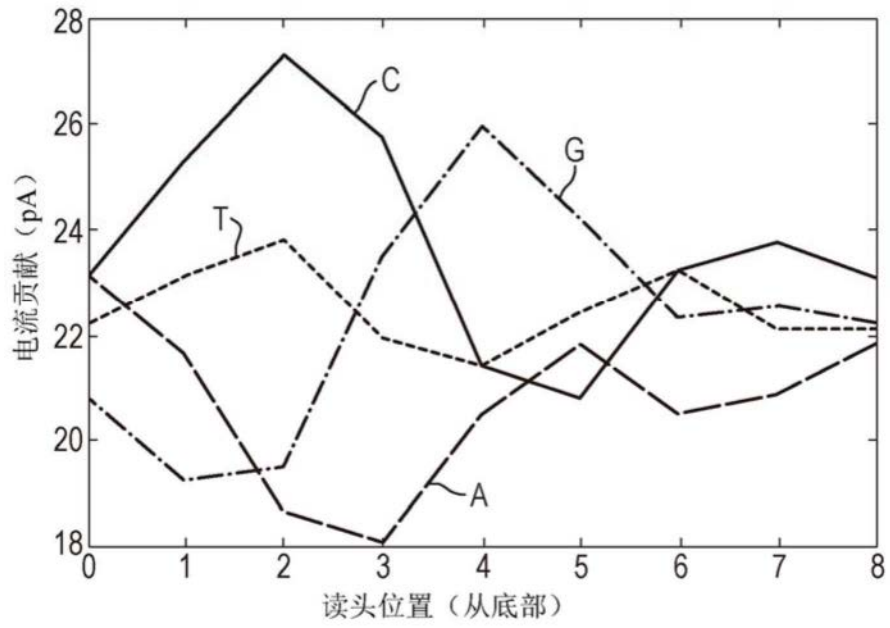


图2

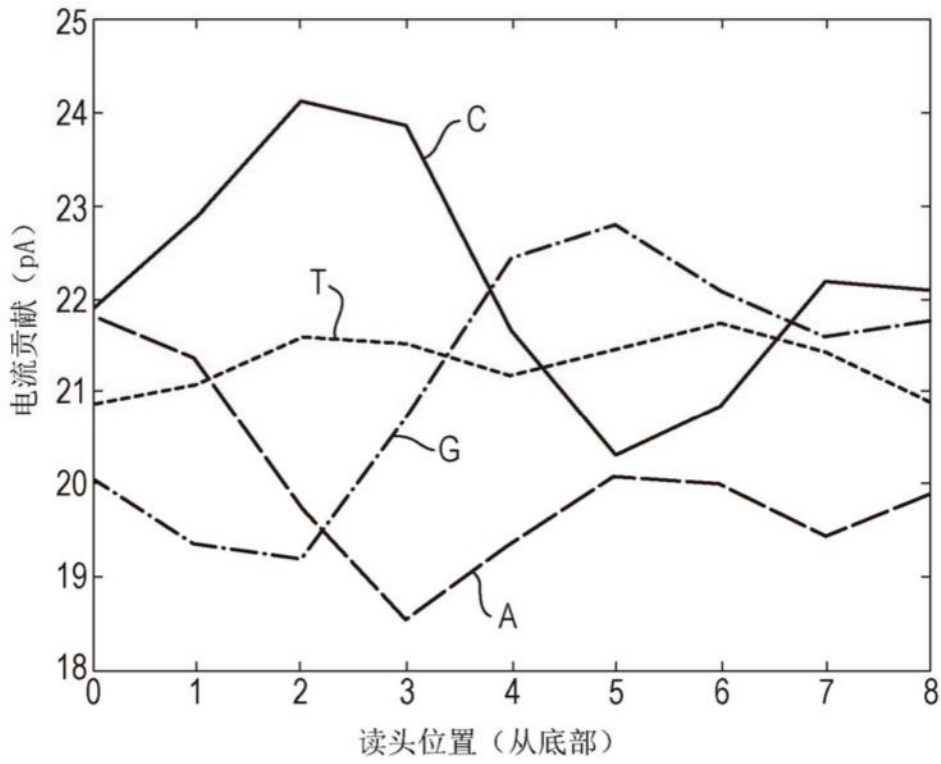


图3

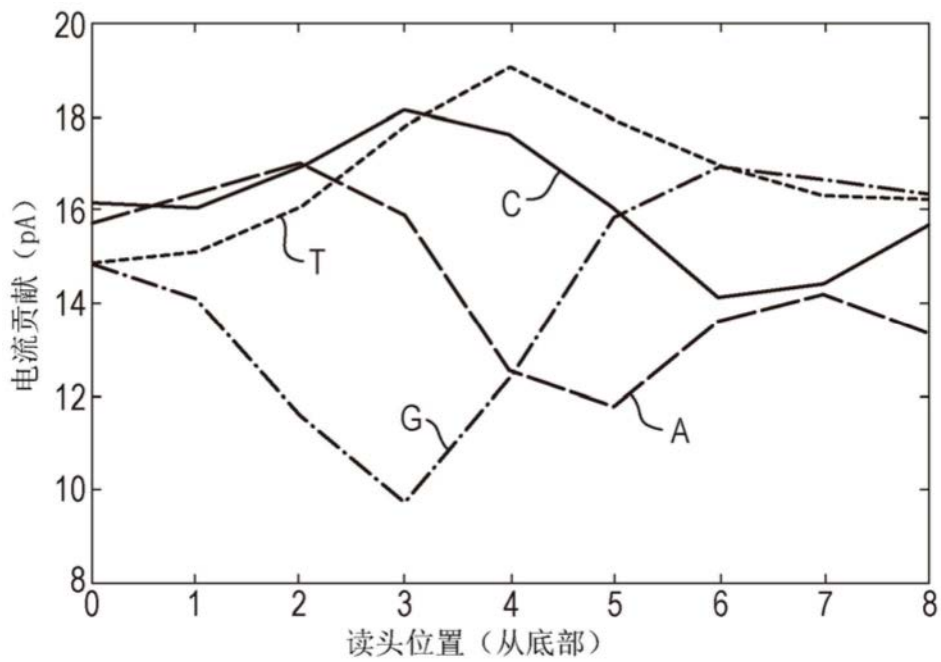


图4

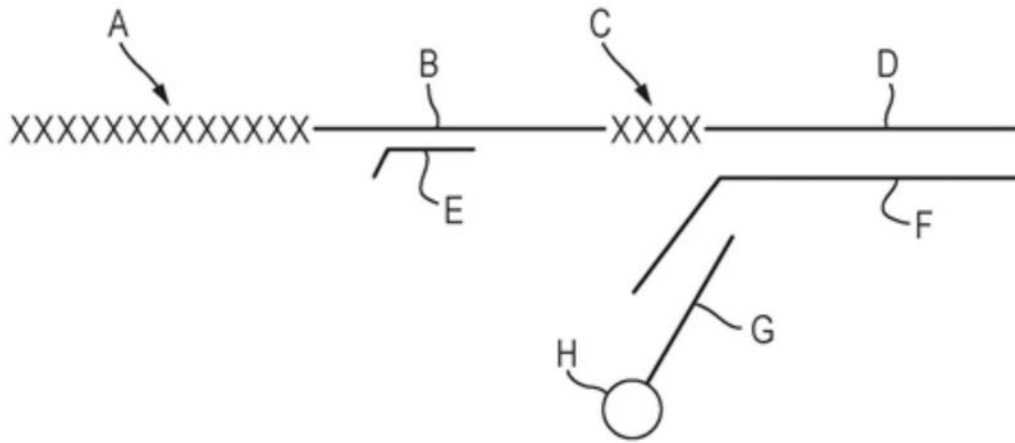


图5



图6

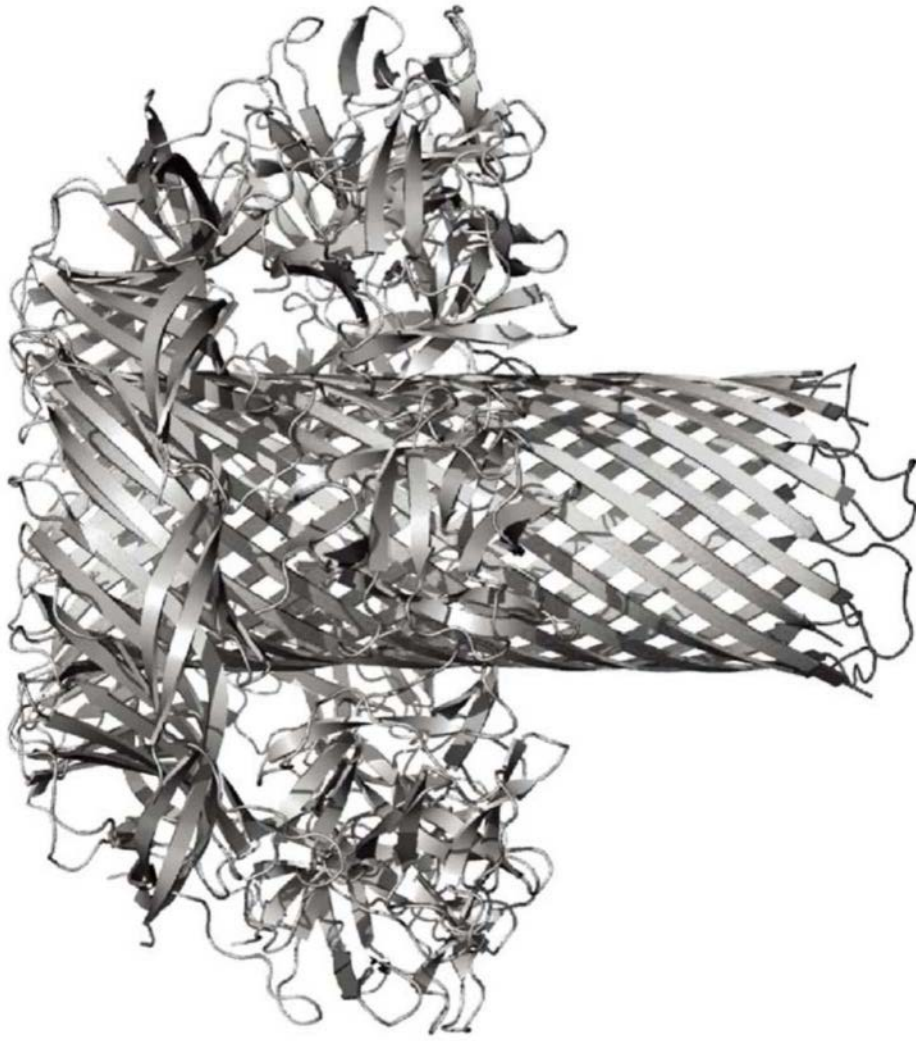


图7A

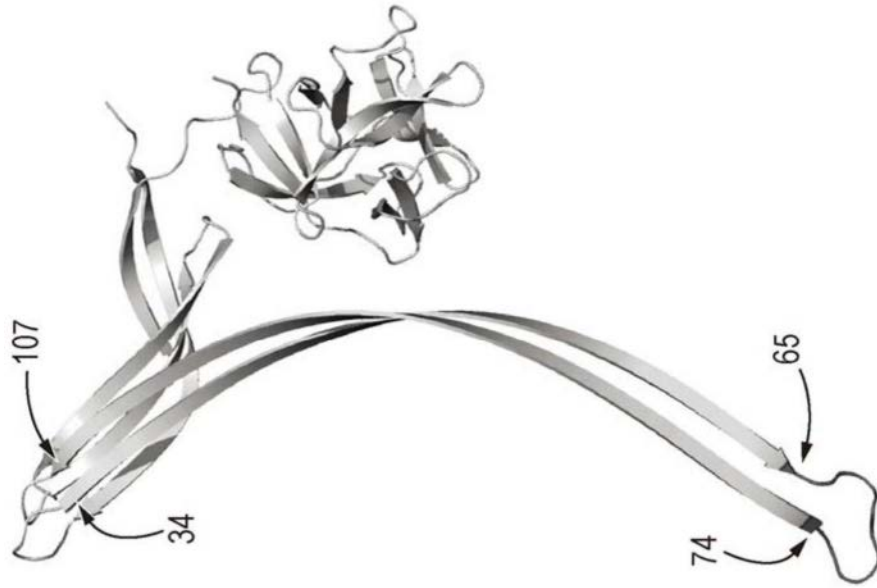


图7B

胞溶素	1M--SAKA--AEGYEQLEVDVVAWVWKEGYNWENRGSTSVDOKAIITKGMKNVNSETRVVTATHSIGSTISTGDAPFELGSWEVSYSHSHEESQVSMTEVEYESKWIE	102
LRP2	1M--SSRAGIAEGYEQLEVDVVAWVWKEGYNWENRGSTSVDOKAIITKGMKNVNSETRVVTATHSIGSTISTGDAPFELGSWEVSYSHSHEESQVSMTEVEYESKEIE	104
LRP3	1MSATAVT--ADGLEEELEVDVVAWVWKEGYNWENRGDTSVEOKIATMTKGMKNVNSETRVVTATHSIGSTISTGDAPFELGSWEVSYSHSHEESQVSMTEVEYESQVIE	104
LRP1	1MSSSTVM--ADGFEELEVDVVAWVWKEGYNWENRGNSVQOKIATMTKGMKNVNSETRVVTATHSIGSTISTGDAPFELGSWEVSYSHSHEESQVSMTEVEYESQVIE	104
胞溶素	103HTITIPPTSKFTRWQLNADVGGADIEYMYLIDEVTPIGGTQSIPOVITSPRAKIIIVGPDZIIILGKTEIRIKHAERREYMTWWSRKSMPAATLGHSHKLFKFLVIEDWGG	208
LRP2	105HTITIPPTSKFTRWQLNADVGGADIEYMYLIDEVTPIGGTLSLPQVIRAKIILVGFELVGEFTEIRIKHADRKEFMTWWSRKSMPAATLGHSHKLYKFLVIEDMYG	210
LRP3	105HTVITIPPTSKFTRWKLNAVGGTDEYMYLIDEVTPISVQTLPOVIRAKIILVGRQTHLGTAVR IKHAERQRYMTVIERKKWPAATLGSNLFKFLVIEDSSG	210
LRP1	105HTVITIPENKKFTRWKLNAVGGTGLEMYLIDEVTAIGADLTLPQVIRAKIILVGRQTHLGEFTEIRIKHAERREYMTVIERKKWPAATLGSNLFKFLVIEDSSG	210
胞溶素	209ERIKTLNTWYSGYEVAYSSDQGGYFDQGTQNPQRWAINKSIDPLRHGDVVTENKRYFTRSGLCYDDGPPATNVCYLDKREDKWIILEVVG--	297
LRP2	211ERIKTLNTWYSGYEVAYSSDQGGYFDQGSQNPQRWAINKSIDPLRHGDVVTENKRYFTRSGLCYDGGPPATNVCYLDKREDKWIILEVW--KP	300
LRP3	211TRIKTLNTWYPGYLEWYSSDQGGYFDESSQNPQRWALSADPLRHGDVVTENKRYFETNSGLCYDDGPPATNVCYLDKREDKWIILEVW--NP	300
LRP1	211IRIKTLNTWYPGYLEWYSSDQGGYFDESSQNPQRWALSAMPPLRHGDVVTENRNNFETNSGMCYDDGPPATNVCYCLEKREDKWIILEVW--NT	300

图8