

孔蛋白单体的突变体、蛋白孔及其应用

技术领域

5 本发明属于靶分析物特性的表征技术领域，特别涉及一种孔蛋白单体的突变体、包含其的蛋白孔以及其检测靶分析物的应用。

背景技术

10 随着对核酸结构和序列的研究，核酸测序技术不断发展，成为生命科学研究的核心理领域，对生物、化学、电学、生命科学、医学等领域的技术发展起到巨大的推动作用。利用纳米孔研究出新型的快速、准确、低成本、高精度及高通量的核酸测序技术是后人类基因组计划的热点之一。

15 纳米孔（Nanopore）测序技术，又被称为第四代测序技术，是一种以单链核酸分子作为测序单元，利用一个能够提供离子电流通道的纳米孔，使得单链核酸分子在电泳驱动下通过该纳米孔，当核酸通过纳米孔时，会减少纳米孔的电流，对产生的不同信号实时读取序列信息的基因测序技术。

20 纳米孔测序主要特点是：读长很长，准确率较高，错误区大都发生在均聚寡核苷酸区域。纳米孔测序不但可实现天然 DNA 和 RNA 测序还可直接获取 DNA 和 RNA 的碱基修饰信息，例如它能够直接读取甲基化的胞嘧啶，而不必像二代测序方法那样需要事先对基因组进行重亚硫酸盐（bisulfite）处理，这对于在基因组水平直接研究表观遗传相关现象有极大推动。纳米孔检测技术作为一个新型平台，具有低成本、高通量、非标记等优势。

25

30 纳米孔分析技术起源于 Coulter 计数器的发明以及单通道电流的记录技术。生理与医学诺贝尔奖获得者 Neher 和 Sakamann 在 1976 年利用膜片钳技术测量膜电势，研究膜蛋白及离子通道，推动了纳米

孔测序技术的实际应用进程。1996 年，Kasianowicz 等提出了利用 α -溶血素对 DNA 测序的新设想，是生物纳米孔单分子测序的里程碑标志。随后，MspA 孔蛋白、噬菌体 Phi29 连接器等生物纳米孔的研究报道，丰富了纳米孔分析技术的研究。Li 等在 2001 年开启了固态纳米孔研究的新时代。受限于半导体和材料工业的发展，固态纳米孔测序进展缓慢。

纳米孔测序技术的关键点之一在于所设计的一种特殊生物纳米孔，孔内缢缩区形成的读取头结构可在当单链核酸（例如 ssDNA）分子通过纳米孔时，造成孔道电流的阻塞，从而短暂地影响流过纳米孔的电流强度（每种碱基所影响的电流变化幅度是不同的），最后高灵敏度的电子设备检测到这些变化从而鉴定所通过的碱基。目前采用蛋白孔作为纳米孔进行测序，孔蛋白主要以大肠杆菌为来源。

目前纳米孔蛋白单一，需要开发替代的纳米孔蛋白实现纳米孔测序技术。孔蛋白也与测序精度密切相关，而且孔蛋白还涉及与控速蛋白的相互作用的模式变化，进一步优化孔蛋白与控速蛋白相互作用界面的稳定性，对提高测序数据的一致性和稳定性有积极影响。纳米孔测序技术的准确率也有待改善，因此，需要开发改进的纳米孔蛋白，以进一步提高纳米孔测序的分辨率。

发明内容

为解决上述问题，本发明实施例的目的在于提供一种替代的孔蛋白单体的突变体、包含其的蛋白孔、及其应用。

第一方面，本发明实施例提供了一种孔蛋白单体的突变体，其中所述孔蛋白单体的突变体的氨基酸包括 SEQ ID NO:1 所示的序列或与其具有至少 99%、98%、97%、96%、95%、90%、80%、70%、60%或 50%同一性的序列，或由其组成，并且所述孔蛋白单体的突变体的氨基酸包括在对应 SEQ ID NO:1 的 T71、S75、F76 中的一个或多个位置处

的突变；

T71、S75、F76 中的一个或多个具体为，T71、S75、F76、T71 和 S75、S75 和 F76、T71 和 S76、或 T71 和 S75 和 F76。

5 优选地，所述孔蛋白单体的突变体的氨基酸包括在对应 SEQ ID NO:1 的 62-175、62-104、68-175、64-79、71-76、或 69-76 的一个或多个位置处的突变。

优选地，所述孔蛋白单体的突变体的氨基酸包括：

10 (1) 对应 SEQ ID NO:1 的 K69、P70、T71、P72、A73、S74、S75、和 F76 的一个或多个位置处具有氨基酸的插入、缺失和/或替换；(2) 对应 SEQ ID NO:1 的 Q64、T65、G66、Q67、Y68、K69、P70、T71、P72、A73、S74、S75、F76、S77、T78、和 S79 的一个或多个位置处具有氨基酸的插入、缺失和/或替换；(3) 对应 SEQ ID NO:1 的 R62、
15 D63、Y68、K69、T71、S75、F76、和 E104 的一个或多个位置处具有氨基酸的插入、缺失和/或替换；或者(4) 对应 SEQ ID NO:1 的 Y68、K69、P70、T71、P72、A73、S74、S75、F76、E171 和 D175 的一个或多个位置处具有氨基酸的插入、缺失和/或替换。

20 在一个实施例中，所述孔蛋白单体的突变体的氨基酸突变选自以下：

(a) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 $M_1M_2M_3M_4M_5M_6M_7M_8$ ，其中 M_1 选自 R、K、H 中的 0 至 3 种； M_2 选自 P 中的 0 至 1 种； M_3 选自 S、G、C、U、T、M、A、V、L、I 中的 0 至 10 种； M_4 选自 P 中的 0 至 1 种； M_5 选自 A、G、V、L、I 中的 0 至 5 种； M_6 选自 S、C、U、T、M 中的 0 至 5 种； M_7 选自 A、T、G、V、L、I、S、C、U、M 中的 0 至 10 种； M_8 选自 Q、D、E、N、K、H、R 中的 0 至 7 种；

25 (b) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 $M_9M_{10}M_{11}M_{12}M_{13}M_{14}M_{15}M_{16}$
30

$M_{17}M_{18}M_{19}M_{20}M_{21}M_{22}M_{23}M_{24}$, 其中 M_9 选自 L、G、A、V、I 中的 0 至 5 种; M_{10} 选自 T、S、C、U、M 中的 0 至 5 种; M_{11} 选自 G、A、V、L、I 中的 0 至 5 种; M_{12} 选自 Q、D、E、N 中的 0 至 4 种; M_{13} 选自 Y、F、W 中的 0 至 3 种; M_{14} 选自 R、H、K 中的 0 至 3 种; M_{15} 选自 P 中的 0 至 1 种; M_{16} 选自 S、C、U、T、M 中的 0 至 5 种; M_{17} 选自 P 中的 0 至 1 种; M_{18} 选自 A、G、V、L、I 中的 0 至 5 种; M_{19} 选自 S、C、U、T、M 中的 0 至 5 种; M_{20} 选自 A、G、V、L、I 中的 0 至 5 种; M_{21} 选自 N、D、E、Q、L、G、A、V、I 中的 0 至 9 种; M_{22} 选自 S、C、U、T、M 中的 0 至 5 种; M_{23} 选自 T、S、C、U、M 中的 0 至 5 种; M_{24} 选自 A、G、V、L、I 中的 0 至 5 种;

(c) 对应 SEQ ID NO:1 的 62 位突变为 S、C、U、T、M 中的 0 至 5 种; 63 位突变为 V、G、A、L、I 中的 0 至 5 种; 68 位突变为 F、W 中的 0 至 2 种; 69 位突变为 R、H 中的 0 至 2 种; 71 位突变为 S、C、U、M 中的 0 至 4 种; 75 位突变为 A、G、V、L、I 中的 0 至 5 种; 76 位突变为 Q、D、E、N 中的 0 至 4 种; 104 位突变为 V、G、A、L、I 中的 0 至 5 种; 和

(d) 对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 $M_{25}M_{26}M_{27}M_{28}M_{29}M_{30}M_{31}M_{32}M_{33}$, 并且 171 位突变为 N、D、Q 中的 0 至 3 种; 175 位突变为 N、E、Q 中的 0 至 3 种, 其中 M_{25} 选自 F、Y、W 中的 0 至 3 种; M_{26} 选自 R、H、K 中的 0 至 3 种; M_{27} 选自 P 中的 0 至 1 种; M_{28} 选自 S、C、U、T、M 中的 0 至 5 种; M_{29} 选自 P 中的 0 至 1 种; M_{30} 选自 A、G、V、L、I 中的 0 至 5 种; M_{31} 选自 S、C、U、T、M 中的 0 至 5 种; M_{32} 选自 A、G、V、L、I 中的 0 至 5 种; M_{33} 选自 Q、D、E、N 中的 0 至 4 种。

25

在一个实施例中, 孔蛋白单体的突变体的氨基酸突变选自以下:

(a) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 $M_1M_2M_3M_4M_5M_6M_7M_8$, 其中 M_1 选自 R、K 或 H; M_2 选自 P; M_3 选自 S、G、C、U、T、M、A、V、L 或 I; M_4 选自 P; M_5 选自 A、G、V、L 或 I; M_6 选自 S、C、U、T 或 M; M_7 选自 A、T、G、V、

30

L、I、S、C、U或M；M₈选自Q、D、E、N、K、H或R；

5 (b) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 M₉M₁₀M₁₁M₁₂M₁₃M₁₄M₁₅M₁₆M₁₇M₁₈M₁₉M₂₀M₂₁M₂₂M₂₃M₂₄，其中 M₉选自 L、G、A、V 或 I；M₁₀选自 T、S、C、U 或 M；M₁₁选自 G、A、V、L 或 I；M₁₂选自 Q、D、E 或 N；M₁₃选自 Y、F 或 W；M₁₄选自 R、H 或 K；M₁₅选自 P；M₁₆选自 S、C、U、T 或 M；M₁₇选自 P；M₁₈选自 A、G、V、L 或 I；M₁₉选自 S、C、U、T 或 M；M₂₀选自 A、G、V、L 或 I；M₂₁选自 N、D、E、Q、L、G、A、V 或 I；M₂₂选自 S、C、U、T 或 M；M₂₃选自 T、S、C、U 或 M；M₂₄选自 A、G、V、L 或 I；

10 (c) 对应 SEQ ID NO:1 的 R62 突变为 R62S、R62C、R62U、R62T 或 R62M；D63 突变为 D63V、D63G、D63A、D63L 或 D63I；Y68 突变为 Y68F 或 Y68W；K69 突变为 K69R 或 K69H；T71 突变为 T71S、T71C、T71U 或 T71M；S75 突变为 S75A、S75G、S75V、S75L 或 S75I；F76 突变为 F76Q、F76D、F76E 或 F76N；E104 突变为 E104V、E104G、E104A、E104L 或 E104I；和

15 (d) 对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 M₂₅M₂₆M₂₇M₂₈M₂₉M₃₀M₃₁M₃₂M₃₃，并且 E171 突变为 E171N、E171D 或 E171Q；D175 突变为 D175N、D175E 或 D175Q，其中 M₂₅选自 F、Y 或 W；M₂₆选自 R、H 或 K；M₂₇选自 P；M₂₈选自 S、C、U、T 或 M；M₂₉选自 P；M₃₀选自 A、G、V、L 或 I；M₃₁选自 S、C、U、T 或 M；M₃₂选自 A、G、V、L 或 I；M₃₃选自 Q、D、E、或 N。

在一个实施例中，孔蛋白单体的突变体的氨基酸突变选自以下：

25 (a) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 RPSPASAQ；

(b) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 KPGPASTK；

30 (c) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 LTGQYRPSPASANSTA；

(d) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 LTGQYRSPASALSTA;

(e) 对应 SEQ ID NO:1 的 R62S、D63V、Y68F、K69R、T71S、S75A、F76Q、和 E104V; 和

5 (f) 对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 FRPSPASAQ, 以及对应 SEQ ID NO:1 的 E171N 和 D175N。

第二方面, 本发明实施例提供了一种蛋白孔, 包括至少一个孔蛋白单体的突变体。

10

第三方面, 本发明实施例提供了一种用于表征靶分析物的复合物, 其特征在于: 所述的蛋白孔及与其结合的控速蛋白。

第四方面, 本发明实施例提供了编码孔蛋白单体的突变体、蛋白孔、或复合物的核酸。

15

第五方面, 本发明实施例提供了包含所述核酸的载体或遗传工程化的宿主细胞。

第六方面, 本发明实施例提供了孔蛋白单体的突变体、其蛋白孔、复合物、核酸、载体或宿主细胞在检测靶分析物存在、不存在或一个或多个特征或制备检测靶分析物存在、不存在或一个或多个特征的产品中的应用。

20

第七方面, 本发明实施例提供了一种产生蛋白孔或其多肽的方法, 包括用所述的载体转化所述的宿主细胞, 诱导所述宿主细胞表达所述的蛋白孔或其多肽。

25

第八方面, 本发明实施例提供了一种用于确定靶分析物存在、不存在或一个或多个特征的方法, 包括:

30

a. 使靶分析物与蛋白孔、复合物、或复合物中的蛋白孔接触，使得所述靶分析物相对于所述蛋白孔移动；以及

b. 在所述靶分析物相对于所述蛋白孔移动时获取一个或多个测量值，从而确定所述靶分析物的存在、不存在或一个或多个特征。

5

在一个实施例中，所述方法包括：所述靶分析物与存在于膜中的所述蛋白孔相互作用从而使得所述靶分析物相对所述蛋白孔移动。

在一个实施例中，靶分析物是核酸分子。

10

在一个实施例中，用于确定靶分析物存在、不存在或一个或多个特征的方法包括将所述靶分析物偶联到膜上；和所述靶分析物与存在于所述膜中的所述蛋白孔相互作用从而使得所述靶分析物相对所述蛋白孔移动。

15

第九方面，本发明实施例提供了一种用于确定靶分析物存在、不存在或一个或多个特征的试剂盒，包括所述的孔蛋白单体的突变体、所述的蛋白孔、所述的复合物、所述的核酸、或所述的载体或宿主，和所述的膜的组分。

20

第十方面，本发明实施例提供了一种用于确定靶分析物存在、不存在或一个或多个特征的装置，包括所述的蛋白孔或所述复合物，和所述的膜。

25

在一个实施例中，所述靶分析物包括多糖、金属离子、无机盐、聚合物、氨基酸、肽、蛋白、核苷酸、寡核苷酸、多核苷酸、染料、药物、诊断剂、爆炸物或环境污染物；

优选地，所述靶分析物包括多核苷酸，

更优选地，所述多核苷酸包括 DNA 或 RNA；和/或，所述一个或多个特征选自 (i) 所述多核苷酸的长度；(ii) 所述多核苷酸的一致性；

30

(iii) 所述多核苷酸的序列；(iv) 所述多核苷酸的二级结构和(v) 所述多核苷酸是否经修饰；和/或，所述复合物中所述控速蛋白包括多核苷酸结合蛋白。

5 附图说明

所描述的附图仅是示意性的而非限制性的。

图 1 示出了根据一个实施例的纳米孔的基本工作原理。

图 2 示出了根据一个实施例的 DNA 测序的示意图。

10 图 3 示出了根据一个实施例核苷酸穿过蛋白孔时相应的堵孔信号。

图 4A、4B 和 4C 示出了根据一个实施例的野生型蛋白孔通道表面结构和飘带图模型。图 4A 为表面结构模型侧视图，图 4B 为飘带结构模型俯视图，及图 4C 为飘带结构模型。

15 图 5A 和图 5B 示出了根据一个实施例的突变孔 1 的氨基酸模型图，图 5A 为顶视图，图 5B 为侧视图。

图 6 示出了根据一个实施例的突变孔 1 各部分尺寸信息。

图 7 示出了根据一个实施例的孔蛋白突变体 1 的单体氨基酸模型图，(b) 是 (a) 放大显示的缢缩区 (eye loop) 的核心氨基酸组成。

图 8 示出了根据一个实施例的孔蛋白突变体 1 的负染色电镜结果。

20 图 9A 示出了根据一个实施例的孔蛋白突变体 1 的冷冻电镜照片图 9B 示出了 2D 分类结果。

图 10A 和图 10B 示出了根据一个实施例的孔蛋白突变体 1 的局部精修傅里叶壳相关系数(FSC)结果，a: rIn FSC 未屏蔽图；b: rIn FSC 相位随机屏蔽图；c: rIn FSC 修正图；d: rIn FSC 屏蔽图。

25 图 11 示出了根据一个实施例的孔蛋白突变体 1 冷冻电镜三维重构 2.2 Å 分辨率的电子密度图。

图 12 示出了根据一个实施例的孔蛋白突变体 1 在 2.2 Å 下的电子密度映射图。

图 13 示出了根据一个实施例的 DNA 构建体 BS7-4C3-SE1 的结构。

30 图 14 示出了根据一个实施例的 DNA 构建体 BS7-4C3-PLT 的结构。

图 15A 示出了根据一个实施例的突变孔 1 在 $\pm 180\text{mV}$ 电压下开孔电流及其门控特征。

图 15B 示出了根据一个实施例的突变孔 1 在 $+180\text{mV}$ 电压下的核酸过孔情况。

5 图 16A 和 16B 示出了根据一个实施例当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 1 移位时的示例电流轨迹。

图 17 是图 16B 实施例单独一条信号的区域放大显示图。

10 图 18A 示出了根据一个实施例的突变孔 2 在 $\pm 180\text{mV}$ 电压下开孔电流及其门控特征。

图 18B 示出了根据一个实施例的突变孔 2 在 $+180\text{mV}$ 电压下的核酸过孔情况。

15 图 19A 和 19B 示出了根据一个实施例当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 2 移位时的示例电流轨迹。

图 20 是图 19A 和 B 实施例单独一条信号的区域放大显示图。

图 21A 示出了根据一个实施例的突变孔 3 在 $\pm 180\text{mV}$ 电压下开孔电流及其门控特征。

20 图 21B 示出了根据一个实施例的突变孔 3 在 $+180\text{mV}$ 电压下的核酸过孔情况。

图 22A 和 22B 示出了根据一个实施例当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 3 移位时的示例电流轨迹。

图 23 是图 22A 实施例单独一条信号的区域放大显示图。

25 图 24A 示出了根据一个实施例的突变孔 4 在 $\pm 180\text{mV}$ 电压下开孔电流及其门控特征。

图 24B 示出了根据一个实施例的突变孔 4 在 $+180\text{mV}$ 电压下的核酸过孔情况。

30 图 25A 和 25B 示出了了根据一个实施例当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 4

移位时的示例电流轨迹。

图 26 是图 25A 和 25B 实施例单独一条信号的区域放大显示图。

图 27 示出了根据一个实施例的突变孔 5 在±180mV 电压下开孔电流及其门控特征。

5 图 28 示出了根据一个实施例当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 5 移位时的示例电流轨迹。

图 29 是图 28 实施例单独一条信号的区域放大显示图。

图 30A 示出了根据一个实施例的突变孔 6 在±180mV 电压下开孔电流及其门控特征。

10 图 30B 示出了根据一个实施例的突变孔 6 在+180mV 电压下的核酸过孔情况。

图 31A 和 31B 示出了根据一个实施例当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 6 移位时的示例电流轨迹。

15 图 32 是图 31A 实施例单独一条信号的区域放大显示图。

图 33 示出了根据一个实施例突变体 1 的 SDS-PAGE 电泳检测结果。

图 34 示出了根据一个实施例突变体 1 蛋白尺寸排阻色谱图。

20 具体实施方式

应理解，所公开的产品和方法的不用应用可根据所属领域的特定需要来调适。还应理解，本文所用的术语仅出于描述本发明的特定实施例的目的，并且不打算限制性的。

25 另外除非上下文另外明确规定，否则如本说明书和权利要求书中所使用，单数形式“一”和“所述”包括多个。举例来说，提及“核苷酸”包括两个或更多个核苷酸，提及“一个解旋酶”包括两个或更多解旋酶。

30 如本文所使用的，术语“包括”是指必须包括任何所列举的要素，

并且也可以任选地包括其他元素。“由...组成”是指不包括所有未列举的元素。由这些术语中的每一个定义的实施例在本发明的范围内。

如本文所用的“核苷酸序列”、“DNA 序列”或“核酸分子”是指任何长度的核苷酸(核糖核苷酸或脱氧核糖核苷酸)的聚合形式。该术语仅指分子的一级结构。因此, 该术语包括双链和单链 DNA 和 RNA。

本文所用的术语“核酸”是指单链或双链共价连接的核苷酸序列, 其中每个核苷酸上的 3'和 5'末端通过磷酸二酯键连接。核苷酸可以由脱氧核糖核苷酸碱基或核糖核苷酸碱基组成。核酸可以包括 DNA 和 RNA, 并可以在体外合成制备或从自然资源中分离。核酸可以进一步包括修饰的 DNA 或 RNA, 例如甲基化的 DNA 或 RNA, 或经过翻译后修饰的 RNA, 例如用 7-甲基鸟苷进行的 5'-盖帽, 3'-端加工, 例如裂解和多腺苷化, 以及拼接。核酸还可以包括合成核酸(XNA), 例如己糖醇核酸(HNA), 环己烯核酸(CeNA), 苏糖核酸(TNA), 甘油核酸(GNA), 锁核酸(LNA)和肽核酸(PNA)。核酸(或多核苷酸)的大小通常用双链多核苷酸的碱基对(bp)数目表示, 或在单链多核苷酸的情况下用核苷酸的数目(nt)表示。1 千个 bp 或 nt 等于一个千碱基对(kb)。长度小于约 40 个核苷酸的多核苷酸通常称为“寡核苷酸”, 并且可以包含用于 DNA 操作(例如通过聚合酶链式反应(PCR))中的引物。

多核苷酸, 例如核酸, 是包含两个或多个核苷酸的大分子。所述多核苷酸或核酸可以包含任意核苷酸的任意组合。所述核苷酸可以是天然存在的或人工合成的。所述多核苷酸中的一个或多个核苷酸可以被氧化或甲基化。所述多核苷酸中的一个或多个核苷酸可以被损伤。例如, 所述多核苷酸可以包含嘧啶二聚体。这种二聚体通常与由紫外线造成的损伤有关并且是皮肤黑色素瘤的主要成因。所述多核苷酸中的一个或多个核苷酸可以被修饰, 例如用常规的标记或标签。所述多核苷酸可以包含一个或多个无碱基的(即缺少核碱基)、或缺少核碱基和糖(即为 C3)的核苷酸。

所述多核苷酸中的核苷酸可以任意方式相互连接。所述核苷酸通常通过其糖基和磷酸基团连接，如在核酸中一样。所述核苷酸可以通过其核碱基连接，如在啼啉二聚体中一样。

5

多核苷酸可以是单链或双链的。多核苷酸的至少一部分优选是双链的。多核苷酸可以是核酸，例如脱氧核糖核酸(DNA)或核糖核酸(RNA)。多核苷酸可以包含一条 RNA 链，所述 RNA 链杂合到一条 DNA 链。多核苷酸可以是任意现有技术已知的合成核酸，例如肽核酸(PNA)，
10 甘油核酸(GNA)，苏糖核酸(TNA)，锁核酸(LNA)或具有核苷酸侧链的其他合成聚合物。所述 PNA 骨架是由通过肽键连接的重复的 N- (2-氨基乙基)-甘氨酸单元组成。所述 GNA 骨架是由通过磷酸二酯键连接的重复乙二醇单元组成。所述 TNA 骨架是由通过磷酸二酯键连接在一起的重复苏糖基组成。LNA 由上述核糖核酸形成，具有连接核糖部分中
15 2' 氧和 4' 碳的额外桥连结构。桥连的核酸(BNA)是修饰的 RNA 核苷酸。它们也可以称为限制的或不可接近的 RNA¹³BNA 单体可以含有 5 元，6 元或甚至 7 元桥连结构并带有 “固定的” C3' -内糖折叠结构 (C3'-endo sugar pucker)。所述桥连结构被合成引入核糖的 2', 4' 位，以产生 2', 4'-BNA 单体。

20

多核苷酸最优选核糖核酸(RNA)或脱氧核糖核酸(DNA)。多核苷酸可以为任意长度。例如，多核苷酸的长度可以是至少 10，至少 50，至少 100，至少 150，至少 200，至少 250，至少 300，至少 400 或至少 500 个核苷酸或核苷酸对。所述多核苷酸的长度可以为 1000 个或更多
25 个核苷酸或核苷酸对，5000 个或更多个核苷酸或核苷酸对或 100000 个或更多个核苷酸或核苷酸对。

任意数量的多核苷酸可以被研究。例如实施例的方法可以涉及表征 2、3、4、5、6、7、8、9、10、20、30、50、100 个或更多个多核苷酸。如果两个或更多个多核苷酸被表征，它们可以是不同的多核苷
30

酸或相同多核苷酸的情形。

多核苷酸可以是天然存在的或人工合成的。例如，所述方法可用于验证所制备的寡核苷酸的序列。所述方法通常在体外进行。

5

在本公开的上下文中，术语“氨基酸”以其最广义的意义使用，并且意指包括含有胺(NH₂)和羧基(COOH)官能团以及每种氨基酸所特有的侧链(例如 R 基团)的有机化合物。在一些实施方案中，氨基酸是指天然存在的 L α -氨基酸或残基。本文使用天然存在的氨基酸的常用单字母和三字母缩写：A=Ala; C=Cys; D=Asp; E=Glu; F=Phe; G=Gly; H=His; I=Ile; K=Lys; L=Leu; M=Met; N=Asn; P=Pro; Q=Gln; R=Arg; S=Ser; T=Thr; V=Val; W=Trp; 和 Y=Tyr (Lehninger, A.L., (1975) Biochemistry, 第2版, 第71-92页, Worth Publishers, New York)。通用术语“氨基酸”还包括 D-氨基酸、逆-反氨基酸以及经化学修饰的氨基酸(诸如氨基酸类似物)，通常不并入蛋白质中的天然存在的氨基酸(诸如正亮氨酸)及具有本领域已知是氨基酸特征的性质的化学合成化合物(诸如 β -氨基酸)。例如，在氨基酸的定义中包括苯丙氨酸或脯氨酸的类似物或模拟物，这些类似物或模拟物允许如同天然 Phe 或 Pro 一样对肽化合物进行相同的构象限制。此类类似物和模拟物在本文中称为相应氨基酸的“功能等效物”。Roberts 和 Vellaccio, The Peptides: Analysis, Synthesis, Biology, Gross 和 Meierhofer 编辑, 第5卷第341页, Academic Press, Inc., N.Y. 1983 列出了氨基酸的其他实例，其通过引用并入本文。

10

15

20

术语“蛋白质”、“多肽”和“肽”在本文中进一步可互换使用，是指氨基酸残基的聚合物以及氨基酸残基的变体和合成类似物。因此，这些术语适用于其中一个或多个氨基酸残基是合成的非天然存在的氨基酸，诸如相应天然存在的氨基酸的化学类似物的氨基酸聚合物，以及适用于天然存在的氨基酸聚合物。多肽还可经历成熟或翻译后修饰过程，这些过程可以包括但不限于：糖基化、蛋白水解裂解、脂化、

25

30

信号肽裂解、前肽裂解、磷酸化等。

5 蛋白质的“同源物”涵盖相对于所讨论的未修饰的或野生型蛋白质具有氨基酸取代、缺失和/或插入并且具有与它们所来源的未修饰的蛋白质相似的生物和功能活性的肽、寡肽、多肽、蛋白质和酶。如本文所用，术语“氨基酸同一性”是指在比较窗口中，在氨基酸-氨基酸的基础上，序列相同的程度。因此，通过以下方式计算“序列同一性百分比”：在比较窗口中比较两个最佳比对的序列，确定两个序列中出现相同氨基酸残基(例如，Ala、Pro、Ser、Thr、Gly、Val、Leu、Ile、10 Phe、Tyr、Trp、Lys、Arg、His、Asp、Glu、Asn、Gln、Cys 和 Met)的位置数量以得到匹配位置数量，将匹配位置数量除以比较窗口中的位置总数(即窗口大小)，并将结果乘以 100 得到序列同一性百分比。

15 序列同一性也可以是全长多核苷酸或多肽的片段或部分。因此，序列可与全长参考序列仅有 50% 的整体序列同一性，但是特定区、结构域或亚基的序列可与参考序列具有 80%、90% 或高达 99% 的序列同一性。

20 术语“野生型”是指从天然存在的来源分离的基因或基因产物。野生型基因是在群体中最常观测到的基因，因此任意地设计为该基因的“正常”或“野生型”形式。相反，术语“经修饰的”、“突变”或“变体”是指与野生型基因或基因产物相比，显示出序列修饰(例如，取代、截短或插入)、翻译后修饰和/或功能性质(例如，特性改变)的基因或基因产物。注意，可以分离天然存在的突变体；这些突变体是通过与野生型基因或基因产物相比，它们具有改变的特征这一事实来鉴定的。引入或取代天然存在的氨基酸的方法是本领域众所周知的。例如，可以通过在编码突变的单体的多核苷酸中的相关位置用精氨酸的密码子(CGT)置换蛋氨酸的密码子(ATG)，用精氨酸(R)取代蛋氨酸(M)。25 引入或取代非天然存在的氨基酸的方法也是本领域众所周知的。例如，30 可以通过在用于表达突变的单体的 IVTT 系统中包括合成的氨酰基

5 -tRNA 来引入非天然存在的氨基酸。可替代地，可以通过在绦虫假单胞菌（*Pseudomonas taeanensis* MS-3）中表达突变的单体来引入非天然存在的氨基酸，绦虫假单胞菌在那些特定氨基酸的合成(即非天然存在的)类似物的存在下对于特定氨基酸而言为营养缺陷型。如果突变的单体是使用部分肽合成法产生的，则它们也可以通过裸连接产生。保守性取代将氨基酸置换为具有相似化学结构、相似化学性质或相似侧链体积的其他氨基酸。引入的氨基酸可以具有与它们所置换的氨基酸相似的极性、亲水性、疏水性、碱性、酸性、中性或电荷。可替代地，保守性取代可以引入另一种芳族或脂肪族氨基酸代替预先存在的芳族或脂肪族氨基酸。保守性氨基酸变化是本领域众所周知的，并且可以根据下表 1 中定义的 20 种主要氨基酸的性质进行选择。在氨基酸具有相似极性的情况下，这也可以参考表 2 中氨基酸侧链的亲水性量表来确定。

15 表 1-氨基酸的化学性质

Ala,A	脂肪族、疏水性、中性	Met,M	疏水性、中性
Cys,C	极性、疏水性、中性	Asn,N	极性、亲水性、中性
Asp,D	极性、亲水性、带电荷(-)	Pro,P	疏水性、中性
Glu,E	极性、亲水性、带电荷(-)	Gln,Q	极性、亲水性、中性
Phe,F	芳族、疏水性、中性	Arg,R	极性、亲水性、带电荷(+)
Gly,G	脂肪族、中性	Ser,S	极性、亲水性、中性
His,H	芳族、极性、亲水性、带电荷(+)	Thr,T	极性、亲水性、中性
Ile,I	脂肪族、疏水性、中性	Val,V	脂肪族、疏水性、中性
Lys,K	极性、亲水性、带电荷(+)	Trp,W	芳族、疏水性、中性
Leu,L	脂肪族、疏水性、中性	Tyr,Y	芳族、极性、疏水性

表 2-亲水性量表

侧链	亲水性
Ile,I	4.5
Val,V	4.2
Leu,L	3.8
Phe,F	2.8
Cys,C	2.5
Met,M	1.9
Ala,A	1.8
Gly,G	-0.4
Thr,T	-0.7
Ser,S	-0.8
Trp,W	-0.9
Tyr,Y	-1.3
Pro,P	-1.6
His,H	-3.2
Glu,E	-3.5
Gln,Q	-3.5
Asp,D	-3.5
Asn,N	-3.5
Lys,K	-3.9
Arg,R	-4.5

众所周知，性质相似氨基酸彼此之间保守性替换通常不会影响肽序列的活性，保守性替换如表 3。

表 3 保守氨基酸替换

类型	氨基酸
脂肪族	甘氨酸(G)，丙氨酸(A)，缬氨酸(V)，亮氨酸(L)，异亮氨酸(I)
水合或含硫/硒	丝氨酸(S)，半胱氨酸(C)，硒半胱氨酸(U)，苏氨酸(T)，甲硫氨酸(M)
环状	脯氨酸(P)
芳香族	苯丙氨酸(F)，酪氨酸(Y)，色氨酸(W)
碱性	组氨酸(H)，赖氨酸(K)，精氨酸(R)
酸性及酰胺类	天冬氨酸(D)，谷氨酸(E)，天冬酰胺(N)，谷氨酰胺(Q)

5 突变或经修饰的蛋白质、单体或肽也可以任何方式在任何位点进行化学修饰。突变或经修饰的单体或肽优选通过分子与一个或多个半胱氨酸的附接(半胱氨酸连接)，分子与一个或多个赖氨酸的附接，分子与一个或多个非天然氨基酸的附接，表位的酶修饰或末端的修饰来进行化学修饰。进行此类修饰的合适方法是本领域众所周知的。经修饰的蛋白质、单体或肽的突变体可以通过任何分子的附接进行化学修饰。例如，经修饰的蛋白质、单体或肽的突变体可以通过染料或荧光团的附接进行化学修饰。在一些实施方案中，用促进包含单体或肽的孔与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化学修饰突变或经修饰的单体或肽。分子衔接子优选为环状分子、环糊精、能够杂交的物质、DNA 结合剂或嵌入剂、肽或肽类似物、合成聚合物、芳族平面分子、带正电荷的小分子或能够氢键键合的小分子。

15 衔接子的存在改善了孔和核苷酸或多核苷酸序列的主-客体化学，从而改善了由突变的单体形成的孔的测序能力。主-客体化学的原理是本领域众所周知的。衔接子对孔的物理或化学性质有影响，这种影响改善了孔与核苷酸或多核苷酸序列的相互作用。衔接子可以改变孔的

桶或通道的电荷，或与核苷酸或多核苷酸序列特异性相互作用或结合，从而促进其与孔的相互作用。

“蛋白孔”是跨膜蛋白结构，其限定了允许分子和离子从膜的一侧易位到另一侧的通道或孔。离子物质通过孔的易位可以由施加到孔任一侧的电位差驱动。“纳米孔”是一种蛋白孔，其中分子或离子所通过的通道的最小直径为纳米级(10^{-9} 米)。在一些实施方案中，蛋白孔可以是跨膜蛋白孔。蛋白孔的跨膜蛋白结构本质上可以是单体或寡聚体。通常，孔包含多个围绕中心轴排列的多肽亚基，从而形成基本上垂直于纳米孔所驻留的膜延伸的蛋白内衬通道。多肽亚基的数量没有限制。通常，亚基的数量为 5 至 30，合适地亚基的数量为 6 至 10。可替代地，亚基的数量不像在产气荚膜梭菌溶素(perfringolysin)或相关大膜孔的情况下那样定义。纳米孔内形成蛋白内衬通道的蛋白亚基部分通常包含可包括一个或多个跨膜 β -桶和/或 α -螺旋部分的二级结构基序。

在一个实施例中，蛋白孔包含一个或多个孔蛋白单体。每个孔蛋白单体可以来自绦虫假单胞菌。在一个实施例中，蛋白孔包括一个或多个孔蛋白单体的突变体（即一个或多个孔蛋白突变的单体）。

在一个实施例中，孔蛋白来自生物界野生型蛋白、野生型同源物、或其突变体。突变体可以成为修饰的孔蛋白或孔蛋白突变体。突变体中的修饰包括但不限于本文公开的任何一种或多种修饰或所述修饰的组合。在一个实施例中，生物界野生型蛋白是来自绦虫假单胞菌的蛋白。

在一个实施例中，孔蛋白同源物是指与 SEQ ID NO:1 所示的蛋白具有至少 99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、85%、80%、75%、70%、65%、60%、55%、50%的完整序列同一性的多肽。

5 在一个实施例中，孔蛋白同源物是指与 SEQ ID NO:2 所示的蛋白的编码多核苷酸具有至少 99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、85%、80%、75%、70%、65%、60%、55%、50% 的完整序列同一性的多核苷酸。所述多核苷酸序列可以包含基于遗传密码的简并性而与 SEQ ID NO:2 不同的序列。

10 多核苷酸序列可以采用本领域的标准方法进行衍生和复制。编码野生型孔蛋白的染色体 DNA 可以从产生孔的生物体如绦虫假单胞菌中提取。编码所述孔亚基的基因可以使用包括特异性引物的 PCR 进行扩增。所述扩增的序列随后可以进行定点突变。定点突变的合适方法是本领域已知的并且包括，例如，组合链式反应。编码实施例的构建的多核苷酸可以采用本领域公知的技术制备，例如在 Sambrook,J.and Russell , D. (2001). Molecular Cloning A Laboratory Manual,3rd
15 Edition.Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 中描述的那些。

20 所得到的多核苷酸序列随后可以被整合到重组可复制载体上，例如克隆载体。所述载体可以用于在相容的宿主细胞中复制所述多核苷酸。因此多核苷酸序列可以通过将多核苷酸引入到可复制载体中，将载体引入相容的宿主细胞中，并在引起载体复制的条件下使宿主细胞生长而进行制备。所述载体可以从所述宿主细胞中回收。

25 纳米孔或蛋白孔的基本工作原理
在一个实施例中，在充满电解液的腔 100 内，带有纳米级小孔的绝缘膜 102 将腔体分成 2 个小室，如图 1 所示，当电压作用于电解液室，离子或其他小分子物质在电场力作用下穿过小孔，形成稳定的可检测的离子电流。掌握纳米孔的尺寸和表面特性、施加的电压及溶液条件，可检测不同类型的生物分子。

30

由于组成 DNA 的四种碱基腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)和胸腺嘧啶(T)的分子结构及体积大小均不同，单链 DNA (ssDNA) 在控速酶和电场驱使下通过纳米级的小孔时，不同碱基的化学性质差异导致穿越纳米孔或蛋白孔时引起的电流的变化幅度不同，从而得到所测核酸例如 DNA 的序列信息。

图 2 示出了 DNA 测序的示意图 200。如图 2 所示，在一个典型的纳米孔/蛋白孔测序实验中，纳米孔是磷脂膜两侧离子通过的唯一通道。控速蛋白例如多核苷酸结合蛋白充当核酸分子例如 DNA 的马达蛋白，拉动 DNA 链使其以单个核苷酸的步长依次通过纳米孔/蛋白孔。每当一个核苷酸穿过纳米孔/蛋白孔，相应的堵孔信号会被记录下来（图 3）。通过相应算法分析这些序列相关的电流信号，可以反推出核酸分子例如 DNA 的序列信息。

在实施例中，孔蛋白通过生物信息学手段和进化角度，从自然界不同物种（主要是细菌和古细菌）进行筛选。在一个实施例中，孔蛋白来自于任何生物体，优选来自于绦虫假单胞菌。通过序列分析，孔蛋白具有完整功能结构域。利用结构生物学手段预测分析孔蛋白 3D 结构模型，选择具有合适读取头架构形式的通道蛋白。之后利用基因工程、蛋白质工程、蛋白质定向进化和计算机辅助蛋白质设计等手段，对候选通道蛋白（或孔蛋白）进行改造、测试和优化，经过几轮迭代，得到同源蛋白突变体多个，优选六个（不同同源蛋白骨架），有不同的信号特征和信号分布模式。

实施例中的孔蛋白可应用于第四代测序技术。在一个实施例中，孔蛋白是纳米孔蛋白。在一个实施例中，孔蛋白可应用于固态孔进行测序。

在一个实施例中，采用新的蛋白骨架，形成新的缢缩区（读取头区域）结构，从而在测序过程中提供全新的作用模式。实施例的孔蛋

白具有良好的跳边分布和与磷脂膜重组的效率。

在一个实施例中，对野生型孔蛋白单体进行基因突变改造形成孔蛋白单体的突变体。在一个实施例中，孔蛋白单体的突变体的氨基酸包括 SEQ ID NO:1 所示的序列或包括与其具有至少 99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、85%、80%、75%、70%、65%、60%、55%、或 50%同一性的序列，并且所述孔蛋白单体的突变体的氨基酸对应 SEQ ID NO:1 的 62-79、104、170 和 175 位的一个或多个位置处具有突变。

在一个实施例中，突变包括氨基酸的插入、缺失和/或替换。在一个实施例中，SEQ ID NO:1 的 62-79、104、170 和 175 位的一个或多个位置处具有突变，是 SEQ ID NO:1 的 62-79、104、170 和 175 位中的一个或多个位置处具有氨基酸插入、缺失和/或替换。

在一个实施例中，所述孔蛋白单体的突变体的氨基酸对应 SEQ ID NO:1 的（1）69-76 位、（2）64-79 位、（3）62、63、68、69、71、75、76、和 104 位、或（4）68-76 位和 171 和 175 位的一个或多个位置处具有突变。

在一个实施例中，所述孔蛋白单体的突变体的氨基酸对应 SEQ ID NO:1 的（1）69-76 位、（2）64-79 位、（3）62、63、68、69、71、75、76、和 104 位、或（4）68-76 位和 171 和 175 位的一个或多个位置处具有氨基酸的插入、缺失和/或替换。

在一个实施例中，所述孔蛋白单体的突变体的氨基酸仅在对应 SEQ ID NO:1 的 69-76 位（即 K69、P70、T71、P72、A73、S74、S75、F76）具有突变，或在一个或多个位置处具有氨基酸的插入、缺失和/或替换。

5

在一个实施例中，所述孔蛋白单体的突变体的氨基酸仅在对应 SEQ ID NO:1 的 64-79 位（即 Q64、T65、G66、Q67、Y68、K69、P70、T71、P72、A73、S74、S75、F76、S77、T78 和 S79）具有突变，或在一个或多个位置处具有氨基酸的插入、缺失和/或替换。

10

在一个实施例中，所述孔蛋白单体的突变体的氨基酸仅在对应 SEQ ID NO:1 的 R62、D63、Y68、K69、T71、S75、F76Q、和 E104 位具有突变，或在一个或多个位置处具有氨基酸的插入、缺失和/或替换。

15

在一个实施例中，所述孔蛋白单体的突变体的氨基酸仅在对应 SEQ ID NO:1 的 68-76 位（即 Y68、K69、P70、T71、P72、A73、S74、S75 和 F76）、E171 和 D175 具有突变，或在一个或多个位置处具有氨基酸的插入、缺失和/或替换。

20

在一个实施例中，对应 SEQ ID NO:1 的位置是指无论是否通过氨基酸插入或缺失或采用同一性序列从而使得序列编号发生变化时，相对位置不变，依然可采用 SEQ ID NO:1 序列的编号。例如，对应 SEQ ID NO:1 的 Q64 可突变为 Q64L，即使 SEQ ID NO:1 序列编号变化或采用与 SEQ ID NO:1 具有本文限定的同一性的序列，相对应于 SEQ ID NO:1 的 64 位的氨基酸 Q（即使在另一序列中并非为 64 位）也可突变为 L，仍在本发明的保护范围内。

25

在一个实施例中，孔蛋白单体的突变体的氨基酸由 SEQ ID NO:1 所示的序列组成，或由与其具有至少 99%、98%、97%、96%、95%、94%、93%、92%、91%、90%、85%、80%、75%或 70%、65%、60%、55%、或 50%同一性的序列组成，并且所述孔蛋白单体的突变体的氨基酸对应 SEQ ID NO:1 的 62-79、104、170 和 175 位的一个或多个位置处具有突变。

30

在一个实施例中，孔蛋白单体的 SEQ ID NO:1 序列来自绦虫假单胞菌。编码 SEQ ID NO:1 氨基酸的核苷酸序列为 SEQ ID NO:2。

5 在一个实施例中，69-76 位的氨基酸 KPTPASSF 突变为 $M_1M_2M_3M_4M_5M_6M_7M_8$ ，其中 M_1 选自 R、K 或 H； M_2 选自 P； M_3 选自 S、G、C、U、T、M、A、V、L 或 I； M_4 选自 P； M_5 选自 A、G、V、L 或 I； M_6 选自 S、C、U、T 或 M； M_7 选自 A、T、G、V、L、I、S、C、U 或 M； M_8 选自 Q、D、E、N、K、H 或 R。

10 在一个实施例中，对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 $M_9M_{10}M_{11}M_{12}M_{13}M_{14}M_{15}M_{16}M_{17}M_{18}M_{19}M_{20}M_{21}M_{22}M_{23}M_{24}$ ，其中 M_9 选自 L、G、A、V 或 I； M_{10} 选自 T、S、C、U 或 M； M_{11} 选自 G、A、V、L 或 I； M_{12} 选自 Q、D、E 或 N； M_{13} 选自 Y、F 或 W； M_{14} 选自 R、H 或 K； M_{15} 选自 P； M_{16} 选自 S、C、U、T 或 M； M_{17} 选自 P； M_{18} 选自 A、G、V、L 或 I； M_{19} 选自 S、C、U、T 或 M； M_{20} 选自 A、G、V、L 或 I； M_{21} 选自 N、D、E、Q、L、G、A、V 或 I； M_{22} 选自 S、C、U、T 或 M； M_{23} 选自 T、S、C、U 或 M； M_{24} 选自 A、G、V、L 或 I。

20 在一个实施例中，对应 SEQ ID NO:1 的 R62 突变为 R62S、R62C、R62U、R62T 或 R62M；D63 突变为 D63V、D63G、D63A、D63L 或 D63I；Y68 突变为 Y68F 或 Y68W；K69 突变为 K69R 或 K69H；T71 突变为 T71S、T71C、T71U 或 T71M；S75 突变为 S75A、S75G、S75V、S75L 或 S75I；F76 突变为 F76Q、F76D、F76E 或 F76N；E104 突变为 E104V、E104G、E104A、E104L 或 E104I。

25 在一个实施例中，对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 $M_{25}M_{26}M_{27}M_{28}M_{29}M_{30}M_{31}M_{32}M_{33}$ ，并且 E171 突变为 E171N、E171D 或 E171Q；D175 突变为 D175N、D175E 或 D175Q，其中 M_{25} 选自 F、Y 或 W； M_{26} 选自 R、H 或 K； M_{27} 选自 P； M_{28} 选自

S、C、U、T 或 M；M₂₉ 选自 P；M₃₀ 选自 A、G、V、L 或 I；M₃₁ 选自 S、C、U、T 或 M；M₃₂ 选自 A、G、V、L 或 I；M₃₃ 选自 Q、D、E、或 N。

5 在一个实施例中，孔蛋白单体的突变体，其中氨基酸突变选自以下：

(a) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 RPSPASAQ；

10 (b) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 KPGPASTK；

(c) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 LTGQYRPSPASANSTA；

(d) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 LTGQYRPSPASALSTA；

15 (e) 对应 SEQ ID NO:1 的 R62S、D63V、Y68F、K69R、T71S、S75A、F76Q、和 E104V；和

(f) 对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 FRPSPASAQ，以及对应 SEQ ID NO:1 的 E171N 和 D175N。

20 在一个实施例中，所述孔蛋白单体的突变体的氨基酸序列包括 SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 或 SEQ ID NO:30，或由其组成。

25 在一个实施例中，蛋白孔包括至少一个孔蛋白单体的突变体（或孔蛋白突变的单体）。在一个实施例中，蛋白孔包括至少两个、三个、四个、五个、六个、七个、八个、九个或十个或更多孔蛋白单体的突变体。在一个实施例中，蛋白孔包括至少两个孔蛋白单体的突变体，所述孔蛋白单体的突变体可以是相同的或不同的。在一个实施例中，蛋白孔包括两个或多个孔蛋白单体的突变体，优选为两个或多个单体的突变体相同。在一个实施例中，蛋白孔的缢缩区孔道直径为

30

0.7nm-2.2nm、0.9nm-1.6nm、1.4-1.6nm 或 15.9Å-20.1Å。

孔蛋白单体的突变体或包括其的蛋白孔用于检测靶分析物存在、不存在或一个或多个特征中的应用。在一个实施例中，孔蛋白单体的突变体或蛋白孔用于检测核酸分子的序列，或表征多核苷酸序列，例如测序多核苷酸序列，因为它们可以高灵敏度区分不同的核苷酸。孔蛋白单体的突变体或包括其的蛋白孔可以区分 DNA 和 RNA 中的四种核苷酸，甚至可以区分甲基化和未甲基化的核苷酸，并且分辨率出人预料的高。孔蛋白单体的突变体或蛋白孔显示对全部四种 DNA/RNA 核苷酸的几乎完全分离。基于在蛋白孔中的停留时间和流过蛋白孔的电流，进一步区分脱氧胞嘧啶单磷酸(dCMP)和甲基-dCMP。

孔蛋白单体的突变体或蛋白孔还可以在一系列条件下区分不同核苷酸。特别地，所述孔蛋白单体的突变体或蛋白孔在有利于核酸表征如测序的条件下区分核苷酸。通过改变施加的电势，盐浓度，缓冲液，温度和添加剂如脲，甜菜碱和 DTT 的存在，可以控制孔蛋白单体的突变体或蛋白孔区分不同核苷酸的程度。这允许孔蛋白单体的突变体或蛋白孔的功能被精细调控，特别是在测序时。孔蛋白单体的突变体或蛋白孔也可以用于通过与一种或多种单体的相互作用而不是在以核苷酸为基础的核苷酸上，来鉴定多核苷酸聚合物。

孔蛋白单体的突变体或蛋白孔可以是分离的，基本上分离的，纯化的或基本纯化的。如果完全不含任何其它组分，例如脂质体或其它蛋白孔/孔蛋白，则实施例的孔蛋白单体的突变体或蛋白孔被分离或纯化。如果孔蛋白单体的突变体或蛋白孔与不会干扰其预期用途的载体或稀释剂混合，则该孔蛋白单体的突变体或蛋白孔基本上被分离。例如，如果孔蛋白单体的突变体或蛋白孔以包含小于 10%，小于 5%，小于 2%或小于 1%的其它组分如三嵌段共聚物，脂质体或其它蛋白孔/孔蛋白的形式存在，则所述孔蛋白单体的突变体或蛋白孔基本上被分离或基本上被纯化。替代地，孔蛋白单体的突变体或蛋白孔可以存在于

膜中。

例如，膜优选为两亲层。两亲层是由两亲分子形成的层，例如，磷脂，其具有亲水性和亲油性。两亲分子可以是合成的或天然存在的。两亲层可以是单层或双层。两亲层通常是平面的。两亲层可以是弯曲的。可以对两亲层进行支撑。膜可以为脂质双层。脂质双层是由脂质的两个相对的层形成的。脂质的两层被排列为使得它们的疏水性尾部基团面向彼此以形成疏水性内部。脂质的亲水性头部基团面向外朝向该双层的每一侧上的含水环境。膜包括固态层。固态层可以由有机和无机材料形成。如果膜包括固态层，则孔通常存在于两亲膜中或固态层内包括的层中，例如，固态层内的孔洞、阱、间隙、通道、沟槽或狭缝中。

分析物的表征

实施例提供一种确定靶分析物的存在、不存在或一种或多种特性的方法。该方法涉及将所述靶分析物与孔蛋白单体的突变体或蛋白孔接触，使得所述靶分析物相对于，例如穿过，所述孔蛋白单体的突变体或蛋白孔移动，并且当所述靶分析物相对于所述孔蛋白单体的突变体或蛋白孔移动时获取一个或多个测量值，从而确定所述靶分析物的存在、不存在或一种或多种特性。所述靶分析物也可以被称为模板分析物或感兴趣的分析物。

靶分析物优选为多糖、金属离子，无机盐，聚合物，氨基酸，肽，多肽，蛋白，核苷酸，寡核苷酸，多核苷酸，染料，药物，诊断剂，爆炸物或环境污染物。所述方法可以涉及确定两种或更多种相同类别的靶分析物的存在、不存在或一种或多种特性，例如，两种以上蛋白，两种以上核苷酸或两种以上药物。或者，所述方法可以涉及确定两种或更多种不同类别的靶分析物的存在、不存在或一种或多种特性，例如，一种或多种蛋白，一种或多种核苷酸和一种或多种药物。

所述方法包括将所述靶分析物与孔蛋白单体的突变体或蛋白孔接触，使得所述靶分析物移动穿过所述孔蛋白单体的突变体或蛋白孔。所述蛋白孔一般包含至少 1 个，至少 2 个，至少 3 个，至少 4 个，至少 5 个，至少 6 个，至少 7，至少 8，至少 9 或至少 10 个孔蛋白突变的单体，例如，7，8，9 或 10 个单体。所述蛋白孔包括相同的单体或不同的孔蛋白单体，优选包含 8 或 9 个相同的单体。所述单体中的一个或多个，例如 2、3、4、5、6、7、8、9 或 10 个，优选如上述讨论的被化学修饰。在一个实施例中，每个单体的氨基酸包括 SEQ ID NO:1 及其上述突变体。在一个实施例中，每个单体的氨基酸由 SEQ ID NO:1 及其上述突变体组成。

实施例的方法可以测量多核苷酸的两个、三个、四个或五个或更多个特征。所述一个或多个特征优选选自 (i) 多核苷酸的长度，(ii) 多核苷酸的身份，(iii) 多核苷酸的序列，(iv) 多核苷酸的二级结构，以及 (v) 多核苷酸是否被修饰。在一个实施例中，可以测量 (i) 至 (v) 的任意组合。

对于 (i)，可以例如通过确定多核苷酸和蛋白单体的突变体/蛋白孔之间相互作用的数量或多核苷酸和蛋白单体的突变体/蛋白孔之间相互作用的持续时间对多核苷酸的长度进行测量。

对于 (ii)，可以以多种方式测量多核苷酸的身份，多核苷酸的身份可以结合多核苷酸序列的测量或不结合多核苷酸序列的测量进行测量。前者较为简单；对多核苷酸进行测序进而进行识别。后者可以通过几种不同方式完成。例如，可以测量多核苷酸中特定基序的存在(无需测量多核苷酸的其余序列)。或者，所述方法中特定的电和/或光信号的测量可以识别出所述多核苷酸来自特定来源。

对于 (iii)，多核苷酸的序列可以如先前所述进行测定。合适的测序方法，特别是使用电学测量方法的测序方法，描述在 Stoddart D et

al. ,ProC Natl Acad Sci ,12; 106 (19) 7702-7, Lieberman KR et al,J Am Chem SoC.2010;132 (50) 17961-72,以及国际申请 W02000/28312 中。

5 对于 (iv), 二级结构可以采用多种方法测量。例如, 如果所述方法涉及电学测量方法, 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所述二级结构。这允许区分单链和双链多核苷酸的区域。

10 对于(v), 可以测量是否存在任何修饰。所述方法优选包括, 测定多核苷酸是否通过甲基化, 氧化, 损伤, 用一种或多种蛋白或用一种或多种标记, 标签或进行无碱基或缺少核碱基和糖的修饰。特定的修饰将导致与所述孔的特异性相互作用, 其可以使用下述方法进行测量。例如, 甲基胞嘧啶可以基于其与每个核苷酸相互作用期间流过所述孔的电流而与胞嘧啶区分开来。

15 所述靶多核苷酸与蛋白单体的突变体/蛋白孔接触, 例如如实施例的蛋白单体的突变体/蛋白孔。所述蛋白单体的突变体/蛋白孔通常存在于膜中。合适的膜如前文所述。所述方法可以使用适合于研究膜/蛋白孔或孔蛋白单体的突变体系统一其中蛋白单体的突变体/蛋白孔存在于膜中的任何装置进行。所述方法可以使用适合用于跨膜孔感侧的任何
20 装置进行。例如, 所述装置包括包含水性溶液的腔室和将腔室分成两个部分的屏障。所述屏障通常具有孔洞, 在孔洞中形成包含孔的膜。或者所述屏障形成膜, 所述膜中存在蛋白单体的突变体/蛋白孔。该方法可以使用描述于国际申请号 PCT/GB08/000562 (WO 2008/102120)中的装置进行。

25 可以进行各种不同类型的测量。这包括但不限于电学测量和光学测量。电学测量包括电压测量、电容测量、电流测量, 阻抗测量, 隧道测量 (tunnelling measurement) (Ivanov AP et al. ,Nano Lett.2011Jan 12;11 (I) :279-85)以及 FET 测量 (国际申请 TO 2005/124888)。光学测量
30 量可以与电学测量结合 (Soni GV et al., Rev Sci Instrum.2010Jan;81

(1)014301)。所述测量可以是跨膜电流测量，例如流过所述孔的离子电流的测量。在一个实施例中，电学测量或光学测量可采用常规的电学或光学测量。

5 电学测量可以使用描述在 Stoddart D et al • , ProC Natl Acad Sci, 12; 106 (19) 7702-7, Lieberman KR et al, J Am Chem SoC.2010;132 (50) 17961-72 和国际申请 WO 2000/28312 中的标准单通道记录设备进行。替代地，电学测量可以使用多通道系统进行，例如如国际申请 W02009/077734 和国际申请 WO 2011/067559 中描述的。

10

所述方法优选采用跨膜施加的电势进行。所述施加的电势可以是电压电势。替代地，所施加的电势可以是化学电势。其一实例为采用跨膜，例如双亲性分子层的盐梯度进行。盐梯度被公开在 Holden et al., J Am Chem SoC.2007Jul 11;129 (27):8650-5 中。在一些情况下，多核苷酸相对所述蛋白单体的突变体/蛋白孔移动时流过所述蛋白单体的突变体/蛋白孔的电流用于估算或确定所述多核苷酸的序列。这就是链测序。

15

所述方法可以包括测量多核苷酸相对所述孔移动时流过所述孔的电流。因此用于所述方法的设备也可以包括能够施加电势并测量穿过膜和孔的电信号的电路。所述方法可以采用膜片钳或电压钳进行。

20

可以包括测量多核苷酸相对所述孔移动时流过所述孔的电流。测量通过跨膜蛋白孔的离子流的合适条件是本领域已知的并且在实施例中公开。所述方法通常通过施加在所述膜和所述孔上的电压进行。所使用的电压通常为从+5V 至-5V，例如从从+4V 至-4V，从+3V 至-3V 或从+2V 至-2V。所使用的电压通常为从-600mV 至+600V 或-400mV 至 + 400mV。所使用的电压优选在具有选自-400mV， -300mV， -200mV， -150mV， -100mV， -50mV， -20mV 和 0mV 的下限和独立地选自+10mV， +20mV， +50mV， + 100mV， + 150mV， +200mV， +300nA^P+ 400mV 的上限的范围内。所使用的电压更优选在 100mV 至 240mV 的范围内

25

30

并且最优选在 120mV 至 220mV 的范围内。通过使用增加的施加电势，可以增加孔对不同核苷酸的识别。

5 所述方法通常在存在任何电荷载体的情况下进行，例如金属盐例如碱金属盐，卤化物盐例如氯化物盐，例如碱金属氯化物盐。电荷载体可以包括离子液体或有机盐，例如四甲基氯化铵，三甲基苯基氯化铵，苯基三甲基氯化铵或 1-乙基-3-甲基咪唑鎓氯化物。在上述示例性装置中，盐存在于所述腔室中的水性溶液中。通常使用氯化钾（KCl），氯化钠（NaCl），氯化铯（CsCl）或亚铁氰化钾和铁氰化钾的混合物。KCl，NaCl 和亚铁氰化钾和铁氰化钾的混合物是优选的。电荷载体在所述膜上可以是不对称的。例如，电荷载体的类型和/或浓度可以在所述膜的每一侧上不同。

15 所述盐的浓度可以是饱和的。所述盐的浓度可以为 3M 或更低，并且通常为 0.1 至 2.5M，0.3 至 1.9M，0.5 至 1.8M，0.7 至 1.7M，0.9 至 1.6M 或 1M 至 1.4M。所述盐的浓度优选为 150mM 至 1M。所述方法优选使用至少 0.3M，例如至少 0.4M，至少 0.5M，至少 0.6M，至少 0.8M，至少 1.0M，至少 1.5M，至少 2.0M，至少 2.5M 或至少 3.0M 的盐浓度进行。高盐浓度提供高的信噪比，并允许通过电流指示在正常电流波动背景下待识别的核苷酸的存在。

20 所述方法通常在存在缓冲液的情况下进行。在上述示例性装置中，所述缓冲液存在于所述腔室中的水性溶液中。任意缓冲液可以用于本发明的方法。通常地，所述缓冲液为磷酸缓冲液。其他合适的缓冲液为 HEPES 或 Tris-HCl 缓冲液。所述方法通常在 pH 为 4.0 至 12.0、4.5 至 10.0、5.0 至 9.0、5.5 至 8.8、6.0 至 8.7、7.0 至 8.8、或 7.5 至 8.5 下进行。使用的 pH 值优选约 7.5。

30 所述方法可以在 0°C 至 100°C，15°C 至 95°C，16°C 至 90°C，17°C 至 85°C，18°C 至 80°C，19°C 至 70°C 或 20°C 至 60°C 温度下进行。所

述方法通常在室温下进行。所述方法任选的在支持酶功能的温度下进行，例如约 37°C。

5 在一个实施例中，用于确定靶分析物（例如多核苷酸）存在、不存在或一个或多个特征的方法包括将所述靶分析物偶联到膜上；和所述靶分析物与存在于所述膜中的所述蛋白孔相互作用（例如接触）从而使得所述靶分析物相对所述蛋白孔移动（例如穿过所述蛋白孔）。在一个实施例中，测量所述靶分析物相对于所述蛋白孔移动时通过所述蛋白孔的电流，从而确定所述靶分析物的存在、不存在或一个或多个特征（例如为多核苷酸的序列）。

多核苷酸结合蛋白

15 实施例的表征方法优选包括使多核苷酸与多核苷酸结合蛋白接触，使得所述蛋白控制多核苷酸相对于蛋白单体的突变体/蛋白孔的移动，例如，通过所述蛋白单体的突变体/蛋白孔。

20 更优选地，所述方法包括(a)使多核苷酸与蛋白单体的突变体/蛋白孔和多核苷酸结合蛋白接触，使得所述蛋白控制多核苷酸相对于蛋白单体的突变体/蛋白孔的移动，例如，通过蛋白单体的突变体/蛋白孔，和(b)当多核苷酸相对于蛋白单体的突变体/蛋白孔移动时获取一个或多个测量值，其中，所述测量值指示多核苷酸的一个或多个特征，从而表征多核苷酸。

25 更优选地，所述方法包括(a)使多核苷酸与蛋白单体的突变体/蛋白孔和多核苷酸结合蛋白接触，使得所述蛋白控制多核苷酸相对于蛋白单体的突变体/蛋白孔的移动，例如，通过蛋白单体的突变体/蛋白孔，和(b)当多核苷酸相对于蛋白单体的突变体/蛋白孔移动时测量通过蛋白单体的突变体/蛋白孔的电流，其中，所述电流指示多核苷酸的一个或多个特征，从而表征多核苷酸。

30

多核苷酸结合蛋白可以是能够结合多核苷酸并控制其移动通过孔的任何蛋白。多核苷酸结合蛋白通常与多核苷酸相互作用并改性多核苷酸的至少一种性质。蛋白可以通过裂解多核苷酸以形成各单个核苷酸或核苷酸的短链(例如,二核苷酸或三核苷酸)来对其进行改性。蛋白可以通过使多核苷酸定向或将其移动到特定位置来对其进行改性,即,控制它的移动。

多核苷酸结合蛋白优选衍生自多核苷酸处理酶。多核苷酸处理酶是能够与多核苷酸相互作用并改性多核苷酸的至少一种性质的多肽。所述酶可以通过裂解多核苷酸以形成各单个核苷酸或核苷酸的短链(例如,二核苷酸或三核苷酸)来对其进行改性。所述酶可以通过使多核苷酸定向或将其移动到特定位置来对其进行改性。多核苷酸处理酶不需要显示酶活性,只要其能够结合多核苷酸并控制其通过孔的移动即可。例如,可以对所述酶进行改性以去除其酶活性,或者可以在防止其用作酶的条件下进行使用。

多核苷酸处理酶优选为聚合酶、外切核酸酶、解旋酶和拓扑异构酶,例如,促旋酶。在一个实施例中,所述酶优选为解旋酶,例如 Hel308Mbu、Hel308Csy、Hel308Tga、Hel308Mhu、Tral Eco、XPD Mbu、Dda 或其变体。实施例中可以使用任何解旋酶。

在一个实施例中,可以使用任何数量的解旋酶。例如,可以使用 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 或更多个解旋酶。在一些实施例中,可以使用不同数目的解旋酶。

实施例的方法优选包括使多核苷酸与两个或更多个解旋酶接触。所述两个或更多个解旋酶通常是相同的解旋酶。所述两个或更多个解旋酶可以是不同的解旋酶。

所述两个或更多个解旋酶可以是上述解旋酶的任意组合。所述两

个或更多个解旋酶可以是两个或更多个 Dda 解旋酶。所述两个或更多个解旋酶可以是一种或多种 Dda 解旋酶和一种或多种 TrwC 解旋酶。所述两个或更多个解旋酶可以是相同解旋酶的不同变体。

5 所述两个或更多个解旋酶优选地彼此连接。所述两个或更多个解旋酶更优选地彼此共价连接。解旋酶可以以任何顺序并使用任何方法连接。

试剂盒

10 本发明还提供一种用于表征靶分析物（例如靶多核苷酸）的试剂盒。试剂盒包含实施例的孔和膜的组分。膜优选地由组分形成。孔优选地存在于膜中。试剂盒可包含上文所公开的任一个膜(如两亲层或三嵌段共聚物膜)的组分。试剂盒可进一步包含多核苷酸结合蛋白。可使用上文所论述的任一个多核苷酸结合蛋白。

15

在一个实施例中，膜为两亲层、固态层、或脂双层。

试剂盒可进一步包含用于使多核苷酸与膜偶联的一或多个锚。

20 试剂盒优选地是用于表征双链多核苷酸，并优选地包含 Y 衔接子和发夹环衔接子。

25 Y 衔接子优选地具有所连接的一或多个解螺旋酶，且发夹环衔接子优选地具有所连接的一或多个分子制动器。Y 衔接子优选地包含用于使多核苷酸与膜偶联的一或多个第一锚，发夹环衔接子优选地包含用于使多核苷酸与膜偶联的一或多个第二锚，且发夹环衔接子与膜偶联的强度优选地大于 Y 衔接子与膜偶联的强度。

30 试剂盒可另外包含使得能够进行上文提到的任一个实施例的一或多个其它试剂或仪器。此类试剂或仪器包括以下中的一或多个：合适

缓冲液(水性溶液)、从个体获得样本的装置(如包含针的容器或仪器)、用于扩增和/或表达多核苷酸的装置,或电压或贴片钳设备。试剂可以干态形式存在于试剂盒中,使得流体样本再悬浮试剂。试剂盒还可任选地包含使得能够用本发明的方法使用试剂盒的说明书或关于何种生物体可使用所述方法的详情。

设备(或装置)

本发明还提供了一种用于表征靶分析物(例如,靶多核苷酸)的设备。所述设备包括单个或多个蛋白单体的突变体/蛋白孔、和单个或多个膜。所述蛋白单体的突变体/蛋白孔优选存在于所述膜中。孔和膜的数量优选相等。优选地,每个膜中存在单个孔。

所述设备优选地还包括用于实施实施例中方法的指令。所述设备可以是任一用于分析物分析的常规设备,例如,阵列或芯片。结合实施例的所述方法所讨论的任一实施例同样适用于所述设备。所述设备还可以包括本述试剂盒中存在的任何特征。用于实施例的设备具体可为齐碳科技基因测序仪 QNome-9604。

上述提及的现有技术以全文引用的方式并入本文。

以下实施例用以阐述本发明,但不具有限制作用。

实施例 1

在实施例中,野生型孔蛋白来自绦虫假单胞菌,并且该野生型孔蛋白的氨基酸序列是 SEQ ID NO:1,编码此氨基酸序列的核苷酸序列由 SEQ ID NO:2 所示。孔蛋白单体的突变体 1 是野生型孔蛋白在对应 SEQ ID NO:1 的 69-76 位处具有突变,具体为 69-76 位的 KPTPASSF 替换为 RPSPASAQ。包括孔蛋白单体的突变体 1 的蛋白孔为突变孔 1。蛋白单体的突变体 1 的氨基酸序列如 SEQ ID NO:24 所示,核酸序列如 SEQ ID NO:25 所示。

实施例 2

在实施例中，野生型孔蛋白来自绦虫假单胞菌，并且该野生型孔蛋白的氨基酸序列是 SEQ ID NO:1，编码此氨基酸的序列的核苷酸序列由 SEQ ID NO:2 所示。孔蛋白单体的突变体 2 是野生型孔蛋白在对应 SEQ ID NO:1 的 69-76 位处具有突变，具体为 69-76 位的 KPTPASSF 被替换为 KPGPASTK。包括孔蛋白单体的突变体 2 的蛋白孔为突变孔 2。蛋白单体的突变体 2 的氨基酸序列如 SEQ ID NO:26 所示。

10 实施例 3

在实施例中，野生型孔蛋白来自绦虫假单胞菌，并且该野生型孔蛋白的氨基酸序列是 SEQ ID NO:1，编码此氨基酸的序列的核苷酸序列由 SEQ ID NO:2 所示。孔蛋白单体的突变体 3 是野生型孔蛋白在对应 SEQ ID NO:1 的 64-79 位处具有突变，具体为 64-79 位的 QTGQYKPTPASSFSTS 替换为 LTGQYRSPASANSTA。包括孔蛋白单体的突变体 3 的蛋白孔为突变孔 3。蛋白单体的突变体 3 的氨基酸序列如 SEQ ID NO:27 所示。

实施例 4

20 在实施例中，野生型孔蛋白来自绦虫假单胞菌，并且该野生型孔蛋白的氨基酸序列是 SEQ ID NO:1，编码此氨基酸的序列的核苷酸序列由 SEQ ID NO:2 所示。孔蛋白单体的突变体 4 是野生型孔蛋白在对应 SEQ ID NO:1 的 64-79 位处具有突变，具体为 64-79 位的 QTGQYKPTPASSFSTS 替换为 LTGQYRSPASALSTA。包括孔蛋白单体的突变体 4 的蛋白孔为突变孔 4。蛋白单体的突变体 4 的氨基酸序列如 SEQ ID NO:28 所示。

实施例 5

30 在实施例中，野生型孔蛋白来自绦虫假单胞菌，并且该野生型孔蛋白的氨基酸序列是 SEQ ID NO:1，编码此氨基酸的序列的核苷酸序列

由 SEQ ID NO:2 所示。孔蛋白单体的突变体 5 是野生型孔蛋白在对应 SEQ ID NO:1 的以下位置突变：R62S、D63V、Y68F、K69R、T71S、S75A、F76Q、和 E104V。包括孔蛋白单体的突变体 5 的蛋白孔为突变孔 5。蛋白单体的突变体 5 的氨基酸序列如 SEQ ID NO:29 所示。

5

实施例 6

在实施例中，野生型孔蛋白来自绦虫假单胞菌，并且该野生型孔蛋白的氨基酸序列是 SEQ ID NO:1，编码此氨基酸的序列的核苷酸序列由 SEQ ID NO:2 所示。孔蛋白单体的突变体 6 是野生型孔蛋白在对应 SEQ ID NO:1 的 68-76、171 和 175 位处具有突变，具体为 68-76 位的 YKPTPASSF 替换为 FRPSPASAQ、171 位的 E 替换为 N 和 175 位的 D 替换为 N。包括孔蛋白单体的突变体 6 的蛋白孔为突变孔 6。蛋白单体的突变体 6 的氨基酸序列如 SEQ ID NO:30 所示。

10

实施例 7

采用 SWISS MODEL 对野生型孔蛋白进行同源建模，野生型孔蛋白单体的氨基酸由 SEQ ID NO:1 所示。图 4A 是预测蛋白结构模型的侧视图 400，其中颜色较深的部分显示的为一个蛋白单体 402。图 4B 是飘带结构模型俯视图 404，其中颜色较深的部分显示的为一个蛋白单体 406。图 4C 为飘带结构模型图 408，颜色较深部分为蛋白单体 410。

15

采用 SWISS MODEL 对突变孔 1 进行同源建模。图 5A 和图 5B 显示的是突变孔 1 的氨基酸模型图，其中加号“+”表示水分子。

20

图 6 显示了孔蛋白单体的突变体 1 的各部分尺寸，两个孔蛋白单体的突变体（即两个突变的孔蛋白单体）602 和 604 中间的缢缩区孔道直径最大为 20.1 Å，其次为 17.2 Å，最小直径为 15.9 Å。两个突变的孔蛋白单体头尾相距分别为 52.4 Å 和 36.9 Å。突变的孔蛋白单体全长 94.6 Å，突变的孔蛋白单体头距缢缩区孔道高为 41.7 Å。缢缩区孔道部分高度如图 6 所示为 12.2 Å 和 4.3 Å。

25

30

图 7 显示了孔蛋白的突变体 1 的单体氨基酸模型，放大显示的是缢缩区结构的氨基酸组成，即 Gln76、Ser74 和 Ser71。

5 实施例 8

孔蛋白的突变体 1 的负染色电镜结果如图 8 所示，由负染色 EM 结果可见：突变体 1 颗粒均匀，聚集少，可见很多明显正确的蛋白颗粒。

10 孔蛋白的突变体 1 的冷冻电镜照片和 2D 分类结果如图 9A 和 9B 所示，其中 2D 结果只展示了较好分类部分。

图 10A 和 10B 显示了局部精修傅里叶壳层相关系数(FSC)结果，可见，孔蛋白的突变体 1 不同区域分辨率结果及其最终冷冻电镜单颗粒重构分辨率为 2.2 Å。重构分辨率是基于黄金标准 FSC 0.143 标准和高分辨率噪音替代确定的。图 11 显示了孔蛋白的突变体 1 冷冻电镜三维重构 2.2 Å 分辨率的电子密度图。图 12 显示了孔蛋白的突变体 1 在 2.2 Å 下的电子密度映射图。映射图显示了通道 eye-loop 区域，并且被覆盖在最终细化的模型上。

20 孔蛋白的突变体 1 的冷冻电镜数据如表 4 所示。

表 4 孔蛋白的突变体 1 的冷冻电镜数据

数据收集	
电子显微镜设备	Titan Krios G3i
电压（kV）	300
检测器	Gatan K3
放大	64k
像素尺寸（Å）	1.08
电子剂量（e-/Å2）	51.8

散焦范围（μm）	1.6-2.0
显微照片	5361
重建	
软件	RELION3.1
细化颗粒数	835258
对称	C9
分辨率范围	2.1-2.8
重 构 分 辨 率 屏 蔽 0.143（后处理后，Å）	2.2
地 图 锐 化 B 因 子 （Å ² ）	-85

实施例 9-制备 DNA 构建体

制备两个 DNA 构建体，分别为 BS7-4C3-SE1 和 BS7-4C3-PLT。

BS7-4C3-SE1 的结构如图 13 所示，序列信息如下所示：

- 5
- a: 30*C3
- b: 5'-TTTTT TTTTT-3'（即 SEQ ID NO:3）
- c: 控速蛋白
- d: 4*C18
- e: 5'-AATGT ACTTC GTTCA GTTAC GTATT GCT-3'（即 SEQ ID
- 10
- NO:4）
- f: 5'P-GC AATAC GTAAC TGAAC GAAGT
- TCACTATCGCATTCTCATGA-3'（即 SEQ ID NO:5）
- g: 胆固醇标签
- h: 5'-TCATG AGAAT GCGAT AGTGA -3'（即 SEQ ID NO:6）
- 15
- i: 5'-AAAAA AAAAA AAAAA AAAAA AAAAA AAAAA
- AAAAA AAAAA AAGCA ATACG TAACT GAACG AAGTA CATTA
- AAAAA AAAAA AAAAA AAAA-3'（即 SEQ ID NO:7）
- j: 5'-ATCCT TTTTT TTTTT TTTTT TTTT-3'（即 SEQ ID NO:8）
- k:5'-AATGT ACTTC GTTCA GTTAC GTATT GCTTT TTTTT

TTTTT TTTTT TTT-3' (即 SEQ ID NO:9)

l: dSpacer

m: 5'-TTTTT TTTTT TTTTT TTTTT-3' (即 SEQ ID NO:10)

5 BS7-4C3-PLT 的结构如图 14 所示, 序列信息如下所示:

a: 30*C3

b: 5'-TTTTT TTTTT-3' (即 SEQ ID NO:11)

c: 控速蛋白

d: 4*C18

10 e: 5'-AATGT ACTTC GTTCA GTTAC GTATT GCT-3' (即 SEQ ID NO:12)

f: 5'-P-GC AATAC GTAAC TGAAC GAAGT
TCACTATCGCATTCTCATGA-3' (即 SEQ ID NO:13)

g: 胆固醇标签

15 h: 5'-TCATG AGAAT GCGAT AGTGA -3' (即 SEQ ID NO:14)

i: 5'-AAAAAAAAAAAAAAAAAAAAAAAAAAAAA (即 SEQ ID NO:15) /dSpacer/ AAAAAAAAAAAAAA (即 SEQ ID NO:16) /dSpacer/
AAAAAAAAAAAAAAAAATCTCTGAATCTCTGAATCTCTGAATCTCTAA
AAAAAAAAAGAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAATAAAAA
20 AAAAAAAAAAGCAATACGTAACGTAACGAAGTACATTAAAAAAAAA
A (即 SEQ ID NO:17) -3'

j: 5'

-ATCCTTTTTTTTTTTAATGTACTTCGTTACGTTACGTATTGCT -3'
(即 SEQ ID NO:18)

25 k: 5'-P-TTTTTTTTTTTTATTTTTTTTTTTTGTTTTTTTTTTTTCTT
TTTTTTTTTTAGAGATTCAGAGATTCAGAGATTCAGAGATTTTTTT
TTTTTTT (即 SEQ ID NO:19) /dSpacer/ TTTTTTTTTTTT (即 SEQ ID
NO:20) /iSpC3/ TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (即 SEQ ID
NO:21) -3'

30 C3、C18、dSpacer 及 iSpC3 是指示孔测序分辨率特征而引入的标

记 (marker) 序列。

在本实施例中，图 13 和图 14 中的 c 控速蛋白为解旋酶 Mph-MP1-E105C/A362C(具有突变 E105C/A362C)，氨基酸序列为 SEQ ID NO:22，核酸序列为 SEQ ID NO:23。

实施例 10

突变孔 1 作为蛋白孔，采用单孔测序的技术方法进行检测。在将氨基酸序列为突变体 1 的单个孔蛋白插入磷脂双分子层之后，使缓冲液(625mM KCl, 10mM HEPES pH 8.0, 50mM MgCl₂)流经该系统，以除去任何过量的突变体 1 纳米孔。将 DNA 构建体 BS7-4C3-SE1 (数据未显示) 或 BS7-4C3-PLT(1~2nM 终浓度)加入所述突变体 1 纳米孔实验系统中，混匀后，使缓冲液(625mM KCl, 10mM HEPES pH 8.0, 50mM MgCl₂)流经该系统，以除去任何过量的 DNA 构建体 BS7-4C3-SE1 或 BS7-4C3-PLT。然后将解旋酶(Mph-MP1-E105C/A362C, 15nM 终浓度)、燃料(ATP 3mM 终浓度)预混物加入所述单个突变体 1 纳米孔实验系统中，并在+180mV 电压下监测突变体 1 孔蛋白的测序情况。

突变孔 1 在±180mV 电压下开孔。图 15A 显示突变孔 1 在±180mV 电压下开孔电流及其门控特征。图 15B 显示突变孔 1 在+180mV 电压下的单链核酸过孔情况。核酸可以过孔。加入单链核酸后，向下的线显示的核酸过孔信号。

采用单孔测序技术方法，通过突变孔 1 对 DNA 构建体 BS7-4C3-PLT 进行测序，完成嵌孔后添加测序体系出现的核酸测序信号。图 16A 和 16B 示出了当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 1 移位时的示例电流轨迹。根据该信号特征，突变孔 1 可以用来核酸测序。

图 17 是将图 16B 的部分显示出电流轨迹的放大结果。具有虚线框

和箭头的图（中间图）为原始信号滤波处理后的结果(两条轨迹的 y 轴坐标=电流(pA)，x 轴坐标=时间(s))。虚线箭头指示部分显示了电流轨迹的放大结果。此单独一条信号的区域放大显示图，进一步证明突变孔 1 可以用来核酸测序。

5

实施例 11

与实施例 10 类似，实施例 11 采用突变孔 2 进行空测和过孔检测。

10

图 18A 显示突变孔 2 在±180mV 电压下开孔电流及其门控特征。
图 18B 显示突变孔 2 在+180mV 电压下的单链核酸过孔情况。核酸可以过孔。加入单链核酸后，向下的线显示的核酸过孔信号。

15

采用单孔测序技术方法，通过突变孔 2 对 DNA 构建体 BS7-4C3-PLT 进行测序，完成嵌孔后添加测序体系出现的核酸测序信号。图 19A 和 19B 示出了当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 2 移位时的示例电流轨迹。根据该信号特征可以得出突变孔 2 测序分辨率，稳定性，信号一致性等相关特征。该孔台阶清晰，跳变分布明显，具备高精度测序能力。从信号特征来看，测序信号一致性较高。

20

图 20 显示出部分电流轨迹的放大结果。具有虚线框和箭头的图为原始信号滤波处理后的结果(两条轨迹的 y 轴坐标=电流(pA)，x 轴坐标=时间(s))。虚线箭头指示部分显示了电流轨迹的放大结果。此单独一条信号的区域放大显示图，表明该突变孔对核酸测序具备高分辨率。

25

实施例 12

与实施例 10 类似，实施例 12 采用突变孔 3 进行空测和过孔检测。

30

图 21A 显示突变孔 3 在±180mV 电压下开孔电流及其门控特征。
图 21B 显示突变孔 3 在+180mV 电压下的单链核酸过孔情况。核酸可以

过孔。加入单链核酸后，向下的线显示的核酸过孔信号。

采用单孔测序技术方法，通过突变孔 3 对 DNA 构建体 BS7-4C3-PLT 进行测序，完成嵌孔后添加测序体系出现的核酸测序信号。图 22A 和 22B 示出了当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 3 移位时的示例电流轨迹。根据该信号特征，突变孔 3 可以用来核酸测序。

图 23 是将图 22A 部分显示出电流轨迹的放大结果。具有虚线框和箭头的图为原始信号滤波处理后的结果(两条轨迹的 y 轴坐标=电流(pA)，x 轴坐标=时间(s))。虚线箭头指示部分显示了电流轨迹的放大结果。此单独一条信号的区域放大显示图，进一步证明突变孔 3 可以用来核酸测序。

实施例 13

与实施例 10 类似，实施例 13 采用突变孔 4 进行空测和过孔检测。

图 24A 显示突变孔 4 在±180mV 电压下开孔电流及其门控特征。图 24B 显示突变孔 4 在+180mV 电压下的单链核酸过孔情况。核酸可以过孔。

采用单孔测序技术方法，通过突变孔 4 对 DNA 构建体 BS7-4C3-PLT 进行测序，完成嵌孔后添加测序体系出现的核酸测序信号。图 25A 和 25B 示出了当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 4 移位时的示例电流轨迹。根据该信号特征，突变孔 4 可以用来核酸测序。

图 26 是将图 25A 和 25B 实施例部分显示出电流轨迹的放大结果。具有虚线框和箭头的图为原始信号滤波处理后的结果(两条轨迹的 y 轴坐标=电流(pA)，x 轴坐标=时间(s))。虚线箭头指示部分显示了电流

轨迹的放大结果。此单独一条信号的区域放大显示图，进一步证明突变孔 4 可以用来核酸测序。

实施例 14

5 与实施例 10 类似，实施例 14 采用突变孔 5 进行空测和过孔检测。

图 27 显示突变孔 5 在±180mV 电压下开孔电流及其门控特征。

10 采用单孔测序技术方法，通过突变孔 5 对 DNA 构建体 BS7-4C3-PLT 进行测序，完成嵌孔后添加测序体系出现的核酸测序信号。图 28 示出了当解旋酶 Mph-MP1-E105C/A362C 控制 DNA 构建体 BS7-4C3-PLT 穿过突变孔 5 移位时的示例电流轨迹。根据该信号特征，突变孔 5 可以用来核酸测序。

15 图 29 是将图 28 实施例部分显示出电流轨迹的放大结果。具有虚线框和箭头的图为原始信号滤波处理后的结果(两条轨迹的 y 轴坐标＝电流(pA)，x 轴坐标＝时间(s))。虚线箭头指示部分显示了电流轨迹的放大结果。此单独一条信号的区域放大显示图，进一步证明突变孔 5 可以用来核酸测序。

20

实施例 15

与实施例 10 类似，实施例 15 采用突变孔 6 进行空测和过孔检测。

25 图 30A 显示突变孔 6 在±180mV 电压下开孔电流及其门控特征。图 30B 显示突变孔 6 在+180mV 电压下的单链核酸过孔情况。核酸可以过孔。加入单链核酸后，向下的线显示的核酸过孔信号。

30 采用单孔测序技术方法，通过突变孔 6 对 DNA 构建体 BS7-4C3-PLT 进行测序，完成嵌孔后添加测序体系出现的核酸测序信号。图 31A 和 31B 示出了当解旋酶 Mph-MP1-E105C/A362C 控制 DNA

构建体 BS7-4C3-PLT 穿过突变孔 6 移位时的示例电流轨迹。根据该信号特征可以得出突变孔 6 测序分辨率，稳定性，信号一致性等相关特征。该孔台阶清晰，跳变分布明显，具备高精度测序能力。从信号特征来看，测序信号一致性较高。

5

图 32 是将图 31A 部分显示出电流轨迹的放大结果。具有虚线框和箭头的图为原始信号滤波处理后的结果(两条轨迹的 y 轴坐标=电流(pA)，x 轴坐标=时间(s))。虚线箭头指示部分显示了电流轨迹的放大结果。此单独一条信号的区域放大显示图，表明该突变孔对核酸测序具备高分辨率。

10

实施例 16

将含有孔蛋白单体的突变体 1 核酸序列 (SEQ ID NO:25) 的重组质粒通过热击法转化到 BL21(DE3)感受态细胞，加入 0.5ml LB 培养基经 30℃ 培养 1h 后取适量菌液涂布于氨苄抗性固体 LB 平板，37℃ 过夜培养，次日挑取单克隆菌落，接种至 50ml 含有氨苄抗性的液体 LB 培养基中 37℃ 培养过夜。按 1% 的接种量转接至氨苄抗性的 TB 液体培养基中进行扩大培养，37℃、220rpm 条件下培养，并连续不断的测量其 OD600 值。当 OD600=2.0-2.2 时，将 TB 培养基中的培养液冷却至 16℃，并添加异丙基硫代半乳糖苷 (Isopropyl β-D-Thiogalactoside, IPTG) 诱导表达，使得终浓度达到 0.015mM。诱导表达 20-24h 后，离心收集菌体。菌体用破碎缓冲液重悬后高压破碎，通过 Ni-NTA 亲和层析方法进行纯化，收集目的洗脱样品。孔蛋白单体的突变体 2-6 按如上方法纯化得到。

15

20

25

示例性的，图 33 示出了突变体 1 的蛋白纯化结果，1-4 泳道显示的是分离的不同组分的 SDS-PAGE 电泳检测结果。图 34 示出了突变体 1 蛋白尺寸排阻色谱 (SEC) 的示例 (25mL superose-6 GE healthcare, x 轴标记=洗脱体积 (mL)，Y 轴标记=吸光度 (mAu))。

30

权 利 要 求 书

1. 一种孔蛋白单体的突变体，其中所述孔蛋白单体的突变体的氨基酸包括 SEQ ID NO:1 所示的序列或与其具有至少 99%、98%、97%、
5 96%、95%、90%、80%、70%、60%或 50%同一性的序列，并且所述孔蛋白单体的突变体的氨基酸包括在对应 SEQ ID NO:1 的 T71、S75、F76 的一个或多个位置处的突变。

2. 根据权利要求 1 所述的孔蛋白单体的突变体，所述孔蛋白单体的突变体的氨基酸包括在对应 SEQ ID NO:1 的 62-175、62-104、68-175、
10 64-79、71-76、或 69-76 的一个或多个位置处的突变。

3. 根据权利要求 1 或 2 所述的孔蛋白单体的突变体，所述孔蛋白单体的突变体的氨基酸包括：

15 (1) 对应 SEQ ID NO:1 的 K69、P70、T71、P72、A73、S74、S75、和 F76 的一个或多个位置处具有氨基酸的插入、缺失和/或替换；(2) 对应 SEQ ID NO:1 的 Q64、T65、G66、Q67、Y68、K69、P70、T71、P72、A73、S74、S75、F76、S77、T78、和 S79 的一个或多个位置处具有氨基酸的插入、缺失和/或替换；(3) 对应 SEQ ID NO:1 的 R62、
20 D63、Y68、K69、T71、S75、F76、和 E104 的一个或多个位置处具有氨基酸的插入、缺失和/或替换；或者(4) 对应 SEQ ID NO:1 的 Y68、K69、P70、T71、P72、A73、S74、S75、F76、E171 和 D175 的一个或多个位置处具有氨基酸的插入、缺失和/或替换。

25 4. 根据前述权利要求任一项所述的孔蛋白单体的突变体，其中所述 SEQ ID NO:1 所示的序列来源于绦虫假单胞菌。

5. 根据前述权利要求任一项所述的孔蛋白单体的突变体，其中所述孔蛋白单体的突变体的氨基酸突变选自以下：

30 (a) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为

5 $M_1M_2M_3M_4M_5M_6M_7M_8$, 其中 M_1 选自 R、K、H 中的 0 至 3 种; M_2 选自 P 中的 0 至 1 种; M_3 选自 S、G、C、U、T、M、A、V、L、I 中的 0 至 10 种; M_4 选自 P 中的 0 至 1 种; M_5 选自 A、G、V、L、I 中的 0 至 5 种; M_6 选自 S、C、U、T、M 中的 0 至 5 种; M_7 选自 A、T、G、V、L、I、S、C、U、M 中的 0 至 10 种; M_8 选自 Q、D、E、N、K、H、R 中的 0 至 7 种;

10 (b) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 $M_9M_{10}M_{11}M_{12}M_{13}M_{14}M_{15}M_{16}M_{17}M_{18}M_{19}M_{20}M_{21}M_{22}M_{23}M_{24}$, 其中 M_9 选自 L、G、A、V、I 中的 0 至 5 种; M_{10} 选自 T、S、C、U、M 中的 0 至 5 种; M_{11} 选自 G、A、V、L、I 中的 0 至 5 种; M_{12} 选自 Q、D、E、N 中的 0 至 4 种; M_{13} 选自 Y、F、W 中的 0 至 3 种; M_{14} 选自 R、H、K 中的 0 至 3 种; M_{15} 选自 P 中的 0 至 1 种; M_{16} 选自 S、C、U、T、M 中的 0 至 5 种; M_{17} 选自 P 中的 0 至 1 种; M_{18} 选自 A、G、V、L、I 中的 0 至 5 种; M_{19} 选自 S、C、U、T、M 中的 0 至 5 种; M_{20} 选自 A、G、V、L、I 中的 0 至 5 种; M_{21} 选自 N、D、E、Q、L、G、A、V、I 中的 0 至 9 种; M_{22} 选自 S、C、U、T、M 中的 0 至 5 种; M_{23} 选自 T、S、C、U、M 中的 0 至 5 种; M_{24} 选自 A、G、V、L、I 中的 0 至 5 种;

15 (c) 对应 SEQ ID NO:1 的 62 位突变为 S、C、U、T、M 中的 0 至 5 种; 63 位突变为 V、G、A、L、I 中的 0 至 5 种; 68 位突变为 F、W 中的 0 至 2 种; 69 位突变为 R、H 中的 0 至 2 种; 71 位突变为 S、C、U、M 中的 0 至 4 种; 75 位突变为 A、G、V、L、I 中的 0 至 5 种; 76 位突变为 Q、D、E、N 中的 0 至 4 种; 104 位突变为 V、G、A、L、I 中的 0 至 5 种; 和

20 (d) 对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 $M_{25}M_{26}M_{27}M_{28}M_{29}M_{30}M_{31}M_{32}M_{33}$, 并且 171 位突变为 N、D、Q 中的 0 至 3 种; 175 位突变为 N、E、Q 中的 0 至 3 种, 其中 M_{25} 选自 F、Y、W 中的 0 至 3 种; M_{26} 选自 R、H、K 中的 0 至 3 种; M_{27} 选自 P 中的 0 至 1 种; M_{28} 选自 S、C、U、T、M 中的 0 至 5 种; M_{29} 选自 P 中的 0 至 1 种; M_{30} 选自 A、G、V、L、I 中的 0 至 5 种; M_{31} 选自 S、C、U、

25

30

T、M 中的 0 至 5 种；M₃₂ 选自 A、G、V、L、I 中的 0 至 5 种；M₃₃ 选自 Q、D、E、N 中的 0 至 4 种。

5 6. 根据前述权利要求任一项所述的孔蛋白单体的突变体，其中所述孔蛋白单体的突变体的氨基酸突变选自以下：

(a) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 RPSPASAQ；

(b) 对应 SEQ ID NO:1 的 69-76 位的氨基酸 KPTPASSF 突变为 KPGPASTK；

10 (c) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 LTGQYRPSPASANSTA；

(d) 对应 SEQ ID NO:1 的 64-79 位的氨基酸 QTGQYKPTPASSFSTS 突变为 LTGQYRPSPASALSTA；

15 (e) 对应 SEQ ID NO:1 的 R62S、D63V、Y68F、K69R、T71S、S75A、F76Q、和 E104V；和

(f) 对应 SEQ ID NO:1 的 68-76 位的氨基酸 YKPTPASSF 突变为 FRPSPASAQ，以及对应 SEQ ID NO:1 的 E171N 和 D175N。

20 7. 根据前述权利要求任一项所述的孔蛋白单体的突变体，其中所述孔蛋白单体的突变体的氨基酸序列包括 SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:27、SEQ ID NO:28、SEQ ID NO:29 或 SEQ ID NO:30，或由其组成。

25 8. 一种蛋白孔，包括至少一个前述权利要求任一项所述的孔蛋白单体的突变体。

9. 根据权利要求 8 所述的蛋白孔，其中所述蛋白孔包括至少两个所述的孔蛋白单体的突变体。

30 10. 根据权利要求 8-9 任一项所述的蛋白孔，其中所述蛋白孔的缢

缩区孔道直径为 0.7nm-2.2nm、0.9nm-1.6nm、1.4nm-1.6nm 或 15.9Å-20.1Å。

5 11. 一种用于表征靶分析物的复合物，其特征在于：包括权利要求 8-10 任一项所述的蛋白孔及与其结合的控速蛋白。

12. 一种核酸，其编码权利要求 1-7 任一项所述的孔蛋白单体的突变体、权利要求 8-10 任一项所述的蛋白孔、或权利要求 11 所述的复合物。

10

13. 根据权利要求 12 所述的核酸，其中孔蛋白单体的核苷酸序列为 SEQ ID NO:2 所示的序列。

14. 包含权利要求 12-13 任一项所述的核酸的载体或遗传工程化的宿主细胞。

15

15. 权利要求 1-7 任一项所述的孔蛋白单体的突变体、权利要求 8-10 任一项所述的蛋白孔、权利要求 11 所述复合物、权利要求 12-13 任一所述核酸、或权利要求 14 所述载体或宿主细胞、在检测靶分析物存在、不存在或一个或多个特征或制备检测靶分析物存在、不存在或一个或多个特征的产品中的应用。

20

16. 一种产生蛋白孔或其多肽的方法，包括用包含权利要求 14 所述的载体转化权利要求 14 所述的宿主细胞，诱导所述宿主细胞表达权利要求 8-10 任一所述的蛋白孔或其多肽。

25

17. 一种用于确定靶分析物存在、不存在或一个或多个特征的方法，包括：

a. 使靶分析物与权利要求 8-10 任一项所述的蛋白孔、权利要求 11 所述复合物、或权利要求 11 所述复合物中的所述蛋白孔接触，使得

30

所述靶分析物相对于所述蛋白孔移动；以及

b. 在所述靶分析物相对于所述蛋白孔移动时获取一个或多个测量值，从而确定所述靶分析物的存在、不存在或一个或多个特征。

5 18. 根据权利要求 17 所述的方法，其中所述方法包括：

所述靶分析物与存在于膜中的所述蛋白孔相互作用从而使得所述靶分析物相对所述蛋白孔移动。

10 19. 一种用于确定靶分析物存在、不存在或一个或多个特征的试剂盒，包括权利要求 1-7 任一项所述的孔蛋白单体的突变体、权利要求 8-10 任一项所述的蛋白孔、权利要求 11 所述的复合物、权利要求 12-13 任一项所述的核酸、或权利要求 14 所述的载体或宿主，和权利要求 18 所述的膜的组分。

15 20. 一种用于确定靶分析物存在、不存在或一个或多个特征的装置，包括权利要求 8-10 任一项所述的蛋白孔或权利要求 11 所述复合物，和权利要求 18 所述的膜。

20 21. 根据权利要求 15-20 中任一所述的应用、方法、试剂盒或装置，其中所述靶分析物包括多糖、金属离子、无机盐、聚合物、氨基酸、肽、蛋白、核苷酸、寡核苷酸、多核苷酸、染料、药物、诊断剂、爆炸物或环境污染物；

优选地，所述靶分析物包括多核苷酸，

25 更优选地，所述多核苷酸包括 DNA 或 RNA；和/或，所述一个或多个特征选自 (i) 所述多核苷酸的长度；(ii) 所述多核苷酸的一致性；(iii) 所述多核苷酸的序列；(iv) 所述多核苷酸的二级结构和 (v) 所述多核苷酸是否经修饰；和/或，所述复合物中所述控速蛋白包括多核苷酸结合蛋白。

30