

一种 Mmup 单体变体及其应用

技术领域

本发明涉及核酸特性的表征技术领域，具体涉及一种 Mmup 单体变体、包含 Mmup 单体变体的孔蛋白及构建体，及应用所述 Mmup 单体变体或孔蛋白进行表征目标多核苷酸的方法。

背景技术

纳米孔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场力驱动下通过该纳米孔，当多核苷酸通过纳米孔易位时，会由于物理占位效应而产生相对应的阻塞电流，对产生的不同信号实时读取进而分析得到多核苷酸序列信息的基因测序技术。纳米孔测序技术具有以下优势：在无需扩增的情况下，即可简便的建库；阅读速度快，单链分子的阅读速度能够达到每小时数万个碱基；阅读长度更长，通常可以达到数千个碱基；可以直接进行甲基化的 DNA 或 RNA 的测量。

但是，每一个或一系列核苷酸在电场力的作用下通过纳米孔蛋白时，会产生一种特定的阻塞电流，此时记录的电流信号与多核苷酸的序列虽是对应关系，却通常是 3-4 个核苷酸控制某些级别的电流水平，所以仍然需要提高精确度。目前，可以通过改变多核苷酸结构、在纳米孔处的持续时间以及开发新的纳米孔从而控制多核苷酸的易位来提高精确度。

例如：专利 US20150065354A1 公开了一种使用 XPD 解旋酶表征目标多核苷酸的方法，所述方法利用孔和 XPD 解旋酶。该发明所述的 XPD 解旋酶可以控制目标多核苷酸穿过所述孔的运动。

专利 US20170268055A1 公开了一种用于多核苷酸测序的组合物和方法，该方法利用目标多核苷酸通过孔的易位分步易位步骤来表征目标多核苷酸，包括表征多核苷酸的序列的方法和组合物。

专利 CN106103741A 公开了将一个或多个多核苷酸结合蛋白连接到靶多核苷酸的方法，该发明还涉及表征靶多核苷酸的新方法。

专利 CN102216783B 公开了一种耻垢分枝杆菌孔蛋白（Msp）纳米孔以及使用该纳米

孔进行测序，其中，将野生型 Msp 的 90 或 91 位点突变以提高分析物在测序时的电导，并降低分析物在测序中的转位速度。

然而，现有技术中并未提到 Mmup 单体变体及在测序中的应用，且现有技术中可以用于测序的孔蛋白种类很少。

因此，本发明进一步提供了一种新的纳米孔蛋白，通过将不能被用于测序的野生型 Mmup 蛋白进行突变制备了 Mmup 单体变体，并证实了该 Mmup 单体变体在测序中的功能。

发明内容

本发明证明了制备 Mmup 突变型蛋白特定位点突变的 Mmup 单体变体，可以用于纳米孔测序，然而野生型 Mmup 单体并没有该功能。并且，应用本发明所述的孔蛋白进行纳米孔测序，可以明显的看出各种不同核苷酸电流信号的差别，具备较高的测序精确度。

本发明所述的“Mmup”来源于产粘液分枝杆菌。优选的，所述的“Mmup”来源于 *Mycobacterium mucogenicum*。

具体的，本发明的第一方面，提供了一种 Mmup 单体变体，所述的 Mmup 单体变体包含 SEQ ID NO: 1 第 91-99 位任意一个或多个氨基酸突变的氨基酸序列。

优选的，所述的 Mmup 单体变体包含第 91 位天冬氨酸 (D) 和/或第 99 位丙氨酸 (A) 的突变。

在本发明的一个具体实施方式中，所述的 Mmup 单体变体包含第 91 位天冬氨酸 (D) 的突变。

在本发明的另一个具体实施方式中，所述的 Mmup 单体变体包含第 99 位丙氨酸 (A) 的突变。

在本发明的还一个具体实施方式中，所述的 Mmup 单体变体包含第 91 位天冬氨酸 (D) 和第 99 位丙氨酸 (A) 的突变。

优选的，所述的 Mmup 单体变体包含至少一种如下突变：

D91 突变为：脯氨酸 (P)、色氨酸 (W)、精氨酸 (R)、谷氨酰胺 (Q)、赖氨酸 (K)、苯丙氨酸 (F)、丝氨酸 (S)、天冬酰胺 (N)、半胱氨酸 (C)、异亮氨酸 (I)、

亮氨酸（L）或缬氨酸（V），或者，非天然氨基酸；或者，

A99 突变为：脯氨酸（P）、苯丙氨酸（F）、异亮氨酸（I）、亮氨酸（L）、缬氨酸（V）、赖氨酸（K）或精氨酸（R），或者，非天然氨基酸。

进一步优选的，所述的 Mmup 单体变体包含 D91K 和/或 A99K 的突变。

在本发明的一个具体实施方式中，所述的 Mmup 单体变体包含 D91K 的突变。

在本发明的另一个具体实施方式中，所述的 Mmup 单体变体包含 A99K 的突变。

在本发明的还一个实施方式中，所述的 Mmup 单体变体包含 D91K 和 A99K 的突变。

优选的，所述的 Mmup 单体变体还包含 SEQ ID NO: 1 第 80-90 位和/或 100-120 位任意一个或多个氨基酸突变的氨基酸序列。

进一步优选的，所述的 Mmup 单体变体还包含 SEQ ID NO: 1 第 1-79 位和/或 121-186 位任意一个或多个氨基酸突变的氨基酸序列。

优选的，所述的 Mmup 单体变体还包含第 89 位亮氨酸（L）的突变、第 110 位天冬酰胺（N）的突变、第 120 位天冬氨酸（D）的突变、第 136 位天冬酰胺（N）的突变或第 141 位丝氨酸（S）的突变中的一种或两种以上的组合。

进一步优选的，所述的 Mmup 单体变体包含至少一种如下突变：

L89 突变为：天冬酰胺（N）、丙氨酸（A）、谷氨酰胺（Q）、甘氨酸（G）、丝氨酸（S）、苏氨酸（T）、赖氨酸（K）或脯氨酸（P），或者，非天然氨基酸；或者，

N110 突变为：脯氨酸（P）、丙氨酸（A）、异亮氨酸（I）或亮氨酸（L），或者，非天然氨基酸；或者，

D120 突变为：精氨酸（R）、赖氨酸（K）、天冬酰胺（N）、谷氨酰胺（Q）、丙氨酸（A）、丝氨酸（S）、甘氨酸（G）或苏氨酸（T），或者，非天然氨基酸；或者，

N136 突变为：精氨酸（R）、赖氨酸（K）、谷氨酰胺（Q）、丙氨酸（A）、丝氨酸（S）或苏氨酸（T），或者，非天然氨基酸；或者，

S141 突变为：精氨酸（R）、赖氨酸（K）、谷氨酰胺（Q）、丙氨酸（A）、丝氨酸（S）或苏氨酸（T），或者，非天然氨基酸。

优选的，所述的 Mmup 单体变体还可以包含除上述突变类型以外的其他突变型，只要所述的突变不影响多核苷酸通过孔蛋白时对不同多核苷酸的区分即可。

优选的，所述的 Mmup 单体变体还可以包括引入半胱氨酸的突变，以连接用于测序的分子，例如核酸结合蛋白等。

优选的，所述的 Mmup 单体变体可以只包含孔蛋白形成结构域的收缩区和环形区片段序列，且保留孔形成活性。可以去除多余的残基或者增加其他氨基酸残基，且保留孔形成活性。所述的片段长度可以为至少 12、20、40、50、100 或 150 个氨基酸。

优选的，所述的 Mmup 单体变体可以是经过修饰的，以便于鉴定或纯化。例如：通过添加天冬氨酸残基（asp 标签）、链霉亲和素标签、flag 标签或组氨酸残基（His 标签）。

优选的，所述 Mmup 单体变体可带有显示标记物。例如：荧光分子、放射性同位素 ^{125}I 、放射性同位素 ^{35}S 、多核苷酸、生物素、抗原或抗体。

优选的，所述的 Mmup 单体变体还包括分子发动机。优选的，所述的分子发动机是酶。进一步优选的，所述的酶是聚合酶、外切核酸酶或 Klenow 片段。

本发明的第二方面，提供了一种包含至少一个本发明任一所述 Mmup 单体变体的构建体。其中，所述的构建体保留了形成孔的能力。

优选的，所述的构建体包含 1-50 个 Mmup 单体变体，其中，所述的 Mmup 单体变体相同或不同。具体的，所述的构建体包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49 或 50 个所述 Mmup 单体变体，其中，所述的 Mmup 单体变体相同或不同。

进一步优选的，所述的构建体包含 1-20 个 Mmup 单体变体，其中，所述的 Mmup 单体变体相同或不同。具体的，所述的构建体包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个 Mmup 单体变体，其中，所述的 Mmup 单体变体相同或不同。

优选的，所述的构建体还包含野生型 Mmup 单体。

进一步优选的,所述的构建体包含 1-50 个野生型 Mmup 单体。具体的,所述的构建体包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49 或 50 个野生型 Mmup 单体。

在本发明的一个具体实施方式中,所述的构建体包含 1-20 个野生型 Mmup 单体。具体的,所述的构建体包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个野生型 Mmup 单体。

最为优选的,所述的构建体包含 4-10 个相同或不同的 Mmup 单体变体。

在本发明的一个具体实施方式中,所述的构建体包含 4、6、8、10 个相同或不同的 Mmup 单体变体。

优选的,所述的 Mmup 单体变体与 Mmup 单体变体、野生型 Mmup 单体与野生型 Mmup 单体、Mmup 单体变体与野生型 Mmup 单体之间通过共价连接。

优选的,所述的 Mmup 单体变体与 Mmup 单体变体、野生型 Mmup 单体与野生型 Mmup 单体、Mmup 单体变体与野生型 Mmup 单体之间是遗传上融合的。

本发明的第三方面,提供了一种包含至少一个 Mmup 单体变体的孔蛋白,所述的 Mmup 单体变体包含 SEQ ID NO: 1 第 91-99 位任意一个或多个氨基酸突变的氨基酸序列,所述的突变导致当多聚核苷酸单链通过所述包含至少一个 Mmup 单体变体的孔蛋白时,产生由于不同种类的核苷酸的物理或化学性质的不同而造成的孔内电阻的差异。

优选的,所述的突变导致电荷性质或者氨基酸疏水性性质变化。

优选的,所述的电阻的差异是指可用于表征多核苷酸的特征,所述的特征包括多核苷酸的来源、长度、大小、分子量、同一性、序列、二级结构、浓度或目标多核苷酸是否被修饰。进一步优选的,所述的电阻的差异是指可用于表征多核苷酸的序列特征,即所述的孔蛋白可以用于测序,精确的区分多核苷酸的不同碱基。

优选的,所述的多核苷酸可以是天然存在的或人工合成的。进一步优选的,所述的多核苷酸可以是天然的 DNA、RNA 或者经过修饰的 DNA 或 RNA。

更进一步优选的，所述的目标多核苷酸中的一个或多个核苷酸可以是经过修饰的，例如甲基化、氧化、损伤、脱碱基的、蛋白标记、带有标签或多核苷酸序列中间连接一段间隔物。

更进一步优选的，所述人工合成的核酸选自肽核酸（PNA）、甘油核酸（GNA）、苏糖核酸（TNA）、锁定核酸（LNA）或其他具有核苷侧链的合成聚合物。

优选的，所述的目标多核苷酸为单链、双链或至少一部分是双链的。

优选的，所述的 Mmup 单体变体包含第 91 位天冬氨酸（D）和/或第 99 位丙氨酸（A）的突变。

在本发明的一个具体实施方式中，所述的 Mmup 单体变体包含第 91 位天冬氨酸（D）的突变。

在本发明的另一个具体实施方式中，所述的 Mmup 单体变体包含第 99 位丙氨酸（A）的突变。

在本发明的还一个具体实施方式中，所述的 Mmup 单体变体包含第 91 位天冬氨酸（D）和第 99 位丙氨酸（A）的突变。

进一步优选的，所述的 Mmup 单体变体包含至少一种如下突变：

D91 突变为：脯氨酸（P）、色氨酸（W）、精氨酸（R）、谷氨酰胺（Q）、赖氨酸（K）、苯丙氨酸（F）、丝氨酸（S）、天冬酰胺（N）、半胱氨酸（C）、异亮氨酸（I）、亮氨酸（L）或缬氨酸（V），或者，非天然氨基酸；或者，

A99 突变为：脯氨酸（P）、苯丙氨酸（F）、异亮氨酸（I）、亮氨酸（L）、缬氨酸（V）、赖氨酸（K）或精氨酸（R），或者，非天然氨基酸。

更进一步优选的，所述的 Mmup 单体变体包含 D91K 和/或 A99K 的突变。

在本发明的一个具体实施方式中，所述的 Mmup 单体变体包含 D91K 和 A99K 的突变。

优选的，所述的 Mmup 单体变体还包含 SEQ ID NO: 1 第 80-90 位和/或 100-120 位任意一个或多个氨基酸突变的氨基酸序列。

进一步优选的，所述的 Mmup 单体变体还包含 SEQ ID NO: 1 第 1-79 位和/或 121-186

位任意一个或多个氨基酸突变的氨基酸序列。

优选的，所述的 Mmup 单体变体还包含第 89 位亮氨酸 (L) 的突变、第 110 位天冬酰胺 (N) 的突变、第 120 位天冬氨酸 (D) 的突变、第 136 位天冬酰胺 (N) 的突变或第 141 位丝氨酸 (S) 的突变中的一种或两种以上的组合。

进一步优选的，所述的 Mmup 单体变体还包含至少一种如下突变：

L89 突变为：天冬酰胺 (N)、丙氨酸 (A)、谷氨酰胺 (Q)、甘氨酸 (G)、丝氨酸 (S)、苏氨酸 (T)、赖氨酸 (K) 或脯氨酸 (P)，或者，非天然氨基酸；或者，

N110 突变为：脯氨酸 (P)、丙氨酸 (A)、异亮氨酸 (I) 或亮氨酸 (L)，或者，非天然氨基酸；或者，

D120 突变为：精氨酸 (R)、赖氨酸 (K)、天冬酰胺 (N)、谷氨酰胺 (Q)、丙氨酸 (A)、丝氨酸 (S)、甘氨酸 (G) 或苏氨酸 (T)，或者，非天然氨基酸；或者，

N136 突变为：精氨酸 (R)、赖氨酸 (K)、谷氨酰胺 (Q)、丙氨酸 (A)、丝氨酸 (S) 或苏氨酸 (T)，或者，非天然氨基酸；或者，

S141 突变为：精氨酸 (R)、赖氨酸 (K)、谷氨酰胺 (Q)、丙氨酸 (A)、丝氨酸 (S) 或苏氨酸 (T)，或者，非天然氨基酸。

优选的，所述包含至少一个 Mmup 单体变体的孔蛋白包含至少 1-50 个 Mmup 单体变体，其中，所述的 Mmup 单体变体相同或不同。

具体的，所述包含至少一个 Mmup 单体变体的孔蛋白包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49 或 50 个所述 Mmup 单体变体，所述的 Mmup 单体变体相同或不同。

进一步优选的，所述包含至少一个 Mmup 单体变体的孔蛋白包含至少 1-20 个 Mmup 单体变体，其中，所述的 Mmup 单体变体相同或不同。

具体的，所述包含至少一个 Mmup 单体变体的孔蛋白包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个所述 Mmup 单体变体，所述的 Mmup

单体变体相同或不同。

优选的，所述包含至少一个 Mmup 单体变体的孔蛋白还包含野生型 Mmup 单体。

优选的，所述包含至少一个 Mmup 单体变体的孔蛋白包含 1-50 个野生型 Mmup 单体。

具体的，所述包含至少一个 Mmup 单体变体的孔蛋白包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49 或 50 个野生型 Mmup 单体。

进一步优选的，所述包含至少一个 Mmup 单体变体的孔蛋白包含 1-20 个野生型 Mmup 单体。

具体的，所述包含至少一个 Mmup 单体变体的孔蛋白包含 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19 或 20 个野生型 Mmup 单体。

在本发明的一个具体实施方式中，所述包含至少一个 Mmup 单体变体的孔蛋白包含 4-10 个相同或不同的 Mmup 单体变体。

在本发明的一个具体实施方式中，所述包含至少一个 Mmup 单体变体的孔蛋白包含 4、6、8、10 个相同或不同的 Mmup 单体变体。

优选的，所述的 Mmup 单体变体与 Mmup 单体变体、野生型 Mmup 单体与野生型 Mmup 单体、Mmup 单体变体与野生型 Mmup 单体之间通过共价连接。

优选的，所述的 Mmup 单体变体与 Mmup 单体变体、野生型 Mmup 单体与野生型 Mmup 单体、Mmup 单体变体与野生型 Mmup 单体之间是遗传上融合的。

优选的，所述包含在包含至少一个 Mmup 单体变体的孔蛋白中的 Mmup 单体变体相同或不同。例如：所述包含至少一个 Mmup 单体变体的孔蛋白可以包含八个相同或者不同的 Mmup 单体变体。优选的，所述包含至少一个 Mmup 单体变体的孔蛋白包含一个 Mmup 单体变体和七个相同单体，其中，所述 Mmup 单体变体与相同单体不同。或者，所述包含至少一个 Mmup 单体变体的孔蛋白包含两个相同或不同的 Mmup 单体变体和六个相同单体，其中，所述的 Mmup 单体变体与相同单体不同。或者，所述包含至少一个 Mmup 单体变体

的孔蛋白包含三个相同或不同的 Mmup 单体变体和五个相同单体, 其中, 所述的 Mmup 单体变体与相同单体不同。或者, 所述包含至少一个 Mmup 单体变体的孔蛋白包含四个相同或不同的 Mmup 单体变体和四个相同单体, 其中, 所述的 Mmup 单体变体与相同单体不同。或者, 所述包含至少一个 Mmup 单体变体的孔蛋白包含五个相同或不同的 Mmup 单体变体和三个相同单体, 其中, 所述的 Mmup 单体变体与相同单体不同。或者, 所述包含至少一个 Mmup 单体变体的孔蛋白包含六个相同或不同的 Mmup 单体变体和两个相同单体, 其中, 所述的 Mmup 单体变体与相同单体不同。或者, 所述包含至少一个 Mmup 单体变体的孔蛋白包含七个相同或不同的 Mmup 单体变体和一个单体, 其中, 所述的 Mmup 单体变体与一个单体不同。

优选的, 所述包含至少一个 Mmup 单体变体的孔蛋白包含八个相同或不同的所述的 Mmup 单体变体。

优选的, 所述包含至少一个 Mmup 单体变体的孔蛋白可以为同源的或异源的。

优选的, 所述包含至少一个 Mmup 单体变体的孔蛋白包含收缩区和环形区。

优选的, 所述包含至少一个 Mmup 单体变体的孔蛋白的收缩区孔道直径小于野生型 Mmup 单体组成的孔蛋白的收缩区孔道直径。进一步优选的, 所述包含至少一个 Mmup 单体变体的孔蛋白的收缩区孔道直径小于 15.5Å、15.4Å、15.3Å、15.2Å、15.1Å、15.0Å、14.9Å、14.8Å、14.7Å、14.6Å、14.5Å、14.4Å、14.3Å、14.2Å、14.1Å、14.0Å、13.9Å、13.8Å、13.7Å、13.6Å、13.5Å、13.4Å 或 13.3Å; 更进一步优选的, 所述包含至少一个 Mmup 单体变体的孔蛋白的收缩区孔道直径小于 13.5Å、13.4Å、13.3Å、13.2Å、13.1Å、13.0Å、12.9Å、12.8Å、12.7Å、12.6Å 或 12.5Å。

在本发明的一个具体实施方式中, 所述包含至少一个 Mmup 单体变体的孔蛋白的收缩区孔道直径约等于 13.3Å 或 13.0Å。

优选的, 所述的 Mmup 单体变体还可以包含除上述突变类型以外的其他突变型, 只要所述的突变不影响多核苷酸通过孔蛋白时对不同多核苷酸的区分即可。

优选的, 所述的单体变体还可以包括引入半胱氨酸的突变, 以连接用于测序的分子,

例如核酸结合蛋白等。

优选的，所述的 Mmup 单体变体可以只包含孔蛋白形成结构域的收缩区和环形区片段序列，且保留孔形成活性。可以去除多余的残基或者增加其他氨基酸残基，且保留孔形成活性。所述的片段长度可以为至少 12、20、40、50、100 或 150 个氨基酸。

优选的，所述的 Mmup 单体变体可以是经过修饰的，以便于鉴定或纯化。例如：通过添加天冬氨酸残基（asp 标签）、链霉亲和素标签、flag 标签或组氨酸残基（His 标签）。

优选的，所述 Mmup 单体变体可带有显示标记物。例如：荧光分子、放射性同位素 ^{125}I 、放射性同位素 ^{35}S 、多核苷酸、生物素、抗原或抗体。

优选的，所述的 Mmup 单体变体还包括分子发动机。优选的，所述的分子发动机是酶。进一步优选的，所述的酶是聚合酶、外切核酸酶或 Klenow 片段。

优选的，所述包含至少一个 Mmup 单体变体的孔蛋白还包括帽状体形成区和/或桶状体形成区域。

优选的，所述包含至少一个 Mmup 单体变体的孔蛋白允许水合离子在施加的电势的驱动下从膜的一侧流向膜的另一层。其中，所述的膜为双层膜，进一步优选为脂质双层膜。

本发明的第四方面，提供了一种编码本发明任一所述包含至少一个 Mmup 单体变体的孔蛋白、本发明任一所述的 Mmup 单体变体或者本发明任一所述构建体的核苷酸序列。

优选的，所述的编码所述 Mmup 单体变体的核苷酸序列与 SEQ ID NO: 2 或 SEQ ID NO: 11 所示序列具有 70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.5%、99.9% 同源性，且编码所述 Mmup 单体变体的核苷酸的序列。

在本发明的一个具体实施方式中，所述的编码所述 Mmup 单体变体的核苷酸序列为 SEQ ID NO: 11 所示。

本发明的第五方面，提供了一种包含编码本发明所述的 Mmup 单体变体、本发明所述的孔蛋白或者本发明所述构建体的核苷酸序列的载体。

优选的，所述的载体可以提供有复制起点、任选用于表达所述核苷酸序列的启动子以及任选所述启动子的调节信号基因的质粒、病毒或噬菌体载体。所述载体可含有一个或多个

个选择性标记基因，例如四环素抗性基因。启动子和其他表达调节信号可被选择为与所述表达载体设计用于的宿主细胞相容。所述的启动子选自 T7、trc、lac、ara 或 λ L 启动子。

本发明所述的 Mmup 单体变体可以采用化学合成或者重组方式制备，优选为重组方式制备。

优选的，所述的载体包含与编码本发明任一所述包含至少一个 Mmup 单体变体的孔蛋白、本发明任一所述的 Mmup 单体变体或者本发明任一所述构建体的核苷酸序列可操作性连接的启动子。

进一步优选的，所述的启动子为诱导型启动子或组成型启动子，其中，所述的诱导型启动子为乙酰胺诱导型启动子。

优选的，所述编码包含至少一个 Mmup 单体变体的孔蛋白的核苷酸序列包含至少一条编码 Mmup 单体变体的核苷酸序列。

进一步优选的，所述编码包含至少一个 Mmup 单体变体的孔蛋白的核苷酸序列还包含至少一条编码野生型 Mmup 单体的核苷酸序列。

更进一步优选的，所述编码 Mmup 单体变体的核苷酸序列与编码 Mmup 单体变体的核苷酸序列，编码 Mmup 单体变体的核苷酸序列与编码野生型 Mmup 单体的核苷酸序列，或者，编码野生型 Mmup 单体的核苷酸序列与编码野生型 Mmup 单体的核苷酸序列之间通过编码氨基酸连接体序列连接。

本发明的第六方面，提供了一种表达本发明任一所述的 Mmup 单体变体、本发明任一所述构建体或本发明任一所述包含至少一个 Mmup 单体变体的孔蛋白的突变细菌，所述的细菌包含：

(a) 野生型 Mmup 单体的缺失；和 (b) 本发明任一所述的载体。

优选的，所述的细菌包含与编码 Mmup 单体变体、包含 Mmup 单体变体的构建体或者包含 Mmup 单体变体的孔蛋白的核苷酸序列可操作性连接的启动子的载体。

进一步优选的，所述的 Mmup 单体变体包括 Mmup 单体变体的旁系同源物或同系物。

进一步优选的，所述的包含 Mmup 单体变体的构建体包括 Mmup 单体变体的旁系同源

物或同系物构建体或单体。

进一步优选的，所述的包含 Mmup 单体变体的孔蛋白包括 Mmup 单体变体的旁系同源物或同系物孔蛋白或单体。

优选的，所述的细菌还可以包含与编码野生型 Mmup 单体、包含野生型 Mmup 单体的构建体或者包含野生型 Mmup 单体的孔蛋白的核苷酸序列可操作性连接的启动子的载体。

进一步优选的，野生型 Mmup 单体包括野生型 Mmup 单体的旁系同源物或同系物单体。

进一步优选的，所述包含野生型 Mmup 单体的构建体为野生型 Mmup 单体旁系同源物或同系物构建体或单体。

进一步优选的，所述包含野生型 Mmup 单体的孔蛋白为野生型 Mmup 单体的旁系同源物或同系物孔蛋白或单体。

优选的，所述的细菌为产粘液分枝杆菌。

本发明的第七方面，提供了一种产生 Mmup 孔蛋白的方法，所述的方法包括用包含本发明任一所述的载体转化本发明任一所述的细菌，诱导细菌表达 Mmup 孔蛋白。

本发明的第八方面，提供了一种 Mmup 单体变体的制备方法，所述的载体可引入适合的宿主细胞中，通过将所述编码 Mmup 单体变体的核苷酸序列插入载体中，将所述载体引入相容的细菌宿主细胞并且在允许所述核苷酸表达的条件下培养所述宿主细胞来产生本发明所述的 Mmup 单体变体。

本发明的第九方面，提供了一种包含本发明所述核苷酸序列或所述载体的细胞。

优选的，所述的细胞可以为大肠杆菌等等。更优选的，所述的细胞为 dam⁺型菌株（例如 DH5 α 菌株）。

本发明的第十方面，提供了一种表征目标多核苷酸的方法，包括：

（a）将目标多核苷酸与本发明任一所述包含至少一个 Mmup 单体变体的孔蛋白接触，使得所述目标多核苷酸序列穿过包含至少一个 Mmup 单体变体的孔蛋白；和

（b）获取目标多核苷酸穿过包含至少一个 Mmup 单体变体的孔蛋白时核苷酸与包含至少一个 Mmup 单体变体的孔蛋白相互作用的一个或多个特征，以表征所述目标多核苷酸。

优选的，重复步骤（a）、（b）一次或多次。

优选的，所述的步骤（a）中目标多核苷酸可以与衍生自多核苷酸处理酶结合，从而控制转位速度。进一步优选的，所述的多核苷酸处理酶是能够与多核苷酸相互作用并修饰其至少一种性质的多肽。其中，所述的多核苷酸处理酶可以具有酶活性也可以不具有酶活性，只要该酶结合多核苷酸且控制其在孔中的转位速度即可。其中，所述的核酸可以与一个或多个多核苷酸处理酶。

优选的，所述的多核苷酸处理酶为溶核酶。进一步优选的，所述的多核苷酸处理酶包括但不限于核酸结合蛋白、解旋酶、聚合酶、核酸外切酶、端粒酶、反转录酶、转位酶或拓扑异构酶。

在本发明的一个具体实施方式中，所述的多核苷酸处理酶为促旋酶。

优选的，所述步骤（a）中还包括目标多核苷酸与核酸结合蛋白、解旋酶、核酸外切酶、端粒酶、拓扑异构酶、反转录酶、转位酶和/或聚合酶中的一种或两种以上的组合接触的步骤，使得所述目标多核苷酸序列穿过孔蛋白的转位速度小于核酸结合蛋白、解旋酶、核酸外切酶、端粒酶、拓扑异构酶、反转录酶、转位酶和/或聚合酶不存在时的转位速度。

进一步优选的，所述的核酸结合蛋白包括但不限于修饰或者野生的真核单链结合蛋白、细菌单链结合蛋白、古生单链结合蛋白、病毒单链结合蛋白或双链结合蛋白中的一种或两种以上的组合。所述的核酸结合蛋白包括但不限于来自 *Escherichia coli* 的 SSB_{Eco}、来自 *Bartonella henselae* 的 SSB_{Bhe}、来自 *Coxiella burnetii* 的 SSB_{Cbu}、来自 *Thermatoga maritima* 的 SSB_{Tma}、来自 *Helicobacter pylori* 的 SSB_{Hpy}、来自 *Deinococcus radiodurans* 的 SSB_{Dra}、来自 *Thermus aquaticus* 的 SSB_{Taq}、来自 *Mycobacterium smegmatis* 的 SSB_{Msm}、来自 *Sulfolobus solfataricus* 的 SSB_{Sso}、来自 *Sulfolobus solfataricus* 的 SSB_{Sso7D}、来自 *Homo sapiens* 的 SSB_{MHsmt}、来自 *Mycobacterium leprae* 的 SSB_{Mle}、来自 *Bacteriophage T4* 的 gp32_{T4}、来自 *Bacteriophage RB69* 的 gp32_{RB69} 或来自 *Bacteriophage T7* 的 gp2.5_{T7}。

进一步优选的，所述的解旋酶可以为任一 Hel308 家族解旋酶及修饰的 Hel308 家族解旋酶、RecD 解旋酶及其变体、TrwC 解旋酶及其变体、Dda 解旋酶及其变体、TraI_{Eco} 及

其变体、XPD Mbu 及其变体。

进一步优选的，所述的聚合酶包括但不限于修饰或者野生的 DNA 聚合酶，包括但不限于 Phi29 DNA 聚合酶、Tts DNA 聚合酶、M2DNA 聚合酶、VENT DNA 聚合酶、T5DNA 聚合酶、PRD1DNA 聚合酶、Bst DNA 聚合酶或 REPLI-gscDNA 聚合酶。

进一步优选的，所述的核酸外切酶包括但不限于修饰或者野生的来自大肠杆菌的核酸外切酶 I、来自大肠杆菌的核酸外切酶 III、来自噬菌体 λ 核酸外切酶或者来自嗜热栖热菌的 RecJ。在本发明的一个具体实施方式中，所述步骤 (a) 中包括目标多核苷酸与解旋酶接触的步骤，所述的解旋酶为 EF8813，所述的解旋酶的氨基酸序列为 SEQ ID NO: 3 所示，所述解旋酶的核苷酸序列为 SEQ ID NO: 4 所示。优选的，所述的目标多核苷酸可以与一个或多个解旋酶接触。进一步优选的，所述的目标多核苷酸可以与 2-20 个解旋酶、甚至更多的解旋酶接触。其中，所述与目标多核苷酸结合的解旋酶可以相同也可以不同。且与目标多核苷酸结合的多个解旋酶彼此共价连接。

优选的，所述的一个或多个特征选自目标多核苷酸的来源、长度、大小、分子量、同一性、序列、二级结构、浓度或目标多核苷酸是否被修饰。

在本发明的一个具体实施方式中，所述的特征为序列。

优选的，所述步骤 (b) 中所述一个或多个特征通过电测量和/或光学测量进行。

进一步优选的，通过电测量和/或光测量产生电信号和/或光信号，而每种核苷酸对应一种信号水平，继而将电信号和/或光信号转化为核苷酸的序列特征。

本发明所述的电测量选自电流测量、阻抗测量、场效应晶体管 (FET) 测量、隧道测量或风洞测量。

本发明所述的电信号选自电流、电压、隧穿、电阻、电位、电导率或横向电测量的测量值。

在本发明的一个具体实施方式中，所述的电信号为穿过所述孔的电流。即所述电流以核苷酸特异性方式通过所述孔，如果检测到与核苷酸相关特征性电流流经所述孔，则存在所述的核苷酸。反之，则不存在。然而，对于相似核苷酸或者修饰的核苷酸之间的区分，

则根据电流的幅度确定。

优选的，采用本发明所述的孔蛋白进行多核苷酸的表征过程中产生的电导高于野生型 Mmup 单体形成的孔。

优选的，所述的方法还包括横跨目标多核苷酸接触的孔蛋白施加电势差的步骤。其中，所述电势差足以将目标多核苷酸从孔蛋白的通道中转位。

优选的，所述的目标多核苷酸可以是天然的 DNA、RNA 或者经过修饰的 DNA 或 RNA。

本发明所述的目标多核苷酸为含有一个或多个核苷酸的大分子。

本发明所述的目标多核苷酸可以是天然存在的或人工合成的。优选的，所述的目标多核苷酸中的一个或多个核苷酸可以是经过修饰的，例如甲基化、氧化、损伤、脱碱基的、蛋白标记、带有标签或多核苷酸序列中间连接一段间隔物。优选的，所述人工合成的核酸选自肽核酸（PNA）、甘油核酸（GNA）、苏糖核酸（TNA）、锁定核酸（LNA）、或其他具有核苷侧链的合成聚合物。

优选的，所述的孔蛋白允许水合离子在施加的电势的驱动下从膜的一侧流向膜的另一层。其中，所述的膜可形成离子、核苷酸和核酸流的屏障。进一步优选的，所述的膜为双层膜，进一步优选为脂质双层膜。所述的脂质双层膜包括但不限于磷脂、糖脂、胆固醇、分枝菌酸中的一种或两种以上的混合物。

优选的，所述的孔蛋白通道位于第一导电性液体介质与第二导电性液体介质之间，其中，至少一种导电性液体介质包含目标多核苷酸，并且第一导电性液体介质与第二导电性液体介质可以相同，也可以不同，只要可以达到分析目标多核苷酸一个或多个特征的目的即可。

优选的，所述的目标多核苷酸为单链、双链或至少一部分是双链的。

在本发明所述的一个具体实施方式中，所述的目标多核苷酸为至少一部分是双链的。其中所述的双链部分构成 Y 衍体结构，所述的 Y 衍体结构包含优先螺入所述孔蛋白的前导序列，所述的前导序列 3' 端连接硫醇、生物素或胆固醇，用来与脂质双层膜的一层膜结合，以为目标多核苷酸指向正确的方向并具有拉动的作用。

在本发明的一个具体实施方式中，所述的前导序列 3' 端连接胆固醇，用来与脂质双层膜的一层膜结合。

调节表征目标多核苷酸过程中的电压、盐浓度、缓冲液、添加剂或温度可以控制本发明所述的孔蛋白在表征目标多核苷酸中对不同核苷酸的区分程度。其中，所述的添加剂选自 DTT、脲或甜菜碱。

优选的，所述的电压范围为-250mV 至+250mV。进一步优选的，所述的电压选自 -250mV、-210mV、-180mV、-140mV、-110mV、-90mV、-70mV、-40mV、0mV、+40mV、+70mV、+90mV、+110mV、+140mV、+180mV、+210mV、+250mV。

在本发明的一个具体实施方式中，所述的电压为-180mV 至+180mV。

在本发明的一个具体实施方式中，所述的方法包括：将所述的孔蛋白插入膜中，然后将目标多核苷酸与所述的孔蛋白、核酸结合蛋白、聚合酶、核酸外切酶、端粒酶、拓扑异构酶、反转录酶、转位酶和/或解旋酶接触，向横跨目标多核苷酸接触的孔蛋白施加电势差，使得所述目标多核苷酸序列穿过孔蛋白；和

获取目标多核苷酸穿过孔蛋白时核苷酸与孔蛋白相互作用的电流特征，以鉴别多核苷酸为是否存在、为何种核苷酸或者是否经过修饰。

优选的，所述的孔蛋白插入膜中的方法可以为任何本领域已知的可以达到表征多核苷酸目的的方法。进一步优选的，所述孔蛋白可以以纯化形式悬浮于含有脂双层的溶液中，使得其扩散到所述脂双层并且通过结合到所述脂双层并组装为有功能的状态而插入到所述脂双层中。

本发明的第十一方面，提供了一种本发明任一所述包含至少一个 **Mmup** 单体变体的孔蛋白、本发明任一所述的 **Mmup** 单体变体、本发明任一所述构建体、本发明所述的核苷酸序列、本发明任一所述的载体或者本发明任一所述的突变细菌在表征目标多核苷酸中的应用。

本发明的第十二方面，提供了一种表征目标多核苷酸的试剂盒，所述的试剂盒包括本发明任一所述包含至少一个 **Mmup** 单体变体的孔蛋白、本发明任一所述的 **Mmup** 单体变体、

本发明任一所述构建体、本发明所述的核苷酸序列、本发明任一所述的载体或者本发明任一所述的突变细菌。

优选的，所述的 **Mmup** 单体变体、所述的构建体、所述的核苷酸序列、所述的载体、所述的细胞或所述的孔蛋白均可以为多个。

优选的，所述的试剂盒还包括一个或多个核酸结合蛋白、解旋酶、核酸外切酶、端粒酶、拓扑异构酶、反转录酶、转位酶和/或聚合酶中的一种或两种以上的组合。

优选的，所述的试剂盒还包括脂质双层的芯片，所述的孔蛋白横跨脂质双层。

优选的，所述的试剂盒包含一个或多个脂质双层，每个脂质双层包含一个或多个所述的孔蛋白。

优选的，所述的试剂盒还包括实施表征目标多核苷酸的试剂或装置。进一步优选的，所述的试剂包括缓冲剂、PCR 扩增所需的工具。

本发明的第十三方面，提供了一种表征目标多核苷酸的装置，所述的装置包括本发明任一所述包含至少一个 **Mmup** 单体变体的孔蛋白、本发明任一所述的 **Mmup** 单体变体、本发明任一所述构建体、本发明所述的核苷酸序列、本发明任一所述的载体或者本发明任一所述的突变细菌。

优选的，所述的装置还包括一个或多个核酸结合蛋白、解旋酶、核酸外切酶、端粒酶、拓扑异构酶、反转录酶、转位酶和/或聚合酶中的一种或两种以上的组合。

优选的，所述的装置还包括支撑所述孔蛋白并可传输孔蛋白与多核苷酸相互作用的信号的传感器，至少一个用于存储目标多核苷酸的存储器，和实施表征过程所需的溶液。

优选的，所述的装置还包括膜片钳放大器和/或数据获取装置。

本发明的第十四方面，提供了一种表征目标多核苷酸的传感器，所述的传感器包括本发明任一所述包含至少一个 **Mmup** 单体变体的孔蛋白、本发明任一所述的 **Mmup** 单体变体、本发明任一所述构建体、本发明所述的核苷酸序列、本发明任一所述的载体或者本发明任一所述的突变细菌。

本发明所述的“非天然氨基酸”为非在蛋白质中天然发现的包含氨基和羧基的化合物。

优选的，所述的非天然氨基酸为本领域已知的任何非天然氨基酸。进一步优选的，所述的非天然氨基酸包括但不限于 N-乙基天冬氨酰、羟基赖氨酸、3-羟基脯氨酸、2-氨基丁酸、 β -丙氨酸、 β -氨基丙氨酸、2-氨基己二酸、3-氨基己二酸、4-氨基丁酸、6-氨基己酸、2-氨基庚酸、别-异亮氨酸、异锁链赖氨酸、4-羟基脯氨酸、别-羟基赖氨酸、2-氨基异丁酸、N-甲基甘氨酸、N-甲基异亮氨酸、3-氨基异丁酸、6-N-甲基赖氨酸、2,4-二氨基丁酸、N-甲基缬氨酸、鸟氨酸、正亮氨酸、正缬氨酸、锁链素、2,2'-二氨基庚二酸、2,3-二氨基丙氨酸、N-乙基甘氨酸或 2-氨基庚二酸等等。

本发明所述的“修饰的.....基酸”为侧链被化学修饰的氨基酸。例如：翻译后修饰的氨基酸，或者，侧链包含新型官能团（如硫氢基、氨基或羧基），或者，侧链包含产生信号的部分（如荧光基团或放射性标记）。

本发明所述的“核苷酸”包括但不限于：腺苷单磷酸(AMP)、鸟苷单磷酸(GMP)、胸苷单磷酸(TMP)、尿苷单磷酸(UMP)、胞嘧啶核苷单磷酸(CMP)、环状腺苷单磷酸(cAMP)、环状鸟苷单磷酸(cGMP) 脱氧腺苷单磷酸(dAMP)、脱氧鸟苷单磷酸(dGMP)、脱氧胸苷单磷酸(dTMP)、脱氧尿苷单磷酸(dUMP) 和脱氧胞苷单磷酸(dCMP)。优选的，所述核苷酸选自 AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP 或 dCMP。

本发明所述的“和/或”包括择一列出的项目以及任何数量的项目组合。

本发明所述的“包括”是开放式的描述，含有所描述的指定成分或步骤，以及不会实质上影响的其他指定成分或步骤。

本发明所述的“约”用于表示该数值和用于测定该数值的装置或方法所允许的标准差。

本发明所述的“同源性”，是指在使用蛋白序列或核苷酸序列的方面，本领域技术人员可以根据实际工作需要调整对序列进行调整，使使用序列与现有技术获得的序列相比，具有（包括但不限于）1%，2%，3%，4%，5%，6%，7%，8%，9%，10%，11%，12%，13%，14%，15%，16%，17%，18%，19%，20%，21%，22%，23%，24%，25%，26%，27%，28%，29%，30%，31%，32%，33%，34%，35%，36%，37%，38%，39%，40%，41%，42%，43%，44%，45%，46%，47%，48%，49%，50%，51%，52%，53%，54%，55%，

56%, 57%, 58%, 59%, 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9%的序列一致性。

本发明所述的“Mmup 单体变体”是指与野生型 Mmup 单体具有至少或至多 70、75、80、85、90、91、92、93、94、95、96、97、98、99、99.5 或 99.9%或更多、或可来自其间的任何范围、但小于 100%的同一性并且当与一个或多个其他 Mmup 单体变体或野生型 Mmup 单体结合时保持形成通道的能力的 Mmup 单体变体。任选地，Mmup 单体变体被进一步确定为在促进完全形成的通道形成孔蛋白的收缩区和/或环形区的形成的序列部分中包含突变。Mmup 单体变体可以是例如重组蛋白。Mmup 单体变体可包含本文中描述的任何突变。

本发明所述的“Mmup 单体变体的旁系同源物或同系物孔蛋白”是指与野生型 Mmup 单体的旁系同源物或同系物孔蛋白具有至少或至多 70、75、80、85、90、91、92、93、94、95、96、97、98、99、99.5 或 99.9%或更多，或可来自其间的任何范围，但小于 100%的同一性并且保持通道形成能力的 Mmup 单体变体的旁系同源物或同系物孔蛋白。任选地，Mmup 单体变体的旁系同源物或同系物孔蛋白被进一步确定为在序列的该部分包含突变，所述部分促进完全形成的通道形成孔蛋白的收缩区和/或环形区的形成。Mmup 单体变体的旁系同源物或同系物孔蛋白可以例如是重组蛋白质。任何 Mmup 单体变体的旁系同源物或同系物孔蛋白可以任选地用于本文中的任何实施方案。

本发明所述的“Mmup 单体变体的旁系同源物或同系物构建体”是指与野生型 Mmup 单体的旁系同源物或同系物构建体具有至少或至多 70、75、80、85、90、91、92、93、94、95、96、97、98、99、99.5 或 99.9%或更多，或可来自其间的任何范围，但小于 100%的同一性并且保持通道形成能力的 Mmup 单体变体的旁系同源物或同系物构建体。任选地，Mmup 单体变体的旁系同源物或同系物构建体被进一步确定为在序列的该部分包含突变，所述部分促进完全形成的通道形成孔蛋白的收缩区和/或环形区的形成。Mmup 单体变体的旁系同源物或同系物构建体可以例如是重组蛋白质。任何 Mmup 单体变体的旁系同源物或

同系物构建体可以任选地用于本文中的任何实施方案。

附图说明

以下，结合附图来详细说明本发明的实施例，其中：

图 1: Mmup-(D91K/A99K)单体变体蛋白分子排阻层析的纯化结果，其中，1-6 泳道显示的是分子筛分离的不同组分的 SDS-PAGE 电泳检测结果。

图 2: 包含 Mmup 单体变体的纳米孔蛋白 (D91K/A99K) 的棍棒模型，图中主要显示了通道孔道收缩区和环形区的氨基酸分布特征，尤其是收缩区和环形区的关键氨基酸残基分布，指向孔道中心的氨基酸残基 91 位的赖氨酸和 92 位的天冬酰胺，孔道直径约为 13Å，99 位的赖氨酸参与通道复合物的正确组装，其中，同源建模通过 SWISS MODEL 完成，模板 pdb 位 1uun。

图 3: 包含野生型 Mmup 单体的纳米孔通道的棍棒模型，图中主要显示了孔道收缩区和环形区的氨基酸分布特征，尤其是收缩区和环形区的关键氨基酸残基，主要有 91 位的天冬氨酸，92 位的天冬酰胺，93 位的缬氨酸和 94 位的丝氨酸，由 D91 和 N92 形成的收缩区直径分别为 15.5Å 和 13.5Å，其中，同源建模通过 SWISS MODEL 完成，模板 pdb 位 1uun。

图 4: 包含野生型 Mmup 单体的纳米孔基于同源建模的卡通示意图。其中，区域 1 对应于帽状体形成区，区域 2 对应于桶状体形成区域，区域 3 对应于收缩区和环形区。

图 5: 待测 DNA 构建体 X2&cX2-80-15 的结构图，其中区段 a 对应于 SEQ ID NO: 7，b 对应于解旋酶 EF8813-1(含有 N 端组氨酸标签及其融合有 TOPV-HI 结构域的变体蛋白，SEQ ID NO: 3-4)，所述解旋酶可以结合到标记为 a 的区段，区段 c 对应于 SEQ ID NO: 6，区段 d 对应于 SEQ ID NO: 5，区段 e 对应于 SEQ ID NO: 8，其 5' 端 45 个碱基与测试链 c 段区域互补配对，3' 端含有 40 个胸腺嘧啶和对应于 g 的 3' 胆固醇 TEG 标记，区段 f 对应于 SEQ ID NO: 9。

图 6: 待测 DNA 构建体 S1T&S1MC 的结构图，其中区段 a 对应于 SEQ ID NO: 10，b 对应于解旋酶 EF8813-1 (含有 N 端组氨酸标签及其融合有 TOPV-HI 结构域的变体蛋白，SEQ ID NO: 3-4)，所述解旋酶可以结合到标记为 a 的区段，区段 h 指的是仅保留磷酸骨

架的 dspace, 标注为 x, 区段 c 对应于 SEQ ID NO: 12, 区段 d 对应于 SEQ ID NO: 13, 区段 e 对应于 SEQ ID NO: 14, 其 5'端 45 个碱基与测试链 c 段区域互补配对, 3'端含有 20 个胸腺嘧啶和对应于 g 的 3'胆固醇 TEG 标记, 区段 f 对应于 SEQ ID NO: 15。

图 7: 包含野生型 Mmup 单体的纳米孔通道在 $\pm 180\text{mV}$ 电压的单通道行为特征, 其中, y 轴坐标 = 电流(pA), x 轴坐标 = 时间(s)。

图 8: 包含 Mmup 单体变体 (D91K/A99K) 的纳米孔蛋白在 $+180\text{mV}$, 0mV 和 -180mV 电压下开孔电流及其门控特征, 其中, y 轴坐标 = 电流(pA), x 轴坐标 = 时间(s)。

图 9: 包含 Mmup 单体变体 (D91K/A99K) 的纳米孔蛋白在 $+180\text{mV}$ 和 0mV 电压下核酸通过纳米孔的信号特征, 其中, y 轴坐标 = 电流(pA), x 轴坐标 = 时间(s)。

图 10: 当解旋酶(EF8813-1)控制 DNA 构建体 X2 穿过包含 Mmup 单体变体(D91K/A99K)的纳米孔蛋白移位时的部分示例电流轨迹, 其中, 两条轨迹的 y 轴坐标 (左-20 至 180, 右 10 至 80) = 电流(pA), x 轴坐标 (左 15:04:25.5 至 15:04:37.5, 右 15:04:34.1 至 15:04:37.1) = 时间(s), 右图为左图虚线部分显示了电流轨迹的放大结果。

图 11: 当解旋酶(EF8813-1)控制 DNA 构建体 X2 穿过包含 Mmup 单体变体(D91K/A99K)的纳米孔蛋白移位时的电流轨迹的全过程电流变化结果, 其中, y 轴坐标 = 电流(pA), x 轴坐标 = 时间(s)。

图 12: 当解旋酶(EF8813-1)控制 DNA 构建体 S1T 穿过包含 Mmup 单体变体 (D91K/A99K)的纳米孔蛋白移位时的部分示例电流轨迹, 其中, 两条轨迹的 y 轴坐标 = 电流(pA), x 轴坐标 = 时间(s), 右图为左图虚线部分显示的电流轨迹的放大结果, 箭头指示的电流最大值显示的是 dspace 的特征峰。

图 13: 当解旋酶(EF8813-1)控制 DNA 构建体 S1T 穿过包含 Mmup 单体变体 (D91K/A99K)的纳米孔蛋白移位时的电流轨迹的全过程电流变化结果, 其中, y 轴坐标 = 电流(pA), x 轴坐标 = 时间(s)。

具体实施方式

下面将结合本发明实施例中的附图, 对本发明实施例中的技术方案进行清楚、完整地

描述，显然，所描述的实施例仅是本发明的部分实施例，而不是全部。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

实施例 1 Mmup 单体变体的制备

一、质粒的构建

Mmup 单体变体蛋白质序列通过对应氨基酸的密码子优化，并在基因两端添加合适的限制性内切酶酶切位点，具体的 5' 端添加 NcoI 位点 ccatgg，3' 端添加 xhoI 位点 ctcgag，之后进行基因合成，合成后的基因克隆到表达载体 pET24b 中。

二、目的基因的定点突变，制备 Mmup 单体变体核苷酸序列

诱导突变基因（PCR 反应）以待突变的质粒为模板，用设计的引物及 KOD plus 高保真酶进行 PCR 扩增反应，诱导目的基因突变。

具体步骤如下：

1、设计点突变引物，准备模板质粒 DNA（包含 SEQ ID NO: 2 的质粒 DNA），进行 50μL PCR 反应体系扩增。用 DH5α 菌株作为宿主菌。在 end+ 型菌株中常有克隆数低的现象，但是对突变效率没有影响。其中，提取模板质粒 DNA 使用 QIGEN 质粒提纯试剂盒。

点突变引物：

SEQ ID NO: 16

GACCCCGAACATTTTACTGAAAAATGTGAGCCCGCGACCCTGAAAGGTGTTGGTT
GGGGTGGC

SEQ ID NO: 17

GCCACCCCAACCAACACCTTTCAGGGTCGCCGGGCTCACATTTTTCAGTAAAATGTT
CGGGGTC

50μL PCR 反应体系：

10×KOD plus Buffer	5μL
模板质粒 DNA 60ng	2μL

正向引物 (20pmol/μL)	2μL
反向引物 (20pmol/μL)	2μL
dNTP mixture (each 2.5mM)	2μL
MgSO ₄	2.5μL
KOD plus 酶	1μL
ddH ₂ O	33.5μL

PCR 扩增反应

循环温度反应时间:

1 循环	95°C	5min
18 循环	95°C	30s
	72°C	30s 6min
	55°C	30min
1 循环	72°C	10min

PCR 扩增反应完成, 获得 Mmup 单体变体核苷酸序列 (SEQ ID NO: 11), 冰浴 5min, 然后置于室温 (避免反复冻融)。

2、模板消化, 提取 Mmup 单体变体基因

PCR 反应结束后使用 DpnI 酶消化甲基化质粒从而选择突变质粒 DNA (含有 SEQ ID NO: 11 的质粒)。准备 PCR 反应产物。具体步骤为: 加入 1μL (10U/μL) DpnI 酶 37°C 温育 2 小时。(当质粒 DNA 用量过多时 DpnI 酶可能发生与样品反应不完全的现象。如果突变率低, 可以适当延长反应时间或增加 DpnI 酶用量)

3、转化, 获得含有 Mmup 单体变体基因的菌株

反应完毕后在质粒 DNA 上会产生缺口, 因此当把这个质粒 DNA 转入 E.coli 中时选择 DH5α。具体步骤为: 将 4μL 突变质粒 DNA 样品加到 50μL DH5α 感受态细胞里, 然后放置在冰上 30min, 42°C 热击 90s, 之后立即冰浴 2min, 加入 500μL SOC 培养基 37°C 培养 1 小时, 最后取 100μL 菌液涂布抗性筛选平板。

4、测序验证

挑取 4 个转化子培养测序，选取突变正确的阳性转化子提取质粒保存备用。

三、制备 Mmup 单体变体

将测序验证正确的 Mmup 单体变体质粒转入 BL21(DE3)中培养。然后蛋白纯化，其中，蛋白纯化用试剂配方见表 1。

吸取 20 μ L 包含 Mmup 单体变体质粒的 BL21(DE3)甘油菌接于 20mL (1: 1000) 含有终浓度 50mg/mL 卡那霉素的新鲜 LB 培养基，37 $^{\circ}$ C，200rpm 摇菌过夜活化；次日按 1%接种量扩大培养至含有终浓度 50mg/mL 卡那霉素的 2L 的 LB 培养基中。37 $^{\circ}$ C，220rpm 培养至 OD₆₀₀=0.6-0.8 后冰浴迅速降温，之后向培养体系中加入 IPTG 至终浓度 1mM，15 $^{\circ}$ C，220rpm 诱导表达过夜。次日 6000rpm，4 $^{\circ}$ C 离心 15min 收集菌体，按照菌体：裂解缓冲液=1: 10 (m/v) 比例重悬菌体，之后加入混合蛋白酶抑制剂和终浓度 2%的 Triton X-100，高压破碎至菌液变清。

室温，搅拌增溶 1~2h，14000rpm，4 $^{\circ}$ C 离心 30min，收集上清液。上清用 0.45 μ m 滤膜过滤后用阴离子交换柱纯化，离子柱预先用 Buffer B 平衡，上清使用 5mL/min 流速过柱，收集穿透液。之后用 Buffer B 洗脱杂蛋白，最后用 Buffer C: 0-1 M 盐浓度线性梯度洗脱，收集洗脱组分。收集的穿透液样品加入终浓度 40%的硫酸铵，冰浴沉淀 2h，之后 14000rpm，4 $^{\circ}$ C 离心 30min，收集漂浮的沉淀物。漂浮的沉淀物加入一定体积含有 0.5% C8E4（四甘醇单辛基醚）去污剂的分子筛缓冲液重新溶解，4 $^{\circ}$ C 孵育过夜。

次日将过夜孵育的样品 47000g，4 $^{\circ}$ C 离心 30min，收集上清液。上清液进行最后一步的分子排阻层析纯化，收集目的组分即为 Mmup 单体变体。分子排阻层析结果见图 1。

表 1 蛋白纯化用试剂的配方

试剂名称	组成成分
裂解缓冲液	50mM Tris-HCl 8.0
	150mM NaCl

	10% 甘油
Buffer B	25mM HEPES-Na 7.5
	0.5% OPOE
Buffer C	25mM HEPES-Na 7.5
	1M NaCl
	0.5% OPOE

实施例 2 孔蛋白的制备

吸取 20 μ L 包含 Mmup 单体变体质粒的 BL21(DE3)甘油菌接于 20mL (1: 1000) 含有终浓度 50mg/mL 卡那霉素的新鲜 LB 培养基, 37 $^{\circ}$ C, 200rpm 摇菌过夜活化; 次日按 1% 接种量扩大培养至含有终浓度 50mg/mL 卡那霉素的 2L 的 LB 培养基中。37 $^{\circ}$ C, 220rpm 培养至 OD₆₀₀=0.6-0.8 后冰浴迅速降温, 之后向培养体系中加入 IPTG 至终浓度 1mM, 15 $^{\circ}$ C, 220rpm 诱导表达过夜。次日 6000rpm, 4 $^{\circ}$ C 离心 15min 收集菌体, 按照菌体: 裂解缓冲液 =1: 10 (m/v) 比例重悬菌体, 之后加入混合蛋白酶抑制剂和终浓度 2% 的 Triton X-100, 高压破碎至菌液变清。

室温, 搅拌增溶 1~2h, 14000rpm, 4 $^{\circ}$ C 离心 30min, 收集上清液。上清用 0.45 μ m 滤膜过滤后用阴离子交换柱纯化, 离子柱预先用 Buffer B 平衡, 上清使用 5mL/min 流速过柱, 收集穿透液。之后用 Buffer B 洗脱杂蛋白, 最后用 Buffer C: 0-1 M 盐浓度线性梯度洗脱, 收集洗脱组分。收集的穿透液样品加入终浓度 40% 的硫酸铵, 冰浴沉淀 2h, 之后 14000rpm, 4 $^{\circ}$ C 离心 30min, 收集漂浮的沉淀物。漂浮的沉淀物加入一定体积含有 0.5% C8E4 去污剂的分子筛缓冲液重新溶解, 4 $^{\circ}$ C 孵育过夜。

次日将过夜孵育的样品 47000g, 4 $^{\circ}$ C 离心 30min, 收集上清液。上清液进行最后一步的分子排阻层析纯化, 收集目的组分即为 Mmup 单体变体。

实施例 3 孔蛋白的测序应用

在缓冲液 (400mM KCl, 10mM HEPES pH 8.0, 50mM MgCl₂) 中, 将单个纳米孔蛋白插入磷脂双分子层中, 并从单个纳米孔蛋白获得电测量值。

具体步骤如下:

在将氨基酸序列为 D91K/A99K 突变的 SEQ ID NO: 1 的单个孔蛋白 (Mmup 单体变体孔蛋白, 棍棒模型如图 2 所示) 插入所述磷脂双分子层之后, 使缓冲液 (400mM KCl, 10mM HEPES pH 8.0, 50mM MgCl₂) 流经该系统, 以除去任何过量的 Mmup 单体变体孔蛋白。将 DNA 构建体 X2&cX2-80-15 或 S1T&S1MC (1~2nM 终浓度) 加入所述 Mmup 单体变体孔蛋白实验系统中, 混匀后, 使缓冲液 (400mM KCl, 10mM HEPES pH 8.0, 50 mM MgCl₂) 流经该系统, 以除去任何过量的 DNA 构建体 X2&cX2-80-15 或 S1T&S1MC。然后将解旋酶 (EF8813-1, 15nM 终浓度)、燃料 (ATP 3mM 终浓度) 预混物加入所述单个 Mmup 单体变体的孔蛋白实验系统中, 并在 +180 mV 电压下监测 Mmup 单体变体的孔蛋白的测序情况。

对照组与上述步骤相同, 仅将 Mmup 单体变体的孔蛋白替换为野生型 Mmup 单体的纳米孔 (棍棒模型及立体结构如图 3、4 所示)。其中, 野生型 Mmup 单体的纳米孔的棍棒模型显示了孔道收缩区和环形区的氨基酸分布特征, 尤其是收缩区和环形区的关键氨基酸残基, 主要有 91 位的天冬氨酸, 92 位的天冬酰胺, 93 位的缬氨酸和 94 位的丝氨酸, 由 D91 和 N92 形成的收缩区直径分别为 15.5Å 和 13.5Å。与野生型 Mmup 单体的纳米孔对比, 包含 Mmup 单体变体的纳米孔蛋白 (棍棒模型如图 2 所示) 棍棒模型显示了突变后通道孔道收缩区和环形区的氨基酸分布特征, 尤其是收缩区和环形区的关键氨基酸残基分布, 指向孔道中心的氨基酸残基 91 位的赖氨酸和 92 位的天冬酰胺, 孔道直径约为 13Å, 99 位的赖氨酸参与通道复合物的正确组装。

其中, X2&cX2-80-15 (具体结构如图 5) 具体序列如下:

X2:

TGGTTTTTGTGTTGTTTTAGAAATTTTTTTTACACTACCACTGCTAGCATTTTTTCA

(SEQ ID NO: 5)

TTTCTCACTATCCCGTTCTCATTTGGTGCACCATCTTTTTTTGGTT (SEQ ID NO: 6)

TTTTTGCAGCAGCAT (SEQ ID NO: 7)

cX2-80-15:

AACCAAAAAAAGATGGTGCACCAATGAGAACGGGATAGTGAGAAA (SEQ ID NO: 8)

TT (SEQ ID NO: 9)

其中, S1T&S1MC (具体结构如图 6) 具体序列如下:

S1T:

TTTTTTTTTTTTTTTCCTTCC (SEQ ID NO: 10)

X (区段 h)

TTCTTTTCCCGTCCGCTCGT (SEQ ID NO: 12)

TCGCGCCTGTCTGCTTGTTTTTTTTTTTTCTTTTTTTTTTTCTCACTATCGCATTCTC
ATGCAGGTCGGTGGTCGCAGTA (SEQ ID NO: 13)

S1MC:

ACGAGCGGACGGGAAAAGAA (SEQ ID NO: 14)

TT (SEQ ID NO: 15)

测试结果见图 7-13, 图 7 示出了包含野生型 Mmup 单体的纳米孔通道在 $\pm 180\text{mV}$ 电压的单通道行为特征。测试体系中包含野生型 Mmup 单体的纳米孔通道在 $+180\text{mV}$ 条件下全开放电流约为 380pA , 门控明显, 有强烈的残留核酸过孔信号; -180mV 条件下全开放电流接近 -350pA , 门控更强, 对膜的扰动很剧烈。显然, 包含野生型 Mmup 单体的纳米孔通道无法满足测序纳米孔蛋白的要求, 无法完成测序目的。

图 8 示出了包含 Mmup 单体变体 (D91K/A99K) 的纳米孔蛋白在 $+180\text{mV}$, 0mV 和 -180mV 电压下开孔电流及其门控特征, 其正向门控消失, 孔蛋白可以在外加电压下保持稳定的开放状态。在 180mV , 400mM KCl 盐浓度下可以产生 160pA 左右的开孔电流。图 9 示出了包含 Mmup 单体变体 (D91K/A99K) 的纳米孔蛋白在 $+180\text{mV}$ 和 0mV 电压下核酸通过纳米孔的信号。

当解旋酶(EF8813-1)控制 DNA 构建体 X2 穿过包含 Mmup 单体变体(D91K/A99K)的纳

米孔蛋白移位时的部分示例电流轨迹，右图为左图（15:04:25.5 至 15:04:37.5）中虚线部分显示的时间为 15:04:34.1 至 15:04:37.1 的电流轨迹的放大结果（参见图 10）。当解旋酶(EF8813-1)控制 DNA 构建体 X2 穿过包含 Mmup 单体变体(D91K/A99K)的纳米孔蛋白移位时的电流轨迹的全过程电流变化结果参见图 11。

当解旋酶(EF8813-1)控制 DNA 构建体 S1T 穿过包含 Mmup 单体变体(D91K/A99K)的纳米孔蛋白移位时的部分示例电流轨迹，右图为左图虚线部分显示的电流轨迹的放大结果，其中箭头指示的电流最大值显示的是 dspacer 的特征峰(参见图 12)。当解旋酶(EF8813-1)控制 DNA 构建体 S1T 穿过包含 Mmup 单体变体(D91K/A99K)的纳米孔蛋白移位时的电流轨迹的全过程电流变化结果参见图 13。

以上详细描述了本发明的优选实施方式，但是，本发明并不限于上述实施方式中的具体细节，在本发明的技术构思范围内，可以对本发明的技术方案进行多种简单变型，这些简单变型均属于本发明的保护范围。

另外需要说明的是，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合，为了避免不必要的重复，本发明对各种可能的组合方式不再另行说明。