

一种基于铜离子修饰 MspA 纳米孔的多肽成分分析方法

[1].本申请要求 2021 年 08 月 30 日提交的中国发明专利申请【CN2021110067648】、名称为“基于纳米孔检测多种氨基酸的方法及多肽组成成分分析方法”的优先权，该优先权发明专利申请以引用方式全文并入。

技术领域

[2].本发明属于纳米孔检测领域，具体涉及一种基于铜离子修饰 MspA 纳米孔的多肽成分分析方法。

背景技术

[3].纳米孔传感是一种单分子传感技术，有着类似于库尔特计数器（Kurt counter）的检测原理。该技术具有在单分子水平上实时和直接监测的特点，且一般不需要对分析物进行标记或修饰。这些优点使纳米孔成为生物传感和生物检测的新兴技术。

[4].由于氨基酸通过纳米孔时，易位过程中信噪比低和构象变化难以预测，这对于纳米孔直接区分并检测氨基酸仍是一项挑战。现有技术无法直接对 20 种氨基酸进行检测并实现良好区分，往往需要对纳米孔进行化学修饰或对氨基酸进行衍生化处理等。

发明内容

[5].有鉴于此，本发明提供了一种基于 M2MspA-N91H 纳米孔的多肽成分分析方法，其特征在于，包括以下步骤：

S1 将外肽酶与待检测多肽混合均匀并反应，得到水解产物；

S2 将所述水解产物加入纳米孔系统中，所述纳米孔系统包括：M2MspA-N91H 纳米孔、绝缘膜、第一介质、第二介质，其中所述 M2MspA-N91H 纳米孔嵌入在所述绝缘膜中，所述 M2MspA-N91H 纳米孔提供连通所述第一介质和所述第二介质的通道，所述水解产物被加入至所述第一介质；

S3 向所述第一介质和所述第二介质之间施加驱动力，所述水解产物中游离氨

基酸穿过所述 M2MspA-N91H 纳米孔并产生电信号；

S4 对所述电信号进行数据分析，获得所述游离氨基酸的氨基酸种类信息。

[6].在一个实施方案中，所述游离氨基酸包括天然氨基酸和非天然氨基酸。

[7].在一个实施方案中，所述天然氨基酸包括甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、苯丙氨酸、色氨酸、酪氨酸、天冬氨酸、组氨酸、天冬酰胺、谷氨酸、赖氨酸、谷氨酰胺、甲硫氨酸、精氨酸、丝氨酸、苏氨酸、半胱氨酸、脯氨酸中的一种或多种。

[8].在一个实施方案中，所述非天然氨基酸包括瓜氨酸、L-正缬氨酸、鹅膏氨酸中的一种或多种。

[9].在一个实施方案中，所述外肽酶包括氨肽酶和/或羧肽酶。

[10].在一个实施方案中，在所述水解产物加入所述纳米孔系统之前，将所述 M2MspA-N91H 纳米孔与铜离子结合。

[11].在一个实施方案中，所述第一介质和所述第二介质的 pH 值包括 6.5-7.5。

[12].在一个实施方案中，所述待检测多肽溶液的终浓度大于 10 μM 。

[13].在一个实施方案中，S4 中所述数据分析步骤进一步包括：

- i) 建立标准信号数据库；
- ii) 设定信号过滤条件；
- iii) 排除部分电信号，获得保留的电信号；
- iv) 通过标准信号数据库中所述标准信号的阻塞率，判断所述保留的电信号对应的氨基酸种类；
- v) 获得所述游离氨基酸的氨基酸种类信息。

[14].在一个实施方案中，所述信号过滤条件为排除 SD 值大于 SD 阈值的电信号，其中所述电信号的 SD 值按照下列公式进行计算：

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

式中， x_i 为信号每个采集点的电流值， $\bar{x}$ 为信号电流点的平均值，n 为信号采集点的数量。

[15].在一个实施方案中，所述方法进一步包括 S5 将所述游离氨基酸的氨基酸种类信息与所述待检测多肽的预期序列信息进行对比，以判断所述待检测多肽的实际序列与所述待检测多肽的预期序列是否存在差异；或

S5 将所述游离氨基酸的氨基酸种类信息结合所述外肽酶的信息与数据库中的多肽序列信息进行对比，以预测所述待检测多肽的实际序列。

[16].有益效果

[17].现有技术的纳米孔往往只能检测特定官能团的氨基酸，而本发明实验表明，M2MspA-N91H 纳米孔（源自耻垢分枝杆菌（*Mycobacterium smegmatis*）的孔蛋白 A）在检测氨基酸方面具有通用性，即可以在单分子水平上稳健且灵敏地检测和区分 20 种常见氨基酸和非天然氨基酸。此外，无需向 M2MspA-N91H 纳米孔系统中额外引入酶和/或对氨基酸进行修饰（例如，添加发光标记），20 种常见氨基酸和非天然氨基酸都可以阻塞 M2MspA-N91H 纳米孔并引起相应的阻塞电流信号。通过分析不同电流信号可以实现对多种氨基酸的直接检测并区分。

[18].基于此，结合外肽酶（例如羧肽酶）的水解特性，本发明提供一种基于 M2MspA-N91H 纳米孔的多肽成分分析方法。可以根据建立的标准信号数据库中标准信号的阻塞率，判断检测到的游离氨基酸的信号对应的氨基酸种类，进而一定程度上实现对待检测多肽的成分分析。本发明提供的多肽成分分析方法适用于肿瘤新生抗原的表达和多肽药物的合成质量的判断，以及未知多肽序列的预测等多种应用场景。

附图说明

[19].为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍。显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其它的附图。

[20].图 1 为 M2MspA-N91H 纳米孔的结构与电生理特性图；

[21].图 2 为加入不同浓度的铜离子时，M2MspA-N91H 纳米孔单孔电流信号图（1 M KCl，pH 6）；

[22].图 3 为 M2MspA-N91H 纳米孔在 3 M KCl 浓度下进行氨基酸检测的电流信号图；

[23].图 4 为氨基酸体积 v.s.电流信号阻塞率的散点图（标注字母为氨基酸缩写）；

[24].图 5 为氨基酸电流信号阻塞率 v.s.阻塞时间图（标注字母为氨基酸缩写）；

[25].图 6 为 Clampfit 提取单氨基酸信号的实验图(A 为 Clampfit 单通道统计结果; B 为信号的 Amp S.D.值; C 为根据 SD 值区分散点图中的单台阶信号(从上到下分别为同一次实验中空白组、asp 实验组、asp+glu 实验组统计 2 分钟的所有信号); D 为-5 pA 背景信号来源; E 为-12 pA 背景信号来源);

[26].图 7 为羧肽酶水解产物加入纳米孔体系后测得的电流信号图(A 为羧肽酶在 0.5 M KCl、10 mM MOPS、pH 7.5 条件下, 10 U 羧肽酶水解 10 mM EF 多肽 20 分钟后加入纳米孔体系检测得到的电流信号图; B 为 A 涉及实验中统计得到的所有电流信号的阻塞率及过孔时间的散点图; C 为纳米孔分别检测三种多肽(P1, P2, P3)的水解产物);

[27].图 8 为 M2MspA-N91H 纳米孔检测非天然氨基酸的实验图(A 为三种非天然氨基酸统计结果; B 为同一次实验中空白组(Blank)和加入鹅膏氨酸后, 1 分钟内的电流阻塞信号散点图);

[28].图 9 为在 3 种 pH 值条件下, M2MspA-N91H 纳米孔加入铜离子后的电流轨迹图;

[29].图 10 为本发明的纳米孔检测多肽水解产物示意图;

[30].图 11 为本发明实施例中建立标准信号数据库的流程图;

[31].图 12 为本发明实施例中数据分析的流程图。

具体实施方式

[32].为使本发明实施例的目的、技术方案和优点更加清楚,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,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作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。

[33].需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置

中还存在另外的相同要素。

[34].如在本说明书中使用的，术语“大约”，典型地表示为所述值的 $\pm 5\%$ ，更典型的是所述值的 $\pm 4\%$ ，更典型的是所述值的 $\pm 3\%$ ，更典型的是所述值的 $\pm 2\%$ ，甚至更典型的是所述值的 $\pm 1\%$ ，甚至更典型的是所述值的 $\pm 0.5\%$ 。

[35].在本说明书中，某些实施方式可能以一种处于某个范围的格式公开。应该理解，这种“处于某个范围”的描述仅仅是为了方便和简洁，且不应该被解释为对所公开范围的僵化限制。因此，范围的描述应该被认为是已经具体地公开了所有可能的子范围以及在此范围内的独立数字值。例如，范围 1~6 的描述应该被看作已经具体地公开了子范围如从 1 到 3，从 1 到 4，从 1 到 5，从 2 到 4，从 2 到 6，从 3 到 6 等，以及此范围内的单独数字，例如 1，2，3，4，5 和 6。无论该范围的广度如何，均适用以上规则。

[36].纳米孔系统

[37].“纳米孔系统”包括具有纳米级尺寸的孔（简称为“纳米孔”）、绝缘膜、第一介质和第二介质。在本发明的一个实施例中，所述具有纳米级尺寸的孔为 M2MspA-N91H 纳米孔。所述具有纳米级尺寸的孔允许所述待测物从所述绝缘膜的一侧易位到另一侧。另外，本领域技术人员还可以根据实际情况，对上述 M2MspA-N91H 纳米孔进行修饰（例如，任何的突变、截短、融合、化学修饰等），来获得相应变体，且修饰手段都是本领域公知的。

[38].在本发明的一个实施例中，所述具有纳米级尺寸的孔被嵌入所述绝缘膜中，所述绝缘膜（也可以理解为，所述纳米级尺寸的孔和所述绝缘膜的复合体）将所述第一介质与所述第二介质分隔开，所述具有纳米级尺寸的孔的孔道提供连通所述第一介质与所述第二介质的通道；向所述第一介质和所述第二介质之间施加驱动力后，位于所述第一介质的待测物与所述 M2MspA-N91H 纳米孔相互作用以形成电流（即电信号）。在本发明中，“第一介质”是指所述待测物被加入所述纳米孔系统时位于的介质；“第二介质”则是指被所述绝缘膜分隔开的两部分介质中，“第一介质”的另一侧。在本发明中，驱动力是指通过电势、电渗流、浓度梯度等方式驱动待测物与所述纳米孔相互作用的力。

[39].所述第一介质和所述第二介质可以相同或不同，并且所述第一介质和所述第二介质可以包括电导液。所述电导液为碱金属卤化物水溶液，具体为氯化钠

(NaCl)、氯化锂(LiCl)、氯化铯(CsCl)、氯化钾(KCl)、溴化钠(NaBr)。在本发明一个实施例中,所述第一介质和所述第二介质含有的电导液的浓度是不同的,换句话说,所述第一介质和所述第二介质中电导液的浓度存在差值,进而使得所述绝缘膜两侧的渗透压存在差值。所述第一介质和/或所述第二介质还可以包括缓冲液,例如 HEPES、MOPS。所述第一介质和/或所述第二介质的浓度范围可以是 0.3 M-3 M。所述第一介质和/或所述第二介质的 pH 范围为 6-9 (优选为 6.5-7.5)。

[40].绝缘膜是指具有搭载纳米孔并阻塞非纳米孔通过的离子电流的能力的膜。所述绝缘膜可以包括磷脂膜和/或高分子膜。示例性的磷脂膜包括 DPHPC、DOPC、E.coli lipid, 示例性的高分子膜包括三嵌段共聚物高分子膜。

[41].在本发明的一个具体实施例中,所述纳米孔系统包括两个电解液室,其被绝缘膜分隔开而形成反式(-trans)隔室和顺式(-cis)隔室,所述 M2MspA-N91H 通道的孔嵌入绝缘膜中,绝缘膜上只有 M2MspA-N91H 通道来连通上述两个电解液室。当向上述两个电解液室施加电势时,电解液室中溶液中的电解质离子通过电泳移动并穿过所述 M2MspA-N91H 通道。

[42].待测物

[43].所述待测物是荷电物质。如果待测物带有净电荷则它是荷电的。所述待测物可以荷负电也可以荷正电。如果待测物带有净负电荷则它是荷负电的。如果待测物带有净正电荷则它是荷正电的。合适的待测物优选为氨基酸。

[44].在本发明的一个实施例中,所述氨基酸可以为天然氨基酸。天然氨基酸是天然存在的氨基酸及其翻译后修饰的变体,例如 20 种常见的氨基酸(甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、甲硫氨酸、脯氨酸、色氨酸、丝氨酸、酪氨酸、半胱氨酸、苯丙氨酸、天冬酰胺、谷氨酰胺、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸和组氨酸)等。

[45].在本发明的另一个实施例中,所述待测物可以为非天然氨基酸。“非天然氨基酸”是指不是蛋白氨基酸或其翻译后修饰的变体的氨基酸(即不是 20 种常见氨基酸或吡咯赖氨酸或硒代半胱氨酸或其翻译后修饰的变体的氨基酸)。在本发明的一个具体实施例中,所述非天然氨基酸为瓜氨酸、L-正缬氨酸、鹅膏氨酸。

[46].所述纳米孔和待测物之间的相互作用

[47].所述待测物可与所述纳米孔在所述绝缘膜两侧的任一侧接触。所述待测物可以与所述绝缘膜两侧中的任一侧相接触,使得所述待测物穿过所述纳米孔的通道以到达所述绝缘膜的另一侧。在这种情况下,所述待测物在其经由所述孔的通道穿过所述绝缘膜时,与所述纳米孔相互作用。或者,所述待测物可与所述绝缘膜的侧面接触,所述绝缘膜的侧面可使所述待测物与所述纳米孔相互作用,使其与所述纳米孔分离并停留在所述绝缘膜的同一侧。所述待测物可以以任何方式并在任何位点与所述纳米孔相互作用。所述待测物还可以撞击到所述纳米孔,与所述纳米孔相互作用,使其与所述纳米孔分离并停留在所述绝缘膜的同一侧。

[48].在所述待测物与所述纳米孔相互作用的过程中,所述待测物会以该待测物特异性的方式影响流过所述纳米孔的电流,即流经所述纳米孔的电流对特定待测物是特征性的。可进行对照实验以测定特定待测物对流过所述纳米孔的电流的效应,然后以鉴定样本中的特定待测物或测定样本中是否存在特定待测物。更具体地,可以根据通过检测待测物所获得的电流模式与在相同的条件下使用已知的待测物获得的已知的电流模式进行比较,以鉴定待测物的存在与否或浓度等。

[49].本发明的纳米孔系统还可以包括一个或多个测量流过所述纳米孔的电流的测量装置,例如膜片钳放大器或数据采集设备。

[50].一种基于 M2MspA-N91H 纳米孔的多肽成分分析方法

[51].本发明还提供了一种基于 M2MspA-N91H 纳米孔的多肽成分分析方法,其特征在于,包括以下步骤:

S1 将外肽酶与待检测多肽混合均匀并反应,得到水解产物;

S2 将所述水解产物加入纳米孔系统中,所述纳米孔系统包括: M2MspA-N91H 纳米孔、绝缘膜、第一介质、第二介质,其中所述 M2MspA-N91H 纳米孔嵌入在所述绝缘膜中,所述 M2MspA-N91H 纳米孔提供连通所述第一介质和所述第二介质的通道,所述水解产物被加入至所述第一介质;

S3 向所述第一介质和所述第二介质之间施加驱动力,所述水解产物中游离氨基酸穿过所述 M2MspA-N91H 纳米孔并产生电信号;

S4 对所述电信号进行数据分析,获得所述游离氨基酸的氨基酸种类信息。

[52].“多肽”是指含有通过肽键连接的两个或多个氨基酸的肽或蛋白。多肽可以含有天然的、修饰的或合成的氨基酸。多肽也可以被天然修饰(如通过翻译后加

工)或化学修饰(如酰胺化酰化、交联等)。在本发明中,所述多肽的序列信息可以是已知的或未知的。

[53].“外肽酶”是指通过切割肽键去除多肽的末端氨基酸的水解酶。末端氨基酸是在多肽的 N-末端或 C-末端的大约 10 个氨基酸内的氨基酸。因此,“外肽酶”分为从多肽的 N-末端逐个水解氨基酸的氨肽酶和从多肽的 C-末端逐个水解氨基酸的羧肽酶。示例性的氨肽酶包括半胱氨酰氨肽酶、脯氨酰氨肽酶、精氨酰氨肽酶。示例性的羧肽酶包括羧肽酶 A、B、C、Y。在本发明中,所述外肽酶可以为羧肽酶和/或氨肽酶。

[54].在适宜的条件下,所述外肽酶可以将待检测多肽完全水解或部分水解。例如,如图 10 所示,羧肽酶 A1 可以切割 C-末端的除了赖氨酸(K)、精氨酸(R)、脯氨酸(P)的氨基酸(即水解的停止点为 R、K、P)。换句话说,如果待检测多肽中不含 R、K、P 这三种氨基酸,则所述待检测多肽将被羧肽酶 A 完全水解,得到的水解产物中的游离氨基酸的氨基酸种类信息则为所述待检测多肽的氨基酸组成。如果待检测多肽中含 R、K、P 这三种氨基酸中的一种或多种,则所述待检测多肽将被羧肽酶 A 部分水解(在停止点停止水解),得到的水解产物中的游离氨基酸的氨基酸种类信息则为所述待检测多肽在停止点前的氨基酸组成。

[55].当所述待检测多肽的序列已知时,可以根据得到的水解产物中游离氨基酸的氨基酸种类信息与所述待检测多肽的预期序列信息进行对比,以判断所述待检测多肽的实际序列与所述待检测多肽的预期序列是否存在差异(例如,某个(些)应该出现在水解产物的氨基酸没有出现,某个(些)不应该出现在水解产物的氨基酸却出现了),适用于肿瘤新生抗原的表达和多肽药物的合成质量的判断等应用场景。

[56].当所述待检测多肽的序列未知时,可以根据得到的水解产物中游离氨基酸的氨基酸种类信息并结合所用外肽酶的信息(例如,羧肽酶或氨肽酶,是否具有“水解停止点”),通过机器学习等方式在数据库中进行比对,进而预测所述待检测多肽的序列或在一定程度上缩小所述待检测多肽的可能范围(例如,在 N 端或 C 端开始,氨基酸的组成(包括游离氨基酸的确定种类以及“水解停止点”对应的可能种类))。

[57].**实施例一**

[58].材料与方法

[59].氨基酸检测:

[60].使用经典的垂直脂质双层装置 (Warner Instruments) 进行单通道实验, 包括通过 150 μm 直径孔径连接的两个室 (顺式 *cis* 和反式 *trans*)。两个 Ag-AgCl 电极用于施加跨膜电压并测量跨膜离子电流。顺式室处于电压接地。氨基酸检测的缓冲液条件为:

- 1) 缓冲液: 0.3 M-3 M 的 KCl 和 NaCl 等盐溶液 (优选为 1 M 的 KCl 水溶液), 10 mM MOPS;
- 2) 铜离子浓度范围: 1 nM 到 200 μM ;
- 3) pH 值范围: 6-9 (优选为 6.5-7.5)。

[61].随着缓冲液盐浓度或 pH 进一步降低, 氨基酸检测效率和区分度可能会下降。所有实验温度可设置为 15°C-60°C (优选为 22.0°C $\pm$ 0.5°C)。通过在连接两个腔室的孔上涂上溶解在癸烷 (40 mg/ml) 中的 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) (Avanti Polar Lipids) 薄膜, 形成平面脂质双层膜。之后将蛋白溶液加入至顺式腔, 在蛋白嵌入磷脂膜形成单通道后, 将氯化铜加入到反式腔, 并将氨基酸加入到顺式腔, 轻轻混匀溶液后记录电流信号。

[62].数据处理:

[63].使用 Axopatch 200B 放大器 (Molecular Devices) 放大单通道电流记录, 并使用内置四孔低通贝塞尔滤波器以 2 kHz 进行过滤。数据由 Digidata 1550B 转换器 (Molecular Devices) 以 100 kHz 的采样率进行数字化。提取采样得到的电流信号文件, 获得氨基酸信号数据。可通过信号阻塞率、阻塞时间、SD 等数据对单氨基酸信号进行识别和定义。

[64].背景噪音的处理:

[65].针对单个氨基酸信号的电流值分布, 计算其 SD 值, 得到信号各个采样点的电流值与总体平均值的偏离程度, 从而将单氨基酸信号与背景信号, 多氨基酸的叠加信号区分。

[66].多肽组成成分分析方法:

[67].实验的缓冲液条件如上所述。实验温度可设置为 15°C-60°C (优选为 25°C)。

在纳米孔体系外，将外肽酶（例如羧肽酶）与多肽按照一定浓度比例进行混合，如 10 U 羧肽酶、0.5 M KCl、10 mM MOPS、pH 7.5、10 mM 多肽（图 7A），25°C 反应 20 分钟后，加入纳米孔样品槽进行检测。根据出现的氨基酸信号种类和数量判断多肽成分。

[68]. 实施例二：利用 M2MspA-N91H 纳米孔检测 20 种天然氨基酸

[69]. 使用铜离子修饰纳米孔的结构和特性：

[70]. 如图 1 所示，M2MspA-N91H 纳米孔为 8 聚体结构，孔道最窄处的 8 个氨基酸残基朝向 M2MspA-N91H 纳米孔的中心轴线。将 Cu^{2+} 添加到反式室中，观察电流轨迹， Cu^{2+} 可显著降低电流波动（图 1）。电流水平的降低与 Cu^{2+} 的添加浓度有关（图 2），铜离子浓度范围可以是 1 nM 至 200 μM 。

[71]. 20 种天然氨基酸的检测：

[72]. 本实施例利用 M2MspA-N91H 纳米孔检测了 20 种天然氨基酸。其中，MspA 单体（MspA 由八个相同的单体组装形成通道）的氨基酸序列为：
GLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAG
PGADEFEGTLELGYQIGFPWSLGVGINFSYTTPNILINHGNITAPPFGLNSVITP
NLFPGVVISARLGNGPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAGGVLL
RPFARLIASGDSVTTYGEPWNMN（SEQ ID NO:1）

[73]. M2MspA-N91H 的氨基酸序列为：
GLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAG
PGADEFEGTLELGYQIGFPWSLGVGINFSYTTPNILINHGNITAPPFGLNSVITP
NLFPGVVISARLGNGPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAGGVLL
RPFARLIASGDSVTTYGEPWNMNGLDNELSLVDGQDRTLTVQQWDTFLNGV
FPLDRNRLTREWFHSGRAKYIVAGPGADEFEGTLELGYQIGFPWSLGVGINFS
YTTPNILINHGNITAPPFGLNSVITPNLFPGVVISARLGNGPGIQEVATFSVRVSG
AKGGVAVSNAHGTVTGAAGGVLLRPFARLIASGDSVTTYGEPWNMNGLDN
ELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPGAD
EFEGTLELGYQIGFPWSLGVGINFSYTTPNILINHGNITAPPFGLNSVITPNLFP
GVVISARLGNGPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAGGVLLRPFAR
LIASGDSVTTYGEPWNMNGLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDR
NRLTREWFHSGRAKYIVAGPGADEFEGTLELGYQIGFPWSLGVGINFSYTTPNI

LINHG NITAPPFGLNSVITPNLFPGV SISARLGNGPGIQEVATFSVRVSGAKGGV
AVSNAHGT V TGAAGGVLLRPFARLIAS TGD SVTTYGEPWNMNGLDNELSLVD
GQDR TLT VQ QWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPGADEFEGTL
ELGYQIGFPWSLGVGINFSYTT PNILINHG NITAPPFGLNSVITPNLFPGV SISAR
LGNGPGIQEVATFSVRVSGAKGGVAVSNAHGT V TGAAGGVLLRPFARLIAS TGD
DSVT TYGEPWNMNGLDNELSLVDGQDR TLT VQ QWDTFLNGVFPLDRNRLTR
EWFHSGRAKYIVAGPGADEFEGTLELGYQIGFPWSLGVGINFSYTT PNILINHG
NITAPPFGLNSVITPNLFPGV SISARLGNGPGIQEVATFSVRVSGAKGGVAVSNA
HGT V TGAAGGVLLRPFARLIAS TGD SVTTYGEPWNMNGLDNELSLVDGQDR T
LT VQ QWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPGADEFEGTLELGYQ
IGFPWSLGVGINFSYTT PNILINHG NITAPPFGLNSVITPNLFPGV SISARLGNGP
GIQEVATFSVRVSGAKGGVAVSNAHGT V TGAAGGVLLRPFARLIAS TGD SVTT
YGEPWNMNGLDNELSLVDGQDR TLT VQ QWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPGADEFEGTLELGYQIGFPWSLGVGINFSYTT PNILINHG NITAPP
FGLNSVITPNLFPGV SISARLGNGPGIQEVATFSVRVSGAKGGVAVSNAHGT V T
GAAGGVLLRPFARLIAS TGD SVTTYGEPWNMN (SEQ ID NO:2)

[74].当目标待测物通过 Cu^{2+} 螯合的 M2MspA-N91H 时，产生特定的电流信号，如图 3 所示。在 3 种 pH 值条件下，加入铜离子后的电流轨迹图，如图 9 所示。结果表明，在 pH 7.9 时铜离子结合不稳定，而在 pH 7.5 和 6.5 条件下铜离子稳定结合。鉴于较高的溶液 pH 值有利于更多氨基酸带负电，从而提高检测效率。故本实施例选择在溶液中添加 MOPS 缓冲盐并将溶液 pH 值稳定至 7.5 左右，以期获得更优的检测效果。

[75].从氨基酸体积与电流信号阻塞率的散点图（图 4）可以看出，氨基酸引起的阻塞率与其疏水体积存在正相关的联系。

[76].而从氨基酸电流信号阻塞率与阻塞时间图（图 5）可以得到，由 20 种天然氨基酸引起的阻塞率分别为：甘氨酸（G）： 0.1207 ± 0.0006 ；丙氨酸（A）： 0.1471 ± 0.0025 ；缬氨酸（V）： 0.1904 ± 0.0012 ；亮氨酸（L）： 0.1995 ± 0.0003 ；异亮氨酸（I）： 0.2072 ± 0.0003 ；苯丙氨酸（F）： 0.2201 ± 0.0007 ；色氨酸（W）： 0.2274 ± 0.0002 ；酪氨酸（Y）： 0.2127 ± 0.0040 ；天冬氨酸（D）： 0.2140 ± 0.0028 ；

组氨酸 (H) : 0.2465 ± 0.0009 ; 天冬酰胺 (N) : 0.1651 ± 0.0002 ; 谷氨酸 (E) : 0.2452 ± 0.0016 ; 赖氨酸 (K) : 0.1687 ± 0.0031 ; 谷氨酰胺 (Q) : 0.1882 ± 0.0009 ; 甲硫氨酸 (M) : 0.1977 ± 0.0005 ; 精氨酸 (R) : 0.1714 ± 0.0022 ; 丝氨酸 (S) : 0.1313 ± 0.0002 ; 苏氨酸 (T) : 0.1610 ± 0.0004 ; 半胱氨酸 (C) : 0.1862 ± 0.0023 ; 脯氨酸 (P) : 0.2183 ± 0.0010 。根据氨基酸的阻塞率以及其标准差进行判断, 互不重叠的氨基酸有 G, S, A, T, N, R, V, Q, M, L, I, D, F, W, E, 共 15 种氨基酸的信号互不重叠, 可互相区分。

[77].背景噪音的处理:

[78].上述直接检测 20 种氨基酸的方法, 可能会产生背景噪音干扰后续的分析。基于此, 本发明实验人员针对单个氨基酸信号的电流值分布, 用下述的公式计算其 SD 值, 得到信号各个采集点的电流值与总体平均值的偏离程度。从而将单氨基酸信号与背景信号、多氨基酸的叠加信号区分。

$$[79].SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

[80].式中, x_i 为信号每个采集点的电流值, $\bar{x}$ 为信号电流点的平均值, n 为信号采集点的数量。

[81].如图 6C 所示, 上、中、下图分别为同一次实验中空白组 (Blank)、加入天冬氨酸 (Asp) 和加入天冬氨酸与谷氨酸混合物 (Asp+Glu) 后, 2 分钟内的电流阻塞信号散点图。可以设定 SD 阈值来过滤信号, 例如, 可以将 SD 阈值设定为 2, 其中, SD 值小于 2 的信号 (蓝色) 被认为是有效信号; SD 值大于 2 的信号, 被认为是背景信号或多氨基酸同时通过产生的叠加信号, 因此可将其过滤。过滤后, 只看蓝色的信号, 可以清楚地看到加入天冬氨酸与谷氨酸后, 新产生的信号峰。利用本背景噪音处理方法, 可更准确地对氨基酸信号进行识别和统计。

[82].为了更好地展示处理过程, 本发明实验人员使用了 Clampfit 软件来提取单氨基酸信号。首先, 利用拟合 SD 值, 将背景信号及双台阶信号过滤掉, 并保留单氨基酸信号。如图 6A 所示, 在 Clampfit 单通道统计结果中 (只设定 level 0 和 level 1), “Amp S.D.” 值表示信号拟合效果。具体信号对应的 SD 值如图 6B 所示; 在所示的 6 个信号中可以看出, 双台阶信号的 SD 值普遍大于 4。单台阶信号是指电信号的电流密度分布曲线呈现单峰分布, 且电流在 0 与峰值的 60% 之间没有分布。而双台阶信号是指电信号的电流密度分布曲线呈现双峰分布。

[83].SD 阈值设定在 0-2 范围内,可过滤大部分多台阶信号。然而,在-5 pA 以及 -12 pA 附近仍然存在背景信号干扰。-5 pA 附近的信号来源如图 6D 所示,因为其对应的 level 0 偏离基线过多,可在后期被过滤掉。-12 pA 附近的信号(图 6E)是正常的背景信号,阻塞时间短。背景信号只是会干扰到附近的 gly、ser、ala 的信号,而这三种氨基酸的信号是可以区分的。

[84].基于此,可以使用已知检测数据,建立标准信号数据库,具体流程如图 11 所示:首先,通过 M2MspA-N91H 纳米孔检测标准品游离氨基酸,获得电信号数据。第二,设置过滤条件(例如将电信号的 SD 阈值设定为 2)。第三,根据过滤条件,将部分电信号排除,获得保留的电信号。第四,判断保留的电信号是否为单台阶电信号:如果保留的电信号为单台阶电信号,则将该单台阶电信号纳入标准信号数据库,且将该单台阶电信号视为该氨基酸的标准信号;如果保留的电信号并非单台阶电信号(例如双台阶电信号),则重新设置过滤条件(例如将 SD 阈值上调),并重新获得保留的电信号,进而再次判断重新获得的保留的电信号是否为单台阶电信号。最后,使用单台阶电信号建立标准信号数据库。

[85].**实施例三:利用 M2MspA-N91H 纳米孔检测 3 种非天然氨基酸**

[86].基于上述方法,本实施例进一步检测了 3 种代表性非天然氨基酸,分别为瓜氨酸、L-正缬氨酸、鹅膏氨酸。如图 8A 所示,这 3 种非天然氨基酸引起的阻塞率分别为:瓜氨酸: 0.1912 ± 0.0011 ; L-正缬氨酸: 0.1806 ± 0.0017 ; 鹅膏氨酸: 0.2333 ± 0.0029 。根据氨基酸阻塞率以及其标准差进行判断,3 种非天然氨基酸的信号互不重叠,可相互区分。

[87].如图 8B 所示,上图和下图分别为同一次实验中空白组(Blank)和加入鹅膏氨酸后,1 分钟内的电流阻塞信号散点图。SD 值小于 2 的信号被认为是有效信号,过滤之后,可以清楚地看到加入鹅膏氨酸之后产生新的信号峰。

[88].多肽可以作为底物参与酶促反应,使用上述检测氨基酸的方法可以对酶促反应实现连续监测:通过检测反应体系中游离氨基酸的种类,可以判断酶促反应是否发生以及酶促级联反应的进展。

[89].**实施例四:基于 M2MspA-N91H 纳米孔的多肽成分分析方法:**

[90].在 M2MspA-N91H 纳米孔蛋白嵌入后,立即将电压调整至+50 mV,记录 10 分钟。接着,在样品槽反式面加入 1 μ L 20 mM 氯化铜溶液,记录 10 分钟。同时,

取 8 μL 浓度为 2 mM 的多肽溶液置于 PCR 管中, 37°C 孵育 5 分钟。加入 2 μL 羧肽酶 A1 溶液, 充分混匀, 在 37°C 反应 5 分钟。反应完成后, 将反应产物全部加入样品槽顺式面, 充分混匀, 在 +50 mV 电压下记录 2 小时。检测原理如图 10 所示, 羧肽酶 A1 (外肽酶的一种) 可以从多肽 C-末端逐一水解氨基酸, 水解得到的游离氨基酸产物被纳米孔捕获从而被检测, 根据出现的氨基酸信号的种类和数量可以判断多肽成分。

[91]. 由于 1 U 外肽酶在 1 min 内能够切割 1 μmol 多肽的末端 1 个氨基酸, 可以根据以下公式计算多肽溶液的终浓度 (假设该多肽能够被完全水解):

$$[92]. b = \frac{c \times d}{a - 1}$$

[93]. 式中, a 为多肽长度, b 为物质的量 (μmol), d 为外肽酶酶活 (U), c 为反应时间 (min)。

[94]. 在纳米孔系统中, 游离氨基酸的浓度为 1 μM 时, 检测效果较好; 如果游离氨基酸的终浓度低于 1 μM 时, 可能会使游离氨基酸的信号减少, 进而影响检测效果。考虑到将水解产物 (反应产物) 加入到 M2MspA-N91H 纳米孔系统后, 可能会被稀释至少 10 倍, 因此, 基于上述公式, 反应体系中多肽溶液终浓度应大于 1 μM 的 10 倍 (即 10 μM)。

[95]. 利用外肽酶水解多肽的末端氨基酸, 获得水解产物 (反应产物)。上述反应产物加入 M2MspA-N91H 纳米孔系统并施加驱动力后, 反应产物中游离氨基酸穿过所述 M2MspA-N91H 纳米孔并产生电信号, 并获得相应的电信号数据。具体的数据分析流程如图 12 所示: 首先, 通过 M2MspA-N91H 纳米孔检测上述反应产物, 获得电信号数据。第二, 建立标准信号数据库 (如上所述)。第三, 设置过滤条件 (电信号的 SD 阈值)。第四, 根据过滤条件, 将部分电信号排除, 获得保留的电信号。第五, 通过标准信号数据库中所述标准信号的阻塞率, 判断所述保留的电信号对应的氨基酸种类。第六, 获得所述游离氨基酸的氨基酸种类信息。

[96]. 图 7 显示了基于上述多肽成分分析方法和数据分析方法分析四种不同多肽的组成。图 7A 为在 0.5 M KCl、10 mM MOPS、pH 7.5 条件下, 10 U 羧肽酶水解 10 mM EF 多肽 20 分钟后, 将反应液加入纳米孔体系检测得到的电流信号图。图 7B 为图 7A 涉及实验中统计得到的所有电流信号的阻塞率及过孔时间的散点

图。可观察到 SD 值位于 1-2 之间的信号（红色）有 3 个分布峰，阻塞率 0.4 附近的为 EF 多肽的信号；左边两个峰与谷氨酸（E）和苯丙氨酸（F）的阻塞率最为接近，故判断为氨基酸 E 和 F 的信号。

[97].图 7C 显示了 M2MspA-N91H 纳米孔分别检测三种多肽(P1: APRLRFYSL; P2: DRVYIHPEHL; P3: RPVKVYPNGAEDESAEAFPLEE；其中下划线表示多肽 C 端的这些氨基酸能够被羧肽酶 A1 水解为游离氨基酸）的水解产物的实验结果。上面的分析结果表明，本发明提供的多肽成分分析方法能够对多肽末端氨基酸组成进行分析。

[98].本发明提供的上述多肽成分分析方法，可以应用于多种不同场景，例如已知多肽的鉴定和未知多肽的预测：

[99].肿瘤新生抗原作为在正常组织中不表达，而仅在肿瘤组织表达的抗原，是理想的免疫治疗靶标。每位患者都有其独特的新抗原谱，可以用上述多肽成分分析方法来筛选肿瘤新生抗原肽：首先，利用高特异性的提取方法，提取并纯化已知肿瘤细胞表面抗原多肽，纯化后的混合物理论上仅含有少量且成分已知的抗原肽；其次，使用上述多肽成分分析方法对纯化后的混合物进行检测；最后，将检测结果中检测出的氨基酸种类与已知抗原肽的氨基酸种类进行对比，可以判断某些氨基酸是否产生或缺少，进而判断肿瘤新生抗原是否可能存在于肿瘤细胞表面。

[100].目前，一些多肽药物已被开发并应用于临床，但由于合成工艺的限制，导致多肽药物的产品质量和纯度等方面达不到要求。通过上述多肽组成成分分析方法，可以实现对多肽药物的纯度的检测：将待检测的多肽药物的氨基酸成分检测结果，与预期合成的氨基酸种类进行对比，能够一定程度上判断该多肽药物的合成质量。

[101].此外，本发明提供的多肽成分分析方法还可以与机器学习结合，来预测未知多肽：将待检测的未知多肽用外切酶水解并检测游离氨基酸；使用机器学习的方法，结合外切酶的种类、水解特点以及氨基酸成分检测结果，与数据库中已知多肽进行比对，能够在一定程度上预测该未知多肽的序列，同时起到缩小未知多肽范围的作用。

[102].上面结合附图对本发明的实施例进行了描述，但是本发明并不局限于上述的具体实施方式，上述的具体实施方式仅仅是示意性的，而不是限制性的，本领域

域的一般技术人员在本发明的启示下，在不脱离本发明宗旨和权利要求所保护的范围情况下，还可做出很多形式，这些均属于本发明的保护之内。

权利要求书

- 1.一种基于 M2MspA-N91H 纳米孔的多肽成分分析方法，其特征在于，包括以下步骤：
 - S1 将外肽酶与待检测多肽混合均匀并反应，得到水解产物；
 - S2 将所述水解产物加入纳米孔系统中，所述纳米孔系统包括：M2MspA-N91H 纳米孔、绝缘膜、第一介质、第二介质，其中所述 M2MspA-N91H 纳米孔嵌入在所述绝缘膜中，所述 M2MspA-N91H 纳米孔提供连通所述第一介质和所述第二介质的通道，所述水解产物被加入至所述第一介质；
 - S3 向所述第一介质和所述第二介质之间施加驱动力，所述水解产物中游离氨基酸穿过所述 M2MspA-N91H 纳米孔并产生电信号；
 - S4 对所述电信号进行数据分析，获得所述游离氨基酸的氨基酸种类信息。
- 2.如权利要求 1 所述的方法，其特征在于，所述游离氨基酸包括天然氨基酸和非天然氨基酸。
- 3.如权利要求 2 所述的方法，其特征在于，所述天然氨基酸包括甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、苯丙氨酸、色氨酸、酪氨酸、天冬氨酸、组氨酸、天冬酰胺、谷氨酸、赖氨酸、谷氨酰胺、甲硫氨酸、精氨酸、丝氨酸、苏氨酸、半胱氨酸、脯氨酸中的一种或多种；所述非天然氨基酸可选地包括瓜氨酸、L-正缬氨酸、鹅膏氨酸中的一种或多种。
- 4.如权利要求 1 所述的方法，其特征在于，所述外肽酶包括氨肽酶和/或羧肽酶。
- 5.如权利要求 1 所述的方法，其特征在于，在所述水解产物加入所述纳米孔系统之前，将所述 M2MspA-N91H 纳米孔与铜离子结合。
- 6.如权利要求 1 所述的方法，其特征在于，所述第一介质和所述第二介质的 pH 值包括 6.5-7.5。
- 7.如权利要求 1 所述的方法，其特征在于，所述待检测多肽溶液的终浓度大于 10 μM 。
- 8.如权利要求 1 所述的方法，其特征在于，S4 中所述数据分析步骤进一步包括：
 - i) 建立标准信号数据库；
 - ii) 设定信号过滤条件；
 - iii) 排除部分电信号，获得保留的电信号；
 - iv) 通过标准信号数据库中所述标准信号的阻塞率，判断所述保留的电信号

对应的氨基酸种类；

v) 获得所述游离氨基酸的氨基酸种类信息。

- 9.如权利要求 8 所述的方法，其特征在于，所述信号过滤条件为排除 SD 值大于 SD 阈值的电信号，其中所述电信号的 SD 值按照下列公式进行计算：

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

式中， x_i 为信号每个采集点的电流值， $\bar{x}$ 为信号电流点的平均值，n 为信号采集点的数量。

- 10.如权利要求 1 所述的方法，其特征在于，所述方法进一步包括 S5 将所述游离氨基酸的氨基酸种类信息与所述待检测多肽的预期序列信息进行对比，以判断所述待检测多肽的实际序列与所述待检测多肽的预期序列是否存在差异；

或

S5 将所述游离氨基酸的氨基酸种类信息结合所述外肽酶的信息与数据库中的多肽序列信息进行对比，以预测所述待检测多肽的实际序列。