



## 纳米孔蛋白质测序\*

李梓溢<sup>1,2)</sup> 伊雅琨<sup>1,2)</sup> 刘蕾<sup>3)</sup> 吴海臣<sup>1,2)\*\*</sup>

<sup>(1)</sup> 中国科学院化学研究所活体分析化学重点实验室, 北京分子科学国家实验室, 北京 100190;

<sup>(2)</sup> 中国科学院大学化学科学学院, 北京 100049; <sup>(3)</sup> 中国科学院高能物理研究所纳米生物效应与安全性重点实验室, 北京 100049)

**摘要** 基因组计划的成功不仅深化了人类对基因组结构的理解, 也促使研究重点逐渐转向对蛋白质组学的解析。鉴于蛋白质在调节生物体内众多生理过程中的核心作用, 蛋白质测序技术的发展显得尤为重要。DNA 测序领域取得的巨大成就已经激发了研究者对纳米孔技术在蛋白质测序应用中的兴趣, 特别是其在快速、便捷及成本效益方面的潜力。本文综述了当前蛋白质测序的主要方法, 并特别聚焦于纳米孔蛋白质测序面临的挑战、最新的进展以及重大突破。随着纳米孔技术的持续改进, 该技术有望在蛋白质组学领域提供更高效和更精准的测序解决方案。

**关键词** 蛋白质测序, 纳米孔, 电化学传感, 单分子分析

**中图分类号** O6, Q6

**DOI:** 10.16476/j.pibb.2024.0295

从20世纪90年代初开始, 科学家们用了十多年时间完成了人类基因组计划。伴随着该计划的巨大成功, 人类全基因组已经被破译, 这标志着进入了后基因组时代<sup>[1-2]</sup>。尽管被解码的基因组数量持续增长, 结构基因组学项目正在揭示许多功能未知的蛋白质结构, 但解析蛋白质的结构和功能、全面了解其生理作用机制仍然是科学研究的前沿<sup>[3]</sup>。蛋白质在生命体中扮演着基石的角色, 参与了众多代谢过程和生理反应, 如信息传递、维持生理环境中的pH和体液平衡, 为细胞和生物体提供结构支撑等<sup>[4]</sup>。遗传组学中, 基因信息通过中心法则编码流向RNA, 再将RNA翻译成密码子对应的氨基酸, 这些氨基酸按照既定的顺序排列生成多肽链, 构成蛋白质的一级结构。然而, 蛋白质中蕴含的信息不仅仅是氨基酸的排列顺序, 还有更丰富的信息深埋在蛋白质的后加工过程中, 包括翻译后修饰、蛋白质的拼接与降解等<sup>[4-6]</sup>, 这些信息无法通过简单的基因解码获得。因此, 亟需一种快速、灵敏且高通量的分析技术来探索蛋白质组学的更深层次秘密。

### 1 现有的蛋白质测序技术

#### 1.1 基于Edman降解法的测序

对蛋白质测序的尝试可以追溯到20世纪中叶,

当时Edman<sup>[7]</sup>提出了一种测定纯化多肽氨基酸序列的方法。这种方法的基本原理是使用异硫氰酸苯酯(phenyl isothiocyanate, PITC)与多肽N端的 $\alpha$ 氨基偶联生成苯氨基硫甲酰肽(phenyl thiocarbonyl-AA, PTC-AA), 然后在强酸(三氟乙酸(trifluoroacetic acid, TFA))条件下, 第一个氨基酸从多肽链上裂解下来, 切割下来的氨基酸随后被送往高效液相色谱(high performance liquid chromatography, HPLC)中进行鉴定和识别(图1a)。在进行了多个循环后, 多肽的氨基酸序列可以被逐一识别。Edman降解法的优势在于它能够非常准确地识别短肽, 几乎不需要预处理即可开始测试。它可以直接分析氨基酸信息, 并且不需要借助任何蛋白质数据库, 因此, Edman测序可以对全新的没有数据库信息的蛋白质进行N端从头测序。同时, 对于具有相同/相近质量的氨基酸, 如亮氨酸与异亮氨酸、赖氨酸与谷氨酰胺, Edman降解法也可轻松区分。借助色谱法的定量分析以及控制降解分子的定量加入, 它也具有定量分析的潜力。然而这种方法在大规模蛋白质测序中难以应用。首先,

\* 国家自然科学基金(22025407)资助项目。

\*\* 通讯联系人。

Tel: 010-82614948, E-mail: haichenwu@iccas.ac.cn

收稿日期: 2024-07-02, 接受日期: 2024-08-24

获得高纯度的多肽是一大难点;其次,每一轮修饰与切割都需要花费大量时间(约1 h/轮)。此外,由于反应体系的复杂性,在进行多次降解后,体系中的原料和氨基酸残余会大大干扰检测的准确度,从而将Edman降解法的读长限制在50个氨基酸以内,这远远不够覆盖大多数生物蛋白质的长度<sup>[8]</sup>。同时,翻译后修饰的存在将大大降低Edman降解的测序准确度,因为一旦N端没有游离的 $\alpha$ 氨基,切割循环就会被迫停止<sup>[9]</sup>。再者,Edman降解需要大约100 pmol的纯肽<sup>[10]</sup>,这一点也极大阻碍了这种方法的普适性。

## 1.2 基于质谱方法的测序

质谱方法至今仍是蛋白质组学分析方面的金标准,质谱方法通过在离子源中使分析样品的各组分电离,生成具有不同质荷比的带电离子。这些离子在电场的作用下进入质量分析器,通过对电离离子的精密分析,可以获得蛋白质的序列信息。目前发展的主流蛋白质测序的质谱方法分为自下而上(bottom-up)的方法和自上而下(top-down)的方法,其中自下而上的方法目前被普遍应用。

所谓自下而上,即从细胞或者组织中提取出待测蛋白质,再通过使用蛋白酶(通常是胰蛋白酶)将蛋白质裂解为多肽碎片,对这些多肽进行电离随后分析检测这些离子化的肽,并进行质量分析(图1b)<sup>[11-12]</sup>。测得的碎片肽的质量可以作为待测蛋白质的指纹,通过比对搜索Mascot或者Sequest等数据库得到蛋白质的相关信息。自下而上的测序方法

发展至今已经非常灵敏,它不是对于蛋白质氨基酸的逐个测量,而更像是一种基于指纹的推测,这就带来因为异构体的准确度缺失,例如亮氨酸和异亮氨酸,它们拥有相同的分子质量却有着不同的结构。这种方法的另一个局限性是需要把蛋白质裂解为0.8~3 ku的碎片,而且数据库匹配的过程只是将碎片和蛋白质进行对照,最后用各种方法对匹配到的蛋白质进行加权评分,得到最可能的结果,这时,同源蛋白质相互之间的干扰就会很大。

另一种质谱测序的方法是自上而下法,这种方法对比自下而上法不需要酶切的步骤,它直接从细胞或者组织中提取目标蛋白质,通过电喷雾离子化将完整的蛋白质离子引入气相,然后在质谱仪中通过碰撞诱导解离或更简单的电子捕获解离或电子转移解离对其进行碎片化,从而得到蛋白质和碎片离子的质量(图1c)<sup>[13]</sup>。经过足够多的碎片分析后,蛋白质的序列信息以及修饰信息都可以被完全破译出来。这种方法相对于自下而上法来说更加便捷,代价是灵敏度仅为自下而上法的1/100左右,因为相比多肽碎片离子分析,大分子质量蛋白质离子的分析对于仪器灵敏度的要求更高。同时由于蛋白质复杂的多级结构,一旦蛋白质过大,将其电离成完整的蛋白质离子就变得非常困难,因此这种方法将待测蛋白质的分子质量通常限制在70 ku甚至50 ku之下。同时,这种方法也需要数据库的搜索与比对,数据库的不完整性、低信噪比事件和翻译后修饰都会影响质谱测序的准确度。

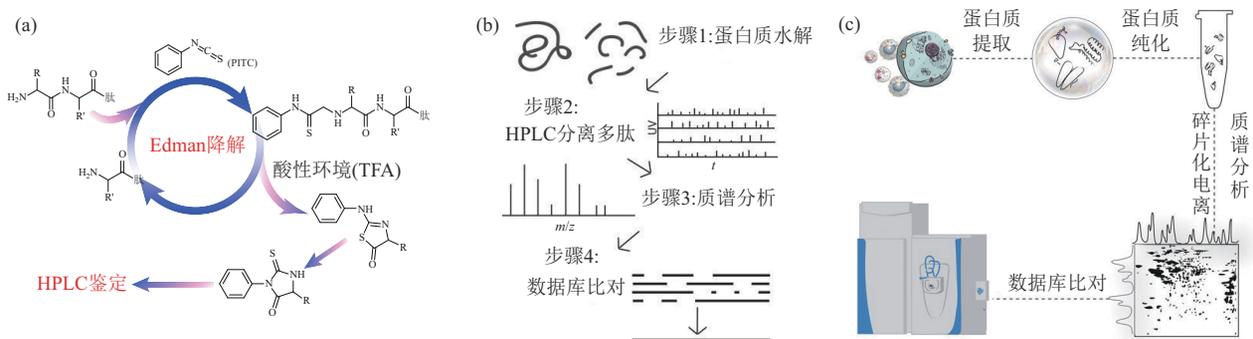


Fig. 1 Traditional protein sequencing methods

图1 传统蛋白质测序方法

(a) Edman降解的循环流程。纯化的多肽N端的游离氨基与PITC反应后在酸性环境下N端的第一个氨基酸从多肽链上裂解下来,随后通过HPLC对氨基酸进行鉴定并循环上述步骤。(b) 自下而上的质谱测序流程图。在纯化出目标蛋白质后,用不同的蛋白酶消化分离出的样品得到多肽,然后结合HPLC和质谱法对肽段进行鉴定,最后用消化产物的序列来计算组装全长蛋白质序列。经授权后改编自文献<sup>[12]</sup>。(c) 自上而下的质谱测序流程图。完整的蛋白质离子被引入气相并被破碎,然后通过质谱分析来确定蛋白质和蛋白质离子碎片的质量,然后匹配数据库对其进行解析,以揭示蛋白质序列。经授权后改编自文献<sup>[13]</sup>。

在蛋白质测序领域，质谱方法依然保持着领先地位，不仅仅是因为其高准确度、高通量以及可重复性，还有强大的全面的且有针对性的前端样品制备工作流程，与后端的质谱技术的巨大发展相辅相成，仪器发展逐渐成熟，研究方法被逐步完善。然而质谱方法也存在一些局限性，包括动态范围不足（通常为 $10^4\sim 10^5$ ，而蛋白质组学分析需要 $10^8\sim 10^9$ ）<sup>[14-16]</sup>、灵敏度不够高（通常检出限为10 amol/L，而蛋白质组学分析可能需要1~1 000 fmol/L的蛋白质）<sup>[17-18]</sup>、读长较短（待测蛋白质分子质量限通常限制在70 ku以下）<sup>[19]</sup>以及序列覆盖度低（自上而下的方法在测序时会损失接近75%的序列信息）<sup>[20-21]</sup>等问题，因此要完全揭秘人类蛋白质组学，需要发展一种准确度高、灵敏度高、通量大且性价比高的新方法<sup>[22]</sup>。

### 1.3 蛋白质测序的新兴技术

为了克服上述挑战，实现蛋白质测序，近十年来发展出多种新技术，包括隧穿电流法<sup>[23-25]</sup>、荧

光指纹法<sup>[26-30]</sup>和实时动态荧光法（表1）<sup>[31]</sup>。隧穿电流法第一次用于蛋白质测序相关尝试的是2014年，Lindsay小组将其应用于氨基酸和短肽的单分子识别<sup>[23]</sup>。在他们的实验体系中，两个相距2 nm的金属电极上覆盖着一层与电极紧密结合的识别分子。这些识别分子与目标分析分子形成较弱的非共价接触，当在电极间施加一个小的电压时，这些被非共价作用捕获的分子会以kHz的频率产生一串随机的电流尖峰（图2a），这与分子的热振动有关，其分布与经典光谱学类似，包含了隧道结中的特征信息。通过分析电流信号的时间、频率以及振幅即可以得到目标分子的指纹信息，再通过使用机器学习解码这些指纹信息即可高效地得到目标分子的性质。利用这种技术，Lindsay小组<sup>[23]</sup>实现了对氨基酸DL异构体、同分异构体以及甲基化氨基酸的识别。同年，Ohshiro等<sup>[24]</sup>利用相同的策略实现了12种不同的氨基酸和翻译后修饰磷酸氨基酸的识别，并且进一步实现了短肽的部分测序。

Table 1 Emerging protein sequencing methods

表1 新兴的蛋白质测序方法

新兴的蛋白质测序方法	优势	劣势	理论适用范围	参考文献
隧穿电流	单分子方法；待测蛋白质无需化学标记；从头测序	分辨率不够高，无法实现20种氨基酸的识别	全长测序	[23-24]
荧光指纹测序	基于Edman降解的荧光指纹法	只能识别序列中被标记的氨基酸；单个循环的时间较长；多轮降解后出现副产物和原料的干扰；需要数据库比对	30个氨基酸以下的多肽	[26-27]
	基于FRET的荧光指纹法	准确测定蛋白质中被标记的氨基酸的位置与数量；信号响应快	只能识别序列中被标记的氨基酸；需要数据库比对	全长测序 [28-30]
实时动态荧光法	实时原位监测；可以同时识别7种氨基酸的荧光信号；待测蛋白质无需化学标记	酶对不同氨基酸的水解速率不均一；单个氨基酸的水解时间较长	全长测序	[31]

伴随蛋白质组学的逐渐完善，指纹识别逐渐被扩展到其他测序方法中，其中荧光指纹测序就是一种可靠的高通量方法。这种方法的最早提出可以追溯到2015年，Macrotte小组<sup>[26]</sup>提出了结合Edman降解和荧光标记的荧光指纹测序的理论设想，并通过蒙特卡罗（Monte Carlo）计算机模拟探讨了其可行性，预测了最可能出现的实验误差，量化了其潜在影响，并讨论了高通量肽测序技术的潜在用途。这种方法选择性地对多肽上的氨基酸进行荧光标记，然后通过Edman降解连续循环去除肽的N端残基，并对单个肽分子荧光强度的相应下降进行成像，由此产生的阶梯式荧光衰减模式通常足以

反映肽的序列，从而通过与参考蛋白质组进行比较，对多肽进行独特的鉴定。同时这种方法还是一种单分子测序的手段，通过对数以亿计的多肽单分子进行高通量的平行测量，对结果进行修正计算，可以得到准确的测序数据。在2018年，Macrotte小组<sup>[27]</sup>将理论转化为了实验，通过在蛋白质片段的半胱氨酸和赖氨酸残基上选择性的标记荧光基团，同时把标记的肽固定，然后使用全内反射荧光（total internal reflection fluorescent, TIRF）对多肽荧光进行定量。重复循环的Edman降解和成像可显示每个分子上的荧光染料强度在每个循环中的下降，从而确定序列中是否存在被荧光标记的氨基酸

及其位置 (图2b) [27]。

Joo小组同期提出了另一种荧光指纹测序的方法, 他们将马达蛋白ClpXP与荧光共振能量转移(FRET)结合, 当马达蛋白拖拽着多肽运动时, 由于标记荧光给体的氨基酸与荧光受体的距离发生变化, 荧光强度也随之改变 (图2c), 通过对荧光谱图的分析, 也可以确定被标记的半胱氨酸和赖氨酸的位置 [28-29]。他们小组通过使用这种方法测得多肽中半胱氨酸和糖基化修饰的位置来确定蛋白质的序列 [30]。这两种方法对于蛋白质测序都有非常大的潜力, 发达的蛋白质组学给它们提供了足够多的数

据库进行比对, 荧光修饰化学提供了可选择的多样化的荧光基团, 广泛研究的Edman降解与蛋白酶给它们提供了良好的测序工具。

近期, Quantum-Si公司提出了一种实时动态荧光的测序方法, 与之前的方法不同, 这种方法将待测多肽固定在纳米腔室中, 并使用氨肽酶对其进行酶解, 每当一个氨基酸被酶切下来, 外面的荧光识别分子则会与之结合产生荧光信号, 通过实时测量荧光强度、寿命等参数, 可以对蛋白质进行从N端的逐一识别 (图2d) [31]。

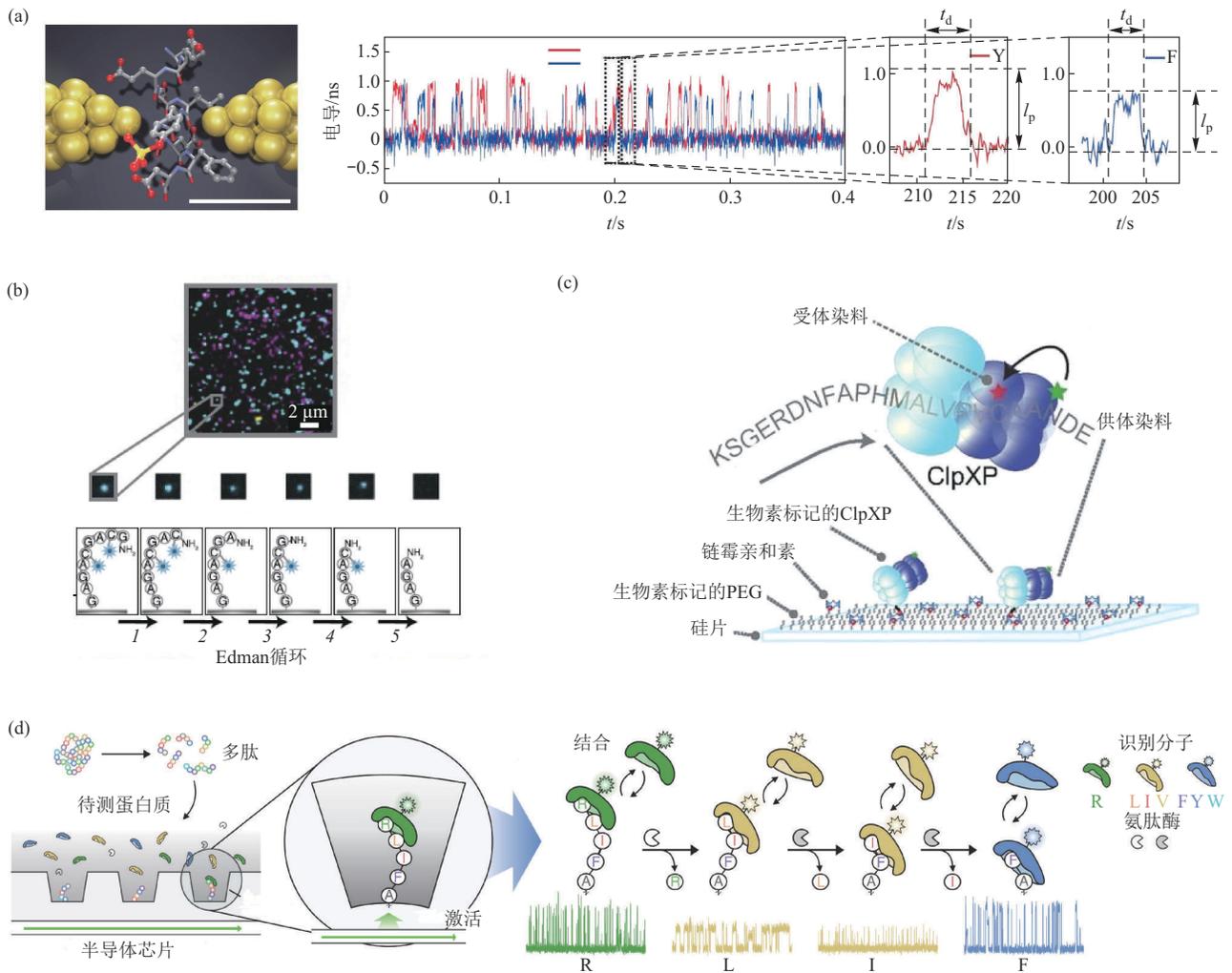


Fig. 2 Emerging protein sequencing methods

图2 新兴的蛋白质测序方法

(a) 隧穿电流法用于蛋白质测序。左: 工作原理示意图, 待测分子夹在两个金纳米间隙电极之间。右: 测量氨基酸Y和F时获得的电导-时间关系曲线。经授权后改编自文献 [24]。(b) 基于Edman降解的荧光指纹法用于蛋白质测序。肽链中特定氨基酸被特异性荧光染料共价标记, 并固定在TIRF单分子显微镜平台灌注室中。通过TIRF对每条肽进行成像, 并通过Edman降解对其N端氨基酸进行化学切除, 从而使每条肽缩短一个氨基酸, 并重新生成其自由N端。重复化学循环 (每去除1个氨基酸) 和成像可显示每个分子中荧光染料的位置。经授权后改编自文献 [27]。(c) 基于FRET的荧光指纹法用于蛋白质测序。在ClpXP酶的作用下, 多肽中被荧光受体标记的氨基酸与荧光给体距离发生变化, FRET的强度随之改变, 由荧光强度可以识别被标记氨基酸的种类和位置。经授权后改编自文献 [29]。(d) 实时动态荧光法用于蛋白质测序。蛋白质样品被消化成肽片段, 固定在纳米反应室中, 与自由扩散的N端氨基酸 (N-terminal amino acid, NAA) 识别器和氨肽酶混合物一起孵育, 进行测序。标记的识别器会在其同源的NAA在N端暴露时与肽结合, 从而产生特征性的脉冲模式。NAA被氨肽酶裂解, 露出下一个氨基酸供识别。经授权后改编自文献 [31]。

新方法为蛋白质测序注入了新活力，它们都具有巨大的发展潜力。隧穿电流法的单分子检测特性使其有望突破现有方法的灵敏度限制，实现单分子蛋白质从头测序，但其仍需解决的问题是如何提高其准确度，能够准确识别20种氨基酸及其翻译后修饰；荧光指纹法只需测出几个被标记的氨基酸的具体位置与数量，对比数据库即可获得蛋白质的部分信息，然而它还是会损失大部分的信息，发展新型的荧光标记以获取更多种类氨基酸的信息是必要的，此外，结合Edman降解的荧光方法仍需克服Edman降解时间长，读长短的缺陷；实时动态荧光法结合氨肽酶从蛋白质N端逐一切割的性质，能够观察到短肽的实时原位测序过程，但是氨肽酶水解速率的不可控性以及对于蛋白质中不同氨基酸的水解速率差异性依旧是一项重大的挑战。

## 2 纳米孔蛋白质测序

### 2.1 纳米孔技术

纳米孔技术是一种基于电化学传感的单分子技术，其基本原理是监测分析物通过纳米孔时产生的电流波动。其基本步骤包括磷脂双分子层的构建、纳米孔的插入以形成离子通道、施加电压产生离子基流和驱使分析物通过纳米孔产生特征阻滞电流信号。通过分析电流信号的振幅变化和平均停留时间，可以推断分析物的特性，而结合事件的频率则能估算其浓度<sup>[32]</sup>。纳米孔技术起源于20世纪80年代末，Deamer等<sup>[33]</sup>提出了DNA穿过纳米孔时ATGC 4种碱基产生不同的阻滞信号并借此对其测序的设想。结合Bayley等<sup>[34]</sup>的研究，Deamer等<sup>[35]</sup>1996年首次在 $\alpha$ 溶血素( $\alpha$ -hemolysin,  $\alpha$ HL)中观测到了DNA与RNA均聚物在纳米孔中产生的特征电流信号。直到2012年，两个小组结合Phi29 DNA聚合酶与纳米孔技术，突破性地实现了对DNA穿孔信号的解析，至此，DNA纳米孔测序的基础性研究初步完成<sup>[36-37]</sup>。同年，牛津纳米孔公司(Oxford Nanopore Technologies, ONT)提出了第一款纳米孔测序仪并于2015年成功实现商业化<sup>[38]</sup>。纳米孔技术在DNA测序方面取得了巨大成功，这激发了其在蛋白质测序中的应用探索。

### 2.2 纳米孔蛋白质测序存在的挑战及应对策略

#### 2.2.1 蛋白质的解折叠

与DNA测序相比，蛋白质测序在纳米孔系统中的复杂性呈指数级增长(表2)。蛋白质不仅具有跨度广泛的尺寸范围和稳定的三维结构，还通过

共价和非共价作用形成特定的空间构象和聚合物。DNA测序常用的生物纳米孔为 $\alpha$ HL与耻垢分枝杆菌膜蛋白A(*Mycobacterium smegmatis* porin A, MspA)。前者由1个开口为2.9 nm、直径为4.2 nm的类球状前庭和1个平均内径为2 nm的 $\beta$ 桶组成，其连接处收缩最窄至1.4 nm；后者则由开口为4.8 nm的锥形前庭和两个 $\beta$ 桶组成，其收缩区域宽为1.2 nm，长为0.6 nm。除此之外，其他生物纳米孔的直径通常为1~2 nm，无法容纳复杂的蛋白质结构<sup>[39]</sup>。而孔径较大的固态纳米孔则在三维尺寸可重复性和识别能力上存在不足。因此，要利用与DNA测序相同的策略，必须首先设法将蛋白质解除三维结构，转变为线性分子，以便其能够穿过纳米孔并逐一识别氨基酸。

DNA通常是双螺旋结构，在DNA解旋酶的作用下，可以很容易以单链形式通过纳米孔，与此类似，蛋白解折叠酶也可以达到这样的效果。在2013年，Akeson小组<sup>[40]</sup>提出了利用AAA+解折叠酶ClpX控制蛋白质通过 $\alpha$ HL纳米孔的解折叠和转运，当蛋白质通过长约5 nm的跨膜孔腔时，可以根据对应的电流信号特征分辨出每种蛋白质的区段(图3a)。

使蛋白质变性也是解折叠的另一种思路，提高温度是其中最便捷的方法，Payet等<sup>[41]</sup>研究了麦芽糖结合蛋白(maltose binding protein, MBP)的突变体MA1E219在20~70°C的穿孔行为(图3b)，他们发现，在温度较低时，由于蛋白质聚集体无法穿过纳米孔，几乎不发生电流阻滞信号，伴随着温度的升高，蛋白质链穿孔的信号明显增多，且测得的蛋白质解链温度大约在45°C，这与圆二色谱测得的结果是一致的。利用强电场也是一种高效的物理方法，Kim等<sup>[42]</sup>观察到了SAP97 PDZ2蛋白在高电压(200~800 mV)下的逐渐伸展行为(图3c)，并将施加电场的解折叠效果与尿素相对比，得出400 mV的电压与3.1 mol/L的尿素对所选蛋白质解折叠效果相当。然而，强电压不适用于所有纳米孔体系，例如在使用生物纳米孔时，当电压大于350 mV，磷脂双层膜可能会被击穿。

另一种效果显著的方法是加入变性剂，常用的变性剂包括尿素<sup>[42-43]</sup>、盐酸胍<sup>[44]</sup>，它们都具有形成氢键的能力，当它们的浓度较高时，可以破坏水的氢键，从而使蛋白质分子内部的疏水残基伸展并增加其溶解性(图3d)。金属离子也是变性剂的一种，Schiopu等<sup>[45]</sup>提出了利用Cu<sup>2+</sup>和组氨酸的结

合, 使淀粉样 $\beta$ 肽解折叠并增加其捕获率的方法。化学方法中, 调节溶液的离子强度和pH也不失为一种行之有效的解折叠手段, Mereuta等<sup>[46]</sup>发现, 通过改变pH和离子强度, 可以改变氨基酸残基之

间的静电相互作用, 进而改变蛋白质的折叠行为, 他们发现, 在酸性条件下, 含组氨酸多肽形成的发夹结构会遭到破坏, 从而增加其穿孔行为(图3e)。

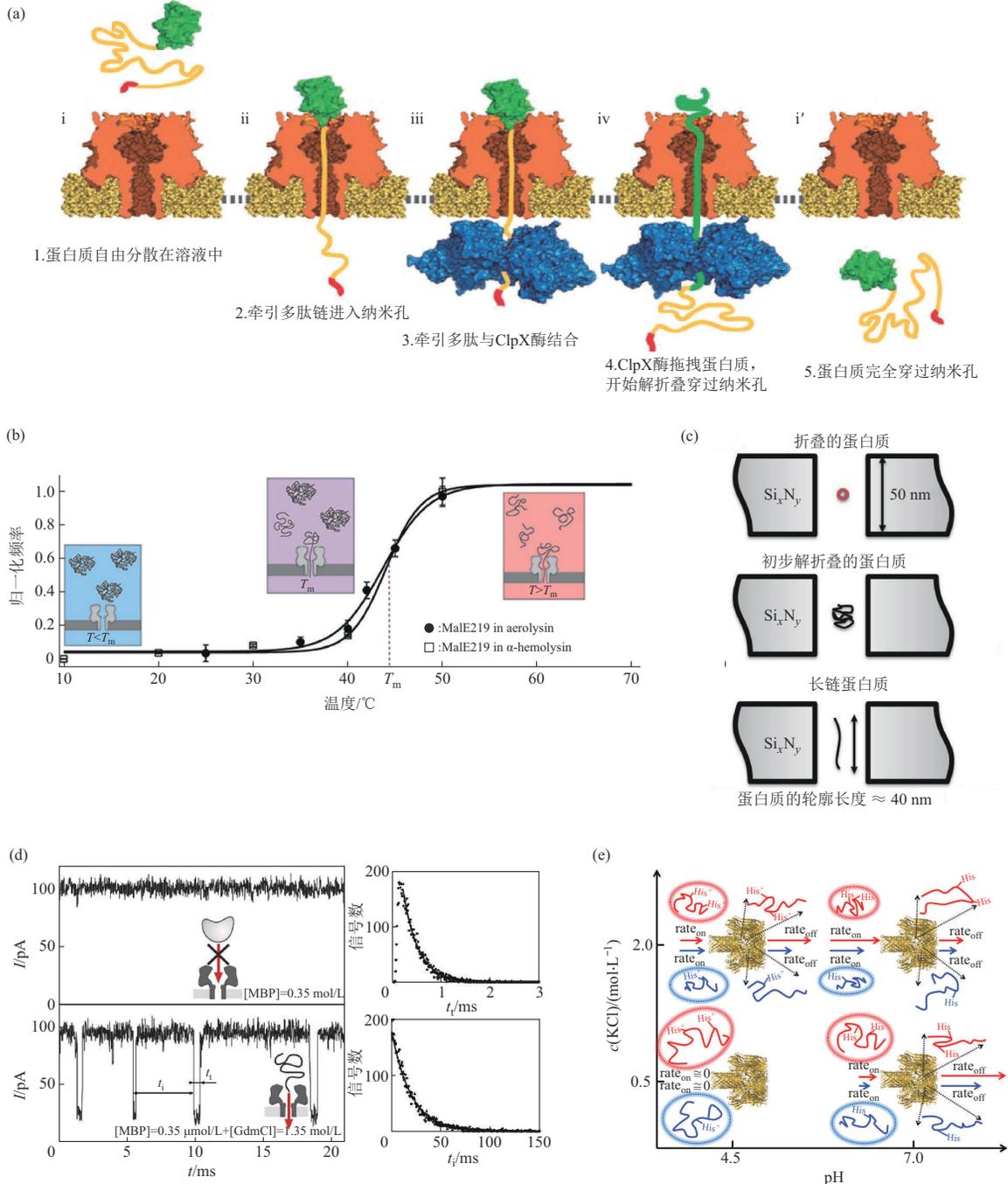


Fig. 3 Protein unfolding for nanopore sequencing

图3 纳米孔测序中蛋白质解折叠策略

(a) 折叠酶ClpX介导的蛋白质 Smt3 的折叠和转运。经授权后改编自文献 [40]。(b) MalE219的热解折叠。随着温度升高, MalE219解折叠穿过纳米孔的信号频率增高, 在大约45°C时达到转变温度。经授权后改编自文献 [41]。(c) 施加高电压使蛋白质解折叠。在高电压(200~800 mV)的情况下,  $Si_xN_y$  纳米孔中的蛋白质逐渐发生解折叠。经授权后改编自文献 [33]。(d) 利用盐酸胍(GdmCl)使MBP解折叠。上: 在没有GdmCl的情况下, 电流恒定在100 pA, MBP无法进入孔道; 下: 在GdmCl为1.35 mol/L的情况下, 当MBP分子进入孔道时, 电流下降到20 pA。经授权后改编自文献 [44]。(e) 调节pH与盐浓度解折叠。CAMA 3 (2个组氨酸残基)和CAMA 1 (1个组氨酸残基)多肽在不同盐浓度和pH值条件下的动力学和折叠行为。经授权后改编自文献 [46]。

### 2.2.2 多肽的定向穿孔

将蛋白质解折叠后面临的第二个挑战是如何让线性多肽定向地通过纳米孔产生电流信号。与DNA不同的是，DNA由均匀带负电的核苷酸组成，在电场的作用下会定向穿过纳米孔。然而，组成蛋白质的20种天然氨基酸仅有4种带有电荷，使蛋白质具有不均一带电性，这导致蛋白质在电场驱动下的穿孔方向和速度不均匀，从而阻碍了氨基酸的顺序识别。为了解决这一难题，需要给多肽施加一个力，让它能够以一个均匀的、可控的速度定向穿过纳米孔。具有均匀带电性的寡聚核苷酸就是一个非常好的选择，Bayley小组<sup>[47]</sup>早在2013年就尝试了将30个胞嘧啶组成的聚核苷酸连接在大肠杆菌硫氧还蛋白上，在电场力的作用下利用聚核苷酸链拖拽着蛋白质边解折叠边穿过纳米孔。Zhang和Lindsay小组<sup>[48]</sup>拓展了这一思路，开发了一种点击化学反应，将聚核苷酸PolyT20高效且具有选择性地连接在多肽的N端，使之穿孔（图4a）。除了核苷酸以外，一些均聚肽也常被作为控制多肽穿孔的牵引序列。Luchian小组<sup>[49-50]</sup>最早通过多肽前后分别连接十二聚精氨酸和十二聚谷氨酸使模型肽在纳米孔的识别区域停留时间增长并对其进行识别（图4b）。近期，Wanunu小组<sup>[51]</sup>将十聚天冬氨酸连接在MBP上，利用机器学习，证明了蛋白质单次穿孔事件包含足够丰富的信息，可以区分其运输方向并根据其指纹信息辨别其身份（图4c）。

连接牵引序列是一个常用的方法，但是它要求牵引序列与待测多肽的高效连接，一些无需反应的控制多肽穿孔的方法也被发展了起来。在纳米孔体系中，电渗流（electroosmotic flow, EOF）常被用于调控物质运输，它不依赖于待测物本身的性质，而是与纳米孔的内部电荷分布情况相关。Luchian及其合作者<sup>[52]</sup>最早通过调节pH来改变 $\alpha$ HL纳米孔的电渗流，他们的实验证明，在pH2.8的溶液环境下，电渗流可以足够大以抵抗电泳力（electrophoretic force, EPF），使多肽在纳米孔中有更长的停留时间，获得更丰富的多肽信息（图4d）。如前文所述，电渗流更多地依赖于纳米孔的内部电荷，因此对纳米孔进行定点突变，将孔中的部分氨基酸残基突变为带电残基，可以更大程度增强电渗流。Maglia小组<sup>[53]</sup>将野生型曲霉毒素C（FraC）纳米孔突变为D10R K159E FraC，极大增强了多肽的捕获能力，并利用纳米孔电流来分辨25 ku至1.2 ku的肽和蛋白质生物标记物，包括只

有一个氨基酸差异的多肽。近期，他们还还对细胞毒素K（CytK）纳米孔的多个位点进行突变，产生了在-100 mV时强度约为20 pN电渗流，使多肽甚至可以对抗电泳力来穿孔（图4e）<sup>[54]</sup>。

### 2.2.3 单个氨基酸的识别

蛋白质测序面临的第3个主要挑战是精确识别单个氨基酸。20种氨基酸的体积范围相差甚微，从0.06 nm<sup>3</sup>到0.23 nm<sup>3</sup>，大小最接近的两种氨基酸体积之差低至0.001 nm<sup>3</sup>，这使得单纯通过体积排阻的原理来区分各个氨基酸极为困难。此外，氨基酸侧链与纳米孔的相互作用的多样性，以及同源氨基酸之间（如谷氨酸与天冬氨酸，仅相差1个亚甲基）的微小差异进一步加剧了这一挑战。还有，蛋白质翻译后修饰也大大增加了蛋白质测序的复杂性。迄今为止，已发现超过300种不同的翻译后修饰，如甲基化、乙酰化、磷酸化、糖基化、硝基化和泛素化等。因此，与DNA测序相比，蛋白质测序需要更高分辨率的识别技术。

无标记的方法是最理想的识别单个氨基酸的方法，但是它要求纳米孔具备极高的空间分辨率。Di Muccio等<sup>[55]</sup>首先对20种氨基酸在 $\alpha$ HL中的电流阻滞情况进行了平衡全原子模拟，他们的研究表明，氨基酸的体积排阻作用在对氨基酸的检测中占主导地位，疏水作用和电荷作用也对电流信号有一定程度的影响，这为纳米孔识别氨基酸提供了理论的可能性。利用半胱氨酸的巯基残基（-SH）与气溶素纳米孔壁上的氢键作用，Ying小组<sup>[56]</sup>实现了对半胱氨酸的特异性识别（图5a）。利用超短单壁碳纳米管的高速离子传输性能，我们小组对3对同源氨基酸进行了区分，获得了足够高的分辨率，为氨基酸的准确识别提供了一种新的器件（图5b）<sup>[57]</sup>。Feng小组<sup>[58]</sup>通过利用工程化的亚纳米级MoS<sub>2</sub>纳米孔识别了20种天然氨基酸中的16种，并且可以识别氨基酸的磷酸化修饰（图5c）。Oukhaled小组<sup>[59]</sup>通过野生型气溶素纳米孔对5~10个氨基酸长度的均聚精氨酸短肽进行实验，发现不同长度的均聚肽产生的电流信号具有独特性，并且电流信号的强度与氨基酸数量呈现线性关系。当混合的赖氨酸和精氨酸肽放入体系中时，也展示了明显的区分度（图5d）。

通过对待测氨基酸进行化学修饰也是一种可行的策略，半胱氨酸的侧链巯基提供了很好的反应位点。Lu等<sup>[60]</sup>将4聚腺嘌呤作为修饰分子，识别了半胱氨酸与高半胱氨酸（图5e）。Dekker及其合作

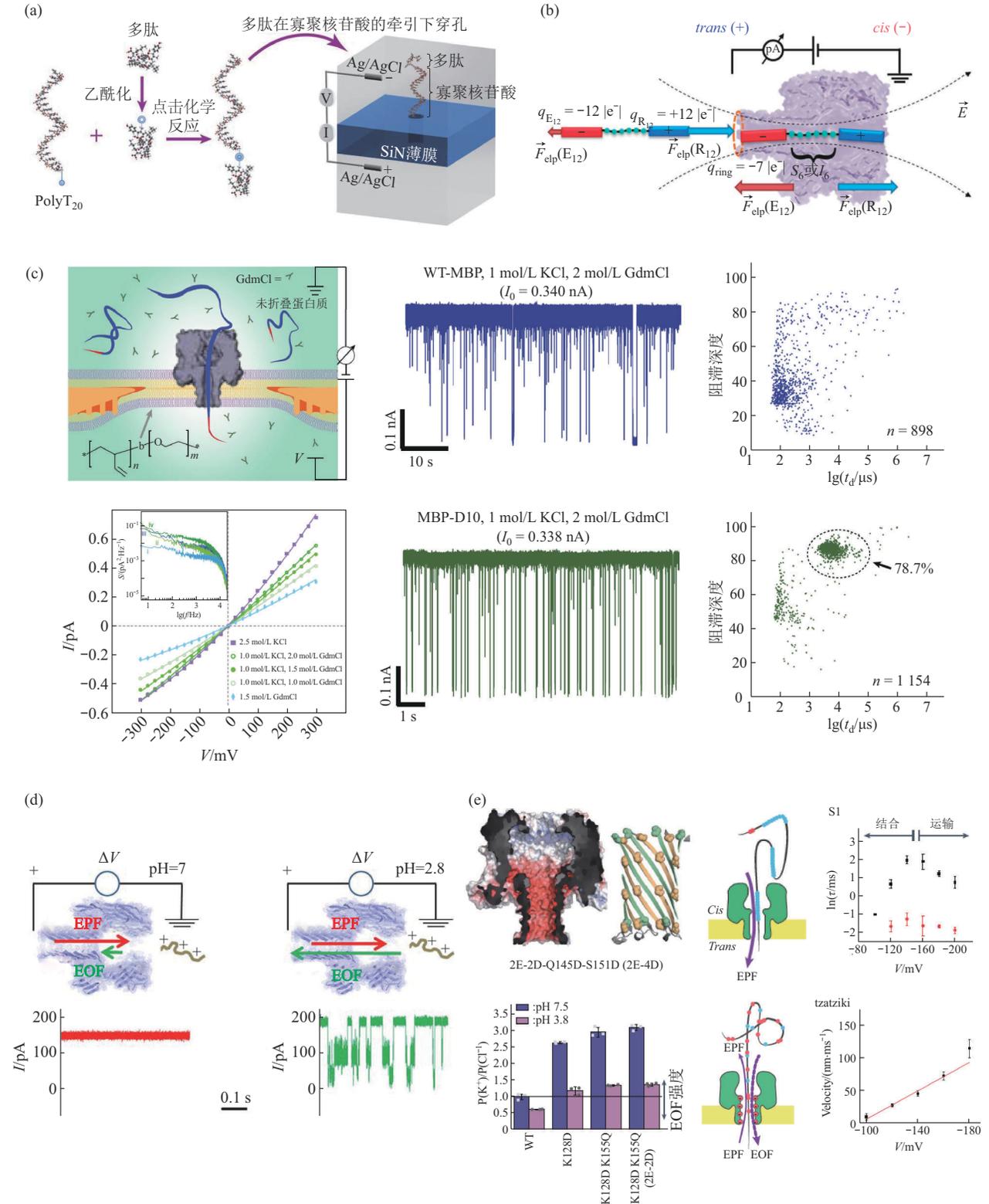


Fig. 4 Controlled peptides translocation for nanopore protein sequencing

图4 纳米孔蛋白质测序中调控多肽定向穿孔的策略

(a) 连接聚核苷酸使多肽定向穿孔。多肽的N端氨基首先被叠氮乙酸酐乙酰化，后与修饰偶氮苯并环辛烷 (aza-dibenzocyclooctyne, ADIBO) 的聚二十胸腺嘧啶通过点击化学反应连接形成多肽-核酸复合物。经授权后改编自文献 [48]。

(b) 连接带电荷的均聚多肽控制多肽穿孔。多肽头尾分别连接E<sub>12</sub>和R<sub>12</sub>，它们分别带大量负电荷和正电荷，当多肽穿孔时产生相反的作用力从而延长多肽与纳米孔的作用时间。经授权后改编自文献 [50]。

(c) 连接聚十天冬氨酸控制MBP定向穿孔。当MBP没有连接D<sub>10</sub>，电流信号呈现为阻滞深度较小、频率较低且阻滞时间较短的碰撞信号；连接D<sub>10</sub>后大部分为蛋白质穿孔信号。经授权后改编自文献 [51]。

(d) pH调节电渗流控制多肽穿孔。在酸性pH值下，EOF会增强以对抗EPF，从而导致肽移动速度减慢，能够检测到带正电荷的多肽。经授权后改编自文献 [52]。

(e) 纳米孔突变调节EOF控制多肽穿孔。将CytK纳米孔进行定点突变，得到带有大量负电的2E-4D-CytK纳米孔，在pH7.5时存在强EOF，可以抵抗EPF促使多肽穿过纳米孔。经授权后改编自文献 [54]。

者<sup>[61]</sup>在多肽的不同位置的半胱氨酸上修饰了6种不同的标记分子，他们都展现出了明显差异化的电流信号，且电流信号的阻滞深度与持续时间都与标记分子的分子质量与电荷呈线性关系。通过将氨基酸连接在7精氨酸均聚肽上，Oukhaled与Aksimentiev等开发了一种可以识别13种天然氨基酸的方法，在他们的实验中，带电的精氨酸肽提供了足够强的电泳力使氨基酸穿过纳米孔，同时聚精氨酸肽链与气溶素纳米孔的相互作用也呈现出了具有高识别度的信号（图5f）<sup>[62]</sup>。受到Edman降解的启发，通过N端的衍生化反应修饰并识别氨基酸也是一个有效的策略，Liu和Wang小组<sup>[63]</sup>尝试了

4种衍生化试剂，异硫氰酸苯酯（PITC）、2-萘基异硫氰酸酯（NITC）、邻苯二甲醛（OPA）和2,3-萘二甲醛（NDA），修饰A、F、Y、H、D 5种氨基酸，且NITC表现出最好的性能，能够完全区分这5种氨基酸（图5g）。后续他们将这种方法进一步优化，通过NDA和NITC的衍生化反应，进一步区分了9种氨基酸（图5h）<sup>[64]</sup>。近期，他们改用异硫氰酸金刚烷酯（ADIC）作为衍生化试剂，相比于之前的衍生化策略，ADIC具有更大的体积，这可以放大氨基酸之间的分辨率，同时增强修饰基团与纳米孔的相互作用（图5i）<sup>[65]</sup>。

Table 2 Challenges and strategies for nanopore protein sequencing

表2 纳米孔蛋白质测序存在的挑战及应对策略

挑战	策略	优势	劣势	参考文献
蛋白质的解折叠	生物方法 解折叠酶解折叠	生物相容性好	具有序列依赖性；解折叠速率不均一	[40]
	物理方法 热解折叠	易于实施	高温使磷脂膜不稳定	[41]
		高电压解折叠	高电压提供更强的电泳力	高电压下磷脂膜被击穿；不适用于生物纳米孔
	化学方法 尿素	解折叠效率高；成本低	需要大量加入	[43]
		盐酸胍	解折叠效率高；成本低	引入过多的H <sup>+</sup> 影响纳米孔的稳定性
	改变pH和盐浓度	易于实施	具有序列依赖性	[46]
多肽定向穿孔	连接牵引序列	电泳力作用下穿孔效率高；可以根据不同的目标分析物改变牵引序列	修饰的普适性不强	[47-52]
	调控电渗流	待测多肽无需修饰；方向可调控；可以抵抗电泳力控制多肽运输	需要对纳米孔突变或改变实验条件	[52-54]
单个氨基酸的识别	无标记识别法	直接分析待测氨基酸	识别精度较低；不同氨基酸的捕获频率范围跨度较大；需要更高分辨率的纳米孔	[56-59]
	化学修饰法	修饰后的分辨率极大提高；修饰的分子可提供电泳力控制氨基酸定向运输	修饰过程繁琐；修饰普适性不强	[60-65]

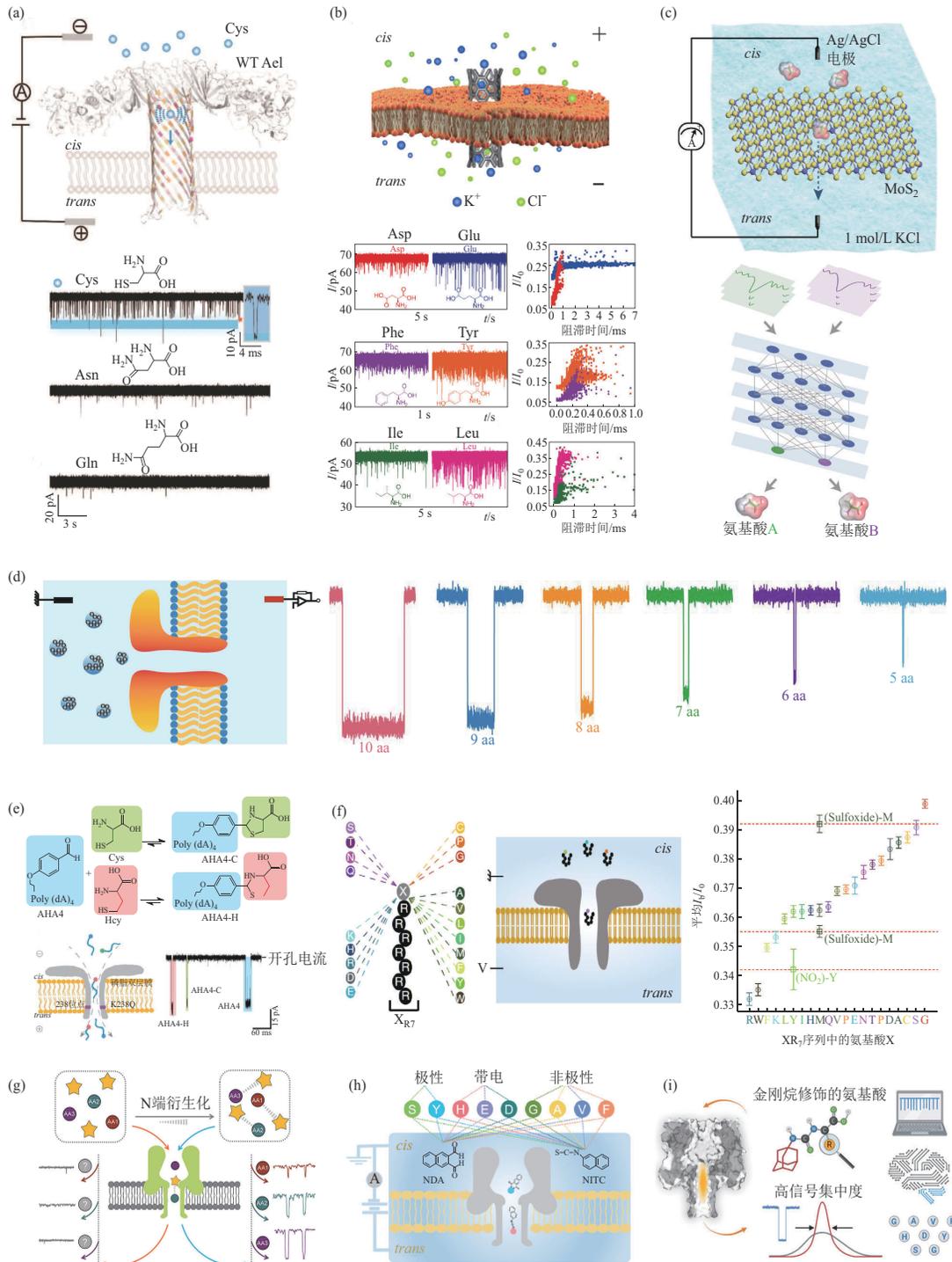
翻译后修饰极大增加了蛋白质测序的复杂性，再过去的几十年里，不少研究者都尝试了利用纳米孔技术进行翻译后修饰的检测，已经有多种翻译后修饰被检测出来。磷酸化是最常见的翻译后修饰，早在2014年，Bayley小组<sup>[66]</sup>就通过 $\alpha$ HL区分了硫氧还蛋白上的两个磷酸化位点（图6a）。Restrepo-Pérez等<sup>[67]</sup>在模型肽上同时进行了糖基化和磷酸化的修饰与检测，显著的区分效果再一次展示了FraC纳米孔在蛋白质测序方面的巨大潜力（图6b）。在2020年和2022年Long小组<sup>[68-69]</sup>实现了对tau蛋白上磷酸化与乙酰化的区分（图6c）。甲基化一直是翻译后修饰的一大挑战，因为它带来的结构改变很小，只有一个—CH<sub>3</sub>的改变，Ensslen等<sup>[70]</sup>利用工程化的气溶素纳米孔的高分辨率攻克了这一

困难（图6d）。除此之外硝基化、氧化（图6e）和泛素化修饰（图6f）也已经通过纳米孔方法得到了检测<sup>[71-73]</sup>。

## 2.3 纳米孔蛋白质测序的方法进展

### 2.3.1 纳米孔全链长测序

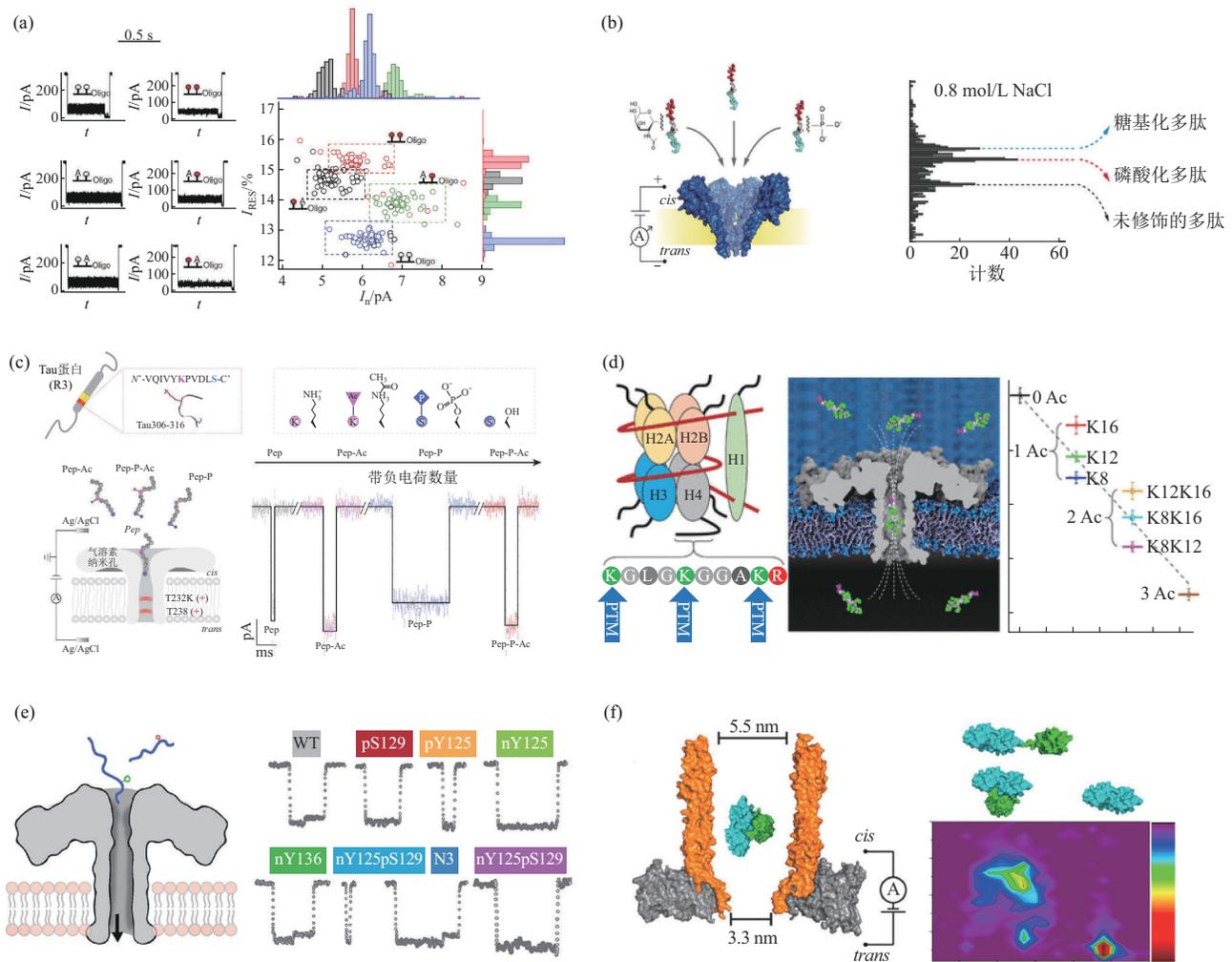
受到DNA测序的启发，实现蛋白质测序的一个理想方法是将蛋白质解折叠为单链多肽，在分子马达的驱使下使其在纳米孔中线性移动，当多肽链经过纳米孔时由于不同氨基酸与纳米孔壁之间的相互作用的区别，会产生特定的电流信号，从而得到蛋白质的序列信息。这种全链长的测序方式需要稳定速率的分子马达，同时需要纳米孔的分辨率足够高以读取每一个氨基酸的易位信息。Huang课题组<sup>[74]</sup>将待测多肽通过化学方法连接到寡聚核苷酸



**Fig. 5 Identification of individual amino acids for nanopore protein sequencing**

**图5 纳米孔蛋白质测序中对单个氨基酸的识别**

(a~d) 使用无标记方法识别氨基酸。(a) 利用Ael纳米孔分析半胱氨酸 (Cys)、天冬酰胺 (Asn) 和谷氨酰胺 (Gln)。经授权后改编自文献 [56]。(b) 利用单壁碳纳米管分析单个氨基酸。不同氨基酸的电流信号与信号散点图均用对应颜色标注, Asp (红色)、Glu (蓝色)、Phe (紫色)、Tyr (橙色)、Ile (绿色) 和Leu (玫红色)。经授权后改编自文献 [57]。(c) 在MoS<sub>2</sub>纳米孔中检测单个氨基酸。上: 实验装置示意图; 下: 使用深度学习网络鉴定氨基酸的流程图。经授权后改编自文献 [58]。(d) 使用Ael纳米孔检测由5、6、7、8、9和10个精氨酸组成的均聚肽 (从右到左)。经授权后改编自文献 [59]。(e~i) 化学修饰法识别氨基酸。(e) 使用Ael对5'-苯甲酰聚腺嘌呤修饰的半胱氨酸 (Cys) 与高半胱氨酸 (Hcy) 的高灵敏度区分。经授权后改编自文献 [60]。(f) 多肽XR<sub>7</sub>中氨基酸 (X) 的识别。左: XR<sub>7</sub>的结构, 其中待测氨基酸 (X) 与1个多阳离子载体R<sub>7</sub>相连; 右: 20种XR<sub>7</sub>的电流阻滞深度。经授权后改编自文献 [62]。(g) 使用PITC、NITC、NDA、OPA对Ala、Asp、Phe、Tyr、His 5种氨基酸进行衍生化并利用αHL对其进行鉴别。经授权后改编自文献 [63]。(h) 使用NDA和NITC对9种氨基酸进行衍生化并采用αHL对其进行鉴别。经授权后改编自文献 [64]。(i) 使用ADIC对9种氨基酸进行衍生化并利用αHL对其进行鉴别。经授权后改编自文献 [65]。



**Fig. 6 Identification of post-translational modifications for nanopore protein sequencing**

**图6 纳米孔蛋白质测序中对翻译后修饰的识别**

(a) 不同磷酸化状态的Trx单分子纳米孔检测。4种不同磷酸化状态的Trx均表现出明显的电流信号差距。经授权后改编自文献 [66]。(b) 使用FraC纳米孔对带电模型肽上磷酸化和糖基化修饰的检测，无修饰、磷酸化修饰和糖基化修饰的模型肽均展现出不同的电流阻滞深度。经授权后改编自文献 [67]。(c) 使用AeL纳米孔对野生型、乙酰化、磷酸化和磷酸化-乙酰化tau蛋白的区分，它们都展现出不同的  $I/I_0$ 。经授权后改编自文献 [69]。(d) 使用R220S突变AeL纳米孔对全长人类H4蛋白片段中乙酰化和甲基化的检测。经授权后改编自文献 [70]。(e) AeL纳米孔对 $\alpha$ 突触核蛋白中磷酸化、硝基化与氧化的检测。经授权后改编自文献 [73]。(f) 工程化ClyA纳米孔对E2F2和泛素E2结合物的识别。经授权后改编自文献 [71]。

上，形成多肽-寡聚核苷酸嵌合体 (POC)，POC 上的核苷酸链与互补链通过碱基互补配对后进入测试体系中，电泳力驱动 POC 通过纳米孔并将配对的 DNA 链解开 (图 7a)。测试体系中的 phi29 DNA 聚合酶以 POC 上的单链 DNA 作为模板进行双链合成，在这个过程中 DNA 聚合酶充当分子马达边合成双链边将 POC 末端的多肽对抗电泳力拖拽通过纳米孔的收缩窄口，此时会产生多肽序列依赖性的信号。同时，在 POC 中存在一个信号标志物，在信号分析的过程中，POC 的特征信号可以作为一个锚点，其后产生的电流信号即为多肽穿过纳米孔

的测序信号。类似的，Chen 等 [75] 利用 Hel308 家族成员 MTA 对单链 DNA 的结合能力，控制 DNA-肽-DNA 偶联物在 MspA 中的易位，也获得了肽的测序信号 (图 7b)。

同期，Dekker 小组 [76] 采用类似的策略进行了尝试，与其他两个小组不同的是，他们改用了 Hel308 DNA 解旋酶，这种酶相比于 phi29 DNA 聚合酶具有更小的步进距离，每次拖拽核苷酸链的距离只有大约半个核苷酸的长度，这意味着它可以为单个氨基酸的识别提供更高的空间分辨率 (图 7c)。除此之外，他们利用胆固醇将互补链锚定在

孔附近, 同时添加高浓度的解旋酶, 使酶拖拽 POC 解旋-酶与 POC 配对脱离-POC 向下运动-酶拖拽 POC 解旋这一过程可以反复发生, 这样可以连续控制同一个 POC 在纳米孔中反复穿梭, 实现对同一个分子的多次读取。更高的空间分辨率以及更多的数据量, 再结合机器学习的高速自动化信号解读, 序列读取的错误率小于  $10^{-6}$ 。近期, Dekker 小组 [77] 用这种策略相继实现了对蛋白质和多肽上翻译后修饰的检测 (图 7d)。

这些工作的发表展现了纳米孔蛋白质链测序的极大潜力, 作为一种单分子分析手段, 它还存在着一些待解决的挑战: a. 由于 MspA 纳米孔前庭的长度

限制 (~10 nm), 导致从开始读取多肽序列信号到结束, POC 最多只能步进 10 nm, 这意味着这种方法的读长会被限制在 17 个氨基酸以内, 这对于真正实现蛋白质的从头测序是远远不够的; b. 在 DNA 测序的过程中, 每一个信号台阶被认为是纳米孔收缩区域附近 4 个碱基的共同贡献, 结合 4 种构成 DNA 的碱基, 这就可能导致  $4^4$  (256) 种信号台阶的产生 [37], 而蛋白质由 20 种氨基酸组成, 且氨基酸的平均体积小于核苷酸, 估计会产生大约  $20^8$  的信号种类 [58, 76], 这对于信号的分析 and 处理是一个极大的挑战。

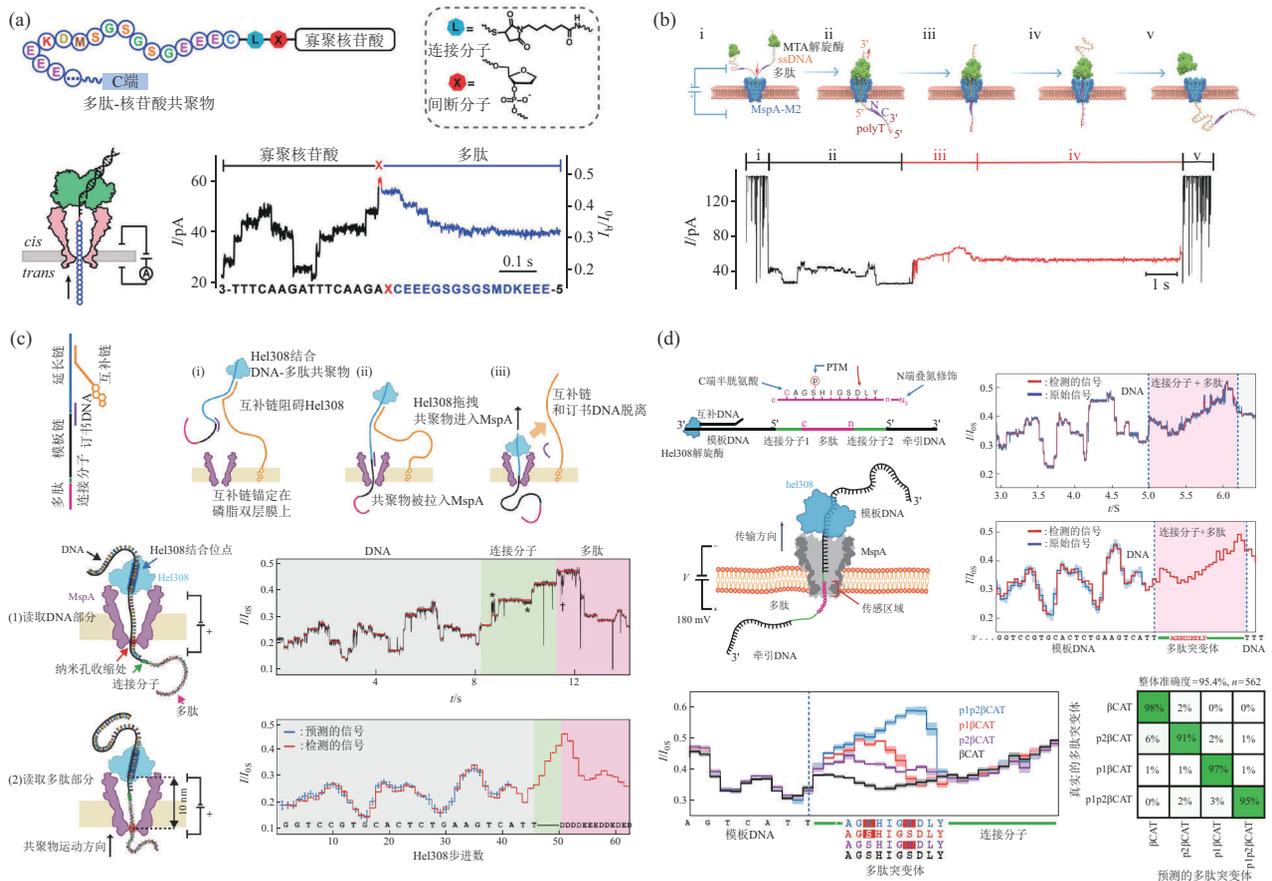


Fig. 7 Full-length chain sequencing methods for nanopore protein sequencing

图7 纳米孔蛋白质全长链测序

(a) DNA 聚合酶驱动多肽-核苷酸共聚物通过 MspA 纳米孔实现肽测序的示意图。缺碱基位点 (X) 作为信号标记将寡聚核苷酸和肽分开。经授权后改编自文献 [74]。(b) MTA 解旋酶驱动 DNA-肽共聚物通过 MspA 纳米孔实现肽测序的示意图。下方为 DNA-多肽共聚物的测序信号。经授权后改编自文献 [75]。(c) Hel308 DNA 解旋酶驱动 DNA-肽共聚物通过 MspA 纳米孔实现肽测序的示意图。中间为运动过程中 DNA 部分通过纳米孔产生的信号。下方为多肽部分通过纳米孔产生的信号。经授权后改编自文献 [76]。(d) Hel308 DNA 解旋酶驱动 DNA-肽共聚物通过 MspA 纳米孔实现多肽中磷酸化位点的识别。经授权后改编自文献 [77]。

### 2.3.2 酶切辅助的纳米孔蛋白质测序

全链长的纳米孔测序方法被开发的同时, 酶切

辅助的纳米孔测序方法也是研究人员攻坚的另一个方向。在理想状况下, 可以利用蛋白酶将蛋白质水

解为单个的氨基酸，被酶切割下来的氨基酸逐一通过纳米孔并产生电流信号，特征的电流信号对应于不同的氨基酸分子，从而实现对目标蛋白质从头测序。在这种方法中，首先是要完全区分组成蛋白质的20种氨基酸，Huang课题组<sup>[78]</sup>利用工程化的MspA纳米孔实现了这一目标（图8a）。受到Bayley等<sup>[79]</sup>利用单巯基突变的 $\alpha$ HL纳米孔结合 $\text{Cu}^{2+}$ 并检测DL氨基酸的工作的启发，他们对孔口更窄的MspA纳米孔进行了单巯基突变，并进一步用点击化学反应在孔口处修饰了次氨基三乙酸用于络合 $\text{Ni}^{2+}$ ，利用 $\text{Ni}^{2+}$ 与氨基酸游离的氨基和羧基的配位作用，当氨基酸通过纳米孔时则会产生特征电流阻滞信号。结合机器学习对这些信号进行分析，他们实现了对20种氨基酸高达98.8%准确度的识别。为了更接近蛋白质测序，他们测试了亮氨酸氨肽酶对至多包含8种不同氨基酸的多肽水解产物并成功获得了这8种氨基酸的独立信号。除此之外，该系统还可以用于甲基化、乙酰化、糖基化和磷酸化的翻译后修饰氨基酸的检测。这种方法展现出了极高的氨基酸分辨率，标志着纳米孔蛋白质测序又往前迈出了一大步，然而它仅仅获得了组成多肽的氨基酸的种类信息，遗失了氨基酸的序列信息。无独有偶，Geng课题组<sup>[80]</sup>采用组氨酸突变的MspA络合 $\text{Cu}^{2+}$ 并进一步络合识别20种氨基酸，也达到了类似的区分效果，他们还将氨基酸的浓度和捕获频率进行了相关性处理，并最后利用归一化的酶切产物的信号频率占比得到了部分多肽的序列信息（图8b）。

我们小组近期发表的一篇论文获得了一些突破性的进展<sup>[81]</sup>。在之前的工作中，我们发现苯丙氨酸（F）和葫芦[7]脌（CB[7]）之间的主客体相互作用能使分子穿孔时产生持续且均一的信号，从而显著提高纳米孔的识别能力<sup>[82-84]</sup>。通过实验优化，我们设计了FGXD<sub>8</sub>模型肽，其中“X”代表20种天然氨基酸中的任何一个。带有负电荷的天冬氨酸链（D<sub>8</sub>）为电场中的转运提供了必要的驱动力，而FG和CB[7]之间的强大相互作用确保了肽CCB[7]复合物在 $\alpha$ HL纳米孔收缩处的稳定定位。这种特殊配置使“X”位于纳米孔收缩处附近，实现了最佳分辨率，结合该探针和其他突变蛋白质，可以明确区分FGXD<sub>8</sub>探针中的20种天然氨基酸（图8c）。在这些结果的激励下，我们将这种方法用于检测游离氨基酸。将游离氨基酸的氨基与FGGCD<sub>8</sub>上的巯基连接，从而形成FGGC(X)D<sub>8</sub>探针，这种修饰几

乎可以将所有天然氨基酸高效地连接到探针上，通过使用纳米孔突变体和改变一些实验条件，我们成功地实现了对20种游离氨基酸的显著区分。鉴定游离氨基酸是蛋白质测序的第一步。难点在于获取氨基酸序列信息。我们的策略是使用FGGC(X)D<sub>8</sub>探针鉴定蛋白酶水解的氨基酸。由于羧肽酶A和B在水解多肽时，酶水解速度过快而导致无法得到单个氨基酸，我们采用了另一种思路，利用FGGC(X)D<sub>8</sub>探针的相对丰度来推断对应氨基酸在肽链中的位置，并验证了其相关性。利用这一方法实现了短肽的精准测序，然而对于超过8~10氨基酸的多肽，就可能出现信号紊乱。为了解决这一问题，我们采用了分步酶切法，利用羧肽酶A和B的特性通过引入羧肽酶A以确定肽链中精氨酸（R）之前的序列，随后加入羧肽酶B特异性切割R残基，通过重复这些循环大大提高了对更长多肽精确测序的能力。

尽管这几项工作都展现出了巧妙的构思用于蛋白质测序，但是它们与真正的单分子蛋白质从头测序还有一段距离，因为它们测试的都是纯化后的多肽。对于Huang课题组和Geng课题组的策略来说，后续的难点首先在于如何提高对游离氨基酸的捕获，因为在他们的体系中加入的氨基酸浓度高达 $\mu\text{mol/L}$ 甚至 $\text{mmol/L}$ 级，这在实际应用中很难达到，其次就是如何平衡20种氨基酸不同的捕获率，从而使信号的产生与多肽有序列相关性。对于我们小组的酶切-偶联-测序的策略来说，尽管使用D<sub>8</sub>牵引极大程度上平衡了氨基酸的不均一的捕获率，但是探针的分辨率还有一定的局限，同时探针的偶联也要耗费较长的时间，除此之外，对于蛋白酶的选择和设计使之能够持续稳定且以合适速率逐一水解氨基酸也是实现蛋白质测序必须克服的一大难题。

### 2.3.3 纳米孔蛋白质指纹测序

除去上述两种方法，纳米孔指纹测序的方法近年来也吸引了研究人员的注意力。纳米孔指纹测序的方法受启发于自下而上的质谱测序方法，与质谱方法相似，这种方法首先用蛋白酶将待测蛋白质水解为短肽，然后对产生的短肽进行纳米孔测试，这些短肽穿过纳米孔时会产生一系列阻滞深度不同的电流信号，将这些电流信号统计并绘制成阻滞深度-密度频率的图谱，并对照已有的数据库即可得到蛋白质的序列信息。Maglia小组<sup>[85]</sup>在2021年的实验结果展现出了类似质谱图谱的纳米孔阻滞电流信号肽图谱。他们使用工程化的FraC纳米孔

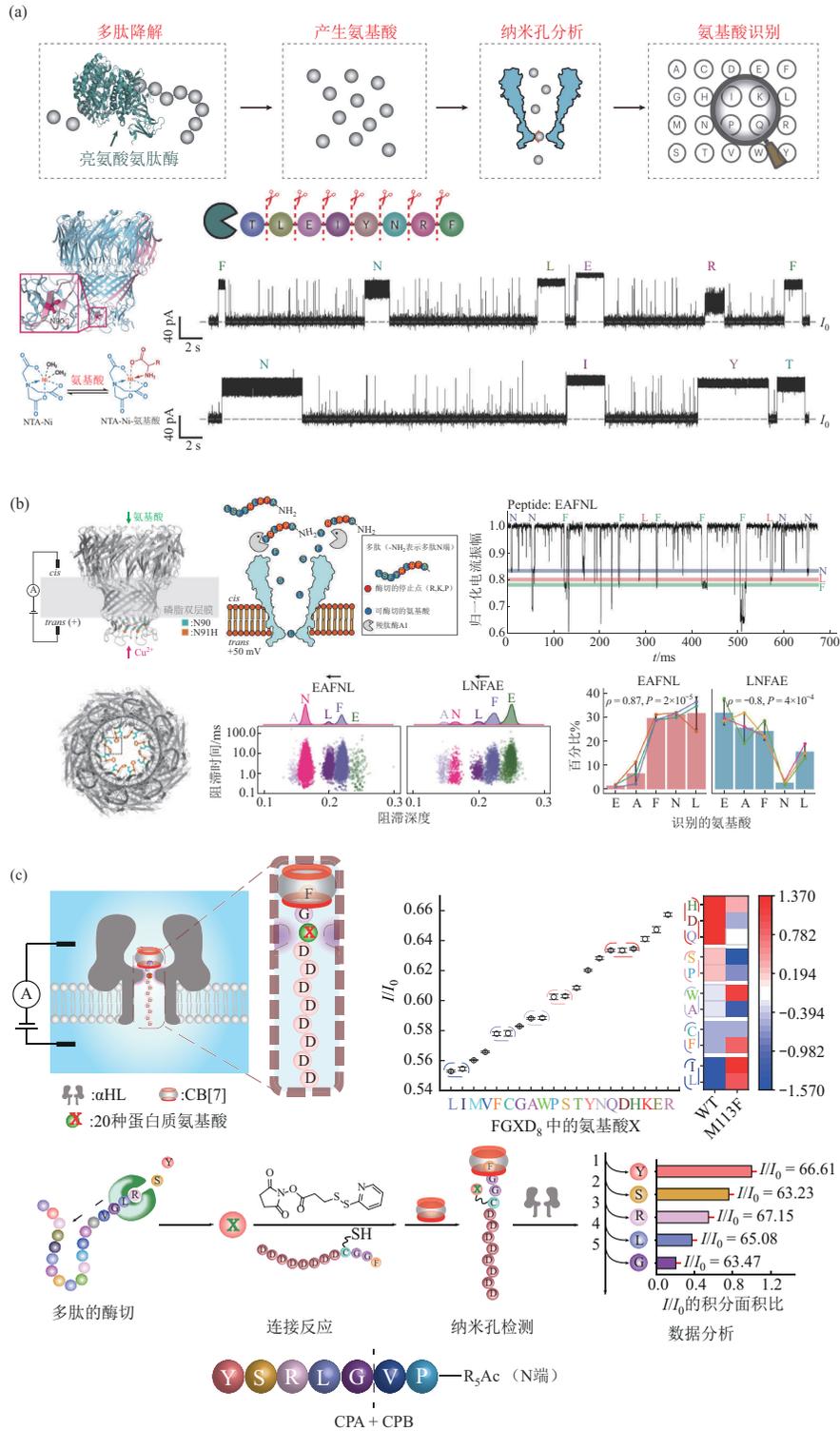


Fig. 8 Enzyme-assisted nanopore protein sequencing

图8 酶切辅助的纳米孔蛋白质测序

(a) MspA N90C突变纳米孔用于纳米孔蛋白质测序。上：亮氨酸氨肽酶水解多肽后用纳米孔识别的流程图。左下：MspA N90C突变纳米孔结合NTA与Ni<sup>2+</sup>并络合氨基酸示意图。右下：亮氨酸氨肽酶水解多肽后产生的氨基酸在纳米孔中的信号图。经授权后改编自文献 [78]。

(b) MspA N91H突变纳米孔用于纳米孔蛋白质测序。左：纳米孔检测装置示意图与纳米孔俯视图。右上：羧肽酶A水解多肽后用MspA N91H识别的示意图与电流信号图。右下：多肽EAFNL和LNFAE水解后的信号散点图与两种肽中已识别氨基酸的平均丰度。经授权后改编自文献 [80]。

(c) 基于酶切与主客体相互作用辅助的多肽测序。上：利用FGXD<sub>8</sub>对20种氨基酸进行识别，X代表待测20种天然氨基酸的一种。下：使用探针肽 FGGCD<sub>8</sub>和混合酶消化策略进行肽测序的过程示意图。模型肽用羧肽酶A和B的混合物消化。经授权后改编自文献 [81]。

G13F-FraC检测了分子质量在12.4~66.5 ku范围内的10种蛋白质，通过使用胰蛋白酶，这些蛋白质被水解为分子质量为500~2 000 u的多肽，这些多肽在纳米孔中输运时，通过统计阻滞电流信号的百分比及其概率密度关系图作为蛋白质纳米孔肽图谱(图9a)，这种方法获得的肽图谱与质谱方法高度一致且具有极高的测试准确度。无独有偶，Oukhaled小组<sup>[86]</sup>用气溶素纳米孔也绘制出了细胞色素C、溶菌酶以及肌红蛋白的纳米孔指纹图谱，并对其进行了精准检测(图9b)。Maglia及其合作者<sup>[87]</sup>后续拓展了这一方法在定量检测蛋白质糖基

化的应用(图9c)，再次证明了纳米孔指纹测序方法的巨大潜力。

纳米孔指纹测序目前还处于发展阶段，它具有与质谱测序相似的原理，虽然目前它还远比不上质谱的高分辨率并缺乏成熟的前置分离装置，但它具有费用低、可便携、检测快速和检测单分子水平的独特优势，纳米孔指纹图谱显示的不仅仅是单纯的多肽体积信息，它还包含了分析物与孔的相互作用，在未来数据库和方法足够完善时，可能用于揭示一些质谱方法难于获得的信息。

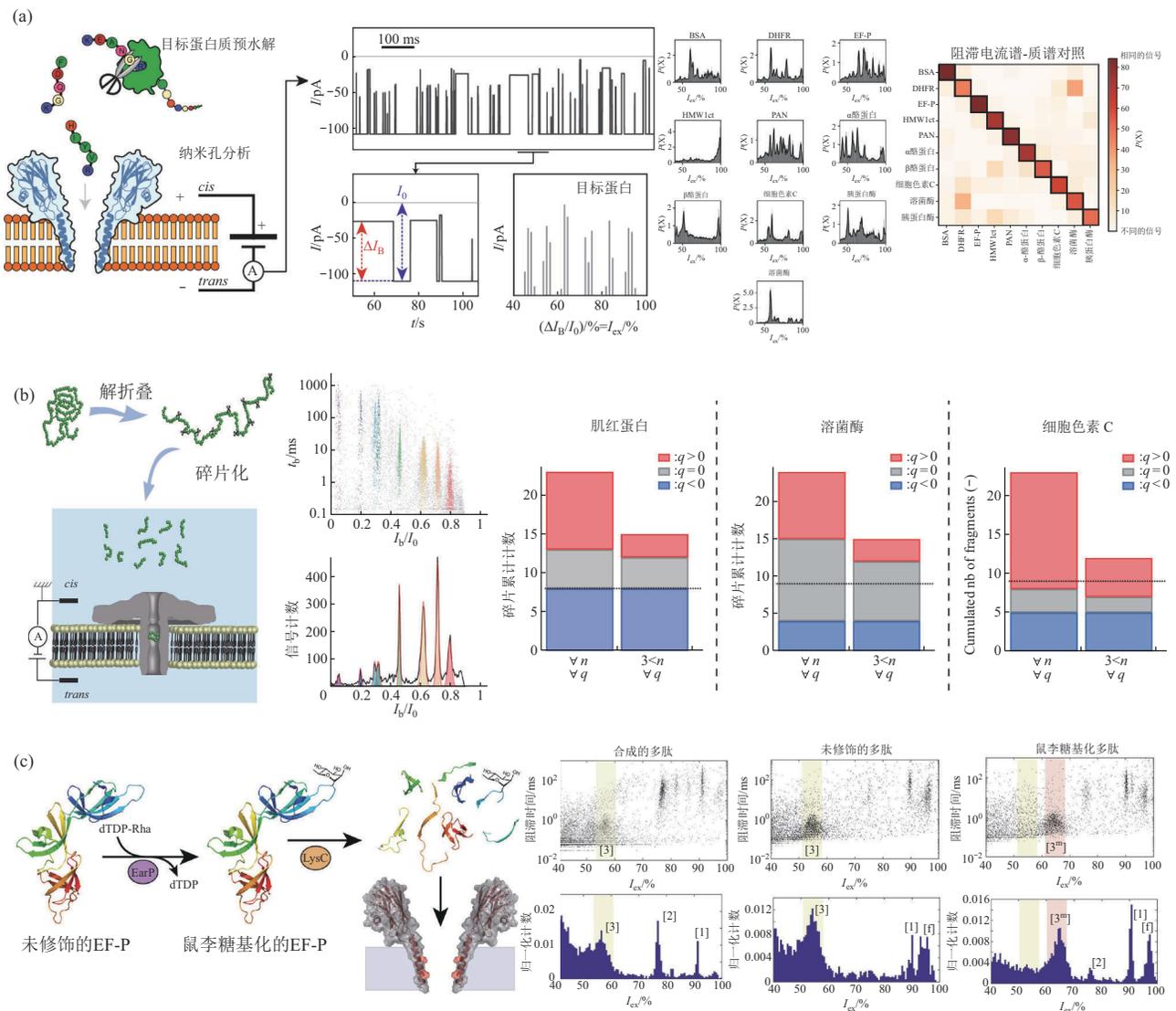


Fig. 9 Fingerprinting methods for nanopore protein sequencing

图9 纳米孔蛋白质指纹测序

(a) 利用FraC纳米孔对胰蛋白酶水解后的10种蛋白质的指纹鉴定。经授权后改编自文献 [85]。(b) 利用AeL纳米孔对胰蛋白酶水解后的3种蛋白质的指纹鉴定。经授权后改编自文献 [86]。(c) 利用FraC纳米孔对胰蛋白酶水解后的蛋白质及其糖基化蛋白的指纹鉴定。经授权后改编自文献 [87]。

### 3 总结与展望

近年来, 纳米孔单分子蛋白质测序技术取得了显著进展。分子马达驱动多肽穿过纳米孔的初步实现, 以及20种氨基酸的精准识别, 标志着这一领域的重大突破。通过纳米孔方法获取短肽序列的尝试也已获得成功, 越来越多的新方法不断涌现。然而这项技术依旧面临巨大的挑战: 对于全长链测序的策略, 读长较短, 信号解析困难依旧是必须突破的难关; 对于酶切辅助测序的方法, 在精确识别20种氨基酸的同时, 如何保证其捕获率以提高装置的灵敏度是要克服的难题, 酶水解速率的可控性以及稳定性也亟需解决。对于纳米孔指纹测序的方案, 对于前期分离装置的优化以及数据库的完善也需要投入大量精力。

在未来的发展中, 跨学科的合作和多学科的融合具有解决这些难题的潜力: 人工改造的蛋白酶将取代DNA解旋酶作为新的马达分子, 并构建复合的纳米孔-蛋白酶体系从而实现逐步对蛋白质链的从头测序; 结合先进的数据分析算法和机器学习技术, 将有效应对蛋白质测序时产生的海量数据分析挑战; 基因工程技术将带来全新的、识别能力更强的纳米孔以保证测量的准确度, 通过基因工程改造的蛋白酶将具有更稳定可控的水解性能, 从而扩展测序的长度; 集成的多维度纳米孔阵列芯片将确保该方法的高通量和高灵敏度识别能力, 更精妙的探针设计将显著缩短测序时间; 前端的样品制备流程的加入将更大限度发挥纳米孔技术单分子检测的优势; 此外, 将纳米孔技术与质谱、荧光、拉曼光谱和原子力显微镜等其他技术结合, 可多维度解析蛋白质组学的奥秘。

纳米孔蛋白质测序技术的应用不仅限于基础研究, 在临床领域也具有巨大潜力。该技术通过快速、精准地识别与疾病相关的生物标志物, 可以实现疾病的早期检测, 从而有望彻底改变诊断疗法。它的实时分析能力可以快速检测出特定的生物标志物, 这对癌症等疾病至关重要, 因为早期干预可以显著提高存活率。通过提供详细的蛋白质图谱, 纳米孔蛋白质测序技术还有助于制定个性化的治疗方案, 从而根据个人病情的分子特征进行治疗, 可以显著提高治疗效果, 改善患者预后; 此外, 这项技术的精准性还能确保即使是低丰度的蛋白质也能被检测到, 而低丰度蛋白质对于诊断罕见疾病和早期疾病具有重大意义。其多功能性也适用于多种疾

病, 如癌症、传染病、遗传性疾病和神经退行性疾病等。随着小型便携化装置的普及, 纳米孔测序的现场诊断和低成本的使用性能使其有望成为临床环境中一种便捷高效的工具。

尽管前进的道路充满挑战, 纳米孔蛋白质测序展示了光明的未来。它有潜力揭示蛋白质编码的完整图景, 为理解生命的分子机制提供前所未有的视角, 并引领生物医学新一轮技术突破。

### 参 考 文 献

- [1] Lander E S, Linton L M, Birren B, *et al.* Initial sequencing and analysis of the human genome. *Nature*, 2001, **409**(6822): 860-921
- [2] Venter J C, Adams M D, Myers E W, *et al.* The sequence of the human genome. *Science*, 2001, **291**(5507): 1304-1351
- [3] Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol*, 2007, **8**(12): 995-1005
- [4] Thomas P J, Qu B H, Pedersen P L. Defective protein folding as a basis of human disease. *Trends Biochem Sci*, 1995, **20**(11): 456-459
- [5] Valastyan J S, Lindquist S. Mechanisms of protein-folding diseases at a glance. *Dis Model Mech*, 2014, **7**(1): 9-14
- [6] Li Y, Liu X, Yu L, *et al.* Covalent LYTAC enabled by DNA aptamers for immune checkpoint degradation therapy. *J Am Chem Soc*, 2023, **145**(45): 24506-24521
- [7] Edman P. A method for the determination of amino acid sequence in peptides. *Arch Biochem*, 1949, **22**(3): 475
- [8] Brocchieri L, Karlin S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res*, 2005, **33**(10): 3390-3400
- [9] Steinke L, Cook R G. Identification of phosphorylation sites by edman degradation//Smith B J. *Protein Sequencing Protocols*. Totowa, NJ: Humana Press, 2003: 301-307
- [10] Ross F E, Zamborelli T, Herman A C, *et al.* Detection of acetylated lysine residues using sequencing by edman degradation and mass spectrometry//Marshak D R. *Techniques in Protein Chemistry*. Amsterdam: Elsevier, 1996: 201-208
- [11] Resing K A, Ahn N G. Proteomics strategies for protein identification. *FEBS Lett*, 2005, **579**(4): 885-889
- [12] Callahan N, Tullman J, Kelman Z, *et al.* Strategies for development of a next-generation protein sequencing platform. *Trends Biochem Sci*, 2020, **45**(1): 76-89
- [13] Nickerson J L, Baghalabadi V, Rajendran S R C K, *et al.* Recent advances in top-down proteome sample processing ahead of MS analysis. *Mass Spectrom Rev*, 2023, **42**(2): 457-495
- [14] Zubarev R A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics*, 2013, **13**(5): 723-726
- [15] Ponomarenko E A, Poverennaya E V, Ilgisonis E V, *et al.* The size of the human proteome: the width and depth. *Int J Anal Chem*,

- 2016, **2016**: 7436849
- [16] Huang B, Wu H, Bhaya D, *et al.* Counting low-copy number proteins in a single cell. *Science*, 2007, **315**(5808): 81-84
- [17] Savaryn J P, Catherman A D, Thomas P M, *et al.* The emergence of top-down proteomics in clinical research. *Genome Med*, 2013, **5**(6): 53
- [18] Valaskovic G A, Kelleher N L, McLafferty F W. Attomole protein characterization by capillary electrophoresis-mass spectrometry. *Science*, 1996, **273**(5279): 1199-1202
- [19] Restrepo-Pérez L, Joo C, Dekker C. Paving the way to single-molecule protein sequencing. *Nat Nanotechnol*, 2018, **13**(9): 786-796
- [20] Angel T E, Aryal U K, Hengel S M, *et al.* Mass spectrometry-based proteomics: existing capabilities and future directions. *Chem Soc Rev*, 2012, **41**(10): 3912-3928
- [21] Griss J, Perez-Riverol Y, Lewis S, *et al.* Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. *Nat Methods*, 2016, **13**(8): 651-656
- [22] Timp W, Timp G. Beyond mass spectrometry, the next step in proteomics. *Sci Adv*, 2020, **6**(2): eaax8978
- [23] Zhao Y, Ashcroft B, Zhang P, *et al.* Single-molecule spectroscopy of amino acids and peptides by recognition tunnelling. *Nat Nanotechnol*, 2014, **9**(6): 466-473
- [24] Ohshiro T, Tsutsui M, Yokota K, *et al.* Detection of post-translational modifications in single peptides using electron tunnelling currents. *Nat Nanotechnol*, 2014, **9**(10): 835-840
- [25] Liu Z, Li X, Masai H, *et al.* A single-molecule electrical approach for amino acid detection and chirality recognition. *Sci Adv*, 2021, **7**(10): eabe4365
- [26] Swaminathan J, Boulgakov A A, Marcotte E M. A theoretical justification for single molecule peptide sequencing. *PLoS Comput Biol*, 2015, **11**(2): e1004080
- [27] Swaminathan J, Boulgakov A A, Hernandez E T, *et al.* Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat Biotechnol*, 2018, **36**: 1076-1082
- [28] Yao Y, Docter M, van Ginkel J, *et al.* Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys Biol*, 2015, **12**(5): 055003
- [29] van Ginkel J, Filius M, Szczepaniak M, *et al.* Single-molecule peptide fingerprinting. *Proc Natl Acad Sci USA*, 2018, **115**(13): 3338-3343
- [30] Filius M, Wee R V, de Lannoy C, *et al.* Full-length single-molecule protein fingerprinting. *Nat Nanotechnol*, 2024, **19**(5): 652-659
- [31] Reed B D, Meyer M J, Abramzon V, *et al.* Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science*, 2022, **378**(6616): 186-192
- [32] Gu L Q, Braha O, Conlan S, *et al.* Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature*, 1999, **398**(6729): 686-690
- [33] Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol*, 2016, **34**(5): 518-524
- [34] Walker B, Kasianowicz J, Krishnaswamy M, *et al.* A pore-forming protein with a metal-actuated switch. *Protein Eng*, 1994, **7**(5): 655-662
- [35] Kasianowicz J J, Brandin E, Branton D, *et al.* Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci USA*, 1996, **93**(24): 13770-13773
- [36] Cherf G M, Lieberman K R, Rashid H, *et al.* Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nat Biotechnol*, 2012, **30**(4): 344-348
- [37] Manrao E A, Derrington I M, Laszlo A H, *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol*, 2012, **30**(4): 349-353
- [38] Jain M, Olsen H E, Paten B, *et al.* The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*, 2016, **17**(1): 239
- [39] Haque F, Li J, Wu H C, *et al.* Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA. *Nano Today*, 2013, **8**(1): 56-74
- [40] Nivala J, Marks D B, Akeson M. Unfoldase-mediated protein translocation through an  $\alpha$ -hemolysin nanopore. *Nat Biotechnol*, 2013, **31**(3): 247-250
- [41] Payet L, Martinho M, Pastoriza-Gallego M, *et al.* Thermal unfolding of proteins probed at the single molecule level using nanopores. *Anal Chem*, 2012, **84**(9): 4071-4076
- [42] Freedman K J, Haq S R, Edel J B, *et al.* Single molecule unfolding and stretching of protein domains inside a solid-state nanopore by electric field. *Sci Rep*, 2013, **3**: 1638
- [43] Pastoriza-Gallego M, Gibrat G, Thiebot B, *et al.* Polyelectrolyte and unfolded protein pore entrance depends on the pore geometry. *Biochim Biophys Acta*, 2009, **1788**(6): 1377-1386
- [44] Oukhaled G, Mathé J, Biance A L, *et al.* Unfolding of proteins and long transient conformations detected by single nanopore recording. *Phys Rev Lett*, 2007, **98**(15): 158101
- [45] Schiopu I, Iftemi S, Luchian T. Nanopore investigation of the stereoselective interactions between Cu(2+) and D, L-histidine amino acids engineered into an amyloidic fragment analogue. *Langmuir*, 2015, **31**(1): 387-396
- [46] Mereuta L, Asandei A, Seo C H, *et al.* Quantitative understanding of pH- and salt-mediated conformational folding of histidine-containing,  $\beta$ -hairpin-like peptides, through single-molecule probing with protein nanopores. *ACS Appl Mater Interfaces*, 2014, **6**(15): 13242-13256
- [47] Rodriguez-Larrea D, Bayley H. Multistep protein unfolding during nanopore translocation. *Nat Nanotechnol*, 2013, **8**(4): 288-295
- [48] Biswas S, Song W, Borges C, *et al.* Click addition of a DNA thread to the N-termini of peptides for their translocation through solid-state nanopores. *ACS Nano*, 2015, **9**(10): 9652-9664
- [49] Asandei A, Rossini A E, Chinappi M, *et al.* Protein nanopore-based discrimination between selected neutral amino acids from polypeptides. *Langmuir*, 2017, **33**(50): 14451-14459
- [50] Asandei A, Dragomir I S, di Muccio G, *et al.* Single-molecule dynamics and discrimination between hydrophilic and

- hydrophobic amino acids in peptides, through controllable, stepwise translocation across nanopores. *Polymers*, 2018, **10**(8): 885
- [51] Yu L, Kang X, Li F, *et al.* Unidirectional single-file transport of full-length proteins through a nanopore. *Nat Biotechnol*, 2023, **41**(8): 1130-1139
- [52] Asandei A, Schioppa I, Chinappi M, *et al.* Electroosmotic trap against the electrophoretic force near a protein nanopore reveals peptide dynamics during capture and translocation. *ACS Appl Mater Interfaces*, 2016, **8**(20): 13166-13179
- [53] Huang G, Willems K, Soskine M, *et al.* Electro-osmotic capture and ionic discrimination of peptide and protein biomarkers with FraC nanopores. *Nat Commun*, 2017, **8**(1): 935
- [54] Sauciu A, Morozzo della Rocca B, Tadema M J, *et al.* Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force. *Nat Biotechnol*, 2024, **42**: 1275-1281
- [55] di Muccio G, Rossini A E, di Marino D, *et al.* Insights into protein sequencing with an  $\alpha$ -hemolysin nanopore by atomistic simulations. *Sci Rep*, 2019, **9**(1): 6440
- [56] Yuan B, Li S, Ying Y L, *et al.* The analysis of single cysteine molecules with an aerolysin nanopore. *Analyst*, 2020, **145**(4): 1179-1183
- [57] Peng W, Yan S, Zhou K, *et al.* High-resolution discrimination of homologous and isomeric proteinogenic amino acids in nanopore sensors with ultrashort single-walled carbon nanotubes. *Nat Commun*, 2023, **14**(1): 2662
- [58] Wang F, Zhao C, Zhao P, *et al.* MoS<sub>2</sub> nanopore identifies single amino acids with sub-1 Dalton resolution. *Nat Commun*, 2023, **14**(1): 2895
- [59] Piguet F, Ouldali H, Pastoriza-Gallego M, *et al.* Identification of single amino acid differences in uniformly charged homopolymeric peptides with aerolysin nanopore. *Nat Commun*, 2018, **9**(1): 966
- [60] Lu Y, Wu X Y, Ying Y L, *et al.* Simultaneous single-molecule discrimination of cysteine and homocysteine with a protein nanopore. *Chem Commun*, 2019, **55**(63): 9311-9314
- [61] Restrepo-Pérez L, Huang G, Bohländer P R, *et al.* Resolving chemical modifications to a single amino acid within a peptide using a biological nanopore. *ACS Nano*, 2019, **13**(12): 13668-13676
- [62] Ouldali H, Sarthak K, Ensslen T, *et al.* Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat Biotechnol*, 2020, **38**(2): 176-181
- [63] Wei X, Ma D, Zhang Z, *et al.* N-terminal derivatization-assisted identification of individual amino acids using a biological nanopore sensor. *ACS Sens*, 2020, **5**(6): 1707-1716
- [64] Wei X, Ma D, Jing L, *et al.* Enabling nanopore technology for sensing individual amino acids by a derivatization strategy. *J Mater Chem B*, 2020, **8**(31): 6792-6797
- [65] Wei X, Ma D, Ou J, *et al.* Narrowing signal distribution by adamantane derivatization for amino acid identification using an  $\alpha$ -hemolysin nanopore. *Nano Lett*, 2024, **24**(5): 1494-1501
- [66] Rosen C B, Rodriguez-Larrea D, Bayley H. Single-molecule site-specific detection of protein phosphorylation with a nanopore. *Nat Biotechnol*, 2014, **32**(2): 179-181
- [67] Restrepo-Pérez L, Wong C H, Maglia G, *et al.* Label-free detection of post-translational modifications with a nanopore. *Nano Lett*, 2019, **19**(11): 7957-7964
- [68] Li S, Wu X Y, Li M Y, *et al.* T232K/K238Q aerolysin nanopore for mapping adjacent phosphorylation sites of a single tau peptide. *Small Meth*, 2020, **4**(11): 2000014
- [69] Huo M Z, Hu Z L, Ying Y L, *et al.* Enhanced identification of Tau acetylation and phosphorylation with an engineered aerolysin nanopore. *Proteomics*, 2022, **22**(5/6): e2100041
- [70] Ensslen T, Sarthak K, Aksimentiev A, *et al.* Resolving isomeric posttranslational modifications using a biological nanopore as a sensor of molecular shape. *J Am Chem Soc*, 2022, **144**(35): 16060-16068
- [71] Wloka C, Van Meervelt V, van Gelder D, *et al.* Label-free and real-time detection of protein ubiquitination with a biological nanopore. *ACS Nano*, 2017, **11**(5): 4387-4394
- [72] Xiong Y, Li M, Cao Y, *et al.* Nanofluidic device for detection of lysine methylpeptides and sensing of lysine methylation. *Anal Chem*, 2023, **95**(19): 7761-7769
- [73] Cao C, Magalhães P, Krapp L F, *et al.* Deep learning-assisted single-molecule detection of protein post-translational modifications with a biological nanopore. *ACS Nano*, 2024, **18**(2): 1504-1515
- [74] Yan S, Zhang J, Wang Y, *et al.* Single molecule ratcheting motion of peptides in a *Mycobacterium smegmatis* porin A (MspA) nanopore. *Nano Lett*, 2021, **21**(15): 6703-6710
- [75] Chen Z, Wang Z, Xu Y, *et al.* Controlled movement of ssDNA conjugated peptide through *Mycobacterium smegmatis* porin A (MspA) nanopore by a helicase motor for peptide sequencing application. *Chem Sci*, 2021, **12**(47): 15750-15756
- [76] Brinkerhoff H, Kang A S W, Liu J, *et al.* Multiple rereads of single proteins at single-amino acid resolution using nanopores. *Science*, 2021, **374**(6574): 1509-1513
- [77] Nova I C, Ritmejeris J, Brinkerhoff H, *et al.* Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat Biotechnol*, 2024, **42**(5): 710-714
- [78] Wang K, Zhang S, Zhou X, *et al.* Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat Methods*, 2024, **21**(1): 92-101
- [79] Boersma A J, Bayley H. Continuous stochastic detection of amino acid enantiomers with a protein nanopore. *Angew Chem Int Ed*, 2012, **51**(38): 9606-9609
- [80] Zhang M, Tang C, Wang Z, *et al.* Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nat Methods*, 2024, **21**(4): 609-618
- [81] Zhang Y, Yi Y, Li Z, *et al.* Peptide sequencing based on host-guest interaction-assisted nanopore sensing. *Nat Methods*, 2024, **21**(1): 102-109

- [82] Sheng Y, Zhou K, Liu L, *et al.* A nanopore sensing assay resolves cascade reactions in a multienzyme system. *Angew Chem Int Ed*, 2022, **61**(20): e202200866
- [83] Liu L, You Y, Zhou K, *et al.* A dual-response DNA probe for simultaneously monitoring enzymatic activity and environmental pH using a nanopore. *Angew Chem Int Ed*, 2019, **58**(42): 14929-14934
- [84] Liu L, Li T, Zhang S, *et al.* Simultaneous quantification of multiple cancer biomarkers in blood samples through DNA-assisted nanopore sensing. *Angew Chem Int Ed*, 2018, **57**(37): 11882-11887
- [85] Lucas F L R, Versloot R C A, Yakovlieva L, *et al.* Protein identification by nanopore peptide profiling. *Nat Commun*, 2021, **12**(1): 5795
- [86] Afshar Bakshloo M, Kasianowicz J J, Pastoriza-Gallego M, *et al.* Nanopore-based protein identification. *J Am Chem Soc*, 2022, **144**(6): 2716-2725
- [87] Versloot R C A, Lucas F L R, Yakovlieva L, *et al.* Quantification of protein glycosylation using nanopores. *Nano Lett*, 2022, **22**(13): 5357-5364

## Nanopore Protein Sequencing\*

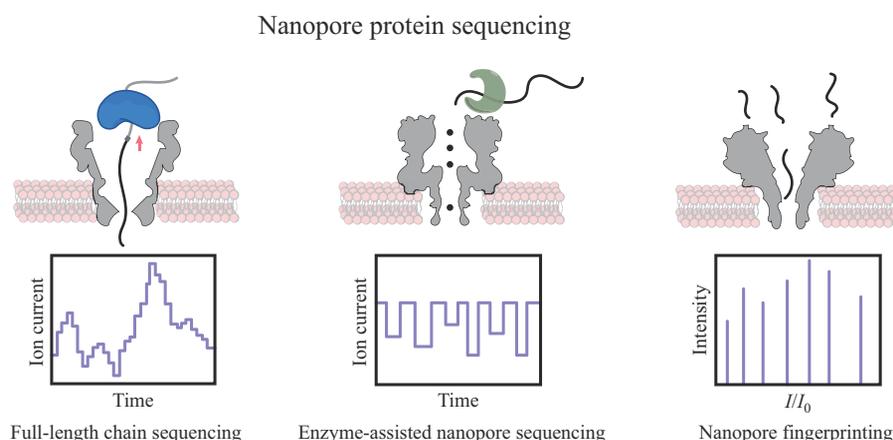
LI Zi-Yi<sup>1,2)</sup>, YI Ya-Kun<sup>1,2)</sup>, LIU Lei<sup>3)</sup>, WU Hai-Chen<sup>1,2)\*\*</sup>

<sup>1)</sup>Beijing National Laboratory for Molecular Sciences, Key Laboratory of Analytical Chemistry for Living Biosystems, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China;

<sup>2)</sup>School of Chemical Science, University of Chinese Academy of Sciences, Beijing 100049, China;

<sup>3)</sup>Key Laboratory for Biomedical Effects of Nanomaterials & Nanosafety, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China)

### Graphical abstract



**Abstract** The success of the Human Genome Project has significantly deepened our understanding of genomics and catalyzed a growing focus on proteomics, as researchers aim to decipher the complex relationship between genes and proteins. Given the central role of proteins in regulating physiological processes—including DNA replication, metabolic reactions, signal transduction, pH balance, and cellular structure—developing advanced protein sequencing technologies is critical. Proteins are fundamental to nearly all biological activities, making their detailed study essential for understanding cellular functions and disease mechanisms. The Edman degradation method, developed in the 1950s, was a breakthrough in sequencing short peptides. However, its limitations in read length (fewer than 50 amino acids) and slow cycle time fall short of modern demands. Mass spectrometry has since emerged as the gold standard in protein sequencing due to its high accuracy, throughput, and reproducibility. The method is enhanced by a robust sample preparation workflow and advances in mass spectrometry technology. Despite these strengths, mass spectrometry faces limitations in dynamic range, sensitivity, read length, and sequence coverage, hindering complete *de novo* protein sequencing. These technological gaps underscore the need for innovative methods to provide more detailed and accurate protein

\* This work was supported by a grant from The National Natural Science Foundation of China (22025407).

\*\* Corresponding author.

Tel: 86-10-82614948, E-mail: haichenwu@iccas.ac.cn

Received: July 2, 2024 Accepted: August 24, 2024

sequence data. In the past decade, new protein sequencing methods, including tunneling current, fluorescence fingerprinting, and real-time dynamic fluorescence, have shown significant developmental potential. However, these methods are not yet ready for widespread application, as each still faces technical hurdles. Meanwhile, advances in nanopore DNA sequencing have sparked interest in applying nanopore technology to protein sequencing, particularly owing to its speed, convenience, and cost-effectiveness. Unlike DNA sequencing, protein sequencing presents greater challenges due to proteins' complex three-dimensional structures, heterogeneous electrical charges, difficulties in directional movement, and diverse amino acid compositions, further complicated by post-translational modifications. Researchers have made significant strides in addressing these challenges, such as using unfolding enzymes, high temperatures, high voltage, and deformers to unravel protein structures, and employing charged sequences and electroosmotic flow to control peptide translocation. The latest strategies for nanopore protein sequencing can be broadly categorized into three approaches: strand sequencing, enzyme-assisted nanopore sequencing, and nanopore fingerprinting. In strand sequencing, dragging a protein-oligonucleotide conjugate through a nanopore with the aid of protein motors generates stepped current signals produced by the peptide strand. In enzyme-assisted nanopore sequencing, 20 proteinogenic amino acids and various post-translational modifications have been distinguished using nanopores, and sequencing of short peptides has also been demonstrated. In nanopore fingerprinting, polypeptide fragments resulting from protease digestion of a protein can be identified through nanopore sensing. Despite these advances, further improvements in protein engineering, data processing, identification accuracy, and read length are needed to make these strategies practically useful. This review provides an overview of the current major approaches to nanopore protein sequencing, emphasizing the strategies, recent advances, breakthroughs and challenges in nanopore protein sequencing. As nanopore technology continues to evolve, it is expected to offer more efficient and accurate sequencing solutions in proteomics, potentially leading to new technological breakthroughs in biochemistry and biomedicine.

**Key words** protein sequencing, nanopore, stochastic sensing, single molecule analysis

**DOI:** 10.16476/j.pibb.2024.0295