

## 基于生物质谱的蛋白质 C 末端研究进展

刘恣博<sup>1,2</sup>, 陆豪杰<sup>1,2</sup>

1 复旦大学化学系, 上海 200433

2 复旦大学生物医学研究院, 上海 200032

刘恣博, 陆豪杰. 基于生物质谱的蛋白质 C 末端研究进展. 生物工程学报, 2014, 30(7): 1083–1093.

Liu MB, Lu HJ. C-terminal proteomics: strategies for characterization of protein C-terminus using MS-based techniques. Chin J Biotech, 2014, 30(7): 1083–1093.

**摘要:** 蛋白质的 C 末端在蛋白质进行各项生命活动过程中都起着极其重要的作用。它不仅标志着 DNA 转录翻译成蛋白质过程的初步完成, 更是参与和调控了蛋白质的各种生理功能。研究蛋白质的 C 末端不仅有利于完整蛋白质的鉴定, 对于在分子水平理解蛋白质的信号传导和生化功能是十分必要的。文中结合我们的研究工作, 综述了近年来基于生物质谱的蛋白质 C 末端研究的相关进展, 包括了 C 末端的识别、鉴定以及蛋白质 C 末端肽段富集的新方法和新技术。

**关键词:** 蛋白质组学, 蛋白质 C 末端, 生物质谱, 富集

## C-terminal proteomics: strategies for characterization of protein C-terminus using MS-based techniques

Minbo Liu<sup>1,2</sup>, and Haojie Lu<sup>1,2</sup>

1 Department of Chemistry, Fudan University, Shanghai 200433, China

2 Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China

**Abstract:** C-termini of proteins often play an important role in various biological processes, such as the transcription and translation from DNA to protein and also participating in various biological regulations. The determination of protein C-terminus is so crucial because it provides not only distinct functional annotation, but also a way to monitor the proteolysis-modified proteins. Based on the biological mass spectrometry, a series of novel methods and technologies were developed both for qualitative and quantitative analyses of protein C-terminus. These methods or technologies can be applied to accurate and effective protein C-terminus profiling, including the sequences and quantitative information of C-termini, which reveals the biological function of C-termini in life's activities and provides a better understanding of the

**Received:** January 17, 2014; **Accepted:** April 9, 2014

**Supported by:** National Basic Research Program of China (973 Program) (No. 2012CB910602), National Science Foundation of China (No. 21025519).

**Corresponding author:** Haojie Lu. Tel: +86-21-54237416; Fax: +86-21-54237486; E-mail: luhaojie@fudan.edu.cn

国家重点基础研究发展计划 (973 计划) (No. 2012CB910602), 国家自然科学基金 (No. 21025519) 资助。

degradation of mature proteins. Combined with our research, this review highlights the improvements in C-terminal proteomics study in the past decades, including the methodologies for recognition and identification of C-terminus, as well as the enrichment strategies for protein C-terminus.

**Keywords:** proteomics, protein C-terminus, biological mass spectrometry, enrichment

DNA 中的基因通过转录和翻译合成蛋白质, 随后通过进一步的加工和翻译后修饰形成具有各种构象和生物功能的蛋白质。当 mRNA 由基因转录得到后, 在核糖体上被用作蛋白质合成的模板, 从而进行蛋白质的翻译合成。在生物体中的蛋白质合成总是从 N 端到 C 端。前者意味着合成的开始, 而后者意味着合成的终止。所以一个成熟蛋白质的 C 端在蛋白质的合成中处于相当重要的地位, 并且在该蛋白质参与生物体各项生理活动中更是起了积极的作用。由于蛋白质的天然构象决定了蛋白质的生物功能, 而其天然构象又是由该蛋白质的一级结构决定的。所以如果要了解一个蛋白质的高级结构及其重要的蛋白质功能, 那必须首先要明确其一级结构, 即蛋白质的氨基酸序列。而 C 末端作为蛋白质合成的终止端, 它序列的确定则显得尤为重要。随着蛋白质组学的不断发展, 人们在蛋白质水平对于蛋白质 C 端的研究越来越深入, 蛋白质 C 端测序工作也逐渐发展起来, 随着一些新蛋白质 C 末端测序方法的建立, 测序的灵敏度和分析通量都得到提高, 而且可以实现 *De novo* 测序。特别是生物质谱技术应用于 C 端测序后, 使得大规模高通量的 C 端测序成为可能, 并能够在蛋白质组水平应用。本文旨在阐述研究蛋白质 C 末端的重要意义, 并系统地综述了以质谱测序法为主的蛋白质 C 末端的鉴定策略以及与质谱测序法相关的 C 末端肽段的特异性富集方法。

## 1 蛋白质 C 末端的研究意义

蛋白质的 C 末端在蛋白质参与生命体各项生理活动中起着非常重要的作用。首先, 蛋白质的 C 末端包含了丰富的与生物功能相关的信息, 对于蛋白质定位、蛋白质识别和信号传导等方面发挥重要的作用<sup>[1-3]</sup>。与蛋白质降解密切相关的泛素化途径中, 泛素就是与非特异性泛素激活酶 E1 的半胱氨酸残基共价结合, 形成 E1-泛素复合物, 并经过一系列的生化反应, 最后将泛素 C 末端甘氨酸上的羧基连接到靶蛋白赖氨酸残基的  $\epsilon$ -氨基基团上; 又例如人源的  $\beta$ -淀粉样蛋白 (Amyloid  $\beta$ -protein) C 末端发生调节紊乱是导致 Alzheimer 病的主要元凶之一<sup>[3]</sup>; 萤火虫荧光素酶 C 端的 3 个氨基酸不仅构成了序列进化保守的 C 末端, 同时也作为过氧化物酶的信号靶标<sup>[4]</sup>。第二, 蛋白质 C 末端涉及到的翻译后修饰也对蛋白质的生物学功能产生重要的影响<sup>[5-6]</sup>。类似于蛋白质 N 端乙酰化的功能, 在 C 端羧基发生的酰胺化可以改变羧基所带的电荷性质, 将原先带有负电荷的羧基转变为接近中性的酰胺基团, 从而在某种程度上提高了蛋白质或者肽段的稳定性<sup>[7]</sup>。在一些分子量相对较小的多肽激素中, 经过酰胺化修饰的肽段比例非常高。又比如通过在 C 端附近的半胱氨酸异戊烯化修饰可以将蛋白质锚定在脂质膜上<sup>[8]</sup>, 以及 C 末端的糖基磷脂酰肌醇修饰也可以用于锚定蛋白质于细胞质膜上<sup>[9-10]</sup>。第

三,蛋白质的 C 末端涉及多方面的生物体酶解作用。有些蛋白质在某些酶的作用下会发生降解,如趋化因子的调控<sup>[11]</sup>、多肽激素的成熟以及纤维蛋白溶解<sup>[12]</sup>等。此外,还有一些酶在平时并不发挥功能,但只要从某一位置将保护的基团切除之后,便可以具有相应的生物活性。比如某些分泌蛋白通过一段信号肽可以从细胞内传输到细胞外,随后这一段信号肽被移除后,该蛋白便被激活<sup>[13]</sup>。这些降解过程都会产生新的蛋白质 C 末端(neo-C-termini),而研究这些新产生的蛋白质 C 端能够更好地对酶解过程进行功能化的注释,这其中包括酶和底物的识别,酶的作用位点,底物的酶解位点等<sup>[14]</sup>。

鉴定蛋白质 C 末端的氨基酸序列对于蛋白质的定性分析同样发挥着重要的作用。蛋白质的末端氨基酸序列具有很高的特异性,仅仅对末端少数几个氨基酸的序列测定就可以实现绝大多数蛋白质的可靠的分析鉴定<sup>[15]</sup>。Wilkins 研究小组通过对理论数据库中各物种的理论末端肽的特异性进行统计分析,结果表明,有 43%–83%的蛋白质(依物种而异)可以通过分析 N 端的前 4 个氨基酸序列来确定。若测定蛋白质 N 端的前 5 个氨基酸序列,则可以将鉴定成功率提高至 78%–97%<sup>[16]</sup>。相对于 N 端鉴定,C 端序列的测定对于整个完整蛋白质的鉴定有着其独特的优势:第一,作为 N 端鉴定的互补,C 端鉴定有着更高的特异性,同样是 Wilkins 小组对于理论数据库中蛋白质的 C 末端进行了统计分析,显示出仅仅检测 C 末端的前 4 个氨基酸,有 74%–97%的蛋白质能够被特异性地成功鉴定<sup>[15]</sup>;第二,在真核细胞中,有 30%–80%的成熟蛋白的 N 端发生了翻译后修饰(如 N-端乙酰化修饰<sup>[17]</sup>),N 端上的  $\alpha$ -氨基由于这些修饰而

被封闭,从而无法进行有效的 Edman 降解,所以实际上能通过 N 端测序而确定具体蛋白质的比例远远低于上述理论计算得到的比例。相比之下,绝大多数蛋白质的 C 端是自由游离状态的,仅仅可能发生为数不多的翻译后修饰,比如在一些小分子肽激素中的酰胺化<sup>[18]</sup>。这使得蛋白质 C 端的序列测定有着更广泛的应用范围和前景。此外,对蛋白质 C 末端的序列分析,在检验重组表达蛋白的纯度、确定具体的蛋白序列方面也是一个很好的补充<sup>[5]</sup>。

综上所述,对蛋白质 C 末端序列的研究不仅有利于蛋白质的鉴定,更重要的是可以分析蛋白质的高级构象,揭示蛋白质的生物学功能。由此看来,一套完善的蛋白质 C 末端分析平台和流程对于整个蛋白质 C 末端的分离分析和功能研究显得至关重要。

## 2 基于质谱技术的蛋白质 C 末端鉴定技术

早在 1926 年,Schlack 和 Kumpf 便提出了(异)硫氰酸法用于蛋白质 C 端的测定。类似于在 N 端测序中被广泛使用的 Edman 降解法,这是一种化学测序法。该方法是指蛋白质或者多肽样品与化学试剂反应后,从 C 末端逐步发生专一性裂解,对裂解得到的氨基酸进行分析,最后还原 C 末端的氨基酸序列的方法。C 端的化学测序法发展至今已有 80 余年,随着机理的深入探究以及分析手段的不断发展,目前比较具有代表性的方法主要有烷基化学法<sup>[19-20]</sup>和过氰酸酐法<sup>[21]</sup>。但是相比于 Edman 降解法,羧基端的碳原子与氨基端氮原子化学活性相差很大,使得化学法的偶联产率很低,往往只能进行 3–5 个重复单元,并且反应起始量要求较高

(20–100 pmol)。此外,这类化学法无法应用于 C 末端被修饰封闭的蛋白质或多肽,也无法作用于脯氨酸,所以并不能作为一种有效的 C 末端测序方法<sup>[22-23]</sup>。

近 30 年随着质谱技术的飞速发展,因为其出色的分辨率以及超高的灵敏度而受到了广泛地重视,特别是在蛋白质组学领域中,结合色谱和串联质谱的联用,更使其成为了一种常规的检测分析手段。相比于氨基,羧基存在一定的化学反应惰性,但是通过质谱以及相关的衍生化手段,可以达到鉴定 C 末端的目的。对于序列已知的蛋白质,可以根据序列和酶切位点计算出 C 末端肽段的理论分子量,进而在质谱分析中选择相对应分子量的肽段通过二级质谱序列分析以确证 C 端的正确性。对于未知序列的蛋白质,则可以通过一系列方法使 C 末端肽段产生特征峰,在质谱中识别并进行测序分析。主要有羧肽酶法、轻重同位素标记法、Lys-C 结合 Lys-N 酶解比较法等。这些方法都可以在一级质谱中对 C 末端肽段进行识别,从而能够选取 C 末端肽段母离子进行二级质谱的序列分析。

## 2.1 羧肽酶测序法

羧肽酶 (Carboxypeptidase, CP) 是一类催化水解蛋白质或者多肽链含羧基末端氨基酸的酶。它能够特异性地识别和作用于末端为自由  $\alpha$ -羧基的氨基酸,经过一定时间的水解,C 末端的氨基酸会逐个解离下来,通过不同的检测方式,可以检测逐个释放得到的氨基酸,来还原得到蛋白质的 C 末端序列。常用的氨基酸检测方法有两种:一种是色谱法,用于直接检测水解得到的氨基酸;一种是质谱法,通过分析比较丢失末端氨基酸的蛋白质或多肽残片与水解

前的分子量差值,从而得到氨基酸分子量信息。

由于目前生物质谱对于分子量较大的蛋白质分析灵敏度和准确度有限,所以最初结合质谱的羧肽酶测序法仅仅局限于多肽<sup>[24-26]</sup>或是小蛋白水平的 C 端分析<sup>[27-28]</sup>。例如 Patterson 等优化了羧肽酶的催化水解条件,通过改变羧肽酶的浓度,在不同的酶解时间点来采集数据,动态地得到多肽的 C 末端氨基酸序列。基于此方法,一共鉴定了 22 条多肽,最多测定了多肽 (ACTH 7-38 片段,质荷比为 3 695.15) 的 C 末端 19 个氨基酸的序列<sup>[25]</sup>。基于此,Beeumen 研究小组发展了一种羧肽酶水解用于 C 末端测序的新方法<sup>[29]</sup>。该方法结合了溴化氰化学裂解蛋白质以及羧肽酶对 C 末端特异性的降解。首先将分离纯化得到的蛋白质通过溴化氰裂解。该裂解方式作用于甲硫氨酸,最大的特点就是酶解产生的 C 末端都形成了高丝氨酸内酯的衍生物。这样使得除了自由 C 末端肽段外,其他由酶解产生的肽段都不再含有自由的  $\alpha$ -羧基。羧肽酶并不能识别 C 末端为高丝氨酸内酯的肽段,因此当溴化氰裂解产物中加入羧肽酶之后,经过一定时间酶切,C 末端肽段的氨基酸会逐个解离下来,可以通过质谱检测不同时间酶切得到的肽段,便可以得到阶梯式的序列信息,最后将所有结果还原成 C 末端肽段的序列。此方法的兼容性较好,也可以测定分子量较大的蛋白质,最大测定到分子量为 68 kDa 的蛋白,并且最多测定到 12 个氨基酸残基。但是该方法的不足在于只适用于简单体系,面对复杂体系时,必须经过蛋白质的分离处理。此外,该方法依赖于最后一个甲硫氨酸在整个蛋白序列中的位置,因为只有酶切后处于合适分子量的 C 末端肽段才适合该方法。

## 2.2 轻重同位素标记法

轻重同位素标记法是指通过酶促手段或者化学方法,能够将轻重同位素分别标记在肽段上。针对酶解后产生的肽段和自由的 C 末端肽在化学性质上存在一定的差异,可以利用同位素的方法将它们区分,从而在质谱图中识别和鉴定。

其中较为典型的方法是 Nakamura 小组基于这样的思路,发展了 50%  $^{18}\text{O}$  水解的方法用于 C 末端肽段的鉴定<sup>[30]</sup>。在酶解过程中,新产生肽段的 C 末端经过 50%  $^{16}\text{O}$  和 50%  $^{18}\text{O}$  的酶促标记,会在质谱图上形成一对分子量相差 2 Da 的对峰,而原来蛋白质的 C 末端,由于不参与酶解反应之中,所以不会被  $^{18}\text{O}$  标记,在质谱图上只会以单峰的形式存在。由此能够在一级质谱图中很容易地识别出 C 末端肽段,并选择此肽段进行二级质谱分析得到 C 端序列。基于傅里叶变换离子回旋共振高分辨质谱 (FT-ICR) 的检测,可以将灵敏度提高至 ppm 级别,与二维凝胶电泳结合后可以快速进行 C 端分析。此外,Young 小组也发展了多种同位素标记策略用于 C 端分析的新方法<sup>[31]</sup>。比如基于乙酰化反应的策略,蛋白质经过 Lys-C 酶解后,酶解混合物平均分为两份分别与乙醛及氘代乙醛反应,使得氨基发生了乙酰化修饰,随后再 1:1 等比例混合后进行质谱分析。此时由酶解得到的肽段在首尾两端分别带有  $\alpha$ -氨基和  $\epsilon$ -氨基两个乙酰化的位点,所以轻标的乙酰化肽段和重标的乙酰化肽段在分子量上有 6 Da 的差别,在质谱图中呈现差 6 Da 的同位素对峰,而蛋白质 C 末端的肽段往往只带有一个  $\alpha$ -氨基,所以在质谱图中以分子量差 3 Da 的对峰形式存在。类似的还有酯化反应策略,蛋白质经过

Glu-C 酶解后,等分成两份并分别与甲醇以及氘代甲醇发生酯化反应,随后再以 1:1 等比例混合。由酶解得到的肽段 C 末端为谷氨酸,含有一个  $\alpha$ -羧基与一个侧链羧基,所以可以和两分子的甲醇反应,同样在质谱图上呈现分子量差 6 Da 的对峰,而蛋白质 C 末端的肽段往往只带有一个  $\alpha$ -羧基,所以在质谱图中以分子量差 3 Da 的对峰形式存在。

通过上述的轻重同位素标记的方法,可以很直观地在一级质谱图中辨识出蛋白质的 C 末端肽段,而且步骤简单,可操作性高,非常适合用于单个蛋白质的 C 端鉴定。但是该类轻重同位素标记方法的不足之处在于通过轻重同位素标记,谱图的复杂程度提高了一倍,对于谱图的解析会带来许多困难。

本课题组在此基础上,发展了一种基于恶唑酮化学的同位素标记策略,实现在复杂的生物样品中同时进行蛋白质 C 端的鉴定和相对定量<sup>[32]</sup>。该方法结合了恶唑酮化学法和双标的精氨酸同位素标记。恶唑酮化学被认为是为数不多的可以有效区分  $\alpha$ -羧基与侧链羧基的反应<sup>[33]</sup>。该反应首先通过  $\alpha$ -羧基与甲酸和乙酸酐反应脱水形成类似恶唑酮的中间产物,随后又与氨基试剂反应,得到含有酰胺键的产物。基于恶唑酮反应,有一系列的功能化试剂被用于与蛋白质 C 端进行衍生反应<sup>[33-36]</sup>。通过恶唑酮反应,将精氨酸特异性地与蛋白质的  $\alpha$ -羧基结合;将蛋白质酶解之后,带有同位素标签的 C 末端肽段在质谱中呈现出对峰的峰形,从而与其他酶解肽段区分开来,通过进一步的串联质谱分析,我们可以得到其序列信息。精氨酸作为衍生化试剂,可以增强 C 末端肽段的碱性,从而大大提高了肽段在质谱中的离子化效率。此外,

双标的同位素标记还可以用于蛋白质 C 端肽段的相对定量。当标有轻重同位素标记的样品混合后进行质谱分析,同位素峰的信号强度之比即代表了来源于不同样本之间 C 端肽段的量的比例。该方法有着良好的动态线性范围,研究表明在 2 个数量级的范围内都保持一致的线性和很高的重现性。我们将此方法运用于腾冲嗜热菌的研究中,考察了在不同温度条件培养下的腾冲嗜热菌的蛋白质 C 端的表达水平的差异。一共有 68 条 C 端肽段被成功鉴定到,其中有 53 条在两个温度下的腾冲嗜热菌中都被鉴定到并有相对定量信息。在这些有表达差异的 C 端肽段中,大部分都归属于各种酶蛋白、ATP 功能相关蛋白以及核糖体蛋白,这些蛋白质的变化都有可能和外界条件(比如温度)有关。所以我们推想,温度确实是一个影响腾冲嗜热菌生物活性的重要因素。此外,还有 24 条非 C 端肽段作为可能是内源性新产生的 C 端肽段被成功鉴定。在鉴定到的 24 个内源性新 C 端肽段中,大部分都定位于完整蛋白质的 N 端或者 C 端附近,说明了蛋白质的 N 端和 C 端有着非常高的生物活性,在生命活动中起着非常重要的作用。

### 2.3 其他基于质谱鉴定的 C 末端测序方法

此外,也有其他小组报道通过不同的酶解方式或者化学衍生手段对蛋白质的 C 末端进行序列分析。大多数的策略出发点仍然基于 C 末端肽段较之其他肽段在一级质谱中呈现出各类差异性的特征,对 C 末端肽段进行识别,进而对其进行序列的分析。比如 Murphy 小组结合了溴化氰化学裂解以及在酸性条件下的甲醇衍生化反应,对于马的 Glutathione S-Transferase 蛋白进行了 C 端的序列分析<sup>[37]</sup>。非 C 末端的酶解后得到的肽段末端都以高丝氨酸内酯的形式存

在,在酸性条件下和甲醇作用,发生开环加合反应,使得此类肽段加上了 32 Da 的修饰;而唯独酶解后的 C 末端肽段在 C 端以  $\alpha$ -羧基的形式存在,与甲醇发生酯化反应,有 14 Da 的分子量增加。通过比较反应前后的质谱图,便能够在一级质谱图中辨识得到蛋白质的 C 末端肽段。

Tanaka 小组结合了 Lys-C 和 Lys-N 的酶解特点,发展了 Lys-C 和 Lys-N 酶解比较的新方法用于蛋白质 C 端的分析<sup>[38]</sup>。该方法使用 Lys-C 和 Lys-N 这两种肽段内切酶,对蛋白质分别进行酶解。Lys-C 产生的都是 C 末端为赖氨酸的肽段,而 Lys-N 产生的则都是 N 末端以赖氨酸开始的肽段。将由两种不同酶切得到的肽指纹谱图进行比较可以发现,蛋白质 N 末端肽段在 Lys-C 酶解体系中比 Lys-N 的体系中多了一个赖氨酸,相反蛋白质 C 末端肽段在 Lys-C 酶解体系中比 Lys-N 的体系中少了一个赖氨酸。而所有蛋白质中间段产生的肽段在分子量上都没有差别,只是序列上略有不同,Lys-C 体系中的赖氨酸在 C 末端,而 Lys-N 体系中的赖氨酸在 N 末端。所以该方法在简单体系中可以通过比较 Lys-C 与 Lys-N 酶解产物的质谱图,同时对蛋白质的 N 端和 C 端进行分析,也具有一定的适用性。但是该方法仍然受限于酶的酶切位点,此外也没有自动化的平台进行高通量的蛋白末端鉴定分析。

## 3 蛋白质 C 末端肽段的特异性富集策略

以上各种方法各有优势和适用的体系,针对不同的蛋白质可以选取合适的方法进行 C 末端的序列分析。但是这些方法都有一个共同的不足,那就是没有对蛋白质 C 末端的肽段进行特异性富集。这样带来的不利影响就是我们

需要从海量的数据中首先寻找到 C 末端肽段, 随后才能对其进行鉴定分析。当 C 末端肽段淹没于大量的非 C 末端肽段时, 就必须耗费大量的时间和精力去识别, 并且有极大的可能受到非目标肽段的干扰, 而造成鉴定失败或者信息丢失, 所以此前的方法都只适用于一些简单的体系, 比如二维凝胶电泳分离后的蛋白质或者是表达纯化过的蛋白质。理想的分析手段必须排除这些干扰, 并且能应用于高通量大规模的蛋白组学研究中, 所以我们需要引进蛋白质 C 末端肽段的富集策略。

### 3.1 C 末端肽段反向富集法

相比与蛋白质 N 端的富集策略, 蛋白质 C 端的富集办法不是很多。这其中有两个方面的原因: 首先是因为与氨基相比, 羧基的反应活性相对较低, 很少有方法能够与羧基反应达到较高的产率; 其次, 能够特异性地区分 C 末端  $\alpha$ -羧基与氨基酸残基上羧基的反应不多, 不像 N 端的  $\alpha$ -氨基有很多方法可以与赖氨酸残基上的  $\epsilon$ -氨基区分开来<sup>[39-40]</sup>, 所以很难有效地进行选择性的反应。目前, 大部分的 C 末端富集都是基于反向富集的策略。简单来说, 就是将所有非 C 末端肽段通过富集手段去除, 最后溶液中剩下的就是 C 末端肽段。本文选取几个代表性的例子来了解目前 C 末端肽段富集策略的研究状况。

Chait 研究小组利用脱水胰蛋白酶 (Anhydro-trypsin) 首次实现了对 C 末端肽段的富集<sup>[41]</sup>。脱水胰蛋白酶是一种失去催化活性的胰蛋白酶, 它不具备水解蛋白质的作用, 但是仍然能够识别 C 末端是赖氨酸或者精氨酸的肽段并与之结合。所以当蛋白质首先通过 Lys-C 酶解后, 产生的肽段 C 末端都是赖氨酸, 这样除了蛋白质的自由 C 末端肽段外, 其余的都能被

连在固相磁珠上的脱水胰蛋白酶所捕获。将磁珠与溶液分离后, 溶液中残留的就是蛋白质的 C 末端肽段。Tsunasawa 小组对该方法进行的优化<sup>[42]</sup>。他们针对于脱水胰蛋白酶磁球价格昂贵、反应效率低的弊端, 用 DITC 树脂代替了脱水胰蛋白酶磁球与赖氨酸上的氨基反应, 同时对 N 端的氨基进行了 TMPP 修饰, 以提高肽段在质谱中的响应信号。基于这样的方法, 作者测定了 3 个标准蛋白以及 PfuRPA 3 个亚单元的 C 端序列并且将该方法用于 De novo 测序。

前两种方法都是基于固相材料上的功能化基团与肽段的氨基发生的修饰。Overall 研究小组发展了针对于酸性氨基酸修饰的新方法<sup>[43]</sup>。蛋白样品经过还原烷基化后首先将所有的氨基通过甲基化保护起来, 随后通过 EDC/NHS 反应, 将所有羧基 (包括谷氨酸以天冬氨酸残基上的羧基以及蛋白质 C 端的  $\alpha$ -羧基) 用乙醇胺封闭。经过胰蛋白酶水解后, 再一次将酶解生成的  $\alpha$ -氨基通过甲基化封闭。此时只有酶解新产生的肽段有自由的  $\alpha$ -羧基, 而蛋白质 C 末端的肽段已经不含有自由的  $\alpha$ -羧基。此时将含有丰富氨基的高分子材料与这些  $\alpha$ -羧基进行 EDC/NHS 偶联反应, 通过超滤后, 溶液中保留的便是衍生化后的蛋白质 C 末端肽段。研究小组将该方法用于 *E. coli* 的分析鉴定中, 这是首次将 C 末端反向富集的方法真正应用于复杂生物体系中。结果显示在富集之前, 所有鉴定结果中只有 2.2% 归属于 C 末端肽段, 而富集之后, C 末端肽段的比例达到了 72.3%。基于此类反向富集的策略, 该研究小组引入同位素标签, 发展了“底物 C 末端氨基同位素标记法 (C-terminal amine-based isotope labeling of substrates, C-TAILS)”的方法, 用于研究由酶作用降解后产

生的新的 C 端。通过同位素标记分别在对照样品与酶促作用后的蛋白质标上轻标同位素以及重标同位素, 等比例混合后经过酶解和 C 末端肽段反向富集。在质谱图中, 蛋白质正常的 C 端由于在两组样品中都存在, 所以以强度相同的对峰形式存在, 而经过酶解作用降解得到的新的蛋白质 C 端, 则只产生带有重同位素标记的质谱峰。通过这样的方法可以寻找内源性降解产生的新蛋白质 C 端, 从而对酶作用的底物的位点进行研究分析。

但是相对于正向富集策略, 反向富集的实现取决于捕获反应的效率和完全性。一般来说这样的富集策略效率偏低, 所以目前大部分的研究工作还主要集中在简单体系中。

### 3.2 C 端肽段正向富集法

目前基于正向富集策略的 C 端富集方法极其有限, Jaffrey 小组利用酶促法将蛋白质的 C 末端特异性地标上带有生物素的标签 (Profiling protein C-terminal by enzymatic labeling, ProC-TEL), 用于 C 末端肽段的正向富集<sup>[44]</sup>。这是第一篇有关 C 末端正向富集策略的研究。羧肽酶 Y 不仅具有催化 C 末端氨基酸水解的活性, 同时也具有转肽酶的活性, 可以用带有标签的氨基酸置换 C 末端原有的氨基酸。特别是当 C 末端羧基形成酯的时候, 转肽酶的活性可以大大提高<sup>[45-47]</sup>。基于此, 研究者首先在蛋白水平上将蛋白质所有的羧基进行酯化反应, 通过羧肽酶 Y 的转肽酶作用, 使蛋白质的 C 末端带上含有生物素的标签, 而侧链上被封闭的羧基在进一步的酯水解作用下还原得到游离的羧基。经过胰蛋白酶酶解后, 只有 C 末端肽段带上了生物素的标签, 通过其与亲和素的亲和作用, 最后达到 C 末端肽段的富集目的。该

方法用于 *E. coli* 的研究中, 一共有 70 条 C 末端肽段被鉴定得到。

## 4 展望

蛋白质 C 末端的研究越来越受到人们的关注。它不仅是作为蛋白质 N 末端研究的补充, 为我们提供了更全面的蛋白质一级结构信息, 同时它也为我们开拓了一个崭新的领域, 帮助我们去了解和研究蛋白质 C 末端在各项生命活动中发挥的重要生物学功能。例如蛋白质 C 末端的序列, 翻译后修饰, 表达水平以及内源性降解产生的新的蛋白质 C 末端等都值得我们去深入地探索和思考。基于生物质谱的蛋白质 C 末端研究正是提供了这样的研究分析平台, 尽管目前的方法各有利弊, 都不是十分成熟, 但是仍然可以看到其在研究蛋白质 C 末端中潜在的巨大应用前景。因此, 发展能够应用于复杂生物体系的蛋白质 C 末端富集、鉴定和定量方法, 对于进一步深入理解蛋白质的 C 末端生物学功能具有重大意义。

## REFERENCES

- [1] Zhang CX, Weber BV, Thammavong J, et al. Identification of carboxyl-terminal peptide fragments of parathyroid hormone in human plasma at low-picomolar levels by mass spectrometry. *Anal Chem*, 2006, 78: 1636-1643.
- [2] Johnson KA, Kari PF, Tangarone BS, et al. Cation exchange-HPLC and mass spectrometry reveal C-terminal amidation of an IgG1 heavy chain. *Anal Biochem*, 2007, 360: 75-80.
- [3] Selkoe DJ. The cell biology of  $\beta$ -amyloid precursor protein and presenilin in Alzheimer's disease. *Trends Cell Biol*, 1998, 8: 447-453.
- [4] Gould SJ, Keller GA, Hosken N, et al. A conserved



- tripeptide sorts proteins to peroxisomes. *J Cell Biol*, 1989, 108: 1657–1664.
- [5] Murphy CM, Fenselau C. Recognition of the carboxy-terminal peptide in cyanogen bromide digests of proteins. *Anal Chem*, 1995, 67: 1644–1645.
- [6] Munch-Petersen B, Knecht W, Lenz C, et al. Functional expression of a multisubstrate deoxyribonucleoside kinase from *Drosophila melanogaster* and its C-terminal deletion mutants. *J Biol Chem*, 2000, 275: 6673–6679.
- [7] Stromstedt AA, Pasupuleti M, Schmidtchen A, et al. Evaluation of strategies for improving proteolytic resistance of antimicrobial peptides by using variants of EFK17, an internal segment of LL-37. *Antimicrob Agents Chemother*, 2009, 53: 593–602.
- [8] Glomset JA, Gelb MH, Farnsworth CC. Protein prenylation in eukaryotic cells: a new type of membrane anchor. *Trends Biochem Sci*, 1990, 15: 139–142.
- [9] Ferguson MA, Low MG, Cross GA. Glycosyl-sn-1,2-dimyristyl-phosphatidylinositol is covalently linked to *Trypanosoma brucei* variant surface glycoprotein. *J Biol Chem*, 1985, 260: 14547–14555.
- [10] Elortza F, Mohammed S, Bunkenborg J, et al. Modification-specific proteomics of plasma membrane proteins: identification and characterization of glycosylphosphatidylinositol-anchored proteins released upon phospholipase D treatment. *J Proteome Res*, 2006, 5: 935–943.
- [11] Cox JH, Dean RA, Roberts CR, et al. Matrix metalloproteinase processing of CXCL11/I-TAC results in loss of chemoattractant activity and altered glycosaminoglycan binding. *J Biol Chem*, 2008, 283: 19389–19399.
- [12] Reznik SE, Fricker LD. Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci*, 2001, 58: 1790–1804.
- [13] Kishimoto T, Kondo J, Takai-Igarashi T, et al. Accurate mass comparison coupled with two endopeptidases enables identification of protein termini. *Proteomics*, 2011, 11: 485–489.
- [14] López-Otín C, Overall CM. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol*, 2002, 3: 509–519.
- [15] Wilkins MR, Gasteiger E, Tonella L, et al. Protein identification with N and C-terminal sequence tags in proteome projects. *J Mol Biol*, 1998, 278: 599–606.
- [16] Wilkins MR, Ou K, Appel RD, et al. Rapid protein identification using N-terminal "sequence tag" and amino acid analysis. *Biochem Biophys Res Commun*, 1996, 21: 609–613.
- [17] Plevoda B, Sherman F. N<sup>α</sup>-terminal acetylation of eukaryotic proteins. *J Biol Chem*, 2000, 275: 36479–36482.
- [18] Eipper BA, Stoffers DA, Mains RE. The biosynthesis of neuropeptides: peptide alpha-amidation. *Annu Rev Neurosci*, 1992, 15: 57–85.
- [19] Boyd VL, Bozzini M, Zon G, et al. Sequencing of peptides and proteins from the carboxy termini. *Anal Biochem*, 1992, 206: 344–352.
- [20] Samyn B, Hardemen K, Van der Eychen, et al. Applicability of the alkylation chemistry for chemical C-terminal protein sequence analysis. *Anal Chem*, 2000, 72: 1389–1399.
- [21] Tsugita A, Takamoto K, Kamo M. A novel partial acid hydrolysis and analysis by mass spectrometry. *Eur J Biochem*, 1992, 206: 691–696.
- [22] Bailey JM. Chemical methods of protein sequence analysis. *J Chromat A*, 1995, 705: 47–65.
- [23] Hardeman K, Samyn B, Van der Eychen, et al. An improved chemical approach toward the C-terminal sequence analysis of proteins containing all natural amino acids. *Protein Sci*, 1998, 7: 1593–1602.
- [24] Schär M, Börnsen KO, Gassmann E. Fast protein sequence determination with matrix-assisted laser desorption and ionization mass spectrometry. *Rapid Commun Mass Spectrom*, 1991, 5: 319–326.
- [25] Patterson DH, Tarr GE, Regnier FE, et al.

- C-terminal ladder sequencing via matrix-assisted laser desorption mass spectrometry coupled with carboxypeptidase Y time-dependent and concentration-dependent digestions. *Anal Chem*, 1995, 67: 3971–3978.
- [26] Bonetto V, Bergman AC, Jornvall H, et al. C-terminal sequence analysis of peptides and proteins using carboxypeptidases and mass spectrometry after derivatization of Lys and Cys residues. *Anal Chem*, 1997, 69: 1315–1319.
- [27] Huang RH, Zhang Y, Wang DC. Primary structural determination of N-terminally blocked peptides from the bark of *Eucommia ulmoides Oliv* by mass spectrometric analysis. *Rapid Commun Mass Spectrom*, 2003, 17: 903–908.
- [28] Huang RH, Wang DC. Mass spectrometric strategy for primary structure determination of N-terminally blocked peptides. *J Chromatogr B*, 2004, 803: 167–172.
- [29] Samyn B, Sergeant K, Castanheira P, et al. A new method for C-terminal sequence analysis in the proteomic era. *Nat Meth*, 2005, 2: 193–200.
- [30] Kosaka T, Takazawa T, Nakamura T. Identification and C-terminal characterization of proteins from two-dimensional polyacrylamide gels by a combination of isotopic labeling and nano-electrospray fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem*, 2000, 72: 1179–1185.
- [31] Julka S, Dielman D, Young SA. Detection of C-terminal peptide of proteins using isotope coding strategies. *J Chromatogr B*, 2008, 874: 101–110.
- [32] Liu M, Zhang L, Zhang L, et al. Approach for identification and quantification of C-terminal peptides: incorporation of isotopic arginine labeling based on oxazolone chemistry. *Anal Chem*, 2013, 85: 10745–10753.
- [33] Yamaguchi M, Oka M, Nishida K, et al. Enhancement of MALDI-MS spectra of C-terminal peptides by the modification of proteins via an active ester generated in situ from an oxazolone. *Anal Chem*, 2006, 78: 7861–7869.
- [34] Nakazawa T, Yamaguchi M, Nishida K, et al. Enhanced responses in matrix-assisted laser desorption/ionization mass spectrometry of peptides derivatized with arginine via a C-terminal oxazolone. *Rapid Commun Mass Sp*, 2004, 18: 799–807.
- [35] Nakajima C, Kuyama H, Nakazawa T, et al. C-terminal sequencing of proteins by MALDI mass spectrometry through the specific derivatization of the  $\alpha$ -carboxyl group with 3-aminopropyl-tris(2,4,6-trimethoxyphenyl)phosphonium bromide. *Anal Bioanal Chem*, 2012, 404: 125–132.
- [36] Kim JS, Shin M, Song JS, et al. *De novo* analysis of protein N-terminal sequence utilizing MALDI signal enhancing derivatization with Br signature. *Anal Biochem*, 2011, 419: 211–216.
- [37] Murphy CM, Fenselau C. Recognition of the carboxy-terminal peptide in cyanogen bromide digests of proteins. *Anal Chem*, 1995, 67: 1644–1645.
- [38] Kishimoto T, Kondo J, Takai-Igarashi, et al. Accurate mass comparison coupled with two endopeptidases enables identification of protein termini. *Proteomics*, 2011, 11: 485–489.
- [39] Chan AO, Ho C, Chong H, et al. Modification of N-terminal  $\alpha$ -amino groups of peptides and proteins using ketenes. *J Am Chem Soc*, 2012, 134: 2589–2598.
- [40] Qin H, Wang F, Zhang Y, et al. Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. *Chem Commun*, 2013, 48: 6265–6267.
- [41] Sechi S, Chait BT. A method to define the carboxyl terminal of proteins. *Anal Chem*, 2000, 72: 3374–3378.
- [42] Kuyama H, Shima K, Sonomura K, et al. A simple and highly successful C-terminal sequence analysis of proteins by mass spectrometry. *Proteomics*, 2008, 8: 1539–1550.
- [43] Schilling O, Barre O, Huesgen PF, et al. Proteome-wide analysis of protein carboxy termini:

- C terminomics. Nat Meth, 2010, 7: 508–511.
- [44] Xu G, Shin SBY, Jaffrey SR. Chemoenzymatic labeling of protein C-termini for positive selection of C-terminal peptides. ACS Chem Biol, 2011, 6: 1015–1020.
- [45] Berne PF, Schmitter JM, Blanquet S. Peptide and protein carboxyl-terminal labeling through carboxypeptidase Y-catalyzed transpeptidation. J Biol Chem, 1990, 265: 19551–19559.
- [46] Lin S, Lowe CR. C-terminal labeling of immunoglobulin G with a cysteine derivative by carboxypeptidase Y catalyzed transpeptidation. Anal Biochem, 2000, 285: 127–134.
- [47] Berne PF, Blanquet S, Schmitter JM. Carboxypeptidase Y-catalyzed transpeptidation of esterified oligo- and polypeptides and its use for the specific carboxyl-terminal labeling of proteins. J Am Chem Soc, 1992, 114: 2603–2610.

(本文责编 郝丽芳)



### 《生物工程学报》对摘要的写作要求

1. 研究报告摘要：基本要素包括研究目的、方法、结果和结论（不用单列标题书写）。目的（Purpose）：主要说明作者写此文章的目的，或说明本文主要要解决的问题；方法（Methods）：重点说明作者的主要工作过程及使用的方法。应用性文章如需要，可注明条件、使用的主要设备和仪器。结果（Results）：本文最后得出的结果（实验数据部分）。结论（Conclusions）：如系基础研究，应写明本文的创新之处，及文章在讨论部分表述的观点；如系应用性研究，应尽可能提及本文结果和结论的应用范围和应用情况或应用前景。

2. 综述摘要：包括论述内容的发展水平、自己的评论及展望，尤其要注意结合自己的研究工作。

3. 英文摘要的撰写要点：英文摘要的内容应与中文摘要一致，但比中文摘要更详尽。英文摘要完成后，务必请英文较好、且专业知识强的专家审阅定稿后再返回编辑部。凡不符合要求的，即使学术上可以达到刊出的水平，本刊也将推迟发表。

(1) 建议使用第一人称，尽量不使用第三人称和被动语态。

(2) 建议用主动语态，被动语态表达拖拉模糊尽量不用，这样可以免好多长句，以求简单清晰。

(3) 尽量使用清晰简练的短句，避免很长的句子。注意正确使用英文写作习惯和语法。

(4) 摘要应当使用过去时态，语法正确，句子通顺。

(5) 摘要中避免使用缩写语，除非是那些人人皆知的（如DNA、ATP等），或者确实是非常长，而且出现多次的短语才允许用缩写语，并且在第一次出现时要写出全称。

(6) 在英文摘要中，不要使用任何汉字字符，包括标点、括号、温度、希腊字母等。

(7) 句子的开头处最好不要使用数字。