

## 综 述

# 蛋白质水解酶和样品预处理优化技术在蛋白质组学研究中的应用

吴飞林<sup>1,2</sup>, 赵明治<sup>2</sup>, 张瑶<sup>2,3</sup>, 熊智<sup>1</sup>, 徐平<sup>2</sup>

1 西南林业大学 生命科学学院, 云南 昆明 650224

2 军事医学科学院放射与辐射医学研究所 蛋白质组学国家重点实验室 国家蛋白质科学中心(北京) 北京蛋白质组研究中心 蛋白质药物国家工程研究中心, 北京 102206

3 中国科学院微生物研究所, 北京 100101

吴飞林, 赵明治, 张瑶, 等. 蛋白质水解酶和样品预处理优化技术在蛋白质组学研究中的应用. 生物工程学报, 2016, 32(3): 306–316.

Wu FL, Zhao MZ, Zhang Y, et al. Application of optimized multi-enzyme combination and sample pretreatment in proteomics. Chin J Biotech, 2016, 32(3): 306–316.

**摘 要:** 蛋白质组学是全景式鉴定、定量蛋白质, 并研究蛋白质功能的学科。基于高分辨质谱的鸟枪法蛋白质组学研究技术首先利用不同的位点特异性蛋白酶对复杂蛋白质样品进行酶解, 进而利用质谱获得蛋白质相关的定性和定量信息。为了获得高质量的质谱信息, 前期的样品处理和质谱数据采集同样重要。本文对蛋白质组学中常用的 Trypsin、Lys-C、Glu-C 等位点特异性蛋白酶的酶切特点进行了总结, 并综述了目前常用的几种酶切组合策略和样本预处理技术在提高蛋白质组学研究效率中的作用。

**关键词:** 蛋白质组学, 蛋白质水解酶, 多种酶组合技术, 过滤器辅助样品制备技术

**Received:** August 20, 2015; **Accepted:** November 4, 2015

**Supported by:** National High Technology Research and Development Program of China (863 Program) (Nos. 2011AA02A114, SS2012AA020502, 2014AA020900, 2014AA020607), National Basic Research and Development Program of China (973 Program) (Nos. 2011CB910600, 2013CB911200), National Natural Science Foundation of China (Nos. 31470809, 31170780), Beijing Natural Science Foundation (No. 5152008), Special Research Foundation of the Forestry Public Welfare (No. 201304105).

**Corresponding authors:** Ping Xu. Tel: +86-10-83147777-1314; Fax: +86-10-80705155; E-mail: xupingghy@gmail.com

Zhi Xiong. Tel: +86-871-63825322; E-mail: zhix65.swfc@gmail.com

国家高技术研究发展计划(863计划)(Nos. 2011AA02A114, SS2012AA020502, 2014AA020900, 2014AA020607), 国家重点基础研究发展计划(973计划)(Nos. 2011CB910600, 2013CB911200), 国家自然科学基金(Nos. 31470809, 31170780), 北京市自然科学基金(No. 5152008), 林业公益性行业科研专项(No. 201304105)资助。

# Application of optimized multi-enzyme combination and sample pretreatment in proteomics

Feilin Wu<sup>1,2</sup>, Mingzhi Zhao<sup>2</sup>, Yao Zhang<sup>2,3</sup>, Zhi Xiong<sup>1</sup>, and Ping Xu<sup>2</sup>

<sup>1</sup> Life Science College, Southwest Forestry University, Kunming 650224, Yunnan, China

<sup>2</sup> State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Engineering Research Center for Protein Drugs, National Center for Protein Sciences, Beijing Institute of Radiation Medicine, Beijing 102206, China

<sup>3</sup> Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

**Abstract:** Proteomics is a powerful subject focusing on large-scale study of protein structures and functions. A complete enzymatic digestion of protein complexes is the key step in modern high-resolution and high-throughput mass spectrometry (MS)-based identification and quantification. To achieve MS analysis, both peptide sample pretreatment and data acquisition are prerequisite in proteomic studies. In this paper, we summarized both the enzymatic proprieties of three common proteolytic enzymes, Trypsin, Lys-C and Glu-C, the optimization of multi-enzyme combination and an advanced sample pretreatment in proteomics research.

**Keywords:** Proteomics, proteolytic enzyme, multi-enzyme combination, filter-aided sample preparation

蛋白质组 (Proteome) 的狭义定义是指生物体的细胞、组织或者器官所有蛋白质, 包括基因转录、翻译后直接产物及其翻译后修饰 (PTM) 的蛋白质的总和。蛋白质是生物学功能的直接执行者, 表达失调可能会导致疾病的发生和发展。因此, 对蛋白质表达水平、功能、相互作用、细胞内定位信息的获取和生物体本身蛋白质基础的了解, 是研究这些蛋白质生物学功能的前提和基础<sup>[1-3]</sup>。蛋白质组学是系统鉴定和定量生物体蛋白质, 并研究蛋白质生物学功能的新兴学科。

蛋白质组学研究的初期阶段, 复杂蛋白样品的分离主要是通过双向 (2D) 凝胶电泳技术。质谱技术的发展和软电离技术的发展, 使得基于凝胶电泳分离的蛋白质鉴定成为可能<sup>[4]</sup>。

蛋白质组学研究中, 胶内消化的局限性在一定程度上限制了蛋白质检测的动态范围。相比较而言, 蛋白质溶液消化预处理的方法逐渐

地崭露头角<sup>[5]</sup>。随着液相色谱分离技术、质谱鉴定技术和生物信息解析蛋白质序列技术的发展, 鸟枪法蛋白质组 (主要是 Bottom-up 蛋白质组) 逐渐成为基于质谱的蛋白质组定性鉴定和定量的核心技术。但多年来位点特异性蛋白酶的开发及其在蛋白质组样品制备中的优化、使用没有得到应有的重视, 一定程度上影响了蛋白质组学学科的发展。

本文主要从 3 种常见的蛋白水解酶出发, 对它们的优缺点和蛋白质组学样品预处理等方面进行阐述, 并对蛋白质水解酶的发展趋势和应用前景做出展望。

## 1 蛋白质组学研究常用的蛋白酶

蛋白水解酶种类繁多, 在生物体内通过对蛋白样品的消化 (如氨基酸肽键之间水解), 实现蛋白质的降解、吸收及激活等复杂而重要的功能<sup>[6]</sup>。蛋白水解酶主要涉及多种生理过程, 包

括组成蛋白质氨基酸的分选和回收、细胞周期、细胞分化和迁移、组织的形态发生和重塑等过程<sup>[7]</sup>。蛋白质水解过程的异常或失调和许多病理过程密切相关,例如癌症、关节炎和心血管疾病等<sup>[8]</sup>。

根据国际酶学委员会 (EC) 分类,蛋白水解酶属于 3.4 亚类 (水解主要作用在肽键上),分为肽段内切酶和外肽酶。肽段内切酶作用于蛋白质内部的肽键,而外肽酶酶切的氨基酸肽键主要来自于蛋白质 N-端或 C-端,因此也称为氨肽酶或羧肽酶。蛋白酶的进一步分类主要是基于它们的催化机制、特异性和其他性质。如大多数已知蛋白酶的活性位点为半胱氨酸、丝氨酸、天冬氨酸、谷氨酸、苏氨酸、天冬酰胺或者是和水结合的金属离子,这些活性位点也可以作为蛋白酶类型的定义。蛋白水解酶的特异性往往比较低,只有少量酶切位点特异性的蛋白酶可以用于蛋白质组学研究。目前常用的蛋白酶主要有胰蛋白酶 (Trypsin)、谷氨酰胺内切酶 (Glu-C)、赖氨酰胺内切酶 (Lys-C) 和天冬氨酸酰胺内切酶 (Asp-N) 等。

胰蛋白酶是一种稳定的蛋白质水解酶,其水解位点位于蛋白质底物 C 端侧链的精氨酸和赖氨酸相连的肽键。Lys-C 是一种丝氨酸蛋白酶,它能特异水解赖氨酸的羧基端 (与精氨酸相连时除外),可以产生比胰蛋白酶更长的酶解肽段。Lys-C 也比胰蛋白酶更为稳定,在一些苛刻的条件下,例如 8 mol/L 尿素中,活性不受影响,因此 Lys-C 经常在胰蛋白酶酶切之前使用。Glu-C 也是一种常见的胞内蛋白酶,能特异切割蛋白质底物 C 端的天冬氨酸及谷氨酸相邻的肽键,发挥酶活力的最适 pH 范围在 4.0-9.0 之间,

且酶切特异性易受缓冲液的影响。在碳酸氢铵和醋酸铵缓冲液中,该酶对谷氨酸残基的酶切特异性更高;而在磷酸盐缓冲液中,谷氨酸和天冬氨酸残基均可被酶切。三种蛋白水解酶单独使用的时候,都可以特异性地酶切蛋白质的序列。采用不同蛋白水解酶的组合,对蛋白序列的确认和一些翻译后修饰位点的鉴定都是非常有帮助的。

### 1.1 胰蛋白酶

胰蛋白酶 (EC 3.4.21.4),是一种丝氨酸蛋白酶,在脊椎动物和无脊椎动物中广泛表达。胰蛋白酶在表达时通常以无活性的酶原 (胰蛋白酶原) 形式表达,经肠激酶或者胰蛋白酶本身活化后形成具有活性的胰蛋白酶。该酶可特异地在蛋白质底物的赖氨酸和精氨酸的羧基端进行酶切水解,但当赖氨酸和精氨酸后的氨基酸是脯氨酸时,水解不会发生。胰蛋白酶消化可产生含有一个碱性精氨酸或者赖氨酸结尾的短肽,比较适合肽段的碎裂,形成较强的 b 和 y 离子系列对,适于现有的鉴定算法。另外,该酶价格合理,在实践中得到了广泛应用。

目前常见的胰蛋白酶以猪源、牛源和人源胰蛋白酶为主。以人为例,胰腺分泌阳离子型胰蛋白酶原 (Cationic trypsinogen)、阴离子型胰蛋白酶原 (Anionic trypsinogen) 和胰蛋白酶原 (Mesotrypsinogen) 等 3 种胰蛋白酶原。这些胰蛋白酶原的编码基因分别为 PRSS1、PRSS2 和 PRSS3。胰蛋白酶原在胰脏中表达后被分泌到胰液,并随胰液转运到小肠。在肠激酶的作用下,胰蛋白酶原的 N 端保护序列 (识别位点为 DDDDK) 被切除并激活。被激活的胰蛋白酶被称为  $\beta$ -胰蛋白酶,可进一步激活更多的酶原。

对猪阳离子胰蛋白酶的结构分析发现经过 EK 酶(肠激酶, Enterokinase)活化, 露出第 16 位氨基酸残基新的 N 端序列。该序列可插入靠近第 195 位的丝氨酸的位置, 并且跟相邻的天冬氨酸形成离子对, 从而实现其他氨基酸的结构重排, 其中 Gly193 被重排到合适的位置而形成氧阴离子穴, 最终完成胰蛋白酶的激活, 形成的活性位点的关键氨基酸包括 His57、Asp102 和 Ser195<sup>[9]</sup>。

$\beta$ -胰蛋白酶除了可以对底物蛋白进行酶切外, 也可以发生自消化, 比较典型的是在 Lys131-Ser132 之间发生自切而产生  $\alpha$ -胰蛋白酶。 $\alpha$ -胰蛋白酶通过分子内二硫键维持其完整的分子状态, 但其活性显著降低。因此通过化学修饰降低胰蛋白酶自消化是维持其高蛋白酶活性的重要方法。位点特异性修饰如乙酰化<sup>[10]</sup>和甲基化<sup>[11]</sup>是目前通用的抑制胰蛋白酶自消化策略。

此外, 胰蛋白酶与胰凝乳蛋白酶(Chymotrypsin)在胰脏中共表达。在纯化过程中这 2 种蛋白酶很难完全分开, 因此在纯化的胰蛋白酶中难于完全去除胰凝乳蛋白酶的活性, 这使得动物来源的胰蛋白酶具有非特异性酶切活性。因此, 需要对纯化产物进行甲苯磺酰苯丙氨酰氯甲酮(TPCK)处理, 以达到淬灭胰凝乳蛋白酶的作用。

## 1.2 赖氨酰基内切酶(Lys-C)

Lys-C 是特异酶切蛋白 C 端赖氨酸残基的胞内蛋白酶, Masaki 等第一次从无色杆菌 *Achromobacter lyticus* M497-1 的肉汤培养基中分离纯化得到<sup>[12]</sup>。该酶具有很强的特异性, 特异水解蛋白质 C 端赖氨酸的肽键, 其中也包括赖氨酸和脯氨酸形成的肽键<sup>[12-13]</sup>, 其反应的 pH

范围比较广, 在 8.5 到 10.7 之间<sup>[14]</sup>。除了底物的专一性特性, 该酶还具有其他特异的特征, 如蛋白水解活性高于牛源的胰蛋白酶几个数量级<sup>[14]</sup>, 在 5 mol/L 尿素(甚至更高)和 0.1% SDS(十二烷基磺酸钠)的溶液中也能表现出高效的活性, 这些特点弥补了胰蛋白酶在上述溶液中酶切活性低的缺陷。由于 Lys-C 的这些优点, Masaki 等成功地利用 *Achromobacter* 蛋白酶作为工具, 开展蛋白的氨基酸序列研究<sup>[15]</sup>。

自从 Lys-C 发现以来, 主要基于原始菌株对该蛋白酶的作用机理和蛋白酶的纯化进行研究。Lys-C 主要从 *A. lyticus*<sup>[11]</sup>、绿脓假单胞菌 *Pseudomonas aeruginosa*<sup>[16]</sup> 和产酶溶杆菌 *Lysobacter enzymogenes*<sup>[17]</sup> 等菌株合成。Kuhlman 等<sup>[18]</sup>以 *L. enzymogenes* ATCC 27796 作为研究对象, 基于不含动物源成分的培养基生产的纯 Lys-C 蛋白酶, 该蛋白酶对于赖氨酸残基的特异性酶切更强, 且内毒素低、耐存储、并可反复冻融, 更适于在蛋白质组学研究中广泛使用。

## 1.3 谷氨酰基内切酶(Glu-C)

金黄色葡萄球菌胞内蛋白酶 Glu-C 是丝氨酸蛋白酶<sup>[19]</sup>, 它可以特异地酶切与蛋白 C 端谷氨酸或者天冬氨酸残基结合的肽键, 但是与前者相比, 对后者的酶切速率要慢大约 300 多倍<sup>[20]</sup>。该酶在 pH 3.5–9.5 范围内均有活性, 最佳活性的 pH 在 4.0 到 7.8 之间。研究表明氟磷酸盐可以特异地抑制其活性<sup>[21]</sup>。通过对蛋白质序列的优化并且经过重组表达, 分别实现大肠杆菌 *Escherichia coli*<sup>[22]</sup>、枯草芽胞杆菌 *Bacillus subtilis*<sup>[23]</sup>、灰色链霉菌 *Streptomyces griseus*<sup>[24]</sup> 等宿主中的重组表达和纯化。和大多数蛋白水解酶一样, 大肠杆菌中表达 Glu-C (Glu V8) 也

存在自身降解现象<sup>[25]</sup>。Ono 等对前肽序列研究表明其 5 个位点突变可以使 Glu-C 抑制其自身降解现象<sup>[26]</sup>，这是由于前导肽序列中 4 个氨基酸发挥了分子内伴侣的作用<sup>[25]</sup>，有效地保护自身的序列。但是对于去除前导肽序列的重组表达 Glu-C 的方法至今还没有人报道。

在蛋白质组学研究中，Glu-C 单一酶切可以很好地应用于分离和富集磷酸化肽段的鉴定<sup>[27]</sup>。与单一酶切相比，Glu-C 与胰蛋白酶联用不仅有效地提高蛋白质鉴定的序列覆盖度，而且提升了基于质谱检测的可信度<sup>[28]</sup>。

Asp-N、糜蛋白酶和胃蛋白酶等蛋白酶虽然在一些实验中有采用，但与 Trypsin、Lys-C、Glu-C 这 3 种常见的酶相比，这些非常见蛋白酶在酶切条件和兼容性方面，都不是很成熟，例如 Asp-N 是一种金属蛋白内切酶，从莓实假单胞菌 *Pseudomonas fragi* 的突变菌株中分离出来，水解的蛋白 N 端天冬氨酸和半胱氨酸残基，平均分子量为 24.5 kDa，酶切的最佳 pH 为 6.0 和 8.5<sup>[29-30]</sup>。

糜蛋白酶 (Chymotrypsin) 别名胰凝乳蛋白酶、 $\alpha$ -糜蛋白酶、甲-糜蛋白酶，是胰液中消化酶的一种，主要在十二指肠中发挥水解酶的作用，酶切蛋白和多聚肽<sup>[31]</sup>。商品化的糜蛋白酶主要来源于牛胰腺，糜蛋白酶优先酶切的是酰胺键 C 端的疏水性氨基酸 (例如酪氨酸、苯丙氨酸和色氨酸)，肽键的水解位置含有其他氨基酸 (亮氨酸和蛋氨酸) 水解速率将会降低<sup>[32-33]</sup>。

胃蛋白酶和胰蛋白酶、糜蛋白酶一样是消化系统中 3 种主要的蛋白降解或者蛋白水解酶之一，是天门冬氨酸水解酶家族成员之一。在消化的过程中，这些酶特异酶切氨基酸的肽键，

共同作用把膳食蛋白酶切成它们的共同组成成分，例如肽段和氨基酸，促进肠膜吸收。胃蛋白酶酶切位点比较广泛，其中最有效的酶切序列是疏水性氨基酸之间的肽键，更适宜的是芳香族氨基酸，例如酪氨酸、苯丙氨酸和色氨酸<sup>[34]</sup>。该酶在酸性环境中具有较高活性，其最适 pH 值约为 3。在中性或碱性 pH 值的溶液中，胃蛋白酶会发生解链而丧失活性，相对于蛋白质组学实验中的酶切条件，这种条件比较苛刻。

LysargiNase 蛋白水解酶主要酶切蛋白 N 端的赖氨酸和精氨酸残基<sup>[35]</sup>，虽说这种性质在蛋白质组学中备受关注，但是其来源比较少，酶活性也不理想，难于使用。在我们的研究中也观察到此现象，结果待发表。因此，目前这些酶在蛋白质组学大规模数据集中应用相对比较少。

## 2 蛋白质组学常用胰蛋白水解酶的不足及优化

商品化胰蛋白酶主要来源于猪或牛的胰脏，经分离纯化得到。这种方法得到的胰蛋白酶虽然产量很大，但是存在病原污染等危险，其安全性和活性不能得到保障，从而限制了它在生物制药中的应用。此外，由于动物胰脏中的胰糜蛋白酶与胰蛋白酶的性质极其相似，提取纯化过程中二者较难分开，因此胰蛋白酶常伴有糜蛋白酶活性。糜蛋白酶为非特异性酶，具有催化酶切酪氨酸、苯基丙氨酸和色氨酸等氨基酸的活性。这使得酶切位点过多，产生的肽段太短，不利于质谱鉴定。此外，不加修饰的胰蛋白酶会自身酶解成肽段，影响蛋白质测序和肽段的质谱分析。

胰蛋白酶原具有自身活化和自身降解的特

性,在真核系统中表达,难于得到完整的胰蛋白酶原,并且其活化为有活性的胰蛋白酶会对宿主细胞造成一定的毒性<sup>[36]</sup>。本实验室利用大肠杆菌作为表达体系,实现了猪胰蛋白酶原的重组表达。胰蛋白酶原以包涵体的形式表达,经过复性和乙酰化修饰等步骤,成功获得稳定性和活性远远高于动物来源胰蛋白酶的重组乙酰化胰蛋白酶,目前该酶已用于大规模蛋白质组学研究<sup>[10]</sup>。

蛋白质组学基于高精度现代质谱仪可对复杂蛋白样品进行鉴定、定量以及翻译后修饰中的位点分析。在鸟枪法蛋白质组学中,上述分析多数基于胰蛋白酶进行酶切实现的,但是并不是所有的序列都可以通过这种方法检测到<sup>[38]</sup>。由于整个蛋白质序列中精氨酸和赖氨酸残基分配不均和数量的限制,可能导致特定区段序列偏长,不利于质谱鉴定。在一些感兴趣的序列区域,例如,转膜区包含丰富的翻译后修饰位点,容易形成漏切,胰蛋白酶的酶切不完全,影响了蛋白质的有效鉴定<sup>[39]</sup>,不利于重要蛋白的后续功能分析。为提高特异性消化的效率,在实践中利用 Lys-C 和胰蛋白酶的级联消化可获得更完全的消化肽段<sup>[38]</sup>。另一方面,在大多数酶切产物中胰蛋白酶完全酶切所产生的肽段,56%以上的长度少于 6 个氨基酸残基。这些肽段太小,也不利于质谱的鉴定,从而限制了蛋白质肽段的覆盖<sup>[40]</sup>。

### 3 蛋白质组学样品的溶液酶切的劣势和优化

目前蛋白质酶解成肽段的策略主要包括胶内消化和溶液酶切两种。胶内消化涉及去污剂

对蛋白质的溶解,首先通过 SDS-PAGE 对蛋白质的分离接着进行胶内消化<sup>[41]</sup>。溶液酶切主要通过溶解能力较强且酶切兼容性的试剂抽提蛋白质,例如尿素和硫脲进行蛋白质的变性,实现蛋白质的有效溶出。随后通过降低尿素或者其他解离试剂浓度,实现变性条件环境下蛋白质的溶液消化('in-solution' digestion)<sup>[42]</sup>。胶内消化的优势是基于强去污剂 SDS 溶解蛋白比较完全,但是其缺点是肽段的抽提效率难以保证、操作复杂和不易进行自动化处理等。溶液消化易实现自动化,可以进行大容量平行样品处理,为了实现对复杂样品的质谱前预分离,需要采用 HPLC (SCX 或者高 pH-C18) 对消化肽段进行分离<sup>[43-44]</sup>,但缺点是很难保证蛋白样品的完全溶解、消化时容易被其他干扰物质影响等。因此如何实现蛋白质的充分溶解,排除表面活性剂的干扰是实现溶液酶切成功的关键。为了得到更充分的酶切,多种技术应运而生,并得到业界的广泛使用。

#### 3.1 膜辅助样品制备 (Filter-aided sample preparation, FASP) 技术在蛋白质样品预处理中的应用

SDS 可以对细胞和组织样品来源的蛋白样品溶解,是目前溶解能力最强的表面活性剂。但是 SDS 浓度高于 0.1% 时会影响胰蛋白酶的消化,进而影响肽段在质谱检测中的离子化。因此,SDS 的去除是蛋白质组学研究尤其是溶液酶切中非常重要的因素。FASP 技术<sup>[45]</sup>是最近几年发展起来的一种对质谱样品前期处理的方法。该方法首先通过 SDS 完全溶解蛋白质组样品,然后通过超滤装置(如 3 kDa, 10 kDa 超滤管)用 8 mol/L 尿素缓冲液将 SDS 缓冲液替换出

来,避免 SDS 对蛋白质消化和质谱分析的干扰。超滤装置被认为是一个“蛋白质组反应器”,可除去去污剂,实现缓冲液的替换,减少化学修饰。随后通过两步酶切,从而实现蛋白的完全溶液酶切。

FASP 方法包括 4 步:1) 先用 SDS 对蛋白样品进行完全溶解,并对蛋白中的二硫键进行还原;2) 通过超滤实现 8 mol/L 尿素替换 SDS 和对还原的半胱氨酸进行烷基化修饰;3) 在 8 mol/L 尿素条件下充分利用 Lys-C 耐受高浓度尿素的特性,首先进行 Lys-C 过夜消化,有效避免功能重要但易聚集沉淀蛋白质的丢失,进而将溶液稀释至 2 mol/L 尿素后,添加胰蛋白酶继续消化 4 h;4) 肽段的洗脱和脱盐。

FASP 方法一个重要的特征是其过滤膜可以保留高分子量的物质(蛋白质和 DNA),过滤低分子量的物质(杂质和消化的肽段)。选择一个理想分离能力的超滤管是非常重要的。消化结束后通过超滤收集滤过液并真空干燥和脱盐获得的肽段,适合后续的 LC-MS/MS 分析。该方法操作简单,并且其蛋白质鉴定效率和覆盖度高于胶内消化,因而得到了越来越多的应用。

FASP 方法可以耐受来自于任何起源生物材料中的蛋白质裂解液(含 SDS),因此不论哪个亚细胞定位的细胞或者蛋白质都可以有效地增溶,膜蛋白也不例外。随着超微体积过滤装置的出现,FASP 可以处理微克级以上的蛋白。在 FASP 的原始文献初次报道中<sup>[45]</sup>,作者采用 20 mg 小鼠组织的蛋白质组样品,其中包括脑组织、眼睛组织、肝脏组织、脾脏组织和骨骼肌组织,单个流程中最多可以鉴定超过 1 000 个蛋白,其中膜蛋白数量为 797±43 个。并且对蛋白

的序列覆盖度和对疏水性较高的蛋白如膜蛋白的鉴定数量上有明显优势。在哺乳动物的复杂蛋白质组深度覆盖分析中,FASP 技术也是一个重要的样品处理技术。本实验室利用 FASP 方法,进一步优化膜蛋白的提取和溶解缓冲液的方法并进行质谱分析,获得了比胶内消化更多的膜蛋白,充分证明了 FASP 在处理特殊样本方面可替代胶内消化,如膜蛋白等特殊样本中有很好的应用前景。同时对于一些微量级、珍贵的样本(如染色体和核蛋白等),应该也是很不错的选择。

### 3.2 蛋白水解酶组合及其在提高消化效率和蛋白质鉴定覆盖度中的应用

多种蛋白酶组合酶切蛋白质样品的策略已被广泛地应用。不同的研究者通过不同酶切组合技术,旨在实现样本鉴定中的高覆盖和更多的蛋白鉴定。

基于单一蛋白酶测序常采用的策略是通过使用蛋白的馏分简化样品复杂度和对目的蛋白富集,进而增加蛋白鉴定数量和提高序列覆盖度<sup>[46-47]</sup>。近来,通过选择多种蛋白酶组合酶切来提高蛋白序列覆盖研究已取得了新的进展。由于具有高的蛋白水解酶活性和酶切位点特异性,胰蛋白酶仍然是质谱技术中应用最普遍的酶。但是胰蛋白酶对底物蛋白质消化不完全,而且折叠紧密的蛋白质具有胰蛋白酶抵抗性,降低了胰蛋白酶对蛋白鉴定数量和序列覆盖度的贡献。

Yates 实验室<sup>[48]</sup>发明了一种组合消化的策略,通过 Lys-C 与胰蛋白酶互补进行消化,弥补了这些缺点。胰蛋白酶紧接着 Lys-C 消化,克服了紧密折叠蛋白对水解的抑制。Saveliev

等<sup>[49]</sup>也发明了一种 Trypsin/Lys-C Mix 的方法使蛋白质的质谱分析更高效。相对于单一胰蛋白酶消化, Trypsin/Lys-C Mix 使蛋白质鉴定数量增加了 20%以上, 而且实验变异系数为 0.5%, 表明消化效率的提高可以在重复实验中重现, 这使蛋白质的定量更加准确。这个研究结果也被 Glatter 等证实<sup>[50]</sup>。Biringer 等<sup>[51]</sup>发现胰蛋白酶和 Glu-C 消化人的脑脊液样品可增加蛋白的鉴定, 也提高了序列覆盖度。Swaney 等<sup>[40]</sup>开展了多种位点特异性蛋白酶 (Trypsin、Lys-C、Arg-C、Asp-N 和 Glu-C) 的组合酶切蛋白质组学研究, 结果显示尽管这个方法对于蛋白数量的增加不显著, 但是一定程度上提高了平均序列覆盖 (从 24.5%到 43.4%)。Nagaraj 等<sup>[52]</sup>首先对样品进行预分离, 接着进行胰蛋白酶、Lys-C 和 Glu-C 酶切两种不同方法的组合对 HeLa 细胞系进行高覆盖蛋白质组学研究, 几乎获得了人细胞系的高覆盖, 共鉴定了 10 255 个蛋白和 166 420 个肽段。但上述研究都是基于不连续的酶消化。最近 Mann 实验室开发了一种新的策略<sup>[53]</sup>, 使用连续的消化并过滤, 进行两步和三步消化, 对胰蛋白酶、Lys-C、Glu-C、Arg-C 和 Asp-N 进行多种组合分析。与单一胰蛋白酶消化相比, Lys-C 和胰蛋白酶的连续使用, 能够鉴定多达 40%的蛋白和磷酸化位点。Mirzaei 等<sup>[54]</sup>用 HeLa 细胞系通过 1 种、2 种和 3 种酶组合的共计 48 种消化方式, 筛选出了特定物种最佳消化组合的实验方法, 结合强阴离子分馏技术, 有效地提高了蛋白质组鉴定的序列覆盖度, 共鉴定到 8 539 个蛋白和 419 952 个独特的肽段序列, 平均序列覆盖度达到 44.7%。这个数量是在单一起始样品中鉴定的最高记录。Jacek 和 Mann<sup>[53]</sup>在研究连续酶切中使用 FASP 技术, 可以实现肽段

最大程度的分离, 进一步增加了蛋白和肽段的数量。在低于微克级的蛋白质组样品, Lys-C 和胰蛋白酶连续使用实现比单一胰蛋白酶消化多鉴定 40%以上的蛋白和磷酸化位点, 有效提升了蛋白鉴定效率, 提高了序列覆盖度。

通过不同酶切技术组合提高蛋白和肽段鉴定的数量, 对于深入研究生物体功能和新基因的发现提供了新方向。

## 4 展望

随着人类基因组计划的完成, 生命科学研究已进入了后基因组时代。生命科学的主要研究对象是功能基因组学。近几年的发展, 蛋白质组学获得了高速发展, 不论是从软件方面还是从硬件方面都有了很大的提升。位点特异性蛋白酶的开发和利用, 进一步提升了蛋白质组鉴定技术的效率。最近 Overall 实验室通过蛋白质重组表达的技术第一次从乙酸甲烷八叠球古菌 *Methanosarcina acetivoran* 纯化表达得到了新的蛋白水解酶 lysargiNase<sup>[35]</sup>, 可以特异地酶切蛋白 N 端的赖氨酸和精氨酸残基, 促进了蛋白 C 端肽段的鉴定。不仅如此, 该酶切还可以发生在赖氨酸或精氨酸单双甲基化的位点, 可促进这些位点表观遗传学修饰的检测。因此对新型蛋白酶和多酶切组合技术的开发, 将促进低丰度蛋白质及其翻译后修饰位点的鉴定。低丰度蛋白及其翻译修饰位点的确定可能对疾病机制研究、发展预警和诊断治疗都有重要的意义。基于目前的文献报道, 多个蛋白酶的组酶切, 对蛋白鉴定数量的提高没有明显的改善, 但是不同蛋白酶的组酶切对蛋白的序列覆盖有比较大的贡献, 这对蛋白翻译后修饰以及疾病相关的突变研究非常有利。胶内消化操作虽然相对成



熟,但是其操作复杂性以及酶兼容性差等问题,阻碍了该方法在大规模蛋白组学样品制备中的应用。目前 FASP 方法的应用越来越多,因为其操作的便捷性、更高的蛋白覆盖率和更多的蛋白鉴定率,受到蛋白质组学研究者的推崇。但是有其自身弱势(如低分子量蛋白的低鉴定率,尿素对某些蛋白的溶解能力有限等),促使研究者对该方法做进一步的优化。

## REFERENCES

- [1] Altelaar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*, 2013, 14(1): 35–48.
- [2] Wilhelm M, Schlegl J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. *Nature*, 2014, 509(7502): 582–587.
- [3] Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol*, 2010, 11(6): 427–439.
- [4] Shevchenko A, Tomas H, Havlis J, et al. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc*, 2006, 1(6): 2856–2860.
- [5] Zhang YY, Fonslow BR, Shan B, et al. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*, 2013, 113(4): 2343–2394.
- [6] Barrett AJ, Woessner JF, Rawlings ND. *Handbook of Proteolytic Enzymes*. Vol. 1. 3rd ed. London: Elsevier, 2012.
- [7] Puente XS, Sánchez LM, Overall CM, et al. Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet*, 2003, 4(7): 544–558.
- [8] López-Otín C, Bond JS. Proteases: multifunctional enzymes in life and disease. *J Biol Chem*, 2008, 283(45): 30433–30437.
- [9] Petersson U, Appelros S, Borgström A. Different patterns in immunoreactive anionic and cationic trypsinogen in urine and serum in human acute pancreatitis. *Int J Pancreatol*, 1999, 25(3): 165–170.
- [10] Zhao MZ, Wu FL, Xu P. Development of a rapid high-efficiency scalable process for acetylated *Sus scrofa* cationic trypsin production from *Escherichia coli* inclusion bodies. *Protein Expr Purif*, 2015, 116: 120–126.
- [11] Rice RH, Means GE, Brown WD. Stabilization of bovine trypsin by reductive methylation. *Biochim Biophys Acta (BBA)-Protein Struct*, 1977, 492(2): 316–321.
- [12] Masaki T, Nakamura K, Isono M, et al. A new proteolytic enzyme from *Achromobacter lyticus* M497–1. *Agric Biol Chem*, 1978, 42(7): 1443–1445.
- [13] Tsunasawa S, Sugihara A, Masaki T, et al. Amino acid sequence of thermostable direct hemolysin produced by *Vibrio parahaemolyticus*. *J Biochem*, 1987, 101(1): 111–121.
- [14] Masaki T, Fujihashi T, Nakamura K, et al. Studies on a new proteolytic enzyme from *Achromobacter lyticus* M497–1 II. Specificity and inhibition studies of *Achromobacter* protease I. *Biochim Biophys Acta (BBA)-Enzymol*, 1981, 660(1): 51–55.
- [15] Drapeau GR. The primary structure of staphylococcal protease. *Can J Biochem*, 1978, 56(6): 534–544.
- [16] Northrop JH, Kunitz M. Isolation of protein crystals possessing tryptic activity. *Science*, 1931, 73(1888): 262–263.
- [17] Jekel PA, Weijer WJ, Beintema JJ. Use of endoproteinase Lys-C from *Lysobacter enzymogenes* in protein sequence analysis. *Anal Biochem*, 1983, 134(2): 347–354.
- [18] Cheng G, Zheng SY. Construction of a high-performance magnetic enzyme nanosystem for rapid tryptic digestion. *Scient Rep*, 2014, 4: 6947.
- [19] Prasad L, Leduc Y, Hayakawa K, et al. The structure of a universally employed enzyme: V8 protease from *Staphylococcus aureus*. *Acta Crystallogr D: Biol Crystallogr*, 2004, 60(2): 256–259.

- [20] Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, 2003, 422(6928): 198–207.
- [21] Drapeau GR, Boily Y, Houmard J. Purification and properties of an extracellular protease of *Staphylococcus aureus*. *J Biol Chem*, 1972, 247(20): 6720–6726.
- [22] Yabuta M, Ochi N, Ohsuye K. Hyperproduction of a recombinant fusion protein of *Staphylococcus aureus* V8 protease in *Escherichia coli* and its processing by OmpT protease to release an active V8 protease derivative. *Appl Microbiol Biot*, 1995, 44(1/2): 118–125.
- [23] Kakudo S, Kikuchi N, Fujiwara T, et al. Purification, characterization, cloning, and expression of a glutamic acid-specific protease from *Bacillus licheniformis* ATCC 14580. *J Biol Chem*, 1992, 267(33): 23782–23788.
- [24] Yuji S, Masayuki Y, Kazuhiro O. Cloning and expression of the gene encoding the glutamic acid-specific protease of *Streptomyces griseus* ATCC10137. *Gene*, 1994, 150(1): 149–151.
- [25] Nemoto TK, Ohara-Nemoto Y, Ono T, et al. Characterization of the glutamyl endopeptidase from *Staphylococcus aureus* expressed in *Escherichia coli*. *FEBS J*, 2008, 275(3): 573–587.
- [26] Ono T, Nemoto TK, Shimoyama Y, et al. An *Escherichia coli* expression system for glutamyl endopeptidases optimized by complete suppression of autodegradation. *Anal Biochem*, 2008, 381(1): 74–80.
- [27] Seeley EH, Riggs LD, Regnier FE. Reduction of non-specific binding in Ga (III) immobilized metal affinity chromatography for phosphopeptides by using endoproteinase glu-C as the digestive enzyme. *J Chromatogr B*, 2005, 817(1): 81–88.
- [28] Prabakaran S, Tepp W, DasGupta BR. Botulinum neurotoxin types B and E: purification, limited proteolysis by endoproteinase Glu-C and pepsin, and comparison of their identified cleaved sites relative to the three-dimensional structure of type A neurotoxin. *Toxicon*, 2001, 39(10): 1515–1531.
- [29] Wittmann-Liebold B, Salnikow J, Erdmann VA. Advanced methods in protein microsequence analysis. Berlin: Springer-Verlag, 1986.
- [30] Drapeau GR. Substrate specificity of a proteolytic enzyme isolated from a mutant of *Pseudomonas fragi*. *J Biol Chem*, 1980, 255(3): 839–840.
- [31] Wilcox PE. [5] Chymotrypsinogens—chymotrypsins. *Methods Enzymol*, 1970, 19: 64–108.
- [32] Appel W. Chymotrypsin: molecular and catalytic properties. *Clin Biochem*, 1986, 19(6): 317–322.
- [33] Berger A, Schechter I. Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Phil Trans Roy Soc Lond. B*, 1970, 257(813): 249–264.
- [34] Dunn BM. Overview of pepsin-like aspartic peptidases. *Curr Protoc Protein Sci*, 2001, chapter 21: 21.3.1–21.3.6.
- [35] Huesgen PF, Lange PF, Rogers LD, et al. LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification. *Nat Methods*, 2015, 12(1): 55–58.
- [36] Ye W, Liu JW, Wang HY, et al. Cloning, expression, purification, and characterization of a glutamate-specific endopeptidase from *Bacillus licheniformis*. *Protein Expr Purif*, 2012, 82(1): 138–143.
- [37] Nash DC, Chase HA. Comparison of diffusion and diffusion-convection matrices for use in ion-exchange separations of proteins. *J Chromatogr A*, 1998, 807(2): 185–207.
- [38] Sanders WS, Brudges SM, McCarthy FM, et al. Prediction of peptides observable by mass spectrometry applied at the experimental set level. *BMC Bioinform*, 2007, 8 Suppl 7: S23.
- [39] Eichacker LA, Granvogl B, Mirus O, et al. Hiding behind hydrophobicity: transmembrane segments in mass spectrometry. *J Biol Chem*, 2004, 279(49): 50915–50922.
- [40] Swaney DL, Wenger CD, Coon JJ. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res*, 2010, 9(3): 1323–1329.

- [41] Shevchenko A, Wilm M, Vorm O, et al. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem*, 1996, 68(5): 850–858.
- [42] Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 2001, 19(3): 242–247.
- [43] Xu P, Duong DM, Peng JM. Systematical optimization of reverse-phase chromatography for shotgun proteomics. *J Proteome Res*, 2009, 8(8): 3944–3950.
- [44] Ding C, Jiang J, Wei JY, et al. A fast workflow for identification and quantification of proteomes. *Mol Cell Proteom*, 2013, 12(8): 2370–2380.
- [45] Nagaraj N, Lu AP, Mann M, et al. Detergent-based but gel-free method allows identification of several hundred membrane proteins in single LC-MS runs. *J Proteome Res*, 2008, 7(11): 5028–5032.
- [46] MacCoss MJ, McDonald WH, Saraf A, et al. Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci USA*, 2002, 99(12): 7900–7905.
- [47] Dator RP, Gaston KW, Limbach PA. Multiple enzymatic digestions and ion mobility separation improve quantification of bacterial ribosomal proteins by data independent acquisition liquid chromatography-mass spectrometry. *Anal Chem*, 2014, 86(9): 4264–4270.
- [48] McDonald WH, Ohi R, Miyamoto DT, et al. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int J Mass Spectrom*, 2002, 219(1): 245–251.
- [49] Saveliev S, Bratz M, Zubarev R, et al. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nat Methods*, 2013, 10(11), doi: 10.1038/nmeth.f.371.
- [50] Glatter T, Ludwig C, Ahrné E, et al. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J Proteome Res*, 2012, 11(11): 5145–5156.
- [51] Biringer RG, Amati H, Harrington MG, et al. Enhanced sequence coverage of proteins in human cerebrospinal fluid using multiple enzymatic digestion and linear ion trap LC-MS/MS. *Brief Funct Genom Proteom*, 2006, 5(2): 144–153.
- [52] Nagaraj N, Wisniewski JR, Geiger T, et al. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol Syst Biol*, 2011, 7: 548.
- [53] Wiśniewski JR, Mann M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal Chem*, 2012, 84(6): 2631–2637.
- [54] Guo XF, Trudgian DC, Lemoff A, et al. Confetti: a multiprotease map of the *HeLa* proteome for comprehensive proteomics. *Mol Cell Proteom*, 2014, 13(6): 1573–1584.

(本文责编 陈宏宇)