

蛋白质 N 末端组学研究进展

王志强^{1,2}, 张瑶^{2,3}, 洪学传¹, 徐平^{1,2}

1 武汉大学药学院 组合生物合成与新药发现教育部重点实验室, 湖北 武汉 430072

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

王志强, 张瑶, 洪学传, 等. 蛋白质 N 末端组学研究进展. 生物工程学报, 2016, 32(8): 1001–1009.

Wang ZQ, Zhang Y, Hong XC, et al. N-terminomics: proteomic strategies for protein N-terminal profiling. Chin J Biotech, 2016, 32(8): 1001–1009.

摘要: 蛋白质的 N 末端作为合成的起始, 其氨基酸序列组成及翻译后修饰直接影响着蛋白质的活性、稳定性和细胞内定位, 调控着细胞内的信号转导, 甚至决定了这些蛋白质的命运。对蛋白质 N 末端组学的系统研究不仅可以揭示 N 末端区域对整个蛋白质的重要作用, 有助于我们深入地了解蛋白质在各种生命活动中所扮演的角色, 同时在实现蛋白质组高覆盖、基因组重注释等方面也有着重要的价值。本文结合我们的现有工作, 综述了近年来蛋白质 N 末端组学的研究进展, 尤其是一些重要的基于质谱的 N 末端富集技术和方法。

关键词: 蛋白质组学, 蛋白质 N 末端, 质谱, 富集

Received: October 20, 2015; **Accepted:** November 24, 2015

Supported by: National Natural Science Foundation of China (Nos. 31400697, 31470809, 31400698), National Natural Science Foundation of Beijing (Nos. 5144027, 5152008), Key Projects in the National Science & Technology Pillar Program (No. 2012BAF14B00).

Corresponding author: Ping Xu. Tel: +86-10-61777113; Fax: +86-10-80705155; E-mail: xupingghy@gmail.com

国家自然科学基金 (Nos. 31400697, 31470809, 31400698), 北京市自然科学基金 (Nos. 5144027, 5152008), 国家重大仪器专项 (No. 2012BAF14B00) 资助。

网络出版时间: 2016-01-06

网络出版地址: <http://www.cnki.net/kcms/detail/11.1998.Q.20160106.1106.003.html>

N-terminomics: proteomic strategies for protein N-terminal profiling

Zhiqiang Wang^{1,2}, Yao Zhang^{2,3}, Xuechuan Hong¹, and Ping Xu^{1,2}

1 Key Laboratory of Combinational Biosynthesis and Drug Discovery (Wuhan University), Ministry of Education, School of Pharmaceutical Science, Wuhan University, Wuhan 430072, Hubei, China

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

3 Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

Abstract: Protein N-termini, as the beginning of translation, has a major impact on protein's biological functions. Its sequence and various post-translational modifications often affect protein activation, stability and cellular-localization, regulate the signal transduction, and even determine protein's final destiny. The systematic study of protein N-termini can clarify the vital function of the N-terminus, and provide in-depth knowledge of the multifunctional roles that protein has played in diverse biological processes. In addition, N-terminal research may help us to achieve high-coverage proteomics and re-annotate genomics. Combined with our own research, this review highlights recent progress of N-terminomics, especially some important enrichment strategies and technologies based on mass spectrometry.

Keywords: proteomics, protein N-termini, mass spectrometry, enrichment

蛋白质的合成是以 mRNA 为模板, 在核糖体、tRNA 以及多种酶的参与下, 经过翻译后修饰、信号肽切除等过程形成成熟的蛋白质。N 末端作为蛋白质合成的起始, 其序列组成及发生的翻译后修饰等影响着整个蛋白质的结构和生物学功能。同时, N 末端序列有着很高的特异性, 大约有 78%–97% 的蛋白可以通过测定其前 5 个氨基酸来实现鉴定^[1], 这样极大地减小了蛋白质组的复杂程度, 为实现蛋白质组的高覆盖提供了可能。随着蛋白质基因组学的兴起, 利用蛋白质组学数据对基因组进行重注释逐渐成为一个研究热点。运用蛋白质组学的方法对蛋白质 N 末端进行大规模的鉴定和分析, 有助于我们验证和校正已注释基因, 甚至发现一些新的基因。近 20 年, 质谱技术经历了快速的发展, 人们对蛋白质 N 末端组学的研究也更加深

入, 尤其是一些 N 末端富集技术和方法的建立、改进, 使得蛋白质 N 末端的大规模测序逐步成熟。本文对近年来蛋白质 N 末端组学的研究方法特别是一些重要的富集策略进行了综述。

1 蛋白质 N 末端的生物学意义

蛋白质的合成起始于 N 末端, 有时在翻译结束后 (或者过程中) N 末端序列会不同程度地发生翻译后修饰、信号肽切除、蛋白质水解等过程, 而这些变化和 N 末端肽段本身的序列决定着蛋白质的生物学功能, 如蛋白质的活性、稳定性, 甚至最终的成熟^[2]。第一, N 末端氨基酸序列对蛋白质的稳定性和半衰期有着重要影响。早在 1986 年, N 末端降解因子就被发现广泛存在于多种生物体中^[3]。第二, 蛋白质 N 末端发生的翻译后修饰会影响蛋白质的生物学功

能。比如, N 末端乙酰化现象在真核生物体内普遍存在, 其在蛋白质降解^[4]、蛋白质转运和定位^[5-6]、蛋白质复合体形成^[7]等方面发挥着重要作用。而很多蛋白质 N 末端的谷氨酸残基或者谷氨酰胺残基会在相应环化酶的作用下转变成焦谷氨酸残基, 这种翻译后修饰能够保护蛋白质不被氨肽酶降解^[8], 同时调控一些肽类激素和神经肽的生物活性^[9]。第三, 一些蛋白质的 N 末端含有一段信号肽, 它们会影响蛋白质的转运和定位, 通常在蛋白质到达指定位点后便会被切除^[10]。此外, 一些蛋白质需要发生水解才能具备生物活性。例如我们常用的胰蛋白酶, 它需要切除其保护基团后才能暴露活性位点, 进而发挥生物学功能^[11]。

因此, 对蛋白质 N 末端的测定和分析不仅可以帮助我们明确其一级结构及可能存在的变体形式, 为蛋白质及其变体的生物学功能研究及在生理、病理条件下生物学意义的阐述提供条件, 同时在基因组重注释方面也有助于我们验证和校正已注释基因, 发现新的基因编码特征, 从而完善并促进基因组学的研究。

2 蛋白质 N 末端的非质谱测序

2.1 化学测序

早在 1949 年, 瑞典有机化学家 Edman 就提出了著名的蛋白质 N 末端测序方法——Edman 降解法^[12]。其原理是蛋白质的 N 末端氨基酸残基首先和异硫氰酸苯酯 (PITC) 耦合, 形成苯氨基硫甲酰 (PTC) 衍生物后用三氟乙酸 (TFA) 从多肽链上切下修饰的残基, 再经层析鉴定, 余下的多肽链被回收再进行下一轮降解循环。虽然 Edman 降解法被其他一些研究者进一步发展和完善^[13], 但是其方法本身的限制无法忽视。

例如, 该法仅适用于一些纯度较高的简单蛋白, 面对复杂样本便无能为力; 其次, Edman 降解是以 PITC 对蛋白质 N 末端氨基的修饰为基础的, 因而难以对 N 末端已发生修饰的蛋白进行测序; 而随着大数据时代的到来, 作为经典的 N 末端测序方法, Edman 降解已无法满足研究者对蛋白质 N 末端的大规模测序的需求。

2.2 算法预测

随着基因组测序技术的日趋成熟, 对基因组进行自动化的注释也成为可能。目前已有不少预测蛋白质 N 末端的算法能够整合翻译起始位点、翻译后修饰以及信号肽切除等信息, 例如 ProTISA、EasyGene 和 GeneMarkS 等^[14-15], 但是并不完善。在真核生物中, 翻译起始位点相当保守, 但一般的算法仍然难以对其进行准确地预测^[16]。这个问题在原核生物中更为复杂, 因为算法必须考虑多个可能的起始密码子和核糖体结合序列^[14], 如 Nielsen 等^[17]对 143 个原核生物的基因组进行研究后发现, 有些基因组的注释错误达到 60%, 而蛋白质 N 末端的注释不对等现象更为突出^[18]。随着宏基因组研究的出现, 这些软件在面对复杂的原核生物群时, 会遇到更大的挑战^[19]。另一方面, 不少算法也被专门用来预测信号肽切除位点, 例如 SignalP、Signal-3L 和 Signal-BLAST^[20-21]等, 这些算法通常根据 N 端的疏水区域和附近的中性氨基酸残基来定位, 因而并不能够有效区分信号肽与信号锚以及跨膜螺旋结构域^[20,22]。SignalP 4.0 被认为是目前较为可靠的信号肽预测软件^[23], 但通常情况下我们难以实现生物学实验的有效验证。

虽然上述这些预测算法在不断地改进, 但是面对生物体内复杂的环境以及多种可能的生

化反应，结果的准确性仍旧存在很大的疑问。随着近 20 年来质谱技术的飞速发展，其高分辨率、高灵敏度、高通量的特性为很多研究者提供了有力的技术支持，一些新的基于质谱的 N 末端测序手段也不断地被提出并广泛应用。

3 蛋白质 N 末端的特异性富集策略

目前，大多数鉴定蛋白质 N 末端的方法是基于鸟枪法的研究手段。即首先将蛋白质混合物酶解为大量肽段，随后进行串联质谱检测，最后利用生物信息的方法对样品中的肽段和蛋白信息进行分析。但是，由于复杂样品中高丰度蛋白的存在，N 末端肽段很容易淹没在大量高丰度肽段中，这使得直接鉴定蛋白质的 N 末端序列变得困难，甚至无法鉴定。因此，近年来发展出了不少特异性富集 N 末端肽段的方法，大大提高了鉴定 N 末端肽段的有效性，并

且已经应用到高通量的蛋白组学研究中。根据 N 末端肽段富集策略的不同，我们大致可以将其分为正向富集和反向富集两类 (表 1)。

3.1 蛋白质 N 末端的正向富集

正向富集方法主要是利用 N 末端肽段与酶解后产生的中间肽段的差异来实现两者的分离。一类正向富集的策略是在蛋白质的 N 末端加一个化学标签，随后通过特异结合标签手段富集到 N 末端肽段。TIMMER 等^[24]首先对样本进行还原烷基化处理，再利用胍基化反应封闭蛋白质中所有赖氨酸残基的 ϵ 氨基，随后使用 NHS-SS-biotin (sulfosuccinimidyl-2-(biotinamido) ethyl-1,3-dithiopropionate) 同 N 末端的 α 氨基反应，酶解后用链霉亲和素富集到带有生物素的 N 末端肽段。他们将这种策略分别应用到酵母、大肠杆菌、鼠、人胚胎肾细胞和人血清等 5 种样本中，研究了其蛋白质水解现象，比如甲硫

表 1 蛋白质 N 末端的富集策略

Table 1 Strategies for N-terminal enrichment

	Enrichment method	Label	Capture	Reference
Positive strategies	Affinity capture	NHS-SS-biotin	Streptavidin	[24]
	Affinity capture	SATA	Thiol-affinity resin	[25]
	N-CLAP	NHS-SS-biotin	Avidin-based resin	[26]
	Affinity capture	TMPP	Anti-TMPP antibody	[27]
	SCX	Dimethyl	None	[30]
Negative strategies	COFRADIC	Acetyl	None	[31]
	COFRADIC	TMPP	None	[32]
	ChaFRADIC	Dimethyl	None	[33]
	TAILS	iTRAQ	HPG-ALD	[35]
	TAILS	Dimethyl	HPG-ALD	[36]
	DICAS	Dimethyl	POROS-AL	[39]
	PTAG	Dimethyl	TiO ₂	[40]
Other strategies	N-TOP	TMPP	None	[41]
	dN-TOP	TMPP	None	[44]

氨酸切除、信号肽切除以及线粒体运输肽的切除等。另一种类似的做法是用 SATA (N-succinimidyl S-acetylthioacetate) 代替 NHS-SS-biotin 修饰蛋白的 N 末端, 随后用一种可和硫醇反应的试剂实现富集^[25]。与链霉亲和素富集法相比, 该法的反应效率较高, 富集的特异性达到 97%。研究者利用该方法在曲霉菌 *Aspergillus niger* 中共鉴定到 1 672 个 N 末端肽段和蛋白质水解位点。Xu 等^[26]在 Edman 降解法原理的基础上开发了一种 N-CLAP (N-terminalomics by chemical labeling of the α -amine of proteins) 的 N 末端富集方法。研究者同样使用异硫氰酸苯酯 (PITC) 封闭所有蛋白质的氨基 (包括 N 末端的 α 氨基和赖氨酸残基支链上的 ϵ 氨基), 接着使用三氟乙酸 (TFA) 处理使得蛋白 N 末端暴露出 α 氨基, 和 NHS-SS-biotin 反应后通过亲和纯化得到 N 末端肽段。这种方法同 Edman 降解一样仅适用于 N 末端未发生修饰的蛋白, 同时反应步骤较多, 增大了 N 末端肽段损失的可能性, 因而富集效果有限。而 Bland 等^[27]则使用一种可以特异性修饰蛋白 N 末端的试剂——TMPP (N-tris (2,4,6-trimethoxyphenyl) phosphonium acetyl) 来修饰玫瑰杆菌 *R. denitrificans* 的全蛋白, 酶解后通过反相液相色谱分为多个组分, 再向每个组分中加入偶联在磁珠上的 TMPP 抗体来富集 N 末端肽段。研究者共制备了 10 种 TMPP 抗体 (anti-TMPP1-anti-TMPP10), 经过筛选与评估, 最终选择 anti-TMPP6 用作后续研究。结果共鉴定到 269 个 N 末端序列, 同时发现了 3 个新基因。这种方法的优点在于做法直接, 且 TMPP 抗体富集的特异性达到 97%, 但是前期抗体的制备与筛选较为繁琐, 使得整个富集流程较长,

难于推广。

另一类正向富集 N 末端肽段的方法是利用 N 末端肽段与中间肽段电荷的不同, 通过强阳离子交换 (SCX) 的手段分离得到 N 末端肽段。这种富集方法通常用于研究 N 末端乙酰化现象, 流程简单、可操作性强。在真核生物中, 大约有 80% 的蛋白会发生 N 末端乙酰化修饰^[28]。由于乙酰化修饰的存在, 酶切后的 N 末端肽段比中间肽段少带一个正电荷, 这样可在强阳离子交换柱中实现分离。同时, 还可以结合不同酶 (例如 Lys-C 和 Lys-N) 的特点来提高 N 末端肽段的覆盖度^[29]。Chen 等^[30]在上述基础上引入了双甲基化标记的方法, 用于分析 HepG2 细胞的 N 末端乙酰化情况, 结果发现绝大多数 N 末端乙酰化肽段在前两个 SCX 组分 (共 8 个组分) 被鉴定到, 同时发现了一些仅在 N 末端序列上存在差异的蛋白质变体, 如 ERK1/ERK2、 β -actin/ γ -actin 等。相比化学修饰的正向富集策略, 强阳离子交换的分析手段更为简捷, 但该方法仅对如 N 末端乙酰化这样特定的翻译后修饰形式适用, 难于成为一种 N 末端蛋白质组研究的通用方法。

3.2 蛋白质 N 末端的反向富集

尽管蛋白质 N 末端的正向富集方法已有不少应用, 鉴于其策略本身的局限性 (如大多数正向富集法依赖于蛋白质 N 末端未被修饰的 α 氨基), 近年来关于反向富集方法的研究和应用更为广泛。反向富集方法通常是在酶解后去除中间肽段来获得 N 末端肽段。其中使用较多的是 Gevaert 和 Vandekerckhove 在 2003 年开发的 COFRADIC (Combined fractional diagonal chromatography) 法^[31]。该方法首先在常规的还原烷基化后封闭蛋白质所有的氨基, 随后第一次反相液相色谱分离

为若干组分,再利用一种带有较强疏水性的试剂 TNBS (2,4,6-trinitrobenzenesulfonic acid) 同中间肽段的 α 氨基反应,二次反相液相色谱分离后得到 N 末端肽段。该法虽然需要经过两次反相液相色谱分离,但其灵敏度较高,加上与其他技术和方法的交叉,已发展成为一种主要的 N 末端富集方法。例如, Bland 等^[32]在 COFRADIC 方法中引入了 TMPP,对 *R. denitrificans* 进行研究,结果共鉴定到 534 个 N 末端肽段,包括 5 个新基因和 41 个新的翻译起始位点。Venne 等^[33]还在 COFRADIC 方法的基础上结合强阳离子交换的分离手段开发了 ChaFRADIC (Charge-based fractional diagonal chromatography) 方法,后者更加简便易行,灵敏度也更高。在总量为 50 μg 的样本中,仅从 40%的组分就鉴定到 1 459 个 N 末端肽段,最终共有 36 个 icp55 的底物被鉴定到,包括一个新的底物 Isa2。

另一种较为成熟的反向富集策略是 TAILS (Terminal amine isotopic labeling of substrates) 方法^[34-36],该方法首先用同位素化合物标记蛋白质中所有的氨基,酶解后利用一种富含醛基的聚合物 HPG-ALD 除去中间肽段,剩余即为 N 末端肽段。基于稳定同位素标记通常有 dimethyl-TAILS^[36]和 iTRAQ-TAILS^[35]两种定量研究 N 末端蛋白质组的方法,这两种方法的标记效率均比较高,且 HPG-ALD 聚合物清除中间肽段的能力也较强,因此,TAILS 方法被应用到多个研究中。例如, Prudova 等^[35]利用 iTRAQ-TAILS 法在鼠胚胎成纤维细胞中研究了两个结构相似的基质金属蛋白酶变体 MMP-2 和 MMP-9,结果共鉴定到 201 个 MMP-2 的酶切产物和 19 个 MMP-9 的酶切产物。同时,这种方法还被用作研究 MMP-10^[37]和鉴定二肽酶

DP8 和 DP9 的内源性酶切底物^[38]。而 DICAS (Dimethyl isotope-coded affinity selection) 方法^[39]的设计同 TAILS 法类似,不同点在于利用另一种含有醛基的基质 (POROS-AL) 除去中间肽段,但是该法的清除效率较低,因此应用前景有限。此外,还有 PTAG (Phospho tagging) 法^[40],这种方法主要是在中间肽段的 N 末端添加一个磷酸化标签,再利用镍柱除去中间肽段。研究者在奈瑟氏菌 *Neisseria meningitidis* 和酵母 *Saccharomyces cerevisiae* 中分别鉴定到 753 个和 928 个 N 末端肽段。

3.3 其他富集策略

上述 COFRADIC 方法和 TAILS 方法是目前较为主流的 N 端富集方法,而另一种鉴定 N 末端的方法 N-TOP (N-terminal-oriented proteomic) 也有一定应用^[41]。该法是 Gallien 等在 2009 年开发,首先利用 TMPP 修饰耻垢分枝杆菌 *Mycobacterium smegmatis* 全蛋白,随后进行 1D SDS-PAGE 分离,分离后一半使用 trypsin 酶切,一半使用 Asp-N 酶切,结果共鉴定到 443 个 N 末端序列。严格意义上来讲,N-TOP 方法并不存在富集,但其与 TMPP 的结合同样能够实现 N 末端肽段的有效鉴定。TMPP 之所以会在 N 末端富集策略中备受青睐,原因大致有以下几点:1) 在 pH 8.2 时, TMPP 能够特异地和蛋白质 N 末端的 α 氨基反应,为后续的富集奠定了基础;2) TMPP 本身带有 +1 价,有效提高了肽段的离子化效率^[42],弥补了 N 末端肽段丰度低的劣势;3) TMPP 修饰肽段的疏水性增强,多在反相液相色谱中保留时间较大的位置出现,降低了肽段的复杂程度;4) TMPP 修饰后的肽段在质谱中的二级碎裂主要以 a、b 离子为主,谱图较为干净,利于鉴定^[43]。之后 Bertaccini 等^[44]

在此基础上发展出了 dN-TOP (Doublet N-terminal oriented proteomics) 的方法,即利用轻、重同位素标记的 TMPP 分别修饰蛋白质的 N 末端,1:1 混合后进行 LC-MS/MS 检测。这样不仅提高了 N 末端蛋白质组鉴定的准确性,还能够达到定量的目的。

作者所在的实验室也开展了 N 末端蛋白质组的研究工作。如我们将泛素 (Ubiquitin) 作为模型蛋白,分析 TMPP 修饰前后其 N 末端序列 (MQIFVK) 在色谱和质谱行为上的变化。结果发现, TMPP 修饰后的 N 末端序列 (TMPP-MQIFVK) 在反相液相色谱中的保留时间明显增大,这表明 TMPP 基团能够显著增强 N 末端序列的疏水性;同时,在高能碰撞诱导碎裂 (HCD) 模式下,其二级碎裂结果不同于常规的 b、y 离子序列,而主要表现为 a、b 离子序列。TMPP 的这两个特点为我们下一步在反相液相色谱梯度优化、质谱参数优化以及谱图过滤等方面提供了重要信息,有利于 N 末端肽段的分离、鉴定和确认。随后,我们利用 TMPP 修饰模式生物 *S. cerevisiae* 的全蛋白,并尝试了不同的蛋白酶 (如 Lys-N) 对上述样本进行酶切,结合强阳离子交换的分离手段,探索新的高效分离 N 末端肽段的方法,实现 N 末端肽段的正向富集 (另文发表)。

综上所述,虽然近年来对蛋白质 N 末端组学的研究在持续发展,各种富集策略也在不断地被提出、改进,但是从总体情况来看,N 末端的覆盖度仍旧比较低^[21]。归纳起来,大致有内源和外源两方面的原因。内源性的原因主要有以下几点:1) 相对于整个蛋白质组来说,N 末端肽段的比例较低;2) 有些 N 末端肽段特殊的理化性质使其难以被质谱检测,如分子量小、

离子化效率低等因素^[40];3) N 末端肽段存在的多种翻译后修饰不利于化学标记。外源性的原因主要是在富集过程中产生的,有三个方面:1) 样品会在提取、制备以及标记过程中发生降解;2) 富集过程会增加样品的复杂程度和 N 末端肽段的损失;3) 不完全的标记和富集会影响 N 末端肽段最终的收率。

4 小结与展望

目前,基于蛋白质组学的方法研究蛋白质 N 末端逐渐地走向成熟,并且已经展现出其在蛋白质 N 末端的鉴定、定量以及揭示蛋白质生物学功能方面的重要价值。这不仅帮助我们清楚地了解蛋白质 N 末端的氨基酸序列、所发生的翻译后修饰等信息,同时也提示我们进一步探究更为复杂的生命现象。虽然当前研究蛋白质 N 末端组学的技术和富集手段并不完善,但是随着质谱技术的不断发展,针对蛋白质 N 末端组学开发出更加有效的材料、试剂甚至全新的理论和策略,尤其是一些预测算法的逐步改进和成熟,我们一定会用更短的时间得到更加可靠、更加全面的 N 末端信息,从而使我们能够更加深入地了解蛋白质 N 末端在整个生命活动中所发挥的重要作用。

REFERENCES

- [1] 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, 221(3): 609-613.
- [2] Lai ZW, Petrera A, and Schilling O. Protein amino-terminal modifications and proteomic approaches for N-terminal profiling. *Curr Opin Chem Biol*, 2015, 24: 71-79.
- [3] Bachmair A, Finley D, and Varshavsky A. *In vivo*

- half-life of a protein is a function of its amino-terminal residue. *Science*, 1986, 234(4773): 179–186.
- [4] Persson B, Flinta C, Heijne G, et al. Structures of N-terminally acetylated proteins. *Eur J Biochem*, 1985, 152(3): 523–527.
- [5] Forte GM, Pool MR, and Stirling CJ. N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. *PLoS Biol*, 2011, 9(5): 1102.
- [6] Dikiy I and Eliezer D. N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound α -synuclein and increases its affinity for physiological membranes. *J Biol Chem*, 2014, 289(6): 3652–3665.
- [7] Scott DC, Monda JK, Bennett EJ, et al. N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science*, 2011, 334(6056): 674–678.
- [8] Van Coillie E, Proost P, Van Aelst I, et al. Functional comparison of two human monocyte chemotactic protein-2 isoforms, role of the amino-terminal pyroglutamic acid and processing by CD26/dipeptidyl peptidase IV. *Biochem*, 1998, 37(36): 12672–12680.
- [9] Shih YP, Chou CC, Chen YL, et al. Linked production of pyroglutamate-modified proteins via self-cleavage of fusion tags with TEV protease and autonomous N-terminal cyclization with glutaminyl cyclase *in vivo*. *PLoS ONE*, 2014, 4: e94812.
- [10] Zybailov B, Rutschow H, Friso G, et al. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE*, 2008, 3(4): e1994.
- [11] Naldini L, Tamagnone L, Vigna E, et al. Extracellular proteolytic cleavage by urokinase is required for activation of hepatocyte growth factor/scatter factor. *EMBO J*, 1992, 11(13): 4825.
- [12] Edman P. A method for the determination of amino acid sequence in peptides. *Arch Biochem*, 1949, 22(3): 475.
- [13] Chen W, Yin X, Mu J, et al. Subfemtomole level protein sequencing by Edman degradation carried out in a microfluidic chip. *Chem Commun*, 2007, (24): 2488–2490.
- [14] Hu G.Q, Zheng X, Yang YF, et al. ProTISA: a comprehensive resource for translation initiation site annotation in prokaryotic genomes. *Nucleic Acids Res*, 2008, 36(suppl 1): D114–D119.
- [15] Hyatt D, Chen GL, LoCascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 2010, 11(1): 119.
- [16] Van der Burgt A, Severing E, Collemare J, et al. Automated alignment-based curation of gene models in filamentous fungi. *BMC Bioinformatics*, 2014, 15(1): 19.
- [17] Nielsen P and Krogh A. Large-scale prokaryotic gene prediction and comparison to genome annotation. *Bioinformatics*, 2005, 21(24): 4322–4329.
- [18] Raghavan R, Sage A, and Ochman H. Genome-wide identification of transcription start sites yields a novel thermosensing RNA and new cyclic AMP receptor protein-regulated genes in *Escherichia coli*. *J Bacteriol*, 2011, 193(11): 2871–2874.
- [19] Liu Y, Guo J, Hu G, et al. Gene prediction in metagenomic fragments based on the SVM algorithm. *BMC Bioinformatics*, 2013, 14(Suppl 5): S12.
- [20] Ivankov DN, Payne SH, Galperin MY, et al. How many signal peptides are there in bacteria? *Environ Microbiol*, 2013, 15(4): 983–990.
- [21] Hartmann EM and Armengaud J. N-terminomics and proteogenomics, getting off to a good start. *Proteomics*, 2014, 14(23–24): 2637–2646.
- [22] Nielsen H, Engelbrecht J, Brunak S, et al. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*, 1997, 10(1): 1–6.
- [23] Petersen TN, Brunak S, von Heijne G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, 2011, 8(10): 785–786.
- [24] Timmer J, Enoksson M, Wildfang E, et al. Profiling constitutive proteolytic events *in vivo*. *Biochem J*, 2007, 407: 41–48.
- [25] Kim JS, Dai Z, Aryal UK, et al. Resin-assisted enrichment of N-terminal peptides for characterizing proteolytic processing. *Anall Chem*, 2013, 85(14): 6826–6832.
- [26] Xu G, Shin SBY, and Jaffrey SR. Global profiling of protease cleavage sites by chemoselective labeling of protein N-termini. *Proc Natl Acad Sci USA*, 2009, 106(46): 19310–19315.
- [27] Bland C, Bellanger L, and Armengaud J. Magnetic

- immunoaffinity enrichment for selective capture and ms/ms analysis of n-terminal-TMPP-labeled peptides. *J Proteome Res*, 2013, 13(2): 668–680.
- [28] Arnesen T, Van Damme P, Polevoda B, et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci USA*, 2009, 106(20): 8157–8162.
- [29] Hennrich ML, Groenewold V, Kops GJPL, et al. Improving depth in phosphoproteomics by using a strong cation exchange-weak anion exchange-reversed phase multidimensional separation approach. *Anal Chem*, 2011, 83(18): 7137–7143.
- [30] Chen SH, Chen CR, Chen SH, et al. Improved N- α -acetylated peptide enrichment following dimethyl labeling and SCX. *J Proteome Res*, 2013, 12(7): 3277–3287.
- [31] Gevaert K, Goethals M, Martens L, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat Biotechnol*, 2003, 21(5): 566–569.
- [32] Bland C, Hartmann EM, Christie-Oleza JA, et al. N-terminal-oriented proteogenomics of the marine bacterium *Roseobacter denitrificans* Och114 using (N-Succinimidylloxycarbonylmethyl) tris (2, 4, 6-trimethoxyphenyl) phosphonium bromide (TMPP) labeling and diagonal chromatography. *Mol Cell Proteomics*, 2014, 13(5): 1369–1381.
- [33] Venne AS, Vogtle FN, Meisigner C, et al. Novel highly sensitive, specific, and straightforward strategy for comprehensive N-terminal proteomics reveals unknown substrates of the mitochondrial peptidase Icp55. *J Proteome Res*, 2013, 12(9): 3823–3830.
- [34] Kleifeld O, Doucet A, auf dem Keller U, et al. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol*, 2010, 28(3): 281–288.
- [35] Kleifeld O, Doucet A, Prudova A, et al. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc*, 2011, 6(10): 1578–1611.
- [36] Wilson CH, Indarto D, Doucet A, et al. Identifying natural substrates for dipeptidyl peptidases 8 and 9 using terminal amine isotopic labeling of substrates (TAILS) reveals *in vivo* roles in cellular homeostasis and energy metabolism. *J Biol Chem*, 2013, 288(20): 13936–13949.
- [37] Schlage P, Egli FE, Nanni P, et al. Time-resolved analysis of the matrix metalloproteinase 10 substrate degradome. *Mol Cell Proteomics*, 2014, 13(2): 580–593.
- [38] Shahinian H, Loessner D, Biniousek ML, et al. Secretome and degradome profiling shows that Kallikrein-related peptidases 4, 5, 6, and 7 induce TGF β -1 signaling in ovarian cancer cells. *Mol Oncol*, 2014, 8(1): 68–82.
- [39] Shen PT, Hsu JL, and Chen SH. Dimethyl isotope-coded affinity selection for the analysis of free and blocked N-termini of proteins using LC-MS/MS. *Anal Chem*, 2007, 79(24): 9520–9530.
- [40] Mommen GPM, van de Waterbeemd B, Meiring HD, et al. Unbiased selective isolation of protein N-terminal peptides from complex proteome samples using phospho tagging (PTAG) and TiO₂-based depletion. *Mol Cell Proteomics*, 2012, 11(9): 832–842.
- [41] Gallien S, Perrodou E, Carapito C, et al. Ortho-proteogenomics: multiple proteomes investigation through orthology and a new MS-based protocol. *Genome Res*, 2009, 19(1): 128–135.
- [42] Chen W, Lee PJ, Shion H, et al. Improving de novo sequencing of peptides using a charged tag and C-terminal digestion. *Anal Chem*, 2007, 79(4): 1583–1590.
- [43] An M, Dai J, Wang Q, et al. Efficient and clean charge derivatization of peptides for analysis by mass spectrometry. *Rapid Commun Mass Sp*, 2010, 24(13): 1869–1874.
- [44] Bertaccini D, Vaca S, Capaito C, et al. An improved stable isotope N-terminal labeling approach with light/heavy TMPP to automate proteogenomics data validation: dN-TOP. *J Proteome Res*, 2013, 12(6): 3063–3070.

(本文责编 陈宏宇)