

## 基于精氨酸酶切的蛋白质 C 端肽段富集方法的优化及评估

赵晓晓<sup>1,2</sup>, 胡昊<sup>2</sup>, 赵雯思<sup>2,3</sup>, 刘萍<sup>1,2</sup>, 谭敏佳<sup>1,2,3\*</sup>

(1. 南京中医药大学新中药学院, 江苏南京 210023; 2. 中国科学院上海药物研究所新药研究国家重点实验室, 上海 201203; 3. 中国科学院大学, 北京 100049)

**摘要:** 基于聚合物的蛋白质 C 端反向富集策略是用于研究蛋白质 C 端最为广泛的策略之一。目前, 基于胰蛋白酶 (trypsin) 切割精氨酸残基 C 端 (ArgC 型酶切) 的蛋白质 C 端组学方法对蛋白质 C 端的鉴定深度仍有待提高。为解决这一问题, 该研究对此方法进行了优化和评估: 建立了基于“V 型”过滤装置的“一锅法”富集流程, 避免了副反应的干扰, 缩短了样本的制备时间; 优化了蛋白水平乙酰化反应条件, 最大限度地降低了丝氨酸、苏氨酸、酪氨酸残基上的副反应, 提高了肽段鉴定的可信性; 优化了基于固相萃取枪头膜片过滤柱 (StageTip 柱) 的样品分离过程, 使 C 端肽段的鉴定深度增加至原来的 4 倍。通过以上优化, 按照肽段水平错误发现率 (FDR) < 0.01、离子分数 (ion score)  $\geq 20$ , 且 C 端带有乙醇胺修饰的数据筛选标准, 从 HEK 293T 细胞中共鉴定出 696 个蛋白质 C 端。若仅要求肽段水平 FDR < 0.01, 鉴定数目进一步增加到 933 个, 这是基于聚合物富集策略的蛋白质 C 端组学方法所得的最大数据集之一。探索了胰蛋白酶镜像酶 (LysargiNase) 切割精氨酸残基 N 端 (ArgN 型酶切) 与不同肽段 N 端衍生修饰组合对蛋白质 C 端鉴定数目和种类的影响, “LysargiNase 酶切+肽段 N 端乙酰化”新策略在原有“胰蛋白酶酶切+肽段 N 端二甲基化”策略的基础上将鉴定蛋白质 C 端的种类提升了 47%。综上, 该研究通过对基于 Arg 型酶切的蛋白质 C 端组学方法的优化, 提升了 C 端肽段的鉴定深度, 扩大了 C 端肽段鉴定的覆盖范围。该方法将有望成为系统性表征蛋白质 C 端的有力工具。

**关键词:** 化学乙酰化; 蛋白质 C 端组学; 反向富集; 固相萃取枪头膜片过滤柱; 生物质谱

**中图分类号:** O658      **文献标识码:** A      **文章编号:** 1000-8713(2022)01-0017-11

## Optimization and evaluation of protein C-terminal peptide enrichment strategy based on arginine cleavage

ZHAO Xiaoxiao<sup>1,2</sup>, HU Hao<sup>2</sup>, ZHAO Wensi<sup>2,3</sup>, LIU Ping<sup>1,2</sup>, TAN Minjia<sup>1,2,3\*</sup>

(1. School of Chinese Materia Medica, Nanjing University of Chinese Medicine, Nanjing 210023, China; 2. State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** As unique biomarkers, protein C-termini are involved in various biological processes such as protein trafficking, subcellular relocation, and signal transduction. Dysregulation of protein C-terminal status is critical during the development of various diseases, including cardiovascular, neurodegenerative, and metabolic diseases and cancer. Thus, global profiling of protein C-termini is of great value in providing mechanistic insight into biological or pathological processes, as well as for identifying potential new targets for therapeutic treatment.

Polymer-based negative enrichment is a prominent C-terminomics strategy with advantages of universal applicability and parallel sample preparation. Compared with other methods of such a strategy, the profiling depth of the approaches based on enzymatic cleavage of Arg residues still

收稿日期: 2021-04-03

\* 通讯联系人. Tel: (021) 50800172, E-mail: mjtan@simmm.ac.cn.

基金项目: 国家自然科学基金(21907100); 国家科技重大专项“重大新药创制”(2018ZX09711002-004); 中国科学院青年创新促进会。

**Foundation item:** National Natural Science Foundation of China (No. 21907100); National Science & Technology Major Project “Key New Drug Creation and Manufacturing Program” (No. 2018ZX09711002-004); Youth Innovation Promotion Association, Chinese Academy of Sciences.

needs to be improved. This greatly limits our understanding of the physiological functions and molecular mechanisms of C-termini. To add a more powerful tool for C-terminomics, Arg cleavage-based negative enrichment C-terminomics was optimized and evaluated.

First, the sample preparation process was optimized. A one-pot enrichment platform based on a V-shaped filter was established, which reduced sample loss, avoided cross-contamination between reactions, and shortened sample preparation time. In addition, the protein-level acetylation conditions were investigated with the optimal labeling conditions as follows: triple coupling using 5 mmol/L Ac-NHS at pH 7.0 and 500 mmol/L ammonium for 15 min provided minimized acetylation rates (acetylation labeling efficiencies of Ser, Thr, and Tyr were lower than 4%, 2%, and 1%, respectively), along with the highest peptide-spectrum match number and satisfactory Lys labeling efficiency (up to 98%). These optimized conditions would not only minimize acetylation, but also facilitate the identification of C-terminal peptides.

Second, it was speculated that the unexpected low identification rate was primarily caused by the interference of the large number of organic compounds accumulated during the peptide-level reactions, including reagents, organic buffering agents, and their complex side-reaction products. Therefore, the conditions for StageTip-based fractionation, including pH, the amount of Empore C18 beads, and the number of fractions, were optimized. As a result, by separating the sample enriched from 300  $\mu\text{g}$  proteome into seven fractions, sample complexity was largely decreased and a total of 696 C-termini were identified in duplicates from strict data filtration, that is, percolator false discovery rate (FDR) < 0.01, ion score  $\geq 20$ , and C-terminal amidation by ethanolamine. If only peptide FDR < 0.01 was considered, the identified C-termini further increased to 933, which was among the largest C-terminome datasets obtained from the polymer-based strategy. Furthermore, compared with the results of a previous study, the optimized method would be a practical strategy for broader C-terminome coverage.

Finally, to further broaden the coverage of the sub-C-terminome generated by Arg-specific cleavage, this study explored a new method in which ArgN-specific cleavage (cleavage at the N-terminal of Arg by LysargiNase) was combined with different N-terminal protections (dimethylation and acetylation). Among all the combinations, the additional use of the “LysargiNase + N-terminal acetylation” method increased 47% more identifications of unique C-termini on the basis of “trypsin + N-terminal demethylation” and the two covered 87% of the total C-termini. Therefore, the parallel use of the two methods would further expand the coverage of Arg-cleaved C-terminal peptides. With the analysis of the physicochemical properties of the peptides identified by the two methods, the reason why the C-terminal peptides identified by different strategies are complementary was explained.

In conclusion, in this study, the optimized C-terminomics platform can deeply profile Arg cleavage-generated C-terminal peptides using a polymer-based approach. This method provides a powerful tool for the global characterization of protein C-termini.

**Key words:** chemical acetylation; C-terminomics; negative enrichment; StageTip cartridge; biological mass spectrometry

引用本文: 赵晓晓, 胡昊, 赵雯思, 刘萍, 谭敏佳. 基于精氨酸酶切的蛋白质 C 端肽段富集方法的优化及评估. 色谱, 2022, 40(1): 17-27.

ZHAO Xiaoxiao, HU Hao, ZHAO Wensi, LIU Ping, TAN Minjia. Optimization and evaluation of protein C-terminal peptide enrichment strategy based on arginine cleavage. Chinese Journal of Chromatography, 2022, 40(1): 17-27.

蛋白质 C 端及其翻译后修饰 (PTMs, 如脂质化、乙酰化、磷酸化等) 参与多种生物学过程, 如蛋白质定位、蛋白质相互作用、维持蛋白质稳定性等<sup>[1-4]</sup>。蛋白质 C 端异常会引起代谢性疾病、神经退行性疾病、心血管疾病等<sup>[5-8]</sup> 多种疾病。因此, 蛋白质 C 端的系统性表征对于生物学机制的解析具有重要意义。

尽管“鸟枪法”蛋白质组学在全蛋白质组分析方面有很大优势, 但由于 C 端肽段占比少、离子化效率低, 蛋白质 C 端的分析效率仍比较低。近年来, 基于“鸟枪法”蛋白质组学发展形成的“蛋白质 C 端组学 (C-terminomics)” 技术使系统表征蛋白质 C 端成为可能。该技术主要有两种策略: 1) 通过对蛋白质 C 端羧基进行选择标记并进行亲和富集, 直接捕获蛋白质 C 端的正向富集策略<sup>[9-11]</sup>; 2) 通过化学衍生化和多种分离技术去除蛋白 N 端和内部肽段, 进一步洗脱获得蛋白质 C 端的反向富集策略<sup>[12, 13]</sup>。后者不仅可以富集普通 C 端肽段, 还可同时富集蛋白质 C 端  $\alpha$ -COOH 上含有 PTMs 的 C 端肽段。尤其是利用基于聚合物的反向富集策略<sup>[14]</sup> 进行蛋白质 C 端的研究, 可以实现多个样本平行操作, 大大提高了样本的制备通量, 并且不需专门的仪器设备, 在普通生化实验室即可完成, 应用范围更加广泛。

近年来, 研究人员基于聚合物富集策略, 发展了多种相关方法, 通过不断优化, 改善了 C 端肽段的鉴定深度<sup>[15-20]</sup>。Zhang 等<sup>[18]</sup> 通过胰蛋白酶 (trypsin) 切割精氨酸残基 C 端 (ArgC 型酶切) 的方式, 获得了 369 个蛋白质 C 端; Du 等<sup>[16]</sup> 通过重组赖氨酸肽链内切酶 (LysC) 切割赖氨酸残基 C 端 (LysC 型酶切) 的方式, 获得了 781 个蛋白质 C 端; 在本课题组前期 LAACter 策略研究<sup>[15]</sup> 中, 利用胰蛋白酶镜像酶 (LysargiNase) 切割赖氨酸和精氨酸残基 N 端 (LysN/ArgN 型酶切), 获得了 834 个蛋白质 C 端。这些方法使用不同的特异性酶切方式, 因此理论上对 C 端肽段有着各自独特的鉴定范围, 具有一定的互补性。而相比之下, 基于精氨酸切割 (Arg 型酶切) 的方法对 C 端肽段的鉴定深度仍有待提高。随着蛋白质 C 端在疾病中的作用越来越受到关注, 为了鉴定一些重要的蛋白质 C 端, 基于 Arg 型酶切的蛋白质 C 端组学方法在鉴定深度上亟待进一步突破。

为解决这一问题, 本研究对基于 Arg 型酶切方

法进行了优化和评估。建立了基于“V 型”过滤装置的“一锅法”富集平台, 并对蛋白水平乙酰化的反应条件进行了系统性的优化。提出了过量有机小分子可能是聚合物反向富集策略中干扰 C 端肽段鉴定的关键因素, 针对性地优化了基于固相萃取枪头膜片过滤柱 (StageTip) 的样品分离过程, 并和已报道的研究结果进行了对比和评估; 探索了使用不同酶切方式与不同肽段衍生化修饰的组合对蛋白质 C 端鉴定数目和种类的影响, 旨在利用“LysargiNase 酶切+肽段 N 端乙酰化”策略进一步扩大蛋白质 C 端的鉴定范围。

## 1 实验部分

### 1.1 仪器、试剂与材料

JY92-IIN 细胞粉碎机购于宁波新芝科技股份有限公司; 5424R 高速离心机购于德国 Eppendorf AG 公司; SPD111V-230 真空离心浓缩仪、Q-Exactive 质谱、Orbitrap Fusion 质谱、纳升液相色谱 EASY-nLC 1000 购于美国 Thermo Fisher Scientific 公司。

蛋白酶抑制剂购于瑞士 Roche 公司; *N*-羟基琥珀酰亚胺乙酸酯 (Ac-NHS) 和乙醇胺 (EA) 购于北京百灵威科技有限公司; LysargiNase 和胰蛋白酶购于北京华利世科技有限公司; 氯化钙 ( $\text{CaCl}_2$ )、盐酸胍 (GnHCl)、*N*-羟基琥珀酰亚胺 (NHS) 和 *N*-(3-(二甲基氨基)丙基)-乙基二亚胺盐酸 (EDC) 购于国药集团化学试剂有限公司; 4-(2-羟乙基) 哌嗪-1-乙磺酸 (HEPES)、2-(*N*-吗啉代) 乙磺酸 (MES)、氰基硼氢化钠 ( $\text{NaBH}_3\text{CN}$ )、二硫苏糖醇 (DTT)、碘乙酰胺 (IAA)、三氟乙酸 (TFA)、聚丙烯基胺 (PAA)、甲醛 (HCHO)、甲酸铵 (AF) 均购于美国 Sigma 公司。Durashell C18 填料购于天津博纳艾杰尔科技有限公司; Empore C18 固相萃取膜片购于美国 3M 公司; Amicon Ultra-0.5 filters (10 kDa 截止) 超滤管和 ZipTip C18 脱盐柱购于美国 Millipore 公司。

### 1.2 蛋白质提取与还原烷基化

将人 HEK 293T 细胞培养于含有 10% 胎牛血清、100 U/mL 青霉素和 100 mg/L 链霉素双抗的 DMEM 培养基中。向获得的湿细胞沉淀中加入 10 倍体积裂解液 (3 mol/L GnHCl、100 mmol/L HEPES、2% (质量分数) 蛋白酶抑制剂, pH 7.0), 冰上静置 30 min 后使用细胞破碎仪进行超声处理, 以 15 000 r/min 离心 10 min, 获取上清。按照 BCA 方法进行蛋白质浓度定量。加入终浓度为 5 mmol/L

的 DTT, 于 56 °C 反应 30 min, 最后加入终浓度为 15 mmol/L 的 IAA, 室温避光反应 30 min。

### 1.3 蛋白水平乙酰化

取 100 μg 还原烷基化后的样本, 用 3 mol/L GnHCl 和 100 mmol/L HEPES, pH 7.0 的缓冲液将样本稀释至 1 μg/μL。然后加入终浓度为 5 mmol/L 的 Ac-NHS (pH 7.0), 室温反应 30 min, 重复反应 3 次。再加入终浓度为 500 mmol/L 的氨水 (NH<sub>3</sub> · H<sub>2</sub>O), 室温反应 15 min。按蛋白质与酶的质量比为 50:1 添加胰蛋白酶, 调 pH 8.0, 于 37 °C 孵育 16 h。样品转干后, 取 10 μg 样本用 ZipTip C18 脱盐柱进行除盐。每组样本进行 2 次技术重复。

### 1.4 蛋白水平酰胺化

取 300 μg 还原烷基化后的样本, 置于 0.5 mL 超滤管中, 加入缓冲液 A (100 mmol/L HEPES, 3 mol/L GnHCl, pH 7.0) 至 500 μL, 在 11 600 r/min 的转速下超滤至 50 μL, 重复两次。使用缓冲液 A 调节样本质量浓度至 1 μg/μL, 加入终浓度为 5 mmol/L 的 Ac-NHS, 室温反应 30 min, 重复反应 3 次。再加入缓冲液 B (1 mol/L EA, 2 mol/L GnHCl, 0.2 mol/L MES, pH 6.0) 至 500 μL, 在 11 600 r/min 的转速下超滤至 50 μL。使用缓冲液 B 调节样本质量浓度至 1 μg/μL, 加入终浓度为 20 mmol/L 的 NHS 和 100 mmol/L 的 EDC, 于 37 °C 反应 2 h。再次补加相同终浓度的 EDC, 反应 2 h。

### 1.5 酶切消化

在上述样本中加入胰蛋白酶酶切体系 (20 mmol/L HEPES, 0.4 mol/L GnHCl, pH 8.0) 至 500 μL, 在 11 600 r/min 的转速下超滤至 50 μL, 重复 3 次。使用酶切体系调节样本质量浓度至 1 μg/μL, 按蛋白质与酶的质量比为 50:1 添加胰蛋白酶, 于 37 °C 孵育 16 h。再按蛋白质与酶的质量比为 100:1 添加胰蛋白酶, 于 37 °C 孵育 4 h。

在研究不同酶切方式和不同肽段衍生生化修饰的组合对 C 端肽段鉴定数目和种类影响的实验中, 引入 LysargiNase 代替胰蛋白酶进行酶切消化, 其酶切体系为 20 mmol/L HEPES 和 10 mmol/L CaCl<sub>2</sub>, pH 8.0, 其他条件与胰蛋白酶酶切处理方式相同。

### 1.6 肽段水平衍生生化修饰与蛋白质 C 端富集

对胰蛋白酶酶切消化产生的肽段分别进行乙酰化和二甲基化反应。LysargiNase 酶切后的样本进行同样的修饰反应。对于乙酰化修饰, 加入终浓度为 5 mmol/L 的 Ac-NHS, pH 7.0, 室温反应 30

min, 重复反应 3 次; 对于二甲基化修饰, 加入终浓度为 20 mmol/L 的 HCHO 和 10 mmol/L 的 NaBH<sub>3</sub>CN, 调节 pH 6.0, 于 37 °C 反应 2 h, 再次加入相同浓度的 HCHO 和 NaBH<sub>3</sub>CN, 反应 2 h。每组样本进行 2 次技术重复。

将反应完毕后的样本在 SPD111V-230 真空离心浓缩仪中浓缩至 150 μL, 各自加入 150 μL 2 mol/L PAA, 50 μL 乙腈以及终浓度为 20 mmol/L 的 NHS 和 100 mmol/L 的 EDC, 于 37 °C 反应 2 h, 补加相同终浓度的 EDC, 反应 2 h。在 11 600 r/min 的转速下离心收集滤液, 加入 15% (v/v) 乙腈水溶液至 300 μL, 相同转速下离心收集滤液, 重复两次。将滤液合并转干冻存。

### 1.7 基于 StageTip 的样品分离

配制碱性溶液体系: 溶液 A 为 98% (v/v) 乙腈水溶液 (含 5 mmol/L AF), 溶液 B 为 5 mmol/L AF 水溶液 (pH 10.0)。

碱性 StageTip 柱的制备: 将 Empore C18 固相萃取膜片置于 200 μL 枪头底部, 称取 6 mg Durashell C18 填料, 用 200 μL 溶液 B 混匀, 加入上述枪头内, 以 1 000 r/min 离心 5 min, 制成 StageTip。然后加入 200 μL 溶液 B, 以 1 000 r/min 离心 5 min 洗涤 StageTip 柱, 再加入溶液 A-溶液 B (1:1, v/v) 的混合液 200 μL, 以 1 000 r/min 离心 5 min 洗涤 StageTip 柱, 最后加入 200 μL 溶液 A, 以 1 000 r/min 离心 10 min 洗涤 StageTip 柱。

基于 StageTip 柱的样品分离: 将胰蛋白酶酶切后进行二甲基化修饰的样本用 200 μL 溶液 A 溶解后, 加入碱性 StageTip 柱, 以 1 000 r/min 离心 10 min, 使样本充分结合在 StageTip 柱中的 Durashell C18 填料上。最后使用 5%、10%、15%、20%、25%、30% 和 80% (v/v) 的乙腈水溶液 (均含 0.1% TFA) 各 200 μL 进行洗脱, 将收集的滤液转干后, 用 ZipTip C18 脱盐柱进行除盐。

### 1.8 液相色谱-质谱分析

#### 1.8.1 乙酰化优化实验

EASY-nLC 1000 高效液相色谱串联 Q-Exactive 用于乙酰化优化实验中的质谱检测。自制 C18 (粒径 3 μm, 孔径 9 nm, 美国 Dikma Technologies 公司) 毛细管柱 (100 mm × 75 μm)。流动相 A: 2% (v/v) 乙腈水溶液 (含 0.1% (v/v) FA); 流动相 B: 90% (v/v) 乙腈水溶液 (含 0.1% (v/v) FA); 流速: 300 nL/min。梯度洗脱程序: 0 ~ 50 min, 5% B ~



28% B; 50~53 min, 28% B~48% B; 53~56 min, 48% B~80% B; 56~60 min, 80% B。将 1.3 节中除盐转干后的样本溶于 10  $\mu$ L 流动相 A 中,以 15 000 r/min 高速离心,取 4.2  $\mu$ L 上清进行上样分析。

Q-Exactive 质谱参数为:一级分辨率为 70 000 ( $m/z$  200);一级自动增益控制 (AGC) 为  $1 \times 10^6$ ;二级自动增益控制为  $1 \times 10^6$ ;最大注射时间为 60 ms;扫描范围为  $m/z$  350~1 300;电荷状态 2~5 价;碰撞归一化能量为 30%;动态排除时间 60 s;数据依赖采集模式为“TOP N” ( $N=16$ )。

### 1.8.2 蛋白质 C 端富集实验

EASY-nLC 1000 高效液相色谱-串联 Orbitrap Fusion 用于蛋白质 C 端富集实验中的质谱检测。流速 300 nL/min。梯度洗脱程序:0~13 min, 5% B~7% B; 13~33 min, 7% B~10% B; 33~88 min, 10% B~25% B; 88~110 min, 25% B~45% B; 110~113 min, 45% B~80% B; 113~120 min, 80% B。将 1.7 节中除盐转干后的样本溶于 10  $\mu$ L 流动相 A 中,以 15 000 r/min 高速离心,取 4.2  $\mu$ L 上清进行上样分析。

Orbitrap Fusion 质谱参数为:一级分辨率为 120 000 ( $m/z$  200);一级自动增益控制 (AGC) 为  $5 \times 10^5$ ;二级自动增益控制为  $7 \times 10^3$ ;最大注射时间为 50 ms;扫描范围为  $m/z$  300~1 300;电荷状态 1~6 价;碰撞归一化能量为 32%;动态排除时间 60 s;数据依赖采集模式为“TOP speed”(循环时间 3 s)。

### 1.9 数据分析

质谱产生的原始数据由可以调用 Mascot 2.3.0 搜索引擎的 Proteome Discoverer 2.2 按照 UniProt (*homo sapiens*) 数据库 (版本 2018.7, 95 128 条序列) 进行搜库。

乙酰化优化实验中,统计乙酰化标记效率的搜库参数设置:酶切方式为胰蛋白酶,最大漏切位点为 3 个,固定修饰为 Cys 还原烷基化,可变修饰为 Lys/Ser/Thr/Tyr 乙酰化、蛋白质 N 端乙酰化、Met 氧化。统计蛋白质 C 端数目的搜库参数设置如下:酶切方式 ArgC,最大漏切位点为 2 个,固定修饰为 Cys 还原烷基化、Lys 乙酰化,可变修饰为蛋白质 N 端乙酰化、Met 氧化。其他参数均为一级离子质量偏差 20 ppm ( $10^{-6}$ ),二级碎片离子的质量偏差 0.02 Da。

蛋白质 C 端富集实验中,搜库参数设置如下:“胰蛋白酶酶切+二甲基化”策略酶切方式为 ArgC,“LysargiNase 酶切+二甲基化”策略酶切方式为

ArgN,两者的固定修饰均为 Met 还原烷基化、Asp/Glu 乙醇胺修饰、Lys 乙酰化、肽段 N 端二甲基化。“胰蛋白酶酶切+乙酰化”策略酶切方式为 ArgC,“LysargiNase 酶切+乙酰化”策略酶切方式为 ArgN,两者的固定修饰均为 Met 还原烷基化、Asp/Glu 乙醇胺修饰、Lys 乙酰化、肽段 N 端乙酰化。以上 4 种策略中,可变修饰均为蛋白质 N 端乙醇胺修饰、Met 氧化,最大漏切位点为 2 个,肽段一级离子的质量偏差设为 10 ppm,碎片离子的质量偏差设为 0.02 Da。

数据筛选标准:肽谱匹配、肽段及蛋白水平错误发现率 (FDR)  $< 0.01$ ,离子分数 (ion score)  $\geq 20$ ;乙酰化标记实验中,通过磷酸化修饰功能节点 (ptmRS)  $\geq 0.75$  的标准筛选总肽谱匹配 (peptide-spectrum matches, PSMs) 计算标记效率。

## 2 结果与讨论

### 2.1 基于“V 型”过滤装置的“一锅法”样品制备流程的建立

基于聚合物的蛋白质 C 端组学工作流程涉及多个化学衍生化步骤 (见图 1a),过量的化学试剂能

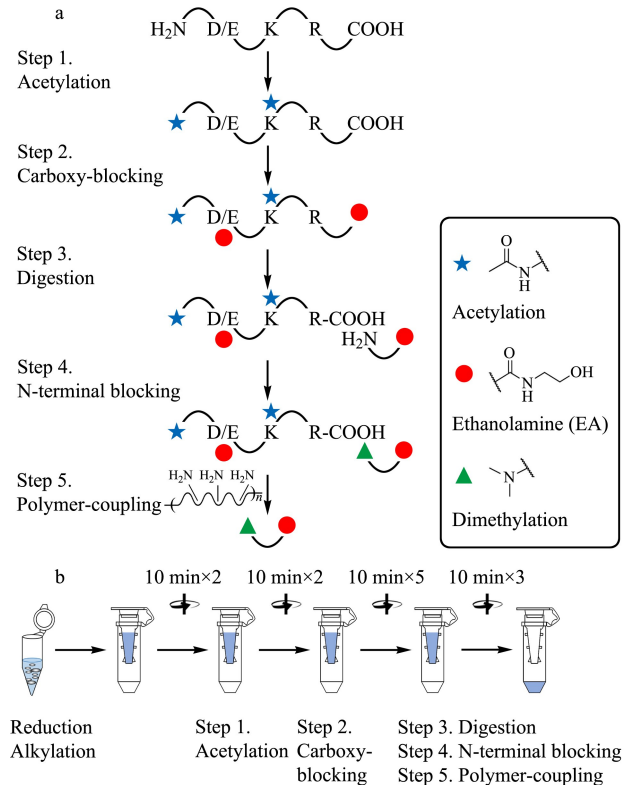


图 1 (a) 基于聚合物的反向富集流程及 (b) C 端肽段的“一锅法”富集平台

Fig. 1 (a) Workflow of polymer-based negative enrichment and (b) platform of one-pot enrichment of C-terminal peptides

够保证反应完全进行,但可能也会影响后续反应。为了避免反应之间的交叉干扰,通过对 Zhang 等<sup>[18]</sup>鉴定蛋白质 C 端实验流程的改进,建立了“一锅法”的富集平台(见图 1b),即在还原烷基化后立即应用超滤辅助样品制备方法(FASP<sup>[21]</sup>)来取代猝灭步骤,从而避免过量的猝灭剂(如二硫苏糖醇或半胱氨酸等)对下一步乙酰化反应的影响。

为了提高样本的制备效率,尝试使用滤膜面积更大的“V 型”过滤装置代替原本的平底过滤装置。当样本体积较大(如 500  $\mu\text{L}$ ),使用原本的平底过滤装置,单次溶剂交换过程需要 40~50 min,并且浓缩过程中产生的絮状物或沉淀容易造成滤膜堵塞。而使用“V 型”过滤装置,单次溶剂交换过程仅需约 10 min,从而将整个样品制备时间缩短了 4~5 h。并且滤膜位于装置侧壁,样本制备过程中不易发生堵塞。对于 200~400  $\mu\text{g}$  的蛋白质组,两种装置可以提供相似的回收率<sup>[21,22]</sup>。因此,以下整个研究过程中使用基于“V 型”过滤装置的“一锅法”样品制备流程。

## 2.2 乙酰化反应条件优化

蛋白质水平赖氨酸乙酰化反应是本研究重要的步骤之一,丝氨酸、苏氨酸、酪氨酸残基上的副反应会导致较高的样品复杂度和错误鉴定率。为了降低副反应的发生并提高肽段的鉴定数目,考虑从以下几个方面对乙酰化反应条件进行系统性优化:1)通过减少反应中 Ac-NHS 的用量以降低副反应的标记效率;2)根据以往的报道,蛋白水平乙酰化标记均是在 pH 8.0 的条件下进行,而碱性条件中组氨酸可能会促进相邻的羟基与 Ac-NHS 发生反应<sup>[23]</sup>,尝试调节反应的 pH 值,以减少副反应的干扰;3)通过在乙酰化标记后加入终浓度为 500 mmol/L 的  $\text{NH}_3 \cdot \text{H}_2\text{O}$  进行猝灭反应,在碱性较强的条件下逆转副反应的标记。基于以上思路,在进行乙酰化反应时,分别加入终浓度为 10 mmol/L 的 Ac-NHS (pH 8.0) (条件 A)、终浓度为 10 mmol/L 的 Ac-NHS (pH 7.0) (条件 B) 和终浓度为 5 mmol/L 的 Ac-NHS (pH 7.0) (条件 C)。结果表明,在条件 B 和条件 C 中,通过将反应 pH 调至 7.0,并使用  $\text{NH}_3 \cdot \text{H}_2\text{O}$  进行猝灭反应,能够在保持 Lys 较高标记效率(>98%) (见图 2a) 的同时,大幅降低副反应的发生(<4%) (见图 2b~2d)。由于条件 C 的 Ac-NHS 用量相对较少,因此后续实验选择条件 C 作为乙酰化修饰的条件。

此外,与以往报道<sup>[15,18,24]</sup> 中将 Ac-NHS 作为乙

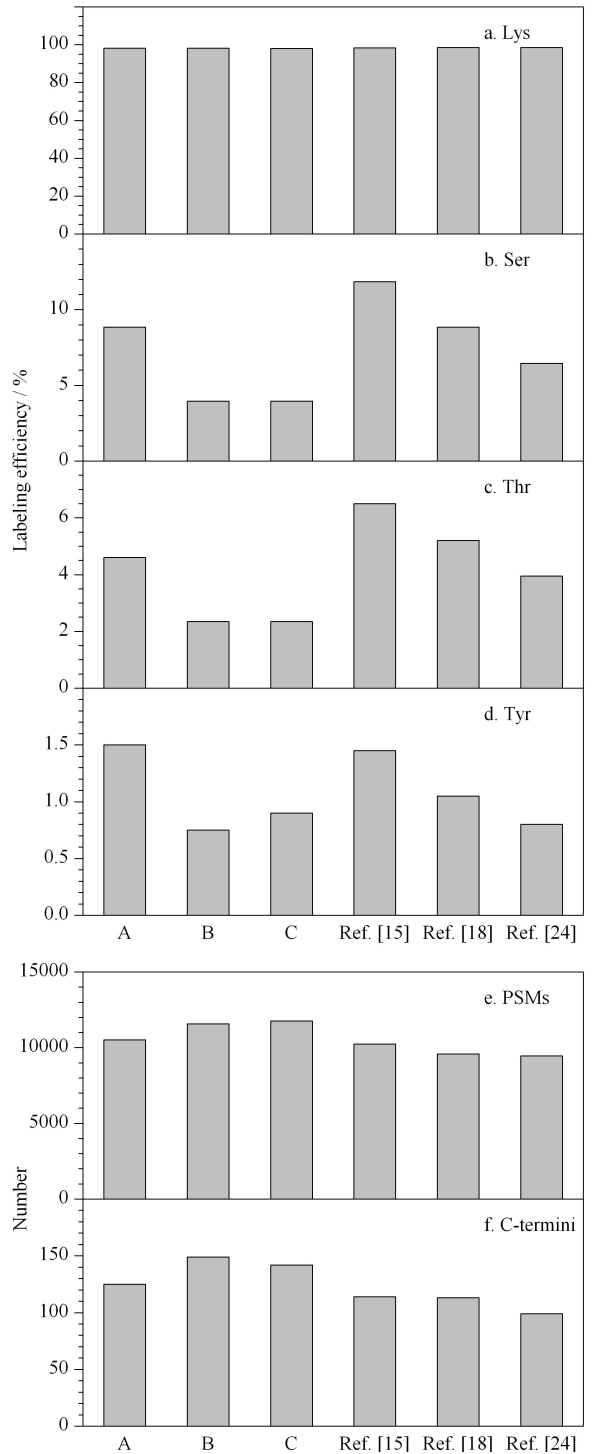


图 2 (a~d) 氨基酸残基上的乙酰化标记效率以及 (e) PSMs 和 (f) 蛋白质 C 端的鉴定数目

Fig. 2 (a~d) Acetylation labeling efficiencies on amino acid residues and numbers of identified (e) peptide-spectrum matches (PSMs) and (f) protein C-termini

A: 10 mmol/L Ac-NHS, pH 8.0; B: 10 mmol/L Ac-NHS, pH 7.0; C: 5 mmol/L Ac-NHS, pH 7.0. The experiments in entry Ref. [18], [15], and [24] were performed according to the conditions of the original study.

酰化修饰试剂的方案进行了比较(见图 2b~2d)。结果显示,LAACter 方法<sup>[15]</sup>和 Zhang 等<sup>[18]</sup>方法中乙酰化的副反应标记效率均较高,尤其是 Ser 的标记效率高达 8%~12%。Zhou 等<sup>[24]</sup>方法中乙酰化的副反应标记效率较低,但 PSMs 数目明显减少(见图 2e)。并且使用条件 C 进行乙酰化反应比 3 种已报道的方案多鉴定到约 15% 的 PSMs,比 Zhou 等<sup>[24]</sup>的方法多鉴定到约 50% 的蛋白质 C 端(见图 2f)。优化后的条件能够使副反应的标记效率降到最低,且不会影响肽段的鉴定。值得注意的是,这一优化不仅有助于 C 端肽段的可靠鉴定,而且有助于其他 Lys 酰化的研究,如 N 端组学<sup>[25,26]</sup>、乙酰化位点占据率计算<sup>[27,28]</sup>、全局组蛋白修饰分析<sup>[29,30]</sup>,以及使用胰蛋白酶进行 ArgC 型酶切的蛋白质组学研究<sup>[31,32]</sup>等。

### 2.3 样品分离过程的优化

为了提高富集样本中蛋白质 C 端的鉴定深度,在乙酰化优化结果的基础上,以 300  $\mu\text{g}$  蛋白质为起始量对基于 StageTip 样品分离的过程进行了优化。首先用 200  $\mu\text{L}$  含 0.1% TFA 的水溶液(pH 3.0)溶解富集后的样本,加至含 0.6 mg Durashell C18 填料的酸性 StageTip 柱(以 98% (v/v) 乙腈水溶液 A' 和 0.1% TFA 水溶液 B' 代替溶液 A、B 制备的 StageTip 柱)进行蛋白质 C 端的鉴定实验。利用 200  $\mu\text{L}$  20% (v/v) 乙腈水溶液(含 0.1% TFA)将肽段从 StageTip 柱中洗脱,在两次重复实验中,C 端肽段富集效率达 80% 和 73%,平均鉴定出 133 条 C 端肽段,对应于 113 个蛋白质 C 端(见图 3a 中 TFA-0.6 w/o a-ion)。通过添加 a 离子进行搜库(该方式已被证明可以提高 C 端肽段的鉴定数目<sup>[15,33]</sup>),蛋白质 C 端的鉴定数目有所增加(见图 3a 中 TFA-0.6),因此后续的数据分析中均考虑添加 a 离子进行搜库,但对比 LAACter<sup>[15]</sup>和 Du 等<sup>[16]</sup>的报道,以此方法鉴定的蛋白质 C 端数目仍然较少。

考虑到肽段水平反应中积累了大量有机物(包括乙酰化试剂、有机缓冲剂及其他副产物),尤其是用于肽段与聚合物偶联的 EDC 总量达 10~20 mg,这些有机物对肽段的鉴定可能存在多方面的影响,如:1)在脱盐和 nano-LC 分离过程中,与 Durashell C18 填料竞争性结合,从而导致样本的严重损失;2)碱性 EDC 会对肽段信号产生压制,从而影响检测灵敏度;3)在数据依赖的采集模式中通过占据 MS<sup>2</sup> 的扫描时间,从而减少对肽段的扫描。因此,需要通过

对样品分离过程进行优化以减轻这些非肽有机物造成的影响。

由于蛋白质 C 端的  $\alpha\text{-COOH}$  被酰胺化修饰,推测在碱性环境中进行样品分离,C 端肽段去质子化而极性变小,可能与 Durashell C18 填料有更高的亲和力,因此在洗脱的滤液中能够鉴定到更多的 C 端肽段。结果表明,在 1.7 节描述的碱性环境下进行样品分离,C 端肽段的鉴定数目约比酸性环境增加了 44%(见图 3a 中 AF-0.6 与 TFA-0.6),因此后续基于 StageTip 柱的样本分离过程均在碱性条件下进行。

为了实现肽段和非肽类物质更好的分离,同时提升蛋白质 C 端的鉴定数目,进一步优化了 StageTip 柱的 Durashell C18 填料用量。结果表明,相比于使用 0.6 mg Durashell C18 柱填料进行样品分离,在使用 6 mg 柱填料时,蛋白 C 端的鉴定数目没有明显减少(见图 3a, AF-6),且填料柱体积相对较大,更利于进行将样本洗脱为多个组分的实验操作;而使用 18 mg 柱填料时,大量填料可能由于非特异性吸附作用引起样品损失,蛋白质 C 端的鉴定数目约下降了 43%(见图 3a, AF-18),不利于后续进一步的优化实验。

因此,通过使用含 6 mg Durashell C18 填料的 StageTip 柱,将样本依次用 5%、10%、15%、20%、25%、30% 和 80% (v/v) 乙腈水溶液(均含 0.1% TFA)各 200  $\mu\text{L}$  洗脱为 7 个组分,约 93% 的蛋白质 C 端仅在一个或两个组分中鉴定到,表明将样本洗脱为多个组分后,降低了样品的复杂度,达到了比较好的分离效果,更加利于蛋白质 C 端的鉴定。两次重复实验中平均鉴定到 616 条 C 端肽段,对应 547 个蛋白质 C 端(见图 3b 中 7Fr),总数目约为酸性条件下的 4 倍。在两次重复实验中,约 73% 的蛋白质 C 端被鉴定到两次,皮尔森相关系数达 0.91,表明实验的重复性较好。考虑到蛋白质 C 端数目约占总 PSMs 的 1%(见图 2e 和 2f),以 300  $\mu\text{g}$  的蛋白质组为起始量进行样品分离过程的优化实验,估计 C 端肽段仅约为 1~3  $\mu\text{g}$ ,但将洗脱后的 7 个组分组合成 4 个组分(5%、10%+25%、15%+30%、20%+80%)后进行质谱分析,鉴定数目约下降了 30%(见图 3b 中 4Fr),证明了富集后样本的仍具有较高的复杂性及需要更高分离度的必要性。

综上所述,通过将 300  $\mu\text{g}$  蛋白质组富集的样品洗脱为 7 个组分,在两次重复实验中共鉴定出 732



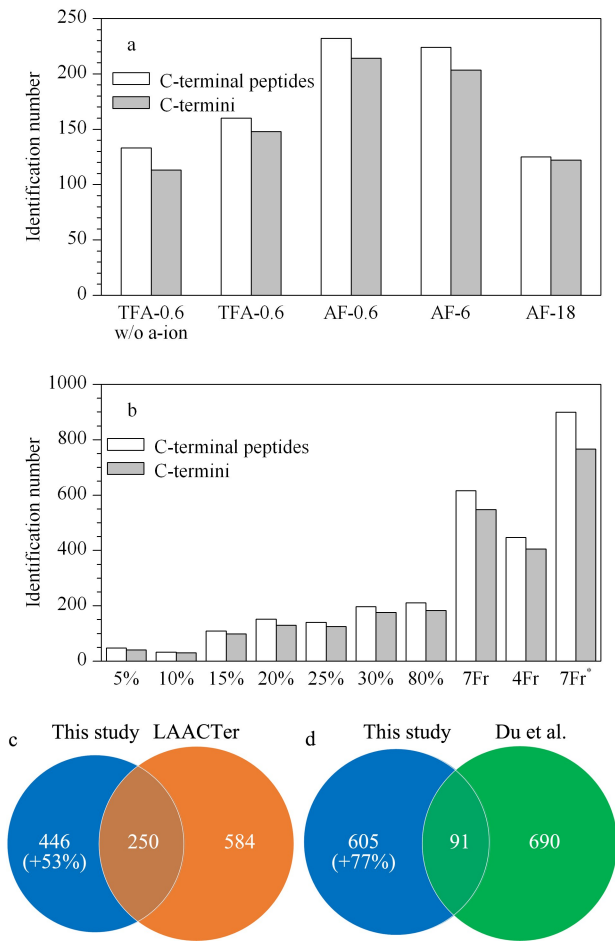


图 3 (a, b) 不同条件下 C 端肽段和蛋白质 C 端的鉴定数目以及 (c, d) 不同策略之间的互补性分析

Fig. 3 (a, b) Identification numbers of C-terminal peptides and (c, d) C-termini and complementary analysis of different strategies

a. TFA-0.6 w/o a-ion; 0.1% trifluoroacetic acid (TFA), 0.6 mg Durashell C18, without a-ion; TFA-6; 0.1% TFA, 0.6 mg Durashell C18, with a-ion; AF-0.6; 5 mmol/L ammonium formate (AF), 0.6 mg Durashell C18, with a-ion; AF-6; 5 mmol/L AF, 6 mg Durashell C18, with a-ion; AF-18; 5 mmol/L AF, 18 mg Durashell C18, with a-ion. b. 5%–80%; 5%–80% (v/v) acetonitrile aqueous solution (0.1% TFA); 7Fr: the above seven fractions; 4Fr: the seven fractions combined into four fractions (5%, 10%+25%, 15%+30%, 20%+80%); 7Fr\*: only false discovery rate (FDR)<0.01 was required for seven fractions. c. overlap of C-termini identified in this study and LAACter<sup>[15]</sup>; d. overlap of C-termini identified in this study and Du's study<sup>[16]</sup>.

条 C 端肽段, 对应 696 个蛋白质 C 端, 达到本课题组前期 LAACter 方法<sup>[15]</sup>和 Du 等<sup>[16]</sup>的鉴定深度 (分别鉴定到 834 和 781 个蛋白质 C 端)。

与 LAACter 策略<sup>[15]</sup> (切割 Lys/Arg 残基 N 端, 产生以 Lys 或 Arg 开头且  $\alpha$ -N 端带有乙酰化修饰的 C 端肽段) 的研究结果相比, 本研究的策略 (切割

Arg 残基 C 端, 产生  $\alpha$ -N 端带有二甲基化修饰的 C 端肽段) 鉴定到的蛋白质 C 端种类新增约 53%, 且重叠部分低于 20%, 两者总数目达 1 200 以上 (见图 3c); 与 Du 等<sup>[16]</sup>的策略 (切割 Lys 残基 C 端, 产生  $\alpha$ -N 端带有二甲基化修饰的 C 端肽段) 相比, 本研究鉴定到的蛋白质 C 端新增约 77%, 重叠部分低于 7%, 两者总数目达 1 380 以上 (见图 3d)。与 Du 等<sup>[16]</sup>的研究 (先将蛋白质组进行样品分离, 再对每个组分进行富集) 相比, 本研究中的策略是先进进行目的肽段富集再完成样品分离, 避免了样品损失, 大大简化了操作流程。不同策略中使用不同的蛋白质水平衍生生化修饰、酶切方式以及肽段 N 端衍生生化方式, 产生了含不同  $\alpha$ -N 末端结构的 C 端肽段, 因此理论上不同策略对蛋白质 C 端的鉴定有着各自独特的偏好性, 两两之间的鉴定结果有所不同, 这不仅实现了对蛋白质 C 端的深度鉴定, 而且在鉴定范围上具有独特性, 将为蛋白质 C 端机制的深入研究提供新的有力工具。

值得注意的是, 696 个蛋白质 C 端是经过严格的数据筛选得到, 即 1) 肽段水平 FDR<0.01; 2) ion score  $\geq 20$ ; 3) C 端带有乙醇胺修饰。若按 Du 等<sup>[16]</sup>所报道的 FDR<0.01 的置信标准, 两个重复实验中平均鉴定到 766 个蛋白质 C 端, 总数目可达 933 个 (见图 3b, 7Fr\*), 这是基于聚合物富集 C 端肽段策略获得的最大数据集之一<sup>[15–20]</sup>。

## 2.4 利用新型 ArgN 型酶切与肽段 N 端衍生生化进一步扩大蛋白质 C 端鉴定范围

LysargiNase 是一种可以特异性切割 Arg 和 Lys 残基 N 端的胰蛋白酶镜像酶<sup>[34]</sup>, 可以在酶切所得的肽段 N 端留下两种碱性氨基酸。LysargiNase 与胰蛋白酶在全蛋白质组肽段的鉴定覆盖范围具有一定的互补性<sup>[34,35]</sup>, 但二者在鉴定 Arg 特异性酶切所产生的 C 端肽段上的互补性鲜有研究。同时, 考虑到肽段 N 端的化学保护基会影响肽段的色谱保留和质谱检测, 推测 C 端肽段 N 端上不同的衍生生化修饰在进行蛋白质 C 端鉴定时可能会存在一定的互补性。

因此, 为了进一步扩大蛋白质 C 端鉴定范围, 本研究探索了两种酶 (LysargiNase 和胰蛋白酶) 和两种肽段 N 端衍生生化修饰的组合对 C 端肽段鉴定数目和种类的影响。在两次重复实验中, 通过对未分级分离的 C 端肽段样品进行单针质谱分析, 即仅使用 200  $\mu$ L 20% (v/v) 乙腈水溶液 (含 0.1% TFA)



将富集后的样品从 StageTip 柱洗脱。在两次重复实验中, 4 种策略共鉴定到 372 个蛋白质 C 端。其中“胰蛋白酶+二甲甲基化”策略鉴定到的蛋白质 C 端

数目共计 220 个(见图 4a 和 4b)。新型“LysargiNase+乙酰化”策略在前者策略的基础上, 鉴定到的蛋白质 C 端种类新增 105 个(见图 4b), 两种策略鉴

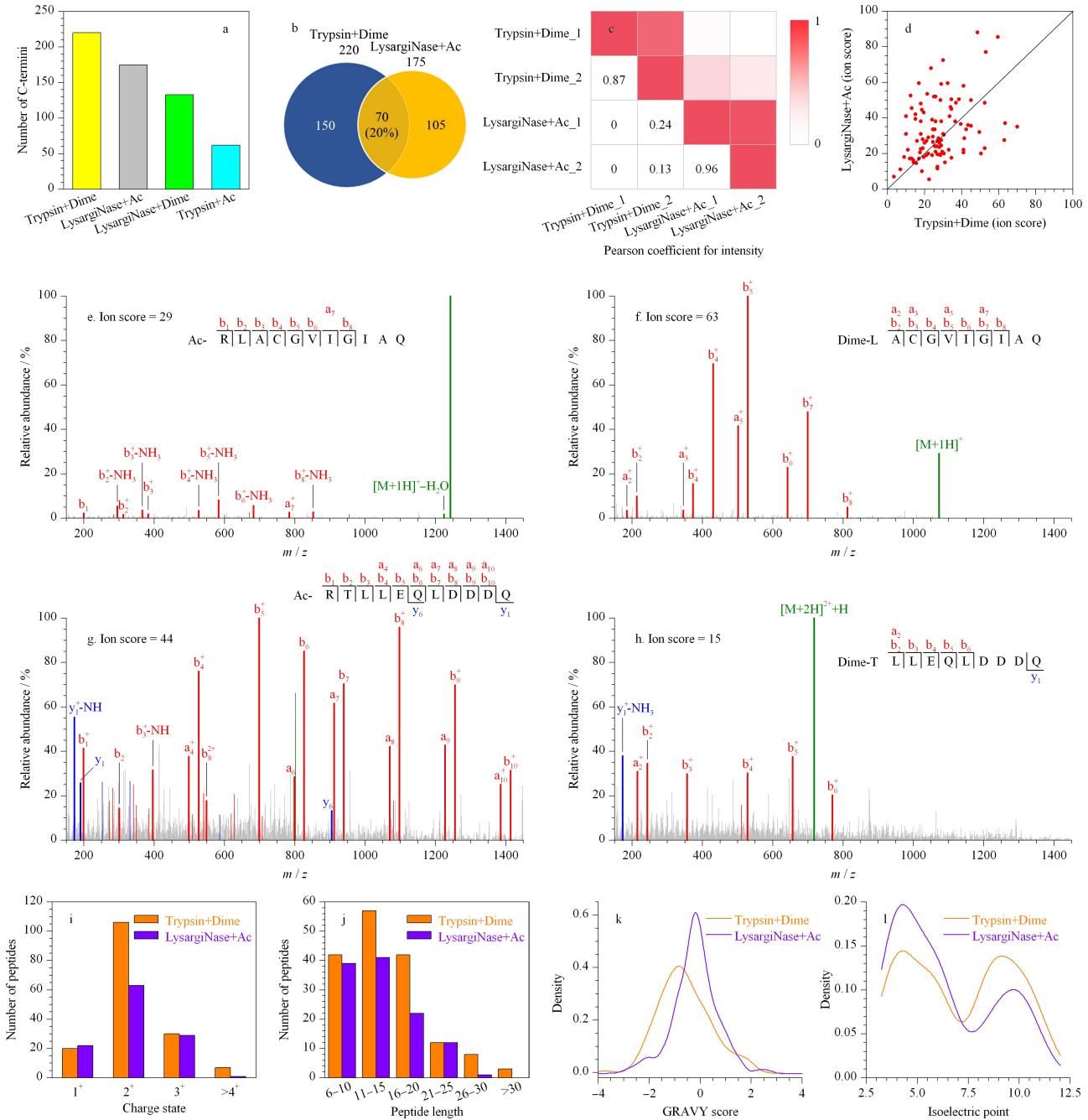


图 4 不同蛋白酶和 N 端衍生化组合对鉴定蛋白质 C 端的互补性分析

**Fig. 4 Complementarity of different combinations of proteases and N-terminal modifications on C-terminal identification**

a. identified number of C-termini by two repetitions under different strategies; b. overlap of totally identified C-termini; c. correlation of intensity between different experiments; d. comparison of ion scores for peptides commonly identified from indicated methods; e. the MS<sup>2</sup> spectra of RLCGVIGAIQ identified by “LysargiNase+acetylation (Ac)” ; f. the MS<sup>2</sup> spectra of LACGVIGAIQ identified by “trypsin+dimethylation (Dime)” ; g. the MS<sup>2</sup> spectra of RTLLEQLDDQ identified by “LysargiNase+Ac” ; h. the MS<sup>2</sup> spectra of TLLEQLDDQ identified by “trypsin+Dime” ; i. frequency of charge states of the C-terminal peptides; j. frequency of length of the C-terminal peptides; k. frequency of GRAVY score (hydrophilicity is represented by negative value and hydrophobicity is represented by positive value) of the C-terminal peptides; l. frequency of isoelectric point of the C-terminal peptides.

定的蛋白质 C 端数目占总数目的 87%,且两种策略鉴定蛋白质 C 端种类的重叠部分仅为 20% (见图 4b),说明两种策略对于蛋白质 C 端的鉴定范围具有很强的互补性。

对两种策略下鉴定到的 C 端肽段进行对比分析,发现共有肽段的强度相关性较差(见图 4c),且离子分数差异较大(见图 4d),表明两种方法对序列相同的 C 端肽段具有不同的偏好性。这种偏好性可归因于两种方法产生了不同的 N 末端结构:乙酰化的  $\alpha$ -N 端(由“LysargiNase+乙酰化”策略产生)和二甲基化的  $\alpha$ -N 端(由“胰蛋白酶+二甲基化”策略产生)。一方面,“LysargiNase+乙酰化”策略中,Arg 上胍基强碱性的优势在于较高的离子化效率利于肽段的检测<sup>[36]</sup>;但该基团的局限可能在于碱性太强而更难碎裂<sup>[37,38]</sup>,从而影响 C 端肽段的鉴定(见图 4e 和 4f)。另一方面,“LysargiNase+乙酰化”策略中  $\alpha$ -N 端上的乙酰化修饰通过增强肽段主链解离和稳定 N 端碎片离子<sup>[39,40]</sup>,能产生更多的 a 离子和 b 离子(见图 4g 和 4h)。此外,乙酰化的  $\alpha$ -N 端 Arg 基团的疏水性远低于二甲基化的  $\alpha$ -N 端(XLOGP3<sup>[41]</sup>值分别为-2.84 和-0.2),这在很大程度上影响了肽段的色谱保留,从而影响了蛋白质 C 端的鉴定。

为了深入了解两种策略对鉴定蛋白质 C 端的偏好性,分析了两种策略各自“特有的”C 端肽段的特征。结果显示,“胰蛋白酶+二甲基化”策略更倾向于鉴定长度为 11~15 的 2<sup>+</sup>肽段(见图 4i 和 4j)。而“LysargiNase+乙酰化”策略更倾向于鉴定长度为 6~15 的 2<sup>+</sup>肽段,以及疏水性较低且含有较多酸性残基的肽段(见图 4k 和 4l)。

综上所述,通过使用基于 Arg 型酶切的新策略“LysargiNase+乙酰化”,在原有策略“胰蛋白酶+二甲基化”的基础上进一步扩大了蛋白质 C 端的覆盖范围。

### 3 结论

本研究通过对基于 Arg 型酶切的反向富集策略的系统性优化,缩短了样本准备时间,提升了 C 端肽段的鉴定深度,并扩大了 C 端肽段的鉴定范围。此外,通过与其他鉴定 C 端肽段的策略相比,本研究在鉴定范围上具有独特性,进一步提示,可以采用多种策略联用的方式进行蛋白质 C 端的鉴定,用以提升蛋白质 C 端的鉴定“深度”和“广度”。通过将

优化后的方法与其他策略结合,不仅可以高效地用于基于 Arg 型酶切的蛋白质 C 端组学深度分析,还可以对不同细胞、组织来源的蛋白质 C 端状态进行系统表征,获得蛋白质 C 端状态的全景,将为蛋白质 C 端在生理病理中的分子机制研究提供新的信息。

### 参考文献:

- [1] Koren I, Timms R T, Kula T, et al. *Cell*, 2018, 173(7): 1622
- [2] Yang X, Coulombe-Huntington J, Kang S, et al. *Cell*, 2016, 164(4): 805
- [3] Sharma S, Schiller M R. *Crit Rev Biochem Mol Biol*, 2019, 54(2): 85
- [4] Chen B E, Sun Y, Niu J X, et al. *Cell Chem Biol*, 2018, 25(7): 817
- [5] Chen Y C, Taylor A J, Verchere C B. *Diabetes Obes Metab*, 2018, 20(S2): 64
- [6] O'Brien R J, Wong P C. *Annu Rev Neurosci*, 2011, 34: 185
- [7] Hill A, DeZern A E, Kinoshita T, et al. *Nat Rev Dis Primers*, 2017, 3: 17028
- [8] Klein T, Eckhard U, Dufour A, et al. *Chem Rev*, 2018, 118(3): 1137
- [9] Niedermaier S, Huesgen P E. *Biochim Biophys Acta Proteins Proteom*, 2019, 1867(12): 140138
- [10] Tanco S, Gevaert K, Van Damme P. *Proteomics*, 2015, 15(5/6): 903
- [11] Marino G, Eckhard U, Overall C M. *ACS Chem Biol*, 2015, 10(8): 1754
- [12] Xu G, Shin S B, Jaffrey S R. *ACS Chem Biol*, 2011, 6(10): 1015
- [13] Liu M B, Fang C Y, Pan X W, et al. *Anal Chem*, 2015, 87(19): 9916
- [14] Schilling O, Barré O, Huesgen P E, et al. *Nat Methods*, 2010, 7: 508
- [15] Hu H, Zhao W S, Zhu M D, et al. *Anal Chem*, 2019, 91(22): 14522
- [16] Du X X, Fang C Y, Yang L G, et al. *Anal Chem*, 2019, 91(10): 6498
- [17] Zhang Y, Li Q Q, Huang J N, et al. *Proteomics*, 2018, 18(1): 1700034
- [18] Zhang Y, He Q Z, Ye J Y, et al. *Anal Chem*, 2015, 87(20): 10354
- [19] Chen L F, Shan Y C, Weng Y J, et al. *Anal Bioanal Chem*, 2016, 408(14): 3867
- [20] Somasundaram P, Koudelka T, Linke D, et al. *J Proteome Res*, 2016, 15(4): 1369
- [21] Wisniewski J R. *Anal Chem*, 2016, 88(10): 5438
- [22] Lipecka J, Chhuon C, Bourderioux M, et al. *Proteomics*, 2016, 16(13): 1852
- [23] Wiktorowicz J E, English R D, Wu Z, et al. *J Proteomics Res*, 2012, 11(3): 1512
- [24] Zhou T, Chung Y H, Chen J J, et al. *J Proteomics Res*, 2016, 15(3): 1103

- [25] Prudova A, Gocheva V, Auf Dem Keller U, et al. *Cell Rep*, 2016, 16(6): 1762
- [26] Klein T, Fung S Y, Renner F, et al. *Nat Commun*, 2015, 6: 8777
- [27] Li Y N, Evers J, Luo A, et al. *Angew Chem Int Ed Engl*, 2019, 58(2): 537
- [28] Hansen B K, Gupta R, Baldus L, et al. *Nat Commun*, 2019, 10(1): 1055
- [29] Liao R J, Wu H P, Deng H B, et al. *Anal Chem*, 2013, 85(4): 2253
- [30] Chen J J, Gao J L, Peng M, et al. *Anal Chim Acta*, 2015, 886: 107
- [31] Golghalyani V, Neupärtl M, Witting I, et al. *J Proteome Res*, 2017, 16(2): 978
- [32] Schröder C U, Moore S, Goodarzi A A, et al. *Anal Chem*, 2018, 90(15): 9077
- [33] Li Q Q, Zhang Y, Huang J N, et al. *Anal Chem*, 2020, 92(7): 4742
- [34] Huesgen P F, Lange P F, Rogers L D, et al. *Nat Methods*, 2015, 12(1): 55
- [35] Tsiatsiani L, Giansanti P, Scheltema R A, et al. *J Proteome Res*, 2017, 16(2): 852
- [36] Tabb D L, Huang Y, Wysocki V H, et al. *Anal Chem*, 2004, 76(5): 1243
- [37] Carr S R, Cassady C J. *J Mass Spectrom*, 1997, 32(9): 959
- [38] Paizs B, Suhai S. *Mass Spectrom Rev*, 2005, 24(4): 508
- [39] Harrison A G, *Mass Spectrom Rev*, 2009, 28(4): 640
- [40] Waldera-Lupa D M, Stefanski A, Meyer H E, et al. *Biochim Biophys Acta*, 2013, 1834(12): 2843
- [41] Cheng T J, Zhao Y, Lin X, et al. *J Chem Inf Model*, 2007, 47(6): 2140