

一种基于疏水基团标记和反相色谱分离的富集策略 及其在含赖氨酸多肽分析中的应用

何宇¹, 单亦初^{2*}, 张丽华², 张振宾¹, 李洋^{1,2*}

(1. 宁波大学新药技术研究院, 浙江 宁波 315211;

2. 中国科学院大连化学物理研究所, 中国科学院分离分析化学重点实验室, 辽宁 大连 116023)

摘要: 赖氨酸(K)已被广泛用于靶向赖氨酸的交联剂设计、蛋白质复合物的结构解析以及蛋白质-蛋白质相互作用等研究领域。在基于液相色谱-串联质谱(LC-MS/MS)联用技术的“鸟枪法”蛋白质组学研究中,复杂生物样品中的蛋白质被酶切成上万条肽段,为直接分析含有K的肽段带来了巨大挑战。鉴于目前缺乏针对含K多肽的有效富集方法,本工作发展了一种基于疏水标记试剂C10-S-S-NHS和反相色谱分离的方法(简称HYTARP),实现对复杂样品中含K多肽的高效富集和鉴定。合成的C10-S-S-NHS试剂可以高效标记含不同数目K的标准肽段,且对HeLa细胞蛋白酶解肽段的标记效率高达96%。通过考察标记肽段在反相色谱中的保留行为,发现大部分被标记的含K肽段在流动相中乙腈比例升高至约57.6%(v/v)时开始洗脱。进一步优化色谱洗脱梯度,发现阶梯式洗脱能够实现对复杂样品酶解肽段中标记的K肽段的高效分离和富集。富集后样品中含K肽段的占比>90%,较富集前提高了35%。富集的含K肽段对应蛋白质的丰度动态范围跨越了5~6个数量级,实现了对复杂样品中低丰度蛋白质的鉴定。综上,本工作发展的HYTARP策略为降低样品复杂度、提高含K肽段及低丰度蛋白质的鉴定覆盖率提供了一种简单、高效的方法。

关键词: 疏水基团标记;富集;液相色谱-串联质谱;含赖氨酸多肽

中图分类号:O658

文献标识码:A

An enrichment strategy based on hydrophobic tagging and reversed-phase chromatographic separation for the analysis of lysine-containing peptides

HE Yu¹, SHAN Yichu^{2*}, ZHANG Lihua², ZHANG Zhenbin¹, LI Yang^{1,2*}

(1. Institute of Drug Discovery Technology, Ningbo University, Ningbo 315211, China;

2. Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China)

Abstract: Lysine (K) is widely used in the design of lysine-targeted crosslinkers, structural elucidation of protein complexes, and analysis of protein-protein interactions. In “shotgun” proteomics, which is based on liquid chromatography-tandem mass spectrometry (LC-MS/MS), proteins from complex samples are enzymatically digested, generating thousands of peptides and presenting significant challenges for the direct analysis of K-containing peptides. In view of the lack of effective methods for the enrichment of K-containing peptides, this work developed a method which based on a hydrophobic-tag-labeling reagent C10-S-S-NHS and reversed-phase chromatography (termed as HYTARP) to achieve the efficient enrichment and identification of K-containing peptides from complex samples. The C10-S-S-NHS synthesized in this work successfully labeled standard peptides containing various numbers of K and the labeling efficiency achieved up to 96% for HeLa cell protein tryptic digests. By investigating the

收稿日期:2024-02-22

* 通讯联系人.E-mail:shanyichu@163.com(单亦初);E-mail:liyong@nbu.edu.cn(李洋).

基金项目:宁波市顶尖人才项目(215-432094250);宁波大学王宽诚幸福基金.

Foundation item: Ningbo Top Talent Project (No. 215-432094250); the K. C. Wong Magna Fund in Ningbo University.

retention behavior of these labeled peptides in C18 RP column, we found that most K-labeled peptides were eluted once when acetonitrile percentage reached 57.6% (v/v). Further optimization of the elution gradient enabled the efficient separation and enrichment of the K-labeled peptides in HeLa digests via a stepwise elution gradient. The K-labeled peptides accounted for 90% in the enriched peptides, representing an improvement of 35% compared with the number of peptides without the enrichment. The dynamic range of proteins quantified from the enriched K-containing peptides spans 5–6 orders of magnitude, and realized the detection of low-abundance proteins in the complex sample. In summary, the HYTARP strategy offers a straightforward and effective approach for reducing sample complexity and improving the identification coverage of K-containing peptides and low-abundance proteins.

Key words: hydrophobic tagging; enrichment; liquid chromatography-tandem mass spectrometry (LC-MS/MS); lysine-containing peptides

在蛋白质组学研究领域,由 Yates 提出的“鸟枪法”(shotgun)已被广泛用于蛋白质组的定性定量、蛋白质的翻译后修饰以及蛋白质-蛋白质相互作用等分析中^[1-5]。该方法首先将蛋白质混合物酶解成多肽,再利用液相色谱-串联质谱(LC-MS/MS)技术对多肽进行分离和鉴定。然而,复杂样品中的蛋白质丰度是高度动态的(高达 9 个数量级),蛋白质酶解成肽段后进一步增加了样品的复杂程度,高丰度蛋白质/肽段会严重影响对低丰度蛋白质/肽段的鉴定^[6]。因此,对复杂样品中的蛋白质/肽段直接进行 LC-MS/MS 分析难以实现对蛋白质组的深度覆盖。

为了解决这个问题,对蛋白质酶解肽段中含有某类特征氨基酸残基的多肽进行富集和分析是降低样品复杂度、提高蛋白质组鉴定覆盖度的有效手段。目前已经发展了针对多种氨基酸残基的多肽富集策略,如半胱氨酸^[7-9]、赖氨酸(K)^[10,11]、色氨酸^[12,13]、甲硫氨酸^[14]、组氨酸^[15]以及 N 端丝氨酸、苏氨酸、脯氨酸、甘氨酸^[16-18]等。这些方法主要利用可以特征标记目标氨基酸残基的活性基团,并将其共价键合于固相功能材料,从而实现对含特征氨基酸蛋白质/多肽的选择性富集。由于大部分蛋白质(超过 96%)含有半胱氨酸残基,Li 等^[7]开发了基于固相烷基化硅胶材料的蛋白质组样品制备方法,该方法对含有半胱氨酸残基蛋白质的富集效率高达 93%,具有较高的通量和抗基质干扰能力,提高了临床蛋白质样品的回收率。Qi 等^[10]发展了一种固定化蛋白

质的酶解方法,利用可以选择性标记氨基的三氟代功能化微球对蛋白质进行快速固定和高效酶解,蛋白质回收率>90%,显著提高了复杂样品中肽段的鉴定数目和蛋白质序列覆盖率。类似地,Shah 等^[11]将可与氨基反应的醛基基团键合于固相材料,并用于福尔马林固定和石蜡包埋组织样品的蛋白质固定、干扰物去除和蛋白质酶解,蛋白质回收率高达 95%。为了降低样品复杂度,Li 等^[17]开发了一种选择性富集 N 末端脯氨酸肽段的方法,首先采用邻苯二甲醛封闭酶解肽段的伯胺基团,再采用戊二醛在还原环境下标记肽段 N 末端的脯氨酸残基——仲胺,利用固相酰肼微球选择性富集标记后的 N 末端脯氨酸肽段,该方法对 N 末端脯氨酸肽段的富集效率高达 93.7%,提高了复杂样品中低丰度蛋白质的鉴定。以上方法均采用了功能化材料富集具有特定氨基酸残基的蛋白质/肽段样品,然而由于蛋白质复杂的空间结构引入的空间位阻效应和固-液反应效率低等原因,样品中的蛋白质难以被完全捕获。此外,功能化材料表面存在大量的非特异性吸附位点,导致样品损失和引入杂质。

K 由于在蛋白质中分布较广(约 6%),侧链氨基反应特异性高且广泛分布在溶剂可接触的蛋白质表面,已被广泛用于 K 靶向的交联剂设计、蛋白质复合物的结构解析以及蛋白质-蛋白质相互作用等研究领域^[19-21]。能够与 K 侧链的伯胺发生反应的试剂众多,包括卤化物^[10]、醛类^[11]、亚胺酸酯类^[22]、

引用本文:何宇,单亦初,张丽华,张振宾,李洋.一种基于疏水基团标记和反相色谱分离的富集策略及其在含赖氨酸多肽分析中的应用.色谱,2024,42(7):721-729.

HE Yu, SHAN Yichu, ZHANG Lihua, ZHANG Zhenbin, LI Yang. An enrichment strategy based on hydrophobic tagging and reversed-phase chromatographic separation for the analysis of lysine-containing peptides. Chinese Journal of Chromatography, 2024, 42(7): 721-729.

胍基类^[23]和琥珀酰化类^[24]等。其中,琥珀酰化类试剂由于反应特异性高、反应条件温和、生成物结构稳定而被广泛用于化学交联蛋白质复合物的结构及其相互作用研究^[25,26]。最常用的化学交联试剂,如双琥珀酰亚胺辛二酸酯(DSS)、二(磺基琥珀酰亚胺)辛二酸酯(BS³)和双琥珀酰亚胺亚砷(DS-SO)^[27],都是由N-羟基丁二酰亚胺(NHS)官能团组成的氨基靶向试剂。然而,针对含K蛋白质/肽段的选择性富集方法却鲜有报道。一方面由于蛋白质的K残基被大部分试剂标记后,由于空间位阻效应,胰蛋白酶无法酶切标记后的K位点,导致酶解肽段过长,不利于MS鉴定^[28];另一方面可能由于大部分氨基反应试剂可以同时标记蛋白质/肽段N末端的 α -氨基与K侧链的 ε -氨基,导致富集过程中含K多肽与其他不含K多肽难以分离。

针对目前功能化富集材料存在的不足以及缺乏可以有效富集含K多肽的方法,本工作发展了一种基于疏水基团标记和反相色谱分离的方法(hydrophobic tagging and reversed-phase chromatography separation,简称HYTARP),以实现复杂样品中含K多肽的高效富集和鉴定。首先利用合成的标记试剂2,5-二氧化吡咯烷-1-基-3-(癸基二硫代)丙酸酯(C10-S-S-NHS)对胰蛋白酶酶解肽段中的氨基(包括肽段N末端和K的侧链氨基)进行C10疏水基团标记。由于含K肽段比不含K肽段多标记至少一个C10烷基链,其在反相色谱中的保留将明显增强。通过优化反相色谱的洗脱梯度,实现了对含K肽段的选择性富集。再进一步断裂标记试剂中的二硫键,释放被标记肽段的C10烷基链,保证含K肽段在LC-MS/MS中的高效分离和鉴定。该方法在肽段水平标记含K多肽,避免了胰蛋白酶的漏切问题;并且利用具有高分辨率的反相色谱可以有效提高对含K多肽的富集选择性,同时避免了大量功能化材料的使用,提高了样品回收率。

1 实验部分

1.1 仪器、材料与试剂

用于反相色谱分离疏水标记多肽的LC-20AD泵和SPD-20A检测器构建的液相色谱仪购自日本岛津公司。Ascend 500 MHz核磁共振仪和Ultraflex III型基质辅助激光解吸电离飞行时间(MALDI-TOF)质谱仪购自德国Bruker公司。EASY-nLC 1000色谱仪与Q-Exactive质谱仪联用系统、Ulti-

Mate 3000 RSLC 色谱仪与Orbitrap Fusion Lumos Tribrid质谱仪联用系统购自美国Thermo Fisher公司;04714-50细胞超声破碎仪购自美国Cole-Parmer公司。用于测定蛋白质浓度的全波长酶标仪购自美国BioTek公司。用于nanoRPLC-ESI-MS/MS分析的熔融石英毛细管(150 μm i. d. \times 375 μm o. d.)购自美国Polymicro Technologies公司。Reprosil-Pur C18-AQ硅胶填料(1.9 μm , 12 nm)购自德国Dr. Maisch公司。Venusil XBP C18硅胶填料(5 μm , 12 nm)购自天津博纳艾杰尔公司。

尿素(urea)、蛋白酶抑制剂(Cocktail)、三(2-羧乙基)膦盐酸盐(TCEP)、碘乙酰胺(IAA)、4-(2-羧乙基)哌嗪-1-乙磺酸(HEPES)、甲酸(FA)、三氟乙酸(TFA)购自美国Sigma-Aldrich公司。磷酸盐缓冲液(1 \times PBS)、细胞培养用的MEM培养基、胎牛血清(FBS)和链霉素(10 mg/mL)/青霉素(10 kU/mL)购自美国Gibco公司。胰蛋白酶(trypsin)购自美国Promega公司。2,2'-二吡啶二硫醚和1-癸硫醇购自日本Tokyo Chemical Industry公司。3-巯基丙酸购自上海阿拉丁生化科技股份有限公司。NHS和N-(3-二甲氨基丙基)-N'-乙基碳二亚胺盐酸盐(EDC·HCl)购自北京百灵威科技有限公司。碳酸氢铵(NH₄HCO₃)、二氯甲烷(DCM)购自天津科密欧化学试剂有限公司。乙腈(ACN, HPLC级别)购自德国Merck公司。BCA蛋白质浓度测定试剂盒购自上海碧云天生物技术有限公司。所有实验用水均通过Milli-Q系统(美国Millipore公司)纯化。其他化学品和溶剂均为分析级。所有标准多肽购自上海强耀生物科技有限公司。

1.2 标记试剂C10-S-S-NHS的合成

标记试剂C10-S-S-NHS的合成步骤参考文献[29]。主要合成过程如下:(1)氮气保护下,将2,2'-二吡啶二硫醚(3.3 g, 15 mmol)溶于25 mL无水乙醇中,依次加入1.25 mL乙酸和3-巯基丙酸(1.06 g, 10 mmol),室温搅拌过夜。减压旋蒸去除有机溶剂后,用碱性氧化铝层析柱纯化产物,先用DCM-EtOH(体积比3:2)淋洗,再用DCM-EtOH-醋酸(体积比30:20:2)洗脱并冻干,得目标产物3-(2-吡啶二硫代)丙酸(2.1 g, 收率98%);(2)氮气保护下,将3-(2-吡啶二硫代)丙酸(2 g, 9.3 mmol)和1-癸硫醇(1.46 g, 8.4 mmol)加入23 mL甲醇中,室温搅拌1 h。减压旋蒸后,用硅胶层析柱纯化产物,先用DCM淋洗,再用DCM-乙酸乙酯(体积比

15:1)洗脱并冻干,得目标产物 1-癸烷基二硫代丙酸(淡黄色油状物,1.15 g,产率 49%);(3)氮气保护下,将 1-癸烷基二硫代丙酸(1.15 g, 4.1 mmol)、NHS (0.95 g, 8.3 mmol) 和 EDC·HCl (1.58 g, 8.3 mmol) 依次加入 18.4 mL DCM 中,室温搅拌过夜。饱和食盐水洗涤有机层后,用硫酸钠干燥。减压旋蒸后,用硅胶层析柱纯化产物,用 DCM-石油醚(体积比 3:2)洗脱并冻干,即得目标产物 C10-S-S-NHS(淡黄色固体,0.72 g,收率 46%)。

1.3 细胞培养

用含有 10% FBS 和 1% 青霉素/链霉素的 MEM 培养基复苏 HeLa 细胞。将复苏后的细胞置于 37 °C 恒温培养箱中,5% CO₂ 气体条件下培养,每隔 3 天传代一次。细胞用预冷的 1×PBS 清洗 3 次后,加入含 0.05% 胰蛋白酶的 EDTA 溶液于 37 °C 消化 5 min,使贴壁细胞从培养皿上脱离,再加入含有血清的培养基终止胰酶消化。收集细胞悬液,于 4 °C 以 500 g 离心 3 min 收集细胞。最后用预冷的 1×PBS 清洗细胞 3 次。

1.4 蛋白质提取和酶解

将 HeLa 细胞悬浮于由 8 mol/L 尿素和 1% (v/v) 蛋白酶抑制剂组成的细胞裂解液中(50 mmol/L NH₄HCO₃, pH 8.0),并在冰浴下超声破碎 60 s(80% 能量,超声频率为 5 s 开,5 s 关)。匀浆液在 4 °C 下以 16 000 g 离心 30 min,收集上清。采用 BCA 法测定蛋白质浓度。将提取的 HeLa 细胞蛋白质进行还原(10 mmol/L TCEP, 56 °C, 1 h)和烷基化(30 mmol/L IAA, 室温遮光, 30 min)。用 50 mmol/L NH₄HCO₃ 溶液(pH 8.0)将样品稀释 8 倍后,以 1:40(m/m)的酶-蛋白质比例加入胰蛋白酶,37 °C 反应 12 h。向溶液中加入终体积分数为 1% 的 TFA,终止酶解反应。将酶解产物上样到自制的 C18 捕集柱(10 mm×4.6 mm, Venusil XBP C18 硅胶填料)上,去除溶液中过量的盐。流动相 A: 2% (v/v) 乙腈水溶液(含 0.1% (v/v) TFA),流动相 B: 98% (v/v) 乙腈水溶液(含 0.1% (v/v) TFA),流速 1 mL/min。梯度设置为阶梯式洗脱程序:0~5 min, 2% B; 5.01~15 min, 80% B; 收集在 80% B 相中不保留的肽段洗脱液,即纯化后的 HeLa 酶解肽段,冷冻干燥。

1.5 疏水基团标记

分别将 10 μg 标准肽段 LVVSTQTALA、MIFVGIK 和 KSLSLSPGK 重溶于 10 μL 50 mmol/L

HEPES 缓冲液(pH 8.3)中,加入 40 μL 含 5 g/L C10-S-S-NHS 的 DMF 溶液,40 °C 反应 2 h。将 50 μg HeLa 酶解肽段重溶于 20 μL 50 mmol/L HEPES 缓冲液(pH 8.3)中,加入 80 μL 含 12.5 g/L C10-S-S-NHS 的 DMF 溶液,肽段的终质量浓度为 0.5 g/L,40 °C 反应 4 h。加入终浓度 50 mmol/L NH₄HCO₃ 终止标记反应。

1.6 反相色谱分离

采用自制的 XBP C18 分离柱(250 mm×4.6 mm, 5 μm)对疏水基团标记的 HeLa 酶解肽段进行分级。流速 1 mL/min,紫外检测波长为 214 nm。

采用线性洗脱程序时流动相 A 为 2% (v/v) 乙腈水溶液(含 0.1% (v/v) TFA),流动相 B 为 80% (v/v) 乙腈水溶液(含 0.1% (v/v) TFA),梯度设置如下:0~15 min, 60% B; 15~25 min, 60% B~70% B; 25~55 min, 70% B~80% B; 55~70 min, 80% B~95% B; 70~80 min, 95% B。

采用阶梯式洗脱程序时流动相 A 为 2% (v/v) 乙腈水溶液(含 0.1% (v/v) TFA),流动相 B 为 98% (v/v) 乙腈水溶液(含 0.1% (v/v) TFA),梯度设置如下:0~10 min, 32% B; 10.1~20 min, 60% B; 20.1~30 min, 95% B。依据保留时间,收集肽段流出液,并冷冻干燥。将各个肽段级分重溶于含 5 mmol/L TCEP 的 50 mmol/L NH₄HCO₃ 溶液中,56 °C 反应 1 h。

1.7 MALDI-TOF MS 分析

将 1 μL 疏水标记前后的标准肽段与 1 μL 的 α-氰基-4-羟基肉桂酸基质(CHCA, 7 g/L, 溶于含 0.1% TFA 的 60% ACN 溶液中)依次点于 MALDI 靶板上。采用固体激光 SmartBeam 技术(355 nm)及正离子反射模式进行 MS 分析。采用 FlexControl(3.0)采集 MS 数据,采用 FlexAnalysis(3.0)提取所有信噪比大于 3 的信号峰进行分析。

1.8 nanoRPLC-ESI-MS/MS 分析

采用 EASY-nLC 1000 色谱仪与 Q-Exactive 质谱仪联用系统对采用反相色谱线性梯度分离的肽段级分进行分析。将肽段级分上样于自制的 C18 捕集柱(30 mm×150 μm, 5 μm),再通过自制 C18 分离柱(150 mm×150 μm, 1.9 μm)进行分离,流速为 600 nL/min。使用流动相 A(2% (v/v) 乙腈水溶液(含 0.1% (v/v) FA)) 和 B(98% (v/v) 乙腈水溶液(含 0.1% (v/v) FA)) 建立 75 min 线性梯度(0~45 min, 6% B~23% B; 45~65 min, 23% B~40% B;

65~70 min, 40% B~80% B; 70~75 min, 80% B)。喷雾电压为 2.4 kV, 离子传输管温度为 250 ℃。在正离子模式下, 采用数据依赖模式 (DDA), MS 分辨率设为 70 000, 扫描范围为 m/z 300~1 800, 自动增益控制 (AGC) 设为 3.0×10^6 , 离子最大注入时间 (M.I.T.) 为 60 ms。选择电荷价态从 +2 到 +6 的母离子肽段, 并采用高能碰撞碎裂模式 (HCD) 对 20 个响应最强的母离子进行碎裂, 碰撞能量设为 28。MS² 的分辨率设为 17 500, AGC 设为 5.0×10^4 , M.I.T. 设为 60 ms, 动态排除窗口时间为 12 s。

采用 UltiMate 3000 RSLC 色谱仪与 Orbitrap Fusion Lumos Tribrid 质谱仪构建的纳升级系统对采用反相色谱阶梯式梯度分离的肽段级分进行分析。将肽段上样于 C18 捕集柱 (20 mm×75 μm, 5 μm), 再通过 C18 分离柱 (250 mm×75 μm, 1.9 μm) 进行分离, 流速为 350 nL/min。使用流动相 A (2% (v/v) 乙腈水溶液 (含 0.1% (v/v) FA)) 和 B (80% (v/v) 乙腈水溶液 (含 0.1% (v/v) FA)) 建立 60 min 线性梯度 (0~6 min, 2% B~5% B; 6~43 min, 5% B~18% B; 43~48 min, 18% B~22% B; 48~51 min, 22% B~35% B; 51~54 min, 35% B~95% B; 54~60 min, 95% B)。喷雾电压为 2.1 kV, 离子传输管温度为 320 ℃。在正离子模式下, 采用 DDA 模式每 3 s 完成一次全扫描循环。选择轨道阱 (OT) 采集 MS 数据, 分辨率设为 60 000。扫描范围为 m/z 350~1 800, AGC 设为 4.0×10^5 , M.I.T. 为 50 ms。选择电荷价态从 +1 到 +7 的母离子肽段, 并采用 HCD 模式对其进行碎裂, 碰撞能量设为 30。OT 采集 MS² 数据, 分辨率设为 15 000, AGC 设为

5.0×10^4 , M.I.T. 设为 30 ms, 动态排除窗口时间为 20 s。

1.9 数据分析

将质谱数据文件 (*.raw) 采用嵌入了 Mascot (版本 2.3.2) 搜索引擎的 Proteome Discoverer (PD) (版本 2.1.1.21) 软件进行二级质谱数据库搜索, 数据库为从 Uniprot 网站下载的人源蛋白质数据库 (共 20 360 个蛋白质条目)。检索参数设置如下: 胰蛋白酶完全酶切, 最多允许 2 个漏切位点; 固定修饰为半胱氨酸的还原烷基化, 可变修饰有 K 和肽段 N 末端的硫酰化, 蛋氨酸的氧化和蛋白质 N 末端的乙酰化; 一级母离子质量允许偏差为 10×10^{-6} (10 ppm), 二级碎片离子的质量偏差为 0.02 Da。检索结果控制蛋白质和肽段的假阳性率不大于 1%; 蛋白质鉴定至少包含一个唯一性肽段。采用 Max-LFQ 算法结合 MaxQuant (version 1.6.5.0) 搜索引擎用于蛋白质的无标记定量分析, 搜库条件与 PD 保持一致。

2 结果与讨论

2.1 基于 HYTARP 策略的富集流程

基于 HYTARP 策略富集含 K 多肽的流程如图 1 所示, 首先采用试剂 C10-S-S-NHS 对胰蛋白酶酶解肽段中 K 侧链的自由氨基进行疏水基团 C10 烷基链的标记, 同时肽段 N 末端的自由氨基也会被标记。采用反相色谱对标记后的肽段进行分离, 由于 C10 疏水链的引入, 酶解肽段在 C18 柱上的保留较标记前将整体增强。通过优化反相色谱的分离梯度, 实现对含 K 肽段的选择性富集。再通过二硫键

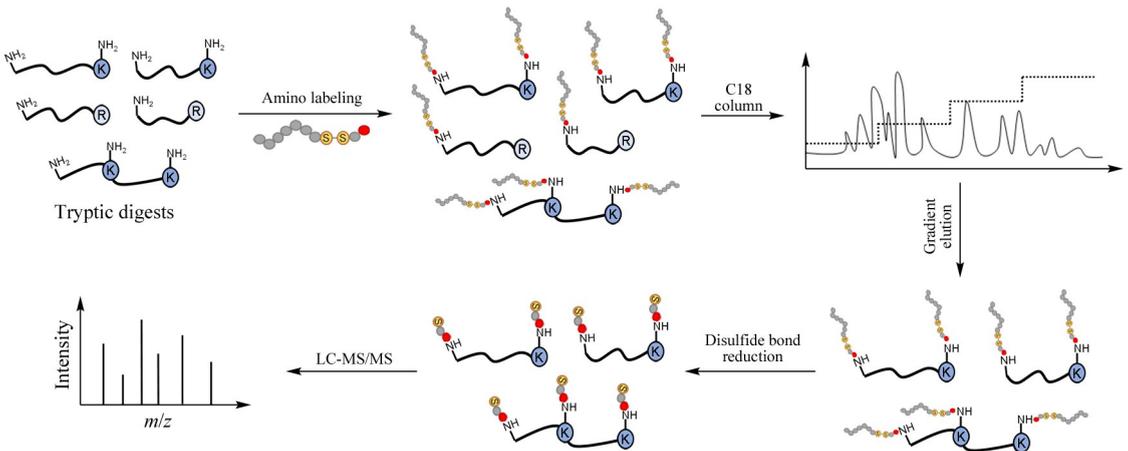


图 1 基于 HYTARP 策略的含 K 多肽的富集流程示意图

Fig. 1 Workflow of lysine-containing peptide enrichment using the hydrophobic tagging and reversed-phase chromatographic separation (HYTARP) strategy

还原反应, 释放 K 侧链和肽段 N 末端的疏水基团 C10 烷基链, 保证肽段的离子化效率, 从而提高 nanoLC-MS/MS 对含 K 肽段的分析和鉴定能力。

C10-S-S-NHS 的合成路线如图 2 所示 (C10-S-S-NHS 及中间产物的核磁共振氢谱表征见附图 1, www.chrom-China.com), 其主要包含 3 个功能基团: 可以高效且特异标记酶解肽段中游离氨基的 NHS、可以提高肽段疏水性的长碳链 (C10) 以及通过还原反应释放 C10 疏水链的二硫键 (S-S)。

2.2 标记效率考察

为了考察 C10-S-S-NHS 的标记效率, 选取 3 条含有不同数目 K 的标准肽段, 包括不含 K 的肽段 LA-10 (序列 LVVSTQTALA), 含有一个 K 的肽段

MK-7 (序列 MIFVGIK), 以及含有两个 K 的肽段 KK-9 (序列 KLSLSPGK)。采用 MALDI-TOF MS 对标记前后肽段的相对分子质量进行分析, 结果如图 3 所示, 标记后原始的 LV-10 肽段质谱峰 (m/z 1024.8, 图 3a) 消失, 同时产生新的肽段质谱峰 (m/z 1284.5, 图 3b), 增加了一个 C10 链的相对分子质量 ($\Delta m = 260$ Da), 表明 LV-10 肽段的 N 末端氨基被 C10-S-S-NHS 完全标记。同样地, 标记后的 MK-7 和 KK-9 的原始肽段 (m/z 807.7, 图 3c; m/z 916.7, 图 3e) 都几乎消失, 且新生成的肽段分别被标记上 2 个和 3 个 C10 链 (m/z 1350.2, 图 3d; m/z 1720.1, 图 3f, +Na 峰), 表明 C10-S-S-NHS 试剂可以高效标记肽段 N 末端和 K 侧链的自由氨基。

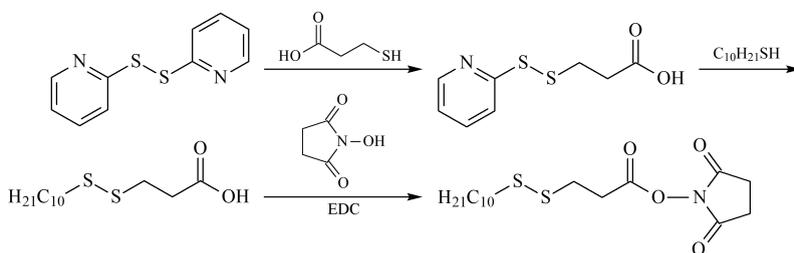


图 2 标记试剂 C10-S-S-NHS 的合成

Fig. 2 Synthesis of 2,5-dioxopyrrolidin-1-yl-3-(decyldisulfanyl) propanoate (C10-S-S-NHS)
EDC: *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbo.

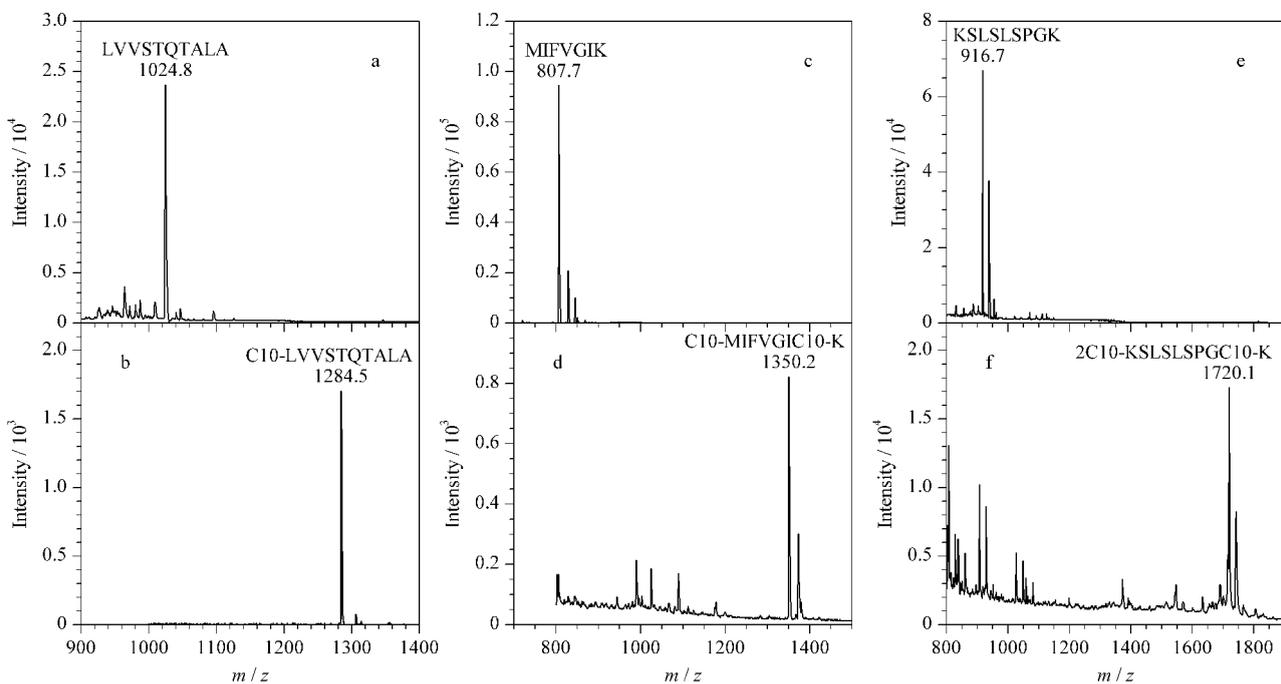


图 3 标准肽段被 C10-S-S-NHS 标记前后的 MALDI/TOF MS 图

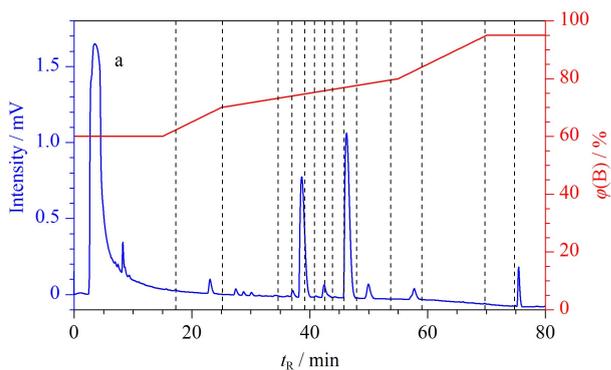
Fig. 3 MALDI/TOF MS profiles of standard peptides before and after C10-S-S-NHS labeling

a, b. peptide of LA-10 (sequence: LVVSTQTALA) before (a, m/z 1024.8) and after (b, m/z 1284.5) labeling; c, d. peptide of MK-7 (sequence: MIFVGIK) before (c, m/z 807.7) and after (d, m/z 1350.2, Na⁺) labeling; e, f. peptide of KK-9 (sequence: KLSLSPGK) before (e, m/z 916.7) and after (f, m/z 1720.1, Na⁺) labeling.

进一步考察了 C10-S-S-NHS 对复杂样品 HeLa 酶解肽段的标记效率。将二硫键断裂释放 C10 链后,标记肽段上的氨基生成一个硫酰基 ($-\text{COCH}_2\text{CH}_2\text{SH}$) 修饰。通过统计标记后肽段样品中氨基被硫酰基修饰的比例,得到 C10-S-S-NHS 对 HeLa 酶解肽段的标记效率为 96.2%,其中被标记 1 个、2 个、3 个 C10 链的肽段比例分别为 46.6%、49.1%、0.6% (图 4)。分别选取被标记 1 个、2 个、3 个 C10 链肽段的 MS/MS 图 (附图 2),发现肽段的 N 末端和 K 侧链的氨基均被标记上硫酰基,且匹配到了大部分 b、y 离子,表明标记后的含 K 多肽可以被 MS 高效鉴定。值得一提的是,在 3.8% 未被标记的酶解肽段中,有 85.7% 的氨基位于蛋白质的 N 末端且发生了乙酰化修饰,进一步说明了 C10-S-S-NHS 对复杂样品中自由氨基的高效标记,为含 K 多肽与不含 K 多肽在后续反相色谱中的有效分离提供保障。



图 4 HeLa 酶解肽段标记 0、1、2 和 3 个 C10 烷基链的百分比分布
Fig. 4 Percentage distribution of HeLa digests labeled by zero, one, two, and three C10 alkyl chains
w/o: without.



2.3 分离梯度考察

将 HeLa 酶解肽段采用 C10-S-S-NHS 进行疏水基团标记后进行反相色谱分离。首先采用线性梯度洗脱,收集不同保留时间的肽段流出液,并将二硫键还原释放 C10 链后,进行 nanoLC-MS/MS 分析。色谱分离图如图 5a 所示,由于标记 C10 链后酶解肽段的疏水性整体增强,选择流动相中 ACN 含量为 48% (v/v) 作为初始洗脱浓度 (即 60% B, 流动相 B 为 80% (v/v) 乙腈水溶液),此时样品中不保留的肽段主要为未被 C10 标记和 N 端被标记一个 C10 (即不含 K) 的肽段 (占比 99.1%, 0~17 min, 图 5b); 当升高 ACN 浓度至约 57.6% (72% B, 35 min), 不保留肽段开始以被标记 2 个及以上 C10 的含 K 肽段为主 (>50%), 并且含 K 肽段的比例随着洗脱浓度升高呈递增趋势。当 ACN 浓度增加至 62.4% (78% B, 48 min) 时含 K 肽段的比例 >90%。样品中的标记试剂 C10-S-S-NHS 由于过量且疏水性较强,在约 46 min 时产生较强的紫外吸收峰; 部分 C10-S-S-NHS 水解后,疏水的 NHS 基团变成亲水的羧基 (COOH),由于疏水性降低,在 C18 色谱柱中的保留前移了 8 min (约 38 min)。综上,以 48% ACN (v/v) 为初始洗脱浓度,此时未被 C10 标记的肽段可以被完全洗脱; 逐渐升高洗脱浓度至约 57.6% ACN, 大部分 N 端被标记一个 C10 的不含 K 肽段被洗脱,此后洗脱的肽段以被标记 2 个 C10 及以上的含 K 肽段为主,以上结果表明含 K 肽段在反相色谱中可以与不含 K 肽段实现有效分离。

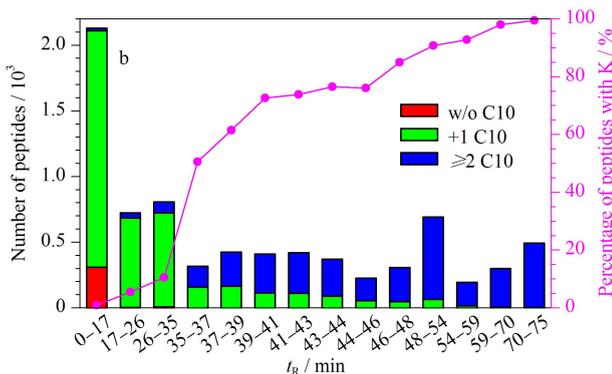


图 5 (a) HeLa 酶解肽段被 C10-S-S-NHS 标记后的反相分离色谱图, (b) 各级分中被标记 0 个、1 个和 ≥ 2 个 C10 链的肽段数 (堆积柱形图) 以及含 K 肽段所占百分比 (折线图)

Fig. 5 (a) RP separation chromatogram of HeLa digests after C10-S-S-NHS labeling, (b) number of peptides labeled by zero, one, two or more C10 chains (stacked bar chart) and percentage of peptides containing K (line chart) in each fraction

Separation gradient: 0-15 min, 60% B; 15-25 min, 60% B-70% B; 25-55 min, 70% B-80% B; 55-70 min, 80% B-95% B; 70-80 min, 95% B (mobile phase A: 2% (v/v) acetonitrile aqueous solution containing 0.1% trifluoroacetic acid (TFA); mobile phase B: 80% (v/v) acetonitrile aqueous solution containing 0.1% TFA); C18 column: 250 mm \times 4.6 mm, 5 μ m; flow rate: 1 mL/min; detection wavelength: 214 nm.

根据上述考察的分离梯度,进一步考察采用阶梯式梯度的洗脱方式对 HeLa 酶解肽段中含 K 多肽的富集效果。流动相 A 为 2% (v/v) 乙腈水溶液 (含 0.1% (v/v) TFA), 流动相 B 为 98% (v/v) 乙腈水溶液 (含 0.1% (v/v) TFA)。4 个梯度设置如下, #1: 0~10 min, 32% B; 10.1~20 min, 56% B; 20.1~30 min, 95% B; #2: 0~10 min, 32% B; 10.1~20 min, 60% B; 20.1~30 min, 95% B; #3: 0~10 min, 32% B; 10.1~20 min, 64% B; 20.1~30 min, 95% B; #4: 0~10 min, 32% B; 10.1~20 min, 68% B; 20.1~30 min, 95% B。

收集 4 个洗脱梯度下的不保留肽段馏分, 还原二硫键释放 C10 链后, 进行 nanoLC-MS/MS 分析。如附图 3a 所示, 在 32% B 洗脱条件下, 鉴定到的肽段以未被 C10 标记的肽段为主, 其中 N 端被乙酰化等修饰的肽段占比 88.9%, 含 K 肽段仅占 0.4%。当 B 相浓度分别提高至 56%~68% (#1~#4), 鉴定到的肽段均以 N 端被标记 1 个 C10 的肽段为主 (>97.5%)。当继续提高 B 相浓度至 95% (附图 3b), 鉴定到的肽段以被标记 2 个及以上 C10 的含 K 肽段为主, 且含 K 肽段的比例在 #3 条件下最高 (占比

85.9%), 其次为 #2、#1 和 #4 (含 K 肽段占比分别为 84.7%、74.3% 和 72.5%), 以上结果说明采用阶梯式洗脱梯度可以实现对复杂酶解肽段样品中含 K 多肽的有效富集。由于鉴定到的含 K 肽段在 #2 条件下最多, 后续将采用 #2 的洗脱梯度作为含 K 多肽的富集条件。

2.4 HeLa 酶解肽段中含 K 多肽的富集及分析

综上, 为了实现对 HeLa 酶解肽段中含 K 多肽的高效富集, 将 HeLa 酶解肽段进行 C10-S-S-NHS 疏水基团标记后, 采用 #2 阶梯式梯度进行反相色谱分离, 收集不同洗脱浓度下的不保留肽段级分, 将标记肽段上的二硫键还原释放 C10 链后, 进行 nanoLC-MS/MS 分析。由于氨基被硫酰基修饰, 肽段的带电数目可能降低。为了提高对含 K 肽段的鉴定覆盖度, 在质谱采集时将携带 +1 价及以上电荷的母离子进行二级碎裂。鉴定结果如图 6a 所示, 在 32% ACN 洗脱条件下, 不保留肽段以未被 C10 标记的肽段为主 (占比 95.7%), 其中含 K 肽段仅占 0.04%。当洗脱梯度升高至 60% ACN, 高达 99.0% 肽段为 N 端被标记一个 C10 的不含 K 肽段, 含 K 肽段比例仅为 0.6%。继续升高梯度至 95% ACN,

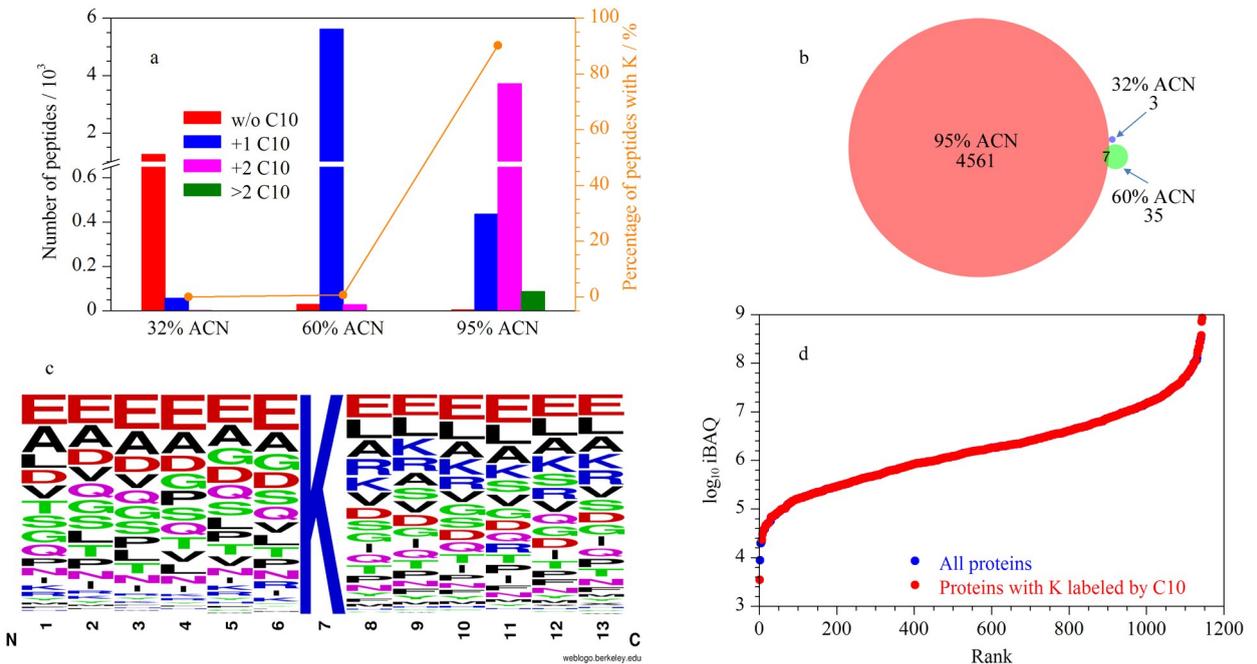


图 6 (a) 各级分中被标记 0 个、1 个、2 个和 2 个以上 C10 链的肽段数 (堆积柱形图) 以及含 K 肽段所占百分比 (折线图), (b) 各级分间含 K 肽段的鉴定结果 Venn 图, (c) 富集的含 K 肽段的序列特征, (d) 定量到的全部 (蓝色圆圈) 以及含有被 C10 标记的 K 的 (红色圆圈) HeLa 细胞蛋白质的丰度范围

Fig. 6 (a) Number of peptides labeled by zero, one, two or more C10 chains (stacked bar chart) and percentage of peptides containing K (line chart) in each fraction, (b) Venn diagram of K-containing peptides identified in each fraction, (c) sequence logo of enriched K-containing peptides, (d) dynamic range of all quantified HeLa cell proteins (blue circles) and K-containing proteins labeled by C10 (red circles)

iBAQ: intensity based absolute quantification.

不保留肽段以被标记 2 个及以上 C10 的含 K 肽段为主,占比为 90.2%,较富集前提高了 35%。在平行两次的 95% ACN 中共鉴定到 4 568 条含 K 肽段,序列重合度为 62.9% (附图 4)。对不同洗脱梯度下鉴定到的含 K 肽段的序列重合度进行比较(图 6b),发现 95% ACN 下单独鉴定到的含 K 肽段占总含 K 肽段的 99.0% (4 561/4 606),表明高分辨率的反相色谱对复杂样品酶解肽段中的含 K 肽段实现了高效分离和富集。进一步对富集的含 K 肽段的序列进行特征分析,如图 6c 所示,K 位点附近的氨基酸分布以酸性氨基酸谷氨酸和侧链为非极性的氨基酸丙氨酸、亮氨酸为主,反映了含 K 肽段具有一定的序列一致性。对富集的含 K 肽段定量到的蛋白质丰度分布进行分析,如图 6d 所示,含 K 肽段定量到的蛋白质丰度动态范围(红色圆圈)跨越了 5~6 个数量级,几乎覆盖了全部肽段定量到的所有蛋白质(蓝色圆圈),且能够鉴定到更多低丰度的蛋白质。

3 结论

本工作发展了一种基于疏水基团标记和反相色谱分离的方法,对复杂样品中含 K 多肽进行高效富集和鉴定。合成的试剂 C10-S-S-NHS 对 HeLa 细胞酶解肽段的标记效率高达 96%。标记后的含 K 肽段与不含 K 肽段在反相色谱实现了高效分离。富集后样品中含 K 肽段的比例高达 90%,较富集前提高了 35%,对应蛋白质的丰度动态范围几乎覆盖了全部肽段定量到的蛋白质,且能够鉴定到更多低丰度的蛋白质。本文发展的 HYTARP 策略为降低样品复杂度,提高对含 K 肽段和低丰度蛋白质的鉴定覆盖率提供了一种简单、高效的方法。

参考文献:

[1] Yates J R. *J Mass Spectrom*, 1998, 33(1): 1
[2] Zhang Y, Fonslow B R, Shan B, et al. *Chem Rev*, 2013, 113(4): 2343
[3] Prus G, Satpathy S, Weinert B T, et al. *Cell*, 2024, 187: 1
[4] Mund A, Brunner A-D, Mann M. *Mol Cell*, 2022, 82(12):

2335

[5] Stukalov A, Girault V, Grass V, et al. *Nature*, 2021, 594(7862): 246
[6] Aebersold R, Mann M. *Nature*, 2003, 422: 198
[7] Li Y, Yuan H, Cheng M, et al. *Sci Bull*, 2022, 67(16): 1628
[8] Ren D Y, Julka S, Inerowicz H D, et al. *Anal Chem*, 2004, 76(15): 4522
[9] Wisniewski J R, Prus G. *Anal Chem*, 2015, 87(13): 6861
[10] Qi Q, Yan G, Deng C, et al. *Anal Bioanal Chem*, 2016, 408(29): 8437
[11] Shah P, Zhang B, Choi C, et al. *Anal Biochem*, 2015, 469: 27
[12] Foettinger A, Leitner A, Lindner W. *J Proteome Res*, 2007, 6(9): 3827
[13] Yu Y, Liu M, Yan G, et al. *Talanta*, 2011, 85(2): 1001
[14] Reid G E, Roberts K D, Simpson R J, et al. *J Am Soc Mass Spectrom*, 2005, 16(7): 1131
[15] Mesmin C, Domon B. *J Proteome Res*, 2014, 13(12): 6160
[16] Yao Y, Huang J, Cheng K, et al. *Anal Chem*, 2015, 87(22): 11353
[17] Li Z, Wang Q, Mao J, et al. *Anal Chim Acta*, 2021, 1142: 48
[18] Cao T, Lv J, Zhang L, et al. *Anal Chem*, 2018, 90(24): 14303
[19] Yu C, Huang L. *Anal Chem*, 2018, 90(1): 144
[20] Zhu Y, Akkaya K C, Ruta J, et al. *Nat Commun*, 2024, 15(1): 3290
[21] Han R N, Zhao L L, An Y X, et al. *Chinese Journal of Chromatography*, 2022, 40(3): 224
韩若楠, 赵丽丽, 安雨馨, 等. 色谱, 2022, 40(3): 224
[22] Dutton A, Adams M, Singer S J. *Biochem Biophys Res Commun*, 1966, 23(5): 730
[23] Zheng H, Mao J W, Pan Y B, et al. *Chinese Journal of Chromatography*, 2014, 32(4): 376
郑昊, 毛家维, 潘彦博, 等. 色谱, 2014, 32(4): 376
[24] Chen F, Gerber S, Korkhov V M, et al. *J Am Soc Mass Spectrom*, 2015, 26(3): 493
[25] Paramelle D, Miralles G, Subra G, et al. *Proteomics*, 2013, 13(3/4): 438
[26] Madler S, Bich C, Touboul D, et al. *J Mass Spectrom*, 2009, 44(5): 694
[27] Kao A, Chiu C L, Vellucci D, et al. *Mol Cell Proteomics*, 2011, 10(1): M110 002212
[28] Zhao L, Zhao Q, Shan Y, et al. *Anal Chem*, 2019, 92(1): 1097
[29] Hu Y, Li Y, Gao H, et al. *Anal Bioanal Chem*, 2019, 411(18): 4159