基于二辛基化标记的蛋白质组N末端肽富集方法

陈玲凡^{1,3}, 单亦初¹, 张丽华^{1*}, 张玉奎¹, 范润龙²

1. 中国科学院大连化学物理研究所, 中科院分离分析化学重点实验室, 国家色谱研究分析中心, 大连 116023;

2. 中国地质科学院地质研究所北京离子探针中心, 北京 102206

3. 福建医科大学, 福建省新药安全性评价中心, 福州 350108;

* 联系人, E-mail: lihuazhang@dicp.ac.cn

收稿日期: 2017-09-16; 接受日期: 2017-10-27; 网络版发表日期: 2018-02-05

国家自然科学基金(批准号: 21475127, 21705020)、国家重点研发计划(批准号: 2017YFA0505004, 2017YFA0505003)和北京离子探针中心开放课题资助

摘要 蛋白质末端的研究不仅对于蛋白质的结构和功能注释十分必要, 还能够为蛋白酶作用机理的揭示提供重要信息. 针对传统蛋白质组N末端肽富集需要采用大量去除材料导致N末端肽回收率低, 以及步骤繁琐、样品损失等问题, 本文发展了一种基于二辛基化标记的蛋白质组N末端肽反向富集方法, 可以在单次反相色谱分离中实现非N末端肽的去除. 以酵母蛋白质酶解产物为样品对方法进行考察, 结果表明, 该方法具备较高的二辛基化标记效率(93%), 且非N末端肽经二辛基化标记后能够显著偏移至强保留区. 将其应用于酵母蛋白质N末端组分析, 共鉴定到237条蛋白原始N末端肽和133 neo-N末端肽, 较富集前分别提高了2.1和3.3倍. 此外, 将该方法结合多种酶切方式, 使N末端肽和neo-N末端肽的鉴定数目进一步提高37%和60%, 促进了蛋白质N末端组鉴定覆盖度的提高.

关键词 蛋白质N末端组学, N末端肽富集, 二辛基化标记, 反相色谱, 蛋白质水解

蛋白质的N末端存在丰富的共翻译和翻译后修饰, 其与蛋白的稳定性, 定位和活性等各项生理功能密切相关^[1-4]. 蛋白酶调控的蛋白水解作为生物体内最常见的翻译后修饰之一, 在众多生理和病理过程中发挥着重要的作用, 如信号传导、细胞凋亡、癌症发生发展等^[5,6]. 人类基因组中, 570个基因用于蛋白酶的编码, 然而超过1/2的蛋白酶的作用底物尚不明确^[7]. 由于蛋白酶调控的蛋白水解促使蛋白产生新的N末端(neo-N末端), 因此通过neo-N末端的鉴定即可追踪蛋白酶的作用底物和位点^[8-13]. 因此, 蛋白质N末端组研究逐渐

受到人们的关注. 由于蛋白质组样品的复杂性, 为了提高蛋白质N末端组的鉴定覆盖度和灵敏度, N末端肽的选择性富集至关重要.

N末端肽的富集策略可分为正向富集与反向富集. 正向富集通过在蛋白质N末端特异性引入亲和标签来实现N末端肽的选择性富集. 然而, 在众多的生物体内超过50%的蛋白质其N末端存在内源性的修饰^[6], 阻碍了亲和标签的引入. 因此, 反向富集策略引起了广泛的关注. 在该策略中, 对蛋白质上的氨基封闭后进行酶解, 非N末端肽会暴露出新的自由氨基, 进而可以通过

引用格式: 陈玲凡, 单亦初, 张丽华, 等. 基于二辛基化标记的蛋白质组N末端肽富集方法. 中国科学: 生命科学, 2018, 48: 215–223

Chen L F, Shan Y C, Zhang L H, et al. Enrichment strategy of protein N-termini based on dioctyl labeling (in Chinese). Sci Sin Vitae, 2018, 48: 215–223, doi: 10.1360/N052017-00154

自由氨基的捕获来去除非N末端肽, 实现N末端肽的反向富集. 氨基反应活性材料, 包括N-羟基琥珀酰亚胺酯活性琼脂糖^[14]、溴化氰活性琼脂糖^[15]、异氰酸树脂^[16]等功能化材料^[17-19], 已被用于非N末端肽的简便快速去除. 然而, 受限于固-液反应效率的不足, 非N末端肽难以被完全去除. 为此, 人们利用液-液反应效率高的优点, 采用氨基反应活性试剂在非N末端肽上引入亲和标签, 然后基于亲和标签与亲和材料的选择性结合, 实现非N末端肽的高效去除^[19-21]. 但是, 以上方法均需要使用大量的功能化材料进行非N末端肽的去除, 难以避免N末端肽由于非特异性吸附造成的严重损失. Kleinfeld等人^[22]则发展了一种基于水溶性高分子聚合物的反向去除方法, 称为TAILS (terminal amine isotopic labeling of substrates). 该方法通过合成一种醛基功能化的高分子聚合物, 利用醛基与氨基的反应将非N末端肽连接于聚合物链上, 采用超滤膜分离聚合物, 从而去除非N末端肽. 由于聚合物的醛基含量高且水溶性好, 因此非N末端肽得到了高效的去除, 且减少了N末端肽的非特异吸附. 此外, Gevaert课题组^[23]发展了基于反相对角线色谱法(COFRADIC)的N末端肽分离方法. 该方法首先对蛋白质N末端氨基和侧链氨基进行乙酰化, 蛋白质酶解后, 使用反相色谱对肽段混合物进行分离, 并收集馏分; 然后, 对每个馏分中的氨基进行修饰, 以改变非N末端肽的疏水性; 最后, 将每个馏分分别进行第二次反相色谱分离, 非N末端肽由于疏水性的变化而发生保留偏移, 从而与N末端肽分离. Zahedi课题组^[24]发展了基于强阳离子交换的对角线色谱法的N末端肽分离方法. 该方法采用化学修饰改变非N末端的带电性, 使非N末端肽在强阳离子色谱上的保留减弱, 从而在第二次强阳离子交换色谱中与N末端肽分离. 基于对角线色谱的N末端肽分离策略虽然避免了功能化材料的引入, 然而分离过程繁琐, 耗时耗力, 而且较多的样品纯化过程导致了样品的损失. Schilling课题组^[25]则通过在非N末端肽标记上4个磺酸根, 采用强阳离子交换色谱实现非N末端肽的快速去除, 但是该方法会造成弱碱性N末端肽的丢失. 本课题组^[26]则发展了一种基于双十六烷基标记的蛋白质组N末端肽反向富集方法, 称为HYTANE (hydrophobic tagging-assisted N-termini enrichment). 该方法基于疏水基团与C18材料的强相互作用, 可以在常规的样品除盐过程中对非N末端肽进行简便快速的去除.

本文发展了一种基于二辛基标记的N末端肽反向富集方法. 该方法通过二辛基标记使非N末端肽偏移至强保留区, 从而在单次反相色谱分离中实现N末端肽的分离. 将该方法应用于酵母蛋白质N末端组的富集, 提高了N末端肽的鉴定效率. 此外, 将该方法与多酶酶切结合, 进一步提高了N末端组的鉴定覆盖度.

1 材料与方法

1.1 实验材料

二硫苏糖醇(dithiothreitol, DTT)、丙烯酰胺、牛血清蛋白(bovine albumin, BSA)、甲醛(37%水溶液)、氰基硼氢化钠(NaCNBH_3)、盐酸胍、三乙基碳酸氢铵缓冲液(triethylammonium bicarbonate, TEAB)、4-(2-羟乙基)哌嗪-1-乙磺酸N-(2-羟乙基)哌嗪-N'-(2-乙磺酸)(4-(2-hydroxyethyl)piperazine-1-erhaesulfonic acid, HEPES)、蛋白酶抑制剂(cocktail)、辛醛均购于Sigma-Aldrich (美国); 胰蛋白酶购自Promega (美国); BCA试剂盒购自碧云天(上海); 超滤离心管(截留分子量10 kD)购自赛多利斯(德国); Daiso C18填料(5 μm , 100 Å)购自Daiso公司(日本); 色谱纯乙腈(ACN, HPLC级别)购自Merck (德国), 所有实验用水均经过Milli-Q system (Millipore, 美国)纯化; 其他试剂均为分析纯.

1.2 细胞培养和蛋白质样品提取

酵母细胞在37℃下于YPD培养基中培养48 h, 然后在4℃下4000×g离心2 min. 沉淀的细胞采用1×PBS清洗3遍后, 悬浮于由6 mol/L盐酸胍和1% (v/v)蛋白酶抑制剂组成的细胞裂解液中, 并在冰浴下超声破碎360 s (100%能量, 10 s开, 10 s关). 匀浆液在4℃下20000×g离心40 min, 收集上清, 采用BCA试剂盒测定所得蛋白浓度.

1.3 蛋白质水平氨基二甲基化

往6 mol/L盐酸胍溶解的蛋白质(~0.1 mg)中加入TEAB缓冲液(pH 8.2)至终浓度100 mmol/L, 并加入10 μL 100 mmol/L DTT, 56℃水浴还原1 h. 冷却至室温, 加入10 μL 300 mmol/L丙烯酰胺, 室温烷基化1 h后, 再次加入10 μL 100 mmol/L DTT消耗残余的丙烯酰胺. 将样品溶液转移至超滤膜上, 14000×g离心去除离心液, 采用100 mmol/L TEAB缓冲液清洗一遍后, 加

入由100 mmol/L TEAB, 100 mmol/L 甲醛, 60 mmol/L NaCNBH₃和6 mol/L 盐酸胍组成的反应溶液, 于37℃孵育2 h. 最后, 加入10倍过量的甘氨酸消耗残余的甲醛和NaCNBH₃.

1.4 蛋白质酶解

超滤膜上的样品经离心去除离心液, 采用由50 mmol/L NH₄HCO₃, 1 mol/L 甘氨酸, 6 mol/L 盐酸胍组成的清洗液清洗3遍, 50 mmol/L 磷酸氢二钠缓冲液(pH 8)清洗一遍后, 采用50 mmol/L 磷酸氢二钠缓冲液溶解膜上的蛋白. 以酶:蛋白质的质量比为1/25的比例加入胰蛋白酶/Glu-C/糜蛋白酶, 37℃酶解12 h后, 14000×g离心收集离心液, 得到酶解肽段.

1.5 二辛基标记

将100 μg肽段样品溶解于200 μL H₂O中, 依次加入100 μL 50 mmol/L HEPES缓冲液(pH 9.0), 40 μL 600 mmol/L NaCNBH₃ (溶于ACN), 1 mL 1% (v/v) 辛醛(溶于ACN), 50℃孵育8 h. 然后, 再次加入40 μL 600 mmol/L NaCNBH₃ (溶于ACN), 并采用真空冷冻干燥冻干去除乙腈. 将冻干样品溶解于A相(2% ACN/0.1% TFA), 12000×g离心15 min后, 去除沉析出的辛醛, 收集水溶液. 除盐冻干, 待质谱分析.

1.6 nanoRPLC-ESI-MS/MS分析

采用Accela 600泵, Accela自动进样器和LTQ Orbitrap Velos质谱仪(Thermo Fisher, 美国)构建的nano-RPLC-ESI-MS/MS系统对样品进行分析. Nano-RPLC的分离条件如下: C18毛细管分离柱(75 μm i.d.×14 cm); C18捕集柱(150 μm i.d.×3 cm); 流动相A: 98% H₂O/0.1% FA; 流动相B: 98% ACN/0.1% FA; 分离柱流速: 300 nL/min; 分离梯度: 5% B (0 min)-35% B (100 min)-80% B (110 min)-80% B (120 min).

质谱分析在正离子模式下进行, 喷雾电压为2.3 kV, 离子传输毛细管温度为250℃. 总离子色谱图和质谱图采用Xcalibur软件(版本2.1.0)对m/z在300~1800的范围内进行数据采集. 在正离子模式下, 每个MS和MS/MS的微扫描数为1. MS/MS谱图采用数据依赖模式进行采集. 每一个MS采集后面针对该MS全扫描中15个最强的母离子进行MS/MS扫描. MS的分辨率设为60000. 动态排除设定为: 重复次数为1; 重复

容忍时间40 s; 动态排除时间40 s.

1.7 数据分析

质谱数据文件(*.raw)采用Proteome Discoverer (版本1.4, Thermo Fisher Scientific)中整合的Mascot (版本2.4, Matrix Science)进行数据库检索. 酵母蛋白数据库下载于UniProt, 下载日期为2015年12月24号, 数据库包含6720个蛋白序列. 采用反库控制结果的假阳性率(FDR). 数据检索参数为: Arg-C/Glu-C/糜蛋白酶全酶切, 最多允许两个漏切位点; 固定化修饰: 半胱氨酸烷基化(+71.0371 Da), 赖氨酸二甲基化(+28.0313 Da); 可变修饰: 甲硫氨酸的氧化(+15.9949 Da), 蛋白质N末端乙酰化(+42.0105 Da), 蛋白质N末端二甲基化(+28.0313 Da), 肽段N末端辛基化(+112.1252 Da), 肽段N末端二辛基化(+224.2504 Da); 母离子质量允许偏差为10 ppm, 碎片离子质量允许偏差为0.5 Da. 检索neo-N末端时, 数据检索参数设置为: Arg-C/Glu-C/糜蛋白酶半酶切; 可变修饰为甲硫氨酸的氧化(+15.9949 Da), 肽段N末端二甲基化(+28.0313 Da)和肽段N末端乙酰化(+42.0105 Da), 肽段N末端辛基化(+112.1252 Da), 肽段N末端二辛基化(+224.2504 Da), 其余参数同上一致. 采用FDR<1%, 离子得分(ion score)≥25处理数据库检索结果, 以得到高可信的鉴定结果.

2 结果与讨论

2.1 富集方法的流程

该方法富集N末端肽的原理如图1所示: 首先, 在蛋白质水平上对蛋白质N末端以及侧链自由氨基进行二甲基化. 然后, 将二甲基化蛋白进行酶解, 肽键的水解使中间肽段于氨基端重新带上一个自由氨基, 而N末端肽不含自由氨基. 基于氨基与醛基的Schiff碱还原反应, 在中间肽段的氨基端引入两个辛基. 最后, 将样品上样至C18反相柱进行色谱分离, 由于二辛基基团的引入, 中间肽段偏移至强保留区, 而蛋白质N末端肽则仍处于常规保留区内, 从而实现N末端肽与中间肽段的分离.

2.2 标记效率考察

蛋白质N末端区域的氨基如果封闭不完全, 将导致N末端肽在后续的去流程中发生共去除, 因此蛋

图1 基于二辛基化标记的N末端肽富集流程示意图(网络版彩图)

蛋白质水平的二甲基化效率直接影响N-端肽段的富集效率。此外,蛋白质水平的二甲基化对N末端肽的验证也至关重要。N末端带有二甲基化标签的肽段可被确认为蛋白质的原始N端肽段和蛋白质的内源性水解所产生的neo-N端肽段,而非蛋白质的体外降解、酶的非特异酶切和肽段的源内裂解所产生的假阳性N末端肽^[27]。为了考察蛋白质水平二甲基化效率,将标准蛋白BSA进行二甲基化和酶解,并采用ESI-LC/MS/MS对酶解肽段进行分析。首先,选取BSA蛋白中的一段序列(SLHTLFGDELCKVASLR)进行分析。当此序列K位点上的氨基被封闭完全时,胰酶对K位点无酶切活性,此序列以完整肽段的形式被鉴定,反之则以碎片肽段(SLHTLFGDELCK)的形式被鉴定。提取这两种肽段的离子色谱图,结果如网络版附图1所示,此序列主要以完整肽段的形式存在。进一步,将所采集的二甲基化BSA酶解产物的质谱图进行数据库匹配,所匹配到的谱图中 $92.7\% \pm 0.12\%$ ($n=3$)对应于氨基完全二甲基化的肽段。以上结果均表明,蛋白质水平的二甲基化具有较高的标记效率。

为了考察二辛基标记的效率,以标准肽段RVYVHPI为样品,采用MALDI-TOF MS对标记后的肽段进行分析。结果如网络版附图2所示,标记后原始肽段质谱峰消失的同时产生+224 Da的质谱峰,归属于肽段的二辛基化标记,且未发现单辛基化肽段的质谱峰,表明肽段发生了几乎完全的二辛基化标记。进一步,以酵母提取蛋白酶解产物为样品,考察复杂样品中二辛基化标记的效率。采用ESI-LC/MS/MS对标记

后的样品进行分析,经数据库检索,~93%的肽段发生了完全的二辛基化标记,说明在复杂样品中二辛基化仍具备很高的标记效率。

2.3 二辛基化标记对肽段反相色谱保留的影响

为了考察肽段标记二辛基后在反相色谱上的保留偏移,以酵母蛋白酶解产物为样品,将标记前和标记后的肽段混合后进行ESI-LC/MS/MS分析。如图2A所示,肽段经二辛基标记后显著偏移至强保留区而与未标记肽段分离。进一步对不同性质肽段的保留偏移进行分析,如图2B所示,肽段的保留时间偏移与其本身的亲疏水性相关。在0.75% ACN/min的梯度下,亲水肽段(保留时间 ≤ 20 min)经二辛基化标记后肽段的平均保留时间偏移量为32 min,疏水肽段(保留时间 > 20 min)

图2 二辛基化标记对肽段反相色谱保留的影响

A: 肽段标记二辛基前后的保留时间分布; B: 疏水性不同肽段的保留时间偏移量。分离梯度: 5% B (0 min) -95% B (120 min)。流动相A为2% ACN/0.1%甲酸, 流动相B为98% ACN/0.1%甲酸

的偏移量为25 min, 表明亲水肽段标记二辛基后保留时间偏移较大, 该性质有助于将亲水肽段偏移至强保留区. 以上结果均说明, 肽段在经过二辛基化标记后能够有效地偏移至强保留区, 从而实现干扰肽段的高效去除.

2.4 酵母蛋白质组N末端肽的富集

为了评价方法对N末端肽的富集能力, 以等量的酵母提取蛋白为样品, 对富集前和富集后的样品分别进行nano-RPLC-ESI-MS/MS分析. 如图3A所示, 样品

经富集后, 中间肽段显著偏移至强保留区, 在常规保留区域内的鉴定数目大幅度下降. 如图3B所示, 消除中间肽段的干扰后, N-末端肽的鉴定数目得到显著提高, 共370条N末端肽得到鉴定. 所鉴定到的N末端肽中, 原始N末端肽237条, neo-N末端肽133条, 较富集前分别提高了2.1和3.3倍(网络版附表1). 此外, 对富集前后共同鉴定的N末端肽的肽段谱图匹配数(PSMs)和肽段得分也进行了比较. 如图3C和D所示, 富集后N末端肽的PSMs的中位值由2提高至4, 肽段谱图得分的中位值由39提高至55, 表明富集后N末端肽在质谱中的鉴定几

图3 富集前后的鉴定结果比较

A: 富集前后中间肽段的鉴定数目分布; B: 富集前后N末端肽的鉴定数目分布; C: 富集前后共同鉴定的肽段的谱图数分布; D: 富集前后共同鉴定的肽段的谱图得分分布

率得到提高. 由此可见, 本方法能够高效去除中间肽段的干扰, 大大地降低样品的复杂度, 从而有效地提高N末端肽的鉴定效率和覆盖度.

与文献报道的方法相比, 反相对角线色谱法(CO-FRADIC)^[28]采用丁基或三硝基苯基标记使中间肽段的保留发生偏移, 由于偏移量较小, 该方法需要将样品分隔成多个馏分, 并进行~15次反相色谱分离和~120次质谱分析, 大大地限制了样品的分析通量; 同样的, 强阳离子对角线色谱(ChaFRADIC)^[24]需要进行~6次强阳离子对角线色谱和繁琐的样品除盐过程; 本实验采用二辛基标记, 显著地提高肽段的保留偏移量, 通过单次的反相色谱分离就能够去除大部分中间肽段的干扰, 避免了繁琐的分离过程, 提高了样品的分析通量. 但是, 由于部分亲水性肽段经二辛基标记后偏移至疏水区域, 干扰了疏水N末端肽的鉴定, 因此根据亲水性肽段的平均保留偏移量, 在富集前于C18捕集柱上采用30%和80% ACN将样品分成2个馏分, 以拆分亲水区域和疏水区域的肽段, 避免亲水区域肽段经二辛基标记后干扰疏水区域N末端肽的鉴定, 可以进一步提高N末端肽的鉴定覆盖度和选择性. 在先前的研究中本课题组采用了双十六烷基标记^[26], 可以更大程度地提高肽段的疏水性, 从而实现除盐过程中非N末端肽的高效快速去除, 进一步提高了样品分析通量. 相比于双十六烷基标记, 二辛基标记过程中残余标记试剂不会造成液相系统的堵塞, 因此可以使用低/高pH反相分离, 在去除N末端肽的过程中实现样品的分级.

通过对酵母蛋白质数据库进行理论胰蛋白酶酶切, 发现仅约50%的N末端肽具备适合质谱鉴定的长度. 此外, 部分胰蛋白酶酶切所得的N末端肽由于具有较差的离子化和碎裂效率而难以鉴定. 为了进一步提高N末端肽鉴定的覆盖度, 另外采用Glu-C蛋白酶和糜蛋白酶对酵母提取蛋白进行酶解, 酶解产物经富集后进行nano-RPLC-ESI-MS/MS分析. 结合3种酶解方式的鉴定结果, 总共鉴定到361条原始N末端肽和257条neo-N末端肽(网络版附表2), 较单独采用胰酶酶切分别提高了50%和90%. 对3种酶切方式所得鉴定结果的重合度分析显示三者共同鉴定到的N末端肽仅占有N末端肽的0.6%, 任意两者之间的重合度不超过7%(图4), 表明3种酶切方式具有很大的互补性. 以上结果表明, 采用多酶酶切方法对蛋白质的N端肽进行分析, 以获得性质互补的N末端肽, 将进一步促进蛋白质N末

图4 3种酶切方式的N末端肽鉴定结果维恩图

端组鉴定覆盖度的提高.

鉴定到的361条原始N末端肽中71% (258/361)发生乙酰化修饰, 而且大量蛋白质N末端甲硫氨酸发生剪切(78%, 282/361). 另外, 根据257个neo-N末端的起始位点与序列信息, 验证了酵母蛋白质数据库中12个蛋白质信号肽切除位点^[29,30], 并发现了7个符合N末端甲硫氨酸剪切和N末端乙酰化规则的neo-N末端, 可能为蛋白质的候选翻译起始位点(网络版附表3)^[31~33]. 以上结果表明, 本方法能够为蛋白质加工的注释提供有效技术手段.

3 结论

本文发展了一种基于二辛基化标记的蛋白质组N末端肽反向富集方法, 可以在单次反相色谱分离中实现非N末端肽的去除, 避免了传统蛋白质组N末端肽富集需要采用大量去除材料导致N末端肽回收率低, 以及步骤繁琐、样品损失等问题. 该方法具备较高的二辛基化标记效率(93%), 且非N末端肽经二辛基化标记后能够显著偏移至强保留区. 将其应用于酵母蛋白质N末端组的分析, 共鉴定到237条蛋白原始N末端肽和133 neo-N末端肽, 较富集前分别提高了2.1和3.3倍. 此外, 将该方法结合多种酶切方式, 共鉴定到361条原始

N末端肽和257条neo-N末端肽, 较单独采用胰酶酶切分别提高了50%和90%, 促进了蛋白质N末端组鉴定覆盖度的提高. 上述结果表明, 基于二辛基化标记的N末

端肽富集方法在蛋白质组N末端的研究中具有广泛的应用前景, 有望为蛋白质结构和功能的注释以及蛋白水解相关生物过程的研究提供实用技术.

参考文献

- Shih Y P, Chou C C, Chen Y L, 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, 9: e94812
- Scott D C, Monda J K, Bennett E J, et al. N-terminal acetylation acts as an avidity enhancer within an interconnected multiprotein complex. *Science*, 2011, 334: 674–678
- Wadas B, Borjigin J, Huang Z, et al. Degradation of serotonin *N*-acetyltransferase, a circadian regulator, by the N-end rule pathway. *J Biol Chem*, 2016, 291: 17178–17196
- Calvo S E, Julien O, Clauser K R, et al. Comparative analysis of mitochondrial N-termini from mouse, human, and yeast. *Mol Cell Proteomics*, 2017, 16: 512–523
- Lai Z W, Petrer A, Schilling O. The emerging role of the peptidome in biomarker discovery and degradome profiling. *Biol Chem*, 2015, 396: 185–192
- Rogers L D, Overall C M. Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol Cell Proteomics*, 2013, 12: 3532–3542
- Quesada V, Ordóñez G R, Sánchez L M, et al. The degradome database: mammalian proteases and diseases of proteolysis. *Nucleic Acids Res*, 2008, 37: D239–D243
- Schilling O, Overall C M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol*, 2008, 26: 685–694
- Schlage P, Egli F E, Nanni P, et al. Time-resolved analysis of the matrix metalloproteinase 10 substrate degradome. *Mol Cell Proteomics*, 2014, 13: 580–593
- Prudova A, Gocheva V, Auf dem Keller U, et al. TAILS N-terminomics and proteomics show protein degradation dominates over proteolytic processing by cathepsins in pancreatic tumors. *Cell Rep*, 2016, 16: 1762–1773
- Shimbo K, Hsu G W, Nguyen H, et al. Quantitative profiling of caspase-cleaved substrates reveals different drug-induced and cell-type patterns in apoptosis. *Proc Natl Acad Sci USA*, 2012, 109: 12432–12437
- Demir F, Niedermaier S, Villamor J G, et al. Quantitative proteomics in plant protease substrate identification. *New Phytol*, 2017, in press, doi: 10.1111/nph.14587
- 陈玲凡, 单亦初, 张丽华, 等. 蛋白质组末端肽富集策略研究进展. *色谱*, 2017, 35: 229–236
- McDonald L, Beynon R J. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat Protoc*, 2006, 1: 1790–1798
- Zhang X, Ye J, Engholm-Keller K, et al. A proteome-scale study on *in vivo* protein *N*-acetylation using an optimized method. *Proteomics*, 2011, 11: 81–93
- Mikami T, Takao T. Selective isolation of N-blocked peptides by isocyanate-coupled resin. *Anal Chem*, 2007, 79: 7910–7915
- Shen P T, Hsu J L, Chen S H. Dimethyl isotope-coded affinity selection for the analysis of free and blocked N-termini of proteins using LC-MS/MS. *Anal Chem*, 2007, 79: 9520–9530
- Kuyama H, Sonomura K, Nishimura O, et al. A method for N-terminal *de novo* sequence analysis of proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal Biochem*, 2008, 380: 291–296
- Zhao L, Zhang Y, Wei J, et al. A rapid isolation and identification method for blocked N-terminal peptides by isothiocyanate-coupled magnetic nanoparticles and MS. *Proteomics*, 2009, 9: 4416–4420
- Mommen G P M, van de Waterbeemd B, Meiring H D, 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: 832–842
- Li L, Wu R, Yan G, et al. A novel method to isolate protein N-terminal peptides from proteome samples using sulfhydryl tagging and gold-

- nanoparticle-based depletion. *Anal Bioanal Chem*, 2016, 408: 441–448
- 22 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: 281–288
- 23 Staes A, Van Damme P, Helsens K, et al. Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics*, 2008, 8: 1362–1370
- 24 Venne A S, Vögtle F N, Meisinger 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: 3823–3830
- 25 Lai Z W, Gomez-Auli A, Keller E J, et al. Enrichment of protein N-termini by charge reversal of internal peptides. *Proteomics*, 2015, 15: 2470–2478
- 26 Chen L, Shan Y, Weng Y, et al. Hydrophobic tagging-assisted N-termini enrichment for in-depth N-terminome analysis. *Anal Chem*, 2016, 88: 8390–8395
- 27 Wang H, Tang H Y, Tan G C, et al. Data analysis strategy for maximizing high-confidence protein identifications in complex proteomes such as human tumor secretomes and human serum. *J Proteome Res*, 2011, 10: 4993–5005
- 28 Helsens K, Van Damme P, Degroeve S, et al. Bioinformatics analysis of a *Saccharomyces cerevisiae* N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation. *J Proteome Res*, 2011, 10: 3578–3589
- 29 Vögtle F N, Wortelkamp S, Zahedi R P, et al. Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability. *Cell*, 2009, 139: 428–439
- 30 Emanuelsson O, Brunak S, von Heijne G, et al. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, 2007, 2: 953–971
- 31 Gawron D, Ndah E, Gevaert K, et al. Positional proteomics reveals differences in N-terminal proteoform stability. *Mol Syst Biol*, 2016, 12: 858–858
- 32 Lange P F, Huesgen P F, Nguyen K, et al. Annotating N termini for the human proteome project: N termini and N α -acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome. *J Proteome Res*, 2014, 13: 2028–2044
- 33 Prudova A, Serrano K, Eckhard U, et al. TAILS N-terminomics of human platelets reveals pervasive metalloproteinase-dependent proteolytic processing in storage. *Blood*, 2014, 124: e49–e60

Enrichment strategy of protein N-termini based on dioctyl labeling

CHEN LingFan^{1,3}, SHAN YiChu¹, ZHANG LiHua¹, ZHANG YuKui¹ & FAN RunLong²

1 CAS Key Laboratory of Separation Sciences for Analytical Chemistry, National Chromatographic R. & A. Center, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China;

2 The Beijing SHRIMP Center, Institute of Geology Chinese Academy of Geological Sciences, Beijing 102206, China;

3 Fujian Province New Drug Safety Evaluation Center, Fujian Medical University, Fuzhou 350108, China

The analysis of protein N-termini provides valuable information for protein structure and function annotation, and helps the profiling of proteases substrates and cleavage sites. However, most of the current N-terminal enrichment approaches involve multiple chromatographic separation processes and require a large amount of scavenger materials, which make it time and labor consuming and may induce significant sample loss. Herein, we develop a negative N-termini enrichment strategy based on dioctyl labeling. With yeast lysate digests as the sample, the strategy showed a high efficiency in dioctyl labeling and retention shift. Such a strategy was applied for the enrichment of N-terminal peptides from yeast cell lysates and enabled the identification of 237 original N-termini and 133 neo-N-termini, which was improved by 100% and 480% compared with direct analysis. Furthermore, with the combination of multi-proteases digestion, the number of protein N-termini and neo-N-termini was improved by 50% and 90%, respectively, facilitating the in-depth N-terminome analysis.

terminome, N-termini enrichment, dioctyl labeling, reversed-phase separation, proteolysis

doi: [10.1360/N052017-00154](https://doi.org/10.1360/N052017-00154)