

液质联用技术中不同蛋白质鉴定策略的比较

卫军营, 张养军, 陈 明, 覃培斌, 王京兰, 蔡 纲, 钱小红

(军事医学科学院放射与辐射医学研究所, 北京蛋白质组研究中心, 蛋白质组学国家重点实验室, 北京 102206)

摘要: 在常用的多维液相色谱-质谱联用技术中, 采取延长梯度洗脱时间、进行重复实验、采用质谱分段扫描等多种方法, 提高对蛋白质的鉴定效率。为了系统评价这些方法在复杂生物样本分析中的效果, 应用酵母的蛋白提取物作为样本, 在 LCQ 质谱仪上进行一系列的比较实验。结果表明, 在一定范围内, 随着梯度时间的延长, 被鉴定的非冗余肽段(unique peptide)数量显著增加, 相对应的蛋白质簇(group protein)数量也随之增加。同样, 进行重复实验和采用质谱分段扫描的方法均能提高蛋白鉴定的覆盖率, 而采用质谱分段扫描的策略具有更为显著的效果, 因此在规模化蛋白质组分析中, 应当选择更为合适的、互补的研究策略以提高结果的完整性。

关键词: 酵母; 液相色谱-串联质谱联用(LC-ESI MS/MS); 质谱分段扫描

中图分类号: O 657.63 文献标识码: A 文章编号: 1004-2997(2008)04-226-05

Effect of Different Experimental Measures in a Nano-scale High Performance Liquid Chromatography-Electrospray Tandem Mass Spectrometry on Protein Discovery Rate

WEI Jun-ying, ZHANG Yang-jun, CHEN Ming, QIN Pei-bin, WANG Jing-lan,
CAI Yun, QIAN Xiao-hong

(State Key Laboratory of Proteomics, Beijing Proteome Research Center,
Beijing Institute of Radiation Medicine, Beijing 102206, China)

Abstract: In order to achieve a higher protein discovery rate, many measures were taken to capture as much peptides as possible during the analysis of a practical biological sample. To study the contribution of gradient elution time, repeated runs and mass ranges in the analytical process to the identification rates of proteome in a sample by a shotgun method, yeast lysate was used as a model sample, and a series of experiments were performed on a nano-scale capillary reversed-phase chromatography interfaced to LCQ mass spectrometer to compare the effect between them on achieving the completeness of proteome. The results show that totally identified unique peptides, which finally result in the identification of protein groups, increased remarkably with longer gradient time, repeated runs of a sample and seg-

mented mass ranges, but employing segmented mass ranges is more effective than using longer gradient time or simply repeated runs for improving coverage of yeast protein identifications. The conclusion also suggests that employing complementary analytical strategies should be a better choice in a large-scale proteomic analysis.

Key words: yeast; high performance liquid chromatography-electrospray tandem mass spectrometry (LC-ESI MS/MS); mass range

蛋白质组学不仅研究特定细胞、组织、体液及亚结构中的蛋白质,而且还研究蛋白质异构体、翻译后修饰及其相互作用等,其中基因组、转录组和蛋白质组之间的关联也是蛋白质组学研究的一项重要内容。此项研究的前提之一是能否有效地鉴定出高覆盖率的可靠蛋白质,而这与目前所采用的分析策略息息相关,分析策略的改进有助于提高蛋白质的鉴定效率^[1-2]。虽然常用的2-DE技术可以有效地结合胶上提取酶切和质谱技术,实现对几千种蛋白的分离和分析^[3-6],但是由于其有限的动态范围^[7]及对极端蛋白的歧视^[8-10]等因素,限制了该方法的进一步发展。因此,基于多维液相色谱-串联质谱联用的shotgun技术^[11-14]迅速发展,并广泛应用于蛋白质表达谱、修饰谱及蛋白复合物等的研究中^[15-21]。

在shotgun技术应用中,随着待分析生物样本复杂程度的增加,质谱分析时的离子抑制效应^[22]越发明显,到一定程度时,许多肽段检测不出来。为了解决这个问题,除采用复杂的多维液相分离外^[23-24],还发展了许多方法以提高分离和质谱检测的效率。这些方法包括延长梯度洗脱时间^[25-26]、进行重复实验^[27]、采用质谱分段扫描等^[28-34],虽然这些方法在提高蛋白鉴定的效率方面均有一定的作用,但究竟哪些方法在规模化蛋白质组分析中更为有效,更有利于提高蛋白鉴定的覆盖率和可靠性,它们之间系统性的比较实验还未见报道。因此本实验应用酵母的蛋白提取物作为样本,通过在LCQ质谱仪上进行的一系列比较实验,对上述三种分析策略在蛋白质鉴定时的效率进行了评价。

1 实验部分

1.1 仪器和试剂

纳升级毛细管高效液相色谱-电喷雾离子阱质谱仪:美国Thermo公司产品;超纯水由Millipore纯水系统制备。

测序级胰蛋白酶(trypsin)、二硫苏糖醇

(DTT):美国Promega公司产品;乙酰胺(IAA):比利时Acros公司产品;乙腈(HPLC级):美国J.T.Baker公司产品;甲酸(FA):德国Fluka公司产品; NH_4HCO_3 及尿素(分析纯):北京化工厂产品;酵母(*saccharomyces cerevisiae*):美国Sigma公司产品。

1.2 样本制备

将200 mg酵母置于1 mL含50 mmol·L⁻¹ NH_4HCO_3 的裂解液中,在冰浴上超声裂解10 s(每超声0.1 s,暂停2 s,重复100次)。室温静置30 min后,将裂解液在40 kg离心60 min,接着再将上清液于40 kg离心60 min。最后用Bradford法测定蛋白浓度为1.46 g·L⁻¹。

在蛋白提取物(100 μg,68.4 μL)中加入25 μL溶于50 mmol·L⁻¹ NH_4HCO_3 的8 mol·L⁻¹尿素溶液,进行蛋白变性后,加入2.78 μL 100 mmol·L⁻¹的二硫苏糖醇,在37 °C水浴中反应4 h,再加入1.46 μL 1 mol·L⁻¹碘乙酰胺,在室温下置于暗处反应1 h。再用50 mmol·L⁻¹ NH_4HCO_3 溶液稀释,使脲的最终浓度小于1 mol·L⁻¹,按照酶与蛋白的质量比为1:50,将胰蛋白酶加入到蛋白溶液中,在37 °C水浴中孵育过夜。酶解液经浓缩后重溶于2%乙腈、0.1%甲酸溶液中,并使其最终浓度为0.24 g·L⁻¹。

1.3 纳升级HPLC-MS分析

将2.4 μg酵母蛋白提取物酶切混合物分别进行1、2、3、4、5、7、10 h的梯度分离(其中经3 h梯度分离的实验重复两次),线性梯度洗脱程序为95%A相(2%乙腈、0.1%甲酸溶液)-50%B相(98%乙腈、0.1%甲酸溶液);流速0.2 μL·min⁻¹,毛细管柱为C₁₈反相柱(Biobasic,100×0.18 mm×5 μm,300 Å)。电喷雾电压3.2 kV,正离子模式,毛细管温度160 °C,归一化碰撞能量设置为35%,质量扫描范围m/z 300~1 800,在一级质谱图中选取最强的3个母离子进行串级质谱分析,动态排除时间为5 min。

对照实验如下:(1)将等量样品分为 3 份,即每份 $0.8 \mu\text{g}$,分别进行 1 h 梯度分离;(2)将等量样品分为 3 份,即每份 $0.8 \mu\text{g}$,分别进行 1 h 梯度分离,但在质谱分析时分别采用如下的质量扫描范围 $m/z 300 \sim 800$ 、 $700 \sim 1200$ 、 $1100 \sim 1800$;(3)最后将对照(2)的实验重复进行 1 次。

1.4 数据库检索

采集的质谱数据用 Bioworks 软件检索酵母数据库。检索时,母离子质量容许误差 1.4 u ,子离子质量容许误差 3.0 u ,漏切位点为 1,半胱氨酸设置为碘乙酰氨乙酰化可变修饰,蛋氨酸设置为氧化可变修饰。蛋白质鉴定标准采用经正、反库评价^[35]后的 95% 置信水平时的鉴定结果, $\Delta C_n \geq 0.1$, $R_{\text{sp}} \leq 4$ 。

2 结果与讨论

2.1 数据置信水平的评价

假阳性率问题普遍存在于规模化蛋白鉴定中^[36],文献^[35]报道假阳性率与许多因素相关,如样品类型、实验条件、动态检测范围等。而在比较性实验中,只有当不同数据间的假阳性率,即对应的置信水平近似时,相应结果之间才具有真正的可比性。因此本实验根据 Qian^[35] 等发展的方法,首先对不同实验条件下的鉴定结果进行置信水平评价,即对质谱数据分别经正、反数据库检索后,利用在一定过滤条件下得到的来自正反库的肽段数目,按式(1)计算假阳性率,得到相同置信水平下所对应的卡值(X_{corr} 值)。

假阳性率 =

$$\frac{\text{鉴定到的来自反库的 unique 肽段数}}{\text{鉴定到的来自正库的 unique 肽段数}} \quad (1)$$

采用不同梯度时间分离的 +1,+2,+3 电荷肽段的 X_{corr} 值列于表 1。可以看出,采用不同梯度时间分离的相同样品,使得鉴定结果均在 95% 可信度时的卡值存在一些差别,这说明在比较性实验中,有必要对不同结果的置信水平进行严格的再评价。在相同置信水平下的蛋白鉴定结果列于表 2。

2.2 应用不同时间梯度洗脱时的蛋白质鉴定结果

如表 2 所示,随着梯度时间的延长,鉴定的非冗余肽段和蛋白簇的数量也随之显著增加。当采用 1 h 梯度时间分离时,鉴定了 146 个蛋白簇;而当梯度时间为 10 h 时,可鉴定 510 个蛋白

表 1 不同梯度时间分离的 +1,+2,+3 电荷肽段的 X_{corr} 值(95% 可信度)

Table 1 X_{corr} values of identified peptides with charges +1, +2 or +3 for yeast samples separated with different gradient elution time at confidence level of more than 95%

梯度时间/h	电荷肽段的 X_{corr} 值		
	+1	+2	+3
1	2.0	2.4	2.6
2	1.8	2.3	2.6
3	2.0	2.2	2.7
4	1.7	2.2	2.7
5	1.9	2.3	2.7
7	1.9	2.2	2.6
10	1.8	2.2	2.7

表 2 不同梯度时间分离时的蛋白鉴定结果(95% 可信度)

Table 2 Peptides and the corresponding protein groups identified employing different gradient elution time at more than 95% confidence level

梯度时间 /h	unique 肽段	非冗余蛋白 group	双肽段以上匹配的蛋白 group 的比例/%
1	369	146	46.54
2	463	217	44.67
3	561	239	46.04
4	963	344	49.96
5	1 049	357	54.89
7	1 333	460	53.04
10	1 654	510	57.46

簇。同样鉴定的非冗余肽段也显著增加,这表明延长梯度时间能够显著提高蛋白鉴定的覆盖率。同时随着梯度时间的延长,双肽段以上匹配蛋白簇的比例也随之增加,从 1 h 梯度分离时的 46.54% 增加到 10 h 梯度分离时的 57.46%,这同样表明梯度时间的延长有助于提高蛋白鉴定的覆盖率。虽然如此,但也不能断定这种趋势会无限发展下去,这是由于梯度时间的延长并不能带来柱容量的无限增加,同时,梯度时间延长造成被质谱瞬间采集的样品浓度的递减。虽然在一定限度内对蛋白鉴定效率没有显著影响,但由于质谱仪灵敏度的限制,也会显著影响梯度时间的继续延长。

2.3 不同分析策略间的比较

不同分析策略的蛋白鉴定结果列于表3,可以看出,使用 $0.8\text{ }\mu\text{g}$ 样本的3次实验平均鉴定391个肽段,对应于155个蛋白簇,而采用相同的梯度时间分析 $2.4\text{ }\mu\text{g}$ 样本鉴定了369个肽段,对应于146个蛋白簇,两者之间没有显著差别,这样就消除了对鉴定结果可能的影响,增强了比较结果的可信度。

表3 不同分析策略的蛋白鉴定结果

Table 3 Peptide and protein group identifications at different analytical parameters

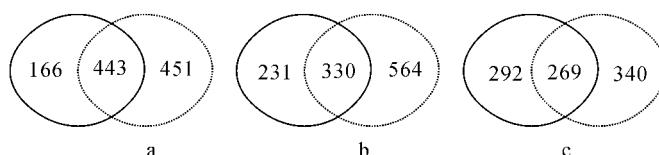
样品量 / μg	梯度时间 /h	质量扫描范围 /(m/z)	unique 肽段	非冗余蛋白 group
2.4	1	300~1 800	369	146
2.4	3	300~1 800	561	239
0.8	1	300~1 800	385	154
0.8	1	300~1 800	394	166
0.8	1	300~1 800	395	145
三次重复总的鉴定肽段数		609	—	
0.8	1	300~800	354	193
0.8	1	700~1 200	464	213
0.8	1	1 100~1 800	303	150
三次分段扫描实验总的鉴定肽段数		894	—	

同时,将 $2.4\text{ }\mu\text{g}$ 样本经3 h梯度分析鉴定的561个肽段,与上述3次重复实验总的鉴定肽段数(609个)相比,两者之间的差别也不显著,这说明将固定量样品平分,并进行简单重复实验的方法,并不能显著提高蛋白的鉴定效率。而在上述重复实验的基础上,将质量扫描范围分为

m/z 300~800、700~1 200和1 100~1 800分析时,3次实验总的鉴定肽段数(894个)有了明显的增加,这说明采用质谱分段扫描的策略在提高蛋白鉴定效率方面具有更为显著的效果,这可能与质量扫描范围细分后,致使质谱扫描速度相对增加有关。另外,与简单重复实验相比,每个分段扫描实验鉴定结果之间的差别十分明显,尤其是当采用质量扫描范围 m/z 700~1 200进行分析时,鉴定的肽段数量最多,这可能与大部分肽段的质荷比正好分布在此范围之间有关。

3 结论

由2.3中的比较可以看出,采用质谱分段扫描的策略在提高蛋白鉴定效率方面具有更为明显的效果;同时,将3次质谱分段扫描实验总的鉴定肽段数(894个)与采用较长梯度时间(3 h)分析时的鉴定肽段数(561个)进行比较,同样可以看出前者具有更为明显的效果;而当将3次简单重复实验总的肽段鉴定情况与采用较长梯度时间分析的肽段鉴定情况进行比较时,发现两者之间没有显著差别,这充分说明在这些提高蛋白鉴定效率的常用策略中,采用质谱分段扫描的策略具有更为显著的效果。尽管如此,互补的研究策略依然是十分必要的。不同分析策略间蛋白鉴定效率的比较示意图示于图1。从图中可以看出,任何一种分析策略的鉴定结果都没有完全覆盖另外一种的鉴定结果,这种互补性有利于提高分析的完整性。因此,在规模化蛋白质组分析中,本着节约时间、提高效率的原则,应当选择更为合适的、互补的研究策略以充分提高分析的完整性。



- a. 三次重复实验(各1 h)总的鉴定肽段(左)与三次分段扫描实验(各1 h,质量扫描范围不同)总的鉴定肽段(右)的比较示意图;
- b. 一次3 h实验的鉴定肽段(左)与三次分段扫描实验(同上)总的鉴定肽段(右)的比较示意图;
- c. 一次3 h实验的鉴定肽段(左)与三次重复实验(同上)总的鉴定肽段(右)的比较示意图

图1 不同分析策略间蛋白鉴定效率的比较示意图

Fig. 1 Comparisons of the identification efficiency by employing different experimental conditions

参考文献:

- [1] TYERS M, MANN M. From genomics to proteomics[J]. *Nature*, 2003, 422: 193-197.
- [2] LIU H B, SADYGOV R G, YATES J R 3rd. A Model for random sampling and estimation of relative protein abundance in shotgun proteomics[J]. *Anal Chem*, 2004, 76(14): 4 193-4 201.
- [3] RIEDEL K, AREVALO-FERRO C, REIL G, et al. Analysis of the quorum-sensing regulon of the opportunistic pathogen *burkholderia cepacia* H111 by Proteomics[J]. *Electrophoresis*, 2003, 24(4): 740-750.
- [4] GÖRG A, OBERMAIER C, BOGUTH G, et al. The current state of two-dimensional electrophoresis with immobilized pH gradients[J]. *Electrophoresis*, 2000, 21(6): 1 037-1 053.
- [5] FREEMAN W M, HEMBY S E. Proteomics for protein expression profiling in neuroscience[J]. *Neurochem Res*, 2004, 29(6): 1 065-1 081.
- [6] CORDWELL S J, NOUWENS A S, WALSH B J. Comparative proteomics of bacterial pathogens [J]. *Proteomics*, 2001, 1(4): 461-472.
- [7] CORTHALS G L, WASINGER V C, HOCHSTRASSER D F, et al. The dynamic range of protein expression: a challenge for proteomic research [J]. *Electrophoresis*, 2000, 21(6): 1 104-1 115.
- [8] BOSSI A, RIGHETTI P G, VECCHIO G, et al. Focusing of alkaline proteases (subtilisins) in pH 10-12 immobilized gradients[J]. *Electrophoresis*, 1994, 15(1): 1 535-1 540.
- [9] SANTONI V, MOLLOY M, RABILLOUD T. Membrane proteins and proteomics: un amour impossible? [J]. *Electrophoresis*, 2000, 21(6): 1 054-1 070.
- [10] GYGI S P, CORTHALS G L, ZHANG Y N, et al. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology [J]. *Proc Natl Acad Sci*, 2000, 97(17): 9 390-9 395.
- [11] YATES J R 3rd. Mass spectrometry and the age of the proteome[J]. *J Mass Spectrom*, 1998, 33 (1): 1-19.
- [12] LINK A J, ENG J, SCHIELTZ D M, et al. Direct analysis of protein complexes using mass spectrometry[J]. *Nat Biotechnol*, 1999, 17: 676-682.
- [13] WASHBURN M P, WOLTERS D, YATES J R 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification tech-
- nology[J]. *Nat Biotechnol*, 2001, 19: 242-247.
- [14] WU C C, MACCOSS M J. Shotgun proteomics: tools for the analysis of complex biological systems[J]. *Curr Opin Mol Ther*, 2002, 4(3): 242-250.
- [15] SCHIRMER E C, FLORENS L, GUAN T L, et al. Nuclear membrane proteins with potential disease links found by subtractive proteomics [J]. *Science*, 2003, 301: 1 380-1 382.
- [16] CHEESEMAN I M, ANDERSON S, JWA M, et al. Phospho-regulation of kinetochore-microtubule attachments by the aurora kinase Ipl1p [J]. *Cell*, 2002, 111: 163-172.
- [17] FLORENS L, WASHBURN M P, RAINES J D, et al. A proteomic view of the plasmodium falciparum life cycle [J]. *Nature*, 2002, 419: 520-526.
- [18] PENG J M, SCHWARTZ D, ELIAS J E, et al. A proteomics approach to understanding protein ubiquitination [J]. *Nat Biotechnol*, 2003, 21: 921-926.
- [19] IDEKER T, THORSSON V, RANISH J A, et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network[J]. *Science*, 2001, 292: 929-933.
- [20] LASONDER E, ISHIHAMA Y, ANDERSEN J S, et al. Analysis of the plasmodium falciparum proteome by high-accuracy mass spectrometry [J]. *Nature*, 2002, 419: 537-542.
- [21] ANDERSEN J S, WILKINSON C J, MAYOR T, et al. Proteomic characterization of the human centrosome by protein correlation profiling [J]. *Nature*, 2003, 426: 570-574.
- [22] MIRZAEI H, REGNIER F. Enhancing electrospray ionization efficiency of peptides by derivatization[J]. *Anal Chem*, 2006, 78(12): 4 175-4 183.
- [23] JACOBS J M, MOTTAZ H M, YU L R, et al. Multidimensional proteome analysis of human mammary epithelial cells[J]. *Journal of Proteome Research*, 2004, 3(1): 68-75.
- [24] SHEN Y F, JACOBS J M, CAMP D G II, et al. Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome[J]. *Anal Chem*, 2004, 76(4): 1 134-1 144.

(下转第236页)

的产地、采收期、药材生长时期的气候以及分析测试条件的不同所引起的。

此外,超临界 CO₂ 萃取的荜茇挥发油,运用 TLC 及 HPLC 法均能检测出药效成分胡椒碱,采用 GC/MS 法同样能检测出胡椒碱的存在,且含量占整个挥发油组分的 29.274%。由于胡椒碱对强光、高温均不稳定,易分解散失,这可能也是 SD 法提取的荜茇挥发油中未能检测到胡椒碱存在的原因之一。

4 结 论

超临界 CO₂ 萃取时间短、温度低、出油率高、无溶剂残留,是提取荜茇挥发油的理想方法。但根据目标产物的不同,可采取不同的提取方法对荜茇中的某些成分进行选择性提取。

参考文献:

- [1] 国家中医药管理局《中华本草》编委会. 中华本草: 第八卷 [M]. 上海: 上海科学技术出版社, 1999; 434.
- [2] 国家药典委员会. 中国药典:一部 [M]. 北京: 化学工业出版社, 2005; 163.
- [3] 吴知行, 杨尚军, 巴图仑, 等. 荜茇挥发油的成分分析 [J]. 中草药, 1994, (9): 5.
- [4] 孙友富, 樊菊芬. 荜茇挥发油化学成分的研究 [J]. 中医杂志, 1981, (12): 65-66.
- [5] 李熙灿, 赵小军, 谢学明, 等. 荜茇挥发油清除自由基作用及其与分子结构的关系 [J]. 中药新药与临床药理, 2006, 17(3): 218-221.
- [6] 国家药典委员会. 中国药典:一部 [M]. 北京: 化学工业出版社, 2000; 附录 XD, 附录 19.

(上接第 230 页)

- =====
- [25] SHEN Y F, ZHANG R, MOORE R J, et al. Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics [J]. Anal Chem, 2005, 77(10): 3 090-3 100.
 - [26] SHEN Y F, TOLIC N, MASSELON C, et al. Ultrasensitive proteomics using high-efficiency on-line micro-SPE-NanoLC-NanoESI MS and MS/MS[J]. Anal Chem, 2004, 76(1): 144-154.
 - [27] DURR E, YU J Y, KRASINSKA K M, et al. Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture[J]. Nat Biotechnol, 2004, 22: 985-992.
 - [28] BLONDER J, HALE M L, LUCAS D A, et al. Proteomic analysis of detergent-resistant membrane rafts[J]. Electrophoresis, 2004, 25 (9): 1 307-1 318.
 - [29] BLONDER J, RODRIGUEZ-GALAN M C, LUCAS D A, et al. Proteomic investigation of natural killer cell microsomes using gas-phase fractionation by mass spectrometry[J]. Biochim Biophys Acta, 2004, 1 698(1): 87-95.
 - [30] BRECI L, HATTRUP E, KEELER M, et al. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing[J]. Proteomics, 2005, 5(8): 2 018-2 028.
 - [31] ELIAS J E, HAAS W, FAHERTY B K, et al. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations[J]. Nature methods, 2005, 2: 667-675.
 - [32] BODNAR W M, BLACKBURN R K, KRISE J M, et al. Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage[J]. J Am Soc Mass Spectrom, 2003, 14(9) : 971-979.
 - [33] SPAHR C S, DAVIS M T, MCGINLEY M D, et al. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry I. profiling an unfractionated tryptic digest[J]. Proteomics, 2001, 1(1) : 93-107.
 - [34] YI E C, MARELLI M, LEE H, et al. Approaching complete peroxisome characterization by gas-phase fractionation [J]. Electrophoresis, 2002, 23(18): 3 205-3 216.
 - [35] QIAN W J, LIU T, MONROE M E, et al. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome [J]. Journal of Proteome Research, 2005, 4(1): 53-62.
 - [36] ADAMSKI M, BLACKWELL T, MENON R, et al. Data management and preliminary data analysis in the pilot phase of the HUPO plasma proteome project [J]. Proteomics, 2005, 5(13): 3 246-3 261.