

基质辅助激光解吸电离质谱和电喷雾电离质谱 在辣根过氧化物酶糖肽结构分析中的应用

陈瑶函^{1,2}, 晏国全¹, 周新文², 杨芃原^{1,2*}

(1. 复旦大学化学系, 上海 200032; 2. 复旦大学生物医学研究院, 上海 200032)

摘要 :糖链结构的质谱解析是今后糖蛋白分析中的重要研究内容,其中完整糖肽的分析,由于可以同时获得糖基化位点和对应糖链的结构信息,更具有重要意义和研究前景。本工作对质谱软电离技术在完整糖肽分析中的应用进行了研究,其中包括了基质辅助激光解吸电离(matrix-assisted laser desorption ionization, MALDI)和电喷雾电离(electrospray ionization, ESI)技术。通过平行使用两种串联质谱(tandem mass spectrometry, MS/MS)分析策略:MALDI-MS/MS和ESI-MS/MS对目标糖蛋白——辣根过氧化物酶进行分析,并讨论了其互补性。结果表明,MALDI和ESI技术各有优劣,结合串联质谱分析,可获得糖肽的糖链结构信息,两条路线互补使用,在揭示蛋白质糖基化修饰(位点和结构)的研究中十分必要。

关键词 :基质辅助激光解吸电离质谱;电喷雾电离质谱;串联质谱;液相色谱;辣根过氧化物酶;糖肽

中图分类号 :O658 文献标识码 :A 文章编号 :1000-8713(2010)02-0135-05

Combination of matrix-assisted laser desorption ionization and electrospray ionization mass spectrometry for the analysis of intact glycopeptides from horseradish peroxidase

CHEN Yaohan^{1,2}, YAN Guoquan¹, ZHOU Xinwen², YANG Pengyuan^{1,2*}

(1. Department of Chemistry, Fudan University, Shanghai 200032, China;

2. Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China)

Abstract : The mass spectrometric analysis of glycan composition and structure is a difficult but essential part in future study following the present glycoprotein identification. Intact glycopeptides analysis is an attracting field, in considering its capability to provide glycosite and corresponding glycan structure information at the same time. Mass spectrometry has been proven to be an important and a key tool for glycan analysis over the past few years. Making use of two soft ionization techniques—matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), through tandem mass spectrometry (MS/MS) analysis, the information for glycans of the glycopeptides can be obtained in the investigation. Meanwhile, by comparing the MALDI-MS/MS and ESI-MS/MS approaches using model glycoprotein (horseradish peroxidase, HRP), the complementary results have been verified experimentally. We believe that the combination of the two techniques is necessary and will provide useful information for the understanding of protein glycosylation.

Key words : matrix-assisted laser desorption ionization (MALDI); electrospray ionization (ESI); tandem mass spectrometry (MS/MS); liquid chromatography (LC); horseradish peroxidase; glycopeptide

蛋白质的糖基化是生物体中最普遍和最重要的蛋白质翻译后修饰之一,人体中大约有50%的蛋白

质发生糖基化^[1]。糖基化与蛋白质的物理化学性质、生物功能密切相关,涉及蛋白质折叠、定位、信号

* 通讯联系人:杨芃原,教授,博士生导师。Tel : (021) 54237416, E-mail : pyyang@fudan.edu.cn.

基金项目 :国家自然科学基金课题(Nos. 20735005, 20975024), 国家“863”课题(Nos. 2006AA02A308)和上海市科委课题(No. 07JC140030).

收稿日期 2009-10-17

传导、细胞黏附等等生物过程的众多方面。一些疾病也和蛋白质的糖基化水平、糖链结构的变化相关^[2,3]。糖蛋白的研究正得到越来越多的关注。

目前研究较多的糖基化蛋白质主要是 *N*-链接糖蛋白和 *O*-链接糖蛋白。前者的糖基化发生在氨基酸序列 Asn-Xxx-Ser/Thr (Xxx ≠ Pro) 的 Asn 残基上,糖链具有由 2 个乙酰葡糖胺 (*N*-acetylglucosamine, GlcNAc) 和 3 个甘露糖 (mannose, Man) 组成的五糖核心结构;后者发生在 Ser 或 Thr 残基上,没有明显的氨基酸序列特征,且糖链核心结构不单一。蛋白质的糖基化具有宏观和微观的不均一性,即糖基化位点的多样性以及同一位点糖链结构的多样性,这都增加了糖蛋白研究的难度。

质谱分析作为蛋白质组学的关键技术,同样可为其重要分支——翻译后修饰蛋白质组学研究提供大量的信息。两种软电离技术——电喷雾电离 (electrospray ionization, ESI) 和基质辅助激光解吸电离 (matrix-assisted laser desorption ionization, MALDI) 技术,分别和与其兼容的各种上游分离纯化技术结合,成为蛋白质组学研究的重要路线。尽管糖蛋白在生物体中比例较高,但是从肽段水平上来看,糖肽的比例仅能占到 2% ~ 5%^[4]。由于大量的非糖肽段的存在,并且非糖肽段的离子化效率远远高于糖肽,所以导致在对复杂体系进行研究时,糖肽的信号被严重抑制,大大降低了结构鉴定的可能。因此,对于糖蛋白,一般在质谱分析时需要经过前期的分离富集等处理。目前,结合各种富集方法,高通量的糖蛋白鉴定已得到实现。糖基化位点,尤其是 *N*-糖基化位点的鉴定,也已通过糖链切除伴随位点标记的方式(如 PNGase F 酶切在 *N*-糖基化位点上引起氨基酸残基 Asn 到 Asp 的变化)实现了通量化^[5-8]。

糖链结构的质谱分析方式主要有两种:一种是将已从蛋白质或肽段上切除的糖链进行衍生化,通过凝集素、色谱、质谱等手段进行分析,即糖组学 (glycomics) 分析;另一种方式不将糖链与蛋白质或肽段分离,直接进行完整的糖肽的串联质谱分析 (tandem mass spectrometry, MS/MS), 获得肽段序列和糖链结构信息。前者可以针对性地对某类糖型 (pattern) 进行研究,或者对整个体系中某些糖型的变化趋势进行反映,现已形成大量的研究成果^[9-11];后者可以将位点和糖链结构对应起来^[12,13], 提供了解糖链与蛋白质功能关系的重要信息,但是由于受到分离方式、富集特异性以及糖肽在质谱中响应等的限制,目前该领域还处在不断的发

展之中,实现高通量的分析仍有距离^[14]。

鉴于从完整糖肽获得糖基化位点和结构信息的重要性,我们对最常用的两种软电离离子化方式在完整糖肽分析中的应用进行了比较。采用 MALDI-MS/MS 和 LC/ESI-MS/MS 两条路线,用一个具有多个糖基化位点的糖蛋白——辣根过氧化物酶 (horseradish peroxidase, HRP) 作为测试对象,目标是准确、详尽地获得酶解产物中的糖基化信息,以对实际样品中目标糖蛋白的研究进行指导。结果表明, MALDI 和 ESI 两种模式均有其优点,但要全面获得糖肽信息,两种方法的结合使用是十分必要的。

1 实验部分

1.1 试剂与仪器

标准蛋白质辣根过氧化物酶、酶解用碳酸氢铵 (NH_4HCO_3)、2,5-二羟基苯甲酸 (2,5-dihydroxybenzoic acid, DHB) 均购自 Sigma-Aldrich 公司。测序级胰蛋白酶 (sequencing grade modified trypsin) 购自 Promega 公司。甲酸、三氟乙酸购自 Fluka 公司。乙腈购自 Fisher Scientific 公司。实验用水为 Milli-Q Filtration System (Millipore 公司) 二次纯化水。其他试剂若无特别说明均为国产分析纯。

AXIMA-QIT™ MALDI TOF MS 购自 Shimadzu Biotech 公司 (Kyoto, Japan)。LTQ-Orbitrap XL System 购自 Thermo Electron 公司 (Bremen, Germany)。

1.2 样品前处理

标准蛋白质 HRP 溶解于 25 mmol/L NH_4HCO_3 中,使蛋白质的终质量浓度为 50 ~ 100 ng/ μL ,热变性^[15]。溶液冷却后,加入胰蛋白酶(按照胰蛋白酶与标准蛋白质质量比为 1:50 的比例加入),于 37 °C 下酶解过夜。将酶解溶液真空离心干燥,以备后续实验使用。

1.3 色谱和质谱条件

MALDI 质谱分析在 AXIMA-QIT™ MALDI TOF MS 上进行。DHB 溶解于含 0.1% (体积分数,下同) 三氟乙酸的 30% 乙腈水溶液中,终质量浓度为 12.5 mg/mL,作为基质。点样方式选择“基质-样品-基质”式^[15]。另外,为了提高每次实验样品点之间的重复性,在最后一层基质结晶以后,用小于基质体积的乙醇对样品点进行重结晶,以提高结晶的均一性^[16]。MALDI 质量扫描范围分为 m/z 750 + , m/z 2 000 + , m/z 3 000 + 几个区域,依据母离子和碎片离子的质量选择最佳的质量扫描范围。

ESI 质谱分析在 LTQ-Orbitrap XL Systems 上

进行。真空离心干燥的蛋白酶解肽段溶于 0.1% 甲酸水溶液后上样,通过 Waters 公司的 nanoACQUITY UPLC BEH C18 柱(200 mm × 75 μm, 1.7 μm)分离后,进行在线电喷雾串联质谱分析。流动相 A 为含 0.1% 甲酸的水,流动相 B 为含 0.1% 甲酸的乙腈,梯度洗脱程序为 50 min 内流动相 B 的体积分数由初始的 5% 上升至最终的 50%,流速为 300 nL/min。肽段串级分析为 data-dependent 模式(data-dependent MS/MS acquisition),选择信号强度在前 6 位的母离子,动态排除 0.5 min。

2 结果与讨论

由于研究目标是准确、详尽地获得目标蛋白的糖基化信息以对实际样品的研究进行指导,因此在对 HRP 进行分析时,将其看作一个经过分离得到的、确定发生 *N*-糖基化的目标蛋白质,且蛋白质的序列已知(质谱鉴定,此处略),后续的糖肽结构分析在此基础上进行。

2.1 MALDI-MS/MS 糖肽分析

对于已知蛋白质的糖肽, MALDI-MS/MS 分析中未进行糖肽水平上的富集,关键是在肽质谱指纹(peptide mass fingerprint, PMF)图谱中寻找糖肽并手动进行串联质谱分析。研究从已知的蛋白质序列入手,找出包含潜在糖基化位点的理论酶解肽段,在大于这些理论肽段相对分子质量(M_r)的范围内进行 PMF 的筛选。对于 *N*-糖基化,糖链的存在一般会造成至少 892 (五糖核心 M_r)的 M_r 增量。基于这种思路,我们对大量的在一级谱图中有足够信号强度的质谱峰进行了串联质谱分析。结果表明,具有特征糖链碎裂的 MS/MS 谱图基本来自 m/z 2 000⁺ 范围。值得说明的是,HRP 无法用常见的糖苷酶 PNGase F 切除糖链的 *N*-糖蛋白,因此寻找它的糖肽比较繁琐(或需选用其他的糖苷水解酶)。对于其他 PNGase F 酶可以发挥作用的糖蛋白,对比其切除糖链前后的 PMF 谱图,可以较快捷地寻找其糖肽。

在 CID (collision induced dissociation) 模式下,糖苷键比肽键更易碎裂。糖肽的最终确定是根据 MALDI-MS/MS 分析中糖链的碎裂特征实现的^[17-19]。图 1 列举了两张糖肽串联质谱图,它们反映了将 MALDI-MS/MS 用于糖肽分析时产生的普遍特征。组成糖链的单糖具有特定的 M_r , 产生了特征质谱峰间距(*N*-乙酰葡萄糖胺(GlcNAc, 203),甘露糖(mannose, 162),海藻糖(fucose, 146),木糖(xylose, 132)等),如图 1 中两谱图峰间水平箭头上方数值标示。与糖基化位点 Asn 直接连接的 GlcNAc

(inner-most GlcNAc),在串联质谱分析时发生跨环断裂(cross-ring cleavage)会形成具有特定的质量差(83 和 120)的 3 个连续峰,如图 1 中两谱图中标注为 P, P*, P-GlcNAc 的质谱峰,分别代表了上述 3 个特征连续峰。 M_r 最小者(P)即可指示该糖肽肽段部分的 M_r ,若与预测的潜在糖基化位点所在肽段吻合,则初步获得了糖基化位点的信息(一般适于只含有一个潜在位点的肽段)。另外,在低 M_r 段也能发现少量对应于该肽段的部分酰胺键碎裂产生的离子(b 和 y 离子),可进一步地证实糖基化位点。

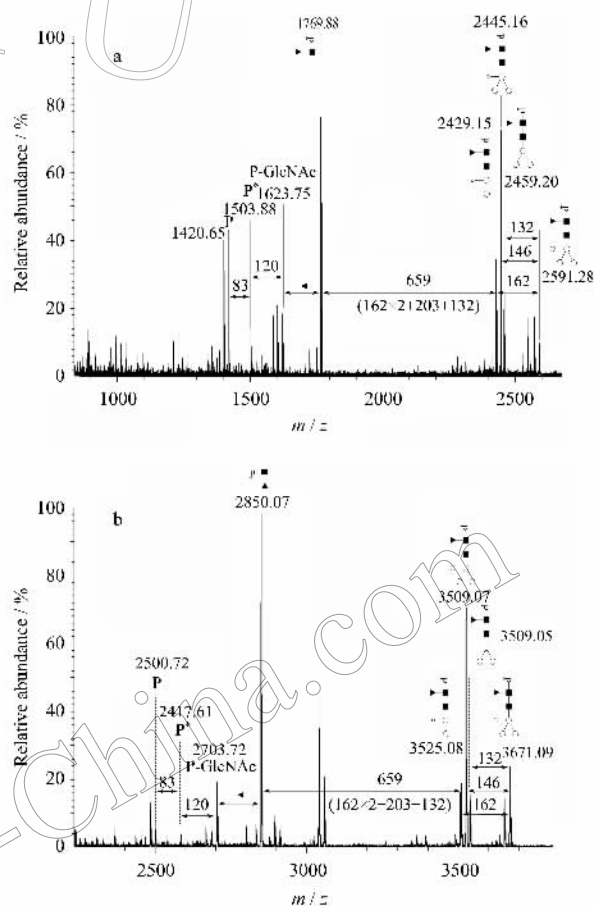


图 1 糖肽((a) m/z 2591.3 和 (b) m/z 3671.1)的 MALDI-MS/MS 谱图

Fig. 1 Tandem mass spectra of two glycopeptides of (a) m/z 2591.3 and (b) m/z 3671.1 by MALDI analysis

■ GlcNAc; ○ mannose (Man); ▲ fucose (Fuc); ☆ xylose (Xyl).

2.2 HPLC/ESI-MS/MS 糖肽分析

LC 连接 ESI 质谱在线分析,理论上来说,只要样品量足够多和液相色谱的分离效果足够好,自动获取的串联质谱图中即包含了糖肽的信息。与 MALDI-MS/MS 类似, CID 模式下 ESI-MS/MS 分析中的糖肽谱图也具有一定的特征,除了同样具有反映组成糖链的单糖的 M_r 的特征质谱峰间距之外,

在谱图的低分子质量端,还具有特征地从糖肽主体碎落的寡糖碎片离子(glycan fragment ions,如 $[\text{GlcNAc} + \text{H}]^+$ (204), $[\text{GlcNAcMan} + \text{H}]^+$ (366), $[\text{GlcNAcManMan} + \text{H}]^+$ (528)等)。通过提取离子色谱图(extracted ion chromatogram, XIC),可以初步确定糖肽的色谱保留时间,对产生这些特征寡糖碎片离子的串联质谱图进行筛选,可进一步缩小寻找范围。

图 2 列举了两张糖肽 ESI-MS/MS 串联质谱图。带多电荷的母离子在碎裂过程中大致产生了两类容易识别的子离子,即寡糖碎片离子和糖肽相关离子(包括丢失部分寡糖的糖肽)。前者(图 2 中实线方框标识部分)即用作 XIC 筛选,后者在谱图中表现为具有单糖特征的质量差值的一组质谱峰(图 2 中虚线方框标识部分)。另外,仅带最内侧乙酰葡糖胺的糖肽峰(图 2 中椭圆框标识部分)在实验中往往呈现出较高的峰强,该离子常被挑选作为第三级质谱分析对象,以进一步确定糖基化位点。

2.3 MALDI 和 ESI-MS/MS 解析糖肽的比较

我们对两种方式产生的串联质谱图进行了详尽的筛选和确认,以尽可能获得全面的目标蛋白质的糖基化信息,结果如表 1 所示,得到确认的位点(包括其糖链结构)为 7 个。根据 NetNGlyc(蛋白质的 N-型糖基化位点预测软件(<http://www.cbs.dtu.dk/services/NetNGlyc/>))的预测,HRP 氨基酸序列中的 9 个 NXS/T 结构发生糖基化的可能性(表 1 中的 potential)各有不同,例如, N_{287} 在进行蛋白质鉴定时即已被覆盖,说明其未发生糖基化(或仅部分发生糖基化),这与预测结果相符(“-”),表现合理。另外,位点 N_{187} (“++”)所在糖肽实际上也被检测到,但是由于该糖肽包含两个发生糖基化的位点(另一个为 N_{199}),我们未能通过手工分析从其谱图中获取足够的信息说明该位点的糖链结构。

表 1 HRP 糖基化位点和相应的糖链结构及 NetNGlyc 预测结果*
Table 1 Glycosites and corresponding glycan structures and prediction results by NetNGlyc*

Position	Potential	Jury agreement	N-Glyc result	Glycan structure
N_{14} VSN	0.7981	9/9	+++	
N_{58} TTS	0.7719	9/9	+++	
N_{159} RSS	0.7112	9/9	++	
N_{187} FSN	0.6657	9/9	++	NA
N_{199} TTY	0.4969	4/9	-	
N_{215} LSA	0.6559	9/9	++	
N_{256} ATD	0.6101	7/9	+	
N_{269} STQ	0.6109	7/9	+	
N_{287} ITP	0.1544	9/9	---	NA

* Any potential crossing the default threshold of 0.5, represents a predicted glycosylated site. The "potential" score is the averaged output of nine neural networks. For further information, the Jury agreement column indicates how many of the nine networks support the prediction. The N-Glyc result column shows one of the following outputs for the predictions indicating glycosylated sites: +, Potential > 0.5; ++, Potential > 0.5 and Jury agreement (9/9) or Potential > 0.75; +++, Potential > 0.75 and Jury agreement; +++, Potential > 0.90 and Jury agreement and non-glycosylated sites; -, Potential < 0.5; --, Potential < 0.5 and Jury agreement (Potential < 0.5 for all the nine); ---, Potential < 0.32 and Jury agreement; NA, not available. The symbols are the same as in Fig. 1.

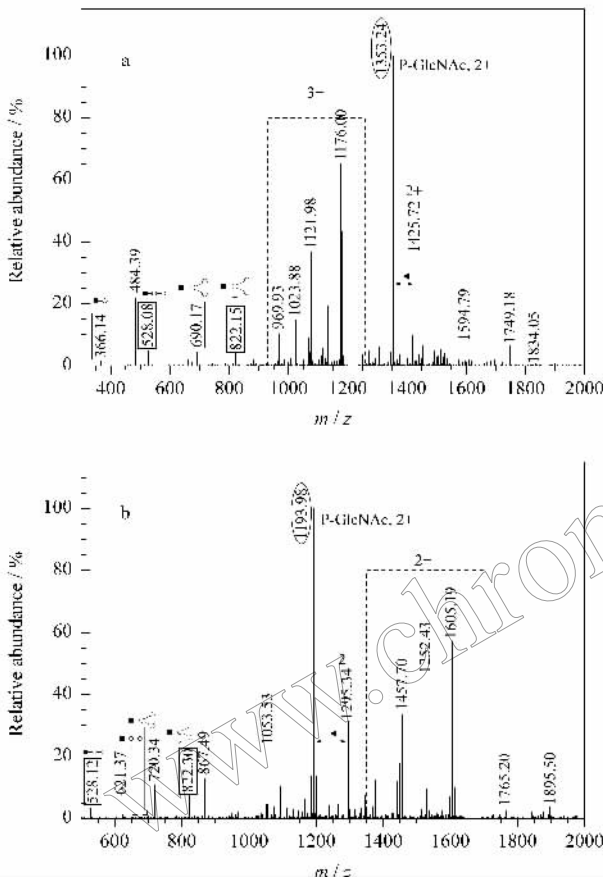


图 2 糖肽((a) m/z 1225.2 正三价离子(3^+)和(b) m/z 1678.2 正二价离子(2^+))的 ESI-MS/MS 谱图

Fig. 2 Tandem mass spectra of two glycopeptides of (a) m/z 1225.2, charge state 3^+ and (b) m/z 1678.2, charge state 2^+ by ESI analysis

The symbols are the same as in Fig. 1.

MALDI 和 ESI 质谱在 CID 模式下产生的糖肽串联质谱图均主要表现为糖链的碎裂,而肽段部分碎裂处于次要。在涉及具体的结构解析和位点确定时,两种方式各有优劣。首先,对于未在肽段水平上进行富集样品,在定位糖肽方面,LC-ESI 方式较 MALDI 更客观、方便;ESI-MS/MS 低分子质量端呈现的寡糖碎片是对糖肽进行定位和确认的重要依据之一,XIC 方式可便捷地提供这样的信息,这较于 MALDI 方式(受到基质峰和扫描范围限制)只能在 PMF 中手工寻找更为方便。其次,对于糖链结构的解析,ESI 方式提供的糖链组成信息更完全:糖肽相关峰特别是糖链最内侧 GlcNAc 后端糖链丢失后的相关的质谱峰(如图 3 中虚线框内部分),在 ESI-MS/MS 中表现为有特征质量差值的峰连续出现(如图 2 虚线框标识);而 MALDI-MS/MS 中呈现出一个断层(如图 1 中 659 的质量差),虽然单糖有特征质量,可以从质量差上面推导出其组成,但若遭遇结构比较复杂的糖肽其可靠性则有待考察。再次,对于糖基化位点的确定,MALDI 方式能较为准确地指示肽段的 M_r :MALDI-MS/MS 引起最内侧 GlcNAc 跨环断裂,产生具有特征质量差值的 3 个连续峰(P , P^* , P -GlcNAc)是 N -糖基化肽碎裂的普遍特征,而 ESI 方式在肽段部分虽然也有一定的规律(如图 2),在实验中 P -GlcNAc 均表现为强度最强的峰,但仅凭强度来判断比较牵强,易受干扰,一般需要三级质谱确定(三级谱图中会出现类似 MALDI 二级谱图的 3 个特征连续峰,图略)。

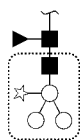


图 3 HRP 糖链结构示意图

Fig. 3 Glycan structure of HRP
The symbols are the same as in Fig. 1.

另外,两者同样都遇到谱图解析的问题:首先,此处 HRP 糖链较短,结构容易根据生物学合成方式推断,对于其他带较长糖链的肽段,连接方式的各种可能性大量增加,仅靠串联质谱图中的单糖特征质量差来判断,信息量显然不足,这需要软件的协助以及结合其他分析手段(如气相色谱-质谱联用等);其次,谱图中存在的可以进一步证明肽段来源和糖基化位点的 b 和 y 离子(包括带糖链或不完整糖链的 b 和 y 离子),在人工分析不完善的情况下,若有软件协助挖掘,则可充分利用串联质谱图获得更可靠的糖基化信息。

3 结论

完整糖肽结构的分析因其重要性备受科研工作者的关注。由于目前没有一种方法能全面地获得糖肽结构的所有信息,所以各种手段的联合使用显得尤为重要。除了互补地使用 MALDI 和 ESI 两种软电离技术之外,负离子检测模式^[20]以及 CID 和 ETD (electron transfer dissociation) 两种离子化方式的联合^[21],虽还未形成通量化的成果,但也正在越来越多地应用到这个领域。另外,糖肽解析软件也亟待开发完善,以便快速、客观、全面地获得糖肽结构信息。

参考文献:

- [1] Apweiler R, Hermjakob H, Sharon N. *Biochim Biophys Acta*, 1999, 1473(1):4
- [2] Dube D H, Bertozzi C R. *Nat Rev Drug Discov*, 2005, 4(6):477
- [3] Taylor A D, Hancock W S, Hincapie M, et al. *Genome Med*, 2009, 1(6):57
- [4] Sun B Y, Ranish J A, Utleg A G, et al. *Mol Cell Proteomics*, 2007, 6(1):141
- [5] Liu T, Qian W J, Gritsenko M A, et al. *Mol Cell Proteomics*, 2006, 5(10):1899
- [6] Schiess R, Mueller L N, Schmidt A, et al. *Mol Cell Proteomics*, 2009, 8(4):624
- [7] Cao J, Shen C P, Wang H, et al. *J Proteome Res*, 2009, 8(2):662
- [8] Chen R, Jiang X N, Sun D G, et al. *J Proteome Res*, 2009, 8(2):651
- [9] Alvarez-Manilla G, Warren N L, Abney T, et al. *Glycobiology*, 2007, 17(7):677
- [10] Toyoda M, Ito H, Matsuno Y, et al. *Anal Chem*, 2008, 80(13):5211
- [11] Kim Y G, Gil G C, Jang K S, et al. *J Mass Spectrom*, 2009, 44(7):1087
- [12] Seipert R R, Dodds E D, Levrilla C B. *J Proteome Res*, 2009, 8(2):493
- [13] Peltoniemi H, Joenvaara S, Renkonen R. *Glycobiology*, 2009, 19(7):707
- [14] Dalpathado D S, Desaire H. *Analyst*, 2008, 133(6):731
- [15] Zhou W, Liu X H, Zhou X W, et al. *Chinese Journal of Chromatography* (周玮, 刘晓慧, 周新文, 等. 色谱), 2007, 25(5):623
- [16] Morelle W, Michalski J C. *Nat Protoc*, 2007, 2(7):1585
- [17] Wührer M, Catalina M I, Deelder A M, et al. *J Chromatogr B*, 2007, 849(1/2):115
- [18] Wada Y, Tajiri M, Yoshida S. *Anal Chem*, 2004, 76(22):6560
- [19] Domon B, Costello C E. *Glycoconjugate J*, 1988, 5(4):397
- [20] Motari E, Zheng X C, Su X D, et al. *Am J Biomed Sci*, 2009, 1(1):1
- [21] Perdivara I, Petrovich R, Allinquant B. *J Proteome Res*, 2009, 8(2):631