



论文

人血清内源性多肽无标记定量分析新策略

朱俊, 王方军, 董小莉, 叶明亮, 邹汉法*

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, 国家色谱研究分析中心, 大连 116023

*通讯作者, E-mail: hanfazou@dicp.ac.cn

收稿日期: 2010-01-18; 接受日期: 2010-02-07

英文全文见: *Science China Chemistry*, 2010, 53(4): 759–767

摘要 多肽组学(Peptidomics)以其在疾病标志物发现中所起的重要作用正受到越来越多的关注。然而, 目前尚缺少针对多肽组学的高效定量方法。本文建立了一种基于纳升液相色谱与四极杆飞行时间质谱联用(μ LC-Q-TOF-MS/MS)系统的色谱峰面积无标记多肽定量方法。该方法首先对两组样品中的多肽进行相对定量, 然后选取相对量大于 2 或小于 0.5 的多肽, 用纳升液相色谱与串联二级质谱(μ LC-MS/MS)的选择性数据依赖模式(DDA)进行序列鉴定。将不同量标准样品牛血清白蛋白(BSA)酶解液加入到相同量的正常人血清样品中, 然后用该方法进行定量分析, 得到的定量结果平均相对偏差为 6.42%。该方法分别被用于健康人血清与肝癌细胞(HCC)以及乳腺癌病人血清中内源性多肽的定量比较分析, 并成功鉴定了血清中一些与 HCC 和乳腺癌相关的内源性多肽。由于 Q-TOF-MS/MS 的高分辨率和宽质量检测范围, 价态高达 +5 和分子量超过 4000 的多肽也可被可靠地进行定性和定量分析。实验结果进一步表明在多肽组学分析中 μ LC-Q-TOF-MS/MS 比传统的 MALDI-TOF-MS 具有更高的检测灵敏度。

关键词内源性肽
血清
无标记定量
肝细胞癌
乳腺癌

1 引言

随着基因组学和蛋白质组学的发展, 目前已经能够对病人进行分子水平的诊断并进行个性化的治疗, 并且在治疗过程中根据其不同的反应而对其进行分类^[1-4]。多肽在人体生理和病理过程中起着重要的作用, 比如肽类激素, 神经肽, 细胞因子和酶抑制剂等^[5]。为了对这些多肽进行研究, 本世纪初出现了一个新兴学科—多肽组学^[6-8]。它的目标在于发现重大疾病早期诊断的多肽类生物标志物, 如癌症标志物。血液作为人类细胞, 组织和器官之间的主要枢纽, 可能存在着能够预示人体生理和病理变化的多肽类物质^[9]。因此, 简单的血液样品, 如血清和血浆, 被广泛地应用于多肽组学的研究^[10-13]。

在“鸟枪法”蛋白质组学(shotgun proteomics)中, 随着液相色谱-串联质谱(LC-MS/MS)等分析技术的

发展, 单次进样分析就能鉴定到成千上万个蛋白质酶解肽段^[14, 15]。然而, 内源性多肽的分析方法相对于“鸟枪法”蛋白质组分析有一些不同, 例如二维凝胶电泳质谱联用技术就不适于分子量小于 10 kDa 多肽的分离分析; 在分析蛋白酶解样品时广泛使用的离子阱质谱仪由于分辨率限制, 对高价态内源性多肽的检测有局限性^[5]。因此, 针对生物样品中的多肽物质, 发展出了各种定性和定量的分析方法。如 Petricoin 等人应用表面增强激光解吸离子化-飞行时间质谱(SELDI-TOF-MS)研究了卵巢癌病人血清中多肽和蛋白的降解通路, 并发展出一种用于区分癌症和非癌症的生物信息学工具^[16]。但是, 用于癌症诊断的几个荷质比(m/z) 为 534, 989, 2111, 2251 和 2465 离子峰在一级质谱不能进行有效的鉴定, 只能采用基质辅助激光解吸离子化-飞行时间串联质谱(MALDI-TOF-MS/MS)对内源性多肽进行鉴定并发现了一些

潜在的多肽类生物标志物^[17, 18].

虽然 MALDI/SELDI-TOF-MS/MS 能直接用于临床样品的高通量多肽鉴定, 但是该方法受离子化效率, 基质抑制效应以及最佳激光解吸样品点选择的影响, 只能被认为是一种半定量方法^[19~21]. 因此, 我们采用同位素标记(iTRAQ)的方法来对 MALDI-TOF-MS/MS 鉴定到的血浆内源性磷酸肽进行相对定量, 证明用 iTRAQ 定量磷酸化肽的结果和 MALDI 结果基本一致^[22].

无标记定量方法也同样被用于多肽组学分析, 并且更适合于生物标志物的发现, 因为该方法能够对任意数目的样品进行定量分析. 相比于各种同位素标记定量方法, 无标记定量方法更加简单廉价, 不同样品中大量多肽的相对变化值通过比较相应多肽的 LC-MS/MS 质谱峰离子流峰强度就能够直接得到^[23]. 通常, 该方法需要一套完整的统计方法以及计算软件来对质谱离子流峰进行检测和比较^[24~26]. Quintana 等人报道了慢性移植肾失功(CAD)病人尿液多肽组无标记定量研究, 并采用虚拟多反应监测(pseudo MRM)对定量结果进行验证^[27]. Silva 等采用纳升超高效液相色谱-四极杆飞行时间质谱联用技术(μ UPLC-Q-TOF-MS)得到高重复的色谱分离和质量分辨率, 这两点对准确定量至关重要^[28]. 该方法中质谱采用高低能量交替扫描(MS/MS^E)的平行碰撞诱导解离方式, 在 μ UPLC- MS/MS^E 实验中能够给出不同离子更精确的单离子流强度值^[29], 而精确质量和保留时间数据对(EMRT)直接反应样品中相应多肽的相对含量. 虽然这种定量策略已经商品化并且成功应用于蛋白酶解多肽的鉴定和定量^[30~34], 但对非酶解的内源性多肽却存在无法鉴定的问题(Waters 公司针对 MS/MS^E 数据的搜库软件不支持非酶切的内源性多肽搜索).

在本工作中, 我们利用介孔硅胶材料 MCM-41 高度有序孔道结构所具有的体积排阻效应, 选择性地提取人血清中的低分子量多肽^[35, 36]. 随后, 富集到的多肽样品采用 μ LC- MS/MS^E 系统分析并用商品化的 PLGS 系统(Waters, USA)对不同样品中的每个 EMRT 进行相对定量分析^[28]. 最后, 选择相对量大于 2 或小于 0.5 的 EMRT 进行 μ LC- MS/MS 选择性 DDA 模式多肽定性, 并通过比较该实验与 μ LC- MS/MS^E 实验所得的质量精度和保留时间的数值对结果进行验

证. 将该方法用于标准样品牛血清白蛋白(BSA)的定量分析, 得到平均相对偏差小于 10%的定量结果. 随后将该方法用正常人血清和 HCC 以及乳腺癌病人血清中内源性多肽的定量比较研究, 并成功鉴定了血清中一些与 HCC 和乳腺癌相关的潜在多肽类生物标志物.

2 实验部分

2.1 试剂

胰蛋白酶、牛血清白蛋白(BSA)购自 Sigma(MO, USA). 二硫苏糖醇(DTT)以及碘代乙酰胺购自华美生物工程公司(北京). 谷氨酸纤维多肽 B(Glu-Fibrinopeptide B, GFP)购自 Waters 公司(Manchester, UK). 甲酸(质谱纯)购自 Fluka 公司; 乙腈(色谱纯)购自德国 Merck 公司; 其他试剂均为分析纯. 所有实验用水均由 Mill-Q(Millipore Co., USA)超纯水净化系统制备.

2.2 多肽提取和样品预处理

取 10 mg MCM-41 分散在 0.5 mL 蒸馏水中, 与 0.5 mL 稀释过的人血清样品(0.2 mL 血清用蒸馏水稀释至 0.5 mL)混合, 在室温下(25 °C)振荡 1 h. 然后 15000 g 下离心 5 min, 去除上清液, 留下的介孔材料分别用 1 mL 蒸馏水清洗三次. 随后介孔材料上富集的多肽用 1 mL 50%乙腈水溶液进行洗脱, 离心收集上清液冻干. 将冻干的样品用 50 μ L 0.1%甲酸水溶液复溶, 并保存在-20 °C下备用.

1 mg BSA 用 1 mL 蛋白变性液(8 M 尿素, 50 mM Tris/HCl, pH 8.2)溶解, 5 mM DTT 37 °C 还原 2 h, 接着加入 10 mM 碘代乙酰胺并在室温避光反应 40 min. 所得溶液用 50 mM Tris/HCl (pH 8.2)稀释至 1 M 尿素, 加入胰蛋白酶(1/25, w/w), 在 37 °C 下酶解过夜. 最后 BSA 酶解液用实验室自制的 C18 固相萃取柱纯化, 冻干后用 0.1%甲酸水溶液复溶备用.

我们使用两种标准样品来确定该定量方法的可靠性. 标准样品 1 由 30 μ L 健康人血清多肽、6 μ L BSA 酶解液(1 μ M)、1 μ L GFP 水溶液(400 fM)和 7 μ L 蒸馏水混合组成. 标准样品 2 由 30 μ L 健康人血清多肽、2 μ L BSA 酶解液(1 μ M)、8 μ L GFP 水溶液(400 fM)和 4 μ L 蒸馏水组成.

2.3 液相色谱分析条件

所有的样品都是采用 Waters 公司无分流的 μ UPLC 系统进行分离, 它的最大操作压力为 10 kpsi. 流动相 A 是 0.1% 甲酸水溶液, 流动相 B 是乙腈(包含 0.1% 甲酸). 反相 C18 预柱(2 cm $\times$ 180 μ m i.d., 填充 5 μ m C18 填料)(Waters, USA)被用作上样柱. 不同的样品通过自动进样器用流动相 A 以 10 μ L/min 的速度经 2 min 冲到预柱上. 然后预柱上所保留的多肽使用实验室填充的 C18 毛细管分析柱(15 cm $\times$ 75 μ m i.d., 填充 3 μ m C18 填料)进行分离, 流速为 300 nL/min, 该 C18 分析柱通过不锈钢三通和一根 7 cm $\times$ 25 μ m i.d. 的 C12 整体柱电喷雾(ESI)喷头相连^[37], 喷雾电压为 3.2 kV. 多肽的分离梯度如下: 起始梯度为 2% 的流动相 B, 90 min 内上升到 35% 的流动相 B, 然后 5 min 内从 35% 上升到 90%, 并维持 10 min, 最后系统用 100% 流动相 A 平衡 15 min. 考虑到系统死体积对梯度洗脱的延迟作用, 在分离梯度开始 15 min 后采集质谱图.

2.4 质谱分析条件

本文所用 μ UPLC 系统和 Q-TOF Premier MS(Waters, USA)相连, 接口是实验室制备的 C12 整体柱 ESI 喷头. 本文所有的质谱实验都是在正离子 V 模式下进行的, 分辨率大于 8000. 质谱检测器 m/z 50 到 1950 的区间用 GFP MS/MS 碎片离子进行校正. 整个质谱分析过程中的质量精确度通过参照物 GFP(150 fmol/ μ L)的平行校正喷雾来维持, 每 30 s 喷雾一次. 当进行 EMRT 定量实验时, LC-MS 和 LC-MS^E 数据分别在离子碎裂能量 4 eV 和 25~45 eV 下采集. 谱图的采集时间设置为 0.9 s, 扫描间隔 0.05 s. 进行选择 EMRT 鉴定实验时, LC-MS/MS 设置为 DDA 模式, 所选择的 EMRT 质量容忍偏差为 200 mDa, 保留时间(RT)容忍偏差为 2 min. m/z 400~1990 区间的 MS 扫描谱图采集时间设置为 0.9 s, m/z 50~3500 区间的 MS/MS 采集时间为 1.2 s, 扫描间隔为 0.05 s.

2.5 EMRTs 定量

所有的 LC-MS/MS^E 数据采用 ProteinLynx GlobalServer (PLGS) v2.3 软件(Water, USA)进行处理. 虽然非酶解的多肽不能用该软件进行鉴定, 但是 EMRTs 的定量信息却能够使用该软件中的 Expression

Analysis 工具来获得. 在定量分析之前, 实验中测定的 EMRTs 的值先进行归一化. 至少在三次分析的两次中得到该定量结果才应用于不同样品的相对定量. 定量的 P 值 ≥ 0.95 表示该多肽表现为上调, ≤ 0.05 则为下调($P=1$ 表示完全上调, $P=0$ 表示完全下调).

2.6 选择性 EMRTs 鉴定

选择相对量 >2 或者 <0.5 的 EMRTs 进行 LC-MS/MS 的 DDA 模式分析, 采集到的 raw 文件用 PLGS v2.3 软件来处理, 所得 MS/MS 数据以 *pk1 格式输出, 最后多肽鉴定使用 Mascot v2.2 (Matrix science, London, United Kingdom)完成. 人 IPI 蛋白数据库(v3.17)用于搜索 MS/MS 数据, 并将 BSA 和 GFP 的序列信息加入该数据库中. 对于 BSA 的数据库检索, 设为胰蛋白酶切限制, 半胱氨酸残基上设置 57.0215 Da 的固定修饰, 甲硫氨酸残基上设置 15.9949 Da 的可变修饰, 可容忍漏切位点数设为 1; 而对于 GFP 和人血清样品, 设置无酶切和固定修饰, 可变修饰为甲硫氨酸残基的氧化 15.9949 Da. 母离子的质量容忍设为 20 ppm, 碎片离子为 0.1 Da. 在数据搜库时用 Mascot 自动生成的随机库评价鉴定的假阳性率(FDR), 通过设置 Mascot 的打分值控制假阳性率 $<1\%$.

3 结果与讨论

3.1 定量方法的验证

所使用的定量分析方法流程如图 1 所示. 首先, 通过 μ UPLC-Q-TOF-MS/MS^E 对两种样品进行相对定量分析, 然后对相对量 >2 或 <0.5 的 EMRT 所对应的质量数进行 μ UPLC-Q-TOF-MS/MS 选择性 DDA 分析. 通过各重复三次的平行试验, 所得到的 MS/MS 数据用 Mascot 进行数据库搜索鉴定, 设置 Mascot 打分值 >30 以控制 FDR $<1\%$. 最后控制在 μ UPLC-Q-TOF-MS/MS 和 μ UPLC-Q-TOF-MS/MS^E 两次分析中质量的相对偏差小于 20 ppm, 保留时间的相对偏差小于 1.5 min.

如上所述, 两个标准样品被用于无标记定量方法的验证. 为了防止错误鉴定, 每个多肽应该在三次平行分析中至少有两次被鉴定到且 Mascot 打分值 >30 才认为是可信的. 在定量和定性分析中同一 EMRT 的精确分子量和保留时间相对偏差应分别小于 20 ppm 和 1.5 min. 在此筛选条件下, 所有从标准样品中

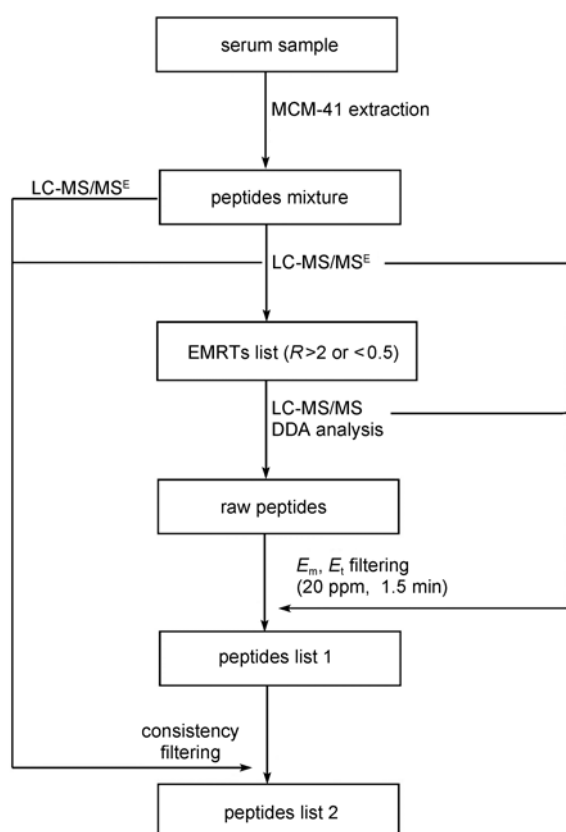


图1 多肽无标记定量策略流程图

鉴定到的血清多肽被完全排除了,理论上血清中的内源性多肽没有相对量的变化,不会被鉴定.最终,21条BSA肽段和GFP被成功定量并定性,结果如表1所示.由表中我们可以看出,BSA肽段的相对量(标准样品1/标准样品2)在2.22到4.00之间变化,平均值为2.81(RSD=18.50%, $n=21$),平均相对偏差6.42%.GFP相对量为0.10,平均相对偏差为22.18%.因此,该方法能够同时定量上调和下调的多肽,并且每个蛋白所鉴定到肽段越多,定量越准确.

当采用 μ UPLC-Q-TOF-MS/MS^E对非酶切多肽进行大规模分析时,由于PLGS软件的功能限制,多肽序列不能进行直接鉴定.因此,本文我们运用 μ UPLC-Q-TOF-MS/MS^E生成存在明显量变的目标多肽的EMRTs列表,用于指导随后的选择性DDA多肽序列鉴定.但标准样品中所添加的BSA为酶解肽段,能够用 μ UPLC-Q-TOF-MS/MS^E直接鉴定,并由PLGS软件计算出其相对量(标准样品1/标准样品2)为2.63,与理论值3.00的相对偏差仅为12.28%.这可能归功于在我们的策略中只有得到准确鉴定的多肽

才用于定量分析.

在我们的策略中只有存在明显量变的目标多肽才会被选取进行下一步的DDA肽段序列鉴定.因此,很多丰度很高但与对照组相比缺少量变的EMRTs将不被鉴定,这有利于低丰度EMRTs的鉴定和定量并提高DDA分析的检测灵敏度.当采用DDA直接对标准样品2中的BSA进行鉴定时,只有7条非冗余肽段的Mascot鉴定打分值大于30,而采用我们的选择性DDA鉴定方法时,能够成功鉴定到14条非冗余肽段.

3.2 HCC病人和健康人血清内源性多肽的定量比较

实验所用的血清样品分别采集自12个HCC病人和12个健康志愿者.首先,用MCM-41从两种血清中分别萃取出内源性肽,然后分别进行 μ LC-MS/MS^E分析,其色谱基峰图如图2所示.虽然图中两种样品的多肽峰很相似,但是两图中具有不同质量的EMRTs可以通过PLGS系统(Waters公司)找到并列.选择其中相对量>2或者<0.5的进行DDA分析,并采用上文所述的流程和标准进行鉴定和假阳性控制.最后再一次 μ LC-MS/MS^E分析对所定量定性到的多肽的相对量进行验证,最终找到了9个相对定量结果一致的多肽,如表2所示.从表2中我们可以看出这9个多肽来源于4个载脂蛋白(A-1, C-1, C-3和E)前体.由于肝脏是人体合成、储存、转运和降解某些载脂蛋白的主要器官^[38],因此肝脏病变会对这类蛋白产生影响^[39].据报道,载脂蛋白A-1前体在诸如HCC的严重肝脏损伤时表现为增强趋势,因为肝脏参与载脂蛋白A-1前体向载脂蛋白A-1转变^[40, 41].表2中,我们的结果显示,HCC病人血清中来源于四个载脂蛋白(A-1, C-1, C-3和E)前体的多肽浓度也呈上升趋势.

从表2中还可以看出,这9个多肽的价态分布在3到5之间,分子量从1638到4066.由于质量分辨率和检测范围的限制,这些多肽很难用离子阱质谱仪进行鉴定.而使用Q-TOF质谱仪,如图3(a)所示,对带五个电荷的多肽离子(R.AELQEGARQKLHEL-
QEKLSPLGEEMRDRA.R)都能检测到.此外,由于Q-TOF质谱仪具有更宽的检测范围,分子量超过4000的多肽(K.SWFEPLVEDMQRQWAGLVEKVQ-AAVGTSAPVPSDNH.- + Oxidation (M))仍然能够被检测到,如图3(b)所示,它的碎片离子很好地分布在 m/z 0~4000之间,这样搜库时能很大程度上增加鉴定的可信度.由于质谱检测的动态范围有限,当采用

表1 标准样品中鉴定和定量到的 BSA 酶解和 GFP 多肽

蛋白	[MH] _{dda}	[MH] _{mse}	<i>E_m</i>	<i>T_{dda}</i>	<i>T_{mse}</i>	<i>E_t</i>	打分值	多肽序列	<i>R_(1/2)</i>
BSA	789.47	789.46	15	38.21	38.35	-0.14	32	K.LVTDLTK.V	2.63
	886.42	886.41	14	35.07	35.47	-0.4	31	K.DDSPDLPK.L	2.63
	1017.57	1017.58	-7	55.96	56.35	-0.39	39	Y.GFQNALIVR.Y	2.70
	1138.49	1138.49	2	34.29	35	-0.71	65	K.CCTESLVNR.R	2.86
	1163.63	1163.61	17	61.11	61.59	-0.48	39	K.LVNELTEFAK.T	2.94
	1170.61	1170.62	-5	49.36	50.01	-0.65	38	R.RHPEYAVSVLL.L	2.63
	1238.59	1238.59	1	25.4	25.97	-0.57	62	L.SHKDDSPDLPK.L	3.33
	1249.61	1249.60	7	38.87	39.76	-0.89	65	R.FKDLGEEHFK.G	2.44
	1305.70	1305.7	2	48.43	48.88	-0.45	51	K.HLVDEPQNLIK.Q	2.78
	1351.67	1351.67	3	28.85	29.42	-0.57	91	F.LSHKDDSPDLPK.L	2.70
	1419.68	1419.68	3	60.55	61.02	-0.47	59	K.SLHTLFGDELCK.V	2.56
	1439.81	1439.8	8	52.46	52.93	-0.47	60	R.RHPEYAVSVLLR.L	2.33
	1443.62	1443.62	3	36.14	36.83	-0.69	62	K.YICDNQDTISSK.L	3.85
	1463.59	1463.57	11	24.12	24.53	-0.41	96	K.TCVADESHAGCEK.S	4.00
	1478.51	1478.51	-3	38.51	39.28	-0.77	61	R.ETYGDMADCCEK.Q	3.85
	1479.79	1479.78	4	70.66	71.04	-0.38	104	K.LGEYGFQNALIVR.Y	2.27
	1502.61	1502.60	5	42.18	42.75	-0.57	65	K.EYEATLEECCA.K.D	2.78
	1511.84	1511.84	-3	57.79	57.75	0.04	36	K.VPQVSTPTLVEVSR.S	2.38
	1532.77	1532.77	-1	33.09	33.79	-0.7	43	K.LKECCDKPLLEK.S	2.22
	1576.77	1576.76	5	54.29	54.54	-0.25	94	K.LKPDPNLTLCDEFK.A	2.44
GFP	1749.65	1749.64	4	42.74	43.43	-0.69	86	K.ECCHGDLLECADDR.A	2.63
	1570.66	1570.66	1	57.43	57.87	-0.44	95	EGVNDNEEGFFSAR	0.10

注: 表中 [MH]_{mse} 表示 μ LC-MS/MS^E 分析中多肽的[MH]; [MH]_{dda} 表示 μ LC-MS/MS DDA 模式分析中多肽的[MH]; *E_m* (ppm)表示 [MH]_{mse} 和 [MH]_{dda} 之间的相对偏差; *T_{mse}* 和 *T_{dda}* 分别表示 μ LC-MS/MS^E 分析和 μ LC-MS/MS DDA 模式分析中多肽的保留时间; *E_t* (min)表示 *T_{mse}* 和 *T_{dda}* 之间的相对偏差; *R_(1/2)* 表示标准样品 1 和标准样品 2 中相应多肽的相对量。

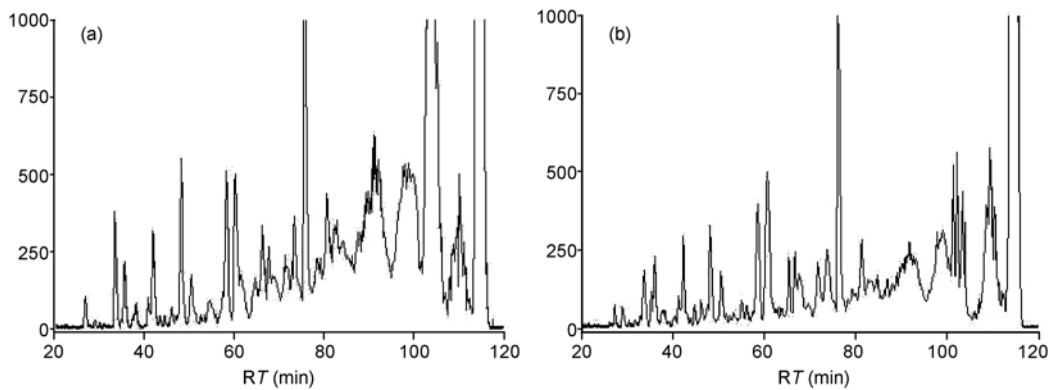


图2 血清多肽梯度 μ LC-MS/MS^E 分析的色谱基峰图. (a) 健康人血清多肽基峰图; (b) HCC 病人血清多肽基峰图

MALDI-TOF质谱对HCC病人和健康人血清多肽进行比较研究时, 如图 4(a)所示, 一些低丰度的多肽如 F.LSALEEYTKKLNTQ.- (MW=1638) 和 K.SWFEPLVEDMQRQWAGLVEKV.Q + Oxidation (M) (MW = 2563)不能被检测到. 因此, 在质谱检测前使用 μ LC可以富集和分离多肽, 随后的Q-TOF-MS就能很容易地对多肽离子进行检测和定量, 结果如图 4(b)所示. 因此, μ LC-Q-TOF-MS系统对高价态、高分子量多肽

具有更好的鉴定能力和更高的检测灵敏度, 更适合于内源性多肽的分析.

3.3 乳腺癌病人和健康人血清内源性多肽的定量比较

乳腺癌病人和健康志愿者血清样品采用同样定量策略进行分析, 并最终定量定性到了 26 条多肽, 如表 3 所示. 由于 Q-TOF-MS 的宽质量范围和高分辨

表 2 HCC 病人和健康人血清中鉴定和定量到的差异多肽

来源蛋白	[MH] _{dda}	[MH] _{mse}	<i>E_m</i> (ppm)	<i>T_{dda}</i> (min)	<i>T_{mse}</i> (min)	<i>E_t</i> (min)	打分值	多肽序列	<i>R_(C/H)</i> ¹	<i>R_(C/H)</i> ²
Apo-A-I	1637.86	1637.86	-1	54.77	55.66	-0.89	61	F.LSALEEYTKKLNTQ.-	2.44	2.63
	2138.03	2138.04	-3	47.36	48.08	-0.72	37	L.HELQEKLSPGGEEMRDRA.R	3.57	3.57
	3361.67	3361.67	1	67.43	68.17	-0.74	52	R.AELQEGARQKLHELQEKLSPGGEEMRDRA.R	3.23	4.00
	2976.50	2976.51	-3	97.54	98.82	-1.28	123	G.TPDVSSALDKLKEFGNTLEDKARELIS.R	3.57	3.85
Apo-C-I	2691.38	2691.39	-2	97.54	98.91	-1.37	79	P.DVSSALDKLKEFGNTLEDKARELIS	2.78	2.94
	2778.41	2778.41	1	97.78	99.01	-1.23	47	P.DVSSALDKLKEFGNTLEDKARELIS.R	3.45	3.33
Apo-C-III	2548.15	2548.15	1	53.33	54.3	-0.97	70	A.SEAEDASLLSFMQGYMKHATKTA.K + 2 Oxidation (M)	5.26	8.33
Apo-E	2563.24	2563.25	-4	89.98	91.07	-1.09	54	K.SWFEPLVEDMQRQWAGLVEKV.Q + Oxidation (M)	3.13	4.00
	4065.93	4065.91	4	98.26	99.39	-1.13	91	K.SWFEPLVEDMQRQWAGLVEKVQAAVGTSAAPVPSDNH.-+Oxidation (M)	3.23	3.45

注: 表中 $R_{(C/H)}^1$ 表示第一次 μ LC-MS/MS^E 实验所得的 HCC 病人和健康人血清中多肽的相对量; $R_{(C/H)}^2$ 表示第二次 μ LC-MS/MS^E 确认实验所得的 HCC 病人和健康人血清中多肽的相对量; 其他符号同表 1 中.

表 3 乳腺癌病人和健康人血清中鉴定和定量到的差异多肽

来源蛋白	[MH] _{dda}	[MH] _{mse}	<i>E_m</i> (ppm)	<i>T_{dda}</i> (min)	<i>T_{mse}</i> (min)	<i>E_t</i> (min)	打分值	多肽序列	<i>R_(C/H)</i> ¹	<i>R_(C/H)</i> ²
ITI4	2184.08	2184.06	11	54.63	55.68	-1.05	121	R.QLGLPGPPDVPDHAAYHPFR.R	11.11	14.29
	1052.55	1052.56	-9	24.88	26.07	-1.19	43	K.SHALQLNNR.Q	9.09	12.50
	1389.72	1389.72	-1	36.8	36.99	-0.19	87	K.GSEMVVAGKLQDR.G	5.26	5.56
	2027.99	2027.97	9	61.03	61.52	-0.49	75	R.QLGLPGPPDVPDHAAYHPF.R	0.13	0.18
	2271.09	2271.09	2	55.67	56.15	-0.48	97	S.SRQLGLPGPPDVPDHAAYHPF.R	0.04	0.04
	1786.83	1786.83	1	55.5	56.42	-0.92	85	L.GLPGPPDVPDHAAYHPF.R	0.02	0.04
	3272.59	3272.58	2	63.14	64.33	-1.19	87	R.MNFRPGVLSSRQLGLPGPPDVPDHAAYHPF.R	0.02	0.03
	2724.36	2724.34	6	61.52	62.50	-0.98	106	R.PGVLSSRQLGLPGPPDVPDHAAYHPF.R	0.01	0.02
	1171.58	1171.57	8	46.15	46.87	-0.72	37	L.GLPGPPDVPDHA.A	H	H
	2657.35	2657.32	13	60.14	60.59	-0.45	49	R.MNFRPGVLSSRQLGLPGPPDVPDHA.A	H	H
	2109.08	2109.08	2	57.09	57.51	-0.42	104	R.PGVLSSRQLGLPGPPDVPDHA.A	H	H
	1412.72	1412.72	-3	54.44	54.83	-0.39	37	R.QLGLPGPPDVPDHA.A	H	H
	1077.53	1077.53	2	44.52	44.47	0.05	32	E.GDFLAEGGGVR.G	C	C
	1206.58	1206.57	8	45.95	45.84	0.11	47	G.EGDFLAEGGGVR.G	C	C
FGAf	1263.61	1263.6	6	47.42	47.30	0.12	43	S.GEGDFLAEGGGVR.G	C	C
	2911.41	2911.42	-2	68.51	69.80	-1.29	152	K.GNTEGLQKSLAELGGHLDQQVEEFRR.R	0.11	0.12
	2599.24	2599.23	4	73.55	73.94	-0.39	48	K.GNTEGLQKSLAELGGHLDQQVEEF.R	H	H
	2762.33	2762.31	6	70.5	70.89	-0.39	42	R.SLAPYAQDTQEKLNHQLEGLTFQM.K	H	H
C3pf	1864.98	1864.97	4	46.63	47.13	-0.5	76	R.SSKITHRIHWESASLL.R	0.28	0.33
	1348.71	1348.71	2	48.39	49.45	-1.06	48	T.HRIHWESASLL.R	0.11	0.15
	1605.85	1605.85	3	42.07	42.73	-0.66	60	I.THRIHWESASLL.R.S	0.06	0.08
	3505.57	3505.56	4	42.5	42.93	-0.43	54	V.SETESRGSESGIFTNTKESSSHHPGIAEFPSRG.K	0.08	0.10
IFGAp	4151.12	4151.12	-1	89.44	90.15	-0.71	44	R.TLLVFEVQQPFLFVLWDQQHKFPVFMGRVYDPR.A.-	0.24	0.26
S/CPIf	2113.06	2113.06	0	32.88	33.58	-0.7	81	E.KNPLPSKETIEQEKQAGES.-	0.38	0.46
Tl3	2378.20	2378.19	5	71.5	71.41	0.09	74	K.DDPDAPLQPVTPQLFEGRRN.R	9.09	11.11
C4ap	2111.17	2111.18	-4	73.98	75.32	-1.34	53	R.DLKVEDIPLARITLPDFR.L	C	C

注: 表中 $R_{(C/H)}^1$ 表示第一次 μ LC-MS/MS^E 实验所得的乳腺癌病人和健康人血清中多肽的相对量; $R_{(C/H)}^2$ 表示第二次 μ LC-MS/MS^E 确认实验所得的乳腺癌病人和健康人血清中多肽的相对量; 其他符号同表 1 中.

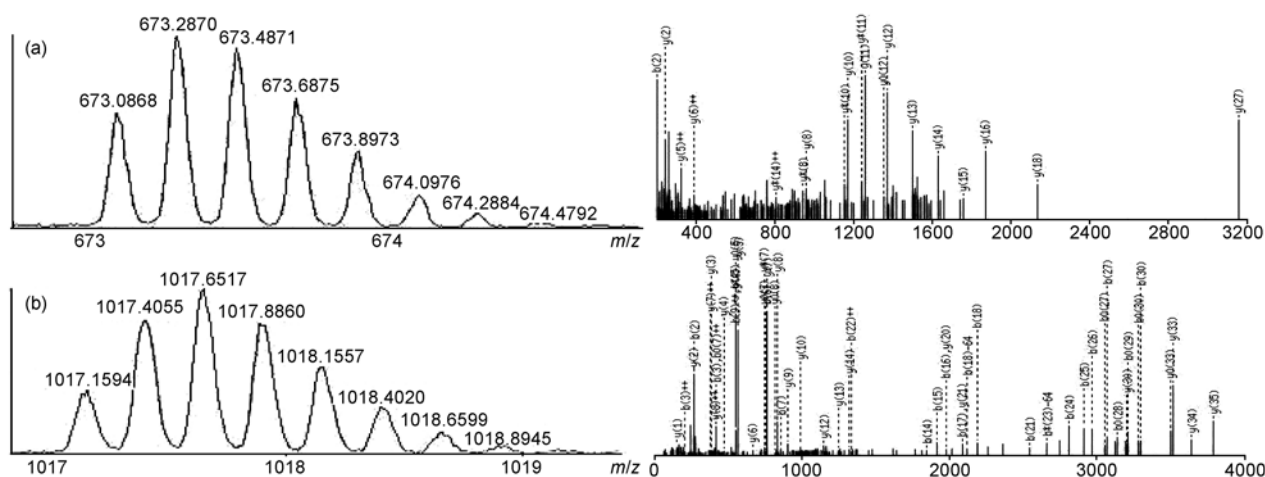


图 3 (a) 五价多肽 R.AELQEGARQKLHELQEKLSPLGEEMRDRA.R 的 MS 和 MS/MS 谱图; (b) 四价多肽 K.SWFEP-LVEDMQRWAGLVEKVQAAVGTSAAPVPSDNH.-(Oxidation (M)) 的 MS 和 MS/MS 谱图

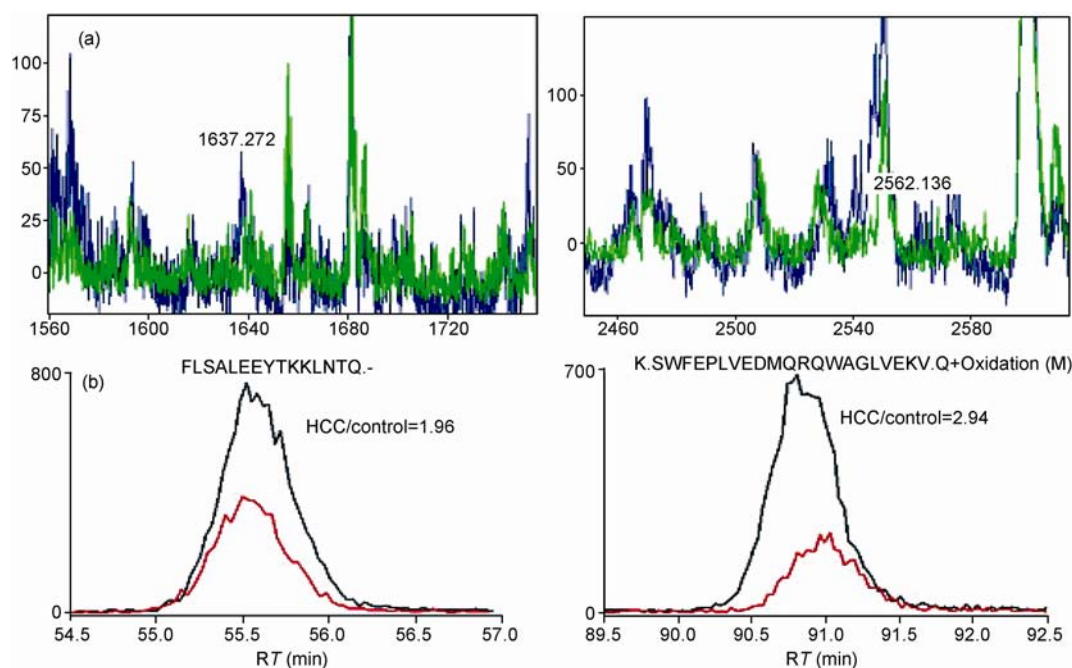


图 4 (a) 血清多肽的 MALDI-TOF-MS 谱图; (b) μ LC-MS/MSE 分析中多肽 FLSALEEYTKKLNTQ.- 和 K.SWFEP-LVEDMQRWAGLVEKV.Q (Oxidation (M)) 的单离子流图. (a) 图中蓝色表示 HCC 病人血清样品, 绿色表示健康人血清样品; (b) 图中黑色表示 HCC 病人血清样品, 红色表示健康人血清样品

率的检测能力, 鉴定到的多肽分子量分布于 1050 到 4150 之间, 与上文一致. 26 条多肽来源于 9 个蛋白质: 间- α 球蛋白抑制因子 H4 变种碎片 (ITI4), 纤维蛋白原 α 链前体碎片 (FGAf), 载脂蛋白 A-4 前体 (apo-A-4), 补体 C3 前体碎片 (C3pf), 纤维蛋白原 α 链前体第二亚型 (IFGAp), 丝氨酸/半胱氨酸蛋白酶抑制剂碎片

(S/CPIf), 类胸腺肽 3 (TI3), 补体 C4-A 前体 (C4ap), 载脂蛋白 B-100 前体 (apo-B-100p). Hamm 等人曾报导, ITI4 在多种人体实体瘤包括乳腺癌中表现为下调^[42]. 我们的结果也显示大多数来源于 ITI4 的血清多肽在 HCC 病人的血样中也表现为下调. 另外, C3pf 的定量结果也与 Villanueva 等人报道的结果一

致^[17]. 由于 μ LC-Q-TOF-MS/MS系统比MALDI-TOF-MS有更高的灵敏度, 与Villanueva等人的工作相比较, 我们的定量策略能定量检测到更多的内源性多肽^[17].

3.4 讨论

结合 μ LC-MS/MS系统的无标记定量方法被广泛地应用于蛋白质组学的相对或绝对定量分析^[23-34]. 然而, 针对多肽组学的此类无标记定量方法研究甚少. 多肽组作为连接蛋白质组和代谢组的桥梁, 对疾病的早期诊断具有重要作用^[6-8]. 我们发展了内源性多肽无标记定量分析的策略, 将多肽定量和鉴定分为两部分, 首先用 μ LC-MS/MS^E分析对两组样品中的多肽离子进行相对定量, 然后选择定量差异明显的离子进行串联质谱DDA模式的多肽鉴定, 得到相应的多肽序列信息. 选择性的DDA鉴定使得两组样品中许多高峰度的多肽离子由于没有相对量的变化而不进行定性分析, 提高了DDA鉴定的检测灵敏度.

例如, 当HCC病人血清样品直接进行DDA分析时, 多肽P.DVSSALDKLKEFGNTLEDKARELI.S, P.DVSSALDKLKEFGNTLEDKARELI.S.R, 和A.SEAEDA-SLLSFMQGYMKHATKTA.K (2 Oxidation (M))都不能被鉴定到(数据未给出). 而采用先定量再定性的定量策略, 三条多肽均能被成功地鉴定和定量.

4 结论

本文发展了一种基于 μ LC-Q-TOF-MS/MS系统的先定量再定性的无标记定量策略, 并将它用于人血清内源性多肽的定量研究. 对BSA酶切样品进行定量分析时, 该方法的平均相对偏差小于10%. 随后, 该方法被应用于正常人血清和HCC以及乳腺癌病人血清中内源性多肽的定量比较研究. 结果表明, 该定量方法检测灵敏度高, 对高价态和高分子量的内源性多肽能可靠地鉴定和定量. 我们发现了血浆中一些与HCC和乳腺癌相关的内源性多肽.

致谢 本工作得到国家自然科学基金(20735004, 20975101)、国家重点基础研究项目(2005CB522701, 2007CB914102)、国家高技术研究发展计划(2006AA02A309, 2008ZX10002-017)、中国科学院知识创新工程重要方向项目(KJCX2.YW.HO9, KSCX2-YW-R-079)、大连化物所知识创新项目、国家高技术研究发展计划(2008ZX1002-020)和国家自然科学基金(20605022, 90713017)项目资助, 特此一并致谢.

参考文献

- 1 Wulfskuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. *Nat Rev Cancer*, 2003, 3: 267—275
- 2 Chung CH, Levy S, Chaurand P, Carbone DP. Genomics and proteomics: emerging technologies in clinical cancer research. *Crit Rev Oncol Hematol*, 2007, 61: 1—25
- 3 Bermúdez-Crespo J, López JL. A better understanding of molecular mechanisms underlying human disease. *Proteomics Clin Appl*, 2007, 1: 983—1003
- 4 Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR, Edgerton M, Gonzalez A, Nadaf S, Moore JH, Caprioli RM, Carbone DP. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet*, 2003, 362: 433—439
- 5 Schrader M, Schulz-Knappe P. Peptidomics technologies for human body fluids. *Trends Biotechnol*, 2001, 19 (Suppl): S55—S60
- 6 Soloviev M, Finch P. Peptidomics: bridging the gap between proteome and metabolome. *Proteomics*, 2006, 6: 744—747
- 7 Diamandis EP. Peptidomics for cancer diagnosis: present and future. *J Proteome Res*, 2006, 5: 2079—2082
- 8 Bischoff R, Luidert TM. Methodological advances in the discovery of protein and peptide disease markers. *J Chromatogr B*, 2004, 803: 27—40
- 9 Ahn S, Simpson RJ. Body fluid proteomics: prospects for biomarker discovery. *Proteomics Clin Appl*, 2007, 1: 1004—1015
- 10 Villanueva J, Lawlor K, Toledo-Crow R, Tempst P. Automated serum peptide profiling. *Nat Protoc*, 2006, 1: 880—891
- 11 Villanueva J, Philip J, DeNoyer L, Tempst P. Data analysis of assorted serum peptidome profiles. *Nat Protoc*, 2007, 2: 588—602
- 12 Rosenzweig CN, Zhang Z, Sun X, Sokoll LJ, Osborne K, Partin AW, Chan DW. Predicting prostate cancer biochemical recurrence using a panel of serum proteomic biomarkers. *J Urology*, 2009, 181: 1407—1414
- 13 Wang N, Wen X, Wang J, Li A, Li W, Li T, Li Y, Zhang X, Wang H, Yang S, Zhou T, Gong W, Tian Y, He K. Breast cancer-specific serum peptide profiles. *Chin Sci Bull*, 2009, 54(3): 421—429

- 14 Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 1999, 17: 676—682
- 15 Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 2001, 19: 242—247
- 16 Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, Liotta LA. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet*, 2002, 359: 572—577
- 17 Villanueva J, Schaffer DR, Phillip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lijia H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst PJ. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *Clin Invest*, 2006, 116: 271—284
- 18 Koomen JM, Li D, Xiao L, Liu TC, Coombes KR, Abbruzzese J, Kobayashi R. Direct tandem mass spectrometry reveals limitations protein profiling experiments for plasma biomarker discovery. *J Proteome Res*, 2005, 4: 972—981
- 19 Chang WC, Huang LCL, Wang Y, Peng W, Chang HC, Hsu NY, Yang WB, Chen CH. Matrix-assisted laser desorption/ionization (MALDI) mechanism revisited. *Anal Chim Acta*, 2007, 582:1—9
- 20 Uttenweiler-Joseph S, Moniatte M, Lagueux M, Dorselaer AV, Hoffmann JA, Bulet P. Differential display of peptides induced during the immune response of drosophila: a matrix-assisted laser desorption ionization time-of-flight mass spectrometry study. *Proc Natl Acad Sci USA*, 1998, 95: 11342—11347
- 21 Jiménez CR, Dreisewerd K, Mansvelter HD, Brussaard AB, Reinhold BB, Van der Schors RC, Karas M, Hillenkamp F, Burbach JPH, Costello CE, Geraerts WPM. Pattern changes of pituitary peptides in rat after salt-loading as detected by means of direct, semiquantitative mass spectrometric profiling. *Proc Natl Acad Sci USA*, 1997, 94: 9481—9486
- 22 Hu LH, Zhou HJ, Li YH, Sun ST, Guo LH, Ye ML, Tian XF, Gu JR, Yang SL, Zou HF. Profiling of endogenous serum phosphorylated peptides by titanium (IV) immobilized mesoporous silica particles enrichment and MALDI-TOFMS detection. *Anal Chem*, 2009, 81(1): 94—104
- 23 Wang GH, Wu WW, Zeng WH, Chou CL, Shen RF. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: reproducibility, linearity, and application with complex proteomes. *J Proteome Res*, 2006, 5: 1214—1223
- 24 Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem*, 2003, 75: 4818—4826
- 25 Radulovic D, Jelveh S, Ryu S, Hamilton G, Foss E, Mao Y, Emili A. Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics*, 2004, 3: 984—997
- 26 Prakash A, Piening B, Whiteaker J, Zhang H, Shaffer SA, Martin D, Hohmann L, Cooke K, Olson JM, Hansen S, Flory MR, Lee H, Watts J, Goodlett DR, Aebersold R, Paulovich A, Schwikowski B. Assessing bias in experiment design for large scale mass spectrometry-based quantitative proteomics. *Mol Cell Proteomics*, 2007, 6: 1741—1748
- 27 Quintana LF, Campistol JM, Alcolea MP, Bañon-Maneus Solé-González A, Cutillas PR. Application of label-free quantitative peptidomics for the identification of urinary biomarkers of kidney chronic allograft dysfunction. *Mol Cell Proteomics*, 2009, 8: 1658—1673
- 28 Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, McKenna T, Nold MJ, Richardson K, Young P, Geromanos S. Quantitative proteomic analysis by accurate mass retention time pairs. *Anal Chem*, 2005, 77: 2187—2200
- 29 Purvine S, Eppel JT, Yi EC, Goodlett DR. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics*, 2003, 3: 847—850
- 30 Hughes MA, Silva JC, Geromanos SJ, Townsend CA. Quantitative proteomic analysis of drug-induced changes in mycobacteria. *J Proteome Res*, 2006, 5: 54—63
- 31 Silva JC, Gorenstein MV, Li GZ, Vissers JPC, Geromanos SJ. Absolute quantification of proteins by lcms: a virtue of parallel ms acquisition. *Mol Cell Proteomics*, 2006, 5: 144—156
- 32 Silva JC, Denny R, Dorschel C, Gorenstein MV, Li GZ, Richardson K, Wall D, Geromanos SJ. Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome. *Mol Cell Proteomics*, 2006, 5: 589—607
- 33 Vissers JPC, Langridge JI, Aerts JMFG, Gerom. Analysis and quantification of diagnostic serum markers and protein signatures for Gaucher disease. *Mol Cell Proteomics*, 2007, 6: 755—766
- 34 Levin Y, Schwarz E, Wang L, Leweke FM, Bahn S. Label-free LC-MS/MS quantitative proteomics for large-scale biomarker discovery in complex samples. *J Sep Sci*, 2007, 30: 2190—2197
- 35 Tian RJ, Zhang H, Ye ML, Jiang XG, Hu LH, Li X, Bao XH, Zou HF. Selective extraction of peptides from human plasma by highly ordered mesoporous silica particles for peptidome analysis. *Angew Chem Int Ed*, 2007, 46: 962—965

- 36 Tian RJ, Ren LB, Ma HJ, Li X, Hu LH, Ye ML, Wu RA, Tian ZJ, Liu Z, Zou HF. Selective enrichment of endogenous peptides by chemically modified porous nanoparticles for peptidome analysis. *J Chromatogr A*, 2009, 1216: 1270—1278
- 37 Wang FJ, Ye ML, Dong J, Tian RJ, Hu LH, Han GH, Jiang XN, Wu RA, Zou HF. Improvement of performance in label-free quantitative proteome analysis with monolithic electrospray ionization emitter. *J Sep Sci*, 2008, 31: 2589—2597
- 38 Lewis GF, Rader DJ. New insights into the regulation of HDL metabolism and reverse cholesterol transport. *Circ Res*, 2005, 96: 1221—1232
- 39 Phillips GB. The lipid composition of serum in patients with liver disease. *J Clin Invest*, 1960, 39: 1639—1650
- 40 Matsuura T, Koga S, Ibayashi H. Increased proportion of proapolipoprotein A-I in HDL from patients with liver cirrhosis and hepatitis. *Gastroenterol Jpn*, 1988, 23(4): 394—400
- 41 Jiang JT, Nilsson-Ehle P, Xu N. Influence of liver cancer on lipid and lipoprotein metabolism. *Lipids Health Dis*, 2006, 5: 4
- 42 Hamm A, Veeck J, Bektas N, Wild PJ, Hartmann A, Heindrichs U, Kristiansen G, Werbowetski-Ogilvie T, DelMaestro R, Knuechel R, Dahl E. Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis. *BMC Cancer*, 2008, 8: 25

A strategy with label-free quantification of the targeted peptides for quantitative peptidome analysis of human serum

ZHU Jun, WANG FangJun, DONG XiaoLi, YE MingLiang & ZOU HanFa

Key Laboratory of Separation Sciences for Analytical Chemistry, Chinese Academy of Sciences; National Chromatographic Research and Analysis Center; Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

Abstract: Peptidomics draws more and more attention in discovering useful biomarkers for early diagnosis of disease. However, there is lack of efficient quantification strategy in peptidome analysis. In this study, a strategy with label-free quantification of the targeted endogenous peptides based on peak intensity using μ UPLC-Q-TOF-MS/MS was developed for quantitative peptidome analysis of human serum. Different amounts of standard BSA tryptic digesting peptides were added into the same serum extracts for evaluation of the developed strategy, and it was observed that the average relative error of the targeted peptides was 6.42%, which was superior to the result obtained directly by commercially available software PLGS. It was also demonstrated this quantification strategy could obviously increase the detection sensitivity of the peptide by DDA analysis. Then, this strategy was applied to comparatively analyze the peptides extracted from the serum of HCC or breast cancer patients and healthy individuals, respectively. Peptides with charge states up to 5 and molecular weight over 4000 can be reliably identified and quantified. And this quantitative analysis method based on μ UPLC-Q-TOF-MS/MS exhibited superior sensitivity than that by MALDI-TOF-MS commonly used in peptidome analysis. Finally, some interesting endogenous peptides related to corresponding diseases were successfully obtained.

Keywords: peptidomics, human serum, label-free quantification, hepatocellular carcinoma, breast cancer