

基于液相色谱质谱联用的蛋白质组非标记定量研究策略的建立及应用*

陈明^{1,2)} 应万涛¹⁾ 方勤美¹⁾ 孙薇¹⁾ 贺福初¹⁾ 钱小红^{1)**}

¹⁾蛋白质组学国家重点实验室, 北京蛋白质组研究中心, 军事医学科学院放射与辐射医学研究所, 北京 100850;

²⁾军事医学科学院卫生学与环境医学研究所, 天津 300050)

摘要 定量蛋白质组研究是蛋白质组研究的热点和难点, 而液相色谱质谱技术已经被广泛地应用于蛋白质的定性和定量研究. 该研究建立和优化了一种基于液相色谱质谱联用技术的蛋白质组非标记定量方法, 并对两种肽段质谱检测计数的归一化算法进行了比较, 结果发现 ASC 法要优于 R_{sc} 法. 最后, 将建立的方法应用于肝癌细胞模型 HepG2 和 HepG2-HBx 细胞系的差异蛋白质组表达研究. 质谱鉴定结果用聚类分析软件 Cluster3.0 进行分析, 最后鉴定出 107 个重叠蛋白, 其中 9 个蛋白质表达上调(Ratio > 1.75), 6 个蛋白质表达下调(Ratio < 0.5), 这些蛋白质均与肝癌发生和恶化密切相关. 结果表明, 该技术操作简单、方便, 具有较高的灵敏度和动态范围, 利用该方法进行差异蛋白质组研究和发现生物标志物在理论和临床上具有十分重要的意义.

关键词 液相色谱质谱联用, 非标记定量, 生物标志物

学科分类号 Q51

定量蛋白质组学是蛋白质组研究的重要内容, 通过对基因表达产物——蛋白质的定量分析, 以了解基因在不同的生理和病理条件下的表达情况. 二维凝胶电泳(2-DE)及其相关的染色技术是目前进行蛋白质组学定量研究的主要方法, 通过凝胶染色可以确定蛋白质的丰度信息及高低表达状况, 通过胶内蛋白质酶解和质谱分析可以进一步鉴定差异表达蛋白质, 但是该策略不仅难以检测到蛋白质表达的痕量变化, 而且其测量灵敏度远低于具有生物学意义的蛋白质表达量的变化范围, 重要的是由于方法费时费力, 而且重现性不理想, 很难实现通量化与自动化^[1~5]. 液相色谱法(LC)作为一种替代 2-DE 的分离方法在蛋白质组研究中广泛应用, 基于液相色谱质谱联用(LC-MS)的分析方法是典型的“自下而上(Bottom-Up)”分析方法, 通常要用特异的蛋白酶(如胰蛋白酶)将蛋白质裂解为肽段. 与 2-DE 不同, LC-MS 对肽的定量和鉴定是同时进行的, 例如, 根据二级质谱图谱, 可以获得肽片段的序列信息, 而利用一级质谱图的信息(如总离子流峰面积或母离子信号强度), 可获得对应肽段或蛋白质的定量信息. 有研究认为, 由于色谱共流出离子的抑

制效应存在, 质谱信号强度并不总是反映蛋白质的丰度^[6~8]. 因此, 在已发表的定量分析文章中, 大多数 LC-MS 分析是与稳定同位素标记联合使用的, 如同位素亲和标签试剂(ICAT、iTRAQ)、 $^{18}\text{O}/^{16}\text{O}$ 和 $^{15}\text{N}/^{14}\text{N}$ 标记等^[9~12]. 虽然稳定同位素标记技术准确、灵敏, 但也有缺点, 例如标记试剂昂贵、标记反应复杂等, 除非是在代谢水平, 否则标记效率很难保证. 另外, 标记反应产生的成对离子峰使质谱图谱更加复杂, 而且对标记稳定同位素的成对质量峰的区分要求质谱仪器具有很高的分辨率^[13].

正是因为稳定同位素标记定量研究策略存在诸多问题, 直接基于质谱数据进行无标记定量分析的方法日益受到重视. 质谱数据中不仅包含着蛋白质的序列信息, 通过对质谱信息的提取也可以获得蛋白质的定量信息. 基于液相色谱质谱技术的非标记

* 国家重点基础研究发展规划项目(2001CB510201, 2004CB518707), 国家自然科学基金(20405017, 20505018, 20505019)资助项目, 北京市科技计划重大项目(H030230280190).

** 通讯联系人.

Tel: 010-80727777-1231, E-mail: qianxhl@yahoo.com.cn

收稿日期: 2007-08-24, 接受日期: 2007-11-07

策略有多种定量模型, 有的以肽段的色谱积分面积为基础, 通过比较一对生物样品中相对应蛋白质酶解多肽的色谱积分面积而得到两者的相对丰度^[14], 有的以肽段被质谱检测的计数为基础, 通过归一化来表征被检测蛋白质的相对丰度^[15]. 非标记技术认为肽段在质谱中被捕获检测的频率与其在混合物中的丰度成正相关, 因此蛋白质被质谱检测的计数反映了蛋白质的丰度, 通过适当的数学公式可以将质谱检测计数与蛋白质的量联系起来, 从而对蛋白质进行定量^[16~19]. 本研究的目的是发展一种快速的、基于液相色谱质谱联用技术平台的非同位素标记定量技术, 以方便地应用于生物标志物的发现和差异蛋白质组学研究.

1 材料与方法

1.1 化学试剂和药品

标准蛋白质牛血清白蛋白(bovine serum albumin, BSA), 乙醇脱氢酶(ADH), 甘油醛-3-磷酸-脱氢酶(glyceraldehyde-3-phosphate-dehydrogenase, G3P), 细胞色素 c, 二硫苏糖醇(DTT)等购于 Amersham Bioscience 公司(Uppsala, Sweden); 色谱纯乙腈(acetonitrile, ACN)购自 JTBaker 公司(NJ, USA); 色谱纯水购自 Millipore 公司(Billerica, MA, USA); 碘乙酰胺(Iodoacetamide, IAA)购自 Acros 公司(New Jersey, USA); 甲酸(FA), 酵母粉购自 Sigma 公司(St Louis, MO, USA); 胰酶(修饰、测序级)购自 Promega 公司(Madison, WI, USA), 三氟醋酸(TFA)购自 Fluka 公司; 其他所用试剂均为色谱纯、分析纯或生化纯试剂.

HepG2, HepG2-Hbx 细胞系由北京蛋白质组研究中心方勤美博士惠赠.

1.2 实验方案

实验分两部分进行. 首先, 用标准蛋白质和相对简单的酵母体系进行实验, 对非标记定量方法进行考察. 第二部分, 将该方法用于一对有差异的实际样本.

1.3 样品准备

收获的 HepG2 和 HepG2-HBx 细胞用预冷的 PBS 缓冲液洗净(在 4℃ 条件下进行), 离心, 弃上清, 往沉淀中加入 5 ml 裂解缓冲液(9.5 mmol/L 脲, 1% DTT, 0.2 mmol/L Na₂VO₃, 1 mmol/L NaF, 1 mmol/L 蛋白酶抑制剂混合物), 超声 1 min, 混合物溶液于室温下旋涡混匀 30 min, 然后于 25 000 g

离心 1 h, 取上清. 提取出的蛋白质用丙酮沉淀. 具体方法是, 按 1:3(体积比)往蛋白质溶液中加入预冷的丙酮, 于 -20℃ 冰箱中放置过夜, 21 000 g 离心 15 min, 弃去上清, 沉淀冷冻干燥于 -20℃ 保存备用. 蛋白质的浓度采用 RCDC 法测定(Bio Rad 公司).

酵母蛋白质的提取同上.

1.4 蛋白质的溶液酶切

取等量的 HepG2 和 HepG2-HBx 细胞提取蛋白质(含 9.5 mol/L 脲, 65 mmol/L DTT), 加入 100 μl 50 mmol/L NH₄HCO₃, 于 37℃ 水浴 4 h, 然后加入 1 mol/L IAA, 使终浓度达到 50 mmol/L, 于室温下暗处放置 1 h, 再加入 ACN(终浓度为 10%)和 50 mmol/L NH₄HCO₃, 以 100:1 加入胰酶混匀, 37℃ 温浴. 为了保证酶切效果, 2 h 后以 100:1 加入胰酶, 混匀后继续温浴共 24 h. 酶切完成后, 反应液用 5% FA 酸化, 酶切混合物冷冻干燥并于 -20℃ 保存.

1.5 LC-ESI-MS/MS 分析

所用系统为 Thermo Finnigan™ LTQ 系统, 配置为高通量分析模式. 两根 C18 预柱(100 μm×5 mm)连接于十通阀上, 当一根预柱上样时, 另一根预柱进行反相洗脱; 分析柱为 PicoFrit™ 反相柱(BioBasic® C18, 5 μm, 75 μm×10 cm, 15 μm spray tip, New Objective, Woburn, MA, USA). 毛细管色谱系统在 LC-Packing 系统上运行, 由 Famos 上样器在 UserProg 模式下上样, 上样体积为 19 μl, 上样流速为 10 μl/min. 色谱洗脱在 Ultimate 系统上进行, 洗脱下来的组分经过 ESI 离子源直接进入质谱, 液相色谱条件为: 流动相 A, 0.1% FA-98% 水溶液; 流动相 B, 0.1% FA-80% ACN 溶液. 洗脱条件: 0~90 min, 流动相比比例由 2% B 线性升至 40% B; 90~105 min, 流动相比比例由 40% B 线性升至 100% B; 维持 15 min 后, 以 100% 流动相 A 平衡色谱柱 30 min. 流动相流速为 200 nl/min.

从反相柱流出的洗脱组分由纳升级电喷雾接口喷出, 电喷雾电压 1.8 kV, 离子传输毛细管温度为 200℃. 质谱采用全扫描(Mass range, m/z 350~1 600)一级质谱数据依赖的二级质谱扫描模式(Data Dependent MS/MS Scan), 依次选取一级质谱中离子强度最强的 5 个离子进行 CID 串联质谱, 归一化碰撞能量为 35%. 采用串联质谱扫描的动态排除功能(dynamic exclusion), 设置排除时间为 90 s.

1.6 数据库检索

二级质谱图的数据检索通过整合 SEQUEST 引擎的 Bioworks3.1 工作站进行. 所用数据库为酵母数据库和 IPI3.07 人蛋白质 Fasta 数据库. 检索设定肽序列氨基酸可变修饰为 C (+57.02 u)、M(+15.99 u), 肽质量允差为 1.5 u. 反转数据库建立: 把每个蛋白质的氨基酸顺序倒转. 用 BuildSummary 软件对得到的 Sequest 检索结果进行整合和比较. 用 Cluster3.0 进行聚类分析^[20].

1.7 统计分析

实验至少重复 3 次, 采用 *t* 检验.

2 结果与讨论

2.1 非标记技术的色谱重现性

利用非标记技术对蛋白质进行定量时, 色谱的

重现性是非常重要的, 特别是在比较一对差异样品时, 如果色谱本身的重现性不能保证就容易引进差异. 因此, 我们以酵母蛋白质组为标准体系对所用的液相色谱体系进行考察. 图 1a, b, c 显示的是酵母蛋白质组酶切产物重复分离 3 次的总离子流色谱图, 从图 1 中可以看出, 3 次重复的峰容量、色谱保留时间、峰强度及色谱峰形状非常一致. 为了方便比较, 我们将图 1a, c 通过软件转化为二维图 (图 1d, e). 图 1d, e 中, 有颜色的点表示色谱峰, 不同的颜色表示不同的峰强度. 图 1f, g 和 h, i 分别表示图 1d, e 中画圆圈处的放大图, 其中标箭头处是我们随机选取的两个色谱保留时间点的色谱峰强度和峰容量的比较. 我们的结果表明, 通过严格的实验设计和实验条件控制, 色谱的重现性可以得到很好的保证.

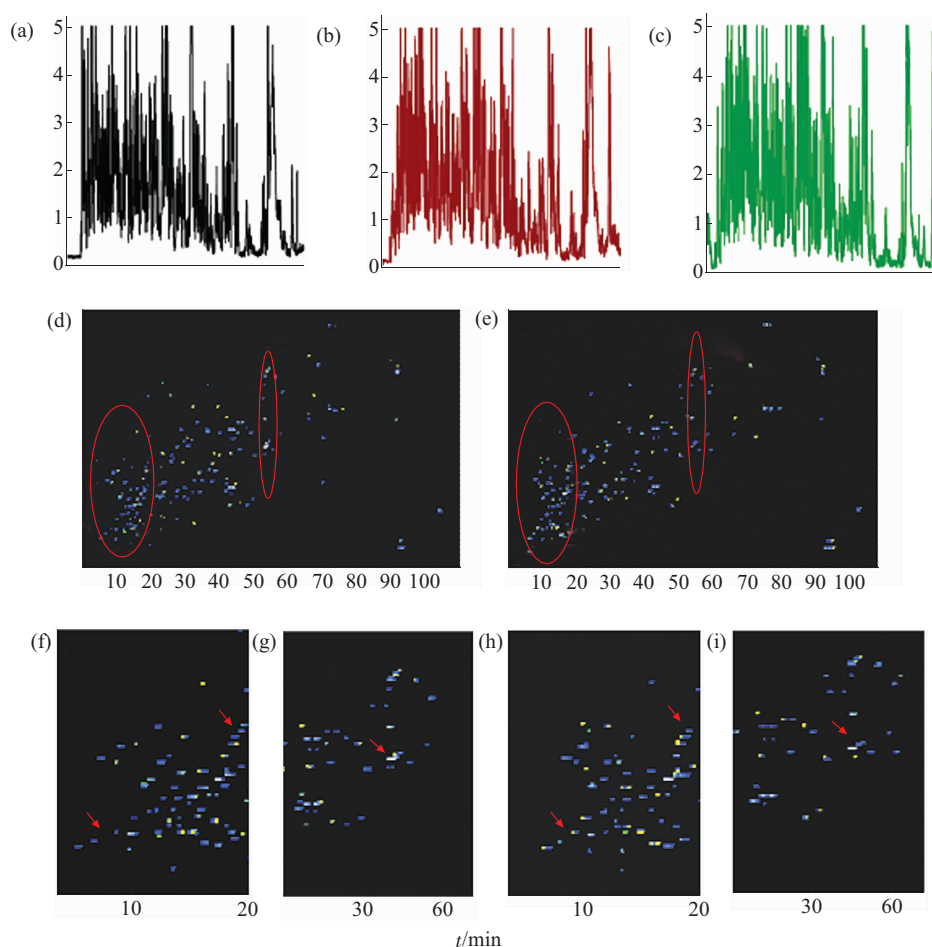


Fig. 1 Repeatability of liquid chromatography

TIC of chromatography in 3 independent experiments (a, b and c) and transformed by Xcalibur (d, e). (f), (g), (h), (i) represent the amplification of circled region in (d) and (e), respectively. The *x*-axis represents the retention time (RT) of liquid chromatography. The *y*-axis represents the intensity of chromatography peaks (a, b, c) and the mass to charge ratio (d, e, f, g, h, i), respectively.

那么质谱鉴定的重现性又怎么样呢? 我们对 3 次重复的质谱鉴定结果进行了分析 (图 2). Run1,

Run2, Run3 分别鉴定酵母蛋白质 158、165 和 162 个 (95% 可信度). Run1 与 Run2, Run1 与 Run3,

Run2 与 Run3 之间的重叠蛋白质数分别是 125、133 和 137, 重叠率在 75%以上, 这说明在该实验条件下重复鉴定的蛋白质在数量上是稳定的. 我们

对任意两次鉴定重叠蛋白的质谱检测频次进行分析, 发现同一蛋白质在任意两次 Run 中的质谱检测频次数基本一致, 相关系数达到 0.98 以上.

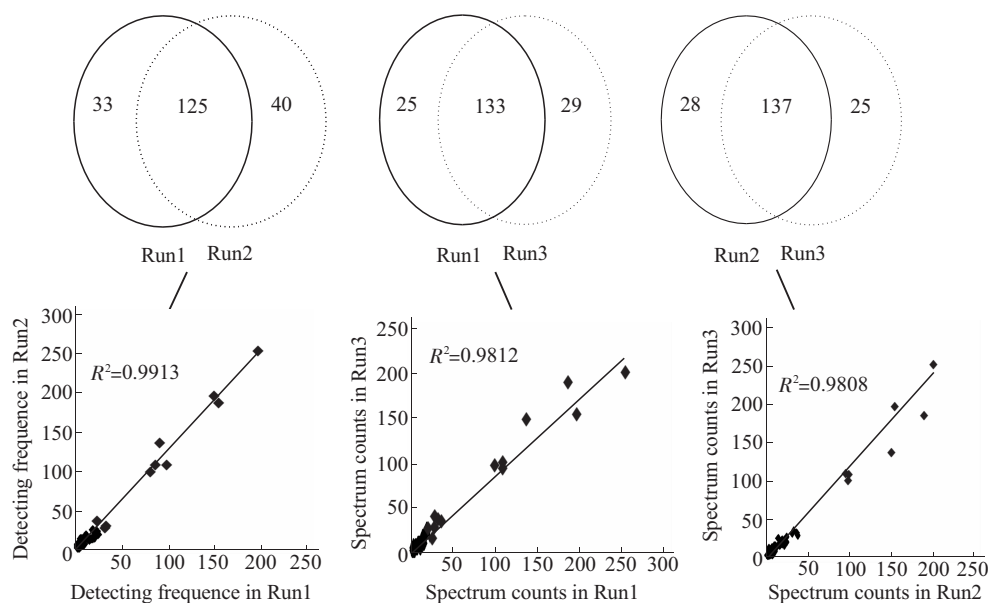


Fig. 2 The overlap of identified protein in 2 independent LC-ESI-MS/MS runs

The coefficient correlation of any two ESI-LCMS/MS runs was range of 0.98~0.99.

2.2 蛋白质的质谱检测计数(spectra counts)与浓度的关系

对蛋白质的质谱检测计数的定义为: 蛋白质的酶切肽段被质谱鉴定到的次数. 因此, 蛋白质的质

谱检测计数在数字上与蛋白质鉴定的冗余肽段数是一致的^[21]. 以 BSA、ADH、G3P 等不同蛋白质为标准品, 考察蛋白质的量与其被质谱检测计数的关系. 如图 3a 所示, 随着蛋白质浓度的增加, 蛋白

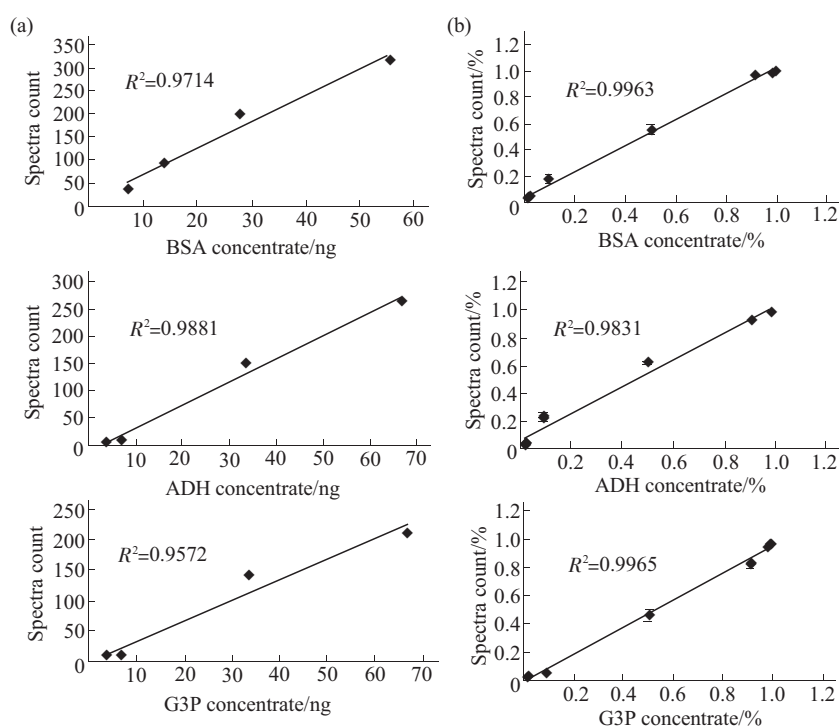


Fig. 3 Correlation between spectra counts with protein abundance

A linear correlation was observed between spectral count with concentrate of standard proteins (a) and relative abundance of protein marker added in yeast (i.e., % of protein markers added), in this experiment 0.01%, 0.05%, 0.1%, 0.5%, 1% standard protein were spiked into yeast proteome.

质被鉴定的质谱检测计数也是呈线性增加的, 而且两者的相关性非常好, 相关系数在 0.95 以上. 我们将不同浓度的标准蛋白质加入到酵母体系中, 考察标准蛋白质在复杂体系中其蛋白质被质谱检测计数与浓度的关系, 结果如图 3b 所示. 可以看出, 随着酵母中加入的标准蛋白质的百分比增加, 标准蛋白质的质谱检测计数所占整个复杂体系的质谱检测计数的百分比是不断增加的. 这说明质谱检测计数可以用来表征样品中某一蛋白质的量.

2.3 蛋白质质谱检测频次的归一化处理

为了使不同样品中蛋白质的质谱检测计数具有可比性, 我们对蛋白质的质谱检测计数进行归一化处理. 主要采用了两种方法. 一种是以蛋白质的平均质谱检测计数(ASC)为参照进行归一化^[22], 一种是“ R_{sc} ”法^[23].

$$\text{Normalize} = SC_n / ASC_n \quad (1)$$

其中 Normalize 为每个蛋白质的质谱检测计数的归一化值, SC_n 为该蛋白质在某一次重复实验中的质谱检测计数, ASC_n 为该蛋白质在多次(一般为 3 次重复)重复实验中的质谱检测计数的平均值.

$R_{sc} = \log_2[(n_2 + f)/(n_1 + f)] + \log_2[(t_1 - n_1 + f)/(t_2 - n_2 + f)]$ (2)
 R_{sc} 表示样品 1 和样品 2 中某一蛋白质的丰度比的对数($\log_2$ ratio); n_1 和 n_2 表示某一蛋白质在两样品中被质谱检测的质谱检测计数, t_1 , t_2 分别为两样品中所有鉴定蛋白质的质谱检测计数之和; f 为校正系数, 在本文中为 1.25.

按照以上方法, 我们对酵母的 3 次重复实验中鉴定的蛋白质的质谱检测计数进行了归一化处理. 理论上同一样品的不同 Run 中蛋白质的丰度比值应该为 1. 但实际上由于系统误差的存在, 该比值是在一定区域内波动的.

如图 4 所示, 图 4a 是用平均质谱检测计数进行归一化后, 3 次 Run 中共同鉴定的 100 个蛋白质的相对丰度的比值. 可以看出比值分布在 1 附近, 最大值为 1.75, 最小值为 0.5. 图 4b 是用 R_{sc} 法进行归一化后的结果, 可以看到几乎所有 R_{sc} 值都在 ± 1 之间, 即所有的比值在 2 倍的范围内波动. 3 次 Run 中蛋白质的平均 Ratio 分别为 0.9589 ± 0.1946 , 1.0043 ± 0.1947 , 1.0367 ± 0.2061 , 这说明用该方法具有很好的灵敏度. 该方法用于两组差异样品时, 当鉴定蛋白质的 Ratio 值 ≥ 1.75 或 ≤ 0.5 时, 可以认为是真正的差异. 为了进一步考察这两种归一化方法的特点, 将酵母样品按 1、3、5、10 和 15 倍的上样量进行上样, 考察这两种方法能否

有效地鉴定出其中的差异. 对于方法(1)当取 Ratio 值大于 1.75 时可以区分出 60 % 以上的差异, 随着差异的增大区分度越来越好(图 5a). 而对于方法(2)则鉴定不出差异, 甚至连 15 倍的差异也很难区分(图 5b), 但随着差异的增大区分度会有所好转. 这说明方法(1)对差异的区分度要优于方法(2).

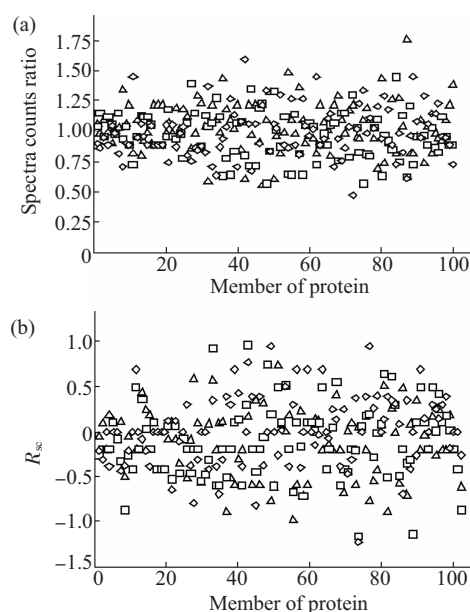


Fig. 4 Distribution of spectra counts ratio of identified proteins in LC-ESI-MS/MS

Spectra counts were normalized by average spectra counts (a) and R_{sc} (b), respectively. (a) $\triangle$:Run1; $\square$:Run2; $\diamond$:Run3. (b) $\triangle$:R2/R1; $\square$:R3/R1; $\diamond$:R3/R2.

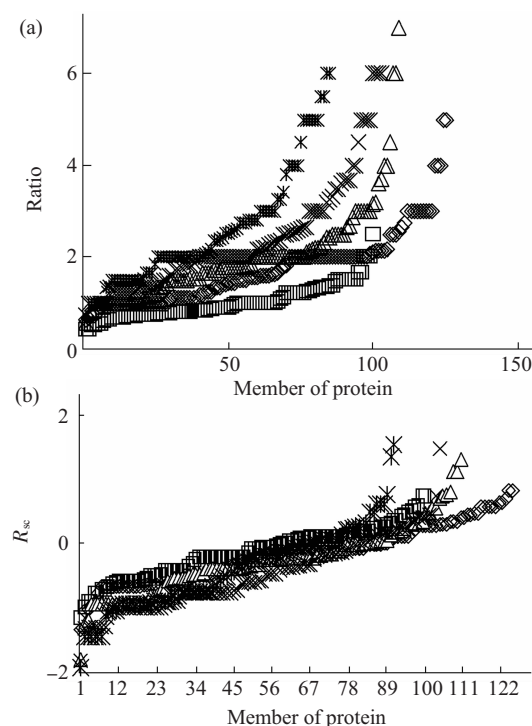


Fig. 5 Comparison of two normalization system

Spectra counts were normalized by average spectra counts (a) and RSC (b). $\square$:CX; $\triangle$:d3; $\diamond$:d5; $\times$:d10; $*$:d15.

2.4 HepG2 和 HepG2-HBx 细胞系的差异表达

乙型肝炎病毒(hepatitis B virus, HBV)是引起包括肝细胞癌 (hepatocellular carcinoma, HCC) 在内的慢性肝脏疾病的重要致病因素, 在人类 HBV 携带者中, HCC 常发生于慢性活动性肝炎(chronic active hepatitis, CAH)和肝硬化后. 慢性乙肝患者发生肝癌的机制目前仍不清楚, 病毒和宿主在肝癌发生中都有作用. HBx 基因是 HBV DNA 中最小的一个, 编码基因位于 1 374~1 838 核苷酸之间, 全长为 465 bp. 含有 3 对直接重复序列(DR1、DR2; DR3L、DR3R; DR4L、DR4R)和 2 对反向重复序列(IR1L、IR1R; IR2L、IR2R), 这些重复序列为 HBx 基因的保守序列, 为 HBx 基因功能所必需, 在整个 HBV 基因组中都有重要的功能^[24].

运用非同位素标记的定量方法分析了肝癌细胞 HepG2 转染 HBx 基因前后的蛋白质组变化. 取等量的正常 HepG2 细胞蛋白质提取物和转染 HBx 基因的 HepG2 细胞(HepG2-HBx)蛋白质提取物, 经溶液酶切后进行 LC-ESI- MS/MS 分析, 每个样品重复鉴定 3 次. 无论是从色谱的峰形, 还是色谱峰强度等参数来比较, 色谱的重现性都非常好(数据未列出).

两个样品分别经过 3 次重复实验, 共鉴定 107 个重叠的蛋白质. 对每个蛋白质被鉴定的质谱检测计数采用方法(1)进行归一化处理, 以转染前的正常细胞蛋白质的质谱检测计数为对照进行比较, 计算出 HepG2 细胞转染 HBx 基因后的蛋白质表达变化值, 并进行聚类分析(图 6).

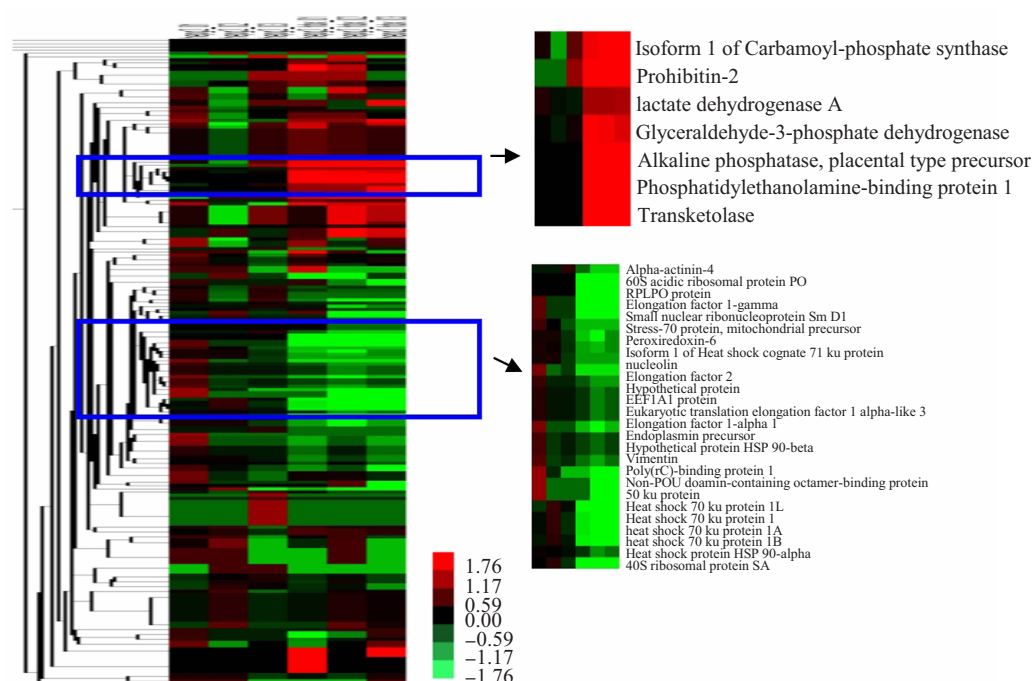


Fig. 6 Cluster analysis of identified proteins

The 107 overlap proteins in HepG2 and HepG2-Hbx cell lines were clustered, the cluster diagram on the left shows the hierarchical clustering of proteomic data from 3 replicate LC-ESI-MS/MS runs of each cell line, the diagram on the right show the up or down regulate proteins. The color indicates the spectra count ratio of proteins.

我们鉴定到的上调蛋白质 (Ratio > 1.75) 有 9 个, 分别为: protein NDRG1, fatty acid synthase, prohibitin-2, FK506-binding protein 4, Glyceraldehyde-3-phosphate dehydrogenase, brain acid soluble protein 1, enolase 1, fructose-

bisphosphate aldolase A, transketolase. 下调蛋白质 (Ratio < 0.5) 有 6 个, 分别为: 60S ribosomal protein L4, poly(rC)-binding protein 1, endoplasmic precursor, elongation factor 2, heat shock 70 ku protein 1L, heat shock 70 ku protein 1(表 1).

Table 1 Partial proteins identified as up-regulate and down-regulate in HepG2 and HepG2-HBx cell lines									
Protein ID	Spectra counts			Average	Spectra counts			Ratio (average)	P-value
	HepG2-1	HepG2-2	HepG2-3		HepG2-Hbx-1	HepG2-Hbx-2	HepG2-Hbx-3		
IPI00000875.5	4	2	2	2.667	2	1	1	0.5	0.148 148 148
IPI00003918.5	2	3	5	3.333	1	1	1	0.3	0.057 235 231
IPI00008530.1	2	2	2	2	1	1	1	0.5	#
IPI00016610.2	5	3	2	3.333	2	1	1	0.4	0.101 191 507
IPI00022078.3	3	4	3	3.333	4	7	6	1.7	0.068 587 106
IPI00026781.2	3	3	3	3	6	5	6	1.888 889	0.001 323 897
IPI00027230.3	7	7	7	7	3	2	3	0.380 952	0.000 202 04
IPI00027252.6	1	1	2	1.333	4	4	6	3.5	0.011 056 493
IPI00186290.5	13	10	11	11.33	5	5	5	0.441 176	0.001 991 489
IPI00219005.2	2	1	1	1.333	3	2	3	2	0.047 420 656
IPI00219018.6	51	47	53	50.33	102	98	92	1.933 775	0.000 158 616
IPI00299024.8	1	2	2	1.667	3	4	4	2.2	0.013 235 6
IPI00301277.1	9	12	8	9.667	5	3	4	0.413 793	0.013 158 073
IPI00302850.4	3	1	2	2	1	1	1	0.5	0.158 302 423
IPI00304596.3	4	2	2	2.667	2	1	1	0.5	0.148 148 148
IPI00304925.3	11	13	6	10	4	3	4	0.366 667	0.039 778 363
IPI00444262.1	5	5	4	4.667	3	1	3	0.5	0.035 168 453
IPI00453473.5	3	4	3	3.333	3	1	1	0.5	0.089 009 342
IPI00465248.5	56	58	60	58	105	101	104	1.781 609	1.086 36×10 ⁻⁵
IPI00465439.4	7	5	3	5	15	10	10	2.333 333	0.030 272 878
IPI00643920.2	6	6	6	6	12	13	13	2.111 111	3.688 31×10 ⁻⁵

2.5 部分差异蛋白质功能的初步讨论

NDRG1 蛋白质是一种由组织缺氧诱导的蛋白质，它的表达与癌症的进程有关^[25]。Sibold 等^[26]的研究表明人肝癌细胞中，NDRG1 蛋白在细胞质和质膜中的表达量会显著上调。NDR 家族蛋白的确切功能目前尚不清楚，但可以肯定该蛋白质与细胞分化有关，该蛋白质表达受 N-myc 的调节。正常情况下其表达受原癌基因 N-myc 的抑制。它在 HepG2-HBx 细胞中高表达的意义值得关注。

Prohibitin-2 蛋白。PHB2 是细胞增殖的负调控蛋白，该蛋白质亚细胞定位于线粒体的内膜，也分布于细胞核中，最初发现其具有雌激素依赖的转录活性抑制，后来又发现 PHB2 具有多种功能，包括细胞周期调节、跨膜信号转导等^[27]。RNA 干扰实验表明，剔除 HeLa 细胞中的 PHB2 基因将会导致 caspase 依赖的细胞凋亡，而且 Prohibitin-2 蛋白在线粒体与核的相互作用中起重要作用^[28]。

转羟乙醛酶(transketolase)。在有氧条件下，恶性肿瘤中的糖代谢通过磷酸戊糖途径进行，而在缺氧条件下，则是由 Transketolase 来控制反应的。Foldi 等^[29]的研究表明，Transketolase 在乳腺癌组织中是高表达的，在正常组织中则正常表达，他们认为可以将 Transketolase 作为乳腺癌的标志物之一。

Enolase 1(烯醇酶)。烯醇酶也是参与糖酵解途

径的酶之一，催化 2- 磷酸甘油酸脱水生成磷酸烯醇型丙酮酸，在能量代谢过程中发挥作用。在关于乳腺癌、肝内胆管癌、小细胞肺癌的蛋白质组研究中发现肿瘤细胞中此酶表达上调。

热休克蛋白(heat shock proteins, HSPs)广泛存在于原核及真核细胞中，是细胞受刺激后被诱导产生的一组应激蛋白。除热损伤外，能由多种损伤因素诱导，具有维持细胞蛋白自稳、调节细胞周期等多种重要功能。HSP70 在正常细胞中表达水平较低，而在应激状态下可显著性升高。在病毒转化和化学诱导的肿瘤细胞中，HSP70 的表达比正常细胞高几倍。HSP70 对细胞凋亡有保护作用，HSP70 的高表达增强了肿瘤细胞抗凋亡能力，使细胞的生长抑制效应和死亡减弱^[30, 31]。但是，我们的结果表明，在转染 HBx 基因的 HepG2 细胞中 HSP70 的表达量反而降低。

内质网素(endoplasmin precursor)。内质网素在蛋白质的合成与折叠中起重要作用，内质网素能与 Ca²⁺结合，维持内质网的静态结构，在胚胎细胞的发育与生长过程中起重要作用^[32]。Jang 等^[33]证明 TNF-α 能够诱导 endoplasmin precursor 表达量上调。

3 结 论

我们对基于液相色谱质谱联用的非标记定量技

术进行了研究, 利用标准蛋白质的胰酶酶切肽段验证了蛋白质的质谱检测频次与蛋白质的量呈正相关, 利用在混合体系中插入不同浓度的标准蛋白质, 验证了在复杂体系中质谱检测频次数, 可以用来表征样品中某一蛋白质的量。

我们对不同的质谱检测频次归一化方法进行了比较, 一种以重复实验中共同鉴定到的蛋白质的平均质谱检测频次作为参照进行归一化, 一种以实验中总的质谱检测频次作为参照进行归一化。两种方法都有较高的灵敏度, 但基于蛋白质的平均质谱检测计数的归一化方法有较好的分辨差异的能力。将该方法用于 HepG2 和 HepG2-HBx 细胞系的差异表达研究, 发现了一系列的差异蛋白质, 这些蛋白质多与肿瘤发生和恶化相关。

因此, 我们建立和优化了一种基于液相色谱质谱联用的蛋白质组非标记定量方法。该方法的优点在于: 实验简单, 不需要任何同位素标记, 动态范围大。利用该方法进行差异蛋白质组研究和发现生物标志物在理论和临床上具有十分重要的意义。

参 考 文 献

- 1 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, 2003, **422**(6928): 198~207
- 2 Lambert J P, Ethier M, Smith J C, *et al.* Proteomics: from gel based to gel free. *Anal Chem*, 2005, **77**(12): 3771~3788
- 3 Link A J, Eng J, Schieltz D, *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol*, 1999, **17**(7): 676~682
- 4 Gygi S P, Rochon Y, Franza B R, *et al.* Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*, 1999, **19**(3): 1720~1730
- 5 Lilley K S, Razzaq A, Dupree P. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol*, 2002, **6**(1): 46~50
- 6 Wang W, Zhou H, Lin H, *et al.* Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal Chem*, 2003, **75**(18): 4818~4826
- 7 Wolters D A, Washburn M P, Yates J R. An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem*, 2001, **73**(22): 5683~5690
- 8 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 Am Soc Mass Spectrom*, 2003, **14**(9): 971~979
- 9 Shiio Y, Aebersold R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat Protoc*, 2006, **1**(1): 139~145
- 10 Dean R A, Overall C M. Proteomics discovery of metalloproteinase substrates in the cellular context by iTRAQ™ labeling reveals a diverse MMP-2 substrate degradome. *Mol Cell Proteomics*, 2007, (4): 611~623
- 11 Nelson C J, Hegeman A D, Harms A C, *et al.* A quantitative analysis of *arabidopsis* plasma membrane using trypsin catalyzed ¹⁸O labeling. *Mol Cell Proteomics*, 2006, **5**(8): 1382~1395
- 12 Conrads T P, Alving K, Veenstra T D, *et al.* Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and ¹⁵N-metabolic labeling. *Anal Chem*, 2001, **73**(9): 2132~2139
- 13 Griffin T J, Lock C M, Li X, *et al.* Abundance ratio dependent proteomics analysis by mass spectrometry. *Anal Chem*, 2003, **75**(4): 867~874
- 14 Anderson L, Hunter C L. Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. *Mol Cell Proteomics*, 2006, **5**(4): 573~588
- 15 Gao J, Opitck G J, Friedrichs M S, *et al.* Changes in the protein expression of yeast as a function of carbon source. *J Proteome Res*, 2003, **2**(6): 643~649
- 16 Ishihama Y, Oda Y, Tabata T, *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*, 2005, **4**(9): 1265~1272
- 17 Chen M, Ying W, Song Y, *et al.* Analysis of human liver proteome using replicate shotgun strategy. *Proteomics*, 2007, **7**(14): 2479~2488
- 18 States D J, Omenn G S, Blackwell T W, *et al.* Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol*, 2006, **24**(3): 333~338
- 19 Ono M, Shitashige M, Honda K, *et al.* Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Mol Cell Proteomics*, 2006, **5**(7): 1338~1347
- 20 Powell D W, Weaver C M, Jennings J L, *et al.* Cluster analysis of mass spectrometry data reveals a novel component of SAGA. *Mol Cell Biol*, 2004, **24**(16): 7249~7259
- 21 Liu H, Sadygov R G, Yates J R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem*, 2004, **76**(14): 4193~4201
- 22 Wang G, Wu W W, Zeng W, *et al.* 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**(5): 1214~1223
- 23 Old W M, Meyer-Arendt K, Aveline-Wolf L, *et al.* Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics*, 2005, **4**(10): 1487~1502
- 24 Marusawa H, Matsuzawa S, Welsh K, *et al.* HBXIP functions as a cofactor of survivin in apoptosis suppression. *EMBO J*, 2003, **22**(11): 2729~2740
- 25 Cangul, Hypoxia upregulates the expression of the NDRG1 gene leading to its overexpression in various human cancers. *BMC Genet*, 2004, **5**: 27

- 26 Sibold S, Roh V, Keogh A, *et al.* Hypoxia increases cytoplasmic expression of NDRG1, but is insufficient for its membrane localization in human hepatocellular carcinoma. *FEBS Lett*, 2007, **581**(5): 989~994
- 27 Wang S, Nath N, Adlam M, *et al.* Prohibitin, a potential tumor suppressor, interacts with RB and regulates E2F function. *Oncogene*, 1999, **18**(23): 3501~3510
- 28 Kasashima K, Ohta E, Kagawa Y, *et al.* Mitochondrial functions and estrogen receptor-dependent nuclear translocation of pleiotropic human prohibitin 2. *J Biol Chem*, 2006, **281**(47): 36401~36410
- 29 Foldi M, Stickeler E, Bau L, *et al.* Transketolase protein TKTL1 overexpression: A potential biomarker and therapeutic target in breast cancer. *Oncol Rep*, 2007, **17**(4): 841~845
- 30 Gurbuxani S, Schmitt E, Cande C, *et al.* Heat shock protein 70 binding inhibits the nuclear import of apoptosis-inducing. *Oncogene*, 2003, **22**(43): 6669~6678
- 31 Ravagnan L, Gurbuxani S, Susin S A, *et al.* Heat-shock protein 70 antagonizes apoptosis-inducing factor. *Nature Cell Biology*, 2001, **3**(9): 839~843
- 32 Luo S, Mao C, Lee B, *et al.* GRP78/BiP is required for cell proliferation and protecting the inner cell mass from apoptosis during early mouse embryonic development. *Mol Cell Biol*, 2006, **26**(15): 5688~5697
- 33 Jang W G, Kim H S, Park K G, *et al.* Analysis of proteome and transcriptome of tumor necrosis factor alpha stimulated vascular smooth muscle cells with or without alpha lipoic acid. *Proteomics*, 2004, **4**(11): 3383~3393

The Evaluation and Application of a Label-free Quantitative Proteomic Strategy Based on LC-ESI-MS/MS*

CHEN Ming^{1,2)}, YING Wan-Tao¹⁾, FANG Qin-Mei¹⁾, SUN Wei¹⁾, HE Fu-Chu¹⁾, QIAN Xiao-Hong^{1)**}

¹⁾ State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 100850;

²⁾ Tianjin Institute of Hygiene and Environmental Medicine, Tianjin 300050)

Abstract Mass spectrometry is being widely applied to identify and quantify proteins in complex mixtures. Quantification of small molecules by integration of LC-MS extracted ion chromatogram (XIC) peaks has a long history in analytical chemistry. Similar quantification techniques applied to proteolytic protein digests have also been previously described. A comprehensive approach for label-free quantification using yeast proteome as a model have been developed. Based on spectra counts of peptides, the relative protein quantification from LC-MS/MS experiments of proteolytic protein digests was performed. Unlabeled protein samples were digested with trypsin and separated by one-dimensional nano-flow HPLC (RPLC), and mass spectra were obtained by using the survey mode of LTQ mass spectrometer with dynamic exclusion. The correlating relationship between concentrate of protein and spectra counts was confirmed. Two algorithms to normalize spectra counts ratio from different samples were compared, and the results suggested that algorithm based on average spectra count (ASC) was eximious. The method was used to biomarkers discovery in HepG2 and HepG2-HBx cell lines. The identified proteins were analyzed and classified by cluster software (Version 3.0). Finally, 107 overlap proteins were identified, among them, 9 proteins were identified up-regulated (Ratio > 1.75) and 6 proteins down-regulated (Ratio < 0.5). Further research indicated that these proteins were related with liver cancer. Altogether, the results indicated that the strategy was operable and convenient with high sensitivity and wild dynamic range, and will be significant to deliver biomarker discovery either in theory or in clinic.

Key words LC-ESI-MS/MS, label-free, biomarker

*This work was supported by grants from The National Key Program for Basic Research of China (2001CB510201, 2004CB518707), The National Natural Science Foundation of China (20405017, 20505018, 20505019) and The Science and Technology Foundation of Beijing Municipality (H030230280190).

**Corresponding author.

Tel: 86-10-80727777-1231, E-mail: qianxhl@yahoo.com.cn

Received: August 24, 2007 Accepted: November 7, 2007