

# 质谱蛋白全谱分析和指纹分析

卞利萍 杨凡原 岳贵花 许崇峰

(复旦大学化学系生物分析研究室 上海 200433)

**[摘要]** 质谱蛋白全谱分析和指纹分析可用于鉴别蛋白。应用酶或化学切解、修饰及序列反应, 并与液相色谱或串联质谱联用, 可给出进一步的结构信息, 从而测定蛋白的组分、蛋白一级结构、表征序列、测定蛋白修饰位、基因变异位及缺失位, 这两种方法可用于医疗诊断、蛋白的功能分析和蛋白组学研究中。

**关键词:** 质谱 蛋白全谱分析 指纹分析

## 1 引言

近几年来, 随着质谱技术的发展和新型高性能仪器的出现, 在生物学领域中, 生物质谱代替一些传统的分析手段, 在常规生物分析方面显示出越来越大的优越性。本文主要介绍在蛋白质和 DNA 分析中的两种质谱方法: 质谱蛋白全谱(MS charting)和质谱指纹(MS mapping)分析。

首先, 我们通过比较这两种方法来定义这两个概念<sup>[1]</sup>。质谱蛋白全谱分析指的是组分(componential)分析, 它的对象是完整的组织、体液或其提取物, 其目的在于识别出尽可能多的肽和蛋白混合物中的组分。质谱指纹分析指的是组成(compositional)分析, 它的对象是单个的、纯化后的肽或蛋白, 其目的在于测定它们的一级结构、蛋白修饰和鉴别遗传差异等。蛋白全谱分析中, 测得未破坏的组分的分子质量, 并和已知或预期的蛋白分解产物的相应分子质量比较。而指纹分析中, 测得酶解或化学方法降解后的产物的分子质量, 并和已知或预期的蛋白分解产物的相应分子质量比较。两种方法都包含一个数据库搜索过程, 即将实验数据同蛋白或 DNA 数据库中的氨基酸或核酸序列比较、匹配。显然, 两种方法是互补的。比如, 蛋白全谱分析可以鉴别脑垂体中预测的神经肽, 然后, 再用指纹分析来研究脑垂体中每个独立组分的结构。

## 2 质谱蛋白全谱分析

九年前, Gottfried 等人最早使用质谱分析内分泌组织中的神经肽<sup>[1]</sup>。通常, 质谱蛋白全谱分析并不要求将蛋白样品从混合物中分离出来, 这样, 便可节省大量的时间和精力。因为, 理想上, 研究人员都想不经任何分离, 用组织或组织切片直接做质谱蛋白全谱分析。质谱蛋白全谱分析只要求提取或简单地分馏组分。它的适用范围远远超过放射性免疫检

定<sup>[2,3]</sup>和化学检定<sup>[4]</sup>的范围,这两种方法都只局限于特异系列的肽和蛋白。放射性免疫检定还要求至少知道分析物的部分一级结构,特别是在特异的抗原测定中。另外,在原则上,质谱蛋白全谱分析可在许多方面代替二维电泳(2D PAGE),并且由于质谱提供精确的分子量信息,所以又具有双向电泳所不具备的优点。

在应用上,质谱蛋白全谱分析可分为两大类。第一类,它可用于定性鉴别给定组织的肽和蛋白,从而可以检测出肽和蛋白的变异和缺失。这种技术在生物工程方面非常有用。例如,由于不同基因型的变异,可能产生一种被抑制的或缺失的酶,用蛋白全谱分析的方法可以检测到这种酶<sup>[5]</sup>。

质谱蛋白全谱分析的第二类用途是定量研究在对疾病、药物治疗、内源调节子或环境压抑(化学、物理或社会的)的响应中,肽或蛋白的整体响应轮廓的规则。对神经免疫内分泌复合物来说,其中的神经、免疫、内分泌系统共同形成一个错综复杂的通讯网络<sup>[6-9]</sup>,通过整个网络,响应可被单一的、局域的、规则的事件触发,因此,响应过程中需监测所有的肽和蛋白,而不仅仅是监测个别的组分。定量的质谱蛋白全谱分析可用以建立未鉴别组分在不同实验条件下的上限和下限规则,这样的信息有助于决定:应对哪一种未鉴别的肽和蛋白进行更深入的研究。这在蛋白组学研究中是极其有用的。

## 2.1 质谱蛋白全谱分析的步骤

质谱蛋白全谱分析通过测定肽和蛋白的分子量,加上已知蛋白和DNA 序列的信息,来试探性地确认给定组织或体液中存在的肽和蛋白。通常,质谱蛋白全谱分析可分为四步:(1)从生物组织提取多肽和蛋白并分馏提取液;(2)测定提取液中各组分的分子量;(3)从肽或蛋白的计算机库中搜索被测的分子量是否符合特定的肽或蛋白的质量;(4)再确认分析,如部分的N-或C-端的测序分析,氨基酸分析,或反应MS/MS分析。这些再确认分析的对象应是那些在(1)-(3)步中仍不能十分确认的特定肽或蛋白。

## 2.2 对仪器的要求

质谱蛋白全谱分析对仪器有一定的要求:离子源应当能提供非破坏性的、气态的、包含提取物中的所有多肽和蛋白的质量信息,最好在物种间没有歧视效应,及不形成碎片离子。离子化中质量的上限及质量分析器的质量上限必须超过待测物可能的最大质量。定量MS的离子化过程必须足够稳定,质荷比信号波动<10%。目前,基体辅助激光解析(MALDI)和电喷雾电离(ESI)作为质谱蛋白全谱分析的新型离子源,具有令人满意的性能。

对实验室质谱而言,要完成蛋白全谱分析,仪器的质量范围、分辨率、质量精度和灵敏度必须满足一定的标准。在实际操作中,测定还受被分析样品的自然状况、进样方式、离子化方式、电荷状态等条件影响。多年来,质谱蛋白全谱分析已在实践上获得了关于上述条件的一些经验。仪器的质量上限以100,000-200,000为佳,可用于大多数神经肽、激素、抗体和其它有信号响应的分子。在现有的几种质谱仪中,离子源的组合有效性、质量分析器的性能和检出器的响应可给出pmol级(或更低)样品量的响应灵敏度。傅里叶质谱仪已可获得相当于分子同位素轮廓宽度的质量分辨率,即在半峰高处大约0.1%的相对分子质量(RMM),或相当于在最大质荷比( $m/z$ )处1000左右的分辨能力。

由于蛋白或DNA 数据库的不断丰富,原则上,将实验得到的质量同库中数据对照便

可推断未知的肽或蛋白。但一般来讲,数据库也许永远不可能完备,因为作为一个理想的数据库,应没有错误的DNA和蛋白信息,有所有最终产物的信息,所有相关的转译后修饰和所有相关的人工修饰信息。而且,即使经过搜索后“匹配”了分子质量,也不能保证分子的确认,因两种肽或蛋白可能有相同的分子质量。可幸的是,如果除了分子量的信息之外还有一些结构信息的话,搜索问题就不是那么复杂。例如,如果2-5个氨基酸序列及位置是知道的话,搜索的准确度会提高几个数量级。<sup>[10,11]</sup>这需要利用串联质谱(MS/MS)的功能,在不影响质谱蛋白全谱分析的情况下,来获得部分结构信息。另外,色谱-串联质谱联用(LC/MS/MS)的方法可以测定氨基酸序列,并与蛋白库中的数据很好吻合。当然,有条件的话,总可以事先用酶切、碎片、测序反应的方法加上蛋白全谱质量谱来获得结构信息。

### 2.3 质谱蛋白全谱分析的应用

Jimenez等将已知的蜗牛的单个神经元放在MALDI基体液中,并将单元粉碎,不再经其它处理,用于MALDI-TOFMS分析,发现了一些新的神经肽的表达和操作模式。这个例子表明MALDI灵敏度高、适用性强,可以成功地用于全谱分析。但只用MALDI-TOFMS,给出的信息较少。

较之只显示单电荷离子峰的MALDI-MS,ESI-MS谱图显示的是多电荷离子峰,谱图较难解释。但ESI-MS易与HPLC在线连接,故先用HPLC对样品进行一定分离后,用ESI-MS分析,成为蛋白全谱分析的标准方法。此外,ESI离子源非常稳定,分析离子信号与分析物浓度成正比,所以可用毛细管液相色谱(dLC)获得超低检出限,并适用于高精度的定量差分蛋白全谱分析。

考察一下可分析的最低浓度和检出限,dLC可分析 $\mu\text{mol}$ 的肽和蛋白流出液,用这种方法(dLC/ESI-MS)已经分析了鼠脑垂体和脑提取液中100多种肽。质谱蛋白全谱分析中,只需用单个脑垂体提取液的1/10-1/20,就催乳素和生长激素而言,只要1%的提取液,就可得到清晰的谱图(图1)<sup>[11]</sup>。

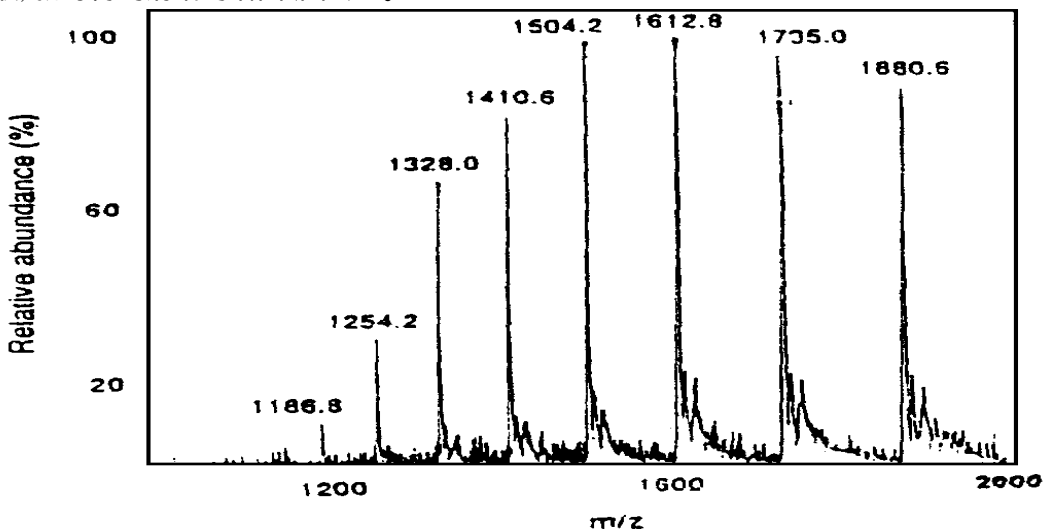


图1 鼠的催乳激素的dLC/ESI-MS谱,样品是从一只成年母鼠的脑垂体前叶中得到的提取液的1%

定量的差分 (differential) 质谱蛋白全谱分析和在线 LC/MS/MS 可挑选出那些未知物进行进一步的细致分析。LC/ESI/MS 和 LC/ESI/MS/MS 的灵敏度相当高, 可以利用老鼠模型定量研究人类疾病<sup>[12-14]</sup>。对照正常的和非正常的鼠的质谱, 只选择那些在生物上相关的未知肽作细致分析。另外, 质谱蛋白全谱分析在鉴定未知肽或蛋白方面的作用被日益重视。Feistner 鉴别了鼠  $\beta$ -LPH 的两种未知的变异体<sup>[11]</sup>。

### 3 质谱指纹分析

指纹分析广泛地用于肽、蛋白和 DNA 的鉴别。原先使用毛细管电泳 (CE)<sup>[15]</sup>, 现在, 质谱由于其灵敏、通用、高生产率等特点, 正日益成为指纹分析强有力的工具。纯化后的蛋白被酶或化学裂解成肽片段, 根据蛋白一级结构和所用裂解试剂, 给出不同蛋白的特征性指纹图谱, 用作鉴别的依据。指纹分析可用于测定蛋白的一级结构、表征氨基酸序列、测定一些位点和转译后修饰位和氨基酸变异位、缺失位。利用不同物种的保留特性可作功能分析, 还可用于基因诊断。

#### 3.1 质谱仪器要求

指纹分析可用激光解析 (PD)、离子喷雾电离 (ESI)、基体辅助激光解析 (MALDI) 做离子源<sup>[16]</sup>, 采用多种模式, 如线性质谱、反射质谱、正离子或负离子质谱等<sup>[17]</sup>。相比其它离子化方法, MALDI 有以下几个优点: (1) 制样简便, 对纯化的要求低; (2) 可与时间飞行质谱 (TOF-MS) 联用; (3) 宽的质量范围 (> 400kDa); (4) 灵敏; (5) 快速; (6) 采用延迟采样 DE 技术后, 更大大提高了分辨率和质量精度, 使蛋白和 DNA 的识别更容易。

#### 3.2 分析步骤

质谱指纹分析一般包括以下几步: (1) 纯化蛋白或 DNA 样品; (2) 用酶或化学的方法切解蛋白或 DNA; (3) 用质谱分析切解后的产物; (4) 用质谱数据进行数据库搜索, “匹配”的数据作为识别的依据, “非匹配”的数据可鉴定一些结构的变化。 (5) 对各个多肽片断进行 MS/MS 分析, 获得序列信息。

纯化最常用的方法是十二烷基磺酸钠-聚丙烯酰胺凝胶电泳 (SDS-PAGE)。但是, SDS-PAGE 由于其有限的分辨率常导致蛋白谱带的相对位移或重叠。即使应用双向电泳, 由于分子大小、电荷相近, 蛋白的转译后修饰或疏水蛋白的“条纹”会使蛋白点模糊, 也给分析带来了蛋白混合物的问题。解决这个问题要求获得大量的肽的序列信息, 这些序列分别在质量指纹谱和序列数据库中寻找。另外一种解决方法是提高质谱的质量精度 (< 30ppm), Perkin-Elmer (PE) 公司的质谱仪已可获得几个 ppm 的质量精度, 并应用延迟取样 (Delayed Extraction 或 DE) 技术<sup>[18]</sup>, 大大增强了质谱方法在指纹分析中的应用能力。

Jensen<sup>[19]</sup>等推荐使用非冗余的蛋白序列数据库, 它允许的最大质量错误率为 50ppm, 识别一个蛋白至少要求匹配 7 个肽质量, 并考虑每个肽的两种不完全断裂, 对蛋白的等电点无限制。找到的假设蛋白按所匹配的肽质量数排序, 然后列表分析。

#### 3.3 指纹分析的应用

Jensen 提出蛋白混合物分析的几点规则: (1) 质谱测得的肽质量有多少符合数据库中的肽质量? (2) 匹配过程中质量精度如何? (3) 匹配中覆盖整个序列的百分比是多少?

(4) 匹配中同背景有多大差别? (5) 匹配后是否有所期望的蛋白及期望的分子量? (6) 匹配得到的蛋白序列是否重复出现? 是否有不完全的末端 (ragged end) 形式? 大多数的离子信号是否得到了匹配?

从数据库中获得一个假定的匹配物后, Jensen 提出一个“二次路径搜索” (second pass search) 策略, 即在测得的和推测的肽分子量间进行细节比较。在这一步中, 应考虑不完全的断裂和肽修饰, 如甲硫氨酸及 *s*-脲基甲基半胱氨酸的氧化。二次路径搜索能给出匹配肽的一些重要信息, 比如它们在蛋白序列中的分布。经实践证明, 这种方法非常直接, 在识别蛋白混合物方面很有用。Jensen 用此法识别了多组蛋白的混合物, 并指出, 质量精度越高, 搜索的确定性越高。例如, 在识别两个 280kDa 的蛋白混合物中, MALDI 肽质量指纹谱共有 143 个信号 (见图 2), 第一次搜索得到鼠的  $\alpha$ -spectrin, 对应 68 个肽质量, 二次路径搜索对应 89 个肽质量, 覆盖率 39%, 则  $\alpha$ -spectrin 被识别。第二次搜索得到鼠的  $\beta$ -spectrin, 对应 28 个肽质量, 二次路径搜索对应 34 个肽质量, 覆盖率 17%, 则  $\beta$ -spectrin 被识别。第三次搜索没有显著匹配的蛋白。这样, 解释了 143 个信号中的 123 个 ( $89 + 34 = 123$ ), 并且所有的强信号都得到对应。

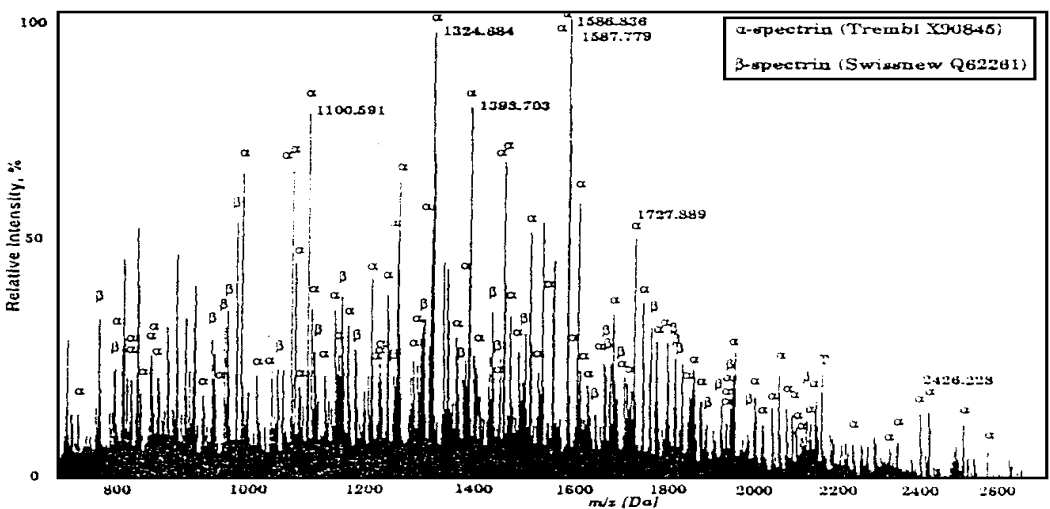


图 2 两个分子量约为 280kDa 的鼠蛋白混合物的 MALDI 肽质量指纹图谱

直接用肽的质量与序列库比较, 识别或表征蛋白结构, 有时仍有混淆, 比如两个肽有非常相近的质量, 同时酶解也会产生未预期的断裂。用 MS/MS 可解决这类问题。K. Hirayama<sup>[20]</sup>等在作鼠的单克隆免疫球蛋白 (IgG2b) 的指纹谱时, 对酶解后的一段肽段无法区分是 TDSFSCNVRHEGLKNYYLK (平均分子量 2332.6u) 还是 GLVRA PQVY-LPPPAEQLSRK (平均分子量 2332.8u), 做此肽段的 MS/MS 谱 (见图 3), 根据断裂的序列信息, 可推知其为后者。在指纹谱中, 由于酶解产生未预期的断裂, 这类肽段很难在序列库中找到匹配的质量。MS/MS 分析, 可得到正确的序列, 并指出酶切的又一个可能点。

质谱指纹分析还可测定不同物种间的保留特性, 从而推断分子的功能<sup>[21]</sup>。由生物多样性和进化上远离引起的氨基酸残基取代, 显示了蛋白中的特征功能区。Cordwell 等用延伸因子 (EF), 比较不同物种的肽片段, 检测分子的氨基酸序列同一性。这里, 指纹被用

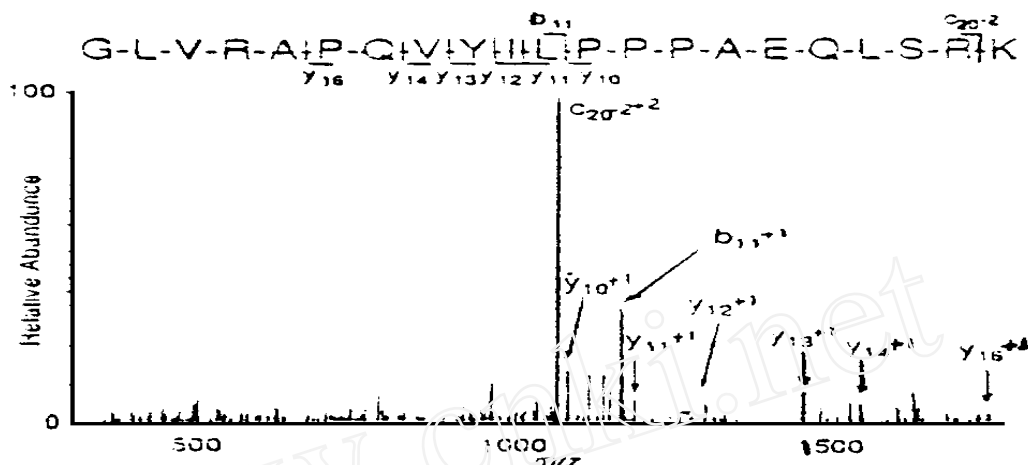


图 3

于识别不同物种的保留蛋白全谱质量和蛋白全谱邻近区域的序列, 将这些质量和序列同现有的数据库比较, 从而识别出保留的氨基酸序列。一旦观察到这个序列在三个以上物种中出现, 即假设已检测到可能与功能相关的一个保留特性。除了在基因组测序中将推测的功能与检测到的蛋白对应起来, 这种方法还可能在变异基因研究中识别分子内的变异区域。

在一个蛋白消解物中, 质谱指纹分析可用来检测在化学或酶处理前后的“非匹配”(即和预测片段不符)的肽, 从而表征蛋白的修饰。例如, 用特殊的内/外糖苷酶分解糖蛋白, 然后用质量指纹谱检测相对于假定的糖基化肽的质量位移, 可以测出聚糖结构的位置和形式。并且, 比较 DTT 还原前后的指纹谱能表征蛋白中的二硫键。Kusmann<sup>[17]</sup>等人分别用正离子、负离子、线性和反射模式, 研究了在 *in situ* (直接在膜上或凝胶中的消解) 条件下和溶液中的 neurolin 的 cDNA 序列和 neurolin 的氨基酸序列同一性, 并检出了糖基化位。

Wade<sup>[22]</sup>等用质量指纹图谱对比分析了正常人和先天脑积水病人限制性酶解前后的 DNA 互补链, 并检测到病人 LICAM 基因的变异。

Zhao 和 Chait<sup>[23]</sup>利用质谱作单元性抗体的线性抗原决定基的快速指纹谱, 在抗原保持原始构型的条件下, 测定了构型非连续的抗原决定基。

#### 4 总结与展望

质谱蛋白全谱分析和指纹分析目前的主要限制是: 数据库的错误和不完全, 以及目前仪器的质量范围和分辨率的限制。随着数据库的迅速扩充和完善, 以及仪器性能的提高 (PE 公司报道其质谱的质量精度已可达 1ppm), 这两种技术一定会在生物领域中发挥越来越大的作用。

#### 参 考 文 献

1 G J Feistner *et al* J Mass Spectrom, 1995, 30: 519- 530

- 2 A N Serafini *J Nucl Med*, 1993, 34: 533
- 3 T Chard In *An Introduction to Radioimmunoassay and Related Techniques*, edited by R H Brudon and P H van Knippenberg, *Laboratory Techniques in Biochemistry and Molecular Biology*, 1990, Vol 6, Part II, Elsevier, New York
- 4 K Takamatsu *et al Neurochem, Res*, 1992, 17: 239
- 5 R E Hill *Clin Chim Acta*, 1993, 217: 3
- 6 D A Nib *et al J Immunol*, 1993, 150: 5281
- 7 B L Spangelo *et al TEM*, 1990, 2: 408
- 8 D A Weigent *et al Immunol Rev*, 1987, 100: 79
- 9 E J Goetzl *et al FA SEB J*, 1992, 6: 2646
- 10 M Mann In *Approaches to the Practical Use of MS/MS in a Protein Sequence Facility*, eds A L Burlingame and S A Carr
- 11 M Mann *et al Anal Chem*, 1994, 66: 4390
- 12 T J I Gill *et al Science*, 1989, 245: 269
- 13 P Zelazowski *et al Neuroendocrinology*, 1992, 56: 474
- 14 K Timmers *et al Metabolism*, 1990, 39: 378
- 15 罗国安等 色谱, 1997, 15(4): 305
- 16 A L Burlingame *et al Anal Chem*, 1996, 68: 629R
- 17 M Kussmann *et al J Mass Spectrom*, 1997, 32: 483- 493
- 18 Eric Winter *Biospectrometry: 蛋白质研究和新药发现技术进展讲座* (PE 公司应用生物系统部) 1998: 52
- 19 Ole N Jensen *et al Anal Chem*, 1997, 69: 4741- 4750
- 20 K Hirayama *et al Anal Chem*, 1998, 70: 2718- 2725
- 21 S J Cordwell *et al J Mass Spectrom*, 1997, 32: 370- 378
- 22 Yoshinao Wada *J Mass Spectrom*, 1998, 33: 187- 192
- 23 A F R Huhmer *et al Anal Chem*, 1997, 69: 40R

## MS Charting and Mapping

Bian Liping, Yang Pengyuan, Yue Guihua, Xu Chongfeng

(Department of Chemistry, Fudan University, Shanghai 200433, China)

Received 1999-01-25

### Abstract

MS charting and mapping can be used in identifying protein. By using enzymatic or chemical cleavage, modification and sequencing reaction, with LC-MS or LC-MS/MS, scientists can gain more structural information so as to detect the composition and primary structure of protein, the sites of decoration in protein, mutation and deletion in genes. These two methods can be used in medical diagnosis and functional analysis of protein, as well as in proteome study.

Key Words: mass spectrometry or MS, MS charting, MS mapping