

蛋白质组与生物质谱技术

钱小红

(军事医学科学院国家生物医学分析中心 北京 100850)

[摘要]蛋白质组是基因组研究的继续。蛋白质组的研究是为了识别及鉴定一个基因组、一个细胞或组织所表达的全部蛋白质以及它们的表达模式。以基质辅助激光解吸附飞行时间质谱和电喷雾质谱为代表的现代生物质谱技术,为蛋白质组的研究提供了必要的技术手段,被称为蛋白质组研究的三大关键性支撑技术之一。而质谱-质谱联用、质谱与其它技术联用以及高产出筛选技术的应用和发展,将使蛋白质组的研究在高准确度、高灵敏度以及大规模化水平上的发展成为可能。

关键词:蛋白质组 质谱

随着基因组计划的快速、大规模推进,成千上万种基因序列的不断问世,使生物学家们面临着一个新的巨大的挑战,这就是如何在生理或生理病理条件下观察细胞或生物组织数千种基因的全部表达,即如何阐明基因的结构与其功能的相互关系^[1]。问题的提出是基于下列的思考,单独的 DNA 序列信息无法解释 1. 基因的翻译在何时发生以及它是否发生了;2. 基因产物的表达量是多少;3. 翻译后修饰的程度如何?;4. 基因剔除与过量表达的后果;5. 小于 300 碱基对的基因或开放阅读框架的检测;6. 外部环境如药物、不同细胞周期、个体发育、衰老、应激和疾病条件下的基因表达等^[2]。为了回答上述问题,蛋白质组(Proteome)的概念应运而生。

1 蛋白质的定义与发展

蛋白质组的概念是由澳大利亚的科学家 Wikings 和 Wasinger 等于 1995 年提出的^[1,3]。其定义是指一个基因组、一个细胞或组织所表达的全部蛋白质成分。与基因组概念不同的是,蛋白质组作为一个整体,在不同的条件下,在一个生物体的不同组织中是不相同的。而一个生物体仅有一个特定的基因组。蛋白质组是一个基因组的直接产物。然而一个蛋白质组中的蛋白质数量可能超过基因提供的数量,这是由于基因的拼接和翻译后的修饰造成的。因此,蛋白质组的研究是为了识别及鉴定一个细胞或组织所表达的全部蛋白质以及它们的表达模式。

自蛋白质的概念问世以来,在近三年左右的时间里,由澳大利亚的悉尼大学和马克奎

1998-10-25 收

尔大学作为先导,已发展到目前有 10 多个国家的多所大学和研究机构参与。研究所涉及的生物体已达十多种,包括人心脏和肝脏在内的 10 余个蛋白质组数据库已初步建立,其中包括第一个完整的蛋白质组数据库-酵母蛋白质组数据库(YPD),含蛋白质 6021 种。包括人体在内的 4 种生物体的 16 个二维凝胶(2-D PAGE)数据库也已建立。研究所涉及到的应用范围包括蛋白质组作图、蛋白质组分鉴定、蛋白质数据库构建、新蛋白质的寻找、蛋白质翻译后修饰的分析、细胞周期、发育、肿瘤发生、发展、环境影响、新药研究等多个领域^[4-12]。

蛋白质组研究的技术路线流程见图 1。蛋白质组研究的基本技术包括二维凝胶电泳(2-DE)技术、图象及数据处理技术及质谱(MS)技术。这三大支撑技术在各自的领域都已发成为相对成熟的技术,但是如何将它们有效地连接与应用,是蛋白质组研究在技术方法上的关键所在。二维凝胶电泳技术是由以等电点对蛋白质进行分离的一相等电聚焦电泳和以分子量对蛋白质进行分离的二相 SDS-聚丙烯酰胺凝胶电泳组成^[13]。由于固相化 pH 梯度电泳胶(IPG)的引进,使得二维凝胶电泳的分辨率和重现性大大改善,从而使在不同时间和不同实验室所得到的电泳胶可以进行比较^[14]。目前,二维凝胶电泳已发展到可将复杂样品中多达 10,000 个蛋白质点进行有效分离的水平^[15]。高分辨的图象扫描和计算

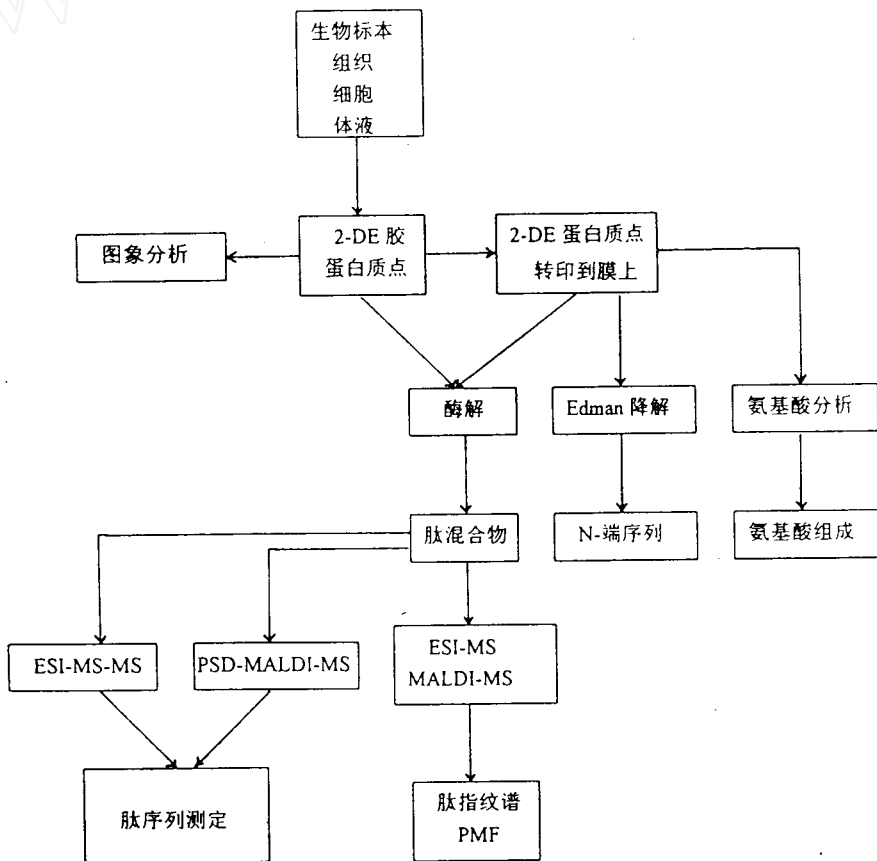


图 1. 蛋白质组研究的技术路线流程图

机软件数据处理系统,可对二维电泳分离的蛋白质点进行扫描和数据处理,不仅可揭示表达蛋白质的量,而且可对不同时间,不同细胞周期,以及正常和病理条件下的电泳胶进行定量对比并构建数据库^[16-18],鉴定 2-DE 分离的蛋白质点的技术手段包括质谱技术(Mass Spectrometry, MS)、Edman 降解 N-端序列测定技术以及自动氨基酸组成分析技术等。其中以质谱技术应用最为广泛,被称为蛋白质组研究的三大关键性技术之一。

2 生物质谱技术

诞生于八十年代末期的两项软电离质谱技术——电喷雾质谱技术(Electrospray Ionization Mass Spectrometry, ESI-MS)和基质辅助激光解吸附质谱技术(Matrix Assisted Laser Desorption/Ionization, MALDI)使传统的主要用于小分子物质研究的质谱技术发生了革命性的变革^[19-20]。电喷雾质谱技术和激光解吸附质谱技术所具有的高灵敏度和高质量检测范围,使得在 pmol(10^{12})甚至 fmol(10^{-15})的水平上准确地分析分子量高达几万到几十万的生物大分子成为可能,从而使质谱技术真正走入了生命科学的研究领域,并得到迅速的发展。

2.1 ES-MS^[21]

ESI 是在毛细管的出口处施加一高电压,所产生的高电场使从毛细管流出的液体雾化成细小的带电液滴,随着溶剂蒸发,液滴表面积缩小,导致分析物以单电荷或多电荷离子的形式进入气相。电喷雾离子化的特点是产生高电荷离子而不是碎片离子,使质量电荷比(m/z)降低到多数质量分析仪都可以检测的范围,因而大大扩展了分子量的分析范围。离子的真实分子质量可根据质荷比及电荷数算出。电喷雾质谱技术可以耐受少量的缓冲溶液、盐和去垢剂,这些物质可与分析物形成加成物而增加分子质量测定的难度,也可抑制分析物离子的形成。因而最佳的离子形成是在无缓冲溶液,无盐和去垢剂的情况下。一个方便的方法是在离子化前采用高效液相色谱(HPLC)技术将被分析物与杂质分离。电喷雾质谱的优势就是它可以方便地与多种分离技术联用,如液-质联用(LC-MS)等。

2.2 MALDI^[22]

MALDI 的基本原理是将分析物分散在基质分子(尼古丁酸及其同系物)中并形成晶体,当用激光(337nm 的氮激光)照射晶体时,由于基质分子经辐照所吸收的能量,导致能量蓄积并迅速产热,从而使基质晶体升华,导致基质和分析物膨胀并进入气相。MALDI 的最大优点是它能耐受较高浓度的缓冲溶液、盐和去垢剂的存在。但是高质量质谱图的获得也需要尽可能去除盐和去垢剂等。一个方便的方法是采用冷水对已形成的晶体进行简单的清洗以去除盐及其杂质。MALDI 所产生的质谱图多为单电荷离子,因而质谱图中的离子与多肽和蛋白质的质量有一一对应关系。由于 MALDI 产生的离子常用飞行时间(time-of-flight, TOF)检测器来检测,所以 MALDI 常与 TOF 连在一起,称为基质辅助激光解吸附飞行时间质谱(MALDI-TOF-MS)。

2.3 串联质谱技术^[23-25]

随着多肽蛋白质质谱分析技术的发展,各种质量分析技术被引进这个领域,使串联质谱(MS-MS)技术在阐明蛋白质结构中起到重要的作用。串联质谱技术是通过离子在运动过程中发生的自然或人为的质量或电荷的变化,研究母离子和子离子的关系,从而获得碎

裂过程的信息,并进而推测多肽或蛋白质的结构。三级四极(Triple quadrupole)质谱、飞行时间(Time of flight)质谱和四极离子阱(Quadrupole ion trap)质谱是目前应用最广泛的三种串联质谱技术。

3 质谱技术在蛋白质组研究中的应用

质谱用于2-DE分离的蛋白质点的鉴定主要通过多肽、蛋白质的质量测定、肽指纹谱测定以及氨基酸序列测定等方法来完成。其中以肽指纹谱的MALDI-TOF-MS测定与串联质谱所进行的序列测定为目前采用最多的两种方法

3.1 肽指纹谱(Peptide mass fingerprinting, PMF)测定

肽指纹谱的测定是对2-DE分离的蛋白质点采用蛋白酶解或化学降解,对所得多肽混合物进行质谱分析的方法。最常采用的策略有两个:一是对2-DE分离的蛋白质点进行胶上原位降解,提取、收集降解产物(必要时做适当的脱盐或浓缩处理),做质谱分析。另一种是将2-DE分离的蛋白质进行膜转印(常用PVDF膜),对转印到膜上的蛋白质做蛋白酶解或化学降解,再做质谱分析,对质谱分析所得肽片与多肽蛋白数据库中蛋白质的理论肽片进行比较,从而判别所测蛋白是已知还是未知。由于不同的蛋白质具有不同的氨基酸序列,因而不同蛋白质所得肽片具有指纹的特征^[26]。

由于MALDI所具有的高灵敏度以及它对盐和去垢剂等耐受性,在肽指纹谱的鉴定中,常采用MALDI-TOF-MS的方法。

采用肽指纹谱的方法已对酵母、大肠杆菌、人心肌等多种蛋白质组进行了研究。其中对人肥大心肌细胞的蛋白质研究达到fmol的水平。对大肠杆菌经PVDF膜转印的蛋白质在亚-pmol的水平上进行的研究表明,三个肽片既可达到对蛋白质的正确识别。而采用原位酶解的方法对酵母蛋白质组研究的结果显示,约90%的蛋白质被正确识别,多于32个新蛋白质被发现,而这些蛋白是酵母基因组研究中未能识别的开放阅读框架。采用基因删除与蛋白质组研究相结合的方法,对大肠杆菌蛋白质组进行的研究表明,蛋白质组的分析不仅可以发现基因组中约40%无同源性的开放阅读框架,而且可以了解生物体对环境因素影响的反应。对啤酒酵母的蛋白质组研究同样表明了环境因素改变,如热休克、热敏突变、突变影响翻译后的修饰等与2-DE蛋白质谱图的对应关系。对嗜血流感菌的蛋白质组的研究显示,肽指纹谱的方法比氨基酸组成分析更为可靠,这是因为MALDI测定肽质量的准确度0.1%—0.05%,而氨基酸组成分析的准确度为10%。另外MALDI可以耐受少量杂质的存在。对肽指纹谱的方法所进行的系统研究表明,大于5000Da的肽片很容易丢失,因而对蛋白质的检索与识别贡献较小,而由胰酶降解产生的小于3000Da的肽片是最有用的^[27-32]。

对肽指纹谱分析最大的挑战就是计算机检索前测定数据的判读。由于蛋白酶的自降解及杂质的存在,使峰的识别及谱图的分析变得复杂。另外,由于翻译后的修饰,使真核生物蛋白质的鉴定增加了难度。电泳过程中两种人为修饰(半胱氨酸丙烯酰胺加成物和甲硫氨酸的氧化)的引入,也会影响肽质量测定的准确^[33]。

总之,肽指纹谱测定的成功取决于质量测定、数据库检索的准确性以及蛋白质的纯度。因此,有时需要第二种方法的确证,如N-端测序、氨基酸分析以及串联质谱MS-MS)

测序的方法。

3.2 肽序列串联质谱分析

串联质谱技术用于肽序列测定是通过采用不同的质谱技术选择具有特定质荷比的离子,并对其进行碰撞诱导解离,通过推断肽片的断裂,可将肽序列导出。通常在低能气相碰撞的条件下,肽离子碎片沿着骨架在酰胺键处断裂,产生一个序列离子阶梯。如果电荷保留在碎片离子的 N-端,将产生 b-形离子。如果电荷留在 C-端,将产生 y-形离子。通过减掉相邻序列离子的质量,由一种或两种离子系列可推测出氨基酸的序列^[34]。对断裂机理的研究表明,碱性氨基酸残基(精氨酸和赖氨酸)在肽链骨架结构中的位置,将决定序列离子单一类型的多寡。如果精氨酸或赖氨酸在 N-端,串联质谱得到的主要是 b-形离子,如果碱性氨基酸残基出现在 C-端,y-形离子将占主导地位。然而,如果碱性氨基酸出现在肽链骨架中部的某个部位,两种类型的离子都将产生,并增加了谱图的复杂性^[35-38]。

超高灵敏度的纳升电喷雾串联质谱技术(nano-electrospray MS-MS)已被成功地用于 SDS 电泳胶分离的蛋白质。在这个技术中,由原位胰酶降解产生的肽混合物被提取后,通过一个约含 10 μ l 吸附剂的毛细管,毛细管吸附的肽段被逐步洗脱,并以约 1 μ l 的体积进入质谱的离子源。由于 nano-electrospray MS 离子源的特殊结构,使其产生的液滴比普通电喷雾质谱产生的液滴小 100 倍,因而有效的利用了样品,减少了损失,提高了灵敏度。此种方法特别适用于微量样品的质谱分析。采用此方法先对 5ng(80fmol)BSA 经胰酶原位降解产生的肽段进行了序列测定,进而对支原体感染的含有强烈抑制内皮细胞增殖成分的肿瘤细胞培养基进行观察,寻找血管生成抑制剂。对经 SDS-PAGE 分离的蛋白质采用上法进行分析,获得了 7 个肽片的序列信息,计算机检索发现,其序列与从支原体获得的精氨酸脱亚氨基酶高度同源^[39]。

采用串联质谱对 2-DE 分离的 *H. influenzae* 蛋白质组进行分析,鉴定了 260 个蛋白质点,其中有 26 个点含有两个或两个以上的蛋白质。三个蛋白质被发现其基因含有框架移动误差(frameshift error)。一个蛋白质点在 *H. influenzae* 数据库找不到与串联质谱数据相对应的蛋白质,而从对大肠杆菌数据库的检索中,却发现了一个与之能很好对应的蛋白质-色氨酸酶^[40]。

一种被称为猎枪(Shotgun)法的串联质谱技术已经被用于混合物中蛋白质的鉴定。其基本过程是通过用串联质谱鉴定由蛋白酶解产生的肽片来识别混合物中的蛋白质。三种基本技术使得这种方法成为可能,一是与 ESI 联用的 LC 分离技术,使得复杂混合物中的成分在进入串联质谱前获得分离,而通过数据依赖控制(data-dependent control)所提供的即时数据采集功能,大大提高了数据采集的效率,最后是直接而自动的肽串联质谱与计算机算法及数据库分析联用。运用此方法已成功地对大肠杆菌 2-DE 谱图进行了分析,并鉴定了含有两个以上蛋白质的蛋白质点^[41]。

3.3 高产出筛选蛋白质组研究技术(High-throughput screening, HTS)^[2]

随着蛋白质组研究的快速推进,现有技术方法已不能满足高效快速大规模样品处理的要求。世界上各大制药公司和许多著名企业纷纷出资开展高产出筛选蛋白质组技术的研究。一个基于对 2-DE 分离、PVDF 膜转印的蛋白质进行自动处理分析的机器人系统已经诞生。其最低设计 requirements 是每天完成 1000 个蛋白质的分析。一个理想的 HTS 系统应当

达到1. MALDI-TOF-MS以每天完成1000个蛋白质鉴定的速度对2-DE分离,胰酶或其他蛋白酶降解的蛋白质进行鉴定;2. ESI-MS-MS或猎枪法以每台机器每天分析几十个蛋白质标记序列的速度对第一步不能识别的蛋白质进行鉴定;3. 串联质谱对全新蛋白质或特别有兴趣的蛋白质肽片的全序列测定。

4 结语

蛋白质组的研究已开辟了一个广阔而重大的生命科学的崭新研究领域,其研究成果势必大大丰富基因组计划的研究结果,并将使人类对生命的本质、其发生发展过程的认识达到一个前所未有的崭新的高度。

参考文献

- 1 Wilkins M R Sanchez J-G,Gooley A A et al Biotechnology and Genetic Engineering Reviews,1995,13,19-50
- 2 Humphery-Smith I,Cordwell S J,Blackstock W P Electrophoresis 1997,18,1217-1242
- 3 Wasinger V C,Cordwell S J,Cerpa-Poljak A et al. Electrophoresis 1995,16,1090-1094
- 4 Cordwell S J,Basseal D J et al. Electrophoresis,1997,18,1335-1346
- 5 Link A J,Eng J et al. American Laboratory,1996,28-30
- 6 Humphery-Smith I,Blackstock W. J Protein Chemisttry,1997,16,537-544
- 7 Fey S J,Nawrocki A et al. Electrophoresis,1997,18,136-1372
- 8 O'Connor C D,Farris M et al. Electrophoresis,1997,18,1483-1490
- 9 Sazuka T,Ohara O. Electrophoresis,1997,18,1252-1258
- 10 Bini L,Heid H et al. Electrophoresis,1997,18,557-562
- 11 VanBogelen R A,Abshire K Z et al. Electrophoresis,1997,18,1243-1251
- 12 Nowak R. Science,1995,270,369-370
- 13 O'Darrell P H. J Biol Chem 1975,250,4007
- 14 Righetti P G. Laboratory Techniques in Biochemistry and Molecular Biology,R H Burdon and P H van Knippenberg,eds Elsevier,Ameterdam,1990
- 15 Klose J,Kobalz U. Electrophoresis,1995,16,1034
- 16 Olsen A D,Miller M J. Analytical Biochemistry,1988,169,49-70
- 17 Wirth P J,Luo L D et al. Electrophoresis,1993,14,1199-1215
- 18 Appel R D,Hochstrasser D F et al. Electrophoresis,1991,12,722-735
- 19 Fenn J B,Mann M et al. Science,1989,246,64
- 20 Karas M,Hillenkamp F. Anal Chem 1988,60,2299
- 21 Roepstorff P. Current Opinion in Biotechnology,Hensley P and Smith L Meds 1997,18,6-13
- 22 Yates J R. J Mass Spectrometry,1998,33,1-19
- 23 Yates J R. Methods Enzymol 1996,271,351
- 24 March R E. J Mass Spectrom 1996,32,351
- 25 Guilhaus M. J Mass Spectrom 1995,30,1519
- 26 Wise M J,Littlejohn T G et al. Elelectrophoresis,1997,1399-1409
- 27 Shevchenko A,Jensen O N et al. Proc Natl Acad Sci USA,1996,14440-14445

- 28 Garrels J I, McLaughlin C S et al. Electrophoresis, 1997, 18, 1347-1360
- 29 Dainese P, Staudenmann W et al. Electrophoresis, 1997, 18, 432-442
- 30 Ogorzalek Loo R R, Mitchell C et al. Electrophoresis, 1997, 18, 382-390
- 31 Langen H, Gray C et al. Electrophoresis, 1997, 18, 1184-1192
- 32 VanBogelen R, Abshire K. Z et al. Electrophoresis, 1997, 18, 1243-1251
- 33 Hess D, Covey T C et al. Protein Protein Science, 1993, 2, 1342-1351
- 34 Hunt D F, Buko A M et al. Biomed Mass Spectrom 1981, 8, 397
- 35 Hunt D F, Yates J R et al. Proc Natl Acad Sci USA, 1986, 83, 6233
- 36 Biemann K. Methods Enzymol 1990, 193, 455
- 37 Roepstorff P, Fohlmann J. Biomed Mass Spectrom 1984, 11, 601
- 38 McCormack A L, Somogyi A et al. Anal Chem 1993, 65, 2859
- 39 Wilm M, Shevchenko A et al. Nature, 1996, 379, 466
- 40 Link A J, Hays L G et al. Electrophoresis, 1997, 18, 1314-1334
- 41 McCormack A L, Schieltz D M. Anal Chem, 1997, 69, 767

Proteome and Mass Spectrometry

Qian Xiaohong

(National Center of Biomedical Analysis, the Academy of
Military Medical Sciences, Beijing 100850, China)

Received 1998-10-25

Abstract

A proteome is the entire protein complement expressed by a genome, or by a cell or tissue type. The proteomic research is to identify and characterize the proteins in a cell or tissue and define their patterns of expression. Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI/TOF/MS) and electrospray ionization mass spectrometry (ESI/MS) have been proved the powerful techniques for protein and peptide study in the last decade. In this review, the principle of ionization and mass analysis for peptide and protein analysis by MALDI/TOF/MS and ESI/MS is described. The current methods employing MALDI/TOF/MS and tandem mass spectrometry for peptide-mass fingerprinting (PMF) and peptide sequencing are also discussed. A high-throughput screening technique based on MALDI/TOF/MS and ESI/MS/MS for Large-scale proteomics is mentioned.

Key Words: proteome, mass spectrometry