

后基因组时代的蛋白质组学及其在药学研究中的应用

周兴旺*

(厦门大学生命科学学院, 福建 厦门 361005)

关键词: 蛋白质组学; 蛋白质组; 药学; 后基因组

中图分类号: Q78 文献标识码: A 文章编号: 0513 - 4870(2002)10 - 0828 - 05

直接为人类自身健康服务的药学研究一直是现代高新技术应用的前沿领域。近十几年来,随着“组合化学”^[1]和“高通量活性筛选”技术的发展^[2],建立在生化与分子生物学基础上的现代生物医学的进步,以新药开发为中心的药学研究也不断取得新的突破。但与此同时,一些严重威胁人类健康的重大疾病(心脑血管病,糖尿病,癌症等)的致病机理迄今仍未完全阐明,更缺少有效的治疗药物。处于自然界进化树顶端的人体生命奥妙无穷,疑难疾病更往往是多因素相关的和动态变化的,因而异常复杂。而现有的多偏重于局部和静态的研究手段,在面对这些难题时就显得力不从心。近年来,随着一门从整体水平,高通量(High-throughput)地研究功能蛋白质的新技术——“蛋白质组学(Proteomics)”的出现,给药学研究带来了崭新的思路和手段。蛋白质组学与药学的学科交叉也逐渐形成新的研究领域——药物蛋白质组学(Pharmaceutical Proteomics)^[3]。

人类基因组草图 2001 年的正式发表^[4]和预计 2003 年的最终完成,标志着生命科学研究进入后基因组时代。后基因组时代的生命科学重心,已从解析生命的全套遗传信息(遗传密码)转移到系统研究这些遗传密码代表的实际意义(生物学功能)上来。尽管在 mRNA 水平也能一定程度地研究基因表达,如基因表达连续分析法(serial analysis of gene expression, SAGE)^[5]和微阵列芯片(microarray chip)^[6]。但是,近年来越来越多的证据表明,基因表达的中间产物 mRNA 水平的研究并不能取代蛋白质水平(基因最终表达产物)的研究。因为, mRNA

与表达蛋白质的相关性不高(相关系数低于 0.5)^[7,8],并且体内蛋白质还存在多种多样的与其功能密切相关的翻译后加工修饰。因此,要精确地研究基因的功能,还是要回到执行生命功能的蛋白质本身上来。与此前的蛋白质研究技术不同,后基因组时代的蛋白质研究是从整体水平上解析基因组表达的全部蛋白质的结构功能,这就需要有和基因组大规模测序技术类似的高通量研究手段。90 年代中期,包括人类在内的多种生物基因组计划的完成累积了大量基因序列数据,并且这时候新出现的物质谱技术已能对蛋白质进行高效灵敏地分析和测序,再加上计算机数据处理能力的大幅提高和互联网技术的逐渐应用,使得科学家一直期待的这种研究开始成为现实可能,并在此基础上诞生了蛋白质组学。对应基因组学(Genomics), Wilkins 和 Williams^[9]首次将蛋白质组学命名为 Proteomics(此后得到承认和沿用)。

蛋白质组学自 1994 年诞生以来,由于其意义重大,引起了科学界的高度重视:世界权威的生物医学检索工具 PubMed 收录的有关蛋白质组学的科学论文从最早的全年 3 篇(1995 年)快速增长到现在的全年 687 篇(2001 年),Science 和 Nature 杂志也多次对此专栏述评^[10,11]。国际上新成立的人类蛋白质组组织(Human Proteome Organization, HUPO),正酝酿发起与“人类基因组计划”类似的“人类蛋白质组计划”,以系统研究人类基因组表达的所有蛋白质^[12]。

蛋白质组学与健康保健业更是密切相关,美国的制药公司和生物技术公司现在纷纷行动,投入巨资进行这方面的研究。最近,因为独立开展人类基因组测序而闻名的美国 Celera 公司和英国的 Oxford Glycoscience 合作,宣布进行人类疾病相关蛋白的蛋白质组研究,引起科学界的巨大反响^[13]——Celera

收稿日期: 2001-12-28.

* 通讯作者 现址: 615 N. Wolfe St. Rm W4202, Johns Hopkins University, Baltimore, MD 21205.

E-mail: xwzhou_2001@hotmail.edu

拟通过差异比较健康和致病人体的蛋白质组,用蛋白质组技术发现新的诊断蛋白和药靶蛋白。

我国科学界对此前沿新动向也很敏感,这方面已有多篇述评报道^[14-16]。本文述评蛋白质组学及其最新发展和前沿动向。

1 蛋白质组学及其特点

与基本固定不变的基因组不同,蛋白质组是个动态的变化中的概念。因而,对蛋白质组(proteome)也有着不同的定义。一般来说,蛋白质组泛指的是基因组表达的全部蛋白质。蛋白质组更为详细清楚的表述是,在特定时间/空间,特定的细胞/组织/体液,或特定的环境条件下(生理、药理和病理等)基因组表达的全部蛋白质。蛋白质组的研究则称之为“蛋白质组学(proteomics)”^[17]。广义的蛋白质组学指的是一切大规模水平的蛋白质研究(包括用高通量的晶体衍射技术进行大规模的蛋白质结构研究)^[18],但通常提到的蛋白质组学(包括本文在内)是用其狭义的定义,即用蛋白质组研究的三项核心技术(二维凝胶电泳、生物质谱和生物信息学)大规模地研究某一特定的蛋白质群体的组成、结构、活动规律和生物功能。即便是狭义的定义,蛋白质组学涵盖的范畴也很广,不但包括蛋白质的定性鉴定和定量分析,也包括蛋白质的加工修饰(翻译后剪切、磷酸化、糖基化、烷基化、泛蛋白化等)、亚细胞定位、相互作用研究以及最终的功能研究。

目前,蛋白质组学通常的研究方法是:二维凝胶电泳分离蛋白质组分,然后利用生物质谱仪分析从胶上分离的蛋白质斑点:应用激光解析电离质谱(MALDI-MS)得到蛋白质酶解后的肽指纹图谱(Peptide Mass Fingerprint, PMF),或者,用电喷雾串联质谱(ESI-MS/MS)中得到肽片段的进一步裂解谱图。最后,用生物信息学软件(ProteinProspector, SEQUEST等)处理质谱数据——主要通过检索蛋白质或基因数据库鉴定蛋白质。全新蛋白质的鉴定则可通过串联质谱谱图,直接推导出该蛋白质部分肽段的序列,并据此序列设计相应的引物克隆该蛋白。应用上述的蛋白质组技术,仅需 sub-pmol 数量级的样品就可有效鉴定蛋白质。并且,这种高通量的技术,能够在一周内鉴定数百个甚至更多的蛋白质,这是以前的蛋白质化学家不可想象的。此外,利用专门的 2-DE 图像处理软件(如 PDQuest)分析 2-DE 上的蛋白质斑点,还可对鉴定的蛋白质进行定量分析。最近,又出现了同位素亲和标记(isotopic-coded affinity tags, ICAT)结合质谱分析的蛋白质组技术^[19,20],可对蛋白

质组进行精确的定量分析。此外,多维色谱技术、蛋白 N-末端微量测序技术、酵母双杂交技术等生化分析技术在蛋白质组学中也得到广泛应用,并逐渐视为蛋白质组学的组成技术。基因剔除(knock out)、免疫分析等生化与分子生物学技术,也用于蛋白质组的深入的功能研究。利用生物芯片(蛋白质芯片)研究蛋白质组,也是蛋白质组学的新方向之一^[21]。

2 蛋白质组学在药学研究中的应用

2.1 应用蛋白质组学发现新的功能蛋白(药靶蛋白、疾病相关蛋白、药用蛋白等)

药物多是通过特异地作用于体内的靶蛋白质(受体、酶、离子通道等)而重新调整病人的生理状态达到治疗目的。新的药靶蛋白,在药学研究中有重要作用。例如,80年代末开始,研究人员用 HIV 蛋白酶作为爱滋病药物设计的新药靶,成功研究出一系列针对此药靶蛋白的抑制剂药物,显著降低了发达国家爱滋病的死亡率^[22]。目前,制药工业已发现的常用药靶蛋白共约 500 种,而人体中全部的药靶蛋白为一万到两万种^[23]。利用高通量的蛋白质组技术发现新的药靶蛋白,在新药设计和开发中将具有重大作用。

应用蛋白质组技术,通过比较一组正常和患病组织(细胞)的蛋白质组的差异表达图谱,可用来发现疾病相关的标记蛋白质(marker protein),用于临床诊断。如 Jungblut 等^[24]在 2-DE 上比较了来源于正常和结肠癌上皮细胞的蛋白质组的表达,利用 2-DE 图象分析软件分析了 2-DE 上的 800 多个蛋白质斑点后,发现了一种只在肿瘤组织中特异表达的蛋白质(分子量 13 ku,等电点 5.6),经质谱和数据库检索最后将该蛋白质鉴定为 Calgranulin B。目前 Calgranulin B 在肿瘤发生过程中的功能还在进一步研究中,这种结肠癌细胞特异表达的蛋白质的发现可能在结肠癌的早期诊断中有重要作用。

蛋白质芯片技术也用于发现疾病标记蛋白并据此诊断疾病。例如,美国 Stanford 大学 Brown 教授领导的小组致力于将来可用于疾病诊断的蛋白质芯片的研究。最近,他们通过锚定在玻片上的抗体蛋白芯片特异性地捕捉和准确测定溶液中的某些低浓度蛋白质,尽管研究结果还不甚理想(芯片上的数百个点样抗体中只有约 20%有预期效果),但已是这方面的良好开端^[25]。

基因在表达过程一般存在多种剪切修饰和翻译后的修饰,并因此产生新的功能蛋白质(这些加工产生的功能蛋白质大多是现有的基因专利不能涵盖

的)。相对人类基因组的 3~4 万种蛋白编码基因而言,人类蛋白质组的蛋白质总数在 20~200 万^[21],其中可能蕴藏着一些尚未发现的像干扰素那样的药用蛋白质或可供疫苗设计使用的新抗原蛋白)。可以说,人类蛋白质组是一个尚待开发的金矿。此外,利用大通量的蛋白质组技术从人类以外的其他生物(如海洋生物)中开发新的药用蛋白也很有前途。

2.2 药物作用机理研究和新药筛选

目前,许多药用的活性化合物在体内的作用机理并不清楚。应用蛋白质组技术,分析经这些活性化合物处理过的细胞/组织或体液表达的蛋白质组,并比较处理(治疗)前后的蛋白质组的表达差异,鉴定其中发生相应变化的蛋白质,可揭示药物的作用机理。如 Anderson 等^[31]通过分析抑制素类降胆固醇化合物对小鼠肝脏蛋白质组的影响,从药物治疗前后表达变化的蛋白质中鉴定出 HMG-CoA 合成酶(胆固醇合成途径的关键酶之一),从而阐明了该类降胆固醇药物的作用机理。Page 等^[26]分析了抗肝癌化合物 OGT 719 处理过的肿瘤细胞株的蛋白质组表达图谱,同时将其和 5-FU(5-氟尿嘧啶)处理过的肿瘤细胞株的蛋白质组表达图谱比较,发现它们二者 2-DE 的表达图谱发生了类似的变化(暗示二者具有类似的作用机制)。其中一种发生显著变化的蛋白质,经蛋白质组技术鉴定为核糖体 RNA 结合蛋白(嘧啶合成途径的一种重要蛋白质)。这表明,OGT 719 确实与已知的 5-FU 的作用机制类似。

同样,此种方法也可应用于新药筛选。目前,组合化学技术的成熟可以源源不断地大量提供新药开发的新化合物。但是,由于筛选技术的相对滞后,使得先导化合物的发现和优化,仍是制约新药研发的最大瓶颈。而且,现有的筛选手段,还有其他的明显局限:如一次只能使用一种与目标疾病对应的药理筛选指标,必须预先了解药物作用的模式和机制等。而利用整体性的蛋白质组技术进行新药筛选,由于可以同时观察活性化合物对细胞/组织中数千种蛋白质表达的影响,并进一步推断出该化合物作用的目标靶蛋白质,因而,在这方面具有明显的优势。具体地说,可通过分析比较化合物处理(治疗)前后模型细胞或组织的蛋白质组的表达图谱,并和该细胞或组织的数据库中的标准蛋白质组表达图谱对照,快速提取该化合物的有效性和毒性方面有价值信息,并将之用于大量新化合物的筛选。随着蛋白质组技术的不断发展成熟(自动化、高通量、标准蛋白质组图谱库的扩大和完善等),蛋白质组筛选技术将

在今后的新药开发中起着越来越重要的作用。

2.3 毒理学研究

蛋白质组学也可用于药物的毒理学研究。例如,免疫抑制剂环孢菌素 A(Cyclosporine, CsA)是用于器官移植的临床药物。环孢菌素 A 具有肾脏毒性的副作用,但其肾毒性机理一直不甚清楚。Steiner 等^[27]通过环孢菌素 A(Cyclosporine, CsA)对小鼠肾脏组织影响的蛋白质组学研究,首次阐明了环孢菌素 A 肾毒性机理。研究人员发现:在用药(环孢菌素)后的小鼠肾脏组织的蛋白质组中,钙结合蛋白 Calbindin D_{28k}(细胞内的一种钙转运蛋白)明显消失,这解释了用该药后肾小管的钙累积和由此造成的肾小管毒性。此外,药物扰动后的肝脏蛋白质组的表达变化,特别是肝脏药物代谢酶系的表达变化,是药物毒副作用的敏感指标,目前,世界上一些研究小组正致力于这方面的研究,并取得进展^[28]。

3 现状和趋势

蛋白质组学从出现到现在不过 6 年左右,但发展速度很快。目前,著名的“Science”和“Nature”杂志几乎每年都有这方面的专述,“Electrophoresis”杂志辟有 Proteomics 的专栏,2001 和 2002 年国际上还分别创刊了名为“Proteomics”和“Molecular and Cell Proteomics”的新学术期刊。蛋白质组技术也逐渐从定性鉴定和粗略定量发展到现在的精确定量研究^[20]。同时,蛋白质组分析过程,从两相电泳、蛋白质斑点的识别和切割、原位酶解、质谱分析、数据库检索,自动化程度不断提高^[29]。现在,基于两相电泳的蛋白质组技术已能从一块二维胶上鉴定数百个蛋白质,如 Langen 等^[30]从 *Haemophilus influenzae* 中鉴定了 502 个蛋白。Yates 等^[31]则用基于多维色谱的蛋白质组技术从酵母细胞溶解液中鉴定了 1 484 个蛋白。

但是,现有以 2-DE-MS 为基础的蛋白质组技术仍存在一些局限性。

首先,极酸碱性和疏水性蛋白质在 2-DE 胶上分离效果不佳。再者,基于 2-DE 的蛋白质组技术对于低丰度蛋白(密码偏倚值小于 0.1)的鉴定较为困难^[32]。此外,由于大规模的蛋白质组研究的要求的分离能力很高(如人类蛋白质组中蛋白质的总数在 20~200 万之间),目前的 2-DE 技术显然还不能适应这种要求。而且现有的 2-DE-MS 的蛋白质组的技术流程联用困难,自动化程度不高。因此,对于今后蛋白质组研究如何进一步发展,目前科学界还有较大争议。一种意见主张用其他技术完全取代现有的

2-DE MS 为基础的蛋白质组技术。如发展蛋白质组芯片技术^[24], 这样就不用通过 2-DE 的分离步骤, 直接在芯片上检测数百甚至数千个蛋白质。尽管最近的这方面的研究有所进展, 但由于蛋白质结构功能复杂(其功能取决于三维结构)和处理困难(分子量太大, 没有 PCR 那样的有效扩增手段)以及目前尚未有很好的检测方法, 对比基因芯片, 蛋白质芯片的研究难度大了很多, 还有很长的路要走, 总的看是现在还刚刚起步, 短期内似乎不太可能成为蛋白质组研究的主流技术。另一种是主张在 2-DE MS 的基础上进一步改良, 提高自动化程度和检测能力。例如, 通过近年来的努力, 2-DE 的检测能力已有显著提高: 窄 pH 范围 IPG 胶条的使用显著提高了蛋白质的上样量和分离能力(改良后的 2-DE 技术能够较好地分离约 3 000 种的蛋白质), 通过使用特殊的溶剂和缓冲液预先处理样品(或先沉淀该组分蛋白质再重新溶解)改善了疏水性和极酸性蛋白质的分离^[33]。应用基于多维色谱-质谱的蛋白质组技术则大大提高了低丰度蛋白质的鉴定能力^[31](前述的同位素亲和标记(ICAT)的蛋白质组技术理论上也可克服 2-DE 技术的局限, 适合用于低丰度蛋白质的分析)。考虑到实际研究中多通过各种样品预处理(Prefractionate)聚焦于更少杂蛋白组成的蛋白质组分, 而且, 尽管人类蛋白质组总数庞大, 但对于某种特定的细胞来说, 在一定条件下表达的蛋白质组的蛋白质数量一般都在数千种以内。所以, 现有的 2-DE 的分离技术已经基本能满足实验室规模的蛋白质组研究需要。此外, Hochstrasser 等^[33]发明的分子扫描器技术, 通过将 2-DE 胶的蛋白质样品同时平行地原位酶解、电转膜和随后的质谱扫描分析, 大大提高了以 2-DE MS 蛋白质组技术的自动化程度。

作者认为, 将来有望研制出与质谱仪器很好地联用的完全自动化的“两相电泳机器”, Figeys 等研制的 Microfluidic device(微缩流体装置)就是这方面的雏形^[35]。甚至将来还可能融合色谱的分离手段出现高效的“三相分离机器”, 并与质谱分析和生物信息学技术完全联用并实现自动化。另外, 生物质谱仪器的灵敏度和准确度还在继续不断提高, 将使得将来的肽指纹图谱的鉴定方法更加专一有效。成像质谱(Image MS)^[36]的出现和今后的进一步发展, 有可能使得将来不需要分离步骤而可直接从活体细胞组织研究蛋白质组。甚至, 随着组合化学技术、基因组学和蛋白质组学三项技术的进一步发展和相互联系, 将来有可能出现融合这三项技术的机器人用

于大规模的新药筛选和开发。

REFERENCES:

- [1] Hall DG, Manku S, Wang F. Solution- and solid-phase strategies for the design, synthesis, and screening of libraries based on natural product templates: a comprehensive survey [J]. *J Comb Chem*, 2001, **3**(2): 125 - 150.
- [2] Cox B, Denyer JC, Binnie A, et al. Application of high-throughput screening techniques to drug discovery [J]. *Prog Med Chem*, 2000, **37**: 83 - 133.
- [3] Steiner S, Anderson NL. Pharmaceutical proteomics [J]. *Ann NY Acad Sci*, 2000, **919**: 48 - 51.
- [4] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome [J]. *Nature*, 2001, **409**(15): 860 - 921.
- [5] Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression [J]. *Science*, 1995, **270**: 484 - 487.
- [6] Schena M, Shalon D, Davis RW, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray [J]. *Science*, 1995, **270**: 467 - 470.
- [7] Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver [J]. *Electrophoresis*, 1997, **18**: 533.
- [8] Gygi SP, Bochon Y, Franza BR, et al. Correlation between protein and mRNA abundance in yeast [J]. *Mol Cell Biol*, 1999, **19**: 1720 - 1730.
- [9] Wasinger VC, Cordwell SJ, Cerpa Poljak A, et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium* [J]. *Electrophoresis*, 1995, **16**: 1090 - 1094.
- [10] News Briefing. A post-genomic challenge: learning to read patterns of protein synthesis [J]. *Nature*, 1999, **402**: 715 - 720.
- [11] Fields S. Proteomics in Genomeland [J]. *Science*, 2001, **291**: 1221 - 1224.
- [12] Abbott A. Workshop prepares ground for human proteome project [J]. *Nature*, 2001, **413**: 763.
- [13] Service RF. Can Celega do it again? [J] *Science*, 2000, **287**: 2136.
- [14] Wang CC, Tsou CL. Post-genome study: Proteomics [J]. *Acta Biochim Biophys Sin* (生物化学与生物物理学报), 1998, **30**(6): 533 - 539.
- [15] Li BL. Functional proteomics [J]. *Chem Lié* (生命的化学), 1998, **18**(6): 1 - 4.
- [16] Hu ZY, He FC. Progress in proteomics research [J]. *Prog Biochem Biophys* (生物化学与生物物理进展), 1999, **26**(3): 202 - 205.
- [17] Anderson NL, Matheson AD, Steiner S. Proteomics: application in basic and applied biology [J]. *Curr Opin Biotechnol*, 2000, **11**: 408 - 412.
- [18] Christendat D, Yee A, Dharamsi A, et al. Structural proteomics of an archaeon [J]. *Nat Struct Biol*, 2000, **7**(10): 903 - 909.
- [19] Gygi SP, Rist B, Aebersold R. Measuring gene expression by

- quantitative proteome analysis [J]. *Curr Opin Biotech*, 2000, **11**:396 - 401 .
- [20] Yuan Q, Zhao FK. New frontiers in the proteomics research: quantitative proteomics [J]. *Acta Biochim Biophys Sin* (生物化学与生物物理学报), 2001, **33**(5):477 - 482 .
- [21] Newsfocus. Proteomics, high-speed biologists search for gold in proteins [J]. *Science*, 2001, **294**:2074 - 2082 .
- [22] Babine RE, Bender SL. Molecular recognition of protein-ligand complexes: application to drug design [J]. *Chem Rev*, 1997, **97**:1359 .
- [23] Service RF. Structural genomics offers high-speed look at proteins [J]. *Science*, 2000, **287**:1954 - 1956 .
- [24] Jungblut PR, Zimny-Arndt U, Zeindl-Eberhart E, et al. Proteomics in human disease: cancer, heart and infectious diseases [J]. *Electrophoresis*, 1999, **20**:2100 - 2110 .
- [25] Haab BB, Dunham MJ, Brown P. Protein microarray for highly parallel detection and quantitation of specific protein and antibodies in complexes solution [J]. *Genome Biol*, 2001, **2**(2):research/0004 .
- [26] Page MJ, Amess B, Rohlf C. Proteomics: a major new technology for the drug discovery process [J]. *Drug Discov Today*, 1999, **4**(2):55 - 62 .
- [27] Steiner S, Aicher L, Raymackers J, et al. Decrease in kidney calbindin-D 28 kDa as a possible mechanism mediating cyclosporine A [J]. *Biochem Pharmacol*, 1997, **53**:723 - 731 .
- [28] Anderson NL, Esquer R, Richardson F, et al. The effects of peroxisome proliferators on protein abundances in mouse liver [J]. *Toxicol Appl Pharmacol*, 1996, **137**:75 - 89 .
- [29] Quadroni M, James P. Proteomics and automation [J]. *Electrophoresis*, 1999, **20**:664 - 677 .
- [30] Langen H, Takacs B, Evers S, et al. Two dimensional map of the proteome of *Haemophilus influenzae* [J]. *Electrophoresis*, 2000, **21**:411 - 429 .
- [31] Washburn WP, Wolters D, Yates JR. Large scale analysis of the yeast proteome by multidimensional protein identification technology [J]. *Nat Biotechnol*, 2001, **19**:242 - 247 .
- [32] Gygi SP, Corthals GL, Zhang Y, et al. Evaluation of two dimensional electrophoresis gel-based proteome analysis technology [J]. *Proc Natl Acad Sci USA*, 2000, **97**:9390 - 9395 .
- [33] Gorg A, Obermaier C, Boguth G, et al. The current state of 2-DE with immobilized pH gradients [J]. *Electrophoresis*, 2000, **21**:1037 - 1053 .
- [34] Binz P, Muller M, Hochstrasser DF, et al. A molecular scanner to automate proteomic research and to display proteome images [J]. *Anal Chem*, 1999, **71**:4981 - 4988 .
- [35] Figeys D, Pinto D. Proteomics on a chip: promising developments [J]. *Electrophoresis*, 2001, **22**:208 - 216 .
- [36] Stoeckli M, Chaurand P, Hallahan DE, et al. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues [J]. *Nat Med*, 2001, **7**:493 .

PROTEOMICS IN POST-GENOME ERA AND ITS APPLICATION IN PHARMACY

ZHOU Xing-wang

(School of Life Sciences , Xiamen University , Xiamen 361005 , China)

KEY WORDS: proteomics ; proteome ; pharmacy ; post-genome