

蛋白质组学进展

甄 朱*

(山东大学生物化学与分子生物学系, 济南 250100)

摘要 在蛋白质水平上定量、动态、整体性研究生物体的蛋白质组学,将在后基因组时代大大增进我们对基因功能的理解。简要介绍了蛋白质组学的概念、研究手段,及最新进展。

关键词 蛋白质组, 蛋白质组学, 酵母双杂交系统, 质谱, 表面等离子体共振

中图分类号 Q51 **文献标识码** C **文章编号** 1000-3061(2001)05-0491-03

蛋白质组(Proteome)一词最早由澳大利亚学者 Wilkins^[1]等于 1994 年提出,指的是由一个基因组 genOME 或一个细胞、组织表达的所有 PROTEin。从这个定义看,蛋白质组内蛋白质的数目应该等于基因组内编码蛋白的基因的数目(准确的说是 ORF 的数目),但在生物体内这样的蛋白质组是不存在的。从基因表达的角度看,蛋白质组中蛋白质的数目总是少于基因组中开放阅读框(ORF)的数目,但从蛋白质修饰的角度看,蛋白质的数目又远远大于这个数字。基因组基本上是固定不变的,然而蛋白质组是动态的,具有时空性和可调节性,能反映出特定基因的表达时间、表达量,以及蛋白翻译后的加工修饰和亚细胞分布等。

与过去将大量时间花在某个特定的蛋白质上不同,蛋白质组学(Proteomics)是在蛋白质水平上定量、动态、整体性地研究生物体。同基因组学一样,蛋白质组学不是一个封闭的、概念化的、稳定的知识体系,而是一个领域。它旨在阐明生物体全部蛋白质的表达模式及功能模式,其内容包括蛋白质的定性鉴定、定量检测、细胞内定位、相互作用研究等,最终揭示蛋白质功能,是基因组 DNA 序列与基因功能之间的桥梁。

1 蛋白质组学的研究内容及手段

总体上看,蛋白质组研究可分为两个方面。一个是对蛋白质表达模式(或蛋白质组组成)的研究,另一方面是对蛋白质组功能模式(目前主要集中在蛋白质相互作用网络关系)的研究。对蛋白质组组成的分析鉴定是蛋白质组学中的与基因组学相对应的主要内容。它要求对蛋白质组进行表征,即实现所有蛋白质的分离、鉴定及其图谱化。双向凝胶电泳(2-DE)和质谱(Mass spectrometry)技术是当前分离鉴定蛋白质的两大支柱技术。通过分析一个蛋白质是否跟功能已知的蛋白质相互作用可得到揭示其功能的线索。利用大规模酵母双杂交系统,建立相互作用关系的网络图,是目前蛋白质组学领域的研究热点。

1.1 蛋白质的分析鉴定

1.1.1 样品制备及蛋白质分离:目前双向凝胶电泳一般只能分辨到 1000~3000 个蛋白质点(Spot)^[2],而一个细胞系可以表达上万个基因,加上蛋白质加工修饰的多样性,一个样品中的蛋白种类可达 10^6 种^[3]。因此,对于一些低拷贝蛋白,通常先将蛋白粗提液通过亲和柱纯化,再由一维或二维电泳分离。单向电泳的优点在于几乎所有蛋白质均溶于 SDS,分子量在 10 000~300 000 范围内均能有效分离,极酸、极碱蛋白极易检出^[3]。

双向凝胶电泳(2-dimensional gel electrophoresis)原理简明,第一向进行等电聚焦,蛋白质沿 pH 梯度分离,至各自的等电点;随后,在沿垂直的方向进行分子量的分离。80 年代固相化 pH 梯度(Immobilized pH gradients, IPG)的发明和完善,解决了 pH 梯度不稳的问题,使 2-DE 的重复性和上样量大为改善^[4]。

1.1.2 质谱鉴定:质谱(Mass Spectrometry)的基本原理是样品分子离子化后,根据不同离子间的质荷比(m/z)的差异来分离并确定分子量。在离子化方法上,多采用电喷雾离子化(Electrospray ionization, ESI)和基质辅助的激光解吸离子化(Matrix-assisted laser desorption ionization, MALDI)的“软电离”方法。即样品分子电离时,保留整个分子的完整性,不会形成碎片离子。

(1) 肽质图谱(Peptide-mass mapping):

用特定蛋白酶(最常用胰酶)将分离蛋白在胶上或膜上于精氨酸或赖氨酸的 C 末端处断裂,通过 MALDI-MS 或 ESI-MS 来得到精确的肽分子量^[5]。

(2) 肽片段的部分测序:

为进一步鉴定蛋白质,可将液相中的肽段经电喷雾电离后,进入串联质谱(Tandem-MS),肽链中的肽键断裂,形成 N-端碎片离子系列(B 系列)和 C-端碎片离子系列(Y 系列)。综合分析这些碎片离子系列,可得出肽段的氨基酸序列。这样,联合肽片段的分子量和肽段序列信息将足以鉴定一个蛋

白质⁶。当然,尽管肽质指纹谱(Peptide Mass Fingerprint)已成为蛋白质鉴定的支柱技术,N-末端 Edman 降解、氨基酸组分分析仍然是目前鉴定蛋白质的常用手段。

1.1.3 翻译后修饰

磷酸化、糖基化及其他修饰方式是蛋白质执行生理功能所必需的,它们的变化往往与疾病的发生有关。目前,对蛋白质磷酸化的研究已取得初步进展⁷。由于许多受体介导的信号通路都发生酪氨酸磷酸化过程,因此,翻译后修饰的鉴定将对信号通路的阐明有着重大意义⁸。

1.2 蛋白质组差异展示

1.2.1 二维凝胶电泳:通过比较病理-正常状态的 2-DE 图谱,可发现特定的蛋白质表达差异,作为诊断标记或治疗靶目标。

当前 2-DE 比较图谱面临的挑战有:(1)疏水膜蛋白和大分子蛋白不易进入凝胶的第二向。(2)只能显示出表达量最丰富的蛋白。(3)有时一个蛋白点(Spot)实际包含不止一种蛋白。(4)重复性仍然不理想。(5)因蛋白多样性的存在,很难确定一张“正常”状态的图谱作为“病理”状态的对照。

某些情况下,在 mRNA(cDNA)水平上检测基因表达,发现大量的差异表达基因显得更方便快捷有效。但当 mRNA 丰度与蛋白质含量不成正相关或牵涉到蛋白质修饰等问题时,2-DE 无疑是首选方案。

1.2.2 蛋白质芯片:将“诱饵”蛋白(通常是抗体)固定在经特殊处理的表面上,检测样品“猎物”。只有那些与特定抗体特异结合的蛋白留在了芯片上。蛋白质芯片技术实际上是 ELISA 酶联免疫吸附技术的一个大规模应用。它的独特之处在于,选择合适的单克隆抗体,可以将不同翻译后修饰的蛋白区分开来。而且,多样的单克隆抗体可以通过重组其编码 DNA 无限获得。许多人认为找到一种新的大规模制备高质量抗体的方法可能对蛋白质组学产生第二次革命⁹。

1.2.3 质谱定量分析:质谱法与同位素标记方法相结合可以跳过电泳分离蛋白质这一步骤而直接定量分析蛋白质表达水平的差异。将具有不同质量的同位素亲和标签(Isotope-coded affinity tags, ICATs)标记处不同状态下的细胞中的蛋白质,利用串联质谱技术,能非常准确的比较出两份样品蛋白质表达水平的不同¹⁰。这项研究目前还处于初级探索阶段,但无疑具有巨大应用价值。

1.3 蛋白质-蛋白质相互作用

1.3.1 酵母双杂交系统:酵母双杂交系统是研究蛋白质-蛋白质相互作用的有力工具¹¹。它的建立得力于对真核生物调控转录起始过程的认识。细胞起始基因转录需要有反式转录激活因子的参与。DNA 结合结构域(DB)和转录激活结构域(AD)是转录激活因子发挥功能所必需的¹²。将“诱饵”与 Gal4(酵母半乳糖代谢调控系统中的一种反式作用因子)的 DB 结构域融合,“猎物”与 Gal4 的 AD 结构域的酸性区域融合,通过对报道基因表达产物的检测,可判别作为“诱饵”和“猎物”的两个蛋白质之间是否存在相互作用。在研究蛋白质的结构功能特点、作用方式过程中,有时还要通过突变、加抑制剂等手段破坏蛋白质间的相互作用。针对实际工作中的这种需要,人们进一步发展出了逆双杂交系统¹³。

对酵母的蛋白质间的相互作用分析主要有两种策略。一是阵列筛选法(Array screening)。将表达不同“猎物”蛋白的酵母单克隆分别加在微滴定板上,与带有不同的“诱饵”蛋白的酵母株一一接合形成二倍体细胞。“猎物”蛋白与“诱饵”蛋白的相互作用通过报道基因的表达而被鉴定。另一种方法是文库筛选法(Library screening)。该方法与前一种方法的区别是将表达“猎物”蛋白的酵母细胞混在一起构成文库,再将这个文库分别与表达不同“诱饵”蛋白的酵母细胞接合,再进一步筛选阳性克隆,即“诱饵”与“猎物”发生相互作用的克隆。相比之下,阵列筛选法更为有效,而文库筛选法的长处是通量大^{14,15}。

1.3.2 噬菌体展示:将编码噬菌体的外壳蛋白基因上连上一个单克隆抗体的基因序列。当噬菌体生长时,表面就会表达出相应的单抗。将噬菌体过柱,由于柱上含目的蛋白,会特异性结合相应抗体¹⁶。同酵母双杂交系统相比,噬菌体展示技术同样具有简便、高通量的优点,不同的是,反应在溶液中进行,而不是在酵母细胞核中。另外,由于有些蛋白质本身有激活转录活性,不经过双杂交相互作用就能激活报道基因的表达,因此酵母双杂交系统不适用。而噬菌体展示能很好的解决这个问题。

1.3.3 表面等离子体共振技术:目前,表面等离子体共振(Surface Plasmon Resonance)技术已成为当今研究蛋白质之间相互作用的一种新的手段。典型的代表是瑞典的 BIACORE 的单元蛋白质芯片。

表面等离子体共振 SPR 技术通常是将“诱饵”蛋白作为配基,固化在几十纳米厚的金属(金、银等)膜表面,加入含“猎物”蛋白的溶液。若“诱饵”蛋白与“猎物”蛋白之间存在相互作用,则会特异性结合形成蛋白质复合物。两者的结合将使金属膜与溶液界面的折射率上升,从而导致共振角度改变。由此,可以检测出蛋白-蛋白之间的相互作用¹⁷。

SPR 技术的特点是不需标记物或染料,测定快速、安全。除应用于蛋白-蛋白外,还可灵敏检测蛋白-核酸及其它生物大分子之间的相互作用。SPR DNA 生物传感器可用于基因突变的检测,PCR 产物的测定以及病毒和其他微生物的检测等方面的研究。

2 蛋白质组学的研究意义及应用前景

随着大量 DNA 序列数据的不断涌现,人们意识到仅仅靠基因组的序列来试图阐明生命现象是远远不够的。一个细胞通常依赖许多代谢调节途径,而其执行者是蛋白质。基因与蛋白质之间并不存在严格的一一对应的线形关系。基因组中开放阅读框的存在并不意味着功能基因的存在。生物信息学准确预测基因及推测蛋白质结构的成功率仍然不高。另外,mRNA 水平与蛋白质含量不一定呈正相关,蛋白质的修饰加工、细胞定位,都只能在蛋白质水平上研究。因此,注解基因组的最好方法就是从蛋白质水平出发。

蛋白质组学研究的一大特色是,从一开始就呈现出基础研究与实际应用并驾齐驱的趋势。许多实验室、公司和药厂很早就开始进行与应用有关的蛋白质组研究。目前,蛋白质组技术已用于肝癌、膀胱癌、前列腺癌等研究中。美国国立

肿瘤研究所(National Cancer Institute)最近还建立了人类肿瘤基因索引(Human Tumor Gene Index)。

另外,生物信息学的迅猛发展使得依据基因组序列分析预测病原微生物中的疫苗候选基因成为可能。如今,许多制药公司利用蛋白质组技术快速鉴定这些基因表达产物,来辅助基因组学筛选疫苗^[18]。

蛋白质芯片技术提高了蛋白质鉴定的速度与重复性。目前已用于某些疾病的蛋白质分子标记的检测,如阿尔茨海默氏病(进行性老年性痴呆)和子宫癌^[19]。

我国关于蛋白质组研究的国家自然科学基金重大项目从1999年起开始启动。随着对蛋白质组学研究的不断深入,将为我国重大疾病机理阐明、基因功能研究、分子诊断和新药开发提供重要的理论基础,把学科建设的基础性研究与满足国家需求的应用性研究结合在一起。

3 结束语

蛋白质组学为直接在蛋白质水平上大规模研究基因功能提供了有力工具。利用质谱技术研究凝胶分离的蛋白质引发了蛋白质功能研究的一场革命。蛋白质鉴定将在高通量、高灵敏度、完整性等方面进一步完善,分析手段将向自动化、微量化、平行化方向发展。

21世纪将是一个整体细胞生物学(Holistic cellular biology)的时代,DNA和RNA的信息加上相应的蛋白质水平上的信息的补充和提高,才是完全的细胞分子生物学的研究。

REFERENCES(参考文献)

- [1] Wilkins M R *et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Bio Technology*, 1996, **14**:61 ~ 65
- [2] YU L R(俞利荣), ZENG R(曾嵘), XIA Q C(夏其昌). Techniques for study of proteome and their progresses. *Chemistry of Life* (生命的化学), 1998, **18**(6):4 ~ 9
- [3] Pandey A, Mann M. Proteomics to study genes and genomes. *Nature*, 2000, **405**:837 ~ 846
- [4] WAN J H(万晶宏), HE F C(贺福初). Progresses in proteomics

- technology. *Chinese Science Bulletin* (科学通报), 1999, **44**(9):904 ~ 910
- [5] Henzel W J *et al.* Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA*, 1993, **90**:5011 ~ 5015
- [6] Shevchenko A *et al.* MALDI quadruple time-of-flight mass spectrometry: powerful tool for proteomic research. *Anal Chem*, 2000, **72**:2132 ~ 2141
- [7] Neubauer G & Mann M. Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations. *Anal Chem*, 1999, **71**:235 ~ 242
- [8] Pandey A *et al.* Analysis of receptor signaling pathways by mass spectrometry: identification of Vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci USA*, 2000, **97**:179 ~ 184
- [9] Lueking A *et al.* Protein microarrays for gene expression and antibody screening. *Anal Biochem*, 1999, **270**:103 ~ 111
- [10] Gygi S P *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol*, 1999, **17**:994 ~ 999
- [11] Fields S, Song O K. A novel genetic system to detect protein-protein interactions. *Nature*, 1989, **340**:245 ~ 246
- [12] Ma J, Ptashne M. A new class of transcriptional activators. *Cell*, 1987, **51**:113 ~ 119
- [13] Vidal M, Endoh H. Prospects for drug screening using the reverse two-hybrid system. *Trends Biotechnol*, 1999, **17**:374 ~ 381
- [14] Uetz P *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 2000, **403**:623 ~ 627
- [15] Ito T *et al.* Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA*, 2000, **97**:1143 ~ 1147
- [16] Zozulya S *et al.* Mapping signal transduction pathways by phage display. *Nature Biotechnol*, 1999, **17**:1193 ~ 1198
- [17] GAO Z X(高志贤), CHAO F H(晁福寰). Advance on DNA biosensor. *Letters In Biotechnology*(生物技术通讯), 2000, **11**(1):45 ~ 53
- [18] Chakravarti D N *et al.* Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine*, 2000, **19**(6):601 ~ 612
- [19] Jain K K. Applications of proteomics in oncology. *Pharmacogenomics*, 2000, **1**(4):385 ~ 393

Progress in Proteomics

ZHEN Zhu*

(Department of Biochemistry & Molecular Biology, Shandong University, Jinan 250100, China)

Abstract Proteomics, the large-scale analysis of proteins, will contribute greatly to our understanding of gene function in the post-genome era. Proteomics uses a combination of sophisticated techniques including two-dimensional(2D) gel electrophoresis, mass spectrometry, yeast two-hybrid system, bioinformatics, etc. to characterize, to quantify, and to analyze cellular proteins in a global way. The application of proteomics provides major opportunities to elucidate disease mechanisms and to identify new therapeutic targets.

This review aims to explain the definition of proteomics and then to outline proteomic techniques. Finally, applications to the study of human disease are briefly discussed.

Key words proteome, proteomics, yeast two-hybrid system, mass spectrometry, surface plasmon resonance

Received: February 12, 2001

* Corresponding author. E-mail: pearlzz2000@163.net