

学科发展

蛋白质组学研究进展与趋势*

曾 嵘 夏其昌

(上海生命科学研究院生物化学与细胞生物学研究所 上海 200031)

摘要 介绍了蛋白质组学的意义、策略与范围,评述了蛋白质组学研究技术与发展趋势,并根据我国该领域的研究现状对其进一步发展提出了呼吁。

关键词 蛋白质组学研究,进展,趋势



如果在五年前提到蛋白质组学(Proteomics),恐怕知之者甚少,而在略知一二者中,部分人还抱有怀疑态度。但是,2001年的 *Science* 已把蛋白质组学列为六大研究热点之一,其“热度”仅次于干细胞研究,名列第二。蛋白质组学的受关注程度

如今已令人刮目相看。

1 蛋白质组学的意义和背景

随着人类基因组计划的实施和推进,生命科学已进入了后基因组时代。在这个时代,生命科学的主要研究对象是功能基因组学,包括结构基因组研究和蛋白质组研究等。尽管现在已有多个物种的基因组被测序,但在这些基因组中通常有一半以上基因的功能是未知的。目前功能基因组中所采用的策略,如基因芯片、基因表达序列分析(Serial analysis of gene expression, SAGE)等都是从细胞中 mRNA 的角度来考虑的,其前提是细胞中 mRNA 的水平反映了蛋白质表达的水平。但事实并不完全如此,从 DNA $\rightarrow$ mRNA $\rightarrow$ 蛋白质,存在三个层次的调

控,即转录水平调控(Transcriptional control)、翻译水平调控(Translational control)、翻译后水平调控(Post-translational control)。从 mRNA 角度考虑,实际上仅包括了转录水平调控,并不能全面代表蛋白质表达水平。实验也证明,组织中 mRNA 丰度与蛋白质丰度的相关性并不好,尤其对于低丰度蛋白质来说,相关性更差。更重要的是,蛋白质复杂的翻译后修饰、蛋白质的亚细胞定位或迁移、蛋白质-蛋白质相互作用等则几乎无法从 mRNA 水平来判断。毋庸置疑,蛋白质是生理功能的执行者,是生命现象的直接体现者,对蛋白质结构和功能的研究将直接阐明生命在生理或病理条件下的变化机制。蛋白质本身的存在形式和活动规律,如翻译后修饰、蛋白质间相互作用以及蛋白质构象等问题,仍依赖于直接对蛋白质的研究来解决。虽然蛋白质的可变性和多样性等特殊性质导致了蛋白质研究技术远比核酸技术要复杂和困难得多,但正是这些特性参与和影响着整个生命过程。

传统的对单个蛋白质进行研究的方式已无法满足后基因组时代的要求。这是因为:(1) 生命现象的发生往往是受多因素影响的,必然涉及多个蛋白质。(2) 多个蛋白质的参与是交织成网络的,或平行发生,或呈级联因果。(3) 在执行生理功能时

* 收稿日期:2002 年 4 月 25 日

蛋白质的表现是多样的、动态的,并不像基因组那样基本不变。要对生命的复杂活动有全面和深入的认识,必然要在整体、动态、网络的水平上对蛋白质进行研究。因此,在上世纪 90 年代中期,国际上产生了一门新兴学科——蛋白质组学(Proteomics),它是以细胞内全部蛋白质的存在及其活动方式为研究对象。可以说,蛋白质组研究的开展不仅是生命科学研究进入后基因组时代的里程碑,也是后基因组时代生命科学研究的核心内容之一。

虽然第一次提出蛋白质组概念是在 1994 年,但相关研究可以追溯到上世纪 80 年代中期甚至更早,尤其是 80 年代初,在基因组计划提出之前,就有人提出过类似的蛋白质组计划,当时称为 Human Protein Index 计划,旨在分析细胞内的所有蛋白质。但由于种种原因,这一计划被搁浅^[1]。90 年代初期,各种技术已比较成熟,在这样的背景下,经过各国科学家的讨论,才提出蛋白质组这一概念。

国际上蛋白质组研究进展十分迅速,不论基础理论还是技术方法,都在不断进步和完善。相当多种细胞的蛋白质组数据库已经建立,相应的国际互联网站也层出不穷。1996 年,澳大利亚建立了世界上第一个蛋白质组研究中心: Australia Proteome Analysis Facility (APAF)。丹麦、加拿大、日本也先后成立了蛋白质组研究中心。在美国,各大药厂和公司在巨大财力的支持下,也纷纷加入蛋白质组的研究阵容。去年在瑞士成立的 GeneProt 公司,是由以蛋白质组数据库“SWISSPROT”著称的蛋白质组研究人员成立的,以应用蛋白质组技术开发新药物靶标为目的,建立了配备有上百台质谱仪的高通量技术平台。而当年提出 Human Protein Index 的美国科学家 Anderson 也成立了类似的蛋白质组学公司,继续其多年未实现的梦想。2001 年 4 月,在美国成立了国际人类蛋白质组研究组织(Human Proteome Organization, HUPO),随后欧洲、亚太地区都成立了区域性蛋白质组研究组织,试图通过合作的方式,融合各方面的力量,完成人类蛋白质组计划(Human Proteome Project)^[2]。

2 蛋白质组研究的策略和范围

蛋白质组学一经出现,就有两种研究策略。一种可称为“竭泽法”,即采用高通量的蛋白质组研究

技术分析生物体内尽可能多乃至接近所有的蛋白质,这种观点从大规模、系统性的角度来看待蛋白质组学,也更符合蛋白质组学的本质。但是,由于蛋白质表达随空间和时间不断变化,要分析生物体内所有的蛋白质是一个难以实现的目标。另一种策略可称为“功能法”,即研究不同时期细胞蛋白质组成的变化,如蛋白质在不同环境下的差异表达,以发现有差异的蛋白质种类为主要目标。这种观点更倾向于把蛋白质组学作为研究生命现象的手段和方法。

早期蛋白质组学的研究范围主要是指蛋白质的表达模式(Expression profile),随着学科的发展,蛋白质组学的研究范围也在不断完善和扩充。蛋白质翻译后修饰研究已成为蛋白质组研究中的重要部分和巨大挑战。蛋白质-蛋白质相互作用的研究也已被纳入蛋白质组学的研究范畴。而蛋白质高级结构的解析即传统的结构生物学,虽也有人试图将其纳入蛋白质组学研究范围,但目前仍独树一帜。

3 蛋白质组学研究技术

可以说,蛋白质组学的发展既是技术所推动的,也受到技术的限制。蛋白质组学研究成功与否,很大程度上取决于其技术方法水平的高低。蛋白质研究技术远比基因技术复杂和困难。不仅氨基酸残基种类远多于核苷酸残基(20/4),而且蛋白质有着复杂的翻译后修饰,如磷酸化和糖基化等,给分离和分析蛋白质带来很多困难。此外,通过表达载体进行蛋白质的体外扩增和纯化也并非易事,从而难以制备大量的蛋白质。蛋白质组学的兴起对技术有了新的需求和挑战。蛋白质组的研究实质上是在细胞水平上对蛋白质进行大规模的平行分离和分析,往往要同时处理成千上万种蛋白质。因此,发展高通量、高灵敏度、高准确性的研究技术平台是现在乃至相当一段时间内蛋白质组学研究的主要任务。当前国际蛋白质组研究技术平台的技术基础和发展趋势有以下几个方面:

3.1 蛋白质组研究中的样品制备

通常可采用细胞或组织中的全蛋白质组分进行蛋白质组分析。也可以进行样品预分级,即采用各种方法将细胞或组织中的全体蛋白质分成几部

分,分别进行蛋白质组研究。样品预分级的主要方法包括根据蛋白质溶解性和蛋白质在细胞中不同的细胞器定位进行分级,如专门分离出细胞核、线粒体或高尔基体等细胞器的蛋白质成分。样品预分级不仅可以提高低丰度蛋白质的上样量和检测,还可以针对某一细胞器的蛋白质组进行研究。

对临床组织样本进行研究,寻找疾病标记,是蛋白质组研究的重要方向之一。但临床样本都是各种细胞或组织混杂,而且状态不一。如肿瘤组织中,发生癌变的往往是上皮类细胞,而这类细胞在肿瘤中总是与血管、基质细胞等混杂。所以,常规采用的癌和癌旁组织或肿瘤与正常组织进行差异比较,实际上是多种细胞甚至组织蛋白质组混合物的比较。而蛋白质组研究需要的通常是单一的细胞类型。最近在组织水平上的蛋白质组样品制备方面也有新的进展,如采用激光捕获微解剖(Laser Capture Microdissection, LCM)方法分离癌变上皮类细胞^[3]。

3.2 蛋白质组研究中的样品分离和分析

利用蛋白质的等电点和分子量通过双向凝胶电泳的方法将各种蛋白质区分开来是一种很有效的手段。它在蛋白质组分离技术中起到了关键作用。如何提高双向凝胶电泳的分离容量、灵敏度和分辨率以及对蛋白质差异表达的准确检测,是目前双向凝胶电泳技术发展的关键问题。国外的主要趋势有第一维电泳采用窄 pH 梯度胶分离以及开发与双向凝胶电泳相结合的高灵敏度蛋白质染色技术,如新型的荧光染色技术。

质谱技术是目前蛋白质组研究中发展最快,也最具活力和潜力的技术。它通过测定蛋白质的质量来判别蛋白质的种类。当前蛋白质组研究的核心技术就是双向凝胶电泳-质谱技术,即通过双向凝胶电泳将蛋白质分离,然后利用质谱对蛋白质逐一进行鉴定。对于蛋白质鉴定而言,高通量、高灵敏度和高精度是三个关键指标。一般的质谱技术难以将三者合一,而最近发展的质谱技术可以同时达到以上三个要求,从而实现对蛋白质准确和大规模的鉴定。

3.3 蛋白质组研究的新技术

做过双向凝胶电泳的人一定会抱怨它的繁琐、

不稳定和低灵敏度等缺点。发展可替代或补充双向凝胶电泳的新方法已成为蛋白质组研究技术最主要的目标。目前,二维色谱(2D-LC)、二维毛细管电泳(2D-CE)、液相色谱-毛细管电泳(LC-CE)等新型分离技术都有补充和取代双向凝胶电泳之势。另一种策略则是以质谱技术为核心,开发质谱鸟枪法(Shotgun)、毛细管电泳-质谱联用(CE-MS)等新策略,直接鉴定全蛋白质组混合酶解产物。随着对大规模蛋白质相互作用研究的重视,发展高通量和高精度的蛋白质相互作用检测技术也被科学家所关注。此外,蛋白质芯片的发展也十分迅速,并已在临床诊断中得到应用。

3.4 蛋白质组生物信息学

蛋白质组数据库是蛋白质组研究水平的标志和基础。瑞士的 SWISS-PROT 拥有目前世界上最大、种类最多的蛋白质组数据库。丹麦、英国、美国等也都建立了各具特色的蛋白质组数据库。生物信息学的发展已给蛋白质组研究提供了更方便、有效的计算机分析软件。特别值得注意的是,蛋白质质谱鉴定软件和算法发展迅速,如 SWISS-PROT、Rockefeller 大学、UCSF 等都有自主的搜索软件和数据管理系统。最近发展的质谱数据直接搜寻基因组数据库,使得质谱数据可直接进行基因注释,判断复杂的拼接方式。随着基因组学的迅速推进,会给蛋白质组研究提供更多更全的数据库。另外,肽序列标记的从头测序软件也十分引人注目。

4 蛋白质组学发展趋势

在基础研究方面,近两年来蛋白质组研究技术已被应用到各种生命科学领域,如细胞生物学、神经生物学等。在研究对象上,覆盖了原核微生物、真核微生物、植物和动物等范围,涉及到各种重要的生物学现象,如信号转导、细胞分化、蛋白质折叠等等。在未来的发展中,蛋白质组学的研究领域将更加广泛。

在应用研究方面,蛋白质组学将成为寻找疾病分子标记和药物靶标最有效的方法之一。在对癌症、老年性痴呆等人类重大疾病的临床诊断和治疗方面,蛋白质组技术也有十分诱人的前景。目前国际上许多大型医药公司正投入大量的人力和物力进行蛋白质组学方面的应用性研究。

在技术发展方面,蛋白质组学的研究方法将出现多种技术并存,各有优势和局限的特点,而难以像基因组研究那样形成比较一致的方法。除了发展新方法外,更强调各种方法间的整合和互补,以适应不同蛋白质的不同特征。另外,蛋白质组学与其它学科的交叉也将日益显著和重要。这种交叉是新技术、新方法的活水之源,特别是蛋白质组学与其它大规模科学如基因组学、生物信息学等领域的交叉,所呈现出的系统生物学(System Biology)研究模式,将成为未来生命科学最令人激动的新前沿。

5 中国蛋白质组学的现状与趋势

国内从 1997 年开始开展蛋白质组研究。其中,中国科学院上海生命科学研究院生物化学与细胞生物学研究所建立的蛋白质组研究中心在 2000 年的 *Electrophoresis* 上发表了我国第一篇蛋白质组研究论文;并在生命科学研究院生物信息中心的协助下建立了我国第一个大型的蛋白质组数据库并实现了网络共享(www.sibs-proteome.org)。军事医学科学院和湖南师范大学也于 2001 年相继发表了蛋白质组研究论文。

国家自然科学基金委员会于 1999 年支持了蛋白质组重大项目,经过 4 年的发展,在国内已形成一支蛋白质组研究队伍并拥有主要的研究技术和

装备。目前已建立起基本蛋白质组平台技术的有中国科学院上海生命科学研究院蛋白质组研究中心和军事医学科学院。已开展相关研究的单位还有湖南师范大学生命科学院、中国科学院大连化学物理研究所、复旦大学等。

2001 年我国的蛋白质组研究又进入了一个迅速发展的新阶段,其主要标志是国家科技部将蛋白质组研究确立为“973”和“863”的项目。今后,只要我国能集中力量、加大投入,建立高效、高通量的蛋白质组研究技术平台,对与重大疾病相关的蛋白质组进行重点突破,就一定能推动我国的蛋白质组研究进程,在新一轮蛋白质组学技术平台的建立和蛋白质组研究方面做出中国人的贡献,并在国际的蛋白质组研究领域占有一席之地。

主要参考文献

- 1 Mary F, Lope Z. Better approaches to finding the needle in a haystack: Optimizing proteome analysis through automation. *Electrophoresis*, 2000, 21: 1 082- 1 093.
- 2 <http://www.hupo.org>
- 3 Emmer-Buck M R, Gillespie J W, Paweletz C P, Ornstein D K, Basrur V, Appella E, Wang Q H, Huang J, Hu N, Taylor P, Petricoin E F 3rd., An approach to proteomic analysis of human tumors. *Mol. Carcinog*, 2000, 27(3): 158- 165.

The Progress and Development of Proteomic Research

Zeng Rong Xia Qichang

(Key Lab for Proteomics, Institutes of Biochemistry and Cell Biology,
Shanghai Institutes for Biological Sciences, CAS, 200031 Shanghai)

In a post-genome era, researchers in life sciences will pay more attention to the functional genomics. One of the most important areas in functional genomics is proteomics, which analyzes all the proteins in a genome. Two major strategies have been used for proteomic research. One is to study all the expressed proteins at the same time, the other is to compare proteomes under different conditions in order to identify the differential proteins. The main techniques used in proteomics are two-dimensional electrophoresis and mass-spectrometry.

曾 嵘 中国科学院上海生命科学研究院生物化学与细胞生物学研究所副研究员,蛋白质组学重点实验室副主任,蛋白质组研究分析中心副主任。1995 年毕业于湖南师范大学生物系。2000 年在中国科学院上海生物化学研究所取得博士学位。1997 年以来,发表有关蛋白质组学论文 10 余篇。