

21 世纪的生物学

蛋白质组学的进展*

李 林

(中国科学院上海生命研究院, 上海生物化学研究所, 分子生物学国家重点实验室, 上海 200031)

摘要 蛋白质组学是在细胞的整体蛋白质水平上进行研究、从蛋白质整体活动的角度来认识生命活动规律的一门新学科。简要介绍蛋白质组学的科学背景及其最新发展。

关键词 蛋白质组, 蛋白质组学

学科分类号 Q7

当今生物学研究已经进入了这样一个时代, 在生物大分子的整体水平上将不同的研究技术与手段有机结合以攻克生物学的难题。这一点突出体现在基因组学 (genomics) 以及随之产生的蛋白质组学 (proteomics) 等新学科领域的研究中。

1 蛋白质组学产生的科学背景

众所周知, 始于 20 世纪 90 年代初的庞大的人类基因组计划业已取得了巨大的成就, 几个物种 (包括人类) 的基因组序列已经或即将完成。生命科学已实质性地跨入了后基因组时代, 研究重心已开始从揭示生命的所有遗传信息转移到在分子整体水平对功能的研究上。这种转向的第一个标志是产生了功能基因组学 (functional genomics) 这一新学科, 即从基因组整体水平上对基因的活动规律进行阐述。如在 mRNA 水平上通过 DNA 芯片技术检测大量基因的表达模式。而第二个标志则是蛋白质组学的兴起。

蛋白质组 (proteome) 一词是澳大利亚 Macquarie 大学的 Wilkins 和 Williams 在 1994 年首次提出, 最早见诸于文献是在 1995 年 7 月的《Electrophoresis》杂志上^[1~3]。它是指基因组表达的全部蛋白质及其存在方式。蛋白质组学旨在阐明生物体全部蛋白质的表达模式及功能模式, 其内容包括鉴定蛋白质的表达、存在方式 (修饰形式)、结构、功能和相互作用等。国内已有多篇综述文章介绍了蛋白质组学的产生背景与科学意义^[1~3]。从蛋白质组的定义上就可以清楚看出, 蛋白质组学不同于传统的蛋白质学科之处在于它的研究是在生物体或其细胞的整体蛋白质水平上进行的, 它从一个机体或一个细胞的蛋白质整体活动的角度来揭示

和阐明生命活动的基本规律。

2 蛋白质组研究的重要性与艰难性

时至今日, 不少人对蛋白质组学仍然存在两个基本疑问, 一是针对其意义, 为什么有了基因组研究特别是功能基因组研究还要有蛋白质组研究? 二是针对其可行性, 现有的技术手段能否实现蛋白质组学的目标?

第一个疑问里面实际上还存在两个问题, 一个是对基因组学 (或者功能基因组学) 的内容如何界定的问题, 也许不少人认为基因组学 (或者功能基因组学) 已包括了蛋白质组学的内容, 它们之间的关系就如同文具和铅笔刀的关系。如果这样认为, 这个疑问当然就不需要解答了, 这个问题也就变得没有意义了。解答上述疑问当然是针对第二个问题, 在这里基因组学 (或者功能基因组学) 与蛋白质组学被看成是“姐妹”关系, 只是不少人不理解为什么当我们已经能够在 mRNA (cDNA) 水平上简便、快速和大通量地检测基因的表达, 发现基因表达的规律以及据此发现大量的差异表达基因等, 还需要进行蛋白质组研究? 而且, 对蛋白质表达的检测要费时费力得多。

要解答这个疑问也许不是很困难。首先, 我们知道, 生物功能的主要体现者或执行者是蛋白质, 蛋白质的表达水平、存在方式及相互作用等直接与生物功能相关。其次, DNA 芯片技术虽然能给出生物体所启动基因的相关信息, 但它并不能反映蛋白质合成的情况, 即 mRNA 的水平 (包括 mRNA

* 国家自然科学基金重大项目资助 (39990600)。

Tel: (021) 64374430, E-mail: lil@sunm.shnc.ac.cn

收稿日期: 2000-04-18

的种类和含量) 并不能反映蛋白质的表达水平, 它们之间的相关性很差. 再一点, 蛋白质具有自身特有的活动规律, 如蛋白质的修饰加工、转运定位、相互作用等, 这一切都与生物学功能密切联系, 对其了解恐怕必须要在蛋白质水平上进行研究. 再加上前面谈到的蛋白质组学与传统蛋白质研究的区别, 不难理解蛋白质组研究的重要性和必要性.

但是, 也应该看到蛋白质组研究的艰难性. 首先, 蛋白质组具有多样性和可变性. 对于一个机体而言, 基因的数量是稳定不变的, 而蛋白质的种类和数量在同一机体的不同细胞中是各不相同的, 即使同一种细胞在不同时期、不同条件下, 其蛋白质组也是在不断的改变之中. 而且在病理或治疗过程中, 细胞蛋白质的组成及其变化, 与正常生理过程的也不同. 其次, 蛋白质组具有复杂性. 由于蛋白质存在翻译后修饰, 所以一个 mRNA 往往对应多个蛋白质. 也就是说, 蛋白质的数量远远多于基因的数量, 对于人类, 这个数目起码是三倍. 再一点, 由于蛋白质无法象 DNA 一样被“扩增”, 蛋白质组中含量低的蛋白质在大规模的检测中很难被检测到, 而有些对细胞功能产生重大影响的蛋白质其表达量往往是比较低的. 因此可以看出, 面对生物体在其生命的不同时期合成的比基因组更庞大的整套蛋白质, 实现鉴定其这一目标显得更庞大、更复杂和更富有挑战性.

总体上看, 蛋白质组研究可分为两个方面, 一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, 即所有蛋白质的分离与鉴定及其图谱化. 具体有两个基本步骤: 第一步是从样品中分离蛋白质, 第二步是鉴定被分离的蛋白质. 双向凝胶电泳 (2DGE) 和质谱 (MS) 技术是当前上述两个步骤中的二大支撑技术. 应该看到, 目前这些技术手段还有相当的局限性, 如蛋白质分离的分辨率、检出的灵敏度以及鉴定的自动化等还有待提高和解决, 其难度是很大的. 因此, 上面提到的关于蛋白质组学的第二个疑问在科学家中是普遍存在的, 实际上也是困扰直接从事蛋白质组学工作的研究人员的一个问题. 不过, 回想起当初人类基因组计划刚提出的情形, 科学家们也是分为满怀激情、怀疑和冷漠三种态度. 其关注的核心问题恐怕也是对技术手段与目标关系

的考虑. 许多人甚至因为人类基因组测序工作的庞大而感到徘徊和沮丧, 投资问题也是一再探讨. 然而通过实践表明所忧虑的一切都是可以解决的. 有人认为蛋白质组学的技术困难不会持续太久, 只要对该项目进行投资, 一些技术问题很快就能解决^[4].

近来, 蛋白质组学有关的技术和生物信息学在不断的并迅速的开发和发展中^[5]. 作为蛋白质组学当前的关键技术之一的双向电泳技术的灵敏度和分辨率不断得到提高, 最高可达到 11 000 个蛋白质点的分辨率. 与此同时, 一种基于芯片技术的蛋白质检测的更先进的方法——抗体与蛋白质阵列技术 (antibody and protein array technologies) 也在迅速发展, 可望能够快速并且平行定量分析蛋白质的分布^[6]. 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. 抗体可以象 cDNA 微阵列一样定位于芯片上, 蛋白质组的鉴定将更加简便、高通量和高自动化. 此外, 大规模酵母双杂交技术已发展成为蛋白质组功能模式研究的主要手段之一^[7], 诸如质谱法与同位素标记方法相结合定量分析蛋白质表达水平的差异^[8]、蛋白质组中膜蛋白的分析鉴定^[9]等一系列蛋白质组学的新方法也在不断地涌现出来.

无论我们对蛋白质组学采取什么态度, 它已经并还正在迅速地发展中, 国际上对它的重视和投入与日俱增, 这一点在下面还会谈到. 同时, 蛋白质组学以其特有的思想方法和技术手段在解决生物学重大问题上已开始显示出其强大的威力.

3 蛋白质组学的研究进展

蛋白质组学强调的是针对蛋白质的一个整体思路. 从整体的角度来看, 蛋白质组研究大致可以分为两种类型. 一种是针对细胞或组织的全部蛋白质, 也就是说着眼点就是整个蛋白质组; 而另一种则是以与一个特定的生物学问题或机制相关的全部蛋白质为着眼点, 在这里整体是局部性的. 针对细胞蛋白质组的完整分析的工作已经比较全面地展开, 不仅如大肠杆菌、酵母等低等模式生物的蛋白质组数据库在建立之中, 高等生物如水稻和小鼠等的蛋白质组研究也已开展, 人类一些正常和病变细胞的蛋白质组数据库也在建立之中. 与此同时, 更多的蛋白质组研究工作则是将着眼点放在蛋白质组的变化或差异上, 也就是通过对蛋白质组的比较分析, 首先发现并去鉴定在不同生理条件下或不同外

界条件下蛋白质组中有差异的蛋白质组分。限于篇幅, 本文不对这方面的工作做进一步论述。

本文接下来重点介绍近期发表的关于蛋白质组学的几个工作, 从中可以看到蛋白质组学的思想方法在蛋白质整体(或局部整体)水平上是如何解决生物学的一些重要问题的。

1999年11月《Nature》杂志发表了一篇用蛋白质组学方法研究蛋白质折叠的研究论文^[10]。在这篇文章中, Houry 等报道了在大肠杆菌胞质中的2 500种新生多肽链中只有近300种以GroEL作为分子伴侣来帮助其折叠成正确构象。在以往的相关研究中, 通常只是针对某个或某些特定的蛋白质, 观察它(们)在折叠过程中是否需要诸如GroEL等分子伴侣的帮助。而在这个工作中, 研究是从一个整体的思路出发, 首先通过免疫共沉淀的方法获得所有与GroEL结合的肽链, 再通过二维电泳和数据库比较等蛋白质组研究的手段对这些肽链进行分析鉴定, 从而实现了大肠杆菌近2 500条新生多肽链与分子伴侣GroEL的关系的全面分析。在这个工作中, 研究者还通过对其中50种与GroEL作用的肽链的鉴定, 进一步地揭示了决定这些蛋白质能与GroEL相互作用的关键结构特征。应该说, 这个工作很好地体现了蛋白质组学的思想方法和技术手段的应用。

过去在细胞生物学领域还没有得到过一个主要亚细胞结构的完整的分子图。核孔复合体是一个巨大的跨核膜的八角形结构, 是控制大分子在胞质和核质间运输的通道。多年来, 很多方法被用来分析这一复合物的组成成分。虽然这些工作取得了很大的进展, 但究竟在多大程度上反映了这一复合体的分子原貌仍然是一个未知数。最近, 通过使用蛋白质组学的手段, Rout 等^[11]鉴定了完整的酵母核孔复合体所有能检测到的多肽, 并系统地对每种可能的蛋白质组分在细胞中定位, 结合免疫电镜的方法将各组分在复合体内定位并定量, 从而揭示了酵母核孔复合体的完整分子构造, 并在此基础上揭示了其工作原理。这个工作可以说是蛋白质组学解决结构生物学问题的一个典范, 为揭示其他巨大分子机器的“构造”和工作原理指出了一条新路^[12]。

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。因为经验告诉我们, 如果两个蛋白质相互作用, 那么它们一般参与相同或相关的细胞活动^[13]。从近期国际上蛋白质组学研究的发展动向可以看出, 揭示蛋白质之间

的相互作用关系、建立相互作用关系的网络图, 已成为揭示蛋白质组复杂体系与蛋白质功能模式的先导, 业已成为蛋白质组学领域的研究热点。2000年初, 《Science》登载了一篇应用蛋白质组学的大规模双杂交技术研究线虫生殖器发育的文章^[14]。在这个工作中, Walhout 等以线虫的生殖发育过程作为研究对象, 从已知的27个与线虫发育相关的蛋白质出发, 构造了一个大规模的酵母双杂交系统, 得到了100多个相互作用的结果, 初步建立了与线虫生殖发育相关的蛋白质相互作用图谱, 从而为深入研究和揭示线虫发育的机制等提供了丰富的线索。这个工作不同于一般的应用酵母双杂交技术进行研究的地方在于, 它出于对一个生物学问题的整体思考, 尽可能地从所有已知的蛋白质而不只是个别的蛋白质为研究的出发点。这一工作为以前专注于信号转导过程中单个蛋白质作用的科学家们提供了一个新的思路, 即将整个途径的相关蛋白质一起考虑。

那么, 能否通过酵母双杂交系统来分析一种细胞或特定组织的所有可能的蛋白质间的相互作用呢? 在今年初, 《Nature》发表了一篇通过大规模双杂交技术研究酵母近6 000个蛋白质之间相互作用的论文^[15]。啤酒酵母基因组DNA的全序列业已测定, 这为通过双杂交技术来鉴定酵母基因组编码的全部6 000种左右的蛋白质间的可能相互作用提供了非常有利的条件。在这个工作中, 研究人员采用了两种不同的策略对酵母的蛋白质间的相互作用做了全面分析。一是所谓的阵列筛选法(array screening)。在此方法中, 6 000株表达不同“猎物”蛋白的酵母单克隆分别加在微滴定板上, 带有不同的“诱饵”蛋白的酵母株与前面6 000株细胞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。这篇文章中报道了192种不同的“诱饵”蛋白与近6 000种“猎物”蛋白的相互作用关系的结果。另一种方法是文库筛选法, 该方法与前一种方法的区别是, 将表达6 000种不同“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6 000株表达不同“诱饵”蛋白的酵母细胞接合, 再进一步筛选鉴定阳性克隆, 即“诱饵”与“猎物”发生相互作用的克隆。根据这篇报道, 上述两种策略得到了不同的结果, 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。这一工作的重要意义在于我们已经看到, 在基因组序列被了解的基础

上, 可以利用大规模双杂交技术全面地, 当然也是初步地, 分析一个物种或其细胞、组织的所有蛋白质之间的相互作用关系. 相信类似的工作将很快针对其他物种开展, 特别是针对基因组序列已被揭示的物种.

由此可见, 蛋白质组学已经开始从建立数据库走向解决生命科学的重大问题, 成为研究生物学问题或机制的强有力手段.

4 结束语

如果说蛋白质组学刚诞生时没有得到国际生物学主流的重视, 那么近两年情况已有了巨大的改变. 美国国立卫生研究院 (NIH) 所属的国立肿瘤研究所 (NCI) 投入了大量经费支持蛋白质组研究, 其中最大的一项资助是投入一千万美金, 用蛋白质组的方法对肺癌, 肝癌, 乳腺癌和子宫癌进行数据采集, 以补充基因组和转录组 (transcriptomics) 的结果, 进一步的目标是构建一个包含双向电泳的蛋白质组数据库. 同时, NCI 和美国食品与药物管理局 (FDA) 联合开发可用于临床的蛋白质组技术. FDA 特别感兴趣的是测试药物的疗效和毒性. 美国能源部不久前也启动了一个蛋白质组项目, 旨在研究涉及环境和能源的微生物和低等生物的蛋白质组. 欧共体目前正在资助酵母蛋白质组研究. 英国生物技术和生物科学研究委员会最近也资助了三个研究中心, 对一些已完成或即将完成全基因组测序的生物开展蛋白质组研究. 在法国, 五个研究不同模式生物的实验室得到为期三年的资助, 每年约为 500 万美元平均分配到基因组、转录组和蛋白质组研究中. 德国也没有忽略蛋白质组研究, 去年联邦政府投资了 730 万美元开展蛋白质组和相关技术的研究, 并建立一个蛋白质组学中心. 1998 年澳大利亚政府开始着手建立第一个全国性的蛋白质组研究网 APAF (Australian Proteome Analysis Facility). APAF 将为该国的有关实验室提供一流的仪器设备, 并把它们整合在一起进行大规模的蛋白质组研究. 我国关于蛋白质组研究的国家自然科学基金重大项目也从去年起开始启动. 应该说, 国家自然科学基金委在启动我国蛋白质组学的研究上是有远见和魄力的.

蛋白质组研究领域的另一个特色是, 许多实验室、公司和药厂等很早就已经开始进行与应用前有关的蛋白质组研究. 如膀胱癌、早老性痴呆症的蛋白质组研究; 利用蛋白质组技术筛选疫苗; 蛋白

质组技术在临床诊断上的应用等. 更有许多药厂和公司通过蛋白质组分析尝试药物筛选, 并从中获得药效和其副作用的信息. 有专家认为蛋白质组学方法的应用足以使得医药公司的药物开发模式发生根本改变, 现在我们可以预测药物的毒性, 而不需要在自愿者身上作试验^[4]. 极大多数知名制药公司都有自己的中等规模的蛋白质组研究, 并且常常与科研机构进行有关的合作研究. 近期的《Science》上的一篇文章介绍说, Celera 公司已经着手开始“蛋白质组”计划. 这个公司在 1998 年提出了在 3 年内完成人类基因组序列测定的雄心勃勃的计划, 看来这个目标近期内将实现. 现在, 该公司又筹集了近 10 亿美元的股金将主要用于蛋白质组学^[16]. 这充分显示出, 在蛋白质组学这个新领域, 基础研究与实际应用呈现出并驾齐驱的趋势.

虽然蛋白质组学还处在一个初期发展阶段, 但我们相信随着其不断地深入发展, 蛋白质组 (学) 研究在揭示诸如生长、发育和代谢调控等生命活动的规律上将会有所突破, 对探讨重大疾病的机理、疾病诊断、疾病防治和新药开发将提供重要的理论基础.

致谢 感谢吴家睿、李伯良教授对本稿提出有益的意见和建议.

参 考 文 献

- 1 李伯良. 功能蛋白质组学. 生命的化学, 1998, 18 (6): 1~ 4
Li B L. Chemistry of Life, 1998, 18 (6): 1~ 4
- 2 王志珍, 邹承鲁. 后基因组-蛋白质组研究. 生物化学与生物物理学报, 1998, 30 (6): 533~ 539
Wang Z Z, Tsou C L. Acta Biochimica Biophysica Sinica, 1998, 30 (6): 533~ 539
- 3 李 林, 吴家睿, 李伯良. 蛋白质组学的产生及其重要意义. 生命科学, 1999, 11 (2): 49~ 50
Li L, Wu J R, Li B L. Life Sci, 1999, 11 (2): 49~ 50
- 4 Abbott A. A post-genomic challenge: learning to read patterns of protein synthesis. Nature, 1999, 402 (6763): 715~ 720
- 5 Dutt M J, Lee K H. Proteomic analysis. Curr Opin Biotechnol, 2000, 11 (2): 176~ 179
- 6 Dalton R, Abbott A. Can researchers find recipe for proteins and chips? Nature, 1999, 402 (6763): 718~ 719
- 7 杨齐衡, 李 林. 酵母双杂交技术及其在蛋白质组研究中的应用. 生物化学与生物物理学报, 1999, 31 (3): 221~ 225
Yang Q H, Li L. Acta Biochimica Biophysica Sinica, 1999, 31 (3): 221~ 225
- 8 Mann M. Quantitative proteomics? Nature Biotechnol, 1999, 17 (10): 954~ 955
- 9 le Coutre, Whitelegge J, Gross J P, et al. Proteomics on full-length membrane proteins using mass spectrometry. Biochemistry, 2000, 39 (15): 4237~ 4242
- 10 Houry W A, Frishman D, Eckerskorn C, et al. Identification of

- in vivo* substrates of the chaperonin GroEL. 1999, *Nature*, **402** (6758): 147~ 154
- 11 Rout M P, Aitchison J D, Suprapto A, *et al.* *J Cell Biol*, 2000, **148** (4): 635~ 651
- 12 Blobel G, Wozniak R W. Proteomics for the pore. *Nature*, 2000, **403** (6772): 835~ 836
- 13 Oliver, S. Guilt-by-association goes global. *Nature*, 2000, **403** (6770): 601~ 603
- 14 Walhout A J, Sordella R, Lu X, *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science*, 2000, **287** (5450): 116~ 122
- 15 Uetz P, Giot L, Cagney G, *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 2000, **403** (6770): 623~ 627
- 16 Service R F. Proteomics. Can Celera do it again? *Science*, 2000, **287** (5461), 2136~ 2138

Progress in Proteomics. LI Lin (*State Key Laboratory of Molecular Biology, Shanghai Institute of Biochemistry, Shanghai Institute for Biological Sciences, The Chinese Academy of Sciences, China*).

Abstract Proteomics is an emerging research area that focuses on the repertoires of entire set of proteins of the cells and the activities of global cellular proteins, thus providing a new view on cell activities. A brief review of the background and current progress of proteomics is given.

Key words proteome, proteomics

中国生物化学与分子生物学会关于征集 2001 年度 英国皇家奖学金人选的通知

英国皇家学会 (Royal Society) 奖学金项目 (Royal Fellowships) 是英国女王伊莉莎白二世于 1986 年 10 月对我国进行国事访问时宣布设立的。其内容是英国皇家学会资助由其中方合作机构推荐的科技人员到英国对口单位进行合作研究一年。奖学金的基金来源于英国海外开发署和英国公司的赞助以及私人捐赠。

中国获得该奖学金的学者在英国期间,除进行正常的课题研究工作外,还可到欧洲其他国家参加一次学术会议,并参观访问英国其他研究机构和有关公司。

一 人选条件和合作课题

1. 已获得博士学位、现主持或参与国家或部级(或省(自治区、直辖市)级)重点科研项目者(需提供证明材料);
2. 年龄在 40 岁以下,有一定的实际工作经验和较强的研究工作能力;
3. 较熟练地掌握英语;
4. 拟在英国进行的合作研究课题应属于英国较先进的领域。学科主要是理、工、农科,重点是应用科学,医科原则不予考虑。

二 选拔、推荐及录取

1. 各地选送人员材料经中国生物化学与分子生物学会初评后,选出 1~2 名候选人推荐给中国科协。中国科协根据各单位推荐的人选情况择优向英国皇家学会推荐。英国皇家学会的专家小组将一一评审科协推荐的人选,对录取者英国皇家学会将发给正式录取书。申请者不得同时再申请其他英国皇家学会以外的奖学金项目。如有违反者,一经发现,将立即取消其申请资格。

2. 被推荐的人选应来自非中国科学院、中国医学科学院、教委系统的研究机构或大学等。

3. 被推荐人选需提供国家外语水平考试(EPT)成绩,并将考试成绩单(中英文)作为申请材料附件。(EPT 成绩两年之内有效)。曾在外国工作、学习 6 个月以上或有相当于 EPT 英语成绩者可免试。

三 出国手续

中国科协将协助英国皇家学会正式录取的人员办理出国手续。

四 出访费用

赴英单程旅费由本人所在单位解决,在英费用由英国皇家学会提供(伦敦 800 英镑/月,外地 750 英镑/月)。

如想申请英国皇家学会奖学金者,请于 2000 年 9 月 30 日前与中国生物化学与分子生物学会秘书处联系,索取正式表格及申请程序。

联系地址:北京朝阳区大屯路 15 号 中国生物化学与分子生物学会秘书处 邮编:100101

电话:(010) 64889892 E-mail: csbmb@sun5. ibp. ac. cn