

综述与专论

蛋白质组研究进展*

胡志远 贺福初¹⁾

(北京放射医学研究所, 北京 100850)

摘要 蛋白质组研究技术是后基因组时代的重要研究手段, 综述了流感嗜血杆菌、细菌、酵母、线虫、果蝇蛋白质组的研究进展以及蛋白质组研究技术在人类疾病研究中的应用。

关键词 蛋白质组, 双向电泳, 数据库

学科分类号 Q7

随着基因组计划的迅猛推进, 近 50 种病毒和原核生物基因组的工作接近或已经完成, 与此同时一些更高等生物的基因组研究也取得了日新月异的进展, 其中包括很重要的模式生物如酵母、线虫、果蝇、小鼠等, 而最为世人瞩目的人类基因组计划预计将在 2003 年提前完成, 那时人们将能读到全部 30 亿个碱基的序列, 人类将向了解自己的生命奥秘这一目标迈进一大步。但是, 人类对这些遗传密码编码基因的功能所知还甚少, 因此越来越多的人已经开始致力于这方面的研究, 许多新的技术、新的手段将被应用来阐明基因的功能, 如基因表达连续分析法 (serial analysis of gene expression, SAGE)^[1]和微阵列分析法 (microarray analysis)^[2]。虽然利用这些手段能够同时检测到成千上万个基因的表达, 但由于基因表达水平与蛋白质水平之间并不完全相关, 我们仍然得不到完整的信息。其解决办法之一是直接研究基因的产物——蛋白质, 即某一物种、个体、器官、组织乃至细胞的全部蛋白质。对应于基因组, 人们称后者为蛋白质组。与以往蛋白质化学的研究不同, 蛋白质组研究的对象不是单一或少数的蛋白质, 它着重的是全面性和整体性, 需要获得体系内所有蛋白质组分的物理、化学及生物学参数, 如分子质量、等电点、表达量等。目前蛋白质组研究主要方法之一是利用双向聚丙烯酰胺凝胶电泳 (2D PAGE) 来分离复杂的蛋白质组分, 并利用专门的计算机软件对所得的图象进行数据采集和分析。然后采用氨基酸组成分析、微量蛋白质序列分析、质谱分析等技术将从胶上回收的蛋白质斑点进行精细的鉴定, 获得有关蛋白质性质、表达变化及翻译后加工等方面的大规模信息。

尽管“蛋白质组”这一名称是由两位澳大利亚科学家 Wilkins 和 Williams 于 1994 年在意大利的 Siena 召开的双向电泳会议上首次提出, 但实际上它所涉及的研究方法和研究思路早已应用于生物学及医学的许多方面, 并且已经取得一些开拓性的成果。

1 原核及简单真核生物的蛋白质组研究

蛋白质组研究目前还主要集中于原核生物及一些简单的真核生物, 尤其是一些基因组序列被完全搞清或大部分已知的生物。如支原体、细菌、酵母等。

1.1 流感嗜血杆菌的蛋白质组研究

流感嗜血杆菌 (*Haemophilus influenzae*) 是第一个获得基因组全序列的生物。瑞士的一个小组通过 2D PAGE 研究了流感嗜血杆菌的蛋白质组分, 在 pH3~10 的范围内鉴定了约 300 个蛋白质组分, 然后用有两个尿素浓度的麦黄酮 (tricine) 胶分离了很难分离到的 5~20 ku 范围内的蛋白质, 还鉴定了其中的 80 种^[3,4]。另外用肝素亲和层析将碱性蛋白质富集后, 在 pH6~11 的范围内又鉴定了 102 个蛋白质, 其中许多是核酸结合蛋白尤其是核糖体蛋白^[5]。华盛顿大学的另一个小组将双向电泳得到的流感嗜血杆菌 303 个蛋白质斑点进行质谱分析, 确定了 263 个蛋白质, 其中大部分是外膜蛋白, 以及与能量代谢和大分子合成相关的蛋白质。另外发现了几种不能在基因组中找到对应序列

* 国家杰出青年科学基金 (39625014) 与国家自然科学基金重点项目 (39730270) 资助。

¹⁾ 稿件联系人。

收稿日期: 1998-09-17, 修回日期: 1999-01-25

的蛋白质. 令人惊奇的是, 大约 22% 的被鉴定了的蛋白质表现出与预计不同的等电点与分子质量, 显示它们可能经过了翻译后加工^[6].

1.2 大肠杆菌的蛋白质组研究

目前大肠杆菌全部 4000 多个基因的 DNA 序列已被测定, 人们想到如果把双向电泳得到的蛋白质谱与基因组分析结果联合起来, 就会得到更多有用的信息. 基于此想法, 建立蛋白质-基因联合数据库, 希望籍此来回答以下几方面的问题: a. 所有编码基因的产物在双向图谱上的位置, b. 各种蛋白质的丰度, c. 不同条件下各种蛋白质表达水平及合成速率的变化, d. 各种蛋白质在细胞内的定位, e. 某些蛋白质翻译后修饰的方式及水平. 目前, 大肠杆菌的联合数据库已更新到第六版^[7], 大约包括 1 600 个蛋白质斑点的数据, 其中大约 400 个蛋白质斑点已与大约 350 个基因相对应 (有些基因可能有多个蛋白质产物)^[8], 对于这些蛋白质, 这个数据库可以提供基因名称、蛋白质名称、E. C. 编号、功能范畴、Swiss-Prot 数据库编号、Genbank 序列号、基因图谱的位置、染色体上的转录方向以及一些生理信息 (如在不同生长条件下该蛋白质在细胞内的丰度).

双向电泳技术的最大优势在于它可以对细胞内复杂的蛋白质组分进行整体性的定量分析, 从而可以分析当条件变化时, 一系列蛋白质而不仅仅是某一蛋白质的表达变化. 在基因水平, “调节子”被用来描述受控于一个调节蛋白的一系列基因; 在蛋白质和基因水平, 刺激子 (stimulon) 用来描述一系列蛋白质, 他们对同一刺激物起反应^[9]. 例如在大肠杆菌中, 人们已经研究了碳、氮、磷及硫元素限制导致的细胞内蛋白质谱变化; 在对营养成分磷限制的研究中, 发现当磷饥饿时有 137 个蛋白质的合成速率明显改变, 其中大部分 (118 个) 的表现诱导合成, 其他则被抑制^[10].

1.3 致病微生物的蛋白质组研究

蛋白质组的一个重要应用是在阐明新抗生素作用机理的研究上. 当今, 细菌对大多数抗生素都有了抗性, 这导致制药公司和研究机构去寻找新的抗生素. 最近 10 年中, 用于临床的抗生素几乎都是原有抗生素的衍生物. 目前这种状况有所改变, 寻找作用于细菌内新的靶子的研究工作已经展开, 找到了许多有效的抗菌化合物. 但目前遇到的困难在于难以揭示新的化合物的作用靶子及其作用机理. 有时虽在体外发现新化合物能够使某种蛋白质失

活, 但在体内是否有这种现象和这是否是抗菌的主要机制仍然未知. 以往主要利用遗传学手段, 即通过分离和确定抗性突变株来阐明作用机理, 现在双向电泳分析提供了一个有效手段^[11]. 例如在研究抑制核糖体类抗生素对细菌的作用机制时, 对 12 种作用于翻译过程的不同阶段和核糖体内不同分子的抗生素加以考察, 发现其中 8 种诱导冷休克反应 (cold-shock response) 的一系列蛋白质, 另 4 种诱导一系列热休克反应的蛋白质, 因此预计新的作用于核糖体的化合物也是诱导这两种反应, 并且籍此可以推测大肠杆菌对冷热的反应发生于核糖体水平上^[12].

蛋白质组研究的重要优势在于能够从整体水平上分析不同条件下蛋白质谱的变化. 例如作为一种差异显示技术, 这一技术已被用于比较结核分枝杆菌与牛结核分枝杆菌蛋白的不同, 结果发现一些在基因水平上很类似的蛋白在蛋白质水平上有很大差别, 这可能部分解释近年来牛痘作为结核病疫苗会出现失败的现象^[13].

1.4 酿酒酵母的蛋白质组研究

利用双向电泳技术分析酵母蛋白谱的工作也早于蛋白质组概念的提出. 酿酒酵母 (*Saccharomyces cerevisiae*) 基因组的完成及其蛋白质数据库在互联网上的建立, 大大加速了这一工作^[14]. 最新版的数据库包括 6 000 多页, 每页代表一个已知或推测的酵母蛋白. 它包含以下几方面的信息: a. 以序列为基础的已知和推测的酵母蛋白的特征信息, 如分子质量、等电点、氨基酸组成、多肽片段大小等; b. 一些研究得到的关于各种蛋白质在翻译后加工、亚细胞定位、功能分类方面的信息; c. 从以往的超过 5 000 篇关于酵母蛋白研究的文章中获得的有关各种蛋白质功能、相互作用、突变表型的信息. 借助数据库的有力推动, 在双向电泳蛋白质谱上最明显的蛋白质斑点的大部分得到鉴定.

正在丹麦进行的一项研究计划分别作出野生型和将基因系统缺失的缺陷型酵母的双向蛋白质图谱. 初步的研究表明, 缺失单一基因将导致的结果并不只是缺乏此基因编码的蛋白质, 而是能导致其他一系列蛋白质质量的改变. 一些蛋白质减少而另一些蛋白质增多, 当然还有一些蛋白质被加工修饰而导致性状改变. 单一基因缺失的结果总是引起蛋白质组全局性的变化. 大约 20% 的酵母基因缺失是致死性的, 另外 80% 则不然, 这时机体能通过调节蛋白质的表达来对抗这种内源性的变化. 这种基

因缺失的研究可能会告诉我们一些关于细胞内蛋白质分子间相互“对话”和“作用”的重要信息,如蛋白质间的物理联系(这些蛋白质可能会组成蛋白质复合物)、信号传递途径,以帮助我们了解细胞是如何构成的以及是如何协同工作的。

2 多细胞真核生物的蛋白质组研究

2.1 线虫的蛋白质组研究

利用双向电泳技术, Bini 等^[15]对线虫(*C. elegans*)进行了蛋白质组分析,在等电点 3.5~9 和分子质量 10~200 ku 的范围内可分辨 2 000 个以上的蛋白质斑点,然后利用 Edman 微量测序技术对其中 24 个斑点进行分析,得到其中 12 个蛋白质的 N 端序列。其余的由于 N 端封闭而未能测出序列。已测出的 12 个中有 1 个未能找到与其匹配的基因。另外 11 个与能量代谢、酸性核蛋白、G 蛋白等对应。*C. elegans* 的 19 000 个基因已于 1998 年 12 月全部测出,预计会对蛋白质组的研究提供帮助。

2.2 果蝇的蛋白质组研究

不同性别果蝇(*Drosophila melanogaster*)成虫的头、胸、腹部的蛋白质组图谱已被分别作出,总共约有 1 200 个蛋白质被检出,其中的大多数在头、胸、腹中是相同的,但也发现了一部分其部位、性别特异的蛋白质^[16]。

3 人类的蛋白质组研究

人类的蛋白质组研究吸引了最多的注意力。由于人有着大量的组织、细胞类型和发育阶段,对人蛋白质组的研究主要聚焦在特异的组织、细胞和疾病上^[17]。已有的证据表明,尽管人的不同组织有着很大的功能差别,但其中的许多蛋白质是看家蛋白,因此它们的双向图谱可能也是近似的。这点与简单生物不同,简单生物在不同的发育阶段有着极不相同的蛋白质组。因此对于人类来说,一个高质量的的基本的双向图谱是非常有用的,它可以作为其他组织、细胞的参照图谱^[18]。

人的各种组织、器官、细胞乃至各种细胞器已被广泛研究^[19]。人的各种体液(血液、淋巴、脊髓、乳汁和尿等)都被用于研究与某些疾病的关系。最近澳大利亚科学家利用双向电泳技术研究眼泪中的蛋白质与生理状态的关系。他们发现了一种新的蛋白质,这个蛋白质非常相似于在乳腺癌细胞里高表达的另一种蛋白质^[20]。这个发现可能会提

供疾病诊治的新的手段。在一项利用蛋白质组研究技术进行的酒精对人体毒性的研究中发现,乙醇会改变血清蛋白糖基化作用,导致许多糖蛋白的糖基缺乏,如转铁蛋白^[21]。

肿瘤至今仍是人类的一个顽敌。癌细胞不仅有许多特异蛋白质产物,而且这些产物还会影响其他蛋白质的翻译后加工及许多蛋白质的表达水平。肿瘤的蛋白质组研究能提供一些以往其他技术所不能提供的早期诊断依据,例如利用不同细胞组织的特征蛋白质谱,可以判别一些难以判定来源的肿瘤细胞的来源。丹麦的一个小组利用蛋白质组研究技术结合组织冰冻切片技术分析 150 例膀胱癌病人的组织,发现在所有的鳞状细胞瘤的尿液中均能发现一种化学引诱剂——银屑素(psoriasin),推测其极可能作为这种肿瘤的早期标志物^[22]。

许多科学家致力于各种肿瘤组织与正常组织之间蛋白质谱差异的研究,这些组织包括脑、胸腺、乳腺、肺、直肠、肾、膀胱、卵巢、骨髓等,这些研究已经找到了一些肿瘤特异的标志物,可能会对揭示肿瘤发生的机制有所帮助。例如,在对肾癌的研究中发现有 4 种蛋白质存在于正常肾组织而在肾癌细胞中缺失。其中两种分别是辅酶 Q 蛋白色素还原酶和线粒体泛醌氧化——还原复合物 I。这提示线粒体功能低下可能在肿瘤发生过程中起重要作用^[23]。

基因组计划无疑为医疗、医药领域带来一场革命,但也应该看到,单纯的遗传分析很难诊断多因素的疾病。复杂的基因间相互作用,细胞内活动和环境的影响都会影响基因的表达及蛋白质的翻译后加工。因此,可靠的诊断和治疗应基于机体渐进发展过程的调控及失调,并且必须考虑到环境因素的影响。蛋白质组的研究正是探索这一领域的有力武器。人们不难预期,基因组计划的不断推进会给蛋白质组研究提供更多更全的数据库;生物信息学的发展会给蛋白质组计划提供更方便有效的计算机分析软件;国际互联网会使各国各领域科学家有关蛋白质组研究的成果出现新的集成;新的技术会不断涌现,蛋白质组研究方法会象 PCR 技术一样易于操作,并渗透到人类活动的方方面面,对工业、农业、医疗卫生各行各业带来新的革命。

参 考 文 献

- 1 Velculescu V E, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. Science, 1995, 270 (5235): 484~487
- 2 Schena M, Shalon D, Davis R W, et al. Quantitative monitoring

- of gene expression patterns with a complementary DNA microarray. Science, 1995, **270** (5235): 467~ 470
- 3 Langen H, Gray C, Roder D, *et al.* From genome to proteome: protein map of *Haemophilus influenzae*. Electrophoresis, 1997, **18** (7): 1184~ 1192
 - 4 Fountoulakis M, Juranville J F, Roder D, *et al.* Reference map of the low molecular mass proteins of *Haemophilus influenzae*. Electrophoresis, 1998, **19** (10): 1819~ 1827
 - 5 Fountoulakis M, Takacs B, Langen H. Two-dimensional map of basic proteins of *Haemophilus influenzae*. Electrophoresis, 1998, **19** (5): 761~ 766
 - 6 Link A J, Hays L G, Carmack E B, *et al.* Identifying the major proteome components of *Haemophilus influenzae* type strain NCTC 8143. Electrophoresis, 1997, **18** (8): 1314~ 1334
 - 7 VanBogelen R A, Abshire K Z, Pertsemliadis A, *et al.* *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. Washington DC: ASM Press, 1996. 2067~ 2117
 - 8 Pasquali C, Frutiger S, Wilkins M R, *et al.* Two-dimensional gel electrophoresis of *Escherichia coli* homogenates: the *Escherichia coli* SWISS-2DPAGE database. Electrophoresis, 1996, **17** (3): 547~ 555
 - 9 Neidhardt F C, Savageau M A. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. Washington DC: ASM Press, 1996. 1310~ 1324
 - 10 VanBogelen R A, Olson E R, Wanner B L, *et al.* Global analysis of proteins synthesized during phosphorus restriction in *Escherichia coli*. J Bacteriol, 1996, **178** (15): 4344~ 4366
 - 11 VanBogelen R A, Abshire K Z, Moldover B, *et al.* *Escherichia coli* proteome analysis using the gene-protein database. Electrophoresis, 1997, **18** (8): 1243~ 1251
 - 12 VanBogelen R A, Neidhardt F C. Ribosomes as sensors of heat and cold shock in *Escherichia coli*. Proc Nat Acad Sci USA, 1990, **87** (15): 5589~ 5593
 - 13 Urquhart B L, Atsalos T E, Roach D, *et al.* 'Proteomic contigs' of *Mycobacterium tuberculosis* and *Mycobacterium bovis* (BCG) using novel immobilised pH gradients. Electrophoresis, 1997, **18** (8): 1384~ 1392
 - 14 Hodges P E, Payne W E, Garrels J I. The Yeast Protein Database (YPD): a curated proteome database for *Saccharomyces cerevisiae*. Nucleic Acids Res, 1998, **26** (1): 68~ 72
 - 15 Bini L, Heid H, Liberatori S, *et al.* Two-dimensional gel electrophoresis of *Caenorhabditis elegans* homogenates and identification of protein spots by microsequencing. Electrophoresis, 1997, **18** (3~ 4): 557~ 562
 - 16 Ericsson C, Petho Z, Mehlin H. An on-line two-dimensional polyacrylamide gel electrophoresis protein database of adult *Drosophila melanogaster*. Electrophoresis, 1997, **18** (3~ 4): 484~ 490
 - 17 Arnott D, O'Connell K L, King K L, *et al.* An integrated approach to proteome analysis: identification of proteins associated with cardiac hypertrophy. Anal Biochem, 1998, **258** (1): 1~ 18
 - 18 Wilkins M R, Williams K L, Appel R D, *et al.* Proteome Research: New Frontiers in Functional Genomics. Berlin Heidelberg: Springer-Verlag, 1997. 187~ 220
 - 19 Rabilloud T, Kieffer S, Procaccio V, *et al.* Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. Electrophoresis, 1998, **19** (6): 1006~ 1014
 - 20 Molloy M P, Bolis S, Herbert B R, *et al.* Establishment of the human reflex tear two-dimensional polyacrylamide gel electrophoresis reference map: new proteins of potential diagnostic value. Electrophoresis, 1997, **18** (15): 2811~ 2815
 - 21 Gravel P, Walzer C, Aubry C, *et al.* New alterations of serum glycoproteins in alcoholic and cirrhotic patients revealed by high resolution two-dimensional gel electrophoresis. Biochem Biophys Res Commun, 1996, **220** (1): 78~ 85
 - 22 Ostergaard M, Rasmussen H, Nielsen H, *et al.* Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. Cancer Research, 1997, **57** (18): 4111~ 4117
 - 23 Sarto C, Marocchi A, Sanchez J C, *et al.* Renal cell carcinoma and normal kidney protein expression. Electrophoresis, 1997, **18** (3~ 4): 599~ 604

Progress in Proteome Research. HU Zhi-Yuan, HE Fu-Chu (Institute of Radiation Medicine, Beijing 100850, China).

Abstract Proteome research techniques are the important tools in the Post-Genome Era. A review is given about the progress in proteome research of bacteria, *Haemophilus influenzae*, *Saccharomyces cerevisiae*, *C. elegans*, *Drosophila melanogaster* and human diseases.

Key words proteome, two dimension electrophoresis, database

高效的骨诱导生长因子——成骨蛋白-1*

王 健 吴惠联¹⁾ 韩金祥

(山东省医药生物技术研究中心, 济南 250062)

摘要 成骨蛋白-1 (OP-1) 又称骨形态发生蛋白-7 (BMP-7), 属转化生长因子 β (TGF- β) 超家族成员。重组人 OP-1 (rhOP-1) 在体内和体外都显示了高效的骨诱导活性, 可使多种实验动物的骨缺损满意愈合, 有良好的临床应用前景。

* 山东省重大攻关项目 (981154506)。 ¹⁾ 山东省医学科学院基础所微生物研究室。

收稿日期: 1998-01-17, 修回日期: 1998-06-08