中国科学院科学出版基金资助出版

生命科学前沿丛书

蛋白质组学：理论与方法

钱小红　贺福初　主编

科学出版社

北　京

内 容 简 介

蛋白质组学是当今生命科学热点与前沿——功能基因组学中的重要研究领域。本书从蛋白质组与蛋白质组学的基本概念入手，重点介绍了这一崭新领域的诞生与发展，并以具体的研究成果为例，详细介绍了相关技术及应用进展。全书共13章，包括蛋白质组学基础知识与研究技术、生物信息学、肿瘤发生与发展的比较蛋白质组学、细胞凋亡的蛋白质组研究、蛋白质组与新药开发等。资料系统、新颖而实用。

本书可供分子生物学、生物化学、细胞生物学以及医学、药学等领域的科研、教学人员及研究生参考。

图书在版编目(CIP)数据

蛋白质组学：理论与方法/钱小红、贺福初 主编 .—北京：科学出版社，2003

（生命科学前沿丛书）

ISBN 7-03-010864-7

Ⅰ.蛋… Ⅱ.①钱… ②贺… Ⅲ.蛋白质-研究 Ⅳ.Q51

中国版本图书馆 CIP 数据核字（2002）第 068866 号

责任编辑：马学海 余和芬/责任校对：钟 洋

责任印制：刘士平/封面设计：王 浩

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

印刷

科学出版社发行 各地新华书店经销

*

2003 年 6 月第 一 版 开本：B5（720×1000）

2005 年 7 月第三次印刷 印张：22

印数：6 001—8 000 字数：435 000

定价：45.00 元

（如有印装质量问题，我社负责调换〈新欣〉）

编 写 人 员

（按汉语拼音排序）

蔡　耘　贺福初　姜　颖
蒋代凤　刘尚义　钱小红
万晶宏　王　建　王京兰
王　妍　杨何义　应万涛
张令强　朱云平

目　录

第一章　功能基因组与蛋白质组

基因研究是20世纪生命科学的主线。20世纪的上半叶，以遗传学为代表，生命科学通过对基因分离、独立分配、连锁及化学属性等的研究，最后以作为遗传信息载体的DNA双螺旋结构的提出而告捷；20世纪的下半叶，以分子生物学为代表，生命科学通过对基因复制、转录、翻译及遗传密码的分析与破译，最终以统一生命世界各层次、生命科学各分支的“中心法则”的问世而集成；20世纪90年代，随着全球性基因组计划尤其是人类基因组计划（HGP）规模空前、速度惊人的推进，基因研究已接近“登峰造极”，人类对生命世界的理性认识达到了前所未有的深度与广度。

人类基因组计划被誉为20世纪的三大科技工程之一。其划时代的研究成果——人类基因组序列草图的完成，宣告了一个新的纪元——“后基因组时代”的到来。其中，功能基因组学（functional genomics）成为研究的重心，蛋白质组学（proteomics）则是其“中流砥柱”。正因为如此，*Nature*、*Science*分别在2001年2月15日、16日公布人类基因组草图的同时，分别发表了“And now for the proteome”（Nature 409：747，2001）、“Proteomics in genomeland”（Science 291：1221，2001）的述评与展望，将蛋白质组学的地位提到前所未有的高度，认为是功能基因组学这一前沿研究的战略制高点，蛋白质组学将成为新世纪最大战略资源——人类基因尤其是重要功能基因争夺战的重要“战场”。

人们在欢呼基因组计划辉煌业绩之时，亦愈来愈清醒地意识到一项更艰巨、更宏大的任务即基因组功能的阐明已经摆在面前，生命科学几乎在转瞬之间开始了新的征程——蛋白质组研究[1~3]，进入了一个新的纪元——后基因组时代（postgenome era）[4,5]。人类经过一个世纪的跋涉，重返现代生命科学的发源地之一，蛋白质——这一生命活动的执行体。当然，这不是简单的回归，而是一次真正的黑格尔式的“重返”。

第一节　人类基因组计划及蛋白质组研究的历史背景

一、基因组计划的成就

以人类基因组计划为代表的基因组计划是20世纪仅次于曼哈顿原子弹研制计划与阿波罗登月计划的重大科技工程。其中，HGP旨在完成人基因组24条染

色体上5万左右基因的作图（遗传图与物理图）和30亿碱基的DNA全序列的测定。此计划自1990年实施以来进展神速：1994年人基因组全套遗传连锁图发表[6]，1995年全基因组覆盖率高达94%的物理图问世[7]；同年，汇集了人基因组初步全物理图，3、12、16、22号染色体高密度物理图以及30余万左右cDNA（EST）序列信息的“人基因组指南”经 *Nature* 结集出版[8]；2000年6月26日，宣告人类基因组序列草图测定完成；2001年2月15日、16日，国际人类基因组计划与美国Celera公司分别在 *Nature*、*Science* 公布人类基因组草图[9]。与此同时，模式生物与致病微生物等的基因组研究亦如火如荼地展开。自1995年支原体（*Mycoplasma genitalium*）和流感嗜血杆菌（*Hemophilus influenzae*）基因组全序列发表[10,11]以来，已相继有10余种原核生物的基因组全序列发表，如大肠杆菌（K-12）[12]。更令人振奋的是，第一个真核生物——酵母的基因组全序列于1996年完成[13]，次年，Nature再次推出（酵母基因组指南）专辑[14]。此外，多细胞真核生物线虫（*Caenorhabditis elegans*）的全基因组序列测定也取得长足进展[15]。截至2001年底，已有不少于75种生物的基因组全序列测定完成，基因组计划已经进入全面收获的“金秋时节”（表1-1综合了共75个目前已经完成全基因组序列测定的生物），而“海量”的基因序列数据为生命科学多层次、多分支的研究提供了丰富的宝藏。

表1-1　目前已经完成全基因组序列测定的生物

（共75个，截至2002年初）

生物种类	基因组大小/kb	ORF数目	测序机构
古细菌（12个）			
Methanococcus jannaschii DSM 2661	1664	1750	美国伊利诺伊州立大学和TIGR公司
Methanobacterium thermoautotrophicum delta H	1751	1918	基因组治疗公司和俄亥俄州立大学
Archaeoglobus fulgidus DSM4304	2178	2493	美国伊利诺伊州立大学和TIGR公司
Pyrococcus horikoshii（*shinkaj*）OT3	1738	1979	东京大学和日本NITE（National Institute of Technology and Evaluation）
Aeropyrum pernix K1	1669	2620	日本NITE（National Institute of Technology and Evaluation）
Pyrococcus abyssi GE5	1765	1765	法国Genoscope中心

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Halobacterium sp. NRC-1	2014	2058	华盛顿大学和麻省州立大学
Thermoplasma acidophilum	1564	1478	德国 Max Planck 生化研究所和 Medigenomix 公司
Thermoplasma volcanium GSS1	1584	1524	日本 AIST（National Institute of Advanced Industrial Science and Technology）
Sulfolobus solfataricus P2	2992	2977	欧盟和加拿大生物信息资源中心
Sulfolobus tokodaii 7	2694	2826	日本 NITE（National Institute of Technology and Evaluation）
Pyrobaculum aerophilum IM2	2222	2587	加利福尼亚理工学院（California Institute of Technology）和加州大学
细菌（57 个）			
Haemophilus influenzae KW20	1830	1850	TIGR 公司
Mycoplasma genitalium G-37	580	468	TIGR 公司
Synechocystis sp. PCC6803	3573	3168	日本 Kazusa DNA Research Institute（KDRI）
Mycoplasma pneumoniae M129	816	677	德国海德堡大学
Escherichia coli K12- MG1655	4639	4289	美国威斯康辛大学
Helicobacter pylori 26695	1667	1590	TIGR 公司
Bacillus subtilis 168	4214	4099	法国 Pasteur 研究所和日本京都大学
Borrelia burgdorferi B31	1230	1256	Brookhaven Natl 实验室和 TIGR 公司
Aquifex aeolicus VF5	1551	1544	美国伊利诺伊州立大学
Mycobacterium tuberculosis H37Rv (lab strain)	4411	4402	英国 Sanger 中心
Treponema pallidum subsp. pallidum Nichols	1138	1041	美国德克萨斯州和 TIGR 公司

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Chlamydia trachomatis serovar D	1042	896	美国斯坦福大学和加利福尼亚大学伯克莱分校
Rickettsia prowazekii Madrid E	1111	834	瑞典 Uppsala 大学
Helicobacter pylori J99	1643	1495	基因组治疗公司和美国 Astra 公司
Chlamydia pneumoniae CWL029	1230	1052	美国斯坦福大学和加利福尼亚大学伯克莱分校
Thermotoga maritima MSB8	1860	1877	TIGR 公司
Deinococcus radiodurans R1	3284	3187	TIGR 公司
Ureaplasma urealyticum serovar 3	751	650	Eli Lilly 公司和美国 Alabama 大学、Perkin Elmer 公司
Campylobacter jejuni NCTC 11168	1641	1654	英国 Sanger 中心和英国 Leicester 大学、伦敦卫生与热带医学学校
Chlamydia pneumoniae AR39	1229	1052	英国 Sanger 中心和加拿大 Manitoba 大学
Chlamydia trachomatis MoPn Nigg	1069	924	英国 Sanger 中心和加拿大 Manitoba 大学
Neisseria meningitidis MC58 (serogroup B)	2272	2158	TIGR 公司
Neisseria meningitidis Z2491 (serogroup A)	2184	2121	英国 Sanger 中心和牛津大学、德国 Max-Planck 分子遗传学研究所
Bacillus halodurans C-125	4202	4066	日本海洋科学技术中心

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Chlamydia pneumoniae J138	1228	1070	日本 Yamaguchi 大学和 KYUSHU 大学
Xylella fastidiosa CVC 8.1.b clone 9.a. 5.c	2679	2904	巴西 ONSA 中心
Vibrio cholerae serotype O1, Biotype El Tor, strain N16961	4000	3885	TIGR 公司
Pseudomonas aeruginosa PAO1	6264	5570	美国华盛顿大学和 Chiron 公司
Buchnera sp. APS	640	564	日本东京大学和 RIKEN 研究所（The Institute of Physical and Chemical Research）
Mesorhizobium loti MAFF303099	7596	6752	日本 Kazusa DNA Research Institute（KDRI）
Escherichia coli O157: H7 EDL933	4100	5283	美国威斯康辛大学
Mycobacterium leprae TN	3268	1604	英国 Sanger 中心和法国 Pasteur 研究所
Escherichia coli O157: H7. Sakai	5594	5448	日本京都大学
Pasteurella multocida Pm70	2250	2014	美国明尼苏达州立大学
Caulobacter crescentus	4016	3737	TIGR 公司

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Streptococcus pyogenes SF370（M1）	1852	1696	美国俄克拉荷马州立大学
Lactococcus lactis IL1403	2365	2266	法国 Genoscope 中心
Staphylococcus aureus N315	2813	2594	日本 NITE（National Institute of Technology and Evaluation）和 Juntendo 大学、Tsukuba 大学、东京大学、KYUSHU 大学等
Staphylococcus aureus Mu50	2878	2697	日本 NITE（National Institute of Technology and Evaluation）和 Juntendo 大学、Tsukuba 大学、东京大学、KYUSHU 大学等
Mycobacterium tuberculosis CDC 1551	4403	4187	TIGR 公司
Mycoplasma pulmonis	963	782	法国 Genoscope 中心
Streptococcus pneumoniae TIGR4	2160	2094	基因组治疗公司
Sinorhizobium meliloti 1021	6690	6205	欧盟和美国斯坦福大学
Streptococcus pneumoniae R6	2038	2043	Eli Lilly 公司
Rickettsia conorii Malish 7	1268	1374	法国 Genoscope 中心
Yersinia pestis CO-92 Biovar Orientalis	4653	4012	英国 Sanger 中心和 MDS 公司、Dstl 实验室、帝国学院
Salmonella typhi CT18	4809	4600	英国 Sanger 中心和帝国学院

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Salmonella typhimurium, *LT2* SGSC1412	4857	4597	美国华盛顿大学
Listeria innocua Clip11262, rhamnose-negative	3011	2981	法国 Pasteur 研究所
Listeria monocytogenes EGD-e	2944	2855	法国 Pasteur 研究所
Nostoc sp. PCC 7120	6413	5366	日本 Kazusa DNA Research Institute (KDRI) 和美国密西根州立大学
Agrobacterium tumefaciens C58-Cereon	4915	5299	美国 Monsanto 公司和 Cereon 公司、Richmond 大学
Agrobacterium tumefaciens C58-DuPont	4915	5402	美国华盛顿大学和 DuPont 公司、巴西 Campinas 大学
Ralstonia solanacearum GMI1000	5810	5120	法国 Genoscope 中心和 INRA 公司、CNRS 公司
Brucella melitensis 16M	3294	3197	美国 Scranton 大学和 Integrated Genomics 公司
Clostridium perfringens 13	3031	2660	日本 Tsukuba 大学和 Kitasato 大学、Kyushu 大学
真核生物 (9 个)			
Saccharomyces cerevisiae S288C	12 069	6294	国际合作
Caenorhabditis elegans	97 000	19 099	美国华盛顿大学和英国 Sanger 中心
Drosophila melanogaster	137 000	14 100	美国 Celera 公司和加大伯克莱分校 DGP (Drosophila Genome Project)、Baylor College of Medicine、欧洲 DGP

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
Arabidopsis thaliana	115 428	25 498	国际合作
Homo sapiens	约 3 000 000	35 000～50 000	国际合作
Guillardia theta	551	464	加拿大大不列颠哥伦比亚大学和 Canadian Institute for Advanced Research（CIAR）、德国 Philipps 大学
Leishmania major Friedlin Chromosome 1	257	79	美国 Seattle Biomedical Research Institute（SBRI）
Plasmodium falciparum 3D7 Chromosome 2	947	205	TIGR 公司
Plasmodium falciparum 3D7 Chromosome 3	1060	220	英国 Sanger 中心

二、基因组计划的局限

任一生物基因组计划［此处按经典含义指结构基因组学（structural genomics）分析］的完成均标志着三套完整数据的获得：遗传图、物理图、全序列图。理论上，这三套数据将提供此生物所有基因在染色体上的精确定位、基因内部序列结构与所有基因间隔序列。但是，由于真核生物中基因结构的复杂性以及现有基因识别（gene identification）理论与技术发展的严重不足，此情况只适用于原核生物或低等真核生物。正因为如此，即使 HGP 能在 2001 年完成，也并不表明此时人类对自身基因组的所有基因及其间隔序列已完全确定。真核生物尤其是高等真核生物已测定基因组中 ORF（可读框，或名开放阅读框架）的确定仍是未解决的重大问题，而一个基因在 ORF 确定前很难从分子水平上进行实质性的功能分析。

基因调控研究表明，即使是简单的微生物（如大肠杆菌），其基因组的所有基因也不同时表达。通常情况下，生物的基因组只表达少部分基因，而且表达的基因类型及其表达程度随生物生存环境及内在状态的变化而表现极大的差别，且

此差别存在严格调控的时空特异性。基因组计划即使已确定某生物基因组内的全部基因，也不能告诉人们哪些基因在何时何地以何种程度表达，而生命过程的精确机制很大程度上正是基于这类基因的精细调控。为了弥补基因组计划这一天然的局限，近年人们相继引进一系列大规模基因表达检测技术，如微阵列法（microarray）[16]、DNA 芯片（DNA chips）[17]及 SAGE（serial analysis of gene expression）[18]等。这些方法虽然能够定性、定量且大规模地检测基因的表达产物 mRNA，但 mRNA 由于自身存在贮存、转运、降解、翻译调控及产物的翻译后加工，难以准确地反映基因的最终产物/基因功能的真正执行体——蛋白质的质与量。

基因与其编码产物蛋白的线性对应关系只存在于新生肽链而不是最终的功能蛋白中。30 多年前，人们即已普遍发现新生肽链合成后存在多种加工、修饰过程；更有甚者，近些年来人们发现蛋白质间亦存在类似于 mRNA 分子内的剪切、拼接，并证明其基本元件“intein”广泛存在于多种蛋白质中[19]。此类过程的存在无疑进一步扩大了基因编码的蛋白质与其最终的功能蛋白间所存在的序列差距。而大量蛋白尤其是重要调控蛋白的化学修饰（如糖基化、磷酸化）、剪切加工（如酶原降解、结构域拼接）不但可改变其立体结构，而且是实施其功能与调节的重要结构基础。这些均不能从其基因编码序列中预测，而只能通过对其最终的功能蛋白进行分析。

从上可见，基因虽是遗传信息的源头，而功能性蛋白是基因功能的执行体。基因组计划的实现固然为生物有机体全体基因序列的确定，为未来生命科学研究奠定了坚实的基础，但是它并不能提供认识各种生命活动直接的分子基础，其间必须研究生命活动的执行体——蛋白质这一重要环节。

任一层次的生命活动均是非线性复杂系统中各种功能单元协同、整合的结果，生命活动的最小单元——“细胞”即是多类“蛋白机器”（protein machine）的有机组合[20]。人类对于蛋白质的研究已逾百年，但以往的视角只是针对生命活动中某一种或某几种蛋白质，这样难以形成一种整体观，难以系统透彻地阐释生命活动的基本机制。因此，无论是从基因组计划的局限、还是从蛋白质研究的自身发展而言，大规模、全方位的蛋白质研究均是势在必行。

蛋白质是生物细胞赖以生存的各种代谢和调控途径的主要执行者，因此蛋白质不仅是多种致病因子对机体作用最重要的靶分子，并且也成为大多数药物的靶标乃至直接的药物。药靶，来源于对生命活动的生理病理过程的研究；药靶，又形成制药业的发展源头。蛋白质组学正是近年来新发展起来的强有力的发现药靶的技术平台，作为一个新的学科发展领域，它对所有及时进入的国家都将提供巨大的机会。机不可失，时不我待。

一项科学统计表明：在 20 世纪 90 年代中期，全世界制药业用于找寻新药的药靶共约 483 个，它们主要是蛋白质（受体占 45%，酶占 28%等）；而当时全世

界正在使用的药物总数约是 2000 种，其中 85%都是针对上述 483 种药靶。这 483 种药靶分子构成了全世界药厂的最重要的发展源泉。从功能基因组学的角度，人们认为每种疾病平均与 10 个左右基因相关，而每种基因又与 3～10 种蛋白质相关，如果以人类主要的 100～150 种疾病进行计算，则应该有 3000～15 000种蛋白质具有成为药靶的可能。也就是说还可能有几千到上万种的新药靶将被发现，这将是功能基因组研究有可能带来的一笔巨大的科学、经济财富；不容置疑，这也是为什么蛋白质组学作为发现药靶的主要技术平台在 20 世纪 90 年代末期以来越来越受国际巨型跨国制药集团垂青的重要原因所在。

三、蛋白质研究技术方法的突破

蛋白质的研究在 20 世纪 70 年代以前一直优于核酸。其后，由于 DNA 重组、测序、PCR 等新方法的不断涌现，核酸研究后来居上，并远远超出而成为生命科学的主导，但蛋白质研究尤其是相关技术的发展并未停滞不前。其中 O'Farrel PH 于 1975 年建立的二维电泳（又名双向电泳，2-DE）技术使蛋白分辨达到成千上万种，因而完全可以用于组织与细胞中大规模蛋白质的分离[21]；近年开发的多种图像分析系统与软件以及大规模样品处理系统更使其如虎添翼。20 世纪 80 年代末期 Hillenkamp F 发展的激光解吸质谱、Fenn J 设计的电喷雾质谱可以高效、精确地测量生物大分子的质量并测定部分序列[22]，进而用于数据库的检索；Mann M 等则在此基础上通过建立“肽质量指纹图谱与肽序列标签”等技术，实现了质谱准确、快速、自动化、大规模鉴定蛋白质的飞跃[23]。

第二节　蛋白质组研究的开端及“蛋白质组”含义

一、蛋白质组研究的开端

“proteome”（蛋白质组）一词由 Marc Wilkins 于 1994 年在意大利 Siena 的一次 2-DE 电泳会议上首次提出。其导师澳大利亚 Macquarie 大学的 Keith Williams 于同年向澳政府提出一项建议：通过对某一种生物的所有蛋白质全部进行质谱筛选与序列分析，以一种不同于 DNA 快速测序的途径对其提供分子水平的全面分析。1995 年，悉尼大学 Humphery Smith I 实验室与 Williams 等 4 家实验室合作，对至今已知最小的自我复制生物 *Mycoplasma genitalium*（一种支原体）进行了蛋白质成分的大规模分离与鉴定，并在文献中首次公开使用“proteome”一词，同时指出该文所采用的技术体系对于大规模鉴定并分析基因对应的产物以及发现新型蛋白均具有十分重要的意义[1]。

二、蛋白质组的含义

根据 Wilkins MR 等[24]的定义，“proteome”一词源于“PROTEin”与“genOME”的杂合，意指“一种基因组所表达的全套蛋白质”；Swinbanks[3]则指出“proteome”代表一完整生物的全套蛋白质。与此同时，Kahn P 则认为“proteome”反映不同细胞的不同蛋白质组合[2]。由此可见，“proteome”有三种不同的含义：一个基因组、一种生物或一种细胞/组织所表达的全套蛋白质。

三、蛋白质组研究的技术路线与相关技术

一般细胞含有数千种乃至上万种蛋白质。蛋白质组研究的宗旨是将组织或细胞所有蛋白质（至少是大部分）分离与鉴定。为达到目的，它引进了下列技术[21]：双向电泳（2-DE），如 ISO-DALT、IPG-DALT 或 NEPHGE；图像分析系统，如 ELSIE 4 & 8、gellab Ⅰ&Ⅱ、MELANIE Ⅰ & Ⅱ、QUEST Ⅰ&Ⅱ与 PDQUEST、TYCHO & KEPLAR；蛋白质鉴定方法，如氨基酸组成分析、序列测定、肽质量指纹图（peptide mass fingerprinting，PMF）、相对分子质量精确测定；HTS（high throughput system）系统与大规模样品处理机器人；数据库设置与检索系统[25]。其中，蛋白质鉴定采用了新近出现的新型质谱（MS）技术，如 MALDI-TOF（matrix assisted laser desorption ionisation-time of flight）MS 与 ESI（electrospray ionisation）/MS/MS。

大规模蛋白质组分析过程包括样品制备、图像分析、蛋白质成分的分析与鉴定。其技术路线及数据处理见有关章节。其中“层次分析”（hierarchial analysis）包括[26]：氨基酸分析、肽质量指纹图、氨基酸分析与 PMF 联合、序列标签途径、N 端 Edman 降解蛋白与微量测序、蛋白质内肽微量测序、MS（MALDI-TOF，ESI）微量测序、“Ladder”测序、MS 对 PVDF 膜或电泳胶上低拷贝分子的系统筛选、基因组文库中克隆片段的倾向性表达。

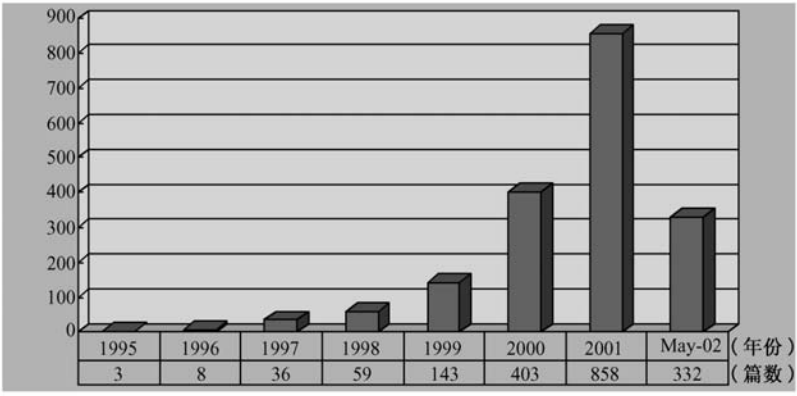
四、蛋白质组研究的国内外现状

1. 发展速度与规模

自 1995 年“proteome”一词问世以来，截至 1997 年底相关文献已达 41 篇，其中研究论文 28 篇，各类评述综述 13 篇。研究论文中，1995 年 1 篇，1996 年 4 篇，1997 年 23 篇，近 4 年更是以指数式增长。其论文增长速度见表 1-2。1995 年首倡“proteome”的两家澳大利亚实验室，1997 年分别挂牌“Centre for Proteome Research and Gene Product Mapping，National Innovation Centre”和“Aus-

tralia Proteome Analysis Facility"。同年，丹麦成立"Centre for Proteome Analysis in Life Sciences"，美国成立专事此类研究与开发的两家公司："Proteome，Inc. Beverly，MA"和"Large Scale Biology Corp，Rockville，MD"。参与的国家，1995年只有澳大利亚，1997年则发展到美国、丹麦、瑞士、英国、法国、日本、瑞典、意大利、德国等10国。国际著名学府如哈佛、斯坦福、耶鲁、密执安、华盛顿大学、欧洲分子生物学实验室、巴斯德研究所、瑞士联邦工业学院等均挤身此类研究。其中，澳大利亚悉尼大学与Macquarie大学发展势头强劲，但美欧多家实验室已奋起直追，正是"群雄纷起，逐鹿中原"。

表1-2　蛋白质组研究论文增长速度



2．研究材料

1995年，Wasinger等[1]在第一篇蛋白质组研究文章中研究的对象为目前已知最小但能自主复制的原核微生物——支原体 *Mycoplasma genitalium*。1996年，研究对象即扩展到单细胞真核生物——酵母[27]以及人体正常组织及病理标本[28]，进而突破了早期人们普遍认为的"蛋白质组研究只适用于基因组计划已完成的生物"的界限[1,24]。因此，1997年，研究对象一下扩展到14种生物，其中虽然绝大多数为原核生物，但也包含多细胞真核生物如线虫[29]。近4年来，蛋白质组研究对象已无任何限制：无需基因组计划完成（当然完成者更好），无原核生物/真核生物、单细胞/多细胞、组织之分。

3．研究范围

"蛋白质组"不仅其研究已成为具有重大战略意义的科学命题，而且其分析已成为一种十分有效且应用广泛的研究手段。正因如此，蛋白质组的研究与分析

其范围在短时间内即扩展到令人惊诧的程度。据不完全统计，目前至少已涉及如下方面：① 蛋白质，如蛋白质组作图[30]、蛋白质组成分鉴定[31]、蛋白质组数据库构建[32]、新型蛋白质发掘[33]、蛋白质差异显示[34]、同工体（isoform）比较[35,36]；② 基因，如功能基因组计划[35~37]、基因产物识别[38]、基因功能鉴定[39]、基因调控机制分析[40,41]；③ 重要生命活动的分子机制，包括细胞周期[26]、细胞分化与发育[29,32]、肿瘤发生与发展[42,43]、环境反应与调节[32,37,41]、物种进化等[44~46]；④ 医药靶分子寻找与分析，靶分子类型包括新型药物靶分子[33,34,44]、肿瘤恶性标志[42,43]、人体病理介导分子[28]、病原菌毒性成分[33,44]。由此不难看到，“蛋白质组”研究与分析已涉足生命科学中一系列热点领域。

蛋白质组作图是早期蛋白组研究的主要领域。经过最初三年的努力，在1995年已测定10种蛋白质组图（2-DE）[24]的基础上，又测定了11种生物或组织的蛋白质组图，其中3种生物（线虫[29]、豆科植物根瘤菌属[34]、*Ochrobactrum anthropi*[47]）的蛋白质组图超过1600点，3种蛋白质组图（大肠杆菌[39]、盘基网柄菌[32]、人正常组织与病理组织[28]）建立数据库并上互联网；各类生物中第一个完整的蛋白质组数据库（YPD）完成[48]，含6021种蛋白；提出“蛋白差异显示”概念，并用于环境应激、基因突变、病理进程等研究[34]；建立“proteomic contig”方法，进而使蛋白质组图分辨率提高10倍[47]；提出“蛋白连锁图”（protein linkage map）概念，改进双杂交系统，用于蛋白质组相互作用网络的分析[49,50]；联合液体自动取样器与LC/ESI-MS技术，使蛋白质鉴定速度达20点/日[35]；建立2-DE中糖蛋白与膜蛋白微量鉴定方法[51,52]；建立“Streamlined样品处理”胶转膜技术，突破了大规模蛋白测序与氨基酸组成分析这两个严重影响蛋白质组分析中蛋白质鉴定的限速步[53]。

4. 国内外研究现状

国外研究现状

基因、尤其是基因组研究形成了20世纪生命科学研究一道亮丽的风景线。根据2001年5月23日所公布的结果，已公开发表67个生物体的基因组全序列（含染色体），其中包括古细菌11个，细菌41个，真核生物15个，此外有208个原核生物和152个真核生物或其染色体的基因组正在测序（http://igweb.integratedgenomics.com/GOLD/）。基因组研究所产生的“海量”数据和大量新型技术以前所未有的深度与广度极大地推动了生物医学多学科的飞速发展。

但是，随着大量生物体全基因组序列的获得、特别是人类基因组序列草图的完成，基因组研究的战略重点不可避免地从结构基因组学（structural genomics）转向功能基因组学（functional genomics），而蛋白质组学（proteomics）正是作为功能基因组研究的重要支柱在20世纪90年代中期应运而生[1]。也正因如此，*Nature*、*Science*在2001年2月公布人类基因组草图的同时，分别发表了“And

now for the proteome"[54]、"Proteomics in genomeland"[55]的述评与展望，将蛋白质组学的地位提到前所未有的高度，认为是功能基因组学这一前沿研究的战略制高点，蛋白质组学将成为新世纪最大战略资源——人类基因争夺战的重要"战场"。近年，*Nature*、*Science* 和 *Cell* 以及其他一些重要杂志接连刊登有关蛋白质组学的评论文章或原始论著，明确表明了蛋白质组学已经成为新世纪生命科学研究的前沿[56~79]。

蛋白质组学的第一篇原始论著 1995 年发表于国际上并不著名的 *Electrophoresis*[1]。如果说蛋白质组学刚诞生时没有立即得到国际生命科学界主流的高度重视，那么近三年已发生巨大的变化。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，其中一千万美元用于在密执安医学院建立一个有关肺、直肠、乳腺和卵巢等肿瘤的蛋白质组数据库；此外，国立肿瘤研究所和美国食品与药物管理局联合，投入数百万美元，资助建立一个有关癌症不同发病阶段和治疗阶段的蛋白质组数据库。美国能源部不久前也启动了蛋白质组项目，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进行蛋白质组研究。在法国，新成立了五个区域性遗传基地（genopoles），它们将得到为期三年的资助，每年约为五百万美元，这些经费将平均用于基因组、转录组（transcriptome）和蛋白质组研究。德国的联邦研究部提供七百多万美元，在东德的 Rostock 建立了一个蛋白质组学中心。1997 年澳大利亚政府即着手建立第一个全国性的蛋白质组研究网 APAF（Australian Proteome Analysis Facility）。APAF 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。日本的科学与技术委员会也已先期由政府出资三百万美元开展蛋白质组研究。由此可见，蛋白质组学虽然问世不到 10 年，但鉴于其战略的重要性和技术的先进性，西方主要发达国家在这一新型领域争先恐后、均已投入巨资全面启动此领域的研究。

更有甚者，由于蛋白质组学研究比基因组学研究更接近实用，具有着巨大的市场前景，企业与制药公司纷纷斥巨资开展蛋白质组研究。如独立完成人类基因组测序的 Celera 公司已宣布投资上亿美元于此领域[78]；又如日内瓦蛋白质组公司与布鲁克质谱仪制造公司联合成立了国际上最大的蛋白质组研究中心。不难看到，蛋白质组学已成为西方各主要发达国家、各跨国制药集团竞相投入的"热点"与"焦点"。

蛋白质组学的前沿大致分为三大方向：① 针对有基因组或转录组数据库的生物体或组织/细胞，建立其蛋白质组或亚蛋白质组（或蛋白质表达谱）及其蛋白质组连锁群，即 compositional proteomics 研究；② 以重要生命过程或人类重大疾病为对象，进行重要生理/病理体系或过程的比较蛋白质组学研究，即 compar-

ative proteomics 研究；③ 蛋白质组学支撑技术平台和生物信息学的研究。

国外大部分蛋白质组表达谱的研究论文发表于 2000 年下半年以后，且大多建立在已完成基因表达谱的基础上，表明目前在基因组或转录组基础上开展蛋白质组表达谱的研究是一个新的方向[79]。人类重大疾病的蛋白质组研究通常采用比较蛋白质组分析方法。近年来，蛋白质组学技术在研究细胞的增殖、分化、异常转化、肿瘤形成等方面进行了有力的探索，涉及到白血病、乳腺癌、结肠癌、膀胱癌、前列腺癌、肺癌、肾癌和神经母细胞瘤等，鉴定了一批肿瘤相关蛋白，为肿瘤的早期诊断、药靶的发现、疗效判断和预后提供了重要依据[80]。高通量、高灵敏度和规模化的双向凝胶电泳-质谱是目前最流行和较可靠的技术平台[81,82]；酵母双杂交技术已被用于研究蛋白质连锁群和蛋白质功能网络系统，且分别发表于 *Nature* 与 *Science*[83,84]。生物信息学方法在蛋白质组学研究领域亦得到有效的利用，其中突出的代表是 Eisenberg D 等联合采用 phylogenetic profiles 法、Rosetta stone 法和 gene neighbour 法，成功地建立了酵母 SIR（silencing information regulator）作用网络和酵母 Prion 功能连锁网络[85]。

存在的问题

方法学上，二维凝胶电泳-质谱仍是目前最流行和较可靠的技术平台，但其通量、灵敏度和规模化均有待进一步加强。二维凝胶电泳有分离容量的先天限制，染色转移等环节操作困难费时，低丰度蛋白难以辨别，和质谱技术的联用已成为瓶颈。因此国际上开始重视研究以色谱/电泳-质谱为主的技术平台。另一方面，酵母双杂交技术虽已被用于研究蛋白质连锁群和蛋白质功能网络系统，但仍缺乏快速、高效的手段获取复杂蛋白质相互作用的多维信息。蛋白质组的生物信息学研究，虽然已有成功的先例，但其应用范围与准确率仍需提高，所面临的更大挑战是如何进行信息综合，准确分析蛋白质的相互作用，界定相互作用连锁群。

学术上，在基因组、转录组基础上的蛋白质组全谱研究，微生物已有成功的报道，但是在高等生物尤其是哺乳动物中未见报道，人类组织或细胞的蛋白质组全谱研究则基本未涉及。而由于物种演化中进化上的差别，人类基因组、转录组、蛋白质组的全景式比较对于不同层次上人类基因表达调控规律的认识必不可少；此外，人类基因组草图虽已公布，但是所估计的 3.5 万左右基因中一半以上纯属理论推测，需要从蛋白质组水平予以检验与确认，因此开展人类组织或细胞的蛋白质组表达谱的分析势在必行。

受基因调控的细胞内各种信号转导途径之间是相互交错和彼此关联的。虽然近年来人们对转导途径以及相互关系的认识取得了不少进展，但是针对任一生物体或组织/细胞开展全方位的蛋白质组相互作用网络的分析鲜有报道。而此类相互作用网络的揭示对于深刻认识重要生理、病理过程的机制不可缺少。

国内研究现状

在1995年国际上发表第一篇蛋白质组学的研究论文后不久，国家自然科学基金委即酝酿并于1997年设立了重大项目“蛋白质组学技术体系的建立”。在此前后，中国科学院生物化学研究所、军事医学科学院与湖南师范大学迅速启动了蛋白质组研究，建立并分别组合了二维电泳蛋白质组分离技术、2D-PAGE图像分析技术和蛋白质鉴定的质谱技术；先后举办了三次全国性的蛋白质组学术研讨会，并在国际上较早提出了功能蛋白质组学的研究战略。

经过几年努力，中国科学院上海生命科学研究院、军事医学科学院与复旦大学相继成立了专门的蛋白质组学研究中心。整体上，我国蛋白质组学技术平台的建设有了飞跃的发展，若干研究单位重点建立了技术平台，并在方法学的跟踪与创新上做了不少工作，使现有技术平台已经达到国际先进水平。我国科学家已经在蛋白质组分析技术与方法、在重大疾病如肝癌、维甲酸诱导白血病细胞凋亡启动模型及维甲酸定向诱导胚胎干细胞向神经系统分化的模型等的比较蛋白质组研究以及一些重要生理和病理体系的蛋白质组成分研究方面获得了不少成就，并在国际蛋白质组学核心刊物上发表了系列高水平的论文[86~92]。对不断涌现的新技术，如衍生出的多种分离和鉴定模式，我国已经进行了很好的跟踪和发展，在某些方面孕育着新的突破[93~105]。

我国的“疾病基因组学”研究已取得明显的成就，在神经系统遗传病致病基因、肝癌、心血管疾病、白血病等相关基因、造血干/祖细胞、下丘脑-垂体-肾上腺轴系统、海马体、胎肝、心血管和神经系统等组织或细胞的基因表达谱（即转录组）方面均做出了与国际前沿水平相当的工作[106~117]，且有我国的特色与优势。所取得的丰富数据将直接成为开展对应蛋白质组学研究的基础与出发点。我国在重大疾病的功能蛋白质组学研究方面也取到了良好的起步，已进行肝癌细胞系及正常肝细胞蛋白质组的比较分析研究，发现了两者间不同的蛋白表达群[90,91]。此外，还进行了我国自行建立的肝癌高/低转移细胞系、肺癌高/低转移细胞系、原位食管癌/转移食管癌间的比较蛋白质组研究，初步发现了一批与肿瘤转移相关的蛋白质群。同时，心肌细胞与应激损伤的心肌细胞的比较蛋白质组研究也已有一定的基础。上述研究一方面证明我国已有的蛋白质组学技术平台已能支撑一定规模的研究，另一方面亦为我国在国际蛋白质组学研究领域争得了一席之地，同时为未来的发展奠定了良好的基础。

五、蛋白质组研究的发展趋势及我国的应对策略

蛋白质组研究的整体状况：相关技术已基本配套且基本达到实用化水平；已形成一定规模的专业队伍及专业机构，因此，其草创阶段已结束，发展阶段已开始；蛋白质组研究已成为后基因组时代最重大的生命科学命题之一；蛋白质组分

析已作为专门的技术体系广泛用于生命科学众多领域尤其是热点领域的研究；蛋白质组的基础研究与分析应用正以指数增长方式发展，其对现代生命科学的介入与贡献将可能使生命科学工作者从核酸时代逐渐回归蛋白质时代，使对生命系统与活动的分子机制的认识由间接的基因、核酸层次深入到生命的直接执行体——蛋白质层次，更将蛋白质研究无论是规模还是深度均推进到前所未有的程度。人们不难看到，它是基因组计划由结构走向功能的必然与必需，是生命科学由分析走向综合的必经之路，是连接微观分子系统、运行机制与宏观生物系统、生命活动的桥梁。因此，人们不难预期，蛋白质组研究将成为21世纪生命科学的重要支柱之一，其发展未可限量。

人们在憧憬蛋白质组研究美好未来的同时，也应看到“蛋白质组”作为“新生”领域在许多方面还很稚嫩，如2-DE的灵敏度虽然已达飞摩（fmol）级水平，但仍难将细胞组织内多种痕量调控蛋白分离显示出来，而此类蛋白对于基础与应用研究都极为重要（甚至比高含量结构蛋白更为重要），此方面仍需进一步改进；此外，现有质谱技术虽然在蛋白质组成分的鉴定中高效、灵敏、特异，但所用仪器价格十分昂贵，十倍于DNA自动测序仪，国内外只有极少数单位有能力购置，因而其推广受到很大的限制，此方面技术与仪器如不改善，将使蛋白质组研究与进展神速的基因组计划严重脱节；此外，蛋白质组研究不能局限于对已完成基因组计划的理论预测的蛋白质组进行实证分析，还应开辟“战场”，一方面对未完成或根本未进行基因组计划的重要生物进行前瞻性蛋白质组研究，以推动其基因组研究，另一方面大力加强重要生命活动中比较蛋白质组研究和重要组织的蛋白质组研究，后一方面将会有更大作为。

蛋白质组研究在国际上正如火如荼、轰轰烈烈，我国则刚启动。鉴于蛋白质组研究代表着生命科学研究中一个新的“制高点”，我国目前的状况应引起我们的高度警觉。虽然国家自然科学基金委已将“蛋白质组研究”作为重大项目立项，国家科技部亦将其列为国家基础研究重大项目，但相对国际上的大力突入与强大的竞争而言，这些只能是“九牛一毛”。21世纪生命科学在整个自然科学中的主导作用，我们注定要面对；蛋白质组研究在后基因组时代的支柱作用我们也必定要面对。我们与其到时“亡羊补牢”，不如现在就奋起直追。为此，综合考虑国际基因组与蛋白质组研究的现状与趋势以及我国基因组研究与蛋白质组研究相关技术的基础，作者认为我国的蛋白组研究不能重复或追随国际已有的工作，而应与我国现行的基因组研究及其他有我国特色或优势的生命科学研究紧密结合（但决不能仅仅融入这两大方面的项目中），走出自己的路。因此，建议开展如下工作：

1）*重大生命活动中蛋白质组的比较研究*　选取1～2类重大生命活动（如重大疾病）中几个相继的重要阶段，分别进行蛋白质组作图，进行系统的定性、定量比较，进而对差别蛋白进行鉴定，然后从核酸、蛋白两层

次对差别蛋白进行性能分析，从而确定重大生命活动的蛋白质组基础。

2）1～2种组织或细胞蛋白质组的系统研究　选择1～2种具有重要生物学意义或性状、且我国已完成cDNA大规模测序的组织或细胞，制备其高分辨率蛋白质组图并建立其蛋白质组图数据库，进而联合应用其cDNA大规模测序的数据，规模化研究其蛋白质组成。

3）蛋白质组的生物信息学研究　蛋白质组成员的序列、结构、功能及定位分类；基于生化途径、遗传网络等，构建蛋白质组功能系统即蛋白连锁图；建立人或其他哺乳动物蛋白质组数据库；高等生物基因组中蛋白质编码基因的识别及算法研究；基因翻译产物的结构、功能预测；基于蛋白质数据库与知识库的知识与规律发现。

4）蛋白质组分析的支撑技术研究　新型蛋白质结构、功能预测方法及程序；大规模蛋白质相互作用分析技术；蛋白质分析鉴定中新型质谱技术的发展及应用；基于蛋白质序列、结构及蛋白质组数据库的知识与规律发现理论与方法；HTS（high throughout system）系统及蛋白质组分析自动操作系统（蛋白质组分析机器人）等。

总之，我国在蛋白质化学与蛋白质科学（protein science）领域曾取得举世瞩目的成就，近年在蛋白质组学这一新兴领域已有了良好的基础，一批研究结果已在国际蛋白组学的核心期刊发表，技术平台的部分指标达到国际先进水平。但是，现有的规模与层次难以提供我国医药卫生事业与生物技术产业迅猛发展所急需的、强有力的蛋白质组学学术与技术支撑，难以适应我国基因组学等现代生命科学前沿领域对蛋白质组学的广泛需求，难以应对国际在蛋白质组学这一战略“高地”的激烈竞争。因此，国家层面上对蛋白质组学的部署不容迟缓。

参 考 文 献

1 Wasinger V C, Humphery Smith I, Williams K L, *et al*. Progress with gene product mapping of the molicutes: *Mycoplasma genitalium*. Electrophoresis, 1995, 16: 1090～1094

2 Kahn P. From genome to proteome: Looking at a cell's proteins. Science, 1995, 270: 369～370

3 Swinbanks D. Government backs proteome proposal. Nature, 1995, 386: 653

4 Nowak R. Entering the postgenome era. Science, 1995, 270

5 Chait B T. Trawling for proteins in the post-genome era. Nature Biotech, 1996, 14: 1544

6 Murray J C, Buetow K H, Weber J L, *et al*. A comprehensive human map with centimorgan density. Science, 1994, 256: 2049～2054

7 Hudson T J, Stein L D, Gerety S S, *et al*. An STS based map of the human genome. Science, 1995, 270: 1945～1954

8 Chumakow I M, Rigoult P, Le Gall I, *et al*. A YAC contig map of the human genome. Nature, 1995, 377 (Suppl): 175～297

9 Venter J V, Smith H, Hood L. A new strategy for genome sequencing. Nature, 1996, 318: 364～366

10 Fleischmann R D, Adams M D, White O, *et al*. Whole genome random sequencing and assembly of *Haemophilis influenzae* Rd. Science, 1995, 269: 496～512

11 Fraser C M, Gocayne J K, Whit O, *et al*. The minimal gene complement of Mycoplasma genitalium. Science, 1995, 270: 297～403

12 Blattmer F B, G Plunkett Ⅲ, Bloch C A, *et al*. The complete genome sequence of E. coli K 12. Science, 1997, 277: 1453～1462

13 Goffeau, A, Barrel B G, Bussey H, *et al*. Life with 6000 genes. Science, 1996, 274: 546～567

14 Mewes H W, Albermann K, Bahr M, *et al*. Overview of the yeast genome. Nature, 1997, 387 (suppl): 7～65

15 Wilson R, Ainscough R, Anderson K, *et al*. 2.2 Mb of contiguous nucleotide sequence from chromosome Ⅲ of *C. elegans*. Nature, 1994, 368: 32～38

16 Wodicka L, Dong H, Mittmann M, *et al*. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nature Biotech, 1997, 15: 1359～1367

17 Ramsay G. DNA chips: State-of-the-art. Nature Botech, 1998, 16: 40～44

18 Velculescu F B, Zhang L, Vogelstein B, *et al*. Serial analysis of gene expression. Science, 1995, 270: 484～4872

19 Perler F B, Olsen G J, Adam E. Compilation and analysis of intein sequences. Nucleic Acid Res, 1997, 25: 1087～1093

20 Alberts B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. Cell, 1998, 92: 291～294

21 O'Farrell P H. High resolution two dimensional electrophoresis of proteins. J Biol Chem, 1975, 250: 4007～4021

22 Strupat K, Karas M, Hillenkamp F, *et al*. Matrix-assisted laser desorption ionization mass spectrometry of proteins electroslotted after PAGE. Anal Chem, 1994, 66: 464～470

23 Mann, M, Hojrup P, Roepstorff P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol Mass Spectrometry, 1993, 22: 238～245

24 Wilkins M R, Sanchez J C, Gooley A A, *et al*. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotech & Genetic Engineering Rev, 1995, 13: 19～50

25 Geisow M. Proteomics. One small step for a digital computer, one giant leap for humankind. Nature Biotech, 1998, 16: 206

26 Humphery Smith I, Cordwell S J, Blackstock W P. Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. Electrophoresis, 1997, 18: 1217～1242

27 Shevchenko A, Mann M, Boucherie H, *et al*. Linking genome and proteome by mass spectrometry: large scale identification of yeast proteins from two dimensional gels. PNAS, 1996, 93: 14440～14445

28 Celis J E, Rasmussen H H, Vanderckhove J, *et al*. Human 2-D PAGE databases for proteome analysis in health and disease: http://biobose. dik/egibin/celis.FEBS Lett, 1996, 398: 129～134

29 Bin L, Zwilling R, Pallin V, *et al*. Two dimensional gel electrophoresis of *Caenorhabdilis elegans* homogenates and identification of protein spots by microsequencing. Electrophoresis, 1997, 18: 557～562

30 Langen H, Fountoulakis M, Takacs B, *et al*. From genome to proteome: protein map of *Haemophinlus influenzae*. Electrophoresis, 1997, 18: 1184～1192

31 Cordwell S J, Humphery Smith I, Shaw D C, *et al*. Characterization of basic proteins from *spiiroplasma melliforum* using novel immobilized pH gradients. Electrophoresis, 1997, 18: 1393～1398

32 Yan J X, Williams K L, Hochstrasser O F, *et al*. The *Dictyostelium discoidenm* proteome the SWISS-2D

PAGE database of the multicellular aggregate (Slug). Electrophoresis, 1997, 18: 491～497

33 O'Connor C D, Qi S Y, Fowler R, *et al*. The proteome of *Salmonella* enteria serovar typhimurium: current progress on its determination and some applications. Electrophoresis, 1997, 18: 1483～1490

34 Guerreiro N, Djordievic M A, Rolfe B G, *et al*. New *Rhizobium leguminosarum* flavonoid induced proteins revealed by proteome analysis of differentially displayad proteins. Mol Plant Microbe Interact, 1997, 10: 506～516

35 Link A J, Yates J R 3rd, Carmack E B, *et al*. Identifying the major proteome components of *Haemophilus influenzae* type strain NCTC 8143. Electrophoresis, 1997, 18: 1314～1334

36 Link A J, Church G M, Robison K. Comparing the predicted and observed properties of proteins encoded in the genome of *Escheriichia coli* K-12. Electrophoresis, 1997, 18: 1259～1313

37 Garrels J I, Payne W E, Mesquita Fuentes R, *et al*. Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. Electrophoresis, 1997, 18: 1347～1360

38 Sazuka T, Ohara O. Towards a proteome project of *cyanobacteriium Syoechocysitis* sp strain PCC6803: linking 130 protein spots with their respective genes. Electrophoresis, 1997, 18: 1252～1258

39 Van Bogelen R A. Neidhardt F C, Olson E R, *et al*. *Escherichia coli* proteome analysis using the gene-protein database. Electrophoresis, 1997, 18: 1243～1251

40 Dainese P, James P, Kertesz M, *et al*. Probing protein function using a combination of gene knockout and proteome analysis by mass spectrometry. Electrophoresis, 1997, 18: 832～842

41 Blomberg A. Osmoresponsive proteins and functional assessment strategies in *Saccharomyces cerevisiae*. Electrophoresis, 1997, 18: 429～440

42 Ostergaard M, Celis J E, Wolf H, *et al*. Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. Cancer Res, 1997, 57: 4111～4117

43 Wimmer K, Hanash S M, Thoraval D, *et al*. Two-dimensional separations of the genome and proteome of neuroblastoma cells. Electrophoresis, 1996, 17: 1741～1751

44 Urquhart B L, Haumphery-Smith I, Brithon W L, *et al*. "Proteomic contig" of *Mycobacterium tuberculosis* and *mycobacterium bovis* (bcg) using novel immobilized pH gradients. Electrophoresis, 1997, 18: 1384～1392

45 Corduell S J, Humphery-Smith I, Basseal D J. Proteome analysis of *Spiroplasma melliferum* (A56) and protein characterization across species boundaries. Electrophoresis, 1997, 18: 1335～1346

46 Cordwell S J, Humphery-Smith I. Evaluation of algorithms used for cross-species proteome characterization. Electrophoresis, 1997, 18: 1410～1417

47 Wasinger V C, Humphery-Smith I, Bjellqist B. "Proteomic contig" of *Ochroboctrum anthropi*, application of extensive pH gradients. Electrophoresis, 1997, 18: 1373～1383

48 Payne W E, Garrels J I. Yeast protein database (YPD): a database for the complete proteome of *Saccharomyces cerevisiae*. Nucleic Acids Res, 1997, 25: 57～62

49 Fromont-Racine M, Legrain P, Rain J C. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature Genetics, 1997, 16: 277～282

50 Lecenier N, Foury F, Goffeau A. Two-hybrid systematic screening of the yeast proteome. Bio-Essays, 1996, 20: 1～6

51 Packer N H, Wiilliams K L, Redmond J W, *et al*. Proteome analysis of glycoforms: a review of strategies for the microcharacterisation of glycoproteins separated by two-dimensional PAGE. Electrophoresis, 1997, 18: 452～460

52 Qi S Y, O'Connor C D, Moir A. Proteome of *Salmonella typhimurium* SL1344: identification of novel

abundant cell envelope proteins and assignment to a two-dimensional reference map. J Bactoriol, 1996, 178: 5032～5038

53 J X, Hochstrasser D F, Pasqual C, *et al*. Large scale amino-acid analysis for proteome studies. J Chromatogr, 1996, 736: 291～302

54 Abbott A. And now for the proteome. Nature, 2001, 409: 747

55 Fields S. Proteomics in genomeland. Science, 2001, 291: 1221

56 Yuen Ho, *et al*. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature, 2002, 415: 180～183

57 Akhilesh Pandey, Matthias Man. Proteomics to study genes and genomes. Nature, 2000, 405: 837～846

58 Potter Wickware, Paul Smaglik. Proteomics to study genes and genomes. Nature, 2001, 413: 869～875

59 Anton J, Ioannis Iliopoulos, Nikos C K, Christos A O. Protein interaction maps for complete genomes based on gene fusion events. Nature, 1999, 402: 86～90

60 Anne-Claude Gavin, *et al*. Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002, 415: 141～147

61 Service, Robert F. PROTEOMICS: A Proteomics Upstart Tries to Outrun the Competition. Science, 2001, 294: 2079～2080

62 Service, Robert F. PROTEOMICS: Proteomics 2.0: The View Ahead. Science, 2001, 294: 2076

63 Ideker, Trey, Thorsson, Vesteinn, Ranish, Jeffrey A., Christmas, Rowan, Buhler, Jeremy, Eng, Jimmy K., Bumgarner, Roger, Goodlett, David R., Aebersold, Ruedi, Hood, Leroy. Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. Science, 2001, 292: 929～934

64 Gerstein, Mark, Lan, Ning, Jansen, Ronald. PROTEOMICS: Enhanced: Integrating Interactions. Science, 2002, 295: 284～287

65 Marshall, Eliot. PROTEOMICS: A Physicist-Turned-Biologist. Science 2001, 294: 2085

66 Service, Robert F. PROTEOMICS: Gene and Protein Patents Get Ready to Go Head to Head. Science, 2001, 294: 2082～2083

67 Service, Robert F. PROTEOMICS: Searching for Recipes for Protein Chips. Science, 2001, 294: 2080～2082

68 Service, Robert F. PROTEOMICS: High-Speed Biologists Search for Gold in Proteins. Science, 2001, 294: 2074～2077

69 Miène Fauchon, Gilles Lagniel, Jean-Christophe Aude, Luis Lombardia, Pascal Soularue, Cyrille Petat, Cérard Marguerie, André Sentenac, Michel Werner, and Jean Labarre. Sulfur Sparing in the Yeast Proteome in Response to Sulfur Demand. Molecular Cell, 2002, 9: 713～723

70 Mary B D, Robert F O. The search for predictive patterns in ovarian cancer: Proteomics meets bioinformatics. Cancer Cell, 2002, 1: 111～112

71 Archa H. Fox, Yun Wah Lam, Anthony K. L. Leung, Carol E. Lyon, Jens Andersen, Matthias Mann, and Angus I. Lamond. Paraspeckles: A Novel Nuclear Domain. Current Biology, 2002, 12: 13～25

72 Jens S. Andersen, Carol E. Lyon, Archa H. Fox, Anthony K. L. Leung, Yun Wah Lam, Hanno Steen, Matthias Mann, and Angus I. Lamond. Directed Proteomic Analysis of the Human Nucleolus. Current Biology, 2002, 12: 1～11

73 Jason K. Ospina and A. Gregory Matera. Proteomics: The Nucleolus Weighs In. Current Biology, 2002, 12: R29～R31

74 Miroslav Dundr and Tom Misteli. Nucleolomics: An Inventory of the Nucleolus. Molecular Cell, 2002, 9: 5～7

75 Piyanun Harnpicharnchai, Jelena Jakovljevic, Edward Horsey, Tiffany Miles, Judibelle Roman, Michael Rout, Denise Meagher, Brian Imai, Yurong Guo, Cynthia J. Brame, Jeffrey Shabanowitz, Donald F. Hunt, and John L. Woolford Jr. Composition and Functional Characterization of Yeast 66S Ribosome Assembly Intermediates. Molecular Cell, 2001, 8: 505～515

76 Mark R. Bray, Stuart Bisland, Subodini Perampalam, Wai-May Lim, and Jean Garè py. Probing the surface of eukaryotic cells using combinatorial toxin libraries. Current Biology, 2001, 11: 697～701

77 Timothy S. Lewis, John B. Hunt, Lauren D. Aveline, Karen R. Jonscher, Donna F. Louie, Jennifer M. Yeh, Theresa S. Nahreini, Katheryn A. Resing, and Natalie G. Ahn. Identification of Novel MAP Kinase Pathway Signaling Targets by Functional Proteomics and Mass Spectrometry. Molecular Cell, 2000, 6: 1343～1354

78 Robert F. Proteomics, can Celera do it again? Science, 2000, 287: 2136

79 Haynes PA. Proteome profiling-pitfalls and progress. Yeast, 2000, 17: 81

80 Seow T K, *et al*. Two-dimensional electrophoresis map of the human hepatocellular carcinoma cell line, HCC-M, and identification of the separated proteins by mass spectrometry. Electrophoresis 2000, 21: 1787

81 Bakhtiar R, *et al*. ESI-MS and MALDI-MS——Emerging technologies in biomedical science. Biochem Pharmacol, 2000, 8: 891

82 Szallasi Z. Gene expression patterns and cancer. Nature Biotech, 1998, 16: 1292

83 Albertha J M, *et al*. Protein interaction mapping in *C. elegens* using protins involved in vulval development. Science, 2000, 287: 116

84 Houry W A, *et al*. Identification of in vivo substrates of the chaperonin. Nature, 1999, 402: 147

85 Eisenberg D, *et al*. Protein function in the post-genomic era. Nature, 2000, 405: 823

86 Xia Q C, *et al*., Application of free-flow electrophoresis to the purification of trichosanthin from a crude product of acetone fractional precipitation. Electrophoresis, 1998, 19: 1097

87 Xia Q C, *et al*. Determination of sialic acid content in glycoproteins by capillary zone electrophoresis. Electrophoresis, 1999, 20: 2930

88 Wan J H, *et al*. Proteomic analysis of apoptosis initiation induced by all-trans retinoic acid in human acute promyelocytic leukemia cells. Electrophoresis, 2001, 21: 3026

89 Guo X X, *et al*. Proteomic Characterization of Early Stage Differentiation of Mouse Embryonic stem Cells into Neural Cells Induced by Retinoic Acid in vitro. Electrophoresis, 2001, 21: 3067

90 Yu L R, Xia Q C, *et al*. Identification of differentially expressed proteins between human hepatoma and normal liver cell lines by two-dimensional electrophoresis and liquid chromatography-ion trap mass spectrometry. Electrophoresis, 2000, 21: 3058

91 Yu L R, Xia Q C, *et al*. Proteome altration in human hepatoma cells transfected with antisense EGF receptor sequence. Electrophoresis, 2000, 22: 3001

92 Zou H F, Zhang Y K, *et al*. Quantitative Study of Competitive Binding of Drugs to Protein by Microdialysis/HPLC. J Chromatogrphy A, 1999, 849: 599～608

93 Zou H F, Zhang Y K, *et al*. Comparative study on the distribution of ovalbumin glycoforms by capillary electrophoresis. Anal Chem, 1998, 70: 373

94 Zou H F, Zhang Y K, *et al*. Study of Physically Adsorbed Stationary Phase for Open-Tubular Capillary Electrochromatography. Electrophoresis, 1999, 20: 2891

95 Liang S P, *et al*. Identification of venom proteins of spider Shuwena on two-dimensional electrophoresis gel by N-terminal microsequencing andmass spectrometric peptide mapping. J Protein Chem, 2000, 3: 225～230

96 Liang S P, *et al*. Analysis of recombinant and modified proteins by capillary zone electrophoresis/electrospray ionization-tradem mass spectrometry. J Chomatography A, 1999, 855: 695

97 Zou H F, Zhang Y K, *et al*. Capillary Electrochromatography Using a Strong Cation-Exchange Column with a Dynamically Modified Cationic Surfactant. Anal Chem, 2000, 72: 616

98 Zou H F, Zhang Y K, *et al*. On-Column Enrichment In Capillary Electrochromatography. Anal Chem, 2000, 72: 23

99 Yang P Y, *et al*. Pulsed-electrospray mass spectrometry. Anal Chem, 2001, 73: 4748

100 Zhang X M, *et al*. Comprehensive Two-Dimensional Capillary LC and CE for Resolution of Neutral Components in Traditional Chinese Medicines. J Sep Sci, 2001, 24: 10

101 Zhang X M, *et al*. Single Step On-Column Frit Making For Capillary High Performance Liquid Chromatography Using Sol-Gel Technology. J Chromatogr A, 2001, 910: 13

102 Kong J L, *et al*. Direct Electrochemistry of Co-factor Redox Sites in a Bacterial Photosynthetic Reaction Center Protein. J Am Chem Soc, 1998, 120: 7371

103 Huo K K, *et al*. Two DNA replication origin identified in the K1 killer toxin gama subunit gene promoter. Current Genetics, 1999, 35: 335

104 Huo K K, *et al*. Protein disulfide isomerase genes of *Kluyveromyces lactis*. Yeast, 2000, 16: 329

105 Hu R M, Han Z G, Song H D, *et al*. Gene expression profiling in the human hypothalamus-pituitary-adrenal axis and full-length cDNA cloning. Proc Natl Acad Sci USA, 2000, 97: 9543～9548

106 Yu Y T, Zhang C G, Zhou G Q, Wu S F, Qu X H, Wei H D, Xing G C, Zhai Y, Wan J H, Ouyang S G, Li L, Zhang S W, Wu C T, He F C. Gene Expression Profiling in Human Fetal Liver and Identification of Tissue- and Developmental Stage-specific Genes through Compiled Expression Profiles and Efficient Cloning of Full-Length cDNAs. Genome Res, 2001, 11 (8): 1392

107 Mao M, *et al*. Identification of genes expressed in human CD34+ hematopoietic stem/protgenitor cells by expressed sequence tags and efficient full-length cDNA cloning. Proc Natl Acad Sci USA, 1998, 95: 8175

108 Huo K K *et al*. Isolation of a novel candidate oncogene within a frequently amplified region at 3q26 in ovarian cancer. Cancer Research, 2001, 61: 3806

109 Li Y, *et al*. Stimulation of the mitogen-activated protein kinase cascade and tyrosine phosphorylation of the epidermal growth factor receptor by hepatopoietin. J Biol Chem, 2000, 275: 37443

110 Wang G, *et al*. Identification and characterization of receptor for the mammalian hepatopoietin that is homologous to yeast ERV1. J Biol Chem, 1999, 274: 11469

111 Liu X Q, *et al*. Characterization of ARF GAP family and its first member derived from human being. FEBS LETT, 2001, 490: 79～83

112Zhang C G, *et al*. Characterization, Chromosomal Assignment, and Tissue Expression of a Novel Human Gene Belong to ARF GAP Family. Genomics, 2000, 63 (3): 400

113 Qu X H, *et al*. Characterization and Tissue Expression of a Novel Human Orthologue of Mouse npdc1. Gene, 2001, 264: 37

114 Zhu F X, *et al*. cDNA transfection of amino-terminal fragment of urokinase efficiently inhibits cancer cell invasion and metastasis. DNA & Cell Biology, 2001, 20: 297

115 Luo T H, *et al*. A genome-wide search for type Ⅱ diabetes susceptibility genes in Chinese Hans. Diabetologia, 2001, 44: 501

116 Tang J G, *et al*. Anti-autolysis of trypsin by modification of autolytic site Arg117. Biochem Biophys Res Commun, 1998, 250: 235

117 Tang J G, *et al*. Separation and characterization of trypsin and carboxypeptidase B digested products of Met-Lys-human proinsulin. Applied Biochemistry and Biotechnology, 1999, 76: 107

（贺福初）

第二章　蛋白质组研究方法

第一节　概　　述

蛋白质组的研究远比基因组研究复杂。一方面，蛋白质的数目远大于基因的数目，这是由于基因的拼接和翻译后的修饰造成的。人的基因组有 25 000～40 000个编码基因[1]，其表达的蛋白质可以达十几万甚至更多。另一方面，基因是相对静态的，一种生物体仅有一个基因组，而蛋白质是动态的，随时间、空间的变化而变化。一个细胞中的蛋白质可多达上万种，而它们的拷贝数可能相差几百倍到几十万倍。组成蛋白质的氨基酸有 20 种，加上修饰的氨基酸就更多，而 DNA 仅由 4 种核苷酸组成。从技术手段上讲，人们无法采用在基因研究中所普遍采用的 PCR 技术来使微量的蛋白质得到扩增，至今尚没有一种蛋白质的测序技术可与在基因组研究中起关键作用的自动化的 DNA 测序技术相媲美。DNA 微阵列技术的发展与应用，使得基因的筛选实现了高通量，而目前的蛋白质分析技术离工业规模的高通量还有较大差距（表 2-1）。

所幸的是，尽管核酸研究的技术近年来取得了突飞猛进的发展，蛋白质的研究技术并未停滞不前。由 O'Farrel 等在 1975 年建立的双向电泳技术（two dimensional electrophoresis，2-DE）可同时分离数千种蛋白[2]，20 世纪 80 年代固相化 pH 梯度凝胶的引进使得双向电泳的重复性和加样量得到巨大的改善，从而使今天的蛋白质组研究得以实施[3]。以计算机技术为基础的多种图像分析与大规模数据处理软件的问世，使得科学家们处理复杂的类似“满天星”样的蛋白质图谱并建立相应的数据库得心应手。80 年代后期出现的两种软电离质谱技术-基质辅助激光解吸电离飞行时间质谱（matrix assisted laser desorption ionization time of flight mass spectrometry，MALDI-TOF-MS）[4]与电喷雾电离质谱（electro-spray ionization mass spectrometry，ESI-MS）[5]，可精确测定生物大分子的相对分子质量及多肽序列，使得微量快速的蛋白质鉴定得以实现。回想 10 年以前，一个蛋白质化学家一年鉴定 2～3 个蛋白质，而现在，蛋白质组研究技术加上基因组提供的信息使得一个科学家一个星期可以鉴定几百个蛋白质。

双向电泳技术、计算机图像分析与大规模数据处理技术以及质谱技术被称为蛋白质组研究的三大基本支撑技术。蛋白质组研究的技术路线如图 2-1。从细胞、体液或组织等生物样品中提取的蛋白质，经 2-DE 分离，染色，得到蛋白质表达谱。采用计算机图像分析技术，可对图谱上的蛋白质点进行定位、定量、图谱比