



科学出版社



生命科学前沿丛书

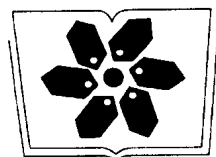
北京生物技术和新医药产业促进中心

北京生物工程学会 合作策划

科学出版社生命科学编辑部

钱小红 贺福初 主编

蛋白质组学： 理论与方法



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>

西源印刷厂 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)

2003年6月第一次印刷 印张: 22

印数: 1—3 000 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〉)

《生命科学前沿丛书》专家委员会

主任委员：吴 旻

委 员：（按汉语拼音排序）

陈永福 陈 竺 范云六 贺福初

黄大昉 李家洋 李衍达 马大龙

强伯勤 沈倍奋 王琳芳

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	(1)
第一节 人类基因组计划及蛋白质组研究的历史背景	(1)
第二节 蛋白质组研究的开端及“蛋白质组”含义	(10)
第二章 蛋白质组研究方法	(24)
第一节 概述	(24)
第二节 大规模的蛋白质分离技术	(27)
第三节 高通量蛋白质鉴定技术	(39)
第三章 二维电泳的蛋白质提取与样品制备	(48)
第一节 细胞裂解的方法	(49)
第二节 蛋白质的分步提取技术	(57)
第三节 亚细胞分离与蛋白质提取	(59)
第四章 二维电泳与细胞蛋白质的分离	(68)
第一节 概述	(68)
第二节 一维等电聚焦电泳	(72)
第三节 二维 SDS-聚丙烯酰胺凝胶电泳	(80)
第四节 胶上蛋白的检测	(81)
第五节 存在问题	(87)
第五章 图像分析与细胞蛋白谱的建立	(91)
第一节 蛋白电泳图像分析系统	(91)
第二节 二维电泳蛋白谱数据库	(97)
第六章 生物质谱技术与蛋白质鉴定	(101)
第一节 基质辅助激光解吸电离飞行时间质谱	(101)
第二节 液相色谱-电喷雾-串联质谱	(105)
第三节 肽质量指纹谱鉴定蛋白质技术	(110)
第四节 串联质谱数据鉴定蛋白质技术	(116)
第七章 蛋白质翻译后修饰的鉴定	(130)
第一节 磷酸化蛋白质的鉴定	(130)
第二节 糖基化蛋白质的鉴定	(145)

第八章 定量蛋白质组学研究技术	(165)
第一节 利用荧光染料进行定量的蛋白质组分析技术	(165)
第二节 运用质谱进行定量的蛋白质组分析技术	(168)
第三节 蛋白质芯片技术	(171)
第九章 蛋白质组研究中的生物信息学	(174)
第一节 生物信息学简介	(174)
第二节 数据库的构建	(180)
第三节 蛋白质组研究中常用的网站及数据库	(188)
第十章 细胞器与蛋白质复合体的组成分析	(203)
第一节 细胞器的分离	(203)
第二节 细胞器的组成分析	(214)
第三节 蛋白复合体分离及组成分析	(217)
第十一章 蛋白质间连锁图的建立	(242)
第一节 酵母双杂交技术及其应用	(242)
第二节 免疫共沉淀技术	(251)
第三节 细胞共定位技术	(257)
第十二章 多能干细胞定向分化的蛋白质组研究	(266)
第十三章 肿瘤发生与发展的比较蛋白组研究	(285)
第一节 蛋白质组与肿瘤	(285)
第二节 肺癌转移的蛋白质组研究	(285)
第三节 辐射致癌相关蛋白质分子标志物的研究	(296)
第十四章 细胞凋亡的蛋白质组研究	(304)
第一节 细胞凋亡的信号传导	(304)
第二节 细胞凋亡的蛋白质组研究	(315)
第十五章 蛋白质组与新药开发	(330)
第一节 药物分子靶标的高通量筛选体系	(332)
第二节 药物的高通量筛选体系	(337)
第三节 药物作用的监测评价体系	(340)

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

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)			
<i>Methanococcus jannaschii</i> DSM 2661	1664	1750	美国伊利诺伊州立大学和 TIGR 公司
<i>Methanobacterium thermoautotrophicum</i> delta H			
<i>Archaeoglobus fulgidus</i> DSM4304	2178	2493	美国伊利诺伊州立大学和 TIGR 公司
<i>Pyrococcus horikoshii</i> (shinkaj) OT3			
<i>Aeropyrum pernix</i> K1	1669	2620	日本 NITE (National Institute of Technology and Evaluation)
<i>Pyrococcus abyssi</i> 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
<i>Halobacterium</i> <i>sp.</i> NRC-1	2014	2058	华盛顿大学和麻省州立大学
<i>Thermoplasma</i> <i>acidophilum</i>	1564	1478	德国 Max Planck 生化研究所和 Medigenomix 公司
<i>Thermoplasma</i> <i>volcanium</i> GSS1	1584	1524	日本 AIST (National Institute of Advanced Industrial Science and Technology)
<i>Sulfolobus</i> <i>solfatarius</i> P2	2992	2977	欧盟和加拿大生物信息资源中心
<i>Sulfolobus</i> <i>tokodaii</i> 7	2694	2826	日本 NITE (National Institute of Technology and Evaluation)
<i>Pyrobaculum</i> <i>aerophilum</i> IM2 细菌 (57 个)	2222	2587	加利福尼亚理工学院 (California Institute of Technology) 和加州大学
<i>Haemophilus</i> <i>influenzae</i> KW20	1830	1850	TIGR 公司
<i>Mycoplasma</i> <i>genitalium</i> G-37	580	468	TIGR 公司
<i>Synechocystis</i> <i>sp.</i> PCC6803	3573	3168	日本 Kazusa DNA Research Institute (KDRI)
<i>Mycoplasma</i> <i>pneumoniae</i> M129	816	677	德国海德堡大学
<i>Escherichia</i> <i>coli</i> K12- MG1655	4639	4289	美国威斯康辛大学
<i>Helicobacter</i> <i>pylori</i> 26695	1667	1590	TIGR 公司
<i>Bacillus</i> <i>subtilis</i> 168	4214	4099	法国 Pasteur 研究所和日本京都大学
<i>Borrelia</i> <i>burgdorferi</i> B31	1230	1256	Brookhaven Natl 实验室和 TIGR 公司
<i>Aquifex</i> <i>aeolicus</i> VF5	1551	1544	美国伊利诺伊州立大学
<i>Mycobacterium</i> <i>tuberculosis</i> H37Rv (lab strain)	4411	4402	英国 Sanger 中心
<i>Treponema</i> <i>pallidum</i> subsp. <i>pallidum</i> 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
<i>Chlamydia trachomatis</i> serovar D	1042	896	美国斯坦福大学和加利福尼亚大学伯克莱分校
<i>Rickettsia prowazekii</i> Madrid E	1111	834	瑞典 Uppsala 大学
<i>Helicobacter pylori</i> J99	1643	1495	基因组治疗公司和美国 Astra 公司
<i>Chlamydia pneumoniae</i> CWL029	1230	1052	美国斯坦福大学和加利福尼亚大学伯克莱分校
<i>Thermotoga maritima</i> MSB8	1860	1877	TIGR 公司
<i>Deinococcus radiodurans</i> R1	3284	3187	TIGR 公司
<i>Ureaplasma urealyticum</i> serovar 3	751	650	Eli Lilly 公司和美国 Alabama 大学、Perkin Elmer 公司
<i>Campylobacter jejuni</i> 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
<i>Chlamydia pneumoniae</i> AR39	1229	1052	英国 Sanger 中心和加拿大 Manitoba 大学
<i>Chlamydia trachomatis</i> MoPn Nigg	1069	924	英国 Sanger 中心和加拿大 Manitoba 大学
<i>Neisseria meningitidis</i> MC58 (serogroup B)	2272	2158	TIGR 公司
<i>Neisseria meningitidis</i> 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
<i>Bacillus halodurans</i> 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
<i>Chlamydia</i> <i>pneumoniae</i> J138	1228	1070	日本 Yamaguchi 大学和 KYUSHU 大学
<i>Xylella</i> <i>fastidiosa</i> CVC 8.1.b clone 9.a. 5.c	2679	2904	巴西 ONSA 中心
<i>Vibrio</i> <i>cholerae</i> serotype O1, Biotype El Tor, strain N16961	4000	3885	TIGR 公司
<i>Pseudomonas</i> <i>aeruginosa</i> PAO1	6264	5570	美国华盛顿大学和 Chiron 公司
<i>Buchnera</i> <i>sp.</i> APS	640	564	日本东京大学和 RIKEN 研究所 (The Institute of Physical and Chemical Research)
<i>Mesorhizobium</i> <i>loti</i> MAFF303099	7596	6752	日本 Kazusa DNA Research Institute (KDR1)
<i>Escherichia</i> <i>coli</i> O157: H7 EDL933	4100	5283	美国威斯康辛大学
<i>Mycobacterium</i> <i>leprae</i> TN	3268	1604	英国 Sanger 中心和法国 Pasteur 研究所
<i>Escherichia</i> <i>coli</i> O157: H7. Sakai	5594	5448	日本京都大学
<i>Pasteurella</i> <i>multocida</i> Pm70	2250	2014	美国明尼苏达州立大学
<i>Caulobacter</i> <i>crescentus</i>	4016	3737	TIGR 公司

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
<i>Streptococcus pyogenes</i> SF370 (M1)	1852	1696	美国俄克拉荷马州立大学
<i>Lactococcus lactis</i> IL1403	2365	2266	法国 Genoscope 中心
<i>Staphylococcus aureus</i> 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
<i>Staphylococcus aureus</i> 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
<i>Mycobacterium tuberculosis</i> CDC 1551	4403	4187	TIGR 公司
<i>Mycoplasma pulmonis</i>	963	782	法国 Genoscope 中心
<i>Streptococcus pneumoniae</i> TIGR4	2160	2094	基因组治疗公司
<i>Sinorhizobium meliloti</i> 1021	6690	6205	欧盟和美国斯坦福大学
<i>Streptococcus pneumoniae</i> R6	2038	2043	Eli Lilly 公司
<i>Rickettsia conorii</i> Malish 7	1268	1374	法国 Genoscope 中心
<i>Yersinia pestis</i> CO-92 Biovar Orientalis	4653	4012	英国 Sanger 中心和 MDS 公司、Dstl 实验室、帝国学院
<i>Salmonella typhi</i> 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
<i>Salmonella</i> <i>typhimurium</i> , LT2 SGSC1412	4857	4597	美国华盛顿大学
<i>Listeria</i> <i>innocua</i> Clip11262, rhamnose-negative	3011	2981	法国 Pasteur 研究所
<i>Listeria</i> <i>monocytogenes</i> EGD-e	2944	2855	法国 Pasteur 研究所
<i>Nostoc</i> <i>sp.</i> PCC 7120	6413	5366	日本 Kazusa DNA Research Institute (KDRI) 和美国密西根州立大学
<i>Agrobacterium</i> <i>tumefaciens</i> C58-Cereon	4915	5299	美国 Monsanto 公司和 Cereon 公司、Rich- mond 大学
<i>Agrobacterium</i> <i>tumefaciens</i> C58-DuPont	4915	5402	美国华盛顿大学和 DuPont 公司、巴西 Campinas 大学
<i>Ralstonia</i> <i>solanacearum</i> GM11000	5810	5120	法国 Genoscope 中心和 INRA 公司、CNRS 公司
<i>Brucella</i> <i>melitensis</i> 16M	3294	3197	美国 Scranton 大学和 Integrated Genomics 公 司
<i>Clostridium</i> <i>perfringens</i> 13	3031	2660	日本 Tsukuba 大学和 Kitasato 大学、Kyushu 大学
真核生物 (9 个)			
<i>Saccharomyces</i> <i>cerevisiae</i> S288C	12 069	6294	国际合作
<i>Caenorhabditis</i> <i>elegans</i>	97 000	19 099	美国华盛顿大学和英国 Sanger 中心
<i>Drosophila</i> <i>melanogaster</i>	137 000	14 100	美国 Celera 公司和加大伯克莱分校 DGP (<i>Drosophila</i> Genome Project)、Baylor Col- lege of Medicine、欧洲 DGP

续表

生物种类	基因组大小/kb	ORF 数目	测序机构
<i>Arabidopsis thaliana</i>	115 428	25 498	国际合作
<i>Homo sapiens</i>	约 3 000 000	35 000~50 000	国际合作
<i>Guillardia theta</i>	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
<i>Leishmania major</i> Friedlin Chromosome 1	257	79	美国 Seattle Biomedical Research Institute (SBRI)
<i>Plasmodium falciparum</i> 3D7 Chromosome 2	947	205	TIGR 公司
<i>Plasmodium falciparum</i> 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'Farrell 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]。