

|生|命|科|学|前|沿|

蛋白质组学研究

—— 概念、技术及应用

〔澳大利亚〕 M.R. 威尔金斯

〔瑞士〕 R.D. 阿佩尔

〔澳大利亚〕 K.L. 威廉斯

〔瑞士〕 D.F. 霍赫斯特拉塞尔

著

张丽华 梁 振 张玉奎 等 译

(原书第二版)

PROTEOME RESEARCH
Concepts, Technology and Application
(Second Edition)



科学出版社
www.sciencep.com

生命科学前沿

蛋白质组学研究——概念、技术及应用

(原书第二版)

Proteome Research **Concepts, Technology and Application** (Second Edition)

[澳大利亚] M. R. 威尔金斯

[瑞士] R. D. 阿佩尔

[澳大利亚] K. L. 威廉斯

[瑞士] D. F. 霍赫斯特拉塞尔

著

张丽华 梁 振 张玉奎等 译

科学出版社

北 京

图字: 01-2008-2166

内 容 简 介

在蛋白质组学水平上,对生命活动的功能执行体——蛋白质进行深入研究,不仅有助于全景式地揭示生命活动的本质,而且对于研究疾病机制、发展预警、诊断和治疗方法均具有重要意义。本书在综述蛋白质组学近10年的发展历程基础上,探讨了蛋白质组在样品预处理、质谱鉴定、定量分析、成像分析、数据处理、相互作用等方面的技术发展,并介绍了蛋白质组学在生物、医学领域的应用。此外,还对蛋白质组学未来的发展方向进行了展望。

本书可供生物、医学、化学、药学等领域的研究人员参考,也可作为高等院校相关专业本科生和研究生的参考书。

Translation from the English language edition;

Proteome Research edited by Marc R. Wilkins, Ron D. Appel, Keith L. Williams, and Denis F. Hochstrasser

Copyright ©Springer-Verlag Berlin Heidelberg 2007

Springer is a part of Springer Science+Business Media

All Rights Reserved

图书在版编目(CIP)数据

蛋白质组学研究:概念、技术及应用/(澳)威尔金斯(Wilkins, M. R.)等著;张丽华,梁振,张玉奎等译. —北京:科学出版社,2010
(生命科学前沿)

Proteome Research: Concepts, Technology and Application, Second Edition
ISBN 978-7-03-027218-8

I. ①蛋… II. ①威…②张… III. ①蛋白质-基因-研究 IV. ①Q51

中国版本图书馆CIP数据核字(2010)第065139号

责任编辑:罗 静 席 慧/责任校对:赵桂芬

责任印制:钱玉芬/封面设计:耕者设计工作室

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>

盛 大 印 刷 厂 印刷

科学出版社发行 各地新华书店经销

*

2010年5月第 一 版 开本:787×1092 1/16

2010年5月第一次印刷 印张:11 1/2 插页:4

印数:1—3 000 字数:249 000

定价:48.00元

(如有印装质量问题,我社负责调换)

译者序

随着人类基因组全序列测定的完成，人类基因的注释与功能确认已成为生命科学面临的最重要任务之一。在蛋白质组学水平上，对生命活动的功能执行体——蛋白质进行深入系统的研究，不仅有助于全景式地揭示生命活动的本质，而且对于研究疾病机制、发展预警、诊断和治疗方法均具有重要意义。因此，基于组学技术发现具有重要生物学意义的蛋白质已成为 21 世纪最重要的研究领域之一，也是国际生物科技关注的战略制高点和竞争焦点。

自 20 世纪以来，我国的蛋白质组学研究发展十分迅速，并取得了显著成绩。此外，《国家中长期科学和技术发展规划纲要》指出，“对蛋白质复杂多样的结构功能、相互作用和动态变化的深入研究，将在分子、细胞和生物体等多个层次上全面揭示生命现象的本质，是后基因组时代的主要任务。同时，蛋白质科学研究成果将催生一系列新的生物技术，带动医药、农业和绿色产业的发展，引领未来生物经济”。因此，我们相信，随着蛋白质组学研究成果的不断涌现，必将促进更多的科研人员投入到该领域的研究之中。

译者希望通过对《蛋白质组学研究——概念、技术及应用》一书的翻译，将原著作者——一些活跃在国际蛋白质组研究领域前沿的科研人员，对近 10 年来蛋白质组研究在样品预处理、质谱鉴定、定量分析、成像分析、数据处理、相互作用等方面的技术发展、生物医学领域的应用，以及未来发展方向的展望介绍给大家。以期能为国内从事蛋白质组研究的研究者提供一些信息，并吸引更多的青年学者加入研究队伍。

本书统稿人张丽华研究员、梁振博士和张玉奎院士对参与本书翻译和校对的人员深表感谢。也非常感谢科学出版社在本书出版过程中给予的理解、支持和帮助。

张丽华 梁 振 张玉奎
中国科学院大连化学物理研究所
2010 年 1 月

再 版 序

10 年前，本书的主编出版了蛋白质组学的第一本著作。书的名字不是《蛋白质组学》，而是《蛋白质组学研究——功能基因组学的新前沿》。旨在延续基因组分析计划，尤其是人类基因组测序计划。然而，我们中很多人已经清楚地意识到在生物学领域正在掀起一场新革命——研究重点将从 DNA 测序转移到蛋白质的结构和功能、蛋白质之间以及蛋白质与其他分子（包括 DNA）间的相互作用等。随着人类和其他物种 DNA 测序的完成，这场科学革命变得越来越清晰。正如科学史上的许多案例一样，以 DNA 技术为主导的科学研究发展最终颠覆了自己。蛋白质组学，即研究细胞和组织中全部蛋白质的科学，已和更常用的“生物多样性”以及“系统生物学”一起，被用来命名这一新的时代。但是人们不要将蛋白质组误解为基因组的简单延续。尽管 DNA 序列是决定蛋白质一级结构必不可少的数据来源，但是这只是显示了一个新故事的开始。相同的蛋白质在不同组织中，甚至是在同一细胞的不同定位和细胞活性状态下，其功能也是不同的。DNA 序列既不能反映蛋白质的三级结构，也不能反映细胞微环境下的蛋白质修饰，更不能揭示蛋白质的合成过程、激活以及失活等。而这些都和蛋白质的功能密切相关。蛋白质组的知识并不仅局限于从基于 DNA 微阵列获得的 DNA 序列来推断表达蛋白质的模式。蛋白质组促进了整个生物学界思维方式的改变。自 20 世纪 60 年代 DNA 结构和功能被揭示以来，在过去的几十年里，人们一直在沿用分子遗传学和基因组学来解释诸如新陈代谢和生命繁衍等与生命本质相关的世纪难题。

这些解释主要基于类似由 DNA 序列编写的“程序”——基因程序。尽管这种方法有些过于简单，但是由于很有启发价值，大多数的生物学家对这些解释还是认可的。由于 DNA 技术简单易行，并且在当时看来很有发展前途，蛋白质物理化学，这个在 20 世纪 50 年代很活跃的研究领域，逐渐失去了活力，甚至走向了被抛弃的边缘。在这种情况下，本书第一版和第二版的作者仍旧在不断发展用于蛋白质以及细胞和组织中蛋白质表达分析的二维凝胶电泳与质谱技术。当人们逐渐认识到基因组学本身只能提供一维结构信息，且几乎不能提供功能信息时，这些作者的研究已经处于生物学研究的最前沿了。

本书阐述了此领域内近 10 年来的进步。不仅在生物信息学和 DNA 数据库的帮助下，发展了各种新技术，而且在接近体内环境下，逐步揭示蛋白质修饰和功能等难题。人们利用不断发展的各项技术，研究了在特定组织中正常和病理状态下细胞内的蛋白质差异。与此同时，也更加清楚地意识到如果蛋白质的结构和生物功能密切相关，必须尽可能地考虑到蛋白质的翻译后修饰。除了研究磷酸化、糖基化、甲基化以及其他共价键相关的修饰外，还应该研究分子间的弱相互作用。目前人们已开始开展蛋白质-蛋白质相互作用网络图的研究工作。所有这些方面的研究产生了大量的数据，这也进一步揭示了功能调控的复杂性。这些相互作用网络如何被调控？它们如何协同作用从而产生可见

的功能化？对于这个问题，不会有类似于基因编码那样，可以用“从细菌到大象均相同”这个谚语来简单地回答。而是针对具体问题会有特定的模型。已经有一些模型应用于临床和药物开发。将来，当个性化的基因组和个性化蛋白质组都变为现实后，人们将发展个性化的医疗。

总而言之，蛋白质组学属于后基因组学的范畴。随着这一全新领域的发展，越来越多的问题被提出，蛋白质组学也展示出其令人吃惊的复杂性。就人类而言，在 200 多种细胞类型中，有着不同的蛋白质组需要研究，并且蛋白质的表达随时间和条件而不断变化。

正如 George Klein 在一次“癌症中细胞通路可能被扰乱”的讨论会中提到的那样“生物学家不仅要接受与复杂性共存的事实，更应该爱上复杂性”。他同时引用了 Tony Pawson 在细胞信号转导中的一句话“我们看到的复杂性在其本身真正的复杂性面前显得微不足道”。

本书中讲述的蛋白质组学将帮助生物学家爱上这种“复杂性”。本书将通过讲述新出现的问题、介绍新的技术和理论工具以激励生物学家发展更高效的方法来解决这一领域内的诸多新问题。

Henri Atlan

“Emeritus 教授”荣誉获得者（生物物理学，巴黎和耶路撒冷）

耶路撒冷 Hadassah 大学人类生物学研究中心主任

巴黎高等社会科学研究学校研究中心主任

2007 年 2 月

再版前言

来自澳大利亚悉尼和瑞士日内瓦的两个研究组的长期密切合作促成了这本书的出版。1997 年我们出版了这本书的第一版。这也激励着我们继续前进，力争再版。许多第一版的作者又加入到再版的工作中，并且还带来了最近 10 年的研究成果。在这 10 年中，本书的作者发展了许多蛋白质组学的新技术和新手段；通过不同途径将诸多新技术商品化；成立了蛋白质组学及生物信息学公司，并将蛋白质组学知识应用于解决重要的学术、临床以及工业问题。我们相信这些不寻常的、独特的亲身经历将使这本书能够满足生物学和医学领域内蛋白质组学研究者的需求。

我们要感谢所有章节作者付出的辛勤而细致的努力，是他们的认真工作才使此书得以顺利出版。我们感谢来自澳大利亚和瑞士的大学与研究所，以及瑞士生物信息研究所、Proteome Systems 公司和 Genveva Bioinformatics (GeneBio) 公司给予的大力支持，是他们组织起众多的作者参与此书的写作。同时感谢近年来澳大利亚国家健康与医学研究委员会 (NHMRC) 以及瑞士国家自然科学基金委员会 (SNSF) 对我们研究工作的资助。

最后，我们要感谢全球蛋白质组学研究者的努力，是他们为我们提供了可以讲述、讨论或者偶尔批评的工作。蛋白质组学是一个快速发展，却也充满着挑战的领域。我们希望此书能够进一步促进本领域的发展，并鼓励无论是资深还是初级的科研人员投身到此领域，去做更多的工作。

Marc Wilkins

Ron Appel

Keith Williams

Denis Hochstrasser

2007 年 4 月

目 录

译者序

再版序

再版前言

第 1 章 蛋白质组学的 10 年	1
1.1 蛋白质组学导引	1
1.1.1 蛋白质组学的含义是什么?	1
1.1.2 事情是否会有所不同?	2
1.2 技术推动了蛋白质组学的发展	3
1.2.1 蛋白质分离	3
1.2.2 质谱分析	4
1.2.3 数据解析	4
1.3 蛋白质组学传达了什么信息?	6
1.4 什么尚不清楚?	7
1.5 本书内容和相关结论	8
参考文献	9
第 2 章 基于电泳的蛋白质组样品制备与预分级技术	11
2.1 引言	11
2.2 常规的样品制备	12
2.3 人为修饰	13
2.3.1 半胱氨酸的还原与烷基化	13
2.3.2 半胱氨酸的 β -消除	14
2.3.3 赖氨酸的氨甲酰化	15
2.4 蛋白质组学的多元化研究方法	16
2.5 预分级技术	18
2.5.1 离心分级	18
2.5.2 色谱技术	18
2.5.3 电泳技术	20
2.6 用于样品预分级的其他方法	22
2.6.1 高丰度蛋白质的去除	22
2.6.2 均衡器颗粒	23
2.7 结论	26
参考文献	27

第3章 蛋白质组学中的蛋白质鉴定	31
3.1 引言	31
3.2 蛋白质鉴定中有用的属性	31
3.2.1 来源种类	31
3.2.2 蛋白质等电点	32
3.2.3 蛋白质分子质量	32
3.2.4 部分序列或序列标签	32
3.2.5 蛋白质氨基酸组成	32
3.3 蛋白质质谱鉴定技术	34
3.3.1 用于蛋白质鉴定的 top-down 和 bottom-up 策略	34
3.3.2 MS 概述	35
3.3.3 基于肽质量指纹谱鉴定蛋白质	38
3.3.4 基于多级质谱的蛋白质鉴定	42
3.4 工具和网址列表	49
3.5 结语	49
参考文献	50
第4章 定量蛋白质组学	52
4.1 引言	52
4.2 非质谱定量分析方法	53
4.3 质谱相对定量方法	55
4.3.1 绝对定量还是相对定量？	57
4.3.2 利用化学标签引入稳定同位素	57
4.3.3 酶催化引入稳定同位素	59
4.3.4 通过生物学的代谢标记引入稳定同位素	60
4.3.5 不使用稳定同位素标记的相对定量法	60
4.3.6 基于质谱的绝对定量法	61
4.4 分析已知的翻译后修饰	61
4.4.1 糖基化	61
4.4.2 磷酸化	62
4.4.3 泛素化	64
4.5 结语	64
参考文献	65
第5章 一个基因，多个蛋白质	70
5.1 引言	70
5.2 修饰总论：有哪些修饰？发生在哪里？	72
5.3 怎样发现翻译后修饰？	73
5.3.1 蛋白质异构体的分离	73
5.3.2 共翻译和翻译后修饰的检测	74

5.3.3 修饰分析的策略: top-down 与 bottom-up 的比较	75
5.3.4 翻译时和翻译后修饰的 MS 分析	75
5.4 特定修饰的分析	77
5.4.1 乙酰化	77
5.4.2 磷酸化	77
5.4.3 泛素化和小泛素化	78
5.4.4 糖基化	78
5.5 蛋白质翻译后修饰的功能: 比所见到的要多?	79
5.6 一些有趣的修饰故事	81
5.6.1 促红细胞生成素的故事	81
5.6.2 脱辅基脂蛋白 E 的故事	81
5.6.3 早衰症的故事	83
5.6.4 流感的故事	83
5.7 未来的方向	84
参考文献	84
第 6 章 蛋白质组成像技术	89
6.1 引言	89
6.2 2-DE 成像技术	90
6.2.1 凝胶成像分析的前期步骤	91
6.2.2 不同蛋白质组方法中的应用实例	92
6.3 液相色谱-质谱	94
6.3.1 液相色谱-质谱成像分析技术的前期步骤	94
6.3.2 不同蛋白质组学方法中的应用	95
6.4 分子扫描器	97
6.5 成像质谱	101
6.5.1 成像质谱技术	101
6.5.2 成像质谱应用	102
6.6 结语	103
参考文献	103
第 7 章 蛋白质组学的数据集成	106
7.1 引言	106
7.2 集成中信息的采集和交联	108
7.2.1 资源的选择和量化	108
7.2.2 生物学上的交联	109
7.2.3 蛋白质组学工作流程的集成单元	111
7.2.4 联合集成机制	113
7.3 信息融合集成	115
7.3.1 文本信息	115

7.3.2 本体	116
7.3.3 融合资源的可视化工具举例	117
7.3.4 从数据集成到系统生物学	117
7.4 结语	118
参考文献	119
第8章 蛋白质的相互作用	123
8.1 引言	123
8.2 人类疾病中的蛋白质相互作用：蛋白质的错误连接导致病变	124
8.3 探索蛋白质相互作用	125
8.3.1 表征生物体中所有编码序列	127
8.3.2 监测双向作用：酵母双杂交系统	127
8.3.3 亲和纯化和质谱方法分析蛋白质复合物	129
8.3.4 发光标记哺乳动物相互作用组定位方法	130
8.3.5 蛋白质生物芯片	131
8.3.6 数据质量	131
8.4 生物和生物医学应用	132
8.4.1 探索病理学和药理学相关的作用途径	132
8.4.2 从酵母相互作用全谱分析中得到的启示	132
8.4.3 一个新应用：蛋白质相互作用的小分子抑制剂的发展	134
8.5 未来发展方向	135
参考文献	136
第9章 蛋白质组学在生物医学中的应用	142
9.1 引言	142
9.2 蛋白质组学在医学上的应用	143
9.3 基于体液的疾病诊断	144
9.4 血管疾病	145
9.4.1 简介	145
9.4.2 蛋白质组学在血管疾病和动脉硬化症中的应用	145
9.4.3 蛋白质组学在心血管疾病中的应用	146
9.4.4 蛋白质组学在脑血管疾病中的应用	147
9.4.5 结论	148
9.5 神经退行性疾病	148
9.5.1 脑蛋白质组	148
9.5.2 神经退行性疾病的蛋白质表达谱	149
9.5.3 脑脊液（CSF）蛋白质标记物	150
9.6 蛋白质组学与癌症	151
9.6.1 肿瘤蛋白质组学中生物标记物的筛选	152
9.6.2 肿瘤学中的蛋白质组表达谱	153

9.6.3 通过蛋白质组学确认组织学起源.....	154
9.6.4 结论.....	154
9.7 药理毒理学：以 2 型糖尿病为例	155
9.7.1 糖尿病概述.....	155
9.7.2 2 型糖尿病的病理学	155
9.7.3 2 型糖尿病的治疗	156
9.7.4 蛋白质组学筛选 2 型糖尿病治疗靶标.....	156
9.8 药物蛋白质组学的现状及将来方向	157
9.8.1 分析前的问题.....	157
9.8.2 分析过程中的问题.....	158
9.8.3 分析后的问题.....	158
9.9 目前和未来的方向	159
参考文献	159
第 10 章 蛋白质组学：下一步在哪里？	163
10.1 引言.....	163
10.2 组学与生物学的相关性	164
10.3 蛋白质组学技术的发展	164
10.3.1 修饰表征	165
10.3.2 全组织分析	165
10.4 蛋白质组学的下一个阶段：诊断学和药学	166
10.4.1 诊断	166
10.4.2 药物	167
10.5 结语.....	167
参考文献	167

图版

第 1 章 蛋白质组学的 10 年

Marc R. Wilkins and Ron D. Appel

摘要

到目前为止，蛋白质组学概念的提出已 10 年有余。每个周年纪念日都是回顾和思考我们称之为蛋白质组学的这个领域取得的新进展的良好时机。哪些方面做得好？哪些方面做得不好？哪些方面取得了新的突破？哪些方面我们还不太清楚？本章将简要对蛋白质分离、质谱鉴定，以及蛋白质组生物信息学等方面的问题进行探讨。

1.1 蛋白质组学导引

本书编者从事蛋白质组学研究皆已超过 20 年。他们发展了分析蛋白质和蛋白质表达谱的技术 (Williams et al. 1991; Hochstrasser et al. 1988)，以及对应用这些技术得到的结果进行解释的软件算法和工具 (Appel et al. 1988; Wilkins et al. 1995)。尽管随着二维凝胶电泳的出现，研究人员可以一次性观察到基因组所能表达的全部蛋白质，而不再是一种蛋白质，然而蛋白质组的概念直到 1994 年才由 Marc Wilkins 在意大利 Siena 的一次会议上提出 (第一届 Siena 会议，二维电泳：从蛋白质图谱到基因组，1994 年 9 月 5~7 日)。该术语是他和他的博士导师 Keith Williams 在 1994 年初创造的。此后不久，使用这一术语的第一批论文发表了 (Wilkins et al. 1995; Wasinger et al. 1995)，关于蛋白质组学的第一本书也于 1997 年出版 (Wilkins et al. 1997)。自第一本书出版后，已经过去的 10 多年里，每个周年纪念日都是我们回顾和思考被称之为蛋白质组学的这个领域所取得的新进展的好时机。哪些方面做得好？哪些方面做得不好？哪些方面取得了新的突破？哪些方面我们还不太清楚？在这里我们尝试就这些问题给出答案。同时，也会对本书的内容加以简单评述。

1.1.1 蛋白质组学的含义是什么？

“蛋白质组”和“蛋白质组学”的概念已被生物学界广泛采用。自 10 年前被提出后，其使用频率快速增加 (图 1.1)。事实上，仅在 2005 年，就有超过 4000 篇关于蛋白质组学的研究报告和综述文章发表。针对该领域的杂志种类的增加是导致这种现象产生的主要原因，这类杂志包括 *Proteomics*、*Proteomics-Clinical Application*、*Practical Proteomics*、*Journal of Proteome Research*、*Molecular and Cellular Proteomics*、*Proteome Science*、*Current Proteomics*、*Genomics and Proteomics*、*Briefings in Functional Genomics and Proteomics*、*Genomics Proteomics Bioinformatics* 和 *Expert Review of Proteomics*。此外，有些关于蛋白质组学研究的论文也发表在其他杂志上。这预示

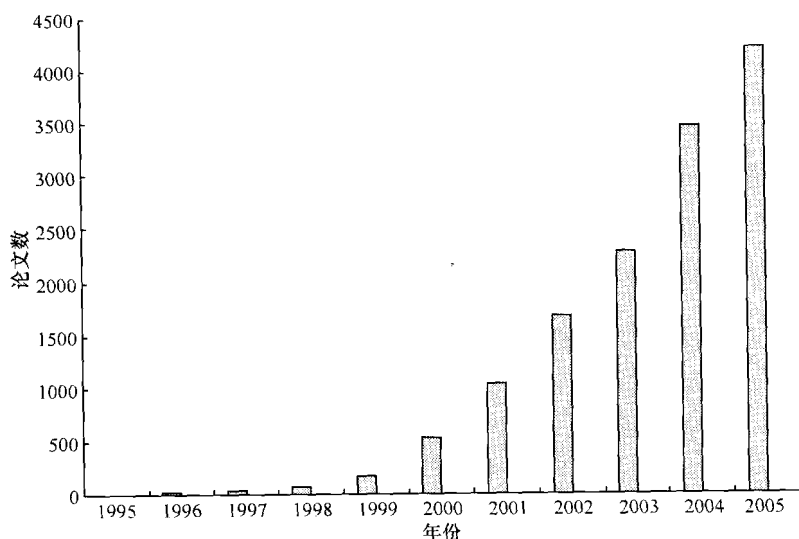


图 1.1 过去的 10 年中，在蛋白质组学和蛋白质组研究领域的论文数量迅速增长。这个数据是查询每年的 NCBI PubMed 数据库时使用关键词“蛋白质组”或者“蛋白质组学”得到的。但是按照这种方法，有一些文章可能被计算了两次。

着蛋白质组学已成为探索复杂生物系统的一种有价值的手段。

假如我们通过某一领域所使用的语言来简单地衡量其发展状况，或许我们会问蛋白质组学的快速发展是否仅仅是所谓的组学革命的一个体现？这能代表该领域的真实发展吗？与之相比，另外两个新的组学领域——代谢组学和糖组学的文献相对较少。2005 年发表的论文数量分别是 433 篇和 115 篇。显然，蛋白质组学的研究更为广泛和深入。

1.1.2 事情是否会有所不同？

这样看来，如果“蛋白质组学”的概念没有被提出，这个世界会有所不同吗？一些评论家认为分离技术（基于凝胶电泳和色谱）和质谱技术的发展，结合由基因组测序计划获得的海量信息，显著提高了研究人员对蛋白质化学的兴趣（Blackstock 2004）。

也有一些评论家认为，蛋白质组概念的提出为以前被基因组和其他核酸研究方法所掩盖的蛋白质化学带来了新的复苏机会。它将生物化学的思维方式从曾经的“一次一个蛋白质”转向一个更加全面的考虑。从语言学的角度讲，如果没有语言，思想也就不复存在（Ferdinand de Saussure，日内瓦大学语言学教授，1901～1913）。蛋白质组和蛋白质组学就是一个很好的实例。其他随后提出的组学术语也是如此 [见 Chitty (2006) 的新组学列表]。新的语言和术语的出现提升了分析技术在科学和文献方面的地位。其他领域的新语言也同样会促进技术发展和凝练思维方式并帮助这些领域的研究得到资助。

1.2 技术推动了蛋白质组学的发展

假如要问到目前为止蛋白质组学领域哪些方面做得较好,人们可能会对新技术的发展和推广给予特别的关注。在过去的10年中,许多重大的技术进步共同使得蛋白质化学转变为蛋白质组科学。重要的是,正是概念的突破与分离、质谱、蛋白质化学和生物信息学领域技术进步的结合才使得这种转变成为可能。序列数据库中可获得的大量核酸序列和基因组信息是另外一个重要的先决条件。

1.2.1 蛋白质分离

最初,蛋白质组学研究人员有一个目标,即从整个或者酸性和碱性范围的双向聚丙烯酰胺凝胶上发现一个蛋白质组样品中的所有蛋白质。电泳在20世纪80年代晚期,人们对可能看到一个蛋白质组样品中的所有蛋白质感到非常兴奋。然而,人们不久就意识到从一个蛋白质组样品中分离并看到全部的蛋白质并非易事。在20世纪90年代中期,人们将第一个基因组序列和预测的蛋白质组结合,实现了对蛋白质组双向凝胶分离的理论计算,从而显示每个蛋白质应该出现的斑点位置(Urquhart et al. 1998)。结果显示蛋白质主要分布在两个区域,一部分为等电点(pI)在4~6.5的蛋白质,另一部分为等电点在8~12的蛋白质。而且大多数蛋白质的分子质量小于100 kDa。将理论上的图谱与实际的双向凝胶电泳的分离图谱进行比较,很快就突显了双向凝胶在极酸性、极碱性或极高分子质量的蛋白质分离方面的不足。将根据大肠杆菌、酿酒酵母和枯草芽孢杆菌的基因组得到的理论图谱与由双向凝胶电泳得到的图谱进行再分析,人们又发现了双向凝胶电泳的另外两个缺陷(Wilkins et al. 1998)——大部分疏水性蛋白质在双向凝胶电泳上缺失以及受双向凝胶电泳的上样量和染色法检测的灵敏度限制,每个细胞中低于1000个拷贝数的低丰度蛋白质很可能检测不到。

从那以后,一系列重要的技术突破开始帮助人们看到蛋白质组中的更多种蛋白质。有关双向凝胶电泳的新进展将在第2章中探讨。广义上说,人们发展了一系列新策略,包括运用窄pH范围的胶条以放大蛋白质组的特定区域;将样品按照生物学(如细胞器)或者物理化学性质的差异(如膜蛋白质)进行分级;从样品中富集或提取感兴趣的蛋白质,以及将新的增溶作用和凝胶电泳技术结合,辅助分析更多难以分析的蛋白质。值得一提的是,分级技术提供了一种可以将样品中更多感兴趣的组分上样到双向凝胶电泳的方法,从而有助于低丰度蛋白质的检测。

为了完全应对复杂蛋白质混合物分析所带来的诸多挑战,人们可能是从1998年Venter等发展的DNA鸟枪测序法得到灵感,提出了一个全新的策略——鸟枪法。它首先采用具有已知特异性切割位点的蛋白酶将复杂的蛋白质混合物或者全蛋白质组解成肽段。然后再将与蛋白质相比虽然数量更多,但在物理化学性质上更相近的肽段,用二维液相色谱和串联质谱进行分析。最后将肽段碎片与序列数据库(Wolters et al. 2001)相比对,从而确定样品中有哪些蛋白质。尽管该方法也具有局限性,尤其是蛋白质异构体的丢失(见第5章),但是它提供了一种有别于凝胶电泳分析的分离和鉴定蛋白质组

中大量蛋白质的方法。

1.2.2 质谱分析

在最近 20 年中，质谱技术的快速发展推动了蛋白质组学的建立。尽管质谱仪器很昂贵，但它不仅具有高质量精确性和分辨率，而且可以分析飞摩尔级的肽段和蛋白质。此外，自动化程度也在逐步提高。目前蛋白质和肽段最常用的离子化方式是电喷雾和基质辅助激光解析离子化。它们可以与多种质量分析器和检测器相匹配（见第 3 章）。

质谱作为蛋白质的鉴定方法几乎完全取代了 Edman 降解法。现在有两种技术，即肽质量指纹谱和肽段碎片，可用于蛋白质的鉴定。肽质量指纹谱已被应用在许多研究中。例如，在酵母蛋白质组的大规模分析中，采用该方法分析了 20 000 多个蛋白质（Gavin et al. 2002）。然而，与高可信度的肽段碎片法相比，采用肽质量指纹谱无法将同一蛋白质碎裂为多个肽段。不过，需要注意的是质谱通常本身不能对肽段或蛋白质进行测序。它们只允许通过将肽段碎片数据与序列数据库相匹配来推断出蛋白质序列。从头测序依然较为复杂，相关工作仍在进行中（见第 3 章）。

除了蛋白质鉴定，对两个或者更多个样品进行定量分析的质谱新方法已被发展。将定量结果加以比较不仅有助于发现生物标记物，而且可以研究内外因微扰引起的蛋白质组的内在变化。早期在两个或者更多样品中蛋白质表达差异的比较主要基于二维凝胶电泳和计算机图像分析（见第 4.2 节）。该方法已被成功用于大量的研究中，现在仍然被广泛使用。新的基于质谱的方法是一项重要突破，其原理为使用不同的稳定同位素来标记两个样品或者多个样品中的蛋白质（Gygi et al. 1999），而后将样品混合并共同分析。质谱的高质量准确性不仅可使同位素得以分离，而且能够进行相对定量。这个概念已经被发展成多种方法（见第 4.3 节）。尽管尚不太完善，但其为两个或者多个样品的比较分析提供了一条新途径。

在蛋白质表征方面，质谱也发挥了重要的作用。由于已证实蛋白质的翻译后修饰对细胞内的多个过程起到关键的调控作用，因而该领域的研究近年来备受关注。最近，蛋白质翻译后修饰在蛋白质-蛋白质相互作用网络中的作用也引起了人们的兴趣（Pawson and Nash 2003）。这已成为该领域的研究热点。现在已经发展了多种分析翻译后修饰的成熟策略（见第 5 章），并被应用于蛋白质组的规模分析。蛋白质的磷酸化分析就是焦点之一（Beausoleil et al. 2004）。这些翻译后修饰分析尽管规模较大且尚不完善，然而它们却是蛋白质组翻译后修饰动态分析的最初尝试。

1.2.3 数据解析

毫无疑问，蛋白质组分析的新策略，以及分离分析技术的发展增加了蛋白质组数据的产出量和复杂性。但是，正是这些分析方法与新的成熟的生物信息学的结合，才能使得研究人员可以更好地产生、分析和研究蛋白质组数据，从而更好地了解生物系统的本质。

在二维凝胶上定量分析蛋白质表达的软件，特别是与新的荧光染色法相关的软件，已经极大地提高了发现两块胶板或两个样品的定性和定量表达差异的能力（见第 4.6

节)。对蛋白质组研究诸多方面都至关重要的蛋白质鉴定软件与统计学方法结合,可以确保鉴定结果的可信度。贝叶斯统计学和非贝叶斯统计学已经被用于解决在基于肽质量指纹图谱鉴定蛋白质时遇到的难题(Perkins et al. 1999)。对于鸟枪法蛋白质组实验结果,数千种蛋白质的鉴定不可能靠人工完成。通过搜索“正常”或者“随机”序列数据库可以评估结果的假阳性率和总体鉴定结果的置信度。人们对蛋白质鉴定置信度的问题争议较大。蛋白质组学杂志现已发布了蛋白质鉴定的准则,希望科研工作者据此规范发表的数据(Wilkins et al. 2006; Carr et al. 2004)。除了改进蛋白质的鉴定策略外,用于大规模蛋白质组实验产生的成千上万质谱数据的峰提取和峰匹配的自动化数据处理流程也得到了发展。此外,人们还发展了用于对海量数据进行管理和储存的工作区(Rauch et al. 2006)。

用于蛋白质翻译后修饰研究的生物信息学技术近年来也取得了极大的进展。在质谱数据中发现蛋白质修饰的软件工具已被用于肽质量和肽段碎片数据分析(见第3章和第5章),以期找出甲基化、乙酰化、氧化和磷酸化等修饰蛋白质。最近,人们已经开发了糖链结构数据库和糖链质量指纹图谱的结构归属工具,因此将加快分析大量复杂质谱碎片谱的蛋白质糖基化的进程。

蛋白质组生物信息学最具意义深远的进步在于其弥补了技术和生物学之间的鸿沟。采用质谱对细胞和组织进行激光扫描成像后,利用生物信息学技术能够将结果可视化。这个惊人的新进展将有助于洞察在细胞中蛋白质表达的微观和宏观不均一性(见第6

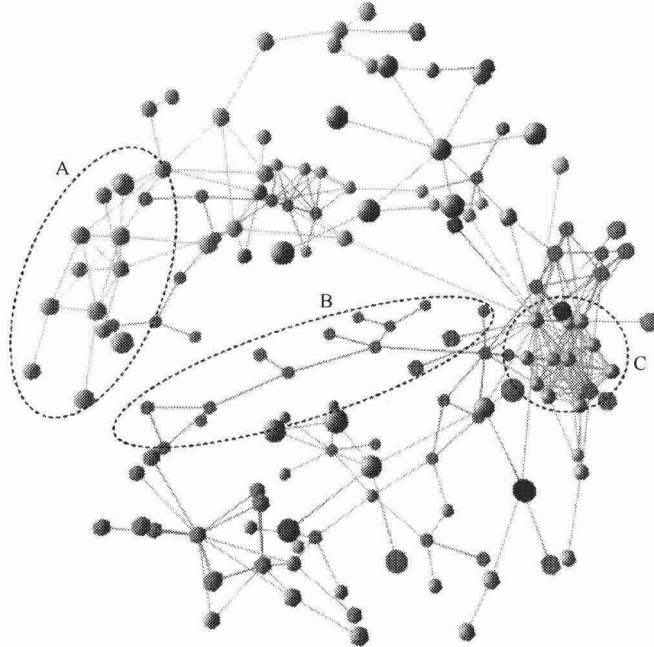


图 1.2 蛋白质-蛋白质相互作用和蛋白质功能的共同可视化。有直接相互作用的一组蛋白质标有相同的颜色,表明有共同的分子功能。这些组中显示的分子功能包括 A: RNA 结合(黄色); B: 结构分子活性(绿色); C: 蛋白质结合(黄色)。(引自 Ho 2006)(另见图版)