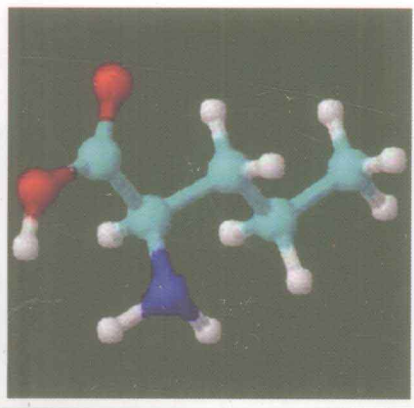




高等院校生命科学类“十二五”规划教材

何华勤 主编

# 简明蛋白质组学



CONCISE PROTEOMICS

中国林业出版社

## 内容简介

本教材共9章,即绪论、蛋白质组学研究技术、蛋白质双向电泳技术、生物质谱鉴定蛋白质技术、蛋白质翻译后修饰、蛋白质组学的定量研究技术、蛋白质结构分析技术、互作蛋白质组研究技术、蛋白质组信息学。教材力争理论和实验技术的结合,使学生不但懂理论,而且会实践。同时,每个章节后附有思考题,便于学生复习和思考。

本教材可供生物科学、生物信息学等专业本科生和研究生使用。

## 图书在版编目(CIP)数据

简明蛋白质组学/何华勤主编. —北京:中国林业出版社, 2011. 1

高等院校生命科学类“十二五”规划教材

ISBN 978-7-5038-5994-6

I. 简… II. 何… III. 蛋白质-基因组-高等学校-教材 IV. ①Q51

中国版本图书馆CIP数据核字(2010)第224989号

## 中国林业出版社·教材建设与出版管理中心

策划、责任编辑:杜建玲

电话:83282720 83220109

传真:83220109

---

出版发行 中国林业出版社(100009 北京市西城区德内大街刘海胡同7号)

E-mail:jaocaipublic@163.com 电话:(010)83224477

http://lycb.forestry.gov.cn

经 销 新华书店

印 刷 北京市昌平百善印刷厂

版 次 2011年6月第1版

印 次 2011年6月第1次

开 本 850mm×1168mm 1/16

印 张 14.25

字 数 352千字

定 价 25.00元

---

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

**版权所有 侵权必究**

## 《简明蛋白质组学》编写人员

**主 编** 何华勤

**编 者** (按姓氏笔画排序)

王经源 (福建农林大学)

王彦芹 (塔里木大学)

庄振宏 (福建农林大学)

何文锦 (福建师范大学)

何华勤 (福建农林大学)

余爱丽 (河北农业大学)

张子丁 (中国农业大学)

张连茹 (厦门大学)

周 倩 (湖南农业大学)

高媛媛 (福建师范大学)

20 世纪人类在基因组学研究上取得了巨大成就,也推动着 21 世纪生命科学研究跨入后基因组时代。蛋白质组学 (Proteomics) 是后基因组时代研究的核心,是从整体水平上分析一个有机体、细胞或组织的蛋白质组成及其活动规律的科学。经过数十年的积淀,蛋白质组学研究取得了丰硕的研究成果,尤为突出的是“人类蛋白质组研究计划”。这些研究成果的取得,促进了一些优秀蛋白质组学专著或教材的出版,如 D. Liebler (1999) 著 *Introduction to Proteomics: Tools for the New Biology*, R. Twyman (2006) 著 *Principles of Proteomics (Advanced Texts)*, 钱小红和贺福初 (2003) 主编的《蛋白质组学理论与方法》,陈主初和肖志强主编的《疾病蛋白质组学》等。这些专著或教材各具特色,要么简明扼要,要么内容全面。随着蛋白质组学研究的不断深入,蛋白质组学专著或教材的内容也需要不断充实、修正和更新。为此,在中国林业出版社的支持下,编写了《简明蛋白质组学》教材,以适应生命科学教育改革与本科人才培养的需要。

本书编写组中聚集了一批具有高学历的科研人员 (每一位编写者成员都已取得博士学位),并且都在从事蛋白质组学的教学与科研实践。编写组成员在分析国内外相关专著和教材特点的基础上,整理了在长期教学与科研实践中积累下来的丰富资料,逐步清晰了本书的编写思路并架构了本书的主要内容。本书具有三个较为显著的特征:一是舍粗存精,编者从庞大的蛋白质组学研究内容中抽取出部分核心内容并进行组装,使本书能适应生命科学类专业的教学需求;二是深入浅出,本书用简明的语言由浅及深逐步展开了蛋白质组学的主要内容,尽量避免复杂的原理与计算,使本书通俗易懂,便于读者自学;三是信息量大,本书的每一章都列出了主要的参考文献,以扩大信息量,指导读者深入学习。同时,在每一章都配有一定数量的复习思考题,帮助读者回顾和总结章节内容。

本书共有 9 章。编写分工如下:第 1 章由福建农林大学生命科学学院何华勤和庄振宏编写,第 2 章由福建农林大学生命科学学院何华勤编写,第 3 章由湖南农业大学生物安全科学技术学院周倩编写,第 4 章由福建师范大学生命科学学院何文锦编写,第 5 章由福建师范大学生命科学学院高媛媛编写,第 6 章

由河北农业大学生命科学学院余爱丽编写，第7章由中国农业大学生物学院张子丁编写，第8章由福建农林大学生命科学学院王经源编写，第9章由厦门大学生命科学学院张连茹编写。

本书编写人员在许多方面都进行了尝试，使其在体系构建、内容选择和思考题的配置等方面与现有同类书籍有较多不同之处。但由于编者水平有限，书中难免有疏漏和不当之处，希望各位同仁和广大读者不吝批评指正。

编写者  
2010年6月于福州

## 前 言

|                                |      |
|--------------------------------|------|
| <b>第1章 绪 论</b> .....           | (1)  |
| 1.1 蛋白质组学的产生 .....             | (1)  |
| 1.1.1 蛋白质组学的产生 .....           | (1)  |
| 1.1.2 蛋白质组学与传统蛋白质化学 .....      | (4)  |
| 1.2 蛋白质组学的发展 .....             | (5)  |
| 1.2.1 蛋白质组学的发展特色 .....         | (5)  |
| 1.2.2 蛋白质组学的研究内容 .....         | (6)  |
| 1.2.3 蛋白质组学发展新趋势 .....         | (8)  |
| 本章小结 .....                     | (9)  |
| 思考题 .....                      | (9)  |
| 参考文献 .....                     | (10) |
| <br><b>第2章 蛋白质组学研究技术</b> ..... | (11) |
| 2.1 表达蛋白质组学 .....              | (12) |
| 2.1.1 蛋白质提取技术 .....            | (13) |
| 2.1.2 蛋白质分离技术 .....            | (15) |
| 2.1.3 蛋白质鉴定技术 .....            | (28) |
| 2.2 功能蛋白质组学 .....              | (31) |
| 2.2.1 酵母双杂交 .....              | (31) |
| 2.2.2 蛋白质芯片技术 .....            | (33) |
| 2.3 结构蛋白质组学 .....              | (34) |
| 2.3.1 X射线晶体衍射图谱法 .....         | (34) |
| 2.3.2 核磁共振 .....               | (35) |
| 本章小结 .....                     | (37) |
| 思考题 .....                      | (37) |
| 参考文献 .....                     | (38) |

|                            |          |
|----------------------------|----------|
| <b>第3章 蛋白质双向电泳技术</b>       | (39)     |
| 3.1 蛋白质样品制备的基本方法           | (39)     |
| 3.1.1 样品的破碎与裂解             | (39)     |
| 3.1.2 蛋白质的沉淀               | (41)     |
| 3.1.3 蛋白质组分的纯化             | (42)     |
| 3.1.4 裂解液的组成成分及其作用         | (44)     |
| 3.1.5 蛋白质的定量               | (45)     |
| 3.2 蛋白质分步提取及亚细胞蛋白质提取       | (46)     |
| 3.2.1 蛋白质一步提取法             | (47)     |
| 3.2.2 蛋白质分步顺序提取法           | (47)     |
| 3.2.3 亚细胞蛋白质提取             | (48)     |
| 3.3 一维等电聚焦电泳               | (50)     |
| 3.3.1 蛋白质是两性电解质            | (50)     |
| 3.3.2 pH梯度的形成              | (52)     |
| 3.3.3 等电聚焦电泳               | (54)     |
| 3.3.4 等电聚焦技术类型及其特点         | (55)     |
| 3.3.5 固相pH梯度技术             | (55)     |
| 3.4 二维SDS-聚丙烯酰胺凝胶电泳        | (58)     |
| 3.4.1 二维SDS-聚丙烯酰胺凝胶电泳的原理   | (58)     |
| 3.4.2 双向凝胶电泳技术的局限性及改进方法    | (60)     |
| 3.5 凝胶的染色、图形获取与分析          | (62)     |
| 3.5.1 凝胶的染色                | (62)     |
| 3.5.2 2-DE数字化图像的处理         | (64)     |
| 3.5.3 蛋白质胶上酶解              | (72)     |
| 本章小结                       | (73)     |
| 思考题                        | (73)     |
| 参考文献                       | (74)     |
| <br><b>第4章 生物质谱鉴定蛋白质技术</b> | <br>(76) |
| 4.1 概 述                    | (76)     |
| 4.1.1 生物质谱仪的基本组成           | (76)     |
| 4.1.2 生物质谱仪的关键性能指标         | (76)     |
| 4.2 基质辅助激光解吸电离质谱           | (77)     |
| 4.2.1 离子化源                 | (78)     |
| 4.2.2 TOF质量检测器             | (80)     |
| 4.3 电喷雾离子化质谱               | (82)     |
| 4.3.1 ESI工作原理              | (82)     |
| 4.3.2 质谱与其他技术的串联使用         | (84)     |

|            |                                |              |
|------------|--------------------------------|--------------|
| 4.4        | 肽质量指纹图谱鉴定蛋白质技术 .....           | (85)         |
| 4.4.1      | 肽质量指纹图谱鉴定蛋白质 .....             | (85)         |
| 4.4.2      | 肽质量指纹图谱的优点及其局限性 .....          | (86)         |
| 4.4.3      | PMF 检索工具 .....                 | (86)         |
| 4.4.4      | PSD 肽片段的部分测序技术 .....           | (89)         |
| 4.5        | 串联质谱数据鉴定蛋白质技术 .....            | (90)         |
| 4.5.1      | 串联质谱测定多肽序列原理 .....             | (91)         |
| 4.5.2      | 串联质谱的优点及局限性 .....              | (92)         |
| 4.5.3      | 肽序列标签鉴定技术 .....                | (92)         |
| 4.5.4      | $^{18}\text{O}$ 标记从头测序技术 ..... | (93)         |
| 4.5.5      | 其他蛋白质鉴定技术 .....                | (94)         |
|            | 本章小结 .....                     | (95)         |
|            | 思考题 .....                      | (95)         |
|            | 参考文献 .....                     | (96)         |
| <b>第5章</b> | <b>蛋白质翻译后修饰 .....</b>          | <b>(98)</b>  |
| 5.1        | 磷酸化蛋白质的鉴定 .....                | (99)         |
| 5.1.1      | 概述 .....                       | (99)         |
| 5.1.2      | 磷酸化蛋白质的检测方法 .....              | (100)        |
| 5.1.3      | 磷酸化蛋白质或多肽的分离与富集 .....          | (101)        |
| 5.1.4      | 磷酸化蛋白质的质谱检测与分析 .....           | (104)        |
| 5.2        | 糖基化蛋白质的鉴定 .....                | (109)        |
| 5.2.1      | 概述 .....                       | (109)        |
| 5.2.2      | 糖基化蛋白质的种类 .....                | (110)        |
| 5.2.3      | 糖基化蛋白质的分离与富集技术 .....           | (113)        |
| 5.2.4      | 糖基化蛋白质的解析 .....                | (114)        |
|            | 本章小结 .....                     | (116)        |
|            | 思考题 .....                      | (117)        |
|            | 参考文献 .....                     | (117)        |
| <b>第6章</b> | <b>蛋白质组学的定量研究技术 .....</b>      | <b>(119)</b> |
| 6.1        | 荧光定量蛋白质分析技术 .....              | (119)        |
| 6.1.1      | 荧光染料染色技术 .....                 | (119)        |
| 6.1.2      | 2D-DIGE 蛋白质组技术 .....           | (121)        |
| 6.2        | 基于质谱的蛋白质组定量分析技术 .....          | (125)        |
| 6.2.1      | 代谢标记 .....                     | (126)        |
| 6.2.2      | 提取后标记 .....                    | (130)        |
| 6.2.3      | 非同位素标记定量分析 .....               | (137)        |

|                               |              |
|-------------------------------|--------------|
| 6.2.4 用 AQUA 肽绝对定量 .....      | (137)        |
| 6.3 蛋白质芯片技术 .....             | (138)        |
| 6.3.1 蛋白质芯片的分类 .....          | (139)        |
| 6.3.2 蛋白质芯片的应用 .....          | (139)        |
| 本章小结 .....                    | (143)        |
| 思考题 .....                     | (143)        |
| 参考文献 .....                    | (143)        |
| <b>第7章 蛋白质结构分析技术 .....</b>    | <b>(147)</b> |
| 7.1 蛋白质的结构 .....              | (147)        |
| 7.1.1 蛋白质结构的组织层次 .....        | (147)        |
| 7.1.2 蛋白质结构的实验测定及结构基因组学 ..... | (150)        |
| 7.1.3 蛋白质结构的分类 .....          | (152)        |
| 7.1.4 蛋白质结构与功能的关系 .....       | (154)        |
| 7.2 蛋白质结构比对 .....             | (155)        |
| 7.2.1 蛋白质结构比对的原理 .....        | (156)        |
| 7.2.2 常用的蛋白质结构比对方法 .....      | (157)        |
| 7.3 蛋白质二级结构预测 .....           | (159)        |
| 7.3.1 常用的二级结构预测方法 .....       | (160)        |
| 7.3.2 不同二级结构预测方法的评价 .....     | (163)        |
| 7.3.3 二级结构预测的展望 .....         | (164)        |
| 7.4 蛋白质三级结构预测 .....           | (165)        |
| 7.4.1 同源模拟 .....              | (165)        |
| 7.4.2 折叠识别 .....              | (168)        |
| 7.4.3 从头计算法 .....             | (172)        |
| 本章小结 .....                    | (176)        |
| 思考题 .....                     | (176)        |
| 参考文献 .....                    | (176)        |
| <b>第8章 互作蛋白质组研究技术 .....</b>   | <b>(178)</b> |
| 8.1 蛋白质-蛋白质相互作用的离体研究技术 .....  | (179)        |
| 8.1.1 蛋白质亲和层析 .....           | (179)        |
| 8.1.2 表面等离子共振技术 .....         | (179)        |
| 8.2 酵母双杂交技术 .....             | (180)        |
| 8.2.1 酵母双杂交技术的原理 .....        | (180)        |
| 8.2.2 酵母双杂交技术的试验流程 .....      | (181)        |
| 8.2.3 酵母双杂交技术的优缺点 .....       | (183)        |
| 8.2.4 双杂交技术的新进展 .....         | (183)        |

---

|            |                          |              |
|------------|--------------------------|--------------|
| 8.3        | 免疫共沉淀技术 .....            | (188)        |
| 8.3.1      | 免疫共沉淀法的原理 .....          | (188)        |
| 8.3.2      | 免疫共沉淀法的实验流程 .....        | (189)        |
| 8.3.3      | 免疫共沉淀技术的应用 .....         | (190)        |
| 8.4        | 蓝色非变性胶技术 .....           | (191)        |
| 8.4.1      | BN-PAGE 技术的原理 .....      | (191)        |
| 8.4.2      | BN-PAGE 技术的实验流程 .....    | (193)        |
| 8.4.3      | BN-PAGE 的应用 .....        | (195)        |
|            | 本章小结 .....               | (196)        |
|            | 思考题 .....                | (196)        |
|            | 参考文献 .....               | (196)        |
| <b>第9章</b> | <b>蛋白质组信息学 .....</b>     | <b>(198)</b> |
| 9.1        | 蛋白质组信息学简介 .....          | (198)        |
| 9.1.1      | 蛋白质组信息学的产生与发展 .....      | (198)        |
| 9.1.2      | 蛋白质组信息学的研究内容 .....       | (199)        |
| 9.2        | 蛋白质组信息学资源 .....          | (200)        |
| 9.2.1      | 蛋白质序列数据库 .....           | (200)        |
| 9.2.2      | 蛋白质模式模体数据库 PROSITE ..... | (203)        |
| 9.2.3      | 蛋白质结构数据库 .....           | (204)        |
| 9.2.4      | 蛋白质结构分类数据库 .....         | (205)        |
| 9.2.5      | 蛋白质互作网络数据库 .....         | (206)        |
| 9.2.6      | 蛋白质功能信息学数据库 .....        | (208)        |
| 9.3        | 蛋白质组信息学技术的应用 .....       | (211)        |
| 9.3.1      | 序列比对 .....               | (211)        |
| 9.3.2      | 结构预测 .....               | (213)        |
| 9.3.3      | 药物筛选及设计 .....            | (214)        |
|            | 本章小结 .....               | (215)        |
|            | 思考题 .....                | (215)        |
|            | 参考文献 .....               | (216)        |

# 第1章 绪 论

“蛋白质组”(proteome)一词源于蛋白质“PROTEin”与基因组“genOME”两个词的杂合,意指“一个基因组所表达的全套蛋白质”。蛋白质组学(proteomics)是以蛋白质组为研究对象,从整体水平上分析一个有机体、细胞或组织的蛋白质组成及其活动规律的科学。其中“-omics”是“组学”的意思,代表对生物体生命活动规律的一种全局研究策略,即从整体的角度来研究一个生物体、细胞或组织。蛋白质组学的主要研究内容包括蛋白质表达存在方式(修饰形式)的鉴定、结构与功能分析、蛋白质定位、蛋白质差异表达以及蛋白质间的相互作用分析等方面。

“蛋白质组学”和“蛋白质组”是澳大利亚学者 Wilkins 和 Williams 等人 1994 年在意大利 Siena 举行的双向凝胶电泳会议上提出,并首次出现在 1995 年 7 月的电泳(Electrophoresis)杂志上。此后生命科学的研究逐渐进入到一个以蛋白质组为研究对象的全新研究领域。经过 10 多年的积淀,蛋白质组学研究已经进入蓬勃发展时期,一批高水平的研究成果陆续在顶尖刊物发表,如 *Cell*、*Nature* 和 *Science* 等。特别指出的是,中国在这其中扮演着重要的角色,有学者认为,“蛋白质组学研究已成为我国生命科学领域能与发达国家保持同步的少有几个领域之一”。2003 年第二届国际蛋白质组大会上确定由中国领导、16 个国家 80 多家实验室共同参与启动“人类肝脏蛋白质组研究计划”(human liver proteome project, HLPP),这是我国有史以来领导的第一项重大国际合作计划,也是第一个人类组织/器官蛋白组计划。

## 1.1 蛋白质组学的产生

### 1.1.1 蛋白质组学的产生

蛋白质组学是在 20 世纪基因组学研究取得巨大成就的基础上发展起来的。基因组学研究促进了蛋白质组学研究的发展,蛋白质组学研究又延伸了基因组学研究的深度。两者结合在一起,深层次揭示生命活动的规律。

#### 1.1.1.1 基因组研究的成就

20 世纪是生命科学迅猛发展的世纪,基因研究为其中的一条主线。为纪念 Waston 与 Crick 提出 DNA 双螺旋结构 50 周年, *Nature* 杂志在 2003 年 4 月 24 日发表了以 DNA 双螺旋结构为背景的基因组研究成就图(图 1-1)。图 1-1 的上半部分为 1900 年到 1990 年基因组学研究的成就。20 世纪初,生命科学的研究以遗传学为代表。孟德尔遗传定律在 1900 年被重新发现,1913 年 Alfred Henry Sturtevant 绘制出第一张线式基因图谱,1929 年 Phoebus Levene 提出 DNA 的化学成分和基本结构,1944 年 Oswald Avery、Colin Macleod 和 Maclyn McCarty 指

## Landmarks in genetics and genomics

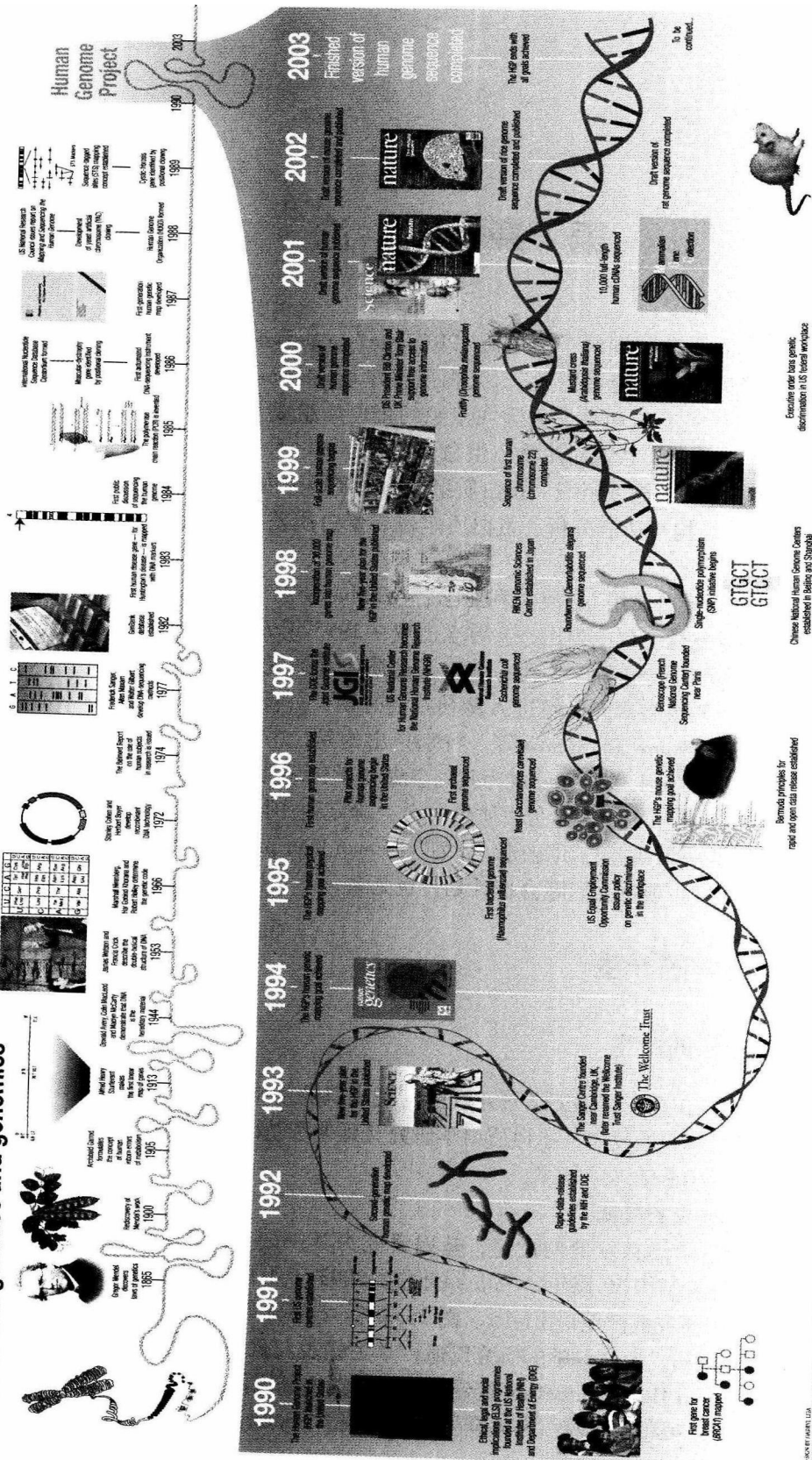


图 1-1 20 世纪基因组学的里程碑

出 DNA 是遗传信息的载体, 并开展从基因分离、独立分配、连锁及化学属性等方面的研究, 直至 1955 年 Waston 与 Crick 提出 DNA 的双螺旋结构, 基因组学走过半个世纪的辉煌。在下半世纪, 生命科学研究以分子生物学为代表, 1966 年 Marshall Nirenberg、Har Gobind Khorana 和 Robert Holley 揭示了遗传密码子; 1972 年 Herbert Boyer 和 Stanley Cohen 发展了重组 DNA 技术, 发现改造后的 DNA 分子可在体外细胞中复制并表达; 1983 年 Kary Mullis 发展聚合酶链式反应 (PCR) 技术; 直至 1990 年美国正式启动人类基因组计划, 生命科学研究达到前所未有的深度和广度。

图 1-1 的下半部分是人类基因组计划的里程碑。人类基因组计划被誉为生命科学领域的“阿波罗登月计划”, 是人类生命科学史上最伟大的工程之一, 是人类第一次系统、全面地解读和研究人类的遗传物质 DNA。1990 年提出人类的基因组计划 (Human Genome Project, HGP); 2000 年, 国际人类基因组计划的科学家发布了人类基因组的“工作框架图”; 2003 年, 国际人类基因组计划宣布绘制完成了更加精确的人类基因组序列图; 2006 年 5 月 18 日, 英美科学家宣布完成了人类第一号染色体的基因测序图, 使已经进行了 16 年的人类基因组计划画上了一个圆满的句号。与此同时, 包括水稻在内的多个物种的基因组测序工作也相继完成。

人类基因组计划的重大研究成果——人类基因组序列图谱的完成, 宣告了生命科学随着新世纪的到来也进入了一个新的纪元——“后基因组时代” (postgenome era)。功能基因组学 (functional genomics) 成为研究的重心, 蛋白质组学又是功能基因组学的核心。对此, 2001 年 2 月, *Nature* 和 *Science* 杂志在公布人类基因组序列草图的同时, 分别发表了题为“*And now for the proteome*” (呼唤蛋白质组) 和“*Proteomics in genome land*” (基因组中的蛋白质组学) 的述评与展望, 对蛋白质组学研究发出了时代性呼唤。

### 1.1.1.2 基因组研究的局限及蛋白质组学的产生

面对庞大的遗传信息, 人们开始关注这些序列信息与生命活动之间直接或间接的联系, 基因的功能是什么? 它们又是如何发挥这些功能的? 也即“后基因组计划”, 又称为功能基因组学研究。完成基因组全长序列的测定只是完成了第一步的工作——结构基因组学, 接下来去完成第二步工作——功能基因组学。正如美国北卡罗来纳州杜克大学负责第一号染色体测序项目的西蒙·格雷戈里博士说, “我们正迈入下一阶段, 那就是弄清楚基因的作用以及如何相互影响。”

尽管已有多个物种的基因组被测序, 但在这些基因组中通常有一半以上基因的功能是未知的。为了分析基因的功能, 功能基因组学采用诸如基因芯片、基因表达序列分析 (serial analysis of gene expression, SAGE) 等的研究策略。但这些研究技术基本上都是以细胞中 mRNA 为研究对象, 其前提是细胞中 mRNA 的水平反映了蛋白质的表达水平。但事实并非完全如此。

基因→mRNA→蛋白质, 三位一体, 构成了遗传信息的流程图, 这是传统的中心法则。但 mRNA 的表达水平不能完全反映蛋白质的表达水平, 原因有 3 个方面。一是基因与蛋白质之间并非一一对应关系, 一个基因并不只存在一个相应的蛋白质, 可能会有几个, 甚至几十个。什么情况下会有什么样的蛋白, 这不仅取决于基因, 还与机体所处的周围环境以及机体本身的生理状态有关。二是组织中 mRNA 的表达丰度与蛋白质丰度的相关性并不好。蛋

蛋白质的动态修饰和加工并非必须来自基因序列。在 mRNA 水平上有许多细胞调节过程是难以观察到的, 因为许多调节是在蛋白质的结构域中发生的。从基因到 mRNA 再到蛋白质, 存在 3 个层次的调控, 即转录水平调控(transcriptional control)、翻译水平调控(translational control)和翻译后水平调控(post-translational control)(图 1-2)。蛋白质的这种修饰是动态的、可逆的, 而且修饰的种类和部位通常不能由基因序列决定。经过 3 个层次调控的蛋白质还通过一系列的运输过程, 定位到组织细胞的适当位置, 才能发挥正

常的生理功能。许多蛋白质还要与其他分子结合后才具有活性。基因不能完全决定蛋白质的后期加工、修饰以及转运定位全过程。用 mRNA 的表达水平代表蛋白质表达水平, 实际上仅考虑了转录水平调控。实践中也已经证明, 组织中 mRNA 丰度与蛋白质丰度的相关性并不好, 尤其对于低丰度蛋白质来说, 相关性更差。更重要的是, 蛋白质复杂的翻译后修饰、蛋白质的亚细胞定位或迁移、蛋白质-蛋白质间相互作用等几乎无法从 mRNA 水平来判断。三是蛋白质组能动态反映生物系统所处的状态。细胞周期的特定时期、分化的不同阶段、对应的生长和营养状况、温度、应激和病理状态, 这些状态所对应的蛋白质组都是有差异的。

毋庸置疑, 蛋白质是生理功能的执行者, 是生命现象的直接体现者, 对蛋白质结构和功能的研究将直接阐明生命在生理或病理条件下的变化机制。蛋白质本身的存在形式和活动规律, 如翻译后修饰、蛋白质间相互作用以及蛋白质构象等问题, 仍依赖于直接对蛋白质的研究来解决。蛋白质组学的研究可提供精确、详细的有关细胞或组织状况的分子描述。

### 1.1.2 蛋白质组学与传统蛋白质化学

值得一提的是, 蛋白质组学研究与传统的蛋白质化学研究不同。蛋白质化学包括研究蛋白质的结构和功能, 通常涉及物理生物化学或机械酶学。研究工作通常包括完整序列测定、结构测定以及进行结构控制功能的模型研究。物理生物化学家和酶学家在同一时间内只研究 1 个蛋白质或多亚基蛋白质复合物。

蛋白质组学研究多蛋白质系统, 重点研究作为一个大系统或部分网络的组成的多个不同蛋白质的相互作用。蛋白质组学需进行复杂混合物的分析, 不是通过完整序列测定进行鉴定, 而是在数据库匹配工具帮助下进行部分序列测定。蛋白质组学的内容是系统生物学, 而不是结构生物学。换句话说, 蛋白质组学的要点是鉴定系统的行为而不是任何单一组分的行为。

因此, 传统蛋白质的研究主要是针对单个蛋白质, 这种研究方式显然无法满足后基因组时代的要求。具体原因有 3 个方面: 其一, 生命现象的发生往往是多因素影响的, 必然涉及

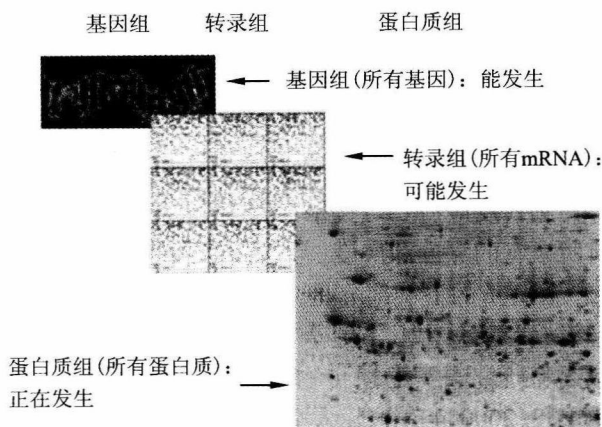


图 1-2 基因组学、转录组学和蛋白质组学间的关系示意图

多个蛋白质；其二，多个蛋白质的参与或是交织互作形成网络，或是平行发生，或呈级联因果；其三，在执行生理功能时，蛋白质的表现是多样的、动态的，并不像基因组那样基本固定不变。单个蛋白质的功能分析显然不足以揭示生命现象的本质。因此，蛋白质组学应运而生。它是以细胞内全部蛋白质的存在及其活动方式为研究对象。可以说，蛋白质组研究的开展不仅是生命科学研究进入后基因组时代的里程碑，也是后基因组时代生命科学的核心内容之一。2001 年国际权威杂志 *Science* 把蛋白质组学列为六大研究热点之一，其“热度”仅次于干细胞研究，名列第二。蛋白质组学的发展备受关注。

## 1.2 蛋白质组学的发展

### 1.2.1 蛋白质组学的发展特色

在 20 世纪 70 年代以前，蛋白质的研究一直优于核酸研究；之后，由于核酸研究新技术的不断涌现，如 DNA 重组、测序、PCR 等技术，使核酸研究后来居上，并远远超出蛋白质的研究水平而成为 20 世纪生命科学研究的主线。但蛋白质的研究尤其是相关技术的发展并未停滞不前。其中 O'Farrell P. H. 于 1975 年根据不同组分之间的等电点和分子量差异建立的蛋白质双向凝胶电泳 (two-dimension electrophoresis, 2-DE) 技术使蛋白质的分辨率能达到成千上万个，完全可以用于组织与细胞中大规模蛋白质的分离，使蛋白质组学的研究进入发展快车道。

国际上蛋白质组研究发展迅速，无论是基础理论还是技术方法，都在不断进步和完善。如果说蛋白质学刚诞生时没有得到国际生物学主流的重视，那么近若干年情况已有了巨大的改变。美国、加拿大、欧盟、日本、韩国等国家和地区都将蛋白质组学作为优先支持发展的领域，相继启动各具特色的大型蛋白质组学研究计划，大力推动本国蛋白质组学的发展，力图在这场 21 世纪最激烈的生命科学竞争中取得先机。如美国国立卫生研究院 (NIH) 提出的未来 15 年发展纲要——NIH 线路图 (NIH Roadmap) 中计划投入大量经费支持蛋白质组研究；美国能源部 2003 年底出炉的《美国未来二十年大型科学设施展望》明确把蛋白质组学大型设施作为其发展重点之一，资助经费在 50 000 万美元以上。欧共体先期资助了酵母蛋白质组研究并取得了重要进展，随后在“第六框架计划” (Sixth Framework Programme) 中将蛋白质组学研究列为优先资助的重要领域。日本启动了“蛋白质 3000 计划” (Protein 3000 Project)，在结构蛋白质组研究项目上已经投入 7 亿美元的研究经费。在法国，5 个研究不同模式生物的实验室得到为期 3 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，联邦政府投资了 730 万美元开展蛋白质组和相关技术研究，并建立了一个蛋白质组中心。1998 年澳大利亚政府着手建立第一个蛋白质组研究网 (Australian Proteome Analysis Facility, APAF)，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。欧洲、亚太地区都成立了区域性蛋白质组研究组织，试图通过合作的方式，融合各方面的力量，完成人类蛋白质组计划 (Human Proteome Project, HPP)。我国自 1999 年开始启动关于蛋白质组研究的国家自然科学基金重大项目。

与人类基因组计划相呼应,2001年4月,在美国成立了国际人类蛋白质组研究组织(Human Proteome Organization, HUPO),提出了人类蛋白质组计划(<http://www.hupo.org/>)。人类蛋白质组计划是继人类基因组计划(Human Genome Project, HGP)之后生命科学领域最大规模的国际性科技工程,也可能是21世纪第一个重大国际合作计划。由于蛋白质组研究的复杂性和艰巨性,人类蛋白质组计划将按人体组织、器官和体液分批启动的策略实施。首批行动计划包括由美国科学家牵头的人类血浆蛋白质组计划和由中国科学家牵头的人类肝脏蛋白质组计划。随后由英国科学家牵头的蛋白质组标准化计划、由德国科学家牵头的人类脑蛋白质组计划、由瑞典科学家牵头的人类抗体计划、由日本科学家牵头的糖蛋白质组计划,以及由加拿大科学家牵头的人类疾病的小鼠模型蛋白质组计划相继启动。最近, HUPO 正酝酿启动重要疾病生物标志物计划,致力于利用蛋白质组学技术寻找重要疾病的生物标志物,以提高其预警、早期诊断和治疗水平。

蛋白质组研究领域的另一个特色是许多企业或药厂开展了一些应用性研究,如膀胱癌、早老年痴呆症的蛋白质组研究、利用蛋白质组技术筛选疫苗等。在美国,各大药厂和公司在巨大财力的支持下,纷纷加入蛋白质组的研究阵容。据报道,Myriad 公司将与美国 Oracle 公司、日本日立股份、瑞士 Friedli 基金组织合作推出“蛋白质组”研究计划;由 Myriad 公司控股,4 家公司共同投资 1.85 亿美元成立 Myriad 蛋白质组学股份有限公司。新成立的公司以鉴定人体中 30 万种以上的蛋白质为目标,并力争弄清各种蛋白质之间相互作用的机制。对此,Myriad 公司的首席执行官 Peter Meldrum 说:“我们将力争在分子水平上去揭示生命过程的奥秘”。Myriad 公司的计划分为两部分:第一部分是在酵母中表达人体的每一种蛋白质的同时研究这些蛋白质的相互作用;第二部分则把目标放在分析人体蛋白质复合体的组成及其中各蛋白质组分的功能与调控机制上。总之,两个方面的研究将帮助科学家们了解蛋白质如何实现正常的细胞功能以及如何抵抗疾病的侵袭。但 Myriad 公司所面临的竞争也是非常激烈。曾在人类基因组计划中发挥了重要作用的 Celera 公司也不甘示弱,宣布投资上亿美元于蛋白质组学领域。而当年提出人类蛋白质索引(Human Protein Index)的美国科学家 Norms G. Anderson 也成立了类似的蛋白质组学公司,继续其多年未实现的梦想。国际上最大的蛋白质组公司——GeneProt 于 2001 年在瑞士成立,该公司由以开发蛋白质组数据库“SWISS-PROT”著称的研究人员发起,投资 1.22 亿美元建立大规模蛋白质组研究中心,以应用蛋白质组技术开发新药物靶标为目的,建立了配备有上百台质谱仪的高通量技术平台。

### 1.2.2 蛋白质组学的研究内容

自 1995 年“proteome”一词问世到 1997 年底,发表的相关文章只有 41 篇,到 2008 年 12 月 31 日发表的文章达到 7 365 篇;近 10 年来文章数量以指数的方式增长。其间出现了一些标志性的研究成果。1999 年 11 月 *Nature* 杂志上发表了一篇用蛋白质组学技术分析蛋白质折叠的研究论文,揭示了蛋白质与分子伴侣间相互作用的关键结构特征,这项工作很好地体现了蛋白质组学研究技术的特点。2000 年 Rout 等运用蛋白质组学技术分离了完整的酵母核孔蛋白复合体,检测到所有的多肽,并系统地对每种可能的蛋白质组分进行复合体内定位与定量分析,从而描绘了酵母核孔蛋白复合体的完整分子结构,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。

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谱,已成为揭示蛋白质组复杂体系与蛋白质功能模式的先导,已成为蛋白质组学领域的研究热点。2000年初, *Science* 刊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官发育的文章,初步建立了与线虫生殖发育相关的蛋白质相互作用图谱,从而为深入研究与揭示线虫发育的机理等提供了丰富的线索。这一工作为以前专注于信号转导过程中单个蛋白质作用机制研究的科学家们提供了一个新的思路。总体而言,近年来蛋白质组学研究已经在蛋白质鉴定、蛋白质翻译后修饰和蛋白质-蛋白质相互作用等领域取得了显著进展。

### 1.2.2.1 蛋白质的分离与鉴定

一些蛋白质组学研究新技术的出现促进了蛋白质的分离与鉴定研究。其中双向凝胶电泳技术与物质谱技术已被广泛应用于可溶性蛋白及细胞膜外表面蛋白的分离与鉴定。对于难溶解的膜蛋白的分离,可采用色谱技术或分步溶解提取技术的研究策略。对于鉴定复合物中的蛋白质,可先对蛋白质复合物进行一维电泳或二维电泳分离,然后结合 Western-blotting (免疫印迹) 技术进行鉴定。也可利用蛋白质芯片,如抗体芯片,或免疫共沉淀等技术对蛋白质进行分离与鉴定。

### 1.2.2.2 蛋白质互作或蛋白质复合体分析

细胞中的调控过程大都是以蛋白质复合体或蛋白质-蛋白质间互作网络协同作用的形式来实现的。蛋白质互作或复合体的研究技术主要有非变性凝胶电泳技术(blue native-PAGE, BN-PAGE)、各种不同类型的色谱技术及免疫共沉淀技术等。从细胞提取物中快速鉴定某一蛋白复合体,特别是低丰度的蛋白质复合体,较为有效的策略是先对转基因生物体进行表位标记(epitope tagging),然后进行亲和沉淀纯化分析。

### 1.2.2.3 蛋白质翻译后修饰研究

许多蛋白质在翻译中或翻译后都要经历一个共价加工的过程,翻译后修饰是蛋白质功能调节的重要方式。蛋白质的翻译后修饰不仅是一个“装饰”,它调节着蛋白质的活性状态、定位、折叠以及蛋白质与蛋白质之间的交互作用等。因此,蛋白质翻译后修饰的研究对阐明蛋白质的功能具有重要作用。

### 1.2.2.4 蛋白质功能鉴定

鉴定蛋白质的功能,如酶的活性及酶底物分析、配基-受体结合的分析,其研究技术有基因敲除和反义基因表达等。另外,对蛋白质翻译后在细胞内的定位研究也在一定程度上有助于蛋白质功能的了解。Clontech 的荧光蛋白表达系统就是研究蛋白质在细胞内定位的一个很好的工具。

### 1.2.2.5 蛋白质复合体的结构分析

对蛋白质复合体结构的研究可揭示复合体与其配体间相互作用的机理,为阐明蛋白质复合体在机体内的生理功能以及指导相关药物的设计和开发提供理论基础。蛋白质经过限制性酶解后,结合交叉连接(cross-linking)或同位素交换技术(isotope exchange),可以分析蛋白质复合体的结构。Tidow H(2010)等人运用X射线衍射技术对PMCA(质膜  $\text{Ca}^{2+}$ -ATPases)的调控区(domain)与钙调蛋白(calmodulin)的复合体晶体体进行分析,检测到的复合体结构分