现代生物技术前沿

蛋白质组学导论
——生物学的新工具

〔美〕D. C. 利布莱尔　著

张继仁　译

高友鹤　校

科学出版社

北　京

图字：01-2004-2025

内 容 简 介

本书介绍了分析蛋白质和肽的各种方法，重点阐述了不同质谱仪及相关数据库检索算法的基本原理和使用方法，详细描述了质谱在蛋白质组学中的应用。蛋白质组学领域最权威的科学家之一 John R. Yates 教授称本书是蛋白质组学的极好的导论和综述，可供有生物化学背景的学生和科学家使用，书写流畅，深入浅出。

本书可用作从事生命科学和医学研究的专业人员的参考书，也可用作学习生命科学和生物技术的本科生和研究生的教材。

图书在版编目（CIP）数据

蛋白质组学导论：生物学的新工具/（美）利布莱尔（D. C. Liebler）著. 张继仁译. —北京：科学出版社，2005.
（现代生物技术前沿）
ISBN 7-03-014258-6

Ⅰ. 蛋… Ⅱ. ①利…②张… Ⅲ. 蛋白质-研究 Ⅳ. Q51

中国版本图书馆 CIP 数据核字（2004）第 098071 号

责任编辑：莫结胜 丁顺华 卢庆陶 / 责任校对：陈丽珠
责任印制：钱玉芬/ 封面设计：王 浩 陈 敬

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

中国科学院印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2005 年 1 月第 一 版 开本：B5（720×1000）
2005 年 1 月第一次印刷 印张：8 1/2
印数：1—3 000 字数：154 000

定价：17.00 元

（如有印装质量问题，我社负责调换〈科印〉）

译者的话

在后基因组时代，生物学家正在对基因组的功能进行研究。蛋白质组学是生物技术的最新领域，它将对基因组功能的研究做出巨大贡献。基因只含有制造蛋白质的指令，而细胞的各种各样的生理功能是由蛋白质来完成的。蛋白质组学将组织或细胞中的蛋白质作为一个系统来研究，而不像在蛋白质化学中那样只研究蛋白质的单一组分。在不同的细胞中会有不同的蛋白质的表达，在相同细胞的不同状况中（如处在疾病状态的细胞中），也会有某些不同的蛋白质的表达。细胞只有一个基因组，但可能有许多不同的蛋白质组。蛋白质组比基因组更复杂。这种复杂性还表现在蛋白质有很多难以预测的翻译后修饰、蛋白质-蛋白质相互作用以及折叠成各种形状的三维结构。从蛋白质的氨基酸序列不一定能推测蛋白质的功能，而蛋白质组学的研究可以揭示蛋白质的表达和功能。各国科学家目前正在用蛋白质组学对人类全部蛋白质进行分类，研究人类蛋白质的相互作用，以便开发更有效的药物。

蛋白质组学面临的挑战是必须研究复杂的蛋白质体系。这要求我们分析各种各样的蛋白质，这些蛋白质大部分以修饰的形式存在并且是低丰度的。《蛋白质组学导论——生物学的新工具》一书描述了应对这个巨大挑战的工具和方法。本书介绍了分析蛋白质和肽的各种方法，重点阐述了不同质谱仪及相关数据库检索算法的基本原理和使用方法，详细描述了质谱在蛋白质组学中的应用。本书作者D.C. 利布莱尔（Daniel C. Liebler）教授是有丰富蛋白质组学研究和教学经验的科学家，尤其在使用质谱技术及相关算法鉴定蛋白质方面有很高造诣。他特别重视蛋白质组学的应用，发表了一系列用质谱技术研究蛋白质修饰的文章。作者力图使本书成为一本可供有生物化学背景的学生和科学家使用的入门教材，书写流畅，深入浅出。蛋白质组学领域最权威的科学家之一 J.R. 耶茨（John R. Yates）教授称本书是蛋白质组学的极好的导论和综述。

本书共分三部分。第Ⅰ部分用两章描述了蛋白质组学和蛋白质组的定义，阐述了蛋白质组学诞生和发展的基础以及在新生物学中的地位，讨论了蛋白质组与基因组的关系。第Ⅱ部分介绍了蛋白质组学的工具和方法。在第 4 章和第 5 章讨论了蛋白质和肽的分离方法以及蛋白质的消化技术。蛋白质和肽的分离包括使用二维 SDS 聚丙烯酰胺凝胶电泳（2D-SDS-PAGE）、制备等电聚焦、HPLC、串联液相层析和毛细管电泳。第 6 章详细描述了 MALDI-TOF 质谱仪和 ESI 串联质谱仪的结构和工作原理以及它们的优缺点，讨论了在蛋白质组学研究中如何选用不同的质谱仪。第 7 章的主题是用肽质量指纹谱鉴定蛋白质，讨论通过测定的

肽质量与数据库理论肽质量的比较进行蛋白质鉴定的方法。第 8 章到第 10 章主要阐述如何用串联质谱分析肽序列，描述了从串联质谱谱图鉴定蛋白质的软件工具（主要介绍 Sequest），也介绍了如何使用 SALSA 算法采集串联质谱数据特征。第Ⅲ部分详细介绍了质谱和相关技术在蛋白质组学中的应用，描述了在蛋白质采集和在蛋白质表达谱的研究中，2D-SDS-PAGE 和 MALDI-TOF 质谱以及肽的多维层析和 LC-串联质谱分析的应用（第 11 章、第 12 章）；讨论了在鉴定蛋白质-蛋白质相互作用和蛋白质复合物以及鉴定蛋白质修饰中，质谱以及 Sequest 和 SALSA 算法的应用（第 13 章、第 14 章）。最后一章指出了蛋白质组学的新的发展方向，包括新质谱仪、自动化和蛋白质微阵等。

相信本书将为从事生命科学和医学研究的专业人员，以及学习生命科学和生物技术的本科生和研究生助一臂之力。

张继仁

序

自 1958 年克劳斯·比曼（Klaus Biemann）教授首先用质谱仪分析氨基酸以来，质谱技术已有了长足发展。Biemann 最初的实验所面临的棘手问题是怎样将非极性分子引入质谱仪产生离子。1958 年以后出现的几种新型电离技术和样品导入方法，促进了生物分子的分析，如化学电离、电场解吸、场致电离、等离子体解吸以及快原子轰击（FAB）等新型电离技术，鉴定肽和蛋白质的方法也得以发展。1987 年由于在生物分子中引入了基质辅助激光解吸电离（MALDI）以及电喷雾电离（ESI），质谱技术有了跃进。这两种电离技术也给肽和蛋白质分析带来极大飞跃，其中一个关键质谱技术是串联质谱。

在 20 世纪 80 年代早期唐纳德·亨特（Donald Hunt）教授开始在肽和蛋白质序列分析中发展和应用串联质谱。FAB 是一项软电离技术，可产生完整的质子化分子，使得用于肽序列分析的方法得以改进。FAB 是肽序列测定的主要突破，该技术可以使肽稳定电离，无需通过其他方法增加肽的挥发性。FAB 与串联质谱的联用，形成了快速肽序列测定方法学。处理复杂肽混合物时，大多数方法采用离线 HPLC 进行分离。人们通过这种方法对许多蛋白质进行了序列测定，并发展了许多重要的方法。然而分离方法与 FAB 的在线结合一直未能发展出可靠易行的方法。直到电喷雾电离使分离技术与质谱仪直接联用，这个问题才得到解决。分析灵敏度的增加以及样品处理的简化和自动化使肽和蛋白质分析的各个方面都得以提高。

质谱的这些进展与全球协同进行的人类基因组序列测定很好地衔接在一起。基因组序列测定工作包括人类基因组及许多模式生物的基因组，并已产生大量的序列信息。1993 年，几个研究小组发现质谱数据可用来检索数据库，以鉴定所研究的蛋白质。1994 年发展了用串联质谱数据检索序列数据库的方法，使研究者能在“书的后面看到答案”。如果“书”是已得到序列分析的生物基因组，答案基本上肯定是在书后面的部分。翻译后修饰和氨基酸序列改变等复杂问题可通过研究从基因组序列推出的蛋白质序列得以解决。

20 世纪 90 年代在生物科学中人们对质谱的兴趣和应用迅速增长，质谱在新千年将会像 SDS-PAGE 一样普遍和重要。生物学家将依赖质谱判断其实验结果。如果生物学家需要使用质谱技术来分析实验，那么他们怎样了解质谱艺术和蛋白质组学的方法呢？D. C. 利布莱尔（Daniel C. Liebler）教授的《蛋白质组学导论——生物学的新工具》一书可以指导我们了解质谱并且在蛋白质组学研究中使用质谱。这本书描述了通常使用的质谱仪和基本的电离技术，这对于确定特定研

究中如何选用质谱仪的类型是重要的。对于非专业研究人员来说，使用质谱数据检索数据库是重要的改进，这样就不再需要掌握解释质谱图的技巧。本书描述了对基础检索算法的基本理解，并阐述了其局限性，最后描述了质谱在蛋白质组学中的应用。本书为研究生和所有对迅速发展的蛋白质组学基础知识感兴趣的生物学家提供了极好的蛋白质组学导论和综述。

J. R. 耶茨（John R. Yates）
Scripps Research Institute
La Jolla，CA

前　言

本书是蛋白质组学这个新领域的导论，侧重描述怎样分析研究蛋白质和蛋白质组。尽管人们对蛋白质组学的兴趣日益浓厚，但是对蛋白质组学的工具和技术了解还很少。本书注重向生物学家介绍新工具和新方法，对生物学学生和有经验的生物学家都适用。任何学过研究生生物化学课程的人都可以很容易地理解什么是蛋白质组学以及如何研究蛋白质组。有经验的生物学家会发现本书大部分内容是熟悉的，但是这些内容被重新整合并围绕蛋白质组的研究展开阐述。

基因组序列测定、分析仪器、计算能力和易于使用的软件工具等方面的重大进展已不可逆地改变了生物学的发展方向。过去我们一直研究生物系统的单个组分，而现在可以综合地并且在精确的分子细节上研究生物系统本身。我们面对的任务是有效地利用新技术和处理大量的数据，更重要的是我们需要调整思想去理解与单一组分相对的复杂体系。

《蛋白质组学导论——生物学的新工具》这本书最早是用质谱进行肽序列分析的短期课程讲义，这门课程是由 Donald F. Hunt 博士 1998 年在北卡罗来纳州 Durham 的生物医学资源设施协会会议上讲授的。那时我的同事 Tom McClure 博士和我在亚利桑那大学毒理中心和亚利桑那癌症中心建立了一个新的蛋白质组学研究机构。Tom 参加了 Hunt 的课程。他回来后，将授课内容传授给我们几个。我们于 1998 年 8 月在亚利桑那大学开设了一个为期四天的关于质谱和蛋白质组学的训练班，共有 50 个学员接受了培训，学员包括研究生、实验室工作人员和教授。对这个训练班的热烈反响反映了以某些易接受方式将蛋白质组学的新技术和在研究中的潜在应用介绍给科学家的需要。这个训练班促进了本书的诞生。

本书是写给初学者的。我的目的是让他们熟悉蛋白质组学的重要工具和应用，所以对某些仪器和应用的描述并不是非常严谨的。本书不是实验室手册或最新技术汇编。有几本很好的书更详细地描述了蛋白质分析技术、质谱仪器和技术，以及这些技术的应用。在这个领域中研究方法的发展和应用是非常迅速的，没有哪本书是真正时新的。在我将蛋白质组学介绍给同事后令我兴奋的是同事们创造性地运用这些新技术，这将促进蛋白质组学的发展。

本书分成三部分。第Ⅰ部分介绍蛋白质组学主题，描述它在新生物学中的位置，分析蛋白质组的性质。第Ⅱ部分介绍蛋白质组学研究的工具，解释它们怎样工作。第Ⅲ部分解释这些工具怎样用来解决不同类型的生物学问题。

感谢 Jeanne Burr、Laura Tiscareno、Julie Jones、Dan Mason、Beau Hansen、Hamid Badghisi、Linda Manza、Richard Vaillancourt、Tom McClure、Arpad Somogyi 和 George Tsaprailis，他们提出了很好的建议，阅读书中各章的草稿并给出评语，提供某些图解的样品数据。感谢 Elizabeth Hedger 杰出的秘书工作。最后，感谢我的妻子 Karen 和儿子 Andrew 对我写作的支持。

D. C. 利布莱尔（Daniel C. Liebler），PhD

目　录

Ⅰ　蛋白质组学和蛋白质组

1　蛋白质组学和新生物学

1.1　新生物学

蛋白质组学是研究与基因组相对应的蛋白质组的学科。术语“蛋白质组学”（proteomics）和“蛋白质组”（proteome）是 Marc Wilkins 及其同事在 20 世纪 90 年代早期提出的，对应于描述生物中全部基因的术语“基因组学”（genomics）和“基因组”（genome）。这些“-omics”术语代表了对怎样思考生物学和生物体系工作方式的重新定义（图 1.1）。直到 20 世纪 90 年代中期，生物化学家、分子生物学家和细胞生物学家还在研究单独的基因和蛋白质或与生物化学途径相关的少量组分。那时可用的技术有 Northern 印迹法（用于基因表达分析）和 Western 印迹法（用于蛋白质分析），利用这些技术研究和分析多基因或多蛋白质是非常困难的。

三项进展形成了新生物学的基础，改变了生物学研究前景。第一项是 20 世纪 90 年代基因、表达序列标签（EST）和蛋白质序列数据库的发展。作为许多生物基因的部分信息库，这些资源很有价值。20 世纪 90 年代后期的基因组序列测定工作，阐明了细菌、酵母、线虫和果蝇的完整基因组序列，最近阐明了人类基因组完整序列。植物和其他广泛研究的动物基因组的序列最近也已完成或接近完成。这些基因组数据库是我们最终从中获取对生物体系理解的信息库。

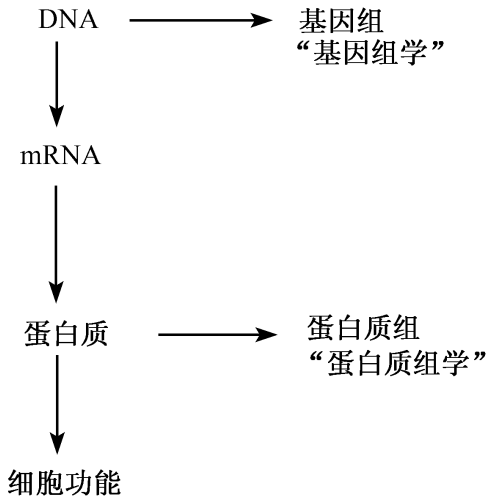

图 1.1　基因组学和蛋白质组学的生物化学关系

第二项重要进展是引入易于操作的、基于浏览器的生物信息学工具。利用这些工具从上述数据库中获取信息。现在可以在几秒钟内从完整的基因组内检索特定的核酸或蛋白质序列。这样的数据库检索工具与其他工具和数据库结合利用，根据已存在的特定功能结构域和基序可以预测蛋白质产物的功能。这样一批基于互联网的免费工具使生物学家可以通过台式电脑检测基因和基因产物的结构与功能，探索大量感兴趣的生物化学问题。

第三项重要进展是寡核苷酸微阵。微阵含有在玻片上或芯片上的一系列基因专一性寡核苷酸或 cDNA 序列。将感兴趣样品的 DNA 混合物荧光标记后与微阵进行杂交，可以一次探测几千个基因的表达。一个微阵可以代替几千个 Northern 印

迹法分析，可以在做一次 Northern 印迹的时间内完成。通过使用双色荧光探针标记，两个不同样品的基因表达谱可直接在一个玻片或芯片上进行比较。

图 1.2 是一块含有酿酒酵母基因组中 6 000 个基因单一序列的玻片。这样的一个微阵可以测定酵母基因组所有基因的表达。显然，这使我们面临新生物学的巨大挑战。我们可以观看整个系统，但这几千个数据点所包含的信息超出了我们直观解释的能力。新的组合算法、自组作图和类似的工具等最新的方法有助于生物学家理解这些数据。

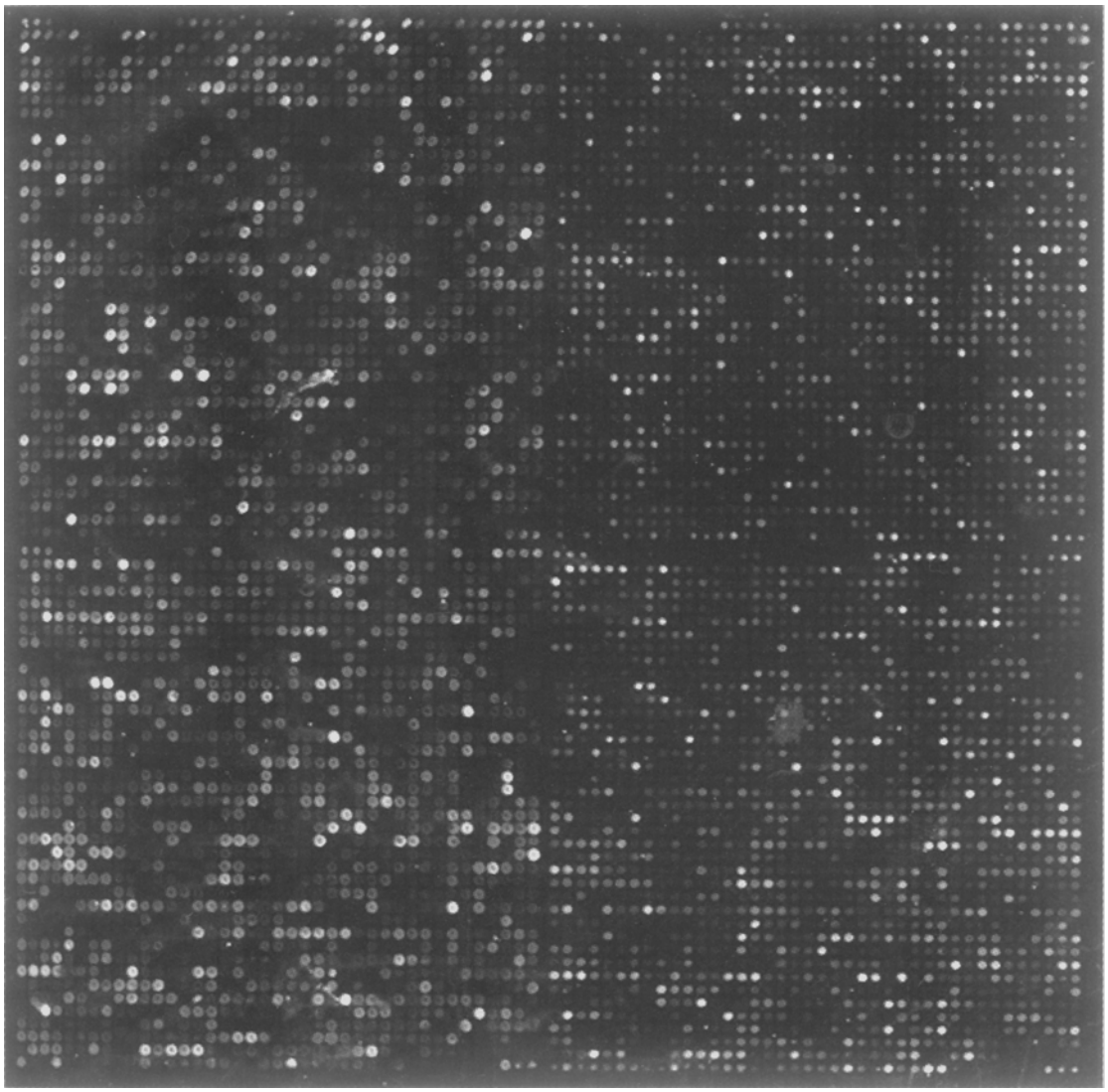

图 1.2　酵母基因组芯片

该酵母 cDNA 微阵由斯坦福大学 Patrick Brown 博士的实验室制作（http：//cmgm. stanford. edu/pbrown/）。

微阵带来的最重要的变化是使生物学家进行“宏观”思考。细胞有成千上万的以不同组合方式进行的基因表达。细胞的生死是由这些基因的表达和其蛋白质产物的活性决定的。无论是跨膜受体、转录因子、蛋白激酶或是伴侣分子，每种蛋白质所表达的功能只有在同一细胞内其他蛋白质的功能和活性同时表达时才有意义。因此生物学家正在努力做宏观思考，去理解系统而不只是理解组分，寻求复杂性的意义。

1.2 蛋白质组学？它不过是我们过去称之为蛋白质化学的学科！

对新思想、术语和方法人们通常说它们其实不是新发展来的，因此解释蛋白质组学与蛋白质化学的不同是很重要的。表 1.1 对各自的主要特征作了小结。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一个蛋白质或多亚基蛋白质复合物。

表 1.1 蛋白质化学和蛋白质组学的不同

蛋白质化学	蛋白质组学
单一蛋白质	复杂混合物
全序列分析	部分序列分析
强调结构与功能	强调通过数据库匹配进行蛋白质鉴定
结构生物学	系统生物学

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。

1.3 我们能测定基因表达，为什么还要有蛋白质组学？

基因微阵提供了细胞中大量或全部基因表达的快速检测。然而从 mRNA 水平并不一定能预测细胞中相应蛋白质的水平。各种 mRNA 不同的稳定性和不同的翻译效率能够影响新蛋白质的产生。蛋白质形成后，在稳定性和转换速度上有很大不同。许多参与信号传导、转录因子调节和细胞周期控制的蛋白质迅速转换，这是其活性调节的一种方式。mRNA 水平没有告诉我们相应蛋白质的调节状态，蛋白质的活性和功能常有一些内源翻译后的改变，也会因环境因素而改变。

1.4 蛋白质组学：对分析的挑战

如何同时测定一个生物中大量或全部基因的表达似乎已通过引入 cDNA 或寡核苷酸微阵得以解决。用 DNA 微阵和相关方法分析基因表达依赖于两个重要工具：聚合酶链反应（PCR）和寡核苷酸与互补序列的杂交。但是没有类似的工具用于蛋白质分析。首先，没有蛋白质 PCR 等价物。目前不可能有多肽分子以类似于核苷酸通过 PCR 复制的方式复制。少量的寡核苷酸可以通过 PCR 进行扩增，而少量的蛋白质必须在没有任何扩增的情况下进行测定和分析。

第二，蛋白质不能专一性与互补氨基酸序列杂交。Watson-Crick 碱基配对

允许寡核苷酸与互补序列杂交。一个特定的互补寡核苷酸序列可以作为高度专一性探针，一个特定的 mRNA 或其他核酸片段可以与之结合。这种专一性允许在微阵上有一个特定的点以便识别单一序列。尽管抗体和寡核苷酸结合子（aptamer，也称适体）可以识别特定的肽或蛋白质，但是这种识别不能简单地根据序列来预测，而寡核苷酸的杂交则可以根据序列来预测。

另一个蛋白质组学的特有问题是细胞中每一个蛋白质产物并不一定只有一种分子实体。这是由于蛋白质有翻译后修饰。修饰的内容和变化随不同的蛋白质、细胞的调节机制和环境因子而变化，许多蛋白质以多种形式存在。对任何特定基因的多种蛋白质产物进行检测和区分的必要性使蛋白质组学在分析方面更具挑战性。

蛋白质组的分析需要一套不同于基因表达分析的工具，能够对修饰和非修饰的蛋白质进行检测和定量的分析。我们怎样应对这项任务是这本书的主题。

1.5 蛋白质组学的工具

尽管上面描述了分析蛋白质组学的不利条件，但是鉴定蛋白质组及其组分实际上可以完成。这是由于以下 4 种重要工具的发展和结合使用给研究人员提供了灵敏性和专一性较高的识别和鉴定蛋白质的方法。

第一种工具是数据库。蛋白质、EST 和基因组序列数据库共同提供了生物表达全部蛋白质的完整数据库目录。例如，根据对果蝇的所有编码序列的分析，我们知道有 110 个果蝇基因编码具有 EGF 类结构域的蛋白质，87 个基因编码具有酪氨酸激酶催化结构域的蛋白质。在进行果蝇蛋白质组学研究时，我们检索大量已知的可能蛋白质结构域。当用限定的序列信息，甚至原始质谱数据（见下文）进行检索时，我们可以根据质谱数据与数据库的匹配情况鉴定蛋白质组分。

第二种工具是质谱（MS）。质谱仪的使用在过去十年中有了极大的革新，在发展为分析生物分子，特别是分析蛋白质和肽的高灵敏度和高可靠性上达到顶点。MS 仪器的使用可提供三类分析，这三类分析在蛋白质组学分析中都非常有用。首先，MS 可以进行 100 kDa 或更大完整蛋白质的精确质量测定。估计蛋白质质量的最好方法是 MS 分析，而不是测定蛋白质在十二烷基硫酸钠-聚丙烯酰胺凝胶电泳（SDS-PAGE）的迁移。高精度蛋白质质量测定的应用有限，因为它们往往不够灵敏。净质量对精确鉴定蛋白质往往是不充分的。其次，MS 也能对蛋白质水解消化产生的肽进行精确的质量测定。相对于完整蛋白质质量测定，肽质量测定可以有高灵敏度和高质量准确度。可以直接用肽质量测定数据在数据库中进行检索，这样常常可以确切鉴定靶蛋白质。最后，MS 可以对蛋白质水解消化得到的肽序列进行分析。目前认为 MS 是肽序列分析中的最新技术。MS 序列数据为蛋白质鉴定提供了最有力和最精确的方法。

蛋白质组学的第三个必要工具是对数据库中特定蛋白质序列与 MS 数据进行比对的各种软件。前面提到从 MS 数据可以测定序列，但是这种从头开始分析成

百上千的谱图时是一项费时费力的任务。蛋白质组学软件将未分析的 MS 数据，在特定算法的帮助下与蛋白质、EST 和基因组序列数据库的序列相比对，自动检测大量用于蛋白质序列匹配的 MS 数据。然后研究人员检查自动检测的结果，估计数据的质量，所用的时间比手工解释每一张谱图要少得多。

蛋白质组学的第四种必需工具是蛋白质的分析分离技术。在蛋白质组学中蛋白质分离有两个目的。第一，通过将蛋白质混合物分离成单一蛋白质或蛋白质小组以简化复杂蛋白质混合物。第二，蛋白质的分离分析可以比较两个样品蛋白质的不同表观，研究者可以标记用于分析的特定蛋白质。2D-SDS-PAGE 是最广泛用于蛋白质组学的技术。2D 凝胶电泳也许是在复杂样品中分离蛋白质的最好单项技术。其他的蛋白质分离技术，包括 1D-SDS-PAGE、高效液相层析（HPLC）、毛细管电泳（CE）、等电聚焦（IEF）和亲和层析，也都是分析蛋白质组学的有用工具。最有力的技术是将不同的蛋白质和肽分离技术结合为多维技术。例如，离子交换液相层析（LC）与反相（RP）-HPLC 的串联是分离复杂肽混合物的有力工具。

这四种工具的结合形成了蛋白质组学当前的技术，每一种工具在技术上都发展迅速。在本书的后面几章我们将讨论每一种分析工具。

1.6 蛋白质组学的应用

蛋白质组学技术确实很新颖，但是鉴定蛋白质组究竟是为了什么呢？根据目前的实践，蛋白质组学包括 4 项主要应用，它们是：①采集；②蛋白质表达谱；③蛋白质网络谱；④蛋白质修饰谱。下面对上述每一项进行简短定义，在本书其后各章将详细讨论。

采集是鉴定样品中所有（或尽可能多）的蛋白质。采集主要是直接对蛋白质组进行分类，而不是通过基因表达（如通过微阵）数据来推断蛋白质组的组成。蛋白质组学中采集需要耗费大量的劳动，以使蛋白质得到最大程度的分离，然后使用 MS 和相关的数据库以及软件工具进行鉴定。有几种采集方法，每一种都有其优点。这些方法的联用可直接分析证实那些只能从基因表达数据推断的数据。

蛋白质表达谱是鉴定生物或细胞特定状态（如分化、发育状态或疾病状态）下蛋白质的表达或药物、化学或物理刺激下蛋白质的表达。表达谱其实是特殊的采集形式，分析中比较一个特定系统的两种不同状态。例如，比较正常细胞和病理细胞中哪些蛋白质有不同的表达。这种信息对于检测药物治疗的潜在靶子极为有用。

蛋白质网络谱是在生物系统中测定蛋白质之间相互作用的蛋白质组学方法。大多数蛋白质在执行功能时与其他蛋白质密切相关。这些相互作用决定蛋白质功能网络（如信号传导级联过程和复杂的生物合成或降解途径）的功能。大多数蛋白质-蛋白质相互作用是通过体外纯化的蛋白质和用酵母双杂交系统获得。通过亲和俘获配对技术与分析蛋白质组学方法相结合，蛋白质组学可以鉴定更复杂的

蛋白质网络。蛋白质组学方法已用来鉴定多蛋白质复合物的组分。在细胞中多蛋白质复合物与点到点的信号传导途径有关。蛋白质网络谱可以测定途径中所有参与者的状态。蛋白质网络谱是蛋白质组学最具远大前景的应用之一。

蛋白质修饰谱是鉴定蛋白质怎样和在何处被修饰的。许多蛋白质翻译后的修饰控制着蛋白质的靶向、结构、功能和转换。此外，许多环境化学因素、药物和内源化学因素可产生修饰蛋白质的活性亲电体。研究人员已开发了各种分析工具用以鉴定修饰蛋白质和修饰的性质。修饰蛋白质可用抗体测定（如用特定磷酸化氨基酸残基的抗体），但是一个特定修饰的精确序列位点往往是未知的。蛋白质组学方法是研究翻译后修饰的性质和序列专一性的最好方法。这项方法的扩展允许在一个网络中同时鉴定调节蛋白质的修饰状态，这是蛋白质组学技术的重要扩充。这些方法用新方式回答蛋白质组的化学修饰怎样影响生物系统的问题。

推荐读物

Brown, P. O. and Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21**, 33—37.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680—686.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Pros. Natl. Acad. Sci. USA* **95**, 14, 863—14, 868.

Fields, S. (2001) Proteomics in genomeland. *Science* **291**, 1221—1224.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860—921.

Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F., Gentile, C., Hwang, S. Y., et al. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* **94**, 13, 057—13, 062.

Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405**, 837—846.

Venter, J. C., Adams, M. D., Myers, E. M., Li, P. W., Mural, R. J., et al. (2001) The sequence of the human genome. *Science* **291**, 1304—1351.

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., and Williams, K. L. (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **13**, 19—50.

2 蛋白质组

2.1 蛋白质组与基因组

每一个人类细胞中含有制造一个完整的人所需的全部信息。然而，并不是全部基因在所有细胞中都表达。编码细胞实现基本功能（如葡萄糖代谢、DNA 合成）必需酶的基因在所有细胞中表达，而具有高度专一功能的基因只在特定类型的细胞中表达（如视紫红质在视网膜色素上皮细胞中表达）。一个细胞中表达两类基因：①必需功能蛋白质的基因；②行使细胞专一性功能蛋白质的基因。因此，一种生物有一个基因组，但有许多蛋白质组。

任何细胞的蛋白质组是所有可能基因产物的某种子集，但这并不意味着蛋白质组比基因组简单。事实正相反，任何蛋白质，即使只是同一个基因的产物，也可能存在多种形式。在一个特定的细胞内或在不同的细胞之间，蛋白质的存在形式可能不同，大多数蛋白质都以几种不同的修饰形式存在。这些修饰影响着蛋白质的结构、定位、功能和转换。

在这一章从 5 个方面讨论蛋白质组。首先，扼要讨论蛋白质的“生命周期”，从蛋白质作为翻译产物在核糖体上出现，到翻译后的多种修饰，再到最终降解。第二，讨论可根据蛋白质的序列基序、结构域的结构和生化功能分成具有不同标准组件结构的蛋白质。第三，讨论功能蛋白质家族在基因组中的分布。第四，通过基因组序列讨论在生物系统中有多种功能和冗余功能的蛋白质组。最后，讨论在特定时间内影响一个蛋白质在细胞中存在数量的各种因素，并讨论这些因素给蛋白质组学分析方法带来的困难。

2.2 蛋白质的生命周期

蛋白质的合成是通过在核糖体上将 mRNA 翻译成多肽来进行的。大多数情况下，最初的多肽翻译产物要进行某种类型的修饰后才具有功能。这些修饰统称为“翻译后修饰”，包括各种可逆和不可逆化学反应。已报告有接近 200 种不同类型的翻译后修饰。图 2.1 表明一个原型蛋白质的生命周期，并总结了一些修饰作用。

蛋白质是 mRNA 序列通过核糖体翻译产生的。半胱氨酸的巯基形成二硫键的折叠和氧化，使多肽随机卷曲具有二级结构。若干“永久”修饰，如谷氨酸的羧化和去除N端甲硫氨酸，在多肽生成的早期发生。在高尔基体的进一步加工

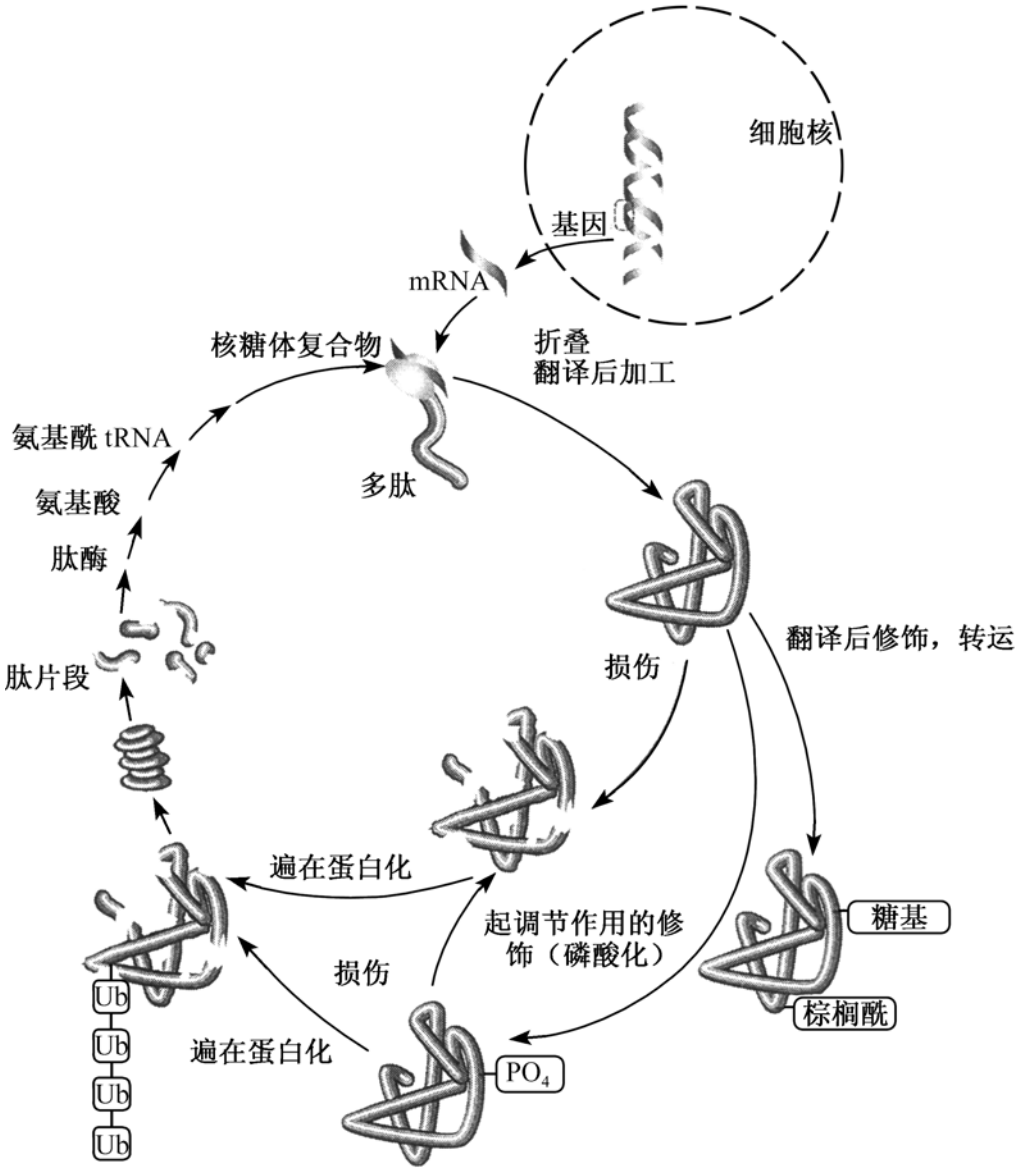

图 2.1　蛋白质的生命周期

产生糖基化。蛋白质到特定的亚细胞或细胞外区域的专一性运送往往需要有前导肽或信号肽序列，这些前导肽或信号肽在完成定位后被蛋白水解酶切割除去。某些蛋白质上还进行辅基加成作用。一个特定蛋白质可与其他蛋白质结合形成多亚基复合物。半胱氨酸残基的棕榈酰化或异戊二烯化辅助蛋白质进入膜内或嵌在膜上。这些不同程度的“永久”性修饰和运送使功能蛋白质进入细胞的特定位置。

蛋白质在细胞内的特定部位执行其功能。翻译后修饰调控许多蛋白质的活性，最重要的和研究较深入的翻译后修饰是丝氨酸、苏氨酸或酪氨酸残基的磷酸化。磷酸化可以使酶活化或失活、改变蛋白质-蛋白质相互作用和连接、改变蛋白质结构、引起蛋白质降解。蛋白质的磷酸化以各种形式调节蛋白质的功能，是信号传导级联反应、细胞周期调控和其他重要细胞功能的快速调控中的关键开关。

蛋白质也易受损伤。在生物系统中普遍存在的自由基和其他氧化剂导致蛋白质的氧化损伤。半胱氨酸巯基对氧化特别敏感。甲硫氨酸、色氨酸、组氨酸和酪氨酸残基容易发生氧化。氨基酸也易受脂类和糖类氧化产物的攻击，这包括活化

的α，β-不饱和羰基化合物。除了这些引起蛋白质修饰的内源因素外，辐射、化学和药物等环境因素也能引起蛋白质的共价修饰或氧化修饰。许多修饰作用可引起蛋白质失活。各种修饰因素都可改变某些蛋白质的结构。

蛋白质修饰对于引发蛋白质降解的过程很重要。某些蛋白质磷酸化后迅速与泛素连接，并被26S蛋白酶体复合物降解。细胞中其他因素也可导致蛋白质的泛素化，包括氧化损伤和蛋白质的其他修饰。蛋白质也可以通过溶酶体酶降解。

图2.1表明了蛋白质组的一个要点：在细胞中任何时间任何蛋白质都可能以多种形式存在，从而使蛋白质组变得异乎寻常地复杂。另一方面，蛋白质组的状态反映了细胞的所有功能状态。

2.3 具有标准组件结构的蛋白质

另一种研究蛋白质的方式是把它们想像成具有标准组件或标准组件拼接的结构。某些氨基酸序列倾向于形成如α螺旋或β折叠的二级结构，或随机卷曲结构。特定的氨基酸序列和从这些序列产生的二级结构具有特定的性质和功能。可以认为氨基酸序列片段是功能的建筑部件或组件。大自然用这些组件建造了工具箱，通过这个工具箱，可以建造多种具有相关功能的蛋白质。

具有特定性质和功能的蛋白质标准组件单位称为“基序”或“结构域”。不同蛋白质中存在的结构域，其可识别序列有类似性质或功能。一般在使用时，这些术语往往可替换使用。在某些情况下，基序或结构域的氨基酸序列高度保守，不随蛋白质的不同而变化。还有一些情况下，一个序列中某些关键氨基酸以重复形式存在，而另一些氨基酸可有各种变化。

即使某些短序列也能赋予蛋白质某些专一性修饰。例如，*N*-糖基化的蛋白质序列中多含有三肽序列“Asn-Xaa-Ser/Thr”。在这个序列中，靶子天冬酰胺之后可以是任何氨基酸，接下来的氨基酸是丝氨酸或是苏氨酸。如果Xaa是脯氨酸，则不能进行糖基化修饰。尽管这个三肽序列并不能保证一定进行*N*-糖基化修饰，但是它提供了一种基序，这种基序提供可能的生物化学作用。

较长氨基酸序列常形成结构域，赋予蛋白质特定的性质或功能。某些结构域结构只是赋予一个肽段重要物理性质，如跨膜结构域，一般形成跨脂双层膜的α螺旋。另一些结构域能够为重要酶底物和辅基提供氢键或其他接触。例如，真核丝氨酸/苏氨酸蛋白激酶有一个富甘氨酸的核心结构域。这个富甘氨酸区域位于与ATP结合的赖氨酸残基和作为催化中心的保守的天冬氨酸残基周围。在许多情况下，结构域由二级结构单位的结合组成，如螺旋-环-螺旋结构域。

基序和结构域对蛋白质组的意义是它们代表从蛋白质序列到蛋白质功能的翻译。当已知性质和功能的结构域和基序出现在未知功能的蛋白质中时，可以推测其细胞功能。简言之，分析蛋白质组学可以确定序列，序列可以确定功能。

2.4 功能蛋白质家族

另一种研究蛋白质组的方式是将它分成具有类似功能的蛋白质家族。例如，某些蛋白质具有结构作用，另一些蛋白质参与信号传导途径，还有一些蛋白质调控如核酸合成或糖代谢等必需代谢途径。根据结构域及其相关功能作用的分类，Venter 及同事推测出人类基因组所编码蛋白质的功能分布（图 2.2）。

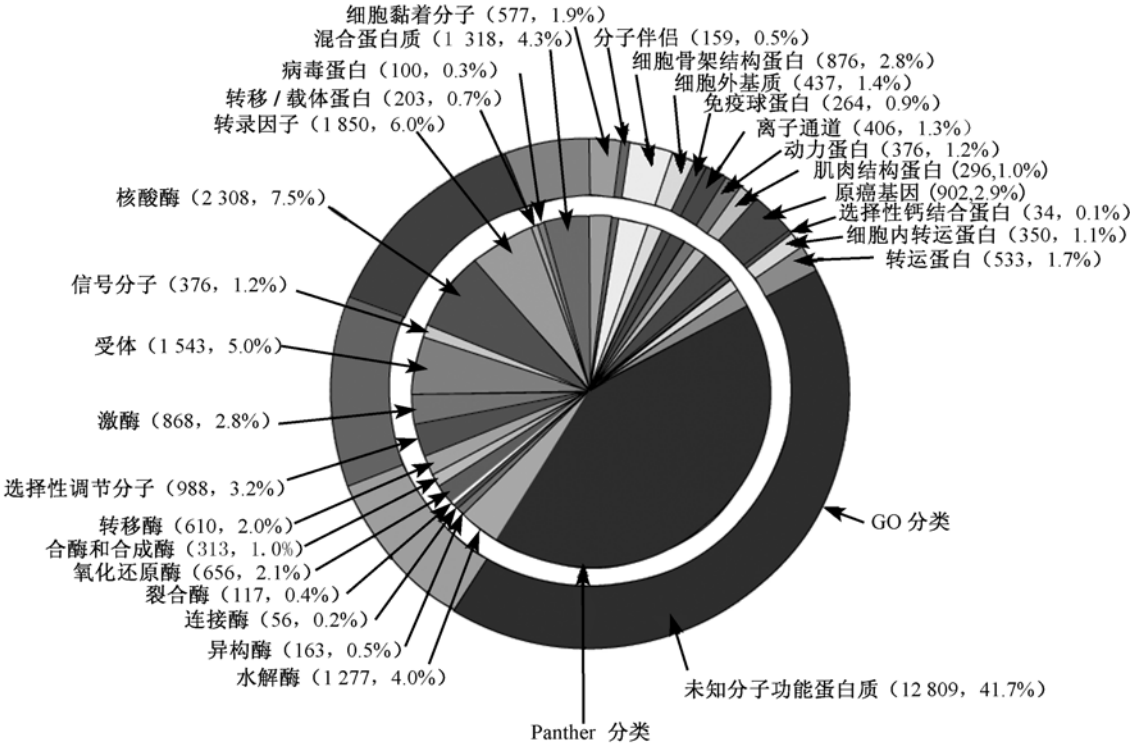

图 2.2 推测的人类基因组蛋白质产物的功能

[Venter et al.（2001）Science，291：1304～1351]

与中间代谢和核酸代谢有关的酶占蛋白质组的 15%。与结构和蛋白质合成与转换相关的蛋白质（细胞骨架蛋白、核糖体蛋白、分子伴侣和蛋白质降解相关因子）总共占 15%～20%。信号传导蛋白和 DNA 结合蛋白占 20%～25%。尽管这些数字给由蛋白质功能来解析基因组提供了有价值的参考数据，但我们并不能确定在细胞中某一特定时间某一特定蛋白质或某类特定蛋白质的表达量。接近 40%的基因组编码的蛋白质功能尚属未知。确定这些基因产物的功能是人类功能基因组学面临的最基本的挑战。

2.5 从基因组推算蛋白质组

对鉴定生物基因组的研究人员来说一个最有趣的问题是“有多少基因?”。这个问题的答案可以使我们得到共有多少蛋白质存在于蛋白质组中的概念。几种生物的全部基因组序列已经测序完成，这些数据允许分析家预测所有基因的产物。根据推测的每一个基因产物的氨基酸序列，按照蛋白序列中所含的结构域和序列

基序已对它们进行分类。例如，酿酒酵母有 119 个基因编码具有真核蛋白激酶结构域的蛋白质，另外有 47 个基因编码具有 C2H2 类型锌指结构域的蛋白质。结构域序列特征与基因组序列的比较表明生物基因组为各种类型蛋白质编码。

最近对酿酒酵母、线虫和果蝇的分析揭示出这些生物的基因组大小与预测的蛋白质组容量之间有着非常有趣的关系。Gerald Rubin 和同事根据已存在的特定的结构域，对通过流感嗜血杆菌、酿酒酵母、线虫和果蝇基因组预测的蛋白质产物进行了分类（图 2.3）。比较所有预测的蛋白质产物，结果表明基因组中存在序列与其他物种蛋白质序列稍有不同的蛋白质。通过对这些冗余蛋白质产物（称为平行进化同源物）的校正可以计算每一种生物的“核心蛋白质组”。这种核心蛋白质组代表生物的不同蛋白质家族的汇集。

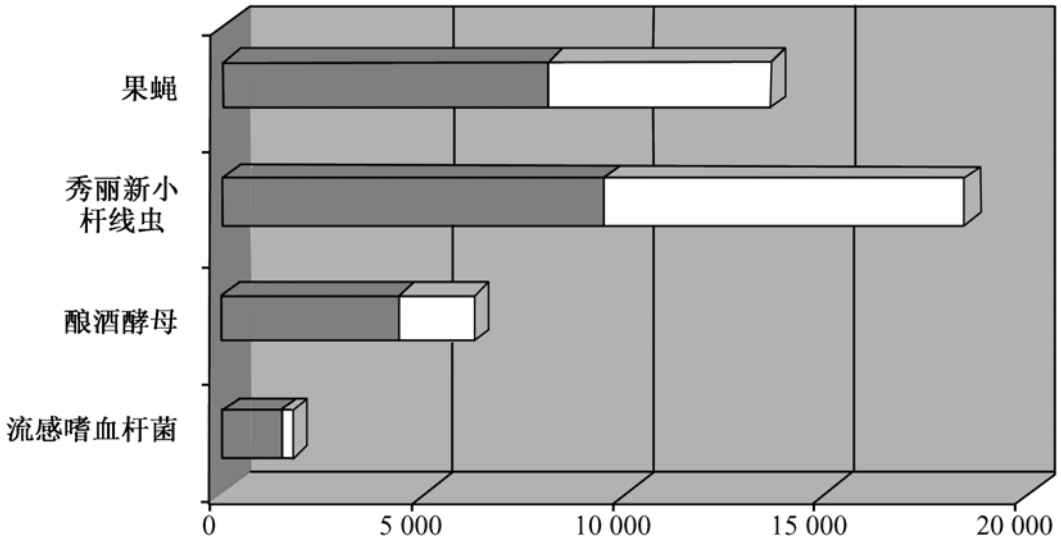

图 2.3　根据流感嗜血杆菌（1 709 个基因）、酿酒酵母（6 241 个基因）、秀丽新小杆线虫（18 424 个基因）和果蝇（13 601 个基因）基因预测的蛋白质产物

实心柱表明为单一蛋白质编码的基因，空心柱表示为平行进化同源物编码的基因。

对这些生物的核心蛋白质组的研究揭示出两个令人感兴趣的方面。首先，一种生物的复杂性和基因组中基因数量之间的关系是复杂的。当然，酵母要比细菌的基因数目多，比蠕虫和蝇类少。然而，蝇类（果蝇）是比蠕虫（线虫）要复杂得多的生物，它却有较少的基因（蝇类 13 601 个，蠕虫 18 424 个），有较少的核心蛋白质组（蝇类有 8 065 个特有蛋白质，蠕虫则有 9 543 个）。这说明生物复杂性并不是因为数量较大的基因数目，而更为复杂的基因调节和蛋白质产物的功能可用以说明蝇类的更高的复杂性。第二，平行进化同源物的数量在蠕虫和蝇类中显著增加。这反映了蠕虫和蝇类几乎大约一半的基因是其他基因的复制。含有复制基因的基因家族常常形成相同染色体上的基因簇。

最近完成的人类基因组序列测定表明人类基因组有 30 000～40 000 个基因。人类的复杂性与蠕虫相比有极大的不同，人类基因组仅是蠕虫编码基因的两倍，这确实令人感到惊奇。人类基因组中的单一基因与平行进化同源物的数目还不能

确切估计。然而，已经公认的是人类的复杂性在于人类蛋白质组的多样性，而不是人类基因组的大小。

2.6 基因表达、密码子偏倚和蛋白质水平

研究蛋白质组的关键课题之一是细胞中某个特定蛋白质的表达水平。蛋白质表达水平变化极大，从几个拷贝到多于百万个拷贝。但细胞中蛋白质表达水平与其意义大小无关。中间代谢的关键酶或结构蛋白质在每个细胞中有几千拷贝或更多，而与细胞周期调节有关的某些蛋白激酶在每个细胞中仅有几十个拷贝。酿酒酵母有 6 000 个基因，根据 mRNA 水平推测，有大约 4 000 个基因在任何时间都表达。

细胞中某一蛋白质在某一特定时间的表达由下列因素控制：①基因的转录速度；②mRNA 翻译成蛋白质的效率；③细胞中蛋白质的降解速度。基因表达确实在很大程度上决定蛋白质水平。然而，有些研究表明基因表达自身并不与蛋白质水平紧密相关。这一发现也证明了前面提到的 mRNA 的翻译效率和蛋白质降解速度对细胞内蛋白质表达水平的影响，同时也指出了基因表达分析（如微阵）的局限性。

许多基因受可诱导转录因子的调节，这些转录因子又是受多种外界环境影响因素的调节。许多基因表达水平也受内部决定因素——“密码子偏倚”现象的影响。“密码子偏倚”描述了一个生物为相同氨基酸编码时偏向于使用某些密码子。所以，含有不常使用的密码子的基因倾向于较低水平表达。根据酵母基因计算的密码子偏倚值约从 0.2 到 1.0。密码子偏倚值为 1.0 时有最高水平基因表达。大多数酵母基因的密码子偏倚值小于 0.25，表明这些基因在相对低水平表达。

比较酵母某些蛋白质的蛋白质水平、mRNA 表达和密码子偏倚值，尽管在某些方面不完全一致，但可以概括如下。

- 具有低密码子偏倚值的基因倾向于低水平表达，无论根据 mRNA 表达或蛋白质水平的分析都是如此。
- 当基因的密码子偏倚值为 0.25 或更低时（如大多数基因），mRNA 水平与蛋白质水平的相关性很差（$r<0.4$）。对大多数高表达的基因（密码子偏倚值大于 0.5 的基因），mRNA 水平与蛋白质水平的相关性要高很多（$r>0.85$）。
- 长寿蛋白质比短命蛋白质（迅速降解的蛋白质）以明显较高的丰度存在。

因而尽管基因表达测定可能表明蛋白质水平的改变，但很难从基因表达水平推断蛋白质的水平。

2.7 分析蛋白质组学的结论和意义

在任何生物中，蛋白质组是所有基因产物的 30%～80%的汇集。虽然某些蛋白质以较高水平表达（每个细胞 10^4～10^6），但是大多数蛋白质都以相对低的

水平表达（每个细胞 10^1～10^2），与基因表达的绝对量无关，大多数蛋白质以多种翻译后修饰形式存在。这样就给蛋白质组学提出了巨大挑战：我们必须找到测定大量不同蛋白质种类的方法。这些蛋白质大多以相对低水平存在，很多以多种修饰形式存在。本书的下一部分描述可以用来应对这个令人生畏的问题的工具。

推荐读物

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., et al. (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37－40.

Coghlan, A. and Wolfe, K.H. (2000) Relationship of codon bias to mRNA concentration and protein length in Saccharomyces cerevisiae. *Yeast* **16**, 1131－1145.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* **19**, 1720－1730.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., et al. (2000) Comparative genomics of the eukaryotes. *Science* **287**, 2204－2215.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., et al. (2001) The sequence of the human genome. *Science* **291**, 1304－1351.

Ⅱ　蛋白质组学的工具

3　分析蛋白质组学概述

在详细讨论分析蛋白质组学之前，我们先概述一些基本方法。蛋白质分析鉴定建立在这样一个基本事实上：大多数含有 6 个或 6 个以上氨基酸的肽序列在一个生物的蛋白质组中是惟一的。换句话说，我们可以将一个 6 氨基酸肽定位于单一基因产物中。因而，如果能得到肽的序列，或者能精确测定肽的质量，就可以通过与蛋白质序列数据库的匹配来鉴定肽片段的蛋白质来源（图 3.1）。当然某些 6 肽可能定位于多个蛋白质，但典型的多次“命中”来自相关蛋白质的高度保守区域（如在第 2 章讨论的平行进化同源物）。如果可以得到定位于相同蛋白质的几个肽序列，这将加强匹配的准确性。因而分析蛋白质组学的本质是将蛋白质转换成肽，得到肽的序列，然后根据在数据库中的序列匹配鉴定相关的蛋白质。

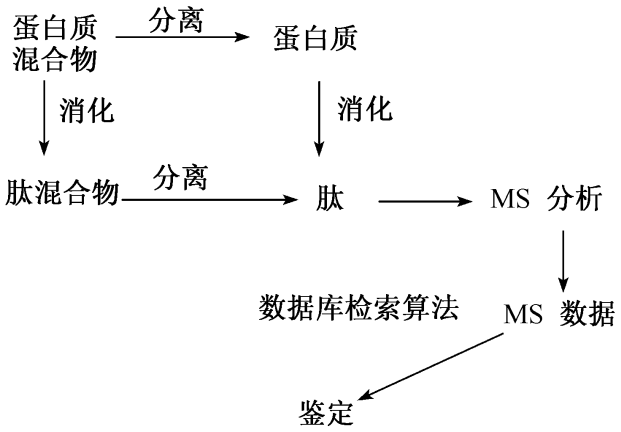

图 3.1　蛋白质组分析的基本流程图

图 3.1 描述了分析蛋白质组学的要素。大多数分析蛋白质组学的问题始于蛋白质混合物。混合物中含有不同分子质量、不同修饰作用和不同溶解度的完整蛋白质。蛋白质必须经切割消化成多肽片段，因为质谱仪往往不能直接对完整的蛋白质进行质量和序列的测定。现代 MS 仪器可以分析复杂的肽混合物，但是组分相对简化的肽混合物更有利于收集数据和分析。

因而，利用 MS 分析蛋白质混合物时，需将含有多种组分的复杂混合物分离，以获得组分较少的简单混合物。先分离出完整的蛋白质，然后消化切割成肽；也可以先将蛋白质切割成肽，在分析前分离肽。第 4 章和第 5 章描述了蛋白质和肽的分离以及将蛋白质切割成肽。

有两种类型的质谱仪可用于分析多肽。第一种类型属于基质辅助激光解吸电离-飞行时间（MALDI-TOF）仪器，主要用以测定多肽的质量。第二种类型属于电喷雾电离（ESI）-串联 MS 仪器，用于分析多肽的序列数据。在第 6 章将对这些仪器进行描述。

在特定软件的帮助下，将质谱仪数据与数据库中的肽序列进行比对以鉴定肽和肽序列。这样可基本上确定混合物中蛋白质的特性。进行这种类型的匹配比对，不需经过直接的 MS 数据分析。第 7 章至第 9 章描述这些软件工具的使用和

蛋白质鉴定方法。

分析蛋白质组学总的来说是一个测定过程。在测定中，蛋白质混合物转换成肽混合物，得到肽 MS 数据，在软件帮助下进行数据库检索，鉴定相关的蛋白质。蛋白质组学之所以作用强大是因为这种测定可用于从各种实验设计产生的多种不同蛋白质样品。蛋白质组学之所以是多用途的是因为通过这种测定可以分析用各种“前端”实验得到的样品。这些前端实验和它们的应用是本书第Ⅲ部分的主要内容。

4　蛋白质和肽的分析分离

4.1　概　　述

这一章描述用于 MS 分析的蛋白质样品的分离方法。在蛋白质组分析的这一阶段必须考虑以下两个方面（图 4.1）。首先，需将完整蛋白质转换成肽。常利用蛋白水解酶对蛋白质进行消化。第二，需将高度复杂的蛋白质和肽混合物分离形成较简单的混合物。这样 MS 仪器能更好地获得混合物各组分的有用数据。这两个步骤没有固定的次序。可以先分离蛋白质，然后消化并分析肽。也可以先将复杂的蛋白质混合物消化成肽，然后分离肽。在这里将讨论每一种方法的优缺点。

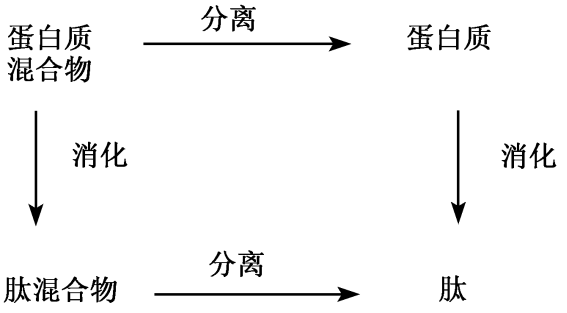

图 4.1　蛋白质组学分析中蛋白质的分离和消化

4.2　复杂蛋白质和肽混合物

在讨论分离和消化方法之前，让我们先考虑复杂蛋白质混合物的问题。MS 仪器能够从相对复杂的混合物样品中获得肽数据。然而，当样品混合物的复杂程度降低时，MS 能鉴定很多的肽片段。样品复杂性和如何处理样品的复杂性可以比作印刷一本书。假如将书中所有的字都印在一张纸上，会产生一张被油墨弄得基本上全黑的纸；而把要印刷的字分配到多张纸上，降低复杂性，可以很容易地阅读每一张纸上的字。对蛋白质和肽的分离我们采用类似的方法，基本上是以“一次一张纸”的方式将肽混合物引入 MS，让仪器尽可能地阅读纸上的内容。

在分离不同类型的蛋白质和肽时，应该先考虑在蛋白质组分析中需要处理的不同蛋白质和肽的数量。根据已知人类基因组的数量，一个人细胞中含有约 20 000个不同表达水平的蛋白质。假定其平均大小为 50 kDa，含有平均水平的赖氨酸和精氨酸，那么每个蛋白质产生约 30 个胰蛋白酶肽。这样一个细胞的蛋白质可产生约 600 000 个胰蛋白酶肽。下文中将看到，这一数字即使对最有效的多维蛋白质和肽分离策略也是一个极大的挑战。

4.3　从生物样品抽提蛋白质

在实际研究中，我们首先是收集生物样品：一块组织、一个平板上培养的细

胞、一瓶细菌、一片叶子等等。样品通常经研磨、匀浆、超声破碎或其他的破裂方法，产生在水缓冲液或悬浮液中含有细胞、亚细胞组分和其他细胞碎片的羹汤般黏稠物质。通过若干技术从这种羹汤般黏稠物中抽提蛋白质。对于蛋白质组分析，目标是回收尽可能多的蛋白质，并尽可能地减少其他生物材料（如脂类、纤维素、核酸等）的污染。抽提蛋白质时常用到以下试剂：

• 去污剂（如 SDS、3-［（3-胆胺丙基）二甲基氨］-1-丙烷磺酸酯（CHAPS）、胆酸盐、吐温），这些试剂有助于溶解膜蛋白质，并有助于膜蛋白质与脂类的分离。

• 还原剂［如二硫苏糖醇（DTT）、巯基乙醇、硫脲］，用于还原二硫键或防止蛋白质氧化。

• 变性剂（如尿素和酸），用于改变溶液离子强度和 pH，破坏蛋白质-蛋白质相互作用，破坏蛋白质的二级和三级结构。

• 酶（如 DNase，RNase），用于消化污染的核酸、糖和脂类。

上面列出的这些试剂可以不同的方式结合使用，生物学各个领域的研究者发展了从各种样品类型（如叶子和培养的细胞）中抽提蛋白质的方法。在某些方法中，通常使用蛋白酶抑制剂防止蛋白酶降解蛋白质。简言之，有多种从生物样品中抽提蛋白质的方法。

必须注意某些试剂可能干扰蛋白质组分析。例如，丝氨酸蛋白酶抑制剂苯甲基磺酰氟（PMSF）常用来在组织加工时防止蛋白质降解。然而，在某些蛋白质样品中残存的 PMSF 可能抑制胰蛋白酶的消化作用。类似地，去污剂可能干扰蛋白质分离和蛋白酶解消化。了解样品的制备过程，对样品分析的成败是很重要的。

4.4 完整蛋白质的分离

广泛使用的，用于完整蛋白质分离的三种主要方法是 1D-SDS-PAGE、2D-SDS-PAGE 和制备等电聚焦（IEF）。还有一些其他方法，特别是 HPLC［反相（RP）、大小排阻、离子交换或亲和层析］，也用于完整蛋白质的分离。分离完整蛋白质是利用了完整蛋白质的物理性质（特别是等电点和分子质量）的不同。样品混合物可以分成较少的组分（如在 1D-SDS-PAGE 和制备 IEF 中），或分成多种组分（在 2D-SDS-PAGE 中有许多点）。这些组分分别进行蛋白酶消化，然后进一步分离肽片段或直接进行肽 MS 分析。

4.5 1D-SDS-PAGE

这种在蛋白质化学中广泛使用的单向分析分离方法也可用于蛋白质组分析。在 1D-SDS-PAGE 中，蛋白质样品溶解在通常含有巯基还原剂（巯基乙醇或

DTT）和 SDS 的上样缓冲液中（图 4.2）。基本原理是 SDS 与蛋白质结合，以与分子质量约恒定的比例将负电荷（来自 SDS 硫酸基团）传递给蛋白质。高电压下，蛋白质-SDS 复合物在交联的聚丙烯酰胺凝胶上迁移，其速度取决于它们穿越凝胶孔基质的能力。蛋白质按照分子质量次序分离形成条带。

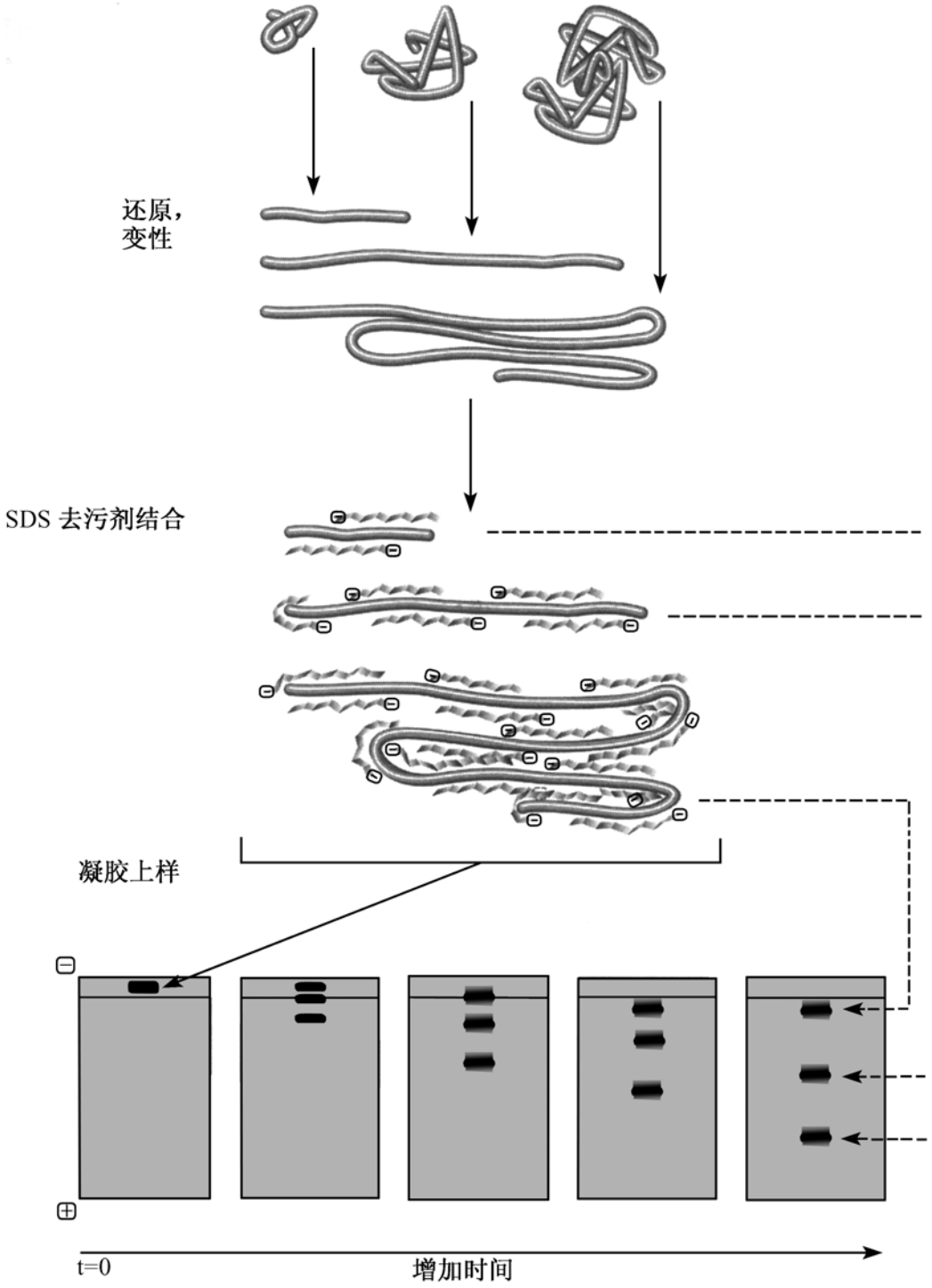

图 4.2　1D-SDS-PAGE

样品的 1D-SDS-PAGE 在交联（即丙烯酰胺的聚合）度为 5%～15%的凝胶上电泳，较大蛋白质在低度交联凝胶中较容易通过。可以根据样品中蛋白质的预测特征来选择交联度。例如，含有低分子质量蛋白质的样品可以在较高交联度的

凝胶上得到较好分离。也可以选择梯度凝胶，交联度从凝胶的顶部到底部逐渐增加。梯度凝胶可对大范围分子质量的蛋白质进行较好的分离。

用 1D-SDS-PAGE 得到的分离度是相当有限的，显示含有单一蛋白质的条带实际上可能含有多种蛋白质分子。细胞粗提物的电泳凝胶上一个跨越约 5 kDa 范围的凝胶切片可能含有几十到几百个不同的蛋白质。甚至一个“纯化的蛋白质”也可能含有多种蛋白质分子形式。比较蛋白质样品的 1D 和 2D-SDS-PAGE 可以很明显的看到这一点。1D-SDS-PAGE 分析常显示出看上去纯净的单一条带，而相同样品的 2D-SDS-PAGE 可将相同分子质量条带分解成具有不同等电点的多个点。这可能反映蛋白质的翻译后修饰，而化学修饰几乎不影响 SDS 结合或在聚丙烯酰胺凝胶上的迁移。

由于蛋白质分离的目的是降低蛋白质混合物的复杂性，据前面提到的 1D-SDS-PAGE 的局限性，使其似乎在蛋白质组分析中用处不大。而实际上这种分离方法的使用取决于样品的复杂性。大多数 1D-SDS-PAGE 在长度为 5～15 cm 的泳道分离蛋白质。这可以很容易将凝胶切割成 5～50 条凝胶切片。对于高度复杂的蛋白质混合物，如完整细胞抽提物，每一凝胶切片可能仍含有许多不同的蛋白质，这样获得的样品仍不够简化。然而，用于蛋白质组分析的许多样品并不是完整细胞抽提物或类似的复杂混合物。例如研究蛋白质-蛋白质相互作用（将在后面几章讨论）的蛋白质组学方法可能含有相对较少的蛋白质。同样，许多生物体液［如脑脊液（CSF）、肺内衬液（lung-lining fluid）］含有较少蛋白质，对于这些混合物的预先分离，1D-SDS-PAGE 可能是相当合适的。

4.6 2D-SDS-PAGE

这种分离方法与蛋白质组学是同义的，一直是分离高度复杂蛋白质混合物的最好方法。2D-SDS-PAGE 实际上是两种不同分离方法的结合。首先，根据等电点用 IEF 分离蛋白质。第二，在聚丙烯酰胺凝胶上电泳进一步分离聚焦的蛋白质（图 4.3）。2D-SDS-PAGE 在第一向电泳根据等电点的不同，第二向电泳根据分子质量的不同分离蛋白质。

尽管 2D-SDS-PAGE 是分离复杂蛋白质混合物的最有效方法，但是在 20 世纪 70 年代早期采用后的许多年来，没有得到广泛使用。这反映了：①进行 IEF 步骤的技术相对困难；②使等电聚焦的蛋白质进入 SDS-PAGE 凝胶的困难。在 2D-SDS-PAGE 的最初形式中，IEF 步骤依赖于“管式凝胶”，这种凝胶需要许多技巧进行操作。而且管式凝胶中的 pH 梯度很难重复。细软管式凝胶中含有的等电聚焦蛋白质很难有效地转移到 SDS-PAGE 平板凝胶上。因而，2D-SDS-PAGE 难于操作，重复性差。

新式 2D-SDS-PAGE 系统的引入大大改善了这一状况。新系统使用固相 pH 梯度（IPG）胶条和相对简易的硬件，能很容易地从 IPG 胶条将蛋白质转