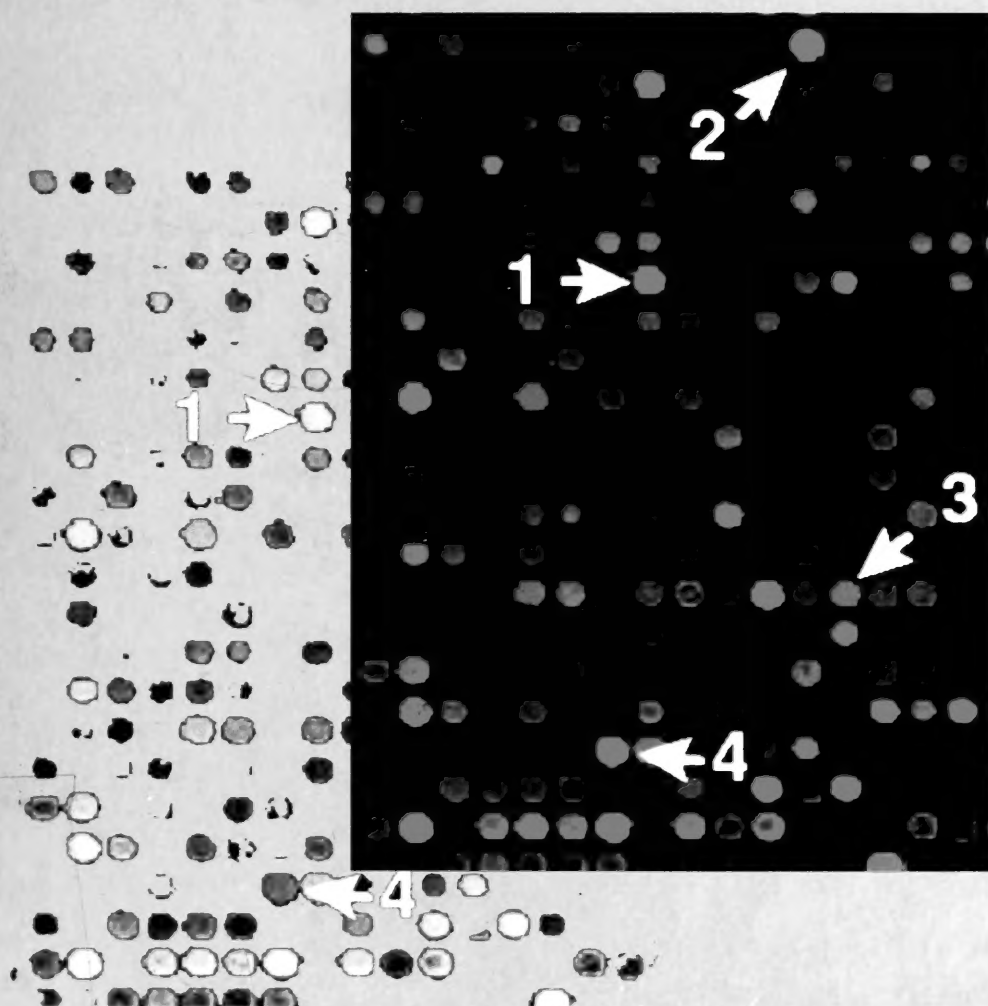



现代生物技术译丛

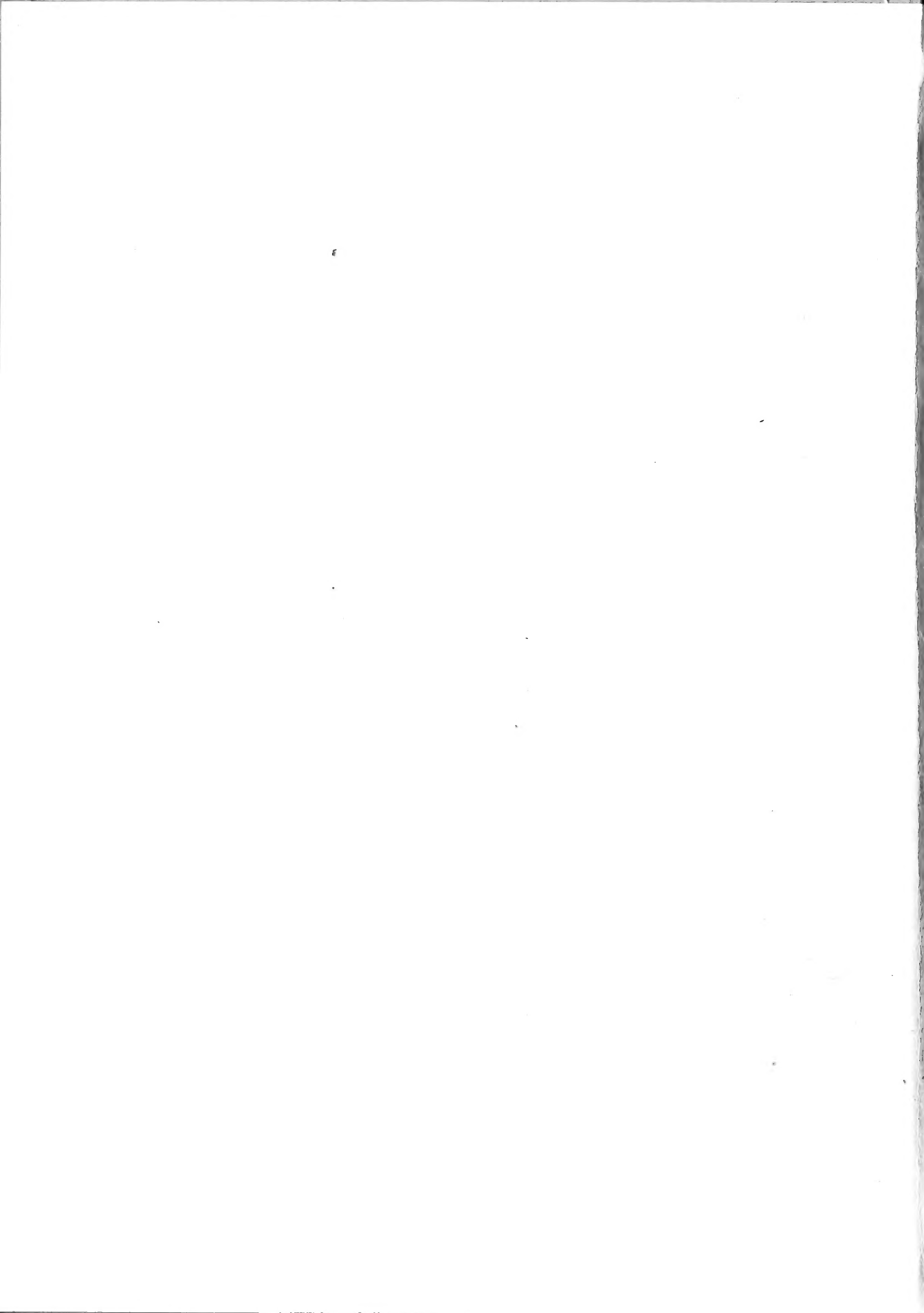
Proteomics: from Protein Sequence to Function

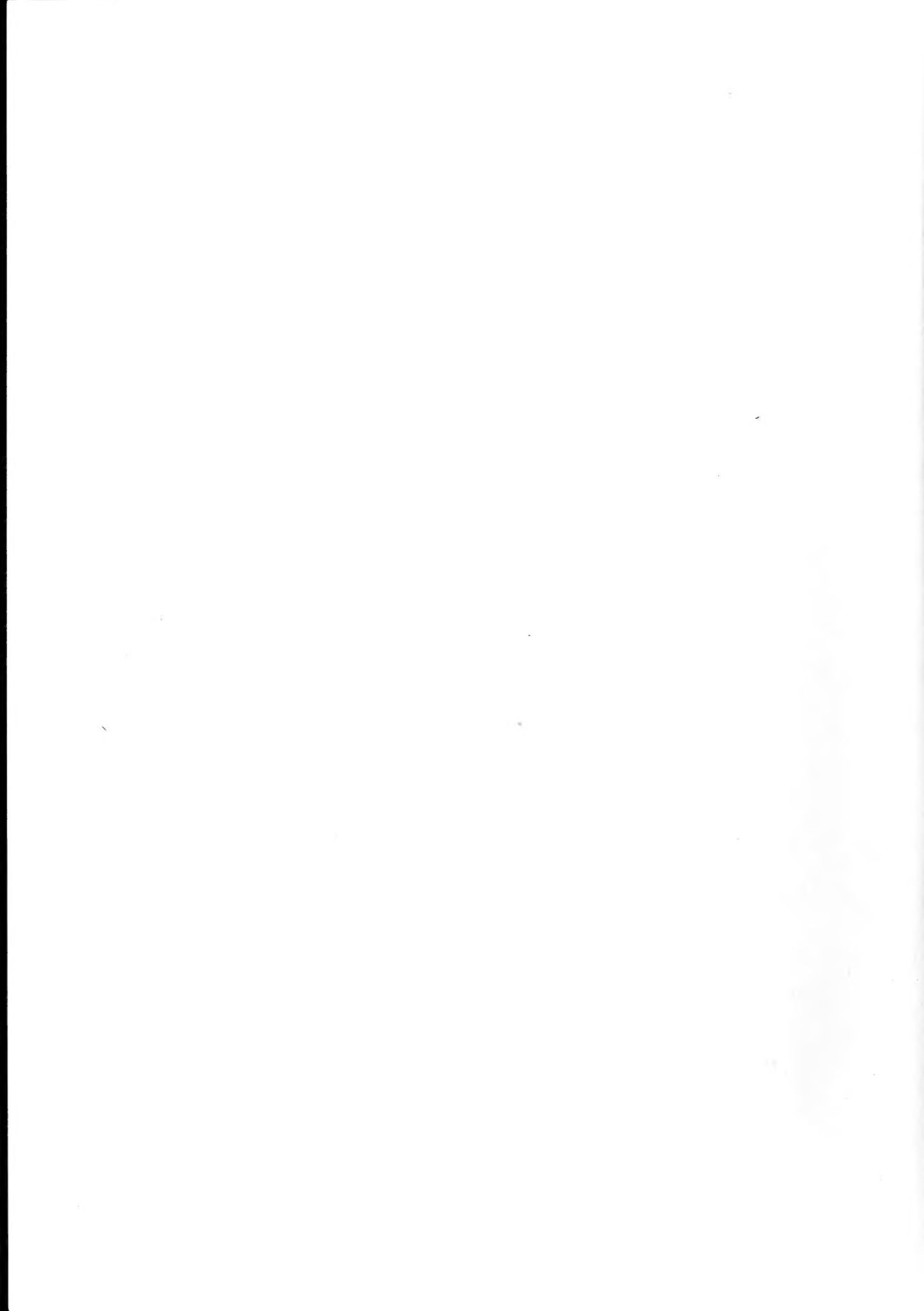
蛋白质组学：从序列到功能

[英] S.R.Pennington M.J.Dunn 编著
钱小红 等译



 科学出版社





中科院植物所图书馆



S0000099

501522
125

现代生物技术译丛

蛋白质组学

从序列到功能

[英]S. R. Pennington M. J. Dunn 等著

钱小红 等译



科学出版社

2002

27481

内 容 简 介

蛋白质鉴定技术的发展与基因组测序计划的实施与相继完成,使蛋白质组学得以诞生和发展。蛋白质组学是在基因组学的基础上研究蛋白质的表达与功能的科学,是建立在从 cDNA 阵列 mRNA 表达谱的基因功能分析,基因组范围的酵母双杂交,蛋白质与蛋白质相互作用分析到蛋白质表达、测序和结构分析等诸多不同实验方法相互融合基础上的科学。

本书着重探讨如下内容:基因组学与蛋白质组学的相互关系; mRNA 表达谱现行的研究方法;蛋白质组研究方法,如双向聚丙烯酰胺凝胶电泳、蛋白质检测、蛋白质鉴定的质谱方法、蛋白质表达定量分析方面质谱方法与图像分析方法的进展;蛋白质组研究的自动化;蛋白质组分析在药物开发中的应用;噬菌体抗体在蛋白质组研究中的应用;糖生物学与蛋白质组学;蛋白质组技术体系在院校实验室的建立;蛋白质组学应用于植物遗传学与育种。

本书适用于高等院校生命科学和医学专业师生及生物技术产业研发人员。

Original edition published in English under the title of *Proteomics: from Protein Sequence to Function*

©BIOS Scientific Publishers Limited, 2001

图字 01-2002-1992

图书在版编目(CIP)数据

蛋白质组学:从序列到功能/(英)彭宁顿(Pennington, S. R.)等著;钱小红等译. —北京:科学出版社,2002.9

(现代生物技术译丛)

ISBN 7-03-010747-0

I. 蛋… II. ①彭…②钱… III. 蛋白质-研究 IV. Q51

中国版本图书馆 CIP 数据核字(2002)第 060808 号

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>

源海印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2002年9月第一版 开本:787×1092 1/16

2002年9月第一次印刷 印张:16 3/4 插页:1

印数:1—3 000 字数:363 000

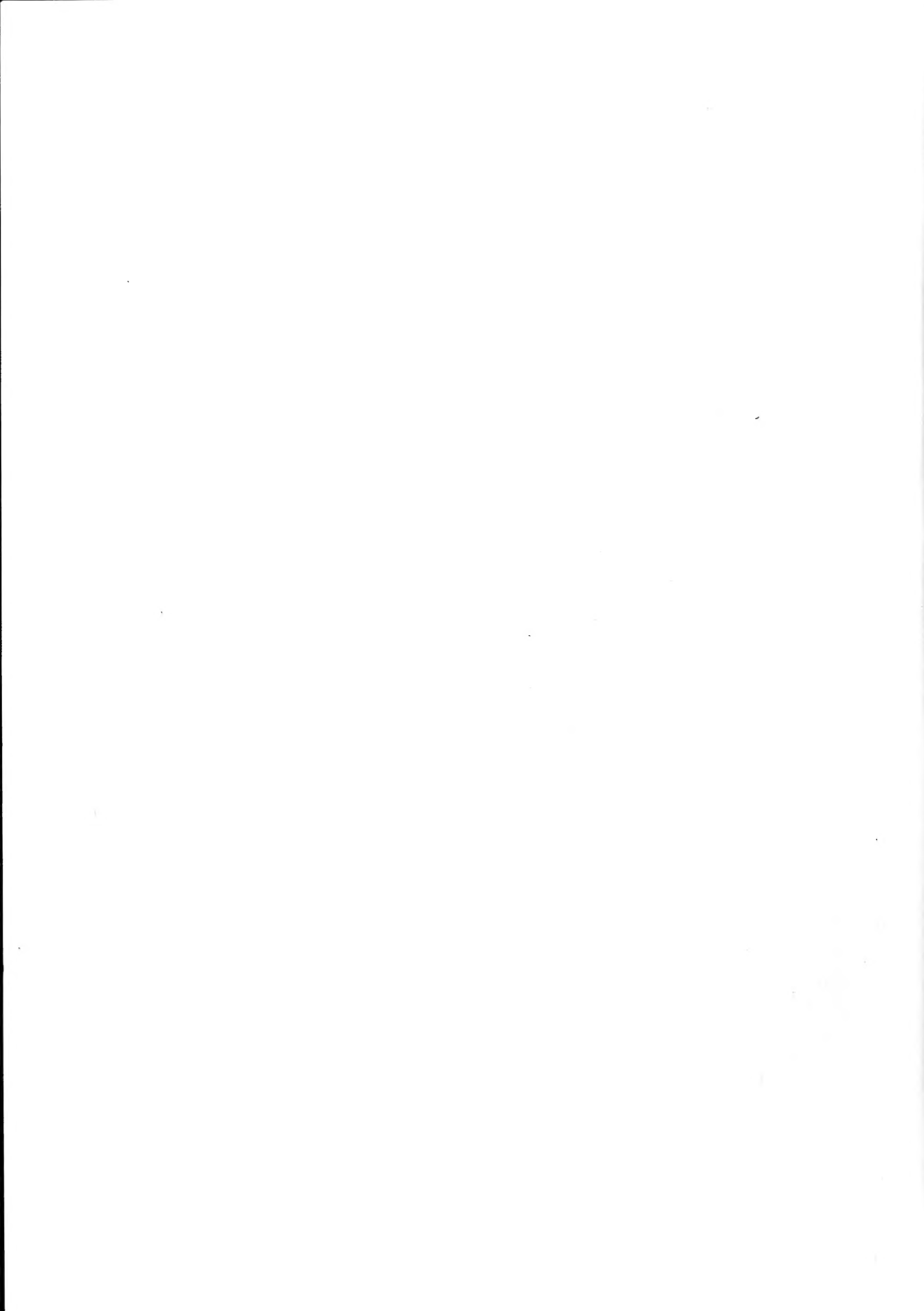
定价:45.00元

(如有印装质量问题,我社负责调换〈新欣〉)

译 者 名 单

(按汉语拼音排序)

蔡 耘	贺福初	蒋代凤	刘尚义
马晓芸	钱小红	万晶宏	王京兰
王清明	杨何义	应万涛	鱼咏涛
张令强			



编者名单

- Aebersold, Ruedi.** Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195, USA
- Blakemore, Steve.** Genomics Unit, Glaxowellcome Medicine Research Centre, Gunnels Wood Road, Stevenage, SG1 2NY, UK
- Cahill, Dolores J.** Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, D-14195 Berlin-Dahlem, Germany
- Cahill, Michael A.** Department of Molecular Biology, Institute for Cell Biology, University of Tuebingen, Auf der Morgenstelle 15, Tuebingen, D-72076, Germany
- Davison, Matthew.** 13S28 Mereside, Alderley Park, Macclesfield, SK10 4TG, UK
- Dunn, Michael J.** National Heart and Lung Institute, Imperial College of Science, Technology and Medicine, Heart Science Centre, Harefield Hospital, Harefield UB9 6JH, UK
- Eickhoff, Holger.** Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany
- Evers, Stefan.** F. Hoffmann-La Roche Ltd, PRPI-D, Building 69/16, Basel, CH-4070, Switzerland
- Goodlett, David R.** Department of Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195, USA
- Görg, Angelika.** Technische Universität München, Institut für Lebensmitteltechnologie und Analytische Chemie, FG Protetik, D-85350 Freising, Weihenstephan, Germany
- Gray, Christopher.** F. Hoffmann-La Roche Ltd, PRPI-D, Building 69/16, Basel, CH-4070, Switzerland
- Humphrey-Smith, Ian.** Department of Microbiology, Centre for Proteome Research and Gene-Product Mapping, Suite G12, National Innovation Centre, Australian Technology Park, Eveleigh, NSW, Australia
- James, Peter.** Wallenberg Laboratory, Lund University, Lund, Sweden
- Jenkins, Rosalind.** Department of Human Anatomy and Cell Biology, New Medical School, Ashton Street, Liverpool L69 3GE, UK
- Johnson, Kevin.** The Science Park, Melbourne, Cambridge SG8 6JJ, UK
- Keatinge, Lucy.** Department of Biological Sciences, Division of Environmental and Life Sciences, Macquarie University, NSW 2109, Australia
- Klose, Joachim.** Humboldt-Universität, Charité, Virchow-Klinikum, Institut für Humangenetik, Augustenburger Platz, 13353 Berlin, Germany
- Lehrach, Hans.** Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany

- Lopez, Mary.** 22 Alpha Road, Chelmsford, MA 01824-4, USA
- Nordheim, Alfred.** Department of Molecular Biology, Institute for Cell Biology, University of Tuebingen, Auf der Morgenstelle 15, Tuebingen, D-72076, Germany
- Nordhoff, Eckhard.** Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany
- O'Brien, John.** Max-Planck-Institute of Molecular Genetics, Ihnestrasse 73, Berlin, D-14195, Germany
- Oswald, Helmut.** German Heart Institute, Augustenburger Platz, Berlin, D-13353, Germany
- Packer, Nicole H.** Proteome Systems Ltd, 1/35-41 Waterloo Road, North Ryde, Sydney, NSW 2113, Australia.
- Patterson, Scott D.** Amgen Inc., Amgen Centre, MS 14-2-E, One Amgen Centre Drive, Thousand Oaks, California, CA 91320-1789, USA
- Patton, Wayne F.** Molecular Probes Inc., 4849 Pitchford Avenue, Eugene, OR 97402, USA
- Pennington, Stephen R.** Department of Human Anatomy & Cell Biology, New Medical School, University of Liverpool, Liverpool L69 3GE, UK
- Pleissner, Klaus-Peter.** Institute for Plant Genetics and Crop Plant Research, Corrensstr 3, Gatersleben, D-06466, Germany
- Plomion, Christophe.** Laboratoire de Génétique et Amélioration des Arbres Forestiers, INRA, France
- Pritchard, Kevin.** Cambridge Antibody Technology Ltd, The Science Park, Melbourne, Cambridge SG8 6JJ, UK
- Quadroni, Manfredo.** Institute of Biochemistry, University of Lausanne, Epalinges, Switzerland
- Thiellement, Herve.** Department de Biologie Vegetale, Universite de Geneve, 3 Place de l'Universite, Geneve 4, CH-1211, Switzerland
- Trower, Michael K.** Genomics Unit, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK
- Valge-Archer, Viiia.** Cambridge Antibody Technology Ltd, The Science Park, Melbourne, Cambridge SG8 6JJ, UK
- Wallace, Don M.** Medicines Research Centre, Glaxo Wellcome Research and Development, Stevenage SG1 2NY, UK
- Wegner, Susan.** German Heart Institute, Augustenberger Platz 1, Berlin, D-13353, Germany
- Wilkins, Marc.** Proteome Systems Ltd, Locked Bag 2073, North Ryde, Sydney, NSW 1670, Australia
- Zivy, Michel.** Station de Génétique Végétale, Gif-sur-Yvette, France

前 言

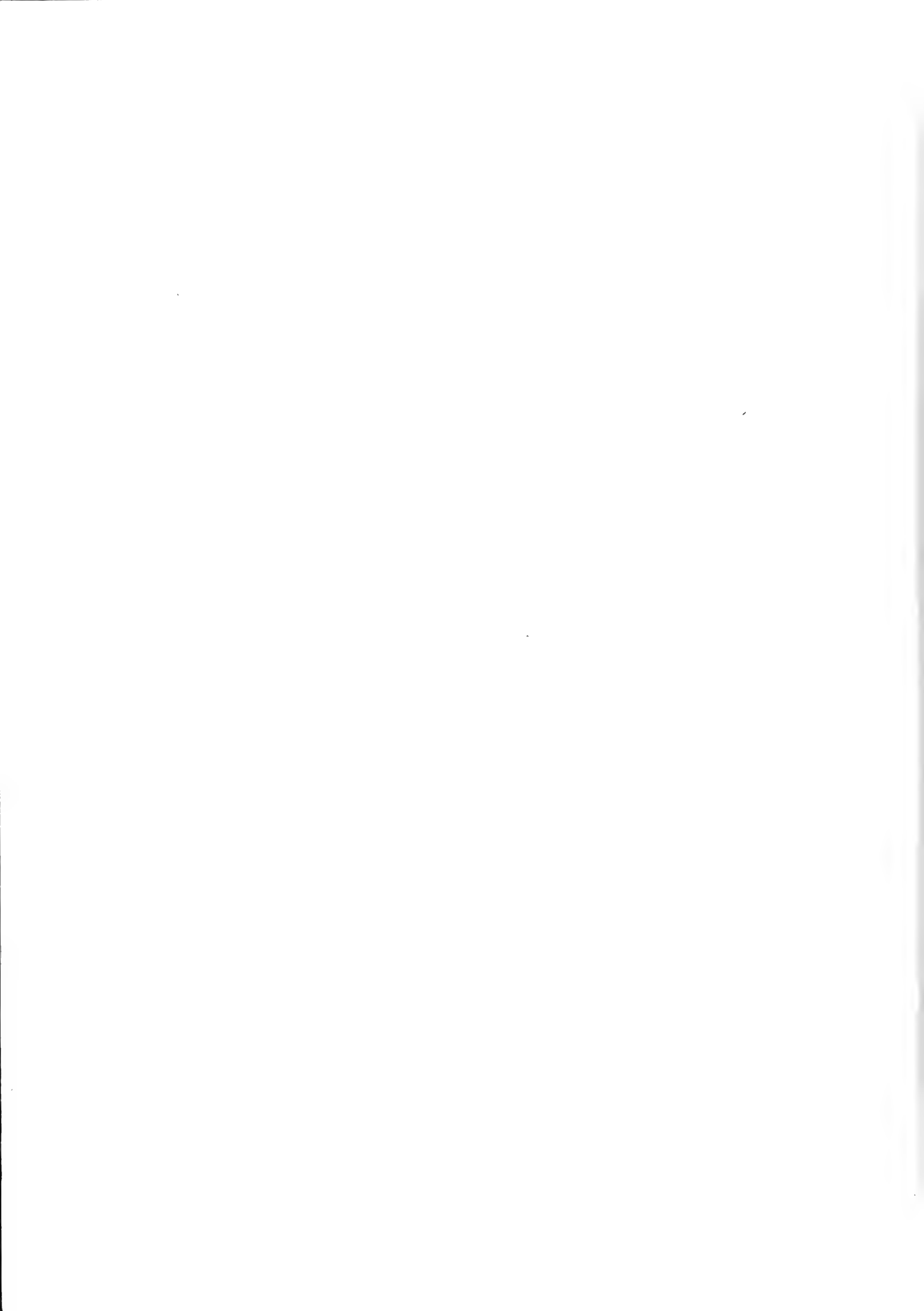
蛋白质鉴定技术的发展与基因组测序计划的实施与相继完成,使得“蛋白质组学”这一全新的研究领域得以诞生和发展。蛋白质组学是在基因组学的基础上研究蛋白质的表达与功能的科学,是建立在从 cDNA 阵列 mRNA 表达谱的基因功能分析,基因组范围的酵母双杂交,蛋白质与蛋白质相互作用分析到蛋白质表达、测序和结构分析等诸多不同实验方法相互融合基础上的科学。

《蛋白质组学:从序列到功能》一书包含了多方面的内容。主要介绍了用于蛋白质表达谱分析的二维聚丙烯酰胺凝胶电泳,用于蛋白质识别与鉴定的质谱技术以及相关的信息学。本书的作者包括了本领域的诸多领衔级人物,其中包括在二维聚丙烯酰胺凝胶电泳的发展,蛋白质的检测,凝胶图像分析,蛋白质的质谱分析,噬菌体展示抗体技术的产生与应用以及这些方法的整合及其在蛋白质组学中的应用中起过关键作用的科学家。本书力图作为对蛋白质组学研究感兴趣的人们的一本指导书,包括初学者和有实际经验的人。

本书从两个重要的篇章开始,第一章介绍基因组学与蛋白质组学的相互关系,第二章描述 mRNA 表达谱现行的研究方法。建立在从蛋白质组研究获得的信息与从 DNA 研究获得的信息相互整合基础上的这两章,为接下来有关蛋白质组研究方法描述的几章提供了必需的平台。有关蛋白质组研究方法的几章包括二维聚丙烯酰胺凝胶电泳、蛋白质检测、蛋白质鉴定的质谱方法、蛋白质表达定量分析方面质谱方法与图像分析方法的进展。现行的蛋白质组研究方法基本上还是费时、费力的,其产生的数据达到了必须采用计算机软件进行处理的规模。因而,接下来两章的主题是关于蛋白质组研究自动化的问题。尽管二维凝胶电泳是目前最有力的分析简单或复杂生物体蛋白质表达谱的平台,但它也确实有局限性,接下来一章即是有关其他潜在分离方法的介绍。最后几章分别是关于蛋白质组分析在药物开发中的应用、噬菌体抗体在蛋白质组研究中的应用、糖生物学与蛋白质组学、蛋白质组技术体系在学院实验室的建立以及蛋白质组学用于植物遗传学与育种。

诸多人的帮助使得本书得以完成。我们真诚地感谢他们。我们也感谢所有的作者,感谢他们的兴趣与热情。我们特别感谢 Scott Patterson, Mare Wilkins 和 Peter James,他们为本书提供了无价的指导和帮助。我们也感谢 BIOS 的全体人员以及那些为我们提供“后勤”支持的人们,包括 Jenni Brown, Lisa Crimmins 和 Jane Hamlett.

(钱小红 译)



目 录

编者名单	i
前言	iii
缩略语	xiii
概述	1
第一章 基因组学与蛋白质组学	6
1. 引言	6
2. cDNA 表达文库的制取、机器化阵列、PCR 膜构建、寡核苷酸指纹非冗余套的制备、DNA 芯片上的复杂杂交	8
2.1 cDNA 文库的构建	9
2.2 大规模热循环和高密度阵列的构建	10
2.3 实验室自动化和高密度阵列的新进展	11
2.4 文库归一化和用寡核苷酸指纹产生 Unigene 套	12
2.5 通过在 DNA 微阵列上的复杂杂交获得不同的差异表达谱	12
2.6 自动化图像分析	12
3. 蛋白阵列。cDNA 表达库制备, 自动化技术在蛋白阵列和芯片制备中的应用, 及蛋白芯片在蛋白质组学中的应用	13
3.1 Uniprotein 套和蛋白阵列的产生与应用	14
4. 通过高分辨二维电泳(2-DE)鉴定特定细胞类型或组织在特定时间的蛋白组成	15
4.1 蛋白质组学与基因组学	15
4.2 二维聚丙烯酰胺凝胶电泳(2-DE)	16
4.3 联合的 2-DE 数据库	17
4.4 2-DE 分析自动化	17
5. 用质谱连接当前蛋白质组学和基因组学的方法	17
6. 前景和发展	18
6.1 基因组学	18
6.2 蛋白组学	18
6.3 尚需发展之处	18
7. 总结	19
参考文献	19
第二章 基因表达的检测方法	23
1. 绪论	23
2. DNA 阵列杂交	24
2.1 cDNA 阵列	26

2.2 寡核苷酸阵列	27
3. 不基于 DNA 阵列的 mRNA 定量方法	28
3.1 基于序列测定的方法学	29
3.2 差异显示 PCR(DD-PCR)	32
4. 结论	34
参考文献	35
第三章 蛋白质组分析中的二维聚丙烯酰胺凝胶电泳	38
1. 引言	38
2. 二维聚丙烯酰胺凝胶电泳发展史	38
3. 固相 pH 梯度(immobilized pH gradient, IPG)二维电泳	39
4. 样品制备	39
4.1 可溶性样品	39
4.2 组织样品	40
4.3 细胞	40
4.4 样品分级	40
5. 溶解	41
6. 还原	41
7. 一维固相 pH 梯度等电聚焦(IEF with IPG)	42
7.1 IPG 凝胶制备	42
7.2 IPG 胶条重泡胀	42
7.3 加样和运行	43
7.4 IPG IEF pH 梯度的选择	45
7.5 聚焦时间的优化	48
8. 两维间平衡	50
9. 二维 SDS-PAGE	50
10. 分辨率	50
11. 重复性	51
12. 展望	51
参考文献	52
第四章 聚丙烯酰胺凝胶和转印膜上蛋白质的检测	55
1. 引言	55
2. 有机染料和银染	56
2.1 有机染料	56
2.2 银染	57
3. 负染	58
3.1 金属盐染料	58
3.2 锌-咪唑染料	59
4. 胶体扩散染料	59
4.1 印度墨水染料	59

4.2 胶体金属染料	60
5. 有机荧光团染料	61
5.1 共价结合荧光团	61
5.2 非共价结合荧光团	62
6. 金属螯合染料	63
6.1 比色金属螯合染料	63
6.2 发光金属螯合染料	64
7. 结论	65
参考文献	66
第五章 蛋白质鉴定和磷酸化位点分析的质谱方法	72
1. 引言	72
1.1 蛋白质鉴定途径	72
2. 质谱基础知识	73
2.1 MALDI-MS	74
2.2 ESI-MS	76
3. 以质谱为基础的相关鉴定策略	80
3.1 肽质量检索(peptide-mass searching)	80
3.2 未解析碎片离子的检索	83
4. 用质谱数据对肽段从头测序(<i>de novo</i> sequencing)	86
5. 磷酸化位点分析的各种方法	88
5.1 磷酸化分析的原理	88
5.2 磷酸化分析策略	88
5.3 磷蛋白的检测和分离	90
5.4 磷酸肽的分离	90
5.5 确定磷酸化氨基酸类型	92
5.6 磷酸化位点的确定	93
5.7 酶分析法及磷蛋白分析的未来方向	96
6. 目前及将来面临的机遇和挑战	97
参考文献	99
第六章 二维凝胶电泳的图像分析	107
1. 介绍	107
2. 数据获取	108
2.1 获取图像设备	108
3. 数字化图像加工	108
4. 蛋白质斑点检测和定量	110
5. 凝胶配比	112
6. 数据分析	114
7. 数据呈递	115
8. 数据库	117

8.1	2-DE 数据库的构建	117
8.2	比较网上分布的 2-DE 数据库	117
9.	结论	118
	参考文献	119
第七章	增强高通量蛋白质组分析:稳定同位素标记的影响	122
1.	前言	122
2.	样品制备	122
3.	二维凝胶分离和分析	123
3.1	自动化二维聚丙烯酰胺凝胶电泳(2-DE-PAGE)	123
3.2	蛋白质的扫描检测和定量	123
3.3	通过稳定同位素标记定量 2-DE 凝胶中的蛋白质	124
3.4	2-DE 凝胶蛋白点的分离	126
3.5	蛋白质酶解/碎裂	128
4.	质谱:使用质谱数据鉴定蛋白质	128
4.1	质量指纹肽谱	128
4.2	小序列标签多肽指纹谱	129
5.	质谱:使用串联质谱数据鉴别蛋白质	130
5.1	串联质谱数据采集	130
5.2	串联质谱数据库检索	131
5.3	自动化串联质谱解释和数据库检索	132
5.4	数据解释工具	133
	参考文献	134
第八章	蛋白质组学的自动化:高通量蛋白分析的技术和信息学途径	138
1.	历史回顾	138
1.1	蛋白化学家面临的困难	138
1.2	蛋白质组学的发展史	139
2.	蛋白质组学的自动化——技术解决途径	139
2.1	2-DE 胶的自动化	140
2.2	2-DE 胶的全自动	140
2.3	2-DE 自动化	143
3.	蛋白质组学的信息解决途径	144
3.1	蛋白质组数据库	145
3.2	蛋白质组学工具	148
3.3	生物信息学展望	149
4.	结语	149
	参考文献	150
第九章	瓶颈:自动化蛋白质组分析的完整途径	155
1.	引言	155

2. 二维电泳结合质谱:蛋白质组的典型范例?	155
3. 自动化概念	156
4. 二维凝胶电泳的显色	157
4.1 二维凝胶	157
4.2 富集和分部收集技术	158
4.3 显像技术	158
4.4 自动染色	160
5. 图像分析	160
6. 自动化机器人	161
6.1 斑点切取	161
6.2 斑点酶切和肽提取	162
6.3 肽脱盐、纯化,在 MALDI 靶上点样	162
7. 质谱的接口策略	163
8. 机器人模块,图像分析软件和信息数据库之间的链接	164
9. 结论	164
参考文献	164
第十章 蛋白质表达分析新方法	167
1. 引言	167
2. 功能蛋白质组研究范围	167
3. 蛋白质组分析:基于 2-DE 的战略	168
4. 除二维凝胶电泳以外的蛋白质表达分析方法	170
4.1 依赖于分离的方法	170
4.2 不依赖分离的方法:“蛋白质芯片”	172
5. 结语	176
参考文献	176
第十一章 蛋白质组分析在药物开发和毒理学中的应用	181
1. 引言	181
2. 数据的使用	182
2.1 比较分析	183
2.2 生物标志物蛋白质的检测	185
2.3 详细的基因表达调控分析	185
2.4 蛋白质功能预测	186
3. 结论与展望	186
参考文献	188
第十二章 噬菌体抗体作为蛋白质组学研究工具	190
1. 蛋白质组学方法	190
1.1 来源于 EST 数据库的蛋白质数据	190
1.2 抗体是蛋白质特异分子探针的合理选择	191
1.3 关键技术概观	191

2. 噬菌体抗体库的构建和应用	192
2.1 抗体片段在噬菌体上的展示	192
2.2 抗体库的构建	192
2.3 库的大小和多样性	192
2.4 从噬菌体抗体库中筛选特异抗体	193
2.5 高通量的筛选	194
2.6 作为试剂的噬菌体抗体	194
2.7 来源于噬菌体库中的可溶性抗体	194
2.8 以 IgG 的形式表达	195
3. 噬菌体抗体在功能基因组研究中的应用	195
3.1 “ProAb TM ”方法	195
3.2 采用生物信息学方法了解 EST 的意义	195
3.3 合成抗原	196
3.4 高通量筛选识别肽抗原的噬菌体抗体	196
3.5 高通量的 ICC 分析	196
3.6 生物信息学	197
3.7 ProAb TM 计划的应用实例	198
4. 噬菌体抗体在蛋白质组学中的应用——ProxiMol TM 技术	199
4.1 常规的选择抗膜蛋白的噬菌体抗体	199
4.2 提高膜蛋白筛选的导向选择	199
4.3 ProxiMol TM 方法及应用	200
4.4 ProxiMol TM 方法在蛋白质组学研究中的应用	203
参考文献	204
第十三章 糖生物学与蛋白质组学	205
1. 引言	205
2. 什么是糖蛋白?	206
2.1 N-糖链结构	206
2.2 O-连接结构	207
2.3 糖基磷脂酰肌醇(GPI)固定结构	207
3. 为什么蛋白质组学对糖蛋白感兴趣?	207
3.1 糖蛋白与受精	207
3.2 糖蛋白与疾病	208
4. 蛋白质组中的糖分析	209
4.1 重组糖蛋白分析	210
4.2 经分离的胶上糖蛋白	210
5. 结论	215
参考文献	216
第十四章 学术环境中的蛋白质组学	220
1. 引言	220

2. 为什么不在学术环境中进行蛋白质组研究?	220
3. 为什么要在学术环境中进行蛋白质组研究?	220
3.1 大学能否承担得起蛋白质组的研究经费?	221
3.2 如果没有学术机构,企业能承担蛋白质组学研究吗?	223
3.3 目前蛋白质组学的研究尚不健全吗?某些人看到局限性	223
3.4 学术机构的蛋白质组学研究对企业是有益的	225
4. 大学的蛋白质组学研究提供许多优势	225
4.1 大学能对企业的资源进行迅速而灵活的反应,且价格合理	225
4.2 大学愿意与临床样本接近	225
4.3 大学不仅研究医学应用	226
4.4 大学的基础研究将被公众接近	226
4.5 大学培养学生	226
5. 大学的“让产易股(spun-off)”公司	226
6. 结论	226
参考文献	227
第十五章 作为植物遗传和育种工具的蛋白质组学	229
1. 引言	229
2. 遗传多样性分析	230
2.1 种内和种间的遗传差异	230
2.2 品种、品系和栽培品种的区别	231
3. 基因组的表达	232
3.1 突变特征	232
3.2 器官和发育阶段的变异性	233
3.3 非生物胁迫反应蛋白的鉴定与分析	233
4. 遗传作图和候选蛋白	235
4.1 蛋白标记的遗传作图	235
4.2 蛋白数量位点(PQL)和候选蛋白	236
5. 方法学进展和植物蛋白数据库	238
6. 总结、前景展望	239
参考文献	239

缩 略 语

ABA	abscissic acid	脱落酸
AFLP	amplified fragment length polymorphism	扩增片段长度多态性
ANS	1-anilino-8-naphthalene sulphonate	1-苯胺基-8-萘炔磺酸酯
APAF	Australian Proteome Analysis Facility	澳大利亚蛋白质组分析实验室
ASRI	ABA/water stress/ripening related protein	脱落酸/水分胁迫/催熟相关蛋白
bis-ANS	bis(8-toluidino-1-naphthalene sultonate)	双(8-甲苯氨基-1-萘磺酸盐)
BLAST	basic local alignment search tool	基本的碱基配对搜索软件
BSA	bovine serum albumin	牛血清白蛋白
CAD	collision-activated dissociation	碰撞活化解离
Cam	Chloramphenical	氯霉素
CCD camera	charge-coupled device camera	电荷耦合照相装置
CDNA	complementary DNA	互补 DNA
CE	capillary electrophoresis	毛细管电泳
CG	candidate gene	候选基因
CHAPS	3 [(cholamidopropyl) dimethylammonio]-1-propane sulphonate	3-[(3-胆酰胺丙基)-二乙胺]-丙磺酸
CID	collision-induced dissociation	碰撞诱导解离
CP	candidate protein	候选蛋白
CS	Chinese spring	中国春(小麦品种)
CSF	cerebrospinal fluid	脑脊液
CyA	cyclosporine A	辐孢孢属 A
CZE	capillary zone electrophoresis	毛细管区带电泳
DbEST	expressed sequence tag database	表达序列标签数据库
DD-PCR	differential display PCR	差异显示 PCR
1-DE	one-dimensional polyacrylmide gel electrophoresis	一维聚丙烯酰胺凝胶电泳
2-DE	two-dimensional polyacrylmide gel electrophoresis	二维聚丙烯酰胺凝胶电泳
DHFR	dihydrofolate reductase	二氢叶酸还原酶
DGE	differential gene expression	基因表达差异
2D-PP	two-dimensional phosphopeptide	二维磷酸肽作图

DMAA	<i>N, N</i> -dimethyl acrylamide	<i>N, N</i> -二甲基丙烯酰胺
DNA	deoxyribonucleic acid	脱氧核糖核酸
DSN	database spot number	数据库中点数目
DTT	dithiothreitol	二硫苏糖醇
EDTA	ethylenediaminetetraacetic acid	乙二胺四乙酸
ELISA	enzyme-linked immunosorbant assay	酶联免疫吸附测定
ESI	electrospray ionization	电喷雾电离
EST	expressed sequence tags	表达序列标签
FACS	fluorescence-activated cell sorter	荧光激活细胞分离器
FITC	fluorescein isothiocyanate	异硫氰酸荧光素
FPR	formyl peptide receptor	甲酰肽受体
FSH	follicle stimulatiting hormome	卵泡刺激素
FT-ICR-MS	fourier transform ion cyclotron resonance mass spectrometry	傅里叶变换离子回旋共振质谱
Fus	fusidic acid	羧链孢酸
FWHM	full-width half maximum	半峰宽
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	磷酸甘油醛脱氢酶
GPI	glycosyl phosphatidylinositol	糖基磷脂酰肌醇
hCG	human chorionic gonadotrophin	人绒毛膜促性腺激素
HCl	hydrochloric acid	盐酸
HIS-tagged	histidine-tagged	组氨酸标签
HIV	human immunodeficiency virus	人类免疫缺陷病毒
HPLC	high performance liquid chromatography	高效液相色谱
HRP	horseradish peroxidase	辣根过氧(化)物酶
HSP	heat shock protein	热激蛋白, 热休克蛋白
HUVEC	human umbilical vein endothelial cells	人脐静脉内皮细胞
ICAT	isotope-coded affinity tagging	同位素编码亲和标签
ICC	immunocytochemistry	免疫细胞化学
ICE	interleukin converting enzyme	白细胞介素转换酶
ICL	instrument control language	仪器控制语言
IEF	isoelectric focusing	等电聚焦
IgG	immunoglobulin G	免疫球蛋白 G
IMAC	immobilized metal affinity chromatography	固相化金属亲和色谱
IMAGE	integrated molecular analysis of genomes and their expression	基因组及其表达产物综合分 析
IMS	ion mobility spectrometry	离子迁移谱
IEF	isoelectric point	等电点
IPG	immobilized pH gradiednt	固相 pH 梯度

IPG-DALT	2-D with immobilized pH gradients in the first dimension	固相 pH 梯度为第一向的二维凝胶电泳
IPTG	isopropyl- β -D-thiogalactopyranoside,	异丙基硫代半乳糖苷
IR	infrared	红外
IRMPD	infrared multiphoton dissociation	红外多光子解离
IT	ion trap	离子阱
Kan	kanamycin	卡那霉素
KCl	potassium chloride	氯化钾
LC	liquid chromatography	液相色谱
LCM	laser capture microdissection	激光俘获微切割
LC-MS/MS	liquid chromatography tandem mass spectrometry	液相色谱-串联质谱联用
LEA	late embryogenesis-abundant	胚胎形成晚期丰富蛋白
LH	lutening hormone	黄体激素
LIMS	laboratory information management	实验室信息管理
LMW	low molecular weight	低分子质量
LOG	Laplacian of Gaussian	高斯拉普拉斯
m/z	mass-to-charge ratio	质核比
MALDI-TOF	matrix-assisted laser desorption-ionization time-of-flight	基质辅助激光解吸飞行时间质谱
MAP	mitogen-activated protein	有丝分裂激活蛋白
MDPF	2-methoxy-2, 4-diphenyl-3(2H)furanone	2-甲氧基-2, 4-二苯基呋喃
MeCN	acetonitrile	乙腈
MIP	molecularly imprinted polymer	分子印记聚合物
MPI	minimal protein identifier	最小蛋白鉴定量
Mr	relative molecular mass	相对分子质量
mRNA	message RNA	信使 RNA
MS	mass spectrometry	质谱
MS/MS	tandem mass spectrometry	串联质谱
NCBI	National Center for Biotechnology Information	全美生物技术信息中心
NEPHGE	non-equilibrium pH gradient electrophoresis	非平衡 pH 梯度电泳
NP-40	Nonidet P40	非离子去垢剂 P-40
OD	optical density	光密度值
PBS	phosphate buffered saline	磷酸盐缓冲溶液
P/A	presence/absence	存在/缺乏
PAGE	polyacrylamide gel electrophoresis	聚丙烯酰胺凝胶电泳
PCR	polymerase chain reaction	聚合酶链反应

PDA	piperazine diacrylamide	二酰基哌嗪
ppm	parts per million	百万分之几
PQL	protein quantity loci	蛋白质数量位点
PS	position shift	位置迁移
PSD	postsorce decay	源后裂解
PVDF	polyvinylidene difluoride	聚偏二氟乙烯
QCM	quartz crystal microbalance	石英结晶微平衡
QTL	quantitative trait loci	数量性状遗传位点
RAPD	random amplified polymorphic DNA	随机扩增 DNA 多态性
RCE	relative collision energy	相对碰撞能量
RF	radio frequency	射频
RFLP	restriction fragment length polymorphism	限制性酶切片长度多态性
rHu G-CSF	human granulocyte colony-stimulating factors	人粒细胞集落刺激因子
RNA	ribonucleic acid	核糖核酸
RP-HPLC	reversed-phase high performance liquid chromatography	反相高效液相色谱
RT-PCR	reverse transcriptase polymerase chain reaction	反转录酶聚合酶链反应
SAGE	serial analysis of gene expression	基因表达系列分析
SAM	self-assembled monolayer	自组装单层物
SB	sulphobetaine	磺基三甲铵乙内酯
SCA	synthetic carrier ampholytes	合成载体两性电解质
scFv	single chain variable fragment	单链片段多样性
SCX	cation exchange	阳离子交换
SDS	sodium dodecyl sulfate	十二烷基硫酸钠
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis	十二烷基硫酸钠-聚丙烯酰胺凝胶电泳
SELDI	surface-enhanced laser desorption/ionization	表面增强激光解吸离子化
SIM	selected ion monitoring	选择性离子检测
SP	candidate protein	候选蛋白
SPR	surface plasmon resonance	表面等离子体共振
SSP	standard spot number	标准点数目
Str	streptomycin	链霉素
TBP	tributyl phosphine	三丁基磷
TCA	trichloroacetic acid	三氯乙酸
TEA	triethylamine	三乙基胺
TEMED	<i>N, N, N', N'</i> -tetramethylethylenediamene	<i>N, N, N', N'</i> -四甲基乙二胺

Tet	tetracycline	四环素
TFA	trifluoroacetic acid	三氟乙酸
TGF	transforming growth factor	转化生长因子
THF	tetrahydrofuran	四氢呋喃
TIC	total ion count	总离子数
TIGR	The Institute of Genomics Research	基因组学研究所
TLC	thin-layer chromatography	薄层色谱
TM	transmembrane	跨膜
Tmp	trimethoprim	三甲氧苄二嘧啶
TNF	tumor necrosis factor	肿瘤坏死因子
TQ	triple quadrupole	三级四极杆
tRNA	transfer RNA	转运 RNA
UV	ultraviolet	紫外
VCAM	vascular cell adhesion molecule	血管细胞黏附分子
WST	watershed transformation	转化分界线
WWW	World Wide Web	万维网
ZP	zona pellucida	透明带

(杨何义译 钱小红校)



概 述

蛋白质组学——迎接后基因组时代的挑战

Stephen R. Pennington and Michael J. Dunn

蛋白质组学的出现,为我们进一步了解整个生物体系——从简单的单细胞生物到复杂的生物体(如人类)提供了一种可靠的方法。用于检测蛋白表达的二维电泳技术(2-DE)和用于蛋白质鉴定的质谱技术对蛋白质组学的产生起了至关重要的作用。而对这一领域越来越多的关注加速了对全基因组蛋白质表达以及功能分析方法的不断改进与完善。

蛋白质组学的诞生实质上是依赖于基因组测序计划的成功,该计划虽未对揭示生物体的本质提供更多的信息,但是它为更加广泛而有效的实验方法的产生提供了基本的平台,而这些实验方法将为鉴定基因组编码的基因,并最终理解这些基因产物在生命活动中的调控作用提供支撑。显然,基因组学与蛋白组学的相互协调发展将为生物学家提供最好的帮助。

整个基因组测序计划的成功既令人瞩目又令人兴奋。1995年首次公布了第一个非寄生生物的全基因组序列(Fleischmann *et al.*, 1995),而到2000年7月,在公共数据库中已经有40个生物体的全基因组序列,另外还有127个原核生物和95个真核生物的序列正在分析中[wit.integratedgenomics.com/GOLD/]。迄今为止,人类基因组的第5、16、19、21、22号染色体的测序已经完成,人类基因组的精确图谱预期在2003年公布。上述的成就可能开始对医学诊断产生影响,一个多国合作用来建立单核苷酸多态性(SNP)图谱的计划已经启动(见表1的网址)。已完成全基因组测序的生物体数目正以加速度持续增加,公众对数据库、搜索工具和相关电子信息资源的访问达到空前的规模(见表1)。而这种对序列信息的利用在十年前几乎是不可想像的。例如,比较基因组学通过对某一特定生物的全基因序列与其他所有生物的基因组序列进行比较,来鉴定两者之间的差异,从而解释某些特定和重要的功能,如病原性。而结构基因组学这一崭新领域是通过X射线晶体衍射、核磁共振和其他方法来确定蛋白质的结构,并将结构与基因的序列和功能相关联。另外,基因组测序及其相关计划已经产生了mRNA表达、转录组学分析的新策略(见第二章)(Chee *et al.*, 1996; Eisen *et al.*, 1998; Gerhold *et al.*, 1999)以及利用酵母双杂交分析检测蛋白之间相互作用的方法(Fields and Song, 1989; Fromont-Racine *et al.*, 1997; Lecrenier *et al.*, 1998)。值得注意的是,用于高通量mRNA表达分析(这种分析有时是在基因组规模上)的DNA芯片和微阵列技术对基因功能的研究产生了显著的影响(Hughes *et al.*, 2000; Young *et al.*, 2000)。微阵列技术的不断改进已经使获取基因序列信息和差异基因表达数据的速度大大加快(Bowtell *et al.*, 1999; Young *et al.*, 2000)。总之,这些方法很可能使人类能识别和鉴定生物体的各种属性(包括需要的和不需要的)。

表 1 基因系列数据库与分析工具网址

网址	机构	可获得信息
ensemble. ebi. ac. uk	德国海德堡的欧洲分子生物学实验室 英国剑桥桑格中心	人类、小鼠、虫基因组注释
genome. wustl. edu/gsc/	美国圣路易斯华盛顿大学医学院, 基因组测序中心	人类、模式生物测序计划、EST 计划、操作流程和技术帮助
www. hgmp. mrc. ac. uk/	人类基因组图谱项目资源中心, 英国剑桥 Wellcome Trust 基因组中心	序列数据库和搜索引擎, 系统发生连锁分析, 可进行友好链接
www. hgsc. bcm. tmc. edu/	美国得克萨斯州休斯敦 Baylor 医学院, 人类基因组测序中心	人类、小鼠、果蝇测序计划, 人类转录数据库
wit. integratedgenomics. com/ GOLD	美国, 伊利诺伊州芝加哥联合基因组公司	调控全部基因组测序计划或进展、链接相关的网址和公共数据库
www. jgi. doe. gov/	美国加利福尼亚州 Walnut Creek 基因组共同体研究所	人类、微生物测序和作图, 功能基因组学计划
Star. scl. genome. ad. jp/ kegg	日本东京大学化学研究所, 东京基因与基因组研究中心	序列数据库和分子相互作用的相关信息
www. ornl. gov/hgmis/	美国田纳西州 Oak Ridge 国家实验室, 能源部生命科学院	链接相关研究进展报告、公共数据库及会议, 特别是关于人类基因组计划
www. ncbi. nlm. nih. gov/genome/ seq/	美国马里兰州贝斯达 NIH, 国立生物技术信息中心	序列, SNP 和文献数据库, 数据分析工具, 人、小鼠的遗传和物理图谱
www. sanger. ac. uk/	桑格中心, 英国剑桥 Wellcome Trust 基因组研究中心	有关人类基因组测序计划及其他原核、真核生物测序计划的进展
www. tigr. org/tdb/	美国马里兰州 Rockville, 基因组研究所	有关微生物、植物及人类的基因序列, 功能和分类学数据库
snp. cshl. org	美国纽约州冷泉港, 冷泉港实验室, SNP 有限公司	人类基因组的 SNP 图谱
www-genome. wi. mit. edu/	美国马萨诸塞州剑桥 Whitehead 基因组研究中心	有关人类、小鼠、大鼠及人类 SNP 数据库的遗传和物理图谱

应用于转录组学研究的 cDNA 阵列具有下列几个重要的特征: 对复杂体系的分析变得相对容易; 反应、杂交和扫描可以自动化操作; 具备可用于结果分析的有效软件。更重要的是, 一旦在实验条件下 mRNA 表达发生改变, 即可直接采用多种简便的分子生物学工具来阐明所识别基因的表达和功能。然而, 这种方法也有局限: ① mRNA 的丰度不总是与蛋白的丰度相一致; ② 现有的检测高丰度 mRNA 的方法, 其灵敏性和动力学范围不适于检测低丰度的 mRNA (具有编码最重要的调控蛋白的潜力); ③ 由 mRNA 所编码蛋白活性的调控并不取决于 mRNA 或蛋白的表达水平, 而是受其他因素的调节, 例如蛋白的亚细胞定位和 (或) 它们进行翻译后修饰的程度, 而这两者均不能通过 mRNA 丰度的测量来确定。此外, 有许多重要的生物标本, 特别是那些可能被用于人类疾病诊断的样品, 如尿液、脑脊液和血清不含有 mRNA。另外, 对人体活组织和尸体解剖标本的 mRNA 表达分析, 始终受到标本采集时间的拖延与 mRNA 降解的挑战。

cDNA 阵列的应用还有其他的局限。cDNA 阵列可被用于基因组测序已完成模式

生物,但使用该方法构建大量非模式生物的 cDNA 阵列的可行性还存在疑虑。有一些生物,如镰状疟原虫, cDNA 阵列的适用性还有待于确定。尽管存在这些局限, cDNA 阵列以及其他基于 DNA 序列的方法在基因表达和功能分析中的应用,即所谓的功能基因组学,仍将大大加强我们对生物体的了解 (Alizadeh *et al.*, 2000; DeRisi *et al.*, 1996; Perou *et al.*, 1999)。

显然,应用基因组序列数据是一件令人兴奋的事。然而,在兴奋之余,一种新的认识和看法逐渐产生,即这种以 DNA 序列为基础的方法必须以直接分析基因编码的产物——蛋白质作为补充。例如已知一个基因组的序列并不代表可以识别这个基因组所编码的蛋白。因此,尽管基因组序列能够被用于预测可读框,但这些预测通常不准确。同样,基于 DNA 序列数据的应用以及现存的有关 DNA 序列与蛋白质功能关系的知识,很难单凭 DNA 序列信息揭示基因编码的蛋白质的功能。即使有可能,确定每个基因所对应的蛋白质的数量仍是很困难的。DNA 序列信息也不能说明蛋白质的翻译后修饰。一个蛋白可能有 400 多个化学修饰来完成其功能、定位和活性。因此, DNA 序列信息在帮助我们了解一个基因组所编码的全部蛋白质成分——蛋白质组方面仍有明显的局限性。对人类基因组来说,所预测的 50 000~100 000 个基因能编码多达 250 000~500 000 个蛋白质,因此所面临的发现蛋白质的工作量是巨大的。最终,蛋白功能的揭示需要详细而直接地分析基因组编码的蛋白质的表达、定位(组织、细胞和亚细胞器)和结构 (Dove *et al.*, 1999; Pennington *et al.*, 1997; Wilkins *et al.*, 1995)。

由于蛋白质组是高度动态的,因此很难对它进行定义。一方面,可以认为蛋白质组学是对基因组编码的所有蛋白质的鉴定;另一方面,可以认为它是对不同转录条件下蛋白表达变化的复杂分析。在很多方面,基因组学和蛋白质组学研究的策略是相同的,但这两个领域的研究结果确有很大的差异。目前,还没有一种理想的方式能够揭示最简单生物体的蛋白质组,而用于蛋白质表达比较分析的方法明显落后于 mRNA 表达的分析方法。

目前,许多研究工作的起点是通过 2-DE 对复杂混合物中的蛋白质进行分离来观察蛋白质表达的改变(见第三章) (Andersen and Andersen, 1996) 以及采用强有力的阵列分析方法进行鉴定(见第五章) (Andersen *et al.*, 1996; Eckerskorn *et al.*, 1997; Figeys *et al.*, 1998; Pappin *et al.*, 1993; Roepstorff *et al.*, 1997; Scheler *et al.*, 1998; Yates *et al.*, 1998)。近年来在蛋白质鉴定方面的进步应归功于基质辅助激光解析/电离质谱(MALDI)和电喷雾质谱(ESI)分别应用于肽质量指纹谱的分析和序列分析(见第五章和七章) (Andersen *et al.*, 1996; Figeys *et al.*, 1998; Pappin *et al.*, 1993; Roepstorff *et al.*, 1997; Yates *et al.*, 1998)。2-DE 能同时分离数千个蛋白 (Klose 1999),尽管这个方法本身存在明显的局限性(第十章),但它在构建大量蛋白质的表达谱并用于蛋白质的鉴定方面有着不可替代的作用。

蛋白质组的方法已应用于后基因组研究,如何真正达到研究目的是需要解决的问题。一个平台技术已经出现,即 2-DE,但它仍需要不断的完善。我们试图通过本书各章的介绍使读者对现有方法及其局限性有一个全面的了解。对现有方法以及在蛋白质检测(第四章)、图像分析(第六章)、定量质谱(第七章)、2-DE 与蛋白质鉴定的自动化(第八章、第九章)以及与信息学的整合(第九章)方面的发展和應用均有丰富且通俗

易懂的介绍。对蛋白质组学在新药开发、毒理学（第十一章）、植物遗传学（第十五章）中的应用以及蛋白质组学与糖生物学（第十三章）也进行了阐述。对未来可能替代2-DE（第十章）的方法、蛋白质组平台补充技术的发展，如噬菌体展示抗体技术（第十二章）也有介绍。

我们相信，本书既对刚刚涉足蛋白质组学研究领域的读者会有所帮助，同时也为在此领域已积累一定经验的读者进一步了解蛋白质组这一不断发展的新兴领域提供详细而有价值的信息。为此，我们向这些读者表示真挚的谢意。

（万晶宏译 钱小红校）

参考文献

- Alizadeh, A.A., Eisen, M.B., Davis, R.E., *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**: 503–511.
- Andersen, J.S., Svendsen, B. and Roepstorff, P. (1996) Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: Powerful analytical tools in recombinant protein chemistry. *Nature Biotechnol.* **14**: 449–457.
- Anderson, N.G. and Anderson, N.L. (1996) Twenty years of two-dimensional electrophoresis: Past, present and future. *Electrophoresis* **17**: 443–453.
- Bowtell, D.D.L. (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nature Genet.* **21**: 25–32.
- Chee, M., Yang, R., Hubbell, E., *et al.* (1996) Accessing genetic information with high-density DNA arrays. *Science* **274**: 610–614.
- DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**: 457–460.
- Dove, A. (1999) Proteomics: translating genomics into products? *Nature Biotechnol.* **17**: 233–236.
- Eckerkorn, C., Strupat, K., Schleuder, D., Hochstrasser, D., Sanchez, J.C., Lottspeich, F. and Hillenkamp, F. (1997) Analysis of proteins by direct scanning infrared-MALDI mass spectrometry after 2D PAGE separation and electroblotting. *Anal. Chem.* **69**: 2888–2892.
- Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**: 14 863–14 868.
- Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245–246.
- Figgs, D., Gygi, S.P., Zhang, Y., Watts, J., Gu, M. and Aebersold, R. (1998) Electrophoresis combined with novel mass spectrometry techniques: powerful tools for the analysis of proteins and proteomes. *Electrophoresis* **19**: 1811–1818.
- Fleischmann, R.D., Adams, M.D., White, O., *et al.* (1995) Whole-genome random sequencing and assembly of haemophilus-influenzae Rd. *Science* **269**: 496–512.
- Fromont-Racine, M., Rain, J.C. and Legrain, P. (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**: 277–282.
- Gerhold, D., Rushmore, T. and Caskey, C.T. (1999) DNA chips: promising toys have become powerful tools. *TIBS* **24**: 168–173.
- Hughes, T.R., Marton, M.J., Jones, A.R., *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* **102**: 109–126.
- Klose, J. (1999) Large-gel 2D electrophoresis. *Meth. Mol. Biol.* **112**: 147–172.
- Leclercq, N., Foury, F. and Goffeau, A. (1998) Two-hybrid systematic screening of the yeast proteome. *Bioessays* **20**: 1–5.
- Pappin, D.J.C., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* **3**: 327–332.
- Pennington, S.R., Wilkins, M.R., Hochstrasser, D.F. and Dunn, M.J. (1997) Proteome analysis: from protein characterisation to biological function. *Trends Cell Biol.* **7**: 168–173.
- Perou, C.M., Jeffrey, S.S., van de Rijn, M., *et al.* (1999) Distinctive gene expression patterns in

- human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* **96**: 9212–9217.
- Roepstorff, P.** (1997) Mass spectrometry in protein studies from genome to function. *Curr. Opin. Biotechnol.* **8**: 6–13.
- Scheler, C., Lamer, S., Pan, Z., Li, X.P., Salnikow, J. and Jungblut, P.** (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* **19**: 918–927.
- Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F. and Williams, K.L.** (1995) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology* **13**: 19–50.
- Yates, J.R.** (1998) Mass spectrometry and the age of the proteome. *J. Mass Spectrometry* **33**: 1–19.
- Young, R.A.** (2000) Biomedical discovery with DNA arrays. *Cell* **102**: 9–15.

第一章 基因组学与蛋白质组学

Dolores J. Cahill, Eckhard Nordhoff, John O'Brien, Joachim Klose,
Holger Eickhoff and Hans Lehrach

1. 引言

蛋白质组学可以被广泛定义为生物样本中蛋白的系统分析与存档。这是一个最近两年才突显重要的领域，被认为将对生物技术产生重要影响。Blackstock 和 Weir 最近对此进行了综述 (1999)。

蛋白质组学可视为分子生物学的大规模筛选技术，目的在于归类细胞中的蛋白的整体分布，鉴定并分析感兴趣的个别蛋白，最终阐明它们的关系与功能。由于基因组学对基因的研究不能充分预测蛋白的结构或动力学，这种在蛋白水平上的分析显得尤为重要，因为绝大多数调控过程、疾病的原发过程以及大多数药物靶标均是发生在蛋白水平。这两种概念在互补的水平上（蛋白质与基因）研究了细胞的分子构架，二者又都在各自的水平上提供了增强对方效应的信息，二者存在着很强的且带有系统性的相互关系。这一章将综述当前用于高密度分析基因和蛋白表达的技术，以及我们如何运用这些技术把基因组学与蛋白质组学连接起来。本章分成如下 4 个部分：

(i) cDNA 表达文库的产生，它们的机器化 (robotic) 阵列，PCR 膜的构建，由寡核苷酸指纹产生的非冗余套 (set) 及在 DNA 芯片上的复杂杂交。

(ii) cDNA 表达亚套 (subset) 和非冗余 Unigene-Uniprotein 套的产生，这些蛋白在大肠杆菌内的高通量表达及高密度蛋白阵列及微阵列的产生。

(iii) 通过高分辨率二维电泳描述一个特定时间内、特定细胞类型或组织内的蛋白成分的特征。

(iv) 连接目前的蛋白质组学与基因组学技术，利用质谱技术鉴定经二维电泳分离的天然蛋白质群的组成，并且将它们与 Uniprotein 套与亚套所含的蛋白相关联。

目前，可以利用序列数据库内质谱确定的肽谱、序列标记或个别蛋白酶切产物的碎片-离子指纹来鉴定蛋白。显然，这种技术只对少数已知蛋白最为可靠，突破这种局限，我们提出一个新的概念：通过质谱，每一个蛋白可特异地认为包含一小套结构信息，这里，我们将这一套信息称作最小蛋白鉴定器 (MPI)。MPI 包含酶切产物的精确分子质量、碎片-离子数据及利用 MALDI-MS 获得的信息。MPI 可以由切割的二维电泳斑点和重组蛋白产生，可用于鉴定二维电泳胶上的同源蛋白 (见图 1.1)。

一旦记录在档，MPI 容许对已知和未知的基因产物进行快速识别。同时，MPI 可以容许蛋白在序列数据库中进行鉴定。具备了这些特征，MPI 可以对不同的生物样品进行二维电泳胶比较，而不依赖于这些样品的形式、分辨率及使用的分离技术。这种技术使得更多蛋白被可靠地确定。因为从二维电泳胶点上测量的 MPI 可与从表达蛋白测得的 MPI 进行比较，而不是与 DNA 序列预测的 MPI 比较 (见图 1.2)。另外，重组蛋

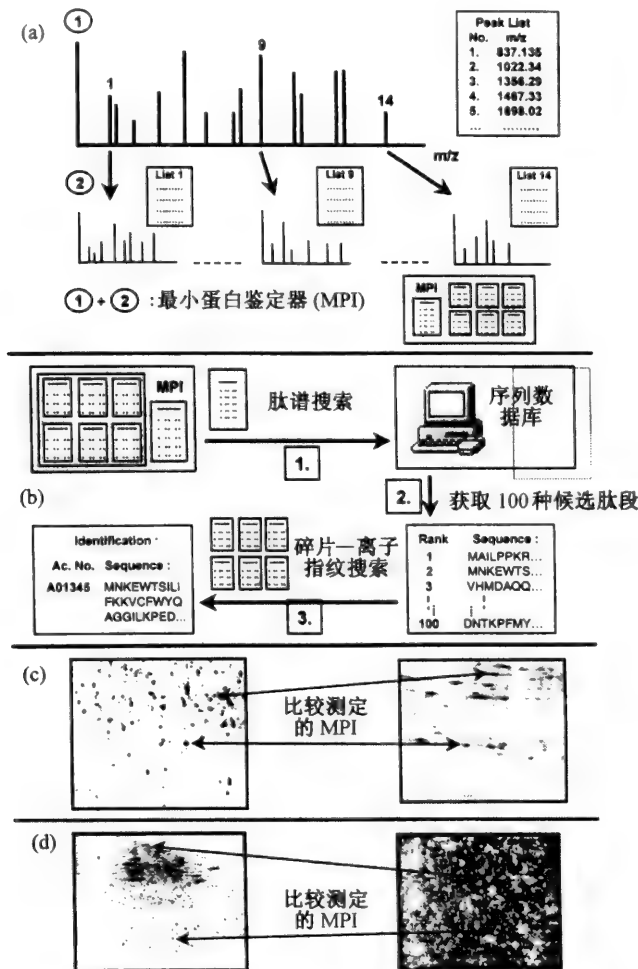


图 1.1

(a) 通过 MALDI-MS 获取 MPI。蛋白被一种特异的蛋白酶（如胰酶）裂解，切割产物的分子质量可被确定，随后，每种蛋白的较明显的切割肽段的碎片-离子谱被记录，从第一个波谱提取的肽质量图谱提供了蛋白一级结构的指纹，而碎片-离子峰列表提供的是肽氨基酸序列的指纹，这些数据被整合、贮存成 MPI，每种蛋白一个 MPI。(b) 建议采用的通过序列数据库鉴定蛋白。对一个特定的肽质量图谱搜索数据库，可获得供候选蛋白序列的列表（如 100 条），这个列表再去搜索可与记录下的碎片-离子指纹匹配的切割肽段且排好序。由于第二轮筛选只用于整个数据库的一小部分亚套，这种序列策略的优点在于搜索特异性高，搜索时间短。(c) 建议的比较二维电泳的策略。可以在计算机上比较它们的 MPI，这种比对不依赖于使用的胶形式、应用的分离技术及随后的二维电泳技术。(d) 二维电泳蛋白斑点模式和排好序的蛋白微阵列相关。将全部重组蛋白点到阵列上，MPI 已经被记录并贮存存在一个数据库内。由二维电泳分离的天然蛋白现在可以与它们的重组衍生物进行比对分析，通过已知的 MPI 与数据库比较。

白文库提供了一种对于质谱定量稳定同位素标记的二维电泳的蛋白非常好的来源（如大肠杆菌生长在¹⁵N标记的培养基中）。这种技术在第七章中有阐述。

这种方法需要产生一个 Unigene-Uniprotein 套（第 3 部分，见图 1.3），它不仅提供了产生 cDNA 微阵列的永生生化来源，并可通过复杂杂交（第 2 部分）估计 mRNA 水平。

此外, Uniprotein 套的每种蛋白都含有储存在数据库中的 MPI。

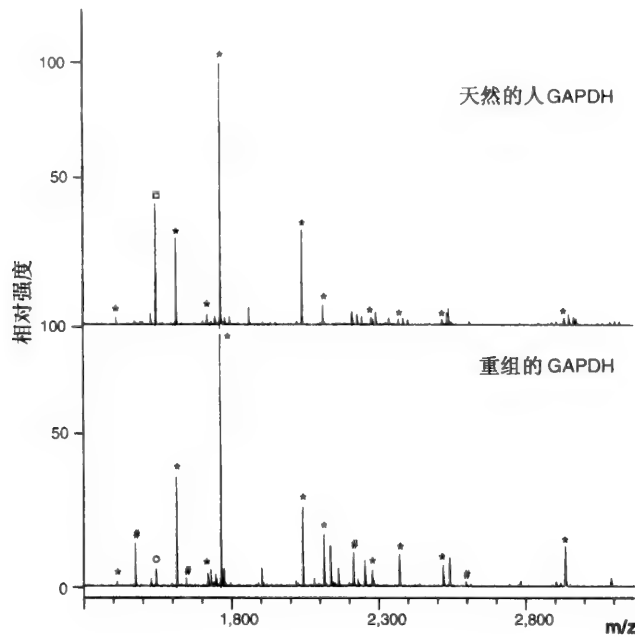


图 1.2 MALDI-TOF-MS 胰酶解天然与重组人 GAPDH 所得的肽图谱。

天然 GAPDH 从人总胎脑蛋白抽提物经大规模 2-DE 与原位切割而分离得到。上半部分是 5 μ l 纯化的过夜酶切上清的图谱。重组的人 GAPDH 由 N 端带有 RGSHis₆ 标签的基因在大肠杆菌内表达获得。利用 NTA-配基固定的琼脂糖 (Qiagen, Germany) 在变性条件下将细胞粗提物进行金属螯合层析从而获得纯化的标记蛋白。纯化的蛋白原位酶切。下半部分是从 150 μ l 总酶切上清中取 0.5 μ l 进行分析得到的图谱。标记的符号: * 天然 GAPDH 的酶切上清中检测到的胰酶切肽段。GAPDH 序列参考 NCBI 数据库注册号 120649, 1999 年 5 月 5 号公布。所有的肽段均可在重组的 GAPDH 酶切上清中检测到。# 重组 GAPDH 酶切上清中检测到的特异的胰酶切肽段。

o 在两种酶切上清中均能检测得到的不能与 GAPDH 和任何胰酶自身降解产物比对的肽段。

完成这种程序不仅需要已有技术的应用, 如高效微阵列、大规模蛋白表达及高分辨率二维电泳, 而且需要发展新技术, 包括自动化的二维电泳、样品纯化、质谱分析及封闭环数据采集与加工的软件。

2. cDNA 表达文库的制取、机器化阵列、PCR 膜构建、寡核苷酸指纹非冗余套的制备、DNA 芯片上的复杂杂交

随着人类基因组计划顺利地进行, 终结的日期临近, 人们面临着如何理解新发现的这些基因的功能。测定如此大而复杂基因组的初衷集中在表达区即 cDNA 序列。估计整个人类基因组的基因总数在 60 000 ~ 140 000 左右, 人类基因的大部分已经体现在现有的 dbEST 库内, 然而只有一小部分被揭示出功能。如 2000 年 7 月公布的消息, 人类的 entries 数量是 2 121 173 (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>) (Wolfsberg and Landsman, 1997), UniGene 套内含有 81 967 个簇 (cluster) (www.ncbi.nlm.nih.gov/uniGene/Hs.stats.shtml), 而只有 13 407 个包含至少一个已知的基因。最直接的解

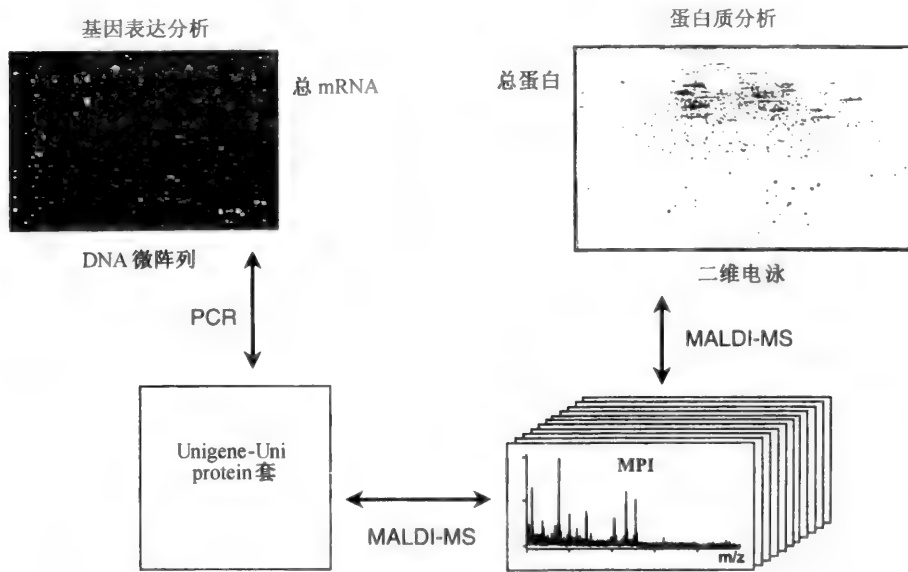


图 1.3 建议的概念，“桥”

天然蛋白与它们的基因与 RNA 水平相关，利用由质谱确定的 MPI 从 cDNA 文库提取出的 Unigene-Uni protein 套不但可以通过方便的 PCR 扩增得到单一基因而在 cDNA 微阵列上进行基因表达分析，而且可以通过亲和层析获得相应的 (His₆) 融合蛋白产物。对纯化的蛋白酶解，进行 MALDI 分析。从细胞或组织中提取的天然蛋白群可以分离，经 2-DE 鉴定，然后原位蛋白降解，MALDI-MS 测定。得到的 MPI 与从重组蛋白文库得到的 MPI 比较，成千上万的生物活性基因产物因此与它们的基因联系起来。这种联系不依赖于任何序列信息。

决结构与功能脱节的方法看起来是建立起组织的功能地位与一套基因的表达之间的直接相关性。在基因表达的最先层次即转录水平上，最能够直接反映基因的活性，因此给出 mRNA 的表达谱 (profile) 可以与研究基因功能相互补。分析基因表达的技术包括基因表达系列分析 (SAGE) (Velculescu *et al.*, 1995)、寡聚核苷酸指纹 (Herwig *et al.*, 1999; Meier-Ewert *et al.*, 1993, 1998; Poustka, 1999; Radelof *et al.*, 1998)、用复杂探针进行测序和杂交 (Schena *et al.*, 1998)。后者可以在 DNA 芯片上对遗传材料进行高通量分析，研究基因表达和发现新的疾病相关基因。高密度的 DNA 和蛋白阵列是袖珍的自动化装置，包含小的用于生物实验的平面。目前，DNA 芯片可以通过将表达分布图与在健康的和疾病的组织中观察到的表达方式比较，来研究一种给定组织的全部基因的相对表达水平。

从基因组和蛋白质组的水平对一种组织与器官的表达方式有全面的了解，需要尽可能快地平行筛选许多样品。在这个部分，在我们实验室已经被自动化与小型化的步骤可以完成高通量及高平行杂交。Clark 和 Eickhoff 等最近对此作了综述，将在下面介绍 (Clark *et al.*, 1999; Eickhoff, 1998)。

2.1 cDNA 文库的构建

分析来自许多不同组织和发育阶段的大的 cDNA 文库可能在某些脊椎动物基因组测

序完成之前让我们了解大多数基因,因为在转录之后 mRNA 总是在体内加工(剪接、编辑),cDNA 总是代表了决定正确剪接转录物的序列和正确的蛋白产物的惟一可能性。除了鉴定新基因和它们正确的剪接产物,基于来自每条基因的 cDNA 克隆的片段,cDNA 库也在分析起始材料中不同 mRNA 的丰度水平分析中起着重要的作用。最近 Clark 等综述了 cDNA 库的产生和分析(Clark *et al.*, 1999)。虽然自动化使处理和分析成百上千的克隆容易了,但即使如此大的库也总不足以鉴定代表稀有转录物的所有克隆。因此使用阵列后的 cDNA 库应能由其他的技术加以补充,以鉴定代表稀有转录物的基因。这些技术包括筛选非常大、随机的平板库,通过 cDNA 选择或外显子捕获和使用细胞/组织特异的库,以及从预测的外显子设计的引物进行 RT-PCR 所得产物的克隆。先不管它们的缺点,从一系列组织/发育阶段制取的足够大的阵列 cDNA 文库常常足以鉴定和分析有机体的大多数基因。

2.2 大规模热循环和高密度阵列的构建

对于 DNA 阵列,PCR 技术得以很好的建立并在大规模基因组分析中扮演了重要角色,商品化的循环设备能平行处理达 1536 个样本 [4×384 孔的微量滴定板 (MTP)]。我们在柏林进行分子遗传学研究的 Max-Planck 研究所已经研制了实验用的基于大量水浴的热循环样机可用来高通量扩增 DNA。一个装满 135×384 孔 MTP (51 840 个反应)的套栏 (basket) 随着空气驱动的 x/z-滑动期在三个设定好温度的水浴之间移动。

对于蛋白阵列的产生,我们已克隆 cDNA 库至组氨酸 (His) 标记的表达系统,能在 PVDF 膜上产生大的蛋白阵列 (Büssow *et al.*, 1998; 第 2 部分)。几个分格的机械手也已建成,它能在高密度格上将克隆、DNA、蛋白从微量滴定板上转移到尼龙膜或 PVDF 膜及玻璃片上。这个点样的小器械在一个 servo 控制的三轴的线性驱动系统下载有 384 钉 (pin) 的头,能精确定位至 5 μm , 可达到近似 300 点/ cm^2 的密度,枪头的直径取决于点的应用并在 150 和 450 μm 之间可变 (Lueking *et al.*, 1999)。

对于常规应用,27 648 个克隆、PCR 产物或蛋白可在单张 22.2cm×22.2cm 滤纸上点样为双份。可平行产生多达 15 张滤纸拷贝。这些点以 5×5 排列,它们每一个由主导点 (通常在中心) 和 12 对双点组成 (见图 3)。对于 DNA 阵列,鲑鱼精 DNA 总是用来做主导点,而对于蛋白阵列用墨水,这些主导点和双点消除了点识别的错误,也使得图像分析更容易。

克隆或 PCR 产物被分格至尼龙膜上的 DNA 滤膜,可重复使用至少 20 次而信号强度不会显著丢失。PVDF 膜上的蛋白滤膜阵列通常只能用一次。

在 DNA 杂交过程中湿膜难处理,这样大大增加了实验时间,它们也成为大规模筛查过程的障碍。为了解决这一点,我们发展了简化处理高密度膜的方法,使用坚硬的塑料薄片提供膜的刚性,以使这些膜更容易处理并反过来提高了实验工作的速度和效率 (Bancroft *et al.*, 1997)。

我们通常有两条途径点 DNA 或蛋白以制成高密度阵列 (Clarke *et al.*, 1999)。一个方法使用针头或尖点样,含 DNA 片段的液体通过支持物传递到无污染的钢尖。第二种方法应用压力喷墨技术,例如 cDNA 不接触表面而转移。我们已经应用了多头的压力喷射微阵列系统,它可以在一系列表面构建大的微阵列。由于用这个系统获得的点的密

度大于 2000 个克隆/cm²，通过与 Perkin Elmer 合作，研制了一种基于激光扫描原理的高分辨率检测系统。作为对传统针头点样的进一步改进，一种称为按需点滴的技术建立起来。基因样品用多道的微处理机械手吸取，它与喷墨式打印机的工作原理一样，可将杂交阵列的面积减少一到两个数量级。完整的图像分析程序决定是否产生了一个合适的点 (drop)。如果点形成的不好，喷嘴枪头将自动被清洗。一个整合的照相机决定自动处理的位置，例如在硅薄片上填洞。每个可处理单个或多点的头可容纳 100 pl。在处理器内部最近发展了基于磁珠的纯化系统，允许点样探针在处理之前的浓缩和纯化。点样的大小依赖于处理液体的表面可以在直径 100~120 μm 之间变化。阵列的密度可提高到 3000 点/cm²。这个微处理系统能够处理 “on-the-fly” 而且不超过 3 min 即可在一个方形上处理 100×100 个直径 100 μm 且每两个点中心之间的距离为 230 μm 的点。以这种密度，在载玻片表面固定一个由 14 000 个克隆组成的小型 cDNA 文库是完全可能的。由于玻璃片比膜更坚固更容易处理，这便提高了自动化的程度。

另一个微处理技术的应用是在 MALDI-MS 靶平板上准备高密度阵列探针，质谱带来了 DNA 和蛋白分析中的高速度、高通量应用。在商品化的 MS 设备中，1 s 可分析一个样品，理论上每天可详细地研究成百上千的样本。直到目前，只有样机 (prototype) 问世，使用在一张 1 cm² 的 MALDI-MS 靶上固定有几千个 DNA 片段或蛋白的微处理阵列 (见第 4 部分)。

2.3 实验室自动化和高密度阵列的新进展

分析大的基因库的第一步是从凝胶平板上挑出随机分布的克隆并将这些克隆制成阵列至微量滴定板上。我们实验室已发展了挑取机械手，它每小时挑 3500 个克隆。这些克隆通过图像分析检测以对挑选进行定位。发展的软件可以识别克隆位置并将其翻译成机械手的移动信息。最新版本的这种软件可识别直径小至 0.5 mm 的克隆并且不依赖于选择的算法，蓝白斑筛选的精确率几乎可达 100%。在进一步小型化和微阵列技术上也有了新的发展 (Eickhoff, 1998)。我们已描述分析荧光标记的生物分子如 DNA 所用的全自动多毛细管电泳系统的构建和操作 (Behr *et al.*, 1999)。在这个设备中，一个特殊的检测系统允许同时对所有 96 个毛细管进行光谱分析，它没有移动部分，但是自动的并完全可与现有系统相媲美。这个设备不需要人工干预即可处理多达 40 个微量滴定板 (96 和 384 孔)，这意味着在再次加样之前已经处理了 15 000 个样本。

为便于这些数据的系统化产生、收集和重分配，于 1989 年本系开始从一些不同文库中制备分布滤膜分格 (distribution filter grids) (Lehrach *et al.*, 1990)。为收集产生于这些材料的数据，建立了一个独立的数据库，参考文库数据库 (Reference Library Database) (Zehetner and Lehrach, 1994)。这个系统继续发展成人类基因组计划的资源中心并成为其他在阵列库上尝试系统化生成数据的模型 (如图像文库收集) (Lennon *et al.*, 1996)。逻辑上，下一步将对差异表达的 cDNA 克隆编码的蛋白产物研究其分布情况。截至目前，我们发展了与蛋白表达分析高度平行的方法，包括在合适的载体系统同时表达较大数量的 cDNA 克隆和蛋白产物的高速阵列 (Büssow *et al.*, 1998)。这些自动化系统已被用来产生高密度 DNA 和蛋白微阵列 (Lueking *et al.*, 1999) 并将在第 3 部分进一步描述。

2.4 文库归一化和用寡核苷酸指纹产生 Unigene 套

高等生物尤其是人类基因组序列的确定, 现在看来是一个可达到的目标, 但仅代表了遗传学复杂性分析的一个层次。生物学信息表达谱的确定代表了复杂性的另一个层次, 它对理解有机体功能与测序一样重要。为了能鉴定这些基因, 并计算它们的丰度, 每个 cDNA 克隆可通过典型的测序途径对应于特定的基因。基于短的寡核苷酸杂交的高效的部分测序技术叫做寡核苷酸指纹 (oligonucleotide fingerprinting), 已经发展起来用于 cDNA 库的鉴定 (Herwig *et al.*, 1999; Meier-Ewert *et al.*, 1998; Poustka *et al.*, 1999; Radelof *et al.*, 1998)。这相当于 200 个寡核苷酸序列依次与扩增的和阵列化的 cDNA 插入片段进行杂交。与不同已知探针的杂交可确定一个未知的靶, 每一克隆的杂交谱产生一个基于 DNA 序列的指纹, 相关基因因此可以组群或成簇。构建来自于 DNA 序列数据库的理论上的指纹之后, 比较每簇的保守指纹, 可以鉴定已知和未知基因的 cDNA。如果在寡核苷酸指纹之前文库克隆在表达载体中, 不仅可产生一个 Unigene 套, 也可产生一个 Uniprotein 套。

2.5 通过在 DNA 微阵列上的复杂杂交获得不同的差异表达谱

基于产生 Unigene 套, PCR 产物按 2.2 部分描述以高密度阵列点于玻璃片上。这些微阵列可以同时测量表达水平, 因此就表明所有呈现在点阵上所检测的任何组织中的基因样本在某一时间点的活性水平。当来自于不同组织或发育阶段的 RNA 或 cDNA 复杂混合物与这些 DNA 芯片杂交时, 可能得到差异的基因表达谱 (Schena *et al.*, 1998)。自动化技术使得可以通过分析复杂的表达模式, 高通量地监测这些基因的活性。发展于高密度滤膜的 DNA 微阵列现在经常是小而密致足以称其为基因芯片 (Lennon and Lehrach, 1991; Jordan, 1998; Ramsay, 1998)。使用微阵列, 已获得拟南芥 (Schena *et al.*, 1995)、酿酒酵母 (Cho *et al.*, 1998; Chu *et al.*, 1998; DeRisi *et al.*, 1997; Shalon *et al.*, 1996), 包含大规模双杂交筛选 (Uetz *et al.*, 2000), 人淋巴细胞和组织如骨髓、脑、前列腺和心脏 (Schena *et al.*, 1996) 的基因表达谱。cDNA 微阵列已被用来对诸如类风湿关节炎和尤文氏肉瘤等复杂疾病做分布分析 (DeRisi *et al.*, 1996; Heller *et al.*, 1997; Trent *et al.*, 1997; Welford *et al.*, 1998)。

在我们系进行的复杂杂交如下进行: 不像在一张玻璃片进行两个不同标记的 mRNA 的杂交, 我们的方法是只用一种标记 mRNA, 在至多 4 张不同的玻璃片上进行每种组织的表达分析。对于复杂杂交, 若不超过 10 000 个基因, 选择的玻璃规格是 2.5cm × 7.5cm。如必须分析更多数量的基因, 玻璃片的选择规格是 8cm × 12cm, 至多 50 000 条基因可被共价吸附于玻璃表面。这使重复利用 DNA 微阵列至少 5 次成为可能。杂交到每个阵列基因的信号被用来确定相应于每个基因样品中 mRNA 的相对含量。内对照包括用短的寡核苷酸的回复杂交、在阵列上使用持家基因、在阵列上使用拟南芥对照克隆以及用拟南芥尖端的组织 RNA 探针。

2.6 自动化图像分析

高通量实验室的一个重要特点是自动化大规模鉴定文库中的阳性克隆。这种自动化

分析系统的主要要求首先是自动化在滤膜上将点样模式分格，其次是确定阳性克隆。不幸的是，不同杂交滤膜显示不同的杂交特性。从这些滤膜获得的图形特征受几个因素影响，存在的问题包括由于不均一尼龙膜或高的杂交背景造成的图形上点的不平均分布。虽然人的判断离确定一个克隆是否阳性的最好方法还很远，自动化找点的算法在我们系已有发展。在当前阶段，接近 80% 的所有阳性克隆可被自动化获得。用我们的高密度蛋白阵列筛选表达的 His 标签蛋白（27 648 种蛋白复制到一张 22.2cm×22.2cm 大小的 PVDF 膜上），80% 的重阵列（rearranged）表达亚套的蛋白显示表达。使用相同的机械手技术和相同的图像分析软件，我们能分析 DNA 蛋白阵列和微阵列，并且我们目前发展的这些系统可分析 2-DE 电泳，提供的高度平行的数据获取和分析系统可提高实验工作的速度，并可进行诸如在 RNA 水平上的基因表达和固定于蛋白芯片上的储积在血清中的抗体谱之间有意义的比较，并可确定通过 2-DE 模式所揭示的相对蛋白含量。

3. 蛋白阵列。cDNA 表达库制备，自动化技术在蛋白阵列和芯片制备中的应用，及蛋白芯片在蛋白质组学中的应用

蛋白将基因组序列翻译为功能，这样使生物学过程成为可能。因此，充分理解在基因组和蛋白组水平某种组织或有机体的表达谱要求尽可能快速地平行筛选许多样本。因为即使在最简单有机体的自稳状态中也有大量的蛋白参与，这就要求用高通量的自动化方法进行蛋白分析（Uetz *et al.*，2000）。我们发现 cDNA 库能在高密度滤膜上筛选蛋白表达（Büssow *et al.*，1998），高密度膜可用于平行的 DNA 杂交、蛋白表达和抗体筛选。使用机械手技术（见第 2 部分），人胎脑 cDNA 表达库（hEx1）被阵列在微孔板上，且细菌克隆分格在 PVDF 膜上。原位表达重组融合蛋白被诱导并用抗含 His₆ 标签表位的抗体检测。这个方法扩展到从液体表达培养物中蛋白阵列自动化点样，其中使用一个固定到平床点样机械手上的转移标记（Lueking *et al.*，1999），这种方法显示，蛋白微阵列提供了非常敏感的基因表达和高通量抗体特异性筛选的方法。使用这种直径为 150 μm 而不是 450 μm 的转移标记，4800 个样本可放置于载玻片上并同时筛选，却应用最少量的介质。灵敏的和定位好的信号可以检测 250×10^{-18} mol 或 10pg 的点样测试蛋白（GAPDH）。cDNA 库的蛋白表达克隆可被可靠地检测，表达不正确的阅读框蛋白的假阳性克隆的比率可降到 11% 的低水平。这些改进的蛋白微阵列可以进行最敏感的蛋白表达和抗体特异性筛选。由于蛋白微阵列是在全基因组水平连接基因表达分析和分子结合研究的最有力工具，如果差异表达的基因可以用 cDNA 微阵列识别，相同的克隆可被同时在不同的细胞系统或通过体外转录/翻译做蛋白表达分析。在相同的蛋白微阵列上，表达克隆可被筛选，用于结合其他蛋白（如抗体）或结合从核酸到小分子配体的不同分子。这种多功能性使得蛋白微阵列成为诊断上的多用途工具。最近我们描述了它们作为高通量配体-受体相互作用研究、诊断及抗体特异性判定等有力工具的用途（Walter *et al.*，2000）。

在这部分，我们将描述非冗余蛋白阵列即所谓的 Uniprotein 阵列的产生，它们在基因组学和蛋白质组学的应用将在第 5 部分描述。高密度蛋白阵列和其他蛋白质组学技术的应用已被描述为“第二代蛋白质组学”（Humphery-Smith，1999）。这已经被定义为：阵列技术用于检测全基因组的相应全蛋白，而不依赖于分离科学（如 2-DE，质谱，

柱层析或毛细管电泳), 但更多应用更传统的分子生物学技术来进行整体细胞蛋白分析。上述描写的蛋白阵列在蛋白质组学中的应用是有效的, 它不同于在这儿描述的蛋白阵列在连接基因组学和蛋白质组学方面的应用。我们的应用包括针对 Unigene-Uniprotein 套编码的全部基因产物的 MPI 目录的产生, 和它在连接高密度表达阵列与 2-DE 实验方面的用途。

3.1 Uniprotein 套和蛋白阵列的产生与应用

直到最近, 仍然不可能使用同样的高密度阵列和自动化的技术来分析蛋白。我们通过产生 cDNA 表达文库、高密度蛋白阵列和微阵列将自动化技术应用到高通量和大规模蛋白分析中去 (Büssow *et al.*, 1998; Cahill *et al.*, 1998; Lueking *et al.*, 1999)。我们的方法是制备由 cDNA 克隆蛋白产物用于进一步的分析和使用, 如产生蛋白阵列。这需要—个高度平行的技术做蛋白表达分析, 包括大量 cDNA 克隆在一个合适的载体系统内的同步表达及高速度蛋白产物的排布, 最近由 Eichhoff 等进行了综述 (Eichhoff *et al.*, 2000)。使用机器 (Lehrach *et al.*, 1997), 可把一个人胎脑 cDNA 表达文库 (hEx1) 排布在微量滴定板上, 或把菌落网格转到 PVDF 膜上。诱导重组融合蛋白的原位表达, 用抗 6 × His 标签的抗体进行检测。使用我们的方法, 文库里的基因可在 DNA 和蛋白水平上同时进行研究, 同时提供重组基因与蛋白以制备 DNA 与蛋白芯片。这种技术同样可以通过制取与排布 cDNA 表达文库、直接建立 DNA 序列信息与单个克隆的蛋白产物之间的联系来获得重组蛋白的大规模、系统化功能研究。反过来, 可以在整个基因组水平上同样操作 (Büssow *et al.*, 1998)。这使得翻译的基因产物直接服从于高通量实验需要并且在蛋白表达与 DNA 序列数据之间建立起直接的联系。

正如第 2 部分所描述的那样, 制备一个非冗余的单克隆 (Uniclone) 或一个单基因-单蛋白对应组 (Unigene-Uniprotein set) 包含这样一些过程: 将已经克隆到蛋白表达载体 (如 Qiagen 公司生产的 pQe 载体) 中的 cDNA 文库代表的表达组“寡核苷酸指纹化”, 再将这个文库重排列成非冗余的 Unigene-Uniprotein 对应组 [Cahill and Lehrach, 1998 (专利)]。这些蛋白可以在大肠杆菌内高通量表达, 因而产生一个高密度的蛋白质阵列与微阵列。Unigene 套可以用于高密度 DNA 阵列的产生, 然后是在 DNA 微阵列上使用复杂杂交、消减杂交等进行表达谱分析。有了分化表达谱, 就可以通过分析基因与基因网络以及表达谱进行生物学功能的复杂研究。全部 Uniprotein 阵列的蛋白都可以高通量地得以表达与纯化。它们随后可以排布在一个 MS 靶上, 产生每一种蛋白的 MS 分布图及它在数据库中的储存信息。同样可以应用 2-DE 给出蛋白质组的分布图, 胶上的蛋白质点被切下来, 酶切, 质谱鉴定。它们的 DNA 序列可以按常规方法确定, 即使对大的蛋白复合物也可以 (Neubauer *et al.*, 1998)。

这种技术还被用来产生蛋白阵列的第二种类型, 包含利用新的固定到平床点样机械手上的转移标记, 对从液体表达培养物中纯化的蛋白微阵列进行自动化点样 (Lueking *et al.*, 1999)。这样的蛋白微阵列提供了一种用于高通量的、非常灵敏的基因表达与抗体特异筛选方法, 以及提供了抗体鉴定、血清分布 (Cahill, 2000)、高通量配体-受体相互作用研究以及此处提到的蛋白质组应用的有力工具。

在哺乳动物细胞中克隆 cDNA 及表达蛋白产物与胞内蛋白的 2-DE 模式相匹配已经

在前描述。cDNA 文库中的混合克隆在体外转录/转译表达、二维电泳分析,将在第 5 部分中描述。然而,我们相信联合蛋白阵列与质谱技术来产生 MPI 是一种连接基因组学与蛋白质组学的新方法。

4. 通过高分辨二维电泳 (2-DE) 鉴定特定细胞类型或组织在特定时间的蛋白组成

在某一特定组织或(和)发育的某一阶段中蛋白表达的功能分析是理解生物学过程的关键步骤。这些蛋白一旦产生,就能潜在地(potentially)在其承担细胞活动的功能之前,执行翻译后修饰。这些修饰可能是因为加了其他分子(如磷酸化、糖基化)、蛋白加工和蛋白分解切割。由于任何一个蛋白的结构修饰和任何量的改变不都是由相同的基因控制,各种蛋白的活性形式不能参照任何单一基因确定。因此,即使所有基因组中的基因已经测序并翻译成蛋白质,识别和鉴定在某种条件的特殊细胞类型中每种蛋白的不同功能形式仍是一个主要任务。蛋白结构或量的变化能导致疾病。为了阐明疾病机制,就必须理解不同蛋白如何表达及在特殊生物学背景下如何发挥功能。通过确定并比较出现在疾病组织和相应正常组织的蛋白,就可以鉴定与特殊疾病相关的特异蛋白。这些疾病特异蛋白具有作为疾病标记或作为药物靶分子的商业潜力。它们的相应基因序列可能在筛选易感性方面具有重要价值。研究和发掘药物的主要前沿就是研究和分析蛋白质组,因为使用这种方法,可能鉴定与特殊疾病相关的特异蛋白。

4.1 蛋白质组学与基因组学

蛋白质组学分析一种基因组或一个细胞或组织类型所有蛋白的表达(Wilkins *et al.*, 1996),最近有这方面的综述(Blackstock and Weir, 1999)。由于同一时间特定的细胞、组织或有机体中并非所有的蛋白都表达,因此蛋白质组是动态的。蛋白质组这个词来源于蛋白质组,蛋白质组与同类术语基因组相一致,由澳大利亚悉尼 Macquarie 大学的 Marc Wilkins 和 Keith Williams 于 1994 年提出。蛋白质组的名称由 Wasinger 等于 1995 年第一次在出版物中使用,他将之定义为“一种基因组的全部蛋白”。相类似的有关蛋白质组的定义如“由基因组编码的一整套蛋白质”可能在文献中更经常看到。然而,从严格的意义上讲,那些定义蛋白质组为基因组的蛋白编码物的所有定义都不确切,因为基因组不能明确地编码有机体中蛋白的整体结构和多样性。这一点同样存在于其他的有关蛋白质组的概念中,例如,有机体的整套蛋白质,同样没有给出有机体的 DNA 和蛋白分布之间简单的线性关系。

基因组在几乎所有的细胞中都是完整的,与之不同,蛋白质组具有很高的细胞特异性:一个有机体的不同细胞表达它的全蛋白的不同亚套。同时,基因组储存了稳定的信息,其成分不随时间变化,而蛋白质组是高度动态的,根据细胞的发育和生理状态蛋白表达的亚套一刻不停地变化。再强调一下,以此类推功能基因组学的概念也是很中肯的,因为功能基因组学研究 mRNA,它像蛋白质一样,是细胞类型和细胞状态特异的,并且是动态表达的。蛋白质组动态的和细胞类型特异的特点提出了系统分析的主要挑战,要完全记录多细胞有机体的蛋白质组,就必须记录每种细胞类型在它可能的生理状态和发育阶段的蛋白分布图。

4.2 二维聚丙烯酰胺凝胶电泳 (2-DE)

复杂蛋白质组的鉴定要求分离和检测成百上千的不同的蛋白种类。这个方法的目标是基于功能参数的广谱的许多蛋白的鉴定。包括这些蛋白要正确地出现在不同机体、不同胚层、出生后和成年阶段（包括老年阶段）、在不同的细胞部位和细胞的细胞器（核、线粒体）以及在不同的性别中（Bowden *et al.*, 1996; Dean *et al.*, 1998; Römer *et al.*, 1997）。通过分析共翻译和翻译后修饰（如糖基化蛋白、磷酸化蛋白），也需要单个蛋白生化特征的确定（Janke *et al.*, 1996）。蛋白质组分析是通过高分辨率的二维聚丙烯酰胺凝胶电泳（2-DE）、分析与比较得到斑点的模式、结合原位蛋白水解和质谱技术鉴定候选蛋白来进行（Gauss *et al.*, 1999; Jungblut *et al.*, 1998）。所需的高分辨率由大胶的 2-DE 技术产生，它能在一块胶上显示多于 10 000 个蛋白点（Klose and Kobalz, 1995; Klose, 1999）。这个技术基于最初的 2-DE 技术（Klose, 1975; O'Farrell, 1975），并在目前现有的文献报道 2-DE 技术中仍然显示了最高的分辨率。对于样品制备，一个详细的高度标准化的操作手册已经建立（Klose, 1999），这个操作方法避免了蛋白种类的特殊组的任何丢失。根据这个操作手册，某一特定组织的蛋白提取物分成三个部分（细胞质、质膜、染色质），这三个部分的 2-DE 模式呈现了这个组织（10 000~15 000 个不同的蛋白点）的绝大部分。为了检测低浓度蛋白要加上来自纯化的细胞器的 2-DE 模式并且可能显示一千或两千个蛋白点以补充以上三个基本组合。

来自 2-DE 模式的蛋白点从胶中切下，酶消化并且用质谱分析消化的肽。每一个点将根据惟一的 MPI（见图 1.3）被特征化；MPI 将被储存在数据库中。为了将几千个单一蛋白按照提到的标准特征化，蛋白的不同生物和生化类型依据 2-DE 点的出现与否而被匹配及消减，并且依据蛋白不同的表达水平，一个蛋白可以显示在不同的组织中。对 2-DE 模式的评估是通过激光扫描及软件辅助的斑点识别和特征化进行（见第六章）。可用高度灵敏的银染技术进行蛋白模式的出现/缺失分析。对于定量分析，可以应用考马斯亮蓝或新近发展的荧光染色技术（见第四章）。为检测翻译后修饰，可用糖基染色和磷酸染色。匹配可通过计算机用两种方法进行：①使用来自标准蛋白类型模式的 2-DE 图像（见 4.3 部分）；②使用 MPI 数据库，它可不依赖于胶模式而进行蛋白匹配。在每个 2-DE 模式中，蛋白点的 MPI 数据也使检测容易，它代表了同一蛋白的修饰情况。每个 2-DE 模式中包括所有未知蛋白在内的尽可能多的蛋白点，可能通过使用质谱 MPI、2-DE 的计算机模式得以特征化，来自同一模式的 MPI 数据是独立匹配和消减的。通过这种方法，几千种蛋白根据生物和生化参数得以鉴定。

必须结合这些数据和基因表达谱（第 2 部分）来充分说明在某一特殊细胞或组织在正常、疾病和激活（如药物处理）条件下发生的事件。这些信息可通过比较测量的 MPI 和那些相应的来自蛋白表达库的数据，与来自同一组织的 mRNA 表达谱（第 2 部分）相联系。在 MPI 数据的基础上我们可以联系起 2-DE 模式中的各自的蛋白种类与蛋白表达阵列的同源蛋白，并且这种方法可联系起相应的 cDNA。结果，基因开始与它们的蛋白相联系，那是功能特征化的。然而，已经知道基因组学、功能基因组学和蛋白质组学的结果之间的相关性较低（Anderson and Seihamer, 1997）。我们认为此处描述的方法使用 MPI 可以为蛋白质组学和基因组学架起桥梁，也对解释和分析那些巨大的数据

库和高通量方法有用。

然而还有一个基本的问题，即检测作用于蛋白调控和修饰的基因。基因组学针对此问题的方法是基于出现在蛋白表型的结构和量的多态性，开展遗传学连锁研究 (Breen *et al.*, 1994; Klose, 1999; Klose *et al.*, 稿件正在准备; Nock *et al.*, 1999)。研究蛋白之间相互作用的方法如酵母双杂交系统 (Wanker *et al.*, 1997) 也可发现除了编码蛋白的结构基因外，相关的其他基因。

4.3 联合的 2-DE 数据库

在最近几年，联合 2-DE 数据库已经建立，使得世界范围的实验室共享 2-DE 数据。然而实际上，在不同的实验室匹配 2-DE 模式证明很困难甚至不可能，因为使用的技术不同 [载体两性电解质、固相 pH 梯度 (IPG)、胶形式、染色过程和样品制备]。然而将来这个问题将由于更广泛的使用不同的质谱技术得到解决，使得各实验室在 MS 数据水平而不是 2-DE 图式的匹配比较结果。尤其是 MPI 数据库将成为转换来自 2-DE 模式数据的平台 (见图 1.1)。结果，标准化世界范围的实验室使用的 2-DE 技术将不再是共享数据的要求。

4.4 2-DE 分析自动化

在不久的将来每个实验室每天将可能分析成百的蛋白点，或者为了分析整块的胶，要将完整的 2-DE 胶分隔成小片。这个发展迫切要求 2-DE 自动化。Pharmacia 在某种范围上简化了固定化电解质技术，但它离自动化还很远。而且，我们发现固定化电解质胶不能达到解决复杂蛋白质组所必须的高分辨率水平。

另一个感兴趣的焦点是在蛋白质组分析中改进高通量技术，例如发展从 2-DE 胶识别和切取蛋白点的自动化机械手、酶消化和将消化好的 2-DE 胶的点转移至用于 MS 分析和产生 MPI 的质谱靶 (见第八章和第九章)。

5. 用质谱连接当前蛋白质组学和基因组学的方法

我们的策略旨在连接基因组学和蛋白质组学。可以总结如下：产生一套 Unigene，每个代表惟一的基因。作为重组蛋白表达，它的 MPI 可用质谱决定。另一方面，相应于 Unigene 套的来自于 2-DE 模式的蛋白被 MPI 特征化。这提供了由 2-DE 特征化蛋白与其相应 mRNA 和基因 (cDNA) 联系的桥梁。所有 2-DE 胶收集的 MPI 将在计算机上和获得于重组蛋白库的 MPI 比较，反之也可。因此，成百上千的具生物活性的基因产物与它们的基因相联系。这种联系不依赖于任何序列信息并因此也吸引了其他有机体的功能蛋白质组分析。我们的策略与现存策略相比另一个优点是蛋白识别更可靠，因为是质谱数据与质谱数据相比，而不是与从 DNA 或蛋白序列预测的数据相比。后一方法的主要缺陷是没有考虑依赖于底物的蛋白酶行为、肽的溶解性和质谱分析信号抑制。

正如已经描述的，MPI 包括酶切产物的精确分子质量和离子碎片数据。将被用于从预测的蛋白或实验记录的 MPI 数据库鉴定蛋白。后者由两个主要目录组成：①由 2-DE 分离的体内表达蛋白；②重组 cDNA、基因组 DNA 表达产物。

为达到广泛应用，将产生一个 Unigene 套，最初是人类基因，每个成员代表惟一的

基因（见第 2 部分），这些基因可表达为重组蛋白（见第 3 部分），它们的 MPI 将被确定并存储于中心数据库。我们预测这将建立起一个新的桥梁，即将天然蛋白群体（蛋白质组）与 mRNA 池（转录组）和它们隐伏的基因（基因组）之间直接联系。

6. 前景和发展

6.1 基因组学

随着生物技术微型化进程的发展，必须发展新的硬件工具。所有微型杂交方法必须的处理步骤，尤其是现存的检测系统，也要求加以改进。越小的斑点尺寸导致所需要的靶、探针和样品的数量的减少，同样也隐含着：随着空间分辨率的增加，需要发展更灵敏的检测系统。然而在过去，放射性标记的方法是占主导地位的生物技术，检测方法的下一代将基于光学的原理，因为用³²P、³³P或³⁵S标记的化合物在放射自显影时显出弥散的信号，并且发射的射线释放（delivered）大于 360°。由于这个限制，光学方法例如针对大面积（22.2cm×22.2cm）激光扫描设备或其他的显微方法，包括对于面积小至几个 mm² 的共聚焦激光扫描显微镜，都是不久的将来可以选择的方法。

6.2 蛋白组学

为了获得通量化，每天必须记录几千个 MPI。为此，急需发展一个裂解平行出现在凝胶基质（gel matrix）或结合至颗粒（重组融合蛋白亲和纯化）的多种蛋白的完全自动化工作站。这个工作站应包括样品清洗、浓缩和随后的质谱样品准备。而且，质谱设备要求每天能产生几千个 MPI。就通量化和全部耗费而言，MALDI-TOF-MS 是目前可选择的质谱技术。另外，新的质谱方法将需要探索高的产品输出，包括改进的 PSD 分析工具及用于离子捕获的 MALDI。然而有一些滞后可能限制 MALDI-MS 在蛋白质组和基因组计划中的应用。最引人注目的是盐和去污剂。例如来自 SDS-PAGE 电泳的 SDS 或染色试剂，大大增加了在测试过程中的背景。因此这就要求进行质谱之前样品的清洗（参阅第九章）。

6.3 尚需发展之处

这种方法的成功需要进一步发展一些技术，包括：

- (1) 能够大规模亚克隆 cDNA 的新方法，同时可修饰带标签的表达载体，例如开发和改进用于提高在酿酒酵母（*S. cerevisiae*）和裂殖酵母（*S. pombe*）中表达和分泌的载体系统，以及用于解决在 *E. coli* 中不容易表达的蛋白表达的载体系统（Lueking *et al.*, 2000）。
- (2) 一个全自动的用于纯化 His-标记的蛋白的工作站（如 BioRobot 8000, Qiagen）。
- (3) 2-DE 过程及 2-DE 的学术与工业研究中更自动化的发展，作为蛋白组学的中心技术要求简化的大规模 2-DE 与技术改进、集中解决三个问题的自动化仪器，在长毛细管和大的胶装置中灌制薄胶。

其他的关键问题包括：灌制胶、清洗试管和玻璃板、应用于等电聚焦（IEF）电泳的样品制备、转 IEF 胶到 SDS 胶。包括发展从 2-DE 胶取点的机械手，它能从 2-DE 胶

上自动切取成百上千的蛋白点。在研究分子遗传学的 Max-Planck 研究所, 已经研制了一个可进行 8 个管穿孔的胶取点设备, 但需要进一步改进 (如释放胶碎片) 及适用于 20cm×30cm 胶的机械手系统。

7. 总结

来自 2-DE 的 MPI 与重组蛋白库的 MPI 的记录的对比直接连系了 Unigene-Uniprotein 库与所有观察到的基因产物, 反之亦然。而且, 蛋白鉴定更为可靠, 因为来自 2-DE 电泳的测试与来自表达蛋白 MPI 的测试对比, 代替了 DNA 序列预测 MPI (酶切产物结合离子片段的的结果)。优势在于这种联系独立于任何序列信息, 因此也吸引了其他有机体的功能蛋白组分析, 尤其是那些还未测序的生物。一旦被记录, MPI 可以用于迅速地识别已知的和未知的基因产物。同时利用 MPI 可以在序列数据库中鉴定蛋白。有了这些特点, MPI 能够保存下不同生物样品 2-DE 电泳的比较结果, 而不依赖于它们的形式、分辨率和应用的分离技术。来自 2-DE 的 MPI 与重组蛋白库记录的 MPI 之间的对比直接连系了 Unigene-Uniprotein 库与所有观察到的基因产物, 反之亦然。因此将基因组学和蛋白质组学连接起来。此外, 直接进入相应的重组蛋白表达文库打开了一条从 2-DE 凝胶进行蛋白质谱鉴定的完美的途径。稳定同位素标记 (如 *E. coli* 细胞在 ^{15}N 培养基中生长) 是这种方法的关键, 第八章将有描述。

(张令强译 贺福初校)

参考文献

- Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**: 533–537.
- Aparicio, S.A.J.R. (2000) How to count human genes. *Nature Genet.* **25**: 129–130.
- Bancroft, D. R., O'Brien, J. K., Guerasimova, A. and Lehrach, H. (1997) Simplified handling of high-density genetic filters using rigid plastic laminates. *Nucleic Acids Res.* **25**: 4160–4161.
- Behr, S., Matzig, M., Levin, A., Eickhoff, H. and Heller, C. (1999) A fully automated multicapillary electrophoresis device for DNA analysis. *Electrophoresis* **20**: 1492–507.
- Blackstock, W.P. and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17**: 121–127.
- Bowden, L., Klose, J. and Reik, W. (1996) Search for parent-specific gene expression in early embryos and embryonic stem cells in the mouse using high-resolution two-dimensional electrophoresis of proteins. *Int. J. Dev. Biol.* **40**: 499–506.
- Breen, M., Deakin, L., MacDonald, B. et al. (1994) Towards high resolution maps of the mouse and human genomes – a facility for ordering markers to 0.1 cM resolution. *Human Mol. Genet.* **3**: 621–628.
- Büssow, K., Cahill, D., Nietfeld, W., Bancroft, D., Scherzinger, E., Lehrach, H. and Walter, G. (1998) A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids. Res.* **26**: 5007–5008.
- Büssow, K., Nordhoff, E., Lübbert, C., Lehrach, H. and Walter, G. (2000) Genomics A human cDNA library for high-throughput protein expression screening.
- Cahill, D.J. (2000) In Blackstock, M.M. (Ed.). *Proteomics: A Trend Guide*, pp. 49–53. Elsevier Science, Oxford.
- Cho, R.J., Campbell, M.J., Winzler, E.A. et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell.* **2**: 65–73.

- Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I. (1998) The transcriptional program of sporulation in budding yeast. *Science* **282**: 699–705.
- Clark, M. D., Panopoulou, G. D., Cahill, D. J., Bussow, K. and Lehrach, H. (1999) Construction and analysis of arrayed cDNA libraries. *Methods Enzymol.* **303**: 205–233.
- Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J.J., Reik, W. and Feil, R. (1998) Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development* **125**: 2273–2282.
- DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**: 457–460.
- 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.
- Eickhoff, H. (1998) *Drug Discovery Today*, Vol. 3, pp. 148–149.
- Eickhoff, H., Ivanov, I. and Lehrach, H. (2000) In Saluz, H.P. (Ed.). *Technical System Management in Microsystem Technology: a powerful tool for biomolecular studies*. Birkhäuser Verlag.
- Gauss, C., Kalkum, M., Lowe, M., Lehrach, H. and Klose, J. (1999) Analysis of the mouse proteome. (I) – Brain proteins: Separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation. *Electrophoresis* **20**: 575–600.
- Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis, R.W. (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* **94**: 2150–2155.
- Herwig, R., Poustka, A., Müller, C., Bull, C., Lehrach, H. and O'Brien, J. (1999) Large-scale clustering of cDNA-fingerprinting data. *Genome Res.* **9**: 1093–1105.
- Humphery-Smith, I. (1999) Replication-induced protein synthesis and its importance to proteomics. *Electrophoresis* **20**: 653–659.
- Janke, C., Holzer, M., Klose, J. and Arendt, T. (1996) Distribution of isoforms of the microtubule-associated protein tau in gray and white matter areas of human brain – a two-dimensional gel-electrophoretic analysis. *FEBS Lett.* **379**: 222–226.
- Jordan, B.R. (1998) Large-scale expression measurement by hybridization methods: From high-density membranes to 'DNA chips'. *J. Biochem.* **124**: 251–258.
- Jungblut, P.R., Otto, A., Favor, J., Lowe, M., Müller, E.C., Kastner, M., Sperling, K. and Klose, J. (1998) Identification of mouse crystallins in 2D protein patterns by sequencing and mass spectrometry. Application to cataract mutants. *FEBS Lett.* **435**: 131–137.
- Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. *Humangenetik* **26**: 231–243.
- Klose, J. (1999a) Genotypes and phenotypes. *Electrophoresis* **20**: 643–652.
- Klose, J. (1999b) Fractionated extraction of total tissue proteins from mouse and human for 2-D electrophoresis. *Methods Mol. Biol.* **112**: 67–85.
- Klose, J. (1999c) Large-gel 2-D electrophoresis. *Methods Mol. Biol.* **112**: 147–172.
- Klose, J. and Kobalz, U. (1995) Two-dimensional electrophoresis of proteins: An updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**: 1034–1059.
- Klose, J., Nock, C., Büssow, K., Löwe, M., Schalkwyk, L.C., Rastan, S., Brown, S., Himmelbauer, H. and Lehrach, H. (2000) Gene mapping by protein mapping. Connecting proteome and genome of the mouse. In preparation.
- Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Monaco, A.P., Nizetic, D., Zehetner, G. and Poustka, A. (1990) Hybridisation fingerprinting in genome mapping and sequencing. *Genome Anal. Genet. Phys. Map.* **1**: 39–81.
- Lehrach, H., Bancroft, D. and Maier, E. (1997) Robotics, computing, and biology. An interdisciplinary approach to the analysis of complex genomes. *Interdisciplinary Science Reviews* **22**(1): 37–44.
- Lennon, G.G. and Lehrach, H. (1991) Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* **7**: 314–317.
- Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M.B. (1996) The IMAGE consortium: An Integrated Molecular Analysis of Genomes and their Expression. *Genomics* **33**: 151–152.
- Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H. and Walter, G. (1999) Protein microarrays for gene expression and antibody screening. *Anal. Biochem.* **270**: 103–111.

- Meier-Ewert, S., Maier, E., Ahmadi, A., Curtis, J. and Lehrach, H. (1993) An automated approach to generating expressed sequence catalogues *Nature* **361**: 375–376.
- Meier-Ewert, S., Lange, J., Gerst, H. *et al.* (1998) Comparative gene expression profiling by oligonucleotide fingerprinting. *Nucleic Acids Res.* **26**: 2216–2223.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998) Mass spectrometry and EST database searching allows characterisation of the multi-protein spliceosome complex. *Nature Genet.* **20**: 46–50.
- Nock, C., Gauss, C., Schalkwyk, L.C., Klose, J., Lehrach, H. and Himmelbauer, H. (1999) Technology development at the interface of proteome research and genomics: Mapping nonpolymorphic proteins on the physical map of mouse chromosomes. *Electrophoresis* **20**: 1027–1032.
- O'Farrell, P.H. (1975) High resolution two-dimensional gel electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007–4021.
- Poustka, A. J., Herwig, R., Krause, A., Hennig, S., Meier-Ewert, S. and Lehrach, H. (1999) Toward the gene catalogue of sea urchin development: The construction and analysis of an unfertilized egg cDNA library highly normalized by oligonucleotide fingerprinting. *Genomics* **59**: 122–133.
- Radelof, U., Hennig, S., Seranski, P., Steinfath, M., Ramser, J., Reinhardt, R., Poustka, A., Francis, F. and Lehrach, H. (1998) Preselection of shotgun clones by oligonucleotide fingerprinting: an efficient and high throughput strategy to reduce redundancy in large-scale sequencing projects. *Nucleic Acids Res.* **26**: 5358–5364.
- Ramsay, G. (1998) DNA chips: state-of-the-art. *Nature Biotechnol.* **16**: 40–44.
- Römer, I., Reik, W., Dean, W. and Klose, J. (1997) Epigenetic inheritance in the mouse. *Curr. Biol.* **7**: 277–280.
- Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. and Davis, R.W. (1996) Parallel humq genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci.* **93**: 10614–10619.
- Schena, M., Heller, R. A., Theriault, T. P., Konrad, K., Lachenmeier, E. and Davis, R. W. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol.* **16**: 301–306.
- Shalon, D., Smith, S.J. and Brown, P.O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* **6**: 639–645.
- Trent, J.M., Bittner, M., Zhang, J. *et al.* (1997) Use of microgenomic technology for analysis of alterations in DNA copy number and gene expression in malignant melanoma. *Clin. Exp. Immunol.* **1**: 33–40.
- Uetz, P., Giot, L., Cagney, G. *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**: 623–627.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**: 484–487.
- Walter, G., Büsow, K., Cahill, D., Lueking, A. and Lehrach, H. (2000) Protein arrays for gene expression and molecular interaction screening. *Curr. Opin. Microbiol.* **3**: 298–302.
- Wanker, E.E., Rovira, C., Scherzinger, E., Hasenbank, R., Walter, S., Tait, D., Colicelli, J. and Lehrach, H. (1997) Hip-I – a Huntingtin interacting protein isolated by the yeast two-hybrid system. *Human Mol. Genet.* **6**: 487–495.
- Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, K.L. and Humphery-Smith, I. (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**: 1090–1094.
- Welford, S.M., Gregg, J., Chen, E., Garrison, D., Sorensen, P.H., Denny, C.T. and Nelson, S.F. (1998) Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res.* **26**: 3059–3065.
- 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. Engin. Rev.* **13**: 19–50.

Wolfsberg, T.G. and Landsman, D. (1997) A comparison of expressed sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Res.* **25**: 1626–1632.

Zehetner, G. and Lehrach, H. (1994) The Reference Library System – sharing biological material and experimental data. *Nature* **367**: 489–491.

第二章 基因表达的检测方法

S. J. Blakemore, D. M. Wallace and M. K. Trower

1. 绪论

信使 RNA (messenger RNA, mRNA) 一词是 20 世纪 60 年代初期产生的, 用来描述一类特定的细胞 RNA, 它们作为运送者, 将遗传信息从细胞核送至细胞质的蛋白质合成器 (核糖体) 内, 在此被翻译成蛋白质。长期以来, 一直认为 mRNA 在决定细胞产生的蛋白质的种类和数量过程中起关键作用 (见图 2.1)。事实上, 不同种类细胞中蛋

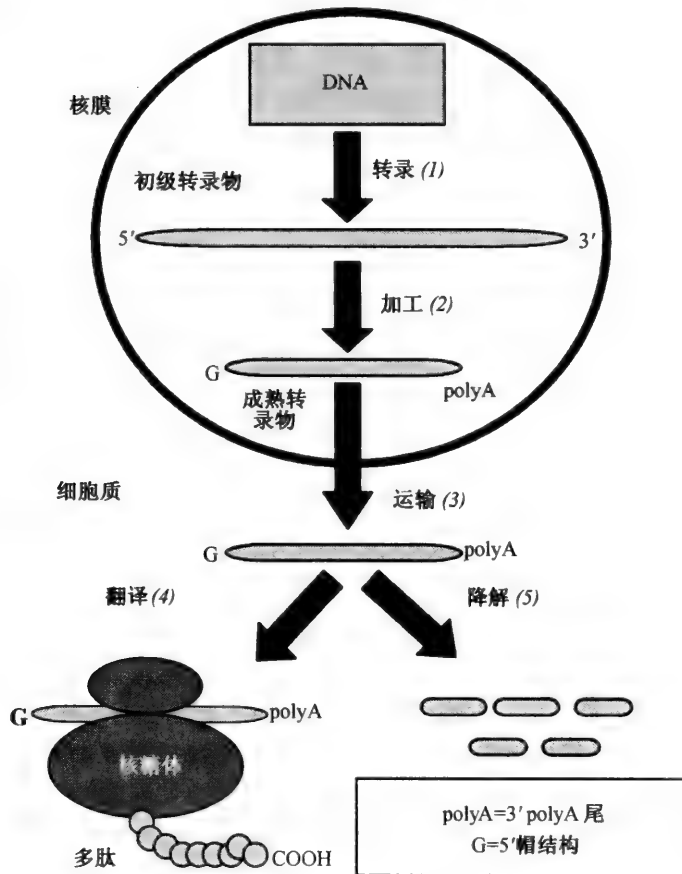


图 2.1 遗传信息从细胞核内的 DNA 传送到蛋白质合成的过程。细胞质中可以被翻译的 mRNA 决定了那些可以在任何指定时间由细胞及时“生产”出来的蛋白质的状态。此 mRNA 池的特性是由特定的细胞生命过程的效率决定的 (1~5)。

白质的差别反映了细胞生长发育过程中表达的 mRNA 的种类和数量（丰度）的差别。核酸杂交动力学研究提示了这一点，研究发现 cDNA（与 mRNA 互补的 DNA 拷贝）与同一组织来源的 cDNA 的复性速度远快于与不同组织来源的 cDNA 的复性速度（Hastie and Bishop, 1976）。一旦认识到不同组织/细胞类型的 mRNA 群体差异，那么确定这种差异到底是什么就变得重要了。因此很明显，我们需要能对生物样品中特定种类的 mRNA 进行精确定量的方法。以杂交为基础的实验方法因为核酸的碱基互补配对单一选择性（如 Northern 杂交）而简单易行。应用这种一个一个基因的分析方法发现，一些转录物在特定的组织中高丰度表达而在其他组织中不表达，某些基因在许多、甚至所有的组织中的表达量都相对恒定。作为 Northern 杂交的补充，后来发展了其他一些 mRNA 定量方法，比如核糖核酸酶保护法、定量反转录酶 PCR（RT-PCR）。然而，每种方法都局限在只能针对相对少数的基因。最近 5 年中技术的进步产生了一些能在基因组水平（大量）定量分析 mRNA 的方法，与过去一次分析一个基因相比，它们能同时分析成百上千个基因。发展这些技术的初衷是希望能洞察生物学变化的根本原因——基因组水平的整体行为（一个典型的哺乳动物细胞大约表达 15 000 个不同的 mRNA）。各种高通量 mRNA 定量技术的发展得益于 PCR、大规模 cDNA 文库测序和核酸从头合成等技术方法的进步。差异显示 PCR（DD-PCR）、基因表达系列分析（serial analysis of gene expression, SAGE）和 DNA 阵列杂交等方法在检测灵敏度和凭借 RNA 量来检测基因数等方面比 Northern 杂交具有明显的优势。本章将讨论这些方法的细节、它们之间相对的优缺点。

需要指出的是，目前所有 mRNA 定量技术都只提供了实验性的、而非稳定状态下 mRNA 绝对量的信息；不过这对于检测 mRNA 表达水平的变化来说，已经足够了。

在关于蛋白质组学的论著中需要指出的是，检测两个生物样品中特定 mRNA 表达丰度的差异并不一定决定于对应的蛋白质的量的差异。早在 20 世纪 60 年代，人们已经认识到原核生物中几乎所有的基因表达调控都发生在转录水平。尽管对于真核生物而言，通常情况下伴随着 mRNA 的增加（或减少），其编码的蛋白质也相应地增加（或减少），但是蛋白质改变的程度和动力学常常是不可预见的。因此当研究对象是真核生物时就需要精心明辨，不能仅仅以 mRNA 的数据为基础就得出蛋白质水平的结论。蛋白质水平的变化需要通过酶联免疫吸附、Western 杂交或蛋白质组学研究来确认。

2. DNA 阵列杂交

判定 mRNA 水平最常用的方法是将标记的核酸（代表一种 mRNA 样品）与点在固相支持物上的 cDNA 或寡核苷酸序列进行杂交。杂交后每个点的标记强度就代表了特定序列的表达水平。用不同来源的 mRNA 所制备的探针与相同的阵列进行杂交后可以比较对应点的杂交信号强度。这就是检测某个点的 DNA 序列（基因）特异性的 mRNA 的表达差异的方法。通过比较两个阵列所有对应点的杂交信号强度，就可以同时监测数千个基因表达的改变。用两种可识别的不同的标记来标记需要研究的两组 mRNA，就能在同一个阵列上同时进行杂交（见本书彩插图 2.2）；比较每个点两种信号的强度便是差异杂交法。阵列上序列的覆盖范围限制了这些方法的使用，这样如果一个特定的基因没有对应序列来代表，那么显然它就不可能被检测到。所以制作阵列时选择何种 DNA 对

于任何研究工作都是至关重要的。这种基于阵列的 mRNA 定量方法称作“封闭体系”。

使用 DNA 阵列检测 mRNA 丰度非常类似于同时进行多个逆向 Northern 杂交。Northern 杂交将 RNA 固定于支持物上, 然后与基因特异性的被标记的 DNA 进行杂交, 而 DNA 阵列法是将基因特异性的 DNA 固定在支持物上, 然后与来源于 RNA 的被标记探针杂交。阵列技术的优势来源于这样一个事实, 即数千个基因(点状印记)能同时检测。具有代表性的是, 这数千个基因是以高密度自动化地印记到尼龙膜或玻璃片上的。目前, 玻璃片上能实现的点密度更高, 从而在不牺牲检测灵敏度的前提下, 增加检测基因的数量。DNA 阵列按照用以构建阵列的 DNA 的化学本质可分为变性的 PCR 产物(源自 cDNA 文库)和寡核苷酸两类。所以构建阵列需要 cDNA 克隆或序列信息(用以设计有代表性的寡核苷酸)。因此阵列的设计受到我们现有的关于构成某个物种基因组的基因知识的限制(杂交按照物种特异的方法进行)。对于那些整个基因组都已完成序列测定的物种, 比如酵母、大肠杆菌、结核杆菌(完成序列测定的基因组列表见基因与基因组京都百科全书, <http://www.genome.ad.jp/kegg/kegg2.html>), 基于阵列的技术能涵盖整个基因组, 是有效的“开放系统”。应用 DNA 阵列技术对酵母(Wodicka *et al.*, 1997)、大肠杆菌(Richmond *et al.*, 1999)、结核杆菌(Behr *et al.*, 1999)进行的基因组范围的表达谱实验结果已经公开发表。然而对于人、小鼠和大鼠等物种, 要等它们各自的基因组序列完成后才能开展基因组范围的阵列研究。

当你计划开始基于 DNA 阵列的实验时, 另外一个主要问题是, 怎样将 RNA 转换成标记探针并进行杂交。以 mRNA 或总 RNA 为起始模板标记探针的方法有很多种, 这其中包括 cDNA 第一链合成(以锚定的 oligo-dT 或随机十寡核苷酸为引物), 或者结合进一步的用 T7RNA 聚合酶或 PCR 进行核酸扩增。用放射性同位素(如³²P或³³P)或荧光素(如 Cy3 和 Cy5)标记的核苷酸来标记探针, 前者通常用于尼龙膜; 后者用于玻璃片阵列。每种方法有各自最佳的 RNA 用量, 比如 cDNA 第一链合成约需 2~5 μg PolyA RNA, 而 PCR 扩增法最少用 25 ng 的总 RNA 就足够了(Gonzalez *et al.*, 1999)。既然关于目前的条件或参数的系统比较未见报道, 就没有一种“标准方法”。PCR 扩增法值得关注, 包括对高丰度的转录物进行非线性扩增的可能性, 可以达到数量上的亚最佳标记的序列特异性扩增。然而最近的证据显示或许这些关注是没有理由的(Endege *et al.*, 1999)。不考虑技巧问题, 不同阵列的灵敏度看来差不多——使用 PolyA RNA 为模板, 能辨别 10⁵ mRNA 中的 1 个(约每个细胞 2~5 个 mRNA 拷贝)。

当使用一种标记来鉴别不同样品中差异表达的基因 mRNA 时, 操作过程相对简单: 标记后的 RNA(来源于两个或多个样品)用于相同的 DNA 阵列杂交, 然后比较影像之间对应点的强度(考虑到比活的差异先归一化)。另一个常用的技术是使用两种可辨别的不同标记(如分别用 Cy3 和 Cy5 标记两种 mRNA 样品), 在同一个玻璃片阵列上用不同标记制备的探针进行杂交。两种有各自特征性发射波长的荧光标记能彼此区分开, 使在同一个阵列上检测差异杂交成为可能。每个点的荧光比例给出了序列特异的竞争性杂交结果。与以前的方法相比, 单一阵列技术主要的优势在于可以减少由于不同阵列间细微的差别产生的影响。然而, 使用单一阵列双重杂交的缺点是得到的比率结果只在特定的数据中有效。比如, 对于不同时间点的取样用 Cy5 标记, 分别和以 Cy3 标记的零时间点的参照 mRNA 一起与不同的阵列进行杂交。由于不同的研究使用不同的参照样本,

把不同研究的结果整合起来的能力就打折扣了。

用单一或双重标记方法获得的强度和（或）比率值被储存在数据库中，进行计算机分析和不同数据组间的比较。使用特定的以基因阵列为基础的技术其主要优势在于对于数据进行多元统计分析时相对简单（例如时间效应、剂量效应和其他多样本数据）。例如在时间效应的情况下，将表达动力学近似的基因进行聚类的方法能使我们观察到基因表达的协同调控（Iyer *et al.*, 1999）。统计分析应用于多基因表达谱数据的另一个例子是用于重新划分淋巴瘤的类型和亚类（Golub *et al.*, 1999）。对基因表达数据进行多元分析可能发现基因表达的相关性，而这仅仅通过比较二元的杂交数据组是无法实现的（例如疾病组织与对应的正常组织，处理和未处理的细胞）。迄今为止，已经公开发表的多元基因表达分析的例子主要是基于阵列分析数据（Alizadeh *et al.*, 1999; Fambrough *et al.*, 1999; Golub *et al.*, 1999; Heller *et al.*, 1997; Iyer *et al.*, 1999; Perou *et al.*, 1999），惟一的例外是应用 RT-PCR 提取的数据。事实上，与其他技术平台（如 SAGE、数字基因表达谱分析和差异显示 PCR）相比，基于阵列的方法其优势是能相对容易地对数据进行多元统计分析。

随着数据库中基因表达数据大量累加，人们希望对于基因调控的发现能够更加准确和灵敏。也就是说这不仅提供了使人们能更好地理解基因表达调控的方法，还提供了基因的生物学功能网络的线索。相似疾病（如炎症）的基因表达谱的确定可以对它们的分子病理学更精确地理解。然而，必须先确定一些标准化的形式以便进行不同组间结果的交叉比较。对于所有不同技术平台产生“标准化表达数据组”的最佳方法仍在争论中。最近，欧洲生物信息学研究所（<http://www.ebi.ac.uk/arrayexpress/>）和美国国家生物工程信息中心（NCBI; Marshall, 1999）都表示计划提供一个公共数据库对基于阵列的杂交数据进行储存和分析。

2.1 cDNA 阵列

cDNA 阵列由源自 cDNA 文库的 PCR 产物组成，由机器人点到固体支持物的 cDNA 阵列可以用不明确的 cDNA 文库（cDNA 序列未知）或明确的 cDNA 文库（cDNA 序列已知）来建立。使用明确的 cDNA 文库有两个主要优势：第一，能在阵列上检测是否存在对于某特定研究来说“感兴趣的基因”（如阳/阴性对照）；第二，快速将杂交数据映射到基因注释加速了结果分析。使用另一种不明确的 cDNA 来源需要对那些表现出差异杂交的克隆进行测序来确定基因的差异表达。此过程费时且效率低，因为同一个基因可能被重复检测（尤其是用“标准”的冗余 cDNA 文库作为阵列的原始材料）。当设计 cDNA 阵列时通常使用 3' cDNA 序列来代表基因。它们可能源自非蛋白编码区，在基因家族的成员之间，非蛋白编码区通常比编码区差别更大。选择这样的 3' cDNA 减少了基因家族相关成员交叉杂交的可能性，尤其在序列相似性很强的情况下。

建立一个 cDNA 阵列生产、杂交和基因表达数据分析的便捷流程需要相当的生物信息学支持。精确和可靠的计算机系统是整个过程中各个阶段的基础。这包括选择用于阵列的 cDNA，样品跟踪以确定 cDNA 位于正确的阵列对应点上，测量所有点杂交信号强度，把杂交结果保存到数据库中（把杂交信号与基因注释相连），以及后续的数据分析。由 NCBI 开发并正在用于 cDNA 阵列数据分析的 ArrayDB（Ermolaeva *et al.*, 1998），就

是此类系统的一个例子。发展这样一个 cDNA 阵列工具目前很少实验室涉及, 因为机器人技术和生物信息学基础建设耗资巨大。因此, 尽管 cDNA 阵列技术原创的发展主要在依托于大学的 Hans Lehrach 实验室 (Gress *et al.*, 1996; Maier-Ewett *et al.*, 1993; Zehetner and Lehrach, 1994) 和 Pat Brown 实验室 (Schena *et al.*, 1995, 1996; Shalon *et al.*, 1996), 目前大多数阵列设备分布于生物技术公司、大型制药集团和一些专门的研究中心。不幸的是, 如果不考虑 cDNA 阵列应用于检测差异表达基因的潜力, 这意味着大多数研究人员目前还不能方便地用到这一技术。不过, 这种情形会改变, 伴随着更多的实验室应用这一技术, 成本可能下降。事实上, Pat Brown 实验室的网站 (<http://cmgm.stanford.edu/pbrown/>) 已经提供了一种阵列仪的产品信息, 价格比 DNA 自动测序仪低得多 (US\$ 23 000)。对于研究人员来说, 拥有一套便捷的阵列工具的基本的优势是能根据需要构建阵列。另外, 通过从商业公司购买阵列分析软件, 例如 Research Genetics Inc. (<http://www.researchgenetics.com/>)、Genome Systems Inc. (<http://www.genomesystems.com/>), Clontech (<http://www.clontech.com/>) 或 NEN (<http://www.nen-lifesci.com/>), 或者通过与商业的 cDNA 阵列服务提供商合作 [如 Incyte Pharmaceuticals Inc. (<http://gem.incyte.com/gem/index.shtml>) 或 Research Genetics Inc.], 以 cDNA 阵列为基础的研究在不用投资于所有技术的情况下仍可进行。

该领域论文数量的日益增长证明了 cDNA 阵列在提供新的生物学信息方面的效用, 例如肿瘤基因表达谱的研究 (DeRisi *et al.*, 1996, Golub *et al.*, 1999; Moch *et al.*, 1999; Perou *et al.*, 1999), 一个关键的胰岛素抗性基因的发现 (Aitman *et al.*, 1999), 生理周期中神经系统基因表达变化的检测 (DeRisi *et al.*, 1996, Golub *et al.*, 1999; Moch *et al.*, 1999; Perou *et al.*, 1999), 哮喘中淋巴细胞基因表达改变的检测 (Syed *et al.*, 1999)。作为这一技术能提供新的生物学洞察力的一个证据是, 利用一个含有 8613 基因的阵列研究静止培养的人成纤维细胞对于血清的反应 (一种成熟的细胞周期控制模型) (Iyer *et al.*, 1999)。许多被发现早期产生表达变化的是那些对于血清产生增殖反应的相关基因 (如转录因子 *fos/jun*)。然而, 在较晚的时间点, 典型的伤口愈合反应的基因表达变化出人意料地出现了, 例如诱导产生血管内皮生长因子 (启动血管生成) 和成纤维细胞生长因子 7 (刺激上皮再形成)。作者推断这些细胞暴露于血清中的主要效应是伤口愈合反应, 建议以此作为此类研究的体外模型。因此, 通过多基因杂交, 这一研究在成纤维细胞对于血清的生物学反应方面提供了以往未知的重要新发现。

概括起来讲, cDNA 阵列能用来构建任何物种的转录图谱 (只要有合适的 cDNA 文库), 同时监控成千上万个基因序列, 简化完整的数据分析, 建立结果数据库。然而 cDNA 阵列仅能对那些全基因组序列已知的物种提供基因组范围的检测。阵列产品和建立生物信息学体系的成本令大多数实验室望而却步, 进而导致这一技术使用的限制。

2.2 寡核苷酸阵列

在作用和一般性能方面, 寡核苷酸阵列与 cDNA 阵列相似 (见 2.1 部分)。两者都是由以高密度点在固体矩阵 (如玻璃片) 上的 DNA 组成, 用于基于杂交的基因表达谱的分析。与排列来源于 cDNA 克隆的 PCR 产物不同的是, 寡核苷酸阵列由合成的基因特异性的寡核苷酸组成。本章节重点分析这两种形式的 cDNA 阵列之间的优势和弱点。

寡核苷酸阵列的基本要素是短的寡核苷酸（通常 20~25 个核苷酸）含有足够的序列复杂性来实现选择性地杂交单一转录物。在实际操作过程中，代表一个基因不同区域的多个寡核苷酸被点到阵列上。显然，寡核苷酸阵列的建立需要事先了解表达的序列情况，这就把它的应用范围限制在基因组表达序列测定完成的物种。cDNA 阵列则不同，一个完全未知的 cDNA 文库可以用来作表达谱分析。不过寡核苷酸阵列的使用意味着无须小心翼翼地保存 cDNA 克隆和 PCR 产物，从而简化了精确制作阵列的后勤工作。事实上，直接在芯片上合成寡核苷酸能把阵列误差的风险降到最低点。有一些辅助阵列设计的方法能为一种转录物提供惟一的序列特异性 (Wodicka *et al.*, 1997)。这些程序能在已有的基因组序列信息基础上进行序列的最佳选择，从而减少由于交叉杂交造成的假象。交叉杂交的影响在 cDNA 阵列并不容易去除。

一旦选定了寡核苷酸来代表需要研究的基因，有两种基本的方法制作阵列：与 cDNA 阵列相似由机器人将预先合成的寡核苷酸点成阵列；或直接将 DNA 合成在阵列表面，这一技术有一家美国的生物技术公司 Affymetrix (<http://www.affymetrix.com/>) (Lockhart *et al.*, 1996) 开发，Affymetrix 的方法能制作高点密度的阵列 (10^6 点/cm²) (McGall *et al.*, 1996)，较机器人点寡核苷酸方法 (3×10^5 点/cm²) (Yershov *et al.*, 1996) 高。目前，用以前的方法设计和制作阵列相当昂贵，限制了潜在用户的使用，此法在以照相平版技术直接将 DNA 合成到阵列过程中产生阵列特异性的遮盖物，使得该技术无法灵活应用。然而，最近一篇报道讲述了一种能降低直接在玻璃片上合成寡核苷酸成本的方法，因为无须产生阵列特异性的遮盖物 (Singh-Gasson *et al.*, 1999)。寡核苷酸阵列区别于 cDNA 阵列的一个特点是能包括与所要代表的基因完全匹配和不完全匹配的寡核苷酸。Affymetrix 的阵列应用了这一技术，包括一个单核苷酸错配的寡核苷酸和完全匹配的寡核苷酸。以此来研究和报告杂交特异性，为每个基因的检测提供了更高的定量准确率。

需要指出的是，寡核苷酸阵列直接提供了可用于杂交的单链模板，而 PCR 产物必须首先变性以提供可用于杂交的单链模板。为使 cDNA 阵列间有最好的重复性，变性操作过程保持一致很重要——这一点在寡核苷酸阵列中不必考虑。

用 mRNA 或总 RNA 标记探针的方法与 cDNA 阵列相同，最常用的是荧光标记线性扩增（如 T7 RNA 聚合酶）产生足够的具有高特异活性的探针。数据分析和数据库流程与 cDNA 阵列相同（见 2.1 部分）。

应用寡核苷酸阵列实例包括同时检测所有酵母基因的表达 (Holstege *et al.*, 1998; Wodicka *et al.*, 1997)，发现小鼠受体酪氨酸激酶激活的信号转导通路的冗余 (Famborough *et al.*, 1999)，以及产热限制对于小鼠骨骼肌衰老的影响 (Lee *et al.*, 1999)。

总的来说，寡核苷酸阵列提供了一种与 cDNA 阵列相似的技术，虽然昂贵了点，但具有更高的基因特异性杂交的准确率。随着基因组测序计划的完成，设计寡核苷酸阵列时需要事先了解的基因序列信息将会变得随手可及，不再成为一个主要问题。

3. 不基于 DNA 阵列的 mRNA 定量方法

基于 DNA 阵列的 mRNA 定量方法只能检测阵列上存在的基因，与此不同的是能够检测生物样品中所有转录物的技术通常被称为“开放体系”。理论上讲，方法的灵敏度

和可供实验使用的资源是该方法的限制因素。应用开放体系的实验可以用于任何物种而无须事先了解物种特异性的序列信息。

3.1 基于序列测定的方法学

cDNA 文库的一过性序列测定 (First-pass sequencing of cDNA libraries) 作为人类基因组计划的一部分, 基因组和表达产物综合分子生物学分析联盟 IMAGE (<http://www-bio.llnl.gov/bbrp/image/image.html>) 构建了数百个来源于不同组织的 cDNA 文库, 并测定其外源 cDNA 片段序列。一过性测定的序列信息 (只读一次的 DNA 序列通常测定克隆插入片段的两端——平均读取 200 ~ 300 bp) 存放在公共数据库中 (由 NCBI 管理的 Genbank 中 dbEST)。也有类似的私有 (如 Incyte 制药有限公司和人类基因组科学有限公司, <http://www.hgsi.com>) 数据库。这些努力的主要目的是尽快为所有的人类基因确定参考序列, 每个一过性测定的序列代表一个基因的表达序列标签 (expressed sequence tag, EST) (与完整的全序列相对)。小鼠和大鼠类似的计划在全球各基因组中心开展。截至 1999 年 11 月 (GenBank 版本号 115), dbEST 中独立的登录条目人类有约 1 300 000, 小鼠 513 000 和大鼠 115 000。cDNA 文库中每一个被测定的克隆代表用于构建文库的原始 mRNA 中的一个独立取样。因此, 通过测定文库中数千个这样的克隆然后计算特定基因序列出现的频率, 就有可能估计出每种转录物的相对丰度。由于这一方法计算代表基因的 EST (或聚类后的 EST) 数量, 因此给出的资料是数字化的。理论上讲, 精确 mRNA 定量中可实现的线性范围是没有界限的。转录物丰度精确定量方法的灵敏度是测定的 cDNA 文库克隆数量的函数。这个方法的魅力在于它能够用于任何可以建立 cDNA 文库的生物样品, 并且是任何有高通量 DNA 测序能力的实验室都可以实现的。当然, 也有一些缺点。较难克隆的 cDNA 无法准确地出现在 cDNA 文库中, 从而低估了它们的丰度。其次, 随机取样使高丰度基因被频繁地反复测序, 甚至连精确检测每个 cDNA 文库中的中丰度的转录物也需要数千个测序反应。NCBI 提供了基于网络的分析根据 xProfiler (<http://www.ncbi.nlm.nih.gov/ncicgap/cgapxpsetup.cgi>) 和数字差异显示 (<http://www.ncbi.nlm.nih.gov/UniGene/ddd.cgi?ORG=Hs>), 可以用来比较来源于数百个 cDNA 文库的序列数据。统计学分析方法也被开发出来以帮助解释此类比较的结果 (Audic and Claverie 1997)。简单地说, 这些方法表明要检测两个 cDNA 文库中转录物丰度统计学上的显著差异, 一个给定的文库中的特定的转录物必须被测序两次以上。所以要检测一个在 20 000 个转录物中有一个拷贝的转录物丰度统计学上的显著差异, 至少要对 cDNA 文库完成 60 000 个独立的测序反应。这个测序量既费时又费钱, 而且可能只提供了中丰度表达基因的信息 (细胞内低丰度转录物在 100 000 拷贝中只有 1 个或更少)。这些实际因素限制了这一用于 mRNA 定量和发现差异表达基因的技术的使用。不过, 通过对 cDNA 文库中 EST 频率分析获得的定量信息的作用已经被以下研究工作证实: 人前列腺特异基因表达的检测 (Vasmatzis *et al.*, 1998), 神经生长因子处理后 PC-12 细胞差异表达基因的确定 (如超氧化物歧化酶和 synapsin 2) (Lee *et al.*, 1995)。在人类造血梯次分化体系中也得到证实 (Claudio *et al.*, 1998)。在所有的事例中原始的实验发现均得到了 Northern 杂交的确认。

总之, 虽然启始于 cDNA 文库的一过性测序数据对 mRNA 定量确实有效, 但因为

它的高成本和低灵敏度，在大多数情况下它并不是首选。假设许多 cDNA 文库的信息 (dbEST) 以及比较和解释此类资料的根据都能免费得到，计划要开展差异 mRNA 表达研究的人应该把对这些 EST 的“电子比较”的应用看作第一个“沿途停靠的港口”。显然，只有当序列数据来源于与研究者相关的 cDNA 文库，这一方法才有意义。这一方法仅适用于标准的 cDNA 信息，而不适用于归一化的 cDNA 文库资料，因为归一化过程造成文库中 cDNA 的数量表征发生根本变化。

基因表达系列分析——SAGE 基因表达系列分析 (serial analysis of gene expression, SAGE) 方法由 Bert Voglestein 实验室 (Velculescu *et al.*, 1995) 于 1995 年首次发表，是一种基于 DNA 测序的用于 mRNA 表达丰度定量的技术。SAGE 克服了通过文库插入片段序列测定进行基因表达分析技术的主要缺点——必须测定上万个 cDNA 克隆以保证量值的精确度 (见前述)。SAGE 的基本原理包括分离代表每个转录物特定区域的独特的短序列标签 (9~14 个碱基)。然后将它们串联、克隆、测序 (见图 2.3)。所有标签中某种特定序列标签出现的频率就是原始材料中相应的 mRNA 频率的度量。理论上讲一个含有 9~14 个碱基的序列标签能提供足够的序列信息来确定特定的 mRNA (Velculescu *et al.*, 1995)。比如，假设随机分布成立，10 个碱基的排列 (4^{10}) 产生 1 048 576 种可能的组合，10 倍于组成人类基因组的基因的数量 (Fields *et al.*, 1994)。因此，通过将 DNA 序列长度降到有信息量的最低限度，使效率比 cDNA 文库测序方法有所提高：测定每个“序列标签串联体”的序列得到 30~50 倍的基因信息 (Bertelsen and Velculescu, 1998)。

像 cDNA 文库测序一样，SAGE 资料是数字化的，理论上线性范围是无限的。基于模拟 (Zhang *et al.*, 1997) 和统计分析 (Audic and Claverie, 1997) 的方法已经开发出来，用于评估来源于两个生物样品的序列标签丰度差异的意义。使用独立的方法证实 SAGE 序列标签频率在小于 0.01% 到大于 0.1% 的范围内是 mRNA 转录物丰度的能精确度量 (Madden *et al.*, 1997; Velculescu *et al.*, 1995, 1997)。SAGE 的应用包括发现结肠癌细胞系凋亡前 P53 诱导基因的发现 (Polyak *et al.*, 1997) 和正常与癌细胞基因表达谱的分析 (Zhang *et al.*, 1997)。更多的 SAGE 应用实例能在以下两个网站找到：NCBI (<http://www.ncbi.nlm.nih.gov/SAGE/>) 和 Genzyme Oncology (一个提供 SAGE 作为商业服务的生物技术公司) (<http://www.genzyme.com/prodserv/molecular-oncology/sage/welcome.htm>)。

“标准” SAGE 方法需要近似于可供 cDNA 文库构建所需的 RNA 量 (2.5~5.0 μg mRNA)。这把 SAGE 的应用限制在起始的生物材料很多的情况下：细胞培养体系和动物模型，而非人临床活组织检查样品和微切割样品。不过，最近微 SAGE (Datson *et al.*, 1999) 的出现提供了一种改良的 SAGE 技术，能利用少至 5 ng 的 mRNA。

SAGE 技术有两个主要的缺陷。第一，由于用短序列标签辨别每个转录物，高质量是准确确认序列的根本。在含有 10 个核苷酸的标签中 1 个碱基的读取错误对于转录物确定的影响 (通过 BLAST 分析) 远比一个 200 碱基、含有 10% 错误率的一过性测定序列产生的影响要大。其次，用标签成功确定原始转录物的能力直接取决于每个物种的数据库中存储序列 (特别是 3' 端的序列) 的数量。因此，目前人类样品的 SAGE 资料解释要比小鼠或大鼠的容易。

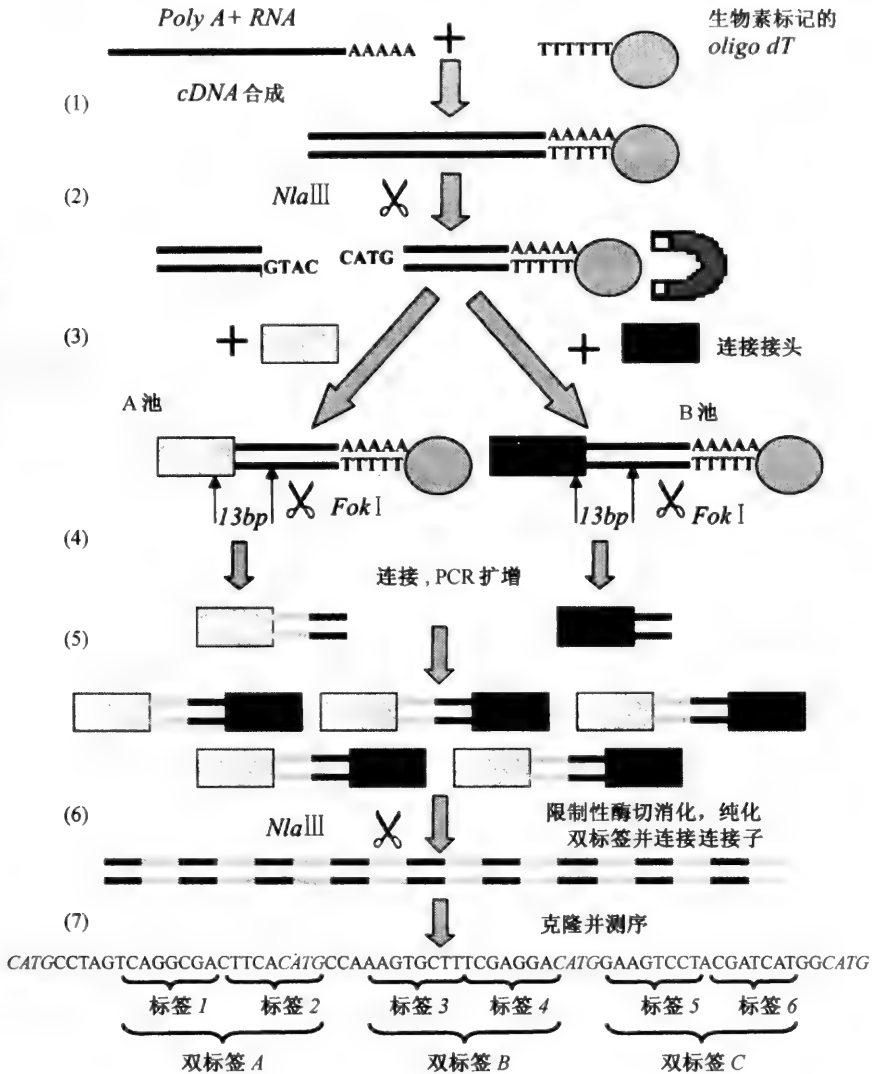


图 2.3 基因表达系列分析 (SAGE) 方法示意图

步骤 1: 用生物素-oligo dT 引物将 PolyA+ 的 RNA 转换为 cDNA。步骤 2: 合成的 cDNA 用高频切割的限制性核酸内切酶, “锚定酶”(如 *Nla* III, 4bp 识别位点, 平均 256 核苷酸一个酶切位点) 消化。步骤 3: cDNA 的 3' 端被生物素包裹的磁珠捕获。cDNA 池被一分为二 (“A 池” 和 “B 池”), 不同的接头被连接到每个 cDNA 池中。步骤 4: 这些接头含有 II_s 型限制性核酸内切酶位点; tagging enzyme (如此命名是因为它们的切割位点离识别位点有确定的碱基数, 比如 *Fok* I 在它的识别位点下游 13 个碱基处切割) 被用来切割 cDNA/接头杂合子, 释放出附着在接头上的 cDNA 短序列标签。步骤 5: 接头-标签杂合分子尾对尾相连, PCR 扩增 (通过接头序列)。步骤 6: 用锚定酶将接头从双标签上去除, 纯化双标签并将它们彼此连接形成连接子 (concatemer)。步骤 7: 连接子克隆与测序, 发现每个标签的特征。

总之，SAGE 被广泛应用于任何生物学体系。结合自动 DNA 序列分析和足够的生物信息学支持，它是一种用于 mRNA 丰度定量的有效而精确的方法

3.2 差异显示 PCR (DD-PCR)

差异显示 PCR 是一种确定两个生物学样品差异表达的 cDNA 片段的方法 (Liang and Pardee, 1992) (见图 2.4a)。这一方法的基础是用两个寡核苷酸引物从 mRNA 产生 cDNA 片段，一个是与转录物的 polyA 尾互补 [如 oligo-d(T)₁₁VN]，另一个是短的随机核苷酸序列 (例如 10 寡核苷酸)。当有足够的引物组合时，DD-PCR 能确定一个生物学样品中所有的转录物。cDNA 合成后，片段在 PCR 扩增过程中被标记 (放射性同位素或荧光标记)。产物通过测序凝胶电泳分离，扩增后的 cDNA 片段的组成模式就显现出来。

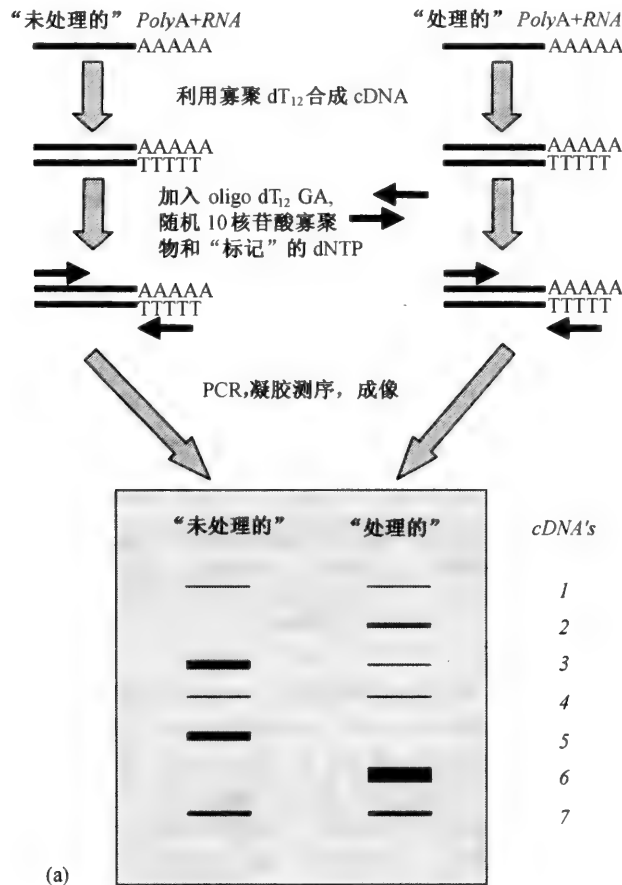


图 2.4 差异显示方法学示意图

(a) 反转录 PCR 产物差异显示的标准方法 (Liang and Pardee, 1992)。来源于两种生物学样品的 RNA (如调节剂处理与未处理的细胞系) 用 oligo-d (T)₁₂ 反转录, 得到的 cDNA 分别用一系列锚定 oligo dT 引物 [此处用 d (T)₁₂GA] 结合一个 10 核苷酸随机引物 (结合不同 cDNA 的不同位置) 进行 PCR 扩增。用测序胶电泳可观察 PCR 产物 (PCR 过程中实现放射性同位素或荧光标记)。同样电泳迁移率的条带的有/无提示了条带所代表的 cDNA 来源基因的表达差异。

被标记条带的强度反应了相应转录物的相对丰度。当使用相同的引物对时，来源于两个生物学样品 cDNA 条带的主要差别显示出它们之间差异表达的转录物。对 cDNA 条带进行克隆和测序就能鉴定基因。早期报告显示 DD-PCR 产生大量的假阳性 (Liang *et al.*, 1994; McClelland *et al.*, 1995)。显然这大大降低了它的效率。因为这些缺点，人们开始尝试各种技术改进，诸如优化引物组合、提高 PCR 引物复性温度、反转录前用 DNA 酶处理 RNA 和使用非变性凝胶 (Bauer *et al.*, 1993; Liang and Pardee, 1995)。或许最有意义的改进是用限制性内切核酸酶消化双链 cDNA (用 mRNA 制备)，然后通过连接接头介导 cDNA 片段 3' 端选择性 PCR 扩增 (Prashar and Weissman, 1996; 见图 2.4b)。限制性内切核酸酶的选择使用增加了能显示的转录物的覆盖率。这些改进通过让引物在 56℃ 退火进行严格的 PCR，实现了比标准 DD-PCR 条件 (40℃ 下非严格的引物退火) 显著的改良。改良的 DD-PCR 据说可以提供基因表达的近似定量的信息；每个转录物

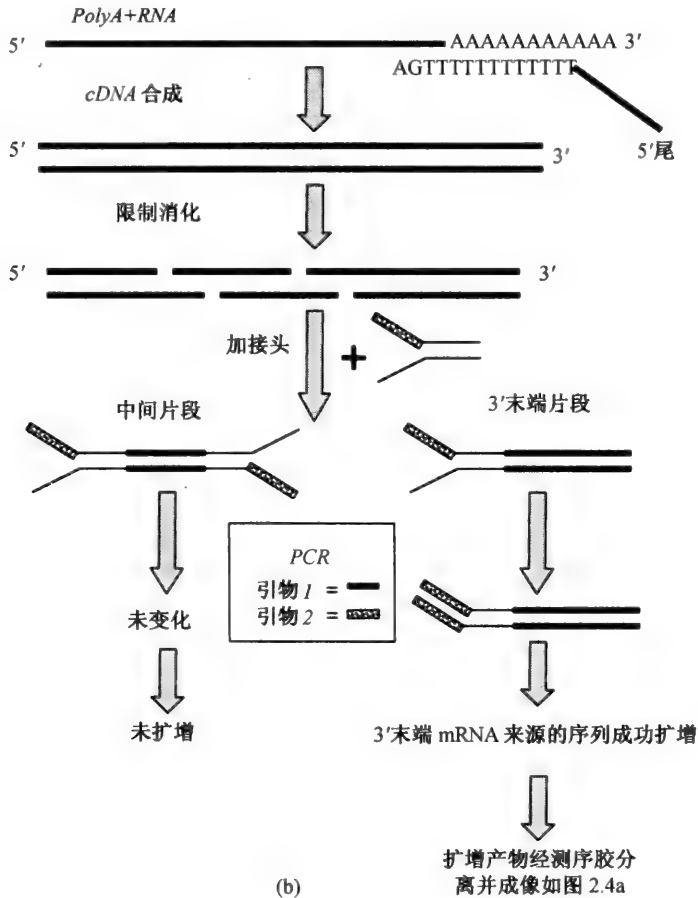


图 2.4 (续)

(b) 为增强重复性而对差异显示技术的改良 (Prashar and Weissman, 1996)。以含有 5' 尾 (粗体) 的锚定 dT 引物，就是 CTCGCAATCGGGGCTCGTCG (T)₁₂GA 来驱动第一链 cDNA 的合成。经过标准的 cDNA 第二链合成后的产物经限制性核酸内切酶消化后与 Y 型的接头连接。PCR 反应有选择地扩增 cDNA 片段的 3' 端。PCR 产物能用测序凝胶电泳分离和鉴定 (如前所述)。

每种限制性内切核酸酶产生单一条带，降低了显示样式的复杂性。还有，序列已知的基因产生的条带大小可以预测，能在把条带从凝胶上回收之前临时确认基因 (Prashar and Weissman, 1996)。这一技术的改良由几家独立的生物技术公司完成 [Curagen 公司, <http://www.curagen.com/> (Shimkets *et al.*, 1999); 数码基因技术公司, <http://www.dgt.com/index.html> 和显示系统生物技术公司, <http://www.displaysystem.com/>], 用来提供快速、高通量的 DD-PCR 服务。

DD-PCR 被用于许多不同的研究领域: 比如新型药靶的发现 (Shiue, 1997; Wang and Feuerstein, 1997), 风湿性关节炎差异调控基因的发现 (White and Petkovich, 1998) 和环境刺激物对于细菌基因表达影响的评估 (Fislage, 1998)。事实上, DD-PCR 是目前确定差异表达基因的技术中最广泛发表的一种。这或许反映了这样一个事实: DD-PCR 不需要专门的昂贵设备或生物信息学工具, 促使它能进入许多实验室。DD-PCR 的一个重要优势是相对较少的 RNA 用量。最初需要几百纳克 RNA (Liang and Pardee, 1992)。最近的改良使该技术能对来源于单个细胞的 RNA 进行 DD-PCR (Renner *et al.*, 1998)。因此, RNA 的产量不会阻碍 DD-PCR 的应用。不过 DD-PCR 也存在缺点。如前所述, 这一技术受假阳性之苦。还有, 一条被确认是差异表达的条带可能并不总是只包括单一的分子, 因为多个转录物能产生近似大小片段在电泳胶上一起泳动。造成确认基因的困难。

总之, DD-PCR 能成功地用于发现任何物种的任何组织的差异表达基因, 只要能分离高质量的 RNA。该技术简单的操作和相对廉价使它广为使用。然而结果通常更是定性的而非定量的, 与其他技术如阵列和 SAGE 产生的数据比较错误似乎更多。

4. 结论

分子生物学和基因组序列分析领域技术的进步促使 mRNA 定量方法由一个接一个基因的手段向基因组范围的检测发展。在过去的 5 年里, 一些彼此互补的技术同时发展起来。每一项技术在以下几方面存在各自相对的优缺点: 检测的灵敏度、通量和基因组的范围, 数据分析的可靠性、成本和简易性/能力。所有方法都需要有意义的实验输入和后续的工作来验证新发现。由于这个原因, 目前尚无公开发表的例子对由两种技术 (如 cDNA 阵列和 SAGE) 得到的结果进行全方位的对比。因此, 很难就每种技术的效用提供彻底的比较。不过毫无疑问, 用于 mRNA 定量的 SAGE、基于阵列的杂交、DD-PCR 和电子差异基因表达等方法的效用已经被用独立的方法 (RT-PCR 或 Northern 杂交) 进行发现的确认所证实。另外一些确定差异表达基因的方法, 如消减抑制杂交 (Diatchenko *et al.*, 1996) 和代表性差异分析 (Hubank and Schatz, 1994) 等未在此讨论, 因为它们起初是为确定 mRNA 表达量的差异 (与 SAGE、阵列等相比) 而设计的, 或者还未像诸如 DD-PCR 等定量方法那样被广泛使用。

(鱼咏涛译 贺福初校)

参考文献

- Aitman, T.J., Glazier, A.M., Wallace, C.A. *et al.* (1999) Identification of *CD36 (Fat)* as an insulin-resistant gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21**: 76–83.
- Alizadeh, A.A., Eisen, M.B., Davis, R.E. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* **403**: 503–511.
- Audic, S. and Claverie J-M. (1997) The significance of digital gene expression profiles. *Genome Res.* **7**: 986–995.
- Bauer, D., Muller, H., Reich, J., Riedel, H., Ahrenkeil, Warthoe, P. and Strauss, M. (1993) Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Res.* **21**: 4272–4280.
- Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S. and Small, P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**: 1520–1523.
- Bertelsen, A.H. and Velculescu, V.E. (1998) High-throughput gene expression analysis using SAGE. *Drug Discov. Today* **3**: 152–158.
- Claudio, J.O., Liew, C.C., Dempsey, A.A., Cukerman, E., Stewart, A.K., Na, E., Atkins, H.L., Iscove, N.N. and Hawley, R.G. (1998) Identification of sequence-tagged transcripts differentially expressed within the human hematopoietic hierarchy. *Genomics* **50**: 44–52.
- Datson, N.A., van der Perk-de Jong, J., van den Berg, M.P., de Kloet, E.R. and Vreugdenhill, E. (1999) MicroSAGE: a modified procedure for the serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res.* **27**: 1300–1307.
- DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**: 457–460.
- Diatchenko, L., Lau, Y.F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. (1996) Suppression subtraction hybridisation – a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA* **93**: 6025–6030.
- Endege, W.O., Steinmann, K.E., Boardman, L.A., Thibodeau, S.N. and Schlegel, R. (1999) Representative cDNA libraries and their utility in gene expression profiling. *BioTechniques* **26**: 542–550.
- Ermolaeva, O., Rastogi, M., Pruitt K.D., Schuler, G., Bittner, M.L., Chen, Y., Simon, R., Meltzer, P., Trent, J.M. and Boguski, M.S. (1998) Data management and analysis for gene expression arrays. *Nature Genet.* **20**: 19–23.
- Fambrough, D., McClure, K., Kazlauskas, A. and Lander, E.S. (1999) Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent sets of genes. *Cell* **97**: 727–741.
- Fields, C., Adams, M.D., White, O. and Venter, J.C. (1994) How many genes in the human genome? *Nature Genet.* **7**: 345–346.
- Fislag, R. (1998) Differential display approach to quantitation of environmental stimuli on bacterial gene expression. *Electrophoresis* **19**: 613–616.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D. and Lander, E.S. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**: 531–537.
- Gonzalez, P., Zigler Jr, J.S., Epstein, D.L. and Borras, T. (1999) Identification and isolation of differentially expressed genes from very small tissue samples. *BioTechniques* **26**: 884–892.
- Gress, T.M., Muller-Pillasch, F., Geng, M., Zimmerhackl, F., Zehetner, G., Friess, H., Buchler, M., Adler, G. and Lehrach, H. (1996) A pancreatic cancer-specific expression profile. *Oncogene* **13**: 1819–1830.
- Gygi, S.P., Rochon, Y., Franz, B.R. and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**: 1720–1730.
- Hastie, N.D. and Bishop, J.O. (1976) The expression of three abundance classes of messenger mRNA in mouse tissues. *Cell* **9**: 761–774.
- Heller, R., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis,

- R.W. (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl Acad. Sci. USA* **94**: 2150–2155.
- Holstege, F.C.P., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner C.J., Green, M.R., Golub, T.R., Lander, E.S. and Young, R.A. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**: 717–728.
- Hubank, M. and Schatz, D.G. (1994) Identifying differences in mRNA expression by representational analysis of cDNA. *Nucleic Acids Res.* **22**: 5640–5648.
- Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C.F., Trent, J.M., Staudt, L.M., Hudson Jr, J.H., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D. and Brown, P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283**: 83–87.
- Lee, C-K., Klopp, R.G., Weindruch, R. and Prolla, T.A. (1999) Gene expression profile of aging and its retardation by caloric restriction. *Science* **285**: 1390–1393.
- Lee, N.H., Weinstock, K.G., Kirkness, E.F., Earle-Hughes, J.A., Fuldner, R.A., Marmaros, S., Glodek, A., Gocayne, J.D., Adams, M.D., Kerlavage, A.R., Fraser, C.M. and Venter, J.C. (1995) Comparative expressed-sequence tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. *Proc. Natl Acad. Sci. USA* **92**: 8303–8307.
- Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**: 967–971.
- Liang, P. and Pardee, A.B. (1995) Recent advances in differential display. *Curr. Opin. Immunol.* **7**: 274–280.
- Liang, P., Zhu, J., Zhang, X., Guo, Z., O'Connell, R.P., Averboukh, L., Wang, F. and Pardee, A.B. (1994) Differential display using one-base anchored oligodT primers. *Nucleic Acids Res.* **22**: 5763–5764.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**: 1675–1680.
- Madden, S.L., Galella, E.A., Zhu, J., Bertelsen, A.H. and Beaudry, G.A. (1997) SAGE transcript profiles for p53-dependent growth regulation. *Oncogene* **15**: 1079–1085.
- Maier-Ewert, S., Maier E., Ahmadi, A., Curtis, J. and Lehrach, H. (1993) An automated approach to generating expressed sequence catalogues. *Nature* **361**: 375–376.
- Marshall, E. (1999) Do-it-yourself gene watching. *Science* **286**: 444–447.
- McClelland, M., Mathieu-Daude, F. and Welsh, J. (1995) RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends Genet.* **11**: 242–246.
- McGall, G., Labadie, J., Brook, P., Wallraff, G., Nguyen, T. and Hinsberg, W. (1996) Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc. Natl Acad. Sci. USA* **93**: 13 555–13 560.
- Moch, H., Schrami, P., Bubendorf, L., Mirlacher, M., Kononen, J., Gasser, T., Mihatsh, M.J., Kallioniemi, O.P. and Sauter, G. (1999) High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal carcinoma. *Am. J. Pathol.* **154**: 981–986.
- Patten C., Clayton, C.L., Blakemore, S.J., Trower, M.K., Wallace, D.M. and Hagan, R.M. (1998) Identification of two novel diurnal genes by screening a rat brain cDNA library. *Neuroreport* **9**: 3935–3941.
- Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenshikov, A., Williams, C.F., Lee, J.C.F., Lashkari, D., Shalon, D., Brown, P.O. and Botstein, D. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **16**: 9212–9217.
- Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W. and Vogelstein, B. (1997) A model for p53 induced apoptosis. *Nature* **389**: 300–305.
- Prashar, Y. and Weissman, S. M. (1996) Analysis of differential gene expression by display of 3' end restriction fragments of cDNA's. *Proc. Natl Acad. Sci. USA* **93**: 659–663.
- Renner, C., Trumper, L., Pfitzenmeier, J-P., Loftin, U., Gerlach, G., Stehle, I., Wadle, A. and Pfreundschuh, M. (1998) Differential mRNA display at the single cell level. *BioTechniques* **24**: 720–724.
- Richmond, C.S., Glasner, J.D., Mau, R., Jin, H. and Blattner, F.R. (1999) Genome-wide expression profiling in *Escherichia coli* K-12. *Nucleic Acids Res.* **27**: 3821–3835.

- Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. and Davis, R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl Acad. Sci. USA* **93**: 10 614–10 619.
- Shalon, D., Smith, S.J. and Brown P.O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridisation. *Genome Res.* **6**: 639–645.
- Shinkets, R.A., Lowe, D.G., Tsu-Ning Tai, J., Sehl, P., Jin, H., Yang, R., Predki, P.F., Rothberg, B.E.G., Murtha, M.T., Roth, M.E., Shenoy, S.G., Windemuth, A., Simpson, J.W., Simons, J.W., Daley, M.P., Gold, S.A., McKenna, M.P., Hillan, K., Went, G.T. and Rothberg, J.A. (1999) Gene expression analysis by transcript profiling coupled to a gene database query. *Nature Biotechnol.* **17**: 798–803.
- Shiue, L. (1997) Identification of candidate genes for drug discovery by differential display. *Drug Devl. Res.* **41**: 142–159.
- Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R. and Cerrina, F. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnol.* **17**: 974–978.
- Syed, F., Blakemore, S.J., Wallace, D.M., Trower, M.K., Johnson, M., Markham, A.F. and Morrison, J.F.J. (1999) CCR7 (EBI 1) receptor down-regulation in asthma: differential gene expression in human CD4+ T lymphocyte. *Q. J. Med.* **92**: 463–471.
- Vasmatazis, G., Essand, M., Brinkmann, U., Lee, B. and Pastan, I. (1998) Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proc. Natl Acad. Sci. USA* **95**: 300–304.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**: 484–487.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997) Characterization of the yeast transcriptome. *Cell* **88**: 243–251.
- Wang X. and Feuerstein G.Z. (1997) The use of mRNA differential display for discovery of novel therapeutic targets in cardiovascular disease. *Cardiovasc. Res.* **35**: 414–421.
- Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L. and Somogyi, R. (1998) Large-scale gene expression mapping of central nervous system development. *Proc. Natl Acad. Sci. USA* **95**: 334–339.
- White, J.A. and Petkovich, M. (1998) Identification and cloning of RA-regulated genes by mRNA-differential display. *Methods Mol. Biol.* **89**: 389–404.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. and Lockhart, D.J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol.* **15**: 1359–1367.
- Yershov, K., Barsky, V., Belgovisky, A., Kirillov, E., Kreindlin, E., Ivanov, I., Parinov, S., Guschin, D., Drobishev, A., Dubiley, S. and Mirzabekov, A. (1996) DNA analysis and diagnostics on oligonucleotide microchips. *Proc. Natl Acad. Sci. USA* **93**: 4913–4198.
- Zehetner, G. and Lehrach, H. (1994) The Reference Library System – sharing biological material and experimental data. *Nature* **367**: 489–491.
- Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B. and Kinzler, K.W. (1997) Gene expression profiles in normal and cancer cells. *Science* **276**: 1268–1272.

第三章 蛋白质组分析中的二维聚丙烯酰胺凝胶电泳

Michael J. Dunn and Angelika Görg

1. 引言

对来自于全细胞、组织或生物体中包含多达几千种蛋白质的蛋白混合物的分离、检测和分析是蛋白质组分析的首要要求。最近,多种用于蛋白质组学研究的蛋白质分离方法得到了发展(见第十章),如基于芯片技术的应用(Merchant and Weinberger, 2000; Nelson *et al.*, 2000)、蛋白复合物的质谱直接分析(Link *et al.*, 1999)、亲和标签的使用(Gygi *et al.*, 1999; Rigaut *et al.*, 1999)以及大规模酵母双杂交筛选系统(Uetz *et al.*, 2000)。但是,二维聚丙烯酰胺凝胶电泳(two-dimensional polyacrylamide gel electrophoresis, 2-DE)依然是大多数蛋白质组研究中分离复杂蛋白混合物所选择的核心技术,这是由于它在同时分离成千蛋白质时所具有的无可比拟的分离能力,以及随后的可用于计算机定量分析差异表达蛋白质的高灵敏度的显色技术,还有对2-DE胶上蛋白质用高灵敏度微量化学方法进行鉴定和表征相对比较容易。本章将要介绍的是2-DE方法的最新进展,这些进展使得2-DE的分辨能力逐渐提高,成为操作相对简便的具有高重复性的分离技术。

2. 二维聚丙烯酰胺凝胶电泳发展史

最早的二维电泳应归功于Smithies和Poulik(1956)的工作,他们将纸电泳和淀粉凝胶电泳结合来分离血清蛋白质,随后在电泳技术上的研究开发,如聚丙烯酰胺支持介质和聚丙烯酰胺浓度梯度,很快应用于二维分离(Dunn, 1987)。特别是等电聚焦(isoelectric focusing, IEF)技术应用于二维分离使基于蛋白质电荷属性的一向分离成为可能。IEF和二维十二烷基硫酸钠聚丙烯酰胺凝胶电泳(sodium dodecyl sulfate polyacrylamide gel electrophoresis, SDS-PAGE)的结合形成了根据两个独立参数(电荷和分子质量大小)来分离蛋白质的二维方法。通过在IEF中应用尿素和非离子性去污剂,可将该方法应用于各种溶解性能差别很大的样品。至此,到1975年,正式形成了应用于全细胞或组织蛋白质混合物分析的二维电泳系统(Iborra and Buhler, 1976; Klose, 1975; Scheele, 1975)。

基于以上发展,O'Farrell发表了用于分离大肠杆菌的2-DE优化方法。该方法在一维利用4%T、5%C柱状PAGE胶,内含8mol/L尿素和2%(m/V)NP-40,在二维利用Laemmli(1970)的非连续SDS-PAGE凝胶系统。这种基于电荷分离(等电点, pI)和分子质量大小分离(相对分子质量, M_r)的正交组合使得样品蛋白质点分布于整个二维凝胶图谱上。过去25年来大部分2-DE研究都以该方法为基础,已有数千篇相关的文献发表。

3. 固相 pH 梯度 (immobilized pH gradient, IPG) 二维电泳

对于蛋白质组研究来说, 2-DE 在分离蛋白质时必须具有高的重复性和分辨率。不久前这仍是一个很大的问题, 原因在于生成 IEF pH 梯度的合成载体两性电解质 (synthetic carrier ampholyte, SCA) 的特性。SCA 是通过复杂的合成过程得到的, 很难控制其重复性, 由此产生相当大的批次之间变化, 限制了 2-DE 分离的重复性。可能更为重要的是, SCA 分子相对较小, 难以固定在 IEF 胶内, 因此在 IEF 过程中由水引起的电渗流将导致 SCA 分子向阴极迁移。这种被称为“阴极漂移”的现象将导致 pH 不稳定。当利用管状 IEF 胶时, 由于玻璃毛细管壁上负电荷基团的作用将加剧阴极漂移。实际上, 在采用 O'Farrell 方法时, pH 梯度很少超出 7 太多, 所带来的后果则是碱性区蛋白质的丢失。注意到这种现象, O'Farrell 变更了方法, 引入非平衡 pH 梯度电泳 (non-equilibrium pH gradient electrophoresis, NEPHGE), 用 2-DE 分离碱性蛋白质 (O'Farrell *et al.*, 1977)。这种方法在快速形成的 pH 梯度中根据蛋白质移动能力进行分离, 但其重现性很难控制。幸运的是, 在 Immobiline 试剂 (Amersham Pharmacia Biotech) 基础上开发的固相 pH 梯度 (immobilized pH gradient, IPG) 技术 (Bjellqvist *et al.*, 1982) 解决了这个问题。虽然早期 IPG 技术在 2-DE 尝试遇到很多问题, 但它们被逐渐克服 (Görg *et al.*, 1988, 2000) 并使得 IPG IEF 成为目前在蛋白质组应用中大多数 2-DE 首选的方法。这种称为 IPG-DALT 的 2-DE 方法, 将在本章详细讨论, 最近的一篇文章也对 SCA IEF 在 2-DE 中的应用进行了详细讨论 (Monribot and Boucherie, 2000)。

4. 样品制备

由于 2-DE 所分析样品的多样性, 没有一种制备方法可以普遍适用于各种样品, 但有点考虑一定要提出: 很有必要尽量减少可能在 2-DE 谱中导致假点 (artificial spot) 的蛋白质修饰, 在实验中, 含尿素的样品一定不要加热, 因为加热后尿素分解产生的异氰酸盐可能对蛋白质产生氨基甲酰化修饰而引起很大的电荷不均一性。此外, 样品中的蛋白酶可以很容易地生成假点, 因此, 样品处理步骤应尽量少并保持冷冻环境。样品中可以加入蛋白酶抑制剂的混合物, 但切记这些试剂可修饰蛋白, 导致电荷假象。

4.1 可溶性样品

可溶性液体样品如血清、血浆、尿样、脑脊液以及细胞和组织的水溶性提取物, 经很少的前处理便可进行 2-DE 分析。含蛋白质浓度较高的样品, 如血浆和血清, 可用适量样品溶解缓冲液 (见第 5 部分) 稀释后直接分析。但是, 诸如白蛋白和免疫球蛋白一类的高丰度蛋白质在 2-DE 图谱上常干扰其他表达量少的成分, 将这些高丰度蛋白质去除可以克服这个问题 (Dunn, 1987; Lollo *et al.*, 1999), 但同时可能非特异性去除其他蛋白质。

其他类型的液态样品可能含有较低的蛋白质浓度或较高盐浓度, 将会干扰 IEF 时蛋白质的分离, 这种样品可在 2-DE 分离前用透析或液相色谱脱盐, 样品可以通过冻

干、对聚乙二醇的透析或用 TCA/丙酮沉淀的方法进行浓缩。研究发现 TCA/丙酮沉淀对导致蛋白酶的失活进而减少蛋白质降解非常有用，同时可以去除干扰化合物，富集极碱性蛋白质，如从全细胞裂解液中富集核糖体蛋白 (Görg *et al.*, 1997)。

4.2 组织样品

固体组织样品常在溶解缓冲液 (见第 5 部分) 中破碎，最好的方法是在组织冷冻及液氮温度的条件下破碎。包裹在铝箔并贮存于液氮中的较小组织样品，可以夹在两个冷研块或基体中间用杵和研钵在液氮环境下压碎，大的组织块可在溶解缓冲液中用旋转叶片型匀浆器进行匀浆处理，但尽量避免加热和起泡沫。

组织样品的一个特殊问题是样品的异质性，组织样品和来源于疾病的样品包含许多种不同类型的细胞，不太可能分别收集组织的疾病区域和非疾病区域。几种用于 2-DE 分析的方法可以解决这个问题，通常用基于免疫亲和技术的正性或负性选择，可以实现对特殊细胞群的富集，这些样品可以直接应用于 2-DE 分析或是在 2-DE 上样前进行短期培养以增加细胞数目。此外，微量切割技术可用来自组织切片中获取纯的细胞群，其中最受关注的是激光俘获微切割 (laser capture microdissection, LCM) 技术 (Simone *et al.*, 1998)。在该技术中，在倒置显微镜下将激光束聚焦在所选择的组织切片区域，用激光束活化与组织切片接触的转移膜来选择细胞。激光束准确击中的区域转移膜与紧贴着的组织结合在一起，因此，可以将这些组织中细胞从周围组织中提取出来。可以用 PCR 技术扩增这些微量标本中所含核酸，所以 LCM 技术很容易地应用到核酸分析中。与此对应的，缺乏相应的蛋白质扩增手段将是蛋白质组分析中 LCM 分离细胞所遇到的主要挑战。而 Banks 等 (1999) 则证实了 LCM 在蛋白质组分析中的应用，在他们的研究中，通过对一例正常人宫颈表皮组织切割分析，与整个组织切割分析对比，结果显示出一一些蛋白质由 LCM 而得到富集。

4.3 细胞

对体外悬浮培养生长的细胞或周期性细胞样品 (如红细胞、淋巴细胞)，理想的方法是通过离心收集。用磷酸盐缓冲液清洗并溶于样品缓冲液 (见第 5 部分)。对于在固体基质中培养的细胞，应首先去除培养基质，用 PBS 或等渗蔗糖溶液清洗细胞层，后者是减少可能干扰一向 IEF 的盐的特别有效的方法。然后通过刮擦方法收获细胞，但应避免使用蛋白裂解酶，或者，加入少量体积溶解缓冲液，将附着在基质上的细胞直接裂解。如果细胞核酸含量过高而导致样品黏度太大时，可用不含蛋白质酶的 DNase/RNase 混合物来处理样品 (Garrels, 1979)。

4.4 样品分级

由于真核组织蛋白质表达的高动态范围和多样性，有时必须对蛋白质进行分级处理，以降低样品的复杂性并富集低拷贝蛋白质。蛋白质预分级可以用亚细胞分级、液相电泳、吸附色谱和选择性沉淀等方法来实现 (Corthals *et al.*, 2000)。此外，还有一种方法是根据蛋白质在一系列溶解能力逐渐增强的缓冲液中溶解性能不同，而实现对细胞或组织蛋白质的分级提取 (Cordwell *et al.*, 2000; Molloy *et al.*, 1998)。

5. 溶解

理想的 2-DE 溶解应能达到破坏所有非共价结合蛋白质复合物和聚积体, 形成各个多肽的溶解液。如果不能达到这一点, 样品中结合牢固的蛋白质复合物可能使 2-DE 中出现新的蛋白质点, 相应地表示单个多肽的点强度将下降。此外, 溶解方法必须允许可能干扰 2-DE 分离的盐、脂类、多糖和核酸等物质的去除。最后, 样品中蛋白质在 2-DE 过程中必须保持可溶性。鉴于上述原因, 样品溶解是 2-DE 成功分离蛋白质的最关键因素之一。

最为普遍应用的 2-DE 蛋白质溶解方法仍然是最早被 O'Farrell 所描述的以下几种物质的混合液(所谓“裂解缓冲液”): 9.5mol/L 尿素、4% (m/V) 的 NP-40、1% (m/V) 的还原剂二硫苏糖醇(dithiothreitol, DTT) 以及 2% (m/V) 相应 pH 范围的 SCA。该配方适用于许多类型的样品, 但并不是一种万能方法, 其中在膜蛋白质的提取上就对此方法提出了挑战(Santoni *et al.*, 2000)。此后发现两性去污剂 CHAPS 可以有效溶解膜蛋白质(Perdew *et al.*, 1983), 特别是在浓度为 4% (m/V), 与 2mol/L 硫脲和 8mol/L 尿素混合物共用时(Molloy *et al.*, 1998; Rabilloud *et al.*, 1997)。硫脲是一种比尿素具有更强能力的变性剂, 但由于其水溶性差, 不能单独使用, 但在尿素浓溶液中其可溶性更好, 因此尿素-硫脲混合物提高了溶解样品的能力。线性磺基三甲胺乙内酯(sulphobetaine) 去污剂, 如 SB3-10 或 SB3-12, 也是有效的溶解剂, 但这些试剂与高浓度尿素不兼容(Rabilloud *et al.*, 1997), 当它们以 2% (m/V) 的浓度与 5mol/L 尿素、2mol/L 硫脲和 2% CHAPS 结合在一块使用时可以克服这些困难(Herbert, 1999; Rabilloud *et al.*, 1997)。

去污剂 SDS 能破坏大多数非共价蛋白质相互作用, 对于膜蛋白的溶解非常有效, 但其阴离子特性使它不可能应用于 IEF 凝胶。但是, SDS 可用于 2-DE 前的前期溶解过程, 在该方法中, 样品首先用 1% (m/V) SDS 溶解, 然后用常用溶解液(如裂解液)稀释, 目的是为了用非离子或两性去污剂从蛋白质中置换出 SDS, 从而使蛋白质保持在溶解状态。为了有效地溶解样品, 并减少 SDS 在 IEF 中的负作用, 必须认真控制蛋白质与非离子或两性去污剂的比例(1:3) 及 SDS 与非离子或两性去污剂的比例(1:8)(Dunn, 1987)。

核酸对于 IEF 可能造成的问题, 是因为它会导致样品黏度的增加且在某些情况下可与样品蛋白质形成复合物, 进而导致假象迁移和条纹的出现。如果怀疑上述问题可能出现, 最好向样品溶解液中加入适当纯的(不含蛋白酶)内切核酸酶来降解核酸, 或是利用 SCA 和核酸结合形成复合物的能力通过超速离心来去除复合物(Rabilloud *et al.*, 1986)。

6. 还原

蛋白质二硫键通常用 DTT 或 β -巯基乙醇等含自由巯基的试剂还原。但是, DTT 之类的试剂带电, 因此在 IEF 过程中迁移到胶外, 导致样品蛋白质的重氧化使得电泳过程中蛋白质溶解性丧失, 最近报道用不带电荷的还原剂如三正丁基膦取代含巯基试剂大大增强一向 IEF 中蛋白质溶解能力, 并使得一维到二维转移效率提高(Herbert *et*

al., 1998; Herbert, 1999)。

7. 一维固相 pH 梯度等电聚焦 (IEF with IPG)

IPG IEF 胶用 Immobilines (Amersham Pharmacia Biotech) 制备。Immobilines 是拥有 $\text{CH}_2=\text{CH}-\text{CO}-\text{NH}-\text{R}$ 结构的 8 种丙烯酰胺衍生物系列, 其中 R 包含羧基或叔氨基团, 它们构成了分布在 pH3~10 不同值的缓冲体系。根据公布的配方计算后, 将适宜的 IPG 试剂添加至混合物中用于凝胶聚合, 在聚合中, 缓冲基团通过乙烯键共价聚合至聚丙烯酰胺骨架中, 形成 pH 梯度。通过这种方式生成的 IPG, 不会发生电渗透作用, 因而可以进行特别稳定的 IEF 分离, 达到真正的平衡状态。最初的 IPG 技术在 2-DE 分离中的应用遇到了一些问题 (Görg *et al.*, 1988), 幸运的是, 这些问题现在已经解决, 并使 IPG IEF 成为 2-DE 中一维所选择的分离方法 (Görg *et al.*, 2000)。

7.1 IPG 凝胶制备

Görg 等 (1988) 的方法如图 3.1 所示。首先, 根据文献中广泛的配方库 (Görg *et al.*, 2000), 将所需 pH 范围的 IPG 平板胶在 GelBond PAGfilm (图 3.1b) 上聚合, 之后, 凝胶用去离子水全面冲洗 ($6 \times 10\text{min}$), 浸入在 2% (m/V) 甘油中 (30min), 室温下在洁净橱中干燥, 然后用塑料膜覆盖, 如果凝胶不立即使用, 可在 -20°C 贮存一年。用前, 用切纸机从 IPG 平板胶切下所需数目的干胶条 (图 3.1c)。此外, 各种 pH 范围商品化的预制 IPG 胶条可从 Amersham Pharmacia Biotech (Immobiline SryPlate 或 Immobiline DryStrip) 和 Bio-Rad (ReadyStrip) 购得。可用不同长度的 IPG 胶条, 但要切记的是, 通常分离区域越大, 能分辨的蛋白质越多。18 或 24cm 胶条 (Görg *et al.*, 1999) 通常用于高分辨分离, 而短胶条 (4cm, 7cm 或 11cm) 则应用于快速筛选。有大量不同宽或窄 pH 梯度范围条件可供选用于 IPG IEF, 这将在 7.4 中讨论。

7.2 IPG 胶条重泡胀

用于 2-DE 时, 胶条应该置于重泡胀池 (图 3.1d) 中在含 8mol/L 尿素的溶液 (或 2mol/L 硫脲和 5mol/L 尿素) 中泡胀。溶液其他成分为 0.5% 到 4% 非离子 (NP-40, Triton X-100) 或两性 (CHAPS) 去污剂、15mmol/L DTT 和 0.5% 适宜范围的 pH 值的 SCA。然后将胶条直接放在水平平板电泳装置的冷却板表面 (如 Multiphor, Amersham Pharmacia Biotech) (图 3.1g)。另外一种便利的方法是用 Pharmacia 公司的泡胀托架 (图 3.1h), 这种托架用波纹塑料板塑造, 上面所含的凹槽正好适于 IPG 的排列摆放。此外, 该设备配备有带有电极条和框架 (bar) 以及配备加样杯的支架, 可以在胶条表面任意位置加样。必要时, 可以将托架用硅油覆盖, 以在 IEF 时保护胶条使之不受大气影响。在 2-DE 图谱碱性端常可以见到水平条纹, 特别是当在一向使用碱性 IPG 梯度时, 在 IPG 胶表面沿阴极区额外叠加一个用 20mmol/L DTT 浸泡过的电极条, 可以解决这个问题 (Görg *et al.*, 1995)。该方法的优点是, IEF 时胶内 DTT 向阳极迁移, 从阴极区外加电极条释放出来的 DTT 可以不断进行补充。另外一种避免阴极区条纹的方法是利用不带电荷的还原剂 TBP, 它在 IEF 时不发生迁移, 已发现在 IEF 中 TBP 可以大大提高蛋白质的溶解能力, 特别是羊毛蛋白 (Herbert *et al.*, 1998; Herbert, 1999)。

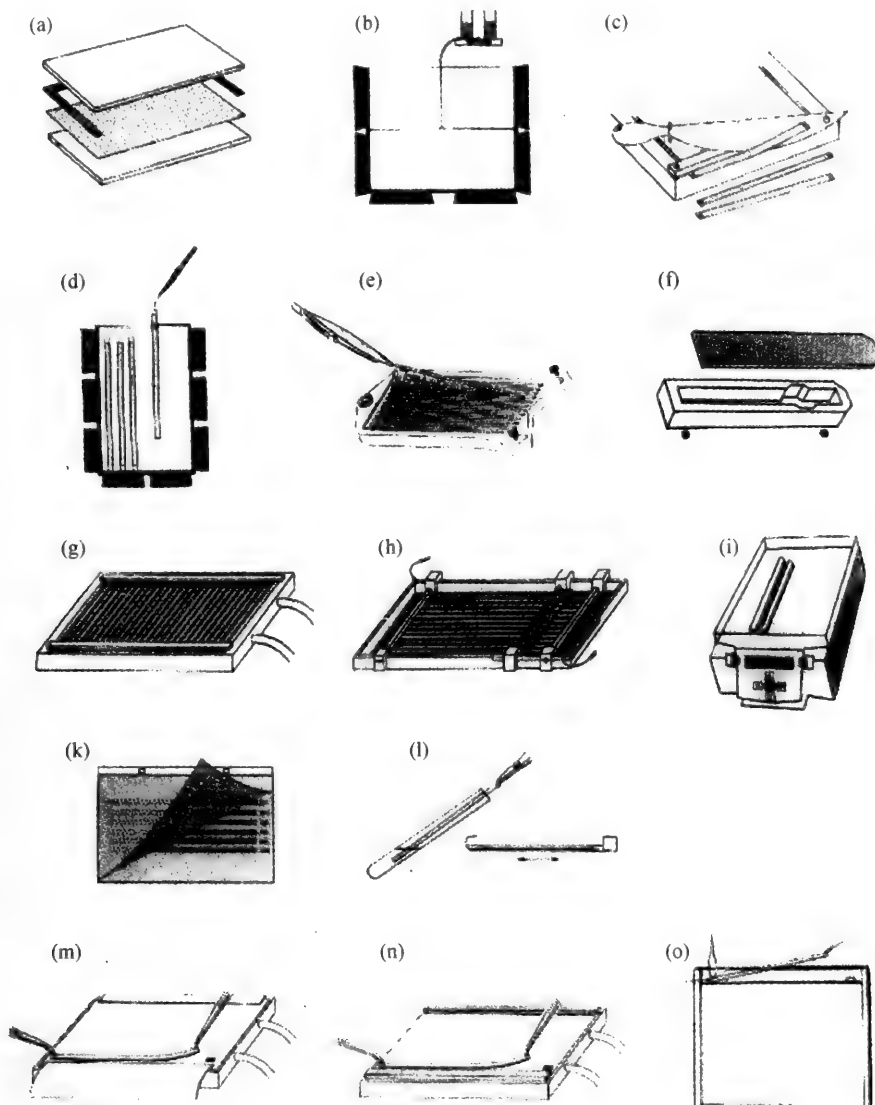


图 3.1 IPG-Dalt 的流程 (Görg *et al.*, 2000)

(a) 在塑料支持膜上聚合 IPG 及 SDS 胶之前组装聚合槽 (玻璃板, GelBond PAGfilm 支持膜, 0.5mm 厚的 U 形框)。(b) IPG 和 (或) SDS 梯度多孔胶的聚合。(c) 清洗并干燥后的 IPG 平板胶 (或 Immobiline DryPlates) 切成单个的 IPG 胶条, 单个 IPG 胶条的重泡胀。(d) 在垂直重泡胀槽内泡胀。(e) 在重泡胀托盘中泡胀, 以及 (f) 在 IPGphor 胶条泡胀池中泡胀。单个 IPG 胶条的 IEF。(g) 直接放置在 IEF 配件的冷却板上, (h) 在 DryStrip 盒中, 以及 (i) 在 IPGphor 上。(k) IEF 后 IPG 胶条的贮存。(l) SDS-PAGE 前 IPG 胶条的平衡。平衡后的 IPG 胶条转移 (m) 沿阴极电极条放置在自制的水平 SDS 胶表面 (n) 或沿阴极缓冲条放置在预制 SDS 胶表面。(o) 平衡过的 IPG 胶条放置在垂直 SDS 胶上。

7.3 加样和运行

样品通常是加在放置在 IPG 胶阳极区或阴极区的硅橡胶框内或特殊加样杯内 (图

3.1g, h), 对不同类型的样品, 最好的加样位置由经验值确定。最初电压应限制在 150V、30min, 从而使尽量多的样品进入加样杯, 然后逐渐增加电压至 3500 V。运行时间决定于几个不同的因素, 包括样品类型、蛋白上样量、IPG 胶条长度及所用 pH 梯度。因为温度低时尿素可能会结晶, IEF 应当在 20℃ 运行, 不同的温度下在 2-DE 图谱中一些蛋白质点的相对位置会发生变化 (Görg *et al.*, 1991)。表 3.1 中列出了一些典型的运行条件。

表 3.1 利用 Multiphor 的运行条件 (Görg *et al.*, 2000)

胶条长度	180mm	
温度	20℃	
最大电流	每个胶条 0.05mA	
最大功率	每个胶条 0.2W	
最大电压	3500V	
I 分析型 IEF		
初始化 IEF		
加样杯上样 (20~50 μ l)	胶内泡胀 (350 μ l)	
150V, 1h	150V, 1h	
300V, 1~3h	300V, 1~3h	
600V, 1h		
到达 3500V 稳定态后 IEF		
1~1.5 个 pH 单位	4 个 pH 单位	7 个 pH 单位
如 IPG5~6 24h	IPG4~8 10h	IPG3~10L 6h
如 IPG4~5.5 20h	IPG6~10 10h	IPG3~10NL 6h
3 个 pH 单位	5~6 个 pH 单位	8~9 个 pH 单位
IPG4~7 12h	IPG4~9 8h	IPG3~12 6h
IPG6~9 12h	IPG6~12 8h	IPG4~12 8h
II 加长分离距离 (240mm)		
到达 3500V 稳定态后 IEF		
IPG3~12 8h		
IPG4~12 12h		
IPG5~6 40h		
III 微量制备 IEF		
初始化 IEF		
加样杯上样 (100 μ l)	胶内泡胀 (350 μ l)	
50V, 12~16h	50V, 12~16h	
300V, 1h	300V, 1h	
到达 3500V 稳定态后 IEF		
相应分析胶 IEF 时间延长约 50%		

上面的加样方法会导致蛋白质沉淀, 上样量较大时 (1mg 或更多) 更为显著。将 IPG 胶条直接在含所分析样品泡胀液中泡胀可以克服这个问题 (Rabilloud *et al.*, 1994; Sanchez *et al.*, 1997)。利用该方法可以成功分离很高样品量 (>10mg), 但可能存在高分子量蛋白、极碱性和膜蛋白的选择性丢失。最近研制成功的集成化设备 IPGphor (Amersham Pharmacia Biotech), 简化了一向 IPG IEF。该设备的特点是在加样品或不加样品时在持胶器 (strip holder) 中对胶条进行重泡胀、选择上样量以及随后的 IEF,

所有这些都可以在胶条放入容器后设置程序自动进行。仪器可以同时放置 12 个持胶器，并配备珀尔帖 (Peltier) 电子晶体管自动冷却装置、高达 8000V 电压的程序化设置和 1.5mA 的电源供应 (图 3.1i)，表 3.2 列出一些利用 IPGphor 的典型的 IPG IEF 运行条件。在重泡胀时加低电压可以提高高分子质量蛋白质的渗入效率。适用于胶内泡胀上样和运行一向 IPG IEF 的设备也可从 Genomic Solutions (IPGpHaser) 和 Bio-Rad (Protean IEF cell) 获得。

表 3.2 利用 IPGphor 的运行条件 (Görg *et al.*, 2000)

胶条长度	180mm		
温度	20℃		
最大电流	每个胶条 0.05mA		
最大电压	8000V		
I 分析型 IEF			
重泡胀	• 30V, 12~16h		
初始化 IEF	• 200V, 1h		
	• 500V, 1h		
	• 1000V, 1h		
达到稳定态后 IEF	• 30min 内由 1000V 梯度上升至 8000V		
	• 8000V 达到稳定态, 依所用胶条 pH 间隔而定		
1~1.5 个 pH 单位	4 个 pH 单位	7 个 pH 单位	
如 IPG5~6 8h	IPG4~8 4h	IPG3~10L	3h
如 IPG4~5.5 8h		IPG3~10NL	3h
3 个 pH 单位	5~6 个 pH 单位	8~9 个 pH 单位	
IPG4~7 4h	IPG4~9 4h	IPG3~12	3h
		IPG4~12	3h
II 微量制备 IEF			
重泡胀	• 30V, 12~16h		
达到稳定态后 IEF	• 相应分析胶 IEF 时间延长约 50%		

7.4 IPG IEF pH 梯度的选择

对一新样品, 常最先使用宽范围、线性 pH3.5~10 梯度。但对大多数样品, 这样做会丧失 pH4~7 区域的分辨率, 许多蛋白质的 pI 值在该区域。利用非线性 pH3.5~10 IPG IEF 胶在一定程度上缓解了这个问题 (Bjellqvist *et al.*, 1993)。pH3.5~10 非线性胶条在 pH4~7 区域的梯度要比 pH7~10 更为平坦, 保证在大部分碱性蛋白质都能得到分辨的前提下, pH4~7 区域能得到更好的分离 (图 3.2)。但用 pH4~7 的 IPG IEF 胶则能在该范围得到更好的分离效果 (图 3.3)。这些 pH 范围的胶条目前已经商品化 (Amersham Pharmacia Biotech, Bio-Rad)。实验室自制 pH4~9 的 IPG 胶也能覆盖来自许多不同种类样品的大多数蛋白质。这些梯度的 IPG 胶与加样杯上样或胶内泡胀兼容, 在 Multiphor 和 IPGphor 系统都工作良好。均适用于分析胶 (上样量 50~100 μ g) 或微量制备胶 (上样量至 1mg)。典型运行条件如表 3.1 和 3.2 所示。

当分析真核细胞提取物一类复杂的样品时, 由于空间分辨率所限, 而且高丰度蛋白质存在时低拷贝数蛋白质显色困难, 使得单一宽范围 pH 梯度 IEF 只能显示该细胞提取物全蛋白质组的一小部分。可以克服真核及组织蛋白质表达高动态范围和多样性问题

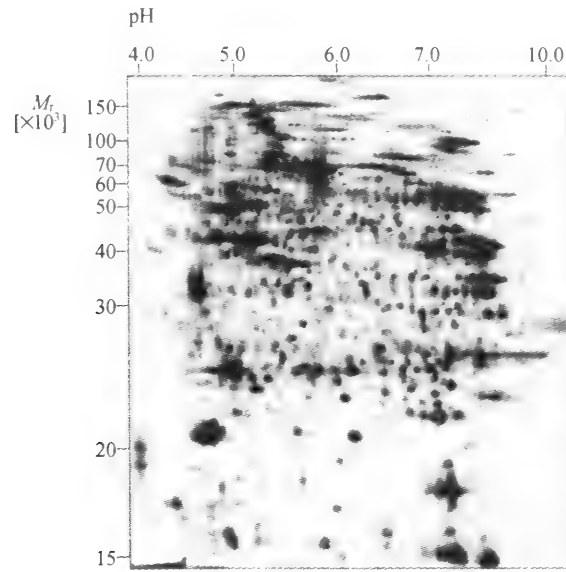


图 3.2 80 μ g 心室蛋白质的 2-DE 分离
一维应用 18cm 非线性 pH3--10 IPG IEF 凝胶, 二维为 21cm 12% SDS-PAGE 胶。银染检测。顶端标尺显示一维 IPG 胶条的非线性 pH 范围, 由此可以估计出分离蛋白的表观等电点。Mr 标尺用以估计蛋白分子质量。

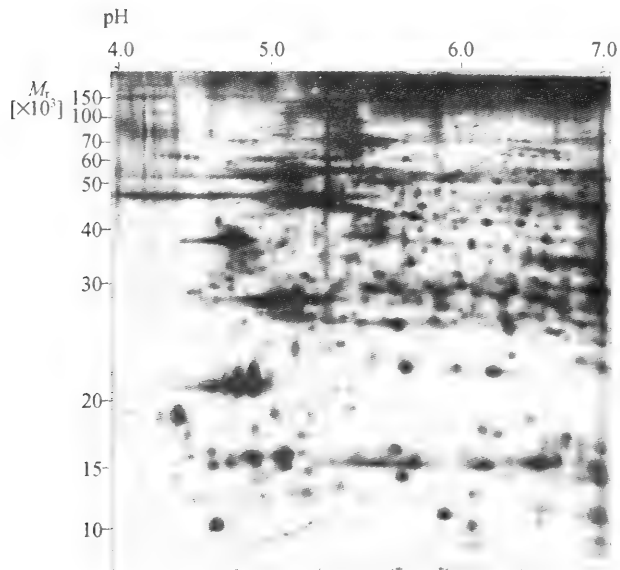


图 3.3 80 μ g 心室蛋白质的 2-DE 分离
一维为 18cm 线性 pH4~7 IPG IEF 凝胶, 二维为 21cm 12% SDS-PAGE 胶。银染检测。顶端标尺显示一向 IPG 胶条的线性 pH 范围, 由此可以估计出分离蛋白的表观等电点。Mr 标尺用以估计蛋白分子质量。

一种方法是进行样品前期分级（见 4.4）。另外一种有效的方案是利用复合重叠窄范围 IPG 凝胶，跨度在 1~1.5pH 单位，该方法被称为“放大胶”（zoom gel）（Wildgruber *et al.*, 2000）、复合胶或亚蛋白质组学（subproteomics）（Cordwell *et al.*, 2000）。此外，也可用 24cm 的宽胶条（Görg *et al.*, 1999）。pH4~7 之间的窄范围胶可用胶内泡胀或加样杯上样，在 Multiphor 或 IPGphor 仪器上运行（图 3.4）。这些凝胶是用于几个毫克上样量微量制备的理想选择，但制备胶需要加长聚焦时间（见表 3.1 和表 3.2）。

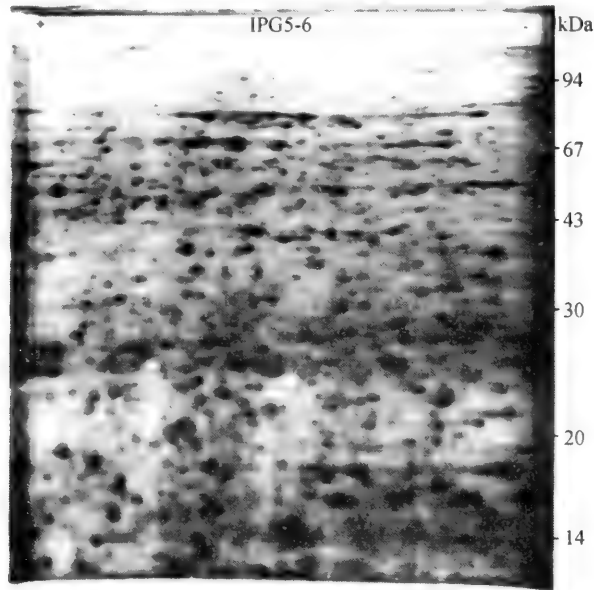


图 3.4 鼠肝蛋白质的微量制备 IPG-Dalt

一维：IPG 5~6（预制胶）。分离距离 18cm。胶内泡胀法上样（500 μ g）。IEF 在 IPGphor 上运行。条件：8000V 最大电压下 12h（详见表 1）。二维：垂直 SDS-PAGE（13% T 连续）。银染。

强碱性蛋白质如核糖体和核蛋白的 pI 值在 10.5 到 11.8 之间，可以用 pH10~12 或 pH9~12 窄范围 IPG（图 3.5；Görg *et al.*, 1997）分离，要获得高重复性的 2-DE 图谱，必须对 pH 梯度生成和凝胶组成进行不同的优化，如用 N, N'-二甲基丙烯酰胺取代丙烯酰胺，并在 IPG 泡胀液中加入异丙醇，从而抑制容易导致大量胶上条纹的反向电渗流（Görg *et al.*, 1997）。必须要用加样杯在阴极上样，且 IEF 要在硅油下进行。

从基因组序列（Link *et al.*, 1997）计算获得的理论 2-DE 图谱显示全细胞裂解所获得的大部分蛋白质等电点在 pH4~9，但同时相当一部分蛋白质等电点可以达到 pH12。在经典 2-DE 中这些蛋白质只能用 NEPHGE 分离（见 3），但用 IPG IEF 很容易将这些蛋白质分开。具体作法是用上文所提到的窄范围 pH10~12 或 pH9~12 的 IPG，但为了要获得某个细胞或组织蛋白质组的全貌，要先制备覆盖 pH3~12 和 pH4~12 的宽范围 IPG（Görg *et al.*, 1998, 1999）。目前已获得细胞和组织提取及 TCA/丙酮沉淀蛋白质时通常在裂解缓冲液提取中缺失的 pI 值大于 10 的碱性蛋白质（Görg *et al.*, 1998, 1999）以及肺支原体中不溶于 Triton X-100 的细胞成分（Hermann *et al.*, 已投

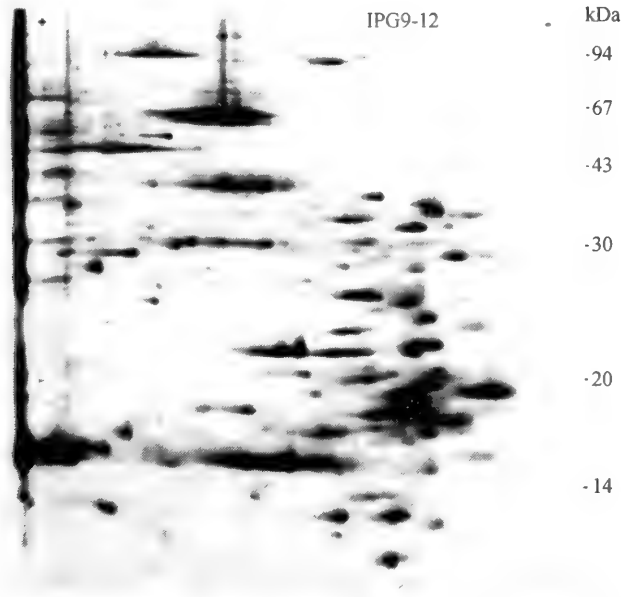


图 3.5 鼠肝极碱性蛋白 (TCA/丙酮沉淀) 的分析型
IPG-DALT (Görg *et al.*, 1997)

一维: IPG9-12 (5000V 最大电压 8h)。分离距离 18cm。在阳极端
用加样杯上样。二维: 垂直 SDS-PAGE (13% T 连续)。银染。

稿) 的良好的 2-DE 图谱。此外, 在 pH9 和 pH12 间含一个平缓区的 pH4~12 IPG 可以对如核糖体蛋白质类强碱性蛋白质获得极好分离 (Görg *et al.*, 1998)。这些宽梯度可以在不含异丙醇的标准条件下运行, 因为此时超过 pH10 的窄 IPG 胶中水合离子向阴极的强烈转移 (反向电渗) 特性可以忽略不计。此外, 高分子质量蛋白质的渗入和聚焦也得到了明显改善 (图 3.6)。

利用经典 O'Farrell 系统可以在宽分离距离 (每个方向上超过 30cm) 上获得复杂蛋白质谱的最大分辨率 (Klose, 1999)。但是, 若用 IEF 管胶技术, 其成功则伴随着凝胶操作和处理过程中方便性的丧失。与此对比, 在塑料支持膜上聚合的长 IPG 胶的处理不需要特殊注意之处。如覆盖 pH3~12 和 pH4~12 范围的 24cm IPG 胶条已成功使用 (图 3.6; Görg *et al.*, 1999)。此外, 利用窄范围 pH 值 (如 pH5~6) 的 24cm 长 IPG 胶条可以获得高分辨 2-DE 图谱 (图 3.7)。

7.5 聚焦时间的优化

理论上讲, 要获得最好的图谱质量和重复性所需最佳时间是 IEF 分离达到稳定态所需的时间 (Görg *et al.*, 1988, 1997)。若聚焦时间太短, 会导致水平和垂直条纹, 但要避免过度聚焦。虽然与经典 O'Farrell 法相比, 不会导致蛋白质向阴极的迁移 (阴极漂移), 但却会因为活性水转运而导致过多水在 IPG 胶表面渗出 (电渗), 会造成蛋白图谱变形。在胶条碱性端产生水平条纹以及蛋白质丢失。最佳聚焦时间必须根据不同蛋白质样品、蛋白上样量和所用特定 pH 范围及 IPG 胶条长度通过经验来确定。作为指

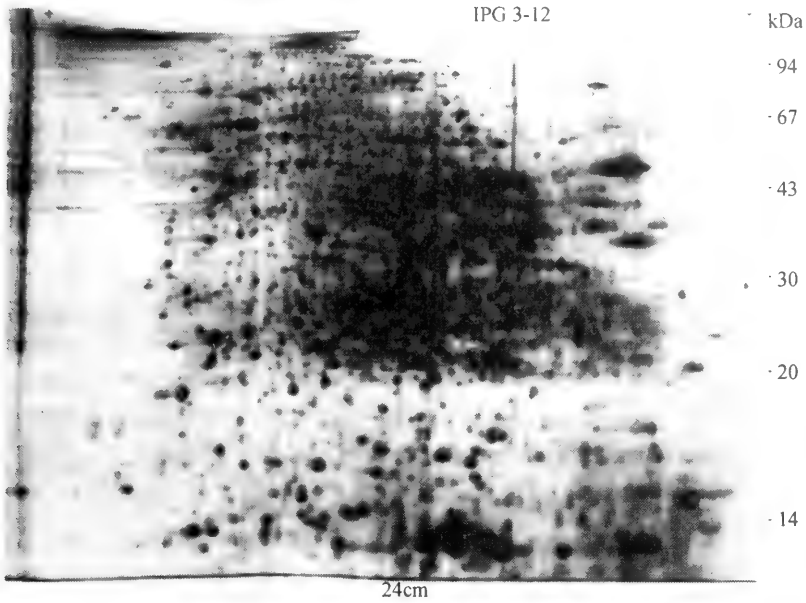


图 3.6 Görg 等 (1999) 加长距离的宽范围 IPG 鼠肝蛋白质的分析型 IPG-Dalt (裂解缓冲液提取)。一维: IPG3~12。分离距离 24cm。阳极端加样杯上样。运行条件: 3500V 最大电压 6h (见表 1)。二维: 垂直 SDS-PAGE (13% T 连续胶)。银染。

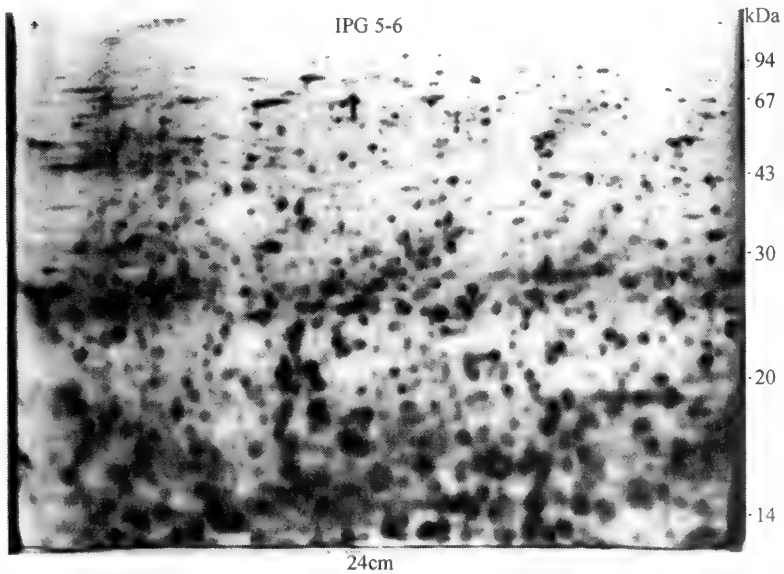


图 3.7 加长距离的窄范围 IPG 根据 Görg 等 (2000) 的鼠肝蛋白质 (裂解液提取) 的分析型 IPG-Dalt。一维: IPG5~6 (实验室自制)。分离距离 24cm。阳极端加样杯上样。运行条件: 3500V 最大电压 40h (见表 1)。二维: 垂直 SDS-PAGE (13% T 连续胶)。银染。

导, 表 3.1 和表 3.2 给出了针对一系列不同宽和窄 pH 范围 IPG 的优化聚焦时间。

8. 两维间平衡

在一维 IPG IEF 后, 胶条可马上用于二维, 也可保存在两片塑料膜间于 -80°C 存放几个月时间。在二维分离前, 必须要平衡 IEF 胶, 以便于被分离蛋白质与 SDS 完整结合, 从而在 SDS-PAGE 过程中正常迁移 (图 3.11)。建议方案是将 IPG IEF 胶在 50mmol/L Tris 缓冲液, $\text{pH}8.8$, 含 $2\% m/V$ SDS, $1\% m/V$ DTT, 6mol/L 尿素和 30% 甘油中孵育 15min , 尿素和甘油是用于减缓电渗效应, 否则会使一维到二维的蛋白质转移率降低 (Görg *et al.*, 1988)。接下来同样是在缓冲液中平衡 15min , 只不过此时用 $5\% m/V$ 碘乙酰胺取代 DTT, 这一步是用来烷基化自由 DTT, 否则自由 DTT 在二维 SDS-PAGE 胶迁移, 会产生点条纹的假象, 在银染后可以观察出来。另外一种方法, 可以一步完成平衡, 是在平衡液中用 5mmol/L TBP 取代 DTT, TBP 不带电荷, 在 SDS-PAGE 过程中不迁移 (Herbert *et al.*, 1998)。平衡后上二维胶之前, IPG 胶应沿边缘用滤纸吸干 1min 以去除多余液体。

9. 二维 SDS-PAGE

2-DE 第二维最常用的是 Laemmli (1970) 非连续缓冲体系, 二维胶是单一浓度或含线性或非线性梯度的凝胶, 其覆盖范围可以使不同分子质量大小蛋白质得到有效分离。由于 IPG 胶黏合在弹性塑料支持膜上, 其处理要比用经典 O'Farrell 体系中脆性 IEF 管胶容易。平衡后, IPG 胶条可直接加在二维 SDS-PAGE 胶的表面, 其方式可以是垂直 (图 3.1m, n) 或水平 (图 3.1o)。

水平 SDS-PAGE 可利用预制胶 (ExcelGel, Amersham Pharmacia Biotech) 或依据 Görg 等 (1988) 所描述的超薄孔状梯度胶在 GelBond PAGfilm (Amersham Pharmacia Biotech) 上自制平板胶上运行。必须要用 $6\% T$ 的浓缩胶, 然后是均一 (如 12%) 或梯度 (如 $12\% \sim 15\%$) 分离胶, 该方法在 Görg 等 (2000) 中有详细描述。水平二维系统的优点是凝胶附着在塑料支持膜上, 在染色过程中可以防止凝胶大小发生变化。此外, 由于水平胶厚度减少 (一般在 0.5mm , 而垂直胶为 1 或 1.5mm), 因而可以施加较高电压, 减短运行时间, 减少蛋白质扩散, 使得水平胶分离蛋白质点的边缘要比垂直胶清晰。

大规模蛋白质组分析通常需要成批二维 SDS-PAGE 同时电泳以获得最好的图谱重复性。此外, 凝胶格式需与计算机控制的机器人切割设备匹配, 用多块胶垂直二维 SDS-PAGE 设备如 DALT 系统 (Amersham Pharmacia Biotech) (Anderson and Anderson, 1978a, b)、Investigator (Genomic Solutions) (Patton *et al.*, 1990) 或 Protean II xi (Bio-Rad) 可以满足大规模蛋白质组分析的需求。在垂直电泳系统中不需要浓缩胶, 因为在 IPG 胶条中蛋白质区已得到浓缩, 可以认为非限制性 IEF 胶 (低浓度聚丙烯酰胺胶) 充当了浓缩胶 (Dunn, 1993)。

10. 分辨率

2-DE 的分离能力取决于在两维上的长度, 通常认为与有效的分离面积成正比。利

用 18cm 长的 IPG IEF 与 20cm 长二维 SDS-PAGE 胶结合, 对于全细胞和组织裂解物所获得的蛋白混合物, 可以很轻易分离得到 2000 个点。利用小胶只能得到几百个蛋白质点, 但小胶运行时间较短, 可用于快速筛选。对特别复杂的样品, 要获得最大分辨率, 用特大胶可以从全细胞裂解液中分离得到 5000~10 000 个蛋白质点 (Klose, 1999)。但利用多块重叠窄范围放大 IPG 胶来生成复杂样品的复合 2-DE 蛋白图谱具有明显优势 (见 7.4)。

11. 重复性

直到最近重复性仍然是限制 2-DE 广泛应用的主要问题。利用经典 O' Farrell 管胶技术, 即使在同一实验室内要获得特定类型样品的重复性分离也很困难。在不同实验室之间所获得的 2-DE 图谱间比较更不可能。2-DE 的专用设备如 ISO-DALT 和 Investigator 系统可以在重复性控制条件下同时运行大量数目的电泳 (5~20 块) 缓解了这个问题。更为重要的是, 不同类型样品 (心肌、大麦、酵母) 实验室间研究明确证实利用 IPG IEF 所获得的 2-D 蛋白分离具有很高的空间位置和定量重复性 (Blomberg *et al.*, 1995; Corbett *et al.*, 1994)。

12. 展望

蛋白混合物的重复性高分辨分离是一个成功的蛋白质组学研究的关键。虽然最近对蛋白质组学中蛋白分离的可替代方法的研究有所进展 (见 1 和第十章), 但仍然没有实用的方法可替代 2-DE 能同时分离和显现混合物中几千个蛋白质。与基于 SCA 产生的 pH 梯度的经典 O' Farrell 2-DE 方法相比, IPG IEF 2-DE 更加灵活, 能满足蛋白质组分析的需求, 特别是高至 pH12 的碱性 IPG 方便了极碱性蛋白的分析。重叠窄范围 IPG, 允许低丰度蛋白的更高分辨率的分离分析, 预制 IPG 胶条的应用和集成化运行设备如 IPGphor 使 2-DE 向自动化方向前进了一步, 虽然 2-DE 的自动化是蛋白质组学中一个关键问题, 但在实际中要设计能完成整个过程的自动化系统是不可能的。事实上, 蛋白质组流程中一个主要的瓶颈效应是 2-D 凝胶图像的分析 and 数据开发, 在该领域的自动化和效率的提高可能会对蛋白质组项目产生最大的影响 (Lopez, 2000)。目前, 限制性因素依然是疏水蛋白及膜蛋白的分析、以及灵敏和可靠的蛋白质定量方法的缺乏, 即使最近在荧光染料技术的进展为解决这个问题指出了方向 (Patton, 2000)。总之, 虽然 2-DE 技术并不完美, 但在目前和相当长的一段时间内, 它将仍然是蛋白质组项目中分离蛋白混合物所要选择的核心理技术。

(应万涛译 钱小红校)

参考文献

- Anderson, N.G. and Anderson, N.L. (1978) Analytical techniques for cell fractions. XXI. Two-dimensional analysis of serum and tissue proteins: multiple isoelectric focusing. *Anal. Biochem.* **85**: 331–340.
- Anderson, N.G. and Anderson, N.L. (1978) Analytical techniques for cell fractions. XXII. Two-dimensional analysis of serum and tissue proteins: multiple gradient-slab gel electrophoresis. *Anal. Biochem.* **85**: 341–354.
- Banks, R.E., Dunn, M.J., Forbes, M.A., Stanley, A., Pappin, D., Naven, T., Gough, M., Harnden, P. and Selby, P.J. (1999) The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis – preliminary findings. *Electrophoresis* **20**: 689–700.
- Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Görg, A., Westermeier, R. and Postel, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J. Biochem. Biophys. Methods* **6**: 317–339.
- Bjellqvist, B., Sanchez, J.C., Pasquali, C., Ravier, F., Paquet, N., Frutiger, S., Hughes, G.J. and Hochstrasser, D.F. (1993) A nonlinear wide-range immobilised pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis* **14**: 1357–1365.
- Blomberg, A., Blomberg, L., Norbeck, J. et al. (1995) Interlaboratory reproducibility of yeast protein patterns analyzed by immobilized pH gradient two-dimensional gel electrophoresis. *Electrophoresis* **16**: 1935–1945.
- Corbett, J.M., Dunn, M.J., Posch, A. and Görg, A. (1994) Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: an interlaboratory comparison. *Electrophoresis* **15**: 1205–1211.
- Cordwell, S.J., Nouwens, A.S., Verrills, N.M., Basseal, D.J. and Walsh, B.J. (2000) Sub-proteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional gels. *Electrophoresis* **21**: 1094–1103.
- Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F. and Sanchez, J.C. (2000) The dynamic range of protein expression: A challenge for proteomic research. *Electrophoresis* **21**: 1104–1115.
- Dunn, M.J. (1987) Two-dimensional polyacrylamide gel electrophoresis. In: *Advances in Electrophoresis* (ed. Chrambach, A., Dunn, M.J. and Radola, B.J.), Vol. 1, VCH, Weinheim, pp. 1–109.
- Dunn, M.J. (1993) *Gel Electrophoresis: Proteins*. BIOS Scientific Publishers, Oxford.
- Garrels, J.I. (1979) Two dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. *J. Biol. Chem.* **254**: 7961–7977
- Garvik, B.M. and Yates, J.R. 3rd. (1999) Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**: 676–682.
- Görg, A., Postel, W. and Günther, S. (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**: 531–546.
- Görg, A., Postel, W., Friedrich, C., Kuick, R., Strahler, J.R. and Hanash, S.M. (1991) Temperature-dependent spot positional variability in two-dimensional polypeptide patterns. *Electrophoresis* **12**: 653–658.
- Görg, A., Boguth, G., Obermaier, C., Posch, A. and Weiss, W. (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): the state of the art and the controversy of vertical versus horizontal systems. *Electrophoresis* **16**: 1079–1086.
- Görg, A., Obermaier, C., Boguth, G., Csordas, A., Diaz, J.J. and Madjar, J.J. (1997) Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. *Electrophoresis* **18**: 328–337.
- Görg, A., Boguth, G., Obermaier, C. and Weiss, W. (1998) Two-dimensional electrophoresis of proteins in an immobilized pH 4–12 gradient. *Electrophoresis* **19**: 1516–1519.
- Görg, A., Obermaier, C., Boguth, G. and Weiss, W. (1999) Recent developments in two-dimensional gel electrophoresis with immobilized pH gradients: wide pH gradients up to pH 12, longer separation distances and simplified procedures. *Electrophoresis* **20**: 712–717.
- Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R. and Weiss, W. (2000)

- The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**: 1037–1053.
- Ygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R.** (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**: 994–999.
- Herbert, B., Molloy, M.P., Gooley, A.A., Walsh, B.J., Bryson, W.G. and Williams, K.L.** (1998) Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as a reducing agent. *Electrophoresis* **19**: 845–851.
- Herbert, B.** (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**: 660–663.
- Iborra, F., Buhler, J.M.** (1976) Protein subunit mapping. A sensitive high-resolution method. *Anal. Biochem.* **74**: 503–511.
- Klose, J.** (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**: 231–243.
- Klose, J.** (1999) Genotypes and phenotypes. *Electrophoresis* **20**: 643–652.
- Laemmli, U.K.** (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **22**: 680–685.
- Link, A.J., Robison, K. and Church, G.M.** (1997) Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *Electrophoresis* **18**: 1259–1313.
- Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R. 3rd.** (1999) Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**: 676–682.
- Lollo, B.A., Harvey, S., Liao, J., Stevens, A.C., Wagenknecht, R., Sayen, R., Whaley, J. and Sajjadi, F.G.** (1999) Improved two-dimensional gel electrophoresis representation of serum proteins by using ProtoClear. *Electrophoresis* **20**: 854–859.
- Lopez, M.F.** (2000) Better approaches to finding the needle in a haystack: Optimizing proteome analysis through automation. *Electrophoresis* **21**: 1082–1093.
- Merchant, M. and Weinberger, S.R.** (2000) Recent advances in surface enhanced laser-desorption/ionization time-of-flight mass spectrometry. *Electrophoresis* **21**: 1165–1177.
- Molloy, M.P., Herbert, B., Walsh, B.J., Tyler, M.I., Traini, M., Sanchez, J.C., Hochstrasser, D.F., Williams, K.L. and Gooley, A.A.** (1998) Extraction of membrane proteins by differential solubilization for separation using two-dimensional electrophoresis. *Electrophoresis* **19**: 837–844.
- Monribot, C. and Boucherie, H.** (2000) Two-dimensional electrophoresis with carrier ampholytes. In: *Proteome Research: Two-Dimensional Gel Electrophoresis and Identification Methods* (ed. Rabilloud, T.), Springer, Berlin, pp. 31–55.
- Nelson, R.W., Nedelkov, D. and Tubbs, K.A.** (2000) Biosensor chip mass spectrometry: a chip-based approach. *Electrophoresis* **21**: 1155–1163.
- O'Farrell, P.H.** (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007–4021.
- O'Farrell, P.Z., Goodman, H.M. and O'Farrell, P.H.** (1977) High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell* **12**: 1133–1141.
- Patton, W.F., Pluskal, M.G., Skea, W.M., Buecker, J.L., Lopez, M.F., Zimmermann, R., Belanger, L.M. and Hatch, P.D.** (1990) Development of a dedicated two-dimensional gel electrophoresis system that provides optimal pattern reproducibility and polypeptide resolution. *BioTechniques* **8**: 518–527.
- Patton, W.F.** (2000) A thousand points of light: The application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* **21**: 1123–1144.
- Perdew, G.H., Schaup, H.W. and Selivonchick, D.P.** (1983) The use of a zwitterionic detergent in two-dimensional gel electrophoresis of trout liver microsomes. *Anal. Biochem.* **135**: 453–455.
- Rabilloud, T., Hubert, M. and Tarroux, P.** (1986) Procedures for two-dimensional electrophoretic analysis of nuclear proteins. *J. Chromatogr.* **351**: 77–89.
- Rabilloud, T., Valette, C. and Lawrence, J.J.** (1994) Sample application by in-gel rehydration

- improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis* **15**: 1552–1558.
- Rabilloud, T., Adessi, C., Giraudel, A. and Lunardi, J.** (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**: 307–316.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B.** (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**: 1030–1032.
- Sanchez, J.C., Rouge, V., Pisteur, M., Ravier, F., Tonella, L., Moosmayer, M., Wilkins, M.R. and Hochstrasser, D.F.** (1997) Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients. *Electrophoresis* **18**: 324–327.
- Santoni, V., Molloy, M. and Rabilloud, T.** (2000) Membrane proteins and proteomics: Un amour impossible? *Electrophoresis* **21**: 1054–1070.
- Scheele, G.A.** (1975) Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins. *J. Biol. Chem.* **250**: 5375–5385.
- Simone, N.L., Bonner, R.F., Gillespie, J.W., Emmert-Buck, M.R. and Liotta, L.A.** (1998) Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genetics* **14**: 272–276.
- Smithies, O. and Poulik, M.D.** (1956) Two-dimensional electrophoresis of serum proteins. *Nature* **177**: 1033.
- Uetz, P., Giot, L., Cagney, G. et al.** (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**: 623–627.
- Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fet, S.J., Larsen, P.M. and Görg, A.** (2000) Towards higher resolution: 2D-Electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow IPGs. *Electrophoresis* **21**: (in press).

第四章 聚丙烯酰胺凝胶和转印膜上蛋白质的检测

Wayne F. Patton

1. 引言

本章主要概括当前应用于凝胶电泳和电转印的全蛋白质呈现技术，特别是这些技术在蛋白质组学研究中的适用性，蛋白质显色技术发展简史如表 4.1。已有许多不同的关于胶内和膜上蛋白质显色的精彩的综述性文章和比较学研究，有兴趣的读者也可以参阅文献 (Brush, 1998; Dunn, 1999; Fernandez-Patron *et al.*, 1998; Fowler, 1994; Li *et al.*, 1989; Merril, 1987; Neuhoff *et al.*, 1985; Patton *et al.*, 1999a, b; Rabilloud, 1990; Rabilloud *et al.*, 1994; Wirth and Romano, 1995)。

表 4.1 凝胶电泳和转膜蛋白质显色试剂 50 年简史

年代	开发者	染料类型	大事记
1950	Durrum	有机染料	在滤纸上用溴酚蓝染色
1952	Grassman and Hannig	有机染料	在滤纸上用氨基黑染色
1963	Fazakas de St. Groth <i>et al.</i>	有机染料	醋酸纤维素膜、琼脂糖或淀粉胶的考马斯蓝 R-250 染色
1965	Meyer and Lamberts	有机染料	聚丙烯酰胺凝胶 (PAGE) 的考马斯蓝 R-250 染色
1967	Chrumbach <i>et al.</i>	有机染料	PAGE 胶用考马斯蓝 R-250 在三氯醋酸溶液中无背景染色
1969	Hartman and Udenfriend	有机荧光团	利用苯胺萘磺酸盐对蛋白质疏水位点的非共价染色
1971	Talgot and Yphantis	有机荧光团	凝胶电泳前用丹磺酰氯共价衍生蛋白质
1972	Diezel <i>et al.</i>	有机染料	三氯醋酸溶液中考马斯蓝 R-250 改善后凝胶无背景染色。 注意到染料转化为胶体
1972	Kerenyi and Gallyas	银	琼脂糖胶中蛋白质基于铁氰化钾的组织化学银染
1974	Wallace <i>et al.</i>	负染	SDS 胶冷却以沉淀去污剂、净化蛋白区
1978	Graham <i>et al.</i>	金属螯合盐	凝胶的亚铁红非绕啉二磺酸盐染色
1979	Switzer <i>et al.</i>	银	利用连续碱性和酸性银染步骤对 PAGE 胶染色
1980	Hager and Burgess	负染	用 KCl 进行凝胶负染
1980	Oakley <i>et al.</i>	银	基于组织化学的 PAGE 胶银染 (碱性/银二氨法)
1981	Merril <i>et al.</i>	银	不含二氨、基于化学的 PAGE 胶银染 (酸性/硝酸银法)
1982	Zapolski <i>et al.</i>	金属螯合盐	凝胶的放射性、亚铁红非绕啉二磺酸盐染色
1983	Hancock and Tsang	胶体扩散	转膜蛋白质的印度墨水染色 (胶体金染色的先驱)
1985	Neuhoff <i>et al.</i>	有机染料	证实胶体考马斯蓝的胶体属性, 确定多种酸/乙醇混合物的 化学式
1985	Rohringer and Holden	胶体扩散	转膜蛋白质的胶体金染色
1985	Moeremans <i>et al.</i>	胶体扩散	转膜蛋白质的高灵敏度胶体金、铁和银染
1987	Lee <i>et al.</i>	负染	胶内蛋白质的氯化铜负染
1988	Dzandu <i>et al.</i>	负染	胶内蛋白质的氯化锌负染
1988	Jackson <i>et al.</i>	有机荧光团	在等电聚焦胶内用 MDPF 共价衍生蛋白质, 进一步二向分 离
1991	Daban <i>et al.</i>	有机荧光团	利用尼罗红对 PAGE 胶内蛋白-SDS 复合物荧光染色
1992	Ortiz <i>et al.</i>	负染	高灵敏度氯化锌/咪唑负染

续表

年代	开发者	染料类型	大事记
1994	Patton <i>et al.</i>	金属螯合盐	比色法、锌铁-亚铁金属螯合盐染色 (硝酸纤维素, PVDF)
1996	Steinberg <i>et al.</i>	有机荧光团	利用 SYPRO Red 和 Orange 的蛋白-SDS 复合物高灵敏度荧光染色
1997	Lim <i>et al.</i>	荧光金属螯合盐	发光钨红菲绕啉二碘酸盐染色 (硝酸纤维素, PVDF, PAGE 胶)。SYPRO Rose 染料
1997	Unlu <i>et al.</i>	有机荧光团	两个蛋白样品用丙基-Cy3 和甲基-Cy5 染料分别衍生并在同一 2-DE 胶上同时显色
1999	Berggren <i>et al.</i>	荧光金属螯合盐	高灵敏度基于发光钨的金属螯合盐染料 (硝酸纤维素, PVDF, PAGE 胶)。SYPRO Ruby 染料

2. 有机染料和银染

用于聚丙烯酰胺凝胶电泳 (PAGE) 分离蛋白质常规检测和定量最普遍的方法是考马斯蓝染色和银染 (Brush, 1998; Merril, 1987; Wirth and Romano, 1995)。测序技术和质谱灵敏度的提高使蛋白质显色的地位从简单的显色试剂转换为集成化的蛋白质微量化学表征过程中的关键步骤。随着高灵敏度蛋白质分析方法和电泳后鉴定技术的结合, 考马斯蓝染色和银染用于蛋白质组研究的局限性日益暴露出来。

2.1 有机染料

自 20 世纪 60 年代早期传统考马斯蓝染色技术在甲醇和醋酸水溶液中用于 PAGE 胶染色后, 即得到广泛应用 (Fazekas de St. Groth *et al.*, 1963; Meyer and Lamberts, 1965)。至少已尝试了 600 种关于 PAGE 胶考染染色的线性和检测灵敏度的不同方法 (Neuhoff *et al.*, 1985)。考马斯蓝染料可以检测到 30~100ng 蛋白质, 但其灵敏度远低于银染或荧光检测 (Brush, 1998)。

胶体考马斯蓝应用于 PAGE 胶分离蛋白质的无背景检测可能是考染方法中最有意义的进展 (Chrumbach *et al.*, 1967; Diezel *et al.*, 1972; Neuhoff *et al.*, 1985, 1988)。酸性紫 17、Serva 紫 49 和坚牢绿 FCF 等相关染料也可在强酸性溶液中形成胶体, 检测 PAGE 胶内蛋白质, 且只产生很低的背景 (Neuhoff *et al.*, 1990; Patestos *et al.*, 1988)。估计其染色机制是因为在胶体染料和溶液中自由扩散的染料间形成平衡态, 少量溶液自由染料穿透凝胶基质, 有选择性地对蛋白质染色, 而胶体染料排阻在胶外, 阻止了背景生成 (Brush, 1998; Neuhoff *et al.*, 1988)。胶体考马斯蓝检测蛋白质的极限是 8~10ng (Brush, 1998)。染色过程通常是在高浓度的三氯醋酸、高氯酸或磷酸中进行, 并辅以甲醇或乙醇。质谱对蛋白质修饰的研究表明用含三氯醋酸和乙醇的考马斯蓝染色的蛋白质易导致谷氨酸羧基侧链的不可逆酸催化酯基化 (Haebel *et al.*, 1998)。这样就使所得肽谱数据复杂化, 但也可通过在分析软件中加入算法来考虑这种修饰。

氨基黑是最早应用于蛋白质电泳后显色的有机染料 (Grassman and Hannig, 1952)。现在常用于中度灵敏度聚偏二氟乙烯 (PVDF) 和硝酸纤维素膜的电转印蛋白质比色检测 (Dunn, 1999)。虽然氨基黑和考马斯蓝都可用于显色电转印蛋白质, 但考马斯蓝特

异性地染色硝酸纤维素膜,因而会产生高背景,但是,改进后的硝酸纤维素膜上蛋白质低背景染色程序可能会使考马斯蓝染料得到有效的利用 (Metkar *et al.*, 1995)。两种染色方法都适用于 PVDF 膜上蛋白质检测,虽然它们与下游微量化学表征兼容,但常干扰随后的免疫检测和照像备份程序 (Eynard and Lauriere, 1998; Pryor *et al.*, 1992)。一个方法是在免疫检测前先脱色 (Pryor *et al.*, 1992)。用结合碱性磷酸化酶的二级抗体和比色底物如坚牢红/萘酚 AS MX 磷酸盐或 5-溴-4-氯-吡啶磷酸盐/氮蓝四唑以及考马斯蓝染料检测全蛋白质可以同时显色特异性蛋白质 (Zeindl-Eberhart *et al.*, 1997)。该方法特点是双向电泳时需要大的上样量 (400 μ g), 必须用 PVDF 膜进行转印,而且所检测的目标蛋白质丰度要相当的高。目前,这种双色显色程序必须针对每一样品进行逐一条件摸索,以确定是否适用于目标蛋白质的检测。

2.2 银染

银染的根源可以追溯到 19 世纪初所开发的照相术及组织学技术 (Merril, 1987; Merrill *et al.*, 1984; Rabilloud, 1990)。Kerenyi 和 Gallyas (1972) 首先设计一种通用的针对电泳、免疫电泳和免疫扩散的蛋白质银染检测方案,其检测灵敏度报道比氨基黑高,但该方案只是优化用于琼脂糖而非 PAGE 胶。此法随后得到提高并在琼脂糖凝胶分离 DNA 和蛋白质中均显示良好的灵敏度 (Peats, 1984)。PAGE 胶分离蛋白质银染是在首次琼脂糖胶方法 7 年后引入的 (Switzer *et al.*, 1979)。据报道新方法灵敏度至少比考马斯蓝高 100 倍,由此使银染成为一项创新性的蛋白质检测技术,其动态线性范围扩展到 40 倍浓度。因此完全可以和有 20 倍线性范围的考马斯蓝染色对比 (Dunn, 1987)。银染技术的引入,使蛋白质检测的灵敏度由微克级升为纳克级 (除了早期胶体考马斯蓝高灵敏度染色)。

已报道 100 多种银染方法 (Rabilloud, 1992), 这个巨大数字也证明了一个事实,即没有一种方法能集便利、快速和高灵敏度为一体 (Rabilloud, 1992, 1999)。许多银染方法所拥有的一个缺点是对某些种类蛋白质如钙结合蛋白质和糖蛋白,染色效果相当差 (Jay *et al.*, 1990; Schleicher and Watterson, 1983; Wirth and Romano, 1995)。研究发现,对于糖蛋白来说,银染的强度与分子质量组成中糖类所占的比例负相关 (Jay *et al.*, 1990)。在银染前用阴离子染料对蛋白质预染色可以很好的改善糖蛋白的检测 (Dubray and Bezar, 1982; Jay *et al.*, 1990)。

银染虽可检测低纳克级的蛋白质,但银离子可与半胱氨酸残基相互作用,而且银染中用了戊二醛和甲醛,这两者会对蛋白质的 α 和 ϵ -氨基进行烷基化。这样妨碍了胶上蛋白质 Edman 蛋白质测序的进一步分析。虽然去除戊二醛后,蛋白质可用质谱分析来鉴定 (Christiansen and Houen, 1992; Shevchenko *et al.*, 1996; Wilm *et al.*, 1996), 但这种改变的程序也伴随着银染灵敏度和均一性的降低以及凝胶背景的升高。利用这种修改后的银染程序,肽质量指纹谱中的低序列覆盖率抵消了银染相对于考染和锌-咪唑染色的高检测灵敏度 (Scheler *et al.*, 1998)。估计银染蛋白质的低序列覆盖率是源于染色过程中甲醛对蛋白的修饰 (Scheler *et al.*, 1998), 但甲醛还不能从目前的银染方案中去

除。

两种用得最普遍的银染方式是碱性/二氨合银法和酸性/硝酸银法 (Rabilloud,

1999)。酸性/硝酸银法的基础是照相过程,将凝胶在酸性 pH 值下的银离子中浸泡,随后在碱性 pH 下用甲醛将银离子还原为金属银显色 (Wirth and Romano, 1995)。目前普遍认为酸性/硝酸银法与碱性/二氨银法相比,对碱性蛋白质的检测灵敏度稍低,而对酸性蛋白质的检测灵敏度较高 (Rabilloud, 1999; Rabilloud *et al.*, 1994)。酸性/硝酸银法也易产生异染色假象,但用机械韧性强的 PAGE 胶基质 Duracryl (Genomic Solutions, Ann Arbor, MI) 取代标准聚丙烯酰胺后可大大降低这种效果 (Patton *et al.*, 1992)。酸性/硝酸银法适于蛋白质定量,凝胶显色过程约限于 10 min (Patton, 1995)。

碱性/二氨银法基于组织学方法,利用氢氧化铵形成可溶性的二氨银复合物 (Wirth and Romano, 1995),随后在酸性显色液中用甲醛还原自由银离子使蛋白质显色。用哌嗪双丙烯酰胺取代 *N,N'*-亚甲基双丙烯酰胺后可优化碱性/二氨银法 (Hochstrasser *et al.*, 1988; Rabilloud, 1999)。碱性/二氨银法对 Tris/tricine 或 *N*-三(羟乙基)甘氨酸缓冲液体系凝胶,如商品化的长贮存期预制胶以及用于低分子质量蛋白质分离的凝胶,染色效果较差。用固相 pH 梯度胶分离蛋白质碱性/二氨银法染色效果也差 (Rabilloud, 1999)。此外,带有塑料支撑膜的 PAGE 胶常会产生银镜假象,使蛋白质染色效果大打折扣 (Rabilloud, 1999)。由于氨气的易挥发性,碱性/二氨银染色不同批次实验间重复性差 (Rabilloud, 1999)。非计量染色属性也使得该方法不太适于定量分析 (Rodriguez *et al.*, 1993)。

3. 负染

标准考染方法所存在的一个缺点是染色步骤中含蛋白质固定过程,银染法除了固定过程外,还有增敏步骤,这些步骤会将蛋白质提取至胶外,而减少蛋白质产量,使得用于随后微量化学表征的样品量更少了。负染的开发则是专门为了提高 PAGE 胶上蛋白质的回收率。这些方法不能用于对印渍膜染色,通常也适于进行准确的蛋白质定量。染色结果在凝胶表面产生半透明背景,其中蛋白质以透明的形式被检测出来。凝胶显示黑色背景或相应的底色 (Wirth and Romano, 1995)。染色过程很快,约需 5~15 min 即可完成,蛋白质的生物学活性经常得以保持。一旦用络合剂如 EDTA 或 Tris/甘氨酸转移缓冲液络合金属离子,就可进行提取来转移蛋白质 (Castellanos-Serra *et al.*, 1996; Lee *et al.*, 1987)。负染适用于蛋白质显色、完整蛋白质的胶上被动提取以及质谱分析 (Cohen and Chait, 1997)。

3.1 金属盐染料

目前已有一系列 SDS-PAGE 胶分离蛋白质显色的负染方法,包括低温 (Wallace *et al.*, 1974)、阴离子去污剂 (Takagi *et al.*, 1977)、乙酸钠 (Higgins and Dahmus, 1979)、KCl (Hager and Burgess, 1980)、银 (Merril *et al.*, 1984)、胶体金 (Casero *et al.*, 1985)、氯化铜 (Lee *et al.*, 1987)、氯化锌 (Dzandu *et al.*, 1988)、乙酸钴 (Dzandu *et al.*, 1988)、氯化镍 (Dzandu *et al.*, 1988)、锌-咪唑 (Ortiz *et al.*, 1992) 以及三氯乙酸甲酯法 (Candiano *et al.*, 1996)。在一个类似于锌-咪唑染色的专利方法中,平面染料如若丹明、吖啶橙、曙红或荧光黄与金属盐如重铬酸钾、铬酸钡、氯化镁或氯化钙形成复合物,从而达到使凝胶呈现可见色,蛋白质仍保持透明的高灵敏度负染

效果 (Berube, 1990)。

三种用得最多的金属盐负染法是氯化钾法、氯化铜法和氯化锌法。氯化铜染色灵敏度较高, 而氯化钾染色灵敏度远低于考马斯蓝 (Adams and Weaver, 1990)。氯化铜染色对细胞色素 C、组蛋白和其他碱性蛋白质灵敏度低于偏酸性蛋白质 (Vanflerteren and Peeters, 1990)。目前为止, 氯化锌法是最灵敏的负染方法 (Fernandez-Patron *et al.*, 1998)。

3.2 锌-咪唑染料

利用特定的锌-咪唑染色方法, 核酸、糖脂和蛋白质呈现出类似的染色行为 (Fernandez-Patron *et al.*, 1998)。在咪唑存在时, 自由或弱结合的锌离子以锌-咪唑形式很好地沉淀下来, 而大部分结合牢固的离子则难以沉淀下来, 这就导致在半透明背景上的空白区域, 也就是负染过程特点。一定要注意控制染色时间, 因为染色过度可能掩盖条带, 且偶尔还会出现负染和阳性染色的复合图谱 (Fernandez-Patron *et al.*, 1998)。锌-咪唑对 SDS-PAGE 分离蛋白质的检测限是每条带 10~20ng, SDS-PAGE 分离的脂多糖为每条带 1~10ng, 尿素变性 PAGE 胶分离小分子寡核苷酸 (28bp~1.3kbp) 为 10~100ng (Fernandez-Patron *et al.*, 1998)。但对非变性 PAGE 胶分离的牛血清白蛋白其检测限较差, 约每条带 40~80ng。

锌-咪唑染色的凝胶可以干燥而染色图谱并不发生变化 (Ferrerias *et al.*, 1993)。虽然蛋白质检测线性范围限制在 10~100ng, 仍可通过光密度仪来估计蛋白量 (Ferrerias *et al.*, 1993; Matsui *et al.*, 1997)。锌-咪唑染色的凝胶可通过在转移缓冲液中加入络合剂来进行电转印 (Fernandez-Patron *et al.*, 1995), 之后可进一步进行 Edman 测序 (Fernandez-Patron *et al.*, 1995)。锌-咪唑染色与胶内酶切、基质辅助激光解吸附-飞行时间质谱 (MALDI-TOF/MS) 测定肽质量和序列的方法兼容 (Matsui *et al.*, 1997)。

4. 胶体扩散染料

胶体扩散染料含有与蛋白质具有高亲和力的精细颗粒, 这种染料主要是用于高灵敏度检出电转印、点转印或狭缝转印至硝酸纤维素膜或 PVDF 膜上的蛋白质, 虽然有文献报道胶体金负染的应用 (Casero *et al.*, 1985), 但此类染料通常不用于 PAGE 胶内蛋白质的检出。此外, 可以在蛋白质混合物中, 加入表面接有有机外壳的胶体钯或铂颗粒, 进行 PAGE 电泳, 再用一个银增强步骤显色 (Powell, 1998)。对印渍膜来说, 最好的胶体扩散染料——胶体金染料, 其灵敏度与 PAGE 胶的银染类似。

4.1 印度墨水染料

稀释的自来水笔印度墨水是首次成功用于电转印蛋白质显色的胶体扩散染料 (Hancock and Tsang, 1983)。该方法的首要优点是简单便宜。自来水笔墨水是包含金属化的酸性染料 (磺酸盐、氯磺化盐、氨磺酰盐) 以及碱性染料盐的专用配方 (Rohde *et al.*, 1998)。此外, 也加入碳、酞菁酮、甘醇、甘油表面活性剂、抗氧化剂和黏度调节剂等精细分散剂以形成水溶性墨水 (Rohde *et al.*, 1998)。常用于电转膜染色的黑墨水 (Pelikan, Hannover, Germany) 经毛细管电泳分离, 并结合紫外/可见吸收和激光诱导

荧光检测表明,在该配方中至少有7种不同种类的染料(Rohde *et al.*, 1998)。Hancock和Tsang声称Pelikan自来水笔墨水是蛋白质显色最灵敏的墨水,而其他人则用Higgins黑墨水893(Faber-Castell Corporation, Newark, NJ)或Speedball黑墨水3211(Hunt Manufacturing Company, Statesville, NC)(Hughes *et al.*, 1988; Lek *et al.*, 1995)获得了更好的结果。

在印度墨水染色同期出现了蛋白质直接比色法或放射自显影免疫检测(Glenney, 1986; Mobbs *et al.*, 1989; Ono and Tuan, 1990)。如果在免疫检测前进行印度墨水染色,染色时间必须限制在2~3h,以避免干扰随后抗体结合步骤。用化学发光免疫检测与印度墨水染色相结合获得了较好效果(Eynard and Lauriere, 1998)。Mobbs则将印度墨水和免疫检测结合进行染色,而用于Edman测序的膜片仍然需用考马斯蓝染色(Mobbs *et al.*, 1989)。

与有机染料和胶体金属染料相比,印度墨水染色显示出较高的蛋白间变异(Li *et al.*, 1989)。很多蛋白质用印度墨水获得很差的染色效果,包括 β -乳球蛋白、胰蛋白酶抑制剂、杆菌肽、C-大麦醇溶蛋白、促甲状腺素释放激素(TRH)、卵白蛋白、卵类黏蛋白、碳酸脱氢酶铁蛋白和胰岛素(Dunn, 1999; Eynard and Lauriere, 1998; Li *et al.*, 1989; Rohringer and Holden, 1985)。与其他胶体扩散染料类似,印度墨水在染色膜片时也须加入Tween-20来减除背景,这是个很大的缺点,因为去污剂的加入将从转印膜上洗脱部分蛋白质(Li *et al.*, 1988)。染色过程中Tween-20的加入所带来的相关问题将在4.2中详细讨论。最后要说明的是,自来水笔墨水是设计用来书写,而不是用来进行与蛋白质染色性能相关的质量控制分析。用印度墨水染色需要针对膜、狭缝膜类型、印度墨水品牌和墨水进行实验室内优化(Hughes *et al.*, 1988)。

4.2 胶体金属染料

印度墨水开始染色电转印蛋白质后不久,其他同样用途的分散颗粒即进行了优化(Moeremans *et al.*, 1985, 1986; Rohringer and Holden, 1985)。Janssen Pharmaceuticals(Beerse, Belgium)先后开发了灵敏的胶体金(AuroDye)、铁(FerriDye)以及银染方法。由于FerriDye包含有毒的卡可基酸并会生成氰化物气体而停止使用。胶体银染没有商品化。80年代对常用胶体扩散染料的直接比较显示胶体金染料的检测灵敏度要优于印度墨水或胶体铁染料,但其有用线性范围特别有限(1~20ng)(Hunter and Hunter, 1987; Li *et al.*, 1989),蛋白质浓度的对数转换使胶体金染料的有用线性范围扩展至100ng左右(Li *et al.*, 1989)。

胶体金染色可用于传统的比色免疫印渍或化学发光免疫检测(Chevallet *et al.*, 1997; Egger and Bienz, 1987; Schapira and Keir, 1988)。但是,金属的氧化使得在化学发光过程中胶体金染色图褪色。同时,由于图谱上蛋白质化学发光变弱,带来的问题就是如何在反应较弱时分辨非特异性背景。从电转印的蛋白质上去除胶体金是不太可行的,可以用10%的盐酸、重铬酸钾的水溶液氧化作用来脱色(Nelson, 1993),染色强度也可通过应用银染增强步骤来加强(Moeremans *et al.*, 1984)。

虽然胶体金染色可以获得特别高的灵敏度,但与其他胶体扩散染色方法一样,它需要用诸如Tween-20一类试剂对膜进行前处理。随后在用去污剂稳定化的胶体溶液中孵

育 2~18h (Moeremans *et al.*, 1985)。如先前所述, 非离子表面活性剂去除了部分膜上蛋白质, 导致用于随后表征的蛋白质回收率降低 (Flanagan and Yost, 1984; Li *et al.*, 1988; Miranda *et al.*, 1993)。用氢氧化钾或戊二醛一类试剂固定或 100℃ 下烘烤会减少由去污剂提取作用所引起的蛋白质丢失, 但经常会使蛋白质不适于其他应用 (Li *et al.*, 1988)。去污剂也会影响 MALDI-TOF 和电喷雾离子化 (ESI) 质谱图谱的高质量采集。因为来源于去污剂簇峰的背景离子信号主导了质谱图, 去污剂也抑制了来自蛋白质的信号 (Gharahdaghi *et al.*, 1996; Loo *et al.*, 1994)。例如已知 Tween-20 在 ESI-MS 谱图中可以产生聚合物背景 (质荷比在 500~1200) 和严重的信号抑制, 并和蛋白质形成加合物 (Loo *et al.*, 1994)。文献中尚没有胶体金染色用于质谱鉴定的成功报道。此外, 胶体金染料一直干扰蛋白质 Edman 降解序列分析, 如很低的初始和重复序列产率 (Christiansen and Houen, 1992), 脱色并不能提高产率 (Nelson, 1993)。

正是由于胶体金染料与下游蛋白质微量化学方法不兼容, 促使了其他可替代胶体染料研究的进展。一种不含去污剂的“胶体”银染方法取代胶体金染色, 可用来检测用于液相色谱——电喷雾离子化串联质谱分析的电转膜蛋白质 (van Oostveen *et al.*, 1997)。该染液实际上并非胶体, 只是微粒的混合浆液, 我们用银浆液的经验是在很高的灵敏度对原始文献中提及的测试蛋白质碳酸酐酶和牛血清白蛋白进行检测, 但对商品化的分子质量标准中其他蛋白质染色效果相当差。另外一种用于电转膜蛋白质染色的悬浊液是碱性碘化铜溶液 (Root and Reisler, 1989; Root and Wang, 1993)。与胶体金染色一样, 铜染可以用银增强步骤来加强 (Root and Wang, 1993)。染色可用 EDTA 轻易去除, 脱色膜可用于随后比色免疫检测 (Root and Reisler, 1989), 但是, 该方法目前与其他现代微量化学方法仍不兼容。

5. 有机荧光团染料

电泳后对荧光检测蛋白质正获得普遍的应用, 特别是在致力于大规模蛋白质组研究的实验室。几种新的荧光染色适于与进行蛋白质整体表达分析的集成化蛋白质组学平台相结合, 利用荧光染色检测的动态线性范围通常优于比色法染色。

5.1 共价结合荧光团

胶内蛋白质和转膜蛋白质的荧光检测有多种方法, 分别是利用荧光胺、荧光素异硫氰酸盐、monobromobimane、2-甲氧基-2,4-二苯基-3(2H)咪唑酮 (MDPF)、Dichlorotriazinylamino-fluorescein 或丹磺酰氯 (Houston and Peddie, 1989; Jackson *et al.*, 1988; Ragland *et al.*, 1974; Szewczyk and Summers, 1987, 1992; Talbot and Yphantis, 1971; Urwin and Jackson, 1993; Vandekerckhove *et al.*, 1985; Vera and Rivas, 1988)。最近, 利用丙基 Cy3 和甲基 Cy5 两种染料分别对两个不同蛋白质样品进行荧光标记, 并在一块 2-D 胶上同步显示两个样品的差异 (Unlu *et al.*, 1997)。

上述方法的缺点包括需要对蛋白质进行共价修饰、改变了标记蛋白质的移动性能和由于荧光探针光漂白作用所引起的快速衰减。蛋白与蛋白之间由于可被荧光团修饰的功能基团数目不同而使得灵敏度变化很大, 荧光团的加入同时会降低蛋白质溶解性能 (Unlu *et al.*, 1997)。可以通过标记样品中一小部分功能基团来补偿, 但标记蛋白质与

未标记蛋白质之间轻微的分子质量差异可能会导致分离后蛋白质的 Edman 降解和质谱鉴定出现偏差。在运行 2-DE 或等电聚焦 (IEF) 时, 由前期荧光分子衍生可能导致异常蛋白质迁移 (Urwin and Jackson, 1993)。通常情况下, 样品中衍生的氨基酸总数很少, 对下游 Edman 测序或肽谱研究等氨基酸分析的影响不大 (Alba and Daban, 1998; Stephens, 1975; Unlu *et al.*, 1997)。

研究开发的一种转膜蛋白质检测方法是用由双 (2,4,6-三氯苯基) oxylate 和过氧化氢进行非酶化学荧光反应, 来显色用 MDPF 共价标记的电转印或点印渍蛋白质 (Alba and Daban, 1997)。后来发现利用 300nm 紫外透射仪对 PVDF 膜上 MDPF 标记蛋白质进行直接荧光检测, 比原始化学荧光法更简单、耗时短、更灵敏 (Alba and Daban, 1998)。直接荧光法可以实现 5~10ng 蛋白质检测并与化学荧光免疫法兼容。但是, 由于干燥 PVDF 膜会致使荧光信号降低 500 倍, 因此用 MDPF 标记的蛋白质必须在湿 PVDF 膜上观察, 目前没有适用于显色转印于硝酸纤维素膜上蛋白质的 MDPF 标记方法, 如果蛋白质要用于 N 端测序, 为了在反应中尽量少的消耗一级胺, 染色时染料浓度要低于最佳浓度, 染色时间要短, 在这种条件下荧光染色灵敏度不比标准考马斯蓝色高 (Alba and Daban, 1998)。

5.2 非共价结合荧光团

许多荧光团如 1-苯胺基-8-萘磺酸盐 (ANS)、双 (8-甲苯氨基-1-萘磺酸盐) (bis-ANS)、9-二乙氨基-5H-苯并 [α] 吩噁嗪-5-酮 (尼罗红, Nile red)、SYPRO Orange 和 SYPRO Red 染料可与蛋白质形成非共价键相互作用 (Aragay *et al.*, 1985; Bermudez *et al.*, 1994; Daban and Aragay, 1984; Daban *et al.*, 1991a, 1991b; Hartman and Undenfriend, 1969; Horowitz and Bowman, 1987; Pina *et al.*, 1985; Steinberg *et al.*, 1996a, 1996b, 1997)。实际上, 这些荧光团在水溶液中是不发荧光的, 而在非极性溶剂或与 SDS-蛋白质复合物相互作用时显示荧光。SDS 与蛋白质的结合计量常数比较稳定, 保持在 1.4 到 1, 提示, 基于与去污剂的相互作用进行定量要比基于与一级胺的作用更为可靠。

ANS 可用 300nm 透射仪激发, 并在 480nm 发射, 已用于检测 SDS-PAGE 胶内蛋白质, 但其灵敏度明显低于考染 (Aragay *et al.*, 1985; Daban and Aragay, 1984; Pina *et al.*, 1985)。相关染料——bis-ANS 与 2mol/L KCl 组合可以获得稍好的灵敏度, 其检测限大约 100ng 蛋白质 (Horowitz and Bowman, 1987)。尼罗红染料可在 300 或 540nm 处激发, 在 640nm 处发射 (Daban *et al.*, 1991a), 其 20~100ng 的检测灵敏度要稍好于考马斯蓝, 而且, 该染料与标准 Edman 测序与免疫检测程序兼容 (Alba and Daban, 1998; Bermudez *et al.*, 1994; Daban *et al.*, 1991b)。

SYPRO Red 和 SYPRO Orange 染料可对 SDS-PAGE 胶内蛋白质进行简单的一步染色, 约 30~60min 即可完成, 可以检测少至 2~10ng 的蛋白质, 可与银染相媲美。染色后胶的备份用标准的实验室 300nm 紫外透射仪照相即可轻松完成, 或者通过基于商品化的 CCD 相机图像分析工作站或激光扫描仪进行定量, 可以获得三个数量级的线性范围。SYPRO Red、Orange 的 Ruby 染料是仅有的商品化染料, 其电泳染色结果与在酵母中通过基因表达序列分析 (SAGE) 所获得的基因表达水平动态范围相匹配 (每个细胞

0.3~200 个转录拷贝; Verculescu *et al.*, 1997)。SYPRO 染料可以很好地抗光淬灭作用, 在 Tris/甘氨酸转印缓冲液中染色后, 蛋白质可被转印至膜上并进行免疫染色或 Edman 测序来鉴定蛋白质 (Hamby, 1996, 1997; Steinberg *et al.*, 1996b)。最后要指出的是, SYPRO Orange 和 SYPRO Red 染料可用于对 SDS-毛细管凝胶电泳分离蛋白质的 pmol 定量检测 (Harvey *et al.*, 1998)。与 SYPRO 染料相比, 利用尼罗红的检测灵敏度较差, 染色不稳定, 染料在水溶液中易于沉淀, 使随后处理很困难, 尼罗红染色对光不稳定的性质也加大了通过图像分析进行蛋白质定量的难度。

6. 金属螯合染料

金属螯合染料是与现代蛋白质组学研究相兼容的、相对较新的蛋白质显色试剂, 其设计专门与常用微量化学表征过程兼容。这些染色不包含影响表征过程的其他试剂如戊二醛、甲醛或 Tween-20。发光金属螯合染料很容易和集成化蛋白质组学平台相结合, 包括自动化凝胶染色仪、图像分析工作站、机器人剪切仪器、蛋白质酶解工作站和质谱仪。最为灵敏的发光染料——SYPRO Ruby 比最好的碱性二氨银和酸性硝酸银染色的灵敏度高, 且线性范围宽。

6.1 比色金属螯合染料

最早报道的用于凝胶电泳蛋白质检测的金属螯合剂为红菲绕啉二磺酸盐/亚铁复合物 (Graham *et al.*, 1978)。报道其灵敏度大约在 600ng。由于它较已经实践应用的考马斯蓝染色没有明显优势, 因而没得到广泛认可。将放射性⁵⁹Fe 引入红菲绕啉二磺酸盐复合物可以检测到 10~25ng 蛋白质, 但按今天的标准也只是和胶体考马斯蓝的灵敏度接近 (Gersten *et al.*, 1984; Zapolski *et al.*, 1982), 放射性环境下工作的危害性和获准使用放射性同位素所增加的麻烦排除了其广泛使用的可能性。酞菁四磺酸酮金属螯合物, 可以在进行 PAGE 胶染色时达到与标准考马斯蓝染色相同的灵敏度 (Bickar and Reid, 1992)。该染料也可用于固定在硝酸纤维素上蛋白质的染色, 其检测限在 10~20ng 蛋白质。

虽然灵敏度稍低, 氨基黑、丽春红 S 和考马斯蓝仍是最常用的检测转膜蛋白质的方法, 因为它们与下游微量化学表征技术兼容。我们的研究小组针对灵敏度问题设计了一系列金属螯合染料以检测转膜蛋白质 (Lim *et al.*, 1996; Patton *et al.*, 1994, 1999a, b; Shojaee *et al.*, 1996)。在 Chung 成功地用 ferreneS 检测 PAGE 胶内铁结合蛋白质的基础上, 我们推测如果外加的亚铁离子附着在蛋白质上, 如果遇到铁指示剂如 ferrene S 或锌铁, 在转印膜上即可得到全蛋白质显色图 (Chung, 1985; Patton *et al.*, 1994)。我们优化了方法, 简化到只要在转膜蛋白质孵育前混合亚铁离子和螯合剂即可, 但直到发表时我们仍确信与蛋白质的结合主要由金属离子独自调节完成 (图 4.1) (Patton *et al.*, 1994)。在开发可替代性金属螯合染料的过程中筛选了大量金属螯合物后我们才意识到与蛋白质的结合是由金属螯合物上的磺酸基团所介导的 (Shojaee *et al.*, 1996), 因此, 金属螯合染料与 1978 年开发的红菲绕啉二磺酸盐/亚铁染料属于同一家族。

锌铁/亚铁和焦酚红/钼酸盐染料的灵敏度和氨基黑不相上下, 通常可以检测到 20~50ng 蛋白质。锌铁/亚铁染色的优点是通过增加溶液 pH 值可以容易脱色, 并且可

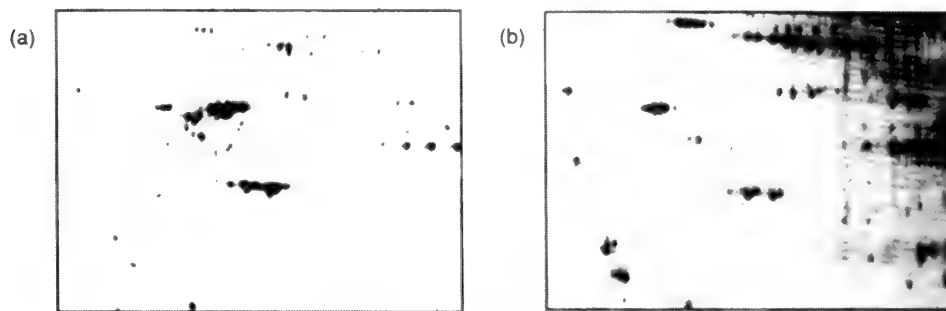


图 4.1 2-DE 凝胶染色区域

(a) 蛋白质凝胶 SYPRO Ruby 染色 (Molecular Probes, Eugene, OR); (b) 酸性/硝酸银染色试剂盒 (Genomic Solutions, Ann Arbor, MI) 染色。一维 IEF 胶上样为鼠成纤维细胞裂解蛋白质, 约 $5\mu\text{g}$ 。为了与银染胶容易比较, SYPRO Ruby 染色胶的灰度值进行了转换。

以很容易地在锌铁/亚铁溶液中孵育来加强信号 (Lim *et al.*, 1996; Patton *et al.*, 1994), 此时其检测限可达到 $5\sim 10\text{ng}$ 蛋白质 (Patton *et al.*, 1994)。金属螯合染料还和 Edman 测序、质谱和免疫印渍兼容 (Patton *et al.*, 1994)。

6.2 发光金属螯合染料

由于对比色染料的检测灵敏度不太满意, 我们对发光金属螯合染料进行了研究 (Lim *et al.*, 1997; Patton *et al.*, 1999a, 1999b)。从原理上讲, 检测光散射要比检测光吸收灵敏一些, 由于光吸收受有色复合物摩尔消光系数的限制 (Gossling, 1990), 因此, 利用金属螯合物与特定过渡金属离子如铈、铽、钇、镱复合而形成的发光蛋白质染料, 与上述比色法相比可以得到更高灵敏度。

红非绕啉二磺酸盐/铈 (SYPRO Rose Protein Blot stain, Molecular Probes, Eugene, OR) 是一种中等灵敏度的染料, 通常可以检测到由狭缝印渍所得的 $2\sim 4\text{ng}/\text{mm}^2$ 的蛋白质上样量。依据常规的电转印实验, 该数值转换成胶内上样为 15 到 30ng。该染料的显色可以用 300nm 紫外表面照射而在 590nm 和 615nm 发射。在中性或弱碱性环境中孵育可以洗脱染料。由于与免疫印渍、凝集素印渍、Edman 测序和质谱的兼容性, 使 SYPRO Rose 染料成为近乎与丽春红 S 染料一样简便而优良的常规蛋白质检测试剂 (Lim *et al.*, 1997)。其缺陷是可以对核酸像蛋白质一样进行染色, 脱色步骤很像传统考马斯蓝染色, 许多激光凝胶扫描仪不能用, 因为该染料不能被可见光激发。

SYPRO Ruby 是一种专利的基于钇的金属螯合染料, 其开发是为了弥补 SYPRO Rose 染料的缺陷。SYPRO Ruby 蛋白印渍染料 (Molecular Probes, Eugene, OR) 可以对硝酸纤维素或 PVDF 转膜蛋白质永久染色, 在狭缝印渍时其检测灵敏度为 $0.25\sim 1\text{ng}/\text{mm}^2$, 常规大约 $2\sim 8\text{ng}$ 的蛋白质可通过电转印检测, 对照比较实验证实它具有和胶体金染色一样的灵敏度 (AuroDye, Amersham Pharmacia Biotech, Arlington Heights, IL)。而胶体金染色需要 $2\sim 4\text{h}$, SYPRO Ruby 染料染色在 15 min 内即可完成。SYPRO Ruby 蛋白印渍染色的动态线性范围大大优于胶体金染色, 扩展了约 1000 倍。该染料可用 300nm 紫外透射仪激活或利用装配有 450、473、488 乃至 532nm 激光的成像系统。

不像胶体金染色, SYPRO Ruby 染料不干扰质谱或免疫检测过程, 又与 SYPRO Rose 不同, 它对核酸不染色。

SYPRO Ruby 2-DE 和 IEF 染色仅需一步即可获得 PAGE 内蛋白质的低背景染色, 而且不需要进行长时间脱色步骤。这些染料的线性动态范围扩展到超过三个数量级之大, 在性能上超过了银染和考马斯蓝染色, 对等电点从 3.5 到 9.3 范围的 11 个蛋白质标准的评估表明 SYPRO Ruby IEF 凝胶染色的灵敏度比高灵敏银染要高 3~30 倍 (Steinberg *et al.*, 2000)。用银染效果很差的蛋白质往往用 SYPRO Ruby 染料很容易检测到 (图 4.1)。虽然比 SYPRO Orange 和 Red 染料更为灵敏, 其最佳染色要稍慢一些, 约需 4h。与胶体考马斯蓝染料类似, 但与银染不同, SYPRO Ruby 染料的染色是终点式 (end-point) 染色, 因此, 染色时间不严格, 可以过夜染色且不会显色过度。

7. 结论

通过大规模基因组测序以及基因芯片的辅助, DNA 和 mRNA 水平上生物学过程的分析正在迅速完成, 人们的注意力正转向蛋白质的动态表达。蛋白质组研究致力于观察一个特定基因组所编码的全蛋白质组分, 以解决不能单纯通过核酸序列来回答的生物学问题 (De Francesco, 1999)。随着蛋白质组学发展成为研究整体蛋白质调控的高通量方法, 对蛋白质显色方法提出了新的需求 (表 4.2)。旧的蛋白质染色如考马斯蓝和银染已经不能满足人们的期望。有幸的是, 较新的染料如锌-咪唑及非共价荧光染料 (SYPRO 染料) 正在用来满足日益增长的现代蛋白质微量表征技术的需求。

表 4.2 理想通用的蛋白质染料性质; 优良染色方法应具备的 7 个 S

安全 (Safety): 该染色方法不应该含有害材料如放射性同位素或氰化物气体、卡列基酸等强毒性化学品。从而保证人身安全并消除耗时的监控、清洗和处理等问题
灵敏 (Sensitivity): 该染色方法可以检测到亚纳克级蛋白质, 染色的线性动态范围要宽, 最少要扩展至三个数量级
简单 (Simplicity): 该染色方法只需凝胶或膜在一种溶液中简单的孵育过程, 对显色时间的要求不能太苛刻, 因为染色应可以达到一个终点而不会显色过度
特异性 (Specificity): 该染色方法应该特异性检测蛋白质, 而不检测脂类、糖类或核酸, 对不同类型蛋白质如钙结合蛋白、脂蛋白、糖蛋白和磷酸化蛋白质均匀检测
快速 (Speed): 该染色方法可以快速检测蛋白, 不需脱色。只要不需监控显色过程或是变换溶液, 采取较长的孵育时间往往较为灵活, 如过夜染色
稳定 (Stability): 染液应当具有化学稳定性, 保存期长, 无需用前新鲜配制, 染液在室温下贮存应该保持稳定。染色后的胶应可以在几天至几个月时间内保持稳定, 而染色信号不会变差
兼容性 (Synergy): 染色方法应该和多种电泳/分离技术 (载体两性电解质 IEF、固相 pH 梯度 IEF、Tris-甘氨酸 SDS 胶、Tris-tricine 预制 SDS 胶、非变性凝胶、2-DE 凝胶、PVDF 膜、硝酸纤维素膜、TLC 板) 兼容。染色过程不应导致蛋白质的不可逆共价修饰。染液中不应含通常会干扰 Edman 测序和质谱的外加试剂如戊二醛、甲醛或 Tween-20

(应万涛译 钱小红校)

参考文献

- Adams, L. and Weaver, K. (1990) Detection and recovery of proteins from gels following zinc chloride staining. *Appl. Theor. Electrophoresis* **1**: 279–282.
- Alba, F. and Daban, J. (1997) Nonenzymatic chemiluminescent detection and quantitation of total protein on Western and slot blots allowing subsequent immunodetection and sequencing. *Electrophoresis* **18**: 1960–1966.
- Alba, F. and Daban, J. (1998) Rapid fluorescent monitoring of total protein patterns on sodium dodecyl sulfate–polyacrylamide gels and Western blots before immunodetection and sequencing. *Electrophoresis* **19**: 2407–2411.
- Aragay, A., Diaz, P. and Daban, J. (1985) A fluorescent method for the rapid staining and quantitation of proteins in sodium dodecyl sulfate–polyacrylamide gel electrophoresis. *Electrophoresis* **6**: 527–531.
- Bermudez, A., Daban, J., Garcia, J. and Mendez, E. (1994) Direct blotting, sequencing and immunodetection of proteins after five-minute staining of SDS and SDS-treated IEF gels with Nile red. *BioTechniques* **16**: 621–624.
- Berggren, K., Steinberg, T., Lauber, W., Carroll, J., Lopez, M., Chernokalskaya, E., Zieske, L., Diwu, Z., Haugland, R. and Patton, W. (1999) A luminescent ruthenium complex for ultrasensitive detection of proteins immobilized on membrane supports. *Anal. biochem.* **276**: 129–143.
- Berube, G. (1990) Visualization of proteins on electrophoresis gels using planar dyes. US patent no. 4,946,749, 7 August.
- Bickar, D. and Reid, P. (1992) A high-affinity protein stain for western blots, tissue prints, and electrophoretic gels. *Anal. Biochem.* **203**: 109–115.
- Brush, M. (1998) Dye hard; protein gel staining products. *Scientist* **12**: 16–22.
- Candiano, G., Porotto, M., Lanciotti, M. and Ghiggeri, G. (1996) Negative staining of proteins in polyacrylamide gels with methyl trichloroacetate. *Anal. Biochem.* **243**: 245–248.
- Casero, P., Del Campo, G. and Righetti, P. (1985) Negative Aurodye for polyacrylamide gels: the impossible stain. *Electrophoresis* **6**: 362–372.
- Castellanos-Serra, L., Fernandez-Patron, C., Hardy, E. and Huerta, V. (1996) A procedure for protein elution from reverse-stained polyacrylamide gels applicable at the low picomole level: an alternative route to the preparation of low abundance proteins for microanalysis. *Electrophoresis* **17**: 1564–1572.
- Chevallet, M., Procaccio, V. and Rabilloud, T. (1997) A nonradioactive double detection method for the assignment of spots in two-dimensional blots. *Anal. Biochem.* **251**: 69–72.
- Chrambach, A., Reisfeld, R., Wyckhoff, M. and Zaccari, J. (1967) A procedure for rapid and sensitive staining of protein fractionated by polyacrylamide gel electrophoresis. *Anal. Biochem.* **20**: 150–154.
- Christiansen, J. and Houen, G. (1992) Comparison of different staining methods for polyvinylidene difluoride membranes. *Electrophoresis* **13**: 179–183.
- Chung, M. (1985) A specific iron stain for iron-binding proteins in polyacrylamide gels: application to transferrin and lactoferrin. *Anal. Biochem.* **148**: 498–502.
- Cohen, S. and Chait, B. (1997) Mass spectrometry of whole proteins eluted from sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels. *Anal. Biochem.* **247**: 257–267.
- Daban, J. and Aragay, A. (1984) Rapid fluorescent staining of histones in sodium dodecyl sulfate–polyacrylamide gels. *Anal. Biochem.* **138**: 223–228.
- Daban, J., Samsó, M. and Bartolomé, S. (1991a) Use of Nile red as a fluorescent probe for the study of the hydrophobic properties of protein–sodium dodecyl sulfate complexes in solution. *Anal. Biochem.* **199**: 162–168.
- Daban, J., Bartolomé, S. and Samsó, M. (1991b) Use of the hydrophobic probe Nile red for the fluorescent staining of protein bands in sodium dodecyl sulfate–polyacrylamide gels. *Anal. Biochem.* **199**: 169–174.
- De Francesco, L. (1999) One step beyond; going beyond genomics with proteomics and two-dimensional gel technology. *Scientist* **13**(1): 16–18.
- Diezel, W., Kopperschlager, G. and Hofman, E. (1972) An improved procedure for protein

- staining in polyacrylamide gels with a new type of Coomassie brilliant blue. *Anal. Biochem.* **48**: 617–620.
- Dubray, G. and Bezar, G.** (1982) A highly sensitive periodic acid-silver stain for 1,2-diol groups of glycoproteins and polysaccharides in polyacrylamide gels. *Anal. Biochem.* **119**: 325–329.
- Dunn, M.** (1987) Two-dimensional polyacrylamide gel electrophoresis. In: *Advances in Electrophoresis*, Vol. 1 (eds A. Chrambach, M. Dunn and B. Radola). VCH Publishers, New York, pp. 4–109.
- Dunn, M.** (1999) Detection of total proteins on Western blots of 2-D polyacrylamide gels. In: *Methods in Molecular Biology*, Vol. 112, *2-D Proteome Analysis Protocols* (ed. A. Link). Humana Press, Totowa, NJ, pp. 319–329.
- Durrum, E.** (1950) A microelectrophoretic and microionophoretic technique. *J. Am. Chem. Soc.* **72**: 2943–2948.
- Dzandu, J., Johnson, J. and Wise, G.** (1988) Sodium dodecyl sulfate-gel electrophoresis: staining of polypeptides using heavy metal salts. *Anal. Biochem.* **174**: 157–167.
- Egger, D. and Bienz, K.** (1987) Colloidal gold staining and immunoprobings of proteins on the same nitrocellulose blot. *Anal. Biochem.* **166**: 413–417.
- Eynard, L. and Lauriere, M.** (1998) The combination of Indian ink with chemiluminescence detection allows precise identification of antigens on blots: application to the study of glycosylated barley storage proteins. *Electrophoresis* **19**: 1394–1396.
- Fazekas de St. Groth, S., Webster, R. and Datyner, A.** (1963) Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. *Biochim. Biophys. Acta* **71**: 377–391.
- Fernandez-Patron, C., Calero, M., Garcia, J., Madrazo, J., Musacchio, A., Soriano, F., Estrada, R., Frank, R., Castellanos-Serra, L. and Mendez, E.** (1995) Protein reverse staining: high-efficiency microanalysis of unmodified proteins detected on electrophoresis gels. *Anal. Biochem.* **224**: 203–211.
- Fernandez-Patron, C., Castellanos-Serra, L., Hardy, E., Guerra, M., Estevez, E., Mehl, E. and Frank, R.** (1998) Understanding the mechanism of the zinc-ion stains of biomacromolecules in electrophoresis gels: generalization of the reverse-staining technique. *Electrophoresis* **19**: 2398–2406.
- Ferreras, M., Gavilanes, J. and Garcia-Segura, J.** (1993) A permanent Zn^{2+} reverse staining method for the detection and quantification of proteins in polyacrylamide gels. *Anal. Biochem.* **213**: 206–212.
- Flanagan, S. and Yost, B.** (1984) Calmodulin-binding proteins: visualization by ^{125}I -calmodulin overlay on blots quenched with Tween 20 or bovine serum albumin and poly(ethylene oxide). *Anal. Biochem.* **140**: 510–519.
- Fowler, S.** (1994) The detection of proteins on blots using gold or immunogold. *Methods Mol. Biol.* **32**: 239–255.
- Gersten, D., Zapolski, E. and Ledley, R.** (1984) Radioactive staining of gels to identify proteins. US Patent no. 4,459,356, 10 July.
- Gharahdaghi, F., Kirchner, M., Fernandez, J. and Mische, S.** (1996) Peptide-mass profiles of polyvinylidene difluoride-bound proteins by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in the presence of nonionic detergents. *Anal. Biochem.* **233**: 94–99.
- Glennay, J.** (1986) Antibody probing of western blots which have been stained with India ink. *Anal. Biochem.* **156**: 315–319.
- Gossling, J.** (1990) A decade of development in immunoassay methodology. *Clin. Chem.* **36**: 1408–1427.
- Graham, G., Nairn, R. and Bates, G.** (1978) Polyacrylamide gel staining with Fe^{2+} -bathophenanthroline sulfonate. *Anal. Biochem.* **88**: 434–441.
- Grassman, W. and Hannig, K.** (1952) Ein quantitatives verfahren zur analyse der serumproteine durch papierelektrophorese. *Hoppe-Seylev's Z. Physiol. Chem.* **290**: 1–27.
- Haebel, S., Albrecht, T., Sparbier, K., Walden, P., Korner, R. and Steup, M.** (1998) Electrophoresis-related protein modification: alkylation of carboxy residues revealed by mass spectrometry. *Electrophoresis* **19**: 679–686.
- Hager, D. and Burgess, R.** (1980) Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results with

- sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* **109**: 76–86.
- Hamby, R.** (1996) A sensitive fluorescent gel stain. *Am. Biotechnol. Lab.* **14**: 12.
- Hamby, R.** (1997) A fluorescent gel stain for detection of proteins following 1-D, 2-D and native polyacrylamide gel electrophoresis. *Biotechnol. Int.* **1**: 339–343.
- Hancock, K. and Tsang, V.** (1983) India ink staining of proteins on nitrocellulose paper. *Anal. Biochem.* **133**: 157–162.
- Hartman, B. and Udenfriend, S.** (1969) A method for immediate visualization of proteins in acrylamide gels and its use for preparation of antibodies to enzymes. *Anal. Biochem.* **30**: 391–394.
- Harvey, M., Bandilla, D. and Banks, P.** (1998) Subnanomolar detection limit for sodium dodecyl sulfate–capillary gel electrophoresis using a fluorogenic, noncovalent dye. *Electrophoresis* **19**: 2169–2174.
- Higgins, R. and Dahmus, M.** (1979) Rapid visualization of protein bands in preparative SDS–polyacrylamide gels. *Anal. Biochem.* **93**: 257–260.
- Hochstrasser, D., Patchornik, A. and Merrill, C.** (1988) Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Anal. Biochem.* **173**: 412–423.
- Horowitz, P. and Bowman, S.** (1987) Ion-enhanced fluorescence staining of sodium dodecyl sulfate–polyacrylamide gels using bis(8-*p*-toluidino-1-naphthalenesulfonate). *Anal. Biochem.* **165**: 430–434.
- Houston, B. and Peddie, D.** (1989) A method for detecting proteins immobilized on nitrocellulose membranes by in situ derivatization with fluorescein isothiocyanate. *Anal. Biochem.* **177**: 263–267.
- Hughes, J., Mack, K. and Hamparian, V.** (1988) India ink staining of proteins on nylon and hydrophobic membranes. *Anal. Biochem.* **173**: 18–25.
- Hunter, J. and Hunter, S.** (1987) Quantification of proteins in the low nanogram range by staining with the colloidal gold stain AuroDye. *Anal. Biochem.* **164**: 430–433.
- Jackson, P., Urwin, V. and Mackay, C.** (1988) Rapid imaging, using a cooled charge-coupled-device, of fluorescent two-dimensional polyacrylamide gels produced by labelling proteins in the first-dimensional isoelectric focusing gel with fluorophore 2-methoxy-2,4-diphenyl-3(2H)furanone. *Electrophoresis* **9**: 330–339.
- Jay, G., Culp, D. and Jahnke, M.** (1990) Silver staining of extensively glycosylated proteins on sodium dodecyl sulfate–polyacrylamide gels: enhancement by carbohydrate-binding dyes. *Anal. Biochem.* **185**: 324–330.
- Kerenyi, L. and Gallyas, F.** (1972) A highly sensitive method for demonstrating proteins in electrophoretic, immunoelectrophoretic and immunodiffusion preparations. *Clin. Chem. Acta* **38**: 465–467.
- Lee, C., Levin, A. and Branton, D.** (1987) Copper staining: a five-minute protein stain for sodium dodecyl sulfate–polyacrylamide gels. *Anal. Biochem.* **166**: 308–312.
- Lek, L., Yang, E., Wang, D. and Cheng, L.** (1995) Comparison of the sensitivity of different India inks staining of electro-blotted proteins on filter membranes. *J. Biochem. Biophys. Methods* **30**: 9–20.
- Li, K., Geraerts, W., van Elk, R. and Joose, J.** (1988) Fixation increases sensitivity of India ink staining of proteins and peptides on nitrocellulose paper. *Anal. Biochem.* **174**: 97–100.
- Li, K., Geraerts, W., van Elk, R. and Joose, J.** (1989) Quantification of proteins in the subnanogram and nanogram range: comparison of the AuroDye, FerriDye and India ink staining methods. *Anal. Biochem.* **182**: 44–47.
- Lim, M., Patton, W., Shojaee, N. and Shepro, D.** (1996) Solid-phase metal chelate assay for quantifying total protein: resistance to chemical interference. *BioTechniques* **21**: 888–897.
- Lim, M., Patton, W., Shojaee, N., Lopez, M., Spofford, K. and Shepro, D.** (1997) A luminescent europium complex for the sensitive detection of proteins and nucleic acids immobilized on membrane supports. *Anal. Biochem.* **245**: 184–195.
- Loo, R., Dales, N. and Andrews, P.** (1994) Surfactant effects on protein structure examined by electrospray ionization mass spectrometry. *Protein Sci.* **3**: 1975–1983.
- Matsui, N., Smith, D., Clausner, K., Fichmann, J., Andrews, L., Sullivan, C., Burlingame, A. and**

- Epstein, L. (1997) Immobilized pH gradient two-dimensional gel electrophoresis and mass spectrometric identification of cytokine-regulated proteins in ME-180 cervical carcinoma cells. *Electrophoresis* **18**: 409–417.
- Merril, C. (1987) Detection of proteins separated by electrophoresis. In: *Advances in Electrophoresis*, Vol. 1 (eds A. Chrambach, M. Dunn and B. Radola). VCH Publishers, New York, pp. 111–140.
- Merril, C., Goldman, D., Sedman, S. and Ebert, M. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**: 1437–1438.
- Merril, C., Goldman, D. and Van Keuren, M. (1984) Gel protein stains: silver stain. *Methods Enzymol.* **104**: 441–447.
- Metkar, S., Mahajan, S. and Sainis, J. (1995) Modified procedure for nonspecific protein staining on nitrocellulose paper using Coomassie Brilliant Blue R-250. *Anal. Biochem.* **227**: 389–391.
- Meyer, T. and Lamberts, B. (1965) Use of Coomassie brilliant blue R250 for the electrophoresis of microgram quantities of paratoid saliva proteins on acrylamide-gel strips. *Biochim. Biophys. Acta* **107**: 144–145.
- Miranda, P., Brandelli, A. and Tezon, J. (1993) Instantaneous blocking for immunoblots. *Anal. Biochem.* **209**: 376–377.
- Mobbs, C., Berman, J., Marquardt, M. and Pfaff, D. (1989) Comprehensive polypeptide analysis of microdissected rat brain areas; combining 2-dimensional gel electrophoresis with 2-dimensional HPLC and immunoanalysis and sequencing procedures. *J. Neurosci. Methods* **29**: 5–15.
- Moeremans, M., Daneels, G., Van Dijk, A., Langanger, G. and De Mey, J. (1984) Sensitive visualization of antigen-antibody reactions in dot and blot immune overlay assays with immunogold and immunogold/silver staining. *J. Immunol. Methods* **74**: 353–360.
- Moeremans, M., Daneels, G. and De May, J. (1985) Sensitive colloidal (gold or silver) staining of protein blots on nitrocellulose membranes. *Anal. Biochem.* **145**: 315–321.
- Moeremans, M., De Raemaeker, M., Daneels, G. and De Mey, J. (1986) Ferridyde: colloidal iron binding followed by Perls' reaction for the staining of proteins transferred from sodium dodecyl sulfate gels to nitrocellulose and positively charged nylon membranes. *Anal. Biochem.* **153**: 18–22.
- Nelson, T. (1993) Destaining of nitrocellulose blots after staining with silver or colloidal gold. *Anal. Biochem.* **214**: 325–328.
- Neuhoff, V., Stamm, R. and Eibl, H. (1985) Clear background and highly sensitive protein staining with Coomassie Blue dyes in polyacrylamide gels. *Electrophoresis* **6**: 427–448.
- Neuhoff, V., Arold, N., Taube, D. and Ehrhardt, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**: 255–262.
- Neuhoff, V., Stamm, R., Pardowitz, I., Arold, N., Ehrhardt, W. and Taube, D. (1990) Essential problems in quantification of proteins following colloidal staining with Coomassie Brilliant Blue dyes in polyacrylamide gels, and their solution. *Electrophoresis* **11**: 101–117.
- Oakley, B., Kirsch, D. and Morris, N. (1980) Simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Anal. Biochem.* **105**: 361–363.
- Ono, T. and Tuan, R. (1990) Double staining of immunoblot using enzyme histochemistry and India ink. *Anal. Biochem.* **187**: 324–327.
- Ortiz, M., Calero, M., Fernandez-Patron, C., Patron, C., Castellanos, L. and Mendez, E. (1992) Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. *FEBS Lett.* **296**: 300–304.
- Patestos, N., Fauth, M. and Radola, B. (1988) Fast and sensitive protein staining with colloidal Acid Violet 17 followed by isoelectric focusing in carrier ampholyte generated and immobilized pH gradients. *Electrophoresis* **9**: 488–496.
- Patton, W. (1995) Biologist's perspective on analytical imaging systems as applied to protein gel electrophoresis. *J. Chromatogr. A* **698**: 55–87.
- Patton, W., Lopez, M., Barry, P. and Skea, W. (1992) A mechanically strong matrix for protein electrophoresis with enhanced silver staining properties. *BioTechniques* **12**: 580–585.
- Patton, W., Lam, L., Su, Q., Lui, M., Erdjument-Bromage, H. and Tempst, P. (1994) Metal

- chelates as reversible stains for detection of electroblotted proteins: application to protein microsequencing and immunoblotting. *Anal. Biochem.* **220**: 324–335.
- Patton, W., Lim, M. and Shepro, D.** (1999a) Protein detection using reversible metal chelate stains. In: *Methods in Molecular Biology*, Vol. 112, *2-D Proteome Analysis Protocols* (ed. A. Link). Humana Press, Totowa, NJ, pp. 331–339.
- Patton, W., Lim, M. and Shepro, D.** (1999b) Image acquisition in 2-D electrophoresis. In: *Methods in Molecular Biology*, Vol. 112, *2-D Proteome Analysis Protocols* (ed. A. Link). Humana Press, Totowa, NJ, pp. 353–362.
- Peats, S.** (1984) Quantitation of protein and DNA in silver-stained agarose gels. *Anal. Biochem.* **140**: 178–182.
- Pina, B., Aragay, A., Suau, P. and Daban, J.** (1985) Fluorescent properties of histone-1-anilino-naphthalene 8-sulfonate complexes in the presence of denaturant agents: application to the rapid staining of histones in urea and Triton-urea-polyacrylamide gels. *Anal. Biochem.* **146**: 431–433.
- Powell, R.** (1998) Small organometallic probes. US patent no. 5,728,590, 17 March.
- Pryor, J., Xu, W. and Hamilton, D.** (1992) Immunodetection after complete destaining of Coomassie blue-stained proteins on immobilized-PVDF membrane. *Anal. Biochem.* **202**: 100–104.
- Rabilloud, T.** (1990) Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis. *Electrophoresis* **11**: 785–794.
- Rabilloud, T.** (1992) A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis* **13**: 429–439.
- Rabilloud, T.** (1999) Silver staining of 2-D electrophoresis gels. In: *Methods in Molecular Biology*, Vol. 112, *2-D Proteome Analysis Protocols* (ed. A. Link). Humana Press, Totowa, NJ, pp. 297–305.
- Rabilloud, T., Vuillard, L., Gilly, C. and Lawrence, J.** (1994) Silver staining of proteins in polyacrylamide gels: a general overview. *Cell. Mol. Biol.* **40**: 57–75.
- Ragland, W., Pace, J. and Kemper, D.** (1974) Fluorimetric scanning of fluorescamine-labeled proteins in polyacrylamide gels. *Anal. Biochem.* **59**: 24–33.
- Rodriguez, L., Gersten, D., Ramagli, L. and Johnston, D.** (1993) Towards stoichiometric silver staining of proteins resolved in complex two-dimensional electrophoresis gels: real-time analysis of pattern development. *Electrophoresis* **14**: 628–637.
- Rohde, E., Vogt, C. and Heineman, W.** (1998) The analysis of fountain pen inks by capillary electrophoresis with ultraviolet/visible absorbance and laser-induced fluorescence detection. *Electrophoresis* **19**: 31–41.
- Rohringer, R. and Holden, D.** (1985) Protein blotting: detection of proteins with colloidal gold, and of glycoproteins and lectins with biotin-conjugated and enzyme probes. *Anal. Biochem.* **144**: 118–127.
- Root, D. and Reisler, E.** (1989) Copper iodide staining of protein blots on nitrocellulose membranes. *Anal. Biochem.* **181**: 250–253.
- Root, D. and Wang, K.** (1993) Silver-enhanced copper staining of protein blots. *Anal. Biochem.* **209**: 15–19.
- Schapira, A. and Keir, G.** (1988) Two-dimensional protein mapping by gold stain and immunoblotting. *Anal. Biochem.* **169**: 167–171.
- Scheler, C., Lamer, S., Pan, Z., Li, X., Salnikow, J. and Jungblut, P.** (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* **19**: 918–927.
- Schleicher, M. and Watterson, D.** (1983) Analysis of differences between Coomassie blue stain and silver stain procedures in polyacrylamide gels: conditions for the detection of calmodulin and troponin C. *Anal. Biochem.* **131**: 312–317.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M.** (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal. Chem.* **68**: 850–858.
- Shojaee, N., Patton, W., Lim, M. and Shepro, D.** (1996) Pyrogallol red-molybdate: a reversible, metal chelate stain for detection of proteins immobilized on membrane supports. *Electrophoresis* **17**: 687–693.

- Steinberg, T., Jones, L., Haugland, R. and Singer, V. (1996a) SYPRO Orange and SYPRO Red protein gel stains: one-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal. Biochem.* **239**: 223–237.
- Steinberg, T., Haugland, R. and Singer, V. (1996b) Applications of SYPRO Orange and SYPRO Red protein gel stains. *Anal. Biochem.* **239**: 238–245.
- Steinberg, T., White, H. and Singer, V. (1997) Optimal filter combinations for photographing SYPRO Orange or Red dye-stained gels. *Anal. Biochem.* **248**: 168–172.
- Steinberg, T., Chernokalskaya, E., Berggren, K., Lopez, M., Diwu, Z., Haugland, R. and Patton, W. (2000) Ultrasensitive fluorescence protein detection in isoelectric focusing gels using a ruthenium metal chelate stain. *Electrophoresis* **21**: 486–496.
- Stephens, R. (1975) High-resolution preparative SDS–polyacrylamide gel electrophoresis: fluorescent visualization and electrophoretic elution-concentration of protein bands. *Anal. Biochem.* **65**: 369–379.
- Switzer, R., Merrill, C. and Shifrin, S. (1979) A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal. Biochem.* **98**: 231–237.
- Szewczyk, B. and Summers, D. (1987) Fluorescent staining of proteins transferred to nitrocellulose allowing for subsequent probing with antisera. *Anal. Biochem.* **164**: 303–306.
- Szewczyk, B. and Summers, D. (1992) Fluorescent protein staining on nitrocellulose with subsequent immunodetection of antigen. In: *Methods in Molecular Biology*, vol. 10: *Immunochemical Protocols* (ed. M. Manson). Humana Press, Totowa, NJ.
- Takagi, T., Kubo, K. and Isemura, T. (1977) Simple visualization of protein bands in SDS–polyacrylamide gel electrophoresis by the insoluble complex formation between SDS and a cationic surfactant. *Anal. Biochem.* **79**: 104–109.
- Talbot, D. and Yphantis, D. (1971) Fluorescent monitoring of SDS gel electrophoresis. *Anal. Biochem.* **44**: 246–253.
- Unlu, M., Morgan, M. and Minden, J. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071–2077.
- Urwin, V. and Jackson, P. (1993) Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: imaging of the fluorescent protein spot patterns using a cooled charge-coupled device. *Anal. Biochem.* **209**: 57–62.
- Vandekerckhove, J., Bawn, G., Puype, M., Van Damme, J. and Van Montagu, M. (1985) Protein-blotting on Polybrene-coated glass-fiber sheets. A basis for acid hydrolysis and gas-phase sequencing of picomole quantities of protein previously separated on sodium dodecyl sulfate/polyacrylamide gel. *Eur. J. Biochem.* **152**: 9–19.
- Vanfleteren, J. and Peeters, K. (1990) Chromatographic recovery of polypeptides from copper-stained sodium dodecyl sulfate polyacrylamide gels. *J. Biochem. Biophys. Methods* **20**: 227–235.
- van Oostveen, L., Ducret, A. and Aebersold, R. (1997) Colloidal silver staining of electroblotted proteins for high sensitivity peptide mapping by liquid chromatography–electrospray ionization tandem mass spectrometry. *Anal. Biochem.* **247**: 310–318.
- Velculescu, V., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M., Bassett, D., Hieter, P., Vogelstein, B. and Kinzler, K. (1997) Characterization of the yeast transcriptome. *Cell* **88**: 243–251.
- Vera, J. and Rivas, C. (1988) Fluorescent labeling of nitrocellulose-bound proteins at the nanogram level without changes in immunoreactivity. *Anal. Biochem.* **173**: 399–404.
- Wallace, R., Yu, P., Dieckert, J. and Dieckert, J. (1974) Visualization of protein–SDS complexes in polyacrylamide gels by chilling. *Anal. Biochem.* **59**: 24.
- Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T. and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**: 466–469.
- Wirth, P. and Romano, A. (1995) Staining methods in gel electrophoresis, including the use of multiple detection methods. *J. Chromatogr. A* **698**: 123–143.
- Zapolski, E., Gersten, D. and Ledley, R. (1982) [⁵⁹Fe]Ferrous bathophenanthroline sulfonate: a radioactive stain for labeling proteins in situ in polyacrylamide gels. *Anal. Biochem.* **123**: 325–328.
- Zeindl-Eberhart, E., Jungblut, P. and Rabes, H. (1997) A new method to assign immunodetected spots in the complex two-dimensional electrophoresis pattern. *Electrophoresis* **18**: 799–801.

第五章 蛋白质鉴定和磷酸化位点分析的质谱方法

Scott D. Patterson, Ruedi Aebersold and David R. Goodlett

1. 引言

在形成蛋白质组学领域的过程中，质谱已经成为一个有重要意义的推动技术。先进的质谱仪的发明使质谱具有电离多肽、蛋白质及其他一些生物大分子的能力，同时还有极高的灵敏度。更多的科学家可以接触到质谱仪，并进一步扩展了其应用范围。现已开发出的计算机算法可以用质谱数据（肽或碎片离子的质量）通过相关分析途径（correlative-based approach）在蛋白序列数据库中鉴定蛋白质。公共数据库中有越来越多的来自于包括人在内的各个种属的核酸序列信息，这些核酸序列信息与质谱仪的新功能结合起来，显而易见，加快了发现/鉴定蛋白质组成成分的步伐。

本章将讲述质谱在以下分析方面的应用：利用质谱产生的数据鉴定蛋白质的计算手段，蛋白质的翻译后修饰位点鉴定，特别是磷酸化修饰。本章重点将不会放在对质谱数据产生步骤的细节描述上，因为这些内容在最近的许多综述和专著中均有论述（Aebersold and Patterson, 1998; Courchesne and Patterson, 1998; Lamond and Mann, 1997; Moritz *et al.*, 1995）。

1.1 蛋白质鉴定途径

科学的许多方面都是由技术进步来推动的，蛋白质的鉴定就是一个很好的例子。自20世纪80年代中期自动化的蛋白测序仪问世以来，Edman测序法就成为蛋白质鉴定的主要方法（Hewick *et al.*, 1981）。Edman测序法对蛋白质从N末端开始实施逐步的化学降解，并对每一步释放出的衍生氨基酸，用反相HPLC分离，紫外检测器检测，根据其保留时间与标准氨基酸的保留时间相比较来鉴定。许多蛋白质的N末端氨基被修饰（封闭），与Edman化学试剂不反应，所以分析完整蛋白时没有数据产生（Brown, 1979; Brown and Robert, 1976）。凝胶电泳是蛋白质分离的最佳选择方法，应用二维凝胶电泳可以对复杂混合物的分离达到非常高的分辨率。将凝胶分离的蛋白质电转印到PVDF膜上，可以直接用于Edman测序化学反应，电转印方法使电泳和Edman测序方法直接联系起来（Aebersold *et al.*, 1986; Matsudaira, 1987; Pluskal *et al.*, 1986）。到20世纪80年代末，开始用各种化学或酶的方法降解凝胶分离后转印到各种膜上的微量蛋白，降解产物经分离及纯化后用于Edman测序分析（Aebersold *et al.*, 1986, 1987; Vandekerckhove *et al.*, 1985）。如果蛋白质的N末端封闭，不适用于Edman测序，上述方法可以提供机会选择蛋白质中间部分肽段用于氨基酸序列分析，而且中间肽段测序的序列覆盖率明显高于单独使用N末端序列法。使用蛋白质不同区域的氨基酸序列设计寡核苷酸引物也便于用PCR反应进行基因克隆（Tempst *et al.*, 1990）。在90年代初期，开始致力于凝胶分离蛋白质鉴定方法的改进，例如转膜和水解过程中获取蛋白和释

放多肽等步骤 (Abersold and Leavitt, 1990; Patterson, 1994;), 以及提高 Edman 测序仪的灵敏度。Edman 测序法是对氨基酸序列从头测定, 因此在蛋白质鉴定过程中不需要从实验数据到氨基酸序列的相关分析, 而从质谱途径鉴定蛋白质需要从实验数据到氨基酸序列相关分析的辅助, 这也是本章的主题内容。因此, Edman 测序法仍将在种属基因组未明了的蛋白质鉴定中发挥重要作用。

以对全部人基因组测序为目标的人类基因组计划的出现, 促进了对其他许多种属基因组的测序。通过政府及私人公司两方面的努力, 已完成了许多与临床和基础研究相关的物种的基因组测序, 例如 21 种细菌 [见 Fraser and Fleischmann (1997) 的综述和基因组研究所微生物数据网站 (The Institute for Genomics Research (TIGR) Microbial Database web site: <http://www.tigr.org/tdb/mdb.html>]、酿酒酵母 (*Saccharomyces cerevisiae*)、秀丽新小杆线虫 (*Caenorhabditis elegans*)、果蝇 (*Drosophila melanogaster*) 的全基因测序已经完成。公共数据库中已经公布了超过 200 万个人和 100 万个鼠的表达序列标志 (expressed sequence tag, EST: 从不同组织或细胞中分离出的 mRNA 反转录成的 cDNA 序列, 大约有 300~500 个碱基对长度), 政府和私人公司的 EST 序列都已经涵盖了至少 3 倍的人基因组全序列。已经有了能够利用质谱测定的蛋白质一级结构数据在蛋白质序列数据库 (从核酸序列翻译成的蛋白质序列) 中迅速鉴定蛋白质的软件算法。这种蛋白质鉴定新技术在 90 年代中期建立, 伴随着技术发展的新浪潮, 还建立了用于 Edman 测序的凝胶分离蛋白质产生肽段的纯化、分离技术 (Lamond and Mann, 1997; Patterson and Abersold, 1995)。质谱及其接口技术也在这一阶段及后来的时期得到发展, 例如纳喷技术 (nanospray) (Wilm and Mann, 1994, 1996) 和微流控技术 (microfabricated devices) (Figeys *et al.*, 1998), 可参见相关综述 (Figeys and Abersold, 1998)。这些技术上的进步已经汇聚成了一个成熟、自动化、灵敏并且快速的技术方法, 极大地提高了鉴定蛋白质的能力, 并构成了正在形成中的蛋白质组学领域的基础。

2. 质谱基础知识

通过质谱数据与序列数据库相关分析的方法鉴定蛋白质的基础在于能够得到一些限制参数, 利用这些参数可以从数据库中所有的序列中特异地识别与这些参数相关的序列, 这些参数包括: 蛋白质的氨基酸组成、氨基酸序列、蛋白质和肽段的质量以及肽碎片的质量。参数可以单独使用或与其他参数结合使用, 其中分子的精确质量特别具有吸引力, 因为这一参数具有高度的限制性, 并且能够利用质谱仪 (MS) 快速、灵敏、高精度地测定。MS 只限于检测能够形成离子并且离子能够被传送入真空系统的分子的质量, 随着 80 年代末两种软电离技术: 电喷雾电离 (ESI) (Fenn *et al.*, 1989) 和基质辅助激光解吸/电离 (MALDI) (Karas and Hillenkamp, 1998) 的发明, 质谱变得更适用于分析生物大分子聚合物, 蛋白质、核酸和糖类。这两种方法之所以被称为“软”电离技术, 是因为它们在离子化过程中不会 (最多只部分) 破坏分子结构, 因此, 就能实现分子质量大于 10 000 质量单位 (u) 的生物大分子的质量分析。

质谱仪由三部分组成: 离子源、质量分析器、检测器。质谱仪在真空状态下分析离子的质荷比 (m/z), 有多种不同的离子源与质量分析器 (检测器) 的结构组合。三种

最常用于蛋白质和多肽分析的商业化仪器类型是：MALDI-飞行时间（TOF）质谱仪、ESI-三级四极杆（TQ）质谱仪和ESI离子阱质谱仪（IT），这三种类型的质谱仪将在下文详述。

2.1 MALDI-MS

线性 MALDI-MS: MALDI-MS 质谱仪恐怕是概念上和设计上都最简单的质谱仪，分子（分析物）经 MALDI 方式离子化，离子的质荷比（ m/z ）由飞行时间质量分析器检测（见图 5.1）。测量的质量精确度与 m/z 有关，25ku 以下可达到约 0.01% ~ 0.05% 的精确度，到 300ku 精确度达到 $\pm 0.05\% \sim 0.3\%$ 。MALDI 也能分析溶解在磷酸盐，Tris 等生物缓冲液中的分析样品，或含有少量脲、非离子变性剂及一些碱金属盐的样品溶液。

蛋白质和多肽的 MALDI 分析包括以下一系列步骤：

- (1) 被分析物与基质（小分子芳香化合物）混合。基质通常溶解在酸性有机溶剂中，以超过样品 1000 倍的量与样品混合，将样品和基质加到金属片/靶上，溶剂在空气中挥发后，形成了样品-基质共结晶。
- (2) 将加有样品-基质混合结晶的靶送入质谱仪的真空室（ $10^{-5} \sim 10^{-8}$ Torr）。
- (3) 在金属片/靶上加 +20 ~ 30kV 的高电压（产生正离子），同时有短的激光脉冲照射在干燥的样品上。
- (4) 基质晶体吸收特定波长的激光能量（紫外激光 337nm，红外激光 2.94 μ m），又以热的形式将能量释放，引了解吸附。这一迅速的放热过程又使基质晶体升华，基质和样品分子气化进入质谱仪的气相。
- (5) 通过在气相中发生质子化/去质子化，附着阳离子/脱离阳离子，或氧化/还原等过程实现离子化。产生的离子从靶表面（保持 +20 ~ 30kV 高电压）被推斥并加速进入一系列的离子透镜（保持接地电压），离子透镜聚焦离子进入飞行时间质量分析器的无场漂移区（50 ~ 300cm）。检测器安装在无场漂移区末端，通过飞行管道的离子被检测器记录。
- (6) MALDI 离子化过程送入质量分析器的离子实质上具有相同的终动能。在动能一定时，离子的速度与离子的质荷比 m/z 成反比。因此，在线性检测器中，离子到达飞行管道另一端检测器的时间反映了被检测离子的 m/z 。
- (7) 离子通过无场漂移区的飞行时间由激光脉冲触发检测器记录。因此，每一离子化过程产生的每一批离子到达检测器的飞行时间都被记录下，并被换算为质荷比 m/z （Carr and Annan, 1997）。

离子越小，飞行速度越快，因此质荷比 m/z 可用已知质量和电荷状态的化合物的飞行时间作校正并用经验公式计算出来（通常用一个基质离子和一个蛋白质或多肽离子作校正，被分析物的质量包括在此质量范围内）。离子动能（ E ）与其质量（ m ）和速度（ v ）的关系可被表述为 $E = \frac{1}{2}mv^2$ ，因此，对于一个单电荷离子，其飞行时间（由速度决定）与分子质量的平方根成反比。

MALDI 离子化过程本身存在一个问题，即离子化过程中动能有微小的分散。由单个激光脉冲产生的离子的动能的微小差别削弱了 MALDI-MS 的分辨力。多肽的质量分

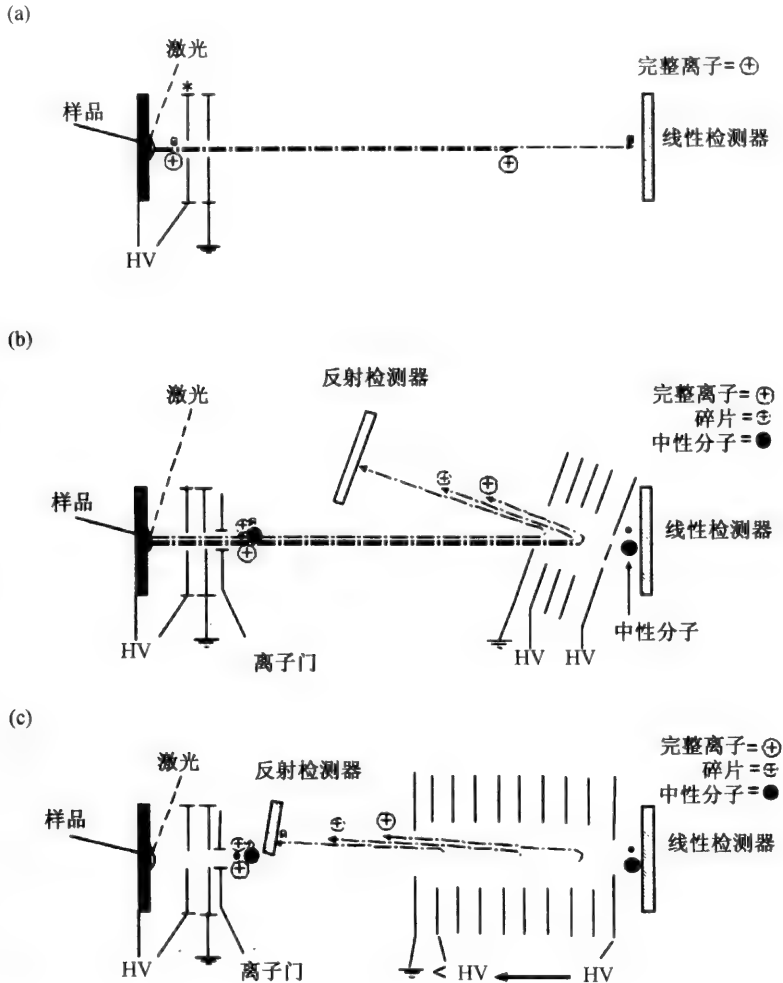


图 5.1 基质辅助激光解吸/电离质谱仪

(a)MALDI-MS;(b)反射 MALDI-MS;(c)环形电场反射 MALDI-MS。具体细节见文中。

分辨率 ($m/\Delta m$) 通常为 40~800 FWHM (峰高一半处的峰宽), 蛋白质的质量分辨率通常为 50~400 FWHM, 这样的质量分辨率即使是多肽的同位素分布也不能测定。初始能量的分散是与质量有关的, 离子质量越大, 峰越宽。现已发展了两种技术以提高 MALDI-MS 的质量分辨率: 使用离子镜 (ion mirror) / (反射器 reflectron) 和 (或) 采用时间延迟聚焦 (time-lag-focusing) 处理 (图 5.1b 和 5.1c)。如图所示, 离子镜/反射器在一个特定区域内将离子反射回检测器。这通过以下步骤实现: 在飞行管末端放置的反射器上施加高电压, 其电压值比离子源的电压值稍高, 进入反射器的离子在电场作用下方向偏转, 直至停止飞行, 然后在反射器内被重新加速飞出反射器进入第二检测器 (图 5.1b)。反射器充当了一个能量聚焦装置, 因为动能稍小的离子 (速度慢), 进入反

*译者注: 这条线应该去掉; 原文插图有误。

射器的距离短, 折返较早, 动能稍大的离子进入反射器的距离长, 折返较晚, 所以速度慢的离子赶上了速度快的离子。反射器的应用减少了离子初始动能的分散, 因此提高了质量分辨率。

第二种用来修正 MALDI 离子化过程中初始能量分散的技术是在离子源内实现的, 由 Wiley 和 McLaren 在 1953 年提出 (Wiley and McLaren, 1953), 称为时间延迟聚焦 (time-lag-focusing), 最近被称为延迟提取 (delayed extraction) (Vestal *et al.*, 1995)。在图 5.1b 和 c 中看到, 相对于连续离子提取 (图 5.1a), 延迟提取是在样品靶和接地的连续离子提取透镜之间又加了二级离子透镜 (施加电压), 延迟提取时, MALDI 源在无电场区域产生离子, 在施加提取电压使离子加速进入飞行管道之前, 离子可以四处分散。这样显著降低了离子的能量分散, 同时也限制了峰的展宽, 峰的展宽通常是由于在连续离子提取模式下源内离子碰撞产生不稳定的解离所致。延迟提取使线性 MALDI 质谱仪分析多肽的质量分辨率达到 2000~4000, 使反射 MALDI 质谱仪分析多肽的质量分辨率达到 3000~6000 (Brown and Lennon, 1995; Vestal *et al.*, 1995)。

PSD-MALDI-MS: 为了在 MALDI-MS 仪器上获取一级结构信息, 必须选择分析某一离子及其碎片离子的质量。这一过程发生在源内离子化之后, 所以被称为源后衰变 (post source decay, PSD) (Chaurand *et al.* 1999)。对多肽进行 PSD 分析时, 其碎片离子主要是沿多肽骨架断裂形成, 产生的一系列碎片离子在理论上包含着这一肽段的氨基酸序列信息 (Kaufmann *et al.*, 1994)。离子源内产生的离子被加速后在飞行过程中产生的碎片离子是亚稳离子, 所有的碎片离子都具有与其前体离子相同的速度, 但只具有前体离子动能的一部分。所有的离子, 包括前体离子和碎片离子, 由于具有相同的表观质荷比, 都将在同一时间到达线性检测器, 所以在线性 TOF 分析器中, 不能分辨碎片离子与前体离子。但是使用离子镜 (反射器) 可以分辨碎片离子与前体离子的动能差异, 现在有两种反射器: 双程 (dual-stage) 反射器和环场 (curved-field) 反射器。双程反射器只能在有限的动能范围内聚焦离子, 为了使较宽动能范围的离子都被聚焦, 反射器电压要逐步减少, 连续降 7~14 步, 聚焦动能越来越小的离子, 收集到最后一张谱图时, 反射器电压大致是初始电压的 5%~10%。由数据系统整合出连续、完整的碎片离子谱图, 并根据每一反射电压下的碎片离子校正谱图进行校正。环场反射器由 Cornish 和 Cotter (1994) 最先提出, 能在一次试验中分辨出所有的碎片离子。环场反射器是由一系列电压逐渐减小的离子透镜组成的长反射器 (图 5.1c), 因此聚焦不同质荷比的碎片离子所需的电压一直存在, 但是在空间上是可以分辨的。在同一激光脉冲下如果产生多个肽段离子, 在没有质量过滤器时, 不能区分不同肽段产生的碎片离子, 也不能区分碎片离子与未断裂的肽段离子。因此, 装备有以上任一种反射器的质谱仪都安装了一个离子门 (ion gate), 可以在低分辨率情况下 ($\pm 2.5\%$) (图 5.1b、c) 选择前体离子/碎片离子。所以, 在一个未经分离的肽混合物中, 可以选择多个肽段得到其序列特异的碎片离子信息。

2.2 ESI-MS

在 ESI 源中, 含有被分析样品 (多肽/蛋白质) 的溶液流经一个细细的进样针, 针头上加高电压 (+1000~5000V) 用来产生正离子, 见图 5.2a。高电压导致样品液流分

散为呈喷雾状的带高电荷的微小液滴，质谱仪入口端的有孔平板上加有 +100~1000V 的低电压，引导离子通过入口 (orifice)，这一入口是离子源和质谱仪的连接处，离子源处于大气压环境，质谱仪处于真空系统内。当液滴从针尖通往入口时，在定向惰性气流 (有时加热) 的作用下溶剂蒸发，使液滴缩小，液滴电荷密度增加，最终导致液滴爆裂，离子从液相中释放出来进入气相 (Carr and Annan, 1997)。离子通过入口后，在一个只加有射频场 (radio frequency, RF) 的四极杆控制下进入质量分析器，可以是四极杆、四极杆-飞行时间、或离子阱质量分析器。ESI 产生的多肽离子的特征是带多电荷，电荷数与肽分子中可电离的基团数目有关。如果用 ESI-MS 正离子模式分析胰酶酶切肽段，大多数肽段都会带至少两个电荷，一个在 N 末端的 NH_2 基团上，另一个在胰酶酶切位点，肽段 C 末端碱性氨基酸的侧链上。

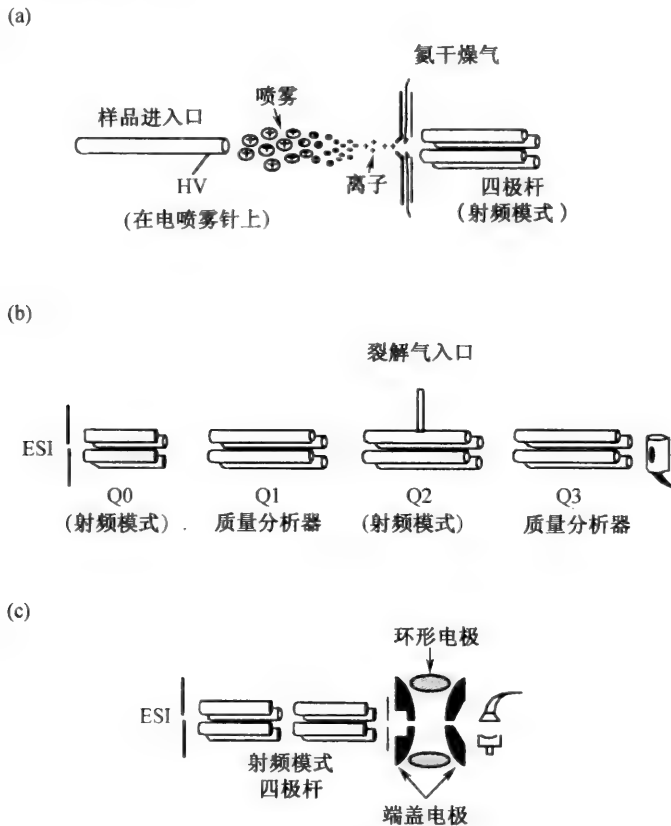


图 5.2 电喷雾离子化质谱仪

(a) ESI-MS; (b) ESI-TQ-MS; (c) ESI-IT-MS。具体细节见文中。

ESI-三级四极质谱仪 如其名所述，三级四极质谱仪由三套四极杆组成 (图 5.2b)。四极杆滤质器既可以传送混合物中所有的离子 (以射频 RF-only 方式工作)，又可以充当质量过滤器，只允许特定质荷比的离子通过。四极杆由 4 个平行的圆棒构成，相向的一对圆棒上施加电极相反的直流和交流电压。当离子漂移通过四极杆圆棒的中间区域时，改变四极杆上施加的电压，使具有特定 m/z 质荷比的离子才能通过，所有其

他质荷比的离子改变了线性飞行路径, 从分析器中消失。如果四极杆上施加的电压连续变化(扫描), 由通道-电子倍增器记录通过滤质器的离子数目, 并与特定的离子 m/z 质荷比建立函数关系, 就得到一张质谱图。四极质量分析器通常可分析的质量范围是 2500~4000u, 因为记录的是 m/z , 而不是多肽质量本身, 而且 ESI 通常产生带多电荷的肽离子, 所以, 使用 ESI 源的四极质量分析器能用来分析远远超出其质量分析范围的蛋白质。

在三级四极质谱中, 第一级 Q1 和第三级 Q3 四极杆是质量过滤器, 第二级四极杆 Q2 仅施加射频电压, 充当产生碎片离子的碰撞室, 从 Q1 传送来的肽离子在碰撞室内经重的惰性气体如 Ar 和 N₂ 的碰撞诱导, 产生正离子, 这一过程称为碰撞诱导解离 (collision-induced dissociation, CID)。

三级四极质谱仪有 5 种工作模式:

- (1) MS 模式 这种模式用于分析未解离的离子质量。扫描 Q1, Q2 中未引入裂解气体, Q3 为射频电场模式。也可以使 Q1 处于射频模式, 扫描 Q3 电场实现质量测定。
- (2) MS/MS 模式 (也称为产物离子扫描法, product ion scan) 在一个给定的时间点, Q1 设定为仅传输某一选定质荷比的离子, 此离子进入 Q2 后, 经碰撞诱导解离, 产生的碎片离子通过扫描 Q3 进行检测。这样得到与 Q1 选择的初始肽段离子相关的碎片离子质谱图。Q1 选择母离子, 在 Q2 内进行 CID, Q3 分析碎片离子, 这一分析过程不断循环, 用以选择分析不同质量的肽段离子。实际上, 某些仪器如 TSQ 7000 (Finnigan, MAT, San Jose, CA) 可以设定程序自动在 MS 和 MS/MS 模式间切换, 将 MS 检测到的离子进行 CID 分析, 并记录碎片离子的谱图, 整个过程不需要用户介入 [Ducret *et al.*, 1998; Stahl *et al.*, 1996; ICL procedure on our web site (<http://depts.washington.edu/~ruedilab/aegersold.html>)]。这一过程称为数据依赖的 CID (data-dependent CID)。产物离子扫描法可用来测定肽段的氨基酸序列。
- (3) 中性丢失扫描模式 (neutral loss scan mode) 此模式中 Q1 和 Q3 的电压同步扫描, 但保持一个特定 m/z 值的电压差 (offset)。Q2 是碰撞室, Q1 和 Q3 的电压差值与 Q2 内碰撞消除的一个特定中性分子的质量一致。因此, 在离子混合物中, 只有碰撞裂解后能丢失这一特定基团的离子才会被 Q3 传输到检测器。例如, 在磷酸肽与非磷酸肽混合物中, 通过对 50 m/z 的 H₃PO^{*} 的中性丢失扫描, 可以检测出磷酸肽的双电荷峰 (Covey *et al.*, 1991)。在离子被传送到检测器时, 已经知道 Q1 和 Q3 上各自施加的特定电压, 所以可以确定发生中性丢失的肽段质量。
- (4) 前体离子 (母离子) 扫描 (precursor/parent ion scanning) 此模式中扫描 Q1 电压, Q3 设定为只传输某一特定质荷比离子, Q2 为碰撞室。与中性丢失扫描相似, 前体离子扫描检测离子混合物中在 Q2 碰撞室丢失特定基团的离子。与中性丢失扫描不同的是, 前体离子扫描模式直接检测断裂的基团 (报告离子,

*译者注: 原文有误; 应为 49 m/z 的 H₃PO₄。

reporter ion)。检测器检测到报告离子时, 已知 Q1 的电压, 所以可以确定产生这一特定碎片的前体离子的质量。因此前体离子扫描模式被用来检测肽混合物中某些含有特定结构特征的肽段。

- (5) 源内 CID (in-source CID) 裂解发生在 ESI 源内由大气压气体碰撞产生的高压区内。可以选择碰撞条件使肽链骨架断裂 (Katta *et al.*, 1991) 或使一些相对不稳定的基团如磷酸基团选择性断裂 (Huddleston *et al.*, 1993)。所带电荷数大于 1 的肽段离子更容易断裂, 因为在这些离子上分配的动能远大于仅仅将这些离子聚焦到 Q1 分析器所需的动能。这一方法提供了在三级四极质谱仪中实现 MS/MS/MS 分析的机会, 及在单级四极质谱仪中实现 MS/MS 分析的机会 (Loo *et al.*, 1990)。以上任一种扫描类型都适用于源内 CID 产生的离子。

为了鉴定蛋白及对肽段从头测序, 仪器通常以 MS/MS 模式运行。对选择离子施加的碰撞能量可以通过改变碰撞气体的压力而变化, 如果时间允许, 可以对裂解过程实现很好的控制。在数据依赖操作模式 (data-dependent operation) 中, 质谱仪由程序控制, 对 MS 模式检测到的超过一预先设定的离子流阈值的离子自动进行碰撞诱导解离, 碰撞能量匀速变化 (ramped), 以期在不同碰撞能量下得到的谱图中至少得到一张最佳裂解碎片谱 (Stahl *et al.*, 1996)。但是, 当样品量极少, 样品溶液通过微量或纳喷 (nanospray) 装置引入源内, 且样品未脱盐时, 得所质谱图的化学背景和杂质峰占据主导地位 (Wilm and Mann, 1996)。因此, 为了在这样的混合物中找到肽段离子, 质谱仪可以前体离子扫描模式运行, 将 Q3 分析器设定为 86 m/z , 即亮氨酸 Leu 和异亮氨酸 Ile 的亚氨离子 (immonium ion) 质量, 通常在许多肽段序列中都存在这两种氨基酸 (Wilm *et al.*, 1996)。

ESI-离子阱质谱仪 离子阱质谱仪是一种串联质谱仪, 在分析前先将离子聚集, 储存。离子阱是仪器的核心部分, 既可以作为质量分析器, 又可以作为碰撞室。四极离子阱使用射频方式工作的四极杆引导离子从离子源进入离子阱。离子阱由两种电极构成, 一个环形电极, 两个端盖 (end-cap) 电极 (图 5.2c), 离子进出离子阱都经过端盖电极上的小孔。环形电极上施加的高射频 (RF) 电压将离子捕获在其产生的电场中, 此射频电压的大小决定了被捕获离子的频率及运动。大于某一特定 m/z 值的离子留在离子阱中, 被残留氦气碰撞冷却, 使离子的正弦振荡坍塌, 在很短时间内, 运动轨道由大变小。通过线性变化 (ramp) RF 电压, 同时对端盖电极施以一小电压, 使连续质荷比的离子共振发射, 得到完整的质谱图。共振发射是指在离子阱中变得不稳定的离子被发射到端盖电极的同轴方向, 穿过端盖电极并被电子倍增检测器的转换电极检测, 转换电极采用离轴 (off-axis) 位置, 是为了除去中性物质的背景 (Jonscher and Yates, 1997; Schwartz and Jardine, 1996)。

通过离子捕获和离子选择发射的过程, 具特定质荷比的离子可以从离子阱中分离出来。如果这一离子在离子阱中裂解, 而且碎片离子也被共振发射, 就得到了 MS/MS 串联质谱图。如果碎片离子中除某一特定质荷比的离子外, 其他所有碎片离子均被共振发射, 这一碎片离子可以重复进行裂解/共振发射过程, 这样就得到了多级串联质谱 (MS^n)。曾有文献报告做到 MS^{12} (Louris *et al.*, 1990), 而 MS^3 已成为常规实验, 特别是在糖肽的结构分析中 (Reinhold *et al.*, 1995)。在离子阱中对某一特定 m/z 比值离

子的分离和裂解包括：共振发射所有质荷比大于和小于这一比值的所有离子，这样在离子阱中只剩下选定 m/z 比值的离子（被分离的离子），这时改变电极电压使离子能量增加，这样离子就会与残留的中性气体（He）碰撞而裂解，产生的碎片离子用上文所述的扫描方式测定。被捕获离子上所加的能量大小可通过改变参数而调节，在某些仪器中这一参数被称为相对碰撞能量（relative collision energy, RCE）。如上文 PSD-MALDI-MS 所述，单电荷离子的裂解比双电荷离子裂解需要更多的能量。在某些离子阱实验中，RCE 可以手动调节产生非常好的效果（Davis and Lee, 1997）。离子阱质谱仪的其他一些可以改变并影响仪器性能的参数包括在任何一次捕获中允许进入离子阱的离子数目，以及这种捕获超过多少次后信号将被平均以增加信噪比。如果有太多 m/z 比值非常接近的离子被捕获，会观察到被称为空间电荷效应（space charging）的现象。由于 m/z 比值非常接近的离子之间运动轨道相互影响，所以空间电荷效应降低了实验测量质量的精确度。因此在设置参数时必须注意不会对数据起负面影响（Courchesne *et al.*, 1998）。商业化的离子阱质谱仪的仪器控制软件支持数据依赖裂解（data-dependent fragmentation）功能，因此非常适用于 LC-MS/MS 分析。

3. 以质谱为基础的相关鉴定策略

3.1 肽质量检索 (peptide-mass searching)

MACDI-MS 质谱仪商品化后，表现出其分析混合物中肽段质量的能力，尤其是对凝胶分离蛋白质的酶切肽段的分析能力，很快有许多研究小组开始研发蛋白质鉴定的软件算法，这些算法都是基于用实验测得的肽质量与蛋白序列数据库中蛋白质的肽质量计算值进行相关性分析（图 5.3）（Henzel *et al.*, 1993; James *et al.*, 1993; Mann *et al.*, 1993; Pappin *et al.*, 1993; Yates *et al.*, 1993）。这一途径已经成为凝胶分离蛋白质鉴

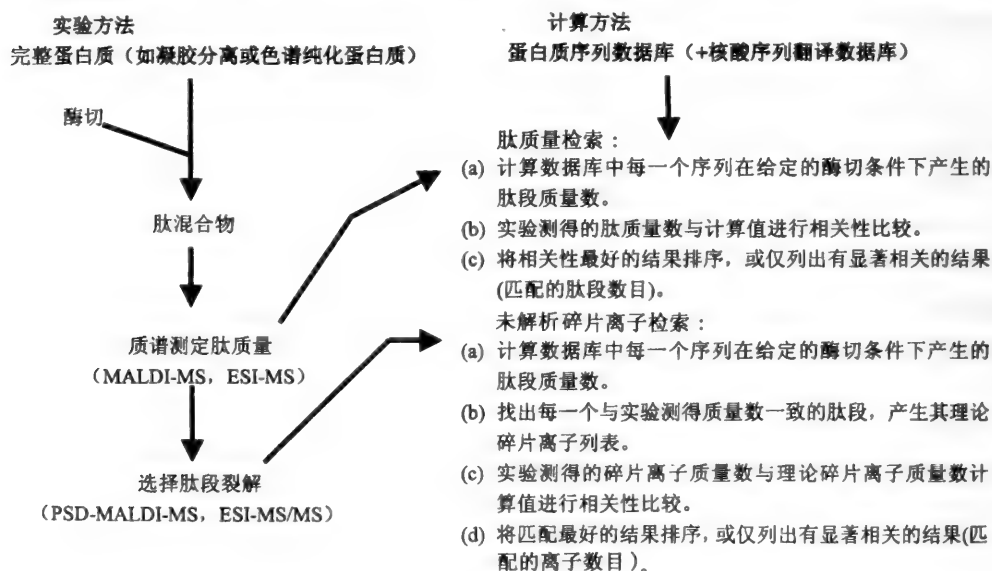


图 5.3 质谱鉴定蛋白质

定的最常用手段 (Aebersold and Patterson, 1998)。

鉴定数据库中已有序列的蛋白质的算法基于以下一些理念。

- (1) 肽段是由切点专一的试剂水解蛋白质产生的, 通常用已知切点专一的蛋白酶。
- (2) 这些肽段的质量由 MALDI-MS 或 ESI-MS 精确测定。
- (3) 计算蛋白序列数据库中的每一序列被实验中所用水解试剂水解后产生的理论肽段的质量。
- (4) 用试验得到的肽质量与数据库中的肽质量计算值相匹配, 计算得分或排序值。

许多做开创性工作的小组及后来加入进来的研究小组已将他们的检索软件放在因特网上供大家使用, 并有例子予以说明 (见表 5.1)。这种检索途径有一个明显的限制, 就是被鉴定的蛋白质必须已经在序列数据库 (翻译的核酸序列或蛋白序列数据库) 中存在。因此, 这一手段非常适用那些基因特性非常清楚的物种, 特别是全基因组已知的物种 (例如见 TIGR 数据库, <http://www.tigr.org/tdb>), 亦适用那些已经建立详尽的蛋白质或 cDNA 序列数据库的物种。肽质量检索鉴定蛋白质方法的依据是实验中获得的蛋白质的多个肽段的质量数据与同一蛋白质肽段质量计算值之间的相关性。因此, 这一技术既不适用于检索表达序列标签数据库 (EST) 翻译序列, 也不适用于鉴定 2 个以上的蛋白混合物。但是, 对某些蛋白质可进行种属间鉴定, 已经有了为这一特殊目的而编写的程序 (Wilkins *et al.*, 1998)。EST 仅代表蛋白编码序列的一部分, 其长度不足以产生供肽质量鉴定实验所需的足够的肽数目。如果水解蛋白质混合物, 只产生其中部分特定肽质量的蛋白质的鉴定就不明显。但如果每一个观察到的肽段都能与序列数据库中的某一蛋白精确相关, 混合物中蛋白质的鉴定仍然可行。方法是从测得的肽质量中去掉与已被鉴定的蛋白质相关的所有肽段, 用剩余的质量数检索数据库, 将这一步骤重复循环进行。但是, 由于下面将要提到的一些原因, 所有检测到的肽质量都与蛋白质的序列相

表 5.1 用质谱数据鉴定蛋白质的检索工具网址

资源	网址	特点
CBRG, ETH-Zurich, Switzerland	cbrg.inf.ethz.ch/MassSearch.html	肽质量检索软件
EMBL, Heidelberg, Germany	www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html	肽质量和碎片离子检索软件
ExPasy	www.expasy.ch/tools/#proteome	肽质量和碎片离子检索软件
Mascot	www.matrix-science.com/cgi/index.pl?page=/home.html	肽质量和碎片离子检索软件
Rockefeller University New York, USA	prowl.rockefeller.edu/	肽质量和碎片离子检索软件
SEQNET, Daresbury, UK	www.seqnet.dl.ac.uk/Bioinformatics/welapp/mowse	肽质量和碎片离子检索软件
University of California, San Francisco, USA	prospector.ucsf.edu	肽质量检索软件 (MS-Fit) 和碎片离子检索软件 (MS-Tag)
University of Washington	thompson.mbt.washington.edu/sequest/	碎片离子检索软件 SEQUEST 应用指导

符的情况很少发生,这不仅使混合物中各组分的鉴定变得复杂,还使单一的纯蛋白质的鉴定也复杂化,因此,了解那些未匹配的肽质量数的性质非常重要。下面将讲述与匹配正确与否有关的这些表现伪质量 (spurious masses) 存在的可能原因 (Jensen *et al.*, 1997, 1998; Patterson and Abersold, 1995)。

- (1) 蛋白质被正确鉴定,但是由于翻译后修饰或人为修饰及翻译后加工(如 N 或 C 末端加工)产生了另外的质量。尽管某些质量的差异可能与这些修饰情况相符,但除非经实验证明,否则仅仅是假设。
- (2) 蛋白质被正确鉴定,但存在非特异性的蛋白水解,或有杂蛋白酶存在。通过排除酶切位点专一的假设,确定候选蛋白质是否会产生具有那些质量数的肽段,这样可以接受这种可能性。
- (3) 蛋白质被正确鉴定,但有部分杂质蛋白质混在其中,如 2-DE 胶上的蛋白点可能由多个蛋白质组成。如果有足够多的未匹配肽质量数,就可以用这些数据单独进行肽质量检索,以证实存在另外的蛋白质。当然要做酶自切的对照实验,以鉴别所有的酶自切产物。
- (4) 被鉴定的蛋白质或许是数据库中已登录蛋白的同源序列,或者是其序列的剪切变异体。某些检索程序允许实现跨种间检索,如果能得到另外确实的数据,这一方法还是很有用的。因为从基因特征不清楚的种属得到的蛋白质也许可以在基因特征清楚的种属的蛋白序列库中得到匹配鉴定 (Cordwell *et al.*, 1995)。
- (5) 蛋白质的鉴定结果是假阳性。如果没有其他数据,特别是实验数据的质量数精确度不高时,这一糟糕的结果很难被证实或证伪。如果所用检索程序是用鉴定结果的得分值排序,而所得检索结果得分值又比较低时,该结果的证实就更加困难了。在某些例子中,通过比较检索结果中排序第一和第二位(及后面)的得分值之差,得到进一步确认。此蛋白质也可能在蛋白数据库中不存在。

用肽质量数据鉴定蛋白质所控制的实验参数中,最具决定性的参数是肽质量测量的精确度 (Fenyo *et al.*, 1998; Jensen *et al.*, 1996a, 1997)。质量精确度越高,结果判定越可信。另一决定性参数是所用酶(或化学试剂)的专一性。酶越纯,专一性越高,检索结果越可信。最常用的酶是胰蛋白酶。但是要注意的是,即便是高纯度的胰蛋白酶也会在非 Lys 或 Arg 的氨基酸残基 C 端酶切(其后不是 Pro)。所有蛋白酶都存在的问题是,如果蛋白序列中有 2 个或更多个连续氨基酸残基是酶的切点,蛋白酶会对底物酶切不完全,留下漏切的位点和不完全的末端,许多检索程序把漏切的位点数目作为输入的一个检索参数。内肽酶 LysC 是另一个具高度专一性的酶,此酶产生的自切产物比胰蛋白酶少,但同样也会有漏切位点,也会在非 Lys 位点酶切(其后不是 Pro)。同样也要注意,不是所有的蛋白质,不论是否经凝胶分离,都适用于胰蛋白酶或其他专一性蛋白酶切,但是,凝胶分离的蛋白质,由于其处于变性状态,通常酶切十分有效。

蛋白酶的性能很难改变,与之不同的是肽质量测量精确度依赖于所使用的仪器及其操作。最佳的肽质量检索应该用内标校正的单同位素质量。在 MALDI-MS 质谱仪中,通常只有安装时间延迟聚焦/延迟提取(见前文)的仪器,能够达到肽质量的同位素分布分辨率。这种仪器达到 5ppm 的质量准确度,可确定单同位素质量,并在肽质量检索程序中设立非常小的质量误差,这样高精度的数据减少了错误匹配的可能性 (Jensen *et*

al., 1996a, 1997; Takach *et al.*, 1997)。如果没有安装延迟提取选项, 质谱仪应该用内标校正, 且内标质量范围应包含被检测肽段的质量。没有延迟提取装置的仪器不能分辨同位素峰, 仅得到平均分子质量。单同位素质量比平均分子质量更精确, 因为单同位素质量是由单一同位素 C12 决定的, 而平均质量是由 C12 和不同数目的同位素 C13 质量的平均值决定的 (其他元素的同位素相似)。

为了进一步增加肽质量的可信度, 开发了“正交法”对单个肽段的氨基酸序列或组成提供额外的, 或限制性的信息。这种方法中, 样品除用于原始质量测量外, 还分出一部分用于正交分析。正交分析信息限制了用肽质量在数据库中检索鉴定蛋白质产生的多个肽段与同一质量匹配的情况。正交方法可分为以下几类。

- (1) 特异位点的化学修饰 甲酯化, 在每个羧基基团上增加 14 个质量单位 [如酸性氨基酸残基 (Asp, Glu) 侧链或肽的 C 末端 (Pappin *et al.*, 1995)]; 碘化反应, 在酪氨酸上增加 126 个质量单位 (Craig *et al.*, 1995); 及用 1:1 混合的乙酰氨/氘代乙酰氨在每个半胱氨残基上进行同位素标记 (Sechi and Chait, 1998)。
- (2) 测定肽段的部分氨基酸组成 有两种方法: 如果肽段中的可交换氢在氘溶液中被交换, 那么每一个氨基酸有其特定的交换数目, 每个残基为 0 至 5 个质量单位。因此, 肽质量增加的总数就反映了肽的氨基酸组成 (James *et al.*, 1994)。肽段部分氨基酸组成也可通过鉴别 MS/MS 质谱图中的氨离子 (immonium ion) 峰来确定。
- (3) 鉴定 N 末端氨基酸残基 混合物中肽段的 N 末端氨基酸确定可用化学法, 一步 Edman 反应 (Jensen *et al.*, 1996b); 或酶法, 用氨肽酶除去末端氨基酸残基 (Woods *et al.*, 1995)。
- (4) 肽内不同酶切位点的鉴定 为了确定段肽中一个特定残基的存在及相对位置, 有人采用酶切位点不同的酶对一级水解产物进行二次水解 (次级水解) (Pappin *et al.*, 1995)。同样也可把一份样品分为两等分用切点不同的酶平行水解 (James *et al.*, 1994)。
- (5) 鉴定 C 末端残基 与 N 末端残基鉴定相似, 用羧肽酶除去 C 末端氨基酸残基, 某些情况下, 会有一个或多个额外的残基被水解了 (Patterson *et al.*, 1995; Woods *et al.*, 1995)。但是当使用高度专一性的酶时, 除去一个残基不能提供太多信息。

Fenyo *et al.* (1998) 做了出色的研究来评价正交信息在肽质量检索中的好处, 检索中使用的数据库是酵母全基因组数据库, 读者可以参考这篇文章以了解此方法的优点的详细内容。

3.2 未解析碎片离子的检索

正如数据库中检索肽质量一样, 数据库同样可以检索一个单个肽段及其碎片离子的质量。肽段质量同它的碎片离子质量一起提供了非常具有鉴别力的判断标准, 可用于在序列数据库中检索, 包括由表达序列标签翻译的序列数据库 (图 5.3)。这一手段已被广泛用于凝胶分离蛋白质酶水解后产生的肽段鉴定 (Aebersold and Patterson,

1998; Lamond and Mann, 1997; Patterson, 1997)。肽段通常用上文所述的一些方法, 在质谱仪气相中被诱导产生碎片离子, 例如 PSD-MALDI-MS, 或三级四极质谱及离子阱中的 CID。在这些仪器中, 肽段离子断裂的方式具有序列特异性, 主要是沿肽骨架的肽键断裂, 见图 5.4 所示。如果肽离子的正电荷保留在碎片离子的 N 末端, 此离子被称为 *b* 系列离子, 用下标数字代表这一碎片离子中的氨基酸残基数目, 从 N 末端起为 1。因此, *b* 离子代表 C 末端缺失, N 末端完整的碎片离子。如果电荷保留在 C 末端, 离子被称为 *y* 系列离子 (与 *b* 系列离子一样有下标数字, 但从 C 末端数起)。因此 *y* 离子代表 N 末端缺失, C 末端完整的碎片离子 (Bieman 命名法, 1990)。CID 谱图是由几千种独立的裂解过程产生数据的混合图, 即 *b* 系列和 *y* 系列离子的混合图。谱图中也存在一些其他离子: 氨基酸侧链 (Gln, Lys, Arg) 中性丢失氨, 产生少 17 个质量数的离子; 氨基酸侧链 (Ser, Thr, Asp, Glu) 中性丢失 H₂O 产生少 18 个质量单位的离子; 以及从 *b* 系列离子中性丢失 CO 产生少 28 个质量单位的离子, 即 *a* 系列离子; 另外, 如果一个肽段离子发生多次裂解, 就会产生中间碎片, 包括酰基离子 (acyl, 由至少 2 个氨基酸残基组成, 见图 5.4b) 和亚氨离子 (见图 5.4c 和表 5.2)。亚氨离子代表单个氨基酸的质量, 因此提供了肽段的部分氨基酸组成。并不是所有的肽键在 CID 条件下都具有相同的断裂倾向, 在同一张 MS/MS 谱图中产生的碎片离子的强度也有显著差异。最常发生的断裂是在脯氨酰 (prolyl) 键上, 通常产生 CID 谱图中的最强离子峰。另外, 不稳定的脯氨酰 (prolyl) 键断裂还导致产生了最常见的中间碎片酰基离子 (acyl), 并从脯氨酸残基延伸到 C 末端。

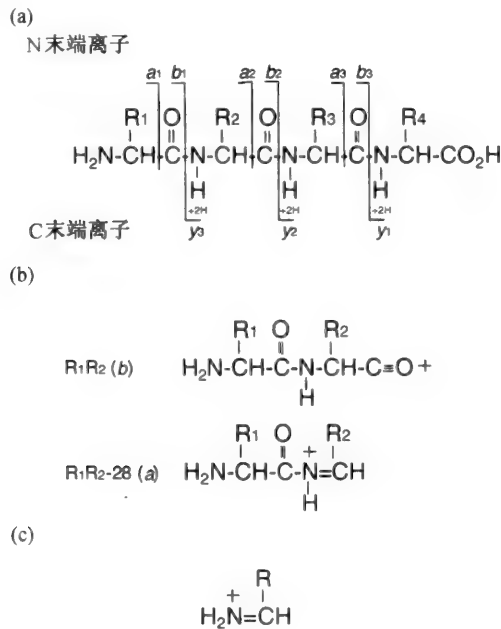


图 5.4 常见的带正电荷的碎片离子命名
 (a) N 末端和 C 末端离子; (b) 中间酰基离子 (internal acyl ions); (c) 中间亚氨离子 (internal immonium ions)。

表 5.2 20 种常见氨基酸的残基和亚氨基离子质量^a

氨基酸名称	3 字母简称	残基质量 ^b	亚氨基离子质量 ^b
丙氨酸	Ala (A)	71	44
天冬酰胺	Asn (N)	114	87
天冬氨酸	Asp (D)	115	88
精氨酸	Arg (R)	156	129, <u>112</u> , 100, 87, <u>70</u> , 43 ^c
半胱氨酸	Cys (C)	103	76
谷氨酰胺	Gln (Q)	128	101 (w/o 84)
谷氨酸	Glu (E)	129	102
甘氨酸	Gly (G)	57	30
组氨酸	His (H)	137	<u>110</u>
异亮氨酸	Ile (I)	113	<u>86</u>
亮氨酸	Leu (L)	113	<u>86</u>
赖氨酸	Lys (K)	128	101, 84, 129 ^c
甲硫氨酸	Met (M)	131	104
苯丙氨酸	Phe (F)	147	120
脯氨酸	Pro (P)	97	<u>70</u> (w/o 112, 100 and 87)
丝氨酸	Ser (S)	87	60
苏氨酸	Thr (T)	101	74
色氨酸	Trp (W)	186	159
酪氨酸	Tyr (Y)	163	<u>136</u>
缬氨酸	Val (V)	99	<u>72</u>

^a 此值来自于 Jardine (1990); ^b 给出的所有质量值均为整数, 划线的值表明其丰度较高; w/o = without 无; ^c 精氨酸和赖氨酸显示有多个亚氨基离子, 均予以列出。

如上所述, 肽段裂解可由 PSD-MALDI-MS 产生, 也可由三级四极或离子阱质谱的 CID 诱导产生。这三种方法中, 由 PSD-MALDI-MS 导致的裂解最不易控制, 在 PSD-MALDI-MS 中仅有的可变参数是激光通量和基质。另外, 在 MALDI-MS 中只产生单电荷离子, 与电喷雾离子化产生的双电荷和三电荷离子相比, 需要更多的能量使其断裂。因此 PSD-MALDI-MS 的成功率很低, 在很多情况下, 同一样品的 PSD 谱图质量比 ESI 仪器的 CID 分析谱图差很多。

碎片离子谱图包含有重复的信息, 例如重叠的 *b* 和 *y* 系列离子以及同一肽段的多个中间离子。这种重复使碎片离子谱成为极丰富的序列特异信息来源, 但也使序列的解析更加复杂, 因为通常不能直接判断某一离子归属于哪一种离子系列。除了人工解析碎片离子谱 (在下文的从头测序中要讲到) 外, 还有两种基本检索手段用于匹配数据库中序列产生的碎片离子谱。第一种方法依靠人工解析部分谱图来鉴别某一特定系列 (*b* 或 *y*) 离子的连续离子元素 (例如鉴别出相互之间相差一个氨基酸残基质量的碎片离子), 解析出部分序列信息, 再结合以下参数: ① 鉴定出的系列中第一个离子的质量数; ② 肽段质量与最后一个离子质量之差 (这两个质量数限制了此序列两端未知残基的总质量); ③ 此完整肽段的质量; ④ 所用蛋白水解酶的酶切特性 (在 N 端、C 端都切还是只切一端)。这五种信息组成了肽序列标签 “PST” (peptide sequence tag, PST) (Mann and Wilm, 1994)。在序列数据库中可以检索出符合 PST 各元素内容的肽序列。第二种类型的碎片离子检索程序是由 Yates 及其合作者开发的 (Eng *et al.*, 1994), 是真正的用未

解析的碎片离子信息检索的程序,称为 SEQUEST。其检索手段首先是在数据库中的每一个序列中寻找与实验测得肽离子质量一致的肽段(在一定误差范围以内),肽段质量计算值是其连续氨基酸质量之和,在检索时如有必要,还要利用蛋白酶的酶切位点特性信息。对检索到的每一个候选肽,程序计算出其在相同 CID 实验条件下预期产生的碎片离子质量数。然后用簇分析算法(cluster-analysis algorithms)将前 500 个计算值与实验测得的碎片离子谱比较。每一个比较结果都有一个得分值,报告出得分最高的结果,如果得分值有显著性意义,则蛋白质鉴定成功。此方法的鉴定基础是未经任何解析的肽段碎片离子质谱图。在数据依赖的(data-dependent) LC-MS/MS 分析中可以自动化地应用这一技术。分析过程中,肽段经反相 HPLC 分离,电喷雾进入联机质谱仪,此质谱仪可进行数据依赖的 MS/MS 分析,即将离子流超过一定阈值的所有离子打碎进行 MS/MS 分析,并记录所有的 CID 谱,自动提交检索。

这种自动化手段有许多优点, Patterson 和 Aebersold (Patterson and Aebersold, 1995) 曾经综述如下:

- (1) 复杂的蛋白质混合物可以被逐个鉴定,因为每一个做 MS/MS 分析的肽段产生的碎片离子数据都独立进入数据库检索 (McCormark *et al.*, 1997)。
- (2) 在这种类型的分析中,来自于每一个肽段的碎片离子谱都代表一个独立的数据库检索。如果一个单一的蛋白质被酶解,蛋白质鉴定结论基本上能够自动确定。因为在多肽段检索结果中,应该是同一个蛋白位于最高排位 (Ducret *et al.*, 1998)。
- (3) 如上所述,这一方法的自动化程度高 (Yates *et al.*, 1995b)。
- (4) 通过指导程序可以预测某一特定残基上的特定修饰,此方法适用于寻找有特定翻译后修饰的肽段,同样可以鉴定产生这一翻译后修饰肽段的蛋白质 (Yates *et al.*, 1995a, 1995b)。
- (5) 以上这些研究者及其他研究者开发的蛋白质鉴定资源的网址见表 1。

4. 用质谱数据对肽段从头测序 (*de novo* sequencing)

前面两部分已经讲述了利用质谱或串联质谱数据鉴定蛋白质的相关手段。这些方法适用于鉴定那些数据库中已有其序列的蛋白质。用 MS 和 MS/MS 方法从头测序的方法已经开发出来。单级质谱测序是分析肽段的阶梯序列,即相邻肽段间相差一个氨基酸残基,用质谱分析肽阶梯,通常用 MALDI-TOF-MS 分析。MS/MS 测序用的是上文所述用于数据库检索的 CID 谱,但碎片离子由人工完全解析。

用于 MS 序列分析的肽阶梯是由化学或酶法降解肽段产生,由 C 末端或 N 末端断裂。测定 N 末端序列的化学方法应用 Edman 化学反应,用 MALDI-MS 读出产生的变短的肽。这一方法通常被称为“阶梯测序”。在这种方法上发展了两种变化方式。第一种方法是在每一测序循环开始时,在多肽/蛋白质底物内加 Edman 试剂。每一个循环中有一小部分的底物被 N 末端封闭,末端封闭后的肽在下一个循环中不会参加反应,这样就产生了序列阶梯。最终反应产物混合物经 MALDI-MS 分析,其质量数之差反映了氨基酸残基的质量数 (Chait *et al.*, 1993; Wang *et al.*, 1994)。第二种方法,也称为“nested peptide sequencing”,是在每一个 Edman 降解循环中多次加入多肽/蛋白质底物。

在此法中加入过量易挥发的 Edman 试剂以使反应完全，每一循环中的过量试剂可通过蒸发被除去。经 MALDI-MS 分析得到肽序列阶梯，实现阶梯测序 (Bartlett-Jones *et al.*, 1994)。这两种方法都需要底物蛋白质为自由 N 末端，并且不能分辨同质异核氨基酸残基亮氨酸和异亮氨酸。谷氨酰胺 (残基质量 128.13) 和赖氨酸 (残基质量 128.17)，在 CID 谱中很难分辨，但在这两种方法中可轻易分辨。因为赖氨酸侧链的氨基会被反应试剂修饰。理论上讲，这两种方法都能分析肽混合物，只要肽段之间质量差异明显，不会与降解产物质量重叠 (Wang *et al.*, 1994)。除了逐步化学降解法，氨肽酶和羧肽酶也可以截短肽段产生肽阶梯序列。此方法的好处是可以将很少量的起始反应物直接放在 MALDI 样品靶的表面反应，只要加入基质就可以终止反应 (Patterson *et al.*, 1995; Woods *et al.*, 1995)。通过控制反应时间可以读出较长的序列，此方法不能分辨 Gln 和 Lys 及 Ile 和 Leu。

如果 CID 能产生高质量的谱，例如谱图含有完整的 *y* 和 *b* 系列离子，那么，通过对碎片离子的从头解析可以确定肽段的序列，可从 Yates 的文章了解其概要 (Yates *et al.*, 1994)。但是，通常对 *y* 和 *b* 离子的判断具有很低的置信度，使得读出最终序列很困难。使 CID 谱图解析变得复杂的主要因素包括①碎片离子丢失，如产生不完整的离子系列；②将观察到的信号归属为某一特定离子系列较困难。通过在酶切过程中对 C 末端离子的特异性同位素标记，可以使 *y* 系列离子的鉴别变得较为容易。这一方法利用了胰蛋白酶的水解特性，即在 C 末端新生成的羧基中从水分子中结合了一个氧。因此，如果在胰蛋白酶水解缓冲液中含有 $H_2^{18}O$ ，那么所有的肽段 (除了 C 末端肽段) 都会有一个同位素标记的 C 末端羧基团 (Schnolzer *et al.*, 1996)。通过在含 50% $H_2^{18}O/H_2^{16}O$ 的缓冲液中进行胰蛋白酶酶切反应，所有的新生肽段 (除了 C 末端肽段) 都将表现为一对峰，两峰质量数相差 2 个质量单位，同时他们的离子强度也只有在 $H_2^{16}O$ 溶液中正常水解产生肽段离子强度的一半。胰蛋白酶切点在赖氨酸和精氨酸之后的肽键，通过激活胰蛋白酶 195 位丝氨酸的 OH 基团，与赖氨酸或精氨酸的羧基形成共价结合的键，同时释放新形成的 N 末端。活性水分子攻击过渡态使酯键断裂，释放新生肽段，肽段的羧基上含有一个从溶剂水中获取的氧，溶剂水为 $H_2^{18}O$ 或 $H_2^{16}O$ 的机率相等。这样，当这两种离子都被质谱仪选择为母离子时，只有具完整 C 末端的离子系列 (*y* 系列离子及其中性丢失离子) 会以相差 2u 的双峰存在。这样就区别出了 *y* 系列离子并可计算出序列差异，对肽段进行从头测序。虽然这种方法简化了对特定离子系列的确认，但也减小了每一个 *y* 离子的信号强度，因此使高灵敏度的测序更为困难。使用高分辨的质谱仪如四极杆-飞行时间串联质谱仪可以实现在单次实验中对肽段谱图的完全解析 (Shevchenko *et al.*, 1997)，同时这种方法已成功应用于离子阱质谱仪 (Qin *et al.*, 1998)。另外一种便于人工解析 CID 谱及从头测序的化学方法是对肽段的羧基实行甲酯化。这个反应使肽段每一个羧基质量增加 14u，包括未修饰的 C 末端、Asp 和 Glu 上的羧基 (Hunt *et al.*, 1986)。如果肽段中没有酸性残基，同时 C 末端没有被修饰，那么只有 *y* 系列离子及其中性丢失离子会表现出质量数的变化，比较衍生化与未衍生化两份样品的碎片离子谱就会鉴定出 *y* 系列离子。如果肽段中存在酸性残基，这个方法的價值就极其有限了，因为只能鉴定中间碎片。对 N 末端残基特异性标记从而鉴定 *b* 系列离子的方法也有尝试。在这些实验中肽段的氨基被衍生化，衍生化试剂在气相中

含有永久正电荷 (Huang *et al.*, 1999; Roth *et al.*, 1998)。虽然这些方法在概念上引人注目,但在实践上遇到困难,主要由于在质谱分析衍生物之前必须先除去过量的试剂,同时这个反应不能区别 N 末端氨基和赖氨酸侧链上的氨基。

5. 磷酸化位点分析的各种方法

5.1 磷酸化分析的原理

在 Krishna 和 Wold (Krishna and Wold, 1998) 所述的上面几种蛋白质修饰类型中,至今只发现少数几种是可逆的或在生命活动中具有重要的调控作用。其中研究最为详细的是蛋白质的磷酸化。蛋白质磷酸化已经被广泛地研究,用以揭示蛋白激酶和磷酸酶各自对磷酸化和去磷酸化的影响,以及靶蛋白结构、功能被修饰后产生的结果 (Charbonneau and Tonks, 1992; Fischer and Krebs, 1989; Hunter, 1987)。研究的最常见的磷酸化类型是丝氨酸、苏氨酸和酪氨酸的磷酸酯化。已知发生的磷酸化还有精氨酸、组氨酸和赖氨酸的亚酰胺化,以及天冬氨酸和谷氨酸的酰化。这些修饰中的其中一些是化学不稳定的,通常观察不到,除非采取特殊的保护措施防止他们在蛋白质分离过程中消失 (Duclos *et al.*, 1991)。举例来说,组氨酸磷酸化是一个相对常见的修饰,至少在原核生物中是这样,但是,在聚丙烯酰胺凝胶常用的酸性染色条件下,磷酸化组氨酸完全消失。因此,在凝胶分离的蛋白质中通常观察不到磷酸化组氨酸,其在细胞中出现的频率也很难估计。

蛋白质磷酸化水平是由两个作用相反的酶系——磷酸酶和激酶来调控的。这些酶中大部分的结构、专一性、调控已被深入研究过,已经有相关综述 (Charbonneau and Tonks, 1992; Fischer and Krebs, 1989; Hunter, 1987)。据估计有几百种蛋白激酶/磷酸酶,它们有不同的底物、动力学特性、组织分布及其与调控通路的关系。通过对酿酒酵母全基因组 DNA 序列中被认为代表蛋白激酶 (和磷酸酶) 的序列基元 (motif) 的分析表明,酵母能够表达 123 种不同的蛋白激酶 (40 种蛋白磷酸酶),这就暗示 2% 的酵母蛋白参与了蛋白质磷酸化过程。读者可以参看许多综述以了解蛋白质磷酸化/去磷酸化在许多受调控的生命系统中的重要性 (Charbonneau and Tonks, 1992; Fischer and Krebs, 1989; Hunter, 1987)。了解蛋白质磷酸化对功能的影响可以深入理解生命系统如何在分子水平进行调控,例如,一个单一的磷酸化过程可以调节糖原磷酸化酶的活性 (Johnson and Barford, 1990)。对蛋白质酪氨酸激酶 p56^{lck} 的研究工作表明,在单个丝氨酸上的磷酸化 [可能是由促分裂原活化蛋白 (MAP) 激酶所致 (Watts *et al.*, 1993)], 能够改变一个酶的底物专一性 (Joung *et al.*, 1995)。酪氨酸磷酸盐诱导的蛋白质相互作用已经表明其提供了稳定蛋白质复合物所需的支架,这种蛋白质复合物表现出复杂的生物功能,特别是在细胞间的信号传导过程中 (Wange *et al.*, 1995; Watts *et al.*, 1994)。因此可以明显看出可逆的蛋白质磷酸化是一种非常重要的蛋白质修饰,它能改变酶的催化活性和底物的专一性,调节具有调控功能的蛋白质复合物的稳定性和亚细胞定位,从而扩展生物活性和功能。

5.2 磷酸化分析策略

蛋白质磷酸化研究的主要目的有三点:第一,定位一个给定蛋白质在体内的磷酸化

氨基酸残基位点，此给定蛋白质位于某一特定状态下的细胞内；第二，鉴定与此磷酸化过程有关的激酶；第三，分析所观察到的磷酸化现象对功能的影响。在这些目的中，第一个可以直接用质谱分析，因此是本章的主题。第二和第三个目标需要应用各种不同的生化、遗传及药理学方法，由于篇幅所限，有关这些的详细情况就不再详述了。

细胞内的许多磷酸化蛋白的含量非常少，几乎不可见。即使蛋白质的表达量处于相对较高的水平，磷蛋白的分析也很困难，因为蛋白磷酸化的化学计量值常常很低，某一个蛋白质中仅有一小部分被磷酸化，同一个蛋白质中也会存在多种不同的磷酸化类型。由于这些原因，即使用当前最灵敏的分析方法，也很难分离到足够量用于分析的体内磷酸化蛋白质。因此许多蛋白质的磷酸化研究是在体外利用激酶反应使蛋白质被磷酸化修饰，通常能够扩大反应规模以产生大量的磷酸化蛋白。在体外研究确定的磷酸化位点被认为具有生物学意义之前，必须证实体内存在相同的磷酸化位点。通常通过比较体内和体外的同一磷酸化蛋白质的二维磷酸肽图得知（图 5.5）。体外磷酸肽产生的量大，经化学或质谱分析作为参考证实体内微量磷酸肽的鉴定。当各自的磷酸肽在谱图中共迁移时则可认为体内、体外磷蛋白有相同的磷酸化位点（例如 Watts *et al.*, 1994, 1996b）。因此，这个方法通过与一个结构清楚的参考样品肽共迁移在某种程度上间接证实了一个体内磷酸化位点的确定。

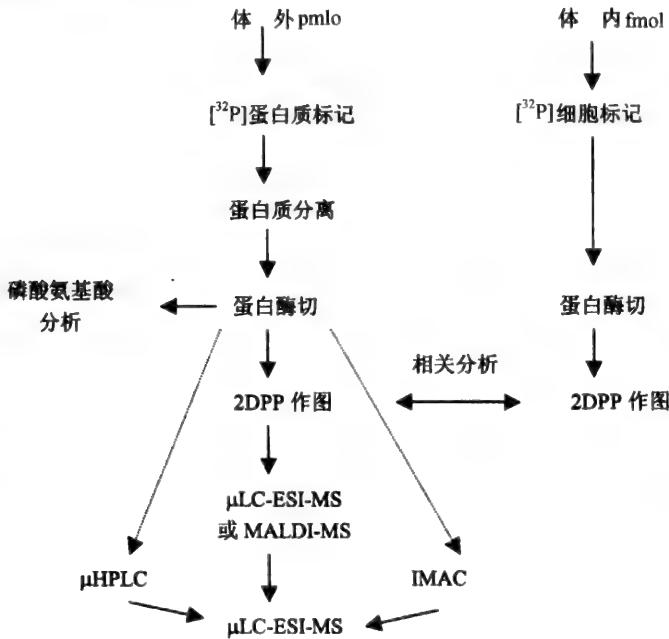


图 5.5 磷酸化位点分析策略。

即使在体内能够产生足够量的磷酸肽用于质谱分析的情况下，实验成功的关键也依赖于所用的样品制备方法。因此，以下部分简要介绍一下常用的，从非磷酸化多肽占主要成分的多肽混合物中检测和分离磷酸化蛋白及多肽的方法。因为这些方法与质谱没有直接关系，所以只是简单讨论一下。读者可以从 Aebersold 的文章中得到这一主题内容

的更详细的论述 (Aebersold and Patterson, 1998)。

5.3 磷蛋白的检测和分离

本质上,任何分析蛋白质磷酸化位点的方法都要依靠对所研究磷酸化蛋白质的纯化,用酶或化学方法水解磷蛋白,分离、分析产生的磷酸肽。幸运的是,前面部分所述的用于处理丙烯酰胺凝胶电泳分离蛋白质的方法,同样适用于磷蛋白的分析。所以,用于产生磷酸肽样品的方法是:用凝胶电泳分离蛋白质,水解蛋白质条带或蛋白斑点,提取产生的磷酸肽混合物作进一步分析。虽然磷蛋白在 SDS-PAGE 上的迁移通常比其非磷酸化的蛋白慢,但是,在一维胶上看到两条迁移非常接近的条带,或在 2-DE 上观察到一系列有相同分子质量但不同等电点的斑点,这样的现象并不足以确定某一个蛋白是磷酸化蛋白。用 ^{32}P 标记后放射自显影,或磷储屏分析 (storage phosphorimaging),或用 Western 印迹方法鉴定磷蛋白的结果较为确定、可信。用现有的方法在常规实验中就可将 ^{32}P 放射标记物掺入体内或体外磷酸蛋白质中 (Patterson and Garrels, 1994)。分子放射标记方法用于体内磷酸化蛋白质标记,要标记的细胞或组织与 $^{32}\text{PO}_4$ 孵育一段时间,使细胞 ATP 库与 ^{32}P 平衡。放射标记的 ATP 被蛋白激酶用于磷酸化其底物。体外蛋白质磷酸化的激酶反应用 $[\gamma\text{-}^{32}\text{P}]$ ATP 作为放射标记物来源,粗分的细胞裂解液或纯化的激酶作为激酶活性来源。Western 印迹在检测酪氨酸磷酸化蛋白方面特别成功,因为已经开发出了一系列非常灵敏的酪氨酸-磷酸盐特异性抗体,例如 4g10 (Upstate Biotechnology, Lake Placid, NY); py2 和 RC10 (Transduction Laboratories, Lexington, KY),同时也因为此方法没有放射性同位素标记。泛磷酸-特异性抗体或磷酸-丝氨酸和磷酸-苏氨酸特异性抗体的开发还不太成功。

5.4 磷酸肽的分离

即使纳喷 MS/MS 技术已经成功用于直接分析蛋白酶解产生的磷酸肽,但是出于以下一些原因,通常建议在质谱分析前作一些分离。

- (1) 磷酸肽在蛋白酶切产物中通常只占少数,因此容易淹没在低丰度离子产生的总背景中,肽分离技术可浓缩分析物,因此会提高信噪比。
- (2) 如果磷酸肽被放射性标记,并具有相同比活度,那么每一个被分离的肽段相应的活度计数就表明这一组分中磷酸肽的相对量。如果已知所用放射性标记物的比活,就能容易地算出磷酸肽的绝对量。
- (3) 放射性标记的磷酸肽的分离谱可以用来定量检测不同时间或细胞状态下蛋白质磷酸化状态的变化。
- (4) 肽分离方法也有效地除去了非肽类杂质,更有利于检测和分析低丰度磷酸肽。

在现有的分离技术中,2D-PP, RP-HPLC, 高分辨凝胶电泳和固相化金属亲和色谱 (immobilized metal affinity chromatography, IMAC) 被优先选择用于分离磷酸肽。虽然要阐明一个明确的策略用于鉴定所有磷蛋白很困难 (图 5.5),但做一个一般性策略的论述还是有价值的。但即使是图 5.5 策略中的内容也没有包含所有可能的途径。任何一个鉴定方案的细节都与许多因素相关,例如个人对检测和分离模式的喜好,可得到的样品及可用的仪器。因为是一个简要的讨论,我们假设要着重表明体内、体外 2D-PP 谱

之间的相似性, 并成功进行证明。总规则是先要得到序列信息, 可通过 MS 或 MS/MS 方法对体外 2D-PP 上的点直接分析, 两者都能得到有用的信息。如果某些点或全部点的序列分析不成功, 那么重新水解蛋白质, 肽段经 HPLC 分段收集, 然后用 MS/MS 分析含³²P 的组分, 可用自动或直接分析方法。如果由于 MS 信噪比糟糕而不能得到 2D-PP 上的点的序列信息, 那么 HPLC 收集的组分本质上会更干净, 也许会得到好的数据。如果磷酸化的化学计量值非常低, 这种 HPLC 制备方法也许会失败, 可能会观察到磷酸肽, 但是柱上浓度会很低, 以致不能产生可用于解析的碎片离子。这种情况下再重复水解蛋白质, 用 IMAC 富集磷酸肽, 用 IMAC 的收集组分做 MS/MS 实验。不管第一步使用了什么方法, 通常从蛋白点到蛋白水解这一步都要进行多次重复, 直到所有感兴趣的磷酸肽都被测序。要记住, 某些肽序列不会以易于解析的方式断裂, 为了证实确切的磷酸化位点, 有必要使用合成肽。这一标准品用于比较研究以证实确切的磷酸化位点。当³²P 不能掺入蛋白质时, 要用下面提到的诊断离子扫描来检测用于序列分析的磷酸肽。这里提到的及图 5.5 所示的每一个分离技术都与下面提到的肽的进一步化学、酶及质谱分析方法相适用。

2D-PP 分离磷酸肽 在 2D-PP 谱中, 肽段的一维分离是在薄层纤维板上进行电泳, 二维分离是在相同的板上进行薄层色谱 (Boyle *et al.*, 1991)。分离到的磷酸肽经放射自显影或磷储屏检测。放射自显影看到的点的数目可用来估计磷酸化位点的最大数目, 点的强度可用来指示肽之间磷酸化的相对化学计量值。此外, 即使并非能够分辨所有的肽, 但样品中每一个磷酸肽都被计数。观察到的点的数目没必要与磷酸化位点的数目直接相联系, 因为磷蛋白中经常可观察到不完全的蛋白水解。但是, 2D-PP 图谱也提供了从其他方法中得不到的有关蛋白质磷酸化位点的重要信息, 包括:

- (1) 磷酸化位点的最大数目。由于蛋白酶水解差异, 谱图产生的点通常多于磷酸化位点。
- (2) 放射自显影强度提供了在所有磷酸肽中磷酸化的相对化学计量。
- (3) 电泳和 TLC 的正交分离提供了磷酸肽之间亲水性的相对状态。

2D-PP 作图的另一个优点是能产生纯化的磷酸肽, 从平板上提取后可用 MS 分析 (Affolter *et al.*, 1994)。如果试图用 2D-PP 作图法作为 MS/MS 分析的样品制备方法, 那么所加蛋白酶的量少, 只要能发挥作用即可。如使用过量的蛋白酶, 在 MS 分析磷酸肽时, 蛋白酶的自切产物会占据图谱的主导, 抑制了磷酸肽的分析。此外, 2D-PP 作图法的灵敏度高且重现性好, 因为检测磷酸肽是通过整合相当长时间内的放射性衰变, 所以方法非常灵敏, 并且灵敏度只依赖于放射性标记物的比活。2D-PP 所获得图谱的高度重现性使其被选择用于分析不同环境, 例如不同时间过程, 不同活性诱导状态下蛋白质的磷酸化状态。

RP-HPLC 分离磷酸肽 反相 HPLC 分段收集磷酸肽重现性好, 简单, 且不需要特殊装备。在 RP-HPLC 中, 磷酸肽根据其疏水性被分离。分离后收集的组分经 Cerenkov 计数检测。对 Cerenkov 计数和时间作图, 显示具放射性活性的组分的计数。纯化的磷酸肽用 MS 分析, 可用 2.2 部分所述不同的 MS 和 MS/MS 方法。相对于 2D-PP, RP-HPLC 的缺点是高亲水性的磷酸肽不会在柱上停留, 而是直接流过柱子。相反, 高疏水性的肽直到最高梯度才会被洗脱, 甚至不会被洗脱。因此, 有可能样品中有些磷酸肽不

会被检测到。总体上说, RP-HPLC 的分辨力不如 2D-PP 作图法的分辨力好。需要注意的另一点是, 磷酸肽会附着在金属表面, 因此, 如果用标准的金属注射器, 会发生明显的样品丢失。使用 RP-HPLC 分析磷酸肽的一个很大的吸引力是 RP-HPLC 系统可以与 ESI 质谱方便地在线连接。使用在线连接到 HPLC 系统的质谱仪使样品混合物中磷酸肽的检测和鉴定成为可能, 即使分析样品没有被放射性标记也可分析, 这需要质谱仪具有特定的扫描功能, 如上文所述包括前体离子扫描、中性丢失及源内解离。因此, 用 LC-MS/MS 分析磷酸肽适用于非同位素标记样品, 在分析人组织中分离的磷酸肽时, 具有明显的优势。

高分辨凝胶电泳分离磷酸肽 在聚丙烯酰胺凝胶上用 2-DE 纯化磷酸肽是最近发表的很有前途的制备技术, 其结果与 2D-PP 结果相似 (Gatti and Traugh, 1999)。非变性凝胶等电聚焦结合碱性 40% PAGE 胶用于磷酸肽分离及比较分析。 ^{32}P 标记样品用放射自显影或磷储屏检测。用 Edman 测序方法鉴定蛋白质并确定磷酸化位点, 但也许更适于用 MS 鉴定。同 HPLC 一样, 但出于不同的原因, 此方法也会丢失特异的磷酸肽。但是, 因为凝胶电泳装置很普通, 所以此方法对于缺乏 2D-PP 作图装置的人来说还是非常具有吸引力的。

IMAC 分离/富集磷酸肽 磷酸肽分析的常见困难是由磷酸化的低化学计量值导致的, 样品中相同序列肽段的磷酸肽的含量要比非磷酸化肽段含量少得多, 这种情况下即使 ^{32}P 标记的 2D-PP 上的点, 经全蛋白质水解及 HPLC 组分收集, 已经确定磷酸肽存在, 用质谱技术也难以鉴定磷酸肽。数据依赖 (data-dependent) 的 MS/MS 方法不能鉴定样品中含量少的成分, 因为 CID 优先选择丰度高的离子进行分析, 为了解决这一问题, 常用 IMAC 选择性富集磷酸肽。这一技术将 Fe^{3+} 或 Ga^{3+} 等金属离子螯合到含有亚氨基二乙酸 (iminodiacetic acid) 或次氨基三醋酸 (nitrilotriacetic acid) 的色谱固定相上 (Neville *et al.*, 1997; Nuwaysir and Stults, 1993; Tempst *et al.*, 1998)。磷酸肽与固定相有高亲和力, 被选择性地吸附在上面。用磷酸盐或高 pH 溶液洗脱, 洗脱组分即为富集的磷酸肽。虽然这一方法对磷酸肽有选择性, 但其他肽, 特别是那些富含酸性氨基酸链的肽段也容易被富集。IMAC 柱上洗脱下的组分可用本章所述的不同的 MS 方式分析。本方法的优点是每一个可溶磷酸肽, 不管其长度如何, 都能被富集, 而且 IMAC 柱洗脱下的样品可直接用于 RP-HPLC 分析。可以建立串联的 IMAC/RP 柱结构, 将肽的富集和分离整合到一起 (Affolter *et al.*, 1994)。

5.5 确定磷酸化氨基酸类型

了解肽或蛋白质中磷酸化氨基酸的类型, 可以限定可能的磷酸化位点, 并因此简化了对多肽链中磷酸化残基的确认。磷酸化残基类型通常由磷酸氨基酸分析或磷酸氨基酸特异性免疫检测确定。

磷酸氨基酸分析是对肽键进行气相或液相水解, 此水解条件下要至少保留一段磷酸酯键完整。 ^{32}P 标记的磷酸蛋白质或磷酸肽的水解产物与磷酸氨基酸标准品混合, 并用 TLC 分离。未标记的标准品由茚三酮 (ninhydrin) 染色检测, ^{32}P 标记氨基酸由放射自显影检测。蛋白质中存在的磷酸氨基酸类型通过染色标准品与放射标记种类的相关分析确定 (Hildebrandt and Fried, 1989)。如果能得到大量样品, 磷酸氨基酸类型可以通过

茚三酮染色直接确定。用 2D-PP 图谱上提取的磷酸肽进行磷酸氨基酸分析尤其有用,因为这些肽段通常只含有一种磷酸氨基酸类型。了解磷酸化氨基酸的性质及其他参数,例如肽段分子质量,裂解蛋白质所用的酶的特性,在某些情况下可以鉴定,或至少可以推测蛋白质中的磷酸化位点。

利用特殊磷酸氨基酸的抗体特异性可以确定地鉴别磷酸化氨基酸的类型,但是这种方法并不能代替磷酸氨基酸分析。酪氨酸磷酸化蛋白质经凝胶分离后,很容易通过与酪氨酸磷酸盐抗血清的免疫印记实验检测,抗血清包括:4g10 (Upstate Biotechnology, Lake Placid, NY) 和 py20, 也有如 RC10 一样的细菌片段抗体 (Transduction Laboratories, Lexington, KY)。这些抗体已经商业化,提供了一个快速、灵敏地检测蛋白质酪氨酸磷酸化的方法。磷酸化丝氨酸和苏氨酸的特异性抗体已有叙述 (Heffetz *et al.*, 1989) 并已商业化。用这些试剂得到的实验结果在某种程度上是不确定的,建议在使用他们的时候要注意,但是最近的研究表明其质量已有所提高 (Soskic *et al.*, 1999)。

5.6 磷酸化位点的确定

磷酸化位点的确定也使用一些通用方法:磷酸化蛋白质被纯化,如果可能蛋白质要均一;磷蛋白被特异性的化学或酶反应切断,产生肽混合物,其中含有一到两个磷酸肽。不同之处在于其分离磷酸肽的策略是为了确定氨基酸序列,定位肽段内磷酸化的残基。上面所述的分离方法,2D-PP 作图、HPLC 或 IMAC 方法,也同样适用于微量制备磷酸肽用于确定磷酸化位点。

磷酸肽的化学测序 过去确定磷酸化位点测序最常用的是逐步化学降解法,如 Wettenhall 等的丝氨酸和苏氨酸磷酸化分析方法 (Wettenhall *et al.*, 1991), Aebersold 等的酪氨酸磷酸化分析方法 (Aebersold *et al.*, 1991)。最近,此方法已经主要被质谱方法所取代。虽然本章主要集中于质谱方法,但仍然有必要提一下非放射活性和放射活性磷酸肽测序的方法,这两种方法可以补充,在某些情况下可以替代质谱方法来确定磷酸化位点 (Watts *et al.*, 1996a)。非放射活性方法是由 Meyer 等发展并综述的 (Meyer *et al.*, 1993), 用于检测磷酸丝氨酸和磷酸苏氨酸,包括 β -消除,与巯基乙醇反应形成 S-乙基化半胱氨酸或 β -甲基-S-乙基半胱氨酸,它们很容易以 PTH 衍生物的形式从自动测序仪中提取出来。放射活性方法是由 Ceerenkov 计数探测器检测在 Edman 化学降解测序循环中释放出的放射性活度 (Aebersold *et al.*, 1991; Wettenhall *et al.*, 1991)。如果测丝氨酸或苏氨酸磷酸盐残基,磷酸盐被除去,可观察到³²P-磷酸盐和各自的 PTH 衍生物 (Wettenhall, 1991)。在酪氨酸磷酸盐情况下,磷酸酯键更稳定,分析的提取物是 PTH 磷酸酪氨酸。如果得到的磷酸肽的量非常有限,也许不可能用化学降解法测肽序列,但是在降解循环后,放射活性从肽中释放,通常有可能确定肽内磷酸化位点。

磷酸肽的质谱分析 已经发展了许多不同的策略和方法用质谱确定磷酸肽中的磷酸化位点,其区别在于质谱与肽分离技术的接口方式,以及所用的质谱方法。如果一个磷酸肽有 1pmol 或更多的量,磷酸肽就能够从 2D-PP 谱中提取出来,并用数据依赖的自动化的微毛细管 LC-MS/MS 分析 (Watts *et al.*, 1994; Whalen *et al.*, 1996)。此方法快速,可以初步估计一个 2D-PP 斑点中磷酸肽的相对量,通过对 2D-PP 谱图的配比,有可能定性、定量地了解细胞或组织在不同状态下分离的磷酸化蛋白在特定位点的磷酸化

过程。对于样品量大大低于 1pmol 量的样品, 磷酸肽信号通常被平板中提取的杂质所干扰而看不到, 这种情况下建议用毛细管液相色谱或凝胶电泳分离肽段。

用于确定磷酸肽中磷酸化位点的质谱方法有两种不同的基本原理, 第一种方法依赖于在质谱仪条件下, 如在 ESI 质谱仪的碰撞室或离子源中, 或在 MALDI-MS 的 PSD 过程中, 磷酸酯键的化学稳定性。因此, 磷酸肽可通过磷酸酯键断裂产生的诊断离子鉴定。第二种方式依靠检测磷酸酯基团使肽段增加的质量数。通常, 在蛋白质磷酸化研究中, 所研究的蛋白质序列是已知的, 因此, 从理论上可以通过对磷酸基团加合在丝氨酸、苏氨酸或酪氨酸残基上引起的 80u 的质量数差的分析而检测蛋白产生的磷酸肽。因此, 得到磷酸化蛋白质的酶切肽质量后, 通常由 MALDI-MS 分析, 如果质量测定值与预期值一致, 就可鉴定混合物中的磷酸肽。这两种方法不能鉴定出肽内的磷酸化位点。磷酸化位点的确定要靠能够有效测序的串联质谱的产物离子扫描技术。但是, 在肽序列中只有一个可能的磷酸化位点情况下, 如果能确实证明此肽段为磷酸肽, 磷酸化残基也能有效定位。如果磷酸化氨基酸类型已用 5.5 部分所述方法确定, 无须离子扫描即能鉴定磷酸化位点的机会也会大大提高。

下面讲的 MS 扫描方式, 在前文已大概叙述过, 已经成功用于磷酸肽的分析。这里我们具体讲一下他们在磷酸肽分析中的应用, 并评价一下其各自的优缺点。总之, 在不能掺入 ^{32}P 或标记物已经衰变到检测限以下的情况下, 产生磷酸盐特异离子 (例如一个诊断离子) 的方法是很有用的。如果希望的话这种扫描模式也能用于分析放射性标记的磷酸肽, 因为由放射性同位素磷酸盐对磷酸肽贡献的质量非常小, 可忽略不计。需要注意的是, 虽然有些研究者担心质谱仪会被 ^{32}P 样品污染, 但这种可能性可以避免, 只要在质谱实验之前等候足够数量的半衰期 (^{32}P 的半衰期为两周)。需要标记的实验, 例如 2D-PP 作图, 可以在放射性活度足够强的时候做。

源内 CID 可用 ESI 源内 CID 产生的负离子形式的诊断离子, H_2PO_4^- (97u), PO_3^- (79u), PO_2^- (63u), 来检测鉴定磷酸肽 (Huddleston *et al.*, 1993; Hunter and Games, 1994)。源内 CID 与在线 HPLC 结合, 既可鉴别磷酸肽的洗脱时间, 得到其色谱标识, 又能得到其分子质量 (Huddleston *et al.*, 1993)。具体做法是, 在 Q1 之前的两个 skimmer 上施加高的锥孔 (cone) 电压, 扫描低质荷比的诊断离子。然后将 cone 电压恢复正常值 (不会诱导产生解离) 扫描高质荷比的离子。仪器的第一个 skimmer 位置使用加热的毛细管的质谱仪也能做类似的实验 (Aebersold *et al.*, 1998)。这个方法使用了交替扫描技术, 即在高八极杆补偿电压下进行诊断离子的选择性离子监测 (selected ion monitoring, SIM), 之后进行两个全扫描。第一个全扫描使用与 SIM 实验相同的高补偿电压, 产生去质子化的磷酸肽分子离子及去磷酸的磷酸肽分子离子。最后在正常八极杆补偿电压下进行第二次全扫描, 产生一个参考图, 与高八极杆补偿电压下的全扫描比较。在 LC 分离过程中, 这种二次质谱扫描系列连续地重复进行。这个实验与 Carr 及其合作者的实验方法提供了相同的信息, 但是因为使用了 SIM, 而不是扫描来检测诊断离子, 每一个扫描循环更快速, 也可能更灵敏。高八极杆补偿电压下的第一次全扫描与低补偿电压下的全扫描相比较, 在多个肽段被共洗脱时, 可以提供线索找到磷酸肽离子, 肽混合物被微毛细管柱分离时, 经常会有肽段共洗脱。这一技术分析磷酸肽标准品通常可低至几个 fmol 上柱量的样品, 但对于体内、体外的实际样品, 灵敏度通常在 pmol 范围。

因为负离子的 CID 谱通常不能产生足够用于序列分析的碎片离子, 如果能用微毛细管柱分离, 在线连接 ESI-MS 系统, 以负离子形式检测磷酸肽, 然后切换到正离子模式的 CID 分析, 当然会非常方便。目前采用诊断离子扫描模式的质谱仪这样做在技术上还有困难, 也许将来使用非扫描模式的质谱仪可以达到这一目的。在四极杆这样的扫描质谱仪上做这一实验的困难在于不能达到正、负离子模式转换所要求时间。

中性丢失扫描 用于磷酸肽检测和分析的中性丢失扫描最早见于 Covey 等所述 (Covey *et al.*, 1991), 后来被 Huddleston 等进一步发展 (Huddleston *et al.*, 1993), 是在三级四极质谱仪上用 ESI 正离子模式分析的。Q1 没有用来为 Q2 中的裂解选择离子, 而是用来扫描。Q3 也用来扫描, 但质量扫描范围与 Q1 不同。确切地说, 两个四极杆以不同的质量范围扫描, 它们之间的差值与丢失的中性分子质荷比一致。对于从 $[M + 2H]^{2+}$ 磷酸肽上丢失的中性磷酸盐, 其质量差值为 $m/z 49$ 。从机理上看, 只有磷酸丝氨酸和磷酸苏氨酸会发生 β -消除, 产生 98 的中性丢失 (Gibson and Cohen, 1990)。此方法的应用并不如上面提到的源内 CID 方法普遍, 因为会出现假阳性, 而且还必须知道丢失磷酸盐的离子带电荷的状态。此方法的优点是以正离子模式进行分析, 可以在同一次实验中通过检测磷酸盐的中性丢失启动 CID, 得到数据依赖的 CID 扫描。

前体离子扫描 此方法是在 ESI 负离子模式下对 Q1 进行连续扫描。所有的离子都在 Q2 中裂解, Q3 中仅能通过一种离子, 对磷酸肽来说通常是丢失的 PO_3^- 基团 ($m/z 79$)。因此, 质谱结果仅显示会丢失 $m/z 79$ 的离子的谱峰 (Wilm *et al.*, 1996)。这大大简化了混合物的分析, 在纳喷离子源直接进样过程中可以分析得很好。如磷酸肽的诊断离子扫描一节所述, 此方法还是存在测序的问题, 即负离子模式检测磷酸丢失后, 立即切换为正离子模式测序。

产物离子扫描 由上文所述的特定扫描法得到的信息通常还不足以鉴定磷酸肽中的磷酸化残基。实际上, 以上方法如源内 CID, 中性丢失和母离子扫描, 是用来将磷酸肽从非磷酸肽中区别出来, 而不是提供序列信息。只有在肽序列已知, 且三种含羟基氨基酸只有一种存在的情况下, 以上三种方法才能成功鉴定磷酸化残基。此外, 只有在肽序列中只含有一个丝氨酸、苏氨酸或酪氨酸残基的情况下, 这种预先设定对答案的简化才起作用。在另一种情况下也可以提供肽序列中磷酸化位点的答案, 即序列已知, 磷酸化的氨基酸分析表明在此磷酸蛋白中只含有三种羟基氨基酸的一种时, 再次简化了答案。当这两种情况都不存在时, 通常需要产生 CID 谱, 并予以解析以确定磷酸肽的共价结构。观察磷酸肽的低能 CID 过程, 其一般趋势是, 磷酸丝氨酸上的磷酸比磷酸苏氨酸更易丢失, 磷酸苏氨酸比磷酸酪氨酸更易丢失磷酸。

较短的磷酸肽比较长的磷酸肽更易于丢失磷酸, 大概是由于在相同碰撞能量下, 短肽的化学键上分布了更多的能量。有趣的是, 很少能观察到磷酸氨基酸的亚氨离子, 磷酸氨基酸的 immonium 离子是由磷酸丢失后形成的脱羟基丙氨酸和脱羟基-2-丁酸的断裂形成的。然而, 用离子阱质谱 (DeGnore and Qin, 1998) 监测磷酸肽的 CID 谱, 在肽碎片离子谱中观察到脱羟基-2-丁酸取代了苏氨酸的位置。

源后衰变 这种实验在 MALDI-TOF 质谱仪上进行, 在前文已讲述了在 single-stage 型仪器中, 通过观察亚稳裂解提供肽段序列信息。这一方法已成功用于磷酸肽的序列分析 (Annan and carr, 1996; Neville *et al.*, 1997)。

用酶和化学方法去磷酸化 这种方法并未得到特别关注,但它能提供磷酸肽中磷酸化氨基酸的定位,可以在非磷酸肽占主导地位的混合物中鉴别磷酸肽,此方法使用的是磷酸酶 (Zhang *et al.*, 1998)。用简单的 MALDI-TOF 仪器就可以得到磷酸化蛋白水解后产生肽段的质量,同样的样品用磷酸酶处理后,除去了丝氨酸、苏氨酸和酪氨酸上的磷酸,再分析其质量。任一个减少了 80u 的质量数,都将预示着这一肽段被磷酸化了。MALDI 做这一实验的优点是产生的肽段离子更倾向于单质子化,使谱图解析比 ESI 更容易。用不同的化学方法处理丝氨酸或苏氨酸磷酸肽,既除去了磷酸又使发生磷酸化修饰的丝氨酸或苏氨酸残基脱水,实际上标记了先前的修饰位点 (Patterson, 1998)。磷酸肽用碱处理后发生磷酸的 β -消除,同时伴随氨基酸侧链的脱水,使氨基酸残基表现出独特的质量 (比正常值少 18u)。这一手段结合 MS/MS 技术,已应用于高磷酸化的 profilaggrin 磷酸化位点的鉴定 (Resing *et al.*, 1995)。

5.7 酶分析法及磷蛋白分析的未来方向

目前磷酸化蛋白质分析的局限性围绕在体内³²P标记蛋白的直接分析困难,有两个问题造成这些困难。第一是磷酸化通常是低化学计量值,要求在质谱分析前对磷酸肽进行一定程度的富集。第二是不能产生足够的体内³²P标记肽用于质谱鉴定。第二种情况必须使用体外³²P标记肽,要求体内和体外磷酸蛋白在 2D-PP 作图中表现一致。这就使磷蛋白从纯化到磷酸肽测序这一全过程耗时太长。而且是假设激酶在已知条件下可以在体外重现体内磷酸化过程的。因此,还是需要能直接分析体内磷酸化蛋白质的方法。

虽然产生足够量体内³²P标记蛋白这一问题目前还不可能直接解决,但越来越灵敏的质谱仪可以减少样品的消耗量。例如使用 MALDI 或 ESI 源的 FT-ICR-MS,对多肽的检测灵敏度已证明远大于标准的三级四极和离子阱技术。这一技术可常规分析 MALDI靶上几十 fmol 的多肽样品,精确测定其分子质量 (Solouki *et al.*, 1995)。虽然在 ICR 室中实际样品的 MS/MS 实验灵敏度在 fmol 范围,但对 HPLC 柱上洗脱的单个肽用宽带解吸法 (broad-band dissociation) 可以做到 amol 水平。例如一个 50 μ m 内径的微毛细管 HPLC 柱可将 9amol 的磷酸化血管紧张素 II 导入 ESI 离子源,此 ESI 源连接 7T 的 FT-ICR-MS,可以用宽带解吸法裂解并鉴定此标准肽 (Bruce J. E., Goodlett, D. R., Smith, R. D. and Abersold, R., 1998。未发表数据,来自 Battelle Memorial 研究所,环境与分子科学实验室, Richland, WA)。这种灵敏度为体内有限样品的处理提供了方便。

仪器控制能力的提高对磷酸化蛋白质分析性能的提高也有贡献。举例来说,如果自动化的 LC-MS/MS 的单次实验操作中能够检测样品中存在的磷酸肽,确定肽段质量,还能够得到 CID 谱,这样将非常理想。串联质谱仪如果能实现从负离子 MS 到正离子 MS/MS 模式快速切换的功能,上面的想法也许能够实现。在这种质谱仪中,可用前面提到的其中一种负离子检测方式如源内 CID 法检测磷酸肽并测定其质量数。检测到的磷酸肽信号触发仪器转向正离子模式并选择检测到的磷酸肽离子作 CID 分析。我们希望将来这种仪器能成为现实。

新兴的蛋白质组学领域中一个重要的部分就是需要鉴定混合物样品中所有蛋白质的修饰状态。因此,磷酸化蛋白质分析的一个重要延伸及今后方向是有关样品中磷酸化蛋

白的总体及定量分析。目前,所有用于磷酸化蛋白分析的方法都集中于对单一蛋白质的分析。将来有必要对目前方法加以简化和自动化以适应高通量研究,或者开发完全不同的仪器和方法以适用于蛋白质组范围的磷酸化研究。

6. 目前及将来面临的机遇和挑战

蛋白质的鉴定和结构确认是任何一个蛋白质组学努力探索的领域中必不可少的部分。为了使蛋白质鉴定更有效,必须采用高通量手段,具有高效率,高准确度和灵敏度。如果是 2-DE 分离的蛋白质,实现这一目标的设想是采用机器人系统。此系统具有专门设计的切胶工具,用于从胶上切下指定的蛋白点,还具有计算机控制的蛋白质酶切模具。图像分析系统检测胶上或转印到膜上被染色的蛋白质,图像系统还具有模式消减功能,因此能够标出两个不同样品之间有定量或定性差别的蛋白质。有分析价值的斑点被标记上,切点工具被自动引导到指定的坐标,切下蛋白质斑点。切下来的斑点被转移到容器中用于酶切反应。全部的酶切过程至少要包括胶片清洗,加入酶解缓冲液,肽段提取及样品脱盐,这一过程通常是由独立的液体输送机器人完成的,可以自动平行处理多个样品 (Traini *et al.*, 1998)。同一机器人还能从每一样品的酶切液中取样,点到 MALDI 的靶上,然后进行自动化的 MALDI-MS 分析,分析过程也不需要人工参与 (Onnerfjord *et al.*, 1999; Traini *et al.*, 1998)。MALDI-MS 分析得出的质量同样也可实现自动高通量肽质量检索。任何未被这一策略鉴定出的蛋白质,取其剩余的酶切产物用于较低通量(但仍然是自动化)的技术分析,例如 LC-MS/MS,用数据依赖方式产生碎片离子谱,并自动在序列数据库中进行检索 (Ducret *et al.*, 1998)。剩余的酶切溶液还可用纳喷 ESI-MS/MS 分析,这一技术仍然是较低通量的,但是灵敏度很高 (Shevchenko *et al.*, 1996)。LC-MS/MS 和 nano-ESI-MS/SM 都能从 MALDI-MS 分析已经确定存在的质量数中受益。这里反复强调的一个策略是,首先用高通量手段鉴定尽可能多的蛋白质,然后再使用低通量(并且更贵)的技术手段。MALDI-MS 分析也能够将那些在标准酶切条件下不能产生肽段的蛋白排除在进一步分析之外。这种模式的系统非常昂贵,需要大量样品才能体现其价值,所以只限于安装在核心实验室或专门的实验室。另外,它们特别适用于分析全基因组测序已经完成的物种(微生物)中的蛋白质。用 MALDI-MS 数据在 EST 数据库中鉴定蛋白质还存在困难,目前限制了这一系统对高级真核生物,特别是人和鼠蛋白质的鉴定。

高通量 MS/MS 鉴定蛋白质方法的进展还很慢,主要是由于仪器和数据处理过程的复杂性。ESI-MS/MS 仪器的微流控 (microfabricate) 样品输送系统的发展 (Figeys *et al.*, 1998), 以及 MS/MS 无需纯化分离就能鉴定较复杂蛋白质混合物各组分的可能 (McCormack *et al.*, 1997), 是很有希望的技术进展,表明用 MS/MS 路线大规模鉴定蛋白质的技术正在迅速进步。

如果蛋白质能从胶上直接分析,或转到膜上后直接分析(用 MALDI-MS 分析),这样也可实现高通量,并且不需要从胶上切下蛋白点及大量的液体输送系统。但是这种方法需要在 MALDI-MS 分析之前将酶解液和基质液运送到胶上/膜上,并且不能牺牲凝胶电泳分离的空间分辨力。有许多小组已经在这一构想上有了进展,他们用 UV-MALDI-MS 从胶上直接分析凝胶分离的蛋白质 (Loo *et al.*, 1996), 另一些小组用 UV 或 IR-

MALDI-MS (Eckerskorn *et al.*, 1997; Patterson, 1995; Strupat *et al.*, 1994; Vestling and Fenselau, 1995) 直接分析转到膜上的蛋白质。在一系列 proof-of-concept 实验中, Loo, *et al.* (1996) 生成了真正的二维凝胶, 他们用 UV-MALDI-MS 从 IEF 胶上直接分析经等电聚焦分离的蛋白质, 生成了分子质量二维谱。基质点在凝胶表面, 并将分散度减少到最小程度, 因此保持了空间分辨力。这种分析法必须采用超薄胶 (0.35mm) 以保持有效的离子化。同一小组还报道了 CNBr 水解的蛋白质也适用于这种分析方法 (Loo *et al.*, 1997)。但是 PSD-MALDI-MS 通常要求被分析肽段小于 CNBr 裂解产生的肽段, 所以还需要阐明是否 PSD-MALDI-MS 可直接用于胶上分析。

许多有离子化抑制效应的污染物通过将蛋白质转移到膜上可以被洗脱, 因此有很多小组评估了凝胶分离转膜后蛋白质的离子化状态。采用这一技术时, 在引入基质溶液时要特别小心, 因为有机溶剂 (在基质溶液中常用) 会导致蛋白质分散, 损失空间分辨力 (如上所述, Vestling and Fenselau, 1994a, 1994b)。系统评价不同的膜后发现, PVDF 膜或其衍生物最适用此类实验 (Vestling and Fenselau, 1995)。UV 和 IR 激光都被用于膜上蛋白的离子化, 两者都已成功, 但是 IR 激光对埋在膜内的蛋白质的提取显得更有效 (Schreiner *et al.*, 1996; Strupat *et al.*, 1994)。Eckerskorn 等 (1997) 已经报道了用 IR-MALDI-MS 对凝胶电泳分离并转印到 PVDF 膜上的蛋白质的离子化是可能的, 他们将数据转化为轮廓线来反应每种蛋白质的丰度, 因此在 MALDI-MS 分析前不需将蛋白显像。但是, 在这种分析方法用于整个 2-DE 胶上斑点之前, 还需要解决数据处理方面的重要问题。另外, 凝胶分离蛋白质点的完整分子质量的测定是研究蛋白质特性的重要方面, 但是对于蛋白质鉴定并没有明显帮助。为了鉴定, 蛋白质必须被酶解 (肽质量检索策略), 并且/或产生的肽段必须裂解 (未解析碎片离子检索策略)。膜上蛋白的肽水解也已实施, 并用 UV 和 IR-MALDI-MS 都分析过 (Eckerskorn *et al.*, 1997; Vestling and Fenselau, 1994b)。但是, 目前为止仅有 UV-MALDI-MS 的 PSD 数据发表 (Fabris *et al.*, 1995), 没有 PSD-(IR)-MALDI-MS 的数据。

Hochstrasser (1998) 引入了“分子扫描” (molecular scanner) 的概念, 即 2-DE 分离的蛋白质电转印到膜上, 用 MALDI-MS 以自动化模式分析膜上所有的点。最终, Hochstrasser 的小组发明了对 2-DE 胶上所有的点进行大规模平行酶切的技术。他们将胰蛋白酶固定化到膜上, 再将此膜放在凝胶和 PVDF 膜之间, 在电转印过程中施加一个脉冲电场, 蛋白质从胶上被慢慢转移, 并通过覆盖有胰蛋白酶的膜被水解。水解剩余的完整蛋白同此蛋白质的酶切肽段一起被 PVDF 膜吸附, 并由 MALDI-MS 分析, 用 UV 或 IR 激光都获成功 (Hochstrasser, 1998)。尽管所有这些技术都处于开发的早期阶段, 但它们仍显示出这种分析方式是可行的。一旦膜上肽段的 PSD-MALDI-MS 分析的所有难题都被解决, 最重要的是这一过程的自动化问题的解决, 这一整体策略的力量就能被充分证明。目前这一技术的最大问题是, 产生的肽段的数目有限, 这些肽段在进行 PSD 分析时必须能够产生足够数量的碎片离子并能成功进行数据库检索。但是, 通过使用化学修饰步骤来增强离子化和肽段的裂解, 这一问题将会被很好地解决 (Spengler *et al.*, 1997)。

尽管由于 2-DE 具有高分辨力, 是蛋白质组研究最常使用的蛋白质分离技术, 但是以色谱为基础的分离技术的研究也正在开发中, 因为色谱技术更易与 MS/MS 在线整

合。Opiteck 等已经用 2-D 色谱(阳离子交换-反相,用两套独立的液相泵)分离了如大肠杆菌裂解液中的蛋白质混合物,并进行在线 ESI-MS 分析,所有过程在两小时内完成(Opiteck *et al.*, 1997b)。他们对诱导和非诱导大肠杆菌裂解液分别分析,鉴定出一个过度表达的蛋白质。同一小组还报道使用这一技术分离了肽混合物(Opiteck *et al.*, 1997a)。Hsieh 等(Hsieh *et al.*, 1996)使用了由计算机控制的配件,包括一个自动进样器,5 个柱子,3 个 10 通阀,使一系列步骤在线进行,避免任何用手工操作进行的物质转移。这一策略用于从混合物中免疫亲和提取鉴定蛋白质,包括以下一系列步骤:免疫亲和色谱;在一混合填充的强离子交换吸附柱上进行脱盐和缓冲液交换;在一个固定化胰蛋白酶柱上进行酶切;在一短毛细管灌注反相柱上提取肽段;最后在一个分析型的反相柱上进行分离并与 MS/MS 在线连接分析(Hsieh *et al.*, 1996)。

不管最初的蛋白质分离采用 2-DE 还是靶向色谱技术,为了最准确地鉴定蛋白质,还是需要生成序列特异性的碎片离子谱。在这一点上,尽管未解析碎片离子检索程序在目前格式下的作用很大,但它们并没有利用碎片离子谱中包含的另外的信息:碎片离子的相对丰度信息。目前的检索程序是将实验测得的质量值与预期值进行比较,当一个科学家在评价一个图谱时,要考虑各种肽键裂解的倾向大小,因为碎片离子谱可被看作每一肽键裂解的特性曲线,某些氨基酸残基旁的肽键在气相中更易于断裂。当检索程序能利用碎片离子相对强度信息时,鉴定的置信水平会提高,使其能真正实现自动化。这是任何大规模蛋白质组学研究要努力达到的一个重要目标。

致谢:

作者 D. R. G. 和 R. A. 感谢国家科学基金会分子生物技术科学和技术中心和 Kinetek Pharmaceuticals 的基金资助,感谢国立卫生研究院对作者 R. A. 的研究基金资助(No. 1R01 AI 41109-01)。作者 D. R. G. 感谢 J. D. Watts 博士帮助讨论磷酸化蛋白质分析。

(王京兰译 钱小红校)

参考文献

- Aebersold, R. and Leavitt, J. (1990) Sequence analysis of proteins separated by polyacrylamide gel electrophoresis: towards an integrated protein database. *Electrophoresis* **11**: 517-527.
- Aebersold, R. and Patterson, S.D. (1998) Current problems and technical solutions in protein biochemistry. In: *Proteins: Analysis and Design* (ed. R.H. Angeletti). Academic Press, San Diego, CA, pp. 3-120.
- Aebersold, R.H., Teplow, D.B., Hood, L.E. and Kent, S.B.H. (1986) Electroblothing onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. *J. Biol. Chem.* **261**: 4229-4238.
- Aebersold, R.H., Leavitt, J., Saavedra, R.A., Hood, L.E. and Kent, S.B. (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl Acad. Sci. USA* **84**: 6970-6974.
- Aebersold, R., Watts, J.D., Morrison, H.D. and Bures, E.J. (1991) Determination of the site of tyrosine phosphorylation at the low picomole level by automated solid-phase sequence analysis. *Anal. Biochem.* **199**: 51-60.
- Aebersold, R., Figeys, D., Gygi, S., Corthals, G., Haynes, P., Rist, B., Sherman, J., Shang, Y. and

- Goodlett, D. (1998) Towards an integrated technology for the generation of multidimensional protein expression maps. *J. Protein Chem.* **17**: 533–535.
- Affolter, M., Watts, J.D., Krebs, D.L. and Aebersold, R. (1994) Evaluation of 2-dimensional phosphopeptide maps by electrospray-ionization mass-spectrometry of recovered peptides. *Anal. Biochem.* **223**: 74–81.
- Annan, R.S. and Carr, S.A. (1996) Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal. Chem.* **68**: 3413–3421.
- Bartlet-Jones, M., Jeffery, W.A., Hansen, H.F. and Pappin, D.J.C. (1994) Peptide ladder sequencing by mass-spectrometry using a novel, volatile degradation reagent. *Rapid Commun. Mass Spectrom.* **8**: 737–742.
- Biemann, K. (1990) Nomenclature for peptide fragment ions (positive ions). [Appendix 5]. *Methods Enzymol.* **193**: 886–887.
- Boyle, W.J., van der Geer, P. and Hunter, T. (1991) Phosphopeptide mapping and phosphoamino acid analysis by two-dimensional separation on thin-layer cellulose plates. *Methods Enzymol.* **201**: 110–149.
- Brown, J.L. (1979) A comparison of the turnover of amino-terminally acetylated and non-acetylated mouse L-cell proteins. *J. Biol. Chem.* **254**: 1447–1449.
- Brown, J.L. and Roberts, W.K. (1976) Evidence that approximately eighty percent of the soluble proteins from Ehrlich ascites cells are amino-terminally acetylated. *J. Biol. Chem.* **251**: 1009–1014.
- Brown, R.S. and Lennon, J.J. (1995) Mass resolution improvement by incorporation of pulsed ion extraction/ionization linear time-of-flight mass spectrometry. *Anal. Chem.* **67**: 1998–2003.
- Carr, S.A. and Annan, R.S. (1997) Overview of peptide and protein analysis by mass spectrometry. In: *Current Protocols in Molecular Biology* (eds F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl). Wiley, New York, pp. 10.21.1–10.21.27.
- Chait, B.T., Wang, R., Beavis, R.C. and Kent, S.B.H. (1993) Protein ladder sequencing. *Science* **262**: 89–92.
- Charbonneau, H. and Tonks, N.K. (1992) 1002 Protein phosphatases? *Annu. Rev. Cell Biol.* **8**: 463–493.
- Chaurand, P., Luetzenkirchen, F. and Spengler, B. (1999) Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **10**: 91–103.
- Cordwell, S.J., Wilkins, M.R., Cerpa-Poljak, A., Gooley, A.A., Duncan, M., Williams, K.L. and Humphery-Smith, I. (1995) Cross-species identification of proteins separated by two-dimensional gel electrophoresis using matrix-assisted laser desorption ionization/time-of-flight mass spectrometry and amino acid composition. *Electrophoresis* **16**: 438–443.
- Cornish, T.J. and Cotter, R.J. (1994) A curved field reflectron time-of-flight mass-spectrometer for the simultaneous focusing of metastable product ions. *Rapid Commun. Mass Spectrom.* **8**: 781–785.
- Courchesne, P.L. and Patterson, S.D. (1998) Identification of proteins by matrix-assisted laser desorption/ionization mass spectrometry using peptide and fragment-ion masses. *Methods Mol. Biol.* **112**: 487–511.
- Courchesne, P.L., Jones, M.D., Robinson, J.R., Spahr, C.S., McCracken, S., Bentley, D.L., Luethy, R. and Patterson, S.D. (1998) Optimization of capillary chromatography ion-trap mass spectrometry for identification of gel-separated proteins. *Electrophoresis* **19**: 956–967.
- Covey, T.R., Huang, E.C. and Henion, J.D. (1991) Structural characterization of protein tryptic peptides via liquid chromatography/mass spectrometry and collision induced dissociation of their doubly charged molecular ions. *Anal. Chem.* **63**: 1193–1200.
- Craig, A.G., Fischer, W.H., Rivier, J.E., McIntosh, J.M. and Gray, W.R. (1995) MS based scanning methodologies applied to *Conus* venom. In: *Techniques in Protein Chemistry VI* (ed. J.W. Crabb). Academic Press, San Diego, CA, pp. 31–38.
- Davis, M.T. and Lee, T.D. (1997) Variable flow liquid chromatography-tandem mass spectrometry and the comprehensive analysis of complex protein digest mixtures. *J. Am. Soc. Mass Spectrom.* **8**: 1059–1069.
- DeGnove, J.P. and Qin, J. (1998) Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **9**: 1175–1188.

- Duclos, B., Marcandier, S. and Cozzone, A.J. (1991) Chemical properties and separation of phosphoamino acids by thin-layer chromatography and/or electrophoresis. *Methods Enzymol.* **201**: 10–21.
- Ducret, A., van Oostveen, I., Eng, J.K., Yates, J.R. III and Aebersold, R. (1998) High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* **7**: 706–719.
- Eckerskorn, C., Strupat, K., Schleuder, D., Hochstrasser, D., Sanchez, J.-C., Lottspeich, F. and Hillenkamp, F. (1997) Analysis of proteins by direct-scanning infrared-MALDI mass spectrometry after 2D-PAGE separation and electroblotting. *Anal. Chem.* **69**: 2888–2892.
- Eng, J.K., McCormack, A.L. and Yates, J.R. III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**: 976–989.
- Fabris, D., Vestling, M.M., Cordero, M.M., Doroshenko, V.M., Cotter, R.J. and Fenselau, C. (1995) Sequencing electroblotted proteins by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **9**: 1051–1055.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**: 64–71.
- Fenyo, D., Qin, J. and Chait, B.T. (1998) Protein identification using mass spectrometric information. *Electrophoresis* **19**: 998–1005.
- Figey, D. and Aebersold, R. (1998) High sensitivity analysis of proteins and peptides by capillary electrophoresis-tandem mass spectrometry: recent developments in technology and applications. *Electrophoresis* **19**: 885–892.
- Figey, D., Gygi, S.P., McKinnon, G. and Aebersold, R. (1998) An integrated microfluidics-tandem mass spectrometry system for automated protein analysis. *Anal. Chem.* **70**: 3728–3734.
- Fischer, E.H. and Krebs, E.G. (1989) Commentary on “the phosphorylase b to a converting enzyme of rabbit skeletal muscle”. *Biochim. Biophys. Acta* **1000**: 297–301.
- Fraser, C.M. and Fleischmann, R.D. (1997) Strategies for whole microbial genome sequencing and analysis. *Electrophoresis* **18**: 1207–1216.
- Gatti, A. and Traugh, J.A. (1999) A two-dimensional peptide gel electrophoresis system for phosphopeptide mapping and amino acid sequencing. *Anal. Biochem.* **266**: 198–204.
- Gibson, B.W. and Cohen, P. (1990) Liquid secondary ion mass spectrometry of phosphorylated and sulfated peptides and proteins. *Methods Enzymol.* **193**: 480–501.
- Heffetz, D., Fridkin, M. and Zick, Y. (1989) Antibodies directed against phosphothreonine residues as potent tools for studying protein phosphorylation. *Eur. J. Biochem.* **182**: 343–348.
- Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C. and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl Acad. Sci. USA* **90**: 5011–5015.
- Hewick, R.M., Hunkapiller, M.W., Hood, L.E. and Dreyer, W.J. (1981) A gas-liquid solid phase peptide and protein sequencer. *J. Biol. Chem.* **256**: 7990–7997.
- Hildebrandt, E. and Fried, V.A. (1989) Phosphoamino acid analysis of protein immobilized on polyvinylidene difluoride membrane. *Anal. Biochem.* **177**: 407–412.
- Hochstrasser, D.F. (1998) Proteome in perspective. *Clin. Chem. Lab. Med.* **36**: 825–836.
- Hsieh, Y.L., Wang, H., Elicone, C., Mark, J., Martin, S.A. and Regnier, F. (1996) Automated analytical system for the examination of protein primary structure. *Anal. Chem.* **68**: 455–462.
- Huang, Z.-H., Shen, T., Wu, J., Gage, D.A. and Watson, J.T. (1999) Protein sequencing by matrix-assisted laser desorption ionization-postsource decay-mass spectrometry analysis of the *N*-Tris(2,4,6-trimethoxyphenyl)phosphine-acetylated tryptic digests. *Anal. Biochem.* **268**: 305–317.
- Huddleston, M.J., Annan, R.S., Bean, M.F. and Carr, S.A. (1993) Selective detection of phosphopeptides in complex mixtures by electrospray liquid chromatography/mass spectrometry. *J. Am. Soc. Mass Spectrom.* **4**: 710–717.
- Hunt, D.F., Yates, J.R. III, Shabanowitz, J., Winston, S. and Hauer, C.R. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **83**: 6233–6237.
- Hunter, A.P. and Games, D.E. (1994) Chromatographic and mass spectrometric methods for the identification of phosphorylation sites in phosphoproteins. *Rapid Commun. Mass Spectrom.* **8**: 559–570.
- Hunter, T. (1987) 1001 protein kinases. *Cell* **50**: 823–829.

- James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993) Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* **195**: 58–64.
- James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1994) Protein identification in DNA databases by peptide mass fingerprinting. *Protein Sci.* **3**: 1347–1350.
- Jardine, I. (1990) Molecular weight analysis of proteins. *Methods Enzymol.* **193**: 441–455.
- Jensen, O.N., Podtelejnikov, A. and Mann, M. (1996a) Delayed extraction improves specificity in database searches by matrix-assisted laser desorption/ionization peptide maps. *Rapid Commun. Mass Spectrom.* **10**: 1371–1378.
- Jensen, O.N., Vorm, O. and Mann, M. (1996b) Sequence patterns produced by incomplete enzymatic digestion or one-step Edman degradation of peptide mixtures as probes for protein database searches. *Electrophoresis* **17**: 938–944.
- Jensen, O.N., Podtelejnikov, A.V. and Mann, M. (1997) Identification of the components of simple protein mixtures by high accuracy peptide mass mapping and database searching. *Anal. Chem.* **69**: 4741–4750.
- Jensen, O.N., Larsen, M.R. and Roepstorff, P. (1998) Mass spectrometric identification and microcharacterization of proteins from electrophoretic gels: strategies and applications. *Proteins* **2** (Suppl.): 74–89.
- Johnson, L.N. and Barford, D. (1990) Glycogen phosphorylase. The structural basis of the allosteric response and comparison with other allosteric proteins. *J. Biol. Chem.* **265**: 2409–2412.
- Jonscher, K.R. and Yates, J.R. III (1997) The quadropole ion trap mass spectrometer – a small solution to a big challenge. *Anal. Biochem.* **244**: 1–15.
- Joung, I., Kim, T., Stolz, L.A., Payne, G., Winkler, D.G., Walsh, C.T., Strominger, J.L. and Shin, J. (1995) Modification of Ser59 in the unique N-terminal region of tyrosine kinase p56lck regulates specificity of its Src homology 2 domain. *Proc. Natl Acad. Sci. USA* **92**: 5778–5782.
- Karas, M. and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal. Chem.* **60**: 2299–2301.
- Katta, V., Chowdhury, S.K. and Chait, B.T. (1991) Use of a single-quadropole mass spectrometer for collision-induced dissociation studies of multiply charged peptide ions produced by electrospray ionization. *Anal. Chem.* **63**: 174–178.
- Kaufmann, R., Kirsch, D. and Spengler, B. (1994) Sequencing of peptides in a time-of-flight mass spectrometer: evaluation of postsource decay following matrix-assisted laser desorption ionization (MALDI). *Int. J. Mass Spectrom. Ion Processes* **131**: 355–385.
- Krishna, R.G. and Wold, F. (1998) Posttranslational modifications. In: *Proteins: Analysis and Design* (ed. R.H. Angeletti). Academic Press, San Diego, CA, pp. 121–206.
- Lamond, A.I. and Mann, M. (1997) Cell biology and the genome projects – a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends Cell Biol.* **7**: 139–142.
- Loo, J.A., Edmonds, C.G. and Smith, R.D. (1990) Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. *Science* **248**: 201–204.
- Loo, R.R.O., Stevenson, T.I., Mitchell, C., Loo, J.A. and Andrews, P.C. (1996) Mass-spectrometry of proteins directly from polyacrylamide gels. *Anal. Chem.* **68**: 1910–1917.
- Loo, R.R.O., Mitchell, C., Stevenson, T.I., Martin, S.A., Hines, W., Juhasz, P., Patterson, D., Peltier, J., Loo, J.A. and Andrews, P.C. (1997) Sensitivity and mass accuracy for proteins analyzed directly from polyacrylamide gels: implications for proteome mapping. *Electrophoresis* **18**: 382–390.
- Louris, J.N., Brodbelt-Lustig, J.S., Cooks, R.G., Glish, G.L., van Berkel, G.J. and McLuckey, S.A. (1990) Ion isolation and sequential stages of mass spectrometry in a quadropole ion trap mass spectrometer. *Int. J. Mass Spectrom. Ion Processes* **96**: 117–137.
- Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**: 4390–4399.
- Mann, M., Hojrup, P. and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* **22**: 338–345.
- Matsudaira, P. (1987) Sequence analysis from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**: 10035–10038.

- McCormack, A.L., Schieltz, D.M., Goode, B., Yang, S., Barnes, G., Drubin, D. and Yates, J.R. III (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**: 767–776.
- Meyer, H.E., Eisermann, B., Heber, M., Hoffmann-Posorske, E., Korte, H., Weigt, C., Wegner, A., Hutton, T., Donella-Deana, A. and Perich, J.W. (1993) Strategies for nonradioactive methods in the localization of phosphorylated amino acids in proteins. *FASEB J.* **7**: 776–782.
- Moritz, R.L., Eddes, J., Ji, H., Reid, G.E. and Simpson, R.J. (1995) Rapid separation of proteins and peptides using conventional silica-based supports: identification of 2-D gel proteins following in-gel proteolysis. In: *Techniques in Protein Chemistry VI* (ed. J.W. Crabb). Academic Press, San Diego, CA, pp. 311–319.
- Neville, D.C., Rozanas, C.R., Price, E.M., Gruis, D.B., Verkman, A.S. and Townsend, R.R. (1997) Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. *Protein Sci.* **6**: 2436–2445.
- Nuwaysir, L.M. and Stults, J.T. (1993) Electrospray ionization mass spectrometry of phosphopeptides isolated by on-line immobilized metal-ion affinity chromatography. *J. Am. Soc. Mass Spectrom.* **4**: 662–669.
- Ommerfjord, P., Ekstrom, S., Bergquist, J., Nilson, J., Laurell, T. and Marko-Varga, G. (1999) Homogenous sample preparation for automated high throughput analysis with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **13**: 315–322.
- Opiteck, G.J., Jorgenson, J.W. and Anderegg, R.J. (1997a) Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. *Anal. Chem.* **69**: 2283–2291.
- Opiteck, G.J., Lewis, K.C., Jorgenson, J.W. and Anderegg, R.J. (1997b) Comprehensive on-line LC/LC/MS of proteins. *Anal. Chem.* **69**: 1518–1524.
- Pappin, D.J.C., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**: 327–332.
- Pappin, D.J.C., Rahman, D., Hansen, H.F., Bartlett-Jones, M., Jeffery, W. and Bleasby, A.J. (1995) Chemistry, mass spectrometry and peptide-mass databases: Evolution of methods for the rapid identification and mapping of cellular proteins. In: *Mass Spectrometry in the Biological Sciences* (eds A.L. Burlingame and S.A. Carr). Humana Press, Totowa, NJ, pp. 135–150.
- Patterson, D.H., Tarr, G.E., Regnier, F.E. and Martin, S.A. (1995) C-Terminal ladder sequencing via matrix-assisted laser-desorption mass-spectrometry coupled with carboxypeptidase-Y time-dependent and concentration-dependent digestions. *Anal. Chem.* **67**: 3971–3978.
- Patterson, S.D. (1994) From electrophoretically separated protein to identification: strategies for sequence and mass analysis. *Anal. Biochem.* **221**: 1–15.
- Patterson, S.D. (1995) Matrix-assisted laser-desorption/ionization mass spectrometric approaches for the identification of gel-separated proteins in the 5–50 pmol range. *Electrophoresis* **16**: 1104–1114.
- Patterson, S.D. (1997) Identification of low to subpicomolar quantities of electrophoretically separated proteins: towards protein chemistry in the post-genome era. *Biochem. Soc. Trans.* **25**: 255–262.
- Patterson, S.D. (1998) Protein identification and characterization by mass spectrometry. In: *Current Protocols in Molecular Biology* (eds F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl). Wiley, New York, pp. 10.22.1–10.22.24.
- Patterson, S.D. and Aebersold, R. (1995) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis* **16**: 1791–1814.
- Patterson, S.D. and Garrels, J.I. (1994) Two-dimensional gel electrophoresis of post-translational modifications. In: *Cell Biology: a Laboratory Handbook* (ed. J.E. Celis). Academic Press, New York, pp. 249–257.
- Pluskal, M.G., Przekop, M.B., Kavonian, M.R., Vecoli, C. and Hicks, D.A. (1986) ImmobilonTM PVDF transfer membrane: a new membrane substrate for western blotting of proteins. *BioTechniques* **4**: 272–283.
- Qin, J., Herring, C.J. and Zhang, X. (1998) *De novo* peptide sequencing in an ion-trap mass spectrometer with ¹⁸O labeling. *Rapid Commun. Mass Spectrom.* **12**: 209–216.
- Reinhold, V.N., Reinhold, B.B. and Costello, C.E. (1995) Carbohydrate molecular weight profiling.

- sequence, linkage, and branching data: ES-MS and CID. *Anal. Chem.* **67**: 1772–1784.
- Resing, K.A., Johnson, R.S. and Walsh, K.A.** (1995) Mass spectrometric analysis of 21 phosphorylation sites in the internal repeat of rat profilaggrin, precursor of an intermediate filament associated protein. *Biochemistry* **34**: 9477–9487.
- Roth, K.D.W., Huang, Z.H., Sadagopan, N. and Watson, J.T.** (1998) Charge derivatization of peptides for analysis by mass spectrometry. *Mass Spectrom. Rev.* **17**: 255–274.
- Schnolzer, M., Jedrzejewski, P. and Lehmann, W.D.** (1996) Protease-catalyzed incorporation of O-18 into peptide-fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass-spectrometry. *Electrophoresis* **17**: 945–953.
- Schreiner, M., Strupat, K., Lottspeich, F. and Eckerskorn, C.** (1996) Ultraviolet matrix-assisted laser-desorption ionization-mass spectrometry of electroblotted proteins. *Electrophoresis* **17**: 954–961.
- Schwartz, J.C. and Jardine, I.** (1996) Quadrupole ion trap mass spectrometry. *Methods Enzymol.* **270**: 552–586.
- Sechi, S. and Chait, B.T.** (1998) Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **70**: 5150–5158.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M.** (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci USA* **93**: 14 440–14 445.
- Shevchenko, A., Chernushevich, I., Ens, W., Standin, K.G., Thomson, B., Wilm, M. and Mann, M.** (1997) Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **11**: 1015–1024.
- Solouki, T., Marto, J.A., White, F.M., Guan, S. and Marshall, A.G.** (1995) Attomole biomolecule mass analysis by matrix-assisted laser desorption/ionization Fourier transform ion cyclotron resonance. *Anal. Chem.* **67**: 4139–4144.
- Soskic, V., Gortlach, M., Poznanovic, S., Boehmer, F.D. and Godovac-Zimmerman, J.** (1999) Functional proteomics analysis of signal transduction pathways of the platelet-derived growth factor β receptor. *Biochemistry* **38**: 1757–1764.
- Spengler, B., Luetzenkirchen, F., Metzger, S., Chaurand, P., Kaufmann, R., Jeffery, W., Bartlett-Jones, M. and Pappin, D.J.C.** (1997) Peptide sequencing of charged derivatives by postsource decay MALDI mass spectrometry. *Int. J. Mass Spectrom. Ion Processes* **169/170**: 127–140.
- Stahl, D.C., Swiderek, K.M., Davis, M.T. and Lee, T.D.** (1996) Data-controlled automation of liquid-chromatography tandem mass-spectrometry analysis of peptide mixtures. *J. Am. Soc. Mass Spectrom.* **7**: 532–540.
- Strupat, K., Karas, M., Hillenkamp, F., Eckerskorn, C. and Lottspeich, F.** (1994) Matrix-assisted laser desorption ionization mass spectrometry of proteins electroblotted after polyacrylamide gel electrophoresis. *Anal. Chem.* **66**: 464–470.
- Takach, E.J., Hines, W.M., Patterson, D.H., Juhasz, P., Falick, A.M., Vestal, M.L. and Martin, S.A.** (1997) Accurate mass measurements using MALDI-TOF with delayed extraction. *J. Protein Chem.* **16**: 363–369.
- Tempst, P., Link, A.J., Riviere, L.R., Fleming, M. and Elicone, C.** (1990) Internal sequence analysis of proteins separated on polyacrylamide gels at the submicrogram level: improved methods, applications and gene cloning strategies. *Electrophoresis* **11**: 537–553.
- Tempst, P., Posewitz, M.C. and Erdjument-Bromage, H.** (1998) Micro-sample handling as a front-end to mass spectrometry. In: *Proceedings of the 46th American Society for Mass Spectrometry Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, 31 May–4 June, 1998, p. 437.
- Traini, M., Gooley, A.A., Ou, K., Wilkins, M.R., Tonella, L., Sanchez, J.-C., Hochstrasser, D.F. and Williams, K.L.** (1998) Towards an automated approach for protein identification in proteome projects. *Electrophoresis* **19**: 1941–1949.
- Vandekerckhove, J., Bauw, G., Puype, M., Van Damme, J. and Van Montagu, M.** (1985) Protein-blotting on polybrene-coated glass-fiber sheets. *Eur. J. Biochem.* **152**: 9–19.
- Vestal, M.L., Juhasz, P. and Martin, S.A.** (1995) Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **9**: 1044–1050.
- Vestling, M.M. and Fenselau, C.** (1994a) Poly(vinylidene difluoride) membranes as the interface

- between laser desorption mass spectrometry, gel electrophoresis, and in situ proteolysis. *Anal. Chem.* **66**: 471–477.
- Vestling, M.M. and Fenselau, C.C. (1994b) Protease digestions on PVDF membranes for matrix-assisted laser desorption mass spectrometry. In: *Techniques in Protein Chemistry V* (ed. J.W. Crabb). Academic Press, San Diego, CA, pp. 59–67.
- Vestling, M.M. and Fenselau, C. (1995) Surfaces for interfacing protein gel electrophoresis directly with mass spectrometry. *Mass Spectrom. Rev.* **14**: 169–178.
- Wang, R., Chait, B.T. and Kent, S.B.H. (1994) Protein ladder sequencing: towards automation. In: *Techniques in Protein Chemistry V* (ed. J.W. Crabb). Academic Press, San Diego, CA, pp. 19–26.
- Wange, R.L., Isakov, N., Burke, T.R., Otaka, A., Roller, P.P., Watts, J.D., Aebersold, R. and Samelson, L.E. (1995) F-2(Pmp)(2)-Tam-Zeta(3), a novel competitive inhibitor of the binding of Zap-70 to the T-cell antigen receptor, blocks early T-cell signaling. *J. Biol. Chem.* **270**: 944–948.
- Watts, J.D., Welham, M.J., Kalt, L., Schrader, J.W. and Aebersold, R. (1993) IL-2 stimulation of T lymphocytes induces sequential activation of mitogen-activated protein kinases and phosphorylation of p56lck at serine-59. *J. Immunol.* **151**: 6862–6871.
- Watts, J.D., Affolter, M., Krebs, D.L., Wange, R.L., Samelson, L.E. and Aebersold, R. (1994) Identification by electrospray ionization mass spectrometry of the sites of tyrosine phosphorylation induced in activated Jurkat T cells on the protein tyrosine kinase ZAP-70. *J. Biol. Chem.* **269**: 29 520–29 529.
- Watts, J.D., Brabb, T., Bures, E.J., Wange, R.L., Samelson, L.E. and Aebersold, R. (1996a) Identification and characterization of a substrate specific for the T cell protein tyrosine kinase Zap-70. *FEBS Lett.* **398**: 217–222.
- Watts, J.D., Affolter, M., Krebs, D.L., Wange, R.L., Samelson, L.E. and Aebersold, R. (1996b) Electrospray ionization mass spectrometric investigation of signal transduction pathways: determination of sites of inducible protein phosphorylation in activated T-cells. In: *Biochemical and Biotechnological Applications of Electrospray Ionization Mass Spectrometry* (ed. A.P. Snyder), ACS Symposium Series. American Chemical Society, Washington, DC, pp. 381–407.
- Wettenhall, R.E., Aebersold, R.H. and Hood, L.E. (1991) Solid-phase sequencing of ³²P-labeled phosphopeptides at picomole and subpicomole levels. *Methods Enzymol.* **201**: 186–199.
- Whalen, S.G., Gingras, A.C., Amankwa, L., Mader, S., Branton, P.E., Aebersold, R. and Sonenberg, N. (1996) Phosphorylation of eIF-4E on serine 209 by protein kinase C is inhibited by the translational repressors, 4E-binding proteins. *J. Biol. Chem.* **271**: 11 831–11 837.
- Wiley, W.C. and McLaren, I.H. (1953) Time-of-flight mass spectrometer with improved resolution. *Rev. Sci. Instrum.* **26**: 1150–1157.
- Wilkins, M.R., Gasteiger, E., Wheeler, C.H., Lindskog, I., Sanchez, J.-C., Bairoch, A., Appel, R.D., Dunn, M.J. and Hochstrasser, D.F. (1998) Multiple parameter cross-species protein identification using MultiIdent – a World Wide Web accessible tool. *Electrophoresis* **19**: 3199–3206.
- Wilm, M.S. and Mann, M. (1994) Electrospray and Taylor–Cone theory, Dole’s beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Processes* **136**: 167–180.
- Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion-source. *Anal. Chem.* **68**: 1–8.
- Wilm, M., Neubauer, G. and Mann, M. (1996) Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **68**: 527–533.
- Woods, A.S., Huang, A.Y.C., Cotter, R.J., Pasternack, G.R., Pardoll, D.M. and Jaffee, E.M. (1995) Simplified high-sensitivity sequencing of a major histocompatibility complex class I-associated immunoreactive peptide using matrix-assisted laser-desorption ionization mass spectrometry. *Anal. Biochem.* **226**: 15–25.
- Yates, J.R. III, Speicher, S., Griffin, P.R. and Hunkapiller, T. (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**: 397–408.
- Yates, J.R. III, McCormack, A.L., Hayden, J.B. and Davey, M.P. (1994) Sequencing peptides derived from the Class II major histocompatibility complex by tandem mass spectrometry. In: *Cell Biology: a Laboratory Handbook* (ed. J.E. Celis). Academic Press, New York, pp. 380–388.
- Yates, J.R. III, Eng, J.K. and McCormack, A.L. (1995a) Mining genomes – correlating tandem mass-spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* **67**: 3202–3210.

- Yates, J.R. III, Eng, J.K., McCormack, A.L. and Schieltz, D.** (1995b) Method to correlate tandem mass-spectra of modified peptides to amino-acid-sequences in the protein database. *Anal. Chem.* **67**: 1426–1436.
- Zhang, X., Herring, C.J., Romano, P.R., Szczepanowska, J., Brzeska, H., Hinnebusch, A.G. and Qin, J.** (1998) Identification of phosphorylation sites in proteins separated by polyacrylamide gel electrophoresis. *Anal. Chem.* **70**: 2050–2059.

第六章 二维凝胶电泳的图像分析

Klaus-Peter Pleissner, Helmut Oswald and Susan Wegner

1. 介绍

二维电泳技术可根据蛋白质的等电点和分子质量从一个复杂的蛋白质溶液中分离成百上千的多肽。我们利用染色的方法，在胶上可通过大小、明暗度和位置等各异（见第四章）的蛋白斑点检测到蛋白质。几种类型的凝胶和染色方法是可行的，但在实验室之间存在差异。通常的结果是获得二维电泳的凝胶图谱，并开始进一步加工以揭示图像所含有的上千个蛋白质的信息。大量的数据以及在客观性、重复性、定量分析的需要使计算机辅助的图像分析方法成为必需。这些特殊性质（即蛋白斑点的凝胶图像）可利用常规和特殊的图像分析技术加工。2-DE 凝胶图像分析的典型流程包括以下几个步骤：①凝胶图像的扫描（数据的获取）；②图像加工；③斑点检测和定量；④凝胶配比；⑤数据分析；⑥数据呈递和解释；⑦2-DE 数据库的建立。为进一步进行以计算机为基础的图像加工，通过扫描，可将凝胶图像的“类似物”转换为数字化形式。明确从蛋白斑点中提取内在信息所需的参数是很重要的。随后图像加工的整体目标是检测胶上斑点的精确位置和决定以蛋白丰度为测量标准的蛋白斑点形状和强度。计算机图像和 2-DE 凝胶图像的分析就这样作为一种工具来提供某些信息。这种计算机图像包括：①去除噪音和人工假象的图像预加工；②每个蛋白斑点的检测和定量。除了从一块凝胶上获得的蛋白斑点信息外，同一蛋白斑点在一组凝胶上可能发生的变化对于生物学分析意义重大。如果进行比较，那么要求通过设定足够标志物的图像注册以完成比较。标志物被用来确定蛋白斑点几何转化的多型性，进而蛋白斑点可以进行相互配比。以这种方式，可以搜索到一组凝胶上相应的蛋白斑点。密度谱显示胶内斑点强度的变化。通过分析密度谱后，可以推测斑点变化的趋势。谱可以用图、表、散点图和图形表示。之后，分析的结果保存在数据库中，并与原始图像的信息链接。

对于以蛋白质组研究小组间的合作促进协同数据的共享和交换，根据其开放和联合的原则，应用软件和数据库系统是必需的。凝胶及其图像，甚至从中提取的信息数据库务必以一定可接近和整合的格式存在。这里，与国际互联网接触能帮助克服每个实验室的界线。

目前存在的所有计算机辅助的凝胶分析系统，多数发展起步于 70 年代末，依据上述所提出的策略来提取 2-DE 凝胶图像的数据。今天，许多非商品化和商品化的系统，如 QUEST/QUEST II/PDQUEST, TYCHO/KEPLER, GELLAB II, BioImage, HERMeS, PHORETIX 和 ELSIE/MELANIE II, 都是可以应用的，它们已在 2-DE 这一科学领域被众多人等所接受。支持这些系统的相关硬件在过去 20 多年间经历了戏剧性的变化，并且仍旧迅速演变，但其应用的基本算法是一致的。计算机凝胶分析前人已提到 (Anderson and Anderson, 1996; Anderson *et al.*, 1981; Appel *et al.*, 1991, 1997a,

b; Bossinger *et al.*, 1979; Collins and Blose, 1992; Garrels, 1979, 1983, 1989; Lemkin and Lipkin, 1981a ~ c; Lemkin *et al.*, 1979, 1982; Miller *et al.*, 1982; Olson and Miller, 1988; Prehm *et al.*, 1987; Serpico and Vernazza, 1987; Skolnick, 1982; Vincens and Tarroux, 1987a, b; Vo *et al.*, 1981)。本章节主要描述 2-DE 图像分析的概况及其主要的特征, 对于更好的理解图像分析是重要的。

2. 数据获取

2.1 获取图像设备

通过扫描 2-DE 凝胶来获取数据, 并将其转换成一个数字化的文件呈递。很明显, 数据获取明显预示最终的分析质量。两个技术参数对于得到令人满意的转换至关重要: ①数字化的空间分辨率; ②灰度值的密度分辨率。一些图像的获取设备可将凝胶转换成数据, 并可获得较好的质量, 例如数码相机、扫描仪、磷储屏和荧光图像分析仪(见表 6.1)。

表 6.1 用于二维电泳的主要图像获取设备

信息载体	扫描装置			
	数码相机/CCD	密度仪(白/激光)	磷储屏图像分析仪	荧光图像分析仪
X 光片(放射标记的样品)	×	×	×	
磷储屏			×	
银染或彩染凝胶	×	×	×	
荧光标记的凝胶	×			×

各种类型的照相机, 大多数情况下不但用于凝胶的“文件化”, 而且用于扫描。照相机的质量依靠其输入和输出的特性。以白光和激光来区分的密度仪或扫描仪多数情况下适合于获取从 2-DE 凝胶图像的数字化的数据。它们具有高密度、高空间分辨率及线性特征, 并可用于银染或考马斯亮蓝染色的干、湿凝胶。放射性标记(^{35}S , ^3H , ^{14}C , ^{32}P)样品的二维电泳图谱通过 X 光片或磷储屏图像分析仪获取。当应用 X 光片的时候, 在密度或灰度分辨率上必须扩展观察其动态变化, 即通过同一凝胶的多次曝光来获得尽可能多的蛋白斑点。不同曝光的片子可融合成一个数字化的 2-DE 凝胶图像。每一组灰度值以线性方式结合以增加动态的密度范围。为获得与放射性相关的光密度(disintegration/min, d. p. m.), 需要同时测量放置在 X 光片上的校正条, 并产生标准曲线。磷储屏图像分析仪提供了一种代替 X 光片和各种曝光底片融合的方法, 因为它们具有高动态范围, 跨越 5 个数量级, 进行自显影(Patterson and Latter, 1993)。磷储屏图像分析系统联合 QUEST II 系统的应用前述已提及(Amess and Tolkovsky, 1995)。然而, 磷储屏图像分析仪受限于放射性标记的蛋白混合物。

3. 数字化图像加工

图像增强被广泛应用于计算机图像领域, 平滑操作、增强对比度、边缘检测和背景消

减是图像加工最熟悉的工具。这些操作的重要性依赖于所应用的领域。然而,在某些领域,这些操作主要用于图像增强,以提高图像的光学质量,分析科学可依靠这些操作提供明确的数据。例如,在 2-DE 凝胶的数字化图像上分析蛋白斑点。这里,平滑操作、增强对比度、边缘检测和背景消减可允许以一种精确的方式来决定蛋白斑点的形状和大小。

不同的操作正常定义在一个图像点或像素“邻居”区域。通常所应用的邻居是 3×3 、 5×5 或 7×7 个像素。例如,在一个代表数字化图像的有规律格子基础上,一个 3×3 的邻居是由界定图像中心点的那些像素组成。

平滑操作被用来清除图像中的统计噪音。主要策略是在一个图像点所定义的邻近区

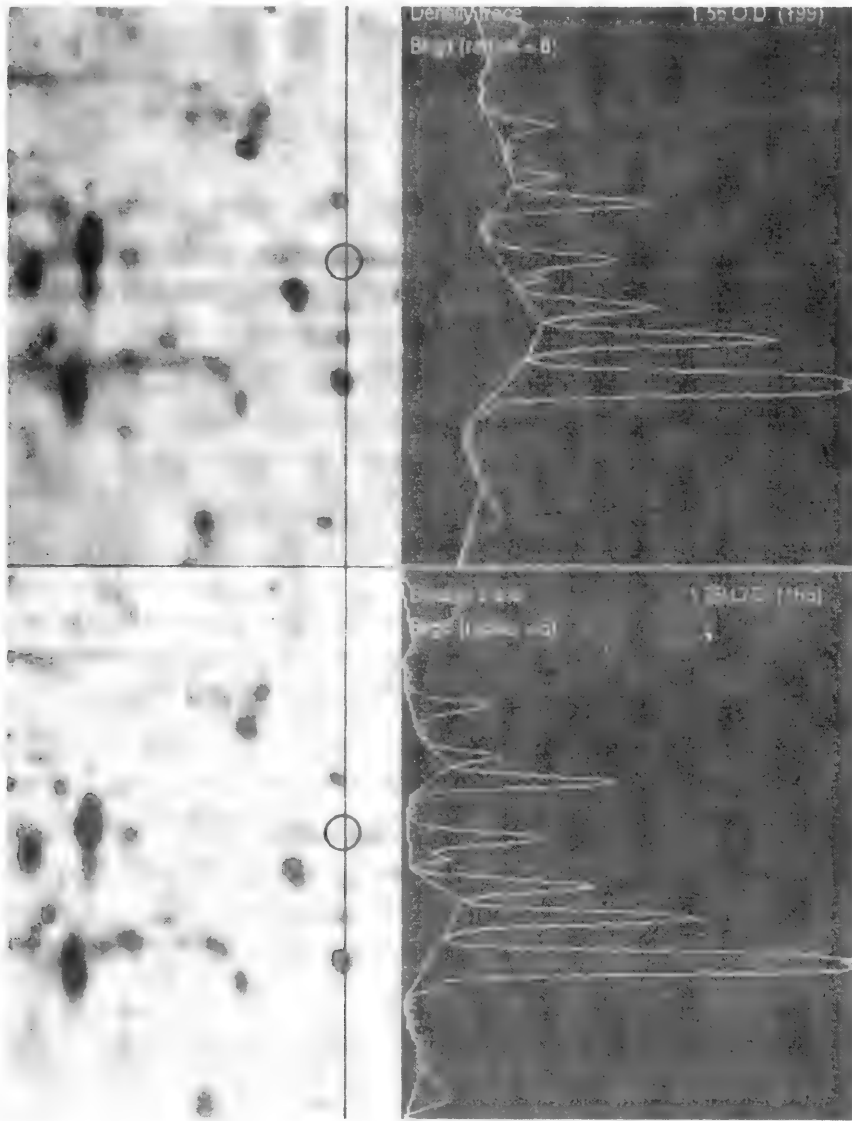


图 6.1 利用“滚动圆盘”的几何模型进行背景消减和条纹清除
球的半径与斑点的大小相适应。本图的上部显示沿着选定线上斑点的背景和密度踪迹。下部说明背景消减后同样的特征。

域内,高斯平滑、平均值或中数的确定。增强对比度重新定义一个图像的灰度,以获得斑点和背景之间较高的对比。边缘检测的目的是限定某一图像中的灰度值变化。

通常,图像功能的第一或第二代衍生物是计算机化的。各种操作被用在图像点周围的小邻近区域居以决定衍生物。典型技术是,在第一代中应用梯度操作,而在第二代衍生物中应用拉普拉斯(Laplace)操作。衍生系列对于噪声敏感性好,因此,通常与高斯平滑联合应用。

背景消减常用于清除背景的无意义变化,例如图像中不同区域的不均一或各种背景水平。一个策略是取背景最黑和最亮区域的平均值,并应用于整个图像。另一个比较复杂的方法被称为“滚动的圆盘”(Sternberg, 1986),可以除去图像中的背景和条纹(见图6.1)。因此一个小半径不但可产生很清晰的图像,而且还可除去最弱的斑点,从而影响这种清除效率是可能的。

4. 蛋白质斑点检测和定量

请记住这次练习的主要目的是分析2-DE蛋白凝胶、斑点检测及其形状模建,这是一个重要任务。斑点检测产生 x/y 位置的 n 数组、形状参数和整合的斑点强度。蛋白斑点有几个光学特征。它们可能分别出现,但在原始凝胶图谱及其随后的计算机图像中是互相接触或重叠的。它们也可能出现在由于二维电泳和凝胶染色所导致的技术问题或背景不同所致的条纹和污点内。它们可能没有清晰的轮廓,因此看来是一种圆形或椭圆形的连续分布。斑点形态依靠于凝胶制备、染色步骤和凝胶分析系统的数字化误差。每一个斑点、背景的灰度变异和缺乏同质性要求令人满意的算法来进行斑点检测和定量。确实存在这种应用广泛的算法,并被分为高斯拟合(Gaussian fitting)和高斯拉普拉斯(Laplacian of Gaussian, LOG)变换斑点检测(Anderson and Anderson, 1981; Appel *et al.*, 1997b; Lemkin and Lipkin, 1981a, b; Wu *et al.*, 1994)。

另一种斑点检测的方法是由Garrels最先提出的线性分析和链式分析算法而完成(1979)。在这个方法中,每一个垂直的扫描线被用来细察密度高峰。邻近扫描线的峰值可组装成“链状”,并且可决定斑点的中心位置。定位斑点之后,其丰度(整合的斑点强度)可通过对斑点形状进行数学模型拟合来决定。二维的高斯模建经常被应用。显著偏离单一高斯图形的斑点可以被认为一组高斯形状的重叠。

另外一种主要应用于数码相机图像分析的斑点检测方法已经被描述(Prehm *et al.*, 1987)。预加工包括底纹校正和非回归的、分离的、低流通的平滑滤镜。一个分割的步骤假定斑点属于具有凸出曲率的灰度表面区域。当应用分割模板(9×9 凸面检测器)时,如果中心像素 9×9 附近的曲率大于零,则像素描述了一个斑点的中心。否则,图像上的斑点属于背景。

相互位置极端靠近的斑点、或者那些肩靠肩、或背包样(一个小斑点附着在大斑点上)的斑点,通常难以分离,因为斑点间的灰度区域不能表明凹面曲率。然而,在这些区域的曲率大小低于斑点区域的曲率。因此,为了分离这些斑点,一个 7×7 像素的模板被应用,并且可计算在中心像素和以 3×3 邻近的8个方向 7×7 模板边缘的像素之间的曲率大小的差异。如果两个曲率大小之差小于至少一个方向上的固定阈值,那么,中心的像素则被认为是背景。通过这个分割过程所发现的斑点中心通过两个像素加以扩大。斑点丰度取

决于扩大斑点中心内所有像素的光密度之和。

其他进行斑点检测的方法基于转化分界线(Watershed transformation, WST)(Bettens *et al.*, 1997; Meyer and Beucher, 1990; Pleissner *et al.*, 1999; Vincent, 1993; Wegner *et al.*, 1998)。将分水岭这一地理学的原理应用于图像,灰度图片被认为是拓扑样地貌,由数值代表着某一点的升高。在计算算法中,我们的解释是一种浸入过程。这可以假设为在地貌的每个最小区域进行钻孔,并且将表面浸入湖泊中,并确保在山谷中所有灰度水平峰值有一个恒定的水平面。从最低海拔开始,水将充满不同的山谷。在这个浸入过程结束的时候,每一个最小值完全被分水岭所包围,以划界相关的山谷。这山谷相对应的区域,也就是分水岭定义某一图像的最佳轮廓。在图像中用来进行梯度分割的分水岭必须以地貌来解释。假设图像中均一的区域以低对比度,即低梯度为特征;反之梯度图像的高数值是由原始图像高对比度的轮廓产生。这些区域可以利用 WST 在梯度图像上产生。随后,每一区域是以完成一个马赛克图像的特征来描述。所选择的特征应该与原始图像的灰度水平一致,例如原始图像上区域灰度值的平均值。

WST 产生封闭的轮廓。每一个最小值(峰底)描绘一个由局部最高点(汇流)所围成的区域。这样一种分割的结果很可能是一种过度分割。那意味着,与相关区域的轮廓一样,主要由图像噪音所致的轮廓也会被发现。

在过度分割的马赛克图像内,以下两种类型斑点区域将被用来鉴定所要求的蛋白斑点:与一完整斑点相对应的区域或者部分不完整斑点的区域。为了减少在马赛克图像内候选蛋白斑点的数目,应用了两个阈值,这基于以下假设,斑点区域的灰度值显著高于背景的灰度值,并且它们与背景区域有一个界线。所要求的阈值可自动适应每一个新的图

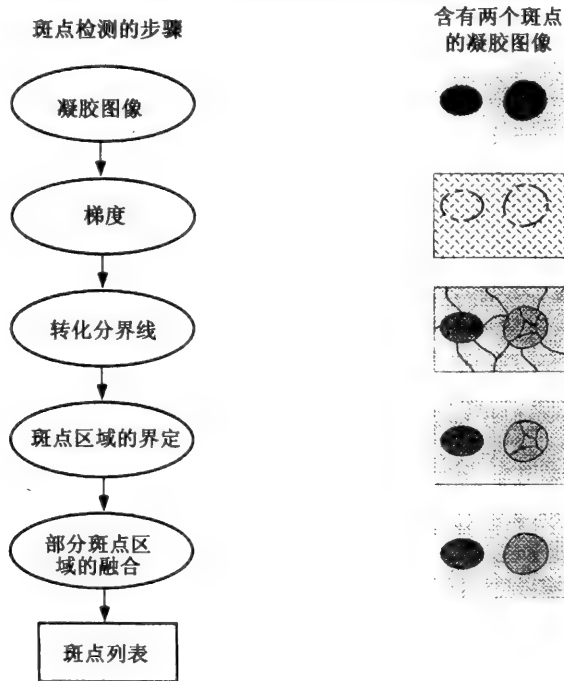


图 6.2 利用分水岭转化进行斑点检测的主要步骤

像。

应用这些标准后,仍存在一些区域,既不是斑点区域,也不是部分斑点区域。把一凝胶图像认为是一个表面时,斑点的形状是明显凸起的,部分斑点区域也是如此。由于第二代计算机衍生系列描述了分水岭拓扑地貌的曲率,斑点和部分斑点区域就可以采用数学方法求解。斑点区域和部分斑点区域具有相同的凸起曲率。最后,如果它们来源于同一斑点,邻近斑点组的所有部分斑点区域被检查、融合。斑点形状应类似于椭圆,并且一个斑点的部分斑点区域在其分水岭线的邻近区域内应有相应的凸起曲率。这些斑点特征的结合被用来作为融合的标准。这种斑点检测算法的主要步骤见图 6.2。

5. 凝胶配比

除了每一块凝胶图像的斑点检测和定量外,一组凝胶上同一斑点的比较是重要的。因此,在各种实验条件下蛋白表达的比较性分析成为一个重要的问题。凝胶的不均一性发生是由于以下差异所导致的,如样品制备、电泳条件或温度的变化以及在不同凝胶上电泳的不同迁移率。上述原因导致无法产生完美无暇的凝胶图谱。因此,为了分析蛋白的变化,斑点图谱必须配比。

进行电泳或斑点图谱配比的前提是注册。注册是一种通过转化的方式使数据进行比较的步骤。正常情况下,一块凝胶上蛋白斑点的位置很大程度依赖于各种可变的参数,如凝胶的质地和密度。尽管许多研究者雄心勃勃以获得“完美”的凝胶,但这些凝胶在化学和物理特征上仍有变化。因此,当进行凝胶配比的时候,蛋白位置并不十分相同,因此它们的信息值不是直接可比的。使图像变形的参数需要从图像中清除,将其转变成可比较的图像。这已通过转化的方式完成,并且随后可进行配比。

配比有两种方式:①在像素水平;②在斑点水平。在第一种情况下,两种凝胶图像的像素绘成图形;在第二种情况下,仅对两块或多块凝胶上的蛋白斑点进行配比,以达到整个凝胶的配比。上述两种方法,即局部或全局的,多项式的转化被用来绘成一个源图像或源斑点图案,或者被用来绘成靶图像或靶斑点图案。

利用标志物(或 passpoints)在图像和亲合转化上进行注册,如调整大小、伸展、旋转和变换。例如标志物是在两张凝胶上都被鉴定的特征性斑点。转化的方法对整个图像的应用可为两块凝胶图像在像素水平上的定量差异提供可视化的比较(Appel *et al.*, 1997b; Lemkin, 1997a; Pleissner *et al.*, 1997a)。

在斑点水平的应用基于斑点检测的信息:位置、形状和丰度可经常在“斑点列表”中发现。斑点水平的配比策略见图 6.3。正确配比的前提是图像或图像的区域具有很高的相似性。配比的算法主要首先假定为在一个比较性研究中凝胶图像属于同一条件下产生的一组凝胶,许多算法可用于这种凝胶的配比,它们是建立在几何、统计、图形理论和错列组合图谱识别法的基础之上。

正像上面所提到的,几何方法将特征性的蛋白斑点作为标志物,是一个特殊的方式通过扩散来配比。这里,从标志物而来的配比信息在它们附近延伸到邻近的斑点。对任何一对凝胶而言,在标志物附近发现最近的邻居是可能的,对邻近的斑点,其相对于标志物的在 x, y 轴上的位移是可确定的。随后,在其他的凝胶上利用与相应的标志物相关的同样位移进行搜索。如果一个斑点在这个位置被发现,即配比成功。通过扩散进行的配比

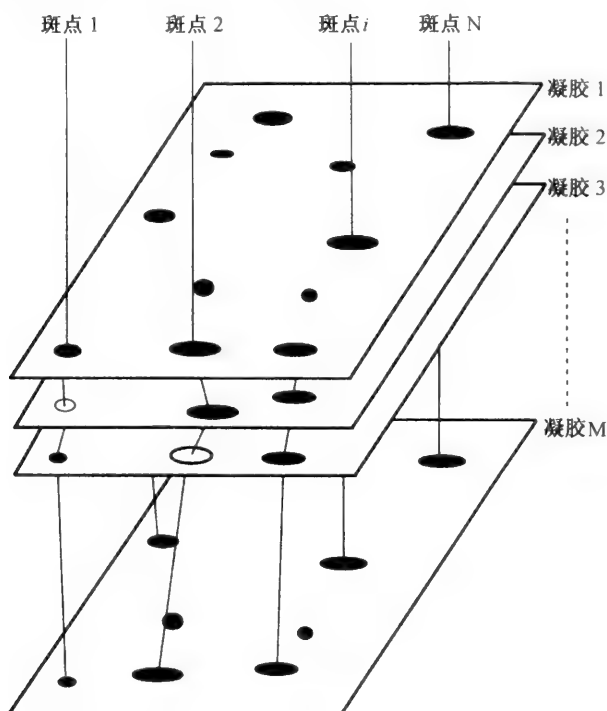


图 6.3 配比的原理(在斑点水平上)

是可回归的,这意味着通过扩散所配比的斑点能成为进一步扩散的中心。

配比的质量可通过两块凝胶之间进行直接配比的结果与所有凝胶的间接配比所进行的交叉配比来提高。此外,加入斑点也可以提高配比的精确度和质量。如果一个斑点出现在源图像,但在靶图像中消失,那么这一斑点将被加入到靶图像的相应位置上(Garrels, 1979, 1989)。

在统计学方法基础上利用相似性原则的一种强有力的配比算法已由 Olson 和 Miller (1988)提出,并且在 MELANIE II(Appel *et al.*, 1997b)这种软件中得到提高和完成。进一步的配比方法是以图形为基础(Kuick *et al.*, 1991; Skolnick, 1982; Skolnick and Neel, 1986)和利用错列组合图案识别方式的语法为基础的技术(Tarroux *et al.*, 1987; Vincens and Tarroux, 1987a, 1987b)。

对于实际的配比方法,凝胶需要额外聚集在一组凝胶中(match sets)。配比的过程导致单个斑点的追溯。斑点及其强度的变化可以通过一系列凝胶进行追寻。凝胶配比在凝胶的图像分析中是最关键的步骤之一。只要凝胶很相似,即当它们在同一条件下产生于同一实验室,多数的算法是成功的。从不同来源的凝胶图像的比较经常导致不相容性。

在某些系统中,数据分析通过配比组得以支持,它由几组凝胶组成(对照组、实验组),它们已经通过配比的加工。有时候附加的特征通过衍生于模板图像的标准图像来提供。包括高斯模建斑点的标准图像含有所有系列凝胶的斑点和数据(均包含于配比组中)。

6. 数据分析

数据分析主要由 3 个问题所驱动:①质的确证性的评估;②量的确定;③蛋白表达中定量变化的决定。每块凝胶中有效的蛋白斑点数目、强度饱和的斑点数目和配比的斑点数目的信息在质量评估中是基本的。应在所有凝胶中存在斑点总数和配比斑点之间的关系,以明确定义质的原则;斑点的总数越高,某一特殊的蛋白斑点不能在所有凝胶中发现的重要性就越小。

为了确定蛋白表达中的量变和质变,在一个可见的 $N \times M$ 的矩阵表中列出了一组凝胶中所有单个斑点的强度值(图 6.4)。 N 代表斑点的数目(行), M 代表配比的凝胶斑点数目(列)。在最初 $N \times M$ 矩阵上的原始数据以斑点的强度来代表。因为凝胶图像的平均强度不同,例如不同的曝光时间,所以在定量分析中必须进行校正(Burstin *et al.*, 1993)。单个斑点强度的校正基于相关的凝胶图像的平均强度。

		对照组				实验组	
		凝胶 1	凝胶 2	凝胶 M-1	凝胶 M
斑点	斑点 1	$I_{1,1}$	$I_{1,2}$			$I_{1,M-1}$	$I_{1,M}$
	斑点 2	$I_{2,1}$	$I_{2,2}$			$I_{2,M-1}$	$I_{2,M}$
	▪						
	▪						
	▪						
	▪						
斑点 N		$I_{N,1}$	$I_{N,2}$			$I_{N,M-1}$	$I_{N,M}$

图 6.4 配比之后的数据分析表(矩阵)

斑点的原始数据(整合的光密度)在 $N \times M$ 矩阵内储存,在此 N 代表斑点的数目(行数),而 M 代表凝胶的数目(列数)。

为了决定蛋白斑点定量变化,人们搜寻在强度上增高、下降或不变的斑点。定量分析可利用不同的统计方法进行,例如通过平均值计数的组间参数变化和斑点强度的偏离。进而,组间的变化可以通过单方差的形式进行统计检验(例如 Student's t -test, Log t -test 和 Mann-Whitney tset)或者多方差的方式(相关分析)。

然而,单方差检验考虑每一行,并不依赖于 $N \times M$ 数据阵列;而多方差方法考虑斑点的相关性(行),相关分析能检测凝胶间的相似性,并鉴定其典型的斑点特征。用于 2-DE 的相关分析的数学背景首先由 Pun 等(1988)提出。这个方法目的是决定最代表这一凝

胶特征的斑点。Schmid 等(1995)已在细节上描述了相关分析,利用 2-DE 蛋白图谱,将之应用到细胞分类。我们已经对从 PDQUEST 系统所产生的数据进行了额外的相关分析(Pleissner *et al.*, 1998)。整体上,单方差和多方差的统计方法对于 2-DE 凝胶中的蛋白表达变化的分析是重要的工具。

对凝胶组之间蛋白表达的定性分析通过寻找仅在一个组中出现的斑点即可。这些斑点经常被称为“开/关”斑点,因为这些斑点在某些特定的实验条件下开与关,并且它们是表明蛋白变化的明显候选者,所以非常重要。

一个凝胶分析系统分析能力的重要特征是使用者的数据处理,这是使用者定义的斑点的集合。此外,不同的分析步骤的结果可以被联合,例如交集操作(boolean operators)。以这种方式,在所有的凝胶中,决定可被发现和被配比上的非饱和斑点是可能的。此外,在使用者中已配比的斑点由于蛋白的变化可能增加或降低,很可能被发现具有统计意义。这种数据分析工具的应用已经在人心脏疾病的分析中描述(Pleissner *et al.*, 1997b)。

7. 数据呈递

为了精确呈递结果,需要考虑一些前提条件。首先,每个被检测的蛋白斑点应该有一个分子量(M_r)和等电点(pI)。因此,利用某些特征的蛋白斑点(参考胶上的斑点),一种 M_r /pI 晶格形式就覆盖在凝胶上,从而推测 M_r /pI 值。这些参考蛋白斑点是已知蛋白,并易在胶上检测。通常, M_r /pI 值在供应者目录排列,只要蛋白名称或蛋白注册号已知,它们也能从 SWISS-PROT (ExPASy 分子生物学服务器)上获得。例如:在人心脏样品的二维凝胶中,血清蛋白、肌动蛋白、肌球蛋白和肌动蛋白轻链 1 和肌红蛋白明显易被识别,并且可作为产生 M_r /pI 晶格的标志物(Knecht *et al.*, 1994)。

利用这些标志物蛋白,所有其他蛋白的 M_r /pI 值可以通过假设沿着 pI 轴(水平轴)的一种线性外推和沿着 M_r 轴(垂直轴)的一种指数外推得到评估。很清楚, M_r /pI 值的精确性主要依赖于用于计算 M_r /pI 晶格的算法。

曲线图或条形图表常被用来比较斑点的量变,反应斑点变化的趋势。理想上,这种变化应来源于基本的蛋白表达变化。此外,在检测非配比斑点和在定量上可疑高于平均值的斑点,曲线图或条形图表也是有用的。条形图表在 x 轴上代表凝胶的斑点数目,在 y 轴上代表斑点强度(丰度)。此外,如果凝胶形成了复制(replicate)组,组内平均强度值和标准偏离可以条形图表的形式说明。利用这种数据呈递,组间的变化趋势可以清晰识别。因此,这些图表能提供一个有力的图形工具,以代表储存于配比表格中的数据。这些图形中的某些类型,显示对照组和实验组间的斑点差异在图 6.5 中说明。此外,以表格形式的斑点强度输出,通常以 EXCEL、Stat View、SSP、SAS 或其他形式应用而进一步评估。

为了评估斑点强度间的相似性,散点图经常被用来显示两块凝胶或两组凝胶间的强度相关性。应用这种图表,依据强度的相关性,相关系数可以被计算。例如,10 个对照组与实验组凝胶被分析,可组建一个研究高血压影响的散点图(Pleissner *et al.*, 1998)。当比较两组中所有已配比斑点的平均强度时,可确定两组的明显相似性(图 6.6)。

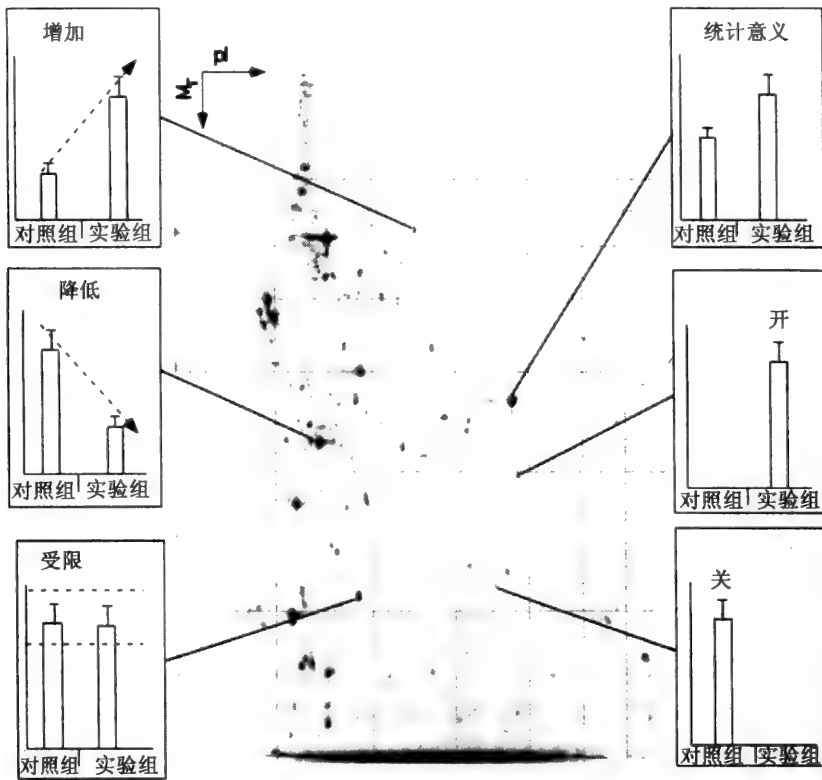


图 6.5 某些类用来表达蛋白变化的图表

每组(对照、实验)由几块凝胶组成,例如:条形图代表组内斑点强度的平均值和标准差。

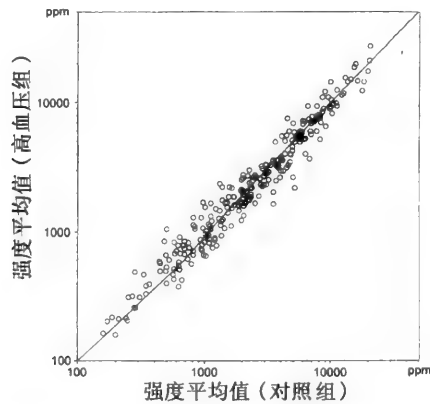


图 6.6 表示实验组(高血压)和对照组间斑点强度平均值依赖性的散点图
已配比的每块凝胶的斑点数目为 269;凝胶的数目为 20。

8. 数据库

8.1 2-DE 数据库的构建

前面所描述的分析提供了关于蛋白与凝胶图像中蛋白斑点间的许多描述性信息,即蛋白名称、 M_r /pI 值和目录性信息和其他文献。通常,信息在2-DE 数据库中储存。数据库正常情况下对所分析的一组图像的代表图像给予明确的注释。这个代表性图像能成为被分析的凝胶图像之一或者是产生一块综合性的标准图谱。这种图像-数据库和数据库-图像的链接提示:①如果一个蛋白斑点在图像中标出,应该出现其相应的描述信息;②如果这个蛋白的名称或其他描述性信息已知,那么必须显示这个斑点在图像上的位置并加以解释。

一些凝胶分析系统包括它们自己的内部数据库。例如在 PDQuest 系统中,除了标准斑点号(SSP)以外,通常一个数据库的斑点号(DSN)表明了这个斑点的位置,因为这个号码基于 M_r /pI 晶格所排序。可每个斑点归于不同的注释类别(annotation category)。这种注释是直接和标准胶上斑点的位置有关,并能通过简单查询检索得到。尽管这些内部的数据库(Lipkin, 1980; Jungblut *et al.*, 1994)是有帮助的,但它们是系统特异的,并不允许凝胶图像和描述信息共享。

通过网络技术的接触可帮助克服实验室的界限。在 WWW(world wide web)网上所产生的 2-DE 数据库已描述(Appel *et al.*, 1993; Boutell *et al.*, 1994; Evans *et al.*, 1997; Latter *et al.*, 1995; Lemkin, 1997b; Monardo *et al.*, 1994; Miller *et al.*, 1996; Pleissner *et al.*, 1996)。目前,大约 30 个 WWW 可及的 2-DE 数据库可在 ExPASy 服务器的 WORLD-2DPAGE 上检索到(<http://www.expasy.ch/ch2d/2d-index.html>)。对于 2-DE WWW 数据库的构建、交互界面和整合的概况已由 Appel 描述(1997c)。

目前,某些组建 2-DE 数据库的软件包可免费应用,如 Make2dbb (Hoogland *et al.*, 1997)。这个软件包可以不依赖 MELANE II 系统而应用,还有一个新工具,它使凝胶的 TIFF 图像、含有 x/y 参数坐标的蛋白斑点的 text 文件和 SWISS-PROT 的注册号转换成特定的图像格式(<http://www.expasy.ch/ch2d/tiff2mel.html>)。

8.2 比较网上分布的 2-DE 数据库

伴随着信息转换,通过因特网进行的凝胶图像比较在蛋白质组研究中是一项重要的任务,并可支持或替代昂贵的蛋白质鉴定。产生数据库的 master 凝胶,并利用某一蛋白的 SWISS-PROT 注册号,这些斑点能在联邦 2-DE 数据库中定位和鉴定(Appel *et al.*, 1996)。在 2-DE 中间数据库中,以两块凝胶图像间可视化的实验室间比较为特征,利用翘曲(warping)转换(如配比)、增强对比、平滑过滤和颤动(flickering)技术得以在 Flicker 服务器上实现(<http://www.lmn.ncicrf.gov/flicker>; Lemkin, 1997a)。颤动意味着可以交替看见两块图像(源与靶图像)。

Pleissner 等已详细描述在 WWW 数据库中可利用的自动化比较/配比的 2-DE 凝胶图像的问题(1999)。在这种情况下,当利用一种被称为计算几何学的方法时,斑点参数(斑点位置和强度)被用来作为点图谱的配比。此外,通过用户界面的方式,求助于任何一

个具有 Java 能力的因特网浏览器,斑点检测和配比策略可直接在网上进行。这种配比方法通过检查两图像间点的边缘斜率和长度,也考虑斑点的强度关系,来比较源图像和靶图像的点图。配比算法的一个中心目标来源于利用增量德劳内(Delaunay)三角测量,即从计算几何学而知(Alt and Guibas, 2000)。将强度绝对值向不连续强度整数的转化和不连续阈值的应用,使这种较弱地依赖于灰度值的配比方法成为可能(绝对的密度分辨率)。

这种配比方法已经在 CAROL 软件系统得以完成(<http://gelmatchy.inf.fu-berlin.de>)。这个程序能够通过整个因特网配比 2-DE 数据库中的凝胶图像。使用者从任何一个 2-DE 数据库中打开 GIF 图像,进行斑点检测,并在源与靶图像中进行局域的配比。图 6.7 展示在比较从 WORLD-2 DPAGE 检索的 2-DE 数据库 HEART-2DPAGE (左,源)和 HSC-2DPAGE(右,靶)而来的凝胶图像时, CAROL 用户界面的应用。

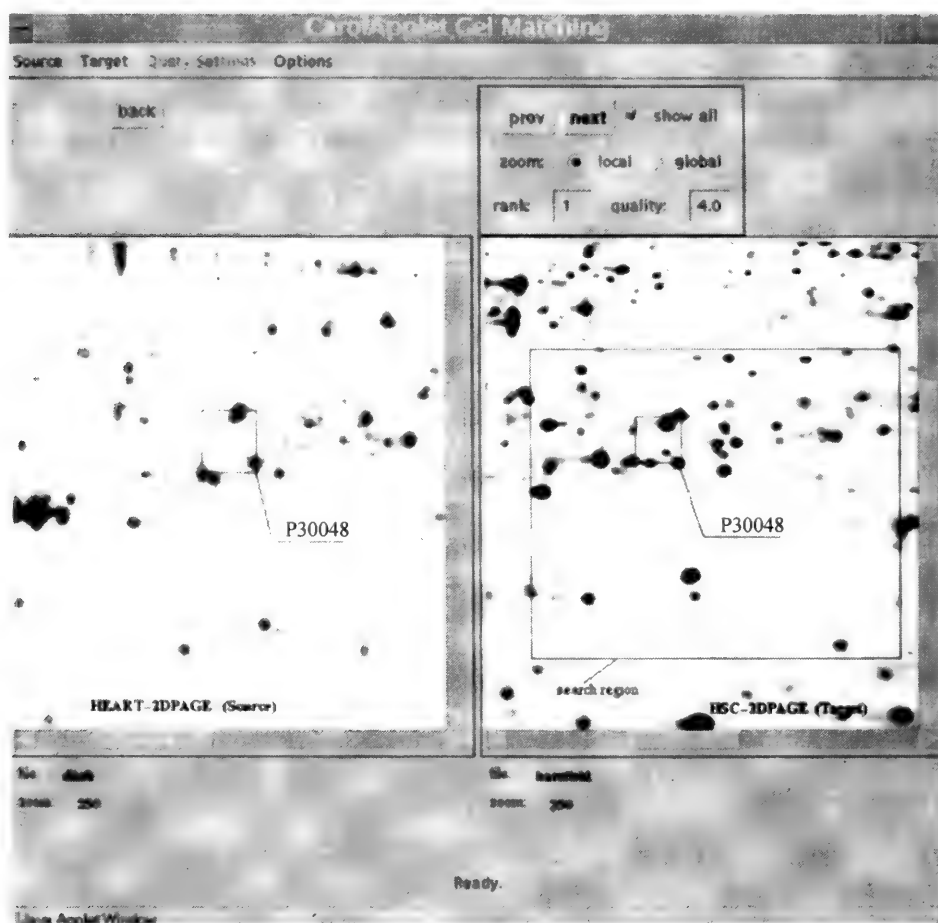


图 6.7 被用来比较 HEART-2DPAGE(源)和 HSC-2DPAGE(靶)数据库的 CAROL 用户界面
搜索图案(+)在源图像中设定,相应的图案在配比后设定。当进行配比时,在两个数据库发现的
SWISS-PROT AC 被显示出来以说明配比结果的正确性。

9. 结论

本章已概括了计算机图像分析 2-DE 凝胶的主要原理,以及 2-DE 数据库的建立和在

WWW 网上的比较。毫无疑问,如果没有计算机图像分析,蛋白斑点的精确而可重复性的评估很难进行。各种各样的商品化凝胶分析系统目前是应用的,用来决定在特定状况下的蛋白斑点变化。一旦在蛋白变化方面具有足够的实验数据,即可检测特定的蛋白标志物,并进行进一步研究。数据库的构建和比较的概念也被描述,蛋白数据和凝胶图像可通过网络被储存、共享和比较。在这个背景下,数字化的图像加工和分析提高了不同 2-DE 数据库的可比性。

感谢

作者感谢德国柏林心脏研究所的 Vera Regitz-Zagrosek、Bernhard Krudewagen、Eckart Fleck 和柏林自由大学的 Frank Hoffmann、Klaus Kreigel、Carola Wenk、Helmut Alt 的科研协作和支持。

(万晶宏译 贺福初校)

参考文献

- Alt, H. and Guibas, L. (2000) Discrete geometric shapes: matching, interpolation, and approximation, a survey. In: *Handbook for Computational Geometry* (ed. J. Urrutia and J.-R. Sack). North Holland, Amsterdam, pp. 121–154.
- Amess, B. and Tolkovsky, A.M. (1995) Programmed cell death in sympathetic neurons: a study by two-dimensional polyacrylamide gel electrophoresis using computer image analysis. *Electrophoresis* **16**: 1255–1267.
- Anderson, N.G. and Anderson, N.L. (1996) Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* **17**: 443–453.
- Anderson, N.L., Taylor, J., Scandora, A.E., Coulter, B.P. and Anderson, N.G. (1981) The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin. Chem.* **27**: 1807–1820.
- Appel, R.D. (1997) Interfacing and integrating databases. In: *Proteome Research: New Frontiers in Functional Genomics* (eds M.R. Wilkins, K.L. Williams, R.D. Appel and D.F. Hochstrasser). Springer, Berlin, pp. 149–175.
- Appel, R.D., Hochstrasser, D.F., Funk, M., Vargas, J.R., Pellegrini, C., Muller, A.F. and Scherer, J.-R. (1991) The MELANIE project: from a biopsy to automatic protein map interpretation by computer. *Electrophoresis* **12**: 722–735.
- Appel, R.D., Sanchez, J.-C., Bairoch, A., Golaz, O., Miu, M., Vargas, J.R. and Hochstrasser D.F. (1993) SWISS-2DPAGE: a database of two-dimensional electrophoresis images. *Electrophoresis* **12**: 1232–1238.
- Appel, R.D., Bairoch, A., Sanchez, J.-C., Vargas J.R., Golaz, O., Pasquali, C. and Hochstrasser, D.F. (1996) Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. *Electrophoresis* **17**: 540–546.
- Appel R.D., Palagi, P.M., Walther, D., Vargas, J.R., Sanchez, J.-C., Ravier, F., Pasquali, C. and Hochstrasser, D.F. (1997a) Melanie II – a third-generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface. *Electrophoresis* **18**: 2724–2734.
- Appel, R.D., Vargas, J.R., Palagi, P.M., Walther, D. and Hochstrasser, D.F. (1997b) Melanie II – a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. *Electrophoresis* **18**: 2735–2748.
- Bettens, E., Scheunders, P., VanDyck, D., Moens, L. and Van Osta, P. (1997) Automatic segmentation and modelling of two-dimensional electrophoresis gels. *Electrophoresis* **18**: 792–798.
- Bossinger, J., Miller, M.J., Vo, K.-P., Geiduschek, P. and Xuong, N. (1979) Quantitative analysis of two-dimensional electrophoretograms. *J. Biol. Chem.* **254**: 7986–7998.
- Boutell, T., Garrels, J.I., Franza, B.R., Monardo, P.J. and Latter, G.I. (1994) REF52 on the global

- gel navigator: an internet accessible two-dimensional gel electrophoresis database. *Electrophoresis* **15**: 1487–1490.
- Burstin, J., Zivy, M., de Vienne, D. and Damerval C.** (1993) Analysis of scaling methods to minimize experimental variations in two-dimensional electrophoresis quantitative data: application to the comparison of maize inbred lines. *Electrophoresis* **14**: 1067–1073.
- Collins, P.J. and Blose, S.H.** (1992) Today's densitometer – multi-application densitometry for biotechnology. *Int. J. Biotechnol.* **9**: 202–205.
- Evans, G., Wheeler, C.H., Corbett, J.M. and Dunn, M.J.** (1997) Construction of HSC-2DPAGE: a two-dimensional gel electrophoresis database of heart proteins. *Electrophoresis* **18**: 471–479.
- Garrels, J.I.** (1979) Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by clonal cell lines. *J. Biol. Chem.* **254**: 7961–7977.
- Garrels, J.I.** (1983) Quantitative two-dimensional gel electrophoresis of proteins. *Methods Enzymol.* **100**: 411–423.
- Garrels, J.I.** (1989) The QUEST system for quantitative analysis of two-dimensional gels. *J. Biol. Chem.* **264**: 5259–5282.
- Hoogland, C., Baujard, V., Sanchez, J.-C., Hochstrasser, D.F. and Appel, R.D.** (1997) Make2ddb: a simple package to set up a two-dimensional electrophoresis database for the World Wide Web. *Electrophoresis* **18**: 2755–2758.
- Jungblut, P., Otto, A., Zeindl-Eberhart, E., Pleissner, K.-P., Knecht, M., Regitz-Zagrosek, V., Fleck, E. and Wittmann-Liebold, B.** (1994) Protein composition of human heart: the construction of a two-dimensional electrophoresis database. *Electrophoresis* **15**: 685–707.
- Knecht, M., Regitz-Zagrosek, V., Pleissner, K.-P., Jungblut, P., Hildebrandt, A. and Fleck, E.** (1994) Computer-assisted analysis of endomyocardial biopsy protein patterns by two-dimensional gel electrophoresis. *Eur. J. Clin. Chem. Clin. Biochem.* **32**: 615–624.
- Kuick, R.D., Skolnick, M.M., Hanash, S. and Neel, J.V.** (1991) A two-dimensional electrophoresis-related laboratory information processing system: spot matching. *Electrophoresis* **12**: 736–746.
- Latter, G. I., Boutell, T., Monardo, P.J., Kobayashi, R., Futcher, B., McLaughlin, C.S. and Garrels J.I.** (1995) A *Saccharomyces cerevisiae* Internet protein resource now available. *Electrophoresis* **16**: 1170–1174.
- Lemkin, P.F.** (1997a) Comparing two-dimensional electrophoretic gel images across the Internet. *Electrophoresis* **18**: 461–470.
- Lemkin, P.F.** (1997b) The 2DWG meta-database of two-dimensional electrophoretic gel images on the Internet. *Electrophoresis* **18**: 2759–2773.
- Lemkin, P.F. and Lipkin, L.E.** (1981a) GELLAB: a computer system for 2D gel electrophoresis analysis I. Segmentation of spots and system preliminaries. *Comput. Biomed. Res.* **14**: 272–297.
- Lemkin, P.F. and Lipkin, L.E.** (1981b) GELLAB: a computer system for 2D gel electrophoresis analysis. II. Pairing spots. *Comput. Biomed. Res.* **14**: 355–380.
- Lemkin, P.F. and Lipkin, L.E.** (1981c) GELLAB: a computer system for two-dimensional gel electrophoresis analysis. III. Multiple two-dimensional gel analysis. *Comput. Biomed. Res.* **14**: 407–446.
- Lemkin, P., Merrill, C., Lipkin, L., Van Keuren, M., Oertel, W., Shapiro, B., Wade, M., Schultz, M. and Smith, E.** (1979) Software aids for the analysis of 2D gel electrophoresis images. *Comput. Biomed. Res.* **12**: 517–544.
- Lemkin, P.F., Lipkin, L.E. and Lester E.P.** (1982) Some extensions to the GELLAB two-dimensional electrophoretic gel analysis system. *Clin. Chem.* **28**: 840–849.
- Lipkin, L.E.** (1980) Data-base techniques for multiple two-dimensional polyacrylamide gel electrophoresis analyses. *Clin. Chem.* **26**: 1403–1412.
- Meyer, F. and Beucher, S.** (1990) Morphological segmentation. *J. Visual Commun. Image Repres.* **1**: 21–46.
- Miller, M.J., Vo, K.-P., Nielsen, C., Geiduschek, E.P., Xuong, N.H.** (1982) Computer analysis of two-dimensional gels: semi-automatic matching. *Clin. Chem.* **28**: 867–875.
- Monardo, P.J., Boutell, T., Garrels, J.I. and Latter, G.I.** (1994) A distributed system for two-dimensional gel analysis. *Comput. Appl. Biosci.* **10**: 137–143.
- Müller, E.-C., Thiede, B., Zimny-Arndt, U., Scheler, C., Prehm, J., Müller-Werdan, U., Wittmann-Liebold, B., Otto, A. and Jungblut, P.** (1996) High-performance human myocardial two-dimensional electrophoresis database: edition 1996. *Electrophoresis* **17**: 1700–1712.

- Olson, A.D. and Miller, M.J.** (1988) Elsie4: quantitative computer analysis of sets of two-dimensional gel electrophoretograms. *Anal. Biochem.* **169**: 49–70.
- Patterson, S.D. and Latter, G.I.** (1993) Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. *BioTechniques* **15**: 1076–1083.
- Pleissner, K.-P., Sander, S., Oswald, H., Regitz-Zagrosek, V. and Fleck, E.** (1996) The construction of the World Wide Web-accessible myocardial two-dimensional gel electrophoresis protein database 'HEART-2DPAGE': a practical approach. *Electrophoresis* **17**: 1386–1392.
- Pleissner, K.-P., Sander, S., Oswald, H., Regitz-Zagrosek, V. and Fleck, E.** (1997a) Towards design and comparison of World Wide Web-accessible myocardial two-dimensional gel electrophoresis protein databases. *Electrophoresis* **18**: 480–483.
- Pleissner, K.-P., Söding, P., Sander, S., Oswald, H., Neuss, M., Regitz-Zagrosek, V. and Fleck, E.** (1997b) Dilated cardiomyopathy-associated proteins and their presentation in a WWW-accessible two-dimensional gel protein database. *Electrophoresis* **18**: 802–808.
- Pleissner, K.-P., Regitz-Zagrosek, V., Krüdwagen, B., Trenkner, J., Hocher, B. and Fleck, E.** (1998) Effects of renovascular hypertension on myocardial protein patterns – analysis by computer-assisted two-dimensional gel electrophoresis. *Electrophoresis* **19**: 2043–2050.
- Pleissner, K.-P., Hoffmann, F., Kriegel, K., Wenk, C., Wegner, S., Sahlström, A., Oswald, H., Alt, H. and Fleck, E.** (1999) New algorithmic approaches to protein spot detection and pattern matching in 2-DE gel databases. *Electrophoresis* **20**: 255–265.
- Prehm, J., Jungblut, P. and Klose, J.** (1987) Analysis of two-dimensional electrophoretic protein patterns using a video camera and a computer II. Adaptation of automatic spot detection to visual evaluation. *Electrophoresis* **8**: 562–572.
- Pun, T., Hochstrasser, D., Appel, R., Funk, M., Villars-Augsburger, V. and Pellegrini C.** (1988) Computerized classification of two-dimensional gel electrophoretograms by correspondence analysis and ascendant hierarchical clustering. *Appl. Theor. Electrophoresis* **1**: 3–9.
- Schmid, H.-R., Schmitter, D., Blum, P., Miller, M. and Vonderschmitt, D.** (1995) Lung tumor cells: a multivariate approach to cell classification using two-dimensional protein pattern. *Electrophoresis* **16**: 1961–1968.
- Serpico, S.B. and Vernazza, G.** (1987) Problems and prospects in image processing of two-dimensional gel electrophoresis. *Opt. Engng* **26**: 661–668.
- Skolnick, M.M.** (1982) An approach to completely automatic comparison of two-dimensional electrophoresis gels. *Clin. Chem.* **28**: 979–986.
- Skolnick, M.M. and Neel, J.V.** (1986) An algorithm for comparing two-dimensional electrophoretic gels with particular reference to the study of mutation. *Adv. Hum. Genet.* **15**: 55–160.
- Sternberg, S.R.** (1986) Gray scale morphology. *Comput. Vision Graph. Image Processing* **35**: 333–355.
- Tarroux, P., Vincens, P. and Rabilloud, T.** (1987) HERMeS: a second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part V: data analysis. *Electrophoresis* **8**: 187–199.
- Vincens, P. and Tarroux, P.** (1987a) HERMeS: a second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part III: spot list matching. *Electrophoresis* **8**: 100–107.
- Vincens, P. and Tarroux, P.** (1987b) HERMeS: a second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part IV: data base organization and management. *Electrophoresis* **8**: 173–186.
- Vincent, L.** (1993) Morphological grayscale reconstruction in image analysis: applications and effective algorithms. *IEEE Trans. Image Processing* **22**: 176–201.
- Vo, K.-P., Miller, M.J., Geiduschek, P., Nielsen, C., Olson, A. and Xuong, N.** (1981) Computer analysis of two-dimensional gel. *Anal. Biochem.* **112**: 258–271.
- Wegner, S., Sahlström, A., Pleissner, K.-P., Oswald, H. and Fleck, E.** (1998) Eine hierarchische Wasserscheidentransformation für die Spotdetektion in 2D Gel-elektrophorese-Bildern. In *Informatik Fachberichte Heft 10, Bildverarbeitung in der Medizin*. Springer, Berlin, pp. 134–138
- Wu, Y., Lemkin, P.F. and Upton, K.** (1994) A fast spot detection algorithm for 2D electrophoresis analysis. *Electrophoresis* **14**: 1341–1356.

第七章 增强高通量蛋白质组分析： 稳定同位素标记的影响

Manfredo Quadroni and Peter James

1. 前言

蛋白质组分析所关注的问题是蛋白质表达的整体变化,实现这种分析最常用的技术是二维电泳和质谱分析技术。最近,商业利益促使了几种集成化系统的发展,即意图建立从 2-DE 样品分离到质谱蛋白质鉴定的工作流水线。在界面友好、适用性强的基础上使步骤简化和标准化可能会为更多的实验室成功运用这项技术开辟一条道路。同时,新型标记方法的引入能大大地简化定量和数据的解释,也可能使凝胶电泳的步骤被省略掉而代之以简单的色谱步骤。本章中,我们将概述一种独特的蛋白质组分析方法的各个阶段,并概括 1998 年末我们对这一领域进行综述以来的新进展(Quadroni and James, 1999)。

全基因组序列(Devine and Wolfe, 1995)和大量表达序列标签库(EST, Adams *et al.*, 1991)的应用使定义生物体内全部可能的蛋白质成分(蛋白质组)成为可能。现在的兴趣集中在努力去解释大量涌现的新序列数据,以了解细胞中大批化学分子在建立细胞的分子机器过程中是怎样相互作用的。因而解答生物学问题的中心也必须从一个简化论者向全局性方法转移。可以开展基于监测整个基因组表达的更为细微的分析方法,而不是分割开整个过程去鉴定一个可能的效应子。这样一种通过 mRNA 表达的定性和定量分析来描述生物学系统成为可能(Lashkari *et al.*, 1997; Velculescu *et al.*, 1997),原因是由于各种 DNA 表达分析方法和全基因组范围(genome-wide)研究的发展。大规模的 mRNA 表达研究和基因组 DNA 测序的前提是高度自动化操作。可重复性是对这些复杂的数据集(datasets)的统计分析的基础,而自动化和高通量的数据积累对这样大的工作来说必不可少。DNA 和 mRNA 物化性质很一致,易于操作,可通过聚合酶链反应方法扩增,因此适于自动化。与核酸相反,蛋白质多种多样,不可能找到通用的蛋白质处理方法。当前惟一系统分析细胞中多数蛋白质表达状态的方法是基于 2-DE 的方法(Klose, 1975; O'Farrell, 1975; O'Farrell *et al.*, 1977)。最近,同位素标记技术(Gygi *et al.*, 1999; Oda *et al.*, 1999)被开发出来,从而使 2-DE 分离蛋白质的定量更加可靠。这些方法为开发可替代 2-DE 的技术开辟了道路,并大大拓宽了蛋白质表达检测的动态范围。本章我们将概述在复杂蛋白质组分析中所尝试使用的自动化蛋白质鉴定和定量的方法及其缺陷。

2. 样品制备

可重复的样品制备可能是获得有效的蛋白质组分析中最重要的一点。为了得到有代表性的蛋白质样品,细胞或组织必须进行有效破碎,细胞内成分的溶解要具有重复性。在

使用基于尿素的混合溶液(其中含有非离子去污剂、还原剂和蛋白酶抑制剂)进行蛋白质提取前,常使用物理破碎法碎裂细胞或组织,比如超声法、机械驱动急剧压力变化法、匀浆和剪切技术。细胞膜和细胞骨架蛋白质的回收率不稳定,在一些蛋白质可能完全被提取,而仍有高达 10% 的蛋白质在提取后残留在沉淀颗粒中。可使样品在一天内运行完成(从上样到胶的染色和扫描)的小 2-DE 凝胶的开发是很重要的一步。这就可以同时快速评价多种不同的样品制备方法。为了降低被分析的真核生物组织蛋白质混合物的高度复杂性,使得其中含量较少的组分可以被分辨和分析,常要求对样品进行预分离。处理类似组织这样的混合细胞组分时,使用荧光活化细胞分类器(FACS)(Madsen *et al.*, 1988)对细胞进行预分离能获得小的亚细胞群,从而大大提高分析的灵敏度。更重要的进展是自由流动电泳分离细胞器/组件方法的开发(Anderson, 1967; Burggraf *et al.*, 1995)。这是蛋白质表达谱研究的重要进展,它将显示蛋白质是在细胞的哪个部分、在何种条件下发现的(Blackstock and Weir, 1999)。这就使我们可以去开发一个理论模型,通过该模型,细胞状态可以根据一套相互作用的分子机器来定义。

3. 二维凝胶分离和分析

3.1 自动化二维聚丙烯酰胺凝胶电泳(2-DE-PAGE)

自20多年前建立以来,标准的变性2-DE 凝胶系统仍然保持其必要的手工操作,而且具有相同动态范围与分辨能力的蛋白质分离技术也没有有效改变。一个使用半自动化仪器样机的详细研究显示了机械操作的可重复性优点(Harrington *et al.*, 1993),但是,直到最近,2-DE 胶的大规模的自动化产品研制开发才由商业企业开始进行,如 Large Scale Biology 和 Oxford Glycosciences。目前已有三家公司宣布了其在蛋白质组分析集成化系统的发展:Amersham-Pharmacia, BioRad 和 Genomic Solutions。我们将在本节最后概述这些,但必须强调,这些是非常新的系统,有时仅部分完成,对这些系统的体验非常有限。

3.2 蛋白质的扫描检测和定量

通常蛋白质可以通过其 UV 吸收呈现,不过,聚丙烯酰胺在同一紫外范围有吸收,所以必须对蛋白质进行染色。为保持蛋白定量和进一步分析步骤的一致性,使用的染色方法应该对全部蛋白质都有效。到目前为止,最常用的蛋白染色方法是考马斯亮蓝染色和银染。两种方法的灵敏度都与蛋白质的序列有关(尤其是银染),对大多数蛋白质来说,二者都有较宽的动态线性范围。另一个选择是使用荧光标记,可以在一向到二向的平衡步骤利用与蛋白共价结合,也可以使用荧光标记 SDS 类似物或荧光染色(Steinberg *et al.*, 1996)。尽管荧光染料具有非常高的甚至在亚 attomol 水平的灵敏度,但响应的线性和通用性仍然是限制其应用的因素。任何染色方法都不会有干扰进一步分析的副作用,即使是银染也只有很少的化学修饰(见第四章)(Shevchenko *et al.*, 1996)。通常,通过激光密度仪在吸光模式下扫描,所得到的线性响应范围是 0~3.5 个光密度单位(OD)。基于放射性同位素标记的检测方法非常灵敏(优于 attomole),但并不适合于自动化操作,因为很难将胶片(或者来自 BaFBr:Eu²⁺ 储屏图像的计算机打印输出)上点的位置与凝胶基体相互关联。

除受所使用的显现方法影响外,2-DE 凝胶分离还受到几个限制因素的干扰。超高灵敏度的质谱蛋白质鉴定技术的出现使宽 pH 范围的 2-DE 凝胶分离的缺点显露无余,这种单个凝胶快速提供蛋白质组状态的全貌,但很多蛋白质得不到充分的分离。部分重叠的蛋白点可以通过计算机处理的方法分辨,但现在逐渐明确了很多均匀的蛋白点也包含有两个以上的蛋白质。对于原核生物,大约 20% 的蛋白点包含一个以上的蛋白质(Link *et al.*, 1997),而对于真核细胞来说,这一数字接近 40%。Proteome Inc. 与 Perseptive Biosystems 的合作研究(Parker *et al.*, 1998)详细说明了酵母提取物 2-DE 凝胶分离的蛋白点的交叉污染程度和实质。很多蛋白点包含了 3 个甚至更多蛋白质。尽管质谱技术足以对蛋白质混合物进行分辨(Arnott *et al.*, 1998),但每个蛋白点含有多个蛋白质使得很多实验结果难以解释。

3.3 通过稳定同位素标记定量 2-DE 凝胶中的蛋白质

体内标记 蛋白质分析中常用的技术是使用放射标记的氨基酸(如³⁵S-Met 和较少用的³⁵S-Cys)增加 2-DE 的检测灵敏度,并利用脉冲示踪实验测定蛋白质合成和降解速率。最近稳定同位素在蛋白质质谱分析中的应用引起人们的巨大兴趣。一种方法是使用同位素缺乏介质(重同位素丰度较低的介质如¹⁵N、²H和³H以及¹³C)。这使得一个离子的同位素簇峰衰减成一个单峰,以增加灵敏度和准确度(Marshall *et al.*, 1997)。整体细胞的同

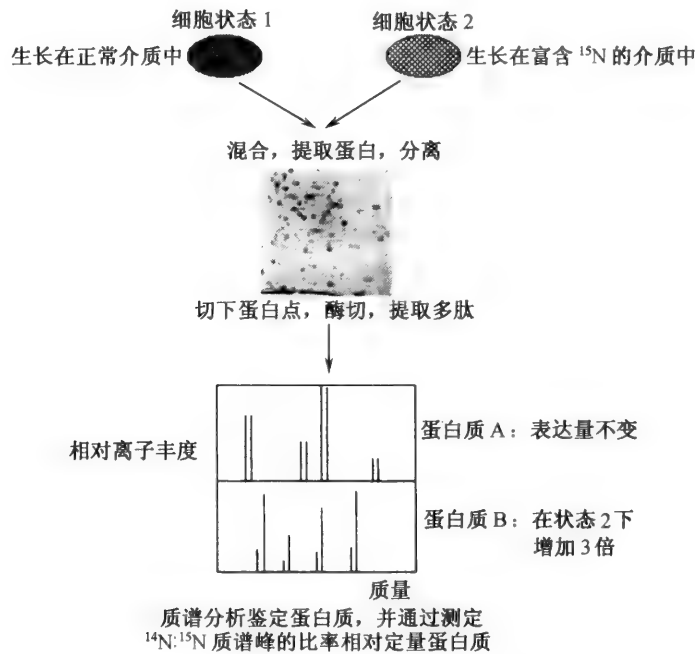


图 7.1 2-DE 凝胶分析的体内稳定同位素标记

细胞分别在生长正常的修饰条件和富含¹⁵N的介质中。蛋白质从细胞中提取,并在 2-DE 分离前混合。蛋白点被切下、酶解,再进行 MALDI-MS 分析。离子对的¹⁴N与¹⁵N同位素之比可以定量两种状态之间的表达水平变化。

位素标记已经用于蛋白质表达和修饰相对变化的定量(图 7.1)(Oda *et al.*, 1999)。细胞生长在由正常同位素组成的“轻的”化合物介质中,或是富含¹⁵N的“重的”介质中;合并从生长在不同条件下的细胞中得到的蛋白质提取物,一部分通过 2-DE 分离。切下重要的蛋白点,酶解后使用质谱分析提取的肽。¹⁵N结合的肽比正常肽大一个质量单位,使得来自两个细胞混合物中的肽得到区分。这样肽质谱峰就会成对出现,可以根据峰的相对高度加以定量。研究发现二者比率的线性值跨越两个数量级。这个步骤也可很好地用于确定翻译后修饰水平的变化(图7.2)。实际上,这种方法受到材料成本的限制,不可能对整个动物进行标记,而且它需要来自小牛胚胎血清的血清衍生因子,通常不用于真核细胞培养。

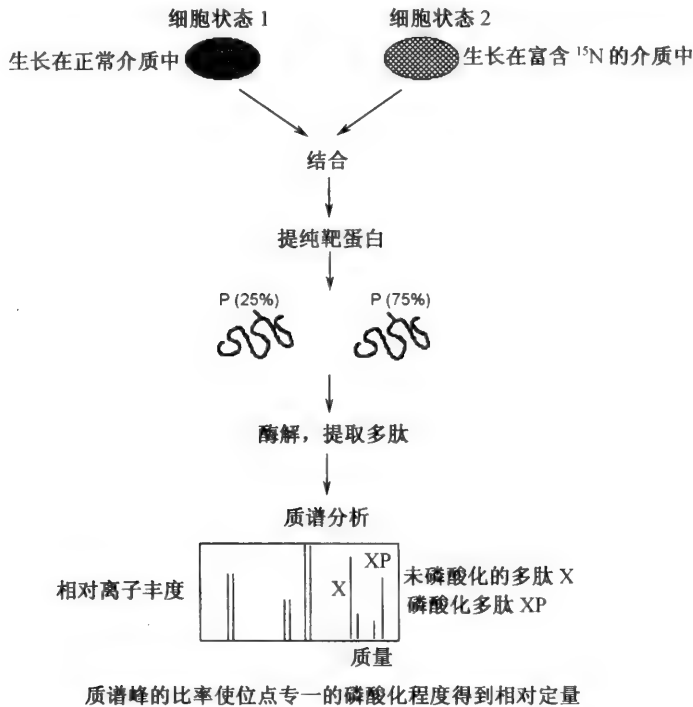


图 7.2 同位素标记和翻译后修饰的鉴定
能用于测定翻译后修饰(如磷酸化)的变化水平的同位素标记和表达定量的基本方法。

肽后标记 另外可以选择使用分离后同位素标记的方法进行相对定量(图 7.3)。最近 Aebersold 小组描述了一种称为同位素编码亲和标签(ICAT)的方法,非常有望成为替代 2-DE 凝胶分析的比较蛋白质组全面方法(Gygi *et al.*, 1999)。它是标记经 2-DE 分离所得到的不同条件下的细胞提取物的蛋白质,选出单个蛋白点,切下后 Lys 残基被琥珀酰化,随后蛋白质用 Asp/GluC 蛋白酶酶解。从状态 1 得到的蛋白质 N 末端用正常的试剂 H4NicNHS[1-(H4-烟酰氧基)琥珀酰亚胺]专一标记,来自状态 2 的蛋白质用氘代试剂 D4NicNHS 予以标记。基质辅助激光解吸飞行时间质谱(MALDI-MS)分析结合酶解可以通过单个肽的 D4/H4 比率进行蛋白质的相对定量(图 7.4)。这种方法的优点是它能用于全部类型的蛋白质提取物,与细胞培养同位素缺乏或富集介质方法相比,费用低廉。这

种技术的一项重要应用是对 1-D 胶条和 2-D 未能完全分离的含多个蛋白质的点进行蛋白质的相对定量。这对只能在 1-DE 上简单地部分分离而不能在 2-DE 上运行的膜蛋白质研究非常有用。标记也有助于简化标记多肽的串联质谱图谱的解释。

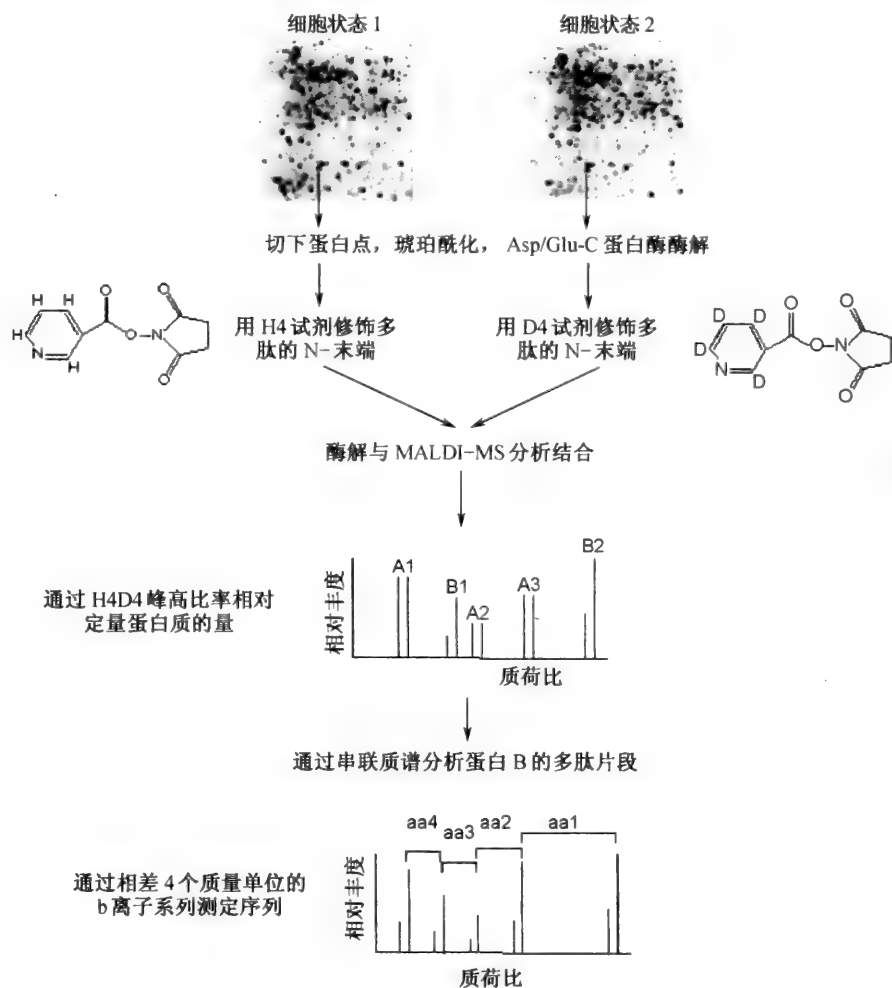


图 7.3 2-DE 凝胶分析的提取后蛋白质标记

为了解决体内蛋白质标记的问题,蛋白质可以在 2-DE 分离后进行同位素标记。多肽的 N 末端被标记后,这种方法可以简化串联质谱图谱的解释。

3.4 2-DE 凝胶蛋白点的分离

如果分析一系列的凝胶,用于进一步分析的蛋白点可以用自动切胶系统(robotic spot picking system)进行单个蛋白点的后处理(Walsh *et al.*, 1998)。澳大利亚蛋白质组分析机构(The Australian Proteome Analysis Facility, APAF)已经与 ARRM(Advanced Rapid Robotics Manufacturing, Kent Town, South Australia)合作开发了这种商业化的系统。机械手切下凝胶或 PVDF 膜上的蛋白点,置于用作自动化蛋白酶解的 96 孔板中酶解,后者由第二台自动化系统(由 Canberra Packard, Downer's Grove, II, USA 销售)完成,它随后

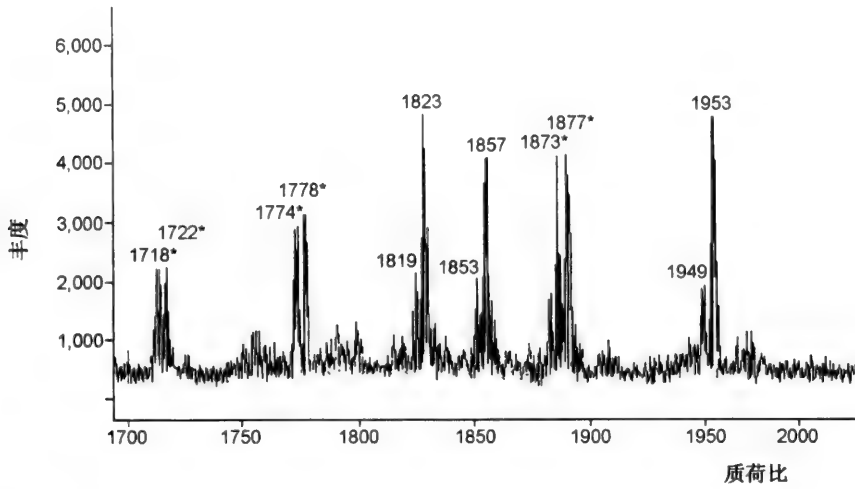


图 7.4 在单个点中发现的两种蛋白质的 MALDI-MS 图谱

酶解和标记后,不同条件下的蛋白质相对表达水平可被测定。在本图谱中,星号标记的峰表示表达水平保持一致的蛋白质。但同一蛋白点中,另一蛋白质(未标星号的多肽)的表达水平随三因素之一增加。

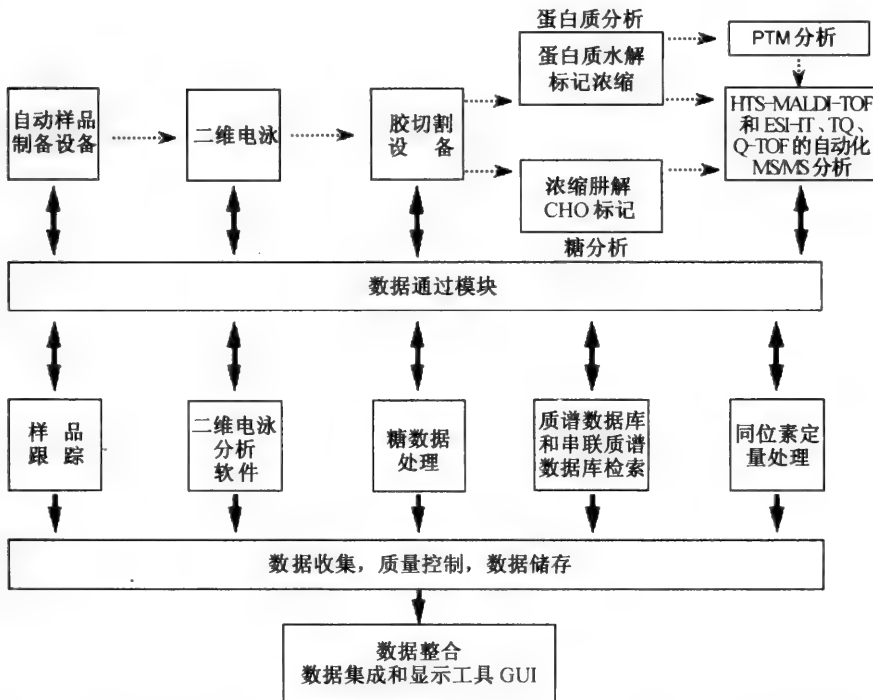


图 7.5 标准的蛋白质组分析系统

两级蛋白质组分析工作站软件控制类似 2-DE 运行、蛋白点切割、酶解和包括数据分析和样品跟踪的质谱分析机械装置。

将酶解液与基质混合后在多样品靶上点靶。在 10 个工作日内成功地完成了来自单个 2-DE 膜的 288 个蛋白点的分析 (Walsh *et al.*, 1998)。理想情况下, 2-DE 分析软件应该选择所分析的蛋白点, 并与机器人后来切下的点相匹配。这对 PVDF 膜来说比较简单, 但凝胶容易变形, 尤其是在点切割时容易发干, 难于操作。因此用于切割的凝胶或膜必须具有机械稳定性。此外也有其他一些自动酶解装置 (Davis *et al.*, 1995a; Houthaeve *et al.*, 1997)。最近三种商用“蛋白质组系统”已分别由 Amersham-Pharmacia, BioRad 和 Genomic Solutions 发布。这三种系统的基本组成部分非常相似, 其中包括一个胶切割机, 它根据用户定义的标准(如可见点和不可见点)得到的计算机辅助 2-DE 胶分析结果切下染色胶中有意义的点。这种“完全的”系统的基本结构如图 7.5 所示。

3.5 蛋白质酶解/碎裂

当前生物蛋白质和多肽分析的主要制约因素是样品的处理。蛋白质浓度低至 500 fmol/ μl 时, 酶解效率很低。这是因为蛋白酶在 5~50 pmol/ μl 范围内仅表现 50% 的最高活性。50 pmol/20 μl BSA 酶解得到约 60 个主要的多肽, 摩尔回收百分比约为 95%。而 100 fmol/20 μl BSA 仅得到 6 个多肽, 摩尔回收百分比约为 5%。可以选择开展化学方法(但仍存在样品处理损失)或直接质谱完成。完整蛋白的串联质谱分析可通过多种质谱技术进行, 如碰撞诱导解离 (collision-induced dissociation, CID) (Senko *et al.*, 1994), 红外多光子解离 (infrared multiphoton dissociation, IRMPD) (Little *et al.*, 1994), 紫外激光诱导碎裂 (ultraviolet laser-induced fragmentation) 和表面诱导碎裂等 (surface-induced fragmentation) (Williams *et al.*, 1990)。这些技术都能用傅立叶变换离子回旋加速共振 (FT-ICR) 完成, 其高质量准确度、分辨率和能进行 n 级串联质谱实验的能力使得人们可以理解质谱复杂的碎裂方式。已经证实, 与精确质量和 n 级质谱外推的系列序列标签结合, 完整蛋白质能通过 FT-ICR 的碎裂加以鉴定 (Mortz *et al.*, 1996)。最近 Li 和 Marshall 用 LC/ESI FT-ICR 质谱示范了蛋白质的在线鉴定。常规扫描首先用于提取完整蛋白质的精确质量, 接下来选择扫描 IRMPD 用于选择蛋白质的 m/z 离子。完整的质量是用于宽分子质量范围的数据库子集的选择。通常 b 和 y -碎片离子和从 IRMPD 获得的小序列标签与这一套方法相匹配识别蛋白质。FT-ICR-MS 仪器极其昂贵, 而且还未完成自动化运行。

4. 质谱: 使用质谱数据鉴定蛋白质

为完成高通量蛋白质鉴定, 必须采取分阶段的步骤。第一阶段是非常快速, 鉴别大批量蛋白质的低信息含量的数据采集, 接下来的第二阶段是详细说明其余蛋白质的低通量、高信息含量的数据产生阶段。最后的第三阶段是完成说明翻译后修饰的大工作量的数据积累。在实际的操作中, 基于 MALDI 的多肽质量指纹谱是理想的第一步, 紧接着的串联质谱肽片段指纹谱部分测序是第二步。

4.1 质量指纹肽谱

1993 年, 5 个研究小组分别提出质量指纹肽谱的概念 (Henzel *et al.*, 1993; James *et al.*, 1993; Mann *et al.*, 1993; Pappin *et al.*, 1993; Yates *et al.*, 1993)。这个概念是指特定的蛋白酶对蛋白质进行酶解后的质谱分析得到一套多肽质量, 这种特性像指纹一样,

每种蛋白都具有特定的质量肽谱。因此,一套质量数据能用来查寻序列已被计算的指纹质量替换的蛋白质数据库,从而发现相似的模式。得到酶解肽质量最快的方法是通常与飞行时间(TOF)结合的MALDI-MS。商业化的MALDI-TOF仪器装有最多1024个的样品靶位,软件可在无人时进行数据采集。样品制备对这项技术的成功与否相当重要。如果样品同源,使用最佳的肽、底物比例,几次激光发射完全可以不搜寻样品位置而获得较好的谱图。最近一项使用基质晶体沉淀的母型结晶层的样品制备技术推进了MALDI-MS的使用(Onnerfjord *et al.*, 1999)。样品均质性使得每个样品在15s内使用16次激光发射就可获得质谱。使用流动高压电动微量分布器在高密度靶盘点靶,用晶种层方法点样约100 μl 样品,得到直径为500 μm 的样品点。事实上,所有1000以上的峰都是单个带电肽,容易得到数据,也能获得较高的信噪比。现在使用在800~3000 m/z 质量范围的“自然的”内标(如胰蛋白酶的自切峰)能将误差保持在5ppm以下。如果数据库中提供指纹肽谱,那就能很快鉴定蛋白质,但在检索基因组或EST数据库时,这种技术就不太成功。使用两种以上数据库交叉检索可以减少随机的噪音因素(James *et al.*, 1994)。也可以结合来自不同特性的蛋白酶的酶解数据(如AspN和LysC)。单个酶解可在自然状态下测定,再在样品靶上完成化学修饰(如甲基化、乙酰化或氘代)。这大大地增加可信度,但需要双倍的分析时间。

4.2 小序列标签多肽指纹谱

使用双重酶解,或自然状态和化学修饰酶解可以开发已存在于简单肽指纹谱中的序列信息。胰蛋白酶在Lys和Arg残基位酶解,二者都是蛋白质中存在最丰富的氨基酸。因此蛋白质中存在大量的重复序列,如KK、RR和KKR。胰蛋白酶随机地切开这些位点,还有固有的较低的蛋白外切酶活性,也就产生不规则的C末端。平均每个肽质量谱可以找到2.5个这样的序列标签(Jensen *et al.*, 1996)。连续的肽链内切酶和肽链外切酶酶解能有效得到不规则的N或C末端序列,产生的数据可用于检索序列数据库(Korostensky *et al.*, 1998),蛋白质鉴别的可信度远比简单的指纹谱高。不过,这种方法有酶解同样的缺点,随着肽浓度的降低和肽数量的增加,肽链端解酶的活性也很快下降。另外,由于肽链端解酶酶解的不可预测性,序列标签可能对应于母离子或开始的5个氨基酸之后的序列,使得在复杂的酶解中难于将序列归属给特定的母离子,很难得到序列标签。

使用连续的化学降解步骤可以替代产生不规则N或C末端序列的肽链内切酶和肽链外切酶酶解。已经证明多肽混合物的一步手工Edman降解得到单个N末端氨基酸标签的方法的有效性(Jensen *et al.*, 1996),但衍生化导致灵敏度降低(降低因子约为10),洗脱步骤也会导致大量的样品流失。实际上,“阶梯测序”是这种方法的简略形式,使用异硫氰酸苯酯的自动Edman降解是在低浓度的序列终止剂(硫氰酸苯酯)的存在下完成的,得到的阶梯肽片段用MALDI-MS进行分析(Chait *et al.*, 1993)。这种方法的改进形式不用序列终止剂,通过在每个循环加入等量的起始肽完成偶联和裂解反应产生多肽阶梯(Bartlett-Jones *et al.*, 1994)。多肽末端氨基的保留使得随后的碱性四级铵NHS酯化可以提高灵敏度,但它的主要缺点是额外增加了处理步骤。

已经开发了替代的使用水溶性试剂的化学降解方法,它是将多肽吸附在固定于聚四氟乙烯膜的C-18反相颗粒上,进行基于硫酯的降解。这种方法能在固定在MALDI靶的

单个聚四氟乙烯膜上同时完成 100 个蛋白质的降解,随后可直接进入质谱仪进行分析。实际上蛋白质首先经过化学降解或蛋白酶解,再固定在植入 C-18 颗粒的聚四氟乙烯膜上。降解液分成两部分,一部分进行完整肽混合物的 MALDI-TOF-MS 分析(图 7.6a)。另一部分进行是化学降解后的测定(图 7.6b)。降解是基于使用类似硫代乙酰巯基乙酸的硫代酰基硫酯(图 7.6c)。三甲胺蒸汽首先通过膜以建立碱性条件,然后 *N*-乙基吗啡啉缓冲液中的硫代乙酰巯基乙酸再透过膜,与多肽反应形成 *N*-硫代乙酰衍生物。在通过三氟乙酸蒸汽切下第一个氨基酸前, Teflon 膜先用水洗,再用 0.1% 三氟乙酸/水洗涤。通过小心选择水蒸汽的量,使仅 25% 的氨基酸裂解,另外 75% 进行硫代乙酰基水解。重复数次,序列标签和母离子组成的肽阶梯产品可通过 MALDI 分析读取,并用于数据库检索。这种方法的优点是数据收集速度快,肽降解物在降解后立即被固定,样品损失降低,反应也可平行完成。另外非序列专一酶解和化学降解可以产生用于数据库检索的数据。这样 MALDI-MS 与自动化酶解和其他多肽修饰化学结合给蛋白质鉴定提供了高通量的手段。现在大多数商业化质谱仪都包括多个步骤,如样品靶的装载、样品定位、图谱采集和质谱峰数据的数据库检索等。

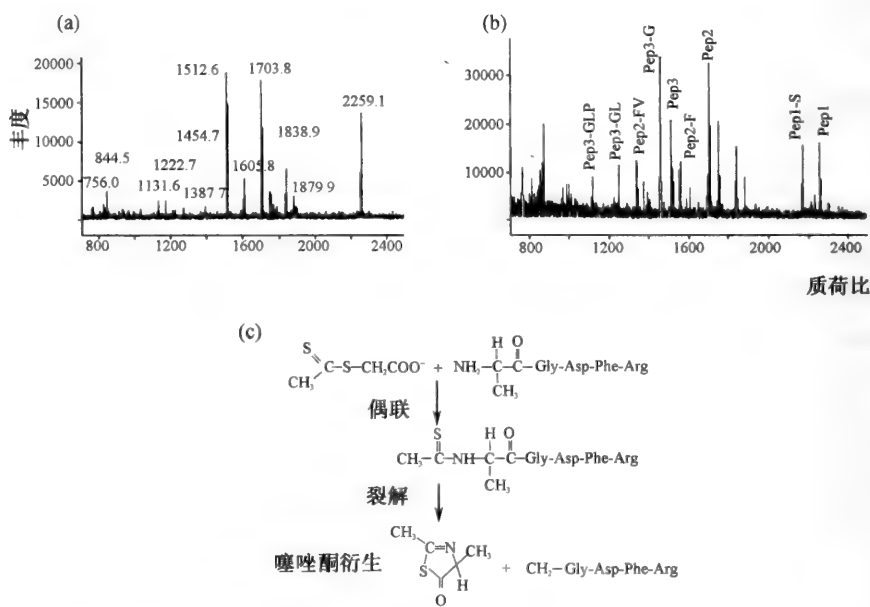


图 7.6 Chemtag 化学和 MALDI-MS

(a) 胰蛋白酶酶解从 2-DE 凝胶分离的蛋白点的 MALDI-MS 谱图。尽管只有几种多肽存在,而且蛋白点中只有两个蛋白质,质量指纹肽谱仍不能鉴定存在的蛋白质。(b)使用硫代乙酰巯基乙酸(thioacetylthioglycolic acid)降解 3 个循环后酶解。序列标签及其母离子如图所示,它们可用来鉴定蛋白点中的两个蛋白质。

5. 质谱:使用串联质谱数据鉴别蛋白质

5.1 串联质谱数据采集

使用串联质谱碎裂方式进行单个多肽鉴定的概念 1994 年被提出(Eng *et al.*, 1994,

Mann and Wilm, 1994), 它类似于质量指纹肽谱的蛋白质鉴定方法。20 世纪 80 年代 Don Hunt 小组提出的低能量、三级四级杆质谱仪和 Klaus Biemann (1990) 提出的高能量、四个扇形的磁质谱开创了多肽串联质谱的历史。第一个质量扫描阶段用于离子通过含有碰撞气体(如氩气)的高压区加速前分离单个多肽离子, 第二个质量扫描阶段用来测定碰撞诱导/活化解离(CID)产生的多肽碎片质量。利用装有离子闸门和反射镜的 TOF 仪器, 增加激光能量(获得正常质谱所需的两种因素之一)可获得源后衰变(PSD)谱。在股流膨胀(plume expansion)和离子加速阶段多肽离子与基质多次碰撞的结果是, 大部分 MALDI 解吸离子到达检测器前要经受“延迟”碎裂。其中任一种技术产生的一组离子都能用于测定多肽的序列。多肽的碎片质量组扮演着多肽指纹的角色, 能用来在序列数据库中搜寻类似的多肽。PSD 分析应易于自动化操作, 并使场反射曲线化, 使得更简单地在一次扫描中可以收集全部碎裂谱。不过它的缺点是很难从多肽混合物得到 PSD 谱, 而且一些分级分离形式应该在分析之前完成(Gevaert *et al.*, 1996)。

虽然 CID-MS/MS 碎裂可获得信息量较大的数据, 但其数据收集比简单的质量指纹肽谱要费时。事实上, 两种获得高灵敏度串联质谱数据的方法可通过使用电喷雾接口到三级四级杆、离子阱或四级杆-TOF 质谱实现: 酶解液作为混合物从毛细管 1~10 μm 的出口末端以非常低的流速直接喷雾(< 50 nl/min, Andren *et al.*, 1994; Gale and Smith, 1993; Valaskovic *et al.*, 1995; Wahl *et al.*, 1993; Wilm and Mann, 1994), 或纳升电喷雾在线分离方法如电泳毛细管区带电泳(CZE)或 HPLC。可通过程序将离子挑选出来自动对多肽混合物进行串联质谱分析, 其优点是使用查询表可以滤出自切碎片和来自凝胶污染物的杂质(Davis *et al.*, 1995b)。这种方法已成功地用于完整蛋白质的测序(Piccinni *et al.*, 1994)。动态的 CZE/HPLC 方法比静态的纳升电喷雾方法灵敏度要高得多, 前者的多肽会浓缩 50 倍甚至更高(Davis and Lee, 1997)。两种方法的共同优点是, 他们都能与母离子扫描相结合。当这些离子在质谱中的信噪比为 1 时, 多肽的共同碎裂片段如 immonium 或磷酸盐离子的母离子扫描仍可确定多肽离子的质量。不过动态方法的最大的优点是能与自动进样器接口, 完成自动进样。

5.2 串联质谱数据库检索

使用未经解析的串联质谱谱图的自动数据库序列检索的运算法则已在过去的 4 年里得到发展。也许应用范围最广的是由西雅图华盛顿大学的 John Yates 和 Jimmy Eng 开发的 SEQUEST 数据库(Eng *et al.*, 1994)。其最初的目的是搜索未修饰的多肽串联质谱碎裂片段的蛋白质数据库, 但后来扩展到搜索翻译后修饰的多肽(Yates *et al.*, 1995b)、DNA 数据库(Yates *et al.*, 1995a)和高能量 CID 或 PSD 数据(Griffin *et al.*, 1995, Yates *et al.*, 1996)。事实上程序搜寻数据库(可以但不一定明确蛋白酶)中所有与母离子(在定义的质量范围内)有相同质量的多肽, 然后将预测的串联质谱谱图和实验测得的谱图相匹配, 最终对得分最高的多肽进行交叉相关分析以获得最好的比较。现在万维网还有其他一些用未经解析的谱图数据检索的数据库检索程序, 如纽约洛克菲勒研究所的 Fragfit(Prowl 软件组的一部分; <http://prowl.rockefeller.edu/>), 加州大学旧金山分校的 MS-Tag(ProteinProspector 组的一部分; <http://prospector.ucsfedu.edu/>)。

另一个序列数据库中蛋白质鉴定的算法 Peptide-Search 是由 Matthias Mann 开发的

端,因为难以区分肽的 α 氨基(N末端)和 ϵ 氨基(Lys),而且去除基本位点后检测灵敏度下降,所以并没有完全成功。我们最近介绍了一种很容易从串联质谱中推导序列的方法。标记的方法已在前面的“胶后标记”部分讨论。简单说来,使用50:50的H4或D4的烟酸混合物标记蛋白质,其中的赖氨酸侧链 ϵ 氨基在衍生前被琥珀酰化,所以N末端的 α -氨基反应绝对专一。MALDI图谱表明酶解得到的全部多肽都有质量相差4Da的成对质谱峰出现。必须选择足够宽的质量范围以包括全部的母离子同位素峰,产生的串联质谱图以成对质谱峰出现的来自N末端的离子为主(b 和一些 a 离子)(图7.7)。所有从C末端得到的离子(y 离子或中间的离子)都是单峰。所以相当容易就获取到蛋白质的序列。即使存在未知的翻译后修饰,也不影响用序列(包括在适当的位置的X)在(T)FASTA中进行序列同源性搜索的能力。自动化序列推断相当容易,使得蛋白质的鉴定非常快速,而且具有很高的可信度。

5.4 数据解释工具

一旦数据库检索开始执行,自动化处理就不停。高通量的蛋白质分析要求开发工具软件帮助实验者解释数据得到有意义的倾向和结论。就HPLC或CZE-自动串联质谱来说,其产生的数据文件量非常庞大,而人工查看只能获得很少的一部分。为了比较各种的文件,强大的数据分析软件包是必不可少的。Terry Lee的实验室为了解释从蛋白质的质谱分析中得到的数据,开发出Macpro-Mass(Lee and Vemuri, 1990)程序,它应用广泛、方便用户,是最早开发的程序之一。第二代程序Sherpa是在西雅图华盛顿大学由Alex Taylor(Taylor *et al.*, 1996)开发的。程序能同时搜索多个蛋白质序列,同时比较两个LC-MS文件,搜索糖基化和磷酸化多肽,给予数据和预测之间的匹配程度简单的评价。这种分析工具给分析带来巨大的帮助。最近John Yates已经描述的一种不同的SEQUENT程序,用于比较多肽的碰撞诱导解离谱库检索和LC-MS/MS实验得到的串联质谱差减分析(Yates *et al.*; 1998)。

因此大部分蛋白质组分析领域已经实现部分自动化。还需要给予格外重视的主要领域有:极小量的原料(>10fmol)酶解鉴别翻译后修饰的通用方法,尤其是糖基化的复杂领域;样品酶解经快速高效地浓缩后可进行肽指纹质谱分析,如果需要,还可进行相同样品的串联质谱分析(MALDI离子阱应该是非常理想的)。但不幸的是缺少一个能以简单可行的方式将各种形式的所有数据结合在一起,进行高级(high-order)数据分析的完整的实验室信息管理系统。稳定的同位素标记简化了很多难题,如单个蛋白点中含有多个蛋白质及其定量问题,它也是简化串联质谱数据文件里信息内容的简单方法。这一领域内需要做的还很多,但大部分问题无足轻重,纳米技术(nanotechnology)和更好的数据开发工具不久将革新这一领域。

(刘尚义译 钱小红校)

参考文献

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. *et al.* (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252**: 1651–1656.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Altschul, S.F., Boguski, M.S., Gish, W. and Wootton, J.C. (1994) Issues in searching molecular sequence databases. *Nature Genet.* **6**: 119–129.
- Anderson, N.G. (1967) Preparative zonal centrifugation. *Methods Biochem. Anal.* **15**: 271–310.
- Andren, P.E., Emmet, M.R. and Caprioli, R.M. (1994) Micro-electrospray: zeptomole/attomole per microliter sensitivity for peptides. *J. Am. Mass Spectrom.* **5**: 867–869.
- Arnott, D., Henzel, W.J. and Stults, J.T. (1998) Rapid identification of comigrating gel-isolated proteins by ion trap-mass spectrometry. *Electrophoresis* **19**: 968–980.
- Bartlett-Jones, M., Jeffery, W.A., Hansen, H.F. and Pappin, D.J. (1994) Peptide ladder sequencing by mass spectrometry using a novel, volatile degradation reagent. *Rapid Commun. Mass Spectrom.* **8**: 737–742.
- Biemann, K. (1990) Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol.* **193**: 455–479.
- Blackstock, W.P. and Weir, M.P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **17**: 121–127.
- Burggraf, D., Weber, G. and Lottspeich, F. (1995) Free flow-isoelectric focusing of human cellular lysates as sample preparation for protein analysis. *Electrophoresis* **16**: 1010–1015.
- Chait, B.T., Wang, R., Beavis, R.C. and Kent, S.B. (1993) Protein ladder sequencing. *Science* **262**: 89–92.
- Davis, M.T. and Lee, T.D. (1997) Variable flow liquid chromatography tandem mass spectrometry and the comprehensive analysis of complex protein digest mixtures. *J. Am. Soc. Mass Spectrom.* **8**: 1059–1069.
- Davis, M.T., Lee, T.D., Ronk, M. and Hefta, S.A. (1995a) Microscale immobilized protease reactor columns for peptide mapping by liquid chromatography/mass spectral analyses. *Anal. Biochem.* **224**: 235–244.
- Davis, M.T., Stahl, D.C., Hefta, S.A. and Lee, T.D. (1995b) A microscale electrospray interface for on-line, capillary liquid chromatography/tandem mass spectrometry of complex peptide mixtures. *Anal. Chem.* **67**: 4549–4556.
- Devine, K.M. and Wolfe, K. (1995) Bacterial genomes: a TIGR in the tank. *Trends Genet.* **11**: 429–431.
- Eng, J.K., McCormack, A.L. and Yates, J.R. (1994) Correlating tandem mass spectral data of peptides to sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**: 976–989.
- Gale, D.C. and Smith, R.D. (1993) Small volume and low flow rate electrospray ionisation mass spectrometry of aqueous solutions. *Rapid Commun. Mass Spectrom.* **7**: 1017–1021.
- Gevaert, K., Verschelde, J.L., Puype, M., Van Damme, J., Goethals, M., De Boeck, S. and Vandekerckhove, J. (1996) Structural analysis and identification of gel-purified proteins, available in the femtomole range, using a novel computer program for peptide sequence assignment, by matrix-assisted laser desorption ionization-reflectron time-of-flight-mass spectrometry. *Electrophoresis* **17**: 918–924.
- Griffin, P.R., MacCoss, M.J., Eng, J.K., Blevins, R.A., Aaronson, J.S. and Yates, J.R. III (1995) Direct database searching with MALDI-PSD spectra of peptides. *Rapid Commun. Mass Spectrom.* **9**: 1546–1551.
- Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**: 994–999.
- Harrington, M.G., Lee, K.H., Yun, M., Zewert, T., Bailey, J.E. and Hood, L. (1993) Mechanical precision in two-dimensional electrophoresis can improve protein spot positional reproducibility. *Appl. Theor. Electrophoresis* **3**: 347–353.
- Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C. and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl Acad. Sci. USA* **90**: 5011–5015.

- Hines, W.M., Falick, A.M., Burlingame, A.L. and Gibson, B.W. (1992) Pattern based algorithm for peptide sequencing from tandem high energy collision induced dissociation mass spectra. *J. Am. Soc. Mass Spectrom.* **3**: 326–336.
- Houthaeve, T., Gausepohl, H., Ashman, K., Nilsson, T. and Mann, M. (1997) Automated protein preparation techniques using a digest robot. *J. Protein Chem.* **16**: 343–348.
- Hunt, D.F., Yates, J.R. III, Shabanowitz, J., Winston, S. and Hauer, C.R. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl Acad. Sci. USA* **83**: 6233–6237.
- James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1993) Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* **195**: 58–64.
- James, P., Quadroni, M., Carafoli, E. and Gonnet, G. (1994) Protein identification in DNA databases by peptide mass fingerprinting. *Protein Sci.* **3**: 1347–1350.
- Jensen, O.N., Vorm, O. and Mann, M. (1996) Sequence patterns produced by incomplete enzymatic digestion or one-step Edman degradation of peptide mixtures as probes for protein database searches. *Electrophoresis* **17**: 938–944.
- Johnson, R.S. and Biemann, K. (1989) Computer program (SEQPEP) to aid in the interpretation of high-energy collision tandem mass spectra of peptides. *Biomed. Environ. Mass Spectrom.* **18**: 945–957.
- Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**: 231–243.
- Korostensky, C., Staudenmann, W., Dainese, P., Hoving, S., Gonnet, G. and James, P. (1998) An algorithm for the identification of proteins using peptides with ragged N- or C-termini generated by sequential endo- and exopeptidase digestions. *Electrophoresis* **19**: 1933–1940.
- Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl Acad. Sci. USA* **94**: 13 057–13 062.
- Lee, T.D. and Vemuri, S. (1990) MacProMass: a computer program to correlate mass spectral data to peptide and protein structures. *Biomed. Environ. Mass Spectrom.* **19**: 639–645.
- Link, A.J., Hays, L.G., Carmack, E.B. and Yates, J.R. III (1997) Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis* **18**: 1314–1334.
- Little, D.P., Speir, J.P., Senko, M.W., O'Connor, P.B. and McLafferty, F.W. (1994) Infrared multiphoton dissociation of large multiply charged ions for biomolecule sequencing. *Anal. Chem.* **66**: 2809–2815.
- Madsen, P.S., Hokland, M., Ellegaard, J., Hokland, P., Ratz, G.P., Celis, A. and Celis, J.E. (1988) Major proteins in normal human lymphocyte subpopulations separated by fluorescence-activated cell sorting and analyzed by two-dimensional gel electrophoresis. *Leukemia* **2**: 602–615.
- Mann, M., Hojrup, P. and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* **22**: 338–345.
- Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**: 4390–4399.
- Marshall, A.G., Senko, M.W., Li, W., Li, M., Dillon, S., Guan, S. and Logan, T.M. (1997) Protein molecular mass to 1 Da by ¹³C ¹⁵N double depletion and FT-ICR mass spectrometry. *J. Am. Chem. Soc.* **119**: 433–434.
- Mortz, E., O'Connor, P.B., Roepstorff, P., Kelleher, N.L., Wood, T.D., McLafferty, F.W. and Mann, M. (1996) Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence data bases. *Proc. Natl Acad. Sci. USA* **93**: 8264–8267.
- Oda, Y., Huang, K., Cross, F.R., Cowburn, D. and Chait, B.T. (1999) Accurate quantification of protein expression and site-specific phosphorylation. *Proc. Natl Acad. Sci. USA* **96**: 6591–6596.
- O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007–4021.
- O'Farrell, P.Z., Goodman, H.M. and O'Farrell, P.H. (1977) High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell* **12**: 1133–1141.
- Onnerfjord, P., Ekstrom, S., Bergquist, J., Nilsson, J., Laurell, T. and Marko-Varga, G. (1999) Homogeneous sample preparation for automated high throughput analysis with matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*

- 13: 315–322.
- Pappin, D.J.C., Hojrup, P. and Bleasby, A.J.** (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**: 327–332.
- Parker, K.C., Garrels, J.I., Hines, W., Butler, E.M., McKee, A.H., Patterson, D. and Martin, S.** (1998) Identification of yeast proteins from two-dimensional gels: working out spot cross-contamination. *Electrophoresis* **19**: 1920–1932.
- Patterson, S.D., Thomas, D. and Bradshaw, R.A.** (1996) Application of combined mass spectrometry and partial amino acid sequence to the identification of gel-separated proteins. *Electrophoresis* **17**: 877–891.
- Piccinni, E., Staudenmann, W., Albergoni, V., De Gabrieli, R. and James, P.** (1994) Purification and primary structure of metallothioneins induced by cadmium in the protists *Tetrahymena pigmentosa* and *Tetrahymena pyriformis*. *Eur. J. Biochem.* **226**: 853–859.
- Quadroni, M. and James, P.** (1999) Proteomics and automation. *Electrophoresis* **20**: 664–677.
- Senko, M.W., Speir, J.P. and McLafferty, F.W.** (1994) Collisional activation of large multiply charged ions using Fourier transform mass spectrometry. *Anal. Chem.* **66**: 2801–2808.
- Shevchenko, A., Wilm, M., Vorm, O. and Mann, M.** (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**: 850–858.
- Shevchenko, A., Chernushevich, I., Ens, W., Standing, K.G., Thomson, B., Wilm, M. and Mann, M.** (1997) Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **11**: 1015–1024.
- Steinberg, T.H., Haugland, R.P. and Singer, V.L.** (1996) Applications of SYPRO orange and SYPRO red protein gel stains. *Anal. Biochem.* **239**: 238–245.
- Takao, T., Hori, H., Okamoto, K., Harada, A., Kamachi, M. and Shimonishi, Y.** (1991) Facile assignment of sequence ions of a peptide labelled with ^{18}O at the carboxyl terminus. *Rapid Commun. Mass Spectrom.* **5**: 312–315.
- Taylor, J.A. and Johnson, R.S.** (1997) Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **11**: 1067–1075.
- Taylor, J.A., Walsh, K.A. and Johnson, R.S.** (1996) Sherpa: a Macintosh-based expert system for the interpretation of electrospray ionization LC/MS and MS/MS data from protein digests. *Rapid Commun. Mass Spectrom.* **10**: 679–687.
- Valaskovic, G.A., Kelleher, N.L., Little, D.P., Aaserud, D.J. and McLafferty, F.W.** (1995) Attomole-sensitivity electrospray source for large-molecule mass spectrometry. *Anal. Chem.* **67**: 3802–3805.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E. Jr, Hieter, P., Vogelstein, B. and Kinzler, K.W.** (1997) Characterization of the yeast transcriptome. *Cell* **88**: 243–251.
- Wahl, J.H., Goodlett, D.R., Udseth, H.R. and Smith, R.D.** (1993) Use of small-diameter capillaries for increasing peptide and protein detection sensitivity in capillary electrophoresis-mass spectrometry. *Electrophoresis* **14**: 448–457.
- Walsh, B.J., Molloy, M.P. and Williams, K.L.** (1998) The Australian Proteome Analysis Facility (APAF): assembling large scale proteomics through integration and automation. *Electrophoresis* **19**: 1883–1890.
- Williams, E.R., Henry, K.D., McLafferty, F., Shabanowitz, J. and Hunt, D.F.** (1990) Surface-induced dissociation of peptide ions in Fourier-transform mass spectrometry. *J. Am. Soc. Mass Spectrom.* **1**: 413–416.
- Wilm, M. and Mann, M.** (1994) Electrospray and Taylor–Cone theory, Dole's beam of macromolecules at last. *Int. J. Mass Spectrom. Ion Processes* **136**: 167–180.
- Wilm, M., Neubauer, G. and Mann, M.** (1996) Parent ion scans of unseparated peptide mixtures. *Anal. Chem.* **68**: 527–533.
- Yates, J.R. III., Speicher, S., Griffin, P.R. and Hunkapiller, T.** (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**: 397–408.
- Yates, J.R. III., Eng, J.K. and McCormack, A.L.** (1995a) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* **67**: 3202–3210.
- Yates, J.R. III., Eng, J.K., McCormack, A.L. and Schieltz, D.** (1995b) Method to correlate tandem

mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67**: 1426–1436.

Yates, J.R. III., Eng, J., Clauser, K.R. and Burlingame, A.L. (1996) Searching sequence databases with uninterpreted high-energy CID spectra of peptides. *J. Am. Soc. Mass Spectrom.* **7**: 1089–1098.

Yates, J.R. III., Morgan, S.F., Gatlin, C.L., Griffin, P.R. and Eng, J.K. (1998) Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Anal. Chem.* **70**: 3557–3565.

第八章 蛋白质组学的自动化： 高通量蛋白分析的技术和信息学途径

Marc R. Wilkins

1. 历史回顾

诞生于 20 世纪 70 年代中期的二维电泳 (2-DE) 描述了以等电聚焦为基础的一相电泳和根据分子质量进行分离的二相电泳的结合 (Kenrich and Margolis, 1970; Klose, 1975; O'Farrell, 1975; Scheele, 1975), 其分辨能力之强, 可以在一块胶上分离出 1000 多个大肠杆菌蛋白 (O'Farrell, 1975), 这就给研究者提示可以通过用 2-DE 技术对生物系统内的复杂蛋白进行定性和定量分析, 因此这一研究技术的巨大潜力迅速被一些人接受, 用于有机体的多种蛋白的研究。此外, 差异显示 2-DE 技术也被用于寻找蛋白的氨基酸改变或蛋白的转移后修饰 (Goodman, 1976; Steinberg *et al.*, 1977)。

介绍完 2-DE 蛋白分离技术后, 将介绍针对不同类型样品及不同分析目的的 2-DE 的应用。通过制备不同条件下的样品可反应基因表达水平的差异 (O'Farrell and Ivarie, 1979); 另外, 通过比较生物体的正常和突变株也可反应基因表达水平的差异 (Reeh *et al.*, 1976)。也有一些研究者利用 2-DE 技术寻找蛋白复合物中蛋白间的相互作用 (Brigggs and Capaldi, 1977)。1981 年, 出现了第一个用于 2-DE 胶分析的软件包 (Anderson *et al.*, 1981)。因此, 从许多方面来讲, 2-DE 作为研究复杂样品的手段在 20 年前就已经是可行的了。如今, 随着对组织的 2-DE 电泳分离为基础的蛋白质组概念的提出, 这些想法得到了进一步的发挥 (Anderson and Anderson, 1982)。

1.1 蛋白化学家面临的困难

2-DE 的分离技术在 80 年代还未得到实现的原因有以下几方面: 首先, 数据库中没有足够的核酸和蛋白序列信息, 因此就无法通过数据库中序列信息很好地了解各种生物体的多种物理和生理活动。这是因为当时分子生物学技术还处于幼稚期, 其后诞生了一些新技术, 如用手工操作测核酸序列的双脱氧链终止法 (Sanger *et al.*, 1977), 用 PCR 技术扩增 DNA (Saiki *et al.*, 1985) 等。而依赖于荧光染料化学技术的全自动测序仪 (Prober *et al.*, 1987) 于 1992 年首次完成了对真核生物染色体的成功测序 (Oliver *et al.*, 1992)。

其次, 2-DE 的潜力还未得到充分发挥的重要原因在于该技术在蛋白鉴定方面的应用的有限性。Towbin (Towbin *et al.*, 1979) 建立的将蛋白转印至尼龙膜的方法以及 Burnette (Burnette, 1981) 建立的用抗体检测转印蛋白的方法不失为一个灵敏的鉴定方法。然而, 这些方法的不足之处在于只有少数蛋白具有相应的抗体, 此外单抗或多抗的制备是一个漫长且费时的工作。全自动 Edman 降解测序 (Edman and Begg, 1967)

要么直接测定新蛋白的序列，要么同数据库中的基因序列相比较，从而鉴定该蛋白。然而，早期的液相测序仪所需的蛋白量是 2-DE 胶上纯化蛋白的几百倍。直到 1987 年 Matsudaira (1987) 将蛋白转印至 PVDF 膜的技术和气相测序技术相结合，使得蛋白分析的灵敏度达到了 pmol 水平。就拿模式生物来讲，蛋白鉴定的最有效的方法仍然是将未知蛋白同已知的纯蛋白在 2-DE 胶上共分离 (Philips *et al.*, 1980)。

由于蛋白鉴定的复杂性，许多科学家在 20 世纪 80 年代就将研究重点转移到了研究在不同条件下蛋白表达量的改变，建立了大量的共刺激和共调节的蛋白数据库，如大肠杆菌 (Van Bogelen *et al.*, 1990)、兔 REF52 细胞株 (Garrels and Franza, 1989) 及小鼠的胚胎 (Latham *et al.*, 1991) 等。同理，通过监测在多种毒剂作用下小鼠肝蛋白的改变可进行毒理学研究 (Anderson *et al.*, 1987)。在这些实验中，通过不同的处理，可观察到 300~1600 个相关和相匹配的蛋白，更关键的是这些研究反应了蛋白质组群间的相互作用，且早于同类的微阵列核酸杂交技术若干年 (Schena *et al.*, 1995; Wodicka *et al.*, 1997)。然而，共调节和共抑制蛋白的科学价值的实现还必须依赖于这些组群中蛋白的鉴定，若做不到这一点，这些蛋白也仅仅是胶上的一些点而已。

1.2 蛋白质组学的发展史

20 世纪 90 年代在不同领域出现的各种进展相互结合形成了一门新学科，伴随者 IPG 胶条的广泛应用，2-DE 技术才真正成为了一种微量制备技术 (Bjellqvist *et al.*, 1993)，随着模式生物和其他生物体基因组序列的完成，(如支原体、大肠杆菌、酵母、线虫)，使得基因和蛋白序列的数据库迅速增加。而质谱也从一种专门的工具变成了一种每个蛋白化学实验室的常规仪器。肽质指纹图技术 (Henzel *et al.*, 1993)，因其简化了数据库中基因序列和蛋白序列的联系，大大提高了蛋白鉴定的灵敏度和速度 (Kuster and Mann, 1998; Yates, 1998 a, 1998 b)。伴随着以上技术的发展，1995 年由 Wilkins 提出的蛋白质组的概念应运而生 (Wilkins *et al.*, 1995)。同分子生物学取得的进展相比，蛋白质组这一概念主要强调和证实以前未获得足够重视的科学。

新千年伊始，蛋白质组学获得了前所未有的发展。随着果蝇基因组的完成及 2000 年人类基因组计划的即将完成，其他的一些模式生物的测序任务也即将开始，如小鼠和拟南芥。为了充分利用这些生物体的基因组序列提供的价值，就有必要制订生物学的下一个大的目标，即功能基因组学或功能蛋白质组学。功能基因组学着重依赖于高密度的微阵列技术 (Schena *et al.*, 1995; Wodicka *et al.*, 1997)，该技术的优点在于其高通量的分析能力，且可同时检测 mRNA 的表达水平。而蛋白质组学则主要解决蛋白的表达水平、加工和修饰、蛋白间的相互作用及蛋白的功能 (Wilkins *et al.*, 1997a)。目前，蛋白质组学面临的挑战有以下两点：第一，全自动、高通量的蛋白质组技术的完成；第二，建立大量的蛋白数据库和能够有效分析处理蛋白质组数据的分析软件。本综述主要介绍以上两点。

2. 蛋白质组学的自动化——技术解决途径

直到最近，蛋白的研究仍然停留在一次解决一个蛋白的基础上。研究者的研究重点往往局限在单个蛋白，通常是一个一个地纯化蛋白，再一个一个地研究。而蛋白质组学

同传统的研究方法有所不同,无论是蛋白复合物的鉴定也好 (Rigaut *et al.*, 1999),还是生物体所表达的全蛋白也好 (Fountoulakis *et al.*, 1997, 1998a, 1998b; Langen *et al.*, 1997),通常是同时研究多个蛋白,但前提是要满足蛋白鉴定的速度、灵敏度和精密度。这就使得蛋白研究者可充分利用 2-DE 技术,同时平行地纯化多个蛋白。

最近对蛋白质组学影响最大的当属质谱,其中 ESI 或 MALDI-TOF 结合四极管、离子阱或飞行时间质量检测器等,使得蛋白的分析变得简单易行 (Yates, 1998b)。更重要的是,质谱技术使 2-DE 胶上银染可见的蛋白也能得到分离鉴定 (Willms *et al.*, 1996)。但目前仍困扰研究者的问题是,尽管可通过平行的 2-DE 胶分离纯化成百上千的蛋白,但仍然无法采取一种简单易行的办法来解决质谱鉴定前大量蛋白的制备。况且,对于质谱来讲,每天分析的样品是有限的,尤其是 ESI 质谱。面临着这些挑战,蛋白分析的自动化已经得到了一定发展,可分为三组技术:2-DE 胶技术、1-DE 胶技术和不以胶为基础的分选技术。我们将依次讨论这些技术以及为实现 2-DE 分离过程的自动化所做的努力。

2.1 2-DE 胶的自动化

考虑到 2-DE 仍然是蛋白分离的首选方法,故实现 2-DE 的自动化对于高通量蛋白质组学是必需的。在过去的几年内,已经实现了多种形式的自动化,如小胶的分离 (50mm×43mm) 已完全自动化 (Brewer *et al.*, 1986),但我们认为这种分离方式还不能满足蛋白质组分析所需要的蛋白量。因此人们致力于实现大胶的自动化 (Harrington *et al.*, 1993; Nokihara *et al.*, 1992)。一种由 Large Scale Biology Corporation 花了多年时间建立的名叫 BioMetre™ 的平台有望实现整个 2-DE 过程的自动化。并且,连接一相和二相分离的方法已获专利保护 (Hochstrasser, 1989, 1998),该方法可产生二维一体的具有自动化潜力的 2-DE 胶。尽管作出了多种努力,但目前还不能实现 2-DE 的全自动。相反,制造商和研究人员把主要精力放在了实现单步骤的自动化上(如银染),或通过平行方法提高通量,而不是把重点放在整个步骤的全自动,因此就产生了用于一相和二相胶的分选装置,可同时制备、分离多个一相或二相的胶,从而提高了分离的重复性。

2.2 2-DE 胶的全自动

蛋白质组经 2-DE 分离后,可获得成百上千感兴趣的蛋白点去作质谱分析。针对 2-DE 胶上蛋白或多肽进行的质谱分析的自动化研究,已有多种不同的解决办法,有的已商业化。

用机器人取点进行 MALDI-TOF 鉴定 以前由人工操作的蛋白鉴定步骤,现在已实现了自动化的流水线(图 8.1)。文献中已有报道采用机器人装置从胶上获取蛋白点 (Traini *et al.*, 1998),也有商品化的产品供应(如 Flexys Robotic Workstation from Genomic Solutions)。一般的,机器人通过数码相机给胶照相,用剪头切胶,并将胶块送入 96 孔板,紧接着液相处理装置开始洗胶块并加入蛋白内切酶,这也可以通过改进的液相处理工作站完成 (Packard Multiprobe 104 in Traini *et al.*, 1998),最后由同一机器人将酶解后的肽段送到 MALDI 样品板进行分析或由新的机器人完成(如 ProGest robot

from Genomeic Solutions, 已由 Houthaeve 等商业化, 1997)。ProGest 机器人与称为 ProMS 的第二组机器人联合, 主要完成肽的脱盐, 并送至 MALDI 样品板。目前的 MALDI 质谱对多个样品的分析则相对较容易, 因为大多数生产厂家已完成该方法的自动化。根据仪器的不同, MALDI 样品板可放 100~384 个样品, 可以一直不间断地进行大量样品的分析。早期的自动化 MALDI-MS 的分析可参见 Schevchenko 等 (1996)。

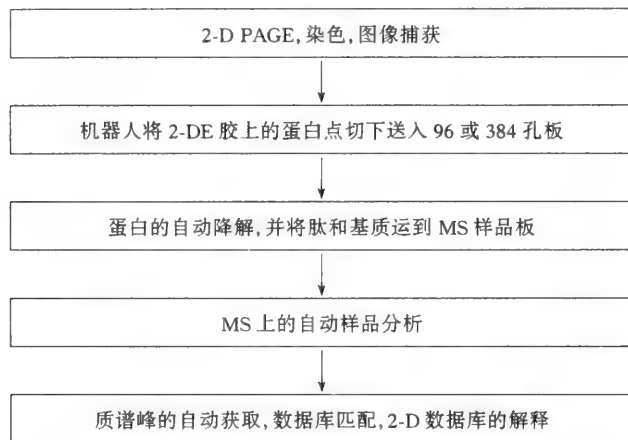


图 8.1 蛋白质组学的自动化流水线进程
该方法实现了手工操作和技术的自动化

以上方法的优点在于质谱的数据获取简单易行。胶上蛋白点的 x , y 坐标首先用 TIF 或其他图像格式记录下来, 然后将图像上的点与 96 孔板中的特定位置依次对应, 96 孔板的编号又可作为 MALDI 分析时样品号。我所在的实验室所采用的技术与上面介绍的相似, 作用强大且可实现高通量。用两天的时间, 一台 MALDI-TOF 就可完成 300 个蛋白的制备和分析鉴定, 这样的高通量是其他的自动化系统所无法比拟的。

用于 MALDI-MS 蛋白鉴定的 2-DE 胶的网割 Hoffmann-La Roche 的科学家对以上的方法作了一些改进 (Langen *et al.*, 1998)。该方法不是采用一个一个蛋白点分别切的方法, 而是将整块 2-DE 胶用网格工具切割成小块, 如 $1\text{mm} \times 1\text{mm}$ 大小, 胶块随后被自动送到 MS 分析前的 96 或 384 孔板作酶解, 该技术从哲学角度讲, 具有完整性, 因为不管胶块上有无蛋白, 每一块都得到了酶解和分析, 因此即便是机器人无法检测或切到的低峰度点也能进行质谱分析。但不足之处在于, 通过网格切割的胶块往往会出现多个蛋白出现在一块胶上或者一个蛋白出现在多块胶上的情况, 这就给数据分析提出了挑战, 需将单个蛋白点的数据进行正确编辑。然而, 由此引出了一个有意思的问题, 可将每一丙烯酰胺胶块上的光谱总离子数 (TIC) 绘制成 2-DE 胶分析图谱, 然而, 其分辨率取决于切下的丙烯酰胺胶块的大小。

通过膜和 MALDI-MS 进行 2-DE 分离蛋白的鉴定 尽管该方法被证明是切实可行的, 但以上两种蛋白鉴定的自动化方法对胶块的处理复杂, 且这些技术在某种程度上对所需步骤的选择不够精练。基于此, 一些研究组试着用膜为基质来吸附并分析蛋白 (图 8.2)。早期的研究表明, 当蛋白从 PVDF 膜上解吸下来后, 可通过 MALDI-MS 测定所有蛋白的质量, 包括将膜贴到样品孔, 加入基质, 通过紫外或红外线激光将膜表面的蛋

白解吸下来 (Eckerskorn *et al.*, 1992; Vestling and Fenselau, 1994)。然后将膜在 MS 内进行高精度扫描, 如 30×30 像素, 从而获取转印蛋白的 2-DE 图谱 (Eckerskron *et al.*, 1997; Stoeckli *et al.*, 1999)。随着蛋白降解技术的发展, 可使酶解后的肽段集中出现于膜上 (Bienvenut *et al.*, 1999), 然后对膜进行扫描就可获得 2-DE 胶上蛋白点的真实图谱, 并且所获得的肽质量有助于蛋白鉴定 (Binz *et al.*, 1999)。由于 MALDI-MS 可自动扫描膜, 就提高了蛋白鉴定的通量。

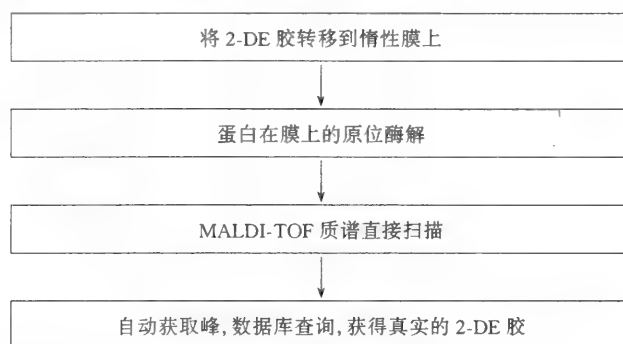


图 8.2 蛋白质组学的自动化包括将 2-DE 胶转移到膜上, 酶解成肽段, MALDI-TOF 质谱扫描与图 8.1 表述的手工操作步骤的简单自动化有所不同

膜扫描技术的优点在于可平行地将膜上的蛋白进行原位消化 (Binz *et al.*, 1999; Schleuder *et al.*, 1999), 只有在样品处理时需将膜放到质谱板上, 此外, 吸附到膜上的蛋白稳定时间长, 膜容易保存。然而, 目前扫描技术的缺点在于: 第一, 速度慢, 举个例子来讲, 目前分析一块 $16\text{cm} \times 16\text{cm}$ 的 2-DE 胶要花 36 天 (Binz *et al.*, 1999)。第二, 蛋白从膜上消化和扫描 MALDI-TOF-MS 技术所得到的肽片段数均低于标准的消化方法和 MALDI-TOF-MS 技术, 这就限制了蛋白鉴定的效率和蛋白转移后的修饰的鉴定。然而, 可通过改进计算机和质谱的软件和硬件, 提高该方法的效率。

LC-MS 对 2-DE 胶上分离蛋白的自动化鉴定 2-DE 胶上的蛋白经消化后, 除了可用 MALDI-TOF-MS 分析外, 还可以采用液相色谱质谱 (LC-MS) 技术, 后者不仅可检测肽质量, 还可用诱导碰撞解离 (CID) 技术提供断裂的可能性, 从而产生用于蛋白鉴定的肽序列标签 (Mann and Wilm, 1994), 而肽段只有经过有效的断裂才可产生肽序列 (Qin *et al.*, 1998)。

一旦样品送入 LC-MS, 就能有效地产生并分析数据, 已有多篇报道涉及到 2-DE 胶上蛋白的 LC-MS 分析 (Gygi *et al.*, 1999)。然而, 该方法的通量比 MALDI-TOF-MS 至少低一个数量级, 每天只能分析 20 个样品 (Link *et al.*, 1997)。这是因为该技术需要与毛细管电喷射或称之为纳流喷射的仪器相结合以提高分析的灵敏度和样品的数据量 (Wilm and Mann, 1996), 毛细管和纳流喷射装置易引起堵塞, 并且难以实现自动化。然而, 位于 96 孔板和 LC 系统间的自动加样器接口已商业化, 可显著提高样品通量。更令人兴奋的是, 研究者已研制出了用微纤维装置生产出的与离子捕获或四极质谱相连接的电渗泵 (Figeys *et al.*, 1997, 1998), 这有助于实现从胶到 LC-MS 的自动化, 但目前还不清楚该系统究竟有多大潜能。

2.3 2-DE 自动化

对蛋白质组学的关注,使得研究者开始寻找单细胞或组织蛋白质组的最佳分离鉴定方法。2-DE 是目前蛋白分离技术中分辨率最高的,无论采用大尺寸的胶还是采用窄的 IPG 胶条多次分离,目前可分离接近 10 000 个蛋白 (Klose, 1999),样品的分步提取技术把疏水性作为分析的第三相 (Herbert, 1999; Klose, 1999),该技术对蛋白的鉴定可增加 20%。除了这些特点,2-DE 胶作为一种平行的分离技术,有其局限性,最值得注意的是对低丰度蛋白,极疏水的蛋白和极碱性蛋白的无能为力 (Wilkins *et al.*, 1996, 1998a)。即便是数据结果显示具有高重复性,许多实验者仍旧关注实验室内和实验室间分离的重复性 (Corbett *et al.*, 1994; Lopez and Patton, 1997)。这些问题已经激起对 2-DE 胶替代分离技术的研究,这些技术可分为两大类:1-DE 胶;以非胶为基础的分

离技术。
与 1-DE 胶联用的自动化技术 MALDI-TOF-MS 用于等电聚焦胶表面的直接扫描可产生真实的 2-DE 图谱 (Ogorzalek Loo *et al.*, 1997; Loo *et al.*, 1999)。1-DE 的一相分离通过计算蛋白的等电点在 IPG 胶条上的位置而得以实现,以分子质量为基础的二相分离则由 MS 完成,从而取代了 SDS-PAGE 分离。该技术的优点在于 IPG 胶条是以平衡为基础进行的分离,而 MS 显然比非平衡的 SDS-PAGE 二相分离更准确。一旦数据的获取和分析实现了自动化,1-DE 就可产生大量蛋白高准确度的分子质量数据和真实的 2-DE 图谱。

从技术上讲,1-DE 首先在 IPG 胶条上用标准的等电聚焦技术分离蛋白,然后进行洗涤、在白芥子酸基质中浸泡、风干,然后将 IPG 胶条直接送入 MALDI-TOF-MS,选择与 IPG 胶条间隔距离相近的扫描间距 (如 $30\mu\text{m}$),各种蛋白离子化后,在不同的位置检测到其分子质量,扫描胶条产生的系列光谱可提供同 2-DE 相似的图谱 (Ogorzalek Loo *et al.*, 1999)。

目前很难比较 IPG 胶条直接扫描的结果和 2-DE 分离的结果,因为 2-DE 胶图谱的结果来源于整个 IPG 胶条,但 1-DE 也有一些缺陷。首先,只有一些特定的蛋白才能在 MALDI-TOF 中很好离子化,一些高分子质量的蛋白则很难离子化,这就影响了蛋白的定量。其次,无法判定直接 MALDI-TOF 的分辨能力是否能与 SDS-PAGE 胶比美。最后,由于扫描技术可产生高准确度的 2-DE 图谱,但无法通过肽质指纹图鉴定感兴趣的蛋白,只能将蛋白从 IPG 胶上切下来后,用溴化氰消化成肽段 (Loo *et al.*, 1999),但蛋白混合物可出现在 IPG 胶条的任何位置。此外,可通过小型生物的基因组信息及 IPG 扫描技术的参数准确鉴定蛋白的分子质量和等电点,从而鉴定该蛋白 (Cavalconi *et al.*, 1997; Wilkins *et al.*, 1998b)。

不以胶为基础的蛋白质组分离的自动化 目前出现的一些新的可取代以胶为基础的分

离技术指的是一相或二相色谱结合紫外和 (或) 电散射质谱。一些研究者结合自动化的大小排阻色谱和反向色谱,根据一相的蛋白大小和二相的蛋白疏水性分离蛋白而获得紫外色谱图 (Opiteck *et al.*, 1998),从而分离蛋白,甚至是全蛋白质组 (Opiteck *et al.*, 1998; Isobe *et al.*, 1991)。该技术分离过程根据操作条件不同需 6~17h。用该方法的目的就是为了获得蛋白质组成分,在大肠杆菌的细胞裂解液可分离出 250 种蛋白组

分，比银染的 2-DE 分辨率低一个数量级 (Tonella *et al.*, 1998)。很明显，需要再进行第三相的分离才能满足蛋白质组分析的目的。

一些研究者以样品中蛋白的肽为分析目的，丰富了以蛋白为分析目的 2-D 色谱的应用 (图 8.3)。该技术利用三级、离子阱或四级飞行时间质谱使肽裂解，然后将所获信息同序列数据库比较，从而鉴定该蛋白 (Yates, 1998b)。此方法可鉴定相对简单的样品，如蛋白复合物 (Link *et al.*, 1999) 或全蛋白质组 (S. D. Patterson, 未发表) 中的蛋白成分。该方法首先用胰蛋白酶等消化样品中的所有蛋白，然后将肽上样到反向色谱柱中进行分离，最后将肽碎片直接收集到电喷射质谱中。质谱能够从肽混合物中选择感兴趣的离子，将其自动打断，产生序列标签或完整序列，如此反复将其他离子打断。一次液相色谱运行即可断裂数以百计的肽，然后自动地用 SEQUEST 等软件将数以百计的肽同序列相联系 (Yates *et al.*, 1995)。目前，在这一领域作得比较出色的当属 Amgen Inc 对人尿的分析 (S. D. Patterson, 未发表)，一次实验可鉴定约 100 个蛋白的肽。该技术的优点在于其高通量和内在的冗余使得可观察每个蛋白的多个肽段，但该技术产生了大量的数据，这就给实验结果的数据分析带来了麻烦。而且，由于该技术是建立在对肽而非蛋白的基础上，就使得该技术无法鉴别结构相近的异构体，也无法对一个单一的蛋白进行进行详细鉴定。

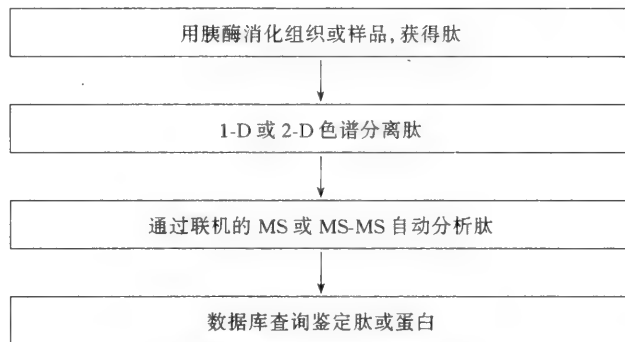


图 8.3 自动化蛋白质组的分析

将全部样品消化成肽，然后通过 HPLC 分离样品中的所有肽，最后用 MS 或 MS-MS 分析。

然而，该技术提供的样品信息不如 2-D 胶技术。

3. 蛋白质组学的信息解决途径

蛋白质组学从许多方面来讲都是一门信息科学。就蛋白自身而言，蛋白质组学涉及到序列信息、加工和修饰的信息、表达信息、功能信息、结构信息及蛋白如何同其他蛋白形成小的或大的蛋白复合物的信息。重要的是，蛋白质组学考虑到了时间和环境的影响，因此可以在不同的发育时间点和不同的环境条件下观察以上的蛋白参数。从更广泛的角度讲，有关样品的记录和结果的分类是蛋白质组研究的另一重要内容。高通量的筛选需要实验室信息管理系统 (LIMS) 来记录样品的分析进程及其分析结果，然后将相关的结果建立成有效的数据库以便于将来的查询。蛋白质组工具在蛋白质组大量数据的分析中发挥重要作用，有助于蛋白的鉴定。我们都知道，即便是酵母这样小的生物也能

合成 6000 多种蛋白，那么更复杂的生物可能有 200 种不同的组织型，每一种合成 10 000 种蛋白。因此，从这方面来讲，蛋白质组学面临的任务是艰巨的。故我们在本综述后续部分写了有关蛋白质组信息资源的种类，分为蛋白质组数据库和蛋白质组工具，我们将介绍一些在目前对研究者有用的知识，其他的以后陆续讨论。

3.1 蛋白质组数据库

从某种程度上来讲，基因组比较好定义，指的是全部基因的碱基序列。根据基因组序列就可以推导出可读框、外显子、内含子、启动子、增强子和染色体的结构区域（如中心粒和端粒）。相反，由于涉及到的数据类型不同和生物系统的蛋白的动态变化，就使得很难给蛋白质组下一个定义。完整蛋白质组资源可来源于数据库中核酸序列编码的蛋白序列。因此，目前没有一个数据库能包含蛋白质组的各个领域，但出现了蛋白质组学相关领域的各种数据库。

目前在蛋白质组学领域用得最广泛的数据库当属 SWISS-PROT (Bairoch and Apweiler, 2000; <http://www.expasy.ch/>)，这是一个由日内瓦的瑞士生物信息学研究所、瑞典和欧洲生物信息学研究所建立的蛋白注释性的数据库。该数据库除了含蛋白序列信息外，还包含了大量的注释（图 8.4），如蛋白表达、功能、结构域、突变体、蛋白加

The screenshot shows a Netscape browser window titled "NiceProt View of SWISS-PROT: P00001 - Netscape". The address bar shows the URL <http://www.expasy.ch/cgi-bin/protsearch.fcgi?P00001>. The main content area displays the SWISS-PROT entry for P00001, which is human cytochrome c. The entry includes a "Keywords" section with terms like "Mitochondrion, Electron transport, Respiratory chain, Heme, Acetylation, Polymorphism". Below this is a "Features" section with a table of features and their counts, and a list of specific features. The "Sequence information" section provides details on the protein's length (104 AA), molecular weight (11617 Da), and a checksum (CRC64 D47C9B513DF1C5C2). The amino acid sequence is displayed in a grid format, with residues grouped by position (10, 20, 30, 40, 50, 60, 70, 80, 90, 100). The sequence is: GDVEKGGKIF IMKCSQCHTV EKGGRKHTGP NLHGLFGRKT GQAPGYSYTA ANKNKGIWV BDTLMEYLEN PKKYIPGTKM IFVGIKKKBE RADLIAYLKK ATNE. The entry is identified as "P00001 m" in "FASTA format".

Feature	Count	Count
INIT_MET	0	0
MOD_RES	1	1
BINDING	14	14
BINDING	17	17
METAL	18	18
METAL	80	80
VARIANT	65	65

ACETYLATION.
HEME (COVALENT).
HEME (COVALENT).
IRON (HEME AXIAL LIGAND).
IRON (HEME AXIAL LIGAND).
M -> L (IN 10% OF HUMAN).
/FTId=VAR_002204.

Length 104 AA Molecular weight 11617 Da CRC64 D47C9B513DF1C5C2 [This is a checksum on the sequence]

10 20 30 40 50 60
GDVEKGGKIF IMKCSQCHTV EKGGRKHTGP NLHGLFGRKT GQAPGYSYTA ANKNKGIWV
70 80 90 100
BDTLMEYLEN PKKYIPGTKM IFVGIKKKBE RADLIAYLKK ATNE

P00001 m
FASTA format

图 8.4 人细胞色素 C (注册号 P00001) 的部分 SWISS-PROT 入口，显示了特征表和蛋白序列。注意，该蛋白的特征表提供了蛋白的翻译后修饰、金属结合位点和一个已知的突变体。

在较早的入口中显示了蛋白的功能信息。

工和转移后的修饰等信息。该数据库同其他数据库, 如 Medline、GeneBank、PDB 结构数据库及其他专业性的数据库, 如图谱数据库、代谢或生物体特异的数据库间可超级链接。由于该数据库有大约 84 000 个入口, 同 GeneBank 这样的序列储藏所相比, 就显得太小了, 但由于其信息的高质量, 就使之成为了蛋白数据库的金标准。

但目前 SWISS-PROT 面临着挑战, 即基因组测序和发表的速度制约了 SWISS-PROT 无法即时收录所有生物所发表的相关信息, 更不要说 SWISS-PROT 内成千上万的物种包含的一些信息。因此, Proteome Inc. 公司采取了一个补救措施以弥补 SWISS-PROT 的不足, 即重点关注一些模式生物或有商业价值的生物。Proteome Inc. 公司建立的酵母蛋白质组数据库 (YPD); (<http://www.proteome.com/>) 是目前最完善的酿酒酵母的数据库, 蛋白质组内的所有蛋白均可通过 SWISS-PROT 的格式进行解释。此外, 还可提供其他信息, 如密码子偏倚、染色体定位、基因结构。该数据库还提供了有关酵母知识的广泛的综述和参考文献及有关的蛋白功能和验证实验的详尽信息, 转录图谱, 表达图谱的搜寻和分析及每个蛋白的序列相似性报告。Proteome Inc 还建立了一些与酵母相似的数据库, 如白色念珠菌生物, 秀丽线虫, 并获得了资助以进一步研究建立病原体肺孢子囊虫、烟曲霉菌及真菌模式生物构巢曲霉、非洲粟酒裂殖酵母等的数据库。从大众化的角度讲, 一些生物特异性的数据库资源的完整性和实用性通常反应了一个生物研究团体的组织有效性。一些比较好的大众化的生物数据库包括 FlyBase (<http://flybase.bio.indiana.edu>) 及 *Arabidopsis thaliana* 数据库 AAtDB (<http://genome-www.stanford.edu/Arabidopsis/>)。

一种不同类型的蛋白质组数据库就是 2-DE 蛋白表达数据库 (图 8.5), 此类数据库以生物、细胞或组织的 2-DE 胶分离为基础, 数据库中包含了 2-DE 胶图谱, 相关的研究样品及胶上鉴定的蛋白的信息。大多数 2-DE 蛋白表达数据库的网址如下: <http://www.expasy.ch/ch2d/2d-index.html>。MS, 此类网址的出现使得 2-DE 胶上大量蛋白点的鉴定成为可能, 多数 2-DE 数据库包含了数百个鉴定的蛋白。某些数据库, 如 SWISS-2DPAGE 数据库, 涉及到 2-DE 胶上每个鉴定的蛋白点的大量文本信息, 包括蛋白的分子质量、等电点、鉴定方法、蛋白在其他组织表达的资料及病理信息。最近, 2-DE 蛋白表达数据库已经开始反应出一个样品采用多块胶分离的必要性, 一些数据库现在开始对样品采取不同的分离方法, 如窄等电点范围分析。但目前 2-DE 胶数据库面临的主要挑战在于其可靠性, 体现在 2-DE 分离水平和数据处理方法两方面。由于近几年在 2-DE 样品制备和电泳条件的标准化方面做出了诸多努力, 且对蛋白质组学兴趣的增长使得人们采取了前所未有的多种多样的技术用于制备 2-DE 参考图谱。这些技术在实验室内具有高重复性, 但在其他实验室的重复性值得商榷。由于应用于 2-DE 数据库的图谱、文本和超级连接的贮存和描述方法有一定的一致性 (Appel *et al.*, 1996), 就具备了一定的实用性。但只要数据库中 2-DE 胶采用不同的分离方法, 就无法进行实验室间的比较, 2-DE 数据库资源的应用就会受到限制。然而, 值得注意的是广泛应用于鉴定 2-DE 分离蛋白的肽质谱信息提供了不依赖于实验室的蛋白质参数, 并且这些信息对建立数据库非常有用。

对蛋白质组来说, 它还有许多其他重要的数据库。由于数据库日新月异, 在此不可能全部作评价。然而理论上它们可包括在下面几个方面的数据库内: 核酸序列, 蛋白序

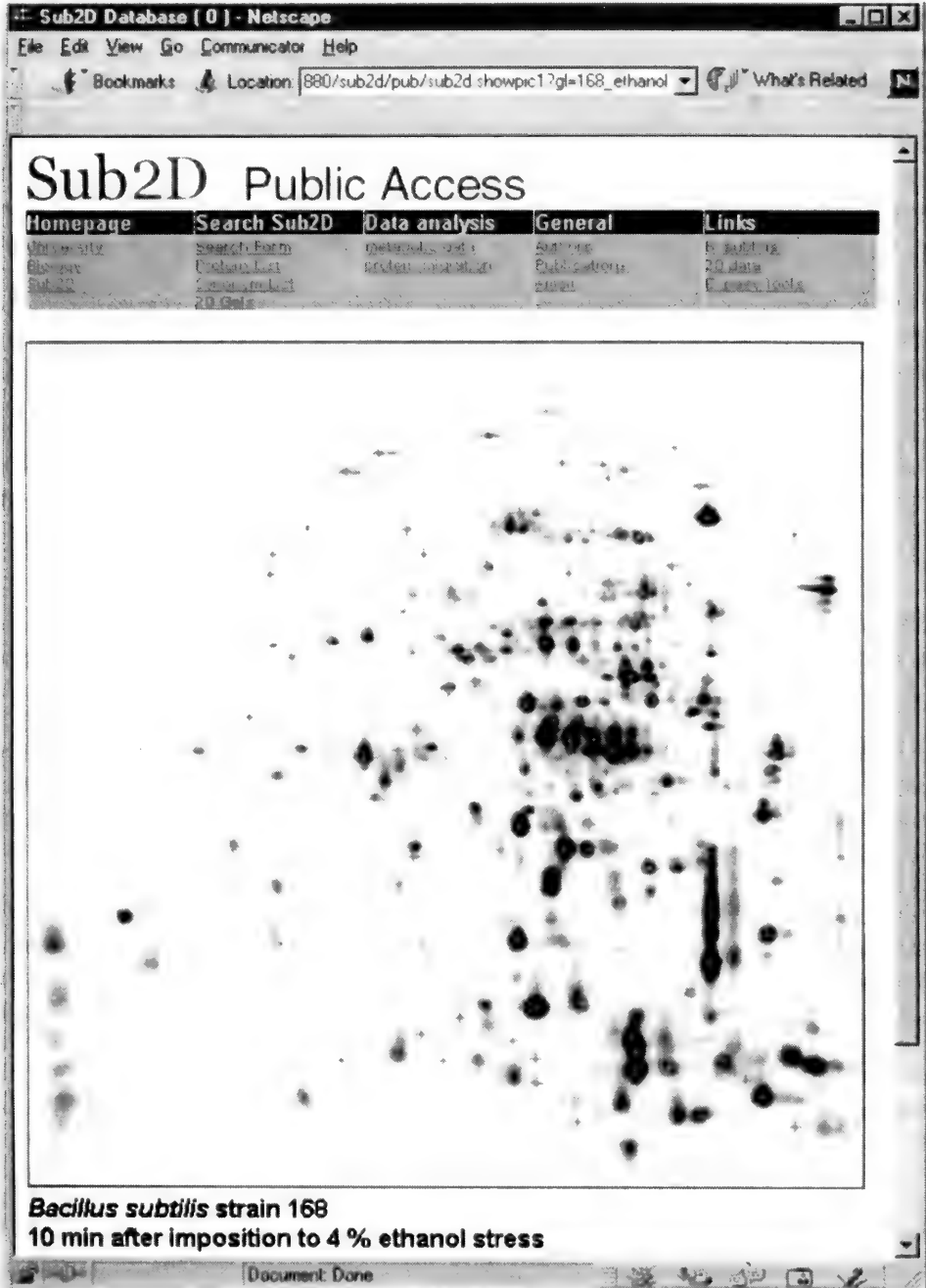


图 8.5 枯草芽孢杆菌的 Sub2D 数据库中的一块注解的 2-D 胶

图中的交叉点代表可点击链接至特定蛋白质的详细说明页面，如蛋白质序列、表达水平等。

大多数 2-D 胶数据库都使用类似的点击链接系统。

列，蛋白图谱、表达谱和结构域，蛋白翻译后修饰，蛋白质的三维结构，蛋白质代谢数据库。由 Amos Bairoch 管理的网站毫无争议地是最好的蛋白质数据库网站，它对蛋白质数据库进行分门别类，其目录下有 1000 多个相关网站，网址为 <http://>

www.expasy.ch/alinks.html.

3.2 蛋白质组学工具

随着许多相关的蛋白质组数据库的建立,近年来出现了许多蛋白质组辅助分析的信息化工具。这些工具解释和注释了 2-DE 电泳图谱,并且可以对从 MS-MS 质谱中产生的多肽从头测序。我们将简短讨论一下这些工具并且评价它们对高通量蛋白质组筛选的应用。

2-DE 胶成像分析软件 1981 年起,已经有辅助分析胶图像的的软件。这些软件需要执行大量的核心任务,包括消减背景,对点进行检测和定量,比对胶与胶之间的图像,通过图像和统计来进行胶与胶的匹配,通过一个或更多的胶重合产生一个系统精细的图像,构建能够存取蛋白质点的数据库,许多的 2-DE 图像分析软件包开始是由学术研究机构开发的(例如 ELSIE, TYCHO, MELANIE 和 QUEST),其中大多数软件包现在性能大大提高了,并且已商品化,运用于 PC 平台上(例如 KEPLAR, MELANIE II 和 PDQUEST)。同时,基于 internet 网和内联网的分析软件也得到进一步发展(Lemkin *et al.*, 1999)。然而,值得注意的是尽管商业化软件包已升级,但自动化的胶分析软件还不能令人满意,2-DE 胶的复杂性意味着精确的图像分析耗时、耗力,但是如果首先把精力放在胶的制备、样品的准备、电泳分离和染色上,可以节省许多分析时间。如需要详细地解释计算机分析图像问题可以参看 Appel 和 Hochstrasser 的文献(1999)。

蛋白鉴定分析软件 由分析产生的大量蛋白参数可用于鉴定蛋白质。这些包括蛋白质量大小、等电点、蛋白质序列标签、肽质指纹图、肽序列标签和蛋白质氨基酸组成。通常在网可获得单参数和多参数相结合的软件,但是在某些情况下,可作为独立的软件包存在(例如 Micromass 公司的 ProteinLynx 质谱分析软件包)。

随着质谱在蛋白质组的广泛应用,大量的软件包服务于此种类型的研究。它们通常包括下列软件:通过肽质指纹图和不同的种属、蛋白质分子质量大小、等电点来鉴定蛋白质;通过 MS/MS 来鉴定蛋白质(MS 来源于 CID 或 MALDI-TOF PSD 技术);预测通过内源性蛋白酶消化肽片段或 MS/MS 的肽片段大小的工具;从头预测来自 MS/MS 数据的肽序列的工具。下列工具包就是这些软件的典范;Protein Prospector (<http://prospector.ucsf.edu>); PROWL (<http://prowl.reckefeller.edu>); MASCOT (<http://www.matrixscience.com>);来自华盛顿大学的 SEQUEST (Yates *et al.*, 1995)和在 ExPASy 服务器上的蛋白质组工具(<http://www.expasy.ch/tools/#proteome>);(Wilkins *et al.*, 1999a)。所有这些软件在鉴定一个蛋白质时是有效的,但是怎样应用于鉴定大量的样本,作到存在一定仍困难。看起来更复杂的是,如何自动挑取质谱峰送到蛋白质鉴定处,将结果存到其他数据库如二维电泳图数据库。一些研究小组已部分或全部解决这个问题(如 Berndt *et al.*, 1999; Gras *et al.*, 1999; Link *et al.*, 1999; Traini *et al.*, 1998)。很明显。随着硬件的发展,这些软件的出现意味着蛋白质鉴定的自动化几乎完成。

蛋白特征化软件 尽管蛋白质鉴定越来越自动化,但检测蛋白同工聚体、蛋白加工过程、翻译后修饰仍存在很多难题。然而研究者开始认识到,2-DE 胶的分辨率和质谱的准确度与解析度可应用于蛋白质性质分析和鉴定。SWISS-PROT 提供的蛋白质处理

和修饰数据库也可用来分析蛋白质的性质。PeptideMass 软件（预测内源性蛋白酶消化的肽质谱大小的工具）涉及到来自 SWISS-PROT 的蛋白质加工和修饰的信息，给出氨基酸修饰位置和修饰后肽的合适分子质量（Wilkins *et al.*, 1997b）。PeptIdent 工具 (<http://www.expasy.ch/tools/peptides.html>); (Wilkins *et al.*, 未发表) 是一个肽质指纹鉴定引擎，它能使使用者输入肽序列，在 SWISS-PROT 的修饰肽库和未修饰的肽库中搜寻，从而鉴定出蛋白的可能修饰的肽段。其他肽质指纹引擎也有相似的功能但只搜寻质量大小而没有考虑到现在库中存在的蛋白质修饰。一个在肽质谱数据库中预测新的翻译后修饰的工具已被报道（Wilkins *et al.*, 1999b），使用者指明蛋白序列和它的质谱数据，这个程序能搜寻对应于某一特定修饰的不同大小的蛋白质。那么就可以运用一系列规则去预测潜在的修饰肽序列中哪一个氨基酸可能被修饰。现有多于 20 种预测修饰的规则，一些规则非常简单，而另一些却是复杂的序列。GlycoMod 是 ExPASy 服务器上很好的蛋白质性质分析工具（N. Packer，私人通信），它能预测蛋白质糖基化位点。由于它与糖蛋白结构数据库相连，提供了一种研究糖蛋白的有力工具。

3.3 生物信息学展望

上述章节对蛋白质组生物信息资源作了概括，显然，这些资源在短时间内的的发展令人惊奇，然而，研究者们面对的主要挑战是在蛋白质组中需要的数据库和工具如何能被有机结合，并且重要的是在某一方面允许我们提出相关的生物学问题。例如，就像疾病状态经常是困难所在，我们需要追踪表达水平，以及在同一时间内修饰和拥有大量蛋白。这些可能在一个有机体的一种组织里（如肝脏），也可能存在于一种以上的组织（如肝和肾）。并且，我们需要筛选成千上万种微生物，检查与病原有关的蛋白。这些需要基于 mRNA 芯片的研究看是否表达了新蛋白，或者已经表达的蛋白是否经过不同的修饰或加工。另一个主要的挑战是大量的蛋白质组数据如何能被最好地显现。显然，一万或者 200 种不同人类组织的蛋白不能在一张 2-DE 电泳上展示。代替的是需要多维可视化软件，对有序及有颜色的排列是必需的，就像 Weinstein 等在一个人类细胞库中规模化筛选大量复合物所显示的那样（1997），提高数据库和软件的标准化是获得以上结果的必要条件，也是研究功能基因组的必要条件，有人提倡大规模 RNA 的分析可能是解决蛋白质组中的基本问题。

4. 结语

上述总结包括了蛋白质组的许多方面，但实际上只涉及到自动化领域及蛋白质组信息学，毫无疑问，它的重点在于技术领域。然而，我们必须时刻记住任何一种软件仅仅应用于某一研究领域才能产生预想的效果。鸟枪克隆法证实，没有哪一种软件是绝对正确的，而许多技术的结合才能产生生物学突破性的结果，蛋白质组现在正处于研究的中心地位，但我们必须问这样的问题，蛋白质组技术能够产生这样的效果吗？对于这个问题，我们不得不等待。

参考文献

- Anderson, N.G. and Anderson, L. (1982) The human protein index. *Clin. Chem.* **28**: 739–748.
- Anderson, N.L., Taylor, J., Scandora, A.E., Coulter, B.P. and Anderson, N.G. (1981) The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clin. Chem.* **27**: 1807–1820.
- Anderson, N.L., Giere, F.A., Nance, S.L., Gemmell, M.A., Tollaksen, S.L. and Anderson, N.G. (1987) Effects of toxic agents at the protein level: quantitative measurement of 213 mouse liver proteins following xenobiotic treatment. *Fundam. Appl. Toxicol.* **8**: 39–50.
- Appel, R.D. and Hochstrasser, D.F. (1999) Computer analysis of 2-D images. *Methods Mol. Biol.* **112**: 363–381.
- Appel, R.D., Bairoch, A., Sanchez, J.-C., Vargas, J.R., Golaz, O., Pasquali, C. and Hochstrasser, D.F. (1996) Federated two-dimensional electrophoresis database: a simple means of publishing two-dimensional electrophoresis data. *Electrophoresis* **17**: 540–546.
- Bairoch, A. and Apweiler, R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**: 45–48.
- Berndt, P., Hobohm, U. and Langen, H. (1999) Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis* **20**: 3521–3526.
- Bienvénut, W.V., Sanchez, J.-C., Karmime, A., Rouge, V., Rose, K., Binz, P.-A. and Hochstrasser, D.F. (1999) Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. *Anal. Chem.* **71**: 4800–4807.
- Binz, P.-A., Müller, M., Walther, D., Bienvénut, W., Gras, R., Hoogland, C., Bouchet, G., Gasteiger, E., Fabbretti, R., Gay, S., Palagi, P., Wilkins, M.R., Rouge, V., Tonella, L., Paesano, S., Rossellat, G., Karmime, A., Bairoch, A., Sanchez, J.-C., Appel, R.D. and Hochstrasser, D.F. (1999) A molecular scanner to automate proteomic research and to display proteome images. *Anal. Chem.* **71**: 4981–4988.
- Bjellqvist, B., Sanchez, J.C., Pasquali, C., Ravier, F., Paquet, N., Frutiger, S., Hughes, G.J. and Hochstrasser, D. (1993) Micropreparative two-dimensional electrophoresis allowing the separation of samples containing milligram amounts of proteins. *Electrophoresis* **14**: 1375–1378.
- Brewer, J., Grund, E., Hagerlid, P., Olsson, I. and Lizana, J. (1986) *Electrophoresis '86* (ed. M.J. Dunn). VCH, Weinheim, pp. 226–229.
- Briggs, M.M. and Capaldi, R.A. (1977) Near-neighbor relationships of the subunits of cytochrome c oxidase. *Biochemistry* **16**: 73–77.
- Burnette, W.N. (1981) “Western Blotting”: electrophoretic transfer of proteins from sodium dodecyl sulfate–polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112**: 195–203.
- Cavalcoli, J.D., VanBogelen, R.A., Andrews, P.C. and Moldover, B. (1997) Unique identification of proteins from small genome organisms: theoretical feasibility of high throughput proteome analysis. *Electrophoresis* **18**: 2703–2708.
- Cooper, C.A., Wilkins, M.R., Williams, K.L. and Packer, N.H. (1999) BOLD—a biological O-linked glycan database. *Electrophoresis* **20**: 3589–3598.
- Corbett, J.M., Dunn, M.J., Posch, A. and Gorg, A. (1994) Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: an interlaboratory comparison. *Electrophoresis* **15**: 1205–1211.
- Eckerskorn, C., Strupat, K., Karas, M., Hillenkamp, F. and Lottspeich, F. (1992) Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis* **13**: 664–665.
- Eckerskorn, C., Strupat, K., Schleuder, D., Hochstrasser, D., Sanchez, J.C., Lottspeich, F. and Hillenkamp, F. (1997) Analysis of proteins by direct-scanning infrared-MALDI mass spectrometry after 2-D PAGE separation and electroblotting. *Anal. Chem.* **69**: 2888–2892.
- Edman, P. and Begg, G. (1967) A protein sequenator. *Eur. J. Biochem.* **1**: 80–91.
- Figgeys, D., Ning, Y. and Aebersold, R. (1997) A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry. *Anal. Chem.* **69**: 3153–3160.
- Figgeys, D., Lock, C., Taylor, L. and Aebersold, R. (1998) Microfabricated device coupled with an

- electrospray ionization quadrupole time-of-flight mass spectrometer: protein identifications based on enhanced-resolution mass spectrometry and tandem mass spectrometry data. *Rapid Commun. Mass Spectrom.* **12**: 1435–1444.
- Fountoulakis, M., Langen, H., Evers, S., Gray, C. and Takacs, B.** (1997) Two-dimensional map of *Haemophilus influenzae* following protein enrichment by heparin chromatography. *Electrophoresis* **18**: 1193–1202.
- Fountoulakis, M., Takacs, B. and Langen, H.** (1998a) Two-dimensional map of basic proteins of *Haemophilus influenzae*. *Electrophoresis* **19**: 761–766.
- Fountoulakis, M., Juranville, J.F., Roder, D., Evers, S., Berndt, P. and Langen, H.** (1998b) Reference map of the low molecular mass proteins of *Haemophilus influenzae*. *Electrophoresis* **19**: 1819–1827.
- Garrels, J.I. and Franza, B.R.** (1989) The REF52 protein database. Methods of database construction and analysis using the QUEST system and characterizations of protein patterns from proliferating and quiescent REF52 cells. *J. Biol. Chem.* **264**: 5283–5298.
- Gras, R., Muller, M., Gasteiger, E., Gay, S., Binz, P.-A., Bienvenut, W., Hoogland, C., Sanchez, J.-C., Bairoch, A., Hochstrasser, D.F. and Appel, R.D.** (1999) Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring algorithm and an optimized peak detection. *Electrophoresis* **20**: 3535–3550.
- 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.
- Harrington, M.G., Lee, K.H., Yun, M., Zewert, T., Bailey, J.E. and Hood, L.E.** (1993) Mechanical precision in two-dimensional electrophoresis can improve spot positional reproducibility. *Appl. Theor. Electrophor.* **3**: 347–353.
- Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C. and Watanabe, C.** (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl Acad. Sci. USA* **90**: 5011–5015.
- Herbert, B.** (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**: 660–663.
- Hochstrasser, D.F.** (1989) Pre-cast gel systems for two-dimensional electrophoresis. US patent no. US4874490.
- Hochstrasser, D.F.** (1998) Two-dimensional electrophoresis device. US patent no. US5773645.
- Houthaeve, T., Gausepohl, H., Ashman, K., Nillson, T. and Mann, M.** (1997) Automated protein preparation techniques using a digest robot. *J. Protein Chem.* **16**: 343–348.
- Isobe, T., Uchida, K., Taoka, M., Shinkai, F., Manabe, T. and Okuyama, T.** (1991) Automated two-dimensional liquid chromatographic system for mapping proteins in highly complex mixtures. *J. Chromatogr.* **588**: 115–123.
- Kenrick, K.G. and Margolis, J.** (1970) Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique. *Anal. Biochem.* **33**: 204–207.
- Klose, J.** (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**: 231–243.
- Klose, J.** (1999) Genotypes and phenotypes. *Electrophoresis* **20**: 643–652.
- Kuster, B. and Mann, M.** (1998) Identifying proteins and post-translational modifications by mass spectrometry. *Curr. Opin. Struct. Biol.* **8**: 393–400.
- Langen, H., Gray, C., Roder, D., Juranville, J.F., Takacs, B. and Fountoulakis, M.** (1997) From genome to proteome: protein map of *Haemophilus influenzae*. *Electrophoresis* **18**: 1184–1192.
- Langen, H., Fountoulakis, M. and Berndt, P.** (1998) Modified in-gel protein digestion procedure allows high throughput MS analysis following 2-D gel gridding. Abstract from the 3rd Siena 2-D Electrophoresis Meeting 'From Genome to Proteome', 31 August–3 September, 1998.
- Latham, K.E., Garrels, J.I., Chang, C. and Solter, D.** (1991) Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages. *Development* **112**: 921–932.
- Lemkin, P.F., Myrick, J.M., Lakshmanan, Y., Shue, M.J., Patrick, J.L., Hornbeck, P.V., Thornwal, G.C. and Partin, A.W.** (1999) Exploratory data analysis groupware for qualitative and quantitative electrophoretic gel analysis over the Internet-WebGel. *Electrophoresis* **20**: 3492–3507.

- Link, A.J., Hays, L.G., Carmack, E.B. and Yates, J.R. III (1997) Identifying the major proteome components of *Haemophilus influenzae* type-strain NCTC 8143. *Electrophoresis* **18**: 1314–1334.
- Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R. III (1999) Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**: 676–682.
- Loo, J.A., Brown, J., Critchley, G., Mitchell, C., Andrews, P.C. and Ogorzalek Loo, R.R. (1999) High sensitivity mass spectrometric methods for obtaining intact molecular weights from gel-separated proteins. *Electrophoresis* **20**: 743–748.
- Lopez, M.F. and Patton, W.F. (1997) Reproducibility of polypeptide spot positions in two-dimensional gels run using carrier ampholytes in the isoelectric focusing dimension. *Electrophoresis* **18**: 338–343.
- Mann, M. and Wilm, M. (1994) Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**: 4390–4399.
- Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**: 10035–10038.
- Nokihara, K., Morita, N. and Kuriki, T. (1992) Applications of an automated apparatus for two-dimensional electrophoresis, Model TEP-1, for microsequence analysis of proteins. *Electrophoresis* **13**: 701–707.
- O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007–4021.
- O'Farrell, P.Z. and Goodman, H.M. (1976) Resolution of simian virus 40 proteins in whole cell extracts by two-dimensional electrophoresis: heterogeneity of the major capsid protein. *Cell* **9**: 289–298.
- O'Farrell, P.H. and Ivarie, R.D. (1979) The glucocorticoid domain of response: measurement of pleiotropic cellular responses by two-dimensional gel electrophoresis. *Monogr. Endocrinol.* **12**: 189–201.
- Ogorzalek Loo, R.R., Mitchell, C., Stevenson, T.I., Martin, S.A., Hines, W.M., Juhasz, P., Patterson, D.H., Peltier, J.M., Loo, J.A. and Andrews, P.C. (1997) Sensitivity and mass accuracy for proteins analyzed directly from polyacrylamide gels: implications for proteome mapping. *Electrophoresis* **18**: 382–390.
- Ogorzalek Loo, R.R., Loo, J.A. and Andrews, P.C. (1999) Obtaining molecular weights of proteins and their cleavage products by directly combining gel electrophoresis with mass spectrometry. *Methods. Mol. Biol.* **112**: 473–485.
- Oliver, S.G., van der Aart, Q.J., Agostoni-Carbone, M.L., Aigle, M., Alberghina, L., Alexandraki, D., Antoine, G., Anwar, R., Ballesta, J.P. and Benit, P. (1992) The complete DNA sequence of yeast chromosome III. *Nature* **357**: 38–46.
- Opitck, G.J., Ramirez, S.M., Jorgenson, J.W. and Moseley, M.A. III (1998) Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal. Biochem.* **258**: 349–361.
- Phillips, T.A., Bloch, P.L. and Neidhardt, F.C. (1980) Protein identifications on O'Farrell two-dimensional gels: locations of 55 additional *Escherichia coli* proteins. *Bacteriology* **144**: 1024–1033.
- Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A. and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**: 336–341.
- Qin, J., Herring, C.J. and Zhang, X. (1998) De novo peptide sequencing in an ion trap mass spectrometer with ¹⁸O labeling. *Rapid Commun. Mass Spectrom.* **12**: 209–216.
- Reeh, S., Pedersen, S. and Friesen, J.D. (1976) Biosynthetic regulation of individual proteins in relA⁺ and relA strains of *Escherichia coli* during amino acid starvation. *Mol. Gen. Genet.* **149**: 279–289.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**: 1030–1032.
- Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.

- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**: 5463–5467.
- Scheele, G.A. (1975) Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins. *J. Biol. Chem.* **250**: 5375–5385.
- Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Schleuder, D., Hillenkamp, F. and Strupat, K. (1999) IR-MALDI-mass analysis of electroblotted proteins directly from the membrane: comparison of different membranes, application to on-membrane digestion, and protein identification by database searching. *Anal. Chem.* **71**: 3238–3247.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci. USA* **93**: 14440–14445.
- Steinberg, R.A., O'Farrell, P.H., Friedrich, U. and Coffino, P. (1977) Mutations causing charge alterations in regulatory subunits of the cAMP-dependent protein kinase of cultured S49 lymphoma cells. *Cell* **10**: 381–391.
- Stoeckli, M., Farmer, T.B. and Caprioli, R.M. (1999) Automated mass spectrometry imaging with a matrix-assisted laser desorption ionization time-of-flight instrument. *J. Am. Soc. Mass Spectrom.* **10**: 67–71.
- Tonella, L., Walsh, B.J., Sanchez, J.C., Ou, K., Wilkins, M.R., Tyler, M., Frutiger, S., Gooley, A.A., Pescaru, I., Appel, R.D., Yan, J.X., Bairoch, A., Hoogland, C., Morch, F.S., Hughes, G.J., Williams, K.L. and Hochstrasser, D.F. (1998) '98 *Escherichia coli* SWISS-2DPAGE database update. *Electrophoresis* **19**: 1960–1971.
- Towbin, H., Staehlin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl Acad. Sci. USA* **76**: 4350–4354.
- Traini, M., Gooley, A.A., Ou, K., Wilkins, M.R., Tonella, L., Sanchez, J.-C., Hochstrasser, D.F. and Williams, K.L. (1998) Towards an automated approach for protein identification in proteome projects. *Electrophoresis* **19**: 1941–1949.
- VanBogelen, R.A., Hutton, M.E. and Neidhardt, F.C. (1990) Gene-protein database of *Escherichia coli* K-12: edition 3. *Electrophoresis* **11**: 1131–1166.
- Vestling, M.M. and Fenselau, C. (1994) Polyvinylidene difluoride (PVDF): an interface for gel electrophoresis and matrix-assisted laser desorption/ionization mass spectrometry. *Biochem. Soc. Trans.* **22**: 547–551.
- Weinstein, J.N., Myers, T.G., O'Connor, P.M., Friend, S.H., Fornace, A.J. Jr, Kohn, K.W., Fojo, T., Bates, S.E., Rubinstein, L.V., Anderson, N.L., Buolamwini, J.K., van Osdol, W.W., Monks, A.P., Scudiero, D.A., Sausville, E.A., Zaharevitz, D.W., Bunow, B., Viswanadhan, V.N., Johnson, G.S., Wittes, R.E. and Paull, K.D. (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**: 343–349.
- Wilkins, M.R. and Gooley, A.A. (1997) Protein identification in proteome projects. In: *Proteome Research: New Frontiers in Functional Genomics* (eds M.R. Wilkins, R.D. Appel, K.L. Williams and D.F. Hochstrasser). Springer, Berlin, pp. 35–64.
- Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphrey-Smith, I., Hochstrasser, D.F. and Williams, K.L. (1995) 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.
- Wilkins, M.R., Sanchez, J.-C., Williams, K.L. and Hochstrasser, D.F. (1996) Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* **17**: 830–838.
- Wilkins, M.R., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (eds) (1997a) *Proteome Research: New Frontiers in Functional Genomics*. Springer, Berlin.
- Wilkins, M.R., Lindskog, I., Gasteiger, E., Bairoch, A., Sanchez, J.C., Hochstrasser, D.F. and Appel, R.D. (1997b) Detailed peptide characterization using PEPTIDEMASS – a World-Wide-Web-accessible tool. *Electrophoresis* **18**: 403–408.
- Wilkins, M.R., Gasteiger, E., Sanchez, J.-C., Bairoch, A. and Hochstrasser, D.F. (1998a) The limitations of two-dimensional gel electrophoresis for proteome projects: the effects of protein

- copy number and hydrophobicity. *Electrophoresis* **19**: 1501–1505.
- Wilkins, M.R., Gasteiger, E., Tonella, L., Ou, K., Sanchez, J.-C., Tyler, M., Gooley, A.A., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (1998b) Protein identification with N- and C-terminal sequence tags in proteome projects. *J. Mol. Biol.* **278**: 599–608.
- Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.-C., Williams, K.L., Appel, R.D. and Hochstrasser, D.F. (1999a) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* **112**: 531–552.
- Wilkins, M.R., Gasteiger, E., Gooley, A.A., Herbert, B.R., Molloy, M.P., Binz, P.A., Ou, K., Sanchez, J.-C., Bairoch, A., Williams, K.L. and Hochstrasser, D.F. (1999b) High-throughput mass spectrometric discovery of protein post-translational modifications. *J. Mol. Biol.* **289**: 645–657.
- Wilm, M. and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**: 1–8.
- Wilm, M., Shevchenko, A., Houthaev, T., Breit, S., Schweigerer, L., Fotsis, T. and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**: 466–469.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. and Lockhart, D.J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**: 1359–1367.
- Yates, J.R. III (1998a) Mass spectrometry and the age of the proteome. *J. Mass. Spectrom.* **33**: 1–19.
- Yates, J.R. III (1998b) Database searching using mass spectrometry data. *Electrophoresis* **19**: 893–900.
- Yates, J.R. III., Eng, J.K. and McCormack, A.L. (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal. Chem.* **67**: 3202–3210.

第九章 瓶颈：自动化蛋白质组分析的完整途径

Mary F. Lopez

1. 引言

蛋白质组分析(Wasinger *et al.*, 1995)是一个正在迅速发展的领域, 已经成为致力于药物开发的制药公司和生物技术公司关注的焦点(Blackstock and Weir, 1999; Lopez 1998, 1999a), 与蛋白质组学相关的出版物数量以指数级数增长(图 9.1), 可以反映出其受关注的程度。最近的研究表明, 细胞中实际蛋白质水平与转录水平之间缺乏相关性(Anderson and Anderson, 1998; Gygi *et al.*, 1999), 这一结论刺激了大部分蛋白质组领域内的研究兴趣。这些研究使人们清楚了, 对细胞中蛋白质的分析是对其基因组分析的补充, 如果不结合蛋白质组学平台, 药物开发计划就不会成功。许多疾病都与蛋白质的翻译后修饰(如糖基化)机能障碍有关, 这是促使人们监测蛋白质谱的另一个原因, 这些生物学上的错误发生在代谢水平上, 在基因组水平上找不到线索。

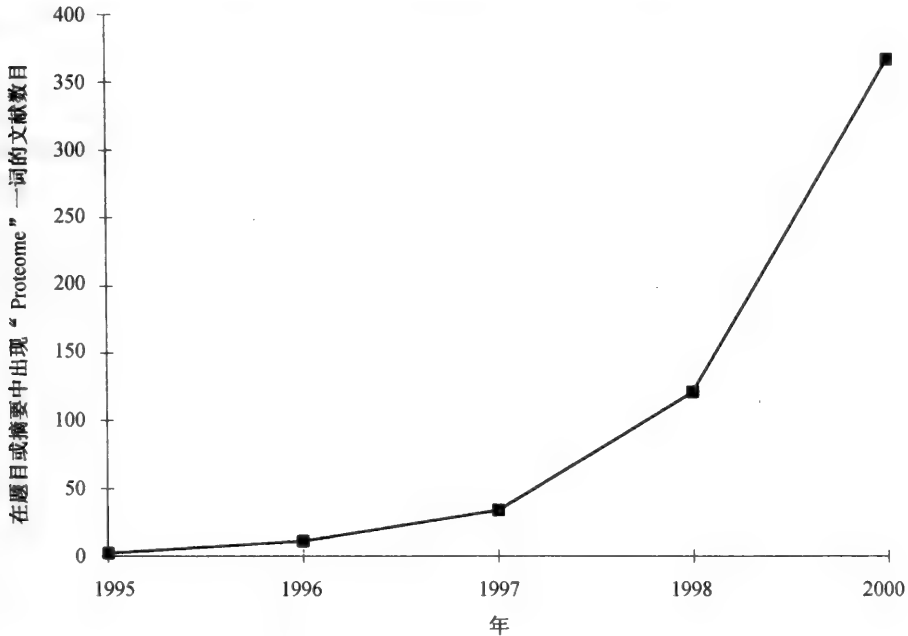


图 9.1 自 1995 年以来题目和摘要中出现“proteome”一词的文献数量增长情况

2. 二维电泳结合质谱：蛋白质组的典型范例？

全面的蛋白质组分析可以用各种方法来进行。但是, 目前人们更喜欢的手段包括了

以下这些技术的结合：二维凝胶电泳(2-DE)(O'Farrel 1975; Patton *et al.*, 1990), 随后进行谱图比较和消减分析, 鉴别可能的药物或诊断靶标, 将目标蛋白质从胶上切下, 水解, 提取肽片断用质谱测定肽质量指纹谱鉴定 (Henzel, 1993; James *et al.*, 1993; Mann *et al.*, 1993; Yate *et al.*, 1993), 或用串联质谱对单个肽段测序 (Hunt *et al.*, 1986; Jonsher *et al.*, 1997; Yates *et al.*, 1999) (第五章)。

不幸的是, 由于蛋白质固有的复杂性, 使蛋白质的分析不能像核酸那样易于高通量分析, 研究复杂蛋白质混合物的方法, 例如 2-DE, 很难或不可能成功地实现自动化 (Lopez *et al.*, 1999a; Nakihara *et al.*, 1992; Quardroni and James, 1999)。人们曾经满怀热情地寻找蛋白质组研究中双向电泳技术的替代技术, 但是都没有获得成功 (Lopez *et al.*, 1999a; Skipp, 1999)。最主要的问题是 2-DE 凝胶分离具有出色的并行 (parallel) 特性, 用大的 2-DE 凝胶配合灵敏的染色方法, 一块 2-DE 凝胶能很容易地同时分辨 1000 个, 多至 10 000 个不同的蛋白质 (Klose, 1999; Kobalz, 1995)。目前发展的技术中没有别的技术或技术的结合能够在相同的时间段内用同样量的起始物质传达出如此详细的内容。除了高分辨外, 2-DE 凝胶还能够定量, 尽管芯片技术有可能取代 2-DE 胶 (第十章), 但是仍然有许多问题有待于解决, 如非变性结合及磷酸化现象等 (Carr, 1999)。

3. 自动化概念

令人惊异的是, 尽管跑 2-DE 胶是一种劳动密集型的, 大部分需人工操作的过程, 但它却不是典型的高通量蛋白质组分析的瓶颈 (Table 1 in Lopez, 2000)。2-DE 胶的分析仍然有一定程度的主观性, 需要人参与编辑。对于获取得到的海量数据进行生物信息学处理是耗时的, 而且不是很有效。但是, 蛋白质组分析过程中许步骤能够被全部或半自动化。这样就能显著地提高效率和产率, 也随之提高了鉴定出靶标蛋白的可能性 (Lopez, 2000)。尽管某些核心蛋白质组实验室正在将自动化技术整合到其中一些分析步骤中去, 就在最近已经有了商业化的蛋白质组分析系统。许多公司正在积极发展蛋白质组分析技术平台和组件, 用于支持药物和靶标的开发, 如 BioRad, APB, PE Biosystem, GSI, Proteome Systems, Teca 等其他公司。

蛋白质组学理想的高通量全系统方法应该包括以下部分: 样品收集技术、高分辨电泳装置、自动化凝胶染色仪、半自动化图像分析软件、机器人模式的蛋白斑点处理过程、MALDI-TOF 和串联质谱仪, 并且具有整合的生物信息学软件, 可以链接到单个模块并进行畅通的样品处理, 追踪, 数据分析和数据归档。图 9.2 的例子说明了这样一种系统, 其部件都是目前商业可得的。在理想状态下, 每一个部件都应该与其他部件相配合, 以在开发路径中实现有效的数据和样品信息传递。理论上, 各个部件都应该能与一个相关数据库链接, 以整合过程中每一步的追踪信息 (例如用条形码)。以这种方式, 每一个样品获得的所有信息最终都会链接到最初胶上的蛋白斑点。

在下面的部分将大概描述一下一些系统部件各自的特性, 这些特性都是为满足高通量蛋白质组分析而设计的。

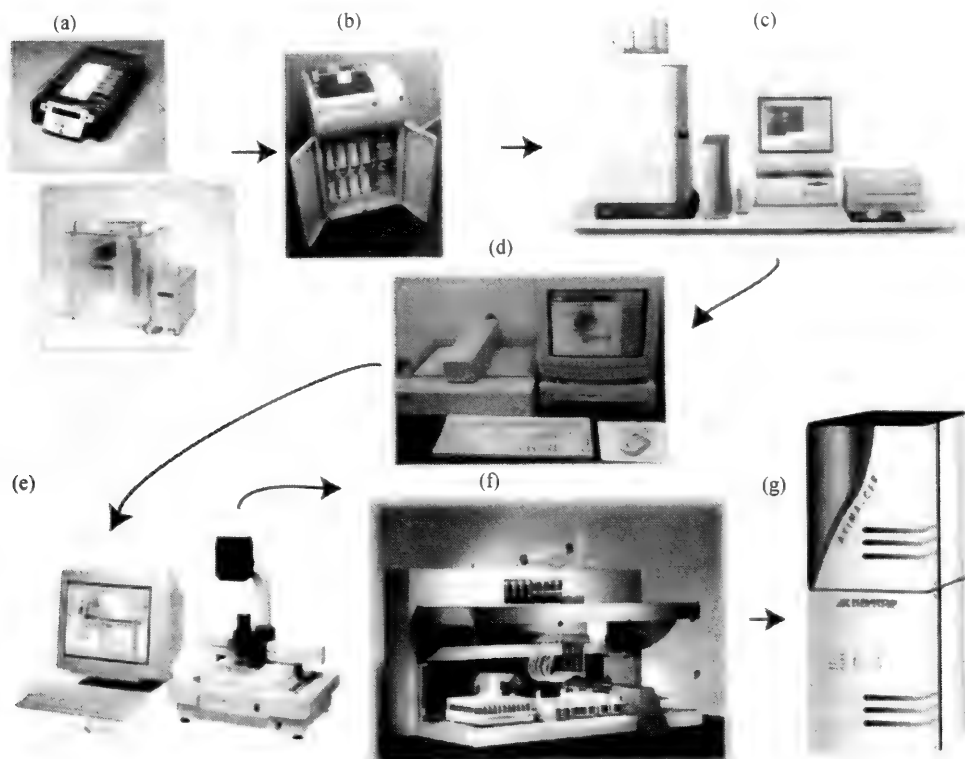


图 9.2 由目前商业化的各部件装配出的完整的蛋白质组研究系统

(a) IPGphor, Hofer DALT SDS-PAGE 凝胶电泳系统, APB; (b) Investigator Gel Processor, GSI; (c) LAS-1000 数字图像获取系统, Fujimed; (d) Investigator HT 2-D 图像分析软件系统, GSI; (e) Protean 2-D Spot Cutter, BioRad; (f) Genesis RSP 酶切和 MALDI 点靶机器人系统, Tecan; (g) Axima-CFR MALDI-TOF 质谱仪, Kratos。

4. 二维凝胶电泳的显色

4.1 二维凝胶

2-DE 凝胶电泳的质量直接影响到任何蛋白质组计划的结果。完美的二维电泳系统将会产生出大型、高重复性的凝胶。等电聚焦对离子和其他污染物极度灵敏，所以要使用高质量的试剂和预制胶将变异减小到最小程度 (Lopez *et al.*, 1991; Patton *et al.*, 1991a; Walsh *et al.*, 1998)。大型胶系统是一个典型的 2-DE 胶专用系统，并可被调整为高通量。每一个单元都应具有同时运行 10~12 块大型胶的能力，多系统模式的加入可适应更高通量的实验。迷你胶同样可成功用于 2-DE 分离，只要在 2-DE 运行之前对样品预先分离收集以减少样品的复杂性 (Harry *et al.*, 2000)。由于迷你胶电泳时操作更方便，所以比大型胶更适合于自动化操作。

第一维和第二维电泳都使用预制胶会提高分离的一致性。选择固定化或载体两性电解质作为一维化学物质使等电聚焦过程更具灵活性。载体两性电解质管胶更适用与疏水

性或膜蛋白的分离，而固定化梯度胶条更方便，常规使用更容易 (Lopez, 1999b; Patton *et al.*, 1999b)。所有这些参数结合起来使胶与胶之间的斑点位置相关性很好，不管是在 IEF (等电聚焦) 还是分子质量分离，变异都小于 1% (Fishmann, 1999; Lopez and Patton, 1997)。图 9.3 显示的是 6 块不同的血清 2-DE 胶在同一区域的放大照片，在疾病样品中可看到出现了在对照样品中没有的新斑点。除非胶与胶之间的模式 (pattern) 重现性非常好，否则是不可能进行这种比较的。

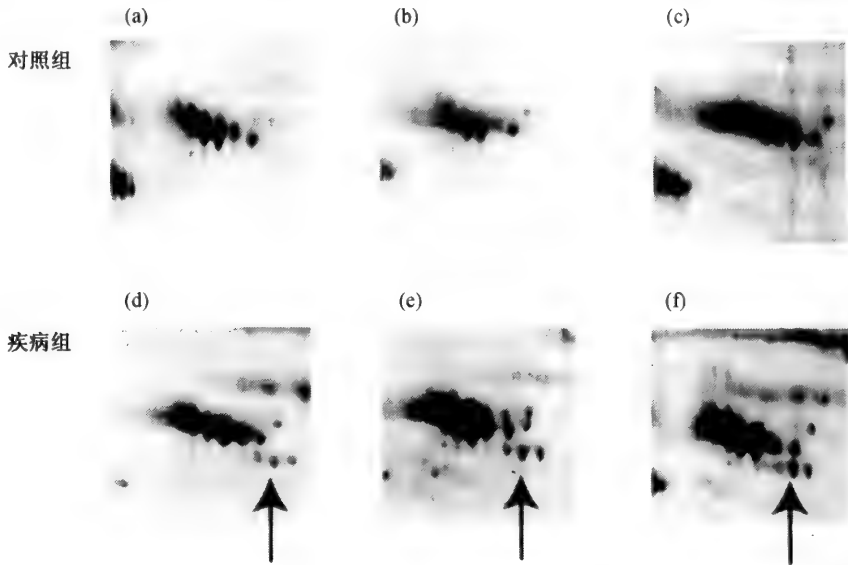


图 9.3 血清蛋白 2-DE 凝胶图像配比区域放大图

(a-c) 对照样品。(d-f) 疾病样品。凝胶分离和分析条件见 Lopez and Patton, 1997，不同之处在于使用了 BioImage 2-DE 分析软件 6.2 版 (GSI)。箭头所指处为在疾病样品中差异表达的蛋白质。

(蒙 S. Rasheed, UCLA 惠允)

4.2 富集和分部收集技术

尽管 2-DE 是目前分离蛋白质混合物的具最高分辨率的技术，它仍然不能分辨在任一细胞中所有表达的蛋白质 (>10 万个)，大多数研究者用富集或预分布收集技术来寻求解决这一问题的办法，如亲和色谱，溶液中分部收集或在 2-DE 前进行亚细胞分部收集 (Harry *et al.*, 2000; Lopez, 1999a; Lopez *et al.*, 2000a; Quadroni and James, 1999)。已经开发出的另一种分部收集手段是在一维等电聚焦分离中使用窄范围的固定化 pH 梯度胶 (1~2 个 pH 单位) (Harry *et al.*, 2000; APB, BioRad, Proteome Systems, GSI)。

4.3 显像技术

目前 2-DE 胶上蛋白质的显像技术只有有限的几种。因为有越来越多的实验室不能使用放射性同位素，所以最灵敏并可定量的发射活性标记法已经变得相当没有吸引力

了。放射性标记法还要求待研究的活细胞和组织在分析前与放射性同位素共培养，因此许多重要的样品明显不能采用这个方法，如人的组织和体液。最常用的染色方法是考马斯蓝染色和银染 (Merril, 1987; Patton *et al.*, 1999; Ramsby and Makowski, 1998; Wirth and Romano, 1995)。考马斯蓝的灵敏度很低，使其仅限于研究丰度最高的看家蛋白。尽管银染蛋白的胶上酶切肽质量指纹谱的肽序列覆盖率要相对低一些 (Gharahdaghi *et al.*, 1999)，但因其非常灵敏 (定量到 ng 级)，在一些例子中可定量 (Patton, 1995)，并能与质谱和测序技术相配合，因此银染通常是最好的选择。不幸的是，银染也有一些非常严重的缺点。通常银染定量的动力学范围 (dynamic range) 不

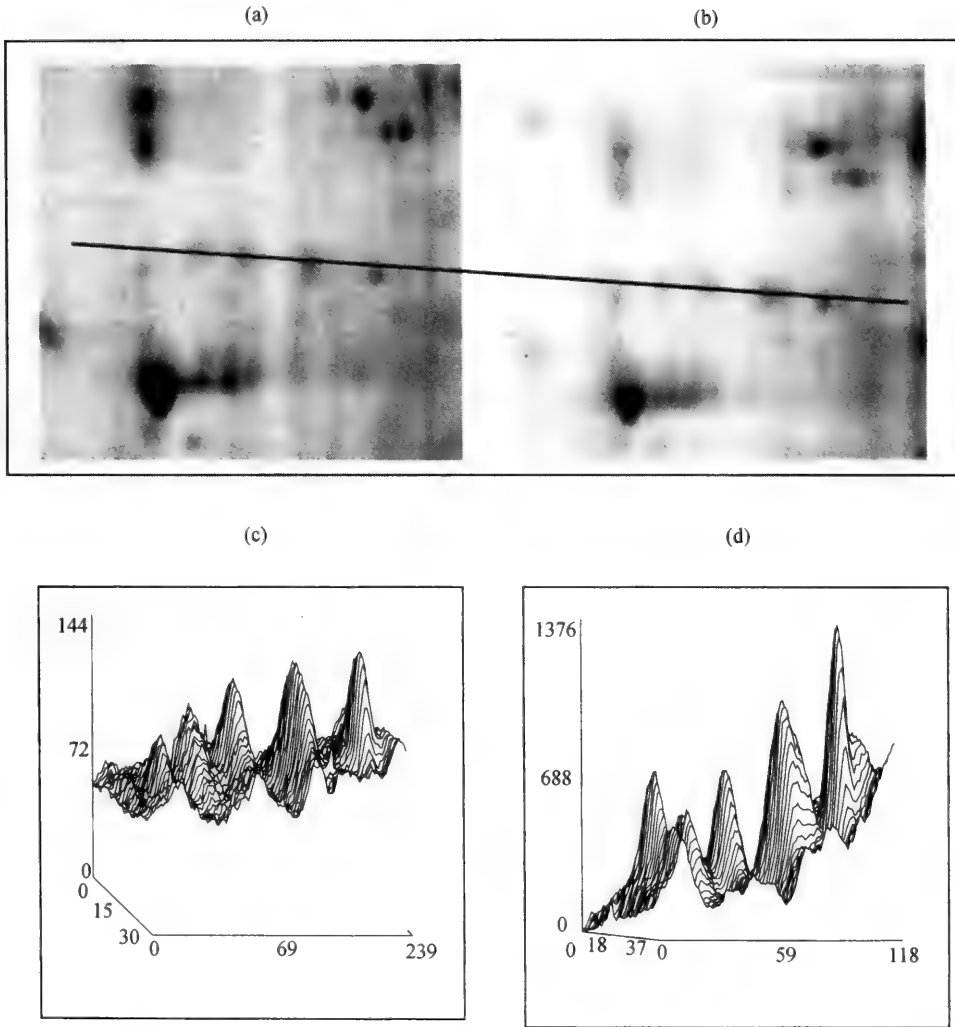


图 9.4 银染和 SYPRO Ruby 染色的 2-DE 凝胶相同区域内象素强度表面轮廓线
 (a) 2-DE 凝胶银染放大图；(b) 2-DE 凝胶 SYPRO Ruby 染色放大图，与 (a) 的区域相同。取直线所标处的蛋白斑点的表面轮廓线；(c) 图 (a) 中直线所标区域的表面轮廓线；(d) 图 (b) 中直线所标区域的表面轮廓线。凝胶分离和分析条件见 Lopez and patton, 1997，不同之处在于使用了 BioImage 2-DE 分析软件 6.2 版 (GSI)，及使用银染或 SYPRO Ruby Protein gel stain (Molecular probes) 染色。

大, 较高表达的丰度的蛋白质非常容易达到饱和。银染不是一个终点染色方法, 这一事实说明其定量是不准确的。另一种染色方法是荧光染色。市场上有许多非常好的荧光染色剂, 灵敏度与银染相当, 但还没有被广泛接受使用。最近, 一种新的荧光染色剂, SYPRO^R Ruby Protein Gel Stain (Molecular Probes) 被引入蛋白质组分析应用。SYPRO Ruby Protein Gel Stain 的灵敏度有所提高, 有更广泛的动力学(线性)范围, 用于质谱分析的胶上肽段的回收率比银染高 (Berggren *et al.*, 2000; Lopez *et al.*, 2000a; Patton 2000, 2000a)。图 9.4 显示的例子是成纤维细胞 2-DE 胶, 上样量相同, 分别用银染和 SYPRO Ruby 染色, 用图示的方式表现出在相同区域内象素强度的表面轮廓线, 表明了两种染色方式的线性差异, 显示出荧光染色用于定量比较的优势。

4.4 自动染色

不管选用什么染色方法, 使用自动化的染色系统将会减少人工劳动, 并提高实验的一致性。目前有少量商业化的 2-DE 胶自动染色系统可供选择, 如 APB, GSI。适用于高通量蛋白质组分析的系统应该能够按照不同使用者设定的步骤, 同时染多块电泳胶。

5. 图像分析

图像分析或许是高通量蛋白质组分析程序中作重要的元素了, 因为它决定了从 2-DE 胶的谱图数据库中提取的数据的质量, 并指导着疾病特异蛋白标志物搜寻的路径 (Wirth and Romano, 1995; 参考第六章)。在蛋白质组学开发路径中, 图像分析似乎越来越成为了重要的瓶颈步骤 (见 Lopez, 2000 中表 1)。第一步是用数字的形式获取凝胶图像。市场上有许多可用于获取凝胶图像的高分辨 CCD 照相系统 (Perkin Elmer, Fuji, APB, BioRad, Boehringer Mannheim)。CCD 照相系统的优点是方便, 灵活, 与包括荧光染色在内的许多染色步骤相适用 (Patton, 2000, 2000a)。激光扫描仪同样能成功用于从白光可见染色 (white light visible stains) 的凝胶中获取图像。这方面发表了许多综述 (Patton, 1995, 2000)。

在 UNIX 基础上的软件通常更适用, 因为 2-DE 胶的模式分析是计算性很强的工作。软件必须能够分析复杂的包含 1500~10 000 个蛋白质斑点的凝胶, 同时也能够处理包含几百凝胶数据的数据库。强大的数据库查询系统应该能够筛选出已有斑点的信息, 已匹配斑点的定量比、已整合的斑点强度、分子质量、等电点、面积以及用户限定的斑点或图像的特性。UNIX 平台的多线程 (multi-threading) 功能使其能够迅速处理复杂的、多水平的查询和非常大的凝胶数据库。在对每一个新加入数据库的凝胶谱图之间判定差异点的配比编辑处理中加入决定点 (decision point), 是可以提高分析效率和速度的一个策略 (图 9.5)。有许多商业化的高级 2-DE 凝胶图像分析软件包, 例如, Imagemaster, APB; Advanced 2-D Software, Phoretix; MELANIE II, Swiss Institute of Bioinformatics, GenBio; PDQUEST, BioRad; HT Investigator, GSI。图像分析软件必备的一个特性是与完整蛋白质组系统中其他部件整合, 精确切下并处理指定的蛋白质斑点。

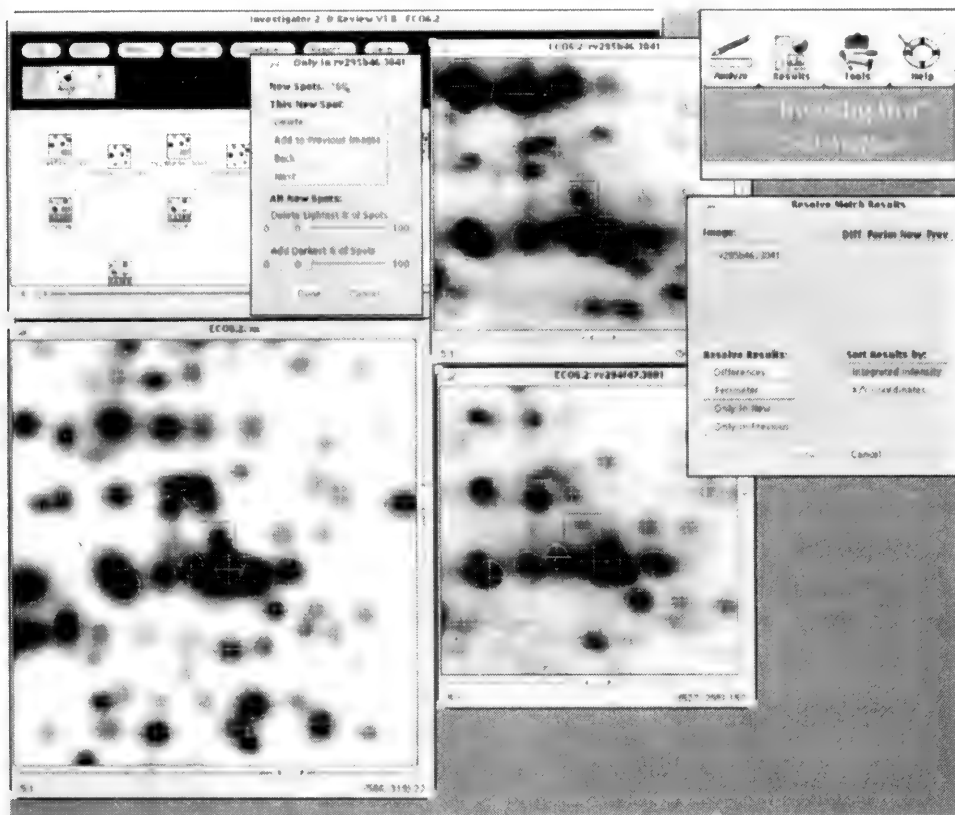


图 9.5 Investigator 2-D 分析软件 (GSI) 配比分辨率选项图示

此选项可以提高效率, 也即提高了 (凝胶) 配比编辑的速度, 这是 2-DE 凝胶图像分析的一个重要瓶颈。

6. 自动化机器人

6.1 斑点切取

一旦图像分析鉴别出了 2-DE 胶上有潜在研究价值的蛋白质斑点, 那他们必须被鉴定。通常, 斑点从凝胶上被切下, 并进行胶上蛋白质酶切 (如用胰蛋白酶)。很明显, 这一过程是非常费事的, 如果完全由手工处理大量样品, 则非常有可能从皮肤上污染角蛋白。现在已经开发出了多种能从 2-DE 胶上自动切点的机器人, 并且已经商品化 (BioRad; APB; GSI; Proteome Systems)。有些公司正在开发多项合一 (all-in-one) 的平台, 将斑点切取, 酶切和 MALDI-TOF 点靶整合为一体 (Proteome Systems)。一个完全自动化的斑点切取系统要包括一个 CCD 照相机, 用于从 2-DE 胶上检测和切取斑点。切点机器人应该与 2-DE 分析软件整合, 以指导其从 2-DE 胶上切取目标斑点或全部斑点。机器人应该有最大程度的灵活性, 能够适用于银染, 考马斯蓝染色或荧光染色的胶。目前最高通量的系统大约每小时可以切下 200 个斑点。应该使用没有反应活性的塑料或不锈钢的采集头来切取斑点, 并将切下的斑点放入微升板内传送到酶切机器人处 (见下文), 图 9.6 显示的是自动化斑点切取机器人切取的 2-DE 凝胶分离的血浆蛋白斑点。

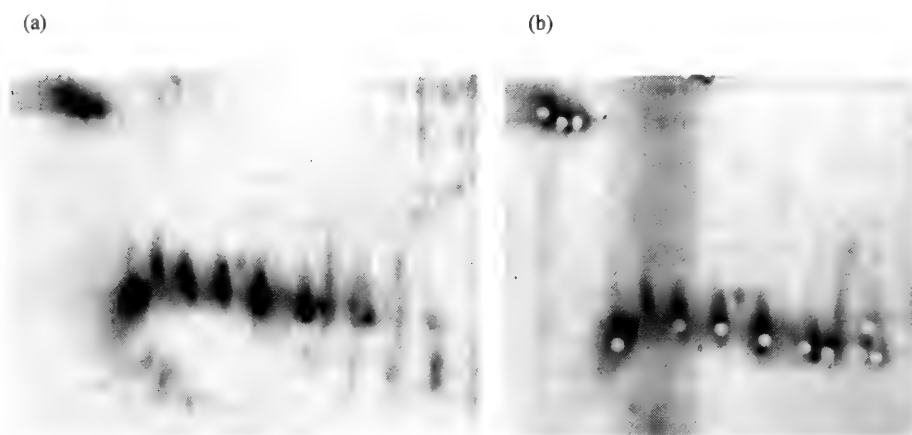


图 9.6 用 Flexys 斑点切取机器人 (GSI) 处理前后的银染凝胶放大图
 (a) 处理前的凝胶; (b) 斑点切下之后的凝胶。Flexys 机器人从 2-D 分析软件产生的报告中选择要切斑点的位置。斑点切取使用 2mm 的塑料收集头。凝胶分离和分析条件见 Lopez and Patton, 1997, 不同之处在于使用了 BioImage 2-DE 分析软件 6.2 版 (GSI)。

6.2 斑点酶切和肽提取

蛋白斑点从 2-DE 凝胶上切下来后, 用蛋白酶水解 (通常用胰蛋白酶), 产生的肽混合物从凝胶中提取出来。同样, 对这一过程也进行了自动化处理的努力。一个完全自动化的酶切机器人应该自动进行从酶切到肽提取过程中的所有步骤, 使用高通量的胶上蛋白胰蛋白酶酶切, 最多需要 8h。所有样品的清洗和提取步骤都应该在微升板中自动进行, 以易于处理并减少样品的交叉污染。并且全过程结束后, 提取的肽段应该很容易传送到下一步的 MALDI-TOF 自动点靶机器人处, 或传送到串联质谱仪的自动进样器上。如果能用一个 PC 网络远程控制多个酶切机器人, 或许能够实现超高通量。用户应该能够按照自己的要求制定酶切步骤, 并在开始操作前将其下载到每一台仪器。

6.3 肽脱盐、纯化, 在 MALDI 靶上点样

肽段从凝胶中提取出来后, 必须除去盐和去污剂这样的污染物, 在进行质谱分析前, 肽还要被浓缩。如果用 MALDI-TOF-MS 分析肽段, 则需与基质混合 (通常用 CCA) 后在 MALDI 靶上点样。有多种商业化的机器人可以完成这一过程中的部分或全部任务 (PE Biosystems; GSI; BioRad; Tecan; Proteome Systems), 其中一些机器人平台可以使用装有疏水性树脂填料的 ZipTip (Millipore) 来纯化肽段 (Courchesne and Patterson, 1997; Jensen *et al.*, 1999)。很重要的一点是机器人要将 $0.5 \sim 1 \mu\text{l}$ 或更少体积的样品排列精确地点在 MALDI 靶上, 而且此机器人平台还要能与大多数制造商生产的 MALDI-TOF 样品靶相适用。能通过 PC 网络远程控制 MALDI-TOF 机器人同样是一个令人满意的特性。图 9.7 显示的是用上述处理达到的从 2-DE 胶上切下的斑点的肽混合物质谱图, 通过检索 SWISS PROT 数据库, 此斑点被鉴定为 50S 核糖体蛋白 L9。

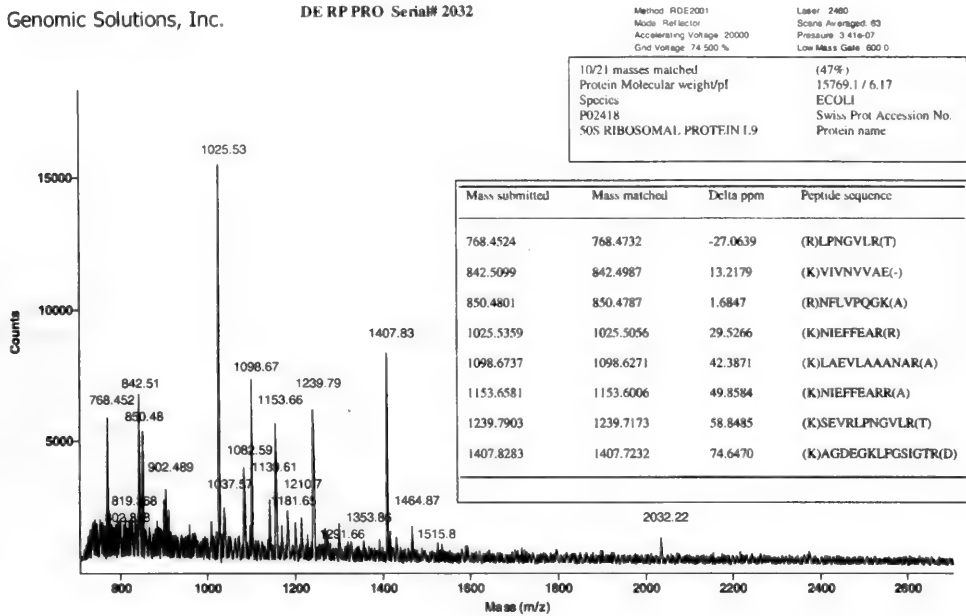


图 9.7 从 2-DE 胶上切下斑点的胶上酶切肽质谱图

斑点用 ProGest 及 proMS MALDI 机器人 (GSI) 酶切处理。蛋白质大约 1pmol, 考马斯蓝染色。质谱图由 Voyager DE-RP Pro MALDI-TOF 质谱仪 (PE BioSystems) 分析 (见 Parker *et al.*, 1998)。用 Protein prospector 程序在万维网上检索 SWISS PROT 数据库 (<http://prospector.ucsf.edu>), 插图显示的是质量精确度及每一个匹配肽段的序列, 也显示了蛋白鉴定的可能结果。蛋白凝胶分离和分析条件见 Lopez and Patton, 1999, 不同之处在于使用了 BioImage 2-D 分析软件 6.2 版 (GSI)。

7. 质谱的接口策略

目前有两种将质谱整合入蛋白质组研究程序的方法。为了鉴定那些在各种基因组和蛋白质组数据库中不存在的蛋白质, 必须要确定其氨基酸序列。以往是用蛋白质序列仪通过 Edman 化学降解方法来实现的。这一步骤耗时长, 且需要相当大量的蛋白质 (Tempst *et al.*, 1999)。最近多数研究者已经采用四极杆或电喷雾串联质谱仪 (Micro-mass, SxiEX/Perkin Elmer, Finnigan, Bruker) 来确定单个肽段的氨基酸序列, 并用这些短的序列标签在公共或私营 EST 库和序列数据库中检索 (Yate *et al.*, 1993, 1999)。这些测序技术很准确, 但是仍然需要相对大量的样品 (picomole), 并且速度慢, 限制了他的高通量。为了高通量鉴定 2-DE 胶上的蛋白质, 用 MALDI-TOF 质谱仪测肽质量指纹谱是更适合的技术 (Henzel *et al.*, 1993; James *et al.*, 1993; Pappin *et al.*, 1993)。尽管 MALDI-TOF 不能定量, 但它是一个非常灵敏, 具高分辨力的技术, 灵敏度达到 fmol~amol。同时也非常适合于自动化和高通量操作。另外, 因为许多物种的基因组测序正以惊人的速度在进行, 所以在万维网上的公共数据库中有丰富的序列数据可用于质量剖析 (mass profiling) 检索 (Lopez, 1998, 1999a)。这些技术在有效的, 自动化系统控制下, 可以每天迅速鉴定几十到几百个蛋白质斑点。在过去几年中, 许多厂家已经制造出了精密的台式 MALDI-TOF 仪器, 如 PE Biosystems, Bruker, CIPHERGEN, Micro-

mass, Kratos。

8. 机器人模块, 图像分析软件和信息数据库之间的链接

用上面几段所述的一整套系统处理大量的样品将会产生大量的数据, 这对样品的追踪是一个挑战。因此, 自动化蛋白质组分析程序中最重要的是对软件的整合, 在分析过程中各种仪器使用的软件及样品信息和数据管理所用软件的整合。开发满足这一需求的生物信息学系统是高通量蛋白质组研究计划首先要考虑的问题。一个生物信息学解决方案应该建立在相关数据库的基础上, 以进行各系统部件间样品追踪和信息数据的交换。样品追踪应该使用条形码, 淘汰人工追踪, 因为人工追踪常常会引入误差。中央数据库是核心部分, 应该能使其他系统部件, 如图像分析软件、斑点切取机器人、酶切器、质谱仪等存储和读取信息。数据库还应该能使用户在查询和报告时方便地读取数据。此外, 样品制备信息和 2-DE 凝胶运行参数都应该有记录。最后, 从质谱仪得到的数据信息应该再链接回到凝胶图像, 这样, 只要鼠标一点击, 就能得到每一个斑点的制备、处理及鉴定数据。网络浏览器应该是这种类型的生物信息学软件包理想的使用界面。目前有许多公司正在开发用于蛋白质组分析的生物信息学软件包。

9. 结论

要解释复杂的生物学系统需要理解其表达的蛋白质及其基因组结构图。蛋白质组分析试图获取任一时刻机体或组织的蛋白表达谱。对蛋白表达谱的比较能够提供一些与鉴定潜在的药物或诊断试剂相关的特殊信息。对蛋白质组分析各种步骤进行自动化及整合对提高效率并进而促进靶标鉴定非常有必要, 高通量蛋白质组分析策略中瓶颈问题的解决正在使药物和诊断标志物开发的途径更便利。未来蛋白质组学的挑战将是基因组学和小分子代谢信息的整合, 这种整合产生的最终结果将对细胞代谢过程有更深入和广泛的理解。显而易见, 蛋白质组学的出现已经鼓励了许多生物学家的思想跳出了基因组学的框框。

(王京兰译 钱小红校)

参考文献

- Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*, **19**: 1853-1861.
- Berggren, K, Steinberg, T.H., Lauber, W.M. et al. (1999) A luminescent ruthenium complex for ultrasensitive detection of proteins immobilized on membrane supports. *Anal. Biochem.* **276**: 129-143.
- Blackstock, W. and Weir, M. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotech.* **17**: 121-127.
- Carr, F. (1999) Proteins on chips. Presentation to IBC Conf., London, Feb. 17-19.
- Courchesne, P.L. and Patterson, S.D. (1997) Manual microcolumn chromatography for sample clean-up prior to mass spectrometry. *Biotechniques* **22**: 244-250.
- Fischmann, J. (1999) Advantages of immobilized pH gradients. In: *2-D Proteome Analysis Protocols* (ed. A.J. Link). Humana Press, Totowa, NJ, pp. 173-174.
- Gharahdaghi, F., Weinberg, C.R., Meagher, D.A., Imal, B.S. and Mische, S.M. (1999) Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. *Electrophoresis* **20**: 601-605.

- Gygi, S.P., Rochon, Y., Franza, R. and Aebersold, R. (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**: 1720–1730.
- Harry, J.L., Wilkins, M.R., Herbert, B.R., Packer, N.H., Gooley, A.A. and Williams, K.L. (2000) Proteomics: Capacity versus utility. *Electrophoresis* **21**: 1071–1081.
- Henzel, W.J., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C. and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **90**: 5011–5015.
- Hunt, D.F., Yates, J.R. II, Shabanowitz, J., Winston, S. and Hauer, C.R. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA* **83**: 6233–6238.
- James, P., Quadroni, M., Carafoli, E. and Gonnè, G. (1993) Protein identification by mass profile finger printing. *Biochem. Biophys., Res. Commun.* **195**: 58–64.
- Jensen, O.N., Wilm, M., Shevchenko, A. and Mann, M. (1999) Sample preparation methods for mass spectrometric peptide mapping directly from 2-DE gels. In: *2-D Proteome Analysis Protocols* (ed. A.J. Link). Humana Press, Totowa, NJ, pp. 513–530.
- Jonscher, K.R., Yates, I. and John R. (1997) The quadrupole ion trap mass spectrometer—a small solution to a big challenge. *Anal. Biochem.* **244**: 1–15.
- Klose, J. (1999) Genotypes and phenotypes. *Electrophoresis* **20**: 643–652.
- Klose, J. and Kobalz, U. (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**: 1034–1059.
- Lopez, M.F. (1998) Proteomic databases: roadmaps for drug discovery. *Am. Biotechnol. Lab.* July.
- Lopez, M.F. (1999a) Advantages of carrier ampholyte IEF. In: *2-D Proteome Analysis Protocols* (ed. A.J. Link). Humana Press, Totowa, NJ, pp. 109–110.
- Lopez, M.F. (1999b) Proteome analysis I. Gene products are where the biological action is. *J. Chromatogr. B* **722**: 191–202.
- Lopez, M.F., (2000) Better approaches to finding the needle in a haystack: Optimizing Proteome analysis through automation. *Electrophoresis* **21**: 1082–1093.
- Lopez, M.F. and Patton, W.F. (1997) Reproducibility of polypeptide spot positions in two-dimensional gels run using carrier ampholytes in the isoelectric focusing dimension. *Electrophoresis* **18**: 338–343.
- Lopez, M.F., Patton, W.F., Utterback, B.L., Chung-Welch, N., Barry, P., Skea, W.M. and Cambria, R.P. (1991) Effect of various detergents on protein migration in the second dimension of two-dimensional gels. *Analytical Biochem.* **199**: 35–44.
- Lopez, M.F., Kristal, B.S., Chernokalskaya, E., Lazarev, A., Shestopalov, A.I., Bogdanova, A. and Robinson, M. (2000a) High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation. *Electrophoresis* (in press).
- Lopez, M.F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M. and Patton, W.F. (2000b) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* (in press).
- Lopez, M.F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M. and Patton, W.F. (2000c) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* (in press).
- Mann, M., Hojrup, P. and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass. Spectrom.* **22**: 338–345.
- Merril, C. (1987) Detection of proteins separated by electrophoresis. In: *Advances in Electrophoresis, vol. I* (eds. A. Chrambach, M. Dunn and B. Radola). VCH, Federal Republic of Germany, pp. 111–139.
- Nokihara, K., Morita, N. and Kuriki, T. (1992) Applications of an automated apparatus for two-dimensional electrophoresis. *Electrophoresis* **13**: 701–707.
- O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007–4021.
- Orntoft, T.F. and Vestergaard, E.M. (1999) Clinical aspects of altered glycosylation of glycoproteins in cancer. *Electrophoresis* **20**: 362–371.
- Packer, N.H. and Harrison, M.J. (1998) Glycobiology and proteomics: is mass spectrometry the

- holy grail? *Electrophoresis* **19**: 1872–1882.
- Parker, K.C., Garrels, J.I., Hines, W., Butler, E.M., McKee, A.H.Z., Patterson, D. and Martin, S.** (1998) Identification of yeast proteins from two-dimensional gels: Working out spot cross-contamination. *Electrophoresis* **19**: 1920–1922.
- Pappin, D.J.C., Hojrup, P. and Bleasby, A.J.** (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**: 327–332.
- Patton, W.F.** (1995) Biologist's perspective on analytical imaging systems as applied to protein gel electrophoresis. *J. Chromatogr. A.* **698**: 55–87.
- Patton, W.F.** (2000) A thousand points of light: the application of fluorescence detection technologies to two dimensional gel electrophoresis and proteomics. *Biotechniques* **28**: 944–957.
- Patton, W.F.** (2000a) Making blind robots 'see'; the synergy between fluorescent dyes and imaging devices in automated proteomics. *Electrophoresis* **21**: 1123–1144.
- Patton, W.F., Pluskal, M.G., Skea, W.M., Buecker, J.L., Lopez, M.F., Zimmermann, R., Belanger, L.M. and Hatch, P.D.** (1990) Development of a dedicated two-dimensional gel electrophoresis system that provides optimal pattern reproducibility and polypeptide resolution. *Biotechniques* **8**: 518–524.
- Patton, W.F., Lopez, M.F., Barry, P. and Skea, W.M.** (1991a) A mechanically strong matrix for protein electrophoresis with enhanced silver staining properties. *Biotechniques* **12**: 580–585.
- Patton, W.F., Chung-Welch, N., Lopez, M.F., Cambria, R.P., Utterback, B.L. and Skea, W.M.** (1991b) Tris-tricine and tris-borate buffer systems provide better estimates of human mesothelial cell intermediate filament protein molecular weights than the standard tris-glycine system. *Anal. Biochem.* **197**: 24–29.
- Patton, W.F., Lim, M.J. and Shepro, D.** (1999) Protein detection: using reversible metal chelate stains. In: *2-D Proteome Analysis Protocols* (ed. A.J. Link). Humana Press, Totowa, NJ, pp. 331–339.
- Quadroni, M. and James, P.** (1999) Proteomics and automation. *Electrophoresis* **20**: 664–667.
- Ramsby, M.L. and Makowski, G.S.** (1999) Differential detergent fractionation of eukaryotic cells: analysis by two-dimensional gel electrophoresis. In: *2-D Proteome Analysis Protocols* (ed. A.J. Link). Humana Press, Totowa, NJ, pp. 53–66.
- Scheler, C.** (1998) Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis* **19**: 918–927.
- Skipp, P.** (1999) Proteome analysis without gels: current status of capillary isoelectric focusing-electrospray ionization (CIEF-ESI)-MS. *Biochemical Soc. Trans.* **27**: 68.
- Tempst, P., Link, A.J., Riviere, L.R., Fleming, M. and Elicone, C.** (1990) Internal sequence analysis of proteins separated on polyacrylamide gels at the sub-microgram level: improved methods, applications and gene cloning strategies. *Electrophoresis* **11**: 537–553.
- Walsh, B.J., Molloy, M.P. and Williams, K.L.** (1998) The Australian proteome analysis facility (APAF): assembling large scale proteomics through integration and automation. *Electrophoresis* **19**: 1883–1890.
- Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L. and Humphery-Smith, I.** (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**: 1090–1094.
- Wirth, P. and Romano, A.** (1995) Staining methods in gel electrophoresis, including the use of multiple detection methods. *J. Chromatogr. A.* **698**: 123–143.
- Yates, J.R. III, Carmack, E., Hays, L., Link, A.J. and Eng, J.K.** (1999) Automated protein identification using microcolumn liquid chromatography-tandem mass spectrometry. In: *2-D Proteome Analysis Protocols* (ed. A.J. Link). Humana Press, Totowa, NJ, pp. 553–558.
- Yates, J.R. III, Speicher, S., Griffin, P.R. and Hunkapiller, T.** (1993) Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* **214**: 397–408.

第十章 蛋白质表达分析新方法

Rosalind E. Jenkins and Stephen R. Pennington

1. 引言

基因组测序方法已取得了巨大进步,而且这些新方法对全世界研究者完全开放(<http://www.tigr.org/tdb/> 和 http://www.ornl.gov/TechResources/Human_Genome/home.html)。一个完整的基因组序列对人们了解这种生物体的生命过程提供了一个很好的机会。基因组测序方法还衍生出一些新的方法和新技术,如对表达 mRNA 的分析方法(Chee *et al.*, 1996; Eisen *et al.*, 1998; Gerhold *et al.*, 1999)和用双杂交方法来综合分析蛋白质间的相互作用。运用这些方法人们获得了海量的不同生物体的信息,包括酵母(Dujon, 1996; Goffeau *et al.*, 1996; Uetz *et al.*, 2000)、线虫(Chalfie, 1998; *C. elegans* consortium, 1998; Plasterk, 1999)和人类(Chee *et al.*, 1996)。最近,由于微阵列技术的发展,更进一步加快了人们获得基因序列和基因差异表达分析的步伐(Bowtell, 1999)。这些方法的综合运用使人们得到了许多功能基因组学的信息,并且正在把这些信息转化为对生命过程的研究(Alizadeh *et al.*, 2000; Alon *et al.*, 1999; DeRisi *et al.*, 1996; Perou *et al.*, 1999; Pollack *et al.*, 1999; Wang *et al.*, 1999)。然而,随着这些技术的广泛应用,它们逐渐变为一种常规技术。现在人们已逐渐认识到上述这些技术将成为对基因或 mRNA 产物——蛋白质直接分析的一种补充手段。最终,对基因组表达产物蛋白质量的分析、细胞定位、空间结构分析将决定一个细胞或生物体活性(Dove, 1999; Parekh, 1999; Pennington *et al.*, 1997; Wilkins *et al.*, 1995)。然而,蛋白质从本质上比其相对应的 mRNA 或 DNA 序列复杂得多。因此,在基因组水平分析蛋白质方法落后于飞速发展的核酸的分析方法,这不足为怪。

2. 功能蛋白质组研究范围

功能蛋白质组学包含有几个重要却又相互补充的方面。截止到现在,对蛋白质组研究的注意力主要集中于试图分析整个蛋白质表达变化特别是在一些特殊条件下蛋白质组表达变化,如在疾病状态下,单个基因敲除后、药物治疗或在改变细胞外界环境的状况下。单个基因表达的 mRNA 水平常与其表达的蛋白质水平不相关。因此,这样的分析是合理的。然而,在一个细胞、器官和组织中,单个蛋白质表达水平同样也不反映这种蛋白质生物活性,这种现象也是普遍存在的。蛋白质活性是一个复杂的综合过程,它涉及到蛋白质翻译后修饰、在组织、细胞和亚细胞中的定位以及它与别的蛋白质间的相互作用。因此,分析蛋白质活性必须考虑每个方面并把所得各部分的数据综合起来。当然,对每个方面现有各种方法的讨论在一章中是不可能完成的。因此在本章我们讨论以下几个方面:①介绍一下现行通用蛋白质表达分析法的局限性;②介绍一些正在发展的新方法;③讨论一些分析蛋白质表达谱的新方法。在讨论之前,我们有必要再重述一

下,那就是对蛋白质定位(第十二章)、蛋白结构分析(Brünger, 1997; Cusack *et al.*, 1998; Gerstein and Hegyi, 1998; Robson, 1999; Sali and Kuriyan, 1999; Swindells *et al.*, 1998; Wery and Schevitz, 1997)、蛋白质功能的预测(Bork *et al.*, 1998; Marcotte *et al.*, 1991a, b; Pellegrini *et al.*, 1999)和蛋白质相互作用图谱方面一些高通量的技术已取得了很大的进展,如二维杂交法(Utez *et al.*, 2000)和一些计算机分析法(Enright *et al.*, 1999)。另外,还包括一些在基因组范围内对基因功能分析的方法(RossMacdonald *et al.*, 1999)。

3. 蛋白质组分析: 基于 2-DE 的战略

目前,大部分蛋白质组分析计划起初都是用二维电泳分离蛋白质复合物(第三章)。这包括基于等电点对蛋白质混合物的一向分离,紧接着依赖于蛋白质分子质量进行垂直分离。在一块二维凝胶上,将有近 10 000 个蛋白质点,利用几块胶,可以得到有关蛋白质不同表达的数据。值得指出的是,一个单一蛋白质可以在二维胶上呈现几个点,这种多样性至少代表该蛋白质几种不同翻译后修饰方式。一段时间以来,由于不能够鉴定二维凝胶电泳分离的蛋白质点,这大大限制了二维凝胶电泳的潜在应用。尽管如此,还是有少数研究者,如著名的 Celis 和他的同事,在二维凝胶电泳分离的蛋白质点的鉴定、二维凝胶电泳图及其数据库的制作方面取得了巨大的进展(Celis *et al.*, 1994, 1995),但是这些只是个别的例子。依赖于质谱的蛋白质鉴定方法(包括从二维凝胶电泳分离的蛋白质点和利用其他方法所分离的蛋白质点(图 10.1 和第五章))的发展,使得以二维凝胶电泳作为分析蛋白质表达谱的工具得到了很大的发展(Celis *et al.*, 1998; Fitcher *et al.*, 1999; Li *et al.*, 1999; Shevchenko *et al.*, 1996)。

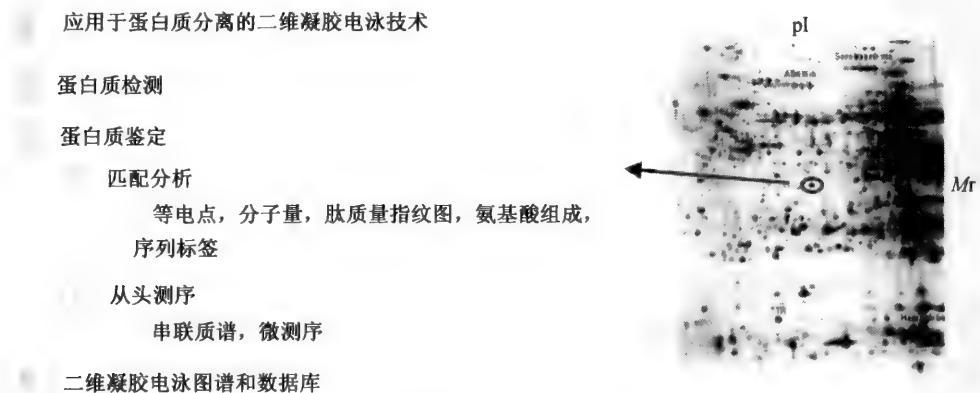


图 10.1 用二维凝胶电泳进行蛋白质表达分析和鉴定

二维凝胶电泳作为分析蛋白质谱的技术平台能否持续发展下去,这个问题一直在争论之中。二维凝胶电泳技术本身所具有的缺点和局限性,使得这种讨论更为热烈。其中最为明显的缺点如图 10.2 所示。即使在最先进的实验室,对胶的控制包括单一蛋白质点在空间的重复性和所染色蛋白质点的量的重复性,是很难达到的。对于在同一个实验室和不同实验室之间的凝胶所分离的点的一致性,更是有许多争论。但是,具有一定的

经验且仔细操作，能够取得很好的重复性，这一点是不容怀疑的 (Corbett *et al.*, 1994)。然而，还是有许多人在评估获得好的二维凝胶电泳重复性的艰难性及对所显示单个蛋白质点在统计学上有明显差异所需二维凝胶的数目。此外，目前通用的二维凝胶电泳方法在其二维凝胶电泳图上所显示大多是那些高丰度和中等丰度的蛋白质 (第四章)。这儿列出了二维凝胶电泳所有的局限性，但有的局限性是可以被克服的，如在进行二维凝胶电泳以前富集样品可以检测到一些低丰度蛋白质。可以预测，人们可以发现一些新荧光染料来检测蛋白质。最感兴趣的是用不同的荧光染料分别标记两种样品，混合后，进行二维凝胶电泳 (Unlü *et al.*, 1997)。根据单一蛋白质所发出的不同的荧光量可以测得这个蛋白质在两种样品中的相对丰度。虽然这种方法具有许多内在的局限性，但是它有一个很重要的优点就是它克服了“不同凝胶间重复性”问题。

应用于蛋白质组分析的二维凝胶电泳技术平台

应用于蛋白质组分析的二维凝胶电泳技术平台

重复性

灵敏度

显示范围

数目

对照

“困难”蛋白质 - 难于在二维凝胶电泳图谱显示的蛋白质

数据分析

需要熟练的技术，花费较多的时间和精力

难于自动化

图 10.2 作为蛋白质表达分析二维凝胶电泳技术局限性

尽管有许多文献报道一块胶上能够显示 20 000~50 000 个点，但是一块胶最多可检测到 10 000 个蛋白质点 (Klose, 1999)。这是由于许多蛋白质在二维凝胶电泳图上可以有好几个点。二维凝胶电泳图上实际含有蛋白质数目少于所显示的“蛋白质点”的数目。值得注意的是，每块胶不能够显示样品中所有的蛋白质点。有的蛋白质，包括膜蛋白，在“正常条件下”不能显示在凝胶电泳图谱上。同样可以采取多种办法克服这个缺点：可以跑多块窄 pH 范围和窄分子质量范围的凝胶；可以改变变性剂的浓度提高膜蛋白质的溶解度；可以采取分步抽提的方法来制备样品。但是这些方法都需要较多的样品量。对一些样品，包括人类活组织，样品量不足以用来做多次电泳。因此，二维凝胶电泳更适合分析一些简单的生物体，可以在二维凝胶电泳合适的范围内用它分析总“蛋白质组”中的主要部分。

二维凝胶电泳得到的图像含有复杂的 3-D 信息，对它的分析是对研究人员一种巨大的挑战 (第六章)。目前对二维凝胶电泳图像的有效分析费时较多，因此需要进一步改进。总的来说，对一个运用二维凝胶电泳方法的研究人员，二维凝胶电泳过程和对凝胶分析过程不仅需要技术，而且工作密度较大，费时较多。这个过程的自动化能克服这些缺点，因此将成为这个领域研究者所需解决的首要问题。目前，仅仅只在跑胶和图像分

析过程实现了部分自动化, 这还远远不够。

完全自动化包括对二维凝胶电泳图像分析、蛋白质点的切割和用质谱鉴定蛋白质的过程, 其正在发展过程中。几家商业公司正在研制或对已有的技术进行改进, 包括凝胶分析软件包、用机械臂来切割蛋白质点和一些可用于自动化操作的质谱仪(包括能够获得肽指纹图谱(PMF)的MALDI-TOF和可进行复杂分析的电喷雾(ESI)串联质谱仪(第五章和第八章))。现对所选择二维凝胶电泳点质谱分析数据收集能力(Traini *et al.*, 1998)和对蛋白质库序列自动化搜索的能力都已得到提高(Jensen *et al.*, 1997; Nicola., 1998; 第八章)。

4. 除二维凝胶电泳以外的蛋白质表达分析方法

蛋白质组分析对我们了解生物体具有很大的潜在影响, 这是加快发展一些可用来分析研究蛋白质表达功能新技术的动力。这些新技术大体上可分为两大类: 一种是依赖于分离的, 像二维凝胶电泳一样; 另一种是不依赖于分离来检测蛋白质表达。

4.1 依赖于分离的方法

在这里描写的方法都需要在分析一个蛋白质混合物以前, 对它进行不同程度的分离和纯化。二维凝胶电泳分离蛋白质在于它运用两种相互独立的物理化学特性—— pI 和 M_r , 来进行垂直分离。在叙述一些已经存在的别的利用垂直分离方法之前, 先讨论两种依赖于二维凝胶电泳分离的方法。利用红外MALDI-TOF-MS来替代常用的紫外MS, 其已被应用于对样品的高能量分析。二维凝胶电泳分离的蛋白质被转移到PVDF膜上后, 泡在基质溶液中, 然后用红外MALDI-TOF-MS进行扫描(Eckerskorn *et al.*, 1997a)。另一个方法与上述方法极其相似, 它是用“分子扫描”的办法直接分析二维凝胶电泳分离的蛋白质(Bienvenut *et al.*, 1999; Bin *et al.*, 1999)。在这个方法中, 整个二维凝胶电泳胶被电转印到PVDF膜上, 在电转过程中, 用蛋白酶来消化蛋白, 然后用MS来扫描PVDF膜。它不需要对蛋白质染色, 分析凝胶电泳图像和对单个蛋白点进行酶解反应。但是它要求质谱仪具有较高的灵敏度和分析能力。这个方法尚处于孕育阶段, 但是如果商家对其产业化(Buter, 2000), 将能使其得到飞速发展。

一步分离法 由于电喷雾质谱能够分析蛋白质混合物, 使得这种方法具有可行性, 并引起研究人员很大兴趣。Lamond和Mann(1997)提出了这种想法, 即对蛋白质混合物初步分离后, 对蛋白质复合物进行快速分析, 并已取得了很大的成功(Neubauer *et al.*, 1998; Rappsilber *et al.*, 2000; Rigaut *et al.*, 1999; Rout *et al.*, 2000; Wigge *et al.*, 1998; Zachariae *et al.*, 1998)。在这种方法中, 利用1-D纯化得到一些蛋白质复合体, 如剪接体, 然后对相应的带(可能含有几种蛋白)进行酶解, 用质谱分析。对这类蛋白复合物进行快速和详细的鉴定。已证明这种方法是可行的。并且用这种方法已获得了一些新的以前不被认识的成分。对一些少数蛋白质组成的复合物, 如由免疫亲和方法得到的复合物, 也可以用这种方法进行分析。Ciphergen公司提出了一种新蛋白分析法, 运用特定的抗体获取目标蛋白, 可以达到fmol级(10^{-15} 级), 然后运用表面增强激光解吸电离质谱来检测它们的分子质量(www.ciphergen.com)。该公司还推出了一系列蛋白质“芯片”, 其实就是在MALDI靶上涂了一层色谱树脂(如阳离子交换树脂), 使其可

以用于蛋白质表达谱实验。该方法主要过程包括先使复杂的蛋白质混合物与靶结合，然后在不同的条件下清洗，最后用质谱获得剩余蛋白分子质量，由此来表征蛋白。虽然这种方法有很多局限性，但是已可以得到感兴趣的“芯片”板。

垂直分离法 为了避免使用二维凝胶电泳，现在正在检测一些新的方法，综合运用两种或多种已经使用的分离方法，使其能够达到二维凝胶电泳的分离能力与通量。由于液相色谱与毛细管电泳易实现自动化，引起了人们特殊的兴趣，现在正在实验其潜在的“高通量”能力。

Link 和他的同事用二维色谱的方法，对酵母核糖体中蛋白质成分进行了研究 (Link *et al.*, 1999)。先用蛋白酶消化纯化的酵母 80S 核糖体，用阳离子交换 (SCX) 柱进行纯化，采用 12 步逐步洗涤液，再进入反相色谱柱中进行纯化，最后对抽提液用串联质谱进行分析。在一个简单的循环中，所预测的 78 个核糖体蛋白中的 75 个被鉴定出来了。令人感兴趣的是，用同样的方法对整个酵母细胞抽提物进行分析时，鉴定出了 78 个蛋白中的 71 个。运用一种新的毛细管柱，对用胰蛋白酶酶解的蛋白质，这种方法的灵敏度可达 10fmol 级水平。

一段时间以来，高效液相色谱 (HPLC) 被用于作为分离蛋白质和肽的一种方法。最近，由于使用毛细管柱，这种方法在微量化和自动化方面得到了改进。人们一直试图把反相毛细管 HPLC 和成分收集结合起来，然后直接加到 MALDI-TOF 靶上、Edman 测序膜或硝酸纤维素膜上 (Grimm and Grasser, 1998)。超高压毛细管液相色谱可以使 HPLC 的分析速度提高 2~3 倍，同时在流量相同的情况下，可以提高被鉴定分析物的数目 (MacNair *et al.*, 1999)。

有报道对上面所提到的红外-MALDI-MS 扫描方法进行了改进 (Eckerskorn *et al.*, 1997b)。用 HPLC 分离蛋白质复合物，按照流出时间，把洗液直接收集到 PVDF 膜上成带，PVDF 膜放在基质溶液中，然后取出进行红外-MALDI-MS 扫描。有报道说用冰作为基质，能够得到较好的图谱。这为运用扫描红外-MALDI-MS，对冰冻组织进行原位蛋白分析开辟了一条道路 (Berkenkamp *et al.*, 1996)。

Opitck 等人 (1998) 一直在寻求对二维-HPLC 改进并且已经研制出一套自动化程度很高的二维-HPLC 方法，其分离原理与二维凝胶电泳相似。在他们的方法中，先根据蛋白质大小用排阻色谱分离蛋白，馏分自动进入反相液相色谱，这样就得到了二维色谱图。分析物被转移到 96 孔板上，然后自动加到 MALDI-MS 靶上进行鉴定。一个典型的分离过程可以得到 576 个组分，同时可以在线分析 MS 结果。因此，在这种分析过程中，人的干预是很少的。利用这种方法在转基因的大肠杆菌裂解液中成功鉴定出两种基因产物，非受体酪氨酸激酶和 β -内酰胺酶 src 同源结构域 (Opitck *et al.*, 1998)。这种方法有许多优点，样品分离可以重复多次、加样过程、分离过程和馏分收集都可以自动化，洗涤液中的蛋白可以直接用质谱分析。目前，这种方法的分离能力还不高，还有一段很长的路要走。但是，通过不断的改进和电喷雾串联质谱分辨能力的提高，最终这些馏分中的蛋白可能被逐步鉴定。

对于临床蛋白样品的检查，毛细管电泳由于它具有快速、高度重复性、灵敏度高等优点，被证明是一种非常有效的方法。例如，它已被应用于诊断腺苷基琥珀酸裂解酶 (adenylosuccinase) 缺乏症、肾病综合征 (nephrotic syndrome) 和癌症恶病质 (Choud-

hary *et al.*, 1999)。运用 CE 和 MALDI-TOF 质谱在癌症恶病质病人尿样中, 鉴定出一种糖蛋白, 它在与恶病质相关肌肉蛋白的酶解中起重要的作用。MS 的灵敏性使得它能与 CE 联用。

还有一些研究人员运用二维液相电泳结合质谱对人类脑脊液中低拷贝蛋白进行了研究 (Davidsson *et al.*, 1999)。这种方法需要大量的蛋白质混合物加到一维上以便于在流出液中有足够的低拷贝蛋白, 可用作 MALDI 分析。对于血清和脑脊液这类的样品, 由于它们中含有大量的血清蛋白和免疫球蛋白, 在许多分析中, 可能屏蔽一些别的重要蛋白的出现。所以, 检测低拷贝蛋白时, 更要加大样品量。

由于“芯片”的使用, 毛细管电泳在微小化方面正在取得重大的进展 (Becker *et al.*, 1998; Kopp *et al.*, 1997; Yao *et al.*, 1999)。如 Yao 等人 (1999) 通过运用一个长 4~5cm, 深为 40nm 通道形成的“芯片”, 缩短了毛细管电泳过程, 对流出物检测它们在 214nm 光吸收值或激光激发荧光量 (对于能衍生出荧光的样品)。并预言微阵列和芯片形式的二维凝胶电泳不久将变为可能 (Yao *et al.*, 1999)。

在将这种微型分离与基于质谱蛋白质鉴定连接在一起时, 一个主要的障碍是怎样才能将分离的蛋白充分转移进入质谱。低浓度少量蛋白质由于被吸收到表面上而易于丢失。如果人工过多干预又易被污染。由于这些原因人们设计了一种装置, 通过它可以直接将微量制备液相系统与 ESI-MS 直接联系起来。以减少液体向表面吸附过程中损失及人对少量样品的控制 (Figcys and Aebersold, 1999)。同样还设计出一种装置, 它可以在样品进入 ESI-MS 以前, 通过 C18 反相柱头, 对样品进行富集, 这种装置可把灵敏度提高到 fmol 级水平 (Figcys and Aebersold, 1999)。

4.2 不依赖分离的方法: “蛋白质芯片”

尽管在微型化和运用多种方法综合起来鉴定蛋白方面取得了进展, 但是上面描述的依靠分离的技术相对来说仍需要较多的蛋白质样品。并在一系列过程中需要多步操作。运用“分子识别”是一种很好的方法, 它不仅能用于这类分析, 并且能减少分析时所需样品的量。其原理是用分子间特定的识别机制, 来分离、检测和确定目标分子。

利用分子识别最成功的例子是“DNA 芯片”技术的使用, 它利用固定在靶上的碱基与荧光标记核酸探针上碱基间相互作用来快速分析成千上万的不同序列 (Bowtell, 1999; Cheung *et al.*, 1999)。cDNA 或合成的寡核苷酸序列紧密排列在玻璃芯片上。一个 3.6cm^2 芯片可以容纳 10 000 个基因 (Gerhold *et al.*, 1999)。运用不同的颜色来标记 cDNA 探针, 同时与相应的芯片杂交, 可以比较健康组织与疾病组织间的 RNA 表达谱, 现在这种方法已经应用于诸多方面的研究 (Alizadeh *et al.*, 2000; Alon *et al.*, 1999; DeRisi *et al.*, 1996; Perou *et al.*, 1999; Pollack *et al.*, 1999; Wang *et al.*, 1999)。

依赖于分子识别的“蛋白质芯片”类似于“DNA 芯片”, 将在蛋白质表达谱方面产生重大的影响。所以, 人们在发展“蛋白质芯片”方面有很大的兴趣。“蛋白质芯片”技术有两个关键因素: ①空间上固定能辨别单一蛋白成分分子; ②具有检测混合物中单个蛋白与它们相对应的识别分子间相互作用的方法。像人们所预料的那样, 有几个实验室已经开始发展类似这样的方法。然而, 直接效仿“DNA 芯片”的方法有许多困难。例如, 用不同的荧光, 像标记 cDNA 那样来标记蛋白质有好多内在的困难。首先, 荧光

染料与蛋白质的结合过程没有像核酸分子结合的那样易于控制，不同的荧光染料与不同样品中同一蛋白结合效率的微小差别将导致定量比较的失败。用单一位点标记特异性位点虽然能克服上述缺点，但是将导致灵敏度的降低。而且，与核酸结合可以预测杂交后荧光染料的变化，可是却不能预测与蛋白质结合后荧光染料的变化。这样，固定的分子识别目标蛋白的能力可能受到结合的荧光集团的影响，而这种影响很难预测或很难控制，这意味着固定在空间特定位置用以识别单一蛋白的蛋白质芯片序列需要新的方法来产生信号分子，通过它来检测与目标分子蛋白间的相互作用。下面我们列出一些正在发展的替代方法，这些方法有的已经得到了很好的发展，但是在蛋白质表达分析应用方面还不能与“蛋白质芯片”相提并论（例如，同时分析成千上万种蛋白），仅可以用于一些中等的蛋白质表达分析，其他的方法处于原始阶段，仅可潜在应用于“蛋白质芯片”方法。

蛋白质识别分子 最明显的能够有效作为识别蛋白的分子是抗体，这种分子识别方式已被广泛应用于许多方面。已使用的抗体有单抗和多抗。如果用抗体作“蛋白质芯片”必须遵循以下几个重要原则：首先，必须有支持获得和选择单个蛋白的方法，这个蛋白甚至是经翻译后修饰的。这个抗体必须是高度特异的，这意味着它们能够在一个复杂混合物中识别和区分单个蛋白。其次，混合物是单个蛋白与它们的抗体结合必须在相似的条件，后者在一定程度上可以通过仔细控制抗体产生过程中筛选步骤（screening step）来达到。很明显，现行的这套产生抗体的方法不适用于基因组范围“蛋白质芯片”技术的发展，然而，它适用于中等程度蛋白质表达分析。例如，诊断蛋白质芯片能够用于检测几十种蛋白表达和翻译后修饰，这是对现行诊断技术的一个重大改进。

嗜菌体显示抗体技术的产生能够在较大范围内获得蛋白质识别分子，促进了“蛋白质芯片”技术的发展。目前利用这项技术产生抗体的应用参看第十二章。

现在有多种方法应用于蛋白质（包括抗体）的排列方面，有文章综述了这方面的进展（Blawas and Reichert, 1998）。最近，Rowe 等人提出了一种具有应用潜力的排列方法，他们将抗体以足够的间隔点在显微带上，以便能够检测多价抗原（multi-bound antigens）。简单地说，他们在一个经过硅烷化和酯化的玻璃表面上，涂了一层 NeutrAvidin，这就形成了一个流动微型组件，把标记了生物素的俘获性抗体点在 6 个垂直的道上，然后按照正确的角度把样品加到抗体上，这样可以同时检测 6 个样品中的分析物。对俘获的分析物可以用“夹层”分析法检测出来，即用荧光标记抗体检测器和基于 CCD 的读数计（Rowe *et al.*, 1999）。

另一种方法用疏水胶作为模子，把蛋白质分子膜放到镀金表面上或把蛋白质固定排列在硅化的玻璃面上的聚丙烯酰胺垫上（Guschin *et al.*, 1997）。后者使用 3% 酰胺溶液，其中含有甘油、亚甲基蓝、N, N', N'-四甲基乙二胺，光照使其聚合，然后在其表面覆盖一层不透明的铬膜后，在其表面形成格子。用镀金的光学显微针给每一个胶垫上加上 1nl 戊醛和鼠免疫球蛋白（IgG1）、兔子免疫球蛋白或牛血清蛋白溶液，使基质活化。作者已经用荧光标记鼠 IgG1 抗体特定鉴定出在细小的玻璃条上排列的鼠 IgG1。

抗体并不是惟一地用在“蛋白质芯片”上的蛋白质识别分子。任何分子集团只要它具有相应的亲合力并能特定辨认出单个蛋白质分子，都可以固定起来，形成“蛋白质芯

片”。运用抗体作为蛋白质识别分子有许多缺点。最大的缺点是芯片保存期不能太长，且在分析的蛋白质混合样中，任何具有蛋白酶活性的物质可能“降解”芯片。运用组合化学获得许多化合物组成的大型文库为找到更好的蛋白质识别分子提供了源泉，但是切记还需要得到一个有效的扫描蛋白质识别分子与单个蛋白分子相结合的方法。

“Aptamers”是单链寡核苷酸，它对于一些具有三维构象的分子（例如蛋白质）具有高的亲和力。长度上将近有几百个核苷酸组成的 Aptamers 的文库固定在合适的表面上，对靶蛋白进行扫描，可得到一些有关靶蛋白的构象、氢键的位置和其他一些有助于构建靶蛋白三维结构的数据 (Brody *et al.*, 1999; Ellington and Szostak, 1990; Hermann and Patel, 2000; Roming *et al.*, 1999)，这个方法具有很大的应用潜力。最近运用分子印迹聚合物 (molecularly imprinted polymers) 得到了一些人造受体或抗原结合位点。

分子印迹聚合物 (molecularly imprinted polymer) 分子印迹 (molecularly imprinting) 通过模仿聚合模上模板分子的形状在其表面上形成了人造识别位点，这样就形成了分子印迹聚合物。推测任何通过形状起作用的生化位点，如抗体-抗原相互作用、底物-酶的结合和受体-配体结合，都可以被分子印迹聚合物来模仿 (Ansell and Mosbach, 1998; Hauptand Mosbach, 1998; Ye *et al.*, 1998)

现用于构建分子印迹聚合物有两种通用方法。第一种方法首先混合模板分子形成聚合物单体，然后硬化基质并用溶剂把模板洗掉，这使得聚合物表面凹凸不平与模板分子表面相吻合，能够用来俘获样品中靶分子。目前，可用于制作聚合物的单体与交联剂有多种，不仅能够符合聚合物所要求的特异性和亲合性，而且还能有一定的柔韧性，耐高温和高压，与酸、碱还有一些活性溶剂不反应 (Owens *et al.*, 1999)。例如 Takeuchi 和同事通过保持除草剂含量不变，改变异丁基酸和三氟甲基乙酸的比例，合成出 triazine 除草剂组合文库。加入乙腈，释放出除草剂，就合成出了分子印迹聚合物 (Takeuchi *et al.*, 1999)。通过比较分子嵌合聚合物孵育前后自由 herbicide 浓度来评价形成的分子印迹聚合物的好坏。同样的，别的研究人员在 1, 3-二甲基黄嘌呤 (theophylline) 存在的情况下，使异丁基酸聚合，得到了另外一种分子印迹聚合物，将该聚合物碾成粉末，用甲醇-乙酸抽提除去 1, 3-二甲基黄嘌呤，用很小的分子印迹聚合物颗粒来装制成柱子，这种柱子能特异结合 1, 3-二甲基黄嘌呤。有人建议用磁性的氧化铁来修饰这些分子印迹聚合物，那么在一定的磁场强度下可把这些分子印迹聚合物从溶液中钩出来 (Ansell and Mosbach, 1998)。

另一种制备 MIP 的方法是用云母作为吸收底物，云母是一种自动展平的物质 (Shi *et al.*, 1999)。把蛋白质涂在云母的表面上，再在其上覆盖一层双糖，它能够模仿蛋白质表面构象。随着辉光放电等离子体沉积，六氟丙烯多聚膜沉积在糖表面上，然后除去云母和蛋白质层，留下了表面镀糖的该蛋白印迹物 (Shi *et al.*, 1999)。用此方法合成出了抗生蛋白链菌素 (streptavidin) 的 MIP，在牛血清蛋白与生物素形成混合物中，它主要与生物素结合。

很明显，如果这些 MIP 要发展成为真正适合蛋白质芯片认知表面，还需要做大量的工作。有三个关键问题。第一，包含有识别单一蛋白独立的区域的 MIP 表面是否完整，第二是没有一种好的方法来检测蛋白质与 MIP 结合的方法不明确。最后一个也是最重要的一个问题是模仿一个完整蛋白而产生的 MIP 直接对应于这个蛋白中的所有的

构象表位。然而人们正逐渐认识到, 由于蛋白质的空间结构是它们自己“设计”的模型。因此, MIP 是否能有效地与单一蛋白特异的结合, 这个问题并不清楚。

检测方法 快速、平行和灵敏地检测蛋白质与“蛋白质芯片”上认知分子间相互作用的方法, 也是对人们一个重大的挑战。现在发展的许多方法非常有效并且也不需要复杂的装置, 但是并不适用于“蛋白质芯片”。如下面所示, 例如, 如上所述的用第二种荧光标记认知分子检测俘获的蛋白的方法并不适用。对基于荧光的光学纤维传感器, 尽管有其潜在的缺点, 但人们还是很感兴趣的。这当中俘获抗体直接被光学纤维探针所围绕 (Tempelman *et al.*, 1996)。Tempelman 和他的同事发展了一个体系, 能同时检测 4 个样品中单一分析物, 对葡萄球菌肠毒素, 其最佳检测范围是 5~200 ng/ml, 别的研究者提出了不同于上述的另外一种依赖荧光的检测方法, 抗体被基质所覆盖, 并且和荧光标记的抗原充分结合。测试样品中未被标记的抗原可以取代荧光标记的抗原, 然后检测释放的荧光量。美国联邦药物管理局用这种方法定量分析尿中可卡因 (cocaine) 的代谢物 (Yu *et al.*, 1996), 而海军研究院用它来分析爆炸物 (Marang *et al.*, 1998)。这些研究所报告的最小检测量在 fmol~pmol (10^{-15} ~ 10^{-12}) 水平。尽管目前这些方法并不能称作“蛋白质芯片”, 人们可以预测一套体系, 将设计好的蛋白质识别分子排列在一个支持面上, 它们都结合等量荧光, 当其结合单一蛋白时释放所结合的荧光, 单一位点上所剩的荧光量与样品混合物相应的蛋白量成反比。

生物传感器, 包括等离子体共振 (SPR) 成像生物传感器, 能够对蛋白质与别的蛋白或配体间相互作用进行实时检测 (Gizeli and Lowe, 1996), 这些装置有一层镀有金属的玻璃面, 通常为金, 再在其上覆盖一层基质自组装层 (SAM), 如羧甲基水凝胶 (Laricchia Robbil *et al.*, 1996), 羧甲基葡聚糖 (Vikinge *et al.*, 1998; Watts *et al.*, 1994) 或硫代烷烃混合物 (Dubrovsky *et al.*, 1999; Lahiri *et al.*, 1999)。对基质的化学性质要求是能把蛋白质和配体固定在传感器上, 当目标蛋白结合上后, 基质的反射指数发生变化, 同时用一些二进制检测器反映这些变化 (Lahiri *et al.*, 1999)。很明显, 如果把这些装置表面积扩大化, 意味着它能容有更多的蛋白质识别位点, 而且将其微小化, 这些都意味着生物传感器具有更大的应用潜力。

石英结晶微平衡 (quartz crystal microbalance) 装置也是一种检测抗体—抗原相互作用的微妙方法。在这个方法中, 抗体通过氨基与乙二醇高聚膜与石英晶体相结合 (Nakanishi *et al.*, 1996), 然后, 置于样品溶液中, 催化, 使抗体与目标抗原相结合, 导致晶体表面质量微小变化, 这些变化可以通过检测两根金电极间振动频率的变化来检测出来, 这个方法的灵敏度使得它可以用于检测单个蛋白分子与单个抗体分子间相互作用。

别的能用于检测单个分子间相互作用的方法包括扩增力生物传感器 (force-amplified biological sensor), 由 Colton 和他的同事发明 (Baselt *et al.*, 1997)。它是由原子力显微镜改造过来的, 运用微小化磁珠与微型压阻悬臂结合起来检测抗体抗原相互作用。最后, 通过结合在微电循环中的生物传感器来检测, 这种生物传感器是用分子离子通道作为检测装置 (Cornell *et al.*, 1997)。在这个方法中, 以合成脂膜为离子通道, 在其下层一个金电极与一个短杆菌肽分子相接触, 在其上层有流动的短杆菌肽分子, 流动的短杆菌肽分子与抗体结合。如果抗原与抗体结合则改变了该通道的电阻。这种方法有许多优点, 包括可以把检测信号进行放大; 这样单一抗原—抗体反应导致了每秒钟上万的离子流的移动 (Turner, 1997)。

5. 结语

尽管二维凝胶电泳有它本身的局限性,但是到目前为止,就分辨率而言,任何一种别的方法都不能同它相提并论,即能同时分析几千个蛋白质(Abbott, 1999),这一点没有人怀疑。而且,已经有了一些半自动和高通量的方法来鉴定2-DE分离的蛋白质并且这种方法还在一直提高。因此,这意味着2-DE技术和它的一些支持技术在很长一段时间内将得到广泛应用。但是这个技术的局限性使得人们有理由去发展一些新的技术来研究蛋白质表达谱。我们综述了这方面的一些进展,并且认识到未来的高通量研究蛋白质表达技术必须有多方面的方法,一些新的技术还需进一步发展。目前,研究机构和商业企业对快速增长的“蛋白质组学”表现出了浓厚的兴趣,这些将加快一些新技术发展的进程。

(杨何义译 钱小红校)

参考文献

- Abbott, A. (1999) A post-genomic challenge: learning to read patterns of protein synthesis. *Nature* **402**: 715-720.
- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.G., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L.M., Marti, G.E., Moore, T., Hudson, J., Lu, L.S., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O. and Staudt, L.M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**: 503-511.
- Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D. and Levine, A.J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **96**: 6745-6750.
- Ansell, R.J. and Mosbach, K. (1998) Magnetic molecularly imprinted polymer beads for drug radioligand binding assay. *Analyst* **123**: 1611-1616.
- Baselt, D.R., Lee, G.U., Hanse, K.M., Chrisey, L.A. and Colton, R.J. (1997) A high-sensitivity micromachined biosensor. *Proc. IEEE* **85**: 672-680.
- Becker, H., Lowack, K. and Manz, A. (1998) Planar quartz chips with submicron channels for two-dimensional capillary electrophoresis applications. *J. Microchem. Microengng* **8**: 24-28.
- Berkenkamp, S., Karas, M. and Hillenkamp, F. (1996) Ice as a matrix for IR-matrix-assisted laser desorption/ionisation: mass spectra from a protein single crystal. *Proc. Natl Acad. Sci. USA* **93**: 7003-7007.
- Bienvenut, W.V., Sanchez, J.C., Karmine, A., Rouge, V., Rose, K., Binz, P.A. and Hochstrasser, D.F. (1999) Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. *Anal. Chem.* **71**: 4800-4807.
- Binz, P.A., Muller, M., Walther, D., Bienvenut, W.V., Gras, R., Hoogland, C., Bouchet, G., Gasteiger, E., Fabbretti, R., Gay, S., Palagi, P., Wilkins, M.R., Rouge, V., Tonella, L., Paesano, S., Rossellat, G., Karmine, A., Bairoch, A., Sanchez, J.C., Appel, R.D. and Hochstrasser, D.F. (1999) A molecular scanner to automate proteomic research and to display proteome images. *Anal. Chem.* **71**: 4981-4988.
- Blawas, A.S. and Reichert, W.M. (1998) Protein patterning. *Biomaterials* **19**: 595-609.
- Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M. and Yuan, Y. (1998) Predicting function: from genes to genomes and back. *J. Mol. Biol.* **283**: 707-725.
- Bowtell, D.D.L. (1999) Options available - from start to finish - for obtaining expression data by microarray. *Nature Genet.* **21**: 25-32.
- Brody, E.N., Willis, M.C., Smith, J.D., Jayasena, S., Zichi, D. and Gold, L. (1999) The use of aptamers in large arrays for molecular diagnostics. *Mol. Diagnostics* **4**: 381-388.
- Brünger, A.T. (1997) X-ray crystallography and NMR reveal complementary views of structure

- and dynamics. *Nature Struct. Biol.* **4** (Suppl.): 862–865.
- Buter, D.** (2000) Celera in talks to launch private sector human proteome project. *Nature* **403**: 815–816.
- C. elegans consortium** (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium. *Science* **282**: 2012–2018.
- Celis, J.E., Rasmussen, H.H., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Gromov, P., Vorum, H., Vassilev, A. et al.** (1994) The human keratinocyte two-dimensional protein database (update 1994): towards an integrated approach to the study of cell proliferation, differentiation and skin diseases. *Electrophoresis* **15**: 1349–1458.
- Celis, J.E., Rasmussen, H.H., Gromov, P., Olsen, E., Madsen, P., Leffers, H., Honoré, B., Dejgaard, K., Vorum, H., Kristensen, D.B. et al.** (1995) The human keratinocyte two-dimensional gel protein database (update 1995): mapping components of signal transduction pathways. *Electrophoresis* **16**: 2177–2240.
- Celis, J.E., Ostergaard, M., Jensen, N.A., Gromova, I., Rasmussen, H.H. and Gromov, P.** (1998) Human and mouse proteomic databases: novel resources in the protein universe. *FEBS Lett.* **430**: 64–72.
- Chalfie, M.** (1998) Genome sequencing. The worm revealed. *Nature* **396**: 620–621.
- Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P.A.** (1996) Accessing genetic information with high-density DNA arrays. *Science* **274**: 610–614.
- Cheung, V.G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R. and Childs, G.** (1999) Making and reading microarrays. *Nature Genet.* **21**: 15–19.
- Choudhary, G., Chakel, J., Hancock, W., Torres-Duarte, A., McMahon, G. and Wainer, I.** (1999) Investigation of the potential of capillary electrophoresis with off-line matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for clinical analysis: examination of a glycoprotein factor associated with cancer cachexia. *Anal. Chem.* **71**: 855–859.
- Corbett, J.M., Dunn, M.J., Posch, A. and Görg, A.** (1994) Positional reproducibility of protein spots in two-dimensional polyacrylamide gel electrophoresis using immobilised pH gradient isoelectric focusing in the first dimension: an interlaboratory comparison. *Electrophoresis* **15**: 1205–1211.
- Cornell, B.A., Braach-Maksyvtis, V.L.B., King, L.G., Osman, P.D.J., Raguse, B., Wiczorek, L. and Pace, R.J.** (1997) A biosensor that uses ion-channel switches. *Nature* **387**: 580–583.
- Cusack, S., Belrhali, H., Bram, A., Burghammer, M., Perrakis, A. and Riek, C.** (1998) Small is beautiful: protein micro-crystallography. *Nature Struct. Biol.* **5** (Suppl.): 634–637.
- Davidsson, P., Westman, A., Puchades, M., Nilsson, C.L. and Blennow, K.** (1999) Characterization of proteins from human cerebrospinal fluid by a combination of preparative two-dimensional liquid-phase electrophoresis and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Chem.* **71**: 642–647.
- DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A. and Trent, J.M.** (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. [See comments.] *Nature Genet.* **14**: 457–460.
- Dove, A.** (1999) Proteomics: translating genomics into products? *Nature Biotechnol.* **17**: 233–236.
- Dubrovsky, T.B., Hou, Z., Stroeve, P. and Abbott, N.L.** (1999) Self-assembled monolayers formed on electroless gold deposited on silica gel: a potential stationary phase for biological assays. *Anal. Chem.* **71**: 327–332.
- Dujon, B.** (1996) The yeast genome project: what did we learn? *Trends Genet.* **12**: 263–270.
- Eckerskorn, C., Strupat, K., Schleuder, D., Hochstrasser, D., Sanchez, J.C., Lottspeich, F. and Hillenkamp, F.** (1997a) Analysis of proteins by direct scanning infrared-MALDI mass spectrometry after 2D PAGE separation and electroblotting. *Anal. Chem.* **69**: 2888–2892.
- Eckerskorn, C., Strupat, K., Kellermann, J., Lottspeich, F. and Hillenkamp, F.** (1997b) High sensitivity peptide mapping by micro-LC with on-line membrane blotting and subsequent detection by scanning IR-MALDI mass spectrometry. *J. Protein Chem.* **16**: 349–362.
- Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D.** (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**: 14 863–14 868.
- Ellington, A.D. and Szostak, J.W.** (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**: 818–822.

- Enright, A.J., Iliopoulos, I., Kyripides, N.C. and Ouzounis, C.A. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402**: 86–90.
- Fields, S. and Song, O. (1989) A novel genetic system to detect protein–protein interactions. *Nature* **340**: 245–246.
- Figeys, D. and Aebersold, R. (1999) Microfabricated modules for sample handling, sample concentration and flow mixing: application to protein analysis by tandem mass spectrometry. *J. Biomech. Engng* **121**: 7–12.
- Fromont-Racine, M., Rain, J.C. and Legrain, P. (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16**: 277–282.
- Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S. and Garrels, J.I. (1999) A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**: 7357–7368.
- Gaber, B.P., Martin, B.D. and Turner, D.C. (1999) Create a protein microarray using a hydrogel ‘stamper’. *Chemtech* **29**: 29–24.
- Gerhold, D., Rushmore, T. and Caskey, C.T. (1999) DNA chips: promising toys have become powerful tools. *Trends Biochem. Sci.* **24**: 168–173.
- Gerstein, M. and Hegyi, H. (1998) Comparing genomes in terms of protein structure: surveys of a finite parts list. *FEMS Microbiol. Rev.* **22**: 277–304.
- Gizeli, E. and Lowe, C.R. (1996) Immunosensors. *Curr. Opin. Biotechnol.* **7**: 66–71.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H. and Oliver, S.G. (1996) Life with 6000 genes. *Science* **274**: 562–567.
- Grimm, R. and Grasser, K.D. (1998) Nanogram scale separations of proteins using capillary high-performance liquid chromatography with fully-automated on-line microfraction collection followed by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry, protein sequencing and Western blot analysis. *J. Chromatogr. A* **800**: 83–88.
- Guschin, D., Yershov, G., Zaslavsky, A., Gemell, A., Shick, V., Proudnikov, D., Arenkov, P. and Mirzabekov, A. (1997) Manual manufacturing of oligonucleotide, DNA, and protein microchips. *Anal. Biochem.* **250**: 203–211.
- Haupt, K. and Mosbach, K. (1998) Plastic antibodies: developments and applications. *Trends Biotechnol.* **16**: 468–475.
- Hermann, T. and Patel, D.J. (2000) Biochemistry-adaptive recognition by nucleic acid aptamers. *Science* **287**: 820–825.
- Jensen, O.N., Mortensen, P., Vorm, O. and Mann, M. (1997) Automation of matrix-assisted laser desorption/ionization mass spectrometry using fuzzy logic feedback control. *Anal. Chem.* **69**: 1706–1714.
- Klose, J. (1999) Large-gel 2D electrophoresis. *Methods Mol. Biol.* **112** (2-D Proteome Analysis Protocols): 147–172.
- Kopp, M.U., Cräbtree, H.J. and Manz, A. (1997) Developments in technology and applications of microsystems. *Curr. Opin. Chem. Biol.* **1**: 410–419.
- Lahiri, J., Isaacs, L., Tien, J. and Whitesides, G.M. (1999) A strategy for the generation of surfaces presenting ligands for studies of binding based on an active ester as a common reactive intermediate: a surface plasmon resonance study. *Anal. Chem.* **71**: 777–790.
- Lamond, A.I. and Mann, M. (1997) Cell biology and the genome projects – a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends Cell Biol.* **7**: 139–142.
- Laricchia-Robbio, L., Liedberg, B., Platou-Viking, T., Rovero, P., Befly, P. and Revoltella, R.P. (1996) Mapping of monoclonal antibody- and receptor-binding domains on human granulocyte-macrophage colony-stimulating factor (rhGM-CSF) using a surface plasmon resonance based biosensor. *Hybridoma* **15**: 343–350.
- Lecrenier, N., Foury, F. and Goffeau, A. (1998) Two-hybrid systematic screening of the yeast proteome. *BioEssays* **20**: 1–5.
- Li, X.P., Pleissner, K.P., Scheler, C., Regitz-Zagrosek, V., Salnikow, J. and Jungblut, P.R. (1999) A two-dimensional electrophoresis database of rat heart proteins. *Electrophoresis* **20**: 891–897.
- Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R. (1999) Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**: 676–682.

- MacNair, J.E., Patel, K.D. and Jorgenson, J.W. (1999) Ultrahigh-pressure reversed-phase capillary liquid chromatography: isocratic and gradient elution using columns packed with 1.0 μm particles. *Anal. Chem.* **71**: 700–708.
- Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O. and Eisenberg, D. (1999a) Detecting protein function and protein–protein interactions from genome sequences. *Science* **285**: 751–753.
- Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O. and Eisenberg, D. (1999b) A combined algorithm for genome-wide prediction of protein function. *Nature* **402**: 83–86.
- Mullett, W.M. and Lai, E.P.C. (1999) Molecularly imprinted solid phase extraction microcolumn with differential pulsed elution for theophylline determination. *Microchem. J.* **61**: 143–155.
- Nakanishi, K., Muguruma, H. and Karube, I. (1996) A novel method of immobilising antibodies on a quartz crystal microbalance using plasma-polymerized films for immunosensors. *Anal. Chem.* **68**: 1695–1700.
- Narang, U., Gauger, P.R., Kusterbeck, A.W. and Ligler, F.S. (1998) Multianalyte detection using a capillary-based flow immunosensor. *Anal. Biochem.* **255**: 13–19.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A. and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet.* **20**: 46–50.
- Nicola, A.J., Gusev, A.I., Proctor, A. and Hercules, D.M. (1998) Automation of data collection for matrix assisted laser desorption/ionization mass spectrometry using a correlative analysis algorithm. *Anal. Chem.* **70**: 3213–3219.
- Opiteck, G.J., Ramirez, S.M., Jorgenson, J.W. and Moseley, M.A. III (1998) Comprehensive two-dimensional high-performance liquid chromatography for the isolation of overexpressed proteins and proteome mapping. *Anal. Biochem.* **258**: 349–361.
- Owens, P.K., Karlsson, L., Lutz, E.S.M. and Andersson, L.I. (1999) Molecular imprinting for bio- and pharmaceutical analysis. *Trends Anal. Chem.* **18**: 146–154.
- Parekh, R. (1999) Proteomics and molecular medicine. *Nature Biotechnol.* **17** (Suppl.): BV19–BV20.
- Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D. and Yeates, T.O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **96**: 4285–4288.
- Pennington, S.R., Wilkins, M.R., Hochstrasser, D.F. and Dunn, M.J. (1997) Proteome analysis: from protein characterisation to biological function. *Trends Cell Biol.* **7**: 168–173.
- Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C., Lashkari, D., Shalon, D., Brown, P.O. and Botstein, D. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96**: 9212–9217.
- Plasterk, R.H. (1999) The year of the worm. *BioEssays* **21**: 105–109.
- Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D. and Brown, P.O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**: 41–46.
- Rappsilber, J., Siniosoglou, S., Hurt, E.C. and Mann, M. (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal. Chem.* **72**: 267–275.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M. and Seraphin, B. (1999) A generic protein purification method for protein complex characterisation and proteome exploration. *Nature Biotechnol.* **17**: 1030–1032.
- Robson, B. (1999) Beyond proteins. *Trends Biotechnol.* **17**: 311–315.
- Romig, T.S., Bell, C. and Drolet, D.W. (1999) Aptamer affinity chromatography: combinatorial chemistry applied to protein purification. *J. Chromatogr. B* **731**: 275–284.
- RossMacdonald, P., Coelho, P.S.R., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., Heldtman, M., Nelson, F.K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G.S. and Snyder, M. (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**: 413–418.
- Rout, M.P., Aitchison, J.D., Suprpto, A., Hjertaas, K., Zhao, Y.M. and Chait, B.T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**: 635–651.

- Rowe, C.A., Scruggs, S.B., Feldstein, M.J., Golden, J.P. and Ligler, F.S. (1999) An array immunosensor for simultaneous detection of clinical analytes. *Anal. Chem.* **71**: 433–439.
- Sali, A. and Kuriyan, J. (1999) Challenges at the frontiers of structural biology. *Trends Genet.* **15**: M20–M24.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl Acad. Sci. USA* **93**: 14440–14445.
- Shi, H., Tsai, W.-B., Garrison, M.D., Ferrari, S. and Ratner, B.D. (1999) Template-imprinted nanostructured surfaces for protein recognition. *Nature* **398**: 593–597.
- Swindells, M.B., Orengo, C.A., Jones, D.T., Hutchinson, E.G. and Thornton, J.M. (1998) Contemporary approaches to protein structure classification. *BioEssays* **20**: 884–891.
- Takeuchi, T., Fukuma, D. and Matsui, J. (1999) Combinatorial molecular imprinting: an approach to synthetic polymer receptors. *Anal. Chem.* **71**: 285–290.
- Tempelman, L.A., King, K.D., Anderson, G.P. and Ligler, F.S. (1996) Quantitating Staphylococcal enterotoxin B in diverse media using a portable fibre-optic biosensor. *Anal. Biochem.* **233**: 50–57.
- Traini, M., Gooley, A.A., Ou, K., Wilkins, M.R., Tonella, L., Sanchez, J.C., Hochstrasser, D.F. and Williams, K.L. (1998) Towards an automated approach for protein identification in proteome projects. *Electrophoresis* **19**: 1941–1949.
- Turner, A.P.F. (1997) Switching channels makes sense. *Nature* **387**: 555–557.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., QureshiEmili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamar, G., Yang, M.J., Johnston, M., Fields, S. and Rothberg, J.M. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**: 623–627.
- Unlü, M., Morgan, M.E. and Minden, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**: 2071–2077.
- Viking, T.P., Askendal, A., Liedberg, B., Lindahl, T. and Tengvall, P. (1998) Immobilized chicken antibodies improve the detection of serum antigens with surface plasmon resonance (SPR). *Biosens. Bioelectron.* **13**: 1257–1262.
- Wang, K., Gan, L., Jeffery, E., Gayle, M., Gown, A.M., Skelly, M., Nelson, P.S., Ng, W.V., Schummer, M., Hood, L. and Mulligan, J. (1999) Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. *Gene* **229**: 101–108.
- Watts, H.J., Lowe, C.R. and Pollard-Knight, D.V. (1994) Optical biosensor for monitoring microbial cells. *Anal. Chem.* **66**: 2465–2470.
- Wery, J.P. and Schevitz, R.W. (1997) New trends in macromolecular X-ray crystallography. *Curr. Opin. Chem. Biol.* **1**: 365–369.
- Wigge, P.A., Jensen, O.N., Holmes, S., Souès, S., Mann, M. and Kilmartin, J.V. (1998) Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* **141**: 967–977.
- Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F. and Williams, K.L. (1995) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology* **13**: 19–50.
- Yao, S., Anex, D.S., Caldwell, W.B., Arnold, D.W., Smith, K.B. and Schultz, P.G. (1999) SDS capillary gel electrophoresis of proteins in microfabricated channels. *Proc. Natl Acad. Sci. USA* **96**: 5372–5377.
- Ye, L., Ramstrom, O., Mansson, M.O. and Mosbach, K. (1998) A new application of molecularly imprinted materials. *J. Mol. Recog.* **11**: 75–78.
- Yu, H., Kusterbeck, A.W., Hale, M.J., Ligler, F.S. and Whelan, J.P. (1996) Use of the USDT flow immunosensor for quantification of benzoylcegonine in urine. *Biosens. Bioelectron.* **11**: 725–734.
- Zachariae, W., Shevchenko, A., Andrews, P.D., Ciosk, R., Galova, M., Stark, M.J., Mann, M. and Nasmyth, K. (1998) Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins. *Science* **279**: 1216–1219.

第十一章 蛋白质组分析在药物开发和毒理学中的应用

Stefan Evers and Christopher P. Gray

1. 引言

良好的药物应该具有有效性和专一性。一般来说，它们应该首先与其靶标相互作用而对其他任何分子仅产生最低限度的效应。药物研发的主要要求之一就是确证候选化合物对潜在靶标的抑制作用（目标有效性），而不产生预计之外的效应。药物发展和毒理学分析研究中共同存在的缺点是：只注重表型的观察和单个大分子或酶解活性。对毒性或药理活性的观测常维持在分子水平上的简单理解，而除被检测的效应外，候选药物作用于潜在靶分子上的直接或间接的效应依然未知。为了弥补药理效应或毒性观察与分子水平上对效应理解之间的差距，完全有必要发展能更大规模提出生物学问题的技术。与迄今为止通过传统测试得到的结果相比，这些技术应该更深邃地洞察疾病和药理学或毒理学效应的分子过程。理解这些机理能改善在药物发展重大进程中决策的作出，同时能提高研发过程的工作效率。在这些技术的发展中，DNA 测序方法的发展是一个重要的进展，这一进展揭示了原核生物和真核生物的基因和相应蛋白质的全部功能。基因组全序列的问世揭开了生物学研究领域的新篇章。但是，特定生物体的基因组序列本身的价值毕竟有限。“功能基因”这个词涉及一个新的研究领域，这一领域使用蛋白质组分析和高密度的 DNA 阵列等技术，并充分使用在基因序列中所包含的丰富信息。

在单次实验中，蛋白质组分析可以检测并定量分析成百上千个蛋白质。这为定量和定性评价具有不同表型或基因型的组织细胞之间、正常与病理状态、给药与不给药时蛋白质的表达差异提供了可能。根据所使用的模型系统，对不同的处治方法或疾病的不同阶段的反应可以随时跟踪，可以实现对药效动力学研究或对疾病进程相关的变化特征进行有效识别。

表面上看，蛋白组学研究似乎是多余的技术。在 mRNA 水平研究基因表达似乎是一种比蛋白质组更简单而且更灵敏的手段。但是，在生物学样品的鉴定中有大量的证据说明蛋白质组分析的重要性：

(1) 通过 mRNA 浓度的测定得到的基因调控研究是间接的，仅能代表一种间接的测量，还有很多现象未能检测。

(2) 在翻译水平的基因表达调控在多数情况下是起决定性作用的，如核糖体蛋白、密码子使用和 *E. coli* 中氨基酸生物合成酶的调节衰减机理。这类调控通过 mRNA 的定量分析检测不到。

(3) 测定细胞中蛋白的浓度时，一个常被忽视的因素是它的降解速度，而蛋白质的降解可能受调控的影响 (Tobias *et al.*, 1991)。

(4) 有一点很重要，翻译中和翻译后的共价修饰（如磷酸化、酰基化或糖基化）扮演重要的生物学的角色，通常完全改变蛋白质的性质（见第五章和第十三章）。若干研

究结果显示，毒性效应可以通过蛋白质修饰的方式显示出来 (Anderson *et al.*, 1992; Witzmann *et al.*, 1998)，很多修饰的蛋白质异构体可以通过二维凝胶电泳分开，例如凝胶中成串的蛋白点代表具有不同糖基化状态的蛋白质。因而蛋白质组为研究人员提供了独特而强大的检测和研究这些修饰（比如在疾病过程中）的工具。

对于药物发展，应该认识到在绝大多数情况下基因编码的功能高分子都是蛋白质，而且药物的靶标几乎全都是蛋白质。最近的研究证实，转录组与蛋白质组的结合能够全面理解细胞内的分子过程。蛋白质组实验表明，作为致癌基因转染的结果，未被加工的无活性 ICE3 聚集在转染的恶性细胞中 (图 11.1)。其后的 mRNA 定量分析显示，活化 ICE3 的蛋白酶合成在这些细胞中剧烈下调 (Wang *et al.*, 1996)，这也就解释了 2-DE 胶上所观察到的现象 (Certa *et al.*, 1999)。

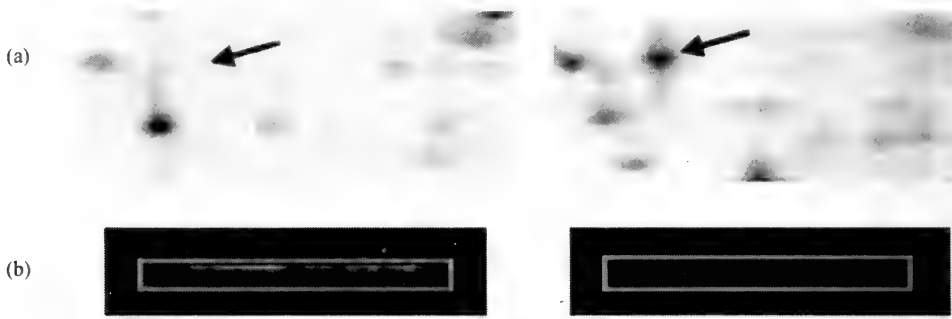


图 11.1 上图分别选自正常老鼠乳突细胞 (图左) 和癌基因 ras 转染细胞 (图右) 得到的蛋白质和 mRNA 提取物的 2DE-PAGE 胶 (a) 和 Affymetrix 基因芯片™ (b)

2-DE 胶上的箭头表示未被加工 ICE3 的位置。芯片区域表示粒酶 B 基因被包埋。粒酶 B 使 ICE3 水解激活 (Wang *et al.*, 1996)。这些结果显示在 H-ras 致癌基因的存在下，粒酶 B 的产生被抑制，导致未被加工的 ICE3 前体的聚集。似乎蛋白酶与凋亡通路有关，没有蛋白酶时细胞就会恶性地无控制地生长 (Certa *et al.*, 1999)。

蛋白质合成速度的调控是细胞对外界环境条件改变后产生反应的一部分。另外，通过细胞循环和分化的细胞成熟也由基因表达的调控所决定。复杂的调控网络多级控制蛋白质的浓度和酶活性（如 mRNA 合成，翻译，修饰，转运，蛋白水解过程和降解）。这样，蛋白组反映了环境条件的变化。蛋白质组的比较研究有助于理解细胞的调控网络。因而它是药物靶标识别和确认、药物的行为机理研究和药理学或毒性效应的检测和推断的有用工具。

2. 数据的使用

通过蛋白质组实验得到的数据量需要借助计算机分析。为了验证观测资料，使得预测可靠，必须进行重复样品的重复凝胶试验，同时得到储存、分析。实验的变化可能非常大，特别是在引入遗传变异的情况下，比如使用远系杂交种动物的毒理学研究 (Steiner *et al.*, 1995)。很明显，统计学对阐明从二维凝胶电泳得到的大量数据（常有错误倾向）极其重要。除标准二维凝胶电泳分析工具的传统统计分析（平均值，标准

差,可信度)外,还有两种基本方法可以检查从这些实验得到的数据:

(1) 蛋白质组在整体上可被看作是反映环境变化的生物传感器。像“化合物 A 的行为是否与化合物 B 类似”这样的问题可以通过这种方法解决。

(2) 从更细致的分析可以得知,化合物作用于单个蛋白质或一组蛋白的效应与其代谢途径或不同的功能类别有关。这使得观察到的药效更合理,疾病、药理学活性或毒性的潜在诊断标记更准确。

下面,将详细地讨论说明两种方法的实例,尤其着重于药物研发和毒物学方面。

2.1 比较分析

可以假定两种具有相同性质的化合物在细胞中引发相似的反应。摄药(响应)下的蛋白质数量与未摄药(空白)的相对比值可以用一系列数值代表(“摄药比”)。最后得到的表达谱可以通过各种条件来评价,提供所研究化合物性质的有价值信息。这种方法可以通过抗菌研究的实例加以验证。

筛选二氢叶酸还原酶(DHFR)抑制剂时得到一种类似三甲氧苄二氨嘧啶(Tmp)的2,4-二氨基嘧啶的衍生物Ro-64-1874。使用流感嗜血杆菌(*Haemophilus influenzae*)作为模型系统的蛋白质组分析得到一套较好表现抗菌效果的数据库比较的数据。按以下方法对Ro-64-1874的响应与标准的抗生素响应加以比较:根据摄药比,蛋白点分成三组①增加的点(摄药比 > 1.25)②减少的点(摄药比 < 0.8)③不变的点($0.8 < \text{摄药比} < 1.25$)。选择摄药比增加或者减少的点作为进一步分析的目标。通过两种条件下诱导/抑制的蛋白点数与总的点数比值、每种条件下的诱导/抑制的蛋白点数来衡量两种响应的相似程度(图11.2)。这些值的比率是衡量两种效应相似程度和重要性的较好标准。将Ro-64-1874显示的结果表达谱与数据库(包括转录、翻译,四氢叶酸合成,tRNA合成,脂肪酸合成和DNA促旋酶的抑制反应)中的其他表达谱,得到与Tmp和磺胺二甲噁唑(sulfamethoxazole, Smz)的较好的相关性。预想的Ro-64-1874的行为模式就被证实。我们最近选择一种体外活性能对抗促旋酶引发的体内次级反应的化合物,通过样品的比较,显示蛋白质合成的抑制。这种活性可以通过体外转录/翻译分析证实。研究结果显示一些化合物的响应是依赖于时间和浓度的(VanBogelen *et al.*, 1987)。

因此,加入抗生素后,需要在不同时间点和多个浓度下检查行为模式尚不清楚的化合物的响应。

这种方法应当成为抗菌研究、药学和毒物学、特别是结构-功能关系研究领域中的不同的应用手段之一(如Myers *et al.*, 1997)。对于这种比较,蛋白质点的鉴别并不重要。这种比较成功的前提是有一个来自行为方式已知的化合物的响应数据库,但这一般是不可能实现的。具有新颖行为模式的化合物不会与任何标准物质有意义的相关值,对于这些化合物更详细的分析要求对它们的性质有更好的理解。一种有趣的靶标确认方法是将用药后引起的基因表达谱的变化与药物靶标编码的基因失活或下调引起的变化相比较。cyclosporine A (CyA)和FK506在酵母细胞上的效应与编码其靶蛋白质的cal-cineurin和FPR1基因的失调效应有关,证实使用DNA微分析这种方法的有效性(Mar-ton *et al.*, 1998)。

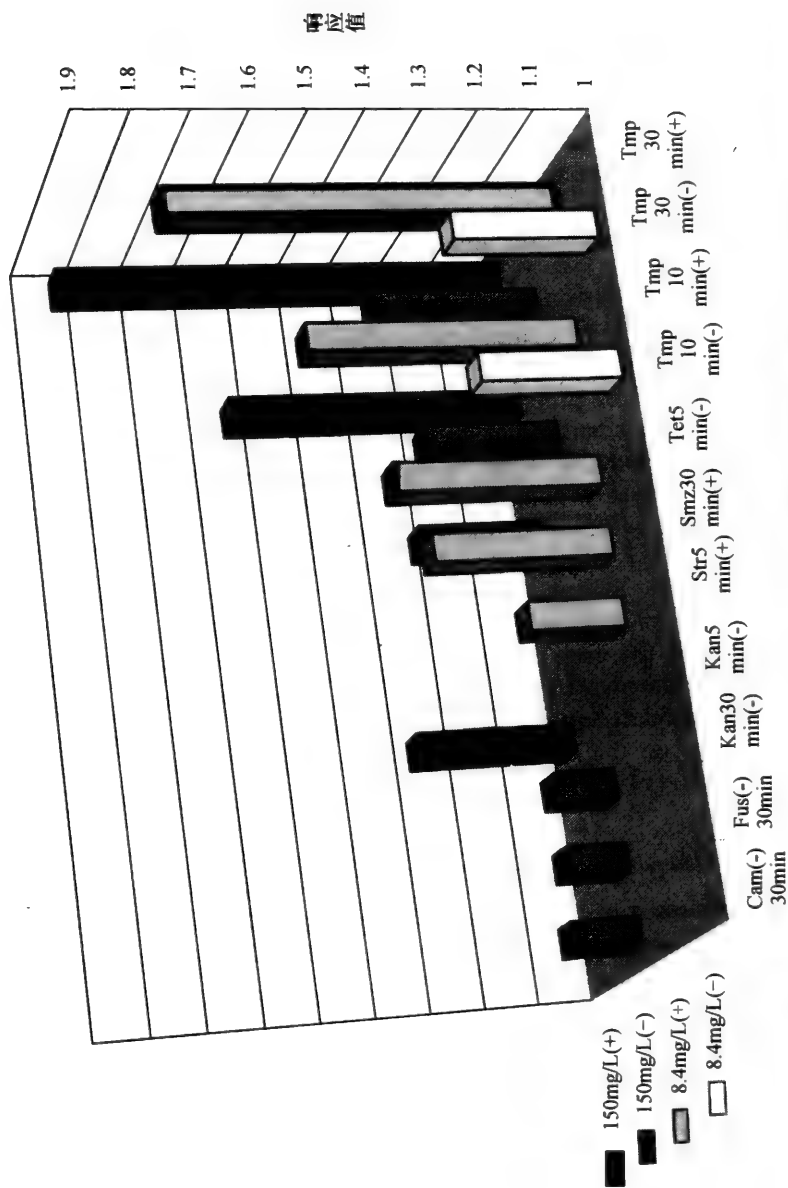


图 11.2 8mg/L 和 150 mg/L 的 Ro-64-1874 在培养的流感嗜血杆菌 (*H. influenzae*) 细胞上的效应与其他抗生素 [THF 合成抑制剂(两种化合物), 细胞壁合成(2), 转录(7), tRNA 合成(1), DNA 促旋酶(2) 和脂肪酸合成(1) 等的抑制剂] 效应的比较结果。(+) 和 (-) 分别表示合成点的增加率和减少率(诱导率 > 1.25 和 < 0.8)。使用的抗生素的缩写和各自的浓度如下, Cam: 氯霉素(2 mg/L); Fus: fusidic acid (32 mg/L); Kan: 卡那霉素(32 mg/L); Smz: 磺胺二甲嘧啶(120 mg/L); Str: 链霉素(24 mg/L); Tet: 四环素(2 mg/L); Tmp: 甲氧苄氨嘧啶(1 mg/L)。培养细胞在 Ro-64-1874 中暴露 60 分钟。其余的暴露时间也被标出。有效值的计算是通在 8mg/L Ro-64-1874 和 Tmp(用粗线表示的柱状图) 诱导下增加点的比较加以证明; 58% 的 Tmp 诱导点也被 Ro-64-1874 诱导, 另一方面, 仅有 31% 的点被 Ro-64-1874 诱导。这两个百分比之间的比率 1.84 可作为两种响应相似性有效值的衡量。仅显示匹配比率在 1.1 以上的部分。

2.2 生物标志物蛋白质的检测

蛋白质组数据库能根据蛋白质点的强度反映一种疾病的过程变化或对一组药物反应的变化。这样,蛋白质可以用作特定反应或疾病的替代标志物。也许蛋白质组在毒理学领域成功应用最好的例子是 CyA 效应的研究。通过检查诱导方式,可以假定 CyA 的毒性机制。CyA 是一种应用广泛的免疫抑制剂,但应用中也伴随着许多副作用,其中最值得注意的是肾毒。用 CyA 处治后最明显的结果是,部分蛋白点消失,这些蛋白点后来从大鼠肾脏蛋白质模式中被鉴定为 calbindin-D 的 28 kDa 蛋白 (Benito *et al.*, 1995; Steiner *et al.*)。使用酶联免疫吸附分析的进一步研究证实,大鼠肾脏中 calbindin-D 的 28kDa 蛋白量持续减少 (Steiner *et al.*, 1996)。这种减少伴随着尿钙排泄物和皮层脊髓管内石灰化的增加 (Aicher *et al.*, 1997)。就像以前提出的那样 (Feher, 1983),如果 calbindin-D 28kDa 蛋白扮演钙运输的促进器和钙缓冲器的角色,观察到的效应就能得到解释。CyA 毒性标志物的鉴定,为其毒效提供了可靠的说明,这是蛋白质组与其他的成功应用结果。第二个生物标记蛋白质的成功鉴别的例子来自膀胱鳞状细胞癌进程的长期研究 (综述见 Celis *et al.*, 1998)。在此研究中,定义蛋白质标志物和癌损害的分化程度能被监测 (Celis *et al.*, 1996; Ostergaard *et al.*, 1997)。

这种生物标记物识别的方法在制药产业的不同领域,从临床前研究到诊断、药理学和毒物学等领域都有巨大的应用潜力。蛋白质组在检测一个蛋白质表达谱是否可以用于疾病诊断方面,主要依赖于用来提取信息的数据库的规模和质量。因此,几个存储从 2-DE 胶产生的大量表达谱和质谱数据的公共数据库已经构建 (见第八章)。不同类型的多元统计分析,如主要组分分析或模型预测,均可用于阐述生物标志物蛋白。

2.3 详细的基因表达调控分析

对胶上蛋白质点的识别鉴定匹配越多,对实验数据的详尽分析就越有可能。与特定的通路或调控子有关的大多数蛋白质的蛋白水平都可以定量。目前,仅仅从微生物的蛋白质识别中获得了足够的信息,可以详细地分析不同生长条件和抗生素作用下对微生物基因表达的影响。F.C.Neidhardt 小组最先建立了大肠杆菌的 2-DE 谱并开始对这些细菌对不同环境条件的反应进行系统的研究、分类 (VanBogelen *et al.*, 1997)。这样,大量不同的调控子已经被发现并表征 (VanBogelen *et al.*, 1997)。几种微生物,如枯草杆菌 (*Bacillus subtilis*) (Antelmann *et al.*, 1997), 流感嗜血杆菌 (*Haemophilus influenzae*) (Cash *et al.*, 1997; Langen *et al.*, 1997), 以及酿酒酵母 (*Saccharomyces cerevisiae*) (Boucherie *et al.*, 1996; Hodges *et al.*, 1999) 的蛋白质组表达谱已经发表。微生物全基因序列的完成大大促进了蛋白质的识别进程,并使代谢途径的构建成为可能。但应该记住,在大多数情况下,功能的指认是以序列同源为基础,传统的生物化学对这些途径的完成和确认至关重要。我们小组已经报道了一个生物化学途径详尽分析的例子 (Evers *et al.*, 1998)。如上所述,我们已经研究了流感嗜血杆菌 (*H. influenzae*) 对四氢叶酸合成抑制剂 Tmp 和 Smz 的反应。其中最重要的变化是代表与蛋氨酸合成有关的酶蛋白点的强度增加。这些结果显示编码这些蛋白质的基因以 *E. coli* 中配对物类似的方式被调控 (Weissbach and Brot, 1991)。因为 2-DE 胶上很多相关的蛋白质还未被鉴

定, 现在还不可能研究 Tmp 和 Smz 作用于嘌呤和胸苷生物合成酶的效应, 它们同样受四氢叶酸合成抑制影响。在对酵母的类似研究中, 进行了嘌呤和组氨酸生物合成蛋白的合成率调控研究 (Denis *et al.*, 1998), 结果证实了这种方法在真核细胞中的可行性。

这种详尽的分析有助于对数据库中没有类似物的化合物建立行为模式假说。因而有助于发现药理学活性物质的新靶标。另外, 在这些分析中观察到的次要效应也有助于识别药物介入的潜在靶标。结合其他方法 (例如重组基因失活), 可以更完整地描述药物效应。结合正在进行的基因组项目产生的大量信息, 这个方法将可能扩展到更复杂的有机体中去。

2.4 蛋白质功能预测

对许多蛋白质来说, 以序列比较为基础指认蛋白质的功能是不可能的。这些蛋白质代表了未被利用的药物潜在的靶标, 因此是医药工业的特殊兴趣所在。“调控同源”(regulatory homology) 的概念是由 Anderson 和 Anderson (1998) 最近提出的: 当面临环境条件的改变时, 以相似的方式变化的蛋白质相对丰度也许参与共同的过程。这个概念对预测蛋白质功能是有用的。另外, 缺乏或过表达“孤独”(orphan) 基因的重组生物体的蛋白质组变化可以显示这类蛋白质的可能功能。在硫酸盐影响 *E. coli* 基因表达效果的研究实验中, 通过 2-DE 凝胶的对比, 鉴定了 8 个与硫酸盐调控有关的蛋白质 (Dainese *et al.*, 1997)。对其中一个蛋白 TauD 的基因重组失活的实验结果显示, 细菌在惟一的硫磺源——牛磺酸上生长不足 (Dainese *et al.*, 1997)。这样, 通过蛋白质组技术鉴定了与硫酸盐同化作用有关的蛋白质, 而重组技术证明了其中一个蛋白, TauD 在牛磺酸的去磺化中的作用。

3. 结论与展望

目前蛋白质组对药物的研究和开发十分有用的观点已被广泛地接受。大多数药物公司为改善与药物行为相关的生物学材料的表征, 发现相关生物标志物和潜在的靶标, 已经纷纷建立了蛋白质组分析小组。Large Scale Biology Corp, 现在是 Biosource Technologies Inc. 的分部, 是分子药理学和毒理学的 2-DE 应用领域的先驱, 构建了全面的药物的分子效应 (Molecular Effects of DrugsTM) 数据库 (Anderson and Anderson, 1998), 为药物产业提供 2-DE 服务, 为公众提供他们所拥有的蛋白质谱数据库和 2-DE 谱图。几个其他的公司随即跟进, 也为药物发现和发展提供这些服务或开发相应的蛋白质组研究项目 (见第十一章“商业蛋白质组”)。

蛋白质组被用于药物发现和发展过程中的各个阶段。如上所述, 它对于早期靶标的识别和确认非常有用。在发展的后期, 其实用性已经在药理学和毒理学领域中得到证明。尤其在后两个领域, 从蛋白质组研究结果的重要意义以及与传统生理学和组织学试验得到的观测资料的相关性需要细心评价。这些研究有希望鉴别新型的生物标志物, 有助于发展更迅速、更经济的诊断试验。这些生物标志物的检测应用于药物发展的早期阶段, 可减少动物实验的需求。

为了充分地开发蛋白质组的潜力, 目前技术发展仍需与方法学结合, 以解决其局限性。2-DE 的使用是众多不满意的因素之一, 它操作不方便而且耗时, 阻碍了蛋白质组

作为高通量技术的使用。在巧妙的技术结合中，2-DE 可用于鉴别标志物蛋白，而使用别的更适当的方法进行高通量检测（例如 ELISA，基因系统报告，生物化学试验）。2-DE 的研究重点是自动操作和图像分析的改善，这样就能加速数据的获得。医药产业对高通量的需求可能使得 2-DE 胶最终被别的更易于自动化分离的系统所取代。现在，仅能对含量丰富的可溶性蛋白质进行研究，较大的疏水蛋白因其灵敏度较低、难于溶解而难以研究。因此蛋白质组研究也就必须偏向于这些通常是结构蛋白质或代谢酶的蛋白质（表 11.1）。用于蛋白质修饰的检测、鉴别、定量的蛋白质组技术正在开发中。目前仅有极少数蛋白质修饰被完全识别、表征的例子。技术发展加大了前进的步伐，也期望质谱分析的灵敏度和准确度的增加。随着分析软件的改进，将允许进行蛋白质修饰的常规分析并拓宽蛋白质组研究的范围。

表 11.1 流感嗜血杆菌可溶性蛋白质提取物的 2-DE 凝胶制备蛋白质的不同功能类型

功能分类 ^a	每种类别的鉴别百分比/%
氨基酸生物合成	44
辅助因子，非肽基团和载体的生物合成	27
细胞膜	22
细胞内加工	30
主要中间物代谢	37
能量代谢	41
脂肪酸和磷脂代谢	40
假定的功能	19
其他类别	23
嘌呤，嘧啶，核苷和核苷酸	40
调控功能	16
复制	25
转录	33
翻译	49
转运和结合蛋白	20
合计	29

^a According to Fleischmann *et al.* (1995)。

大多数蛋白质组研究或 mRNA 分析观察到的现象难于解释。因为调控通路是相互关联的，一种效应可能发生在表面上无关联的、不同功能类别的蛋白质中。一级、二级或三级效应相互交错，形成非常复杂的表达方式。在生物学研究中，这些方法适合产生一些假说，而这些假说需要用其他技术验证。另外，来自不同来源的信息，如代谢物的定量信息，有助于对细胞生物化学的基本理解。生物信息学在数据储存和分析中扮演着必不可少的角色，把我们目前对基因调控方面的知识与从蛋白质组及其他来源的发现有有机结合起来，形成对生物体的更全面、“整体的”看法，并且得到易于理解的数据。

致谢

感谢 Hanno Langen and Uli Certa 提供图 11.1，感谢 W. Keck 对原稿建设性的阅读。

(刘尚义译 钱小红校)

参考文献

- Aicher, L., Meier, G., Norcross, A., Jakubowski, J., Varela, M., Cordier, A. and Steiner, S. (1997) Decrease in kidney calbindin-D 28 kDa as a possible mechanism mediating cyclosporine A- and FK-506-induced calciuria and tubular mineralization. *Biochem. Pharmac.* **53**: 723–731.
- Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**: 1853–1861.
- Anderson, N.L., Copple, D.C., Bendele, R.A., Probst, G.S. and Richardson, F.C. (1992) Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. *Fundam. Appl. Toxicol.* **18**: 570–580.
- Antelmann, H., Bernhardt, J., Schmid, R., Mach, H., Volker, U. and Hecker, M. (1997) First steps from a two-dimensional protein index towards a response-regulation map for *Bacillus subtilis*. *Electrophoresis* **18**: 1451–1463.
- Benito, B., Wahl, D., Steudel, N., Cordier, A. and Steiner, S. (1995) Effects of cyclosporine A on the rat liver and kidney protein pattern, and the influence of vitamin E and C coadministration. *Electrophoresis* **16**: 1273–1283.
- Boucherie, H., Saggiocco, F., Joubert, R., Maillet, I., Labarre, J. and Perrot, M. (1996) Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* **17**: 1683–1699.
- Cash, P., Argo, E., Langford, P.R. and Kroll, J.S. (1997) Development of a *Haemophilus* two-dimensional protein database. *Electrophoresis* **18**: 1472–1482.
- Celis, J.E., Rasmussen, H.H., Vorum, H., Madsen, P., Honore, B., Wolf, H. and Orntoft, T.F. (1996) Bladder squamous cell carcinomas express psoriasin and externalize it to the urine. *J. Urol.* **155**: 2105–2112.
- Celis, J.E., Ostergaard, M., Jensen, N.A., Gromova, I., Rasmussen, H.H. and Gromov, P. (1998) Human and mouse proteomic databases: novel resources in the protein universe. *FEBS Lett.* **430**: 64–72.
- Certa, U., Hochstrasser, R., Langen, H., Buess, M. and Moroni, C. (1999) Biosensors in biomedical research: development and application of gene chips. *Chimia* **53**: 57–61.
- Dainese, P., Staudenmann, W., Quadroni, M., Korostensky, C., Gonnet, G., Kertesz, M. and James, P. (1997) Probing protein function using a combination of gene knockout and proteome analysis by mass spectrometry. *Electrophoresis* **18**: 432–442.
- Denis, V., Boucherie, H., Monribot, C. and Daignan-Fornier, B. (1998) Role of the myb-like protein bas1p in *Saccharomyces cerevisiae*: a proteome analysis. *Mol. Microbiol.* **30**: 557–566.
- Evers, S., Di Padova, K., Meyer, M., Fountoulakis, M., Keck, W. and Gray, C.P. (1998) Strategies towards a better understanding of antibiotic action: folate pathway inhibition in *Haemophilus influenzae* as an example. *Electrophoresis* **19**: 1980–1988.
- Feher, J.J. (1983) Facilitated calcium diffusion by intestinal calcium-binding protein. *Am. J. Physiol.* **244**: C303–307.
- Fleischmann, R.D., Adams, M.D., White, O. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496–512.
- Hodges, P.E., McKee, A.H.Z., Davis, B.P., Payne, W.E. and Garrels, J.I. (1999) The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res.* **27**: 69–73.
- Langen, H., Gray, C., Röder, D., Juranville, J.-F., Takacs, B. and Fountoulakis, M. (1997) From genome to proteome: protein map of *Haemophilus influenzae*. *Electrophoresis* **18**: 1184–1192.
- Marton, M.J., DeRisi, J.L., Bennett, H.A., Vishwanath, I.R., Meyer, M.R., Roberts, C.J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett, D.E.J., Hartwell, L.H., Brown, P.O. and Griend, S.H. (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4**: 1293–1301.
- Myers, T.G., Anderson, N.L., Waltham, M., Li, G., Buolamwini, J.K., Scudiero, D.A., Paull, K.D., Sausville, E.A. and Weinstein, J.N. (1997) A protein expression database for the molecular pharmacology of cancer. *Electrophoresis* **18**: 647–653.
- Ostergaard, M., Rasmussen, H.H., Nielsen, H.V., Vorum, H., Orntoft, T.F., Wolf, H. and Celis, J.E. (1997) Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. *Cancer Res.* **57**: 4111–4117.
- Steiner, S., Wahl, D., Varela, M.C., Aicher, L. and Prieto, P. (1995) Protein variability in male and female Wistar rat liver proteins. *Electrophoresis* **16**: 1969–1976.

- Steiner, S., Aicher, L., Raymackers, J., Meheus, L., Esquer-Blasco, R., Anderson, N. and Cordier, A. (1996) Cyclosporine A decreases the protein level of the calcium-binding protein calbindin-D 18 kDa in rat kidney. *Biochem. Pharmac.* **51**: 253-258.
- Tobias, J.W., Shrader, T.E., Rocap, G. and Varshavsky, A. (1991) The N-end rule in bacteria. *Science* **254**: 1374-1377.
- VanBogelen, R.A., Kelley, P.M. and Neidhardt, F.C. (1987) Differential induction of heat shock, SOS, and oxidation stress regulons and accumulation of nucleotides in *Escherichia coli*. *J. Bacteriol.* **169**: 26-32.
- VanBogelen, R.A., Abshire, K.Z., Moldover, B., Olson, E.R. and Neidhardt, F.C. (1997) *Escherichia coli* proteome analysis using the gel-protein database. *Electrophoresis* **18**: 1243-1251.
- Wang, S., Miura, M., Jung, Y., Zhu, H., Gagliardini, V., Shi, L., Greenberg, A.H. and Yuan, J. (1996) Identification and characterization of Ich-3, a member of the interleukin-1 beta converting enzyme (ICE)/Ced-3 family and an upstream regulator of ICE. *J. Biol. Chem.* **271**: 20 580-20 587.
- Weissbach, H. and Brot, N. (1991) Regulation of methionine synthesis in *Escherichia coli*. *Mol. Microbiol.* **5**: 1593-1597.
- Witzmann, F.A., Daggett, D.A., Fultz, C.D., Nelson, S.A., Wright, L.S., Kornguth, S.E. and Siegel, F.L. (1998) Glutathione S-transferases: two-dimensional electrophoretic protein markers of lead exposure. *Electrophoresis* **19**: 1332-1335.

第十二章 噬菌体抗体作为蛋白质组学研究工具

Kevin Pritchard, Kevin S. Johnson and Ulla Valge-Archer

1. 蛋白质组学方法

1.1 来源于 EST 数据库的蛋白质数据

为了更详尽地理解基因的生物学功能，人们在基因组测序方面付出了巨大的努力。尽管全长序列数据库在不断增长，但更多的有用信息来源于被称为表达序列标签 (EST) 的短 DNA 序列 (Adams *et al.*, 1991)，尽管是一些片段，但它们却可能代表了人类表达基因组的大部分。解释这些数据的生物学相关性是一个挑战，它远比 DNA 序列测定的难度大。很明显，我们需要了解 EST 所编码的蛋白质的表达谱，以作为功能研究的线索。因而，现在基因组测序的努力与建立和描述更为复杂的蛋白质组目录的努力同时进行。

当前，蛋白质组学着重依赖于两项互补的技术，即二维凝胶电泳和质谱，从选择的组织和器官中显示、鉴定代表部分蛋白质组的蛋白质，并编纂蛋白质目录。按照这一途径，研究者从级分的蛋白质找到其相应的 DNA 序列。这一方法的局限性在于建立的数据库具有严重的偏性，其中高丰度的蛋白质出现频率高。和 DNA 和 RNA 不同，蛋白质不能被扩增。具有重要功能低丰度的蛋白常常检测不到，现在，还不能定义一个完全的蛋白质组。

还有另外一种基于 DNA 序列数据而制备蛋白质特异探针的方法，利用这些探针可以鉴定特定的蛋白，研究其在细胞提取物、完整细胞、组织以及活的有机体中的功能。与传统的方法相比，这种靶向性的方法很可能可以更快地产生全面、有用、可解释的数据。利用靶向分子探针进行蛋白质组的分析依赖于与基因组计划相匹配的制备特异性探针的速度和规模。下面我们将描述如何通过噬菌体抗体库实现这一想法。

噬菌体展示技术 (Barbas *et al.*, 1991; McCafferty *et al.*, 1990) 是惟一一种在数量和速度上有能力产生满足能处理基因组测序计划所产生的信息流的特异性抗体制备技术。传统的方法，如二维电泳，是利用一般的染色剂检测蛋白质，而特异性的探针，如噬菌体抗体，则可以在从 Western 印迹到完整生物体的几乎任何生物学环境中揭示和量化特异蛋白。我们将描述如何高通量地实现这一想法的计划，即 ProAb™ 计划。其独特之处在于 ProAb™ 方法不像斑点印迹和二维电泳那样从提取物中分析蛋白质的表达，而是从组织学水平上直接分析蛋白质的表达，因此，可以积累组织、细胞及亚细胞水平上每种蛋白质的分布细节。

蛋白质并不是孤立地起作用，而是与其他蛋白及小分子配体相互作用而发挥功能。从关联蛋白及多分子蛋白复合体拓扑学得到的信息常常是了解蛋白质功能的关键。我们将描述一种被称之为 ProxiMol™ 的技术，通过这项技术可以获得针对靶分子的特异抗

体, 而此项技术的关键是识别靶分子的导向分子, 这些导向分子可以是任何一种类型的配体分子 (Osborn *et al.*, 1998)。ProxiMol™ 技术为我们提供了获得一种蛋白质特异抗体的方法, 这种方法并不需要从其自然状态中纯化和克隆靶蛋白。而且, 识别蛋白质表位的一种抗体原则上可被用作导向分子, 进而获得一组新抗体, 其中一些抗体分子甚至具有针对抗原的直接功能, 如中和配体的结合作用。这一项技术具有巨大的潜力, 因为目标蛋白分布的原始数据很快就可以得到来自于细胞和组织功能分析数据的印证。

1.2 抗体是蛋白质特异分子探针的合理选择

数十年来, 从动物中获得的抗体是生物化学研究中不可缺少的工具, 并且是多功能的试剂, 它们可用于特异、灵敏、定量地检测组织和提取物中的蛋白质。它们可以揭示特定蛋白在组织中的亚细胞定位; 并且通过阻止它所识别的目标蛋白与其天然配体的结合可以证实这些目标蛋白特异的生物学功能。从功能基因组学到蛋白质组学, 抗体都具有巨大的应用潜力, 因为基于噬菌体展示 (McCafferty *et al.*, 1990) 而产生抗体的新技术可以高通量、自动化地产生高质量的抗体试剂, 以满足对基因组计划所产生的大量潜在靶蛋白研究的需要。下一节将概述这些方法, 它们将构成基于抗体的功能基因组学研究平台的基础。

1.3 关键技术概观

噬菌体展示抗体 抗体的功能结构域可以被展示在丝状噬菌体的表面 (McCafferty *et al.*, 1990; Winter and Milstein, 1991)。大的噬菌体抗体库 ($>10^{11}$ sequences) 包含能结合任何给定抗原的高亲和力抗体 (Vaughan *et al.*, 1996)。编码抗体的基因被包裹在噬菌体颗粒中, 这样就将抗体的特异性与其基因序列联系起来。通过目标抗原的亲合作用从噬菌体抗体库中筛选所要的特异性抗体。因此, 噬菌体库相当于一个体外免疫系统。而传统的免疫方法则需要整体动物 (每种抗原至少一只动物), 用数周的时间, 通过多次免疫才能使免疫系统产生出有用的抗体。与传统的免疫方法相比, 噬菌体展示系统是一种体外系统, 产生抗体只需要几天的时间。因此, 噬菌体展示技术是一种最有效的制备抗体的方法。

噬菌体作为试剂 展示抗体功能结构域的噬菌体颗粒被常规地用于 ELISA 检测、免疫细胞化学 (ICC)、流式细胞术及其他的一些筛选技术。由于不需要进行蛋白质纯化, 因此, 大大加速了筛选的过程。

高通量地产生抗体 利用噬菌体展示技术, 抗体的分离可以在一个工作周中完成。将抗原排列在微孔板中, 可以手工完成针对数百种不同抗原的抗体筛选。噬菌体展示技术易于自动化, 其产生抗体的速度完全可以应付 DNA 序列数据库中现存的大量 DNA 序列。

抗原 在 ProAb™ 技术中联系着基因序列和蛋白质 经过筛选从 EST 序列中找出感兴趣的序列。肽序列由特定的可读框所决定, 其相应的抗原肽在 96 孔板中由自动合成仪自动合成。之后, 这些抗原被用于所期望的特异性抗体的筛选。来源于 DNA 数据库中的信息被直接用于产生表达序列的抗体探针, 而不需要全长 DNA 序列, 也不需要蛋白质抗原的表达。

抗原：通过 Proximol™ 技术探测未克隆的蛋白质及其关联者。特异识别完整细胞或组织中蛋白质的噬菌体抗体可以被分离出来，但是，如何进行针对完整细胞或其他多蛋白分子复合物中的一种特定靶分子的筛选？只要存在任何一种导向分子，一种被成为 Proximol™ 的技术就可以拯救特异识别导向分子结合的靶分子或其关联者的噬菌体。

抗原：解决自身耐受问题 噬菌体展示系统绕过免疫系统。那些针对自身抗原（包括序列完全保守的抗原）的抗体都可以直接从噬菌体抗体库中分离出来（Griffiths *et al.*, 1993）。从我们的人源抗体库中分离识别人蛋白的抗体还从来没有失败过。

检测蛋白表达及其细胞分布 ICC 能够展示蛋白质在细胞和组织中分布的精细画面。而目前面临的挑战是如何高通量地进行 ICC 分析，因为要解释这些数据需要专门的知识，并且可视的结果和解释必须被存放于易于进入且用户界面友好的数据库中。

生物信息学推动这一进程 在过去，个别的科学家很容易理解和解释低通量的技术所产生的信息流。而现在，没有复杂的生物信息学工具，没有人能够处理大量而复杂的数据。从 EST 到 ICC 图，整个发现过程都需要生物信息学的专门知识。

2. 噬菌体抗体库的构建和应用

2.1 抗体片段在噬菌体上的展示

功能抗体片段可融合于噬菌体次要外壳蛋白 g3p 的 N 末端而展示于丝状噬菌体的表面。大多数的噬菌体颗粒仅展示不超过一种的抗体拷贝（McCaffery *et al.*, 1990）。噬菌体表面所展示的抗体片段的形式为单链抗体（SC-Fv），由 IgG 的重链可变区和轻链可变区（V_H 和 V_L）构成，中间通过柔性的连接肽连接。编码抗体结构域的 DNA 序列可以被克隆到噬菌体基因组中（一种噬菌体载体）；然而，人们一般更喜欢选用噬菌粒载体，因为噬菌粒载体易于操作，并且转化效率高。（噬菌粒是一种质粒，带有质粒和噬菌体两种复制起点。）这一章描述的工作采用噬菌粒载体 pCANTAB6（McCafferty *et al.*, 1994）和 M13 噬菌体（Yanisch *et al.*, 1985）。具体操作可参考更好的操作手册（Kay *et al.*, 1996；McCaffery *et al.*, 1996）。

2.2 抗体库的构建

人淋巴细胞是理想的 IgGs 的 V_H 和 V_L DNA 序列的来源（图 12.1）。利用引物家族，通过 PCR 方法扩增抗体的可变区结构域（Marks *et al.*, 1991）。V_H 和 V_L 基因序列再通过组装 PCR 连接成完整的 scFv 抗体 DNA 序列。V_H 和 V_L 随机配对，非天然的配对增加了抗体库的功能多样性。scFv 抗体 DNA 序列通过 PCR 进行扩增，同时在引物中引入限制酶切位点，以利于将其克隆到噬菌粒载体中。连接到噬菌粒载体中后，通过电穿孔法转化 *E. coli*。在辅助噬菌体的辅助下，噬菌粒 DNA 被包装，通过拯救培养液中的噬菌粒获得噬菌体抗体库（McCaffery *et al.*, 1996）。

2.3 库的大小和多样性

特异性和亲和性是任何抗体都有的品质。库越大，它包含一种具有高亲和力、高特异性抗体的几率越大（Perelson and Oster, 1979）。尽管用于抗体库构建的供体细胞被

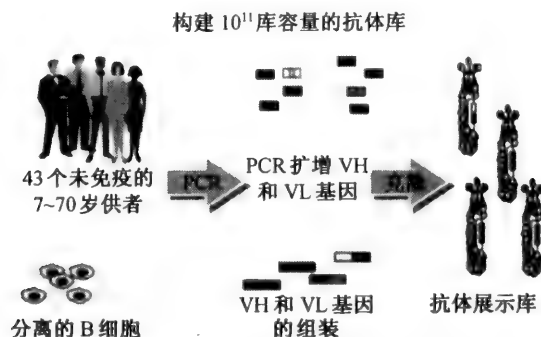


图 12.1 构建噬菌体抗体库的原理

编码人抗体抗原结合域的 DNA 被包装于噬菌体颗粒中，这种噬菌体颗粒能够在其末端展示具有功能的抗体蛋白。这种抗体的抗原结合特异性和亲和力可用于分离其相应的抗原。

描述为非免疫的，但人具有的巨大的 V-基因库仍然是经过繁殖剔除的，这些细胞一直接触许多不同种类的抗原。人 B 细胞可以产生针对自身抗原的抗体，而这些抗原是人正常的蛋白，这些细胞克隆在免疫系统监视下被剔除，但是这种局限性在噬菌体抗体技术中并不存在。人源抗体库能够产生至今为止所试验的针对任何一种抗原的抗体。一个库容量超过 10^9 克隆的抗体库 (Vaughan *et al.*, 1996) 被用于制备针对各种不同种类抗原的 scFv，并且系统地研究了其所产生抗体的亲和力。很明显，这个抗体库包含有亲和力超过 10^8M^{-1} 的针对各种不同种类抗原的 scFv，有的甚至达到 10^{10}M^{-1} 。这项工作表明，通过噬菌体抗体技术产生的抗体的亲和力和通常通过免疫方法制备的抗体的亲和力一样高 (Foote and Eisen, 1995)。尽管随着库容量的增大，从统计学角度验证抗体品质的提高将更加费力，但经验证明，与一个小库相比，从一个更大的库中更容易筛选到高亲和力的抗体。此外，检测高于 10^{10}M^{-1} 的亲和力从技术上变得更具挑战性。当前，作者所知道的最大的库包含有 10^{11} 以上不同序列。但构建这样一个库可不是一项简单的工作。

2.4 从噬菌体抗体库中筛选特异抗体

最基本的原则就是亲和捕获。被包裹在噬菌体颗粒中编码抗体的基因序列决定所展示的 scFv 的特异性，通过亲和捕获，蛋白质和基因被同时选择。

对纯化抗原进行的筛选 被广泛采用的筛选技术是淘洗。抗原被包被在固相表面，如微孔板，然后，加入噬菌体抗体库与抗原温育。经淘洗之后，留下那些表面展示有能力结合目标抗原的抗体的噬菌体。将特异噬菌体感染的 *E. coli* 进行繁殖；将细菌培养在培养板上使之形成克隆，每个克隆来源于一个细胞，产生一种单克隆抗体。

对抗原复合物进行的筛选 噬菌体抗体库常被用于对整个细胞或亚细胞组份如膜制备物进行抗体筛选，也能够获得单克隆的 scFv，但是，这种使用抗原复合物进行的筛选将导致难以确定筛选抗体的特异性。ProxiMolTM 技术 (见后) 的出现克服了这一困难。

2.5 高通量的筛选

从噬菌体抗体库中筛选抗体的程序很简单,操作的体积也很小(100 μ l),并且可在多孔微板中并行进行。不同的抗原分别被包被在不同的微孔中,一些早期的手工实验可在一个微孔板中同时筛选针对83种不同抗原的抗体。大多数的药物研究实验室采用96孔或是384孔板,用机器人进行自动化操作。同样地,商业化的现货供应类的生产工厂(如Tecan和GRI)一周能够分离多达针对200种不同抗原的抗体。

2.6 作为试剂的噬菌体抗体

展示scFv的噬菌体颗粒是研究的有用工具。在基因组水平上,研究和发现大量靶蛋白需要尽可能高通量地制备相应的试剂,而噬菌体抗体就满足了这种需要。下面我们将描述如何从噬菌体抗体库中方便地获得可溶性的抗体蛋白;然而,我们要强调的是,在所有的上游检测分析中,这种噬菌体形式的抗体都具有较好的应用。噬菌体抗体已被成功用于蛋白质、多肽、完整细胞及亚细胞组份的ELISA检测、流式分析和免疫细胞化学分析(ICC)。噬菌体颗粒主要外壳蛋白g8p有近3000多个拷贝,因此,抗噬菌体的抗血清可以高灵敏度地检测结合到目标抗原上的噬菌体。在接种辅助噬菌体后,96孔板中的*E. coli*培养上清可被直接用作试剂(McCafferty *et al.*, 1996)。

2.7 来源于噬菌体库中的可溶性抗体

scFv的应用 一般地,最快的获得数据的方式就是采用噬菌体颗粒直接进行检测。然而,在目的蛋白的下游研究实验中,最好还是采用可溶性的scFv蛋白。如亲和力和动力学的测定、细胞中和实验以及活体研究。也许还有一些其他的过程,在这些过程中应用大的噬菌体颗粒将会阻碍它向抗原的扩散。而且,噬菌体的大小限制了抗体的最大浓度约为10nmol/L(6×10^{12} phage/ml)。不论在活体还是体外,scFv被证明是主要的细胞生物学研究的良好试剂。下面的部分将会说明如何制备可以满足实验需要的scFv(μ g \rightarrow mg)。

大量地制备scFv 满足检测用量的scFv可以在不离开多孔板的情况下直接表达。为了诱导可溶性scFv表达细菌的外周腔中,含有pCANTAB质粒的细胞克隆用IPTG诱导表达,表达的scFv将会部分地泄漏到培养液中,特别是在过夜诱导表达之后。多年的经验证明,这一简单方案适合于检测分析(McCafferty *et al.*, 1996)。

更多的scFv可通过标准程序裂解细胞,从细菌的外周质中获得。细菌的外周质提取物可直接用于许多分析工作。而且,pCANTAB质粒载体在scFv的C末端还加了一个含6个组氨酸残基的小肽标签,因此,表达的scFv很容易通过固相金属亲和层析(IMAC)进行纯化。而IMAC磁珠(Qiagen)使得从96孔板中分离scFv变得更加方便、快捷。

若需要100 μ g到mg级的scFv(如体外研究),则需要扩大到摇瓶进行培养。通过IMAC纯化的scFv的量可以适合于大多数的分析;如果需要,通过简单的凝胶过滤可以得到纯度90%的scFv。

2.8 以 IgG 的形式表达

完整的 IgG 分子比 scFv 具有更长的药代动力学半衰期，并且是非常稳定的蛋白质。因此，IgG 也许是体内或体外用于长期研究较合适的抗体形式。噬菌体形式展示的抗体很容易被克隆到 IgG 载体中，因此，研究用的 IgG 可以直接用于生产治疗用的 IgG。已经证明，有许多方法可用于从哺乳动物细胞中表达毫克到克级的人源 IgG。这使得“从基因到临床”的计划变为现实。

3. 噬菌体抗体在功能基因组研究中的应用

3.1 “ProAbTM” 方法

维持基因组测序计划的巨大资源和巨大的蛋白质组数据库的出现，主要是由于人们认为下一代药物计划而发现新的靶蛋白将会很快获利。从制药企业和医学研究的观点看，对任何新蛋白最有兴趣的信息就是在某种疾病状态下，在某些特殊的组织中，它的丰度有明显的变化。合乎理想的情况是研究者喜欢通过高通量的方法将基因的表达和疾病的指征相联系。这就是 ProAb 计划的主题思想。这一过程的示意图见图 12.2。其策略是将 EST 序列翻译成多肽抗原，用这些多肽抗原筛选噬菌体抗体，这些抗体可用于在 ICC 中检测相应的蛋白质抗原。

这样，ProAb 计划就成为从 EST 到疾病相关性的可能的最直接途径。下面的说明描述了高通量实现这一过程的技术步骤。

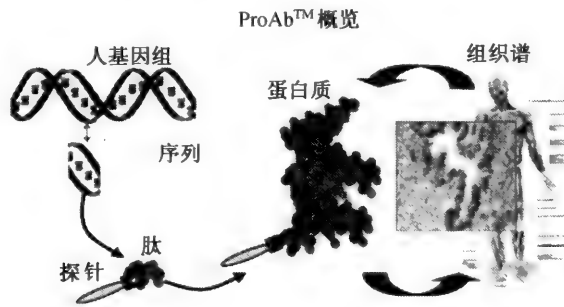


图 12.2 从 ProAbTM 方法到功能基因组的原理

从 EST 数据库中发掘出来的感兴趣的 DNA 片段被翻译成蛋白质。通过化学合成法合成一些肽段来代表这种蛋白质的序列。作为抗原，这些肽段被用于筛选噬菌体抗体。这些被筛选出来的抗体可用作免疫细胞化学检测试剂以显示正常或疾病组织中这种基因表达产物的分布和丰度。

3.2 采用生物信息学方法了解 EST 的意义

EST 是当今最丰富的 DNA 序列资源。如在 dbEST 数据库 (Boguski *et al.*, 1993) 中，有超过 10^6 的 EST 序列，因此，理论上，从统计学的角度看，在 EST 数据库中应该包含有几倍冗余的开放阅读框架。然而，很明显提高这一发现过程的产量应该优先选

择一些 EST 序列进入该程序。最好的策略就是瞄准一种具有令人感兴趣的生物学功能的已知蛋白（并且，最好结构已知），去寻找这一家族的其他成员（见 3.6）。如果已经找到感兴趣的 EST，就应该选择其抗原肽。有效的选择方法能提高成功率。抗原肽的决定取决于设计用于预测抗原肽可能性的算法而得来的权重指数，包括天然样结构，天然蛋白的溶剂可接近程度和抗体易结合的序列特征。生物信息学参与这一过程的第一步，即确定哪一段 EST 序列所翻译的特异肽段能代表这一蛋白质。

3.3 合成抗原

预测的抗原肽序列可以采用自动 96 孔固相多肽合成仪（Advanced Chem Tech）进行合成，此合成仪器采用 fmoc 化学法在微摩尔水平上（5~100mg 产物）进行。这一步，应该寻找那些少数的合成有疑问的肽序列（基于经验）。通常，15 个氨基酸残基的肽段是合适的，但是，事实上，更长的肽（如 30 氨基酸残基）也可以通过自动合成仪高产量的合成。一些翻译后的修饰（如磷酸化），也可以在合成肽中实现。合成结束后，合成的肽被去保护，并被从珠体上切下，然后，通过沉淀法使之与合成试剂分离。洗涤后的沉淀被重新溶解，每一合成肽再通过反相色谱和电喷雾质谱进行质量控制，以确保用作抗原的合成肽纯度大于 85%。将液-质系统与生物信息系统相连，根据预测的肽的分子量检查合成肽的氨基酸序列。合成的抗原肽的 N 端常常为半胱氨酸残基，以便使其可以方便地连接到载体蛋白质上。这样有利于截断的合成产物不会污染抗原肽（肽的合成从 C 末端开始）。N 端的生物素化也是基于同样的原因。载体蛋白常常是牛血清白蛋白，其赖氨酸残基可以被双功能交联剂（succinimide maleimide）所活化。SDS-PAGE 和 MALDI 质谱常用于检查蛋白质的交联效率。一般地，每个蛋白交联 20~30 肽是理想的。值得提及的是，一个商业用途的机器可以合成高质量肽的数量是空前的（每周多达 300 个）。这一成果是通过优化软件、优化交联时间以及精心选择化学方法及高质量的化学合成试剂而取得的。

3.4 高通量筛选识别肽抗原的噬菌体抗体

最基本的筛选方法就是上面所描述的（见 2.4）淘洗。肽-蛋白质交联物被包被到 96 孔微板上，特异抗体的筛选程序如上面所述（见 2.4，2.5），但不同的是高通量筛选过程，包括筛选、结合噬菌体的洗脱、感染、克隆挑选、克隆在 96 孔板中的排列、抗原肽特异抗体的 ELISA 检测等过程全部由商用机器人自动完成。这意味抗体的产生不再是限速步骤。一般地，每个肽序列挑选 12 个抗体克隆。由线性编码，通过生物信息学系统将阳性克隆和 EST 序列联系起来。在抗体选择的最后阶段，ELISA 检测阳性的 *E. coli* 被置于甘油中，在 -70℃ 保存。

3.5 高通量的 ICC 分析

ICC 分析是阐明抗原与疾病相关性的理想技术。与斑点印迹和二维电泳不同，这两项技术必须将组织磨碎并溶解，而 ICC 技术则保持了蛋白质在细胞中的分布。与将所有细胞匀浆而产生的平均效果不同，ICC 技术可以显示蛋白质在细胞水平上的差别。也许有人会问，这样一项能够提供丰富信息的技术为什么没有被广泛采用？问题可能在于高

通量地进行此项技术操作存在明显的困难。当然，要想每月完成 1000 个样品，有许多困难需要克服。在下面一些领域应付出特别的努力。

获取组织 与保存各种不同组织的临床中心保持联系是非常重要的；这些组织在取材之后必须立即进行保存。然而，当组织中目标 mRNA 序列所剩无几时，应时常用蛋白质试剂检测组织中的抗原。

理想的过程 为了提高产量，被仔细挑选出来的各种不同的组织在被切片之前，被切割并被包装起来。除提高速度之外，还增加了过程的一致性，由于确保了一批中不同的组织被进行完全相同的处理，使得不同组织间易于比较。为了保持抗原性，在切法和石蜡包埋切片法时常常优先采用干切法；用冷丙酮的温和固定方法优先于交联固定方法。

在 ICC 中使用噬菌体抗体 在 ICC 中，噬菌体抗体具有较好的灵敏度和空间分辨率。96 孔板中的克隆上清可直接用于标记组织切片。

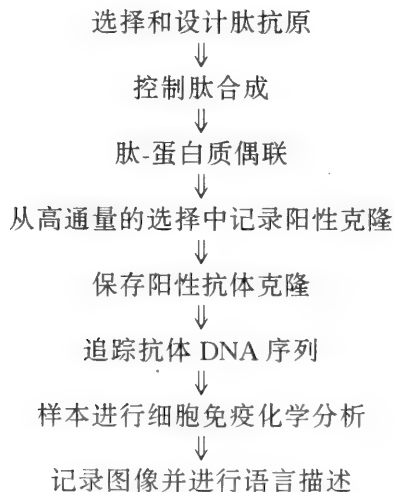
筛选 ICC 样本 当前还没有什么可以替代人类的专门知识来解释染色的图案。团队的方式是必须的。在预筛选中，引起注意的图案被挑选出来，必须再经过有经验的组织化学家作进一步的检查；必须有一个可交互检查解释的可追踪的系统。

记录数据 与显微镜相连的数字相机可以拍摄很好的图像，这些图像被存于数据库中。如果没有专家的解释所拍摄的图像几乎没有什么用处，只有有经验的细胞病理学家才能给出专门的解释。

3.6 生物信息学

我们始终强调生物信息学在 ProAbTM计划的各个阶段都非常重要。没有一个现成的软件是合适的，必须由内部人士自己写专门的软件。能概括所发现的各个方面的要点是非常有用的。

在发现阶段生物信息学的作用 上面我们简要概括了如何利用生物信息学系统在实验策略中的每一个阶段进行控制和检测。下面是流程：



选择和设计肽抗原 利用搜索引擎,由统计学给出的一个相关序列家族的代表序列可被用作搜索 EST 数据库的模板。这是一种灵敏的发现一个家族中新 EST 序列的方法。这种应用就是 Prolemy。

从 ICC 中记录图像和进行语言解释 通过数字相机存储图像。组织病理学家的解释被记录下来,并由声音识别软件进行转录 (Plato),然后被编辑到注释系统中,与数据库中的图像一起存放。

生物信息学和数据分析 ProAbTM计划的产品是一个数据库,在数据库中研究者可以从 EST 序列与细胞病理学的相关模式中获得生物学洞察力。这些信息必须被组织成一个对非专业人士也是一个界面友好的数据库。

应用体制 Java 软件(被称为可移植性软件)确保数据库中各个部分的信息和程序能够在任何硬件平台上进行存取。在可移植性软件所提供的外壳中,各个部分可以通过 CORBA 进行通讯。

用户界面 ProAbTM数据库的界面叫做 Voyager。一般用户可以在 Voyager 界面中通过询问方式获取特定的数据。要想更深入地挖掘数据,必须有一个交互性更好的界面。一种有用的策略是将 ICC 数据还原为真实组织的点印迹。组织染色后常常被分为四个区域,即膜、细胞质、结缔组织和非结缔组织。这些区域表现为不同颜色的点,颜色的深度被分级,反映出组织学家对染色深度的评判。

ProAbTM数据库 在可移植性软件框架内,通过友好的 Voyager 用户界面进行存取是 ProAbTM数据库所特有的。这就是 Oracle 8 数据库,它整合了所有的实验数据,包括数字图像文件、染色图像分析、组织学解释、从 EST 到肽设计的数据追踪、每一特定序列的合成和质量控制数据、抗体分离数据以及与 SWISS-PROT 和 EST 数据库的链接。

3.7 ProAbTM计划的应用实例

在 ProAb 过程中,一些深入了解的蛋白质被用做内对照,这被证实可以有效地鉴别正确的 EST 表达谱。

实例 1 一个 EST 片段 (dbEST number 845032) 被翻译成 VRSSRTPSDKP; 针对相应的合成肽抗原的噬菌体抗体被选择出来; ICC 显示在扁桃体中染色较深 (图 12.3 (a))。通过 Blast 程序搜索 EST 数据库发现与此肽序列匹配的 cDNA 序列来源于人扁桃体细胞文库,通过流式分选证明,此序列在生发中心的 B 细胞中含量丰富。注册人推定此序列为人肿瘤坏死因子前体 (TNF α)。ELISA 检测证明,抗此肽的抗体确实能特异识别人天然 TNF α 。

实例 2 肽序列 ESRKLEKGIQVEIY 被用于代表人内皮细胞黏附分子 (VCAM-1),它对应于从脾 cDNA 文库中得到的 EST 序列-EST20523。此 ESTs 序列被注释为内皮细胞黏附分子-1 类似序列。用选择得到的抗此肽的噬菌体抗体染色扁桃体生发中心,结果与报道的 VCAM-1 在树突状细胞上的分布一致 (图 12.3 (b))。

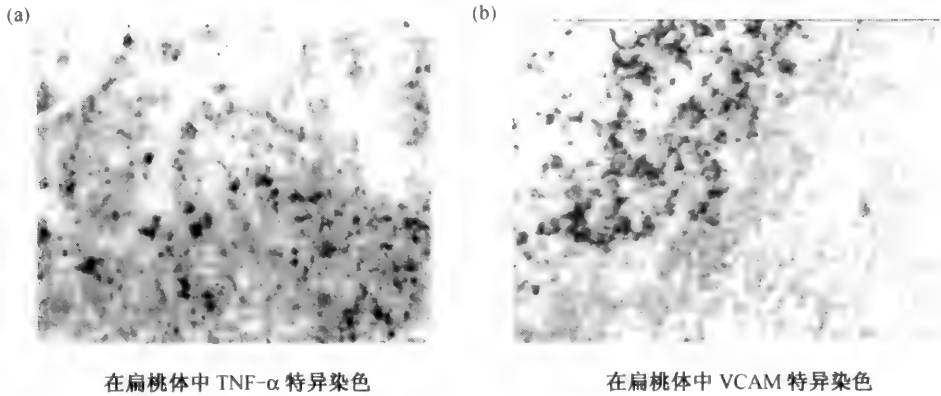


图 12.3 利用 ProAbTM方法得到的代表性结果

通过利用针对代表已知蛋白抗原的多肽片段的抗体在免疫细胞化学中的应用，这一过程得到了证实。免疫细胞组化得到的抗原的组织分布与已有的文献报道的一致。(a) 利用抗多肽 VRSSSRTPSDKP 的噬菌体抗体进行的 TNF- α 特异染色；(b) 利用抗多肽 ESRKLEKGIQVEIY 的噬菌体抗体进行的 VCAM 特异染色

4. 噬菌体抗体在蛋白质组学中的应用——ProxiMolTM技术

4.1 常规的选择抗膜蛋白的噬菌体抗体

不论是用体内方法还是用噬菌体展示技术，筛选针对膜蛋白的抗体都是一种挑战。这主要是因为被用作免疫原的这些蛋白质非常难以纯化。简单的跨膜蛋白常能可溶性表达而被成功用于噬菌体抗体的筛选。对于那些没有可溶性表达形式的膜蛋白，则必须纯化天然的膜蛋白，而众所周知，想在没有一些蛋白质降解的情况下完成纯化很困难。整合膜蛋白，如 7 次跨膜的趋化因子受体，则代表了另一种难度：它们不可能被以可溶性的形式表达，不论是天然的还是重组的，要纯化一定数量的此种蛋白且不变性很难。而当前的方法认为至少蛋白质的生物化学信息是可用的，而不是蛋白质本身。

利用未纯化的抗原混合物来制备抗体有时是可能的，并且常常有人尝试。完整的细胞可以被直接用作免疫原，这样问题就转移到下游的筛选上。必须筛选大量的抗体，而不能保证其目的抗原特异性，因为这些抗体是被随机选择出来的。一种提高针对细胞表面特异分子导向性的方法是用一般细胞表面蛋白进行反筛。这可以通过采用将少量表达所希望细胞表面蛋白的细胞与大量不表达所希望细胞表面蛋白的细胞混合来实现。通过加一种细胞特异的抗体将感兴趣的目的细胞进行标记。将噬菌体抗体库加入细胞混合物，靶细胞将结合一些噬菌体，然后，采用细胞特异的抗体将靶细胞从细胞混合物中分离出来。这种方法采用靶细胞群的特异抗体，当然，最好是被用于那些已经进行了很好的细胞亚群分类的细胞，如血细胞。

4.2 提高膜蛋白筛选的导向选择

如果噬菌体抗体库中的抗体可被导向那些感兴趣的蛋白，那么，筛选针对那些只能

在混合物中存在的蛋白质的抗体，成功的机会更大，特别是细胞膜蛋白，如受体。一种理想的方式就是采用受体的同种配体，同时也是一种证明靶分子具有活性构象的方法。结合到导向分子附近的噬菌体可以被分离出来，从而富集那些被选择出来的带有特异识别靶受体的噬菌体群体。这就是 ProxiMol™ 的方法原理。

4.3 ProxiMol™ 方法及应用

酶催化沉淀报告法 在 ICC 和一些其他的技术中，常常通过报告酶催化产生的沉淀物将信号进行放大 (Bobrow *et al.*, 1989, 1992)。辣根过氧化物酶 (HRP) 试剂被直接或间接地结合到靶抗原上。在过氧化氢和生物素酪胺存在下，HRP 将催化产生短命的生物素酪胺自由基，它将共价结合并生物素化靠近辣根过氧化物酶的蛋白质。沉积于蛋白质上的生物素分子可通过链亲和素偶联试剂检测，如辣根过氧化物酶，这样初始的信号被放大，因为新沉淀的生物素分子远超过了原始靶分子或辣根过氧化物酶分子的数目。

ProxiMol™ 方法的原理 在 ProxiMol™ 方法中，催化报告沉淀法被用于分离靠近生物素化导向分子的噬菌体抗体 (Osborn *et al.*, 1998)。接触导向分子有两个条件。第一，它必须与靶分子具有一定的亲和力或是与靠近靶分子的分子有亲和力。第二，它必须能够直接或间接偶联 HRP。间接偶联 HRP 包括将偶联 HRP 的抗体结合到导向分子上，或是生物素化导向分子，然后将导向分子结合到链亲和素化的 HRP 上。导向分子必须保持足够的生物学活性，以保持它与靶分子的结合。然而，与导向配体竞争结合受体的噬菌体抗体不能被附着在偶联于配体上的 HRP 附近。

实践中的 ProxiMol™ 方法 用 ProxiMol™ 方法进行针对原位靶抗原的筛选很简单 (图 12.4)。对于细胞表面抗原，导向分子 (通常是生物素化的靶分子配体或是偶联 HRP 的抗体) 与人抗体库中的噬菌体一起被加入到完整的细胞中。温育之后，没有结合的噬菌体被洗掉。如果需要，可加入 HRP 标记的第二种试剂，温育，然后洗涤。加入生物素酪胺和过氧化氢，温育 10min，然后再洗涤。通常用 TEA 将结合的噬菌体洗脱下来，生物素化的噬菌体通过链亲和素磁珠进行分离，然后，感染大肠杆菌。通过筛选，选择出那些特异结合靶分子的噬菌体抗体及结合靶分子附近蛋白质分子或导向分子的噬菌体。

ProxiMol™ 方法的应用范围 ProxiMol™ 方法可被应用于各种不同的靶分子及不同形式的蛋白质。导向分子除抗体或天然的蛋白质配体如细胞因子外，还包括糖类和脂肪酸配体。存在于完整的细胞上 (附着的或可溶的)，冷冻的组织切片中或分离的细胞膜上的靶抗原不需要进行蛋白质的分离。

生物素沉淀的范围 由于生物素酪胺自由基的半衰期很短，生物素分子被沉淀在附近的 HRP 分子上。为了了解生物素沉淀的范围，抗基因 3 的抗体被用作导向分子去生物素化 M13 噬菌体。用 5nm 的链亲和素包被的金珠，通过电子显微镜探测了沉淀的生物素分子。生物素沉淀的范围可达到 25nm，平均在 10~15nm。这使我们相信生物素沉淀的范围在几个导向分子直径范围内。

ProxiMol™ 技术用于抗原表位步移：扩大结合 TGF- β 克隆库 以前选择到的针对 TGF- β 的主要抗原表位或克隆被称为 31G9。ProxiMol™ 方法被用于扩大抗 TGF- β 的噬

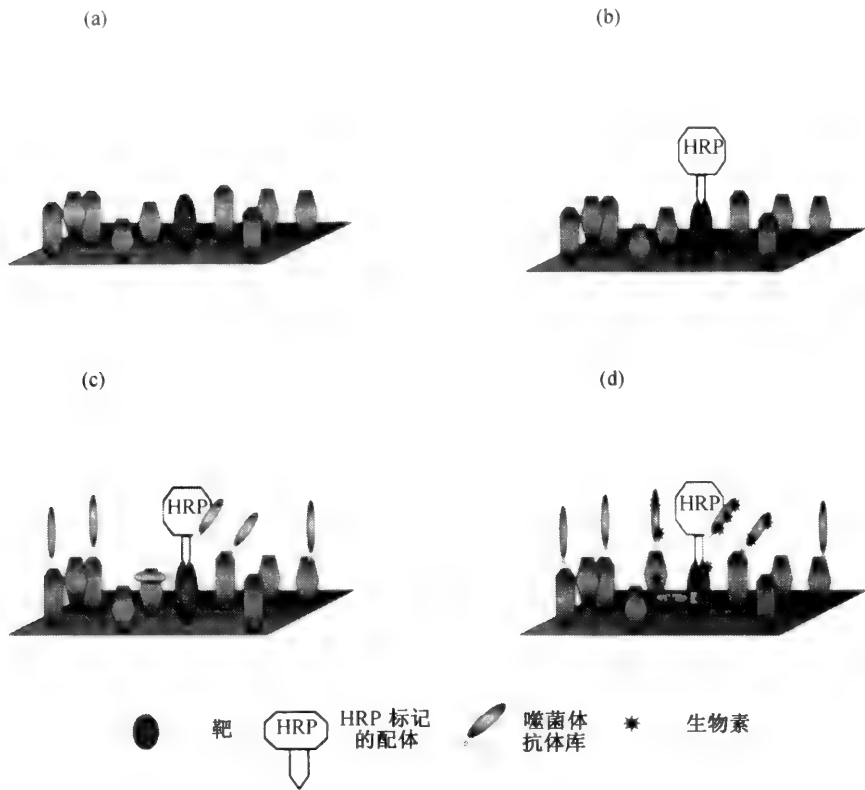


图 12.4 导向筛选噬菌体抗体的 ProxiMol™方法

(a) 在细胞表面处于正常环境条件中的感兴趣抗原；(b) 导向配体被偶联了 HRP，并特异结合到其相应的靶蛋白上；(c) 加入噬菌体抗体库，不同的抗体分别特异结合到导向配体、靶蛋白及其他的细胞表面抗原上；(d) 加入生物素化的酪胺试剂，产生活化生物素；结合到导向配体周围 25nm 范围内的噬菌体被共价标记上生物素。这些识别靶蛋白及其周围蛋白的噬菌体抗体可通过包被链亲和素的磁珠进行分离。

菌体抗体库，导向新的抗原表位。采用 ProxiMol™方法进行的选择在固定有 TGF-β 的 BIAcore 芯片上进行；导向分子为 HRP 偶联的 31G9。噬菌体的结合和非结合噬菌体的洗脱如前所述，然后，加入过氧化氢和生物素酪胺。被生物素化的噬菌体被分离出来，然后侵染大肠杆菌。经过两轮的筛选，分离出来的 13.5% 的克隆是 TGF-β 阳性克隆，并且均为新克隆。与此对照，在没有导向分子存在下筛选到的所有阳性克隆均为 31G9 (Osborn *et al.*, 1998)。因此，ProxiMol™方法通过对非主要抗原表位的噬菌体抗体的筛选，扩大了针对纯化抗原的抗体库。

ProxiMol™方法用于筛选针对细胞表面特异黏附分子的抗体 ProxiMol™技术最大的应用是在筛选针对细胞表面特异分子的抗体方面。为了靶向选择素（表达于活化的内皮细胞表面的黏附分子），采用了选择素的配体和抗体。

用 TNF-α 活化人脐静脉内皮细胞，以诱导细胞表达 P 选择素和 E 选择素。采用 P 选择素和 E 选择素的配体 sialyl Lewis X 作为导向分子，利用 ProxiMol™方法，在完整的单层培养的细胞上进行筛选。经过两轮的筛选，13.7% 的克隆为 E 选择素阳性克隆，

6.3%的克隆为P选择素阳性克隆 (Osborn *et al.*, 1998)。与此对照, 用HRP标记的抗E选择素的抗体作为导向分子, 只筛选到E选择素特异的噬菌体。因此, 通过配体特异的导向筛选, 有可能在完整细胞上进行原位天然膜结合抗原的抗体筛选。

采用 ProxiMol™方法获得抗细胞表面 7-TM 受体的抗体 深埋于细胞膜中的穿膜蛋白, 如 7 次穿膜的蛋白 (7TM), 很难以天然状态进行纯化, 它们是原位筛选的理想靶标。趋化因子 CC5R 表达于 T 淋巴细胞上, 并被认为是 HIV 感染的辅助受体。利用生物素化的 CC5R 的配体 MIP-1 α 作为导向分子, 采用 ProxiMol™方法对人 T 淋巴细胞进行筛选。将生物素化的 MIP-1 α 和噬菌体与细胞进行温育, 然后, 加入链亲和素标记的 HRP, 采用上面所述的 ProxiMol™方法进行筛选。第一轮筛选的产物通过基于完整细胞及蛋白质的 ELISA 方法进行了分析。随机挑选的 95 个克隆中, 30 个克隆可以识别 T 细胞, 其中, 13 个克隆结合 CCR5, 剩下的 17 个克隆结合 T 细胞表面的与 CCR5 相关

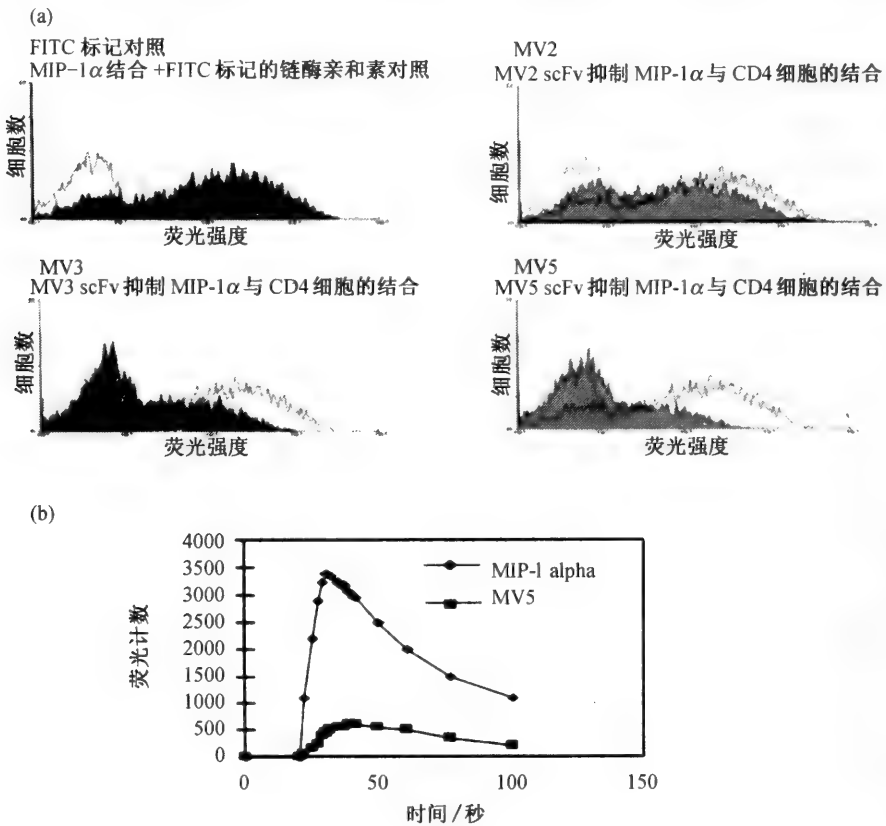


图 12.5 通过回筛 ProxiMol™方法得到的活性抗体

(a) 通过流式细胞术, 回筛得到的 scFv 能抑制 MIP-1 α 结合到 CD4 阳性细胞上。每个峰的阴影部分代表 MIP-1 α 结合到细胞上 (将生物素化的 MIP-1 α 加入细胞中, 然后加入 FITC 标记的链亲和素)。每个峰的非阴影部分代表不加生物素标记的 MIP-1 α , 只加 FITC 标记的链亲和素得到的信号。与对照 (上面左图), 三种 scFv-MV2, MV3 和 MV5 均能抑制 MIP-1 α 的结合; (b) scFv-MV5 封闭了由 MIP-1 α 结合到其受体 CCR5 上而产生的 Ca^{2+} 信号; 将 50nmMIP-1 α 加入 CCR5 转染的 CHO 细胞中, 利用 2-呋喃的荧光, 通过标准方法检测细胞内自由 Ca^{2+} 浓度的升高。

的或附近的分子表位。这些抗体被用于筛选 cDNA 表达文库，得到了几个潜在的令人感兴趣的蛋白质，其中包括磷酸酪氨酸磷酸酶。因此，ProxiMol™方法不仅可使我们定向特异的细胞表面受体，而且，还能用于鉴定一些可能与此受体相关的蛋白质。

回筛 在 ProxiMol™筛选方法中，用受体的配体作为导向分子的筛选，只能筛选到那些与配体识别不同位点的抗体，这些抗体不能中和配体与受体的相互作用。为了克服这种局限性，采用两步筛选法。以受体的配体分子为导向分子，按常规 ProxiMol™方法进行筛选，生物素化的噬菌体被分离出来。这些被分离出来的噬菌体不是被用于感染大肠杆菌，而是被用于利用新的抗体库进行第二轮 ProxiMol™筛选的导向分子。这使得那些与原始指导分子/配体识别相同位点的噬菌体被分离出来。当然，这必将产生各种不同的产物，因为那些与第一轮全部筛选产物所识别的蛋白质附近所有结合的噬菌体均被分离出来。先检测第一轮筛选产物，鉴别出受体特异的噬菌体，然后，用这些特异的噬菌体作为指导分子进行第二轮的回筛，可以增加筛选的特异性。

CCR5 的回筛策略是第一轮的 ProxiMol™筛选，采用 MIP-1 α 作为导向分子，筛选的产物作为回筛的导向分子。那些识别 CCR5 的克隆被通过 ELISA 方法鉴别出来。利用生物素化的 MIP-1 α 和 scFv，通过流式细胞仪将抑制 MIP-1 α 结合的噬菌体克隆鉴别出来（图 12.5a）。其中，一个具有抑制 MIP-1 α 结合活性的克隆同时能抑制 MIP-1 α 诱导的 CCR5 表达细胞的钙离子流（图 12.5b）。因此，经过简单的两步筛选策略，特异结合细胞表面 7-TM 受体的配体结合位点的抗体被分离出来。

4.4 ProxiMol™方法在蛋白质组学研究中的应用

来源于 ProAb™的抗体作标签指导新的筛选 在 ProAb™筛选中得到的令人感兴趣的噬菌体抗体不可能在生物学分析中表现出活性。但这些抗体是疾病特异的抗原或表位的有用标签，可被用于在 ProxiMol™筛选中的导向分子。因为 ProAb™数据库已显示了噬菌体抗体的组织特异性，合适的组织切片就可成为 ProAb™筛选的靶抗原的来源。作为导向分子，最简单的就是生物素化的 ProAb™噬菌体抗体，还有一种就是 scFv。

筛选 ProxiMol™方法得到的产物，可以找到满足所期望生物学活性的抗体。低通量的染色筛选是在组织切片上进行；然而，必须鉴别与原始 ProAb™抗体染色图案匹配的模式。高通量的筛选策略是利用 ELISA 方法，或是用对特定细胞的流式细胞术。这种筛选方法也可以用特定激活状态的细胞，这样有可能深入了解在原始组织切片中细胞的状态。噬菌体或 scFv 可直接用于离体的细胞分析。这样，ProAb™抗体就可被用于分离具有生物学活性的抗体，并有可能被用于界定靶细胞的活化状态。

用 ProxiMol™技术鉴定新的蛋白质成员或蛋白质/DNA 复合物 在 ProxiMol™技术中的生物素标记步骤中，生物素沉淀的范围在 3~4 个蛋白质分子直径范围内。这样 ProxiMol™技术将产生结合靶分子附近其他蛋白质的反应物。对 CCR5 的工作就是一个很好的例子；当 MIP- α 被用作导向分子时，得到了一些并不针对 CCR5 的噬菌体抗体。这其中的一些抗体被用于筛选表达文库，发现这些抗体所识别的一些蛋白质，它们被认为存在于膜中，靠近 CCR5。这导致发现了一些可能令人感兴趣的蛋白质，包括磷酸酪氨酸磷酸酶。因此，ProxiMol™技术不仅使我们能定向细胞表面特异受体，而且，还可使我们鉴定可能与受体相关的其他蛋白质。这导致了一种可能性，那就是通过连续几轮

的 ProxiMolTM 筛选和检测, 将可描绘出蛋白质多分子复合物的轮廓。

致谢

上面所描述的工作依赖于团体的努力, 这些努力远不止来自于剑桥抗体技术方面的科学家, 还要感谢以下个人对著述本书所提供的专门技术。他们是 Simon Brocklehurst 及生物信息小组的全体成员, 在肽合成方面作出杰出成就的 George Masih, John McCafferty, Tony Pope 及 MAb 分离小组的全体成员, 以及在 ICC 方面提供专门指导的 Tony Warford。

(王清明 译 贺福初 校)

参考文献

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F. *et al.* (1991) Complementary DNA sequencing, expressed sequence tags and human genome project. *Science* **252**: 1651–1656.
- Barbas, C.F., Kang, A.S., Lerner, R.A. and Benkovic, S.J. (1991) Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proc. Natl Acad. Sci. USA* **88**: 7978–7982.
- Bobrow, M.N., Harris, T.D., Shaughnessy, K.J. and Litt, G.J. (1989) Catalyzed reporter deposition, a novel method of signal transduction. Application to immunoassays. *J. Immunol. Methods* **125**: 279–285.
- Bobrow, M.N., Litt, G.J., Shaughnessy, K.J., Mayer, P.C. and Conlon, J. (1992) The use of catalysed reporter deposition as a means of signal amplification in a variety of formats. *J. Immunol. Methods* **150**: 145–149.
- Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) dbEST – database for “expressed sequence tags”. *Nature Genet.* **4**: 332–333.
- Foote, J. and Eisen, H.N. (1995) Kinetic and affinity limits on antibodies produced during immune responses. *Proc. Natl Acad. Sci. USA* **92**: 1254–1256.
- Griffiths, A.D., Malmqvist, M., Marks, J.D., Bye, J.M., Embleton, M.J., McCafferty, J., Baier, M., Holliger, K.P., Gorick, B.D., Hughes-Jones, N.C., Hoogenboom, H.R. and Winter, G. (1993) Human anti-self antibodies with high specificity from phage display libraries. *EMBO J.* **12**: 725–734.
- Kay, B., Winter, J. and McCafferty, J. (1996) *Phage Display of Peptides and Proteins: a Laboratory Manual*. Academic Press, London.
- Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D. and Winter, G. (1991) By-passing immunization: human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**: 581–597.
- McCafferty, J., Griffiths, A.D., Winter, G. and Chiswell, D.J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**: 552–554.
- McCafferty, J., FitzGerald, K.J., Earnshaw, J., Chiswell, D.J., Link, J., Smith, R. and Kenten, J. (1994) Selection and rapid purification of murine antibody fragments that bind a transition-state analog by phage display. *Appl. Biochem. Biotechnol.* **47**: 157–173.
- McCafferty, J., Hoogenboom, H.R. and Chiswell, D.J. (1996) *Antibody Engineering – a Practical Approach*. IRL Press, Oxford.
- Osborn, J.K., Derbyshire, E.J., Vaughan, T.J., Field, A.W. and Johnson, K.S. (1998) Pathfinder (ProxiMol) selection: *in situ* isolation of novel antibodies. *Immunotechnology* **3**: 293–302.
- Perelson, A.S. and Oster, G.F. (1979) Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self non-self discrimination. *J. Theor. Biol.* **81**: 645–670.
- Vaughan, T.J., Williams, A.W., Pritchard, K., Osborn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R., Wilton, J. and Johnson, K.S. (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage-display library. *Nature Biotechnol.* **14**: 309–314.
- Winter, G. and Milstein, C. (1991) Man-made antibodies. *Nature* **349**: 293–299.
- Yanisch, P.C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103–119.

第十三章 糖生物学与蛋白质组学

Nicolle H. Packer and Lucy Keatinge

1. 引言

蛋白质组学的发展对各种生物体系的研究已经并将产生意义深远的影响，二维凝胶电泳作为探索细胞中蛋白质组分变化的平行效应的一种有效方法已经形成。目前，从胶及膜上快速、准确以高通量方式在 fmol 数量级上鉴定及定性蛋白质已成为可能，但是当这种方法被延伸至分析细胞中糖蛋白时，结果并不令人乐观。真核细胞的基因组与蛋白质组间的特征区别之一就是绝大多数基因产品是经过翻译后修饰的，其中糖基化是最重要的修饰方式。从已知的 DNA 序列导出蛋白质的氨基酸序列，并以此来确定其特殊的功能通常是不够的，这一点已日趋明朗，发生在蛋白表达之中或之后的修饰对蛋白质的功能影响较大，因此，分析这些修饰对不同蛋白表达模型的理解是至关重要的，目前，这一需求可以通过蛋白质组技术加以实现。

我们所观察到的糖基化变化是糖基转移酶活性的反映，糖基转移酶合成寡糖并将其附着于蛋白质，糖苷酶修整及降解多糖。糖蛋白质组学研究的目的是鉴定作为特定表型标志的糖蛋白，鉴定与表型变化相关的特定的糖基化的改变，以及追踪与这些变化相对应的代谢过程。

比较考马斯亮蓝染色及一种新的荧光糖蛋白染色技术用二维凝胶电泳分离的人血清蛋白（图 13.1），我们发现许多丰度高的蛋白是糖基化的，并且依糖基形式不同它们的等电点及分子量不同。普遍认为我们应该研究为什么在蛋白质的糖基化过程中有如此多的不均一性以及对于这些不均一性必须加以分析。尽管如此，相对而言在蛋白质组学研究中，用上述提到的平行进程的方法分析蛋白质中的糖组分这一领域并未得到足够的重视。

我们相信，在细胞中糖蛋白组分研究方面的延误有两个原因：第一，尽管有大量文献报导在疾病的发生、发展过程中糖基化的变化，我们也只是停留在对这些糖基化变化如何对有机体产生生物学效应方面的了解。其次，研究多糖的分析方法传统地被认为是基于糖化学家的实验室工作，这就造成了一种印象，即分析糖蛋白需要蛋白质化学家们所难以胜任的特殊知识与技术。本章将在蛋白质组学的论述中阐述这两个问题。

为了建立有机体新陈代谢中糖基化的规律，仅仅对蛋白质中多糖结构定性是不够的。因此，尽管分析蛋白质的糖基化变化有利于新陈代谢变化（如疾病的发生）的诊断和监测，但这些变化并不足以描述糖基赋予蛋白质的功能。此外，糖链的丢失对蛋白质活性的影响并不总是可以观察到的，一些重组糖蛋白在其天然糖基化被丢失后仍具有活性。普遍认为，在许多情况下，特定的糖基化对于蛋白的正确折叠及作用是至关重要的；或者，由于糖基化的多样性使蛋白质的功能表现为不规律性，二者必具其一。Var-ki (1993) 的综述对这些明显矛盾的观点予以了论述。

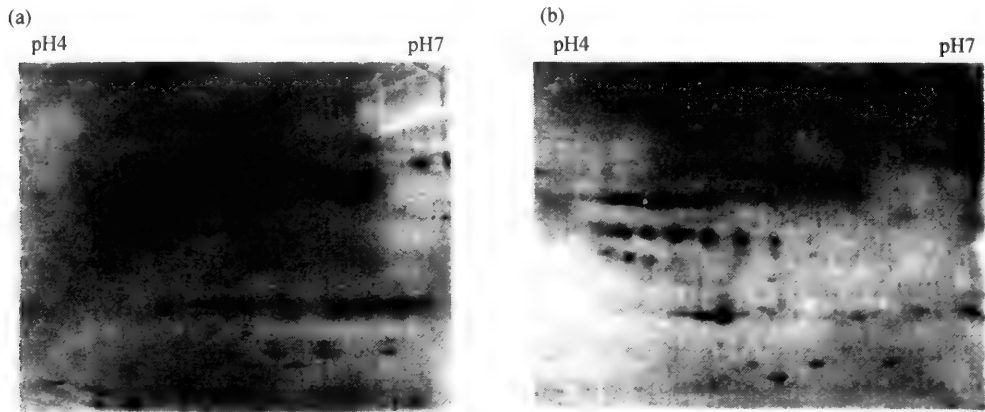


图 13.1 PVDF 膜上糖蛋白的荧光染色法

(a) 考马斯蓝染色；(b) 糖荧光素染色。1 μ l 血清样品经重泡胀上样，经 2-DE 分离在 PVDF 膜上显出斑点，用 Bio-Red 糖蛋白检测试剂盒，即高碘酸盐氧化糖的荧光染色法代替夹层抗体染色法，改进方法为用 0.05mg/ml 氨基脲在 50% (V/V) 甲醇，100mM 醋酸钠溶液。

糖蛋白糖基组分固有的多样性能导致蛋白质在形状、电荷及体积等方面发生微妙的变化，这种变化有可能暂时性或者在空间上修饰蛋白质的功能。这种多样性需要有相关复杂的物理及化学技术对其组成、连接、分支、异构体及连接位点等进行确定。由于仅限于单糖残基结构及多糖连接方式的研究，糖蛋白的分析过于简单化，通常碳水化合物分析化学被发展成为复杂多糖的结构测定，蛋白质中多糖的分析就是从这些技术中选择性地开发的。这就促使一些商业性的分析多糖结构的工具出现 (Oxford GlycoSciences, UK; Glyko Inc, USA; Bio-Rad, USA; Beckman, USA)。综合糖苷外切酶、商业化的生物毛细管电泳及质谱，目前已能够定性分析蛋白质中多糖的微不均一性，而蛋白质组技术接受了综合这些技术从而适应大量糖蛋白分析的挑战。

本章我们给出一些领域中的例子，说明糖基化有其重要的规律，同时可以有力地使读者确信蛋白质糖基形式的分析是理解细胞蛋白质组的基础。下面我们描述糖分析技术，它使适应糖蛋白质组研究的高通量分析成为可能。

2. 什么是糖蛋白？

糖蛋白是由酶将一个或多个糖元加合在最初的多肽链上，修饰范围从蛋白质上连接单糖 (比如胶原糖蛋白及核糖蛋白)；到密集、复杂、有侧链结构的多糖 (如黏液素)；以及具有复杂的氨基葡聚糖侧链的蛋白多糖。多糖与蛋白质相联有三种形式：N-连接、O-连接及糖基磷酸酯酞基醇固定结构。在同样一个蛋白质上可以发现这三种修饰方式共存。具有相同氨基酸序列但不同糖基侧链的蛋白质异构形式通常可以在二维凝胶电泳分离上看到，由于它们的等电点及分子质量不同表现为斑点的拖尾。

2.1 N-糖链结构

N-糖链结构通常发生于内质网及高尔基体上，在这里 $\text{GlcNAc}_2\text{Man}_9\text{Glc}_3$ 为普遍结

构, 在此基础上, 由糖苷酶及糖基转移酶将其进行多样化整理和修饰, 几乎所有的 N-连接多糖内侧都具有 $\text{GlcNAc}_2\text{Man}_3$ 的核心结构, 它与出现在特定序列 Asn-Xaa-Ser/Thr (Xaa 是除 Pro 以外的任何氨基酸) 中的天冬酰胺残基相连接。按照单糖的位置可将 N-连接结构分为高甘露糖型、混合型或复杂型多糖。不同的糖链可连于同一位点 (微不均一性), 同时在同一蛋白质上也可有不同的位点连接糖链 (显著不均一性)。

所有定性的糖蛋白都具有相同的 N-连接五糖核心, 与此同时多糖链内或链间的整理与修复使它在位置、侧链、大小及异构配置上发生变化。

2.2 O-连接结构

O-连接寡糖通常与丝氨酸或苏氨酸的羟基部分连接, 有时也与羟赖氨酸及羟脯氨酸连接。尽管 Hansen 等 (1998) 已经建立了一种神经网络可以预测黏蛋白型糖基化, 但并无一个简单的规律认为含羟基残基可以作为连接位点。黏蛋白型 O-连接糖基化最初是由完全折叠的蛋白质中丝氨酸或苏氨酸与 GalNAc 残基连接, 这是由糖蛋白的一级、二级甚至三级结构所决定的, 由特殊转移酶来实现的单糖残基继续延伸是被高度控制的 (Van den Steen *et al.*, 1998a), 与较普通的黏蛋白型 O-连接糖基化相比, 一些特殊的 O-糖基化形式, 如 O-连接的岩藻糖及葡萄糖, 是依赖于蛋白质的氨基酸序列以及八个不同的还原型末端与如前所述的丝氨酸与苏氨酸相连 (Hounsell *et al.*, 1996)。多数哺乳动物的分泌蛋白上可发现 O-糖基化, 但是单一的 GlcNAc 组分在细胞质及核膜孔蛋白 (Hart *et al.*, 1996) 上发现, 而糖部分与蛋白质之间的磷酸二酯连接方式在黏物质及一些寄生动物或者更为广泛的一些生物体上可以见到 (Hanynes, 1998)。

2.3 糖基磷脂酰肌醇 (GPI) 固定结构

糖基磷脂酰肌醇结构具有将细胞膜的疏水性磷酸双分子层固定于不同蛋白质的功能 (Kinoshita *et al.*, 1997)。它们具有相同的 $\text{Man}_3\text{GlcNH}_2$ 糖核, 它是细胞膜内磷脂与蛋白质之间的桥梁, 这一翻译后修饰行为发生在内质网中信号肽 C 末端切割之后。

3. 为什么蛋白质组对糖蛋白感兴趣?

蛋白质组学是一个以蛋白质为基础的整体生物学方法, 这种方法是对 DNA 及 mRNA 的表达产物而不是 DNA 和 mRNA 本身进行分析。绝大多数情况下, 一个成熟蛋白的生物学活性是被翻译后修饰的, 而这些修饰又是高度动态的。对哺乳动物来说, 蛋白质的多糖组分的作用包括免疫反应的修饰、蛋白识别及定位、蛋白质折叠、血清清除率及信号调控。讨论碳水化合物体内功能的详细综述见文献 (Dwek, 1995; Kukuruzinska and Lennon, 1998; Lis and Sharon, 1993; Varki, 1993)。

人类生物学最感兴趣的实验体系包括在受精及疾病的发生过程中糖蛋白的涉入。我们在此对这两个例子进行讨论目的在于揭示理解那些被表达的基因产品, 其翻译后修饰功能的重要性。

3.1 糖蛋白与受精

对于动物体, 从卵母细胞与精子细胞之间的识别开始, 到一个成熟的有机体发展到

顶点, 在此进程中糖蛋白起着至关重要的作用。对于哺乳动物的受精, 包围在卵子周围的细胞外糖蛋白外衣 [透明带 (ZP)] 是关键部分。在渗透过滤泡细胞之后, 精子捆绑于透明带 (ZP), 它是由三种硫酸化糖蛋白 ZP1、ZP2 及 ZP3 组成的, 这些糖蛋白对识别与捆绑是关键性的 (Zara and Naz, 1998)。在精子表面的 β -1, 4-半乳糖转移酶是研究卵壳 ZP3 的最佳受体。对于转基因小鼠, 这种酶的过表达可在许多方面对精-卵相互作用产生影响 (Youakim *et al.*, 1994), 此外, 在 β -1, 4-半乳糖转移酶基因丢失中幸存下来的一些小鼠的精子, 仅能艰难地渗透过透明带, 在体外不能与卵母细胞受精 (Lu and Shur, 1997)。在避孕与受精的研究中, 糖蛋白质组学方法可以监测和定位受精过程中糖蛋白的变化。在避孕研究中令人感兴趣的一种糖蛋白是调节黄体酮的子宫 glycodelins, 它的氨基酸序列与 β -乳球蛋白完全相同, 所不同的只是其独一无二的多糖结构。源于子宫内膜的 glycodelin A 阻止精卵结合, 同时, 与其糖基化不同的源于血浆的精液中 glycodelin S 却无此功能 (Seppala *et al.*, 1998)。Glycodelin A 与 S 的基因相同 (胎盘蛋白 14), 但是由于翻译后糖基化的不同导致了蛋白质的功能及组织定位不同。

3.2 糖蛋白与疾病

在疾病的最初阶段, 糖蛋白通常在结构与功能上显示出变化, 而这些变化涉及到与蛋白相连的多糖结构。这种变化能够描绘疾病的进程, 在这一领域有大量文献及两本书发表 (Brockhausen and Kuhns, 1997; Montreuil *et al.*, 1996), 其中描述了糖蛋白在人类疾病进程中的作用。最近 Brockhausen 等 (1998) 的综述讨论了在人类疾病过程中糖基转移酶缺失情况下这些糖蛋白交互作用的机理。我们所缺乏的是对于疾病相关或疾病导致的糖基化变化与真正引起疾病的糖基化变化二者之间差异的理解。很清楚, 后者是我们感兴趣的, 因为它们干涉治疗的位点。

蛋白质组方法测定包括疾病进程初期, 疾病与正常组织之间所表现出来的差异蛋白质的鉴定。这些蛋白以后可作为诊断及控制疾病的靶点。许多糖蛋白, 特别是在糖基化方面发生变化的糖蛋白已经用于反应人类疾病的过程 (表 13.1)。表 13.1 的例举虽然并不完全, 但从中也可以看到一些异常新陈代谢与糖基化变化之间的联系。由于血清通常用作临床诊断的对象, 同时易于获得, 因此, 有许多报导是关于在疾病过程中, 血清中糖蛋白的糖基化变化的, 我们推断在固体病变组织中表现出的糖基化变化也会有相类似的差异性。

糖基化差异也被用于构建多糖类癌症疫苗。例如, 正常上皮细胞上的黏蛋白表面就密集地包围着混合型多糖结构, 在癌细胞上, 隐藏的糖基抗原决定基 (Tn 及唾液酸化 Tn 抗原) 暴露于黏蛋白上, 同时, 成为了抗癌治疗疫苗的靶点。这些以特殊的多糖与合成多肽结合, 用于免疫治疗的不同策略已在临床试用 (Kudryashov *et al.*, 1998; Lo-Man *et al.*, 1999), 同时, 它已经成为这一领域的研究兴奋点。

普遍认为二维凝胶电泳是展示蛋白质差异表达的首选方法。许多蛋白质被同时展示, 而所观察到的差异可与其表型差异相关联。蛋白质组并不需要以显示单个蛋白为目的, 而是要观察对任何所选择的生物体系的网络途径的影响, 所以蛋白质组要求对多种种群群体成员进行筛选, 以揭示那些发生在蛋白表达中与个体水平相类似的群体水平上的变化, 如果一个糖蛋白对表型变化有影响, 那么它可以用作疾病的诊断标记; 或者直

接采用凝胶电泳分析方法；或者利用其他分析方法对其进行分析，在对成骨细胞原始细胞系分泌的细胞外胶原酶（Hankey *et al.*, 1992）、糖缺乏综合症的多重血清蛋白（Yuasa *et al.*, 1995）及胰腺癌病人的血清及黏蛋白的分析中，2-DE 已经成功地用于鉴定糖蛋白差异，已报导众多的血清糖蛋白做为肾细胞癌的标记物（Delahunt *et al.*, 1996）。

表 13.1 糖基化类型的变化与人类疾病的关系

疾病	糖蛋白	组织	变化	参考文献
碳水化合物缺乏症	转铁蛋白, 抗凝血酶, α -酸性蛋白	血清	N-连接糖蛋白形式不同	Winchester <i>et al.</i> (1995)
肝癌	α -胎球蛋白	血清	N-连接糖蛋白形式不同	Seregin <i>et al.</i> (1995) Taketa (1998)
炎症, 癌, 爱滋病	α 1-酸性糖蛋白	血清	N-连接糖蛋白形式不同	Van Dijk <i>et al.</i> (1995); Mackiewicz and Mackiewicz (1995)
免疫紊乱	CD43	T 细胞	O-连接糖蛋白形式不同	Ellies <i>et al.</i> (1996)
类风湿关节炎	IgG	血清	N-连接端基半乳糖降低	Delves (1998)
Creutzfeldt-Jakob 症	朊蛋白	CSF	糖基化程度不同	Furakawa <i>et al.</i> (1998)
血吸虫病	循环阴/阳抗原	血清, 尿	O-连接糖蛋白形式不同	Van Dam <i>et al.</i> (1996)
糖尿病妊娠, 绒毛膜癌	hCG	尿	N-及 O-连接糖链过多分支	Elliott <i>et al.</i> (1997)
酗酒	转铁蛋白	血清	脱唾液酸	Tagliaro <i>et al.</i> (1998)
癌症	黏蛋白	肿瘤	特殊短 O-连接多糖的过表达或暴露	Taylor-Papadi Mitriou and Epenetos (1994); Kim <i>et al.</i> (1996)
溶菌酶储藏症	溶菌酶类多糖	尿	异常碳水化合物的分泌	Starr <i>et al.</i> (1994)
绝经	FSH, LH	血清	酸性增加, 复杂性降低	Anobile <i>et al.</i> (1998)

从表 13.1 可以看出，许多发生在目标蛋白糖基化中的变化，有些是由于糖基化程度不同而导致分子量的变化，有的是由于相应的唾液酸化而引起等电点变化，同时，多糖的磷酸化及硫酸化也会使糖蛋白的等电点发生变化。糖蛋白的这些固有属性使它们在 2-DE 上理想地排列，这一阵列非常适合检测糖基化过程中的特征性变化，目前还没有其他技术能够以这种形式展示糖蛋白的不同。

4. 蛋白质组中的糖分析

糖分析最初是以复杂的、大分子质量的多糖的化学及物理结构测定为焦点的。这些技术从前非常适于对于那些分子量较小，结构较简单的多糖的分析，而至今已逐渐成为蛋白质生物化学家手中的一种更快、更简易的分析方法。能够完成小分子多糖及糖肽的质量分析的质谱仪的到来，加之许多专一的外切糖苷酶的克隆，促使我们能够分析许多糖蛋白上寡糖链的位点、连接方式及序列（Packer and Harrison, 1998），由于经纯化过的被表达的重组糖蛋白在样品量的提供方面占有优势，所以通过对它们的分析，使得绝大多数分析方法得以优化。（10nmol 或 100 μ g 蛋白量）相对而言，分析 2-DE 分离的糖蛋白所需的样品量日益减小（通常 10 pmol 或 100ng）。

4.1 重组糖蛋白分析

由重组 DNA 生产蛋白质的方法是生物技术产业中日益增长的领域。许多治疗及诊断蛋白质都是糖基化的，因此糖基化形式的快速分离及鉴定就十分必要，由于多糖链的变化能够诱导不良的免疫学反应，导致越来越多的官方法规要求对糖蛋白的多糖部分进行严格的测定。

对于细胞系和培养条件的依赖，使表达体系不可能完全控制蛋白的糖基化，从而产生了如此多的差异 (Lifely *et al.*, 1995; Wright and Morrison, 1997)。发酵过程的条件可监测不同糖基结构是怎样及何时产生的。液相色谱是监测批次间差异及下游进程的有效方法，而电泳技术由于其高通量、对比性分析及分辨率等方面的优势，应用越来越广泛 (Taverna *et al.*, 1998)。应用电泳技术如等电聚焦 (IEF)、毛细管电泳 (CE) 及二维凝胶电泳 (2-DE) 等使快速常规分析及产品监测日益简单化了。

糖基化的出现与缺失也能调节一些重组蛋白的功能，比如人粒细胞集落刺激因子 (rHG-CSF; Hoglund, 1998)、人淋巴细胞毒素 (Funahashi *et al.*, 1993)、白细胞介素 2 (Voshol *et al.*, 1996)、甲状腺过氧化物酶 (Fayadat *et al.*, 1998) 及 $\alpha 1$ -酸性糖蛋白 (Shiyan and Bovin, 1997) 都是糖基化后而具有生物活性的重组蛋白。其他一些重组蛋白天然就存在糖基化，如白细胞介素-3 (Zitener *et al.*, 1994)、血小板生成素 (Jagerschmidt *et al.*, 1998) 及人生长激素 (Becker and Hsiung, 1986)，它们甚至于在无糖基化时也有活性。蛋白质组学方法可以监测这些蛋白的体内变化，提供对天然糖基形式不均一性及功能方面的了解。一个具有生物学意义的糖蛋白在体外重组和表达之前，了解相关的知识是相当重要的。

4.2 经分离的胶上糖蛋白

由于纯化蛋白所提供的量是有限的，所以用一维或二维凝胶电泳分离蛋白从而实现糖基化结构变化检测的方法受到了挑战。质谱 (包括基质辅助激光解吸电离质谱——MALDI-MS，及电喷雾电离质谱——ESI-MS) 及毛细管区带电泳 (CIE)，是两种用于解决出现在糖蛋白异构体中糖链结构的不均一性问题的最有前途的技术。

糖蛋白物理及化学特性造成了用目前蛋白质组技术对其进行分析的困难，这一点已越来越清楚了，糖链的连接，特别是大型的 N-连接糖链，由于其庞大的分子量及亲水性，赋予了蛋白质不同的特性，为了分析 2-DE 分离的糖蛋白，蛋白质组研究需要在空间分辨 (阵列)、定性及生物信息学方面加以优化。

空间分辨 (阵列) 血浆糖蛋白的 2-DE 已做为血清蛋白图谱发表于网上 (Sanchez *et al.*, 1995)。在我们的工作中发现，采用重泡胀技术将纯化的糖蛋白在一维固相 pH 梯度 (IPG) 胶条上上样，导致进入胶条的样品量较低，杯状进样法可以使糖蛋白进入 IPG 胶条的量大大增加，小量大肠杆菌蛋白抽提物与重组糖蛋白经重泡胀上样产生同样的结果。图 13.2 显示了用标准糖蛋白——牛胎球蛋白所表现的这种效果，而对非糖基化的蛋白上样时无此类困难。糖蛋白这种效应的原因只能如上推断，但是它确实能够反映多糖对蛋白质表面积大小，电荷等方面的影响，同时，它也有助于揭示糖蛋白物理特性的不同。

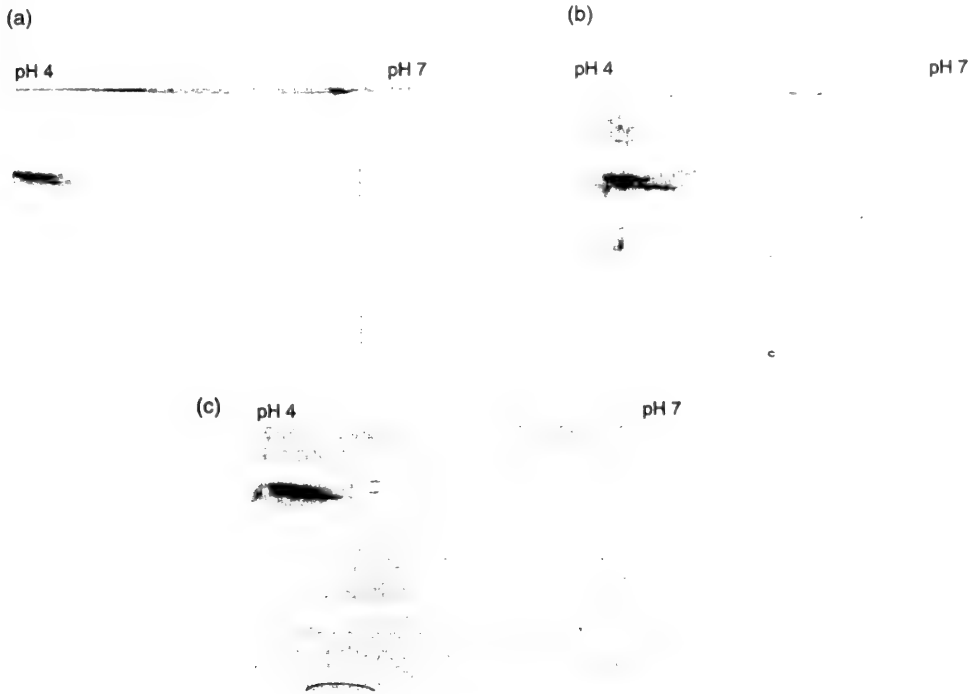


图 13.2 纯化糖蛋白的二维电泳

糖蛋白溶于 5mol/L 尿素, 2mol/L 硫脲, 2mmol/L TBP, 2% CHAPS, 2% 硫代甜菜碱 3-10, 0.5% 两性电解质 3-10, 40 mmol/L Tris, 上样于 pH 4~7 的 IPG 胶条, 2-DE 分离, 考马斯蓝染色。(a) 10 μ g 牛胰胎球蛋白一维重泡胀上样 (Rabilloud *et al.*, 1994); (b) 10 μ g 牛胰胎球蛋白一维阳极加样杯上样 + 1% pH 2.5~5 两性电解质 (Pharacia); (c) 10 μ g 牛胰胎球蛋白 + 5 μ g *E. coli* 全蛋白一维重泡胀技术加样。

糖蛋白一经分离, 下一步是将其显色, 许多用考马斯亮兰染色和银染方法来染糖蛋白的人们发现, 糖蛋白的染色非常不理想, 或者根本染不上色。为了对 2-DE 胶上蛋白质混合物中的糖蛋白着色, 最好的办法是染色其特殊的多糖部分, 那么, 灵敏的凝集素 (Golaz *et al.*, 1995) 及半抗原识别法可以实现这一过程, 目前市场上 (Bio-Rad 多糖检测试剂盒, Boehringer-Mannheim DIG 检测试剂盒) 已经出现了许多基于糖的高碘酸盐氧化作用, 以及夹心抗体或生物素/链霉亲和素等显色方法发展的产品。我们对染色方法加以改进, 采用荧光替代氧化糖上的颜色标签, 从而使 PVDF 膜上的糖蛋白快速、灵敏地显现。

定性 多糖的不均一性是导致 2-DE 胶上单个糖类型分离不明显的原因, 重组的截短的免疫球蛋白超家族糖蛋白 CD4, 它包括一个 N-糖基化位点, 在 2-DE 胶上分离为三个不同的糖型式 (图 13.3), 通常我们以唾液酸化的不同来解释这种在等电点上的差异。但是, 在分离实验中, N-连接糖蛋白的质谱分析对另一个糖蛋白——人 α 1-抗胰蛋白酶的解释显示, 这并非是直接原因。而其真正的原因是由于在不同的点上, 糖基化是不均一的 (Packer *et al.*, 1998)。在对 α 1-抗胰蛋白酶的单糖分析中, 我们发现在每个糖基形式中, 中性糖的唾液酸化有相同的比率, 糖基化的明显增加导致了等电点的降低 (Packer *et al.*, 1996)。

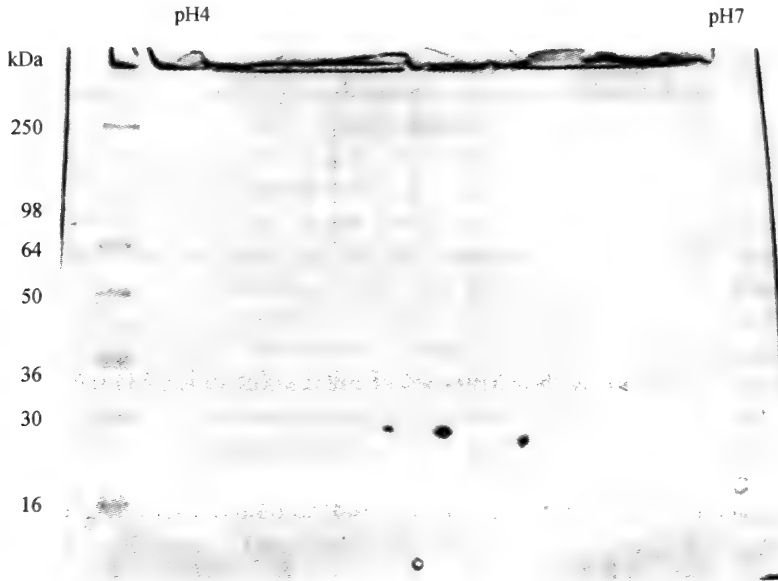


图 13.3 CD4.c 截型糖蛋白异构体的二维电泳

2-DE 分离 CD4.c 截型糖蛋白，一维用 pH4~7 的 IPG 胶条，蛋白以阳极杯上样于 1% (pH 2.5~5) 的两性电解质，二维用 10%~20% SDS-PAGE 分离，未修饰蛋白具有理论分子量 20 429Da，等电点 5.66。

ESI 与 MALDI 的快速发展已经给蛋白质的分析带来了革命性变化，同时，对多糖的分析也存在着同样的应用潜力 (Costello, 1997; Packer and Harrison, 1998)，脱糖基化、衍生化释放的多糖正相 HPLC、酶切测序以及 ESI 和 MALDI 等各种技术已经发展到对重组糖蛋白进行分析的阶段，同时目标已经指向了对胶上分离的糖蛋白完全结构鉴定。

N-连接多糖的糖肽，由于其质量 (包括多肽部分) 通常大于 3500Da，在 MALDI 上难以看到，加之以不均一化的趋势，最多只能在线性方式下看到一个宽的信号峰。O-连接多糖在总的趋势上比 N-连接多糖分子要小，做为胶上酶解糖蛋白的胰酶糖肽碎片，在质谱上有时可以看到 (Packer and Harrison, 1998)。将胶上电泳分离并用 PVDF 膜转印的血型糖蛋白 A (Packer *et al.*, 1998) 以及胶上荧光标记病毒蛋白 (Medina *et al.*, 1998) 进行碱性 β -消除释放糖链，也可对 O-连接多糖进行定性分析。胶上糖蛋白的分析还有胍解及胶上分离的明胶酶 B 变体的 O-糖链质谱分析 (Van den Steen *et al.*, 1986b)。O¹⁸ 标记的 N-连接位点谱及肽质量指纹谱分析 (Kuster and Mann, 1999)。

在重组糖蛋白异构型的高通量监测方面，毛细管区带电泳也是十分有用的，这一技术可以解决等电点及结构上的微小差异，同时对这些差异的定性可以通过与 MS 连接来实现。这一技术及其在重组多糖及糖肽谱图方面的应用由 Susuk 和 Honda (1998) 以及 Taverna 等 (1998) 的两篇详细综述报导。用这一技术分析 2-DE 分离的大量蛋白的难点在于它要求被分析物有一定的浓度 ($\mu\text{mol/L}$)。

首先，我们对蛋白质高通量分析方法的选择，是应用灵敏，快速，简单的 MALDI

自动方法从一个已知基因组对蛋白进行鉴定，这种方法已用于只有一些小的考染蛋白斑点的胰酶降解碎片。一旦蛋白在数据库上鉴定，许多被修饰的蛋白的微小特性也就相应显现（图 13.4）。接下来我们用较慢一点的 ESI-MS-MS 技术做为第二梯队，对我们感兴趣于其修饰的特殊蛋白加以定性分析。为了彻底确定糖基化性质，微不均一性及宏观差异理论上都应该进行测定。

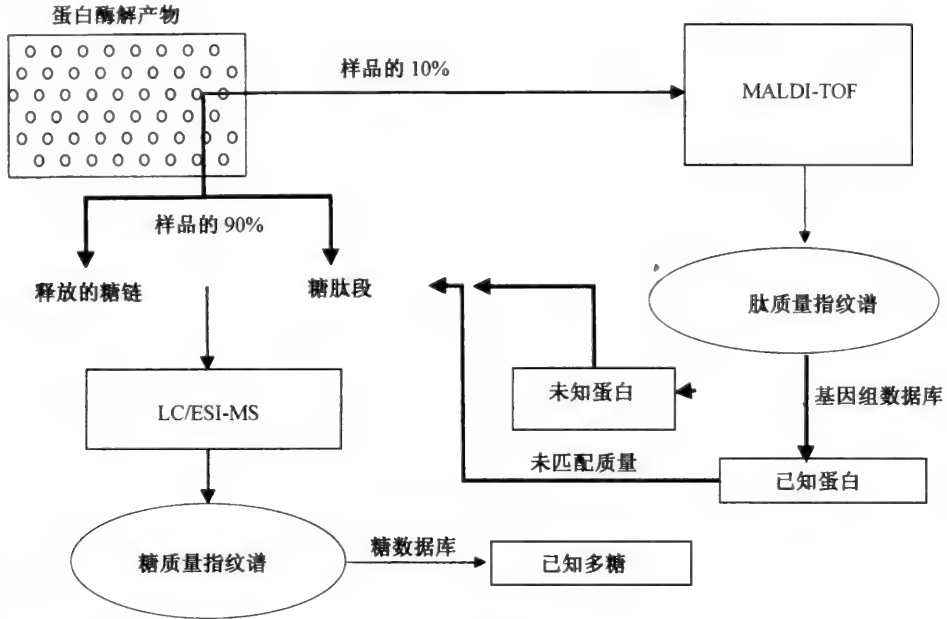


图 13.4 糖蛋白组策略

2-DE 分离糖蛋白结合自动化高通量质谱分析

我们认为 ESI 在糖蛋白或糖肽——严格地说是那些具有微不均一性的 N-连接糖基化的糖蛋白——的离子化方面比 MALDI-TOF 更为成功，这可能是由于在 MALDI 中必须应用基质造成的，那些具有微不均一性的多糖、大分子的 N-连接糖链以及难以观察的多糖唾液酸化，使它们与基质难以发生衍生化 (Powell and Harvey, 1996)。我们对不完全的、CHO 表达的重组大鼠 CD4 (未修饰 MDa=20 429.31Da) 进行分析，它包含一个 N-糖基化位点，图 13.5 列出了这两种质谱在糖蛋白分析方面的不同，在全部分子被脱唾液酸化后，具有 8 个支链的 N-连接多糖可以被定性分析 (Ashford *et al.*, 1993)，而它们是包括在 2 个 N-连接糖基化位点上的，在 MALDI-MS 图上，由于其广泛的不均一性，在如此高的质量范围内，没有个别质量能被分开 (图 13.5a)。ESI-MS 可以分辨单一一位点上 9 个可能的糖基化形式 (图 13.5b)，由于我们用的是未加修饰的蛋白质量，因此，主要的宽峰可被视为杂合形多糖结构 [22 497 Da = 核心 (GlcNAc)₂ (Man)₃ + (Gal)₁ (GlcNAc)₁ (Man)₅] 及混合形多糖结构 [22 489Da = 核心 (GlcNAc)₂ (Man)₃ + (Gal)₂ (GlcNAc)₂ (Fuc)₁ (NeuAc)₁] 的混合物，质量偏差是由相关的微不均一性造成的。

质谱可以测定不同糖基化形式中单糖位点。通过碰撞诱导解离的特殊方式，这一技术还可以提供各个残基的序列、连接方式方面的信息，文献中报导的绝大多数碎片的数

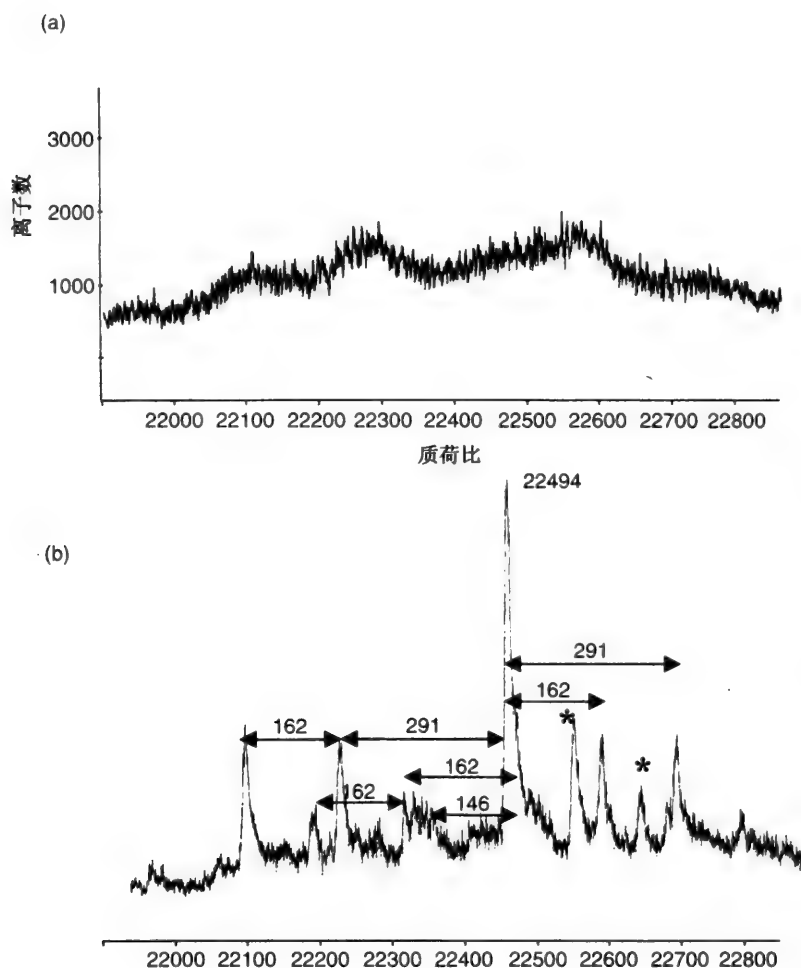


图 13.5 CD4.c 糖蛋白的质谱图

(a) MALDI-TOF (Perseptive Voyager): 20pmol CD4.c 糖蛋白, 用 Poros 50R1 (相当于 C4 反相性质) 清洗 (Kussman 等, 1997), 用 2 μ l 2% TFA 洗脱, 80% MeCN 含 10mg/ml 芥子酸直接上靶; (b) ESI-TOF-MS (Micromass), 100pmol CD4.c (同批) 脱盐用 HPLC C4 柱 (Aquapore Bu-300 \AA 100 \times 2.1mm, PE Brownlee), 梯度为 0.05% TFA 至 80% 乙腈及 0.05% TFA, 30min, 蛋白大约在 22min 洗脱, 检测波长 214nm (主峰加合 TFA)。

据是由单甲基化的多糖的快原子轰击 (FAB) 质谱获得的。MALDI 及 ESI 这两种技术适合于在高灵敏度地对大分子量范围内的生物分子进行测定, 因而它们可用于探索打碎糖复合物, 目前已经获得了在对蛋白或多肽释放的非衍生化多糖结构分析方面的较好数据。MALDI-TOF 的源后衰变 (PSD) (Naven *et al.*, 1997) 或 CIO-ESI-TOF 的源内碰撞-诱导解离 (Packer and Harrison, 1998) 或 ESI-MS-MS (Oxley, 1998) 等技术可以打碎糖苷键。由于微孔及毛细管液相谱的接口适合于被分离的糖基化形式的碰撞诱导碎片分析, 目前电喷雾电离技术已成为蛋白糖基化结构分析的最佳选择 (Bulingame, 1996), 不同离子化技术与更好的检测方法之间结合是质谱的发展方向, 我们感到在蛋

白糖基化形式的分析方面,这一方法将比别的方法出色。

糖蛋白质组的质谱常规应用还存在大量有待于解决的问题。对于 MALDI-MS 分析,未来的工作是需要发现一些适合于大分子亲水性糖肽的特殊基质,同时, MALDI 分析对于糖肽的上游富集的选择将证明是有帮助的。对于 ESI-MS,经由源内 CID,糖肽可以在 LC 上分离,而肽上的多糖碎片则难以被解释,对这种不带电荷的多糖的检测灵敏度需要进一步提高,同时,需要一个易离子化基团对目标糖进行衍生化。这种化学反应必须是反应完全并简单的,同时能够适应高通量的分析,在质谱之前被采用的柱分离及电泳分离步骤需要加快速度及自动化进程。

可是,对蛋白质上糖类型的完全结构分析,其要求远远超过仅仅对糖进行分子质量测定。与用于肽质量指纹谱中的氨基酸质量相比,许多单糖残基有着相同的分子质量,同时,连接方式及序列难以用质谱测定。尽管如此,由于各种单糖在其结构和框架范围内有一定的守恒性,使糖蛋白的分析简化。因此,可以通过对一些已知的特殊组织或种属方面的知识,来约束可能的具有某种特定质量的多糖结构。

信息学 基因组及蛋白质组领域所获得的大量数据给生物信息学领域带来了巨大的冲击,数据及开发数据库的工具广泛地应用于 DNA 的翻译和蛋白序列以及修饰,在分析不断进展的同时,需要构建数据库及信息工具适应糖基化数据的翻译。

Carbank (<http://128.192.9.29/Carbbank/Carbbank.html>) 自 1995 年来未收到输入数据的基金, O-Glycbase (<http://www.cbs.dtu.dk/databases/OGLYCBASE/>) 以神经网络预测 O-糖基化位点,至今也仅仅发表了一些有限的糖蛋白寡糖链信息的数据库。我们最近已经整理了一个糖蛋白 O-连接位点的数据库 (Cooper *et al.*, 1999) 并且正在准备更加广泛的生物信息学工具及数据库,以用于糖蛋白质组学研究。

SWISS-PROT 及相类似的许多数据库已经给生物化学家们提供了信息的贮藏地及开发的工具,我们也需要类似于它的完整的信息库,以适应糖分析的发展。糖肽或被释放的糖的质量测定,可以较容易地转化成单糖构架,这些构架与先前报导的结构相匹配,同时通过设计特殊的糖苷外切酶实验,可以检测结构的可能性。从特殊结构质谱碎片模型(糖质量指纹谱)的知识还可以推测序列,我们需要综合糖蛋白中大量有价值的知识,以使用它们来分析从糖蛋白质组分析中获取的数据。

5. 结论

蛋白的翻译后修饰,包括糖基化,将证明是对于发生在生物学体系中新陈代谢的“极好控制”,当基因产品在蛋白活性组织中被修饰时,这一点尤为明显。多糖结构中可利用的广泛的差异性使其成为微调新陈代谢的最佳部分。Clark 等(1997)已将这种糖基化的特殊功能假定为从母体免疫应答的人胚胎的保护基础。并且,这些特殊多糖的获得也许意味着,病原体能够搅乱宿主的免疫应答。2-DE 展示的糖蛋白阵列提供一个谱图,由它我们可以探索疾病发生、发展过程中的变化。重组糖蛋白由于其能够提供足够的样品量,对于它们的分析工具已经较完善了。目前要求这些可达到 pmol 灵敏度的检测方法,能够象蛋白质的鉴定及定性的一些方法一样,适应高通量的糖蛋白的鉴定与定性。从已开发的数据形式我们可以获得延伸的信息,这将依赖于改进的、更加灵敏的方法学的继续发展,以及生物信息特别是糖生物信息工具的构建。

致谢

感谢 Neily Barclay 友好提供纯化的重组截形 CD4, Ben Herbert 提供 2-DE 方面的帮助, 以及 Keith Williams 和 rew Gooley 共同整理。

(蔡耘译 钱小红校)

参考文献

- Anobile, C.J., Talbot, J.A., McCann, S.J., Padmanabhan, V. and Robertson, W.R. (1998) Glycoform composition of serum gonadotrophins through the normal menstrual cycle and in the post-menopausal state. *Mol. Hum. Reprod.* **4**: 631–639.
- Ashford, D.A., Alafi, C.D., Gamble, V.M., MacKay, D.J.G., Rademacher, T.W., Williams, P.J., Dwek, R.A., Barclay, A.N., Davis, S.J., Somaoza, C., Ward, H.A. and Williams, A.F. (1993) Site-specific glycosylation of recombinant rat and human soluble CD4 variants expressed in Chinese hamster ovary cells. *J. Biol. Chem.* **268**: 3260–3267.
- Becker, G.W. and Hsiung, H.M. (1986) Expression secretion and folding of human growth hormone in *Escherichia coli*. Purification and characterization. *FEBS.* **204**: 145–150.
- Brockhausen, I. and Kuhns, W. (1997) *Glycoproteins and Human Disease*. R.G. Landes, New York.
- Brockhausen, I., Schutzbach, J. and Kuhns, W. (1998) Glycoproteins and their relationship to human disease. *Acta Anat.* **161**: 36–78.
- Burlingame, A.L. (1996) Characterization of protein glycosylation by mass spectrometry. *Curr. Opin. Biotechnol.* **7**: 4–10.
- Ching, C.K. and Rhodes, J.M. (1988) Identification and partial characterization of a new pancreatic cancer-related serum glycoprotein by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and lectin blotting. *Gastroenterology* **95**: 137–142.
- Clark, G.F., Dell, A., Morris, H.R., Patankar, M., Oehninger, S. and Seppala, M. (1997) Viewing AIDS from a glycobiochemical perspective: potential linkages to the human foetoembryonic defence system hypothesis. *Mol. Hum. Reprod.* **3**: 5–13.
- Cooper, C.A., Wilkins, M.R., Williams, K.L. and Packer, N.H. (1999) *Electrophoresis* **20** (in press).
- Costello, C.E. (1997) Time, life ... and mass spectrometry. New techniques to address biological questions. *Biophys. Chem.* **68**: 173–188.
- Delahunt, B., Maxwell, J., Moroni-Rawson, P., Renner, R. M., Nacey, J.N. and Jordan, T.W. (1996) Serum protein variation in patients with renal cell carcinoma. *J. Urol. Pathol.* **5**: 97–107.
- Delves, P.J. (1998) The role of glycosylation in autoimmune disease. *Autoimmunology* **27**: 239–253.
- Dwek, R.A. (1995) Glycobiology: towards understanding the function of sugars. *Biochem. Soc. Trans.* **23**: 1–25.
- Ellies, L.G., Jones, A.T., Williams, M.J. and Ziltener, H. J. (1994) Differential regulation of CD43 glycoforms on CD4+ and CD8+ T lymphocytes in graft-versus-host disease. *Glycobiology* **6**: 885–893.
- Elliott, M., Kardana, A., Lustbader, J. and Cole, L. (1997) Carbohydrate and peptide structure of the alpha- and beta-subunits of human chorionic gonadotrophin from normal and aberrant pregnancy and choriocarcinoma. *Endocrine* **15**–32.
- Fayadat, L., Niccoli-Sire, P., Lanet, J. and Franc, J.L. (1998) Human thyroperoxidase is largely retained and rapidly degraded in the endoplasmic reticulum. Its N-glycans are required for folding and intracellular trafficking. *Endocrine* **139**: 4277–4285.
- Funahashi, I., Wantanabe, H., Abo, T., Indo, K. and Miyaji, H. (1993) Tumour growth inhibition in mice by glycosylated recombinant human lymphotoxin: analysis of tumour-regional mononuclear cells involved with its action. *Br. J. Cancer* **67**: 447–455.
- Furakawa, H., Doh-ura, K., Kikuchi, H., Tateishi, J. and Iwaki, T. (1998) A comparative study of abnormal prion protein isoforms between Gerstmann–Straussler–Scheinker syndrome and Creutzfeldt–Jakob disease. *J. Neurol. Sci.* **158**: 71–75.
- Golaz, O., Gravel, P., Walzer, C., Turler, H., Balant, L. and Hochstrasser, D. F. (1995) Rapid detection of the main human plasma glycoproteins by two-dimensional polyacrylamide gel electrophoresis lectin affino blotting. *Electrophoresis* **16**: 1187–1189.
- Hankey, D.P., Nicholas, R.M. and Hughes, A.E. (1992) Two-dimensional polyacrylamide gel electrophoresis reveals differences between osteoblast and fibroblast extracellular proteins. *Electrophoresis* **13**: 329–332.

- Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **15**: 115–130.
- Hart, G.W., Kreppel, L.K., Comer, F.I., Arnold, C.S., Snow, D.M., Ye, Z., Cheng, X., Della Manna, D., Caine, D.S., Earles, B.J., Akimoto, Y., Cole, R.N. and Hayes, B.K. (1996) O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization. *Glycobiology* **7**: 711–716.
- Haynes, P.A. (1998) Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology* **8**: 1–5.
- Hoglund, M. (1998) Glycosylated and non-glycosylated recombinant human granulocyte colony-stimulating factor (rhG-CSF) – What is the difference? *Med. Oncol.* **15**: 229–233.
- Hounsell, E.F., Davies, M.J. and Renouf, D.V. (1996) O-linked protein glycosylation structure and function. *Glycoconj. J.* **13**: 19–26.
- Jagerschmidt, A., Fleury, V., Anger-Leroy, M., Thomas, C., Agnel, M. and O'Brein, D.P. (1998) Human thrombopoietin structure-function relationships: identification of functionally important residues. *Biochem. J.* **333**: 729–734.
- Kim, Y.S., Gum, J.R. and Brockhausen, I. (1996) Mucin glycoproteins in neoplasia. *Glycoconj. J.* **13**: 693–707.
- Kinoshita, T., Ohishi, K. and Takeda, J. (1997) GPI-anchor synthesis in mammalian cells: genes, their products, and a deficiency. *J. Biochem. (Tokyo)* **122**: 251–257.
- Kudryashov, V., Kim, H.M., Ragupathi, G., Danishefsky, S.J., Livingston, P.O. and Lloyd, K.O. (1998) Immunogenicity of synthetic conjugates of Lewis(y) oligosaccharide with proteins in mice: towards the design of anticancer vaccines. *Cancer Immunol. Immunother.* **45**: 281–286.
- Kukuruzinska, M.A. and Lennon, K. (1998) Protein N-glycosylation: molecular genetics and functional significance. *Crit. Rev. Oral Biol. Med.* **9**: 415–448.
- Kussman, M., Nordhoff, E., Rahbek-Nielsen, H., Haebel, S., Rossel-Larsen, M., Jakobsen, L., Gobom, J., Mirgorodskaya, E., Kroll-Kristensen, A., Palm, L. and Roepstorff, P. (1997) Matrix-assisted laser desorption/ionisation mass spectrometry sample preparation techniques designed for various peptide and protein analytes. *J. Mass Spec.* **32**: 593–601.
- Kuster, B. and Mann, M. (1999) 18O-labeling of N-glycosylation sites to improve the identification of gel-separated glycoproteins using peptide mass mapping and database searching. *Anal. Chem.* **71**: 1431–1440.
- Kuster, B., Wheeler, S.F., Hunter, A.P., Dwek, R.A. and Harvey, D.J. (1997) Sequencing of N-linked oligosaccharides directly from protein gels: in-gel deglycosylation followed by matrix-assisted laser desorption/ionization mass spectrometry and normal-phase high-performance liquid chromatography. *Anal. Biochem.* **250**: 82–101.
- Kuster, B., Hunter, A.P., Wheeler, S.F., Dwek, R.A. and Harvey, D.J. (1998) Structural determination of N-linked carbohydrates by matrix-assisted laser desorption/ionization-mass spectrometry following enzymatic release within sodium dodecyl sulphate–polyacrylamide electrophoresis gels: application to species-specific glycosylation of alpha1-acid glycoprotein. *Electrophoresis* **19**: 1950–1959.
- Lifely, M.R., Hale, C., Boyce, S., Keen, M.J. and Phillips, J. (1995) Glycosylation and biological activity of CAMPATH-1H expressed in different cell lines and grown under different culture conditions. *Glycobiology* **5**: 813–822.
- Lis, H. and Sharon, N. (1993) Protein glycosylation: structural and functional aspects. *Eur. J. Biochem.* **218**: 1–27.
- Lo-Man, R., Bay, S., Vichier-Guerre, S., Deriaud, E., Cantacuzene, D. and Leclerc, C. (1999) A fully synthetic immunogen carrying a carcinoma-associated carbohydrate for active specific immunotherapy. *Cancer Res.* **59**: 1520–1524.
- Lu, Q. and Shur, B.D. (1997) Sperm from beta 1,4-galactosyltransferase-null mice are refractory to ZP3-induced acrosome reactions and penetrate the zona pellucida poorly. *Development* **124**: 4121–4131.
- Mackiewicz, A. and Mackiewicz, K. (1995) Glycoforms of serum alpha 1-acid glycoprotein as markers of inflammation and cancer. *Glycoconj. J.* **12**: 241–247.
- Medina, L., Grove, K. and Haltiwanger, R. S. (1998) SV40 large T antigen is modified with O-linked N-acetylglucosamine but not with other forms of glycosylation. *Glycobiology* **8**: 383–391.

- Montreuil, J., Vliegthart, J.F.G. and Schachter, H. (1996) Glycoproteins and disease. In: *New Comprehensive Biochemistry*, Vol. 30. Elsevier, New York.
- Naven, T.J., Harvey, D.J., Brown, J. and Critchley, G. (1997) Fragmentation of complex carbohydrates following ionisation by matrix-assisted laser desorption with an instrument fitted with time-lag focusing. *Rapid Commun. Mass Spectrom.* **11**: 1681–1686.
- Oxley, D., Munro, S.L., Craik, D.J. and Bacic, A. (1998) Structure and distribution of N-glycans on the S7-allele stylar self-incompatibility ribonuclease of *Nicotiana glauca*. *J. Biochem. (Tokyo)*. **123**: 978–983.
- Packer, N.H. and Harrison, M.J. (1998) Glycobiology and proteomics: is mass spectrometry the holy grail? *Electrophoresis* **19**: 1872–1882.
- Packer, N.H., Wilkins, M.R., Golaz, O., Lawson, M.A., Gooley, A.A., Hochstrasser, D.F., Redmond, J.W. and Williams, K.L. (1996) Characterization of human plasma glycoproteins separated by two-dimensional gel electrophoresis. *Bio/Technology* **14**: 66–70.
- Packer, N.H., Lawson, M.A., Jardine, D.R., Sanchez, J.-C. and Gooley, A.A. (1998) Analysing glycoproteins separated by two-dimensional gel electrophoresis. *Electrophoresis* **19**: 981–988.
- Powell, A.K. and Harvey, D.J. (1996) Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **10**: 1027–1032.
- Rabilloud, T., Valette, C. and Lawrence, J.J. (1994) Sample application by in-gel rehydration improves the resolution of two-dimensional electrophoresis with immobilized pH gradients in the first dimension. *Electrophoresis* **15**: 1552–1558.
- Sanchez, J.-C., Appel, R.D., Golaz, O., Pasquali, C., Ravier, F., Bairoch, A. and Hochstrasser, D.F. (1995) Inside SWISS-2DPAGE database. *Electrophoresis* **16**: 1131–1151.
- Seppala, M., Koistinen, H., Koistinen, R., Mandelin, E., Oehninger, S., Clark, G.F., Dell, A. and Morris, H.R. (1998) Glycodelins: role in regulation of reproduction, potential for contraceptive development and diagnosis of male infertility. *Hum. Reprod.* **13**(Suppl. 3): 262–291.
- Seregini, E., Botti, C. and Bombardieri, E. (1995) Biochemical characteristics and clinical applications of alpha-fetoprotein isoforms. *Anticancer Res.* **15**: 1491–1499.
- Shiyan, S. D. and Bovin, N.V. (1997) Carbohydrate composition and immunomodulatory activity of different glycoforms of alpha 1-acid glycoprotein. *Glycoconj. J.* **14**: 631–638.
- Starr, C.M., Klock, J.C., Skop, E., Masada, I. and Giudici, T. (1994) Fluorophore-assisted electrophoresis of urinary carbohydrates for the identification of patients with oligosaccharidosis- and mucopolysaccharidosis-type lysosomal storage diseases. *Glycosyl. Dis.* **1**: 165–176.
- Susuki, S. and Honda, S. (1998) A tabulated review of capillary electrophoresis of carbohydrates. *Electrophoresis* **19**: 2539–2560.
- Tagliaro, F., Crivellente, F., Manetto, G., Puppi, I., Deyl, Z. and Marigo, M. (1998) Optimized determination of carbohydrate-deficient transferrin isoforms in serum by capillary zone electrophoresis. *Electrophoresis* **19**: 3033–3039.
- Taketa, K. (1998) Characterization of sugar chain structures of human alpha-fetoprotein by lectin affinity electrophoresis. *Electrophoresis* **19**: 2595–2602.
- Taverna, M., Tran, N.T., Merry, T., Hovarth, E. and Ferrier, D. (1998) Electrophoretic methods for process monitoring and the quality assessment of recombinant glycoproteins. *Electrophoresis* **19**: 2572–2594.
- Taylor-Papadimitriou, J.J. and Epenetos, A.A. (1994) Exploiting altered glycosylation patterns in cancer: progress and challenges in diagnosis and therapy. *TiBTech.* **12**: 227–233.
- van Dam, G., Bogtsh, B., van Zeyl, R., Rotmans, J. and Deelder, A. (1996) *Schistosoma mansoni*: in vitro and in vivo excretion of CAA and CCA by developing schistosomula and adult worms. *J. Parasitol.* **82**: 557–564.
- Van den Steen, P., Rudd, P.M., Proost, P., Martens, E., Paemen, L., Kuster, B., van Damme, J., Dwek, R.A. and Opdenakker, G. (1998a) Oligosaccharides of recombinant mouse gelatinase B variants. *Biochim. Biophys. Acta* **1425**: 587–598.
- Van den Steen, P., Rudd, P.M., Dwek, R.A. and Opdenakker, G. (1998b) Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**: 151–208.
- van Dijk, W., Havenaar, E.C. and Brinkman-van der Linden, E.C. (1995) Alpha 1-acid glycoprotein

- (orosomucoïd): pathophysiological changes in glycosylation in relation to its function. *Glycoconj. J.* **12**: 227–233.
- Varki, A.** (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**: 97–130.
- Voshol, H., Dullens, H.F., Otter, W.D. and Vliegthart, J.F.** (1996) The glycosylation profile of interleukin-2 activated human lymphocytes correlates to their anti-tumour activity. *Anticancer Res.* **16**: 155–159.
- Winchester, B., Clayton, P., Mian, N., di-Tomaso, E., Dell, A., Reason, A. and Keir, G.** (1995) The carbohydrate-deficient glycoprotein syndrome: an experiment of nature in glycosylation. *Biochem. Soc. Trans.* **23**: 185–188.
- Wright, A. and Morrison, S.L.** (1997) Effect of glycosylation on antibody function: implications for genetic engineering. *TiBTech.* **15**: 26–32.
- Youakim, A., Hathaway, H., Miller, D., Gong, X. and Shur, B.** (1994) Overexpressing sperm surface beta 1,4-galactosyltransferase in transgenic mice affects multiple aspects of sperm-egg interactions. *J. Cell Biol.* **126**: 1573–1583.
- Yuasa, I., Ohno, K., Hashimoto, K., Iijima, K., Yamashita, K. and Takeshita, K.** (1995) Carbohydrate-deficient glycoprotein syndrome: electrophoretic study of multiple serum glycoproteins. *Brain Devl.* **17**: 13–19.
- Zara, J. and Naz, R.K.** (1998) The role of carbohydrates in mammalian sperm-egg interactions: how important are the carbohydrate epitopes? *Front. Biosci.* **3**: d1028–1038.
- Ziltener, H.J., Clark-Lewis, I., Jones, A.T. and Dy, M.** (1994) Carbohydrate does not modulate the *in vivo* effects of injected interleukin-3. *Exp. Hematol.* **22**: 1070–1075.

第十四章 学术环境中的蛋白质组学

Michael A. Cahill and Alfred Nordheim

1. 引言

这本书的作者人数在学院和企业中大致分布相等。但用于学术研究的预算仅相当于同类企业支出的很小一部分。既然新技术的引进是昂贵的，像支持蛋白质组学所要求的那样，一些人已指出学术性的蛋白质组学不可能具有竞争性。尽管如此，我们和其他的许多同行正打算建立一个蛋白质组的基本设施，即匆匆上马组建的 Tuebingen 大学的细胞生物研究所。我们实验室在哺乳类转录调节和信号传导方面的建树得到一致公认 (Arsenian *et al.*, 1998; Cahill *et al.*, 1996; Heidenreich *et al.*, 1999; Hipskind *et al.*, 1991)。蛋白质组学在这一领域的应用潜力促使我们全身心地接受这种技术。蛋白质组学设施的建立在我们研究所至今已有 30 个月，并且筹建蛋白质组研究小组的时候，已经有两个博士后，两个研究生和两个技术助教。我们正同心协力提高技术，并实施适合我们研究所现状的临床和生物学科的计划。我们已经成功地获得了资金，用于购买等电聚焦仪、磷储屏图像分析仪、MALDI-TOF、将酶解的蛋白转到 MALDI 样品阶段的机器人和四级杆飞行时间串联质谱。我们已经与临床和商业密切接触，已申请几个蛋白质组学相关的专利，并且已投稿正在审查的论文。

2. 为什么不在学术环境中进行蛋白质组研究？

蛋白质组分析是一项有潜力改变医学面貌的技术。一些观点认为，由于高额的研究费用，蛋白质组的研究不宜在学院开展。当人们意识到蛋白分析中昂贵的设备和药品的消耗，尤其是高通量的自动化流程都往往自然地认为大量与药物相关的蛋白质组的研究工作应由商业组织开展；但实际上在学术机构内建立蛋白质组研究体系及技术仍具有可观的潜力和价值。

另一种论点认为，高通量的流程是重复的、自动化和智能化低的，而学术研究应具有创新性。对于学术研究人员，如果曾经准备并筛选过 100 个质粒 DNA 微量制备以鉴定正确 DNA 载体构建从而使关键实验得以进行的话，这一观点的谬误之处显而易见了。

学术机构中蛋白质组体系的发展面临的一个重要问题是人才向企业界的流失，因为企业上急需训练有素的蛋白质组研究人员，并且企业界能提供更高的薪金和貌似更强的安全感。这一问题在包括德国在内的一些国家中尤为突出，由于在这些国家学术机构的薪金水准是预先设定和不能更改的，这就使得学术机构很难保留一支高水准的研究队伍。

3. 为什么要在学术环境中进行蛋白质组研究？

总体考虑，许多方面都有力支持大学和非盈利机构参与蛋白质组领域的研究，并认为是很有价值的。下面首先考虑这样一种争论：大学和非盈利机构不具有从事蛋白质组

研究的经济实力。

3.1 大学能否承担得起蛋白质组的研究经费?

当建立一个适当的研究小组的费用满足后,目前的蛋白质组分析费用并不比其他研究的费用高,如小鼠分子遗传学。甚至研究小组的组建费用与其他研究相比也并未高出许多。我们曾在三个方面比较了相同人数的蛋白质组实验室与传统的分子生物学实验室的组建费用。显然,蛋白质鉴定设备花费较大;然而考虑到蛋白质组研究所得数据的质量和容量,以及数据潜在的价值,这些将有力表明:蛋白质组研究的花费比生命科学研究中的其他领域更物有所值。

二维电泳凝胶的费用 图 14.1 是我们的蛋白质组实验室从 1998 年 3~12 月的费用情况(以欧元为单位)。在此期间制备和运转了约 1300 块由浓度为 13%的、标准尺寸为 20cm×20cm、含哌嗪双丙烯酰胺(PDA)作为交联剂的聚丙烯酰胺凝胶。绝大多数凝胶采用银氨进行染色(Hochstrasser *et al.*, 1988; Rabilloud, 1992)。由于多数 PDA(属主要消费项目)是在 1998 年 3 月以前购买的,这将使 3~12 月间的实际花费增加约 2000 欧元。总花费约 40 000 欧元中的大部分用于购置,这是任何一个新建实验室所必需的。在消费高峰的 6 月至 12 月中,实验室的平均流动花费相对稳定在 2500 欧元/月。这笔费用供两名博士后、两名学生和一名技师使用,平均每人每月 500 欧元,完全在基金资助项目研究花费的正常范围内。因此,我们可以认为二维电泳凝胶费用与其他研究项目的花费相差无几。

二维电泳数据分析费用 采用常规的、已公布的蛋白质组分析技术时,实验室中最昂贵的消费品就是雇员的时间。蛋白质组目前是一项劳动密集型的工作。由于凝胶间存在差异,对某一特定样品有必要重复 3~6 次实验,最好采用双份样品。采用商品化二维电泳图像分析软件,这些凝胶图像随后成为数字化文件并排列。我们采用的分析软件有 Image Master 2-D Elite 软件(从 Amersham Pharmacia 生物技术公司购买)和

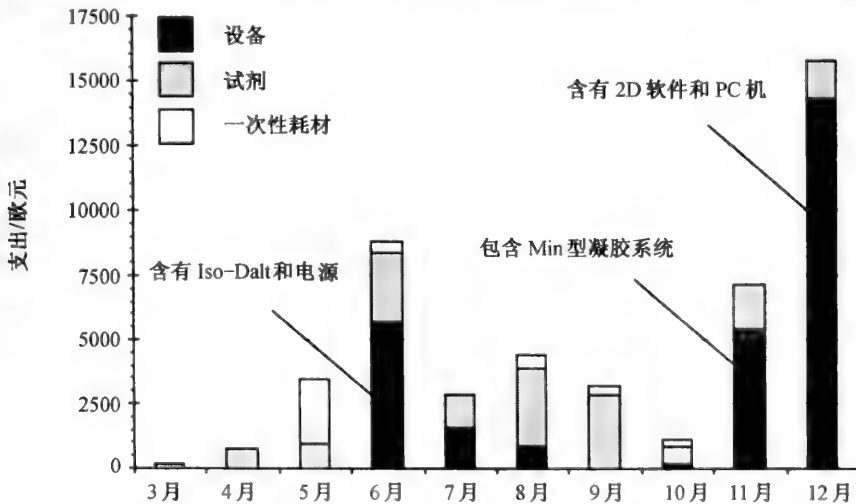


图 14.1 从 1998 年 3 月到 12 月, Tübingen 大学分子生物学系蛋白质组学实验所消耗费用 (设备、试剂和一次性用品)

MELANIE II 软件 (瑞士 2D PAGE 集团, 从 Bio-Rad 公司购买)。软件运行环境: 双处理器奔腾 II 型 PC 机, CPU 主频 350MHz, 10G 硬盘, 内存 500M, 实际内存设置为 3000M, 操作系统为 Windows NT 4.0。在该配置中图像处理时间是可接受的。但根据我们的经验, 上述两种软件都不令人满意。一些用户声称, 这些缺陷不幸正好位于一些公司所给予的技术支持服务之列。其他一些用户也曾向我们表达了类似的看法, 因此购买者一定要小心! 不论购买了其中哪一种软件, 用户都将在每张凝胶上耗费数小时的时间去将凝胶上的蛋白进行对齐和配比, 才能获得满意的结果。例如: 我们研究组利用脑脊液作为研究对象, 选择 pH 范围 4~7 的固相 pH 梯度干胶条 (IPG), 并观察到每块凝胶上约有 1000~1400 个斑点。斑点自动检测可识别凝胶间许多对蛋白作为一个斑点, 但不幸的是在不同的凝胶间并不一致。因此, 研究人员必须经手工证实或使每一个斑点有效, 以确保其准确性。在我们实验室里这个利用自动斑点检测的系统经过编辑后, 每块凝胶需花 4h 的时间, 这仅仅比手工增加斑点快一点, 它需要每块胶 4.5h。在进行脑髓液 (CSF) 的研究中, 有 15 个作为对照组的成员, 5 个明确诊断为有某种疾病。每个病人的样品需要进行 5 次电泳以产生总数达 100 块凝胶的实验。这要求计算机分析 400h。到目前为止, 这是分析中最昂贵的成分。同样的方法也被其他研究人员采用 (Andersen *et al.*, 1997)。

的确, 在企业代表们的讨论中, 我们感到这种劳动耗费, 商业机构与学术单位明显不同, 它是妨碍蛋白质组学工业化发展的主要因素。有几种并未公布的应用软件程序, 运行比商业软件好, 并且, 几个组织正力图制定策略来突破图像分析的瓶颈。然而, 就目前所公布的几种技术的应用而言, 大学与企业相比, 在蛋白质组凝胶分析方面处于明显的优势。而由于这些研究所内部具有多种专业人员, 他们可能较好地提高已有的软件或开发新软件。

蛋白鉴定的花费。蛋白鉴定所要求的特殊酶价格不菲, 但并不比所需要的常规分子生物学试剂贵。至今, 最大的花费是购买精致的蛋白鉴定仪器。质谱则被公认为最好的选择 (Burlbyame *et al.*, 1998; Gygi *et al.*, 1999; James, 1997; Yates, 1998)。MALDI-TOF 质谱特别提供了为高通量筛选的潜力。从 Hoffman-La Roche 公司来的科学家在 1998 年 9 月召开的第 3 届 2-DE 会议上报告了他们利用 MALDI-MS 在一天内可以鉴定 500 个蛋白质。大多数的质谱在将来可以提供这种常规水平的测试。这样一种自动化系统 (即从胶上斑点切割到 MALDI-MS), 将会花费约 500 000 欧元。此外, 需要一个通风良好、完全干净的实验室和至少两个专家小组比较合理。如果某一特定蛋白的 cDNA 序列存在于一个数据库内, 那么它将可能被鉴定。很快所有的人类蛋白都将满足此前提。

对于那些未知序列的蛋白, 例如, 对于鲜为人知的生物或从许多低丰度人类 mRNA 翻译的蛋白而言, 此类蛋白的部分氨基酸序列则必须进行从头合成。这对于具有源后衰变 (PSD) 功能的现代 MALDI-MS 是可能的 (Chaurond *et al.*, 1999), 或者另外的电喷雾质谱可被应用。如三级四极杆 (quadrupole), 离子阱, 或 Q-TOF 质谱 (Burlingame *et al.*, 1998; Kuster and Man, 1998; Rates, 1998; 请见第五章)。另外一种经常被低估, 但明显有用的选择是应用微量-HPLC 进行多肽分级, 随后进行 Edman 降解。现代的毛细管 HPLC Edman 测序仪其灵敏度已达 10~20fmol; 然而, 真正的问

题在于，当在 HPLC 分离过程中混合物酶解成能被测序的单个肽片段时样品的丢失。因此，蛋白质鉴定要求可能的系列仪器价格在 50 000 到 400 000 欧元之间。此方面要具有高度风险性，因为这些仪器会迅速过时。表面上看这一定是无可奈何的事情，即这样一个“机器公园”应该成为一种集中的设施来从资本投资中获取最大的盈利。但如果对一件仪器投资 50 万欧元，可能建成一个高性能、竞争性的设备，如果投资超过 100 万欧元或更多，可能建成一个很优秀的平台。反对大学蛋白质组学研究的批评家们会将这种水平的支出能否承担作为问题来诘难。然而，当利用非自动化的蛋白酶解、价廉的 MALDI-MS 仪器时，蛋白质组设施也能有效建立；并且这对某些实验室已足够 (Zubritsky, 1998)。依我们之见，完全有理由提出一个象德国大小的国家支撑学术圈内拥有上述实力的多中心。对于科学社区和医学、生物技术研究的利益将远大于论证这些耗费。

信息管理的花费 大而复杂数据组的产生随之而来要求完成复杂的数据库和分析软件，更不用说强大的计算机。理论上，蛋白质组数据一定要以数据库的形式产生，文件化和档案化。几个大公司正开发自己的软件和数据库。这些生物信息学引擎被设计为令人信服的文件，并以数据库的格式尽可能的连接许多相关的参数。随后，强有力的算法试图从原始数据中提取图谱，并用来作数据解释。进而应该允许模建和模拟，并最终会产生实验上可证实的预测并揭示这个系统的真正复杂性。

尽管上面已提到特殊的、商品化的 2-DE 分析软件包，但是它们的表现远远落在了这些企业化产品的后面。就我们的知识而言，在开发过程中根本没有相似的数据库管理和分析系统。一种涉及数百万欧元支出的整合程序被要求来生产这样一种产品。目前，似乎仅有几家大企业乐意开始这样的贡献。在此领域有一种可能、也最有理由值得争论的是：如果没有企业上的协助，大学机构不能表现出优秀的蛋白质组分析。

3.2 如果没有学术机构，企业能承担蛋白质组学研究吗？

尽管大制药公司的研究和开发预算是相当乐观的，但是它们不是无限的，且压力日益增加。例如，制药公司目前在美国必须产生大约 15%~20% 的盈利 (p.a.) 才能跟上道琼斯指数。事实上投资者期望能够年年实现利润回报，否则，他们将撤销其投资而转投其他地方。董事会成员必须通过研究和开发决定将来的投资，以及投资多少会使股东满意。在这种财政状况下，制药公司的有限研究经费也必须广泛分配到许多其他的可与蛋白质组竞争的高成本技术上。与之相关的是，蛋白质组学在各商业机构内的分配是否可能足够用来支撑这门目前需要大力优化的技术的发展。如果他们利用更有效的技术，将提高企业投资的有效性。历史上，许多新的技术是由学术机构最初启动和开发的。新技术被转到企业环境，在那里，它们被优化并用于专门的领域。尤其可能的是学术机构将开发出能提高企业效率的创新性技术和工艺，但前提条件是他们必须与合适的机构进行合作。为了实现这种转化过程，至关重要是建立和维持大量的具有专业能力的学术中心。这些中心额外的优势在于能提供有价值的公共信息。

3.3 目前蛋白质组学的研究尚不健全吗？某些人看到局限性

就目前蛋白质组技术而言，我们看到有许多缺点。并且在以下的部分，我们将详细

描述其阻碍蛋白质组研究的主要技术挑战。其他人也意识到这些问题 (Herbert *et al.*, 1997; Jungblut *et al.*, 1996; Wilkins *et al.*, 1996)。就我们目前的知识而言, 蛋白质组学的限制因素还未得到完全评估和充分认识。

蛋白的溶解性 如前所述, 许多疏水性蛋白和高分子量蛋白难以在 2-DE 凝胶图谱上出现 (Wilkins *et al.*, 1996, 1998)。近来在离液剂和表面活性剂的进步已经大大改变了这种状况; 然而问题远未解决 (Blisnick *et al.*, 1998; Chevallet *et al.*, 1998; Rabilloud, 1998, 1999; Rabilloud *et al.*, 1997)。下述观点适用于大约 85% 的可溶蛋白。

流通量 由于 2-DE 凝胶的内在变异性, 同一样品必须进行多次实验。此外, 如前所述 2-DE 的商业分析软件在自动模式下仅能部分有效地鉴定斑点和进行凝胶间的交叉配比。由于进行 2-DE 实验是耗时的, 这是达到高流通量的另一个主要的瓶颈。我们注意到, 利用不同样品在同一凝胶上电泳后得到的蛋白进行多颜色显示处理, 可大大减少这种瓶颈。差异同位素标记蛋白的应用前景也很乐观 (Langen *et al.*, 1998)。

灵敏度 目前除了像临床标本分析这一类材料来源有限的情况外, 大多数蛋白质组研究不是因敏感性而受限, 而是受限于 2-DE 凝胶分离和宽范围蛋白丰度的检测 (丰度的动态范围见下)。当对样品量为 2mg 的全细胞蛋白进行考马斯亮蓝染色, 对样品量为 100 μ g 蛋白进行银氨染色, 或者对样品量为 5ng 的高特异活性³⁵S-蛋氨酸代谢标记的蛋白, 通过磷屏图像分析所得的 2-DE 图谱是相似的。然而, 利用³⁵S-蛋氨酸标记 2mg 蛋白, 却没有优势。因为实际上, 整个凝胶提供了一种信号, 并且, 每个蛋白斑点间的差异消失了。因此, 敏感性在多数情况下是不受限的。实际上, 蛋白的上样量可因在一定敏感性下的凝胶分辨率的能力来优化, 这受限于蛋白丰度的线性动态变化, 也受限于凝胶系统的分辨率窗口的数目。为了评估这种概念的有效性, 应该考虑蛋白丰度的动态范围和凝胶分辨率的相关问题。

丰度的动态范围 一个中心问题在于, 根据实验和样品类型的不同, 目前用于蛋白质组研究的 2-DE 凝胶系统通常分离大约 2 000~10 000 个蛋白斑点 (Fey *et al.*, 1997; Gurg *et al.*, 1988; Jungblut *et al.*, 1996; Klose and Kobalz, 1995)。据预测, 人类基因组可编码 65 000~100 000 个功能基因 (borrebaeck, 1998)。其中的 5000~15 000 被认为是以具有多种 mRNA 的剪切形式和翻译后修饰的单细胞类型表达, 从一个细胞蛋白质组中所期望的不同斑点数目是 50 000 左右。这已经通过观察在一张 2-DE 凝胶上最多达 10 000 个蛋白来得以证实, 它们主要是看家蛋白和结构蛋白 (Jungblut *et al.*, 1996)。通常认为, 细胞内蛋白丰度的动态变化范围可能达 10⁷ 之高。也许, 每个细胞内仅有某些蛋白的几个分子 (如: 端粒酶?) 用来提供相应的生物学功能。而 2000 万以上的蛋白分子仅组成整个蛋白的 1%。因此, 一个理想的分析系统必须能够处理跨越 7 个数量级或更多的动态范围, 并且超过 50 000 个分离层次。我们估计, 目前仅有 5%~20% (2000~10 000/50 000) 的多数可溶性蛋白可通过 2-DE 分离和染色检测。

蛋白分辨率 如果 50 000 个斑点均匀分布在 20cm \times 20cm 的凝胶上, 并且斑点平均占据 2mm², 那么每平方毫米将出现 5 个斑点。因为根据分子量和等电点分配蛋白可以准确模拟倾斜曲线 (Cavalcoli *et al.*, 1997), 所以显然凝胶的某一区域将会被分辨率较差的丰度蛋白斑点聚集。这个问题可以用窄范围的 pH 进行等电聚焦而部分解决; 然而, 这种策略不能解决线性动态问题, 并且加重流通量的问题。

蛋白鉴定问题 利用传统的技术,许多实验室通常要求 picomolar (10^{-12} mol) 到数百个 femtomole 的蛋白量,以完成蛋白的鉴定。很明显,假设每个细胞含 50 000 个斑点,那么对于 5% 丰度的蛋白斑点是可能的。最近的一些参考文献上某些实验室声称,已经对 40 fmol 或 50 amol 为起始量成功进行常规的分析。然而,得到的图谱很不典型。

与“基因组学”方法比较 在阐述了蛋白质组的缺点后,将之与基因组学技术进行简单的比较是合适的。很大程度上由于 PCR 技术的扩增能力使得核酸分析在灵敏性上超过蛋白质研究。然而,应当注意到,高通量的筛选研究也因在 EST 和 SAGE 研究中需检测低丰度的 mRNA 而受限。当与放射性方法相比时,微阵列研究通常所用的荧光分子的检测灵敏性也受限。目前,仍不清楚多少比例的基因中可以通过芯片技术观察 (Bowtell, 1999; Carulli *et al.*, 1998; DeRisi and Iyer, 1999; Drmanac *et al.*, 1998; Iyer *et al.*, 1999; Lipshutz *et al.*, 1999; Watson *et al.*, 1998)。芯片技术能迅速经济地测定 mRNA 的丰度;然而,它不能提供蛋白质丰度的信息。预期将基因组和蛋白质组两种研究策略联合起来,对于获得认识细胞和整个有机生物的全局观点是必需的。很明显,将必须显著提高蛋白质组学技术。

3.4 学术机构的蛋白质组学研究对企业是有益的

前述提到蛋白质组学的局限性,可能有理由争论:在蛋白质组学达到一定阶段之前,因为企业能迅速而高效地筛选大量的细胞蛋白,就像目前他们筛选细胞的 mRNA 一样,所以是否需要相当多的蛋白质组技术进步。若实现这个目标,将提供较现代基因组学技术多得多的建设性的生物学信息。然而,每个商业组织必须为这种技术提高进行投资是不可能的。这任务最可能由公共基金支持的研究中心和一些研究所内的多学科研究部门来执行。

4. 大学的蛋白质组学研究提供许多优势

4.1 大学能对企业的资源进行迅速而灵活的反应,且价格合理

在财政改革的现代国际化氛围内,大公司持续地寻求与大学签订开发计划。企业已经认识到,大学能够从小到中等程度的资本投资中产生更多的效益。如果不建立一个全新的实验室及其相关设施,公司能给大学的短期、高风险的计划提供资金,并且若没有遭受过多的财政损失,则在某一合适的时间退出(如果必须的话)。建立合适的研究机构和培养合适专家的大学要对由企业而来的无论技术还是资金方面的流动利润待机而动。

4.2 大学愿意与临床样本接近

许多大学与世界级的临床研究部门有联系,可以使企业合作者愿意与生物学、医学专家进行前所未有的接近。优秀的病历和经验丰富的标本处理以确保高质量的结果,使相关的样品得以直接比较成为可能。对与临床医学的空间接近和一个可依赖的拥有临床专家的交流基地意味着:样品能在离开手术台数分钟内转到学术机构的实验室,或者被有效和系统地冻存以备后用。与医学专家的个人接触使精确的专业化要求成为可能,并且连续确保获得令人满意的质控。企业寻找的正是这种类型的机构,对于潜在的投资

或合作伙伴而言以最低消耗来维持。在劳动密集型的靶分子开发领域，企业与大学竞争确有可能。然而，无论应用在观察试验药物对临床I和II试验的影响时进行迅速的现场评估，还是应用在组织学家能检测的时间点之前组织移植排斥的启动，附属于专业临床医院的蛋白质组机构为企业合作者提供了一个理想的切入点。这种考虑的重要性显而易见。

4.3 大学不仅研究医学应用

值得人注意的是蛋白质组学的主要目的有可能使健康保险系统破产。除此之外，大学还提供仅有的手段，使得一些基本的生物能通过蛋白质组学进行检查。例如，通过直接应用先进的蛋白质组和基因组研究模式生物来获得对于胚胎和发育的更多理解。目前主要的制药企业认为用来研究这些系统的模式生物，例如果蝇、斑马鱼或蟾。这类计划可能成为学院蛋白质组中心的长处，并扩展和完全互补于人类生物学计划。

4.4 大学的基础研究将被公众接近

生物数据管理系统在今天存在的一个结果是值钱。生物信息能被转化为市场化的制药产品。这与公共数据库的信息沉淀不相容。企业所关注的是下载整个公共资源的内容，并利用它作为基础，依据它来扩大私人公司的数据库。由于蛋白质组分析是大规模筛选、快速流通量的主要部分，则基本上可能由企业界执行，因而许多有价值的生物信息将不能为公共所利用。

4.5 大学培养学生

基本的经济信条表明，任何高科技企业成功生存，一个合适的训练有素的工作组是必须的。这种培训的耗费相当可观，因此，对于公司能够招募合格的员工是有很大的优势的。教学是大学的主要任务，要将蛋白质组领域的技能熟练地传授给学生，大学必须积极参与蛋白质组领域的研究中来。的确，大学已经具有大量必需的基本设备和条件，并需要蛋白化学、细胞生物学和生物信息学间的直接合作，以面对共同的目标——蛋白质组学。

5. 大学的“让产易股 (spin-off)”公司

历史上，德国的大学并不擅长于将其专利成果转化为商业上的利益。现在已认识到这种状况必须改变。学院蛋白质组的另一方面是，我们在准备这部手稿的过程中，已经有了第一手的经验。因为蛋白质组潜在的经济重要性，所以在技术发展上存在强大的经济利益，正是这样，许多学院研究组产生“让产易股”公司，使他们获得的专长商品化。当考虑到某些基因组公司时，例如支撑这些公司（如 Hyseq, Affymetrix 和 Incyte）的技术都从学院研究组而出，这种活动的潜在经济意义变得很明显。我们看到同样的现象将在蛋白质组上发生，并且这本书的几个作者（包括我自己，ProteoMed GmbH 公司）正在尽力实现这个目标。

6. 结论

总之，与某些意见相反，我们相信学院的蛋白质组研究既值得做也是必须的，对于

企业和学院都能提供许多益处。从我们这方面看, 如果已经提供了大部分研究并使其不断发展的欧洲的大学放弃对这一强有力的技术的合适和连续的投资将是不幸的。企业界已经认识到蛋白质组研究提供了巨大、难以预测的潜在价值。至少其中部分必需技术的发展将可能由学术机构进行。这是学术机构为何要支持蛋白质组研究项目的最引人注目的原因。

(万晶宏译 贺福初校)

参考文献

- Andersen, H.U., Fey, S.J., Larsen, P.M., Nawrocki, A., Hejnaes, K.R., Mandrup-Poulsen, T. and Nerup, J. (1997) Interleukin-1beta induced changes in the protein expression of rat islets: a computerized database. *Electrophoresis* **18**: 2091–2103.
- Arsenian, S., Weinhold, B., Oelgeschlager, M., Ruther, U. and Nordheim, A. (1998) Serum response factor is essential for mesoderm formation during mouse embryogenesis. *Embo J.* **17**: 6289–6299.
- Blisnick, T., Morales-Betoulle, M.E., Vuillard, L., Rabilloud, T. and Braun Breton, C. (1998) Non-detergent sulphobetaines enhance the recovery of membrane and/or cytoskeleton-associated proteins and active proteases from erythrocytes infected by *Plasmodium falciparum*. *Eur. J. Biochem.* **252**: 537–541.
- Borrebaeck, C.A. (1998) Tapping the potential of molecular libraries in functional genomics. *Immunol. Today* **19**: 524–527.
- Bowtell, D.D. (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nature Genet.* **21**: 25–32.
- Burlingame, A.L., Boyd, R.K. and Gaskell, S.J. (1998) Mass spectrometry. *Anal. Chem.* **70**: 647R–716R.
- Cahill, M.A., Janknecht, R. and Nordheim, A. (1996) Signalling pathways: jack of all cascades. *Curr. Biol.* **6**: 16–9.
- Carulli, J.P., Artinger, M., Swain, P.M., Root, C.D., Chee, L., Tulig, C., Guerin, J., Osborne, M., Stein, G., Lian, J. and Lomedico, P.T. (1998) High throughput analysis of differential gene expression. *J. Cell Biochem.* **30–31** (Suppl.): 286–296.
- Cavalcoli, J.D., VanBogelen, R.A., Andrews, P.C. and Moldover, B. (1997) Unique identification of proteins from small genome organisms: theoretical feasibility of high throughput proteome analysis. *Electrophoresis* **18**: 2703–2708.
- Chaurand, P., Luetzenkirchen, F. and Spengler, B. (1999) Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **10**: 91–103.
- Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J. and Rabilloud, T. (1998) New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* **19**: 1901–1909.
- DeRisi, J.L. and Iyer, V.R. (1999) Genomics and array technology. *Curr. Opin. Oncol.* **11**: 76–79.
- Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J.D. and Drmanac, R. (1998) Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nature Biotechnol.* **16**: 54–58.
- Fey, S.J., Nawrocki, A., Larsen, M.R., Gorg, A., Roepstorff, P., Skews, G.N., Williams, R. and Larsen, P.M. (1997) Proteome analysis of *Saccharomyces cerevisiae*: a methodological outline. *Electrophoresis* **18**: 1361–1372.
- Görg, A., Postel, W. and Günther, S. (1988) The current state of two dimensional electrophoresis with immobilised pH gradients. *Electrophoresis* **9**: 531–546.
- Gygi, S.P., Han, D.K.M., Gingras, A.-C., Sonenberg, N. and Aebersold, R. (1999) Protein analysis by mass spectrometry and sequence database searching: tools for cancer research in the post-genomic era. *Electrophoresis* **20**: 310–319.
- Heidenreich, O., Neininger, A., Schrott, G., Zinck, R., Cahill, M.A., Engel, K., Kotlyarov, A., Kraft, R., Kostka, S., Gaestel, M. and Nordheim, A. (1999) MAPKAP kinase 2 phosphorylates serum response factor in vitro and in vivo. *J. Biol. Chem.* **274**: 14434–14443.

- Herbert, B.R., Sanchez, J.-C. and Bini, L.** (1997). Two-dimensional electrophoresis: the state of the art and future directions. In: *Proteome Research: New Frontiers in Functional Genomics* (ed. M.R. Wilkins, K.L. Williams, R.D. Appel, and D.F. Hochstrasser). Springer, Berlin, pp. 13–33.
- Hipskind, R.A., Rao, V.N., Mueller, C.G., Reddy, E.S. and Nordheim, A.** (1991) Ets-related protein Elk-1 is homologous to the c-fos regulatory factor p62TCF. *Nature* **354**: 531–544.
- Hochstrasser, D.F., Patchornik, A. and Merrill, C.R.** (1988) Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Anal. Biochem.* **173**: 412–423.
- Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C.F., Trent, J.M., Staudt, L.M., Hudson, J., Jr., Boguski, M.S., Lashkari, D., Shalon, D., Botstein, D. and Brown, P.O.** (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283**: 83–87.
- James, P.** (1997) Of genomes and proteomes. *Biochem. Biophys. Res. Commun.* **231**: 1–6.
- Jungblut, P., Thiede, B., Zimny-Arndt, U., Muller, E.C., Scheler, C., Wittmann-Liebold, B. and Otto, A.** (1996) Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* **17**: 839–847.
- Klose, J. and Kobalz, U.** (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**: 1034–1059.
- Kuster, B. and Mann, M.** (1998) Identifying proteins and post-translational modifications by mass spectrometry. *Curr. Opin. Struct. Biol.* **8**: 393–400.
- Langen, H., Fountalakis, M., Evers, S., Wipf, B. and Berndt, P.** (1998). ¹⁵N- and ¹³C- labeling of cells for identification and quantification of proteins on 2D gels. In: *From Genome to Proteome: 3rd Siena 2D Electrophoresis Meeting*, Siena, Italy.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J.** (1999) High density synthetic oligonucleotide arrays. *Nature Genet.* **21**: 20–24.
- Rabilloud, T.** (1992) A comparison between low background silver diammine and silver nitrate protein stains. *Electrophoresis* **13**: 429–439.
- Rabilloud, T.** (1998) Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis* **19**: 758–760.
- Rabilloud, T.** (1999) Solubilization of proteins in 2-D electrophoresis. An outline. *Methods Mol. Biol.* **112**: 9–19.
- Rabilloud, T., Adessi, C., Giraudel, A. and Lunardi, J.** (1997) Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **18**: 307–316.
- Watson, A., Mazumder, A., Stewart, M. and Balasubramanian, S.** (1998) Technology for micro-array analysis of gene expression. *Curr. Opin. Biotechnol.* **9**: 609–614.
- Wilkins, M.R., Sanchez, J.C., Williams, K.L. and Hochstrasser, D.F.** (1996) Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* **17**: 830–838.
- Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Bairoch, A. and Hochstrasser, D.F.** (1998) Two-dimensional gel electrophoresis for proteome projects: the effects of protein hydrophobicity and copy number. *Electrophoresis* **19**: 1501–1505.
- Yates, J.R. III** (1998) Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33**: 1–19.
- Zubritsky, E.** (1998) MALDI-TOFMS: biomolecules and beyond. *Anal. Chem.* **70**: 733A–737A.

第十五章 作为植物遗传和育种工具的蛋白质组学

Hervé Thiellement, Christophe Plomion and Michel Zivy

1. 引言

自从 O'Farrell (1975) 和 Klose (1975) 最初描述蛋白二维电泳 (2-DE) 技术以来, 植物生物学家已经用于包括遗传和生理学的多种研究。在一块胶上 (图 15.1) 显示几百种基因产物允许以蛋白方式检测基因型之间的差异, 并且以这种方式能够估计遗传距离, 建立系统发育关系。在发育和对各种处理反应中也研究了基因表达的差异, 而此过程中能够鉴定出受这些条件调控蛋白的表达。基因突变在蛋白水平上进行了分类并估计了多效性。编码鉴定出来的蛋白的基因被作图, 它们的蛋白显示出等电点或分子量的变异 (例如位置移动或 PS 变异, 图 15.2)。表达的基因能被加到以核酸为标记建立起来的遗传图谱中。另外, 由于软件的贡献 (Appel *et al.*, 1997), 一种蛋白的数量可以作为数量性状, 适合于在遗传图谱上定位控制这些数量变异的因素数量性状位点 (QTL)。



图 15.1 拟南芥叶蛋白的 2-DE 胶
一维在 pH3.5 到 10 非线性梯度的 IPG 胶条中; 二维 SDS 胶为丙烯酰胺/PDA 均匀胶, 12% T; 蛋白点用银染方法染色。

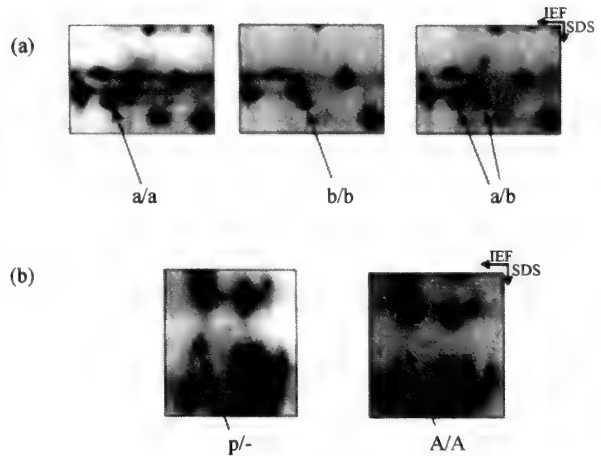


图 15.2
(a) 2-DE 胶的一部分, 显示海岸松 F2 代针叶的位移 (position shift, PS) 变异。a1, 纯合基因型 “a/a”; a2, 纯合基因型 “b/b”; a3, 杂合基因型 “a/b”。(b) 2-DE 胶的一部分, 显示海岸松 F2 代针叶的有/无 (presence/absence, P/A) 变异。b1, 表现型 ‘有’, 基因型 P/-; b2, 表现型 ‘无’, 纯合缺失基因型 “A/A”。第一维 IEF; 第二维 SDS。

在植物生物学中, 描述 2-DE 胶中的蛋白点的技术目前正在发展中。而且, 几种植物蛋白的数据库正在建立过程中。这种植物蛋白质组学的开端将具有重要的意义, 因为

第一个植物(拟南芥)的基因组测序将于2000年底全部完成,蛋白质组与基因组数据的整合将是非常重要的。近来发展起来的“转录组”工具,如高密度cDNA膜、cDNA微阵列或DNA芯片(Castellino, 1997; Lemieux *et al.*, 1998; Marshall and Hodgson, 1998),与蛋白质组学结合将是非常有用的,尽管蛋白质的数量和它的mRNA经常不相关也是值得注意的(Anderdon and Anderson, 1998; Anderson and Seilhammer, 1997; Haynes *et al.*, 1998)。另外,既然蛋白的更新和翻译后修饰不能在cDNA和mRNA水平研究,将来的蛋白质组学是并将持续是非常重要的。

最近出版了几篇关于植物蛋白质组学的综述(de Vienne *et al.*, 1996, 1999; Thiellement *et al.*, 1999),这里将描述蛋白质组学在植物生物学中几种应用的范例,并重点讲述植物蛋白质组学导致的植物遗传学的进展。

2. 遗传多样性分析

2.1 种内和种间的遗传差异

自然群体中遗传变异的程度总是群体遗传学家,进化学家和植物育种家最感兴趣的课题。尽管同工酶和DNA为基础的标记已经并正在被广泛使用,2-DE显示的蛋白很少被用于群体遗传学的研究。但是,2-DE能发现比同工酶电泳更多的标记,Vienne *et al.* (1996)报道2-DE在每个多态性位点能够揭示与同工酶相同的等位变异。与DNA标记相比,2-DE揭示的遗传变异性只涉及表达的基因。在最近的20年中,人们做了许多用2-DE方式在多种分类水平评价遗传差异的实验,下面详细叙述几个范例。

种间的变异性 栽培小麦是多倍体,有两个(硬粒小麦, *Triticum turgidum*, AABB)或三个(软小麦, *Triticum aestivum*, AABBDD)基因组。这些基因组起源于从那时起已经趋异的一个共同祖先,现在仍然存在许多二倍体或多倍体的野生品种,它们具有不同的所谓“部分同源”的基因组,且能够程度不同地杂交。小麦遗传学家根据减数分裂中二价染色体的数目推测栽培小麦不同的基因组可能起源于哪一种现有品种。A基因组起源于 *Triticum urartu*, D基因组起源于 *T. taushii* (Kerby and Kuspira, 1987)。但是,对于哪一个品种的 *Sitopsis* 区段有贡献于B基因组及栽培小麦的细胞质现在仍然存在争议。为了解决这个问题,对不同品种的这一区段在2-DE水平上(*T. longissimum*, *T. sharonense*, *T. bicorne*, *T. searsii* 和 *T. speltoides*)进行了交叉对比并与面包小麦栽培种——中国春(Chinese spring, CS)进行了比较。通过分析共有的蛋白点的数量计算出了每对基因型间的相似性指数。从相似性结果绘制出反应在 *Sitopsis* 区段系统发育关系的树状图。不仅以2-DE为基础的树状图和传统的分类学近乎完美一致,而且还发现 *T. speltoides* 是与栽培小麦B基因组最为相近的野生品种(Bahrman *et al.*, 1988a)这个发现通过分析细胞质蛋白得到了进一步验证:核酮糖二磷酸羧化酶-加氧酶的大亚基和两种形式的 β -ATP酶在中国春和 *T. speltoides* 中有相同的等位形式,而其他的 *Sitopsis* 品种则具有其他的这些叶绿体基因的等位位点(Bahrman *et al.*, 1988b)。

通过比较2-DE方式对系统发育关系的研究已经延伸到携带不同基因组的小麦属品种(Thiellement *et al.*, 1989)。在 *Triticeae* 属不同品种间的另一个实验中,也就是小

麦的 A 和 D 基因组, 大麦的 H 基因组和黑麦的 S 基因组之间得到的树状图很好地反映了已知的系统发育关系 (Zivy et al., 1995), 但是小麦和大麦间共有的蛋白点数目降到 30%, 说明已达到了这种方法的极限。

在欧洲偃麦中, Barreneche 等 (1996) 检测了两个非常相近的品种 *Quercus petraea* 和 *Quercus robur* 之间的关系。他们用 2-DE 的方法, 通过比较部分覆盖欧洲白偃麦自然地理范围的来自 6 个欧洲国家的 23 种偃麦来研究遗传差异。评价了来自幼苗总蛋白中的 530 个点, 其中 101 个是多态性的。两个品种的差异性是 0.36, 而 *Quercus petraea* 种内差异性为 0.35, *Quercus robur* 为 0.33。这种非常相近的种间和种内距离进一步证实了这两个品种之间较低水平的遗传差异性, 正如在同工酶、随机扩增多态性 DNA (RAPD) 和叶绿体 DNA 中所报道的。

种内的变异性 在海岸松中, Bahrman 等 (1994) 用 2-DE 的方法研究自然地域 7 个群体 (原产地) 之间的关系, 评价在不同地理起源内或之间存在的遗传变异性。利用雌配子体 (一种包被针叶松种子胚胎的单倍体营养组织) 中蛋白位点可能存在的不同等位形式, 共评价了 968 个蛋白位点, 其中 84% 是变异性的。在这些信息的基础上, 三个主要的松树群体 (即大西洋、地中海和北非群体) 可以被区分出来, 遗传结构与萜烯数据相符。在另外一个研究中, Petit 等 (1995) 证实对萜烯和同工酶、2-DE 揭示的蛋白三者群体中的遗传差异是相同的, 说明在三种类型的位点中均缺乏 (或同水平) 选择。

Burstin 等 (1994) 通过用 2-DE, 同工酶和 RFLP 标记研究 21 个玉米近交系间的遗传多样性。从 21 个酶位点、59 个 RFLP 探针、70 个位置移动位点 (PS) 计算了距离, 比较了 103 个蛋白的数量变异及共祖率系数的相关性。在 PS 和同工酶位点间每个多态性位点的等位位点没有差异 (各为 2.2 和 2.4), RFLP 位点的稍高 (4.6)。在相关的品系中, 根据同工酶位点, 除了一个, 其余所有的距离都与共祖率系数高度相关。这很有可能是因为只检测了较小数量的同工酶位点: 当把所有的 PS、同工酶和 RFLP 位点都考虑进去时的共祖率系数相关性 (0.85) 比单独考虑某一项时要高。在不相关的品系中, 多肽的数量变异与其他标记之间的相关性下降到 0.14。几种原因可以用说明这个结果: ①进化的压力在控制蛋白数量的基因上和其他位点有所不同; ②上位性和多效性相互作用能导致变异蛋白的数量与控制它们数量的多态性基因间存在非线性相关; ③当只考虑相关品系时, 单基因标记的重要相关性可能因为连锁不平衡。

David 等 (1987) 研究来自于一个小麦群体在不同地点经过 8 年栽培独立进化的 11 个群体间的遗传差异。在这个分析中, 162 个多肽中的 39 个被证明在群体间存在多态性。多变异分析表明, 不同地点进化的群体起源于同一个群体, 造成这些差异的根源是自然选择而不是随机漂移。

2.2 品种、品系和栽培品种的区别

20 世纪 80 年代, Zivy 等 (1983, 1984) 通过 2-DE 用蛋白间质量和数量的差异区分出来 3 个不同的小麦品种。用异质品系 (通过外来细胞质的连续回交得到的同核基因组) 揭示细胞质编码基因 (例如核酮糖二磷酸羧化酶-加氧酶的大亚基) 的差异。二维电泳技术的改进, 尤其是从植物组织中提取蛋白过程的改进 (Bahrman et al., 1985;

Colas des Francs *et al.*, 1985) 以及大一些尺寸胶的使用提高了遗传变异的检出数 (Damerval *et al.*, 1986)。在 2-DE 电泳胶中尽管只有部分编码区的突变是可检出的, 大量的基因产物还是允许研究者准确地分析及分辨基因型, 即使它们属于相关的群体。这在小麦 (Anderson *et al.*, 1985; Dunbar *et al.*, 1985; Picard *et al.*, 1997)、大麦 (Görg *et al.*, 1988, 1992)、水稻 (Abe *et al.*, 1996; Saruyama and Shinbashi, 1992)、甘蔗 (Ramagopal, 1990) 和辣椒 (Poch *et al.*, 1992, 1994) 中均有报道。在木本植物中, Bon (1989) 应用 2-DE 进行分生组织培养分离巨杉、北美红杉、花旗松、欧洲云杉、海岸松、桉树和 *Populus nigra* 的克隆。

当用 2-DE 检测遗传差异时, 研究者可找到足够的差异点来实现其的目标。在许多情况下, 获得的数据允许建立与用其他基于分子或表型特征估计的相一致的遗传距离。

3. 基因组的表达

3.1 突变特征

比较野生型与突变型 2-DE 的模式可提供对突变效果的评价。例如, 它能够检测前面述及的一个突变的多样性。而且, 它能够鉴定出突变基因编码或受其影响的蛋白。在模式植物拟南芥中, 用 2-DE 方式检测了一系列影响第一阶段发育相关的突变 (Santoni *et al.*, 1994)。一种蛋白的数量, 即肌动蛋白的同型, 与下胚轴的长度相关。在番茄中, 野生型和铁缺乏突变型二维电泳方式的比较导致了几种数量不同的酶的鉴定, 这几种酶参与无氧代谢和胁迫防御 (Herbik *et al.*, 1996)。在藓类植物的实验中, Kasten 等 (1997) 比较叶绿体突变与野生型之间的蛋白差异, 是否与细胞分裂素互补。结论是激素对胞核和胞质编码的蛋白均有影响。对一系列影响拟南芥早期发育的突变和激素处理野生型蛋白方式的分析, 阐明了一种与预期非常相符的生化分类法, 未分类的突变可能导致细胞分裂素的超量表达 (Santoni *et al.*, 1997)。这种假设后来被细胞分裂素的剂量进一步证实 (Faure *et al.*, 1998)。比较几种晚花期拟南芥的突变发现, 几个蛋白点相对于野生型出现或消失 (Tacchini *et al.*, 1995)。但是, 在野生型和突变型杂交的 F₂ 代群体中, 没有一个变异蛋白点与花的表现型共分离。除了遗传证实的必要性, 这些实验说明 2-DE 分析能够揭示用于突变产生实验中的生态型或品系的遗传异质性。但是, 必须注意蛋白质组分析的蛋白只是总蛋白的一部分。

蛋白质组检测突变的能力进一步被研究突变的多样性证实。Gottlieb 和 de Vienne (1988) 比较了豌豆的两个近等基因系, 它们携带的决定圆粒和皱粒的 *r* 基因不同。在成熟种子的蛋白质组中, 近 10% 的蛋白点在数量上不同, 证实了这两种类型的种子之间的许多已知的生理差异。在玉米中, Damerval 和 de Vienne (1993) 研究了 *Opaque2* (*O2*) 基因的多样性, 该基因编码一种转录因子。通过对在几个无关背景中的 *O2* 和野生型玉米品系 2-DE 结果的比较, 鉴定出了 *O2* 基因特异的靶点。几种属于不同途径代谢酶的发现进一步证实 *O2* 是一种连接不同谷物代谢途径的调控基因 (Damerval and Le Guilloux, 1998)。

因此, 很明显蛋白质组学确实是区分基因型的强有力工具, 即使是在遗传背景相差无几的情况下。因为这种分析与 DNA 序列并非紧密相关, 在遗传差异非常小的时候也

可能观察到许多差异。同时，这能允许人们去解释单个突变的多种效应。另外，只有在蛋白水平才适合观察翻译后的修饰，而这也是遗传变异的主题。

3.2 器官和发育阶段的变异性

几个研究表明，器官特异的蛋白比非器官特异的蛋白有更大的变异性，并且遗传变异的水平依赖于组织或器官。在海岸松中，Bahrman 和 Petit (1995) 检测了 18 个不相关树种的 3 个器官（针叶，芽和花粉）蛋白的遗传变异性。在 902 个多肽中，通过评价有无变异，染色强度和位置移动变异，27.2% 是多态性的。在 3 个器官共有的点中，只有 18.1% 是变异性的，而器官特异的多肽有很高水平的多态性（平均 56%，针叶特异的点 44.4%，花粉 58.1%，芽 70.4%）。

器官特异的高水平变异性也在玉米中有所发现。通过比较两个近交系间 3 个器官（中胚轴、第二叶的叶鞘和三周龄幼苗的叶耳），Leonardi 等 (1988) 发现 357 种蛋白中遗传数量和质量的变异为 3.6%，器官间没有变异性，而 629 种蛋白中 22.6% 表现为器官特异的变异。然后，这种比较延伸到两种另外的基因型（de Vienne *et al.*, 1988）：59% 的表现器官间差异的蛋白是遗传变异的，18% 为稳定的蛋白。这个假设也被海岸松的结果证实：器官特异的点位置移动变异的频率是分析每一种器官的两倍，而数量则为五倍（Bahrma and Petit, 1995）。

降低的“功能约束”（Kimura, 1983），如 Klose (1982) 所说，可能也解释了与“管家”泛在蛋白相比，在单细胞环境中表达的器官特异性多肽其等位基因变异水平的提高。

3.3 非生物胁迫反应蛋白的鉴定与分析

植物对非生物胁迫的反应引起了人们极大的关注，主要因为应用于栽培品种的可能性。

热休克蛋白 在植物中，象其他有机体一样，升高温度能诱导热休克蛋白（HSP）的合成。它们的合成对耐热性的获得是相关的，也就是忍耐较高温度的能力。植物不同于其他有机体之处在于它们能合成大量的低分子质量 HSP（LMW-HSP），例如，Nover 和 Scharf (1984) 发现了在番茄细胞培养中 48 种 HSP 的诱导。人们通过比较不同处理的效果或研究在发育过程中的诱导，研究耐热性与 HSP 合成的关系：Albernathy *et al.* (1989) 认为 HSP 的合成与第一阶段小麦种子吸涨过程中耐热性的演变相关。人们也研究了不同细胞器 HSP 的合成：Lund 等 (1998) 证实了在玉米线粒体中一种特殊的 LMW-HSP（HSP22）被强烈诱导合成。

鉴于 HSP 具有如下特征：① HS 处理诱导后很容易鉴定（鉴定它们不需要测序）；② 用 2-DE 方式检测，可以放射性标记，也可以银染；③ 一个优先知道的与耐热性相关的蛋白，就早用遗传研究调查 HSP 多态性与耐高温遗传变异的关系，Zivy (1987) 调查了小麦中 HSP 的遗传变异性，检测出了很高水平的多态性：在 3 个近交系 35 个被检测的 HSP 中，超过三分之一属于质量或数量变异，而同一基因型只有 13% 的其他蛋白是多态性的。然后，El Madidi 和 Zivy (1993) 发现热耐受不与 HSP 的质量变异相关，而与两种 LMW-HSP 的数量变异相关，且后者普遍存在于 7 个检测的基因型中。

人们为了寻找特定 HSP 的表达与耐热性遗传变异的合理关系进行了遗传研究 (Krishnan *et al.*, 1989; Vierling and Nguyen, 1990)。Vierling 和 Nguyen (1992) 只检测了单粒小麦耐热和热敏株系之间的数量变异。然后, Jorgensen 和 Nguyen (1995) 研究了 LMW-HSP 在玉米幼苗中的遗传特征, 结果检测出很高水平的多态性: 在这两系中, 29 个被检测的 LMW-HSP 中有 15 个属于质量变异, 3 个表现为数量变异。

冷适应蛋白 暴露在寒冷条件下能够诱导冻结耐受, 但是这种生化机制可能与高温耐受的机制不同, 因为没有特异的已知蛋白是被寒冷诱导的。因此, 2-DE 被用于描述寒冷反应蛋白, 鉴定与寒冷适应相关的调控蛋白: Meza-Basso 等 (1986) 发现了在油菜幼苗中受低温诱导的蛋白。然后 Guy 和 Haskell (1988) 检测了菠菜幼苗中与冷适应相关的蛋白, 发现这些蛋白的合成在获得冻结耐受性期间有所升高, 而在再适应过程中减少。Cabane 等 (1993) 在大豆中鉴定了冷适应相关蛋白, 其中一个为热休克蛋白 (HSP70)。Van Berkel *et al.* (1994) 检测了马铃薯块茎冷处理反应过程中 26 种蛋白数量的改变。Sabehat 等 (1996) 证明番茄的果实通过热休克处理获得的长期冷耐受与 HSP 的持久性相关。Danyluk 等 (1991) 首先用遗传途径研究了 3 个小麦品种对冷处理的反应: 18 种蛋白被瞬时诱导, 另外有 53 种蛋白在诱导冻结耐受所需的四周时间内持续高水平表达。而后者中的 34 种在冻结耐受最好的品系中也有较高水平的表达。

抗旱相关蛋白 缺水引起植物许多形态学、生理学和生物化学反应, 例如, 减少气生部分的生长, 气孔关闭, 叶子卷曲, 叶衰亡和渗透调整。这些反应至少部分受脱落酸 (ABA) 控制, 即一种当植物遭受干旱时浓度升高的植物激素。许多细胞的功能都受影响, 并发现该胁迫诱导出了不同类型的蛋白。其中一些的功能直接与水胁迫有关 (如水孔蛋白), 一些与细胞损伤有关 (如 HSP, 与氧化胁迫、蛋白酶和抑蛋白酶相关), 还有一些与不同的代谢功能相关 (Ingram and Bartels, 1996)。第一个被鉴定出来的水分胁迫蛋白来自于干燥的胚胎 (Aspart *et al.*, 1989; Dure *et al.*, 1981; Sanchez-Martinez *et al.*, 1986)。这类蛋白被称为胚胎形成晚期丰富蛋白 (LEA)。它们的功能未知, 但高度亲水, 并根据重复区段的排列分成不同的组 (Dure *et al.*, 1989)。其中一些后来被证明是不同器官受干旱、ABA 或盐度胁迫诱导的结果。其中的一个, 玉米的 RAB17 (组 2 ABA 反应性 LEA 蛋白) 被证明在核内和胞质内磷酸化的方式不同 (Jensen *et al.*, 1998; Goday *et al.*, 1994)。

在一些情况下, 2-DE 不是用来鉴定蛋白, 而是观察不同条件下蛋白改变的首选途径 (Chen and Tabaeizadeh, 1992; Hurkman and Tanaka, 1988; Ramani and Apte, 1997)。Ramagopal (1987) 证明大麦幼苗的根和茎对盐度胁迫的反应是不同的。Leone 等 (1994) 证明在马铃薯细胞悬浮时严重的渗透胁迫诱导不同系列的蛋白。Leymarie *et al.* (1996) 用植物生长素不敏感突变体研究了植物生长素对拟南芥干旱反应性蛋白的效应。

在其他的一些研究中鉴定了诱导出的蛋白。例如, Claes 等 (1990) 根据水稻中盐胁迫诱导的多肽段微测序的结果设计 DNA 引物, 分离出了一个富含甘氨酸蛋白的 cDNA 序列。同样地, Reviron (1992) 在菠菜叶子中鉴定出了一种受干旱诱导的蛋白酶抑制剂, Moons (1995, 1997) 在水稻中通过微测序和 Western blotting 鉴定出了一种新型的富含甘氨酸的蛋白, 组 3 LEA 蛋白, 一系列的组 2 LEA 蛋白 (也叫脱水素, 如

RAB17)。在 3 个菠菜的品种中，对盐度耐受性最高者合成出最大数量的这些蛋白。Rey 等 (1998) 证明在马铃薯叶绿体中通过水分胁迫诱导出一种硫氧还蛋白-样蛋白，Moons 等 (1998) 从水稻根中诱导出丙酮酸正磷酸双激酶。Costa 等 (1998) 发现在海岸松针叶中通过干旱诱导出 38 种蛋白。其中包括涉及氧化胁迫反应，热休克反应和木质素生物合成的酶类。Riccardi 等 (1998) 在玉米叶子的生长部分发现受干旱影响的 78 种蛋白，其中 50 种是上调的 (图 15.3)。19 种蛋白得到了部分序列信息，使涉及几种代谢途径蛋白的鉴定成为可能。

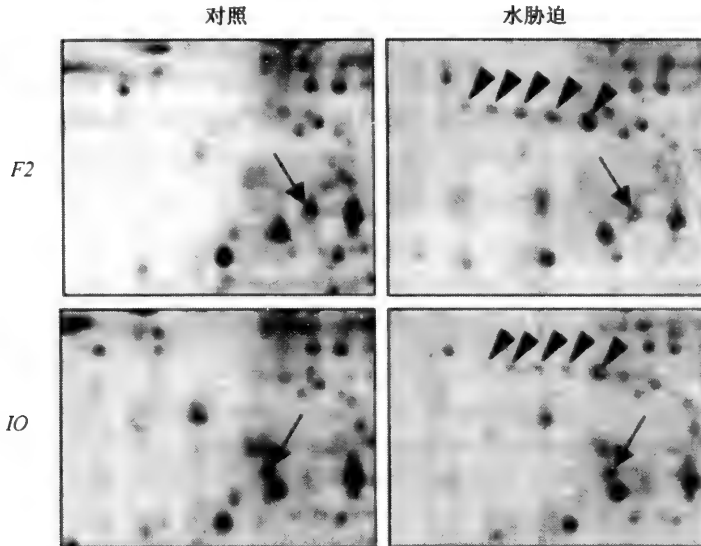


图 15.3 在两个玉米近交系 F2 和 I₀ 根中受水分胁迫诱导的蛋白
小箭头指示一串被诱导的蛋白，在 I₀ 中这一串蛋白质的表观分子量要比 F2 中的稍高些，显示所有的这些蛋白可能都被一个多态性基因编码，只是翻译后修饰不同。大箭头指示非诱导蛋白的位移。

4. 遗传作图和候选蛋白

近几年分子遗传方法在遗传作图和数量性状分离分析 (QTL 作图) (Kearsey and Farquhar, 1998) 中的应用有了巨大的进步。尽管用 DNA 为基础的技术能够快速遍及基因组，2-DE 则不但提供表达基因组作图的有效分子标记，还可能为理解 QTL 生物功能提供“候选蛋白”。

4.1 蛋白标记的遗传作图

代表质量变异的点，或是位置移动 (PS) 或有无 (P/A) (图 2)，已被证明是受单基因控制的 (de Vienne *et al.*, 1996)。PS 变异是指两个有相同特征和染色强度的点，可用以下的标准来鉴定：①它们在 2-DE 胶上的位置较近，有相近的 pI 和 (或) 表观分子量；②它们在分离谱系中以单基因共显性的孟德尔方式遗传 [在单倍体和双单倍体品系中以 1:1 分离 (Plomion *et al.*, 1995; Zivy *et al.*, 1992)；在 F2 自交群体中以

1:2:1分离 (Costa, 1999; Damerval, 1994; Plomion, 1997)]; ③它们对应于同一个蛋白。最后一条标准已在玉米和松中一对等位位点的微量测序得到进一步证实 (Plomion, 1997; Touzet, 1995a)。如果 PS 变异对应于蛋白初级结构的等位基因变异已经很明了, 这种显示 P/A 变异的点的平行性就不是很直接了。在 F₂ 代的 2-DE 胶中, 这种蛋白显示明显的单基因显性遗传方式 (一个点的有无以 3:1 的比例分离), 说明这种多态性是单基因引起的。如果 P/A 变异与 PS 变异 (许多等位变异是检测不到的) 相对应, 那么 P/A 位点就对应于蛋白的结构基因。反过来, 如果一个 P/A 变异对应于一个不能检测的点的数量差异, 那么 P/A 位点对应一个控制蛋白表达水平的主要调控基因。

在海岸松中, Bahrman 和 Petit (1995) 证实 PS 位点变异的水平与分子量成正相关, 也就是说高分子量蛋白的变异性要比低分子量的变异性大的多。这样, 大的一些蛋白就会出现较高水平的多态性, 因为“每个多肽的替换率的变异取决于氨基酸数目或每个位点替换率的不同” (Nei, 1987)。但是, 没有发现 P/A 变异的水平与分子量相关。这两个研究 (单基因方式的遗传以及与分子量的不相关) 暗示 P/A 变异可能来自于主效调控元件的多态性。

蛋白标记的遗传定位主要在小麦、玉米和海岸松中有报道, 尽管也有几个 PS 位点在其他作物中有所报道, 包括大麦 (Zivy *et al.*, 1992) 和豌豆 (de Vienne *et al.*, 1996)。在小麦中, Colas des Francs 和 Thiellement (1985) 报道了对比整倍体和双末端体 35 种蛋白的染色体定位。在海岸松中, Bahrman 和 Damerval (1989) 以及 Gerber *et al.* (1993) 各自报道了 119 和 65 个位点的连锁分析。近来用来自于该物种第三代近交谱系单倍体 (Plomion *et al.*, 1995) 和二倍体 (Costa *et al.*, 1999; Plomion *et al.*, 1997) 组织的 68 种蛋白进行作图。在玉米中, de Vienne *et al.* (1996) 提供了显示 65 个 PS 位点分布的多系连锁图。在松树和玉米中, 能在每一条染色体上找到蛋白位点 (图 15.4), 与其他标记 (玉米中的 RFLP, 松树中的 RAPDs 和 AFLP) 散在分布。

RFLP 标记由 cDNA 文库发展而来, 一般揭示数个不连锁位点 (Devey *et al.*, 1994; Helentjaris *et al.*, 1988; Song *et al.*, 1991), 鉴定表达的基因, 而对于沉默的基因和假基因的鉴定是非常困难的。因此, 对于构建表达基因图谱和建立在不同组织或不同发展阶段哪一个 cDNA 位点在表达, 用 2-DE 方法分离蛋白是非常有趣的技术。例如, 一个玉米蛋白表现为叶、种子和胚芽鞘的 PS 变异, 用微量测序和蛋白成分分析鉴定为磷酸甘油酸变位酶。用于 RFLP 分析的 cDNA 探针作图到 3 条染色体上 4 个不同的位点。每个器官中的 PS 位点只与其中的一个 (也许为表达的一个) 共定位 (de Vienne *et al.*, 1999)。

到现在为止, 单个植物品种中作图的蛋白位点不超过 100 个。这种标记的数目将在以后随着分析同一个体一些生理学上对照的器官或组织 (根、叶、芽、花粉和木质部等) 而增加, 直到遍及整个基因组。这种表达基因图谱将是 QTL 分析中候选基因/蛋白策略的无以比拟的支撑。

4.2 蛋白数量位点 (PQL) 和候选蛋白

候选基因。在近 10 年中, 鉴定出有重要农学性状遗传变异的基因主要是通过分离群体中用无名标记做的连锁图分析, 这种图的连锁不平衡能被最大化 (Tanksley,

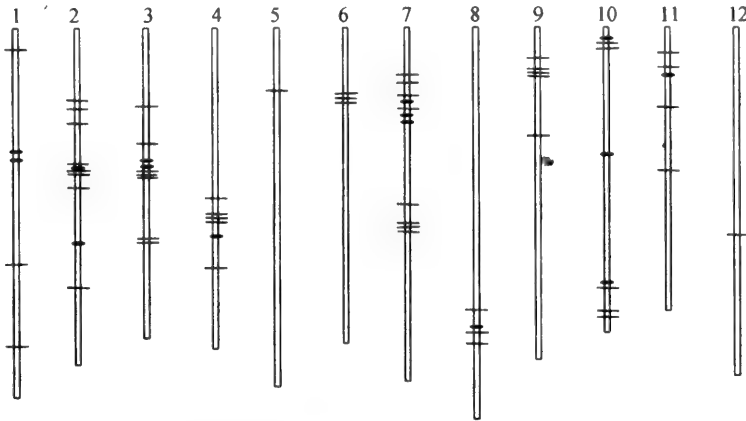


图 15.4 来自与松树雌配子体和针叶 2-DE 胶的 15 个位置变异 (-) 和 53 个有/无变异 (-) 在图谱上的位置模式图

该图用 RAPD 和 AFLP 标记构建, 可通过因特网查询 (<http://www.pierroton.inra.fr/genetics/pinus/>).

1993)。尽管很成功, 但应指出这种 QTL 检测分析存在一定的局限性。QTL 只能被大致定位, 也就是说遗传图谱上的 QTL 的置信区间是很大的 (Mangen *et al.*, 1994)。这意味着 QTL 作图不允许鉴定潜在基因。因此, 性状分割分析, 尽管已应用于标记辅助育种程序中, 可以被认为是鉴定控制部分数量性状变异的第一步。另外, 在总基因组长度和 DNA 含量的基础上, 相当于 1 厘摩 (cM) 的平均物理图距是变化的, 如拟南芥、桉树、玉米、小麦和松树分别是: 230、460、2500 和 12000Kb。这意味着在大部分一年生植物森林树种中用定位克隆 (Rommens *et al.*, 1998) 鉴定 QTL 是非常困难的。从生物学角度, 可以从候选基因 (CG) 和候选蛋白 (CP) 途径获得 QTL。

“功能 CG”是指可能影响性状表达并且测序的基因, 可以基于已知影响目标性状的生化和发育途径标准来推断。另外, 任何一个产生很大影响的突变位点可以被认为是一个 CG, 而这个 CG 的等位位点的效应较小, 可能影响自然发生的数量变异。例如, Touzet 等 (1995b) 推测一个矮化基因涉及赤霉素生物合成途径, 作为高秆生长的 QTL。QTL 和这个突变质量性状的共定位已有报道 (Beavis *et al.*, 1991; Doebley and Stec, 1993)。这种“功能 CG”的主要局限性在于目前只有一小部分基因影响的性状是已知的。

“位置 CG”可以推测所有与目标性状 QTL 定位在同一位置的基因。但是, 该策略是以比较完善的基因图谱为基础的, 只可以应用于少数几个物种。QTL 和已知基因的共定位以前曾有报道 (Byrne *et al.*, 1996; Causse *et al.*, 1995; Goldman *et al.*, 1994; Prioul *et al.*, 1997, 1999)。“位置 CG”的主要局限性在于 CG 和 QTL 的共定位纯属偶然。

蛋白数量位点作为正式候选基因的可能的方法 CG 的证实在这种信息应用于育种程序以前是非常必要的。其证实依赖于功能基因组学工具 (Borchez and Hofte, 1998), 其中 CP 途径是必不可少的步骤。确实, 如果一个 CG 的产物不显示数量和 (或) 活性

1:2:1分离 (Costa, 1999; Damerval, 1994; Plomion, 1997)]; ③它们对应于同一个蛋白。最后一条标准已在玉米和松中一对等位位点的微量测序得到进一步证实 (Plomion, 1997; Touzet, 1995a)。如果 PS 变异对应于蛋白初级结构的等位基因变异已经很明了, 这种显示 P/A 变异的点的平行性就不是很直接了。在 F₂ 代的 2-DE 胶中, 这种蛋白显示明显的单基因显性遗传方式 (一个点的有无以 3:1 的比例分离), 说明这种多态性是单基因引起的。如果 P/A 变异与 PS 变异 (许多等位变异是检测不到的) 相对应, 那么 P/A 位点就对应于蛋白的结构基因。反过来, 如果一个 P/A 变异对应于一个不能检测的点的数量差异, 那么 P/A 位点对应一个控制蛋白表达水平的主要调控基因。

在海岸松中, Bahrman 和 Petit (1995) 证实 PS 位点变异的水平与分子量成正相关, 也就是说高分子量蛋白的变异性要比低分子量的变异性大的多。这样, 大的一些蛋白就会出现较高水平的多态性, 因为“每个多肽的替换率的变异取决于氨基酸数目或每个位点替换率的不同” (Nei, 1987)。但是, 没有发现 P/A 变异的水平与分子量相关。这两个研究 (单基因方式的遗传以及与分子量的不相关) 暗示 P/A 变异可能来自于主效调控元件的多态性。

蛋白标记的遗传定位主要在小麦、玉米和海岸松中有报道, 尽管也有几个 PS 位点在其他作物中有所报道, 包括大麦 (Zivy *et al.*, 1992) 和豌豆 (de Vienne *et al.*, 1996)。在小麦中, Colas des Francs 和 Thiellement (1985) 报道了对比整倍体和双末端体 35 种蛋白的染色体定位。在海岸松中, Bahrman 和 Damerval (1989) 以及 Gerber *et al.* (1993) 各自报道了 119 和 65 个位点的连锁分析。近来用来自于该物种第三代近交谱系单倍体 (Plomion *et al.*, 1995) 和二倍体 (Costa *et al.*, 1999; Plomion *et al.*, 1997) 组织的 68 种蛋白进行作图。在玉米中, de Vienne *et al.* (1996) 提供了显示 65 个 PS 位点分布的多系连锁图。在松树和玉米中, 能在每一条染色体上找到蛋白位点 (图 15.4), 与其他标记 (玉米中的 RFLP, 松树中的 RAPDs 和 AFLP) 散在分布。

RFLP 标记由 cDNA 文库发展而来, 一般揭示数个不连锁位点 (Devey *et al.*, 1994; Helentjaris *et al.*, 1988; Song *et al.*, 1991), 鉴定表达的基因, 而对于沉默的基因和假基因的鉴定是非常困难的。因此, 对于构建表达基因图谱和建立在不同组织或不同发展阶段哪一个 cDNA 位点在表达, 用 2-DE 方法分离蛋白是非常有趣的技术。例如, 一个玉米蛋白表现为叶、种子和胚芽鞘的 PS 变异, 用微量测序和蛋白成分分析鉴定为磷酸甘油酸变位酶。用于 RFLP 分析的 cDNA 探针作图到 3 条染色体上 4 个不同的位点。每个器官中的 PS 位点只与其中的一个 (也许为表达的一个) 共定位 (de Vienne *et al.*, 1999)。

到现在为止, 单个植物品种中作图的蛋白位点不超过 100 个。这种标记的数目将在以后随着分析同一个体一些生理学上对照的器官或组织 (根、叶、芽、花粉和木质部等) 而增加, 直到遍及整个基因组。这种表达基因图谱将是 QTL 分析中候选基因/蛋白策略的无以比拟的支撑。

4.2 蛋白数量位点 (PQL) 和候选蛋白

候选基因。在近 10 年中, 鉴定出有重要农学性状遗传变异的基因主要是通过分离群体中用无名标记做的连锁图分析, 这种图的连锁不平衡能被最大化 (Tanksley,

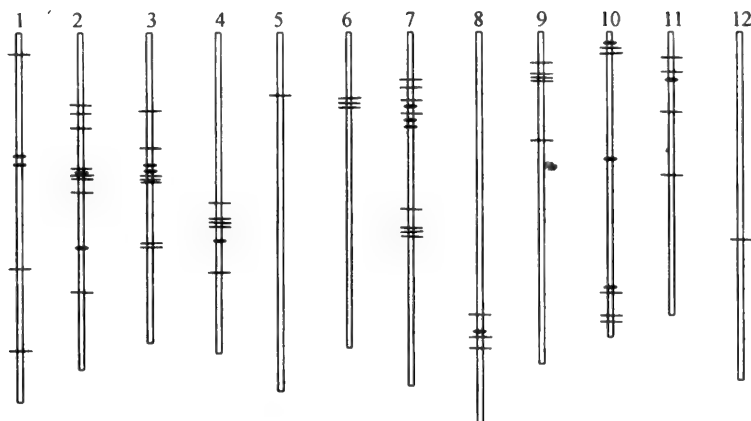


图 15.4 来自与松树雌配子体和针叶 2-DE 胶的 15 个位置变异 (-) 和 53 个有/无变异 (-) 在图谱上的位置模式图

该图用 RAPD 和 AFLP 标记构建, 可通过因特网查询 (<http://www.pierroton.inra.fr/genetics/pinus/>).

1993)。尽管很成功, 但应指出这种 QTL 检测分析存在一定的局限性。QTL 只能被大致定位, 也就是说遗传图谱上的 QTL 的置信区间是很大的 (Mangen *et al.*, 1994)。这意味着 QTL 作图不允许鉴定潜在基因。因此, 性状分割分析, 尽管已应用于标记辅助育种程序中, 可以被认为是鉴定控制部分数量性状变异的第一步。另外, 在总基因组长度和 DNA 含量的基础上, 相当于 1 厘摩 (cM) 的平均物理图距是变化的, 如拟南芥、桉树、玉米、小麦和松树分别是: 230、460、2500 和 12000Kb。这意味着在大部分一年生植物森林树种中用定位克隆 (Rommens *et al.*, 1998) 鉴定 QTL 是非常困难的。从生物学角度, 可以从候选基因 (CG) 和候选蛋白 (CP) 途径获得 QTL。

“功能 CG”是指可能影响性状表达并且测序的基因, 可以基于已知影响目标性状的生化和发育途径标准来推断。另外, 任何一个产生很大影响的突变位点可以被认为是一个 CG, 而这个 CG 的等位位点的效应较小, 可能影响自然发生的数量变异。例如, Touzet 等 (1995b) 推测一个矮化基因涉及赤霉素生物合成途径, 作为高秆生长的 QTL。QTL 和这个突变质量性状的共定位已有报道 (Beavis *et al.*, 1991; Doebley and Stec, 1993)。这种“功能 CG”的主要局限性在于目前只有一小部分基因影响的性状是已知的。

“位置 CG”可以推测所有与目标性状 QTL 定位在同一位置的基因。但是, 该策略是以比较完善的基因图谱为基础的, 只可以应用于少数几个物种。QTL 和已知基因的共定位以前曾有报道 (Byrne *et al.*, 1996; Causse *et al.*, 1995; Goldman *et al.*, 1994; Prioul *et al.*, 1997, 1999)。“位置 CG”的主要局限性在于 CG 和 QTL 的共定位纯属偶然。

蛋白数量位点作为正式候选基因的可能的方法 CG 的证实在这种信息应用于育种程序以前是非常必要的。其证实依赖于功能基因组学工具 (Borchez and Hofte, 1998), 其中 CP 途径是必不可少的步骤。确实, 如果一个 CG 的产物不显示数量和 (或) 活性

的多态性,也就是说如果一个 CG 和一个 QTL 的共定位不与另外一个控制其活性和(或)数量的 QTL 共定位,那么这个 CG 是不能被证实的。

Damerval 等(1994)率先用 QTL 策略调查 2-DE 分离蛋白的数量变异的遗传确定性。他们用玉米 F2 中分离的 RFLPs 和 PS 位点构建的连锁图谱通过对调控因子或解释点强度变异的 PQL 的区间作图(Lander and Botstein, 1989)进行定位。在分析的 72 个蛋白中,42 个蛋白检测到 70 个 PQL,其中 20 含有不只有一个 PQL,分布在整个基因组中。近来,PQLs 控制海岸松针叶蛋白的积累也在 F2 代(Costa *et al.*, 1999)检测到,并得出了相同的结论。这种途径在玉米中(de Vienne *et al.*, 1999)近来也给出一个范例,一个定位在 10 号染色体上的干旱反应性 CG(ASR1 基因,ABA/水分胁迫/催熟相关蛋白)与木质部树液、ABA 含量、叶子衰老和花丝间隔(一种干旱反应性性状)的 QTL、一个在干旱条件下控制 ASR1 蛋白有/无变异的主效 PQL 共定位。

PQL 策略也能鉴定出 CP,也就是控制数量和(或)活性的遗传因子蛋白与已知农艺性状的 QTL 共定位,而结构基因则不能。在此前提下,de Vienne 等(1999)在玉米中证实了 3 个控制单叶蛋白数量的 PQL 与 3 个高秆 QTL 共定位。

5. 方法学进展和植物蛋白数据库

近来对 2-DE 和蛋白分析的改进使植物蛋白质组分析有了相当大的进展。技术的改进,如 IPG 预制条,使得 2-DE 的可重复性有所改善,跑大量的样品变得容易,这些代表了使 PQL 途径更加可行的主要进步。现在通过微量测序、电喷雾电离 MS/MS 或 MALDI-TOF 质谱随后在蛋白或基因数据库中进行序列同源性搜寻(Li *et al.*, 1997),鉴定和分析多肽点已相对较为容易。而且,近几年,在软件和图像分析方面的进步允许分析和比较大量样品中的大量的点(Appel *et al.*, 1997)。这些进步表明 2-DE 是一种功能强大的蛋白方式的技术,能够联系生理学和遗传学方面的问题。

表 15.1 植物基因组数据库

品种	鉴定的蛋白数目 (已鉴定出的百分数)	参考文献	WWW 地址
水稻	50 (58%)	Komatsu <i>et al.</i> (1993)	http://www.rs.noda.sut.ac.jp/~kamom/2-DE/
	59 (56%)	Tsugita <i>et al.</i> (1994)	2d.html*
	51 (53%)	Tsugita <i>et al.</i> (1996)	
拟南芥	57 (63%)	Kamo <i>et al.</i> (1995)	http://www.rs.noda.sut.ac.jp/~kamom/2-DE/
	62 (56.5%)	Tsugita <i>et al.</i> (1996)	2d.html*
	150 (50%)	Santoni <i>et al.</i> (1998)	http://sphinx.rug.ac.be:8080/ppmdb/index.html
玉米	74 (48.5%)	Touzet <i>et al.</i> (1996)	http://moulon.inra.fr/imgd
	80 (67.5%)	Zivy (私人通讯)	
松树	65 (90%)	Costa <i>et al.</i> (1999)	http://www.pierroton.inra.fr/genetics/2D/

*该数据库已不再使用。

尽管蛋白质组学在越来越多的微生物以及人类中广泛研究(Humphery-Smith *et al.*, 1997; Wilkins *et al.*, 1997),蛋白质组在植物中还没有得到很大的关注。表 15.1

中列出了目前网上可用的植物蛋白数据库的特征。现在可用的数据库有：双子叶的拟南芥、单子叶的水稻、重要的农作物如玉米、裸子植物海岸松。这些数据库提供扫描的胶和可点击的图像，特征点以超级链接强调。单个蛋白入口通过激活的交叉文献与其他蛋白数据库如 SWISS-PROT 或 TrEMBL 链接。在基因组表达分析方面，所有这些数据库都包括在各种植物器官（根、叶、茎、芽、种子）或组织（愈伤组织、木质部、韧皮部）或基因型间多态方式的比较。有时候也提供其他的信息如亚细胞定位、生理学数据（海岸松针叶蛋白季节变异或水分缺乏效果）和遗传学数据（对应于海岸松和玉米连锁图谱中基因的定位）。

6. 总结、前景展望

关于群体遗传学研究，选择分子筛选技术分析自然群体中遗传多样性的程度和分布将依赖于许多因素（Karp *et al.*, 1995）是广泛接受的观点。因为每种类型的标记都具有优点和局限性，需要应用多种技术，其中包括 2-DE。这种技术的优点在于蛋白位点从基因组取样区别于许多基于 PCR 的技术，因此对于多样性问题，它提供不同水平的差异。目前，2-DE 没有被用于系统发生关系的建立。但是，如在 *Triticeae* 中展示的，在传统的细胞和分子途径不能区分的物种间的比较方面似乎具有相当大的价值。而且，当研究属于同一个家族的相关物种时发现了共有的蛋白。如果一个蛋白是栽培品种如油菜的育种的目标，而且如果编码电泳中同一蛋白的基因已经在拟南芥中（另一个十字花科植物）测序，这可以极大地缩短建立该基因为目标农艺基因的时间。可以乐观地展望未来这种比较基因组学的发展。在候选基因途径中使用蛋白分析的方法刚刚起步，因为植物育种家的获益将很快发展起来。

2-DE 只能揭示 1000 或 2000 多丰富蛋白的相对局限性，如本书的其他章节中描述的，可以通过改进提取方法和应用多步技术，提高每个样品 2-DE 胶的数目和尺寸（变窄和拓宽 pH 范围等）等方法来解决。像在医学和微生物学中一样，蛋白质组学将很快成为植物生物学中破译基因功能和贡献于育种程序的主要途径，如前面农作物品种和森林树种示例。

致谢

感谢我们在 Réseau I.N.R.A. Protéomes Végétaux 的同事帮助收集文献和丰富讨论。在 Pierroton (I.N.R.A., 法国) 森林树种的工作部分由欧盟项目 FAIR-CT95-0781 和 FAIR-CT98-3953 赞助。

(马晓芸译 贺福初校)

参考文献

- Abe, T., Gusti, R.S., Ono, M. and Sasahara, T. (1996) Variations in glutelin and high molecular weight endosperm proteins among subspecies of rice (*Oryza sativa* L.) detected by two-dimensional gel electrophoresis. *Genes Genet. Syst.* **71**: 63–68.
- Albernathy, R.H., Thiel, D.S., Petersen, N.S. and Helm, K. (1989) Thermotolerance is developmentally dependent in germinating wheat seed. *Plant Physiol.* **89**: 576–596.

- Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new world. *Electrophoresis* **19**: 1853–1861.
- Anderson, L. and Seilhammer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**: 533–537.
- Anderson, N.G., Tollaksen, S.L., Pascoe, F.H. and Anderson, L. (1985) Two-dimensional electrophoresis of wheat seed proteins. *Crop Sci.* **25**: 667–674.
- Appel, R.D., Palagi, P.M., Walther, D., Vargas, J.R., Sanchez, J.-C., Ravier, F., Pasquali, C. and Hochstrasser, D.F. (1997) Melanie II – a third-generation software package for analysis of two-dimensional electrophoresis images: 1. Features and user interface. *Electrophoresis* **18**: 2724–2734.
- Aspart, L., Meyer, Y., Laroche, M. and Penon, P. (1984) Developmental regulation of the synthesis of proteins encoded by stored mRNA in radish embryos. *Plant Physiol.* **76**: 664–673.
- Bahrman, N. and Damerval, C. (1989) Linkage relationships of loci controlling protein amounts in maritime pine (*Pinus pinaster* Ait). *Heredity* **63**: 267–274.
- Bahrman, N. and Petit, R. (1995) Genetic polymorphisms in maritime pine (*Pinus pinaster* Ait) assessed by two-dimensional gel electrophoresis of needle, bud and pollen proteins. *J. Mol. Evol.* **41**: 231–237.
- Bahrman, N., Zivy, M., Damerval, C. and Baradat, P. (1994) Organisation of the variability of abundant proteins in seven geographical origins of maritime pine (*Pinus pinaster* Ait). *Theor. Appl. Genet.* **88**: 407–411.
- Bahrman, N., de Vienne, D., Thiellement, H. and Hofmann, J.P. (1985) Two-dimensional gel electrophoresis of proteins for genetic studies in Douglas fir (*Pseudotsuga mensiesii*). *Biochem. Genet.* **23**: 247–255.
- Bahrman, N., Zivy, M. and Thiellement, H. (1988a) Genetic relationships in the *Sitopsis* section of *Triticum* and the origin of the B genome of the polyploid wheats. *Heredity* **61**: 473–480.
- Bahrman, N., Cardin, M.-L., Seguin, M., Zivy, M. and Thiellement, H. (1988b) Variability of 3 cytoplasmically encoded proteins in the *Triticum* genus. *Heredity* **60**: 87–90.
- Barreneche, T., Bahrman, N. and Kremer, A. (1996) Two dimensional gel electrophoresis confirms the low level of genetic differentiation between *Quercus robur* L and *Quercus petraea* (Matt) Liebl. *For. Genet.* **3**: 89–92.
- Beavis, W.D., Grant, D., Albertsen, M. and Fincher, R. (1991) Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theor. Appl. Genet.* **83**: 141–145.
- Bon, M.C. (1989) A two-dimensional electrophoresis procedure for single meristems of different forest species. *Electrophoresis* **10**: 530–532.
- Bouchez, D. and Höfte, H. (1998) Functional genomics in plants. *Plant Physiol.* **118**: 725–732.
- Burstin, J., de Vienne, D., Dubreuil, P. and Damerval, C. (1994) Molecular markers and protein quantities as genetic descriptors in maize. I. Genetic diversity among 21 inbred lines. *Theor. Appl. Genet.* **89**: 943–950.
- Byrne, P.F., McMullen, M.D., Snooks, M.E., Musket, T.A., Theuri, J.M., Widstrom, N.W., Wiseman, B.R. and Coe, E.H. (1996) Quantitative trait loci and metabolic pathways: genetic control of a concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl Acad. Sci.* **93**: 8820–8825.
- Cabane, M., Calvet, P., Vincens, P. and Boudet, A. M. (1993) Characterization of chilling-acclimation-related proteins in soybean and identification of one as a member of the heat shock protein (HSP 70) family. *Planta* **190**: 346–353.
- Castellino, A. (1997) When the chips are down. *Genome Res.* **7**: 943–946.
- Cause, M., Rocher, J.P., Henry, A.M., Charcosset, A., Prioul, J.L. and de Vienne, D. (1995) Genetic dissection of the relationship between carbon metabolism and early growth in maize, with emphasis on key-enzyme loci. *Mol. Breed.* **3**: 259–272.
- Chen, R. D. and Tabaeizadeh, Z. (1992). Expression and molecular cloning of drought-induced genes in the wild tomato *Lycopersicon chilense*. *Biochem. Cell Biol.* **70**, 199–206.
- Claes B., Dekeyser R., Villaroel R, van den Bulke M., Bauw G., van Montagu M. and Caplan A. (1990) Characterization of a rice gene showing organ-specific expression in response to salt stress and drought. *Plant Cell* **2**: 19–27.

- Colas des Francs, C. and Thiellement, H. (1985). Chromosomal localization of structural genes and regulators in wheat by 2D electrophoresis of ditelosomic lines. *Theor. Appl. Genet.* **71**: 31–38.
- Colas des Francs, C., de Vienne, D. and Thiellement, H. (1985) Analysis of leaf proteins by two-dimensional gel electrophoresis: protease action as exemplified by ribulose biphosphate carboxylase/oxygenase degradation and procedure to avoid proteolysis during extraction. *Plant Physiol.* **78**: 178–182.
- Costa, P., Bahrman, N., Frigerio, J. M., Kremer, A. and Plomion, C. (1998) Water-deficit-responsive proteins in maritime pine. *Plant Mol. Biol.* **38**: 587–596.
- Costa, P., Pot, D., Dubos, C., Frigerio, J.-M., Pionneau, C., Bodénès, C., Bertocchi, E., Cervera, M.-T., Remington, D.L. and Plomion, C. (2000) A genetic map of Maritime pine based on AFLP, RAPD and protein markers. *Theor. Appl. Genet.* **100**: 39–48.
- Damerval, C. and de Vienne, D. (1993) Quantification of dominance for proteins pleiotropically affected by *Opaque-2* in maize. *Heredity* **70**: 38–51.
- Damerval, C. and Le Guilloux, M. (1998) Characterization of novel proteins affected by the *o2* mutation and expressed during maize endosperm development. *Mol. Genes Genet.* **257**: 354–361.
- Damerval, C., de Vienne, D., Zivy M. and Thiellement, H. (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis* **7**: 52–54.
- Damerval, C., Maurice, A., Josse, J.M. and de Vienne, D. (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* **137**: 289–301.
- Danyluk, J., Rassart, E. and Sarhan, F. (1991) Gene expression during cold and heat shock in wheat. *Biochem. Cell Biol.* **69**: 383–391.
- David J.L., Zivy M., Cardin M.L. and Brabant, P. (1997) Protein evolution in dynamically managed populations of wheat: adaptive responses to macro-environmental conditions. *Theor. Appl. Genet.*, **95**: 932–941.
- de Vienne, D., Leonardi, A. and Damerval, C. (1988) Genetic aspects of variation of protein amounts in maize and pea. *Electrophoresis* **9**: 742–750.
- de Vienne, D., Burstin, J., Gerber, S., Leonardi, A., Le Guilloux, M., Murigneux, A., Beckert, M., Bahrman, N., Damerval, C. and Zivy, M. (1996) Two-dimensional electrophoresis of proteins as a source of monogenic and codominant markers for population genetics and mapping the expressed genome. *Heredity* **76**: 166–177.
- de Vienne, D., Leonardi, A., Damerval, C. and Zivy, M. (1999) Genetics of proteome variation as a tool for QTL characterization. Application to drought stress responses in maize. *J. Exp. Bot.* **50**: 303–309.
- Devey, M.E., Fiddler, T.A., Liu, B.-H., Knapp, S.J. and Neale, B.D. (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. *Theor. Appl. Genet.* **88**: 273–278.
- Doebley, J. and Stec, A. (1993) Inheritance of the morphological differences between maize and teosinte: comparison of the results for two F₂ populations. *Genetics* **134**: 559–570.
- Dunbar, B.D., Bundman, D.S. and Dunbar, B.S. (1985) Identification of cultivar-specific proteins of winter wheat (*Triticum aestivum* L.) by high resolution two-dimensional polyacrylamide gel electrophoresis and color-based silver stain. *Electrophoresis* **6**: 39–43.
- Dure, L. III, Greenway, S.C. and Galau, G.A. (1981) Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis. *Biochemistry* **20**: 4162–4168.
- Dure, L. III, Crouch, M., Harada, J., Ho, T.H.D., Mundy, J., Quatrano, R.S., Thomas, T. and Sung, Z.R. (1989) Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol. Biol.* **12**: 475–486.
- El Madidi, S. and Zivy, M. (1993) Variabilité génétique des protéines de choc thermique et thermotolérance chez le blé. In: *Le progrès génétique passe-t-il par le repérage et l'inventaire des gènes?* (eds. H. Chlyah and Y. Demarly). John Wiley Eurotext, London, pp. 173–184.
- Faure, J.-D., Vittorioso, P., Santoni, V., Fraissier, V., Prinsen, E., Barlier, I., Van Onckelen H., Caboche, M. and Bellini, C. (1998) The PASTICCINO genes of *Arabidopsis thaliana* are involved in the control of cell division and differentiation. *Development* **125**: 909–918.
- Gerber, S., Rodolphe, F., Bahrman, N. and Baradat, P. (1993) Seed-protein variation in maritime

- pine (*Pinus pinaster* Ait) revealed by two-dimensional electrophoresis: genetic determinism and construction of a linkage map. *Theor. Appl. Genet.* **85**: 521–528.
- Goday, A., Jensen, A.B., Culianez-Macia, F.A., Mar Alba, M., Figueras, M., Serratos, J., Torrent, M. and Pagès, M. (1994) The maize abscisic acid-responsive protein Rab17 is located in the nucleus and interacts with nuclear localization signals. *Plant Cell* **6**: 351–360.
- Goldman I.L., Rocheford, T.R. and Dudley, J.W. (1994) Quantitative trait loci influencing protein and starch concentration in the Illinois long term selection maize strains. *Theor. Appl. Genet.* **87**: 217–224.
- Görg, A., Postel, W., Domscheit, A. and Gunther, S. (1988) Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (*Hordeum vulgare*): method, reproducibility and genetic aspects. *Electrophoresis* **9**: 681–692.
- Görg, A., Postel, W., Baumer, M. and Weiss, W. (1992) Two-dimensional polyacrylamide gel electrophoresis, with immobilized pH gradients in the first dimension, of barley seed proteins: discrimination of cultivars with different malting grades. *Electrophoresis* **13**: 192–203.
- Gottlieb, L.D. and de Vienne, D. (1988) Assessment of pleiotropic effects of a gene substitution in Pea by two-dimensional polyacrylamide gel electrophoresis. *Genetics* **119**: 705–710.
- Guy, C.L. and Haskell, D. (1988) Detection of polypeptides associated with the cold acclimation process in spinach. *Electrophoresis* **9**: 787–796.
- Haynes, P.A., Gygi, S.P., Figeys, D. and Aebersold, R. (1998) Proteome analysis: biological assay or data archive? *Electrophoresis* **19**: 1862–1871.
- Helentjaris, T., Weber, D. and Wright, S. (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. *Genetics* **118**: 353–363.
- Herbik, A., Giritch, A., Horstmann, C., Becker, R., Balzer, H.-J., Bäumlein, H. and Stephan, U.W. (1996) Iron and copper nutrition-dependent changes in protein expression in a tomato wild type and the nicotianamine-free mutant *chloronerva*. *Plant Physiol.* **111**: 533–540.
- Humphery-Smith, I., Cordwell, S.J. and Blackstock, W.P. (1997) Proteome research: complementarity and limitations with respect to the RNA and RNA worlds. *Electrophoresis* **18**: 1217–1242.
- Hurkman, W.J. and Tanaka, C.K. (1988). Polypeptide changes induced by salt stress, water deficit, and osmotic stress in barley roots: a comparison using two-dimensional gel electrophoresis. *Electrophoresis* **9**: 781–787.
- Ingram, J. and Bartels, D. (1996) The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**: 377–403.
- Jensen, A.B., Goday, A., Figueras, M., Jessop, A.C. and Pagès, M. (1998) Phosphorylation mediates the nuclear targeting of the maize Rab17 protein. *Plant J.* **13**: 691–697.
- Jorgensen, J.A. and Nguyen, H.T. (1995) Genetic analysis of heat shock proteins in maize. *Theor. Appl. Genet.* **91**: 38–46.
- Kamo, M., Kawakami, T., Miyatake, N. and Tsugita, A. (1995) Separation and characterization of *Arabidopsis thaliana* proteins by two-dimensional gel electrophoresis. *Electrophoresis* **16**: 423–430.
- Karp, A., Kresovich, S., Bhat, K.V., Ayad, W.G. and Hodgkin, T. (1995) Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical bulletin no. 82, Rome, Italy, 47pp.
- Kasten, B., Buck, F., Nuske, J. and Reski, R. (1997) Cytokinin affects nuclear- and plastome-encoded energy-converting plastid enzymes. *Planta* **201**: 261–272.
- Kearsey, M.J. and Farquhar, A.G.L. (1998) QTL analysis in plants; where are we now? *Heredity* **80**: 137–142.
- Kerby, K. and Kuspira, J. (1987) The phylogeny of the polyploid wheats *Triticum aestivum* (bread wheat) and *Triticum turgidum* (macaroni wheat). *Genome* **29**: 722–737.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Klose, J. (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**: 231–243.
- Klose, J. (1982) Genetic variability of soluble proteins studied by two-dimensional electrophoresis on different inbred mouse strains and on different mouse organs. *J. Mol. Evol.* **18**: 315–328.

- Komatsu, S., Kajiwara, H. and Hirano, H. (1993) A rice protein library: a data-file of rice proteins separated by two-dimensional electrophoresis. *Theor. Appl. Genet.* **86**: 935–942.
- Krishnan, M., Nguyen, H.T. and Burke, J.J. (1989) Heat shock protein synthesis and thermal tolerance in wheat. *Plant Physiol.* **90**: 140–145.
- Lander, E.S. and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185–199.
- Lemieux, B., Aharoni, A. and Schena, M. (1998) Overview of DNA chip technology. *Mol. Breed.* **4**: 277–289.
- Leonardi, A., Damerval, D. and de Vienne, D. (1988) Organ-specific variability and inheritance of maize proteins revealed by two-dimensional electrophoresis. *Genet. Res. Camb.* **52**: 97–103.
- Leone, A., Costa, A., Tucci, M. and Grillo, S. (1994) Comparative analysis of short- and long-term changes in gene expression caused by low water potential in potato (*Solanum tuberosum*) cell-suspension cultures. *Plant Physiol.* **106**: 703–712.
- Leymarie, J., Damerval, C., Marcotte, L., Combes, V. and Vartanian, N. (1996) Two-dimensional protein patterns of *Arabidopsis* wild-type and auxin insensitive mutants, *axr1*, *axr2*, reveal interactions between drought and hormonal responses. *Plant Cell Physiol.* **37**: 966–975.
- Li, G., Waltham, M., Anderson, N.L., Unsworth, E., Treston, A. and Weinstein, J.N. (1997) Rapid mass spectrometric identification of proteins from two-dimensional polyacrylamide gels after in gel proteolytic digestion. *Electrophoresis* **18**: 382–390.
- Lund, A.A., Blum, P.H., Bhatramakki, D. and Elthon, T.E. (1998) Heat-stress response of maize mitochondria. *Plant Physiol.* **116**: 1097–1110.
- Mangin, B., Goffinet, B. and Rebai, A. (1994) Constructing confidence intervals for QTL location. *Genetics* **138**: 1301–1308.
- Marshall, A. and Hodgson, J. (1998) DNA chips: an array of possibilities. *Nature Biotechnol.* **16**: 27–31.
- Meza-Basso, L., Alberdi, M., Raynal, M., Ferrero-Cadinanos, M.L. and Delseny, M. (1986) Changes in protein synthesis in rapeseed (*Brassica napus*) seedlings during a low temperature treatment. *Plant Physiol.* **82**: 733–738.
- Moons, A., Bauw, G., Prinsen, E., Van Montagu, M. and Van der Straeten, D. (1995) Molecular and physiological responses to abscisic acid and salts in roots of salt-sensitive and salt-tolerant *Indica* rice varieties. *Plant Physiol.* **107**: 177–186.
- Moons, A., Bauw, G., Prinsen, E., Van Montagu, M. and Van der Straeten, D. (1995) Molecular and physiological responses to abscisic acid and salts in roots of salt-sensitive and salt-tolerant *Indica* rice varieties. *Plant Physiol.* **107**: 177–186.
- Moons, A., Gielen, J., Vandekerckhove, J., Van der Straeten, D., Gheysen, G. and Van Montagu, M. (1997) An abscisic-acid- and salt-stress-responsive rice cDNA from a novel plant gene family. *Planta* **202**: 443–454.
- Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press.
- Nover, L. and Scharf, K.D. (1984) Synthesis, modification and structural binding of heat-shock proteins in tomato cell cultures. *Eur. J. Biochem.* **139**: 303–313.
- O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007–4021.
- Petit, R.J., Bahrman, N. and Baradat, P. (1995) Comparison of genetic differentiation in maritime pine (*Pinus pinaster* Ait) estimated using isozyme, total protein and terpenic loci. *Heredity* **75**: 382–389.
- Picard, P., Bourgoïn-Grenéche, M. and Zivy, M. (1997) Potential of two-dimensional electrophoresis in routine identification of closely related durum wheat lines. *Electrophoresis* **18**: 174–181.
- Plomion, C., Bahrman, N., Durel, C.-E. and O'Malley, D.M. (1995) Genomic mapping in *Pinus pinaster* (maritime pine) using RAPD and protein markers. *Heredity* **74**: 661–668.
- Plomion, C., Costa, P. and Bahrman, N. (1997) Genetic analysis of needle protein in Maritime pine. 1. Mapping dominant and codominant protein markers assayed on diploid tissue, in a haploid-based genetic map. *Silvae Genet.* **46**: 161–165.
- Posch, A., van den Berg, B.M., Postel, W. and Görg, A. (1992) Genetic variability of pepper (*Capsicum annum* L.) seed proteins studied by 2-D electrophoresis with immobilized pH gradients. *Electrophoresis* **13**: 774–777.
- Posch, A., van den Berg, B.M., Duranton, C. and Görg, A. (1994) Polymorphism of pepper

- (*Capsicum annuum* L.) seed proteins studied by two-dimensional electrophoresis with immobilized pH gradients: methodical and genetic aspects. *Electrophoresis* **15**: 297–304.
- Prioul, J.L., Quarrie, S., Causse, M. and de Vienne, D.** (1997) Dissecting complex physiological functions into elementary components through the use of molecular quantitative genetics. *J. Exp. Bot.* **48**: 1151–1163.
- Prioul, J.L., Pelleschi, S., Sene, M., Thévenot, C., Causse, M., de Vienne, D. and Leonardi, A.** (1999) From QTLs for enzyme activity to candidate genes in maize. *J. Exp. Bot.* **50**: 1281–1288.
- Ramagopal, S.** (1987) Salinity stress induces tissue-specific proteins in barley seedlings. *Plant Physiol.* **84**: 324–331.
- Ramagopal, S.** (1990) Protein polymorphism in sugarcane revealed by two-dimensional gel analysis. *Theor. Appl. Genet.* **79**: 297–304.
- Ramani, S. and Apte, S. K.** (1997) Transient expression of multiple genes in salinity-stressed young seedlings of rice (*Oryza sativa* L.) cv. bura rata. *Biochem. Biophys. Res. Commun.* **233**: 663–667.
- Reviron, M.P., Vartanian, N., Sallantin, M., Huet, J.C., Pernellet, J.C. and de Vienne, D.** (1992) Characterization of a novel protein induced by progressive or rapid drought and salinity in *Brassica napus* leaves. *Plant Physiol.* **100**: 1486–1493.
- Rey, P., Pruvot, G., Becuwe, N., Eymery, F., Rumeau, D. and Peltier, G.** (1998) A novel thioredoxin-like protein located in the chloroplast is induced by water deficit in *Solanum tuberosum* L. plants. *Plant J.* **13**: 97–101.
- Riccardi, F., Gazeau, P., de Vienne, D. and Zivy, M.** (1998) Protein changes in response to progressive water deficit in maize. Quantitative variation and polypeptide identification. *Plant Physiol.* **117**: 1253–1263.
- Rommens, J.M., Iannuzzi, M.C., Kerem, B.S., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L.C. and Collins, F.S.** (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**: 1059–1065.
- Sabehat, A., Weiss, D. and Lurie, S.** (1996) The correlation between heat-shock protein accumulation and persistence and chilling tolerance in tomato fruit. *Plant Physiol.* **110**: 531–537.
- Sanchez-Martinez, D., Puigdomènech and Pagès, M.** (1986) Regulation of gene expression in developing *Zea mays* embryos. Protein synthesis during embryogenesis and early germination of maize. *Plant Physiol.* **82**: 543–549.
- Santoni, V., Bellini, C. and Caboche, M.** (1994) Use of two-dimensional protein-pattern analysis for the characterization of *Arabidopsis thaliana* mutants. *Planta* **192**: 557–566.
- Santoni, V., Delarue, M., Caboche, M. and Bellini, C.** (1997) A comparison of two-dimensional electrophoresis data with phenotypical traits in *Arabidopsis* leads to the identification of a mutant (cri1) that accumulates cytokinins. *Planta* **202**: 62–69.
- Santoni, V., Rouquié, D., Doumas, P., Mansion, M., Boutry, M., Déhais, P., Sahnoun, I. and Rossignol, M.** (1998) Use of proteome strategy for tagging proteins present at the plasma membrane. *Plant J.* **16**: 633–641.
- Saruyama, H. and Shinbashi, N.** (1992) Identification of specific proteins from seed embryos by two-dimensional gel electrophoresis for the discrimination between *indica* and *japonica* rice. *Theor. Appl. Genet.* **84**, 947–951.
- Song, K.M., Suzuki, J.Y., Slocum, M.K., Williams, P.H. and Osborn, T.C.** (1991) A linkage map of *Brassica rapa* (syn. *campestris*) based on restriction fragment length polymorphism loci. *Theor. Appl. Genet.* **82**: 296–304.
- Tacchini, P., Fink, A., Thiellement, H. and Greppin, H.** (1995) Analysis of leaf proteins in late flowering mutants of *Arabidopsis thaliana*. *Physiol. Plant.* **93**: 312–316.
- Tanksley, S.D.** (1993) Mapping polygenes. *Annu. Rev. Genet.* **27**: 205–233.
- Thiellement H., Seguin M., Bahrman N. and Zivy M.** (1989). Homoeology and phylogeny of the A, S and D genomes of the *Triticinae*. *J. Mol. Evol.* **29**: 89–94.
- Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossignol, M., Santoni, V., de Vienne, D. and Zivy, M.** (1999) Proteomics for genetical and physiological studies in plants. *Electrophoresis* **20**: 2013–2026.
- Touzet, P., Morin, C., Damerval, C., Le Guilloux, M., Zivy, M. and de Vienne, D.** (1995a) Characterizing allelic proteins for genome mapping in maize. *Electrophoresis* **16**: 1289–1294.

- Touzet, P., Winkler, R.G. and Helentjaris, T. (1995b) Combined genetic and physiological analysis of a locus contributing to quantitative variation. *Theor. Appl. Genet.* **91**: 200–205.
- Touzet, P., Riccardi, F., Morin, C., Damerval, C., Huet, J-C., Pernollet, J-C., Zivy, M. and de Vienne D. (1996) The maize two-dimensional gel protein database: toward an integrated genome analysis program. *Theor. Appl. Genet.* **93**: 997–1005.
- Tsugita, A., Kawakami, T., Uchiyama, Y., Kamo, M., Miyatake, N. and Nozu, Y. (1994) Separation and characterization of rice proteins. *Electrophoresis* **15**: 708–720.
- Tsugita, A., Kamo, M., Kawakami, T. and Ohki, Y. (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. *Electrophoresis* **17**: 855–865.
- van Berkel, J., Salamini, F. and Gebhardt, C. (1994) Transcripts accumulating during cold storage of potato (*Solanum tuberosum* L.) tubers are sequence related to stress-responsive genes. *Plant Physiol.* **104**: 445–452.
- Vierling, R.A. and Nguyen, H.T. (1990) Heat-shock protein synthesis and accumulation in diploid wheat. *Crop Sci.* **30**: 1337–1342.
- Vierling, R.A. and Nguyen, H.T. (1992) Heat-shock protein gene expression in diploid wheat genotypes differing in thermal tolerance. *Crop Sci.* **32**: 370–377.
- Wilkins, M.R., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1997) *Proteome Research: New Frontiers in Functional Genomics*. Springer, Berlin, p. 243.
- Zivy, M. (1987) Genetic variability for heat shock proteins in common wheat. *Theor. Appl. Genet.* **74**: 209–213.
- Zivy, M., Thiellement, H., de Vienne, D. and Hofmann, J.-P. (1983) Study on nuclear and cytoplasmic genome expression in wheat by two-dimensional gel electrophoresis. I. First results on 18 alloplasmic lines. *Theor. Appl. Genet.* **66**: 1–7.
- Zivy, M., Thiellement, H., de Vienne, D. and Hofmann, J.-P. (1984) Study on nuclear and cytoplasmic genome expression in wheat by two-dimensional gel electrophoresis. II. Genetic differences between two lines and two groups of cytoplasm at five developmental stages or organs. *Theor. Appl. Genet.* **68**: 335–345.
- Zivy, M., Devaux, P., Blaisonneau, J., Jean, R. and Thiellement, H. (1992) Segregation distortion and linkage studies in microspore derived doubled haploid lines of *Hordeum vulgare* L. *Theor. Appl. Genet.* **83**: 919–924.
- Zivy, M., El Madidi, S. and Thiellement, H. (1995) Distance indices in a comparison between the A, D, and R genomes of the *Triticeae* tribe. *Electrophoresis* **16**: 1295–1300.



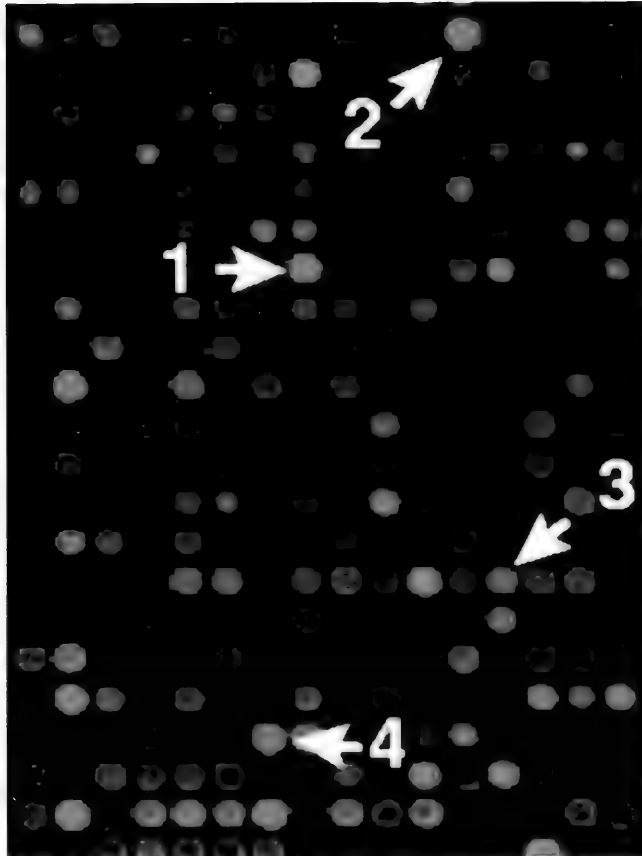


图 2.2 用分离标签区分的两个 mRNA 家族共同杂交后的典型玻璃微阵列，微阵列的一部分包含与来源于无血清 (Cy3) 和血清刺激 (Cy5) 成纤维细胞的 RNA 衍生物杂交的人 cDNA。

(1) 黄点代表两个样品间表达量变化不大的基因。(2 和 4) 在血清刺激的成纤维细胞中表达丰度高的 mRNAs 显红色。(3) 在无血清成纤维细胞中表达丰度高的 mRNAs 显绿色。(Reprinted with permission from Iyer *et al.*, The transcriptional program in the response of human fibroblasts to serum, *Science*, volume 283, copyright 1999 American Association for the Advancement of Sciences).



收到期	2003.1.3
来源	三联书店
书价	45.00元
单据号	1674472
开票日期	2003.1.3

中科院植物所图书馆



S0000099

27481		58.17421	
		125	
蛋白质组学：从序列到功能			
2002			
借借者单位	借者姓名	借出日期	还书日期
2124世第	03.5.13	徐松	04.8.04
	04.7.10	徐松	04.9.06
		徐松	04.2.07
			04.11.30

58.17421
125

注 意

- 1 借书到期请即送还。
- 2 请勿在书上批改圈点，折角。
- 3 借去图书如有污损遗失等情形须照章赔偿。

27481

京卡 0701

责任编辑: 马学海 王 静 盖 宇
封面设计: 孟华丽

生命科学编辑部

联系电话: 010-64012501

<http://www.lifescience.com.cn>

e-mail: info@lifescience.com.cn



全国高技术重点图书·生物技术领域

随着人类基因组序列框架图的宣布和许多其他基因组测序工作的完成, 现在人们的注意力正转向对基因组编码的蛋白质的分析。这是一个迅速发展的研究领域, 《蛋白质组学: 从序列到功能》集中阐述了蛋白质组学的最新进展, 并将之与当前基因组学技术相比较。本书参编人员均为学术界和企业界的专家, 以准确而通俗的风格综述该领域, 全面涵盖并整合了基因组学和蛋白质组学的内容。

本书适合于生物化学、分子生物学和遗传学专业及所有对基因组学和蛋白质组学感兴趣的大学师生、科研人员和研究生参考。

相关图书:



基因组 (*Genomes*)

T. A. Brown 著, 袁建刚、周严、强伯勤 主译

科学出版社2002年8月出版, 定价: 48元

本书以清晰而简明的写作风格将基因组学的新观点与研究基因表达的传统方法相结合, 为基因组作为生命蓝图所起的重要作用提供了最新理解。全书从三个方面论述了基因组学的核心主题: 基因组的研究方法、基因组的功能、基因组的复制和进化。全书采用大量的图表, 形象而简洁, 适合各层次的学生使用, 是一本必备的现代分子生物学教材。

ISBN 7-03-010747-0



9 787030 107473 >

ISBN 7-03-010747-0/Q · 1201

定价: 45.00元