

"十一五"国家重点图书

蛋 | 白 | 质 | 科 | 学 | 与 | 技 | 术 | 丛 | 书

「蛋白质组学 原理」

Principles of Proteomics

[英] R. M. 特怀曼 (R. M. Twyman) 著

王恒樑 袁 静 刘先凯 等译

苏国富 主审

从基因组学到蛋白质组学

蛋白质分离策略

蛋白质鉴定的策略

蛋白质定量方法

蛋白质组学与蛋白质序列分析

结构蛋白质组学

相互作用蛋白质组学

蛋白质组学中的蛋白质修饰

蛋白质芯片和功能蛋白质组学

蛋白质组学的应用



化学工业出版社
生物·医药出版分社

58.17421
487

" 十一五 " 国家重点图书

蛋 | 白 | 质 | 科 | 学 | 与 | 技 | 术 | 丛 | 书

「蛋白质组学 原理」

Principles of Proteomics

[英] R. M. 特怀曼 (R. M. Twyman) 著

王恒樑 袁 静 刘先凯 等译

苏国富 主审



化学工业出版社
生物·医药出版分社

· 北 京 ·

序 言

生命科学在 21 世纪已从基因组的研究进入功能基因组或结构基因组的研究,也就是说要在蛋白质结构的基础上研究基因编码的各种蛋白质的功能,并从传统的单个蛋白质的研究走向细胞内蛋白质群体的研究,从而更加深入地揭示生命活动的奥秘。这方面的工作构成了当今生命科学研究的前沿,即蛋白质科学。

人类对蛋白质重要性的认识源远流长。1878 年恩格斯在《反杜林论》中就指出“生命是蛋白体的存在形式”。不过当时对蛋白质的本质和结构还知之甚少。不久,瑞典的 Theodor Svedberg 利用他首创的超离心技术才知道蛋白质分子是均一的并具有固定的分子量。此后,剑桥大学的 Frederick Sanger 利用纸电泳和纸层析测出了第一个蛋白质——胰岛素的一级结构;Max Perutz 和 John Kendrew 利用他们首创的重原子同晶置换技术测出了第一个蛋白质——肌红蛋白晶体的三维结构。我国对蛋白质的研究起步也不晚。20 世纪二三十年代,北京协和医学院吴宪领导的实验室对蛋白质的变性作用进行了深入系统的研究,提出蛋白质变性的实质是蛋白质分子从紧密而有序的结构变为散漫而无序的结构。吴宪的蛋白质变性学说对当前蛋白质分子折叠的研究仍然具有现实意义。及至 20 世纪五六十年代,中国科学院生物化学研究所、有机化学研究所和北京大学又通过社会主义大协作合成了胰岛素,而且得到了晶体。这不仅是世界上第一次在实验室合成了具有天然构象的蛋白质,而且彻底证明了“一级结构决定高级结构”这一重要原理。随后,我国又独立测出了胰岛素晶体的三维结构。正是由于国内外这些开创性的工作,才有了当前蛋白质科学的蓬勃发展。

回顾蛋白质科学发展的历程,我们可以看到先进技术对科学发展的无比重要性。除了上述的超离心、层析、电泳、重原子同晶置换等技术以外,后来涌现的高效液相色谱、二维凝胶电泳、毛细管电泳、质谱、生物芯片等技术又引领了蛋白质组学的研究,使我们有可能对细胞内的大量蛋白质群体进行综合研究。由此看来,科学与技术是相辅相成的,相互促进的,二者缺一不可。

在人类基因组已被基本破译的基础上,成千上万的蛋白质的结构与功能及其相互作用亟待阐明。蛋白质科学的研究成果将有助于阐明生命现象的本质和活动规律,为多种疾病的致病机理和防治提供理论依据。因此,蛋白质科学是发达国家激烈争夺的制高点,成为近年来生命科学不断取得重大突破的热点领域。我国对蛋白质科学也给予了高度重视,《国家中长期科学和技术发展规划纲要(2006~2020年)》将“蛋白质研究”列为四项重大科学研究计划之一。在广大科研人员的共同努力下,我国在蛋白质研究领域已开始出现与国际先进水平基本同步的良好发展态势,取得了一些重要的理论和应用成果,陆续在“Nature”、“Science”、“Cell”等重要期刊上发表了一批有影响的论文。

正是在这样的背景下,化学工业出版社及时出版的这套《蛋白质科学与技术丛书》很有现实意义。丛书由相互独立而又彼此联系的各分册组成,主题包括“理论”和“技术”两个层面,现阶段以“技术”为主。归纳起来,这套丛书具有以下一些特色。

1. 选题涉及的范围较为广泛。从蛋白质的基础理论,到研究的各种技术方法,以至于前沿的蛋白质组研究,在丛书中都有体现,以满足不同方向、不同层次的科研和教学需要。

2. 编写队伍具有较高水准,由中国科学院、军事医学科学院、北京大学、南京大学等单位活跃在第一线的学术骨干为主编写。

3. 编写人员分布在国内外各研究机构,因而能扬各家之长,并体现国际化的学术合作。

值得欣慰的是,目前这套丛书已通过立项成为国家“十一五”重点图书出版项目,其编写计划也得到了学术界的大力支持,进展顺利。计划两年内陆续出版的的第一批书目包括:

蛋白质化学	蛋白质物理学
蛋白质组学原理	蛋白质化学分析技术
蛋白质电泳技术	蛋白质色谱分离技术
重组蛋白质制备技术	蛋白质光谱技术
蛋白质质谱技术	蛋白质芯片技术

蛋白质科学的研究方兴未艾,今后的发展任重而道远。面对这一挑战,《蛋白质科学与技术丛书》的出版必将促进我国的蛋白质科学在已有的基础上进一步发扬光大。同时,随着这一领域的飞速发展,也希望这套丛书能不断推出后续的新篇和新人新作。

中国科学院院士

张友尚

2007年4月26日

译者的话

基因是遗传信息的携带者，而蛋白质是遗传信息的执行者。在人类基因组计划完成后，后基因组时代的任务就是对蛋白质复杂多样的结构与功能、相互作用和动态变化进行深入研究，从而在分子、细胞和整个生物体等多个层次上全面揭示生命现象的本质。采用研究蛋白质的传统方法是不可能实现的，而蛋白质组学能对蛋白质进行全面分析。另外，蛋白质组学研究成果将催生一系列新的生物技术，带动医药、农业和绿色产业的发展，极大地影响未来生物经济。因此，蛋白质组学研究是目前世界上众多国家激烈争夺的生命科学热点。我国也将“蛋白质研究”列入《国家中长期科学和技术发展规划纲要》的重大科学研究计划中。因此，迫切需要全面介绍蛋白质组学的入门读物。

《蛋白质组学原理》一书主要讲述了蛋白质组学研究的一般原理和所涉及的技术与方法，如双向凝胶电泳、多维液相色谱、质谱学、序列分析、结构分析、研究蛋白质相互作用、修饰、定位和功能的方法等。另外，该书还对蛋白质芯片和蛋白质组学在医药领域中的应用进行了阐述。因此，该书是一本浅显易懂的、理论与实践有机结合的蛋白质组学参考书。通过对该书的阅读可以使读者对蛋白质组学所涉及的范围及其潜能有一个全面的认识，对于进一步普及蛋白质组学知识和拓宽蛋白质组学的研究领域将十分有益。所以，我们在化学工业出版社生物·医药出版分社的大力支持下，组织军事医学科学院生物工程研究所从事蛋白质组学研究的年轻博士们对该书进行翻译，奉献给广大对蛋白质组学研究与应用感兴趣的国内读者，希望能进一步推动蛋白质组学在我国的迅速发展。

尽管参与翻译本书的研究人员在蛋白质组学研究方面都积累了一定的经验，但是对蛋白质组学整个学科知识的了解和掌握还相当有限，所以在翻译过程中难免出现这样那样的错误或不正确之处，希望读者提出宝贵意见。

王恒樑

军事医学科学院生物工程研究所
病原微生物生物安全国家重点实验室

2007年1月于北京

前 言

蛋白质组学，这个词汇被使用还不到十年，现在代表着一个迅速成长和成熟的学科，一个蓬勃发展的产业。蛋白质组学是对蛋白质的全面分析，看来该方法能够获得生命科学中其他大规模工作所得不到的结果：通过直接分析生命细胞的所有功能组分而不是编码它们的基因，来对生命细胞进行全面的描述。虽然蛋白质组学迅速发展起来只有很短时间，但是它可能提供的关于生命系统的信息甚至比十年前开始的“基因组学革命”还要多，这是因为蛋白质组学所包含的信息极其丰富。基因有序列，而蛋白质不仅有序列，还有结构、生化功能和生理功能，并且它们的活性还受到化学修饰、细胞定位以及蛋白质与其他分子的相互作用等因素的影响，其中后者可能是所有因素中最重要。如果把基因比作遗传指令的携带者，那么蛋白质就是执行这些指令的分子；如果基因是整个进化过程中的重要改变因素，那么蛋白质就是确定哪些改变被遗传下来、哪些改变被抛弃的分子。正是从蛋白质我们得知生命细胞和微生物是如何构造和维持生存的，又如何在生命活动的某一环节出现问题时死亡的。

像其他任何新兴的科学领域一样，蛋白质组学对于那些每天都进行大规模的蛋白质分析的人很有意义，但是那些不涉足该领域的人们对它并不是很了解。蛋白质组学有很多术语和缩写词，并且新技术的出现和变化看起来好像每天都在发生，跟上新技术的进步确实很困难，有时候甚至在蛋白质组学某一领域的专家也很难把他们的知识应用到其他的专业领域。我希望这本书对那些需要扩展自己蛋白质组学知识的人有所帮助，告诉他们蛋白质组学能够做什么；它并不是要在某些特殊的专业领域里提供专家性意见，对于那些需要知道诸如操作方法细节等问题的读者应该阅读其他很多专业领域的书籍：比如关于电泳、质谱学以及生物信息学等。然而这本书把有关蛋白质组学的分散的、不同的技术和应用组织在一起，并且按照我所期望的那样，以用户容易掌握使用的简单方式呈现给读者。在第1章的简单介绍后，对各种蛋白质组学技术进行了更详细的讨论：双向凝胶电泳，多维液相色谱，质谱学，序列分析，结构分析，研究蛋白质相互作用、修饰、定位和功能的方法

等。蛋白质芯片是一个新出现的很有前途的技术，近来也被用来进行蛋白质组学研究，它丰富了蛋白质组学研究的手段，将安排在第9章讲解。本书的最后一章介绍了几个如何应用蛋白质组学的实例，尤其是在医药学领域的应用。再次强调，本书并不打算进行详尽的阐述，而只是让读者对蛋白质组学所涉及的范围及其潜能有一个总体的认识。在每一章的末尾都有一个短的参考书目，包括一些经典的研究论文和综述，这对于有意于更深入钻研这个学科的读者会大有裨益。我认为该书适合于具有一定分子生物学和生物化学应用知识的读者。

如果没有一大批同道们的帮助和支持，本书将不可能写成。尤其是Garland/BIOS课题组成员们的耐心、坚韧和面对紧张的截稿期限的乐观精神，促成了这本书的如期完成。我衷心地感谢对本书各章节提出宝贵意见和指出潜在的错误和遗漏的许多朋友和同事们，我尤其要感谢费劳恩霍夫分子生物学和应用生态学研究所、亚琛科技大学和纽约大学生物科学系的全体工作人员。

最后，我也谨以此书奉献给我的父母，Peter 和 Irene；以及我的孩子们，Emily 和 Lucy。

Richard M. Twyman

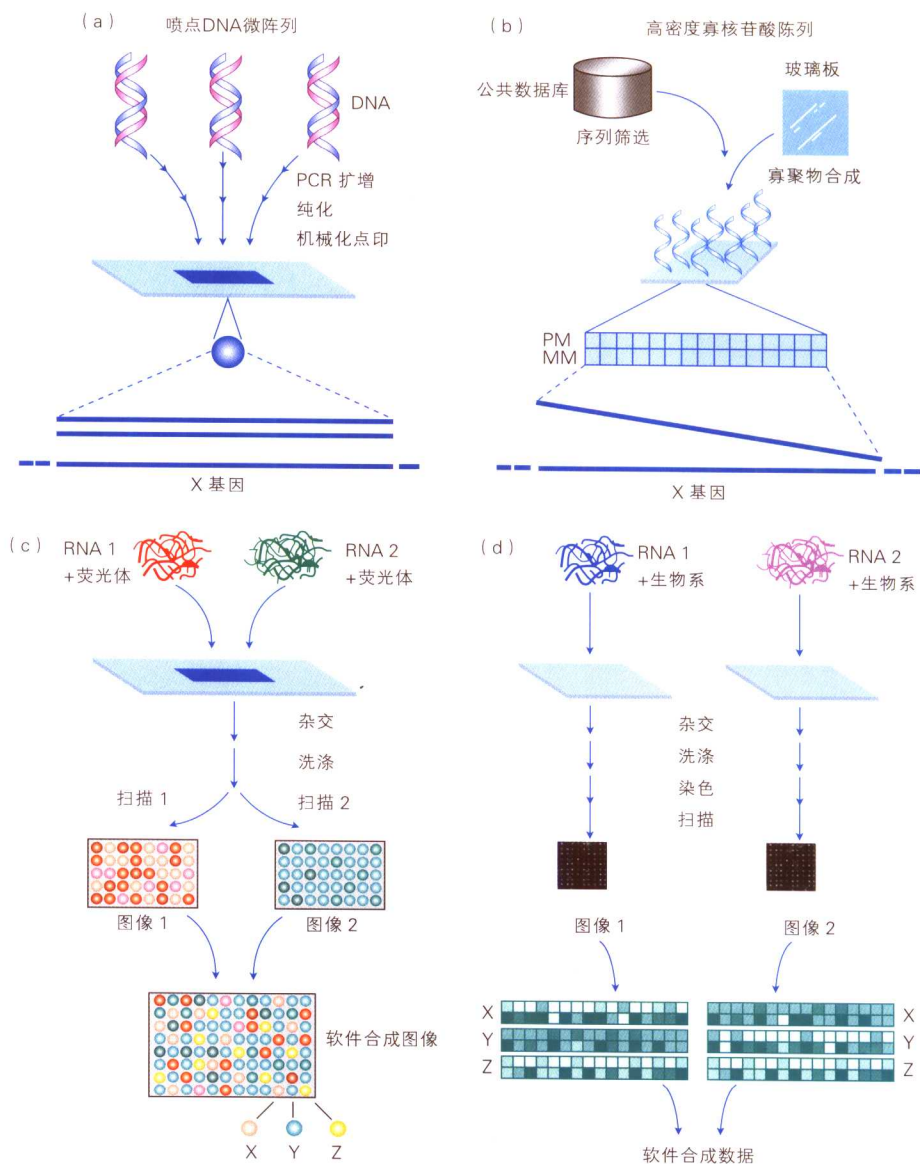
缩略语

2DGE	双向凝胶电泳
3D-PSSM	三维位置特异打分矩阵
AC	亲和色谱
ADP	二磷酸腺苷
AFM	原子力显微镜
ATP	三磷酸腺苷
BCA	二辛可宁酸
BCG	卡介苗
BLAST	局部同源检索工具
CATH	类型、结构、拓扑学和同源超家族
CBP	钙结合蛋白
CCD	电感耦合摄像装置
CD	圆二色
cDNA	互补 DNA
CDS	圆二色谱
CE	毛细管电泳
CF	色谱聚焦
CGE	毛细管凝胶电泳
CHAPS	3-[(3-胆酰胺丙基)-二甲氨基]-1-丙磺酸内盐
CHS	查耳酮合成酶
CID	碰撞诱导裂解
COSY	相关能谱
cRNA	互补 RNA
DALPC	大蛋白复合物的直接分析
DDBJ	日本 DNA 数据库
DHFR	二氢叶酸还原酶
DIGE	差异凝胶电泳

DNA	脱氧核糖核酸
dsRNA	双链 RNA
DTT	二硫苏糖醇
EDC	<i>N,N'</i> -二甲氨基丙基乙基碳二亚胺
EGF	表皮生长因子
EGFR	表皮生长因子受体
EGTA	乙二醇双(2-氨基乙醚)四乙酸
eIF	真核起始因子
ELISA	酶联免疫吸附测定
EM	电镜
ER	内质网
ESI	电喷雾离子化
EST	表达序列标签
FMN	黄素单核苷酸
fn II	纤连蛋白 II 型结构域
FRET	荧光共振能量转移
FT-ICR	傅里叶变换离子回旋共振
GM	基因修饰
GPI	糖基磷脂酰肌醇
GST	谷胱甘肽-S-转移酶
HMM	隐马尔可夫模型
HPLC	高效液相色谱
HTI	高通量成像
ICAT	同位素编码的亲亲和标签
IEC (IEX)	离子交换色谱
IEF	等电聚焦
IMAC	固相金属亲和色谱
IPG	固相 pH 梯度
LC	液相色谱
LCM	激光捕获显微解剖
MAD	多波长异常散射
MALDI	基质辅助激光解吸/离子化
MCAT	质量编码丰度标签
MIP	分子印记聚合物
MIR	多重同晶置换法
μ LC-MS	毛细管液相色谱-质谱分析
MPSS	大规模平行信号测序

mRNA	信使 RNA
MS	质谱
MS/MS	串联质谱
MudPIT	多维蛋白质鉴定技术
MultiD-LC	多维液相色谱
NAD	烟酰胺腺嘌呤二核苷酸
NEPHGE	非平衡 pH 梯度凝胶电泳
NIGMS	美国国家医学研究院
NMR	核磁共振
NOE	核奥佛豪塞效应
NOESY	NOE 谱
NTA	次氮基三乙酸
OPA	邻苯二甲醛
ORF	可读框
PAGE	聚丙烯酰胺凝胶电泳
PAL	苯丙氨酸解氨酶
PAS	过碘酸希夫染色法
PCA	蛋白互补测定
PDB	蛋白质数据库
PDMS	聚二甲基硅氧烷
PGC	多孔石墨碳
PISA	蛋白质原位阵列
PLP	磷酸吡哆醛
PMF	肽质量指纹谱
PNGase	肽 N-糖苷酶
ppi	焦磷酸
PQL	蛋白数量位点
PS	位置迁移
PSD	源后衰变
PSI-BLAST	位点特异性迭代 BLAST
PTM	翻译后修饰
PVDF	聚偏(二)氟乙烯
QTL	定量性状位点
RCA	滚环扩增
RISC	RNA 诱导沉默复合物
RMSD	均方根偏差
RNA	核糖核酸

RNAi	RNA 干扰
RRS	Ras 募集系统
RT-PCR	反转录 PCR
SAGE	基因表达系列分析
scFv	单链抗体
SCOP	蛋白质结构分类数据库
SDS	十二烷基硫酸钠
SEC	尺寸排阻色谱
SELDI	表面增强激光解吸 / 离子化
SEND	表面增强纯解吸
SGA	合成基因阵列
SH2	SRC 同源体 2
SH3	SRC 同源体 3
SILAC	用细胞培养液中的氨基酸进行稳定同位素标记
siRNA	小干扰 RNA
SIRAS	单对同晶置换加异常散射
SNP	单核苷酸多态性
SOS	鸟嘌呤核苷酸释放因子
SPR	表面等离子体共振
SRCD	同步辐射圆二色
SRS	SOS 募集系统
TAP	串联亲和纯化
TEV	烟草蚀纹病毒
TIM	磷酸丙糖异构酶
TLC	薄层色谱
TNF	肿瘤坏死因子
TOCSY	总相关能谱
TOF	飞行时间
TPA	组织纤溶酶原激活物
TQ	三级四极杆
tRNA	转运 RNA
UPA	通用蛋白质阵列
USPS	基于泛素蛋白的裂解蛋白传感器
UV	紫外线
XRC	X 射线晶体学



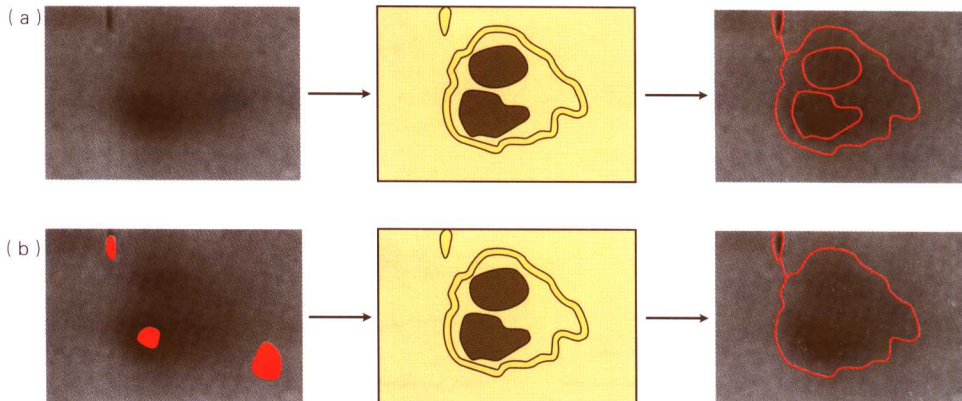
彩色插图1 用DNA微阵列进行表达分析

(a) 喷点微阵列是将扩增的cDNA分子机械地点印在玻璃片上。每个点或每份DNA对应于一个几百个碱基或更长的连续的基因片段；

(b) 高密度寡核苷酸芯片是用光控组合化学合成法 (light-directed combinatorial chemical synthesis) 制备的, 使产生的成千上万的不同序列高度有序地排列在一个小的玻璃芯片上。在阵列上用15~20个不同的寡核苷酸对代表一个基因 (PM-perfectly matched, 完全配对; MM-mismatch, 错配);

(c) 在喷点微阵列上, 比较表达分析通常用不同的荧光基团分别标记两组mRNA或cDNA样本, 将它们与玻片上的每个点杂交, 然后通过扫描来独立测定两种荧光基团。在图像中显示X颜色的点表示在样品1中转录水平增加, 显示Y颜色的点表示在样品2中转录水平增加, 显示Z颜色的点表示在样品1和样品2中转录水平相同;

(d) 在Affymetrix GeneChips芯片上, 结合生物素的cRNA与阵列杂交, 然后用结合有抗生物素抗体的荧光基团染色, 信号用激光扫描进行检测。图中显示了假想基因的几套寡核苷酸对: 在样品1中表达水平提高的假想基因的完全匹配的寡核苷酸显示为X、在样品2中表达水平提高的假想基因的完全匹配的寡核苷酸显示为Y, 在样品1和样品2中表达水平类似的假想基因的完全匹配的寡核苷酸显示为Z。经 Elsevier 许可摘自 Current Opinion in Microbiology, Vol.3, Harrington等。'Monitoring gene expression using DNA microarray', pp285~291, ©2000



彩色插图2 在2D胶图上寻找轮廓的分水线法

(a) 任何一张灰度图都可以被看作一张地形图。由不同的地图图形单位汇集而成，不同来源的水不允许汇合，那么这个图被分成积水盆地和分水线，但实际上这会导致过度分割；

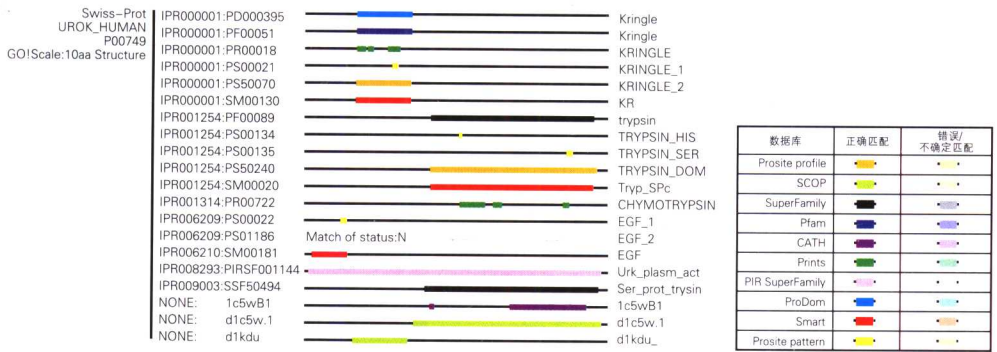
(b) 因此，如果从标记处（红色区）开始汇合，这就明显减少了过度分割。

图片改编自Serge Beucher博士, CMM/Ecole de Mines de Paris的图片集



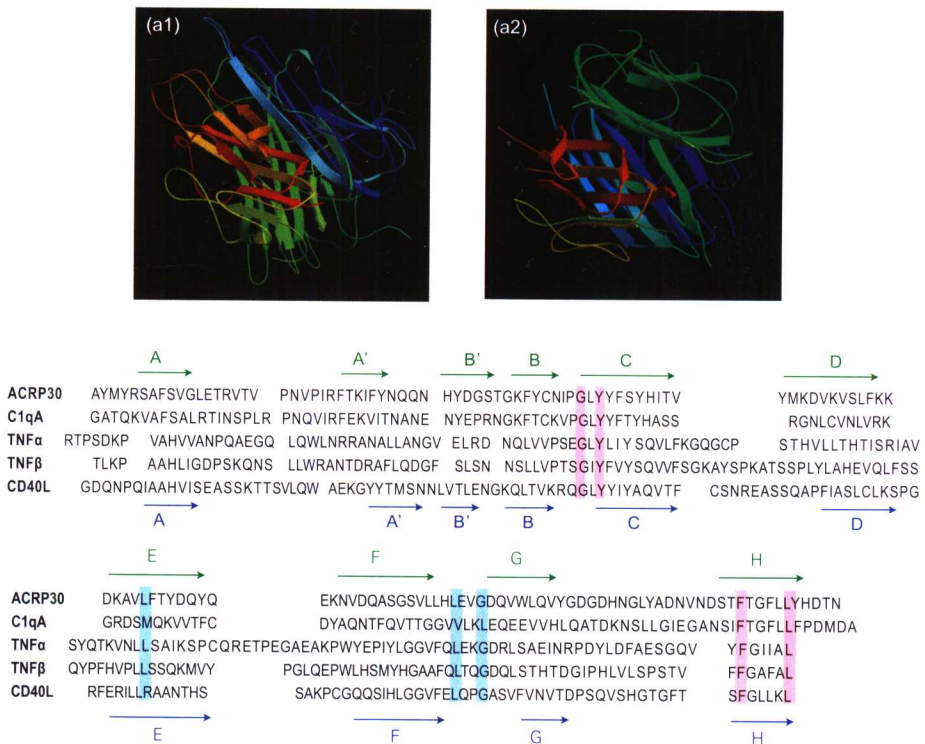
彩色插图3 Cy3（绿色）和Cy5（红色）标记的带有检测内标（spike）蛋白的胡萝卜软腐欧文菌（*Erwinia carotovora*）蛋白的2D DIGE 拟合图

实验中使用的检测内标包括3个伴清蛋白单体（箭头所示）和2个肌红蛋白单体（箭头所示），两个颜色通道间具有相同密度的蛋白点在拟合图上显示黄色。因为内标蛋白在Cy5通道的丰度高8倍，所以在拟合图上显示红色。凝胶的左端是pH酸性段。经Elsevier许可摘自Current Opinion in Biotechnology, Vol.6, Lilly等。‘Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation’, pp46-50. ©2002



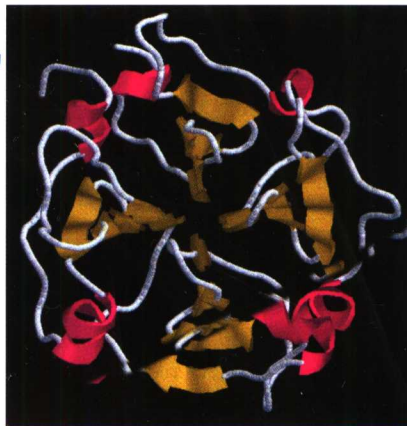
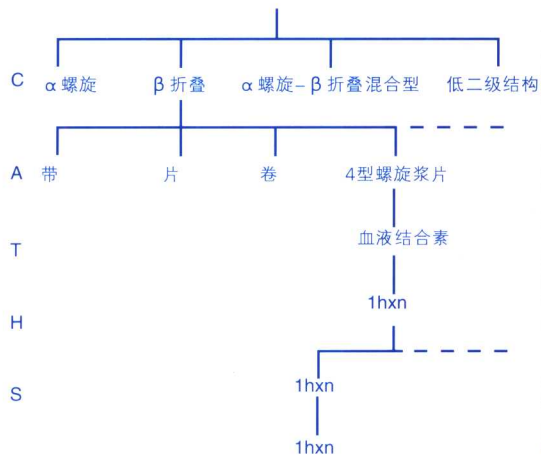
彩色插图4 人尿激酶样纤溶酶原激活物的InterPro分析

不同颜色的柱子代表在不同的数据库如Prosite、SuperFamily、Prints、iProClass (PIR SuperFamily)、ProDom、Smart 以及CATH和SCOP结构数据库中的图形和结构的匹配程度



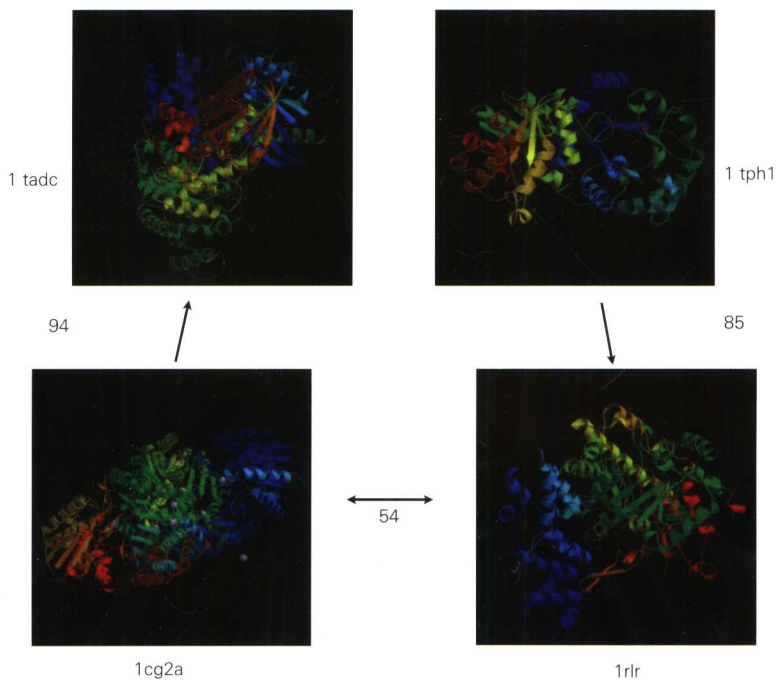
彩色插图5

(a) AdipoQ和TNFα的飘带图比较。两者的结构相似性等同于肿瘤坏死因子家族内部的相似性。(b) TNF家族的几个成员(CD40L、TNFα和TNFβ)和C1q家族的两个成员(C1qA和AdipoQ)之间基于结构的序列比对。高度保守的残基(至少在4个蛋白质中出现)用阴影遮盖,箭头表示的是蛋白质中的β折叠股区域。AdipoQ和TNF之间只有很低的序列相似性(如AdipoQ和TNFα之间的残基一致性率只有9%),因而用BLAST搜索是不能识别到关系的。经Elsevier 许可摘自Current Opinion in Biotechnology, Vol. 11, Shapiro 和 Harris, 'Finding function through structural genomics' pp31-35, ©2000。图像由Protein DataBank惠赠



彩色插图6 利用CATH数据库的蛋白质结构分类

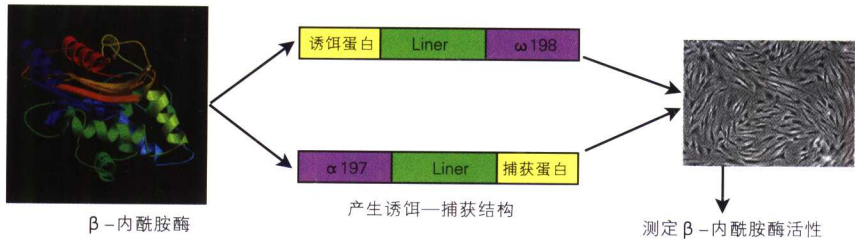
显示的蛋白质是血液结合素，这是一个富含β折叠、只有很少α螺旋的蛋白质。经过Christine Orengo的允许



彩色插图7 俄罗斯卷效应

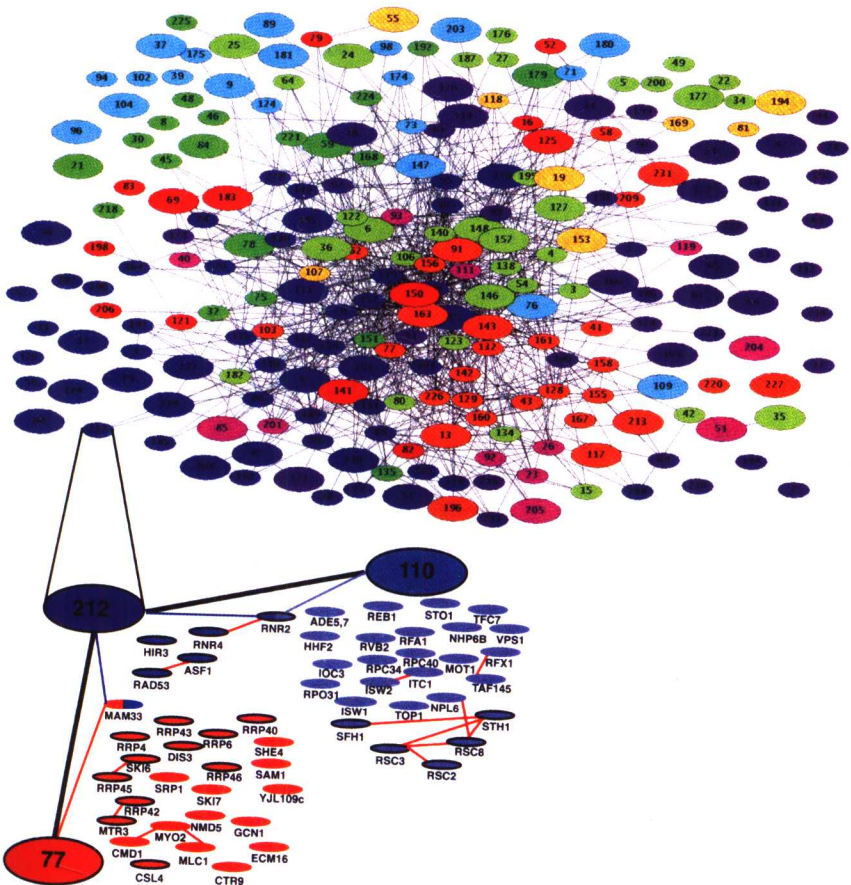
显示了4个蛋白质在折叠模式空间经历的连续的结构转变。每个蛋白质与其最近的邻居至少共同拥有74个等位残基，但是在直接结构比较时两个极端蛋白质显示了只有54个结构等位残基。

1cg2A—carboxypeptidaseG2, 羧肽酶G2; 1tadC—transducin-K, 传递蛋白K; 1tph1—triose phosphate isomerase, 磷酸丙糖异构酶; 1rlr—ribonucleotide reductase protein R1, 核糖核苷酸还原酶蛋白R1。经过欧洲生物化学学会联盟 (Federation of European Biochemical Societies) 的许可摘自The role of protein structure in genomics, Domingues等. FEBS Letters, Vol 476, pp98~102, ©2000



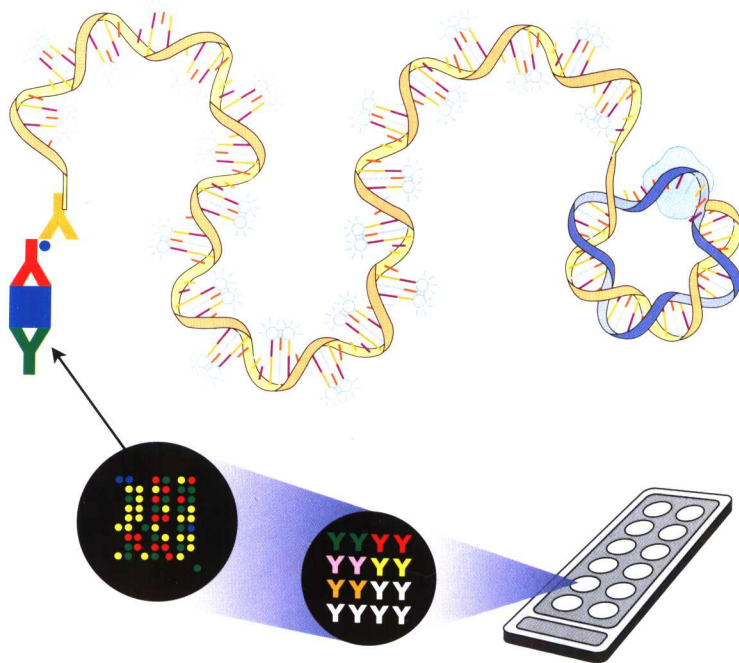
彩色插图8 通过测定β-内酰胺酶活性来绘制哺乳动物蛋白质相互作用图谱

在这种方法中，先构建一个能够融合表达诱饵蛋白与β-内酰胺酶ω198片段的结构，第二个结构能融合表达捕获蛋白与β-内酰胺酶α197片段。在稳定的细胞株中表达这两个蛋白质，如果诱饵蛋白与捕获蛋白相互作用就可以检测到β-内酰胺酶的活性



彩色插图9 蛋白复合物网络和有联系的复合物的分组

在共同包括至少一种蛋白的复合物间建立联系。为了更清楚描述蛋白质间相互作用，将至少在9种复合物中发现的蛋白质排除掉。该图通过一种松弛算法自动生成，该算法将存在联系的节点间距离最小化，而将不存在联系的节点间距离最大化而来发现节点分布中的局部最小值。图中上面部分每个复合物的细胞功能用颜色进行标注：红色，细胞周期；暗绿色，信号传导；深蓝色，转录、DNA修复、染色质结构；粉红色，蛋白质及RNA运输；橘红色，RNA代谢；淡绿色，蛋白质折叠与翻转；褐色，细胞极性化及结构维持；紫色，中间态或能量代谢；浅蓝色，膜的生物发生及膜运输。图下面部分为举例说明。复合物TAP-C212与其他两个复合物（TAP-C77和TAP-C110），通过其共有成分而被联系在一起，这说明了组织中在蛋白质和复合物水平的联系。红色线条表明酵母蛋白质组数据库中所列出的物理相互作用。© Nature Publishing Group, Nature, Vol.415, 114~117, 'Functional organization of the yeast proteome by systematic analysis of Protein Complexes' by Gavin A.-C.等



彩色插图10 使用RCA抗体芯片灵敏地检测蛋白质

芯片被分成16个聚四氟乙烯树脂孔，每个包含256个抗体作为探针的阵列。当蛋白质（用蓝色的方块表示）被其中的一个抗体捕获时，可以通过生物素标记的二级抗体（红色）来识别。接下来可以结合到环状寡聚核苷酸的第三个通用抗体上得以检测。一种链置换DNA聚合酶能使用这个环状的模板生成长的串联体。经Elsevier许可摘自 *Current Opinion in Biotechnology*, Vol.4, Kingsmore 和 Patel, 'Multiplexed protein profiling on antibody-based microarrays by rolling circle amplification', pp74~81, ©2003