

"十一五" 国家重点图书



华夏英才基金学术文库

蛋 | 白 | 质 | 科 | 学 | 与 | 技 | 术 | 丛 | 书

蛋白质 结构模拟与设计

Protein Structure Modeling and Design

李荣秀 魏冬青 冯雁 等编著

蛋白质结构研究的重大历史事件

蛋白质结构基础理论

从序列到结构

蛋白质结构模拟在蛋白研究中的应用实例

锌指核酸酶的构建

诱导多能干细胞的关键转录因子Oct4结构模拟与设计

蛋白质(酶)进化工程



化学工业出版社
生物·医药出版分社

" 十一五 " 国家重点图书



华夏英才基金学术文库

蛋 | 白 | 质 | 科 | 学 | 与 | 技 | 术 | 丛 | 书

「蛋白质 结构模拟与设计」

Protein Structure Modeling and Design

李荣秀 魏冬青 冯雁 等编著



化学工业出版社
生物·医药出版分社

· 北 京 ·

图书在版编目 (CIP) 数据

蛋白质结构模拟与设计/李荣秀等编著. —北京:
化学工业出版社, 2011. 1

(“十一五” 国家重点图书)

(华夏英才基金学术文库)

(蛋白质科学与技术丛书)

ISBN 978-7-122-10183-9

I. 蛋… II. 李… III. 蛋白质-生物结构

IV. Q510. 1

中国版本图书馆 CIP 数据核字 (2010) 第 250908 号

责任编辑: 傅四周

文字编辑: 焦欣渝

责任校对: 边 涛

装帧设计: 关 飞

出版发行: 化学工业出版社 生物·医药出版分社

(北京市东城区青年湖南街 13 号 邮政编码 100011)

印 刷: 北京永鑫印刷有限责任公司

装 订: 三河市万龙印装有限公司

710mm×1000mm 1/16 印张 10 $\frac{3}{4}$ 彩插 3 字数 186 千字 2011 年 5 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888 (传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书, 如有缺损质量问题, 本社销售中心负责调换。

定 价: 49.00 元

版权所有 违者必究

	5	10	15	20	25	30	35	40	45	50	55	60
1	H5N1-NA 62 67 72 77 82 87 92 97 102 107 112 117 122 KAVTSVTLAGNSSLCPI SGWAVHSDKNSIRIGSKGDV FVIREPFI SCSHLECR TFFLTQGALLND											
2	H9-NA(1F8B) 87 92 97 102 107 112 117 122 127 132 137 142 RDFNNLTKGL...CTINSWHIYGKDN A VRIGEDSDVLVTREPYVSCDPDECRFYAL SQGTTIRG											
	70	75	80	85	90	95	100	105	110	115	120	125
1	H5N1-NA 127 132 137 142 147 152 157 162 167 172 177 182 187 KHSNGTVKDRSPHRTL MSCPVGEAPSPYNSRFESVAWSASACHDGT SWLTIGISGPDNGAVAVLK											
2	H9-NA(1F8B) 147 152 157 162 167 172 177 182 187 192 197 202 207 KHSNGTIHDRSQYRALISWPLSSPPTVYNSRVE CIGWSSTSC HDGKTRMSICISGPNNNASAVIW											
	135	140	145	150	155	160	165	170	175	180	185	190
1	H5N1-NA 192 197 202 207 212 217 222 227 232 237 242 247 252 YNGIITDTIKSWRNNILRTQESE CACVNGSCFTVM TDGPSNGQASYKIFKMEKGKVVKSVELDAP											
2	H9-NA(1F8B) 212 217 222 227 232 237 242 247 252 257 262 267 272 YNRRPVTETNTWARNILRTQESE CVC HNGVCPVVF TDGSATGPAETRIYYFKECKILKWEPLAGT											
	200	205	210	215	220	225	230	235	240	245	250	255
1	H5N1-NA 257 262 267 272 277 282 287 292 297 302 307 312 NYHYEECS CYPDAGEITCVC RDNWHG SNRPWVSFNQ-NLEYQIGYICSGVFGDNPRPN-DGTGSC											
2	H9-NA(1F8B) 277 282 287 292 297 302 307 312 317 322 327 332 337 AKHIEECSCYGERAEITCTCRDNWQGSNRPVIRIDPVAMTHTSQYICSPVL TDNPRPNDPTVGKC											
	265	270	275	280	285	290	295	300	305	310	315	320
1	H5N1-NA 317 322 327 332 337 342 347 352 357 362 367 372 377 -GPVSPNGAYGVKGF SFKYGN GYWI GR TKSTNSRSGFEMI WDPNGWTGTDSSF SVKQDIVAITDW											
2	H9-NA(1F8B) 342 347 352 357 362 367 372 377 382 387 392 397 402 NDPYPGNNGVGVKGF SYLDGVNTWLGRTISIASRSGYEMLKVPNAL TDDKSKPTQGQITVLNTDW											
	330	335	340	345	350	355	360	365	370	375	380	385
1	H5N1-NA 382 387 392 397 402 407 412 417 422 427 432 437 442 SGYSGSFVQHPELTGLDCIRPCFWVELIRGRPKESTI-WTSGSS-SFCGVNSDTVGWSWPDGAEL											
2	H9-NA(1F8B) 407 412 417 422 427 432 437 442 447 452 457 462 SCYSGSFMDY-WAE-GE CYRACFYVELIRGRPKEDKVWVTSNSIVSMCSSTEF LGQWDWPDGAKI											
	395	400	405	410	415	420	425	430	435	440	445	450
1	H5N1-NA 447 PFTIDK											
2	H9-NA(1F8B) 467 EYFL											

图 5.1 序列H5N1-NA (Accession No. Q5H89) 和 N9-NA (Accession No. Q84070) 比对
图中氨基酸颜色表示氨基酸的功能：酸性—红色；碱性—蓝色；中性亲水—粉红色；脂肪族—墨绿色；芳香族—青绿色；包含巯基—黄色；亚胺基—橙色

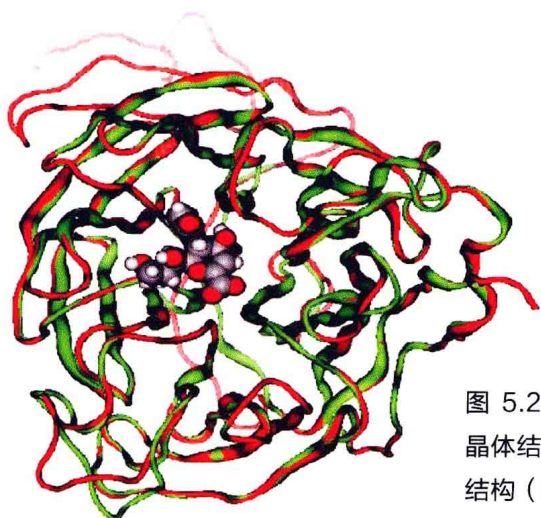


图 5.2 根据结合 DANA 的 N9-NA 复合物晶体结构 (红色) 构建的 H5N1-NA 的三维结构 (绿色)

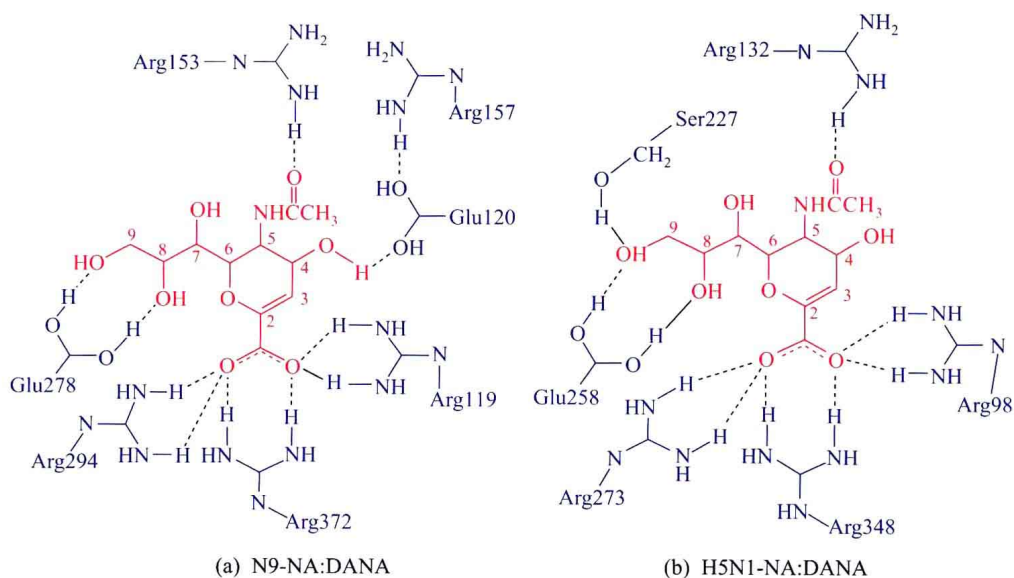


图 5.4 配体 DANA (红色) 分别与受体 N9-NA (a) 和 H5N1-NA (b) 活性位点残基 (蓝色) 相互作用示意图

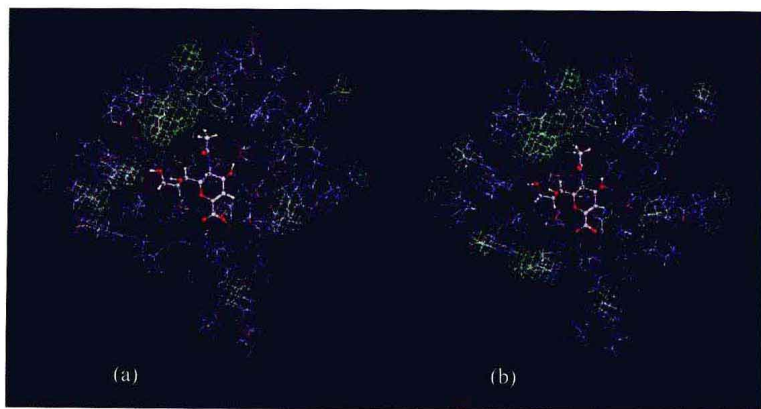


图 5.5 N9-NA (a) 和 H5N1-NA (b) 活性空腔的亲脂和亲水表面示意图
其中绿色表示亲脂性, 蓝色表示亲水性

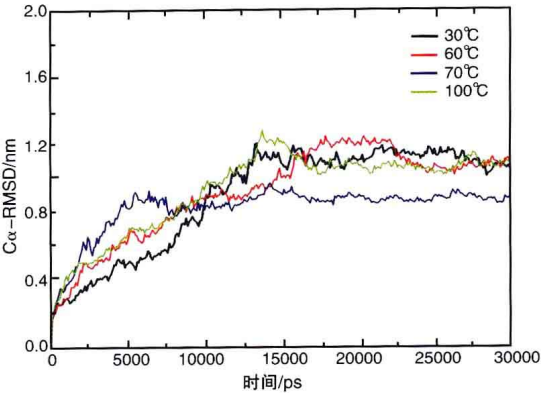


图 5.7 模拟过程中骨架碳原子的均方根偏差随时间变化示意图

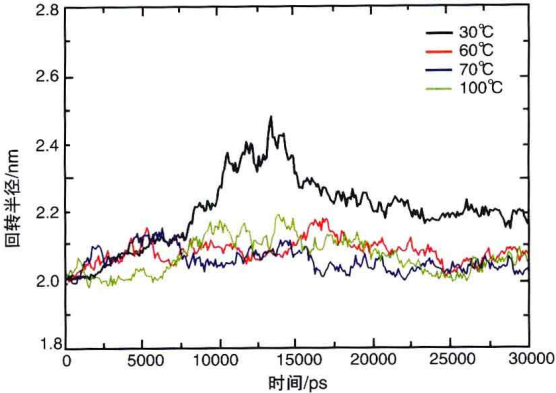


图 5.8 骨架碳原子的回转半径示意图

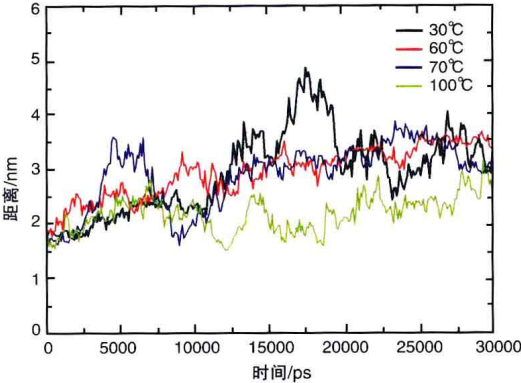


图 5.9 模拟过程中螺旋 6 与螺旋 9 之间的距离

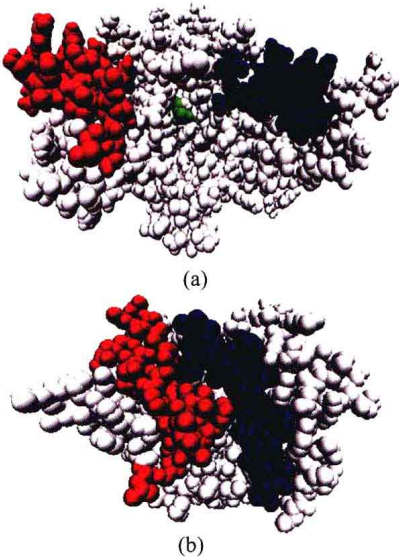


图 5.10 MD 模拟中 T1 脂肪酶打开 (a) 与关闭 (b) 的构象
红色代表螺旋9, 蓝色代表螺旋6, 绿色代表活性中心

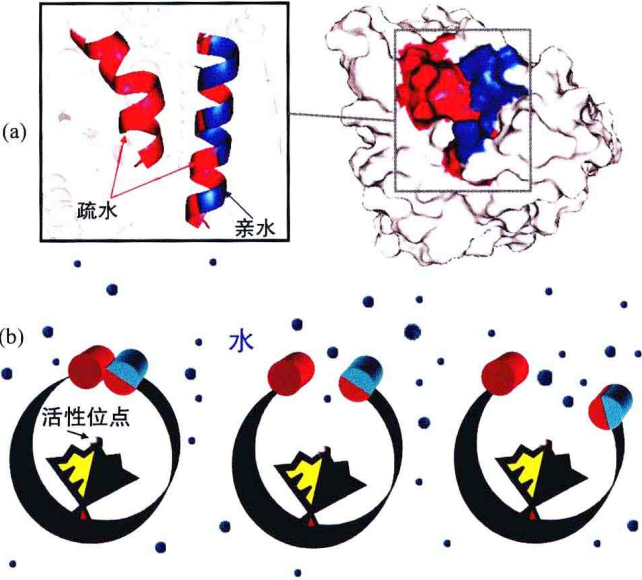


图 5.12 T1 脂肪酶结构变化的分子机理示意图
红色代表疏水, 蓝色代表亲水

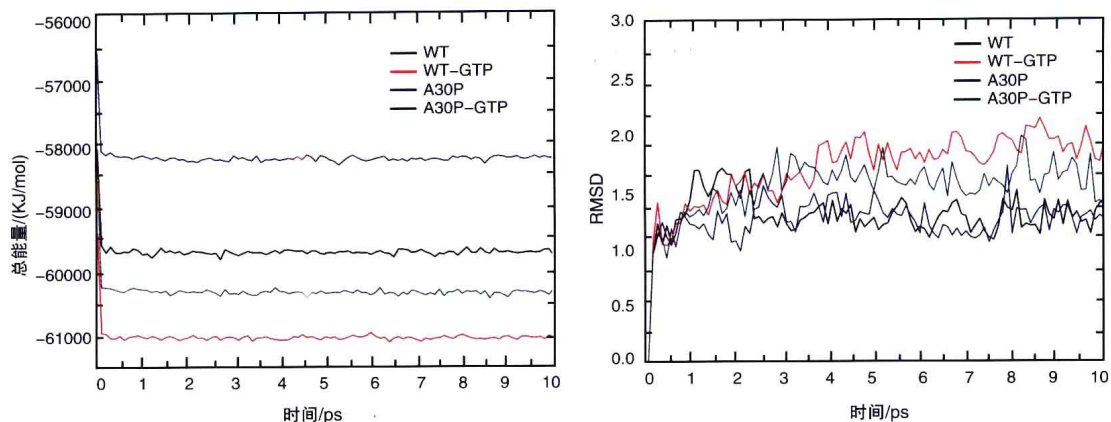


图 5.14 模拟过程中体系总能量和骨架碳原子的均方根偏差随时间变化示意图

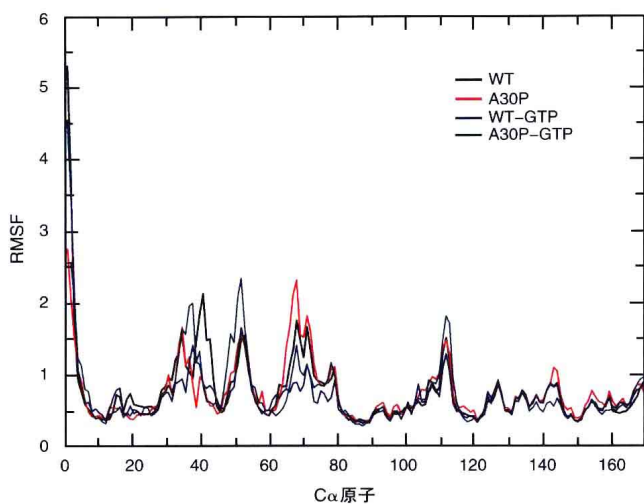


图 5.15 分子动力学过程中骨架碳原子的均方根波动示意图

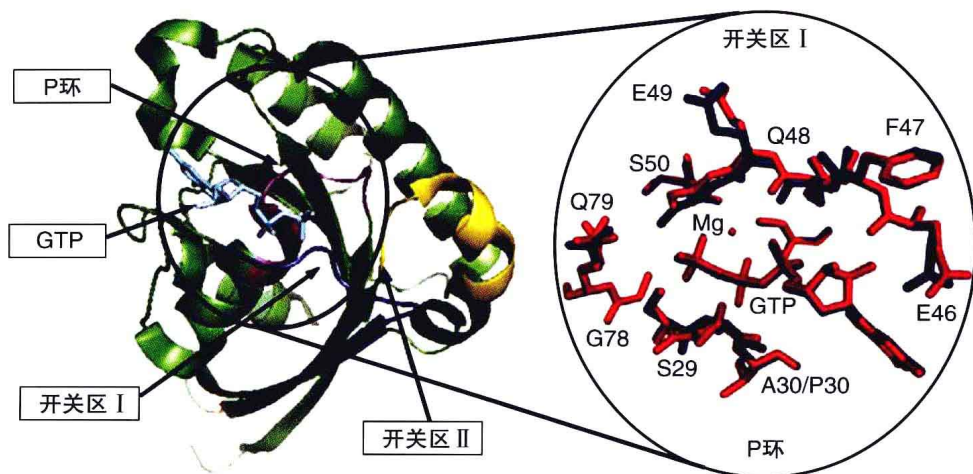


图 5.16 A30P突变所带来的结构变化示意图

图中野生型结构用蓝色棍状模型表示，而突变型结构则是用红色棍状模型表示

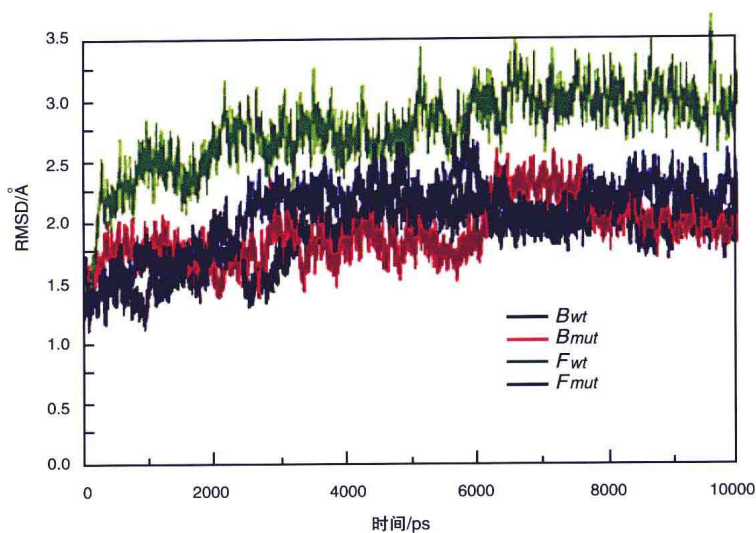


图 5.19 各体系蛋白酶骨架原子的 RMSD 随时间变化规律

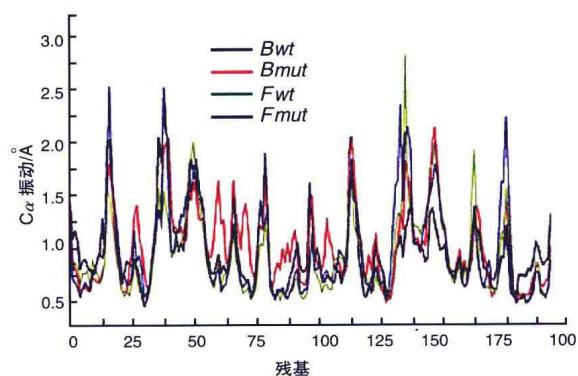


图 5.20 各体系中骨架碳原子在分子动力学模拟中的振动情况



图 5.22 *Bmut* 蛋白酶 (红色) 与 *Bwt* (黄色) 蛋白酶的空间结构叠合

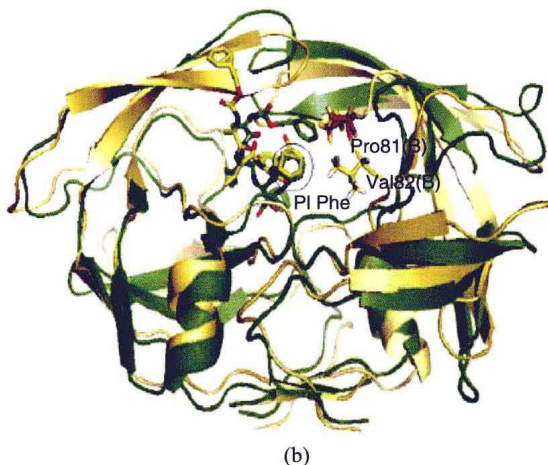
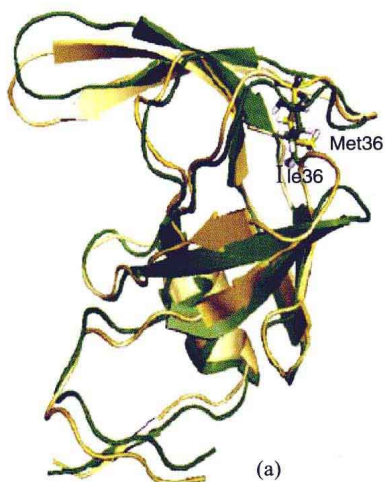


图 5.24 *Fwt* (绿色) 和 *Bwt* (黄色) 蛋白酶B链 (a) 和 S1/S1' 疏水口袋 (b) 的空间结构叠合

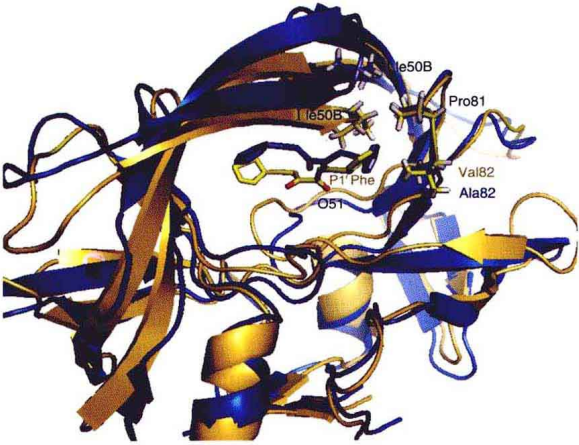


图 5.26 *Fmut* (蓝色) 和 *Bwt* (黄色) 蛋白酶空间结构叠合

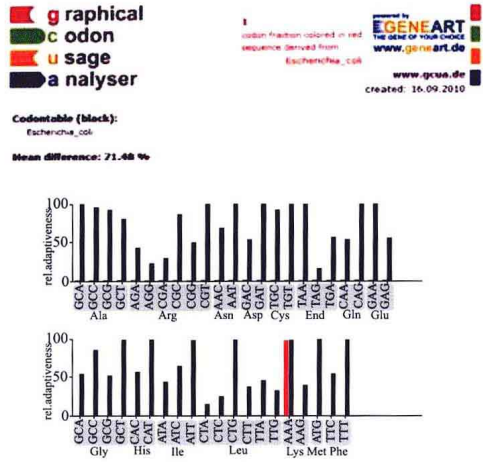


图 6.18 各种密码子在大肠杆菌中的使用频率

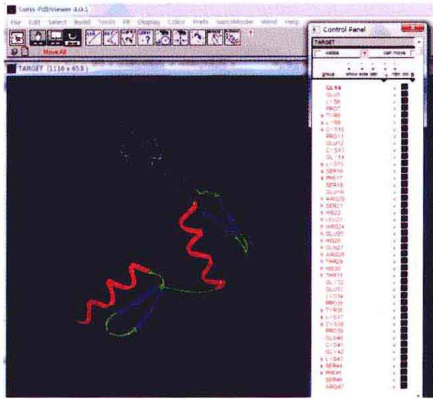


图 6.26 Deep View 软件工作界面

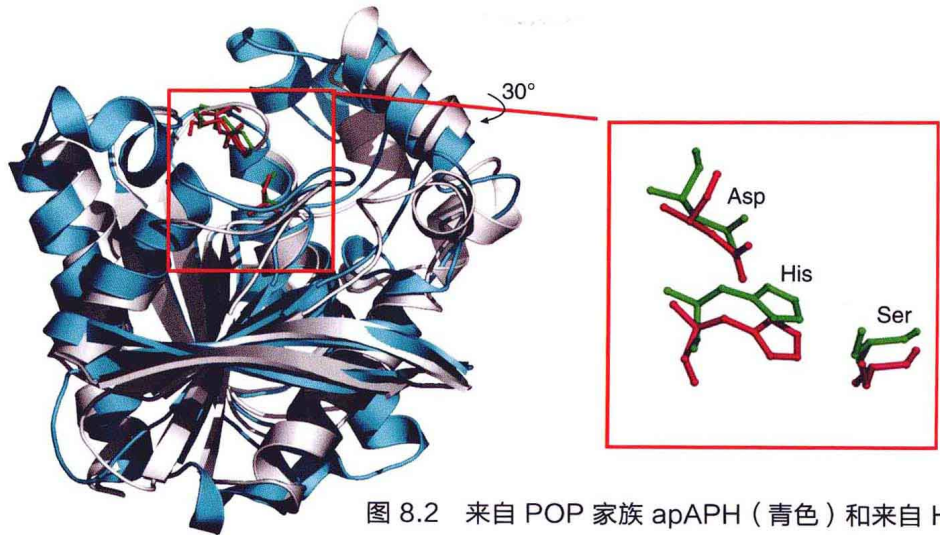


图 8.2 来自 POP 家族 apAPH (青色) 和来自 HSL 家族的 AFEST (灰色) 催化结构域和催化三联体的结构重叠

编 写 人 员

(按姓名汉语拼音排序)

陈仕高 程 超 冯 雁 李荣秀 连 鹏
王靖方 王 莹 魏冬青 仲从浩

丛书序

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、色谱、电泳、重原子同晶置换等技术以外，后来涌现的高压液相层析、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“十一五”国家重点图书出版项目，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第一批书目包括：

蛋白质物理学概论

蛋白质组学原理

蛋白质电泳技术指南

蛋白质色谱分离技术

蛋白质质谱技术

蛋白质微阵列

蛋白质高效表达技术

蛋白质结构模拟与设计

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

前言

蛋白质结构是生物科学的一个重要部分，与人类自身的重要哲学命题“生命是什么”密切相关。回顾这个研究领域的科学问题、研究方法、相应的理论和概念的发展以及相互促进的关系，能够启发对科学研究充满兴趣的青年及早考虑诸如此类的重要事项：什么是科学问题？科学问题是从哪里来的？重大科学发现的逻辑和巧妙是什么？科学问题的解决首先需要大胆的假设和推断，再借鉴和发展新方法和新工具，小心求证理论假设。

蛋白质结构研究的很多突破性发展都是从物理学领域借用研究方法和工具，如X射线衍射、核磁共振（NMR）、同步辐射。因此，科学工作者不仅需要与同行交流，也要时时关注相关科学领域的进展情况，有哪些理论、方法可以“借用”？有哪些启示可以收集？青年科学工作者心理要有所准备，在重大创新的实践活动中，不乏“十年磨一剑”的例子。在磨剑过程中潜心提升自己科学问题凝练能力，设计解决问题的有效方法，积累重要的科学数据作为攀登的“台阶”和“扶手”。此外，还要对自己的发现不能马上被“大师”们承认的风险有所准备。原始创新有时是孤独的，因此是勇敢者的游戏。

蛋白质结构研究中充满着大量的数学问题，请不要被其吓回头！阅读本书重在设想自己处在科学家们“无助”的位置时，会如何思考遇到的问题，能够有解决问题的不同逻辑吗？请读者与我一起来参加到科学家们的游戏中，体验一下他们的甘苦，重温发现的快乐与激动，也体验一下找不到解决问题出路时“无助”的挫折感，以及重大发现未被承认时的孤独感。你，只有及早体验这些感受，才能及早进入创新的角色中。

本书在介绍科学与技术内容的同时，也尽所能提供一些对蛋白质结构研究科学思想发展的节点和重要手段作出重要贡献的科学家的简短介绍。近不惑之年才刚刚悟出这样的道理：蛋白质结构研究手段及其科学思想发展都是本领域的科学家接力完成的。只有具备了某些必要的素质、品质和思维能力，才能接过前人思想的接力棒，进一步设计新的探索方向，提出科学假说，设计新的实验方法，建立新层次的科学思想和创新理论，才能将科学事业推到更高的境界。为什么只有少数人能够成为杰出科学家，为什么这些人才能够传承顶级的科学思想？正如“钱学森之问”：为什么我们的学校总是培养不出杰出人才？目前我们也刚有些朦胧的感觉，核心应该是“人”。

最后，建议读者在了解科技内容的同时，更要提醒自己：“需要培养什么样的思想、素质和能力才能创造新的科学知识？”创新的主体是“人”，读者每读过一章，可以挑一个自己有亲近感的科学家和具体问题，用互联网提供的便利工具，找出自己的“科学家”解决科学问题的每一个台阶和扶手的原始科学文献，看能否拼出其走过路线上的路标，其判断前进方向的逻辑和其中的精妙。也请读者以本书科学家小传内容为跳板，查阅更多关于杰出先辈们思想和行为的文献，找出自己需要培养什么思想品格，需要积累什么经验，进行什么练习？认识到了这些，并坚持身体力行，选定目标，成为大师是时间迟早的问题。

本书成稿仓促，错误难免，欢迎批评指正。

李荣秀

2011年3月

目 录

第 1 章 导论	1
1.1 生命是什么	1
1.2 生物体的有序结构	1
1.3 生物体功能与结构层次	2
1.4 蛋白质是生命活动的载体	2
1.4.1 蛋白质的功能	2
1.4.2 蛋白质结构	4
1.5 蛋白质结构模拟	5
1.5.1 序列比对	6
1.5.2 蛋白质结构比对和预测	6
参考文献	7
第 2 章 蛋白质结构研究的重大历史事件	8
2.1 假设+假设=诺贝尔奖	8
2.1.1 X 射线：波还是粒子？	8
2.1.2 晶体结构之谜	9
2.1.3 诺贝尔奖培训班	9
2.1.4 一石二鸟的精巧实验设计	10
2.2 父子同心：X 射线晶体学	11
2.3 蛋白质晶体学的兴起	12
2.3.1 蛋白质结构	12
2.3.2 多萝西·霍奇金：蛋白质结构测定时代的起始	12
2.3.3 马克斯·佩鲁茨：蛋白质结构测定技术的成熟	13
2.3.4 相位问题	16
2.3.5 又是相位问题：核糖体结构的解析	18
2.4 核磁共振方法	20
2.4.1 傅里叶变换和多维核磁共振	20
2.4.2 测定蛋白质结构的核磁共振方法	22
2.5 蛋白质结构测定的其他辅助工具	24
2.5.1 实验方法	24

2.5.2 计算及结构显示工具	24
2.6 生物分子、复合物及病毒结构测定	25
2.6.1 重要生物学功能蛋白的测定	25
2.6.2 病毒结构的测定	27
2.7 同步辐射光源及发展趋势	28
参考文献	30
第3章 蛋白质结构基础理论	32
3.1 蛋白质结构形成的本质	32
3.2 分子势能	33
3.2.1 量子力学	33
3.2.2 Born-Oppenheimer 近似	34
3.2.3 势能函数	36
3.3 分子内基团识别的要素	37
3.3.1 分子之间相互作用	37
3.3.2 静电相互作用	38
3.3.3 氢键	39
3.3.4 溶剂或疏水相互作用	39
3.3.5 伦敦类型的相互作用	41
3.3.6 原子核在超短距离时的排斥作用	42
3.4 识别的专一性	42
3.5 基团间识别与接触	44
参考文献	45
第4章 从序列到结构	46
4.1 背景介绍	46
4.2 蛋白质二级结构预测	47
4.2.1 Chou-Fasman 方法	47
4.2.2 GOR 方法	47
4.2.3 机器学习	48
4.3 蛋白质三级结构预测	49
4.3.1 蛋白质结构的从头预测	50
4.3.2 蛋白质结构的同源建模	50
4.3.3 同源建模实例	53
4.4 蛋白质四级结构预测	56
4.5 结束语	57
参考文献	57

第 5 章 蛋白质结构模拟在蛋白研究中的应用实例	59
5.1 引言	59
5.2 蛋白质结构的预测	60
5.3 分子对接	61
5.4 蛋白结构预测与分子对接在生物研究中的应用	64
5.5 分子动力学模拟在蛋白结构变化中的应用	67
5.5.1 研究背景	68
5.5.2 实验方法	68
5.5.3 实验结果	69
5.5.4 结论	71
5.6 分子动力学模拟研究氨基酸突变对蛋白结构的影响	72
5.6.1 研究背景	72
5.6.2 方法介绍	73
5.6.3 实验结果与分析	74
5.6.4 结论	76
5.7 靶标蛋白与抑制剂的相互作用研究	77
5.7.1 研究背景	77
5.7.2 实验方法	78
5.7.3 体系分子动力学轨迹分析	79
5.7.4 结构变化分析	81
5.7.5 结论	85
5.8 本章总结	85
参考文献	86
第 6 章 锌指核酸酶的构建	88
6.1 锌指蛋白	88
6.1.1 锌指蛋白的分类	89
6.1.2 锌指蛋白的作用及机制	89
6.1.3 锌指蛋白与免疫性疾病	91
6.2 锌指核酸酶	92
6.2.1 核酸酶	92
6.2.2 锌指核酸酶	93
6.3 锌指核酸酶的设计	94
6.3.1 靶位点的确定	94
6.3.2 ZFN 蛋白序列的设计及优化	95
6.3.3 锌指核酸酶设计举例	95