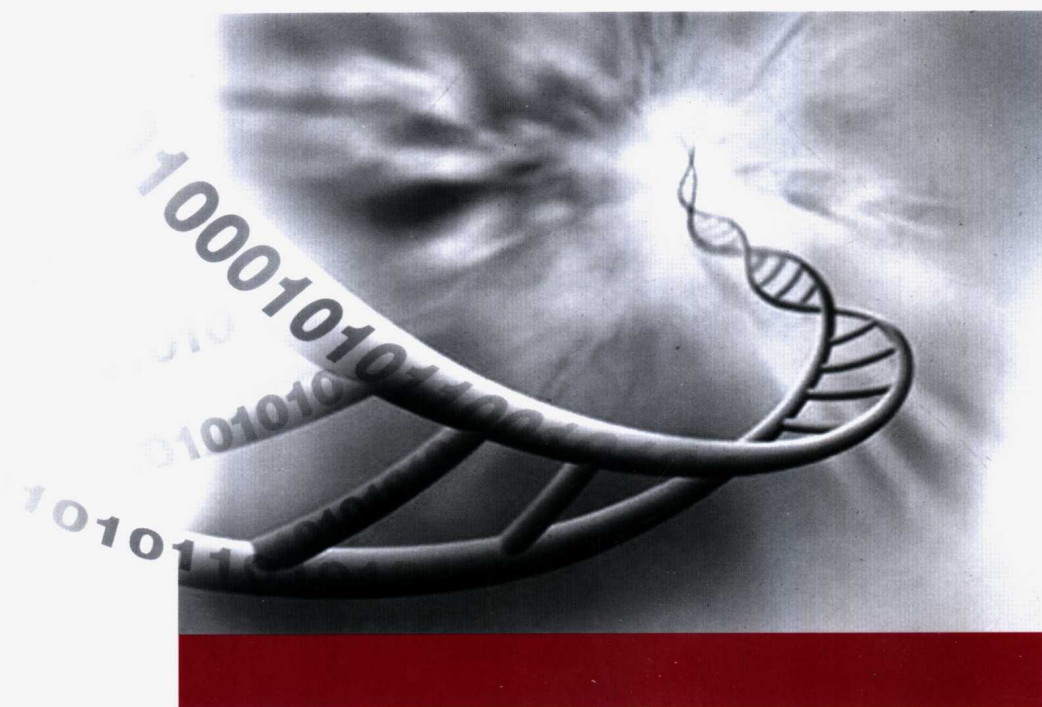


郑 珩 王 非 编著

药物生物信息学



Chemical Industry Press



化学工业出版社
现代生物技术与医药科技出版中心

药物生物信息学

郑 珩 王 非 编著



化学工业出版社
现代生物技术与医药科技出版中心

· 北 京 ·

(京) 新登字 039 号

图书在版编目 (CIP) 数据

药物生物信息学/郑珩, 王非编著. —北京: 化学工业出版社, 2004. 3
ISBN 7-5025-5318-5

I. 药… II. ①郑…②王… III. 药物-生物信息论 IV. R915

中国版本图书馆 CIP 数据核字 (2004) 第 013983 号

药物生物信息学

郑 珩 王 非 编著

责任编辑: 余晓捷 杨燕玲 孙小芳

文字编辑: 李 丽

责任校对: 李 林

封面设计: 关 飞

*

化 学 工 业 出 版 社 出版发行
现代生物技术与医药科技出版中心

(北京市朝阳区惠新里 3 号 邮政编码 100029)

发行电话: (010) 64982530

<http://www.cip.com.cn>

*

新华书店北京发行所经销

聚鑫印刷有限责任公司印刷

三河市延风装订厂装订

开本 787 毫米×1092 毫米 1/16 印张 19 1/4 字数 447 千字

2004 年 4 月第 1 版 2004 年 4 月北京第 1 次印刷

ISBN 7-5025-5318-5/R · 202

定 价: 45.00 元

版权所有 违者必究

该书如有缺页、倒页、脱页者, 本社发行部负责退换

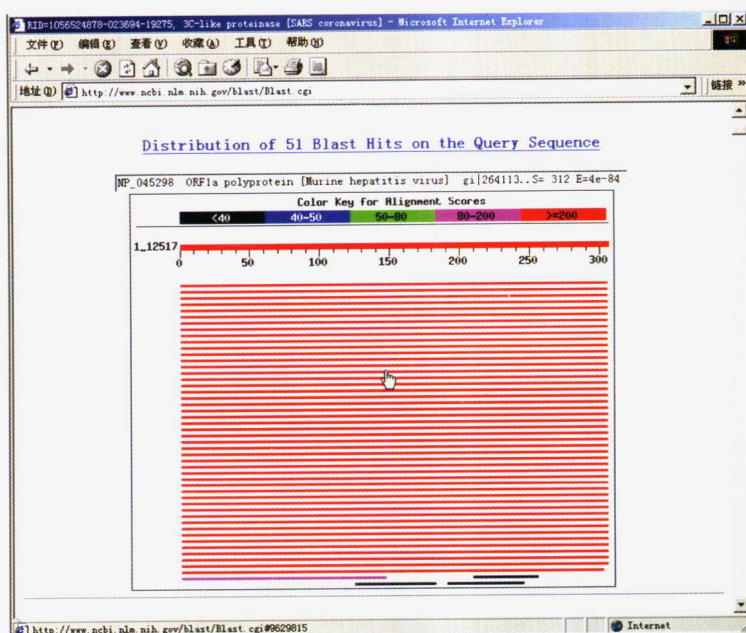


图 6-24 搜索结果总览图

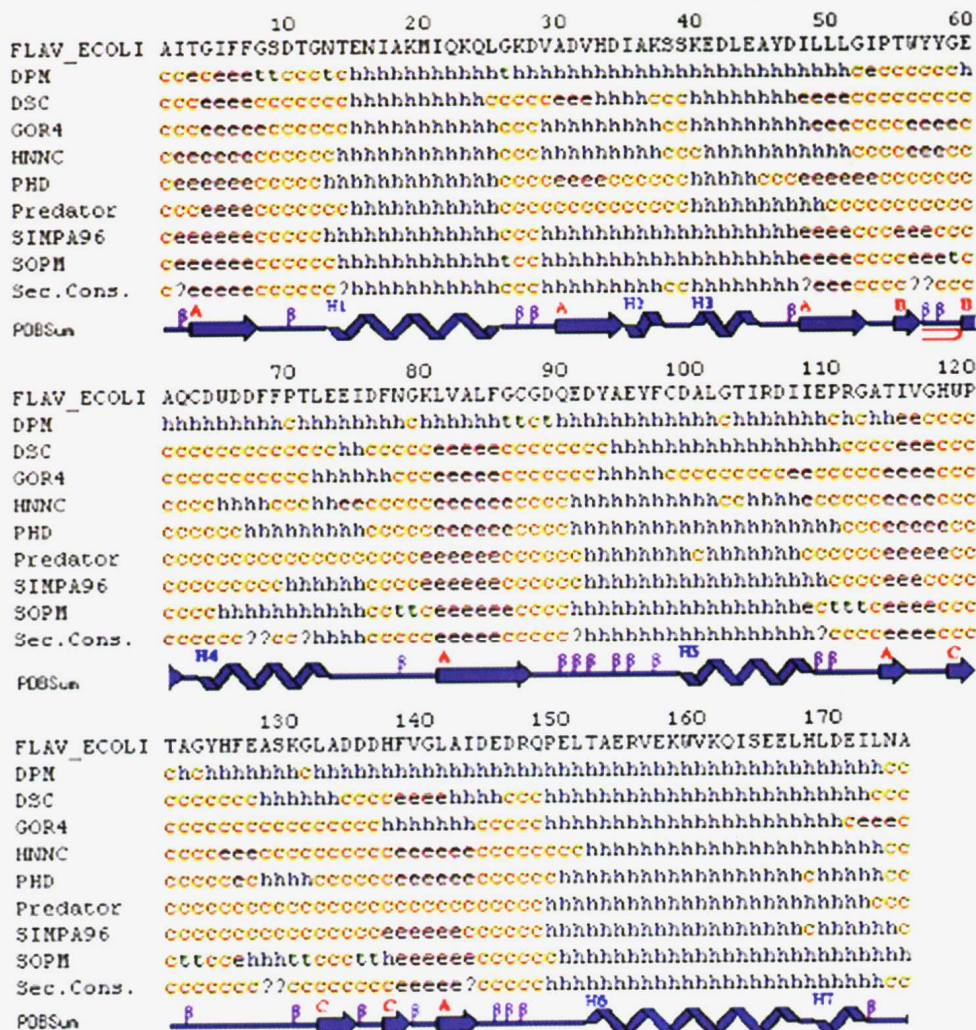


图 8-11 大肠杆菌黄毒素蛋白用NPS@二级结构预测服务预测的结果

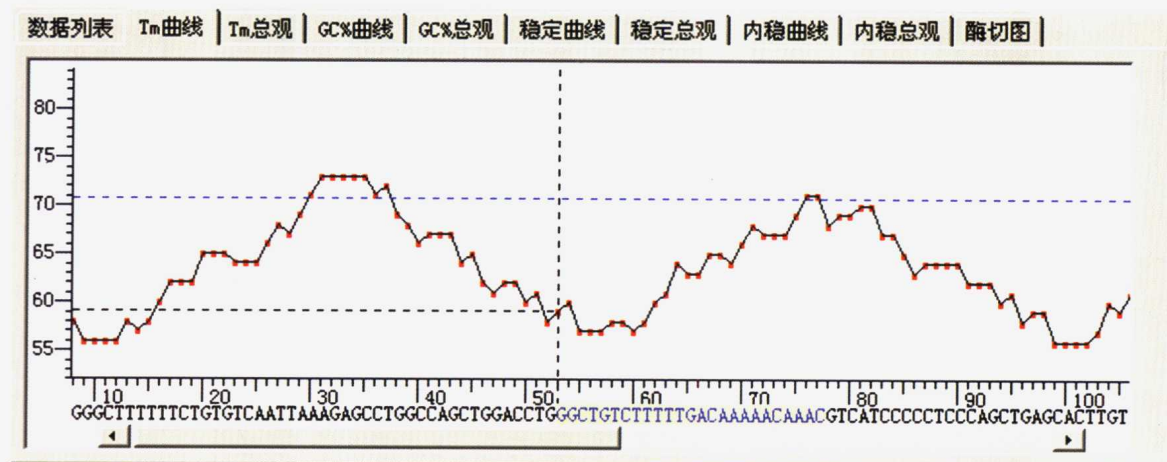


图 9-5 参数曲线显示



图 9-13 DNA序列在3个起始相位下翻译成蛋白质的结果

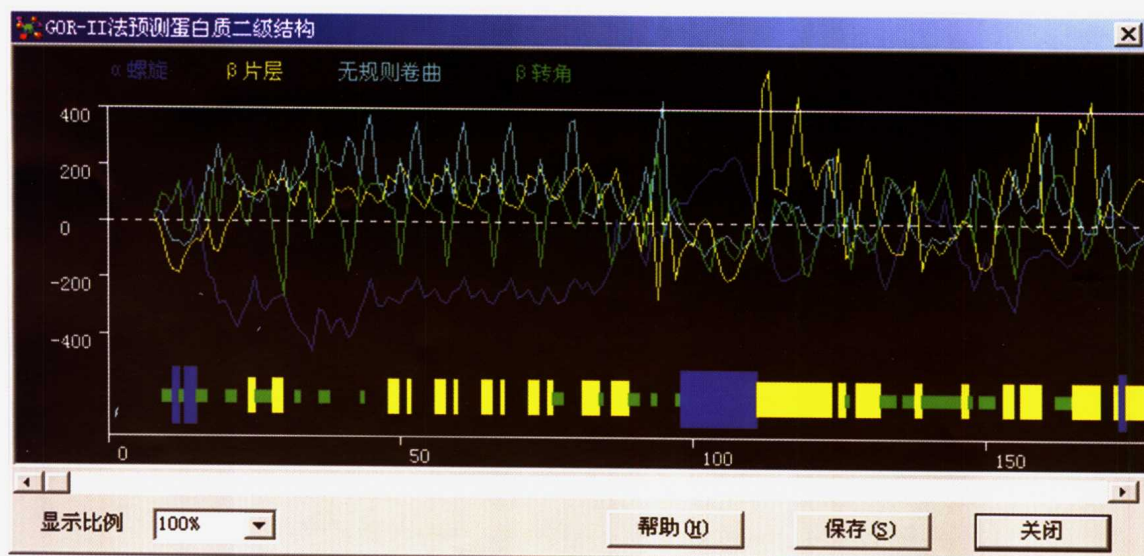


图 9-17 蛋白质二级结构预测

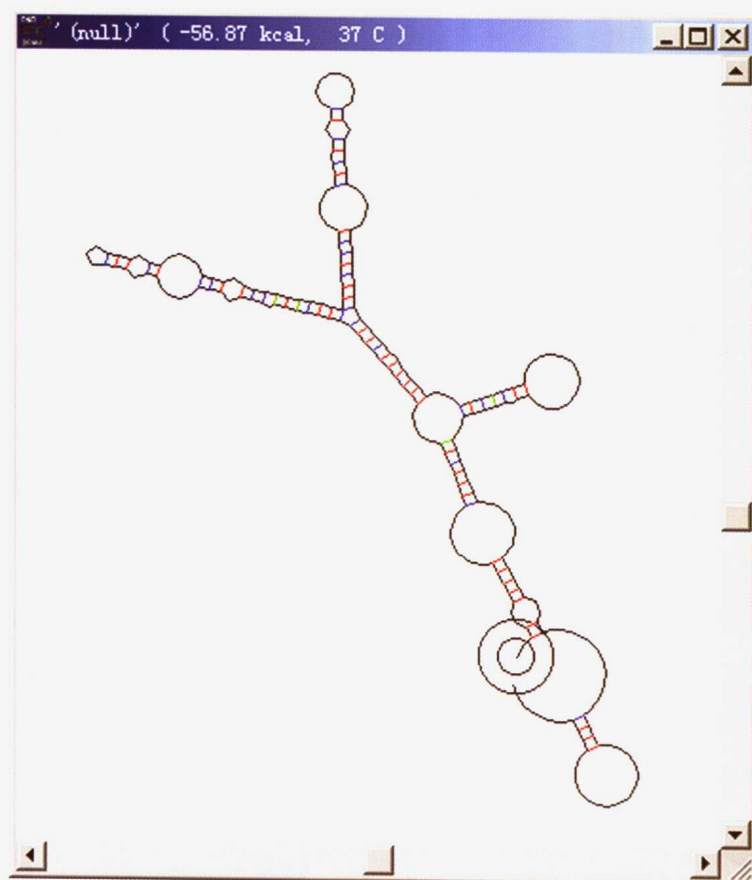


图9-28 某RNA序列的结构预测

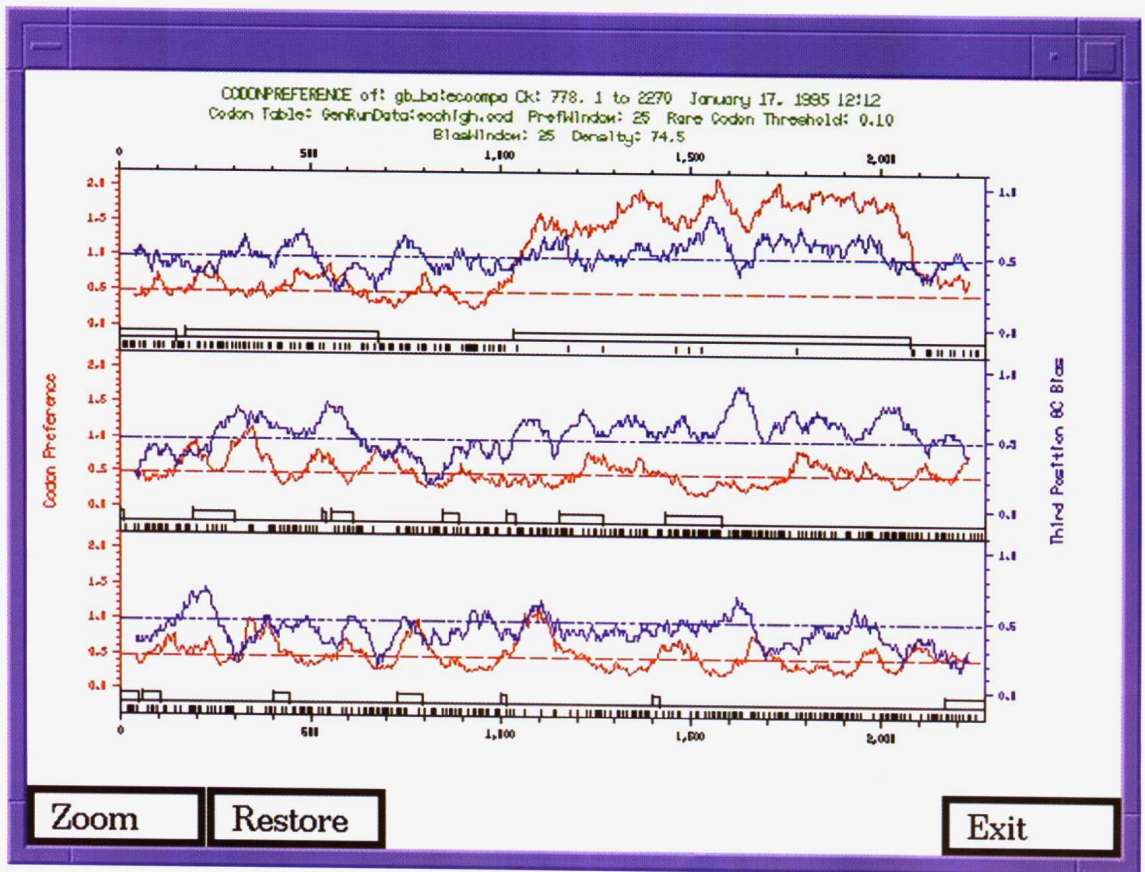


图 9-33 CodonPreference预测蛋白质编码区

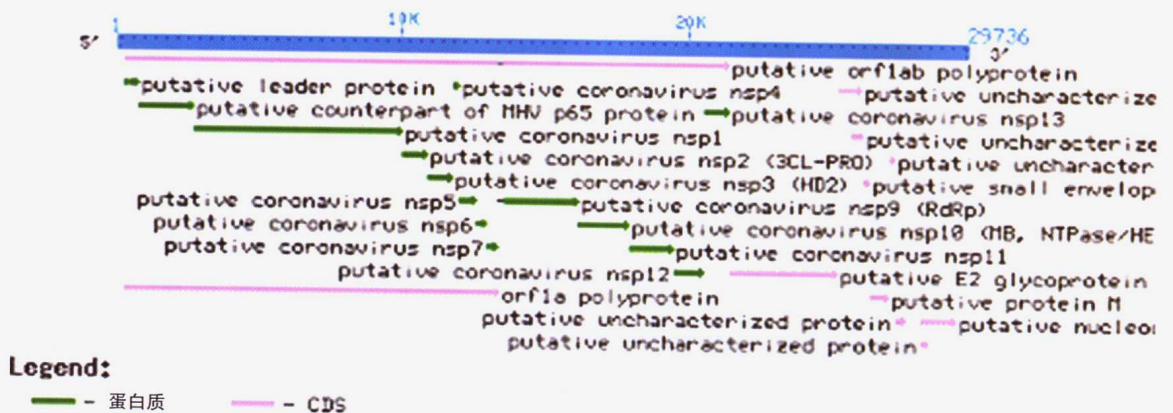


图 10-3 SARS基因组分析结果

序

21 世纪是生命科学的时代，也是信息时代。随着人类基因组计划的实施与完成，有关核酸、蛋白质的序列和结构数据呈指数增长。面对巨大而复杂的数据，运用计算机管理数据、控制误差、加速分析过程势在必行。从 20 世纪 80 年代末开始，生物信息学逐渐兴起并蓬勃发展，业已成为一门工具型学科，广泛渗透到医学、药学、农学等相关领域，从而衍生出一系列交叉性前沿学科。

基因组学和生物信息学研究将极大地改变新药开发的思路与模式。对基因组计划产生的核酸序列分析一方面可从本质上认识疾病尤其是遗传性疾病的发生原因，为这些疾病的诊断、预防和治疗奠定基础；另一方面，这些核酸序列自身就是丰富的药物资源，其中蕴含大量目前尚不明的基因，这些基因产物有的是调节机体机能的蛋白质类药物，有的可作为药物作用潜在的靶点而应用于新的药物开发。更重要的是，从药物作用与人类基因型的关系开拓了药物基因组学这一新的研究领域，为减小药物副作用、实现针对基因型的个性化治疗展示了宏伟蓝图。

对生物大分子信息处理分析为基于结构生物学的合理药物设计奠定了基础。通过偶然发现和随机生理活性筛选的方法已发现许多药物，但其盲目性较大，目前用这些方法发现药物已越来越难。随着分子生物学、分子药理学、计算机科学等的发展，根据药物与其作用靶点相互关系来设计新药成为当前国内外新药研究的重要手段，并已成功应用于酶抑制剂、受体抑制剂的研究开发；在蛋白质和多肽类药物分子设计方面，对新型蛋白质或多肽的分析及计算机辅助结构预测更不可缺少；生物信息的积累同时使计算机辅助抗原性分析、表位预测等成为可能，为降低生物大分子药物免疫原性，或提高疫苗活性等的分子设计研究提供了理论分析手段。

鉴于国内尚缺乏系统性、针对性介绍药物研发中的生物信息学工具参考书，作者编著了适用于我国高等教育的《药物生物信息学》。该书覆盖了核酸和蛋白质分析、计算机辅助药物设计、疫苗设计、生物芯片技术、药物基因组学等，较全面地介绍了生物信息学基本原理和技术，及其在药物研究开发中的应用，使学生通过本门课程的学习，可以较好地掌握药物生物信息学的基本理论、基本知识和基本技能。对于从事药学、生物技术、生命科学及生物信息学研究的科技工作者，也具有重要的参考价值。

吴梧桐
2004 年 1 月

前 言

药物生物信息学是一门崭新的综合性交叉学科。随着人类基因组计划的实施和深入,核酸、蛋白质序列及相关的生物信息数据呈爆炸性增长,这些生物信息为药学研究提供了丰富的资源,如从人类基因组序列中可预测和发现新基因,它们的产物可能是新的蛋白质类药物或作为药物开发的靶标;生物体尤其是病原微生物特异性基因序列分析,可为感染性疾病诊断芯片的设计奠定基础;蛋白质结构相关信息,可为虚拟药物筛选和合理药物设计提供帮助。另一方面,这些生物信息资源的处理分析和开发利用也对药学工作者提出了新的挑战,对这些错综繁杂的海量生物信息的提取和分析,离不开计算机和各种分析模型、算法软件的支持。药学、生物学、物理学、数学和计算机科学等学科的相互交叉渗透,促进了药物生物信息学这一新的研究领域的诞生。简单地说,药物生物信息学就是综合应用生命科学、数学、计算机科学等多学科的理论和方法,对伴随基因组计划产生的海量生物信息进行整理和分析,然后应用于药物的设计和开发,以达到合理药物设计目的的一门交叉性学科。

本书较系统地介绍了药物生物信息学这一新兴学科概况,全书共分13章,前六章主要是生物信息学的基本原理,包括生物信息学概述、相关计算机基础、核酸和蛋白质数据库、常用生物信息和医药信息资源及检索方法、序列比对和数据库搜索原理等。后面七章则偏重于生物信息学原理和技术在药物开发上的应用,如核酸分析、蛋白质分析、新药开发相关生物信息学软件、计算机辅助药物设计、计算机辅助疫苗设计、生物芯片和药物基因组学等。药物生物信息学是一门实践性很强的学科,本书选择了一些应用性实例,如SARS病毒基因序列检索及分析、计算机辅助HIV蛋白酶抑制剂设计等,希望能给读者一些启发,促进我国生物信息学的普及和应用,提高药物研发和创新能力。本书可作为药学、生化药学、生物信息学、生物技术制药等专业的高年级本科生、研究生教材,也可供从事药物研发工作的技术人员参考。

本书编写工作自2001年底始,于2002年9月完成初稿。在初稿完成后,我们印制了试用教材在中国药科大学生物工程和生物技术专业高年级本科生以及生化药学专业研究生所修“药物生物信息学”课程中试用,获得了良好的教学效果及宝贵的反馈信息。与此同时,鉴于药物生物信息学及其相关学

科技发展迅速，新理论和新方法层出不穷，我们数易其稿，以期本书不但能够适用于课程教学，还能够反映该学科的最新进展。

本书编写过程中，李菁博士在蛋白质结构预测部分给予了很大的帮助并编写了蛋白质分析一章中的部分内容，杨欣博士在生物信息软件部分给予了有益的建议并提供了部分参考材料。与远在美国的吴滔博士、胡容博士在生物信息学内容方面的讨论，以及张丽君博士、宋鹏飞博士赠送的《Computational Molecular Biology》一书，都使作者受益匪浅，在此对他们的支持和帮助表示由衷的感谢。同时特别感谢化学工业出版社有关编辑，他们在本书出版过程中所付出的大量而卓有成效的工作保证了本书的顺利出版。

药物生物信息学发展迅速，新理论、新技术、新发现和新成果层出不穷，书中取舍不当、疏漏和错误之处，恳请各位同行和广大读者批评指正。

郑珩 王非

2004 年 1 月

内 容 提 要

药物生物信息学是伴随人类基因组计划的实施和深入而蓬勃发展起来的一门新兴学科,它综合应用生命科学、数学、计算机科学等多学科的理论和方法,对伴随基因组计划产生的海量生物信息进行整理和分析,然后应用于药物的设计和开发,以达到合理药物设计的目的。

药物生物信息学是一门应用性、实践性很强的学科。为突出这一特点,本书偏重于药学相关生物信息的获取、分析和在药物开发上的应用,内容上涵盖了生物信息学概述,必备的信息技术基础,常用的生物信息及医药信息资源,序列比对及数据库搜索基本方法,核酸序列分析方法,蛋白质序列分析方法,新药开发相关生物信息学软件,计算机辅助药物设计,计算机辅助疫苗设计,生物芯片和药物基因组学等,较全面地阐述了生物信息学基本原理、分析方法、常用软件以及在药物开发各环节上的应用。

本书可作为医药学、生命科学相关高年级本科生、研究生教材,也可作为药物研发人员和其他专业师生学习参考。

目 录

第一章 生物信息学概述	1
第一节 生物信息学及其发展	1
一、什么是生物信息学	1
二、生物信息学的诞生和发展	1
三、生物信息学研究的主要推动力	3
第二节 生物信息学的研究领域	5
一、生物信息学在生物学领域的研究内容	5
二、生物信息学在计算机领域的研究内容	13
第三节 生物信息学在药学领域的应用	14
一、初始阶段	14
二、药物设计阶段	15
三、药物开发阶段	17
参考文献	17
第二章 计算机网络基础	19
第一节 Internet 简介	19
一、Internet 的建立和发展	19
二、TCP/IP 协议	20
三、IP 地址与子网掩码	20
四、域名系统	22
第二节 Internet 提供的服务	23
一、World Wide Web	23
二、电子邮件	23
三、文件传输 (FTP)	24
第三节 搜索引擎及其使用	24
一、搜索引擎基本类型	24
二、常用搜索引擎简介	25
三、搜索引擎的使用方法	28
参考文献	30
第三章 核酸序列及基因组数据库	31
第一节 核酸序列数据库	31
一、概述	31
二、GenBank 数据库	33
三、EMBL 数据库	38

第二节 基因组与功能基因组数据库	44
一、GDB	44
二、UniGene	45
三、SGD	45
四、TDB	46
五、AceDB	46
六、dbEST	46
七、TRANSFAC	46
八、TRRD	47
九、COMPEL	47
十、CUTG	48
参考文献	48
第四章 蛋白质序列及结构数据库	50
第一节 蛋白质序列数据库	50
一、概述	50
二、SwissProt 蛋白质序列数据库	50
三、PIR 蛋白质序列数据库	55
四、PROSITE	56
五、PRINTS	58
六、TrEMBL	59
七、NRL-3D	59
第二节 蛋白质序列复合数据库	59
一、NRDB	60
二、OWL	60
三、MIPSX	60
四、SWISS-PROT+TrEMBL	61
第三节 蛋白质结构数据库	61
一、蛋白质结构数据库 PDB	61
二、结构信息显示	65
三、结构浏览器	68
第四节 蛋白质结构分类数据库	70
一、蛋白质结构域的折叠模式	70
二、蛋白质结构分类数据库	71
参考文献	74
第五章 生物医药学信息检索	76
第一节 集成式数据库检索系统	76
一、Entrez 系统	76
二、SRS 系统	85
第二节 生物医药文献资源	90

一、Medline	90
二、医药搜索引擎	91
三、专利全文检索系统	92
四、中文生物医药文献检索	93
参考文献	97
第六章 序列比对方法和相似性搜索	98
第一节 概述	98
第二节 双重序列比对	99
一、序列比对的基本概念	99
二、动态规划算法与全局比对	100
三、局部比对	103
四、打分矩阵和空位罚分	104
第三节 多序列比对	109
第四节 点阵作图法	112
第五节 序列比对的统计学显著性	115
第六节 数据库中的相似性搜索	116
一、概述	116
二、FASTA 搜索	117
三、BLAST 搜索	120
参考文献	127
第七章 核酸序列分析	129
第一节 基因结构分析	129
一、生物体基因结构	129
二、基因识别及预测	132
三、其他位点分析	137
四、基因结构分析的一般步骤	141
第二节 核酸序列组装和基因定位分析	142
一、核酸序列组装及电子克隆	142
二、基因定位分析	144
第三节 新基因的发现及功能预测	145
一、利用 EST 数据库 (dbEST) 发现新基因	145
二、从基因组 DNA 序列中预测新 ORF	146
三、通过核酸序列同源性比较进行功能预测	146
参考文献	146
第八章 蛋白质序列分析	147
第一节 蛋白质组成及基本性质	147
一、分子量及等电点	147
二、蛋白质辨识	147
三、酶切及断裂位点	149

四、疏水性	150
第二节 蛋白质二级结构预测	152
一、二级结构预测方法	152
二、二级结构预测软件	155
三、二级结构预测的准确度	157
第三节 特殊结构或结构特征分析	159
一、跨膜区	159
二、信号肽和蛋白质定位	163
三、卷曲螺旋分析	163
第四节 蛋白质三级结构预测	164
一、同源模建	164
二、折叠模式识别	165
三、从头预测	167
参考文献	168
第九章 新药开发中的生物信息学软件	170
第一节 概述	170
一、核酸和蛋白质序列分析	170
二、蛋白质高级结构分析	171
三、蛋白质组研究图谱的处理	171
四、蛋白质、核酸三维结构显示	171
五、格式转换	172
第二节 PCR 引物设计及其软件介绍	172
一、引物设计的原则	172
二、WONDERFUL 生物信息学系统的引物设计功能介绍	173
第三节 DNA 序列分析软件	177
一、酶切位点分析	178
二、核酸基序分析	178
三、环状序列分析	179
四、开放阅读框 ORF 搜索	179
五、CpG 岛搜索	179
六、编码测试法预测编码区	180
七、序列转换	181
八、序列统计	181
第四节 蛋白质序列分析软件	182
一、蛋白酶切位点分析	182
二、蛋白质基序分析	183
三、蛋白质特性分析	183
四、蛋白质二级结构预测	184
第五节 序列比对软件	184

一、双序列比对·····	185
二、多序列比对·····	187
第六节 RNA 结构预测及其软件介绍 ·····	188
一、RNA 二级结构预测的理论和方法 ·····	188
二、RNA 结构预测软件 ·····	191
第七节 生物信息软件的工业标准 GCG 简介 ·····	194
第八节 生物分子与建模软件 Insight II ·····	199
一、功能及模块·····	199
二、应用领域·····	203
参考文献·····	206
第十章 计算机辅助药物设计 ·····	207
第一节 生物信息学与药物设计·····	207
一、新药研究与开发过程·····	207
二、合理药物设计·····	208
三、生物信息学与新药设计·····	209
第二节 先导化合物的计算机辅助设计·····	212
一、计算机辅助先导化合物设计的基本原理和方法·····	212
二、基于分子可视化和建模技术设计先导化合物·····	213
三、三维结构搜寻·····	214
四、全新药物设计·····	218
五、间接药物设计·····	221
第三节 先导化合物的计算机辅助优化·····	222
一、定量构效关系中的 Hansch 方法 ·····	222
二、定量构效关系研究中的 Free-Wilson 方法 ·····	223
三、分子连接性法·····	223
四、三维定量构效关系 (3D-QSAR) ·····	224
第四节 新型多肽及蛋白质分子设计·····	227
一、定点突变·····	228
二、盒式突变·····	228
三、蛋白质融合·····	229
四、蛋白质的从头设计·····	229
参考文献·····	231
第十一章 计算机辅助疫苗设计 ·····	233
第一节 免疫学基本概念·····	233
一、基本概念·····	233
二、主要组织相容性系统·····	233
三、人工免疫与疫苗·····	235
四、新疫苗的研制和展望·····	236
第二节 免疫应答·····	238

一、概述	238
二、免疫应答的细胞学基础	238
三、T 细胞介导的免疫应答	239
四、B 细胞介导的免疫应答	240
第三节 抗原表位预测方法	240
一、表位测定方法	240
二、基于基序的表位预测方法	241
三、基于人工神经网络的表位预测	243
四、基于分子模拟的表位预测方法	244
第四节 计算机辅助疫苗设计相关软件及应用	244
一、SYFPEITHI	245
二、TEPITOPE	246
三、MHCPre d	246
四、其他软件及数据库	247
参考文献	248
第十二章 生物芯片技术	249
第一节 基因芯片	249
一、基因芯片的制备	250
二、基因芯片的应用	251
第二节 蛋白质芯片	255
一、蛋白质芯片的类型与制备	255
二、蛋白质芯片的应用	256
三、存在问题和发展前景	258
第三节 生物芯片技术在新药研究中的应用	258
一、基因芯片与药物靶标	258
二、药理作用及药物毒性的评价	259
三、药物筛选	260
四、中药标准化鉴定	261
五、诊断芯片	261
第四节 生物芯片与生物信息学	262
一、确定芯片检测目标	262
二、芯片设计	263
三、生物芯片检测数据分析	269
参考文献	270
第十三章 药物基因组学	272
第一节 药物基因组学的发展	272
一、概述	272
二、药物基因组学的产生和发展	273
第二节 药物基因组学的研究内容和方法	274