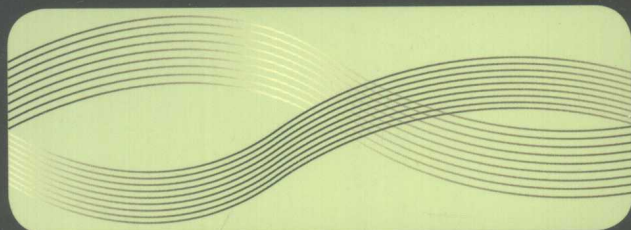




卫生部“十一五”规划教材

全国高等医药教材建设研究会规划教材

全国高等学校教材
供8年制及7年制临床医学等专业用



生物信息学

Bioinformatics

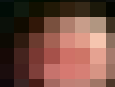
主 编 李 霞

副主编 李亦学 廖 飞



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE



第 1 版 (2013 年 1 月)

第 1 版 (2013 年 1 月) 第 1 次印刷

第 1 版 (2013 年 1 月) 第 1 次印刷

第 1 版 (2013 年 1 月) 第 1 次印刷



生物信息学

第 1 版 (2013 年 1 月) 第 1 次印刷

第 1 版 (2013 年 1 月) 第 1 次印刷

第 1 版 (2013 年 1 月) 第 1 次印刷

- 卫生部“十一五”规划教材
- 全国高等医药教材建设研究会规划教材
- 全国高等学校教材
- 供8年制及7年制临床医学等专业用

生物信息学

Bioinformatics

主 编 李 霞

副 主 编 李亦学 廖 飞

编 委 (以姓氏笔画排序)

田 心 (天津医科大学)	朱 浩 (南方医科大学)
刘建国 (河北大学)	许丽艳 (汕头大学)
李 霞 (哈尔滨医科大学)	李亦学 (同济大学)
吴忠道 (中山大学)	张 岩 (哈尔滨医科大学)
茆灿泉 (西南交通大学)	赵雨杰 (中国医科大学)
胡福泉 (第三军医大学)	童隆正 (首都医科大学)
廖 飞 (重庆医科大学)	魏冬青 (上海交通大学)

学术秘书 汪强虎 (哈尔滨医科大学)



人民卫生出版社

图书在版编目 (CIP) 数据

生物信息学 / 李霞主编. —修订本. —北京: 人民卫生出版社, 2010.8

ISBN 978-7-117-12938-1

I. ①生… II. ①李… III. ①生物信息论—高等学校—教材 IV. ①Q811.4

中国版本图书馆 CIP 数据核字 (2010) 第 105644 号

门户网: www.pmph.com	出版物查询、网上书店
卫人网: www.ipmph.com	护士、医师、药师、中医师、卫生资格考试培训

版权所有, 侵权必究!

本书本印次封底贴有防伪标。请予识别。

生物信息学

主 编: 李 霞

出版发行: 人民卫生出版社 (中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-67605754 010-65264830

010-59787586 010-59787592

印 刷: 北京金盾印刷厂

经 销: 新华书店

开 本: 850×1168 1/16 印张: 31.5

字 数: 932 千字

版 次: 2010 年 8 月第 1 版 2010 年 8 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-12938-1/R·12939

定价 (含光盘): 90.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社销售中心联系退换)

第二版出版说明

全国高等学校八年制临床医学专业规划教材自2005年出版以来,得到了教育部、卫生部等主管部门的认可,以及医学院校广大师生的好评。为了进一步满足教学改革与实践不断推进,以及医学科学不断发展的需要,全国高等医药教材建设研究会和卫生部教材办公室在吴阶平、裘法祖、吴孟超、陈灏珠和刘德培院士等的亲切关怀和支持下于2009年启动了该套教材第二轮的修订工作。

第二轮修订过程中仍坚持“精品战略,质量第一的原则,从精英教育的特点、医学模式的转变、信息社会的发展、国内外教材的对比等角度出发,在注重‘三基’、‘五性’的基础上,从内容到形式都‘更新’、‘更深’、‘更精’,为培养高素质、高水平、富有临床实践和科学创新能力的医学博士服务”的编写宗旨,并根据使用过程中的反馈意见与建议,在第一轮的基础上力求做到:学科体系更加完善,增加了《临床流行病学》、《肿瘤学》、《生物信息学》、《实验动物学》、《医学科学研究导论》和《医学伦理学》;相关学科的交叉与协调更为完善,比如《生物化学》与《医学分子生物学》合并为《生物化学与分子生物学》;内容的选材与框架体系的设计更加注重启发性,强调学生创新能力的培养,并适当给学生留下了思维分析、判断、探索的空间;教材的配套更加健全;装帧设计更为精美。

该套书在修订过程中,得到了广大医学院校的大力支持,作者均来自各学科临床、科研、教学第一线,具有丰富临床、教学、科研和写作经验的优秀专家,作者队伍覆盖了目前国内所有开办临床医学专业八年制及七年制的院校。

修订后的第二版仍以全国高等学校临床医学专业八年制及七年制师生为主要目标读者,并可作为研究生、住院医师等相关人员的参考用书。

全套教材共37种,其中36种于2010年8月出版,1种将于2010年年底出版。

全国高等学校八年制临床医学专业卫生部规划教材 编写委员会

顾问 吴阶平 裘法祖 吴孟超 陈灏珠

主任委员 刘德培

委员 (按姓氏笔画排序)

丰有吉	孔维佳	王卫平	王吉耀	王宇明	王怀经
王明旭	王家良	王鸿利	冯作化	田勇泉	孙贵范
江开达	何 维	吴 江	张永学	张绍祥	李玉林
李甘地	李立明	李 和	李桂源	李 霞	杨世杰
杨宝峰	杨 恬	步 宏	沈 铿	陈孝平	陈 杰
陈 竺	欧阳钦	罗爱静	金征宇	姚 泰	姜乾金
柏树令	赵仲堂	郝希山	秦 川	贾文祥	贾弘禔
高英茂	黄 钢	葛 坚	詹启敏	詹希美	颜 虹
薛辛东	魏于全				

八年制教材目录

*1.《细胞生物学》 第2版(含光盘)	主 编 副主编	杨 恬 左 伋 刘艳平
*2.《系统解剖学》 第2版(含光盘)	主 编 副主编	柏树令 应大君 丁文龙 崔益群
*3.《局部解剖学》 第2版(含光盘)	主 编 副主编	王怀经 张绍祥 张雅芳 胡海涛
*4.《组织学与胚胎学》 第2版(含光盘)	主 编 副主编	高英茂 李 和 李继承 陈晓蓉
*5.《生物化学与分子生物学》 第2版(含光盘)	主 编 副主编	贾弘提 冯作化 屈 伸 药立波 方定志 冯 涛
*6.《生理学》 第2版(含光盘)	主 编 副主编	姚 泰 曹济民 樊小力 王庭槐
*7.《医学微生物学》 第2版(含光盘)	主 编 副主编	贾文祥 陈锦英 江丽芳 黄 敏
*8.《人体寄生虫学》 第2版(含光盘)	主 编 副主编	詹希美 诸欣平 刘佩梅
*9.《医学遗传学》 第2版(含光盘)	主 编 副主编	陈 竺 陆振虞 傅松滨
*10.《医学免疫学》 第2版	主 编 副主编	何 维 曹雪涛 熊思东
*11.《病理学》 第2版(含光盘)	主 编 副主编	陈 杰 李甘地 文继舫 来茂德 孙保存
*12.《病理生理学》 第2版(含光盘)	主 编 副主编	李桂源 吴伟康 欧阳静萍
*13.《药理学》 第2版(含光盘)	主 编 副主编	杨世杰 杨宝峰 颜光美 臧伟进
*14.《临床诊断学》 第2版(含光盘)	主 编 副主编	欧阳钦 吴汉妮 刘成玉
*15.《实验诊断学》 第2版(含光盘)	主 编 副主编	王鸿利 尚 红 王兰兰
*16.《医学影像学》 第2版(含光盘)	主 编 副主编	金征宇 冯敢生 冯晓源
*17.《内科学》 第2版(含光盘)	主 编 副主编	王吉耀 廖二元 黄从新 华 琦
*18.《外科学》 第2版(含光盘)	主 编 副主编	陈孝平 石应康 邱贵兴 杨连粤

*19.《妇产科学》 第2版（含光盘）	主 编 副主编	丰有吉 沈 铿 马 丁 孔北华 李 力
*20.《儿科学》 第2版（含光盘）	主 编 副主编	薛辛东 杜立中 毛 萌
*21.《感染病学》 第2版（含光盘）	主 编 副主编	王宇明 施光峰 宁 琴 李 刚
*22.《神经病学》 第2版（含光盘）	主 编 副主编	吴 江 贾建平 崔丽英
*23.《精神病学》 第2版（含光盘）	主 编 副主编	江开达 于 欣 李凌江 王高华
*24.《眼科学》 第2版（含光盘）	主 编 副主编	葛 坚 赵家良 黎晓新
*25.《耳鼻咽喉头颈外科学》 第2版	主 编 副主编	孔维佳 周 梁 许 庚 王斌全 唐安洲
*26.《核医学》 第2版（含光盘）	主 编 副主编	张永学 黄 钢 匡安仁 李亚明
*27.《预防医学》 第2版（含光盘）	主 编 副主编	孙贵范 凌文华 孙志伟 姚 华
*28.《医学心理学》 第2版（含光盘）	主 编 副主编	姜乾金 马 辛 林大熙 张 宁
29.《医学统计学》 第2版（含光盘）	主 编 副主编	颜 虹 徐勇勇 赵耐青
*30.《循证医学》 第2版（含光盘）	主 编 副主编	王家良 詹思延 许能锋 康德英
*31.《医学文献信息检索》 第2版（含光盘）	主 编 副主编	罗爱静 马 路 于双成
32.《临床流行病学》 （含光盘）	主 编 副主编	李立明 詹思延 谭红专
33.《肿瘤学》	主 编 副主编	郝希山 魏于全 赫 捷 周云峰
34.《生物信息学》 （含光盘）	主 编 副主编	李 霞 李亦学 廖 飞
35.《实验动物学》 （含光盘）	主 编 副主编	秦 川 张连峰 魏 泓 顾为望 王 钊
36.《医学科学研究导论》	主 编 副主编	詹启敏 赵仲堂 刘 佳 刘 强
37.《医学伦理学》 （含光盘）	主 编 副主编	王明旭 尹 梅 严金海

注：全套书均为卫生部“十一五”规划教材，画*者为普通高等教育“十一五”国家级规划教材

八年制教材再版序言

五年来，在大家的热情呵护下，我们共同见证了八年制临床医学教材——这个新生命的诞生与茁壮成长。如今，第二版教材与大家见面，怀纳第一版之精华而不张扬，吞吐众学者之智慧而不狂放，正如医学精英人才所应具备的气质与神韵。在继承中发展，新生才能越发耀眼；切时代之脉搏，思维才能永领潮头。第二版教材已然跨入新的成长阶段，心中唯觉欣喜和慰藉。

回想第一版教材面世之后，得到了各方众多好评，这充分说明了：这套教材将生命科学信息化、网络化以及学科高度交叉、渗透的特点融于一身，同时切合了环境-社会-心理-工程-生物医学模式的转变，诠释了以人为本、协调发展的战略思想。另外，编委构成的权威性和代表性、内容选择、编排体系、印刷装帧质量等，令广大师生耳目一新，爱不释卷。诚然，第一版教材也并非十全十美，比如有的学科仍以介绍知识为主，启发性不强，对学生难以起到点石成金、抛砖引玉的作用，不利于学生创新思维能力的培养；有的学科、章节之间有重复现象，略显冗余，不够干练。另外，随着学科的进展，部分疾病的临床分类、治疗等内容已略显滞后，亟待最新的研究成果加入其中，充实完善。

鉴此，第一版教材的修订工作便提上日程。此次修订，比当初第一版的编纂过程更为艰辛和严谨，从编者的谨慎遴选到教材内容的反复推敲、字斟句酌，可谓精益求精、力臻完美，经过数轮探讨、分析、总结、归纳、整理，第二版教材终于更富于内涵、更具有生命力地与广大师生们见面了。

“精英出精品，精品育精英”是第二版教材在修订之初就一直恪守的理念。主编、副主编与编委们均是各领域内的医学知名专家学者，不仅著作立身，更是德高为范。在教材的编写过程中，他们将从医执教中积累的宝贵经验、体会以及医学精英的特质潜移默化地融入到教材当中。同时，在主编负责制的前提下，主编、副主编负责全书的系统规划，编委会构成团结战斗的团队，各位专家群策群力、扬长补短、集思广益、查漏补缺，为教材的高标准、高质量的修订出版打下了坚实的基础。

注重医学学科内涵的延伸与发展，同时兼顾学科的交叉与融合是第二版教材的一大亮点。此次修订不仅在第一版的基础上增加了《临床流行病学》、《肿瘤学》、《生物信息学》、《实验动物学》、《医学科学研究导论》和《医学伦理学》，同时还合并了《生物化学》与《医学分子生物学》。通过主编顶层设计，相邻学科主编、副主编协调与磋商，互审编写提纲，以及交叉互审稿件等措施，相当程度上实现了突出中心、合理交叉、避免简单重复的要求。

强调启发性以及创新意识、创新思维和创新能力的培养是第二版教材的另一大特色。除了坚持“三基（基础理论、基本知识和基本技能）和五性（思想性、科学性、先进性、启发性和适用性）”，更注重激发学生的思维，让他们成为自己头脑的主人，批判地看待事物，辩证地对待知识，创造性地预见未来。同时，这版教材也特别注重与五年制教材、研究生教材、专科医师培训教材以及参考书的区别与联系。

以吴阶平、裘法祖、吴孟超、陈灏珠为代表的德高望重的老前辈对第二版教材寄予了殷切期望和悉心指导，教育部、卫生部、国家中医药管理局、国家食品药品监督管理局的各位领导的支持是这版教材不断完善动力之源。在这里，衷心感谢所有关心这套教材的人们！正是你们的关注，广大师生手中才会捧上这样一本融贯中西、汇纳百家的精品。

八年制医学教材的第一版是我国医学教育史上的重要创举，相信修订后的第二版将不负我国医学教育改革的使命和重任，为培养高层次的具有综合素质和发展潜能的医药卫生人才做出更大的贡献。诚然，修订过程虽然力求完美，但纰漏与瑕疵在所难免，冀望各位领导、同道及师生不吝赐教，以便于这套教材能够与时俱进，不断完善。

是为序。

中国工程院院士
中国医学科学院院长
北京协和医学院院长

刘德培

于庚寅端午佳节

二〇一〇年六月十六日

生物信息学是 20 世纪 80 年代末伴随着基因组研究而产生的一门新的前沿学科。它在获取、加工、储存、分发海量基因组信息的同时,把基因组 DNA 序列信息分析作为源头,寻找基因组序列中代表蛋白质和 RNA 基因等功能元件的编码序列,并阐明非编码序列的信息实质,破译隐藏在 DNA 序列中的遗传语义规律。

基因组研究的初期,生物信息学在大规模基因组的组装与基因标注上发挥了关键的作用。随着基因组研究的深入,不仅 DNA 序列数据,而且表达序列标签(EST)数据、单核苷酸多态性(SNP)数据、单体型图(HapMap)数据等也大量涌现。进入结构基因组与功能基因组时代后,基因组、蛋白质组、代谢组等组学数据迅速出现。当前,蛋白-蛋白相互作用网络、基因表达调控网络、信号传导网络以及代谢网络的出现与发展,使得生物信息学进入了系统生物学的时代。如果说生物信息学发展的初期面对的主要是序列数据,那么,随着基因组领域研究在质与量上的提高,这一学科面对的数据在类型与本质上都极大地丰富了。为了处理这些数据,生物信息学得到了蓬勃的发展,成为了基因组研究不可或缺的工具。生物信息学自身也成为了一门实用学科。

可以相信,在第二、第三代高通量测序技术的推动下,基因组及相关的成果将很快进入国民经济及人类健康的很多领域,也会直接走到每个人的面前。那时候,生物信息学的很多知识会成为人们生产、生活的基本知识,生物信息学自身也将成为一门基础学科。

在生物信息学不断普及的背景下,李霞和李亦学两位教授组织编写了这本教科书,为生物信息学的教育提供了很好的教材。

本书的特点是:全面地提供了数据库资源,序列对比,表达谱分析,分子进化,芯片数据处理,生物网络,蛋白质结构以及药物设计等几乎生物信息学涉及的所有方面。为了使读者能更好地检验自己掌握程度,每章末还附了习题。由于两位教授长期在生物信息学第一线从事研究与教学,书中也融汇了他们大量的体会与经验。相信这本书能为使用者带来切实的帮助。

陈润生

2010 年 4 月 16 日

《生物信息学》是在国内资深生物学家与医学专家的倡导下,经全国高等医药教材建设研究会、卫生部教材办公室组织有关专家反复论证后决定组织编写的。在教材的具体筹备和编写过程中,全国高等医药教材建设研究会、卫生部教材办公室先后组织三次会议进行研讨。本书正是根据这些会议精神及专家的指导意见组织全国十余所大学的一线教师和学者共同努力编写而成的。

“21 世纪是生命科学的世纪,也是信息科学的世纪”。伴随着人类基因组计划及其他模式生物基因组计划的全面实施,分子生物数据正在呈爆炸性增长。及时、充分、有效地利用不断增长的生物信息资源进行分析和探索,已经成为生物医学领域研究与应用的必备方法。为适应现代生物医学发展和素质教育的需要,我们在编写《生物信息学》教材过程中,突出“三基”(基本知识、基本理论、基本技能),强调五性(思想性、科学性、先进性、启发性、适用性),并力求在内容和形式上有所创新。

全书共分为三篇十六章。第一篇生物信息学基础,含 DNA、RNA 和蛋白质序列信息资源、双序列比对、多序列比对、序列特征分析、分子进化分析、表达序列分析和基因芯片数据分析七章,均系生物医学相关领域发展过程中形成的基础生物信息数据及分析方法;第二篇功能基因组信息学,含基因功能注释、蛋白质分析与蛋白质组学、蛋白质结构分析、转录调控的信息学分析、生物分子网络和计算表观遗传学六章,均系功能基因组研究中颇具特色的生物信息学方法;第三篇生物信息学与人类复杂疾病,含人类复杂疾病与计算系统生物学、单核苷酸多态与人类疾病和 miRNA 与复杂疾病三章,以及光盘附录内容药物生物信息学,均系近几年发展起来的与复杂疾病有关的重要生物信息学方法。

本书各章节相对独立,每一章都反映了生物信息学某一方向的最新成果与发展趋势。为适应不同读者群的需要,各章的布局是统一的。第一节是引言,以简明易懂的语言介绍该章的主要内容,包括能解决什么问题和解决问题的思路;后面各节介绍基本概念和常用生物信息学方法,着重于生物医学实际应用、操作方法和生物医学意义的解释;小结与主要参考文献放在各章最后。

本书的读者面十分广泛。不论是生物医学领域的学生、教师、研究人员还是生物信息学专业人员,都可以各取所需、各有所获。生物信息学思想和技术是医学研究的有效工具,医学研究者阅读本书,不难知道有哪些现代生物信息学方法可以为之所用、基本思路如何、需要怎样的设计和数据,应用的结果如何解释等;生物信息学专业人员阅读本书,不但可以深入地掌握生物信息学的最新研究成果与未来发展方向,而且还有助于提升研究工作的水平。

本书各章作者都是相关研究方向的专家,每一章都凝聚了他们独特的学术思想、研究心得和研究成果。他们在百忙之中精心组织素材,字斟句酌地编写,付出了大量心血。在此我们对全体编委的无私奉献深表谢意!同时,哈尔滨医科大学生物信息科学与技术学院的老师和研究生们也做了大量的协助工作,在此一并致谢!

本教材得到国家高科技“863”项目和哈尔滨医科大学“211”工程重点学科建设经费的资助,特此鸣谢!

在本教材编写过程中,尽管我们努力跟踪学科的新发展、新技术,并尽力把它们纳入到教材中来,以保持本书的先进性和实用性,但由于时间紧迫、能力有限,直至完稿,仍觉有许多不足之处,希望学术同仁不吝赐教,以便再版时改正。

李霞 李亦学

2010年4月20日

目 录

绪论

I

INTRODUCTION TO BIOINFORMATICS

第一节 生物信息学的兴起	1
Section 1 The Rise of Bioinformatics	
一、人类基因组计划	2
二、生物信息学与组学	3
第二节 生物信息学在生命科学中的地位及意义	4
Section 2 The Significance of Bioinformatics in Life Science	
一、生物信息学内涵	4
二、生物信息学在现代生物医学发展中起着重要作用	6

第一篇 生物信息学基础

第一章 DNA、RNA 和蛋白质序列信息资源

9

CHAPTER 1 DNA、RNA AND PROTEIN SEQUENCE INFORMATION RESOURCES

第一节 引言	9
Section 1 Introduction	
第二节 核酸序列数据库	9
Section 2 Nucleic Acid Sequence Databases	
一、GenBank 数据库	10
二、EMBL 数据库	14
三、DDBJ 数据库	15
四、其他数据库	15
第三节 蛋白质序列数据库	16
Section 3 Protein Sequence Database	
一、PIR 数据库	16
二、MIPS 数据库	18
三、其他数据库	18
第四节 NCBI 与 EMBL-EBI	19
Section 4 NCBI and EMBL-EBI	
一、NCBI 简介	19

二、EMBL-EBI 简介	22
三、通过 Entrez Gene 从 NCBI 获取序列信息	24
四、通过 SRS 从 EBI 中获取蛋白质序列信息	32
小结	36

第二章 双序列比对

38

CHAPTER 2 PAIRWISE SEQUENCE ALIGNMENT

第一节 引言	38
Section 1 Introduction	
一、同源、相似与相同	38
二、相似性的定量描述	38
三、空格	40
第二节 替换记分矩阵	40
Section 2 Scoring Matrix	
一、通过点矩阵对序列比较进行记分	40
二、DNA 序列比对的替换计分矩阵	41
三、蛋白质序列比对的替换记分矩阵	41
第三节 双序列比对算法	45
Section 3 Algorithms of Pairwise Sequence Alignment	
一、全局比对的经典算法	46
二、局部比对的经典算法	48
第四节 数据库搜索	48
Section 4 Database Search	
一、BLAST	49
二、数据库搜索实例	50
第五节 比对的统计学显著性	52
Section 5 Statistical Significance of Pairwise Alignment	
一、全局比对的统计学显著性	53
二、数据库搜索的统计学显著性	53
第六节 参数的选择	54
Section 6 Selecting Scoring Parameters	
一、空格罚分参数	54
二、BLAST 的参数	55
三、如何处理太多与太少的数据库搜索返回	55
小结	56

第三章 多序列比对

58

CHAPTER 3 MULTIPLE SEQUENCE ALIGNMENT

第一节 引言	58
Section 1 Introduction	
一、多序列比对具有广泛的应用	58
二、多序列比对存在多种种类	59
第二节 相似性与距离、计分与罚分、替换矩阵	60
Section 2 Similarity and Distance, Scoring Matrix and Substitution Matrix	
一、相似性与距离是序列相似性的两个主要度量	60
二、存在多种方法对比对进行计分与罚分	61
三、精确计算失配计分需要使用核苷酸和氨基酸替换矩阵	62
四、记分方法可显著影响多序列对比	62
五、多序列对比的困难性	62
第三节 主要比对方法与软件	63
Section 3 Methods and Softwares of Multiple Alignment	
一、动态规划法	63
二、渐进多序列比对	65
三、迭代法	67
四、基于一致性的方法	69
五、多序列比对结果编辑器	70
第四节 局部比对、glocal 比对和 syntenic 比对	71
Section 4 Local, Glocal and Syntenic Alignment	
一、局部比对	71
二、glocal 比对	71
三、syntenic 比对	72
第五节 全基因组比对	74
Section 5 Whole Genomic Alignment	
一、全基因组多序列比对	74
二、UCSC 基因组浏览器	74
三、其他方法与软件	76
第六节 软件、参数和比对质量	77
Section 6 Softwares, Parameters and Alignment Quality	
一、软件的选择	77
二、计分等参数的选择	79
三、控制比对质量	79
四、注意事项	80
小结	80

第四章 序列特征分析

85

CHAPTER 4 ANALYSIS OF SEQUENCE CHARACTERISTICS

第一节 引言	85
Section 1 Introduction	
第二节 DNA 序列特征分析	87
Section 2 Analysis of DNA Sequence Characteristics	
一、利用 GENSCAN 识别基因开放阅读框	87
二、利用 POLYAH 预测分析转录终止信号	89
三、利用 PromoterScan 预测分析启动子区域	90
四、利用 CodonW 分析密码子偏好性	91
第三节 蛋白质序列特征分析	94
Section 3 Analysis of Protein Sequence Characteristics	
一、利用 ProtParam 分析蛋白质的理化性质	94
二、利用 ProtScale 分析蛋白质的亲水或疏水性	96
三、利用 Tmpred 分析蛋白质的跨膜区	98
四、蛋白质序列分析软件包 AntheProt	100
第四节 序列综合分析	103
Section 4 Sequence Analysis Software	
一、EMBOSS 软件包	103
二、DNASTar 软件包	104
三、Omega 2.0 软件包	105
四、Vector NTI 软件包	105
小结	106

第五章 分子进化分析

109

CHAPTER 5 MOLECULAR EVOLUTION ANALYSIS

第一节 引言	109
Section 1 Introduction	
第二节 系统发生分析与重建	109
Section 2 Phylogeny Reconstruction	
一、核苷酸置换模型及氨基酸置换模型	109
二、系统发生树的基本概念及搜索方法	114
三、分子钟假说	117
第三节 核苷酸和蛋白质的适应性进化	118
Section 3 Adaptive Evolutions of Nucleotide and Protein	
一、中性与近中性理论	118
二、基因适应性进化的统计学检验方法	119
三、 d_N 或 d_S 检验	121
四、适应性进化基因	123

第四节 分子进化与生物信息学	124
Section 4 Molecular Evolution and Bioinformatics	
一、基因组进化概述	124
二、病毒基因组分析	124
三、原核生物基因组比较	126
四、蛋白质互作网络进化	128
五、代谢网络进化分析	130
小结	132

第六章 表达序列分析 134

CHAPTER 6 ANALYSIS OF EXPRESSED SEQUENCES

第一节 引言	134
Section 1 Introduction	
第二节 EST 数据分析	135
Section 2 Analysis of EST data	
一、cDNA 文库构建与 EST 数据的实验获取	135
二、EST 数据库	136
三、EST 数据分析方法	146
第三节 基因表达系列分析	157
Section 3 Serial Analysis of Gene Expression	
一、SAGE 技术原理简介	157
二、SAGE 技术方案简介	159
三、SAGE 技术的缺陷与改进	160
四、SAGE 技术的应用前景	161
五、SAGE 数据库和分析软件	162
小结	169

第七章 基因芯片数据分析 172

CHAPTER 7 MICROARRAY DATA ANALYSIS

第一节 引言	172
Section 1 Introduction	
第二节 常见的芯片平台与数据库	172
Section 2 General Microarray Platform and Database	
一、cDNA 微阵列芯片	173
二、寡核苷酸芯片	174
三、原位合成芯片	174
四、光纤微珠芯片	176
五、基因表达数据库	177
六、斯坦福微阵列数据库	177