



国家科学技术学术著作出版基金资助出版

生物信息学

计算技术和软件导论

马占山 等 编著

Bioinformatics Computing and Software Zhanshan (Sam) Ma

非
外
借

科学出版社

国家科学技术学术著作出版基金资助出版

生物信息学

计算技术和软件导论

马占山 等 编著



科学出版社

北京

内 容 简 介

如果说 21 世纪是生物学世纪,生物信息学应该是支撑生物学世纪的核心科技之一。而大数据科学和人工智能技术正在将生物信息学推向生命科学和信息科学的前沿。本书分为生物信息学基础篇和生物信息组学技术篇两大部分。生物信息学基础篇从新兴领域切入,介绍生物信息学的计算科学及进化生物学基础(如网络科学与大数据技术、深度学习、计算智能、高维数据分析、马尔可夫链蒙特卡洛法,隐马尔可夫模型,贝叶斯统计、医学生态学、DNA 计算、进化树与溯祖树分析、种群遗传学等)。生物信息组学技术篇除经典内容(基因组、转录组、蛋白质组)外,还包括最新的三代基因测序算法和软件(作者团队研发的 DBG2OLC 和 SPARC)、微生物群系(Microbiome)和宏基因组学(Metagenomics)、非编码 RNA、新药发现、代谢组学(Metabolomics)等热点内容。

本书可作为高等院校生物信息学教材或参考书,也可供相关领域(生物学、生态学、医学、药学、计算科学、农林、食品科学等)的科技工作者阅读参考。

图书在版编目(CIP)数据

生物信息学计算技术和软件导论 / 马占山等编著. —北京:科学出版社, 2017.11

ISBN 978-7-03-042639-0

I. ①生… II. ①马… III. ①生物信息论-计算技术 ②生物信息论-应用软件 IV. ①Q811.4

中国版本图书馆 CIP 数据核字 (2014) 第 279222 号

责任编辑:张展 华宗琪 / 责任校对:葛茂香 侯彩霞

封面设计:墨创文化 / 责任印制:罗 科

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>

成都锦瑞印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2017年11月第 一 版 开本:787×1092 1/16

2017年11月第一次印刷 印张:20 3/4

字数:490千字

定价:137.00元

(如有印装质量问题,我社负责调换)

编著者(团队)及其学术机构^①

Contributors and Affiliations

马占山 叶承羲 李连伟 夏 尧 董 萍 关 琼

中国科学院昆明动物研究所

遗传资源与进化国家重点实验室

计算生物与医学生态学科组

Zhanshan (Sam) Ma, Chengxi Ye, Lianwei Li, Yao Xia, Ping Dong & Qiong Guan

Computational Biology and Medical Ecology Lab

State Key Lab of Genetic Resources and Evolution

Kunming Institute of Zoology

Chinese Academy of Sciences

朱天琪

中国科学院数学与系统科学研究院

应用数学研究所

Tianqi Zhu

Institute of Applied Mathematics

Academy of Mathematics and Systems Sciences

Chinese Academy of Sciences

李海鹏 高 峰 明 晨

中国科学院上海生命科学研究院

中国科学院-马普学会计算生物学伙伴研究所

Haipeng Li, Feng Gao & Chen Ming

CAS-MPG Partner Institute for Computational Biology

Shanghai Institutes for Biological Sciences

Chinese Academy of Sciences

吴东东 邵 永 徐海波 叶凌群

中国科学院昆明动物研究所

① 按章节出现顺序排列, 国内机构排前; 作者名单中第一位(一般)为通讯作者。

遗传资源与进化国家重点实验室
进化与发育转录组学科组
Dongdong Wu, Yong Shao, Haibo Xu & Lingqun Ye
Evolutionary and Developmental Transcriptome Lab
State Key Lab of Genetic Resources and Evolution
Kunming Institute of Zoology
Chinese Academy of Sciences

刘长宁 李 菁 陈 雯 和桃梅
中国科学院西双版纳热带植物园
中国科学院热带植物资源可持续利用重点实验室
生物信息研究组
Changning Liu, Jing Li, Wen Chen & Taomei He
Bioinformatics Lab
Key Lab of Tropical Plant Resources and Sustainable Use
Xishuangbanna Tropical Botanical Garden
Chinese Academy of Sciences

李 慧
中国科学院沈阳应用生态研究所
中国科学院森林生态与管理重点实验室
Hui Li
CAS Key Laboratory of Forest Ecology and Management
Institute of Applied Ecology
Chinese Academy of Sciences

赵勇山 宋勇波
沈阳药科大学
生命科学与生物制药学院
Yongshan Zhao & Yongbo Song
School of Life Sciences and Biological Pharmacy
Shenyang Pharmaceutical University

王 健
沈阳药科大学
制药工程学院
Jian Wang
School of Pharmaceutical Engineering
Shenyang Pharmaceutical University

Chengxi Ye
Department of Computer Science
University of Maryland
College Park, MD 20742
USA

David S Wishart
Professor & Director
National Institute for Nanotechnology (NINT)
Canadian National Research Council (NRC)
Departments of Computing Science & Biological Sciences
University of Alberta, Edmonton
Canada

Ziheng Yang, FRS
R. A. Fisher Professor of Statistical Genetics
Department of Genetics, Evolution and Environment
University College London
London WC1E 6BT
England, UK

序

生物信息学是生命科学与信息(计算)科学的交叉学科,现已成为现代生命科学和生物技术领域的基础学科之一,大数据时代的来临更是将其推向了现代科技的前沿。在国内外生物信息学迅猛发展的今天,《生物信息学计算技术和软件导论》一书的出版,对国内生物信息学的科研和教学具有重要的现实意义。该书主要从计算技术和软件的视角综述了生物信息学中计算、遗传和统计等基础领域的一些最新重要进展(基础篇);同时较全面系统地介绍了各种组学技术所涉及的生物信息数据分析和建模技术(技术篇)。通过该书,读者可了解生物信息学和计算生物学领域的前沿科技。编著者来自中国科学院五个相关研究所(昆明动物研究所、数学与系统科学研究院、上海计算生物学研究所、西双版纳植物园和沈阳应用生态研究所)、沈阳药科大学,以及多家国外著名大学和科研机构(伦敦大学学院、阿尔伯塔大学、加拿大国家科学研究委员会、加拿大国家纳米技术研究所和马里兰大学),在生物信息学和计算生物学领域具有较高的造诣。主编马占山研究员在美国爱达荷大学获得计算机科学和昆虫学双博士,并具有在美国硅谷近十年的计算机高级工程师经历,是中国科学院遗传资源与进化国家重点实验室于2010年从美国爱达荷大学通过“百人计划”引进的PI(Principal Investigator)。引进后在中科院昆明动物研究所建立了“计算生物与医学生态学实验室”,并在三代基因测序软件、人类微生物群系宏基因医学生态学等领域取得了一系列重要突破。

跨学科障碍或许是当今生物信息学领域面临的最大的挑战之一。传统上,生命科学多以实验和归纳推理为主,而计算和数理科学则以理论和演绎推理为主。该书基础篇虽侧重于介绍计算和数理科学领域的前沿发展,但想必也能够为生物学家理解和掌握。而技术篇不仅为生物学家详细介绍了组学中重要的生物信息学分析技术和相关的资源(软件及数据库),同时也为计算和数理学提供了深入了解生物学研究的恰当切入点。该书特别值得一提的是对国内外在生物信息学专业课程设置的比较分析和建议。毕竟,专业人才培养应该是减缓并最终消除跨学科障碍的大计!

20世纪末叶展开的信息革命为开启21世纪生命科学世纪奠定了良好的技术基础。生命的进化史和未来的延续本质上是信息的遗传、变异和进化。生物信息学的重要性就在于它能够对人类认识、保护和永续利用地球上的生物资源,以及为自身的保健和疾病诊治提供有效的信息处理理论、技术和工具。本人愿《生物信息学计算技术和软件导论》一书的出版发行能够为推进中国生物信息学和计算生物学的发展有所贡献!

张亚平

2017年6月于北京

前 言

作为拙作《生物信息学计算技术和软件导论》的主编，笔者必须面对“什么是生物信息学”这一看似简单的问题。而要对此给出一个令人满意的答案也是一项挑战！除自身学疏才浅之外，笔者所感受到的压力主要源于三点：其一，生物信息学的跨学科、综合性等特点使得笔者不能简单地从字面意义去定义生物信息学。其二，生物信息学领域尚未建立起完整的理论体系。据笔者拙见，从内容上看，现阶段的生物信息学主要是一些用于分析组学（特别是基因组）数据的算法和软件技术的集合，这些算法和技术源于计算机科学、数学及统计学，它们与用于分析其他科学数据（如天文学、物理学、化学等）的计算技术并无本质差别，均属于计算科学（computational science），但生物信息分析所获得的信息、知识，乃至它们的理论升华则应该回归到生物科学领域的问题上来，如肿瘤基因组学、分子育种、新药研制、医学生态学等。而这些偏向应用性的内容是否属于生物信息学似乎尚无定论。其三，生物信息学与其姊妹学科（特别是计算生物学和系统生物学）之间存在“竞争”；同时，它们之间，乃至与各种组学技术之间的学科界限常常模糊不清。

幸运的是，生物信息学的先驱们已经发表了诸多可供我们参考的文章和论述，尽管他们之间似乎尚未对生物信息学的内涵达成共识。此处笔者选择简单地介绍荷兰理论生物学家 Paulien Hogeweg 和 Ben Hesperd 所提出的概念。之所以选择介绍他们关于生物信息学的论述，考量在于：据 2002 年版的《牛津英语词典》对“Bioinformatics”（生物信息学）的解释，“生物信息学”一词最早出现在 1978 年 Paulien Hogeweg 和 Ben Hesperd 发表的论文中。其实他们早在 1970 年就建议从荷兰语“Bioinformatica”一词衍生出英语“Bioinformatics”一词，即生物信息学。不过，生物信息学的概念最初是用于泛指“生物系统的信息处理过程”（the study of informatic processes in biotic systems）。Hogeweg 和 Hesperd 的观点是：生命最重要的特征之一就是不同形式表现出来的信息处理（information processing）功能，例如遗传信息的进化、转录、翻译等。显然，用“信息处理”来比喻生命系统对于我们理解生命系统的运作亦十分有益。生物化学、生物物理、生物数学、生物信息等生命科学分支学科之间应该具备各自相对独立的研究范畴。正是基于这些考量，Hogeweg 和 Hesperd 于 1970 年便提出了“生物信息学”这一概念。显然，他们的思想在 1970 年代是非常超前了！

然而，在许多生物信息学的文献中，Hogeweg 和 Hesperd 关于“生物信息学”的概念并没有受到多少关注。他们的概念表面上似乎与目前生物信息学的主流，即研究统计学和计算机科学在分子生物学中的综合应用（如研发用于测序数据、蛋白质结构解析、系统进化发育等数据管理和分析建模的计算方法和软件）并非吻合，甚至有些读者可能会认为 Hogeweg 和 Hesperd 的生物信息学概念，以及他们的理论生物学研究背景并不能反映现今分子测序时代的生物信息学，而仅仅是用词上的巧合。但事实上，Hogeweg 教授不仅于

1977年在荷兰乌得勒支大学创建了世界上首家理论生物学与生物信息学(Theoretical Biology and Bioinformatics)实验室,并且在1984年开发出了一套用于DNA多重序列比对的算法,这表明,我们今天的生物信息学确实是与Hogeweg和Hesperd近半世纪前创立的生物信息学概念一脉相承。另外,Hogeweg和Hesperd于2011年在“The Roots of Bioinformatics in Theoretical Biology”一文中,将1970年代以“信息处理”为核心的生物信息学研究比喻为“生物信息学的树根”(roots of bioinformatics),而将由高通量测序技术催生的、由大数据驱动的、关于功能和进化的研究比作“(主)树干”(trunk of bioinformatics)。因此,笔者在此斗胆建议将由生物信息学所支撑的一些应用性研究领域(如由基因组学为基础的癌症基因组学和分子诊断、由人体菌群宏基因组学为基础的医学生态学、新药设计、分子育种、生物合成等技术)归为生物信息学“树”模型的“枝叶”(branches & leaves)。所以,在本书的构思和写作过程中,我们试图以生物信息学“树”模型作为框架,兼顾根、干和枝叶。

本书分为基础篇(第1~4章)和技术篇(第5~12章)两部分。前者旨在培植生物信息学“树”的“根”,后者旨在培育生物信息学“树”的“主干”。第1章选择性地介绍目前生物信息学中的一些热点议题,包括:生物信息大数据、网络分析、人工智能(深度学习)与计算智能、复杂性科学、医学生态学和DNA计算等。限于篇幅,该章省略了传统生物信息学著作中通常包括的技术性议题,如编程基础、数据库技术、高性能计算与云计算。虽然这些在本书中所省略的内容是生物信息学的重要基础,但读者应该能够参考许多已出版的生物信息学和计算生物学专著,也可寻求计算机科学或IT领域的专业帮助。与此同时,第1章中的“医学生态学”是一探索性议题,旨在起到抛砖引玉的作用;而“DNA计算”可被视为生物信息学对计算科学的回馈。

第2、3章旨在建立生物信息学与进化论之间的桥梁。正如进化生物学家Theodosius Dobzhansky近半个多世纪前的名言金句“Nothing in biology makes sense except in the light of evolution”所预测的那样,生物信息学自然也不会是例外。或许可以认为,计算机科学和数学(特别是概率统计)是生物信息最重要的两大数理科学基础。第4章旨在建立生物信息的另一根基。该章同样省略了传统生物信息或生物统计学专著中普遍涵盖的一些内容(如生物信息学中最为广泛使用的统计编程建模工具R、多元分析等重要内容),将重点放在其他专著介绍较少的方法,包括马尔可夫蒙特卡洛法、隐马尔可夫模型、贝叶斯统计、统计学习和高斯图模型(Gaussian graphical model)。特别是高斯图模型(又名协方差选择模型)适用于变量数大于样本量的情形。由于基因数量(数以万计)一般都超过样本数量,因此,高斯图模型在生物信息学中的重要性是显而易见的。

第5章介绍最新的第三代基因测序(单分子测序技术)组装算法和软件,着重于作者在该领域开发的两款软件:DBG2OLC和SPARC。

第6、11章涉及大量软件、工具和资源。相关软件和资源的详细介绍已编辑为在线附件材料,有兴趣的读者可以通过e-mail(ma@vandals.uidaho.edu)免费获取。其中一些重要软件和资源的介绍,会不定期更新。

第7章介绍RNA-Seq技术的原理、应用以及数据的生物信息学分析技术(包括基于参考基因组的转录组分析、无参考基因组的转录组的从头拼装,以及差异表达基因分析等)。

第8章全面且深入介绍非编码RNA研究常用数据库及软件领域的主要资源。

第9章的主要内容包括：计算蛋白质组学及其应用，以及计算蛋白质组学算法与数据库两大方面。

第10章“新药物发现生物信息学分析”严格讲属于生物信息学应用领域、或者可以认为是生物信息学的“枝叶”议题。

第11章“宏基因组学概述及生物信息学分析”讨论目前生物医学和环境微生物学中最活跃的热点领域——微生物群系(Microbiome)。

第12章“Bioinformatics metabolomics: an introduction”系技术篇的最后一章，也是本书所涉及的“树干”最后一段。关于代谢组一章，笔者借此前言向读者介绍该章作者，加拿大国家纳米研究中心、阿尔伯塔大学 David Wishart 教授关于代谢组应用的一个观点：代谢组在个性化精准医学中具有特殊的重要地位，在实践中若要将所有代谢产物的基因调控和表达研究清楚，可能会太费周折，甚至难以实现；而直接从代谢产物入手则可能诊断许多疾病。换句话讲，“基因组(宏基因组)—转录组(宏转录组)—蛋白组—代谢组学”的一条龙研究或许有“捷径”可走。这一观察在涉及宏基因组(宏转录组)时可能更具有现实意义。这一观察其实也与现代医学实践相吻合，人类代谢组学研究或许可以看作是“医学检验(化验)”在新时代的扩展。David Wishart 教授领导了人类代谢组项目(Human Metabolome Project; 2006 - 2009)的完成，并创建了人类代谢组数据库(<http://www.hmdb.ca/>)，其观点应该具有一定的权威性。

严格来讲，“新药物发现”和“医学生态学”的内容归入应用篇应该更为妥当。但我们只收集到“基础”和“技术”两类题材，故将前两者并入基础篇和技术篇。原因是医学生态学最重要的研究方法和技术：复杂网络和理论生态学分析技术应属于第1章，医学生态学正好可以作为该技术的应用实例来介绍。如果未来有机会更新此书，我们会力争开辟应用篇，在其中加入肿瘤基因组学、分子育种、生物合成及个性化精准医疗等内容，从而为生物信息学这棵大树增添“枝叶”。需要指出的是，虽然我们完全赞同 Hogeweg 和 Hesperd 在 2011 年提出的关于生物信息学应植“根”于理论生物学、计算科学、复杂性科学等领域的沃土中这一理念，并对此作出了相应的努力，但我们并不宣称我们的劳动已经“勾画”出了一棵“根深”、“叶茂”的大树。其原因之一是：生物信息学理论体系仍然处在早期发展阶段，而要扎根则必需有肥沃的土壤基础；而更深层次的原因或许是，目前生物信息学研究主要围绕大数据分析，而忽略了对基础理论的研究。大数据固然重要，但大数据并非生物信息学全部。或许我们尚未完全认识到生物信息学的“根”(即理论基础)的重要性，而将生物信息学作为一门纯技术学科。事实上，目前许多生物信息学教科书和专著(包括本书)都缺乏对“根”的深度讨论，我们期盼本书的讨论能够起到抛砖引玉的作用！而关于“枝叶”，即生物信息学的应用研究，尽管发展迅猛，但仍主要停留在对生物信息学数据分析结果的解释上，缺乏系统化、一般化的知识积累，因而较难将其凝练为专著或教科书。加之凝练的过程需要对应用领域本身(如人体微生物群系(human microbiome)宏基因组医学生态学、新药物设计)有综合的了解，而且不同应用领域生物信息学所要解决问题的性质也可能各异，这也使我们没有能够在本书中设立专门的应用篇。

追溯历史，20 世纪后半叶开启的分子生物学革命和计算机信息革命直接催生了生物信息学。人类基因组计划的完成在依赖于生物信息学分析的同时也极大地促进了生物信息学的发展。时至今日，生物信息学的应用范围早已扩展至生物学、生态学、脑科学、信息

科学、医学(包括法医鉴定)、药学、食品科学、农学、林学等多个领域。生物信息学属于交叉学科,理想情况下,优秀的研究者应同时具备扎实的数学统计、计算机科学和生物学等多方面的背景知识。然而,一方面,目前绝大多数生物信息学工作者系由生物学家“转业”,他们可能对数学和计算科学的背景了解较少;另一方面,即使生物信息学家同时具备了多学科背景,组学技术的快速发展和其应用领域的迅猛扩展所产生的、与日俱增的海量数据也会不断地给生物信息学家提出大量亟待解决的难题。当今生物信息学中的一些难题确已成为现代生物学、生物医学、生物工程等领域发展的瓶颈。本书围绕生物信息学的学科“树”模型,较为全面综合地介绍“树根”(计算科学基础)、“树干”(-omics 大数据分析技术和软件)和部分“枝叶”(生物信息学应用)的“生长发育”现状,不仅可作为高等院校生物信息学课程的教材或参考书,而且对相关领域的科技工作者也应该具有重要参考价值。例如,对计算生物信息学感兴趣的计算机科学家、生物学家、医学家、农学家、林学家等均可不同程度地从本书中找到有用的生物学数据分析技术和方法,从而实现具体的生物学数据分析。

21 世纪是生命科学快速发展的时代,生命科学早已渗透到人们生活的方方面面。据笔者拙见,如果说 21 世纪是生物学的世纪,生物信息学应该是支撑生物学世纪的关键技术。而生物信息学面临的最大挑战可能并不仅仅来自于大数据,简单来讲,大数据可能只是表象,而算法(基础)和机制(应用问题,例如医学生态学机理)的研究可能更具有挑战性!换言之,生物信息学这棵大树必须“根深”、“叶茂”才能有效地为不断生长的“树干”提供养分。从更宽广的视角来看,生物信息学应该不断从理论生物学汲取营养,甚至与理论生物学融合。如此发展,未来生物学可能会像物理学一样,而成就理论生物学与实验生物学的二分世界。来时,理论生物学也会像理论物理学一样,成为自然科学的明珠!

最后,笔者在此对本书各位合作者所付出的辛勤努力表达由衷谢意!并感谢国家科学技术著作出版基金提供资助,以及关琮同学为成功申请该基金、李文迪同学为校稿所做的努力!真诚感谢华宗琪编辑的耐心帮助!衷心感谢张亚平院士的指导和鼓励、并为本书作序!

马占山

2017 年仲夏于昆明

Bioinformatics
Computing and Software
Zhanshan (Sam) Ma *Editor*
Chinese Academy of Sciences

Table of Contents

Preface

Yaping Zhang

Foreword

Zhanshan (Sam) Ma

Part I. The Foundations of Bioinformatics

Chapter 1. Recent advances in computational bioinformatics

Zhanshan (Sam) Ma

Chapter 2. Phylogenetic and coalescent analyses

Tianqi Zhu & Ziheng Yang

Chapter 3. Population genetics: an introduction on software tools

Feng Gao, Chen Ming & Haipeng Li

Chapter 4. Statistical computing for bioinformatics

Zhanshan (Sam) Ma, Ping Dong & Qiong Guan

Part II. The Omics Technologies

Chapter 5. Genome assembly algorithms and software for the 3rdGS technologies

Zhanshan (Sam) Ma & Chengxi Ye

Chapter 6. Genomic data analysis for the NGS technologies

Lianwei Li & Zhanshan (Sam) Ma

Chapter 7. Transcriptomic data analysis

Yong Shao, Haibo Xu, Lingqun Ye & Dongdong Wu

Chapter 8. Software and database resources for non-coding RNA research

Jing Li, Wen Chen, Taomei He & Changning Liu

Chapter 9. Software and database resources for proteomic analysis

Yongshan Zhao & Hui Li

Chapter 10. Software and database resources for drug discovery

Jian Wang & Yongbo Song

Chapter 11. Bioinformatics software pipelines for metagenomics

Yao Xia & Zhanshan (Sam) Ma

Chapter 12. Bioinformatics for metabolomics—an introduction

David S Wishart

目 录

生物信息学基础篇

第1章 生物信息学一些前沿领域简介	3
1.1 生物信息大数据	3
1.2 复杂网络分析概论	11
1.3 复杂网络分析实例：以微生物群系医学生态网络为例	15
1.4 深度学习、计算智能与人工智能	21
1.5 医学生态学	25
1.6 DNA 计算机—生物学对计算机科学的回馈	30
第2章 系统发育树与溯祖分析	38
2.1 树的概念	38
2.2 主要的建树方法	39
2.3 模型选择	50
2.4 贝叶斯方法	54
2.5 溯祖理论	60
2.6 物种树估计	64
第3章 群体遗传学数据分析软件简介	70
3.1 多功能软件比较	70
3.2 理论模型与分析方法的实现方式	72
3.3 软件运行方式与编程语言	79
3.4 总结与展望	79
第4章 生物信息学中重要统计计算方法和模型	85
4.1 计算机模拟技术	85
4.2 马尔可夫蒙特卡罗法	93
4.3 隐马尔可夫模型	98
4.4 贝叶斯统计	105
4.5 统计学习	114
4.6 高斯图模型	120

生物信息组学技术篇

第5章 第三代基因测序组装算法和软件技术	129
5.1 第三代基因测序及组装技术简介	129
5.2 第三代基因组组装算法及软件简介：以 DBG2OLC 和 SPARC 为例	132
5.3 三代基因组组装算法和软件比较	139
5.4 DBG2OLC 和 SPARC 软件使用简介	140
第6章 基因组第二代测序数据的生物信息学分析	145
6.1 基因测序技术简介	145
6.2 基因组组装技术	149
6.3 外显子基因突变检测	154
6.4 单细胞测序数据的基因组组装	156
第7章 转录组数据的生物信息学分析	160
7.1 转录组技术的发展	160
7.2 RNA-seq 数据的质量控制	163
7.3 基于参考基因组的转录组分析	164
7.4 无参考基因组的转录组的从头拼装及拼装质量评估	170
第8章 非编码 RNA 研究常用数据库及软件	175
8.1 非编码 RNA 概述	175
8.2 非编码 RNA 常用数据库	179
8.3 非编码 RNA 研究常用软件	184
第9章 蛋白质组学研究常用软件简介	210
9.1 蛋白质组学简介	210
9.2 计算蛋白质组学的应用	215
9.3 计算蛋白质组学算法与数据库	230
第10章 新药物发现中的生物信息学软件简介	236
10.1 大型药物设计平台	237
10.2 分子视图软件	238
10.3 化学结构编辑程序	242
10.4 分子对接与虚拟筛选软件	245
10.5 配体构象搜索软件	250
10.6 药效团模拟软件	251
10.7 分子动力学模拟软件	254
10.8 在线药物设计资源列表	255
10.9 小结	257
第11章 宏基因组学概述及生物信息学分析	260
11.1 宏基因组学技术简介	260
11.2 宏基因组学研究流程	261

11.3	宏基因测序数据的生物信息学分析	263
Chapter 12	Bioinformatics for Metabolomics: An Introduction	277
	Abstract	277
12.1	Introduction to Metabolomics	277
12.2	Technologies for Metabolomics	280
12.3	Data Formats for Metabolomics	285
12.4	Databases for Metabolomics	287
12.5	General Principles for Metabolomic Data Analysis	292
12.6	From Spectra to Metabolite Lists; Bioinformatics for Metabolite Identification	293
12.7	From Metabolite Lists to Significant Metabolites; Multivariate Statistics	300
12.8	From Significant Metabolites to Pathways; Bioinformatics for Metabolite Interpretation	306
12.9	Conclusion	310

生物信息学基础篇

