科学出版社

生命科学前沿丛书

北京生物技术和新医药产业促进中心
北京生物工程学会 合作策划
科学出版社生命科学编辑部

张成岗 贺福初 编著

生物信息学
方法与实践

内 容 简 介

本书是《生命科学前沿丛书》之一，从应用角度对生物信息学的有关资源进行描述，提供重要软件的用法及其分析原理，介绍如何构建局域网内部的生物信息学综合分析平台。全书共分 4 章，以核酸和蛋白质序列分析为重点进行展开，分别介绍如何从本地化以及网络上使用生物信息学资源。本书面向不同层次读者，是一本实用性极强的操作指南，将进一步促进生物信息学在我国的推广普及、研究与应用。

本书语言亲切，图文并茂，有具体而典型的实例。可供广大生物信息学入门和提高的读者参考使用。

图书在版编目（CIP）数据

生物信息学方法与实践/张成岗，贺福初编著. —北京：科学出版社，2002.6
(生命科学前沿丛书)
ISBN 7-03-010436-6

Ⅰ. 生… Ⅱ. ①张… ②贺… Ⅲ. 生物信息论 Ⅳ. Q811.4

中国版本图书馆 CIP 数据核字(2002)第 030914 号

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
源海印刷厂 印刷

科学出版社发行 各地新华书店经销
*
2002 年 6 月第 一 版 开本：720×1000 B5
2002 年 6 月第一次印刷 印张：19 1/4
印数：1—3 000 字数：355 000

定价：40.00 元

（如有印装质量问题，我社负责调换<北燕>）

《生命科学前沿丛书》专家委员会

主任委员： 吴　旻

委　　员：（按汉语拼音排序）

陈永福　　陈　竺　　范云六　　贺福初

黄大昉　　李家洋　　李衍达　　马大龙

强伯勤　　沈倍奋　　王琳芳

总 序

近年来，生命科学的发展进入了黄金时期，由人类基因组计划衍生出的一批新兴技术——“e-cell”、“克隆技术”、“转基因技术”、“生物信息学”等等，如雨后春笋般呈现在人们面前。

在近 20 余年的现代生命科学发展过程中，跨学科、跨领域融合以及新思想、新方法的不断涌现，使生命科学的各个分支领域的前沿以惊人的速度不断取得突破性进展。现代生命科学所展现的美好未来，也诱使人们以前所未有的热情关注着生命科学的最新发展。

北京生命科学研究力求与世界先进水平同步，密切注视着生命科学前沿研究的最新动态。北京从事生命科学前沿研究队伍的核心力量越来越年轻化，京区青年生命科学家的最新科研成果更新率及知识刷新率亟需提高，为此，我们在《生命科学前沿系列丛书》中将京区从事生命科学前沿研究的优秀青年科学家的最新研究进展展示给大家，对生命科学领域近年来的新技术、新成果、新问题作了精辟的介绍。

愿我国的生命科学能更上一层楼！

北京生物技术和新医药产业促进中心

北京生物工程学会

2002 年 6 月

序

生物信息学目前可以说是生命科学研究领域中一颗璀璨的明珠。由于计算机、网络以及由此导致的信息科学的快速发展，利用信息技术剖析生命现象的本质就成为生命科学研究工作者关注的焦点，尤其是随着人类基因组计划的快速发展，大量核酸和蛋白质序列以指数形式呈现迅猛的骤增态势，更是要求采用最新的信息学技术去探索这些核酸和蛋白质序列所蕴藏的生命意义。生物信息学正是在这一需求和背景下应运而生的。国际上许多单位如美国国家生物技术信息中心(NCBI)、欧洲生物信息学研究所(EBI)等提供了大量的核酸和蛋白质序列以及丰富的基于互联网的生物信息学资源。目前，核酸和蛋白质序列的常规分析策略已经成为生命科学工作者的一个基本技能。然而，国内关于生物信息学方面研究和应用的专著却十分缺乏，在很大程度上限制了生物信息学在我国的推广和应用。本书正是在这一形势下面世的，旨在将目前关于核酸和蛋白质序列分析方面的资源和策略进行有机整合后呈现给读者，对于在我国推广和提高生物信息学的研究与应用水平具有十分重要的价值。

纵观全书内容，主要在于系统地介绍了生物信息学在核酸和蛋白质序列分析方面的方法与实践，较大程度地覆盖了核酸和蛋白质序列分析方面的内容，例如核酸序列的酶切位点分析、PCR 反应中的引物设计、核酸和蛋白质序列的同源性分析、新基因的功能预测以及分子进化分析等。同时，本书还用了相当多的篇幅介绍如何基于 PC 机/Linux 操作系统构建本地化的生物信息学综合分析平台，对于具体实践操作很有帮助，因而本书在很大程度上是一本集生物信息学方法与实践相结合的专业论著，对于国内从事生物信息学研究者具有极其重要的指导作用。

吴昊

2001.8.11

此为试读，需要完整PDF请访问：www.ertongbook.com

前 言

在人类基因组计划的推动下，以生物信息的采集、处理、存储、传布、分析和解释等多个方面为研究内容的生物信息学也得到了很大发展。目前，核酸和蛋白质的序列分析已经成为生命科学工作者的一个不可或缺的基本技能。在作者的科研实践中，倍感缺乏生物信息学介绍和应用的专著。国内为数不多的几本生物信息学论著中，有的因出版时间较早而较少涉及国际互联网上的重要资源，有的则只是对生物信息学资源进行泛泛地介绍，而缺乏生物信息学中“怎么去做”的内容，后者实际上是从事有关研究所急需的内容。根据作者的个人实践，本书则力图从应用角度对生物信息学的有关资源进行描述，提供重要软件的用法及其分析原理，介绍如何构建局域网内部的生物信息学综合分析平台。因而，本书正是面向应用的生物信息学论著，旨在将目前关于核酸和蛋白质序列分析方面的资源和策略进行有机整合后呈现给读者，为广大读者提供一本行之有效的参考读物，或可称之为操作指南。

生物信息学的发展在很大程度上依赖于计算机技术的发展，所以，几乎在生物信息学的每个角落里都有计算机的身影，这是为什么很多人想学习生物信息学但是又不敢贸然涉足的原因之一，也是很多人了解了很多生物信息学的内容但是却又始终无法理解生物信息学的根本含义是什么的原因。本书的目的不想过多地去描述生物信息学理论，因为其中涉及到很多具体的算法，只是希望能够通过一些关键的操作技术让读者自己去体会、去感知生物信息学能够为自己做些什么，以及如何去做。而在此过程中，读者自然会以自己的眼光去领悟和感知生物信息学的一些理论范畴的东西。所以，本书以“生物信息学方法与实践”冠名，重点围绕“核酸和蛋白质序列分析”为中心进行展开，分别从本地化使用生物信息学资源以及使用网络上的生物信息学资源，面向不同层次的读者，应能够对生物信息学在我国的普及、研究与应用起积极的推动作用。

全书共分 4 章。第一章介绍了生物信息学的基础知识，包括相关的计算机及网络知识、生物信息学数据库及部分算法。第二章全面介绍核酸序列分析技术，内容包括电子基因定位、序列相似性分析、ORF 查询和序列提交等。第三章介绍了蛋白质序列分析技术，内容包括氨基酸组成、分子质量计算、蛋白质功能预测、结构预测以及分子进化分析等。第四章则对常用的生物信息学资源进行综合介绍，并以实例方式介绍构建生物信息学综合分析平台的技术策略，尤其重要的是以一个具体而又典型的例子介绍了生物信息学如何应用于新基因的研究工作。

需要指出的是，本书要求作者具有一定的生物医学背景知识，所以在涉及一些软件的分析结果时对其生物学意义仅作扼要介绍。同时，由于作者的水平有限，

书中肯定存在不少错误，敬请读者批评指正。

成功的经验往往需要通过个人实践才能获得，希望本书对读者在实践过程中起到抛砖引玉的作用。

张成岗　贺福初

2001 年 8 月 6 日

目　录

总序
序
前言
第 1 章　生物信息学基础 ……(1)
1.1　计算机网络及计算(机)环境简介 ……(1)
1.1.1　WEB 中的部分生物信息学资源简介 ……(2)
1.1.2　WEB 中的重要搜索工具 ……(2)
1.1.3　WEB 中的部分生物信息学相关新闻组资源参考 ……(3)
1.1.4　使用基于网络的工具……(4)
1.1.5　电子邮件(e-mail)服务 ……(4)
1.1.6　匿名 FTP 服务——获得软件的重要途径 ……(5)
1.1.7　网络规则——索取与奉献……(6)
1.1.8　从事生物信息学研究应掌握的计算机语言……(7)
1.2　生物信息学数据库及其分析……(8)
1.2.1　基本数据库……(8)
1.2.1.1　DNA 数据库 ……(8)
1.2.1.2　基因组数据库 …… (16)
1.2.1.3　蛋白质序列数据库 …… (18)
1.2.1.4　蛋白质结构数据库 …… (25)
1.2.2　常用数据库…… (25)
1.3　基本序列数据库注释及序列格式…… (38)
1.3.1　基本序列数据库注释…… (38)
1.3.2　序列格式…… (41)
1.4　信息检索系统…… (43)
1.4.1　SRS 序列检索系统…… (43)
1.4.2　Entrez 信息检索系统 …… (44)
1.4.3　DBGET/LinkDB 检索工具 …… (46)
1.5　序列对齐分析…… (46)
1.5.1　记分矩阵…… (47)
1.5.2　空位罚分…… (48)
1.5.3　两两对齐分析…… (48)
1.5.4　多重序列对齐分析…… (49)

1.5.5 序列对库的对齐检索分析 …… (50)
1.5.5.1 BLAST 检索服务 …… (52)
1.5.5.2 FASTA 检索服务 …… (57)
1.5.5.3 Blitz 蛋白质序列对库检索服务 …… (62)
1.5.6 同源性有效的意义判据 …… (63)
第 2 章 核酸序列分析 …… (64)
2.1 核酸序列的检索 …… (64)
2.2 核酸序列的基本分析 …… (65)
2.2.1 分子质量、碱基组成、碱基分布 …… (65)
2.2.2 序列变换 …… (65)
2.2.3 限制性酶切分析 …… (66)
2.2.4 克隆测序分析 …… (68)
2.2.4.1 测序峰图查看 …… (68)
2.2.4.2 核酸测序中载体序列的识别与去除 …… (69)
2.2.4.3 其他人工序列的分析与去除 …… (72)
2.3 核酸序列的电子延伸 …… (72)
利用 UniGene 数据库进行电子延伸 …… (74)
2.4 基因的电子表达谱分析 …… (76)
2.4.1 利用 UniGene 数据库进行电子表达谱分析 …… (77)
2.4.2 利用 Tigem 的电子原位杂交服务器进行电子表达谱分析 … (77)
2.5 核酸序列的电子基因定位分析 …… (78)
2.5.1 利用 STS 数据库进行电子基因定位 …… (79)
2.5.2 利用 UniGene 数据库进行电子基因定位 …… (79)
2.5.3 直接利用基因组序列进行电子基因定位 …… (80)
2.6 cDNA 对应的基因组序列分析 …… (81)
2.6.1 通过从 NCBI 查询全部基因组数据库进行基因组序列的分析 …… (82)
2.6.2 通过从 Sanger 中心查询基因组数据库进行基因组序列的分析 …… (86)
2.7 基于核酸序列对齐分析的功能预测 …… (86)
2.7.1 基于 NCBI/Blast 软件的核酸序列同源性分析 …… (86)
2.7.2 两条核酸序列之间的同源性分析 …… (89)
2.7.3 核酸序列之间的多重比对分析及进化分析 …… (90)
2.8 可读框架分析 …… (91)
2.8.1 cDNA 序列的可读框架分析 …… (91)

2.8.2 基因组序列中的编码区/内含子结构分析 …… (93)
2.8.2.1 “断裂”的真核基因 …… (93)
2.8.2.2 真核基因外显子-内含子连接区 …… (94)
2.8.2.3 基因组序列的内含子/外显子分析 …… (95)
2.8.2.4 cDNA 序列与基因组序列的对齐及其显示 …… (96)
2.9 基因启动子及其他 DNA 调控位点分析 …… (100)
2.10 重复序列分析 …… (102)
2.10.1 RepBase …… (102)
2.10.2 利用 RepeatMasker 程序分析重复序列 …… (103)
2.11 引物设计 …… (103)
2.12 向数据库中提交核酸序列 …… (106)
2.12.1 EST 序列的注册 …… (107)
2.12.2 较长或全长 cDNA 序列的注册 …… (108)
2.13 从 IMAGE 协作组索取相关克隆 …… (109)
第 3 章 蛋白质序列分析实践 …… (110)
3.1 蛋白质序列检索 …… (110)
3.1.1 基于网络的序列检索 …… (110)
3.1.1.1 从 NCBI 检索蛋白质序列 …… (110)
3.1.1.2 利用 SRS 系统从 EMBL 检索蛋白质序列 …… (110)
3.1.2 通过 e-mail 进行序列检索 …… (119)
3.2 蛋白质基本性质分析 …… (120)
3.2.1 疏水性分析 …… (120)
3.2.2 跨膜区分析 …… (121)
3.2.3 前导肽和蛋白质定位 …… (123)
3.2.4 卷曲螺旋分析 …… (125)
3.3 蛋白质功能预测 …… (126)
3.3.1 基于序列同源性分析的蛋白质功能预测 …… (126)
3.3.1.1 基于 NCBI/Blast 软件的蛋白质序列同源性分析 …… (127)
3.3.1.2 基于 WU/Blast2 软件的蛋白质序列同源性分析 …… (127)
3.3.1.3 基于 FASTA 软件的蛋白质序列同源性分析 …… (129)
3.3.2 基于 motif、结构位点、结构功能域数据库的蛋白质功能预测 …… (129)
3.3.2.1 motif 数据库——PROSITE …… (130)
3.3.2.2 Profile 数据库 …… (133)
3.3.2.3 蛋白质序列的轮廓(Profile)分析 …… (134)
3.3.2.4 HITS 蛋白质结构域数据库 …… (134)

3.3.2.5 InterProScan 综合分析网站……(135)
3.3.2.6 蛋白质的结构功能域分析 ……(136)
3.4 蛋白质结构预测……(138)
3.4.1 蛋白质结构资源……(138)
3.4.1.1 PDB 数据库 ……(138)
3.4.1.2 NRL-3D 数据库 ……(139)
3.4.1.3 ISSD 数据库 ……(139)
3.4.1.4 HSSP 数据库 ……(139)
3.4.1.5 蛋白质结构分类数据库(SCOP) ……(139)
3.4.1.6 MMDB 蛋白质分子模型数据库 ……(139)
3.4.1.7 Dali/FSSP 数据库 ……(140)
3.4.2 蛋白质二级结构预测……(140)
3.4.3 蛋白质三级结构预测……(140)
3.4.3.1 与已知结构的序列比较 ……(140)
3.4.3.2 同源模建 ……(141)
3.4.3.3 穿针引线(threading)算法和折叠识别 ……(141)
3.5 蛋白质分子进化分析……(142)
3.5.1 蛋白质分类数据库(ProtoMap) ……(143)
3.5.2 蛋白质序列多重对齐分析及进化分析……(143)
第 4 章 常用的生物信息学资源简介及其综合利用 ……(147)
4.1 计算机服务/开发环境的构建 ……(147)
4.1.1 程序开发语言……(147)
4.1.1.1 C 语言 ……(147)
4.1.1.2 Perl 语言 ……(149)
4.1.1.3 PHP 语言 ……(151)
4.1.2 数据库工具……(151)
4.1.2.1 MySQL 数据库工具……(151)
4.1.2.2 AceDB 数据库及管理工具 ……(152)
4.1.3 网络服务器……(153)
4.1.3.1 Linux 下的 Apache 网络服务器 ……(153)
4.1.3.2 Windows 下的 Apache 网络服务器……(154)
4.1.4 操作系统……(154)
4.1.4.1 Linux 操作系统 ……(154)
4.1.4.2 常用的 Linux 命令 ……(158)
4.1.4.3 Linux 与 Windows NT/2000 相比的几个技术优势 …(*161*)

此为试读，需要完整PDF请访问：www.ertongbook.com

4.1.4.4 Linux 与 Windows 系统的集成 ……(163)
4.2 Windows 下的软件资源推荐……(164)
4.2.1 软件的下载与安装……(164)
4.2.2 文件管理软件——Windows commander ……(164)
4.2.3 文件下载——Net Vampire 软件 ……(170)
4.2.4 文件传输协议——FTP 命令……(173)
4.2.5 建立 FTP 服务器——FTP SerV-U 软件 ……(174)
4.2.6 创建网站相关软件——Webzip 软件 ……(175)
4.2.7 图形处理软件——HyperSnap ……(175)
4.2.8 远程登录/远程管理 ……(176)
4.2.8.1 Telnet 服务程序 ……(176)
4.2.8.2 NetTerm 远程登录软件 ……(177)
4.2.9 压缩与解压缩工具……(178)
4.2.9.1 压缩软件——Winzip ……(178)
4.2.9.2 压缩软件——WinRAR ……(179)
4.2.10 超大文本编辑软件——Ultra Edit ……(180)
4.2.11 程序集成开发环境——Visual BASIC ……(182)
4.2.12 网络浏览器——FastBrowser ……(185)
4.3 生物信息学软件资源……(186)
4.3.1 Windows 环境下的生物信息学资源……(186)
4.3.1.1 序列分析软件——DNAMAN……(186)
4.3.1.2 综合序列分析软件——BioEdit ……(196)
4.3.1.3 Vector NTI……(198)
4.3.1.4 引物设计软件——Oligo ……(200)
4.3.1.5 核酸序列分析软件——GeneTool ……(202)
4.3.1.6 蛋白质序列分析软件——PepTool ……(203)
4.3.1.7 序列分析软件——Lasergene99 ……(204)
4.3.1.8 蛋白质三维分子结构显示软件——RasMol ……(205)
4.3.1.9 序列分析与管理软件——Omiga ……(209)
4.3.1.10 序列多重对齐软件——ClustalW ……(212)
4.3.2 Linux 环境下的生物信息学资源 ……(216)
4.3.3 Macintosh 环境下的核酸和蛋白质序列分析 ……(217)
4.3.3.1 MacOS 的部分工具 ……(217)
4.3.3.2 MacOS 下的生物信息学分析资源 ……(219)
4.3.4 综合生物信息学资源——生物软件网……(221)
4.4 资源的综合利用:自建核酸和蛋白质序列分析平台 ……(238)

4.4.1 Windows 下 Blast 软件的本地化实现及其使用 ……(239)
4.4.1.1 下载软件 ……(239)
4.4.1.2 软件解压缩 ……(239)
4.4.1.3 进行系统配置 ……(240)
4.4.1.4 Blast 软件的使用 ……(241)
4.4.1.5 Visual BASIC 程序接口设计及使用示例 ……(242)
4.4.2 Linux 系统下命令行方式 Blast 软件的安装与使用 ……(243)
4.4.3 含有 Web 界面的 Blast 系统的安装与使用 ……(243)
4.4.3.1 Linux 操作系统安装及局域网组建 ……(244)
4.4.3.2 WEB 界面 Blast 软件的安装 ……(244)
4.4.3.3 检索用数据库的准备 ……(245)
4.4.3.4 Blast 软件的配置 ……(246)
4.4.3.5 Blast 分析环境的使用 ……(247)
4.4.4 基于 PC/Linux 的核酸序列电子延伸系统(AutoCTG)的构建及其应用 ……(249)
4.4.4.1 电子序列延伸的生物信息学策略 ……(249)
4.4.4.2 程序设计 ……(249)
4.4.4.3 系统需求 ……(249)
4.4.4.4 体系性能的综合评价 ……(249)
4.4.4.5 数据库预处理 ……(250)
4.4.4.6 程序设计及用法 ……(251)
4.4.4.7 人胎肝来源部分 EST 序列和较长 cDNA 序列的电子延伸分析 ……(252)
4.4.5 基于 PC/Linux 的核酸序列分析系统的构建及其应用 ……(257)
4.4.5.1 本地化核酸序列大规模自动分析系统的构建 ……(258)
4.4.5.2 本地化核酸序列大规模分析体系的使用 ……(264)
4.5 实例分析：人 ADP-核糖基化因子 GTP 酶活化蛋白基因的生物信息学分析 ……(267)
4.5.1 cDNA 序列分析 ……(268)
4.5.1.1 EST 序列的获得 ……(268)
4.5.1.2 利用 Blast 软件进行序列相似性检索 ……(268)
4.5.1.3 确定转录物大小 ……(270)
4.5.1.4 全长 cDNA 序列的获得 ……(271)
4.5.1.5 可读框架分析 ……(272)
4.5.1.6 基因名称确定 ……(274)
4.5.2 蛋白质序列分析 ……(274)

4.5.2.1 基本性质分析 ……(274)
4.5.2.2 功能位点分析 ……(274)
4.5.2.3 结构功能域的确定 ……(275)
4.5.2.4 序列多重对齐分析 ……(275)
4.5.3 基因组结构分析 ……(277)
4.5.3.1 染色体定位分析 ……(277)
4.5.3.2 基因组结构确定 ……(279)
4.5.4 小结 ……(281)
附录
常用词汇与缩略语表 ……(282)
参考文献 ……(287)

第 1 章　生物信息学基础

生物信息学自诞生之日起，计算机似乎就注定要成为其核心工具。早在美国国家生物医学研究基金会(NBRF)的 Dayhoff 于 1966 年开始收集序列信息并发表了她编辑的第一本《蛋白质序列和结构集》后不久，就已经有了采用计算机技术对数据进行管理和分析的思想。后来，由于计算机技术的发展，对于以核酸序列和蛋白质序列为主的生物信息的管理和分析就几乎完全依赖于计算机了。目前，在国际互联网日益发展的情况下，凭借网络的数据传播和数据分析已经成为生物信息学资源利用的主要方式。鉴于生物信息学中涉及到重要的计算机知识，所以，在论述生物信息学之前，我们先对有关的计算机知识作简要介绍，以便于读者对后续内容的理解和掌握。

1.1　计算机网络及计算(机)环境简介

计算机是进行生物信息学分析的必备工具。从计算规模上而言，计算机可分为大型机、中型机、小型机和微机。不同的计算机采用不同的操作系统，例如 UNIX、Linux、MacOS、Windows 95/98/NT/2000 操作系统等。其中，UNIX 操作系统以其稳定的计算环境和良好的多用户支持自然而然地成为企业内部和研究所采用的稳定的计算平台。Macintosh 机由于具有优良的图像支持能力而拥有大量优良的图形界面环境。而微软公司(Microsoft)的 Windows 操作系统则成为个人电脑(personal computer, PC)上的主流操作系统。值得一提的是，近年来 Linux 操作系统(http://www.linux.org)则“异军突起”。它是一个以 Intel 系列 CPU、Macintosh 等硬件为平台，完全免费的 UNIX 兼容系统，在微机上能够实现 UNIX 操作系统的大多数功能，在生物信息学分析中已经显示出巨大的生命力。国内研究院所的用户大多数以微机为主进行工作，但是，在从事生物信息学分析的过程中，用户将发现需要同时安装 Linux 操作系统以满足更为深入的生物信息学分析任务，而且，Windows 9X/2000 和 Linux 具有很好的兼容性(详见第 4 章)。

近年来计算机通讯方面的进步如国际互联网(World Wide Web, 简称 WWW 或 WEB)使得生物信息学资源的存取极为方便。通过联网的电脑，大量的生物信息学资源可以快速、免费地传送到分子生物学家的手中。许多网站还提供基于 UNIX/Linux、Windows、Macintosh 等多种操作系统的生物信息学软件和分析平台。本书假设用户能够联网(采用 Microsoft Internet Explorer 或者 Netscape 等浏览器)。

本节主要对国际互联网上与生物信息学相关的资源进行介绍，使读者能够从中粗窥生物信息学的面貌，而较为详细的内容将在后续章节陆续介绍。

1.1.1 WEB 中的部分生物信息学资源简介

WEB 的发展十分快速。很多人将生物信息学资源进行整理并制备了主页。在对这些主页进行浏览时，用户能够快速地获得有关的生物信息学资源。表 1-1中列出了部分重要资源及其网址。将这些主页加入到浏览器的收藏夹中能够方便后续的工作。然而，该表中所列资源仅仅是一小部分，在网上随便用相应关键词进行搜索就能够得到成千上万的有用的主页。

表 1-1 WEB 中的部分生物信息学资源参考

网址	说明
http://www.expasy.ch/alinks.html	Amos Bairoch 的个人感兴趣的主页——也许是数据库及其资源的最理想集合
http://www.biophys.uni-duesseldorf.de/BioNet/Pedro/research_tools.html	Pedro 的生物分子研究工具——这是一个广为人知的资源，但更新太慢
http://www.expasy.org/tools/	核酸和蛋白质序列分析工具清单
http://ahsc.arizona.edu/~lei/genetics/tools.htm	一个精心构建的主页，和其他站点有广泛的链接、资源描述和 WEB 搜索系统
http://www.incyte.com/index.shtml	精心构建的主页(Incyte 公司)
http://www.ebi.ac.uk/bioworld	生物学相关网站的搜索引擎

1.1.2 WEB 中的重要搜索工具

新的生物信息学网站和主页一直在层出不穷。也许上述网址很快就被更新或

表 1-2 WEB 中的重要搜索工具参考

网址	说明
http://www.lycos.com/	可使用关键词进行搜索的一个良好网站
http://www.yahoo.com/	主页按照层次进行了划分
http://www.infoseek.com	对搜索的结果进行打分，便于用户参考
http://www.excite.com	基于概念(concept)的搜索
http://www.webcrawler.com	支持自然语言方式(natural language)的搜索
http://www.sohu.com	著名的搜狐搜索引擎，可检索中文生物信息学资源
http://www.sina.com.cn	著名的新浪搜索引擎，可检索中文生物信息学资源

不可用了。在国际互联网上有大量不同的搜索引擎便于用户利用关键词快速检索所需的资源。一些有用的搜索资源参见表 1-2。需要注意的是，这些搜索引擎中没有一个是万能的，具体使用时需综合参考。笔者一般使用 lycos (http://www.lycos.com)进行搜索，其优点是搜索结果比较准确，而且可对搜索结果进行二次检索，能够提高检索效率。

1.1.3 WEB 中的部分生物信息学相关新闻组资源参考

参加新闻组(newsgroup)是紧跟生物信息学发展步伐的一个重要方式。为了使用新闻组，用户需要一个特定的新闻阅读软件，并拥有访问特定新闻服务器的权限。一般的网络浏览器如 Internet Explorer 和 Netscape 等均内置了新闻阅读器。新闻阅读软件也可从很多网站下载，例如 http://www.forteinc.com。通常，提供 Internet 访问权限的服务商会同时提供新闻阅读软件。

在 bionet 新闻组中有很多非常好的生物信息学资源(部分见表 1-3)。如果用户使用特定资源有困难，那么可通过向 bionet.software 提问得到回答。新闻服务器通常会及时更新以保持信息的非冗余性。

表 1-3 WEB 中的部分生物信息学相关新闻组资源参考

网址	说明
news:bionet.announce	对生物学家具有广泛影响的信息
news:bionet.biophysics	生物物理方面的内容
news:bionet.celegans	关于线虫(*C. elegans*)研究的讨论
news:bionet.general	生物科学的一般性讨论
news:bionet.glycosci	关于糖类和复合糖分子研究的讨论
news:bionet.immunology	免疫学方面的讨论
news:bionet.infotheory	关于生物信息理论的讨论
news:bionet.jobs	科研工作职位
news:bionet.software	生物软件信息

通常，新闻组服务器上的信息只允许保持较短时间后即被删除。然而，所有信息同时也被保存在 http://www.bio.net。通过选择“存取 BIOSCI/bionet 新闻组”，用户可以搜索某一主题是否曾经被讨论过。这其实是在 Web 上了解生物信息学资源的一个极佳的方法。也许同样的问题已经被咨询了几十遍，但是用户仍然可以向新闻服务器进行提问。当然，应先检索以前的提问与回答以免浪费时间。

此为试读，需要完整PDF请访问：www.ertongbook.com