

■ 主编 吕 军

E-mail

访问网络 生物医学资源 实用指南

62

上海科学技术出版社



E-mail 访问网络 生物医学资源实用指南

主编 吕 军

上海科学技术出版社

图书在版编目(CIP)数据

E-mail 访问网络生物医学资源实用指南 / 吕军主编.
上海: 上海科学技术出版社, 2002. 9
ISBN 7-5323-6535-2

I. E... II. 吕... III. ①因特网-应用-医学-情报检索-指南②因特网-应用-生物学-情报检索-指南 IV.
G252.7-62

中国版本图书馆 CIP 数据核字(2002)第 027589 号

上海科学技术出版社出版发行

(上海瑞金二路 450 号 邮政编码 200020)

同济大学印刷厂印刷 新华书店上海发行所经销

2002 年 9 月第 1 版 2002 年 9 月第 1 次印刷

开本 787 × 1092 1/32 印张 8.4 字数 169 000

印数 1—5 500 定价: 15.00 元

本书如有缺页、错装或损坏等严重质量问题,
请向本社出版科联系调换

内 容 提 要

本书是一本介绍如何通过 E-mail 获取 Internet 上医学和生物学信息资源的基本原理和方法的工具书。包括的内容有:检索 Medline——蛋白质、核酸与三维结构数据库;检索与分析与人类基因组研究相关的一些重要的网络数据库,如序列的同源性、开放阅读框架(ORF)预测、蛋白质二级结构分析、RNA 二级结构预测、分子模建等,这些都是生命科学最前沿的研究热点。现代科学研究的一个显著特点就是综合性大研究,学科之间交叉、联系日益紧密。本书介绍了如何通过 Internet 上邮件列表和新闻组这两种形式在世界范围内与同行或非同行交流、联系、协作、探讨;如何访问 WWW 上的医学生物学资源和 WWW 的各类搜索引擎;如何访问 FTP 资源,寻找和下载有用的一些生物学档案资料;附录介绍的一些使用的 E-mail 技巧使各章介绍的方法事半功倍。

本书的主要读者对象为医学、药学和生命科学等领域的大学生、研究生、医务工作者、教师和研究人員。

编写人员名单

主 编 吕 军

副主编 冯伟华 胡惠民

编 者 (按姓氏笔画为序)

冯伟华 吕 军 赵 东 奚 昊

曹 勇 康 斌 彭 玲 翟海峰

序

近年来,随着人类基因组计划的初步完成及快速序列测定、生物芯片等高通量技术的出现和广泛应用,现代生物医学研究已经进入了以功能基因组学和生物信息学为主题的新领域。同时,随着基因序列和表达谱等生物学实验数据呈爆炸趋势增长,大规模数据的贮存、处理和传输在计算机和国际互联网的迅速发展下越来越容易,现在实验室的研究成果可以迅速进入生物信息网络为全球共享。作为一门新的学科领域,生物信息学以基因组 DNA 序列信息分析作为源头,早期的研究侧重于结构基因组,现在则以功能基因组和蛋白质组为核心,在获得了蛋白质编码区的信息之后进行蛋白质空间结构模拟和预测,然后依据特定蛋白质的功能进行可能的药物设计等等,涉及面非常广泛,同时也是一门综合了生物学、数学、物理、化学、计算机学科等的交叉学科。在应用与发展研究方面,生物信息学汇集与疾病相关的人类基因信息,发展患者样品序列信息检测技术和基于序列信息选择表达载体、PCR 引物的技术,建立相关的数据库以及与大分子设计和药物设计相关的数据库等。

可以预见,生物信息学是未来生物科学研究的重要学科和发展方向之一,有效地掌握一些生物信息学的基本工具,将对课题研究起到事半功倍的效果。本书在这方面做了一些有益的尝试,拉近了学术研究与实际应用方面的距离,尤其对生物医学方面的工作人员,有一定的启发和帮助作用。尤为欣

喜的是,本书编者均为在一线从事本专业研究工作的优秀中、青年专家,具有扎实的理论基础和娴熟的实验操作技术。相信本书的出版,将有助于推动国内生物信息普及相关研究的不断进步。

第二军医大学生物化学与分子生物学教研室

焦炳华

2002年4月20日

前 言

尊敬的读者,非常荣幸您能使用这本《E-mail 访问网络生物医学资源实用指南》。这是特为那些在医学和生命科学领域学习和工作的人们提供的一本利用 Internet 相关信息资源的工具书。当您打开这本指南的时候,也许会感到很惊讶,只是使用 E-mail 这一 Internet 上最为简单的访问方式,竟然可以用到 Internet 上几乎所有的生物医学资源?! 的确如此。

与在线方式相比,E-mail 不仅方便、快捷、可靠,而且可以节省高昂的网络费用。本书力求做到:即使是没有任何网络经验的生物医学工作者,也能迅速地通过 E-mail 找到您想要的资料,而无需付出较多的时间和网费。

Internet 在某种意义上也是一个巨大的 help-each-other-net。在您享受到 E-mail 给您带来方便的同时,请感谢那些日日夜夜无偿提供和维护这些信息资源的人们,是他(她)们组成了 Internet 的灵魂。

这本书是利用业余时间编写而成的,我们试图向大家提供尽可能多的资源访问信息,无奈 Internet 实在是太大了,有关医学和生命科学方面的网络资源也正如这一领域的发展一样增长和更新得很快,加之限于篇幅等因素,因此,只能是抛砖引玉,难免挂一漏万。希望读者朋友们在学习了书中介绍的基本原理和方法后,举一反三,推陈出新。希望能及时得到您的意见和建议,如果您发现了新的方法和资源,请及时与我们交流,以便使本书不断更新,使更多的同行享受到您所带来

的便利。

本书的主要读者对象为医学、药学和生命科学等领域的大学生、研究生、医务工作者、教师和研究人员。

编 者

2001 年 10 月

目 录

绪 论	1
第一章 Medline——蛋白质、核酸与三维结构数据库检索	5
第一节 简介	8
第二节 Entrez 可检索的数据库	12
第三节 词条检索	13
一、字段说明符	14
二、字段解释	15
三、布尔检索	22
四、截词检索	26
五、范围检索	26
第四节 搜索参数选项	28
一、显示方式选项	28
二、其他选项	32
第五节 替代符号	32
第六节 流水号检索	33
一、GenBank 登录号与序列标识符(GI 号)	34
二、用于 UID 中的 Fasta 格式	35
第二章 基因和基因组序列分析	37
第一节 NCBI BLAST 电子邮件服务器的介绍	50
一、访问 NCBI BLAST 电子邮件服务器	50

二、BLAST 电子邮件检索请求示例	50
三、NCBI BLAST 电子邮件服务器检索的可选程序	51
四、NCBI BLAST 电子邮件服务器可用数据库	52
五、NCBI BLAST 电子邮件服务器指令表	54
六、利用 NCBI BLAST 电子邮件服务器检索的序列格式	57
七、序列标识符	59
八、检索请求的呈递与等候	60
九、向 GenBank 呈递新的数据或更新修订	60
第二节 BLAST 程序使用详解	61
一、BLAST 程序算法	61
二、BLAST 序列对比分值方案	62
三、邻域(the neighborhood, T)	64
四、HSP 分值阈	65
五、Karlin-Altschul 统计学	66
六、SUM 统计学和 Poisson 统计学	68
七、P 值、比对比分值及信息	68
八、匹配报告中的选择度	69
九、搜索敏感度	70
十、遗传代码	72
十一、设置参数	73
十二、控制输出	80
十三、环境变量	80
十四、支持功能	81
十五、输出示例与详解	81
十六、原因不明的故障	89
第三节 BLAST 常见问题解答	91
一、最常见的问题	91
二、BLAST 程序	95
三、查询格式	97

四、数据库结构	98
五、选项/参数	98
六、输出	99
七、检索策略	102
八、电子邮件服务器	103
九、其他资源	105
第四节 其他核酸与蛋白质序列分析 E-mail 服务器	
简介	105
一、EMBL FASTA 蛋白质与 DNA 序列同源性检索	105
二、GRAIL & GENQUEST 序列分析、组合与比较系统	108
三、DDBJ FASTA 序列同源性检索	108
四、DDBJ malign 序列多重比对检索	110
五、DDBJ BLAST 序列同源性检索	111
六、DDBJ SSEARCH 序列同源性检索	112
七、DDBJ SSTRHEAD 蛋白质二级结构预测	113
八、NCBI RETRIEVE 序列检索服务	114
九、EMBL BLITZ 蛋白质序列同源性检索服务	116
十、BLOCK 序列分析	116
十一、EBI 网络文件服务器	117
十二、Predict Protein 蛋白质结构预测	119
十三、nnpredict 蛋白质二级结构预测	120
十四、CBRC 序列分析	121
十五、EST 报告	122
十六、未知蛋白质序列功能分析	123
十七、生物医学数据库	126
第三章 访问医学与生物学邮件列表	147
第一节 邮件列表简介	147
第二节 查找订阅医学与生物学邮件列表	148

一、使用 Listserv 列表服务器查找邮件列表	148
二、使用 WWW 列表服务器查找邮件列表	148
三、使用 lizster 列表服务器	149
四、订阅、停订以及检索所有邮递表	149
第三节 Internet 上的医学与生物学邮件列表简介	150
第四章 访问医学与生物学新闻组	154
第一节 新闻组简介	154
第二节 查找新闻组	158
一、使用 reference.com	158
二、使用搜索引擎	159
三、nova.edu	159
四、使用 lisa 搜索	159
五、访问 dejanews	159
六、查询中文新闻组,可使用中文搜索引擎 Cseek	159
七、新闻组的完整列表	160
第三节 阅读新闻组	160
一、使用 Gopher mail 服务器	160
二、使用 Agora 服务器	161
三、通过 WWW 方式	161
四、获取常见问题解答(FAQ)	161
第四节 如何将邮件投递给新闻组	163
第五节 一些主要的医学和生命科学新闻组	163
第五章 访问医学与生物学 WWW 资源	169
第一节 简介 WWW E-mail 服务器	169
第二节 主要 WWW E-mail 服务器	170
一、主要 WWW E-mail 服务器地址与性能表	170

二、几个 WWW E-mail 服务器的具体用法	171
第三节 一些医学与生物学 WWW 网址推荐	172
一、医学网络搜寻器	172
二、免费 Medline 检索	173
三、网络虚拟环境的应用	175
四、网络医学专业期刊	176
五、医学各专业常用网址	176
第六章 通过 E-mail 访问 WWW 的各类检索引擎	196
一、检索 Yahoo	196
二、检索 Altavista	198
三、检索 Excite	198
四、检索 InfoSeek	199
五、检索 Lycos	199
六、检索 DogPile	200
七、用于访问 WWW 检索引擎的 WWW 服务器 简介	201
第七章 访问医学与生物学 FTP 资源	205
第一节 FTP 简介	205
第二节 E-mail 访问匿名 FTP 站点	206
一、使用 Agora, getweb, w3mail 邮件服务器	207
二、使用 FTPmail 服务器	207
三、使用 Cophermail 服务器	208
第三节 注册 FTP 站点的访问	211
第四节 一些 FTPmail 服务器的地址	212
第五节 如何查找 FTP 站点上的文件	214
第六节 一些医学与生物学 FTP 资源介绍	215

【附录】一些实用的 E-mail 技巧	221
一、了解 FTPmail/WEBmail 服务器工作状态	221
二、通过 E-mail 学习从初级到高级的 Internet 教程	221
三、E-mail 常用的脸谱及其含义	221
四、查找研究人员的 E-mail 地址	222
1. 使用 Whois 服务器	222
2. 使用 MIT 服务器	222
3. 使用 Netfind	222
五、使用 Archie 查找软件	223
六、查找医学杂志的文献全文	224
七、查询癌症信息(Cancermail)	224
八、用 E-mail 发传真	226
九、Internet 用国家(地区)和组织代码表	227
十、测试收发 E-mail 的功能	231
十一、韦伯(Webster)字典查询	231
十二、E-mail 常用的缩略词及其含义	231
十三、如何获取最新和最好的生物学文献	232
十四、检索基因和蛋白质序列专利信息	233

绪 论

20 世纪 70 年代诞生的基因工程使现代医学和生命科学迅猛发展,伴随着人类基因组计划(Human Genome Project, HGP)、基因组学和蛋白质组学的出现,一门崭新的学科——生物信息学(bioinformatics)诞生了,它涵盖了利用计算机和网络技术来收集、储存、整理、检索、分析和了解现代医学和生命科学信息的各个方面。

我们都知道,在 Internet 上,可以查询文献资料、进行基因序列的检索、比较,预测蛋白质组成与结构,下载软件、招聘及会议信息、知识介绍,召开学术会议,开展远程医疗、远程教学,与同行交流等等。如何轻松快捷地利用好这些网络资源,已成为医学和生命科学工作者的一项必备素质。本书旨在帮助读者能够最有效率地获取这些信息资源。

Internet 本身发展非常迅速,网络资源也日新月异,我们这里主要介绍几类能够用 E-mail 实现的最主要功能,以帮助大家尽快地各取所需,而不必花大量的时间了解网络背后复杂的原理。

第一章介绍检索 Medline——蛋白质、核酸与三维结构数据库。Internet 有关医学和生命科学的数据库非常多,最权威的当数 Medline 及其相关的数据库,尽管 Medline 数据庞大,但用 E-mail 检索起来却很轻松。

人类基因组计划的实施使生命科学的研究逐渐步入以结构和功能为主题的研究,关于这方面有大量的数据库,供检索

序列的同源性、开放阅读框架(ORF)预测、蛋白质二级结构分析、RNA 二级结构预测、分子建模等。这些都是生命科学最前沿的研究热点,进行这些检索需要更深入的分子生物学和计算机等知识背景,目前这方面的相关资料较少,本书第二章用了很大篇幅进行详细介绍,也是希望促进国内的同行们充分利用好这些信息资源,将自己的工作尽快与国外同步发展,提高创新性和深度。

现代科学研究的一个显著特点就是综合性大研究,学科之间交叉、联系日益紧密,与同行或非同行间的交流、联系、协作、探讨显得尤为重要。第三章和第四章分别介绍了如何通过 Internet 上邮件列表和新闻组这两种形式,在世界范围内与其他医学及生命科学工作者进行广泛交流的技术。

WWW 及其各类搜索引擎(如 Yahoo、Altavista)是重要的医学生物学网络资源,第五章和第六章介绍了如何通过 E-mail 来访问这些资源。这两章的技术结合起来使用,效果更好。访问 FTP 资源可以使您找到许多 WWW 上没有的、有用的医学生物学档案资料 and 软件,这在第七章介绍。

附录介绍的是一些实用的 E-mail 技巧,有的虽并不是直接的医学生物学信息,但使用好这些技术,将使本书各章介绍的方法事半功倍,如查找医学和研究人员的 E-mail 地址,查找医学等专业杂志的文献全文等。

本书的特点之一是:实用性。所有 Internet 信息资源都是可以通过 E-mail 获取的。E-mail 与在线方式相比,快捷、方便、费用低廉、操作简便,已成为网上学术交流最便捷、最流行的方式,传输量占整个 Internet 的一半多,只要在几分钟之内学会收发 E-mail,即使没有任何网络经验的医生和生命科学工作者也可以方便地向国内外著者索取原文和相关研究资