常用医学生物信息学数据库

主编 尚 彤 国强华 景 霞

北京大学医学出版社

常用医学生物信息学数据库

主　编　尚彤　国强华　景霞

编写者　（以姓氏汉语拼音为序）

国强华　景　霞　李　黔

卢　铭　芮　伟　尚　彤

石　磊　孙冬泳　孙贞媛

汤　健　张　丹　张其鹏

赵睿颖　周彩虹　朱晓华

北京大学医学出版社

CHANGYONG YIXUE SHENGWU
XINXIXUE SHUJUKU

图书在版编目(CIP)数据

常用医学生物信息学数据库/尚彤,国强华,景霞主编. —北京:北京大学医学出版社,2003
ISBN 7-81071-434-1

Ⅰ.常… Ⅱ.①尚…②国…③景… Ⅲ.医学:生物学—数据库 Ⅳ.R31

中国版本图书馆 CIP 数据核字(2003)第 018569 号

北京大学医学出版社出版发行
(100083 北京市海淀区学院路 38 号 北京大学医学部院内 电话:010-62092230)
责任编辑:罗德刚
责任校对:李月英 焦 娴
责任印制:郭桂兰
北京市地泰德印刷有限责任公司印刷 新华书店经销
开本:787 mm×1 092 mm 1/16 印张:26.75 字数:678 千字
2003 年 5 月第 1 版 2003 年 5 月第 1 次印刷 印数:1—3000 册
定价:49.00 元

内容简介

医学生物信息学是医学、生物学、计算机科学和信息科学等多学科交叉而形成的一门新兴学科。它对各种医学和生物学的信息、资料、数据进行搜集、储存、整理、计算和分析，形成可再生的资源，为医学科学的发展提供全方位的支持。数据是医学生物信息学的基础，建立以疾病为中心，贯穿病理、药理、基因、蛋白、调控等方面数据的数据库是医学生物信息的核心。

本书以医学生物信息数据库为中心，主要介绍了国内外近百个知名度高、可信度强、应用范围广的生物医学数据库或网站；着重介绍这些数据库的结构、内容、资源、用法及主要特点。全书共分为九章，第一章简要介绍医学生物信息学的基本概念；第二章主要介绍几种医学生物学综合数据库；第三、四章着重介绍核酸和蛋白质相关数据库；第五、六章主要介绍几个细胞生物学和微生物学数据库；第七、八、九章主要介绍一些与疾病相关的临床和药物专业数据库。

本书可作为医学、生物学、计算机科学和信息科学的研究人员和教学人员从事科研、教学、医学和开发的参考书。

本书由
国家重点基础研究发展规划项目(G2000056907,G1998051015)
资助出版

序

21 世纪是生物科学和信息科学飞速发展的世纪。人类基因组和蛋白质组研究的发展以及信息学工具的介入，推动了以序列为核心的生物信息学的诞生。医学信息学，则是生物信息学在医学领域的延伸与发展，是生物信息科学与基础及临床医学科学的结合和交融。它是基于信息科学的原理和方法，运用算法和软件，对医学和生命科学的各种数据和资料进行搜集、储存、整理、归纳、比较和分析，以形成可利用可再生的医学模式和资源，并予以发布、应用的一门新兴学科。它不仅包括基因和蛋白质的序列信息，还包括基因、蛋白质的结构、表达、分布、功能、调节、正常生理、病理异常和临床等系列信息，涉及分子、细胞、器官和整体等多层次、多方面，涵盖了基础医学、临床医学、预防医学、流行病学、药物设计、医学文献、数据存储、数据分析、知识挖掘、网络服务、模拟与预测等各个领域。

医学生物信息学，数据是基础，数据库的建立是核心，利用数据库进行数据分析与知识挖掘，开发软件和建立模型，实现科学的预测和资源共享，促进医学生物学发展，增进人类卫生健康是目的。随着生物医学的发展，人类正面临着“数据灾难”，只有将无数杂乱无序的数据，变为有序可用的数据；将凝固、枯燥的资料，变为生动可以扩展和再生的信息；将个体存放的信息，变为人类共享的资源，才更有意义，才具有生命力。这是医学生物信息学的任务，亦是建立医学生物信息学数据库的主要目的。

数据库的建设是国家的一项重大基础工程，目前在国际上约有 30 多万个/种不同学科和领域的数据库或网站，其中医学生物学相关的数据库约占 3%～5%。但网站或数据库的质量千差万别，容量和范围亦大不相同。现在每天都有数据库或网站新生，每天亦有数据库或网站消亡，优胜劣汰，推陈出新。就生物信息数据库而言，*Nucleic Acids Research* 每年均推荐一批优秀数据库，这些数据库建立时间长、访问率高，数据储量大，应用范围广，更新速度快，具有科学性、实用性和时效性。目前医学生物学数据库正从序列数据库向功能数据库，从综合数据库向专题数据库发展，从单一结构数据库向基础临床相结合的兼容数据库发展。数据库的功能亦从数据的储存和共享向数据挖掘、预测和模拟的方向发展。我们正面临着一个数字化的医学生物学研究的新时代，它将彻底改变人们的学习、研究、防治、开发的传统模式，极大地推动医学生物学的发展。我们希望能为此摇旗呐喊、添砖加瓦、引凤筑巢，以迎接我国即将出台的国家数字化系统工程的来临。

这本参考书以医学生物信息学为中心，主要介绍了国际上近百个知名度高、可信度强、应用范围广的生物医学数据库或网站；着重介绍这些数据库的结构、内容、资源、用法及主要特点。全书共分为九章，第一章简要介绍医学生物信息学的基本概念；第二章主要介绍几种医学生物学综合数据库；第三、四章着重介绍核酸和蛋白质相关数据库；第五、六章主要介绍几个细胞生物学和微生物学数据库；第七、八、九章主要介绍一些与疾病相关的临床和药物专业数据库，其中包括了我们研究室自主建立的心血管疾病相关数据库。

这本参考书是我们研究室一些年轻人，包括医学、生物学、计算机科学和信息科学的研究生、进修生和高年级本科学生自学、编辑和整理的。他们勤奋认真地学习和工作，阅览了大量的文献和资料，求教了许多专家和单位，经过反复查询和修订，历时两年才完成的。其精神是

十分难能可贵的。但鉴于医学生物信息学和数据库建设还是一个年轻和新兴的学科，一个不断新陈代谢的交叉学科，更由于我们的学识和经验的不足，在编辑过程中肯定会有许多疏漏、不足和错误的地方，恳请读者和专家给我们予以批评、指正。

汤　健
北京大学心血管研究所
2002 年 12 月 20 日

目　录

第一章　医学生物信息学

第一节　医学生物信息学的基本概念

一、生物信息学

自从1990年美国启动人类基因组计划以来，迄今为止已经完成了约40多种生物的全基因组测序工作。由此产生的是海量的基因序列和蛋白质序列的数据，怎样存放、管理、利用、解释这些数据就成了最迫切的问题。由此诞生了一门新的学科——生物信息学(Bioinformatics)。生物信息学通过对生物学实验数据的获取、存储、分析，来共享、解释这些数据，并且希望在这些数据的基础上对一些生物学现象进行预测。由于当前生物信息学发展的主要推动力来自分子生物学，尤其是各种基因组工程，生物信息学的研究主要集中于核酸序列和蛋白质序列。美国人类基因组计划实施五年后的总结报告中，对生物信息学作了以下定义：生物信息学是一门交叉科学，它包含了生物信息的获取、处理、存储、分发、分析和解释等在内的所有方面，它综合运用数学、计算机科学和生物学的各种工具，来阐明和理解大量数据所包含的生物学意义。生物信息学研究的内容包括了序列和结构比对、蛋白质结构预测、基因识别、分子进化、比较基因组学、序列重叠群、药物设计、基因芯片、基因表达谱等方面。

生物信息学处理的是大量的生物学数据，包括基因蛋白序列、文献数据、蛋白质空间结构等等方面。这些数据随着以人类基因组为中心的各种基因组工程的完成飞速地增长，如何存放、共享和注释。数据库作为一种有序的数据存放方式，可以很好地实现数据的共享、扩充和分析，所以数据库成为最理想的生物数据的存储形式。同时，几乎在生物学数据大量增长的同时，因特网技术开始发展并成熟起来，从而为这些数据的共享提供了便利。目前这些数据的绝大部分都是以网络数据库的形式存放，并且向全世界免费开放。利用网络数据库，研究人员可以通过因特网在世界的各地，很容易地从数据库中搜索自己研究所需要的数据，并且通过数据库来发布自己发现的基因序列。现在，以GenBank/EMBL/DDBJ、SWISS-PROT、PIR为首的一批大型的基因和蛋白序列数据库已经成为生物学研究中不可缺少的帮手。*Nucleic Acids Research* 杂志连续七年在其每年第一期中详细介绍当年各种生物学数据库。在2002年出版的30卷第一期中详细地介绍了335种通用和专用数据库，包括它们详细的描述和网址。2001年，这个数字是281，而2000年是115。可见近几年来生物学数据库的增长非常的迅速。

有大量的数据并不表示有大量的知识，关键是如何从这些杂乱的数据中挖掘出我们需要的东西。现在，各种基因组工程都已经基本完成，获得数据过程的高峰时期已经过去。当今的生物信息学研究重点开始转向对数据的整理，注释并最终实现数据的应用，也就是知识。与此相应的是大型的基因蛋白序列数据库的增长减缓，并趋于数据的整合。而各种以基因蛋白序列为基础的，以各种生理生化功能或者是疾病为着眼点的专题数据库逐渐兴起。这类数据库不再着重于数据的大量积累，而是开始对已有的数据进行整理、归类和注释，按照各自的专题从大型的序列库以及文献库中搜集需要的信息加以整理分析，总结知识，探索规律。如蛋白相

互作用数据库、细菌脂蛋白数据库、免疫多肽数据库、蛋白信号肽等，这些专题数据库的特点是：

1. 主题明确、内容集中、种类多样。包括序列与结构、表达与鉴别、分子相互作用、代谢通路与信号通路、突变与分子病理等。

2. 数据冗余度低，利用价值高。一般有从事相关专题研究多年的专家参与数据筛选和审核，数据可信度更高，冗余度更小。

3. 规模较小，构建和维护成本相对较低。除少数由大型信息中心建设的数据库之外，多数专题数据库均由大学实验室或相关研究机构构建。

随着数据库的越来越专业，数据的越来越齐全，数据库不再仅仅是一个为科研提供信息和数据服务的成分，而是开始发展出一种不同于实验科学，以数据库和大量生物数据为基础，以计算机为工具的生物学研究方法，通过在计算机上模拟生物某些功能的生理生化过程来对生物学现象进行预测和分析。这种方法的好处是能大量的节约实验成本和时间。

二、医学生物信息学

就目前来说，通过几个基因组工程的数据积累，现在生物学数据已经比较丰富了，生物学数据库也发展得很快。很多的基因的功能已经被阐明，一些生理生化过程也研究得比较清楚，为生物学基础研究服务的数据库也越来越多。如前所述，以基因和蛋白的序列以及结构为主要数据对象的数据库在当今的生物学数据库中占有绝对多数比例。另一方面，医学由于长期的临床观察积累了大量的医学数据，如病人的病历，流行病的发病情况，药物的临床研究，药理、病理等研究形成了一个庞大的数据源。现在，这些数据也被搜集起来输入计算机中，以网络数据库的形式开始为医学研究人员和临床医生提供服务。然而相对于当今制药和医学研究来说这些数据库都有着一些缺点，以基因蛋白序列积累的生物学数据库偏重于对生物学基础研究数据的搜集，有着相当多的基因和蛋白的数据，而缺少病理、药理和临床方面的数据。医学数据库虽然都以疾病为着眼点，但是搜集的数据以临床观察和药理为着眼点，偏重于为大众提供健康方面的信息服务以及为临床大夫提供资料，却缺乏疾病的分子机制方面的数据。

由此我们提出医学生物信息学的概念。顾名思义，医学生物信息学简单地说就是为医学研究服务的生物信息学，致力于搜集、整理和研究以疾病为中心，以基因蛋白序列数据为基础，从宏观的单个疾病到与该疾病相关的分子生物学水平的相关数据。它是医学、生物学、计算机科学和信息科学等多学科交叉而形成的一门新兴学科，是以计算机和信息科学为工具，对各种医学和生物学的信息、资料、数据进行搜集、储存、整理、计算和分析，形成可再生的资源，并加以比较和应用的学科。其研究范围包括临床信息、医学教育信息、医学科研信息、医学文献信息、决策信息、实验信息、流行病学信息、老年康复信息、药品信息、医疗器械信息、医学图像信息、医学遗传工程信息及医药生物模型信息等医学生物信息学。医学生物信息数据库着眼点是建立以疾病为中心，贯穿病理、药理、基因、蛋白、调控等方面数据的数据库。除了为医学工作者提供更加全面细致的服务以外，进而在已经建立好的数据库的基础上尝试构建生物学模型。图 1-1-1 简单地表示了医学生物信息与生物信息和临床信息之间的关系。

经过对文献的筛选、整理和加工，根据特定的疾病从文献库中搜索疾病相关的基因和蛋白，再从各大型生物学数据库中寻找这些基因和蛋白的序列、功能、结构等方面的数据，将这些基因和蛋白数据按照相应疾病分类存放，最终形成一个疾病相关基因数据库。以新建立的心衰相关基因和蛋白数据库为例。首先，通过关键词在 PubMed，Medline，MedlineNew 等文献库中搜索。到 2001 年 7 月 11 日为止共搜集到与心衰、心肌肥厚和心脏发育缺陷相关的文献

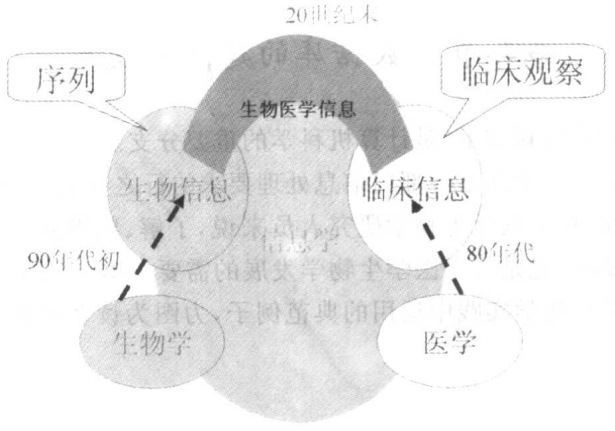

图 1-1-1　医学生物信息与生物信息和临床信息之间的关系

共 47 653 篇，从中挑选与基因和遗传相关的文献 1132 篇，结合 OMIM、Human MitBase、SWISS-CHANGE、EMBLCHANGE、MUTERS 等基因突变和遗传病数据库的查询结果归纳总结出心衰发病基因和蛋白共 407 条。在分别从 GenBank、SWISS-PROT、Protein Information Resource (PIR)、PDB、GDB 的数据库中搜集相关基因和蛋白的数据，按照 Gene Ontology 的方法对基因和蛋白进行分类，最终构成一个心力衰竭疾病为中心的包括相关基因蛋白数据的数据库。

人类对生物学的研究的最终目的之一就是为了治病救人。人体作为一个有机体，是数目庞大的有机分子协同作用并且彼此保持平衡的结果。人体一切的病变都可以归结到人体内生物活性分子的变化。要最终克服一种疾病必须从分子水平搞清楚疾病发生的机制。基因组时代以及后基因组时代分子生物学的飞速发展为病理学的研究提供了很好的基础。越来越多的疾病的分子机制被阐明，各种基因和蛋白的突变、序列、功能、代谢的数据中也包含了大量的人类各种疾病相关基因和蛋白的数据。但是，这些数据仍然是散见于各种文献之中没有进行很好的整理，没有一个很好的信息平台为制药和临床研究提供服务。医学生物信息学正是致力于以各种临床疾病为中心，整合当今基因蛋白数据库中的数据，试图建立一整套按照临床疾病分类的包括从临床数据到相关分子生物学数据的数据库。

参考文献

1. Andreas D Baxecanis. The Molecular Biology Database Collection. *Nucleic Acids Res* 2002; 30(1): 1-12
2. Andreas D Baxecanis. The Molecular Biology Database Collection: an updated compilation of biological database resources. *Nucleic Acids Res* 2001; 29(1): 1-10
3. Andreas D Baxecanis. The Molecular Biology Database Collection: an online compilation of relevant database resources. *Nucleic Acids Res* 2000; 28(1): 1-7
4. Loannis Xenaios, David Eisenberg. Protein interaction database. *Biotechnology* 2001; 12: 334-339
5. Madan Babu M, Sankaran K. DOLOP-database of bacterial lipoproteins. *Bioinformatics* 2002; 18(4): 641-643
6. Martin J Blythe, Lrini A Doychinova, Darren R Flower. JenPep: a database of quantitative functional peptide data for immunology. *Bioinformatics* 2002; 18(3): 434-439
7. Tao Wei, Mary O'Connell. TargetDB: a database of peptides targeting proteins to subcellular locations. *Bioinformatics* 1999; 15(9): 765-766
8. 张其鹏，张丹. 心力衰竭相关基因和蛋白质数据库的初构. 北京大学学报医学版 2002: 276-280

第二节　数据库的建立和发展

数据库是数据管理的最新技术，是计算机科学的重要分支。今天，信息资源已成为各个部门不可短缺的资源。建立一个满足各部门信息处理要求的行之有效的信息系统也成为其生存发展的重要条件。因此，对于医学生物学研究人员来说，了解、熟悉并将数据库技术运用到具体的实践工作中是必要的，也是当今医学生物学发展的需要。本节从数据库的基本概念出发，结合数据库在当今医学生物学实践中运用的典范例子，力图为读者展现一幅当今医学生物学数据库运用的概貌图。

一、基本概念

（一）数据

数据是数据库中存储的基本对象。数据在大多数人头脑中的第一个反应就是数字。其实数字仅仅是最简单的一种数据，是数据的一种传统的狭义上的理解。数据的种类其实很多，文字、图像、声音、商品的库存记录等等都是数据。数据的定义：描述事物的符号记录称为数据。

（二）数据库

数据库是数据的集合，表明了一些事实性的信息，是长期存放在计算机内的、有组织的、可共享的数据集合。数据库中的数据按一定的数据模型组织、描述和存储，具有较小冗余度、较高的数据独立性和易扩展性，并可为各种用户共享。数据库同知识库的区别在于：知识库不但有事实，而且还有一些知识领域的规则或法则、规律（定律或公式）、方法和理论知识及常规知识。数据库只允许授权的每个人或应用程序对数据（或信息）进行存取、更新和修改，而知识库允许应用程序利用一些知识规则进行推理，从而产生许多新的知识和信息。

（三）数据库系统

数据库系统是一个实际可运行的存储、维护和应用系统，是提供数据的软件系统，是存储介质、处理对象和管理系统的集合体。它通常由软件、数据库和数据库管理员组成。其软件主要包括操作系统、各种宿主语言实用程序以及数据库管理系统。数据库是依照某种数据模型组织起来并存放在二级存储器中的数据集合。这些数据为多个应用服务，独立于具体的应用程序。数据库由数据库管理系统统一管理，数据的插入、修改和检索均要通过数据库管理系统进行。数据库管理系统是一种系统软件，它的主要功能是维护数据库并有效地访问数据库中任意部分数据。对数据库的维护包括保持数据的完整性、一致性和安全性。数据库管理员负责创建、监控和维护整个数据库，使数据能被任何有权使用的人有效使用。数据库管理员一般是由业务水平较高、资历较深的人员担任。

数据库系统的个体含义是指一个具体的数据库管理系统软件和用它建立起来的数据库；它的学科含义是指研究、开发、建立、维护和应用数据库系统所涉及的理论、方法、技术所构成的科学。在这一含义下，数据库系统是软件研究领域的一个重要分支，常称为数据库领域。数据库研究跨越于计算机应用、系统软件和理论三个领域，其中应用促进新系统的研制开发，新系统带来新的理论研究，而理论研究又对前两个领域起着指导作用。数据库系统的出现是计算机应用的一个里程碑，它使得计算机应用从以科学计算为主转向以数据处理为主，并使计算机得以在各行各业乃至家庭中普遍使用。在它之前的文件系统虽然也能处理持久数据，但是文件系统不提供对任意部分数据的快速访问，而这对数据量不断增大的应用来说是至关重要

的。为了实现对任意部分数据的快速访问，就要研究许多优化技术。这些优化技术往往很复杂，是普通用户难以实现的，所以就由系统软件（数据库管理系统）来完成，而提供给用户的是简单易用的数据库语言。由于对数据库的操作都由数据库管理系统完成，所以数据库就可以独立于具体的应用程序而存在，从而数据库又可以为多个用户所共享。因此，数据的独立性和共享性是数据库系统的重要特征。数据共享节省了大量人力物力，为数据库系统的广泛应用奠定了基础。数据库系统的出现使得普通用户能够方便地将日常数据存入计算机并在需要的时候快速访问它们，从而使得计算机走出科研机构进入各行各业、进入家庭。

（四）数据库管理系统

数据库管理系统（DataBase Management System，DBMS）是一套软件。它可以完成如下功能：

（1）存储、恢复和修改数据；

（2）保持数据的一致性；

（3）解决数据更新的并发性；

（4）允许按一定的规则访问数据库中的数据。

（五）数据库都有哪些数据模型

总的来说，数据库有三种数据模型，它们分别是：

（1）层次模型：层次模型出现于20世纪60年代，它是“一个父亲多个孩子”型的，其数据结构只能是树。

（2）网状模型：网状模型出现于20世纪70年代，它是“多个父亲多个孩子”型的，其数据结构是任何形式的“图”。

（3）关系模型：关系模型出现于20世纪80年代，是自动建立起来的所有可能的关系，例如：一对多的关系。

二、医学生物学数据库的现状与发展

（一）为何要发展医学生物学数据库

在人类基因组计划及其后续实验工作的推动之下，各种类型的生物学数据库，尤其是分子生物学的数据库，如雨后春笋般涌现出来。如何高效地使用这些数据库，已经成为生命科学工作者最为关心的问题之一；如何建立并维护高质量的数据库，使之加速实验科学的研究进程，也是摆在生物信息学工作者面前亟待解决的课题。

正确理解数据库和实验科学发展之间的关系，充分认识医学生物学数据库发展的意义，是充分利用数据库资源和建设新型高质量数据库的基础。如果仅将数据库理解为一种存储、管理、调用实验数据的介质，那无疑是不够完整的。可以说，实验科学的发展决定了数据库技术的引入，数据库的发展又反过来促进实验科学研究进程的加快和研究模式的变革。

首先，实验科学飞速发展的结果与需求，决定了数据库技术的及时引入与蓬勃发展。一方面，越来越多的高通量自动化技术设备的引入，使得实验科学的数据呈现爆炸式增长。海量数据的存储管理，成为实验进程中的现实问题。人类基因组计划中测序进程的加速带来的数据爆炸，就是最好的实例。另一方面，生命科学的研究正日益呈现“系统化”的趋势。一个简单生物大分子的研究，往往涉及从基因、蛋白质到组织器官乃至整体水平的各个层次的工作。以一个酶的研究为例，它所涉及的信息包括：一级结构（氨基酸序列）、信号序列、折叠信息、三级结构、基因组与染色体定位、基因结构与表达调控、底物产物与辅助因子、该分子在代谢通路中所处的位置、该分子的亚细胞定位、细胞类型与组织分布、物种特异性、结构异常与疾病相关性、

酶分子的命名及相关文献信息等等。上述每一个方面的信息，往往都要用到或者依托于某个或者多个不同类型的数据库。由此可见，正是实验科学发展的自身特点和迫切需求，决定了医学生物学数据库的必然兴起。

其次，数据库的发展，为生命科学实验领域的迅速发展提供了保障，并日益深入地影响和改变着传统的生命科学研究手段。这一点，在蛋白质组学的研究中体现的尤为充分：高通量的2D电泳和质谱分析，大大加快了细胞内蛋白质分析的速度，得到了大量的电泳图谱和序列数据。但要处理加工这些数据，从中获得新的蛋白质的结构与功能信息，则完全需要依赖于数据库存储量的充分扩增和数据项的高效搜索比对。没有已知蛋白质的数据存储和对比，未知蛋白质的发现和进一步研究也就无从谈起。另一方面，有了充分的蛋白质数据储备，就为综合分析蛋白质之间的相互作用，阐明蛋白质的功能和作用规律，提供了新的思路和方法，也为未来信息技术与生命科学技术的更高层次的结合——生命活动的计算机模拟，打下数据基础。*Nature* 新近发表的对酿酒酵母(*Saccharomyces cerevisiae*)蛋白质的功能网络的分析，就是建立在强大的数据库储备基础上的信息学与实验科学手段相结合的典范。美日两国已经开展的虚拟细胞项目，后台也有强大的数据库作为支撑。

从以上两个方面不难看出，数据库的发展与实验科学的发展，互为因果，相辅相成，密不可分，有着不可替代的重要意义。数据库，既是未来生物信息学继续前进的基础，更是生命科学实验领域深入发展的基石。

（二）现有的生物医学数据库是何种发展水平

目前，生物医学数据库，可谓是林林总总、眼花缭乱，其数量和所涵盖的生物医学研究领域类别，都处于日新月异、飞速增长的阶段。为了解当前数据库的总体发展状况，笔者对目前因特网上比较有权威性的四个数据库索引网站所发布的最新版本进行了链接检查及分类统计。统计结果如下列诸表所示。

DBCAT (Catalog of Databases)，是法国生物信息研究中心与EBI的合作项目之一，旨在收集各种生物信息数据库的名称、内容、数据格式、联系地址、网址等详细信息，使用户对目前生物信息数据库有一个详尽的了解；DBCAT本身也是一个具有一定数据格式的数据库(表1-2-1)。

表1-2-1　DBCAT (2001.3)

数据库类型	数量
DNA	87
RNA	29
Protein	94
Genomic	58
Mapping	29
Protein Structure	18
Literature	43
Miscellaneous	153
Summary	511

(URL:http://www.infobiogen.fr/services/dbcat)

英国基因组图谱资源中心(Human Genome Mapping Resource Centere，HGMP)的GenomeWeb所收录的基因组数据库链接。除此之外，GenomeWeb还提供很多知名的专题数据