“十五”国家重点图书

现代生物技术丛书

生物信息学

——智能化算法及其应用

王翼飞　史定华　主编

- 生物信息学
- 智能化算法
- 序列联配与隐马氏模型
- 模体识别与神经网络
- 蛋白质折叠与遗传算法
- RNA结构预测与模拟退火
- 微阵列技术与统计推断
- 基因相互作用与贝叶斯网络

化学工业出版社
现代生物技术与医药科技出版中心

“十五”国家重点图书

现代生物技术丛书

生物信息学
——智能化算法及其应用

王翼飞　史定华　主编

化学工业出版社
现代生物技术与医药科技出版中心
·北京·

本书作为《现代生物技术丛书》的分册之一，旨在为从事生物信息学研究的学子们提供一个可操作的入门性介绍。

生物信息学是一门涵盖生物学、数学、化学、物理学、计算机科学等学科的年轻科学，也是近年来发展非常迅速的研究领域。目前，生物信息学研究工作者大都依据各自的知识背景采用擅长的数学方法，独门利器，庖丁解牛，从初等数学到高等数学，可说是“十八般武艺、各显神通”。本书独辟蹊径，以智能化算法为主线逐一介绍了隐马氏模型、神经网络、遗传算法、模拟退火算法、贝叶斯网络等算法，着重阐述了这些算法在生物信息学研究中的应用，力图探索破译生命奥秘的可行之径。书中介绍的各种算法和生物信息学课题都是笔者多年来实际研究过的工作，相关的论文也都已陆续发表。因此，从一定意义上说，本书是作者多年研究工作的整理和总结。

国内高校和科研院所生物和数学领域中从事生物信息学教学和研究的教师和学生，阅读本书，将会发现它是一本实用的教材和阅读方便的参考书。

图书在版编目（CIP）数据

生物信息学——智能化算法及其应用/王翼飞，
史定华主编. —北京：化学工业出版社，2006.5
（现代生物技术丛书）
“十五”国家重点图书
ISBN 7-5025-8619-9

Ⅰ.生… Ⅱ.①王… ②史… Ⅲ.生物信息论
Ⅳ.Q811.4

中国版本图书馆 CIP 数据核字（2006）第 040758 号

“十五”国家重点图书
现代生物技术丛书
生物信息学
——智能化算法及其应用
王翼飞 史定华 主编
责任编辑：叶 露 傅四周
文字编辑：周 倜
责任校对：陈 静
封面设计：关 飞

*

化 学 工 业 出 版 社
现代生物技术与医药科技出版中心 出版发行
（北京市朝阳区惠新里 3 号 邮政编码 100029）
购书咨询：(010)64982530
(010)64918013
购书传真：(010)64982630
http://www.cip.com.cn

*

新华书店北京发行所经销
大厂聚鑫印刷有限责任公司印刷
三河市万龙印装有限公司装订
开本 787mm×1092mm 1/16 印张 18¾ 彩插 2 字数 445 千字
2006 年 7 月第 1 版 2006 年 7 月北京第 1 次印刷
ISBN 7-5025-8619-9
定 价：35.00 元

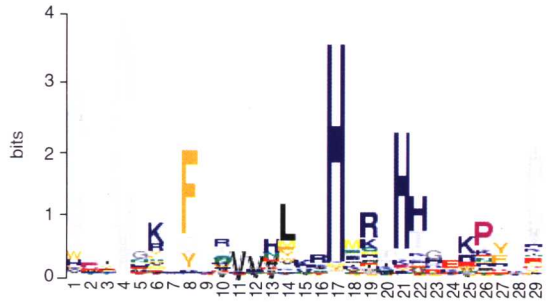

图 3–5
C2H2 锌指蛋白序列家族的标识图
（文见第82页）

迭代次数	1	2	3	4	5
主状态数	57	56	56	56	56
Viterbi得分	−1461.736	−919.818	−895.477	−892.103	−892.103

```
1clf          A Y K I A D S C V S  C - - G A C A S E C  P V N A I S Q G D S
1fca          A Y V I N E A C I S  C - - G A C E P E C  P V D A I S Q G G S
1blu          A L M I T D E C I N  C - - D V C E P E C  P N G A I S Q G D E
FER_BACSC     A Y V I T E P C I G  T K D A S C V E V C  P V D C I H E G E D
FER_BUTME     A Y K I T D E C I A  C - - G S C A D Q C  P V E A I S E G - S
              x x x x x x x x x x  x     x x x x x x x  x x x x x x x x x x

1clf          I F V I D A D T C I  D C G N - - - - - -  C A N V C P V G A P
1fca          R Y V I D A D T C I  D C G A - - - - - -  C A G V C P V D A P
1blu          T Y V I E P S L C T  E C V G H Y E T S Q  C V E V C P V D C I
FER_BACSC     Q Y Y I D P D V C I  D C G A - - - - - -  C E A V C P V S A I
FER_BUTME     I Y E I D E A L C T  D C G A - - - - - -  C A D Q C P V E A I
              x x x x x x x x x x  x x x x              x x x x x x x x x x

1clf          V Q E - -
1fca          V Q A - -
1blu          I K D P S
FER_BACSC     Y H E D F
FER_BUTME     V P E - D
              x x x   x
```

图 3–18
基于启发式方法 Viterbi 得分变化情况和多序列联配结果
（文见第100页）

迭代次数	1	2	3	4
主状态数	57	53	53	53
Viterbi得分	−1461.736	−914.910	−893.102	−893.102

```
1clf          A Y K I A D S C V S  C - - G A C A S E C  P V N A I S Q G D S
1fca          A Y V I N E A C I S  C - - G A C E P E C  P V D A I S Q G G S
1blu          A L M I T D E C I N  C - - D V C E P E C  P N G A I S Q G D E
FER_BACSC     A Y V I T E P C I G  T K D A S C V E V C  P V D C I H E G E D
FER_BUTME     A Y K I T D E C I A  C - - G S C A D Q C  P V E A I S E G - S
              x x x x x x x x x x  x     x x x x x x x  x x x x x x x x x x

1clf          I F V I D A D T C I  D C G N - - - - - -  C A N V C P V G A P
1fca          R Y V I D A D T C I  D C G A - - - - - -  C A G V C P V D A P
1blu          T Y V I E P S L C T  E C V G H Y E T S Q  C V E V C P V D C I
FER_BACSC     Q Y Y I D P D V C I  D C G A - - - - - -  C E A V C P V S A I
FER_BUTME     I Y E I D E A L C T  D C G A - - - - - -  C A D Q C P V E A I
              x x x x x x x x x x  x x x x              x x x x x x x x x x

1clf          - V Q - - E
1fca          - V Q - - A
1blu          I K D P - S
FER_BACSC     - Y H E D F
FER_BUTME     - V P E - D
                x x
```

（a）δ=0.1

迭代次数	1	2	3	4	5
主状态数	57	57	56	56	56
Viterbi得分	−1461.736	−921.418	−895.477	−892.103	−892.103

```
1clf          A Y K I A D S C V S  C - - G A C A S E C  P V N A I S Q G D S
1fca          A Y V I N E A C I S  C - - G A C E P E C  P V D A I S Q G G S
1blu          A L M I T D E C I N  C - - D V C E P E C  P N G A I S Q G D E
FER_BACSC     A Y V I T E P C I G  T K D A S C V E V C  P V D C I H E G E D
FER_BUTME     A Y K I T D E C I A  C - - G S C A D Q C  P V E A I S E G - S
              x x x x x x x x x x  x     x x x x x x x  x x x x x x x x x x

1clf          I F V I D A D T C I  D C G N - - - - - -  C A N V C P V G A P
1fca          R Y V I D A D T C I  D C G A - - - - - -  C A G V C P V D A P
1blu          T Y V I E P S L C T  E C V G H Y E T S Q  C V E V C P V D C I
FER_BACSC     Q Y Y I D P D V C I  D C G A - - - - - -  C E A V C P V S A I
FER_BUTME     I Y E I D E A L C T  D C G A - - - - - -  C A D Q C P V E A I
              x x x x x x x x x x  x x x x              x x x x x x x x x x

1clf          V Q E - -
1fca          V Q A - -
1blu          I K D P S
FER_BACSC     Y H E D F
FER_BUTME     V P E - D
              x x x   x
```

（b）δ=0.5

图 3–19
基于极大化后验算法的 Viterbi 得分变化情况和多序列联配结果
（文见第100页）

迭代次数	1	2	3	4	5	6
主状态数	57	56	55	56	55	55
Viterbi得分	−1461.736	−919.818	−894.226	−894.256	−890.834	−890.834

```
1clf        A Y K I A D S C V S  C - - G A C A S E C  P V N A I S Q G D S
1fca        A Y V I N E A C I S  C - - G A C E P E C  P V D A I S Q G G S
1blu        A L M I T D E C I N  C - - D V C E P E C  P N G A I S Q G D E
FER_BACSC   A Y V I T E P C I G  T K D A S C V E V C  P V D C I H E G E D
FER_BUTME   A Y K I T D E C I A  C - - G S C A D Q C  P V E A I S E G - S
            x x x x x x x x x x  x     x x x x x x x  x x x x x x x x x x

1clf        I F V I D A D T C I  D C G N - - - - - -  C A N V C P V G A P
1fca        R Y V I D A D T C I  D C G A - - - - - -  C A G V C P V D A P
1blu        T Y V I E P S L C T  E C V G H Y E T S Q  C V E V C P V D C I
FER_BACSC   Q Y Y I D P D V C I  D C G A - - - - - -  C E A V C P V S A I
FER_BUTME   I Y E I D E A L C T  D C G A - - - - - -  C A D Q C P V E A I
            x x x x x x x x x x  x x x x              x x x x x x x x x x

1clf        V Q E _ _
1fca        V Q A _ _
1blu        I K D P S
FER_BACSC   Y H E D F
FER_BUTME   V P E - D
            x x x
```

图 3–20　利用 BIC 准则自适应的 Viterbi 得分变化情况和多序列联配结果（文见第101页）

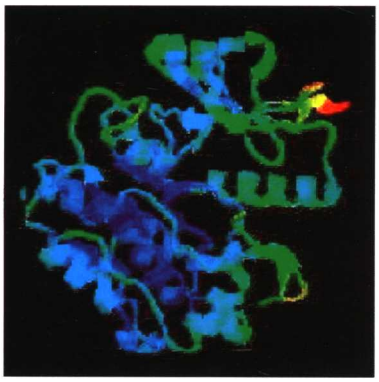

图 5–1
酪氨酸激酶的空间结构图形
（摘自http：// www. equn. com / folding / ）
（文见第138页）

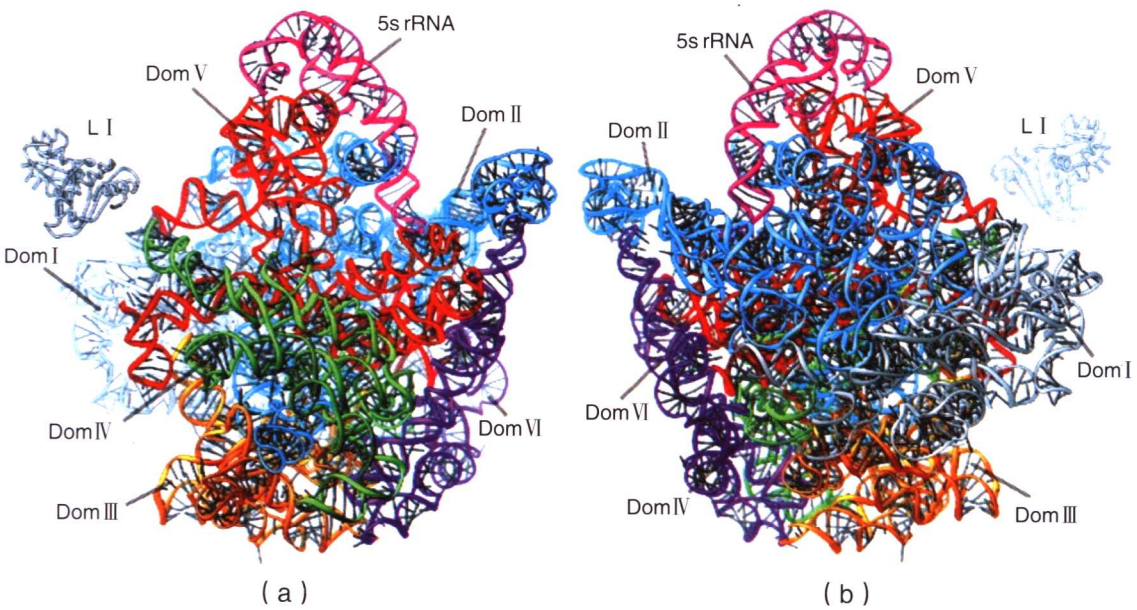

图 6–7　盐杆菌核糖体大亚基中的 rRNA（文见第176页）

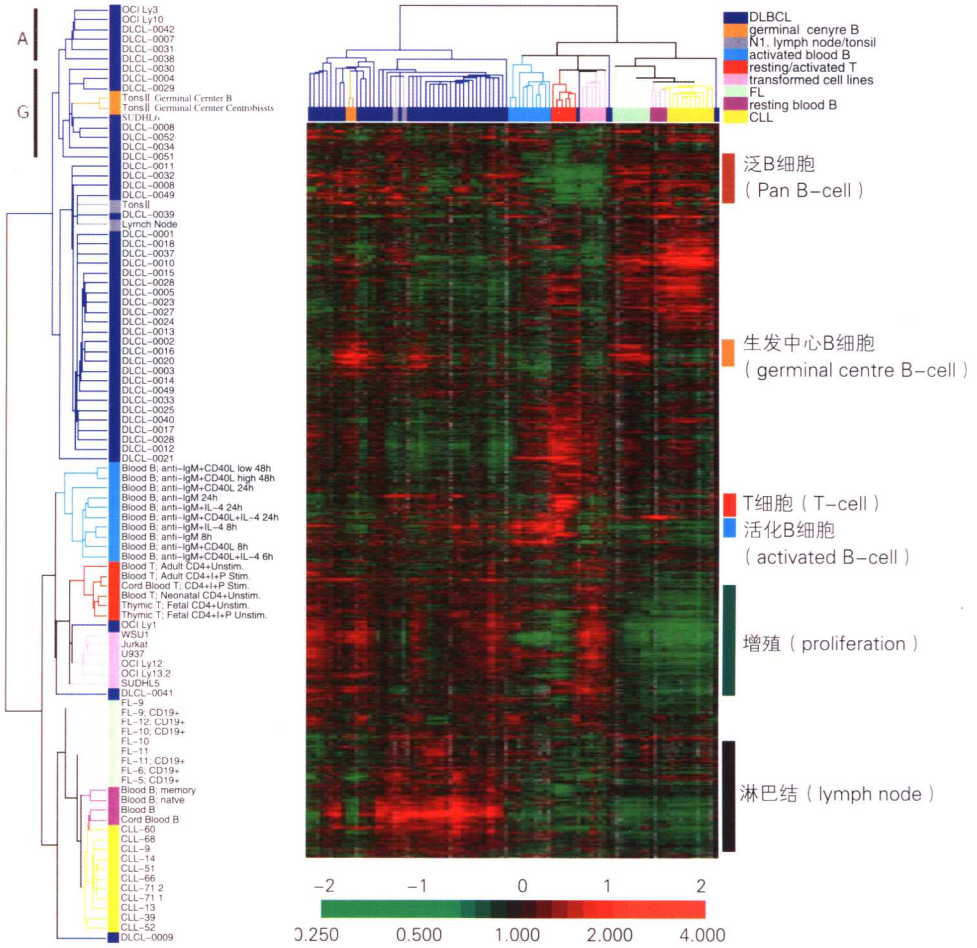

图 7-9　系统聚类法聚类基因表达数据（文见第223页）

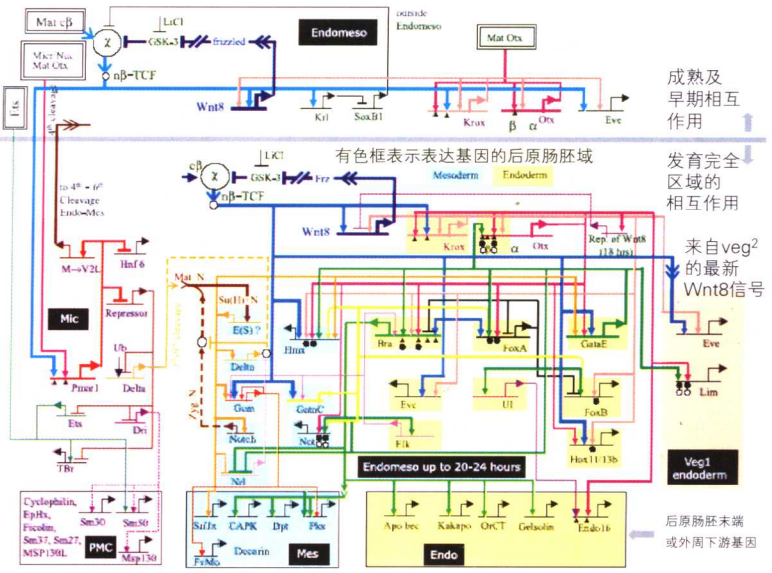

图 8-3　海胆内胚叶的基因调控网络（文见第252页）

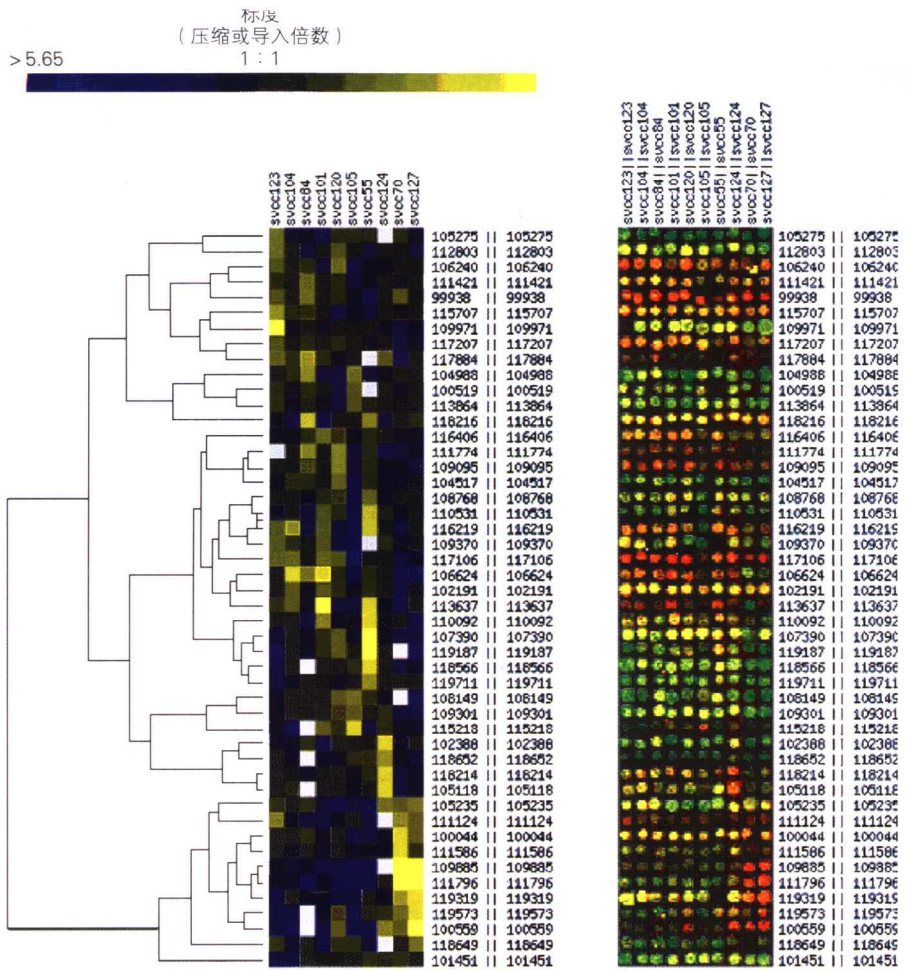

图 8-4　微阵列聚类分析示例（文见第254页）

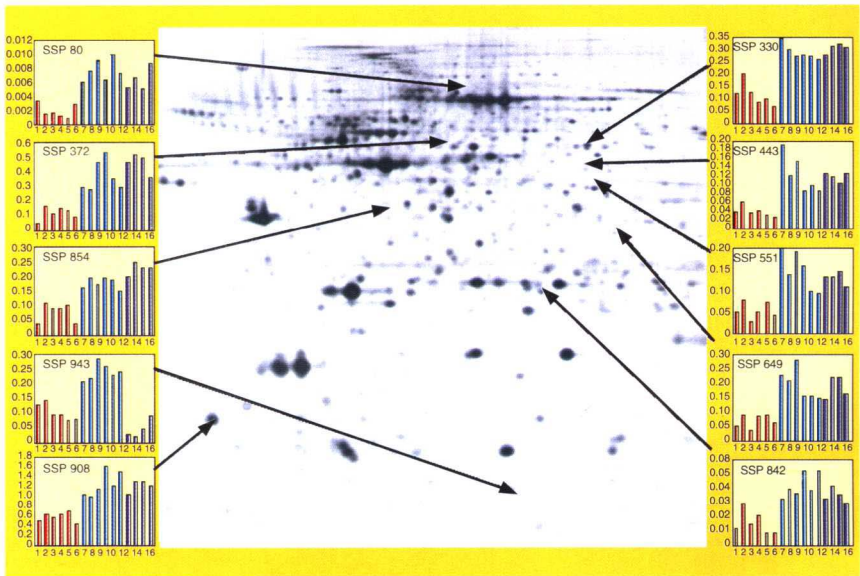

图 8-6　基于质谱数据库的计算机分析（文见第256页）

出版者的话

现代生物技术（生物工程）建立在分子生物学、分子遗传学、生物化学、微生物学、细胞学以及工程技术、计算机技术等基础之上，是21世纪最重要的技术和产业领域之一，正迅速改变着传统的产业格局与人们生活的方方面面。

化学工业出版社一直致力于生物技术类图书的出版工作。早在20世纪80年代末、90年代初，就出版了由我国著名的微生物学家焦瑞身先生组织编写的国内第一套《生物工程丛书》。这是一套普及性科技图书，共有8个分册：《遗传学基础》、《生物化学基础》、《微生物学基础》、《生物工程概论》、《生物化学工程》、《细胞工程》、《酶工程》、《微生物工程》。《生物工程丛书》一经出版，就受到了读者的广泛好评，对促进20世纪90年代我国生物技术的发展起到了积极的推动作用。

进入20世纪90年代后期，生物科学研究更加活跃，新成果层出不穷，并与许多学科交叉融合，涌现了许多新学科、新技术。为此，化学工业出版社于2000年组建了现代生物技术与医药科技出版中心，专门从事生物技术、生物科学及医药科技类图书的出版工作，其宗旨在于：传播生命科学、服务生物产业、促进医药发展。出版中心成立伊始，即着手《生物工程丛书》的修订工作，组成了以焦瑞身先生为首的编委会，在广泛调研、充分论证的基础上，顺应生物技术的发展潮流，对丛书重新设题，推陈出新，更名为《现代生物技术丛书》。

第一批《现代生物技术丛书》共组织了15个分册，其中14个分册已于2000～2006年陆续出版：《基因工程》、《微生物工程》、《酶工程》、《植物细胞工程》、《动物细胞工程》、《蛋白质工程》、《组织工程》、《生物技术与疾病诊断——兼论人类基因治疗》、《环境生物工程》、《生物制药技术》、《生物工程下游技术》（第二版）、《农业生物工程》（第二版）、《生物传感器》、《生物信息学——智能化算法及其应用》，所余分册《糖生物工程》也计划于近期出版。参与编撰《现代生物技术丛书》的专家有180多人，特别是焦瑞身先生尽管年事已高，仍欣然挂帅，多方联系和推荐作者，逐一审订各分册的提纲，并亲自主编凡100多万言的《微生物工程》。老一辈科学家鞠躬尽瘁的奉献精神与严谨务实的科学态度，深深地感染了新一代科研专家和出版者，激发了大家认真高效的工作热情，这是该丛书在较短时间内高质量出版的强大动力与工作基础。已出版的12个分册在首印后陆续重印，得到社会广泛好评，无疑是对众多编者辛勤笔耕的最好回馈！

鉴于现代生物技术日益丰富的内涵、较快的技术更新速度以及读者多样化的需求，化学工业出版社拟将《现代生物技术丛书》的出版之路不断延伸下去，分阶段地补充新技术、新内容，力争使丛书跟上生物技术本身的发展步伐，涵盖生物技术的方方面面。为此，近期将推出《现代生物技术丛书》第二批书目，或吸纳近年发展起来的新技术、交叉学科，或赋予

传统学科以新内涵，包括《生物芯片技术》、《生物化学工程》、《微生物基因组学》等。

作为出版者，我们衷心希望《现代生物技术丛书》能够更好地服务于读者，为我国生物技术乃至生命科学的快速发展作出应有的贡献，我们也将为此付出最大的努力。同时，欢迎广大读者就丛书的后续书目贡献良策，以及就书中存在的不足和问题随时与我们交换意见。请联系：life@cip.com.cn。

化学工业出版社
现代生物技术与医药科技出版中心
2006 年 4 月

序

建立在分子生物学、分子遗传学、生物化学、微生物学、细胞学以及化工、计算技术等基础之上的现代生物技术（生物工程），是20世纪后半期国际上突飞猛进的技术领域之一，它为人类保健、农牧业、食品工业、环境保护以及精细化工等产业的发展提供了前所未有的动力。展望新世纪，可以预料生物技术的前景更为光辉灿烂。本丛书将就该领域的研究动态逐个进行详细介绍，这里我们仅概述其突出进展与读者分享。鉴于各领域发展迅速和编者水平有限，丛书定有遗漏和不足之处，敬请读者指正。

一、基因组和后基因组学

人类基因组计划（HGP）正式启动于1990年，这是一个跨世纪、跨国界的最伟大的生命科学工程，经美国、英国、法国、德国、日本、中国6国的合作和努力，已于2001年完成全部序列测定。这一成就可以与原子弹计划和登月计划相媲美，它将对生命科学和人类健康产生巨大影响。应用各种技术，上千个与疾病相关的基因已被定位，并有近百个疾病基因被克隆。毫无疑问，这将为新药研究设计和疫苗制备提供依据，且已有多个物质进入临床试验。

与此同时，小家鼠、果蝇、线虫、拟南芥、水稻、啤酒酵母，以及多种真菌、细菌的基因组研究相继开展，其中拟南芥基因组的全序列测定业已完成。由于微生物的基因组远小于多细胞真核生物，且细菌和酵母基因中不存在内含子，因而便于分析，迄今已在酵母基因组中发现了一些与人类疾病基因同源的基因，研究这些基因在酵母中的生理功能，将有助于了解相关疾病的发病机理。

今天，一个崭新的领域——生物信息学迅速发展，它将基因的结构、蛋白质功能以及物种的进化在基因信息的基础上统一起来。这一学科的发展，对基因组和后基因组学研究及对人类健康和农业发展将产生深远的影响。

二、基因工程（重组DNA技术）

体外DNA重组技术始于1972年，首先在大肠杆菌中获得成功，继而扩展到其他微生物，生产出了多种新型发酵产品。美国批准上市的基因工程产品有人类胰岛素、人类生长因子、白介素、干扰素、牛型生长激素疫苗等，并不断有新的品种进入临床应用。重组微生物的应用，也为高等生物作为表达外源基因的宿主提供了技术和经验，如哺乳动物细胞株、昆虫细胞株、转基因动物、转基因植物，都有可能作为生产需要糖基化的重组蛋白质的宿主。

我国基因工程研究起步较晚，自1986年“863”计划实施以来，生物技术药物的研究和产业化获得迅猛发展，至1998年已有14种基因工程药物、3个基因工程疫苗和数十个重组诊断试剂投放市场。

三、转基因作物及其他农业生物工程

农业生物技术中最重要的是转基因作物（GMC）。近10年来GMC发展速度极快，1996～2001年全球GMC的种植面积增长了30倍。2000年达4420万公顷，比1999年增长11%，2001年又在2000年的基础上增长19%，达5260万公顷。GMC种植面积占相关作物全球种植面积的比例依次为：大豆46%、棉花20%、油菜11%、玉米7%。

我国 GMC 的种植面积在 13 个国家中居第四位。国产转基因 Bt 抗虫棉的育成和推广，开创了国内基因工程农业应用的成功范例，仅 2001 年种植面积就达 60 万公顷。抗虫棉的杀虫性强，农药用量可减少 70%～80%，既降低了用工成本，又保护了环境。

继获得第一代 GMC（抗除草剂、抗虫、抗病等）之后，第二代转基因作物已呼之欲出，重点是进一步改良作物品质，提高其营养水平（如“金稻米”等），或以植物作为生物反应器生产医疗保健产品（如口服疫苗等）。同时，针对旱、涝、盐碱、低温等恶劣自然环境，培育各类抗逆作物。

此外重组根瘤菌、重组联合固氮菌、抗病杀虫重组微生物的开发和应用也取得了明显的成效。

四、克隆动物及转基因动物

动物体细胞克隆技术的发展为生产蛋白质类药物、器官移植、挽救珍稀濒危动物以及培育优良品种等奠定了基础。有科学家用山羊胚胎的核转入去核未受精的卵母细胞，产生了克隆动物——Dolly 羊，成为科学上的重大突破，并在多种动物中得到重复。

转基因动物的成功引导了一种新型制药工业，即利用转基因山羊、绵羊和乳牛的乳汁来生产治疗人类疾病的蛋白质类药物。转基因动物发展的另一动向是克隆修饰的猪，为人体器官移植提供外源器官，以缓解临床上对人体器官的迫切需求。

体细胞克隆山羊在我国的上海市转基因研究中心及陕西的中国杨凌克隆动物基地都获得了成功。

五、细胞工程和组织工程

多年来我国植物组织培养和细胞工程研究在国际上是领先的。我国学者通过花药和花粉单细胞培养培育出烟草、水稻、小麦、大麦、油菜、甘蔗等作物的新品种、新品系，种植面积逾 100 万公顷。脱病毒快速繁殖的主要作物有香蕉、马铃薯、甘蔗、木薯、香草兰、草莓、柑橘、苹果、葡萄、花卉和观赏植物。紫草、三七等植物细胞已可在发酵罐中大量培养。我国的传统中药涉及 5000 种左右植物，细胞培养是中药资源开发的一个重要方面。

我国学者在动物细胞工程方面也做出了重要贡献。例如亲缘关系远近不同的鱼类可进行各种核质组合，在变种间、属间及科间都获得了具有独特性状的核质重组鱼。

动物发育工程中另一重大进展是干细胞株的建立，这已成为国际上研究的热点。干细胞是指未充分分化、但具有再生为各种组织器官和个体潜在功能的细胞。血液干细胞能够分化、生成整个血液系统，用造血干细胞移植来治疗白血病和一些遗传血液病，是医学界正在探索的课题。最近，以色列科学家首次从胚胎干细胞培养出人类心脏组织，它可以正常跳动，并且有新生心脏组织的电特性和机械特性。波兰科学家用脐血干细胞成功地培育出了脑细胞，有可能被用于帕金森病、脑震荡等疾病的治疗和脑部损伤的修复。美国科学家最近成功地将胚胎干细胞分化成人类骨髓中的造血先驱细胞，并进一步培养成红细胞、白细胞和血小板。这些结果预示着人类有可能获得取之不尽的血源。我国科学家已成功地将干细胞体外培养成胃和肠黏膜组织，这是继利用干细胞原位培养皮肤组织全能修复之后，人类再生组织器官方面的又一重大成果。

六、环境生物工程

我国是环境污染较严重的国家，环境生物工程在防治各种污染中将起重要作用。众所周

知，油轮海上倾油可引起大面积海域污染，国外虽采用“超级细菌”（含有多个降解烃类的质粒）进行海面浮油处理，但其效果尚有待改进。化学农药对土壤的污染虽可用具专一性降解能力的特种细菌处理，但作用也甚缓慢。相对而言，较为先进的方法是采用可被降解的生物农药。此外，河流、湖泊水域的污染防治，酸雨危害以及城市垃圾的处理等，也都是亟待解决的问题。

七、酶工程

酶工程是现代生物技术的重要组成部分，其特点是利用酶、含酶细胞器或细胞（微生物、植物、动物）作为生物催化剂来完成某些重要的化学反应。应用范围包括医药工业、食品工业、化学工业、诊断分析和生物传感器等。涉及的品种不少，诸如糖化酶、淀粉酶、洗涤用酶以及与β-内酰胺抗生素生产有关的青霉素酰化酶、7-ACA 酰化酶等，其市场需求、生产规模和产值均很可观，并已产生巨大的经济效益。随着酶的大量应用，各种酶反应器和固定化技术应运而生，更进一步地推动了酶工程的发展。

当代酶工程发展的趋势之一是寻找耐极端条件的酶，如耐高温、耐酸碱、耐盐等。这些酶存在于嗜高温、嗜酸碱、嗜高盐的细菌中。近年来对这些细菌的研究进展迅速，这将为酶工业提供源源不断的新型酶类。

八、新型能源和清洁能源的开拓

随着化石能源逐年减少，再生能源的研制开发已备受国际关注。虽然我国石油和煤炭储量丰富，但从长远考虑，还需对这一课题予以重视。展望将来，新型能源，特别是清洁能源的开发很有必要。

氢气是无污染的清洁能源，燃烧后不产生二氧化碳、硫、氮氧化物等有害物质，国外的燃氢汽车已研制成功。产氢的微生物甚多，值得重视的是光合细菌，该菌可利用工业废水产氢，同时具有农用肥效的作用。

巴西和美国是燃料乙醇生产技术和商业应用比较成熟的国家。作物秸秆、废报纸等生物材料是生产再生能源的最廉价原料，所生产的燃料乙醇成本可低到每加仑 1.10 美元，虽然仍高于每加仑 0.80～0.90 美元的汽油批发价，但随着技术的改进，生产成本将会逐步降低。

九、新型生物传感器的研制

要研制新型生物传感器，需要新型的酶和生物材料，这些酶需能耐高温、酸、碱或低温。已发现的这类特殊生物材料有嗜盐细菌的紫膜，这是一种光敏材料，可转化光子为 ATP。另一个例子是磁细菌细胞中的微小磁石（Fe_3O_4），对细胞起导航作用。当代正竞相研制 DNA 芯片，以色列学者已用其建成简单的计算机。

生物传感器应用范围广泛，包括临床检测、免疫反应、反应罐过程检测、环保毒物检测等，不胜枚举。

十、生化工程

生化工程包括发酵工艺、过程检测与控制、反应模型建立、反应器的设计和应用，以及产品提取纯化、包装在内的下游加工工艺等方面，这是生物技术产业化的最后重要过程。

本丛书以应用生物技术为主，包括必要的基础知识和前景展望。丛书第一批包括 15 个分册，即《基因工程》、《蛋白质工程》、《酶工程》、《生物信息学——智能化算法及其应用》、《植物细胞工程》、《动物细胞工程》、《微生物工程》、《生物制药技术》、《生物传感器》、《环

境生物工程》、《农业生物工程》（第二版）、《糖生物工程》、《生物技术与疾病诊断——兼论人类基因治疗》、《组织工程》、《生物工程下游技术》（第二版）。

每册均由工作在第一线的专家撰写，概要阐述了国内外生物技术的进展和趋势。期望本丛书的出版能够对推动我国生物技术的研究开发及产业化作出微薄的贡献。

编者衷心寄语青年朋友，认识生物技术的光辉前景，祝愿你们以聪明才智为我国的生物技术作出创新贡献。

2002 年 1 月

前　言

历时 10 年的人类基因组计划，在人类探索生命奥秘的征途上迈出了坚实的一步。它不仅产出了海量的生物学数据，翻开了地球生命的“天书”，而且孕育了一门崭新的交叉学科——解读天书的“生物信息学”。这是一门涵盖生物学、数学、物理学、化学和计算机科学等众多学科的新兴集成学科。随着我国正式加入人类基因组计划的国际协作，生物信息学研究也在我国引起了广泛的重视。现在，相关的研究不仅是遍地开花，而且已经是百花争艳了。许多高等院校和研究机构都为本科生和研究生开设了生物信息学课程，同时也取得了许多研究成果。虽然相关的书籍也出了不少，但对生物信息学研究中用到的数学技术却很少有实质性的剖析。其实，离开了数学理论的支撑，生物信息学的研究是走不了多远的。以数学作为工具而言，研究工作者大都依据各自的知识背景采用擅长的数学方法，独门利器，庖丁解牛，从初等的数学到高等的数学，可说是“十八般武艺各显神通”。现在要想在一本书上汇聚所有的方法，并梳理出头绪来已是难以实现的愿望了。生命蕴涵着智能，生命演化出智能，用智能化算法破译生命的奥秘不失为一条可行之径，虽然崎岖，尚可攀登。多年来，我们将各种智能化算法作为工具开展了生物信息学的研究，有了一些心得和体会。在这本书里，我们就以智能化算法为主线逐一介绍各种算法及其在生物信息学研究中的应用，旨在为有志于生物信息学研究的年轻学子提供一个具有可操作性的入门介绍，使他们可以较快地开展实际的研究工作。生物信息学是一门年轻的学科，也是一门正在急速发展的学科，更是一门青年人可以大显身手的学科。如果将生物信息学比作大海的话，那么本书介绍的方法和问题仅是一束小小的、毫不起眼的浪花而已。我们抛砖引玉，期待更多的有识之士投身于生物信息学的研究。

本书介绍的各种算法和生物信息学课题都是我们多年来实际研究过的，相关的论文也都已陆续发表。因此，本书从某种意义上可以说是一份工作汇报和小结。但限于我们的水平，无论是对算法实质的理解和阐述，还是对相关生物学问题的归纳和处理，都难免会有缺陷，甚至是错误，我们希望读者不吝批评指正，如能将书中的错误和不妥之处用电子邮件方式（yifei_wang@staff.shu.edu.cn）通知我们，将不胜感激。

王翼飞教授和史定华教授拟定了全书的框架和提纲，并且撰写了部分章节。上海大学数学系生物信息学实验室的研究生顾燕红、倪红春、刘海军、徐东、朱伟、忻健、陆巍、邵建林、雷耀山、方茜、刘翔、周晖杰、王斌滨、孟宪花、于彬、刘阳、张冬宁、万虎、蔡传政、张家军、刘祥、沈称意、李冯等参加了本书的研讨和撰稿。全书由王翼飞教授和史定华教授统稿并校订。

中国科学院上海生命科学院的丁达夫研究员和李亦学研究员以及化学工业出版社对本书的编写都给予了热情的支持和帮助，他们有价值的建议帮助我们克服了编写中出现的种种困难。在此，谨向他们表示由衷的敬意和感谢！

王翼飞　史定华

2005 年 12 月于上海大学数学系

目　　录