国　外　药　学　专　著　译　丛

生物信息学

——从基因组到药物

[德] T. 伦盖威尔　主编
Thomas Lengauer
郑珩　王非　译

Bioinformatics

— From Genomes to Drugs

化学工业出版社
现代生物技术与医药科技出版中心

国 外 药 学 专 著 译 丛

生物信息学

——从基因组到药物

[德] T. 伦盖威尔 主编
Thomas Lengauer
郑珩 王非 译

Bioinformatics

— From Genomes to Drugs

化学工业出版社
现代生物技术与医药科技出版中心
·北京·

图书在版编目（CIP）数据

生物信息学——从基因组到药物/[德] 伦盖威尔（Lengauer，T.）主编；郑珩，王非译. —北京：化学工业出版社，2006
书名原文：Bioinformatics—From Genomes to Drugs
（国外药学专著译丛）
ISBN 7-5025-8479-X

Ⅰ. 生… Ⅱ. ①伦…②郑… Ⅲ. 生物信息论 Ⅳ. Q811.4

中国版本图书馆 CIP 数据核字（2006）第 027397 号

国外药学专著译丛
生物信息学
——从基因组到药物
[德] T. 伦盖威尔 主编
郑 珩 王 非 译
责任编辑：杨燕玲 余晓捷
责任校对：洪雅姝
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
北京云浩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 720mm×1000mm 1/16 印张 28¼ 彩插 5 字数 557 千字
2006 年 6 月第 1 版 2006 年 6 月北京第 1 次印刷
ISBN 7-5025-8479-X
定 价：68.00 元

序

药物与广大人民群众的生命健康息息相关，药学科学的发展和医药产业的进步对防病治病、保障人民健康具有举足轻重的作用。伴随着生命科学和信息技术的飞速发展，世界药学科技取得了一系列重大突破，与此同时，我国的药学科技和生产水平也有了长足进步。但我们也应看到，与世界发达国家相比，我国在药学科技、生产、管理等诸多领域还有较大差距，因此，及时、准确、全面地了解国际药学科技的最新成果和管理经验，对加快我国新药研发水平的提高和医药产业的进步都有十分重要的现实意义。

《国外药学专著译丛》正是基于上述原因由化学工业出版社倡议组织出版的。该丛书立足国内需求，瞄准国外药学科技前沿和医药生产中的先进技术，所选择的内容大多是科研、生产中迫切需要解决和提高的关键问题，针对性和指导性强。因此该丛书具有重要的参考价值。

为保证引进图书的水平和翻译质量，化学工业出版社聘请了近 30 位国内药学各专业领域的专家、教授成立了“国外药学专著引进顾问委员会”，旨在推荐、评阅引进图书，推荐译者或亲自组织翻译工作。专家、教授们丰富的学识和严谨的作风对保证该译丛的质量起到了重要的作用。

药学科学的发展日新月异，本译丛也将追踪药学科学的发展不断推出新的分册。相信这套译丛将对提升我国的药学科技和医药生产水平起到促进作用。

2005 年 10 月

前　言

“计算生物学”和“生物信息学”是近年来信息技术和生物学相结合并飞速发展起来的一个交叉领域的术语。这个领域定位于这两门科学与技术学科的交界处，我们认为这两门学科即便对当代科学的创新所起的不是主导作用，也是具有重要意义的作用。在英语中，计算生物学主要指这个领域的科学部分，而生物信息学则主要阐明了基础部分。在其他一些语言中（如德语）生物信息学涵盖了这个领域的各个方面。

这领域的目的是提供计算机方法来处理和解释分子生物学中由各种基因组测序计划和其他新的实验技术揭示的大量的基因组数据。这个领域向我们这个时代提出了一个巨大挑战。由于我们目前还无法在器官甚至分子水平对生物系统有很深入的了解，有大量的基础研究需要进行。同时，因为编码许多生物信息的基因组数据尚未揭示，而解码是可以取得激动人心的科学和商业成功的根本，所以人们迫切要求这个领域能给出解决方案。在以致力于人类基因组序列测序为标志的前基因组时代结束后，我们正在进入致力于收获隐藏在基因组序列中的各种成果的后基因组时代。从宣布进行人类基因组测序开始到完成，前基因组时代总共持续了不到 15 年的时间，与之相比，后基因组时代预计会经历更长甚至可能会是几代人的时间。

虽然本书会包括这个领域的许多基础性和普遍性的内容，但其主旨是指出生物信息学能够开拓新药设计的前景。而正是生物信息学在制药中的应用成为人们对生物信息学产生广泛兴趣的最强劲动力。

在这种背景下，本书希望面向那些具有广泛知识面的读者，成为其进入生物信息学领域的入门书。生物学家、生物化学家、药理学家、药剂师和医生可以从中得到基本并且实用的基于计算机处理和解释基因组数据的知识。特别指出，本书的许多章节提到的生物信息学软件和数据资源是可以从国际互联网上得到的（通常是免费的），而且本书也尝试着对这些资源进行分类和比较。对于计算机科学家和数学家来说，这本书有对生物学基本情况的介绍，还包含一些必要的文字来理解计算机构建复杂的生物化学和生物分子模型时遇到的难题和困惑。

生物信息学是一个快速进步的领域。实验技术和计算机技术正处于蓬勃的发展阶段，本书是对当今这个领域所处阶段的简明描述。

我感谢为本书问世而做出贡献的很多人。Hugo Kubinyi 第一个建议我编写这

本书，而自从那时起，他便恰到好处地给予压力和鼓励促使本书编写完成。Raimund Mannhold 和 Henk Timmerman 作为丛书的另外两位编委支持这个计划。最重要的是，我要感谢各章节的作者，他们在这个领域快速发展的时期内花费大量时间编写好这些经过慎重考虑后产生的章节。Gudrun Walter 和 Frank Weinreich 为本书的出版做出了出色的工作，我也要感谢他们。最后我要深深地感谢我的妻子 Sybille 和我的孩子 Sara 及 Nico，在本书编写过程中他们不得不克服没有我在时他们所遇到的困难。

Thomas Lengauer

原著编写人员

Prof. Ron D. Appel
Swiss Institute of Bioinformatics
Proteome Informatics Group
CMU-1, rue Michel Servet
1211 Geneva 4
Switzerland
ron. appel@isb-sib. ch

Dr. Amos Bairoch
Swiss Institute of Bioinformatics
SWISS-PROT group
CMU-1, rue Michel Servet
1211 Geneva 4
Switzerland
Amos. Bairoch@isb-sib. ch

Dr. Pierre-Alain Binz
Swiss Institute of Bioinformatics
Proteome Informatics group
CMU-1, rue Michel Servet
1211 Geneva 4
Switzerland
Pierre-Alain. Binz@isb-sib. ch

Dr. Christopher S. Carlson
University of Washington
Department of Molecular Biotechnology
Box 357730
Seattle , WA 98195
USA
peterpan@mbt. washington. edu

Prof. Roland L. Dunbrack, Jr.
Institute for Cancer Research
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, PA 19111
USA
rl _ dunbrack@fccc. edu

Dr. Thure Etzold
Lion Bioscience Ltd.
Sheraton House, Castle Business Park
Cambridge CB3 OAX
United Kingdom
etzold@lionbio. uk. com

Dr. Stefanie Fúhrman
Incyte Pharmaceuticals, Inc.
3174 Porter Dr.
Palo Alto CA 94304
USA
sfuhrman@incyte. com

Dr. Elisabeth Gasteiger
Swiss Institute of Bioinformatics
SWISS-PROT group
CMU-1, rue Michel Servet
1211 Geneva 4
Switzerland
Elisabeth. Gasteiger@isb-sib. ch

Dr. David P. Hansen
Lion Bioscience Ltd.
Sheraton House, Castle Business Park
Cambridge CB3 0AX
United Kingdom
David. Hansen@uk. lionbioscience. com

Prof. Dr. Denis F. Hochstrasser
Laboratoire Central de Chimie Clinique
Hôpital Cantonal Universitaire
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland
Denis. Hochstrasser@dim. hcuge. ch

Prof. Xiaoqiu Huang
Department of Computer Science
Iowa State University
226 Atanasoff Hall
Ames, IA 50011
USA
xghuang@cs. iastate. edu

Prof. Gerhard Klebe
Philipps-Universität Marburg
Institut für Pharmazeutische Chemie
Marbacher Weg 6
35032 Marburg
Germany
klebe@mailer. uni-marburg. de

Prof. Thomas Lengauer
Fraunhofer Institute for Algorithms and Scientific Computing
Schloss Birlinghoven
53754 Sankt Augustin
Germany
(present address:
Max-Planck-Institute for Computer Science
Stuhlsatzenhausweg 85
66123 Saarbrücken
Germany
lengauer@mpi-sb. mpg. de)

Dr. Shoudan Liang
Incyte Pharmaceuticals, Inc.
3174 Porter Dr.
Palo Alto, CA 94304
USA
sliang@incyte. com

Dr. Gidon Moont
Imperial Cancer Research Fund
Biomolecular Modelling Laboratory
44 Lincoln's Inn Fields
London WC2A 3PX
United Kingdom
moont@icrf. icnet. uk

Prof. Deborah Nickerson
University of Washington
Department of Molecular Biotechnology
Box 357730
Seattle, WA 98195
USA
debnick@u. washington. edu

Dr. Matthias Rarey
Fraunhofer Institute for Algorithms and Scientific Computing
Schloss Birlinghoven
53754 Sankt Augustin
Germany
matthias. rarey@gmd. de

Dr. Mark J. Rieder
University of Washington
Department of Molecular Biotechnology
Box 357730
Seattle, WA 98195
USA
mrieder@uwashington. edu

Dr. Jean-Charles Sanchez
Laboratoire Central de Chimie Clinique
Hôpital Cantonal Universitaire
24, rue Micheli-du-Crest
1211 Genève 14
Switzerland
Jean-Charles. Sanchez@dim. hcuge. ch

Victor Solovyev
EOS Biotechnology
225A Gateway Boulevard
South San Francisco, CA 94080
USA
solovyev@eosbiotech. com

Dr. Roland Somogyi
Incyte Pharmaceuticals, Inc.
3174 Porter Dr.
Palo Alto, CA 94304
USA
rsomogyi@incyte. com

Dr. Martin Stahl
Pharmaceuticals Division
F. Hoffmann-La Roche AG
4070 Basel
Switzerland

Dr. Michael J. E. Sternberg
Imperial Cancer Research Fund
Biomolecular Modelling Laboratory
44 Lincoln's Inn Fields
London WC2A 3PX
United Kingdom
m. sternberg@icrf. icnet. uk

Dr. Martin Vingron
Max-Planck-Institute of Molecular
Genetics
Ihnestraße 73
14195 Berlin
Germany
vingron@molgen. mpg. deGermany

Dr. Xiling Wen
Incyte Pharmaceuticals, Inc.
3174 Porter Dr.
Palo Alto, CA 94304
USA
xwen@incyte. com

Prof. Dr. Ralf Zimmer
Ludwig-Maximilians-Universität München
Institut für Informatik
Theresienstraße 39
80333 München
Germany
zimmer@bio. informatik. uni-muenchen. de

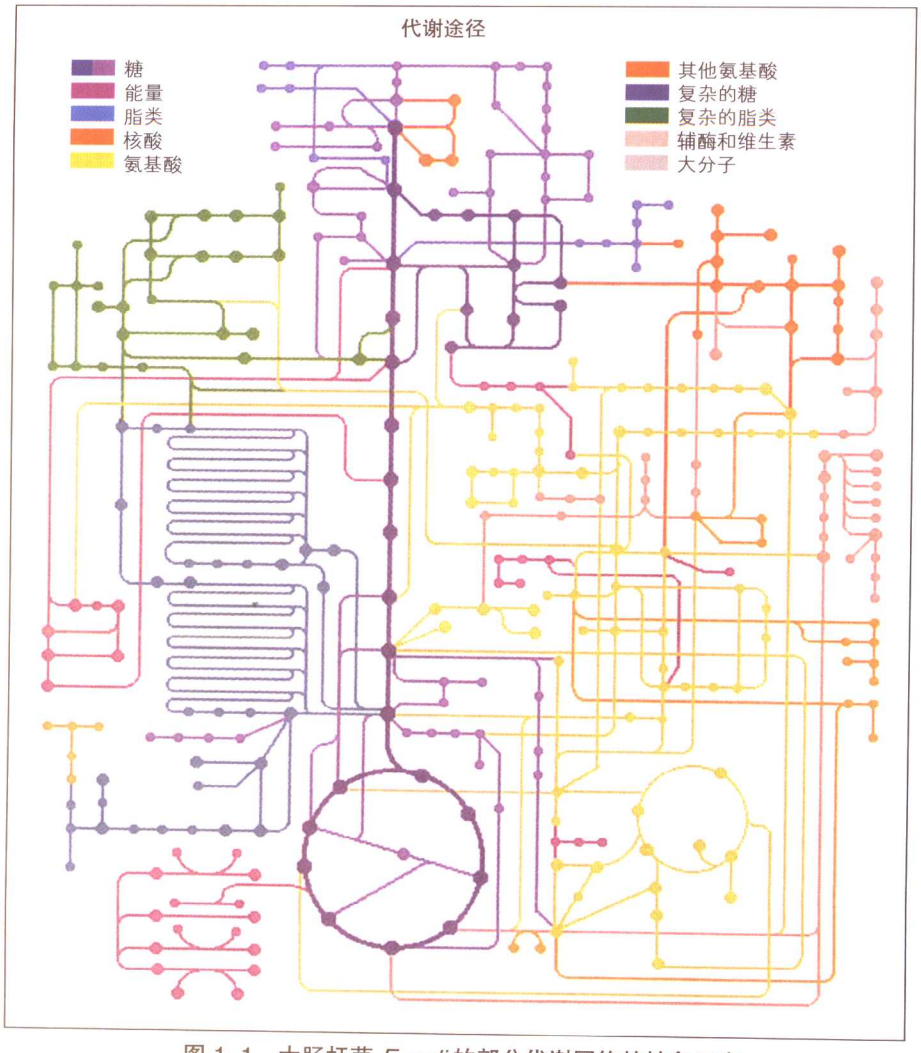

图 1–1　大肠杆菌 *E. coli* 的部分代谢网络的抽象示意
（源自http：//www.genome.ad.jp/kegg/kegg.html）

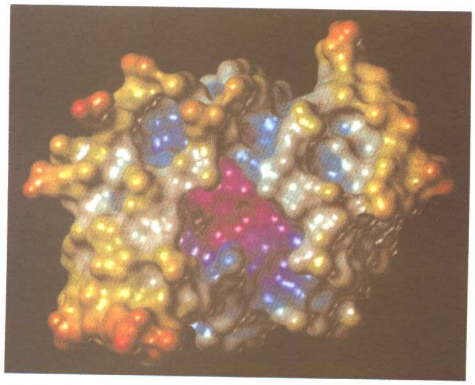

图 1–3　二氢叶酸还原酶分子表面的三维结构

图 1–4　结合了 DHF（绿色）和 NADPH（红色）的 DHFR（灰色）复合物

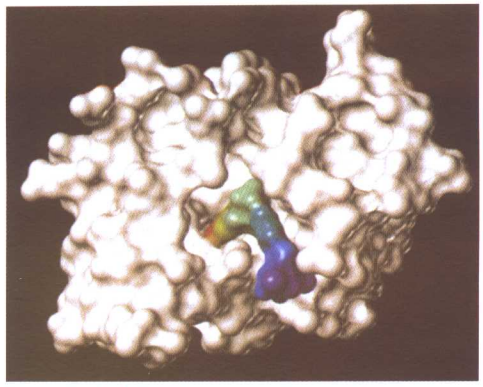

图 1–5　甲氨蝶呤（依据表面电势着色，参见图1–3）及结合的 DHFR（灰色）

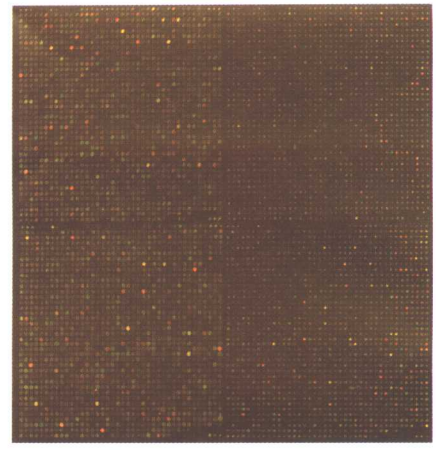

图 1–6　DNA芯片
（摘自http：//www.cmgm.stanford.edu/pbrown/explore/）

图 5–3
BACE 家族成员和人胃蛋白酶

```
                                                  * *
SecStr      : EEEE EEE   EEEEEEEE     EEEEEEEE      EEEE
1pso      1:VDEQPLENYLDMEYFGTIGIGTPAQDFTVVFDTGSSNLWVPSVYCSSLACTNHNRFNPED: 60
hbace1   62:EMVDNLRGKSGQGYYVEMTVGSPPQTLNILVDTGSSNFAVGAAPH----PFLHRYYQRQL:117
mbace1   62:EMVDNLRGKSGQGYYVEMTVGSPPQTLNILVDTGSSNFAVGAAPH----PFLHRYYQRQL:117
rbace1   62:EMVDNLRGKSGQGYYVEMTVGSPPQTLNILVDTGSSNFAVGAAPH----PFLHRYYQRQL:117
bbace1      :                      LNILVDTGSSNFAVGAAPH   --  PFLHRYYQRQL:
sbace       :                                                  YYDSEK:
zbace       :DMINNLKGDSGRGYYMQMIIGTPGQTLNILVDTGSSNFAVAAAAH----PYITHYFNRAL:
hbace2   79:AMVDNLQGDSGRGYYLEMLIGTPPQKLQILVDTGSSNFAVAGTPH----SYIDTYFDTER:134
mbace2   75:AMVDNLQGDSGRGYYLEMLIGTPPQKVQILVDTGSSNFAVAGAPH----SYIDTYFDSES:130

                                  **                                  *
SecStr      :    EEEEE EEEEE    EEEEEEEEEEEEE       EEEEEEEEEEEEE   HHHHH
1pso     61:SSTYQSTSE-TVSITYGTGSMTGILGYDTVQVG----GISDTNQIFGLSETEPGSFLYYAP:116
hbace1  118:SSTYRDLRK-GVYVPYTQGKWEGELGTDLVSIPH-GPNVTV-RANIAAITESDKFFINGSN:175
mbace1  118:SSTYRDLRK-GVYVPYTQGKWEGELGTDLVSIPH-GPNVTV-RANIAAITESDKFFINGSN:175
rbace1  118:SSTYRDLRK-SVYVPYTQGKWEGELGTDLVSIPH-GPNVTV-RANIAAITESDKFFINGSN:175
bbace1      :SSTYRDLRK-GVYVPYTQGKWEGELGTDLVSIPH-GPNVTV-RANIAAITESDKF------:
sbace       :SSSVINSGIPDVDIEYTEGFWKGPLVTDLVSIPEAGLTEQV-RVDIVKITSSKKFFINGSG:
zbace       :SSTYQSTER-AVAVKYTQGEWEGELGTDLITIP
hbace2  135:SSTYRSKGF-DVTVKYTQGSWTGFVGEDLVTIPK-GFNTSF-LVNIATIFESENFFLPGIK:192
mbace2  131:SSTYHSKGF-DVTVKYTQGSWTGFVGEDLVTIPK-GFNSSF-LVNIATIFESENFFLPGIK:188

                  *
SecStr      :  EEEE              HHHHHHHH      EEEEE
1pso     117:FDGILGLAYPSISS--S-GATPVFDNIWNQGLVSQDLFSVYLSADD-----------QSGS:163
hbace1   176:WEGILGLAYAEIARPDD-SLEPFFDSLVKQTHVP-NLFSLQLCGAGFPLNQSEVLASVGG:233
mbace1   176:WEGILGLAYAEIARPDD-SLEPFFDSLVKQTHIP-NIFSLQLCGAGFPLNQTEALASVGG:233
rbace1   176:WEGILGLAYAEIARPDD-SLEPFFDSLVKQTHIP-NIFSLQLCGAGFPLNQTEALASVGG:233
sbace       :WQGIIGLGYDELVRPNNPKVKSFMTSVIENTSVR-NVFSIQLCAA----NTMNFSDVTTG:
hbace2   193:WNGILGLAYATLAKPSS-SLETFFDSLVTQANIP-NVFSMQMCGAGLPVAGS---GTNGG:247
mbace2   189:WNGILGLAYAALAKPSS-SLETFFDSLVAQAKIP-DIFSMQMCGAGLPVAGS---GTNGG:243

                                  *                                   *
SecStr      :EEEE              EEEE        EEEEEEEEEE EEEE      EEEEEEEE
1pso     164:VVIFGGIDSSYYTGSLNWVPVTVEGYWQITVDSITMNGEAIAC----AEGCQAIVDTGTS:219
hbace1   234:SMIIGGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTT:293
mbace1   234:SMIIGGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTT:293
rbace1   234:SMIIGGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTT:293
bbace1   143:----IGGIDHSLYMGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTT:200
sbace       :SLVFGDYDRT--DGTIFRTRIVHEWYYEVIVLGMKV  *       CREFNNDKSIVDSGTT:
hbace2   248:SLVLGGIEPSLYKGDIWYTPIKEEWYYQIEILKLEIGGQSLNLDCREYNADKAIVDSGTT:307
mbace2   244:SLVLGGIEPSLYKGDIWYTPIKEEWYYQIEILKLEIGGQNLNLDCREYNADKAIVDSGTT:303
rbace2      :                                ILKLEIGGQSLNLDCREYNADKAIVDSGTT:

                *
SecStr      : EEE HHHHHHHHHH  EE            EE EEEEEEE   EEEEE
1pso     220:LLTGPTSPIANIQSDIGASENSDGD-------MVVSCSAISSLPDIVFTIN--------:260
hbace1   294:NLRLPKKVFEAAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEV:352
mbace1   294:NLRLPKKVFEAAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEV:352
rbace1   294:NLRLPKKVFEAAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEV:352
bbace1   201:NLRLPKKVFEAAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEV:259
sbace       :NLHLPEKVFN                *
hbace2   308:LLRLPQKVFDAVVEAVARASLIPEFSDGFWTGSQLACWTNSETPWSYFPKISIYLRDEN:366
mbace2   304:LLRLPQKVFDAVVEAVARTSLIPEFSDGFWTGAQLACWTNSETPWAYFPKISIYLRDEN:362
             LRLPQKVFDAVVEAVARTSLIPEFSDGFWTGAQLACWTNSETPWAYFPKISIYLRDEN:
```

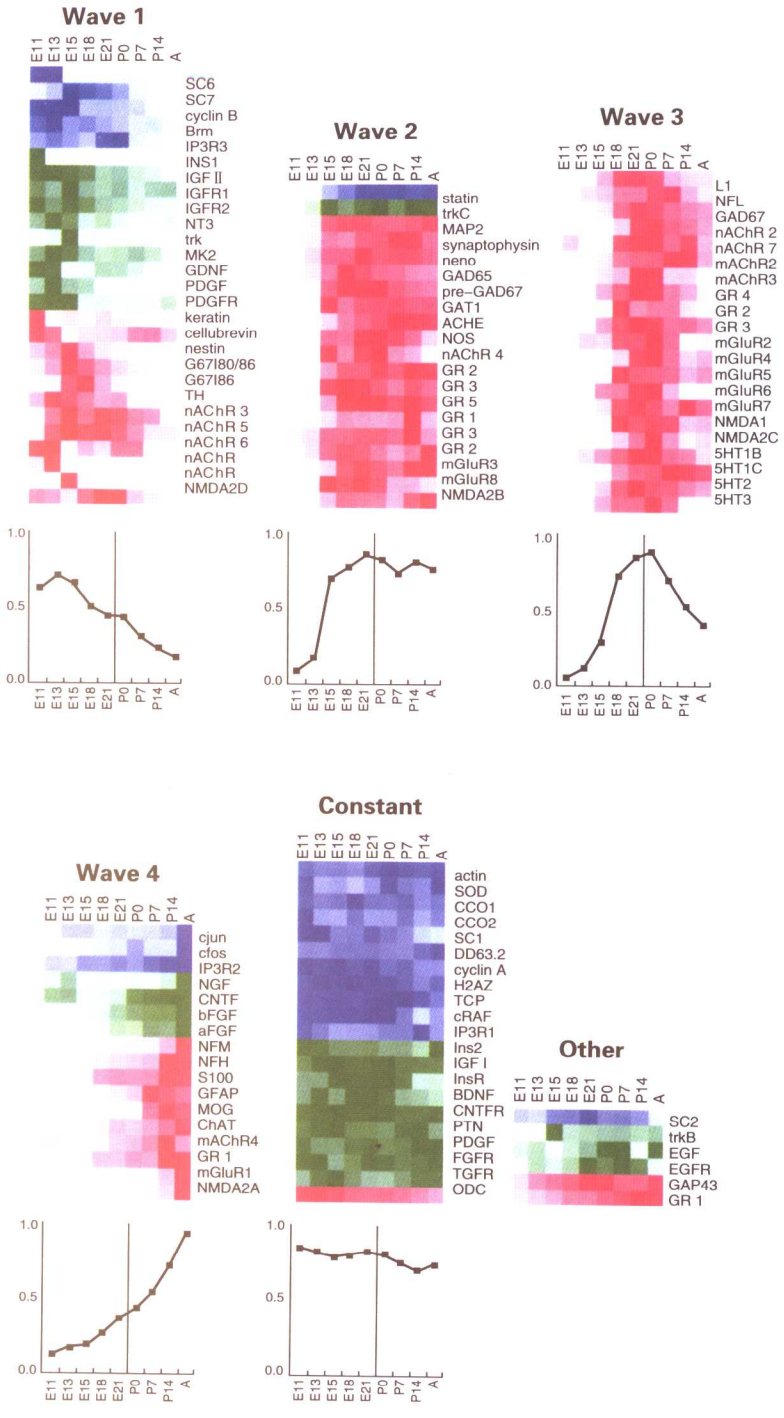

图 13–1　大鼠脊髓发育过程中112个基因时序表达的聚类分析

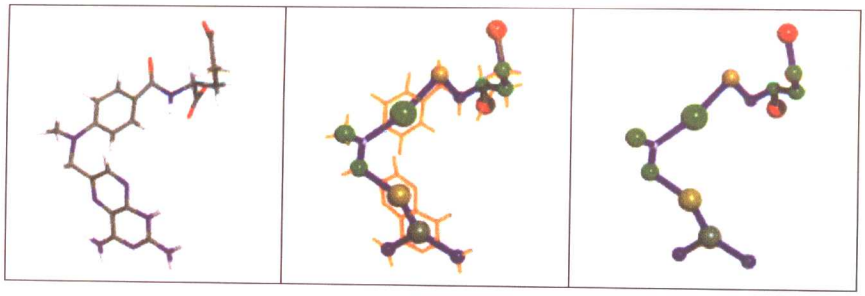

图 14–1　一个分子和它相应的特征树替代

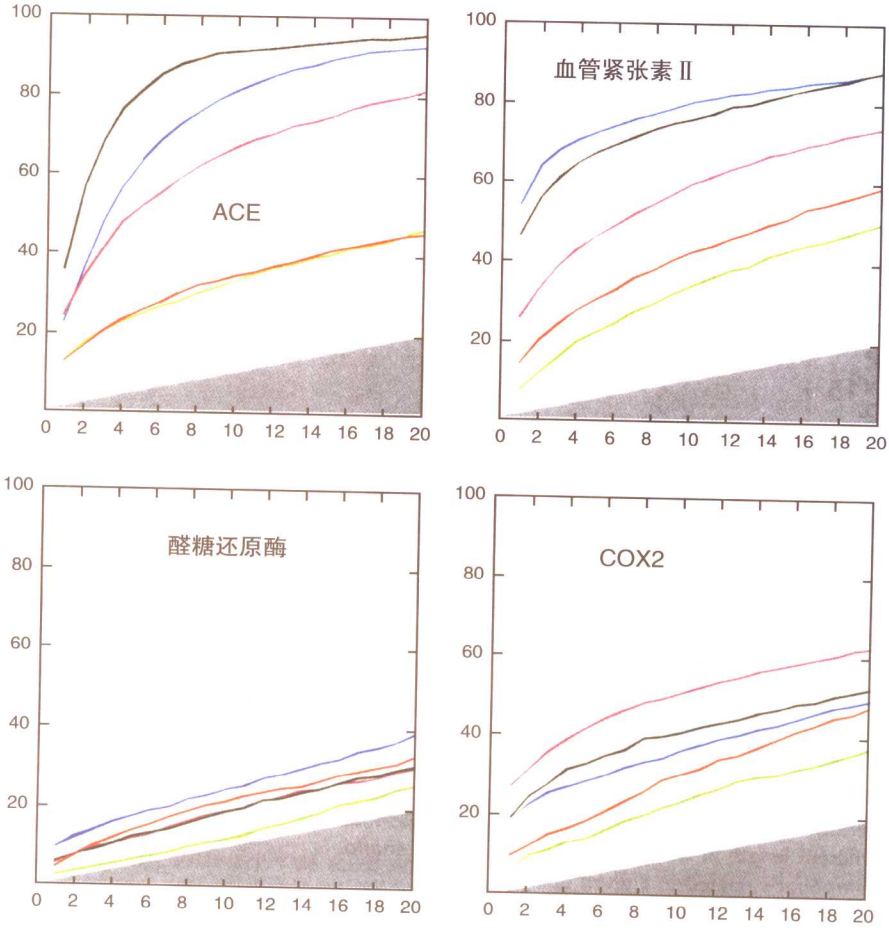

图 14–3

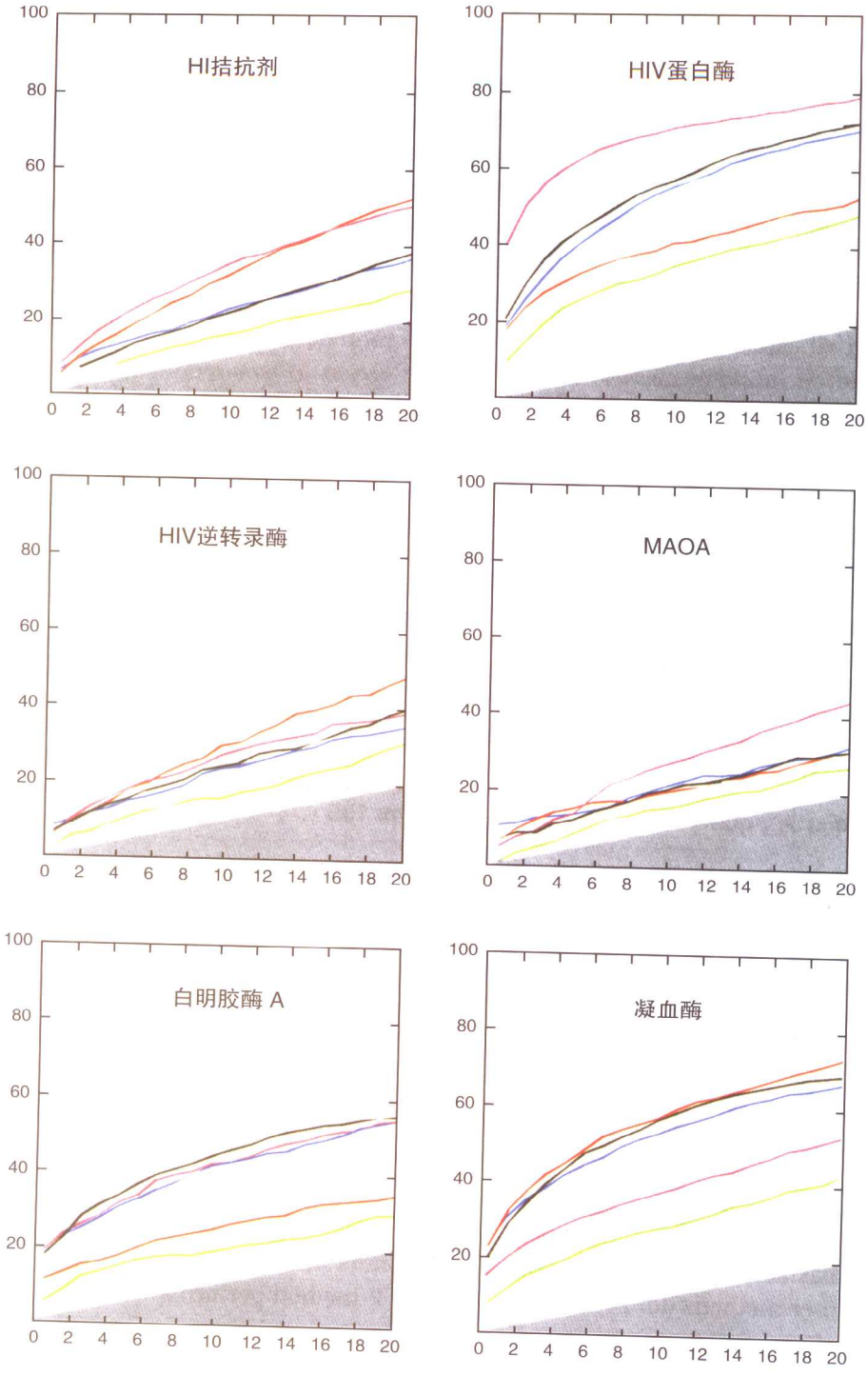
HI拮抗剂
HIV蛋白酶
HIV逆转录酶
MAOA
白明胶酶 A
凝血酶
100
80
60
40
20
0 2 4 6 8 10 12 14 16 18 20

图 14-3

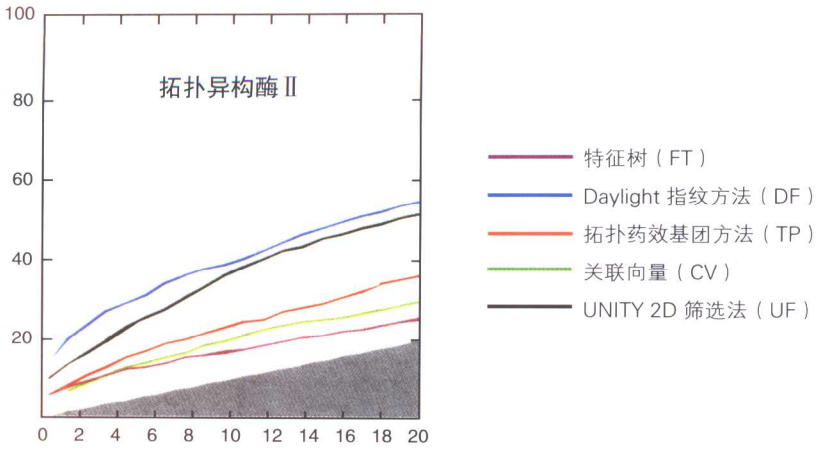

图 14–3　活性物质平均百分比对已排序数据库百分比作图

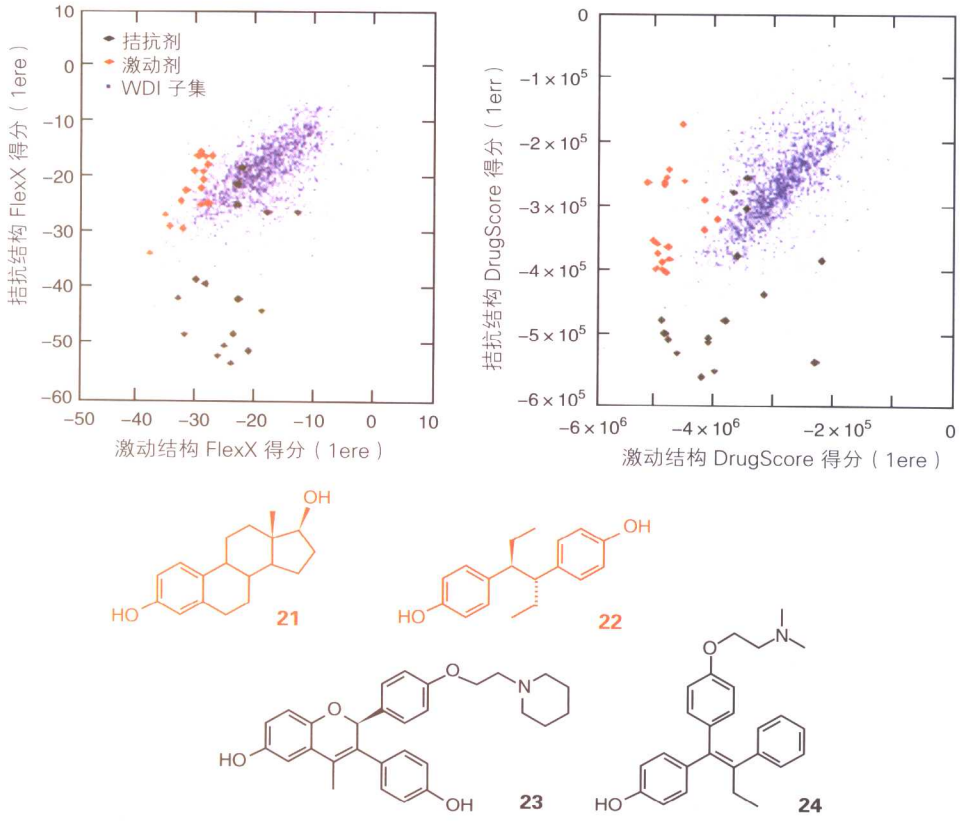

图 14–11　20种激动剂、20种拮抗剂和 WDI 子库化合物和雌激素受体以拮抗形式（PDB 号 Lerr）对接得分对和同一受体激动形式（PDB 号 Lere）对接得分作图

目　录

第1篇　基础技术