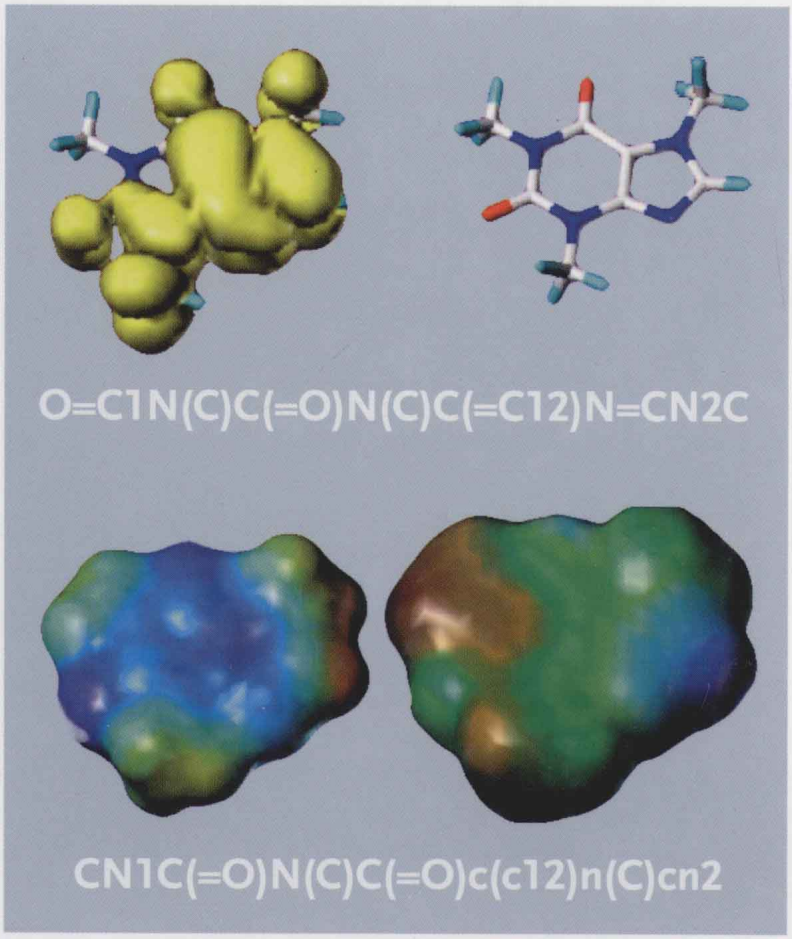
Edited by Tudor I. Oprea

WILEY-VCH

Chemoinformatics in Drug Discovery



Volume 23

Series Editors:
R. Mannhold,
H. Kubinyi,
G. Folkers

Series Editors:

Prof. Dr. Raimund Mannhold
Biomedical Research Center
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätsstrasse 1
40225 Düsseldorf
Germany
raimund.mannhold@uni-duesseldorf.de

Prof. Dr. Hugo Kubinyi
Donnersbergstrasse 9
67256 Weisenheim and Sand
Germany
kubinyi@t-online.de

Prof. Dr. Gerd Folkers
Department of Applied Biosciences
ETH Zürich
Winterthurerstrasse 19
8057 Zürich
Switzerland
folkers@pharma.anbi.ethz.ch

Volume Editor:

Prof. Dr. Tudor I. Oprea
Division of Biocomputing
MSC08 4560
University of New Mexico
School of Medicine
Albuquerque, NM 87131
USA

Library of Congress Card No.: Applied for
British Library Cataloging-in-Publication Data A catalogue record for this book is available from the British Library.

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at http://dnb.ddb.de.

Printed in the Federal Republic of Germany
Printed on acid-free paper

Composition Laserwords Private Ltd, Chennai, India
Printing betz-druck GmbH, Darmstadt
Bookbinding Litges & Dopf Buchbinderei GmbH, Heppenheim

ISBN-13: 978-3-527-30753-1
ISBN-10: 3-527-30753-2

Chemoinformatics in Drug Discovery

Edited by
Tudor I. Oprea

Methods and Principles in Medicinal Chemistry

Edited by R. Mannhold, H. Kubinyi, G. Folkers

Recently Published Volumes:

T. Lengauer (ed.)

Bioinformatics – From Genomes to Drugs
Vol. 14

2001, ISBN 3-527-29988-2

J. K. Seydel, M. Wiese

Drug-Membrane Interactions
Vol. 15

2002, ISBN 3-527-30427-4

O. Zerbe (ed.)

BioNMR in Drug Research
Vol. 16

2002, ISBN 3-527-30465-7

P. Carloni, F. Alber (eds.)

Quantum Medicinal Chemistry
Vol. 17

2003, ISBN 3-527-30456-8

H. van de Waterbeemd, H. Lennernäs, P. Artursson (eds.)

Drug Bioavailability
Vol. 18

2003, ISBN 3-527-30438-X

H.-J. Böhm, G. Schneider (eds.)

Protein-Ligand Interactions
Vol. 19

2003, ISBN 3-527-30521-1

R. E. Babine, S. S. Abdel-Meguid (eds.)

Protein Crystallography in Drug Discovery
Vol. 20

2004, ISBN 3-527-30678-1

Th. Dingermann, D. Steinhilber, G. Folkers (eds.)

Molecular Biology in Medicinal Chemistry
Vol. 21

2004, ISBN 3-527-30431-2

H. Kubinyi, G. Müller (eds.)

Chemogenomics in Drug Discovery
Vol. 22

2004, ISBN 3-527-30987-X

A Personal Foreword

This volume brings together contributions from academic and industrial scientists who develop and apply chemoinformatics strategies and tools in drug discovery. From chemical inventory and compound registration to candidate drug nomination, chemoinformatics integrates data via computer-assisted manipulation of chemical structures. Linked with computational chemistry, physical (organic) chemistry, pharmacodynamics and pharmacokinetics, chemoinformatics provides unique capabilities in the areas of lead and drug discovery. This book aims to offer knowledge and practical insights into the use of chemoinformatics in preclinical research.

Divided in four sections, the book opens with a first-hand account from Garland Marshall, spanning four decades of chemoinformatics and pharmaceutical research and development. Part one sets the stage for virtual screening and lead discovery. Hit and lead discovery via *in silico* technologies are highlighted in part two. In part three, data collection and mining using chemical databases are discussed in the context of chemical libraries. Specific applications and examples are collected in part four, which brings together industrial and academic perspectives. The book concludes with another personal account by Don Abraham, who presents drug discovery from an academic perspective.

The progression hit identification → lead generation → lead optimization → candidate drug nomination is served by a variety of chemoinformatics tools and strategies, most of them supporting the decision-making process. Key procedures and steps, from virtual screening to *in silico* lead optimization and from compound acquisition to library design, underscore our progress in grasping the preclinical drug discovery process, its needs for novel technologies and for integrated informatics support. We now have the ability to identify novel chemotypes in a rational manner, and *in silico* methods are deep-rooted in the process of systematic discovery. Our increased knowledge in a variety of seemingly unrelated phenomena, from atomic level issues related to drug–receptor binding to bulk properties of drugs and pharmacokinetics profiling, is likely to lead us on a better path for the discovery of orally bioavailable drugs, at the same time paving the way for novel, unexpected therapeutics.

I want to acknowledge all the contributors who made this book possible. Their insights, examples and personal accounts move beyond the sometimes dry language of science, turning this volume into an interesting and fascinating book to read.

Finally, I thank Frank Weinreich and Hugo Kubinyi for their encouragement and timely pressure to prepare this book on time.

Albuquerque, January 2005 *Tudor I. Oprea*

Chemoinformatics in Drug Discovery. Edited by Tudor I. Oprea

ISBN: 3-527-30753-2

Preface

The term "chemoinformatics" was introduced in 1998 by Dr. Frank K. Brown in the Annual Reports of Medicinal Chemistry. In his article "Chemoinformatics: What is it and How does it Impact Drug Discovery", he defines chemoinformatics as follows: "*The use of information technology and management has become a critical part of the drug discovery process. Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and organization*".

In fact, Chemoinformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization and use of chemical information. Related terms of chemoinformatics are cheminformatics, chemi-informatics, chemometrics, computational chemistry, chemical informatics, and chemical information management/science.

Reflecting the above given definitions, the present volume on "Chemoinformatics in Drug Discovery"covers its most important aspects within four main sections. After an introduction to chemoinformatics in drug discovery by Garland Marshall, the first section is focused on *Virtual Screening*. T. Oprea describes the use of "Chemoinformatics in Lead Discovery" and M.M. Hann et al. deal with "Computational Chemistry, Molecular Complexity and Screening Set Design". Then, M. Rarey et al. review "Algorithmic Engines in Virtual Screening" and D. Horvath et al. review the "Strengths and Limitations of Pharmacophore-Based Virtual Screening". The next section is dedicated to *Hit and Lead Discovery* with chapters of I.J. McFadyen et al. on "Enhancing Hit Quality and Diversity Within Assay Throughput Constraints", of C.L. Cavallaro et al. on "Molecular Diversity in Lead Discovery", and of C. Ho on "In Silico Lead Optimization". Topics of the third section refer to *Databases and Libraries*. They include chapters on "WOMBAT: World of Molecular Bioactivity" by M. Olah et al., on "Cabinet – Chemical and Biological Informatics Network" by V. Povolna et al., on "Structure Modification in Chemical Databases" by P.W. Kenney and J. Sadowski, and on the "Rational Design of GPCR-specific Combinational Libraries Based on the Concept of Privileged Substructures" by N.P. Savchuk et al.

According to our intention, to provide in this series on "Methods and Principles in Medicinal Chemistry" practice-oriented monographs, the book closes with a section on *Chemoinformatics Applications*. These are exemplified by G.M. Maggiora et al. in a chapter on "A Practical Strategy for Directed Compound Acquisition", by

Chemoinformatics in Drug Discovery. Edited by Tudor I. Oprea

ISBN: 3-527-30753-2

K.-H. Baringhaus and H. Matter on "Efficient Strategies for Lead Optimization by Simultaneously Addressing Affinity, Selectivity and Pharmacokinetic Parameters", by R.A. Goodnow et al. on "Chemoinformatic Tools for Library Design and the Hit-to-Lead Process" and by A. Tropsha on the "Application of Predictive QSAR Models to Database Mining". The section is concluded by a chapter of D.J. Abraham on "Drug Discovery from an Academic Perspective".

The series editors would like to thank Tudor Oprea for his enthusiasm to organize this volume and to work with such a fine selection of authors. We also want to express our gratitude to Frank Weinreich from Wiley-VCH for his valuable contributions to this project.

September 2004

Raimund Mannhold, Düsseldorf
Hugo Kubinyi, Weisenheim am Sand
Gerd Folkers, Zürich

List of Contributors

DONALD J. ABRAHAM
Department of Medicinal Chemistry
Virginia Commonwealth University
800 E. Leigh Street
Richmond, VA 23219-1540
USA

JUAN C. ALVAREZ
Chemical and Screening Sciences
Wyeth Research
200 Cambridge Park Drive
Cambridge, MA 02140
USA

KONSTANTIN V. BALAKIN
Chemical Diversity Labs, Inc.
Computational and Medicinal Chemistry
11575 Sorrento Valley Road
San Diego, CA 92121
USA

MAGDALENA BANDA
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

FRÉDÉRIQUE BARBOSA
Cerep S. A.
128, Rue Danton
92506 Rueil-Malmaison
France

KARL-HEINZ BARINGHAUS
Aventis Pharma Deutschland GmbH
DI&A Chemistry
Computational Chemistry
Industriepark Höchst, Bldg. G 878
65926 Frankfurt am Main
Germany

KONRAD BLEICHER
F. Hoffmann-La Roche Ltd.
PRBD-C
4070 Basel
Switzerland

ALINA BORA
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

CULLEN L. CAVALLARO
Pharmaceutical Research Institute
Bristol-Myers Squibb Co.
Province Line Road
Princeton, NJ 08540
USA

SCOTT DIXON
Metaphorics, LLC
441 Greg Avenue
Sante Fe, NM 87501
USA

Chemoinformatics in Drug Discovery. Edited by Tudor I. Oprea

ISBN: 3-527-30753-2

THOMPSON N. DOMAN
Lilly Research Laboratories
Structural and Computational Sciences
Indianapolis, IN 46285
USA

PAUL GILLESPIE
Hoffmann-La Roche Inc.
Discovery Chemistry
340 Kingsland Street
Nutley, NJ 07110
USA

ROBERT A. GOODNOW, Jr.
Hoffmann-La Roche Inc.
New Leads Chemistry Initiative
340 Kingsland Street
Nutley, NJ 07110
USA

RAFAEL GOZALBES
Cerep S. A.
128, Rue Danton
92506 Rueil-Malmaison
France

DARREN V. S. GREEN
GlaxoSmithKline Research
and Development
Gunnels Wood Road
Stevenage, SG1 2NY
United Kingdom

NICOLETA HADARUGA
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

MICHAEL M. HANN
GlaxoSmithKline Research
and Development
Gunnels Wood Road
Stevenage, SG1 2NY
United Kingdom

CHRIS HO
Drug Design Methodologies, LLC
4355 Maryland Ave.
St. Louis, MO 63108
USA

DRAGOS HORVATH
UMR8525-CNRS
Institut de Biologie de Lille
1, rue Calmette
59027 Lille
France

PETER W. KENNY
AstraZeneca Mereside
Alderley Park
Macclesfield, SK10 4TG
United Kingdom

MICHAEL S. LAJINESS
Lilly Research Laboratories
Structural and Computational Sciences
Indianapolis, IN 46285
USA

ANDREW R. LEACH
GlaxoSmithKline Research
and Development
Gunnels Wood Road
Stevenage, SG1 2NY
United Kingdom

CHRISTIAN LEMMEN
BioSolveIT GmbH
An der Ziegelei 75
53757 Sankt Augustin
Germany

GERALD M. MAGGIORA
Dept. of Pharmacology and Toxicology
University of Arizona
College of Pharmacy
Tucson, AZ 85271
USA

BORYEU MAO
Cerep Inc.
15318 NE 95th Street
Redmond, WA 98052
USA

GARLAND R. MARSHALL
Center for Computational Biology
Washington University
School of Medicine
660 S. Euclid Ave.
St. Louis, MO 63110
USA

HANS MATTER
Aventis Pharma Deutschland GmbH
DI&A Chemistry
Computational Chemistry
Industriepark Höchst, Bldg. G 878
65926 Frankfurt am Main
Germany

IAIN MCFADYEN
Chemical and Screening Sciences
Wyeth Research
200 Cambridge Park Drive
Cambridge, MA 02140
USA

MARIA MRACEC
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

MIRCEA MRACEC
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

IONELA OLAH
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

MARIUS OLAH
Division of Biocomputing, MSC08 4560
University of New Mexico
School of Medicine
Albuquerque, NM 87131
USA

TUDOR I. OPREA
Division of Biocomputing, MSC08 4560
University of New Mexico
School of Medicine
Albuquerque, NM 87131
USA

LILIANA OSTOPOVICI
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

VERA POVOLNA
Metaphorics, LLC
441 Greg Avenue
Sante Fe, NM 87501
USA

RAMONA RAD
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

MATTHIAS RAREY
Center for Bioinformatics (ZBH)
University of Hamburg
Bundesstrasse 43
20146 Hamburg
Germany

SHERRY L. ROGALSKI
Cerep Inc. 15318 NE 95th Street
Redmond, WA 98052
USA

JENS SADOWSKI
AstraZeneca R&D Mölndal
Structural Chemistry Laboratory
SC264
43183 Mölndal
Sweden

NIKOLAY P. SAVCHUK
Chemical Diversity Labs, Inc.
Chemoinformatics
11575 Sorrento Valley Road
San Diego, CA 92121
USA

DORA M. SCHNUR
Pharmaceutical Research Institute
Bristol-Myers Squibb Co.
Province Line Road
Princeton, NJ 08540
USA

MARTIN W. SCHULTZ
Pfizer Global Research and Development
301 Henrietta Street
Kalamazoo, MI 49007
USA

VEERABAHU SHANMUGASUNDARAM
Computer-Assisted Drug Discovery
Pfizer Global Research and Development
2800 Plymouth Road
Ann Arbor, MI 48105
USA

ZENO SIMON
Romanian Academy
Institute of Chemistry
"Coriolan Dragulescu"
Bv. Mihai Viteazul No. 24
300223 Timisoara
Romania

ANDREW J. TEBBEN
Pharmaceutical Research Institute
Bristol-Myers Squibb Co.
Province Line Road
Princeton, NJ 08540
USA

SERGEY E. TKACHENKO
Chemical Diversity Labs, Inc.
Computational and Medicinal Chemistry
11575 Sorrento Valley Road
San Diego, CA 92121
USA

ALEXANDER TROPSHA
Laboratory for Molecular Modeling
University of North Carolina
Chapel Hill, NC 27599
USA

GARY WALKER
Chemical and Screening Sciences
Wyeth Research
401 N. Middletown Road
Pearl River, NY 10965
USA

DAVID WEININGER
Metaphorics, LLC
441 Greg Avenue
Sante Fe, NM 87501
USA

Contents

A Personal Foreword *XV*

Preface *XVII*

List of Contributors *XIX*

1 Introduction to Chemoinformatics in Drug Discovery – A Personal View *1*
Garland R. Marshall

1.1 Introduction *1*
1.2 Historical Evolution *4*
1.3 Known versus Unknown Targets *5*
1.4 Graph Theory and Molecular Numerology *6*
1.5 Pharmacophore *7*
1.6 Active-Analog Approach *8*
1.7 Active-Site Modeling *9*
1.8 Validation of the Active-Analog Approach and Active-Site Modeling *10*
1.9 PLS/CoMFA *11*
1.10 Prediction of Affinity *12*
1.11 Protein Structure Prediction *13*
1.12 Structure-Based Drug Design *15*
1.13 Real World Pharmaceutical Issues *15*
1.14 Combinatorial Chemistry and High-throughput Screens *16*
1.15 Diversity and Similarity *16*
1.16 Prediction of ADME *17*
1.17 Failures to Accurately Predict *17*
1.18 Summary *18*
References *19*

Chemoinformatics in Drug Discovery. Edited by Tudor I. Oprea

ISBN: 3-527-30753-2

Part I **Virtual Screening** *23*

2 **Chemoinformatics in Lead Discovery** *25*
Tudor I. Oprea

2.1 Chemoinformatics in the Context of Pharmaceutical Research *25*
2.2 Leads in the Drug Discovery Paradigm *27*
2.3 Is There a Trend for High Activity Molecules? *29*
2.4 The Concept of Leadlikeness *32*
2.5 Conclusions *37*
References *38*

3 **Computational Chemistry, Molecular Complexity and Screening Set Design** *43*
Michael M. Hann, Andrew R. Leach, and Darren V.S. Green

3.1 Introduction *43*
3.2 Background Concepts: the Virtual, Tangible and Real Worlds of Compounds, the "Knowledge Plot" and Target Tractability *44*
3.3 The Construction of High Throughput Screening Sets *45*
3.4 Compound Filters *47*
3.5 "Leadlike" Screening Sets *48*
3.6 Focused and Biased Set Design *54*
3.7 Conclusion *55*
References *56*

4 **Algorithmic Engines in Virtual Screening** *59*
Matthias Rarey, Christian Lemmen, and Hans Matter

4.1 Introduction *59*
4.2 Software Tools for Virtual Screening *61*
4.3 Physicochemical Models in Virtual Screening *62*
4.3.1 Intermolecular Forces in Protein–Ligand Interactions *63*
4.3.2 Scoring Functions for Protein–Ligand Recognition *66*
4.3.3 Covering Conformational Space *67*
4.3.4 Scoring Structural Alignments *68*
4.4 Algorithmic Engines in Virtual Screening *69*
4.4.1 Mathematical Concepts *69*
4.4.2 Algorithmic Concepts *76*
4.4.3 Descriptor Technology *81*
4.4.4 Global Search Algorithms *85*
4.5 Entering the Real World: Virtual Screening Applications *89*
4.5.1 Practical Considerations on Virtual Screening *89*
4.5.2 Successful Applications of Virtual Screening *91*
4.6 Practical Virtual Screening: Some Final Remarks *99*
References *101*

5 Strengths and Limitations of Pharmacophore-Based Virtual Screening *117*
Dragos Horvath, Boryeu Mao, Rafael Gozalbes, Frédérique Barbosa, and Sherry L. Rogalski

5.1 Introduction *117*
5.2 The "Pharmacophore" Concept: Pharmacophore Features *117*
5.3 Pharmacophore Models: Managing Pharmacophore-related Information *118*
5.4 The Main Topic of This Paper *119*
5.5 The Cox2 Data Set *119*
5.6 Pharmacophore Fingerprints and Similarity Searches *120*
5.7 Molecular Field Analysis (MFA)-Based Pharmacophore Information *123*
5.8 QSAR Models *125*
5.9 Hypothesis Models *125*
5.10 The Minimalist Overlay-Independent QSAR Model *126*
5.11 Minimalist and Consensus Overlay-Based QSAR Models *128*
5.12 Diversity Analysis of the Cox2 Compound Set *131*
5.13 Do Hypothesis Models Actually Tell Us More Than Similarity Models About the Structural Reasons of Activity? *131*
5.14 Why Did Hypothesis Models Fail to Unveil the Key Cox2 Site–Ligand Interactions? *134*
5.15 Conclusions *136*
References *137*

Part II Hit and Lead Discovery *141*

6 Enhancing Hit Quality and Diversity Within Assay Throughput Constraints *143*
Iain McFadyen, Gary Walker, and Juan Alvarez

6.1 Introduction *143*
6.1.1 What Makes a Good Lead Molecule? *144*
6.1.2 Compound Collections – Suitability as Leads *144*
6.1.3 Compound Collections – Diversity *145*
6.1.4 Data Reliability *146*
6.1.5 Selection Methods *149*
6.1.6 Enhancing Quality and Diversity of Actives *153*
6.2 Methods *154*
6.2.1 Screening Library *155*
6.2.2 Determination of Activity Threshold *156*
6.2.3 Filtering *156*
6.2.4 High-Throughput Screen Clustering Algorithm (HTSCA) *157*
6.2.5 Diversity Analysis *160*

6.2.6 Data Visualization *161*
6.3 Results *162*
6.3.1 Peptide Hydrolase *162*
6.3.2 Protein Kinase *167*
6.3.3 Protein–Protein Interaction *168*
6.4 Discussion and Conclusion *169*
References *172*

7 Molecular Diversity in Lead Discovery: From Quantity to Quality *175*
Cullen L. Cavallaro, Dora M. Schnur, and Andrew J. Tebben

7.1 Introduction *175*
7.2 Large Libraries and Collections *176*
7.2.1 Methods and Examples for Large Library Diversity Calculations *177*
7.3 Medium-sized/Target-class Libraries and Collections *181*
7.3.1 Computational Methods for Medium- and Target-class Libraries and Collections *183*
7.4 Small Focused Libraries *189*
7.4.1 Computational Methods for Small and Focused Libraries *190*
7.5 Summary/Conclusion *191*
References *192*

8 *In Silico* Lead Optimization *199*
Chris M.W. Ho

8.1 Introduction *199*
8.2 The Rise of Computer-aided Drug Refinement *200*
8.3 RACHEL Software Package *201*
8.4 Extraction of Building Blocks from Corporate Databases *201*
8.5 Intelligent Component Selection System *203*
8.6 Development of a Component Specification Language *205*
8.7 Filtration of Components Using Constraints *207*
8.8 Template-driven Structure Generation *208*
8.9 Scoring Functions – Methods to Estimate Ligand–Receptor Binding *209*
8.10 Target Functions *212*
8.11 Ligand Optimization Example *214*
References *219*

Part III Databases and Libraries *221*

9 WOMBAT: World of Molecular Bioactivity *223*
Marius Olah, Maria Mracec, Liliana Ostopovici, Ramona Rad, Alina Bora, Nicoleta Hadaruga, Ionela Olah, Magdalena Banda, Zeno Simon, Mircea Mracec, and Tudor I. Oprea

9.1 Introduction – Brief History of the WOMBAT Project *223*
9.2 WOMBAT 2004.1 Overview *224*
9.3 WOMBAT Database Structure *227*
9.4 WOMBAT Quality Control *228*
9.5 Uncovering Errors from Literature *231*
9.6 Data Mining with WOMBAT *234*
9.7 Conclusions and Future Challenges *235*
References *237*

10 Cabinet – Chemical and Biological Informatics Network *241*
Vera Povolna, Scott Dixon, and David Weininger

10.1 Introduction *241*
10.1.1 Integration Efforts, WWW as Information Resource and Limitations *241*
10.1.2 Goals *243*
10.2 Merits of Federation Rather than Unification *243*
10.2.1 The Merits of Unification *244*
10.2.2 The Merits of Federation *244*
10.2.3 Unifying Disparate Data Models is Difficult, Federating them is Easy *245*
10.2.4 Language is a Natural Key *246*
10.3 HTTP is Appropriate Communication Technology *248*
10.3.1 HTTP is Specifically Designed for Collaborative Computing *248*
10.3.2 HTTP is the Dominant Communication Protocol Today *248*
10.3.3 HTML Provides a Universally Accessible GUI *249*
10.3.4 MIME "Text/Plain" and "Application/Octet-Stream" are Important Catch-alls *249*
10.3.5 Other MIME Types are Useful *250*
10.3.6 One Significant HTTP Work-around is Required *250*
10.4 Implementation *251*
10.4.1 Daylight HTTP Toolkit *251*
10.4.2 Metaphorics' Cabinet Library *252*
10.5 Specific Examples of Federated Services *252*
10.5.1 Empath – Metabolic Pathway Chart *253*
10.5.2 Planet – Protein–ligand Association Network *254*
10.5.3 EC Book – Enzyme Commission Codebook *254*
10.5.4 WDI – World Drug Index *254*

10.5.5 WOMBAT – World of Molecular Bioactivity 255
10.5.6 TCM (Traditional Chinese Medicines), DCM (Dictionary of Chinese Medicine), PARK (Photo ARKive) and zi4 255
10.5.7 Cabinet "Download" Service 256
10.5.8 Cabinet Usage Example 256
10.6 Deployment and Refinement 262
10.6.1 Local Deployment 264
10.6.2 Intranet Deployment 264
10.6.3 Internet Deployment 265
10.6.4 Online Deployment 266
10.7 Conclusions 266
References 268

11 Structure Modification in Chemical Databases 271
Peter W. Kenny and Jens Sadowski

11.1 Introduction 271
11.2 Permute 274
11.2.1 Protonation and Formal Charges 274
11.2.2 Tautomerism 275
11.2.3 Nitrogen Configurations 276
11.2.4 Duplicate Removal 276
11.2.5 Nested Loop 276
11.2.6 Application Statistics 277
11.2.7 Impact on Docking 277
11.3 Leatherface 279
11.3.1 Protonation and Formal Charges 279
11.3.2 Tautomerism 280
11.3.3 Ionization and Tautomer Model 281
11.3.4 Relationships between Structures 282
11.3.5 Substructural Searching and Analysis 283
11.4 Concluding Remarks 283
References 284

12 Rational Design of GPCR-specific Combinational Libraries Based on the Concept of Privileged Substructures 287
Nikolay P. Savchuk, Sergey E. Tkachenko, and Konstantin V. Balakin

12.1 Introduction – Combinatorial Chemistry and Rational Drug Design 287
12.2 Rational Selection of Building Blocks Based on Privileged Structural Motifs 288
12.2.1 Privileged Structures and Substructures in the Design of Pharmacologically Relevant Combinatorial Libraries 288