Common generic scaffolds of synthetic bioactive compounds

Common generic scaffolds of natural products
Edited by
Gisbert Schneider

De novo Molecular Design
<table>
<thead>
<tr>
<th>Related Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Brown, N. (ed.)</strong></td>
</tr>
<tr>
<td><strong>Scaffold Hopping in Medicinal Chemistry</strong></td>
</tr>
<tr>
<td>2014</td>
</tr>
<tr>
<td>ISBN: 978-3-527-33364-6</td>
</tr>
<tr>
<td>(also available in digital formats)</td>
</tr>
<tr>
<td><strong>Hoffmann, R., Gohier, A., Pospisil, P. (eds.)</strong></td>
</tr>
<tr>
<td><strong>Data Mining in Drug Discovery</strong></td>
</tr>
<tr>
<td>2014</td>
</tr>
<tr>
<td>ISBN: 978-3-527-32984-7</td>
</tr>
<tr>
<td>(also available in digital formats)</td>
</tr>
<tr>
<td><strong>Brown, N. (ed.)</strong></td>
</tr>
<tr>
<td><strong>Bioisosteres in Medicinal Chemistry</strong></td>
</tr>
<tr>
<td>2012</td>
</tr>
<tr>
<td>ISBN: 978-3-527-33015-7</td>
</tr>
<tr>
<td>(also available in digital formats)</td>
</tr>
<tr>
<td><strong>Sotriffer, C. (ed.)</strong></td>
</tr>
<tr>
<td><strong>Virtual Screening</strong></td>
</tr>
<tr>
<td><strong>Principles, Challenges, and Practical Guidelines</strong></td>
</tr>
<tr>
<td>2011</td>
</tr>
<tr>
<td>ISBN: 978-3-527-32636-5</td>
</tr>
<tr>
<td>(also available in digital formats)</td>
</tr>
<tr>
<td><strong>Comba, P. (ed.)</strong></td>
</tr>
<tr>
<td><strong>Modeling of Molecular Properties</strong></td>
</tr>
<tr>
<td>2011</td>
</tr>
<tr>
<td>ISBN: 978-3-527-33021-8</td>
</tr>
<tr>
<td>(also available in digital formats)</td>
</tr>
<tr>
<td><strong>Matta, Chérif F. (ed.)</strong></td>
</tr>
<tr>
<td><strong>Quantum Biochemistry</strong></td>
</tr>
<tr>
<td>2010</td>
</tr>
<tr>
<td>ISBN: 978-3-527-32322-7</td>
</tr>
<tr>
<td>(also available in digital formats)</td>
</tr>
<tr>
<td><strong>Schneider, G., Baringhaus, K.-H</strong></td>
</tr>
<tr>
<td><strong>Molecular Design</strong></td>
</tr>
<tr>
<td><strong>Concepts and Applications</strong></td>
</tr>
<tr>
<td>2008</td>
</tr>
<tr>
<td>ISBN: 978-3-527-31432-4</td>
</tr>
</tbody>
</table>
De novo Molecular Design
Contents

List of Contributors XV
Foreword XXI
Preface XXIII

1 De Novo Design: From Models to Molecules 1
Gisbert Schneider and Karl-Heinz Baringhaus

1.1 Molecular Representation 1
1.2 The Molecular Design Cycle 9
1.3 Receptor–Ligand Interaction 14
1.4 Modeling Fitness Landscapes 21
1.4.1 Naïve Bayes Classifier 26
1.4.2 Artificial Neural Network 27
1.4.3 Support Vector Machine 27
1.4.4 Gaussian Process 28
1.5 Strategies for Compound Construction 30
1.6 Strategies for Compound Scoring 33
1.6.1 Receptor-Based Scoring 35
1.6.2 Ligand-Based Scoring 37
1.7 Flashback Forward: A Brief History of De Novo Drug Design 37
1.8 Conclusions 43
Acknowledgments 43
References 44

2 Coping with Complexity in Molecular Design 57
Michael M. Hann and Andrew R. Leach

2.1 Introduction 57
2.2 A Simple Model of Molecular Interactions 58
2.3 Enhancements to the Simple Complexity Model 60
2.4 Enumerating and Sampling the Complexity of Chemical Space 61
2.5 Validation of the Complexity Model 65
2.6 Reductionism and Drug Design 67
Contents

2.7 Complexity and Information Content as a Factor in De Novo Design  69
2.8 Complexity of Thermodynamic Entropy and Drug Design  73
2.9 Complex Systems, Emergent Behavior, and Molecular Design  74
Acknowledgments  75
References  75

3 The Human Pocketome  79
Ruben Abagyan and Clarisse Gravina Ricci
3.1 Predicted Pockets  79
3.2 Compilation of the Validated Human Pocketome  83
3.3 Diversity and Redundancy of the Human Pocketome  85
3.4 Compound Activity Prediction by Ligand-Pocket Docking and Scoring  87
3.4.1 Optimizing Pocket Sets for Reliable Docking and Scoring Results  87
3.4.2 Difficult Cases: Unusually Large and Multifunctional Pockets  88
3.5 Pocketome-Derived 3D Chemical Fields as Activity Prediction Models  90
3.6 Clustering the Ligands by Function and Subpockets  92
3.7 Conclusions  94
Acknowledgments  94
References  94

4 Structure-Based De Novo Drug Design  97
Alla Srinivas Reddy, Lu Chen, and Shuxing Zhang
4.1 Introduction  97
4.2 Current Progress in SBDND Methodologies  99
4.2.1 Identification of Binding Site  100
4.2.2 Design of Molecules  101
4.2.2.1 Atom-Based versus Fragment-Based Methods  101
4.2.2.2 Pharmacophore-Based Methods  105
4.2.3 Searching the Chemical Space  105
4.2.3.1 Monte Carlo-Based Methods  106
4.2.3.2 Evolutionary Algorithms  106
4.2.4 Scoring Methods  108
4.2.4.1 Force-Field-Based Scoring Functions  108
4.2.4.2 Empirical Scoring Functions  109
4.2.4.3 Knowledge-Based Scoring Functions  109
4.2.4.4 Consensus Scoring  109
4.2.5 Synthetic Accessibility  110
4.3 Recent Applications of Structure-Based De Novo Design  110
4.4 Perspectives and Conclusion  115
Acknowledgment  116
References  116
5 De Novo Design by Fragment Growing and Docking 125

Jacob D. Durrant and Rommie E. Amaro

5.1 Introduction 125
5.2 Case Study I: High-Throughput Screening with Dr Feils 126
5.2.1 Target Identification 126
5.2.2 Small-Molecule Library Design 126
5.2.2.1 Computer Docking 128
5.2.2.2 Pharmacophore Searching 129
5.2.3 High-Throughput Screening 129
5.2.4 Optimization 130
5.3 Case Study II: Fragment-Based Drug Design with Dr Goode 130
5.3.1 Library Generation 130
5.3.1.1 Computational Techniques for Library Refinement 132
5.3.2 Detection Methods 132
5.3.2.1 Functional/High-Concentration Screening 132
5.3.2.2 Fluorescence-Based Thermal Shift Assay (TSA) 133
5.3.2.3 Surface Plasmon Resonance (SPR) 133
5.3.2.4 Mass Spectrometry (MS) 133
5.3.2.5 Nuclear Magnetic Resonance (NMR) 133
5.3.2.6 X-Ray Crystallography 134
5.3.3 Screening 134
5.3.4 Optimization 135
5.3.5 Final Products 137
5.4 Conclusion 138

6 Hit and Lead Identification from Fragments 143

Michael Mazanetz, Richard Law, and Mark Whittaker

6.1 Introduction to FBDD 144
6.2 Fragment Library Design Incorporating Computational Methods 148
6.2.1 Fragment Library Design Strategies 148
6.2.2 Molecular Attributes and Physicochemical Properties 150
6.2.3 Influence of Screening Method on Library Selection 151
6.2.4 Removal of Undesirable Functionality 152
6.2.5 Size of Library and Diversity 152
6.2.6 Focused Sets 154
6.2.7 Designing in Fragment Optimization 154
6.3 Fragment Screening 155
6.3.1 Screening by X-Ray Crystallography 155
6.3.2 Screening by NMR 156
6.3.3 Screening by SPR 157
6.3.4 Screening by Biochemical Assay 157
6.3.5 Thermal Shift Assays 158
6.3.6 Isothermal Titration Calorimetry (ITC) 160
6.3.7 Other Biophysical Assay Techniques 160
6.3.8 Assay Techniques for Membrane Proteins 161
6.3.9 Fragment Library Screening Using Computational Methods 161
6.3.10 Ligandability Screening Using Fragments 162
6.4 Fragment Prioritization for Optimization 165
6.4.1 Efficiency Metrics 165
6.4.2 Computational and Thermodynamic Methods for Fragment Selection and Prioritization 167
6.5 Fragment Hit Expansion and Fragment Evolution 170
6.6 Fragment Merging Principles 175
6.7 Fragment Linking Principles 177
6.8 Fragment-Assisted Drug Discovery (FADD) 182
6.9 Conclusion 183
Acknowledgments 184
References 184

7 Pharmacophore-Based De Novo Design 201
Wen-Jing Wang, Qi Huang, and Sheng-Yong Yang

7.1 Introduction 201

7.2 A Summary of the Algorithms of PhDD v1.0 202
7.2.1 The Basic Scheme of PhDD 202
7.2.2 Fragment and Linker Databases 203
7.2.3 Mapping of Fragments onto the Locations of Pharmacophore Features of the Pharmacophore Hypothesis 204
7.2.4 Connecting Fragments by Linkers 205
7.2.5 Assessments to the Generated Molecules 206
7.2.5.1 The Drug-Likeness Assessment 207
7.2.5.2 The Estimation of Bioactivity 207
7.2.5.3 The Assessment of Synthetic Accessibility 207
7.3 An Introduction to the Modifications in the Updated Version of PhDD (v2.0) 208
7.3.1 The Use of a Designated Fragment 209
7.3.2 Conformation Optimization in the Process of Molecular Construction 209
7.3.3 Two Pharmacophore Features Share One Fragment 210
7.4 Validation of PhDD 210
7.5 Concluding Remarks 212
Acknowledgment 213
References 213

8 3D-QSAR Approaches to De Novo Drug Design 215
Richard D. Cramer

8.1 Introduction 215
8.2 Current Methods 216
8.3 Leapfrog 217
8.4 Recent Advances 219
8.5 Conclusions 223
Acknowledgments 223
References 223

9 Ligand-Based Molecular Design Using Pseudoreceptors 227
Darren Fayne
9.1 Introduction 227
9.2 Pseudoreceptor Algorithms 231
9.3 Successful Applications Overview 232
9.4 Conclusions 240
Acknowledgments 241
References 241

10 Reaction-Driven De Novo Design: a Keystone for Automated Design of Target Family-Oriented Libraries 245
Markus Hartenfeller, Steffen Renner, and Edgar Jacoby
10.1 Introduction 245
10.2 Reaction-Driven Design: Tackling the Problem of Synthetic Feasibility 247
10.2.1 Exploiting the Valuable Knowledge Stored in Electronic Laboratory Notebooks 249
10.2.2 Assessing the Chemical Space of a Focused Set of Reactions 251
10.3 Successful Applications of Reaction-Driven De Novo Design 254
10.4 Reaction-Driven Design of Chemical Libraries Addressing Target Families 256
10.5 Conclusions 261
References 265

11 Multiobjective De Novo Design of Synthetically Accessible Compounds 267
Valerie J. Gillet, Michael J. Bodkin, and Dimitar Hristozov
11.1 Introduction 267
11.2 Design of Synthetically Accessible Compounds 269
11.3 Synthetic Accessibility Using Reaction Vectors 270
11.4 De Novo Design Using Evolutionary Algorithms 276
11.4.1 Optimizing Multiple Objectives 277
11.4.2 Multiobjective De Novo Design 279
11.4.3 Multiobjective De Novo Design Using Reaction Vectors 280
11.5 Conclusions 282
Acknowledgments 283
References 283
12  **De Novo Design of Ligands against Multitarget Profiles**  287
   *Jérémie Besnard and Andrew L. Hopkins*

12.1 Introduction  287
12.2 Automating the Creativity of Ligand Design  289
12.3 Evolutionary Algorithm  294
12.4 Experimental Validation  295
12.5 Reducing Antitarget Activity  296
12.6 Optimizing D4 Receptor Potency  301
12.7 Designing Novel Ligands to a Defined Profile  301
12.8 Conclusion  304

Acknowledgments  306
References  306

13  **Construction of Drug-Like Compounds by Markov Chains**  311
   *Peter S. Kutchukian, Salla I. Virtanen, Eugen Lounkine, Meir Glick, and Eugene I. Shakhnovich*

13.1 Introduction  311
13.2 FOG Algorithm and Library Generation  313
13.3 Applications  314
13.3.1 Overview  314
13.3.2 Target Class Prediction of FOG Compounds  314
13.3.3 Design of BACE-1 Inhibitors with FOG  316
13.4 Conclusion  319

Acknowledgments  320
References  320

14  **Coping with Combinatorial Space in Molecular Design**  325
   *Florian Lauck and Matthias Rarey*

14.1 Introduction  325
14.2 Chemical Space  326
14.2.1 Size Estimation of Chemical Space  327
14.2.2 Enumeration of Chemical Subspaces  328
14.3 Combinatorial Space  330
14.3.1 Generation of Combinatorial Spaces  332
14.3.1.1 Combinatorial Space from Fragmentation  332
14.3.1.2 Computational Space from Chemical Knowledge  334
14.3.2 Manipulation of Combinatorial Space  335
14.3.3 Querying Combinatorial Spaces  336
14.3.3.1 Fragment Spaces  337
14.3.3.2 Reaction-Based Combinatorial Spaces  339
14.3.4 Other Applications of Combinatorial Space  340
14.3.5 Markush Structures  341
14.4 Visualization  342
14.5 Conclusion  343

References  343
15  **Fragment-Based Design of Focused Compound Libraries**  349
   *Uta Lessel*

15.1  Introduction  349
15.2  General Workflow  351
15.3  Fragment Space  352
15.4  Query  355
15.5  FTrees Fragment Space Search  356
15.6  Scaffold Selection  356
15.7  Design of Focused Libraries  359
15.8  Application Example  360
15.9  Summary and Conclusions  366

Acknowledgments  367
References  367

16  **Free Energy Methods in Ligand Design**  373
   *Yvonne Westermaier and Roderick E. Hubbard*

16.1  Free Energy (FE) Methods in Lead Optimization (LO)  373
16.1.1  FE Methods: An Emerging Tool in Industry?  374
16.1.2  Finding the Needle in a Haystack: The Role of FE Methods in Fine-Tuning Ligand Discovery  375
16.2  The Variety of *In Silico* Binding Affinity Methods  377
16.2.1  Thermodynamic Integration (TI) and Alchemical Transformations  377
16.2.2  Free Energy Perturbation (FEP)  378
16.2.3  Potential of Mean Force (PMF) Calculations  379
16.2.4  Nonequilibrium Approaches  380
16.2.5  Other MM-Based Methods  381
16.2.5.1  Linear Interaction Energy (LIE)  381
16.2.5.2  MM-PBSA and MM-GBSA  382
16.3  The Choice of a Method for Calculating Binding FE  382
16.3.1  MM-PBSA and MM-GBSA versus FEP/TI  383
16.3.2  LIE versus FEP/TI  383
16.3.3  PMF versus FEP  383
16.3.4  PMF versus TI  383
16.3.5  TI versus FEP  384
16.3.6  PMF/TI/FEP: Absolute or Relative Binding FEs?  384
16.3.7  Equilibrium versus Nonequilibrium Methods  385
16.4  Experimental Data  385
16.5  Current Issues  385
16.6  Practical Examples  387
16.6.1  Studies on Model Systems  387
16.6.2  FE Methods Applied to Pharmacologically Relevant Systems  389
16.7 Miscellaneous Issues 395
16.8 Best Practices 396
16.9 Conclusions and Outlook 397
Acknowledgments 398
Abbreviations 398
References 399

17 Bioisosteres in De Novo Design 417
Nicholas C. Firth, Julian Blagg, and Nathan Brown
17.1 Introduction 417
17.2 History of Isosterism and Bioisosterism 418
17.3 Methods for Bioisosteric Replacement 421
17.3.1 Databases 422
17.3.1.1 BIOSTER 422
17.3.1.2 Cambridge Structure Database 422
17.3.1.3 ChEMBL 423
17.3.2 Descriptors 424
17.3.2.1 Physicochemical Properties 424
17.3.2.2 Molecular Topology 426
17.3.2.3 Molecular Shape 426
17.4 Exemplar Applications 427
17.4.1 Information-Based Bioisosteric Replacement 427
17.4.2 Drug Guru 429
17.4.3 SkelGen 431
17.5 Conclusions 433
Acknowledgments 433
References 434

18 Peptide Design by Nature-Inspired Algorithms 437
Jan A. Hiss and Gisbert Schneider
18.1 Template-Based Design 437
18.2 Nature-Inspired Optimization 441
18.2.1 Evolutionary Algorithms 444
18.2.2 Particle Swarm Optimization 446
18.2.3 Ant Colony Optimization 449
18.3 Worked Example: De Novo Design of MHC-I Binding Peptides by Ant Colony Optimization 450
18.4 Chemical Modification 456
18.4.1 Backbone Cyclization 456
18.4.2 Stapling 458
18.4.3 End-Capping 458
18.4.4 Sugar-Coating 459
18.5 Conclusions and Outlook 460
Acknowledgments 461
References 461

19 De Novo Computational Protein Design 467
Jeffery G. Saven
19.1 Introduction 467
19.2 Elements of Computational Protein Design 470
19.2.1 Target Structures 470
19.2.2 Degrees of Freedom: Amino Acids and Side-Chain Conformations 470
19.2.2.1 Amino Acids 470
19.2.2.2 Side-Chain Conformations 471
19.2.3 Energy Functions 471
19.2.4 Solvation 472
19.2.5 Foldability Criteria and Negative Design 472
19.2.6 Sequence Search and Characterization 473
19.2.6.1 Monte Carlo 473
19.2.6.2 Dead-End Elimination 474
19.2.6.3 Mean Field Theory 475
19.2.6.4 Probabilistic Approach 475
19.3 Efforts in Theoretically Guided Protein Design 477
19.3.1 Toward Catalysis, Redox Activity, and Enzymes 477
19.3.2 De Novo Design and Redesign 478
19.3.3 Protein Reengineering 479
19.3.4 Cofactors and Nonbiological Protein Assemblies 480
19.3.5 Membrane Proteins 481
19.3.6 Protein–Protein Interactions and Protein Assemblies 483
19.4 Conclusion 485
Acknowledgments 485
References 486

20 De Novo Design of Nucleic Acid Structures 495
Barbara Saccá, Andreas Sprengel, and Udo Feldkamp
20.1 Introduction 495
20.2 DNA-Branched Structures 499
20.2.1 De Novo Design of DNA Junctions 499
20.2.2 Tile-to-Tile Binding 504
20.3 Scaffolded DNA Origami Design 505
20.3.1 Monolayer DNA Origami 506
20.3.1.1 Two-Dimensional Structures 506
20.3.1.2 Three-Dimensional Structures 509
20.3.2 Multilayer DNA Origami 509
20.4 Alternative DNA Designs: between Junctions and Origami 511
20.5 Conclusions 514
Acknowledgments 515
References 515

21 RNA Aptamer Design 519
Cindy Meyer, Ulrich Hahn, and Andrew E. Torda
21.1 Aptamers and Design 519
21.2 Riboswitches and Aptamers 520
21.3 SELEX 521
21.3.1 Introduction 521
21.3.2 The Method 522
21.3.3 Technical Challenges and Recent Developments in SELEX 526
21.4 Speeding Up SELEX by Computational Methods 526
21.4.1 Design of Structures 529
21.5 Structures and Probing Methods 530
21.6 Functional Analyses (In Vitro and In Vivo) 532
21.7 Problems 533
21.8 Future Perspectives 535
References 536

Index 543
List of Contributors

Ruben Abagyan
University of California
San Diego
Skaggs School of Pharmacy and
Pharmaceutical Sciences
9500 Gilman Drive
La Jolla, CA 92093
USA

Rommie E. Amaro
University of California San
Diego
Department of Chemistry and
Biochemistry
9500 Gilman Drive
Mail Code 0340
La Jolla, CA 92093-0340
USA

Karl-Heinz Baringhaus
Sanofi-Aventis Deutschland
Chemical Science/Drug Design
Gebäude H 831
65926 Frankfurt
Germany

Jérémy Besnard
University of Dundee
Division of Biological Chemistry
and Drug Discovery
College of Life Sciences
Dow Street
Dundee DD1 3DF
UK

Jérémie Besnard
University of Dundee
Division of Biological Chemistry
and Drug Discovery
College of Life Sciences
Dow Street
Dundee DD1 3DF
UK

Jérémie Besnard
Ex Scientia Ltd
14 City Quay
Dundee DD1 3JA
UK

Julian Blagg
Cancer Research UK Cancer
Therapeutics Unit
Division of Cancer Therapeutics
The Institute of Cancer Research
15 Cotswold Road
Sutton SM2 5NG
UK

Michael J. Bodkin
Medicinal Chemistry
Eli Lilly UK
Erl Wood Manor
Windlesham
Surrey GU20 6PH
UK
Nathan Brown
Cancer Research UK Cancer Therapeutics Unit
Division of Cancer Therapeutics
The Institute of Cancer Research
15 Cotswold Road
Sutton SM2 5NG
UK

Lu Chen
University of Texas
MD Anderson Cancer Center
Department of Experimental Therapeutics
Integrated Molecular Discovery Laboratory
1901 East Road
Unit 1950
Houston, TX 77054
USA

Richard D. Cramer
Tripos Associates
1699 South Hanley Road
St. Louis, MO 63144
USA

Jacob D. Durrant
University of California San Diego
Department of Chemistry and Biochemistry
9500 Gilman Drive
Mail Code 0340
La Jolla, CA 92093-0340
USA

Darren Fayne
Trinity College Dublin
School of Biochemistry and Immunology
Trinity Biomedical Sciences Institute
152–160 Pearse Street
Dublin 2
Ireland

Udo Feldkamp
Technical University of Dortmund
Faculty of Chemistry
Otto-Hahn Street 6
D-44227 Dortmund
Germany

Nicholas C. Firth
Cancer Research UK Cancer Therapeutics Unit
Division of Cancer Therapeutics
The Institute of Cancer Research
15 Cotswold Road
Sutton SM2 5NG
UK

Valerie J. Gillet
Information School
University of Sheffield
Regent Court
211 Portobello Street
Sheffield S1 4DP
UK

Meir Glick
Novartis Institutes for BioMedical Research
Center for Proteomic Chemistry
250 Massachusetts Avenue
Cambridge, MA 02139
USA
Ulrich Hahn  
University of Hamburg  
MIN-Faculty, Chemistry Dept  
Institute for Biochemistry and Molecular Biology  
Martin-Luther-King Platz 6  
D-20146 Hamburg  
Germany  

Michael M. Hann  
Chemical Sciences  
Molecular Discovery Research  
GSK Medicines Research Centre  
Stevenage SG1 2NY  
UK  

and  
Ex Scientia Ltd  
14 City Quay  
Dundee DD1 3JA  
UK  

Dimitar Hristozov  
Medicinal Chemistry  
Eli Lilly UK  
Erl Wood Manor  
Windlesham  
Surrey GU20 6PH  
UK  

Qi Huang  
Sichuan University  
State Key Laboratory of Biotherapy and Cancer Center  
West China Hospital  
No. 17, Sec 3  
Renmin Road South, Chengdu  
Sichuan, 610041  
China  

Roderick E. Hubbard  
University of York  
Department of Chemistry  
YSBL  
Heslington  
York YO10 5DD  
UK  

and  
Vernalis (R&D) Ltd  
Granta Park  
Cambridge, CB21 6GB  
UK  

Markus Hartenfeller  
Novartis Pharma AG  
Forum 1  
Novartis Institutes for BioMedical Research  
Novartis Campus  
CH-4056 Basel  
Switzerland  

Jan A. Hiss  
Swiss Federal Institute of Technology (ETH)  
Department of Chemistry and Applied Biosciences  
Institute of Pharmaceutical Sciences  
Wolfgang-Pauli-Strasse 10  
8093 Zurich  
Switzerland  

Andrew L. Hopkins  
University of Dundee  
Division of Biological Chemistry and Drug Discovery  
College of Life Sciences  
Dow Street  
Dundee, DD1 3DF  
UK
List of Contributors

Edgar Jacoby
Novartis Pharma AG
Forum 1
Novartis Institutes for BioMedical Research
Novartis Campus
CH-4056 Basel
Switzerland

Peter S. Kutchukian
Novartis Institutes for BioMedical Research
Center for Proteomic Chemistry
250 Massachusetts Avenue
Cambridge, MA 02139
USA

Florian Lauck
University of Hamburg
Center for Bioinformatics
Research Group for Computational Molecular Design
Bundesstraße 43
D-20146 Hamburg
Germany

Richard Law
Evotec (UK) Ltd
114 Innovation Drive
Milton Park
Abingdon
Oxfordshire OX14 4RZ
UK

Andrew R. Leach
Chemical Sciences
Molecular Discovery Research
GSK Medicines Research Centre
Stevenage SG1 2NY
UK

Uta Lessel
Boehringer Ingelheim Pharma GmbH & Co. KG
Lead Identification and Optimization Support
Computational Chemistry
Birkendorfer Straße 65
D-88397 Biberach an der Riss
Germany

Eugen Lounkine
Novartis Institutes for BioMedical Research
Center for Proteomic Chemistry
250 Massachusetts Avenue
Cambridge, MA 02139
USA

Michael Mazanetz
Evotec (UK) Ltd
114 Innovation Drive
Milton Park
Abingdon
Oxfordshire OX14 4RZ
UK

Cindy Meyer
The Rockefeller University
Howard Hughes Medical Institute
Laboratory of RNA Molecular Biology
1230 York Ave
New York, NY 10065
USA

Matthias Rarey
University of Hamburg
Center for Bioinformatics
Research Group for Computational Molecular Design
Bundesstraße 43
D-20146 Hamburg
Germany
Steffen Renner
Novartis Pharma AG
Forum 1
Novartis Institutes for BioMedical Research
Novartis Campus
CH-4056 Basel
Switzerland

Clarisse Gravina Ricci
State University of Campinas – UNICAMP
Institute of Chemistry
Cx. P. 6154
Campinas
São Paulo 13083–970
Brazil

and

University of California
San Diego
Skaggs School of Pharmacy and Pharmaceutical Sciences
9500 Gilman Drive
La Jolla, CA 92093
USA

Barbara Saccà
University of Duisburg-Essen
Department of Bionanotechnology
Center for Medicinal Biotechnology
Faculty of Biology
Universitätstraße 2
D-44117 Essen
Germany

Jeffery G. Saven
University of Pennsylvania
Department of Chemistry
231 South 34th Street
Philadelphia, PA 19104
USA

Gisbert Schneider
Swiss Federal Institute of Technology (ETH)
Department of Chemistry and Applied Biosciences
Institute of Pharmaceutical Sciences
Wolfgang-Pauli-Strasse 10
8093 Zürich
Switzerland

Eugene I. Shakhnovich
Harvard University
Chemistry and Chemical Biology
12 Oxford Street
Cambridge, MA 02138
USA

Andreas Sprengel
University of Duisburg-Essen
Department of Bionanotechnology
Center for Medicinal Biotechnology
Faculty of Biology
Universitätstraße 2
D-44117 Essen
Germany

Alla Srinivas Reddy
University of Texas
MD Anderson Cancer Center
Department of Experimental Therapeutics
Integrated Molecular Discovery Laboratory
1901 East Road
Unit 1950
Houston, TX 77054
USA
List of Contributors

Andrew E. Torda
University of Hamburg
Center for Bioinformatics
Bundesstrasse 43
D-20146 Hamburg
Germany

Salla I. Virtanen
Harvard University
Chemistry and Chemical Biology
12 Oxford Street
Cambridge, MA 02138
USA

Wen-Jing Wang
Sichuan University
State Key Laboratory of Biotherapy and Cancer Center
West China Hospital
No. 17, Sec 3
Renmin Road South, Chengdu
Sichuan, 610041
China

Yvonne Westermaier
University of York
Department of Chemistry
YSBL
Heslington
York YO10 5DD
UK

Mark Whittaker
Evotec (UK) Ltd
114 Innovation Drive
Milton Park
Abingdon
Oxfordshire OX14 4RZ
UK

Sheng-Yong Yang
Sichuan University
State Key Laboratory of Biotherapy and Cancer Center
West China Hospital
No. 17, Sec 3
Renmin Road South, Chengdu
Sichuan, 610041
China

Shuxing Zhang
University of Texas
MD Anderson Cancer Center
Department of Experimental Therapeutics
Integrated Molecular Discovery Laboratory
1901 East Road
Unit 1950
Houston, TX 77054
USA

and

Universitat de Barcelona
Facultat de Farmàcia
Departament de Fisicoquímica
and Institut de Biomedicina
Computational Biology and Drug Design Group
Avinguda Joan XXIII, s/n
08028 Barcelona
Spain
Foreword

The history of de novo drug design, which is concerned primarily with the use of computers to design new active drug compounds, may as well be called the history of computer-aided drug design (CADD). Quantitative structure–activity relationship (QSAR) studies were a prominent feature of the drug design process until the second half of the 1980s. QSAR studies provide an effective technique for analyzing the correlations between molecular structure and biological activity, and can still be used as a powerful approach during the lead optimization phase of a drug discovery program. For the purposes of lead generation, however, QSAR studies cannot be used, for example, to design a molecule with a different molecular skeleton.

To overcome these issues, de novo drug design was introduced in the 1990s. Although a variety of different de novo drug design software suites have been developed, they are invariably difficult to at the practical level for real drug design. It is noteworthy that successful examples of drug design using these tools could not be found during those early days, and the use and general perception of de novo drug design consequently went into decline following its peak usage in the mid-1990s. This decline in the use of de novo drug design was attributed to scientists focusing on the magnitude of the computational binding strength with the target receptor, while ignoring the drug-like properties and the synthetic tractability of the designed compounds.

Following of from the pivot role of CADD in in silico virtual screening, de novo drug design has reappeared in the form of lead hopping or scaffold hopping during the first half of the 2000s. This reappearance owes a lot to the compound libraries generated using combinatorial chemistry and chemoinformatics technologies. The recent progress of de novo drug design was reviewed by Prof Schneider [1], where 36 kinds of software were classified according to their methodology. Furthermore, in a review by Prof Kunchukian [2], the 20 latest types of software were comprehensively added. According to Kunchukian’s count, the number of the reports published every year from 2005 through 2008 increased by five to six reports until it eventually doubled in 2009. The numbers then continued to increase at the same pace afterward. The big difference in the recent popularity of de novo drug design relative to its initial release in the 1990s relates to the number of research reports in which the compounds designed on the computer were actually
synthesized and evaluated. This shows that the use of *de novo* drug design software for drug development spot has reached a practical level. Interestingly, there are now more *de novo* drug design software suites available than there are *in silico* virtual screening programs. That is, the technology of *de novo* drug design is not fixed and the software has many advantages and disadvantages. In other words, *de novo* drug design has a hidden potential for further developments.

The number of possible combination of atoms in organic compounds (chemical space) is vast, and the number is said to be the sixtieth power of 10. *De novo* drug design is a combination of optimization problem that enables the user to find the most promising compound out of this vast chemical space. A variety of different optimization algorithms have been devised, including the evolutionary algorithm, Monte Carlo simulation, taboo search, depth-first search, breadth-first search, and the A* algorithm. As *de novo* drug design software adopts various algorithms, the software is flooded with many candidates.

Taken against this background, the publication of this special edition of “*De novo* Molecular Design” appears to be particularly timely. This book itself is dedicated to the concepts and ideas for *de novo* drug design. The potentials and limitations of the relevant techniques are critically discussed and comprehensively exemplified in 21 chapters by distinguished authors from both academia and the pharmaceutical industry. A series of well-defined chapters follow the first exciting and challenging chapter, “*De novo* design: from models to molecules”, with examples including structure-, fragment-, pharmacophore-, QSAR-, reaction-, polypharmacology-, combinatorial-, and biosteric-based *de novo* designs. As a scientist keen to recommend the use of *de novo* drug design in the drug discovery process, I am convinced that readers will be able to successfully apply these *de novo* design methods to their own drug design projects and produce many innovative compounds for the pharmaceutical drug market. It is my central hope that this book will be helpful and be used in the same way as an encyclopedia when hints and ideas are needed during the drug design process.

Tokyo, April 2013

*Prof Kimito Funatsu*

*Department of Chemical System Engineering*

*The University of Tokyo, Tokyo, Japan*

**References**


Preface

This book builds on the legacy of many bright minds. It does not claim completeness or truth. Its intention merely is to inspire readers to critically and creatively explore the possibilities of de novo molecular design for drug discovery and chemical biology. I am most grateful to all authors for contributing truly exciting 21 chapters. Their willingness and thoroughness allowed us to compile a formidable collection of ideas and reports on the various aspects of computer-assisted molecular design. Special thanks go to Prof Kimito Funatsu, who shares his thoughts on de novo design in the Foreword. I am equally grateful to my colleagues who agreed to act as impartial reviewers, and through their personal advice helped me not to go over the top. My beloved wife was very lenient toward me during the preparation of this volume (and not just then). Dr Heike Nöthe and Dr Frank Weinreich from Wiley-VCH did a great job supporting me in the editing process and ensured swift and professional book production. Persistent challenge by my research team at ETH helped me focus on some tough scientific questions and come up with hopefully useful answers.

This book starts off with a general overview of the scientific pillars of molecular design. In the subsequent chapters, renowned experts from industry and academia alike provide their views on the drug discovery process and the role of de novo design, receptor- and ligand-based approaches, the nature of macromolecular structure and ligand–receptor interaction, chemical space navigation, combinatorial- and fragment-based design principles, rigorous physical approaches to solve the scoring problem in drug design, and the automated generation of bioactive peptides, proteins, and nucleic acids as potential drugs of the future. I have structured the contributions so that they are attuned to one another and demonstrate the various ideas and technological concepts in a well-defined collation. This book is meant to be read from cover to cover. Nevertheless, all chapters stand on their own, and the interested reader may cherry-pick favorites. Consequently, slight redundancy of contents was unavoidable and has intentionally been kept to ensure that each chapter represents the authors’ individual views on a topic, and at the same time allows the reader to learn about different thoughts and opinions.

As I have had the great privilege to witness the rise of computer-assisted de novo design from its humble beginnings to become mainstream science, I am pleased to see these fascinating techniques now being broadly applied in drug
discovery and chemical biology. There is still much more to come and to expect from the amalgamation of technologies and complementary scientific thinking. I am convinced that only by constantly keeping an open mind for surprising fresh ideas and unexpected revelations will we be able to make continuous progress in molecular design. This also means that some of our “old beliefs” might need a critical overhaul, and some should better be discarded to make way for new and improved concepts that will enable researchers to conceive of innovative algorithms for molecule construction, scoring, and chemical space navigation.

Zürich, April 2013

Gisbert Schneider
Innovative bioactive agents fuel sustained drug discovery and the development of new medicines. Future success in chemical biology and pharmaceutical research alike will fundamentally rely on the combination of advanced synthetic and analytical technologies that are embedded in a theoretical framework that provides a rationale for the interplay between chemical structure and biological effect. A driving role in this setting falls on leading edge concepts in computer-assisted molecular design, by providing access to a virtually infinite source of novel druglike compounds and guiding experimental screening campaigns. In this chapter, we present concepts and ideas for the representation of molecular structure, suggest predictive models of structure–activity relationships, and discuss approaches that have proved their usefulness and will contribute to future drug discovery by generating innovative bioactive agents. We also highlight some of the current prohibitive aspects of fully automated de novo design that will require attention for future methodological breakthroughs. This chapter provides an introduction to important pillars of de novo drug design, whereas the subsequent contributions presented in this book offer in-depth treatments of current trends, methods, and approaches together with numerous practical examples. We are confident that the reading will inspire.

1.1 Molecular Representation

Ever since the first atomic models of molecules have been conceived, scientists have used such models, and their associated concepts and language, to come up with innovative chemical agents that possess sought properties [2]. So far, we tend to think of a molecule in terms of sticks and balls when it comes...
Figure 1.1 Atomic models of molecular structure as depicted in John Dalton’s seminal book entitled *A New System of Chemical Philosophy* (1808). Panel (a) presents the “arbitrary signs chosen to represent the several chemical elements or ultimate particles.” Panel (b) might be considered as an early molecular design study, as it depicts Dalton’s view of various arrangements of water molecules. Note the similarity between these archaic philosophies and contemporary molecular models.

to visualize chemical structure. No doubt, simplistic representations have their justification for describing certain aspects of molecular constitution, configuration, and conformation and provide an intuitive access to “molecular architecture” (Figure 1.1). However, they fall far short of relating functional aspects to the objects we recognize as molecules. In the end, it is the desired function we wish to get from a molecular structure. “Form follows function” – this credo of modern architecture and industrial design is equally valid for molecular design, in particular in medicinal chemistry and chemical biology striving for new chemical entities (NCEs) as biologically active lead compounds and eventually future drugs.

Ideally, one would like to obtain a compound with a desired function directly from a design hypothesis, for example, a mathematical model that serves as a
blueprint, without the need for exhaustive screening and meticulous optimization. In fact, de novo design means generating new molecules with desired properties “from scratch.” The concept of using transition functions that assign new states to objects, thereby observing emergent system properties [3, 4], has been well researched in fields such as complexity analysis, dynamical system, game theory, and systems biology [5]. In molecular design, we use models of the molecular world and expect a trustworthy model to correctly reflect aspects of the real world, so it can be used for predicting new molecules that possess the target property reflected in the model (Figure 1.2a). De novo design theory is tightly related to solving the inverse quantitative structure–activity relationship (SAR) problem or—to paraphrase from a philosophical point of view—finding the “Urbild,” 1) that is, the structural archetype associated with a molecular representation. In terms of mathematics, one tries to find an element $x$ that is related to the value $\xi$: $\xi = f(x)$. In molecular design, $x$ is a molecular structure from the set of all compounds (usually referred to as chemical space) and $\xi$ is the representation (descriptor) of $x$ computed by function $f$ [8]. Typically, the representation of a compound is a real numbered value or set of values (vector representation), although other, for example, symbolic forms of representations have been suggested [9]. It is essential to realize that the representation of a chemical structure is always uniquely defined by the mapping function $f$, while there may exist—if defined—many possibly infinite numbers of molecules that have the exact same descriptor values (Figure 1.2b). As a basic illustration of this important point, consider the total charge descriptor of a molecule containing $N$ atoms, which is computed as $\xi = f(x) = \sum_{i=1}^{N} q_i$, where $q_i$ is the partial charge of atom $i$. Accordingly, it is easy to determine the total charge for a given molecular structure, but it there may be numerous chemically feasible compounds featuring the same total charge.

Generally speaking, molecular de novo design aims at generating new compounds that can be mapped to well-defined, preferred representations, that is, sets of descriptor values that characterize compounds with the desired biological or pharmacological activity. The challenge hereby is twofold, namely to

1) define a set of mathematical functions that characterize compounds with desired properties (i.e., they belong to the same equivalence class), and
2) for a given molecular representation, find corresponding Urbild compounds.

Consequently, as a prerequisite for successful design, we need an adequate representation of molecular structures and their physicochemical properties to allow the extraction of features that are responsible for a certain compound property or pharmacological activity (=function). Ideally, we need to understand the behavior of a molecule in different environments (e.g., in solution and in complex with a receptor) over time. Consequent physical treatment of molecular properties and dynamics can in principle be achieved based on solutions of the

1) The Urbild concept has multiple references and partly different meaning in mathematics and philosophy. See, for example, Refs. [6, 7].
Figure 1.2 (a) Models of chemical space. (Adapted from Ref. [4].) Molecules in chemical space (real world) are lumped into an equivalence class (dotted circle) according to a structure–activity relationship model. In computer-based molecular design, appropriate algorithms act as transition functions so that changes of model states are faithfully reflected in the adaptation of molecular structure and function. (b) Molecular representation and design. A function \( f : X \rightarrow Y \) transforms a molecular structure \( x \) to its corresponding molecular descriptor \( \xi \). One may call \( x \) the “Urbild” of \( \xi \). In molecular design applications, molecules are often mapped to numerical descriptor values by surjective functions, meaning that multiple elements of \( X \) might be turned into the same element of \( Y \) by applying \( f \). This property of many molecular descriptor sets is exploited by \textit{de novo} design, which aims at finding new molecules in \( X \) that can be mapped to pharmacologically meaningful representations.

Schrödinger equation (Eq. (1.1)).

\[
\hat{H} \Psi = E \Psi
\]  

(1.1)

where \( \hat{H} \) is the Hamilton operator defining the operations that need to be performed with the set of wave functions \( \Psi \) (psi) of the particles of a molecular system and