

## Contents

<b>List of Contributors</b>	<i>XVII</i>
<b>Preface</b>	<i>XXIII</i>
<b>A Personal Foreword</b>	<i>XXV</i>

### Part One Principles 1

<b>1</b>	<b>Virtual Screening of Chemical Space: From Generic Compound Collections to Tailored Screening Libraries</b>	<b>3</b>
	<i>Markus Boehm</i>	
1.1	Introduction	3
1.2	Concepts of Chemical Space	4
1.3	Concepts of Druglikeness and Leadlikeness	6
1.4	Diversity-Based Libraries	8
1.4.1	Concepts of Molecular Diversity	8
1.4.2	Descriptor-Based Diversity Selection	9
1.4.3	Scaffold-Based Diversity Selection	12
1.4.4	Sources of Diversity	13
1.5	Focused Libraries	15
1.5.1	Concepts of Focused Design	15
1.5.2	Ligand-Based Focused Design	16
1.5.3	Structure-Based Focused Design	17
1.5.4	Chemogenomics Approaches	18
1.6	Virtual Combinatorial Libraries and Fragment Spaces	20
1.7	Databases of Chemical and Biological Information	21
1.8	Conclusions and Outlook	24
1.9	Glossary	25
	References	26

<b>2</b>	<b>Preparing and Filtering Compound Databases for Virtual and Experimental Screening</b>	<b>35</b>
	<i>Maxwell D. Cummings, Éric Arnoult, Christophe Buyck, Gary Tresadern, Ann M. Vos, and Jörg K. Wegner</i>	
2.1	Introduction	35
2.2	Ligand Databases	36
2.2.1	Chemical Data Structures	36
2.2.2	3D Conformations	38
2.2.3	Data Storage	39
2.2.4	Workflow Tools	39
2.2.5	Past Reviews and Recent Papers	40
2.3	Considering Physicochemical Properties	42
2.3.1	Druglikeness	42
2.3.2	Leadlikeness and Beyond	43
2.4	Undesirables	43
2.4.1	Screening Artifacts	44
2.4.2	Pharmacologically Promiscuous Compounds	45
2.5	Property-Based Filtering for Selected Targets	46
2.5.1	Antibacterials	47
2.5.2	CNS	49
2.5.3	Protein–Protein Interactions	51
2.6	Summary	52
	References	53
<b>3</b>	<b>Ligand-Based Virtual Screening</b>	<b>61</b>
	<i>Herbert Koeppen, Jan Kriegl, Uta Lessel, Christofer S. Tautermann, and Bernd Wellenzohn</i>	
3.1	Introduction	61
3.2	Descriptors	62
3.3	Search Databases and Queries	67
3.3.1	Selection of Reference Ligands	67
3.3.2	Preparation of the Search Database	68
3.4	Virtual Screening Techniques	68
3.4.1	Similarity Searches	69
3.4.1.1	Similarity Measures	69
3.4.1.2	Practice of Similarity Searches	69
3.4.1.3	Selection of Descriptors	71
3.4.1.4	Data Fusion	72
3.4.2	Similarity Searches in Very Large Chemical Spaces	72
3.4.3	Machine Learning in Virtual Screening	74
3.4.3.1	Unsupervised Methods	75
3.4.3.2	Supervised Methods	75
3.4.3.3	Selected Techniques	76
3.4.3.4	Machine Learning Applications for Virtual Screening	78
3.4.4	Validation of Methods and Prediction of Success	78

3.5	Conclusions	79
	References	80
<b>4</b>	<b>The Basis for Target-Based Virtual Screening: Protein Structures</b>	<b>87</b>
	<i>Jason C. Cole, Oliver Korb, Tjelvar S.G. Olsson, and John Liebeschuetz</i>	
4.1	Introduction	87
4.2	Selecting a Protein Structure for Virtual Screening	87
4.2.1	Why Are There Errors in Crystal Structures?	87
4.2.2	Possible Problems That May Occur in a Crystal Structure	91
4.2.2.1	Entirely Incorrect Models	91
4.2.2.2	Sequencing Errors	91
4.2.2.3	Misplaced Side Chains	91
4.2.2.4	Structural Disorder	92
4.2.2.5	Poorly Modeled Cofactors and Ligands	92
4.2.2.6	Erroneous Solvent	94
4.2.3	Structural Relevance	95
4.2.3.1	The Biologically Relevant Unit and Crystal Packing	95
4.2.4	Critical Evaluation of Models: Recognizing Issues in Structures	98
4.3	Setting Up a Protein Model for vHTS	101
4.3.1	Binding Site Definition	101
4.3.2	Protonation	104
4.3.3	Treatment of Solvent in Docking	104
4.3.4	Use of Protein-Based Constraints in Docking	105
4.3.5	Protein Flexibility	106
4.3.5.1	Pose Prediction	107
4.3.5.2	Virtual Screening	108
4.4	Summary	109
4.5	Glossary of Crystallographic Terms	110
4.5.1	R-Factor	110
4.5.2	Resolution	110
4.5.3	2mFo-DFc Map	110
	References	110
<b>5</b>	<b>Pharmacophore Models for Virtual Screening</b>	<b>115</b>
	<i>Patrick Markt, Daniela Schuster, and Thierry Langer</i>	
5.1	Introduction	115
5.2	Compilation of Compounds	116
5.2.1	Chemical Structure Generation	116
5.2.2	Conformational Analysis	116
5.3	Pharmacophore Model Generation	117
5.3.1	State of the Art	117
5.3.2	Structure-Based Methods	117
5.3.3	Ligand-Based Methods	118
5.3.4	Limitations of Ligand-Based Methods	119
5.4	Validation of Pharmacophore Models	119

5.4.1	Chemical Databases for Validation	119
5.4.2	Enrichment Assessment	121
5.4.3	Enrichment Metrics	122
5.4.4	Receiver Operating Characteristic Curve Analysis	124
5.4.5	Area Under the ROC Curve	125
5.5	Pharmacophore-Based Screening	127
5.5.1	DS CATALYST	128
5.5.2	UNITY (GALAHAD/GASP)	128
5.5.3	LIGANDSCOUT	129
5.5.4	MOE	130
5.5.5	PHASE	130
5.6	Postprocessing of Pharmacophore-Based Screening Hits	131
5.6.1	Lead- and Druglikeness	131
5.6.2	Structural Similarity Analysis	131
5.7	Pharmacophore-Based Parallel Screening	132
5.8	Application Examples for Synthetic Compound Screening	133
5.8.1	17 $\beta$ -Hydroxysteroid Dehydrogenase 1 Inhibitors	133
5.8.2	Cannabinoid Receptor 2 (CB <sub>2</sub> ) Ligands	134
5.8.3	Further Application Examples	136
5.9	Application Examples for Natural Product Screening	136
5.9.1	Cyclooxygenase (COX) Inhibitors	139
5.9.2	Sigma-1 ( $\sigma_1$ ) Receptor Ligands	139
5.9.3	Acetylcholinesterase Inhibitors	140
5.9.4	Human Rhinovirus Coat Protein Inhibitors	141
5.9.5	Quorum-Sensing Inhibitors	141
5.9.6	Peroxisome Proliferator-Activated Receptor $\gamma$ Ligands	141
5.9.7	$\beta$ -Ketoacyl-Acyl Carrier Protein Synthase III Inhibitors	142
5.9.8	5-Lipoxygenase Inhibitors	142
5.9.9	11 $\beta$ -Hydroxysteroid Dehydrogenase Type 1 Inhibitors	142
5.9.10	Pharmacophore-Based Parallel Screening of Natural Products	143
5.10	Conclusions	143
	References	144

## 6 Docking Methods for Virtual Screening: Principles and Recent Advances 153

*Didier Rognan*

6.1	Principles of Molecular Docking	153
6.1.1	Sampling Degrees of Freedom of the Ligand	154
6.1.1.1	Generation of Multiconformer Ligand Libraries	154
6.1.1.2	Incremental Construction	154
6.1.1.3	Stochastic Methods	155
6.1.2	Scoring Ligand Poses	156
6.1.2.1	Empirical Scoring Functions	156
6.1.2.2	Knowledge-Based Potential of Mean Force	156

6.1.2.3	Force Fields	157
6.1.2.4	Critical Evaluation of Scoring Functions	157
6.2	Docking-Based Virtual Screening Flowchart	158
6.2.1	Ligand Setup	158
6.2.2	Protein Setup	159
6.2.3	Docking	160
6.2.4	Postdocking Analysis	161
6.3	Recent Advances in Docking-Based VS Methods	162
6.3.1	Novel Docking Algorithms	162
6.3.2	Fragment Docking	164
6.3.3	Postdocking Refinement	164
6.3.3.1	Rescoring with Rigorous Scoring Functions	164
6.3.3.2	Topological Scoring by Protein–Ligand Interaction Fingerprint (IFP)	165
6.3.4	Addressing Protein Flexibility	166
6.3.5	Solvated or Dry?	168
6.4	Future Trends in Docking	168
	References	169

## Part Two Challenges 177

<b>7</b>	<b>The Challenge of Affinity Prediction: Scoring Functions for Structure-Based Virtual Screening</b>	<b>179</b>
	<i>Christoph Sotriffer and Hans Matter</i>	
7.1	Introduction	179
7.2	Physicochemical Basis of Protein–Ligand Recognition	180
7.3	Classes of Scoring Functions	185
7.3.1	Force Field-Based Methods	185
7.3.2	Empirical Scoring Functions	189
7.3.3	Knowledge-Based Scoring Functions	191
7.4	Interesting New Approaches to Scoring Functions	192
7.4.1	Improved Treatment of Hydrophobicity and Dehydration	192
7.4.2	Development and Validation of SFCscore	194
7.4.3	Consensus Scoring	195
7.4.4	Tailored Scoring Functions	196
7.4.5	Structural Interaction Fingerprints	199
7.5	Comparative Assessment of Scoring Functions	200
7.6	Tailoring Scoring Strategies in Virtual Screening	203
7.6.1	Toward a Strategy for Applying Scoring Functions	203
7.6.2	Retrospective Validation Prior to Prospective Virtual Screening	204
7.6.3	Lessons Learned: Improvements in Scoring Evaluations	205
7.6.4	Postfiltering Results of Virtual Screenings	205
7.7	Caveats for Development of Scoring Functions	206
7.7.1	General Points	206
7.7.2	Biological Data	207

7.7.3	Structural Data on Protein–Ligand Complexes and Decoy Data Sets	207
7.7.4	Cooperativity and Other Model Deficiencies	208
7.8	Conclusions	209
	References	210
<b>8</b>	<b>Protein Flexibility in Structure-Based Virtual Screening: From Models to Algorithms</b>	<b>223</b>
	<i>Angela M. Henzler and Matthias Rarey</i>	
8.1	How Flexible Are Proteins? – A Historical Perspective	223
8.1.1	Ligand Binding Is Coupled with Protein Conformational Change	223
8.1.2	Types of Flexibility	224
8.2	Flexible Protein Handling in Protein–Ligand Docking	225
8.2.1	Docking Following Conformational Selection	227
8.2.1.1	Protein Flexibility Analysis and Protein Ensemble Generation	227
8.2.1.2	Ensemble-Based Docking Techniques	228
8.2.2	Induced Fit Docking: Single-Structure-Based Docking Techniques	231
8.2.2.1	Consecutive Ligand and Protein Conformational Change	232
8.2.2.2	Simultaneous Ligand and Protein Conformational Change	234
8.2.3	Integrated Docking Approaches	235
8.3	Flexible Protein Handling in Docking-Based Virtual Screening	236
8.3.1	Efficiency of Fully Flexible Docking Approaches in Retrospective	237
8.3.2	Discrimination of Binders and Nonbinders	238
8.4	Summary	238
	References	239
<b>9</b>	<b>Handling Protein Flexibility in Docking and High-Throughput Docking: From Algorithms to Applications</b>	<b>245</b>
	<i>Claudio N. Cavasotto</i>	
9.1	Introduction: Docking and High-Throughput Docking in Drug Discovery	245
9.2	The Challenge of Accounting for Protein Flexibility in Docking	246
9.2.1	Theoretical Understanding of the Problem	246
9.2.2	Docking Failures Due to Protein Flexibility	247
9.3	Accounting for Protein Flexibility in Docking-Based Drug Discovery and Design	250
9.3.1	Receptor Ensemble-Based Docking Methods	252
9.3.2	Single-Structure-Based Docking Methods	253
9.3.3	Multilevel Methods	256
9.3.4	Homology Modeling	257
9.4	Conclusions	257
	References	258

<b>10</b>	<b>Consideration of Water and Solvation Effects in Virtual Screening</b>	<b>263</b>
	<i>Johannes Kirchmair, Gudrun M. Spitzer, and Klaus R. Liedl</i>	
10.1	Introduction	263
10.2	Experimental Approaches for Analyzing Water Molecules	266
10.3	Computational Approaches for Analyzing Water Molecules	271
10.3.1	Molecular Dynamics Simulations	271
10.3.2	Empirical and Implicit Considerations of Solvation Effects	274
10.4	Water-Sensitive Virtual Screening: Approaches and Applications	275
10.4.1	Protein–Ligand Docking	275
10.4.2	Pharmacophore Modeling	278
10.5	Conclusions and Recommendations	281
	References	282
<b>Part Three</b>	<b>Applications and Practical Guidelines</b>	<b>291</b>
<b>11</b>	<b>Applied Virtual Screening: Strategies, Recommendations, and Caveats</b>	<b>293</b>
	<i>Dagmar Stumpfe and Jürgen Bajorath</i>	
11.1	Introduction	293
11.2	What Is Virtual Screening?	293
11.3	Spectrum of Virtual Screening Approaches	294
11.4	Molecular Similarity as a Foundation and Caveat of Virtual Screening	295
11.5	Goals of Virtual Screening	296
11.6	Applicability Domain	297
11.7	Reference and Database Compounds	299
11.8	Biological Activity versus Compound Potency	300
11.9	Methodological Complexity and Compound Class Dependence	301
11.10	Search Strategies and Compound Selection	302
11.11	Virtual and High-Throughput Screening	304
11.12	Practical Applications: An Overview	306
11.13	LFA-1 Antagonist	307
11.13.1	Similarity Searching	308
11.13.2	Results and Further Calculations	309
11.14	Selectivity Searching	310
11.14.1	Selectivity Searching for Cathepsin K-Selective Inhibitors	311
11.14.2	Selectivity Searching with 2D Fingerprints	312
11.14.3	Identification of Selective Inhibitors	313
11.15	Concluding Remarks	314
	References	315
<b>12</b>	<b>Applications and Success Stories in Virtual Screening</b>	<b>319</b>
	<i>Hans Matter and Christoph Sotriffer</i>	
12.1	Introduction	319
12.2	Practical Considerations	320

12.3	Successful Applications of Virtual Screening	321
12.3.1	Structure-Based Virtual Screening	322
12.3.1.1	Kinases	322
12.3.1.2	Proteases	324
12.3.1.3	Nuclear Receptors	325
12.3.1.4	Short-Chain Dehydrogenases	327
12.3.1.5	G Protein-Coupled Receptors (GPCRs)	327
12.3.1.6	Antiinfectives	331
12.3.1.7	Other Target Proteins	333
12.3.2	Structure-Based Library Design	336
12.3.3	Ligand-Based Virtual Screening	338
12.3.3.1	Ion Channels	339
12.3.3.2	Kinases	340
12.3.3.3	Nuclear Hormone Receptors	341
12.3.3.4	G Protein-Coupled Receptors (GPCRs)	342
12.3.3.5	Other Protein Targets	345
12.4	Conclusion	347
	References	348
<b>Part Four</b>	<b>Scenarios and Case Studies: Routes to Success</b>	<b>359</b>
<b>13</b>	<b>Scenarios and Case Studies: Examples for Ligand-Based Virtual Screening</b>	<b>361</b>
	<i>Trevor Howe, Daniele Bemporad, and Gary Tresadern</i>	
13.1	Introduction	361
13.2	1D Ligand-Based Virtual Screening	362
13.3	2D Ligand-Based Virtual Screening	363
13.3.1	Examples from the Literature	363
13.3.2	Applications at J&JPRD Europe	366
13.4	3D Ligand-Based Virtual Screening	368
13.4.1	Methods	370
13.4.2	3DLBVS Examples	372
13.4.2.1	CRF1 Antagonists	372
13.4.2.2	Ion Channel Antagonism	375
13.4.2.3	Metabotropic Glutamate Receptor	375
13.5	Summary	376
	References	377
<b>14</b>	<b>Virtual Screening on Homology Models</b>	<b>381</b>
	<i>Róbert Kiss and György M. Keserű</i>	
14.1	Introduction	381
14.2	Homology Models versus Crystal Structures: Comparative Evaluation of Screening Performance	382
14.2.1	Soluble Proteins	382
14.2.2	Membrane Proteins	392



14.3	Challenges of Homology Model-Based Virtual Screening	394
14.3.1	Level of Sequence Identity	395
14.3.2	Main-Chain Flexibility	396
14.3.3	Side-Chain Conformation: Induced Fit Effects of Ligands	396
14.3.4	Loop Modeling	397
14.4	Case Studies	399
14.4.1	Virtual Screening on the Homology Model of Histamine H4 Receptor	399
14.4.2	Virtual Screening on the Homology Model of Janus Kinase 2	402
	References	404
<b>15</b>	<b>Target-Based Virtual Screening on Small-Molecule Protein Binding Sites</b>	<b>411</b>
	<i>Ralf Heinke, Urszula Uciechowska, Manfred Jung, and Wolfgang Sippl</i>	
15.1	Introduction	411
15.1.1	Pharmacophore-Based Methods	412
15.1.2	Ligand Docking	412
15.1.3	Virtual Screening	413
15.1.4	Binding Free Energy Calculations	414
15.2	Structure-Based VS for Histone Arginine Methyltransferase PRMT1 Inhibitors	414
15.2.1	Structure-Based VS of the NCI Diversity Set	415
15.2.2	Pharmacophore-Based VS	417
15.3	Identification of Nanomolar Histamine H <sub>3</sub> Receptor Antagonists by Structure- and Pharmacophore-Based VS	422
15.3.1	Generation of Homology Model of the hH <sub>3</sub> R and hH <sub>3</sub> R Antagonist Complexes	423
15.3.2	Validation of the Homology Model by Docking Known Antagonists into the hH <sub>3</sub> R Binding Site	424
15.3.3	Pharmacophore-Based VS	425
15.3.4	Experimental Testing of the Identified Hits	429
15.3.5	Discussion of the Applied VS Strategies	429
15.4	Summary	431
	References	432
<b>16</b>	<b>Target-Based Virtual Screening to Address Protein–Protein Interfaces</b>	<b>435</b>
	<i>Olivier Sperandio, Maria A. Miteva, and Bruno O. Villoutreix</i>	
16.1	Introduction	435
16.2	Some Recent PPIM Success Stories	437
16.3	Protein–Protein Interfaces	438
16.3.1	Interface Pockets, Flexibility, and Hot Spots	440
16.3.2	Databases and Tools to Analyze Interfaces	442
16.4	PPIMs' Chemical Space and ADME/Tox Properties	442

16.5	Drug Discovery, Chemical Biology, and <i>In Silico</i> Screening Methods: Overview and Suggestions for PPIM Search	447
16.6	Case Studies	450
16.6.1	PPI Stabilizers: Superoxide Dismutase Type 1	450
16.6.2	PPI Inhibitors: Lck	452
16.6.3	Allosteric Inhibitors: Antitrypsin Polymerization	455
16.7	Conclusions and Future Directions	457
	References	458
<b>17</b>	<b>Fragment-Based Approaches in Virtual Screening</b>	<b>467</b>
	<i>Danzhi Huang and Amedeo Caflisch</i>	
17.1	Introduction	467
17.2	<i>In Silico</i> Fragment-Based Approaches	468
17.3	Our Approach to High-Throughput Fragment-Based Docking	470
17.3.1	Decomposition of Compounds into Fragments	471
17.3.2	Docking of Anchor Fragments	471
17.3.3	Flexible Docking of Library Compounds	472
17.3.4	LIECE Binding Energy Evaluation	472
17.3.5	Consensus Scoring	475
17.3.6	<i>In Silico</i> Screening Campaigns	475
17.3.7	West Nile Virus NS3 Protease (Flaviviral Infections)	475
17.3.8	EphB4 Tyrosine Kinase (Cancer)	477
17.4	Lessons Learned from Our Fragment-Based Docking	479
17.5	Challenges of Fragment-Based Approaches	481
	References	482
	<b>Appendix A: Software Overview</b>	<b>491</b>
	<b>Appendix B: Virtual Screening Application Studies</b>	<b>501</b>
	<b>Index</b>	<b>511</b>