

## Contents

	<b>Series Editors Preface</b>	<i>xiii</i>
	<b>Raimund Mannhold – A Personal Obituary from the Series Editors</b>	<i>xvii</i>
	<b>A Personal Foreword</b>	<i>xxi</i>
<b>1</b>	<b>Open Access Databases and Datasets for Computer-Aided Drug Design. A Short List Used in the Molecular Modelling Group of the SIB</b>	<b>1</b>
	<i>Antoine Daina, María José Ojeda-Montes, Maiia E. Bragina, Alessandro Cuozzo, Ute F. Röhrig, Marta A.S. Perez, and Vincent Zoete</i>	
	References	<i>30</i>
	<b>Part I Small Molecules</b>	<b>39</b>
<b>2</b>	<b>PubChem: A Large-Scale Public Chemical Database for Drug Discovery</b>	<b>41</b>
	<i>Sunghwan Kim and Evan E. Bolton</i>	
2.1	Introduction	<i>41</i>
2.2	Data Content and Organization	<i>42</i>
2.3	Tools and Services	<i>45</i>
2.3.1	PubChem Search	<i>45</i>
2.3.2	Summary Pages	<i>48</i>
2.3.3	Literature Knowledge Panel	<i>49</i>
2.3.4	2D and 3D Neighbors	<i>50</i>
2.3.5	Classification Browser	<i>51</i>
2.3.6	Identifier Exchange Service	<i>52</i>
2.3.7	Programmatic Access	<i>52</i>
2.3.8	PubChem FTP Site and PubChemRDF	<i>53</i>
2.4	Drug- and Lead-Likeness of PubChem Compounds	<i>54</i>
2.5	Bioactivity Data in PubChem	<i>56</i>
2.6	Comparison with Other Databases	<i>57</i>
2.7	Use of PubChem Data for Drug Discovery	<i>58</i>
2.8	Summary	<i>59</i>
	Acknowledgments	<i>60</i>
	References	<i>60</i>

<b>3</b>	<b>DrugBank Online: A How-to Guide</b>	<i>67</i>
	<i>Christen M. Klinger, Jordan Cox, Denise So, Teira Stauth, Michael Wilson, Alex Wilson, and Craig Knox</i>	
3.1	Introduction	<i>67</i>
3.2	DrugBank	<i>68</i>
3.2.1	Overview of DrugBank	<i>68</i>
3.2.2	DrugBank Datasets	<i>69</i>
3.2.2.1	Drug Cards: An Overview and Navigation Guide	<i>70</i>
3.2.2.2	Identification	<i>70</i>
3.2.2.3	Pharmacology	<i>71</i>
3.2.2.4	Categories	<i>73</i>
3.2.2.5	Properties	<i>73</i>
3.2.2.6	Targets, Enzymes, Carriers, and Transporters	<i>73</i>
3.2.2.7	References	<i>77</i>
3.3	Protocols	<i>77</i>
3.3.1	General Workflows	<i>77</i>
3.3.1.1	Using DrugBank Online's Search Functionality	<i>77</i>
3.3.1.2	Using DrugBank Online's Advanced Search Functionality	<i>80</i>
3.3.1.3	Browsing Drugs Using DrugBank Online's Drug Categories	<i>83</i>
3.3.2	Identifying Chemicals and Relevant Sequences	<i>86</i>
3.3.2.1	Searching Using Chemical Structure Search	<i>86</i>
3.3.2.2	Using Sequence Search to Find Similar Targets	<i>89</i>
3.3.3	Extracting DrugBank Datasets for ML	<i>93</i>
3.4	Research Using DrugBank	<i>94</i>
3.5	Discussion and Conclusions	<i>95</i>
	References	<i>96</i>
<b>4</b>	<b>Bioisosteric Replacement for Drug Discovery Supported by the SwissBioisostere Database</b>	<i>101</i>
	<i>Antoine Daina, Alessandro Cuozzo, Marta A.S. Perez, and Vincent Zoete</i>	
4.1	Introduction	<i>101</i>
4.1.1	Concept of Isosterism and Bioisosterism	<i>101</i>
4.1.2	Classical vs. Non-classical Bioisostere and Further Molecular Replacements	<i>102</i>
4.1.3	Bioisosteric Replacement in Drug Discovery	<i>105</i>
4.2	Construction and Dissemination of SwissBioisostere	<i>106</i>
4.2.1	Intention and Requirements	<i>106</i>
4.2.2	Bioactivity Data	<i>107</i>
4.2.3	Nonsupervised Matched Molecular Pair Analysis	<i>108</i>
4.2.4	Database	<i>108</i>
4.2.5	Web Interface	<i>109</i>
4.3	Content of SwissBioisostere	<i>111</i>
4.3.1	Global Content	<i>111</i>
4.3.2	Biological and Chemical Contexts	<i>112</i>
4.3.3	Fragment Shape Diversity	<i>113</i>

4.4	Usage of SwissBioisostere	115
4.4.1	Website Usage	115
4.4.2	Most Frequent Requests	117
4.4.3	Examples Related to Drug Discovery	117
4.4.3.1	Use Cases	117
4.4.3.2	Replacing Unwanted Chemical Groups	118
4.4.3.3	Optimization of Passive Absorption and Blood–Brain Barrier Diffusion	122
4.4.3.4	Reduction of Flexibility	124
4.4.3.5	Reduction of Aromaticity/Escape from Flatland	128
4.5	Conclusive Remarks	133
	Acknowledgment	133
	References	133

## Part II Macromolecular Targets and Diseases 139

5	<b>The Protein Data Bank (PDB) and Macromolecular Structure Data Supporting Computer-Aided Drug Design</b>	141
	<i>David Armstrong, John Berrisford, Preeti Choudhary, Lukas Pravda, James Tolchard, Mihaly Varadi, and Sameer Velankar</i>	
5.1	Introduction	141
5.2	Small Molecule Data in Protein Data Bank (PDB) Entries	142
5.2.1	What Data are in the PDB Archive?	142
5.2.2	Definition of Small Molecules in OneDep	145
5.3	Small Molecule Dictionaries	146
5.3.1	wwPDB Chemical Component Dictionary (CCD)	146
5.3.2	The Peptide Reference Dictionary	147
5.4	Additional Ligand Annotations in the PDB Archive	148
5.4.1	Linkage Information	148
5.4.2	Carbohydrates	149
5.5	Validation of Ligands in the Worldwide Protein Data Bank (wwPDB)	150
5.5.1	Various Criteria and Software Used for Validating Ligand in Validation Reports	150
5.5.2	Identification of Ligand of Interest (LOI)	151
5.5.3	Geometric and Conformational Validation	152
5.5.4	Ligand Fit to Experimental Electron Density Validation	152
5.5.5	Accessing wwPDB Validation Reports from PDBe Entry Pages	154
5.5.6	Other Planned Improvements to Enhance Ligand Validation	154
5.6	PDBe Tools for Ligand Analysis	155
5.6.1	Ligand Interactions	155
5.6.1.1	Classifying Ligand Interactions	155
5.6.1.2	Data Availability	156
5.6.2	Ligand Environment Component	156
5.6.3	Chemistry Process and FTP	158

5.6.4	PDBChem Pages	158
5.7	Ligand-Related Annotations in the PDBe-KB	158
5.7.1	Introduction to PDBe-KB	158
5.7.2	Data Access Mechanisms for Ligand-Related Annotations	160
5.7.3	Ligand-Related Annotations on the Aggregated Views of Proteins	162
5.8	Case Study: Using PDB Data to Support Drug Discovery	164
5.9	Conclusions and Outlook	165
5.9.1	Upcoming Features and Improvements	166
	References	167
<b>6</b>	<b>The SWISS-MODEL Repository of 3D Protein Structures and Models</b>	<b>175</b>
	<i>Xavier Robin, Andrew Mark Waterhouse, Stefan Bienert, Gabriel Studer, Leila T. Alexander, Gerardo Tauriello, Torsten Schwede, and Joana Pereira</i>	
6.1	Introduction	175
6.2	SMR Database Content and Model Providers	176
6.2.1	PDB	177
6.2.2	SWISS-MODEL	177
6.2.3	AlphaFold Database	179
6.2.4	ModelArchive	180
6.3	Protein Feature Annotation and Cross-References to Computational Resources	181
6.3.1	Structural Features, Ligands, and Oligomers	181
6.3.2	SWISS-MODEL associated tools	182
6.3.3	Web and API Access	183
6.4	Quality Estimates and Benchmarking	188
6.5	Binding Site Conformational States	189
6.6	SMR and Computer-Aided Structure-based Drug Design	190
6.7	Conclusion and Outlook	191
	References	193
<b>7</b>	<b>PDB-REDO in Computational-Aided Drug Design (CADD)</b>	<b>201</b>
	<i>Ida de Vries, Anastassis Perrakis, and Robbie P. Joosten</i>	
7.1	History and Concepts	201
7.1.1	X-ray Structure Models	201
7.1.2	PDB-REDO Development	202
7.1.2.1	First Uniformity	203
7.1.2.2	Automatic Rebuilding of Protein Backbone and Side Chains	203
7.1.2.3	Automated Model Completion Approaches	204
7.1.2.4	Systematic Integration of Structural Knowledge	205
7.1.2.5	Overview of PDB-REDO Pipeline	205
7.2	Structure Improvements by PDB-REDO	206
7.2.1	Parametrization and Rebuilding Effects on Small Molecule Ligands	206
7.2.1.1	Re-refinement Improves Ligand Conformation	206
7.2.1.2	Side Chain Rebuilding Improves Ligand Binding Sites	207

7.2.1.3	Histidine Flip and Improved Ligand Parameterization	208
7.2.2	Building of Protein Loops and Ligands into Protein Structure Models	210
7.2.2.1	Loop Building Completes a Binding Site Region	210
7.2.2.2	Loop Building Results in Improved Binding Sites	211
7.2.2.3	Building new Compounds into Density	212
7.2.3	Nucleic Acid Improvements by PDB-REDO	213
7.2.4	Glycoprotein Structure Model Rebuilding	214
7.2.5	Metal Binding Sites	214
7.2.6	Limitations of the PDB-REDO Databank	216
7.3	Access the PDB-REDO Databank and Metadata	218
7.3.1	Downloading and Inspecting Individual PDB-REDO Entries	218
7.3.2	Data Available in PDB-REDO Entries	220
7.3.3	Usage of the Uniform and FAIR Validation Data	220
7.3.4	Creating Datasets from the PDB-REDO Databank	222
7.3.5	Submitting Structure Models to the PDB-REDO Pipeline	223
7.4	Conclusions	223
	Acknowledgments and Funding	224
	List of Abbreviations and Symbols	224
	References	225

## 8 Pharos and TCRD: Informatics Tools for Illuminating Dark Targets 231

*Keith J. Kelleher, Timothy K. Sheils, Stephen L. Mathias, Duc-Trung Nguyen, Vishal Siramshetty, Ajay Pillai, Jeremy J. Yang, Cristian G. Bologa, Jeremy S. Edwards, Tudor I. Oprea, and Ewy Mathé*

8.1	Introduction	231
8.2	Methods	233
8.2.1	Data Organization	233
8.2.1.1	Target Alignment	234
8.2.1.2	Disease Alignment	234
8.2.1.3	Ligand Alignment	234
8.2.1.4	Data and UI Updates	235
8.2.2	Programmatic Access and Data Download	235
8.2.3	UI Organization	235
8.2.3.1	List Pages	236
8.2.3.2	Details Pages	236
8.2.3.3	Search	238
8.2.3.4	Tutorials	240
8.2.4	Analysis Methods Within Pharos	240
8.2.4.1	Searching for Ligands	240
8.2.4.2	Finding Targets by Amino Acid Sequence	241
8.2.4.3	Finding Targets with Similar Annotations	241
8.2.4.4	Finding Targets with Predicted Activity	241
8.2.4.5	Enrichment Scores for Filter Values	241

8.3	Use Cases	242
8.3.1	Hypothesizing the Role of a Dark Target	242
8.3.1.1	Primary Documentation	242
8.3.1.2	List Analysis	247
8.3.1.3	Downloading Data	251
8.3.1.4	Variations on this Use Case	251
8.3.2	Characterizing a Novel Chemical Compound	251
8.3.2.1	Finding Predicted Targets	252
8.3.2.2	Analyzing Similar Ligands	254
8.3.2.3	Ligand Details Pages	256
8.3.2.4	Variations on this Use Case	257
8.3.3	Investigating Diseases	260
8.4	Discussion	262
	Funding	264
	References	264

### Part III Users' Points of View 269

<b>9</b>	<b>Mining for Bioactive Molecules in Open Databases</b>	<b>271</b>
	<i>Guillem Macip, Júlia Mestres-Truyol, Pol Garcia-Segura, Bryan Saldivar-Espinoza, Santiago Garcia-Vallvé, and Gerard Pujadas</i>	
9.1	Introduction	271
9.2	Main Tools for Virtual Screening	272
9.2.1	ADMET and PAINS Filtering	272
9.2.2	Protein–Ligand Docking	274
9.2.3	Pharmacophore Search	275
9.2.4	Shape/Electrostatic Similarity	276
9.2.5	Protein-Structure Databases	277
9.2.6	The Protein Data Bank	278
9.2.7	The PDB-REDO Databank	278
9.2.8	The SWISS-MODEL Repository	279
9.2.9	The AlphaFold Protein Structure Database	279
9.3	Validating Binding Site and Ligand Coordinates in Three-Dimensional Protein Complexes	280
9.4	Databases for Searching New Drugs	281
9.4.1	COCONUT	281
9.4.2	GDBs	282
9.4.3	ZINC20	282
9.5	Databases of Bioactive Molecules	282
9.5.1	The BindingDB Database	283
9.5.2	PubChem	283
9.5.3	ChEMBL	284
9.6	Databases of Inactive/Decoy Molecules	285

9.6.1	Collecting Experimentally Inactive Compounds from PubChem	285
9.6.2	Collecting Presumed Inactive Compounds from Decoy Databases	285
9.6.3	Building Custom-Based Decoy Sets	286
9.7	Main Metrics for Evaluating the Success of a Virtual Screening	286
9.8	Concluding Remarks	288
	References	289
<b>10</b>	<b>Open Access Databases – An Industrial View</b>	<b>299</b>
	<i>Michael Przewosny</i>	
10.1	Academic vs. Industrial Research	299
10.2	Scaffold-Hopping	310
10.3	Virtual-Screening	311
	Abbreviations	312
	References	313
	<b>Index</b>	<b>317</b>

