

Contents

Part 1: Introduction

Chapter 1	Introduction	3
	1.1 Overview	3

Part 2: Molecular Representations

Chapter 2	Chemistry and Graph Theory	13
	2.1 Overview	13
	2.2 Graph Theory and Chemistry	13
	2.3 Graph Theory in Chemistry	16
	2.4 Mathematical Chemistry and Chemical Graph Theory	17
	2.5 Summary	18
	References	18
Chapter 3	Structure Representation	19
	3.1 Overview	19
	3.2 The Need for Machine-Readable Structure Representations	20
	3.3 Adjacency Matrix	22
	3.4 Connection Table	23
	3.5 Line Notations	25
	3.5.1 WLN: Wiswesser Line Notation	26
	3.5.2 SMILES: Simplified Molecular-Input Line-Entry Specification	26
	3.5.3 InChI: IUPAC International Chemical Identifier	28
	3.6 Summary	29
	References	30

Chapter 4	Molecular Similarity	31
	4.1 Overview	31
	4.2 Molecular Similarity	33
	4.3 Similar Property Principle	35
	4.4 Molecular Descriptors	36
	4.5 Calculation of Molecular Similarity	37
	4.5.1 Similarity Coefficients	37
	4.6 Molecular Diversity	40
	4.7 Summary	40
	References	41
Part 3: Molecular Descriptors		
Chapter 5	Molecular Property Descriptors	45
	5.1 Overview	45
	5.2 Molecular Weight (MW or MWt)	45
	5.3 Octanol/Water Partition Coefficient (ClogP)	46
	5.4 Topological Polar Surface Area (TPSA)	47
	5.5 Hydrogen Bond Acceptors and Donors (HBA and HBD)	48
	5.6 Lipinski's Rule-of-Five	48
	5.7 Summary	50
	References	51
Chapter 6	Topological Descriptors	52
	6.1 Overview	52
	6.2 Topological Indices	52
	6.2.1 Wiener Index	53
	6.2.2 Randić Index	53
	6.2.3 Petitjean Index	55
	6.2.4 Zagreb Indices	55
	6.3 Molecular Fingerprints	55
	6.3.1 Structure-Key Fingerprints	56
	6.3.2 Hash-Key Fingerprints	57
	6.3.3 Ligand-Based Topological Pharmacophores	60
	6.4 Summary	64
	References	64
Chapter 7	Topographical Descriptors	66
	7.1 Overview	66
	7.2 Topographic Descriptors	67
	7.3 Pharmacophores	71
	7.4 ROCS: Rapid Overlay of Chemical Structures	75
	7.5 USR: Ultrafast Shape Recognition	76
	7.6 XED: Cresset Group	77
	7.7 Conformer Generation and the Conformer Problem	77
	7.8 Summary	80
	References	81

Part 4: Statistical Learning

Chapter 8	Statistical Learning	85
	8.1 Overview	85
	8.2 Statistical Learning	85
	8.3 Unsupervised Learning	86
	8.3.1 Overview	86
	8.3.2 Cluster Analysis	87
	8.3.3 <i>k</i> -Means Clustering	88
	8.3.4 Stirling Numbers of the Second Kind	89
	8.3.5 Self-Organising Maps	89
	8.3.6 Principal Component Analysis	91
	8.4 Supervised Learning	92
	8.4.1 Naïve Bayesian Classification	92
	8.4.2 Support Vector Machine	93
	8.4.3 Partial Least Squares	93
	8.5 Best Modelling Practice	94
	8.6 Summary	95
	References	96

Part 5: Modelling Methodologies

Chapter 9	Similarity Searching	101
	9.1 Overview	101
	9.2 Similar Property Principle	102
	9.3 Molecular Similarity and Virtual Screening	104
	9.4 Data Fusion	104
	9.5 Enrichment	106
	9.5.1 Lift Plots	106
	9.5.2 Confusion Matrix	107
	9.5.3 Receiver Operating Characteristic Curves	108
	9.5.4 Enrichment Factors	108
	9.6 Summary	110
	References	110
Chapter 10	Bioisosteres and Scaffolds	112
	10.1 Overview	112
	10.2 A Brief History of Bioisosterism	113
	10.3 Bioisosteric Replacement Methods	115
	10.3.1 Knowledge-Based Bioisosteric Replacements	115
	10.3.2 Information-Based Bioisosteric Replacements	116
	10.4 Scaffold Representations	117
	10.5 Scaffold Diversity Analysis	118
	10.6 Summary	121
	References	122

Chapter 11	Clustering and Diversity	124
	11.1 Overview	124
	11.2 Dissimilarity-Based Compound Selection	125
	11.3 Sphere-Exclusion	127
	11.4 Cell-Based Diversity Selection	127
	11.5 Hierarchical Clustering	128
	11.6 Non-Hierarchical Clustering	129
	11.7 Summary	130
	References	131
Chapter 12	Quantitative Structure–Activity Relationships	132
	12.1 Overview	132
	12.2 Free–Wilson Analysis	132
	12.3 Hansch–Fujita Analysis and Topliss Trees	133
	12.4 QSAR Model Generation	133
	12.5 Feature Selection	136
	12.6 Methods for Estimating Model Validity, Predictive Power, and Applicability Domains	139
	12.7 Automated Model Generation, Validation and Application	140
	12.8 Summary	141
	References	142
Chapter 13	Protein–Ligand Docking	143
	13.1 Overview	143
	13.2 Search Algorithms	143
	13.3 Scoring Functions	144
	13.4 GOLD: Genetic Optimisation for Ligand Docking	145
	13.5 Model Validation	147
	13.6 Docking in Prospective Studies	147
	13.7 Summary	149
	References	151
Chapter 14	<i>De Novo</i> Molecular Design	153
	14.1 Overview	153
	14.2 Receptor-Based Methods	154
	14.3 Fragment-Linking Methods	155
	14.4 Fragment-Growing Methods	156
	14.5 Sampling Chemistry Space	156
	14.6 Atom-Based <i>De Novo</i> Design	157
	14.7 Fragment-Based <i>De Novo</i> Design	158
	14.8 Reaction-Based <i>De Novo</i> Design	158
	14.9 Multiobjective Optimisation	160
	14.10 Summary	161
	References	161

Part 6: Applications in Medicinal Chemistry

Chapter 15	Applications in Medicinal Chemistry	165
	15.1 Overview	165
	15.2 Early Stage Targets	166
	15.3 Hit Generation	168
	15.4 Hits-to-leads	172
	15.5 Lead Optimisation	175
	15.6 Summary	176
	References	177

Part 7: Summary and Outlook

Chapter 16	Summary and Outlook	181
	16.1 The Past	181
	16.2 The Present	182
	16.3 The Future	182
	16.4 Summary	184

Appendices

Appendix A	Glossary of Terms	187
Appendix B	Professional Societies	192
Appendix C	Journals	195
Appendix D	Resources for Computational Drug Discovery	199
	Subject Index	205