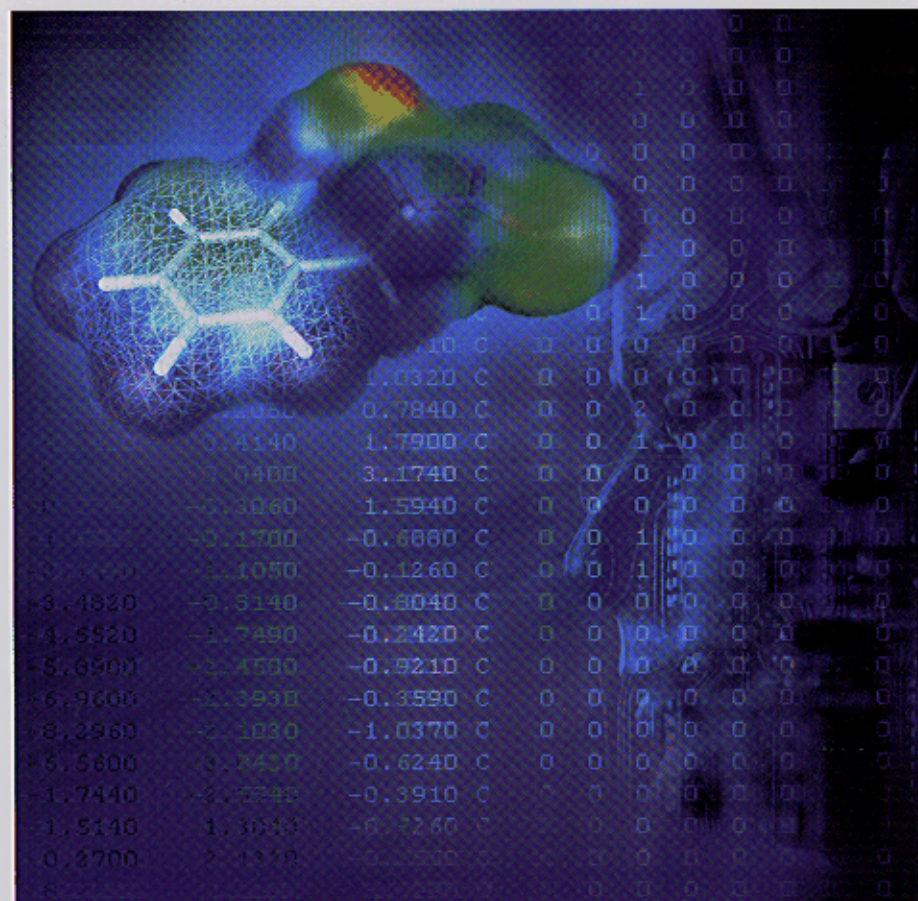


Edited by
Johann Gasteiger and Thomas Engel

 WILEY-VCH

Chemoinformatics

A Textbook



Contents

Foreword VII

Preface XXV

Addresses of the Authors XXVII

1 Introduction 1

- 1.1 The Domain of Chemistry 1
- 1.2 A Chemist's Fundamental Questions 3
- 1.3 The Scope of Chemoinformatics 4
- 1.4 Learning in Chemoinformatics 6
- 1.5 Major Tasks 8
 - 1.5.1 Representation of the Objects 8
 - 1.5.2 Data 9
 - 1.5.3 Learning 9
- 1.6 History of Chemoinformatics 9
 - 1.6.1 Structure Databases 10
 - 1.6.2 Quantitative Structure–Activity Relationships 10
 - 1.6.3 Molecular Modeling 10
 - 1.6.4 Structure Elucidation 11
 - 1.6.5 Chemical Reactions and Synthesis Design 11
- 1.7 The Scope of this Book 11
- 1.8 Teaching Chemoinformatics 13

2 Representation of Chemical Compounds 15

- 2.1 Introduction 15
- 2.2 Chemical Nomenclature 18
 - 2.2.1 Development of Chemical Nomenclature 19
 - 2.2.2 Representation of Chemical Elements 19
 - 2.2.2.1 Characterization of Elements 19
 - 2.2.3 Representation of the Empirical Formulas of (Inorganic) Compounds 20
 - 2.2.3.1 Present-Day Representation 20
 - 2.2.4 Representation of the Empirical Formulas of Organic Compounds 20
 - 2.2.4.1 Present-Day Representation 20

2.2.5	Systematic Nomenclature of Inorganic and Organic Compounds	21
2.3	Line Notations	23
2.3.1	Wiswesser Line Notation	23
2.3.1.1	Applications	25
2.3.2	ROSDAL	25
2.3.2.1	Applications	26
2.3.3	The SMILES Coding	26
2.3.3.1	Applications	27
2.3.4	Sybyl Line Notation	27
2.3.4.1	Applications	30
2.4	Coding the Constitution	30
2.4.1	Graph Theory	31
2.4.1.1	Basics of Graph Theory	32
2.4.2	Matrix Representations	34
2.4.2.1	Adjacency Matrix	35
2.4.2.2	Distance Matrix	36
2.4.2.3	Atom Connectivity Matrix	36
2.4.2.4	Incidence Matrix	36
2.4.2.5	Bond Matrix	36
2.4.3	Connection Table	40
2.4.4	Input and Output of Chemical Structures	42
2.4.5	Standard Structure Exchange Formats	45
2.4.6	Tutorial: Molfiles and SDfiles	47
2.4.6.1	Structure of a Molfile	47
2.4.6.2	Structure of an SDfile	51
2.4.6.3	Libraries and Toolkits	51
2.5	Processing Constitutional Information	53
2.5.1	Ring Perception	53
2.5.1.1	Minimum Number of Cycles	55
2.5.1.2	All Cycles	55
2.5.1.3	Smallest Fundamental Basis	55
2.5.2	Unambiguous and Unique Representations	57
2.5.2.1	Structure Isomers and Isomorphism	57
2.5.2.2	Canonicalization	59
2.5.3	The Morgan Algorithm	59
2.5.3.1	Tutorial: Morgan Algorithm	60
2.6	Beyond a Connection Table	63
2.6.1	Deficiencies in Representing Molecular Structures by a Connection Table	63
2.6.2	Representation of Molecular Structures by Electron Systems	64
2.6.2.1	General Concepts	64
2.6.2.2	Simple Single and Double Bonds	64
2.6.2.3	Conjugation and Aromaticity	65
2.6.2.4	Orthogonality of π -Systems	66
2.6.2.5	Non-bonding Orbitals	67

2.6.2.6	Charged Species and Radicals	67
2.6.2.7	Ionized States	68
2.6.2.8	Electron-Deficient Compounds	68
2.6.2.9	Organometallic Compounds	69
2.6.3	Generation of RAMSES from a VB Representation	69
2.7	Special Notations of Chemical Structures	70
2.7.1	Markush Structures	70
2.7.2	Fragment Coding	71
2.7.2.1	Applications	71
2.7.3	Fingerprints	71
2.7.3.1	Hashed Fingerprints	72
2.7.4	Hash Codes	72
2.7.4.1	Applications	74
2.8	Representation of Stereochemistry	75
2.8.1	General Concepts	75
2.8.2	Representation of Configuration Isomers and Molecular Chirality	75
2.8.2.1	Detection and Specification of Chirality	78
2.8.3	Ordered Lists	79
2.8.4	Rotational Lists	80
2.8.5	Permutation Descriptors	81
2.8.6	Stereochemistry in Molfile and SMILES	82
2.8.6.1	Stereochemistry in the Molfile	82
2.8.6.2	Stereochemistry in SMILES	84
2.8.7	Tutorial: Handling of Stereochemistry by Permutation Groups	85
2.8.7.1	Stereochemistry at Tetrahedral Carbon Atoms	86
2.8.7.2	Stereochemistry at Double Bonds	88
2.9	Representation of 3D Structures	91
2.9.1	Walking through the Hierarchy of Chemical Structure Representation	91
2.9.2	Representation of 3D Structures	92
2.9.3	Obtaining 3D Structures and Why They are Needed	94
2.9.4	Automatic 3D Structure Generation	96
2.9.5	Obtaining an Ensemble of Conformations: What is Conformational Analysis?	103
2.9.6	Automatic Generation of Ensembles of Conformations	105
2.9.7	Tutorial: 3D Structure Codes (PDB, STAR, CIF, mmCIF)	112
2.9.7.1	Introduction	112
2.9.7.2	PDB File Format	112
2.9.7.3	STAR File Format and Dictionaries	120
2.9.7.4	CIF File Format (CCDC)	121
2.9.7.5	mmCIF File Format	123
2.9.7.6	Software	123
2.10	Molecular Surfaces	124
2.10.1	van der Waals Surface	125
2.10.2	Connolly Surface	126
2.10.3	Solvent-Accessible Surface	127

2.10.4	Solvent-Excluded Surface (SES)	128
2.10.5	Enzyme Cavity Surface (Union Surface)	128
2.10.6	Isovalue-Based Electron Density Surface	129
2.10.7	Experimentally Determined Surfaces	129
2.11	Visualization of Molecular Models	129
2.11.1	Historical Review	130
2.11.2	Structure Models	131
2.11.2.1	Wire Frame Model	132
2.11.2.2	Capped Sticks Model	132
2.11.2.3	Balls and Sticks Model	133
2.11.2.4	Space-Filling Model	133
2.11.3	Models of Biological Macromolecules	133
2.11.3.1	Cylinder Model	134
2.11.3.2	Ribbon Model	134
2.11.3.3	Tube Model	134
2.11.4	Crystallographic Models	135
2.11.5	Visualization of Molecular Properties	135
2.11.5.1	Properties Based on Isosurfaces	135
2.12	Tools: Chemical Structure Drawing Software – Molecule Editors and Viewers	137
2.12.1	Introduction	137
2.12.2	Molecule Editors	138
2.12.2.1	Stand-Alone Applications	139
2.12.2.2	Web-Based Applications	144
2.12.3	Molecule Viewers	146
2.12.3.1	Stand-Alone Applications	146
2.12.3.2	Web-Based Applications	155
2.13	Tools: 3D Structure Generation on the Web	157
3	Representation of Chemical Reactions	169
3.1	Introduction	169
3.2	Reaction Types	172
3.3	Reaction Center	173
3.4	Chemical Reactivity	176
3.4.1	Physicochemical Effects	176
3.4.1.1	Charge Distribution	176
3.4.1.2	Inductive Effect	177
3.4.1.3	Resonance Effect	178
3.4.1.4	Polarizability Effect	178
3.4.1.5	Steric Effect	178
3.4.1.6	Stereoelectronic Effects	178
3.4.2	Simple Approaches to Quantifying Chemical Reactivity	179
3.4.2.1	Frontier Molecular Orbital Theory	179
3.4.2.2	Linear Free Energy Relationships (LFER)	179
3.4.2.3	Empirical Reactivity Equations	183

3.5	Reaction Classification	183
3.5.1	Model-Driven Approaches	183
3.5.1.1	Hendrickson's Scheme	183
3.5.1.2	Ugi's Scheme	185
3.5.1.3	InfoChem's Reaction Classification	191
3.5.2	Data-Driven Approaches	191
3.5.2.1	HORACE	191
3.5.2.2	Reaction Landscapes	192
3.6	Stereochemistry of Reactions	196
3.7	Tutorial: Stereochemistry of Reactions	197
4	The Data	203
4.1	Introduction	203
4.1.1	Data, Information, Knowledge	203
4.1.2	The Data Acquisition Pathway	204
4.2	Data Acquisition	206
4.2.1	Why Does the Quality of Data Matter?	206
4.2.2	Data Complexity	207
4.2.3	Experimental Data	208
4.2.4	Data Exchange	209
4.2.4.1	DAT files	209
4.2.4.2	JCAMP-DX	210
4.2.4.3	PMML	211
4.2.5	Real-World Data and their Potential Drawbacks	211
4.3	Data Pre-processing	213
4.3.1	Mean-Centering, Scaling, and Autoscaling	213
4.3.2	Advanced Methods	215
4.3.2.1	Fast Fourier Transformation	215
4.3.2.2	Wavelet Transformation	216
4.3.2.3	Singular Value Decomposition	216
4.3.3	Variable Selection	217
4.3.3.1	Genetic Algorithm (GA)-Based Solutions	217
4.3.3.2	Orthogonalization-Based Solutions	219
4.3.3.3	Simulated Annealing (SA)-Based Solutions	219
4.3.3.4	PCA-Based Solutions	219
4.3.4	Object Selection	220
4.4	Preparation of Datasets for Validation of the Model Quality	221
4.4.1	Training and Test Datasets	221
4.4.2	Compilation of Test Sets	222
5	Databases and Data Sources in Chemistry	227
5.1	Introduction	227
5.2	Basic Database Theory	228
5.2.1	Databases in the Information System	228
5.2.2	Search Engine	230

5.2.3	Access to Databases	230
5.2.4	Types of Database Systems	232
5.2.4.1	Hierarchical Database System	232
5.2.4.2	Network Model	233
5.2.4.3	Relational Model	235
5.2.4.4	Object-Based Model	236
5.3	Classification of Databases	236
5.3.1	Literature Databases	238
5.3.2	Factual Databases	238
5.3.2.1	Numeric Databases	238
5.3.2.2	Catalogs of Chemical Compounds	240
5.3.2.3	Research Project Databases	240
5.3.2.4	Metadatabases	240
5.3.3	Structure Databases	240
5.3.4	Reaction Databases	240
5.4	Literature Databases	241
5.4.1	Chemical Abstracts File	241
5.4.2	SCISEARCH	241
5.4.3	Medline (Medical Literature, Analysis, and Retrieval System Online)	241
5.5	Tutorial: Using the Chemical Abstracts System	242
5.5.1	Online Access	242
5.5.2	Access to CAS with SciFinder Scholar 2002	242
5.5.2.1	Getting Started	242
5.5.2.2	Searching within Various Topics	244
5.6	Property (Numeric) Databases	247
5.6.1	Beilstein Database	248
5.6.2	Gmelin	248
5.6.3	DETERM	249
5.7	Tutorial: Searching in the Beilstein Database	249
5.7.1	Example 1: Combined Structure and Fact Retrieval	249
5.7.2	Example 2: Reaction Retrieval	254
5.8	Spectroscopic Databases	257
5.8.1	SpecInfo	258
5.9	Crystallographic Databases	258
5.9.1	Inorganic Crystal Structure Database (ICSD)	259
5.9.2	Cambridge Structural Database (CSD)	259
5.9.3	Protein Data Bank (PDB)	259
5.10	Molecular Biology Databases	260
5.10.1	GenBank (Genetic Sequence Bank)	260
5.10.2	EMBL	261
5.10.3	PIR (Protein Information Resource)	261
5.10.4	SWISS-PROT	261
5.10.5	CAS Registry	261
5.11	Structure Databases	262
5.11.1	CAS Registry	262

5.11.2	National Cancer Institute (NCI) Database	262
5.12	Chemical Reaction Databases	263
5.12.1	CASREACT	263
5.12.2	ChemInform RX	264
5.13	Tutorial: Searching in the ChemInform Reaction Database	264
5.13.1	Introduction	264
5.13.2	Example 1: Reaction Retrieval	264
5.13.3	Example 2: Advanced Reaction Retrieval	266
5.13.4	Classifying Reactions on a Hit List	267
5.14	Patent Databases	268
5.14.1	INPADOC	269
5.14.2	World Patent Index (WPINDEX)	270
5.14.3	MARPAT	270
5.15	Chemical Information on the Internet	270
5.16	Tutorial: Searching the Internet for Chemical Information	271
5.17	Tutorial: Searching Environmental Information in the Internet	274
5.17.1	Introduction: Difficulties in Extracting Scientific Environmental Information from the Internet	274
5.17.2	Ways of Searching for Environmental Information on the Internet	275
5.17.2.1	Metadatabases and Portals	275
5.17.2.2	Search Engines	275
5.17.2.3	Databases	277
5.18	Tools: The Internet (Online Databases in Chemistry)	278
6	Searching Chemical Structures	291
6.1	Introduction	291
6.2	Full Structure Search	292
6.3	Substructure Search	296
6.3.1	Basic Ideas	296
6.3.2	Backtracking Algorithm	298
6.3.3	Optimization of the Backtracking Algorithm	301
6.3.4	Screening	302
6.4	Similarity Search	303
6.4.1	Similarity Basics	303
6.4.2	Similarity Measures	304
6.4.3	The Similarity Search Process	310
6.4.3.1	Object Selection	310
6.4.3.2	Descriptor Selection and Encoding	311
6.4.3.3	Similarity Measure Selection	312
6.4.3.4	Query Object Specification	312
6.4.3.5	Similarity Scores	312
6.4.3.6	Application Areas	313
6.5	Three-Dimensional Structure Search Methods	313

7	Calculation of Physical and Chemical Data	319
7.1	Empirical Approaches to the Calculation of Properties	320
7.1.1	Introduction	320
7.1.2	Additivity of Atomic Contributions	321
7.1.2.1	Hybridization States	322
7.1.3	Additivity of Bond Contributions	323
7.1.4	Additivity of Group Contributions	323
7.1.5	Effects of Rings	325
7.1.6	Drug-Receptor Binding Energies	326
7.1.7	Attenuation Models	327
7.1.7.1	Calculation of Charge Distribution	329
7.1.7.2	Polarizability Effect	333
7.2	Molecular Mechanics	338
7.2.1	Introduction	338
7.2.2	No Force Field Calculation Without Atom Types	339
7.2.3	The Functional Form of Common Force Fields	339
7.2.3.1	Bond Stretching	340
7.2.3.2	Angle Bending	342
7.2.3.3	Torsional Terms	343
7.2.3.4	Out-of-Plane Bending	343
7.2.3.5	Electrostatic Interactions	345
7.2.3.6	Van der Waals Interactions	346
7.2.3.7	Cross-Terms	348
7.2.4	Available Force Fields	349
7.2.4.1	Force Fields for Small Molecules	349
7.2.4.2	Force Fields for Biomolecules	352
7.3	Molecular Dynamics	359
7.3.1	Introduction	359
7.3.2	The Continuous Movement of Molecules	359
7.3.3	Methods	360
7.3.3.1	Algorithms	361
7.3.3.2	Ways to Speed up the Calculations	362
7.3.3.3	Solvent Effects	363
7.3.3.4	Periodic Boundary Conditions	366
7.3.4	Constant Energy, Temperature, or Pressure?	366
7.3.5	Long-Range Forces	368
7.3.6	Application of Molecular Dynamics Techniques	369
7.4	Quantum Mechanics	376
7.4.1	Hückel Molecular Orbital Theory	376
7.4.2	Semi-empirical Molecular Orbital Theory	381
7.4.3	<i>Ab Initio</i> Molecular Orbital Theory	384
7.4.4	Density Functional Theory	389
7.4.5	Properties from Quantum Mechanical Calculations	390
7.4.5.1	Net Atomic Charges	391
7.4.5.2	Dipole and Higher Multipole Moments	392

7.4.5.3	Polarizabilities	392
7.4.5.4	Orbital Energies	393
7.4.5.5	Surface Descriptors	393
7.4.5.6	Local Ionization Potential	393
7.4.6	Quantum Mechanical Techniques for Very Large Molecules	394
7.4.6.1	Linear Scaling Methods	394
7.4.6.2	Hybrid QM/MM Calculations	395
7.4.7	The Future of Quantum Mechanical Methods in Chemoinformatics	395

8 Calculation of Structure Descriptors 401

8.1	Introduction	401
8.1.1	Definition of the Term "Structure Descriptor"	403
8.1.2	Classification of Structure Descriptors	403
8.2	Structure Keys and 1D Fingerprints	403
8.2.1	Distance and Similarity Measures	405
8.3	Topological Descriptors	407
8.3.1	Some Fundamentals of Graph Theory	407
8.3.2	The Adjacency Matrix	408
8.3.3	The Laplacian Matrix	409
8.3.4	The Distance Matrix	409
8.3.5	The Wiener Index	410
8.3.6	The Randic Connectivity Index	411
8.3.7	Topological Autocorrelation Vectors	411
8.3.8	Feature Trees	411
8.3.9	Further Topological Descriptors	412
8.4	3D Descriptors	412
8.4.1	3D Structure Generation	412
8.4.2	3D Autocorrelation	413
8.4.2.1	Example: Xylene Isomers	413
8.4.3	3D Molecule Representation of Structures Based on Electron Diffraction Code (3D MoRSE Code)	415
8.4.4	Radial Distribution Function Code	415
8.5	Chirality Descriptors	418
8.5.1	Quantitative Descriptions of Chirality	418
8.5.2	Continuous Chirality Measure (CCM)	418
8.5.3	Chirality Codes	419
8.6	Tutorial: Conformation-Independent and Conformation-Dependent Chirality Codes	420
8.6.1	Introduction	420
8.6.2	Conformation-Independent Chirality Code (CICC)	420
8.6.2.1	Preparatory Calculations	420
8.6.2.2	Neighborhoods of Atoms Bonded to the Chiral Center	421
8.6.2.3	Enumeration of Combinations	421
8.6.2.4	Characterization of Combinations	421
8.6.2.5	Generation of the Code	422

8.6.3	Conformation-Dependent Chirality Code (CDCC)	423
8.6.3.1	Overview	423
8.6.3.2	Enumeration of combinations	423
8.6.3.3	Ranking of the Four Atoms in a Combination	423
8.6.3.4	Characterization of Combinations	423
8.6.3.5	Generation of the Code	424
8.6.3.6	Example of an Application	424
8.7	Further Descriptors	427
8.7.1	Comparative Molecular Field Analysis (CoMFA)	428
8.7.2	BCUT Descriptors	428
8.7.3	4D-QSAR	429
8.7.4	HYBOT Descriptors	429
8.8	Descriptors that are not Structure-Based	430
8.9	Properties of Structure Descriptors	431
9	Methods for Data Analysis	439
9.1	Introduction	439
9.2	Machine Learning Techniques	440
9.2.1	Machine Learning Process	441
9.2.2	Unsupervised Learning	441
9.2.3	Supervised Learning	441
9.3	Decision Trees	442
9.4	Chemometrics	442
9.4.1	Multivariate Statistics	443
9.4.2	Correlation	444
9.4.3	Multiple Linear Regression Analysis (MLRA)	446
9.4.4	Principal Component Analysis (PCA)	446
9.4.5	Principal Component Regression (PCR)	448
9.4.6	Partial Least Squares Regression/Projection to Latent Structures (PLS)	449
9.4.7	Example: Ion Concentrations in Mineral Waters	449
9.4.8	Tools: Electronic Data Analysis Service (ELECTRAS)	449
9.5	Neural Networks	452
9.5.1	Modeling the Brain: Biological Neurons versus Artificial Neurons	452
9.5.2	Networks	454
9.5.2.1	Training	454
9.5.2.2	Learning Strategies	455
9.5.3	Kohonen Network	456
9.5.3.1	Architecture	456
9.5.3.2	Training	456
9.5.4	Tutorial: Application of a Kohonen Network for the Classification of Olive Oils using ELECTRAS	458
9.5.5	Counter-propagation Network	459
9.5.5.1	Architecture	459
9.5.5.2	Training	460

9.5.6	Tools: SONNIA (Self-Organizing Neural Network for Information Analysis)	461
9.5.7	Back-propagation Network	462
9.5.7.1	Architecture	462
9.5.7.2	Training	462
9.5.8	Tutorial: Neural Networks	463
9.5.9	Tasks for Neural Networks and Selection of an Appropriate Neural Network Method	464
9.6	Fuzzy Sets and Fuzzy Logic	465
9.6.1	Some Concepts	465
9.6.2	Application of Fuzzy Logic in Chemistry	466
9.7	Genetic Algorithms	467
9.7.1	Representation and Encoding of Chromosomes	468
9.7.2	Initialization of Individuals	468
9.7.3	Fitness and Objective Function	469
9.7.4	Selection Functions	469
9.7.5	Genetic Operators	470
9.7.6	Tutorial: Selection of Relevant Descriptors in a Structure–Activity Study	471
9.7.6.1	Example: Drug Design	471
9.8	Data Mining	472
9.8.1	Classification	473
9.8.2	Clustering and Detection of Similarities	473
9.8.3	Prediction and Regression	473
9.8.4	Association	473
9.8.5	Detection of Descriptions	474
9.8.6	Data Mining in Chemistry	474
9.9	Visual Data Mining	474
9.9.1	Advantages of Visual Data Mining Approaches	475
9.9.2	Information Visualization Techniques	476
9.9.2.1	Data Types	476
9.9.2.2	Visualization Techniques	476
9.9.2.3	Interaction and Distortion Techniques	478
9.10	Expert Systems	478
9.10.1	Architecture of Expert Systems	478
9.10.2	Tasks of Expert Systems	479
9.10.3	Expert Systems in Chemistry	480
9.10.3.1	DENDRAL	480
9.10.3.2	EROS	481
10	Applications	487
10.1	Prediction of Properties of Compounds	487
10.1.1	Introduction	487
10.1.2	Linear Free Energy Relationships (LFER)	489
10.1.3	Quantitative Structure–Property Relationships (QSPR)	489

10.1.3.1	Structure Representation	489
10.1.3.2	Descriptor Analysis	490
10.1.3.3	Model Building	490
10.1.4	Estimation of Octanol/Water Partition Coefficient ($\log P_{ow}$)	492
10.1.4.1	Other Substructure-Based Methods	493
10.1.4.2	QSPR Models	494
10.1.5	Estimation of Aqueous Solubility ($\log S$)	495
10.1.5.1	Solubility Prediction Methods	495
10.1.5.2	Tutorial: Developing Models for Solubility Prediction with 18 Topological Descriptors	498
10.1.5.3	Models with 32 Radial Distribution Function Values and Eight Additional Descriptors	501
10.1.6	Prediction of the Toxicity of Compounds	504
10.1.6.1	How to Quantify Toxicity	504
10.1.6.2	Modeling Toxicity	504
10.1.7	Tutorial: Classifying Compounds into Different Modes of Action	508
10.1.8	Conclusion and Future Outlook	511
10.2	Structure–Spectra Correlations	515
10.2.1	Introduction	515
10.2.2	Molecular Descriptors	516
10.2.2.1	Fragment-Based Descriptors	516
10.2.2.2	Topological Structure Codes	516
10.2.2.3	Three-Dimensional Molecular Descriptors	517
10.2.3	^{13}C NMR Spectra	518
10.2.4	^1H NMR Spectra	520
10.2.4.1	Prediction of Chemical Shifts	520
10.2.4.2	Tools: Prediction of ^1H NMR Chemical Shifts	524
10.2.5	Infrared Spectra	529
10.2.5.1	Overview	529
10.2.5.2	Infrared Spectra Simulation	530
10.2.5.3	Tools: TeleSpec – Online Service for the Simulation of Infrared Spectra	530
10.2.6	Mass Spectra	534
10.2.7	Computer-Assisted Structure Elucidation	535
10.3	Chemical Reactions and Synthesis Design	542
10.3.1	The Prediction of Chemical Reactions	542
10.3.1.1	Introduction	542
10.3.1.2	Knowledge Extraction from Reaction Databases	544
10.3.1.3	Tutorial: Prediction of the Regiochemistry in Pyrazole Synthesis	545
10.3.1.4	CAMEO	549
10.3.1.5	EROS	550
10.3.1.6	Tutorial: Modeling the Degradation of <i>s</i> -Triazine Herbicides in Soil	553
10.3.1.7	Biochemical Pathways	556
10.3.1.8	Tutorial: Multidimensional Searching in Biochemical Pathways	564
10.3.2	Computer-Assisted Synthesis Design	567

10.3.2.1	Introduction	567
10.3.2.2	Basic Terms	569
10.3.2.3	Concepts for Computer-Assisted Organic Synthesis	573
10.3.2.4	Synthesis Design Systems	573
10.3.2.5	Tutorial: Synthesis Design with the WODCA Program	585
10.4	Drug Design	597
10.4.1	Introduction	597
10.4.2	Some Economic Considerations Affecting Drug Design	598
10.4.3	Definitions of some Terms in the Context of Drug Design	599
10.4.4	The Drug Discovery Process	600
10.4.4.1	Target Identification and Validation	600
10.4.4.2	Lead Finding and Optimization	601
10.4.4.3	Preclinical and Clinical Trials	602
10.4.5	Fields of Application of Chemoinformatics in Drug Design	602
10.4.5.1	Subset Selection and Similarity/Diversity Search	602
10.4.5.2	Analysis of HTS Data	603
10.4.5.3	Virtual Screening	603
10.4.5.4	Design of Combinatorial Libraries	604
10.4.5.5	Further Issues	605
10.4.6	Ligand- and Structure-based Drug Design	605
10.4.6.1	Ligand-Based Drug Design	607
10.4.6.2	Structure-Based Drug Design	608
10.4.7	Applications	612
10.4.7.1	Distinguishing Molecules of Different Biological Activities and Finding a New Lead Structure – An Example of Ligand-Based Drug Design	612
10.4.7.2	Examples of Structure-Based Drug Design	615
10.4.8	Outlook – Future Perspectives	616
11	Future Directions	619
	Appendix	627
A.1	Software Development	627
A.1.1	Programming Languages	627
A.1.2	Object-Oriented Programming	628
A.1.3	Universal Modeling Language (UML)	628
A.1.4	Design Patterns	629
A.1.5	Graphical User Interface	629
A.1.6	Source Code Documentation	629
A.1.7	Version Control	630
A.2	Mathematical Excursion into Matrices and Determinants	632
A.2.1	Mathematical Example	633
A.2.2	Chemical Example of an Atom Connectivity Matrix	633
	Index	635