

Muthukumarasamy Karthikeyan • Renu Vyas

Practical Chemoinformatics

 Springer

Contents

1	Open-Source Tools, Techniques, and Data in Chemoinformatics	1
1.1	Chemoinformatics	2
1.1.1	Open-Source Tools	2
1.1.2	Introduction to Programming Languages	3
1.2	Chemical Structure Representation	8
1.3	Code for Including the Editor Applet in JChemPaint	9
1.4	Definition of Templates (Polygons, Benzene, Bond, Atom, etc.)	9
1.5	Free Tools	10
1.6	Academic Programs	11
1.6.1	Marvin Sketch	11
1.6.2	ACD Labs	12
1.7	Commercial Tools	12
1.7.1	ChemDraw	12
1.7.2	Schrodinger	14
1.7.3	MOE (CCG)	14
1.7.4	Accelrys	14
1.8	A Practice Tutorial	15
1.8.1	Interconversion of Name/SMILES to Structure and Vice Versa	15
1.9	Introduction to Chemical Structure Formats	20
1.9.1	Linear Format	20
1.9.2	Graph-based Representation (2D and 3D formats)	21
1.9.3	Connection Tables	22
1.9.4	FILE FORMATS	22
1.10	2D and 3D Representation	30
1.10.1	Code for 3D Structure Generation in ChemAxon	31
1.10.2	A Practice Tutorial	31
1.11	Abstract Representation of Molecules	32
1.12	File Format Exchange	35
1.12.1	A Practice Tutorial	36
1.12.2	Code for Reading a Molecule, checking the Num- ber of Atoms, and Writing a SMILES String	38

1.12.3	Code for Reading a SMILES String in Python	39
1.13	Similarity and Fingerprint Analysis	39
1.13.1	Simple Fingerprints (Structural Keys)	41
1.13.2	Hashed Fingerprints	42
1.13.3	A Practice Tutorial	44
1.14	Molecular Similarity	45
1.14.1	Exact Structure Search	46
1.14.2	Substructure Search	47
1.14.3	Similarity Search	48
1.14.4	Subsimilarity Search	50
1.15	Search for Relationship	51
1.16	Similarity Measures	52
1.17	Molecular Diversity	55
1.18	Advanced Structure-handling Tools	56
1.18.1	CCML	56
1.19	ChemXtreme	56
1.19.1	Barcoding SMILES	57
1.19.2	Chem Robot	57
1.19.3	Image to Structure Tools	58
1.19.4	CLide	59
1.19.5	Advanced Structure Computation Platforms	59
1.20	Virtual Library Enumeration	59
1.21	Clustering	60
1.22	Databases	60
1.22.1	Database Server My SQL	62
1.22.2	Code for Connecting to a MySQL Database	63
1.22.3	A Practice Tutorial	64
1.22.4	Creating and Hosting Database	67
1.22.5	A Practice Tutorial	67
1.22.6	Hosting the Database	71
1.22.7	Chemical Databases	74
1.22.8	Do It Yourself (DIY)	85
1.22.9	Questions	89
	References	89
2	Cheminformatics Approach for the Design and Screening of Focused Virtual Libraries	93
2.1	Introduction to Structure–Property Correlations	93
2.1.1	Descriptors	94
2.1.2	Online Property Prediction Tools	108
2.1.3	Virtual Library Generation (Enumeration)	111
2.1.4	Virtual Screening	121
2.1.5	Thumb Rules for Computing Molecular Properties	128
2.1.6	Do it Yourself	128
2.1.7	Questions	129
	References	129

3 Machine Learning Methods in Chemoinformatics for Drug Discovery	133
3.1 Introduction	133
3.2 Machine Learning Models for Predictive Studies	134
3.3 Machine Learning Methods	136
3.4 Open-Source Tools for Building Models for Drug Design	139
3.4.1 Library for Support Vector Machines (LibSVM)	139
3.4.2 Waikato Environment for Knowledge Analysis (WeKa) ...	141
3.4.3 R Program	151
3.5 Free Tools for Machine Learning	152
3.5.1 An Example of SVR-based Machine Learning	152
3.5.2 Rapid Miner	160
3.6 Commercial Tools for Building ML Models	164
3.6.1 Molecular Operating Environment (MOE)	164
3.6.2 IBM SPSS	176
3.6.3 Matrix Laboratory (MATLAB)	178
3.7 Genetic Programming-Based ML Models	179
3.7.1 A Practical Demonstration of GP-Based Software	185
3.8 Thumb Rules for Machine Learning-Based Modelling	189
3.9 Do it Yourself (DIY)	191
3.10 Questions	191
References	192
4 Docking and Pharmacophore Modelling for Virtual Screening	195
4.1 Introduction	195
4.2 A Practice Tutorial: Docking Using a Commercial Tool	196
4.3 Docking Using Open Source Software	211
4.3.1 Autodock Steps	212
4.3.2 Docking Using AutoDock Vina	220
4.4 Other Docking Algorithms	223
4.4.1 Induced Fit Docking	224
4.4.2 Flexible Protein Docking	225
4.4.3 Blind Docking	226
4.4.4 Cross Docking	226
4.4.5 Docking and Site-Directed Mutagenesis	229
4.5 Protein-Protein Docking	231
4.6 Pharmacophore	234
4.6.1 Pharmacophore Modelling in SCHRÖDINGER	235
4.6.2 Finding Pharmacophore Features Using MOE	248
4.7 Open Source Tools for Pharmacophore Generation	253
4.8 Rules of Thumb for Structure-Based Drug Design	254
4.9 Do it Yourself Exercises	260
4.10 Questions	261
References	267

5 Active Site-Directed Pose Prediction Programs for Efficient Filtering of Molecules	271
5.1 Introduction	271
5.2 A Practice Tutorial for Predicting Active Site Using SiteMap	272
5.3 A Practice Tutorial for Active Site Prediction Using MOE	276
5.4 Free Online Tools for Active Site Prediction	279
5.5 Homology Modelling	282
5.6 A Practice Tutorial for Homology Modelling	285
5.7 Model Validation Using Online Servers	295
5.8 Receptor-Based Pharmacophore	296
5.9 Studies on Active Site Structural Features	298
5.9.1 Application of Active Site Features in Chemoinformatics ...	300
5.10 Thumb Rules for Active Site Identification and Homology Modelling	312
5.11 Do it Yourself Exercises	313
5.12 Questions	313
References	313
6 Representation, Fingerprinting, and Modelling of Chemical Reactions	317
6.1 Introduction	318
6.2 Reaction Representation in Computers	318
6.3 Computational Methods in Reaction Modelling	318
6.3.1 Empirical and Semiempirical Methods	319
6.3.2 Molecular Mechanics Methods	320
6.3.3 Molecular Dynamics Methods	321
6.3.4 Statistical Mechanics and Thermodynamics	321
6.3.5 The Quantum Mechanical/molecular Mechanical Approach	322
6.3.6 Modelling the Transition State of Reactions	322
6.4 TS Modelling of Organic Transformations	324
6.4.1 Name Reactions	324
6.4.2 A Practice Tutorial for Transition State and Intrinsic Reaction Coordinate Modelling	326
6.4.3 A Practice Tutorial Using Maestro–Jaguar	338
6.4.4 A Practice Tutorial Using Spartan	344
6.5 Reaction-Searching Approaches and Tools	347
6.5.1 Chemical Ontologies Approach for Reaction Searching	351
6.5.2 Reaction Searching Using Fingerprints-Based Approach ...	354
6.5.3 Tools for Reaction Searching	359
6.6 Reaction Databases	363
6.6.1 Tools for Reaction Library Enumeration	364
6.6.2 A Practice Tutorial	365
6.7 Artificial Intelligence in Chemical Synthesis	366
6.8 Modelling Enzymatic Reactions	369

6.9	Thumb Rules for Performing Reaction Representation, Fingerprints, and Modelling	369
6.10	Do it Yourself	371
6.11	Questions	371
	References	371
7	Predictive Methods for Organic Spectral Data Simulation	375
7.1	Introduction	376
7.2	Fragment-Based Drug Discovery	378
7.3	Spectra Prediction Methods	384
7.4	Spectra Prediction Tools	384
7.5	Open-Source Tools	385
7.5.1	GAMESS	385
7.6	Proprietary Tools	385
7.6.1	ACD/NMR Predictors	385
7.6.2	Cambridgesoft Chem3D	385
7.6.3	Jaguar	385
7.6.4	Gaussian	390
7.6.5	ADF	391
7.6.6	MestreNova	392
7.6.7	Spartan	396
7.6.8	Spectral Databases	399
7.7	Spectra Viewer Programs	404
7.8	In-House Tools for Spectra Prediction	404
7.9	Code to Generate Proton and Carbon NMR Spectrum	406
7.10	Thumb Rules for Spectral Data Handling and Prediction	409
7.11	Do it Yourself	410
7.12	Questions	411
	References	412
8	Chemical Text Mining for Lead Discovery	415
8.1	What is Text Mining?	416
8.1.1	Text Mining vis-a-vis Data Mining	416
8.1.2	A Snippet of Java Code Using the Above URL	418
8.2	What are the Components of Text Mining?	419
8.3	Text-mining Methods	421
8.3.1	Statistics/ML-based Approach	422
8.3.2	Rule-based Approach	423
8.4	Why Text Mining	424
8.5	General Text-mining Tools	424
8.5.1	A Practice Tutorial with an Open-source Tool	425
8.5.2	R Program for Text Mining	430
8.6	Free Tools for Text Mining	434
8.7	Biomedical Text Mining	434
8.8	Chemically Intelligent Text-mining Tools	435

8.9	In-house Tools for Text-mining Applications for Chemoinformatics	437
8.9.1	Java Code Snippet for Data Distribution	441
8.10	Thumb Rules While Performing and Using Text-mining Results ...	445
8.11	Do it Yourself	445
8.12	Questions	445
	References	445
9	Integration of Automated Workflow in Chemoinformatics	
	for Drug Discovery	451
9.1	What is a Workflow?	451
9.2	Need for Workflows	452
9.3	General Workflows in Bioinformatics	453
9.4	General Workflows in Chemistry Domain	453
9.4.1	Accelrys Pipeline Pilot	453
9.4.2	IDBS Chemsense (Inforsense Suite)	454
9.4.3	CDK Taverna	455
9.4.4	KNIME	455
9.4.5	Workflow Examples	467
9.4.6	Workflow for QSAR (Anti-cancer)	469
9.5	Schrodinger KNIME Extensions	470
9.5.1	A Practice Tutorial	473
9.6	Other KNIME Extensions	481
9.6.1	MOE(CCG)	481
9.6.2	ChemAxon	483
9.7	Protein-Ligand Analysis-Based Workflows for Drug Discovery	483
9.7.1	A Practice Tutorial for Protein-Ligand Fingerprint Generation	486
9.8	Prolix	489
9.9	J-ProLINE: An In-house-developed Chem-Bioinformatics Workflow Application	489
9.10	Targetlikeness Score	496
9.11	Databases and Tools	496
9.12	Thumb Rules for Generating and Applying Workflows	496
9.13	Do it Yourself	497
9.14	Questions	497
	References	497
10	Cloud Computing Infrastructure Development for Chemoinformatics	501
10.1	What is a Portal?	501
10.2	Need for Development of Scientific Portals	502
10.3	Components of a Portal	502
10.4	Examples of Portal Systems	503

10.5	A Practice Tutorial for Portal Creation	504
10.5.1	Custom Database connection and Display Table with Paginator via portlet in Liferay Portal	509
10.6	A Practice Tutorial for Development of Portlets for Chemoinformatics	512
10.6.1	Marvin Sketch Portlet	512
10.6.2	JME Portlet	515
10.6.3	Jchempaint Portlet	515
10.7	Mobile Computing	516
10.7.1	Android Applications for Chemoinformatics	517
10.8	Need of High-Performance Computing in Chemoinformatics	526
10.9	Thumb Rules for Developing and Using Scientific Portals and Mobile Devices for Computing	526
10.10	Do it Yourself Exercises	526
10.11	Questions	527
	References	527
Index	529