

Contents

About the Editor — VII

List of contributing authors — XIX

Fidele Ntie-Kang

Editorial: Chemoinformatics at the service of natural product discovery — 1

References — 3

Part I: Foundational Chemoinformatics Concepts for Natural Product-based Drug Discovery

Berhanu M. Abegaz and Henok H. Kinfe

1 Secondary metabolites, their structural diversity, bioactivity, and ecological functions: An overview — 7

1.1 Introduction — 8

1.1.1 Definitions — 8

1.1.2 The roles of secondary metabolites — 8

1.1.3 Classification of secondary metabolites — 9

1.1.4 Structural complexity of secondary metabolites — 9

1.1.5 Weaning and re-emergence of interest in secondary metabolites — 10

1.2 Alkaloids — 11

1.2.1 Tropane alkaloids — 13

1.2.2 Piperidine alkaloids — 15

1.2.3 Xanthine alkaloids — 15

1.2.4 Benzyl isoquinoline alkaloids — 17

1.3 Fatty acids — 19

1.3.1 Prostaglandins – metabolites of arachidonic acid — 21

1.3.2 Macrolides — 22

1.4 Polyketides — 22

1.5 Phenyl propanoids and aromatic polyketides — 26

1.5.1 Phenyl propanoids — 26

1.5.2 Aromatic polyketides-flavonoids — 28

1.6 Terpenoids — 30

1.6.1 Monoterpenes — 32

1.6.2 Sesquiterpenes — 33

1.6.3 Diterpenes — 36

1.6.4 Triterpenes — 40

1.6.5 Steroids — 43

References — 49

Mohd Athar, Alfred N. Sona, Boris D. Bekono and Fidele Ntie-Kang

2 Fundamental physical and chemical concepts behind “drug-likeness” and “natural product-likeness” — 55

2.1 Introduction — 55

2.1.1 What makes drugs “drug-like”? — 55

2.1.2 Structural features affect ADMET — 56

2.1.3 Christopher Lipinski, the father of “drug-likeness” — 56

2.1.4 Bioavailability *versus* “drug-likeness” — 57

2.1.5 Molecular properties used for predicting bioavailability and simple filters — 58

2.1.6 Property filters for ADME/tox predictions as rough estimates for “drug-likeness” — 59

2.1.7 The distinction between “drug” and “leads” — 60

2.1.8 Rationale of this work — 61

2.2 Re-canvassing drug-likeness “rule of thumbs” for natural products — 62

2.3 The need to increase the chemical space and target space for addressing biological complexity — 63

2.4 Why are natural products distinct from synthetic drugs? — 64

2.4.1 They have a wider chemical space and increased diversity — 64

2.4.2 Unique chemical properties of natural products at the atomic and molecular levels — 66

2.4.3 Greater biocompatibility — 68

2.4.4 Natural product-likeness in terms of natural product derived fragments — 68

2.5 Quantifying “natural product-likeness” — 69

2.6 Natural product space in chemical space — 70

2.7 Drug-like and non drug-like natural products — 72

2.8 Do orally administered drugs in the market violate Lipinski’s “rules”? — 73

2.8.1 Talaprevir — 73

2.8.2 Lipitor — 73

2.9 Conclusions — 74

References — 77

Fidele Ntie-Kang, Kennedy D. Nyongbela, Godfred A. Ayimele and Suhaib Shekfeh

3 “Drug-likeness” properties of natural compounds — 81

3.1 Introduction — 81

3.2 Natural products as drugs — 83

3.2.1	The proportion of natural products in catalogs of drugs —	83
3.2.2	The future of natural product drug discovery —	84
3.3	Natural <i>versus</i> synthetic drugs —	86
3.3.1	The uniqueness and potential of natural products for drug discovery —	86
3.3.2	The complexity and diversity of natural product scaffolds —	87
3.4	Navigating the natural product chemical space —	88
3.4.1	The Universal Natural Products Database <i>versus</i> FDA-approved drugs —	88
3.4.2	Antidiabetic medicinal plant-based bioactive natural products <i>versus</i> known antidiabetic drugs —	90
3.4.3	Analysis of the World of Molecular Bioactivity dataset against the Dictionary of Natural Products —	90
3.5	The design of “drug-like” natural product libraries and implications for drug discovery —	91
3.5.1	Strategy for designing a library with focused properties —	91
3.5.2	“Drug-likeness” prediction on available electronic natural product libraries for chemoinformatics analysis —	92
3.6	Computational methods for estimating drug-likeness and ADME/T —	93
3.7	Computational methods for estimating natural product-likeness —	94
3.7.1	The natural product-likeness score —	94
3.7.2	Implementations of the natural product-likeness score —	94
3.8	Machine-learning methods to classify drugs from nondrugs —	95
3.9	A binary QSAR model to classify natural products from synthetic molecules —	96
3.10	Conclusions —	97
References —		98

Fernanda I. Saldívar-González, B. Angélica Pilon-Jiménez
and José L. Medina-Franco

4 Chemical space of naturally occurring compounds — 103

4.1	Introduction —	103
4.1.1	Importance of chemical space of natural products —	104
4.1.2	Overview of a representative cheminformatic analysis of natural products databases —	108
4.1.3	Novel cheminformatic approaches to navigate the chemical space of natural products —	110
4.2	Coverage of natural product chemical space from diverse sources —	111

4.2.1	Plants, fungi, cyanobacteria, and marine —	111
4.2.2	Comparison of natural products with themselves and other reference libraries —	112
4.3	Diversity in chemical space: quantification and implications —	114
4.3.1	Molecular properties —	115
4.3.2	Structural fingerprints —	116
4.3.3	Scaffolds —	116
4.3.4	Consensus diversity analysis: consensus diversity plots —	118
4.4	Conclusions and future directions —	118
	References —	120

David J. Newman

5 From natural products to drugs — 125

5.1	Introduction —	125
5.2	Natural product and natural product-derived structures as approved drugs 2010–2017 —	126
5.3	Discussion on the selected compounds in Table 5.2 —	129
5.3.1	Romidepsin (Istodax®; Figure 5.1) —	129
5.3.2	Fingolimod (Gilenya®; Figure 5.2) and Derivatives —	130
5.3.3	Eribulin (Halaven®; Figure 5.3) —	130
5.3.4	Fidaxomicin (Difcid®; Figure 5.4) —	131
5.3.5	Ceftaroline fosamil acetate (Teflaro®; Figure 5.5) and Cefozopran —	132
5.3.6	Zucapsaicin (Zuacta®; Figure 5.6) —	132
5.3.7	Brentuximab vedotin (Adcetris®; Figure 5.7) —	133
5.3.8	Ingenol mebutate (Picato®; Figure 5.8) —	133
5.3.9	Omacetaxine mepesuccinate (homoharingtonine) (Ceflatonin®; Figure 5.9) —	134
5.3.10	Cholic acid (Orphacol®; Figure 5.10) —	135
5.3.11	Trastuzumab emtansine (Kadcyla®; Figure 5.11) and Linkers —	135
5.3.12	Dalbavancin (Dalvance®; Figure 5.12) —	136
5.3.13	Cangrelor tetrasodium (Kengreal®; Figure 5.13) —	137
5.3.14	Dronabinol hemisuccinate (Syndros®; Figure 5.14) —	138
5.3.15	Abaloparatide (Tymlos®; Figure 5.15) —	138
5.3.16	Forodesine hydrochloride (Mundesine®; Figure 5.16) —	139
5.3.17	Midostaurin (Rydapt®; Figure 5.17) —	139
5.4	Some selected agents in clinical use and/or trials that are np-based —	140
5.4.1	Antibody-drug conjugates —	140
5.4.2	Nucleoside-based agents —	143
5.4.3	Beta-lactam antibiotics —	147

5.4.4	One S*/NM. pitavastatin magnesium (zipitamag®; Figure 5.40) — 150
5.5	In conclusion — 150
	References — 151

Part II: Chemoinformatics Tools and Methods for Natural Product Structure Elucidation

Aurélien F. A. Moumbock, Fidele Ntie-Kang, Sergi H. Akone, Jianyu Li, Mingjie Gao, Kiran K. Telukunta and Stefan Günther

6	An overview of tools, software, and methods for natural product fragment and mass spectral analysis — 157
6.1	Introduction — 157
6.2	Mass spectral databases — 159
6.2.1	Available databases — 159
6.2.2	The general workflow of MS annotation — 161
6.3	Software used in computational MS — 163
6.3.1	Rule-based fragmentation prediction — 163
6.3.2	Combinatorial fragmentation prediction — 164
6.3.3	Mass spectral classifiers — 166
6.3.4	Fragmentation trees analysis — 167
6.3.5	Fingerprint fragmentation (advanced mass spectral classifiers) — 168
6.3.6	Competitive fragmentation modeling — 168
6.3.7	Quantum chemistry fragmentation — 169
6.4	Conclusions — 171
	References — 171

Marilia Valli, Helena Mannocho Russo, Alan Cesar Pilon, Meri Emili Ferreira Pinto, Nathalia B. Dias, Rafael Teixeira Freire, Ian Castro-Gamboa and Vanderlan da Silva Bolzani

7	Computational methods for NMR and MS for structure elucidation I: software for basic NMR — 177
7.1	Introduction — 177
7.2	NMR predictions — 178
7.3	Computational methods for structure elucidation — 180
7.3.1	CASE programs – examples — 181
7.4	Mass spectrometry and NMR for peptide sequencing — 185
7.4.1	Fundamental aspects in MS of peptides and proteins — 185
7.4.2	Peptide fragmentation and sequencing — 186

7.4.3	Important rules for peptide spectrum interpretation —	187
7.4.4	Software and database tools —	189
7.4.5	Automated <i>de novo</i> sequencing —	191
7.4.6	Protein identification in data banks —	192
7.4.7	Peptide identification using NMR —	193
7.5	Conclusions and future directions —	196
	References —	197

Marilia Valli, Helena Mannocho Russo, Alan Cesar Pilon, Meri Emili Ferreira Pinto, Nathalia B. Dias, Rafael Teixeira Freire, Ian Castro-Gamboa and Vanderlan da Silva Bolzani

8 Computational methods for NMR and MS for structure elucidation II: database resources and advanced methods — 205

8.1	Introduction —	205
8.2	Databases for assisting structure elucidation —	206
8.3	Peak pattern recognition (PPR): a method of analysis and pattern recognition applied to NMR —	213
8.4	Dereplication methods and MN —	218
8.4.1	MN for reinvestigating a determined species —	221
8.4.2	MN for the study of extract libraries —	223
8.4.3	MN for the study of bioactive compounds —	223
8.5	Conclusions and future directions —	225
	References —	226

Gabin T. M. Bitchagno and Serge A. F. Tanemossu

9 Computational methods for NMR and MS for structure elucidation III: More advanced approaches — 229

9.1	Introduction —	229
9.1.1	Nuclear magnetic resonance (NMR) —	229
9.1.2	Mass spectrometry (MS) —	230
9.1.3	Rationale of this work —	230
9.2	Sequence in NMR structure elucidation —	231
9.2.1	Steps in NMR structure determination —	231
9.2.2	Peak picking —	231
9.2.3	Resonance assignment —	232
9.3	Computer-assisted structure elucidation (CASE) —	233
9.3.1	Logic for structure determination (LSD) —	234
9.3.2	SENECA —	235
9.3.3	ACD/structure elucidator —	236
9.4	Computation in MS —	239
9.4.1	Databases in MS: Importance and uses —	240

9.4.2	Empirical methods using rule-based fragmentation model —	241
9.4.3	Empirical methods using combinatorial fragmentation model —	242
9.5	Conclusion —	244
References —		245

Part III: Chemoinformatics Tools and Methods for Lead Compound Discovery and Development

Eleni Koulouridi, Marilia Valli, Fidele Ntie-Kang and Vanderlan da Silva Bolzani

10	A primer on natural product-based virtual screening —	251
10.1	Introduction to databases (definition/construction) —	251
10.2	Guidelines when preparing a compound database —	256
10.2.1	The purpose of the compound database —	256
10.2.2	Molecular properties —	256
10.2.3	3D representations —	257
10.2.4	Format for storing chemical data —	257
10.2.5	Compounds with problematic functional groups —	258
10.2.6	Compounds likely to interfere in assays —	258
10.2.7	Diversity —	259
10.2.8	Tools for library design —	259
10.2.9	Sources —	260
10.3	The use of virtual databases in computational chemistry —	260
10.4	Tools for handling and managing database resources —	263
10.5	Public and private databases —	264
10.6	Databases for natural products —	265
10.7	The purchasable chemical space of currently available natural product libraries —	277
10.8	A watchword when using natural product databases —	278
10.9	Conclusion —	280
References —		281

Rita C. Guedes and Tiago Rodrigues

11	Drug target prediction using chem- and bioinformatics —	291
11.1	Introduction —	291
11.2	Pharmacophore screening —	293
11.3	Molecular docking —	295
11.4	Similarity-based searches —	296
11.5	Statistical learning and machine intelligence —	299
11.6	Outlook —	304
References —		305

Annalisa Maruca, Francesca Alessandra Ambrosio, Antonio Lupia, Isabella Romeo, Roberta Rocca, Federica Moraca, Carmine Talarico, Donatella Bagetta, Raffaella Catalano, Giosuè Costa, Anna Artese and Stefano Alcaro

12 Computer-based techniques for lead identification and optimization I:

Basics — 311

- 12.1 Introduction — 311
- 12.2 Virtual screening: A tool for lead identification — 312
 - 12.2.1 Drug-like properties prediction of natural compounds — 313
 - 12.2.2 Drug-likeness of commercially available compound libraries for drug discovery — 314
 - 12.2.3 Software tools for pre-filtering of the compounds libraries — 315
 - 12.2.4 Compound libraries preparation — 315
 - 12.2.5 Pharmacophore — 316
 - 12.2.6 Successful applications of VS — 321
- 12.3 Molecular mechanics simulations — 323
 - 12.3.1 Monte Carlo simulations — 324
 - 12.3.2 Molecular Dynamics: A classical tool driving drug discovery — 325
 - 12.3.3 MD-derived successful stories of natural compounds — 326
- 12.4 Conclusions — 327
- References — 328

Antonio Lupia, Federica Moraca, Donatella Bagetta, Annalisa Maruca, Francesca Alessandra Ambrosio, Roberta Rocca, Raffaella Catalano, Isabella Romeo, Carmine Talarico, Francesco Ortuso, Anna Artese and Stefano Alcaro

13 Computer-based techniques for lead identification and optimization II:

Advanced search methods — 333

- 13.1 Introduction — 333
- 13.2 Metadynamics: An enhanced sampling method to accelerate rare events — 334
 - 13.2.1 The well-tempered metadynamics and the reweighting procedure — 336
 - 13.2.2 Funnel Metadynamics: a valuable method for an accurate estimation of the ligand–binding free energy — 337
- 13.3 *Dynophores*: The use of molecular dynamics in pharmacophore modeling — 343
 - 13.3.1 *Dynophores*: A case study on the catalytic site modulation of arginase — 344
- 13.4 Pharmacophore models deriving from MD simulation and virtual screening — 348
 - 13.4.1 Virtual screening performance and comparison of pharmacophore models — 350

13.4.2	Common hit approach —	350
13.5	Progress in computational technology: The role of the GPU —	356
13.6	Conclusions —	357
	References —	358

Ricardo Bruno Hernández-Alvarado, Abraham Madariaga-Mazón and Karina Martínez-Mayorga

14	Prediction of toxicity of secondary metabolites —	361
14.1	Introduction —	361
14.2	Predictive models of toxicological endpoint —	361
14.2.1	Toxicological endpoints commonly predicted —	362
14.2.2	Databases of toxic compounds —	363
14.2.3	Methods and software to predict toxicity —	365
14.3	Toxicity prediction in regulatory settings —	369
14.4	Case studies —	371
14.4.1	Toxicity prediction using different software —	371
14.4.2	Toxicity prediction of an entire database of natural products with antidiabetic activity —	372
14.4.3	Toxicity prediction of natural compounds from the African flora —	373
14.5	Conclusions —	374
	References —	374

Part IV: Case Studies

Samuel Egieyeh, Sarel F. Malan and Alan Christoffels

15	Cheminformatics techniques in antimalarial drug discovery and development from natural products 1: basic concepts —	381
15.1	Introduction —	381
15.1.1	Natural products and malaria —	381
15.1.2	Cheminformatics: definitions and background —	382
15.2	Cheminformatics techniques in the development of antimalarial drugs from natural products —	383
15.3	Molecular descriptors —	384
15.4	LE metrics —	385
15.4.1	Ligand efficiency —	385
15.4.2	Ligand lipophilicity efficiency (LLE) —	386
15.4.3	Ligand efficiency dependent lipophilicity (LELP) —	386
15.5	Data mining —	387
15.5.1	Data mining: molecular similarity and exploration of chemical space —	387

15.6	Structure-activity landscape: identifying activity cliffs —	388
15.7	Conclusions —	389
	References —	391

Samuel Egieyeh, Sarel F. Malan and Alan Christoffels

16	Cheminformatics techniques in antimalarial drug discovery and development from natural products 2: Molecular scaffold and machine learning approaches —	397
16.1	Introduction —	397
16.2	Exploration of molecular scaffolds —	398
16.3	Bioactivity predictive models via machine-learning —	399
16.3.1	Some machine-learning algorithms —	402
16.3.2	Platforms and steps for building predictive models via ML —	403
16.3.3	Applicability domain (AD) of a predictive model —	406
16.4	Conclusions —	406
16.5	Case study: Cheminformatic profiling of selected natural product with in vitro antiplasmodial activities (NAA) —	407
16.5.1	Dataset —	407
16.5.2	Analysis of molecular descriptors of NAA and CAD —	407
16.5.3	Molecular similarity and exploration of chemical space of NAA and CAD —	407
16.5.4	Exploration of molecular scaffolds from NAA —	407
16.5.5	Predictive classifier models built from NAA using ML approach —	410
	References —	412

Eleni Koulouridi, Sergi Herve Akone, Marilia Valli, Vanderlan da Silva Bolzani, Fernanda I. Saldívar-González, Angélica Pilón-Jiménez, José L. Medina-Franco, Berhanu M. Abegaz, Hanok Kinfe, Aurélien F. A. Moumbock, Mohd Athar, Gabin T.M. Bitchagno, Serge A.T. Fobofou, David Newman, Rita C. Guedes, Tiago Rodrigues, Maruca Annalisa, Bagetta Donatella, Lupia Antonio, Ricardo B. Hernández-Alvarado, Abraham Madariaga-Mazón, Karina Martinez-Mayorga, Stefan Günther and Fidele Ntie-Kang

Glossary of terms used in chemoinformatics of natural products: fundamental principles —	417
References —	441

Index —	443
----------------	------------