

# Contents

List of contributors — XV

About the Editor — XIX

## Part I: Advanced Chemoinformatics Concepts for Natural Product-based Drug Discovery — 1

Fidele Ntie-Kang, Daniel Svozil

- 1 An enumeration of natural products from microbial, marine and terrestrial sources — 3**
  - 1.1 Introduction — 3
  - 1.2 Attempts to obtain numbers by the generation of universal natural compound libraries — 7
    - 1.2.1 Enumeration of compound annotations in commercial libraries — 7
    - 1.2.2 Academic and other open-access efforts — 10
    - 1.2.3 Enumerating natural product sample collections — 11
  - 1.3 Number of unique molecules in the country and regional datasets — 12
  - 1.4 Number of unique molecules in disease or therapeutic use datasets — 12
  - 1.5 Enumeration by compound types — 12
  - 1.6 Global enumeration by major natural product pools — 15
    - 1.6.1 The challenge of classifying natural products into terrestrial, marine and microbial origins — 15
    - 1.6.2 Natural products of microbial origin — 16
    - 1.6.3 Natural products of marine origin — 17
    - 1.6.4 Natural products from food plants — 19
    - 1.6.5 Human metabolites — 21
  - 1.7 Conclusions — 21
  - References — 22

Joelle Ngo Hanna, Boris D. Bekono, Luc C. O. Owono, Flavien A. A. Toze, James A. Mbah, Stefan Günther and Fidele Ntie-Kang

- 2 A chemoinformatic analysis of atoms, scaffolds and functional groups in natural products — 31**
  - 2.1 Introduction — 31
  - 2.2 An atom-based enumeration of natural product-like virtual libraries — 33

|       |                                                                                         |    |
|-------|-----------------------------------------------------------------------------------------|----|
| 2.3   | Chemoinformatic comparison between natural products of terrestrial and marine origins — | 34 |
| 2.4   | Scaffold analysis in natural products —                                                 | 42 |
| 2.4.1 | The case of the Dictionary of Natural Products (DNP) —                                  | 42 |
| 2.4.2 | The case of natural products from cyanobacteria —                                       | 43 |
| 2.4.3 | The case of the terpenoid chemical class —                                              | 45 |
| 2.5   | Functional group analysis in natural products —                                         | 47 |
| 2.6   | Conclusions —                                                                           | 51 |
|       | References —                                                                            | 52 |

Lena Y. E. Ekaney, Donatus B. Eni and Fidele Ntie-Kang

|          |                                                                                                                          |           |
|----------|--------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>3</b> | <b>Chemical similarity methods for analyzing secondary metabolite structures —</b>                                       | <b>57</b> |
| 3.1      | Introduction —                                                                                                           | 57        |
| 3.2      | Different ways of expressing chemical similarity —                                                                       | 58        |
| 3.2.1    | (2D) Fingerprint-based similarity —                                                                                      | 59        |
| 3.2.2    | Chemical similarity using physicochemical property descriptors —                                                         | 62        |
| 3.2.3    | Virtual screening based similarity search —                                                                              | 62        |
| 3.2.4    | SMILES based chemical similarity search —                                                                                | 63        |
| 3.2.5    | Hypothetical enumeration method for comparing natural products —                                                         | 65        |
| 3.3      | Two-dimensional <i>versus</i> three-dimensional similarity methods —                                                     | 65        |
| 3.3.1    | Enantiomers with “high” 2D similarity but dissimilar activities —                                                        | 67        |
| 3.3.2    | 3D-atom pair fingerprints —                                                                                              | 67        |
| 3.4      | Similarity search algorithms for phylogenetic relationship between plants based on similarity of secondary metabolites — | 68        |
| 3.5      | Towards the collapse of the concept of “chemotaxonomy”? —                                                                | 70        |
| 3.6      | Conclusion —                                                                                                             | 71        |
|          | References —                                                                                                             | 72        |

## Part II: Advanced Chemoinformatics Tools and Methods for Lead Compound Discovery and Development — 75

Abdulkarim Najjar, Abdurrahman Olğaç, Fidele Ntie-Kang and Wolfgang Sippl

|          |                                                                  |           |
|----------|------------------------------------------------------------------|-----------|
| <b>4</b> | <b>Fragment-based drug design of nature-inspired compounds —</b> | <b>77</b> |
| 4.1      | Introduction —                                                   | 77        |
| 4.2      | Methods for fragment screening —                                 | 78        |

|       |                                                                                  |    |
|-------|----------------------------------------------------------------------------------|----|
| 4.3   | Fragment optimization —                                                          | 80 |
| 4.4   | Strategies for fragment optimization —                                           | 81 |
| 4.4.1 | Growing —                                                                        | 81 |
| 4.4.2 | Linking —                                                                        | 82 |
| 4.4.3 | Merging —                                                                        | 82 |
| 4.5   | Characterization of NP-related fragments —                                       | 83 |
| 4.5.1 | Analysis of NP databases regarding the fragment-based definition —               | 84 |
| 4.5.2 | Fragmentation of NPs —                                                           | 85 |
| 4.6   | Case studies —                                                                   | 86 |
| 4.6.1 | The identification of antimalarial NP fragments —                                | 86 |
| 4.6.2 | Fragment-based structural optimization of an NP as an anticancer lead compound — | 90 |
| 4.7   | Challenges and limitations of the NPs by using FBDD —                            | 90 |
| 4.8   | Conclusion —                                                                     | 92 |
|       | References —                                                                     | 93 |

Fatima Baldo

|          |                                                                                            |           |
|----------|--------------------------------------------------------------------------------------------|-----------|
| <b>5</b> | <b>Prediction of modes of action of components of traditional medicinal preparations —</b> | <b>99</b> |
| 5.1      | Introduction —                                                                             | 99        |
| 5.2      | Genetic interaction methods —                                                              | 100       |
| 5.3      | Computational inference —                                                                  | 102       |
| 5.3.1    | Ligand-based target prediction —                                                           | 103       |
| 5.3.2    | Machine learning in ligand-based target prediction —                                       | 103       |
| 5.4      | Pathway annotations —                                                                      | 110       |
|          | References —                                                                               | 112       |

Fidele Ntie-Kang

|          |                                                                                                                                          |            |
|----------|------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>6</b> | <b>Mechanistic role of plant-based bitter principles and bitterness prediction for natural product studies I: Database and methods —</b> | <b>119</b> |
| 6.1      | Introduction —                                                                                                                           | 119        |
| 6.1.1    | Natural products and bitterness —                                                                                                        | 119        |
| 6.1.2    | Distinguishing between sweet, salty, sour, umami and bitter —                                                                            | 120        |
| 6.1.3    | Advantage and disadvantage of bitterness —                                                                                               | 121        |
| 6.1.4    | Rationale of this work —                                                                                                                 | 121        |
| 6.2      | BitterDB -the database of bitter compounds —                                                                                             | 121        |
| 6.3      | In silico methods for predicting bitterness —                                                                                            | 123        |
| 6.3.1    | Structure-based methods —                                                                                                                | 123        |
| 6.3.2    | Ligand-based methods —                                                                                                                   | 127        |
| 6.3.3    | Machine learning approaches —                                                                                                            | 129        |
| 6.4      | Bitterness-toxicity relationship —                                                                                                       | 131        |

|     |                   |
|-----|-------------------|
| 6.5 | Conclusions — 132 |
|     | References — 133  |

Fidele Ntie-Kang

|          |                                                                                                                                                            |
|----------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>7</b> | <b>Mechanistic role of plant-based bitter principles and bitterness prediction for natural product studies II: prediction tools and case studies — 137</b> |
| 7.1      | Introduction — 137                                                                                                                                         |
| 7.2      | Available tools for predicting bitterness — 138                                                                                                            |
| 7.2.1    | A note on machine learning methods — 138                                                                                                                   |
| 7.2.2    | BitterX – a support vector machines bitterness predictor — 143                                                                                             |
| 7.2.3    | BitterPredict – a decision trees-based tool for predicting taste from chemical structure — 145                                                             |
| 7.2.4    | BitterSweetForest – a random forest open access tool — 145                                                                                                 |
| 7.2.5    | e-Bitter – a free software for bitterness prediction — 146                                                                                                 |
| 7.2.6    | BitterSweet – a freely available state-of-the-art software for bitter <i>versus</i> sweet taste prediction — 146                                           |
| 7.3      | Bitter natural products from plants — 148                                                                                                                  |
| 7.3.1    | Why are some bitter plants medicinal? — 150                                                                                                                |
| 7.3.2    | Understanding the role of natural bitter compounds in plant defense mechanisms — 151                                                                       |
| 7.4      | Conclusions — 156                                                                                                                                          |
|          | References — 158                                                                                                                                           |

Abraham Madariaga-Mazón, Ricardo Bruno Hernández-Alvarado, Karla Olivia Noriega-Colima, Adriana Osnaya-Hernández and Karina Martinez-Mayorga

|          |                                                                         |
|----------|-------------------------------------------------------------------------|
| <b>8</b> | <b>Toxicity of secondary metabolites — 161</b>                          |
| 8.1      | Introduction — 161                                                      |
| 8.2      | Secondary metabolites: Sources, classification, and bioactivities — 162 |
| 8.2.1    | Chemical classification and bioactivities — 163                         |
| 8.2.2    | Toxicity of representative secondary metabolites — 167                  |
| 8.3      | Toxicity assessment in agrochemicals and drugs — 169                    |
| 8.3.1    | Toxicity assessment — 169                                               |
| 8.3.2    | Toxicity of drug candidates — 171                                       |
| 8.3.3    | Toxicity of agrochemicals — 171                                         |

|     |                   |
|-----|-------------------|
| 8.4 | Conclusions — 173 |
|     | References — 173  |

## **Part III: Chemoinformatics Tools for Natural Products Discovery in the Modern Age and Case Studies — 177**

Ramsay Soup Teoua Kamdem, Omonike Ogbale, Pascal Wafo, Philip F. Uzor,  
Zulfiqar Ali, Fidele Ntie-Kang, Ikhlas A. Khan and Peter Spiteller

|          |                                                                                                                  |
|----------|------------------------------------------------------------------------------------------------------------------|
| <b>9</b> | <b>Rational engineering of specialized metabolites in bacteria and fungi — 179</b>                               |
| 9.1      | Introduction — 179                                                                                               |
| 9.2      | Roles and definition of secondary metabolites — 180                                                              |
| 9.3      | Main classes of secondary metabolites and their biosynthetic pathway — 182                                       |
| 9.3.1    | Metabolites that originated from the shikimate pathway — 182                                                     |
| 9.3.2    | Altered peptides obtained through the ribosomal and post-translational pathway — 182                             |
| 9.3.3    | Polyketides — 183                                                                                                |
| 9.4      | Secondary metabolic engineering — 184                                                                            |
| 9.5      | Bioinformatic and computer predictions of biosynthetic gene clusters — 185                                       |
| 9.6      | Genomics-driven discovery of secondary metabolites — 187                                                         |
| 9.7      | Engineering approaches to produce novel natural products — 187                                                   |
| 9.7.1    | Pathway rebuilding in heterologous hosts — 187                                                                   |
| 9.7.2    | Genetic manipulation in the natural hosts — 189                                                                  |
| 9.7.3    | Activation of cryptic genes clusters in a native microbial host — 190                                            |
| 9.7.4    | Domain exchange in multi-modular enzymes — 190                                                                   |
| 9.8      | Plug and play approach to cloning transcriptionally silent gene clusters of unknown function from bacteria — 192 |
| 9.9      | Conclusions — 192                                                                                                |
|          | References — 194                                                                                                 |

Sergi Herve Akone, Cong-Dat Pham, Huiqin Chen, Antonius R. B. Ola,  
Fidele Ntie-Kang and Peter Proksch

|           |                                                                                                    |
|-----------|----------------------------------------------------------------------------------------------------|
| <b>10</b> | <b>Epigenetic modification, co-culture and genomic methods for natural product discovery — 197</b> |
| 10.1      | Introduction — 198                                                                                 |
| 10.2      | Production of natural compounds through epigenetic modification — 200                              |

|        |                                                                                      |     |
|--------|--------------------------------------------------------------------------------------|-----|
| 10.2.1 | Fungal cryptic metabolites induced by small molecules epigenetic modifiers —         | 200 |
| 10.2.2 | Bacterial cryptic secondary metabolites induced by epigenetic modifications —        | 201 |
| 10.3   | Induction of natural compounds by co-cultivation —                                   | 202 |
| 10.3.1 | Fungal-bacterial co-cultivation: A powerful tool for cryptic metabolites induction — | 202 |
| 10.3.2 | Fungal-fungal co-cultivation —                                                       | 210 |
| 10.4   | Genomic approaches —                                                                 | 212 |
| 10.5   | Conclusions —                                                                        | 214 |
|        | References —                                                                         | 214 |

## **Part IV: Case Studies — 219**

Conrad V. Simoben, Fidele Ntie-Kang, Dina Robaa and Wolfgang Sippl

|           |                                                                                              |            |
|-----------|----------------------------------------------------------------------------------------------|------------|
| <b>11</b> | <b>Case studies on computer-based identification of natural products as lead molecules —</b> | <b>221</b> |
| 11.1      | Introduction —                                                                               | 221        |
| 11.2      | Drugs approved by the FDA and the proportion of NPs —                                        | 222        |
| 11.3      | CADD of some NPs identified as lead compounds —                                              | 223        |
| 11.3.1    | Structure-based CADD methods —                                                               | 223        |
| 11.3.2    | Ligand-based CADD methods —                                                                  | 230        |
| 11.3.3    | Machine Learning —                                                                           | 233        |
| 11.4      | Conclusion —                                                                                 | 238        |
|           | References —                                                                                 | 240        |

Lucas Paul, Celestin N. Mudogo, Kelvin M. Mtei, Revocatus L. Machunda and Fidele Ntie-Kang

|           |                                                                                |            |
|-----------|--------------------------------------------------------------------------------|------------|
| <b>12</b> | <b>A computer-based approach for developing linamarase inhibitory agents —</b> | <b>247</b> |
| 12.1      | Introduction —                                                                 | 247        |
| 12.2      | Determination of the three-dimension linamarase structure prediction —         | 251        |
| 12.2.1    | Method without database information —                                          | 251        |
| 12.2.2    | Methods with database information —                                            | 251        |
| 12.2.3    | Fold recognition and threading methods —                                       | 252        |
| 12.2.4    | Comparative modeling method and sequence alignment strategies —                | 252        |
| 12.3      | Identification and validation of linamarase inhibitors —                       | 255        |

- 12.3.1 Virtual Screening for the identification of linamarase inhibitors — 255
- 12.3.2 Molecular docking — 256
- 12.4 Conclusion — 260
- References — 261

## Part V: Tutorials and Glossary — 269

Fidele Ntie-Kang, Kiran K. Telukunta, Linda Jahn and Jutta Ludwig-Müller

### 13 Searching natural product databases — 271

- 13.1 Introduction — 271
  - 13.1.1 Literature assignment — 272
  - 13.1.2 Plan of the day — 272
- 13.2 Session I (Compound databases) 1 hour — 272
  - 13.2.1 Objectives — 272
- 13.3 Session II (Secondary metabolite databases) 2 hours — 274
  - 13.3.1 Objectives — 274
- 13.4 Session III (Taxonomy databases) 2 hours — 282
  - 13.4.1 Objectives — 282
- 13.5 Session IV (Chemotaxonomy) 1.5 hours — 284
  - 13.5.1 Objectives — 284

Fidele Ntie-Kang, Linda Jahn and Jutta Ludwig-Müller

### 14 From genes to secondary metabolites — 287

- 14.1 Introduction — 287
  - 14.1.1 Literature assignment — 287
  - 14.1.2 Plan of the day — 288
- 14.2 Session I (Genome-based databases) 1.5 hours — 288
  - 14.2.1 Objectives — 288
  - 14.2.2 The medicinal plant genomics database — 288
- 14.3 Sessions II & III (Databases for gene cluster analysis) 3 hours — 289
  - 14.3.1 Objectives — 289
- 14.4 Session IV (Tools for gene cluster analysis) 2 hours — 293
  - 14.4.1 Objectives — 293
  - 14.4.2 Notes — 294

## **XIV — Contents**

Fidele Ntie-Kang, Abdurrahman Olğaç, Conrad V. Simoben, Abdulkarim Najjar, Sergi Herve Akone, Lucas Paul, Ramsay S. T. Kamdem, Abraham Madariaga-Mazón, Celestin N. Mudogo, Donatus B. Eni, Boris D. Bekono, Fatima Baldo, Dina Robaa, Daniel M. Shadrack, Ricardo B. Hernández-Alvarado, Karina Martinez-Mayorga, José L. Medina-Franco, Daniel Svozil and Wolfgang Sippl

**Glossary of terms used in chemoinformatics of natural products: advanced concepts and applications — 299**

**Index — 319**