

Contents

Chapter 1	Big Data in Predictive Toxicology: Challenges, Opportunities and Perspectives	1
	<i>Andrea-Nicole Richarz</i>	
1.1	Introduction	1
1.2	Big Data in the Area of Predictive Toxicology	2
1.3	The Big Vs of Predictive Toxicology Data	3
1.4	Challenges of Big Data in Predictive Toxicology	6
1.4.1	Need for an Adequate Infrastructure	7
1.4.2	Standardisation and Data Curation	9
1.4.3	Too Big? Identifying the Relevant Data	11
1.4.4	Data Integration Infrastructures	11
1.4.5	Making Sense of the Data	14
1.5	Opportunities Provided by Big Data for Predictive Toxicology	15
1.5.1	More is More – Benefits of a Broader and More Diverse Data Basis	16
1.5.2	Big Data for the Big Picture	18
1.5.3	Creating and Using New Knowledge: Applications for Hot Topics in (Predictive) Toxicology	19
1.6	Conclusions and Perspectives	25
	References	27

Chapter 2	Biological Data in the Light of Toxicological Risk Assessment	38
	<i>Vessela Vitcheva</i>	
2.1	Introduction	38
2.2	Data Generated by <i>In Vivo</i> Toxicity Testing	39
2.2.1	Acute Toxicity Testing	40
2.2.2	Toxicity Testing after Repeated Dose Administration (Short-term, Sub-chronic and Chronic Toxicity Tests)	42
2.2.3	Toxicity Tests for Specific Endpoints	44
2.3	Data Generated by <i>In Vitro</i> Studies	48
2.3.1	<i>In Vitro</i> Test Methods for Genotoxicity/ Mutagenicity Testing	48
2.3.2	<i>In Vitro</i> Test Methods for Carcinogenicity	51
2.3.3	<i>In Vitro</i> and <i>In Chemico</i> Test Methods for Skin Sensitisation	51
2.4	Mechanistic Understanding – Adverse Outcome Pathways	52
2.5	Data from New Approach Methodologies	53
2.6	Toxicokinetic Data	54
2.7	Sources of Toxicological Data	55
2.7.1	TOXicology Data NETwork (TOXNET)	55
2.7.2	Registry of Toxic Effects of Chemical Substances (RTECS)	56
2.7.3	Kyoto Encyclopedia of Genes and Genomes (KEGG)	56
2.7.4	Aggregated Computational Toxicology Resource (ACToR)	57
2.7.5	European Chemicals Agency (ECHA) Data	57
2.7.6	COSMOS Next Generation	57
2.7.7	Data from Assessment Reports	58
2.8	Quality of Data	59
2.9	Data Interpretation	61
2.10	Use of Biological Data for Predictive Toxicology	62
	References	63
Chapter 3	Chemoinformatics Representation of Chemical Structures – A Milestone for Successful Big Data Modelling in Predictive Toxicology	69
	<i>Nikolay Kochev, Nina Jeliazkova and Ivanka Tsakovska</i>	
3.1	Introduction to Chemoinformatics Approaches for Chemical Structure Representation	69
3.2	Structure Representation Characteristics	74

3.3	Constitutional Representations	77
3.4	Topological Representations	78
3.5	Linear Notations	81
3.5.1	SMILES Linear Notation	82
3.5.2	InChI Linear Notation	84
3.5.3	SMARTS Linear Notation	84
3.6	3D and 4D Structure Representations	85
3.6.1	3D Structure Optimisation	87
3.6.2	Conformational Search	89
3.7	Molecular Descriptors, Fingerprints and Hash Codes	91
3.8	Challenging Cases for Structure Representation	95
3.8.1	Handling Aromaticity	97
3.8.2	Unique Structure Representations	98
3.8.3	Tautomerism	99
3.8.4	Miscellaneous Challenges	100
3.9	Chemical Substances	101
3.10	Conclusions	104
	References	104
Chapter 4	Organisation of Toxicological Data in Databases	108
	<i>David Bower, Kevin Cross and Glenn Myatt</i>	
4.1	Introduction	108
4.2	Standards for Exchange and Organising Toxicology Data	111
4.2.1	SD Files	111
4.2.2	ToxML (DB)	113
4.2.3	SEND	120
4.2.4	OECD Harmonised Templates	122
4.2.5	ISA-Tab	123
4.3	Databases and Data Sets of Toxicity Studies	125
4.4	Process Management and Regulatory Compliance Databases	133
4.5	Alternative Approaches Databases	135
4.5.1	ToxBank	135
4.5.2	ToxCast/DSSTox/Tox21	137
4.5.3	PubChem	138
4.5.4	NIEHS/CEBS	138
4.6	Integrated Approaches	138
4.7	Conclusions	144
	Appendix	145
	References	160

Chapter 5	Making Big Data Available: Integrating Technologies for Toxicology Applications	166
	<i>Nina Jeliaskova and Vedrin Jeliaskov</i>	
5.1	Introduction	166
5.2	Data Integration Approaches	168
5.3	Ontologies and Shared Vocabulary	171
5.4	Web Services and Cloud Technology	172
5.5	Standards	172
5.6	Integrative Data Analysis	174
5.7	Data Sharing Infrastructure	174
5.8	Conclusions	175
	References	176
Chapter 6	Storing and Using Qualitative and Quantitative Structure–Activity Relationships in the Era of Toxicological and Chemical Data Expansion	185
	<i>Sulev Sild, Geven Piir, Daniel Neagu and Uko Maran</i>	
6.1	Introduction	185
6.2	Anatomy of Predictive Models	190
6.2.1	Mathematical Representation	190
6.2.2	Data Set Representation	191
6.2.3	Model (and Data) Provenance and Metadata	192
6.3	Model Governance	193
6.4	Data and File Formats for the Representation of Predictive Models	195
6.4.1	QSAR Model Reporting Format	196
6.4.2	QSAR-ML	198
6.4.3	The Predictive Model Markup Language	199
6.4.4	QsarDB Archive Format	201
6.5	Applications and Solutions for the Storage of Predictive Models	204
6.5.1	Repository Approaches	205
6.5.2	Integrated Modelling Environment Approaches	206
6.6	Conclusions	208
	References	209
Chapter 7	Toxicogenomics and Toxicoinformatics: Supporting Systems Biology in the Big Data Era	214
	<i>Terezinha M. Souza, Jos C. S. Kleinjans and Danyel G. J. Jennen</i>	
7.1	Introduction to Toxicogenomics and Toxicoinformatics	214
7.2	Current ‘Omics Technologies	216

<i>Contents</i>	xvii
7.2.1 Transcriptomics	216
7.2.2 (Epi)Genomics	217
7.2.3 Proteomics	218
7.2.4 Metabolomics	219
7.3 Data Handling	220
7.3.1 Toxicoinformatics: Standardisation, Storage and Availability of 'Omics Data	220
7.3.2 From Bits to Annotation	223
7.4 Use of 'Omics Data in Predictive Toxicology	226
7.4.1 Comparative Assessments	226
7.4.2 Unsupervised and Supervised Pattern Recognition	226
7.4.3 Connectivity Mapping (CMap)	227
7.4.4 Mechanistic Analysis	228
7.5 Big ('Omics) Data	229
7.6 Toxicogenomics in the Big Data Era: Challenges, Perspectives and Opportunities	232
References	234
Chapter 8 Profiling the Tox21 Chemical Library for Environmental Hazards: Applications in Prioritisation, Predictive Modelling, and Mechanism of Toxicity Characterisation	242
<i>S. Sakamuru, H. Zhu, M. Xia, A. Simeonov and R. Huang</i>	
8.1 Introduction	242
8.2 The Tox21 Process	243
8.2.1 Tox21 Compound Collection	243
8.2.2 Tox21 Robotic Platform	247
8.2.3 Tox21 Screening Process	247
8.2.4 Tox21 Data Analysis Process Including Data Quality Control	248
8.3 Using the Tox21 Big Data Collection for Predictive Modelling	251
8.3.1 Profiling for Compound Mechanism of Toxicity	251
8.3.2 Modelling Tox21 Data for <i>In Vivo</i> Toxicity Prediction	252
8.4 The Tox21 Data Challenge – New Methods for Data Modelling in the Big Data Era	257
8.5 Conclusion	259
Acknowledgements	259
References	260

Chapter 9	Big Data Integration and Inference	264
	<i>Karen H. Watanabe-Sailor, Hristo Aladjov, Shannon M. Bell, Lyle Burgoon, Wan-Yun Cheng, Rory Conolly, Stephen W. Edwards, Nàtalia Garcia-Reyero, Michael L. Mayo, Anthony Schroeder, Clemens Wittwehr and Edward J. Perkins</i>	
9.1	Introduction: New Toxicology Paradigm and Challenges of Big Data Integration and Application	265
9.2	Structuring Knowledge to Support Data Mining, Modelling, and Decision-making	268
9.2.1	Development of a Globally Accessible Knowledge Base for Data Integration	268
9.2.2	Use of Structured Knowledge	269
9.2.3	Supporting Decision-making	273
9.3	Integration of High-throughput Screening Assays and Curated Databases for Understanding Chemical–Bioactivity Relationships	274
9.3.1	Data Generated from HTS Assays	274
9.3.2	Publicly Available Databases	275
9.3.3	Current Approaches for Integrating HTS Data and Databases and Computational Modelling to Predict <i>In Vivo</i> Toxicity From <i>In Vitro</i> Data	278
9.4	Systems Biology: <i>De Novo</i> Network Inference of Chemical, Gene, Protein, and Metabolite Relationships Underlying Toxicity	279
9.4.1	Network Inference	280
9.4.2	Steady-state Networks	280
9.4.3	Dynamic Networks	281
9.4.4	Networks and Inference of Toxicological Effects	282
9.5	Computationally Predicted and Putative AOP Generation	283
9.5.1	Computational AOP Generation	284
9.5.2	From Computationally Predicted AOPs to Putative AOPs	287
9.6	Use of AOPs to Support Testing and Risk Assessment	287
9.6.1	Identification of Sufficient Key Events in AOP Networks	287
9.6.2	Case Study: Steatosis	289
9.7	Translating Experimental Data into Model Structure: Integration of High-throughput Datasets or Expert Elicitation?	290

<i>Contents</i>		xix
	9.7.1 Mechanistic Modelling Methods	291
	9.7.2 Case Study: Aromatase Inhibition Model Development	292
	9.8 Conclusions	295
	9.9 Future Directions	296
	References	297
Chapter 10	Chemometrical Analysis of Proteomics Data	307
	<i>Marjan Vračko</i>	
	10.1 Introduction	307
	10.2 Chemometrical Methods for Data Analysis	310
	10.2.1 Measure of Similarity	310
	10.2.2 Clustering and Classification	311
	10.2.3 Genetic Algorithm	313
	10.2.4 Chemical Descriptors and Bio-descriptors	313
	10.3 Application Examples of Chemometrical Analyses of Proteomics Data	315
	10.3.1 Clustering and Classification of Proteomics Data	315
	10.3.2 Genetic Algorithm Applied to Proteomics Data	317
	10.3.3 Proteomics as Bio-descriptors	320
	10.3.4 Biomarkers	324
	10.4 Conclusions	326
	References	326
Chapter 11	Big Data and Biokinetics	331
	<i>Miyoung Yoon, Gina Song, Harvey Clewell and Bas Blaauboer</i>	
	11.1 Introduction	331
	11.2 Simple Biokinetic Modelling Approaches for Supporting the Use for Large Data Amounts	333
	11.2.1 Biokinetic Approach for Supporting High- throughput Testing	333
	11.2.2 Biokinetic Approach for Predicting Acute Toxicity	336
	11.2.3 Biokinetics Supporting Chemical Grouping	339
	11.2.4 Minimalist Biokinetics Approach	340
	11.3 Big Data Facilitating the Parameterisation of Biokinetic Modelling to Support the Use of Big Data for Risk Assessment	342
	11.3.1 Example of an Approach for Rapid PBPK Modelling: PLETHEM	343
	11.3.2 Building <i>In silico</i> Humans	344
	11.3.3 Obtaining Chemical Specific Parameters	346

11.4 Conclusion	348
List of Abbreviations	348
Appendix: Examples of Life Stage Physiological Parameters for Human Populations	349
Height	349
Weight	350
Linking the Height and Weight Growth Curves	351
Body Fat Volume	352
References	353
Chapter 12 Role of Toxicological Big Data to Support Read-across for the Assessment of Chemicals	359
<i>Mark T. D. Cronin and Andrea-Nicole Richarz</i>	
12.1 Introduction	359
12.2 Role of Toxicological Big Data to Inform and Support Read-across	364
12.2.1 Supporting a Read-across Justification	364
12.2.2 Investigation of Toxicological Mechanisms of Action	365
12.2.3 Biological Read-across	366
12.2.4 The Use of Omics Technologies for Read-across	368
12.2.5 Making Sense of Big Data Compilations to Form the Bridge With Chemical Structure	370
12.2.6 Read-across for Nanomaterials	372
12.3 The Future – Needs for the Use of Big Data for Read-across	373
12.4 Conclusions	374
References	375
Subject Index	385