

## Contents

<b>1</b>	<b>Crystallography Open Database: History, Development, and Perspectives</b>	<b>1</b>
	<i>Saulius Gražulis, Andrius Merkys, Antanas Vaitkus, Daniel Chateigner, Luca Lutterotti, Peter Moeck, Miguel Quiros, Robert T. Downs, Werner Kaminsky, and Armel Le Bail</i>	
1.1	Introduction	1
1.2	Open Databases for Science	3
1.3	Building COD	6
1.3.1	Scope and Contents	7
1.3.2	Data Sources	7
1.3.3	Data Maintenance	8
1.3.3.1	Version Control	11
1.3.3.2	Data Curation Policies	12
1.3.3.3	Quarterly Releases	13
1.3.4	Sister Databases (PCOD, TCOD)	14
1.4	Use of COD	14
1.4.1	Data Search and Retrieval	14
1.4.1.1	Data Identification	15
1.4.1.2	Web Search Interface	15
1.4.1.3	RESTful Interfaces	15
1.4.1.4	Output Formats	17
1.4.1.5	Accessing COD Records	17
1.4.1.6	MySQL Interface	18
1.4.1.7	Alternative Implementations of COD Search on the Web	20
1.4.1.8	Installing a Local Copy of the COD	21
1.4.1.9	File System-Based Queries	23
1.4.1.10	Programmatic Use of COD CIFs	24
1.4.2	Data Deposition	26
1.5	Applications	27
1.5.1	Material Identification	27
1.5.2	Applications for the Mining Industry	27
1.5.3	Extracting Chemical Information	28
1.5.4	Property Search	30
1.5.5	Geometry Statistics	30

1.5.6	High-Throughput Computations	31
1.5.7	Applications in College Education and Complementing Outreach Activities	31
1.6	Perspectives	32
1.6.1	Historic Structures	32
1.6.2	Theoretical Data in (T)COD	32
1.6.3	Conclusion	32
	Acknowledgments	33
	References	33
<b>2</b>	<b>The Inorganic Crystal Structure Database (ICSD): A Tool for Materials Sciences</b>	<b>41</b>
	<i>Stephan Rühl</i>	
2.1	Introduction	41
2.2	Content of ICSD	42
2.3	Interfaces	46
2.4	Applications of ICSD	46
2.4.1	Prediction of Ferroelectricity	47
2.4.2	Using the Concept of Structure Types	47
2.4.3	Two Examples of Training Machine Learning Algorithms with ICSD Data	48
2.4.4	High-Throughput Calculation	50
2.5	Outlook	51
	References	51
<b>3</b>	<b>Pauling File: Toward a Holistic View</b>	<b>55</b>
	<i>Pierre Villars, Karin Cenzual, Roman Gladyshevskii, and Shuichi Iwata</i>	
3.1	Introduction	55
3.1.1	Creation and Development of the PAULING FILE	57
3.2	PAULING FILE: Crystal Structures	57
3.2.1	Data Selection	58
3.2.2	Categories of Crystal Structure Entries	58
3.2.3	Database Fields	59
3.2.4	Structure Prototypes	62
3.2.5	Standardized Crystallographic Data	63
3.2.5.1	Checking of Symmetry	63
3.2.5.2	Standardization	65
3.2.5.3	Comparison with the Type-Defining Data Set	67
3.2.6	Assigned Atom Coordinates	67
3.2.7	Atomic Environment Types (AETs)	68
3.2.8	Cell Parameters from Plots	72
3.3	PAULING FILE: Phase Diagrams	72
3.4	PAULING FILE: Physical Properties	75
3.4.1	Data Selection	75
3.4.2	Database Fields	76
3.4.3	Physical Properties Considered in the PAULING FILE	76
3.5	Data Quality	80

3.5.1	Computer-Aided Checking	80
3.6	Distinct Phases	81
3.6.1	Chemical Formulas and Phase Names	83
3.6.2	Phase Classifications	84
3.7	Toward a Megadatabase	84
3.8	Applications	89
3.8.1	Products Containing PAULING FILE Data	89
3.8.2	Holistic Overviews Based on the PAULING FILE	91
3.8.3	Principles Defining Ordering of Chemical Elements	92
3.9	Lessons to Learn from Experience	99
3.10	Conclusion	103
	References	104
<b>4</b>	<b>From Topological Descriptors to Expert Systems: A Route to Predictable Materials</b>	<b>107</b>
	<i>Alexander P. Shevchenko, Eugeny V. Alexandrov, Olga A. Blatova, Denis E. Yablokov, and Vladislav A. Blatov</i>	
4.1	Introduction	107
4.2	Topological Tools for Developing Knowledge Databases	108
4.2.1	Why Topological?	108
4.2.2	Topological vs. Other Descriptors of Crystal Structures	110
4.2.3	Topological vs. Crystallographic Databases	111
4.2.4	Deriving Topological Knowledge from Crystallographic Data	116
4.2.4.1	Algorithms for Topological Analysis	116
4.2.4.2	Building Distributions of Descriptors	118
4.2.4.3	Finding Correlations Between Descriptors	123
4.2.5	Universal Data Storage	126
4.3	Applications of Topological Tools in Crystal Chemistry and Materials Science	131
4.3.1	Network Topology Prediction	131
4.3.2	Prediction of Properties	137
4.4	Conclusions	137
	References	138
<b>5</b>	<b>A High-Throughput Computational Study Driven by the AiiDA Materials Informatics Framework and the PAULING FILE as Reference Database</b>	<b>149</b>
	<i>Martin Uhrin, Giovanni Pizzi, Nicolas Mounet, Nicola Marzari, and Pierre Villars</i>	
5.1	Introduction	149
5.1.1	Three Key Developments Opened Up Unprecedented Opportunities	150
5.1.2	Relative Few Inorganic Solids Have Been Experimentally Investigated	151
5.2	Nature Defines Cornerstones Providing a Marvelously Rich but Still Very Rigid Systematic Framework of Restraint Conditions	151
5.3	The First, Second, and Third Paradigms	153

- 5.4 The Realization of the Fourth and Fifth Paradigms Requires Three Preconditions 153
  - 5.4.1 Introduction of the *Prototype Classification* to Link Crystallographic Databases Created by Different Groups 153
  - 5.4.2 Introduction of the *Distinct Phases Concept* to Link Different Kinds of Inorganic Solids Data 154
  - 5.4.3 The Existence of a Comprehensive, Critically Evaluated Inorganic Solids Database Concept (DBMS) of Experimentally Determined Single-Phase Inorganic Solids Data to Be Used as Reference 154
- 5.5 The Core Idea of the Fifth Paradigm 154
- 5.6 Restraint Conditions Revealed by “Inorganic Solids Overview–Governing Factor Spaces (Maps)” Discovered by Data-Mining Techniques 156
  - 5.6.1 Compound Formation Maps 157
  - 5.6.2 Atomic Environment Type Stability Maps for AB Inorganic Solids 158
  - 5.6.3 Twelve Principles in Materials Science Supporting Three Cornerstones Given by Nature 159
- 5.7 Quantum Simulation Strategy 161
- 5.8 Workflows Engine in AiiDA to Carry Out High-Throughput Calculation for the Creation of the Materials Cloud, Binaries Edition 164
  - 5.8.1 AiiDA 164
  - 5.8.2 SSSP (Standard Solid State Pseudopotentials) Library 165
  - 5.8.3 Workflows 166
  - 5.8.4 Workfunctions 166
  - 5.8.5 Workchains 166
  - 5.8.6 Workflows Used in This Project 168
- 5.9 Conclusions 169
  - Acknowledgment 169
  - References 169
- 6 Modeling Materials Quantum Properties with Machine Learning 171**  
*Felix A. Faber and O. Anatole von Lilienfeld*
  - 6.1 Introduction 171
  - 6.2 Kernel Ridge Regression 171
  - 6.3 Model Assessment 173
    - 6.3.1 Learning Curve 173
    - 6.3.2 Speedup 174
  - 6.4 Representations 176
  - 6.5 Recent Developments 177
    - References 178
- 7 Automated Computation of Materials Properties 181**  
*Cormac Toher, Corey Oses, and Stefano Curtarolo*
  - 7.1 Introduction 181

7.2	Automated Computational Materials Design Frameworks	182
7.2.1	Generating and Using Databases for Materials Discovery	182
7.2.2	Standardized Protocols for Automated Data Generation	185
7.3	Integrated Calculation of Materials Properties	187
7.3.1	Autonomous Symmetry Analysis	189
7.3.2	Elastic Constants	191
7.3.3	Quasi-harmonic Debye–Grüneisen Model	193
7.3.4	Harmonic Phonons	195
7.3.5	Quasi-harmonic Phonons	197
7.3.6	Anharmonic Phonons	198
7.4	Online Data Repositories	198
7.4.1	Computational Materials Data Web Portals	198
7.4.2	Programmatically Accessible Online Repositories of Computed Materials Properties	200
7.5	Materials Applications	202
7.5.1	Disordered Materials	202
7.5.1.1	High Entropy Materials	203
7.5.1.2	Metallic Glasses	203
7.5.1.3	Modeling Off-Stoichiometry Materials	204
7.5.2	Superalloys	205
7.5.3	Thermoelectrics	205
7.5.4	Magnetic Materials	208
7.6	Conclusion	209
	Acknowledgments	209
	References	209
<b>8</b>	<b>Cognitive Chemistry: The Marriage of Machine Learning and Chemistry to Accelerate Materials Discovery</b>	<b>223</b>
	<i>Edward O. Pyzer-Knapp</i>	
8.1	Introduction	223
8.2	Describing Molecules for Machine Learning Algorithms	224
8.3	Building Fast and Accurate Models with Machine Learning	234
8.3.1	Squared Exponential Kernel	239
8.3.2	Rational Quadratic Kernel	240
8.4	Searching Through Chemical Libraries	244
8.5	Conclusion	248
	References	249
<b>9</b>	<b>Machine Learning Interatomic Potentials for Global Optimization and Molecular Dynamics Simulation</b>	<b>253</b>
	<i>Ivan A. Kruglov, Pavel E. Dolgirev, Artem R. Oganov, Arslan B. Mazitov, Sergey N. Pozdnyakov, Efim A. Mazhnik, and Alexey V. Yanilkin</i>	
9.1	Introduction	253
9.2	Machine Learning Potential for Global Optimization	258
9.2.1	Lattice Sums Method	258
9.2.2	Feature Vector	261

9.2.3	Feature Vector Analysis	262
9.2.4	Examples of Machine Learning Interatomic Potentials	265
9.2.4.1	Aluminum	265
9.2.4.2	Carbon	267
9.2.4.3	Helium and Xenon	271
9.2.5	Discussion	272
9.3	Interatomic Potential for Molecular Dynamics	273
9.3.1	General Form of the Potential	273
9.3.2	Parameters Selection	274
9.3.3	Thermodynamic Quantities and Phase Transitions	277
9.3.4	Interatomic Potential for System of Two (or More) Atomic Types	281
9.4	Statistical Approach for Constructing ML Potentials	284
9.4.1	Two-Body Potential	284
9.4.2	Three-Body Potential	286
	Acknowledgements	286
	References	286
	<b>Index</b>	<b>289</b>