

Index

- ABO₃ perovskites, orthorhombic 77
- Adrenaline 119
- Agglomerative clustering 17
- α - ω -Alkandihalogenides 122
- Alkanes, melting points 122
- AMBER 91
- Amino acid sequence data 137
- Anion parameter 63
- Anomaly detection 61
- Association analysis 61
- Association rule mining 11
- Atom-atom potentials 103
 - tailor-made force field 92
- Atomic coordinates 43
- 1-Aza-2-cyclobutanone (FEPNAP) 127, 128, 130

- Basic local alignment search tool (BLAST) 143
- Bayesian statistics 136
- Bias vs. variance (fit vs. generalizability) 32
- Biological macromolecular crystallization database (BMCD) 141
- Black tar mystery, ICSD search 52
- Boltzman, L. 10
- Buckingham potentials 126
- BUWDUS 98
- BUWFEE 98
- BUWGEF 98
- BUWGEF01 98
- BUWKEJ 95, 110

- C-C interactions 114
- Caffeine 119
- Cambridge Structural Database (CSD) 102
- Carbon/sulfur interactions, van der Waals parameters 118

- Carbon monoxide 119
- Carboxyhemoglobin 119
- Chemical formulas 42
- Chemical names, IUPAC 42
- Clopidogrel 119
- Classification 22
- Cluster analysis 60, 92, 93
- Clustered crystal structures 97
- Clustering 15
 - algorithms 17, 93
 - hierarchical 17, 93
- COMPACT 94
- COMPASS 91
- Complete linkage clustering 18
- Confidence 12
- Connolly surface 124
- Consensus database mining (CDM) 135, 136, 155, 160
- Correlation, linear 4
- Correlation coefficients 4
- Critical assessment of structure prediction (CASP) 139
- CRYCOM 94
- Crystal structures 59
 - clustering 96
 - databases, overview 39
 - determination 98
 - faulty 109
 - generation 123
 - prediction (CSP) 89, 90, 123
 - ranking 126
 - similarities 89
- CrystalEye 102
- Crystallization data mining 136
 - proteins 140
- Crystallization space 141
- Crystallographic databases 37

- Crystallographic descriptors, geometrical expression 65
- Crystallographic Information File (CIF) 38
- Data classification 64
- Data “cloud” 13
- Data clustering 69
 - topology preserving distance measures 71
- Data collection 37
- Data evaluation 37
- Data fields 40
- Data mining 1, 59, 89, 92, 136
 - crystal structure prediction 92
 - methods in biology 142
- Data mining force field (DMFF), derivation 101
 - application 118
 - descriptors 102
 - parametrization 106
 - validation 111
- Data prediction, structure–property relationships 73
- Data preprocessing 62
- Data systematics 62
- Data training 75
- Database design 37, 49
- Database functionality 37
- Database management systems (DBMS) 3
- DDL (data drive lattice) 70
- Decision trees 28
- Density estimation 4, 12
- Descriptor refinement, data mining 78
- Descriptors 59
- DFT 126
- Diagram comparison 89
- Dihalogenalkanes, melting points/lattice energies 122
- Dilation of anion position 63
- Dimensionality reduction 59
 - techniques 70
- Distance function 15
- DMAREL 91
- DMSP 76
- Docking 119
- Dreiding force field 122
- Drug and enzyme 120
- Drug development 119
- DSSP algorithm 139
- ECEPP 91
- Eigenvalues 31, 67
- Eigenvectors 14, 31, 67
- EMBL algorithm 139
- Entropy 11, 91
- Enumeration schemes 62
- Euclidean distance 16
- FABFOE 112
- Fatty acids 121
- Faulty crystal structures 109
- Feature space 27
- Feed forward neural network 26
- Fisher’s linear discriminant 23
- FlexCryst 125
- Flexible molecules 92
- Force field 89, 91
 - derivation 101
- Fragment database mining (FDM) 136, 149, 155
- Genetic algorithms (GA) 75, 125
- Genome sequencing 137
- Geodesic distance
- Geometrical modeling equations 63
- Goldschmidt diagrams 64
- GOR V 148
 - secondary structure prediction 160
- Helix predictions 138, 158
- Hemoglobin 119
- Hierarchical clustering 17, 93
- Homoatomic pair potentials, validation 114
- Hydrogen bonds 103, 115, 124
- Ice I_c 53
- Independent component analysis (ICA) 13, 15
- Information gain 11
- Inorganic Crystal Structure Database (ICSD) 37, 102
 - access 49
 - Web version, modules/search strategies 49, 50
- Inorganic crystallography, enumeration schemes 62
 - knowledge discovery 60
- Intermolecular interactions 100
- IsoMap technique 70
- K-means 19
- Kernel PCA 15
- Knowledge discovery in databases (KDD) 2, 61
- Latent variables 65
- Lattice constants 77
- Lattice energies 90
- Lattice parameters 63
- LDA (linear discriminant analysis) 23
- Least squares optimization 30

- Lennard-Jones potentials 126
- Ligand–protein docking 120, 123
- Linear correlation 4
- Linkage 94
- Liquids, densities 92
- Locally linear embedding (LLE) 70, 71

- Manhattan distance 16
- Many Body Interaction Types (MBITs) 75
- Melting points 121
- Mineral names 42
- Minimization 125
- Molecular dynamics (MD), force fields 91
 - simulations, formation/breaking of bonds 92
- Molecular mechanics (MM), force fields 91
- Mooser-Pearson plots 64
- Multiple sequence alignments (MSA) 146

- Naive Bayes 13
- Neural networks 26
- Nicotine 119

- Outliers, overestimation/underestimation 5
- Overfitting/overtraining 32, 105
- Oxidation numbers 43

- PCA (principal component analysis) 13, 65, 66, 93
 - nonlinear 13
 - procedural logic 68
 - semiconductors 70
- PDL (pre defined lattice) 70
- Perceptron algorithm 24, 25
- Perovskites, orthorhombic 77
- Pettifor plots 64
- Phase designations 42
- Philips and van Vechten diagrams 64
- Pigment Orange 5 (1-((2,4-dinitrophenyl)azo)-2-naphthol, refcode CICCUN) 99
- Pigment Red 3 (4-methyl-2-triphenylarsine-2-naphthol, refcode MNIPZN) 99
- Pigment Red 170 100
- Pigment Red 181 (PR181) 99
- Polymorph Predictor 94
- Polymorphism 90, 91
- Polymorphs 90, 107, 113, 123, 125
 - clustering 129
 - predicted 127
- Powder 89
 - patterns/diagrams 39, 50, 52, 93
 - indexed 99
 - unindexed 98, 131
- Powder X-ray data 98
- Prediction, quality 143

- Predictive modeling 60
- Progesterone 119
- Protein Data Bank (PDB) 101, 137
- Proteins 92
 - coil 138
 - crystallization 136, 142
 - folding 140
 - secondary structure prediction 136, 145
 - structure prediction 136
 - X-ray structure determination,
 - crystallization data mining 140
- PSI-BLAST 146

- QSAR/QSPR 74, 75

- Radial distribution functions 92
- Rank correlation 5
- Regression 19, 30
 - nonlinear 31
- Reliability index 44
- Retrieval 50
- Root mean squared deviation (RMSD) 17

- Search strategies 37
- Secondary descriptors 78
- Secondary structure prediction 136, 160
- Segment overlap coefficient (SOV) 145
- Self organization maps 70
- Self-assembly (self-recognition) 90
- Shannon, C. 10
- β -Sheet predictions 138, 158
- Similarity functions 15
- Similarity index 95, 99
- Simplex method 125
- Single linkage clustering 17
- Singular value decomposition see also PCA 66
- Space group preferences 11
- Space groups 109
 - CSD 124
 - Hermann–Mauguin symbol 43
- Special name records 42
- Spinel nitrides 64, 67
- Statistical dependence 8
- Structural fragments 136
- Structure generation 126
- Structure maps 62, 69
- Structure prediction, data mining aided 75
- Structure–chemistry relationships 64
- Sublimation energies 111, 113
- Supervised methods 19
- Support vector machines (SVM) 15, 24, 27, 92, 100, 146

- Support vector regression (SVR) methods 77
- “Swiss roll” template 69
- Symmetry records 43

- Tanimoto coefficient 16
- Temperature factors 44
- Text mining 12
- Thermal parameters (atomic displacement factors) 43
- Tiling theory 66

- Unit cell dimensions 43
- Unsupervised methods 4

- Valency effects vs. atomic size effects 63
- Validation 101
- Validity 48
- Valium 119
- Van der Waals parameters 118
- Virtual databases 102

- Water, ICSD search 53
- Wyckoff sequence 44

- X-ray structure determination, proteins 140

- ZnS cluster geometries 77