Index

a
ab intio density calculations 43
ACD/ChemSketch 19
activity cliffs 56
activity landscapes 312
adjacency matrix 112, 120, 130
– spectral moments (SMs) 125
adjusted rand index (ARI) 35
ADMET modeling 42
ADRIANA.Code 20
algebraic graph theory 176–180
alkane molecules 176
– chemical graph for isomers 176
– Wiener index of 179
Angstrom-scale structure 374
antihistamine Fexofenadine
– 2D drawn representation 43
area under the curve (AUC) 35
aromaticity index 159
artificial neural network (ANN) 5, 48, 67
– simplified view 48
– three layers 6
– types 6
atoms in molecules, model of 258, 259
– Li–allyl complex 258
– useful to investigate 259
atoms-in-molecules-based approach 42
average atom connectivity index 93
average connectivity index (CHIA) 136
average vertex sum indices 115
Avogadro’s number 188

b
backpropagation method 7
Balaban distance connectivity index 92
Balaban index 113, 114, 116, 135, 136, 141
balanced error rate (BER) 35
Barysz matrix 122, 126
Barysz weighting scheme 121
Bayes decision rule 15
Bayesian approach, to virtual screening 231–235
– descriptor distributions 231, 232
– estimation of probability densities 234
– Gaussian distributions 232
– m-estimate correction 232
– predicting performance 235, 236
BCUTdescriptors 152
Beilstein database 394
BEIm2 descriptor 168, 169
best practice principles 34
beta-lactamase 67
binned ParaSurf descriptors 300
binormalized quadratic index 94
biochemical systems theoretic (BST) framework 328
biodecriptors 39, 40
biological assays
– spectrum 39
biological screening 307
BMLR method 102
– two-parameter regression equations 103
bond energy 368
Born–Oppenheimer approximation 250
bridge chemical structure 69
Brownian computation (processing). See electronic messages; molecular messages
Burden matrix 88, 121, 125, 126, 131, 134, 168

c
canonical measure of correlation 123, 127
canonical measure of distance 123, 127
– multidimensional scaling calculation 131
CAOS. See computer-assisted organic synthesis (CAOS)
carbon nanotube, SEM image 40
CASREACT database 394
characteristic root index 94
charge density 368
ChemDraw 19
chemical descriptors 33, 34
chemical information analysis
  – from graph-theoretical matrices 124–143
chemical reaction network 329–331
  – differential evolution for searching space 337
    -- basic DE optimization method 338, 339
    -- self-adaptive DE with integer variables 339, 340
chemical reactions
  – automated creation of rules
    -- by learning and reaction database mining 404
  – automatically derived reaction rules 404–406
  – chemical reaction graph 399, 400
  – empirical reaction rules 404
  – formal-logical approach for 397–399
  – functional group transformations 406
  – graph-theoretic reaction rules 397
  – graph transformation rules, for generation of 396, 397
  – substructure-based transformations 406–409
  – Ugi and Dugundji formal theory 400, 401, 403, 404
chemical space 122, 313
  – generation using fingerprint 313, 314
  – influence of molecular representation in 314
  – similarity/diversity 124
  – TPIMs compounds and 314
chemical transformations, retrosynthetic generation of 410, 411
  – complexity-based disconnective strategies 412, 413
  – forward generation 414, 416
  – guiding patterns, recognition of 411, 412
  – isomorphic substructures 411, 412
  – molecular symmetry 411, 412
  – strategic bond tree for disconnections 413, 414
Chemistry Development Kit (CDK) 21
Chi matrix 120, 127, 130
clustering of chemical databases 211
  – classical hierarchical clustering applications 214, 215
  – classical partitional clustering applications 212–214
  – nonclassical and mixed clustering methods 215–217
  – pattern representation, of chemicals structures 211, 212
CODESSA software 24, 102, 104
collinear descriptors 4, 47, 53, 71
combinatorial complexity problem 409, 410
  – chemical transformations
    -- forward generation of 414–419 (See also computer-assisted reaction prediction)
    -- retrosynthetic generation of 410–414
comparative molecular field analysis (CoMFA) 5, 23, 41, 66, 218, 299
  – descriptors, alignment issues 41, 42
compound recall
  – alternative approaches, to prediction of 238–240
  – for Bayesian screening, using continuous numerical descriptors 238
    -- linear regression models, from MDDR database 238, 239
    -- MACCS fingerprint 238
  – practical prediction of 236–238 compounds
    – narcotic effect 66
    – properties 65
comprehensive descriptors for structural and statistical analysis (CODESSA) 24, 72, 102, 103
  – descriptors 102, 103
    – regression model in 103, 104
computer-assisted organic synthesis (CAOS) 393–396
computer-assisted reaction prediction 414, 416
  – disconnective strategies to taxol and taxane ring core, application of 415
    – formal-logical approach to search space of 418, 419
    -- Bertz approach, evaluation of 419
    -- combinatorial learning methods 419
    -- FORWARD program 418
    -- Wender’s principle 419
    -- quantitative models for reactivity prediction 416–418
    -- Beppe program to evaluate 417
    -- measuring regioselectivities 417
    -- ROBIA approach 417
    -- selection of solvent 417
    -- software packages 417
    -- use of reactivity descriptors and 418
computer messages 368, 369, 374, 389.
  See also electronic messages
Concord software 20
Index

consensus analysis (CANCON) 67
continuous stirred tank reactors (CSTRs) 327
CORINA software 19, 20
Coulomb interaction 294
cross-correlation matrix 123
cross-validation methods 4
cyclicity index 93
cytochrome P450 (CYP) 52

D
Database of pyridine compounds 276
data sets 124, 125
– characteristics 124
2D autocorrelation descriptors 93
decision boundary 16
decision forest (DF) 67, 100
– algorithm, flowchart 101
decision tree (DT) 12–14, 49, 50
– drawbacks 13
– for molecular classification 50
– types of nodes 12, 13
de descriptors 152
degree of consensus (DoC)
– benzimidazoles against 321
– computation 320
– high value 321
– low values 320
– measure, to compare SAS maps 320
2D electrophoresis gels 39
density-functional theory (DFT) 247, 252, 258, 296, 351, 358, 359, 361, 362
– values for electric properties of HOOH 350
descriptor-based property prediction
– best practices development 33–57
– biodescriptors 39
– classical QSAR descriptors, and uses 38
– 0D, 1D and 2D computational descriptors 40, 41
– 3D descriptors, and beyond 41, 42
– descriptors, from spectroscopy/spectrometry and microscopy 40
– descriptors, QSARs lexicon 37–44
– experimentally derived descriptors 38–40
– leveraging experimental data, and limitations 36, 37
– local molecular surface property
– descriptor 42
– machine learning methods 44–52
– modeling strategies, definition 52–56
– models interpretation 36
– models validation 35
– posing the question 34
– quantum chemical descriptors 42–44
descriptors 160
– Brownian 372
– calculation
– according to fit of ligand 300
– software packages 72
– 4D: conformational-ensemble-based 299, 300
– derived from local properties 297
– MEP as descriptor for 298
– ParaSurf descriptors 298, 299
– PEST methodology 297, 298
– 3D QSA(P)R descriptors 299
– public logPow dataset 300
detour matrix 93, 121, 127
DF. See decision forest (DF)
DFT. See density-functional theory (DFT)
3D grid-based methods 299, 300
– CoMFA 299
– CoMSIA 299
digital systems 366
dipole moment 65, 70, 112, 202, 252, 351, 355, 359
Discovery Studio 23
dispersion interaction 295
distance matrix 73, 94, 112, 121, 134, 140, 189, 214, 265
diverse weighting schemes 131
diversity-oriented synthesis (DOS) 396
diversity-sampling scheme 37
2D matrix-based descriptors 114–120, 125
– calculation, by Dragon software 126
3D-MoRSE descriptor 169
4D QSAR methods 299
3D QSAR models 300
Dragon data set 131
Dragon software 20, 116, 117, 122, 126, 168, 265
dual activity difference (DAD) map 321

e
eigenvalues 113, 118, 180, 186
– analysis of the covariance matrix 45
– average of 117
– of Rouse matrix 192
eigenvector
– average coefficient 118
– coefficient sum 118
– logarithmic coefficient sum 118
electron density 42, 295
electronegativity, as atomic feature 253
electronegativity-weighted Barysz matrix 132
electronic density 70
electronic messages 366, 369, 372
– average energy 372, 374
– and state space 368
Index

electronic structure 251
– important quantity of molecular structure 254
– observations, and features 252
– perturbation treatment of 257
elementary reaction networks, reconstruction of 331, 332
– network search 331–333
energy
– components of atoms and bonds 256, 257
– and Hückel Approach 255, 256
– partitioning 255
enhanced replacement method (ERM) 150–153
– algorithms for 154
– standard deviation 155
ensemble techniques
– bagging 49
– boosting 49
– stacking 49
entropy 96–98, 258, 376, 387, 389
equilibrium constant 38, 66, 258
Estrada-like indices 118, 127, 142
ethanol molecule
– calculated torsion potential C–C–O–H 261
– C–C–O valence angle 260
– conformational space 260–262
– contour plot 262
– H–C–C–O torsional angle 260
– internal rotation, about C–C bond 260
– MD simulation 260
Euclidean distance 14, 46
expectation value (Q) 252, 263
f
factor analysis 45
feature selection methods 4
feedforward backpropagation neural network 5–7
feedforward neural network 6
fixed-length descriptors 41
flavonoids data sets, scaffold 125
fluorophilicity dataset (FLUOR) 152, 153, 156
folding degree index 94
forward stepwise regression 164
Frog software 20
Fukui’s frontier molecular orbital (FMO) approximation 296
function-oriented synthesis (FOS) 396
g
GAs. See genetic algorithms (GAs)
Gaussian distributions 9, 56, 182, 232–234
Gaussian kernel 8
Gaussian probability distribution 9
Gaussian process (GP) 9, 10
gel permeation chromatograph 175
generalized Wiener indices 114
general regression neural network (GRNN) 7–9, 15, 16
– basic equation 8
– layers 9
genetic algorithms (GAs) 150–153
– solutions in parallel, processing information 155
– using operator 155
genetic function approximation (GFA) 67
GETWAY descriptor 159
Gibbs free energy 347
Gini index 13
Gini’s concentration index 128
goodness-of-fit for regression models 35
graph-energy indices 116, 126
graph invariants, classes 113
graph-theoretical approach
– applications 193, 194
– g-factor 196
– rheological functions and descriptors, relationship 193
– Zimm matrix, dynamics of flexible chains 196
– to chain dynamics, and statistics 182
– intrinsic viscosity 188–190
– radius of gyration 182–185
– relaxation time and 191–193
– rouse dynamics 185–187
– scattering function 190
– isospectral tree graphs 196
– logarithmic plot of reduced intrinsic moduli 195
graph-theoretical matrices 111, 120–122, 125
– comparison 125–132
– dissimilarity analysis 128
GRIND descriptors 44
GRNN. See general regression neural network (GRNN)
h
Hall valence connectivity indices 70
Hamilton operator 252
Hammett equation 66
Hamming distance 123
Hansch equation 66
Harary-like index 115, 126
Hasse diagrams 163, 166
– partially ordered set 162
– for predicting ecotoxicological data (ME) 167
– QSAR based on 165
– for total order 162
Hasse diagram technique (HDT) 150
H-bonding 289
H-depleted molecular graph 73, 112
higher order Wiener numbers 114
high-throughput processes 40
Hosoya-like indices 114, 117
5-HT6 receptor ligands 3
hydrogen peroxide (HOOH) molecule, approach based on metrics 351
– calculated similarity for all TDs 356
– clustering in space of TDs 362
– density functional theory 350
– electric properties, ab initio method 350
– evolution of similarity S(method,CCSD(T)) for 359, 360
– method dependence of
  – anisotropy of dipole polarizability 353
  – mean dipole polarizability 352
  – mean of first hyperpolarizability 354
  – methods and computational strategy 354, 355, 357, 358
– minimum spanning tree (MST) 354
– for the space of TDs 361
– pattern space 353
– single-linkage cluster analysis 354
– theoretical description (TD), 352
hydrophobicity model system 66
HyperChem package 168
hyper-Wiener-like indices 114, 116

i
ideal chain models 180–182
information basis (IB) 159
information entropy analysis 96
intermolecular interactions 293
interpretation strategies 49
ionization energy 248, 253, 278, 296
ISOMAP method 46
isomers of hexane, modeling of
  – physicochemical properties 265
  – correlations
    – between boiling point, and heat of formation 274
    – of boiling point, with number of rotatable bonds 272
    – of heat of vaporization, and heat of formation 274
  – matrix between properties, and descriptors 270
  – database with SMILES notation, and experimental data 267
  – dimer of 2,2-dimethylbutane 273
  – experimental heat of formation
    – and calculated FF energy 270
    – and quantum chemical calculated HoF 271
  – heat of formation
    – and chl 1,268
    – vs. number of rotatable bonds 269
    – and Wiener index 267
    – with Zagreb index 269
  – intercorrelation between different properties 275
  – two molecules of n-hexane 273
  – isometric variant of SPE (ISPE) 46

j
joint probability density function (PDF) 7, 8

k
kernel function 18
kernel methods 50, 51
kernel partial least squares (KPLS) 24, 51
kernel trick 10, 50
Kier flexibility index 70
Kier–Hall connectivity indices 114
Kier index 73, 202
Kirchhoff number 119
KL-divergence to recall, regression curve generation 237. See also compound recall
  – activity class not included in regression analysis 237
  – data points, for regression curve 237
  – fingerprint
    – calculation, and bit frequency determination 237
    – yields low predicted recall 237, 238
    – rate per activity class 237
    – relation with KL-divergence 239
Klopman–Hudson equation 282
k-nearest neighbors (kNN) 14, 15, 47, 48
  – example 47
Konstanz information miner (KNIME) 21, 22
Kullback–Leibler (KL) divergence 230, 239

l
Lagrange multiplier 17, 18
Lagrangian expression 18
Laplacian matrix 119, 120
latent factors 4
  – extraction 5
latent variables. See latent factors
learning process 2
leave-one-out methods 4
lengthy process 70
Lennard–Jones interaction 294
ligand-based virtual screening 240
   – Bayesian approach application 230, 231
   – cumulative recall of active database compounds 229, 230
   – KL-divergence to recall, regression curve 237
   – for predicting compound recovery rates 229
light scattering (LS) measurements 175
linear combination 11
linear discriminant analysis (LDA) 11, 12, 51
   – application 12
linear free energy relationships (LFERs) 44
linearly separable data, binary classification 16
linear QSAR methodology 150–153
liver toxicity knowledge base (LTKB) project 103
loading plots 134–142
local electron affinity 296
local ionization energy 296
locally linear embedding (LLE) 46
logarithmic Randic-like eigenvector-based indices 119
logarithmic spectral positive sum indices 117
logistic regression (LR) 2, 10, 11
   – applications 11
   – resources 25
   – Lovasz–Pelikan index 94
m
MACCS fingerprint 238
machine learning methods 44–52
   – clustering method 46, 47
   – decision trees, and random forests 49, 50
   – factor analysis 45
   – kernel methods 50, 51
   – k-nearest neighbors (kNN) 47, 48
   – multidimensional scaling method 45, 46
   – neural networks method 48, 49
   – nonlinear dimensionality reduction method 45, 46
   – partial least squares regression (PLS) 47
   – principal component analysis method 44
   – ranking methods 52
   – stochastic proximity embedding method 45, 46
materials balance 329
MATLAB software 22, 23
matrix 111
   – dissimilarity, MDS analysis 143
   – operators, comparison 133–137
   – spectral moments 118
Maxwell–Boltzmann distribution 259
MDDR database 238
Metropolis–Hastings algorithm 151
modified RM (MRM) 154
   – evolution of 160
   – vs. RM 156–159
modified Tanimoto index 165
MOE. See molecular operating environment (MOE)
Mohar index 119
Molconn-Z software 21
mold2 descriptors 73
   – application program interface (API) 96
   – Balaban index descriptors 92
   – calculation 94–96
   – flowchart 95
   – constitutional descriptors 94
   – correlations between descriptors 98, 99
   – description 74–92
   – evaluation 96–99
   – information content-based descriptors 94
   – regression model in 103
   – Shannon entropy analysis, information content by 96–98
   – topological descriptors 73–94
molecular commonality, and similarity 263
molecular descriptors 3, 4, 10, 68, 69, 70, 95, 100, 105, 150, 159
   – bridge chemical structure 69
   – calculation 71
   – algebraic operators for 114
   – chemical information analysis
   – from graph-theoretical matrices 124–143
   – data sets 124, 125
   – 2D matrix-based descriptors 114–120
   – graph-theoretical matrices 120–122
   – comparison 125–132
   – linear combination 11
   – matrix operators, comparison 133–137
   – multivariate analysis 111–143
   – multivariate similarity analysis of chemical spaces 122, 123
   – for ranking, selection of 163, 164
   – role of 70, 71
   – selection of 265, 266
   – single operators
   – from different graph-theoretical matrices, comparison 137–143
   – threshold 13
   – types 71
   – vertex weighting schemes 122
   – weighted graph-theoretical matrices, comparison 130–132
molecular docking 67
molecular dynamics (MD) 259
– simulation 259
  -- to detect local interaction centers 289
  -- of ethanol molecule 260
  -- of phenylalanine with a water shell 290
  -- to study flexibility of molecules and molecular electrostatic potential (MEP) 294, 295
molecular flexibility 259
molecular matrix 111, 245
molecular messages
  – analysis of reactions, by Brownian computation 376
  – and brownian computation 370
  – with corresponding frequency, and energy terms 371
  – inhomogeneities and structure correlations 375, 376
  – message mutation 375
  – mutual information 371, 372
  – reactions of cyclohexene 377, 382
  -- binding affinity 386
  -- cyclization chemistry 380
  -- cycloheptanone converted to tropinone 384–386
  -- orientation of methyl groups 382
  -- ozonolysis 378, 379
  -- state points allied with 378, 381, 383
  -- state points and protease inhibitors 388
  – Shannon information 371, 372
  – state points and widths 373
molecular operating environment (MOE) 23, 24, 238, 315
molecular polarizabilities 296
molecular properties, type of 262, 263
molecular surface properties 285–290
  – electron density distribution, of phenylalanine 287
  – H-bonding, and hydrophobic areas 289
  – interaction potential surrounding structure 289
  – MD simulation of phenylalanine with water shell 290
  – molecular electrostatic potential (Z-clip) 288
  – molecular lipophilic potential 288
  – phenylalanine ball-stick model 287
  – phenylalanine CPK model 288
  – molecular topological indices (MTIs) 112
  – molecules, bonding in 254, 255
  – molecules, structure of
    – as Brownian computers 374
    – 3D molecular structure, characteristics of 248
    – and energy 250, 251
  -- gradient (gi) calculation 251
-- potential energy curve 251
  -- graph theory 246
  -- InChIKey 246
  -- MCS clustering of compounds 249
  -- marked clusters 250
  -- tree of that MCS clustering 250
  – molecular modeling software for 247
  – ROSDAL string 246
  – simultaneous optimization 247
  – SMILES 246
  – structure coding 247
  – X-ray diffraction 246
  Monto Carlo method 151
  Moran coefficient 93
  MRM. See modified RM (MRM)
multidimensional scaling (MDS) method 45, 46, 128
  multidimensional scaling plots 129, 132
  multilinear regression method 263
  -- application of 263
  -- descriptor reducing, structural information 265
  -- F-value, measure of quality 301
  -- mathematical models (QSAR) for 263
  -- molecular properties, and computational methods 264
  -- structure–property relations 265
  -- validation of derived mathematical model 264
  multiple instance ranking (MIRank) algorithm 52
  multiple linear regression (MLR) 4, 9, 11, 44, 222, 275, 282, 301
  -- general expression 3
  -- multisigma models 8
  multivariate analysis technique 111–143, 128
  multivariate similarity analysis, of chemical spaces 122, 123
  -- network search

  synaptic effects 37, 240
  network-like similarity graphs (NSG) 309
  network identification 340
  – DE settings 343
  – model selection methodology 343
  – reaction networks 340–342
  – results of running DE 344–346
  – time derivatives, estimation of 342
  network search
  – formulation of objective function for 335, 336
  -- no physical/chemical information 336, 337
Index

– physical/chemical information 336
– as nonlinear integer programming problem 332, 333
– reconstruction of elementary reaction networks 331, 332
neural networks 48, 49
n-fold cross validation 4
N-heterocyclic aromatics, basicity of 283
– descriptor LMO 283
– pKb values 283
noise deficient descriptors, for use in ranking 164
nonlinear dimensionality reduction methods 45, 46
nonlinear mapping (NLM) 46
normalized Randic-like eigenvector-based indices 119
normalized spectral positive sum indices 116

\( o \)
on-pass neural network learning algorithm 7
one-size-fits-all approach 34
Open Babel 19
open-source system. See RapidMiner Orange software 22

\( p \)
PaDEL-Descriptor 21
partial equalization of orbital energy (PEOE) algorithm 253
partial least-squares (PLS) analysis 4, 5, 47, 66, 67
– used for 4
partial least-squares regression (PLSR) 221, 222
partial-order ranking (POR) theory 150, 162, 169
partition coefficient 38, 65, 66, 167
Parzen’s nonparametric estimator 8, 15
PCDD data set, scaffold 125
Pearson’s product-moment correlation coefficient 35
PESD descriptors 56
P-glycoprotein inducers 13
pharmacophores induced fit enzyme–substrate binding 41
Polak–Ribiere algorithm 152
polarizability, defined 295
polarization 296
polychlorinated biphenyls (PCBs) 168
polymer molecules 175, 176. See also graph-theoretical approach
– application of graph theory concepts 176
– flexible polymer
– Gaussian chain 176
– molecular graph
– of randomly branched polymer chain 184
– Wiener indices for 183
– random flight statistics 176
prefabricated descriptor 53
principal component analysis (PCA) 44, 45, 66, 133
principal components regression (PCR) 47, 67
principal quantum number 122
probabilistic neural network (PNN) 15, 16
– advantages 16
– network architecture for 16
probability density function 15
probability distribution 96
property-encoded shape distributions (PESDs) descriptors 42
property-encoded surface translator (PEST) descriptors 43
property space, represented using molecular properties 314
proton affinity, modeling of 275
– basicity of N-heterocyclic aromatics 283–285
– correlation matrix of useful descriptors 284
– predicted PA of substituted imidazole molecules 286
– predicted PA vs. exp. PA 285
– proton affinity of pyridines 275
– data and mechanism 275–277
– models 277–283

\( q \)
QSPR/QSAR models 149, 150, 300
– applied to toxicology 150
– based on POR, application of 168
– based on structural similarity 217–219
– building and validation of 221–223
– dataset representation 220, 221
– dataset selection 219, 220
– design of 150
– employed HD for predicting ecotoxicological data 167, 168
– estimation of 166
– generation of 300–302
– linear modeling algorithms, for analyzing datasets 150
– validation of 166, 300–302
quadratic index 94
quantitative structure-activity relationships (QSARs) 33
Index

- assumption 69
- challenging aspects 68
- failures of 35
- history of 65–67
- leave-one-out cross-validation (LOO-CV) 54
- modeling methods 1–26
  -- ACD/ChemSketch 19
  -- ADRIANA.Code 20
  -- application of 12
  -- ChemDraw 19
  -- classification problems, methods for 10–18
  -- CODESSA 24
  -- Concord 20
  -- CORINA 19
  -- descriptor calculation 20, 21
  -- development workflow 2
  -- development software for 18–24
  -- Discovery Studio 23
  -- 3D structure generation 19, 20
  -- Frog 20
  -- general purpose 23, 24
  -- Konstanz information miner (KNIME) 21
  -- MATLAB 22, 23
  -- Molconn-Z 20
  -- molecular operating environment (MOE) 23, 24
  -- Open Babel 19
  -- Orange 22
  -- PaDEL-Descriptor 21
  -- RapidMiner 22
  -- regression problems, methods for 3–10
  -- R software 23
  -- smi23d 20
  -- structure drawing/file conversion 19
  -- SYBYL 23
  -- TANAGRA 22
  -- tuning and validation 2
  -- validation principles 26
  -- WEKA 22
  -- mold molecular descriptors for 65–105
  -- evaluation 96–99
  -- use 99–105
  -- molecular descriptors, bridge 68–71
  -- validated workflow 54
  quantitative structure–pharmacokinetic relationship (QSPKr) 1
  quantitative structure–property relationships (QSPRs) modeling methods 1–26
  -- classification problems, methods for 10–18
  -- regression problems, methods for 3–10
  quantitative structure–toxicity relationship (QSTR) 1
quantum mechanical energy 253
quantum mechanics (QM) 70, 150, 294
quasi-Wiener index 119
r
radial basis function (RBF) 10
Randic-like eigenvector based-indices 118
Randic-like index 115, 116, 127, 139, 140
random forest (RF) 12–14, 49, 50, 67
ranking methods 52, 150
-- principles of 159
RapidMiner 22
receiver operating characteristic (ROC) curve 35
reciprocal squared distance matrix 120, 126, 130, 133
RECON descriptors 42, 43
redundancy index 94
regression analysis 37, 237
regression coefficients 38, 68
regression models 103
replacement method (RM) 150–153
-- standard deviation 156
-- steps for 154
-- vs. standard deviation MRM 156–159
reverse engineering. See also chemical reaction network; network identification; network search
-- advantages of approaches 346
-- limitations 346
-- as useful description of a network from data 327
R-hopping 308, 312
root mean squared error (RMSE) 35
Rouvray index 93
S
SALI. See structure-activity landscape index (SALI)
Sanderson electronegativity 143
SAR. See structure-activity relationships (SAR)
SARI. See structure-activity relationship index (SARI)
Schrödinger equation 250, 365
Schultz index 94
self-consistent field (SCF) procedure 252
Shannon entropy analysis
-- information content by 96–98
-- results 97
-- of top descriptors 99
-- vs. reverse cumulative probability 98
similarity-based virtual screening 307
single matrix descriptors, PCs for PCAs 133
single matrix operators, role 124
single operators, from different graph-theoretical matrices, comparison 137–143
single-sigma model 8
smallest set of smallest rings (SSSR) 95
smi23d program 20
software Dragon 152
spanning tree number 119
Spearman’s rank coefficient 168
Spearman’s rank correlation coefficient 35
spectral absolute deviations (SpAD) 117, 138
spectral diameters 117
spectral indices 114
spectral mean absolute deviations (SpMADs) 117, 142
spectral positive sum indices 116
statistical learning theory 16
– minimization principle 16
stepwise multiple linear regression (SMLR) 67, 102, 103, 301
steric effects 38
stochastic proximity embedding (SPE) methods 45, 46
structural similarity
– approximate similarity (AS) 205–207
– chemical structural similarity 201–203
– 2D structural similarity 202
– molecular graph 203
– clustering models based on 207–211
– combining approaches 204, 205
– descriptor-based 203, 204
– molecular graph and 203
– TPIMS compounds 315
structure–activity landscape index (SALI) 56, 309, 310, 319, 323
structure–activity relationship index (SARI) 56, 309, 310, 323
structure–activity relationships (SAR) 33, 39, 56, 212, 307, 311
structure–activity similarity (SAS) map 311, 312
– of T. brucei cathepsin B 318, 322
structure–property–activity (SPA) similarities 321
structure property relationships 262
structure–selectivity relationships 311
support vector machines (SVMs) 16–18, 51, 67
– advantage 18
– based ranking 52
– margin and decision boundary 16
SVMs. See support vector machines (SVMs)
SYBYL software 5, 20, 23

\textbf{t}
TANAGRA software 22
Tanimoto coefficient 229, 240, 312
target factor analysis (TFA) 328
\textit{Tetrahymena pyriformis} 152
thermodynamic equilibrium 258
topological charge index 94
topological descriptors 69
topological distance index 93
topological distance matrix 120
topological indices (TIs) 112
– topostructural/topochemical indices 112
total walk count 94
TOX2 dataset 153
trial reaction networks, estimation of rate coefficients 333–335
\textit{Trichomonas vaginalis} 308, 309
\textit{Trypanosoma brucei} 317
– chemical structures of pairs of compounds, SAS maps of 319

\textbf{u}
variable Zagrebi indices 114
vertex weighting schemes 122

\textbf{w}
walk connectivity indices 94
wavefunction 252, 253, 263, 365
weighted graph-theoretical matrices 112
– comparison 130–132
weighted walk degrees 94
WEKA software 22
Whasse software 163
Wiener indices 73, 114, 115, 138, 142, 183, 193, 194

\textbf{y}
y-scrambling 54
– comparison 55

\textbf{z}
Zagrebi index 94, 114
Zauhar shape signatures method 43