Index

A
Absorption, distribution, metabolism, excretion (ADME), 216, 394–396, 489
databases and software systems, 489–495
Atomic properties
binary, 264–265
general, 265–267
B
Bioisosterism, 428
Biological promiscuity, 116–117
C
Chemical graphs, 6–10
hydrogen-suppressed, 7–8
maximum common substructure, 7–9
Chemical space representations, 35–40
binning, 282
dimensionality, 36–37
dimension reduction, 37–40, 281–282
Chemometrics, 171–172
Chirality, 144, 428–429
Chromosomes, 287
Classical sets, 4, 43–44
Clustering, 279–280, 303–304
Compound classification, 285–287, 296, 311–312
Cytochrome P450 isozymes, 450–451
docking models, 459–463
homology models, 450–459
pharmacophore models, 463–473
quantitative structure-activity relationship models, 473–488
D
Database access, 67–72
Data shaving, 94–97
Descriptors
3D-logP, 219–223, 256
3D pharmacophore, 358–360
BCUTs, 18–19, 282–283, 365–367
receptor relevance, 367–368
constitutional, 302, 327
correlation, 267–268, 295
descriptor medians, 293
electron density derived, 401, 411–412
encoding, 268–271
general, 302–303, 339–341
grid cell occupancy, 164
VSA, 265–267
Distance metrics (see Feature vectors)
Diversity (see also Molecular similarity), 51, 58
design, 308–310
selection, 295–296
Docking, 370–371, 439, 459–463
Drug-like features, 340–341, 355, 450
Index

F
- Feature vectors, 10–27
  - continuous-valued, 18–27
  - discrete-valued, 10–18
- distance metrics
  - Euclidean, 11–12
  - Hamming, 11–12
  - Minkowski, 21
  - Tanimoto, 11–12
  - Tversky, 13–16
- Fitness functions (see Scoring)
- Focused (or sequential) screening, 59–60, 326
- Force fields, 144–145, 441–442

G
- G protein coupled receptors (GPCRs), 356–357
- Genetic algorithms, 161–164, 343, 415
  - pattern recognition, 412–419
  - boosting, 417–418
- Gibbs free energy, 440

H
- High-throughput screening, 86–87, 112

K
- Kinases, 369–372
- K-means, 304

L
- Library design, 75–82, 87–88
  - combinatorial, 335–336, 361
  - cherry-picking, 338
  - filtering, 77–79, 338, 345
  - multi-objective evolutionary algorithms, 341–344
  - reactant- vs product-based design, 337–338
  - simulated annealing guided diversity sampling, 382
  - enumeration, 79–81, 338–339
  - focused, 368
  - simulated annealing guided focusing, 383
  - targeted, 355–356
    - cell-based, 365–368
    - pharmacophores, 358–363
    - privileged substructures, 363–364
    - target class-specific, 355–356
    - target structure-based, 368–372
- Linear regression (see Quantitative Structure-Activity Relationship)
- Lipophilicity, 216–217
  - molecular lipophilicity potential, 220

M
- Molecular alignment (see Quantitative Structure-Activity Relationship)
- Molecular complexity, 117–118
- Molecular fields, 27–30, 176–178
  - field-based similarity indices, 31–32
  - Gaussian functions, 28–30
- Molecular fingerprints (see also Feature vectors), 10–11
- Molecular similarity, 1–2, 51–52
  - dissimilarity, 35, 382
Index

Molecular surface, 219–220

N
Neural networks, 178–181

O
Olfactory stimulants, 400–401

P
Pareto ranking, 343
Partial charges, 145, 190
Partial least squares (see also Quantitative Structure-Activity Relationship), 174–176
Partitioning
  cell-based, 282–283
  cell coverage, 304–306
  median, 292–293, 297–298
  recursive, 319–327
    feature selection, 325–326
    prediction, 325–326
    p-value, 320–322
Principal axes, 430–432
  chirality measure, 431
  symmetry, 430
Principal component analysis, 39–40, 172–174, 283–284, 413–415
Principal coordinate analysis, 39–40

Q
Quantitative structure-activity relationship (QSAR)
  2D-QSAR, 136
    linear correlation coefficient, 170–171
    linear regression, 169–170
  3D-QSAR, 136–139
  comparative molecular field analysis (CoMFA), 176–178
  comparative molecular moment analysis (CoMMA), 137–138
  comparative molecular similarity index analysis (CoMSIA), 137
  orthogonal signal correction, 223, 228
  partial least squares regression, 176, 223, 232
  self-organizing molecular field analysis (SOMFA), 137, 194–197
  4D-QSAR, 163–168
    conformational ensemble profile, 163–164
    consensus model, 167–168
    lack of fit function, 165
    binary QSAR, 92–93, 183–184, 267, 276
    bioactivity data, 142–144
      scaling, 143–144
      transforming, 142–143
      descriptor calculation, 157–159
    metabolism prediction, 473–488
    molecular alignment, 150–156
    molecular conformation, 145–148
      conformational search, 148–150
    overfitting, 184
    receptor-dependent, 140–141
    receptor-independent, 135–136
    training and test sets, 156–157
    validation, 184–188
      cross-validation, 185–186
least squares fit, 185
scrambling, 187–188
Quantitative structure-property relationship (QSPR), 271

S
Scaffold hopping, 290
Scoring
functions, 286, 296–297, 342,
381–382, 384, 416, 440
components, 115
consensus, 121–122, 445
tools, 53–54,
empirical, 443
force field approximations, 441–442
knowledge-based, 443–444

Similar property principle, 53
nearest neighbors, 52–53
similarity coefficient, 52–53

Similarity searching, 52–53, 72–75, 88–90
information retrieval, 53–54
cumulative recall graph, 57
enrichment factor, 55
G-H score, 55

precision, 54
recall, 54
Simulated annealing, 160–161, 381–382
Support vector machines, 181–183

R
Random selection, 328
Rule-based evaluation, 91–92

T
Tanimoto coefficient (see Feature vectors)
Three-dimensional (3D) pharmacophores (see also Descriptors), 358–360

V
Van der Waals surface area, 262–264

Virtual screening, 52, 139–140, 297–298, 439

W
Web-based tools, 67, 495