

INDEX

A

Absorption, distribution, metabolism, elimination
(ADME)..... 13, 34, 428, 435
Absorption, distribution, metabolism, elimination and
toxicity (ADMET)..... 11, 44, 377, 378
Active learning. 369, 370, 373–377, 381, 382, 385–387,
389, 391, 392, 394, 396, 397, 399–402, 404, 406
Aldose reductase (ALDR) 185, 186, 188
AlloScore 308, 309, 316, 317, 322, 323
AlloSite 301, 306, 308, 315, 317, 320, 321, 324
Allosteric Database (ASD) 301, 303, 304, 306, 309, 310,
312, 313, 324
Allosteric site, *see* Binding pocket, orthosteric
Analysis of Variance (ANOVA) 229, 261, 272, 273
Anatomical Therapeutic Classification System (ATC) .. 16
ASBench 303, 307, 310, 315, 316
Assay record..... 27, 28, 71–73
Automated retrieval 101–111

B

Bemis-Murcko (BM) scaffold..... 327–336
Benzene 171, 172, 197
Binding pocket 295–324, 366, 431, 440
allosteric 295, 440
orthosteric 281–293, 440
Bioactivity
EC50..... 59
IC50..... 18, 59, 125, 132, 136
Kd..... 59
Ki..... 18, 59, 132, 379
Bioactivity record 26, 140, 142, 158
collection 158
representation..... 141
Bioactivity table, *see* Bioactivity record, collection
Biological effect..... 3, 36

C

Cancer Genome Atlas 419, 440
ChEMBL6, 7, 11–14, 19, 26–29, 59, 99–106, 113, 116,
125, 132, 133, 135, 165, 224, 328, 340, 341,
351, 373, 378, 426, 427, 429, 437, 441
Chemical Abstracts Service (CAS) 29, 204

Chemical Entities of Biological Interest (ChEBI) 6,
14–15, 29–35, 46, 60, 61, 214, 216, 426, 427,
429
Chemically Advanced Template Search (CATS) 173, 191,
193, 197–199, 201, 202
Chemokine 296
ChemProt 6, 7, 16–20, 57, 59, 61, 427, 436
ChemSpider..... 6, 15, 34–42, 59, 60
Cisplatin 43, 45–48, 50, 51, 414
Clustering 7, 10, 12, 23, 117, 180, 221, 227–229,
251–254, 283, 390, 432, 441, 442
hierarchical..... 180, 229, 251, 252, 255
k-means..... 180, 229, 253
Compound
analog series..... 329–331, 345
filter 33, 71, 78, 79, 82, 83, 112
promiscuity 112, 334, 335, 340, 348, 350
scaffold 327–336
visualization 95, 115, 127, 340, 441
Conformer 14, 84–86, 88, 177, 296, 297
Connectivity map 427, 433
Correlation vii, 191, 193, 227–229, 232, 239, 244,
246–248, 250, 251, 257, 269, 274, 357, 358,
360, 362, 382, 387, 405, 431, 433, 435, 436, 443
Cyclin-dependent kinase (CDK) 101, 102, 104, 105, 108,
110–112, 116, 117, 119, 122, 124, 204, 214,
215, 221

D

Dimensionality reduction
multidimensional scaling (MDS) 178, 187
Disease factor 4
Disease-oriented tool 7
Distance measure 180, 193, 196, 204, 205
DNA replication..... 4
DNA structure 4
Docking 60, 127, 182, 323, 356
DrugBank 7, 19, 32, 214, 224, 356, 357, 429
Drug design. 7, 173, 296, 300, 303, 306, 309, 322, 430
Drug discovery vi, 4, 8, 11, 13, 20, 57, 59, 281, 304, 339,
355, 400, 425, 429, 433, 435, 436, 441, 443
DrugEBIity 13
Drug optimization vi, 59, 426, 440

E

- Epidermal growth factor receptor (EGFR) 413–415, 421, 435, 437
- Etoposide 4, 5, 14, 16–24, 29–32, 35–40, 43, 49, 61
- Extensible Markup Language (XML) 10, 61, 76, 309

F

- FASTQ format 416

G

- Gefitinib 413
- Gene Ontology (GO) 23, 428, 429
- Genetic mutation v
- Genetic sequence
 alignment 300, 358, 391, 403, 418
 read quality 417
 variant 416, 417, 422
- Genetic variant 45
- G protein-coupled receptor (GPCR) 13, 59, 132, 135, 160, 161, 163, 165, 187, 296, 303, 311, 356, 358, 370, 373, 377, 378, 391, 398, 401, 402, 438, 439
- dopamine 300, 391

H

- High-content screening (HCS) 355, 432, 433
- High-throughput screening (HTS) v, 281, 339, 355, 426, 432, 433, 435, 436, 442
- Hit-to-lead 4, 438
- Hopkins statistic 229, 251
- Hydrogen bond
 acceptor 41, 85, 173, 176, 191, 192, 201, 232, 239–242, 248, 251
 donor 41, 85, 173, 176, 191, 192, 201, 240–242, 248, 251
- Hypothesis testing 227, 229, 252, 257–259, 261, 270–273

I

- IChem 282, 284, 289, 292
- ImageMagick 116
- Inhibitor 4, 5, 25, 30, 50, 65, 105, 112, 117–121, 124, 185, 303, 313, 327–336, 413, 415, 426–428, 433, 435, 437, 440, 442
- Interactions
 chemical-chemical 7
 chemical-protein 6, 7, 16, 20, 21, 427, 438
 ligand-target 132, 399, 404
 protein-protein 7, 281, 433
- Interactome 7

- International Union of Pure and Applied Chemistry (IUPAC) 14, 127, 211, 427
- IUPHAR 7, 429, 435

K

- Kernel methods 355–366
 local alignment kernel 358, 366, 391
 machine learning (*see* Machine learning)
- Kinase 13, 59, 65, 100–103, 118, 121, 124, 187, 303, 306, 311, 336, 358, 369, 373, 378, 394, 402, 403, 413, 433, 435, 437–442
- Kinase-Ligand Interaction Fingerprints and Structure database (KLIFS) 429, 440
- KNapSAcK database 132, 214
- Konstanz Information Miner (KNIME) 182–188, 190, 191, 194–196, 203, 204
- Kyoto Encyclopedia of Genes and Genomes (KEGG) 23, 32, 132, 212–215, 221, 224, 303, 356–358

L

- Ligand-target database 96, 100–121, 135, 161
- Lipinski
 Alert Index 190
 rule of 5 71, 228
- LogP 11, 190, 238, 240, 242, 246, 248, 251

M

- Machine learning vi, 97, 108, 182, 221, 227, 270, 274, 282, 289, 356, 357, 359, 362, 369, 370, 372, 374, 375, 381, 383, 385, 386, 399
 decision tree 382
 linear classifier 356, 360
 logistic regression 356, 361
 random forest algorithm 370, 374, 382
 scikit-learn 375, 384
 support vector machine 221, 356, 357, 361–364, 370
- Matched molecular pair (MMP) 329–331, 336, 342, 343, 351, 441
- Matthews correlation coefficient (MCC) 372, 386–389, 392, 405
- Merck molecular force field (MMFF) 86, 190
- Molecular descriptor 171–205, 211–224, 228, 234, 251
 calculation 174, 175, 178–180, 182, 185–187, 190–194, 203, 204, 228, 232
 circular fingerprint 190
 dragon software 173, 185, 187, 190, 191, 204
 extended connectivity fingerprint 190, 191, 199
 MACCS key 215, 221
- Molecular Dynamics (MD) 300, 301

- Molecular representation
 dimensionality 174, 179, 180, 191, 193, 203
 graph 175, 188
 InChI key 113
 protein 174, 187, 203
 SMILES 185, 188, 189, 203
 Molecular weight 11, 28, 33,
 41, 172–174, 178, 180, 190, 194, 232, 238–241,
 243, 246, 248, 250, 251, 263, 265, 274
 Molecule search
 feature vector 357, 358,
 360, 363, 365, 403, 432
 fingerprint 18, 84, 115,
 118–120, 127, 128, 173, 178, 185, 190, 191,
 194, 195, 199, 204, 214, 215, 221, 357, 376,
 378, 426, 429, 432, 436, 440, 441
 structure 61
- N**
 NCBI Entrez 66, 68, 79, 80
 Next-generation sequencing (NGS) 415–419,
 421, 422
- O**
 On-line service 4
 OpenBabel 99, 113, 116, 127,
 188, 189, 203, 214, 376, 402, 404
 OpenEye 342, 345, 376, 378, 399, 402, 441
 OpenPHACTS 426, 429
 Orthosteric site, *see* Binding pocket, allosteric
- P**
 Pathway 5, 7, 13, 15, 18, 20, 23,
 30, 43, 52, 59, 132, 211, 215, 221, 296, 298,
 300, 302, 303, 426, 429–433, 435, 436, 439,
 440, 442
 Pharmacogenomics 15, 42,
 52, 54, 56, 61, 413–422
 Pharmacophore 191, 201, 202,
 282, 289, 292, 293, 366, 426, 440
 PharmGKB 6, 7, 15–16, 42–45,
 47, 48, 50, 52–54, 56, 57, 60, 61, 132
 Phylogenetic tree 373, 375,
 390–392, 427, 440, 442
 pIC50, *see* Bioactivity, IC50
 pKi, *see* Bioactivity, Ki
 Platelet-derived growth factor receptor
 (PDGFR) 65, 72, 78–83
 Polypharmacology 340, 438, 439, 442
 Protein
 domain 73, 203, 221, 309, 358, 362, 365
 homology 7, 391, 430, 440
 sequence 16, 21, 64, 99, 101,
 106, 108, 300, 356, 376, 403, 438
 Protein Data Bank (PDB) 108, 281–284,
 286, 288, 291–293, 296–298, 306, 310, 312,
 315, 316, 318, 375, 429, 440
 Protein-protein
 interaction 7, 16, 21, 281, 282, 433
 interface 281
 PubChem vii, 7–11, 24–26, 63
 PubChem BioAssay 8, 10, 11, 24–26,
 69–72, 79–82, 340
 PubChem Compound 8, 9, 12, 67,
 69, 71, 74, 75, 81–83
 PubChem Substance 8, 9
p-value 23, 47, 229, 247,
 250, 251, 261, 269, 271, 273, 274
 Python programming language 97, 133, 375
- Q**
 Quantitative structure activity relationship
 (QSAR) 20, 172, 173, 175,
 177, 181–183, 356, 438
- R**
 RDkit 97, 99, 115–118,
 127, 190, 204, 214, 376, 402
 Resource Description Framework (RDF) 12,
 373, 436
 ROC curve 257
 Boltzmann-Enhanced Discrimination of Receiver
 Operating Characteristic (BEDROC) 181
 Root Mean Squared Error (RMSE) 181
 R programming language 227
- S**
 SARfari 13, 59, 133, 135, 136,
 151, 160, 161, 163, 165, 373, 377, 378, 381,
 387, 391, 394, 398, 401–404, 437
 SAR Matrix (SARM) 339–351,
 433, 437, 439, 441, 442
 Scalable Vector Graphics (SVG) 127
 SciPy 377
 Similarity Ensemble Approach (SEA) 7, 16, 19
 Similarity metric
 Kendall tau 244
 Manhattan distance 195
 Spearman coefficient 228, 244, 248, 250, 269
 Tanimoto coefficient 84–86, 195, 358
 Similarity principle 65, 444
 Similarity search 65, 175,
 180–183, 186, 187, 196, 215, 221, 356, 435,
 436, 441

Standard Query Language (SQL).....131, 132, 135, 373
 Statistical correction
 Benjamini-Hochberg 273
 Bonferroni correction 273
 Statistical test
 Kolmogorov-Smirnov test 388
 Kruskal-Wallis test 272, 273
 two-tailed T test 388
 Wilcoxon rank-sum test 272, 273
 STITCH 6, 7, 21–24, 59, 132, 430, 436
 Structural chemogenomics 440
 Structure-activity relationship (SAR) 10, 11,
 59, 215, 300, 341, 343, 345, 346, 348–351, 425,
 428, 430, 433, 435–437, 439, 441, 442
 Structured data file (SDF) 34, 40,
 61, 76, 84, 113, 114, 118, 119, 203, 212,
 214–217, 224, 228, 404

T

Targets
 allosteric vii, 296, 300, 303,
 304, 310, 316, 317, 440
 Topoisomerase 4, 5, 433

U

Uniprot
 accession id69, 104, 108
 UNIX
 Linux.....96, 97, 123, 127, 133, 215
 shell96, 123, 133

V

Vascular endothelial growth factor receptor
 (VEGFR) 65, 78–80, 82, 83, 88
 Virtual screening 173, 180, 182,
 184, 185, 187, 188, 193, 196, 197, 328, 384,
 393, 437–439
 enrichment curve 186, 187
 enrichment factor 181, 186, 187
 Vitamin D receptor 26, 27, 29
 VolSite286, 289, 290, 293

W

Web database3–61, 66, 73, 97, 99–105, 131, 232

Z

ZINC database 231–234