

Index

a

absorption, distribution, metabolism, and excretion (ADME) profile 218
 Accelrys enhanced stereochemistry labeling 81
 ACD/Labs Name Batch tool 87
 ACD/Name to Structure Batch 90
 Adaboost 235
 ADME 218
 affinity database 11
 algorithm adaptation method 233ff
 Analysis of Cell-Based Screening Data 141
 Analysis of HTS Data 136
 analytical data
 – linking chemical information 91
 ARES (assay registration system) 143
 association rule 246
 atom tracing 109
 automatic processes
 – database 87

b

basic formal ontology (BFO) 60
 batch 77
 – concept 77
 – reassignment 87
 bicluster 276
 biclustering algorithm 276
 BIF, *see* biological impact factor
 binary kernel discrimination (BKD) 30
 binary relevance problem transformation 235
 binary relevance transformation 234ff
 bindability 17
 binding affinity data
 – collection 11
 binding site 17
 – comparison 17
 – detection 17
 – quantitative measure 18

– similarity-based profiling 262
 BindingDB 26f
 – browsing 27, 28
 – compound 28
 – data set 30
 – downloading capabilities 27
 – linking with other databases 29
 – querying 27
 – target 27
 BindingMOAD 38
 Biomarkers Knowledge Area 222ff
 bio-ontology 56
 bioactive compounds
 – data mining 131
 – identification via high-throughput screening 131ff
 bioactivity 26
 bioactivity data
 – linking chemicals 93
 biological entities 228
 Biological Expression Language (BEL) 298
 biological impact factor (BIF) 290, 302ff
 biological network 293ff
 Biological Networks Gene Ontology tool (BiNGO) 68
 biomarker 307, 309
 Biomarkers Knowledge Area 222f
 biomedical ontology 56
 BioPAX 119
 BLAST 124f
 box plot
 – quality control 171
 BRENDA database 123

c

.C 189
 C/C++ library 200
 C-side 193
 calling convention 187

- CAS Registry 108
 - CAS registry number 75, 107f
 - cell-based screening data
 - analysis 141
 - mode of mechanism hypotheses 141
 - ChEBI 61
 - ChemBioBank (CBB) 228ff
 - ChEMBL 26, 31ff, 96
 - browsing 31
 - compound 31
 - Compound Set Enrichment (CSE) and
 - Docking 144ff
 - Conditions (Parameters) for Analysis of HTS
 - Screens 133
 - content 31
 - data set 33
 - downloading capabilities 35
 - linking with other databases 32
 - querying 31
 - targets 31
 - ChemDB 91f
 - chemical
 - calculated properties 63
 - linking to bioactivity data 93
 - Chemical Abstracts Service (CAS) 107
 - chemical cartridge 79
 - Chemical Component Dictionary 7
 - chemical compound 228
 - chemical compound database 26, 108
 - Chemical Entities of Biological Interest (ChEBI) ontology 61
 - chemical information
 - linking to analytical data 91
 - Chemical Information (CHEMINF)
 - ontology 65
 - chemical ontology 55ff
 - chemical probe 228
 - chemical structure 79, 203
 - chemical text mining 66
 - ChemicalTagger 121
 - chemogenomic screening 20
 - chemoinformatics tool 179ff
 - ChemSpider 38, 96, 108
 - chromosomal map 171
 - CID (PubChem compound ID) 108
 - Clinical Studies Knowledge
 - Area 223ff
 - CMPSim 67
 - combinatorial biosynthesis 102
 - command edition 181
 - command system 181ff
 - comodule (CM) 278
 - Common Visualization Tools 157ff
 - Companies & Research Institutions
 - Knowledge Area 224
 - compilation 190ff
 - compound 105
 - ambiguity 81, 90, 93, 118
 - exposure 292
 - identification 144
 - known unknown 87
 - name into structure transformation 89
 - purity 133
 - representation 84
 - undesirable compounds in hit list 136
 - compound activity database 94f
 - compound ambiguity 118
 - compound representation
 - standardization 84
 - compound set enrichment (CSE) 144
 - identification of hit series and SAR 144
 - identification of new compound 144
 - compound testing 289
 - condition
 - external 276
 - internal 276
 - confidence 247
 - Content Development Strategies 211ff
 - copy number variation (CNV) 272
 - copy transformation 236
 - CORINA 149
 - Corporate Chemical Database 75ff
- d**
- data
 - accuracy of the registration 82
 - content 5
 - format 6
 - migration 89
 - mining, *see* data mining (DM) 68
 - quality 6
 - registering 78
 - uniformity 6
 - data aggregation system 135
 - Data Architectures for the Analysis of HTS
 - Data 133ff
 - data management feature 227
 - data mining (DM) 68
 - assay condition 134
 - identification of bioactive compounds via high-throughput screening 131ff
 - integrative and modular analysis approach 273ff
 - knowledge-based 232
 - ligand profiling 257ff

- plant metabolic pathway 101ff
- purity 133
- rule-based method 241ff
- target fishing 257ff
- data set preparation 243
- data production 291
- database
 - automatic process 87
 - coordination 44
 - data quality 44
 - implementation of the platform 88
 - integrating 118
 - linking journals 45
- dendrogram 204f
- descriptor 203
 - electronic 204
- Disease Briefings Knowledge Area 221
- directed R-group combination graph (DRGC) 253
- disease
 - comparing healthy with unhealthy tissue or patients 177
- Disease Briefings Knowledge Area 221
- Disease Understanding 177
- DNA microarray 294
- docking
 - identification of new compound 144
 - protein-ligand 16
- dose–response curve (DRC) 131, 164
- drawing rules for scientists 82
- drug design
 - application 16
 - PDB-related database 10
 - structure-based computer-aided 3
- drug discovery
 - interactive visual analytics 155ff
 - knowledge pyramid 212
 - structural database 3ff
- drug treatment
 - measure effects on a cellular level 177
- drug–receptor interaction 217
- DrugBank 38
- Druggability 17
- Drugs & Biologics Knowledge Area 216
- DSSTox 96

e

- EC number, *see* enzyme commission (EC) number
- EFICAz 124
- EFICAz2 124
- ENZYME database 123
- enzyme commission (EC) number 110, 125

- enzyme function
 - 3D protein structure information 125
- enzyme function prediction
 - protein sequence information 123
- enzyme identification 123
- enzyme information 110
 - pathway prediction 126
- enzyme–target cell interaction 217
- Estate 204
- EU-OPENSREEN project 228, 230
- European Strategy Forum on Research Infrastructures (ESFRI) 230
- Experimental Models Knowledge Area 218
- Experimental Pharmacology Knowledge Area 217
- experimental reproducibility 298
- experimental system 292
- ExpressionView 281
- Extended 204
- external pointer reference 196
- extraction–transformation–load (ETL) system 135

f

- filter
 - OpenEye 183
- final network model 300
- fingerprint 205
- FlexX docking method 259
- flux balance analysis (FBA) 102
- formal concept analysis (FCA) 242
- Fortran 189
- FragFCA 242
- fragment swapping 247
 - hybrid structure 247
- frequent hitter (FH)
 - analysis 136
- Frequent Hitters in Hit Lists 136

g

- GCRMA (GeneChip robust multichip average) algorithm 296
- gene expression 169, 273, 277, 295ff, 302ff
 - heat map 169
- Gene Ontology (GO) 56
- Gene Ontology term 143
- Gene Ontology (GO) term enrichment component 143
- Gene Ontology tree map 174
- gene set enrichment analysis (GSEA) 143
- GenMAPP Pathway Markup Language (GPML) 120

genome-wide association study (GWAS) 271ff
 genomics 168
 – visualization 168
 Genomics Knowledge Area 223
 geometric perturbation index (GPI) 305
 Genomics Visualization Tools 168ff
 Glide 149
 GRAC 38
 Graphical User Interface (GUI)
 facility 114

h

Hadoop project 237
 heat map 164ff
 – gene expression 169
 – hierarchical clustered 168
 – triangular 176
 Het-PDB 9
 HIC-Up 9
 hidden Markov model (HMM) 124
 high-throughput screening (HTS)
 – analysis of data 133ff
 – identification of bioactive compounds 131ff
 histogram 162
 – quality control 171
 Hit-Hub 132, 135, 137
 hit series
 – compound set enrichment 144
 homologous organ group (HOG) 284
 HTS Explorer 150
 hybrid structure 247
 – fragment swapping 247

i

Identification of Bioactive Compounds via
 High-Throughput Screening 131ff, 141
in silico ligand profiling method 258
 InChI, *see* International Chemical Identifier
 integrative and modular analysis
 approach 271ff
 interactive analysis 166
 interactive visual analytics 155ff
 International Chemical Identifier (InChI) 107
 International Union of Pure and Applied
 Chemistry (IUPAC) 107
 ISIDA 183f
 Iterative Signature algorithm (ISA) 277
 IUPAC InChIKeys 136
 Informative Visualization 156
 ISIDA descriptors 183ff
 IUPAC, *see* International Union of Pure and
 Applied Chemistry
 IUPHAR-DB 38

k

k-nearest neighbors algorithm 235
 karyogram 171
 KEGG, *see* Kyoto Encyclopedia of Genes and
 Genomes
 KEM® 242ff
 knowledge area 215ff
 Knowledge-Based Data Mining
 Technologies 232
 Knowledge Challenges in Drug Discovery 212
 knowledge discovery 65
 knowledge pyramid 212
 Kyoto Encyclopedia of Genes and Genomes
 (KEGG) database 109ff
 – database structure 113
 – navigation 113
 – pathway painting 126

l

label powerset transformation 234
 large-scale molecular and organismal
 traits 271ff
 ligand
 – relationship to target 19
 LIGAND database 113ff
 ligand descriptor-based *in silico* profiling 264
 Ligand Expo 9
 ligand profiling 257ff
 ligand–protein interaction 261
 ligand–protein recognition PDB-related
 database 9
 ligandability 17
 LigandScout 261
 Lightweight Directory Access Protocol
 (LDAP) 86
 Literature Knowledge Area 225ff
 literature mining 121
 literature model 298

m

MACCS 204
 mammalian data set 281
 matched molecular pairs (MMP) method 253
 MBRole 68
 Measure Drug Treatment Effects 177
 mechanism of action model 235
 medicinal chemistry data 39ff
 metabolic design and prediction 103
 metabolic flux analysis (MFA) 103
 metabolic pathway 173
 metabolic pathway database 117
 MetaCrop 118
 MetaCyc 116

Microarray Quality Control Phase I (MAQC-I)
 approach 271ff
 minor allele frequency (MAF) 272
 MMAC 235
 MOA 142ff
 mode of mechanism hypotheses 141
 – analysis of cell-based screening data 141
 modular analysis tool 281ff
 module commonality 279
 module visualization 280
 molecular docking 125, 147, 259
 molecular fingerprint 204
 Molecular Libraries Program (MLP)
 228ff, 231
 Molecular Libraries Screening Centers
 Network (MLSCN) 232
 molecular phenotype 96, 273, 275
 MolSMILESSet function 199
 multilabel classification problem 233
 multiple objective optimization 252

n

name transformation into structure 89
 name mangling 187
 natural language processing (NLP) 121
 NB (naive Bayesian) classifier 138
 NDFI (NIBR Data Federation Initiative) 151
 network perturbation amplitude (NPA) 302ff
 normalized Cscores 150
 Novartis Lead Finding Platform 131

o

OBO, *see* Open Biomedical Ontologies
 OEChem library 188
 OEGraphMol method 198
 online enrichment analysis 280
 ontology 55ff
 – biology 57
 – chemical 60
 – medicine 57
 ontology-based enrichment analysis 68
 ontology interoperability 60
 Ontologies Release Tool 58
 Open Biomedical Ontologies (OBO)
 – format 58
 – Foundry 57
 Open PHACTS consortium 151
 OpenEye 188
 Organic Synthesis Knowledge Area 220f
 organism-specific pathway database 122
 organismal phenotype 273, 275
 orthologue 124
 OSCAR3 program 121

– R 181
 OWL 58

p

paralogue 124
 Patents Knowledge Area 225
 pathway 111
 – distinction between pathway and
 superpathway 111
 – format for exchanging data 119
 – obtaining information 116
 – typical size 111
 pathway database
 – adding information 120
 – constructing organism-specific
 database 122
 – manual curation 120
 Pathway/Genome Database (PGDB) 113
 pathway management platform 111
 pathway management software 112
 pathway modeling 102
 pathway painting
 – KEGG reference map 126
 pathway prediction 126
 pathway reconstruction 126
 – Pathway Tools 126
 pathway representation 103
 Pathway Tools platform 113
 – content creation and management 114
 – database management 114
 – pathway reconstruction 126
 – visualization capability 115
 Pathway Tools (PWT) software suite 110
 PDB, *see* Protein Data Bank
 perturbed network 299
 pharmacophore 261
 phenotype 273ff
 – high-dimensional 275
 – molecular 275
 – organismal 275
 phenotypic readouts 301
 Ping-Pong algorithm (PPA) 278ff
 Pipeline Pilot protocol 144
 plant metabolic pathway 101ff
 PlantCyc 116
 Problem Transformation Methods 233
 polypharmacology 241
 polypharmacology data set
 – rule-based methods to data mining 241ff
 polypharmacology space 248
 Poroikov's PASS 236, 264
 principal component analysis
 (PCA) 297

- problem transformation method 233
- profiling profile 162, 242, 263
- Programming in R 179ff
 - Binding to C/C++ libraries 200
 - Chemoinformatics tools integration 179ff
 - Command System 181ff
 - Compilation 190
 - Java/rJava 200ff
 - Name Mangling 187
 - R Internals 194
 - Rcdk package 202
 - SEXP 195
 - Shared Library 185ff
 - System call 180
 - Third party software integration 180
 - Wrapping 191
- PROSITE pattern 124
- Protein Data Bank (PDB) 3ff, 5ff, 260
- protein fold topology (PFT) 20
- Protein-ligand
 - binding site 11, 263
 - complex 12
 - docking 16, 259
 - fingerprint 263
- Protein
 - drug discovery 3ff
 - enzyme function prediction 123
 - hot spot prediction 19
 - sequence information 123ff
 - structural database 3ff
- protein structure 3
- 3D protein structure information
 - enzyme function 125
- Prous Institute's BioEpisteme 236
- PSMDB database 12
- PubChem database 34ff, 96, 108, 232
 - bioassay 34–36
 - browsing 35
 - compound 35, 36
 - content 34
 - dataset 37
 - downloading capabilities 35
 - linking with other databases 37
 - querying 35
 - target 35
- PubChem BioAssay repository 133
- public compound activity database 46
- public domain database
 - medicinal chemistry 25ff

q

- QSAR/QSPR model 204
- quality control 171, 296

- quantitative measurement 271
- quantitative trait loci (qlt) 271

r

- R internals 194
- R-side 193
- ranking by pairwise comparison (RPC)
 - transformation 234
- raw data normalization 295
- rcdk 202ff
- RCR, *see* reverse causal reasoning
- RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) 9
- reaction 109
 - definition 109
- redundancy 118
- reference pathway database 116
- record
 - uniqueness 80
- reference enzyme database 123
- Reference Enzyme Sequence Database (RES D) 123
- registrar 86
- registration
 - data 82
- registration area 86ff
- reproducibility
 - experimental 298
- response profile 302
- reverse causal reasoning (RCR) 300f
- rJava 200
- RMA (robust multichip average)
 - algorithm 296f
- RPAIR 110
- rule
 - generation 248
 - polypharmacology space 248

s

- SAMBA (Statistical-Algorithmic Method for Bicluster Analysis) biclustering method 276
- SAR (structure–activity relationship) 145f, 242ff
 - compound set enrichment 144
- SAR from Primary Screening Data 144
- SAR table 157ff
 - color-coded 158f
- sc-PDB (screening the Protein Data Bank) 12ff
 - content 13
 - database setup 13
- SAR transposition 146
- scientific relevance 297

- selectivity 249f
 - shared library 185ff
 - shared library call 185
 - Signature algorithm 276
 - significance analysis of microarrays (SAM)
 - approach 296
 - Simplified Molecular Input Line Entry
 - Specification (SMILES) 81, 107, 190
 - codes 107
 - parsing function 188
 - single-label classification problem 233
 - small molecule binding
 - database 26
 - small molecule chemistry 76
 - small molecule database 26, 38
 - SMILES, *see* Simplified Molecular Input Line Entry Specification
 - SMIREP 242ff
 - SNP (single-nucleotide polymorphism) 173
 - SpecDB 91f
 - SpecID 92
 - Spotfire 150
 - SRP, *see* systems response profile
 - standardization 64
 - compound representation 84
 - technical 297
 - stereochemistry 81
 - stoichiometry 109
 - structure
 - transformation of name 89
 - structure–activity relationship, *see* SAR
 - structure-based computer-aided drug
 - design 3
 - structure-based ligand profiling 259
 - structure-based pharmacophore
 - profiling 260
 - superligand 19
 - superpathway
 - distinction between pathway and superpathway 111
 - SuperTarget 39
 - support vector machine (SVM) 30, 235
 - systems biology approach 289ff
 - Systems Biology Markup Language (SBML) 119
 - systems response profile (SRP) 294ff
 - systems response profile calculation 296
- t**
- TarFisDock 259
 - target fishing 257ff
 - data mining 257ff
 - TargetDB archive 3
 - Targets & Pathways Knowledge Area 221ff
 - Therapeutic Targets Database 39
 - Thomson Reuters IntegritySM 213ff
 - in industry and academia 227
 - topological features 64
 - trait loci 272
 - transcription module 276
 - Triangular Heat Map 176
- u**
- Undesirable Compounds in Hit Lists 136
 - UniProtKB/Swiss-Prot databases 123
 - Unique Compound and Spectra Database (UCSD) 75ff, 78, 89
 - upper-level ontology 60
 - USAN (United States Adopted Name) 216
- w**
- Web Ontology Language (OWL) 58
 - WikiPathways 120
- z**
- ZINC database 39

