

Index

Note: Page numbers followed by “f” and “t” refer to figures and tables, respectively.

A

- Ab initio
 - annotations, 61
 - folding, 193
 - method, 29, 182–183
 - modeling, 192
- ABCPred server, 322
- ABi/SOLiD sequencing, 95
- Absorption, disposal, metabolism, excretion, and transportation (ADMET), 226, 228, 233–234, 254–255, 269, 307–308
 - analysis, 226
 - computational approaches for ADMET prediction, 311–313
- Absorption, distribution, metabolism, and excretion (ADME), 262–263
 - ADME-Tox
 - analysis, 282
 - predictor, 270
- Absorption distribution, metabolism, excretion, and toxicity. *See* Absorption, disposal, metabolism, excretion, and transportation (ADMET)
- Abundance estimation, 99
- Accelrys’ HipHop algorithm in Discovery Studio, 274
- Accepted mutations, 37
- Activation domain (AD), 390
- Activity flow language, 412
- AdaBoost, 199
 - algorithm, 463
 - classifier, 445
- Adaptation of fields for molecular comparison (AFMoC), 244
- Adapter
 - removal, 110
 - sequences, 110
 - trimming, 52
- Adaptor and quality trimming, 98
- Adjacency matrix, 373
- Adjuvant selection, 325–326
- Advanced genomic techniques, 362–363
- Ag-presenting cells (APCs), 318
- Agile Protein Interaction DataAnalyze (APID), 385
- AgriGO, 146
- Albeit distance phylogeny methods, 44
- Algorithms, 3, 35, 49–50, 352–353
 - algorithm-based tools, 220
- AlgPred 2.0, 326–327
- Alignment score, 34
- Alignment-based method, 375
- Alignment-free method, 375
- Allele-specific expression (ASE), 111, 120
- Allele-Specific Expression analysis in Population (ASEP), 120
- AllerTOP version 2.0, 326–327
- ALLPATHS-LG, 52
- Alpha diversity (α -diversity), 170–171
- Alzheimer’s disease, 215
- Amazon, 425
- AMBER, 180
 - classical force, 192–193
- Ambient ionization inductively coupled plasma ionization (ICP), 391
- AmiGO, 149
- Amino acids, 177–178, 189, 223
 - sequences, 197–198
- AMOS, 51
- Amplicon sequencing, 161
- Amplification
 - by utilizing labeled microsatellite primers, 134
 - by utilizing M13-tailed microsatellite primers and M13-labeled probes, 134
 - by utilizing unlabeled microsatellite primers, 134
- Amplified fragment length polymorphism (AFLP), 131
- Analogous traits or characters, 67
- Animal science, 342
- Annealing algorithm, 237
- Annotation, 28–29
 - of gene/protein structure and function, 467–468
 - package, 149
 - of unknown gene/genome, 147–149
 - AmiGO, 149
 - Blast2GO, 148
 - DEEPred, 149
 - genome assembly and annotation package, 149
 - GO enrichment analysis, 149–153, 151t
 - GO-FEAT, 149
 - GOnet, 149
 - IPRscan, 148
 - OBO-Edit, 149
 - SDNGO, 149
- ANNOVAR, 138–140
- Antibody-mediated immunity, 317–318
- Antigens (Ags), 317
 - selection and immunological databases, 318–319
- AntiJen, 28
- Applicability domain tool (AD tool), 246
- Apriori algorithm, 461, 463
- Arachidonic acid (AA), 255–256
- Arachne assembler, 51
- AraCyc, 393
- Area under PR curve (AUPR), 200
- Area under the curve (AUC), 276–277
- Arg267 residue, 184
- Arg367 residue, 184
- Armour, 121
- ArrayExpress, 26
- Artificial intelligence technique (AI technique), 227–228
- Artificial neural network (ANN), 9, 245, 323–324, 452
- Asn519 residue, 184
- Assembler technology, 50–51
- Assembly, 49–50
- Association, 459
- ASTRO-FOLD, 194
- Atlas of Protein Sequences and Structures, 18
- ATLAS tools, 185
- Attribute, 17
- AUGUSTUS, 61
- Autoantigens, 319
- AutoDock, 254–255, 292
 - AutoDock 4.2, 326
 - AutoDock Bias with Solvent Sites approach, 280–281
 - AutoDock Vina tool, 211–212, 292, 326
 - tool, 211–212
- Automated arrays to microarray technology, 92
- Average of shortest path length, 374–375
- Average path length, 381–382
- Azaindole derivatives, 277–278

B

- B lymphocytes, 318
- B-cell epitope (BCE), 318
 - in silico* method for BCE prediction, 319–322
 - conformational B-cell epitopes prediction, 320
 - linear B-cell epitopes prediction, 321–322
- Backing ways, 237

- Backtranseq, 61
Bankit, 19
Base calling accuracy, 109
Base quality score (BQS), 162
Basic Local Alignment Search Tool (BLAST), 8, 38–40, 61–62, 178–179, 467
Blast+, 62
BLAT *vs.*, 39–40
gapped extension step, 39
seeding step, 39
types, 39, 39*t*
ungapped extension step, 39
Bayes' theorem, 83, 198
Bayesian classifier, 452
Bayesian phylogeny, 83–84
BBB. *See* Blood–brain barrier (BBB)
BcePred server, 321
BepiPred 1.0 tool, 321–322
Betweenness centrality (BC), 373–375
Bibliographic databases, 26
Big Data, 7
in biology, 425–426
several International big data resources, 427*t*
in drug repurposing and systems pharmacology, 434–435
network-based approaches for systems pharmacology, 435
limitations in big data analysis, 437–438
BiGG Model, 357
Binary Alignment Map (BAM), 110
Binary tree, 70
BIND. *See* Biomolecular Interaction Network Database (BIND)
Binding affinity (BA), 323–324
Binding Estimation After Refinement (BEAR), 298–299
Binding site, 211, 219–220, 291
prediction, 185
Binding-free energy calculations, 299–300
binding-free-energy calculation-MM-PBSA, 227
Binning tools in metagenomics, 164–166, 165*t*
statistical analysis, 166
Bioactive conformation, 236
Bioavailability, 256
Biochemical genetic and genomic/biochemical genetic and genomic models (BiGG), 394
Biochips, 91
BioCyc, 340, 385
Database Collection, 393
Bioinformatics, 1, 3–4, 17, 50–51, 161, 305, 337, 457
aim of, 7
application of, 8–11, 9*f*
challenges in, 8
future perspective, 11
historical background, 5–7
important and popular database resources for, 2*t*
method, 221–223
needs, 5
recent development in field of, 7–8
research area in bioinformatics with deep learning, 453
research areas and tools, 466–469
annotation of gene/protein structure and function, 467–468
biological systems modeling and network analysis, 468–469
expressed sequence tag analysis, 469
gene and protein expression analysis, 468
medical and health data analysis and clinical decision support system, 469
microRNA and target prediction, 469
mutation and disease association study, 468
protein structure prediction, docking, and protein–protein interaction analysis, 468
sequence searching, comparison, and evolutionary analysis, 467
scientific discipline and support systems for, 5, 6*f*
software, 4*t*
supervised machine learning in, 444–445
translational, 313
unsupervised learning in, 452
Biological activity data, 242–243
Biological big data, 7
cellular metabolome, 430, 431*t*
different other databases, 432
drug and chemical compound databases, 432
genomics and transcriptomics resources, 430
protein–protein interaction databases, 430–431, 431*t*
proteomics resources, 430, 431*t*
resources, 429–432
Biological data
availability, 463–464
in computer-readable form, 464
cleaning, 464
dimensionality, 464–465
evolution, 463–466
knowledge discovery, 466
mining, 457
data mining algorithms, 462–463
quality, 464
Biological databases, 18, 20*t*
composite database, 21–22
for data submission, 99
database searching and annotation, 28–29
object-oriented databases, 17–19
relational databases, 17
secondary database, 22–23
sequence databases, 19–21
specialized database, 24–28
structural databases, 23–24
Biological function, 338
Biological General Repository for Interaction Datasets (BioGRID), 385
Biological molecules, 425
Biological networks, 367–369, 381
analysis, 370–373, 468–469
adjacency matrices, 373
graph theory, 371–373
applications of network biology, 376
biological significance of network motifs, 375–376
construction, 368–369
based network construction based on gene association/correlation, 369
network construction based on protein interaction, 368
limitations and challenges, 376
motifs and patterns, 369–370
motif discovery and counting, 369–370
properties, 381–382
date and party hub, 382
degree, 382
network motifs, 382
path, average path length, and diameter, 381–382
scale free, 382
small world network, 382
topological parameters, 373–375
types, 382–387
cellular signaling network, 385–386
ecological networks, 383
GCN, 387
gene (genetic) regulatory network, 383–384
metabolic networks, 385
PPIN, 384–385
Biological Objective Solution Search method, 360–361
Biological process (BP), 145, 147
Biological sequence analysis, 33
application of sequence alignment, 44
BLAST, 38–40
multiple sequence alignment, 40–42
phylogenetic analysis, 42–44
scoring matrices, 36–38
sequence alignments, 34–36
Biological significance of network motifs, 375–376
Biological systems, 367
modeling, 468–469
Biological units, 409
Biological-oriented synthesis (BOS), 259–261
Biomass reaction equation, 346
Biome, 381
BioModels, 415
Biomolecular Interaction Network Database (BIND), 385, 397
Biomolecules, maximizing production of, 358–359
BioNetStat, 396
BIOSIS Previews, 26
BioStudies database, 26
Bipartite, 433–434
graph, 434
network analysis, 435
BitSeq12, 388
BLAST-like alignment tool (BLAT), 39–40
Blast2GO tool, 62, 99, 148
BlastKOALA, 62
BLOcks SUBstitution Matrix (BLOSUM matrices), 33, 36–37
BLOSUM62, 37
BLOSUM80, 37

- Blood–brain barrier (BBB), 307–308
 Bootstrapping, 44, 85
 Bowtie, 54, 136–138
 Bowtie2, 33, 54, 99
 Branch, 70
 Braunschweig Enzyme Database (BRENDA), 27
 Burrow–Wheeler Transform, 99
 Burrows–Wheeler Aligner (BWA), 54, 136–138
 Burrows–Wheeler Aligner’s Smith–Waterman Alignment (BWA-SW), 33
- C**
 C-terminal activation domain, 390
 C2H system. *See Candida albicans* two-hybrid system (C2H system)
 C4.5 algorithm, 462
Candida albicans two-hybrid system (C2H system), 390–391
 Cap3 genome assembler, 50–51
 Capillary electrophoresis (CG), 391
 Capillary electrophoresis, 92
 Carbon flux analysis, 337, 352–353
 algorithm, 352–353
 Carbon labeling experiment (CLE), 348–349
 Carcinogenic potency database (CPDB), 311
 Cascade SVM (C-SVM), 323–324
 CastP tool, 185, 211
 Cavity prediction, 211
 Celera assembler, 51, 53–54
 Cell line–cell line networks, 432–433
 Cell membrane, 307–308
 Cell-penetrating peptides (CPPs), 327–328
 CellDesigner, 415
 CellNetAnalyzer tool, 357
 Cellular components (CCs), 146–147
 Cellular metabolome, 430, 431*t*
 Cellular networks, 420
 Cellular signaling networks, 381, 385–386
 Centrality metrics, 435
 Chain termination method, 94, 105
 Character-based methods, 43, 74
 Character-based phylogeny, 68–69
 CHARMM classical force, 192–193
 CHARMS, 180
 ChEMBL database, 211
 ChemDB, 28
 ChemDraw, 312–313
 Chemical compound databases, 432, 432*t*
 Chemical degradation method.
 See Maxam–Gilbert method
 Chemical diversity of natural products, 315
 Chemical Entities of Biological Interest, 340
 Chemical ionization (CI), 391
 ChemSpider, 340
 ChemTree tool, 311
 Chromatin, 428
 Chromatin immunoprecipitation (ChIP), 384
 ChIP-seq, 389, 428, 433, 452–453
 Chromatographic method, 337–339
 Chromosomes, 467
Cis-eQTLs, 120
 Clade, 71
 Cladistics, 68–69
 Cladogram, 43, 68, 73
 Class, Architecture, Topology, and Homologous Superfamily (CATH Superfamily), 18–19, 24
 CATH-Gene3D database, 464
 database, 464
 Classification algorithms, 458, 461
 CLC Genomics Workbench, 122
 Clinical decision support system, 469
 ClusPro server, 211–212
 Clustal2, 41
 ClustalV, 41
 ClustalW, 41–42
 guide tree construction, 42
 pairwise alignment, 41–42
 progressive alignment, 42
 workflow, 42*f*
 ClustalX, 41
 Cluster centroid, 450
 Clustering, 450–451, 459–460
 algorithms, 461
 databases, 24–25
 density-based, 450–451
 distribution-based, 451
 fuzzy, 451
 K-means, 450
 method, 236
 Clustering coefficient (CC), 375
 clusterProfiler, 151
 Clusters of Orthologous Groups (COGs), 24–25, 62, 168–170
 CmaA1, 281
 COACH tool, 211
 COBRApy tool, 357
 Coding sequences (CDS), 20
 CollecTF, 384
 COLOMBOS, 384
 Combinatorial chemistry, 219
 CoMFA, 243
 results, 244–245
 Comparative analysis, 115–118
 Comparative modeling. *See* Homology modeling
 Comparative molecular moment analysis (CoMMA), 244
 Comparative molecular surface analysis (CoMSA), 244
 Comparative residue interaction analysis, 244
 Compare web resource, 122
 Complementary deoxyribonucleic acid (cDNA), 91, 105
 cDNA-based microarray, 92–93
 collection and analysis of microarray data, 93
 collection of samples, 92
 creation of labeled target DNA/cDNA, 93
 extraction and isolation of sample, 93
 hybridization, 93
 cloning to microarray technology, 92
 library, 100
 sequences, 469
 Complementary determining region 3 (CDR3), 325
 Complex Pathway Simulator (COPASI), 416–418
 Composite data, 463
 Composite database, 21–22
 nonredundant database, 22
 UniProt, 21–22
 Computational approaches, 209–214
 for ADMET prediction, 311–313
 binding site and cavity prediction, 211
 computational ligand designing and searching, 211
 lead optimization, 214
 MDS, 212–213
 molecular docking, 211–212
 molecular modeling, 209–210
 pharmacophore modeling, 211
 QSAR, 213–214
 Computational biology, 3–4
 Computational drug repurposing, 314
 Computational lead
 advantages of computational lead designing, 263
 optimization, 263
 Computational model(ing), 410
 approaches, 410–411
 of pathway, 409–411
 type, 410
 Computational molecular phylogeny
 application of molecular phylogeny, 86–87
 cladistics and systematics, 68–69
 convergent and divergent evolution, 67–68
 estimating reliability of phylogenetic tree, 84–85
 phylogenetic tools, 85–86
 phylogenetic trees terminology, 69–73
 tree construction methods, 74–84
 Computational resources, 220
 Computational systems biology, 468–469
 Computational technology, 159, 409
 Computer-aided drug design (CADD), 207–209, 214, 253, 269, 282–284
 future perspectives, 263–264
 LBDD, 209
 methods, 291
 principles, 254–256
 SBDD, 207–209
 techniques used in lead discovery and optimization, 254*f*
 Computer-aided drug discovery, 315
 Computer-aided vaccine development example, 328–329
 Conformational B-cell epitopes prediction, 320
 Congenital hyperinsulinism (CH), 396
 Connectivity. *See* Degree
 Constraint Based Reconstruction and Analysis (COBRA), 211, 346–347, 356
 Constraint types, 349–350
 environmental constraints, 350
 physico-chemical constraints, 349
 topobiological constraints, 350
 Contigs, 49
 Convergent evolution, 67–68

- Convex function, 448
 Convolutional neural networks (CNNs), 197, 295–297, 452
 Core modeling, 179
 Cost function, 447–448
 CRISPR arrays, 7
 CRISPR-Cas systems, 7
 Critical assessment of Protein Structure Prediction (CASP), 209
 CASP6, 178
 proteins, 194
 Cross-transferability of SSR markers, 134–135
 Crystallography (X-ray), 189
 CTLPred method, 325
 Curse of dimensionality, 464–465
 Cutadapt tool, 52, 98
 Cytochrome P 450 (CYP), 309
 Cytometry by time of flight (CyTOF), 392
 Cytoscape software, 415, 435
- D**
- D-Structure quality assessment protocol, 191
 DamID, 384
Danio rerio. *See* Zebrafish fish (*Danio rerio*)
 Data, 463
 cleaning, 464
 preprocessing, 52
 quality control of raw sequencing data, 52
 trimming and filtering of raw reads, 52
 quality, 59, 376
 transformation, 464
 type in systems biology, 426–429
 binding sites and domains, 429
 biological sequences, 428
 gene expression, 428
 mass spectroscopy, 429
 metabolic pathways, 429
 molecular structure, 428
 protein–protein interaction, 429
- Data mining, 457
 applications, 457–458
 bioinformatics research areas and tools, 466–469
 biological, 457
 biological data evolution and related issues, 463–466
 biological data availability, 463–464
 feature selection technique in, 461–462
 limitations, 469–470
 major data mining algorithms applicable to biological data, 462–463
 C4.5 algorithm, 462
 K-means algorithm, 462–463
 support vector machine, 463
 process, 458–461
 association, 459
 case studies using WEKA, 461
 classification, 458
 clustering, 459–460
 description and visualization, 460–461
 estimation, 458–459
 prediction, 459
 tool, 460–461
- Data processing in metabolomics, 339–341
 challenges and affecting factors, 341
 limitations of NMR and MS methods, 340
 nuclear magnetic resonance spectroscopy, 339
 recent advances in MS-and NMR-based metabolomics, 341
 dynamic nuclear polarization, 341
 workflow of MS-based metabolomics, 339–340
- Database
 searching and annotation, 28–29
 annotation, 28–29
 Entrez, 28
 SRS, 28
 systems, 457–458
- Database for Annotation, Visualization, and Integrated Discovery (DAVID), 150
 Database Management Systems (DBMS), 17
 Database of Interacting Proteins (DIP), 26, 385
 Datasets, 159, 461–462
 Date and party hub, 382
 De Bruijn graph assemblers, 51–52
 De novo
 assembly method, 49, 52–54, 98
 modeling, 192
 pathway, 240
 sequencing, 52–53, 105
 transcriptome assembly, 112
- Decaprenyl-phosphoryl-D-ribose 2-epimerase (DprE), 277–278
 DprE1 inhibitors for tuberculosis, 277–278
- Decision tree, 199, 445, 452, 459
 Decoy set, 272
 Deep learning (DL), 197, 199, 295
 algorithms, 452–453
 bioinformatics with, 453
 Deep mRNA sequencing, 388
 Deep semantic text representation (DeepText2GO), 199
 DEEPred, 149
 Degradome sequencing, 119
 Degree, 382
 Delivery methodology, 327–328
 Dendrogram, 43, 73
 Density-based clustering, 450–451
 Deoxyribonucleic acid (DNA), 5–6, 159, 189, 428
 arrays, 91
 chips, 91
 DNA-seq, 452–453
 fragments, 50
 isolation, 159–160
 microarrays, 91, 457
 quality, 58–59
 regulatory elements, 44
 sequencing, 94–96, 452–453
 first generation, 94–95
 second generation, 95
 third generation, 96
- Depression, 342–343
 Descriptors, 279
 DESeq2, 99, 116
 Diameter, 381–382
- Diamond, 62
 Differential expression, 115–118
 Differentially expressed genes (DEGs), 99, 115–116
 Direct infusion, 391
 Direct insertion probe (DIP), 391
 Direct ionization, 391
 Direct neighbor score (DN score), 397–398
 Directed graphs, 372
 Directory of useful decoys (DUD), 237, 272
 Discovery Studio, 254–255, 273
 interaction generation protocol, 275
 Disease, 253
 disease–disease networks, 432–433
 disease–gene networks, 435
 mutation and disease association study, 468
 Disease Ontology, 153–154
 Disease-Causing Gene Finder (DIGNiFI), 397–398
 application of DIGNiFI to rare disorders, 397–398
 DisGeNET database, 153–154
 Distance-based methods, 43, 74
 distant eQTLs, 120
 Distribution-based clustering, 451, 452f
 DIVEIN, 85–86
 Divergent evolution, 67–68
 Diversity-oriented synthesis (DOS), 259–261
 DNA Data Bank of Japan (DDBJ), 18, 20, 99
 DNA-binding domain (DBD), 390, 429
 DNase sequence, 452–453
 DOCK
 program, 292
 tool, 211–212
 Docking, 468
 algorithms, 224–225
 methodologies, 294–295
 docking tools and features, 297, 297f
 flexible docking, 295
 fragment docking, 296
 machine learning in docking, 296–297
 semiflexible docking, 295
 virtual screening of high-throughput docking, 295
 procedure, 239
 Domains, 178
Drosophila melanogaster. *See* Fruit fly (*Drosophila melanogaster*)
 Drug design(ing), 269–270, 305, 418–419
 computational approaches in, 207
 CADD, 207–209
 computational approaches, 209–214
 limitations, 214
 recent trends in drug designing, 214–215
 Drug repositioning. *See* Drug–repurposing
 Drug-set enrichment analysis (DSEA), 433
 Drug(s), 219
 absorption, 308
 candidates, 305
 chemical diversity of natural products, 315
 computational approaches for ADMET prediction, 311–313
 drug repurposing, 313–314
 pharmacodynamics, 311

pharmacokinetics, 307–310
 postgenomic era and drug discovery,
 306–307
 translational bioinformatics, 313
 databases, 432, 432t
 development(al)
 processes, 207
 studies, 229
 discovery, 207, 219, 245, 291
 postgenomic era and, 306–307
 distribution, 308–309
 drug–drug interaction, 310
 drug–drug networks, 432–433
 metabolism, 309–310
 repurposing, 214–215, 313–314
 benefits, 313
 big data in, 434–435
 computational, 314
 pharmacogenomics role in precision
 medicine, 314–315
 target, 209
 maneuvers, 305
 toxicity, 311
 DrugBank, 28
 DSSTox tool, 311
 Duplicate sequences, 163
 Dynamic flux balance analysis (DFBA),
 362–363
 Dynamic nuclear polarization (DNP), 341
 Dynamic programming (DP), 33–34
 Dynamic simulations, 348
 Dynamic Vaxign analysis, 319
 Dynophore approach, 280

E

E-Dragon, 312–313
 E-pharmacophore for virtual screening of drug
 molecules, 281
e-value, 178–179
 EasyGO, 152
 EasyModeller 4.0, 326
 EBseq, 99
 ECEPP, 180
 EcoCyc, 385, 393
 Ecological network analysis (ENA), 383
 Ecological networks, 381, 383
 Ecology, 383
 Edenaare, 53–54
 Edge, 70
 length, 72
 Edge-weighted equation, 397
 EdgeR, 99, 116
 Electron impact (EI), 391
 Electron microscopy, 177, 428
 Electron Microscopy Data Bank entries, 189
 Electrospray ionization (ESI), 391
 Elementary metabolite units (EMU), 357
 Embedded method, 198
 EMBL-Bank, 19
 EMBL-European Bioinformatics Institute
 (EMBL-EBI), 35
 EMBOSS Matcher, 35
 EMBOSS Stretcher, 35

eMOTIF, 18–19
 Empirical energy scores, 180–181
 Empirical scoring functions, 294
 Enamine REAL database, 211
 Encyclopaedia of DNA Elements (ENCODE),
 433
 Endogenous antigens, 319
 Endogenous siRNAs, 113–114
 Endometabolome, 338
 Energy
 energy-based methods, 180
 functions of protein, 192–197
 knowledge-based energy function and its
 application, 194–197
 physics-based energy function and its
 applications, 192–194
 minimization software packages, 180
 Enrichment analysis, 116, 145
 Enrichment factor (EF), 276
 Enrichr, 151
 Ensembl, 62, 112–113
 Ensemble methods, 199
 Entity relationship language, 412
 Entrez database, 28, 467
 Environmental constraints, 350
 Enzyme Commission (EC), 393–394
 Enzyme databases, 27, 393–394
 Enzyme Nomenclature database, 27
 Epicentre, 389–390
 EPISUITE, 312–313
 Epitope-based vaccines, 317–318, 325
 ERATO Kitano Systems Biology Project, 412
 ERGO tool, 395
 ERRAT, 184
 Erythromycin, 418
 Escher-FBA tool, 357
Escherichia coli, 337, 418
 Essential genes identification, 358
 Estimation, 458–459
 European Nucleotide Archive (ENA), 19
 Evolutionary analysis, 467
 Evolutionary genealogy of genes:nonsupervised
 orthologous groups (eggNOGs), 168–170
 Evolutionary studies, 44
 Evolutionary-related proteins, 177–178
 Exogenous antigens, 318
 Exome, 367
 sequencing, 388–389
 Exometabolome, 338
 Exonerate, 61
 ExplorEnz, 394
 Expressed sequence tag analysis (EST), 467,
 469
 Expression databases, 26
 Expression quantitative trait loci (eQTLs),
 120–121
 Extensible Markup Language (XML), 412
 Extracellular Ags. *See* Exogenous antigens
 Extraction, 339–340

F

Facebook, 425
 False discovery rate (FDR), 99

FASTA Clustal, 8
 FASTA tool, 467
 FASTQ format, 97–98, 162
 FastQC, 52
 FASTX-Toolkit, 52
 Feature learning, 199
 Feature selection algorithms (FSAs), 198
 Feature selection technique in data mining,
 461–462
 objective, 462
 Feature vector, 197–198
 FeatureCounts, 99
 FGAP tool, 56
 Filter method, 198
 Filtering programs, 319
 First generation DNA sequencing, 94–95
 Fitch-Margoliash method (FM), 43
 Flexible docking, 224–225, 295
 FlexX tool, 211–212
 Flow cytometry (FC), 392
 Flux balance analysis (FBA), 337, 343–344,
 395–396, 419
 applications, 358–360
 analysis of GSMNs, 359
 flux coupling analyses, 359
 identifying “essential” and “synthetic
 lethal” genes, 358
 maximizing production of biomolecules,
 358–359
 predicting novel regulatory mechanisms,
 360
 refinement of metabolic networks, 360
 applications industrially important
 microorganisms, 361–363
Escherichia coli, 362–363
Lactococcus lactis, 361
Saccharomyces cerevisiae, 361–362
 challenges and future perspectives, 360–361
 biologically relevant objective functions,
 360–361
 impact of high-throughput experiments,
 361
 limitations, 360
 different types of flux balance analysis at
 different conditions, 353–355
 analysis of flux under isotopic nonsteady
 state, 354
 analysis of flux under metabolic and
 isotopic nonsteady state, 355
 flux analysis under metabolic and isotopic
 steady state, 354
 flux analysis under metabolic nonsteady
 state, 355
 flux analysis under metabolic steady-state,
 353–354
 recent advances in metabolic flux analysis,
 355
 steps in, 351
 tools, 355–358
 CellNetAnalyzer, 357
 COBRA Toolbox, 356
 Escher-FBA, 357
 MetaboAnalyst 4.0, 356
 MetaFluxNet, 357–358

- Flux balance analysis (FBA) (*Continued*)
 OpenFlux, 357
 OptFlux, 356–357
 OptGene, 355
 OptKnock, 355
 OptStrain, 356
 SBRT, 357
- Flux coupling analyses, 359
- Fold classification scheme of protein, 190–191
- Fold recognition. *See* Threading, recognition
- Food and herbal medicines, 343
- Forbidden spaces, 271
- Force field (FF), 192–193
 scoring functions, 294
- Fourier-transform ion cyclotron resonance, 391
- Fourth-order Runge–Kutta algorithm, 414
- Fragment docking, 296
- Fragment-based drug design, 283
- Free energy landscape (FEL), 297–298
- Free modeling (FM), 192
- Free-Wilson models, 233–234
- Fruit fly (*Drosophila melanogaster*), 51, 342
- Functional annotation, 62, 398. *See also*
 Structural annotation
 COGs, 62
 gene ontology, 62
 KEGG, 62
 STRING, 62
- Functional assignment, 147
- Functional modules identification, 399
- Functionally oriented synthesis (FOS),
 259–261
- FungalRV database, 319
- Fusion vaccine construct, 3D model and
 validation of, 326
- Fuzzy clustering, 451
- G**
- G-protein-coupled receptors (GPCRs), 279
- g:Profiler, 150–151
- GAMS software, 346–347
- Gapcloser tool, 56
- GapFill algorithm, 347
- GapFiller tool, 56
- Gas chromatography (GC), 337, 391
- GBPM tool, 239
- GBrowse, 61
- GenBank database, 19, 464
- Gene, 428, 467
 analysis, 468
 annotation of gene/protein structure and
 function, 467–468
 arrays, 91
 chips, 91
 density on chromosome, 467
 expression, 26, 428
 gene (genetic) regulatory network, 383–384
 gene-based SSRs, 133
 gene-set library, 433–434
 gene–gene associations, 381
 hubs, 383–384
 of interest, 99
 microarray technology, 91
 network construction based on gene
 association/correlation, 369
 prediction, 60
- Gene Associations File (GAF), 146
- Gene coexpression network (GCN), 381, 387
- Gene Expression Omnibus (GEO), 26, 99, 384
- Gene Ontology (GO), 60, 62, 116, 145–146,
 393
 annotation of unknown gene/genome,
 147–149
 applications, 153–154
 database, 145–146
 enrichment analysis, 149–153, 151*t*
 clusterProfiler, 151
 DAVID, 150
 EasyGO, 152
 Enrichr, 151
 g:Profiler, 150–151
 GOAT, 153
 GOEAST, 153
 GOLEM, 153
 GOrilla, 152
 GOssTo, 153
 NaviGO, 153
 PANTHER, 150
 PoGo, 152
 QuickGo, 152
 REVIGO, 152
 ShinyGO, 152
 ToppGene, 151
 ViSEAGO, 152
 WEGO, 152
 future prospects, 154
 GO–based classification, 146–147
 resources, 146
- Gene Ontology Annotation Tool (GOAT), 153
- Gene Ontology enRiChment anaLysis and
 visuaLizAtion tool (GOrilla), 152
- Gene Ontology Enrichment Analysis Software
 Toolkit (GOEAST), 153
- Gene Ontology Functional Enrichment
 Annotation Tool (GO-FEAT), 149
- Gene Ontology Local Exploration Map
 (GOLEM), 153
- Gene Ontology semantic similarity Tool
 (GOssTo), 153
- Gene regulatory networks (GRN), 367, 381
- Gene set enrichment analysis (GSEA), 148
- Gene Structure Display Server, 61
- Gene3D, 24
- GeneChips, 362–363
- General Pathway Simulator (GEPASI), 416
- Generic version of PLS (G/PLS), 244
- Genetic algorithms (GA), 194, 292
- Genetic diseases, 468
- Genetic polymorphism, 308
- Genome, 381, 467
 finishing/polishing, 56–57
 genome-guided transcriptome assembly,
 111–112
 genome-wide bisulfite sequencing, 389–390
 genome-wide data sources, 306
 sequencing, 362–363
- Genome Analyzer (GA), 106
- Genome Analyzer Toolkit (GATK), 120
- Genome annotation, 60–62, 452, 467
 functional annotation, 62
 structural annotation, 61
- Genome assembly, 49, 56*t*, 149
 algorithms, 51–52
 application and future prospects, 62
 approaches, 52–56
 assembler technology, 50–51
 data preprocessing, 52
 mathematical calculation for depth/coverage,
 60
 pitfall in, 58–60
 reassemble genome after sequencing, 50
 tools and software for, 56–58
 assembly quality assessment and
 validation, 58
 genome finishing/polishing, 56–57
 tools in metagenomics, 164, 165*t*
- Genome-scale metabolic networks (GSMNs),
 359
 analysis, 359
- Genomics, 44, 381, 457
 resources, 430
- Genomics England project, 426
- GenPept, 20
- GFinisher tool, 56
- Gilde tool, 211–212
- Glide, 254–255
- Glimmer, 61
- Global sequence alignment, 35
- GLOBI database, 383
- Globular domain regions, 178
- Glow discharge ionization (GD ionization), 391
- Glycopeptides, 23
- GNU Linear Programming Kit, 357
- GOLD tool, 211–212
- GOnet, 149
- Gradient Boosting, 199
- Gradient descent, 448
- Graph mining techniques, 468
- Graph theory, 371–373, 381
- Graphlet correlation distance method, 375
- GreenGenes, 161
 normalized, 161
- GRID, 244
- GRids of phArmacophore Interaction fieLds
 approach (*GRAIL* approach), 280–281
- GROMOS classical force, 180, 192–193
- GROningen MACHine for Chemical
 Simulations (GROMACS), 326
- GSnap, 33
- Guanine–cytosine content (GC content), 107,
 163
 content per sequence, 110
- Guner-Henry scoring method (GH scoring
 method), 237
- H**
- ¹H NMR-based metabolomics, 342–343
- HADDOCK, 428
- HaloPlex, 388–389
- Hamiltonian path, 51

- HDOCK server, 211–212
Herbal medicines, food and, 343
Herpes simplex virus (HSV), 397
Herpes simplex virus encephalitis (HSE), 397
 application of human gene connectome to, 397
Heterochromatic small-interfering RNAs (hc-siRNAs), 113–114
Heterozygosities, 134
Hexadecahydro-1*H*-cyclopenta[*a*]phenanthrene framework (HHCPF), 279
Hidden Markov model (HMM), 9, 23, 61, 394, 459
Hierarchical algorithm, 194
High Throughput Sequencing Technologies, 388
High-performance computers (HPCs), 193
High-performance liquid chromatography (HPLC), 391
High-scoring segment pair (HSP), 39
High-throughput screening/sequencing (HTS), 92, 219, 233, 381
High-throughput X-ray technique, 190
HipHop algorithms, 237
HISAT2, 99
Hits distribution
 of functional group, 168–170
 of metagenome from database sources, 168
HMM homologous sequence, 29
HMMER tool, 467
 HMMER3 tools, 190
Holistic approach, 425
Hologram quantitative structure activity relationship (HQSAR), 248
Holotomography (HT), 392–393
Homologous characters, 67
Homologous superfamily assignment, 190
Homologous template structure, 178
Homology modeling, 178–180, 221–223
 loop modeling, 180
 model accuracy and validation, 180
 protein structure prediction, 179*f*
 SCR or core modeling, 179
 side chain modeling, 180
Homology Protein Structure Database (CATH), 190
HTSeq, 99
Hubs, 374
Human gene connectome (HGC), 397
 application of HGC to HSE, 397
Human Genome Project, 6–7, 467
Human immunodeficiency virus-1-inhibiting FDA-approved drugs (HIV-1-inhibiting FDA-approved drugs), 220
Human Metabolome Database (HMDB), 338, 340
Human Protein Reference Database (HPRD), 385, 397
HumanCyc, 393
Humoral immunity, 317–318
Hybrid assembly approach, 54
Hybrid models, 411
Hybridization-based approaches, 388–389
Hydrogen bond acceptor (HBA), 269
Hydrogen bond donor (HBD), 269
5-hydroperoxy eicosatetraenoic acid (5-HpETE), 255–256
Hypogen algorithms, 237, 274
Hyponastic leaf-111 (HYL1), 114
Hypothesis, 447–450
Hypothesis-driven metabolomics, 338
- I**
Ideal curve, 277
Ideogram, 467
Illumina, 106
 HiSeq, 54
 Illumina/Solexa sequencing, 95
 Tru-seq SLR technology, 108
IMAGE tool, 56
Immune Epitope Database and Analysis Resource (IEDB), 319
Immune system modulation, 282
Immuno Polymorphism Database, 325
ImMunoGeneTics project (IMGT), 325
Immunological databases, 318–319
In silico biology, 3–4
In silico method
 for B-cell epitope prediction, 319–322
 for T-cell epitope prediction, 322–325
In Vivo/In Silico Metabolites Database, 340
Information technology (IT), 1
Insertions, and deletions (INDEL), 37–38, 97
Integrated relational Enzyme database (IntEnz), 27
Inter-simple sequence repeats (ISSR), 131
Interaction databases, 26
Interactome, 381
Interfeature distances, 269
Intermediate data, 426
Internal node, 70
Internal transcribed spacers, 159–160
International Nucleotide Sequence Database Collaboration (INSDC), 19
International Union of Biochemistry and Molecular Biology (IUBMB), 27, 393–394
International Wheat Genome Sequencing Consortium, 62
InterPro, 23
InterProScan (IPRscan), 62, 148
Intron length polymorphic primer, 131
Ion Personal Genome Machine (PGM), 107
Ion Proton, 107
Ion S5, 107
 XL system, 107
ION torrent, 107
Ion torrent sequencing, 95
Ion trap (IT), 391
Ionization, 340
isomiRs, 115
Isotopic nonsteady state, flux analysis under, 354–355
Isotopic steady state, flux analysis under, 354
Iterative Threading ASSEMBLY Refinement (I-TASSER), 195, 196*f*, 209
IWDB interaction web database, 383
- J**
Jatropha healthy (JH), 435–437
Jatropha Viral (JV), 435–437
Joint Commission on Biochemical Nomenclature (JCBN), 394
Junk DNA, 112
- K**
K-means
 algorithm, 462–463
 clustering, 450
k-mers, 51–53
 value, 110
k-nearest neighbor (k-NN), 198–200, 445, 452, 458
k-tuple method, 41–42
Kernel window, 453
Kirchhoff's laws, 348
Knowledge discovery process, 466
Knowledge-based energy function and its application, 194–197
 I-TASSER, 195
 QUARK, 195–197
 ROSETTA, 195
Knowledge-based scoring functions, 294–295
Kyoto Encyclopedia of Genes and Genomes (KEGG), 60, 62, 133, 340, 385, 393–394, 415, 433
 database, 27
 KEGGtranslator, 395
Kyoto Encyclopedia of Genes and Genomes orthologous groups (KOs), 168–170
- L**
Lambda phage, 411
Lander–Waterman model, 60
Language in pathway modeling, 411–413
Lassa Virus (LASV), 321
Lead molecules, 256–259
 discovery, 256
 natural lead compounds, 256–257
 semisynthetic drugs, 259
 synthetic lead molecules, 257–258
Lead optimization, 214, 234, 256
 bonding strength and selectivity, 262
 functional group interconversion, 261–262
 organic synthetic chemistry, 259–261
 and strategies, 259–263
 structure modification, 261
 structure simplification, 261
 using thermodynamic, pharmacodynamics, and pharmacokinetic parameters, 262–263
Leaf, 69
Leave-one-out method (LOO method), 246–247, 397
Leishmania braziliensis, 321–322
Leu23 residue, 184
Library preparation, 59
 and sequencing, 96–97
Ligand-based designing approaches, 255

- Ligand-based drug designing (LBDD), 209, 233–234. *See also* Structure-based drug designing (SBDD)
- advancement in exploring 3D pharmacophore principles over above limitations, 240
 - application of
 - pharmacophore in virtual screening and de novo design, 239–240
 - QSAR/SAR, 248–250
 - development of new QSAR, 248
 - pharmacophore, 234–235
 - classifications, 238–239
 - mapping, 236–237
 - 3D fingerprints, 235–236
- Ligand-based pharmacophore (LBP), 238–239. *See also* Structure-based pharmacophore (SBP)
- first problem lies with ligand flexibility, 238
 - modeling, 269–270, 273–275
 - using Accelrys' HipHop algorithm in Discovery Studio, 274
 - using LigandScout's espresso algorithm, 274–275
 - molecular alignment in LBP modeling, 238
 - selection of training set of compounds, 239
- LigandScout, 275–276
- espresso algorithm, 274–275
- Linear B-cell epitopes prediction, 321–322
- Linear programming, 350
- Linker selection, 325–326
- Lipidomics, 367
- Lipophilicity, 308
- 5-lipoxygenase-activating protein (FLA protein), 282
- Lipoxygenases (LOXs), 255–256
- inhibition, 255–256
- Liquid chromatography (LC), 337, 391
- Live cell imaging, 392–393
- Living organisms, 337
- Local random walk algorithm (LRW algorithm), 397–398
- Local sequence alignment, 35–36
- Log2fc, 99
- Logistic regression algorithm, 319, 448–450
- Long ncRNAs (lncRNAs), 111
- Long noncoding RNAs, 113
- Loops, 180
 - centroid step, 450
 - modeling, 180
- Losartan, 255
- Lys440 residue, 184
- Lys594 residue, 184
- M**
- Machine learning (ML), 147, 197, 227–228, 241–242, 273, 317, 443, 457–458
- algorithms, 198–199, 319
 - application, 452–453
 - genome annotation, 452
 - protein structure prediction, 452–453
 - research area in bioinformatics with deep learning, 453
 - artificial intelligence components, 443f
 - and categories, 443f
 - clustering, 450–451
 - in docking, 296–297
 - feature engineering and representation, 197–198
 - feature selection, 198
 - ML-based SFs, 295
 - models implementation and evaluation, 199–200
 - generally used metrics applied for protein function prediction, 200t
 - models incorporated with pharmacophore descriptors, 278–279
 - problems
 - in bioinformatics, 453–454
 - to understand supervised learning and unsupervised learning, 446
 - in PSP, 197–200
 - regression, 446–450
 - supervised learning, 443–445, 444f
 - techniques, 324
 - unsupervised learning in bioinformatics, 452
 - unsupervised machine learning, 445–446
- Macromolecule ligand complex, 239
- Madison Metabolomics Consortium Database, 340
- Magnetic sector, 391
- MAKER, 61
- Mangal-Ecological interaction database, 383
- Many Microbe Microarrays Database (M3D), 384
- Mapping/alignment of reads onto assembly, 99
- MAPPP server, 323–324
- Marker genes, 159–160
- Mass balance equations, 348–349
- Mass cytometry, 392
- Mass spectrometry (MS), 337, 429, 466–467
- limitations, 340
 - mass spectrometry-based proteomics, 391–392
 - MS-based techniques in metabolomics, 338–339
 - workflow of MS-based metabolomics, 339–340
 - analysis and identification, 340
 - detection, 340
 - extraction, 339–340
 - ionization, 340
 - sample preparation, 339
 - separation, 340
- Massively parallel signature sequencing (MPSS), 122
- Matched molecular pair (MMP), 233–234
- Matching
 - algorithm, 292
 - symbols, 34
- Mathematica, 415
- Mathematical biology, 413
- Mathematical models, 317, 410
- Mathematical network models, 383
- MATLAB, 415
- Matrix decomposition techniques, 362–363
- Matrix-assisted laser desorption ionization (MALDI), 391
- Maxam–Gilbert method, 95
- Maximum likelihood (ML), 43
- method, 43
 - phylogeny, 82–83
- Maximum parsimony method (MP method), 43, 80–82
- MCDOCK, 292
- MedDRA, 313
- Medical and health data analysis, 469
- Medical science, 342–343
- MEGA, 85
- Messenger mRNA, 96
- Messenger RNAs (mRNAs), 105, 467
- Meta-assembly approach, 54–56
- MetaboAnalyst, 340
- MetaboAnalyst 4.0, 356
- Metabolic control analysis, 352
- data, 352
- Metabolic engineering, 418
- Metabolic flux analysis (MFA), 11, 337
- Metabolic foot printing, 338
- Metabolic networks, 381, 385
- construction and refinement
 - initial model development, 345–346
 - model improvement by experimental validation, 347
 - optimization method to model system, 346–347
 - regulatory controls and dynamic simulations, 348
 - thermodynamic constraints to metabolic model, 348
 - mass balance equations, 348–349
 - and model construction, 344–351
 - optimization, 350–351
 - refinement of, 360
 - steady state, 349
 - steps in FBA, 351
 - types of constraints, 349–350
- Metabolic nonsteady state, flux analysis under, 355
- Metabolic pathways, 429
- Metabolic steady-state, flux analysis under, 353–354
- Metabolism, 385
- Metabolite and Tandem MS Database, 340
- Metabolome, 338
- Metabolome-wide association studies (MWAS), 342
- Metabolomics, 11, 337–338, 381. *See also* Genomics
- applications, 341–343
 - animal science, 342
 - food and herbal medicines, 343
 - in industrially important microorganisms, 361–363
 - medical science, 342–343
 - microbial science, 341
 - plant science, 341–342
 - carbon flux analysis, 352–353
 - data processing in, 339–341
 - fingerprinting, 338

- metabolic control analysis, 352
 - and model construction, 344–351
 - MS-and NMR-based techniques in, 338–339
 - Metabolomics Workbench MetaboLights, 340
 - MetaCore, 340
 - MetaCyc, 340
 - Metadata, 146–147
 - MetaFluxNet tool, 357–358
 - Metagenome dataset, 166
 - Metagenomics, 105, 159–161
 - advantages, 172
 - analysis, 166
 - material, methodology, and outcome, 166–171
 - binning tools in, 164–166, 165*t*
 - challenges in, 161–162
 - chronology of principal breakthroughs in, 160*f*
 - data storage and sharing, 166
 - dataset acquire and processing, 162
 - genome assembly tools in, 164, 165*t*
 - limitations and future perspective, 172
 - QC analysis, 162–163
 - resources in, 161
 - workflow
 - for sequencing and analysis of sample, 160*f*
 - in metagenome analysis, 162
 - MetaSite tool, 311
 - MetaTIGER, 385, 394
 - Methanol–water–chloroform approach, 339–340
 - Methylation sequencing, 105
 - 5-methylcytosine (5mC), 389–390
 - Metropolis criteria, 292
 - MHC
 - class I binder prediction, 323–324
 - class II binder prediction, 324–325
 - gene diversity, 325
 - Michaelis–Menten constant, 349
 - Microarray, 91, 362–363, 387–388, 428
 - advantages of transcriptome sequencing over microarray technology, 100
 - applications of, 99
 - automated arrays and cDNA cloning to microarray technology, 92
 - evolution, 92–94
 - principle of, 92–94
 - tools and utility, 94
 - data analysis, 94
 - pre-processing of microarray data, 94
 - Microarray Experiment format, 26
 - Microbes, 159
 - Microbial science, 341
 - Microbiomics, 367
 - Microorganisms, 159
 - MicroRNA Truncation and Tailing Analysis (miTRATA), 122
 - MicroRNAs (miRNAs), 105, 114–115, 469
 - database, 27
 - miRNAconsTarget, 122
 - miRNA–mRNA interaction network, 435
 - and target prediction, 469
 - Microsatellites, 131–132
 - benefits and limitations of microsatellite markers, 132–133
 - Middle East respiratory syndrome coronavirus (MERS-COV), 322
 - Minimization of metabolic adjustment (MOMA), 350–351
 - Minimum evolution method (ME), 43
 - Minimum information about genome sequence (MIXS), 166
 - MinION, 108–109
 - mir in database, 27
 - miRBase, 27, 121
 - MISO, 388
 - Mixed integer linear programming (MILP), 194, 355
 - Model-Based Expression Index (MBEI), 94
 - Modeled protein, 180
 - ModelSEED, 395–396
 - Modulators, 282
 - MOE, 254–255
 - MOLE db (molecular descriptors database), 312–313
 - Molecular alignment in LBP modeling, 238
 - Molecular docking, 185–186, 211–212, 224–225, 291–295, 326
 - algorithms, 292–293, 293*t*
 - challenges in, 300–301
 - interaction view of uridilate kinase, 212*f*
 - knowledge-based SFs, 294–295
 - scoring functions, 294
 - Molecular dynamics simulation (MDS), 193, 212–213, 226–227, 278, 291, 297–300, 326
 - binding-free energy calculations, 299–300
 - challenges in, 300–301
 - pharmacophore methods in light of, 280–281
 - postdocking refinement, 298–299
 - Molecular evolution techniques, 177
 - Molecular functions (MFs), 146–147
 - Molecular holograms, 248
 - Molecular INTeraction database (MINT), 385
 - Molecular inversion probes (MIPs), 388–389
 - Molecular markers, 131
 - Molecular mechanics (MM), 298–299
 - Molecular mechanics-generalized born and solvent-accessible surface area solvation (MM-GBSA), 299–300
 - Molecular mechanics-Poisson–Boltzmann surface area (MM-PBSA), 299–300
 - Molecular modeling, 209–210, 219
 - Molecular networks, 368
 - Molecular phylogeny
 - application of, 86–87
 - classification, 86–87
 - identifying origin of pathogens, 87
 - species conservation, 87
 - Molecules
 - alignment of, 235
 - superimposition based on atom overlapping, 235
 - MolProbity, 184
 - Monooxygenases, 309
 - Monophyletic taxon, 71
 - Monorepeats, 133
 - Monte Carlo algorithm (MC algorithm), 193–194, 292
 - Morphological markers, 131
 - Motifs, 367
 - discovery and counting, 369–370, 371*t*
 - MouseGene Expression Database, 384
 - MtbWeb, 319
 - Multi-Task Deep Neural Network models (MT-DNN), 199
 - Multinteger linear programming, 350
 - Multilevel models, 410–411
 - Multipartite, 433–434
 - Multiple data sources, 400
 - Multiple linear regression (MLR), 245
 - Multiple sequence alignment (MSA), 6–7, 33, 40–42, 178
 - algorithm, 41
 - ClustalW, 41–42
 - need for, 40
 - tools, 178–179
 - Multiplexing PCR, 388–389
 - Multitarget drugs design, 214–215
 - Multitarget ligand approach, 282
 - Multivariate data, 429
 - Mutation and disease association study, 468
 - MycobacRV, 319
 - Mycobacterium tuberculosis*, 215, 418–419
 - Mycolic acid, 281
 - MYSAP approach, 280
- ## N
- N-terminal DNA-binding domain, 390
 - N50 statistic, 58
 - Naïve Bayes (NB), 458
 - classification, 198
 - classifier, 445
 - National Center for Biotechnology Information (NCBI), 6–7, 19, 62, 99, 161
 - PubChem, 27
 - Natural antisense small-interfering RNAs (nat-siRNAs), 114
 - Natural lead compounds, 256–257
 - Natural products (NPs), 256
 - chemical diversity of, 315
 - NaviGO, 153
 - Needleman–Wunsch algorithm, 35
 - Neighbor joining (NJ), 41, 43, 76–80
 - advantages and disadvantages of, 80
 - working example, 77–80
 - NetChop 3.1, 323–324
 - NetCTL1.2 method, 325
 - Netflix, 425
 - NetMHCcons 1.1 server, 323–324
 - NetMHCIIpan 3.1, 324–325
 - NetMHCpan 4.1 server, 323–324
 - Network Analysis Tools, 396
 - Network biology, 367, 381. *See also* Systems biology
 - applications, 376, 396–399
 - essential protein identification, 398–399
 - functional modules' identification, 399
 - in rare diseases, 396–398

- Network biology (*Continued*)
 pathway determination, 398
 protein function determination, 398
 biological networks, 381
 properties, 381–382
 types, 382–387
 challenges and future perspective, 399–400
 combination of algorithms, 400
 multiple data sources, 400
 pseudo temporal ordering, 399–400
 experimental methods in, 387–393
 ChIP-seq, 389
 deep mRNA sequencing, 388
 exome sequencing, 388–389
 flow and mass cytometry, 392
 genome-wide bisulfite sequencing, 389–390
 live cell imaging, 392–393
 mass spectrometry-based proteomics, 391–392
 microarray, 387–388
 yeast two-hybrid, 390–391
 resources for biological network-based studies, 393–394
 BiGG, 394
 BioCyc Database Collection, 393
 ENZYME, 393–394
 ExplorEnz, 394
 KEGG, 393
 metaTIGER, 394
 STRING, 394
 tools for network pathway analysis, 394–396
 BioNetStat, 396
 ERGO tool, 395
 KEGGtranslator, 395
 ModelSEED, 395–396
 Network Analysis Tools, 396
 OmicsNet, 396
 pathway tools, 395
 Network data, 429
 Network flow-based algorithm, 398
 Network generation and analysis from data sources, 432–434
 Network motifs, 367, 382
 biological significance of, 375–376
 Network-based approaches for systems pharmacology, 435
 Network-based methods, 368
 Neural networks (NN), 196–197, 199–200, 445, 449, 459
 Newbler, 53–54
 NEWLEAD software, 240
 Next-generation sequencing (NGS), 1, 17, 33, 91–92, 105, 159–160, 317, 367, 388
 data analysis, 10
 platforms and technologies, 105–109
 technology role in vaccine design, 328
 Nextera, 388–389
 NG50, 98
 NimbleGen, 388–389
 Node, 69
 degree, 374
 NOISeq, 99
 Nomenclature Committee (NC), 27, 394
 Noncoding DNA, 112
 Noncoding RNAs (ncRNAs), 96, 105
 Nonlinear activation function, 453
 Nonlinear modeling algorithms, 245
 Nonredundant database (NRDB), 22
 Nontargeted metabolomics, 338
 Norfloxacin, 255
 Novel protein-coding genes, identification of, 112–113
 Nuclear magnetic resonance (NMR), 337, 428
 limitations, 340
 NMR-based techniques in metabolomics, 338–339
 spectroscopy, 177, 189, 207–208, 219–220, 339
 Nucleic acids, 221
 Nucleotide, 467
O
 Object-oriented databases, 17–19
 biological data available to research community in single place, 18–19
 biological data in computer-readable form, 18
 stored data, 18
 Object-relational database management system, 17–18
 OBO-Edit, 149
 Observational medical outcomes partnership (OMOP), 313
 Oligonucleotide-based microarray, 93–94
 Omega algorithm, 272–274
 Omics
 data integration, 11
 techniques, 368
 OmicsNet, 396
 OneDep system, 191
 Online Chemical Modeling Environment, 312–313
 Online Mendelian Inheritance in Man (OMIM), 464
 Ontology, 146–147
 techniques, 466–467
 OpenFlux tool, 357
 Operational taxonomic units (OTUs), 69
 OPLS classical force, 192–193
 OptFlux tool, 356–357
 OptGene tool, 355
 Optimization, 350–351
 alternative objective functions, 350–351
 linear programming, 350
 method to model system, 346–347
 multiinteger linear programming, 350
 OptKnock
 algorithm, 358–359
 tool, 355
 OptStrain tool, 356
 Ordinary differential equations (ODEs), 343–344, 410, 413–414
 integration of ODEs, 414
 ORF Finder, 61
 OrfPredictor software, 133
 Orphan disorders (OD), 397
 Orthogonal descriptors, 247
 Overlap-layout-consensus (OLC), 51
 overlap-layout-consensus/string graph assemblers, 51
 Overrepresentation analysis (ORA), 150–151
 Overrepresented k-mers, 163
 Overrepresented sequences, 110, 163
 Oxford Nanopore, 96, 108–109
P
 P-glycoprotein (Pgp), 308–309
 Pacific Biosciences (PacBio), 107–108
 SMRT sequencing, 96, 108
 Padlock probes, 388–389
 Paired-end sequencing, 97
 Pairwise alignment, 41–42
 Pairwise sequence alignment (PSA), 33
 Pan-Dataset Density Analysis, 190
 Parallel algorithms, 369–370
 Parallel Analysis RNA End (PARE).
See Degradome sequencing
 Paraphyletic taxon, 71
 Partial cDNA sequences, 469
 Partial differential equations (PDEs), 410
 Partial least squares (PLS), 244
 Particle swarm optimization (PSO), 292–293
 Party hub, date and, 382
 Path, 381–382
 Pathdock, 428
 Pathogen-associated molecular patterns (PAMPs), 326
 PathoLogic program, 393
 Pathway databases, 27
 Pathway designing tools, 415
 CellDesigner, 415
 Cytoscape, 415
 Pathway determination, 398
 Pathway modeling
 applications and simulations, 418–419
 general diagram and language used in, 411–413
 Pathway simulations analysis, 413–415
 ODEs, 413–414
 stochastic simulation, 414–415
 Pathway tools, 395, 415–416
 Pathway/Genome Databases (PGDBs), 393
 Patient–patient networks, 432–433
 Pattern identification, 44
 PDBsum server, 184
 Peptide toxicity, 326–327
 Per base N content, 163
 Per base sequence content, 163
 Perl script, 435
 Pfam database, 23, 62
 Phagocytic cells, 318
Pharm-IF, 278
 Pharmaceuticals, 305
 Pharmacodynamics of drug candidates, 311
 drug toxicity, 311
 Pharmacogenomics, 10–11
 role in precision medicine, 314–315
 Pharmacokinetics of drug candidates, 307–310

- drug absorption, 308
- drug distribution, 308–309
- drug metabolism, 309–310
- Pharmacology, 307
- Pharmacophore, 234–235, 239–240, 269
 - alignment of molecules, 235
 - applications, 281–283
 - basics, 271–273
 - build pharmacophore hypothesis, 234
 - classifications, 238–239
 - LBP, 238–239
 - SBP, 239
 - computational drug design, 270*f*
 - conformational analysis with three-dimensional structures, 272–273
 - different methods of pharmacophore generation, 273–276
 - division of initial data into diverse datasets, 271–272
 - decoy set, 272
 - test set, 271
 - training set, 271
 - future perspectives, 283–284
 - mapping, 236–237
 - diverse conformation generation, 236–237
 - generation of 3D pharmacophore, 237
 - validation of, 237
 - methods in light of molecular dynamics simulations, 280–281
 - modeling, 211, 212*f*, 269–270
 - pharmacophore-based de novo design, 240
 - pharmacophore-guided drug target identification, 282–283
 - recent trends in pharmacophore generation, 278–281
 - similarity search methods, 235
 - 2D finger prints, 235
 - validation of pharmacophore models, 276–278
- Pharmmaker, 281
- Phenograms, 68–69
- Phenomics study, 419
- Phenotype–phenotype networks, 432–433
- PHrap genome assembler, 50–51
- Phylemon2.0, 85–86
- PhyloGene server, 85–86
- Phylogenetic analysis, 8–9, 34, 42–44
 - algorithms for, 43
 - terminology of phylogenetic tree, 43–44
 - types of trees, 43
- Phylogenetic systematics, 68
- Phylogenetic tools, 85–86
- Phylogenetic trees, 69
 - binary tree, 70
 - clade, 71
 - cladogram, 73
 - dendrogram, 73
 - edge, 70
 - edge length, 72
 - estimating reliability of, 84–85
 - evolutionary inference, 73–74
 - importance of shared derived characters in phylogeny, 74
 - internal node, 70
 - leaf, 69
 - monophyletic taxon, 71
 - node, 69
 - paraphyletic taxon, 71
 - phylogram, 73
 - polyphyletic taxon, 71
 - root, 70
 - rooted tree, 70
 - split, 72
 - subtree, 72
 - taxon, 69
 - terminology, 69–73
 - topology, 70
 - ultrametric tree, 72
 - unrooted tree, 70
- PHYLogeny Inference Package (PHYLIP), 85
- Phylogram, 43, 73
- Physico-chemical constraints, 349
- Physics-based energy function and applications, 192–194
 - coupling MDS with physics-based potentials, 193
 - structure refinement, 193–194
- Physics-based modeling, 192
- Phytozome, 62
- PickPocket 1.1 server, 323–324
- piRNA, 114
- Plant Ontology (PO), 146
- Plant science, 341–342
- Plant small RNA target analysis (psRNA target analysis), 435
- PMRD, 121
- Pocket v.2 tool, 239
- Point Accepted Mutation (PAM), 33, 36–37
- Polymerase chain reaction (PCR), 159, 161, 388–389
 - PCR-based markers, 131
- Polymorphism Information Content (PIC), 134
- Polyphyletic taxon, 71
- Polysaccharides, 23
- Position weight matrix (PWM), 429
- Position-specific scoring matrix, 429
- Position-specific weight matrix, 429
- Postdocking refinement, 298–299
- Postgenomic era and drug discovery, 306–307
- Posttranslational Modifications (PTM), 21
- Pre-ADMET tool, 311
- Precision medicine, pharmacogenomics role in, 314–315
- Precision-recall (PR), 200
- Predicted protein model, 185
 - evaluation and validation of, 10
- Prediction, 459
 - of pharmacokinetic properties, 279
- Prediction of Gene Ontology (PoGo), 152
- Preremunerating, 238
- Primary data, 463
- Primitive traits, 68
- Principal component analysis (PCA), 247, 297–298, 362–363
- PrinSeq tool, 98
- PRINTS database, 22–23
- Probabilistic neural network, 445
- Probabilistic regulation of metabolism (PROM), 348
- Probes, 91
- Process descriptions (PDs), 412
- PROCHECK, 10, 184, 209
- Profiling expression patterns, 115–118
 - comparative analysis and differential expression, 115–118
 - expression level calculation, 115
- Progenesis, 340
- PromethION, 108–109
- Promoter2.0, 61
- ProPred tool, 324–325
- ProPred-1 method, 325
- ProRule, 22
- PROSA web server, 184–185
- PROSITE, 18–19, 22
- Protein ANalysis THrough Evolutionary Relationships (PANTHER), 150
- Protein Data Bank (PDB), 18, 177–178, 189, 220, 464
 - repository data, 189
 - research collaboratory for structural bioinformatics, 23
- Protein Information Resources (PIR), 18, 20–21
- Protein Literature, INformation and Knowledge (ProLINK), 21
- Protein Ontology (PRO), 21
- Protein structure prediction methods (PSP methods), 177–178, 192, 452–453, 468
 - applications of, 185–186
 - binding site prediction, 185
 - mutation studies, 185
 - protein docking and virtual screening, 185–186
 - some protein structure prediction software, 183*t*
 - structure evolution analysis, 186
 - understanding dynamics, 186
 - unfolding studies, 185
 - validation of modeled protein, 184*f*, 185*f*
 - evaluation of, 184–185
 - machine learning in, 197–200
 - method of, 178–183
 - ab initio approach, 182–183
 - homology modeling (comparative modeling), 178–180
 - threading or fold recognition, 180–182
- Protein–protein interaction network (PPIN), 381, 384–385
- Protein–protein interactions (PPIs), 197–198, 283, 367, 381, 429
 - analysis, 468
 - databases, 430–431, 431*t*
- Proteins, 177, 186, 224–225, 467
 - D-structure quality assessment protocol, 191
 - design, 283–284
 - drug target, 185–186
 - expression analysis, 468
 - function
 - determination, 398
 - prediction, 198
 - functionality studies, 279

- Proteins (*Continued*)
 growth of protein structural database, 189–190, 190*f*
 machine learning in PSP, 197–200
 network construction based on protein interaction, 368
 preliminaries, 189
 protein 3D structure prediction, 191–197
 protein-amino acid-dependent factors, 194–195
 protein-coding RNAs, 428
 protein–DNA interaction, 291
 protein–ligand complex, 186, 224–225, 227
 protein–ligand docking, 291
 protein–nucleic acid docking, 291
 protein–protein docking, 291
 protein–RNA interaction, 291
 secondary structure prediction, 9
 structural topology and fold classification scheme, 190–191, 191*f*
 structure, 190
 threading, 180–181
 3D structure prediction, 9–10
 topology-based analysis, 190–191
- Proteome, 381
- Proteomics, 44, 367, 381. *See also* Metabolomics
 resources, 430, 431*t*
- Pseudo temporal ordering, 399–400
- PSI-BLAST homologous sequence, 29, 178–179, 394
- Psychiatric diseases, 342–343
- PTMnet, 21
- PubChem, 340, 394
 BioAssay, 27
 compound, 27
 database, 211
 substance, 27
- PubMed database, 415, 463
- Pulse cytophotometry, 392
- PyRodalgorithm, 280–281
- Q**
- Q-SiteFinder tool, 211
- QikProp tool, 311
- Quadrupole (Q), 391
- Quality and depth of raw sequencing data, 109–110
 adapter sequences, 110
 base calling accuracy, 109
 GC content per sequence, 110
 k-mer value, 110
 overrepresented sequences, 110
- Quality check, 97–98
- Quality control (QC), 109
 analysis, 162–163
 BQS, 162
 duplicate sequences, 163
 GC content, 163
 overrepresented k-mers, 163
 overrepresented sequences, 163
 per base N content, 163
 per base sequence content, 163
- quality analysis and improving software tools, 163
- of raw sequencing data, 52
- RLD, 163
- sequence quality score, 162
- software tools, 164*t*
- Quantitative Phase Contrast microscopy, 392–393
- Quantitative structure–activity relationship (QSAR), 209, 213–214, 213*f*, 233–234, 240–247, 255, 269
 application of QSAR/SAR, 248–250
 drug development and other applications, 244
 prediction of kinetic and thermodynamic parameters, 244
 synthetic organic chemistry and QSAR, 248
- backbone of chemical similarity, 241–242
- classification of 3D QSAR approaches, 243
- concept of applicability domain and QSAR approaches, 246–247, 247*f*, 248*f*
- data, 242
- designation of, 241, 241*f*
- development of new QSAR, 248
- modeling, 245–246
 determination of multitask modeling, 245
 different validation methods, 246
 validation of, 245–246
- molecular interactions and energies, 243–245
 AFMoC, 244
- comparative residue interaction analysis, 244
- detailed analysis of CoMSA, 244
- determination of CoMMA, 244
- GRID, 244
- interaction energies, 245
- representation of COMFA results, 244–245
- shape analysis, description of, 243
- QUARK, 195–197
- Quasi steady-state approximation, 349
- Quenching, 341
- QuickGo, 152
- QXP, 292
- R**
- R tool, 415
- Radiation hybrid (RH), 467
- Radius of gyration (Rg), 298
- Ramachandran plot, 184
- Rampage, 10
- Random amplified polymorphic DNA (RAPD), 131
- Random forest (RF), 295, 445
- Rarefaction curve, 170
- Rat Genome database (RGD), 153–154
- Reactome, 116, 340, 397
- Read length distribution (RLD), 163
- Receiver operating characteristic (ROC), 200
- Receiver operating characteristic curve, 276–277
- Reduce & Visualize Gene Ontology (REVIGO), 152
- Reductionist approach, 425
- Redundancy rate of reads, 111
- Reference-based assembly approach, 49, 54
- Refinement of metabolic networks, 360
- Regression, 446–450
 hypothesis, 447–450, 448*f*
 cost function, 447–448
 gradient descent, 448
 logistic regression algorithm, 448–450
 models, 459
- Regulatory controls, 348
- Regulatory on/off minimization (ROOM), 351
- Relation, 17
- Relational databases, 17
- REM-TrEMBL, 22
- Repeat Explorer, 60
- Repeat Masker, 60–61
- Repeat Modeler, 60
- Repetitive DNA, 59–60
- Replica-Exchange Monte Carlo (REMC), 195
- Resequencing, 105
- Restriction enzyme-based DNA sequencing, 135
- Restriction-site-associated DNA sequencing (RAD-Seq), 135
- Reverse transcriptase polymerase chain reaction (RT-PCR), 91
- Reverse vaccinology (RV), 317
- RF-Score, 296–297
- rFBA techniques, 359
- Ribosomal Database Project (RDP), 161
- Ribosomal RNA (rRNA), 96
- Rigid docking, 224–225
- RNA, 189
 polymerase, 429
 sequence, 452–453
 targets, 221
- RNA-sequencing (RNA-seq), 91, 96–99, 105, 388, 428, 452–453
 analyzing RNA-seq data, 109–111
 applications, 100, 111–121
 databases and software for small RNA analysis, 121–122
 identification
 of novel protein-coding genes, 112–113
 of other classes, 113–115
 library preparation and sequencing, 96–97
 limitations and future perspective of, 100
 next-generation sequencing platforms and technologies, 105–109
 pipeline and usage, 97–99
 software and databases, 116*t*
 steps involved in, 97–99
- rnaSPAdes, 98
- ROBETTA model, 209
- Robust Multi-chip Average (RMA), 94
- Roche 454 sequencing, 54, 95, 106–107
- Root, 70
- Root mean square deviation (RMSD), 185, 193, 297–298
- Root mean square fluctuation (RMSF), 297–298

- Rooted tree, 70
 Rosetta method, 29
 ROSETTA model, 195, 196*f*, 209
 ROSETTALIGAND, 292
 RSEM, 99, 388
- S**
- 16S rRNA, 159–160
 18S rRNA, 159–160
Saccharomyces cerevisiae, 341, 390
 Sailfish, 388
 Sample preparation, 339
 Sampling algorithms, 369–370
 Sanger sequencing, 94, 105
 SARS-CoV-2 proteins, 190
 Saudi Human Genome Program, 426
 SAVES v6.0, 184
 Scaffold hopping, 236
 Scale free, 382
 Schizophrenia, 342–343
 Scoring criteria, 180–181
 Scoring functions (SFs), 292, 294
 empirical SFs, 294
 force-field SFs, 294
 Scoring matrices, 36–38
 BLOSUM matrices, 37
 PAM matrices, 37
 PAM vs. BLOSUM, 37–38
 SDNGO, 149
 Search Tool for Retrieval of Interacting Genes/
 Protein (STRING), 62, 385, 394
 Second generation DNA sequencing, 95
 Second-generation sequencing (SGS), 108
 Secondary data, 463
 Secondary database, 22–23
 InterPro, 23
 Pfam database, 23
 PRINTS database, 22–23
 PROSITE, 22
 Secondary structure (SS), 452–453
 contact list, 190–191
 Semiflexible docking, 295
 Semisynthetic drugs, 259
 Sensitivity, 276
 Sequence Alignment Map (SAM), 110
 Sequence alignments, 33–36
 application of, 44
 construct, 34
 global sequence alignment, 35
 homology of sequences, 35
 local sequence alignment, 35–36
 similarity of sequences, 34–35
 working of alignment algorithm, 36
 initialization of matrix, 36
 matrix filling with maximum score, 36
 traceback approach to find appropriate
 alignment, 36
 Sequence analysis, 8
 Sequence capture, 136
 Sequence data, 429
 Sequence databases, 19–21
 DNA Database of Japan, 20
 ENA, 19
 GenBank, 19
 GenPept, 20
 PIR, 20–21
 Sequence quality score, 162
 Sequence Read Archive (SRA), 19
 Sequence Retrieval System (SRS), 28
 Sequence searching, 467
 Sequencing, 49
 platforms, 136
 quality analysis, 166–167
 Sequencing by Oligonucleotide Ligation and
 Detection (SOLiD), 54, 95
 Ser323 residue, 184
 Ser412 residue, 184
 Serial algorithms, 369–370
 Shannon diversity, 170–171
 SHARCGS method, 53–54
 Shared characters, 74
 SHARKhunt program, 394
 ShinyGO, 152
 Short nuclear polymorphism (SNP),
 161–162
 Short Read Archive (SRA), 162
 Shortest path length, average of, 374–375
 Shortest-path-based equation, 397
 Side chain modeling, 180
 Side effect–side effect networks, 432–433
 Sigmoid function, 449
 Signaling network, 386
 SignalP, 62
 Silent mutations, 37
 SILVA, 161
 Similarity search methods, 235
 Similarity-based methods, 29
 Simple sequence repeats (SSRs), 100, 131
 amplification and evaluation of polymorphic
 potential, 133–134
 analysis of SSR markers, 132–135
 classification and genomic localization, 133
 cross-transferability of SSR markers,
 134–135
 functional annotations of SSR containing
 sequences, 133
 future perspective, 135
 identification, 10
 mining and primer designing, 133
 Simulation tools, 416–418
 COPASI, 416–418
 GEPASI, 416
 SBW, 416
 Single-molecule real-time sequencing approach
 (SMRT), 96, 107–108
 Single-node-type network, 434
 Single-nucleotide polymorphisms (SNPs), 92,
 132, 467
 analysis of SNP markers, 135–140
 approaches for sequence data generation,
 135
 assessment of sequence quality, 136
 functional annotation of variant, 138–140
 postprocessing of mapped reads, 138
 reads assembly and mapping to reference
 genome, 136–138
 sample preparation, 135
 sequencing of complex genomes,
 135–136
 variant calling and filtration, 138
 identification, 10
 Singleplex PCR Target Enrichment Strategies,
 388–389
 Singular value decomposition (SVD), 362–363
 Small molecules
 database, 27–28
 leading to drugs/agrochemical development,
 10
 Small molecules PPI inhibitors (SMPPIIs),
 283
 Small noncoding RNAs, 113–115
 endogenous siRNAs, 113–114
 miRNAs, 114–115
 piRNA, 114
 Small RNA analysis
 Armour, 121
 databases and software for, 121–122
 miRBase, 121
 miRNAconsTarget, 122
 MPSS, 122
 PMRD, 121
 sRNAbench, 122
 sRNAtoolbox, 122
 UEA sRNA workbench, 121
 Small RNA-PARETargets Analyzer, 122
 Small world network, 382
 Small-interfering RNA (siRNAs), 111
 Smith–Waterman algorithm, 35
 SMRTbell, 108
 SNAP, 61
 SNPEff, 138–140
 SOAPdenovo, 52
 Social media, 426
 Social networks, 368
 Software, 147
 SoLid ABI, 107–108
 Soluble epoxide hydrolase (sEH), 282
 SP-TrEMBL, 21
 Spacers, 325–326
 SPAdes software, 53–54
 sPARTA, 122
 Specialized database, 24–28, 25*r*
 bibliographic databases, 26
 clustering databases, 24–25
 enzyme databases, 27
 expression databases, 26
 interaction databases, 26
 microRNA database, 27
 pathway databases, 27
 small molecule database, 27–28
 taxonomy databases, 26
 vaccine design database, 28
 Specificity, 276
 Split, 72
 Squared error cost function, 447–448
 sRNAbench, 122
 sRNAtoolbox, 122
 SSAKE method, 53–54
 SSAP tools, 190
 STAN server, 184
 STAR, 99

- Statistical approaches, 317
- Statistical comparison of biological networks, 375
- Statistical potential based SFs. *See* Knowledge-based scoring functions
- Statistics, 457–458
- Steady state, 349
- Stochastic differential equations, 410
- Stochastic method, 236
- Stochastic simulation, 414–415
- Streptomyces*
- S. erythreus*, 418
 - S. lividans*, 418
- Strongly significant of feature selection technique, 461–462
- Structural analysis and verification server (SAVES), 10
- Structural annotation, 61. *See also* Functional annotation
- AUGUSTUS, 61
 - MAKER, 61
- Structural Classification of Proteins (SCOP), 18–19, 24, 190, 464
- Structural databases, 23–24
- CATH/Gene3D, 24
 - research collaborative for structural bioinformatics protein data bank, 23
 - SCOP, 24
- Structural topology of protein, 190–191
- Structurally conserved regions (SCRs), 178–180
- Structure affinity relationship (SAR), 261
- Structure prediction techniques, 44, 177, 192–193, 221
- Structure refinement, 193–194
- Structure-based drug designing (SBDD), 207–209, 219, 233, 254–255. *See also* Ligand-based drug designing (LBDD)
- background of, 220
 - challenges and limitations, 228
 - drug discovery and development process, 208f
 - future prospective, 228–229
 - process of, 221–227
 - ADMET analysis, 226
 - binding-free-energy calculation-MM-PBSA, 227
 - cavity/binding site prediction, 223
 - ligand structure preparation/retrieval, 223–224
 - MDS, 226–227
 - molecular docking and virtual screening, 224–225
 - target identification, 221
 - target structure determination/prediction, 221–223
 - recent development in, 227–228
- Structure-based pharmacophore (SBP), 238–239. *See also* Ligand-based pharmacophore (LBP)
- feature-based outlook, 239
 - macromolecule ligand complex, 239
 - modeling, 275–276
 - with LigandScout, 275–276
 - structure-based drug discovery with Discovery Studio's interaction generation protocol, 275
- Structure–activity relationships (SAR), 207, 233–234
- Structured Query Language, 17
- Subtree, 72
- Superconductors, 339
- Superimposition based on atom overlapping, 235
- Supervised classification model, 458
- Supervised learning, 443–445
- classification, 444
 - problems to understand, 446
 - supervised machine learning in bioinformatics, 444–445
- Support vector machines (SVM), 295, 319, 445, 452, 458
- algorithms, 200
- Suppressors of gene silencing (SGS3), 114
- SVMTriP server, 321
- Swiss Institute of Bioinformatics (SIB), 21
- Swiss Model, 209
- Swiss PDB viewer, 10
- SWISS-PROT, 18
- Swiss-Prot database, 21
- SwissDock server, 211–212
- SymmDock, 428
- Synapomorphic traits, 67–68
- Synapomorphies, 68–69
- Synapomorphism, 67–68
- SynLinker tool, 325–326
- Synthetic approach, 96
- Synthetic lead molecules, 257–258
- Synthetic lethal genes identification, 358
- Synthetic Long-Read technology (SLR), 107–108
- System Biology Markup Language (SBML), 412–413
- Systematics, 68–69
- search algorithms, 236
- Systemic sclerosis (SSc), 396–397
- Systems biology, 11, 409, 425
- applications of pathway modeling and simulations, 418–419
 - drug designing, 418–419
 - flux balance analysis, 419
 - metabolic engineering, 418
 - phenomics study, 419
 - challenges, 420
 - knowledge gaps between computationalists and experimentalists, 420
 - miscellaneous computational challenges, 420
 - theory development, 420
 - computational modeling of pathway, 409–411
 - data type in, 426–429
 - general diagram and language used in pathway modeling, 411–413
 - SBGN, 411–412
 - SBML, 412–413
 - pathway simulations analysis, 413–415
 - platforms used for modeling and simulations, 415–418
 - pathway designing tools, 415
 - Pathway Tools, 415–416
 - simulation tools, 416–418
- Systems Biology Graphical Notation (SBGN), 411–412
- Systems Biology Research Tool (SBRT), 357
- Systems Biology Workbench (SBW), 415–416
- Systems pharmacology, big data in, 434–435
- ## T
- T-cell epitope (TCE), 318
- in silico* method for T-cell epitope prediction, 322–325
 - MHC class I binder prediction, 323–324
 - MHC class II binder prediction, 324–325
 - MHC gene diversity and importance in TCE prediction, 325
- T-cell receptors (TCRs), 317–318
- Tanimoto coefficient, 235
- Target 3D structure, 223
- Target alignment, 180
- Target identification and de novo ligand design using pharmacophore approaches, 279
- Target model, 178
- Target prediction, miRNAs and, 469
- Target protein, 3D structure of, 209
- Target structure determination/prediction, 221–223
- Target template, 235
- Target-enrichment strategies, 388–389
- Target-oriented synthesis (TOS), 259–261
- Targeted metabolomics, 338
- TargetFinder Perl script, 435
- TargetP, 62
- Taxon, 69
- Taxonomic hits distribution, 170
- Taxonomy databases, 26
- tbl2asn, 19
- Template, 178
- Template-based modeling (TMB), 178, 192
- TF hubs, 383–384
- TFBIND, 61
- Theoretical biology, 413
- Therapeutic properties, 253
- Thermodynamic constraints to metabolic model, 348
- Third generation DNA sequencing, 96
- Third-generation sequencing (TGS), 107–108, 108t
- Threading, 181
- alignments, 195
 - recognition, 180–182
 - twofold recognition approach, 181f
- Three-dimension (3D), 233
- arrays, 429
 - fingerprints, 235–236
 - conformational flexibility, 236
 - fragment-based drug design, 236
 - scaffold hopping, 236
 - 3D similarities depend on alignment, 236
 - macromolecular structure, 233–234

- model and validation of fusion vaccine, 326
 pharmacophore, 237
 constructive phase, 237
 modeling using web platform, 279–280
 optimization phase, 237
 principles over above limitations, 240
 subtractive phase, 237
 protein structure, 181
 QSAR approach, 243–244
 space, 181–182
 structures, 189, 219–220, 233
 data, 429
 of protein, 177
 of target protein, 209
- 3D electron cryo-microscopy (3DEM), 189
 3Omics, 340
 TIGR genome assembler, 50–51
 Time of flight (TOF), 340, 391–392
 Tissue microarray (TMA), 91
 Toll-like receptors (TLRs), 282, 326
 Top-ranking molecules, 235
 Tophat2, 99
 TOPKAT tool, 311
 Topobiological constraints, 350
 Topological parameters of network, 373–375
 average of shortest path length, 374–375
 betweenness centrality, 375
 clustering coefficient, 375
 node degree, 374
 statistical comparison, 375
 Topological properties, 373
 Topology, 70
 topology-driven online resources, 190–191
 ToppGene, 151
 ToxinPred, 326–327
 Trace Archive, 19
 Traceback approach, 33–34
 Training set of compounds, 239
 Trajectory, 297–298
 Trans-ABYSS, 98
Trans-Acting small-interfering RNAs (ta-siRNAs), 114
Trans-eQTLs, 120
 Transcription factors (TFs), 428
 Transcriptome
 assembly, 111–112
 de novo transcriptome assembly, 112
 genome-guided transcriptome assembly, 111–112
 case study related to transcriptome data
 analysis, 435–437
 sequencing, 105, 136
 Transcriptomics, 44, 91, 367
 resources, 430
 Transeq, 61
 Transfer RNA (tRNA), 96
 Translate, 61
 Translated EMBL database (TrEMBL
 database), 21–22
 Translational bioinformatics, 313
 Transporters associated with Ag-processing
 protein (TAP protein), 323–324
 Transposon-mediated fragmentation, 388–389
 Tree construction methods, 74–84
 Bayesian phylogeny, 83–84
 maximum likelihood phylogeny, 82–83
 maximum parsimony, 80–82
 neighbor-joining algorithm, 76–80
 UPGMA, 74–76
 Tree of Life, 73
 Trimming and filtering of raw reads, 52
 Trimmomatic tool, 52, 98
 Trinity, 98
 True positive (TP), 200
 TruSeq Amplicon, 388–389
Trypanosoma brucei, 418–419
 Twitter, 426
 2-dimensional arrays, 429
- U**
 Uber, 425
 UCSC Genome Browser, 62
 UEA sRNA workbench, 121
 Ultrametric tree, 72
 Undirected graphs, 372
 Unicycler software, 53–54
 UniPort database, 464
 UniProt Archive (UniParc), 21–22
 UniProt Knowledgebase (UniProtKB), 21–22
 UniProt Reference Cluster (UniRef), 21–22
 UNited RESidue (UNRES), 194
 Universal Protein Resource (UniProt), 21–22
 UniParc, 22
 UniProtKB, 21–22
 UniRef, 22
 Unrooted tree, 70
 Unsupervised classification model, 458
 Unsupervised learning in bioinformatics, 452
 Unsupervised machine learning, 445–446.
 See also Supervised learning
 problems to understand, 446
 Untargeted metabolomics, 338
 Unweighted Pair Group Method with
 Arithmetic Mean (UPGMA), 43, 74–76
 advantages and disadvantages of neighbor-
 joining method, 80
- V**
 Vaccination, 317
 Vaccine design(ing), 317
 adjuvant and linker selection, 325–326
 antigen selection and immunological
 databases, 318–319
 autoantigens, 319
 endogenous antigens, 319
 exogenous antigens, 318
 building 3D model and validation of fusion
 vaccine construct, 326
 computer-aided vaccine development
 example, 328–329
 database, 28
 in silico method for
 BCE prediction, 319–322
 T-cell epitope prediction, 322–325
 miscellaneous properties, 326–328
 delivery methodology, 327–328
 half-life or stability, 327
 peptide toxicity, 326–327
 next-generation sequencing technology role
 in vaccine design, 328
 Vaccine Investigation and Online Information
 Network (VIOLIN), 28
 Vaccinology, 317
 Val4 residue, 184
 Valency. *See* Degree
 Validation of pharmacophore models, 237,
 276–278
 analysis of Fisher's randomization test, 237
 azaindole derivatives, 277–278
 GH scoring method, 237
 receiver operating characteristic curve,
 276–277
 test set prediction, 237
 Van der Waals forces (vdW forces), 223
 Variants
 calling, 138
 detection, 120
 Vaxign query, 319
 VaxiJen, 319
 VaxiJen 2.0 server, 322
 Vaxjo database, 325
 VCAKE method, 53–54
 Velvet software, 53–54
 Verify 3D, 209
 Vertex similarity (VS), 397
 application of vertex similarity to rare
 diseases, 397
 Virtual screening (VS), 211, 224–225,
 233–234, 239–240, 270
 conformer selection, 239
 exploration of conformational space,
 239–240
 of high-throughput docking, 295
 Virus attack, 435–437
 ViSEAGO, 152
 Visualization, 149
 Volume, velocity, veracity, and variety (four
 Vs), 426
- W**
 Waikato Environment for Knowledge Analysis
 (WEKA), 458, 461
 Weakly significant of feature selection
 technique, 461–462
 Web Gene Ontology (WEGO), 152
 Web Graphics Library technology, 396
 Weighted graph, 372
 Wellcome Trust Sanger Institute, 62
 WhatCheck, 209
 Whole exome sequencing (WES), 388
 Whole genome bisulfite sequencing (WGBS),
 389–390
 Whole genome sequencing, 136, 367
 Worldwide PDB (wwPDB), 189
 Wrapper method, 198
- X**
 X-rays, 177–178

X-rays (*Continued*)

crystallography, 177, 207–208, 219–220, 428
XGBoost algorithm, 199
XML. *See* Extensible Markup Language (XML)

Y

Yeast
tRNA, 18
two-hybrid, 390–391
Yeast one-hybrid (Y1H) system, 384
YeastCyc, 393
Yield of actives (YA), 276

Z

ZDOCK server, 211–212
Zebrafish fish (*Danio rerio*), 342
Zero-mode waveguide (ZMW), 108
Zileuton, 256
Zolmitriptan, 255