

# INDEX

## A

- aaRSs, *see* Aminoacyl-tRNA synthetases
- aaRS-tRNA complexes structure,
  - usage..... 443–444
- ABL-BCR pathway of human, network transition
  - probabilities of..... 263–264
- Acute myeloid leukemia ..... 257
- Additive approximation..... 25, 34
- A-GLAM software package
  - hardware..... 4
  - intergenic regions of cluster files ..... 17–21
  - network comparison operation method
    - functional similarity network construction..... 8–9
    - identification of cis-regulatory elements in
      - promoter regions..... 11–13
    - identification of dense gene clusters ..... 10–11
    - identification of functional motifs..... 5
    - identification of proximal promoter regions..... 11
    - individual score and E-value..... 6–7
    - intersection network construction ..... 9–10
    - size of co-expression networks..... 5
  - software ..... 4–5
- Algorithm comparison and SEBINI..... 560
  - See also* Software Environment for Biological Network Inference
- Alzheimer’s Disease (AD)..... 451
  - protein, interaction of..... 458–459
  - related protein
    - SNP and disease mutation in..... 454
    - SVM and SIFT in ..... 457
- Aminoacyl-tRNA biosynthesis pathway ..... 505
- Aminoacyl-tRNA synthetases ..... 443–446
- AML, *see* Acute myeloid leukemia
- API, *see* Application Programming Interface
- Apolipoprotein, decision trees for ..... 393
- Application Programming Interface..... 512, 523
  - client, role of..... 523–525
  - documentation ..... 526–527
  - role of bioverse ..... 523
  - usage of ..... 525–526
  - versioning of..... 530
- Arabidopsis* Genome Initiative locus identifier
  - (AGI ID)..... 110
- Arabidopsis thaliana* ..... 108, 114–115
- Array data, pathway inference
  - and modeling of..... 232–237
  - See also* Genome-wide data generation, experimental methods for

## B

- Bacillus subtilis* ..... 108, 440
- Bacterial two-component system, data collection
  - and alignments for ..... 428–429, 434–435
  - See also* Specificity determining residues
- Bacteriophages, role of..... 337–338
- Bagging method, in gene clusters..... 487
- Bayesian approach, for gene expression..... 488–490
  - See also* Cluster algorithms, for gene expression
- Bayesian estimation technique, in intron
  - evolution ..... 358–359
- Bayesian Estimator of Protein–Protein Association
  - Probabilities ..... 553
- Bayesian expression network
  - definition of ..... 226
  - physical interaction networks,
    - measurement of..... 240–241
    - TF-gene regulation, usages of..... 239–240
- Bayesian marginal log ..... 5
- Bayesian network approach, features of..... 237
- Bayes nets, *see* Bayesian expression network, definition of
- BellaVista protocol..... 526
- Benjamini & Hochberg False Discovery Rate (FDR)
  - correction ..... 9
- BEPro, *see* Bayesian Estimator of Protein–Protein Association Probabilities
- Bergey’s Manual of Determinative and Systematic Bacteriology ..... 173
- Biased reference set, in GO biological process..... 468
  - See also* Supervised learning
- Biclustering method..... 205–206
- Bidirectional best-hit procedure, of network
  - reconstruction ..... 168
- BIND, *see* Biomolecular Interaction Network Database
- BIND database ..... 69, 84
- BiNGO plug-in ..... 9
- Binomial tau-leaping method..... 322, 323
- Biochemical control theory (BCT), for biochemical
  - network ..... 275–276

- Biochemical control (*continued*)
- dynamic motifs
    - bistable systems ..... 298–299
    - feedback and oscillatory systems ..... 300–304
  - and engineering control theory ..... 290
  - frequency response ..... 291–295
  - stability analysis ..... 295–297
  - linear perturbation analysis
    - control coefficients ..... 278–279
    - elasticities and control coefficient, relationship ..... 279–281
    - elementary processes ..... 276–277
  - pathway motifs, linear analysis of
    - branched pathways ..... 283–285
    - cyclic systems ..... 285–287
    - negative feedback of ..... 287–290
    - straight chains ..... 281–283
- Biochemical network, steady-state solution, for ..... 275
- Biochemical pathways
  - construction of ..... 227–228
  - definition of ..... 226
- Biochemical Systems Theory ..... 276
- Biological network inference and analysis ..... 552–555
  - control genes, identification of ..... 556
  - gene function and target identification ..... 555
  - irrelevant genes, elimination of ..... 555
- Biological networks
  - benefits of ..... 146–147
  - dynamic behavior of ..... 251
  - emergence of ..... 146
  - hierarchical module structure ..... 147–149
  - metabolic networks ..... 147
  - reconstruction, goals of ..... 229
  - structure, determination of ..... 421
  - topological organization of cellular metabolism ..... 147
    - metabolic network of *Escherichia coli*, case study ..... 151–156
    - method for finding ..... 150–151
- Biological replicate ..... 203
- Biological systems
  - computational representation of ..... 535–536
  - kinetic modeling of
    - discrete stochastic simulation, acceleration of ..... 320–324
    - historical perspectives of ..... 311–315
    - network size, scaling properties of ..... 326–327
    - simulation methods for ..... 316–319
    - software tools for ..... 328–330
    - SSA fundamentals ..... 319–320
    - thermodynamic limit ..... 324–325
- BioMoby web services, in bioinformatics tools ..... 536
- Biomolecular Interaction Database ..... 556
- Biomolecular Interaction Network
  - Database ..... 104, 106–107, 129, 514
- Bioverse
  - API
    - client application in ..... 523
    - usage of ..... 523, 525–526
    - XML-RPC protocol in ..... 526
  - audiences in ..... 512
  - data
    - creation and presentation ..... 515
    - organisation ..... 512–513
    - pipeline ..... 515–518
    - protinfo ..... 518
    - sources ..... 513–515
  - database in ..... 518
  - data tracking and ..... 541
  - ETL ..... 546–547
  - function identifier ..... 527–528
  - importance of ..... 511–512
  - integrator ..... 524
  - model design
    - functional data ..... 544–546
    - sequence data ..... 541–542
    - taxonomy data ..... 542–544
  - technology ..... 534
  - web application
    - HTML templates and ..... 520
    - implementation of ..... 519
    - role of ..... 515
  - See also* Application Programming Interface
- Biozon database ..... 69
- BL2SEQ ..... 117
- BLAST ..... 82, 118
  - in enzyme detection ..... 374
- BLASTclust procedure, of network
  - reconstruction ..... 168, 176
- BLOCK matrix ..... 33
- BMP7, *see* Bone morphogenetic protein 7
- BMP7 and ACVR2B interaction, in lung cancer
  - co-expression ..... 260
- BOND database ..... 106–107, 129
- Bone morphogenetic protein 7 ..... 260
- Bottom-Up modelling, experimental design for
  - building of ..... 344–347
  - decouple key components ..... 343–344
  - mutations effects on ..... 344
  - network subset isolation ..... 343
  - in silico experiments, model predictions in ..... 347–348
- B-spline approximation ..... 252
  - See also* Gene co-expression
- B $\tau$ L method, *see* Binomial tau-leaping method
- C**
- C4.5 algorithm, in protein modeling ..... 385
- CABIN, *see* Collective Analysis of Biological Interaction Networks

<i>Caenorhabditis elegans</i> .....	107–110, 113–114, 146
Cancers and normal tissue, differential co-expression of.....	257–261
CCL23, <i>see</i> Chemokine ligand 23	
CCR1, <i>see</i> Chemokine receptor 1	
Cellular network dynamics, quantitative approaches of.....	270–271
Center for Molecular and Cellular Systems.....	567
CERs, <i>see</i> Co-evolving residues	
CHARMM.....	28, 34–36
BLOCK module in.....	31–32
CPT keyword.....	30
and dual-topology.....	31
script file.....	30
CHARMM's HBUILD module.....	28
Chemical Langevin equation.....	325
Chemical master equation.....	314, 319
Chemical transformation, in biological networks.....	276–277
Chemokine ligand 23.....	260
Chemokine receptor 1.....	260
ChIP-chip.....	24
CHIP on-chip data.....	239
CHIP on chip technique, in protein-DNA interactions measurement.....	229
Chromosomal single-copy reporters, usages of.....	340
Chronic myeloid leukemia.....	261
CLE, <i>see</i> Chemical Langevin equation	
ClustalW algorithm.....	85, 118–119
Cluster algorithms in co-regulated genes identification.....	230–232
for gene expression.....	485–488
algorithms.....	494
best clustering, determination of.....	500–501
cellular pathways, association of clusters in.....	502–506
double index assessment.....	488
external index of validity.....	490–491
hierarchical clustering algorithms.....	482–483
individual clusters, assessment of.....	501–502
internal index of validity.....	488–490
noisy data, algorithms for.....	484–485
parameter optimization.....	495–500
partitional clustering algorithms.....	481–482
pre-processing similarity data.....	494
reference set.....	492
similarity measures.....	493
spectral clustering algorithms.....	483–484
CMCS, <i>see</i> Center for Molecular and Cellular Systems	
CME, <i>see</i> Chemical master equation	
CML, <i>see</i> Chronic myeloid leukemia	
<i>cMonkey</i> .....	206
CO4A5_HUMAN protein.....	454
CoD, <i>see</i> Coefficient of determination	
Coefficient of determination.....	250
estimation of.....	252–254
<i>See also</i> Gene co-expression	
Co-evolving residues.....	424
in ATPase and DD domains.....	437–440
in class Ia aaRSs.....	443–446
conservation and variation, sources.....	425
in double alignments, determination.....	433
genomic proximity in.....	426
modeled structures.....	440–441
procedures in determining.....	426–427
protein duplication, consequences of.....	425–426
in RR domain.....	435–437
statistical significance estimation and computation.....	434
CoExMiner algorithm.....	250
biological applications of disease genes identification and drug targets of leukemia.....	261–265
ligand-receptor pairs, co-expression of.....	257–261
Co-expressed genes.....	205
Co-expression network, <i>see</i> Correlation network, definition of	
Collective Analysis of Biological Interaction Networks.....	551
development and advantages of.....	556
exploratory analysis and.....	561–562
find/search functionality and.....	562
inferred network in.....	566
inferred network structures.....	560
java and.....	559
network import in.....	561
network subsets and.....	562
<i>Rhodopseudomonas palustris</i> and.....	568
role of.....	552
and SEBINI system.....	559
Computational modelling, LacZ units in.....	341
Computational systems biology, challenges in.....	249–250
Condition-specific sub-networks, categories of.....	235
Connectivity theorem.....	280
<i>See also</i> Biochemical control theory (BCT), for biochemical network	
Conservation index (C.I.), for network motifs.....	172, 177
Context-specific clustering technique.....	237
Co-regulated genes.....	205
identification, cluster analysis in.....	230–232
<i>See also</i> Genome-wide data generation, experimental methods for	
Correlated mutations, <i>see</i> Co-evolving residues	
Correlation network, definition of.....	226
Cytoscape, usages of.....	4, 10, 233
<b>D</b>	
DAEs, <i>see</i> Differential-algebraic equations	
DAG, <i>see</i> Directed acyclic graph	

- Database of Interacting Proteins ..... 69, 514, 556
- Database of Ligand– Receptor Partners..... 257–258
- Data warehouse..... 536–537
- construction of
- extracting, transforming and loading..... 539–541
- model design ..... 539
- dimensional models ..... 537–539
- relational databases ..... 537
- DBNs, *see* Dynamic Bayesian networks
- DBTBS ..... 104, 108
- DD, *see* Dimerization domain
- Decision tree learning model, for protein modeling
- classification of..... 383–384
- training procedure for ..... 384–387
- Diabetes Mellitus (DM)..... 193
- Differential-algebraic equations ..... 312
- Dimerization domain..... 434
- DIP, *see* Database of Interacting Proteins
- Directed Acyclic Graph (DAG)..... 193, 204, 542
- Discrete stochastic simulation
- acceleration of ..... 320–324
- in biological systems study..... 313
- See also* Biological systems, kinetic modeling of
- DLRP, *see* Database of Ligand– Receptor Partners
- DNA microarrays, usages of..... 230
- Dollo parsimony, in intron evolution..... 358
- Double index assessment, for gene expression ..... 488
- See also* Cluster algorithms, for gene expression
- Drosophila melanogaster* ..... 108, 115, 146, 529
- Dynamical stiffness, definition of..... 312
- Dynamic Bayesian networks..... 241
- Dynamic motifs
- bistable systems ..... 298–299
- feedback and oscillatory systems..... 300–304
- See also* Biochemical control theory (BCT),  
        for biochemical network
- E**
- EBI UniProt database, usage of ..... 227
- EC classification test, for protein domains ..... 400–402
- EC hierarchy..... 376
- See also* Enzyme class prediction
- EC number, *see* Enzyme Commission number
- Elasticity coefficient, definition of ..... 277
- EM-like algorithm, for protein subset selection ..... 80–81
- Engineering control theory and BCT ..... 290
- frequency response ..... 291–295
- stability analysis..... 295–297
- See also* Biochemical control theory (BCT),  
        for biochemical network
- Enzyme class prediction
- EC hierarchy..... 376
- proteins
- alternate representation of ..... 378–379
- multiple aspects of ..... 380–381
- sequence similarity..... 376–377
- structure-based approaches ..... 377–378
- Enzyme Commission number ..... 373, 376
- Enzyme function prediction
- data preparation and feature extraction method
- database characteristics of ..... 382–383
- predicted features of..... 382
- sequence characteristics of ..... 381–382
- decision tree learning model, methods for
- decision tree model ..... 383–384
- decision trees, training procedure for ..... 384–387
- learning algorithm, optimization
- method of..... 390–392
- stochastic decision trees, mixture of
- decision trees, evaluation of ..... 388–390
- stochastic framework ..... 387–388
- Enzymes in organisms, role ..... 374
- eQTL mapping, *see* Expression quantitative trait locus  
    mapping
- Equilibrium..... 2023
- Escherichia coli* ..... 302, 338
- etabolic network, case study
- color-coding of hierarchical tree..... 153–154
- data source ..... 151
- network generation ..... 152–153
- pathways of pyrimidine metabolism ..... 154–156
- Escherichia coli* K-12 ..... 108
- ET, *see* Evolutionary tracing
- ETL, *see* Extraction, Transformation, and Loading
- Euclidean metric, in mRNA expression..... 493
- Eukaryotes ..... 112
- Eukaryotes, prediction and identification of ..... 3–4
- Eukaryotic gene structure
- bayesian estimation technique in ..... 358–359
- expectation-maximization algorithm..... 359
- maximum likelihood estimation techniques in
- materials in..... 360–361
- methodologies in..... 361–368
- properties of ..... 360
- Evolutionary tracing ..... 423
- Expectation-maximization algorithm, in introns
- evolution ..... 359–360
- Expression quantitative trait loci (eQTL), analysis
- building molecular interaction graph..... 213
- drawbacks..... 221
- finding linked loci ..... 213
- identifying linear pathways ..... 214–215
- random walks across an interaction graph..... 215–216
- using gene knockouts..... 216–217
- Expression quantitative trait locus mapping ..... 237
- Extraction, Transformation, and  
    Loading..... 539, 540, 546

**F**

FA8\_HUMAN protein ..... 454  
 FA9\_HUMAN protein ..... 454  
 Fact table  
   characteristics of ..... 539  
   definition of ..... 538  
 False negative/positive ..... 203  
 FASTA files ..... 11  
 Felsenstein's pruning algorithm ..... 364  
 Flux control coefficients ..... 288, 289  
 Flux summation theorem ..... 280  
   *See also* Biochemical control theory (BCT),  
     for biochemical network

**G**

$\beta$ -galactosidase ..... 203  
 $\beta$ -galactosidase, measurement ..... 340  
*GAL* gene expression ..... 232–233  
 Gaussian cumulative distribution function ..... 432  
 Gene co-expression  
   computational model of  
     mixed patterns ..... 251–254  
     pathway dynamics ..... 254–257  
   study of ..... 250  
 Gene expression  
   cluster algorithms for ..... 485–488  
     algorithms ..... 494  
     best clustering, determination of ..... 500–501  
     cellular pathways, association  
       of clusters in ..... 502–506  
     double index assessment ..... 488  
     external index of validity ..... 490–491  
     hierarchical clustering algorithms ..... 482–483  
     individual clusters, assessment of ..... 501–502  
     internal index of validity ..... 488–490  
     noisy data, algorithms for ..... 484–485  
     parameter optimization ..... 495–500  
     partitional clustering algorithms ..... 481–482  
     pre-processing similarity data ..... 494  
     reference set ..... 492  
     similarity measures ..... 493  
     spectral clustering algorithms ..... 483–484  
   Gene functional annotation, current state  
     assessment of ..... 226–227  
   Gene function identification ..... 555  
     *See also* Biological network inference and analysis  
   Gene functions, databases for ..... 464  
   Gene–gene similarity ..... 8  
 Gene network model  
   functional bias effects on ..... 463–465  
     circumventing functional bias ..... 472  
     functional annotations ..... 465–468  
     genomics data set functional bias ..... 470–472

    reference set functional bias ..... 468–470  
 Gene ontology (GO) ..... 465, 466  
   annotations, for yeast gene products ..... 8  
   biological process annotations ..... 467  
   evidence codes ..... 466  
   hierarchy in ..... 472–473  
 Gene regulatory networks, data collection  
   and modeling in ..... 338–340  
   bottom-up modelling, experimental design for ..... 342  
   building of ..... 344–347  
   decouple key components ..... 343–344  
   mutations effects on ..... 344  
   network subset isolation ..... 343  
   in silico experiments, model predictions in ..... 347–348  
   model organisms for ..... 337–338  
   phage  $\lambda$ , promoter regulation in  
     computational modeling of data ..... 350–353  
     experimental design in ..... 348–350  
     prokaryotic promoters, measurement of ..... 340–341  
 Gene-specific model ..... 358  
 Genetic regulatory networks ..... 182  
 Genome-wide data generation, experimental  
   methods for ..... 228  
   physical interaction, measurement methods of ..... 229–230  
   regulation and predictive networks creation,  
     methods for  
       array data, pathway inference and  
       modeling of ..... 232–237  
       co-regulated genes identification,  
       cluster analysis in ..... 230–232  
       gene expression and genetics ..... 237–239  
 Genome-wide genetic mapping, usages ..... 238  
 GGB, *see* Gibson-Bruck method  
 Gibbs sampling algorithm ..... 5  
 Gibson-Bruck method ..... 326  
 Gillespie method ..... 271  
 Glycine biosynthesis pathway ..... 375  
 GO, *see* Gene ontology  
 Graph-based clustering algorithms, *see* Pairwise  
   clustering algorithms  
 Graph Merge plug-in ..... 9

**H**

*Halobacterium NRC-1* ..... 195  
 Heat-cool-cycle ..... 32–33  
*Helicobacter pylori* ..... 146  
 Hierarchical clustering algorithms ..... 482–483  
   *See also* Cluster algorithms  
 Hierarchical clustering, of expression data ..... 497–498  
   *See also* Cluster algorithms  
 High blood glucose (HBG) ..... 193  
 High-throughput protein–protein interaction,  
   measurements of ..... 229  
 Histidine kinase ..... 434

- HK, *see* Histidine kinase  
HMMER algorithm ..... 516  
*Homo sapiens* ..... 108, 112–113  
HPRD, *see* Human Protein Reference Database  
Human Protein Reference Database ..... 69, 514
- I**
- ID3 learning algorithm, characteristics of ..... 385  
IDA, *see* Inferred from direct assay  
I/GDP-mannose metabolism ..... 375  
IGI, *see* Inferred from genetic interaction  
IMAGE ..... 33  
Immunoprecipitation ..... 203  
IMP, *see* Inferred from mutant phenotype  
Inferelator tutorial ..... 197–201  
Inferred from direct assay ..... 465  
Inferred from Electronic Annotation (IEA) ..... 8  
Inferred from genetic interaction ..... 465  
Inferred from mutant phenotype ..... 465  
Inferred from physical interaction ..... 465  
Information gain, definition of ..... 385  
Information impurity, definition of ..... 384  
In silico experiments, model predictions in ..... 347–348  
    *See also* Gene regulatory networks, data collection  
    and modeling in  
Integrator, in network visualisation ..... 524  
Interolog ..... 128  
Interolog method and BIND data ..... 514–515  
Intersection network construction ..... 9–10  
Introns evolution, in eukaryotes ..... 357  
    bayesian estimation technique in ..... 358–359  
    expectation-maximization algorithm ..... 359  
    maximum likelihood estimation  
        techniques in ..... 359  
        materials in ..... 360–361  
        methodologies in ..... 361–368  
    properties of ..... 360  
IPI, *see* Inferred from physical interaction  
Iterative clustering method, of noisy data ..... 485  
    *See also* Cluster algorithms  
Iterative clustering, of expression data ..... 495–496  
    *See also* Cluster algorithms  
IUB Enzyme Commission ..... 376  
IUBMB-Sigma-Nicholson Metabolic Pathways ..... 228
- J**
- Java and CABIN ..... 559  
Java Runtime Environment (JRE) ..... 4  
JavaScript Object Notation ..... 526  
Jensen-Shannon measure ..... 389  
 $\lambda$ -JS divergence, definition of ..... 389  
JSON, *see* JavaScript Object Notation  
Jump Markov process ..... 319
- K**
- K connections ..... 172  
KEGG, *see* Kyoto Encyclopedia of Genes and Genomes  
KEGG2 database ..... 227  
Kidney disease (KD) ..... 193  
Kinetic measurement ..... 203  
Kinetic modeling, of biological systems  
    discrete stochastic simulation,  
        acceleration of ..... 320–324  
    historical perspectives of ..... 311–315  
    network size, scaling properties of ..... 326–327  
    simulation methods for ..... 316–319  
    software tools for ..... 328–330  
    SSA fundamentals ..... 319–320  
    thermodynamic limit ..... 324–325  
Klemm-Eguíluz model ..... 156–157  
K-means algorithm ..... 481–482  
    *See also* Cluster algorithms  
K-means clustering algorithm, for expression  
    data ..... 498–500  
    *See also* Cluster algorithms  
kNN algorithm ..... 82–83  
Knowledge-based network, definition of ..... 226  
Kyoto Encyclopaedia of Genes and Genomes ..... 466, 467
- L**
- L1-shrinkage ..... 207  
*lacZ* gene ..... 340  
*lacZ* transcriptional fusions, usage of ..... 341  
Langevin leaping formula ..... 325  
Laplace transform approach ..... 291  
LASSO ..... 207–208  
Learning algorithm, optimization of ..... 390–392  
Learning system, parts of ..... 381  
Leukemia-related BCR-ABL pathway,  
    analysis of ..... 261  
L-glutamine ..... 156  
Lifestyle-based network similarity  
    index (LSI) ..... 174–175, 178  
Ligand-receptor pairs, co-expression ..... 257–261  
Linear perturbation analysis  
    control coefficients ..... 278–279  
    elasticities and control coefficient,  
        relationship ..... 279–281  
    elementary processes ..... 276–277  
    *See also* Biochemical control theory (BCT),  
        for biochemical network  
Linux ..... 4–5  
    and python ..... 534  
LLS, *see* Log-likelihood score  
Locally homogeneous approach, importance of ..... 314–315  
Log-likelihood score ..... 468, 469

**M**

Machine-learning algorithms, importance of..... 470  
 Markov chain algorithm ..... 251  
 Markov chain model, in intervention analysis..... 256–257  
 Markov-chain Monte Carlo procedure ..... 5  
 MATLAB, usage of ..... 559  
 Matrix2png ..... 169, 177  
 Maximum likelihood estimation techniques,  
     in intron evolution ..... 359  
 MCODE Cytoscape plug-in ..... 10, 15  
 MCODE plug-in for Cytoscape ..... 4  
 MDL, *see* Minimum description length  
 MDL principle, model selection in ..... 500  
 Metabolic Control Analysis, *see* Biochemical  
     Systems Theory  
 Minimum description length..... 410, 486  
 MIPS, *see* Munich Information Center for Protein  
     Sequences  
 Mirror-tree co-evolution-based algorithms ..... 66  
 Mismatch/Novel kernel, function of ..... 379  
 Modularity, concept of ..... 145  
 Module-replacement approach..... 344  
     *See also* Gene regulatory networks, data collection  
         and modeling in  
 Molecular pathway, *see* Biochemical pathways, definition of  
 Molecule\_by\_name API method, confirmation of ..... 527  
 Molecule\_sequence table and sequence data ..... 541–542  
     *See also* Bioverse  
 Monte Carlo (MC)  
     algorithm, usage of ..... 430–432  
     simulation..... 254, 313  
 mRNA expression  
     euclidean metric in..... 493  
     SEBINI and..... 564  
 MSAs, *see* Multiple sequence alignments  
 M $\tau$ L method, *see* Multinomial tau-leaping method  
 “multi-copy” representation..... 31  
 Multinomial tau-leaping method ..... 323  
 Multiple sequence alignments ..... 433  
 Multi-scale chemically reacting systems simulation,  
     hybrid methods in..... 330–331  
 Munich Information Center for Protein  
     Sequences ..... 470, 514  
*Mus musculus* ..... 108  
 Mutation analysis, in protein ..... 451  
     *See also* Protein  
 Mutual information, computation of ..... 429  
     *See also* Specificity determining residues

**N**

Name mapping ..... 111–112  
 NCBI GenBank Identifiers (GIs) ..... 109, 131  
     role of ..... 514

Negative design constraints ..... 442  
 Negative feedback oscillators ..... 300–302  
     *See also* Biochemical control theory (BCT),  
         for biochemical network  
 Network import and CABIN ..... 561  
     *See also* Collective Analysis of Biological Interaction  
         Networks  
 Network inference and SEBINI..... 560  
     *See also* Software Environment for Biological  
         Network Inference  
 Network motifs ..... 170  
 Newton-Raphson algorithm..... 368  
 Normalization ..... 203  
 Normalized cut, definition of ..... 484  
 Nucleic acid..... 5

**O**

Oak Ridge National Laboratory..... 567  
 ODEs, *see* Ordinary differential equations  
 OLAP, *see* Online analytical processing  
 OLTP, *see* Online transaction processing  
 OMIM database and disease-specific proteins..... 453–454  
 Online analytical processing ..... 536–537  
 Online Predicted Human Interaction Database ..... 451  
 Online transaction processing..... 536  
 Operon ..... 202  
 OPHID, *see* Online Predicted Human Interaction  
     Database  
 Oracle DBMS sqldr and protein interactions ..... 452  
     *See also* Protein  
 Ordinary differential equations ..... 270, 312, 345  
 ORNL, *see* Oak Ridge National Laboratory  
 Orthologs ..... 128, 168  
 Orthology detection procedure, of network  
     reconstruction ..... 167  
*Oryza sativa* ..... 108, 114–115  
 Oscillatory systems study, in biochemistry ..... 300–304  
     *See also* Biochemical control theory (BCT),  
         for biochemical network

**P**

Pairwise clustering algorithms..... 482  
 Paralogous pairs ..... 76  
 Particle transport dynamics, role of ..... 314  
 Partitional clustering algorithms ..... 481–482  
     *See also* Cluster algorithms  
 Pathway hole, definition ..... 374  
 Pathway motifs, linear analysis  
     branched pathways..... 283–285  
     cyclic systems ..... 285–287  
     negative feedback of..... 287–290  
     straight chains ..... 281–283  
     *See also* Biochemical control theory (BCT),  
         for biochemical network

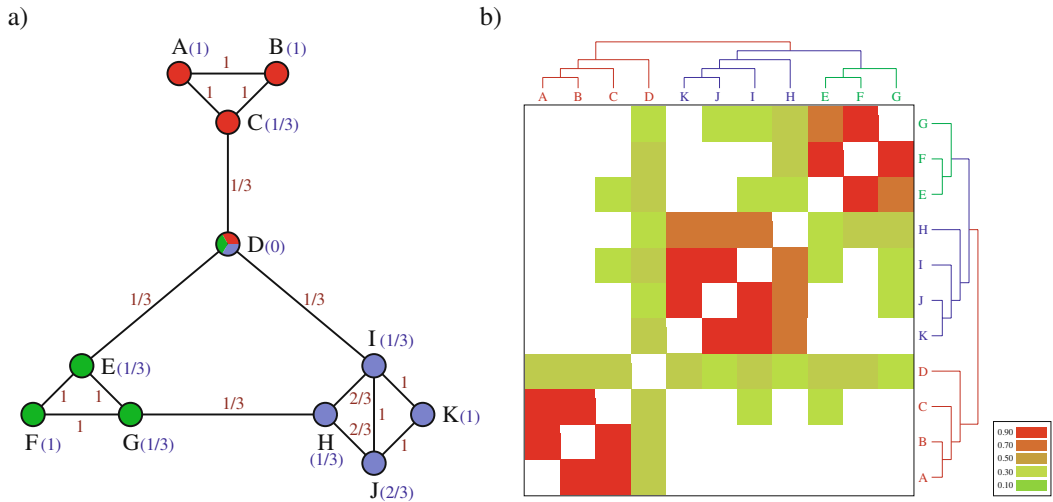
- Pathway prediction, role ..... 374
- PathwayPro algorithm ..... 251, 254
- biological applications of
    - disease genes identification and drug targets of leukemia ..... 261–265
    - ligand–receptor pairs, co-expression of ..... 257–261
- PDB, *see* Protein Data Bank
- Pearson correlation
- coefficient, role of ..... 250, 468
  - for mRNA expression ..... 493
- Pearson correlation coefficient (PCC) values ..... 7, 13–15
- Perl programming language ..... 5
- Pfam classification test, for protein domains ..... 393–397
- Phage  $\lambda$
- bistable switch ..... 338
  - promoter regulation in
    - computational modeling of data ..... 350–353
    - experimental design in ..... 348–350
  - See also* Gene regulatory networks, data collection and modeling in
- Physical interaction network, definition of ..... 226
- Physical models, construction of ..... 235
- Poisson random variable, definition of ..... 322
- Poisson  $\tau$ L method ..... 322
- Position-specific scoring matrix ..... 455
- Position weight matrix (PWM) ..... 25
- Post-genomic research, challenges in ..... 249–250
- PostgreSQL database ..... 558
- PPI, *see* Protein–protein interaction
- Predictor ..... 204
- $P_{RM}$  promoter, autoregulation of ..... 349
- Probabilistic Boolean network algorithm ..... 251
- Probabilistic model, in intron evolution ..... 361–363
- See also* Eukaryotic gene structure
- Probability based MDL postpruning, in model
- assigns ..... 415
- Probability weighted dynamic Monte Carlo method ..... 321
- Prokaryotes ..... 112, 116
- Prokaryotic promoters, measurement of ..... 340–341
- See also* Gene regulatory networks, data collection and modeling in
- Propensity function, definition of ..... 319
- Protein
- bioverse in ..... 516–518
  - classifications of ..... 381
  - data visualization ..... 458–459
  - and DNA interactions, measurement of ..... 229
  - evolution, model of
    - conservation and variation, sources of ..... 425
    - genomic proximity in ..... 426
    - protein duplication, consequences of ..... 425–426
  - family, information content of features in ..... 397–400
  - feature selection in ..... 457–458
  - identification of disease-specific ..... 453–454
  - interactions
    - data integration ..... 452–453
    - data, steps of ..... 451–452
    - genetic variation and ..... 450
    - mutation analysis in ..... 451
    - and ligand interactions ..... 421–422
    - sequence, function prediction in ..... 376–377
    - SNP function prediction ..... 455
    - SVM model ..... 455–457
    - See also* Enzyme class prediction
- Protein Data Bank ..... 514
- Protein domains and family, public databases ..... 134–136
- protein–protein BLAST (BLASTP) ..... 117–118
- Protein–protein interaction ..... 553
- Protein–protein interactions, co-evolution model
- basic model
    - break-up of interactions by evidence codes .... 69–71
    - co-evolution algorithm ..... 67
    - computing correlated divergence ..... 67–72
    - construction of data set of non-interacting proteins ..... 72
    - construction of multiple sequence alignments ..... 72
    - correlated approach ..... 66
    - Pearson correlation coefficient, drawbacks ..... 74–76
    - signals of correlated divergence ..... 72–73
  - co-evolution approaches ..... 64–65
  - hybrid approaches ..... 65
  - improved model
    - organism subset selection ..... 81–83
    - protein subset selection ..... 76–81
  - relational data mining inference ..... 63–64
  - structure-based approaches ..... 62–63
- Protein–protein interactions, computational reconstruction
- BioGRID method ..... 96–97
  - decision tree analysis ..... 94
  - effect of size and bias in experimental interactome maps ..... 95–96
  - genomic features ..... 91–92, 95
  - gold-standard positive and negative
    - data sets ..... 91, 94–95
  - logistic regression analysis ..... 92–93
  - Naïve Bayes assumption ..... 92
  - nonlinear continuous and graph-based
    - features ..... 93–94
  - random forest analysis ..... 94
  - regularization ..... 93
  - support vector machine (SVM)
    - and Adaboosting ..... 93
- Protein–protein interactions, prediction
- materials ..... 44–46
  - methods
    - domain combinations ..... 46
    - high quality interactions ..... 53–57



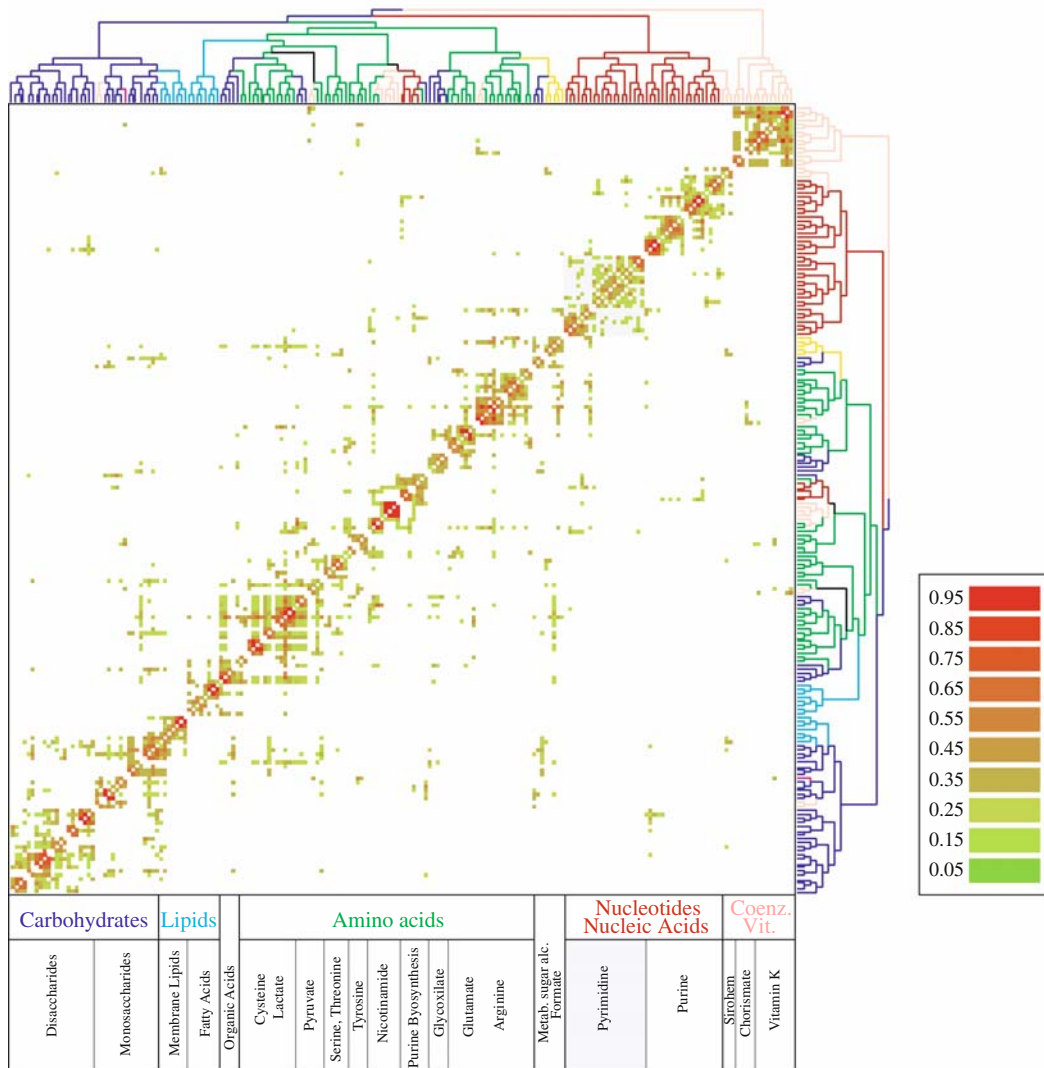
MSSC algorithm .....	51–53	Routh-Hurwitz table, advantage of.....	297
MSSC problem.....	49–51	RR, <i>see</i> Response regulator domain	
of prediction.....	47–48	RREs, <i>see</i> Reaction rate equations	
protein cover problem.....	46		
set-cover problem.....	48–49		
Protein subcellular localization		<b>S</b>	
experiments, database .....	133	<i>Saccharomyces cerevisiae</i> (Yeast) .....	3, 108–109,
sources.....	134	111, 114–115, 130, 146, 227, 463, 465	
ProteoLens database and protein interactions.....	458	<i>Saccharomyces cerevisiae</i> Genome Database .....	227, 514
<i>See also</i> Protein		<i>Saccharomyces</i> Genome Database (SGD).....	8, 11
Protinfo server and protein structure.....	518	SAM model, in protein domains.....	402
PSI-BLAST .....	85	SBML, <i>see</i> Systems Biology Markup Language	
algorithm.....	516	SCOP, <i>see</i> Structural Classification of Proteins	
database and SVM.....	455–456	SCPD.....	104, 106
role of.....	379	SDRs, <i>see</i> Specificity determining residues	
PSI-MI database.....	539	SDT, <i>see</i> Stochastic decision trees	
PSIPRED algorithm .....	516	Search query, molecules in.....	522
PSSM, <i>see</i> Position-specific scoring matrix		<i>See also</i> Web application	
P $\tau$ L methods, <i>see</i> Poisson $\tau$ L method		SEBINI, <i>see</i> Software Environment for Biological	
<i>p</i> -value		Network Inference	
index and k-means algorithm.....	500–501	SELEX .....	24
pathways in .....	505	Sequence analysis	
PW-DMC method, <i>see</i> Probability weighted dynamic		importance of.....	376–377
Monte Carlo method		problems in .....	231
Python programming language		SEQUEST peptide mass spectra	
API in .....	527	analysis algorithm .....	567
operating system of.....	534	SGD, <i>see</i> <i>Saccharomyces cerevisiae</i> Genome Database	
XML-RPC code in.....	526	SGD 267, ORFs verification in .....	227
<b>R</b>		SHAKE .....	30, 33
Random access memory (RAM) .....	4	Shrinkage parameter .....	209
<i>Rattus norvegicus</i> .....	108	Simple genetic circuits.....	292
Reaction rate equations.....	312	cascade of .....	294
development of.....	318	low-pass frequency response of.....	293
Receiver operating characteristic .....	389, 472	Simple interaction file (SIF).....	7, 8
Reciprocal Best BLAST Hit (RBH) algorithm.....	80	Simple linear reaction chains, frequency	
$\gamma$ -Recursion, computation of.....	365	analysis of.....	292–294
Recursive feature elimination .....	457	Simple negative feedback, frequency	
Red Hat Linux.....	558	analysis of.....	294–295
RegulonDB.....	104, 108	Simulation algorithm, for steady-state analysis.....	256
Relaxation oscillators .....	302–303	Size Scalable Model.....	326
<i>See also</i> Biochemical control theory (BCT), for		Smith-Waterman algorithm.....	118
biochemical network		SNP function prediction, in protein.....	455
REMORA web services, in bioinformatics tools.....	536	<i>See also</i> Protein	
Response regulator domain .....	434	Software architecture .....	556–560
Ala-Scan of.....	441	Software Environment for Biological Network	
RFE, <i>see</i> Recursive feature elimination		Inference .....	551
<i>Rhodospseudomonas palustris</i>		advantages of.....	553–554
CABIN role in.....	561	analysis flow of control .....	563–569
protein-protein interaction in.....	565, 568	architecture in .....	556–559
RMA, <i>see</i> Robust Multi-Array Analysis		and CABIN system .....	559
Robust Multi-Array Analysis .....	258	capabilities in .....	560–563
ROC, <i>see</i> Receiver operating characteristic		inference algorithms and.....	553
Routh-Hurwitz criterion .....	297	network subsets and .....	562
		role of .....	552
		Specificity-determining amino acids .....	425

- Specificity determining residues  
  in ATPase and DD domains ..... 437–440  
  in class Ia aaRSs ..... 443–446  
  conservation and variation, sources of ..... 425  
  genomic proximity in ..... 426  
  identification methods ..... 422–424  
  modeled structures ..... 440–441  
  procedures in determining ..... 426  
    evolution pattern of ..... 427  
    multiple alignments and bacterial  
      two-component system alignment ..... 428–429  
    mutual information calculation ..... 429  
    ortholog detection and clustering ..... 427–428  
    statistical significance estimation ..... 430–433  
  protein duplication, consequences of ..... 425–426  
  in RR domain ..... 435–437
- Spectral clustering algorithms ..... 483–484  
  *See also* Cluster algorithms
- Spectral clustering, of expression data ..... 496–497  
  *See also* Cluster algorithms
- SPEL method ..... 424
- SPR, *see* Surface patch ranking
- SQL, *see* Structured Query Language
- SSAs, *see* Stochastic simulation algorithms
- SSM, *see* Size Scalable Model
- Steady-state solution, for biochemical network ..... 275
- Steepest Descent ..... 30
- Stochastic approach, for cellular networks ..... 271
- Stochastic decision trees ..... 379  
  for protein modeling  
    evaluation of ..... 388–390  
    framework of ..... 387–388
- Stochastic differential equation ..... 325
- Stochastic simulation algorithms ..... 314, 317, 319  
  development of ..... 320  
  fundamentals of ..... 319–320  
  future perspectives of ..... 327, 330–332  
  steps of ..... 319–320  
  *See also* Biological systems, kinetic modeling of
- Stoichiometric networks, for biochemical network  
  analysis ..... 271–275
- Strict clustering  
  of expression data ..... 495–496  
  of noisy data ..... 484–485  
  *See also* Cluster algorithms
- Structural Classification of Proteins ..... 514  
  database and enzyme ..... 377
- Structured Query Language ..... 537
- Subset of network, isolation of ..... 343  
  *See also* Gene regulatory networks, data collection  
    and modeling in
- Summation theorem ..... 289
- Sum of pairs (SOP) function ..... 78, 82
- Supervised learning  
  genomics data set functional bias on ..... 470–472  
  reference set functional bias on ..... 468–470  
    *See also* Gene network model
- Supervised learning technique, importance of ..... 464
- Support vector machine ..... 379, 424, 455–456
- Surface patch ranking ..... 424
- SVM, *see* Support vector machine
- Swiss-Prot database, extraction of SNP data  
  and mutation ..... 453
- SWISS-PROT functional assignments ..... 424
- Swiss-Prot mutations and protein interactions ..... 450
- SynTreN network and SEBINI ..... 563
- Systems Biology Markup Language ..... 538–539
- ## T
- TAS, *see* Traceable author statement
- Tau-leaping and PW-DMC methods,  
  difference of ..... 321–322
- Term-term similarity ..... 8
- TF, *see* Transcription factor
- Thermodynamic equilibrium, for biochemical  
  network ..... 275
- TIGR Rice Genome Annotation database ..... 108
- TIP3P ..... 29
- TP53\_HUMAN protein ..... 454
- Traceable author statement ..... 465
- Traceable Author Statement (TAS) ..... 8
- Trajectory analysis ..... 317
- Transcriptional regulation networks, global models  
  challenges ..... 188–190  
  design principle characteristics ..... 187–188  
  methods  
    Bayesian network approaches ..... 193–195, 204  
    ordinary differential equations (ODEs) ..... 192  
    summary of the ten baseline variables ..... 207  
    using Inferelator algorithm ..... 195–202  
  properties of good solution ..... 190–191  
  review of functional-genomics  
    direct assays of protein-DNA interactions ..... 184–185  
    functional annotations ..... 185–186  
    microarrays ..... 183–184  
    protein-protein interactions ..... 186  
    systems biology data quality ..... 186–187
- Transcriptional regulation networks, reconstruction  
  and comparison  
  computational resources  
    genome sequences ..... 176  
    hardware ..... 175  
    software ..... 175  
    templates ..... 175–176  
  procedure  
    analysis of conserved regulatory interactions ..... 170  
    analysis of global network structure ..... 172–173  
    analysis of local network structure ..... 170–172

attributes of lifestyle class of an organism ..	173–175
creation of random networks .....	168–169
gene analysis and regulatory interactions....	169–170
network reconstruction procedure .....	166–168
structure .....	166
Transcription factor .....	229
Transcription factor–binding sites (TFBSs).....	3
Transcription factor–binding sites, prediction	
benchmarking .....	121–127
computational costs and monetary equivalent.....	34–35
computational resources.....	27–28
conversion into sequence logo .....	26–27
determination of .....	24
homology-based.....	103
method	
heating and equilibration of structure .....	30–31
introduction of counter ions .....	30, 37
role of water in binding of protein	
and DNA .....	29–30, 37
simulation of “multi-copy”	
structures .....	31–33, 37–39
simulation protocols for free energy	
calculations.....	31–33
simulation protocols for protein–DNA	
complex .....	28–31
systematic and statistical errors.....	39–40
mutational free energies, calculation of .....	25
Python script for .....	120–121
regulatory protein–DNA and protein–protein	
interactions	
additional data sets.....	116–117
data sets.....	104–116
determination of similarity among protein	
sequences.....	117–120
procedures .....	103–104
relative binding free energy, determination of .....	24–25
similarities in source organism and target	
organism.....	102
sources of experimental.....	105
TRANSFAC <sup>®</sup> 7.0 .....	104–106, 111, 116, 129, 132–133
Transformation .....	203
Transient state, for biochemical network .....	275
TRIPLES database.....	116
<b>U</b>	
UbiA family, tree for.....	396
UCSC Genome Browser .....	130
UniProt Knowledgebase and OPHID proteins .....	452
UNIX workstation .....	4–5
Urinating frequently (UF) .....	193
<b>V</b>	
Variable Connectivity Model .....	326
VCM, <i>see</i> Variable Connectivity Model	
<b>W</b>	
Watson–Crick base pairs.....	25
Web application .....	518–523
WormBase .....	104, 107–108
WormMart .....	104, 108
<b>X</b>	
XML-RPC protocol, in bioverse API.....	526
<i>See also</i> Application Programming Interface	
<b>Y</b>	
YBR059C gene .....	218
Yeast cell-cycle gene co-expression, and GO intersection	
network .....	9
Yeast galactose expression data, mapping of .....	234
Yeast galactose-utilization pathway, cluster	
analysis in .....	232
yFiles Layouts plug-in .....	10
YJR123W gene .....	217

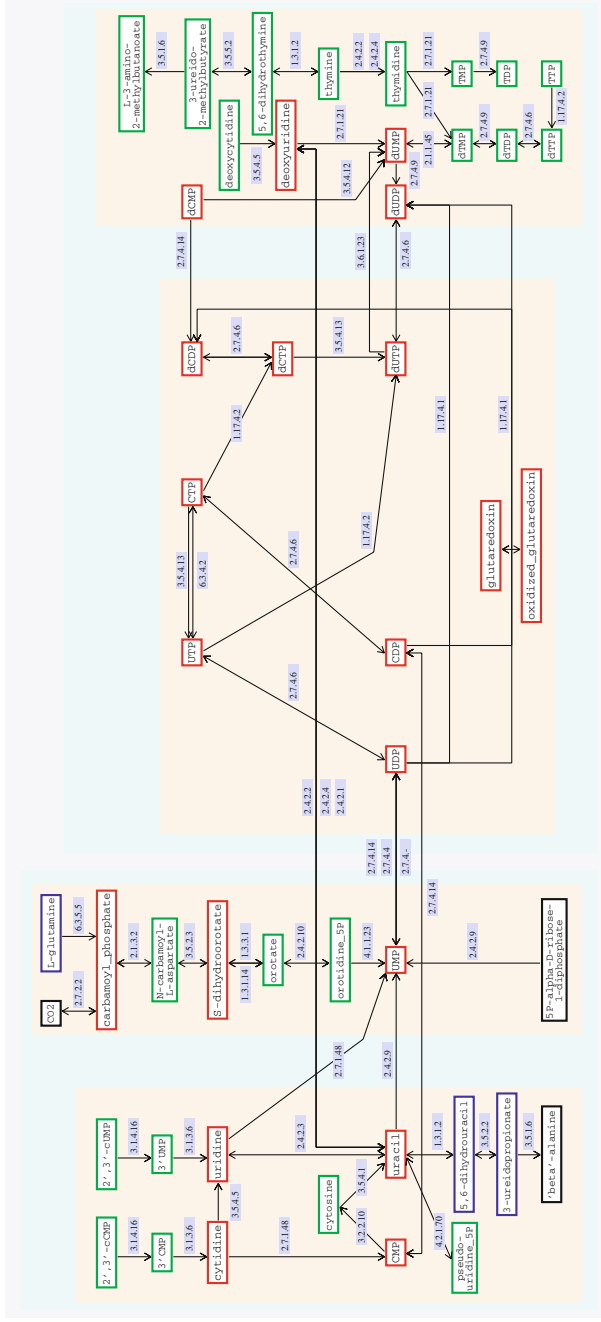
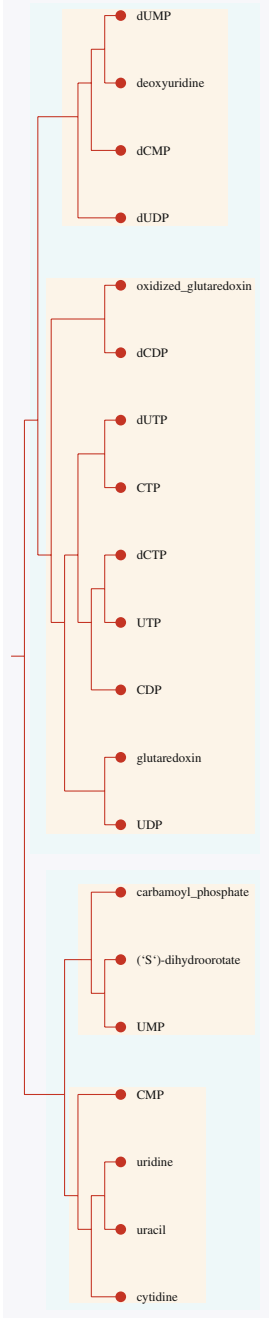


**Color Plate 1.** Uncovering the underlying modularity of a complex network. **(a)** Topological overlap illustrated on a small hypothetical network. On each link, we indicate the topological overlap for the connected nodes; and in parentheses next to each node, we indicate the node's clustering coefficient. **(b)** The topological overlap matrix corresponding to the small network shown in **(a)**. The *rows and columns* of the matrix were reordered by the application of an average linkage clustering method to its elements, allowing us to identify and place close to each other those nodes that have high topological overlap. The *color code* denotes the degree of topological overlap between the nodes. The associated tree reflects the three distinct modules built into the model, as well as the fact that the EFG and HIJK modules are closer to each other in the topological sense than to the ABC module (Chapter 7, Fig. 3; *see* discussion on p. 151).

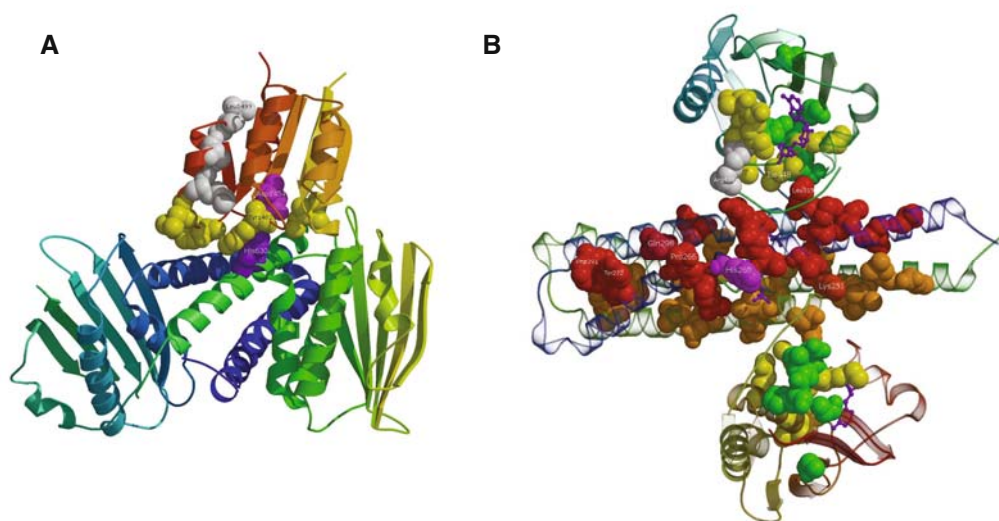


**Color Plate 2.** Topological modules in the *Escherichia coli* metabolism: the topologic overlap matrix, together with the corresponding hierarchical tree (*top and right*) that quantifies the relation between the different modules. The branches of the tree are *color-coded* to reflect the predominant biochemical classification of their substrates. The *color code* of the matrix denotes the degree of topological overlap shown in the matrix. The large-scale functional map of the metabolism, as suggested by the hierarchical tree, is also shown (*bottom*) (Chapter 7, Fig. 5; *see* discussion on p. 153).

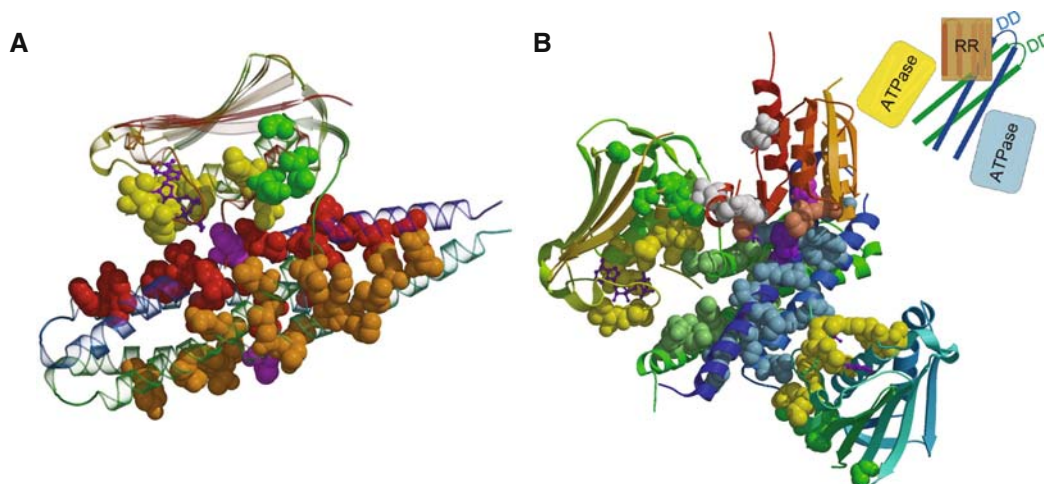
# Pyrimidine Metabolism



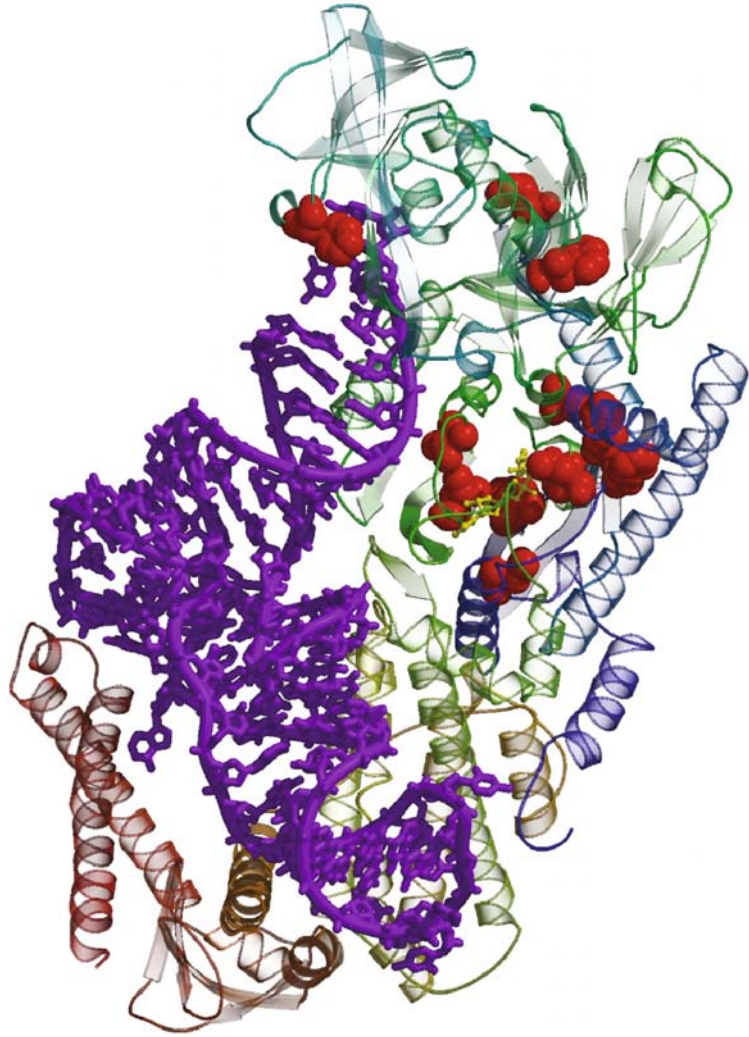
**Color Plate 3.** Enlarged view of the substrate module of pyrimidine metabolism, along with a detailed diagram of the metabolic reactions that surround and incorporate it. The colored boxes in the background denote the first two levels of the three levels of nested modularity suggested by the hierarchical tree. Red-outlined boxes denote the substrates directly appearing in the reduced metabolism and thus on the tree (Chapter 7, Fig. 6; see discussion on p. 154 and full caption on p. 155).



**Color Plate 4.** Structural localization of putative SDRs and CERs in two-component system domains. (a) RR Spo0F (*red-brown ribbon*) bound to structural analog of the DD in Spo0B protein. The conserved His is shown in *purple*, the conserved Asp in RR in *magenta*. SDRs and CERs are shown in *yellow* or, when located on the  $\alpha 4$  helix, in *white* (PDB entry 1F51). (b) The non-catalytic conformation of HK homodimer. ADP is shown as a *purple* wireframe, the phosphate-accepting conserved His residue in *magenta spacefill*. SDRs and CERs on the ATPase are shown in *yellow*, or in *white* if located on the unresolved ATP-lid loop that was superimposed from PhoQ kinase (PDB entry 1ID0), or in *green* in the RR-specific CERs side patch. SDRs and CERs on the DD are shown in *red* on one homodimer and *orange* on another (PDB entry 2C2A) (Chapter 18, Fig. 6; *see* discussion on p. 435).



**Color Plate 5.** Localization of putative SDRs and CERs on computationally obtained models (models provided by Marina et al (27)). (a) HK in the active conformation, the ATPase is docked on the DD so that transfer of the phosphoryl group is possible. SDRs and CERs on the ATPase domain are shown in *yellow* or *green* when located in the RR-specific CERs side patch. SDRs and CERs on the DD are shown in *red* on one homodimer and orange on another. (b) Spo0F computationally docked on HK and subsequently superimposed with RR from OmpR. RR (*brown-red ribbon*) (PDB entry 1KGS) with its  $\alpha 4$  helix swung  $\sim 90^\circ$ ; the phosphorylated Asp in the RR is shown in *magenta*, SDR and CERs are shown in *light red* or, when located the  $\alpha 4$  helix in *white*. DD (*dark blue and dark green ribbon*): SDRs and CERs are shown in *light blue* on one dimer and in *light green* on another. ATPase (*yellow-green ribbon* on the left and *light-blue* on the right): the colors are the same as in (a) (Chapter 18, Fig. 7; *see* discussion on p. 439).



**Color Plate 6.** Valine aminoacyl-tRNA synthetase (PDB entry *1GAX*). The tRNA is shown as a *purple* wireframe structure, SDRs and CERs are *red* balls, and amino acid (valyl-adenylate analog) is in *yellow* wireframe (Chapter 18, Fig. 8; *see* discussion on p. 443).