Math 4/5779: Mathematics Clinic
Pathway Inference: Computational Issues

Supplement: Bibliography

Taught by Harvey J. Greenberg

Assisted by Raphael Bar-Or, Lance Lana, Cary Miller, Tod Morrison and Christiaan van Woudenberg

Sponsored by
DMI BioSciences, Genomica and Tech-X

Spring Semester 2002

May 23, 2002

This supplement contains the complete bibliography, most of which were compiled before the semester began. Student annotations have been left separate, rather than merged. Unsigned annotations are by Harvey J. Greenberg, most of which were done early to serve as examples.

Bibliography

[1] T. Akutsu, S. Miyano, and S. Kuhara. Identification of genetic networks from a small number of gene expression patterns under the boolean network model. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 4, pages 17–28, http://www-smi.stanford.edu/projects/helix/psb01/, 1999. World Wide Web.

> The *boolean network model* infers genetic network architectures from state transition tables, which correspond to time series of gene expression patterns. This article reports the results of computational experiments, suggest that a small number of stage transition (INPUT/OUTPUT) pairs are sufficient in order to infer the original Boolean network correctly. It goes on to discuss the practical usefulness of this model.
> — Dung T. Ngyuen

> This paper proposes an algorithm for inferring genetic network architectures from state transition tables which correspond to time series of gene expression patterns, using the Boolean network model. It is argued that if the indegree of each node is bounded by a constant, only $O(\log n)$ state transition pairs are necessary and suficient to identify, with high probability, the original Boolean network of n nodes correctly. The paper describes the computational experiments executed in order to expose the constant factor involved in $O(\log n)$ notation. The computational results are used to show that a Boolean network of size 100,000 can be identified by their algorithm from about 100 INPUT/OUTPUT pairs if the maximum indegree is bounded by 2. The paper claims that the algorithm is conceptually so simple that it is extensible for more realistic network models.
> — Adolfo Perez-Duran

[2] M. Arita. Metabolic reconstruction using shortest paths. *Simulation Practice and Theory*, 8:109–125, 2000.

> This explains bottom-up fashion of biological models to reconstruct. It considers all possible consequences from observed laboratory data to check their validation. The AMR(automated metabolic reconstruction) is a simulation system for the metabolic reconstruction. The graph-oriented

representation that uses node and edges expresses atoms and enzymatic mappings, so it considers atomic level, not compound level. The shortest path algorithm efficiently searches possible pathways. AMR system generates hypothetical metabolic pathways from compound structures and reaction formulas.
— Min Hong

[3] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

This paper describes the efforts of the Gene Ontology Consortium to construct a universal ontology. The authors point out the strengths of their ontology, one of which is its independency from any type of organism. A controlled language to describe processes, cellular components and molecular functions is provided and discussed in this paper. The general high acceptance of the concepts and axiomes used to build this ontology already make it the most widely used controlled language in molecular biology.
— Jens Eberlein

The authors propose a unifying vocabulary for various aspects of eukaryote biology. They create three distinct hierarchies: biological process, molecular function and cellular component. A biological process uses molecular functions to transform something. A molecular function is the biochemical activity of a gene product. The cellular component refers to the area where the gene product is active. While one could imagine there is a single gene product unifying entries in each hierarchy, a gene product might actually have multiple assignments. It might be involved in various biological processes, perform several molecular functions and do so at different parts of the cell.
— Aaron Gabow

[4] A.R. Asthagiri and D.A. Lauffenburger. A computational study of feedback effects on signal dynamics in a mitogen-activated portein kinase (mapk) pathway model. *Biotechnology Progress*, 17:227–239, 2001.

[5] G. Bader and C. Hogue. BIND: A data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics*, 16(5):465–477, 2000.

The Biomolecular Interaction Network Database (BIND) is a standard data specification that stores information in all its detail and allows efficient cross-platform transfer of data using the internationally standard ASN. Syntax is the key point in this paper. There are three main types of data objects in the BIND specification: interactions, molecular complex and pathway. It also contains useful database management and data exchange objects. A BIND interaction record is based on the interaction between two objects, such as protein, DNA, RNA, ligand, molecular complex or an interaction. An interaction is described by cellular location, experimental conditions used to observe the interaction, conserved sequence, molecular location, chemical action, kinetics, thermodynamics, and chemical state. An interaction contains an NCBI date object, a sequence of updates for an audit trail, an interaction identifier (IID) accession number, two interacting molecules (BIND-objects), a description of the interaction, a series of publications and a private flag. The BIND ID number space is controlled using a unique key server.
— Xuan Le

[6] G.D. Bader, I. Donaldson, C. Wolting, B.F.F. Ouellette, and T. Pawson. BIND — the biomolecular interaction network database. *Nucleic Acids Research*, 29(1):242–245, 2001.

[7] J.E. Bailey. Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. *Biotechnology Progress*, 14:8–20, 1998.

[8] P.G. Baker, C.A. Goble, S. Bechhofer, N.W. Paton, R. Stevens, and A. Brass. An ontology for bioinformatics applications. *Bioinformatics*, 15(6):510–520, 1999.

The authors give an overview of the ontology for the Transparent Access to Multiple Biological Information Sources

(TAMBIS). In particular, they claim that an ontology is needed that reflects the view of the data. Furthermore, they claim that Description Logics (DLs), which are a form of in silico knowledge representation, have the flexibility, consistency, and other necessary properties to construct a bioinformatic ontology that can be used to make inferences. The authors compare DLs to frame representations, as used in the EcoCyc database. They claim that one way in which DLs have an advantage is that not only do they have a mechanism for capturing declarative knowledge, but DLs also have a built-in classifier that allows for reasoning. In a frame representation the hierarchy of concepts is static and built in by the modeler. In a DL representation a concept is inserted into the hierarchy by the classifier and may be reclassified when new information is given about the concept. Additionally, a concept may have more than one parent in a DL, as opposed to just one, giving the DL greater flexibility. This is important because often times in biology one concept can be viewed in many different ways. The TAMBIS ontology was written in the GRAIL language, and this paper gives a very through description how GRAIL enables the TAMBIS ontology to work through the use of assertions, operations, reasoning services, and sanctions. The authors mention that currently the TAMBIS ontology has more breadth than depth and they discuss the known limitations of DLs and of GRAIL in particular. The primary aim of TAMBIS is to develop a system flexible enough to perform some of the tasks of a domain expert.
— Tessa F. Weinstein

[9] M.Y. Becker and I. Rojas. A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, 17(5):461–467, 2001.

In this article, the authors present a dynamically generated graph layout algorithm that is designed to handle cyclic, partially cyclic, linear, and branched metabolic pathways. The authors propose an algorithm that specifically deals with the main nodes of complex pathways without graphing the side reactions. By using a recursive algorithm, the graph representing the metabolic pathway to be displayed is partitioned into subgraphs, which have simple display methods. The authors also describe the use of a spring embedding algorithm to position the primary parts of the graph relative to one

another. While the algorithm was only tested on five pathways the results were promising.
Lance Lana et al.

[10] BioCarta pathways database and software. World Wide Web, http://www.biocarta.com/genes/.

[11] J.M. Bower and H. Bolouri, editors. *Computational Modeling of Genetic and Biochemical Networks.* MIT Press, Cambridge, MA, 2001.

This collection of articles is very diverse. The first, "Modeling the Activity of Single Genes," by M.A. Gibson and E. Mjolsness, is a particularly insightful introduction. Topics include boolean networks, ODE models and stochastic models. Hybrid systems are discussed, along with relative merits of each approach, giving examples where each is preferred. The chapter on model reduction techniques [26] steps the reader through carefully examined examples, pointing out both good and bad points. Other chapters bear further study, as they deal with different aspects of modeling genetic and biochemical networks.
— Cary Miller [HG, ed.]

[12] Cell signaling networks database (CSN). World Wide Web, http://geo.nihs.go.jp/csndb/.

This is not only a database, but also a knowledgebase for signaling pathways of human cells. It compiles the information on biological molecules, sequences, structures, functions and biological reactions that transfer the cellular signals. The source of the data is from articles in *Nature*, *Science* and *Nature Cell Biology*.

[13] C. Chassagnole, D.A. Fell, B. Raïs, B. Kudla, and J.-P. Maza. Control of the threonine synthesis pathway in *Escherichia coli*: a theoretical and experimental approach. *Biochemistry Journal*, 356:433–444, 2001. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[14] C. Chassagnole, B. Raïs, E. Quentin, D.A. Fell, and J.-P. Maza. An integrated study of threonine-pathway enzyme kinetics in *Escherichia coli*. *Biochemistry Journal*, 356:415–423, 2001. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[15] M.W. Covert, C.H. Schilling, and B. Ø. Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213:73–88, 2001.

[16] F.W. Cummings. A model of pattern formation based on signaling pathways. *Journal of Theoretical Biology*, 207:107–116, 2000.

A model of pattern formation in the early embryo is presented that depends on the family of Wnt kinase and RPTP phosphatase signaling pathways. The question asked is how such signaling molecules, consisting of kinases and phatases, are able to make complex patterns. This process is modeled as a system of PDEs, which includes a combination of the Laplacian and nonlinear Helmholtz equations. The Laplacian operator on a curved epithelial surface is called the Beltrami operator. Numerical results are given for the 1-D case, that analyze which regions set apart the binary gene selection of a particular pattern. A discussion highlights similarities between physical and biological processes and how these famous equations apply to both.
— Rico Argentati

[17] E. Davidson. *Genomic Regulatory Systems: Development and Evolution*. Academic Press, Reading, MA, 2001.

This book takes the theory of genetic regulation developed over the past 50 years and reduces it to its most basic essence. When this is done the problem begins to resemble circuit design or reverse engineering of circuits. The basic idea is that a gene is controlled by *cis* and *trans* regulatory elements, with the former being binding binding sites or other DNA elements and the latter being transcription factors. The expression of a gene is determined by the combination regulatory elements acting on it and their orientation with respect to one another. The two running examples throughout the book are embryonic development in *Drosophila* and sea urchin. These two genera represent hugely divergent evolutionary cousins and so if the same things happen in both that is a good indication that the same probably happens in all multi-cellular creatures. These two model systems have

been used to uncover universal principles of animal development.
— Cary Miller

[18] P. D'haeseleer. *Reconstructing Gene Networks from Large Scale Gene Expression Data*. PhD thesis, University of New Mexico, Albuquerque, NM, 2000.

[19] P. D'haeseleer, S. Liang, and R. Somogy. Tutorial: Gene expression data analysis and modeling. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 4, http://psb.stanford.edu/psb99/genetutorial.pdf, 1999. World Wide Web.

This is a succinct introduction to the entitled topics. the color figures are very helpful in giving insight into gene expression data and the associated issues with the high dimensionality.

[20] P. D'haeseleer, S. Liang, and R. Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707–726, 2000.

This paper focuses on generating models that allow them to systematically derive predictions about important biological processes in disease, development and metabolic control. They use clustering of co-expression profiles, which allow them to infer shared regulatory inputs and functional pathways. The reverse engineering has the goal of identifying the causal relationships among gene products that determine important phenotypic parameters. Using the network inference, the goal of this project is to construct a coarse-scale model of the network of regulatory interactions among the genes.
— Xuan Le

[21] J.S. Edwards and B.Ø. Palsson. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *The Journal of Biological Chemistry*, 274(25):17410–17416, 1999.

[22] J.S. Edwards and B. Ø. Palsson. Robustness analysis of the textitEscherichia coli metabolic network. *Biotechnology Progress*, 16:927–939, 2000.

[23] L.B.M. Ellis, S.M. Speedie, and R. McLeish. Representing metabolic pathway information: an object-oriented approach. *Bioinformatics*, 14(9):803–806, 1998.

In this article the authors describe the development of the Compound, Organism, Reaction and Enzyme (CORE) database management system. CORE was programmed in Java and intended to encapsulate and serve the information in the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD). The UM-BBD contains pathway information concerning the biodegradation of compounds. The article details the development of CORE, the hardware and software used to build and access CORE, and the algorithms used. The authors claim that the primary way in which this database differs from other web based metabolic databases in that it contains dynamic search links to other relevant information. This difference becomes evident when we find that CORE produces textual pathway maps, where the text in the pathway provides a link to more information about the topic selected. The system's greatest strength is its flexibility and the ease with which it can be updated to include other information.
— Tessa F. Weinstein

[24] A. Enright and C. Ouzunis. BioLayout — an automatic graph layout for similarity visualization. *Bioinformatics*, 17(9):853–854, 2001.

The authors present a force-based graph layout algorithm solved with simulated annealing. They employ this rather standard algorithm to visualize protein families. Nodes represent proteins and repel each other. However, proteins that have sequence similarities have an edge between them that exerts an attractive force. After about 50 iterations of the algorithm, similar proteins end up clustered together. They also implement a 3-dimensional version that uses OpenGL. However, it appears this version lacks labeling facilities

[25] ENZYME database. World Wide Web, http://expasy.ch/enzyme/.

This is a repository of enzyme nomenclature, based upon the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). It describes each type of characterized enzyme for

which an Enzyme Commission number has been provided. The site contains a User Manual and other services.

[26] B. Ermentrout. Simplifying and reducing complex models. In J.M. Bower and H. Bolouri, editors, *Computational Modeling of Genetic and Biochemical Networks*, pages 307–322, Cambridge, MA, 2001. MIT Press.

[27] D.A. Fell. Systems properties of metabolic networks. In Y. Bar-Yam, editor, *Proceedings of the International Conference on Complex Systems*, pages 21–26, 1997. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[28] D.A. Fell. Signal transduction and the control of expression of enzyme activity. *Adv. Enzyme Reg. XXX must find full name*, 40:35–46, 2000.

[29] D.A. Fell, C. Chassagnole, and J-P. Mazat. Building a computer simulation of threonine synthesis in *Escherichia coli*. In T-M. Yi, M. Hucka, M. Morohashi, and H. Kitanno, editors, *Proceedings of the Second International Conference of Systems Biology*, http://www.icsb2001.org/toc.html, 2001. World Wide Web.

[30] D.A. Fell and A. Wagner. Structural properties of metabolic networks: implications for evolution and modelling of metabolism. In J.-H. S. Hofmeyr, J.M. Rohwer, and J.L. Snoep, editors, *Animating the cellular map*, pages 79–85. Stellenbosch University Press, 2000. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[31] A.E.N. Ferreira. PLAS, Version 1.2. with Voit [112] and at http://correio.cc.fc.ul.pt/~aenf/plas.html, 2000.

[32] A. Finney, V. Gor, B. Bornstein, E. Mjolsness, and H. Bolouri. Systems biology markup language (SBML) level 2 proposal: Array features. Technical report, California Institute of Technology, http://www.cdb.caltech.edu/erato/, 2002.

[33] A. Finney, V. Gor, B. Bornstein, E. Mjolsness, and H. Bolouri. Systems biology markup language (SBML) level 2 proposal: Miscellaneoous features. Technical report, California Institute of Technology, http://www.cdb.caltech.edu/erato/, 2002.

[34] C.V. Forst and K. Schulten. Evolution of metabolisms: A new method for comparison of metabolic pathways. In *RECOMB Proceedings*, pages 174–181, New York, NY, 1999. ACM.

> This introduces a new method based on a combination of sequence information with graph topology for comparison of metabolic pathways. The approach uses phylogenetic analysis to find relationships of different organisms. The relationship between individual biopolymers is measured by distances between substrates and enzymes. In particular, they apply the distance of the difference between ortholog and paralog and gap penalty technique for missing functional roles.
> — Min Hong

[35] J.J. Fox and C.C. Hill. From topology to dynamics in biochemical networks. IGERT program in nonlinear systems, Cornell University, Ithaca, NY, 2001.

[36] C. Friedman, P. Kra, H. Yu, M. Krauthammer, and A. Rzhetsky. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics*, 17(Supplement 1):S74–S82, 2001.

> GeneWays is a program that has the ability mine data from biological literature. The program enacts a literature search associated with a given interest. The program uses grammar as the procedure for data collection. In this manner, GENIES semantically searches and stores interactions and relationships between biological molecules.
> — Ben Perrone
>
> GENIES is a natural-language processing system that has been highly successful in specialized domains. The GENIE system extracts and structures information about cellular

pathways from the biological literature in accordance with a knowledge model that developed earlier. A modification of GENIE by implementing an existing medical natural language processing system is a powerful tool of molecular biology and medical field today. Both of these techniques acquired valuable data from biological journals.
— Dung T. Ngyuen

[37] S. Fuhrman, P. D'haeseleer, and R. Somogyi. Tracing genetic information flow from gene expression to pathways and molecular networks. In D. H. Geschwind, editor, *DNA Microarrays: The New Frontier in Gene Discovery and Gene Expression Analysis*, Short Course Syllabus, pages 57–66, Washington, DC, 1999. Society for Neuroscience, Society for Neuroscience.

[38] M.A. Gibson and E. Mjolsness. Modeling the activity of single genes. In J.M. Bower and H. Bolouri, editors, *Computational Modeling of Genetic and Biochemical Networks*, pages 1–48, Cambridge, MA, 2001. MIT Press.

This is a thorough and clear introduction to what genes are and how they function. It is not light reading, but it is especially useful in beginning to gain a non-superficial understanding of the underlying biology and biochemistry of such pathway inference issues as cell signaling.

[39] P.M. Gleiss, P.F. Stadler, A. Wagner, and D.A. Fell. Small cycles in small worlds. Working paper 00-10-058, The Sante Fe Institute, 1399 Hyde Park Road, Sante Fe, NM 87501, 2000. Note: available at http://www.santafe.edu/sfi/publications/00wplist.html.

[40] S. Goto, H. Bono, H. Ogata, W. Fujibuchi, T. Nishioka, K. Sato, and M. Kanehisa. Organizing and computing metabolic pathway data in terms of binary relations. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 2, pages 175–186, http://www-smi.stanford.edu/projects/helix/psb97/, 1997. World Wide Web.

At present, functional prediction of genes and genomes is done by searching similarities of each in the sequence database or the motif libraries and to extend sequence

similarities to functional similarities. The problem is the sequence-function relationships of molecules do not contain higher level information of how components are connected to form a functional unit, such as a metabolic pathway. Project KEGG (Kyoto Encyclopedia of Genes and Genomes) tries to computerize current knowledge of the information pathways of genes and gene products, which may be considered wiring diagrams of biological systems. The data of interacting molecules or genes is represented using the binary relations that correspond to the pair wise interaction. Interactions involving more than two components are approximated by a collection of pair wise interactions. KEGG has two major purposes, first to establish an integrated view of gene products, namely how they are interacting. Second, it will provide a practical tool for making assignments of enzyme genes from genomic sequences. The major feature of KEGG is its link capabilities, both in terms of linking to the existing databases and different organisms and in terms of the linking to compute a pathway from binary relations. By representing each reaction as a collection of binary relations KEGG tries to reproduce the known pathways.
— Adolfo Perez

[41] H.J. Greenberg. *Mathematical Programming Glossary*. World Wide Web, http://www.cudenver.edu/~hgreenbe/glossary/, 1996–2002.

[42] A.J. Hartemink, D.K. Gifford, T.S. Jaakkola, and R.A. Young. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 6, pages 422–433, http://www-smi.stanford.edu/projects/helix/psb01/, 2001. World Wide Web.

The authors describe a model-driven approach for the analysis of genomic expression data, which permit genetic regulatory networks to be represented in an interpretable form. The models are Bayesian networks that contain latent variables capturing unobserved factors, which describe arbitrarily complex relationships. The graph semantics permit annotated edges and score models.
— Ben Perrone

[43] J. Hasty, D. McMillen, F. Isaacs, and J. Collins. Computational studies of gene regulatory networks: *In Numero* molecular biology. *Nature Reviews Genetics*, 2:268–279, 2001.

Most of the gene regulatory schemes proposed offer qualitative information only. If a deep understanding of regulation is to be achieved, we require quantitative modeling techniques in concert with experiment. This is the basis for the authors' argument: an engineering approach is desirable. The paper goes on to describe the current state of such efforts and provides an extensive bibliography. Much has been achieved in the study of nonlinear dynamics and stochastic process. People are now applying this work to biosystems. Gene circuits are compared to electric circuits. Electric circuits are governed by well-understood physical laws and perterbation of such circuits has predictable effects. It is not yet clear whether gene circuits will ultimately be as predictable but nevertheless people are working on it. The understanding of electric circuits is based on a thourough knowledge of the individual components; resistors, capacitors, etc. That level of detailed knowledge is not there yet for the components of gene circuits; transcription factors, promoters, etc. Theoretical models without a connection to experiment are not very useful. Modelers have sometimes tended to let their imaginations run wild and made very elegant models that are untestable or just not useful for describing real biology. Most of the models here have an experimental component.

While it is important to have our models based in reality we will probably never get a complete understanding of gene circuits through the reverse-engineering approach implied. Now people are experimenting with *synthetic* gene networks. Such a network has the distinct advantage that we know exactly what went into it. Then we can systematically perturb the network and observe the results. It is hoped that such experimentation will yield insights into the workings of natural gene circuits.
— Cary Miller

[44] J. Heidel, J. Maloney, and C. Farrow. Finding cycles in synchronous boolean networks with applications to biochemical systems. *International Journal of Bifurcations and Chaos*, 2002 (to appear).

[45] J-H S. Hofmeyr. Metabolic control analysis in a nutshell. In M. Yi, M. Hucka, M. Morohashi, and H. Kitano, editors, *Proceedings of the Second International Symposium on Systems Biology*, pages 291–300, Madison, WI, 2001. Omnipress. available at http://www.icsb2001.org/Papers/19_icsb2001_distiller.pdf.

This presents a derivation of some of the important theorems of metabolic control analysis. Starting from a general kinetic model of coupled chemical reactions written a set of nonlinear differential equation, the author then considers the special case of the steady-state kinetic model. Using the steady-state model, he investigates how the steady-state variables change with respect to perturbations in one or more of the parameters. This is accomplished with a detail discussion of the differentiation of the steady-state equation, which is then used to derive several basic theorems of metabolic control analysis.
— Tod Morrison

[46] M. Hucka, A. Finney, H. Sauro, H. Bolouri, J. Doyle, and H. Kitano. The ERATO systems biology workbench: architectural evolution. In T-M. Yi, M. Hucka, M. Morohashi, and H. Kitanno, editors, *Proceedings of the Second International Conference of Systems Biology*, http://www.icsb2001.org/toc.html, 2001. World Wide Web.

[47] L. Hunter, editor. *Artificial Intelligence and Molecular Biology*, Cambridge, MA, 1993. MIT Press.

This is now available at http://www.aaai.org//Library/Books/Hunter/hunter.html. Chapter 1, by the editor, provides a good introduction to biology for computer scientists and mathematicians. This is highly recommended for a gentle, informative introduction. Other chapters of direct relevance are by Karp [51] and Mavrovouniotis [72].

[48] T. Ideker, T. Galitski, and L. Hood. A new approach to decoding life: Systems biology. *Annual Reviews of Genomics and Human Genetics*, 2:343–372, 2001.

[49] T. Ideker, V. Thorsson, J. Ranish, R. Christmas, and J. Buhler. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292:929–934, 2001.

[50] T.E. Ideker, V. Thorsson, and R.M. Karp. Discovery of regulatory interactions through perturbation: Inference and experimental design. Technical report, Institute for Systems Biology, Seattle, WA, 2000.

[51] P.D. Karp. A qualitative biochemistry and its application to the regulation of the tryptophan operon. In L.E. Hunter, editor, *Artificial Intelligence and Molecular Biology*, pages 289–324, Cambridge, MA, 1993. MIT Press.

This paper discusses the representation and simulation of biological information so that it can be used in various situations. This is done through the particular example of a bacterial gene regulation system, the tryptophan operon of *E. coli*. The possibility of improving simulation programs that predict the outcome of a gene regulation experiment is explored. The GENISM simulator is held up as a simulator that will efficiently do this. Karp explores its functions, possibilities and limitations. Some results of simulations run through GENISM are presented.
— Jennifer Phillips

The focus of this chapter is on the issues of representation and simulation of the gene regulation system of the trytophan operon of *E. coli*. The author presents a model, GENSIM, which describes the biochemical reactions that determine the expression of the genes, the reactions by which the genes direct the synthesis of enzymes, and the reactions catalyzed by these enzymes. He then presents a detailed discussion of the implementation of this model.
— Tod Morrison

[52] P.D. Karp. An ontology for biological function based on molecular interactions. *Bioinformatics*, 16(3):269–285, 2000.

This paper discusses *functional bioinformatics*, the idea of computation that involves the function of proteins. The ontology is a step toward a controlled, precise vocabulary for biology. Discusses the use of Lisp as the implementation language for the EcoCyc system and gives a series of example queries that could be used with EcoCyc. All example queries are relatively simple. The article suggests that more

complex queries are possible. Distinguishes 'local' and 'integrated' function for proteins.
— Cary Miller et al.

[53] P.D. Karp. Pathway databases: A case study in computational symbolic theories. *Science*, 293:2040–2044, 2001.

This article introduces the concept of pathway genome database (PGDB) as a method of describing biochemical pathways and their component reactions, enzymes and substrates. A PGDB includes pathway information as well as information about the complete genome of the organism. The EcoCyc project is provided as an example of a PGDB. EcoCyc is structured using an ontology of about 1000 classes. The PGDB consists of a network of interconnected frames. Each frame represents a biological object. The labeled connections between the frames represent semantic relationships between the objects. The key to this representation is devising an ontology that clearly defines the meaning of the different PGDB fields and provides ease of extension when new domain concepts are discovered. The authors include a discussion of a program called "PathoLogic," which was developed to predict the metabolic network from the genome of an organism. Prediction of pathway flux rates for the entire metabolic network of an organism is also discussed. The authors stress the importance of using database content in solving these computational problems. There are no known algorithms that can solve these problems without being coupled with an accurate and well-designed pathway DB.
— Lance Lana et al.

[54] P.D. Karp, M. Krummenacker, S. Paley, and J. Wagg. Integrated pathway/genome databases and their role in drug discovery. *Trends in Biotechnology*, 17(7):275–281, 1999.

Reviews the use of EcoCyc and PathoLogic to construct new pathways for prokaryotes other that E. coli. Pathways for new organisms are compared to E. coli and the annotated genome is used to populate the knowledge base. Pathways present in a new organism but not in E. coli are added by hand. Presents methods used for the identification of false information in the annotated genome.
— Cary Miller et al.

[55] P.D. Karp, C. Ouzounis, and S. Paley. HinCyc: A knowledge base of the complete genome and metabolic pathways of *H.influenzae*. In D.J. States, P. Agarwal, T. Gaasterland, L. Hunter, and R.F.Smith, editors, *Proceedings of the Fourth Intelligent Systems for Molecular Biology Conference* (ISMB), pages 116–124, Menlo Park, CA, 1996. AAAI Press.

> *Haemophilius Influenzae* is a rod-shaped bacteria that is the leading cause of meningitis in children, and the second most common cause of bacterial pneumonia. As a bacterial infection, it is treated with antibiotics. A good collection of references for *H.influenzae* can be found at http://www.bacteriamuseum.org/species/Hinfluenzae.shtml. Karp et al. apply the knowledge and methodology learned from creating the EcoCyc knowledge base to construct metabolic pathway representations for *H. influenzae*. They use a bootstrap approach to populate the new KB from gene data at The Institute for Genomic Research (TIGR) and infer pathways. The completed knowledge base contains data about genes, enzymes, reactions, pathways and compounds that occur in *H. influenzae*, and the functional homologues that occur in *E. coli*.
> — Raphael Bar-Or et al.
>
> In this article the authors describe a methodology for predicting which metabolic pathways that are present in *E. coli* are also present in H. *influenzae*. The EcoCyc knowledge base (KB) contains information regarding the known metabolic pathways of *E. coli*. Recently, the genomic sequence of H. *influenzae* was completed, allowing the authors to develop a process by which they compare H. *influenzae* to *E. coli* and store the prediction results in the HinCyc KB. First, the authors created gene objects in HinCyc and compared them to *E. coli* proteins, those with homologous sequences were obtained. Next, they created polypeptide objects and then a protein-complex object. The authors note that these objects were created in a conservative fashion so that the resulting prediction would be minimal but accurate. They then determine what reactions are catalyzed by the H. *influenzae* enzymes, and finally make pathway predictions. They discuss the limitations of their method, related work, and lastly suggest how better information might make pro-

cess more accurate.
— Tessa F. Weinstein

[56] P.D. Karp and S. Paley. Integrated access to metabolic and genomic data. *Journal of Computational Biology*, 3(1):191–212, 1996.

This article addresses the EcoCyc system, which consists of a graphical user interface (GUI) that provides integrated access to a metabolic and genomic knowledge base (KB) for *E. coli*. EcoCyc is the result of joint work between SRI International and the Marine Biology Laboratory (MBL). The authors advocate their choice of implementing the KB as a frame based system called "HyperTheo." They argue that this is a very efficient and rapid way of implementing the KB because it adapts easily to the ever-changing world of biology. The authors also discuss their dynamically generated display of signaling pathways as a mixture of hypertext and gif files, using a tool they developed to retrofit CLIM applications to be displayed on the World Wide Web. Several examples are presented that include an enzyme display for 2-dehydro-3-deoxphosphoheptonate aldolase, graphical depictions of biosynthetic pathways for threonine, phenylalanine and tyrosine, as well as a genomic map display. They further describe the use of a user-friendly but very restricted GUI to query the KB. Only pre-formed options can be used to query the KB from the GUI, which prohibits complex Boolean searches.
— Lance Lana et al.

[57] P.D. Karp and S.M. Paley. Automated drawing of metabolic pathways. In H. Lim, C. Cantor, and R. Robbins, editors, *Proceedings of the Third International Conference on Bioinformatics and Genome Research*, 1994. Note: reprinted in 2000 at SRI.

[58] P.D. Karp and S.M. Paley. Representations of metabolic knowledge: Pathways. In R. Altman, D. Brutlag, P. Karp, R. Lathrop, and D. Searls, editors, *Proceedings of the Second Intelligent Systems for Molecular Biology Conference* (ISMB), pages 203–211, Menlo Park, CA, 1994. AAAI Press.

Karp and Paley discuss the formulation of a database schema and a set of accompanying software tools which form the knowledge base known as EcoCyc. The core problem is one

of path representation; how can one most efficiently store pathway data and still construct complete pathways? The authors discuss the problems encountered during their formulation of the knowledge base and propose solutions to achieve a complete reconstruction of the known pathways in *E. coli*. Relationships between reactions is stored in a predecessor list, containing tuples of a reaction and its predecessor compound in the pathway. Main versus side compounds and direction in individual reactions is determined by the predecessor list along with a list of heuristics.
— Raphael Bar-Or et al.

In this article Karp and Paley tackle the difficult problem of using information in the EcoCyc knowledge base (KB) to obtain a pathway graph, with the ultimate goal of using the pathway graph to create a pathway map. The EcoCyc knowledge base is a database (DB) that has information pertaining to the genes and intermediary metabolism of *E. coli*. Specifically, it contains information about the genes, enzymes, reactions and chemical compounds that participate in the metabolic pathways in *E. coli*. Their goal was to use a minimal amount of information from the KB to accomplish this so that the KB would be easy to maintain and update. The primary contribution they made to this end was what the authors call a predecessor list, which effectively gives reactions in a pathway order. Using the predecessor list and information stored in the KB about reactions, the enzymes that catalyze those reactions and the chemical compounds that are active in those reactions, Karp and Paley develop an algorithm based on production rules and heuristics to turn a predecessor list into a pathway graph. To this end they are somewhat successful.
— Tessa F. Weinstein

In this article the authors present an automated graph layout algorithm that dynamically draws given pathways present in the EcoCyc database. The algorithm determines the topology of the pathway as being cyclic, linear, or branched. Larger groupings of such pathways are handled by predefined layouts that are applied to the subgraphs within the pathway. Facilities for navigating, expanding and collapsing pathways within the user interface are discussed. Complex junctions and super-pathway algorithms are also discussed. It is apparent that one of the goals of the algorithm is depth

of representation, but not necessarily breadth.
Lance Lana et al.

[59] P.D. Karp, M. Riley, M. Saier, I.T. Paulsen, S.M. Paley, and A. Pellegrini-Toole. The EcoCyc and MetaCyc databases. *Nucleic Acids Research*, 28(1):56–59, 2000.

The authors provide an update on the EcoCyc and MetCyc knowledge bases, now unified under a common software toolbox called "The Pathway Tools," released as version 5.0 at the time of publication. While EcoCyc attempts to completely document the metabolic map of *E. coli*, MetaCyc is designed as a reference for metabolic pathways in various organisms, without the detailed genetic information provided in EcoCyc. The breadth of information contained in EcoCyc has been expanded to include membrane transport systems with the same level of detail afforded to the metabolic pathways represented in EcoCyc already. This allows researchers to query the relationships between metabolic pathways and transport systems at the cellular level. MetaCyc is built on the framework of EcoCyc, but it is expanded to contain species information for each pathway reaction. It does not contain any genetic map information.
— Raphael Bar-Or et al.

This article focuses on describing the information contained in, and the available forms of query to access, the EcoCyc and MetaCyc databases. EcoCyc is a database (DB) that contains biochemical information about *E. coli*, such as signal transduction pathways, transports, and its genes. A recent addition to this DB is information regarding membrane transport systems. MetaCyc, on the other hand, aims to describe metabolic pathways from a variety of different species. The information it contains about pathways includes reactions, enzymes and substrate components. However, it does not include information about transport processes like the EcoCyc database does.
— Tessa F. Weinstein

[60] H. Kitano, editor. *Foundations of Systems Biology*. MIT Press, Cambridge, MA, 2001. editor's introduction, *Systems Biology: Toward System-level Understanding of Biological Systems*, is also available at http://www.cds.caltech.edu/erato/.

[61] F. Kolpakov, E. Ananko, G. Kolesov, and N. Kolchanov. GeneNet: A gene network database and its automated visualization. *Bioinformatics*, 14(6):529–537, 1998.

This paper presents an object-oriented database, called GeneNet, and the software for its automated visualization. The authors describe three major models for the gene networks and their dynamics: (1) logical description, in which variables the gene networks components and functions the relationships between the components can take only a limited number of values, typically only 0, 1 and 2; (2) description of the gene network dynamics using a system of non-linear differential equations; and, (3) stochastic model of the gene network, which is most applicable to those parts of the gene network where small events are determined by a very low concentration of transcription factor. The major goal of the object-oriented database model is to ensure semantic data integrity where it reflects the real world and has other merits versus the relational model. The state of a gene network the objects involved, their states and interactions depends on the type of cells, inducers and other factors.
— Xuan Le

The authors describe the automated visualization of gene networks in their GeneNet database through a Java based graphical user interface found at http://wwwmgs.bionet.nsc.ru/systems/MGL/GeneNet/. While most other gene network data bases have been drawn manually, GeneNet's software provides automated construction of gene network diagrams. GeneNet is based on an object-oriented approach with entities such as cell, protein, gene and substance being related to each other by relations such as reaction and regulatory event. The resulting gene network diagram is a graph with nodes corresponding to entities and arrows representing relations. In order to automatically construct the graph, the nodes are first assigned to regions of the graph. Then, the size of the regions are determined and the arrows are drawn. GenNet also provides filters such as the ability to filter out the entities and relationships that have not been experimentally verified. All of the images are interactive allowing a user to click on an image to view more information. The main advantage to automated drawing of gene network diagrams is that the diagrams can be automatically updated

when new data is obtained. The major disadvantage is that the algorithm is not yet capable of drawing the very complex diagrams that humans are able to construct.
— Lance Lana

[62] F. Kose, W. Weckwerth, T. Linke, and O. Fiehn. Visualizing plant metabolomic correlation networks using clique-metabolite matrices. *Bioinformatics*, 17(12):1198–1208, 2001.

The authors use graph theoretic ideas to reduce visual complexity of metabolic network diagrams. Assigns edges between metabolites based on correlation of metabolite concentration. This idea is unusual but may indeed in some cases provide useful compression of data. Uses standard algorithm for finding all maximal cliques. These maximal cliques correspond to metabolites that all vary together. Uses an interesting approach that is very much like viewing the maximal-clique/vertex matrix for easily visualizing all metabolites in a clique and all cliques that contain a given vertex/metabolite.
— Cary Miller et al.

[63] J.R. Koza, G. Lanza W. Mydlowec, J. Yu, and M.A. Keane. Reverse engineering of metabolic pathways from observed data using genetic programming. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 6, pages 434–445, http://www-smi.stanford.edu/projects/helix/psb01/, 2001. World Wide Web.

This paper demonstrates that that it is possible to automatically create a network of chemical reactions from observed time domain data. Genetic programming starts with observed time domain. These concentrations of input substances can automatically create both the topology of the network of chemical reactions. The rates of each reaction with the network would bring concentration of the final product of the automatically created network, which matches the observed time-domain data. Specifically, genetic programming automatically created metabolic pathways involved in the phospholipid cycle and the synthesis and degradation of ketone bodies.
— Felemon Belay

This paper discusses the possibility of automatically creating a network of chemical reactions from observed time-domain

data using genetic programming. The paper walks through a detailed example of creating a metabolic pathway from only the time-domain concentration values of the final product. This involves automatically creating both the topology of the network and the sizing of the network of chemical reactions. There is discussion of representing networks of chemical reactions with programming trees and some background information on genetic programming.
— Amy Rulo

This paper introduces the feature of genetic programming that demonstrates the ability to create networks that have been generated by linear and nonlinear continuous-time differential equations. The article describes the ability of genetic programming to create a set of chemical reactions from observed time-domain data. The ideal result is the creation of metabolic pathways.
— Jennifer Phillips

The major principle addressed by this book is that genetic programming is a method for automatically creating a computer program whose behavior satisfies certain high-level requirements. Genetic programming has been demonstrated to work by authomatically creating both the topology and sizing for analog electrical circuits composed for transistors, capacitors, resistors, and other components merely by specifying the circuit's output.
— Olasumbo Olufunke Adesola

This discusses the application of genetic programming techniques to automatically generate feasible metabolic pathways from observed time-domain data. Such programs start with observed time-domain concentrations and automatically produce a both the topology and reaction rates of the underlying network of chemical reactions. The method presented in the paper establishes a representation of the involved chemical reactions and applies genetic programming methodologies to generate a population of improving pathways. These pathways are interpreted as analog electrical circuits and evaluated for fitness using a well known circuit simulation package.
— Tod Morrison

[64] G. Krauss. *Biochemistry of Signal Transduction and Regulation*. Wiley-VCH, Weinheim, FRG, 2nd edition, 2001.

This book is a must for everybody who wants to model signal transduction pathways. It provides the user with necessary facts and principles about signal transduction and regulation. The only negative aspect of the book is that the literature list is not as extensive and almost only restricted to review articles. Of special interest might be chapter 13 (cell cycle) and chapter 15 (apoptosis).
— Jens Eberlein

[65] A. Kremling, T. Sauter, E. Bullinger, M. Ederer, F. Allgower, and E. D. Gilles. Biosystems engineering: applying methods from systems theory to biological systems. In T-M. Yi, M. Hucka, M. Morohashi, and H. Kitanno, editors, *Proceedings of the Second International Conference of Systems Biology*, http://www.icsb2001.org/toc.html, 2001. World Wide Web.

This is a presentation of the application of the methods and tools from system theory in the analysis and design of mathematical models of biological systems. The author begins with a brief historical introduction to the development of the field of systems biology. He then proceeds to construct a mathematical model to simulate and analyze the cellular system of *E. coli* by progressively combining submodels, based on structured functional units, into higher aggregated models structures. The system is then systematically perturbed to reveal genes which are highly interconnected. Using the described model, methods of analysis and validation are discussed and high level conclusions regarding the usefulness of such an approach inferred.
— Tod Morrison

[66] Kyoto encyclopedia of genes and genomes (KEGG). World Wide Web, http://www.genome.ad.jp/kegg/.

This database contains pictures of pathways, hyperlinked to give information about the parts. Basic references to the literature are cited, and some are available as pdf or postscript files.

[67] S. Liang, S. Fuhrman, and R. Somogyi. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 3, pages 18–29, http://www-smi.stanford.edu/projects/helix/psb98/, 1998. World Wide Web.

The REVerse Engineering ALgorithm (REVEAL) was implemented as a C program to infer a complex regulatory network architecture from input/output patterns of its variables. The algorithm was generalized to include multi-state models, essentially allowing direct application to realistic biological data sets. A genetic network analysis tools can be designed based on generating model systems on which the performance of the tools can be tested on. An example of such model system is the Boolean network. In this network, the wiring of the elements to one another correspond to functional links between genes, and the rules determine the result of a signaling interaction given a set of input values. Genes are expressed as either off or on, which results in binary elements interacting according to Boolean rules.
— Olasumbo Olufunke Adesola

This article attempts to answer the question of whether it is possible in principle to completely infer a complex regulatory network architecture from the input/output patterns of its variables. Using state transition tables to represent gene expression patterns, Information Theory (Shannon Entropy) is used to determine the wiring relationships of the extended genetic network. A program called REVEAL does this, enabling the inference of inputs which control genes in the network.
— Rob Wilburn

[68] Ligand chemical database for enzyme reactions (LIGAND). World Wide Web,http://www.genome.ad.jp/dbget/ligand.html.

This database is designed to provide the linkage between chemical and biological aspects of life in the light of enzymatic reactions. There are three sections: ENZYME [25]; COMPOUND — a collection of metabolic compounds, including substrates, products and inhibitors; and REACTION — a collection of chemical reactions that appear in the pathway diagrams of KEGG [66].

[69] R. Maimon and S. Browning. Diagrammatic notation and computational structure of gene networks. In *Proceedings of the Second International Conference on Systems Biology*, pages 311–317, 2001.

This brings in a formal notation for gene networks. The notations are trying to standardize a higher-level language for

genetic networks in this paper. At first, it describes the simplest nouns, the atoms that are indivisible units of biological function and three types of actions that are the simplest verbs. And then it explains the linkbox that is the first complex grammatical element in the notation and keeps explaining about the likebox, the third elements that details sets of objects which act similarly. Through the examples, this paper proves the notation has two merits. The one is that it can explain modules and individual reactions and the other is that it is translated mechanically into a set of individual reactions. As a result, the notation is much reasonable comparing to the current notation.
— Min Hong

[70] Y. Maki, D. Tominaga, M. Okamoto, S. Watanabe, and Y. Eguchi. Development of a system for the inference of large scale genetic networks. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 6, pages 446–458, http://www-smi.stanford.edu/projects/helix/psb01/, 2001. World Wide Web.

This paper presents an approach to the inference of interrelated mechanisms among genes in a genetic network based on the analysis of gene expression patterns. The approach developed in this paper actually uses a combination of two previously developed methods. The first analysis is performed by a static Boolean network model based on a multi-level digraph approach. The second uses a dynamic network model, the S-system that relies on the analysis of temporal responses of gene expression patterns against perturbations or internal changes. The weakness in the Boolean network model is that relations between genes that affect each other cannot be determined. These genes are assigned to equivalence classes and then the dynamic network model is used to determine the relations within these equivalence classes. The Boolean network model can infer large genetic networks (10,000+ genes) in less than a second, but it cannot determine the relations between genes that belong to the same equivalence classes. The dynamic model based on the S-system can infer the network even if there are equivalence classes, but the run time is $O(n^2)$ so this model is not practical to use with large networks. The authors approach uses the Boolean network model to reduce the size of the network into functional units

that the dynamic model can solve in a "reasonable" amount of time.
— Lance Lana et al.

[71] E. Martz. *Beginner's Guide to Molecular Biology.* World Wide Web, http://www.iacr.bbsrc.ac.uk/notebook/courses/guide/, 2001.

This is a simplified introduction, with online animations to help understand mainstream topics in molecular biology. (The animations are with RasMol, which is free software that runs in an MS Windows environment.)

[72] M.L. Mavrovouniotis. Identification of qualitatively feasible metabolic pathways. In L.E. Hunter, editor, *Artificial Intelligence and Molecular Biology*, pages 325–364, Cambridge, MA, 1993. MIT Press.

This chapter describes an algorithm for the synthesis of biochemical pathways. Biochemical pathway synthesis is the construction of pathways which produce certain target bioproducts, under partial constraints on the available reactants, allowed by-products, etc. Given a set of stoichiometric constraints and a database of biochemical reactions, this algorithm transforms an initial set of available bioreactions into a final set of pathways by and *iterative* satisfaction of constraints. After explaining the design of the algorithm, the author presents a case study of its application to study of the synthesis of biochemical pathways for the production of lysine from glucose and ammonia.
— Tod Morrison

This article discusses the use of an AI method for finding quantitatively feasible metabolic pathways. In order to quantify a pathway's feasibility, the method uses information on the types and amounts of enzymes, ratios of metabolites, and the likelihood of a reactions occurrence in a particular direction within the pathway. The chapter discusses how the AI algorithm works and gives an abstract problem as an example.
— Rob Wilburn

[73] R. McEntire, P. Karp, N. Abernethy, D. Benton, G. Helt, M. DeJongh, R. Kent, A. Kosky, S. Lewis, D. Hodnett, E. Neumann, F. Olken, D. Pathak, P. Tarczy-Hornoch, L. Toldo, and T. Topaloglou. An evaluation of ontology exchange language for bioinformatics. In P. Bourne, M. Gribskov, R. Altman, N. Jensen, D. Hope, T. Lengauer, J. Mitchell,

E. Scheeff, C. Smith, S. Strande, and H. Weissig, editors, *Proceedings of the Eighth Intelligent Systems for Molecular Biology Conference* (ISMB), pages 239–250, Menlo Park, CA, 2000. AAAI Press.

This paper compares some candidate ontology-exchange languages to find the one that best deals with a variety of issues. Ontology exchange languages should exchange using a standardized form that has well-described syntax and semantics to make the sharing of information effective. If the database uses a well-defined ontology, then it can convey more accurate nuances of purpose. On the other hand, coarsely-defined ontologies will convey only superficial facets of information.

This paper compares the following ontology exchange languages: Ontolingua, CycL, OML/CKML, OPM, XML/RDF, UML, ASN.1, and ODL with the following ideal criteria: Language support and standardization, data model/capabilities, performance, pragmatics, and connectivity. It turns out Ontolingua and OML/CKML have enough expressivity, however, Ontolingua does not have XML expressions and OML is not a framed-based system, the author recommends a new language XOL (XML Ontology Language) that has frame-based semantics with XML expressions the author feels that XML is important due to the proliferation of the web and the widespread availability of parsers.
— Min Hong

[74] P. Mendes. *Computer Simulation of the dynamics of biochemical pathways*. PhD thesis, University of Wales, Institute of Biological Sciences, Aberystwyth, Wales, 1994.

[75] P. Mendes. Modeling large scale biological systems from functional genomic data: parameter estimation. In H. Kitano, editor, *Foundations of Systems Biology*, pages 163–186, Cambridge, MA, 2001. MIT Press.

Since new data became available, Mendes states that simultaneous measurements of thousands of cellular components, such as mRNA and proteins, can occur. When these sequences are put together, they form "movies" of the cellular machinery in action, and it should be possible to build models to describe the dynamics. These models, according

to Mendes, will be able to represent large numbers of biochemical reactions at some level of detail.
— Felemon Belay

[76] P. Mendes and D. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14(10):869–883, 1998.

This article discusses the suitability of various optimization methods, used as part of simulation software, to study the kinetics of biochemical reactions. The authors focus on analyzing the ability of these algorithms to find global minima. They recommend that a suite of diverse optimization methods should be available in simulation software, as no single one performs best for all problems. They discuss how they have implemented such a simulation-optimization strategy in the biochemical kinetics simulator Gepasi (http://gepisi.dbs.aber.as.uk/softw/Gepasi.html). They provide an overview of optimization methods and discuss the importance of finding the global minimum. They also discuss computational issues that arise from the nonlinearity. Applications and numerical results are discussed for two areas: simulation of a hypothetical branched biochemical pathway with conserved cofactor and feedback, and parameter estimation. The following methods are compared: L-BFGS-B, Levenberg-Marquardt, Steepest descent, Simulated annealing, Multistart, Random search, Truncated Newton, Evolutionary programming and a genetic algorithm. Many of these methods can be implemented without explicit calculation of derivatives.
— Rico Argentati

[77] G. Michal, editor. *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. John Wiley & Sons, Heidelberg, FRG, 1999.

[78] *MIT Biology Hypertextbook*. Massachusetts Institute of Technology, http://esg-www.mit.edu:8001/esgbio/7001main.html, latest edition, 2002.

This has undergone maturation since its first posting, and it is a fairly complete introductory resource. The chapters of most direct benefit to understanding pathways are *Chemistry Review, Large Molecules, Cell Biology, Enzyme Biochemistry, Glycolysis and the Krebs Cycle,* and *Prokaryotic*

Genetics and Gene Expression. (Of course, the other chapters are also important.) You can stretch your knowledge by working their "practice problems."

[79] J.E. Mittenthal. An algorithm to assemble pathways from processes. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 2, pages 292–303, http://www-smi.stanford.edu/projects/helix/psb97/, 1997. World Wide Web.

This paper discusses algorithms created to assemble molecular networks in which the enzymes remain unchanged. It is shown that this type of algorithm is different than one in which the enzymes are changed. Specific steps are given to implement a method in which the enzymes remain unchanged. A particular example, the regulation of the breakdown and sythesis of glycogen, is given to help the reader to understand the method. The future possibilities of such an algorithm are explored as well.
— Jennifer Phillips

This article covers the steps of an algorithm due to Mavrovouniotis et al. (1990) and gives several examples of what the algorithm does. This can be used to assemble pathways from existing molecules and reactions and can be used to assemble possible pathways from databases of molecules and reactions.
— Amy Rulo

This article discusses how an algorithm developed for the construction of metabolic pathways can be altered in order to construct protein-modifying networks. The method uses lists of reactions, molecules and enzymes, along with a set of constraints to build potential pathways. The method first identifies coupled reactions which are considered partial pathways. These coupled reactions can then be used as building blocks for constructing larger pathways.
— Rob Wilburn

This discusses the application of the algorithm developed by Mavrovouniotis [72] to the construction of protein-modifying networks. Typically in metabolism, reactions do not modify enzymes, but in some networks a process *may* modify enzymes. The author presents an example of one such network and demonstrates that the algorithm developed by

Mavrovouniotis is suitable to these types of metabolic networks as well as the types it was initially designed for. From this the author concludes that the algorithm can assemble many types of intercellular networks, including pathways mediating gene regulation and cell mobility.
— Tod Morrison

[80] C.A. Ouzounis and P.D. Karp. Global properties of the metabolic map of *Escherichia coli*. *Genome Research*, 10:568–576, 2000.

[81] M.G. Poolman and D.A. Fell. Modelling and experimental eveidence for two separate steady states in the photosynthetic calvin cycle. In J.-H. S. Hofmeyr, J.M. Rohwer, and J.L. Snoep, editors, *Animating the cellular map*, pages 249–254. Stellenbosch University Press, 2000. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[82] Protein data bank (PDB). World Wide Web, http://www.rcsb.org/pdb/.

This database contains protein data, including annotations. It highlights a "Molecule of the Month."

[83] B. Raïs, C. Chassagnole, T. Letellier, D.A. Fell, and J.-P. Mazat. Threonine synthesis from aspartate in *Escherichia coli* cell-free extracts: pathway dynamics. *Biochemistry Journal*, 356:425–432, 2001. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[84] R. Ramakrishna, J.S. Edwards, A. McCulloch, and B.Ø. Palsson. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *American Journal of Physiology — Regulatory, Integrative and Comparative Physiology*, 280:R695–R704, 2001.

[85] T.J. Rothenberg. Simultaneous equations models. In J. Eatwell, M. Milgate, and P. Newman, editors, *The New Palgrave: Econmetrics*, pages 229–237, New York, NY, 1990. W.W. Norton & Company.

The book is a series of invited articles by noted econometricians about subjects basic to the theory and doctrine of the field of econometrics. The discussion on this chapter on simultaneous equations gives a clearer view as to how econometricians use this method and the limitations of this method in modelling real economic phenomena. The author conisiders exogenous and endogenous factors to be important in this analysis as tools for building the model. This can be compared to external and internal factors in the behavior of the cell, or some other biological systems.
— Andrew Been

[86] A. Rzhetsky, T. Koike, S. Kalachikov, S.M. Gomez, M. Krauthammer, S.H. Kaplan, P. Kra, J.J. Russo, and C. Friedman. A knowledge model for analysis and simulation of regulatory networks. *Bioinformatics*, 16(12):1120–1128, 2000.

This paper describes an ontology specifically geared toward signal transduction pathways. The ontology is based upon Peter Karp's EcoCyc ontology [59] and is extended to supposingly capture the features of a signal transduction pathways. Like many papers in this area, this paper is not very detailed. In particular, it does not provide any statistical method to prove how well the ontology works. It merely states that the knowledgebase created using this ontology works well, not justifying the claim with any depth. Still, the presented ontology and taxonomies does provide some interesting new relationships geared toward signal transduction.
— Jens Eberlein

An attempt to deal with the enormous volume of scientific publication by using natural language processing techniques to automate knowledge acquisition. Uses a knowledge-base similar to that of EcoCyc but removes the human intervention to the level of assigning scores to authors, journals, etc. Discusses in some detail the 4 aspects of a knowledge-base; taxonomy of concepts (is-a relationship), relationships between concepts (has-a relationship), properties of concepts (slots) and axioms. Makes the important distinction between the logical and biochemical representations of a reaction or pathway.
— Cary Miller et al.

[87] W. Salamonsen, K.Y.C. Mok, and P. Kolatkar. BioJAKE: A tool for the creation, visualization and manipulation of metabolic pathways. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 4, http://www-smi.stanford.edu/projects/helix/psb99/, 1999. World Wide Web.

The authors discuss the Java computer program BioJAKE which allows for the creation, manipulation and visualization of metabolic pathways. A user may create pathways from scratch or from information stored in databases. Pathways are represented as a series of images (molecules) and reaction lines. Users may zoom into specific areas of interest and remote database queries such as Prosite, Entrez, and Blast are possible. The molecules are objects that have a type and a state. A *type tree* is used to describe the possible molecular types such as protein and nucleic acid. States include phosphorylation and capping. Three dimensional representation and the addition of reaction simulations are discussed as possible future enhancements to the program.
— Lance Lana

This article introduces BioJAKE, a Java-based visualization program that enables the user to create and manipulate metabolic pathways. Incorporated into the BioJAKE program is the ability to access other remote databases. Once the databases have been accessed, BioJAKE sorts and stores any queries according to the given metabolic pathway. Ultimately, the information is integrated into a larger, comprehensive database for future references.
— Kimberly A. Somers

This article describes a JAVA program called BioJAKE. BioJAKE provides easy to use mechanisms for the construction of on-screen pathway visualizations. The primary topics in this article are how BioJAKE constructs these visualizations, what information is used, how Types and States are used to identify a molecules present state, and the future development of BioJAKE.
— Bob Wilburn

This paper covers the possibility of completely inferring complex regulatory network architecture from input/output patterns of its variables. This was investigated using the REVEAL algorithm and binary models of genetic networks. The

paper goes into detail about how the algorithm works using M-analysis, mutual information measures, and gives the step-by-step process. The benefit of this algorithm is that it performs rather quickly for networks with low numbers of inputs per gene, however more time is required for larger values. It can also be generalized to include multi-state models allowing direct application to realistic biological data sets.
— Amy Rulo

The BioJAKE program was created for the visualization, creation and manipulation of metabolic pathways. It provides an easy-to-use mechanism to constuct pathways either from scratch, or automatically from information stored within a database. The elements of these pathways contain and represent diverse information including state, location, comments and associated remote database queries. It also computes the reactions in which any given molecule is involved, or the sequence of reactions required for a specific pathway to take place. This program was developed in Java to provide individual independence and maximum extendibility.
— Olasumbo Olufunke Adesola

[88] M. G. Samsonova and V. N. Serov. NetWork: An interactive interface to the tools for analysis of genetic network structure and dynamics. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 4, pages 102–111, http://www-smi.stanford.edu/projects/helix/psb99/, 1999. World Wide Web.

This introduces a Java applet, called NetWork, that enables the visualization of large genetic networks at the level of gene expression. A Boolean network is chosen for the mathematical representation of the genetic networks. In the Boolean network, the state of a gene is either turned on or off. Genes are represented as elements of the Boolean net, and the wiring of the elements represents the functionality of the links between the genes. A detailed explanation of the algorithm is given. In addition to the visualization of a network, the program provides an interactive interface to the tools for network dynamics simulation. A user can simulate a dynamic network by selecting a network, selecting the genes in the network that are turned on at the start, then clicking on the Dynamics button. The weakness of the program

is that regulatory interactions occurring at other than the gene expression level are not considered. The authors plan to modify the program to account for this weakness.
— Lance Lana

[89] H.M. Sauro. Jarnac: A system for interactive metabolic analysis. In J-H. S. Hofmeyr, J. M. Rohwer, and J. L. Snoep, editors, *Animating the Cellular Map*, 9th International BioThermoKinetics Meeting, pages 221–228, http://www.sun.ac.za/biochem/btk/book/Sauro.pdf/, 2000. Stellenbosch University Press.

[90] J.M. Savinell and B. Ø. Palsson. Network analysis of intermediary metabolism using linear optimization I: development of mathematical formalism. *Journal of Theoretical Biology*, 154:421–454, 1992.

The article addresses the benefits of using a stoichiometric matrix as a tool for understanding metabolic behavior of a cell its relationship with external compounds. The metabolic pathways considered in the matrix were the mass and energy factors. Shadow prices acted as the parameters on the matrix, which ultimately helped in identifying any growth limitations. Attempts at linearly optimizing the matrix revealed important cell behavior in its processing of elements, such as carbon, for growth factors. Glucose, glutamine, and glutamate factors were the primary focus upon linear optimization. The linear optimization returned information relating to energy use, growth, and ratios under given conditions such as limited oxygen uptake or minimized production of NADH. Using linear optimization techniques proved successful when attempting to minimize certain productions, which then offered insight to inhibiting some cell functions or deleting genes altogether. Overall, the stoichiometric matrix for analyzing the cell behavior related to metabolic pathways returned good results and offered great promise for further study.
— Kimberly A. Somers

[91] J.M. Savinell and B. Ø. Palsson. Network analysis of intermediary metabolism using linear optimization II: interpretation of hybridoma cell metabolism. *Journal of Theoretical Biology*, 154:455–473, 1992.

The article continues to examine the metabolic network represented as stoichiometric matrix. The simulations of the hy-

bridoma looked at the conditions placed upon the stoichiometric and the associated interactions among the nutrients. Quite a few factors were able to be examined for their effects on growth rate, which included ATP and antibody production rates. Also analyzed were the shadow price effects and reduced cost effects of chosen elements or nutrients on the growth factors in the cell. The sensitivities of some compounds (pyruvate, for example) were also analyzed. Shadow prices rendered the most significant information regarding the stoichiometry as a whole. Overall, further stoichiometric analyses contributed to increased understanding about cell function and cost analyses for conducting such simulations.
— Kimberly A. Somers

[92] J.M. Savinell and B.Ø. Palsson. Optimal selection of metabolic fluxes for *in vivo* experimental determination I: development of mathematical methods. *Journal of Theoretical Biology*, 155:201–214, 1992.

This paper presents a framework for determining internal fluxes for highly branched metabolic networks. The problem is formulated as a linear system which contains both measured and calculated fluxes. The goal of this particular work is to determine an optimal set of fluxes to measure so that the calculated values are as accurate as possible. This is done by anayzing the sensitivity of the system (bounded by a condition number) to the choice of measured vs. calculated fluxes. The authors find an estimation to the optimal set of measured fluxes using a shotgun approach and observing the shape of the scatter plot produced.
— Raphael Bar-Or et al.

[93] J.M. Savinell and B.Ø. Palsson. Optimal selection of metabolic fluxes for in vivo experimental determination II: application to *E. coli* and hybridoma cell metabolism. *Journal of Theoretical Biology*, 155:215–242, 1992.

This paper is an application of the authors previous work to e-coli and Hybridoma cell lines to estimate metabolic fluxes from a set of measured fluxes. They find that there were potentially very poor choices of measured fluxes such that they would lead to a 100,000 fold amplification of experimental errors. They further found that the sensitivity of the systems increased with system complexity and size. They show that by measuring additional fluxes, such that the system

is over determined and computing a least-squared solution, that they could significantly reduce the error. The data they present suggests that relatively few extra measurements can have a significant effect on the final error and that for poorly conditioned systems, that this approach may be useful.
— Raphael Bar-Or et al.

[94] F. Schacherer. *An object-oriented database for the compilation of signal transduction pathways.* PhD thesis, Technical Universität Braunschweig, Braunschweig, FRG, 2001.

[95] F.G.D. Schacherer. A review of biological network visualization. Technical report, Technical Universität Braunschweig, Braunschweig, FRG, 2000.

[96] C.H. Schilling, J.S. Edwards, D. Letscher, and B. Ø. Palsson. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering*, 71(4):286–306, 2001.

[97] C.H. Schilling, D. Letscher, and B. Ø. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203:229–248, 2000.

Uses the well-established, super-powerful techniques of linear algebra to analyze metabolic pathways. Makes use of a subset of all possible metabolic pathways available in a system to define the 'metabolic flux cone'. This is the set of all conceivable metabolic phenotypes. To make the analysis tractable to linear algebraic methods makes the simplifying assumption that reactions and pathways are at a steady-state. This assumption is not totally justifiable but is common in modeling of metabolic pathways. The paper does not address pathway regulation. Some of the descriptive linear "equations" are actually inequalities so concepts of linear programming or convex analysis are used.
— Cary Miller et al.

[98] C.H. Schilling and B. Ø. Palsson. The underlying pathway structure of biochemical reaction networks. *Proceedings of the National Academy of Science*, 95:4193–4198, 1998.

[99] C.H. Schilling and B. Ø. Palsson. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *Journal of Theoretical Biology*, 203:249–283, 2000.

[100] C.H. Schillling, S. Schuster, B. Ø. Palsson, and R. Heinrich. Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress*, 15:296–303, 1999.

[101] I. Shah and L. Hunter. Visualization based on the enzyme commission nomenclature. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 5, pages 275–287, http://www-smi.stanford.edu/projects/helix/psb00/, 2000. World Wide Web.

> Using the EC numbers, Shah and Hunter map the systems hierarchical structure into a visualizable trees. At its simplest, these visualizations indicate related proteins and their quantitative properties. They implement a more complex application that looks at the ROCK (a sensitivity versus specificity test) scores measuring the ability to predict functionality based off of sequence. They could then look at an enzyme class and see for that particular class how good of a predictor the sequence was.
> — Aaron Gabow

[102] G.N. Stephanopoulos, A.A. Aristidou, and J. Nielsen. *Metabolic Engineering: Principles and Methodologies*. Academic Press, London, UK, 1998.

[103] Z. Szallasi and S. Liang. Modeling the normal and neoplastic cell cycle with "realistic boolean genetic networks": Their application for understanding carcinogenesis and assessing therapeutic strategies. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings*

of the Pacific Symposium on Biocomputing (PSB), volume 3, pages 66–76, http://www-smi.stanford.edu/projects/helix/psb98/, 1998. World Wide Web.

[104] A. Tanay and R. Shamir. Computational expansion of genetic networks. *Bioinformatics*, 17(Supplement 1):1–9, 2001.

[105] S. Thomas, P.J.F. Mooney, M.M. Burrell, and D. A. Fell. Metabolic control analysis of glycolysis in tuber tissue of potato (*Solanum tuberosum*): explanation for the low control coefficient of phosphofructokinase over respiratory flux. *Biochemistry Journal*, 322:119–127, 1997. Note: available at http://bms-mudshark.brookes.ac.uk/publicat.html.

[106] M. Tomita, K. Hashimoto, K. Takahashi, Y. Matsuzaki, R. Matsushima, K. Yugi, F. Miyoshi, H. Nakano, Y. Saito, S. Shimizu, and Y. Nakayama. The e-cell project: Towards integrative simulation of cellular processes. In D.J. States, P. Agarwal, T. Gaasterland, L. Hunter, and R.F.Smith, editors, *RECOMB Proceedings*, pages 290–298, New York, NY, 1996. ACM.

This introduces the E-Cell project (http://www.e-cell.org) that models and simulates various cellular processes. It focuses on the simulation of large scale of cellular behavior using same framework. This contains kinetic model of erythrocyte, DNA replication, signal transduction of bacterial chemotaxis, and energy metabolism for mitochondria. Three classes of objects (substance, reactor, and system) are used to represent molecular species, reactions, and physical/functional compartment.
— Min Hong

[107] P. Uetz, T. Ideker, and B. Schwikowski. Visualization and integration of protein-protein interactions. In E. Golemis, editor, *The Study of Protein-Protein Interactions — An Advanced Manual*, Cold Spring Harbor, 2002 (to appear). Cold Spring Harbor Laboratory Press.

The introduction compares protein interaction maps with metabolic pathways and describes the kinds of information in need of visual displays. Besides the usual zoom and pan functions, they point out the need for graph condensation

and expansion, raising issues of recognizable protein classifications. They proceed to descibe the symbolic syntax of their displays and handling supplemental data. In their "Future Directions" the authors point to three important needs:

1. More complex schemes to integrate data and network visualization — for example, small molecules that influence a gene's expression ought to be represented in the network, but presently they are not
2. Better network layout algorithms — for example, group nodes that represent genes of similar function
3. Broader range of information about the network objects — for example, was some interaction predicted or discovered by experiment, and what was the associated information about error?

[108] C. van Gend and U. Kummer. STODE — automatic stochastic simulation of systems described by differential equations. Technical report, Bioinformatics and Computational Chemistry Group, European Media Laboratory, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, FRG, 2001.

[109] J. van Helden, D. Gilbert, L. Wernisch, M. Schroeder, and S. Wodak. Application of regulatory sequence analysis and metabolic network analysis to the interpretation of gene expression data. In O. Gascuel and M-F. Sagot, editors, *Computational Biology*, number 2066 in Lecture Notes in Computer Science, pages 155–172, Heidelberg, FRG, 2001. Springer-Verlag.

[110] J. van Helden, A. Naim, C. Lerner, R. Mancuso, M. Eldridge, and S.J. Wodak. From molecular activities and processes to biological function. *Briefings in Bioinformatics*, 2(1):81–93, 2001.

This article describes the aMAZE data model, based on a conceptual description of the network of interactions between molecular entities. aMAZE is the name for "the labyrinth of interactions between the myriad of molecular entities in the cell." This software has the ambition of embodying the molecular interactions, processes and pathways. Furthermore, it handles information on signal transduction

pathways, bio-ontology, and it can be thought of as defining biological function at the semantic level. This aMAZE database, can deal with information on metabolic pathways, gene regulation, sub-cellular locations and transport. There are four major classes or levels of hierarchy in the aMAZE database:

1. **Biochemical Entity** attributes of the object represent structural properties.
2. **Interaction** contains objects describing molecular activities, which are subdivided into classes specifying particular types of activities called Reaction, Expression, transcriptional regulation and Assembly/Disassembly.
3. **Process/Pathway** is a network, whose nodes are pathway elements, connected by pathway arcs.
4. **Compartment** defines the different levels in the organization of biochemical entities: sub-cellular compartment, cell type, tissue, organ, and organisms themselves can be considered as a special type of compartment.

The Object Oriented aMAZE model is represented by the reference relationship class hierarchy, instead of the inheritance relationship, operating between objects, whereby a given object refers to another object in the database. This relationship is encoded in the objects attributes. Representing the various facets of multifunctionality is not straightforward with this system, and would require adding extra dimensions to the current description. Thus, in this model, information on the function(s) of a gene product is not obtained by looking up catalogues, but is computed on the fly via queries to the aMAZE database, which analyze the network of interactions in which the gene product takes part. Other more elaborate data structures make a clear distinction between entities and interactions. However, except for chemical and catalyst reactions, all other types of interactions are represented in relationship between the objects, rather than objects in their own right as in the aMAZE model. These advanced models, though they may represent entities, interactions and pathways, fall short in representing hierarchical relationships between object classes, and the representation of object taxonomies, as the aMAZE does.

— Anissa Larson

[111] J. van Helden, A. Naim, R. Mancuso, M. Eldridge, L. Wernlsch,

D. Gilbert, and S.J. Wodak. Representing and analyzing molecular and cellular function in the computer. *Biological Chemistry*, 381(9/10):921–935, 2000.

The authors of this article discuss problems with existing biological databases and the need for finding better ways to represent information including the representation of biological function on multiple levels. They discuss how the aMAZE database model will address some of these problems. The aMAZE database model goes beyond the defining and catagorizing tasks of most ontologies by defining proteins and compounds in a functional context and allowing representation of multiple functions for the same molecule. The data model also has the capacity to add new knowledge in a flexible way when needed. — Amy Rulo

[112] E.O. Voit. *Computational Analysis of Biochemical Systems*. Cambridge University Press, Cambridge, UK, 2000.

[113] A. Wagner and D.A. Fell. The small world inside large metabolic network. Working paper 00-07-041, The Sante Fe Institute, 1399 Hyde Park Road, Sante Fe, NM 87501, 2000. Note: available at http://www.santafe.edu/sfi/publications/00wplist.html.

[114] D. Weaver, C. Workman, and G. Stormo. Modeling regulatory networks with weight matrices. In *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 4, pages 112–123, 1999.

The authors describe the TReMM (Transcription Regulation Modeled with Matrices) algorithm to model gene regulatory pathways with a linear weight matrix. Each row represents all of the regulatory inputs for one gene. TReMM was implemented in MATLAB, and it is described as a framework within which more accurate nonlinear modeling components and environmental variables can be included.
— Adolfo Perez

[115] L.F.A. Wessels, E.P. Van Someren, and M.J.T. Reinders. A comparison of genetic network models. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 6, pages 508–519, http://www-smi.stanford.edu/projects/helix/psb01/, 2001. World Wide Web.

Genetic networks were proposed as a possible methodology for modeling genetic interactions. Since then, a wide variety of different models have been introduced. It is unclear what the strengths and weaknesses of each of these approaches are. This article compares and discover what about each network makes them unique but at the same time similar. Also this paper addresses the extraction of the gene regulation matrix from the expressed data.
— Felemon Belay

Several genetic network models have been proposed for modelling genetic interactions. This paper compares many of those different approaches and rates them based on their consistency, ability to infer and predict, stability and computational cost. Each of these properties is discussed. These models serve the purpose of extracting the gene regulation matrix. The paper gives the results of the comparisons, showing which models are best suited for performing large scale gene expression and anlysis.
— Jennifer Phillips

[116] U. Wittig and A. De Beuckelaer. Analysis and comparison of metabolic pathway databases. *Briefings in Bioinformatics*, 2(2):126–142, 2001.

This is a useful, albeit very brief, description of current pathway databases (most of which are static and have no computational methods to support inference): CSNDB, EcoCyc/MetaCyc, ExPASy — Biochemcial Pathways, KEGG, PATHDB, SPAD UM-BBD. (It does not include others, such as BioCarta [10].)

[117] L. Wong. PIES: A protein interaction extraction system. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Proceedings of the Pacific Symposium on Biocomputing* (PSB), volume 6, pages 520–531, http://www-smi.stanford.edu/projects/helix/psb97/, 2001. World Wide Web.

The Protein Interaction Extraction System (PIES) is a data extraction method that uses three separate tools for data collection, analysis, and visualization development. A natural language processing method is utilized and the information is analyzes to create graphical layouts for metabolic and protein pathways. A strength of the PIES program is in its ability to utilize multiple comparison tools.
— Ben Perrone

[118] A. Zien, R. Küffner, R. Zimmer, and T. Lengauer. Analysis of gene expression data with pathway scores. In P. Bourne, M. Gribskov, R. Altman, N. Jensen, D. Hope, T. Lengauer, J. Mitchell, E. Scheeff, C. Smith, S. Strande, and H. Weissig, editors, *Proceedings of the Eighth Intelligent Systems for Molecular Biology Conference* (ISMB), pages 407–417, Menlo Park, CA, 2000. AAAI Press.