

CHAPTER 3

NETWORK INFERENCE

Frank Doyle and Douglas Lauffenburger

INTRODUCTION

As systems biology emerges in the post-genomic era, the emphasis is shifting from annotation of individual genes and gene products to ascertaining how DNA-protein and protein-protein interactions occur within a complex network of structural, metabolic, and regulatory pathways in cells. This goal is, of course, aligned with that pursued in “reductionist” molecular cell biology for the past two decades, in which efforts to identify and characterize pathways typically have proceeded in a component-centric manner, beginning with an initial gene or protein of particular interest and attempting to ascertain other genes and/or proteins involved in the same pathway. However, although component-centric approaches have been successful in assembling most of the available knowledge about pathways to date, they have several inherent difficulties. First is the time required: accurate models of pathway function emerge only after evidence is accumulated over many years, with the work of many researchers at many laboratories. Second, these approaches do not directly reveal how multiple pathways influence each other or reveal this crosstalk only accidentally. Third, the vast amount of information on the various intracellular pathways remains fairly decentralized, buried across primary literature or within narrowly defined reviews.

Systems biology can offer an accelerated approach to this goal of identifying networks by which genes and proteins interact to carry out cellular operational and regulatory functions, using computational mining methods on high-throughput experimental data. The resulting high-level models are finding increasing utilization as tools for drug discovery, both by small companies as well as large pharmaceutical companies such as Eli Lilly and Novartis (Henry, 2005). In general, the resulting “high-level,” topological models can then serve as a foundational prelude, if one wishes, for more familiar (to engineers, physicists, and applied mathematicians, at least) “low-level,” mechanistic models. This categorization has previously been laid out (Ideker & Lauffenburger 2003), and the summary and Fig. 3.1 below are largely taken from their discussion by a common co-author of this chapter and that article.

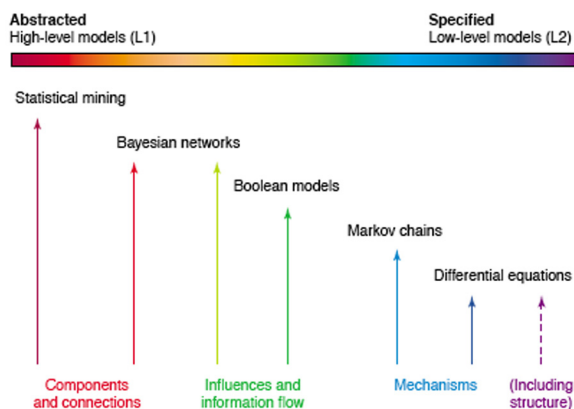


Figure 3.1. A diverse spectrum of high-to-low modeling approaches (Ideker and Lauffenburger, 2003).

Signaling and regulatory pathways consist of some number of components—such as genes, proteins and small molecules—wired together in a complex network of intermolecular interactions. Recent technological developments are enabling us to define and interrogate these pathways more directly and systematically than ever before, using two complementary approaches. First, it is now possible to systematically measure the molecular interactions themselves, by screening for protein-protein, protein-DNA and small molecular interactions. Several methods are available for measuring protein-protein interactions on a large scale—two of the most popular being the yeast two-hybrid system and protein co-immunoprecipitation (coIP) in conjunction with tandem mass spectrometry. Although the vast majority of protein interactions have been generated for the budding yeast *Saccharomyces cerevisiae*, protein interactions are becoming available for a variety of other species including *Helicobacter pylori* and *Caenorhabditis elegans* and are catalogued in public databases such as BIND and DIPTM. A current drawback of these high-throughput measurements is an associated high error rate. One approach for addressing this problem may be to integrate several complementary data sets together (e.g., two-hybrid interactions with coIP data or gene expression profiles) to reinforce the common signal.

Protein-DNA interactions, as commonly occur between transcription factors and their DNA binding sites, constitute another interaction type that can now be measured at high throughput. The relatively new technique of chromatin immunoprecipitation microarraying (so-called “ChIP-chip”) has been used to characterize the complete set of promoter regions bound under nominal conditions for each of the >100 transcription factors in yeast, yielding >5,000 novel protein-DNA interactions in that organism. Additional types of pathway interactions, such as those between proteins and small molecules (carbohydrates, lipids, drugs, hormones and other metabolites), are difficult to measure on a large scale, although protein array technology might enable high-throughput measurement of protein-small molecule interactions in the near future.

In addition to characterizing molecular interactions, a second major approach for interrogating pathways is to systematically measure the molecular and cellular states induced by the interaction wiring. For example, global changes in gene expression are measured with DNA microarrays, whereas changes in protein abundance, protein phosphorylation state, and metabolite concentrations can be quantified with mass spectrometry, Nuclear magnetic resonance (NMR) and other advanced techniques. Of these approaches, measurements made by DNA microarrays are currently the most comprehensive (every mRNA species is detected); high-throughput (a single technician can assay multiple conditions per week); well characterized (experimental error is appreciable, but understood); and cost-effective (whole-genome microarrays are purchased commercially for US \$50 to \$1000, depending on the organism). However, continued advances in protein labeling and separation technology are making the measurement of protein abundance and phosphorylation state almost as feasible, with the primary barrier being the expense and expertise required to set up and manage a mass spectrometry facility. Measurement of metabolite concentrations, an endeavor otherwise known as metabolomics, is currently limited not by detection (thousands of peaks, each representing a different molecular species, are found in a typical NMR spectrum) but by identification (matching each peak with a chemical structure is difficult). Clearly, measuring changes in cellular state at the protein and metabolic levels will be crucial if researchers are to gain insight into not only regulatory pathways, but also those pertaining to the cell’s signaling and metabolic circuitry.

To arrive at a high-level topological model of a cellular network of interest, data on molecular interactions and states can be integrated in a multi-tiered strategy. First, the global molecular interaction scaffold is constructed from systematic measurements of protein-protein interactions, protein-DNA interactions and/or metabolic reactions (as detailed in the previous section). In the case of budding yeast, a minimal set might include 14,941 protein-protein interactions catalogued in the DIPTM database; 5,631 protein-DNA interactions from a combination of TRANSFAC[®] and ChIP-chip; and 599 enzymatic reactions in MetaCyc. Second, this scaffold is filtered against changes in mRNA expression, protein expression and/or post-translational modifications recorded in response to different cellular perturbations. Networks within the interaction scaffold with mRNA or protein states that are significantly activated by perturbation are identified and mapped according to a computational search engine. The interaction pathways and complexes comprising the scaffold constitute topological models, which are then prime candidates for further verification and modeling as important signaling and compensatory mechanisms controlling the cellular perturbation response. The key advance of these searches is that by integrating two complementary global

data sets, it is possible to condense and partition the enormous quantity of data into a small number of relevant pieces suitable for lower level modeling.

Examples of this general scheme have been reported in recent literature. Several groups have applied “co-clustering” approaches to identify groups of proteins that are both differentially expressed under similar sets of conditions and closely connected by the same network of interactions in the scaffold. In many cases, these “expression-activated networks” correspond to well known protein complexes, regulatory pathways or metabolic reaction pathways. Other groups have used probabilistic approaches to match changes in gene expression with the transcription factors that are most likely to regulate them directly. These methods start with a cluster of differentially expressed genes and incrementally choose a small set of transcription factors that, by virtue of their levels and/or interactions in the scaffold, can maximally predict the observed levels of differential expression in the cluster. New transcription factors are added only if they lead to a sufficient increase in predictive power over the transcription factors already in the model.

Several software tools are now available for visualizing interaction scaffolds [Osprey, <http://biodata.mshri.on.ca>; PIMRider®, <http://pim.hybrigenics.com>; GenoMax™, <http://www.informaxinc.com>; Cytoscape, <http://www.cytoscape.org>; Pathway Tools, <http://bioinformatics.ai.sri.com/ptools/>]. For instance, the Cytoscape framework provides network visualization, layout and annotation, as well as clustering of the network against expression data to generate topological network models. The PathwayTools component of the MetaCyc metabolic pathway database can superimpose enzyme expression levels on the map of biochemical reactions for a species, giving a good indication of which reaction pathways are most affected over a panel of growth conditions profiled by microarray.

Because DNA microarray technology is currently much more widespread than technologies for protein or metabolite profiling, the vast majority of these approaches have used gene expression profiling as the primary state measurement. Of course, pathway mapping methods based on mRNA profiling alone capture just one facet of a much larger and complex cellular response. As it becomes possible to measure cellular state at the protein and small-molecule level, researchers expect that algorithms similar to those described above will emerge. Currently, omitting this information from the analysis means that regulatory networks not purely transcriptional in nature remained to be elucidated in this high-throughput computation-aided manner. Indeed, the next obvious challenge beckoning is to push forward from the current focus on transcriptional regulation to regulation occurring in the post-transcriptional and post-translational arenas, and especially arising from extracellular cues.

NETWORK INFERENCE AND MODEL STRUCTURE

In the area of network inference, the models are primarily static interconnection descriptions of collections of proteins, metabolites, and/or genes. The “inference” problem involves the estimation of the interactions of elements in the network, given (possibly time series) data of activities of different nodes (e.g., gene interactions from gene expression data). The goals of the inference problem are multiple, and include: (i) hypothesis generation, (ii) design of experiment, (iii) understanding of cellular function, and (iv) unraveling design principles, among others. The sources of information for these inference problems include large-scale deletion projects, and vast numbers of microarray experiments. In the early years of bioinformatics studies, the structural localization properties were inferred (e.g., which transcription factors regulate the transcription of which genes), although experimental methods now exist for identifying protein-DNA interactions on a genomic scale, such as ChIP assays, that yield structural knowledge.

Given the wide variety of modeling objectives, as well as the heterogeneous sources of data, it is not surprising that the WTEC study team observed many approaches to modeling for network inference in Japan, Europe, and the U.S. For the case of microbial systems biology, the review paper by Stelling (2005) provides a good summary of the spectrum of modeling approaches. He classifies the modeling efforts in two respects: network complexity and level of detail. This chapter of the report focuses on the problems associated with the more complex, less detailed models, while a separate chapter examines the issues associated with more detailed (mechanistic) models.

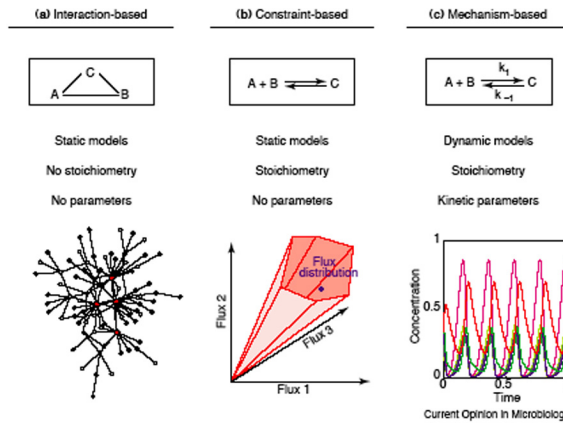


Figure 3.2. Approaches to the mathematical modeling of cellular networks (Stelling, 2005).

The mathematical structures invoked to capture network interactions are numerous, and include:

- *Boolean Networks*—in which the network is represented as a graph of nodes, with directed edges between nodes and a function for each node (e.g., Ideker et al., 2000)
- *Petri Nets*—another graph theoretic structure in which nodes (or places) are connected by arcs and activities are modeled by transitions (e.g., Nagasaki et al., 2004)
- *Bayesian Nets*—combine directed acyclic graphs with a conditional distribution for each random variable (vertices in graph) (e.g., Pe’er et al., 2001)
- *Signed Directed Graphs*—another graph theoretic structure in which a signed directed edge is used to represent activation versus inhibition (depending on sign) (e.g., Kyoda et al., 2004)
- *S-systems*—notably a *dynamic* approach in which polynomial nonlinear dynamic nodes are used to capture network behavior (e.g., Kimura et al., 2005)

A significant challenge in constructing these network models from data, particularly for gene network models, is the fact that the node dimension (number of genes) can be on the order of 10,000—leading to a computationally untenable problem for inference (i.e., determination of 10^8 coefficients of interaction!). In reality the network is tremendously sparse and highly structured, such that there are orders of magnitude fewer “interactions” that must be captured with coefficients. The knowledge that not every gene regulates every other gene, and the fact that not every transcription factor regulates every gene can be exploited to prune significantly the number of coefficients for network identification.

A related concept that can be exploited is the knowledge that the low dimensional connection structures in these networks obey regular hierarchies, which create opportunities for structured model identification. Many biophysical networks can be decomposed into modular components that recur across and within given organisms. One hierarchical classification is to label the top level as a *network*, which is comprised of interacting regulatory *motifs* consisting of groups of 2–4 genes (Lee et al., 2002; Shen-Orr et al., 2002; Zak et al., 2003). At the lowest level in this hierarchy is the *module* that describes transcriptional regulation, of which a nice example is given in Barkai and Leibler, 2000. At the *motif* level, one can use pattern searching techniques to determine the frequency of occurrence of these simple motifs (Shen-Orr et al., 2002), leading to the postulation that these are basic building blocks in biological networks. Many of these components have direct analogs in system engineering architectures. Consider the three dominant network motifs found in *E. coli* (Shen-Orr et al., 2002): (i) feed-forward loop, (ii) single input module, and (iii) densely overlapping regulon. Similar studies in a completely different organism, *S. cerevisiae*, yielded six related or overlapping network motifs (Lee et al., 2002): (i) autoregulatory motif, (ii) feed-forward loop, (iii) multi-component loop, (iv) regulator chain, (v) single input module, and (vi) multi-input module.

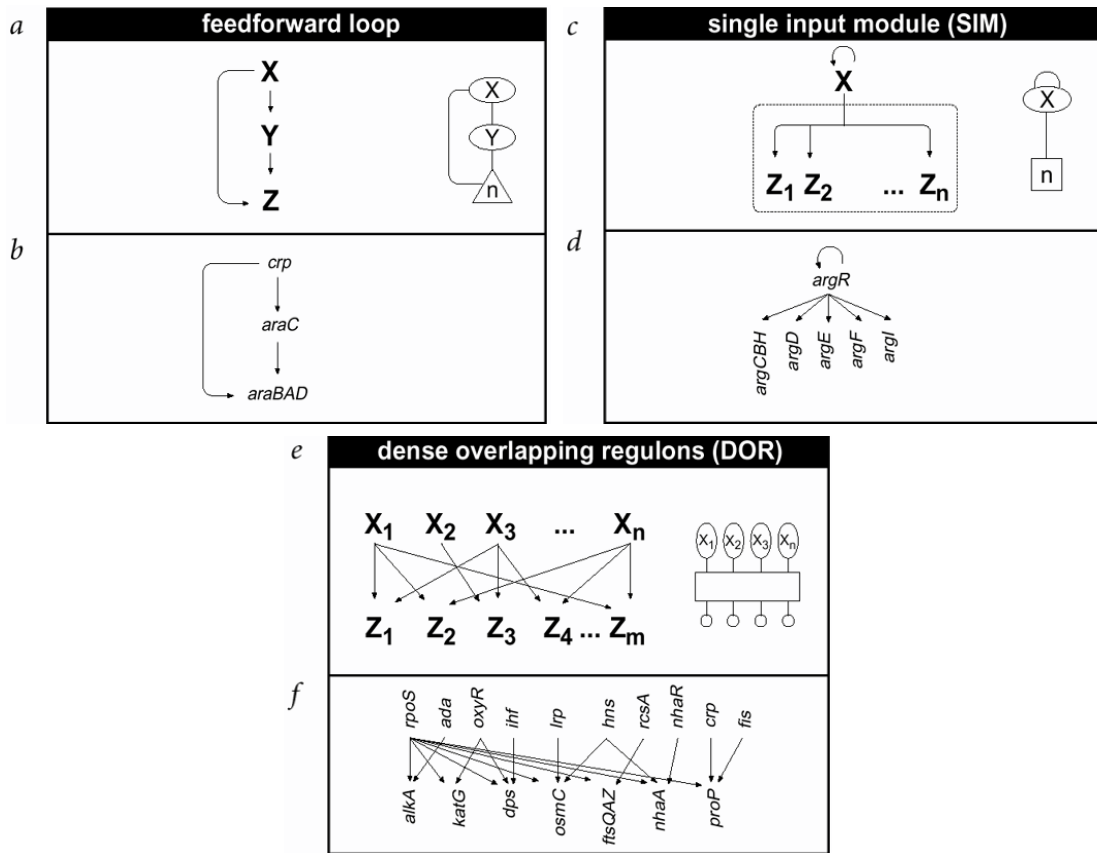


Figure 3.3. Network motifs found in the E.Coli transcriptional regulation network.

Beyond structural classification, one can analyze these motifs for their functional character, as shown by Wolf and Arkin, 2003, and again, one finds the recurring *dynamic* functional motifs in circuits and signal processing: (i) switches, (ii) oscillators, (iii) amplitude filters, (iv) bandpass filters, (v) memory, (vi) noise filters, and (viii) noise amplifiers.

In effect, these studies demonstrate that, in both eukaryotic and prokaryotic systems, cell function is controlled by sophisticated networks of control loops that are cascading onto, and interconnected with, other (transcriptional) control loops. The noteworthy insight here is that the complex networks that underlie biological regulation appear to be constructed of elementary *systems* components, not unlike a digital circuit. This creates opportunity for the network inference methods that incorporate such knowledge via constrained search methods, or exploiting prior knowledge in Bayesian frameworks.

In addition to the two classes of models mentioned previously (based on complexity and detail), there is an intermediate class consisting of optimization-based models. In many respects, this class has a hybrid character of empiricism and fundamental details. The underlying assumption is that cells have been organized over evolutionary time scales to optimize their operations in a manner consistent with mathematical principles of optimality. The cybernetic approach developed by Ramkrishna and co-workers (Varner and Ramkrishna, 1998) is founded on a simple principle; evolution has programmed or conditioned biological systems to optimally achieve physiological objectives. This straightforward concept can be translated into a set of optimal resource allocation problems that are solved at every time step in parallel with the model mass balances (basic metabolic network model). Thus, at every instant in time, gene expression and enzyme activity is rationalized as choice between sets of competing alternatives each with a relative cost and benefit for the organism. Mathematically, this can be translated into an instantaneous objective function. The potential shortcoming is a limited handling of more flexible objective functions that are commonly observed in biological systems. An alternative approach is the Flux Balance Analysis (FBA) (Watson, 1986), in which a suitable linear programming problem is posed and solved (Edwards et al., 1999). The resulting model is not a dynamic model, and does not yield an analytical formulation, but the computational solution

time is modest, and the approach has yielded success for a number of biological examples. Essential to the development of the model are the formulation of the system constraints, consisting of: (i) stoichiometric constraints that represent flux balances; (ii) thermodynamic constraints to restrict the directional flow through enzymatic reactions; and (iii) physicochemical capacity constraints to account for maximum flux through individual reactions. Recent extensions have addressed the problem of regulation by including additional time-dependent constraints in the formulation. The incorporation of transcriptional regulatory events in the FBA framework has extended the validity of the methodology for a number of complex dynamic system responses (Covert et al., 2001). In an alternate formulation, dynamic mechanistic details are incorporated as constraints leading to a dynamic FBA extension (Mahadevan et al., 2002).

As noted several times in this report, *dynamic* behavior is an essential property of complex biophysical networks that must be captured in models of those networks. There are some preliminary ideas in capturing network behavior in the form of dynamic models—both discrete time (Hartemink et al., 2002) and continuous (Zak et al., 2004). Many challenges exist in developing dynamic models from the type of data that is typically generated in the corresponding experiments, including: (i) sampling rate is rarely uniform and (ii) data is often ~~the~~ combined with other labs, introducing a number of biases. The previously noted problems of the curse of dimensionality are more pronounced in the case of dynamic models, if one augments the network interconnection dimensionality with a large number of possible dynamic states (activated, repressed, silenced, etc.), let alone the full continuum of dynamic response.

VALIDATION, ITERATION, DISCRIMINATION, AND IDENTIFIABILITY

One of the major issues in the reverse engineering of a genetic regulatory network is the challenge of *uniquely* identifying the gene interactions (i.e., model parameters) from experimental data, such as gene expression profiling. This issue, known as identifiability in control theory (Ljung, 1999), deals with the information content of the data; the quantity and quality of the measurements with respect to the model parameters. Recent work in the U.S. and in Europe on the identifiability of gene networks revealed that full knowledge of gene interconnections and perfect measurements still could not guarantee full identifiability of gene interactions (Zak et al., 2003), and, furthermore, that improved experimental protocol was far more effective than increased measurements (J. Stelling, unpublished data, 2005). The latter study points to the fact that perturbations should be designed strategically. Typical knockouts involve so-called “direct effects” in which the expression level of various genes is altered in a network arrangement that involves direct connectivity to cis-regulatory elements of downstream genes (possible multiple cascades). An “indirect effect” can also be used in which a mediating component (e.g., mRNA) is introduced to correct an intermediate element in the direct action cascade described previously.

Coupled to this, noise in the measurements and the inherent stochastic nature of gene expression make practical identification of genetic regulatory networks difficult. In practice, the reverse engineering of a gene network should involve a careful design of the experiments using prior knowledge of the system, to obtain the most informative measurements. Further, this process should be iterative in which the result from each trial is used to better design the next experiment. Here, a measure of informativeness of data, such as the Fisher Information Matrix (FIM), can lead to a formal procedure for the optimal design of the experiment. Aside from the aspect of the quality of data, another practical limitation in most (if not all) of the reverse-engineering of a gene network is the limited quantity of data, in terms of sampling frequency and number of independent measurements. For example, although gene expression profiling can provide high-throughput data to estimate interactions among thousand of genes, this method still does not depict the protein-mediated regulatory effects. In many cases, parameter estimation from limited measurements suffers from stringent computational requirement and degeneracy, where many parameter combinations give similar agreement to the observed behavior. Here, measurement selection procedures can help identify the combination of measurements that give the best identifiability.

Given the iterative nature of this framework for model development and refinement of experimental protocols, a termination criterion must be established. In the application domain of systems engineering, it is understood that for certain experimental data, it is not possible to confirm whether the model is really valid; however, one can conclude whether the model is not contradicted by the given data (Poolla et al., 1994). Such model (in)validation tests can be formulated for the network inference problems described in this chapter, and are usually based on the difference between the simulated and measured output and some

statistics about these differences. Typical statistics for the model errors include maximum absolute value, mean value and variance. These methods are slowly migrating from the engineering domain and are likely to find greater application in systems biology as experimental methods are refined, and closer collaborations are developed between modelers and experimentalists.

REGRESSION

As described above, most of the network inference work undertaken to date has been aimed at elucidating relationships among components in a network. However, the operation of a network is generally important only within the context of a physiological function it carries out or regulates at the level of individual cells or cell populations (e.g., tissues). Thus, it is essential to consider efforts to elucidate relationships between network components (or, more appropriately, network component properties such as levels, states, locations, and/or activities) and downstream cellular behavioral functions. Computational models for these kinds of “signal-response” relationships are much more difficult to formulate than for the more commonly-studied “cue-signal” relationships by which network component properties are governed by extracellular (or intracellular) stimuli, because the biochemical and biophysical processes involved in the former are much less well understood as well as less proximal. Nonetheless, several research efforts are underway that take as their objective the development of inferring “signal-response” relationships for network component regulation of cell functions. One class of methods which appear to be useful for inferring dependence of cell functions on network component properties are those founded on principal components analysis (Janes, 2004), including a variant termed network components analysis (Kao, 2004). This class of methods permits determination of the most critical combinations of network component properties for correlation, and even prediction, of functional responses. A second class of methods also being employed for this purpose is decision tree analysis (Hautaniemi, 2005). An advantage of decision tree analysis is that the combinations of network component properties associated with functional responses are explicitly delineated in direct manner, permitting the “logic” of how the network component properties combine to govern functional behavior to be viewed and interpreted easily. Finally, Bayesian network analysis has been used for a similar purpose (Woolf, 2005). Here, the “logic” connecting network component properties to one another and to the functional behavioral response is more complex to interpret, but is nonetheless available.

COMPARISON OF EFFORTS IN EUROPE, JAPAN, AND THE U.S.

Implicit in the preceding sections was a comparative analysis of efforts in the U.S., Japan, and Europe (as well as other regions such as Israel) by virtue of the cited references. For the purposes of the study, we outline from additional specific highlights in this area of network inference. It is worth noting that many of the ideas described in this chapter fall into the area of *Bioinformatics*, which has arguably gained a foothold in all of the geographical regions considered, with consideration to research, education, and infrastructure.

In the region of Japan, there were numerous significant advances in the area of network inference. The Kitano Laboratory (Symbiotic Systems Project) plays a leading role in the development of the Systems Biology Markup Language (SBML) and the establishment of standards in modeling biological systems. In addition, they are conducting research in the area of regression algorithms for network inference. The RIKEN Yokohama Institute was, as noted elsewhere, an extremely large-scale operation, and there were numerous laboratories addressing important problems in network inference, including: (i) inference algorithms (cooperatively coevolutionary), (ii) the formation of a consortium for the study of receptor tyrosine kinase regulatory networks, and (iii) the dynamic profiling of regulation in circadian networks by the Ueda Laboratory (Kobe). The Miyano Laboratory (U. Tokyo) was also developing regression algorithms for network inference of gene regulatory networks in yeast, and notably, they are conducting drug development studies with pharmaceutical companies. The Computational Biology Research Center (Tokyo) was conducting network inference studies for application to lung cancer. The Kanehisa Laboratory (Kyoto University), well known for the Kyoto encyclopedia of genes and genomes (KEGG) database, is conducting research in the reconstruction of dynamic networks via kernel methods, and is also enabling portability of the KEGG database networks to SBML and genomic object net (GON) models formats. This can be useful for the ultimate development of mechanistic large-scale network models. Finally, the Ito Laboratory at the University of Tokyo was utilizing heterogeneous measurements for network inference (MS, FRET, ChiP,

GATC-PCR, etc.), which is viewed as essential to overcome the identifiability issues described earlier in this chapter.

In Europe, the Max Planck Institute for Dynamics of Complex Systems (headed by Prof. Gilles) was one of the few groups that was explicitly addressing the challenges in identifiability, model iterations, perturbations, and design of experiment. At Humboldt University in Berlin, there were efforts described for dynamic modeling from microarray data, with application to the Ras pathway. The Reuss Laboratory (Stuttgart) described bioinformatics studies with application to cytochrome p450. The Armitage Laboratory (Oxford) emphasized bottom-up approaches for pathway analysis in histidine sensing. The Noble Laboratory (Oxford) challenged the strict bottom-up and top-down approaches, advocating a combination that starts in the middle.

A number of groups are vigorously active in the United States in the area of network inference, across all the technical areas outlined above. The Ideker group at UC-San Diego and the Ingber group at Harvard Medical School are pursuing Boolean approaches. The Gifford group at MIT and the Koller group at Stanford are employing Bayesian network approaches for investigation of networks focused on genomic data, and by the Lauffenburger group at MIT for investigation of networks focused on proteomic data. Regression methods, such as principal components analysis, network components analysis, and decision tree analysis are being used by the Liao group at UC-Los Angeles to study genomic networks and their relationship to physiological functions, and by the Lauffenburger group at MIT to study proteomic networks and their relationship to physiological functions. Cybernetic approaches are championed by the Ramkrishna group at Purdue, and flux balance methods—mainly for metabolic networks—by the Palsson group at UC-San Diego and the Stephanopoulos group at MIT.

SUMMARY

In summary, one finds numerous network inference studies in all of the regions described with the U.S., Japanese, and Israeli groups leading in the development of methodologies. All regions showed exciting application studies, with significant potential for “success stories” to emerge in the coming years.

The encouraging trends that were observed included: (i) multiple, complementary approaches to the regression of models for network inference, (ii) motifs and modules being incorporated into network inference methods, (iii) a nice interplay emerging between the classical static network databases and the formats for dynamic systems biology models (e.g., SBML), and (iv) a considerable amount of curricular development in this area (notably in bioinformatics).

Of concern was the fact that the issues of: (i) explicit incorporation of dynamics, (ii) identifiability and (in)validation of models, and (iii) model iterations with design of experiment, were receiving only modest attention in the regions, with noteworthy efforts in the U.S. and Europe (particularly Germany). There were many reported examples of researchers identifying large numbers of parameters from relatively small data sets. However, there appear to be a number of groups working towards solutions to these challenges, and considerable progress can be expected in the next two to three years.

REFERENCES

- Barkai, N. and S. Leibler. 2000. Circadian Clocks Limited by Noise. *Nature* 403: 267–8.
- Covert, M. W., C. H. Schilling and B. Palsson. 2001. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 213: 73–88.
- Edwards J. S., R. Ramakrishna, C. H. Schilling and B. O. Palsson. 1999. Metabolic Flux Balance Analysis. In: *Metabolic Engineering* (S. Y. Lee and E. T. Papoutsakis, Eds.). pp. 13–57. Marcel Dekker.
- Hartemink A. J., D. K. Gifford, T. S. Jaakola, R. A. Young. 2002. Combining Location and Expression Data for Principled Discovery of Genetic Regulatory Network Models. *Proc Pac Symp Biocomput* 7: 437–449.
- Hautaniemi S., S. Kharait, A. Iwabu, A. Wells, and D. A. Lauffenburger. 2005. Modeling of signal-response cascades using decision tree analysis. *Bioinformatics*, in press.
- Henry C. M. 2005. Systems Biology: Measurement and Modeling Approaches Bring a Big-Picture View of Biology and May Improve Drug Discovery and Development. *Chem & Eng News* 83(7): 47–55.

- Ideker T. E., V. Thorsson, and R. M. Karp. 2000. Discovery of Regulatory Interactions Through Perturbation: Inference and Experimental Design. *Pac Symp BioComputing* 5:302–313.
- Janes K. A., J. R. Kelly, S. Gaudet, J. G. Albeck, P. K. Sorger, and D. A. Lauffenburger. 2004. Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J Comput Biol* 11: 544–561.
- Kao K. C., Y. L. Yang, R. Boscolo, C. Sabatti, V. Roychowdhury, and J. C. Liao. 2004. Transcriptome-based determination of multiple transcription regulator activities in *Escherichia coli* by using network component analysis. *Proc Natl Acad Sci U S A* 101: 641–646.
- Kimura S., K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya. 2005. Inference of S-system Models of Genetic Networks using a Cooperative Coevolutionary Algorithm. *Bioinformatics* 21: 1154–1163.
- Kyoda K., K. Baba, S. Onami, and H. Kitano. 2004. DBRF-MEGN Method: An Algorithm for Deducing Minimum Equivalent Gene Networks from Large-Scale Gene Expression Profiles of Gene Deletion Mutants. *Bioinformatics* 20: 2662–2675.
- Lee T. I., N. J. Rinaldi, F. Robert, D. T. Odom, Z. bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford and R. A. Young. 2002. Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science* 298: 799–804.
- Ljung L. 1999. System Identification: Theory for the User, Englewood Cliffs, NJ, Prentice Hall.
- Mahadevan R., J. Edwards and F. J. Doyle III. 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 83: 1331–1340.
- Nagasaki M., A. Doi, H. Matsuno, and S. Miyano. 2004. A Versatile Petri Net Based Architecture for Modeling and Simulation of Complex Biological Processes. *Genome Informatics* 15: 180–197.
- Pe'er D., A. Regev, G. Elidan, and N. Friedman. 2001. Inferring Subnetworks from Perturbed Expression Profiles. *Bioinformatics* 17: S215–S224.
- Poolla K., P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal. 1994. A Time-Domain Approach to Model Validation. *IEEE Trans Automat Contr* 39: 951–959.
- Shen-Orr S. S., R. Milo, S. Mangan, and U. Alon. 2002. Network Motifs in the Transcriptional Regulation Network of *Escherichia coli*. *Nature Genetics* 31: 64–68.
- Stelling J. 2004. Mathematical Issues in Systems Biology. *Current Opinion in Microbiology* 7: 513–518.
- Varner J. and D. Ramkrishna. 1998. Application of cybernetic models to metabolic engineering: Investigation of storage pathways. *Biotech Bioeng* 58: 282–291.
- Watson M. R., 1986. A discrete model of bacterial metabolism. *Comp Appl Biosciences* 2: 23–27.
- Wolf D. M. and A. P. Arkin. 2003. Motifs, Modules, and Games in Bacteria. *Current Opinion in Microbiology* 6: 125–134.
- Woolf P. J., W. Prudhomme, L. Daheron, G. Q. Daley, and D. A. Lauffenburger. 2005. Bayesian analysis of signaling networks governing embryonic stem cell fate decisions. *Bioinformatics* 21: 741–753.
- Zak D., G. Gonye, J. S. Schwaber and F. J. Doyle III. 2003. Importance of Input Perturbations and Stochastic Gene Expression in the Reverse Engineering of Genetic Regulatory Networks: Insights from an Identifiability Analysis of an *In Silico* Network. *Genome Research* 13: 2396–2405.
- Zak D., R. K. Pearson, R. Vadigepalli, G. Gonye, J. S. Schwaber and F. J. Doyle III. 2004. Continuous-time identification of gene expression models. *Omic* 7: 373–386.