Published in final edited form as:
Curr Opin Biotechnol. 2007 August ; 18(4): 365–370.

Biochemical and Statistical Network Models for Systems Biology

Nathan D. Price[+] and **Ilya Shmulevich**[*]
Institute for Systems Biology, 1441 N 34th Street, Seattle, WA 98103

Introduction

The creation of models of the integrated functions of genes and proteins in cells is of fundamental and immediate importance to the emerging field of computational systems biology. Some of the most successful attempts at cell-scale modeling to date have been based on piecing together networks that represent hundreds of experimentally-determined biochemical interactions, while others have been very successful at inferring statistical networks from large amounts of high-throughput data. These networks (metabolic, regulatory, or signaling) can be analyzed, and predictions about cellular behavior made and tested. Many types of models have been built and applied to study cellular behavior and in this review we focus on two broad types: biochemical network models and statistical inference models. Through iterative model prediction, experimentation, and network refinement, the molecular circuitry and functions of biological networks can be elucidated. The construction of genome-scale models that integrate the myriad components that produce cellular behavior is a fundamental goal of systems biology today.

Biochemical Reaction Networks

Biochemical reaction networks represent the underlying chemistry of the system, and thus at a minimum represent stoichiometric relationships between inter-converted biomolecules. The stoichiometry of biochemical reaction networks can now be reconstructed at the genome-scale, and at smaller scale with sufficient detail to generate kinetic models. These biochemical reaction networks represent many years of accumulated experimental data and can be interrogated *in silico* to determine their functional states. Genome-scale models based on biochemical networks provide a comprehensive, yet concise, description of cellular functions.

For metabolism the reconstruction of the biochemical reaction network is a well-established procedure [1–7], while methods for the reconstruction of the associated regulatory [8,9] and signaling networks [10–12] with stoichiometric detail are being developed. The typically used formalism is to reconstruct the stoichiometric matrix, where each row represents a molecular compound and each column represents a reaction. For metabolism, these networks are often focused on just the metabolites, where the existence of a protein that catalyzes this reaction is used to allow that reaction to be present in the network. It is also possible (and truer to the realities in the system) to represent the proteins themselves as compounds in the network, which enables the integration of proteomics, metabolomics, and flux data (Figure 1). For regulatory and signaling networks, the inclusion of the proteins as compounds is essential. This process of reconstructing biochemical reaction networks has been referred to as the “two-dimensional annotation” of genomes [13].

* To whom correspondence should be addressed: ishmulevich@sytemsbiology.org
[+]Present Address: Department of Chemical and Biomolecular Engineering University of Illinois, Urbana-Champaign

Genome-scale models can be generated using a constraint-based approach, as has been reviewed in detail elsewhere [14]. This modeling process involves a three-step procedure. First, the biochemical network is reconstructed, as described above. Second, the physico-chemical constraints under which the reconstructed network operates are represented mathematically. Such constraints are based upon properties such as enzyme capacity, reaction stoichiometry and thermodynamics associated with reaction directionality and biochemical loops. This approach is most commonly used to describe fluxes through biochemical networks at steady state (i.e. flux balance analysis). The statement of constraints leads to the definition of a solution space that contains all non-excluded network states, describing the possible functions of the reconstructed network or all the allowable phenotypes. The third step is the determination of the possible solutions in this space that correspond to physiologically meaningful states. This constraint-based modeling procedure has been successfully utilized to study phenotypes in various model [15–18] and infectious [19,20] microorganisms. Recently, this approach has passed a significant milestone, namely the reconstruction of the human metabolic network [1]. Constraint-based analysis has been applied to selected human systems in the past [21,22], but with this reconstruction in place, this approach is poised to begin to have a much more significant impact on modeling human systems.

Stoichiometric reaction networks can be expanded into dynamic models using standard differential equation models and the addition of information in the form of kinetic rate constants. These systems, unlike the usual constraint-based approach, do not make the steady-state assumption for all compounds but rather can simulate detailed dynamic behavior. Thus, they can model a greater degree of complexity for a given network and simultaneously account for both concentrations of compounds and fluxes through reactions. The disadvantage of these systems compared with constraint-based analysis is that they require many more parameters and are thus more data intensive to create and/or can be more prone to overfitting when large numbers of parameters are unknown. Thus, these models are not typically created at the whole cell scale, with a notable exception being cell-scale models of red blood cell metabolism [23, 24].

In recent years, these types of dynamic models have been used very successfully to provide mechanistic models of, for example, key signaling pathways involved in critical physiological and pathophysiological processes. For example, a differential equation model of NF-kB signaling [25] was updated and used to elucidate the role of tumor necrosis factor in controlling the sustained phase of lipopolysaccharide-induced IKK activation [26,27], demonstrating the usefulness of these models for deciphering functions of biological networks. Another example was the use of a model to study the integrated dynamic behavior of a network of epidermal growth factor receptor family members [28]. Christopher *et al*. have constructed a dynamic simulator of gene expression and signaling networks in a human cancer cell [29]. In the long run, such dynamic models set the stage for personalized medicine by offering the promise of rational drug design and control of the outcome of induced molecular perturbations.

Statistical Influence Networks

A second approach to modeling biological networks also holds tremendous potential for advancing knowledge of biology – namely, statistical influence networks. The growing amounts of microarray gene expression data along with improvements in data fidelity are now making it possible to make robust statistical systems-level inferences about the structure and dynamics of biomolecular control mechanisms, such as transcriptional regulatory networks.

Many approaches attempt to infer relationships between gene expression measurements using deterministic or stochastic formalisms. The fundamental idea behind these approaches is that models that faithfully capture such relationships have predictive capacity as regards system

behavior and can be used to gain insight about system-wide properties, such as steady-state behavior or responses to perturbations or specific stimuli. There are a number of ways in which such relationships can be represented, both in the discrete and continuous domains.

One popular modeling approach that captures the nonlinear multivariate relationships exhibited by biological control circuits, such as gene regulatory networks, is the class of Boolean networks, which owes its inception to the work of Stuart Kauffman in the late 1960s [30]. In the Boolean network model, the variables (e.g., genes or proteins) are binary-valued, meaning that their states can be either on or off, and the relationships between the variables are captured by Boolean functions. Each (target) gene is assigned a Boolean rule that determines its value as a function of the values of a set of other (predictor) genes, possibly including the target gene itself. System dynamics are generated by updating the Boolean functions, either synchronously or asynchronously, causing the system to transition from state to state in accordance with its Boolean update rules, where a state is a binary representation of the activities of all of the variables in the system (i.e., a binary vector representing the genes that are on or off at any given time).

Boolean network models have been constructed and analyzed for a number of developmental and physiological processes. For example, Albert *et al.* constructed a Boolean network model for a subset of genes of the fruitfly *Drosophila melanogaster*, which describes different stable gene expression patterns in the segmentation process of the developing embryo [31]. The steady-state behavior of this model was in excellent agreement with experimentally observed expression patterns under wild type and several gene mutation conditions. This study highlighted the importance of the network topology in determining biologically correct asymptotic states of the system. Indeed, when the segment polarity gene control network was modeled with more detailed kinetic models, such as systems of nonlinear differential equations, exceptional robustness to changes in the kinetic parameters was observed [32].

Boolean networks have also been used to model the yeast and mammalian cell cycle [33,34]. Li *et al*. demonstrated that the cell cycle sequence of protein states, which is a globally attracting trajectory of the dynamics, is extremely robust with respect to small perturbations to the network. The Boolean network formalism was also recently used to model systems-level regulation of the host immune response, which resulted in experimentally validated predictions regarding cytokine regulation and the effects of perturbations [35]. Boolean rules can be learned from gene expression data using methods from computational learning theory [36] and statistical signal processing [37].

A limitation of the Boolean network approach is its inherent determinism. Because of the inherent stochasticity of gene expression and the uncertainty associated with the measurement process due to experimental noise and possible interacting latent variables (e.g. protein concentrations or activation states that are not measured), the inference of a single deterministic function may result in poor predictive accuracy, particularly in the context of small sample sizes (e.g., number of microarrays) relative to the number of genes.

One approach to “absorb” this uncertainty is to infer a number of simple functions (having few variables), each of which performs relatively well, and probabilistically synthesize them into a stochastic model, called a probabilistic Boolean network (PBN) [38]. The contribution of each function is proportional to its determinative potential as captured by statistical measures such as the coefficient of determination, which are estimated from the data [37]. The dynamical behavior of PBNs can be studied using the theory of Markov chains, which allows the determination of steady-state behavior as well as systematic intervention and control strategies designed to alter system behavior in a specified manner [39–41]. The PBN formalism has been used to construct networks in the context of several cancer studies, including glioma [42],

melanoma [41], and leukemia [40]. PBNs, which are stochastic rule-based models, bear a close relationship to dynamic Bayesian networks [43] – a popular model class for representing the dynamics of gene expression.

Bayesian networks are graphical models that have been used to represent conditional dependencies and independencies among the variables corresponding to gene expression measurements [44]. One limitation of Bayesian networks for modeling genetic networks is that these models must be in the form of directed acyclic graphs and, as such, are not able to represent feedback control mechanisms. Dynamic Bayesian networks, on the other hand, are Bayesian networks that are capable of representing temporal processes [45,46] that may include such feedback loops. Since not all causal relationships can be inferred from correlation data, meaning that there can be different directed graphs that explain the data equally well, intervention experiments where genes are manipulated by overexpression or deletion have been proposed to learn networks [47]. The Bayesian network formalism has also been used to infer signaling networks from multicolor flow cytometry data [48].

There exist a number of other approaches for inferring large-scale molecular regulatory networks from high-throughput data sets. One example is a method, called the Inferelator, that selects the most likely regulators of a given gene using a nonlinear model that can incorporate combinatorial nonlinear influences of a regulator on target gene expression, coupled with a sparse regression approach to avoid overfitting [49]. In order to constrain the network inference, the Inferelator performs a preprocessing step of biclustering using the cMonkey algorithm [50], which results in a reduction of dimensionality and places the inferred interactions into experiment-specific contexts. The authors used this approach to construct a model of transcriptional regulation in *Halobacterium* that relates 80 transcription factors to 500 predicted gene targets.

Another method that predicts functional associations among genes by extracting statistical dependencies between gene expression measurements is the ARACNe algorithm [51]. This information-theoretic method uses a pairwise mutual information criterion across gene expression profiles to determine significant interactions. A key step in the method is the use of the so-called data processing inequality, which is intended to eliminate indirect relationships in which two genes are co-regulated through one or more intermediaries. Thus, the relationships in the final reconstructed network are more likely to represent the direct regulatory interactions. The ARACNe algorithm was applied to 336 genome-wide expression profiles of human B cells, resulting in the identification of MYC as a major regulatory hub along with newly identified and validated MYC targets [52].

A method related to the ARACNe algorithm, called the context likelihood of relatedness (CLR), also uses the mutual information measure but applies an adaptive background correction step to eliminate false correlations and indirect influences [53]. CLR was applied to a compendium of 445 *E. coli* microarray experiments collected under various conditions and compared to other inference algorithms on the same data set. The CLR algorithm had superior performance as compared to the other algorithms, which included Bayesian networks and ARACNe, when tested against experimentally determined interactions curated in the RegulonDB database. It also identified many novel interactions, a number of which were verified with chromatin immunoprecipitation [53].

Comparison of Network Analysis Approaches

There are fundamental differences between the biochemical and statistical classes of network modeling described herein. One clear difference is the manner in which these underlying networks are reconstructed. For biochemical networks, reconstruction is typically a work-

intensive process requiring significant biochemical characterization with little network inference done (other than inferences of single gene function for catalyzing a reaction based on sequence homology). Thus, the ability to rapidly generate networks for organisms that are relatively uncharacterized from high-throughput data is an inherent advantage of the inferred statistical networks. One advantage of the biochemical network models is that, once reconstructed, the networks are not as subject to change (other than addition) since many of the links are based directly on biochemical evidence. Inferred networks, on the other hand, can undergo substantial changes in light of additional data. Another common difference, although not fundamental, is that constraint-based biochemical network models have mostly been used to model flux, whereas inference networks have mostly been used to predict substance amounts (e.g. mRNA expression). One way this can be thought of is that the biochemical network models currently link more closely to functional phenotype (i.e. fluxes) [54], while the inferred networks relate more directly to available high-throughput data (i.e. transcriptomes). The kinetic biochemical network models, of course, have the capacity to account for both flux and abundance, but suffer from the limitation that they are by far the most data intensive to reconstruct. Another key advantage of biochemical reaction networks, stemming from their basis in chemistry, is that physico-chemical laws apply, such as mass-energy balance, while such laws are not generally applicable to the inferred networks. Of course, the advantage of the inferred networks is that, since they do not need to be mechanistic or require biochemical detail, they can be applied very broadly to systems that are not yet well characterized and can link very disparate data types as long as underlying correlations exist. In summary, both modeling types are essential to contemporary computational systems biology and provide advantages over each other in different settings.

One interesting challenge going forward is whether hybrid models that take advantage of the strengths of the different modeling approaches can be constructed to move us further towards the goal of predictive whole-cell models and beyond. Early attempts have been done to link Boolean regulatory networks with constraint-based flux models [8], but the extent to which these approaches can be married to provide significant advances in our ability to model biological networks remains an open question.

Acknowledgements

NDP was supported by a postdoctoral fellowship from the American Cancer Society (PF-06-062-01-MGO).

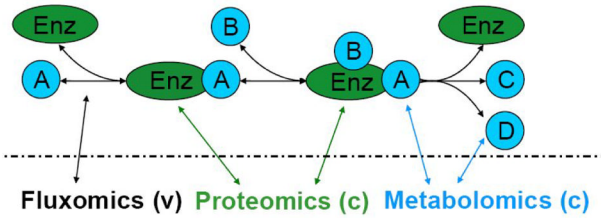
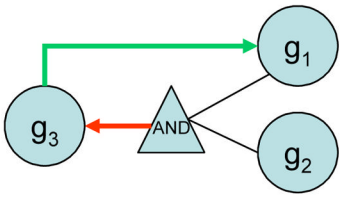
Reference and recommended reading

1. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci U S A 2007;104:1777–1782. [PubMed: 17267599]**A global reconstruction of known metabolism in humans, this opens the door for constraint-based modeling, and eventually kinetic modeling, to be broadly used in studying human systems
2. Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 2004;14:1298–1309. [PubMed: 15197165]
3. Reed JL, Palsson BO. Thirteen years of building constraint-based in silico models of Escherichia coli. J Bacteriol 2003;185:2692–2699. [PubMed: 12700248]
4. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol 2003;4:R54. [PubMed: 12952533]
5. Thiele I, Vo TD, Price ND, Palsson BO. Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. J Bacteriol 2005;187:5818–5830. [PubMed: 16077130]

6. Heinemann M, Kummel A, Ruinatscha R, Panke S. In silico genome-scale reconstruction and validation of the Staphylococcus aureus metabolic network. Biotechnol Bioeng 2005;92:850–864. [PubMed: 16155945]
7. Francke C, Siezen RJ, Teusink B. Reconstructing the metabolic network of a bacterium from its genome. Trends Microbiol 2005;13:550–558. [PubMed: 16169729]
8. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. Nature 2004;429:92–96. [PubMed: 15129285] **Integrates constraint-based modeling and Boolean networks to identify a large number of novel putative regulators in *E. coli*
9. Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO. Matrix formalism to describe functional states of transcriptional regulatory systems. PLoS Comput Biol 2006;2:e101. [PubMed: 16895435]
10. Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and analysis of their properties. Nat Rev Mol Cell Biol 2005;6:99–111. [PubMed: 15654321]
11. Papin JA, Palsson BO. The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis. Biophys J 2004;87:37–46. [PubMed: 15240442]
12. Papin JA, Palsson BO. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. J Theor Biol 2004;227:283–297. [PubMed: 14990392]
13. Palsson B. Two-dimensional annotation of genomes. Nat Biotechnol 2004;22:1218–1219. [PubMed: 15470454]
14. Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol 2004;2:886–897. [PubMed: 15494745]*A comprehensive review of constraint-based modeling and its applications
15. Edwards JS, Ibarra RU, Palsson BO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 2001;19:125–130. [PubMed: 11175725]
16. Ibarra RU, Edwards JS, Palsson BO. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 2002;420:186–189. [PubMed: 12432395]
17. Famili I, Forster J, Nielsen J, Palsson BO. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci U S A 2003;100:13134–13139. [PubMed: 14578455]
18. Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabasi AL. Global organization of metabolic fluxes in the bacterium Escherichia coli. Nature 2004;427:839–843. [PubMed: 14985762]
19. Papin JA, Price ND, Edwards JS, Palsson BB. The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy. J Theor Biol 2002;215:67–82. [PubMed: 12051985]
20. Price ND, Papin JA, Palsson BO. Determination of redundancy and systems properties of the metabolic network of Helicobacter pylori using genome-scale extreme pathway analysis. Genome Res 2002;12:760–769. [PubMed: 11997342]
21. Thiele I, Price ND, Vo TD, Palsson BO. Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. J Biol Chem 2005;280:11683–11695. [PubMed: 15572364]
22. Price ND, Schellenberger J, Palsson BO. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. Biophys J 2004;87:2172–2186. [PubMed: 15454420]
23. Mulquiney, PJ.; Kuchel, PW. Modelling metabolism with Mathematica detailed examples including erythrocyte metabolism. Boca Raton, Fla: CRC Press; 2003.
24. Jamshidi N, Edwards JS, Fahland T, Church GM, Palsson BO. Dynamic simulation of the human red blood cell metabolic network. Bioinformatics 2001;17:286–287. [PubMed: 11294796]
25. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation. Science 2002;298:1241–1245. [PubMed: 12424381]
26. Covert MW, Leung TH, Gaston JE, Baltimore D. Achieving stability of lipopolysaccharide-induced NF-kappaB activation. Science 2005;309:1854–1857. [PubMed: 16166516]**A clear demonstration of the value of kinetic models to elucidate function of important biological networks
27. Werner SL, Barken D, Hoffmann A. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. Science 2005;309:1857–1861. [PubMed: 16166517]**A clear demonstration of the value of kinetic models to elucidate function of important biological networks

28. Hendriks BS, Wiley HS, Lauffenburger D. HER2-mediated effects on EGFR endosomal sorting: analysis of biophysical mechanisms. Biophys J 2003;85:2732–2745. [PubMed: 14507736]
29. Christopher R, Dhiman A, Fox J, Gendelman R, Haberitcher T, Kagle D, Spizz G, Khalil IG, Hill C. Data-driven computer simulation of human cancer cell. Ann N Y Acad Sci 2004;1020:132–153. [PubMed: 15208190]
30. Kauffman, SA. The origins of order: self organization and selection in evolution. New York: Oxford University Press; 1993. **A broad treatment of Boolean networks in the context of developmental and evolutionary biology
31. Albert R, Othmer HG. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. J Theor Biol 2003;223:1–18. [PubMed: 12782112]
32. von Dassow G, Meir E, Munro EM, Odell GM. The segment polarity network is a robust developmental module. Nature 2000;406:188–192. [PubMed: 10910359]
33. Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci U S A 2004;101:4781–4786. [PubMed: 15037758]
34. Faure A, Naldi A, Chaouiya C, Thieffry D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. Bioinformatics 2006;22:e124–131. [PubMed: 16873462]
35. Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R. Modeling Systems-Level Regulation of Host Immune Responses. PLoS Computational Biology 2007:e109.eor. [PubMed: 17559300] preprint
36. Lahdesmaki H, Shmulevich I, Yli-Harja O. On Learning Gene Regulatory Networks Under the Boolean Network Model. Machine Learning 2003;52:147–167.
37. Shmulevich I, Dougherty ER, Zhang W. From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. Proceedings of the IEEE 2002;90:1778–1792.*A review of Boolean and probabilistic Boolean network models, including inference from gene expression data, perturbation analysis, and intervention strategies
38. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics 2002;18:261–274. [PubMed: 11847074]
39. Shmulevich I, Dougherty ER, Zhang W. Gene perturbation and intervention in probabilistic Boolean networks. Bioinformatics 2002;18:1319–1331. [PubMed: 12376376]
40. Li H, Zhan M. Systematic intervention of transcription for identifying network response to disease and cellular phenotypes. Bioinformatics 2006;22:96–102. [PubMed: 16278241]
41. Pal R, Datta A, Bittner ML, Dougherty ER. Intervention in context-sensitive probabilistic Boolean networks. Bioinformatics 2005;21:1211–1218. [PubMed: 15531600]
42. Hashimoto RF, Kim S, Shmulevich I, Zhang W, Bittner ML, Dougherty ER. Growing genetic regulatory networks from seed genes. Bioinformatics 2004;20:1241–1247. [PubMed: 14871865]
43. Lahdesmaki H, Hautaniemi S, Shmulevich I, Yli-Harja O. Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks. Signal Processing 2006;86:814–834. [PubMed: 17415411]
44. Friedman N. Inferring cellular networks using probabilistic graphical models. Science 2004;303:799–805. [PubMed: 14764868]*A review of Bayesian network modeling and inference in the context of gene expression data
45. Kim SY, Imoto S, Miyano S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. Brief Bioinform 2003;4:228–235. [PubMed: 14582517]
46. Zou M, Conzen SD. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 2005;21:71–79. [PubMed: 15308537]
47. Pournara I, Wernisch L. Reconstruction of gene networks using Bayesian learning and manipulation experiments. Bioinformatics 2004;20:2934–2942. [PubMed: 15180938]
48. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. Science 2005;308:523–529. [PubMed: 15845847]*A good example of the use of Bayesian networks for inferring signaling network structure

49. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol 2006;7:R36. [PubMed: 16686963]*An innovative approach based on nonlinear regression and model selection for large-scale transcriptional regulatory network inference
50. Reiss DJ, Baliga NS, Bonneau R. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. BMC Bioinformatics 2006;7:280. [PubMed: 16749936]
51. Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A. Reverse engineering cellular networks. Nat Protoc 2006;1:662–671. [PubMed: 17406294]
52. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. Nat Genet 2005;37:382–390. [PubMed: 15778709]*A powerful approach based on information theory for inferring large-scale transcriptional regulatory networks
53. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol 2007;5:e8. [PubMed: 17214507]
54. Sauer U. High-throughput phenomics: experimental methods for mapping fluxomes. Curr Opin Biotechnol 2004;15:58–63. [PubMed: 15102468]

Biochemical Reaction Networks	Statistical Influence Networks
	
Directly mechanistic	Not generally mechanistic
Require significant knowledge of the system	Can be applied without needing prior knowledge (although can be incorporated)
Broadly applicable where biochemistry is known	Broadly applicable without knowledge of biochemistry
Laws of physics and chemistry can be directly applied	Physico-chemical laws typically not applicable/applied
Relate more closely to phenotype (i.e. fluxes)	Relate more directly to high-throughput data (i.e. transcriptomes)
Once reconstructed from biochemical data, network not likely to change (other than additional reactions)	Additional data can lead to significant network rewiring

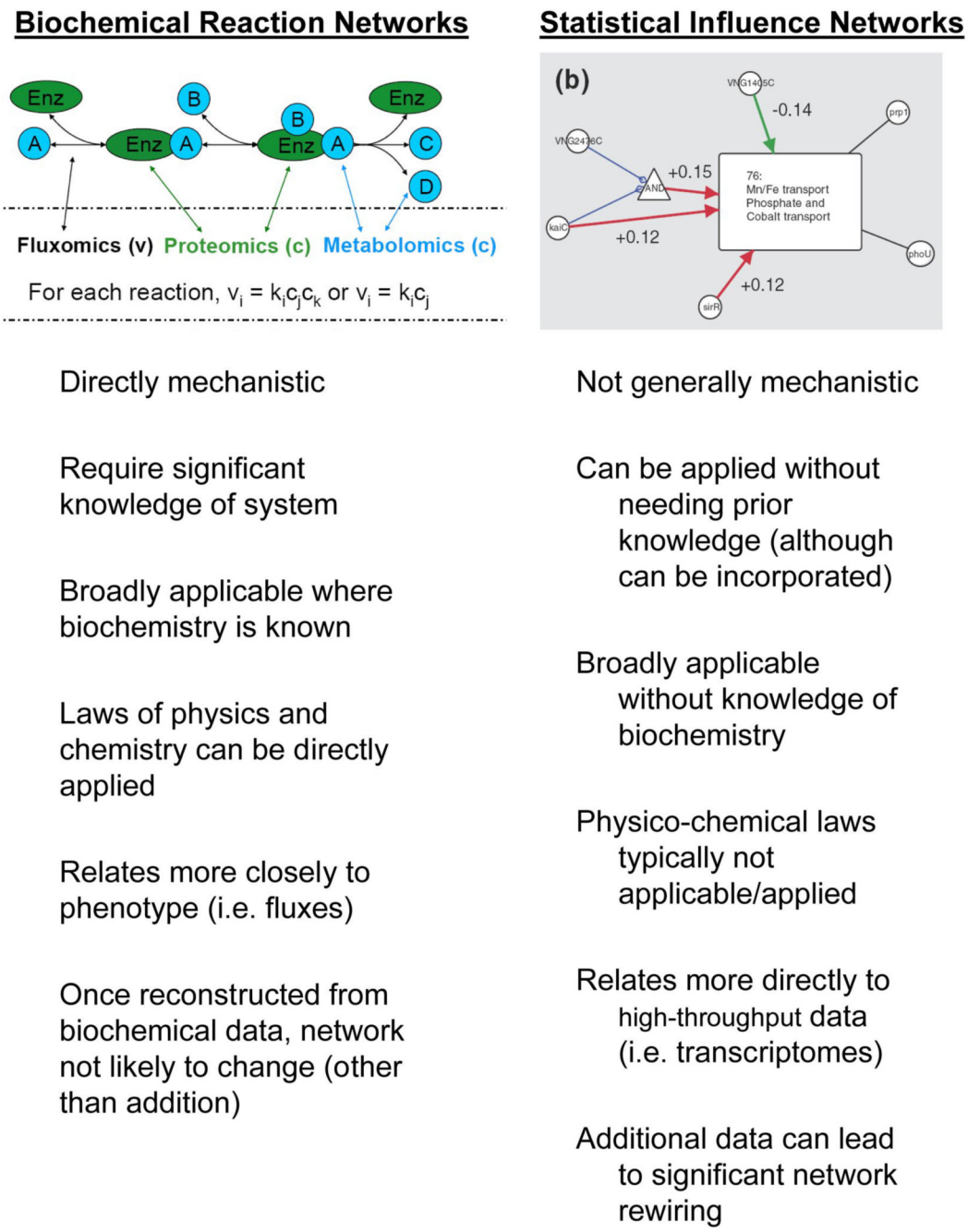


Figure 1.