31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Descriptive and Predictive Applications of Constraint-Based Metabolic Models

Jennifer L. Reed

***Abstract*—**Constraint-based models of metabolism are becoming available for an increasing number of organisms. These models can be used in combination with existing experimental data to describe the behavior of an organism and to analyze experimental observations in the context of a model. Such a descriptive application of the models can also allow for the integration of various types of data. Additionally, these models can be used in a predictive fashion to hypothesize the outcomes of new experiments. Comparing model predictions with experimental results allows for the iterative improvement of developed models and increases our understanding of the organism being studied. A number of recent examples of both descriptive and predictive applications of constraint-based models are discussed.

I. Introduction

A systems approach towards the study of biological networks is driven by both computational and experimental efforts. Diverse high-throughput datasets (including genomic, transcriptomic, proteomic, phenomic and metabolomic), which characterize components, interactions, and network states are becoming readily available. However, with this data comes a need to integrate and analyze these datasets, and computational models can facilitate this process.

Computational models have multiple roles in the field of systems biology, which can be both descriptive and predictive. Models as descriptive tools allow for the integration and analysis of a variety of large experimental datasets, such as genome annotations, gene expression, gene essentiality, phenotype microarrays, ChIP-chip and ChIP-seq, protein interactions, proteomic and metabolomic datasets. Significant challenges exist in integrating these datasets which can be attributed to the both size and non-unique mapping between different types of data (eg. metabolite and protein concentrations). Integration and analysis of data by computational models can help interpret experimental data and generate hypotheses regarding network components or component interactions.

Manuscript received April 23, 2009. Some of the reported work was supported in part by a Faculty Fellowship to J. L. Reed from the University of California, San Diego and funding from the DOE (DE-AC05–76RLO 1830).

J. L. Reed is University of Wisconsin-Madison, Madison, WI 53706 USA (phone: 608-262-0188; fax: 608-262-5434; e-mail: reed@ engr.wisc.edu).

Models, as predictive tools, allow for the simulation of a cell's phenotypic response to either environmental or genetic perturbations. By predicting phenotypic behavior, models can be used as design tools to engineer strains or environments that result in desired phenotypes. In both descriptive and predictive applications, models provide a way to generate and evaluate different hypotheses.

II. Constraint-Based models

A variety of modeling methods can be used to model metabolic and regulatory networks [1-3]. Constraint-based methods can be used to model genome-scale networks since less information (such as metabolite and protein concentrations and kinetic rate constants) is needed to construct them. These models can be developed with substantially less information than kinetic models, making them more tractable for less characterized organisms, where kinetic information is not available.

Constraint-based models of metabolism use three types of constraints to define a set of feasible flux distributions for a given metabolic network. Solutions that do not satisfy any of the imposed constraints are excluded from space of all feasible flux distributions. Application of additional constraints further reduces the solution space and, consequently, reduces the number of possible solutions a cell can utilize. Constraints that have been used in constraint-based metabolic models include steady-state mass balance, thermodynamic (regarding reaction reversibility), and enzymatic capacity constraints (where upper and lower limits restrict individual flux values) [4]. Steady-state mass balance constraints can be represented by the system of linear equations represented by:

$$\mathbf{S} \cdot \mathbf{v} = 0 \qquad \text{(Eq. 1)}$$

where **S** is the stoichiometric matrix describing all the reactions in the network and v is a vector describing fluxes through each of the reactions. Each column of **S** corresponds to an individual reaction and the rows of **S** correspond to the different metabolites. Equation 1 imposes the restriction that the total rate of production for each metabolite in the network must equal the total rate of consumption for that metabolite. In addition to steady-state mass balance constraints (Eq. 1), thermodynamic constraints and enzyme capacity constraints restrict the range of values for individual fluxes (v) in the network (Eq. 2).

$$v_{min} \leq v \leq v_{max} \quad \text{(Eq. 2)}$$

Enzyme capacity constraints provide upper (v_{max}) and lower (v_{min}) limits on the values a given flux can take. Application of thermodynamic constraints further restricts the range of flux values. If a reaction is irreversible then the corresponding flux must be greater than or equal to zero (hence, $v_{min} = 0$); while, reversible reactions can have negative flux values. More recent studies have directly incorporated calculations of the change in Gibbs free energy of reactions into the models [5-7], which allows for constraints on flux directionality using metabolite concentrations [7].

Given a set of constraints, models can be used to characterize the solution space and predict which flux distribution a cell is likely to use. Different constraint-based methods have been developed to characterize the solution space, identify physiologically relevant flux distributions [8-12], evaluate flux inter-dependencies [13-15], incorporate regulatory constraints [16-19], and identify and resolve model-data discrepancies [17, 20-23]. Flux balance analysis is one of the most commonly used constraint-based methods, where an objective function is used to identify flux distributions which have the maximum (or minimum) value for this objective function. A variety of objective functions have been proposed for microbial networks, with maximizing biomass production as the most commonly used [4, 8]. Recently, ATP yields and ATP production normalized to the total flux through metabolism have also been shown to be useful for predicting intracellular flux distributions [24]. To date there are dozens of genome-scale constraint-based models of metabolism available [25] (See Figure 1), and the modeling applications continue to grow as do the number of methods used to analyze the models [4, 26].

III. RESULTS

A. Constraint-Based Models as Descriptive Tools

Constraint-based metabolic models summarize the information available for an organisms metabolic network based on a number of sources, including genome annotations, primary literature, and on-line databases [27]. These genome-scale models account for all of the metabolic pathways in an organism, and as such represent a significant fraction of the genes in an organism (eg. ~28% of the genes in *Escherichia coli* are accounted for in the latest model [28]). Developed models can be useful not only as predictive tools (as described below), but also as descriptive tools that facilitate the evaluation, analysis, and interpretation of experimental data. Constraint-based models have been used to analyze various types of data, including gene essentiality, growth phenotypes, protein expression, and gene expression.

A model of *E. coli* metabolism, for example, was used to analyze lethal and non-lethal phenotypic data for single gene deletion strains from the Keio collection [29]. The large number of consistencies (~91%) between observations and model predictions provides an explanation for the observed phenotypes. Additionally, the discrepancies were used for generating hypothesis about the presence or regulation of alternative pathways in the organism. In addition to growth phenotypes, gene expression data can also be analyzed in the context of a model. A method was recently been developed by Sholmi et al. [12] which incorporates gene expression data in the calculation of intracellular fluxes. Here flux distributions are found by favoring fluxes through reactions of highly expressed genes and avoiding fluxes through reactions of lowly expressed genes. Thus, gene expression data in combination with the models can be used to find metabolic pathways which are likely active in the organism under a given condition.

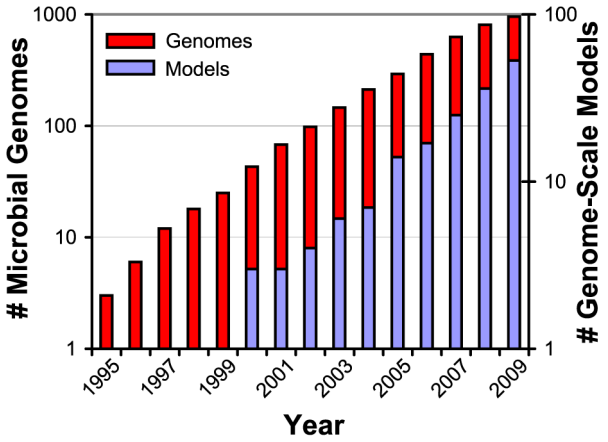

Figure 1.**Available Genome Sequences and Constraint-Based Metabolic Models.** Shown here is the exponential growth in the number of sequenced microbial genomes (red bars) and genome-scale metabolic models (blue bars). Both the number of sequenced organisms and developed models are growing at an exponential rate.

B. Constraint-Based Models as Predictive Tools

Constraint-based metabolic and regulatory models have also been successfully used to predict the metabolic and phenotypic behavior of cells. We have recently reconstructed the metabolic network for *Salmonella typhimurium* LT2 [30]. The corresponding constraint-based metabolic model for *S. typhimurium* was used to predict growth phenotypes (growth or no growth) on various carbon and nitrogen sources based on the presence of enzymes and transporters in the genome. Model predictions matched experimental growth measurements for ~80% of the cases that could be compared, with most errors being cases where the bacterium appears to have the pathways needed for catabolism but does use them experimentally [30]. Transcriptional regulation may explain some of the discrepancies between model predictions and experimental observations, as 13 of 21 incorrectly predicted carbon sources can be used as nitrogen sources experimentally, indicating that the required enzyme(s) may be expressed only during nitrogen limitation [30].

In this study, we also predicted whether *S. typhimurium* mutant strains defective in metabolic enzymes would be

virulent or avirulent and then compared our simulation results to experimental data. We found that model predictions were consistent with experiments for 80% of the cases examined. A large number of the incorrect predictions (6 out of 11) could be corrected by adjusting what nutrients were assumed to be present in the host-cell environment in the simulations [30].

We have recently developed a metabolic model for *Shewanella oneidensis* MR-1, which can use a wide variety of electron acceptors. From chemostat data we were able to estimate the growth associated ATP requirement—this value was substantially higher than those reported for other organisms. The model was then evaluated to identify futile cycles and other less energetically efficient pathways that could explain the high estimate for growth associated ATP requirement. These model predictions, about the possible use of less energetically efficient pathways, can be used to design experiments to test what enzymes may cause the organism's seemingly high ATP requirement (Pinchuk and Reed, unpublished data).

C. Iterative Improvement of Constraint-Based Models

While correct predictions are important for verifying the accuracy of a model, incorrect predictions can also be useful as well. These incorrect predictions can pinpoint problems in model formulations or lead to the discovery of new metabolic or regulatory connections in the networks being studied. For example, by looking at cases where the *S. typhimurium* model made incorrect predictions about mutant virulency we could identify changes needed in the list of nutrients that we had included in our simulations of the host-cell environment.

While discrepancies between model predictions and experimental results can be used to improve models, these discrepancies can also lead to new biological discoveries. We previously used discrepancies between simulated and observed *E. coli* growth phenotypes (model predicted no growth, but the cells were able to grow) to find new metabolic reactions which occur in *E. coli* metabolism [20]. In this study, we developed an algorithm to identify potentially missing metabolic and transport reactions from our model by analyzing growth phenotypes, and then screened knock-out mutants to find proteins responsible for catalyzing these missing reactions. Here models were used to identify what experimental observations were inconsistent with our current knowledge of *E. coli* metabolism and to hypothesize what reactions were missing in *E. coli* metabolism. This integration of computation and experimental efforts improved models of *E. coli* metabolism and expanded our understanding of metabolism in this organism [20].

Once accurate models are developed they can be used in a wide variety of applications, as has been recently reviewed for *E. coli* [26]. One such application is the design of strains for metabolic engineering, where the production of specific compounds is a desirable phenotype. Here models can be used to predict how metabolite production will be affected by changing an organism's metabolic network through removal of reactions. Recent studies have illustrated the usefulness of *E. coli* metabolic constraint-based models as tools for engineering strains for a the production of a variety of compounds, including lycopene [31], lactate [32] and valine [33].

IV. CONCLUSIONS

As illustrated in the above examples constraint-based models of metabolic networks can be used in both a retrospective or descriptive fashion to analyze existing experimental data, or in a prospective or predictive fashion to generate predictions about results for new experiments. The integration of modeling and experimental efforts benefits both model development and experimentation, where experimental data can improve models and models can aide in the interpretation and analysis of data. The development of new computational and experimental approaches will allow for the better understanding of cellular physiology in both the well and less characterized biological systems.

ACKNOWLEDGMENT

Recent research results summarized here is partly from work that was done in collaboration with a number of individuals including: A.Raghunathan, S. Shin, S. Daefler at Mt. Sinai School of Medicine; X. Zhang at UW-Madison, G. Pinchuk, J. Fredrickson, A. Belieav, M. Romine, O. Geydebrekht, E. Hill at PNNL; and A. Osterman, J. De Ingeniis at Burnham Institute.

REFERENCES

[1] G. Stephanopoulos, A. A. Aristidou, and J. Nielsen, *Metabolic engineering : principles and methodologies*. San Diego: Academic Press, 1998.

[2] Z. Szallasi, J. Stelling, and V. Periwal, *System modeling in cell biology : from concepts to nuts and bolts*. Cambridge, Mass.: MIT Press, 2006.

[3] S. P. Ellner and J. Guckenheimer, *Dynamic models in biology*. Princeton, N.J. ; Woodstock: Princeton University Press, 2006.

[4] N. D. Price, J. L. Reed, and B. O. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints," *Nat Rev Microbiol*, vol. 2, pp. 886-97, 2004.

[5] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, "Thermodynamics-based metabolic flux analysis," *Biophys J*, vol. 92, pp. 1792-805, 2007.

[6] C. S. Henry, M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis, "Genome-scale thermodynamic analysis of Escherichia coli metabolism," *Biophys J*, vol. 90, pp. 1453-61, 2006.

[7] A. Kummel, S. Panke, and M. Heinemann, "Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data," *Mol Syst Biol*, vol. 2, pp. 2006 0034, 2006.

[8] A. Varma and B. O. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110," *Appl Environ Microbiol*, vol. 60, pp. 3724-31, 1994.

[9] T. Shlomi, O. Berkman, and E. Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations," *Proc Natl Acad Sci U S A*, vol. 102, pp. 7695-700, 2005.

[10] D. Segre, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proc Natl Acad Sci U S A*, vol. 99, pp. 15112-7, 2002.

[11] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, "In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data," *Nat Biotechnol*, vol. 19, pp. 125-30, 2001.

[12] T. Shlomi, M. N. Cabili, M. J. Herrgard, B. O. Palsson, and E. Ruppin, "Network-based prediction of human tissue-specific metabolism," *Nat Biotechnol*, vol. 26, pp. 1003-10, 2008.

[13] A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas, "Flux coupling analysis of genome-scale metabolic network reconstructions," *Genome Res*, vol. 14, pp. 301-12, 2004.

[14] J. S. Edwards and B. O. Palsson, "Robustness analysis of the Escherichia coli metabolic network," *Biotechnol Prog*, vol. 16, pp. 927-39, 2000.

[15] J. S. Edwards, R. Ramakrishna, and B. O. Palsson, "Characterizing the metabolic phenotype: a phenotype phase plane analysis," *Biotechnol Bioeng*, vol. 77, pp. 27-36, 2002.

[16] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin, "A genome-scale computational study of the interplay between transcriptional regulation and metabolism," *Mol Syst Biol*, vol. 3, pp. 101, 2007.

[17] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson, "Integrating high-throughput and computational data elucidates bacterial networks," *Nature*, vol. 429, pp. 92-6, 2004.

[18] M. W. Covert and B. O. Palsson, "Transcriptional regulation in constraints-based metabolic models of Escherichia coli," *J Biol Chem*, vol. 277, pp. 28058-64, 2002.

[19] M. W. Covert, C. H. Schilling, and B. Palsson, "Regulation of gene expression in flux balance models of metabolism," *J Theor Biol*, vol. 213, pp. 73-88, 2001.

[20] J. L. Reed, T. R. Patel, K. H. Chen, A. R. Joyce, M. K. Applebee, C. D. Herring, O. T. Bui, E. M. Knight, S. S. Fong, and B. O. Palsson, "Systems approach to refining genome annotation," *Proc Natl Acad Sci U S A*, vol. 103, pp. 17480-4, 2006.

[21] M. J. Herrgard, S. S. Fong, and B. O. Palsson, "Identification of genome-scale metabolic network models using experimentally measured flux profiles," *PLoS Comput Biol*, vol. 2, pp. e72, 2006.

[22] M. J. Herrgard, B. S. Lee, V. Portnoy, and B. O. Palsson, "Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae," *Genome Res*, vol. 16, pp. 627-35, 2006.

[23] V. S. Kumar and C. D. Maranas, "GrowMatch: an automated method for reconciling in silico/in vivo growth predictions," *PLoS Comput Biol*, vol. 5, pp. e1000308, 2009.

[24] R. Schuetz, L. Kuepfer, and U. Sauer, "Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli," *Mol Syst Biol*, vol. 3, pp. 119, 2007.

[25] A. M. Feist, M. J. Herrgard, I. Thiele, J. L. Reed, and B. O. Palsson, "Reconstruction of biochemical networks in microorganisms," *Nat Rev Microbiol*, vol. 7, pp. 129-43, 2009.

[26] A. M. Feist and B. O. Palsson, "The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli," *Nat Biotechnol*, vol. 26, pp. 659-67, 2008.

[27] J. L. Reed, I. Famili, I. Thiele, and B. O. Palsson, "Towards multidimensional genome annotation," *Nat Rev Genet*, vol. 7, pp. 130-41, 2006.

[28] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson, "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information," *Mol Syst Biol*, vol. 3, pp. 121, 2007.

[29] A. R. Joyce, J. L. Reed, A. White, R. Edwards, A. Osterman, T. Baba, H. Mori, S. A. Lesely, B. O. Palsson, and S. Agarwalla, "Experimental and computational assessment of conditionally essential genes in Escherichia coli," *J Bacteriol*, vol. 188, pp. 8259-71, 2006.

[30] A. Raghunathan, J. Reed, S. Shin, B. Palsson, and S. Daefler, "Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction," *BMC Syst Biol*, vol. 3, pp. 38, 2009.

[31] H. Alper, Y. S. Jin, J. F. Moxley, and G. Stephanopoulos, "Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli," *Metab Eng*, vol. 7, pp. 155-64, 2005.

[32] S. S. Fong, A. P. Burgard, C. D. Herring, E. M. Knight, F. R. Blattner, C. D. Maranas, and B. O. Palsson, "In silico design and adaptive evolution of Escherichia coli for production of lactic acid," *Biotechnol Bioeng*, vol. 91, pp. 643-8, 2005.

[33] J. H. Park, K. H. Lee, T. Y. Kim, and S. Y. Lee, "Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation," *Proc Natl Acad Sci U S A*, vol. 104, pp. 7797-802, 2007.