Integrated Regulatory and Metabolic Models

Chapter 10

Markus W. Covert
California Institute of Technology, Pasadena, California, USA

ABSTRACT

This chapter describes how to reconstruct functional metabolic and transcriptional regulatory networks, as well as the modeling approaches that allow for simulation of network behavior for networks separately and for networks combined. This process is placed in the context of model-driven biological discovery, and is illustrated with a detailed case study. In this study, a genome-scale model was reconstructed and used in conjunction with experimental data to elucidate the regulatory and metabolic networks in *Escherichia coli*.

I. INTRODUCTION

A major goal of systems biology is to further our understanding of complex biological systems. Using systems biology to facilitate biological discovery may be thought of as a simple expansion of traditional biology, as shown in Figure 10.1. Traditional biology (shaded box) begins with an experimental system of interest. The "inputs" to the system are simply aspects of the system that can be controlled. Thus, the inputs may be external (such as environmental conditions) or internal, such as perturbations to the genetic makeup of the organism (gene knock-outs or knock-ins). The "outputs" to the system are aspects that are changed by the system itself and that are measurable. Outputs can also be external (such as the concentration over time of secreted by-products or biomass) or internal, such as the differential expression of genes or activity of regulatory proteins. If the experiment is

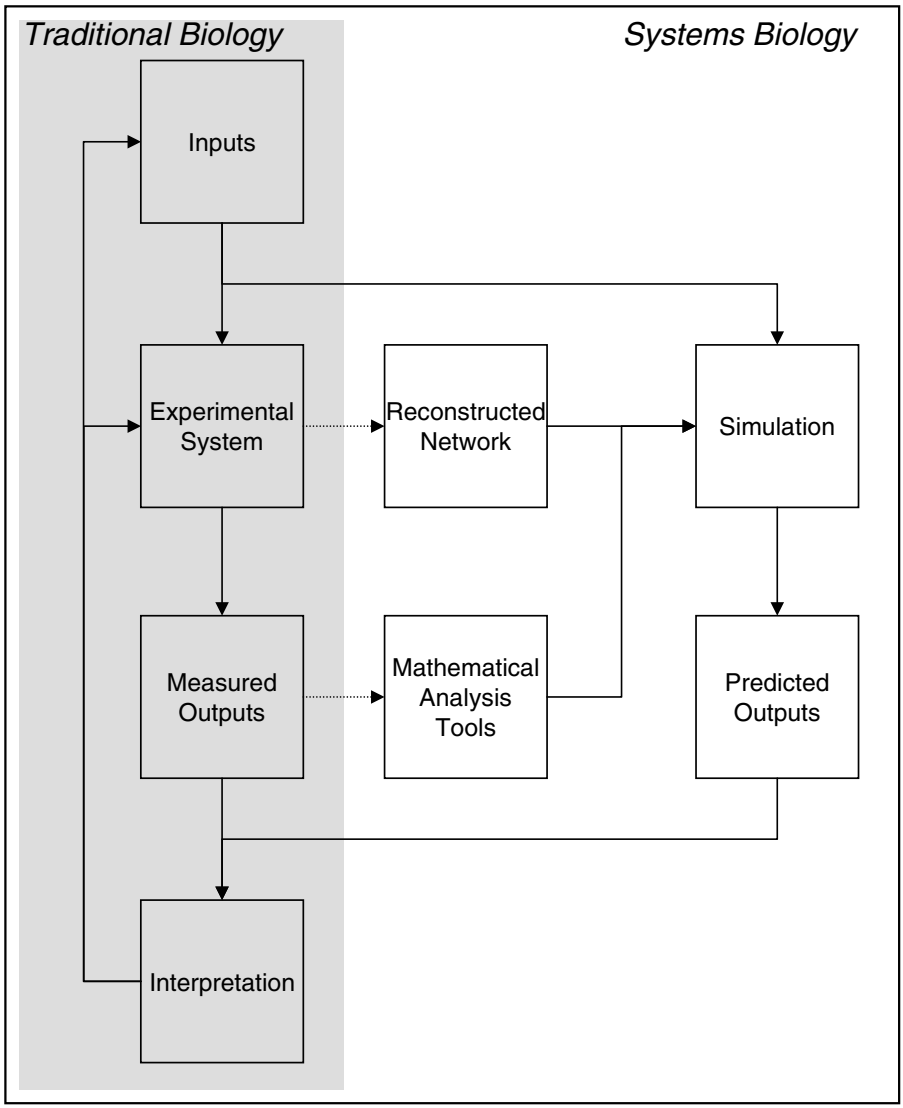

Figure 10.1. The traditional and systems biology approaches to discovery. The systems approach can greatly expedite the discovery process by incorporating the testable predictions of a mathematical model.

well-designed, the investigator can make an interpretation of the measured outputs that (1) gives new insight into the system and (2) suggests new perturbations to the inputs for a subsequent experiment.

This process has had great success over the last several decades and is the foundation for all of the biological knowledge we have. However, it now has the poten-

tial to be greatly enhanced by two major factors. First, the development of high-throughput technologies means that we are now able to vary the inputs and measure outputs many thousandfold faster than before (although arguably with different accuracy). The result is a combinatorial explosion of data that would be impossible to interpret without the aid of a computer. The development of mathematical modeling tools is the second factor, enabling a much more rapid characterization of biological systems.

How do these two factors influence traditional biology? First, experimental systems can be studied more broadly, with much the same detail. Instead of looking at one small part of the organism, we can consider an entire network. Metabolism and transcriptional regulation are currently the networks most feasible, but there is every reason to believe that others (signal transduction networks, for example) will follow. The annotated genome sequence enables us to obtain most of the components of the network, although a substantial minority of components must still be obtained from the traditional biology literature. The type of measured output should drive the choice of mathematical analysis tools, as the predictions made by a mathematical model are of much greater use if they can be directly compared to experimental data. By analyzing the network with the appropriate mathematical tools, it is possible to run simulations that predict outputs given a set of inputs.

In sum, once the inputs have been determined they are applied to the experimental system as well as to the mathematical representation. The predicted and measured outputs are obtained and compared. The reconciliation of experimental and computational results, which may also be automated, is in actuality interpretation of the experimental data on a grand scale. It can lead to the identification of many new components and interactions in the system at once.

The incorporation of these elements (high-throughput technology and mathematical modeling) with the traditional biology process is one definition of systems biology (Cowley 2004). The purpose of this chapter is to show how this integrative approach can be applied to metabolic and transcriptional regulatory networks.

II. METABOLIC NETWORKS

For several reasons, the state of metabolic network reconstruction and modeling is the more advanced. Much of the required information for network reconstruction can be obtained from the annotated genome sequence and enzyme-to-reaction databases, and several organisms are quite well characterized biochemically. Because there is a wealth of literature on this topic, we describe it fairly generally and refer to the reviews for more detail (Covert et al. 2001; Price et al. 2004; Gagneur and Casari G 2005; Patil et al. 2004).

A. Network reconstruction

Metabolic network reconstruction begins by compiling a list of all enzymes and transport proteins identified in the annotated genome sequence of an organism,

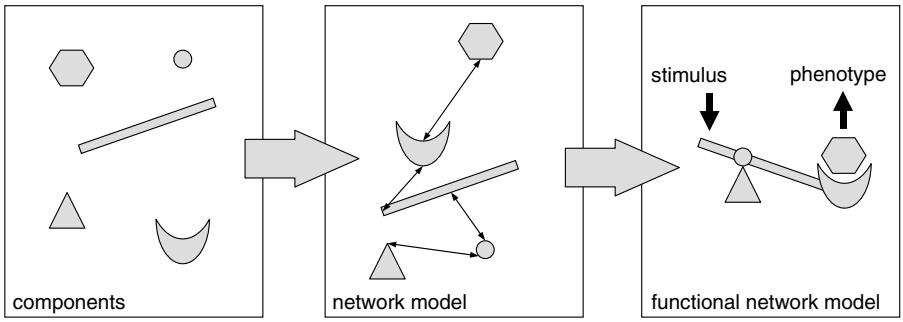

Figure 10.2. Functional network reconstructions. Metabolic and regulatory networks may be reconstructed in terms of component lists or graphs indicating some interactions, but are most useful when integrated in such a way that they actually predict behaviors that can be compared with experimental observations.

as found in a database such as the Comprehensive Microbial Resource (Peterson et al. 2001) or MetaCyc (Krieger et al. 2004). Each protein or protein complex is associated with one or more metabolic reactions or transport processes, using a database such as the ENZYME databank (Bairoch 1994). Missing pieces of the network may also be found in the biochemical literature or identified by comparison with known pathways in other organisms (Overbeek et al. 2000) to obtain a relatively complete reconstruction.

Although such reconstructions can be very useful for some types of network analysis (Jeong et al. 2000; Ma et al. 2004), for applications such as phenotype simulation a more complete network is required. Put simply, the metabolic network must be "functional"; that is, given a set of known and well-characterized behaviors of the organism the reconstruction must contain all proteins necessary to simulate these behaviors (Figure 10.2). For a vegetative cell, the network must allow production or transport of biomass components (e.g., all essential amino acids, nucleotides) given a defined medium, and must be able to take up known substrates and produce known secreted metabolites. The initial reconstruction of a functional metabolic network therefore requires a thorough integration of genomic, biochemical, and phenotypic data.

What is in a reconstruction? This depends on several factors, most importantly the type of analysis to be performed on the network. For a simple graph network analysis, all that is required is a set of nodes (e.g., metabolites) and the interactions between the nodes (e.g., reactions). To enable a metabolic flux analysis, it is also necessary to include the stoichiometry of the reactions as well as some flux information, such as a maximum oxygen or substrate uptake rate. Flux balance analysis, described in more detail later in the chapter, also requires definition of the organism's biomass composition, in terms of how many moles of all amino acids, nucleotides, and so on are contained in one gram dry weight of the organism. For a complete kinetic description, all of the kinetic parameters would need to be

included. However, the parameters would be extremely difficult to obtain (Bailey 2001). Recently, some of the most detailed metabolic network reconstructions have been updated to include complete charge and elemental balancing, in addition to stoichiometry, biomass composition, and some maximum uptake and secretion flux rates (Reed et al. 2003; Duarte et al. 2004).

B. Analysis and simulation

Once the network is reconstructed it may be analyzed, depending on the detail of the reconstruction (as described previously). Because other chapters will discuss graph-based and detailed kinetic modeling approaches, I will focus on the analysis methods currently applicable to large-scale functional networks, under the umbrella term *constraint-based modeling*.

Constraint-based modeling itself has been reviewed thoroughly and frequently over the past several years (Covert et al. 2003; Price et al. 2004). In brief, because of the difficulty of obtaining a complete detailed description of all reaction fluxes in the metabolic networkconstraint-based analysis instead focuses on limiting the ranges these flux values can have, given a set of constraints. These constraints generally include those associated with mass balance and the stoichiometry of biochemical reactions, as well as reaction reversibility and certain maximum flux rates. More recently, the constraints of energy balance have also been added (Beard et al. 2002).

In practical terms, constraint-based analysis begins with mass-balance equations for each metabolite, as shown in Equation 10.1.

$$\frac{dX}{dt} = \sum v_{\text{syn}} - \sum v_{\text{deg}} + \sum v_{\text{trans}} \tag{10.1}$$

Here, X is the metabolite concentration, and v represents reaction fluxes that synthesize (*syn*), degrade (*deg*), or transport (*trans*) metabolites into and out of the system. It is often assumed that the system is at a quasi-steady state with respect to metabolism (i.e., $dX/dt = 0$, described in more detail in material following). Incorporating this assumption and combining all of the mass balance equations yields Equation 10.2.

$$\mathbf{Sv} = \mathbf{0} \tag{10.2}$$

Here, $\mathbf{S}$ is the stoichiometric matrix for the system and $\mathbf{v}$ is a vector of all fluxes in the system. Other constraints—such as the reversibility of metabolic reactions (e.g., $v_i \geq 0$), as well as maximum enzyme/transport capacity of proteins (e.g., $v_i \leq v_{max}$)—are incorporated when known.

Once these constraints have been defined, the overall capabilities of the metabolic network may be determined using extreme pathway analysis or elementary mode analysis (Papin et al. 2004), and flux distributions that optimize network production of cellular biomass components may be determined using flux-balance analysis (Price et al. 2004). Recent years have been extremely fruitful in terms of

developing creative and insightful analysis techniques for studying metabolic networks (Price et al. 2004).

It is often assumed that such analyses are limited because one assumption crucial to all of the approaches discussed here is that the metabolic network is at a steady-state. However, as the time constants relevant to metabolic reactions are on the order of milliseconds (McAdams et al. 1998), behavior of the network may be simulated dynamically. The simulation is simply broken into several time steps just large enough that the metabolic network may be assumed to be at a quasi-steady state, and differential equations are solved to calculate the growth, uptake, and secretion of various metabolites over time (Varma et al. 1994). Such an approach is important when incorporating the highly dynamic behavior of the transcriptional regulatory network.

C. Predicted and measured outputs

Because the utility of a mathematical model depends on how directly model predictions may be compared to experimental data, it is useful to describe the experimental techniques used to study metabolism. Of particular importance are the measured outputs from such techniques, and whether they can be compared to predicted outputs. The measurable outputs for metabolic networks are growth rate, concentrations of external compounds over time, and internal metabolic fluxes. For metabolic networks, we can now assess growth rate under various environmental conditions on 96-well plates using phenotype microarrays (Bochner et al. 2001). Substrate uptake and product secretion rates can be measured using standard chromatography techniques, and high-throughput metabolomic technologies are being developed (Kell 2004). The uptake and metabolism of radiolabeled substrates may also be used to calculate internal metabolic fluxes indirectly (Sauer 2004). Current metabolic network reconstructions allow for direct comparison with all of these data using flux-balance analysis (as described previously).

III. REGULATORY NETWORKS

A. Network reconstruction

Regulatory networks differ from metabolic networks in ways that impact the network reconstruction as well as modeling approaches (Herrgard et al. 2004). First, the components are different. Whereas metabolic networks involve metabolites, enzymes, and transport proteins, regulatory networks involve regulatory proteins and the promoter regions of target genes. Second, most of the metabolic proteins are well conserved across species. Regulatory proteins may also be conserved. However, the *cis* regulatory regions are generally not conserved across species, and transcription factor binding sites are extremely difficult to find in promoter regions due to their short length, although progress is being made (Beer et al. 2004). In addi-

tion, the interactions of transcription factors at one promoter region can be extremely complex (Davidson et al. 2002), and even a single nucleotide difference in a transcription factor binding site can change the specificity of cofactor binding (Leung et al. 2004).

Accordingly, the level of characterization of regulatory networks does not approach that found in metabolic networks. Currently, detailed genome-scale regulatory networks have been reconstructed only for *Saccharomyces cerevisiae* (Lee et al. 2002; Harbison et al. 2004) and *E. coli* (Shen-Orr et al. 2002; Salgado et al. 2004). These reconstructions are qualitative, including the effect of active transcription factors on target genes (whether the factor acts as an inducer, repressor, or both). More detailed reconstructions, which would include some of the dynamics of gene expression, are extremely useful but also far more difficult to obtain (Kalir et al. 2004).

Notwithstanding these challenges to those wishing to study regulation, two high-throughput technologies have made it possible to reconstruct regulatory networks at the large scale. First, microarray analysis enables the determination of the expression profile of an entire genome in one experiment (Gardner et al. 2003). Second, it is now possible to determine with some accuracy where all of the transcription factors are binding in the genome under a given set of experimental conditions (Lee et al. 2002). These two approaches, especially when used in combination with each other or with the existing literature, are a powerful way of characterizing a regulatory network (Hartemink et al. 2002; Herrgard et al. 2003).

B. Analysis and simulation

Regulatory network modeling approaches are significantly different from metabolic network modeling approaches (McAdams et al. 1998; de Jong 2002; Tyson et al. 2003; Herrgard et al. 2004). They include Boolean logic (Thomas 1973), fuzzy logic (Lee et al. 1999), Bayesian models (Hartemink et al. 2002), kinetic models (Kremling et al. 2001; Kalir et al. 2004), and stochastic models (McAdams et al. 1998). In general, the greater the level of detail required by the modeling approach (in terms of the number of parameters) the less complex the network studied, down to the extreme simplicity of engineered regulatory networks (Hasty et al. 2002). On the other hand, the detailed models of small engineered systems have been instrumental in developing our understanding of the effect of noise on network dynamics (Elowitz et al. 2000).

For large-scale modeling, an approach that is qualitative is most advantageous, because of the qualitative nature of the existing literature (Bolouri et al. 2002). The presence of relevant stimuli, activity of regulatory proteins, and expression of target genes can all be described in terms of Boolean logic. This framework was demonstrated to be particularly useful for integrating regulatory and metabolic models, wherein the effects of regulatory events are represented as time-dependent constraints on the metabolic network (Covert et al. 2001).

C. Measured and predicted outputs

The typical outputs of comparing expression of a gene under two conditions using microarrays or quantitative real-time RT-PCR are a *p*-value (derived from appropriate statistical analysis of repeated expression measurements) indicating the probability that a change in expression occurred, and a ratio of expression levels or signal intensities, which assigns a quantitative magnitude of the expression shift. For comparison with genome-wide qualitative gene expression changes, a regulatory network need therefore only be expressed in terms of logical rules. More detailed models also allow comparison with the ratio data for a limited number of genes (Kremling et al. 2001; Kalir et al. 2004).

IV. EXPERIMENTAL AND COMPUTATIONAL DATA INTERPRETATION

Although there are many reasons to build models (Bailey 1998), one of current importance is to elucidate the biology of the modeled network. Specifically, models can be used to identify or indicate the presence of previously unknown components or interactions in the network. This occurs through integration and reconciliation of measured and computationally predicted experimental outcomes. The remainder of this chapter focuses in depth on the use of a combined regulatory-metabolic model in *E. coli*, which was used in coordination with high-throughput experimental studies to facilitate elucidation of the metabolic and regulatory networks (Covert et al. 2004) (Figure 10.3).

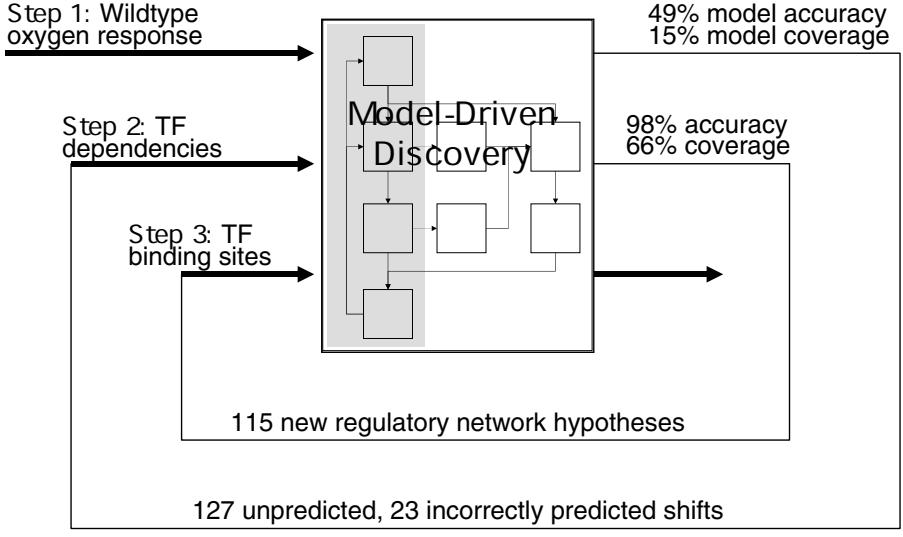

Figure 10.3. Model-driven discovery as applied to *E. coli*. High-throughput experiments and a mathematical model were integrated using the approaches described in this chapter to determine many previously unknown interactions in the transcriptional regulatory network.

The *E. coli* model accounts for the products of 1,010 genes, or roughly one-third of the annotated genes in this organism. It contains 104 regulatory proteins, which regulate the transcription of 479 metabolic genes. There are 906 genes that constitute the metabolic component of the model. The metabolic network is described and simulated using flux-balance analysis, and the regulatory model uses logic statements to describe regulatory events. The two networks interface via the constraint-based framework: regulatory events are interpreted simply by imposing time-dependent constraints on the metabolic network. Such an approach had previously been shown to result in more accurate predictive capability as well as broader scope of prediction (Covert et al. 2002).

The model is able to generate predictions of the following outputs: growth rate, substrate uptake rates, by-product secretion rates, medium concentrations of biomass and metabolites over time, internal flux rates, and shifts in gene expression. In addition, it can predict the effects of internal (i.e., deletion of one or more metabolic or regulatory genes) and external (i.e., change in medium composition, availability of oxygen, and so on) perturbations on the behavior of the system. Experimental data corresponding to all of the predictions listed here can also be obtained with relative ease using standard methods in microbial physiology and gene expression profiling.

Model predictions were compared to two large data sets for the purpose of network elucidation. The first was a large set of phenotype data available from the ASAP database (Glasner et al. 2003). Cells were seeded onto 96-well plates, with each well containing a medium designed to test one feature of microbial metabolic capability (e.g., the ability to utilize glucose as a sole carbon source) and allowed to grow overnight, upon which respiration of the cells was compared to a negative control as an indicator of growth (Bochner et al. 2001). The data that could be compared to model predictions included 110 different growth environments and 125 knockout strains of *E. coli* for a total of 13,750 outcomes.

The predicted and measured outcomes agreed in most (approximately 80%) of the cases. More interestingly, the model failures corresponded to particular environments or strains. Closer examination of the failures led to new hypotheses about *E. coli* metabolism and regulation. In all, comparison of prediction and experiment for 10 environmental conditions and eight knockout strains led to new hypotheses about regulatory interactions or uncharacterized enzymes and metabolic pathways.

As an example, one of the environmental tests was the ability of the cells to grow using thymidine as a sole carbon source. The model predicted that such growth was impossible. However, the measured data showed that each of the knockout strains was able to grow. One possible reason for the model failure is that the reconstructed metabolic network lacks a thymine-reductive pathway (including enzymes with the following E.C. numbers: 1.3.1.2 or 1.3.1.1, 3.5.2.2, and 3.5.1.6). As including this pathway would reconcile the model predictions and measured observations, one can find the most likely open reading frames to encode the pathway using sequence and phylogeny comparison tools such as MAST and MEME (Reed et al. 2003; Covert et al. 2004). In this case, the most likely open reading frames

(ORFs) for the thymine-reductive pathway enzymes are b2106 for 1.3.1.2 and b2873 or b0512 for 3.5.2.2. Such hypotheses have been verified in past metabolic network studies (Covert et al. 2001).

The second set of data was a collection of gene expression profiles generated as part of the study. Based on an earlier study (Herrgard 2003) the aerobic-anaerobic shift was targeted as a portion of the network with an intermediate level of characterization. The gene expression profile was obtained for *E. coli* during exponential growth on M9 glucose minimal medium under aerobic and anaerobic conditions. The model was used to predict the differential gene expression between the profiles, as well as growth rates and the like. In this case, the comparison between model predictions and experimental outcomes involves two measures: the accuracy (where a shift was predicted, it was also observed) and coverage (where a shift was observed, it was also predicted) of model predictions. For the first version of the model, the accuracy was about 49% and the coverage was only about 15%. These measurements indicate first that the regulatory network is much less characterized than the metabolic network, and second that the aerobic-anaerobic part of the network in particular requires more scrutiny to be fully understood.

The discrepancies between experiment and model were examined in more detail by determining the transcription factor dependencies of the differential expression observed in the wild type. This was accomplished via a perturbation analysis (Ideker et al. 2001) (Figure 10.4). Strains in regulatory proteins involved in the molecular response to oxygen were constructed, and their gene expression profiles under conditions identical to the wild type were determined. Using analysis of variance enabled determination of whether a shift in expression observed in the wild type was abolished in the knockout strain. This led directly to description of a logical rule.

For example, the *kgtP* gene (b2587) was listed without a regulatory rule in the original model. However, the microarray data indicated a significant shift with a log2 ratio of 2.05 between the aerobic and anaerobic conditions. The perturbation studies indicated that the differential expression observed in the wild type was abolished in the $\triangle$*arcA* and the $\triangle$*arcA* $\triangle$*fnr* knockout strains. As a result, the rule was rewritten as *kgtP* = IF NOT (ArcA) (ArcA, Fnr, and NarL are regulatory proteins that also have rules that dictate their activity). For the *fdnI* gene (b1476), a rule already existed: *fdnI* = Fnr OR NarL. However, no differential expression was observed. The rule became *fdnI* = NarL. In several cases, the only change made to resolve the model predictions and observations were in the interactions between regulatory proteins (e.g., changing an *AND* to an *OR*, and vice versa). This is an important observation, as the regulatory effects of most regulatory proteins to date have been tested singly and not in combination.

This analysis led to a greatly improved network model. The second-version *E. coli* model predicted 67% of the 151 observed expression shifts (coverage), with a predictive accuracy of 98%. More importantly, reconciliation of the model and the data led to many new hypotheses about the regulatory network in *E. coli* that are readily testable. Finally, the new model was compared to the phenotype microarray study

Bnum	Gene	L2R	Ar	Ap	F	O	S	A/F	Rule Addition
b0033	*carB*	0.62				X			OxyR
b0034	*caiF*	-1.37						X	ArcA and Fnr
b0068	*sfuA*	0.33							Oxygen
b0113	*pdhR*	0.35	X		X			X	Not (ArcA and Fnr)
b0114	*aceE*	0.48	X		X			X	Not (ArcA and Fnr)
b0115	*aceF*	0.48	X		X			X	Not (ArcA and Fnr)
b0116	*lpdA*	1.32	X		X			X	Not (ArcA and Fnr)
b0118	*acnB*	2.63	X					X	Not (ArcA)
b0313	*betI*	1.98	X					X	Not (ArcA)
b0336	*codB*	0.43				X			OxyR
b0401	*brnQ*	-0.65							Not (Oxygen)
b0564	*appY*	-1.87			X				Not (ArcA) and Fnr
b0653	*gltK*	0.73	X		X			X	Not (ArcA and Fnr)
b0683	*fur*	0.26						X	Not (ArcA or Fnr)
b0721	*sdhC*	4.70	X		X			X	Not (ArcA and Fnr)
b0722	*sdhD*	4.63	X		X			X	Not (ArcA and Fnr)
b0723	*sdhA*	3.01	X		X			X	Not (ArcA and Fnr)
b0726	*sucA*	2.17	X					X	Not (ArcA)
b0727	*sucB*	2.07	X					X	Not (ArcA)
b0733	*cydA*	-0.79							Not (Oxygen)
b0734	*cydB*	-0.66							Not (Oxygen)
b0755	*gpmA*	0.84	X		X			X	Not (ArcA and Fnr)
b0776	*bioF*	0.48							Oxygen
b0778	*bioD*	0.43							Oxygen
b0854	*potF*	0.83	X		X			X	Not (ArcA and Fnr)
b0864	*artP*	-0.57							Not (Oxygen)
b0993	*torS*	-0.97							Not (Oxygen)
b1033	*ycdW*	0.42						X	Not (ArcA or Fnr)
b1221	*narL*	0.56						X	Not (ArcA or Fnr)
b1241	*adhE*	-1.44							Not (Oxygen)
b1323	*tyrR*	-0.62							Not (Oxygen)
b1531	*marA*	0.90	X		X	X		X	Not (ArcA and Fnr) or OxyR
b1656	*sodB*	-0.20							Not (Oxygen)
b1676	*pykF*	-0.47							Not (Oxygen)
b1702	*pps*	0.68							Oxygen
b1779	*gapA*	-0.18							Not (Oxygen)
b1827	*kdgR*	-0.47						X	ArcA and Fnr
b1991	*cobT*	-0.27			X			X	Fnr
b1993	*cobU*	-0.17			X			X	Fnr
b2040	*rfbD*	0.16							Oxygen
b2129	*yehX*	0.35	X		X			X	Not (ArcA and Fnr)
b2296	*ackA*	-1.49						X	ArcA and Fnr
b2308	*hisQ*	0.26	X		X			X	Not (ArcA and Fnr)
b2309	*hisJ*	0.44	X		X	X		X	Not (ArcA and Fnr) or OxyR
b2344	*fadL*	0.98	X					X	Not (ArcA)

Bnum	Gene	L2R	Ar	Ap	F	O	S	A/F	Rule Addition
b2463	*maeB*	0.92							Oxygen
b2530	*iscS*	0.80							Oxygen
b2573	*rpoE*	-0.62				X			Not (OxyR)
b2587	*kgtP*	2.05	X					X	Not (ArcA)
b2676	*nrdF*	0.80							Oxygen
b2677	*proV*	1.72						X	Not (ArcA or Fnr)
b2707	*srlR*	-0.36			X			X	Fnr
b2747	*ispD*	0.28							Oxygen
b2904	*gcvH*	0.77						X	Not (ArcA or Fnr)
b2905	*gcvT*	1.38							Oxygen
b2925	*fbaA*	-0.45							Not (Oxygen)
b3089	*sstT*	-1.86							Not (Oxygen)
b3423	*glpR*	0.13							Oxygen
b3453	*ugpB*	0.74						X	Not (ArcA or Fnr)
b3612	*yibO*	-0.90							Not (Oxygen)
b3767	*ilvG_1*	-0.36				X			Not (OxyR)
b3769	*ilvM*	-0.53			X			X	Fnr
b3805	*hemC*	-0.19							Not (Oxygen)
b3806	*cyaA*	-0.54			X			X	Fnr
b3916	*pfkA*	-1.06							Not (Oxygen)
b3917	*sbp*	0.35							Oxygen
b3962	*sthA*	1.74	X					X	Not (ArcA)
b3990	*thiH*	0.69							Oxygen
b3993	*thiE*	0.51						X	Not (ArcA or Fnr)
b3994	*thiC*	0.69						X	Not (ArcA or Fnr)
b4014	*aceB*	0.77	X					X	Not (ArcA)
b4015	*aceA*	1.33	X					X	Not (ArcA)
b4151	*frdD*	-2.23							Not (Oxygen)
b4152	*frdC*	-0.98							Not (Oxygen)
b4153	*frdB*	-2.31							Not (Oxygen)
b4154	*frdA*	-0.80							Not (Oxygen)
b4232	*fbp*	0.46							Oxygen
b4322	*uxuA*	0.40							Oxygen

Legend

- L2R > +1.0
- 1 > L2R > 0.5
- +0.5 > L2R > -0.5
- -0.5 > L2R > -1.0
- -1 > L2R

Figure 10.4. Determining new regulatory rules using the perturbation approach. A list of genes for which the computational model failed to predict observed differential expression (false negatives). The observed aerobic-anaerobic log2 ratio for the wild-type cells (L2R) is shown numerically and color coded, as explained in the legend. The observed wild-type differential expression was abolished in certain transcription factor knockout strains (Ar = △*arcA*, Ap = △*appY*, F = △*fnr*, O = △*oxyR*, S = △*soxS*, A/F = △*arcA* △*fnr*), as indicated by an X. These transcription factor dependencies were used to determine new regulatory rules, as shown. Note that certain transcription factors, such as OxyR, are generally active in the presence of oxygen, whereas others (such as ArcA and Fnr) are active in the absence of oxygen.

described previously, with slight improvement to the predictive capabilities there, and is therefore completely consistent with regard to all of the other available data.

V. CONCLUSIONS

This chapter shows how model-building fits in the context of experimental discovery in terms of metabolism and transcriptional regulation, using a model of *E. coli* as an example. How well this approach can be more broadly applied to organisms and processes more complex and much less understood remains to be seen.

Protein chips to measure outputs of cell signaling processes (Hall et al. 2004) and methods for simulating signaling networks at the large scale (Papin et al. 2005) are also being developed. It can be expected, however, that the success of such efforts will depend on the ability of models to generate predictions that can be directly compared to experimental measurements at a large scale. As can be seen from this case study, such models will have the potential to greatly facilitate biological discovery.

ACKNOWLEDGMENTS

The author is a Robert Black Fellow supported by the Damon Runyon Cancer Research Foundation (DR6-#1835-04), and would also like to thank Markus Herrgard for critical reading of the manuscript.

RECOMMENDED RESOURCES

Davidson, E. (2001). *Genomic Regulatory Systems*. San Diego: Academic Press.

Ptashne, M., and Gann, A. (2001). *Genes and Signals*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Stephanopoulos, G., Aristidou, A., and Nielsen, J. (1998). *Metabolic Engineering*. San Diego: Academic Press.

REFERENCES

Bailey, J. E. (1998). Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. *Biotechnol. Prog.* **14**:8–20.

Bailey, J. E. (2001). Complex biology with no parameters. *Nat. Biotechnol.* **19**:503–504.

Bairoch, A. (1994). The ENZYME data bank. *Nucleic. Acids. Res.* **22**:3626–3627.

Beard, D. A., Liang, S. D., and Qian, H. (2002). Energy balance for analysis of complex metabolic networks. *Biophys. J.* **83**:79–86.

Beer, M. A., and Tavazoie, S. (2004). Predicting gene expression from sequence. *Cell* **117**:185–198.

Bochner, B. R., Gadzinski, P., and Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res.* **11**:1246–1255.

Bolouri, H., and Davidson, E. H. (2002). Modeling transcriptional regulatory networks. *Bioessays* **24**:1118–1129.

Covert, M. W., and Palsson, B. O. (2002). Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *J. Biol. Chem.* **277**:28058–28064.

Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I.I., Selkov, E., and Palsson, B. O. (2001). Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.* **26**:179–186.

Covert, M. W., Schilling, C. H., and Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. *J. Theo. Biol.* **213**:73–88.

Covert, M. W., Famili, I., and Palsson, B. O. (2003). Identifying constraints that govern cell behavior: A key to converting conceptual to computational models in biology? *Biotechnol. Bioeng.* **84**:763–772.
Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., and Palsson, B. O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**:92–96.
Cowley, A. W. Jr. (2004). The elusive field of systems biology. *Physiol. Genomics* **16**:285–286.
Davidson, E. H., Rast, J. P., Oliveri, P., Ransick, A., et al. (2002). A genomic regulatory network for development. *Science* **295**:1669–1678.
de Jong, H. (2002). Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* **9**:67–103.
Duarte, N. C., Herrgard, M. J., and Palsson, B. O. (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750: A fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**:1298–12309.
Elowitz, M. B., and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* **403**:335–338.
Gagneur, J., and Casari, G. (2005). From molecular networks to qualitative cell behavior. FEBS Lett. **579**:1867–1871.
Gardner, T. S., di Bernardo, D., Lorenz, D., and Collins, J. J. (2003). Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**:102–105.
Glasner, J. D., Liss, P., Plunkett, G. III, et al. (2003). ASAP: A systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* **31**:147–151.
Hall, D. A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., and Snyder, M. (2004). Regulation of gene expression by a metabolic enzyme. *Science* **306**:482–484.
Harbison, C. T., Gordon, D. B., Lee, T. I., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**:99–104.
Hartemink, A. J., Gifford, D. K., Jaakkola, T. S., and Young, R. A. (2002). Combining location and expression data for principled discovery of genetic regulatory network models. *Pac. Symp. Biocomput*, 437–449.
Hasty, J., McMillen, D., and Collins, J. J. (2002). Engineered gene circuits. *Nature* **420**: 224–230.
Herrgard, M. J., Covert, M. W., and Palsson, B. O. (2003). Reconciling gene expression data with known genome-scale regulatory network structures. *Genome. Res.* **13**:2423–2434.
Herrgard, M. J., Covert, M. W., and Palsson, B. O. (2004). Reconstruction of microbial transcriptional regulatory networks. *Curr. Opin. Biotechnol.* **15**:70–77.
Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**:929–934.
Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabasi, A. L. (2000). The large-scale organization of metabolic networks. *Nature* **407**:651–654.
Kalir, S., and Alon, U. (2004). Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell* **117**:713–720.
Kell, D. B. (2004). Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* **7**:296–307.
Kremling, A., Bettenbrock, K., Laube, B., Jahreis, K., Lengeler, J. W., and Gilles, E. D. (2001). The organization of metabolic reaction networks: Application for diauxic growth on glucose and lactose. *Metab. Eng.* **3**:362–379.

Krieger, C. J., Zhang, P., Mueller, L. A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S. Y., and Karp, P. D. (2004). MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **32**:D438–D442.

Lee, B., Yen, J., Yang, L., and Liao, J. C. (1999). Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. *Biotechnol. Bioeng.* **62**:722–729.

Lee, T. I., Rinaldi, N. J., Robert, F., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**:799–804.

Leung, T. H., Hoffmann, A., and Baltimore, D. (2004). One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. *Cell* **118**:453–464.

Ma, H. W., Zhao, X. M., Yuan, Y. J., and Zeng, A. P. (2004). Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics* **20**:1870–1876.

McAdams, H. H., and Arkin, A. (1998). Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* **27**:199–224.

Overbeek, R., Larsen, N., Pusch, G. D., D'Souza, M., Selkov, E. Jr., Kyrpides, N., Fonstein, M., Maltsev, N., and Selkov, E. (2000). WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* **28**:123–125.

Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., and Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends Biotechnol.* **22**:400–405.

Patil, K. R., Akesson, M., and Nielsen, J. (2004). Use of genome-scale microbial models for metabolic engineering. Curr Opin Biotechnol. **15**:64–69.

Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., and White, O. (2001). The comprehensive microbial resource. *Nucleic Acids Res.* **29**:123–125.

Papin, J. A., Hunter, T., Palsson, B. O., and Subramaniam, S. (2005). Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* **6**:99–111.

Price, N. D., Reed, J. L., and Palsson, B. O. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**:886–897.

Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. O. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**:R54.

Salgado, H., Gama-Castro, S., Martinez-Antonio, A., et al. (2004). RegulonDB (version 4.0): Transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. *Nucleic Acids Res.* **32**:D303–D306.

Sauer, U. (2004). High-throughput phenomics: Experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.* **15**:58–63.

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31**:64–68.

Thomas, R. (1973). Boolean formalization of genetic control circuits. *J. Theo. Biol.* **42**:563–585.

Tyson, J. J., Chen, K. C., and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell. Biol.* **15**:221–231.

Varma, A., and Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **60**:3724–3731.