Genome-scale microbial *in silico* models: the constraints-based approach

Nathan D. Price[1], Jason A. Papin[1], Christophe H. Schilling[2] and Bernhard O. Palsson[2]

[1]Department of Bioengineering, 9500 Gilman Drive, University of California, San Diego, La Jolla, CA 2093-0412, USA
[2]Genomatica, Inc., 5405 Morehouse Drive, Suite 210, San Diego, CA 92121, USA

Genome sequencing and annotation has enabled the reconstruction of genome-scale metabolic networks. The phenotypic functions that these networks allow for can be defined and studied using constraints-based models and *in silico* simulation. Several useful predictions have been obtained from such *in silico* models, including substrate preference, consequences of gene deletions, optimal growth patterns, outcomes of adaptive evolution and shifts in expression profiles. The success rate of these predictions is typically in the order of 70–90% depending on the organism studied and the type of prediction being made. These results are useful as a basis for iterative model building and for several practical applications.

The value of building mathematical models of cells and simulating their integrated behavior has long been recognized, and computer simulations of complex biological functions began essentially as soon as the computational capability became available [1–3]. Lack of appropriate experimental data and the complexity of living cells have historically hampered these efforts. The first whole-cell metabolic model was developed for the human red blood cell (RBC) [4] as the culmination of two decades of work. Continual model building of the RBC has since taken place [5–7] and versions can now be downloaded [8]. These RBC models are based on kinetic theory and are comprised of ordinary differential and associated algebraic equations. Interestingly, the RBC has emerged as a model system to examine various *in silico* analysis procedures [9,10]. To date, the success with the construction of a whole-cell RBC model has not been reproduced for other cell types although several efforts are emerging [11,12].

The sequencing of the first bacterial genome [13] signified a transition of biology from a data-poor to a data-rich environment. Since then, various 'omics' datasets are becoming available in ever increasing sizes [14–16]. This flood of biological data has underscored the need for systems analysis in biology and necessitated a change in mathematical modeling philosophy [17–19]. Modern biological model building thus needs to meet new sets of criteria: models need to be organism-specific, data-driven, easily scalable, capable of integrating various 'omics' data types and able to account for the inherent uncertainty in biological functions. *In silico* models will, therefore, be informatics intensive calling for the appropriate development of databases and algorithms, as well as the construction of robust quality controlled software processes for their implementation. Many modeling approaches are currently being used to model cellular processes, including kinetic [11,12], stochastic [20,21] and cybernetic approaches [22,23]. Although these methods provide useful results, it is currently difficult to use them to model genome-scale networks because of the large number of parameters needed and the computational complexity. To date, genome-scale models of metabolism have only been analyzed with the constraints-based modeling philosophy, and therefore we focus on this approach. The purpose of this review is to describe genome-scale microbial models that have been built, their uses, and to describe some of the current challenges in this field.

The constraints-based modeling approach

The challenges of genome-scale model building are being met, in part, by constraints-based models [19,24]. This modeling process involves a multi-step procedure (Fig. 1). The first step is to reconstruct the underlying network [25–29]. For metabolism, this is a well-established procedure [25], whereas methods for the reconstruction of the associated regulatory networks are being developed [30,31]. The second step involves the statement of the constraints under which the reconstructed network operates. Such constraints are based on enzyme capacity, reaction stoichiometry and thermodynamics associated with reaction directionality and biochemical loops [32,33]. The statement of constraints leads to the definition of a solution space in which the solution to the network equations must lie. This solution space contains all the possible functions of the reconstructed network or all the allowable phenotypes. For example, all possible steady state flux distributions through a metabolic network are bounded in a solution space formed by a set of unique basis pathways, called the 'extreme pathways [34].' All steady state flux distributions are similarly characterized by the elementary modes [35]. The third step is to determine which of the possible solutions in this space correspond to physiologically meaningful states. Traditionally, linear optimization has been used to predict optimal states such as growth and ATP production [36–38] and newer methods and approaches are being developed to study

Corresponding author: Bernhard O. Palsson (bpalsson@be-research.ucsd.edu).

http://tibtec.trends.com 0167-7799/03/$ - see front matter doi:10.1016/S0167-7799(03)00030-1

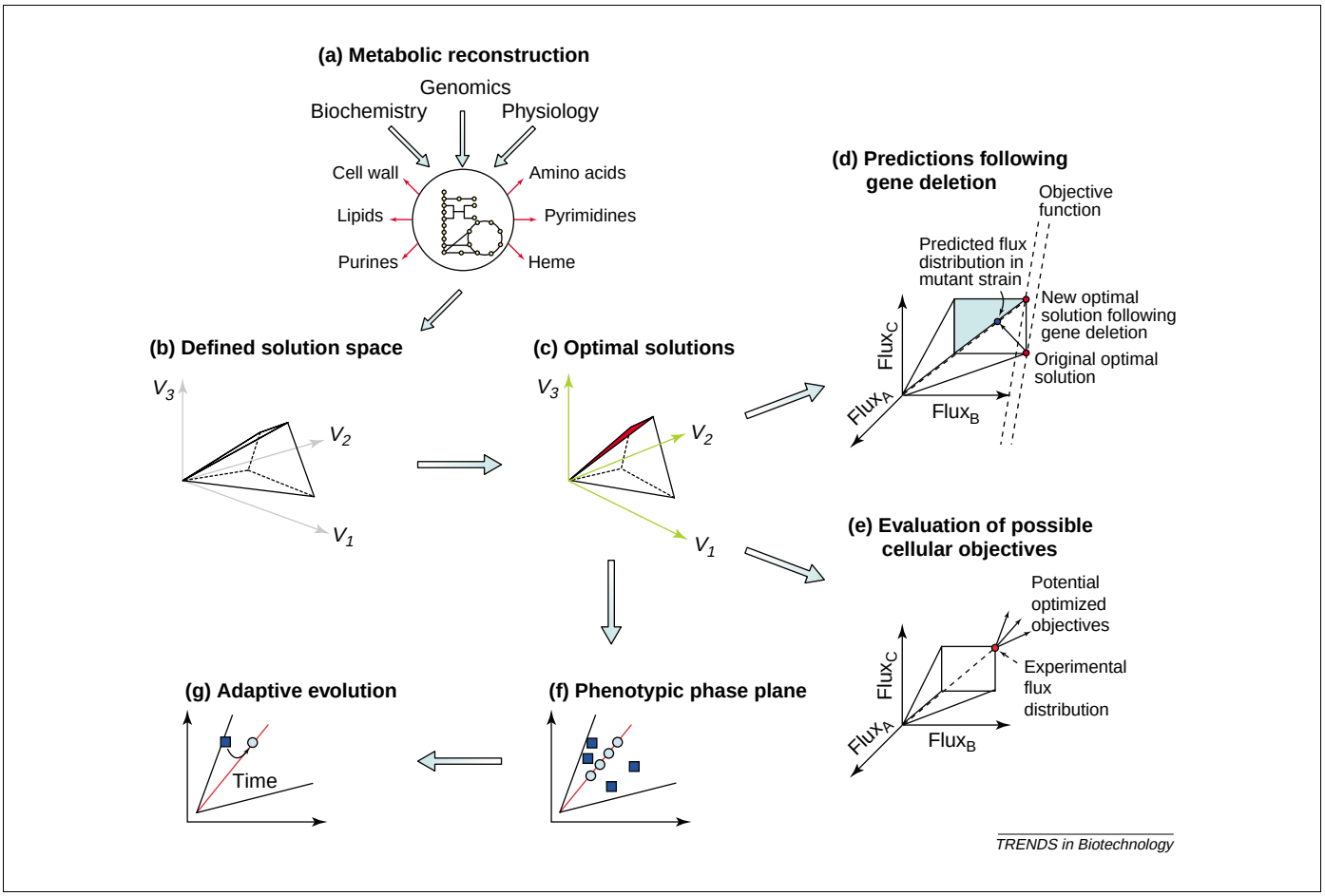

Fig. 1. Process of constraints-based model building and analysis. (a) The process of constraints-based modeling begins with the reconstruction of a metabolic network from known biochemistry, genomics and physiology. (b) Then, governing physico-chemical constraints are used to define a solution space from which are excluded flux distributions that are not allowed to the cell's metabolic network. (c) Within this defined solution space, optimal solution sets can be determined, given a known cellular objective such as the maximization of growth rate. (d) Predictions of flux distributions following a gene deletion can be made based on the assumption that the new strain will maintain a flux distribution as similar as possible to that of the wildtype. Thus, these predictions are not based on optimal growth. (e) Potential cellular objectives that would lead to the observed phenotype can also be evaluated. (f) Phenotypic phase planes show several solution types, including the line of optimal growth (shown in red). (g) Over time, strains can evolve from suboptimal states to predicted optimal growth rates.

the solution space [39–41]. For example, in metabolic networks the region of the space that corresponds to optimal growth has been correlated to experimental data [42]. Methods to study large number of solution types, such as the phenotypic phase plane [43], have been developed and used to drive optimal growth [42] and adaptive evolution experiments [44].

The constraints-based modeling philosophy has been established as an alternative to kinetic theory based models but it is in the early stages of development and we can expect much more progress in this field in the years to come. Results from a constraints-based analysis have recently been compared with results from a kinetic model of human red blood cell metabolism [10]. Some of the current capabilities and potential applications are summarized in Table 2 and we will discuss them in detail.

Current genome-scale metabolic models

Genome-scale constraints-based models of metabolism have been built for several organisms, and some have appeared in the literature, including for *Escherichia coli* [45,46], *Haemophilus influenzae* [47], *Helicobacter pylori*

Table 1. Current dimensions of *in silico* networks

Organism	*E. coli*	*H. influenzae*	*H. pylori*	*S. cerevisiae*
Genome characteristics				
Genome size (Mb)	4.6	1.83	1.66	12.16
Total ORFs	4401	1743	1590	6335
Completion date	1997	1995	1997	1997
Size of current genome-scale networks				
Total metabolic genes	660	412	291	708
Metabolic reactions	720	461	390	1175
Metabolites	436	367	340	723

Three *in silico* networks have been reconstructed and published: *Escherichia coli*, *Haemophilus influenzae*, and *Helicobacter pylori* and *Saccharomyces cerevisiae*. The number of metabolites in *S. cerevisiae* includes metabolites in both the cytosol and the mitochondria. Summary information is provided about each of these organisms and each corresponding model.

Table 2. Genome-scale microbial *in silico* models: methods and applications

In silico methods	Potential applications
Simulated genetic modifications	
Gene deletions or additions	Identification and prioritization of candidate drug targets
Gradual inhibition or enhancement of gene function	Production of low molecular weight products from cellular metabolism
Objective and adaptability studies	
Predictions of optimal growth rate	Direction of adaptive evolution strategies to engineer strains with desired properties
Prediction of flux redistribution following gene deletion	
A priori prediction of the outcome of adaptive evolution	Evaluation of current state of knowledge about metabolic (and regulatory) networks
Redundancy calculations	
Minimal reaction sets	Designing and implementing of experimental programs
Metabolic pathway redundancy	
Most/Least Redundant Subsystems	Analysis of enzyme deficiencies
Accounting for regulation	
Predictions of co-regulated reaction sets	Evaluation of genome annotations
Increased predictive power through incorporation of information about transcriptional regulation	

[48], and *Saccharomyces cerevisiae* [49]. Others, including *Bacillus subtilis*, *Pseudomonas aeruginosa*, and *Pseudomonas putida* have been fully built but not yet published (Sung Park and Jeremy Edwards, unpublished results).

'Genome-scale' is used to describe these models because all of the metabolic reactions that could be determined to take place in an organism based on genome annotation and biochemical literature are included in the model. The number of genes accounted for and the number of reactions included in each of these models is summarized in Table 1. The current models are developed based on 60–70% complete genome annotation. Several lessons have been learned about these organisms as a result of the genome-scale models and many potential biotechnology applications have been identified.

Simulating the results of manipulating gene content and function

There are many experimental methods for changing the gene content of an organism. Genes can be added or deleted, or their functions impaired or enhanced. Reliable computational models that link genotype to phenotype would thus allow for directed manipulation of the gene content of an organism to obtain a desired phenotype. Genome-scale constraints-based models can provide such a link. The ability to predict phenotypic outcomes from genetic inputs would provide a basis for the rational selection of drug targets and the generation of hypotheses about how to metabolically engineer a strain with desired properties.

Gene deletion studies

Individual genes have been deleted from *in silico* models, the consequences of the deletion assessed and the results compared with experimental data. Such *in silico* predictions were found to be ~60% correct for an initial model of *H. pylori* [48] and 86% in *E. coli* [30,45] Interestingly, the failure modes are of great value to the investigator using genome-scale models. Prediction failure means that the *in silico* model is incomplete and is lacking in some way or that the data are potentially incorrect. Thus detailed analysis of the failure modes is important because it will lead to updated and improved *in silico* models for the particular organism and to increased understanding of the organisms' physiology.

Some of the commonly occurring failure modes have been identified. For instance, if a gene deletion is lethal *in silico* and has been experimentally identified as non-lethal, it suggests that there are either unknown isozymes or alternative routes have not been fully characterized. Conversely, if a gene deletion is predicted to be non-essential and is experimentally found to be lethal, it suggests that another factor besides the inability to synthesize biomass is causing the gene deletion to be lethal. Other potential reasons for a gene deletion to be lethal are unmodeled regulatory effects or toxicity of metabolic intermediates.

Making all possible double knockouts is easy *in silico*, but difficult *in vivo*. Thus, if an *in silico* model has been validated for a series of single gene knockouts, it can be used in a prospective fashion to pick what are likely to be informative double knockouts. This approach might prove to be useful by complementing and directing experimental efforts to assess the epistatic consequences of synthetic lethal mutants. There is limited computational experience with double-knockouts [47,50].

Gene additions

The potential of gene additions to the *E. coli* metabolic network [46] to improve the theoretical yield of amino acid production has been evaluated [39]. The optimal production rates of amino acids from a wild-type *in silico* strain were compared with an *in silico* strain that was allowed access to as many as 3400 additional biochemical reactions known to be catalyzed in other species. It was determined that gene additions were able to increase the theoretical maximum yield of seven amino acids in *E. coli* with the addition of known reactions. Interestingly, these increases were generally the result of the addition of only a few genes, often only one or two.

Network robustness

The gradual retardation of enzyme activity can readily be simulated *in silico* by simply constraining the maximum

flux through the reaction. In general, the results show that as the activity level of an enzyme is increased or decreased, overall network function, such as meeting growth requirements, does not change significantly [51]. In other words, the reconstructed network tends to be robust with respect to changes in the activity of individual components. For instance, the flux levels through the enzymes associated with the seven essential *in silico* genes from the central metabolism of *E. coli* were incrementally decreased from their maximum level to zero and the resulting effect on the growth flux was evaluated [51]. Surprisingly, some of these enzymes could be restricted to significantly low levels with essentially zero change in the growth flux.

Studying the objectives and adaptability of microorganisms

Predictions of optimal growth rates

Experimental validation of *in silico* predictions has provided increasing evidence that a primary function of the *E. coli* metabolic network is to maximize growth [42,44]. *E. coli* was grown on several substrates, including acetate, succinate, malate or glucose minimal media, and the corresponding uptake rates, secretion rates and growth rates were experimentally measured. Good correlation was obtained between the growth rates, uptake rates and secretion rates that were experimentally observed and the *in silico* predictions.

Further computational evidence has been obtained supporting the appropriateness of the optimal growth objective for *E. coli* [52]. Objectives for *E. coli* were calculated to best match experimental flux distributions for both aerobic and anaerobic growth. Results showed that the calculated objectives for growth under both aerobic and anaerobic conditions were very similar and corresponded well to biomass production. Flux distribution predictions based on these calculated objective functions were compared with experimental data and it was determined that the internal flux distributions associated with optimal biomass formation matched experimental predictions better than any of the other tested objectives. Analysis using additional constraints on the rate of change of fluxes, called dynamic flux balance analysis, showed that predictions based on the instantaneous optimization of growth match experimental data better than predictions based on overall or end-point optimization of growth [53]. Thus, the instantaneous optimization of growth seems to be the primary objective of *E. coli*'s metabolic network.

Flux distribution following gene deletion

Predictions of growth fluxes and other properties using constrained optimization are based on the hypothesis that cells have evolved to maximize growth rate. However, this expectation might not hold for gene knockout strains. Recently, an analysis method called 'minimal perturbation analysis' has been developed that does not assume that the knockout is initially optimized for growth [41]. Rather, it is assumed that the behavior of knockout strains can be more accurately predicted by assuming that the mutant strain would have an initial flux distribution that is as similar as possible to that of the wildtype strain (subject to additional constraints imposed on the network by the loss of a certain gene) (Fig. 1). Experimental data (both measured internal flux distributions and predicted growth rate) provided evidence for the hypothesis that knockout strains of microorganisms do use their metabolic networks, at least initially, similarly to how they were used in the wildtype strains and thus are not necessarily optimized for growth [41].

A priori predictions of endpoints of adaptive evolution

Evolution can be used to design and improve strain performance [54]. The adaptive evolutionary process presumably leads to optimal strain performance in a defined environment under the applied selection pressure. The prediction of the outcome of adaptive evolution would clearly improve the effectiveness of strain development. A suboptimal growth of *E. coli* on glycerol reproducibly underwent adaptive evolution to achieve its maximal growth rate predicted *a priori* with the constraints-based model. [44]. Thus, *in silico* models show promise in guiding strain engineering and the use of evolutionary pressure.

One interesting aspect of adaptive evolution is the question about the uniqueness of the evolved strain. The evolution of a phosphotransferase system *E. coli* knockout shows that strains with equivalent growth performance can be obtained that utilize their metabolic networks differently [55]. One of the key features of constraints-based models is that the complexity of genome-scale metabolic networks is such that there are multiple equivalent ways in which a particular phenotype can be obtained. Such characteristics of redundancy might become the chief concerns in producing biological designs.

Studying redundancy in metabolic networks

Computing minimal reaction sets

A recent study used the *E. coli* metabolic model to enumerate the number of reactions necessary to maintain the capacity of the metabolic network to produce all of the biomass components necessary for growth [40]. It was discovered that the *E. coli* metabolic network was able to support growth using only 31% of its known metabolic reactions for growth on glucose and only 17% of its metabolic reactions on a rich medium. Thus, *E. coli*'s metabolic network contains many redundant reactions, making *E. coli* robust to changing environmental conditions and failures in many gene products.

Pathway redundancy

Network-based pathways such as extreme pathways and elementary modes circumscribe all optimal and suboptimal steady state flux distributions through a metabolic network, meaning that all possible steady state flux distributions are non-negative linear combinations of either the extreme pathways or the elementary modes [34,35]. Pathway redundancy is a measure of the number distinct routes a metabolic network needs to have to produce a given set of outputs from a given set of inputs, as measured by either the number of elementary modes or extreme pathways. *H. influenzae* amino acid synthesis with minimal medium was calculated to have an average pathway redundancy of 50, whereas amino acid synthesis in *H. pylori* had a pathway redundancy of only two,

indicating a much reduced degree of network robustness [56,57]. The pathway redundancy associated with various growth yields of *E. coli* has been assessed with a model of core metabolism and was shown to correlate the robustness of the organism to gene deletions [58]. Thus, this quantitative measure of redundancy is important for the comparison of robustness in different organisms and in different subsystems, identifying less redundant systems that might be more favorable to drug targeting. Additionally, pathway analysis can more readily identify portions of the network in which the intended effects of genetic manipulations might be negated by the existence of redundant pathways.

Accounting for regulation

The genome-scale models discussed above are based on network topology and the analysis assumes their unfettered use to achieve assumed optimal performance. Cells use complex regulatory networks to achieve their goals that might or might not be consistent with assumptions of optimal performance. Thus, significant need exists to account for regulation in genome-scale models and initial progress in this regard is being made [59].

Systemically correlated reactions

Systemically correlated reactions are reactions that must be used simultaneously under all steady state conditions. The definition and computational identification of correlated reactions in metabolic networks might give insight into regulatory strategies. Metabolic reactions that are always used together in the same ratios have been identified in subsystems and genome-scale models of *H. influenzae* and *H. pylori* [48,60]. Because these correlated reactions must occur simultaneously to maintain homeostasis, the genes that encode the corresponding enzymes are likely candidates for co-regulation.

Transcriptional regulation

A Boolean formalism can be used to describe metabolic events in which transcriptional regulation is a dominant factor [59]. This Boolean formalism is not used to model the regulatory network *per se* but rather is used to set time-dependent constraints on the metabolic network stemming from transcriptional regulation. The inclusion of such regulation can improve mutant phenotype predictions and enable better simulation of the time course of growth and metabolic uptake and

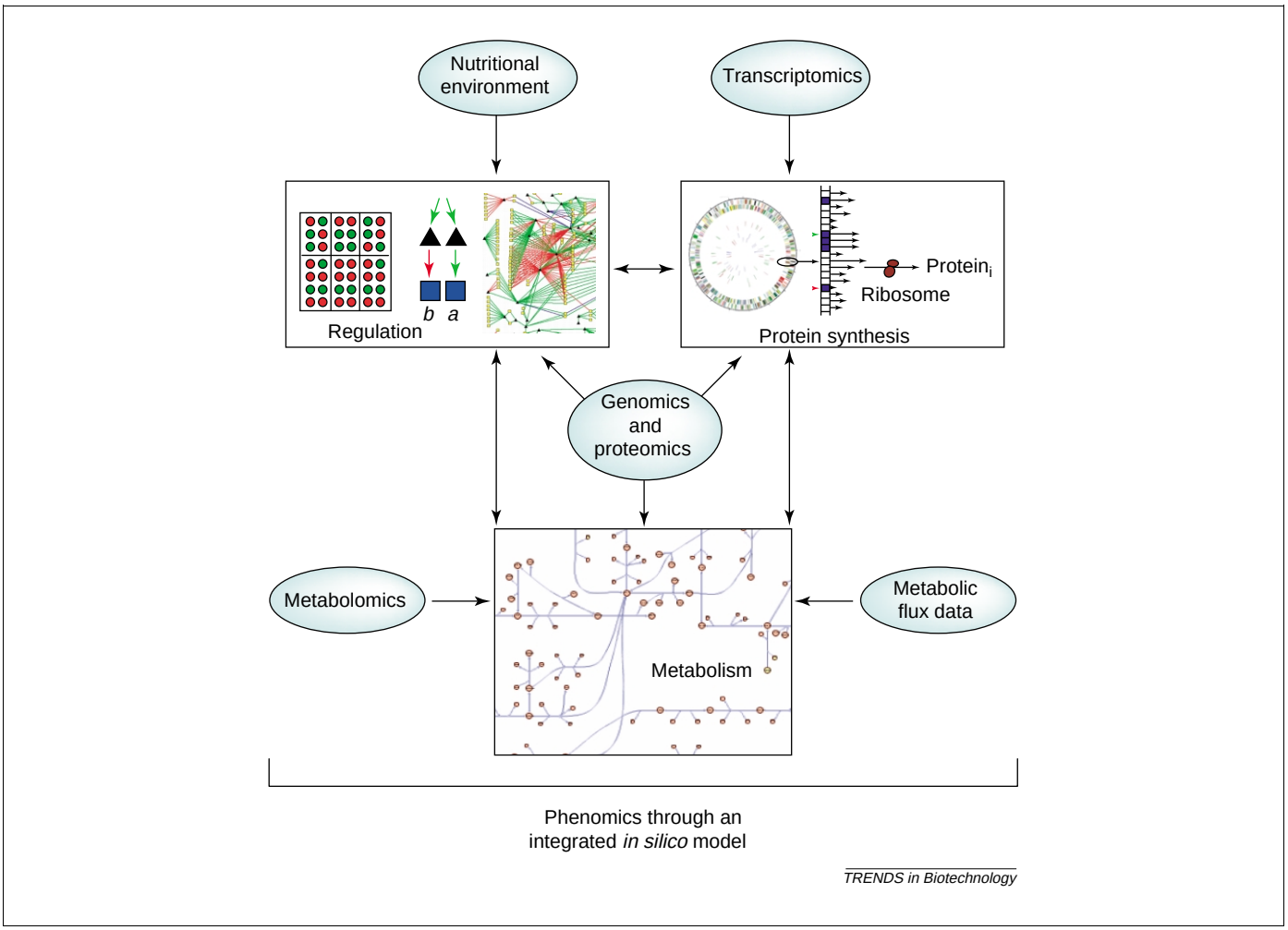

Fig. 2. Integration of 'omics' data into constraints-based models. Constraints-based models are being expanded to form an integrated model of metabolism, transcriptional regulation and protein synthesis. This integrated model provides a framework in which to house various 'omics' datasets. The regulatory state of a cell is set by its nutritional environment. This regulatory state then sets the transcription state of the genome, which controls protein synthesis in the cell. The protein produced then establishes the enzymes present to catalyze metabolic reactions. The costs of synthesizing the RNA transcripts and the proteins will place sequence-based, calculable demands on the metabolic model. The metabolic model can then be used to generate testable hypotheses and interpret experimental results.

secretion rates. Analysis of *E. coli*, without accounting for transcriptional regulation, made correct predictions of experimental data in 97 out of 116 cases considered (83.6%) [30]. However, accounting for transcriptional regulation using the Boolean formalism, 106 out of the 116 cases (91.4%) were correctly predicted and included cases that could not be predicted without the regulatory effects [30,59]. Thus, the incorporation of regulation represents an improvement in the predictive capability of genome-scale *in silico* models. The remaining false predictions are usually attributed to the buildup of toxic substances.

Some future directions and needs

Beyond metabolism: constraints-based modeling as means to integrate 'omics' data

Genome-scale models provide a framework in which high-throughput biological data can be integrated and thus broadens our capacity to predict phenotypes. *In silico* methods are being developed that allow for the integration of heterogeneous data sets, including genomics, proteomics, transcriptomics and metabolomics [17,61] (Fig. 2). In the case of *E. coli*, the inclusion of metabolism, transcription, translation and regulation will lead to models that account for ~2000 open reading frames. Such a model will be achievable shortly using the constraints-based methods described here and if cell replication is included will lead to comprehensive genome-scale models that include all major cellular functions.

From concepts to industrial strength models and simulators

Delivering enhanced efficiency and productivity is the goal of *in silico* modeling technologies in any industry. One of the main challenges confronting the biotechnology community is to couple *in silico* modeling with high-throughput experimental research and development programs leading to efficient biological discovery. As new discoveries are made in the laboratory, the content of organism-specific models needs to be continuously updated to reflect this new knowledge. In this way, a model should serve as the most concise representation of our biological understanding of the organism (Fig. 2). An accurate model can then serve as the basis for prospective research and experiments can be designed around *in silico*-derived hypotheses of the type reviewed herein.

The introduction and scale-up of such an iterative model development and implementation paradigm requires the development of robust computational platforms. These platforms must also allow for the integration of massive amounts of biological data with a comprehensive model to enable integrative analysis, simulation and visualization of the underlying data and modeling predictions. Furthermore, they must be user-friendly so as to broaden the user base and acceptance of *in silico* predictive technologies beyond the advanced modeling community.

To meet this challenge the SimPheny™ modeling platform has been recently developed (Fig. 3). This system supports the efficient implementation of constraints-based modeling technologies to support model-driven systems biology research. In the course of developing such an application to transfer technology from concept to product several issues were addressed. Technical issues include the use of complete elementally and charge balanced reactions (a concept often simplified in the literature), the explicit association logic between genes, proteins and reactions, and the annotation of model content with confidence ratings and complete reference linking to

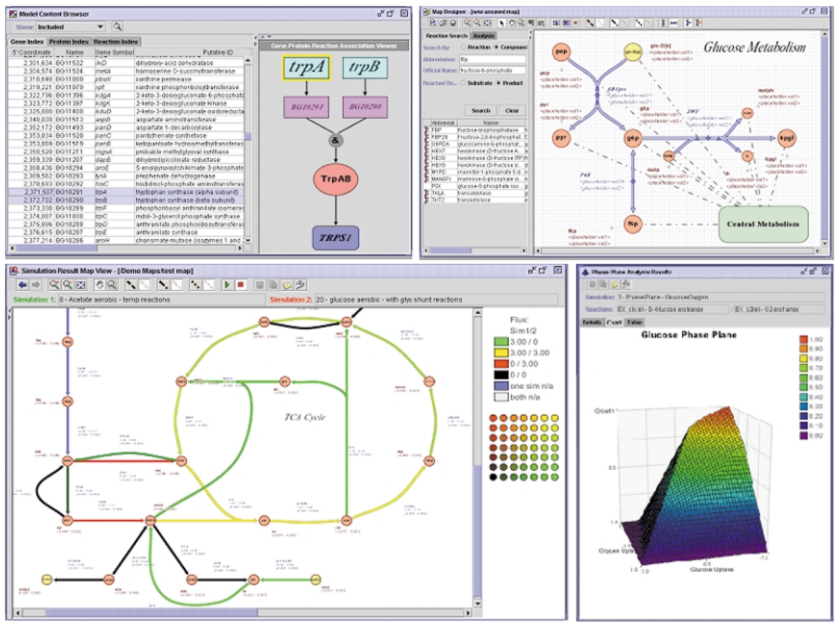

Fig. 3. SimPheny™: software suite for building genome-scale constraints-based models and their use for simulation and visualization. Top left: loading the annotated genome and forming gene–protein–reaction associations. Top-right: the Atlas manger for building maps enabling the visualization of all links and connections in a network and the ability to associate simulated values with map locations. Lower left: the dynamic display of a flux solution on a metabolic map where the flux levels are color-coded. Lower right: the computations of phenotypic functions and the display of optimal growth rates as a function of two variables.

experimental results and the literature. Implementation issues include methods to version validated models as they are iteratively developed, to support multi-user simulation studies and comparative modeling and also to provide animated visualization of pathway flux distributions. The continued development of supporting modeling platforms will be crucial to delivering on the promise and potential of modeling technologies to improve biotechnology research.

In closing

Genome-scale models of metabolism have been developed for several microbial cells. Many useful results have already been derived from these models and as they grow in scope and validation, an even broader set of uses can be anticipated. Perhaps most important is the integration of diverse omics data and the ability to account for the management of the genome and associated regulatory processes. Eventually, such models are expected to become routinely used for a variety of biotechnological applications and the generation of novel biological designs.

Acknowledgements

We acknowledge the support of the National Science Foundation (BES 01–20363), the National Institutes of Health (GM 57089), and the Whitaker Foundation (Graduate Research Fellowship to JP).

References

1 Chance, B. *et al.* (1962) Analogue and digital computer representations of biochemical processes. *Fed. Proc.* 21, 75–86
2 Garfinkel, D. *et al.* (1970) Computer applications to biochemical kinetics. *Annu. Rev. Biochem.* 39, 473–498
3 Reich, J.G. and Sel'kov, E.E. (1981) *Energy Metabolism of the Cell: A Theoretical Treatise*, Academic Press
4 Joshi, A. and Palsson, B.O. (1989) Metabolic dynamics in the human red cell. Part I–A comprehensive kinetic model. *J. Theor. Biol.* 141, 515–528
5 Mulquiney, P.J. *et al.* (1999) Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: *in vivo* kinetic characterization of 2,3-bisphosphoglycerate synthase/phosphatase using 13C and 31P NMR. *Biochem. J.* 342, 567–580
6 Mulquiney, P.J. and Kuchel, P.W. (1999) Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement. *Biochem. J.* 342, 581–596
7 Mulquiney, P.J. and Kuchel, P.W. (1999) Model of 2,3-bisphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: computer simulation and metabolic control analysis. *Biochem. J.* 342, 597–604
8 Jamshidi, N. *et al.* (2001) Dynamic simulation of the human red blood cell metabolic network. *Bioinformatics* 17, 286–287
9 Kauffman, K.J. *et al.* (2002) Description and analysis of metabolic connectivity and dynamics in the human red blood cell. *Biophys. J.* 83, 646–662
10 Wiback, S.J. and Palsson, B.O. (2002) Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.* 83, 808–818
11 Tomita, M. *et al.* (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15, 72–84
12 Loew, L.M. and Schaff, J.C. (2001) The virtual cell: a software environment for computational cell biology. *Trends Biotechnol.* 19, 401–406
13 Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512
14 Drell, D. (2002) The department of energy microbial cell project: a 180° paradigm shift for biology. *Omics* 6, 3–9
15 Lander, E.S. (1999) Array of hope. *Nat. Genet.* 21, 3–4
16 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
17 Palsson, B.O. (2002) *In silico* biology through 'omics'. *Nat. Biotechnol.* 20, 649–650
18 Palsson, B.O. (2000) The challenges of in silico biology. *Nat. Biotechnol.* 18, 1147–1150
19 Bailey, J.E. (2001) Complex biology with no parameters. *Nat. Biotechnol.* 19, 503–504
20 Arkin, A. *et al.* (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648
21 McAdams, H.H. and Arkin, A. (1998) Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* 27, 199–224
22 Varner, J. and Ramkrishna, D. (1999) Mathematical models of metabolic pathways. *Curr. Opin. Biotechnol.* 10, 146–150
23 Guardia, M.J. *et al.* (2000) Cybernetic modeling and regulation of metabolic pathways in multiple steady states of hybridoma cells. *Biotechnol. Prog.* 16, 847–853
24 Edwards, J.S. *et al.* (2002) Metabolic modelling of microbes: the flux-balance approach. *Environ. Microbiol.* 4, 133–140
25 Covert, M.W. *et al.* (2001) Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.* 26, 179–186
26 Selkov, E. Jr *et al.* (1998) MPW: the metabolic pathways database. *Nucleic Acids Res.* 26, 43–45
27 Gaasterland, T. and Selkov, E. (1995) Reconstruction of metabolic networks using incomplete information. *Ismb* 3, 127–135
28 Bono, H. *et al.* (1998) Reconstruction of amino acid biosynthesis pathways from the complete genome sequence. *Genome Res.* 8, 203–210
29 Karp, P.D. (1998) Metabolic databases. *Trends Biochem. Sci.* 23, 114–116
30 Covert, M.W. and Palsson, B.O. (2002) Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* 277, 28058–28064
31 Wyrick, J.J. and Young, R.A. (2002) Deciphering gene expression regulatory networks. *Curr. Opin. Genet. Dev.* 12, 130–136
32 Beard, D.A. *et al.* (2002) Energy balance for analysis of complex metabolic networks. *Biophys. J.* 83, 79–86
33 Price, N.D. *et al.* (2002) Extreme pathways and Kirchhoff's second law. *Biophys. J.* 83, 2879–2882
34 Schilling, C.H. *et al.* (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248
35 Schuster, S. *et al.* (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326–332
36 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
37 Bonarius, H.P.J. *et al.* (1997) Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* 15, 308–314
38 Edwards, J.S. *et al.* (1999) Metabolic flux balance analysis. *Metabolic Engineering* (Lee, S.Y., Papoutsakis, E.T., *et al.* eds), Marcel Deker
39 Burgard, A.P. and Maranas, C.D. (2001) Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol. Bioeng.* 74, 364–375
40 Burgard, A.P. *et al.* (2001) Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.* 17, 791–797
41 Segre, D. *et al.* (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15112–15117
42 Edwards, J.S. *et al.* (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19, 125–130
43 Edwards, J.S. *et al.* (2002) Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* 77, 27–36
44 Ibarra, R.U. *et al.* (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 420, 186–189
45 Edwards, J.S. and Palsson, B.O. (2000) The *Escherichia coli* MG1655

in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5528–5533
46 Pramanik, J. and Keasling, J.D. (1997) Stoichiometric model of *Escherichia coli* metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* 56, 398–421
47 Edwards, J.S. and Palsson, B.O. (1999) Systems properties of the Haemophilus influenzae Rd metabolic genotype. *J. Biol. Chem.* 274, 17410–17416
48 Schilling, C.H. *et al.* (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184, 4582–4593
49 Forster, J. *et al.* (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13, 244–253
50 Schilling, C.H. *et al.* (1999) Towards metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* 15, 288–295
51 Edwards, J.S. and Palsson, B.O. (2000) Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* 16, 927–939
52 Burgard, A.P. and Maranas, C.D. An optimization based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol. Bioeng.* (in press)
53 Mahadevan, R. *et al.* (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 83, 1331–1340
54 Sauer, U. (2001) Evolutionary engineering of industrially important microbial phenotypes. *Adv. Biochem. Eng. Biotechnol.* 73, 129–169
55 Flores, S. *et al.* (2002) Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by (13)C labeling and NMR spectroscopy. *Metab. Eng.* 4, 124–137
56 Papin, J.A. *et al.* (2002) The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.* 215, 67–82
57 Price, N.D. *et al.* (2002) Determination of redundancy and systems properties of *Helicobacter pylori*'s metabolic network using genome-scale extreme pathway analysis. *Genome Res.* 12, 760–769
58 Stelling, J. *et al.* (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190–193
59 Covert, M.W. *et al.* (2001) Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213, 73–88
60 Schilling, C.H. and Palsson, B.O. (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.* 203, 249–283
61 Allen, T.E. and Palsson, B.O. (2003) Sequenced-based analysis of metabolic demands for protein synthesis in prokaryotes. *J. Theor. Biol.* 220, 1–18