cloning of two novel poly(ADP-ribose) polymerase homologues. *Genomics* 57, 442–445
18 Sallmann, F.R. *et al.* (2000) Characterization of sPARP-1. An alternative product of PARP-1 gene with poly(ADP-ribose) polymerase activity independent of DNA strand breaks. *J. Biol. Chem.* 275, 15504–15511
19 Kickhoefer, V.A. *et al.* (1999) The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase. *J. Cell Biol.* 146, 917–928
20 Rome, L. *et al.* (1991) Unlocking vaults: organelles in search of a function. *Trends Cell Biol.* 1, 47–50
21 Kickhoefer, V.A. *et al.* (1996) Vaults are the answer, what is the question? *Trends Cell Biol.* 6, 174–178
22 Kickhoefer, V.A. *et al.* (1999) Vaults and telomerase share a common subunit, TEP1. *J. Biol. Chem.* 274, 32712–32717
23 Nakayama, J. *et al.* (1997) TLP1: a gene encoding a protein component of mammalian telomerase is a novel member of WD repeats family. *Cell* 88, 875–884
24 Harrington, L. *et al.* (1997) A mammalian telomerase-associated protein. *Science* 275, 973–977
25 Kong, L.B. *et al.* (2000) RNA location and modeling of a WD40 repeat domain within the vault. *RNA* 6, 890–900
26 Chugani, D.C. *et al.* (1993) Evidence that vault ribonucleoprotein particles localize to the nuclear pore complex. *J. Cell Sci.* 106, 23–29
27 Kong, L.B. *et al.* (1999) Structure of the vault, a ubiquitous cellular component. *Structure* 7, 371–379
28 Abbondanza, C. *et al.* (1998) Interaction of vault particles with estrogen receptor in the MCF-7 breast cancer cell. *J. Cell Biol.* 141, 1301–1310
29 Kickhoefer, V.A. *et al.* (1998) Vaults are up-regulated in multidrug-resistant cancer cell lines. *J. Biol. Chem.* 273, 8971–8974
30 Schroeijers, A.B. *et al.* (2000) The Mr 193,000 vault protein is up-regulated in multidrug-resistant cancer cell lines. *Cancer Res.* 60, 1104–1110
31 Nugent, C.I. and Lundblad, V. (1998) The telomerase reverse transcriptase: components and regulation. *Genes Dev.* 12, 1073–1085
32 Greider, C.W. and Blackburn, E.H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* 43, 405–413
33 Chong, L. *et al.* (1995) A human telomeric protein. *Science* 270, 1663–1667
34 Broccoli, D. *et al.* (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* 17, 231–235
35 Bilaud, T. *et al.* (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nat. Genet.* 17, 236–239
36 Collins, K. (2000) Mammalian telomeres and telomerase. *Curr. Opin. Cell Biol.* 12, 378–383
37 Griffith, J.D. *et al.* (1999) Mammalian telomeres end in a large duplex loop. *Cell* 97, 503–514
38 Griffith, J. *et al.* (1998) TRF1 promotes parallel pairing of telomeric tracts *in vitro*. *J. Mol. Biol.* 278, 79–88
39 van Steensel, B. and de Lange, T. (1997) Control of telomere length by the human telomeric protein TRF1. *Nature* 385, 740–743
40 Smith, S. and de Lange, T. (2000) Tankyrase promotes telomere elongation in human cells. *Curr. Biol.* 10, 1299–1302
41 Chi, N.W. and Lodish, H.F. (2000) Tankyrase is a golgi-associated MAP kinase substrate that interacts with IRAP in GLUT4 vesicles. *J. Biol. Chem.* 275, 38437–38444
42 d'Adda di Fagagna, F. *et al.* (1999) Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability. *Nat. Genet.* 23, 76–80
43 Liu, Y. *et al.* (2000) Telomerase-associated protein TEP1 is not essential for telomerase activity or telomere length maintenance *in vivo*. *Mol. Cell. Biol.* 20, 8178–8184
44 Smith, S. and de Lange, T. (1999) Cell cycle dependent localization of the telomeric PARP, tankyrase, to nuclear pore complexes and centrosomes. *J. Cell Sci.* 112, 3649–3656
45 Earle, E. *et al.* (2000) Poly(ADP-ribose) polymerase at active centromeres and neocentromeres at metaphase. *Hum. Mol. Genet.* 9, 187–194
46 Lin, W. *et al.* (1997) Isolation and characterization of the cDNA encoding bovine poly(ADP-ribose) glycohydrolase. *J. Biol. Chem.* 272, 11895–11901
47 Zhang, X. *et al.* (1999) Telomere shortening and apoptosis in telomerase-inhibited human tumor cells. *Genes Dev.* 13, 2388–2399
48 Hahn, W.C. *et al.* (1999) Inhibition of telomerase limits the growth of human cancer cells. *Nat. Med.* 5, 1164–1170

Metabolic modeling of microbial strains *in silico*

Markus W. Covert, Christophe H. Schilling, Iman Famili, Jeremy S. Edwards, Igor I. Goryanin, Evgeni Selkov and Bernhard O. Palsson

The large volume of genome-scale data that is being produced and made available in databases on the World Wide Web is demanding the development of integrated mathematical models of cellular processes. The analysis of reconstructed metabolic networks as systems leads to the development of an *in silico* or computer representation of collections of cellular metabolic constituents, their interactions and their integrated function as a whole. The use of quantitative analysis methods to generate testable hypotheses and drive experimentation at a whole-genome level signals the advent of a systemic modeling approach to cellular and molecular biology.

Many high-throughput experimental technologies have been developed in recent years that have enabled full genomic sequences to be obtained, genome-wide expression assessment to be performed and the protein portfolio of particular cells and organisms examined. It is likely that these experimental technologies will only increase in speed and potential in the coming years. Moreover, the development of high-throughput phenotyping technologies is expected. These developments are having a profound impact on the general thinking in the biological sciences. For example, it is becoming universally accepted that cells should be viewed as systems. Such systems represent complex networks of interacting gene products to produce physiological functions.

As in many other fields of science and engineering, the large-scale generation of complex data sets calls for their mathematical analysis and computer simulation. In the past, such endeavors were the curiosity of a few and were hampered by the lack of good data upon which reliable models could be built. However, mathematical model building is now taking the 'center stage' in biology, and its use and importance is likely to grow. How does one begin to build such models? Knowledge of the list of

 0968-0004/01/$ – see front matter PII: S0968-0004(00)01754-0

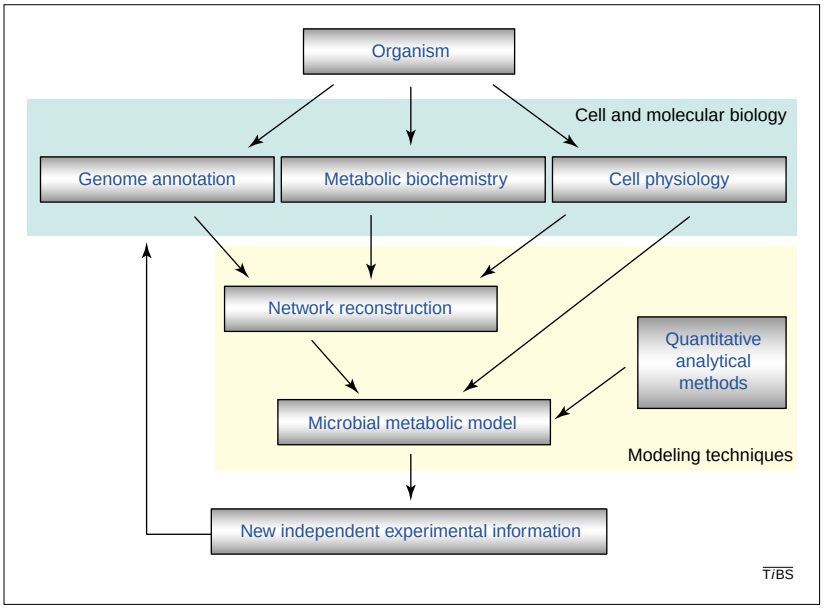

Fig. 1. Integrated process of microbial metabolic model construction. Such construction requires a comprehensive knowledge of the metabolism of an organism. From the annotated genome sequence and the experimentally determined biochemical and physiological characteristics of a cell, the metabolic reaction network can be reconstructed. This network is then modified in the context of other physiological constraints to produce a mathematical model, which can be used to generate quantitatively testable hypotheses *in silico*. As the model is used to direct an experimental plan, it can be important in further re-examining the biological properties of the organism.

components that comprise cells and how they interact is growing as evidenced by the large number of metabolic databases accessible through the World Wide Web. The consequences of these interactions must now be analyzed and determined.

In this paper, the process of building mathematical models of carbon and energy metabolism for microbial organisms is described, and is shown schematically in Fig. 1. From the annotated genome sequence and the experimentally determined biochemical and physiological characteristics for a given organism, the network of metabolic reactions can be reconstructed, as far as it is known. The reconstructed metabolic network is then analyzed using various mathematical modeling techniques. These quantitative analysis methods enable the simulation of microbial growth and behavior *in silico* and therefore have many important applications in the field of metabolic engineering, in which organisms are genetically modified to enhance desirable properties. Achievements in the mathematical analysis of microbial metabolism, as well as issues and challenges in the field, are discussed.

M.W. Covert, C.H. Schilling, I. Famili and B.O. Palsson*
Dept Bioengineering, University of California, San Diego, La Jolla, CA 92093-0412, USA.
*e-mail: palsson@ucsd.edu

J.S. Edwards
Dept Chemical Engineering, University of Delaware, Newark, DE 19716, USA.

I.I. Goryanin
RIS Informatics, GlaxoWellcome, Stevenage, UK SG2 9AR.

E. Selkov
Integrated Genomics, 2201 W. Campbell Park Dr., Chicago, IL 60612, USA.

Reconstruction of metabolic reaction networks

Genome annotation

Reconstruction of metabolic reaction networks in an organism begins with the thorough examination of the genome. The first step in functional annotation of a genome sequence is to identify the coding regions or open reading frames (ORFs) on the sequence. Each ORF is searched initially against databases with the goal of assigning a putative function to it. Established algorithms (e.g. the BLAST and FASTA family of programs) can be used to determine the similarity between a given sequence and gene or protein sequences deposited in sequence databases[1,2]. As the number of sequenced organisms rises, putative gene functions could also be determined by various types of gene clustering[3,4]. A large fraction of the genes for a newly sequenced organism can usually be readily identified by these methods.

Several high-quality genomic database and metabolic network reconstruction web sites that provide access to the annotated genome sequences of many organisms can be found online[5,6]. Two types of metabolic databases are available from the Web: organism-specific and general-purpose databases. Organism-specific databases, such as EcoCyc (Ref. 7), are designed to provide a user-friendly interface for the inspection of the metabolic characteristics (i.e. experimental and sequence data) of a single genome (see Box 1 for a list of Web addresses). The general-purpose databases, such as the Metabolic Pathways Database (MPW)[8,9] and the Kyoto Encyclopedia of Genes and Genomes (KEGG)[10,11], contain sequence data for a large spectrum of organisms. Once the functional assignment of a sequenced genome is complete, a software program such as the What Is There (WIT) system[4] or KEGG can search the general-purpose database for the closest set of metabolic maps complementing the annotated genome. The metabolic maps in WIT and KEGG are then used as templates whereon the metabolic network reconstruction of the organism can be represented as an organism-fitted subset of their pathway collections to be used similarly to an organism-specific database.

The wealth of biochemistry knowledge contained in general-purpose metabolic databases enables the automated metabolic reconstruction of any sequenced organism, including those for which only a partial or 'gapped' genome sequence is available[12]. These reconstructed metabolic networks, based exclusively on genomic data, can form the backbone of an *in silico* organism. However, to build a more comprehensive

Box 1. List of addresses for Web sites relevant to genome annotation

EcoCyc	http://ecocyc.panbio.com/ecocyc/ecocyc.html
Metabolic pathways database (MPW)	http://igweb.integratedgenomics.com/MPW/
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.ad.jp/kegg/
What Is There (WIT)	http://wit.mcs.anl.gov/WIT/
Biology Workbench	http://workbench.sdsc.edu

metabolic network, these automatically constructed networks must be evaluated in the context of experimental data, specifically the biochemical and growth characteristics of the organism.

Metabolic biochemistry

Genome sequencing and annotation have already outpaced the generation of biochemical and physiological data. The time required to make comprehensive experimental and literature surveys of the biochemical and physiological characteristics for each organism of interest can be excessively long. However, many of the organisms whose genomes have been sequenced completely have also been the subjects of extensive biochemical research. After locating all the known metabolic genes on an annotated genome, the additional information gained from experimental data can make the reconstructed metabolic network more complete. Continued experimental investigation of the metabolic biochemistry of an organism is important and has three main purposes. These are: (1) to assign pertinent biochemical reactions to the enzymes found in the genome; (2) to validate and scrutinize information already found in the genome; and (3) to determine the presence of reactions or pathways not indicated by current genomic data.

The use of a reconstructed metabolic network depends largely on its accuracy. Biochemical evidence helps to assign a function to a particular gene, and validates the corresponding links in the reconstructed network. Also, functionality can sometimes be determined more easily by biochemical than by genomic studies. The sequencing of many organisms has shown that 20–30% of all eubacterial genes annotated so far are found to be species specific having, as yet, no known homologs[13]. It follows from this observation that various organisms might have evolved widely different methods of catalyzing similar reactions or pathways. The proteins involved in these reactions would have disparate sequences despite their similar function, and would thus be undetectable by sequence comparison. The substrate specificity of many enzymes can also introduce serious errors into the metabolic reconstruction if genomes are annotated by sequence similarity alone. Combining the findings of experimentalists with the information contained in an annotated genome will reconcile these issues and lead to the most complete reconstruction of the metabolic network.

Cell physiology

At the current state of knowledge in genetics and biochemistry, a number of the metabolic genes that contribute significantly to the metabolic phenotype of an organism cannot be identified. The identification of these additional genes depends on the inclusion of cell-physiological data.

Knowledge of the physiology of an organism gives indirect evidence to the presence or absence of certain metabolic reactions in a cell. For example, if experiments suggest that an organism can grow without a certain essential amino acid, but the reconstructed metabolic network is not able to produce that amino acid *in silico*, perhaps for lack of a single enzyme, then for the metabolic reaction network to have any practical meaning, the missing steps in the pathway must be included. Once a network reconstruction has been developed and evaluated in the context of available biochemical and physiological information, it can be applied to various types of mathematical analysis.

Model construction and analysis

Mathematical models and their computer simulation allow us to examine the integrated function of the reconstructed metabolic network. A well-defined network by itself is not sufficient to describe the behavior of a system quantitatively, as shown in Fig. 2. Here, an analogy is drawn between simulating traffic conditions in a typical city and simulating the behavior of a microbial metabolic network. The first step for both situations is to generate a list of the functional components for the system. For traffic simulation, this could be represented by a list of all the major roads in the city, together with the places that are connected by these roads. For a cell, gene products are discovered and characterized as described earlier. In both situations, the next step is to determine how these functional components are connected. This information can be integrated into a 'map', that is a road map or a reconstructed metabolic map. Once a network has been described in sufficient detail, some qualitative predictions can be made. For example, a road map is used to determine the route in travelling from one place to another. Relative distances can be compared. Similarly, the reconstructed metabolic network can be used to study the connectivity of metabolites and other characteristics of network structure[14–16].

The completed road map, however, has limitations. For example, although the possibility of driving from one destination to another can be ascertained, the actual travel time is unknown. The travel time depends partly on traffic conditions, which in turn depend on the road, the time of day, the weather and several other contributing factors. Obtaining all the necessary data to specify each contributing factor is not feasible. However, by estimating or approximating many of these conditions, thereby creating a realistic model for traffic conditions, it is possible to obtain a reasonable calculation of the travel time. Such a calculation could not be made with a road map alone.

Similarly, without including more information, the reconstructed metabolic map of an organism is limited in its ability to generate quantitative predictions about the phenotype. The behavior of a cell depends on many factors such as temperature, substrate availability, the presence of signaling

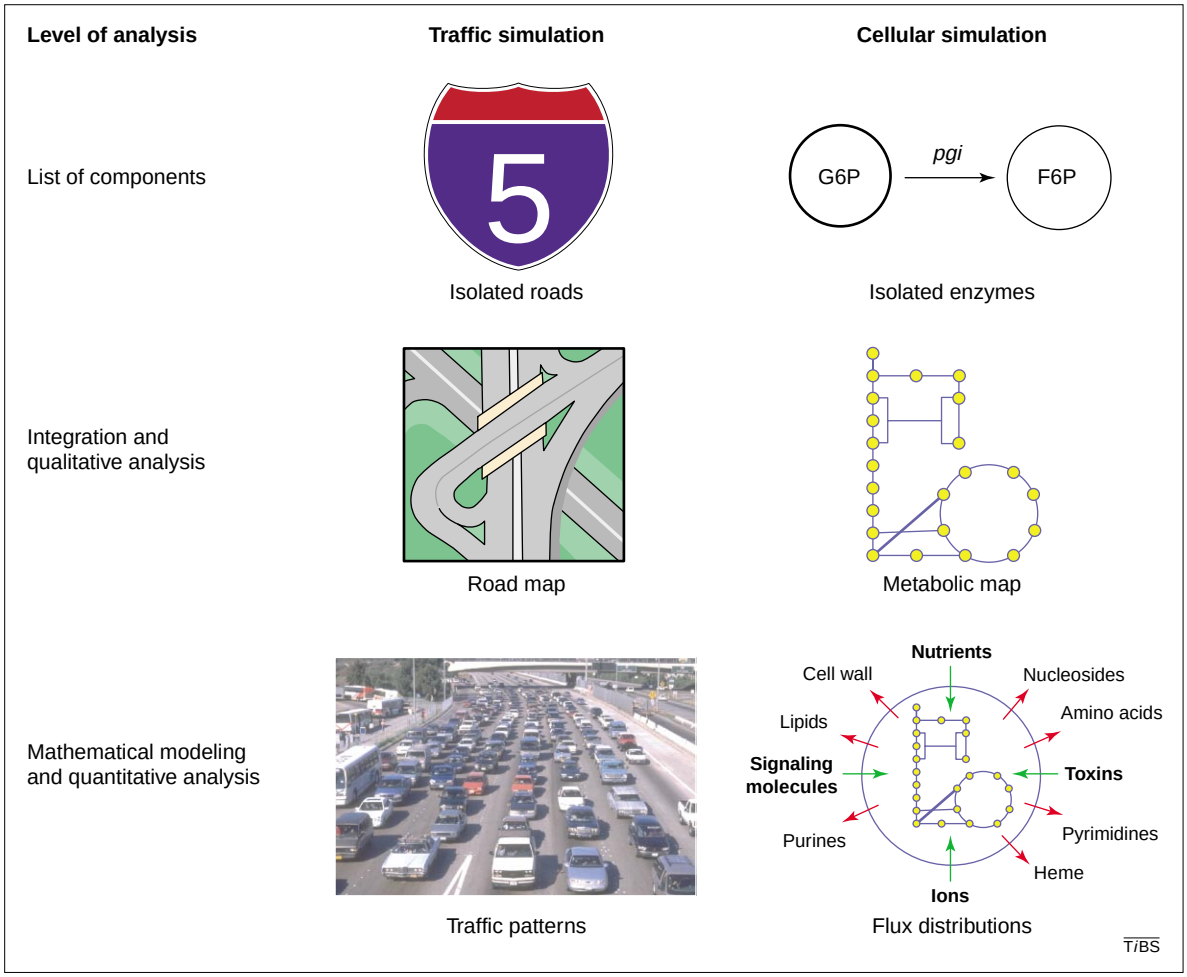

Fig. 2. An analogy between simulation of traffic conditions in a typical city and simulation of a microbial cell using systems analysis. For both simulations, the first step is to generate a list of all the relevant components (e.g. roads or gene products) of the system, after which the integration of these components must be determined and specified. At this point some qualitative predictions can be made about the performance of the system. Finally, mathematical modeling is used to quantitatively analyze the system as it responds to a number of environmental factors or a change in the network. Abbreviations and explanations: F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; pgi (phosphoglucose isomerase) is the enzyme that catalyzes the reaction.

molecules, and other environmental parameters, many of which have yet to be specified completely. Properties such as stoichiometry are relatively easy to establish, whereas kinetic properties of bacterial metabolism are typically much more difficult to obtain under all possible environmental conditions. Several approaches to dynamic cellular modeling have been developed[17–22]. However, it is clear that for detailed dynamic model building to succeed on a whole-genome scale, much progress must be made in estimating kinetic parameters of enzymes, either from first principles[23] or by correlation to known properties of similar enzymes[24].

Although kinetic constants are hard to obtain, the network structure can be established as outlined above. From the traffic analogy given in Fig. 2, it is clear that the study of fluxes through such a network is important and can be accomplished without having detailed kinetic information. Metabolic fluxes can be seen as a fundamental determinant of cell physiology because they show quantitatively the contributions of various pathways to overall cellular functions[17]. A common way of relating cell genotype to phenotype is therefore by analyzing the fluxes in the metabolic network. Some approaches to cellular flux analysis are described in Box 2.

The ability of flux-based analytical techniques to generate quantitative hypotheses has given these techniques a wide range of applications in the field of metabolic engineering, whether in the large-scale microbial generation of valuable substances or in pollutant degradation. For example, they have been used effectively to model penicillin production by *Penicillium chrysogenum*[25,26], to improve the yield of aromatic amino acids by *Escherichia coli*[27] and lysine by *Corynebacterium glutamicum*[17] through metabolic engineering of central metabolism, and to model enhanced biological phosphorus removal for wastewater treatment[28]. Flux-based approaches also lend themselves easily to genotype–phenotype studies *in silico* and can be used to analyze enzyme deficiencies or identify drug targets as has been shown in gene deletion and metabolite connectivity studies for *E. coli*[29] and *Haemophius influenzae*[14,30]. Used in conjunction with an experimental study, the

Box 2. A brief description of flux-based analysis methods

Methods that fall into this category are pathway analysis[a], flux-balance analysis (FBA)[b–d] and metabolic flux analysis (MFA)[e]. All are based on the principle of conservation of mass for the metabolites of a given metabolic network. Pathway analysis is a method for generally defining the structure of the metabolic network as it relates to the overall metabolic capabilities[f,g] of the organism. In contrast, FBA examines the metabolic network from a performance perspective, using linear optimization to determine optimal cellular behaviors under changing environmental and genetic conditions[h]. MFA characterizes the flux distribution in more experimental detail, estimating internal fluxes based on a combination of isotope labeling techniques[i] and mathematical analysis[j].

References

a Schilling, C.H. *et al.* (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* 15, 296–303

b Schilling, C.H. *et al.* (1999) Towards metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* 15, 288–295

c Varma, A. and Palsson, B.O. (1994) Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology* 12, 994–998

d Bonarius, H.P.J. *et al.* (1997) Flux analysis of undetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* 15, 308–314

e Stephanopoulos, G. *et al.* (1998) *Metabolic Engineering*, Academic Press

f Schilling, C.H. and Palsson, B.O. (2000) Assesment of the metabolic capabilities of *Haemophilus influenzae Rd* through a genome-scale pathway analysis. *J. Theor. Biol.* 203, 249–283

g Schuster, S. *et al.* (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* 17, 53–60

h Edwards, J.S. *et al.* (1999) Metabolic flux balance analysis. In *Metabolic Engineering* (Lee, S.Y. and Papoutsakis, E.T., eds), Marcel Deker

i Wiechert, W. and de Graaf, A.A. (1996) *In vivo* stationary flux analysis by 13C labeling experiments. *Adv. Biochem. Eng. Biotechnol.* 54, 109–154

j Christensen, B. and Nielsen, J. (2000) Metabolic network analysis. A powerful tool in metabolic engineering. *Adv. Biochem. Eng. Biotechnol.* 66, 209–231

metabolic outcome and growth of *E. coli* using acetate and succinate as single-carbon sources has been accurately predicted[31]. Further applications of flux-analysis techniques have been reviewed[17,32,33].

Although the goal of developing a completely specified cellular model will require the inclusion of kinetic parameters, the development of flux-analysis methods and other approaches has many applications and will continue to lead to the generation of novel and important quantitative hypotheses about microbial behavior, even in the absence of detailed kinetic information.

Table 1. Comparison of genomic characteristics and *in silico* metabolic model characteristics between three bacterial strains[a]

Properties	Bacterial strain		
	E. coli K-12	***H. influenzae Rd***	***H. pylori* 26695**
Genome characteristics			
Genome length (bp)	4 639 221	1 830 135	1 667 867
G + C content	51%	38%	39%
Open reading frames	4288	1743	1590
Identified database match	2656	1011	1091
No database match	1632	732	499
***In silico* metabolic networks**			
Genes included (% of known ORF)	660 (~25%)	400 (~40%)	290 (~27%)
Associated reactions[b]	697	412	272
Other reactions[c]	42	49	109
Metabolites	442	367	332

[a]Abbreviations: *E. coli*, *Escherichia coli*; *H. Influenzae Rd*, *Haemophilus influenzae Rd*; *H. pylori*, *Helicobacter pylori*.
Reactions included in the network are grouped [b]as associated with a particular gene, in which their inclusion is on the basis of direct genomic or biochemical evidence, or [c]on either indirect cell physiological evidence or inferred by the demands imposed on the metabolic reaction network.

Model characteristics

A comparison of the genomic characteristics and *in silico* metabolic model characteristics for three bacterial strains is shown in Table 1 (Refs 14,29) (C.H. Schilling, PhD Thesis, University of California, 2000). These *in silico* models represent between 25% and 40% of the known ORFs in their *in vivo* counterparts. Figure 3a shows the reaction complement of the gastric pathogen *Helicobacter pylori* 26695 in greater detail. The Venn diagram is used to categorize the inclusion of reactions in the reconstructed network. Many of them have been included with a combination of different kinds of evidence, as shown by the overlap in circles. In better-known organisms, such as *E. coli* and *Saccharomyces cerevisiae*, the overlap between the circles is expected to be much greater.

As shown in the figure, the bulk of reactions in the network were derived from genomic evidence (almost 73% in the case of *Helicobacter pylori*). Approximately half of the remaining reactions in the reconstructed metabolic network were included on the basis of observations found in the literature, whether from direct biochemical evidence or indirect physiological evidence. The remaining reactions, labeled as inferred reactions in Fig. 3a, have been included on the basis of the metabolic demands of the reconstructed network, but without experimental or genome evidence. Each inferred reaction added to the reconstructed metabolic network will eventually require further experimental justification.

Modeling issues

There are two primary issues regarding the construction of microbial metabolic models. First, not all of the reactions suggested by these models are found directly in the databases or the biochemical

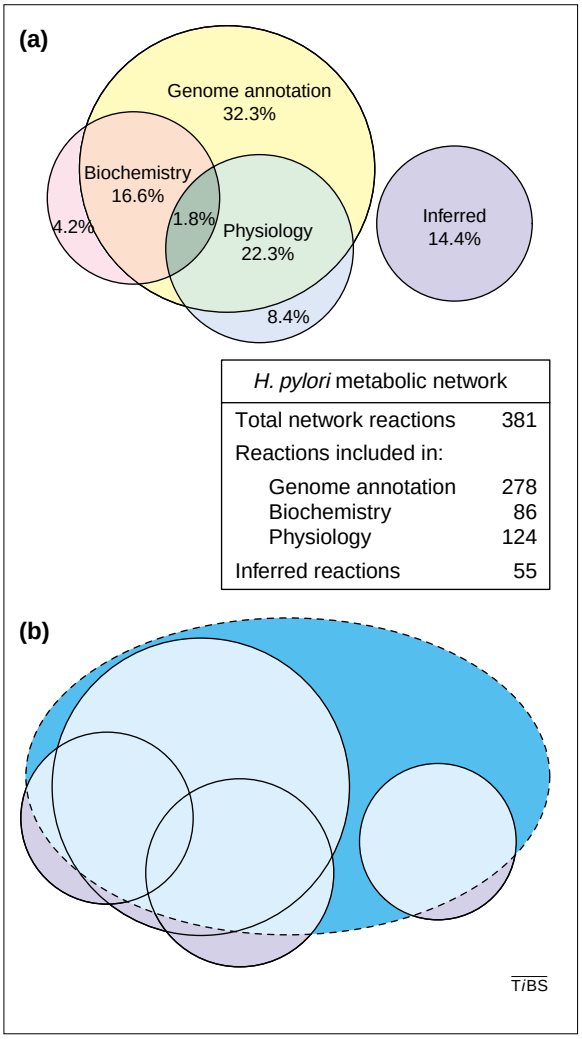

Fig. 3. Venn diagram of reactions in the *Helicobacter pylori* (*H. pylori*) *in silico* metabolic model. (a) Reactions are categorized by the type of evidence used to justify their inclusion in the reconstructed network. The number of reactions, by category, is given in the table (inset) and as a percentage of the total reactions (shown in each colored area). Many of the reactions have been included by evidence from multiple sources, as shown by overlapping circles. (b) This diagram depicts the real metabolic reaction network, illustrating the two major issues associated with metabolic reconstruction. If the actual complete metabolic network is the area enclosed by a dashed line, the reconstructed model probably consists of true reactions (light blue area) and some reactions that are not actually occurring in the cell (i.e. false reactions; purple), and is probably missing some crucial reactions (dark blue area).

literature, and second, not all of the metabolic genes present in the genotype are accounted for – or even noted – in the model, because their functions are as yet undiscovered (Fig. 3b).

However, a 'real metabolic network' exists for the example given in Fig. 3a, that is the actual set of all the relevant reactions that occur in *H. pylori* strain 26695 are included in the model. This network, surrounded by a dashed line, is superimposed on the network defined by our model. The light-blue area is the set of all reactions that are found both in strain 26695 and in our model, the 'true' reactions. The purple area represents 'false' reactions that were included in the model but do not actually occur in *H. pylori* strain 26695. These reactions represent mistaken assumptions used in creating the model.

The secondary issue is the inverse problem of the above: many of the proteins synthesized by the organism are not accounted for in the metabolic reconstruction. These 'missing' reactions are shown by the dark-blue area in Fig. 3b. It is likely that some of the metabolic reactions that are catalyzed by the organism are as yet undiscovered. This implies that functionalities open to the organism are neglected by the model.

These metabolic network reconstruction issues can be resolved in part as the model is applied to various analyses. For example, the metabolic *H. pylori* model was used to re-examine the annotation of the metabolic network. All of the genes that were included in the reconstruction of *H. pylori* metabolism without direct genomic or biochemical evidence can be thought of as hypothetical. The presence of these hypothetical genes can be determined by collecting the sequences of other organisms' copies of the hypothetical genes and using BLAST to compare them with the *H. pylori* genome sequence. The genes that are found to be significantly homologous to loci in the *H. pylori* genome sequence can then be studied experimentally to verify their proposed function on the basis of the reconstruction and BLAST analysis.

One such gene product included in the *H. pylori* model without genomic or biochemical evidence was malate dehydrogenase. A subsequent study indicated that on locus HP0086 of the *H. pylori* genome, an ORF was located that showed significant similarity (36.81%) and identity (25.93%) with a malate:quinone oxidoreductase in glutamic acid bacterium *C. glutamicum*[34]. Thus, the analysis of microbial metabolic models can also have bioinformatic applications, such as functional assignment of ORFs, in addition to the more obvious experimental applications.

There are also significant issues pertaining to the analysis of microbial metabolic models. It has been noted above that flux models can successfully predict the effects of gene knockouts and the metabolic behavior of an organism quantitatively. The specific advantage of the flux-based analyses is that such models do not require experimental information such as enzyme kinetics, regulatory mechanisms, intracellular concentrations or enzyme activity profiles.

However, the attractive simplicity of the models also sets some inevitable limits for their predictions. Simulations of microbial behavior reflect only the topology of reconstructed systems (as contained in the reconstructed network) and the boundary conditions of the system (such as extracellular substrate concentrations). Flux-based models currently incorporate no control mechanisms of any kind, predicting a theoretical metabolic potential that assumes the constitutive expression of all genes in the metabolic reaction network. This assumption could lead to false predictions. Additionally, the flux models

Fig. 4. Applying the 'traditional scientific method' of iterative hypothesis development in the post-genomic era. Once the *in silico* microbial metabolic model has been constructed, it can be used to generate testable hypotheses. These hypotheses are examined, both in traditional experimental studies (right-hand side) and using new bioinformatics or genome sequencing techniques (left-hand side), to discover new attributes of the metabolic network. After these discoveries have been incorporated into the *in silico* metabolic model, it can be used to generate new hypotheses in a subsequent iteration. Adapted, with permission, from Ref. 36.

describe a stable stationary state of metabolism. Any projections based on their analysis toward intracellular metabolic dynamics associated with the cell cycle or cell differentiation must be made very cautiously. It is expected that these issues will be resolved as analysis methods are improved and developed to incorporate additional experimental information as listed above, resulting in models that are more complicated but perhaps more accurate in predicting the dynamic behavior of microbial organisms.

Future challenges

In silico models of metabolic networks will be subjected to an ongoing iterative model-building process just as complex systems in other branches of science and engineering have in the past. This process is illustrated in Fig. 4. Here, the traditional scientific method is depicted in the context of biology in the post-genomic era. Hypotheses based on the metabolic analysis of microbial strains are examined both in terms of an experimental study and using bioinformatics techniques. Experimentalists and bioinformaticists must work cooperatively to provide information to analysts, from which *in silico* representations of microbial metabolism can be created. The analysis of these models will lead to suggestions for bioinformatic and experimental studies which, in turn, will contribute to a more robust characterization of the metabolism of an organism. Once refined in this manner, the metabolic model can be used to generate a new set of hypotheses in a subsequent iteration.

Another important challenge in the improvement of microbial metabolic modeling is the expansion of integrated development environment (IDE) software[35], which combines all of the available tools and methods for model creation, analysis and development, with convenient access to the metabolic and enzymology databases and a user-friendly interface that can be understood by scientists with diverse backgrounds and training. The development of such IDE software will make the *in silico* modeling of microbial metabolism more widespread and facilitate the introduction of quantitative analysis to microbiologists, leading to the generation of new and important experimental hypotheses and industrial developments.

Concurrently, these models need to be expanded to incorporate features of the genome other than simply metabolism. This broadening in scope will occur as a direct result of more-advanced analysis methods. The further development of *in silico* microbial models that quantitatively simulate complexities such as signal transduction, control mechanisms and the dynamic behavior of microorganisms, will be vitally important in the field of metabolic engineering as well as in the effort to model eukaryotic organisms. In the latter, the genome is larger and therefore the percentage of known metabolic genes in the genome is generally far smaller. As the efficacy of analysis increases, the new *in silico* microbial models will add various functions until finally the construction of a whole-cell model has been completed, an accomplishment that would greatly contribute to our understanding of the essential nature of the cell.

References

1 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* 25, 3389–3402
2 Pearson, W.R. *et al.* (1997) Comparison of DNA sequences with protein sequences. *Genomics* 46, 24–36
3 Eisenberg, D. *et al.* (2000) Protein function in the post-genomic era. *Nature* 405, 823–826
4 Overbeek, R. *et al.* (2000) WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucl. Acids Res.* 28, 123–125
5 Karp, P.D. *et al.* (1999) Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* 17, 275–281
6 Karp, P.D. (1998) Metabolic databases. *Trends Biochem. Sci.* 23, 114–116
7 Karp, P.D. *et al.* (1999) Eco Cyc: encyclopedia of *Escherichia coli* genes and metabolism. *Nucl. Acids Res.* 27, 55–58
8 Selkov, E., Jr *et al.* (1998) MPW: the Metabolic Pathways Database. *Nucl. Acids Res.* 26, 43–45
9 Gaasterland, T. and Selkov, E. (1995) Reconstruction of metabolic networks using incomplete information. In *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology* (3) (Rawlings, C. *et al.*, eds), pp. 127–135, American Association for Artificial Intelligence Press, Menlo Park, CA
10 Bono, H. *et al.* (1998) Reconstruction of amino acid biosynthesis pathways from the complete genome sequence. *Genome Res.* 8, 203–210
11 Ogata, H. *et al.* (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl. Acids Res.* 27, 29–34
12 Selkov, E. *et al.* (2000) Functional analysis of gapped microbial genomes: amino acid metabolism of *Thiobacillus ferrooxidans*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 3509–3514
13 Saier, M.H., Jr (1998) Genome sequencing and informatics: new tools for biochemical discoveries. *Plant Physiol.* 117, 1129–1133
14 Edwards, J. and Palsson, B. (1999) Properties of the *Haemophilus influenzae Rd* metabolic genotype. *J. Biol. Chem.* 274, 17410–17416
15 Fell, D.A. and Wagner, A. (2000) Structural properties of metabolic networks: implications for evolution and modelling of metabolism. In *Animating the Cellular Map* (Hofmeyr, J-H.S. *et al.*, eds), pp. 79–85, Stellenbosch University Press

16 Jeong, H. *et al.* (2000) The large-scale organization of metabolic networks. *Nature* 407, 651–654
17 Stephanopoulos, G. *et al.* (1998) *Metabolic Engineering*, Academic Press
18 Heinrich, R. and Schuster, S. (1996) *The Regulation of Cellular Systems*, Chapman & Hall
19 Fell, D. (1996) *Understanding the Control of Metabolism*, Portland Press
20 Reich, J.G. and Sel'kov, E.E. (1981) *Energy Metabolism of the Cell*, Academic Press
21 Varner, J. and Ramkrishna, D. (1999) Mathematical models of metabolic pathways. *Curr. Opin. Biotechnol.* 10, 146–150
22 McAdams, H.H. and Arkin, A. (1998) Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* 27, 199–224
23 Vaseghi, S. *et al.* (1999) *In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metab. Eng.* 1, 128–140
24 Selkov, E. *et al.* (1996) The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucl. Acids Res.* 24, 26–28
25 Jorgensen, H. *et al.* (1995) Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnol. Bioeng.* 46, 117–131
26 Henriksen, C.M. *et al.* (1996) Growth energetics and metabolism fluxes in continuous cultures of *Penicillium chrysogenum*. *J. Biotechnol.* 45, 149–164
27 Liao, J.C. *et al.* (1996) Pathway analysis, engineering and physiological considerations for redirecting central metabolism. *Biotechnol. Bioeng.* 52, 129–140
28 Pramanik, J. *et al.* (1999) Development and validation of a flux-based stoichiometric model for enhanced biological phosphorus removal metabolism. *Water Res.*33, 462–476
29 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
30 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
31 Edwards, J.S. *et al.* *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotech.* 19, 125–130
32 Schilling, C.H. *et al.* (1999) Towards metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* 15, 288–295
33 Gombert, A.K. and Nielsen, J. (2000) Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* 11, 180–186
34 Kather, B. *et al.* (2000) Another unusual type of citric acid cycle enzyme in *Helicobacter pylori*: the malate:quinone oxidoreductase. *J. Bacteriol.* 182, 3204–3209
35 Goryanin, I. *et al.* (1999) Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics* 15, 749–758
36 Palsson, B. (2000) The challenges of *in silico* biology. *Nat. Biotechnol.* 18, 1147–1150

IκB-independent control of NF-κB activity by modulatory phosphorylations

M. Lienhard Schmitz, Susanne Bacher and Michael Kracht

Activation of the transcription factor nuclear factor κB (NF-κB) requires its release from inhibitor of NF-κB (IκB) proteins in the cytoplasm. Much work has focussed on the identification of pathways regulating this cytosolic rate-limiting step of NF-κB activation. However, there is increasing evidence for another complex level of NF-κB activation, which involves modulatory phosphorylations of the DNA-binding subunits. These phosphorylations can control several functions of NF-κB, including DNA binding and transactivation properties, as well as interactions between the transcription factor and regulatory proteins. Although their overall impact on NF-κB function has yet to be determined, modifications of this factor will very probably provide a mechanism to fine tune NF-κB function.

For more than a decade now, the transcription factor nuclear factor κB (NF-κB) has attracted attention because of both its unique activation pathways and its physiological importance as a key regulatory molecule in the immune response, cell proliferation and apoptosis[1]. The DNA-binding form of NF-κB is dimeric. These dimers can be composed of various combinations of the five different DNA-binding subunits – NF-κB1 (p50 and its precursor p105), NF-κB2 (p52 and its precursor p100), c-Rel, RelB and p65 (RelA) – although the most frequently observed form of NF-κB is a p50–p65 heterodimer. All NF-κB family members have a conserved N-terminal Rel-homology domain (RHD), which is responsible for dimerization, DNA binding and interaction with IκBs (inhibitors of NF-κB)[1–4]. The precursor proteins p105 and p100 can be processed by the proteasome to generate p50 and p52, respectively. In addition, p50 can be produced by an alternative pathway, which involves the cotranslational dimerization of the RHD of p50 with p105 (Ref. 5).

In most cell types, NF-κB is maintained in an inactive form in the cytoplasm by association with IκBs. Physical and chemical stresses, viruses, bacteria and pro-inflammatory cytokines [e.g. interleukin (IL)-1 and tumour necrosis factor (TNF)] activate NF-κB by inducing the rapid phosphorylation of IκB and its subsequent ubiquitination and proteolytic degradation. Released NF-κB then translocates to the nucleus, binds to its cognate DNA element and activates transcription of numerous target genes[2]. The inducible phosphorylation of IκB is mediated by recently identified IκB kinases (IKKα, β and ε). The catalytic subunits, IKKα and IKKβ, and the regulatory IKKγ/NEMO (NF-κB essential modulator) subunit, form the prototypic core IκB kinase complex (IKC)[3]. Importantly, this complex serves as an intracellular point of convergence for distinct signals that ultimately activate NF-κB (Refs 1–4).

 0968-0004/01/$ – see front matter PII: S0968-0004(00)01753-9