How Will Bioinformatics Influence Metabolic Engineering?

Jeremy S. Edwards, Bernhard O. Palsson

Department of Bioengineering, University of California-San Diego, La Jolla, California 92093-0412; telephone: 619-534-5668; fax: 619-822-0240; e-mail: palsson@ucsd.edu

Received 7 July 1997; accepted 19 September 1997

Abstract: Ten microbial genomes have been fully sequenced to date, and the sequencing of many more genomes is expected to be completed before the end of the century. The assignment of function to open reading frames (ORFs) is progressing, and for some genomes over 70% of functional assignments have been made. The majority of the assigned ORFs relate to metabolic functions. Thus, the complete genetic and biochemical functions of a number of microbial cells may be soon available. From a metabolic engineering standpoint, these developments open a new realm of possibilities. Metabolic analysis and engineering strategies can now be built on a sound genomic basis. An important question that now arises; how should these tasks be approached? Flux-balance analysis (FBA) has the potential to play an important role. It is based on the fundamental principle of mass conservation. It requires only the stoichiometric matrix, the metabolic demands, and some strain specific parameters. Importantly, no enzymatic kinetic data is required. In this article, we show how the genomically defined microbial metabolic genotypes can be analyzed by FBA. Fundamental concepts of metabolic genotype, metabolic phenotype, metabolic redundancy and robustness are defined and examples of their use given. We discuss the advantage of this approach, and how FBA is expected to find uses in the near future. FBA is likely to become an important analysis tool for genomically based approaches to metabolic engineering, strain design, and development. *Biotechnol Bioeng* **58:** 162–169, 1998.
Keywords: bioinformatics; metabolic engineering; genetic engineering; mathematical analysis; stoichiometry; enzyme kinetics; modal analysis; genetic circuits

INTRODUCTION

Considerable interest in the redirection of metabolic fluxes for medical and industrial purposes has developed in recent years. As a result, the field of Metabolic Engineering has been born (Bailey, 1991; Stephanopoulos & Sinskey, 1993; Stephanopoulos & Vallino, 1991), whose goal is to implement desirable metabolic behavior in living cells. The engineering approach to analysis and design is to have a mathematical or computer model, e.g., a dynamic simulator, of metabolism that is based on fundamental physicochemical laws and principles. The metabolic engineer hopes that such models can be used to systematically ''design'' a new strain. The methods of recombinant DNA technology could then be applied to achieve the desired changes in the genotype of the cell of interest. However, much of the experimental work on strain development has followed the ''controlled experiment'' approach of biology, by introducing or deleting one or a few genes and determining the effects of such changes on cell behavior. Using this approach, one can, in a step-by-step fashion, develop strains that can be used for bioprocessing. This difference has led to a recent review in the field to conclude that ''despite the recent surge of interest in metabolic engineering, a great disparity still exists between the power of available molecular biological techniques and the ability to rationally analyze biochemical networks'' (Stephanopoulos, 1994).

While the interest in the willful manipulation of cellular metabolism is growing, a landmark in biological history has occurred; *Haemophilus influenzae* became the first cell whose genetic sequence is completely known. The field of microbial genetics has thus entered a new era where an increasing number of microbial genomes are being completely sequenced (Fleischmann, Adams, & White, 1995; Fraser, Gocayne, & White, 1995; Koonin, Mushegian, & Rudd, 1996). Currently ten microbial genomes have been fully sequenced, and this number is growing at a rapid pace (Table I). The microbial genomes which have been completely sequenced include: *Mycoplasma genitalium, Helicobacter pylori, Escherichia coli,* and *Methanococcus jannaschii* (Bult, White, & Olsen, 1996). The identification and functional assignment of the open reading frames (ORFs) defines the metabolic genotype of a fully sequenced strain. In the first completely sequenced organisms, the majority of genes found encode for gene products involved in metabolic functions (Ouzounis & Casari, 1996).

The impact that these developments have on metabolic modeling and engineering of microbial strains is potentially

Correspondence to: Bernhard Palsson
Contract grant sponsors: Fulbright and Ib Heinricksen Fellowships; University of California Biotechnology Training Fellowship

 CCC 0006-3592/98/020162-08

Table I. Completely sequenced organisms (TIGR-Web Site, 1997).

Genome	size (Mb)
Haemophilus influenzae	1.83
Mycoplasma genitalium	0.58
Methanococcus jannaschii	1.66
Synechocystis sp. (PCC 6803)	3.57
Mycoplasma pneumoniae	0.81
Saccharomyces cerevisiae	13
Escherichia coli K-12	4.60
Helicobacter pylori	1.66
Borrelia burgdorferii	1.30
Archaeoglobus fulgidus	2.20

quite significant. In this article, we explore this impact and how to meet the challenges that microbial genomics has presented biochemical and bioengineering.

BIOINFORMATICS AND MICROBIAL METABOLISM

The biological information pouring out of laboratories worldwide is rapidly being organized into databases. The number of biological databases has grown to exceed 100 (Karp, 1996a) and many are available on the World Wide Web (Botstein & Cherry, 1997). These databases include information about issues such as DNA sequences, ORF assignments, protein structures, and the proteins expressed in a bacterium under a particular condition.

One important current challenge is to provide for database interoperability (Karp, 1996a). This allows the exploration of the function of a gene in the context of its physiological function, such as an enzyme in intermediary metabolism. A user could link together the sequence of an enzyme, the three-dimensional nature of the protein, and its biochemical function. Interoperability should also allow for the comparison of the genetic content and the biochemical functionality of two cells. Thus, within just a few years one will be able to browse through the genetic and biochemical information available for a variety of species. Presently, this is possible for some of the fully sequenced microbial strains (Karp, 1996b), and an effectively complete database of the genetic and biochemical characteristics are available at the stroke of a key.

From a metabolic engineering standpoint, the promise of these databases is important; soon the complete metabolic genotype of a strain to be engineered may become available. The ambiguities associated with what metabolic transformations that take place in a particular cell of interest will be reduced, if not eliminated. Thus, the scope of a quantitative systems analysis of microbial metabolism will be defined. The selection of the methods that are suitable to carry out such analysis will be chosen based on the biological data that is at hand, and using these methods, realistic models of whole cells may now be achievable. How will such models be formulated?

FLUX BALANCING

In recent years, an approach to the analysis of metabolic behavior has been developed that relies on balancing metabolic fluxes. This approach is based on the fundamental law of mass conservation, and the flux-balance method is able to provide insightful information about the systemic constraints placed on metabolic function. Flux-balance analysis (FBA) is performed under steady-state conditions, and it requires information only about the stoichiometry of metabolic pathways, metabolic demands, and a few strain specific parameters. Pathway flux distributions are obtained using different solution methods, but no information on metabolite concentrations or transient behavior results from this type analysis.

The fundamental principle underlying FBA is the conservation of mass. A flux balance can be written for each metabolite (X_i) within a metabolic system to yield the dynamic mass balance equations that interconnect the various metabolites. Equating the rate of accumulation of X_i to its net rate of production, the dynamic mass balance for X_i is:

$$\frac{dX_i}{dt} = V_{syn} - V_{deg} - (V_{use} - V_{trans}) \tag{1}$$

where the subscripts, *syn* and *deg* refer to the metabolic synthesis and degradation of metabolite X_i. The uptake or secretion flux, V_{trans}, can be determined experimentally. The growth and maintenance requirements, V_{use}, can be accurately estimated from cellular composition (Ingraham, Maalce, & Neidhardt, 1983; Neidhardt, Ingraham, & Schaechter, 1990; Varma & Palsson, 1993). Equation (1) therefore, can be written as:

$$\frac{dX_i}{dt} = V_{syn} - V_{deg} - b_i \tag{2}$$

where b_i is the net transport out of our defined metabolic system. Generally, for a metabolic network that contains m metabolites and n metabolic fluxes, all the transient material balances can be represented by a single matrix equation,

$$\frac{d\boldsymbol{X}}{dt} = \mathbf{S} \cdot \mathbf{v} - \mathbf{b} \tag{3}$$

where $\boldsymbol{X}$ is an m dimensional vector of metabolite amounts per cell, $\mathbf{v}$ is the vector of n metabolic fluxes, $\mathbf{S}$ is the stoichiometric $m \times n$ matrix, and $\mathbf{b}$ is the vector of known metabolic demands. The element S_{ij} is the stoichiometric coefficient that indicates the amount of the i^{th} compound produced per unit flux of the j^{th} reaction.

The time constants characterizing metabolic transients are typically very rapid compared to the time constants of cell growth and process dynamics, and the transient mass balances can be simplified to only consider the steady-state

behavior. Eliminating the time derivative in Equation (3) yields,

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \tag{4}$$

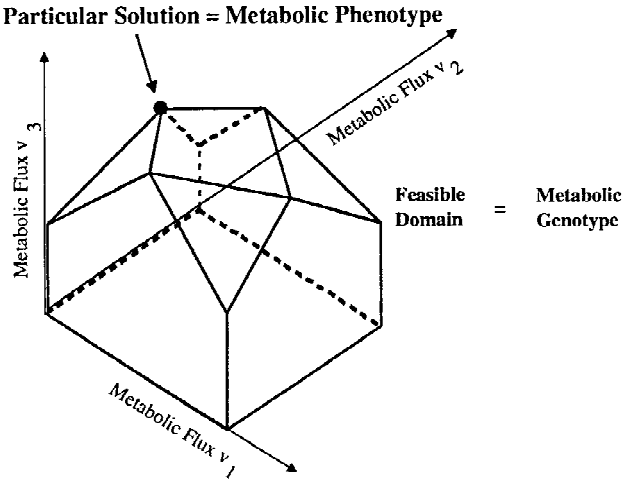
This equation simply states that over long periods of time, the formation of fluxes of a metabolite must be balanced by the degradation fluxes. Otherwise, significant amounts of the metabolite will accumulate inside the metabolic network. Note that this balance equation is formally analogous to Kirchhoff's current law used in electrical circuit analysis.

The flux-balance equation is typically under-determined ($m < n$), and cannot be solved using Gaussian elimination. Thus, additional information is needed to solve for all the metabolic fluxes. Various techniques have been used to solve this equation. Several researchers have made sufficient measurements of external fluxes to either completely determine or over-determine the system (Jorgensen, Nielsen, & Villadsen, 1995; Papoutsakis & Meyer, 1985a; Papoutsakis & Meyer, 1985b; Papoutsakis, 1984; Pons, Dussap, Pequignot, & Gros, 1996; Vallino & Stephanopoulos, 1993). In order for measurements of only the external fluxes to completely determine the system, additional assumptions are required, such as neglecting certain reactions occurring within the cell.

For these completely sequenced organisms, cellular metabolism is defined, and the cellular inventory of metabolic gene products is expressed in the stoichiometric matrix (**S**). The metabolic genotype of an organism then is defined by all the allowable reactions that can occur with a given gene set. Mathematically, the metabolic capabilities of a metabolic genotype is defined as the null space of **S** (Fig. 1). The null space of **S** is typically large, and it represents the flexibility that a cell has in determining the use of its metabolic capabilities. The measurement of the external fluxes alone is not sufficient to uniquely determine the full metabolic flux map.

In addition to the measurement of external fluxes, internal metabolic fluxes have been measured and used to determine metabolic flux distributions (Delgado & Liao, 1997; Sauer et al., 1997; Zupke & Stephanopoulos, 1994, 1995). These measurements are sufficient to determine subsystems of the complete metabolic genotype, and frequently this is adequate to answer the desired questions. However, the measurement of internal fluxes is not always practical, and these measurements can only allow for the determination of the metabolic fluxes in subsystems of the metabolic network.

The measurement of metabolic fluxes is thus important to determine how a cell chooses to use its metabolism under a given condition. Because such measurements are difficult and not generally available, alternative approaches that can be used to explore the capabilities of a metabolic genotype become important. The metabolic capabilities of an organism can be explored using linear programming (LP). The application of LP to metabolic systems has recently been reviewed (Bonarius, Schmid, & Tramper, 1997; Varma & Palsson, 1994a). Linear Programming has been used to determine a number of metabolic functions (Table II), including the biochemical production capabilities and calculate maximum yields, the population stability of genetically engineered strains, prediction of metabolic responses to oxygen availability, and to describe cellular behavior in batch, fed-batch, and continuous culture (Varma & Palsson, 1994b) (Table II). Based on these studies, it has been hypothesized that metabolism operates in a stoichiometrically optimal fashion, and that its behavior in wild-type strains can be predicted based on the metabolic genotype.



Metabolic Genotype:
All allowable flux distributions by a set of metabolic genes. Mathematically speaking, the null space of S. (Varma and Palsson, 1994a)

Metabolic Phenotype:
A particular flux distribution used under given conditions. Mathematically speaking, an optimal solution obtained through linear programming. (Varma and Palsson, 1994a)

Metabolic Redundancy:
The ability of the metabolic circuit to adjust to in the absence of a gene without changes to the phenotype.

Metabolic Robustness:
The ability of the metabolic circuit to adjust to decreased fluxes through essential enzymes without changes to the phenotype.

Figure 1. A schematic illustration of the solution domain that is defined by flux balance constraints. The domain illustrated has been called the *metabolic genotype,* because it represents all possible flux distributions with the set of metabolic enzymes given, whereas a specific solution has been called the *metabolic phenotype,* because it represents a particular flux distribution occurring under a defined set of conditions (Varma & Palsson, 1994a). The assessment of the metabolic flexibility in the use of this domain leads to the indicated definitions of redundancy and robustness.

Thus, with the availability of complete genetic sequences, the fundamental questions regarding the metabolic repertoire of a cell have been answered, and metabolic engineering techniques can be applied to these completely sequenced organisms to generate complete models of single cells. This opens the possibility to further improve industrial processes and, in a broader sense, broaden the understanding of cellular metabolic physiology. The outline of the process that must integrate microbial genomics, metabolic biochemistry, strain specific information, and methods of systems science are outlined in Figure 2. How the process takes place is illustrated via the following example.

Table II. Questions that can be addressed using flux-balance analysis.

Question	Objective	Reference
What are the biochemical production capabilities?	Maximize metabolite product	Varma, Boesch, & Palsson, 1993
What is the maximal growth rate and biomass yield?	Maximize growth rate	Varma & Palsson, 1993; Varma & Palsson, 1994b
How efficiently can metabolism channel metabolites through the network?	Minimize the Euclidean norm	Bonarius et al., 1996
How energetically efficient can metabolism operate?	Minimize ATP production or minimize nutrient uptake	Majewski & Domach, 1990; Savinell & Palsson, 1992; Fell & Small, 1986
What is the tradeoff between biomass production and metabolite overproduction?	Maximize biomass production for a given metabolite production	Varma et al., 1993

EXAMPLE: FLEXIBILITY IN *ESCHERICHIA COLI* CENTRAL METABOLISM

Escherichia coli is arguably the best-studied microorganism. Additionally, the complete genetic sequence for the K-12 strain has recently been established (TIGR-Web Site, 1997). With such extensive knowledge regarding *E. coli*, it has served as a model organism for many studies. A genomically complete stoichiometric model for *E. coli* has been constructed. The model consists of 594 reactions and transport process that involve 334 metabolites. Using this model, the flexibility that *E. coli* has in metabolic flux distributions was examined during growth on glucose. Metabolic flexibility is the manifestation of two principal properties: redundancy and robustness (Fig. 1).

Figure 3A shows the metabolic flux distribution for the complete gene set present in *E. coli* for maximal biomass yield on glucose. The computed value for biomass yield (0.489 g DW/g glucose) compares quantitatively with experimental data. The ability of *E. coli* to respond to the loss of an enzymatic function (through gene mutation or inhibition of activity) can be assessed by removing a gene from the basic gene set. Figures 3B and C show the optimal metabolic flux distribution when the *sucA* gene or the *sdh* gene is removed from the basic gene set, respectively. The *sucA* gene codes for an essential component of the 2-oxoglutarate dehydrogenase complex, and the *sdh* gene codes for the succinate dehydrogenase enzyme.

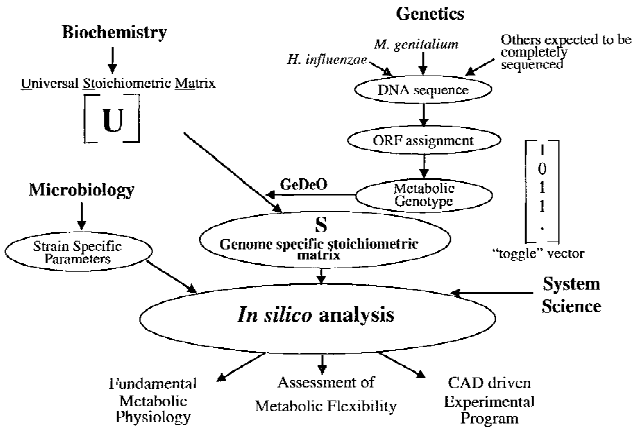


Figure 2. A schematic representing the process of formulating genomically metabolic models for microorganisms.

A mutant defective in the *sdh* gene product has been shown to be able to grow on glucose minimal medium (Creaghan & Guest, 1978). The in silico predicted metabolic flux distribution for optimal biomass synthesis for such a mutant is shown in Figure 3B. The energy needs of this mutant are greatly increased, and for optimal biomass synthesis, the pentose phosphate pathway must support large fluxes to generate redox potential to be used in energy generation.

Figure 3C shows the in silico metabolic flux distribution for maximal biomass synthesis in a *sucA* mutant. The in silico analysis shows that the ability of the *sucA* mutant to synthesize biomass is nearly equivalent to the ability of the complete gene set. Experiment results have shown that mutants in the *sucA* gene are able to grow anaerobically on glucose, but unable to grow aerobically. However, revertants arise that inactivate the *sdh* gene product and are able to grow aerobically on glucose minimal medium (Creaghan & Guest, 1978). Figure 3D shows the in silico predicted metabolic flux distribution for this *sdh sucA* double mutant.

This example shows the stoichiometric redundancy in central metabolism under the conditions considered. The metabolic network has the stoichiometric flexibility to redistribute its metabolic fluxes with remarkably little change in its ability to support biomass synthesis, even if faced with the loss of key enzymes.

The other key metabolic property that leads to metabolic flexibility is robustness. Robustness can be defined as the ability of the metabolic network to adjust to decreased fluxes through a particular enzyme without significant changes in overall metabolic function (Fig. 1). Figure 4 shows how the biomass yield will be effected by decreasing the activity of the succinate dehydrogenase enzyme. The results show how the activity of this enzyme can be significantly reduced without affecting the organisms ability to support biomass synthesis. The results also show how the robustness is decreased and flexibility is lost once the basic gene set is changed.

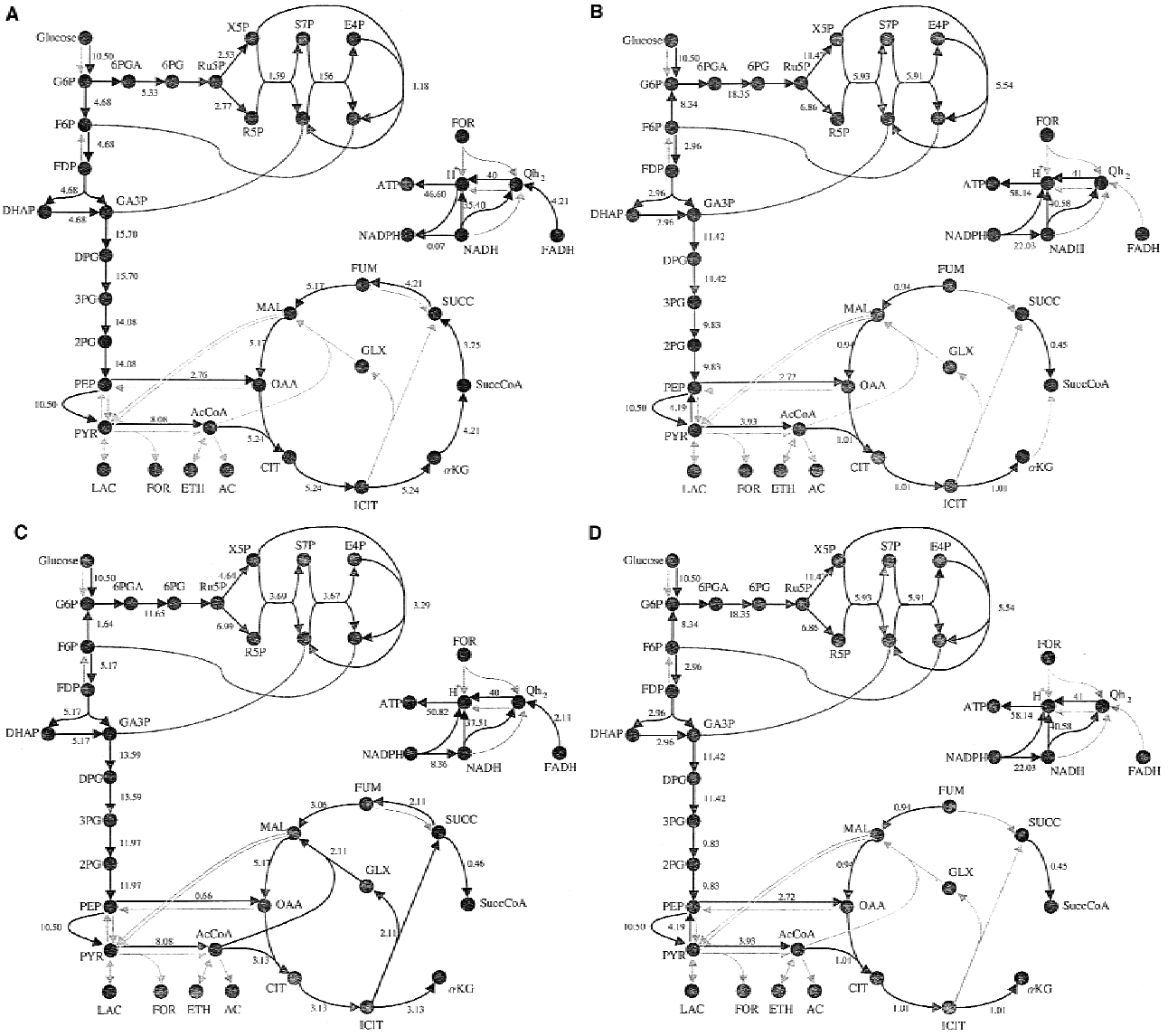


Figure 3. Re-routing of metabolic fluxes when enzyme function is lost. **A.** Flux distribution for the basic gene set. Biomass yield is 0.489 g DW/g glucose. **B.** Flux distribution for maximal biomass yield for *sdh* mutant. Biomass yield is 0.476 g DW/g glucose. **C.** Flux distribution for maximal biomass yield for *sucA* mutant. Biomass yield is 0.489 g DW/g glucose. **D.** Flux distribution for *sdh sucA* double mutant. Biomass yield is 0.476 g DW/g glucose. The solid lines represent enzymes that are being utilized with the corresponding flux value noted. The fluxes are relative to the glucose uptake rate (10.5 mmol glucose/h/g DW) (Varma & Palsson, 1994b). The gray lines represent enzymes that are not being utilized.

THE FUTURE OF FLUX-BALANCE ANALYSIS

As more organisms become fully sequenced and all of the components of their metabolic machinery are characterized, the utilization of FBA for organism-specific analysis and the determination of metabolic phenotypes is sure to grow. This capability will create the need for organized data compilation so as to accurately determine the net transport of metabolites out of the defined metabolic system in organism specific cases, that is defining the **b** vector. Uptake rates, secretion rates, and other related cellular characteristics will need to be determined and compiled into online organism-specific databases providing all of the necessary information to perform FBA for a particular organism. Together, FBA and genomics will aid in the understanding of the genotype-phenotype relationship and elucidation of the underlying interconnectivity which determine this relationship. Perhaps, the greatest asset of FBA is its sole reliance on stoichiometric characteristics and successful neglect of kinetic parameters of the system in question. However, the stoichiometry of the metabolic network does not uniquely specify the fluxes through the cell's pathways, because there is a plurality of feasible solutions in the under-determined cases. The particular flux distribution chosen by the cells is a function of regulatory mechanisms within the cell that determine the kinetic characteristics of cellular enzymes as well as enzyme expression. Thus, the flux-balance constraints placed upon a metabolic system by its stoichiometry define a wider limit of metabolic behavior. These stoichio-

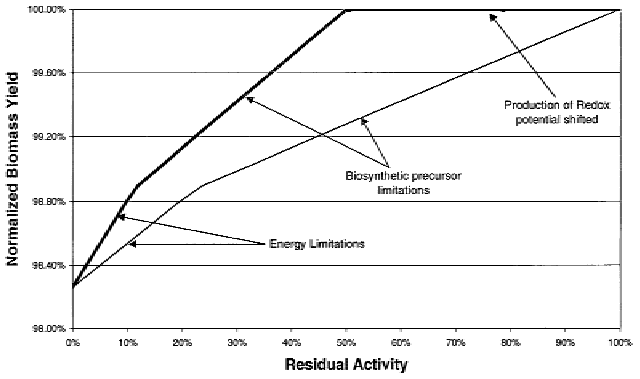


Figure 4. Robustness in metabolism when enzyme function is attenuated. Maximal biomass yields for growth on glucose during restricted flux through the sdh gene product for the basic gene set (thick line), and for the sucA mutant (thin line). The allowable flux through the enzyme was reduced from the value that allowed for maximal biomass synthesis.

metrically set limits are further narrowed by the kinetic and regulatory function of metabolic enzymes. In some cases, computational or in silico results of FBA are in disagreement with experimental data. Our experience shows that these discrepancies can often be accounted for when known regulatory loops are considered. Thus, a future challenge to FBA will be to incorporate methods that account for known regulation of gene expression.

''Cells obey the laws of [physics and] chemistry'' (Watson, 1972), to which we humbly add, that includes conservation of mass as well. Based on this fundamental rationale, FBA has been developed under the assumption that metabolic transients are sufficiently rapid to allow the imposition of the steady-state assumption. This line of reasoning leads to the development of the basic flux-balance equations.

Given the limited amount of information that is needed for FBA and the growing genomic databases from which some of this information can be derived, it seems quite likely that this approach to the analysis of metabolism will expand in scope, and grow in its application to metabolic networks of practical importance.

MODAL ANALYSIS

Flux-balance analysis will define the boundaries of possible flux distributions achievable with a defined basic gene set. However, FBA cannot be used to determine the dynamics of metabolic behavior within these boundaries. Each biochemical event occurring within a living cell proceeds at a characteristic rate. Each of the individual events contributes to the overall response. The interaction of all the events leads to systemic motion on several time scales, which are usually distributed over several orders of magnitude. An approach developed in the mid-1980s allows for the decomposition of dynamic systems into dynamically independent systems (or modes) that move on qualitatively different time scales. This method, known as modal analysis, allows for the incorporation of order of magnitude dynamic information. This technique has been applied to dynamic models of metabolism (Palsson, Joshi, & Ozturk, 1987), kinetic models of enzyme action (Palsson, 1987; Palsson, Jamier, & Lightfoot, 1984; Palsson & Lightfoot, 1984; Palsson, Palsson, & Lightfoot, 1985), complex bacterial growth models (Joshi & Palsson, 1987; Palsson & Joshi, 1987), and plasmid replication (Keasling & Palsson, 1989a & b). As evidenced by numerous examples, modal analysis provides us with a useful method for the analysis and conceptualization of metabolic dynamics.

Unexpectedly, the dynamically independent modes proved to contain some very insightful information. When applied to the Michaelis-Menten mechanism, modal analysis demonstrated the quasi-equilibrium and quasi-steady-state assumptions (Palsson, 1987; Palsson & Lightfoot, 1984). When applied to a complex growth model of *E. coli*, modal analysis lead to a dramatic model simplification by reducing the description of the growth process to the growth of the three pools of macromolecules (Joshi & Palsson, 1987; Palsson & Joshi, 1987).

Perhaps most importantly, modal analysis of red blood cell metabolism (Joshi & Palsson, 1989; Palsson et al., 1987) lead to the discovery of metabolic pools that had direct metabolic physiological significance. These pools included: (1) the adenosine moiety (carrier of phosphate bonds), (2) the high energy phosphate bond on the adenosines, (3) the high energy bonds on the glycolytic intermediates, and (4) the important 2,3 diphosphoglycerol regulator of hemoglobin binding to oxygen. Each pool moved on one or more time scales. Pool 1 moved on a 2.5 d time scale, pools 2 and 3 on a 45 min time scale, and pool 4 moved on a 12 h time scale. All have physiological significance and correspond to experimental observations; the overall energy charge moves on the order of 1 h, oxygen binding to hemoglobin on the order of 12 h, and the loss of the adenosine carrier moves on the order of a few days.

Thus, modal analysis reveals an interesting (metabolic) function vs. (temporal) structure relationship in red blood cell metabolism. The results of these analyses suggest that only approximate kinetic information with precise stoichiometry suffices to set the elements of dynamic pool formation. Therefore, it might be possible to combine FBA and approximate enzyme kinetic information to analyze the basic modalities of metabolic dynamics.

CONCLUSIONS

Biotechnology is about to enter a new era. An era in which bioinformatics will lead to the discovery of biological ''rules'' and ''principles'' upon which design of biological systems will rely. To accomplish this goal, we must complete the analysis of the hierarchical genetics-to-physiology relationship. This accomplishment will rely on the employment of engineering methods for operation and design of integrated systems, of which cells are spectacularly elabo-

rate examples. In the process of applying these methods, key features of biological systems, such as physiological behavior and change through evolution, must be given primary consideration.

Metabolic engineering, in particular, is likely to be at the forefront of these developments. This expectation is based on the fact that the fundamentals of metabolism are well known, complete genetic information is now available for microbial genomes, the capability to change the genomic content of bacteria almost at will is at hand, and the methods for metabolic analysis are being developed. The ability to design bacteria based on biological principles and alteration of underlying genetic components is likely to lead to the realization of the potential that metabolic engineering holds.

Many of the ideas and developments presented here were a result of Bernhard O. Palsson's stay at the Biotechnology Department at the Danish Technical University in 1996. Many thanks to Professors John Villadsen and Jens Nielsen at the DTU, and to Ramprasad Ramakrishna and Christophe Schilling for preparing the figures and proofreading the manuscript.

References

Bailey, J. E. 1991. Toward a science of metabolic engineering. Science **252**: 1668–1675.

Bonarius, H. P. J., Schmid, G., Tramper, J. 1997. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. Trends Biotechnol. **15**: 308–314.

Bonarius, P. J. B., Hatzimanikatis, V., Meesters, K. P. H., de Gooijer, C. D., Schmid, G., Tramper, J. 1996. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. Biotechnol. Bioeng. **50**: 299–318.

Botstein, D., Cherry, J. M. 1997. Molecular linguistics: Extracting information from gene and protein sequences. Proc. Natl. Acad. Sci. **94**: 5506–5507.

Bult, C. J., White, O., Olsen, G. J. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science **273**: 1058–1073.

Creaghan, I. T., & Guest, J. R. 1978. Succinate dehydrogenase-dependent nutritional requirement for succinate in mutants of *Escherichia coli* K-12. J. Gen. Microbiol. **107**: 1–13.

Delgado, J., Liao, J. C. 1997. Inverse flux analysis for reduction of acetate excretion in *Escherichia coli*. Biotechnology Progress. **13(4)**: 361–367.

Fell, D.A., Small, J.R. 1986. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochem. J. **238(3)**: 781–786.

Fleischmann, R. D., Adams, M. D., White, O. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**: 496–512.

Fraser, C. M., Gocayne, J. D., White, O. 1995. The minimal gene complement of *Mycoplasma genitalium.* Science **270**: 397–403.

Ingraham, J. L., Maalce, O., Neidhardt, F. C. 1983. Growth of the bacterial cell. Sinauer Associates, Inc., Sunderland, MA.

Jorgensen, H., Nielsen, J., Villadsen, J. 1995. Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. Biotechnol. Bioeng. **46**: 117–131.

Joshi, A., Palsson, B. O. 1987. *Escherichia coli* growth dynamics: A three pool biochemically based description. Biotechnol. Bioeng. **31**: 102–116.

Joshi, A., Palsson, B. O. 1989. Metabolic dynamics in the human red cell. Part I - A comprehensive kinetic model. J. Theor. Biol. **141**: 515–528.

Karp, P. D. 1996a. Database links are a foundation for interoperability. Trends Biotechnol. **14**: 273–279.

Karp, P. D. 1996b. EcoCyc: An encyclopedia of *Escherichia coli* genes and metabolism. Nucl. Acids Res. **24**: 32–39.

Keasling, J. D., Palsson, B. O. 1989a. ColE1 plasmid replication: A simple kinetic description from a structured model. J. Theor. Biol. **141**: 447–461.

Keasling, J. D., Palsson, B. O. 1989b. On the kinetics of plasmid replication. J. Theor. Biol. **136**: 487–492.

Koonin, E. V., Mushegian, A. R., Rudd, K. E. 1996. Sequencing and analysis of bacterial genomes. Curr. Biol. **6**: 404–416.

Majewski, R. A., Domach, M. M. 1990. Simple constrained optimization view of acetate overflow in *E. coli.* Biotechnol. Bioeng. **35**: 732–738.

Neidhardt, F. C., Ingraham, J. L., Schaechter, M. 1990. Physiology of the bacterial cell. Sunderland, MA, Sinauer Associates, Inc.

Ouzounis, C., Casari, G. 1996. Computational comparisons of model genomes. Trends Biotechnol. **14**: 280–285.

Palsson, B. O. 1987. On the dynamics of the irreversible Michaelis-Menten reaction mechanism. Chem. Eng. Sci. **42**: 447–458.

Palsson, B. O., Jamier, R., Lightfoot, E. N. 1984. Mathematical modeling of dynamics and control in metabolic networks. Part II. Simple dimeric enzymes. J. Theor. Biol. **111**: 303–321.

Palsson, B. O., Joshi, A. 1987. On the dynamic order of structured *Escherichia coli* growth models. Biotechnol. Bioeng. **29**: 789–792.

Palsson, B. O., Joshi, A., Ozturk, S. S. 1987. Reducing complexity in metabolic networks: making metabolic meshes manageable. Fed. Proc. **46**: 2485–2489.

Palsson, B. O., Lightfoot, E. N. 1984. Mathematical modeling of dynamics and control in metabolic networks. Part I. On Michaelis-Menten Kinetics. J. Theor. Biol. **111**: 273–302.

Palsson, B. O., Palsson, H., Lightfoot, E. N. 1985. Mathematical modeling of dynamics and control in metabolic networks. Part III. Linear reaction sequences. J. Theor. Biol. **113**: 231–259.

Papoutsakis, E., Meyer, C. 1985a. Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations. Biotechnol. Bioeng. **27**: 50–66.

Papoutsakis, E., Meyer, C. 1985b. Fermentation equations for propionic-acid bacteria and production of assorted oxychemicals from various sugars. Biotechnol. Bioeng. **27**: 67–80.

Papoutsakis, E. T. 1984. Equations and calculations for fermentations of butyric acid bacteria. Biotechnol. Bioeng. **26**: 174–187.

Pons, A., Dussap, C., Pequignot, C., Gros, J. 1996. Metabolic flux distribution in *Corynebacterium melassecola* ATCC 17965 for various carbon sources. Biotechnol. Bioeng. **51**: 177–189.

Sauer, U., Hatzimanikatis, V., Bailey, J., Hochuli, M., Szyperski, T., Wuthrich, K. 1997. Metabolic fluxes in riboflavin-producing *Bacillis subtilis.* Nature Biotechnol. **15**: 448–452.

Savinell, J. M., Palsson, B. O. 1992. Network analysis of intermediary metabolism using linear optimization: I. Development of mathematical formalism. J. Theor. Biol. **154**: 421–454.

Stephanopoulos. 1994. Metabolic engineering. Curr. Opin. Biotechnol. **5**: 196–200.

Stephanopoulos, G., Sinskey, A. J. 1993. Metabolic engineering-methodologies and future prospects. Tibtech 11 (September). **7(9)**: 392–396.

Stephanopoulos, G., Vallino, J. 1991. Network rigidity and metabolic engineering in metabolite overproduction. Science **252**: 1675–1681.

TIGR-Web Site. 1997. TIGR microbial database. http:\\www.tigr.org.

Vallino, J., Stephanopoulos, G. 1993. Metabolic flux distributions in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol. Bioeng. **41**: 633–646.

Varma, A., Boesch, B. W., Palsson, B. O. 1993. Biochemical production capabilities of *Escherichia coli.* Biotechnol. Bioeng. **42**: 59–73.

Varma, A., Palsson, B. O. 1993. Metabolic capabilities of *Escherichia coli*. II. Optimal growth patterns. J. Theor. Biol. **165**: 503–522.

Varma, A., Palsson, B. O. 1994a. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technol.* **12**: 994–998.

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

Watson, J. D. 1972. Molecular biology of the cell, 1st edition. Garland Publishing Inc., New York.

Zupke, C., Stephanopoulos, G. 1994. Modeling of isotope distributions and intracellular fluxes in metabolis networks using atom mapping matrices. Biotechnol. Prog. **10**: 489–498.

Zupke, C., Stephanopoulos, G. 1995. Intracellular flux analysis in hybridomas using mass balances and in vitro ^{13}C NMR. Biotechnol. Bioeng. **45**: 292–303.