Identifying Constraints that Govern Cell Behavior: A Key to Converting Conceptual to Computational Models in Biology?

Markus W. Covert, Iman Famili, Bernhard O. Palsson

Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093; telephone: 858-534-5668; fax: 858-882-3120; e-mail: palsson@ucsd.edu

Pubilshed online 24 November 2003 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/bit.10849

Abstract: Cells must abide by a number of constraints. The environmental constraints of cellular behavior and physicochemical limitations affect cellular processes. To regulate and adapt their functions, cells impose constraints on themselves. Enumerating, understanding, and applying these constraints leads to a constraint-based modeling formalism that has been helpful in converting conceptual models to computational models in biology. The continued success of the constraint-based approach depends upon identification and incorporation of new constraints to more accurately define cellular capabilities. This review considers constraints in terms of environmental, physicochemical, and self-imposed regulatory and evolutionary constraints with the purpose of refining current constraint-based models of cell phenotype.
Keywords: systems biology; flux balance analysis (FBA); extreme pathway analysis; constraints; computational modeling

INTRODUCTION

The complex composition of a biological system requires the use of computational tools to describe its integrated function. As a result, more biologists are turning to engineers, physicists, and mathematicians, who in turn are scrambling to learn biological fundamentals. Such cross-disciplinary fertilization has led to important studies on regulation of galactose utilization in yeast (Ideker et al., 2001), control of the IκB–NF–κB signaling module (Hoffmann et al., 2002), and mammalian cell cycle regulation (Qu et al., 2003), among others. Taken together, these developments signal an important shift in biology from *conceptual modeling* to *computational modeling*. Conceptual models describe a system in qualitative terms, whereas computational models can quantitatively simulate systemic properties to analyze, interpret, and predict cell behavior. Thanks to the high-throughput generation of ''omics'' data (Blaine-Metting and Romine, 1997), biologists find themselves well positioned to reconstruct fairly complicated conceptual models of metabolic, regulatory, and signaling networks (in order of difficulty) (Davidson et al., 2002a, 2002b; Karp et al., 2002; Salgado et al., 2001), culminating in the development of databases such as KEGG and MetaCyc (Kanehisa and Goto, 2000; Karp et al., 2000).

One current challenge in systems biology is to translate these conceptual models into genome-scale computational models. It is clear that the complexity of biological systems, the difficulty of obtaining kinetic parameters, and the enormous generation of data will require the development of new analytical methods in *in silico* biology (Bailey, 2001; Kitano, 2002) (Fig. 1). The constraint-based approach (Palsson, 2000) to the deduction of phenotype from the genotype and environment directly addresses these challenges. In the constraint-based approach to analyzing metabolic networks, all possible behaviors of a system (e.g., flux distributions through the metabolic network) are considered, as shown in Figure 2. Although cellular behavior may not be completely specified at this point, it is known that cells behave in ways that reflect the constraints imposed by their environments, by physicochemical laws, and by themselves. By successively imposing constraints on conceptual models, as illustrated in Figure 2, the allowable range for each flux in the network is reduced dramatically. The problem of modeling complex biological systems shifts from experimental determination of kinetic and other fundamental parameters—as mentioned, currently an intractable problem—to continued identification of constraints that allow a more specific description of the system—a challenge that is successfully being addressed, as discussed in what follows.

Current constraint-based computational models have focused on microbial organisms. These models are at the ''genome-scale''—referring simply to the amount of genes accounted for in the model, which has now reached approximately 30% to 35% of the annotated genes in an organism (Price et al., 2003). Genome-scale models have focused primarily on metabolism and associated transcriptional

Correspondence to: B. Palsson
Contract grant sponsors: National Science Foundation; National Institutes of Health
Contract grant numbers: BES 01-20363; GM 57089

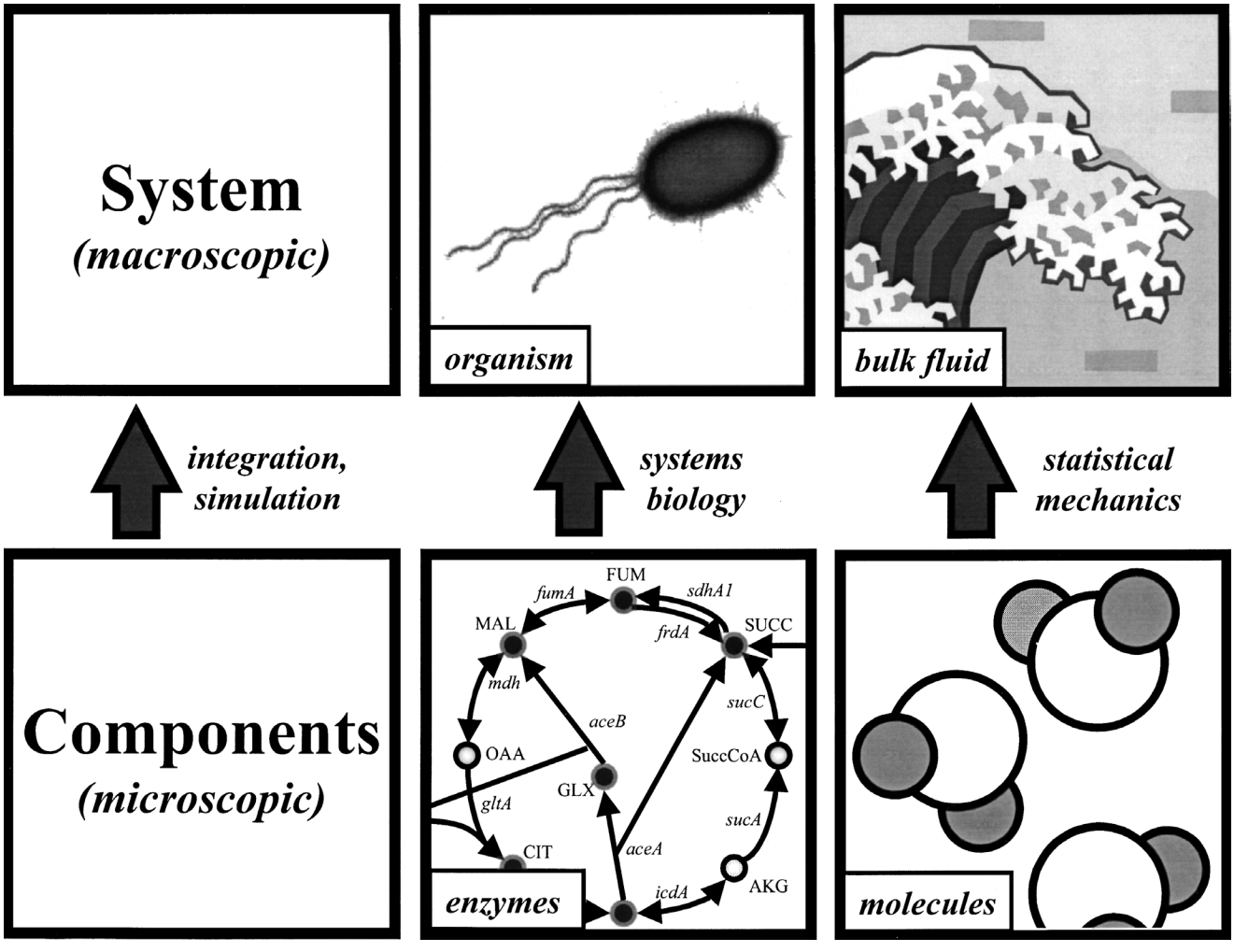


Figure 1. The shift from studying components to studying systems requires the use of computational tools to integrate conceptual data and simulate systemic behavior. The importance of computation has become apparent in the biological sciences, wherein generation of experimental data far outpaces efforts to reconcile these data in terms of a comprehensive model. New mathematical approaches will be required to describe such systems. Such a challenge is not new; the development of statistical mechanics originated as an attempt to integrate the known chemical properties of molecules to simulate the properties of a bulk fluid. *Escherichia coli* image courtesy of www.denniskunkel.com.

regulation, but are aimed at a complete representation of an organism and have already been used to simulate cell behavior under a variety of conditions (Reed and Palsson, 2003). In addition, because these models attempt to capture systems-level behaviors as completely as possible, they are instrumental in identifying and characterizing emergent properties of biological networks, where an observed behavior has been difficult or even impossible to interpret from the cellular ''parts list'' alone (Papin et al., 2003; Price et al., 2003).

The success of the constraint-based approach depends upon identification and incorporation of new constraints to define the cellular capabilities more accurately. This review considers constraints in terms of environmental, physicochemical, and self-imposed regulatory and evolutionary constraints with the purpose of refining current constraint-based models of cell phenotype. Some of the constraints may seem intuitive or basic; however, we aim to illustrate how their consideration leads to nonintuitive modeling consequences (Fig. 3, Table I).

ENVIRONMENTAL CONSTRAINTS

The constraints imposed on cells by their environments—both external and internal—have a major influence on cell behavior. External environments impose constraints on cells in terms of nutrients, physical factors, and neighboring influences. Whether a cell can grow in a given environment depends in part on its ability to obtain or synthesize all necessary biomass components (Neidhardt et al., 1990). The presence or absence of necessary compounds thus represents an environmental constraint on the cell. As an example, development of a ''minimal gene set'' to sustain life must be kept in the context of environmental constraints, as the minimal gene set for life in a complex medium would differ significantly from that required for life in glucose minimal

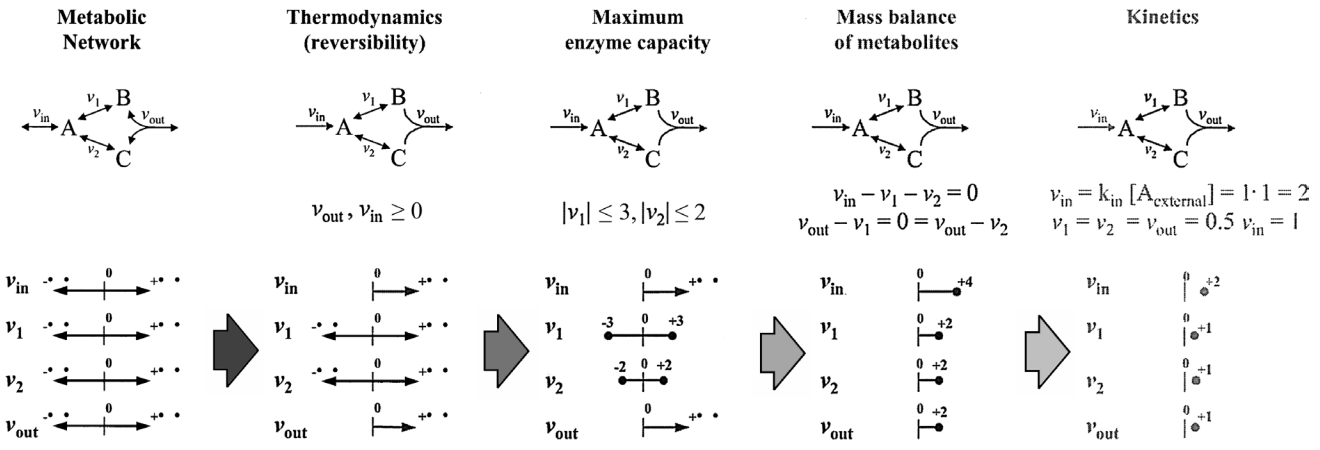


Figure 2. The constraint-based approach applied to metabolic networks. A small network with only two chemical reactions (flux v_1: A→B; flux v_2: A→C) and two transport processes (metabolite A enters the cell with flux v_{in} and B and C exit together via flux v_{out}) is depicted together with the allowable range of each reaction/transport flux. Initially, the flux ranges are unbounded. By incorporating characteristics of the system, such as reaction thermodynamics, maximum enzyme turnover rates, and metabolite mass balance, in terms of constraints, the allowable ranges are reduced significantly. If the system is characterized completely, the ranges are reduced to a single point.

medium (Burgard et al., 2001; Koonin, 2000). The importance of the nutrient constraints imposed by the external environment also underscores the importance of defined media in mathematical simulation of cell behavior. Without adequate knowledge of the nutritional content of the external environment, significant constraints must be ignored or grossly approximated, resulting in incorrect or misleading predictions of cell behavior. The development of high-throughput phenotyping technologies has addressed the inadequacies of studies on undefined media, enabling characterization of the environmental effects on organism growth under thousands of well-defined conditions (Bochner et al., 2001). Physical characteristics of the external environment, such as temperature, pressure, pH, and exposure to light or water, can also limit possible cell behavior and survival. Physical environmental constraints have been used to investigate the possibility of life on Mars (Cockell et al., 2000).

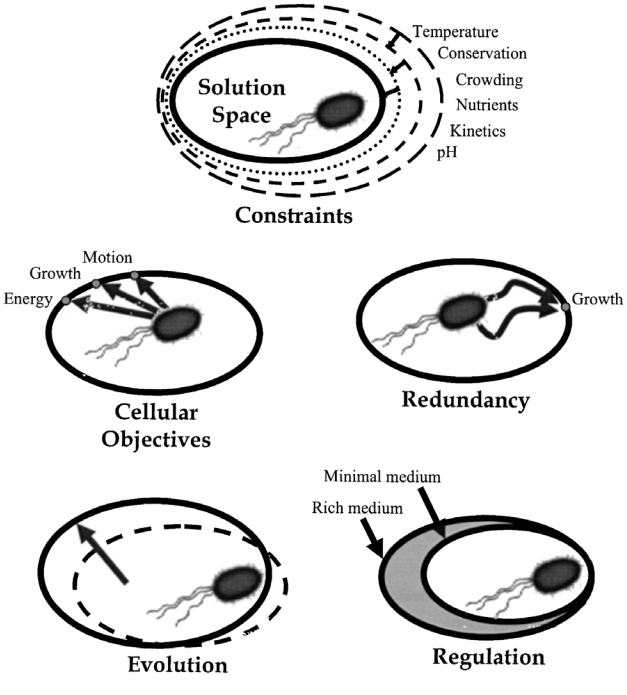
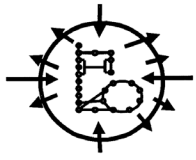
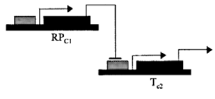


Figure 3. Perspectives generated using the constraint-based approach. The phenotypic potential (set of all behaviors that can be exhibited by a biological system) can be represented as a solution space (shown as an ellipse for simplicity). The identification of constraints that the cell must obey, such as conservation of mass and energy, survivable temperature ranges, etc., further constrain this space (e.g., solutions that violate conservation laws are excluded from the space). Depending on the objective of the cell, it may exhibit a variety of behaviors. Although the cell is constrained by its solution space at one point in time, it is also able to control its behavior over time via processes such as evolution and regulation. *Escherichia coli*: image courtesy of www.denniskunkel.com

The environmental conditions experienced by a cell generally change over time. They may change by the presence of new harmful products (such as the cell's own waste), depletion of nutrients (by the cell itself or by competitors), or other dynamic forces. A tightly packed cellular community necessitates competition for, or exchange of, nutrients and adhesion sites, evolving mechanisms to survive toxin exposure or to move toward scarce nutrients, but also raises the possibility of obtaining new cellular capabilities via gene transfer, or of cooperation and specialization, such as in a tissue, where signaling molecules allow cells to communicate and cooperate (DeLisa et al., 2001). To account for such interactions in a model, the cellular community must therefore be accurately represented (Tsuchiya et al., 1966). Forces such as fluid flow (wind on a plant leaf, juices through the intestines) necessitate that cells adhere to their chosen environments or develop systems that are resilient to changing environments (sporulation, broad substrate utilization). The cell is therefore constrained to the development of biological functions that allow it to thrive in a dynamic environment.

The intracellular environment of a cell also imposes constraints on cellular behavior, notably in terms of its internal components and the physical properties of its interior. Cells are obviously limited by the biochemical components of

Table I. Some biological constraints with important modeling consequences. The constraints listed at the left may seem intuitive, but the consideration of such constraints in modeling biological systems can be vital in terms of correct prediction and simulation.

Biological Constraint		Modeling Consequence
External environment: Cell function is limited by nutrient availability		• "Minimal gene set" is environment dependent • Undefined media of limited use in modeling
Internal environment: Cell "toolbox" determines function		• Importance of "omics" data in modeling
Conservation: Mass is not created or destroyed in the cell		• Cell function can be simulated quantitatively with few parameters
Regulation: Signaling molecules present in nanomolar concentrations		• Stochastic component to regulation modeling
Evolution: Finite number of sites for potential mutation		• Evolution is more limited by environment than by "random" mutation event

which they are comprised; the glycolytic genes must not only be located in the genome but also must be expressed as functional proteins for glycolysis to occur in an organism. The components may be thought of as a ''toolbox'' that the cell requires to use the resources found in the environment to perform necessary functions (growth, signaling, chemotaxis, etc.). This toolbox constantly changes over time, either enabling or limiting the function of the cell. For example, the maximum transport rate of a particular cell moiety will be determined in part by the number of transport proteins specific to that moiety located in the cell membrane. The total number of components that can be contained in the cell are limited by a generalized bounding constraint, cell volume (i.e., the toolbox is also quite small). The last several decades of biological research has focused on identifying cellular components, culminating in the development of high-throughput methods to study the genome (Gaasterland and Oprea, 2001), transcriptome (Devaux et al., 2001), proteome (Naaby-Hansen et al., 2001), and metabolome (Raamsdonk et al., 2001) of organisms under various conditions. Because cell function depends on the action and interaction of various components, the ''omics'' data are of fundamental importance in the effort to model cell behavior.

Physical factors of the cellular interior also impose constraints on the cell. Pictorial models of the interior of *E. coli* depict a crowded, tightly packed, nonhomogeneous cytoplasm (Ellis, 2001; Goodsell, 1993). Such a dense environment has a constraining effect on solute and macromolecular diffusion. One way in which cells may overcome diffusion-related constraints in a crowded environment is via compartmentalization of major cell processes and metabolic channeling (Verkman, 2002). Furthermore, the crowded internal environment of the cell creates an osmotic pressure in relation to the often aqueous external environment that must be balanced. Cells achieve this balance by exchanging molecules with the external environment. The balance of osmotic pressure must be achieved while maintaining an electroneutral environment on both sides of the membrane. Osmotic and electroneutrality constraints can affect the total volume of the cell, and the need to meet these constraints imposes significant energy demands on the cell.

PHYSICOCHEMICAL CONSTRAINTS

Physicochemical laws place demanding constraints on cellular behavior (see examples in Appendix). Cells balance

mass and energy, conform to the laws of thermodynamics and kinetics, and operate under limited enzyme turnover rates and activity of gene products. Physicochemical constraints are generally considered to be ''hard'' constraints and are thought to remain unchanged.

Mass and energy are never created or destroyed in the cell. Elements entering the cell are either incorporated into biosynthetic material for cell growth and replication, utilized to generate energy required for cellular functions, or secreted into the extracellular environment. Excess biochemical byproducts that remain internal to the cell may accumulate over time and result in cellular toxicity and death. Complex systems have evolved to sense and respond to imbalances of mass within the cell. Energy imbalance also has detrimental consequences. Eukaryotic loss of mitochondrial function to generate energy prohibits the cell from driving cellular functions and causes death (Scheffler, 1999). The balance of mass and energy thus poses critical constraints on how the cell must allocate its resources. Mass balance of reactions also imposes stoichiometric constraints on the network. The stoichiometric coefficients of any biochemical reaction are such that the number of elements and charge is conserved in a conversion. Stoichiometric constraints impose restrictions on the network that, apart from the nature of kinetics, define what combinations of species must be present or absent in a steady state (Feinberg, 1987). For example, for a set of compounds with zero concentrations, it is possible to determine which reactions have zero rates (i.e., ''switched off'') and which have positive rates (i.e., ''switched on'') (Feinberg, 1987). The requirement of mass balance exerts such a strong constraint on metabolic network function that flux balance analysis requires virtually only these constraints, with only a handful of strain-specific parameters, for detailed qualitative simulations (Edwards et al., 2002).

The thermodynamics of internal reactions can significantly affect the overall capability and phenotypic properties of the cell. The direction in which reactions proceed is a function of energetic properties of the biochemical conversions and may determine the ability of the cell to reach diverse metabolic states in a variety of internal and external conditions. Furthermore, chemical turnover of uncatalyzed reactions is often very slow. In the presence of enzymes, however, substrates are quickly converted into products. Enzymes also provide means for fast-responding control mechanisms. Feedback and feedforward control mechanisms, including inhibition and activation, and effect of pH and temperature on enzyme activity can influence the rate of biochemical reactions, either linearly or nonlinearly (Bailey, 1998; Desai et al., 1999; Lee et al., 1999; Savageau, 1998). Kinetic constraints are especially important in cells with little or no other means of regulation. Upon maturation, the red blood cell loses its DNA and, consequently, lacks any means of transcriptional regulation. The sole form of regulation in the red cell is thus kinetic regulation, which makes the red cell a suitable model for studying kinetics (Jamshidi et al., 2001).

The maximum throughput or enzyme capacity of biochemical reactions can also force the cell to exhibit more limited behaviors than otherwise. The enzyme turnover rate of reactions in biochemical pathways can impose ''bottleneck'' constraints on the maximum allowable flow achieved in a pathway (Bailey, 2001). Such bottlenecks have been identified and analyzed using metabolic flux analysis and metabolic control analysis (Stephanopoulos et al., 1998). Metabolic engineering of microbial organisms has centered around the premise of removing such bottleneck constraints and achieving a higher production rate of desirable compounds. Such efforts have been successful in a number of cases (Stephanopoulos et al., 1998).

The balance of osmotic pressure and maintenance of electroneutrality also impose constraints on cells. For example, the constraints of osmotic pressure and charged molecule requirements (e.g., nutrients) were found to be driving forces in evolution of cell walls in bacteria (Koch, 2000). Although cell volume regulation is relatively well characterized (Hallows and Knauf, 1994), it is routinely ignored by most models. The consequences of such constraints have been studied in detail in the human red blood cell, because such cells are relatively simple, lacking chromosomes, and are therefore not capable of replication (Joshi and Palsson, 1989; Werner and Heinrich, 1985). Initial assessment of the importance of these constraints using metabolic control analysis shows that they represent dominant regulatory effects (Lee and Palsson, 1991). Clearly, much work is required to ascertain the importance of these constraints and implement them in mathematical models.

SELF-IMPOSED CONSTRAINTS

We have previously discussed the environmental and physicochemical constraints imposed on cells, which are beyond the cells' direct control. To respond to these constraints and still carry out their desired functions (e.g., growth, nitrogen/carbon dioxide fixation, development), cells must impose constraints upon themselves to direct their behavior, selecting the ''best'' or most suitable option from a range of allowable alternatives. Self-imposed constraints are different from other constraints because they respond to—and often change—internal or external environments. Unlike physicochemical constraints, they are time-dependent. Such adaptive constraints may entail regulation in the short term and evolution over longer time scales.

Evolutionary Constraints

Although cells are constrained by the contents of their genome, they are able to change their genome sequence via evolution. The evolutionary process is associated with certain constraints. For example, *E. coli*'s overall error rate for DNA replication is between 1 in 10^{10} and 1 in 10^{11} basepairs (Neidhardt et al., 1990); this error rate may change some-

what in stressful conditions to increase the rate of beneficial mutation and adaptation. Evolution is also constrained by the number of possible mutation sites, as can be illustrated by rolling a six-sided die. With one roll, the probability of rolling a six is only one in six, or 16.7%. However, if the die is rolled 36 times, the probability of rolling at least one six becomes 99.9%. Similarly, genome size and growth rate influence the probability of a favorable mutation (see Appendix, section iii). Although most mutations are selected against, the number of mutations that occur in each generation can drive the occurrence of an evolved mutant to virtual inevitability over time (de Duve, 1996). Thus, after a period of only decades, bacteria, malarial parasites, insects, and weeds have been found to resist man-made antibiotics and toxins (de Duve, 1996). More quantitatively, *Escherichia coli* has been shown to overcome thymineless death by experimental evolution over 10,000 generations (de Crecy-Lagard et al., 2001), and also to exhibit parallel changes in expression of 59 genes after 20,000 generations of evolution on a glucose-limiting medium (Cooper et al., 2003). Consideration of evolutionary constraints also led to the prediction that *E. coli* would optimize its metabolic network for growth on a glycerol minimal medium, which was then demonstrated experimentally over 700 generations (Ibarra et al., 2002).

Regulatory Constraints

Beyond reconfiguring their genomes via mutation and adaptive evolution, organisms are able to control, to a certain extent, which genes are expressed, which proteins are present, and even the activity of proteins in cells. The incorporation of regulatory constraints in considerations of biological systems is vital (Liao and Oh, 1999). For example, recent studies of cellular metabolic networks (Jeong et al., 2000; Wagner and Fell, 2001) have characterized such networks as scale-free or ''small world.'' These studies were based on genomic constraints only (i.e., any gene located in the genome was considered a part of the network). However, the entire metabolic network of a cell is never completely expressed at a given time. For this reason, transcriptomic and proteomic studies are necessary to further constrain and thereby more accurately characterize these networks.

Furthermore, many molecules, notably those involved in the control of regulatory networks, are present and act at nanomolar concentrations in the cell (McAdams and Arkin, 1999). As a result of these low concentrations, there is a substantial amount of noise in gene expression, as has been recently demonstrated (Elowitz et al., 2002). Stochastic modeling approaches have been used to incorporate the constraints imposed by gene noise in models of the lambda phage (McAdams and Arkin, 1999), and experiments have demonstrated how such noisy systems may be controlled by negative feedback (Gardner and Collins, 2000). Recent work has quantified the efficacy of regulatory constraints in reducing metabolic solution spaces using extreme pathway analysis to illustrate the potential dominant effect of regulation on cell phenotype (Covert and Palsson, 2003).

APPLICATION OF CONSTRAINTS TO GENOME-SCALE MODELS

As mentioned earlier, constraint-based approaches have enabled the development of genome-scale models of microorganisms. Thus far, the constraints that have been incorporated into genome-scale simulative models of metabolism, such as those that exist for *E. coli*, *H. influenzae*, *H. pylori*, and *S. cerevisiae*, have been stoichiometric, thermodynamic, enzyme capacity, and energy balance constraints (Price et al., 2003a). Transcriptional regulatory constraints have also recently been added to enable combined simulation of regulatory and metabolic networks (Covert et al., 2001b; Covert and Palsson, 2002, 2003).

The constraint-based approach to modeling metabolic systems has been described by Covert et al. (2001b) and Palsson (2000). A metabolic network is reconstructed for a microorganism, yielding a set of all metabolic reactions available to the cell (Covert et al., 2001a). Once defined, this set of reactions represents the stoichiometric constraints on a cell. The reaction list also contains certain thermodynamic constraints, in the effective irreversibility of some reactions. A useful mathematical representation of all possible cell behaviors is one established geometrically as a solution space, which is effectively capped and reduced as constraints are incorporated (Fig. 3). We are then left with a smaller solution space having general properties that can be studied, or in which certain points (i.e., cellular behaviors) may be examined in more detail.

Pathway analysis, a method of relating the structure of a metabolic network in terms of an organism's overall metabolic capabilities, is based solely on identification of the aforementioned stoichiometric and thermodynamic constraints and requires no other parameters (Papin et al., in press). Using pathway analysis it is possible to determine, for example, possible network ''dead-ends'' (Schilling and Palsson, 2000; Schilling et al., 2002), key regulatory control points (Price et al., 2003b, and the level of redundancy in a metabolic network (Papin et al., 2002a, 2002b; Price et al., 2002).

Incorporation of enzyme capacity constraints enables flux balance analysis (FBA), which may be used to study specific flux distributions in more detail. FBA is associated with linear programming (Chvatal, 1983), wherein the maximum value of an objective function (e.g., synthesis of biomass precursors) is located in the solution space. Some notable recent work in FBA focused on identification and characterization of objective functions (Burgard and Maranas, 2003; Lee et al., 2000; Mahadevan et al., 2002; Segre et al., 2002). Metabolic flux analysis (MFA), closely related to FBA, attempts to constrain the space more completely by determining certain internal fluxes experimentally (Bonarius et al., 1998; Wiechert and de Graaf, 1996); however, it must be noted that such measurements do not represent ''hard''

constraints. It has been shown both computationally using extreme pathway analysis (Papin et al., 2002a; Price et al., 2002) and experimentally (Flores et al., 2002) that a high degree of redundancy exists in metabolic networks; organisms may use any of a number of routes to generate precursors or byproducts (see Fig. 3). The uses of MFA and FBA have been reviewed in detail elsewhere (Edwards et al., 2002; Stephanopoulos et al., 1998).

The elimination of systemically (i.e., not reaction by reaction) thermodynamically infeasible solutions yields an additional set of useful physicochemical constraints to determine allowable behaviors of biochemical reaction networks. Incorporation of the systemic thermodynamic constraints leads to energy balance analysis (EBA) (Beard and Liang, 2002). The importance of adding these constraints into genome-scale models needs to be evaluated. EBA, as presented by Beard et al. (2002), provides a foundation for constraints-based analysis of reaction free energies in large-scale biochemical systems and thus expands the scope of information available from constraints-based modeling of biochemical networks.

Thus far, the constraints noted in connection with genome-scale models have focused primarily on metabolic processes. Such models fail to accurately predict cell behavior when transcriptional regulatory processes have a dominant effect on phenotype (Edwards and Palsson, 2000; Varma and Palsson, 1994). To expand the scope and predictive capability of these models, a formalism for incorporation of regulatory constraints was developed (Covert et al., 2001b). Boolean rules to describe regulation and a time delay for protein synthesis/degradation are specified for each gene in the model. The resulting regulatory network determines the constraint imposed on every reaction in the metabolic network over time. Because the time constants that describe regulation and cell growth are significantly slower than those for metabolism (milliseconds to tens of seconds vs. tens of minutes to hours or days), metabolism may be considered to be at quasi-steady state for appropriately short time periods (Varma and Palsson, 1994). Simulation of cellular metabolism leads to calculation of flux changes and external metabolite/biomass concentrations; these outputs lead to a reinterpretation of the Boolean rules in the regulatory network and, possibly, to a change in the expressed metabolic network (Covert and Palsson, 2002). Integration of a metabolism simulation module (i.e., FBA) and a transcriptional regulation simulation module (i.e., evaluation of Boolean logic) is greatly facilitated by the recognition that regulatory events simply impose constraints.

This framework, which has been called regulatory flux balance analysis (rFBA), has been applied to a central metabolic/regulatory model of *E. coli*, accounting for 149 genes, including 16 regulatory proteins and 73 enzymes (Covert and Palsson, 2002). Comparison with experimentally determined mutant phenotypes and gene expression data has shown that the application of regulatory constraints to regulatory models led not only to more accurate predictive capabilities, but also to a broader predictive scope. For example, mutant regulatory gene phenotypes may now be predicted, as well as qualitative gene expression (Covert and Palsson, 2002). The metabolic/regulatory model of *E. coli* has recently been expanded to the genome scale and now accounts for over 1000 genes in this organism (Reed et al., 2003; Covert et al., submitted).

CONSTRAINTS-BASED MODELS IN THE FUTURE

Several constraints relevant to cells, especially microorganisms, have been enumerated. A framework has been presented whereby impossible behaviors are eliminated and the resulting solution space is used to analyze, interpret, and predict cell phenotype in the context of genotype and external environment. We have attempted to show how the recent incorporation of new constraints, such as energy balance and transcriptional regulation, results in both a broader scope of simulation and more accurate predictive capability.

This leads us to the following question: Which constraints can we incorporate in the future to add more functionality to our models? The primary challenges to integrating new constraints into cell models include obtaining the necessary data to define a constraint mathematically and developing the framework to incorporate the constraint in an existing model (for an insightful discussion and review of such approaches, see Bailey [1998]). For example, most of the constraints discussed herein have involved fluxes through biochemical reactions. One can anticipate that similar analyses will be developed for defining limits on allowable ranges of intracellular concentration as well as bracketing numerical values of kinetic parameters (Lee et al., 1999; Ronen et al., 2002). Importantly, not all constraints are of equal value with respect to modeling, and determination of the factors that are most constraining of cell function under various conditions will result in more accurate and meaningful simulations.

These challenges will likely be answered in a dialogue between experimental biologists, who are generally more aware of biological constraints, and computational biologists, who are equipped to describe constraints mathematically. As such dialogues multiply and grow, we believe that genome-scale constraints-based models will develop to include all major cellular processes and enable prediction and interpretation of large-scale data sets. Such models will be of fundamental importance in understanding biological systems.

APPENDIX 1

Some representative calculations illustrating certain physicochemical constraints in biology are given in what follows. These examples are computationally simple, but have much biological importance.

(i) *Growth rate constraints and chromosome replication.* The growth rate of a cell is limited by the time required to synthesize all necessary cellular components, such as cell

membrane, proteins, and chromosome. Accordingly, the synthesis time of each set of components can be calculated and interpreted as a constraint on growth rate. Chromosome replication in *E. coli* K-12 is considered here. Replication rate, r_{rep}, can be calculated as:

$$r_{rep} = \left(\frac{f \cdot r_f}{l_g}\right)$$

where r_f is the rate of replication for one replication fork, f is the number of replication forks, and l_g is the genome length. Given a genome size of roughly $4.3 \cdot 10^6$ bp, two replication forks, and a fork replication rate of approximately 750 bp per second (Marians, 1996), the average time required for genome replication is approximately 45 minutes. *E. coli* is also capable of increasing its growth rate by multiplexing DNA replication, resulting in an average of 4.2 replication forks per cell and an estimated doubling time of about 23 minutes (Neidhardt et al., 1990). As another example, the S, G_2, and M phases of the cell cycle in mammalian systems last about 8, 3, and 1 hours, respectively, resulting in a doubling time of approximately 12 hours (Alberts et al., 2002). The outcome of this computation leads to the following question: Which constraint dominates when *E. coli* grows at maximum or near-maximum growth rates?

(ii) *Maximum enzyme capacity*. Physicochemical constraints impose an upper limit on the maximum enzyme turnover rate. The enzyme association rate:

$$S + E \xrightarrow{k_1} SE$$

where S and E are the substrate and enzyme, respectively, is generally the limiting step in the enzymatic conversions of metabolites. The maximum turnover rate of this step, $V_{max} = k_1[S][E]$, is determined by the collision rate constant, k_1, as well as metabolite and enzyme concentrations in the cell. The substrate concentration may be approximated to be about 100 μM, assuming that the average compound concentration for 1000 different metabolites is about 100 g/mol molecular weight, and that 1% of the dry weight is composed of metabolic intermediates and that the cell density is about 1 g/cm^3 (Ingraham et al., 1983). The average enzyme concentration can be estimated similarly, assuming 1000 different enzymes in the cell with an average molecular weight of 40,000 g/mol, composing 15% of the dry weight (Ingraham et al., 1983). The enzyme concentration is thereby estimated to be about 1 μM. Using a typical collision rate constant of $k_1 = 10^8\ M^{-1}\ s^{-1}$ and the average enzyme and metabolite concentrations calculated earlier, a maximum enzymatic rate limit of 10^6 molecules $\mu m^{-3}\ s^{-1}$ is computed. This upper limit implies that the enzymatic reactions may never exceed rates faster than 10^6 molecules $\mu m^{-3}\ s^{-1}$. Measured experimental fluxes appear to coincide with this calculation; for example, in *E.* coli, a glucose uptake rate of 15 mmol $g^{-1}\ h^{-1}$ corresponds to about $0.7 \cdot 10^6$ molecules $\mu m^{-3}\ s^{-1}$, which falls within the estimated maximum rate.

(iii) *Evolution and chance*. As explained in the text, the probability of rolling a particular number may be low for one roll, but is driven to virtual inevitability as the number of rolls increases (de Duve, 1996). Similarly, we can ask the question: To what extent does ''chance'' govern the evolutionary process in a cell? A rough estimate may be obtained using simple statistics. The probability of obtaining a mutation in a particular base, P_m, after n replications, may be calculated using the replication error rate, P_{err}, by:

$$P_m = 1 - (1 - P_{err})^n$$

In other words, the chance of *not* getting the mutation of interest after n generations is subtracted from one. Given that the replication error rate of *E. coli* is generally about one error in 10^{10} bases (but can be increased under stress) (Neidhardt et al., 1990), the number of replications required for the mutation to occur with 99.9% probability is $7 \cdot 10^{10}$, corresponding to only 36 generations. Therefore, the appearance of any particular mutation in a population is extremely likely after a relatively short time (e.g., hours to days with *E. coli*), indicating that the environment may be a more constraining factor than the probability of a desired mutation.

The authors thank Markus Herrgard and Jason Papin for helpful discussions and reviews.

References

Alberts B, Johnson A, et al. 2002. Molecular biology of the cell. New York: Garland. 1463 p.

Bailey JE. 1998. Mathematical modeling and analysis in biochemical engineering: Past accomplishments and future opportunities. Biotechnol Progr 14:8–20.

Bailey JE. 2001. Complex biology with no parameters. Nat Biotechnol 19:503–504.

Beard DA, Liang SD, et al. 2002. Energy balance for analysis of complex metabolic networks. Biophys J 83:79–86.

Blaine-Metting F, Romine MF. 1997. Microbial genomics: The floodgates open. Trends Microbiol 5:91–92.

Bochner BR, Gadzinski P, et al. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res 11:1246–1255.

Bonarius HPJ, Timmerarends B, et al. 1998. Metabolite-balancing techniques vs. ^{13}C tracer experiments to determine metabolite fluxes in hybridoma cells. Biotechnol Bioeng 58:258–262.

Burgard AP, Maranas CD. 2003. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng 82:670–677.

Burgard AP, Vaidyaraman S, et al. 2001. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol Progr 17:791–797.

Chvatal V. 1983. Linear programming. New York: W.H. Freeman and Co.

Cockell CS, Catling DC, et al. 2000. The ultraviolet environment of Mars: Biological implications past, present, and future. Icarus 146: 343–359.

Cooper TF, Rozen DE, et al. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. Proc Natl Acad Sci USA 100:1072–1077.

Covert M, Palsson BO. 2003. Constraints-based models: Regulation of gene expression reduces the steady-state solution space. J Theor Biol 221:309–325.

Covert MW, Palsson BO. 2002. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. J Biol Chem 277: 28058–28064.

Covert MW, Reed JL, Knight EM, Herrgard MJ, Palsson BO. Combining high-throughput data and network-based computational models leads to testable hypothesis about transcriptional regulation in *E. coli.* (submitted).

Covert MW, Schilling CH, et al. 2001a. Metabolic modeling of microbial strains *in silico*. Trends Biochem Sci 26:179–186.

Covert MW, Schilling CH, et al. 2001b. Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213:73–88.

Davidson EH, Rast JP, et al. 2002a. A genomic regulatory network for development. Science 295:1669–1678.

Davidson EH, Rast JP, et al. 2002b. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246:162–190.

de Crecy-Lagard VA, Bellalou J, et al. 2001. Long term adaptation of a microbial population to a permanent metabolic constraint: Overcoming thymineless death by experimental evolution of *Escherichia coli*. BMC Biotechnol 1:10.

de Duve C. 1996. The constraints of chance. Sci Am 274:112.

DeLisa MP, Valdes JJ, et al. 2001. Quorum signaling via AI-2 communicates the ''metabolic burden'' associated with heterologous protein production in *Escherichia coli*. Biotechnol Bioeng 75: 439–450.

Desai RP, Nielsen LK, et al. 1999. Stoichiometric modeling of *Clostridium acetobutylicum* fermentations with non-linear constraints. J Biotechnol 71:191–205.

Devaux F, Marc P, et al. 2001. Transcriptomes, transcription activators and microarrays. FEBS Lett 498:140–144.

Edwards JS, Covert M, et al. 2002. Metabolic modelling of microbes: The flux-balance approach. Environ Microbiol 4:133–140.

Edwards JS, Palsson BO. 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 97:5528–5533.

Ellis RJ. 2001. Macromolecular crowding: Obvious but underappreciated. Trends Biochem Sci 26:597–604.

Elowitz MB, Levine AJ, et al. 2002. Stochastic gene expression in a single cell. Science 297:1183–1186.

Feinberg M. 1987. Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems. Chem Eng Sci 42:2229–2268.

Flores S, Gosset G, 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.

Gaasterland T, Oprea M. 2001. Whole-genome analysis: Annotations and updates. Curr Opin Struct Biol 11:377–381.

Gardner TS, Collins JJ. 2000. Neutralizing noise in gene networks. Nature 405:520–521.

Goodsell DS. 1993. The machinery of life. New York: Springer.

Hallows K, Knauf P. 1994. Principles of cell volume regulation. In: Strange K, editor. Cellular and molecular physiology of cell volume regulation. Boca Raton, FL: CRC Press. p 3–29.

Hoffmann A, Levchenko A, et al. 2002. The IkappaB-NF-kappaB signaling module: Temporal control and selective gene activation. Science 298:1241–1245.

Ibarra RU, Edwards JS, et al. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420:186–189.

Ideker T, Thorsson V, et al. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292: 929–934.

Ingraham JL, Maaloe O, et al. 1983. Growth of the bacterial cell. Sunderland, MA: Sinauer Associates.

Jamshidi N, Edwards JS, et al. 2001. Dynamic simulation of the human red blood cell metabolic network. Bioinformatics 17:286–287.

Jeong H, Tombor B, et al. 2000. The large-scale organization of metabolic networks. Nature 407:651–654.

Joshi A, Palsson BO. 1989. Metabolic dynamics in the human red cell. Part II—Interactions with the environment. J Theor Biol 141: 529–545.

Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucl Acids Res 28:27–30.

Karp PD, Riley M, et al. 2002. The EcoCyc database. Nucl Acids Res 30:56–58.

Karp PD, Riley M, et al. 2000. The EcoCyc and MetaCyc databases. Nucl Acids Res 28:56–59.

Kitano H. 2002. Systems biology: A brief overview. Science 295: 1662–1664.

Koch AL. 2000. The bacterium's way for safe enlargement and division. Appl Environ Microbiol 66:3657–3663.

Koonin EV. 2000. How many genes can make a cell: The minimal-gene-set concept. Annu Rev Genom Hum Genet 1:99–116.

Lee B, Yen J, et al. 1999. Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. Biotechnol Bioeng 62:722–729.

Lee I-D, Palsson BO. 1991. A comprehensive model of human erythrocyte metabolism: Extensions to include pH effects. Biomed Biochim Acta 49:771–789.

Lee S, Phalakornkule C, et al. 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. Compr Chem Eng 24:711–716.

Liao JC, Oh MK. 1999. Toward predicting metabolic fluxes in metabolically engineered strains. Metab Eng 1:214–223.

Mahadevan R, Edwards JS, et al. 2002. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. Biophys J 83:1331–1340.

Marians KJ. 1996. Replication fork propagation. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonella*: Cellular and molecular biology. Washington, DC: ASM Press. p 749–763.

McAdams HH, Arkin A. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet 15:65–69.

Naaby-Hansen S, Waterfield MD, et al. 2001. Proteomics—post-genomic cartography to understand gene function. Trends Pharmacol Sci 22:376–384.

Neidhardt FC, Ingraham JL, et al. 1990. Physiology of the bacterial cell. Sunderland, MA: Sinauer Associates.

Palsson BO. 2000. The challenges of in silico biology. Nat Biotechnol 18:1147–1150.

Papin JA, Price ND, et al. 2002a. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. J Theor Biol 215:67–82.

Papin JA, Price ND, et al. 2002b. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. Genome Res 12: 1889–1900.

Papin JA, Price ND, et al. 2003. Metabolic pathways in the post-genome era. Trends Biochem Sci 28:250–258.

Price ND, Papin JA, 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.

Price ND, Papin JA, et al. 2003. Genome-scale microbial in silico models: The constraints-based approach. Trends Biotechnol 21:162–169.

Price ND, Reed JL, et al. 2003. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. Biophys J 84:794–804.

Qu Z, Weiss JN, et al. 2003. Regulation of the mammalian cell cycle: A model of the G1-to-S transition. Am J Physiol Cell Physiol 284: C349–C364.

Raamsdonk LM, Teusink B, et al. 2001. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19:45–50.

Reed JL, Palsson BO. 2003. Thirteen years of building constraint-based in silico models of *Escherichia coli*. J Bacteriol 185:2692–2699.

Reed JL, Vo TD, Schilling CH, Palsson BO. 2003. An expanded genome-

scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol. 4:R54. Epub 2003 Aug 28.

Ronen M, Rosenberg R, et al. 2002. Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. Proc Natl Acad Sci USA 99:10555–10560.

Salgado H, Santos-Zavaleta A, et al. 2001. RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. Nucl Acids Res 29:72–74.

Savageau MA. 1998. Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. Biosystems 47:9–36.

Scheffler IE. 1999. Mitochondria. New York: Wiley-Liss. 367 p.

Schilling CH, Covert MW, et al. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. J Bacteriol 184:4582–4593.

Schilling CH, Palsson BO. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. J Theor Biol 203:249–283.

Segre D, Vitkup D, et al. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99: 15112–15117.

Stephanopoulos G, Nielsen J, et al. 1998. Metabolic engineering. San Diego, CA: Academic Press.

Tsuchiya HM, Fredrickson AG, et al. 1966. Dynamics of microbial cell populations. Adv Chem Eng 6:125–206.

Varma A, Palsson BO. 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.

Verkman AS. 2002. Solute and macromolecule diffusion in cellular aqueous compartments. Trends Biochem Sci 27:27–33.

Wagner A, Fell DA. 2001. The small world inside large metabolic networks. Proc R Soc Lond B Biol Sci 268:1803–1810.

Werner A, Heinrich R. 1985. A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary ''in vivo'' state and of time dependent variations under blood preservation conditions. Biomed Biochim Acta 44: 185–212.

Wiechert W, de Graaf AA. 1996. In vivo stationary flux analysis by 13C labeling experiments. Adv Biochem Eng/Biotechnol 54: 109–54.