# Regulation of Gene Expression in Flux Balance Models of Metabolism

MARKUS W. COVERT*, CHRISTOPHE H. SCHILLING‡ AND BERNHARD PALSSON*†

*Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093-0412, U.S.A. and ‡Genomatica, Inc., 5405 Morehouse, Suite 210, San Diego, CA 92121, U.S.A.*

Genome-scale metabolic networks can now be reconstructed based on annotated genomic data augmented with biochemical and physiological information about the organism. Mathematical analysis can be performed to assess the capabilities of these reconstructed networks. The constraints-based framework, with flux balance analysis (FBA), has been used successfully to predict time course of growth and by-product secretion, effects of mutation and knock-outs, and gene expression profiles. However, FBA leads to incorrect predictions in situations where regulatory effects are a dominant influence on the behavior of the organism. Thus, there is a need to include regulatory events within FBA to broaden its scope and predictive capabilities. Here we represent transcriptional regulatory events as time-dependent constraints on the capabilities of a reconstructed metabolic network to further constrain the space of possible network functions. Using a simplified metabolic/regulatory network, growth is simulated under various conditions to illustrate systemic effects such as catabolite repression, the aerobic/anaerobic diauxic shift and amino acid biosynthesis pathway repression. The incorporation of transcriptional regulatory events in FBA enables us to interpret, analyse and predict the effects of transcriptional regulation on cellular metabolism at the systemic level.

© 2001 Academic Press

## Introduction

The abundance of genomic data currently available has led to the construction of genome-scale models of microbial metabolism (Covert *et al.*, 2001). These models may be used to analyse, interpret and predict cellular phenotype from the genotype under defined environmental conditions. One method for the *in silico* analysis of metabolic networks is the constraints-based approach. This approach is based on the fact that the underlying cellular functions of biochemical reaction networks are subject to certain constraints that limit their possible behaviors (Palsson, 2000). In this approach, "hard" physicochemical constraints are used to define a closed solution space within which the steady-state solution to the flux vector must lie. The "best" solution is then found in the solution space using linear optimization. This analysis method has been called flux-balance analysis (FBA) (Varma & Palsson, 1994a, b; Bonarius *et al.*, 1997; Edwards *et al.*, 1999; Gombert & Nielsen, 2000). FBA has been described in detail and used in a variety of recent applications (Edwards & Palsson, 1999, 2000; Edwards *et al.*, 2001).

FBA models to date have not accounted for the constraints associated with regulation of gene expression nor activity of the expressed gene product. Until now, flux-based simulations have assumed that all gene products in the metabolic reaction network are available to contribute to an optimal solution, unconstrained by regulatory processes. An *in silico* model of *E. coli*, for

---

† Author to whom correspondence should be addressed.

example, accounts for 660 metabolic genes thought to be in the genome (Edwards & Palsson, 2000). However, it has been estimated that about 400 regulatory genes exist in the genome of *E. coli* (Thieffry *et al.*, 1998); of these, 178 regulatory and putative regulatory genes were found during annotation of the K-12 MG 1655 genome (Blattner *et al.*, 1997). Currently, 539 transcription units (sets of contiguously located genes with a common expression condition, promoter and terminator) are identified in RegulonDB, a database of transcriptional regulation and operon organization for *E. coli* (Salgado *et al.*, 2001). The high level of transcriptional regulation in this and other organisms has a significant effect on cell behavior. These regulatory effects have not been accounted for in previous FBA models of *E. coli*, which leads to certain incorrect predictions of cellular-level behavior (Edwards & Palsson, 2000). For FBA to be effectively used to predict cell behavior on a more general scale, these regulatory constraints must be incorporated. Unlike physico-chemical constraints, regulatory constraints are self-imposed by the organism, and presumably represent the result of an optimal evolutionary process.

The ability to model transcriptional regulatory events has several important applications. As high throughput technologies such as gene chips and microarrays have been developed to perform genome-wide expression assessment (DeRisi *et al.*, 1996), there is a need to predict intracellular transcription and transcriptional regulation on a whole-genome scale based on cellular environment and intracellular conditions. Regulatory modeling is also beneficial to the field of metabolic engineering. Using metabolic control analysis, it has been demonstrated that control of biosynthetic fluxes depends on multiple enzymes (Fell, 1998; McAdams & Arkin, 1998). The engineering of cell regulation, rather than of the expression of several related genes, may be a more efficient use of an organism's metabolism to produce a desired product. Experiments have already been described wherein a regulatory, rather than a metabolic, network was manipulated to increase the flux through a particular pathway (Ostergaard *et al.*, 2000). The results of such experiments may lead to significant advances in the large-scale microbial generation of valuable

substances such as pharmaceuticals (Chartrain *et al.*, 2000) and biocommodities (Lynd *et al.*, 1999; Aristidou & Penttilä, 2000; Fotheringham, 2000) or in pollutant degradation (Pieper & Reineke, 2000).

Metabolic regulation and cell dynamics have been modeled using several approaches (Rao & Arkin, 2001; Hasty *et al.*, 2001). For example, Boolean logic may be used to examine the various states of a regulatory circuit (Thomas, 1973; Kauffman, 1993; Somogyi & Sniegoski, 1996), such as the regulatory networks controlling the lambda phage decision circuit (Thieffry & Thomas, 1995) as well as the immune response (Kaufman *et al.*, 1985). Such mathematical descriptions enable the qualitative study of regulatory structure and lead to general analytical insights which can be usefully applied to the analysis of complex metabolic networks, but must be used in connection with other techniques to make truly quantitative predictions. Mixed-integer linear optimization has also been used to predict optimal regulatory structures for metabolic engineering (Hatzimanikatis *et al.*, 1996). Another approach is the use of kinetic theory to solve systems of ordinary differential equations (Reich & Sel'kov, 1981; Shuler & Domach, 1983; Fell, 1996; Heinrich & Schuster, 1996; Stephanopoulos *et al.*, 1998), as has been done to study *E. coli* growth on glucose and lactose (Wong *et al.*, 1997). Kinetic theory may also be combined with Boolean logic as in a hybrid model of the lambda phage (McAdams & Shapiro, 1995), using fuzzy logic as has been done with *E. coli* core metabolism (Lee *et al.*, 1999) or in conjunction with cybernetic principles (Guardia *et al.*, 2000; Varner, 2000). Other approaches to the analysis of genetic regulatory circuits include the use of fractal kinetic theory (Savageau, 1998) and stochastic modeling techniques (McAdams & Arkin, 1997, 1998, 1999; Carrier & Keasling, 1999).

Detailed deterministic and stochastic models require extensive information, such as temperature, substrate availability, the presence of signaling molecules, and other environmental parameters, many of which have yet to be completely specified. The difficulty inherent in constructing detailed deterministic or stochastic models is therefore that most of the parameters required to develop them are typically very

difficult to obtain under all possible environmental conditions, although some progress has been made (Vaseghi *et al.*, 1999). For this reason, none of the above models has been able to be adopted on a scale large enough to represent the entire metabolism of an organism, as recently observed (Bailey, 2001), and thus the systemic regulatory/ metabolic properties of an entire organism have yet to be analysed.

This paper describes a framework whereby the development of a transcriptional regulatory structure may be used together with FBA to generate time profiles as well as steady-state solutions of cell growth, substrate utilization and by-product secretion for organisms for which the metabolism and regulation have been characterized, either from the genome or experimentally.

## Methods

### CONSTRAINTS-BASED ANALYSIS

The constraints-based approach to metabolic network analysis has been recently described (Palsson, 2000) and is illustrated in Fig. 1. The axes represent fluxes through all individual reactions in the metabolic network. Not all the points in this space are attainable because of constraints on the system, such as the interrelatedness of the fluxes, thermodynamics or maximum capacity. By imposing these constraints, one can restrict the behaviors available to the cell to an enclosed solution space which contains a finite number of possibilities (shown on the right-hand side of Fig. 1), from which an optimal solution (gray circle) may be determined using linear programming (Schilling *et al.*, 1999; Gombert & Nielsen, 2000).

### FBA AND THE INTERPRETATION OF MICROBIAL GROWTH EXPERIMENTS

FBA can be used to quantitatively predict the time profiles of cell density as well as substrate and by-product concentrations (Varma & Palsson, 1994a, b). In general, mass balances are written around every metabolite concentration in a metabolic network; the resulting equations may be written in matrix notation as

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mathbf{S} \cdot v, \tag{1}$$

where $\mathbf{S}$ is the stoichiometric matrix, $v$ is a vector which contains the value of all reaction and transport fluxes and $X$ is a vector of metabolite concentration. Since the time constants which describe metabolic transients are fast (on the order of milli-seconds to tens of seconds; McAdams & Arkin, 1998) as compared to the time constants associated with cell growth, on the order of hours to days, the system may be treated by considering the behavior inside the cell to be in a quasi-steady state, and eqn (1) reduces to

$$\mathbf{S} \cdot v = 0. \tag{2}$$

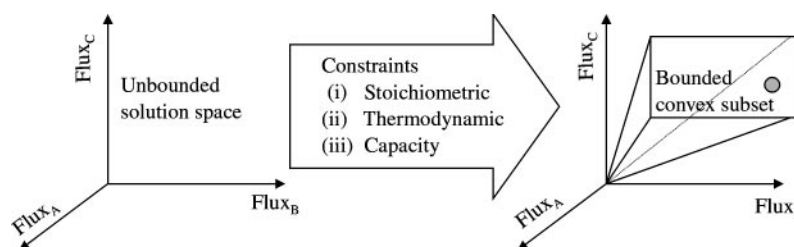The steady-state equation contains the mass balance constraints imposed on the system.



FIG. 1. Constraints-based analysis of metabolic reaction networks. The metabolic network of an organism may be described in geometrical terms. On the left is an unbounded space containing every possible distribution of fluxes through each reaction in the network. Many of the points in the space are unattainable due to constraints on the system. By applying these constraints, such as system stoichiometry, thermodynamics (e.g. the reversibility of reactions) and maximum capacity, the set of all possible behaviors is reduced to a bounded convex subset (right) in which the solution (gray circle) must lie. These flux distributions represent phenotypes which may be exhibited by a reconstructed metabolic network.

Thermodynamic and capacity constraints or experimental data can be used to limit the magnitude of each individual metabolic flux in the network

$$\alpha_i \leqslant v_i \leqslant \beta_i. \qquad (3)$$

Using the above equation, $\alpha_i$ or $\beta_i$ may be set to zero or to another finite value to constrain the direction or magnitude of a flux. Such constraints correspond to thermodynamic constraints (e.g. effective irreversibility of a given reaction due to an extremely high equilibrium constant) or capacity constraints (e.g. maximum uptake rate for a given transport protein), respectively. Experimental data on flux levels, as obtained by isotope labeling (Wiechert & de Graaf, 1996) or other methods, can also be used to set $\alpha_i$ and $\beta_i$ (an approach called metabolic flux analysis, see Stephanopoulos *et al.*, 1998). These equality and inequality constraints define a closed solution space.

### REPRESENTING TRANSCRIPTIONAL REGULATORY CONSTRAINTS

Cells are subject to both invariant (i.e. non-adjustable) and adjustable constraints. The former are physico-chemical in origin and include stoichiometric, capacity and thermodynamic constraints. They can be used to bracket the range of possible behaviors as described above. Adjustable constraints are biological in origin, and they can be used to further limit allowable behavior. These constraints will change in a condition-dependent manner.

Regulatory events impose temporary, adjustable constraints on the solution space as shown in Fig. 2. This figure depicts a solution space and optimal solution (gray circle) for a hypothetical metabolic network. The solution space defined by non-adjustable constraints is shown in Fig. 2(a). If the flux through a certain reaction is repressed due to transcriptional regulation, then one or more extreme vectors that define the boundaries of the solution space are removed and the volume of the space (i.e. the range of allowable cellular behaviors) is reduced. For example, when the vectors are removed as illustrated in Fig. 2(b) and (c), respectively, the
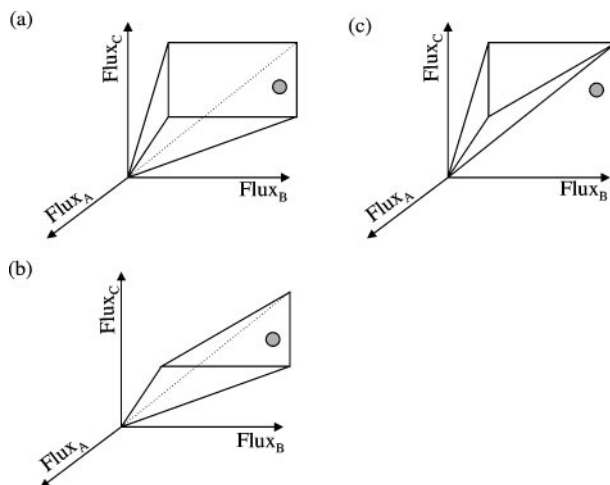


FIG. 2. Regulatory constraints change the shape of the solution space. The hypothetical solution space and solution as given in Fig. 1, defined by various non-adjustable constraints, is shown in (a). The flux through a certain reaction may be constrained by a transcriptional regulatory event, leading to the removal of one or more extreme vectors from the boundaries of the solution space. These constraints further reduce the size of the solution space. After regulatory constraints have been applied, the original solution may either remain in the smaller solution space (b), or may no longer be located inside the space (c), in which case, a new solution (i.e. a new behavior) will be determined by the cell.

solution space is restricted to a smaller space. This restricted space is analogous to a cell with fewer metabolic behavioral possibilities. Note that the optimal solution remains in the subspace shown in Fig. 2(b) but not in the subspace shown in Fig. 2(c). If the optimal solution is no longer in the space, the phenotype that it corresponds to cannot be expressed and a new optimal solution will be determined corresponding to a different behavior exhibited by the cell.

The transcriptional regulatory structure can be described using Boolean logic equations. This approach involves restricting expression of a transcription unit to the value 1 if the transcription unit is transcribed and 0 if it is not. Similarly, the presence of an enzyme or regulatory protein, or the presence of certain conditions inside or outside of the cell, may be expressed as 1 if the enzyme, protein, or a certain condition is present and 0 if it is not.

The Boolean logic representation includes well-known modifiers such as *AND*, *OR*, and *NOT*, which can be used to develop equations
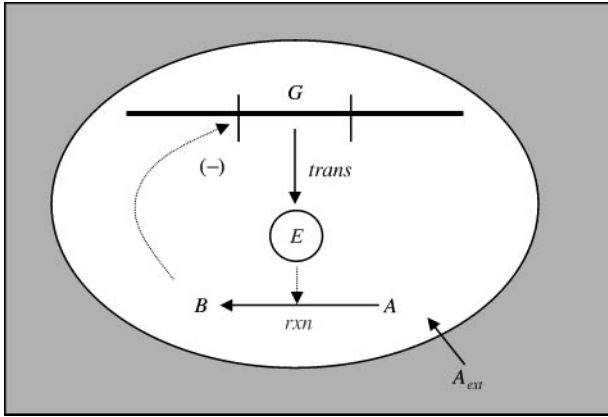
FIG. 3. A simple regulatory circuit. Here, gene $G$ is transcribed by a process *trans* to produce an enzyme $E$. This enzyme then catalyses a reaction *rxn* which converts substrate $A$ into product $B$. Product $B$ then represses transcription of $G$, leading to depletion of $E$.

governing the expression of transcription units. Consider a simple system, as depicted in Fig. 3 (adapted from Thomas, 1973), containing one gene $G$ which is transcribed by a process *trans*, resulting in an enzyme $E$. This enzyme then catalyses the reaction *rxn* which is the conversion of substrate $A$ to product $B$. Product $B$ interacts with a binding site near $G$ such that the transcription process *trans* is inhibited. In other words, the transcription event *trans* will occur if the gene $G$ is present in the genome and the product $B$ is not present to bind to the DNA. A logic equation which describes this circumstance is

$$trans = IF\ (G)\ AND\ NOT\ (B). \qquad (4)$$

After a certain time for protein synthesis has elapsed, progression of the transcription/translation process *trans* will result in significant amounts of enzyme $E$. Similarly, after a certain protein decay time, the absence of process *trans* will result in decay and eventual depletion of $E$.

The requirement for the reaction *rxn* to proceed is the presence of $A$ and of $E$, for which a logical equation can be written as

$$rxn = IF\ (A)\ AND\ (E). \qquad (5)$$

The presence of enzymes or regulatory proteins in a cell at a given point in time depends both on the previous transcription history of the cell and on the rates of protein synthesis and decay. If sufficient time for protein synthesis has elapsed since a transcription event for a particular transcription unit occurred, we say that enzyme $E$ is present in the cell and remains present until the time for $E$ to decay has elapsed without the cell experiencing another transcription event for that specific transcription unit. In other words, dynamic parameters—the time delays of protein synthesis and degradation —are required in addition to the known causal relationships that represent regulation of gene transcription. Under steady-state conditions, the average protein synthesis and degradation times are equal.

Once the presence of all regulated enzymes in the metabolic network has been determined for a given time interval $(t_1 \rightarrow t_2)$, if an enzyme has been determined "not present" for the time interval, then the flux through that enzyme is set to zero. This restriction may be thought of as adding a temporary constraint on the metabolic network

$$v_k(t) = 0, \quad when\ t_1 \leqslant t \leqslant t_2, \qquad (6)$$

where $v_k$ is the flux though a reaction at the given time point $t$. If an enzyme is "present" during a given time interval, the corresponding flux is left unconstrained by regulation and determined instead using FBA. This approach thereby retains the quantitative aspects of FBA while incorporating qualitative regulatory information.

Thus, known regulation of gene expression in cells can be represented by Boolean logic and incorporated into flux balance models. Such regulation represents additional condition-dependent constraints on the $v_{max}$ through a particular reaction.

## TIME COURSE OF GROWTH

The quasi-steady-state assumption on the metabolic network can be used to generate dynamic profiles of cell growth. The experimental time is divided into small time steps, $\Delta t$. Beginning at $t = 0$ where the initial conditions of the experiment are specified, the metabolic model may be used to predict the optimal flux

distribution for the metabolic network. From the transport fluxes, the extracellular concentrations may be calculated in a time-dependent fashion, as previously described (Varma & Palsson, 1994). These concentrations are then used as the initial conditions for the next time step. This type of dynamic modeling was shown to correlate well with the growth of *E. coli* on glucose minimal media under aerobic and anaerobic conditions, predicting quantitatively the uptake of glucose and growth rate as well as by-product secretion (Varma & Palsson, 1994).

The time constants characterizing transcriptional regulation are generally on the order of a few minutes or slower (Zubay, 1973; Rivett, 1986; McAdams & Arkin, 1998), which are slower than the time constants associated with metabolism. Therefore, the FBA generation of time profiles for dynamic cellular behavior may be integrated with a set of transcriptional regulatory rules which are represented by Boolean logic equations. The status of transcription is found from the given conditions at a particular time interval. Specifically, transcription may be altered by the presence or surplus of an intracellular metabolite, an extracellular metabolite, regulatory proteins, signaling molecule, or any combination of these or other factors. The logic equation governing transcription of each transcriptional unit is used to determine whether transcription occurs or does not occur.

We now apply this approach to analyse a simple example that represents the skeleton of core metabolism.

### SAMPLE NETWORK

A simplified metabolic network is represented in Fig. 4; the reactions and regulatory rules are given in Table 1. The network contains 20 reactions, seven of which are regulated by four regulatory proteins. For the purposes of this example, the following instances of transcriptional regulation were examined:

1. Preferential carbon source uptake/catabolite repression (Saier *et al.*, 1996). For this example, *Carbon1* is arbitrarily defined as the preferred carbon source. For our purposes, we say that the presence of *Carbon1* in the

extracellular medium activates a regulatory protein which inhibits the transcription of the gene which encodes a protein for transport of *Carbon2* into the cell, via a transport process *Tc2*. Framed in terms of Thomas' Boolean formalism, the resulting equations for this system are

$$RPc1 = IF\ (Carbon1), \qquad (7)$$

$$tTc2 = IF\ NOT\ (RPc1), \qquad (8)$$

where $RPc1$ is the regulatory protein which senses extracellular *Carbon1*, $tTc2$ is the occurrence of a transcription event (which will eventually result in the protein enabling transport process $Tc2$ and the relaxation of one regulatory constraint, $\leftarrow v_{Tc2} = 0$, on the solution space).

2. Anaerobic growth. The transcription of many enzymes is regulated according to whether or not oxygen is available to the cell (Lynch & Lin, 1996). In this case, the presence of *Oxygen* will inactivate regulatory protein *RPO2*, which inhibits transcription of the genes for *Rres* and *R5a* but induces transcription of the gene for *R5b*. Note that *R5a* and *R5b* are reactions catalyzed by isozymes. The logic equations follow this form:

$$RPO2 = IF\ NOT\ (Oxygen), \qquad (9)$$

$$tRres = IF\ NOT\ (RPO2), \qquad (10)$$

$$tR5a = IF\ NOT\ (RPO2), \qquad (11)$$

$$tR5b = IF\ (RPO2). \qquad (12)$$

3. Amino acid biosynthesis pathway repression. The transcription of amino acid biosynthesis genes is often induced by a low intracellular concentration of the amino acid (Patte, 1996). Intracellular concentrations as yet cannot be obtained using FBA; instead we use fluxes to approximate the regulation. Metabolite *H* represents the "amino acid" in this example, and can be made by the cell via reaction *R8a* or transported from the extracellular media through transport process *Th*. For the regulatory structure, *Th* will be used to activate *RPh* which will repress transcription of the gene encode *R8a*. In other words, we assume that if *Th* is active
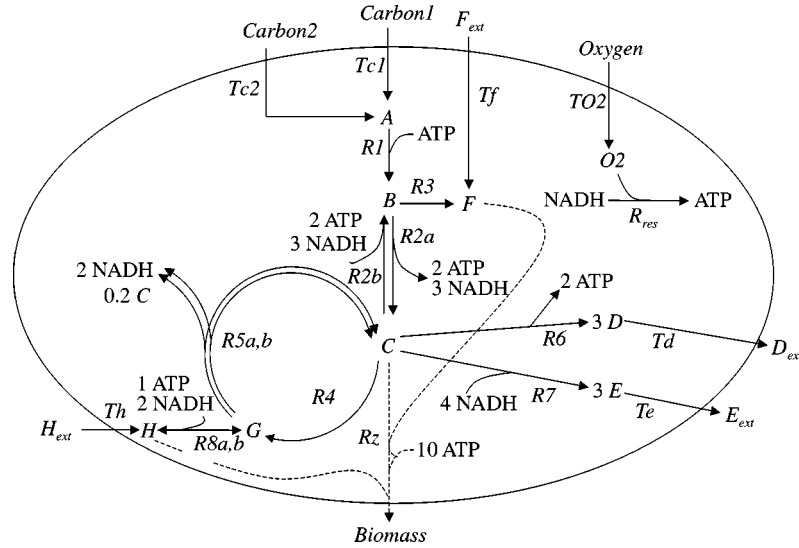
FIG. 4. A simplified core carbon metabolic network, mimicking core metabolism.

(due to the presence of extracellular $H$), the concentration of amino acid is relatively high and therefore that transcription of $H$ biosynthesis genes will not be induced.

$$RPh = IF\ (v_{Th} > 0),\qquad(13)$$

$$tR8a = IF\ NOT\ (RPh).\qquad(14)$$

4. Transcriptional regulation to maintain concentration levels of important metabolites (Saier *et al.*, 1996). The activation or repression of these genes depends on the level of $B$ in the cell. Again, rather than attempting to determine an internal concentration, we may use a flux rather than concentration to turn "off" an enzyme. We choose $R2b$ as the determining factor; it will activate $RPb$ which in turn will inactivate $tR2a$ and $tR7$.

$$RPb = IF\ (v_{R2b} > 0),\qquad(15)$$

$$tR2a = IF\ NOT\ (RPb),\qquad(16)$$

$$tR7 = IF\ NOT\ (RPb).\qquad(17)$$

The theoretical metabolic capabilities of the simplified metabolic network shown in Fig. 4 combined with a regulatory structure as defined in Table 1 were examined using the constraints-based approaches described above. Specifically, at a given time point, a commercially available linear programming package (LINDO, Lindo Sytems, Chicago) was used to identify an optimal metabolic flux distribution within the solution space. The optimal metabolic flux distribution was identified as the flux distribution which maximized the *Growth* flux in Table 1, a flux which represents growth of an organism by removing necessary precursors of growth from the system:

$$C + F + H + 10\ \mathrm{ATP} \rightarrow Biomass.\qquad(18)$$

The hypothesis that microbial cells behave in such a way that their growth is optimized has been verified experimentally under certain conditions (Edwards *et al.*, 2001). Using the resulting flux distribution and the conditions of the system in a previous time step, the conditions of the next time step were calculated on a commercially available spreadsheet package (EXCEL, Microsoft Corporation, Redmond), following a procedure previously described (Varma & Palsson, 1994) to obtain biomass as well as extracellular substrate and by-product concentrations. Numerical values for the parameters used in the simulation, such as the protein transcription and decay time as well as maximum uptake rates for all possible substrates, are shown in Table 2.

TABLE 1

*Reactions and regulatory rules for the simplified metabolic network shown
in Fig. 4. The network contains 20 reactions, seven of which are regulated
by four regulatory proteins*

| Reaction | Name | Regulation |
|---|---|---|
| *Metabolic reactions* | | |
| $-1\ A - 1$ ATP $+ 1\ B$ | *R1* | |
| $-1\ B + 2$ ATP $+ 2$ NADH $+ 1\ C$ | *R2a* | *IF NOT(RPb)* |
| $-1\ C - 2$ ATP $- 2$ NADH $+ 1\ B$ | *R2b* | |
| $-1\ B + 1\ F$ | *R3* | |
| $-1\ C + 1\ G$ | *R4* | |
| $-1\ G + 0.8\ C + 2$ NADH | *R5a* | *IF NOT (RPO2)* |
| $-1\ G + 0.8\ C + 2$ NADH | *R5b* | *IF RPO2* |
| $-1\ C + 2$ ATP $+ 3\ D$ | *R6* | |
| $-1\ C - 4$ NADH $+ 3\ E$ | *R7* | *IF NOT (RPb)* |
| $-1\ G - 1$ ATP $- 2$ NADH $+ 1\ H$ | *R8a* | *IF NOT (RPh)* |
| $+1\ G + 1$ ATP $+ 2$ NADH $- 1\ H$ | *R8b* | |
| $-1$ NADH $- 1\ O2 + 1$ ATP | *Rres* | *IF NOT (RPO2)* |
| *Transport processes* | | |
| $-1\ Carbon1 + 1\ A$ | *Tc1* | |
| $-1\ Carbon2 + 1\ A$ | *Tc2* | *IF NOT(RPc1)* |
| $-1\ F_{ext} + 1\ F$ | *Tf* | |
| $-1\ D + 1\ D_{ext}$ | *Td* | |
| $-1\ E + 1\ E_{ext}$ | *Te* | |
| $-1\ H_{ext} + 1\ H$ | *Th* | |
| $-1\ Oxygen + 1\ O2$ | *To2* | |
| *Maintenance and growth processes* | | |
| $-1\ C - 1\ F - 1\ H - 10$ ATP $+ 1\ Biomass$ | *Growth* | |
| *Regulatory proteins* | | |
| | *RPO2* | *IF NOT(Oxygen)* |
| | *RPc1* | *IF Carbon1* |
| | *RPh* | *IF $(v_{Th} > 0)$* |
| | *RPb* | *IF $(v_{R2b} > 0)$* |

TABLE 2

*Numerical values of parameters used in growth
simulations of the sample metabolic/regulatory
network*

| Parameter | Value |
|---|---|
| *Maximum transport rates* (mmol g-DCW$^{-1}$ hr$^{-1}$) | |
| *Carbon1* | 10.5 |
| *Carbon2* | 10.5 |
| *D* | 12.0 |
| *E* | 12.0 |
| *F* | 5.0 |
| *H* | 5.0 |
| *O2* | 15.0 |
| *Protein synthesis/decay delay* (hr) | 0.25 |

## Results

Using the sample metabolic network shown in Fig. 4, modeling the dynamic growth of a cell with the incorporation of temporary regulatory constraints as shown in Table 1 can be illustrated with several insightful examples. The results of five simulations, chosen to illustrate each regulatory element separately and in a complex medium, are described below.

EXAMPLE 1—DIAUXIE ON TWO CARBON SOURCES

The first example concerns the growth of the cell on two carbon sources, *Carbon1* and *Carbon2*. For this example, the initial concentrations of both carbon sources in the media were set to 10 mM and the simulation was run with oxygen in excess. The simulation was run until both
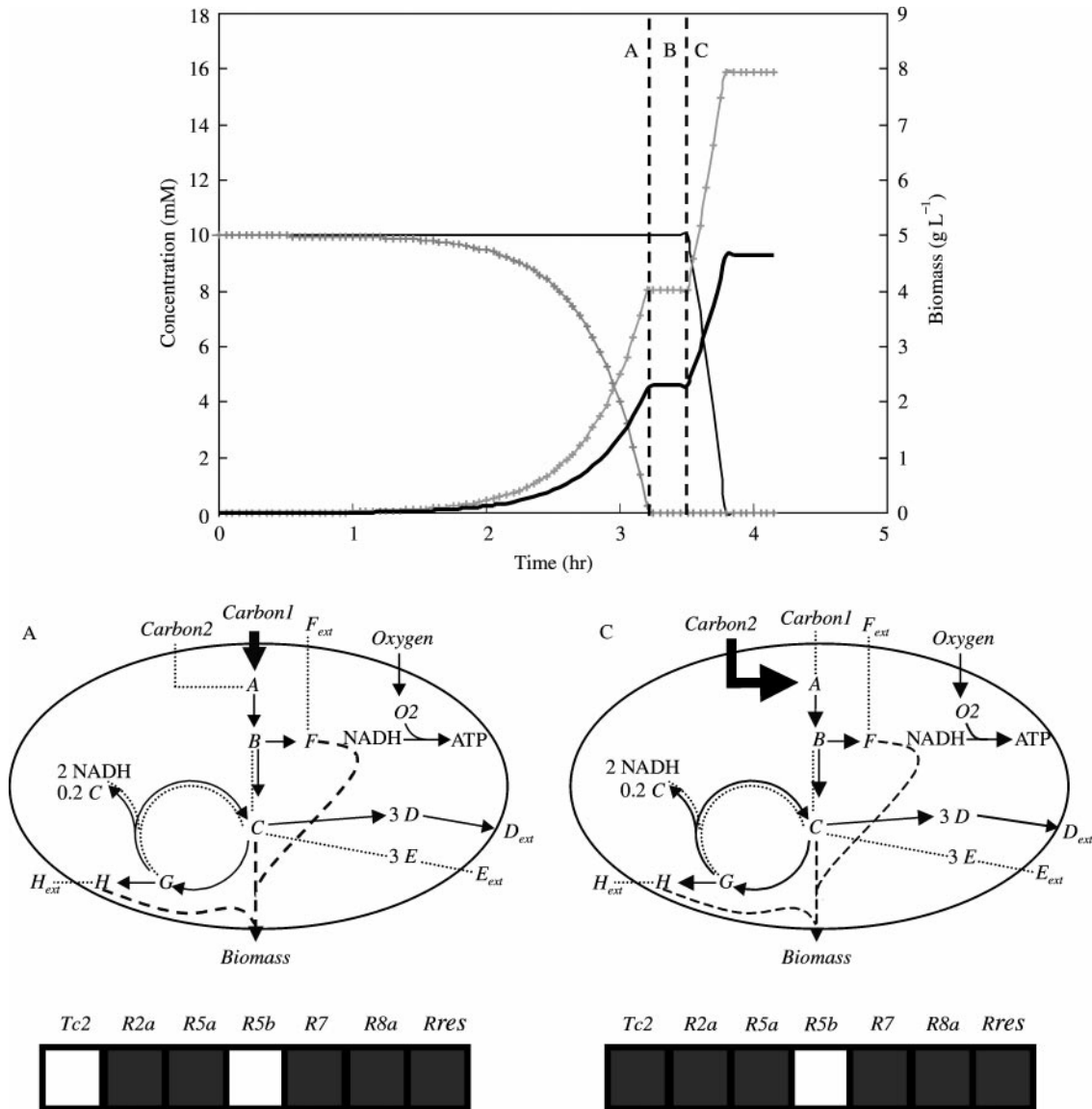
FIG. 5. Catabolite repression in the simplified network. A time profile of calculated growth and metabolism is shown (top), divided into three regions with dotted lines. For regions A and C, the network maps are shown (bottom) with the inactive fluxes denoted by thin dotted lines and the active fluxes shown as solid, or in the case of the biomass flux dashed, black lines. Certain fluxes are emphasized with a thick arrow to indicate the change in flux distributions due to regulation. Underneath each network map is an *in silico* array which shows whether a particular reaction is activated (dark gray) or inactivated (white). Labels correspond to metabolites shown in Table 1, with three additions: *C1* = *Carbon1*, *C2* = *Carbon2*, *X* = *Biomass*. (——) *C1*; (——) *C2*; (——) *D*; (——) *X*.

carbon sources had been completely exhausted. The results are shown in Fig. 5. The upper half of Fig. 5 is a time plot which shows the concentrations of *Carbon1*, *Carbon2*, by-product *D*, and the biomass *X*. The typical diauxic growth curve found in instances of catabolite repression is easily seen. The time plot is divided into three regions with dotted lines. The flux maps generated

using FBA are qualitatively identical in each region, and the bars beneath the flux maps indicate whether the regulated genes are being transcribed (dark gray) or not (white). In region A (flux map shown on bottom left of Fig. 5), the cell grows using *Carbon1* as the preferred carbon source; in region C (bottom right), *Carbon2* is used. Region B (not shown) is a short period where the cell

does not grow while the transport protein which enables the transport process *Tc2* is being up-regulated and synthesized. Without the addition of regulatory constraints, the system would grow on both carbon sources together to maximize production of the biomass conditions and fail to predict a diauxic shift.

### EXAMPLE 2—AEROBIC/ANAEROBIC-DIAUXIE

The second example of transcriptional regulatory modeling is the diauxic shift associated with a sudden removal of available oxygen to the culture (Fig. 6). Again, the dynamic profile may be divided into three areas. First, in region A, the culture grows with *Carbon2* as the primary carbon source. At 2 hr, the oxygen supply to the culture is removed and the culture grows anaer-

obically (regions B and C). It can readily be seen from Fig. 6 that regions B and C have similar flux distributions and that the regulatory constraints, unlike in Fig. 5 where they actually restrict the optimal solution, act redundantly (the regulation is not necessary for the solution). Therefore, the regulatory structure allows unnecessary proteins to decay in this case, without changing the optimal solution.

### EXAMPLE 3—GROWTH ON CARBON AND AMINO ACID WITH CARBON IN EXCESS

In this case, shown in Fig. 7, the simplified amino acid *H* is present in the medium with *Carbon2*, where it is used both to satisfy the *H* biomass requirement and as a supplementary
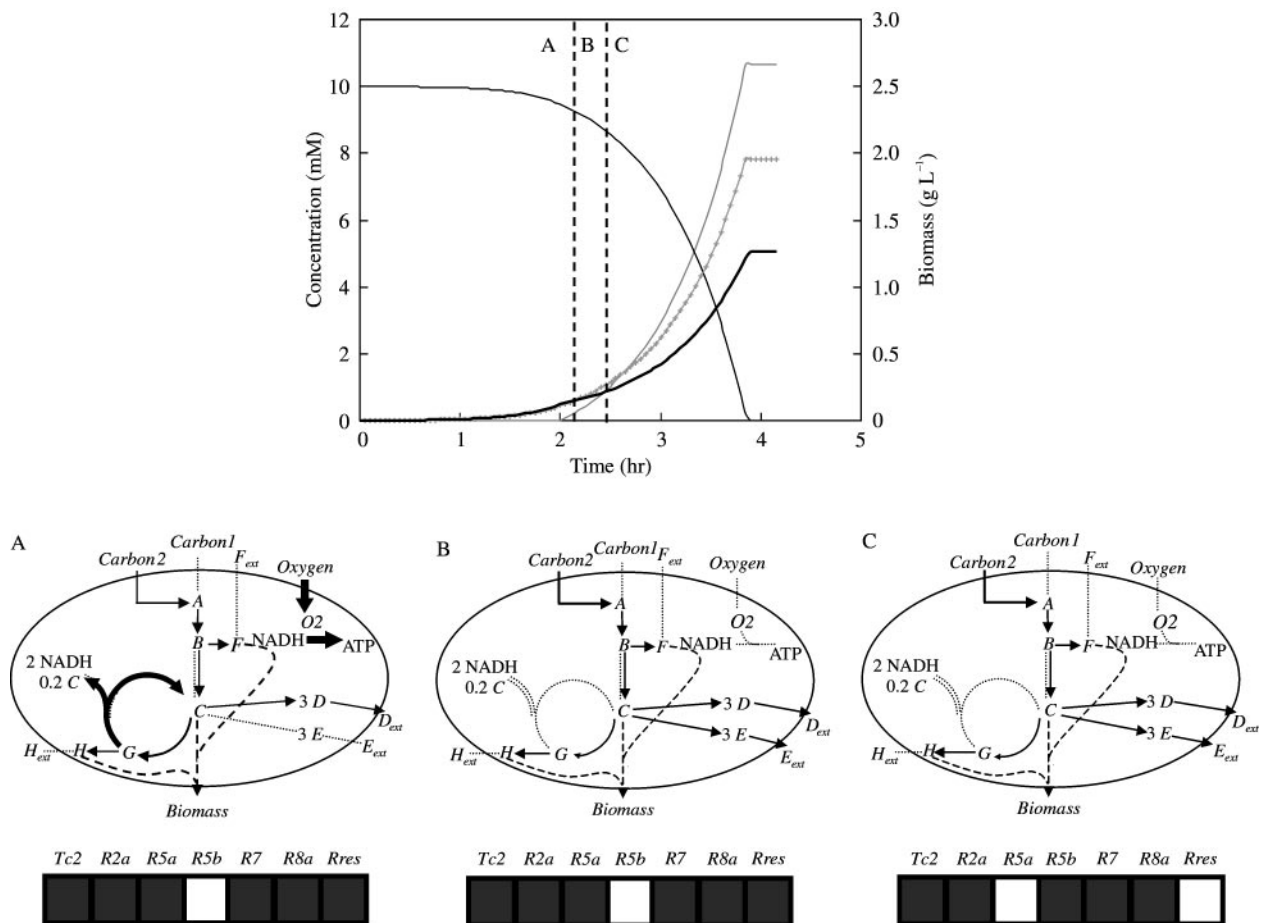


FIG. 6. Aerobic/anaerobic growth calculated using the simplified network. Similar in format to Fig. 5. Again, the dynamic profile may be divided into three areas, with the metabolic network maps and *in silico* arrays shown below. (——) *E*; (——) *C2*; (—+—) *D*; (——) *X*.
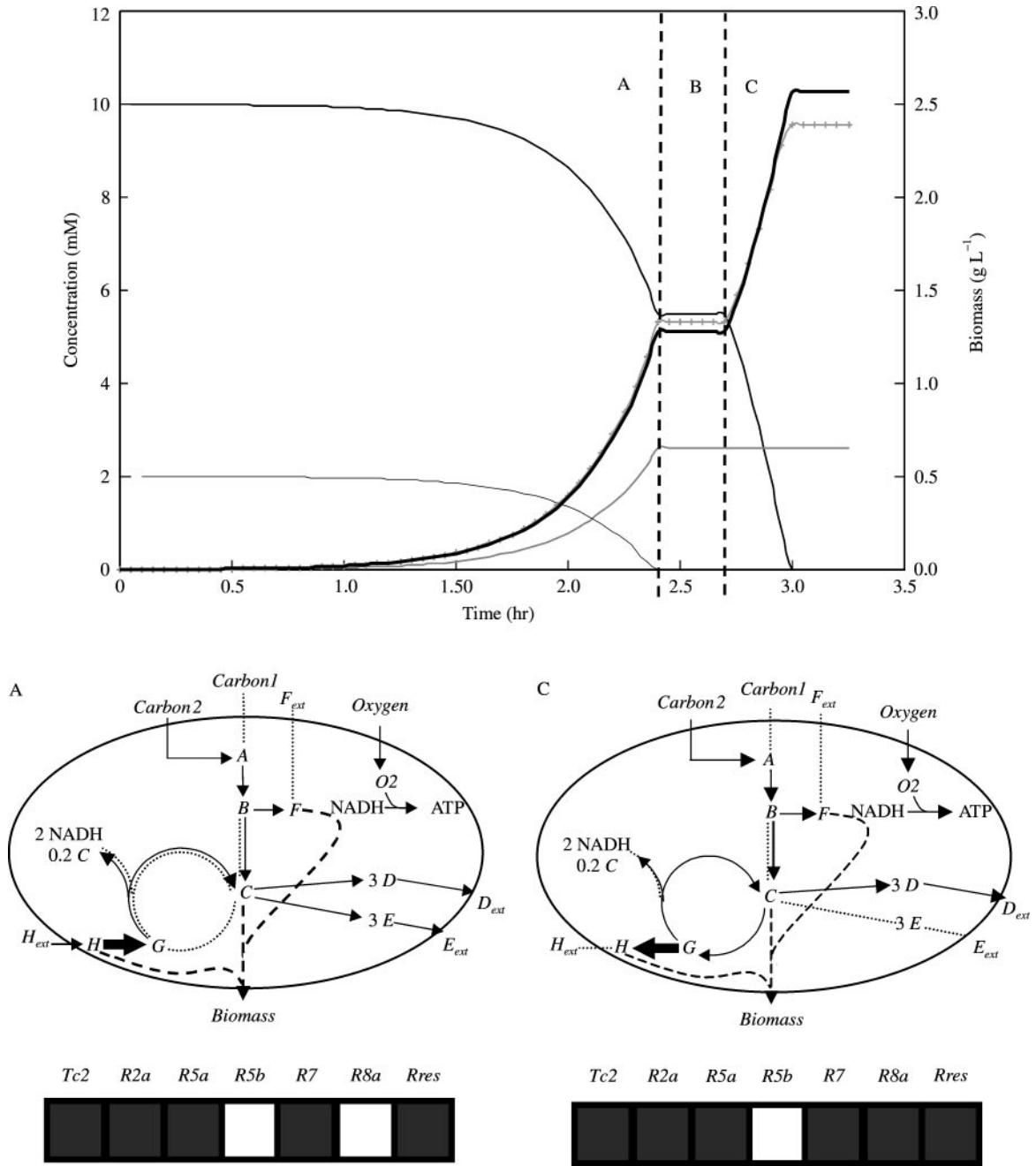
FIG. 7. Amino acid biosynthesis. Similar in format to Figs 5 and 6. (——) $C2$; (—+—) $D$; (——) $E$; (——) $H$; (▬▬) $X$.

source of $C$ (Region A). After the extracellular supply of $H$ is depleted there is a region of no growth (Region B) while the cell upregulates its $H$ biosynthesis machinery to catalyze $R8a$. This example is similar to Example 1 in that there is a pause where no growth occurs while the regulatory structure is expanding the solution space to allow the cell to synthesize $H$.

EXAMPLE 4—GROWTH ON CARBON AND AMINO ACID WITH AMINO ACID IN EXCESS

Figure 8 shows a graph of simulated growth on *Carbon2* and $H$ where $H$ is in excess. This case is designed to demonstrate the regulation of *RPb*; initially, the *Carbon2* and $H$ are taken up together as in Example 3 (Region A). However, as
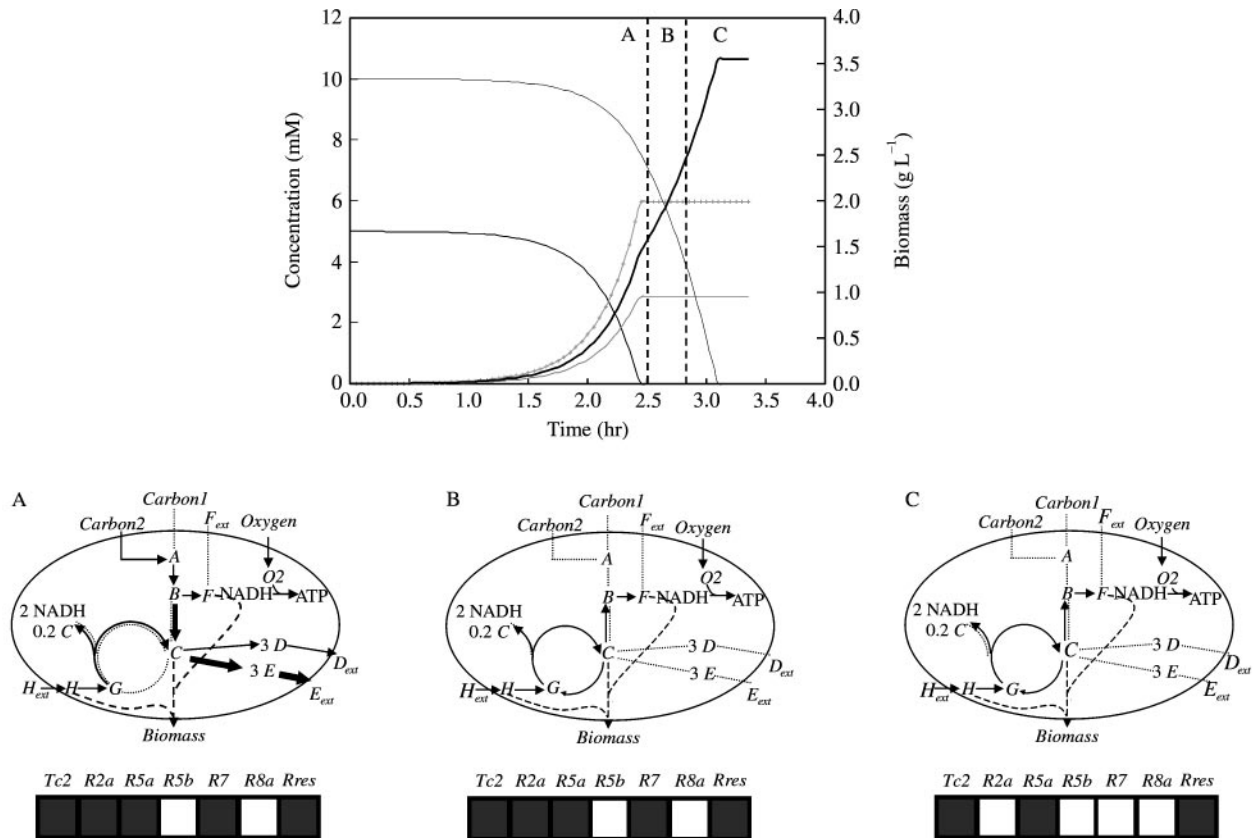
FIG. 8. Growth on carbon and amino acid with amino acid in excess. Similar in format to Figs 5–7. (——) C2; (—+—) D; (——) E; (——) H; (——) X.

the Carbon2 is depleted, RPb is activated and the transcription of enzymes catalyzing R2a and R7 ceases (Region C). As in Example 2, the flux distribution changes as the process moves to Region B and remains constant as regulatory events constrain the cell to stop producing certain unnecessary proteins.

### EXAMPLE 5—COMPLEX MEDIUM

Figure 9 depicts growth of the cell on a complex medium, with initial substrate concentrations strategically chosen so that the status of all regulatory proteins will be changed over the course of the experiment. Carbon1, Carbon2, F and H are all initially present in the medium; at 1 hr, oxygen is also allowed into the culture. The figure illustrates the interplay of several regulatory actions to control the growth of the cell on multiple substrates and under changing condi-

tions. The plot is divided into 15 regions, with Regions 12 and 15 as transitory periods with no growth while certain regulatory changes are occurring. The reactions which are being regulated are thick black arrows, with the exceptions of R5a and R5b which are light and dark gray arrows, respectively. In seven of these (Regions 1, 2, 4–6, 8, 10) the internal flux distribution is qualitatively similar, with minor changes as Carbon1 or F is depleted. Another more interesting change occurs in the transition from Regions 2 to 3. The simplified TCA cycle uses R5b in Regions 1 and 2 and begins using R5a (catalyzed by an isozyme which is expressed under aerobic conditions) in Region 3, after oxygen has been allowed to enter the system. Since R5a and R5b are equivalent stoichiometrically, FBA alone does not favor one reaction over the other and would fail to predict which of the two isozymes are active under given conditions. The growth of the cell on the complex medium also exhibits unusual flux
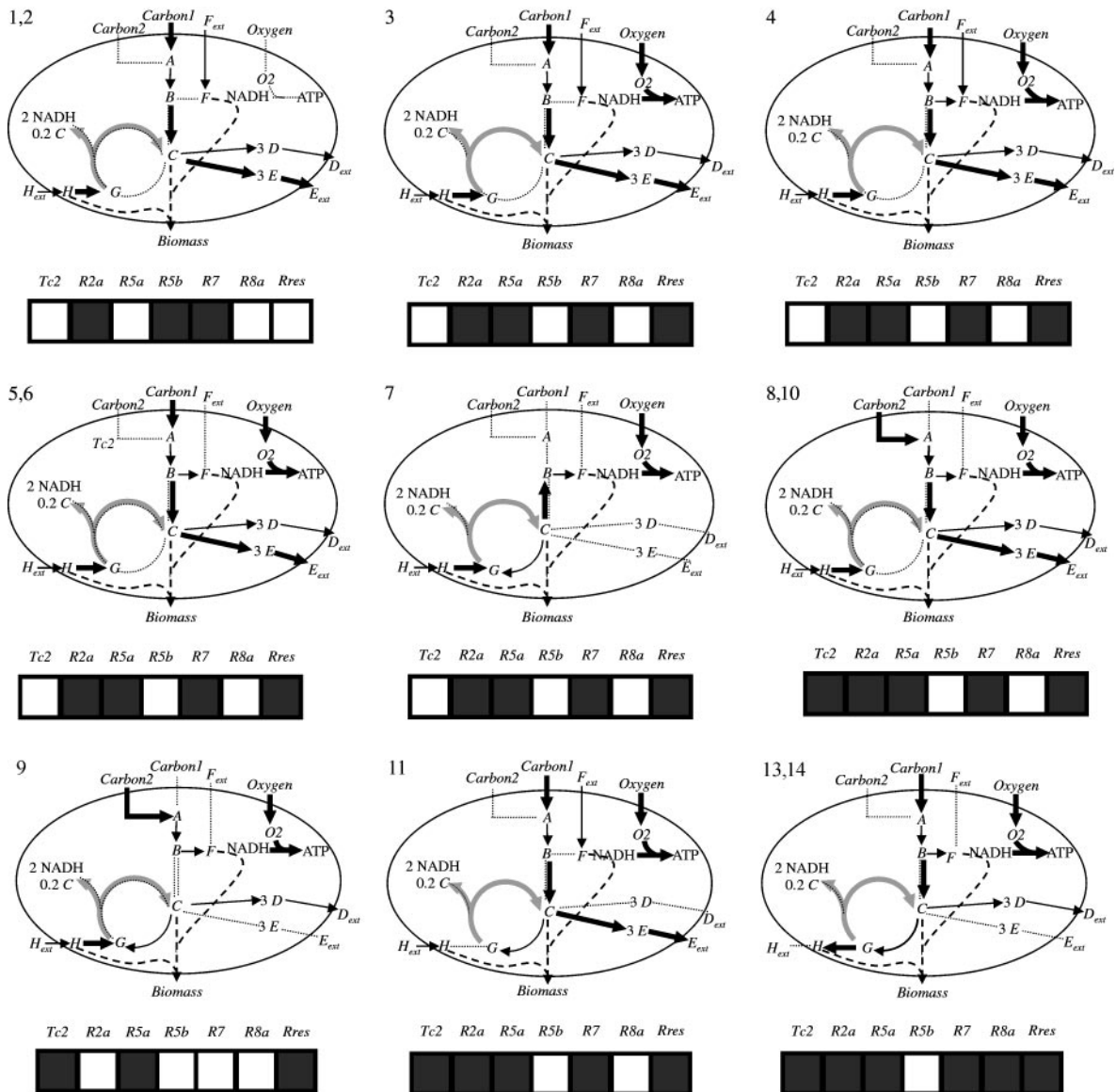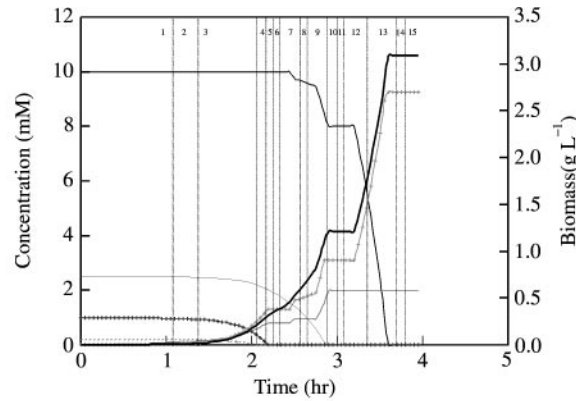
FIG. 9. Growth on complex media. Similar in format to Figs 5–8, except in the case of *R5a* and *R5b*, where the activity of these isozymes has been highlighted with a light-gray arrow in the case of *R5a* and with a dark-gray arrow in the case of *R5b*. Additionally, similar flux distributions have been grouped together as shown. (———) *C1*; (———) *C2*; (———) *D*; (·······) *F*; (———) *H*; (———) *E*; (———) *X*.

distributions, which represent unstable transitory states while the organism's regulation is changing. For example, Region 9 shows a situation where *Carbon2* is used only to fulfill the *F* biomass requirement while *H* is used as a *C*, *H* and energy source, due to the repression of *R2a*.

## Discussion

This manuscript describes the procedure for incorporating transcriptional regulatory structure into FBA to more accurately predict dynamic flux profiles of microbial growth. This procedure has major advantages over FBA in the following areas:

1. quantitative dynamic simulation of substrate uptake, cell growth and by-product secretion;
2. qualitative simulation of gene transcription events and the presence of proteins in the cell; and
3. investigation of the systemic effects of imposing temporary regulatory constraints on the solution space.

The quantitative predictions made by the combined regulatory/metabolic model are completely unpredictable using FBA alone under many conditions. Example 1 illustrates this point. The diauxic growth curve shown in Fig. 5 is a completely different result than would be obtained by FBA alone, which would incorrectly predict the maximal possible uptake of both *Carbon1* and *Carbon2*. It is interesting that the addition of one simple constraint to the solution space ($v_{Tc2} = 0$) results in such a dramatic change in the predicted phenotype. Similarly, from Example 5 it is clear that FBA alone would fail to predict which of two isozymes plays the more active role in catalysis under given conditions. Again, it is the addition of one simple constraint which results in the more correct prediction.

The combined regulatory/metabolic model is also capable of making qualitative predictions about the up- and down-regulation of enzyme production. This capability is shown in Examples 2 and 4, where FBA and the combined regulatory/metabolic approach predict similar flux distributions, but the combined approach also

predicts the down-regulation of certain enzymes which are no longer required to obtain the optimal solution. FBA is unable to consistently make such predictions due to the fact that an enzyme may be present in a cell and still have a zero flux.

Finally, this approach allows an investigation of the systemic effects of transcriptional regulation. The individual operations required to model the combined metabolic/regulatory network as described are simple. However, when combined to represent biological networks of even modest complexity, they reproduce quite complex behavioral patterns. The sample network examined here, although two orders of magnitude smaller than the metabolic networks of commonly studied bacteria, exhibits surprisingly complex behavior, as shown by the unusual intermediate flux distributions during growth on the complex medium. For a more complicated network, the multiple constraints applied to the system can cause the solution space to change dramatically in response to a changing environment.

Besides simply determining whether or not regulatory constraints are implemented, the environment also has an important influence on the regulatory constraints themselves. Unlike the non-adjustable physico-chemical constraints, regulatory constraints are biological in nature and can change with evolution. The demand theory of gene regulation indicates that the evolution of gene regulation may be governed by rules, predicting a correlation between demand for gene expression and the mode of control exhibited (Savageau, 1977, 1998a, b). This correlation has been supported experimentally (Savageau, 1989). The constraints-based approach to modeling regulatory events, by providing a framework for analysing, interpreting and predicting the systemic effects of transcriptional regulation, may therefore also contribute to our understanding of the rules which govern evolution and the subsequent effects of evolutionary forces on an organism's solution space.

The method is presented here in its simplest form. However, the use of Boolean logic to represent genetic regulatory networks qualitatively has grown in sophistication, including such features as multilevel logic variables and asynchronous updating of protein synthesis (Thomas,

1991). These features may be incorporated into this approach at some time in the future. Another important development in this approach will be its application to a real network. Databases have already been developed and made available on-line which detail the known regulation of *E. coli* (Karp *et al.*, 2000; Salgado *et al.*, 2001); such information, made available for this and other organisms, will enable the construction of genome-scale metabolic/regulatory microbial models. The ability of the approach described here to generate quantitative hypotheses which may be experimentally tested will lead to an ongoing iterative model-building process, resulting in advanced models and augmented scientific knowledge.

In summary, FBA has been used to calculate time courses of growth under certain limited conditions; however, the approach described here also accounts for the systemic effects of temporary regulatory constraints on cellular behavior. Such an approach is potentially more versatile and may be used to simulate a wider range of experimental conditions.

## REFERENCES

ARISTIDOU, A. & PENTTILÄ, M. (2000). Metabolic engineering applications to renewable resource utilization. *Curr. Opin. Biotechnol.* **11**, 187–198.

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

BLATTNER, F. R. & PLUNKETT 3RD, G. *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.

BONARIUS, H. P. J. & SCHMID, G. *et al.* (1997). Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* **15**, 308–314.

CARRIER, T. A. & KEASLING, J. D. (1999). Investigating autocatalytic gene expression systems through mechanistic modeling. *J. theor. Biol.* **201**, 25–36, doi:10.1006/jtbi.1999.1010.

CHARTRAIN, M. & SALMON, P. M. *et al.* (2000). Metabolic engineering and directed evolution for the production of pharmaceuticals. *Curr. Opin. Biotechnol.* **11**, 209–214.

COVERT, M. W. & SCHILLING, C. H. *et al.* (2001). Metabolic modeling of microbial strains in silico. *Trends Biochem. Sci.* **26**, 179–186.

DERISI, J. & PENLAND, L. *et al.* (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457–460.

EDWARDS, J. & PALSSON, B. (1999). Properties of the *Haemophilus influenzae* Rd metabolic genotype. *J. Biol. Chem.* **274**, 17 410–17 416.

EDWARDS, J. S. & PALSSON, B. O. (2000). The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences* **97**, 5528–5533.

EDWARDS, J. S. & IBARRA, R. U. *et al.* (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **19**, 125–130.

EDWARDS, J. S. & RAMAKRISHNA, R. *et al.* (1999). Metabolic flux balance analysis. In: *Metabolic Engineering* (Lee, S. Y. & Papoutsakis, E. T., eds). New York: Marcel-Deker.

FELL, D. (1996). *Understanding the Control of Metabolism.* London: Portland Press.

FELL, D. A. (1998). Increasing the flux in metabolic pathways: a metabolic control analysis perspective. *Biotechnol. Bioeng.* **58**, 121–124.

FOTHERINGHAM, I. (2000). Engineering biosynthetic pathways: new routes to chiral amino acids. *Curr. Opin. Chem. Biol.* **4**, 120–124.

GOMBERT, A. K. & NIELSEN, J. (2000). Mathematical modelling of metabolism. *Curr. Opin. Biotechnol.* **11**, 180–186.

GUARDIA, M. J. & GAMBHIR, A. *et al.* (2000). Cybernetic modeling and regulation of metabolic pathways in multiple steady states of hybridoma cells. *Biotechnol. Progr.* **16**, 847–853.

HASTY, J. & MCMILLEN, D. *et al.* (2001). Computational studies of gene regulatory networks, *in numero*: molecular biology. *Nat. Rev. Genet.* **2**, 268–279

HATZIMANIKATIS, V. & FLOUDAS, C. *et al.* (1996). Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J.* **1996**, 1277–1292.

HEINRICH, R. & SCHUSTER, S. (1996). *The Regulation of Cellular Systems.* New York: Chapman & Hall.

KARP, P. D. & RILEY, M. *et al.* (2000). The EcoCyc and MetaCyc databases. *Nucl. Acids Res.* **28**, 56–59.

KAUFFMAN, S. A. (1993). *The Origins of Order.* New York: Oxford University Press.

KAUFMAN, M. & URBAIN, J. *et al.* (1985). Towards a logical analysis of the immune response. *J. theor. Biol.* **114**, 527–561.

LEE, B. & YEN, J. *et al.* (1999). Incorporating qualitative knowledge in enzyme kinetic models using fuzzy logic. *Biotechnol. Bioeng.* **62**, 722–729.

LYNCH, A. S. & LIN, E. C. C. (1996). Responses to molecular oxygen. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F. C., ed.), Vol. 1, pp. 1526–1538. Washington, D.C.: ASM Press.

LYND, L. R. & WYMAN, C. E. *et al.* (1999). Biocommodity engineering. *Biotechnol. Progr.* **15**, 777–793.

MCADAMS, H. H. & ARKIN, A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. U.S.A.* **94**, 814–819.

MCADAMS, H. H. & ARKIN, A. (1998). Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomolecular Structure* **27**, 199–224.

MCADAMS, H. H. & ARKIN, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69.

MCADAMS, H. H. & SHAPIRO, L. (1995). Circuit simulation of genetic networks. *Science* **269**, 651–656.

OSTERGAARD, S. & OLSSON, L. *et al.* (2000). Increasing galactose consumption by *Saccharomyces cerevisiae* through metabolic engineering of the GAL gene regulatory network. *Nat. Biotechnol.* **18**, 1283–1286.

PALSSON, B. O. (2000). The challenges of in silico biology. *Nat. Biotechnol.* **18,** 1147–1150.

PATTE, J.-C. (1996). Biosynthesis of threonine and lysine. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F. C., ed.), Vol. 1, pp. 528–541. Washington, D.C: ASM Press.

PIEPER, D. H. & REINEKE, W. (2000). Engineering bacteria for bioremediation. *Curr. Opin. Biotechnol.* **11,** 262–270.

RAO, C. V. & ARKIN, A. P. (2001). Control motifs for intracellular regulatory networks. *Annu. Rev. Biomed. Eng.* **3,** 391–419.

REICH, J. G. & SEL'KOV, E. E. (1981). *Energy Metabolism of the Cell.* New York: Academic Press.

RIVETT, A. J. (1986). Regulation of intracellular protein turnover: covalent modification as a mechanism of marking proteins for degradation. *Curr. Topics Cellular Regulation* **28,** 291–337.

SAIER, M. H. J. & RAMSEIER, T. M. *et al.* (1996). Regulation of carbon utilization. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology.* (Neidhardt, F. C., ed.), Vol. 1, pp. 1325–1343. Washington, D.C: ASM Press.

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, M. A. (1977). Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl Acad. Sci. U.S.A.* **74,** 5647–5651.

SAVAGEAU, M. A. (1989). Are there rules governing patterns of gene regulation? In: *Theoretical Biology—Epigenetic and Evolutionary Order* (Goodwin, B. C. & Saunders, P. T., eds), pp. 42–66. Edinburgh: Edinburgh University Press.

SAVAGEAU, M. A. (1998a). Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *Biosystems* **47,** 9–36.

SAVAGEAU, M. A. (1998b). Rules for the evolution of gene circuitry. *Pacific Symp. Biocomput.* **7,** 54–65.

SCHILLING, C. H. & EDWARDS, J. S. *et al.* (1999). Towards metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Progr.* **15,** 288–295.

SHULER, M. L. & DOMACH, M. M. (1983). Mathematical models of the growth of individual cells. In: *Foundations of Biochemical Engineering* (Blanch, H. W., Papoutsakis, E. T. & Stephanopoulos, G., eds), p. 101. Washington, DC: American Chemical Society.

SOMOGYI, R. & SNIEGOSKI, C. A. (1996). Modeling the complexity of genetic networks: understanding multigenic and pleitropic regulation. *Complexity* **1,** 45–63.

STEPHANOPOULOS, G. & ARISTODOU, A. *et al.* (1998). *Metabolic Engineering.* New York: Academic Press.

THIEFFRY, D. & THOMAS, R. (1995). Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.* **57,** 277–297.

THIEFFRY, D. & HUERTA, A. M. *et al.* (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli. Bioessays* **20,** 433–440.

THOMAS, R. (1973). Boolean formalization of genetic control circuits. *J. theor. Biol.* **42,** 563–585.

THOMAS, R. (1991). Regulatory networks seen as asynchronous automata: a logical description. *J. theor. Biol.* **153,** 1–23.

VARMA, A. & PALSSON, B. O. (1994a). Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology* **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.

VARNER, J. D. (2000). Large-scale prediction of phenotype: concept. *Biotechnol. Bioeng.* **69,** 664–678.

VASEGHI, S. & BAUMEISTER, A. *et al.* (1999). In vivo dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae. Metabolic Eng.* **1,** 128–140, doi:10.1006/mben.1998.0110.

WIECHERT, W. & DE GRAAF, A. A. (1996). In vivo stationary flux analysis by 13C labeling experiments. *Adv. Biochem. Eng./Biotechnol.* **54,** 109–154.

WONG, P. & GLADNEY, S. *et al.* (1997). Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Progr.* **13,** 132–143.

ZUBAY, G. (1973). In vitro synthesis of protein in microbial systems. *Annu. Rev. Genet.* **7,** 267–287.