ELSEVIER

Chemical Engineering Science 59 (2004) 5041–5049

Chemical Engineering Science

www.elsevier.com/locate/ces

Cybernetic modeling of metabolism: towards a framework for rational design of recombinant organisms

Jamey Young[a], Kristene Henne[b], John Morgan[a], Allan Konopka[b], Doraiswami Ramkrishna[a,*]

[a]*School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b]*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA*

Received 3 March 2004

Abstract

Most cybernetic modeling efforts to date have taken a minimalist viewpoint of metabolism, relying heavily upon process lumping and pathway abstraction to simplify the underlying reaction network. However, such models are unsuitable for metabolic engineering applications because they do not incorporate the level of biological detail that is needed to predict the effects of genetic perturbation. In this paper, we inquire into the formulation and analysis of structured cybernetic models as tools for metabolic engineering. We present fresh theoretical developments leading to a more systematic treatment of global cybernetic variables based on conservation of metabolic resources. Finally, we illustrate these concepts with a cybernetic model-based analysis of anaerobic *Escherichia coli* metabolism. The model is able to accurately describe the growth of wild-type and acetate knockout strains when fitted to experimental data. Although it adequately predicts the end-of-batch concentrations for a strain with intermediate *pta-ackA* activity, it overestimates the measured growth and glucose uptake rates.

Keywords: Cybernetic modeling; *Escherichia coli*; Metabolism; Mathemetical modeling

1. Introduction

Chemical reaction engineers have a unique role to play in the emerging field of systems biology. The same toolset that has been developed for modeling complex reaction networks inside catalysts and chemical reactors is directly transferable to biological systems. By formulating and solving the governing balance equations, we can simulate the dynamic behavior of living cells in much the same way as engineers have traditionally studied industrial chemical processes. Instead of using these models to design a reactor, however, the field of metabolic engineering seeks to design novel organisms for specific commercial or research applications (Bailey, 1991). In most cases, implementation of a particular metabolic engineering design involves steps to alter the native enzyme levels inside a host organism or to confer the ability to produce foreign enzymes that catalyze alternate reactions. Such manipulations are enabled by modern recombinant DNA techniques for inserting, deleting, and controlling the expression of targeted genes.

One major difficulty arises in our attempts to understand and model biological systems that does not fit with our previous analogy to industrial processes, which is the issue of self-regulation. Unlike chemical reactors where we are able to fully prescribe the catalyst levels and other controlled variables, living organisms maintain survival by constantly adjusting enzyme levels and activities in response to their environments. Furthermore, since metabolic engineering objectives are oftentimes at odds with the built-in nutritional objectives of the cell, engineered organisms typically enact regulatory actions that mitigate the imposed changes. The ability to reliably model these regulatory features is prerequisite to the development of computational tools that can guide metabolic engineering design. Cybernetic models

* Corresponding author. Tel.: +1 765 494 4066; fax: +1 765 494 0805.
E-mail address: ramkrish@ecn.purdue.edu (D. Ramkrishna).

0009-2509/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ces.2004.09.037

offer great promise in this regard because they provide a systematic approach for describing metabolic regulation, even when the underlying mechanisms are poorly understood.

Unlike Flux balance analysis (FBA), which relies on pseudo-steady-state approximations and linear programming to describe metabolic fluxes (Reed and Palsson, 2003), cybernetic models are fully dynamic. Therefore, the governing balance equations are expressed as a system of ordinary differential equations, which can be integrated to give the time course of metabolite concentrations, enzyme levels, and reaction rates over a specified time interval. Cybernetic models can also predict how the system might respond to dynamically changing environmental conditions (e.g., feed rates, external concentrations, etc.) or genetic controls (e.g., transient induction of gene expression from an artificial promoter). FBA models, on the other hand, are not fully predictive because they require certain uptake rates to be specified before the remaining fluxes can be determined. As such, their capabilities are limited to predicting product yields, but not the actual production rates. Since the uptake rates are inputs to the FBA model, the model has no way to predict how these rates might be altered in response to genetic or environmental changes.

The original application of cybernetic models was to predict diauxic growth patterns in multiple substrate bacterial cultures. Kompala et al. (1984, 1986) were able to capture the experimentally verified diauxic shifts with low-dimensional models comprising minimal biological structure. Later work focused on describing the effects of maintenance metabolism in bacterial cultures subjected to low-growth situations. The work of Baloo and Ramkrishna (1991a,b) was significant in that it showed how cybernetic models incorporating maintenance effects could be used to simulate the transient response of continuous cultures following a step-up or step-down in dilution rate, both in single and mixed substrate environments. Straight and Ramkrishna (1994) expanded the cybernetic modeling framework to include more general classes of elementary units and nutritional objectives that reflect both substitutable and complementary resource competitions. Recent results by Varner and Ramkrishna (1999) and Varner (2000) have laid the groundwork necessary for extension of cybernetic models to large-scale metabolic networks. Our current work seeks to further develop the cybernetic modeling framework with an eye toward specific applications in the metabolic engineering arena.

2. Cybernetic model formulation

The central feature of cybernetic models is their ability to account for metabolic regulation by viewing the cell from an economic perspective. This idea derives from the observation that certain *limited* resources are required for the synthesis of enzymes. In the same way that a manufacturing facility requires raw materials, energy, and capital equipment to fabricate a product, the cell also requires material inputs from the pool of available amino acids, energetic inputs from the breakdown of ATP, and machinery in the form of ribosomes, RNA polymerases, etc. in order to synthesize new enzymes. Presumably, the cell's regulatory strategies have evolved in such a way as to ensure that said resources are allocated in an efficient manner. The Matching and Proportional Laws of Kompala et al. (1984, 1986), which form the central tenets of cybernetic modeling, embody the mathematical translation of this key insight.

In presenting the model formulation, there are three major classes of objects that we will refer to by name. These are *reactions*, *metabolites*, and *enzymes*. The ith reaction is referred to as R_i, where the index i belongs to the set $I = \{1, 2, \ldots, N_R\}$. Similarly, the jth metabolite is denoted M_j and the kth enzyme is denoted E_k, where the subscripts are described by $j \in J = \{1, 2, \ldots, N_M\}$ and $k \in K = \{1, 2, \ldots, N_E\}$. The concentration of M_j is written m_j. The specific rate of R_i and the specific concentration of E_k are denoted r_i and e_k, respectively.

To construct a cybernetic model, we first decompose the metabolic network into a collection of *elementary units*, which are subsets of reactions presumed to compete for a common pool of cellular resources. We denote the ℓth unit as Φ_ℓ, where the index ℓ belongs to the set $L = \{1, 2, \ldots, N_\Phi\}$. In general, these are topologically connected subnetworks embedded in the overall metabolic network and have traditionally been associated with convergent and divergent pathway nodes, cyclic pathways, or linear pathway segments, although more general network structures are not excluded (Straight and Ramkrishna, 1994). Each elementary unit is attributed a nutritional objective, and a *return on investment* is assigned to each constituent reaction in reference to this objective. The return on investment for a particular reaction is a dynamic quantity that depends on the prevailing reaction rates and metabolite concentrations at each instant in time. The Matching and Proportional Laws establish how enzyme synthesis rates and activities should depend on the computed returns. In particular, these laws specify how cybernetic u- and v-variables are to be determined for each enzyme. The u-variables describe how cellular resource is allocated toward synthesis of needed enzymes, while the v-variables describe the relative activity of these enzymes. It is through the action of these control variables that the cybernetic structure exerts its influence on model behavior, thus mimicking the role of metabolic regulation.

The rate of enzyme E_k synthesis, denoted r_{E_k}, is expressed as the sum of a constitutive and an inducible term. The constitutive rate is constant over time, while the inducible rate varies in proportion to the amount of resource allocated to E_k.

$$r_{E_k} = \alpha^*_{E_k} + \alpha_{E_k} \dot{Q}_T \mathscr{U}_k. \tag{1}$$

In this equation, $\alpha^*_{E_k}$ is the constitutive rate of E_k synthesis, $\dot{Q}_T$ is the total rate at which resource is supplied to the

metabolic network, and α_{E_k} is a proportionality constant relating the inducible synthesis rate to the rate of resource input. The *complete* cybernetic variable $\mathscr{U}_k$ is the overall fraction of $\dot{Q}_T$ allocated to E_k synthesis, and is defined in terms of local u-variables and global U-variables as follows.

$$\mathscr{U}_k \equiv \sum_{i \in I^E(k)} \sum_{\ell=1}^{N_\Phi} u_{i\ell} U_\ell. \tag{2}$$

Here, the set $I^E(k)$ contains the indices of all reactions catalyzed by enzyme E_k. The global U-variable U_ℓ is the fractional allocation of $\dot{Q}_T$ to unit Φ_ℓ. The local u-variable $u_{i\ell}$ describes the fractional sub-allocation of resource $U_\ell \dot{Q}_T$ earmarked for Φ_ℓ to reaction R_i. Thus, we envision a two-tiered process of resource allocation, wherein $\mathscr{U}_k$ is determined by regulatory controls enacted at both global and local levels.

The Matching Law relates the local u-variables within each elementary unit to the rate of return $\dot{p}_{i\ell}$ that reaction R_i contributes to the objective of unit Φ_ℓ.

$$u_{i\ell} = \frac{\dot{p}_{i\ell}}{\sum_{s=1}^{N_R} \dot{p}_{s\ell}}. \tag{3}$$

In order to compute $\dot{p}_{i\ell}$, we associate an objective function $\psi_\ell(\mathbf{m})$ with unit Φ_ℓ, which depends on the vector of metabolite concentrations. The rate of return $\dot{p}_{i\ell}$ reflects the total contribution of R_i toward improving the objective function $\psi_\ell(\mathbf{m})$. For those reactions that compete within unit Φ_ℓ we write

$$\dot{p}_{i\ell} = r_i \sum_{j=1}^{N_M} \max\left(S_{ij} \frac{\partial \psi_\ell}{\partial m_j}, 0\right), \tag{4}$$

where $\mathbf{S}$ is the stoichiometric matrix and the reaction rate r_i is determined by a suitable kinetic expression.

Previously, global cybernetic U-variables have been incorporated into models as an ad hoc mechanism for switching entire pathways "on" or "off" in response to the prevailing needs of the organism. In the current treatment, we employ a more systematic approach wherein each elementary unit is associated with a U-variable that depends explicitly on the unit's objective function value ψ_ℓ. First, we apply the Matching Law to express U_ℓ in terms of the global rates of return.

$$U_\ell = \frac{\dot{P}_\ell}{\sum_{s=1}^{N_\Phi} \dot{P}_s}. \tag{5}$$

Then, we select a function $f_\ell(\psi_\ell)$ to relate the global rate of return $\dot{P}_\ell$ to the objective value ψ_ℓ.

$$\dot{P}_\ell = f_\ell(\psi_\ell). \tag{6}$$

There may be several suitable choices for the function $f_\ell(\psi_\ell)$, so long as they result in $\dot{P}_\ell \geqslant 0$ for all valid values of ψ_ℓ. However, choosing a function that decreases monotonically with increasing ψ_ℓ leads to a complementary resource competition among the various elementary units, thereby diverting resource away from overperforming units in favor of units that are underperforming with regard to their assigned objectives. This behavior reflects the delicate balancing that the cell must carry out in order to simultaneously satisfy multiple nutritional objectives. An example of a useful function with these characteristics is described in Section 3, where it is applied to a cybernetic model of anaerobic *Escherichia coli* metabolism.

Local v-variables and global V-variables are expressed in terms of the aforementioned rates of return, which are combined to give complete $\mathscr{V}$-variables. The complete variable $\mathscr{V}_i$ multiplies r_i to account for variations in relative enzyme activity.

$$v_{i\ell} = \frac{\dot{p}_{i\ell}}{\max_{s \in I}(\dot{p}_{s\ell})}, \tag{7}$$

$$V_\ell = \min\left(\frac{f_\ell(\psi_\ell)}{f_\ell(\bar{\psi}_\ell)}, 1\right), \tag{8}$$

$$\mathscr{V}_i \equiv \frac{\sum_{\ell=1}^{N_\Phi} v_{i\ell} V_\ell}{\sum_{\ell=1}^{N_\Phi} v_{i\ell}^{\max} V_\ell^{\max}}, \tag{9}$$

where $\bar{\psi}_\ell$ is the reference value of ψ_ℓ at which V_ℓ attains its maximum value. Here, $V_\ell^{\max}$ and $v_{i\ell}^{\max}$ are the maximum possible values of V_ℓ and v_ℓ, respectively. As seen in Eqs. (7) and (8), $V_\ell^{\max} = 1\ \forall \ell \in L$ while $v_{i\ell}^{\max} = 1$ if reaction R_i competes within unit Φ_ℓ and $v_{i\ell}^{\max} = 0$ otherwise.

The metabolite balance equations are straightforward and are relegated to the Appendix to maintain brevity. The enzyme balances, however, play a central role in the model formulation and are worth special mention. They take the form

$$\frac{\mathrm{d}e_k}{\mathrm{d}t} = r_{E_k} - (\beta_k + r_G)e_k, \tag{10}$$

where β_k is the first-order degradation constant for E_k and r_G is the overall growth rate. We do not generally possess information about the absolute enzyme levels, so it is more convenient to nondimensionalize the enzyme concentration by defining a characteristic enzyme level $\bar{e}_k$ that arises in a particular metabolic reference state. We choose $\bar{e}_k$ to be the steady-state E_k concentration when $r_G = 0$, $\dot{Q}_T = 1$, and resource is allocated evenly amongst all enzymes in proportion to the number of elementary units in which they participate. In so doing, we can substitute for r_{E_k} using Eq. (1) and rewrite Eq. (10) in terms of the relative enzyme level ε_k.

$$\frac{\mathrm{d}\varepsilon_k}{\mathrm{d}t} = \beta_k \left[\mathscr{U}_k^* + (1 - \mathscr{U}_k^*)\dot{Q}_T \frac{\mathscr{U}_k}{\bar{\mathscr{U}}_k}\right] - (\beta_k + r_G)\varepsilon_k. \tag{11}$$

In this equation, $\mathscr{U}_k^*$ represents the fractional constitutive expression of enzyme E_k, which is defined by

$$\mathscr{U}_k^* \equiv \frac{\alpha_{E_k}^*}{\alpha_{E_k}^* + \alpha_{E_k}\bar{\mathscr{U}}_k}.$$

If we let $n(k)$ represent the number of elementary units in which enzyme E_k participates, the reference value $\bar{\mathscr{U}}_k$ is described by

$$\bar{\mathscr{U}}_k = \frac{n(k)}{\sum_{s=1}^{N_E} n(s)}.$$

The kinetic expression for r_i is assumed to vary directly with the relative enzyme level and to exhibit saturation dependence on all substrates. This relationship can be expressed as

$$r_i = k_i \varepsilon_{z(i)} \prod_{j \in J^-(i)} \frac{m_j}{K_{M_{ij}} + m_j}, \tag{12}$$

where k_i is the rate constant and $K_{M_{ij}}$ is the saturation constant. The set $J^-(i)$ contains the indices of the substrates of reaction R_i. The mapping $z : I \to K$ maps the index of a given reaction to the index of the enzyme that catalyzes it.

3. Cybernetic *E. coli* model

Because of the tight coupling between carbon and energy metabolism (viz., due to the predominance of substrate level phosphorylation over oxidative phosphorylation) and the diverse range of excreted products that are available for measurement, anaerobic *E. coli* cultures serve as an appealing test bed for modeling studies. The metabolic network under consideration is shown in Fig. 1. The reaction steps represent the major pathways of central carbon metabolism that are active in *E. coli* cultures growing anaerobically on glucose, which include glycolysis, pentose phosphate pathway, fermentation reactions, and certain TCA cycle reactions. Biosynthetic processes are combined into a single growth reaction R_{13} that consumes the precursor molecules PEP, PYR, R5P, OAA, AcCoA, and OGA as well as the cofactors ATP, NAD^+, and NADPH in producing the lumped portion of biomass. The precursor requirements for biosynthesis are based on values in Neidhardt et al. (1990), which have been modified to account for lumped metabolites.

The elementary units and objective functions used to construct the *E. coli* cybernetic model are listed in Table 1. Units Φ_1 and Φ_2 are assigned complementary objective functions, which ensure that fluxes leaving the glucose and PEP branch points are modulated to maintain adequate levels of the growth precursors PEP, R5P, PYR, and OAA. The scaling parameters γ_1 and γ_2 are included in these objective functions so that metabolites existing at different concentration scales are handled in a consistent manner. Units Φ_3 and Φ_4 are associated with objective functions that serve to prevent over-accumulation of PYR and OAA intermediates in the model. Units Φ_5, Φ_6, Φ_7, Φ_8, and Φ_9 ensure that levels of OGA, NAD^+, NADPH, AcCoA, and ATP are sufficient to support the needs of growth and overall metabolism. Unit Φ_{10} functions to induce the formate hydrogenlyase pathway as the extracellular concentration of formate rises, which is a well-known physiological trait of anaerobic *E. coli* cultures (Böck and Sawers, 1996). Unit Φ_{11} induces growth when ATP levels are high and represses it when ATP levels fall.

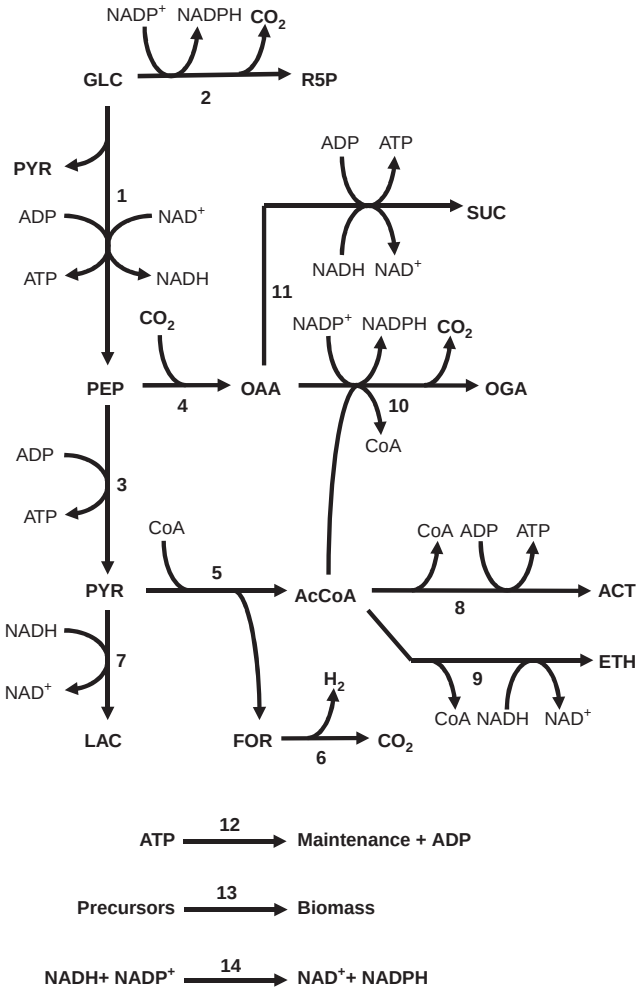

Fig. 1. Anaerobic *E. coli* central carbon metabolism. Metabolite abbreviations and associated indices: AcCoA, acetyl-CoA, 1; ACT, acetate, 2; ADP, 3; ATP, 4; CoA, 5; ETH, ethanol, 6; FOR, formate, 7; GLC, glucose, 8; LAC, lactate, 9; NAD^+, 10; NADH, 11; $NADP^+$, 12; NADPH, 13; OAA, oxaloacetate, 14; OGA, oxoglutarate, 15; PEP, phosphoenolpyruvate, 16; PYR, pyruvate, 17; R5P, ribose-5-phosphate, 18; SUC, succinate, 19.

The global rate of return for unit Φ_ℓ is computed by comparing the value of ψ_ℓ with its "target value" $\bar{\psi}_\ell$. This value serves as a reference for determining the extent to which the unit achieves its assigned objective. The function $f_\ell(\psi_\ell)$ introduced in Eq. (6) is given by

$$f_\ell(\psi_\ell) = \frac{a_\ell(\psi_\ell^{\max} - \psi_\ell)}{(2b_\ell - c_\ell)\psi_\ell + c_\ell\bar{\psi}_\ell - 2b_\ell\psi_\ell^{\min}}, \tag{13}$$

where the coefficients are chosen such that

$$a_\ell = \bar{\psi}_\ell - \psi_\ell^{\min},$$
$$b_\ell = \psi_\ell^{\max} - \bar{\psi}_\ell,$$
$$c_\ell = \psi_\ell^{\max} - \psi_\ell^{\min}.$$

Table 1
Elementary units and objective functions used in the cybernetic *E. coli* model

ℓ	Unit name	$I^{\Phi}(\ell)$	$\psi_\ell(\mathbf{m})$	$\psi_\ell^{\min}$	$\bar{\psi}_\ell$	$\psi_\ell^{\max}$
1	GLC node	{1, 2}	$m_{16}(m_{18})^{\gamma 1}$	0	3.8	∞
2	PEP node	{3, 4}	$m_{17}(m_{14})^{\gamma 2}$	0	3.8	∞
3	PYR node	{5, 7}	$-m_{17}$	$-\infty$	-4	0
4	OAA node	{11}	$-m_{14}$	$-\infty$	-0.1	0
5	OGA node	{10}	m_{15}	0	1	∞
6	NAD^+ node	{9, 11}	m_{10}	0	2	2.2
7	NADPH node	{14}	m_{13}	0	0.1	0.4
8	AcCoA node	{5}	m_1	0	0.5	0.6
9	ATP node	{8}	m_4	0	7	8
10	FOR node	{6}	$-m_7$	$-\infty$	K_{M67}	0
11	Growth	{13}	$-m_4$	-8	-7	0

The relevant component indices are given in Fig. 1. $I^{\Phi}(\ell)$ is the set of reactions competing within the ℓth unit. Parameter values for units Φ_1 and Φ_2 are $\gamma_1 = 0.175$ and $\gamma_2 = 0.025$. All concentration units are μmol/g, except m_7 is in units of mM.

Note that the numerator and denominator of this rational function are both linear in ψ_ℓ. By constructing $f_\ell(\psi_\ell)$ in this way, the resulting function smoothly interpolates the points

$$f_\ell(\psi_\ell^{\max}) = 0,$$
$$f_\ell(\bar{\psi}_\ell) = 0.5,$$
$$f_\ell(\psi_\ell^{\min}) = 1.$$

Hence, each global return is scaled such that $0 \leqslant \dot{P}_\ell \leqslant 1$, with $\dot{P}_\ell$ decreasing gradually as ψ_ℓ increases. In most cases, the parameters $\psi_\ell^{\min}$ and $\psi_\ell^{\max}$ are determined solely by the non-negativity constraints on metabolite concentrations, and Eq. (13) reduces to one of the limiting expressions that follow.

$$\lim_{\psi_\ell^{\min}\to-\infty}\ \lim_{\psi_\ell^{\max}\to 0} f_\ell(\psi_\ell) = \frac{\psi_\ell}{\bar{\psi}_\ell + \psi_\ell}, \tag{14}$$

$$\lim_{\psi_\ell^{\min}\to 0}\ \lim_{\psi_\ell^{\max}\to\infty} f_\ell(\psi_\ell) = \frac{\bar{\psi}_\ell}{\bar{\psi}_\ell + \psi_\ell}. \tag{15}$$

However, the concentration of any cofactor molecule is necessarily bounded by the total availability of the cofactor's backbone moiety, since there is no net production or consumption of this moiety due to reactions other than the growth reaction R_{13}. For example, the total ([ATP] + [ADP]) concentration is constrained by the stoichiometric yields in reaction R_{13} to be no greater than 8 μmol/g. Therefore, the objective functions of units Φ_6, Φ_7, Φ_8, Φ_9, and Φ_{11} have finite upper and lower bounds as given in Table 1.

The balance equations for all metabolites and enzymes included in the model comprise a set of 39 simultaneous ODEs. The model was simulated using the ODE15s routine of Matlab 6.5 (The Mathworks, Inc., Natick, MA). Initial values for saturation constants were taken from the BRENDA website (Schomburg, as accessed February 2004) and from Chassagnole et al. (2002). The degradation half-lives of all enzymes were set to 1 h (Thattai and van Oudenaarden, 2001), and we chose $\mathscr{U}_k^* = 0.1$ for all enzymes except for E_{12}. The maintenance reaction R_{12} is constitutively expressed in the current model, and thus we set $\mathscr{U}_{12}^* = 1$ and $\mathscr{V}_{12} = 1$. The values of γ, $\bar{\psi}$, $\psi^{\max}$, and $\psi^{\min}$ shown in Table 1 were approximated based on the available steady-state measurements provided in Chassagnole et al. (2002) and Jackowski (1996). The value of $\dot{Q}_T$ was assumed to be proportional to the sum of the global returns, as given by the equation

$$\dot{Q}_T = \frac{\sum_{\ell=1}^{N_\Phi} f(\psi_\ell)}{\sum_{\ell=1}^{N_\Phi} f(\bar{\psi}_\ell)}. \tag{16}$$

Anaerobic shake-flask experiments on glucose minimal medium were performed on two different *E. coli* strains to enable parameter estimation, a parent strain (GJT001) that exhibits wild-type glucose metabolism and a derivative *pta-ackA* knockout strain (YBS121) that cannot ferment to acetate. Cell density was monitored spectrophotometrically at 600 nm, and the concentrations of glucose and fermentation products were determined by HPLC with dual UV and refractive index detectors. The 14 rate constants as well as the saturation constants for glucose in reactions R_1 and R_2, formate in R_6, pyruvate in R_7, and ATP in reactions R_{12} and R_{13} were optimized to fit the data in Fig. 2 using the Adaptive Simulated Annealing code of Ingber (1993). The acetate knockout strain was simulated by setting the total synthesis rate and initial concentration of enzyme E_8 to zero.

Deleting the major acetate production pathway in *E. coli* represents a significant perturbation to its metabolism, which is evidenced by a drastically reduced growth rate and a complete redistribution of fermentation products away from acetate/ethanol/formate and into lactate. The model's ability to accurately capture this metabolic shift using a single set of kinetic parameters is an extremely promising result and demonstrates the ability of cybernetic models to successfully describe the effects of genetic manipulation.

4. Application to metabolically engineered strains

We envisage the cybernetic modeling approach to metabolic engineering as depicted in Fig. 3. First, the

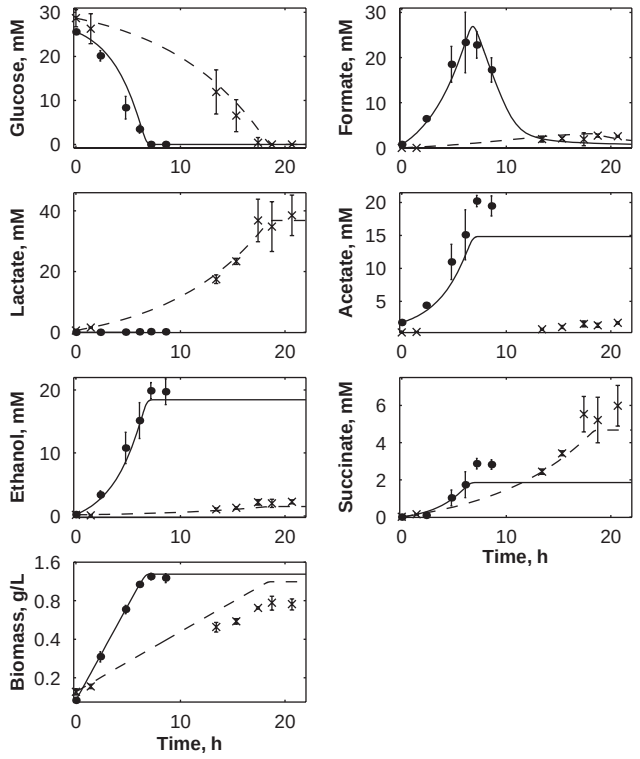

Fig. 2. Simulated behavior of GJT001 parent strain (solid lines) and YBS121 acetate knock-out strain (dashed lines) optimized to fit experimental results (GJT001, •; YBS121, ×). Error bars indicate 90% confidence intervals.

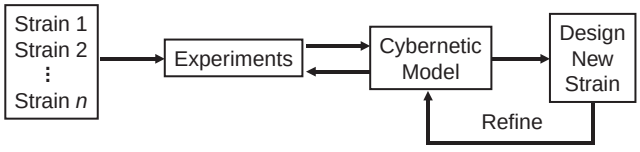

Fig. 3. Overview of cybernetic modeling approach to metabolic engineering.

cybernetic model is identified based on experimental inputs from a parent strain as well as several genetic variants. The resulting model can then be used to simulate the network response evoked by overexpression, underexpression, or complete knockout of specific enzymes. Qualitative insights gained from such an analysis serve to guide the metabolic engineering process while eliminating unnecessary rounds of strain development and characterization. Furthermore, model simulations may be used to provide quantitative prescriptions for how enzyme levels should be "tuned" in order to elicit a desired phenotype. Once an optimized strain has been designed and engineered, its experimental characterization will enable further model refinements.

In order to model recombinant organisms, we must account for differences in gene expression between the host strain and its derivatives by adjusting the enzyme synthesis parameters $\alpha^*_{E_k}$ and α_{E_k} in Eq. (1). Equivalently, one can

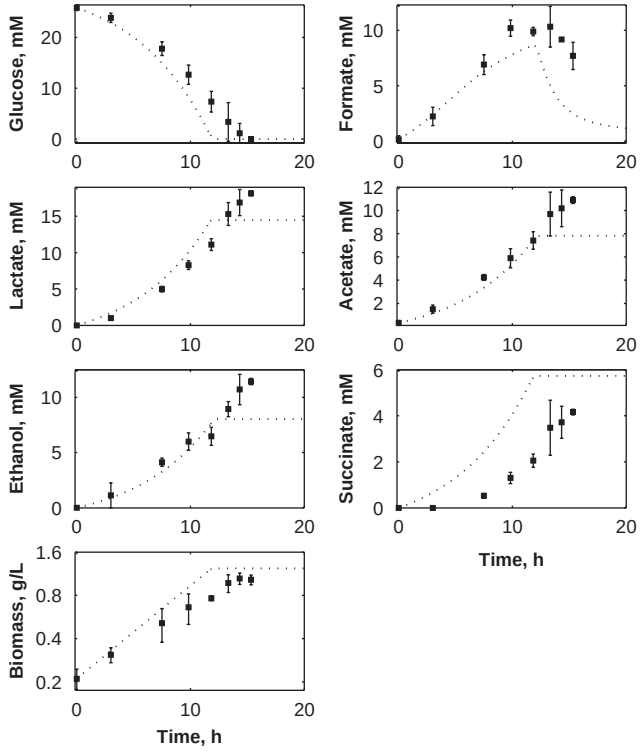

Fig. 4. Simulated behavior of strain 538-1Y (dotted lines) with comparison to experiment. Error bars indicate 90% confidence intervals.

rewrite Eq. (11) in the modified form

$$\frac{\mathrm{d}\varepsilon_k}{\mathrm{d}t} = \beta_k \left[\delta_k^* \mathscr{U}_k^* + \delta_k (1 - \mathscr{U}_k^*) \dot{Q}_T \frac{\mathscr{U}_k}{\bar{\mathscr{U}}_k} \right] - (\beta_k + r_G)\varepsilon_k, \tag{17}$$

where δ_k^* and δ_k represent the relative levels of constitutive and inducible synthesis of enzyme E_k in the recombinant strain. For example, values of $\delta_k^* < 1$ and $\delta_k < 1$ reflect attenuation of gene expression, while values greater than unity reflect overexpression.

As a trial case, we applied the cybernetic *E. coli* model of Section 3 to predict the effects of attenuating *pta-ackA* expression to levels intermediate between that of the wild-type and acetate knockout strains. To check the accuracy of this prediction, *E. coli* strain 538-1Y was constructed by re-inserting the *pta-ackA* genes into YBS121 on a high-copy plasmid, but with the native promoter replaced by a constitutive promoter of reduced strength (Jensen and Hammer, 1998). Results are presented in Fig. 4 comparing the simulated behavior of 538-1Y to experimental results obtained in shake-flasks. The simulation was accomplished by simultaneously setting δ_8 to zero and varying δ_8^* until the predicted distribution of fermentation products best matched the experimental observations, which occurred at $\delta_8^* = 1.7$. The predicted log-phase E_8 levels in 538-1Y are approximately 10.5% of those in GJT001.

The model correctly predicts a product distribution that is intermediate between those of GJT001 and YBS121. The end-of-batch concentrations are predicted to within 20–30% of the measured values in most cases, which is reasonable in light of relative standard deviations that can be 10% or higher in many experiments. However, the model overestimates the experimentally determined rates of growth and glucose uptake for strain 538-1Y, with the predicted batch time of 12 h differing significantly from the observed 15 h batch time. In fact, these rates are no faster than those exhibited by YBS121, although the distribution of fermentation products is markedly different. This is certainly an unexpected result, since the ATP-producing acetate pathway is at least partially active in 538-1Y and is completely inactive in YBS121. Plasmid burden does not seem to be a major factor leading to the less-than-expected growth rate of 538-1Y, as borne out by experiments with a control strain harboring the *pta-ackA*-free cloning vector in the YBS121 background. Ongoing research efforts are focused on understanding the cause of the growth deficiency in derivatives of YBS121, which leads to much of the current mismatch between model and experiment. One possibility is that YBS121 may have acquired additional mutations in genes other than *pta-ackA*, or that other nearby genes or regulatory regions may have been affected by the deletion of *pta-ackA*. We are currently investigating this possibility by constructing a plasmid that, when inserted into YBS121, will express *pta-ackA* at levels near those of GJT001. If the phenotype of this strain differs significantly from that of GJT001, we can conclude that inadvertent differences in the genetic backgrounds of GJT001 and YBS121 are primarily responsible for the diminished growth rate of YBS121 and its derivatives. Another possible source of mismatch might be due to the putative role of acetyl-phosphate as a global regulator involved in signal transduction pathways (Ninfa, 1996). As acetyl-phosphate is an intermediate of the acetate pathway, its level will be strongly affected by manipulation of *pta-ackA* expression.

5. Conclusion

A systematic treatment of global cybernetic variables has been developed and applied to investigate the response of *E. coli* to targeted genetic manipulations. Although it reasonably estimates the end-of-batch concentrations for a strain with intermediate *pta-ackA* activity, it over-predicts the measured growth and glucose uptake rates. This study illustrates a key difficulty in carrying out metabolic engineering design, which is that manipulations intended to boost product yield or titer can have deleterious effects on growth and overall metabolism. The current cybernetic *E. coli* model most likely falls short due to some added regulatory complexity or mutational side effect that has not been included in the model formulation. However, FBA and other approaches based on steady-state approximations cannot even offer predictions of these changing uptake rates because they are *inputs* to the model. The cybernetic approach, on the other hand, ties these uptake rates to the rest of cellular metabolism through their associated $\mathscr{U}$- and $\mathscr{V}$-variables and kinetic rate expressions. Elevating metabolic engineering to a level of precision that is on par with other engineering disciplines will require continued development of modeling strategies that predict not only yields, but overall productivities as well.

Acknowledgements

This material is based upon work supported under a National Science Foundation Graduate Research Fellowship (JY) and a National Science Foundation GOALI Grant (Award No. BES-0000961). The authors would also like to thank Prof. Ka-Yiu San of Rice University for donation of strains GJT001 and YBS121.

Appendix A

A.1. Extracellular liquid phase

In order to describe the time evolution of the various liquid-phase concentrations, we require a balance for each species present in the extracellular liquid.

$$\frac{\mathrm{d}m_j}{\mathrm{d}t} = c\sum_{i=1}^{N_R} S_{ij} r_i \mathscr{V}_i, \tag{A.1}$$

where c is the biomass concentration. The kinetic expression used to compute r_i is presented in Section 2. As also discussed in Section 2, the cybernetic variable $\mathscr{V}_i$ represents the relative activity of the enzyme that catalyzes reaction R_i.

A.2. Total biomass balance

The total biomass balance is a special case of Eq. (A.1).

$$\frac{\mathrm{d}c}{\mathrm{d}t} = c r_G. \tag{A.2}$$

The specific growth rate r_G is defined as the sum of the net specific production rates of all intracellular components.

$$r_G = \sum_{j\in J^{\mathrm{int}}} \sum_{i=1}^{N_R} h_j S_{ij} r_i \mathscr{V}_i. \tag{A.3}$$

The set J^{int} contains the indices of all intracellular metabolites. The conversion factor h_j appearing in Eq. (A.3) is required to convert the net specific production rate of M_j to the appropriate units (viz., grams of M_j per gram dry weight of biomass per unit time).

Table 2
Reaction stoichiometries and kinetic parameters used in the cybernetic *E. coli* model of Section 3

i	Reaction R_i	k_i	$K_{M_{ij}}$
1	GLC + ADP + 2NAD$^+$ → PYR + ATP + PEP + 2NADH	14 ± 2	GLC: 1 ± 3, ADP: 0.3, NAD$^+$: 0.7
2	GLC + 2NADP$^+$ → 2NADPH + CO_2 + R5P	6.774 ± 0.002	GLC: 1 ± 3, NADP$^+$: 0.02
3	PEP + ADP → PYR + ATP	16.379 ± 0.005	PEP: 0.5, ADP: 0.4
4	PEP + CO_2 → OAA	15.233 ± 0.003	PEP: 1, CO_2: 0.2
5	PYR + CoA → AcCoA + FOR	12.748 ± <0.001	PYR: 3, CoA: 0.05
6	FOR → CO_2 + H_2	3.78796 ± 0.00007	FOR: 6.638 ± 0.003
7	PYR + NADH → LAC + NAD$^+$	17.5 ± 1.2	PYR: 37.691 ± 0.011, NADH: 0.02
8	AcCoA + ADP → ACT + ATP + CoA	9.02 ± 0.03	AcCoA: 0.1, ADP: 0.1
9	AcCoA + 2NADH → ETH + 2NAD$^+$ + CoA	11.988 ± 0.003	AcCoA: 0.02, NADH: 0.08
10	OAA + AcCoA +NADP$^+$ → OGA + CO_2 + NADPH + CoA	3.7596 ± 0.0003	OAA: 0.03, AcCoA: 0.2, NADP$^+$: 0.02
11	OAA + 2NADH + ADP → SUC + 2NAD$^+$ + ATP	4.80273 ± 0.00001	OAA: 0.08, NADH: 0.1, ADP: 1
12	ATP → ADP	14.4432 ± < 0.0001	ATP: 33.7955 ± < 0.0001
13	1.9739PEP + 2.8328PYR + 1.7867OAA + 3.7478AcCoA + 1.0789OGA+ 1.6197R5P + 42.7030ATP + 3.5448NAD$^+$ + 14.2250NADPH → B + 3.7484CoA + 42.7110ADP + 3.5470NADH + 14.2254NADP$^+$	7 ± 5	PEP : 0.4, PYR : 0.4, OAA : 0.01, AcCoA : 0.05, OGA : 0.1, R5P : 0.07, ATP : 1.01457 ± < 0.00001, NAD$^+$: 0.2, NADPH : 0.01
14	NADH + NADP$^+$ → NADPH + NAD$^+$	4.361 ± 0.002	NADH: 0.002, NADP$^+$: 0.008

90% confidence intervals are indicated for all estimated parameters. Values of k_i are in units of mmol g^{-1} h^{-1}. Values of $K_{M_{ij}}$ are in units of μmol g^{-1} for intracellular species and mM for extracellular species. The component B represents the lumped portion of biomass.

A.3. Intracellular metabolite balance

The balance equation for intracellular metabolites is given by

$$\frac{\mathrm{d}m_j}{\mathrm{d}t} = \sum_{i=1}^{N_R} S_{ij} r_i \mathscr{V}_i - r_G m_j. \tag{A.4}$$

Note that here, in contrast to Eq. (A.1), m_j represents the *specific* concentration of M_j.

A.4. Model parameters

The reaction network for the cybernetic *E. coli* model is shown in Table 2 along with the values of k_i and K_{Mij} associated with each reaction. For those parameters that were estimated based on experimental data, 90% confidence intervals are provided. Initial concentrations for extracellular species were set based on the available experimental data. For intracellular species, initial concentrations were established by setting [AcCoA] = 0.5 μmol/g, [ADP] = 1 μmol/g, [ATP] = 7 μmol/g, [CoA] = 0.1 μmol/g, [NAD$^+$] = 2 μmol/g, [NADH] = 0.2 μmol/g, [NADP$^+$] = 0.3 μmol/g, [NADPH] = 0.1 μmol/g, [OAA] = 0.1 μmol/g, [OGA] = 1 μmol/g, [PEP] = 4 μmol/g, [PYR] = 4 μmol/g, [R5P] = 0.7 μmol/g, and all relative enzyme levels to unity followed by simulating the model to an internal steady-state. The resulting steady-state values were taken as initial concentrations for the simulated growth curves.

References

Bailey, J.E., 1991. Towards a science of metabolic engineering. Science 252, 1668–1674.

Baloo, S., Ramkrishna, D., 1991a. Metabolic regulation in bacterial continuous cultures: I. Biotechnology and Bioengineering 38 (11), 1337–1352.

Baloo, S., Ramkrishna, D., 1991b. Metabolic regulation in bacterial continuous cultures: II. Biotechnology and Bioengineering 38 (11), 1353–1363.

Böck, A., Sawers, G., 1996. Fermentation. In: Neidhardt, F. (Ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, second ed., vol. 1. ASM Press, Washington, DC, pp. 262–282 (Chapter 18).

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J.W., Mauch, K., Reuss, M., 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. Biotechnology and Bioengineering 79 (1), 53–73.

Ingber, L., 1993. Adaptive simulated annealing (ASA). Global Optimization C-code. Caltech Alumni Association, Pasadena, CA, URL http://www.ingber.com/#ASA-CODE.

Jackowski, S., 1996. Biosynthesis of pantothenic acid in coenzyme A. In: Neidhardt, F. (Ed.), *Escherichia coli and Salmonella*: Cellular and Molecular Biology, second ed., vol. 1. ASM Press, Washington, DC, pp. 687–994 (Chapter 44).

Jensen, P.R., Hammer, K., 1998. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Applied and Environmental Microbiology 64 (1), 82–87.

Kompala, D.S., Ramkrishna, D., Tsao, G.T., 1984. Cybernetic modeling of microbial growth on multiple substrates. Biotechnology and Bioengineering 26, 1272–1281.

Kompala, D.S., Ramkrishna, D., Jansen, N.B., Tsao, G.T., 1986. Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. Biotechnology and Bioengineering 28, 1044–1055.

Neidhardt, F.C., Ingraham, J.L., Schaechter, M., 1990. Physiology of the Bacterial Cell: A Molecular Approach. Sinauer Associates, Sunderland, MA.

Ninfa, A.J., 1996. Regulation of gene transcription by extracellular stimuli. In: Neidhardt, F. (Ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, second ed., vol. 1. ASM Press, Washington, DC, pp. 262–282 (Chapter 18).

Reed, J.L., Palsson, B.O., 2003. Thirteen years of building constraint-based in silico models of *Escherichia coli*. Journal of Bacteriology 185 (9), 2692–2699.

Schomburg, D., as accessed February 2004. BRENDA, the comprehensive enzyme information system. Online, http://www.brenda.uni-koeln.de/.

Straight, J.V., Ramkrishna, D., 1994. Cybernetic modeling and regulation of metabolic pathways. Growth on complementary nutrients. Biotechnology Progress 10 (6), 574–587.

Thattai, M., van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory networks. Proceedings of the National Academy of Sciences of the United States of America 98 (15), 8614–8619.

Varner, J.D., 2000. Large-scale prediction of phenotype: concept. Biotechnology and Bioengineering 69 (6), 664–678.

Varner, J., Ramkrishna, D., 1999. Metabolic engineering from a cybernetic perspective: 1. Theoretical preliminaries. Biotechnology Progress 15, 407–425.