

5 Toward Modeling of Metabolic Networks

In research on bacteria metabolism we have indeed much the same position as an observer trying to gain an idea of the life of a household by careful scrutiny of the persons and material arriving or leaving the house; we keep accurate records of the foods and commodities left at the door and patiently examine the contents of the dust-bin and endeavour to deduce from such data the events occurring within the closed doors.

(Marjorie Stephenson, 1930)

The early development of cybernetic models (Kompala et al., 1984, 1986; Ramkrishna, 1983) began with viewing metabolism with gross networks. The more detailed models that were discussed in Chapter 2 were based on *ad hoc* formulations without a general prescription and hence could not be readily extended to larger networks. With the advent of metabolic engineering in the 1990s, modeling of large metabolic networks became a problem of vigorous engagement. Researchers sought to make genetic changes in microbes to alter their metabolism to produce specific products of interest. This called for a comprehensive understanding of pathways and how their regulatory mechanisms navigated the flow of metabolites through a maze of reactive paths. Clearly, understanding the complexity of such regulated flow is a forbidding task and, not surprisingly, modelers have resorted to various simplifying approaches, some with stated justifications and some without them. In this quest, we state forthwith the need for a dynamic approach without which there would be no way to introduce the role of *productivity*, the quantity of prime interest in any engineering endeavor.

Extending Straight's analysis (1991) of a metabolic network by parsing it into pathway units with regulation based on local goals ran into difficulties as such goals could not relate unambiguously to an overall (global) metabolic goal. The resolution of this difficulty was eventually to be found in the use of network decompositions that have come to be known as *Metabolic Pathway Analysis*. The reader, interested in the general background of this field, has available a spate of publications in the literature (Clarke, 1988; Schilling et al., 2000; Schuster and Hilgetag, 1994; Schuster et al., 2000, 2002; Trinh et al., 2009). Among many different concepts of metabolic pathway, our focus here is on an *elementary (flux) mode* (EM). While mathematically precise definitions exist, we will convey here its conceptual essence without distraction. An EM may be understood as a minimal sequence of reactions beginning with the uptake of a substrate from outside the cell followed by a sequence of intracellular reactions and terminating in an extracellular product with all intracellular intermediates at steady state. The term

“minimal” implies that if any reaction in the sequence is excluded no flux can occur through the sequence. The foregoing definition of an EM is somewhat restrictive in that it is not essential that the reaction sequence must necessarily involve the uptake of a substrate or excretion of a product. However, the discussion of cybernetic models in this chapter is better served by avoiding the nuances associated with the many different characterizations in the literature of EMs and their offshoots which are motivated by various considerations of pathway analysis that can be postponed for the present. Note that because of an EM being a sequential pathway, steady state implies that all the fluxes through it are fully determined up to a multiplicative constant. Thus the uptake rate of the substrate through a specific EM determines all of the fluxes through that pathway, thereby providing a quantitative characterization of the mode as a mathematical vector.

Metabolism involves the interplay of intracellular and extracellular variables. The intracellular variables are generally assumed to be under “quasi” steady state, as described by the following equation

$$\frac{d\mathbf{m}}{dt} = \mathbf{S}_m \mathbf{r} = \mathbf{0} \quad (5.1)$$

where \mathbf{m} is the specific concentration vector of intracellular species, the term “specific” connoting amount per unit biomass; \mathbf{r} is the metabolic flux vector which includes exchange fluxes that feature extracellular species, and intracellular fluxes involving only intracellular metabolites. It is well to remember that Eq. (5.1) has been stripped of the “dilution” term due to growth as it is usually negligible. The steady state theories seek to describe metabolism in terms of the metabolic flux vector \mathbf{r} so that its prime pursuit is the solution of the linear, homogeneous equations (5.1). The extracellular fluxes are generally obtainable by temporal measurements of the (well-stirred) extracellular environment. However, the deficit relating to the considerably smaller number of equations than unknowns in \mathbf{r} remains unresolved at this stage. The different theories to overcome this issue will be dealt with in an upcoming chapter. For the present, we address issues toward the extension of cybernetic models to metabolic networks.

The matrix \mathbf{S}_m contains stoichiometric coefficients of all intracellular reaction species (metabolites). Thus $S_{m,ij}$, the ij^{th} coefficient of \mathbf{S}_m , represents the stoichiometric coefficient of the i^{th} metabolite in the j^{th} reaction. For a metabolic network, one encounters a large number of rows and columns in \mathbf{S}_m . Since the reaction rates are intrinsic, the irreversible reactions form nonnegative components of the rate vector \mathbf{r} while the components from reversible reactions can be positive or negative. Eq. (5.1) shows that for any solution vector \mathbf{r} , the vector $\alpha \mathbf{r}$ is also a solution for any positive α as the distribution of signs among the reaction components is unaffected. Thus Eq. (5.1) satisfies the mathematical definition of a cone. EMs are a superset of the convex basis of the cone. Thus every convex combination of such a set of EMs is a solution to Eq. (5.1).

The cybernetic approach to a metabolic network aims to combine EMs in such a way that the defined metabolic goal such as maximizing growth rate or uptake of substrate is

accomplished. This metabolic goal may be referred to as the organism's *global* goal. In addition, *local* goals applicable to the control of individual reactions in each EM will be employed toward reinforcing the global goal. Identification of the set of EMs is therefore a requirement for the implementation of such a model framework. The computation of EMs has been discussed at length by Gagneur and Klamt (2004). The number of EMs rises precipitously for large networks. Thus genome-scale networks can produce more than millions of EMs.

We will present in the following section the development of the cybernetic framework for a general metabolic network (Young, 2005; Young et al., 2008) followed by its application to anaerobic growth of *E. coli*.

5.1 Cybernetic Modeling of Metabolic Networks

Since the optimal choice for meeting the metabolic goal is by distributing substrate uptake among different EMs, we begin with focusing on control of a single EM so as to maximize its throughput (Young et al., 2008). Toward this end, Young (2005) defines the concept of "extent of advancement" through an EM, so that maximization of its throughput is accomplished through an optimal control problem maximizing the extent of advancement. A further global control is imposed by combining local and global cybernetic variables.

5.1.1 Model Formulation

Recall the state variable vector for metabolism $\mathbf{y} = [\mathbf{x}^T, \mathbf{e}^T, c]^T$ introduced in Section 4.1, where $\mathbf{x} = [\mathbf{s}^T, \mathbf{m}^T]^T$ represents the concentration vector of extracellular metabolites (\mathbf{s}) and specific concentration vector of intracellular metabolites (\mathbf{m}), \mathbf{e} is the vector of enzyme levels, and c is the biomass concentration. The components of \mathbf{e} also represent specific concentrations. The mass balance equations are written as

$$\frac{d\mathbf{s}}{dt} = \mathbf{S}_s \mathbf{r} c \quad (5.2)$$

$$\frac{d\mathbf{m}}{dt} = \mathbf{S}_m \mathbf{r} - \mu \mathbf{m} \quad (5.3)$$

$$\frac{d\mathbf{e}}{dt} = \boldsymbol{\alpha} + \mathbf{diag}(\mathbf{u}) \mathbf{r}_E - \mathbf{diag}(\boldsymbol{\beta}) \mathbf{e} - \mu \mathbf{e} \quad (5.4)$$

$$\frac{dc}{dt} = \mu c \quad (5.5)$$

where \mathbf{r} denotes the fully regulated reaction rate $\mathbf{diag}(\mathbf{v}) \hat{\mathbf{r}}$, \mathbf{S}_s and \mathbf{S}_m represent the rows of the stoichiometric matrix \mathbf{S} corresponding to \mathbf{x} and \mathbf{m} , respectively. Following earlier notation, $\mathbf{diag}(\mathbf{vector})$ represents a diagonal matrix with the components of the vector along its diagonal. The cybernetic variables \mathbf{u} and \mathbf{v} are, as on all previous occasions, concerned with the regulation of enzyme synthesis and enzyme activity respectively. The specific growth rate μ can be expressed as

$$\mu = \mathbf{h}^T \mathbf{S}_m \mathbf{r} \quad (5.6)$$

The vector \mathbf{h} in Eq. (5.6) has as components the conversion factors required to express each metabolite concentration as a weight fraction of biomass. The kinetic expressions of partially regulated reaction rate for components of $\hat{\mathbf{r}}$ and \mathbf{r}_E are given by

$$\hat{r}_j = k_j e_j \prod_{i \in I^-(j)} \frac{x_i}{K_{ij} + x_i} \quad (5.7)$$

$$r_{E,j} = k_{E,j} b \prod_{i \in I^-(j)} \frac{x_i}{K_{ij} + x_i} \quad (5.8)$$

where $I^-(j)$ is the set of metabolite indices associated with the substrates of the j^{th} reaction, i.e., $I^-(j) = \{i : S_{ij} < 0\}$. The parameter b appearing in Eq. (5.8) denotes the fraction of biomass ascribed to the enzyme synthesis machinery. It represents the specific concentration of a “lumped” biomass component B including DNA, RNA, protein, lipid, and other core biomass constituents. A balance equation for b will also be a feature of the model in a manner akin to that for the resource variable R used by Baloo and Ramkrishna (1991a).

Cybernetic Laws

The cybernetic laws remain to be incorporated in the model for which, per the discussion at the beginning of this chapter, the EMs must be identified. This identification would require specification of the metabolic network. For the development of the model framework, however, we may grant that the set of EMs has been determined without having to summon a specific network. The relative flux pattern for the k^{th} EM may be denoted as \mathbf{z}_k . As some of the components of \mathbf{z}_k may be zero (since reactions not included in the EM would have flux zero), it is convenient to define a set $P(\mathbf{z}_k)$ as the “support” of \mathbf{z}_k which will represent reactions in the mode with non-zero flux. The EM matrix \mathbf{Z} will possess the vectors \mathbf{z}_k as its columns. Clearly, the matrix \mathbf{Z} will be determined entirely by the network stoichiometry.

Young’s (2005) control strategy for metabolic regulation first seeks at each instant the optimal distribution of substrate among the various EMs based on their capacity to contribute to the global metabolic goal. This goal was chosen to be the maximization of biomass or a part of it essential for survival such as the component B . The associated cybernetic variables were denoted \mathbf{u}' for enzyme synthesis and \mathbf{v}' for enzyme activation. Second, the resources made available to any EM were shared in such a way as to prevent the possibility of any reaction throttling the expected flux throughput because of a diminished level or activity of the enzyme catalyzing that reaction. The cybernetic variables governing this control of a specific EM are denoted \mathbf{u}'' and \mathbf{v}'' . Thus the overall cybernetic variables \mathbf{u} and \mathbf{v} , appearing in Eqs. (5.2)–(5.5) may be written as

$$\mathbf{u} = \mathbf{U}'' \mathbf{u}' \quad (5.9)$$

where \mathbf{U}'' is a matrix which has along its rows the cybernetic variables \mathbf{u}''_k with k representing the k^{th} mode.

$$\mathbf{v} = \frac{\mathbf{V}''\mathbf{v}'}{\|\mathbf{V}''\mathbf{v}'\|_\infty} \quad (5.10)$$

Eq. (5.10) assures us that $\|\mathbf{v}\|_\infty = 1$ as required. Further, it implies that the cybernetic variables \mathbf{v} are a nonnegative linear combination of the local \mathbf{v}'' control vectors. Thus the resulting (fully regulated) flux distribution is a convex combination of the locally optimized flux vectors. It must be borne in mind here that at any instant the actual fluxes in an EM may not be at steady state and that the targeted role of the cybernetic variables \mathbf{u}'' and \mathbf{v}'' is to facilitate a quick approach to steady state. How this is to be accomplished will be the domain of the optimal control problem to be formulated for both sets of cybernetic variables $(\mathbf{u}', \mathbf{v}')$ and $(\mathbf{u}'', \mathbf{v}'')$ which will be addressed next.

Global Control: For a specific metabolic goal, the regulatory machinery is viewed to select, at each instant t , pathway options (EMs) based on the current metabolic state and availability of nutrients in the environment.

As indicated earlier, we let \mathbf{u}' and \mathbf{v}' be the *global* cybernetic variables which control the enzyme induction and activation of all the EMs. Thus u'_k is the fraction of transcriptional resources allocated to synthesize all enzymes in the k^{th} EM, while v'_k controls the activities of these enzymes.

Toward computing the global cybernetic variables \mathbf{u}' and \mathbf{v}' , we construct a surrogate dynamical system model

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}(\mathbf{y}, \mathbf{u}', \mathbf{v}') \quad (5.11)$$

by replacing the reaction term \mathbf{r} on the right hand sides of Eqs. (5.2) and (5.3) with $\mathbf{Z}\text{diag}(\mathbf{v}')\mathbf{q}$. Eq. (5.11) is to be noted for the replacement of the cybernetic variable \mathbf{v} by the new notation \mathbf{v}' intended for this context and the metabolic vector $\hat{\mathbf{r}}$ by a vector \mathbf{q} whose k^{th} component q_k represents the “composite” flux of the k^{th} EM. This composite flux is meant to represent the contribution of the entire k^{th} EM to the global goal of the organism. Young et al. (2008) modeled this as the harmonic mean of $(\hat{r}_j v'_j / z_{jk}) \forall j \in P(z_k)$ by which is meant

$$q_k(\hat{\mathbf{r}}_k, \mathbf{v}'') = \frac{|P(\mathbf{z}_k)|}{\sum_{j \in P(\mathbf{z}_k)} (z_{jk} / \hat{r}_j v'_j)} \quad (5.12)$$

where $|P(\mathbf{z}_k)|$ represents the cardinality of the set $P(\mathbf{z}_k)$, i.e., the number of non-zero fluxes in the k^{th} EM. The ratio $\hat{r}_j v'_j / z_{jk}$ represents the involvement of the j^{th} reaction in the k^{th} mode; we view this reaction as being subject to local control through cybernetic variables in the vector \mathbf{v}'' , which appears in the argument of q_k besides $\hat{\mathbf{r}}_k$ meant to represent the restriction of $\hat{\mathbf{r}}$ to the k^{th} mode. We will pursue this further but for the present return to the consideration of v'_k , the control on the entire k^{th} EM relative to other EMs. Since this is contingent on the realization of the global metabolic objective, we will evaluate the returns-on-investment from the different EMs by application of cybernetic

laws derived in the previous chapter. Denoting the metabolic objective function by $\psi'(\mathbf{y})$ and following the development in the previous chapter we may write

$$\mathbf{p} = \mathbf{B}^T \nabla_{\mathbf{y}} \psi' \tag{5.13}$$

where the matrix $\mathbf{B} \equiv \nabla_{\mathbf{v}'} \mathbf{F}$ represents the effect of control changes made to the EMs on the system dynamics. The global cybernetic variables \mathbf{u}' and \mathbf{v}' are obtained as (Young, 2005; Young and Ramkrishna, 2007)

$$\mathbf{u}' = \frac{\mathbf{p}^+}{\|\mathbf{p}^+\|_1} \tag{5.14}$$

$$\mathbf{v}' = \frac{\mathbf{p}^+}{\|\mathbf{p}^+\|_\infty} \tag{5.15}$$

where the elements of \mathbf{p}^+ are given by

$$p_k^+ = \max(p_k, 0) \tag{5.16}$$

Next we attend to the local cybernetic variables that are concerned with local control in the EM.

Local Control: This control strategy aims to reinforce the flux throughput across modes preferred by the global control system because of high performance for the global objective. Thus it circumvents problems encountered with extending Straight’s approach to networks by preventing flux-throttling bottlenecks due to low enzyme levels and activities anywhere in the EM. For application of the cybernetic laws we return to the metabolic model restricted to the reactions in the EM. Thus we have

$$\frac{d\mathbf{y}_k}{dt} = \mathbf{f}(\mathbf{y}_k, \mathbf{u}'', \mathbf{v}'') \tag{5.17}$$

The right hand side of Eq. (5.17) is different from the right hand side of Eq. (5.11) because the former has (i) the local cybernetic variables and (ii) the reactions other than those in the EM are excluded, which is also reinforced by the subscript k on \mathbf{y} . This model is augmented by a differential equation for the rate of change in the extent of advancement defined by

$$\frac{d\xi_k}{dt} = q_k(\hat{\mathbf{r}}_k, \mathbf{v}'') \tag{5.18}$$

which is coupled to Eq. (5.17) without reverse coupling. The cybernetic variables \mathbf{u}'' and \mathbf{v}'' have not been affixed with the subscript k as the process is only just afoot to identify them. Let the objective function to be maximized for the k^{th} mode be denoted by $\psi''(\mathbf{y}_k, \xi_k)$. Following the development in the previous chapter for the v -variables, we can obtain the return-on-investment for the k^{th} mode as.

$$\mathbf{p}_k = \mathbf{B}_k^T \mathbf{e}^{\mathbf{A}_k^T \Delta t} (\nabla_{\mathbf{y}} \psi'') \tag{5.19}$$

where $\mathbf{A}_k \equiv \nabla_{\mathbf{y}_k} \mathbf{f}$, and $\mathbf{B}_k \equiv \nabla_{\mathbf{v}''} \mathbf{f}$, distinct from the same symbol used in Eq. (5.11) for global control, can be determined from linearization of the right hand side of Eq. (5.17).

The linearization is performed about $(\mathbf{y}(t), \mathbf{u}''^o, \mathbf{v}''^o)$. The reference control $(\mathbf{u}''^o, \mathbf{v}''^o)$ are chosen so that there is no regulation. Thus

$$u_i''^o = \begin{cases} \frac{1}{|P(\mathbf{z}_k)|}, & \text{if } i \in P(\mathbf{z}_k) \\ 0, & \text{otherwise} \end{cases} \quad (5.20)$$

$$v_i''^o = \begin{cases} 1, & \text{if } i \in P(\mathbf{z}_k) \\ 0, & \text{otherwise} \end{cases} \quad (5.21)$$

Young (2005) observes that the foregoing reference choice ensures that the computed cybernetic variables \mathbf{u}_k'' and \mathbf{v}_k'' (the suffix k has been added as they are associated with the control of the k^{th} EM) are not unduly biased by the reference inputs. If we set the local goal as maximum advancement of flux through the k^{th} mode in time Δt , then we may have $\psi'' = \xi_k$, so that $\nabla_{\mathbf{y}} \psi'' = [0, 0, \dots, 0, 1]^T \equiv \mathbf{1}^T$. The cybernetic variables are obtained from the linearized version of the right hand side of Eq. (5.18). The process, after some extended algebra available in Young (2005), will precisely identify the returns-on-investment vector \mathbf{p}_k^+ in the following expressions for the local cybernetic variables.

$$\mathbf{u}_k'' = \frac{\mathbf{p}_k^+}{\|\mathbf{p}_k^+\|_1}, \quad \mathbf{v}_k'' = \frac{\mathbf{p}_k^+}{\|\mathbf{p}_k^+\|_\infty} \quad (5.22)$$

Young et al. (2008) consider two policies, (i) the temperate policy with $\Delta t > 0$ and (ii) the greedy policy implying $\Delta t = 0$. We discuss two applications. The first is a simple linear pathway and the second to anaerobic growth of *E. coli* using its central carbon metabolism. For the case of $\Delta t > 0$, the matrix exponential was evaluated using the Padé approximation

$$\mathbf{e}^{\mathbf{A}\Delta t} \approx (\mathbf{I} - \frac{1}{2}\mathbf{A}\Delta t)^{-1}(\mathbf{I} + \frac{1}{2}\mathbf{A}\Delta t) \quad (5.23)$$

The time interval Δt for the temperate policy was chosen by Young (2005) as $1/\rho(\mathbf{A})$ where $\rho(\mathbf{A})$ is the logarithmic norm of \mathbf{A} (Ström, 1975; Young and Ramkrishna, 2007).

5.1.2 Modeling of a Simple Linear Pathway

Although our undertaking in this chapter was to apply cybernetic models to metabolic networks, it is useful to begin with a simple linear pathway to show how the methodology in this chapter provides a resolution of the difficulties behind the extension of the pathway unit approach to large networks. Clearly the linear pathway in Figure 5.1 has only one EM and has no room for a global objective.

The intracellular component M is expected to be low in amount relative to B . The model equations are readily written as follows

$$\frac{ds}{dt} = -v_1 \hat{r}_1 c \quad (5.24)$$

$$\frac{dm}{dt} = v_1 \hat{r}_1 - v_2 \hat{r}_2 - \mu m \quad (5.25)$$

Table 5.1 Model parameters for linear pathway of Figure 5.1. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

i	k_i	K_i	$k_{E,i}(h^{-1})$	$\beta_i(h^{-1})$	$e_i(0)$
1	5	1g/L	1	0.05	0.1
2	1	10^{-3} g/g	1	0.05	0.1

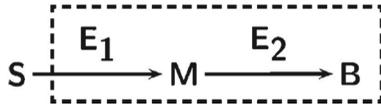


Figure 5.1 A simple linear pathway for cybernetic control. M and B are intracellular. B represents the core biomass component. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

$$\frac{db}{dt} = v_2 \hat{r}_2 - \mu b \tag{5.26}$$

$$\frac{de_1}{dt} = r_{E,1} u_1 - (\beta_1 + \mu) e_1 \tag{5.27}$$

$$\frac{de_2}{dt} = r_{E,2} u_2 - (\beta_2 + \mu) e_2 \tag{5.28}$$

$$\frac{dc}{dt} = \mu c \tag{5.29}$$

with kinetics as given below

$$\hat{r}_1 = k_1 e_1 \frac{s}{(K_1 + s)}, \quad \hat{r}_2 = k_2 e_2 \frac{m}{(K_2 + m)} \tag{5.30}$$

$$r_{E,1} = k_{E,1} b \frac{s}{(K_2 + s)}, \quad r_{E,2} = k_{E,2} b \frac{m}{(K_2 + m)} \tag{5.31}$$

The growth rate μ is the net expansion of the biophase given by

$$\mu = v_1 \hat{r}_1 \tag{5.32}$$

The model parameters chosen by Young and Ramkrishna (2007) and Young et al. (2008) are presented in Table 5.1.

Simulations were reported for initial conditions $s(0) = 5\text{g/L}$, $m(0) = 1 \times 10^{-3}\text{g/g}$, $b(0) = 0.999\text{g/g}$, $c(0) = 0.01\text{g/L}$ for four different model choices. We consider four different control policies for comparison. The first two are after Straight and Ramkrishna (1994a,b) that will be referred to as SR-1 and SR-2, while the remaining two are due to Young and Ramkrishna denoted YR-1 (based on the temperate policy) and YR-2 (based on the greedy policy). SR-1 uses the cybernetic variables

$$\mathbf{u} = \frac{\hat{\mathbf{r}}}{\hat{r}_1 + \hat{r}_2}, \quad \mathbf{v} = \frac{\hat{\mathbf{r}}}{\max(\hat{r}_1, \hat{r}_2)} \tag{5.33}$$

the matching law for enzyme synthesis and the proportional law for enzyme activation. These laws arise from maximizing the *sum* of the amounts of M and B. SR-2 uses the alternative set of cybernetic variables

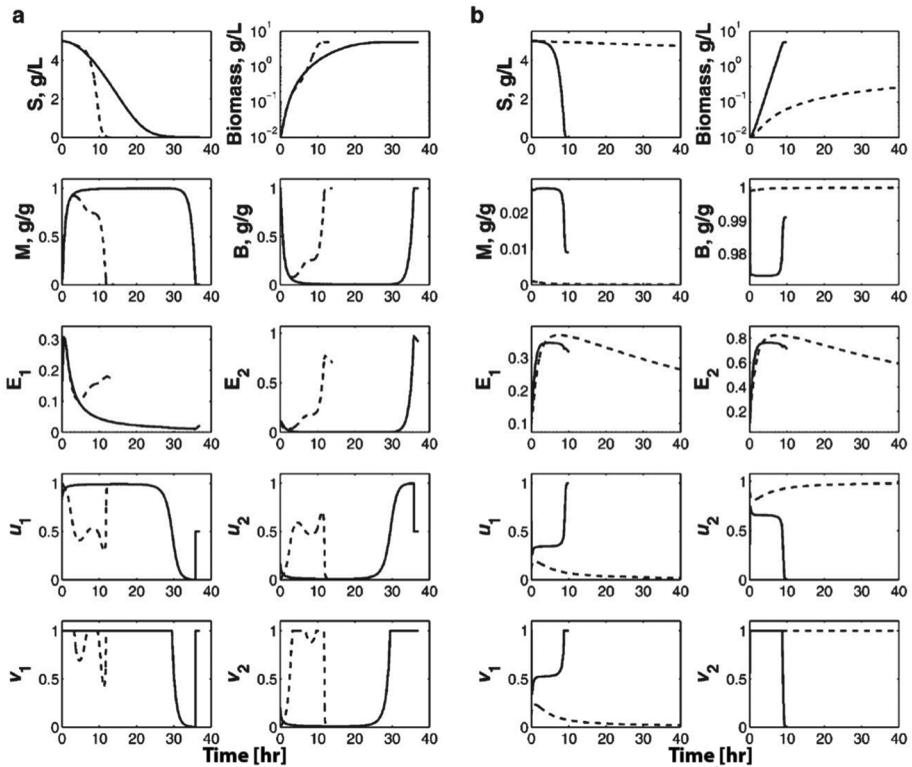


Figure 5.2 Comparison of prediction of a simple linear pathway of Section 5.1.2 by models (a) SR-1 (—) and SR-2 (---) and (b) by YR-1 (—) and YR-2 (---). Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

$$\mathbf{u} = \frac{\mathbf{p}}{p_1 + p_2}, \quad \frac{\mathbf{p}}{\max(p_1, p_2)} \tag{5.34}$$

where the components of \mathbf{p} are given by $p_1 = \hat{r}_1/m$ and $p_2 = \hat{r}_2/b$, which arise from maximizing the *product* of the amounts of M and B. We note, however, that these formulations here were used by Straight and Ramkrishna only for diverging units.

For YR-1 and YR-2, the single EM vector for the pathway in Figure 5.1, given by $\mathbf{z} = [1, 1]^T$ has the composite flux

$$q = \frac{2}{1/\hat{r}_1 + 1/\hat{r}_2} \tag{5.35}$$

The predictions by models SR-1 and SR-2 are presented in Figure 5.2a alongside those by YR-1 and YR-2 in Figure 5.2b. Consider first the predictions of the Y-R models. For the temperate policy, growth is complete in 10hrs before M starts to deplete, while for the greedy policy, growth is stymied as the level of M drops to values too low to recover. On the other hand, SR-1 starts to fill the cell with M as B starts to drop precipitously thus producing unrealistic cell compositions. Figure 5.2 makes for interesting study in various ways but it clearly emerges that YR-1 is more realistic than all the other models.

This merit is due to the anticipatory quality of the returns-on-investment which properly assesses the direction in which control action must be taken. Thus the temperate policy upregulates reaction 1 to maintain a higher level of M to realize the maximum growth potential while the greedy policy makes a premature switch to the second reaction and falls short of producing the highest growth. The enzyme level profiles in particular are markedly different for different models.

5.1.3 Modeling of Anaerobic Metabolism of *Escherichia coli*

We will now be concerned with the application of the cybernetic framework due to Young (2005) (see also Young and Ramkrishna, 2007; Young et al., 2008) to modeling the anaerobic metabolism of *E. coli*. Growth of the bacteria is accompanied by the formation of a variety of fermentation products such as formate, acetate, lactate, succinate, ethanol, CO_2 , and H_2 . The reaction network is that of central carbon metabolism which is presented in Figure 5.3 below.

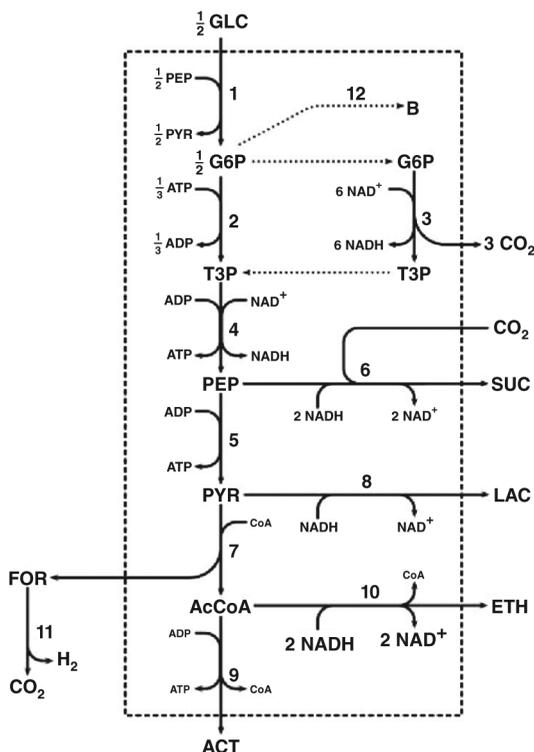


Figure 5.3 Metabolic network (central carbon metabolism) for modeling anaerobic metabolism of *E. coli*. Cofactor molecules are shown in grey. AcCoA: acetyl-CoA; ACT: acetate; ACTN: acetoin; ETH: ethanol; FOR: formate; G6P: glucose-6- phosphate; GLC: glucose; LAC: lactate; OAC: oxaloacetate; PEP phosphoenolpyruvate; PYR: pyruvate; SUC: succinate; T3P: triose-3-phosphate. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

Table 5.2 displays the various reactions in the network. Reactions 1, 2, 4, and 5 are glycolysis reactions, while 3 represents pentose phosphate pathway; reactions 6-11 represent the various fermentation pathways. Reaction 12 is a reflection of the assumption that all of the core biomass arises from G6P. The stoichiometric coefficients for G6P and NADH in this reaction were estimated by letting B be the molecular formula $\text{CH}_{1.8}\text{O}_{0.5}\text{N}_{0.2}$ (Stephanopoulos et al., 1998). The reader is referred to Young (2005) for a more detailed discussion of the reactions in the chosen network.

Model Equations

The mass balance equations Eqs. (5.24)–(5.29) with kinetics as appearing below them are applicable to the model in question. There are, however, some additional considerations that must be made because of kinetic inhibitory effects in reaction (10) which call for modifying both r_{10} and $r_{E,10}$ with the multiplicative (inhibitory) factor $K'_{10,\text{PYR}}/(x_{\text{PYR}} + K'_{x,10,\text{PYR}})$. Justification for such inhibition by pyruvate, otherwise unknown, is available from Yang et al. (2001). Young et al. (2008) also point out the need for introducing into r_{12} , the rate of biosynthesis, inhibition due to G6P, PEP, PYR, and T3P, which is accomplished by multiplying by the four inhibition terms $2x'_j/(x'_j + K'_{12,j})$, $j = \text{G6P, PEP, PYR, and T3P}$ to arrive at r_{12} and $r_{E,12}$. With a view to exploring the model capabilities for describing the behavior of multiple mutants, Young et al. (2008) report on two engineered strains, a *pta-ackA ldhA* double knockout strain (GJT001) that cannot ferment to acetate or lactate, and an *adhE* single knockout strain (YBS121) that cannot ferment to ethanol. Experimental data obtained by Young (2005) based on the fractional carbon yield for different fermentation products are presented in Figure 5.4. The anaerobic network shown in Figure 5.3 has 8 growth-associated EMs (see Table 5.3) which are the only ones that are needed as the global objective involves only maximizing core biomass. Table 5.4 shows the EMs in terms of net conversion of substrate to fermentation products and biomass (Young et al., 2008).

The cybernetic control laws are incorporated per the development in Section 5.1.1 using the global objective of maximizing $\psi' = bc$, which represents maximization of the production of core biomass component. Since reaction 11 is not involved in growth associated modes, its regulation is described by kinetics alone and by setting the cybernetic variables $u_{11} = v_{11} = 1$ as it is not in competition with other network reactions.

The model parameters were fitted by Young (2005) by using growth and fermentation data on both strains GJT001 and YBS121. We omit the details available in Young et al. (2008) in favor of a more general discussion on the topic of parameter estimation elsewhere in this book. Figure 5.5 shows two sets of simulations one for each of the two strains showing biomass, residual glucose and the different fermentation products. The model fits can be seen to be excellent for both species, as the observed metabolic shifts caused by knocking out the acetate pathway genes are faithfully represented by the model using a *single set of kinetic parameters*. For the YBS121 strain, the enzyme synthesis parameter $k_{E,9}$, connected with the single knockout of *adhE* gene, was set to zero.

Table 5.2 Stoichiometric reactions for anaerobic *E. coli* network (Young et al., 2008).

Reaction	Stoichiometry
1	$\text{GLC} + \text{PEP} \rightarrow \text{G6P} + \text{PYR}$
2	$\text{G6P} + \text{ATP} \rightarrow 2 \text{T3P} + \text{ADP}$
3	$\text{G6P} + 6 \text{NAD}^+ \rightarrow \text{T3P} + 3 \text{CO}_2 + 6 \text{NADH}$
4	$\text{T3P} + \text{NAD}^+ + \text{ADP} \rightarrow \text{PEP} + \text{NADH} + \text{ATP}$
5	$\text{PEP} + \text{ADP} \rightarrow \text{PYR} + \text{ATP}$
6	$\text{PEP} + \text{CO}_2 + 2 \text{NADH} \rightarrow \text{SUC} + 2 \text{NAD}^+$
7	$\text{PYR} + \text{CoA} \rightarrow \text{AcCoA} + \text{FOR}$
8	$\text{PYR} + \text{NADH} \rightarrow \text{LAC} + \text{NAD}^+$
9	$\text{AcCoA} + \text{ADP} \rightarrow \text{ACT} + \text{CoA} + 2 \text{NAD}^+$
10	$\text{AcCoA} + 2 \text{NADH} \rightarrow \text{ETH} + \text{CoA} + \text{ATP}$
11	$\text{FOR} \rightarrow \text{CO}_2 + \text{H}_2$
12	$6.775 \text{G6P} + 82.217 \text{ATP} + 4.065 \text{NADH} \rightarrow \text{B} + 82.225 \text{ADP} + 4.065 \text{NAD}^+ + 0.001 \text{CoA}$

Growth requirements in reaction 12 are in mmoles per gram of biomass

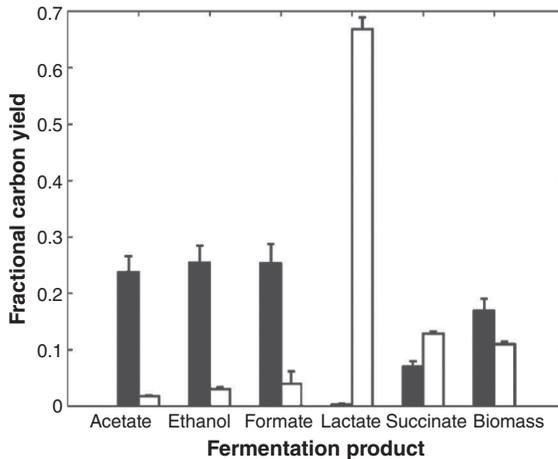


Figure 5.4 Distribution of glucose carbon in fermentation products of GJT001 (black) and YBS121 (white) obtained by Young et al. (2008) in shake flask experiments. Error bars indicate standard errors. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

Figure 5.6 presents the model predictions of intracellular metabolites in GJT001 and YBS121 strains. Literature sources (Chassagnole et al., 2002; Kümmel et al., 2006; Yang et al., 2001) show that the estimated metabolite levels are within the expected range; the results suggest that pyruvate accumulation in the YBS121 strain is responsible for inducing the succinate and lactate production pathways while inhibiting ethanol production.

Table 5.3 Relative flux patterns in EMs of *E. coli* network in Figure 5.3. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

EM	Reaction											
	1	2	3	4	5	6	7	8	9	10	11	12
1	51.6	44.2	0.7	89.0	37.4	0	0	89.0	0	0	0	1
2	89.0	70.9	11.4	153.1	0	64.1	0	89.0	0	0	0	1
3	49.2	42.5	0	84.9	35.7	0	4.1	80.9	4.1	0	0	1
4	49.2	42.5	0	84.9	0	35.7	39.8	9.5	39.8	0	0	1
5	35.8	29.0	0	58.0	22.2	0	58.0	0	31.0	27.0	0	1
6	46.9	40.1	0	80.2	0	33.3	46.9	0	42.1	4.7	0	1
7	59.0	36.7	15.5	89.0	30.0	0	89.0	0	0	89.0	0	1
8	89.0	58.1	24.1	140.4	0	51.4	89.0	0	0	89.0	0	1

EM fluxes are scaled relative to the rate of biomass synthesis.

Table 5.4 Elementary modes of *E. coli* network in Figure 5.3 represented by net conversions of substrates to products. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

EM	Net Conversion of Substrate to Fermentation Products and Biomass
1	51.6 GLC → 89.0 LAC + 2.0 CO ₂ + B
2	89.0 GLC + 30.0 CO ₂ → 89.0 LAC + 64.1 SUC + B
3	49.2 GLC → 4.1 ACT + 4.1 FOR + 80.9 LAC + B
4	49.2 GLC + 35.7 CO ₂ → 39.8 ACT + 39.8 FOR + 9.5 LAC + 35.7 SUC + B
5	35.8 GLC → 31.0 ACT + 27.0 ETH + 58.0 FOR + B
6	46.9 GLC + 33.3 CO ₂ → 42.1 ACT + 4.7 ETH + 46.9 FOR + 33.3 SUC + B
7	59.0 GLC → 89.0 ETH + 89.0 FOR + 46.5 CO ₂ + B
8	89.0 GLC → 89.0 ETH + 89.0 FOR + 51.4 SUC + 20.8 CO ₂ + B

A unique attribute of the cybernetic model, not shared by constraint based models, is its ability to predict both *yields* and *productivities* of the recombinant strains. The simulations show that the deletion of both *ldha* and *pta-ackA* improve ethanol production over and above that of the wild-type strain, a result that has corroboration in the literature (Yang et al., 1999b).

Figure 5.7 displays the global *v*-variables during the exponential phase which indicates the relative preferences of the different EMs. Specifically, EMs 5–7 are shown to be down-regulated in YBS121 in response to the acetate pathway bottleneck. Further, the dominance of EM 5 in the parent strain shifts away to EM1 in the knockout strain resulting in the observed shift to lactate production. Experimental findings by Zhu and Shimizu (2005) showed a small growth rate of *adhE* knockout strain, considerably smaller than depicted in Figure 5.5. Gupta and Clark (1989) found that *E. coli* strains lacking *adhE* cannot grow anaerobically on glucose although significant amounts of acetate and lactate were observed. However, spontaneous *pta* mutations in the organism restored growth to this strain with small amounts of succinate and acetate and large amounts of lactate. Thus adaptive evolution is found to vindicate model predicted

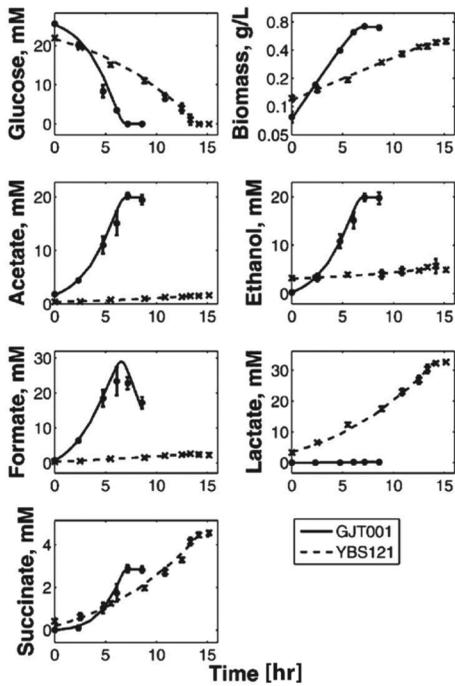


Figure 5.5 Simulated behavior of strains GJT001 (—) and YBS121 (---) fitted to experimental measurements. Error bars indicate standard errors of the measurements. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

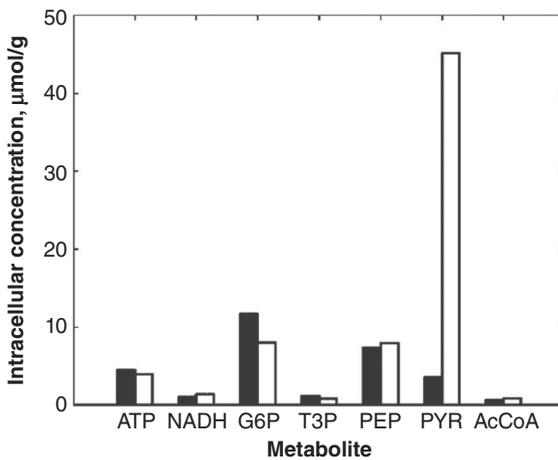


Figure 5.6 Simulated exponential phase intracellular concentrations of various metabolites in GJT001 (black) and YBS121 (white). Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

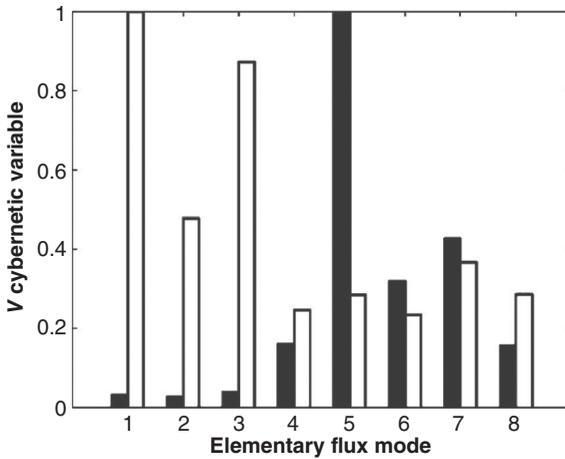


Figure 5.7 The global v cybernetic variables for GJT001 (black) and YBS12 strains (white) during exponential growth. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

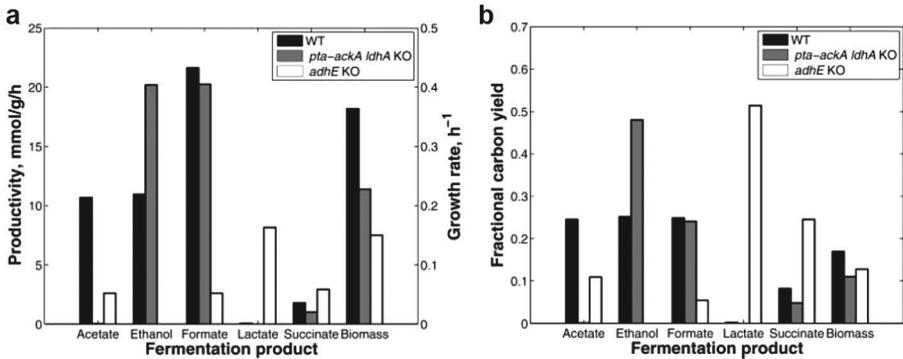


Figure 5.8 Simulated exponential phase profiles of productivity (a) and fractional carbon yield (b) of fermentation products for wild-type, *pta-ackA ldhA* double knockout, *adhE* knockout strains of *E. coli*; (a): Rates of product formation and growth. (b): Distribution of glucose carbon in the fermentation products. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

behavior of recombinant strains leading Young et al. (2008) to conclude that “it is often better to have a model that overestimates the robustness of the organism and thus provides some estimate of its ultimate metabolic capabilities, rather than a model that cannot be extrapolated effectively outside of a limited phenotypic range.”

Young et al. (2008) also present the results of modeling the genetic manipulation with introduction of heterologous genes. Such gene additions lead to additional metabolic routes thus creating additional EMs. The predicted results of three different *E. coli* strains, for which experimental results are available for comparison, are presented in Figure 5.9. The three strains are: S1) the wild type (WT) with an NAD^+ dependent formate dehydrogenase (*fdhI*) gene from the yeast *Candida boidinii*, S2) WT with a

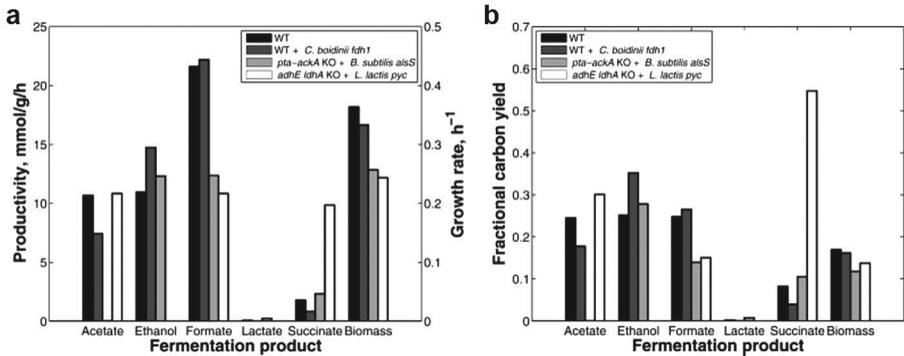


Figure 5.9 Simulated exponential phase fermentation profiles of 4 different strains of *E. coli*: wild-type, wild-type with insertion of *C. boidinii fdh1* gene, *pta-ackA* knockout with insertion of *B. subtilis alsS* gene. Reproduced from Young et al. (2008), with permission, Copyright © 2008 Wiley Periodicals, Inc.

pta-ackA knockout and an acetolactase synthase (*alsS*) gene insertion from *Bacillus subtilis*, S3) WT with an *adhE ldhA* double knockout and pyruvate carboxylase gene (*pyc*) insertion from *Lactobacillus lactis*.

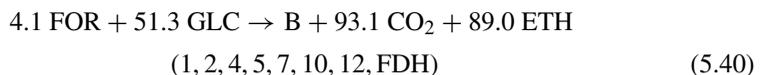
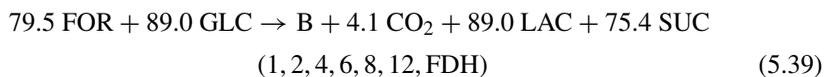
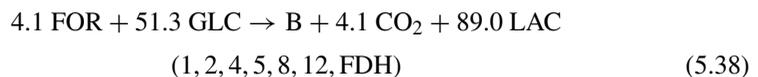
In formulating the model for S1, Young et al. (2008) added the NADH-producing reaction below as a consequence of the addition of *fdh1* gene.

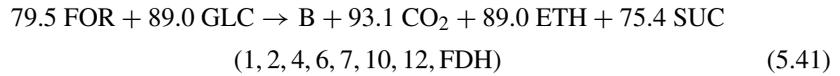


with a fully identified (specific) kinetic rate given by

$$r_{\text{FDH}}(\text{mmol/g/h}) = 100b \times \left(\frac{x_{\text{FOR}}}{x_{\text{FOR}} + 13\text{mM}} \right) \times \left(\frac{x_{\text{NAD}^+}}{x_{\text{NAD}^+} + 0.1\mu\text{mol/g}} \right) \quad (5.37)$$

The saturation constants above were obtained from Schütte et al. (1976). Young et al. (2008) assumed that transcription of heterologous genes was controlled by a non-native promoter and adopted an arbitrary rate constant without involving an enzyme balance thus bypassing the associated *u*-variable. However, they employed the *v*-variable as determined by the level of participation of the FDH reaction in the four additional EMs acquired by the expanded network, reflecting the assumption that the enzyme was subject to native activity control. The net stoichiometry of each FDH-containing reaction EM is reproduced below with the full list of participating reactions in parentheses.





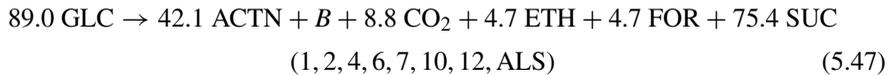
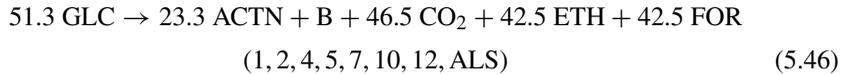
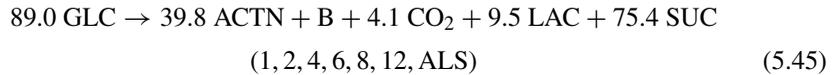
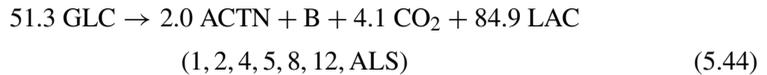
Similarly, the *alsS* gene insertion from *B. subtilis* confers the pyruvate draining reaction



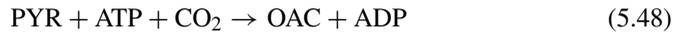
where ACTN refers to actoin, the end product of the ALS reaction. The kinetics of this reaction was assigned the rate expression

$$r_{\text{ALS}} = 100b \times \left(\frac{x_{\text{PYR}}}{x_{\text{PYR}} + 21 \mu\text{mol/g}} \right) \quad (5.43)$$

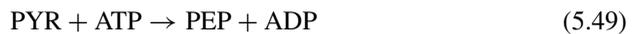
The saturation constant for pyruvate was obtained from Holtzclaw and Chapman (1975) and the rate constant was arbitrarily set to a value of 100mmol/g/h as before. The introduction of the ALS reaction adds four more EMs as given below



Lastly, the addition of *pyc* gene from *L. lactis* confers the pyruvate carboxylase reaction



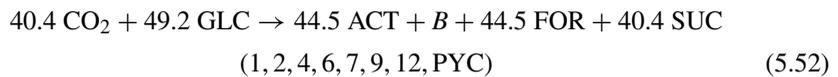
where OAC represents oxaloacetate. Young et al. (2008) observe that since the base network of *E. coli* does not feature oxaloacetate as it is an intermediate in the lumped pathway from PEP to succinate. Toward avoiding a change of the base network of *E. coli*, Young et al. (2008) replace the reaction Eq. (5.48) by



which will account for the recycling of pyruvate in the presence of *pyc*. The rate of reaction Eq. (5.49) was assumed to follow the kinetics

$$r_{\text{PYR}} = 50b \times \left(\frac{x_{\text{PYR}}}{x_{\text{PYR}} + 1 \mu\text{mol/g}} \right) \left(\frac{x_{\text{ATP}}}{x_{\text{ATP}} + 0.1 \mu\text{mol/g}} \right) \quad (5.50)$$

The lack of data on saturation constants led Young et al. (2008) to adopt constants for other microbes on the BRENDA website, www.brenda-enzymes.info/. The value of 50 mmol/g/h. was arbitrarily chosen for the rate constant. Three more EMs are acquired by the introduction of *pyc* to the base *E. coli* network as given by internal cycle of two reactions (5, PYC), and



Only one of the above three EMs is associated with the synthesis of biomass. Retaining the same objective function, and employing the cybernetic variables model predictions of the wild type and the three different strains are shown in Figure 5.9. The results shown are in striking agreement with observations in the published literature. Thus the increased ethanol production due to the insertion of *C. boidinii fdh1* in the wild-type strain is corroborated by the observation of Berrios-Rivera et al. (2002) who showed that introduction of the NAD⁺ dependent *fdh1* leads to more ethanol and less acetate as appearing in Figure 5.9. Next, Yang et al. (1999a) were able to revert a *pta-ackA* knockout strain to make ethanol instead of lactate by introducing the *alsS* gene from *Bacillus subtilis*. The model is able to reproduce this result by restoring ethanol production in this strain to nearly that of the wild-type. Since the underlying mechanism for this effect is the dynamics of intracellular pyruvate levels, this prediction of the cybernetic model is outside the purview of stoichiometric modeling approaches. Finally, the improved production of succinate reported by Sánchez et al. (2005) through the insertion of *pyc* gene from *Lactobacillus lactis* in an *adhE ldhA* double knockout background is successfully reproduced by the model by showing that half of the available carbon is diverted to succinate production.

5.2 Concluding Remarks

It should be evident to the reader that the work of Young (2005) has successfully accomplished the conceptual extension of the cybernetic framework to a metabolic network. The introduction of EM decomposition is a significant aspect in this regard because the underlying structure minimizes the need for introduction of *ad hoc* features often found in the early cybernetic models. The different control policies that appeared in Chapter 4 notably add to the armory of cybernetic models in understanding the regulatory underpinnings of bioinformatic data. In this regard, emergence of the concept of a temperate policy deserves a special note of approbation because of its latent capacity to overcome the potential shortfalls of the “greedy” version. A case in point, as pointed out by Young et al. (2008), is the circumstance of a reaction capable of contributing to

the organism's goal lagging behind other reactions because of substrate depletion that might spur the greedy policy to promote the level and activity of the enzyme for the slow reaction which could be a potentially hasty response.

Multiple ways of attributing costs to regulatory action can also produce a variety of formulational alternatives for the cybernetic model. Recognition of these and other aforementioned attributes considerably expands the scope of the cybernetic approach notwithstanding viewpoints in the literature expressed to the contrary.