A cybernetic modeling framework for analysis of metabolic systems

Abhijit Anand Namjoshi, Doraiswami Ramkrishna

Department of Chemical Engineering, Purdue University, 480 Stadium Mall Drive, West Lafayette, IN 47907, USA

Received 15 January 2004
Available online 14 November 2004

Abstract

We present a cybernetic perspective of metabolic systems, designed to take full account of regulatory processes, which profoundly influence the course of metabolism through dependence not only on the organism’s genomic constitution, but also on abiotic environmental influences. The cybernetic perspective relies on proposing “mechanisms” of choice behavior by which biological systems drive enzyme syntheses and allosteric processes for chemical reactions that promote perceived local or global cellular objectives. This article is an exposition of the framework, attempting to show how experimental data can be used to facilitate model identification. Such an approach is proposed with the view to exploit detailed cellular measurements such as metabolite and mRNA or enzyme levels. We present a modular approach that originates with the metabolic network to be modeled and systematically breaks it down into smaller regulatory units which can be modeled using previously established cybernetic competitions [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; Straight, J. V., & Ramkrishna, D. (1990). Regulation of complex growth dynamics—substitutable and complementary processes. Abstracts of the papers of The American Chemical Society, 200, 4-BIOT]. Experimental data can assist in rapid formulation of numerous specific combinatorial alternatives suggested by the metabolic network. Methods are discussed within this continuously evolving framework for the identification of cybernetic models from data with illustrative examples.

© 2004 Published by Elsevier Ltd.

Keywords: Metabolic systems; Modeling framework; Cybernetic modeling

1. Introduction

Biological systems are extremely complex not only due to the large number of reactions that take place in a single cell, but also due to the extremely non-linear interactions between different metabolites and enzymes. Such interactions stem from the genetic make-up of the cells wherein the response to external stimulus is encoded in the genome of the cells. Thus, perturbations in the extracellular environment can lead to variations in the levels of the mRNA of the cells, and correspondingly in the enzymes. The enzymes themselves act on different intracellular metabolites, thereby changing their levels. The fluxes in different pathways therefore change and such changes can lead to different substrate uptake patterns as well as differences in the growth of cells as well as the syntheses of secondary products. The metabolic engineer must therefore deal with a large number of reactions in order to predict cellular behavior in order to optimize cells as well as cell-based processes. The cybernetic approach has been extremely successful in describing bioreactor behavior [Kompala, Ramkrishna, Jansen, & Tsao, 1986; Namjoshi, Kienle, & Ramkrishna, 2003; Ramkrishna, 1982; Ramakrishna, Ramkrishna, & Konopka, 1996; Straight & Ramkrishna, 1990; Varrer & Ramkrishna, 1999a,1999b] based on the postulate that cells are objective-oriented organisms which change the internal enzyme levels in response to environmental changes.

The cybernetic modeling framework relies on proposing objective functions for intracellular reactions.1

1 This aspect of the formulation of cybernetic models is far from trivial. The cybernetic structure of large metabolic systems is an evolving issue and what is contained in this paper may be considered as relatively primitive.
Once an objective function is proposed, the matching law and proportional law can be employed to derive cybernetic "u" and "v" control variables which modify enzyme synthesis as well as enzyme activity, respectively. As the number of reactions that need to be modeled expand, the given network can be broken down into elementary reaction units in non-unique ways. The existing strategies to guide this include the use of network fluxes calculated from metabolic flux analysis (MFA) (Namjoshi, Hu, & Ramkrishna, 2003) which is the use of network fluxes calculated from metabolic flux analysis as well as enzyme activity, respectively. The enzyme balances are reactions involved in transport. The corresponding "u" variables of the reactions involved in transport. The corresponding "v" variables may simply be set to unity if the reaction is not to be modeled as a part of a cybernetic regulatory structure.

The biomass components are diluted only by growth

\[
\frac{dm_j}{dt} = g_j^m \text{diag}(\nu^t) + g_j^t \text{diag}(\nu^r - r_{\text{out}} m)
\]

where \( r \) is the \( R \)-dimensional-specific (per unit cell mass) reaction rate vector with diag(\( \nu^r \)) as the corresponding cybernetic "u" variables and \( r_{\text{out}} \) the specific growth rate. Also, \( g_j^m \) and \( g_j^t \) are the \( j \)-th columns of \( G^m \) and \( G^t \), respectively, and represent the metabolic profile of the \( j \)-th metabolite. \( G \) is a matrix corresponding to the participation of intracellular metabolites in reactions restricted to the biotic phase. In vector notation,

\[
\frac{dm}{dt} = G^m \text{diag}(\nu^t) + G^t \text{diag}(\nu^r - r_{\text{out}} m)
\]

Regulation is incorporated into the cybernetic framework by the use of cybernetic "u" and "v" variables (Kompala et al., 1986; Ramkrishna, 1982) which modify enzyme synthesis and enzyme activity, respectively. The enzyme balances are thus written as

\[
\frac{dv_i}{dt} = r_{E_i}^u + r_{E_i} - (r_{E_i} + \beta_{E_i}) v_i
\]

Here, \( r_{E_i}^u \) is a constitutive level of enzyme synthesis, with \( r_{E_i} \) being the inductive level and \( \beta_{E_i} \) a first-order enzyme

2. A general cybernetic modeling framework for biosystems

The modeling framework consists of several state variables:

(1) Biomass component vector: We denote by \( X \) the mass concentration (mass per unit volume of culture) of biomass which consists of \( n \) master components \( M_i \) with concentration \( c_i \) and intracellular level \( m_i \). In vector form, the biomass components \( m \) can be related to the concentrations \( c \) as \( \text{vec}(m) = A \text{vec}(c) \).

(2) Enzyme vector: Enzymes are treated separately from the biomass vector, since they serve as catalysts for reactions between species in the biomass vector. We denote by \( U \) the enzyme component vector, which includes all extracellular components \( S_j \) which include substrates as well as excreted products.

### Abiotic vector

(3) Abiotic vector: The abiotic vector \( s \) is composed of the mass concentrations of all extracellular components \( S_j \) which are components of the extracellular metabolite vector \( S \) and its concentration \( s \) is expressed as a mass concentration (mass per unit volume of culture). The corresponding stoichiometric matrix \( G_m^s \) of the extracellular metabolites and \( G_t^s \) of the abiotic phase variables are involved in transport processes between the biotic and the abiotic phases, \( s_i \) is the vector of concentrations of extracellular species at the inlet and \( G_s^t \) is the matrix with diagonal elements as the cybernetic "v" variables of the reactions involved in transport. The corresponding "v" variables may simply be set to unity if the reaction is not to be modeled as a part of a cybernetic regulatory structure.

2.1. Species balances

The abiotic phase variables are involved in transport and reaction processes that may be represented as

\[
\sum_{i=1}^{R_s} G_{i,k}^s s_i + \sum_{j=1}^{R_t} G_{j,i}^t m_j, \quad i = 1, 2, \ldots, R_s
\]

where

- \( G_s^t \) and \( G_t^s \) are the corresponding stoichiometric matrices for abiotic and biotic components, respectively.
- The mass balance on \( s \) for a chemostat can then be written as

\[
\frac{ds}{dt} = G_s^t \text{diag}(\nu^* s) X + D(s_f - s)
\]

where \( D \) is the dilution rate, \( s^* \) is the \( R_s \)-dimensional reaction rate vector that governs the transport processes between the biotic and the abiotic phases. \( s_f \) is the vector of concentrations of extracellular species at the inlet and \( G_s^t \) is the matrix with diagonal elements as the cybernetic "v" variables of the reactions involved in transport. The corresponding "v" variables may simply be set to unity if the reaction is not to be modeled as a part of a cybernetic regulatory structure.

The biomass components are diluted only by growth

\[
\frac{dm_j}{dt} = g_j^m \text{diag}(\nu^t) + g_j^t \text{diag}(\nu^r - r_{\text{out}} m_j)
\]

where \( r \) is the \( R \)-dimensional-specific (per unit cell mass) reaction rate vector with diag(\( \nu^r \)) as the corresponding cybernetic "u" variables and \( r_{\text{out}} \) the specific growth rate. Also, \( g_j^m \) and \( g_j^t \) are the \( j \)-th columns of \( G^m \) and \( G^t \), respectively, and represent the metabolic profile of the \( j \)-th metabolite. \( G \) is a matrix corresponding to the participation of intracellular metabolites in reactions restricted to the biotic phase. In vector notation,

\[
\frac{dm}{dt} = G^m \text{diag}(\nu^t) + G^t \text{diag}(\nu^r - r_{\text{out}} m)
\]

Regulation is incorporated into the cybernetic framework by the use of cybernetic "u" and "v" variables (Kompala et al., 1986; Ramkrishna, 1982) which modify enzyme synthesis and enzyme activity, respectively. The enzyme balances are thus written as

\[
\frac{dv_i}{dt} = r_{E_i}^u + r_{E_i} - (r_{E_i} + \beta_{E_i}) v_i
\]

Here, \( r_{E_i}^u \) is a constitutive level of enzyme synthesis, with \( r_{E_i} \) being the inductive level and \( \beta_{E_i} \) a first-order enzyme...
de/dt = r^∗E + diag(u)E - (diag(r_D) + diag(BE))E (7)

3. Strategy for model formulation using the general modeling framework

This paper presents a cybernetic modeling strategy which modifies considerably one proposed earlier (Namjoshi, Hu, et al., 2003) to include measurements of metabolites as well as enzymes. We present this strategy in this section. Metabolic pathways may comprise numerous reactions, and in many situations, it is possible and sufficient to model a few key reaction steps that are relevant to the phenomena being explained. The first step in the cybernetic modeling process is therefore to identify the pertinent pathway of interest and decide on a level of abstraction to facilitate the identification of cybernetic competitions in the (abstracted) pathway. This is done by considering elementary pathways such as convergent and divergent branch points as well as linear and cyclic structures (Straight & Ramkrishna, 1990) that are modeled from optimality criteria. Subsequently, the competitions need to be identified in these pathways, with due regard to enzymes that are common to multiple pathways. Entire sets of pathways might be turned on or off using global cybernetic variables. Model identification can thus be broken down into the following steps:

1. Pathway abstraction;
2. Identification of elementary pathway structures;
3. Defining cybernetic variables for enzymes in overlapping pathways;
4. Defining global cybernetic variables;
5. Writing species balances for the pertinent reactor configuration;
6. Parameter identification and optimization.

This model identification strategy is depicted in a flowchart shown in Fig. 1. Thus, we begin by abstracting the model suitably, with due regard to the objective of the modeling effort. Metabolic flux measurements can aid considerably in this process. The strategy of using metabolic flux analysis to abstract the framework has been discussed elsewhere (Namjoshi, Hu, et al., 2003) in detail. Continuing deliberations on the strategy at hand, we present in this paper an algorithm-based approach to examine the stoichiometric matrix of the network to identify convergent, divergent, cyclic and linear pathways in which cybernetic competitions can be identified. This process can be aided considerably by MFA. Additionally, a power law-based technique, which we present in this paper, can also be used, if detailed enzyme and metabolite measurements are available. Since the number of elementary pathways identified can be large in general, one can use a few non-discriminatory parameter values to aid the search of a model using statistical search techniques such as genetic algorithms. Once a few candidate models are identified, a full-blown parameter search can be carried out to identify the model parameters.

3.1. Pathway abstraction

Mathematical models that capture a small number of reactions are more tractable compared to ones that capture detailed pathways. Once a simple model is found to explain some of the phenomena being investigated, it can be expanded to capture detailed measurements. This is a “top-down” approach, in contrast with a “bottom-up” approach that would try and model every observable reaction in the cell to explain experimental results, and then reduce it using model reduction techniques to make it computationally attractive. A top-down approach is computationally more attractive as cellular reactions cannot always be easily quantified, although that is changing at a rapid pace. At this initial stage of model development, the metabolic pathway may be suitably abstracted to capture the phenomena to be addressed by the model. The model needs to explore the metabolic network at a level of complexity higher than the experimental data it seeks to explain or predict. For instance, highly abstracted pathways might be enough to explain reactor behavior such as the formation of cell mass and secondary metabolic products, but a more detailed treatment is necessary for developing a genome scale model of primary metabolism.
This approach is discussed in detail elsewhere (Namjoshi, Hu, et al., 2003), where a cybernetic model was constructed to explain the phenomenon of multiple steady states in hybridoma cell continuous cultures. These multiple steady states resulted from culturing the mammalian cells in batch or fed-batch cultures in which the cells were subject to discriminating concentrations of their key substrates glucose and glutamine. The cells cultured as a batch saw higher concentrations of glucose and converted it into lactate, whereas in a fed-batch, the formation of lactate and other waste metabolites such as alanine and ammonia was considerably reduced (Hu, et al., 2003), where a cybernetic model was constructed considering the network reactions in matrix form, i.e.

$$\sum_{j=1}^{n} G_{ij} M_j = 0, \quad i = 1, 2, \ldots, R$$

(8)

where \( n \) is the number of species and \( R \) is the number of reactions in the network. The \( j \)th column \( g \) in the matrix \( G \) gives the “stoichiometric profile” of the metabolite \( M_j \) and this can be used as a representation of the metabolite’s role in the reaction network. \( G \) includes transport as well as intracellular reactions. The metabolic matrix can thus be represented as \( G \equiv [g_1, g_2, \ldots, g_n] \). We can now identify the sets \( I^+ \) and \( I^- \) which represent the reactions that produce and consume, respectively, the metabolite \( M_j \):

$$I^+(j) \equiv \{ i \in (1, 2, \ldots, R) : G_{ij} > 0 \},$$

$$I^-(j) \equiv \{ i \in (1, 2, \ldots, R) : G_{ij} < 0 \}$$

(9)

The cardinalities of both these sets (the number of element in each set) are represented as \( C(I^+(j)), C(I^-)(j) \). Similarly, one can keep track of the metabolites that are produced \( P(i) \) and consumed \( F(i) \) in the \( i \)th reaction, given by

$$P(i) \equiv \{ j \in (1, 2, \ldots, n) : G_{ij} > 0 \},$$

$$F(i) \equiv \{ j \in (1, 2, \ldots, n) : G_{ij} < 0 \}$$

(10)

with cardinalities \( C(P(i)), C(F(i)) \), respectively.

We now consider the problem of identifying the elementary structures identified in the work of Straight and Ramkrishna (1990), i.e. convergent and divergent branch points, as well as cyclic and linear pathways in a reaction network of arbitrary complexity.

1. **Convergent branch points**: Convergent branch points arise where more than one reaction can form the same product. Thus, metabolite \( j \) is said to be at a convergent branch point if \( C(P(j)) \geq 2 \). Thus, the total number of converging units is given by \( C[I \cap C(P(j)) \geq 2] \). The set of all metabolites that are convergent branch points is then given by

$$J^{con} = \{ j : C(I^+(j)) \geq 2 \}$$

(11)

Correspondingly, one can identify the set of convergent pathways given by

$$\Phi = \bigcup_{j \in J^{con}} I^+(j)$$

(12)

2. **Divergent branch points**: Divergent branch points arise where more than one reaction can consume the same reactant. Thus, metabolite \( j \) is said to be at a divergent branch point if \( C(I^-(j)) \geq 2 \). Thus, the total number of diverging units is given by \( C[I \cap C(I^-)(j) \geq 2] \). The divergent branch points can be identified as

$$J^{div} = \{ j : C(I^-)(j) \geq 2 \}$$

(13)

leading to the set of divergent pathways given by

$$\Delta = \bigcup_{j \in J^{div}} I^-$$

(14)

3. **Cyclic pathways**: A flow chart for identification of cycles is presented in Fig. 2. Begin by isolating the metabolites that are both formed as well as consumed by at least one reaction and store them in a master set. Then, start with an arbitrary metabolite from the master set and store it into an active set. Find all the reactions that consume this metabolite and for each of these reactions, track the product of each reaction. For each product, form a path with the first element as the original starting metabolite, and the second element as the newly found product. Similarly, one can form paths in which the reactions that connect the metabolites are tracked. Additionally, each metabolite encountered must be added to the active set. For each of the metabolite paths thus formed, check if any repeated metabolite has been encountered. If not, continue the search starting with each of the second-generation metabolites and appending the original paths with the third-generation products. However, if a metabolite is repeated at any generation, a cycle has been found, since it represents a situation of starting from a metabolite as a reactant and arriving at the same metabolite. If, on the other hand, a metabolite has been reached that is not consumed further, it signifies a dead-end eliminating the need for further pursuit of the path for the detection of a cycle. Alternatively, it may be stored as a likely candidate for linear processes. One can thus potentially find many cycles starting from the same initial metabolite. After all cycles have been found from starting
from one metabolite, or all metabolite paths eliminated as they have reached dead ends, one should start with any left-over metabolite that may not have featured in any path starting from the first metabolite, and continue the search. This set can be easily identified by considering any metabolite that is present in the master set, but absent in the active set. After all the metabolites that are both formed and consumed have featured in at least one metabolite path, the search may be terminated.

Following the set notation established earlier, one can think of this problem as follows. Form a path of metabolites and reactions as

$$j_1 \in E, i_1 \in I^+(j_1), j_2 \in J^+(i_1), \ldots, i_k \in I^-(j_k), j_{k+1} \in J^+(i_k)$$

where

$$E \equiv \{j \in \{1, 2, \ldots, n\} : C[I^+(j)] \geq 1, C[J^-(j)] \geq 1\}$$

The set of cyclic pathways can thus be defined as

$$\Psi \equiv \{\psi_1, \psi_2, \ldots, \psi_M\}$$

where $\psi_k$ are cyclic pathways comprising reactions in each pathway identified by the model identification strategy.

(4) Linear pathways: Linear pathways can be found just like cycles; however, since linear pathways can be subsets of cycles and otherwise intersect with convergent as well as divergent pathways, we choose to model linear pathways among reactions that are left over after the cyclic, convergent and divergent pathways have been identified. This process is carried out similar to the algorithm for the cycles as discussed above, but by restricting the stoichiometric matrix to subsets of reactions that remain after all the reactions participating in cyclic, convergent and divergent pathways are eliminated. We have the linear pathways identified earlier given by

$$\Lambda \equiv \{\lambda_1, \lambda_2, \ldots, \lambda_{\Lambda}\}$$

Example: The same strategy is applied to the more detailed pathway in Fig. 3 which represents the central carbon metabolism found in Escherichia coli and other microorganisms. Nineteen convergent and 18 divergent pathways were identified. Many cyclic pathways were found, which were consolidated into larger cycles. The basis for such consolidation is inclusion of cycles with common reactions as one big cycle. Such consolidated cycles are shown in Fig. 4. The figure shows five reversible reactions which do not have any enzymes common with other pathways, whereas the two consolidated cycles identified represent the pentose phosphate cycle as well as the TCA cycle. These cycles are well doc-
3.3. Identification of competitions in the pathways

The elementary pathways have now been identified. The next step is to associate cybernetic competitions with these elementary pathways. This cannot be always done uniquely and so the same metabolic network can lead to multiple ways in which the competitions can be identified in the pathway. This should not be seen as a drawback of the theory but is, in fact, an advantage as the modeler can test different hypotheses about the regulatory features of the network, using the different ways in which the competitions can be modeled.

Once the elementary pathways are identified, it remains to be decided how these pathways should be modeled. In particular, it is not clear whether each elementary pathway identified should be modeled as a cybernetic competition, or whether simple Michaelis–Menten kinetics is enough to describe the underlying regulation. Moreover, there might be a choice of competitions to be applied. These decisions can be aided by a power law formulation in the case where detailed measurements on enzymes and metabolites are available. The power law formulation can narrow down the choice of competitions, but the combinatorial nature of the problem may not vanish completely. This can then be handled by the use of optimization tools to span the modeling space, such as genetic algorithms. This is discussed further in what follows.

3.4. Defining cybernetic variables for enzymes in overlapping pathways

The modular approach proposed is convenient to use because cybernetic competitions have to be identified only for small segments of the network, and pathways common to these segments derive their cybernetic variables as a combination of their cybernetic variables from different pathways. Varner (2000) proposed the combination to be a product of the cybernetic variables from all the pathways. This product rule is just one of many ways in which common elements of pathways can be handled. Another rule where the cybernetic variables for the common pathway may be the sum or average of the values obtained due to participation in individual pathways may be employed (Namjoshi, 2003). Needless to say, the dynamic as well as steady-state fluxes are a strong function of the combination rule chosen for a given set of model parameters, as demonstrated next.

3.5. Defining global cybernetic variables

Global cybernetic variables have been used (Varner & Ramkrishna, 1998, 1999a) as tools to turn on or switch off entire pathways in response to environmental changes. Traditionally, therefore, the product rule has been used to multiply the global cybernetic variable with the local cybernetic variable for the entry point reaction in a pathway. Once the entry point enzyme is controlled, it leads to increase or decrease in the product it makes, thereby fueling or stopping, respectively, the downstream reactions. Using another rule, such as the sum or the average rule, in this instance would not facilitate the turning off of the pathway, as it would curtail only a part of the enzyme synthesis/activity. The issue of global cybernetic variables is a deep one and our discussion here must be regarded as somewhat rudimentary.

3.6. Coordinated control using cybernetic variables

Genes that are expressed together can be thought to be organized in a cluster of genes. Such gene clusters may be quite conveniently modeled using cybernetic variables. Suppose that pathways 1 and 2 are said to be clustered together and that pathway 1 derives its cybernetic variables from competition with some other pathways. In that case, one can set $u_2 = u_1$ and $v_2 = v_1$, thereby ensuring that enzyme $E_2$ is always made synchronously with $E_1$. This is further illustrated in the case of the cybernetic hybridoma model example (Namjoshi, Hu, et al., 2003), where a pathway for producing lactate from pyruvate is coordinately controlled with one glycolytic enzyme. The rationale behind this is the recycling of the cofactors NAD⁺ and NADH that is known to occur between these two pathways.

4. Power law formulation for determining cybernetic competitions

In the preceding sections, we have discussed how cybernetic models can be formulated, once the elementary units are identified. However, since this can be done in a multiple number of ways, we propose a diagnostic tool for identifying the most likely underlying competition. Such an analysis assumes that measurements of enzymes or mRNAs along with metabolite concentrations are available and is based on past work (Gupta, Varner, & Maranas, 2005; Yeung, Tegner, & Collins, 2002). Yeung et al. (2002) consider transcript array data for systems operating near steady state. They write the balance for an mRNA species...
(36) as

\[ x(t) = -\lambda x(t) + \sum_{j=1}^{N} W_{ij} x_j(t) + \beta_i(t) + \xi_i(t), \]

\[ \forall i = 1, 2, \ldots, N \] (19)

where \( \lambda \) denotes a first-order degradation and \( W_{ij} \) denotes the influence of the \( j \)th gene on the \( i \)th gene, \( \beta_i(t) \) denotes the perturbation induced and \( \xi_i(t) \) denotes measurement noise. For \( M \) different cellular measurements, in vector form, \( \mathbf{x}_{N \times M} \) is given by

\[
\begin{bmatrix}
  x_1^1 \\
  x_1^2 \\
  \vdots \\
  x_1^M \\
  x_2^1 \\
  x_2^2 \\
  \vdots \\
  x_2^M \\
  \vdots \\
  x_N^1 \\
  x_N^2 \\
  \vdots \\
  x_N^M
\end{bmatrix}
\] (20)

leads to

\[
\mathbf{x}_{N \times M} = \mathbf{A}_{N \times N} \mathbf{x}_{N \times M} + \mathbf{B}_{N \times M}
\] (21)

where \( \mathbf{A} = \mathbf{W} - \text{diag}(\lambda) \) is the unknown matrix, \( \mathbf{B} \) captures the perturbations and noise is neglected. In general, \( M < N \) and hence,

\[
(\mathbf{A}^T \mathbf{M}_{N \times N} \mathbf{A})_{N \times N} = (\mathbf{A}^T \mathbf{M}_{N \times N} + (\mathbf{B}^T \mathbf{M}_{N \times M})
\] (22)

will have multiple solutions. Yeung et al. (2002) use singular value decomposition

\[
(\mathbf{A}^T \mathbf{M}_{N \times N} \mathbf{A})_{N \times N} = \mathbf{U}_{M \times N} \mathbf{Z} \mathbf{V}^T \]

where \( \mathbf{U} \) and \( \mathbf{V} \) are the matrices of orthogonal vectors and \( \mathbf{Z} \) is the diagonal matrix of the singular values. A particular solution to \( A \) can then be obtained as

\[
A = \mathbf{A}_0 \equiv (\mathbf{X} - \mathbf{B}) / \text{diag} \left( \frac{1}{|c_j|} \right) \mathbf{V}^T
\] (24)

wherein the \( c_j = 0 \) terms are ignored. A more general solution is given by \( A = \mathbf{A}_0 + \mathbf{C} \mathbf{V}^T \) where \( \mathbf{C} \) is a matrix of coefficients with element \( c_{ij} = 0 \), \( \forall j > L \), \( L = \text{dim} \text{ker} \mathbf{X} \)). The particular solution can be shown to be an optimal solution in the least squares sense. However, Yeung et al. (2002) postulate the biochemical systems are sparse in that a gene expresses control over few other genes in the network and hence, the sparsest solution (with as many \( W_{ij} = 0 \) as possible) is the most likely solution. An optimization is subsequently carried out to find \( \mathbf{C} \) that will lead to the sparsest \( \mathbf{W} \). This strategy has been expanded on by Gupta et al. (2005) to substitute the linear dependence with a log-linear model for \( \mathbf{W} \). In this work, we build further on this approach by including not just mRNA (or enzyme) measurements but also different metabolite measurements and demonstrate how that is key to the correct identification of cybernetic variables using some preliminary cybernetic models.

For the sake of simplicity, we will restrict ourselves to traditional cybernetic models where the mRNAs are not modeled but enzyme synthesis and activity are manipulated directly to achieve regulation. This is largely because the cybernetic models we will use to demonstrate these concepts do not model mRNA levels but carry this out with enzymes instead. The identification problem can change substantially with increasing availability of measurements of intracellular species. The availability of metabolic measurements, in addition to enzyme and mRNA levels, for instance, will considerably aid understanding but also add to the complexity of the identification problem. The model identified will thus depend on the type of measurement used, and the framework should be adapted to the measurement available.

If one has measurements of enzyme levels at different time points, then one can differentiate these to obtain the derivatives \( \text{d}x / \text{d}t \). Also, these help us back out the term \( \beta_i(t) + \xi_i(t) \), since we know the growth rate and can assume a first-order degradation rate for the enzyme. Thus, we can then calculate the levels \( x_{Ei} \). One can then propose a linear model for the various \( r_{Ei} \), \( \forall i = 1, \ldots, N \) as a function of the enzyme levels \( e_j \), \( \forall j = 1, \ldots, N \). If one has \( M \) different measurements of each of the \( N \) enzyme levels over period of time, then one can rescale the values \( r_{Ei} \), \( \forall i = 1, \ldots, M \) by dividing them by the highest \( r_{Ei} \) encountered for some \( i = p, j = q \). Gupta et al. evaluate a linear as well as power law formulation to back out the coefficients of the “model”. Thus, one can write

\[
\begin{bmatrix}
  r_{E1} u_1^1 & r_{E1} u_1^2 & \cdots & r_{E1} u_1^M \\
  r_{E2} u_2^1 & r_{E2} u_2^2 & \cdots & r_{E2} u_2^M \\
  \vdots & \vdots & \ddots & \vdots \\
  r_{EN} u_N^1 & r_{EN} u_N^2 & \cdots & r_{EN} u_N^M
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  1 & 1 & \cdots & 1 \\
  \cdots & \cdots & \cdots & \cdots \\
  \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N} \\
  \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N} \\
  \vdots & \vdots & \ddots & \vdots \\
  \gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{NN}
\end{bmatrix}
\begin{bmatrix}
  e_1 & e_2 & \cdots & e_M
\end{bmatrix}
\] (25)

The linear model mentioned above holds good only for small perturbations around steady state and hence, a better approximation is required.

Fig. 5. Schematic of the modifications to the model of Ramakrishna et al. (1996); dotted line denotes components included as part of cell mass.
imiation is a power law formulation:

\[ r_E u_i^j = \prod_{j=1}^{N} (e_i^j)^{\lambda_{i,j}} \]  

(26)

In vector form,

\[ \log \begin{bmatrix} r_E u_1^1 & r_E u_2^1 & \ldots & r_E u_M^1 \\ r_E u_1^2 & r_E u_2^2 & \ldots & r_E u_M^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_E u_1^N & r_E u_2^N & \ldots & r_E u_M^N \end{bmatrix} \]

\[ = \begin{bmatrix} \gamma_1,1 & \gamma_1,2 & \ldots & \gamma_1,N \\ \gamma_2,1 & \gamma_2,2 & \ldots & \gamma_2,N \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_N,1 & \gamma_N,2 & \ldots & \gamma_N,N \end{bmatrix} \log \begin{bmatrix} e_1^1 & e_1^2 & \ldots & e_1^N \\ e_2^1 & e_2^2 & \ldots & e_2^N \\ \vdots & \vdots & \ddots & \vdots \\ e_N^1 & e_N^2 & \ldots & e_N^N \end{bmatrix} \]

This formulation was extended to include dependencies on all substrates or a chosen subset of all the substrates. This vastly improves the power law fit as demonstrated in the example next. Although one is not keen on using the power law in a predictive manner, if the power law fit is accurate, the coefficients are good representations of the underlying regulation.

4.1. Example: Model of Ramakrishna et al.

The model of Ramakrishna et al. (1996) was proposed to explain the simultaneous utilization of sugars and organic acids by considering two intermediates \( M_1 \) and \( M_2 \) which are formed from the two substrates \( S_1 \) and \( S_2 \) respectively, in enzyme-catalyzed steps 1 and 4. Reactions 2 and 3 allow interconversion between these two intermediates. Simultaneous utilization is addressed by considering competitions between reactions 1 and 2 as well as 3 and 4. This model overcame the drawback of the model of Kompala et al. (1986) to explain such simultaneous utilization due to the latter’s more rigid structure.

The techniques discussed above are extended to the modified version (Namjoshi & Ramkrishna, 2001). In order to demonstrate truly the model identification strategy, the model identification algorithm was employed on the abstracted network of Fig. 5. The reaction that forms \( C \) is presently not considered in the analysis. The pathways obtained are shown in Table 1. The last row of the table shows a pathway called global which is not identified by the model, but can be thought of as two topologically distant reactions, which can compete for producing different parts of cell mass at a global level. For instance, incorporating such a competition and lumping the intermediates \( M_1, M_2 \) in a cyclic pathway. A preliminary examination shows that model C due to its similarity to the Kompala model is incapable of explaining simultaneous utilization of substrates.

Model C, on the other hand, is closer to the diauxic model and links the two external substrates directly while letting the internal reactions produce the two growth precursors in a cyclic pathway. A preliminary examination shows that model C due to its similarity to the Kompala model is incapable of explaining simultaneous utilization of substrates.

Model A was simulated for the glucose–fumarate system (Namjoshi & Ramkrishna, 2001) under batch growth with \( c_1 = 1 \times 10^{-3} \) g/gdw, \( c_2 = 1 \times 10^{-3} \) g/gdw, \( v_j = 1, \ldots, 4 \), \( m_1 = 1 \times 10^{-3} \) g/gdw, \( v_j = 1, 2 \). The initial substrate concentrations were set to 0.1 g/l (glucose, \( S_1 \)) and 5 g/l (fumarate, \( S_2 \)), respectively, while biomass concentration was 0.02 g/l. A batch simulation where cells are grown for 15 h of culture time is used to generate the data. For a system with enzymes 1 and 2 competing with each other, one expects the results to look like

\[ \Gamma = \begin{bmatrix} + & − \\ − & + \end{bmatrix} \]  

(28)

Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Competitions involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1, 2</td>
</tr>
<tr>
<td>B</td>
<td>1, 3</td>
</tr>
<tr>
<td>C</td>
<td>1, 4</td>
</tr>
</tbody>
</table>

Combination of these pathways, with the added constraint that one reaction will be modeled only in one pathway, leads to essentially three different models. These are depicted in Table 2.

A model identification strategy should be able to identify model A since it is the original model that we will use to generate “data”. Model B, on the other hand, considers two linear pathways, which seek to maximize the end product in each pathway. Model C, on the other hand, is closer to the diauxic model and links the two external substrates directly while letting the internal reactions produce the two growth precursors in a cyclic pathway. A preliminary examination shows that model C due to its similarity to the Kompala model is incapable of explaining simultaneous utilization of substrates.
(1) Enzyme-only formulation:

\[
\begin{bmatrix}
 e_1 \\
e_2 \\
e_3 \\
e_4
\end{bmatrix}
\begin{bmatrix}
 r_{E,1} \\
r_{E,2} \\
r_{E,3} \\
r_{E,4}
\end{bmatrix}
\begin{bmatrix}
 -0.153 & -4.202 & 1.797 & 2.927 \\
0.466 & 0.761 & -1.309 & 0.239 \\
0.150 & -1.211 & 0.168 & 1.093 \\
0.370 & -0.558 & -1.127 & 1.478
\end{bmatrix}
\]

The enzyme-only formulation does not consider the models A, B or C, but lets all reactions compete against each other. As can be seen from the coefficients above, the inherent competitions between enzymes 1, 2 and 3, 4 are not revealed by the analysis.

(2) Enzymes and all metabolites:

\[
\begin{bmatrix}
 s_1 & s_2 & m_1 & m_2 & C' & e_1 & e_2 & e_3 & e_4
\end{bmatrix}
\begin{bmatrix}
 r_{E,1} \\
r_{E,2} \\
r_{E,3} \\
r_{E,4}
\end{bmatrix}
\begin{bmatrix}
 0.305 & 0.051 & 0.54 & -0.583 & 61.09 & -0.307 & -0.55 & 2.24 & -1.42 \\
0.098 & 0.308 & 1.27 & -1.64 & 70.85 & -0.255 & 4.1 & 1.48 & -4.91 \\
0.193 & 0.301 & 0.591 & -0.949 & 31.82 & -1.09 & 1.4 & 3.11 & -3.06 \\
0.141 & 0.3 & 0.793 & -1.15 & 39.93 & -0.688 & 2.12 & 1.76 & -2.81
\end{bmatrix}
\]

This approach is a slight improvement over the previous approach, as key metabolite measurements are included in addition to enzyme levels, but the result is the same, in that the coefficients do not reveal the inbuilt competitions.

(3) Enzyme groups and inducing metabolites: In this approach, we make use of enzyme groups that are proposed to compete against each other along with only the inducing substrate for each enzyme. Thus, we are testing out models A, B and C in Table 2 to see if any of the models are appropriately identified.

For model A, therefore, one obtains the following:

\[
\begin{bmatrix}
 s_1 & s_2 & m_1 & m_2 & C' & e_1 & e_2 & e_3 & e_4
\end{bmatrix}
\begin{bmatrix}
 r_{E,1} \\
r_{E,2} \\
r_{E,3} \\
r_{E,4}
\end{bmatrix}
\begin{bmatrix}
 0.344 & 0.000 & 0.00 & 0.00 & 0.102 & -0.239 & 0.00 & 0.00 \\
0.000 & 0.000 & 0.203 & 0.00 & -0.246 & 0.250 & 0.00 & 0.00 \\
0.000 & 0.182 & 0.00 & 0.00 & 0.00 & 0.779 & -0.756 & 0.00 \\
0.000 & 0.000 & 0.00 & 0.00 & 0.00 & -0.165 & 0.342 & 0.00
\end{bmatrix}
\]

This amply reveals that enzymes 1 and 2 are in competition as also are enzymes 3 and 4. Hence, model A seems like a candidate model for this system. This result is satisfying, because model A is used to generate the data. However, in order to test this technique further, we try the two other models. Model B results in:

\[
\begin{bmatrix}
 s_1 & s_2 & m_1 & m_2 & C' & e_1 & e_2 & e_3 & e_4
\end{bmatrix}
\begin{bmatrix}
 r_{E,1} \\
r_{E,2} \\
r_{E,3} \\
r_{E,4}
\end{bmatrix}
\begin{bmatrix}
 0.373 & 0.000 & 0.00 & 0.00 & 0.043 & 0.00 & -0.201 & 0.00 \\
0.000 & 0.000 & 0.281 & 0.00 & 2.370 & 0.00 & -2.468 & 0.00 \\
0.000 & 0.170 & 0.00 & 0.00 & 0.107 & 0.00 & -0.066 & 0.00 \\
0.085 & 0.000 & 0.00 & 0.00 & 0.682 & 0.00 & -0.506 & 0.00
\end{bmatrix}
\]

Clearly, this does not point toward any competition between the enzymes proposed. However, a surprise is in store in the case of model C where enzymes 1 and 4 are proposed to compete with each other along with enzymes 2 and 3. This leads to:

\[
\begin{bmatrix}
 s_1 & s_2 & m_1 & m_2 & C' & e_1 & e_2 & e_3 & e_4
\end{bmatrix}
\begin{bmatrix}
 r_{E,1} \\
r_{E,2} \\
r_{E,3} \\
r_{E,4}
\end{bmatrix}
\begin{bmatrix}
 0.349 & 0.000 & 0.00 & 0.00 & 0.132 & 0.00 & -0.278 & 0.00 \\
0.000 & 0.000 & 0.161 & 0.00 & 0.565 & 0.00 & -0.539 & 0.00 \\
0.000 & 0.161 & 0.00 & 0.00 & 0.464 & 0.00 & 0.464 & 0.00 \\
0.086 & 0.000 & 0.00 & 0.00 & -0.141 & 0.00 & 0.331 & 0.00
\end{bmatrix}
\]

This indeed shows competition between enzymes 1 and 2 as well as 3 and 4. This may seem intriguing at first, but considering that for the conditions tried, the system shows a diauxic behavior, it is not completely surprising that enzymes 1 and 4 which uptake the two substrates seem to be in competition as also enzymes 3 and 4 which are also turned on under complementary set of conditions.
Thus, our analysis has led us to conclude that models A and C are possible candidates, whereas B fares poorly in comparison. More data might help narrow down our selection to one candidate model. Therefore, in order to distinguish between the two models discussed above, another simulation is tried to excite the system to consume both substrates simultaneously. The model (Fig. 5) is simulated with $e_i = 2 \times 10^{-6} \text{g/gdw}, \forall i = 1, 4; e_i = 1 \times 10^{-6} \text{g/gdw}, \forall i = 2, 3; m_i = 1 \times 10^{-3} \text{g/gdw}, \forall j = 1, 2$. The initial concentration of glucose as well as fumarate is set to 5 g/l, while biomass concentration is 0.02 g/l. The simulation is carried out for 15 h to generate the data. By the choice of our initial conditions, we have favored the simultaneous utilization of both substrates, at least initially, by choosing high levels of concentrations. We have thus been able to excite the system sufficiently to reveal its underlying nature by generating more data.

As can be seen from the above matrices, the approach of Gupta et al. (2005) when modified to include the inducing substrates as well as subgroups of enzymes can serve as a guiding tool for the diagnosis of cybernetic competitions.

5. Future recommendations

With progressive understanding gained from more detailed experimental observations, the cybernetic framework is continuing to evolve in its formulation that will appear in future publications. The cybernetic modeling identification strategy presented in this paper may lead to a large number of possible cybernetic structures, some of which overlap. In that case, this might present a formidable optimization challenge which may be tackled from either a rigorous mathematical programming approach (Floudas, 2000; Floudas & Pardalos, 2000), or by the use of statistical approaches such as Genetic Algorithms (Goldberg, 1989; Holland, 1975; Sundaram, Ghosh, Caruthers, & Venkatasubramanian, 2001; Venkatasubramanian, Chan, & Caruthers, 1994). We present here a very preliminary development of how Genetic Algorithms may be employed for this task. Consider the sets of convergent ($\Phi$), divergent ($\Delta$), cyclic ($\Psi$) and linear pathways.
parameters. Each of the elements choose for them a few sample values spanning their range.

\[ \组成部分(Ω) \]

For instance, one can choose 10 different values for the \( k_i \) for each reaction, ranging from, say \( 10^{-3} \) to unity, and identify these hand-in-hand with the model structure.

The parameters are thus identified in conjunction with the model competitions. Getting the right cybernetic structure in a model is very crucial to the success of identifying a good model. Using the tools of Genetic Algorithms (Holland, 1975), one can thus generate an initial population of models, i.e. generate \( \Omega \) for which the least squares error can be determined by comparing model predictions to available experimental data. This first generation population of candidate models is ranked according to its fitness (calculated from the least square error). From this, a selected number of models with the highest as well as the lowest fitness may be transferred directly to the next generation. On the other hand, more candidates can be generated by carrying out crossover and mutation operations among these selected candidates. More candidates can be generated randomly to probe more areas of the search space. This continues from generation to generation until a suitable model is found with acceptable fitness. This tool, although not evaluated in this paper, seems extremely promising for large throughput cybernetic modeling analysis.

References


