Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways

Rui Alves\(^1,2,3\) and Michael A. Savageau\(^1,*\)

\(^1\)Department of Microbiology and Immunology, University of Michigan Medical School, 5641 Medical Science Building II Ann Arbor, MI 48109-0620 USA, \(^2\)Grupo de Bioquimica e Biologia Teoricas, Instituto Rocha Cabral, Calçada Bento da Rocha Cabral 14, 1250 Lisboa, Portugal and \(^3\)Programa Gulbenkian de Doutoramentos em Biologia e Medicina, Departamento de Ensino, Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 1800 Oeiras, Portugal

Received on July 28, 1999; revised on November 11, 1999; accepted on December 3, 1999

\textbf{Abstract}

Motivation: Mathematical models are the only realistic method for representing the integrated dynamic behavior of complex biochemical networks. However, it is difficult to obtain a consistent set of values for the parameters that characterize such a model. Even when a set of parameter values exists, the accuracy of the individual values is questionable. Therefore, we were motivated to explore statistical techniques for analyzing the properties of a given model when knowledge of the actual parameter values is lacking.

Results: The graphical and statistical methods presented in the previous paper are applied here to simple unbranched biosynthetic pathways subject to control by feedback inhibition. We represent these pathways within a canonical nonlinear formalism that provides a regular structure that is convenient for randomly sampling the parameter space. After constructing a large ensemble of randomly generated sets of parameter values, the structural and behavioral properties of the model with these parameter sets are examined statistically and classified. The results of our analysis demonstrate that certain properties of these systems are strongly correlated, thereby revealing aspects of organization that are highly probable independent of selection. Finally, we show how specification of a given behavior affects the distribution of acceptable parameter values.

Contact: Savageau@umich.edu

\textbf{Introduction}

The characterization of large and complex biochemical networks cannot be achieved with the direct intuitive approaches that have been successful for simpler model systems. The more systematic tools provided by mathematical modeling and computer analysis have become essential because they are especially well suited for organizing large amounts of data and representing nonlinear and parallel processes.

The most common method of constructing an appropriate model for a biochemical system has been the reductionist or bottom-up approach. The component parts are isolated and characterized, and then the resulting submodels are assembled into a model of the integrated system. For example, in the study of metabolic pathways, individual enzymes were isolated and kinetically characterized \textit{in vitro}; pathway models were then constructed by assembly of the individual rate laws. The fundamental problems inherent in this approach are three (Ni and Savageau, 1996):

1. failure to identify all the relevant components
2. failure to identify all the relevant interactions
3. failure to determine accurately all the relevant parameter values.

The associated practical problems are the enormous numbers of components and interactions that need to be identified and the difficulty of reproducing the conditions experienced by the components in their natural setting so that their parameter values can be accurately determined \textit{in vitro} (e.g. Clegg, 1984; Moore \textit{et al.}, 1984; Ovadi and Srere, 1996; Savageau, 1992; Sorribas \textit{et al.}, 1993). These problems have limited the success of the bottom-up approach (e.g. Albe and Wright, 1992; Antunes \textit{et al.}, 1996; Curto \textit{et al.}, 1998; Ni and Savageau, 1996; Shiraishi and Savageau, 1993).

An alternative method of constructing an appropriate model is often termed the reverse-engineering or top-down approach. Many of the variables are measured in
the intact system, and then one attempts to reconstruct the underlying model that produced these data. The fundamental problems in this case are:

1. selecting a mathematical representation that is sufficiently general so that one can be assured that it will encompass the system to be characterized

2. the theoretical limits on what can be identified (problem of identifiability: e.g. Chappell and Godfrey, 1992; Feng and DiStefano, 1991; Ginn and Cushman, 1992) when one can only measure a subset of the variables (problem of observability: e.g. Moheimani et al., 1996; Xu et al., 1996).

The practical problems are associated with the limits of the technologies currently available for measuring all the relevant variables. Although the top-down approach has long been applied in simple cases that illustrate the method (e.g. Brown et al., 1990; Diamond, 1975; Domnitz, 1976; Kargi and Shuler, 1979; Quant, 1993; Voit and Savageau, 1982), its use in biology is currently being driven by the new techniques coming out of the Human Genome Project that generate massive data sets (e.g. Brown and Botstein, 1998; Chu et al., 1997; DeRisi et al., 1997; Eisen et al., 1998; Somogyi et al., 1997; Törönen et al., 1999). Although the top-down approach shows considerable promise, it is unlikely that this method alone will provide a satisfactory solution to the problem of modeling large and complex biochemical systems.

If one is interested in modeling a specific system (e.g. tryptophane biosynthetic pathway of Escherichia coli), the best way to proceed is to measure all the necessary parameters of the system in the organism of interest and build the model based on those values. One could productively combine the bottom-up and top-down approaches described above (Bliss et al., 1982; Yanofsky and Horn, 1994). On the other hand, if one is interested in a generic class of systems (e.g. amino acid biosynthetic pathways in general) or if the measurements are impossible to perform with accuracy and precision, even a combination of the two approaches may not be adequate.

In this paper we propose a statistical approach for dealing with generic classes of biochemical systems. We apply this approach to a general three-step unbranched biosynthetic pathway with inhibitory feedback. This pathway is an abstraction from the collection of unbranched pathways responsible for the biosynthesis of amino acids (e.g. see http://www.genome.ad.jp/kegg/dblinks/map/map01150.html). The results of our analysis demonstrate that certain properties of these systems are strongly correlated, thereby revealing aspects of organization that are highly probable independent of selection.

Methods

Amino acid biosynthetic pathways and their regulation have been studied intensively for more than 40 years. There is widespread acceptance among cell physiologists that the principal role of these systems is to provide a homeostatically regulated supply of amino acid for protein synthesis. This role has been characterized in terms of several behaviors that can be described by quantitative criteria (Savageau, 1976) that will be elaborated upon in this paper.

Systemic description and analysis

An unbranched three-step pathway with feedback inhibition is depicted in Figure 1. The independent variable $X_4$ represents the cell demand for the end product $X_3$. If the cell requires large amounts of $X_3$, then the value of $X_4$ will be high; if small amounts of $X_3$ are required, then the value of $X_4$ will be low. The dynamic behavior of such a model can be described by a set of ordinary differential equations, one equation per intermediate. This set of equations can be approximated to the first order in logarithmic space, yielding another set of ordinary differential equations with the canonical form of an S-system (Savageau, 1969):

$$\frac{dX_1}{dt} = \alpha_1 X_0^{g_1} X_3^{g_2} - \alpha_2 \prod_{j=1}^{3} X_j^{g_3}$$

$$\frac{dX_2}{dt} = \alpha_2 \prod_{j=1}^{3} X_j^{g_3} - \alpha_3 \prod_{j=2}^{3} X_j^{g_4}$$

$$\frac{dX_3}{dt} = \alpha_3 \prod_{j=2}^{3} X_j^{g_4} - \alpha_4 X_3^{g_3} X_4^{g_4} .$$

The multiplicative parameters, $\alpha$, can be interpreted as rate constants that are always positive. The exponential parameters, $g$, can be interpreted as kinetic orders that
represent the direct influence of each species on each rate law. If $X_i$ is directly involved in the reactions of the aggregate rate law $V_j$, as either a substrate or a modulator, and if an increase in $X_i$ causes an increase in the rate $V_j$, then the kinetic order will be positive. If an increase in $X_i$ causes a decrease in $V_j$, then the kinetic order will be negative. If $X_i$ is not directly involved in $V_j$, then the kinetic order will be zero. The kinetic orders $g_{i+1, i}(0 \leq i \leq 3)$ in (1) are positive because these are the kinetic orders for substrates of reactions. The kinetic order $g_{44}$ will be set arbitrarily equal to 1 for the remainder of this paper in order to simplify subsequent calculations. This will not affect our results in any significant way since $g_{44}$ is simply a scale factor in logarithmic space. The remaining kinetic orders, which represent negative feedback interactions, are negative.

At a steady state, the rate of production and the rate of consumption will be equal for each intermediate, and (1) reduces to the following matrix equation (Savageau, 1969):

$$
\begin{bmatrix}
    b_1 - g_{10}Y_0 \\
    b_2 \\
    b_3 + Y_4
\end{bmatrix}
= \begin{bmatrix}
    a_{11} & a_{12} & a_{13} \\
    a_{21} & a_{22} & a_{23} \\
    a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
    Y_1 \\
    Y_2 \\
    Y_3
\end{bmatrix}
$$

(2)

where $b_1 = \ln(a_{i+1}/a_i), a_{ij} = g_{ij} - g_{i+1,j}$ and $Y_i = \ln(X_i)$. This linear equation is easily solved, e.g. using Cramer’s rule, to provide a steady-state expression in symbolic form for each $Y_i$.

Other steady-state magnitudes of interest can be calculated in a similar way. Logarithmic gains quantify the influence of each independent variable on each dependent variable; e.g. the logarithmic gain

$$
L(X_i, X_0) = \frac{d \ln(X_i)}{d \ln(X_0)} = \frac{dY_i}{dY_0}
$$

gives the percentage change in an intermediate $X_i$ caused by a percentage change in $X_0$. These logarithmic gains are calculated analytically at the steady state (Savageau, 1971) by differentiating each $Y_i$ with respect to $Y_0$.

Parameter sensitivities quantify the influence of each parameter on each dependent variable of the system; e.g. the sensitivity

$$
S(X_i, p_j) = \frac{d \ln(X_i)}{d \ln(p_j)} = \frac{dY_i}{d \ln(p_j)}
$$

gives the percentage change in the concentration $X_i$ caused by a percentage change in the parameter $p_j$. These parameter sensitivities and those of the steady-state flux are also calculated analytically at the steady state. The parameter sensitivities give important information about the sensitivity of the system to perturbations in its structure.

The steady state for an unbranched biosynthetic pathway should be locally stable; i.e. the system should return to its original steady state after a small perturbation in the variables (as opposed to the parameters) of the system. If this does not occur, the system is dysfunctional. The stability can be determined by using the well-known Routh criteria (Savageau, 1976).

Any of these systemic properties can be analytically determined in the steady state by using the S-systems local representation. However, having an analytical expression for these systemic properties is just the first step in the analysis of a system. Interpretation of these analytical expressions can be problematic because they depend on many parameters and their behavior is too complex for easy visualization. Even when a general qualitative interpretation can be obtained just by looking at the closed-form expressions [e.g. $L(X_3, X_0) < L(X_1, X_0)$], the results are difficult to quantify [e.g. how much larger is $L(X_1, X_0)$?].

Also, there are no general closed-form solutions for the dynamic properties of the system. To analyze these properties one must specify numerical values for the parameters and solve the differential equations (1) using numerical techniques. An example of such a property is the settling time of a system, which is defined as the time required for a system to return to its steady state after a perturbation in the levels of its metabolites. The settling time also gives us an indication of the average transit time for material passing through the system. Short transit times allow a system to respond rapidly to changes in its environment (Savageau, 1972).

**Defining classes of systems for statistical comparison**

If one wishes to understand the general properties of pathways such as the one depicted in Figure 1, then one faces the following dilemma. General results that follow from the closed-form analytical expressions may be too complex to interpret and quantify, and quantitative results for particular values of the parameters do not yield general insights. One way of resolving this dilemma is to study the statistical properties for a class of systems generated by an ensemble of sets of parameter values. We shall consider two different methods for defining the class of interest.

**Structural classes.** Systems that have the same network topology (i.e. have the same pattern of interactions among their elements and the same signs for the interactions) will be defined as members of the same structural class. As a case study for this paper we have chosen the system in Figure 1 and described its local behavior by the S-system representation within the power-law formalism. By so doing we have defined a specific class of systems that share the same network topology. By focusing on such a topology we have limited the study to systems belonging...
to the same structural class. Individual members of this structural class can be generated by sampling the space of parameters that define the class and their characteristics can be obtained from the corresponding solutions of (1).

Behavioral classes. Systems that exhibit a specific type of systemic behavior will be defined as members of the same behavioral class. For example, those systems belonging to the structural class in Figure 1 that have a single locally stable steady state can be defined as members of a behavioral class. Individual members of such a behavioral class cannot be generated directly by sampling at random the space of parameters because some of the parameter sets will produce unstable systems. Instead, they must be generated indirectly, e.g., by sampling at random the space of parameters, testing the sample for the desired behavior, and then retaining only the relevant samples.

In the example above the behavioral class is a subclass within the structural class, but this need not be so. If our only knowledge of the system was that it had three metabolites, we could study an ensemble of models in which each kinetic order might have positive or negative values, which generates models belonging to different structural classes. One could then choose models for study based simply on their behavior, disregarding the signs of the kinetic orders.

Several (elementary) behavioral classes can be combined to define a composite behavioral class whose members are systems that exhibit all of the individual systemic behaviors.

Sampling the parameter space

The regular structure of the local S-system representation facilitates building the ensemble of sets of parameter values. The positive kinetic orders $g_{1+1}$ refer to enzymes binding their substrates. The maximum value for these kinetic orders is given by the number of substrate binding sites on the enzyme.1 In the majority of cases there are less than four such sites (Hlavacek and Savageau, 1995; Voit and Savageau, 1987). Thus we will assume that these kinetic orders have values between 0 and 5. The negative kinetic orders ($g_{13}, g_{22}, g_{23}$ and $g_{33}$) refer to enzymes binding inhibitors. In most cases there are again fewer than four such binding sites per enzyme, and we will assume that these kinetic orders have values between $-5$ and 0. One can always normalize the time scale with respect to one of the rate constants. The others will be assumed to have normalized values within 5 orders of magnitude of 1. Thus, the logarithm of each normalized rate constant will have values between $-5$ and 5.

In building an appropriate ensemble of sets of parameter values one needs to use a representative sample of the allowable parameter space. Since the statistical distribution of parameter values in real-life systems is unknown, the most appropriate approach is to sample the space uniformly. There are several strategies for accomplishing this.

First, one can impose a regular grid on the multidimensional parameter space and use the vertices of that grid to define the set of parameter values. In general, a system with $n$ unknown parameters and the same grid size, $ω$, will require $ω^n$ samples. This exponential increase in number of required samples makes it difficult to maintain a dense grid as the number of parameters increases. Also, maintaining a rigid grid complicates matters when one is studying ensembles of parameter sets that give rise to certain types of systemic behavior. Second, pseudo-random number generators can be used to generate the largest possible sample size without having a rigid grid to sample from. This method facilitates the study of ensembles of parameter sets that give rise to certain types of systemic behavior. Third, strategies based on number theory can be used to generate what are known as quasi-random numbers that are uniformly distributed. Examples include Halton and Sobol sequences (for a review see Bratley and Fox (1988)). Finally, another technique devised for dealing with large parameter spaces is the Latin Hypercube. The Latin Hyper-cube ensures that each parameter will be sampled in every one of its sub-ranges. It has no advantage over the other methods mentioned above if there are important interactions between parameters (for a discussion see Dunn and Clark (1974)). For the results reported below we have used the pseudo-random number generator.

Specifying behavioral classes

Since the system in Figure 1 is an abstraction of an unbranched biosynthetic pathway, the literature was searched and a basic number of desirable characteristics have been found for such systems. The group of all these characteristics was used to define a composite behavioral class. If the model generated by a given set of parameter values did not belong to this class, then the set was discarded and a new random set was tested. In this way we generated ensembles of 5000 for our studies.

The composite behavioral class studied is defined by a collection of six elementary behavioral classes with the following characteristics:

B1. The steady-state concentration of pathway intermediates should be low when compared with the concentration of the final product. The major function of unbranched biosynthetic pathways is production of their end product (e.g., $X_3$ in the example of Figure 1). High concentrations of intermediates per se

---

1This is not always true of reversible reactions operating close to equilibrium. The usual strategy for aggregating fluxes can lead to kinetic orders with extremely large absolute values. This problem can be solved by using an alternative strategy for aggregating fluxes (Sorribas and Savageau, 1989). However, we will not deal with these cases here.
are unnecessary; they would tax the solvent capacity of the cell and potentially interfere in a non-specific way with otherwise unrelated reactions (e.g. Atkinson, 1969; Savageau, 1972; Srere, 1987 and Levine and Ginsburg, 1985, for a general discussion of the subject from different perspectives). For the results presented in the next section, a parameter set was accepted only if the steady-state ratio $(X_1 + X_2)/X_3 \leq 0.1$. This value for the ratio was chosen arbitrarily because there are no reliable measurements on which to base a more accurate estimate.

B2. Changes in the concentration of intermediates caused by changes in demand for the end product should be small. The previous condition ensures that the concentration of intermediates will not saturate the solvent capacity of a cell in a given steady state. However, if the metabolic conditions change and the demand for the end product of the pathway changes, this will cause the concentration of each intermediate to change, which may lead to saturation of the solvent capacity in the new steady state (e.g. Savageau, 1972). This could be prevented in our model if the absolute values of the logarithmic gains for intermediates, $|L(X_i, X_4)|$, are smaller than a predetermined value arbitrarily set at 0.5.

B3. Changes in the concentration of intermediates caused by changes in the initial substrate should be small. This will buffer the intermediate concentrations against changes in metabolism that are reflected in alterations in the level of initial substrate. This could be ensured if the absolute values of the logarithmic gains for intermediates, $|L(X_i, X_0)|$, are smaller than a predetermined value, e.g. $|L(X_i, X_0)| < 0.5 (i = 1, 2, 3)$. This value is chosen arbitrarily because there are no reliable measurements on which to base a more accurate estimate.

B4. Systems should be robust, i.e. insensitive to spurious fluctuations in the parameters that define their structure (Savageau, 1972). We require that each intermediate have an aggregate sensitivity, defined as $\sqrt{\sum_j S(X_i, p_j)^2}$, less than a predetermined value arbitrarily set equal to 5.

B5. Each system should have a locally stable steady state. Systems without such stable steady states are dysfunctional because they are unable to maintain their homeostatic behavior in the face of spurious perturbations. The two margins of stability can be specified in terms of the last two Routh criteria (e.g. Savageau, 1976).

B6. Systems should have a rapid response time. This is related to the inverse of the turnover number (Dixon, 1958; Savageau, 1975), which should therefore be high. We require the turnover number for the pathway, defined as the pathway flux divided by the sum of the intermediate pools $(V/\sum X_i)$, to be larger than a predetermined value arbitrarily set equal to 1.

Results

Bias in the frequency distribution of parameter values

The values for each parameter were originally sampled with a uniform distribution. However, those parameter sets that define systems excluded from the composite behavioral class are rejected, and the frequency distribution of the accepted parameter values is therefore biased. The nature of the bias for each of the parameters can be determined from the histograms presented in Figure 2. We observe that the composite behavioral class has $\alpha_1$ biased towards small values whereas $\alpha_2$, $\alpha_3$, and $\alpha_4$ are biased towards large values. The kinetic order for the substrate of the pathway, $g_{10}$, is biased towards small values. Its frequency increases from $g_{10} = 0$ to $g_{10} = 0.3$ and then decreases exponentially until $g_{10} = 5$. A similar pattern is observed for $g_{32}$, although the frequency increases from 0 to 1.8, and then decrease but not exponentially. The kinetic order $g_{21}$ is biased towards large values, and $g_{43}$ is nearly uniform over its range. The inhibitory kinetic order for overall feedback, $g_{13}$, has a distribution with a central tendency, whereas the other inhibitory kinetic orders are almost uniformly distributed throughout their range of possible values.

We also determined the parameter distributions for each of the elementary behavioral classes (B1–B6 defined above) to see which, if any, might qualitatively reproduce the deviations from a uniform distribution that were observed for the composite behavioral class (Figure 2). Table 1 shows which elementary class is mainly responsible for the shape of each distribution in the composite behavioral class. In some cases, the distribution for the composite behavioral class can be attributed to the dominant influence of a particular elementary class (e.g. B3 in the case of $g_{10}$). In other cases, the distributions for the composite behavioral class can be attributed to the influence of several elementary classes acting in combination, which implies a synergistic influence (e.g. B1–B6 in the case of $\alpha_3$).

Frequency distribution for systemic properties of the ensemble

The frequency distributions for all steady-state properties of our model have long tails. These tails make it difficult to present informative histograms for each of the systemic
Fig. 2. Distribution of parameter values in ensembles of systems selected on the basis of various behavioral classes. Selection involved each of the six elementary behavioral classes (B1–B6) considered separately and the composite class consisting of all six elementary classes considered together. The solid line in each panel is the distribution for the composite behavioral class. Three different patterns are represented. In most cases the distribution for the composite class is closely represented by the distribution for one of the elementary classes (α1, α2, α4, g10, g22, g33). The distributions for the other elementary classes have very different shapes and are not shown. In four cases the distribution for the composite class is closely resembled by two or more of the distributions for the elementary classes (g21, g43, g13, g23). Distributions for only two of the elementary classes are shown. In two cases none of the distributions for the elementary classes is a close match to the distribution for the composite class (α3, g32). In these cases we show only the distribution for the elementary class that most closely resembles the distribution for the composite class.

properties. We chose to cut off the tails and add their frequency to the more extreme classes presented in the histograms. The results in this section are shown as histograms in Figure 3. We did not include histograms for the elementary behavioral classes because, in most cases, they have extremely long tails.

Steady-state concentrations and flux. All steady-state concentrations have frequency distributions that decrease as the concentration increases. At low concentrations the frequency decreases very sharply as the concentrations increase, but then the decrease becomes very small and there is a long tail in the distribution. The modal class for all of the frequency distributions is small. For X1 and X2 the modal class is in the interval [0, 0.2334] with 90% and 75% of all systems in this interval, respectively. The modal class for X3 has a larger value, in the interval [0.234, 0.468]. Also, only 10% of all systems fall within
Fig. 3. Distributions for the different systemic properties. The distributions are constructed from an ensemble of systems selected on the basis of the composite behavioral class. The extreme values at the ends of some distributions are indicative of a long tail that was cut off and included in the last interval of values.
Table 1. Influence of the elementary behavioral classes on the distribution of parameter values for systems in the composite behavioral class

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Elementary behavioral class</th>
</tr>
</thead>
<tbody>
<tr>
<td>g_1</td>
<td>+</td>
</tr>
<tr>
<td>g_2</td>
<td></td>
</tr>
<tr>
<td>g_3</td>
<td>+</td>
</tr>
<tr>
<td>g_4</td>
<td></td>
</tr>
<tr>
<td>g_10</td>
<td>+</td>
</tr>
<tr>
<td>g_13</td>
<td>+ + +</td>
</tr>
<tr>
<td>g_21</td>
<td>+ +</td>
</tr>
<tr>
<td>g_22</td>
<td></td>
</tr>
<tr>
<td>g_23</td>
<td>+ + +</td>
</tr>
<tr>
<td>g_32</td>
<td>+</td>
</tr>
<tr>
<td>g_33</td>
<td>+</td>
</tr>
<tr>
<td>g_43</td>
<td>+ +</td>
</tr>
</tbody>
</table>

+ Signs indicate the major influence. a Two or more classes have similar distributions for this parameter. b None of the individual classes shows a similar distribution for this parameter. The distribution results from a synergism among all the classes.

Logarithmic gains. The logarithmic gains of X_1 and X_2 with respect to X_3 can be positive or negative, depending on parameter values, whereas the logarithmic gain of X_3 with respect to X_4 is always negative. The logarithmic gain of V with respect to X_4 is always positive. The distribution for L(X_3, X_4) is skewed with no values less than 0.1 in magnitude and a modal class in the interval [−0.155, −0.150]. On the other hand, the distributions for L(X_1, X_4), L(X_2, X_4) and L(V, X_4) have a symmetrical central tendency; the modal class for L(X_1, X_4) and L(X_2, X_4) is in the interval [0.025,0.035] and that for L(X_3, X_4) is in the interval [−0.495, −0.505].

All logarithmic gains with respect to X_0 are positive. The distribution for L(X_1, X_0) is almost uniform in its allowable range. The distributions for L(X_2, X_0), L(X_3, X_0) and L(V, X_0) are biased toward low values; L(X_2, X_0) has a modal class in the interval [0.025,0.035], whereas both L(X_3, X_0) and L(V, X_0) have a modal class in the interval [0.005,0.015]. The maximum observed values for these gains was 1.7.

Parameter sensitivities. We determined the frequency distribution of the aggregate sensitivities for each parameter. For the intermediates X_1 and X_2, the modal class of the aggregate sensitivities is in the interval [4.9,5], which includes the largest allowable values, and the percentage of systems with aggregate sensitivities in this interval is smaller for X_2 than for X_1. For the end product X_3, the modal class is in the interval [0.8,0.9]. Taken together, these data show that the aggregate parameter sensitivities of the concentrations decrease along the pathway. The distribution for the aggregate sensitivity of the steady-state flux is similar to that for X_3.

Stability margins. The frequency distributions for both stability margins are skewed. Frequencies increase until the margins reach values of about 10^5 (12 in log base e) for the second stability margin and 0.5 × 10^5 (20 in log base e) for the third stability margin. Beyond these values the frequencies decrease abruptly.

Response times. The frequency distribution for the inverse of the turnover number for the pathway (which is related to response time) appears to decay almost exponentially. The modal class is in the interval [0,0.2].

Correlations between different systemic properties

Moving median plots can reveal the degree of correlation between various magnitudes of interest (Alves and Savageau, 2000). We have built 170 such plots for each behavioral class to look for correlations among the systemic properties of the model in Figure 1. It is not feasible to present this large number of plots here. Instead, we illustrate in Figure 4 the 11 different shapes that we have observed for the curves in these plots. It must be emphasized that the curves in Figure 4 are idealized; the actual plots have considerably more noise. The results for this section are summarized in Table 2 by making reference to these idealized shapes.

Concentrations, fluxes and logarithmic gains. There is, on average, a direct correlation between the concentrations of all metabolites in the system. If one metabolite has a high concentration then, on average, the other metabolites will also have high concentrations. Concentrations are also directly correlated with fluxes.

There is also a direct correlation between the logarithmic gains in concentration for the different metabolites and between the logarithmic gains in concentration and the logarithmic gains in flux. On the other hand, there is an inverse correlation between logarithmic gains (in concentration and flux) and concentrations and between logarithmic gains (in concentration and flux) and flux. As the concentration of metabolites in the system increases, the logarithmic gains (in concentration and flux) decrease. The logarithmic gains
Fig. 4. Qualitatively different shapes for the correlation curves between different systemic properties. The correlation is determined by a plot of the moving median for one property versus the moving median for another constructed from an ensemble of systems selected on the basis of various behavioral classes (see Alves and Savageau, 2000). The 11 shapes (C1–C11) include all the tabulated shapes found by examination of the actual graphs. These shapes are referenced in Tables 2 and 3.

in concentration also decrease along the pathway; i.e. \( \langle L(X_1, X_0) \rangle > \langle L(X_2, X_0) \rangle > \langle L(X_3, X_0) \rangle \).

Parameter sensitivities. There is an inverse correlation between each concentration and each aggregate sensitivity for the individual concentrations. On average, as the concentrations increase, the aggregate sensitivities decrease. This also occurs in the case of aggregate sensitivities for the flux; as the concentrations (and fluxes) increase, the sensitivities decrease.

There is a direct correlation between any two aggregate sensitivities for the metabolites in the system. Thus, if any given \( X_i \) has a high aggregate sensitivity, on average, the other \( X_j \) will also have a high aggregate sensitivity.

Stability margins and response times. There is a direct correlation between the values of the two stability margins. This means that on average if one of the stability margins has a high value so does the other. Also, there is a direct correlation between the stability margins (response times) and the concentrations in the model when the concentrations are low; as the concentrations increase, the correlations disappear or are reversed. This indicates that on average systems with intermediate values for the concentrations have larger values for the stability margins. We also observe an inverse correlation between the stability margins and the aggregate sensitivities for each concentration and flux.

Influence of elementary behavioral class on the correlations between systemic properties

We determined which elementary behavioral classes were primarily responsible for the correlations observed for the composite behavioral class. To do this we defined the same elementary behavioral classes as before (B1–B6) and then determined the correlations for each of these (data not shown). In the case of the correlations involving concentrations, the qualitative nature of the correlations for the composite behavioral class is mainly determined by the elementary class B1, which requires low concentrations of intermediates in the steady state. This is so because most of the correlations are symmetric for the classes B2–B6 (e.g. C8 and C9), whereas they are asymmetric for class B1 (C1 and C2). In all other cases, there is no particular pattern that is apparent.
rate-constant parameters in
Table 3 shows how each of the individual parameters
In Statistical synergisms between different parameters. The shape of the curves C1
Table 2.
Each correlation is determined by a plot of the moving median for the $i^{th}$ systemic property versus the moving median for the $j^{th}$ systemic property constructed from an ensemble of systems selected on the basis of the composite behavioral class. See text for discussion.

<table>
<thead>
<tr>
<th>y axis</th>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>V</th>
<th>L(.,X₀)</th>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>V</th>
<th>L(.,X₄)</th>
<th>X₁</th>
<th>X₂</th>
<th>X₃</th>
<th>V</th>
<th>S(.,p)</th>
<th>2nd</th>
<th>3rd</th>
<th>Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>X₁</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₂</td>
<td>C3</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X₃</td>
<td>C3</td>
<td>C1</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>C1</td>
<td>C1</td>
<td>C1</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(X₁,X₀)</td>
<td>C3</td>
<td>C3</td>
<td>C8</td>
<td>C8</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(X₂,X₀)</td>
<td>C3</td>
<td>C4</td>
<td>C8</td>
<td>C8</td>
<td>C1</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(X₃,X₀)</td>
<td>C3</td>
<td>C3</td>
<td>C8</td>
<td>C1</td>
<td>C3</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(V,X₀)</td>
<td>C3</td>
<td>C3</td>
<td>C3</td>
<td>C3</td>
<td>C1</td>
<td>C3</td>
<td>C3</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(X₁,X₄)</td>
<td>C3</td>
<td>C1</td>
<td>C3</td>
<td>C3</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C2</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(X₂,X₄)</td>
<td>C3</td>
<td>C1</td>
<td>C3</td>
<td>C3</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C2</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(X₃,X₄)</td>
<td>C3</td>
<td>C2</td>
<td>C3</td>
<td>C3</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C11</td>
<td>C3</td>
<td>C3</td>
<td>C3</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L(V,X₄)</td>
<td>C3</td>
<td>C3</td>
<td>C3</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C10</td>
<td>C4</td>
<td>C3</td>
<td>C3</td>
<td>C2</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(X₁,p)</td>
<td>C4</td>
<td>C4</td>
<td>C4</td>
<td>C8</td>
<td>C1</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
<td>C8</td>
<td>C2</td>
<td>C2</td>
<td>C1</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(X₂,p)</td>
<td>C4</td>
<td>C4</td>
<td>C4</td>
<td>C8</td>
<td>C1</td>
<td>C1</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
<td>C1</td>
<td>C4</td>
<td>C2</td>
<td>C1</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(X₃,p)</td>
<td>C4</td>
<td>C3</td>
<td>C8</td>
<td>C1</td>
<td>C1</td>
<td>C1</td>
<td>C1</td>
<td>C10</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
<td>C8</td>
<td>C1</td>
<td>C3</td>
<td>C1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S(V,p)</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C11</td>
<td>C8</td>
<td>C10</td>
<td>C9</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C1</td>
<td></td>
</tr>
<tr>
<td>2nd</td>
<td>C4</td>
<td>C4</td>
<td>C4</td>
<td>C9</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C4</td>
<td>C7</td>
<td>C2</td>
<td>C2</td>
<td>C1</td>
<td>C10</td>
<td>C1</td>
<td>C9</td>
<td>C7</td>
<td>C1</td>
</tr>
<tr>
<td>3rd</td>
<td>C4</td>
<td>C4</td>
<td>C4</td>
<td>C9</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C4</td>
<td>C7</td>
<td>C2</td>
<td>C2</td>
<td>C1</td>
<td>C10</td>
<td>C1</td>
<td>C9</td>
<td>C7</td>
<td>C1</td>
</tr>
<tr>
<td>Turnover</td>
<td>C10</td>
<td>C1</td>
<td>C1</td>
<td>C4</td>
<td>C7</td>
<td>C7</td>
<td>C7</td>
<td>C1</td>
<td>C7</td>
<td>C7</td>
<td>C1</td>
<td>C2</td>
<td>C2</td>
<td>C2</td>
<td>C4</td>
<td>C2</td>
<td>C2</td>
<td>C1</td>
</tr>
</tbody>
</table>

Table 3. Statistical synergisms between different parameters. The shape of the curves C1–C11 are shown in Figure 4. Each statistical synergism is determined by a plot of the moving median for the sensitivity with respect to parameter $p_i$ versus the moving median for parameter $p_j$, constructed from an ensemble of systems selected on the basis of the composite behavioral class.

<table>
<thead>
<tr>
<th>p_j</th>
<th>S(X₁,p₁)</th>
<th>S(X₂,p₂)</th>
<th>S(X₃,p₃)</th>
<th>S(V,p₄)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a₁</td>
<td>a₂</td>
<td>a₃</td>
<td>a₄</td>
<td>8</td>
</tr>
<tr>
<td>a₁</td>
<td>a₂</td>
<td>a₃</td>
<td>a₄</td>
<td>8</td>
</tr>
</tbody>
</table>
| a₁  | C9       | C7       | C8       | C8      | C8     | C2       | C2       | C2       | C1       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1       | C1       | C2       | C2       | C1� of parameters on sensitivities

Table 3 shows how each of the individual parameters statistically influences the parameter sensitivities of systems in the composite behavioral class. In general, the rate-constant parameters influence fewer sensitivities than the kinetic-order parameters.

In Figure 5 $S(X₃,p)$ is plotted as a function of $p_j$.

Influence of parameters on sensitivities

The slope of the curve can be interpreted as a statistical synergism between the parameters $p_i$ and $p_j$. This slope measures the average change in the average sensitivity to $p_i$ caused by a change in the average value of $p_j$. Unlike the differential synergisms defined by Salvador (1999a,b), which are always symmetric, these statistical synergisms are not necessarily symmetric.
of providing an exhaustive statistical characterization of their systemic properties.

In this work we have created large ensembles of randomly generated parameter values for a given structural class of biochemical systems and imposed selection on the basis of particular systemic properties. We then examined the resulting systems for bias in parameter values, bias in unselected systemic properties, and correlations among all their systemic properties.

Selection can be expected to influence the range of parameter values in the resulting systems. Although specific systemic properties have been used for some time as criteria to evolve networks towards optimality (e.g. James et al., 1999), few, if any, attempts have been made to characterize the bias in parameter values that results from such a selection procedure. In fact, the usual view on the subject is that parameter values determine systemic behavior. We have had to take the opposite view to learn how selection for particular systemic behaviors influences the frequency distribution of parameter values. As seen in Figure 2, there are regular patterns of deviation from what was a uniform distribution before imposing selection based on the composite behavioral class. By using each of the elementary behavioral classes as an independent selection criterion we were able to determine whether any given elementary class made a major contribution to the observed bias in the distribution of any given parameter value. In some cases this is true (B3 in the case of \( g_{10} \)), whereas in others the distribution of parameter values for the composite class is the result of interplay among different elementary classes (B1–B6 in the case \( a_3 \)).

Information on the distribution of parameter values is of interest in the design of experiments to measure the parameters in actual systems. By knowing the most probable values of a parameter, one can design experiments to target that range. Also, the use of behavioral classes to study specific kinds of systems provides an effective way to identify the relative importance of various regions in the parameter space of fit systems.

Selection for a particular systemic property may also influence other unselected systemic properties. As seen in Figure 3, selection on the basis of the composite behavioral class produces a frequency distribution for the values of the different systemic properties that is skewed in nearly every case, with a peak at low values and a long tail that decreases in frequency almost exponentially. (This is true of the distributions for the aggregate sensitivities, although it is not evident from the curves for \( X_1 \) and \( X_2 \) because their tails are off the scale.) The exceptions to this general pattern are the distributions for the logarithmic gain \( \log(X_1, X_0) \), which is nearly uniform over the range \([0,5]\), and the logarithmic gains \( \log(X_1, X_4) \), \( \log(X_2, X_4) \), and \( \log(V, X_4) \), which exhibit a symmetric central tendency. We have also determined the

**Discussion**

The study of generic biochemical systems requires a mathematical formalism that is systematically structured and capable of representing rather arbitrary nonlinear phenomena. The power-law formalism provides such a canonical nonlinear representation (Savageau, 1996), and the power-law representation provides a canonical nonlinear representation (Savageau, 1996), and for the work presented in this paper we have focused on the local S-system representation within this formalism. This representation, although nonlinear, has closed-form solutions for the steady state, and these can be used to study systemic properties analytically. However, more often than not, the complexity of the solutions for the properties of interest makes it difficult to analyze systemic behavior without assigning specific values to the parameters. In most cases these values are unknown; when they are known, they limit the interpretation of the results to a specific system and thus prevent generalization of the results. To overcome these limitations statistical studies involving large ensembles of random systems have been performed in a variety of contexts (see, e.g. Bhattacharjya and Liang, 1996; Glass, 1975; Kauffman, 1969a,b, 1993, and references therein). However, to our knowledge this approach has not been applied to continuous models for specific classes of biochemical systems with the objective
Ensembles of metabolic networks

influence of the elementary behavioral classes on these distributions, but the results are less straightforward to interpret. In many cases the distribution for the composite class is quite different from the distribution for any of the elementary classes (data not shown). This indicates a strong synergism between different constraints that determine the distribution of values for the systemic properties.

Selection also can be expected to influence the correlations among the various systemic properties in the resulting systems. We have used a moving median technique (Alves and Savageau, 2000) to determine average correlations between different systemic properties. We found that these correlations exist and are, at least in some cases, dependent on the behavior of the system (Table 2). For example, most aggregate sensitivities are correlated with the concentrations by a symmetric curve of type C8 if no restrictions are imposed on the values of the concentrations in the system (data not shown). However, when we imposed the condition that intermediate concentrations be small (B1), this kind of symmetry breaks down (curves of type C8 and C9 become C3 and C4), because the systems being studied include only those with concentrations that have low values. It is important to note that, as a concentration tends to unity, the sensitivities to the kinetic orders associated with that concentration will tend to zero (due to the properties of the power law in the S-system formalism). This will tend to diminish the aggregate sensitivities that include these kinetic-order sensitivities. Table 2 also shows that in the composite class, robustness of intermediates and stability margins are inversely correlated; systems that have large stability margins have, on average, intermediates with high aggregate sensitivities and are thus less robust.

Finally, the same technique used to determine the correlations between different systemic properties also was used to determine the statistical synergisms between different parameters. The system in Figure 1 has small synergisms for the end product and flux (Table 3), because in many cases (54 out of 120) the sensitivities are not correlated with any parameter (statistical synergism is zero). Thus, the end product and the steady-state flux of the system are, on average, well buffered against second-order perturbations.

The approach illustrated in this paper provides statistical insights. It might be argued that biological systems are optimized and atypical, and thus not compatible with the application of statistical techniques. However, this objection is avoided in our approach. By defining behavioral classes for optimized systems, we are able to study the average behavior of optimized systems and not just the average behavior of random systems.

The methods we have described can in principle be applied to systems with more complex behaviors. For example, suppose we wish to consider biochemical systems that are capable of exhibiting either a single locally stable steady state (nongrowing cells that are viable but quiescent) or a single stable limit cycle (growing cells with a well-defined cycle time), depending only upon the value of an environmental cue. The behavioral classes that we would define for such systems would now include the combined properties of these two different modalities as well as the properties that might be applied to each of the separate modalities. The more complex behavioral class would include a number of dynamic properties (e.g. the period, amplitude, phase, and robustness of the oscillation, and the bifurcation value of the environmental cue for switching between modalities), and the analysis necessary to identify and characterize these behavioral classes would accordingly become more complex. The local S-system representation is capable of describing each of the separate modalities (Lewis, 1991), but not the two of them together with a given set of parameter values. For this purpose we would need the generalized-mass-action representation within the power-law formalism. This representation does not have analytical solutions for the steady state, and so the analysis and comparison of these properties would have to be done by numerical methods. Randomly generated sets of parameter values (which would now include values for a parameter representing the environmental cue) could be generated as before. However, we would now select only those sets of parameter values that satisfy the more complex behavioral class that includes both modalities and the appropriate switching between them in response to the environmental cue. Those sets of parameter values that only yield one of the two modalities would be excluded from consideration. This would ensure that any averaging procedure that is subsequently applied to systems with the randomly generated parameter sets would range over a homogeneous class of systems.

The approach proposed in this work also may be useful in providing information about systems that are poorly characterized. For example, suppose we know the structure of a system, but we are able to determine experimentally only some of the characteristic behaviors of the system. To be more specific, suppose we know that the concentrations of the system are within a given range, that increasing the value of a given independent variable will always cause a decrease in the values for some dependent variables, and that we are able to measure the range of values for the turnover times of the concentrations. With this information, we could generate ensembles of systems with the described characteristics and study them statistically. The results would allow us to make predictions about other systemic properties that might be measured and, for the unknown parameters,
about the range of values most likely to generate systems with the known behavior.

A combination of approaches will surely be needed to advance our understanding of large and complex systems in biology. We need to take advantage of the broad-scale capabilities of the top-down genomic technologies and the structural constraints provided by the more traditional bottom-up methodologies of molecular biology. We also need to identify the systemic regularities that exist even in randomly constructed networks. The approach presented in this paper appears well suited for the determination of such regularities in continuous models. It may facilitate the design of experiments to measure parameters by the bottom-up approach as well as provide a suitable framework to determine classes of models that give a good fit to data obtained by the top-down approach.

Acknowledgments

This work was supported in part by a joint PhD fellowship PRAXIS XXI / BD / 9803 / 96 granted by PRAXIS XXI through Programa Gulbenkian de Doutoramentos em Biologia e Medicina (RA), the US Public Health Service Grant RO1-GM30054 from the National Institutes of Health (MAS), and the US Department of Defense Grant N00014-97-1-0364 from the Office of Naval Research (MAS). We thank Armindo Salvador for critically reading early versions of this manuscript and making useful comments.

References


