Biotechnology and Genetic Engineering Reviews - Vol. 25, 1-40 (2008)

Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways

RUI ALVES[1], ESTER VILAPRINYO[1, 3], BENITO HERNÁNDEZ-BERMEJO[2], ALBERT SORRIBAS[1,*]

[1]Departament de Ciències Mèdiques Bàsiques, Institut de Recerca Biomèdica de Lleida (IRBLLEIDA), Universitat de Lleida, Montserrat Roig, 2, 25008-Lleida, Spain. [2]Universidad Rey Juan Carlos, Madrid, Spain. [3]Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Spain.

Abstract

There is a renewed interest in obtaining a systemic understanding of metabolism, gene expression and signal transduction processes, driven by the recent research focus on Systems Biology. From a biotechnological point of view, such a systemic understanding of how a biological system is designed to work can facilitate the rational manipulation of specific pathways in different cell types to achieve specific goals. Due to the intrinsic complexity of biological systems, mathematical models are a central tool for understanding and predicting the integrative behavior of those systems. Particularly, models are essential for a rational development of biotechnological applications and in understanding system's design from an evolutionary point of view. Mathematical models can be obtained using many different strategies. In each case, their utility will depend upon the properties of the mathematical representation and on the possibility of obtaining meaningful parameters from available data. In practice, there are several issues at stake when one has to decide which mathematical model is more appropriate for the study of a given problem. First, one needs a model that can represent the aspects of the system one wishes to study. Second, one must choose a mathematical

*To whom correspondence may be addressed (albert.sorribas@cmb.udl.cat)

Abbreviations: CPU, Central Processing Unit; S, Stoichiometric Matrix; v, Vector of Fluxes; v_0, Vector of Fluxes at steady state; X, Vector of metabolites; X_0, Vector of metabolites at steady state; SC, Saturating and Cooperative formalism; OP, Operating Point; GMA, Generalized Mass Action model; PL, Power-law formalism; LL, Lin-log and (log)lin model; SS, S-System model

representation that allows an accurate analysis of the system with respect to different aspects of interest (for example, robustness of the system, dynamical behavior, optimization of the system with respect to some production goal, parameter value determination, etc). Third, before choosing between alternative and equally appropriate mathematical representations for the system, one should compare representations with respect to easiness of automation for model set-up, simulation, and analysis of results. Fourth, one should also consider how to facilitate model transference and re-usability by other researchers and for distinct purposes. Finally, one factor that is important for all four aspects is the regularity in the mathematical structure of the equations because it facilitates computational manipulation. This regularity is a mark of kinetic representations based on approximation theory. The use of approximation theory to derive mathematical representations with regular structure for modeling purposes has a long tradition in science. In most applied fields, such as engineering and physics, those approximations are often required to obtain practical solutions to complex problems. In this paper we review some of the more popular mathematical representations that have been derived using approximation theory and are used for modeling in molecular systems biology. We will focus on formalisms that are theoretically supported by the Taylor Theorem. These include the Power-law formalism, the recently proposed (log)linear and Lin-log formalisms as well as some closely related alternatives. We will analyze the similarities and differences between these formalisms, discuss the advantages and limitations of each representation, and provide a tentative "road map" for their potential utilization for different problems.

Introduction: goals of mathematical modeling

Mathematical modeling is an essential tool for Systems Biology. We will briefly discuss some of the more frequent types of biological problems for which modeling is used and the challenges that those problem pose to the modeling process.

For many researchers, a modeling exercise has the fundamental goal of *fitting experimental data to derive parameter values that characterize the processes of interest*. When dealing with such a problem, the mathematical model provides a tool to evaluate if the conceptual description of the system under study is adequate to fit the data that are observed and measured for that system. Such an approach has two types of potential drawbacks. On one hand, a model based on the correct reaction scheme for the system may fail to fit experimental data. There may be two general reasons for such a failure. Either the absence of some unknown regulatory signal in a network structure or the use of inappropriate kinetic representations to write the mathematical model could preclude a conceptual model from explaining the systemic behavior of the network. On the other hand, models that use complex mathematical representations have the flexibility to fit a wider range of dynamic behaviors than models created using simpler mathematical representations. A consequence of this is that, depending on the mathematical representation used for the model, conceptual models that are incorrect can in some cases lead to a "false positive" result. This "false positive" is a model that adequately fits experimental measurements, but that incorrectly represents the mechanism of the processes that underlie the results. A way to decrease the probability of accepting an inappropriate conceptual model is by using well suited alternative mathematical models to predict the behavior of the system under untested conditions. The model's predic-

tions can then be compared to the corresponding wet lab experiments, thus validating which of the alternative conceptual models proposed is better suited for representing the system of interest. It is important to carefully select and design the experiments that can provide the most adequate information for discriminating between alternative models. Nevertheless, one should keep in mind that the ability of a model to fit experimental data is not an ultimate proof that the process that underlies the data generation is clearly understood (Voit, 2002).

Researchers may also be interested in *reconstructing and identifying the topology of reactions and regulation* in biological pathways and circuits. This is an important scientific challenge that requires an integrative use of many different types of tools and information, mined from genomic, proteomic, bibliomic, fluxomic and metabolomic data. Mathematical models can play a central role in this task as they can be used to characterize the dynamic behavior of alternative network structures and compare that behavior to what is observed experimentally (Alves *et al.*, 2004a,b). Mathematical representations used for this type of model building should be easy to manipulate in a systematic and automated way. This is so because alternative processes and metabolites need to be included and excluded in combinatorial ways to generate alternative network topologies.

Molecular Biologists have come to realize the importance of *identifying and analyzing design principles* in gene circuits, metabolic pathways, etc. Mathematical models are fundamental for this research. This type of research has a long tradition in Systems Biology (Savageau, 1972; 1976). The recent field of Synthetic Biology (see Arkin, 2001; Forster and Church, 2007; Greber and Fussenegger, 2007; Luisi, 2007; Meyer *et al.*, 2007; Pleiss, 2006; Saito and Inoue, 2007; Sole *et al.*, 2007 for reviews) heavily depends on the ability to characterize the underlying rules that govern the systemic behavior of a network. These rules are then used to create systems with specific performances. As examples of such applications we have the oscillatory clock designed by Ninfa, Savageau and co-workers in *E. coli* (Atkinson *et al.*, 2003), or the bistable switch designed by Kim et al. (2006), among others (Antunes *et al.*, 2006; Atsumi and Little; 2006, Fung *et al.*, 2005; Greber and Fussenegger, 2007; Haseltine and Arnold, 2007; Kim *et al.*, 2006; Rosenfeld *et al.*, 2007; Sprinzak and Elowitz, 2005; Weber *et al.*, 2007; Yokobayashi *et al.*, 2002). Identifying and analyzing design principles is, perhaps, the application discussed in this review for which mathematical models are more central as a tool (Voit, 2006). For example, one can analyze the response of yeast gene expression patterns to a given stress by experimental means, such as microarray experiments (Alvarez-Vasquez *et al.*, 2005; Sims *et al.*, 2004; Vilaprinyo *et al.*, 2006; Voit, 2003; Voit and Radivoyevitch, 2000). In the response to heat shock, an increase in the concentration of sphyngolipids, trehalose, and chaperones plays an important protective role. Thus, it seems logical that genes involved in the synthesis of these cellular components are over-expressed. However, the question of why a specific increase in gene expression has evolved instead of some other change can hardly be addressed with a wet lab experiment. By using mathematical models one can show that alternative changes in gene expression would hinder specific physiological requirements for the survival of the cell. This type of analysis can ultimately lead to understanding the qualitative and quantitative organization of cellular circuits.

In many biotechnological applications, mathematical models are used for predicting the best way to modify or manipulate a biological circuit in order to *optimize specific*

properties of the system (for example see Marin-Sanguino *et al.,* 2007; Sevilla *et al.*, 2005; Vera *et al.*, 2007). An important goal of metabolic engineering is precisely that of manipulating cell metabolism in order to obtain specific products by capitalizing on modified cell physiology. While such a task has been accomplished in the past using a trial and error strategy, mathematical models can speed-up the process and provide optimal solutions that are hard to find through a less systematic approach (Bailey, 1991; 2001; Bailey *et al.*, 1990; 2002; Prati *et al.*, 2002). Again, the mathematical representation chosen for this type of task must be adequate for optimization purposes. If that representation is too complicated we could face a problem that is similar to the "false positive" cases described above for parameter fitting. Nevertheless, one must also be aware that models built using oversimplified representations (for example linear representations) can fail to capture essential properties of the system that are important, leading to sub-optimal solutions for the optimization problem at hand.

Mathematical models can also be used to understand the relationships between genotype and phenotype (health and disease states, effect of knocking out specific genes, etc.) by creating more or less large scale models of cellular processes. These models can *integrate different levels of the cellular response and create a network that accounts for the dynamic behavior of genes, proteins and metabolites*. In this type of application, one is often faced with highly complex conceptual schemes. Mathematical models that include the entire genome of an organism are becoming common. Most of these models are based on stoichiometric descriptions of the network that allow the use of graph theory to predict certain qualitative phenotypes that may result from modifying genes in the network. For example, these types of model have been used to predict essentiality of genes, by performing genome-wide *in silico* knock-out experiments (Edwards and Palsson, 2000; Edwards *et al.*, 2002). On the other hand, projects like e-cell (Takahashi *et al.*, 2003) aim at producing a more detailed mathematical description of the cell so that one can (semi)quantitatively predict the cellular behavior. Such detailed descriptions are difficult to obtain in most cases. For example, if one is interested in understanding purine metabolism in humans, one is faced with a system that involves different levels of metabolism and for which virtually no experimental data exists (Curto *et al.*, 1997; 1998). Thus, it is often necessary to rely on data obtained for isolated components of the system in order to create a mathematical model that can be use to explore healthy and pathological states known to be dependent on the system being modeled (Boros *et al.*, 2003; Lee *et al.*, 2004; Maher *et al.*, 2003; Orosz *et al.*, 2003; Ramos-Montoya *et al.*, 2006; Rodriguez-Caso *et al.*, 2006; Selivanov *et al.*, 2007; Vera *et al.*, 2007).

The role of approximated kinetic representations in modeling metabolic processes

As stated above, mathematical models provide a tool for investigating the integrated behavior of complex systems. They have been used to do so even before computers were widely available (Chance, 1943). Such mathematical models may grow very quickly in number of variables and parameters when one considers all individual mechanistic steps and species of a molecular system. This is true even for small pathways. The simpler a model is while still being able to predict the behavior of a system, the simpler the analysis of that system will be. Approximate kinetic representations

may be used to build models that reduce the dimensions of the network, providing an appropriate tool for understanding the systemic properties of a network. The use of such formalisms is justified because:

1. In many cases, one cannot find enough information for a detailed description of the mechanisms and for the estimation of individual parameter values. This is especially true while analyzing problems of design principles or while dealing with poorly characterized systems. Thus, instead of arbitrarily choosing a kinetic function to describe each process, one can use an approximated mathematical representation that is theoretically guaranteed to be a good approximation in some range of values. For this, we need mathematical formalisms that can be used to extract meaningful systemic information under such restrictive conditions.

2. The widespread use of rational kinetic functions to reduce the dimensionality of a network, such as the Michaelis-Menten enzyme kinetics description, is accurate only as long as the theoretical conditions that allow the derivation of such functions hold. These functions are usually derived by lumping processes in a time-scale dependent manner and/or assuming large differences between the concentration of the different species and catalysts involved in the reaction. Such conditions are valid for typical *in vitro* reactions catalyzed by enzymes and taking place in a homogenized medium under excess substrate conditions. However, these conditions may break down for example in non-homogeneous environments such as the cellular medium (Anacker and Kopelman 1987; Kopelman 1988; Savageau 1995; Savageau, 1998; Schnell and Turner, 2004). Thus, mathematical representations that depend on the mathematical properties of functions rather than on the physical properties of the processes may have advantages while representing *in vivo* systems.

3. In the process of model simplification, one often lumps different variables and processes together (Curto *et al.*, 1997; 1998). This creates an aggregated process, often like a black box, with a non-traditional and undefined kinetic description. In such a case, the use of approximation theory to derive a description of this process also facilitates the modeling process.

4. Mathematical models based on approximated representations provide a systematic way of building a model from scratch that can be easily automated. This is especially useful for large systems, when exploring unknown network structures, and in generating models automatically.

5. Optimization problems using nonlinear models are a difficult task. Models based on regular approximate representations of the different processes (i.e. structured models) can facilitate this task because optimization methods that take advantage of the mathematical structure can be developed. The caveat here is that the selected mathematical representation must capture the essential properties of the problem.

6. Approximate representations also have the advantage of allowing models to account for ill-characterized regulatory interactions. This is usually done by using qualitative and semi-quantitative information to estimate ranges of parameter values for the relevant interactions.

Analysis of systemic behavior

As stated above, mathematical models are defined and used to study different types of problems. Independently of the purpose of the study, important considerations while creating a model are accuracy of the predictions, the possibility of analytical solutions, and the easiness of model implementation and analysis through the use of computational tools. It is worth it to briefly review some of the main concepts that are important for model analysis:

1. *Steady-state solution:* The steady-state is a situation in which the net flux through all pools of metabolites in a system is 0. This is a dynamic equilibrium that characterizes, at least ideally, the basal working conditions of many pathways or the homeostatic, long term, response of an organism. A steady-state solution for a model is a vector of metabolite concentrations that correspond to this state of zero net flux. This solution can be obtained from the equations of a model, either numerically (if one attributes values to the parameters and uses numeric algorithms to solve the equations) or analytically (if one solves the equations in closed form, obtaining a solution that is independent of parameter values). Finding analytical solutions is often impossible because, in general, the model of interest is a non-linear system of equations for which such a solution cannot be calculated. Analytical solutions are important to understand general systemic design principles and properties that are independent of parameter values. If such solutions are required, one can use the S-system representation within the Power-law formalism. The GMA representation using the (log)linear or the Lin-log formalism has the same solution than the corresponding S-system Power-law representation (Sorribas *et al.*, 2007). When no analytical solution is available, the steady-state values can be calculated using numerical methods.

2. *Steady-state stability:* Stability is an important property of a steady-state that measures the capacity of the system for returning to a steady-state after a perturbation. Mathematical conditions that ensure stability of the steady-state are design principles for those metabolic systems where instability of the steady-state would be incompatible with biological function (Savageau, 1971a,b, 1974, 1976). In general, local stability of the steady-state is analyzed through linearization of the steady-state equations. Steady-state stability can be studied numerically in any model that displays steady-state behaviour.

3. *Steady-state parameter sensitivity (systemic robustness):* Parameter sensitivity is both a measure of biological adequacy of a network and a tool for model validation. Parameter sensitivity measures the effect that an infinitesimal change in the value of a parameter has on the steady-state values of the model. High parameter sensitivities in a model are often indicative that some parts of the system have not been adequately described in the conceptual network. On the other hand, low parameter sensitivity (*robustness*) is expected for most biological systems. A consequence of this low sensitivity is that systemic properties do not critically depend of small variations in the parameter values. High robustness (low sensitivity) is an important design principle (Savageau, 1971b) that is emerging as a central concern in Systems Biology (Kitano, H., 2007). Sensitivity

analysis has also been used to improve and update pre-existing models (Curto *et al.*, 1997; Ni and Savageau, 1996; Shiraishi and Savageau, 1992a,b,c.

4. *Dynamic systemic behaviour:* Numerical simulation is a common procedure for exploring the dynamic systemic behaviour of a model. These simulations are done by attributing values to the parameters and initial concentrations of the model followed by solving the differential equations using numerical algorithms. Most of these algorithms solve the equations, independently of the form of those equations. Nevertheless, as numerical integration of differential equations is a costly procedure in terms of computer time, it is desirable to develop algorithms that speed-up the process. Often, by taking advantage of the regular structure of approximated formalisms, one can develop numerical algorithms that are very efficient in their CPU time usage. Such a method, based on Taylor series expansions of the solution, has been developed for the Power-law formalism (Irvine and Savageau, 1990). Given that it takes advantage of regularities in the calculations of the Taylor series, it is likely that this method could be extended and adapted for other Taylor formalisms. Nevertheless, there is the concern that practical problems may appear for the Taylor method when the systems moves towards low metabolite values in representations derived for the (log)linear and Lin-log formalisms. In such situations, (log)linear and Lin-log models can produce negative values for the velocities (Sorribas *et al.*, 2007). Having negative values for a given flux is a physical and biological impossibility. In addition, such negative values are a problem for any numerical integration based on the Taylor method.

An overview of mathematical formalisms based on stoichiometry and on approximate representations

STOICHIOMETRIC MODELS

One of the simplest mathematical representations that can be used to model a network is derived from the conceptual graph that represents the reactions of the network. The information from the graph can be condensed into the stoichiometric matrix $\mathbf{S}$, in which each element S_{ij} correspond to the stoichiometric coefficient of the metabolite *i* in reaction *j*. Additionally, each reaction *j* is represented by a flux v_j, creating a network flux vector **v**. A dynamical model of a given system can be written as:

$$\dot{\mathbf{X}} = \mathbf{S} \cdot \mathbf{v}, \tag{1}$$

in which $\dot{\mathbf{X}}$ is a vector of derivatives dX_i/dt. At steady-state, the system obeys the equation

$$\mathbf{S} \cdot \mathbf{v}_0 = \mathbf{0}, \tag{2}$$

where v_0 in now the vector of steady-state fluxes. Such a simple representation for the steady-state of the system is known as a stoichiometric model and it can be easily obtained from existing information on the biochemical reactions of the network. This

representation allows certain types of analysis that can shed some light on the effect that a change in a given process may have on flux distribution. The use of Eq. (2) as a description for the flux balance in steady-state is appealing as it overcomes one of the major drawbacks of more complicated models, i.e. the lack of detailed information on kinetic mechanisms and parameters.

KINETIC MODELS BASED ON APPROXIMATED REPRESENTATIONS

The choice of a particular mathematical form for the velocities **v** in Eq. (1) generates a kinetic model that can be used for simulating the dynamical behaviour of the system and for the analysis of the changes in fluxes and concentration of metabolites. Such an analysis is unavoidable if one wants to understand systemic response under different external conditions. Furthermore, these models play a central role if one is interested in investigating network structure and design principles. As a common initial choice, one often considers using traditional enzymatic rate-laws to describe the dynamic behaviour of each of the individual processes v_i present in vector **v**. However, as discussed above, the lack of information about the precise mechanism of each of these processes is a serious limitation for obtaining such a detailed representation. This limitation becomes more dramatic as the models grow and integrate more and more individual processes (Curto *et al.*, 1998; Shiraishi and Savageau, 1992a,b,c; Sims *et al.*, 2004; Voit *et al.,* 2006; Voit, 2002).

An alternative to traditional enzyme kinetics is the use of approximated kinetic representations. Such representations provide all-purpose functional forms that greatly facilitate the modelling process. To obtain such a representation, we consider that each velocity can be expressed as

$$v_i = \Psi(E_i, \mathbf{X}, \theta) \tag{3}$$

where Ψ is a nonlinear function, E_i is the enzyme catalyzing the reaction, $\mathbf{X}$ is a vector of metabolites, effectors, etc., and θ is a vector of parameters. A number of alternative approximated representations of v_i can be derived using mathematical approximation techniques, such as the Taylor series representation of a function (see *Table 1*). The first non-linear, approximated, representation of v_i is the Power-law formalism, which was originally presented in 1969 (Savageau, 1969a,b; 1970). The formalism is a consequence of approximating a function in logarithmic space using a first order Taylor series, followed by a return to Cartesian space. Use of the Power-law formalism has facilitated the development of a very complete set of analytical methods, leading to the framework know as Biochemical Systems Theory (Savageau, 1976; Voit, 2000). The Power-law representation of v_i is given by:

$$v_i = \gamma_i \prod_{j=1}^{n+m} X_j^{f_{ij}^0} \tag{4}$$

where, n indicates the number of dependent metabolites (i.e. those metabolites which are considered as internal metabolites of the systems and whose concentration and dynamics depend on the systemic behaviour), m indicates the number of external metabolites (fixed outside of the system), and f_{ij}^0 (kinetic-order) is the local sensibility

Table 1. Comparison between alternative approximate kinetic representations for a one substrate process. All formalisms are derived by using a Taylor series approximation.

Approximation	Change in enzyme activity	No change in enzyme activity
Linear	$\frac{v_i}{v_{i0}} \cong 1 + f_{iE}\left(\frac{E_i}{E_{i0}} - 1\right) + \sum_{j=1}^{m} f_{ij}\left(\frac{X_j}{X_{j0}} - 1\right)$	$\frac{v_i}{v_{i0}} \cong 1 + \sum_{j=1}^{m} f_{ij}\left(\frac{X_j}{X_{j0}} - 1\right)$
(log)linear	$\frac{v_i}{v_{i0}} \cong 1 + f_{iE} \ln\left(\frac{E_i}{E_{i0}}\right) + \sum_{j=1}^{m} f_{ij} \ln\left(\frac{X_j}{X_{j0}}\right)$	$\frac{v_i}{v_{i0}} \cong 1 + \sum_{j=1}^{m} f_{ij} \ln\left(\frac{X_j}{X_{j0}}\right)$
Linear approximation of $\Psi(x)$	$\frac{v_i}{v_{i0}} = \frac{E_i}{E_{i0}} \times \left\{1 + \sum_{j=1}^{m}\left(f_{ij}\left(\frac{X_j}{X_{j0}} - 1\right)\right)\right\}$	$\frac{v_i}{v_{i0}} = 1 + \sum_{j=1}^{m} f_{ij}\left(\frac{X_j}{X_{j0}} - 1\right)$
Lin-log	$\frac{v_i}{v_{i0}} = \frac{E_i}{E_{i0}} \times \left\{1 + \sum_{j=1}^{m} f_{ij} \ln\left(\frac{X_j}{X_{j0}}\right)\right\}$	$\frac{v_i}{v_{i0}} \cong 1 + \sum_{j=1}^{m} f_{ij} \ln\left(\frac{X_j}{X_{j0}}\right)$
Power-law	$\ln\left(\frac{v_i}{v_{i0}}\right) \cong f_{iE} \ln\left(\frac{E_i}{E_{i0}}\right) + \sum_{j=1}^{m} f_{ij} \ln\left(\frac{X_j}{X_{j0}}\right)$	$\ln\left(\frac{v_i}{v_{i0}}\right) \cong \sum_{j=1}^{m} f_{ij} \ln\left(\frac{X_j}{X_{j0}}\right)$

of v_i to changes in X_j and is defined as

$$f_{ij}^0 = \left(\frac{\partial v_i}{\partial X_j} \right)_0 \frac{X_{j0}}{v_{i0}} \tag{5}$$

In this definition, the index 0 indicates evaluation at a given operating point, defined by a set of values of the metabolites and the corresponding flux values: $(\mathbf{X_0}, \mathbf{v_0})$. Finally, the parameter γ_i (an apparent *rate-constant*) can be calculated from

$$\gamma_i = v_{i0} \prod_{j=1}^{n+m} X_{j0}^{-f_{ij}^0} \tag{6}$$

For technical details on the derivation of this formalism see for instance Voit (2000), Sorribas et al. (2007), and references therein.

The interpretation of kinetic-orders is of special relevance for some of the questions one is usually interested in. According to (5), a kinetic-order f_{ij}^0 will be positive if an increase in X_j leads to an increase in v_i. This will be the case for the substrates of a reaction or for any positive effector of that reaction. A kinetic-order will be negative for any species that inhibits the rate v_i. If a metabolite j has no effect on v_i, then the kinetic-order f_{ij}^0 will be zero. Furthermore, it is often possible to derive approximate numerical values for the kinetic orders, based on qualitative information about the operating point of the approximation. For example, if a reaction is close-to-saturation with respect to a metabolite, then this metabolite will have a kinetic-order with a value that is close to zero. Values of 1 are appropriate for kinetic orders of metabolites while approximating classical enzyme kinetic functions at an operating point well below the K_m for the relevant metabolite. Values higher than 1 for kinetic orders are only possible for cooperative processes, again well below the K_m. If an operating point is near the K_m, the value for the kinetic order is approximately 0.5 for a Michaelis-Menten process.

Different authors have derived alternative representations that are also based on a Taylor series approximation. The (log)linear formalism was introduced by Bailey and Hatzimanikatis (Hatzimanikatis, 1999; Hatzimanikatis and Bailey, 1996; Hatzimanikatis *et al.*, 2004; 2005). In this formalism, which is based on a linear representation of the target function (3), the functional form of the approximation is:

$$\frac{v_i}{v_{i0}} \cong 1 + f_{iE}^0 \ln\left(\frac{E_i}{E_{i0}} \right) + \sum_{j=1}^{m} f_{ij}^0 \ln\left(\frac{X_j}{X_{j0}} \right) \tag{7}$$

In this equation, f_{iE}^0 is the apparent kinetic-order with respect to the enzyme E_i. Its value will, in general, be 1, as velocities are linear with respect to the enzyme concentration. In a similar way, Heijnen and coworkers derived the Lin-log formalism, which has the following form (Heijnen, 2005; Heijnen *et al.*, 2004; Kresnowati *et al.*, 2005; Visser *et al.*, 2004; Wu *et al.*, 2004):

$$\frac{v_i}{v_{i0}} = \frac{E_i}{E_{i0}} \times \left\{ 1 + \sum_{j=1}^{m} f_{ij}^0 \ln\left(\frac{X_j}{X_{j0}} \right) \right\} \tag{8}$$

It is noteworthy that when the concentration of enzymes is constant the (log)linear and log-lin representations reduce to the same expression:

$$\frac{v_i}{v_{i0}} = 1 + \sum_{j=1}^{m} f_{ij}^0 \ln\left(\frac{X_j}{X_{j0}} \right) \rightarrow v_i = v_{i0} + v_{i0} \sum_{j=1}^{m} f_{ij}^0 \ln\left(\frac{X_j}{X_{j0}} \right)$$

Finally, the *Saturable and Cooperativity* (SC) representation (Sorribas *et al.*, 2007) is also an alternative approximated representation for kinetic functions, based on Taylor series. This representation is derived through a procedure that is analogous to that used to derive the Power-law formalism. As stated above, the Power-law representation is based on using a linear Taylor series approximation in log-log space. Making the approximation in logarithmic space increases the range of accuracy of the approximation. This has motivated the search for alternative representations that could combine different, non-Cartesian, spaces in order to further increase this range of accuracy. The formalism is developed from a transformation of coordinates of the form $(w,z) = (v^{-l}, X^c)$. In this coordinate system, c is a nonzero constant that is to be defined for every X_k. Following arguments analogous to those used by Savageau in his derivation of the log-log transformation (Savageau, 1969), the final *Saturable and Cooperativity* (SC) representation is given by

$$v_i = \frac{V_i \prod_{j=1}^{n+m} X_j^{n_{ij}}}{\prod_{j=1}^{n+m} \left(K_{ij} + X_j^{n_{ij}} \right)} \tag{9}$$

where

$$n_{ij} = \frac{f_{ij}}{\left(1 - p_{ij}\right)}$$
$$K_{ij} = \frac{\left(1 - p_{ij}\right)}{p_{ij}} X_{j0}^{n_{ij}} \tag{10}$$

In this representation, p_{ij} is the saturation fraction of flux v_i by substrate or modifier X_j. This is given by the ratio between the rate at the operating point, v_{i0}, and the rate when $X_j \rightarrow \infty$ and all the other metabolites are kept at their operating point values. If X_j is an inhibitor, i.e. $f_{ij} < 0$, then the saturation fraction is given by the ration between the rate at the operating point and the rate when $X_j \rightarrow 0$ (see Sorribas *et al.*, 2007 for further details and examples).

All the formalisms described above have some common features as they are *local approximations* of a function at a given *operating point*. In all cases, the resulting

representation uses, to some extent, a Taylor series approximation and, consequently, has a limited accuracy range. Interestingly, all approximations share the following common parameters: (1) Operating flux, v_{i0}; (2) Operating values for each metabolite, effectors and enzymes, X_{i0}; and (3) Local sensitivity (elasticity), f_{ij}^{0}, at the operating point[1]. The SC formalism introduces an additional parameter, p_{ij}, that accounts for the fraction of saturation of the rate at the operating point.

As a consequence, parameterization of a model in all these formalisms requires the same information about how the rate of a process depends on metabolite concentration about the operating point, although the final representation varies. In addition, the SC formalism requires additional information about the fraction of saturation of the fluxes at the operating point, with respect to the different metabolites that modulate the flux.

BUILDING MATHEMATICAL MODELS USING APPROXIMATED REPRESENTATIONS

Independently of the formalism one is using, the first step in defining a mathematical model for a system with n variables and p processes is to write the node equations:

$$\frac{dX_i}{dt} = \dot{X}_i = \sum_{r=1}^{p} \mu_{ir} v_r \qquad i = 1,...,n \tag{11}$$

where μ_{ir} is a stoichiometric factor, i.e. $\mu_{ir} = 1$ if v_r is producing a molecule of X_i, $\mu_{ir} = -1$ if v_r is degrading a molecule of X_i, and so on and so forth. A specific and dynamic mathematical model is obtained when each v_r is written using a defined functional form (either using a traditional enzyme kinetics rational expression or any of the approximations discussed above). Using the Power-law formalism, we would obtain

$$\frac{dX_i}{dt} = \dot{X}_i = \sum_{r=1}^{p} \left(\mu_{ir} \gamma_r \prod_{j=1}^{n+m} X_j^{f_{rj}^{0}} \right) \qquad i = 1,...,n \tag{12}$$

This particular representation is known as the *Generalized Mass Action* (GMA) representation within the Power-law framework and within Biochemical Systems Theory (Voit, 2000). The equivalent representation using the (log)linear formalism (Hatzimanikatis, 1999; Hatzimanikatis and Bailey, 1996; Hatzimanikatis *et al.*, 2004; 2005) would be

$$\dot{X}_i \cong \sum_{r=1}^{p} \left(\mu_{ir} \left(v_{r0} + v_{r0} f_{rE}^{0} \ln\left(\frac{E_r}{E_{r0}} \right) + v_{r0} \sum_{j=1}^{m} f_{rj}^{0} \ln\left(\frac{X_j}{X_{j0}} \right) \right) \right) \tag{13}$$

[1] Kinetic-orders f_{ij}^{0} as defined in equation (5) are formally equivalent to elasticities ε_{ij}^{0} used in Metabolic Control Analysis and related techniques. We shall maintain the kinetic-order nomenclature to relate the different formalisms.

Eqs. 12 and 13 appear quite different because the operating point concentrations and fluxes appear explicitly in Eq. 13. As discussed in the previous section, these fluxes and concentrations are implicit in the rate-constant of the Power-law formalism (Eq. 12). The (log)linear representation (Eq. 13) can be rewritten as

$$\dot{X}_i \cong \sum_{r=1}^{p}\left(\mu_{ir} v_{r0}\left(1 - f_{rE}^{0} \ln\left(E_{r0}\right) - \sum_{j=1}^{m} f_{rj}^{0} \ln\left(X_{jo}\right) + f_{rE}^{0} \ln\left(E_r\right) - \sum_{j=1}^{m} f_{rj}^{0} \ln\left(X_j\right)\right)\right) \tag{14}$$

Eq. 14 can then be rewritten in a form that makes all the steady-state values implicit to the appropriate parameters, i.e.

$$\dot{X}_i \cong \sum_{r=1}^{p}\left(\mu_{ir} v_{r0}\left(\varphi_r + f_{rE}^{0} \ln\left(E_r\right) - \sum_{j=1}^{m} f_{rj}^{0} \ln\left(X_j\right)\right)\right) \tag{15}$$

where, $\varphi_r = 1 - f_{rE}^{0} \ln\left(E_{r0}\right) - \sum_{j=1}^{m} f_{rj}^{0} \ln\left(X_{jo}\right)$. It is important to note that the same information is required to parameterize Eq. 12 and Eqs. 13-15. Following similar procedures, we can derive the mathematical representation for the Lin-log formalism and for the SC formalism. In fact one can derive a mathematical representation for any other formalism derived from approximating a function through a first order Taylor series in a space that has a one-to-one correspondence to Cartesian space.

Equations for the model can be derived using different strategies for flux and pool aggregation (Sorribas and Savageau, 1989a,b,c). All formalisms discussed here are amenable to these different aggregation strategies. GMA-like representations emphasize individual processes. Alternatively, the basic differential equation for a model can be alternatively written as aggregated node equations

$$\frac{dX_i}{dt} = \dot{X}_i = \sum_{r=1}^{p} \mu_{ir}^{+} v_r - \sum_{r=1}^{p} \mu_{ir}^{-} v_r = V_i^{+} - V_i^{-} \qquad i = 1,\ldots,n \tag{16}$$

In this alternative representation, different processes that contribute for the synthesis of a metabolite are aggregated into a single function and the different processes that contribute for the degradation of that metabolite are aggregated into another single function. In Eq. 16, μ_{ir}^{+} and μ_{ir}^{-} account for the positive and negative stoichiometric factors, i.e. synthesis and degradation terms[2]. Then, each aggregated term is represented using whatever formalism we decide to use. In the case of the Power-law formalism, this strategy leads to the following representation

$$\frac{dX_i}{dt} = \dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - ß_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \qquad i = 1,\ldots,n \tag{17}$$

[2] It is possible to define different aggregation strategies that would result in a different S-system. Independently of the strategy, the original stoichiometric matrix would be divided into two matrices.

This particular representation is known as an S-system (Savageau, 1969a,b; Savageau, 1970; Voit, 2000)[3]. It has a number of interesting properties, such as the possibility of calculating analytical steady-state solutions, which makes it advantageous to choose this representation for different modelling applications. A S-system-like description can be derived within the (log)linear, the Lin-log, or the SC formalism. However, in the later formalism, some of the advantages related to the Power-law representation are lost and no analytical steady-state can be calculated (*Table 2*).

A COMPARATIVE EXAMPLE OF PATHWAYS MODELLING USING DIFFERENT FORMALISMS

Given a conceptual scheme for a system, it is straightforward to obtain a mathematical model for this scheme using any of the mathematical representations discussed in this review. As an illustrative example, we will compare the alternative mathematical models for the system in *Figure* 1. This instructive example represents a branched pathway with two regulatory interactions. One is a positive 'feedforward' and the other is a negative feedback. A stoichiometric description of the dynamic behavior of the system shown in *Figure 1* is given by

$$\frac{dx}{dt} = \mathbf{S} \cdot \mathbf{v} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} \tag{18}$$

Qualitative aspects of the dynamic behavior of a system can sometimes be predicted by analyzing the stoichiometric matrix **S** from Eq. 18. However, the results of such an analysis would be the same for any pathway or circuit with the same set of reactions, independent of any differences in regulation between systems. These regulatory differences may translate into distinct dynamic behavior, throwing off the predictions from the stoichiometric matrix analysis. Techniques such as Boolean network analysis or Feinberg *et al.*'s reaction network theory can be used for this type of analysis (Craciun and Feinberg, 2005, 2006; Craciun *et al.*, 2006, Demongeot *et al.*, 2000; Feinberg, 1985; 1987; 1988; Kaufman and Thomas, 2003; Thomas, 1973; Thomas *et al.*, 1995).

[3] In the S-system representation, α_i rate constants and g_{ij} kinetic orders are used for synthesis while β_i rate constants and h_{ij} kinetic orders are used for degradation.

Table 2. Properties of the different mathematical formalisms

Theoretical Support	Formalisms	Representation	Simplification of model	Closed form steady-state solution	Multiple steady-states possible	Closed form dynamic solution	Steady-state accuracy	Time curve accuracy
		Generalized mass action	Y	N	Y	N	100% (OP)	Variable
Based on Taylor's Theorem	Power-law	S-System	Y	Y	N	N	100% (OP)	Variable
		Half-System	Y	N	N	N		
	Lin-log	Lin-log	Y	Y	N	Y	100% (OP)	Variable
	(log)linear	(log)linear	Y	Y	N	N	100% (OP)	Variable
	Linear	Linear	Y	Y	N	Y	100% (OP)	Variable
	Linear approximation of ψ	Linear approximation of ψ	Y	Y	N	Y	100% (OP)	Variable
	Saturable	Saturable	Y	N	Y	N	100% (OP)	Variable
Based on time scale simplifications	Michaelis Menten-like kinetics	MMLK	Y	N	Y	N	Variable	Variable
Based on statistical mechanics	Mass action	Mass action	N	N	Y	N	100%	100%

100% (OP) The steady-state is exact only at the operating point

Table 2. C'nued

Theoretical support	Formalisms	Representation	Accurate logarithmic gains	Accurate sensitivities	Suitable for classical mathematically controlled comparison	Suitable for statistical mathematically controlled comparison
		Generalized Mass Action	100% (OP)	100% (OP)	N	Y
Based on Taylor's Theorem	Power-law	S-System	100% (OP)	100% (OP)	Y	Y
		Half-System	100% (OP)	100% (OP)	N	N
	Lin-log	Lin-log	100% (OP)	100% (OP)	Y	Y
	(log)linear	(log)linear	100% (OP)	100% (OP)	Y	Y
	Linear	Linear	100% (OP)	100% (OP)	Y	Y
	Linear approximation of ψ	Linear approximation of ψ	100% (OP)	100% (OP)	Y	Y
	Saturable	Saturable	100% (OP)	100% (OP)	N	Y
Based on time scale simplifications	Michaelis Menten-like kinetics	MMLK	Variable	Variable	N	Y
Based on statistical mechanics	Mass action	Mass action	100%	100%	N	Y

100% (OP) The steady-state is exact only at the operating point

Theoretical support	Formalisms	Aggregated models	Individual fluxes models*
Based on Taylor's Theorem	Power-law	S-System Half System	GMA
	Lin-log	ND	Y
	(log)linear	ND	Y
	Linear	ND	Y
	Linear approximation of ψ	ND	Y
	Saturable	Y	Y
Based on time scale simplifications	Michaelis Menten-like kinetics	ND	Y
Based on statistical mechanics	Mass action	ND	Y

*This refers to the possibility of creating models where the flux of each individual process is approximated by a function, rather than the flux going through a pool of metabolites.

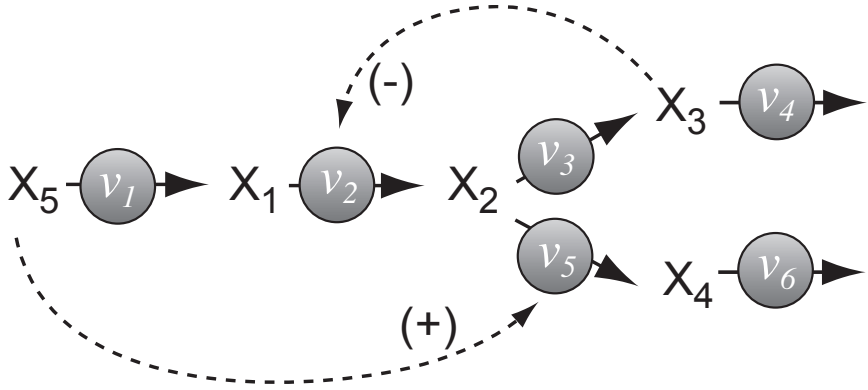

Figure 1. A branched pathway with positive feedforward and negative feedback regulation. This pathway is used to illustrate the implications of modeling a system by different mathematical formalisms. Alternative mathematical models are created for this system using different formalisms (*Table 3*). The performance of the alternatives is evaluated by comparing the behavior predicted by simulation to that of the original system, when that system moves away of the operating point.

A way to include regulatory information into a stoichiometric model is by defining the vector of rate expressions **v** and accounting regulatory factors in the individual rate expressions**.** A GMA Power-law model incorporates such regulatory effects through the kinetic-orders, which can be represented by the following matrices[4]:

$$F_D = \begin{pmatrix} 0 & 0 & 0 & 0 \\ f_{21} & 0 & f_{23} & 0 \\ 0 & f_{32} & 0 & 0 \\ 0 & 0 & f_{43} & 0 \\ 0 & f_{53} & 0 & 0 \\ 0 & 0 & 0 & f_{64} \end{pmatrix} \qquad F_I = \begin{pmatrix} f_{15} \\ 0 \\ 0 \\ 0 \\ f_{55} \\ 0 \end{pmatrix}$$

These regulatory parameters are incorporated into the rates, generating the following model

$$\dot{X}_1 = v_1 - v_2 = \gamma_1 X_5^{f_{15}} - \gamma_2 X_1^{f_{21}} X_3^{f_{23}}$$
$$\dot{X}_2 = v_2 - v_3 - v_5 = \gamma_2 X_1^{f_{21}} X_3^{f_{23}} - \gamma_3 X_2^{f_{32}} - \gamma_5 X_2^{f_{52}}$$
$$\dot{X}_3 = v_3 - v_4 = \gamma_3 X_2^{f_{32}} - \gamma_4 X_3^{f_{43}}$$
$$\dot{X}_4 = v_5 - v_6 = \gamma_5 X_2^{f_{52}} - \gamma_6 X_4^{f_{64}}$$

Steady-state values and dynamic changes can now be computed from these equations after setting numerical values for the different parameters. We could choose any of the alternative formalisms to obtain the corresponding model. Assuming that enzyme concentration is constant, the (log)linear and Lin-log models for *Figure 1* are:

[4]For clarity, in this example we omit the superscript 0 in the kinetic-orders.

$$\dot{X}_1 = v_{10} + v_{10}f_{15}\ln\left(\frac{X_5}{X_{50}}\right) - v_{20} - v_{20}f_{21}\ln\left(\frac{X_1}{X_{10}}\right) - v_{20}f_{23}\ln\left(\frac{X_3}{X_{30}}\right)$$

$$\dot{X}_2 = v_{20} + v_{20}f_{21}\ln\left(\frac{X_1}{X_{10}}\right) + v_{20}f_{23}\ln\left(\frac{X_3}{X_{30}}\right) - v_{30} - v_{30}f_{32}\ln\left(\frac{X_2}{X_{20}}\right) - v_{50} - v_{50}f_{52}\ln\left(\frac{X_5}{X_{50}}\right)$$

$$\dot{X}_3 = v_{30} + v_{30}f_{32}\ln\left(\frac{X_2}{X_{20}}\right) - v_{40} - v_{40}f_{43}\ln\left(\frac{X_3}{X_{30}}\right)$$

$$\dot{X}_4 = v_{50} + v_{50}f_{52}\ln\left(\frac{X_5}{X_{50}}\right) - v_{60} - v_{60}f_{64}\ln\left(\frac{X_4}{X_{40}}\right)$$

Because steady state fluxes of production and consumption of any metabolite are balanced, the previous equations simplify to[5]

$$\dot{X}_1 = v_{10}f_{15}\ln\left(\frac{X_5}{X_{50}}\right) - v_{20}f_{21}\ln\left(\frac{X_1}{X_{10}}\right) - v_{20}f_{23}\ln\left(\frac{X_3}{X_{30}}\right)$$

$$\dot{X}_2 = v_{20}f_{21}\ln\left(\frac{X_1}{X_{10}}\right) + v_{20}f_{23}\ln\left(\frac{X_3}{X_{30}}\right) - v_{30}f_{32}\ln\left(\frac{X_2}{X_{20}}\right) - v_{50}f_{52}\ln\left(\frac{X_5}{X_{50}}\right)$$

$$\dot{X}_3 = v_{30}f_{32}\ln\left(\frac{X_2}{X_{20}}\right) - v_{40}f_{43}\ln\left(\frac{X_3}{X_{30}}\right)$$

$$\dot{X}_4 = v_{50}f_{52}\ln\left(\frac{X_5}{X_{50}}\right) - v_{60}f_{64}\ln\left(\frac{X_4}{X_{40}}\right)$$

Table 3 shows the reference mathematical model for the pathway in *Figure 1*, as well as all alternative mathematical models for that system, based on the formalisms described in *Table 1*. The nominal steady-state that was used as an operating point to calculate parameter values is that of the reference mathematical model when the value for the independent variable X_5 is 0.4. As expected, at this point all the formalisms predict the same steady-state concentrations and velocities.

The dynamic behavior of the alternative models differs when the concentration of X_5 moves away from the operating point. As an example, consider two different conditions, one below ($X_5 = 0.1$) and one above ($X_5 = 1$) the operating point ($X_5 = 0.4$) (*Table 4, Figure 2*). As a result of these changes in X_5, the system will move from the basal state to a different steady-state. The dynamic changes in the concentration of the corresponding variables when they move from one steady-state to another are predicted differently by each model. Remarkable drawbacks for the different representations of this particular system are: (1) Lin-log and (log)linear formalisms velocities become negative when we decrease X_5. (2) When increasing the independent variable from 0.4

[5]This final form of the (log)linear and lin-log models is formally equivalent to a linearized Power-law model around the steady-state.

Table 3. A branched pathway with positive feedforward and negative feedback regulation. The actual formalism is approximated by different mathematical formalisms, all of them equivalent at the operating point. Abbreviations: SC for Saturable and Cooperative; PL for Power-Law; GMA for Generalized Mass Action; SS for S-System; and LL for Lin-log and (log)linear formalisms. *For S-System an additional velocity is considered, the aggregated flux for degradation of X_2.

	Original	PL	SC	LL
v_1	$(16\ X_5)/(1+X_5)$	$8.8\ X_5^{0.714}$	$(16\ X_5)/(1+X_5)$	$4.57\ (1+0.714\ \ln[2.5\ X_5])$
v_2	$(217\ X_1\ (1+0.133\ X_3))/(1+8.33\ X_3+3.33\ X_1\ (1+3.33\ X_3))$	$(14.4\ X_1^{0.788})/X_3^{0.734}$	$(109\ X_1)/((0.613+X_1)\ (5.7+1/X_3^{0.916})\ X_3^{0.916})$	$4.57\ (1+0.788\ \ln[6.06\ X_1]-0.734\ \ln[1.46\ X_3])$
v_3	$(5\ X_2)/(1+X_2)$	$2.62\ X_2^{0.349}$	$(5\ X_2)/(1+X_2)$	$3.26\ (1+0.349\ \ln[0.535\ X_2])$
v_4	$(8\ X3)/(1+X_3)$	$4.07\ X_3^{0.593}$	$(8\ X_3)/(1+X_3)$	$3.26\ (1+0.593\ \ln[1.46\ X_3])$
v_5	$(58\ X_2^3)/(1331\ (X_2^3/1331+(1+X_5^3)/(1+63\ X_5^3)))$	$1.54\ X_2^{2.93}\ X_5^{2.17}$	$(604\ X_2\ X_5^{2.4})/((281.+X_2)\ (1.04+X_5^{2.4}))$	$1.31\ (1+2.93\ \ln[0.535\ X_2]+2.17\ \ln[2.5\ X_5])$
v_6	$(8\ X_4)/(1+X_4)$	$5.12\ X_4^{0.836}$	$(8\ X_4)/(1+X_4)$	$1.31\ (1+0.836\ \ln[5.09\ X_4])$
V_2		$4.1\ X_2^{1.09}\ X_5^{0.625}$		

to 1 Lin-log, (log)linear, and S-System predict a raise in X_2, X_3, v_3, and v_4 while in the reference model the value for all these variables decreases. (3) As we move away from the operating point, the S-System aggregated fluxes differ from the sum of the individual velocities that compose it. In that specific example, the SC formalism has a wider range of accuracy about the operating point, predicting a dynamic behavior that closely follows the behavior of the reference model (*Figure 3*).

Table 4. Predictions of the new steady-state concentrations and velocities.

Approx.	X_5	X_1	X_2	X_3	X_4	v_1 and v_2	v_3 and v_4	v_5 and v_6	*V_2
OP	0.4	0.16	1.87	0.69	0.20	4.57	3.26	1.31	4.57
Original	0.1	0.02	0.41	0.22	4 E-4	1.45	1.45	3 E-3	
SC	0.1	0.02	0.41	0.22	7 E-5	1.45	1.45	6 E-4	
PL(GMA)	0.1	0.02	0.29	0.23	8 E-6	1.70	1.70	3 E-4	
PL(SS)	0.1	0.04	1.67	0.64	4 E-3	1.70	3.13	0.05	1.70
LL	0.1	0.04	1.67	0.64	4 E-3	0.04	3.13	-3.08	
Original	1	0.32	1.57	0.62	1.62	8.	3.06	4.94	
SC	1	0.34	1.68	0.64	1.55	8.	3.13	4.87	
PL(GMA)	1	0.34	1.57	0.62	1.15	8.80	3.06	5.73	
PL(SS)	1	0.39	2.01	0.72	2.77	8.80	3.34	12.	8.80
LL	1	0.39	2.01	0.72	2.77	7.56	3.34	4.22	

Abbreviations: OP for Operating point; SC for Saturable and Cooperative; PL for Power Law; GMA for Generalized Mass Action; SS for S-System; and LL for Lin-log and (log) linear formalisms. *For S-System an additional velocity is considered, the aggregated flux of the degradation of X2, that at the OP it is the sum of v_3 and v_4.

In a similar exercise, Heinjen suggested that the Lin-log formalism is the more accurate formalism to be used in all cases (Heijnen, 2005). Our results show clearly that this is not generally true and that many different issues are at stake when using the different formalisms. One can always build conceptual systems for which a particular formalism will be the most accurate. However, making the general statement that a given formalism is in general the most accurate requires an exhaustive analysis of different systems that, to our knowledge, no one has yet made. In the few examples we have tested, the SC formalism is among the ones with a bigger range of accuracy, especially when reactions take place in a dimensionally restricted space or when saturable and cooperative mechanisms are important for the dynamical behavior of the system.

Mathematical models at work: some examples on the utility of models based on approximated representations

The choice of the formalism to be used to create a model for the analysis of a given problem depends on the available data, on the problem, and on personal preference and training. We shall now discuss guidelines and relevant criteria for selecting the types of representation that are more adequate in each specific type of problem. For simplicity, we shall discuss the relevant criteria to be used in choosing between

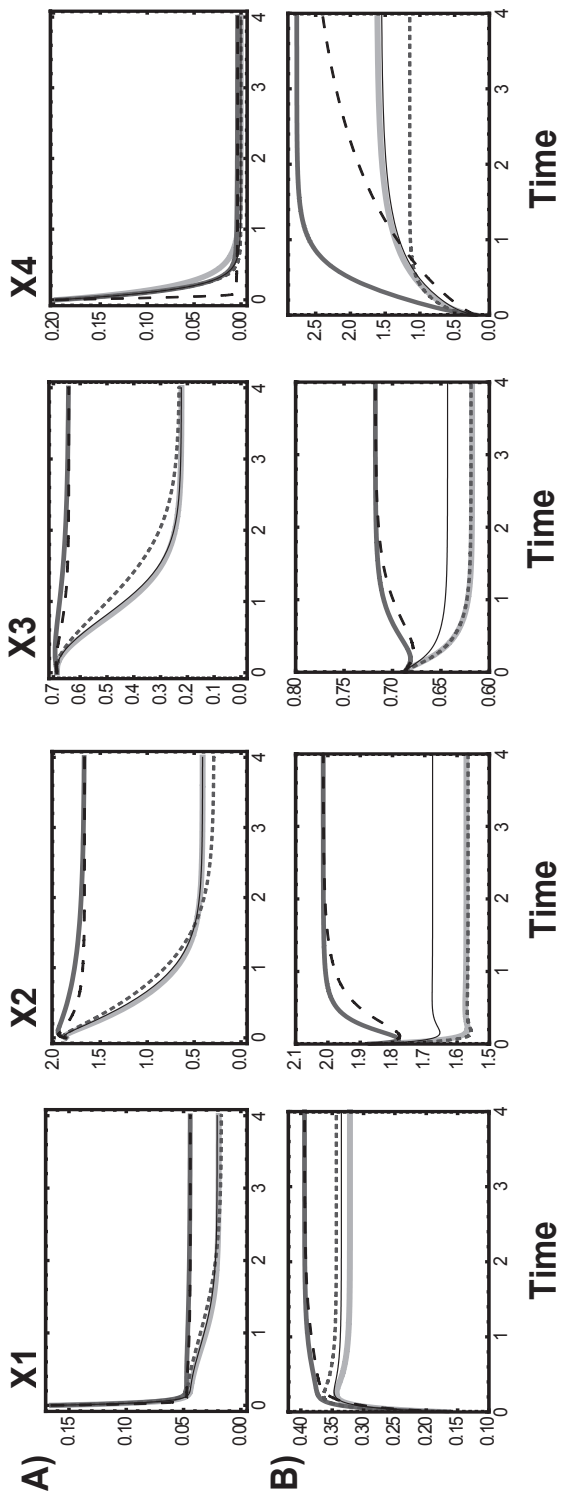

Figure 2. Dynamic responses of the dependent variables as we change the independent variable while keeping the operating point of the approximations constant. A) X_5=0.1 and B) X_5=1. Abbreviations: SC for Saturable and Cooperative; PL for Power-law; GMA for Generalized Mass Action; SS for S-System; and LL for Lin-log and (log) linear formalisms. The x-axis of the plots represents time, while the y-axis of the plots represents concentration of the relevant metabolite. By and large, the SC model is the most accurate in approximating the dynamical behavior of the original system.

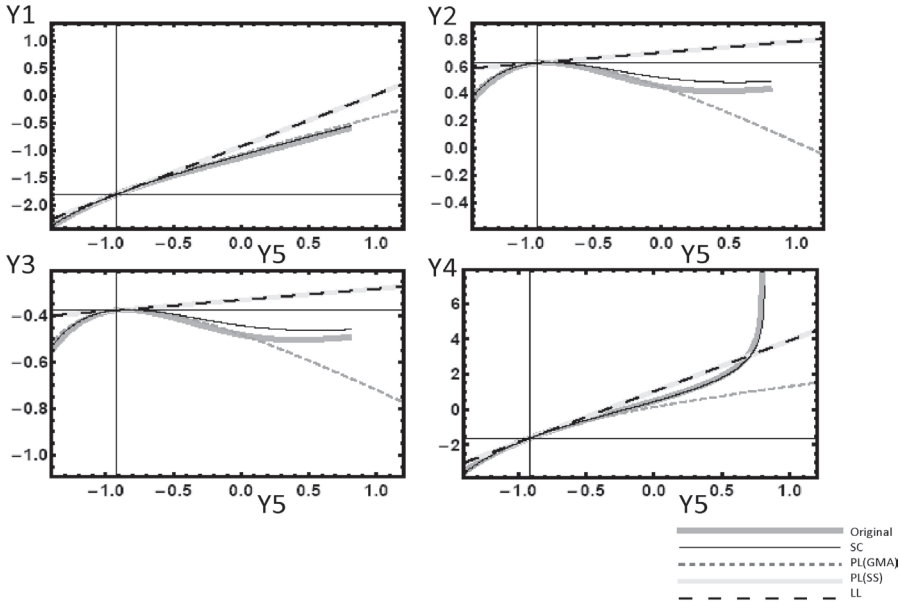

Figure 3. Predictions of the new steady-state concentrations as we change the independent variable X_5 while keeping the operating point of the approximations constant. The cross between the two lines in each plot indicates the operating point of the approximation. Abbreviations: Log(Xi)=Yi; SC for Saturable and Cooperative; PL for Power-law; GMA for Generalized Mass Action; SS for S-System; and LL for Lin-log and (log)linear formalisms. The x-axis of the plots represents time, while the y-axis of the plots represents concentration of the relevant metabolite. By and large, the SC model is the most accurate in approximating the dynamical behavior of the original system.

formalisms and the potential problems of the different formalisms, without technical details that would require a mathematical discussion. We will reference the relevant bibliography for those interested in such details.

GENOME-WIDE ANALYSIS AND PREDICTION OF CELLULAR PHENOTYPES

By far, most genome wide models have been created using linear stoichiometric models (Becker, and Palsson, 2005; Duarte *et al.*, 2004; Feist *et al.*, 2007; Jamshidi and Palsson, 2007; Mo *et al.*, 2007; Oh *et al.*, 2007; Resendis-Antonio *et al.*, 2007; Teusink *et al.*, 2006; Thiele *et al.*, 2005). By assuming that metabolism is at steady state one can analyze how certain changes in the environment or in a given gene propagate through the network, using a strategy known as Flux Balance Analysis (FBA). This analysis can improve our understanding of the relationships between genotype and phenotype (Edwards and Palsson, 1999; 2000a,b,c; Famili *et al.*, 2003; Savinell and Palsson, 1992a,b; Varma *et al.*, 1993; Varma and Palsson, 1994). FBA is a constraint-based approach that uses optimization methods to find appropriate flux distributions that may be compatible with specific stoichiometric matrices (Savinell and Palsson, 1992a,b). The constraints often assume that the organism is using material from the medium at maximum uptake rates, or growing at maximum velocities, etc. These assumptions constrain the solution space for flux optimization and help to define physiologically feasible solutions for the flux distribution upon specific changes to the genome. For example, the essentiality of different genes was predicted by individually knocking

out each gene in a genome-wide stoichiometric model of *Saccharomyces cerevisiae* metabolism. The predicted essentiality of a gene was experimentally confirmed in approximately 80% of the cases (Famili *et al.*, 2003). Applying a similar modeling strategy, Schilling and co-workers (Schilling *et al.*, 2002) developed a genome-scale metabolic model of *Helicobacter pylori* 26695. As stated by the authors, their analysis accurately predicted 10 of 17 gene deletion cases, when compared to actual phenotypes. The failures may be due to regulatory effects that cannot be accounted for using linear FBA models. Also, in its present formulation, FBA cannot predict metabolite levels and dynamic changes in metabolism. Recent proposals for increasing the accuracy of FBA analysis include considering thermodynamic constraints during the modelling and optimization (Feist *et al.*, 2007; Henry *et al.*, 2006; 2007; Hoppe *et al.*, 2007).

These results suggest that linear stoichiometric models and FBA analysis may adequately provide a broad picture of how cellular metabolism works. The need to go beyond FBA while interpreting metabolomic data has been recently emphasized by Lee and colleagues (Lee *et al.*, 2006). If one wants either to understand how specific parts of metabolism work or to increase the success rate in the predicting the association between genotype and phenotype, one requires non linear models that account for regulatory signals.

One natural way to extend FBA models is by using any of the formalisms previously discussed in this review, as they all account for regulatory interactions. Such models can also accurately predict dynamical aspects of the systemic behaviour. However, optimization in non-linear models is a difficult task. This explains why FBA is such a popular tool and why it is used in detriment of more detailed mathematical descriptions. Currently, optimization techniques have been developed only for the Power-law formalism (see section dedicated to optimization issues). Thus, this mathematical description would be the most indicated to extend the FBA to the nonlinear and regulatory domain. The optimization techniques available for Power-law models, either in S-system or GMA forms, takes advantage of the mathematical structure of the Power-law representation and are not easily extensible to the other alternatives.

PATHWAYS RECONSTRUCTION

In silico reconstruction of metabolic and signal transduction pathways and gene circuits is another type of common problem in Systems Biology. Traditionally, such reconstruction is based upon collecting published information for the individual enzymes and reactions that participate in the network one wants to reconstruct. Examples of this type of problem are many fold and range from metabolism to signal transduction, and to gene circuits. Some examples are the reconstructed network of Purine metabolism (Curto *et al.*, 1997; 1998a,b), and the reconstruction of the whole metabolic network of red blood cells (Ni and Savageau, 1996a,b). Mathematical models of signal transduction pathways have provided new information on the basic properties of signalling cascades in connection with their targets (see Klipp and Liebermeister, 2006 for a review). Another area where the use of mathematical models has facilitated the understanding of how a complex network of genes and proteins interacts to regulate and execute cellular functions is that of cell cycle (Alfieri *et al.*, 2007; Allen *et al.*, 2006; Barberis *et al.*, 2007; Brazhnik and Tyson, 2006; Lau *et al.*, 2007; Novak and Tyson, 2003; Sible and Tyson, 2007; Zi and Klipp, 2007).

Currently, a different type of reconstruction problem is emerging. Situations where a) new proteins and genes are found to play unknown roles in what were thought to be well characterized pathways, b) well known proteins play unknown roles in new pathways or c) unknown proteins play unknown roles in new ill-characterized pathways. In such a situation, one asks a) what is the role of the proteins in the pathway, b) what is the structure of the network underlying the pathway or c) what is the structure of the pathway and which role does each protein play?

To answer these questions one can create sets of mathematical models where large scale scanning of network structures and interactions can be done efficiently. In addition, for each network structure, extensive parameter scans must be fairly easy to execute. Comparison of the systemic behaviour of the models representing the alternative network structures to know aspects of the *in vivo* dynamical behaviour of the system can assist in choosing which network structures are more likely. This helps in prioritizing which hypothesis should be tested first regarding the system.

Using this strategy, Alves *et al.* have investigated the iron sulphur cluster biogenesis pathway of *S. cerevisiae* (Alves *et al.*, 2004a,b; Sorribas *et al.*, 2007). A set of alternative network structures was reconstructed from literature, structural, and expert information, and the resulting models were analyzed to compare dynamic predictions with experimental data. As a result, a number of hypotheses on the reaction and regulatory structure of the network that underlies iron sulphur cluster biogenesis could be discarded. Furthermore, specific experiments were suggested for obtaining information that would allow resolving the fine details of the network. Bas Teusink's group has been developing a method where sequence homology analysis is combined with the existence of well curated full metabolic networks to reconstruct the metabolic networks of new genomes (Francke *et al.*, 2005; Notebaart *et al.*, 2006; Teusink *et al.*, 2005; 2006). The major application of these methods has been to the reconstruction of *L. lactis* metabolism (Teusink *et al.*, 2005; 2006). Su *et al.* reconstruct both the pathway of phosphate assimilation and the gene circuits that regulate the expression of that pathway in *Synechococcus,* by using a combination of genomic and interaction information (Su *et al.*, 2006). A combination of literature analysis and microarray data analysis has also been used to derive a regulatory network for *E. coli* and test the consistency of microarray data based predictions (Gutierrez-Rios *et al.*, 2003). The regulatory network of the galactose biosynthesis pathway in yeast has also been reconstructed *in silico* by combining microarray data and protein interaction data (Darvish and Najarian,, 2006). A combination of time series gene expression analysis and *in silico* prediction of transcription factor biding sites has been used to define regulatory modules in the inflammatory response of the macrophage, suggesting novel roles for the transcription factors ATF3 and NRF2 (Nilsson *et al.*, 2006).

The combination of different datasets to generate testable hypothesis regarding the alternative connectivity of pathways is still a fairly manual process. For the most part, this process lacks a well defined structure and only partially allows for automatic combination of the different datasets. Some groups are already structuring various approaches. For example, Su *et al.* propose and apply an integrative approach for gene network reconstruction (Su *et al.*, 2006) as do Alves and Sorribas for the case of iron-sulfur cluster biogenesis (Alves *et al.*, 2004a,b; Sorribas *et al.*, 2007).

Because of the characteristics of this type of work, it is important to have a regular formalism for automated scanning of network structures. Any of the approximated

formalisms can facilitate this task. However, given that the parameter scans will assuredly take the system away from the operating point, it is important to choose an approximation whose range of accuracy is large. This, together with the fact that rates of biological processes often saturate, suggests that the SC formalism might be more appropriate for this task. However, if the network to reconstruct is large, this approximation will significantly increase the number of parameters one has to scan. In practice, the Power-law formalism might be an appropriate initial choice in those cases where the pathway to reconstruct is suspected to have a large number of individual processes that are to be considered. In subsequent, more detailed modelling analysis, one may use the SC formalism for those processes that depend on saturation effects. Thus, a mixed Power-law and SC model may help in characterizing the system's dynamic response. The Lin-log and the (log)linear could be used for the same task instead of the Power-law representation. However, as negative velocities may appear for low metabolite values, these two alternatives seem less appropriate than the Power-law formalism when it comes to the analysis of quantitative results.

DESIGN PRINCIPLES

The study of design principles in the structure of biological networks is a few decades old and was started by Savageau (1972; 1976). He developed a method, known as mathematical controlled comparisons, that allows for the comparison of alternative network structures by applying mathematical controls when comparing the dynamic behavior of models for the alternative networks. These controls ensure that any difference in the dynamic behavior of the system is due exclusively to the differences in topology between the alternative networks and not to other spurious differences. The differences in dynamic behavior are rationalized in terms of the functional requirements for the networks, and often this has implication for the evolution and ecology of the organisms (Savageau, 1974; 1976; 1998).

If one is interested in analyzing qualitative design principles of a network, that is, why a given network structure and not some other is selected to perform a given function, then one needs to compare the dynamic behavior of alternative classes of systems. Then, it is desirable that one is able to calculate the relevant properties of the alternative systems in closed form. Thus, the use of a mathematical formalism that has an analytical solution for those properties is required. This excludes the SC formalism as an option for this type of studies. The more widely used formalism for this purpose is the Power-law formalism (Alves and Savageau, 2000a,b; 2003; Hlavacek and Savageau, 1995; 1996; 1997; Igoshin*, et al.*, 2006; 2007; Irvine and Savageau, 1985a,b; Wall *et al.*, 2003; 2004). The S-system representation within this formalism has an analytical solution at the steady-state. This, together with the normally wider range of accuracy, makes the S-systems representation within the Power-law formalism ideal for use in the comparison of steady-state properties. The Lin-log formalism also provides an analytical solution for the steady-state and could be used for such comparisons.

Mathematical controlled comparisons have been extensively used to explore design principles in gene regulatory networks (Hlavacek and Savageau, 1995; 1996; 1997; Igoshin *et al.*, 2006; 2007; Savageau, 1998; Wall *et al.*, 2003; 2004), in signal transduction networks (Alves and Savageau, 2003), in metabolic networks (Alves and

Savageau, 2000; 2001, Savageau, 1972; 1976), and in immunological networks (De Boer and Hogeweg, 1987, Irvine and Savageau, 1985a,b; Ray and Kirschner, 2006). Often, for more complicated networks, the qualitative differences between relevant properties of alternative network designs are dependent on parameter-values. For example, the ratio between the sensitivity of network design A to some signal and the sensitivity of network design B to the same signal may be smaller or larger than one, depending on the actual parameter values. If this is the case, then a closed form solution does not help in deciding which system is better designed for a specific type of response to that signal. An extension of the method has allowed for the use of statistical mathematically controlled comparison (Alves and Savageau, 2000). This extension compares a large number of equivalent specific instances of alternative network designs and uses statistical criteria to understand which of the alternatives is more likely to be appropriately designed for a specific type of response. If a statistical mathematically controlled comparison is to be done, any of the formalisms can be used because parameters are attributed values and the comparisons are made numerically. In such cases, the formalism that is chosen to represent the model must consider the criteria discussed in this and previous sections**.**

It should be noted that statistical mathematically controlled comparisons can also be used as a subsequent step in the analysis during a traditional mathematically controlled comparison (Alves and Savageau, 2000; Schwacke and Voit, 2004). While the traditional comparison will provide information about the qualitative differences in systemic behavior, the statistical comparison will provide a statistical quantification of those differences.

As a final note we would like to point that Lau et al. have applied a form of uncontrolled comparison to the analysis of a Boolean network model to infer design principles in the network that controls cell cycle regulation and progression (Lau *et al.*, 2007).

OPERATIONAL PRINCIPLES

While analysis of qualitative design principles provides information about the evolution of network structures, analysis of operational principles provides information about fine tuning of parameters, about the evolution of specific dynamic behavior, once a network structure is in place, and the emergence of new regulatory requirements (Voit, 2003a,b). This is essentially a numerical task, although it may also require extensive scanning of network structure and parameter values. Thus, considerations similar to those discussed in the section devoted to pathways reconstruction should be taken into account when choosing a mathematical formalism to create model for this type of study.

In the context of the formalisms discussed in this review, this kind of systems biology problem has mainly been addressed through Power-law models. Operational principles on the adaptive response of yeast to heat shock have been investigated using models created with this formalism. Voit and Radivoyevitch (2000) suggest that the actual gene expression profile after heat shock seem to be an optimal functional solution for the cellular adaptation to heat. Vilaprinyo *et al.* (2006) extended this work and identified a set of functional criteria that explain the adaptive response of yeast to temperature changes.

Using classical kinetics approaches, Klipp *et al.* (2005) have analyzed osmotic response of yeast. Another example of this sort is the analysis of the implications beyond the quantitative values of human erythrocytes enzymes by Salvador and Savageau (2003; 2006). The potential of the SC formalism for application to this type of problems remains to be studied.

We would also like to stress that, while studying operational principles, the choice of formalism is often crucial for the correctness of the analysis. As we have shown for the system in *Figure 1*, the comparison of different formalisms reveal differences in the predicted behavior. A careful interpretation of the results for operational responses and optimization predictions should take into account the limitations and specificities of the selected formalism. This is especially relevant when we predict dynamic responses away from the operating point.

OPTIMIZATION

Trial and error strategies are behind many classical biotechnological applications, such as selecting yeast strains for bread and wine production, in an attempt to optimize the production of specific metabolites by the microbes. Despite clear methodological progress, the need for more efficient strategies in developing such applications is now evident, especially in connection with our ability to measure and manipulate cellular processes and gene expression. Mathematical models are at the core of a more efficient strategy for developing new biotechnological applications. These models are a tool that can be used for predicting the effect of alternative manipulations, thus creating a rank of priorities for which alternatives to implement first.

While using models to assist in the development of cellular strains that optimize production of some metabolite, mathematical optimization techniques play an important role. These techniques identify the changes in the values of systemic parameters that will make the system better achieve the relevant production goals. In the context of dynamic models, global optimization techniques for linear models are well known. In fact, the success of FBA applications relies on those methods. General and global optimization techniques that can be applied to non-linear models do not exist[6]. Several optimization methods may work for such models, but the results one obtains from applying those methods are more likely to be local than global optima. Canonical formalisms, such as the ones discussed in this review, can be helpful for developing global optimization methods for non-linear models.

To our knowledge, global optimization methods that rely on the formalism of the mathematical models have only been specifically developed for Power-law models (Marin-Sanguino and Torres, 2003; Torres and Voit, 2002). These methods are based on the fact that the steady-state equations for an S-system can be written as a linear system in log-log coordinates (Voit, 1992). Some examples that illustrate the advantage of using well structured and canonical models are the optimization of citric acid production (Alvarez-Vasquez *et al.*, 2000), tryptophan production (Marin-Sanguino and Torres, 2000), ethanol production (Vera *et al.*, 2003), and L-carnitine production

[6] Genetic algorithms, simulated annealing and other global optimization methods exist. However, in practice, these methods are only global optimization methods if one allows the optimization to run for an infinite amount of time.

(Alvarez-Vasquez, *et al.*, 2002; Sevilla *et al.*, 2005). The method has also been used to identify potential targets for drug action (Vera *et al.*, 2007). Recently, this method has been adapted through the utilization of geometric programming and it can be used for GMA models (Marin-Sanguino *et al.*, 2007).

In the future, optimization results obtained using non-linear dynamical models should be compared to those obtained through the use of FBA models. If the former results are a part of the set defined by the later results, FBA models could be used as an exploratory tool for optimization, before using more complete descriptions of the target system.

PARAMETER ESTIMATION THROUGH FITTING OF MODELS TO DYNAMIC DATA

Parameter estimation is one of the most difficult problems related to model building and utilization. As discussed previously, one of the important limitations of current data bases is the lack of specific information on kinetic properties of enzymes, which precludes using previous knowledge in the automatic generation of models. Furthermore, although such information can sometimes be retrieved by human curation from classical papers, in most cases the information refers to *in vitro* experiments performed in a plethora of conditions that do not reflect those within the cell.

In systems biology applications, *in vivo* measurements for the dynamical behaviour of the system are the data sets that would provide appropriate information for identification of systemic mechanisms and estimation of parameter values. Any of the formalisms discussed in this review can be used for such estimation purposes, if appropriated estimation procedures are devised. Due to the numerical problems that procedures for estimating parameter values must face (numerical integration of the differential equations, minimization, etc.), developing specific strategies that take advantage of the mathematical structure of each of the formalisms would greatly facilitate the estimation task.

To our knowledge, parameter identification procedures have been developed specifically only for Power-law models. One strategy that facilitates the estimation from dynamic data is the decoupling of the model equations, by estimating the slopes to substitute the derivatives (Lall and Voit, 2005; Veflingstad *et al.*, 2004; Voit and Almeida, 2004). In a different approach, genetic algorithms have been used as a method that can significantly speed-up the search for the best parameter set (Kimura *et al.*, 2005). Hybrid differential evolution has also been tested as a method that could provide a global solution to the estimation problem (Tsai and Wang, 2005). Simulated annealing, has been tested as a method for finding an appropriate data set for S-system models (Gonzalez *et al.*, 2007). All these methods have been developed for S-system models. A procedure that is specific for GMA models and use branch and bound methods has been proposed. This procedure finds a global optimization solution to the fitting problem (Polisetty *et al.*, 2006).

More recently, a new strategy base on alternate regression has been developed to facilitate the estimation task and avoid some of the numerical issues (Chou *et al.*, 2006). This strategy emphasizes the utility of using smoothing techniques for representing the time course of some variables while fitting parameter values. The smoothing techniques reduces the fitting problem to an iterative procedure that fits the values for a few parameters at each step of the iteration (Vilela *et al.*, 2007).

What must be emphasized is that parameter identification from dynamic data is a difficult task that must overcome the following problems:

(1) Collection of appropriate data sets. Ideally, this should include different perturbations and as much data points as possible. Metabolomic methods are the techniques that may provide such a large amount of data.

(2) Consider alternative network structures that may explain the data. A good data fit does not assure that the considered model reflects the actual structure of the biological system.

(3) Use the fitted model(s) to predict systems behaviour to unmeasured perturbations. Expert assessment of the predicted results and evaluation through appropriate experiments may help in finally assessing the best model and data set.

S-system and GMA models are the approximated models for which specific fitting methods have been developed. This makes them logical alternatives to be used for parameter fitting problems. SC models may provide an interesting alternative that could be used in cases where saturation is an issue for the processes under study.

Discussion

As discussed throughout this review, selecting a particular mathematical formalism and representation for a model is not a trivial issue. Each of the alternative formalisms has some advantages and limitations, which are related to being approximated representations of non-linear functions.

Table 2 provides a summary of the properties for the different formalisms and representations, as well as a short list of advantages and disadvantages of using each of the formalisms. There is no type of problem for which one can say "always use this formalism for this type of problem". Nevertheless, either due to the properties of the alternative formalisms and representations or to the tools that are available for a specific formalism, often, one can predict which alternative is more likely to be successful in a specific application, as was discussed above in the sections dedicated to the different formalisms. We would like to conclude by stressing again that we believe that the importance of these and other approximated formalisms in the study of biological systems is bound to increase, and become more central in Systems Biology. The reasons for this are several-fold:

1. A large amount of sequence, genomic, proteomic, metabolomic and fluxomic data is accumulating without an elucidation of the mechanism of individual molecular steps in the organism. A consequence of this is that approximate formalisms are the only available tools for creating mathematical models that use the accumulated data to gain understanding about the integrated workings of the molecular systems that compose a cell.

2. A critical step in the modelling process is the parameterization of a model. Approximate formalisms use a small number of parameters per individual process. Additionally, for some formalisms, such as the power- law formalism, it is fairly easy to obtain reasonable estimates for parameter values using

only qualitative information about the dynamic of the system. Currently, SC, (log)linear and Lin-log models have less specific methods available to estimate parameter values than the power- law formalism. Additionally, the SC formalism uses at least one more parameter per equation than the other described formalisms, which implies that more information is needed to parameterize SC models. The upside is that, if such information is available, SC models are likely to have a higher range of numerical accuracy.

3. Due to the uncertainty associated to many "omics" datasets and to the new pathways that are being discovered, it is important to have the ability to generate alternative models automatically. The use of approximate formalisms facilitates automation of the model set-up process. Because of the regular form of approximate formalisms, scripts that generate the mathematical models automatically based on the information provided by the conceptual schemas are easy to create. Automated model generation can then be connected to automated analysis of systemic behaviour, in a high-throughput manner.

4. The uncertainty of the data and the nature of models created using approximate formalisms allow for the possibility that, in the future, additional data may be generated that will prompt model update and reutilization. Again, the systematization of the modelling process that the use of approximated formalism allows is an advantage in this context.

5. Last but not least, often, biotechnological applications require cells to be functioning in a very restricted range of their operational capacities, because they are kept under very constant environments. Due to their nature, approximate formalisms are ideal for modelling this type of situations, because a) they allow researchers to accurately predict how a system will behave around an operating point and b) they can do so using less than complete information about the intrinsic mechanisms underlying systemic function.

Acknowledgments

This work has been partially supported by grants BFU2005-0234BMC and BFU2007-62772/BMC of the Spanish Ministerio de Educación y Ciencia. RA is supported by a Ramon y Cajal award from the Spanish Ministerio de Educacion y Ciencia.

References

Alfieri, R., Merelli, I., Mosca, E. and Milanesi, L. (2007). A data integration approach for cell cycle analysis oriented to model simulation in systems biology. *BMC Syst Biol*. **1**: 35.

Alvarez-Vasquez, F., Canovas, M., Iborra, J.L. and Torres, N.V. (2002). Modeling, optimization and experimental assessment of continuous L-(-)-carnitine production by Escherichia coli cultures. *Biotechnol Bioeng*. **80**: 794-805.

Alvarez-Vasquez, F., Gonzalez-Alcon, C. and Torres, N.V. (2000). Metabolism of

citric acid production by Aspergillus niger: model definition, steady-state analysis and constrained optimization of citric acid production rate. *Biotechnol Bioeng*. **70**: 82-108.

Alvarez-Vasquez, F., Sims, K.J., Cowart, L.A., Okamoto, Y., Voit, E.O. and Hannun, Y.A. (2005). Simulation and validation of modelled sphingolipid metabolism in Saccharomyces cerevisiae. *Nature*. **433**: 425-430.

Alves, R., Herrero, E. and Sorribas, A. (2004a). Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism. II. Role of glutaredoxin Grx5. *Proteins*. **57**: 481-492.

Alves, R., Herrero, E. and Sorribas, A. (2004b). Predictive reconstruction of the mitochondrial iron-sulfur cluster assembly metabolism: I. The role of the protein pair ferredoxin-ferredoxin reductase (Yah1-Arh1). *Proteins*. **56**: 354-366.

Alves, R. and Savageau, M.A. (2000). Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys J*. **79**: 2290-2304.

Alves, R. and Savageau, M.A. (2000). Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics*. **16**: 786-798.

Alves, R. and Savageau, M.A. (2001). Irreversibility in unbranched pathways: preferred positions based on regulatory considerations. *Biophys J*. **80**: 1174-1185.

Alves, R. and Savageau, M.A. (2003). Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: differences in molecular structure and physiological function. *Mol Microbiol*. **48**: 25-51.

Allen, N.A., Chen, K.C., Shaffer, C.A., Tyson, J.J. and Watson, L.T. (2006). Computer evaluation of network dynamics models with application to cell cycle control in budding yeast. *Syst Biol (Stevenage)*. **153**: 13-21.

Anacker, L.W. and Kopelman, R. (1987). Steady-state chemical kinetics on fractals: Segregation of reactants. *Phys Rev Lett*. **58**: 289-291.

Antunes, M.S., Ha, S.B., Tewari-Singh, N., Morey, K.J., Trofka, A.M., Kugrens, P., Deyholos, M. and Medford, J.I. (2006). A synthetic de-greening gene circuit provides a reporting system that is remotely detectable and has a re-set capacity. *Plant Biotechnol J*. **4**: 605-622.

Arkin, A.P. (2001). Synthetic cell biology. *Curr Opin Biotechnol*. **12**: 638-644.

Atkinson, M.R., Savageau, M.A., Myers, J.T. and Ninfa, A.J. (2003). Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell*. **113**: 597-607.

Atsumi, S. and Little, J.W. (2006). A synthetic phage lambda regulatory circuit. *Proc Natl Acad Sci U S A*. **103**: 19045-19050.

Bailey, J.E. (1991). Toward a science of metabolic engineering. *Science*. **252**: 1668-1675.

Bailey, J.E. (2001). Reflections on the scope and the future of metabolic engineering and its connections to functional genomics and drug discovery. *Metab Eng*. **3**: 111-114.

Bailey, J.E., Birnbaum, S., Galazzo, J.L., Khosla, C. and Shanks, J.V. (1990). Strategies and challenges in metabolic engineering. *Ann N Y Acad Sci*. **589**: 1-15.

Bailey, J.E., Sburlati, A., Hatzimanikatis, V., Lee, K., Renner, W.A. and Tsai, P.S. (2002). Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol Bioeng*. **79**: 568-579.

Barberis, M., Klipp, E., Vanoni, M. and Alberghina, L. (2007). Cell size at S phase initiation: an emergent property of the G1/S network. *PLoS Comput Biol*. **3**: e64.

Becker, S.A. and Palsson, B.O. (2005). Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiol*. **5**: 8.

Boros, L.G., Steinkamp, M.P., Fleming, J.C., Lee, W.N., Cascante, M. and Neufeld, E.J. (2003). Defective RNA ribose synthesis in fibroblasts from patients with thiamine-responsive megaloblastic anemia (TRMA). *Blood*. **102**: 3556-3561.

Brazhnik, P. and Tyson, J.J. (2006). Cell cycle control in bacteria and yeast: a case of convergent evolution? *Cell Cycle*. **5**: 522-529.

Chance, B. (1943) The kinetics of the enzyme-substrate compound of peroxidase. *J. Biol. Chem*. **151**: 553-577.

Craciun, G. and Feinberg, M. (2005). Multiple equilibria in complex chemical reaction networks: I. The injectivity property. *Siam Journal on Applied Mathematics*. **65**: 1526-1546.

Craciun, G. and Feinberg, M. (2006). Multiple equilibria in complex chemical reaction networks: II. The species-reaction graph. *Siam Journal on Applied Mathematics*. **66**: 1321-1338.

Craciun, G., Tang, Y.Z. and Feinberg, M. (2006). Understanding bistability in complex enzyme-driven reaction networks. *Proceedings of the National Academy of Sciences of the United States of America*. **103**: 8697-8702.

Curto, R., Voit, E.O. and Cascante, M. (1998a). Analysis of abnormalities in purine metabolism leading to gout and to neurological dysfunctions in man. *Biochem J*. **329**: 477-487.

Curto, R., Voit, E.O., Sorribas, A. and Cascante, M. (1997). Validation and steady-state analysis of a Power-law model of purine metabolism in man. *Biochem J*. **324**: 761-775.

Curto, R., Voit, E.O., Sorribas, A. and Cascante, M. (1998b). Mathematical models of purine metabolism in man. *Math Biosci*. **151**: 1-49.

Chou, I.C., Martens, H. and Voit, E.O. (2006). Parameter estimation in biochemical systems models with alternating regression. *Theor Biol Med Model*. **3**: 25.

Darvish, A. and Najarian, K. (2006). Prediction of regulatory pathways using rnRNA expression and protein interaction data: Application to identification of galactose regulatory pathway. *Biosystems*. **83**: 125-135.

De Boer, R.J. and Hogeweg, P. (1987). Self-nonself discrimination due to immunological nonlinearities: the analysis of a series of models by numerical methods. *IMA J Math Appl Med Biol*. **4**: 1-32.

Demongeot, J., Kaufman, M. and Thomas, R. (2000). Positive feedback circuits and memory. *C R Acad Sci III*. **323**: 69-79.

Duarte, N.C., Herrgard, M.J. and Palsson, B.O. (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res*. **14**: 1298-1309.

Edwards, J.S. and Palsson, B.O. (1999). Systems properties of the Haemophilus influenzae Rd metabolic genotype. *J Biol Chem*. **274**: 17410-17416.

Edwards, J.S. and Palsson, B.O. (2000a). The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A*. **97**: 5528-5533.

Edwards, J.S. and Palsson, B.O. (2000b). Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*. **1**: 1.

Edwards, J.S. and Palsson, B.O. (2000c). Robustness analysis of the Escherichia coli metabolic network. *Biotechnol Prog*. **16**: 927-939.

Edwards, J.S., Ramakrishna, R. and Palsson, B.O. (2002). Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng*. **77**: 27-36.

Famili, I., Forster, J., Nielsen, J. and Palsson, B.O. (2003). Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci U S A*. **100**: 13134-13139.

Feinberg, M. (1985). Reaction Network Structure and Multiple Steady-States in Complex Isothermal Reactors. *Abstracts of Papers of the American Chemical Society*. **189**: 35-Inde.

Feinberg, M. (1987). Chemical-Reaction Network Structure and the Stability of Complex Isothermal Reactors .1. The Deficiency-Zero and Deficiency-One Theorems. *Chemical Engineering Science*. **42**: 2229-2268.

Feinberg, M. (1988). Chemical-Reaction Network Structure and the Stability of Complex Isothermal Reactors .2. Multiple Steady-States for Networks of Deficiency One. *Chemical Engineering Science*. **43**: 1-25.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V. and Palsson, B.O. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*. **3**: 121.

Forster, A.C. and Church, G.M. (2007). Synthetic biology projects in vitro. *Genome Res*. **17**: 1-6.

Francke, C., Siezen, R.J. and Teusink, B. (2005). Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol*. **13**: 550-558.

Fung, E., Wong, W.W., Suen, J.K., Bulter, T., Lee, S.G. and Liao, J.C. (2005). A synthetic gene-metabolic oscillator. *Nature*. **435**: 118-122.

Gonzalez, O.R., Kuper, C., Jung, K., Naval, P.C., Jr. and Mendoza, E. (2007). Parameter estimation using Simulated Annealing for S-system models of biochemical networks. *Bioinformatics*. **23**: 480-486.

Greber, D. and Fussenegger, M. (2007). Mammalian synthetic biology: engineering of sophisticated gene networks. *J Biotechnol*. **130**: 329-345.

Gutierrez-Rios, R.M., Rosenblueth, D.A., Loza, J.A., Huerta, A.M., Glasner, J.D., Blattner, F.R. and Collado-Vides, J. (2003). Regulatory network of Escherichia coli: Consistency between literature knowledge and microarray profiles. *Genome Research*. **13**: 2435-2443.

Haseltine, E.L. and Arnold, F.H. (2007). Synthetic gene circuits: design with directed evolution. *Annu Rev Biophys Biomol Struct*. **36**: 1-19.

Hatzimanikatis, V. (1999). Nonlinear metabolic control analysis. *Metab Eng*. **1**: 75-87.

Hatzimanikatis, V. and Bailey, J.E. (1996). MCA has more to say. *J Theor Biol*. **182**: 233-242.

Hatzimanikatis, V., Li, C., Ionita, J.A. and Broadbelt, L.J. (2004). Metabolic networks: enzyme function and metabolite structure. *Curr Opin Struct Biol*. **14**: 300-306.

Hatzimanikatis, V., Li, C., Ionita, J.A., Henry, C.S., Jankowski, M.D. and Broadbelt, L.J. (2005). Exploring the diversity of complex metabolic networks. *Bioinformatics*.

21: 1603-1609.

Heijnen, J.J. (2005). Approximative kinetic formats used in metabolic network modeling. *Biotechnol Bioeng*. **91**: 534-545.

Heijnen, J.J., van Gulik, W.M., Shimizu, H. and Stephanopoulos, G. (2004). Metabolic flux control analysis of branch points: an improved approach to obtain flux control coefficients from large perturbation data. *Metab Eng*. **6**: 391-400.

Henry, C.S., Broadbelt, L.J. and Hatzimanikatis, V. (2007). Thermodynamics-based metabolic flux analysis. *Biophys J*. **92**: 1792-1805.

Henry, C.S., Jankowski, M.D., Broadbelt, L.J. and Hatzimanikatis, V. (2006). Genome-scale thermodynamic analysis of Escherichia coli metabolism. *Biophys J*. **90**: 1453-1461.

Hlavacek, W.S. and Savageau, M.A. (1995). Subunit structure of regulator proteins influences the design of gene circuitry: analysis of perfectly coupled and completely uncoupled circuits. *J Mol Biol*. **248**: 739-755.

Hlavacek, W.S. and Savageau, M.A. (1996). Rules for coupled expression of regulator and effector genes in inducible circuits. *J Mol Biol*. **255**: 121-139.

Hlavacek, W.S. and Savageau, M.A. (1997). Completely uncoupled and perfectly coupled gene expression in repressible systems. *J Mol Biol*. **266**: 538-558.

Hoppe, A., Hoffmann, S. and Holzhutter, H.G. (2007). Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst Biol*. **1**: 23.

Igoshin, O.A., Brody, M.S., Price, C.W. and Savageau, M.A. (2007). Distinctive topologies of partner-switching signaling networks correlate with their physiological roles. *J Mol Biol*. **369**: 1333-1352.

Igoshin, O.A., Price, C.W. and Savageau, M.A. (2006). Signalling network with a bistable hysteretic switch controls developmental activation of the sigma transcription factor in Bacillus subtilis. *Mol Microbiol*. **61**: 165-184.

Irvine, D.H. and Savageau, M.A. (1985a). Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes. *J Immunol*. **134**: 2100-2116.

Irvine, D.H. and Savageau, M.A. (1985b). Network regulation of the immune response: modulation of suppressor lymphocytes by alternative signals including contrasuppression. *J Immunol*. **134**: 2117-2130.

Irvine, D.H. and Savageau, M.A. (1990). Efficient Solution of Nonlinear Ordinary Differential-Equations Expressed in S-System Canonical Form. *Siam Journal on Numerical Analysis*. **27**: 704-735.

Jamshidi, N. and Palsson, B.O. (2007). Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol*. **1**: 26.

Kaufman, M. and Thomas, R. (2003). Emergence of complex behaviour from simple circuit structures. *C R Biol*. **326**: 205-214.

Kim, J., White, K.S. and Winfree, E. (2006). Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Mol Syst Biol*. **2**: 68.

Kimura, S., Ide, K., Kashihara, A., Kano, M., Hatakeyama, M., Masui, R., Nakagawa, N., Yokoyama, S., Kuramitsu, S. and Konagaya, A. (2005). Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*. **21**: 1154-1163.

Kitano, H. (2007). Towards a theory of biological robustness. *Mol Syst Biol*. **3**: 137.

Klipp, E. and Liebermeister, W. (2006). Mathematical modeling of intracellular signaling pathways. *BMC Neurosci*. **7 Suppl 1**: S10.

Klipp, E., Nordlander, B., Kruger, R., Gennemark, P. and Hohmann, S. (2005). Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol*. **23**: 975-982.

Kopelman, R. (1988). Fractal Reaction Kinetics. *Science*. **241**: 1620-1626.

Kresnowati, M.T., van Winden, W.A. and Heijnen, J.J. (2005). Determination of elasticities, concentration and flux control coefficients from transient metabolite data using linlog kinetics. *Metab Eng*. **7**: 142-153.

Lall, R. and Voit, E.O. (2005). Parameter estimation in modulated, unbranched reaction chains within biochemical systems. *Comput Biol Chem*. **29**: 309-318.

Lau, K.Y., Ganguli, S. and Tang, C. (2007). Function constrains network architecture and dynamics: a case study on the yeast cell cycle Boolean network. *Phys Rev E Stat Nonlin Soft Matter Phys*. **75**: 051907.

Lee, J.M., Gianchandani, E.P. and Papin, J.A. (2006). Flux balance analysis in the era of metabolomics. *Brief Bioinform*. **7**: 140-150.

Lee, W.N., Guo, P., Lim, S., Bassilian, S., Lee, S.T., Boren, J., Cascante, M., Go, V.L. and Boros, L.G. (2004). Metabolic sensitivity of pancreatic tumour cell apoptosis to glycogen phosphorylase inhibitor treatment. *Br J Cancer*. **91**: 2094-2100.

Luisi, P.L. (2007). Chemical aspects of synthetic biology. *Chem Biodivers*. **4**: 603-621.

Maher, A.D., Kuchel, P.W., Ortega, F., de Atauri, P., Centelles, J. and Cascante, M. (2003). Mathematical modelling of the urea cycle. A numerical investigation into substrate channelling. *Eur J Biochem*. **270**: 3953-3961.

Marin-Sanguino, A. and Torres, N.V. (2000). Optimization of tryptophan production in bacteria. Design of a strategy for genetic manipulation of the tryptophan operon for tryptophan flux maximization. *Biotechnol Prog*. **16**: 133-145.

Marin-Sanguino, A. and Torres, N.V. (2003). Optimization of biochemical systems by linear programming and general mass action model representations. *Math Biosci*. **184**: 187-200.

Marin-Sanguino, A., Voit, E.O., Gonzalez-Alcon, C. and Torres, N.V. (2007). Optimization of biotechnological systems through geometric programming. *Theor Biol Med Model*. **4**: 38.

Meyer, A., Pellaux, R. and Panke, S. (2007). Bioengineering novel in vitro metabolic pathways using synthetic biology. *Curr Opin Microbiol*. **10**: 246-253.

Mo, M.L., Jamshidi, N. and Palsson, B.O. (2007). A genome-scale, constraint-based approach to systems biology of human metabolism. *Mol Biosyst*. **3**: 598-603.

Ni, T.C. and Savageau, M.A. (1996a). Application of biochemical systems theory to metabolism in human red blood cells. Signal propagation and accuracy of representation. *J Biol Chem*. **271**: 7927-7941.

Ni, T.C. and Savageau, M.A. (1996b). Model assessment and refinement using strategies from biochemical systems theory: application to metabolism in human red blood cells. *J Theor Biol*. **179**: 329-368.

Nilsson, R., Bajic, V.B., Suzuki, H., di Bernardo, D., Bjorkegren, J., Katayama, S., Reid, J.F., Sweet, M.J., Gariboldi, M., Carninci, P., Hayashizaki, Y., Hume, D.A., Tegner, J. and Ravasi, T. (2006). Transcriptional network dynamics in macrophage

activation. *Genomics*. **88**: 133-142.

Notebaart, R.A., van Enckevort, F.H., Francke, C., Siezen, R.J. and Teusink, B. (2006). Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics*. **7**: 296.

Novak, B. and Tyson, J.J. (2003). Modelling the controls of the eukaryotic cell cycle. *Biochem Soc Trans*. **31**: 1526-1529.

Oh, Y.K., Palsson, B.O., Park, S.M., Schilling, C.H. and Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*. **282**: 28791-28799.

Orosz, F., Wagner, G., Ortega, F., Cascante, M. and Ovadi, J. (2003). Glucose conversion by multiple pathways in brain extract: theoretical and experimental analysis. *Biochem Biophys Res Commun*. **309**: 792-797.

Pleiss, J. (2006). The promise of synthetic biology. *Appl Microbiol Biotechnol*. **73**: 735-739.

Polisetty, P.K., Voit, E.O. and Gatzke, E.P. (2006). Identification of metabolic system parameters using global optimization methods. *Theor Biol Med Model*. **3**: 4.

Prati, E.G., Matasci, M., Suter, T.B., Dinter, A., Sburlati, A.R. and Bailey, J.E. (2002). Engineering of coordinated up- and down-regulation of two glycosyltransferases of the O-glycosylation pathway in Chinese hamster ovary (CHO) cells. *Biotechnol Bioeng*. **79**: 580-585.

Ramos-Montoya, A., Lee, W.N., Bassilian, S., Lim, S., Trebukhina, R.V., Kazhyna, M.V., Ciudad, C.J., Noe, V., Centelles, J.J. and Cascante, M. (2006). Pentose phosphate cycle oxidative and nonoxidative balance: A new vulnerable target for overcoming drug resistance in cancer. *Int J Cancer*. **119**: 2733-2741.

Ray, J.C. and Kirschner, D.E. (2006). Requirement for multiple activation signals by anti-inflammatory feedback in macrophages. *J Theor Biol*. **241**: 276-294.

Resendis-Antonio, O., Reed, J.L., Encarnacion, S., Collado-Vides, J. and Palsson, B.O. (2007). Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli. *PLoS Comput Biol*. **3**: 1887-1895.

Rodriguez-Caso, C., Montanez, R., Cascante, M., Sanchez-Jimenez, F. and Medina, M.A. (2006). Mathematical modeling of polyamine metabolism in mammals. *J Biol Chem*. **281**: 21799-21812.

Rosenfeld, N., Young, J.W., Alon, U., Swain, P.S. and Elowitz, M.B. (2007). Accurate prediction of gene feedback circuit behavior from component properties. *Mol Syst Biol*. **3**: 143.

Saito, H. and Inoue, T. (2007). RNA and RNP as new molecular parts in synthetic biology. *J Biotechnol*. **132**: 1-7.

Salvador, A. and Savageau, M.A. (2003). Quantitative evolutionary design of glucose 6-phosphate dehydrogenase expression in human erythrocytes. *Proc Natl Acad Sci U S A*. **100**: 14463-14468.

Salvador, A. and Savageau, M.A. (2006). Evolution of enzymes in a series is driven by dissimilar functional demands. *Proc Natl Acad Sci U S A*. **103**: 2226-2231.

Savageau, M.A. (1969a). Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol*. **25**: 365-369.

Savageau, M.A. (1969b). Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a Power-law approximation. *J Theor Biol*. **25**: 370-379.

Savageau, M.A. (1970). Biochemical systems analysis. 3. Dynamic solutions using a Power-law approximation. *J Theor Biol*. **26**: 215-226.

Savageau, M.A. (1971a). Concepts relating the behavior of biochemical systems to their underlying molecular properties. *Arch Biochem Biophys*. **145**: 612-621.

Savageau, M.A. (1971b). Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature*. **229**: 542-544.

Savageau, M.A. (1972). The behavior of intact biochemical control systems. *Current topics of cellular regulation*. **6**: 63-130.

Savageau, M.A. (1974). Genetic regulatory mechanisms and the ecological niche of Escherichia coli. *Proc Natl Acad Sci U S A*. **71**: 2453-2455.

Savageau, M.A. (1974). Optimal design of feedback control by inhibition. *J Mol Evol*. **4**: 139-156.

Savageau, M.A. (1976). Biochemical systems analysis : a study of function and design in molecular biology. Reading, Mass. Addison-Wesley.

Savageau, M.A. (1995). Michaelis-Menten mechanism reconsidered: implications of fractal kinetics. *J Theor Biol*. **176**: 115-124.

Savageau, M.A. (1998). Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of Escherichia coli. *Genetics*. **149**: 1677-1691.

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

Savinell, J.M. and Palsson, B.O. (1992a). Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J Theor Biol*. **154**: 421-454.

Savinell, J.M. and Palsson, B.O. (1992b). Optimal selection of metabolic fluxes for in vivo measurement. I. Development of mathematical methods. *J Theor Biol*. **155**: 201-214.

Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S. and Palsson, B.O. (2002). Genome-scale metabolic model of Helicobacter pylori 26695. *J Bacteriol*. **184**: 4582-4593.

Schnell, S. and Turner, T.E. (2004). Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog Biophys Mol Biol*. **85**: 235-260.

Schwacke, J.H. and Voit, E.O. (2004). Improved methods for the mathematically controlled comparison of biochemical systems. *Theor Biol Med Model*. **1**: 1.

Selivanov, V.A., de Atauri, P., Centelles, J.J., Cadefau, J., Parra, J., Cusso, R., Carreras, J. and Cascante, M. (2007). The changes in the energy metabolism of human muscle induced by training. *J Theor Biol*.

Sevilla, A., Vera, J., Diaz, Z., Canovas, M., Torres, N.V. and Iborra, J.L. (2005). Design of metabolic engineering strategies for maximizing L-(-)-carnitine production by Escherichia coli. Integration of the metabolic and bioreactor levels. *Biotechnol Prog*. **21**: 329-337.

Shiraishi, F. and Savageau, M.A. (1992a). The tricarboxylic acid cycle in Dictyostelium discoideum. I. Formulation of alternative kinetic representations. *J Biol Chem*. **267**: 22912-22918.

Shiraishi, F. and Savageau, M.A. (1992b). The tricarboxylic acid cycle in Dictyostelium discoideum. II. Evaluation of model consistency and robustness. *J Biol Chem*. **267**:

22919-22925.

Shiraishi, F. and Savageau, M.A. (1992c). The tricarboxylic acid cycle in Dictyostelium discoideum. IV. Resolution of discrepancies between alternative methods of analysis. *J Biol Chem*. **267**: 22934-22943.

Sible, J.C. and Tyson, J.J. (2007). Mathematical modeling as a tool for investigating cell cycle control networks. *Methods*. **41**: 238-247.

Sims, K.J., Spassieva, S.D., Voit, E.O. and Obeid, L.M. (2004). Yeast sphingolipid metabolism: clues and connections. *Biochem Cell Biol*. **82**: 45-61.

Sole, R.V., Munteanu, A., Rodriguez-Caso, C. and Macia, J. (2007). Synthetic protocell biology: from reproduction to computation. *Philos Trans R Soc Lond B Biol Sci*. **362**: 1727-1739.

Sorribas, A., Hernandez-Bermejo, B., Vilaprinyo, E. and Alves, R. (2007). Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnol Bioeng*. **97**: 1259-1277.

Sorribas, A. and Savageau, M.A. (1989a). A comparison of variant theories of intact biochemical systems. I. Enzyme-enzyme interactions and biochemical systems theory. *Math Biosci*. **94**: 161-193.

Sorribas, A. and Savageau, M.A. (1989b). A comparison of variant theories of intact biochemical systems. II. Flux-oriented and metabolic control theories. *Math Biosci*. **94**: 195-238.

Sorribas, A. and Savageau, M.A. (1989c). Strategies for representing metabolic pathways within biochemical systems theory: reversible pathways. *Math Biosci*. **94**: 239-269.

Sprinzak, D. and Elowitz, M.B. (2005). Reconstruction of genetic circuits. *Nature*. **438**: 443-448.

Su, Z.C., Mao, F.L., Dam, P., Wu, H.W., Olman, V., Paulsen, I.T., Palenik, B. and Xu, Y. (2006). Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium Synechococcus sp WH 8102. *Nucleic Acids Research*. **34**: 1050-1065.

Takahashi, K., Ishikawa, N., Sadamoto, Y., Sasamoto, H., Ohta, S., Shiozawa, A., Miyoshi, F., Naito, Y., Nakayama, Y. and Tomita, M. (2003). E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics*. **19**: 1727-1729.

Teusink, B., van Enckevort, F.H., Francke, C., Wiersma, A., Wegkamp, A., Smid, E.J. and Siezen, R.J. (2005). In silico reconstruction of the metabolic pathways of Lactobacillus plantarum: comparing predictions of nutrient requirements with those from growth experiments. *Appl Environ Microbiol*. **71**: 7253-7262.

Teusink, B., Wiersma, A., Molenaar, D., Francke, C., de Vos, W.M., Siezen, R.J. and Smid, E.J. (2006). Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model. *J Biol Chem*. **281**: 40041-40048.

Thiele, I., Vo, T.D., Price, N.D. and Palsson, B.O. (2005). Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J Bacteriol*. **187**: 5818-5830.

Thomas, R. (1973). Boolean formalization of genetic control circuits. *J Theor Biol*. **42**: 563-585.

Thomas, R., Thieffry, D. and Kaufman, M. (1995). Dynamical behaviour of biological

regulatory networks--I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol*. **57**: 247-276.

Torres, N.V. and Voit, E.O. (2002). Pathway Analysis and Optimization in Metabolic Engineering CA. Cambridge University Press.

Tsai, K.Y. and Wang, F.S. (2005). Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics*. **21**: 1180-1188.

Varma, A., Boesch, B.W. and Palsson, B.O. (1993). Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. *Appl Environ Microbiol*. **59**: 2465-2473.

Varma, A. and Palsson, B.O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol*. **60**: 3724-3731.

Veflingstad, S.R., Almeida, J. and Voit, E.O. (2004). Priming nonlinear searches for pathway identification. *Theor Biol Med Model*. **1**: 8.

Vera, J., Curto, R., Cascante, M. and Torres, N.V. (2007). Detection of potential enzyme targets by metabolic modelling and optimization: application to a simple enzymopathy. *Bioinformatics*. **23**: 2281-2289.

Vera, J., de Atauri, P., Cascante, M. and Torres, N.V. (2003). Multicriteria optimization of biochemical systems by linear programming: application to production of ethanol by Saccharomyces cerevisiae. *Biotechnol Bioeng*. **83**: 335-343.

Vilaprinyo, E., Alves, R. and Sorribas, A. (2006). Use of physiological constraints to identify quantitative design principles for gene expression in yeast adaptation to heat shock. *BMC Bioinformatics*. **7**: 184.

Vilela, M., Borges, C.C., Vinga, S., Vasconcelos, A.T., Santos, H., Voit, E.O. and Almeida, J.S. (2007). Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinformatics*. **8**: 305.

Visser, D., Schmid, J.W., Mauch, K., Reuss, M. and Heijnen, J.J. (2004). Optimal re-design of primary metabolism in Escherichia coli using linlog kinetics. *Metab Eng*. **6**: 378-390.

Voit, E., Neves, A.R. and Santos, H. (2006). The intricate side of systems biology. *Proc Natl Acad Sci U S A*. **103**: 9452-9457.

Voit, E.O. (1992). Optimization in integrated biochemical systems. *Biotechnol Bioeng*. **40**: 572-582.

Voit, E.O. (2000). Computational analysis of biochemical systems : a practical guide for biochemists and molecular biologists. Cambridge ; New York. Cambridge University Press.

Voit, E.O. (2002). Models-of-data and models-of-processes in the post-genomic era. *Math Biosci*. **180**: 263-274.

Voit, E.O. (2003a). Biochemical and genomic regulation of the trehalose cycle in yeast: review of observations and canonical model analysis. *J Theor Biol*. **223**: 55-78.

Voit, E.O. (2003b). Design principles and operating principles: the yin and yang of optimal functioning. *Math Biosci*. **182**: 81-92.

Voit, E.O. and Almeida, J. (2004). Decoupling dynamical systems for pathway identification from metabolic profiles. *Bioinformatics*. **20**: 1670-1681.

Voit, E.O. and Radivoyevitch, T. (2000). Biochemical systems analysis of genome-wide expression data. *Bioinformatics*. **16**: 1023-1037.

Wall, M.E., Hlavacek, W.S. and Savageau, M.A. (2003a). Design principles for

regulator gene expression in a repressible gene circuit. *J Mol Biol*. **332**: 861-876.

Wall, M.E., Hlavacek, W.S. and Savageau, M.A. (2004b). Design of gene circuits: lessons from bacteria. *Nat Rev Genet*. **5**: 34-42.

Weber, W., Stelling, J., Rimann, M., Keller, B., Daoud-El Baba, M., Weber, C.C., Aubel, D. and Fussenegger, M. (2007). A synthetic time-delay circuit in mammalian cells and mice. *Proc Natl Acad Sci U S A*. **104**: 2643-2648.

Wu, L., Wang, W., van Winden, W.A., van Gulik, W.M. and Heijnen, J.J. (2004). A new framework for the estimation of control parameters in metabolic pathways using lin-log kinetics. *Eur J Biochem*. **271**: 3348-3359.

Yokobayashi, Y., Weiss, R. and Arnold, F.H. (2002). Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A*. **99**: 16587-16591.

Zi, Z. and Klipp, E. (2007). Cellular signaling is potentially regulated by cell density in receptor trafficking networks. *FEBS Lett*. **581**: 4589-4595.