

Approximate kinetic formalisms for modeling metabolic networks: does anything work?

Albert Sorribas

Dep.Ciències Mèdiques Bàsiques
IRBLLLEIDA-Univ.of Lleida (Spain)
Montserrat Roig, 2, 25008-Lleida
34 973 702 2406
albert.sorribas@cmb.udl.cat

Ester Vilaprinyo

Dep.Ciències Mèdiques Bàsiques
IRBLLLEIDA-Univ.of Lleida (Spain)
Montserrat Roig, 2, 25008-Lleida
34 973 702 2406
evilaprinyo@cmb.udl.cat

Rui Alves

Dep.Ciències Mèdiques Bàsiques
IRBLLLEIDA-Univ.of Lleida (Spain)
Montserrat Roig, 2, 25008-Lleida
34 973 702 2406
ralves@cmb.udl.cat

ABSTRACT

Approximated kinetic formalisms provide useful mathematical representations for modeling complex genetic and metabolic networks. Although the different formalisms that currently have a wider use for modeling biological systems have technical similarities, their practical utility differs. A formal comparison of the rationale behind each formalism helps understanding their similarities and differences. From a practical point of view, we discuss basic requirements of a useful mathematical description and provide a rationale for selecting a particular formalism depending on the purpose of the study. Such an analysis, suggests that the power-law formalism has clear advantages if one's purpose is that of analyzing circuit design and systemic properties. The recently proposed Saturable and Cooperative formalism can provide an additional tool that allows for a complementary analysis of the predictions based on the power-law formalism.

Keywords

Power-law formalism, Biochemical Systems Theory, S-system, GMA, Taylor's theorem, Function approximation, Modeling, Systems Biology

1. INTRODUCTION

Driven by the current rise of Systems Biology, there is a renewed interest in obtaining a systemic view of how metabolic processes work. Mathematical models are fundamental in such a task and they are becoming central to many Systems Biology applications.

There are several issues at stake when one has to decide upon which mathematical approximation and model are more appropriate to address a specific problem. First, one needs to choose a model that is able to represent the aspects of the system one wishes to study. Second, one must choose a mathematical approximation that is able to capture the dynamic aspects of interests in the systemic behavior in an accurate way. Third, it

facilitates one's job if the approximation has a regular structure that allows for automatic generation of the equations from the model schema. Fourth, an added bonus of using a regular representation is that it facilitates model transference, recycling and expansion by other researchers and for distinct purposes..

In this context, the choice of an appropriate mathematical formalism is an important issue. Among other possibilities, the use of approximate kinetic representations provides a practical solution.

The power-law formalism, based on a Taylor's series approximation in logarithmic coordinates, was derived by Savageau in the late sixties [1]. A well developed set of tools exist for systemic modeling and analysis using this approach [2]. The application of approximation theory by different groups has resulted in various formalisms that offer alternative strategies for obtaining useful representation for mathematical modeling.

In this contribution we compare the more popular formalisms that have been derived using approximation theory, with a focus on the formalisms that are theoretically supported by the Taylor Theorem. These include the power-law formalism, and the recently proposed (log)linear and Lin-log formalisms (see [3] for a review) as well as the recently suggested Saturable and Cooperative formalism [4]. We highlight similarities and differences between formalisms, and discuss their individual advantages, limitations, and potential utility.

2. APPROXIMATED KINETIC FORMALISMS

Mathematical models provide a tool for investigating the integrated behavior of complex systems. Mathematical formalisms based on different simplifications and on approximated kinetic representations that reduce the mathematical complexity of the model while still reproducing its dynamic behavior help in simplifying the analysis and in providing an appropriate tool for understanding systemic properties of a network. This focus is justified by the following reasons:

(1) In many cases, one cannot find information that can be used for a detailed description of the mechanisms and for the estimation of individual parameter values. This is especially true while dealing with poorly characterized systems. Thus, we need mathematical formalisms that can be used under such restrictive conditions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference'04, Month 1–2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

(2) The widespread use of rational kinetics (e.g. Michaelis-Menten) may be accurate only if the theoretical conditions that support the derivation of such functions hold. It is not clear if they correspond to the properties of *in vivo* conditions. Approximated representations may guarantee a valid representation.

(4) When in the process of model simplification, one often lumps different variables and processes together. This creates aggregated processes, often like black boxes, with non-traditional and unknown kinetic descriptions. Thus, use of approximation theory to derive a mathematical description for the kinetics of this black box also facilitates the modeling process.

(5) Approximate representations provide a systematic and automatable way for building a model from scratch. This is especially useful while modeling large systems or when exploring large alternative sets for fitting to unknown network structures.

(6) 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 representation must capture the essential properties of the problem.

(7) Approximate representations also have the advantage of allowing models to account for regulatory interactions for which not much information is available. Furthermore, qualitative information can be easily incorporated in this class of models.

2.1 Formalisms based on approximation theory

Consider a process which velocity can be expressed as

$$v = \Psi(E, \mathbf{X}, \boldsymbol{\theta}) \quad (1)$$

where \mathbf{X} is a vector of metabolites, effectors, etc., $\boldsymbol{\theta}$ is a vector of parameters, and E is the enzyme. Ψ is a non-linear function that accounts for the kinetics of the process. As in many cases the velocity is linear with respect to the enzyme, one can alternatively consider the expression:

$$v = E \Psi(\mathbf{X}, \boldsymbol{\theta}) \quad (2)$$

The function Ψ could be any of the commonly used kinetic rate-laws. However, selection of a given function will be difficult, unless sound evidence exist on the kinetic characteristics of the target process. In many practical modeling situations Ψ will be unknown, and the use of a particular non-linear function for modeling purposes based on a subjective choice, either by familiarity or by mimetic reason, may be discussable. In such a case, we can use approximation theory (Taylor's series) to develop different approximations to Ψ around an operating point and obtain a simple and structured representation of the original unknown function. The different formalisms are obtained after using Taylor's series at different levels. After considering each case, they can be formally related in the following way:

Approximation strategy	Name
Taylor series approximation of $\Psi(E, \mathbf{X}, \boldsymbol{\theta})$	Linear-1
Taylor series approximation of $\Psi(E, \mathbf{X}, \boldsymbol{\theta})$ and approximation of (y-1) by Log(y)	(log)linear
Taylor series approximation of $\Psi(\mathbf{X}, \boldsymbol{\theta})$	Linear-2
Taylor series approximation of $\Psi(\mathbf{X}, \boldsymbol{\theta})$ and approximation of (y-1) by Log(y)	Lin-log
Taylor series approximation of $\Psi(E, \mathbf{X}, \boldsymbol{\theta})$ in log-log space	Power-law
Taylor series approximation of $\Psi(E, \mathbf{X}, \boldsymbol{\theta})$ in generalized inverse space	Saturable and Cooperative

According to these strategies, the formal expressions obtained for the different formalisms are as follows:

Linear

$$\frac{v_i}{v_{i0}} = 1 + f_{iE} \left(\frac{E_i}{E_{i0}} - 1 \right) + \sum_{j=1}^m f_{ij} \left(\frac{X_j}{X_{j0}} - 1 \right) \quad (3)$$

(log)Linear

$$\frac{v_i}{v_{i0}} = 1 + f_{iE} \log \left(\frac{E_i}{E_{i0}} \right) + \sum_{j=1}^m f_{ij} \log \left(\frac{X_j}{X_{j0}} \right) \quad (4)$$

Linear (2)

$$\frac{v_i}{v_{i0}} = \frac{E_i}{E_{i0}} \times \left\{ 1 + \sum_{j=1}^m \left(f_{ij} \left(\frac{X_j}{X_{j0}} - 1 \right) \right) \right\} \quad (5)$$

Lin-log

$$\frac{v_i}{v_{i0}} = \frac{E_i}{E_{i0}} \times \left\{ 1 + \sum_{j=1}^m f_{ij} \log \left(\frac{X_j}{X_{j0}} \right) \right\} \quad (6)$$

Power-law

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

$$\frac{v_i}{v_{i0}} = \left(\frac{E_i}{E_{i0}} \right)^{f_{iE}} \prod_{j=1}^m \left(\frac{X_j}{X_{j0}} \right)^{f_{ij}}$$

Saturable and Cooperative

$$v_i = \frac{V_i \prod_{j=1}^{n+m} X_j^{n_{ij}}}{\prod_{j=1}^{n+m} (K_{ij} + X_j^{n_{ij}})} \quad (8)$$

$$n_{ij} = \frac{f_{ij}}{(1 - p_{ij})} \quad (9)$$

$$K_{ij} = \frac{(1 - p_{ij})}{p_{ij}} X_{j0}^{n_{ij}}$$

The previous equations show each of the formalisms in a representation that help comparing their relationships. In all these cases, the following features are common:

1. Any of these representations is exact at the operating point.
2. Any of these representations provides an accurate approximation around the operating point.
3. For any of these representations the accuracy decreases when we move away from the operating point.
4. The resulting of any of these representations is systematic in the sense that it has the same mathematical form for any case.

Furthermore, they share the following basic parameters (although they appear with different names and are treated in slightly different manner):

Operational point values of metabolites and fluxes: By definition, the metabolite values and fluxes have the same value at the operating point.

Local sensitivities at the operating point: Local sensitivities, defined as

$$f_{ij} = \left(\frac{\partial v_i}{\partial X_j} \frac{X_j}{v_i} \right)_0 \quad (10)$$

are equal for each of the formalisms.

3. ADVANTAGES AND LIMITATIONS

Although any of the proposed formalisms is useful for modeling purposes, their applicability varies. For instance, the linear formalism is clearly the less adequate as this approximation cannot capture the essential non-linear features of many biological processes.

The power-law formalism is, by large, the most developed of the alternative representations. A wide set of tools exist for analyzing a model developed with this formalism. These include stability analysis, steady-state characterization, parameter estimation, optimization, etc.

Focusing on the non-linear representations, we can identify some limitations that must be considered in applications. The (log)linear and lin-log formalism may approximate saturation in some cases better than the power-law formalism, although by definition none of them saturate. On the other hand, if the system

moves below the operating point values, negative values for the velocities are obtained for the (log)linear and lin-log representations. This problem is, in practice, an important limitation for these approximations. Although the sensitivity of the (log)linear and lin-log approximations to any of the variables changes as we move from the operating point, the obtained values may violate the admissible values. For example, while sensitivities for a Michaelis-Menten equation go from 0 to 1, the corresponding (log)linear or lin-log representation produces values greater than 1 when we move below the operating point. On the contrary, the power-law formalism yields the same sensitivity for any value of the variables. Conversely, using a power law will necessarily mean that one needs to be careful if any concentration goes to zero, because then the whole rate term goes to zero. This can be solved by a mathematical translation of the concentration values.

Besides these limitations, the important use of these formalisms concerns system's representation. The basic representations that may be obtained obey to the following equations:

Node equations

$$\frac{dX_i}{dt} = \sum_{r=1}^p \mu_{ir} v_r \quad (11)$$

Aggregated node equations

$$\frac{dX_i}{dt} = \sum_{r=1}^p \mu_{ir}^+ v_r - \sum_{r=1}^p \mu_{ir}^- v_r \quad (12)$$

where μ_{ir} are stoichiometric coefficients. In each case, substitution of v_r by their approximations will lead to the corresponding model. In the case of the power-law formalism, the first strategy leads to Generalized Mass Action models (GMA), and the second to S-system models. Although no formal development has been made for the rest of formalisms, the same procedure can be applied.

One advantage of the S-system form is that it has an explicit steady-state solution, which greatly facilitates the analysis of alternative system design. Curiously, an analogous of the GMA representation using the (log)linear and lin-log formalisms produces exactly the same solution as the S-System. The lin-log formalism has the advantage of a close form solution for dynamic curves. It is unclear how accurate these solutions are in general. Use of the SC formalism yields models that cannot be algebraically analyzed. In all cases, stability and parameter sensitivity can be obtained.

4. SUGGESTIONS FOR SELECTING AN APPROPRIATED FORMALISM

Mathematical models are required for dealing with complexity. However, the perception of their usefulness and purposeless greatly varies depending of the goals. The selection of a particular formalism would depend on the goals. Based on experience, we may suggest the following practical rules:

Fit experimental data. This is a very common goal that makes people use models. In many cases, this may be a quite descriptive

use without further expectation. If a mathematical model fits the data appropriately, then the conceptual scheme used to derive that model represents a possible explanation of the mechanisms that generate the data. So far, any model based on these formalisms could be used. However, some technical issues may preclude the use of some of them, especially when parameters must be estimated from dynamic data [5]. Power-law models appear as more appropriate than (log)linear and Lin-log models for this task. Although no results have been obtained with the SC formalism, the flexibility of this representation may be an advantage for fitting this kind of data.

Reconstruct and identify the topology of 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, and metabolomic data [6]. Mathematical models 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. Mathematical representations used for this type of model building should be easy to manipulate in a systematic and automated way regarding the inclusion/exclusion of alternative processes and metabolites. Any of the considered nonlinear approximations may be used for this purpose.

Identify design principles in gene circuits, metabolic pathways, etc. This goal requires being able of analyzing class of systems to establish their fundamental properties and design principles. The exploration of design principles by the method of Controlled Mathematical Comparisons is greatly facilitated by using a formalism that provides a symbolic steady-state solution. Thus, the S-system form of the power-law formalism should be the method of election. Once the design principles are derived, the SC formalism can help in exploring the quantitative aspects of these principles. Basically, the SC introduces the degree of saturation as a parameter. This may be useful in identifying their impact on system's behavior, beyond the general predictions of the power-law model.

Optimize specific properties of a system. This is an important problem in many biotechnological applications that aim at optimizing the production of specific metabolites. The ultimate goal of metabolic engineering is precisely to be able of manipulating cell metabolism to obtain specific products by capitalizing on modified cell physiology. Power-law models, either in S-system or GMA forms, are specially suited for optimization. Recently, the method of Geometric Programming (GP) has been used for optimizing GMA models. Specific methods can be developed for the other formalisms, although the requirements of the optimization techniques favor the use of the power-law solution. Through recasting into power-law, models based on the SC formalism can also be optimized using GP. This opens an interesting possibility as the SC formalism incorporates saturation and cooperativity in a systematic way.

Model exchange. Models based on approximated representations greatly facilitate model exchange. In all cases, models can be generated in a systematic way from conceptual maps. This is a valuable advantage for automatic generation of models. Any of these nonlinear approximations could be used and interchange of models between formalisms can be automatically performed.

5. CONCLUSIONS

The use of approximate kinetic representations facilitates modeling and analysis of complex metabolic networks. Models based on the power-law formalism are specially suited for dealing with most of the problems that leads to use mathematical models in Systems Biology. Although other alternatives such as (log)linear and lin-log models can be used, no clear advantage has been demonstrated that may suggest use these models instead of power-law based models. On the contrary, technical limitations related to the properties of the (log)linear and lin-log representations suggest using other alternatives. The SC formalism provides an extension of the power-law formalism that can be useful for complementing some analysis.

6. ACKNOWLEDGMENTS

Grants BFU2005-0234BMC and BFU2007-62772/BMC from the Spanish Ministerio de Educación y Ciencia supported this work. RA is a Ramon y Cajal awardee.

7. REFERENCES

- [1] Savageau MA. (1969) Biochemical systems analysis. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* 25(3):365-9.
- [2] Voit, E.O. (2000) *Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists.* Cambridge University Press, Cambridge, U.K.
- [3] Heijnen JJ. (2005) Approximate kinetic formats used in metabolic network modeling. *Biotechnol Bioeng* 91(5):534-45.
- [4] Vilaprinyo, E.; Alves, R.; 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.
- [5] Vilela M, Borges CC, Vinga S, Vasconcelos AT, Santos H, Voit EO, Almeida JS. (2007) Automated smoother for the numerical decoupling of dynamics models. *BMC Bioinformatics.* 8:305.
- [6] Alves, R.; Sorribas, A. (2006) In silico pathway reconstruction: Iron-sulfur cluster biogenesis in *Saccharomyces cerevisiae*. *BMC Syst Biol* 31;1:10.