

**Dagstuhl Seminar "Integrative Bioinformatics"**

**Challenges for the identification of biological systems from *in vivo* time series data**

**Eberhard O. Voit<sup>1,2,\*</sup>, Simeone Marino<sup>3</sup> and Raman Lall<sup>2</sup>**

<sup>1</sup> The Wallace H. Coulter Department of Biomedical Engineering at Georgia Institute of Technology and Emory University  
313 Ferst Drive, Atlanta, GA 30332, USA  
Email: [Eberhard.Voit@bme.gatech.edu](mailto:Eberhard.Voit@bme.gatech.edu)

<sup>2</sup> Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina  
303K Cannon Place, 135 Cannon Street, Charleston, SC 29425, USA

<sup>3</sup> Department of Microbiology and Immunology, 6730 Medical Science Bldg. II  
The University of Michigan, Ann Arbor, MI 48109-0620

\* Corresponding author

**Edited by E. Wingender; received September 29, 2004; revised and accepted November 17, 2004; published December 23, 2004**

---

## Abstract

Modern methods of high-throughput molecular biology render it possible to generate time series of metabolite concentrations and the expression of genes and proteins *in vivo*. These time profiles contain valuable information about the structure and dynamics of the underlying biological system. This information is implicit and its extraction is a challenging but ultimately very rewarding task for the mathematical modeler. Using a well-suited modeling framework, such as Biochemical Systems Theory (BST), it is possible to formulate the extraction of information as an inverse problem that in principle may be solved with a genetic algorithm or nonlinear regression. However, two types of issues associated with this inverse problem make the extraction task difficult. One type pertains to the algorithmic difficulties encountered in nonlinear regressions with moderate and large systems. The other type is of an entirely different nature. It is a consequence of assumptions that are often taken for granted in the design and analysis of mathematical models of biological systems and that need to be revisited in the context of inverse analyses. The article describes the extraction process and some of its challenges and proposes partial solutions.

*Keywords:* Biochemical Systems Theory, metabolic profile, network identification, pathway analysis, proteomics, S-system, time series

---

## Introduction

The rapid development of systems biology over the past few years has been driven by the advancement of experimental methods that generate high-throughput data characterizing the expression of genomes, activity patterns of proteins, and the simultaneous profiles of select groups of metabolite concentrations *in vivo*. These data usually provide snapshots of a biological system, which may be compared with corresponding snapshots of an alternative system, for instance, with the goal of assessing the differences between stressed and unstressed cells or of associating a disease pattern with specific physiological markers not present in healthy subjects. Some of the methods are furthermore suitable for obtaining *in vivo* measurements in short time sequence, which allows the generation of dynamic profiles. The repertoire of experimental high-throughput tools presently includes nuclear magnetic resonance [e. g., Szyperki, 1998; Neves *et al.*, 2002], mass spectrometry [e. g., Goodenowe, 2003], 2-d gels [Gerner *et al.*, 2002], different types of DNA arrays [e. g., Gupta and Oliver, 2003; Snijders *et al.*, 2003], tissue array analysis [Alizadeh *et al.*, 2001; Simon *et al.*, 2004], transfection arrays [Tominaga, 2004], and protein kinase phosphorylation [Mckenzie and Strauss, 2003]. It is reasonable to expect that many other innovative ideas will provide additional high-throughput tools in the near future.

It is also merely a matter of time before some of these methods will be extended from small numbers of simultaneous measurements and relatively coarse time series to dense time sequences of many concentration or expression values. The resulting time series data will provide unprecedented insights in the functioning of cells *in vivo*. However, gaining these insights will require that methods be available for appropriately analyzing time series data with dynamical models and thus for extracting information on the structure and regulation of biological systems. This task of information extraction will be highly rewarding and require two components beyond the generation of good data. One is the identification of a suitable modeling framework and the other is the establishment and fine-tuning of theory and algorithms supporting the extraction and interpretation of information from time series.

Several articles have discussed in detail why *Biochemical Systems Theory* (BST) [Savageau, 1969a; 1969b; 1976; Voit, 2000; Torres and Voit, 2002] is a

very promising candidate for pathway identification tasks based on time series data [e. g., Voit and Savageau, 1982; Voit, 2000; Maki *et al.*, 2002; Kikuchi *et al.*, 2003; Voit and Almeida, 2004], and we will not repeat these arguments here. Instead, we will only briefly review features of this theory that are particularly pertinent in the present context and place our primary focus on some of the challenges that biological system identification from time profiles harbors.

## Relationships between network structure and S-system parameters

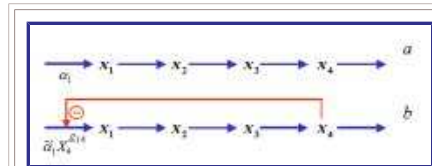
BST comes in two variants that offer alternative options for representing biological systems, the *Generalized Mass Action* and the *S-system* forms [e. g., Shiraishi and Savageau, 1992]. For the identification of structure from time series data, the S-system variant seems particularly useful, especially if not much additional information about the biological system is available. In this representation, the rate of change in each pool (variable) is represented as the difference between influx into the pool and efflux out of the pool. Each flux term is summarily approximated by a product of power-law functions, so that the generic form of any S-system model is

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^n X_j^{g_{ij}} - \beta_i \prod_{j=1}^n X_j^{h_{ij}}, \quad i=1, \dots, n, \quad \alpha_i, \beta_i > 0; g_{ij}, h_{ij} \in \mathbb{R}, \quad (1)$$

where  $n$  is the number of state variables. The exponents  $g_{ij}$  and  $h_{ij}$  are called kinetic orders and quantify the effect of  $X_j$  on the production or degradation of  $X_i$ , respectively. A kinetic order of zero implies that the corresponding variable  $X_j$  does not have an effect on  $X_i$ . If the kinetic order is positive, the effect is activating or augmenting, and if it is negative, the effect is inhibiting. The multipliers  $\alpha_i$  and  $\beta_i$  are rate constants that quantify the turnover rate of the production or degradation, respectively. It is noted that the S-system formulation allows the inclusion of  $m$  independent variables, so that the products in Eq. (1) extend to  $n + m$ . Because these independent variables remain constant during each analysis, they can be absorbed into the rate constants, and it is sufficient for our purposes here to consider the form in Eq. (1).

As an example, consider the two simple pathways in Fig. 1, which differ in the feedback inhibition exerted by end product  $X_4$ . BST associates this inhibition exclusively with the parameter  $g_{14}$ . Because this parameter represents an inhibitory effect, its value is negative, and the magnitude of this value indicates the strength of inhibition. Strong inhibition may be modeled, e. g., with  $g_{14} = -2$ , while weak inhibition corresponds to a negative value close to zero. If inhibition disappears altogether, the value becomes zero, and both the graph in Fig. 1b and the corresponding production term of  $X_1$  become those in Fig. 1a, which is thus a special case of the system in Fig. 1b. This example indicates that structural features of a system are mapped onto S-system parameters in a unique fashion: if  $X_4$  inhibits the production of  $X_1$ ,  $g_{14}$  is negative; if it does not,  $g_{14}$  equals 0. Reversely, if all parameters of a symbolic S-system model can be estimated from (time series) data, their values may be interpreted as structural features of the biological system.

Several articles have appeared in the recent literature describing computational strategies, using S-systems, for this "inverse problem" of extracting information from time series data [e. g., Voit and Savageau, 1982; Maki *et al.*, 2002; Kikuchi *et al.*, 2003; Voit and Almeida, 2004; Veflingstad *et al.*, 2004; Wang, 2004]. These articles primarily emphasized algorithmic issues, which are challenging and increase exponentially in difficulty with the size of the problem. However, algorithmic advances alone will not be sufficient to identify large systems from dynamic data. Recognizing this, Almeida and Voit, 2003, suggested making maximal use of other *a priori* biological information that might be available in addition to the time series data. For instance, the lack of any direct interaction between two variables allows the immediate elimination of all corresponding kinetic orders before the estimation begins.



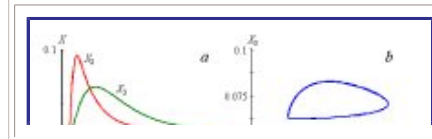
**Figure 1:** Two linear pathways that differ in a feedback signal with which  $X_4$  inhibits the production of  $X_1$ . In pathway  $a$ , the production is constant at a rate  $\alpha_1$ . With inhibition, the production becomes dependent on  $X_4$ , and this dependence is quantified by the parameter  $g_{14}$ , which is negative and has a magnitude that directly reflects the strength of inhibition. The rate constant may also be different, because it reflects the unit of  $X_4$ . If the inhibition decreases,  $g_{14}$  approaches 0, and in the limit, pathways  $a$  and  $b$  along with the representations of the production of  $X_1$  become equivalent, demonstrating that the S-system model for pathway  $a$  is always a numerical special case of the model for pathway  $b$ .

While the use of information beyond the time series data themselves seems to be a very beneficial, if not crucial component of dealing with the complexities of the inverse problems at hand, caution and re-evaluation of some common modeling ideas is needed. We describe some of these here. Specifically, we revisit some tenets of model design that are generally taken for granted in biological systems modeling, but may need to be relaxed when *in vivo* time series data are to be analyzed.

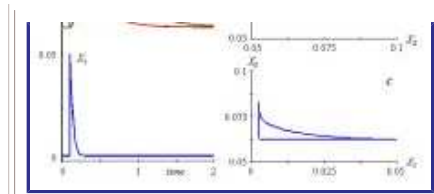
## Effects of model assumptions on structure identification

### Direct versus indirect effects

The generic strategy for designing systems models calls for setting up one equation for each dependent (state) variable and to include in this equation those and only those variables that directly affect this variable. For instance, if  $X_1$  is converted into  $X_2$  and then  $X_3$ , the equation describing the dynamics of  $X_3$  contains  $X_2$  but not  $X_1$ . This strategy is justified by the argument that the effect of  $X_1$  on  $X_3$  is "automatically" mediated through the dynamics of  $X_2$ . Nevertheless, indirect effects can create problems for inverse tasks of parameter and structure identification. After all,  $X_1$  does affect  $X_3$ , even if only indirectly, as is demonstrated in Figure 2, where a bolus increase in  $X_1$  leads to subsequent, staggered rises and falls in  $X_2$  and  $X_3$  (Fig. 2a). The phase-plane plot of  $X_3$  versus  $X_2$  shows the direct substrate-product relationship between the two variables (Fig. 2b), but the analogous plot of  $X_3$  versus  $X_1$  also shows dependence (Fig. 2c). As a consequence, if the pathway structure is to be determined from time series data, one may incorrectly identify  $g_{31} > 0$  and thus deduce a direct effect of  $X_1$  on  $X_3$ .



**Figure 2:** According to conventional wisdom of modeling linear pathways, in which  $X_1$  is converted into  $X_2$  and then  $X_3$ , the equation for  $X_3$  does not include  $X_1$ , because its effect is indirect. Nonetheless,  $X_1$  does affect  $X_3$ ,



as is shown in the time courses (a) and the phase-plane plots on the right-hand side (b and c). The effect of  $X_1$  (c) has entirely different characteristics than that exerted by  $X_2$  (b) and is somewhat less pronounced.

(Numerical model used for this example:  $\dot{X}_1 = 0.25 - 5X_1^{0.5}$ ,  $X_1(0) = 0.0025$ ;

$\dot{X}_2 = 5X_1^{0.5} - 4X_2$ ,  $X_2(0) = 0.0625$ ;  $\dot{X}_3 = 4X_2 - 2X_3^{0.75}$ ,  $X_3(0) = 0.0625$ . At time  $t = 0.1$ ,  $X_1$  is externally set to 0.05.)

In theory, the fitting of the system of differential equations should identify  $g_{31}$  as not significantly different from zero, especially if non-zero kinetic orders are penalized in the estimation process [Kikuchi *et al.*, 2003; Voit and Almeida, 2003], but practical applications to noisy data show that this is not always the case. For simple linear pathways, Vance *et al.*, 2002, and Torralba *et al.*, 2003, devised methods for inferring direct and indirect causality, but these methods are not likely to be applicable to more complex systems and need to be complemented with additional methods. Kikuchi *et al.*, 2003, and others somewhat alleviated the problem by requiring many data sets that described responses to different systematic perturbations, but one has to ask how often such complete data are available. While the consideration of structural parsimony, implemented as a penalty for non-zero kinetic orders [Kikuchi *et al.*, 2003; Voit and Almeida, 2003], or a model construction scheme with minimal numbers of parameters [Marino and Voit, 2004], alleviates the issues to some degree, it seems that the problem is in general unsolved at this point in time.

### Precursor-product relationships

If within an unbranched pathway  $X_i$  is converted directly into  $X_j$ , the loss from  $X_i$  is equal to the production of  $X_j$ , a feature that is called *precursor-product relationship*. The relationship mandates that the structure and parameter values of the loss function in the equation of  $X_i$  equal those of the production function in the equation of  $X_j$ . For the inverse procedure of deducing parameter values from measured metabolic profiles one is therefore inclined to equate the two functions beforehand and to estimate only one common set of parameter values. Under ideal conditions, this assumption is certainly justified. However, real pathways may appear to be linear in idealized representations but are not always truly unbranched *in vivo*. Instead, secondary pathways may divert material from the primary pathway or add material that is produced in a converging path. For instance, a typical glycolysis model assumes that glucose is converted into glucose 6-phosphate (G6P) and that this step is the only source of G6P. In reality there are several branches leading to G6P, such as the reversible phosphoglucosmutase reaction between G6P and G1P. Even if such branches have relatively small fluxes in comparison to the path of interest, they compromise the precursor-product relationships that were taken for granted based on the idealized model. Keeping the parameters of the degradation function of  $X_i$  and the production function of  $X_j$  separate solves the problem in principle: If the parameter values turn out to be the same (within the accuracy of the estimation procedure and the noise in the data), one may conclude that there are no significant branches (or, less likely, that additional incoming and outgoing branches cancel each other). For reasons of practicality and computational efficiency, however, any unnecessary increase in the number of parameters to be estimated leads to undesirable redundancies and should be avoided where possible.

For the same reasons, one has to ask to what degree reversibility of reactions needs to be considered. In principle, very few reactions are truly irreversible, but allowing for the potential of reversibility adds to the system several parameters that need to be estimated [cf. Sorribas and Savageau, 1989].

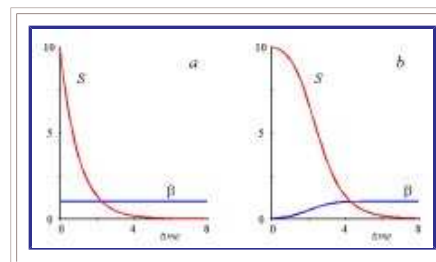
### Intermediate steps and time delays

In some cases one might not be interested in intermediate metabolites of a linear unbranched pathway. For instance, [Curto \*et al.\*, 1998](#), reduced the first ten reaction steps of purine metabolism [cf. [Stanbury \*et al.\*, 1983](#): chapter 50] to a single step representing the overall conversion of phosphoribosylpyrophosphate into inosine monophosphate. For steady-state analyses, such a reduction is inconsequential because the influx, efflux and all intermediate fluxes are equivalent and all intermediate metabolites are in balance. Even dynamic analyses are often only mildly affected as long as there are no branches diverging from or converging to the linear pathway. However, the intermediate steps do lead to some time delay and to a distortion of any signals flowing through the pathway, and these effects may be significant enough to cause difficulties in the estimation of parameter values from *in vivo* time series data. The same is true for other non-enzymatic reactions that affect a metabolic system, such as transport through membranes and transcription and translation.

One should note that the exclusion of intermediate steps, while reducing the number of parameters, is not always desirable. For instance, one might be interested in maximizing the production of certain intermediate metabolites in a linear pathway due to their industrial importance. As an example, diacetyl is a valuable by-product of pyruvate metabolism in *Lactococcus lactis*, and there is great industrial interest in developing metabolic strategies for enhanced production [[Neves, 2001](#)]. The process involves multiple reactions in series, and it is necessary for a useful model to include these steps so that it is possible to maximize the selectivity of diacetyl with respect to other, less desirable metabolites.

### Homogeneity of populations

Most biochemical *in vivo* experiments are executed with populations of cells, and it is commonly assumed that a set of average parameter values is sufficient for modeling their responses. This may or may not be a valid assumption. If cells differ significantly in their numerical characteristics, observed and modeled time courses may differ, thereby complicating the inverse problem of identifying structure from time data. As an example, consider the uptake of a bolus of substrate from the surrounding medium. Because the concentration of substrate is highest at the beginning and subsequently decreases due to uptake by the cells, the time course of substrate concentration should be expected to show the steepest decrease at the beginning. However, one often observes an s-shaped decrease [e. g., [Ramos \*et al.\*, 2004](#)], which is irreconcilably incompatible with the model. The discrepancy must therefore be due to other mechanisms that are not considered in the model. For instance, an s-shaped uptake characteristic is obtained if one supposes that the cells begin taking up substrate with different response times. This is illustrated in [Fig. 3](#) where substrate uptake according to a default linear transport model is compared to a model where the uptake speed is normally distributed among cells.



**Figure 3:** Dynamics of substrate  $S$  in medium due to uptake by cells. In panel a, all cells begin taking up substrate immediately with the same constant rate  $\beta$ , and the decrease in  $S$  is exponential. In panel b, cells begin taking up substrate at different times. If the initiation of uptake is normally distributed ( $\mu = 2$ ,  $\sigma = 1$ ) among cells, the result is a total uptake rate  $\beta$  that is sigmoid, and the consequence is a corresponding sigmoid disappearance of substrate from the medium.

---

### Dynamic changes in independent variables

One typically assumes that enzymes and other independent variables remain constant during a model analysis. However, if the observed data consist of time series beginning with some perturbation and extending over a time period of half an hour or more, it is possible that the perturbation triggers changes in gene expression that alter some of the enzyme activities relevant to the pathway of interest. If so, the dynamics of the activities is inconsistent with the model assumptions, thereby potentially creating difficulties during the fitting of dynamic metabolite data.

---

### Exploration of the validity of model assumptions through decoupling and data smoothing

The vignettes above indicate that idealized models may experience difficulties capturing *in vivo* phenomena exactly. This conclusion is of course nothing new *per se*, because the definition of a model entails that not all discrepancies between a model and reality can ever be resolved. However, the discrepancies caused by simplification and abstraction appear in a new disguise and, ironically, become more blatant the better the data are. In this light it is useful to develop strategies that at least allow the exploration of these discrepancies. For instance, it would be beneficial to answer questions like 'If we could model substrate uptake so that it captured the observations well, would that eliminate other discrepancies between model responses and observed time series or not?' We show in the following that decoupling of equations and the use of smoothed time traces provides such a tool.

Several authors have shown that it is possible to decouple systems of differential equations for purposes of parameter estimation. At first this may seem to violate the integrity of a model system and its mathematical representation, but the procedure has been proven valid if applied correctly [Voit and Almeida, 2004]. In some cases the decoupling was accomplished by estimating slopes of all ( $n$ ) time courses at many ( $N$ ) time points, which reduced the system of  $n$  coupled differential equations to  $n \times N$  algebraic equations [Voit and Savageau, 1982; Voit, 2000: chapter 5; Voit and Almeida, 2004]. In a different implementation, Maki *et al.*, 2002, estimated the parameters of one differential equation at a time, using for the time courses of other variables the raw data or a smoothed analogue.

Another application of decoupling systems of differential equations, also inspired by the need for computational simplification, targeted the numerical calculation of non-steady-state ion transfer caused by diffusion, migration and convection in uni-dimensional electrochemical systems [Volgin *et al.*, 2003]. In this case the authors showed that a set of coupled continuity equations could be solved in the form of independent difference transfer equations for each electrolyte species. Similarly, to simulate the activation dynamics in anisotropic cardiac tissues more efficiently, Clements *et al.*, 2004, decoupled a system of elliptic-parabolic partial differential equations, which resulted in a computationally more convenient reaction-diffusion equation.

Whereas decoupling in these examples has been used to ameliorate computational difficulties, the same type of decoupling may also be used for testing hypotheses regarding the validity of implicit or explicit model assumptions. The generic situation is that some observed time course, say for  $X_k$ , is not represented well with the model and that the discrepancies cannot be resolved through valid parameter adjustments. The idea is then to replace this model equation with the actual data, which in the terminology of differential equations corresponds to considering the data set as forcing function. This

replacement of equations may be implemented with data in either raw or some smoothed form and allows for various options. At one extreme, one may replace all equations with the observed data except for the equation for  $X_k$ . Solving this system shows whether the equation for  $X_k$  in itself is appropriate and the observed discrepancies are due to other components of the model. At the other extreme, one may be unable to model some metabolite appropriately and replace its system equation with the measured time profile as input. One example for this situation may be a transport or substrate uptake process that is more complex than the model allows, as it was described above. Another example is the modeling of ubiquitous metabolites like ATP,  $P_i$ , and NADH. These metabolites are involved in dozens if not hundreds of equations *in vivo*, and it is presently not feasible to model them all. If their time profile has been measured along with other components of the system, the data may be taken as inputs at face value or in a smoothed fashion, thereby allowing explorations of the rest of the system. Thus, by replacing one, several, or all equations except for one with actual data or smoothed analogues, it is possible to explore different types of scenarios that are otherwise very difficult to disentangle.

Ultimately the most important question is of course how well the model is able to predict situations that had not been used for model identification or data fitting, and one would consider a model as more valuable the further it can be extrapolated toward untested conditions, such as changes in substrate concentrations, pH, or oxygen conditions in the surrounding medium. In general, the predictive power of a model can only be assessed by comparison with experimental data. However, as [Martin, 1997](#), demonstrated for a somewhat similar situation in forestry research, if data are available on several related conditions, it is possible to design a model that interpolates validly between these scenarios, thereby facilitating a more comprehensive set of simulation analyses.

---

## Conclusions

Time series data contain enormous information about the system from which they were observed. For the modeler, such data pose a very rewarding, yet difficult challenge. The difficulties of extracting information fall into two categories. First, the computational issues of estimating parameters in large nonlinear systems are daunting, and while several groups have addressed different aspects of the estimation process, they are still far from being solved in general. The second category of challenges is due to the well-recognized fact that models are more or less crude abstractions of reality. In the given context of structure identification, an important consequence is that it is not a priori clear how closely a given model is able to capture observed data even under the best conditions, that is, with optimal parameter values. This question cannot be answered in generic terms, but it was shown here where discrepancies between model and data may emerge and how to explore them. The strategy of decoupling systems of equations provides a tool for such explorations. Originally used to ease some of the computational issues of parameter estimation, decoupling was shown here to provide a modeling tool that allows the investigator to pinpoint where discrepancies might originate. The decoupling and replacement of equations with raw data permits the focused analysis of the effects of a single variable within a system that is taken as observed and no longer modeled. Alternatively, the decoupling and replacement strategy allows partial modeling of a system in the presence of poor data fits that are caused by one or a few system components that are simply not adequately captured by the model. Decoupling and replacement of equations does not resolve the root cause of a poor fit, but it does aid the researcher in pinpointing and diagnosing the topological location of a problematic component within the integrated system.

---

## Acknowledgements



This work was supported in part by a Quantitative Systems Biotechnology grant (BES-0120288; E.O. Voit, PI) from the National Science Foundation and a National Heart, Lung and Blood Institute Proteomics Initiative (Contract N01-HV-28181; D. Knapp, PI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring institutions.

---

## References

- 
- Alizadeh, A. A., Ross, D. T., Perou, C. M. and van de Rijn, M. (2001). Towards a novel classification of human malignancies based on gene expression pattern. *J. Pathol.* **195**, 41-52.

---

  - Almeida, J. S. and Voit, E. O. (2003). Neural-network-based parameter estimation in complex biomedical systems. *Genome Informatics* **14**, 114-123.

---

  - Clements, J. C., Nenonen, J., Li, P. K. and Horacek, B. M. (2004). Activation dynamics in anisotropic cardiac tissue via decoupling. *Ann. Biomed. Eng.* **32**, 984-990.

---

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

---

  - Gerner, C., Vejda, S., Gelbmann, D., Bayer, E., Gotzman, J., Schulte-Hermann, R. and Mikulits, W. (2002). Concomitant determination of absolute values of cellular protein amounts, synthesis rates, and turnover rates by quantitative proteome profiling. *Mol. Cell. Proteomics* **1**, 528-537.

---

  - Goodenowe, D. (2003). Metabolomic analysis with fourier transform ion cyclotron resonance mass spectrometry. *In: Metabolic Profiling: Its role in Biomarker Discovery and Gene Function Analysis*, Goodacre, R., Harrigan, G. G. (eds.), Dordrecht, The Netherlands: Kluwer Academic Publishing, pp. 125-139.

---

  - Gupta, V. and Oliver B. (2003). *Drosophila* microarray platforms. *Brief. Funct. Genomics Proteomics* **2**, 97-105.

---

  - Kikuchi, S., Tominaga, D., Arita, M., Takahashi, K. and Tomita, M, (2003). Dynamic modeling of genetic networks using genetic algorithm and S-system. *Bioinformatics* **19**, 643-650.

---

  - Maki, Y., Ueda, T., Okamoto, M., Uematsu, N., Inamura, Y. and Eguchi Y. (2002). Inference of genetic network using the expression profile time course data of mouse P19 cells. *Genome Informatics* **13**, 382-383.

---

  - Marino, S. and Voit, E. O. (2004). An automated procedure for information extraction with S-systems. International Conference on Molecular Systems Biology (ICMSB'04), Tahoe, CA, August 21-25 (Abstract).

---

  - Martin, P.-G. (1997). The use of canonical S-system modelling for condensation of complex dynamic models. *Ecol. Model.* **103**, 43-70.

---

  - Mckenzie, J. A. and Strauss, P. R. (2003). A quantitative method for measuring protein phosphorylation. *Anal. Biochem.* **313**, 9-16.

---

  - Neves, A. R. (2001). Metabolic Strategies to Reroute Carbon Fluxes in *Lactococcus lactis*, Ph.D. thesis, Universidade Nova de Lisboa, Oeiras, Portugal.
-

- Neves, A. R., Ventura, R., Mansour, N., Shearman, C., Gasson, M. J., Maycock, C., Ramos, A. and Santos H. (2002). Is the glycolytic flux in *Lactococcus lactis* primarily controlled by the redox charge? *J. Biol. Chem.* **277**, 28088-28098.
- Ramos, A., Neves, A. R., Ventura, R., Maycock, C., Lopez, P. and Santos, H. (2004). Effect of pyruvate kinase overproduction on glucose metabolism of *Lactococcus lactis*. *Microbiology* **150**, 1103-1111.
- Savageau, M. A. (1969a). Biochemical systems analysis. 1. 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. 2. The steady-state solutions for an n-pool system using a power law approximation. *J. theor. Biol.* **25**, 370-379.
- Savageau, M. A. (1976). *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, Reading, Mass.
- Shiraishi, F. and Savageau, M. A. (1992). The tricarboxylic acid cycle in *Dictyostelium discoideum*. I. Formulation of alternative kinetic representations. *J. Biol. Chem.* **267**, 22912-22918.
- Simon, R., Mirlacher, M. and Sauter, G. (2004). Tissue microarrays. *Biotechniques* **36**, 98-105.
- Snijders, A. M., Pinkel, D. and Albertson, D. G. (2003). Current status and future prospects of array-based comparative genomic hybridisation. *Brief. Funct. Genomics Proteomics* **2**, 37-45.
- Sorribas, A. and Savageau, M. A. (1989). Strategies for representing metabolic pathways within biochemical systems theory: Reversible Pathways. *Math. Biosci.* **94**, 239-269.
- Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L. and Brown, M. S. (1983). *The Metabolic Basis of Inherited Disease* (5th edition). McGraw-Hill Book Company, Auckland, New York.
- Szyperski, T. (1998).  $^{13}\text{C}$ -NMR, MS and metabolic flux balancing in biotechnology research. *Quart. Rev. Biophys.* **31**, 41-106.
- Tominaga, D. (2004). Analysis system for gene expression time-series data. International Conference on Molecular Systems Biology (ICMSB'04), Tahoe, CA, August 21-25 (Abstract).
- Torralba, A. S., Yu, K., Shen, P., Oefner, P. J. and Ross, J. (2003). Experimental test of a method for determining causal connectivities of species in reactions. *Proc. Natl. Acad. Sci. USA* **100**, 1494-1498.
- Torres, N. V. and Voit, E. O. (2002). *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press, Cambridge, UK.
- Vance, W., Arkin, A. and Ross, J. (2002). Determination of causal connectivities of species in reaction networks. *Proc. Natl. Acad. Sci. USA* **99**, 5816-5821.
- Veflingstad, S. R., Almeida, J. S. and Voit, E. O. (2004). Priming nonlinear searches for pathway identification. *Theor. Biol. Med. Model.* **1**, 8.
- Voit, E. O. (2000). *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, Cambridge, United Kingdom.

- Voit, E. O. and Almeida, J. (2003). Dynamic Profiling and Canonical Modeling: Powerful Partners in Metabolic Pathway Identification. *In: Metabolite Profiling: Its Role in Biomarker Discovery and Gene Function Analysis*, R. Goodacre and G.G. Harrigan (eds.), Kluwer Academic Publishing, Dordrecht, The Netherlands, pp. 257-276.

---

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

---

- Voit, E. O. and Savageau, M. A. (1982). Power-law approach to modeling biological systems; III. Methods of analysis. *J. Ferment. Technol.* **60**, 233-241.

---

- [Volgin, V. M., Volgina, O. V. and Davydov, A. D. \(2003\). Finite difference method of simulation of non-steady-state ion transfer in electrochemical systems with allowance for migration. \*Comput Biol Chem.\* \*\*27\*\*, 185-196.](#)

---

- Wang, F.-S. (2004). Dynamic modeling of biological systems using global optimization. *Bioinformatics in Taiwan Symposium*, National Yang Ming University, Taipei, Taiwan, September 9-11 (Abstract).