COARSE BUT EFFICIENT IDENTIFICATION OF METABOLIC PATHWAY SYSTEMS

IWATA M.[1], SHIRAISHI F.[1] AND VOIT E.O.[2]*

[1]Section of Bio-process design, Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, 6-10-1, Hakozaki, Higashi-Ku, Fukuoka 820-8581, Japan.
[2]The Wallace H. Coulter, Department of Biomedical Engineering at Georgia Institute of Technology and Emory University, 313 Ferst Drive, Suite 4103, Atlanta, GA 30332-0535, U.S.A.
*Corresponding Author: Email- eberhard.voit@bme.gatech.edu

Abstract- The rapid development of high-throughput experiments in molecular biology over the past decade has facilitated the generation of metabolic datasets of unprecedented quantity and quality. These datasets contain enormous information that must be extracted with computational methods. A premier goal of such an extraction is the construction of fully parameterized kinetic models of metabolic pathway systems, which may subsequently be utilized for deepening our understanding of metabolism and for the design of manipulation and optimization strategies toward the production of valuable organics. Numerous methods have been proposed for the conversion of metabolic data into such models, but the nonlinear nature of regulated pathway systems continues to be a challenge for the estimation of reliable model parameters. To alleviate the situation, the present work proposes a coarse yet efficient technique for the estimation of parameter values that is based on linear regression and therefore naturally scales to large systems. The proposed method permits the complete construction of a coarse model, the identification of a starting point for fine-tuned nonlinear estimation, or the characterization of a few fluxes that can subsequently be used for performing Dynamic Flux Estimation (DFE), a method that yields insights into the true nature of the processes that govern metabolic pathway systems. The proposed linear inference method is demonstrated with a regulated branched pathway and with a metabolic system from the literature that describes the biosynthesis of amino acids from aspartate. Both analyses indicate that the method yields satisfactory results .

Keywords- Biochemical Systems Theory (BST), Dynamic Flux Estimation (DFE), Metabolic pathway analysis, Parameter estimation, S-system, Time course data

Citation: Iwata M., Shiraishi F. and Voit E.O. (2013) Coarse but Efficient Identification of Metabolic Pathway Systems. International Journal of Systems Biology, ISSN: 0975-2900 & E-ISSN: 0975-9204, Volume 4, Issue 1, pp.-57-72.

Introduction

The performance of high-throughput analytical instruments has been enhanced so much in recent years that it is becoming feasible to measure comprehensive time course data of metabolite concentrations, in some cases even in living cells [1-5]. The resulting data are very valuable, as they portray the collective functionality of cellular metabolism in unprecedented detail. At the same time, the data in themselves only provide limited insight and mandate the development of corresponding theoretical advances and computer-aided techniques that extract as much information from the metabolic time series as possible. The reward for such developments is substantial: if valid mathematical models can be constructed from experimental time series data, it may become possible to explore with computer simulations of metabolic pathway systems what is actually happening in cells, for instance, in response to some environmental demand or stress [6-12].

The true format of optimal mathematical representations of metabolism is unknown. This fact poses a great challenge, because about any parameter estimation or simulation technique assumes some functional form within which the parameters are optimally adjusted

or the dynamics of a system is computed. Dynamic Flux Estimation (DFE) [13] is an attempt to characterize processes within biological systems with minimal prior knowledge. While intriguing in principle, DFE requires a complete set of time series data and, more limiting, systems that contain as many independent fluxes as system variables. This situation is seldom given, thus requiring the augmentation of DFE with additional *a priori* knowledge or assumptions regarding a subset of fluxes.

Power-law equations within the modeling framework of Biochemical Systems Theory (BST) [14-20] have proven to be useful defaults for the construction of models for metabolic reaction networks. Three reasons for their utility are that BST models are based on rigorous approximation theory, that they can easily be constructed directly from the topology of a metabolic map, and that each parameter in these models has a uniquely defined meaning and interpretation, which facilitates the moving back and forth between biochemistry and mathematics. Whereas it is easy to set up a BST model for a pathway system in a symbolic format, the identification of optimal parameter values continues to be a significant challenge. As a consequence, numerous direct and indirect optimization methods for

BST models have been proposed in recent years; examples include [21-27]. None of them, however, is effective in all practical cases.

One partial strategy for system identification that has turned out to be quite effective is the estimation of slopes from time series data. The advantage of this step is that it converts the estimation of differential equations into an estimation task exclusively involving systems of algebraic equations. As an added benefit, this strategy permits the decoupling of differential equation models into several smaller estimation tasks that may be executed in parallel or sequentially [28-30].

In spite of numerous advances in recent years, the estimation of kinetic parameters, even for moderately sized pathway systems, is still difficult because the task is nonlinear, measured time course data always contain noise, and the most appropriate functions to represent the data are simply not known. To alleviate this problem, we propose here a coarse yet efficient estimation of parameter values that is based on linear regression and may be considered a simplification of an earlier method [21,31]. The method permits either the full construction of an initial model, under the assumption that power-law functions are reasonable approximations, or the characterization of a few fluxes that can subsequently be used to execute DFE. The latter case minimizes the assumption that the power-law form is adequate. In the following we first introduce the concepts and details of the coarse estimation technique and then evaluate this technique against data created with an artificial small system and by analyzing a metabolic pathway described in the literature. In both cases, the method, with its different variations, yields satisfactory fits.

Methods

S-system Differential Equations

BST models come in two main varieties. S-system models consist of differential equations that express the dynamics of dependent valuables (metabolite concentrations) in the following format:

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}, \qquad (i = 1,...n) \qquad (1)$$

Here, the dependent variables X_i ($i = 1, ..., n$) represent metabolite pools or concentrations, and m additional variables may be included as factors or modulators that remain unaffected by the action of the system; t is time. All influxes into each metabolite pool are collectively formulated as a single power-law term, and the analogous is done for all effluxes. The parameters α_i and β_i ($i = 1, ..., n$) are aggregate rate constants of the influx and efflux terms, respectively, while g_{ij} and h_{ij} ($i = 1, ..., n; j = 1, ..., n+m$) are kinetic orders that reflect the strength and directionality of the effect a variable has on a given influx or efflux.

In the related Generalized Mass Action (GMA) system, each influx and each efflux is represented with its own power-law term, so that each equation may contain several positive and several negative terms. The format with analogous parameter names is therefore

$$\dot{X}_i = \sum_{k=1}^{T_i} \pm \gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{ikj}}, \qquad (i = 1,...,n) \qquad (2)$$

where T_i is the number of terms in the ith equation. GMA and S-systems have been discussed extensively in the literature. A recent comprehensive review is [20].

Slope-estimation and Decoupling

It is useful to plot each set of the measured metabolite concentrations $X_i(t_k)$ at the time points t_k ($k = 1, 2, ..., K$) and to estimate the slopes of the trend lines at these time points. Each slope $S_i(t_k)$ ($k = 1, 2, ..., K$) corresponds directly to the derivative dX_i/dt at time t_k, so that the two may be substituted for parameter estimation purposes. As a result, each differential equation may be represented by a set of K algebraic equations. For instance, in the case of an S-system model, the ith equation can be substituted point-wise with the following algebraic equations:

$$S_i(t_1) = \alpha_i \prod_{j=1}^{n+m} X_j(t_1)^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j(t_1)^{h_{ij}} = V_i(t_1) - V_{-i}(t_1)$$

$$S_i(t_2) = \alpha_i \prod_{j=1}^{n+m} X_j(t_2)^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j(t_2)^{h_{ij}} = V_i(t_2) - V_{-i}(t_2)$$

$$\vdots$$

$$S_i(t_k) = \alpha_i \prod_{j=1}^{n+m} X_j(t_k)^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j(t_k)^{h_{ij}} = V_i(t_k) - V_{-i}(t_k) \qquad (3)$$

$$\vdots$$

$$S_i(t_K) = \alpha_i \prod_{j=1}^{n+m} X_j(t_K)^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j(t_K)^{h_{ij}} = V_i(t_K) - V_{-i}(t_K)$$

[28-30]. Since experimental data usually contain noise, it is beneficial to smooth the time series data. Numerous methods are available for this purpose [32-37]. Once the data are smoothed, values of $X_i(t_k)$ and $S_i(t_k)$ ($k = 1, 2, ..., K$) may be obtained from the time series not only at the time points corresponding to actual measurements, but at essentially arbitrary time points, so that more instances are available to construct the set of algebraic equations. Subsequently, these equations are used to estimate the kinetic parameters α_i, g_{ij}, β_i and h_{ij}. It is noted that this technique automatically decouples the differential equations, and that the parameter estimation maybe performed for one differential equation at a time or in parallel. In the following, we will merge independent variables, which are typically constant, with the corresponding rate constants and therefore let the products in the equations run to n rather than $n+m$. This decision does not affect the generality of our methods and results.

Parameter Estimation Through Approximation and Linear Regression

In an S-system model, the efflux from a metabolite pool at each time point t_k is given by

$$V_{-ik} = \beta_i \prod_{j=1}^{n} X_{jk}^{h_{ij}}, \qquad (k = 1,...,K), \qquad (4)$$

where $X_{jk} = X_j(t_k)$. Once the slopes $S_{ik} = S_i(t_k)$ are estimated, the corresponding equation in (3) may equivalently be written as

$$V_{-ik} = \alpha_i \prod_{j=1}^{n} X_{jk}^{g_{ij}} - S_{ik}, \qquad (k = 1,...,K) \qquad (5)$$

Suppose that the efflux in [Eq-4] only depends on the metabolite X_i itself, which is a relatively frequent occurrence. Then, [Eq-4] simplifies to [Eq-6]:

$$V_{-ik} = \beta_i X_{ik}^{h_{ii}}, \qquad (k = 1,...K) \qquad (6)$$

The value of the kinetic order h_{ii} typically falls within a small range between 0 and 1 [18,38,39], while β_i is unconstrained, except that it must be non-negative. It is known that there is a certain degree of compensation between h_{ii} and β_i, so that two power-law functions,

for instance, one with h_{ii} = 0.4 and one with h_{ii} = 0.5, can be made very similar by an adjustment of β_i, as long as the range for X_i is relatively small. Thus, a relatively coarse grid of h_{ii} and β_i covers most essential power-law behaviors, as long as very high accuracy is not required and the variable X_i does not vary too much.

For any combination of assumed grid values for h_{ii} and β_i and for observed concentration values X_{ik}, the terms V_{-ik} and S_{ik} in [Eq-5] and [Eq-6] are real numbers. Thus, substituting these numbers in [Eq-5] and taking logarithms of both sides permits the application of linear regression with the residual function

$$\theta_i = \sum_{k=1}^{K} \left(\ln(V_{-ik} + S_{ik}) - \ln \alpha_i - \sum_{j=1}^{n} g_{ij} \ln X_{jk} \right)^2 \qquad (7)$$

The result of the regression consists of optimized parameter values α_i and g_{ij}, for each set of grid values for h_{ii} and β_i.

If the degradation term contains variables other than X_i itself, [Eq-4] is used as opposed to [Eq-6], and the grid needs to be extended by one or more dimensions for the linear regression. If the procedure is used as a preliminary step for DFE, only a few fluxes have to be estimated and one may choose efflux terms with only one or two dependent variables. We will demonstrate this strategy later.

Estimation Strategies for Pathway Systems

The parameter determination may follow alternative procedures. It may be accomplished backwards, forwards, with or without constraints, or as a preliminary step for DFE. Each variant is presented separately. It is always assumed that the concentration trend lines had been smoothed and that slopes had been obtained. In the following, the methods are generally demonstrated for one equation at a time.

Backward Estimation

Considering that the degradation term, but not necessarily the production term, for a metabolite pool almost always contains the variable itself, it makes sense to estimate pathways in a backward direction. Specifically, the following steps are used for each variable X_i.

1. Determine a suitable grid for β_i and h_{ii}, such as 0.1, 0.5, 1, 5, 10, 50, 100, 500 and 1000 for β_i and 0.2, 0.4, 0.6, 0.8 and 1 for h_{ii}.

2. For each combination of grid points, perform linear regression according to [Eq-7]. The result consists of parameter sets for α_i and g_{ij}.

3. For each set of parameter values, compute the sum of squared errors

$$\text{SSE} = \sum_{k=1}^{K} (X_{ik} - \hat{X}_{ik})^2 \qquad (8)$$

between the actual data and the corresponding estimates and select the smallest SSE for the set of best parameter estimates.

If the degradation term contains variables other than X_i itself, the grid in Step 1 is extended by one or more dimensions.

Forward Estimation

Instead of using a grid for β_i and h_{ii}, one may use a grid for α_i and g_{ij}. The forward estimation strategy is preferable when information is available about influxes to the system. For instance, it may be possible to measure substrate uptake. The result of the forward estima-

tion consists of parameter sets for β_i and h_{ij} for every combination of grid values for α_i and g_{ij}. Typically, the set with the smallest SSE is selected.

Constrained Estimation

So far, the parameters of each equation were computed as if all equations were independent of each other. In reality, systems are much more constrained. For instance, the efflux out of X_2 in [Fig-1] is identical to the influx into X_3. Thus, one possible strategy of constrained estimation is to choose grid points β_3 and h_{33}, estimate the corresponding α and g parameters for the influx of X_3, select the best set and use this set as estimates for the parameters in the efflux of X_2. The influx is then estimated as before. The same strategy is used for X_4. The collective efflux of X_1 is identical to the sum of the influxes to X_2 and X_4. Thus, the estimates should satisfy the equation

$$\beta_1 X_1^{h_{11}} = \alpha_2 X_1^{g_{21}} + \alpha_4 X_1^{g_{41}} \qquad (9)$$

Due to the mathematical features of power-law functions, this equation can, in general, only be satisfied at some chosen operating point or for trivial situations, where $g_{21} = g_{41}$ or where one of the fluxes on the right is zero. At a chosen operating point, the parameters for the efflux of X_1 are related to those of the two influxes by the equation

$$h_{11} = \frac{g_{21}\alpha_2 X_1^{g_{21}} + g_{41}\alpha_4 X_1^{g_{41}}}{\alpha_2 X_1^{g_{21}} + \alpha_4 X_1^{g_{41}}} = \frac{g_{21}V_2^+ + g_{41}V_4^+}{V_2^+ + V_4^+} \qquad (10)$$

where all variables are evaluated at the operating point.

Estimation of Subsets of Fluxes for DFE

DFE consists of two phases. The first phase represents metabolic pathways as flux systems of the type

$$\frac{dX_i}{dt} = \sum_{j=1}^{n} \sigma_{ij} v_{ij} - \sum_{j=1}^{n} \sigma_{ij} v_{ij}, \qquad (i = 1,...,n) \qquad (11)$$

where the σ quantities are stoichiometric coefficients, which are typically known [40-42]. Given slope estimates at K time points, as before, the equations in [Eq-11] constitute one set of n linear equations in the fluxes v_{ij} and v_{ij} at each of the K time points. If the number of non-zero fluxes equals n, each set of equations can be inverted and the result is a numerical value of each flux at each time point. In a second phase of DFE, one attempts to find mathematical representations for these numerical flux estimates.

In most realistic cases, the number of fluxes exceeds the number of dependent variables and the system in [Eq-11] cannot be inverted, unless independent information about some of the fluxes is obtained. Some strategies for obtaining such information have been proposed [43,44], but they are not always sufficient. Here, the proposed coarse estimation, in forward or backward direction, may be used to obtain such fluxes. Thus, if the system contains eight dependent variables and 12 fluxes, four fluxes typically need to be estimated. The choice of these four is not entirely arbitrary, but this step usually permits several degrees of freedom [45]. As the method proposed here assumes validity of the power-law representation, it is *a priori* beneficial to select variables for estimation that vary only modestly within the given data set, as this limited range of variation increases the likelihood that the power-law representation is adequate.

Metabolic Reaction Networks for Illustration of the Method

Generic Branched Pathway Model of Michaelis-Menten Type

The first pathway for illustration is the metabolic reaction network shown in [Fig-1]. In previous work, this system was used to demonstrate estimation techniques for BST systems [24,26,27,30]. Here, we model the same pathway with Michaelis-Menten-type rate laws. The advantage of this artificial system is that we have complete knowledge of all of its features. Thus, we can create "data" with this model, subsequently pretend not to know the model parameters or even the mathematical structure of the equations, and ultimately gauge how well our inference method works and where it fails.

The specific model, with typical parameters, is

$$\frac{dX_1}{dt} = \frac{1.0}{1.0 + \left(X_3 / 15.0\right)} - \frac{X_1}{X_1 + 5.0} - \frac{X_1}{X_1 + 2.0} \tag{12}$$

$$\frac{dX_2}{dt} = \frac{X_1}{X_1 + 5.0} - \frac{X_2}{X_2 + 2.0} \tag{13}$$

$$\frac{dX_3}{dt} = \frac{X_2}{X_2 + 2.0} - \frac{2.0 X_3 X_4}{(X_3 + 10.0)(X_4 + 6.0)} \tag{14}$$

$$\frac{dX_4}{dt} = \frac{X_1}{X_1 + 2.0} - \frac{X_4}{X_4 + 2.0} \tag{15}$$

The steady state of the system is approximately $X_1^* = 1.050$, $X_2^* = 0.420$, $X_3^* = 13.96$ and $X_4^* = 1.050$. As initial values we choose, more or less arbitrarily, $X_{10} = 1.4$, $X_{20} = 2.7$, $X_{30} = 1.2$, $X_{40} = 0.4$.
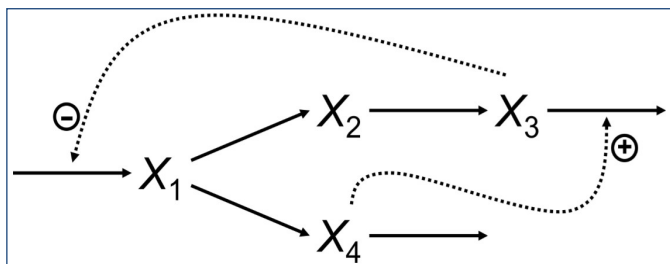

Fig. 1- Reaction diagram of a generic branched pathway.

Biosynthesis of Aspartate-derived Amino Acids

The metabolic reaction network for this example is shown in [Fig-2]. A model of the pathway was originally presented with traditional kinetic rate functions for seven dependent variables, namely X_1 = [AspP], X_2 = [ASA], X_3 = [Lys], X_4 = [Hser], X_5 = [PHser], X_6 = [Thr] and X_7 = [Ile] [46]. Here, we additionally consider X_8 = [Threonyl-tRNA]. The flux equations are

$$\frac{dX_1}{dt} = v_{AK} - v_{ASADH} \tag{16}$$

$$\frac{dX_2}{dt} = v_{ASADH} - v_{DHDPS} - v_{HSDH} \tag{17}$$

$$\frac{dX_3}{dt} = v_{DHDPS} - v_{LKR} \tag{18}$$

$$\frac{dX_4}{dt} = v_{HSDH} - v_{HSK} \tag{19}$$

$$\frac{dX_5}{dt} = v_{HSK} - v_{TS1} \tag{20}$$

$$\frac{dX_6}{dt} = v_{TS1} - v_{TD} - v_{(Thr)tRNAsth} \tag{21}$$

$$\frac{dX_7}{dt} = v_{TD} - v_{(Ile)tRNAsth} \tag{22}$$

$$\frac{dX_8}{dt} = v_{(Thr)tRNAsth} \tag{23}$$

The kinetic formulations of these fluxes are taken directly from [46]; they are reproduced in the Additional Results. With these settings, the steady-state values for the first seven variables are approximately $X_1^* = 0.3440$, $X_2^* = 0.9645$, $X_3^* = 69.38$, $X_4^* = 0.9368$, $X_5^* = 45.27$, $X_6^* = 302.4$, and $X_7^* = 59.25$. X_8 continues to accumulate in this model and has no steady state. As initial values, we start the system either 20% above the steady state or with values that correspond to 20% of the steady state.
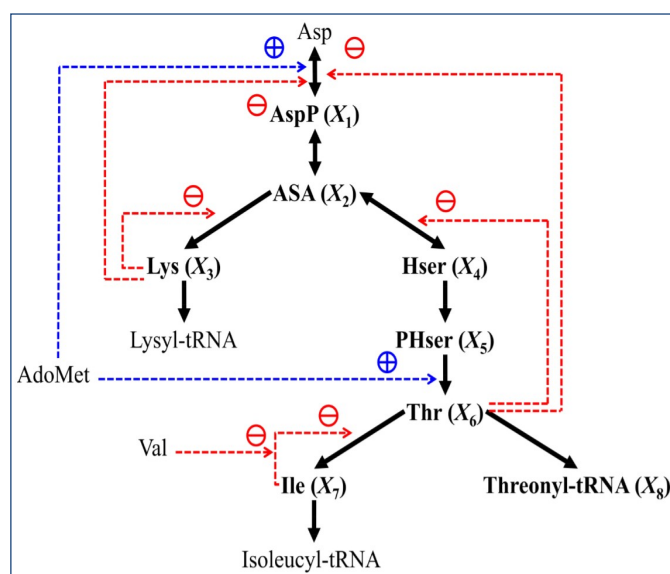

Fig. 2- Metabolic reaction network of the biosynthesis of aspartate-derived amino acids. Abbreviations are: Asp: L-Aspartate, AspP: L-Aspartate-4-phosphate, ASA: L-Aspartate-semialdehyde, Lys: L-Lysine, Hser: Homoserine, PHser: O-Phospho-L-homoserine, AdoMet: S-Adenosylmethionine, Thr: L-Threonine, Ile: L-Isoleucine, Val: L-Valine. Lysyl: tRNA and Isoleucyl-tRNA are shown here as end products, although they are not explicitly included in the model.

Computational Environment

All calculations were performed with Microsoft Visual Studio 2008 on Windows 7, with a computer with Intel® Core™ i7 CPU and mounted 4.00 GB memory.

Results

Performance Evaluations

The performance of the proposed method is first evaluated with the generic branched pathway model consisting of Michaelis-Menten processes [Fig-1], [Eq-12] to [Eq-15].

Approximate S-system Parameters at an Operating Point

In order to assess how well the power-law approximation might perform, we first approximate the original equations with an S-system. As the operating point, we choose the values $X_{1, op} = 1.7$, $X_{2, op} = 1.5$, $X_{3, op} = 5.0$ and $X_{4, op} = 1.3$, which are more or less in the center of the range of variation for each variable. The parameter

values of the approximate S-system are shown in [Table-1]. Although its functions are different, the S-system approximates the original Michaelis-Menten model overall quite well [Fig-S1].

Table 1- Parameter values of an S-system model approximating the Michaelis-Menten model of a generic branched pathway [Fig-1], [Eq-12] to [Eq-15] at the operating point ($X_{1, op}$ = 1.7, $X_{2, op}$ = 1.5, $X_{3, op}$ = 5.0, $X_{4, op}$ = 1.3).

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	β_i	h_{ii}
X_1	1.12E+00	-	-	-2.50E-01	-	5.15E-01	6.14E-01
X_2	1.71E-01	7.46E-01	-	-	-	3.40E-01	5.71E-01
X_3	3.40E-01	-	5.71E-01	-	-	4.06E-02	6.67E-01
X_4	3.45E-01	5.41E-01	-	-	-	3.36E-01	6.06E-01

Unconstrained Backward Estimation

Using backward estimation for various grid point combinations of β and h parameters, we computed linear regressions and SSE values, as well as calculation times. The results for the best β - h combinations are shown in [Table-2a] and [Fig-3]. Although a relatively coarse grid was used, the computed dynamics fits the artificial data rather well. Other β - h combinations lead to much inferior fits [Fig-S2].
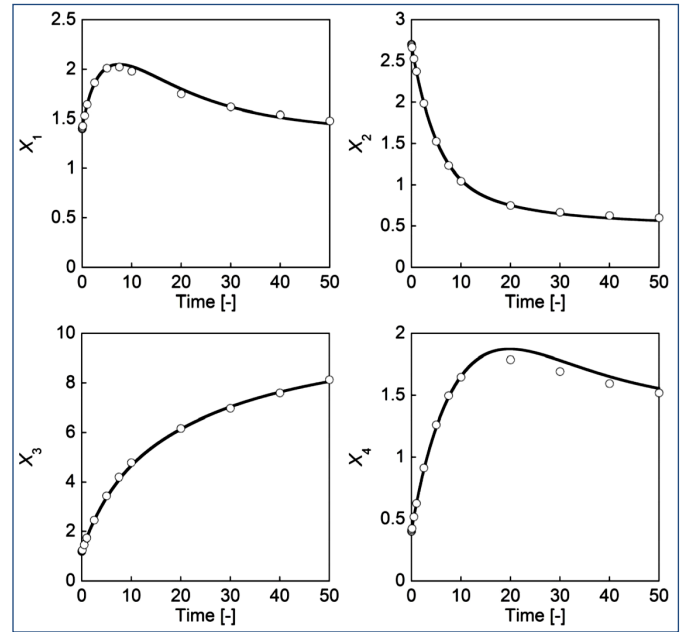

Fig. 3- Comparison between results of the inferred unconstrained S-system model (lines) and those computed with the Michaelis-Menten model of the generic branched pathway in [Fig-1] (symbols). Time and concentrations are given in unspecified units.

Table 2- Parameter values, residual errors (SSE values) and computing times (CT) associated with different inference strategies using artificial data characterizing the branched pathway model [Fig-1], [Eq-12] to [Eq-15]

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	β_i	h_{ii}	SSE	CT (s)
				Backward estimation					
				a) Unconstrained inference					
X_1	9.03E-01	-	-	-2.16E-01	-	5.00E-01	4.00E-01	4.00E-03	1.88E+01
X_2	3.41E-01	3.96E-01	-	-	-	5.00E-01	4.00E-01	2.57E-03	4.60E+01
X_3	3.76E-01	-	5.96E-01	-	-	1.00E-01	4.00E-01	7.06E-02	1.82E+01
X_4	1.01E+00	2.02E-01	-	-	-	1.00E+00	2.00E-01	1.70E-02	1.90E+01
				b) Constrained inference					
X_1	1.72E+00	-	-	-9.98E-02	-	1.31E+00	1.84E-01	4.79E-03	4.68E-01
X_2	3.00E-01	1.20E-01	-	-	-	3.76E-01	5.96E-01	5.24E-02	4.05E-01
X_3	3.76E-01	-	5.96E-01	-	-	1.00E-01	4.00E-01	7.06E-02	1.82E+01
X_4	1.01E+00	2.02E-01	-	-	-	1.00E+00	2.00E-01	1.70E-02	1.90E+01
				Forward estimation					
				c) Unconstrained inference					
X_1	5.00E-01	-	-	-6.00E-01	-	1.06E-01	9.87E-01	8.99E-03	3.56E+01
X_2	1.00E-01	4.00E-01	-	-	-	1.86E-01	9.31E-01	9.11E-04	2.99E+01
X_3	1.00E+00	-	4.00E-01	-	-	9.45E-01	-9.76E-02	9.45E-03	2.10E+01
X_4	5.00E-01	4.00E-01	-	-	-	4.90E-01	4.43E-01	6.39E-04	3.21E+01
				d) Constrained inference (α_2 and g_{21} selected first)					
X_1	5.00E-01	-	-	-6.00E-01	-	1.06E-01	9.87E-01	8.99E-03	3.56E+01
X_2	1.00E-01	-2.00E-01	-	-	-	1.50E-01	1.09E+00	6.70E-03	4.05E-01
X_3	1.50E-01	-	1.09E+00	-	-	9.29E-06	4.01E+00	2.41E+00	4.21E-01
X_4	7.08E-03	1.08E+01	-	-	-	9.69E-01	3.13E+00	4.96E+00	2.23E+00
				e) Constrained inference (α_4 and g_{41} selected first)					
X_1	5.00E-01	-	-	-6.00E-01	-	1.06E-01	9.87E-01	8.99E-03	3.56E+01
X_2	5.78E-03	7.54E-01	-	-	-	4.17E-02	2.29E+00	8.22E-02	3.80E+00
X_3	4.17E-02	-	2.29E+00	-	-	NS	NS	NS	NS
X_4	1.00E-01	1.00E+00	-	-	-	2.91E-02	3.52E+00	3.56E-02	4.40E-01

NS: no solution was obtained

We also created data from the same system and superimposed them with 10% uniformly distributed random noise. While the concentration values thus deviate from smooth trend lines, we assume that appropriate smoothing of these noisy data would lead to correct slopes S_{ik}. [Fig-4] shows results from 300 estimations of α_i, g_{ij} and β_i in pseudo-three-dimension plots and the most illustrative projections. Red circles mark the values shown in [Table-2a]. The results are not connected and do not form clouds, as one encounters them frequently, due to the discrete choices of grid points. As has been observed in other contexts [47], the correlations between α and β parameters for the different results are more or less linear, while correlations between rate constants and kinetic orders are nonlinear and more or less logarithmic.
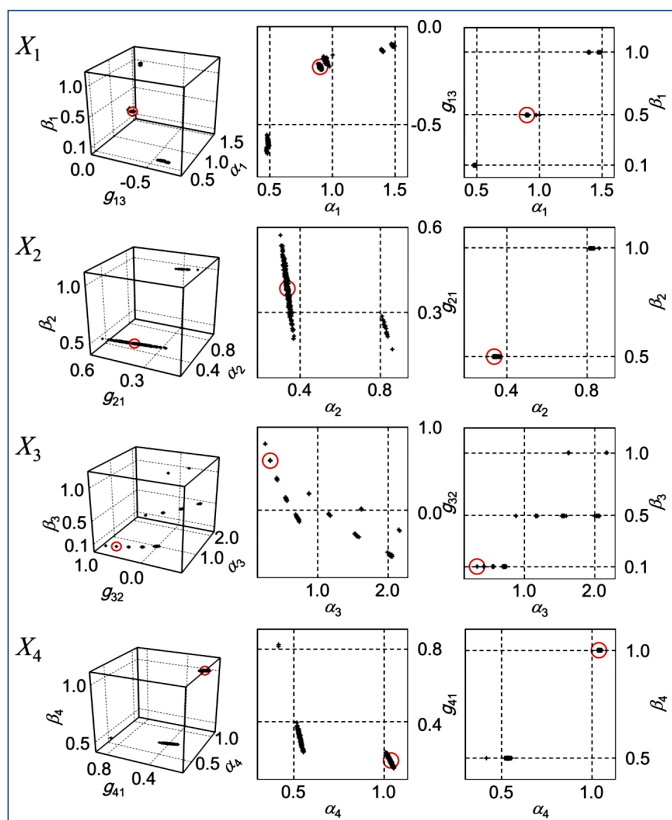
Fig. 4- Distributions of kinetic parameter values, inferred for the Michaelis-Menten type branched pathway model with the unconstrained backward estimation strategy with 300 datasets with ±10% noise. Red circles mark the kinetic parameter values shown in [Table-2a]. Left panel: Pseudo-three-dimensional representations of α_i-g_{ij}-β_i; center panel: α_i-g_{ij} projections; right panel: α_i-β_i projections.
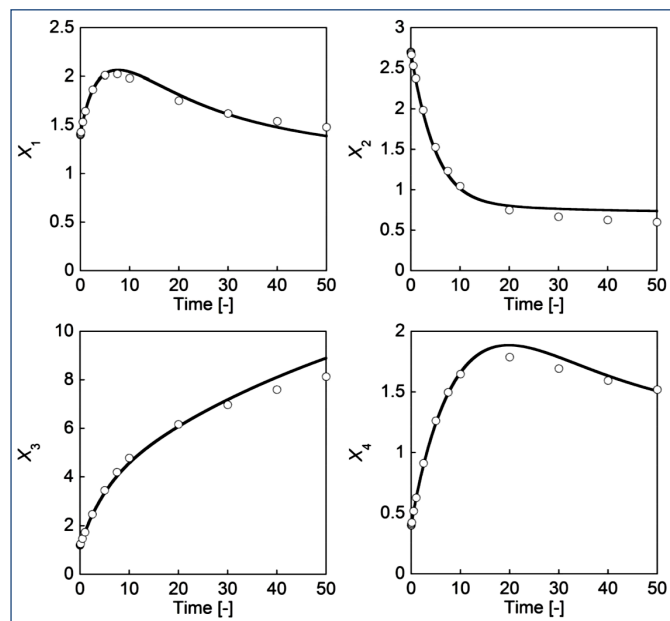
Constrained Backward Estimation

While simulations with the parameters estimated so far fit the given data quite well, an adequate model should satisfy the constraints that derive from the topology of the metabolic reaction network. For instance, the production of X_3 in [Fig-1] constitutes the same process as the degradation of X_2, which mandates

$$\beta_2 X_2^{h_{22}} = a_3 X_2^{g_{32}} \qquad (24)$$

The branched pathway imposes an additional constraint which, outside trivial special cases, can only be satisfied at an operating point:

$$\beta_1 X_1^{*h_{11}} = a_2 X_1^{*g_{21}} + a_4 X_1^{*g_{41}} \qquad (25)$$

(*cf.* [Eq-9] and [Eq-10]). To satisfy these constraints during the backward estimation, kinetic parameters for X_3 and X_4 are determined first, and then the estimation moves upstream the metabolic reaction network. [Table-2b] indicates that the resulting minimum SSE values are slightly higher than before. Nonetheless, simulations with the best parameter values yield good results [Fig-5].

Fig. 5- Comparison between results of the inferred S-system model (lines) and results computed with the Michaelis-Menten model of the generic branched pathway in [Fig-1] (symbols), using fully constrained backward inference. Time and concentrations are given in unspecified units.

Forward Estimation

As described in the Methods section, the estimation may also proceed in forward direction. Without constraints, optimal kinetic parameters, SSEs and calculation times are shown in [Table-2c]. All SSE values indicate satisfactory data fits, which are confirmed in [Fig-6]. As before, we also estimated the parameters from noisy data. The results are shown in [Fig-S3].

Because the pathway is branched, the forward computation of constrained parameter values allows some flexibility, because either α_2 and g_{21} or α_4 and g_{41} may be estimated first from the branch point constraint in [Eq-25]. If α_2 and g_{21} are estimated first, then α_4 and g_{41} are calculated using [Eq-26] and [Eq-27], respectively:

$$g_{41} = \frac{h_{11}\beta_1 X_1^{*h_{11}} - g_{21}\alpha_2 X_1^{*g_{21}}}{\beta_1 X_1^{*h_{11}} - \alpha_2 X_1^{*g_{21}}} \qquad (26)$$

$$a_4 = \frac{\beta_1 X_1^{*h_{11}} - \alpha_2 X_1^{*g_{21}}}{X_1^{*g_{21}}} \qquad (27)$$

In turn, if α_4 and g_{41} are estimated first, then α_2 and g_{21} are determined from the corresponding constraints. Once influx parameters are determined, forward estimation for X_2, X_3 and X_4 is performed and efflux parameters with the lowest SSE values are retained [Table-2d] and [Table-2e]. While the branch point offers some flexibility, not all choices are feasible, because in some cases the constraints (26) and (27) can no longer be satisfied for all values of the

dependent variables. As an example, some choices of a_4 and g_{41} lead to infeasibility for the efflux of X_3. In other cases, the best-fitting kinetic order might be negative, which is inconsistent with the structure of the pathway.
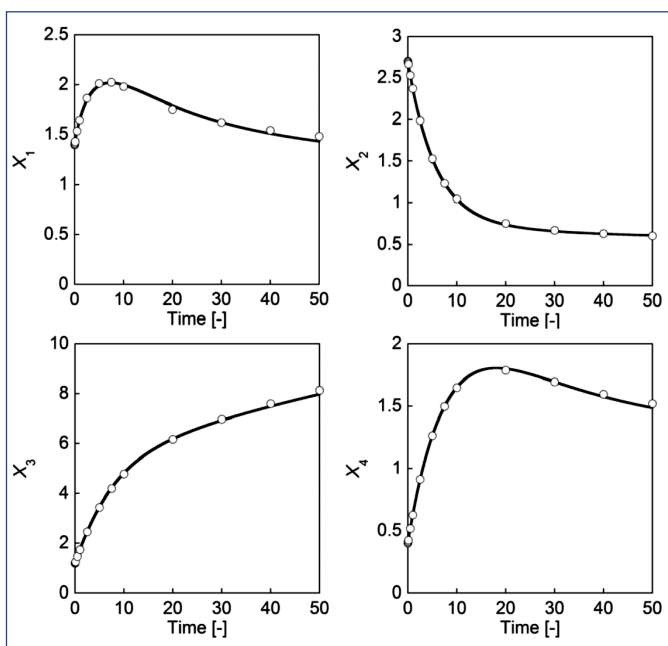
Fig. 6- Comparison between the inferred S-system model (lines) and results computed with the Michaelis-Menten model of the generic branched pathway in [Fig-1] (symbols), using unconstrained forward inference. Time and concentrations are given in unspecified units.

Biosynthesis of Aspartate-derived Amino Acids

To evaluate the scalability of the proposed method, we applied the method to a larger metabolic reaction network model from the literature. Curien et al. [46] constructed a dynamic model for the biosynthesis of aspartate-derived amino acids from kinetic measurements *in vitro*. We pretend here that simulation results with their model are measured experimental time series data, and it is our goal to infer an adequate mathematical model from these "data." In particular, the analysis allows us to compare the inferred fluxes with those used by the authors to construct their model from the bottom up. In contrast to the original article, we merge isozyme catalyzed reactions into single processes.

Analysis I: Initial Values are Set as 120% of the Steady State Values

The first set of "data" corresponds to a simulation with the published model, which starts with initial values that are set as 120% of the steady-state values. Such a deviation is relatively small, which has advantages and disadvantages. On the positive side, the power-law approximation is possibly good enough to capture the dynamics of the system. On the negative side, the data are not very informative. As with the earlier example, we first estimate parameters for unconstrained equations, using the backward strategy. [Table-3a] shows the resulting kinetic parameters and corresponding SSE values. [Fig-S4] compares trend lines calculated by integration of the eight S-system equations with those from the original equations. Some of the results are fairly good, but X_6 is mis-estimated, which through feedbacks affects all other variables after a while. [Fig-S5] shows corresponding estimations for data with noise. The corresponding

forward estimation without constraints also yields good results [Table-3c] and [Fig-S6].

More relevant is an estimation that accounts for constraints in the pathway system. The results of a constrained backward inference are shown in [Table-3b] and [Fig-7]; a corresponding figure with a magnified initial range is presented as [Fig-S7]. Even though the power-law terms are quite different from the functions used to construct the original model, the results are surprisingly good, especially in the initial time range.

Analysis II: Initial Values are Set as 20% of Steady State Values

In contrast to the previous analysis, the deviation from the steady state is now considerable. The estimation method itself is not affected by this strong deviation, but it is more likely that the power-law approximation might exceed the range of valid representation. Nonetheless, the results of an unconstrained forward estimation are actually surprisingly good, except for variable X_8, which is not subject to degradation and only accumulates material and with it all inference errors [Table-4] and [Fig-8]. In an unconstrained backward estimation, X_8 is modeled very well, but the other variables show inferior fits, thus indicating that the power-law formulation leaves its range of valid representation [Table-S1], [Fig-S8] and [Fig-S9].

Analysis III: Estimation of a Minimum Set of Fluxes to Complement DFE

The pathway contains eight dependent variables and ten fluxes. Thus, we need to estimate two independent fluxes. Furthermore, characterizability analysis of the pathway system reveals that even perfect, complete time series data do not allow a full characterization of flux profiles, because the same time series data could be associated with different flux profiles. In particular, it is possible that pairs of fluxes, such as $V_i(t_k)$ and $V_{-i}(t_k)$, permit a shift by an unknown constant c, which disappears in the differential equation, because $V_i(t_k) + c - (V_{-i}(t_k) + c) = V_i(t_k) - V_{-i}(t_k)$. One should note that this possible shift is not a result of the slope-estimation-decoupling strategy or our coarse inference method. It constitutes a genuine estimation challenge for underdetermined flux systems, because the "true" differential equations generate results that are exactly equivalent to those from other differential equations with shifted fluxes.

As a first example, consider "data" obtained with initial values at 120% of the steady state. We use two fluxes, estimated from the earlier constrained backward estimation and use DFE to infer the remaining fluxes. For Case 1, we use the fluxes v_{ASADH} and v_{HSDH} and for Case 2, we use v_{DHDPS} and v_{HSK} [Fig-9].

The main results are shown in [Fig-10]. The green and blue lines are inferred with DFE for Cases 1 and 2, while the red lines correspond to the fluxes obtained from the coarse, constrained backward estimation and the black lines correspond to the fluxes in Curien's original model. Three generic features emerge. First, the inferred DFE fluxes by and large have the correct shape; *i.e.*, they are very similar to the fluxes of Curien's model, with occasional exceptions at the very beginning or end of the plots. Second, most of the fluxes exhibit some shift, indicating a degree of freedom and non-uniqueness that is predicted by characterizability analysis [45]. These shifts are not entirely arbitrary though: for instance, the shift in v_{AK} is the same as the shift in v_{ASADH}, as both fluxes are associated with X_1. Third, the original fluxes and the backward-inferred fluxes have slightly different shapes, even though the metabolite trends [Fig-7]

are overall well modeled. These differences reflect the fact that relatively complex kinetic rate functions were approximated with rather simple power-law functions. The situation is similar for "data" that are initialized at 20% of the steady-state values [Fig-S10].

Table 3- Parameter values and residual errors (SSE values) associated with different inference strategies for the model of aspartate-derived amino acid biosynthesis [Fig-2]. In Analysis I, the initial values were set at 120% of the steady state.

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	g_{i5}	g_{i6}	g_{i7}	g_{i8}	β_i	h_{ii}	SSE
					Backward Estimation							
					a) Unconstrained inference							
X_1	1.26E+01	-	-	-1.21E-01	-	-	-1.63E-01	-	-	5.00E+00	5.00E-01	7.89E-04
X_2	1.52E+01	3.98E-01	-	-	-	-	-	-	-	1.00E+01	4.00E-01	6.33E-03
X_3	1.24E+02	-	8.47E-01	-1.37E+00	-	-	-	-	-	1.00E-01	3.00E-01	2.55E-03
X_4	5.76E+00	-	9.82E-02	-	-	-	-2.51E-02	-	-	5.00E+00	1.00E-01	5.75E-04
X_5	7.37E+00	-	-	-	9.20E-02	-	-	-	-	5.00E+00	1.00E-01	1.12E-01
X_6	1.85E-02	-	-	-	-	1.02E+00	-	-	-	5.00E-01	1.00E-01	1.36E+01
X_7	1.17E+01	-	-	-	-	-	9.49E-01	-2.19E+00	-	1.00E-01	3.00E-01	2.73E-05
X_8	9.19E-02	-	-	-	-	-	2.20E-01	-	-	-	-	9.18E-05
					b) Constrained inference							
X_1	7.51E+01	-	-	-4.11E-01	-	-	-4.72E-01	-	-	4.60E+00	1.58E+00	7.58E-04
X_2	4.60E+00	1.58E+00	-	-	-	-	-	-	-	8.62E-01	1.38E+00	2.79E-02
X_3	1.24E+02	-	8.47E-01	-1.37E+00	-	-	-	-	-	1.00E-01	3.00E-01	2.55E-03
X_4	2.02E+00	-	1.79E+00	-	-	-	-2.46E-01	-	-	5.17E-01	1.62E+00	1.15E-04
X_5	5.17E-01	-	-	-	1.62E+00	-	-	-	-	2.25E-04	1.99E+00	4.41E+00
X_6	2.25E-04	-	-	-	-	1.99E+00	-	-	-	1.33E-02	5.94E-01	3.68E+01
X_7	1.17E+01	-	-	-	-	-	9.49E-01	-2.19E+00	-	1.00E-01	3.00E-01	2.73E-05
X_8	9.19E-02	-	-	-	-	-	2.20E-01	-	-	-	-	9.18E-05
					Forward estimation							
					c) Unconstrained inference							
X_1	1.00E+03	-	-	-1.00E+00	-	-	-1.00E+00	-	-	1.05E+03	9.10E+00	4.23E-04
X_2	1.00E+01	1.00E+00	-	-	-	-	-	-	-	3.48E+00	9.33E-01	6.57E-03
X_3	1.00E+02	-	2.00E-01	-1.00E+00	-	-	-	-	-	1.57E+01	-5.66E-01	2.74E-04
X_4	5.00E+00	-	4.00E-01	-	-	-	-2.00E-01	-	-	1.58E+00	3.97E-01	3.65E-04
X_5	1.00E+00	-	-	-	6.00E-01	-	-	-	-	7.27E-02	6.79E-01	1.41E-01
X_6	5.00E-01	-	-	-	-	4.00E-01	-	-	-	3.16E+00	-5.48E-02	2.16E+01
X_7	5.00E+00	-	-	-	-	-	2.00E-01	-6.00E-01	-	1.40E+00	-9.57E-03	9.64E-03
X_8	5.00E+01	-	-	-	-	-	-8.00E-01	-	-	-	-	1.79E+04

Table 4- Parameter values and residual errors (SSE values) associated with different inference strategies for the model of aspartate-derived amino acid biosynthesis [Fig-2]. In Analysis II, the initial values set at 20% of the steady state.

|	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	g_{i5}	g_{i6}	g_{i7}	g_{i8}	β_i	h_{ii}	SSE
					Forward estimation							
					Unconstrained inference							
X_1	1.00E+01	-	-	4.00E-01	-	-	-6.00E-01	-	-	1.13E+02	4.49E+00	7.14E-03
X_2	1.00E+01	2.00E-01	-	-	-	-	-	-	-	8.40E+00	1.78E-01	7.15E-01
X_3	1.00E+01	-	6.00E-01	-6.00E-01	-	-	-	-	-	1.32E-01	4.13E-01	6.19E+00
X_4	1.00E+00	-	8.00E-01	-	-	-	-2.00E-01	-	-	3.07E-01	1.19E+00	5.50E-02
X_5	1.00E+00	-	-	-	8.00E-01	-	-	-	-	7.68E-02	6.74E-01	9.26E+00
X_6	1.00E+00	-	-	-	-	2.00E-01	-	-	-	1.37E+00	7.77E-02	1.55E+01
X_7	1.00E-01	-	-	-	-	-	2.00E-01	6.00E-01	-	1.10E-01	8.45E-01	2.32E+01
X_8	5.00E+00	-	-	-	-	-	-4.00E-01	-	-	-	-	1.73E+05
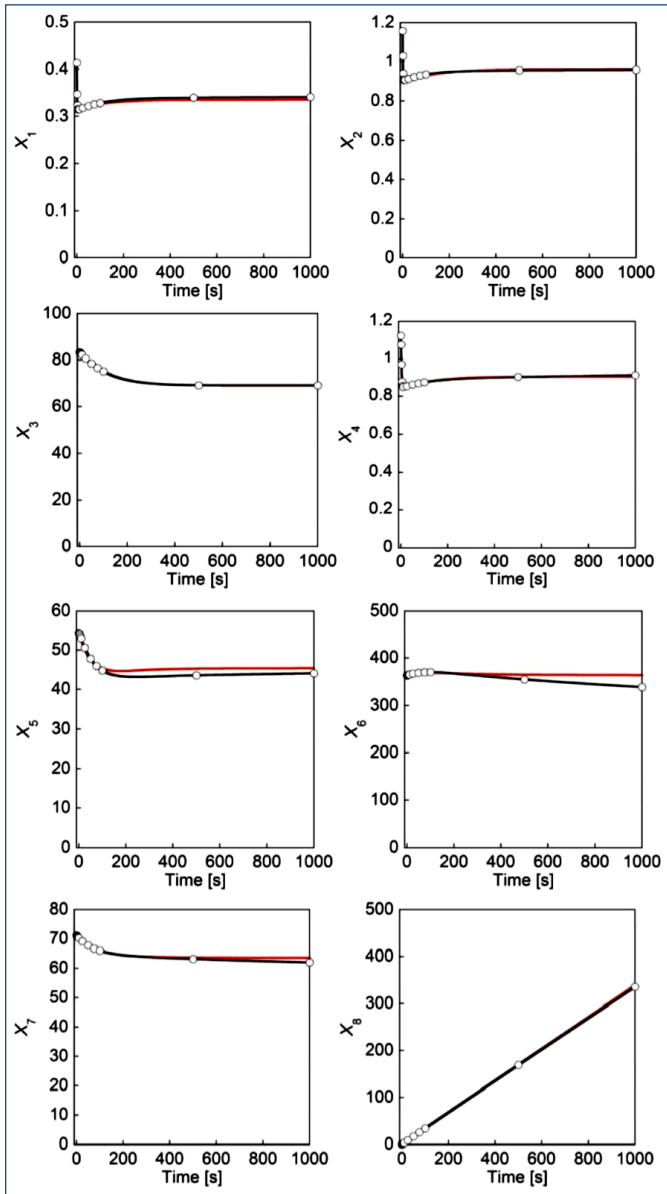
Fig. 7- Comparisons between "data" (symbols and connecting black lines) obtained from the model of aspartate-derived amino acid biosynthesis [46] and results the S-system model with kinetic parameters inferred through constrained backward estimation (red lines). Analysis I: Initial values set at 120% of the steady state. See [Fig-S7] for a magnified representation of the initial time range.

Discussion and Conclusions

The estimation of parameter values for nonlinear models of biological systems is a challenging task. Even harder is the identification of functional forms for the processes governing biological systems. We have shown here that both tasks can be aided by a relatively coarse, but very simple, estimation strategy that exclusively uses linear regression and is therefore easily scalable. The strategy may be used in different ways. First, one may obtain a complete power-law model of a pathway by assuming a relatively coarse grid of parameter values characterizing some fluxes in the system. The result is a fully parameterized model that may be used for further analyses. The model parameters thus inferred may also be used as start values for launching a more refined parameter estimation based directly on the nonlinear differential equations. Assuming that

the coarsely inferred parameter values are acceptable, this nonlinear estimation would start in-or close to-a basin of attraction of the true solution, thereby circumventing many issues with nonlinear optimization methods. The coarsely parameterized model may also be used to augment DFE. In this case, only a few fluxes are retained from the coarse estimation and all other fluxes are computed with methods of DFE. If the retained fluxes correspond to variables that do not vary very much within the given datasets, one might assume that the power-law representation is relatively accurate. As a consequence, this strategy makes the fewest assumptions and yields results in an almost unbiased fashion. A remaining issue is that the topology of the pathway system may allow one or more degrees of freedom within the flux profile, which cannot be resolved with DFE method, unless additional, independent information is available.
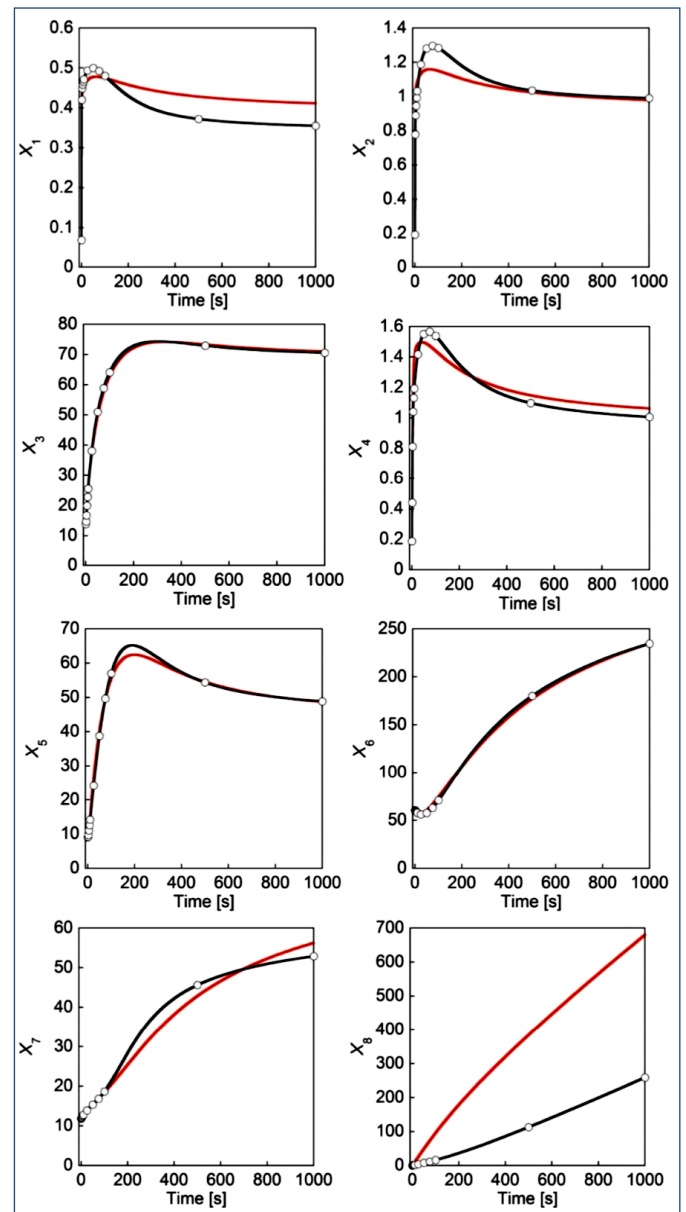
Fig. 8- Comparisons between "data" (symbols and connecting black lines) obtained from the model of aspartate-derived amino acid biosynthesis [46] and results the S-system model with kinetic parameters inferred through unconstrained forward estimation (red lines). Analysis II: Initial values set at 20% of the steady state.
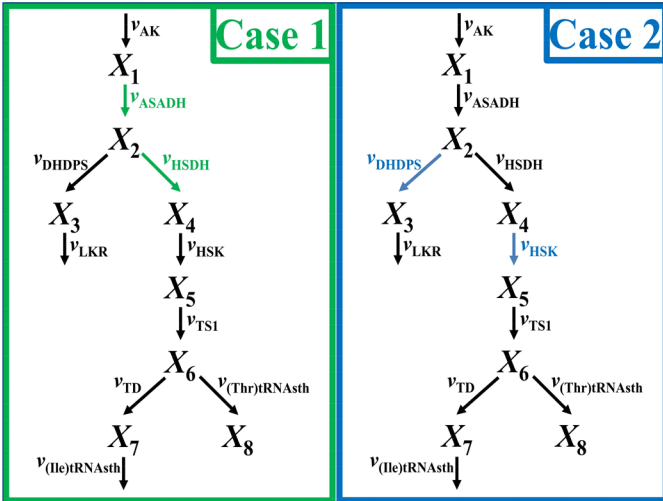
Fig. 9- Two pairs of fluxes were selected to complement DFE for Case Studies 1 and 2. Their values were taken from the constrained backward inference with the model of aspartate-derived amino acid biosynthesis [Fig-2]; [46] and used to perform DFE. Abbreviations are: v_{AK}–Aspartate kinase, v_{ASADH}–Aspartate semi-aldehyde dehydrogenase, v_{DHDPS}–Dihydrodipicolinate synthase, v_{LKR}–Lysine ketoglutarate reductase, v_{HSDH}–Homoserine dehydrogenase, v_{HSK}–Homoserine kinase, v_{TS1}–Threonine synthase, v_{TD}–Threonine deaminase, $v_{(Thr)tRNAsth}$–Threonyl-tRNA synthetase, $v_{(Ile)tRNAsth}$–Isoleucyl-tRNA synthetase.
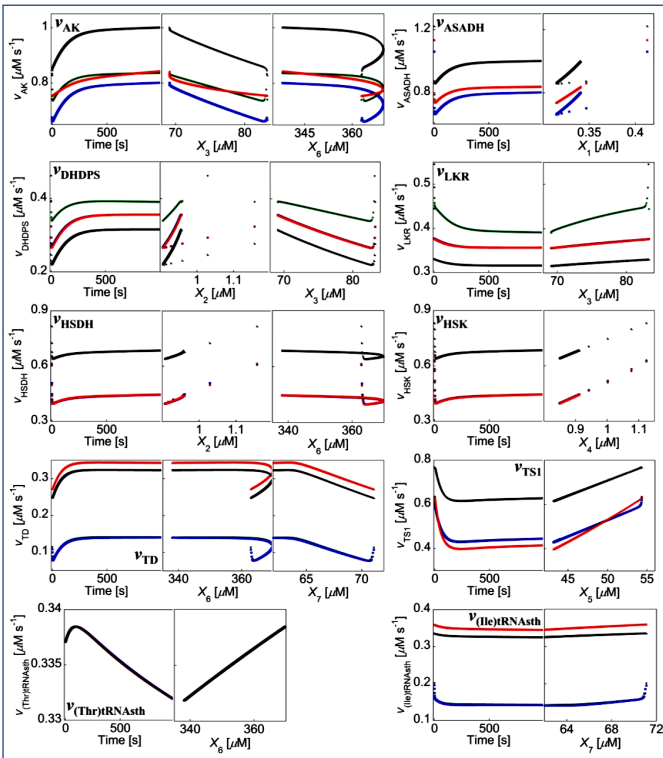
Fig. 10- Results of DFE, augmented with pairs of fluxes as shown in [Fig-9]. The fluxes obtained with DFE (green and blue; sometimes overlapping) by and large have the shapes of the fluxes used by Curien et al. [46] to construct the model of aspartate-derived amino acid biosynthesis, but they are shifted, thus indicating a remaining degree of freedom. The red flux graphs show the corresponding backward-inferred, constrained power-law representations (*cf.* [Fig-7]).

Additional Results

Branched Pathway

To evaluate the performance of the proposed method, a generic, branched pathway model consisting of Michaelis-Menten processes [Fig-1]; [Eq-12] to [Eq-15] was analyzed. The original equations were approximated with an S-system model at an operating point chosen within the range of the variables, as described in the Text. [Table-1] shows the parameter values of the approximate S-system. The model approximates the original reasonably well [Fig-S1].
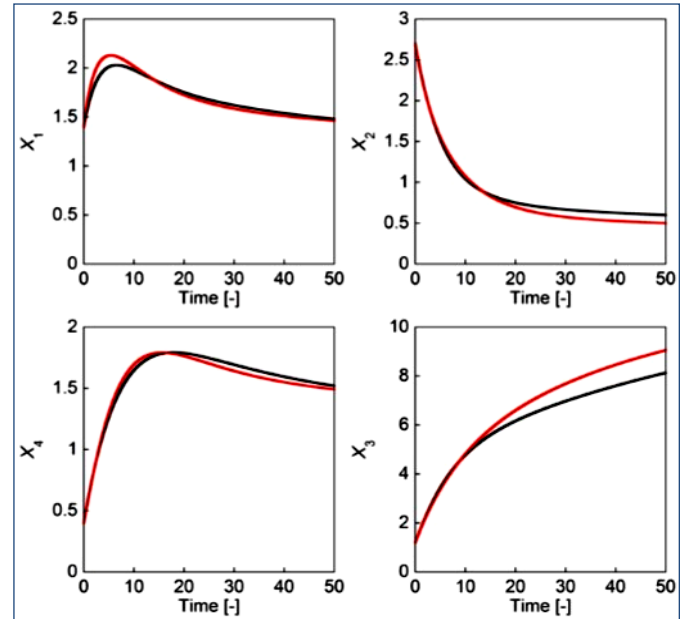

Fig. S1- Comparison between results of the approximate S-system model (red lines) and those computed with the Michaelis-Menten model of the generic branched pathway in [Fig-1] (black lines). Time and concentrations are given in unspecified units.
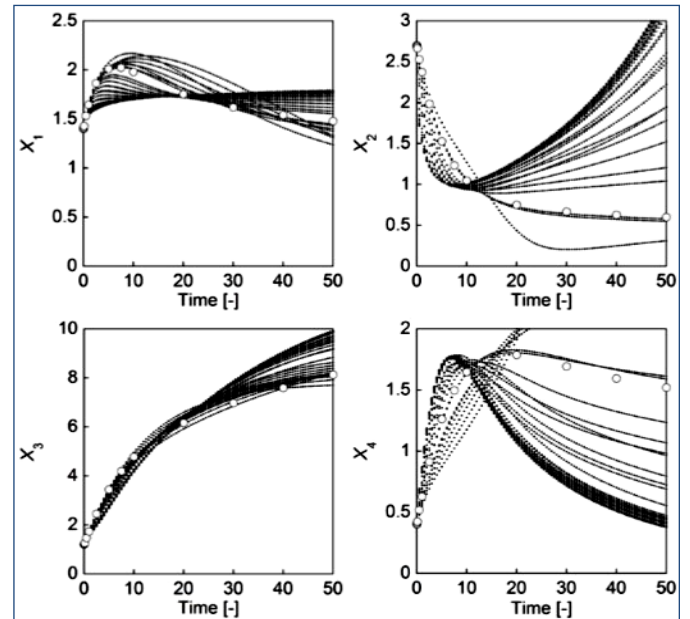
Fig. S2- Comparison between results of the various unconstrained S-system models (lines) and those computed with the Michaelis-Menten model of the generic branched pathway in [Fig-1] (symbols). Time and concentrations are given in unspecified units.

For the unconstrained backward estimation strategy, we chose various grid point combinations of β and h parameters. The dynamic results for the best β - h combination are shown in [Table-2a] and [Fig-3]; results for other combinations are shown in [Fig-S2]. The dynamic representations in these cases are much inferior.

Parameter values were also estimated for the unconstrained forward estimation strategy using data with 10% uniformly distributed random noise. [Fig-S3] shows the distributions of inferred kinetic parameter values.
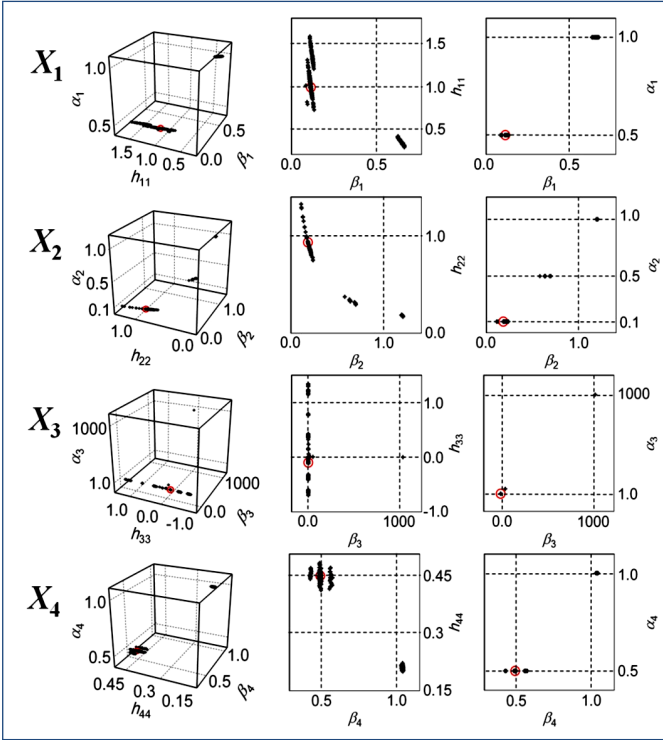
Fig. S3- Distributions of kinetic parameter values, inferred for the Michaelis-Menten-type branched pathway model with the unconstrained forward estimation strategy, using 300 datasets with 10% uniformly distributed noise. Red circles indicate the kinetic parameter values shown in [Table-2c]. Left panel: Pseudo-three-dimensional representations of β_i-h_{ii}-α_i; center panel: β_i-h_{ii} projections; right panel: β_i-α_i projections.

Biosynthesis of Aspartate-Derived Amino Acids

To evaluate the scalability of the proposed method, the method was applied to a larger metabolic reaction network model from the literature [Fig-2]; [Eq-16] to [Eq-23]; [46], as described in the Text. The kinetic representations, as proposed by Curien et al. are shown below; note that our analysis combines isozyme-catalyzed reactions.

$$v_{AK} = v_{AK1} + v_{AK2} + v_{AKI} + v_{AKII}$$

$$= [AK1] \cdot \frac{5.65 - 1.6[AspP]}{1 + \left[[Lys] \middle/ \left(\frac{550}{1 + [AdoMet]/3.5} \right) \right]^2} + [AK2] \cdot \frac{3.15 - 0.86[AspP]}{1 + ([Lys]/22)^{1.1}}$$

$$+ [AKI\text{-}HSDH\,I] \cdot \frac{0.36 - 0.15[AspP]}{1 + ([Thr]/124)^{2.6}} + [AKII\text{-}HSDH\,II] \cdot \frac{1.35 - 0.22[AspP]}{1 + ([Thr]/109)^2}$$

$$v_{ASADH} = [ASADH] \cdot (0.9[AspP] - 0.23[ASA])$$

$$v_{DHDPS} = v_{DHDPS1} + v_{DHDPS2}$$

$$= [DHDPS1] \cdot [ASA] \cdot \frac{1}{1 + ([Lys]/10)^2} + [DHDPS2] \cdot [ASA] \cdot \frac{1}{1 + ([Lys]/33)^2}$$

$$v_{HSDH} = v_{HSDHI} + v_{HSDHII}$$

$$= [AKI\text{-}HSDH\,I] \cdot 0.84 \cdot \left(0.14 + \frac{0.86}{1 + [Thr]/400} \right)$$

$$+ [AKII\text{-}HSDH\,II] \cdot 0.64 \cdot \left(0.25 + \frac{0.75}{1 + [Thr]/8500} \right)$$

$$v_{LKR} = v_{(Lys)tRNAsth} + v_{LKR,\,orig}$$

$$= V^{AaRS} \cdot \frac{[Lys]}{25 + [Lys]} + [LKR] \cdot \frac{3.1[Lys]}{13000 + [Lys]}$$

$$v_{HSK} = [HSK] \cdot \frac{2.8[Hser]}{14 + [Hser]}$$

$$v_{TD} = v_{TD,\,orig} + v_{THA}$$

$$= [TD] \cdot \frac{0.0124[Thr]}{1 + \left[[Ile] \middle/ \left(30 + \frac{74[Val]}{610 + [Val]} \right) \right]^3} + [THA] \cdot \frac{1.7[Thr]}{7000 + [Thr]}$$

$$v_{(Thr)tRNAsth} = V^{AaRS} \cdot \frac{[Thr]}{100 + [Thr]}$$

$$v_{(Ile)tRNAsth} = V^{AaRS} \cdot \frac{[Ile]}{20 + [Ile]}$$

$$v_{TS1} = v_{CGS} + v_{TS1,\,orig}$$

$$= [CGS] \cdot \frac{\left(\frac{30}{1 + 460/[Cys]} \right) \cdot [PHser]}{\left(\frac{2500}{1 + 460/[Cys]} \right) \left(1 + \frac{[Pi]}{2000} \right) + [PHser]}$$

$$+ [TS1] \cdot \frac{\left(\frac{0.42 + 3.5[AdoMet]^2/73}{1 + [AdoMet]^2/73} \right)[PHser]}{\left[\frac{250 \left(\frac{1 + [AdoMet]/0.5}{1 + [AdoMet]/1.1} \right)}{1 + \frac{[AdoMet]^2}{140}} \right] \left(1 + \frac{[Pi]}{1000} \right) + [PHser]}$$

It should be noted that the following flux equations are used in [Eq-21] and [Eq-22]. The values are also used in [Fig-10].

$$v_{TS1} = [TS1] \cdot \frac{\left(\frac{0.42 + 3.5[AdoMet]^2/73}{1 + [AdoMet]^2/73} \right)[PHser]}{\left[\frac{250 \left(\frac{1 + [AdoMet]/0.5}{1 + [AdoMet]/1.1} \right)}{1 + \frac{[AdoMet]^2}{140}} \right] \left(1 + \frac{[Pi]}{1000} \right) + [PHser]}$$

$$v_{TD} = [TD] \cdot \frac{0.0124[Thr]}{1 + \left[[Ile] \middle/ \left(30 + \frac{74[Val]}{610 + [Val]} \right) \right]^3}$$

Analysis I: Initial Values are Set as 120% of the Steady State Values

The "data" for this inference study were generated with initial values set as 120% of the steady-state values. The inferred kinetic parameter values and corresponding SSEs for the unconstrained backward estimation strategy are shown in [Table-3a]. [Fig-S4] compares trend lines calculated by integration of the eight S-system equations with those from the original equations. [Fig-S5] and [Fig-S6] show corresponding estimations for data with noise and the results of an unconstrained forward estimation, respectively. [Fig-S7] exhibits a magnified initial range for the plots shown in [Fig-7] of the Text.
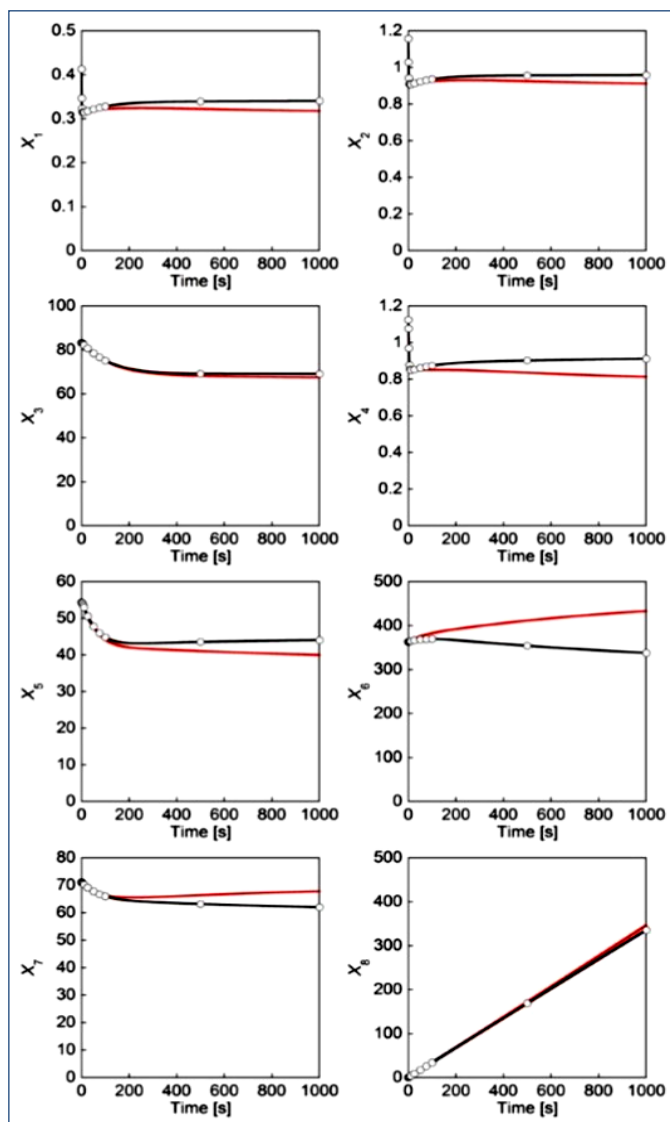
Fig. S4- Comparisons between "data" (symbols connected by black lines) obtained from the model of aspartate-derived amino acid biosynthesis [46] and results from the S-system model (red lines) with kinetic parameters inferred through *unconstrained backward estimation*. Analysis I: Initial values set at 120% of the steady state.
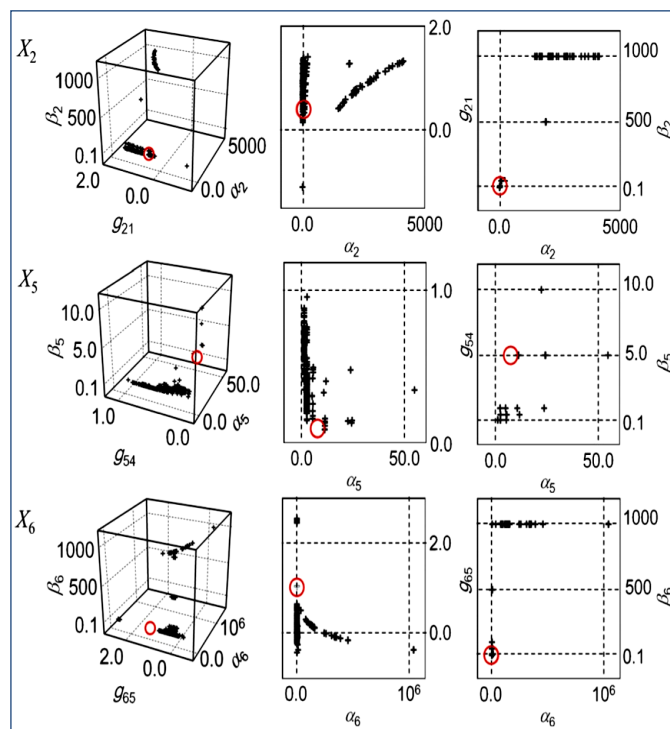

Fig. S5- Distributions of kinetic parameter values, inferred for the aspartate-derived amino acid biosynthesis model with the *unconstrained backward estimation* strategy, using 300 datasets with 10% uniformly distributed noise. Red circles indicate the kinetic parameter values shown in [Table-3a]. Analysis I: Initial values set at 120% of the steady state. Left panel: Pseudo-three-dimensional representations of α_i-g_{ij}-β_i; center panel: α_i-g_{ij} projections; right panel: α_i-β_i projections.

Analysis II: Initial Values are Set as 20% of the Steady State Values

The "data" for this simulation were generated with initial values set as 20% of the steady-state values. The inferred kinetic parameter values and corresponding SSEs of the unconstrained backward estimation strategy are shown in [Table-S1]. [Fig-S8] compares trend lines calculated by integration of the eight S-system equations with those from the original equations. [Fig-S9] shows corresponding estimations for data with noise.

Table S1: Parameter values and residual errors (SSE values) associated with different inference strategies for the model of aspartate-derived amino acid biosynthesis [Fig-2]. In Analysis II, the initial values set at 20% of the steady state.

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	g_{i5}	g_{i6}	g_{i7}	g_{i8}	β_i	h_{ii}	SSE
					Backward estimation							
					Unconstrained inference							
X_1	1.00E+01	---	---	2.32E-02	---	---	-5.73E-02	---	---	1.00E+01	2.00E-01	6.55E-04
X_2	1.19E+01	2.05E-01	---	---	---	---	---	---	---	1.00E+01	2.00E-01	6.26E-01
X_3	2.12E+01	---	8.96E-01	-8.63E-01	---	---	---	---	---	1.00E-01	4.00E-01	3.32E+00
X_4	1.47E+00	---	6.41E-01	---	---	---	-6.87E-02	---	---	1.00E+00	6.00E-01	5.26E-04
X_5	1.03E+00	---	---	---	6.33E-01	---	---	---	---	1.00E-01	6.00E-01	1.01E+01
X_6	3.79E-02	---	---	---	---	7.95E-01	---	---	---	1.00E-01	4.00E-01	9.10E+02
X_7	1.00E+03	---	---	---	---	---	1.02E-05	1.00E+00	---	1.00E+03	1.00E+00	1.41E+01
X_8	2.32E-02	---	---	---	---	---	4.73E-01	---	---	---	---	2.07E+01
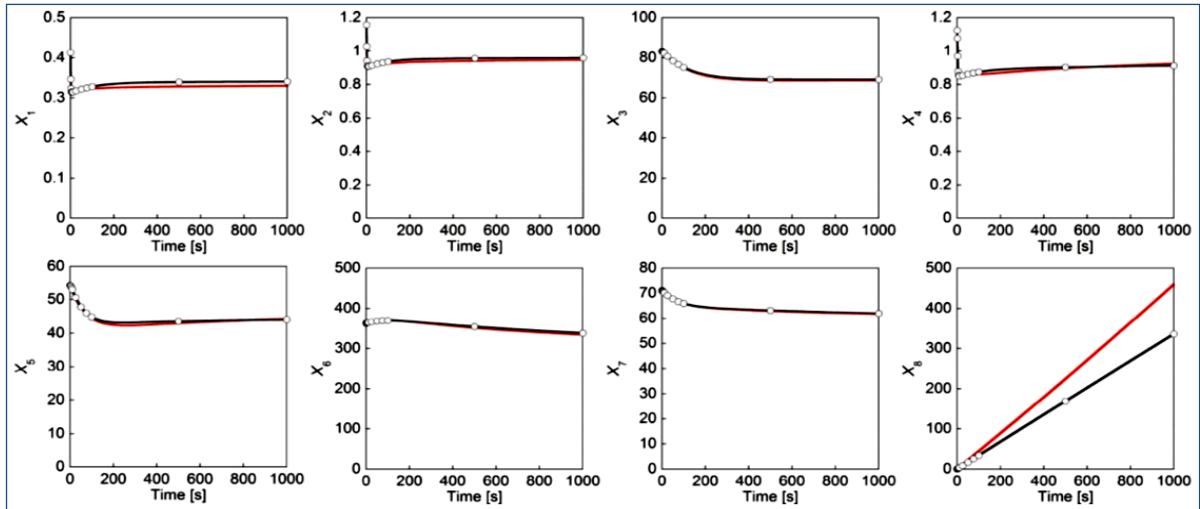
Fig. S6- Comparisons between "data" (symbols connected by black lines) obtained from the model of aspartate-derived amino acid biosynthesis [46] and results from the S-system model (red lines) with kinetic parameters inferred through *unconstrained forward estimation*. Analysis I: Initial values set at 120% of the steady state.
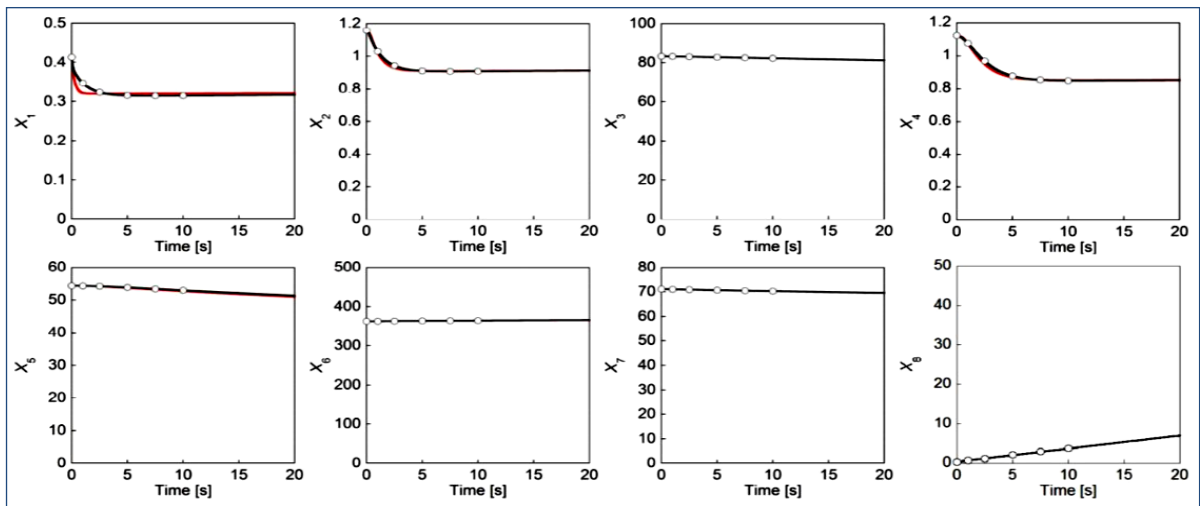

Fig. S7- Comparisons between "data" (symbols and connecting black lines) obtained from the model of aspartate-derived amino acid biosynthesis [46] and results the S-system model with kinetic parameters inferred through *constrained backward estimation* (red lines). In comparison with [Fig-7] of the Text, only the initial time range of 20 seconds is shown. Analysis I: Initial values set at 120% of the steady state.
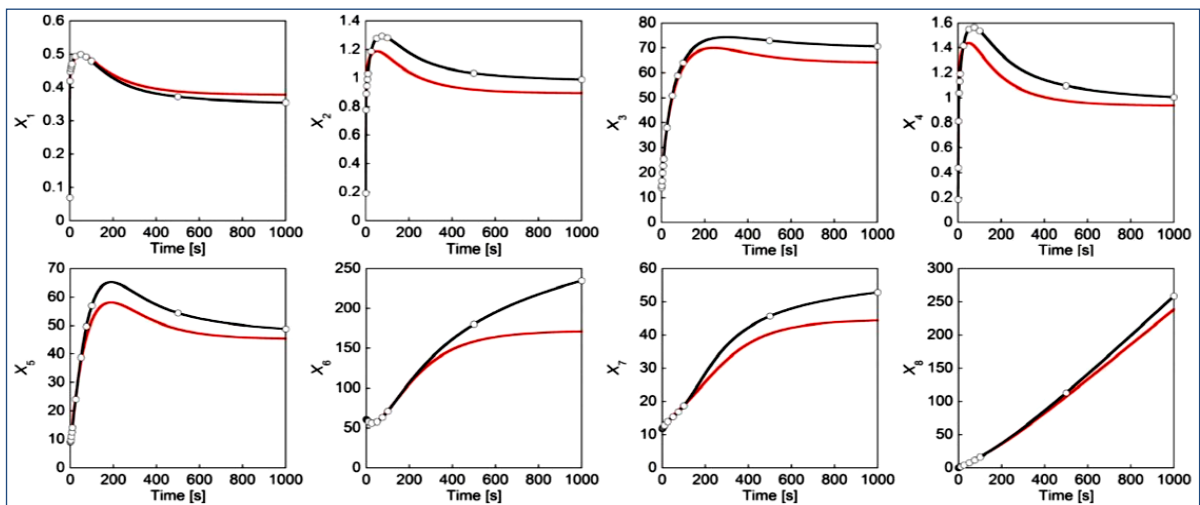
Fig. S8- Comparisons between "data" (symbols connected by black lines) obtained from the model of aspartate-derived amino acid biosynthesis [46] and results from the S-system model (red lines) with kinetic parameters inferred through *unconstrained backward estimation*. Analysis II: Initial values set at 20% of the steady state.

**Analysis III: Estimation of a Minimum Set of Fluxes to Complement DFE**

The pathway contains eight dependent variables and ten fluxes. Thus, we need to estimate two independent fluxes, which we select as described in the Text [Fig-9].
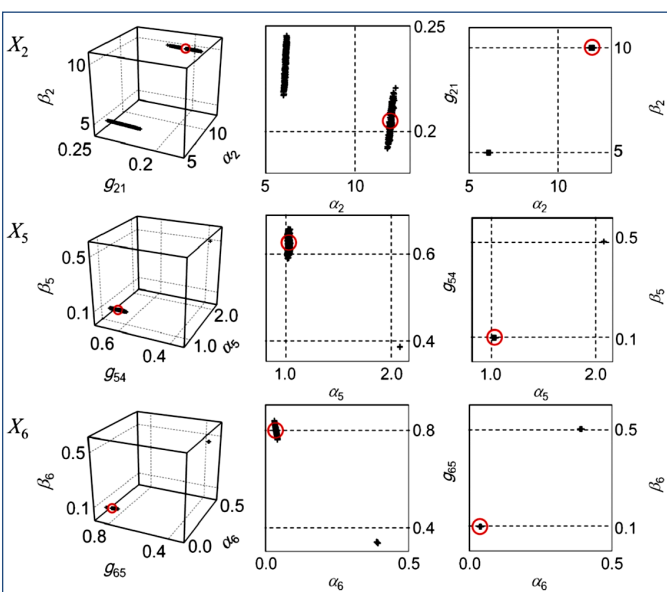
Fig. S9- Distributions of kinetic parameter values, inferred for aspartate-derived amino acid biosynthesis model with the *unconstrained backward estimation* strategy with 300 datasets with ± 10% noise. Red circle indicates the kinetic parameter values shown in [Table-S1]. Analysis II: Initial values set at 20% of the steady state. Left panel: Pseudo-three-dimensional representations of α_i-g_{ij}-β_i; center panel: α_i-g_{ij} projections; right panel: α_i-β_i projections.

The case where "data" were obtained with initial values at 120% of the steady state is shown in the Text. [Fig-S10] shows the corresponding results for "data" that are initialized at 20% of the steady state values in comparison to results from the S-system model with kinetic parameters inferred through constrained backward estimation. One notes that v_{ASADH}, which was used for DFE in Analysis I, cannot be calculated using this result. [Fig-S10], therefore, does not show the corresponding green plot lines. Similarly, red plot lines

cannot be computed for v_{AK} and v_{ASADH}. By contrast, v_{DHDPS} and v_{HSK}, which are now available, are shown as blue lines.
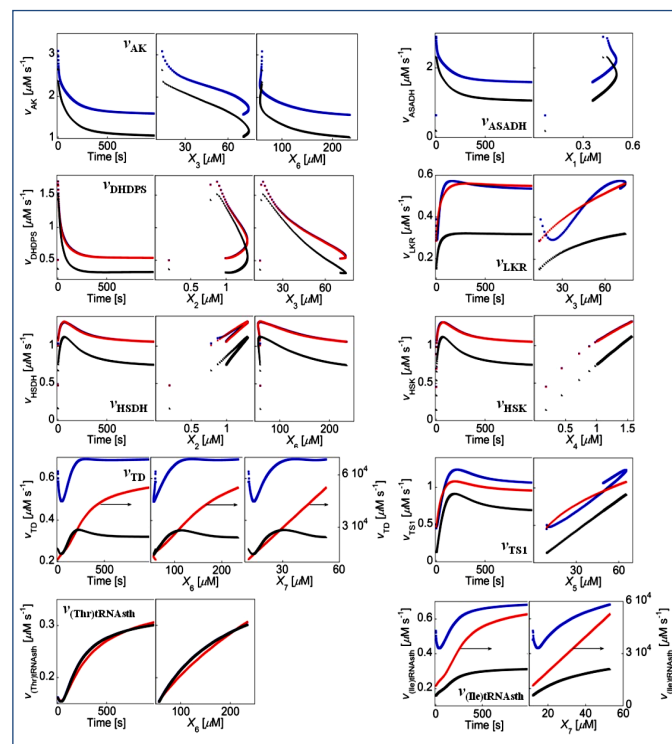
Fig. S10- Results of DFE, augmented with pairs of fluxes as shown in [Fig-9] of the Text. The fluxes obtained with DFE (blue) by and large have the shapes of the fluxes used by Curien et al. [46] to construct the model of aspartate-derived amino acid biosynthesis (black). However, they are shifted, thus indicating a remaining degree of freedom. The red flux graphs show the corresponding backward-inferred, constrained power-law representations (*cf.* [Fig-S7]). The small arrows indicate that the corresponding values are shown on the right y-axis.

[Table-S2] presents numerical results. The parameter values do not necessarily satisfy all constraints, because some parameter values cannot be obtained, as the corresponding fluxes are obtained with DFE.

Table S2- Parameter values and residual errors (SSE values) associated with different inference strategies for the model of aspartate-derived amino acid biosynthesis [Fig-2]. In Analysis II, the initial values set at 20% of the steady state.

	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	g_{i5}	g_{i6}	g_{i7}	g_{i8}	β_i	h_{ii}	SSE
					Backward estimation							
					Unconstrained inference							
X_1	NS	---	---	NS	---	---	NS	---	---	NS	NS	NS
X_2	NS	NS	---	---	---	---	---	---	---	NS	NS	NS
X_3	2.12E+01	---	8.96E-01	-8.63E-01	---	---	---	---	---	1.00E-01	4.00E-01	3.32E+00
X_4	1.50E+00	---	5.37E-01	---	---	---	-6.19E-02	---	---	1.06E+00	5.02E-02	2.44E-03
X_5	1.06E+00	---	---	---	5.02E-01	---	---	---	---	1.87E-01	4.20E-01	2.83E+02
X_6	1.87E-01	---	---	---	---	4.20E-01	---	---	---	7.67E-01	1.29E-05	2.30E+03
X_7	1.00E+03	---	---	---	---	---	1.02E-05	1.00E+00	---	1.00E+03	1.00E+00	1.41E+01
X_8	2.32E-02	---	---	---	---	---	4.73E-01	---	---	---	---	2.07E+01

References

[1] Hirai M.Y., Yano M., Goodenowe D.B., Kanaya S., Kimura T., Awazuhara M., Arita M., Fujiwara T. and Saito K. (2004) *Proceedings of the National Academy of Sciences*, 101(27), 10205-10210.

[2] Hirai M.Y., Klein M., Fujikawa Y., Yano M., Goodenowe D.B., Yamazaki Y., Kanaya S., Nakamura Y., Kitayama M., Suzuki H., Sakurai N., Shibata D., Tokuhisa J., Reichelt M., Gershenzon J., Papenbrock J. and Saito K. (2005) *Journal of Biological Chemistry*, 280(27), 22590-22595.

[3] Sawada Y., Akiyama K., Sakata A., Kuwahara A., Otsuki H., Sakurai T., Saito K. and Hirai M.Y. (2009) *Plant and Cell Physiology*, 50(1), 37-47.

[4] Neves A.R., Ramos A., Nunes M.C., Kleerebezem M., Hugenholtz J., de Vos W.M., Almeida J. and Santos H. (1999) *Biotechnology and Bioengineering*, 64(2), 200-212.

[5] Neves R., Pool W.A., Kok J., Kuipers O.P. and Santos H. (2005) *FEMS Reviews Microbiology*, 29, 531-554.

[6] Fonseca L.L., Sanchez C., Santos H. and Voit E.O. (2011) *Molecullar BioSystems*, 7(3), 731-741.

[7] Fonseca L.L., Chen P.W. and Voit E.O. (2012) *Metabolites*, 2, 221-241.

[8] Sriyudthsak K. and Shiraishi F. (2010) *Journal of Biotechnology*, 149(3), 191-200.

[9] Sriyudthsak K. and Shiraishi F. (2010) *Mathematical Biosciences*, 228(1), 1-9.

[10] Sriyudthsak K. and Shiraishi F. (2010) *Industrial & Engineering Chemistry Research*, 49(20), 9738-9742.

[11] Sriyudthsak K. and Shiraishi F. (2010) *Industrial & Engineering Chemistry Research*, 49(5), 2122-2129.

[12] Chen P.W., Fonseca L.L., Hannun Y.A. and Voit E.O. (2013) *PLOS Computational Biology*, 9(5).

[13] Goel G., Chou I.C. and Voit E.O. (2008) *Bioinformatics*, 24(21), 2505-2511.

[14] Savageau M.A. (1969) *Journal of Theoretical Biology*, 25, 365-369.

[15] Savageau M.A. (1969) *Journal of Theoretical Biology*, 25, 370-379.

[16] Savageau M.A. (1970) *Journal of Theoretical Biology*, 26, 215-226.

[17] Savageau M.A. (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecullar Biology*, Addison-Wesley Pub. Co.

[18] Voit E.O. (2000) *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press.

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

[20] Voit E.O. (2013) *International Scholarly Research Network - Biomathematics*, 1-53.

[21] Chou I.C., Martens H. and Voit E.O. (2006) *Theoretical Biology and Medical Modelling*, 3, 1-11.

[22] Kutalik Z., Tucker W. and Moulton V. (2007) *IET Systems Biology*, 1(3), 174-180.

[23] Liu P.K. and Wang F.S. (2008) *Bioinformatics*, 24(8), 1085-1092.

[24] Chou I.C. and Voit E.O. (2009) *Mathematical Biosciences*, 219(2), 57-83.

[25] Srinath S. and Gunawan R. (2010) *Journal of Biotechnology*, 149(3), 132-140.

[26] Jia G.S., Gregory N. and Gunawan R. (2011) *Bioinformatics*, 27(14), 1964-1970.

[27] Sriyudthsak K., Shiraishi F. and Hirai M.Y. (2013) *PLOS ONE*, 8(1).

[28] Varah J.M. (1982) *SIAM Journal on Scientific and Statistical Computing*, 3(1), 28-46.

[29] Voit E.O. and Savageau M.A. (1982) *Journal of Fermentation Technology*, 60(3), 233-241.

[30] Voit E.O. and Almeida J. (2004) *Bioinformatics*, 20(11), 1670-1681.

[31] Chou I.C., Martens H. and Voit E.O. (2007) *Statistics and Operations Research Transactions*, 31(1), 55-74.

[32] Kimura S., Hatakeyama M. and Konagaya A. (2005) *Chem-Bio Informatics Journal*, 4, 1-14.

[33] Vilela M., Borges C.C.H., Vinga S., Vasconcelos A.T.R., Santos H., Voit E.O. and Almeida J.S. (2007) *BMC Bioinformatics*, 8.

[34] Kimura S., Ide K., Kashihara A., Kano M., Hatakeyama M., Masui R., Nakagawa N., Yokoyama S., Kuramitsu S. and Konagaya A. (2004) *Bioinformatics*, 21(7), 1154-1163.

[35] Almeida J.S. and Voit E.O. (2003) *Genome Informatics*, 14, 114-123.

[36] Whittaker E.T. (1923) *Proceedings of the Edinburgh Mathematical Association*, 78, 81-89.

[37] Eilers P.H. (2003) *Analytical Chemistry*, 75(14), 3631-3636.

[38] Shiraishi F. and Savageau M. (1992) *Journal of Biological Chemistry*, 267(32), 22912-22918.

[39] Curto R., Sorribas A. and Cascante M. (1995) *Mathematical Biosciences*, 130(1), 25-50.

[40] Heinrich R. and Schuster S. (1996) *The Regulation of Cellular Systems*, Chapman & Hall.

[41] Gavalas G.R. (1968) *Nonlinear Differential Equations of Chemically Reacting Systems*, Springer-Verlag.

[42] Palsson B.O. (2006) *Systems Biology: Properties of Reconstructed Networks*, Cambridge University Press.

[43] Voit E.O., Goel G., Chou I.C. and Fonseca L.L. (2009) *IET Systems Biology*, 3(6), 513-522.

[44] Chou I.C. and Voit E.O. (2012) *BMC Systems Biology*, 6.

[45] Voit E.O. (2013) *Mathematical Biosciences*, DOI: 10.1016/j.mbs.2013.01.008.

[46] Curien G., Bastien O., Robert-Genthon M., Cornish-Bowden A., Cardenas M.L. and Dumas R. (2009) *Molecular Systems Biology*, 5.

[47] Sorribas A., March J. and Voit E.O. (2000) *Statistics in Medicine*, 19(5), 697-713.