Species Connectivities and Reaction Mechanisms from Neutral Response Experiments

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · APRIL 2007

Impact Factor: 2.69 · DOI: 10.1021/jp0661793 · Source: PubMed

5 AUTHORS, INCLUDING:

Federico Morán
Complutense University of Madrid
102 PUBLICATIONS **2,282** CITATIONS
SEE PROFILE

Juan Carlos Triviño
Sistemas Genómicos
18 PUBLICATIONS **27** CITATIONS
SEE PROFILE

John Ross
Stanford University
89 PUBLICATIONS **2,459** CITATIONS
SEE PROFILE

Species Connectivities and Reaction Mechanisms from Neutral Response Experiments

F. Morán,[†] M. O. Vlad,[*,‡,§] M. Bustos,[†] J. C. Triviño,[†] and J. Ross[‡]

Departamento de Bioquímica y Biología Molecular I, Universidad Complutense Madrid, 28040 Madrid, Spain, Department of Chemistry, Stanford University, Stanford, California 94305-5080, and Institute of Mathematical Statistics and Applied Mathematics, Casa Academiei Romane, Bucharest, Romania

Received: September 21, 2006; In Final Form: January 4, 2007

We develop a new method for obtaining connectivity data for nonlinear reaction networks, based on linear response experiments. In our approach the linear response is not the result of an approximation procedure but is due to the appropriate design of the response experiments, that is (1) they are carried out with the preservation of constant values for the total (labeled plus unlabeled) input and output fluxes and (2) the labeled compounds obey a neutrality condition (i.e., they have practically the same kinetic and transport properties as the unlabeled compounds). Under these circumstances the linear response equations hold even though the kinetics of the process is highly nonlinear. On the basis of this linear response law, we develop a method for evaluating reaction connectivities in biochemical networks from stationary response experiments. Given a system in a stationary regime, a pulse of a labeled species is introduced (with conservation of the total flux) and then the response of all the species of the network is recorded. The mechanistic information is contained in a connectivity matrix, **K**, which can be evaluated from the response data by means of differential as well as integral methods. The approach does not require any prior knowledge of the reaction mechanism. We carried out a numerical study of the method, based on a two-step procedure. Starting from a known reaction mechanism, we generated response data sets, to which we add noise; then, we use the noisy data sets for retrieving the connectivity matrix. The calculations were done with two programs written in Mathematica: the urea cycle and the upper part of glycolysis are used as sample biochemical networks. Given enough computer power, there are no limitations concerning the number of species involved in the response experiments; on current desktop systems processing responses of teens of species would take a few hours. The method is limited by the occurrence of experimental errors: if experimental errors in the evaluation of fluxes are larger than 10%, the method may fail to reproduce the correct values of some elements of the connectivity matrix.

1. Introduction

In our previous work we carried out a systematic theoretical analysis of a new type of response experiment in physical, chemical, biological kinetics, and population dynamics.[1−3] In our studies the response is linear, even though the underlying evolution equations are generally nonlinear. In the suggested experiments the linearity is due to the use of labeled species, which have the same kinetic and transport properties as the unlabeled species (neutrality condition), and not due to a linearization procedure. We have shown that this type of response experiment is useful and we presented different applications of our approach: for example, the study of fractal response experiments in desorption kinetics,[4] the study of on−off time distributions in single molecule kinetics,[5] and the analysis of geographical spread of mutations in human populations.[5,6]

In this paper we focus on the application of our approach to the qualitative and quantitative analysis of complex reaction systems, more precisely to the determination of the reaction connectivities and mechanisms from neutral response experiments. Our current approach shares common features with other types of response experiments developed recently.[7,8] The structure of the paper is the following. In section 2 we give a short summary of the method of neutral response experiments in homogeneous systems and elaborate a general approach for determining the reaction connectivities. In section 3 we present the simulation approach used in our research. In section 4 we give the results of the simulation approach applied to the urea cycle and the upper part of glycolysis. Finally in section 5 we consider the limits of our procedure as well as the possibilities of extending our approach to the quantitative analysis of reaction networks.

2. Species Connectivities from Response Experiments

We study a complex chemical system and consider a set of S species M_u, $u = 1, ..., S$ which can carry one or more identical molecular fragments; these are unchanged during the process and we refer to these species as "carriers". We limit ourselves to the case of isothermal, well-stirred, homogeneous systems, for which the concentrations $c_u = c_u(t)$, $u = 1, ..., S$, of the chemicals M_u, $u = 1, ..., S$, are space independent and depend only on time.[1] The kinetic equations of the process can be expressed in the following form:

$$dc_u(t)/dt = \mathcal{J}_u^+(t) - \mathcal{J}_u(t) + \rho_u^+(\mathbf{c};t) - \rho_u^-(\mathbf{c};t) \qquad u = 1, ..., S \quad (1)$$

where $\rho_u^{\pm}(\mathbf{c};t)$ are the rates of formation and consumption of the species M_u, respectively, $\mathcal{J}_u^{\pm}(t)$ are the input and output

[†] Universidad Complutense Madrid.
[‡] Stanford University.
[§] Casa Academiei Romane.

10.1021/jp0661793 CCC: $37.00
Published on Web 02/20/2007

fluxes of M_u, respectively, and

$$\mathbf{c}(t) = [c_u(t)]_{u=1,...,M} \tag{2}$$

is the composition vector of the system. Together with the initial condition

$$\mathbf{c}(t=t_0) = \mathbf{c}_0 \tag{3}$$

eq 1 determines the time evolution of the concentration vector.

We denote by z_u the number of fragments in the carrier species u. We use the notation

$$F_u = F(M_u) \qquad u = 1, ..., S \tag{4}$$

for a molecular fragment in the carrier M_u. All fragments F_u in different carriers M_u have the same structure; the label u means that a fragment belongs to a given carrier. The concentrations $f_u(t)$, $u = 1, ..., S$, of the fragments F_u, $u = 1, ..., S$, which belong to different carriers are given by

$$f_u(t) = z_u c_u(t) \qquad u = 1, ..., S \tag{5}$$

The kinetic equation (1) can be expressed in terms of the concentrations $f_u(t)$, $u = 1, ..., S$, of the fragment F_u, $u = 1, ..., S$, in the carrier u; we have

$$df_u(t)/dt = J_u^+(t) - J_u^-(t) + R_u^+(\mathbf{c};t) - R_u^-(\mathbf{c};t) \qquad u = 1, ..., S \tag{6}$$

where

$$R_u^\pm(\mathbf{c};t) = z_u \rho_u^\pm(\mathbf{c};t) \qquad u = 1, ..., S \tag{7}$$

are the rates of formation and consumption of the fragment F_u, in the carrier $u = 1, ..., S$, and

$$J_u^\pm(t) = z_u \mathcal{J}_u^\pm(t) \tag{8}$$

are input and output fluxes of fragment F_u, in the carrier $u = 1, ..., S$, respectively. A fragment is transferred from one carrier to another. These transfer processes involving fragments F_u, $u = 1, ..., S$, among different carriers can be represented as

$$F_u \underset{R_{uu'}}{\overset{R_{u'u}}{\rightleftharpoons}} F_{u'} \qquad u \neq u',\ u, u' = 1, ..., S \tag{9}$$

where $R_{u'u} = R_{u'u}(\mathbf{c};t)$ is the rate of transport of the fragment F_u from a carrier M_u to a carrier $M_{u'}$. The rates $R_{u'u}$ are related to the formation and consumption rates $R_u^\pm(\mathbf{c};t)$ of the fragment F_u in the carrier M_u by means of the balance equations

$$R_u^+(\mathbf{c};t) = \sum_{u' \neq u}^{S} R_{uu'}(\mathbf{c};t) \qquad R_u^-(\mathbf{c};t) = \sum_{u' \neq u}^{S} R_{u'u}(\mathbf{c};t) \qquad u = 1, ..., S \tag{10}$$

If the mechanism and kinetics of the process are known, then the transformation rates $R_{u'u} = R_{u'u}(\mathbf{c};t)$ can be evaluated by using the mass action law or other kinetic laws.

Now we can consider a kinetic tracer experiment by assuming that a fraction α_u, $u = 1, ..., S$, of the "in" flux $J_u^+(t)$ of the fragment F_u is replaced by a labeled fragment F_u^* and suppose that there is no kinetic isotope effect; that is, the rates of the processes involving labeled species are the same as the rates of the processes involving unlabeled species. We assume that the fractions β_u, $u = 1, ..., S$, of the labeled fragments in the output fluxes $J_u^-(t)$, $u = 1, ..., S$, can be measured experimentally.

We introduce the specific rates of transport of a fragment from one carrier to another and in and out of the system:

$$\omega_{uu'}(t) = R_{uu'}(t)/f_{u'} \tag{11}$$

$$\Omega_u^\pm = J_u^\pm(t)/f_u \tag{12}$$

Here $\omega_{uu'}(t)\, dt = dt\, R_{uu'}(t)/f_{u'}$ is the infinitesimal probability of transport of a fragment from the carrier $M_{u'}$ to the carrier M_u at a time between t and $t + dt$; similarly, $\Omega_u^\pm\, dt = dt\, J_u^\pm(t)/f_u$ is the infinitesimal probability that a fragment in the carrier M_u enters or leaves the system at a time between t and $t + dt$. If the kinetic isotope effect is missing, the rate of exchange of the labeled fragments in the system can be completely expressed in terms of these infinitesimal probabilities. We assume that the time dependences of the total rates $R_{u'u} = R_{u'u}(\mathbf{c};t)$ and $J_u^\pm(t) = z_u \mathcal{J}_u^\pm(t)$ attached to the total amounts of fragments from different carriers, labeled and unlabeled, are not changed during the process. We use the notations $f_u^*(t)$, $u = 1, ..., S$, for the concentrations of labeled fragments and $J_u^\pm{*}(t)$, $u = 1, ..., S$, for the input and output fluxes of labeled fragments, respectively. We use the kinetic isotope approach in the form suggested by Neiman and Gal[9] and derive the balance equations[1]

$$df_u^*(t)/dt = J_u^{+*}(t) - \Omega_u^-(t)\, f_u^*(t) + \sum_{u' \neq u}^{S} [\omega_{uu'}(t)\, f_{u'}^*(t) - \omega_{u'u}(t)\, f_u^*(t)] \tag{13}$$

and

$$J_u^{-*}(t) = \Omega_u^-(t)\, f_u^*(t) \tag{14}$$

We define the fractions of labeled fragments in the input fluxes

$$\alpha_u(t) = J_u^{+*}(t)/J_u^+(t) \tag{15}$$

and the fractions of labeled fragments in the output fluxes

$$\beta_u(t) = J_u^{-*}(t)/J_u^-(t) \tag{16}$$

From eqs 13−16 we can derive the following response laws:[1−2]

$$\beta_u(t) = \sum_{u'=1}^{S} \int_{-\infty}^{t} \chi_{uu'}(t;t')\, \alpha_{u'}(t')\, dt' \tag{17}$$

where $\chi_{uu'}(t;t')$ are non-negative susceptibility functions that fulfill the normalization condition:

$$\sum_{u'=1}^{S} \int_{-\infty}^{t} \chi_{uu'}(t;t')\, dt' = 1 \tag{18}$$

Two types of expressions have been derived for the susceptibility functions $\chi_{uu'}(t;t')$. Here we give only one set of relations, which depend on the Green functions $G_{uu'}(t,t') = [\mathbf{G}(t,t')]_{uu'}$, which are the solutions of the matrix differential equation

$$\frac{d}{dt}\mathbf{G}(t,t') = \mathbf{K}(t)\mathbf{G}(t,t') \text{ with } \mathbf{G}(t = t',t') = \mathbf{I} \tag{19}$$

where

$$\mathbf{K}(t) = [(1 - \delta_{uu'})\omega_{uu'}(t) - \delta_{uu'}[\Omega_u^-(t) + \sum_{u'' \neq u} \omega_{u''u}(t)]]_{u,u' = 1,\ldots,S} \quad (20)$$

is a connectivity matrix, which contains useful information about the structure of the reaction mechanism. The expressions for the susceptibility functions are[1−2]

$$\chi_{uu'}(t;t') = \frac{J_{u'}^+(t')\, G_{uu'}(t;t')}{f_u(t)} \quad (21)$$

We introduce the notation $\theta = t - t'$ for the transit time of a fragment and $\varphi_{uu'}(\theta;t)$ for its probability density, which fulfills the normalization condition

$$\sum_{u'=1}^{S} \int_0^t \varphi_{uu'}(\theta;t)\, \mathrm{d}\theta = 1 \quad (22)$$

We have $\varphi_{uu'}(\theta;t) = \chi_{uu'}(t;t-\theta)$; that is, the probability density of the transit time is given by the susceptibility function.

We limit ourselves here to stationary processes for which $\varphi_{uu'}(\theta;t) = \varphi_{uu'}(\theta)$ depends only on the transit time but is independent of current time. Under these circumstances, the response law (17) can be represented by a convolution product:

$$\beta_u(t) = \sum_{u'=1}^{S} \chi_{uu'}(t) \otimes \alpha_{u'}(t) = \sum_{u'=1}^{S} \varphi_{uu'}(t) \otimes \alpha_{u'}(t) \quad (23)$$

where $\otimes$ denotes the temporal convolution product. For stationary processes the reaction rates and the effective rate coefficients are also time independent. From eq 20 it follows that the matrix $\mathbf{K}(t)$ is also time independent, $\mathbf{K}(t) = \mathbf{K}$. From eq 19 it follows that the Green functions $G_{uu'}(t,t') = [\mathbf{G}(t,t')]_{uu'} = G_{uu'}(t-t') = [\mathbf{G}(t-t')]_{uu'}$ depend only on the time difference $t - t'$ and can be evaluated in terms of the exponential of the matrix $\mathbf{K}(t) = \mathbf{K}$. We have

$$\mathbf{G}(t-t') = \exp[\mathbf{K}(t-t')] \quad (24)$$

A real time or frequency response experiment makes it possible to evaluate the probability density $\varphi_{uu'}(t)$ of the transit time, which is at the same time the susceptibility function $\chi_{uu'}(t,t') = \varphi_{uu'}(t - t')$; further on, from eqs 24 we can evaluate the matrix **K**, which bears direct information about the connectivities of the reaction species: if a nondiagonal matrix element $K_{uu'}$ is different from zero, then there is a direct reaction pathway that leads to the transport of a fragment from the carrier $M_{u'}$ to the carrier M_u. In the following we illustrate this approach for evaluating the species connectivities by considering two simple biochemical reaction networks.

3. Simulation Approach

To test the possibilities of extracting connectivity information from the type of experiment suggested in the previous section, we consider a two-step simulation approach. (A) We simulate the response of a chemical system to a neutral perturbation and add noise to the response. (B) We use the simulated noisy data for retrieving the connectivity matrix.

As a simulation tool, we decided to use Mathematica[10] because of its unique ability to combine symbolic manipulations with numerical computations in the same program. Mathematica is a high-level programming language, which makes it possible to change various numeric and symbolic components of a program on the spot without debugging. These features outweigh some shortcomings due to the limited numerical capabilities of Mathematica.

Program A, which generates the noisy simulated response, is based on eqs 10−14 and 19−24. Starting from a set of input and reaction fluxes, we compute the **K** matrix and the matrix of the Green functions **G**; we limit ourselves to the case of stationary states for which the total input, output, and reaction fluxes are constant and only the excitation and response fractions are time dependent. In this case the matrix of the Green functions **G** is the exponential of the connectivity matrix **K** multiplied by the time difference $(t - t')$ (eq 24); fortunately, Mathematica has the ability to compute the exponential of a matrix symbolically. Finally from eqs 17 and 21 we can compute the response of the system to a given excitation; for simplicity, in most cases we assume that the excitations are given by delta functions and thus the responses are simply proportional to the susceptibility functions $\chi_{uu'}(t,t') = \varphi_{uu'}(t-t')$. Finally, we add multiplicative Gaussian noise to the response functions.

Program B, which extracts the connectivity matrix from the noisy response data, is based on the same equations with some adaptations. Starting from the noisy response, we can estimate the matrix of the Green functions from eq 21. To get the connectivity matrix from the Green functions, we apply two different types of methods:

(a) a method based on the observation that eq 24 can be viewed as the solution of the matrix differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{G}(\theta) = \mathbf{K}\mathbf{G}(\theta) \qquad \mathbf{G}(0) = \mathbf{I} \qquad \theta = t - t' \quad (25)$$

from which we obtain

$$\mathbf{K} = \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbf{G}(\theta)\right][\mathbf{G}(\theta)]^{-1} \quad (26)$$

The method based on eq 26 is similar to the differential methods in conventional chemical kinetics; because it involves the evaluation of the derivatives of the Green functions with respect to the transit time $\theta = t - t'$, it is sensitive to experimental errors. To eliminate the noise, we use fit functions for the Green functions. We have used both polynomial fits as well as fits based on the use of a linear combination of exponential functions. The use of exponential fits has a theoretical justification: if the connectivity matrix has simple eigenvalues, then, according to the Sylvester theorem,[11] it follows that the Green functions are combinations of exponential functions: $\sum_u A_u \exp(\lambda_u \theta)$, where A_u are amplitude factors and λ_u are the eigenvalues of the secular equation attached to the connectivity matrix, $\det|\mathbf{K} - \lambda\mathbf{I}| = 0$. The analytic expressions obtained from fitting are introduced into eq 26, which is evaluated symbolically by Mathematica for an arbitrary positive transit time θ. Further on, estimates of the connectivity matrix are evaluated repeatedly from eq 26 for various transit times. Finally, an average value of the connectivity matrix is evaluated from these estimates.

(b) a method based directly on eq 24 written in the following form:

$$\mathbf{K} = \frac{1}{\theta} \ln G(\theta) \quad (27)$$

which is also applied repeatedly for various values of the transit times. Mathematica can evaluate symbolically the logarithm of a matrix. A slightly different approach, which can be applied if

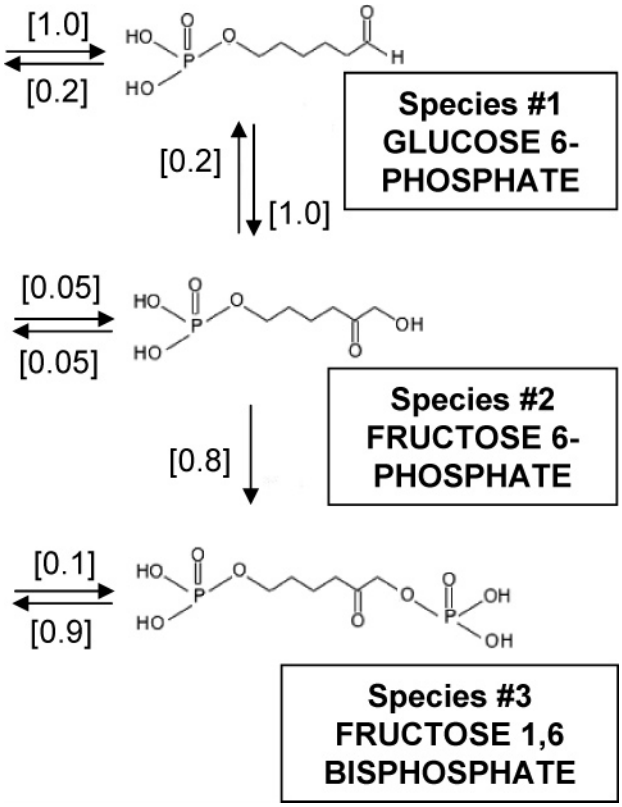

Figure 1. Schematic representation of the upper part of glycolysis. The concentrations of the three species are 2.5, 0.6, and 5.5 mM, respectively. The net flux though the network is 0.8 mM/min. All the fluxes are shown in square brackets [mM/min].

the eigenvalues of the connectivity matrix are simple, is based on the equation

$$\mathbf{K} = \frac{1}{\theta}\sum_u \ln(\mu_u)\prod_{v\neq u}\left[\frac{\mathbf{G}(\theta) - \mathbf{I}\mu_v}{\mu_u - \mu_v}\right] = \sum_u \lambda_u \prod_{v\neq u}\left[\frac{\mathbf{G}(\theta) - \mathbf{I}\exp(\lambda_v\theta)}{\exp(\lambda_u\theta) - \exp(\lambda_v\theta)}\right] \quad (28)$$

where

$$\mu_u = \exp(\lambda_u\theta) \quad (29)$$

are the eigenvalues of the matrix of the Green functions. Equation 28 can be derived by using the Sylvester theorem.[11]

The connectivity information can be easily extracted from the **K** matrix. If the nondiagonal matrix element $K_{uu'}$ is different from zero, then there is a direct reaction pathway, which leads to the transport of a fragment from the carrier $M_{u'}$ to the carrier M_u. The connectivity matrix can be also used for extracting quantitative kinetic information; although this is beyond the scope of the present article, a few possibilities are briefly mentioned in section 5.

Before applying the method to real biochemical networks, we carried out a numerical study of our approach on different reaction networks for various values of the input, output, and reaction fluxes and tried to identify what difficulties may occur. Here is a summary of our observations:

(1) In the case of noisy data, exponential fits work better than the polynomial fits, which tend to produce fake oscillations of the response curves.

(2) The exponential fit has its own limitations, which can be easily identified from eq 24. In some cases, even though from the kinetic point of view the system is at a stationary state, the response in the labeled fractions can display a slight, damped oscillatory behavior; in this case the pure exponential fit should be replaced by an exponential-trigonometric fit. Although rather unlikely, for some symmetrical cases multiple eigenvalues may occur; in such a case the exponential functions should be modulated by polynomial terms. In both cases, the fitting becomes too complicated and polynomial fits may provide a simple solution (provided that the noise is small).

(3) A shortcoming of all fitting techniques is that for large times they produce artificial results. In particular, in the case of differential methods based on eq 26, for large transit times we get spurious singularities, which originate in the canceling of the determinant of the matrix of the Green functions. Fortunately, these spurious singularities can be easily eliminated by examining the plots of the matrix elements as functions of transit time. Usually, for small to moderate transit times there is a region without singularity which can be safely used for evaluating the connectivity matrix

4. Application to the Upper Part of Glycolysis and the Urea Cycle

In this section we illustrate the method developed in the previous section with two biochemical examples: the upper part of the glycolytic pathway and the urea cycle. In experiments any type of excitation can be used. From eq 23, through numerical deconvolution, it is possible to extract the elements $\chi_{uu'}(t-t') = \varphi_{uu'}(t-t')$ of the susceptibility matrix. In particular, for unitary delta excitations of the type

$$\alpha_{u'}(t) = \mathcal{A}^{(0)}_{u'}\delta_{u'u_0}\delta(t) \quad (30)$$

where $\mathcal{A}^{(0)}_{u'}$ are amplitude factors with physical dimension [Time], $\delta_{u'u0}$ are Kronecker symbols and $\delta(t)$ is Dirac's unitary delta function. For an excitation of type (30) the response output fluxes attached to the different species u are proportional to the susceptibility functions. From eqs 23 and 30 we have

$$\beta_u(t)|_{u_0} = \mathcal{A}^{(0)}_{u_0}\chi_{uu_0}(t) \quad (31)$$

that is, for an excitation the species u_0 with a unitary delta impulse, the response of species u is proportional to the susceptibility function $\chi_{uu_0}(t)$ and the proportionality factor is the amplitude $\mathcal{A}^{(0)}_{u_0}$. For simplicity, without loss of generality, for the experiments simulated in this section we assume that the excitation and response functions are given by eqs 30 and 31.

For each example the response curves (the susceptibility functions) are calculated using the method presented in section 2. The curves are generated with a certain amount of noise to simulate actual experimental data. We carried out many simulations to explore the effects of different noise intensities, different fitting methods used as well as various possible approaches for evaluating the connectivity matrix. In this section we present the results of two typical simulations. In the examples shown here, a multiplicative noise factor of average relative intensity of 10% has been used. For larger values of the noise the method fails to reproduce the correct size of a few components of the connectivity matrix (typically, one, maximum two).

Figure 1 shows the scheme of the three reactions taken in the case of the glycolysis. This is a simple reaction network, with three species, and two enzymatic reactions, the first one

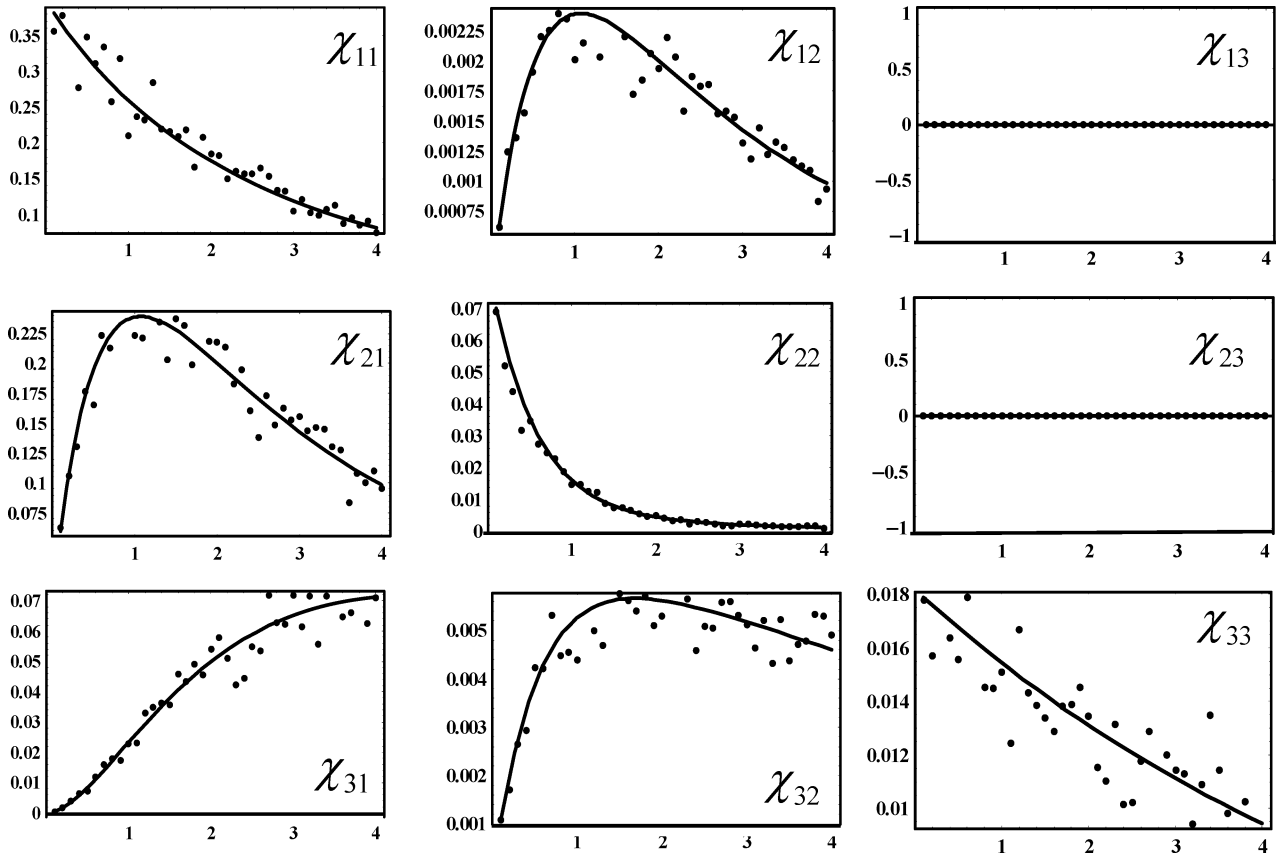

Figure 2. Graphical representation of the elements of the susceptibility matrix [min^{-1}] for the upper part of the glycolysis as functions of the transit time [min]. The susceptibility functions are computed from pulse excitations of the type described by eq 30 for which the responses of different species are proportional to the various elements of the susceptibility matrix (eq 31). Dots represent the simulated noisy data and the continuous curves the corresponding exponential fits.

being reversible. To produce the simulated response curves, it is necessary to know: the reactions rates of transformation of the species of the network; the concentration of each species; and the input and output fluxes of each one. The data for the simulation of a tracer-pulse experiment have been taken from the literature (see Teusink et all[12] and Crawford and Blum[13]) and the database SABIO-RK (Wittig et al.[14]). It is obvious that kinetic and concentration data vary from one biological system to another (species, organism, tissue, experimental or biological conditions, etc.). Here we have taken a typical set of values for these data. (The values in mM are indicated in the caption of Figure 1.) The reaction fluxes are indicated in square brackets in Figure 1, in mM/min.

The original **K** matrix computed by applying our first simulation program (A) is

$$\mathbf{K}_{\text{original}} = \begin{bmatrix} -0.48 & 0.33 & 0.00 \\ 0.40 & -1.75 & 0.00 \\ 0.00 & 1.33 & -0.16 \end{bmatrix} \tag{32}$$

From this original matrix the response curves can be calculated. In Figure 2 the dots represent the simulated susceptibility functions for the different species. The simulated data contain multiplicative Gaussian random errors of a relative average intensity of 10%. The solid lines show the best exponential fitting of the data. From these curves, following the method described above we extract the following connectivity matrix **K**:

$$\mathbf{K}^{\text{exponential}}_{\text{reproduced}} = \begin{bmatrix} -0.46 & 0.31 & 0.00 \\ 0.37 & -1.81 & 0.00 \\ -0.01 & 1.48 & -0.21 \end{bmatrix} \tag{33}$$

For polynomial fitting (not shown) the resulting matrix is

$$\mathbf{K}^{\text{polynomial}}_{\text{reproduced}} = \begin{bmatrix} -0.45 & 0.34 & 0.00 \\ 0.38 & -1.71 & 0.00 \\ -0.01 & 1.46 & -0.18 \end{bmatrix} \tag{34}$$

From these results it can be inferred that there is no reaction connection between species 3 to 1 and 3 to 2. This can be directly noted from Figure 2, where the susceptibility functions χ_{31} and χ_{32} are zero. The matrix element K_{31} has a small negative value but no zero. This can be also interpreted as a nondirect connection between species 1 and 3. In other simulations (not shown) with lower noise factor the matrix element K_{31} is zero.

In Figure 3 we give a schematic representation of the reaction network of the urea cycle. The invariant part of the molecule is represented with solid bonds. In eukaryotic cells the urea cycle takes place in different cell compartments, part of the reactions in the cytoplasm and part in mitochondria (see Maher et al.[15]). In our simulations we ignore the compartments and consider a homogeneous system. All the enzyme reactions of the cycle are reversible. The concentrations of the metabolites of the cycle vary substantially as a function of the organism, tissue, or experimental conditions. Here we have used concentrations from the literature (see Maher et al.[15]) and the data base SABIO-RK (Wittig et al.[14]). The metabolite concentrations (indicated in the caption of Figure 3, mM) have been evaluated by maintaining relative proportions. The fluxes chosen in the simulated experiment lead to a clockwise net flux (shown in Figure 3, in

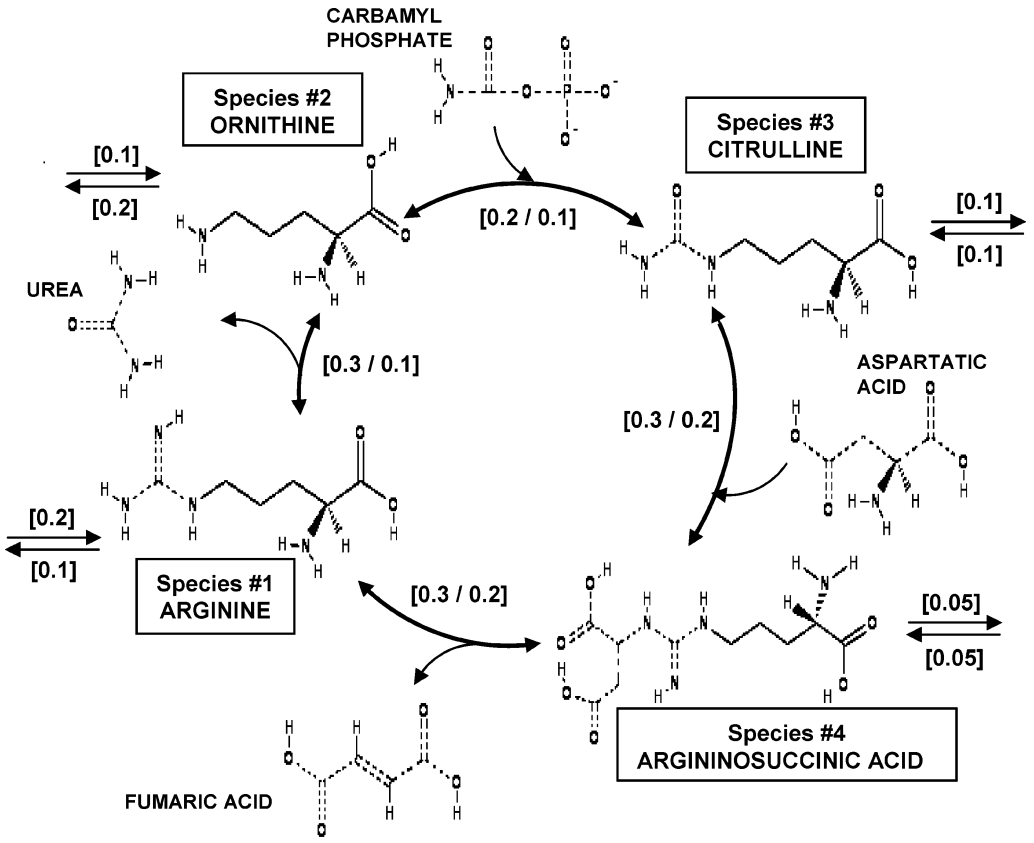

Figure 3. Schematic representation of the urea cycle. The concentrations of the four species of the cycle are 0.5, 1.0, 0.8, and 1.0 mM, respectively. The figure displays pairs of reaction fluxes in the following form: [clockwise-flux/counterclockwise-flux]. The net flux throughout the cycle is 0.1 mM/min and the net production of urea is 0.1 mM/min. The input and output fluxes are shown in square brackets [mM/min].

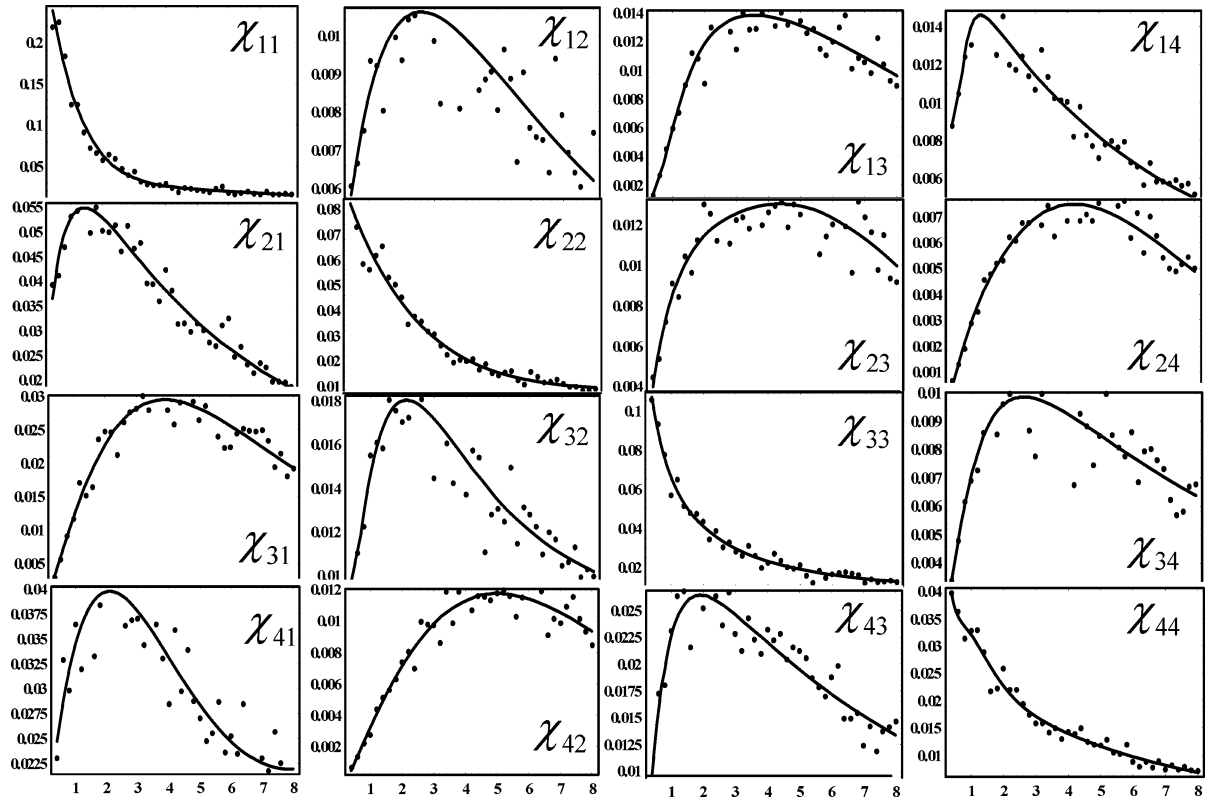

Figure 4. Graphical representation of the elements of the susceptibility matrix [min^{-1}] for the urea cycle example as functions of the transit time [min]. Dots represent the simulated noisy data and the continuous curves the corresponding exponential fits.

mM/min). Most of the arginosuccinic acid is bound to the enzyme (an effect referred to as 'tunneling', by Maher et al.[15]) and thus its concentration is not used explicitly in computations; this is possible because in our simulated experiments this metabolite has low input and output rates and its effective concentration is close to saturation.

For our simulated experiment the original connectivity matrix **K** computed with our first simulation program (A) is

$$\mathbf{K}_{\text{original}} = \begin{bmatrix} -1.2 & 0.1 & 0.0 & 0.3 \\ 0.6 & -0.5 & 0.125 & 0.0 \\ 0.0 & 0.2 & -0.625 & 0.2 \\ 0.4 & 0.0 & 0.375 & -0.55 \end{bmatrix} \quad (35)$$

As in the example of the glycolysis, the elements with value 0.0 show that there is no direct transformation reaction between the corresponding species. From this original connectivity matrix we simulate the response curves. In Figure 4 the dotted curves represent a set of realizations of the 16 susceptibility functions of the system computed for an average Gaussian noise intensity of 10% and the continuous curves represent the corresponding exponential fits.

The connectivity matrix extracted from this simulated experiment is

$$\mathbf{K}^{\text{exponential}}_{\text{reproduced}} = \begin{bmatrix} -1.13 & 0.09 & 0.00 & 0.27 \\ 0.54 & -0.48 & 0.13 & 0.01 \\ 0.00 & 0.19 & -0.73 & 0.21 \\ 0.46 & -0.01 & 0.41 & -0.59 \end{bmatrix} \quad (36)$$

for exponential fitting, and

$$\mathbf{K}^{\text{polynomial}}_{\text{reproduced}} = \begin{bmatrix} -1.17 & 0.10 & 0.00 & 0.31 \\ 0.57 & -0.48 & 0.14 & 0.00 \\ 0.00 & 0.20 & -0.73 & 0.22 \\ 0.46 & -0.01 & 0.40 & -0.53 \end{bmatrix} \quad (37)$$

for polynomial fitting.

Even with 10% noise the connectivity matrix is reproduced with a reasonable accuracy and from it the structure of the reaction network can be determined. The elements K_{31},K_{42}, K_{13}, and K_{24} are zero or negative close to zero: this implies that there is no connection between species 1 and 3 on the one hand and 2 and 4 on the other, which is evidence of a cyclic reaction network with all the reactions reversible.

5. Discussion

We start out by outlining the limitations of our approach. Given enough computing power, there are no limitations regarding the numbers of species involved in a response experiment; we must point out that the time necessary for extracting the connectivity matrix from response data increases with the number of species. We carried out our computations on a 64 Bit AMD computer at 2.4 GHz, under Ubuntu Linux 64Bit and Windows XP 32Bit. For both operating systems a 3−4 species computation with exponential fits varies between 10 and 30 min with an average of 15 min; the computations based on polynomial fits are much faster. Because curve fitting is the slowest computation process, the time required for the evaluation of the connectivity matrix is approximately proportional to the square of the number of species. A more serious limitation is the size of experimental errors; the method works well for experimental errors up to 10% but for higher errors, e.g., up to 20%, fails to reproduce few elements of the connectivity matrix (typically one or two) for errors higher than 20% usually the method fails completely; in general, the exponential fits produce better results than the polynomial fits.

In our current approach the statistical treatment of noisy data is rather crude. For higher errors our method can be improved by using statistical estimation methods,[16] based on likelihood approach or on Bayesian methods; we are currently working on a method based on the likelihood approach.

In addition to qualitative information about a reaction network, the connectivity matrix also contains quantitative information and can be used for obtaining kinetic laws and rate coefficients from experimental data. In general, such studies require the repetition of response experiments for different concentrations of the active species, according to eqs 11 and 20 the nondiagonal components, of the connectivity matrix, $K_{uu'} = \omega_{uu'} = R_{uu'}/f_{u'}$, $u \neq u'$, are specific rates of transfer of a fragment from one carrier to another, which for stationary experiments are time independent and generally concentration dependent. From repeated experiments it is easy to evaluate the concentration dependence of the rates $R_{uu'}$, from which we can evaluate kinetic laws and parameters (reaction orders, rate coefficients, activation energies, etc.).

An alternative approach for kinetic studies is based on the study of the concentration and temperature dependence of the eigenvalues λ_u which are the solutions of the secular equation attached to the connectivity matrix, $\det|\mathbf{K} - \lambda\mathbf{I}| = 0$, and can be determined by means of exponential fits of the response data. As the dependence of the eigenvalues on the concentrations and other parameters, such as temperature, can be extremely complicated, it is easier to study the variations of a set of the tensor invariants[17] built from the eigenvalues λ_m

$$\mathscr{J}_w = \sum_{\nu_1} \dots \sum_{\nu_w} \lambda_{\nu 1} \dots \lambda_{\nu w} \quad (38)$$

Typically, the dependences $\mathscr{J}_w = \mathscr{J}_w$(concentrations) can be expressed by polynomials, where the coefficients of the concentrations depend on the rate coefficients of the process. Starting from different assumed reaction mechanisms, we can derive theoretical expressions for the dependence on the invariants on the concentrations. By checking the validity of these dependences for the experimental data, we can test the validity of the assumed reaction mechanisms. By repeating the experiments at different temperatures, we can determine the activation energies corresponding to different reaction steps, etc.

Acknowledgment. This research has been supported in part by the National Science Foundation, the CEEX-M1−C2-3004/2006 Grant of the Romanian Ministry or Research and Education, and by the Grant BMC2003-06957 from MEC (Spain) and Beca Complutense Del Amo 2005/2006.

References and Notes

(1) Vlad, M. O.; Moran, F.; Rodriguez, Y.; Ross, J. *Int. J. Bif. Chaos* **2004**, *12*, 2599.

(2) Vlad, M. O.; Moran, F.; Tsuchiya, M.; Cavalli-Sforza, L. L.; Oefner, P. J.; Ross, J. *Phys. Rev. E* **2002**, *E65*, 061110(1−17).

(3) Vlad, M. O, Arkin, A.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7223. Vlad, M. O.; Cavalli-Sforza, L. L.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10249. Vlad, M. O.; Szedlacsek, S. E.; Pourmand, N. Cavalli-Sforza, L. L.; Oefner, P. J.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9848.

(4) Vlad, M. O.; Cerofolini, G. F.; Ross, J. *Phys. Rev. E* **2000**, *E62*, 937. Vlad, M. O.; Moran, F.; Ross, J. *J Phys. Chem. B* **2001**, *105B*, 11710.

(5) Vlad, M. O.; Moran, F.; Schneider, F. W.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 12548. Vlad, M. O.; Moran, F.; Ross, J. *Chem. Phys.* **2003**, *287*, 83.

(6) Vlad, M. O.; Cavalli-Sforza, L. L.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10249. Vlad, M. O.; Szedlacsek, S. E.; Pourmand, N.; Cavalli-Sforza, L. L.; Oefner, P. J.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 9848.

(7) Vance, W.; Arkin, A.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5816.

(8) Torralba, A.; Yu, K.; Peidong, S.; Oefner, J. P.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1494.

(9) Neiman, M. B.; Gál, D. *The Kinetic Isotope Method and Its Application*; Akadémic Kiadó: Budapest, 1971.

(10) Wofram, S. *Mathematica Book*, 4th ed.; Wolfram Media: Champaign, IL, 1999.
(11) Frazer, R. A.; Duncan, W. J.; Collar, A.R. *Elementary Matrices*; Cambridge University Press: Oxford, U.K., 1965; Chapters 2−3.
(12) Teusink, B.; Passarge, J.; Reijenga, C. A.; Esgalhado, E.; van der Weijden, C. C.; Schepper, M.; Walsh, M. C.; Bakker, B. M.; van Dam, K.; Westerhoff, H. V.; Snoep, J. L. *Eur. J. Biochem.* **2000**, *267*, 5313.
(13) Crawford, J. M.; Blum, J. J. *Biochem. J.* **1983**, *212*, 595.
(14) Wittig, U.; Golebiewski, M.; Kania, R.; Krebs, O.; Mir, S.; Weidemann, A.; Anstein, S.; Saric, J.; Rojas, I. *Lecture Notes Bioinformatics* **2006**, *4075*, 94.
(15) Maher, A. D.; Kuchel, P. W.; Ortega, F.; Atauri, P.; Centelles, J.; Cascante, M. *Eur. J. Biochem.* **2003**, *270*, 3953.
(16) Frieden, B. R.; *Probability, Statistical Optics and Data Testing*, 3rd ed.; Springer: Berlin, 2001.
(17) Aris, R. *Vectors, Tensors and the Basic Equations of Fluid Mechanics*; Dover: New York, 1990.