

Numerical Matrices Method for Nonlinear System Identification and Description of Dynamics of Biochemical Reaction Networks

Alexey V. Karnaukhov,* Elena V. Karnaukhova,* and James R. Williamson[†]

*Institute of Cell Biophysics Russian Academy of Sciences, Pushchino, Russia; and [†]Departments of Molecular Biology and Chemistry, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California USA

ABSTRACT A flexible Numerical Matrices Method (NMM) for nonlinear system identification has been developed based on a description of the dynamics of the system in terms of kinetic complexes. A set of related methods are presented that include increasing amounts of prior information about the reaction network structure, resulting in increased accuracy of the reconstructed rate constants. The NMM is based on an analytical least squares solution for a set of linear equations to determine the rate parameters. In the absence of prior information, all possible unimolecular and bimolecular reactions among the species in the system are considered, and the elements of a general kinetic matrix are determined. Inclusion of prior information is facilitated by formulation of the kinetic matrix in terms of a stoichiometry matrix or a more general set of representation matrices. A method for determination of the stoichiometry matrix beginning only with time-dependent concentration data is presented. In addition, we demonstrate that singularities that arise from linear dependencies among the species can be avoided by inclusion of data collected from a number of different initial states. The NMM provides a flexible set of tools for analysis of complex kinetic data, in particular for analysis of chemical and biochemical reaction networks.

INTRODUCTION

An extensive theoretical framework has been developed for analysis of nonlinear dynamic systems, in particular, for the case of chemical reaction kinetics (1). A variety of theoretical and computational approaches for studying complex reaction networks have been developed and described (1–3). Most recently, advances in genomics have made the understanding of the complex biochemical reaction networks of metabolism, gene regulation, and cell function an active area of investigation (4,5). Many of these approaches make extensive use of a stoichiometry matrix that describes the set of reactions present in the system. Analysis of the null spaces and the span of a stoichiometry matrix provides extensive information about the system, including steady-state solutions, elementary flux modes, and conservation of mass relations (4,6–9). For cases where the stoichiometry matrix is not known, correlation methods have been applied to deduce the structure of a reaction network from time-dependent data for the chemical species (10–12). A complete understanding of metabolic networks will require methods to determine the network structure, the resulting steady states and fluxes for the network, and the transient response of a system away from the steady state.

In this work, the theoretical basis for a set of numerical tools is developed to practically address two important problems encountered in analysis of biochemical reaction networks. The approach is conceptually distinct from previous work based on the properties of the stoichiometry matrix or correlation methods (4–12). Here, we present an approach

that both solves the reaction identification problem, and provides a least squares analysis to determine the values of the rate constants that best fit a given reaction structure to the time-dependent data. The tools presented here represent a powerful “bottom-up” approach to determine reaction network structure and to quantitatively analyze reaction network dynamics.

The reaction identification problem

Experimentally observed rates and concentrations can be used to identify the underlying nature of the chemical reactions present. Solving the identification problem involves finding the elementary reaction steps and kinetics that give rise to the observed data without any prior knowledge of the reaction chemistry. The basis of most of the modern methods of system identification was developed during the period 1950–1970. These methods have been successfully used to obtain the rate constants for nontrivial chemical reaction systems with several species.

A set of time-dependent data for the concentrations of the species can be represented as $X_j(t_i)$, where the index i runs from 1 to nt , which is the number of time points, and the index j runs from 1 to ns , which is the number of species. In addition, for each species and time point, there is a matrix of observed rates $dX_j(t_i)/dt$ with the same indices. In practice, the rates may be either directly observed, or calculated from the observed concentration data, $X_j(t_i)$. A set of np chemical reactions is generally described by Cauchy equations in the form:

$$\frac{dX_j(t_i)}{dt} = \sum_{p=1}^{np} k_p f_{pj}(X_1(t_i), X_2(t_i) \dots X_{ns}(t_i)), \quad (1)$$

Submitted July 14, 2006, and accepted for publication January 19, 2007.

Address reprint requests to James R. Williamson, E-mail: jrwill@scripps.edu.

© 2007 by the Biophysical Society

0006-3495/07/05/3459/15 \$2.00

doi: 10.1529/biophysj.106.093344

where the k_p are the rate constants, and the f_{pj} are the empirical functions that contain information about structure of the chemical reaction, for each reaction p . A number of approaches have been developed to solve sets of equations such as Eq. 1 for the rate constants, including empirical optimization (13–16), direct integral methods (13,17–22), and the Prony method for quasilinear systems (23).

A direct differential approach has been developed (24–31) to obtain rate constants from observed rates of change of the species using a least squares criterion:

$$\phi = \sum_{i=1}^{ni} \sum_{j=1}^{ns} \left[\left(\frac{dX_j(t_i)}{dt} \right)_{\text{obs}} - \left(\frac{dX_j(t_i)}{dt} \right)_{\text{calc}} \right]^2, \quad (2)$$

where the observed value of the derivatives is obtained from directly or by numerical differentiation of observed value of concentrations and the calculated value of derivatives is obtained from Eq. 1. Differentiation of Eq. 2 with respect to each rate parameter k_p , gives a system of linear equations that can be directly solved for the rate constants k_p .

Chemical reaction network theory

A general formalism for chemical reaction network theory (32,33) has been developed based on the idea of reaction complexes, which are the combinations of species that are reaction products and reactants. Each species is represented by a unit vector \hat{x}_i , linear combinations of species that are products or reactants are represented by complex vectors \hat{y}_j , and reactions are represented by differences between two complex vectors as vectors \hat{r}_q , as outlined in Appendix I in the Supplementary Material. Any network of chemical reactions can be described as a set of reaction vectors that constitute a stoichiometry matrix. The dynamics of a network of species can be described using the stoichiometry matrix, and the properties of the stoichiometry matrix can be analyzed to identify steady states and steady-state fluxes for the reaction network (6).

Despite the extensive developments in nonlinear system identification and chemical reaction network theory, there has been no strong connection made between these two important areas. In this article, we make such a theoretical connection by formulating the Numerical Matrices Method (NMM) for nonlinear system identification in terms of the formalism of reaction complexes, and presenting a general method for de novo determination of the stoichiometry matrix. First, the previously developed method for nonlinear system identification (34,35) has been formulated as the Kinetic Matrix Method (KMM) using the formalism of kinetic complexes that provides a key connection to the other matrix representations of the kinetic equations. Second, a variant of the KMM is described that incorporates varying degrees of prior information as the Representation Matrix Method (RMM), which result in increased accuracy of the determined rate constants. Third, the Stoichiometry Matrix Method (SMM) is presented as a general method for determination of rate constants where the reaction network

structure is known. Fourth, a method for breaking the holonomic conditions arising from linear dependence among species is presented, which is essential for general application of the Numerical Matrices Method to large complex reaction networks. Finally, a general method is presented for construction of a stoichiometry matrix with the Numerical Matrices Method, using only concentration data as a starting point, without any prior knowledge of the reaction network structure. The approach outlined here provides a general and powerful set of analytical tools that will have a wide variety of applications in analysis of chemical reaction dynamics and metabolic networks.

RESULTS

The Numerical Matrices Method for nonlinear system identification

The direct integral, direct differential, and empirical optimization methods are not generally suitable for the task of investigating chemical reaction cascades with strong nonlinear dynamics, a large number of species, and unknown structure of the underlying chemical reactions. A direct differential method using a linearly quadratic kinetic model has been recently developed that can be used to solve the identification problem for these more demanding systems (34,35). Here we extend this approach to a more general set of related matrix methods that constitute a flexible and powerful set of tools for the quantitative analysis of nonlinear dynamic systems.

The linearly quadratic kinetic model

Considering a set of ns chemical species, there are a number of elementary reactions that might occur among them, including unimolecular, bimolecular, and higher order reactions, as well as mass flow in and out of the system. The vast majority of chemical and biochemical reactions can be represented as cascades of uni- and bimolecular reactions. If the unusual cases of trimolecular and higher order reactions are neglected, a general mathematical expression for the time dependence of a particular species j is given by:

$$\left(\frac{dX_j(t_i)}{dt} \right)_{\text{calc}} = \sum_{p=0}^{ns} \sum_{q=0}^p a_{jpq} X_p(t_i) X_q(t_i), \quad (3)$$

where the rate of change is described as a linear sum of flux terms, which are the rate parameters, a_{jpq} , times the appropriate concentrations of species p and q . To conveniently describe the kinetics for unimolecular reactions or mass flow in this quadratic expression, a fictitious species, X_0 , is introduced whose concentration is held constant. In this way, kinetics that depend linearly on a single species can be accounted for. It should be noted that the complete set of rate constants describing the kinetics should contain all of the combinations of indices p and q , but that Eq. 3 contains the additional index j . For the purpose of system identification, it

is necessary to treat each of the species j individually, as described below. The linearly quadratic form of Eq. 3 is very general for common reactions and avoids the ambiguity in specification of the arbitrary functions $f_{pj}(X_1(t_i), X_2(t_i) \dots X_{ns}(t_i))$ in Eq. 1.

Matrix formalisms for chemical reaction dynamics

The vector formalism from chemical reaction network theory provides useful and compact expressions for chemical reaction kinetics, and provides strong connections between different representations of the kinetics. As described in Appendices I and II in the Supplementary Material, the reaction dynamics from Eqs. 1 and 3 can be equivalently expressed in the following three forms:

$$\frac{d\hat{\mathbf{X}}}{dt} = \tilde{\mathbf{A}} \cdot \hat{\mathbf{X}}^{\tilde{\mathbf{Y}}} \quad \text{Kinetic matrix} \quad (4)$$

$$\frac{d\hat{\mathbf{X}}}{dt} = \sum_{p=1}^{np} k_p \tilde{\mathbf{G}}_p \cdot \hat{\mathbf{X}}^{\tilde{\mathbf{Y}}} \quad \text{Representation matrix} \quad (5)$$

$$\frac{d\hat{\mathbf{X}}}{dt} = \tilde{\mathbf{R}} \cdot \text{diag}(\hat{\mathbf{K}}) \cdot \hat{\mathbf{X}}^{\tilde{\mathbf{Y}}} \quad \text{Stoichiometry matrix,} \quad (6)$$

where $\hat{\mathbf{X}}$ is the vector of species, $\tilde{\mathbf{A}}$ is a generalized kinetic matrix, $\tilde{\mathbf{G}}_p$ is a set of representation matrices, $\tilde{\mathbf{R}}$ is a stoichiometry matrix formed from the set of reaction vectors, $\hat{\mathbf{K}}$ is a vector containing the rate constants k_p , and $\tilde{\mathbf{Y}}$ is a matrix formed from the set of complex vectors. The exponentiation of the species vector by the complex matrix, $\hat{\mathbf{X}}^{\tilde{\mathbf{Y}}}$, is a compact expression for the vector of kinetic complexes $\hat{\mathbf{F}}$, as described in Appendices I and III in the Supplementary Material.

The kinetic matrix $\tilde{\mathbf{A}}$ is most useful in the absence of any knowledge about the reaction structure, the representation matrices $\tilde{\mathbf{G}}_p$ are useful when there is partial information about the reaction structure or rates, while the stoichiometry matrix $\tilde{\mathbf{R}}$ is useful when the complete structure of the reaction network is known. Each of these representations offers advantages for analysis of reaction network kinetics in particular situation, and this set of dynamic equations serves as the basis for the Numerical Matrices Method.

The Kinetic Matrix Method

The dynamics of a cascade of uni- and bimolecular reactions is expressed in matrix form in Eq. 4, where $\tilde{\mathbf{A}}$ is a general kinetic matrix containing the elements a_{jpq} , and $\tilde{\mathbf{Y}}$ is the matrix of complexes. Because of the inclusion of the fictitious species X_0 in the linearly quadratic model, there are $ns + 1$ species, and $nk = (ns + 2)(ns + 1)/2$ possible rate constants, each of which corresponds to all possible reaction among the ns species. The dimensions of matrices $\tilde{\mathbf{A}}$, and $\tilde{\mathbf{Y}}$ are ns rows, one for each species, and nk columns, one for each kinetic complex.

The agreement of the observed data and model data calculated with Eq. 4 is quantitated by the least squares dis-

crepancy in analogy to Eq. 2. Differentiation of Eq. 2 with respect to each of the elements A_{ij} gives a linear system of nk equations for the $\hat{\mathbf{a}}_j$ (row j of matrix $\tilde{\mathbf{A}}$) as described in Appendix II in the Supplementary Material. The observed rates in Eq. 2 are calculated from the concentration data using finite differences as described in Appendix II in the Supplementary Material. The set of equations can be solved by defining the elements of matrix $\tilde{\mathbf{B}}$ and vector $\hat{\mathbf{c}}_j$:

$$\begin{aligned} \tilde{\mathbf{B}} \cdot \hat{\mathbf{a}}_j &= \hat{\mathbf{c}}_j \\ \tilde{\mathbf{B}} &= b_{kl} = \sum_{i=1}^{nt} (\hat{\mathbf{X}}(t_i))^{\hat{\mathbf{y}}_k} (\hat{\mathbf{X}}(t_i))^{\hat{\mathbf{y}}_l} \\ \hat{\mathbf{c}}_j &= c_{jk} = \sum_{i=1}^{nt} \left(\frac{dX_j(t_i)}{dt} \right)_{\text{obs}} (\hat{\mathbf{X}}(t_i))^{\hat{\mathbf{y}}_k}. \end{aligned} \quad (7)$$

The desired vector of rates $\hat{\mathbf{a}}_j$ is obtained by inversion of matrix $\tilde{\mathbf{B}}$. It is necessary to solve this system of equations for each of the ns species j separately to ensure the condition of $\tilde{\mathbf{B}}$. The results of this analysis are estimates for the rate constants for each of the possible quadratic combinations of species involved in all possible reactions that affect the concentration of species j . Solving for each $\hat{\mathbf{a}}_j$ in turn allows for the construction of the kinetic matrix $\tilde{\mathbf{A}}$, which we term the Kinetic Matrix Method for nonlinear system identification. Many of the A_{ij} are zero, and the nonzero values give information about which of these many possible reactions are occurring. The elements of $\tilde{\mathbf{B}}$ and $\hat{\mathbf{c}}_j$ are completely defined in terms of the concentration time-series data, the rates that must be obtained by differentiation of the concentration data, and the complex matrix $\tilde{\mathbf{Y}}$.

The Representation Matrix Method

An alternative approach to reducing the number of parameters to be determined from the data is to decompose the kinetic matrix $\tilde{\mathbf{A}}$ into the product of a set of representation matrix $\tilde{\mathbf{G}}_p$ and a vector of nonzero parameters $\hat{\mathbf{k}} = k_p$ that are to be determined. The representation matrices contain information about the complexes that are present in the dynamics. In this case, the dynamics takes the form of Eq. 5, and minimizing the least squares discrepancy function in Eq. 2 leads to the formulas similar to Eq. 7 for the vector of parameters $\hat{\mathbf{k}} = k_p$:

$$\begin{aligned} \tilde{\mathbf{B}} &= b_{pq} = \sum_{i=1}^{nt} \left(\tilde{\mathbf{G}}_p (\hat{\mathbf{X}}(t_i))^{\hat{\mathbf{Y}}} \right)^T \cdot \tilde{\mathbf{G}}_q (\hat{\mathbf{X}}(t_i))^{\hat{\mathbf{Y}}} \\ \hat{\mathbf{c}} &= c_q = \sum_{i=1}^{nt} \left(\left(\frac{d\hat{\mathbf{X}}(t_i)}{dt} \right)_{\text{obs}} \right)^T \cdot \tilde{\mathbf{G}}_q (\hat{\mathbf{X}}(t_i))^{\hat{\mathbf{Y}}} \\ \hat{\mathbf{k}} &= \tilde{\mathbf{B}}^{-1} \cdot \hat{\mathbf{c}}. \end{aligned} \quad (8)$$

The Stoichiometry Matrix Method

For the case of the dynamics in terms of the stoichiometry matrix in Eq. 6, the discrepancy function in Eq. 2 is

differentiated with respect to each of the elements of $\hat{\mathbf{K}}$, giving a system of linear equations in direct analogy to Eq. 7. The sparse nature of the stoichiometry matrix makes it numerically tractable to treat all ns species simultaneously, and the sum over all species j is retained in analogy to Eq. 2. This linear system of equations can be readily solved for $\hat{\mathbf{K}}$ by defining the elements of matrix $\tilde{\mathbf{B}}$ and vector $\hat{\mathbf{c}}$:

$$\begin{aligned}\tilde{\mathbf{B}}(\tilde{\mathbf{R}}) &= b_{pq} = \sum_{i=1}^{nt} \sum_{j=1}^{ns} R_{jp}(\hat{\mathbf{X}}(t_i))^{\hat{y}_{jp}} R_{jq}(\hat{\mathbf{X}}(t_i))^{\hat{y}_{jq}} \\ \hat{\mathbf{c}}(\tilde{\mathbf{R}}) &= c_p = \sum_{i=1}^{nt} \sum_{j=1}^{ns} \left(\frac{dX_j(t_i)}{dt} \right)_{\text{obs}} R_{jp}(\hat{\mathbf{X}}(t_i))^{\hat{y}_{jp}} \\ \hat{\mathbf{K}} &= \tilde{\mathbf{B}}^{-1} \cdot \hat{\mathbf{c}}.\end{aligned}\quad (9)$$

An important difference in Eq. 9 compared to Eq. 7 is that the matrix elements are determined by summing over all ns species. The stoichiometry matrix effectively matches the various complexes to the appropriate rate constants according to the reaction scheme, and ensures the good condition of $\tilde{\mathbf{B}}$. In this way, the set of rate constants that best describes the observed data is obtained in a closed form.

For all three matrix representations, the elements of $\tilde{\mathbf{B}}$ and $\hat{\mathbf{c}}$ are defined in terms of the concentration data, the rates that must be obtained from the concentration data, and the complex matrix $\tilde{\mathbf{Y}}$. These three matrix methods constitute the basic tools of the Numerical Matrices Method. Analysis of reaction kinetics and reaction structure is based on calculation of the Numerical Matrices Method described below.

Application of the Kinetic Matrix Method

In the case of a de novo nonlinear system identification problem, where there is no prior knowledge of the reaction dynamics, the NMM is implemented using the generalized kinetic matrix $\tilde{\mathbf{A}}$ and the dynamics are expressed in Eq. 4, which we term the Kinetic Matrix Method. As an example of applying the KMM, consider the parametric nonlinear oscillator system described by the following set of differential equations:

$$\begin{aligned}\frac{dX_1}{dt} &= k_1 X_2 \\ \frac{dX_2}{dt} &= -k_2 X_1 + k_3 X_2 - k_4 X_2 X_3 \\ \frac{dX_3}{dt} &= k_5 X_1^2 - k_6 X_3.\end{aligned}\quad (10)$$

A synthetic data set with $N = 1000$ points was constructed by numerical integration of Eq. 10, using initial concentrations $X_1(0) = 0$; $X_2(0) = 0.25$; $X_3(0) = 0.0625$, the target rate values $k_{10} = k_{20} = k_{30} = k_{40} = k_{50} = k_{60} = 1$, and introducing noise at a level of 1%, shown in Fig. 1.

All possible reactions among the ns species are considered, and the complete set of complex vectors is considered

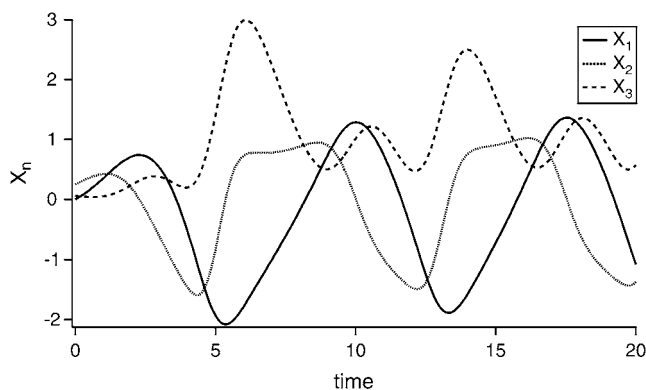


FIGURE 1 Data set for a nonlinear oscillator. The set of differential equations in Eq. 10 was numerically integrated using rate constant values and initial conditions given in the text, and 1% random noise was added to the data.

including the fictitious species X_0 . For a system of ns species, there are $nk = (ns + 1)(ns + 2)/2$ possible combinations of two species p and q . The set of nk possible complex vectors $\hat{\mathbf{y}}_k$ that describe all possible kinetic processes can be assembled into the complete complex matrix $\tilde{\mathbf{Y}}$:

$$\begin{aligned}\tilde{\mathbf{Y}} &= [\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_{nk}] \\ \hat{\mathbf{y}}_k &= \hat{\mathbf{x}}_p + \hat{\mathbf{x}}_q.\end{aligned}\quad (11)$$

The construction of the complete matrix of complexes and the standard mapping between indices $(p, q) \leftrightarrow k$ is given in Appendix III in the Supplementary Material. For the case of three species ($ns = 3$) the matrix $\tilde{\mathbf{Y}}$ is a 3×10 matrix:

$$\begin{aligned}\tilde{\mathbf{Y}} &= [\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \mathbf{y}_4, \mathbf{y}_5, \mathbf{y}_6, \mathbf{y}_7, \mathbf{y}_8, \mathbf{y}_9, \mathbf{y}_{10}] \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 2 \end{bmatrix},\end{aligned}\quad (12)$$

which produces vector of complexes $\hat{\mathbf{F}}$:

$$\begin{aligned}\hat{\mathbf{F}} &= \hat{\mathbf{X}}^{\tilde{\mathbf{Y}}} = \prod_{j=1}^{ns} X_j^{\mathbf{y}_{jk}} \\ &= [1 \ X_1 \ X_2 \ X_3 \ X_1^2 \ X_1 X_2 \ X_1 X_3 \ X_2^2 \ X_2 X_3 \ X_3^2]^T.\end{aligned}\quad (13)$$

These complexes represent all possible unimolecular and bimolecular reactions that can occur among the ns species, and includes a constant term that allows for mass to be added to or taken away from the system.

The elements of the set of $\hat{\mathbf{a}}_j$ are determined using the formulas in Eq. 7 for each of the three species j , giving a set of three solutions that can be assembled into the kinetic matrix $\tilde{\mathbf{A}}$ that represents the dynamics of the system in Eq. 4:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_3 \end{bmatrix} = \begin{bmatrix} -0.0002 & 0.0009 & \mathbf{1.0026} & 0.0005 & 0.0006 & -0.0029 & -0.0012 & -0.0005 & -0.0027 & -0.0007 \\ 0.0006 & \mathbf{-0.9844} & \mathbf{0.9990} & 0.0000 & -0.0001 & -0.0019 & -0.0011 & 0.0005 & \mathbf{-0.9978} & -0.0013 \\ 0.0026 & 0.0009 & -0.0062 & \mathbf{-0.9876} & \mathbf{0.9887} & 0.0124 & 0.0004 & -0.0109 & 0.0013 & 0.0020 \end{bmatrix}. \quad (14)$$

The elements of $\tilde{\mathbf{A}}$ are all nonzero due to the effects of noise on the reconstruction, however, the bold elements that correspond to the processes in the dynamics are all significantly above the noise floor. A significant nonzero value for A_{jp} indicates that the dynamics of species j depends on complex p . Element A_{13} indicates that the dynamics of X_1 depend on complex y_3 , which can be seen from inspection of $\tilde{\mathbf{Y}}$ is X_2 , which is in turn due to the term $k_1 X_2$ in Eq. 10. Element A_{35} indicates that the dynamics of X_3 depend on complex y_5 , which can be seen from inspection of $\tilde{\mathbf{Y}}$ is X_1^2 , which is due to the term $k_5 X_1^2$ in Eq. 10. There are six significant values in $\tilde{\mathbf{A}}$: A_{13} , A_{22} , A_{23} , A_{29} , A_{34} , and A_{35} each of which directly corresponds to one of the flux terms in Eq. 10. Thus, the Kinetic Matrix Method in terms of the complexes formalism successfully extracts the structure of the dynamics of the system from the time-dependent data.

The accuracy of the reconstruction can be determined from the relative deviation of the np fitted nonzero rates to the input rates Δ_s averaged over $S = 40$ noise realizations:

$$\Delta_s = \sqrt{\frac{1}{np} \sum_{p=1}^{np} (k_p^s - k_{0p})^2 / k_{0p}^2}$$

$$\Delta = \sqrt{\frac{1}{S} \sum_{s=1}^S (\Delta_s)^2}. \quad (15)$$

For small numbers of time points ($N \leq 50$), Δ is on the order of 10%, and there is rapid decrease that plateaus at $\Delta < 1\%$ for $N > 500$.

If an entire column p of the reconstructed matrix $\tilde{\mathbf{A}}$ is null, this indicates that complex p does not contribute to the observed dynamics. Thus, the minimal reduced set of complexes required to reconstruct the data can be identified by inspection of $\tilde{\mathbf{A}}$. The accuracy of determination of the rate constants for the system can be improved by subsequently neglecting the complexes that are not present in the dynamics. The complex \hat{y}_m is neglected if $\forall j \in [0, ns]$ that fulfils the condition:

$$|A_{jm}| < \varepsilon_j, \quad (16)$$

where ε_j is a threshold that must be empirically determined. The reduced matrix of complexes $\tilde{\mathbf{Y}}$ for the system of Eq. 10 can be simply obtained from the matrix $\tilde{\mathbf{A}}$:

$$\tilde{\mathbf{Y}} = [\hat{y}_2, \hat{y}_3, \hat{y}_4, \hat{y}_5, \hat{y}_9] = \begin{bmatrix} 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}. \quad (17)$$

Having constructed the reduced matrix $\tilde{\mathbf{Y}}$, it is now straightforward to solve again for the set of rate constant assembled into the reduced kinetic matrix $\tilde{\mathbf{A}}$, using Eq. 7:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_3 \end{bmatrix} = \begin{bmatrix} 0.0009 & \mathbf{1.0026} & -0.0001 & -0.0001 & -0.0000 \\ \mathbf{-0.9984} & \mathbf{0.9990} & -0.0010 & 0.0004 & \mathbf{-0.9950} \\ 0.0009 & -0.0062 & \mathbf{-0.9876} & \mathbf{0.9999} & 0.0012 \end{bmatrix}. \quad (18)$$

Using these reduced matrices reduces the number of possible kinetic complexes that must be identified from the data from 10 to 5, and reduces the number of kinetic parameters from 30 to 15. There is a reduction in the root mean square (rms) error (Δ) of the target rate constants by a factor of at least 2 using the reduced KMM compared to the full KMM, as shown in Fig. 2. It should be noted that if there is prior information about the nature of the complexes present in the dynamics, this information can be used to directly construct the reduced matrix of complexes.

Application of the Representation Matrix Method

Describing the kinetics in terms of a set of representation matrices provides a general and flexible way of including

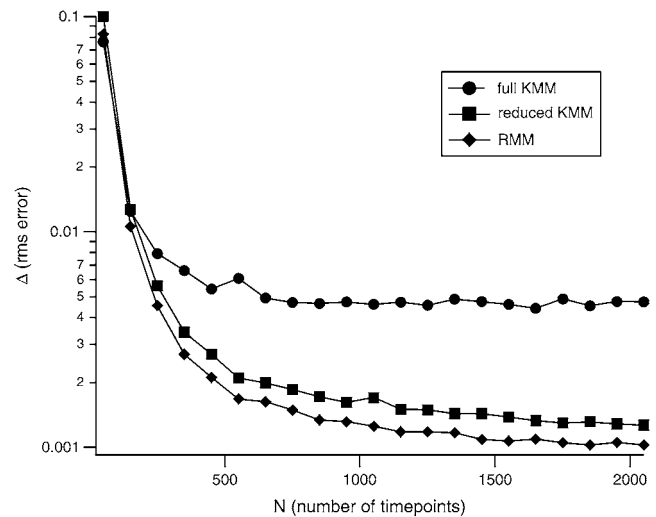


FIGURE 2 Comparison of the accuracy of three variants of the Numerical Matrices Method applied to the nonlinear oscillator data set.

prior information about the structure of the dynamics. In addition, the representation matrices allow description of nonlinear dynamic systems that do not correspond to chemical reaction networks. For the system in Eq. 10, there are a total of six kinetic terms operative, and the complex matrix $\tilde{\mathbf{Y}}$ takes the form in Eq. 17, but the kinetic equations cannot be described with a standard stoichiometry matrix. However, we can use the Representation Matrix Method to determine the rate constants from the time-series data. If all of the rates are nonequal, $k_1 \neq k_2 \neq k_3 \neq k_4 \neq k_5 \neq k_6$, the set of the representation matrices $\tilde{\mathbf{G}}_p$ in Eq. 5 is given by:

$$\begin{aligned} \tilde{\mathbf{G}}_1 &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} & \tilde{\mathbf{G}}_2 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \tilde{\mathbf{G}}_3 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} & \tilde{\mathbf{G}}_4 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \\ \tilde{\mathbf{G}}_5 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} & \tilde{\mathbf{G}}_6 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (19)$$

The values for the rate constants k_p are determined using Eqs. 8 and 19, and the deviation from the data is calculated using Eq. 15. The RMM includes prior information about the reaction structure, which results in a decrease in the number of parameters to be determined, compared to the KMM, and provides an additional increase in the accuracy of determination of the rates, as shown in Fig. 2.

Application of the Stoichiometry Matrix Method

The Stoichiometry Matrix Method can be applied to a large class of problems involving chemical reactions. The struc-

ture of the chemical reaction network and the relationship among the chemical species is given by the stoichiometry matrix, and the dynamical equations are given by Eq. 6. This work was in part motivated by the need to analyze complex kinetic data for assembly of the 30S ribosomal subunit, which is responsible for decoding the mRNA during protein synthesis in bacteria. The 30S subunit is composed of 20 small proteins and a large 16S rRNA that form a large globular structure with a dense RNA interior decorated by the ribosomal proteins (36,37). It is possible to reconstitute 30S subunits from purified components in vitro, which led to an assembly map involving a complex series of parallel and sequential protein binding events (38). Recently, quantitative kinetic data for the binding rates of 30S proteins has been collected using an isotope pulse chase method (39). There is currently no straightforward method to analyze this complex kinetic data to determine the mechanism of assembly and to extract rate constants for the binding reactions.

As a model system to demonstrate the application of the Stoichiometry Matrix Method to assembly of a ribonucleoprotein complex, we consider three RNA binding proteins, A, B, and C, that bind to an RNA, R, to form a quaternary complex RABC. We specify that A and B can bind to the RNA independently, but that binding of C requires prior binding of A (i.e., there is no RC complex), and that no protein-protein complexes are formed among A, B, or C, as shown in Fig. 3.

There are 14 rate constants associated with these seven reactions (seven forward and seven reverse), and nine different species. The species vector $\hat{\mathbf{X}}$ is given by:

$$\begin{aligned} \hat{\mathbf{X}} &= [X_1 \ X_2 \ X_3 \ X_4 \ X_5 \ X_6 \ X_7 \ X_8 \ X_9]^T \\ &= [[A] \ [B] \ [C] \ [R] \ [RA] \ [RAB] \ [RABC] \ [RB] \ [RBC]]^T. \end{aligned} \quad (20)$$

For the reaction scheme shown in Fig. 3, the stoichiometry matrix $\tilde{\mathbf{R}}$, and complex matrix $\tilde{\mathbf{Y}}$ are given by:

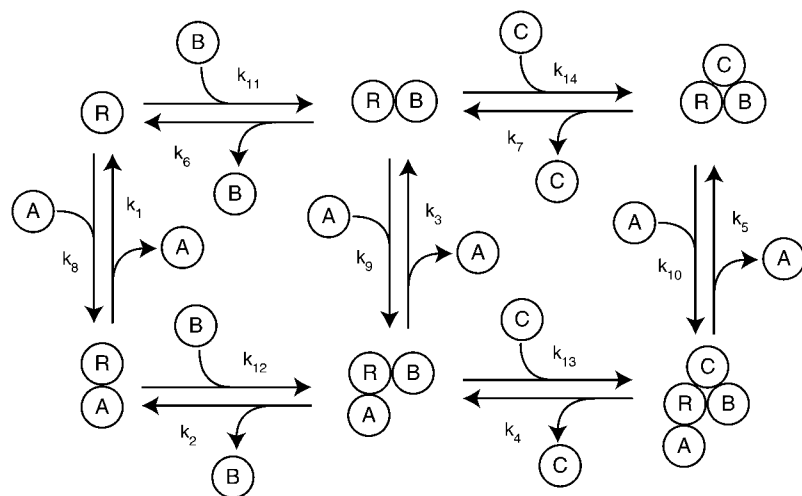


FIGURE 3 Hypothetical assembly mechanism of a quaternary complex between an RNA R, and three proteins (A,B,C). Proteins A and B can bind independently to R, but binding of protein C requires prior binding of protein B.

$$\tilde{\mathbf{R}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{\mathbf{Y}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (21)$$

To demonstrate the application of the SMM, a synthetic data set was constructed by numerical integration of Eq. 6, using the following values for the rate constants $\hat{\mathbf{K}}_0$ and initial concentrations $\hat{\mathbf{X}}(0)$:

$$\hat{\mathbf{K}}_0 = [0.01 \ 0.001 \ 0.02 \ 0.05 \ 1 \ 0.005 \ 0.002 \ 1 \ 6 \ 1 \ 3 \ 3.05 \ 1 \ 0 \ 0.02]^T$$

$$\hat{\mathbf{X}}(0) = [1.0 \ 1.0 \ 1.0 \ 1.0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T. \quad (22)$$

Noise was introduced into the data set at a level of $\sigma = 0.001$ and the data are shown in Fig. 4. Application of Eq. 9 gives values for the set of rate parameters from the model data set shown in Fig. 4. The estimation of the rate constants for the synthetic data using Eq. 9 is shown as the solution vector $\hat{\mathbf{K}}$:

$$\hat{\mathbf{K}} = [0.0100 \ 0.0010 \ 0.0198 \ 0.0499 \ 0.0998 \ 0.0051 \ 0.0021 \ \dots \\ 0.1002 \ 0.5994 \ 1.2963 \ 0.3000 \ 0.0503 \ 0.9996 \ 0.0201]^T, \quad (23)$$

which closely matches the initial values used to generate the data, in Eq. 22.

The SMM procedure was repeated for synthetic data sets with different numbers of time points, and for comparison, reconstructions were also performed using the full KMM and the reduced KMM. The accuracy of the reconstructions were quantified as the root mean square deviation of the reconstructed rate constants ($\hat{\mathbf{K}}$) from the target rate

constants $\hat{\mathbf{K}}_0$, in Eq. 22. A plot of the rms error Δ for the three methods as a function of the number of time points is shown in Fig. 5. Inclusion of the stoichiometry matrix improves the accuracy of the reconstructed rates for data sets of all size.

Breaking holonomic constraints in the KMM by using multiple initial states

Considering the case where no information about the structure of the reaction network is known a priori, we seek to reconstruct the reaction pathways and determine the rate constants given data from the time dependence of the concentrations of the various species. Data from a real

experiment such as that shown in Fig. 4 would be collected after initiating assembly of the RNA-protein complex by mixing equimolar amounts of R, A, B, and C. Using an appropriate measurement method, we would observe the presence of five new RNA-protein complexes, X_5 , X_6 , X_7 , X_8 , and X_9 , whose identity is not known, a priori. We assume for this exercise that we can identify the four input molecules R, A, B, and C using the measurement method.

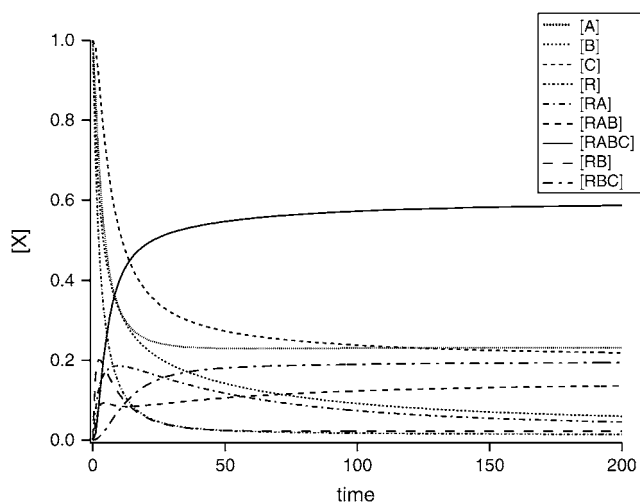


FIGURE 4 Model kinetic data set for assembly of the RABC quaternary complex. Data were generated by numerical integration of Eq. 6 using the matrices in Eq. 21 and the rate constants and initial concentrations in Eq. 22.

There are several classes of mechanisms that might be operative for assembly of the final particle RABC. There might be a required sequential order to binding, there might be completely independent binding of the three proteins in any order, or, as is the case for the mechanism in Eq. 59, a combination of the two. Here we demonstrate the application of the KMM to this model system to determine the mechanism and extract the rate constants. Determining the mechanism allows the stoichiometry matrix to be constructed from the kinetic matrix $\hat{\mathbf{A}}$.

In principle, the KMM can be directly applied as described above for the nonlinear oscillator example. However, in practice, the matrix $\tilde{\mathbf{B}}$ becomes singular using data in Fig. 4 due to linear dependencies among the concentration data for the system of Fig. 3. The linear dependencies arise from conservation of mass in the system of reactions:

$$\begin{aligned} [A] + [RA] + [RAB] + [RABC] &= X_1 + X_5 + X_6 + X_7 = A_{\text{total}} \\ [B] + [RAB] + [RABC] + [RB] + [RBC] &= X_2 + X_6 + X_7 + X_8 + X_9 = B_{\text{total}} \\ [C] + [RABC] + [RBC] &= X_3 + X_7 + X_9 = C_{\text{total}} \\ [R] + [RA] + [RAB] + [RABC] + [RB] + [RBC] &= X_4 + X_5 + X_6 + X_7 + X_8 + X_9 = R_{\text{total}}. \end{aligned} \quad (24)$$

These holonomic constraints are a property of the stoichiometry matrix that describes the reaction network. Analysis of the left and right null spaces and range of the stoichiometry matrix can be used to derive the steady-state fluxes and equilibrium points, and in particular, the conservation of mass relations that are the holonomic constraints on the dynamics of the network. For the purpose of applying the KMM, the structure of the stoichiometry matrix is not known, and thus the nature of the holonomic constraints cannot be known, a priori.

To apply the KMM, it is necessary to break the holonomic conditions arising from the conservation of the quantities

$A_{\text{total}}, B_{\text{total}}, C_{\text{total}}, R_{\text{total}}$ that make the numerical matrix $\tilde{\mathbf{B}}$ singular. One strategy to break the holonomic conditions is to collect the time-series data beginning with a sufficient number of different initial concentrations to increase the rank of $\tilde{\mathbf{B}}$ and to avoid its singularity. For the model system of Fig. 3, it is sufficient to use 16 different time series, using the initial concentrations:

$$\begin{aligned} \hat{\mathbf{X}}^T(1,0) &= [1 \ 1 \ 1.1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(2,0) &= [1 \ 1 \ 1 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(3,0) &= [1 \ 1 \ 1.5 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(4,0) &= [1 \ 1 \ 1.5 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(5,0) &= [1 \ 1.5 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(6,0) &= [1 \ 1.5 \ 1 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(7,0) &= [1 \ 1.5 \ 1.5 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(8,0) &= [1 \ 1.5 \ 1.5 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(9,0) &= [1.5 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(10,0) &= [1.5 \ 1 \ 1 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(11,0) &= [1.5 \ 1 \ 1.5 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(12,0) &= [2.5 \ 1 \ 1.5 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(13,0) &= [1.5 \ 1.5 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(14,0) &= [1.5 \ 1.7 \ 1 \ 1.5 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(15,0) &= [1.5 \ 1.5 \ 1.5 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0] \\ \hat{\mathbf{X}}^T(16,0) &= [1.5 \ 1.5 \ 1.5 \ 1.4 \ 0 \ 0 \ 0 \ 0 \ 0], \end{aligned} \quad (25)$$

where $\hat{\mathbf{X}}^T(m,0)$ is the initial set of concentrations used to generate the time series $X_j(m, t_i)$ using $\hat{\mathbf{K}}^T$ from Eq. 22, and m is the index for the time series that runs from 1 to $nm = 16$

data sets. Noise was introduced into the data with standard deviations $\sigma(X_j(t_i)) = 0.001 \times |X_j(t_i)|$.

From this set of 16 synthetic data sets, the Kinetic Matrix Method of Eq. 7 can be applied to identify the complexes responsible for the kinetics and to extract the rate constants. The kinetic matrix for this system is derived from the $ns = 9$ species, and the $nk = 55$ canonical complexes \mathbf{y}_k given by Eq. 11, resulting in the 9×55 matrix $\tilde{\mathbf{A}} = A_{jk} = \hat{\mathbf{a}}_j$. In addition, the formulas for $\tilde{\mathbf{B}}$ and $\hat{\mathbf{c}}_j$ in Eq. 7 must be modified to include summation over the nm data sets recorded at different initial concentrations:

$$\begin{aligned}\tilde{\mathbf{B}} &= b_{kl} = \sum_{m=1}^{nm} \sum_{i=1}^{nt} (\hat{\mathbf{X}}(m, t_i))^{\hat{y}_k} (\hat{\mathbf{X}}(m, t_i))^{\hat{y}_l} \\ \hat{\mathbf{c}}_j &= c_{jk} = \sum_{m=1}^{nm} \sum_{i=1}^{nt} \left(\frac{dX_j(m, t_i)}{dt} \right)_{\text{obs}} (\hat{\mathbf{X}}(m, t_i))^{\hat{y}_k}.\end{aligned}\quad (26)$$

Most of the elements A_{jk} of the kinetic matrix $\tilde{\mathbf{A}}$ obtained from Eq. 26 are close to zero, but some A_{jk} that correspond to the k_{0i} are close to the initial value used to generate the data.

The quality of reconstruction of the reaction system of Fig. 3 can be estimated from the deviation of reconstructed value of k_i from initial k_{0i} , using Eq. 15, where $np = 14$ is the total number of nonzero rate constants. The value of Δ as function of number of time points used for the reconstruction is shown in Fig. 5.

The full kinetic matrix $\tilde{\mathbf{A}}$ is too large to be conveniently shown, but columns $k = 19 \dots 23$ of matrix $\tilde{\mathbf{A}}$ are shown to illustrate the essential features of the reconstructed matrix:

$$\tilde{\mathbf{A}} = \begin{matrix} k = \\ \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_3 \\ \hat{\mathbf{a}}_4 \\ \hat{\mathbf{a}}_5 \\ \hat{\mathbf{a}}_6 \\ \hat{\mathbf{a}}_7 \\ \hat{\mathbf{a}}_8 \\ \hat{\mathbf{a}}_9 \end{bmatrix} \end{matrix} = \begin{matrix} \begin{matrix} 19 & 20 & 21 & 22 & 23 \end{matrix} \\ \begin{bmatrix} \dots & -\mathbf{1.0896} & -0.0007 & -0.0018 & 0.0008 & -0.0028 & \dots \\ \dots & 0.1613 & -0.0234 & -0.0043 & -\mathbf{0.3008} & -\mathbf{0.0387} & \dots \\ \dots & -0.1419 & -0.0086 & -0.0016 & -0.0006 & -0.0002 & \dots \\ \dots & -0.0337 & -0.0098 & -0.0000 & -\mathbf{0.3034} & -0.0006 & \dots \\ \dots & 0.0393 & 0.0001 & -0.0000 & 0.0001 & -\mathbf{0.0499} & \dots \\ \dots & 0.0563 & -0.0002 & 0.0003 & -0.0001 & \mathbf{0.0500} & \dots \\ \dots & \mathbf{0.7218} & 0.0034 & -0.0002 & -0.0004 & -0.0017 & \dots \\ \dots & -0.0120 & 0.0000 & -0.0004 & \mathbf{0.3004} & -0.0003 & \dots \\ \dots & -\mathbf{0.9461} & -0.0023 & -0.0000 & 0.0004 & 0.0015 & \dots \end{bmatrix} \end{matrix}, \quad (27)$$

where k is the canonical number of complex that is determined using Eq. 4. It is helpful to immediately construct

the reduced complex matrix by inspection of the matrix $\tilde{\mathbf{A}}$, and the reduced matrices $\tilde{\mathbf{Y}}$ and $\tilde{\mathbf{A}}$ are given by:

$$\tilde{\mathbf{Y}} = \begin{matrix} k = \\ \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_3 \\ \hat{\mathbf{a}}_4 \\ \hat{\mathbf{a}}_5 \\ \hat{\mathbf{a}}_6 \\ \hat{\mathbf{a}}_7 \\ \hat{\mathbf{a}}_8 \\ \hat{\mathbf{a}}_9 \end{bmatrix} \end{matrix} = \begin{matrix} \begin{matrix} 6 & 7 & 8 & 9 & 10 & 14 & 18 & 19 & 22 & 23 & 31 & 33 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \end{matrix} \quad (28)$$

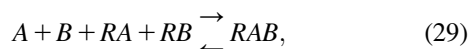
$$\tilde{\mathbf{A}} = \begin{matrix} k = \\ \begin{bmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \\ \hat{\mathbf{a}}_3 \\ \hat{\mathbf{a}}_4 \\ \hat{\mathbf{a}}_5 \\ \hat{\mathbf{a}}_6 \\ \hat{\mathbf{a}}_7 \\ \hat{\mathbf{a}}_8 \\ \hat{\mathbf{a}}_9 \end{bmatrix} \end{matrix} = \begin{matrix} \begin{matrix} 6 & 7 & 8 & 9 & 10 & 14 & 18 & 19 & 22 & 23 & 31 & 33 \end{matrix} \\ \begin{bmatrix} \mathbf{0.0101} & \mathbf{0.0198} & \mathbf{0.1008} & -0.0009 & 0.0001 & -\mathbf{0.1009} & -\mathbf{0.5929} & -\mathbf{1.3036} & 0.0003 & 0.0003 & -0.0100 & 0.0003 \\ -0.0000 & \mathbf{0.0005} & 0.0011 & \mathbf{0.0044} & -0.0000 & -0.0004 & 0.0168 & 0.0039 & -\mathbf{0.3035} & -\mathbf{0.0479} & -0.0286 & 0.0028 \\ 0.0000 & -0.0001 & \mathbf{0.0492} & 0.0000 & \mathbf{0.0020} & -0.0002 & 0.0034 & 0.0123 & -0.0004 & 0.0000 & -\mathbf{1.0039} & -\mathbf{0.0196} \\ \mathbf{0.0100} & -0.0001 & 0.0007 & \mathbf{0.0047} & 0.0000 & -\mathbf{0.0999} & 0.0041 & -0.0040 & -\mathbf{0.3010} & 0.0006 & -0.0080 & 0.0008 \\ -\mathbf{0.0100} & \mathbf{0.0010} & -0.0002 & 0.0001 & -0.0000 & \mathbf{0.1000} & 0.0000 & 0.0027 & -0.0000 & -\mathbf{0.0500} & -0.0005 & -0.0001 \\ -0.0000 & -\mathbf{0.0210} & \mathbf{0.0491} & 0.0005 & -0.0000 & -0.0000 & \mathbf{0.5989} & 0.0096 & -0.0001 & \mathbf{0.0497} & -\mathbf{0.9964} & -0.0003 \\ 0.0000 & 0.0002 & -\mathbf{0.1475} & -0.0014 & 0.0001 & -0.0001 & 0.0004 & \mathbf{1.2678} & 0.0002 & -0.0001 & \mathbf{0.9994} & 0.0003 \\ -0.0000 & \mathbf{0.0199} & -0.0004 & -\mathbf{0.0051} & \mathbf{0.0020} & -0.0000 & -\mathbf{0.5976} & 0.0078 & \mathbf{0.3001} & 0.0001 & -0.0038 & -\mathbf{0.0198} \\ -0.0000 & -0.0001 & \mathbf{0.0990} & 0.0007 & -\mathbf{0.0020} & 0.0000 & -0.0007 & -\mathbf{1.2866} & -0.0000 & 0.0001 & 0.0006 & \mathbf{0.0198} \end{bmatrix} \end{matrix}.$$

At this point, the details of the mechanism for assembly become clear by examining the structure of these two matrices. First, all of the irrelevant possible reactions, such as dimerization of proteins and formation of heterodimeric protein-protein complexes have been eliminated from consideration. Importantly, there is no complex that corresponds to the forward or reverse reaction $R + C \rightleftharpoons RC$. However, there are two elementary reactions that correspond to the complexes $R + A = \hat{y}_{14}$ and $R + B = \hat{y}_{22}$, which indicates that either A or B can bind independently to R. Inspection of columns 1 and 6 of $\tilde{\mathbf{A}}$ reveal that the dynamics of X_5 also depends on \hat{y}_{14} , which clearly identifies X_5 as the RNA-protein complex RA. Similarly, complex \hat{y}_{22} also affects the dynamics of X_8 , identifying it as the RNA-protein complex RB. Complex \hat{y}_7 identifies X_6 as RAB, complex \hat{y}_{10} identifies X_9 as RBC, and complex \hat{y}_{19} identifies X_7 as RABC. It can also be seen from the structure of $\tilde{\mathbf{A}}$, that both forward and reverse reactions are occurring in the dynamics. A great deal of the mechanism is clear from the reconstruction, however there are some ambiguities evident from column 3 of $\tilde{\mathbf{A}}$. Application of Eq. 7 to the reduced matrix of complexes gives rise to the reconstructed reduced kinetic matrix $\tilde{\mathbf{A}}$. The accuracy of the reconstruction of the rates is significantly improved using the reduced KMM, as shown in Fig. 5.

Reconstruction of the stoichiometry matrix

In the previous sections we have described the application of the KMM to identify the processes inherent in kinetic data, and to extract the rate constants for the processes. In the case of a dynamical system that corresponds to a chemical reaction network, we have shown that application of the stoichiometry matrix, as in the SMM, further improves the accuracy of the rate reconstructions. A complete and powerful synthesis of these two approaches would involve application of the KMM to identify the dynamics present, and use of this information to construct de novo a stoichiometry matrix that describes the reactions present in the system.

One of the difficulties encountered in construction of a stoichiometry matrix from the kinetic matrix $\tilde{\mathbf{A}}$ is apparent in column 2 of Eq. 28. It would appear that an apparent reaction of the form:



is taking place, but it is clear that such a complex reaction is a composite of two separate elementary reactions that both give the same product, RAB. It is necessary to describe several of the properties of networks of unimolecular and biomolecular reactions to develop an algorithm to resolve kinetic ambiguities in the structure of a kinetic matrix.

Consider one of the elementary chemical reaction steps from Fig. 3:

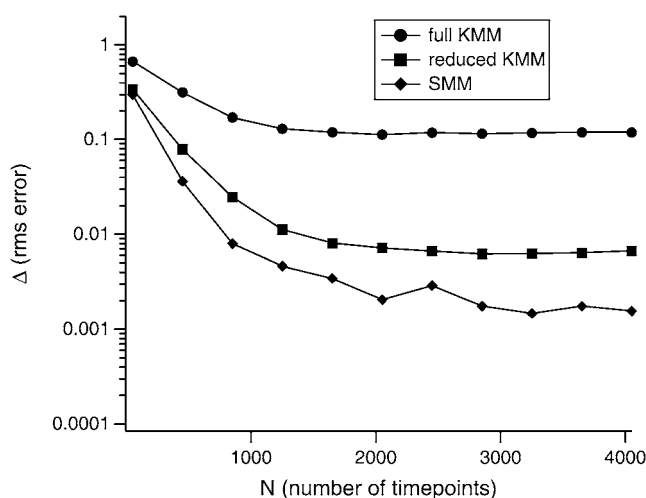
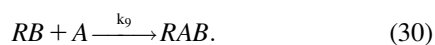
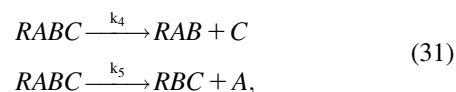


FIGURE 5 Comparison of the accuracy of three variants of the Numerical Matrices Method applied to the quarternary ribonucleoprotein complex.

The reaction in Eq. 30 is represented by a reaction vector $\hat{\mathbf{r}}_9 = \hat{\mathbf{y}}_{18} - \hat{\mathbf{y}}_7$, which is a bimolecular reaction where RAB is the only possible product of the reaction of A and RB, and we define this as a “univariant” reaction. More specifically, there is only one possible product complex, $\hat{\mathbf{y}}_{18}$, arising from reactant complex $\hat{\mathbf{y}}_7$. In contrast, for the dissociation of RABC, there are two possible products:



and we define this as a “bivariant”, or more generally a “multivariant” reaction. There are two possible product complexes that can be formed from the same reactant complex. The observed rate of dissociation of RABC is determined by the composite rate constant:

$$k^{\text{obs}} = k_4 + k_5, \quad (32)$$

as outlined in Appendix IV in the Supplementary Material. The composite reactions contained in the matrix $\tilde{\mathbf{A}}$ due to the presence of bivariant reactions must be decomposed into their elementary reactions.

With these considerations in mind, we can devise a method for reconstruction of the stoichiometry matrix $\tilde{\mathbf{R}} = R_{jk}$ from the primary set of dynamical data $X_j(t_i)$. The first step is to use the KMM to obtain a model of the chemical system in terms of the kinetic matrix $\tilde{\mathbf{A}}$ and a complete set of complexes $\tilde{\mathbf{Y}}$. As described above, matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{Y}}$ will both have dimensions of 9×55 . Second, the reduced set of complexes is formed by elimination of the null columns of $\tilde{\mathbf{A}}$, and the reduced dimension of matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{Y}}$ will be 9×12 . The third step is to use the Representation Matrix Method for extraction of nonzero rates from the reduced kinetic matrix $\tilde{\mathbf{A}}$, which are then organized into a vector of rates $\hat{\mathbf{K}} = k_p$. The kinetic matrix $\tilde{\mathbf{A}}$ can be equivalently

$$\begin{array}{rcl}
 q = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\
 k = & 6 & 7 & 7 & 7 & 8 & 8 & 8 & 9 & 10 & 14 & 18 & 19 & 22 & 23 & 31 & 33 \\
 \tilde{\mathbf{W}} = & \begin{bmatrix}
 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\
 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 \\
 -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\
 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & -1 \\
 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 1
 \end{bmatrix} \\
 \\
 k = & 6 & 7 & 7 & 7 & 8 & 8 & 8 & 9 & 10 & 14 & 18 & 19 & 22 & 23 & 31 & 33 \\
 \tilde{\mathbf{Y}}^w = & \begin{bmatrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
 \end{bmatrix} \\
 \hat{\mathbf{K}}_0^T = & [.01 & .001 & .02 & .0210 & .0492 & .1 & .1475 & .0048 & .002 & .1 & .6 & 1.29 & .3 & .05 & 1. & .02] \quad (33)
 \end{array}$$

expressed in terms of a set of representation matrices, given by Eq. 5. Each of the resulting representation matrices $\tilde{\mathbf{G}}_p$ will have only one nonzero element, as in Eq. 19. Fourth, since the form of the dynamical equations using the set of representation matrices is significantly different from the form using the stoichiometry matrix, as in Eq. 6, and it is necessary to generate an intermediate “collapsed” kinetic matrix $\tilde{\mathbf{W}}$. If the dynamics of the system corresponds to a set of univariant chemical reactions, matrix $\tilde{\mathbf{W}}$ will be identical to the stoichiometry matrix $\tilde{\mathbf{R}}$. However, in general, there are differences between $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{R}}$ in columns that correspond to the multivariant reactions. Inspection of $\tilde{\mathbf{W}}$ allows the identification of the multivariant reactions.

A final step in construction of the stoichiometry matrix $\tilde{\mathbf{R}}$ from the collapsed matrix $\tilde{\mathbf{W}}$ is necessary. Each column of $\tilde{\mathbf{W}}$ may correspond to a true reaction vector for a univariant reaction, or it may correspond to a reactant complex vector for a multivariant reaction. The formal mapping of indices to convert the kinetic matrix $\tilde{\mathbf{A}}$ into the set of representation matrices $\tilde{\mathbf{G}}_p$ and vector of rates $\hat{\mathbf{k}} = k_p$, the construction of the collapsed matrix $\tilde{\mathbf{W}}$, and the decomposition of the multivariant columns of $\tilde{\mathbf{W}}$ into reaction vectors is detailed in

Appendix V in the Supplementary Material. Finally, the newly constructed stoichiometry matrix $\tilde{\mathbf{R}}$ can be used in the SMM to further improve the accuracy of the rate constants for the dynamical system. The outlined procedure provides a robust method for determination of the structure of a chemical reaction network beginning only with the time series data $X_j(t_i)$.

DISCUSSION

The numerical matrices method for nonlinear system identification

We have developed and implemented several variant methods of nonlinear system identification and determination of rate constants for cascades of chemical reactions. All of these methods describe the chemical kinetics in a similar manner using various numerical matrices: the matrix of complexes $\tilde{\mathbf{Y}}$, the kinetic matrix $\tilde{\mathbf{A}}$, the stoichiometry matrix $\tilde{\mathbf{R}}$, the vector of rates $\hat{\mathbf{k}}$, and the set of representation matrices $\tilde{\mathbf{G}}_p$. However, these methods differ from each other in the amount of prior information about the reaction system that is used in the determination. Inclusion of prior information

increases the accuracy of the reconstruction, in large part due to the decrease in the number of unknown parameters that must be determined.

The most basic method is the Kinetic Matrix Method, which uses an arbitrary kinetic matrix $\tilde{\mathbf{A}}$ and the complete set of complexes $\tilde{\mathbf{Y}}$ to describe all possible unimolecular and bimolecular reactions among the species present. No prior information is included about the structure of the reaction network, the input being the identity of the chemical species and the time dependence of their concentrations. The KMM identifies which reactions are taking place, and provides estimates for the rate constants for the reactions. Although the KMM is robust, it can be seen from Figs. 2 and 5 that there is a significant error in the rates that are determined.

If the nature of the reactions are known or have been determined using the KMM, or if some possible reactions can be excluded a priori, it is possible to form a reduced matrix of complexes, which provides increased accuracy of the determined rates using the KMM. In addition, prior information about the structure of the reaction network is included by construction of the reduced matrix of complexes $\tilde{\mathbf{Y}}$ to reflect only the complexes that are present. As can be seen in Figs. 2 and 5, the reduced KMM provides a significant increase in the accuracy of the determined rates compared to the full KMM, which is presumably due to elimination of rate parameters for complexes that are irrelevant to the dynamics of the system.

If the complete structure of the reaction network is known, the dynamics can be represented using the stoichiometry matrix formalism, which effectively pairs up the reactant and product complexes appropriate for each reaction in the system. The stoichiometry matrix $\tilde{\mathbf{R}}$ contains information about the structure of the reaction network, whereas the rate constants are contained in a vector $\tilde{\mathbf{K}}$. This method requires determination of the fewest parameters from the data, and consequently provides the most accurate determination of the rate constants, as can be seen in Fig. 5.

A particularly significant part of the NMM is the Representation Matrix Method. The RMM is the most generalized and universal of the methods, and can work with different levels of prior information, and the KMM and the SMM can be considered as specific implementations of the RMM, where the level of prior information included in the reconstruction is determined by the choice of the set of representation matrices $\tilde{\mathbf{G}}_p$. In the complete absence of prior information, the KMM can be fully and equivalently represented and implemented by choosing the set of representation matrices $\tilde{\mathbf{G}}_p^A$ such that:

$$\frac{d\tilde{\mathbf{X}}}{dt} = \tilde{\mathbf{A}} \cdot \tilde{\mathbf{X}}^{\tilde{\mathbf{Y}}} = \sum_{p=1}^{np} k_p^A \tilde{\mathbf{G}}_p^A \cdot \tilde{\mathbf{X}}^{\tilde{\mathbf{Y}}}. \quad (34)$$

When the full structure of the reaction network is known, the SMM can be similarly represented by the choice of representation matrices $\tilde{\mathbf{G}}_p^R$ such that:

$$\frac{d\tilde{\mathbf{X}}}{dt} = (\tilde{\mathbf{R}} \cdot \text{diag}(\tilde{\mathbf{K}})) \cdot \tilde{\mathbf{X}}^{\tilde{\mathbf{Y}}} = \sum_{p=1}^{np} k_p^R \tilde{\mathbf{G}}_p^R \cdot \tilde{\mathbf{X}}^{\tilde{\mathbf{Y}}}. \quad (35)$$

In addition, the RMM is particularly applicable when some of the rates of reactions are exactly known and others remain to be determined. In this most general form, the dynamics of the system is given by:

$$\frac{d\tilde{\mathbf{X}}}{dt} = \sum_{p=1}^{np} (k_p \tilde{\mathbf{G}}_p + \tilde{\mathbf{Q}}) \cdot \tilde{\mathbf{X}}^{\tilde{\mathbf{Y}}}, \quad (36)$$

where the matrix $\tilde{\mathbf{Q}}$ contains the information about known rate parameters (35).

Determination of reaction network structure

For a given reaction mechanism, it is straightforward to construct the set of differential equations that describe the dynamics of the reaction network, as shown in the top part of Fig. 6. Given initial conditions and values for the rate constants it is also straightforward to either analytically or numerically integrate the system of differential equations to provide the time-dependent concentrations of the species in the network. The NMM described here provides a general method for the reverse process, where the time dependence of the species alone is sufficient to determine the reaction structure, as shown in the lower part of Fig. 6. From the time dependence of the species and the corresponding rates, the KMM is applied to determine the basic reaction structure and provide initial estimates for rate constants in the general kinetic matrix $\tilde{\mathbf{A}}$. In some cases it is possible to deduce the stoichiometry matrix $\tilde{\mathbf{R}}$ directly from $\tilde{\mathbf{A}}$, but in particular for the case where there are multivariant reactions present in the system, additional steps are required. We have outlined general procedure for generating $\tilde{\mathbf{R}}$ from $\tilde{\mathbf{A}}$ for a chemical reaction network via a set of intermediate representation matrices $\tilde{\mathbf{G}}_p$ and an intermediate collapsed matrix $\tilde{\mathbf{W}}$. Thus, beginning only with the time dependence of the species, the structure of the reaction network can be determined using the NMM.

Comparison to correlation metric construction

There are some strong analogies between the NMM and the previously developed correlation metric construction (CMC) method (10–12). For both methods, reaction pathways can be identified from concentration measurements of the species, and a type of correlation matrix is used for the analysis. In addition, the stochastic modulation of the inputs in the CMC is similar in spirit to the use of multiple initial states in NMM. Both the NMM and CMC have the common limitation that data for all species must be considered.

There are significant differences between the NMM and CMC as well, and the NMM has several advantages. First, an important part of NMM is that values for rate parameters are determined, but no such values are produced by CMC, which

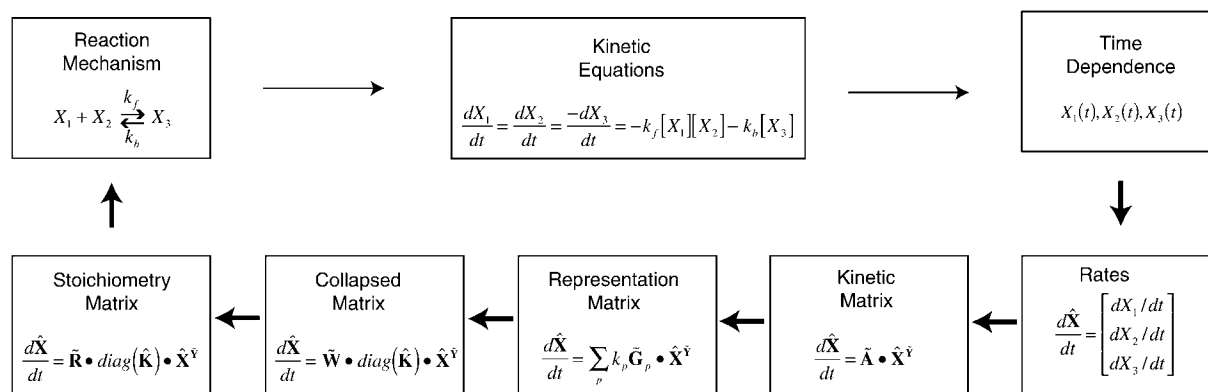


FIGURE 6 Overview of the Numerical Matrices Method to nonlinear system identification. Given a reaction mechanism, it is straightforward to construct a set of differential equations that can be numerically integrated to give time dependence of the species. The NMM formally carries out the reverse of this process in a series of steps involving kinetic matrices.

is limited to identification of the reaction structure. Second, NMM is based on correlations between kinetic complexes such as $(X_p(t)X_q(t))$ and $(X_r(t)X_s(t))$, which are analyzed using linear algebra and statistical methods. In contrast, the CMC is based on analysis of a time-lagged two-species correlation function between $(X_i(t) - \bar{X}_i)$ and $(X_j(t+\tau) - \bar{X}_j)$ that is interpreted using a heuristic algorithm. Finally, there is no way to include prior information about reaction structure using the CMC, which is a strength of the NMM.

Domain of applicability

To apply the methods in the NMM, it is necessary to collect time-series data on all of the species present, and data sets with many time points are required. One advantage of the NMM is that no information is required about the initial conditions. Data collected beginning at any arbitrary time $t > 0$ is sufficient, and there are no boundary conditions imposed for determination of the rate constants. The only requirement is that significant changes in the concentrations occur in the data sets, with respect to the noise. The Numerical Matrices Method can be carried out using not only direct differential methods, but also the direct integral or empirical optimization methods described above. The primary application for these alternative implementations would be when the number of time points is small. For a large number of time points, the condition number of the matrix $\hat{\mathbf{B}}$ is much better for the direct differential method, which is particularly important when the number of parameters to be determined is large.

The linearly quadratic model used here for the kinetics is generally applicable to most chemical reaction networks. However, this choice is not imperative, and other more complex models for the kinetics can be used within the framework of the NMM. In addition, the NMM can be generalized to important cases with nonlinear dependence on parameters, such as the case of Michaelis-Menten enzyme kinetics. A more generalized nonlinear kinetic model of the form:

$$\frac{dX_j(t_i)}{dt} = \frac{\sum_{p=0}^{ns} \sum_{q=0}^p a_{jpq} X_p(t_i) X_q(t_i)}{f_j + \sum_{p=0}^{ns} \sum_{q=0}^p g_{jpq} X_p(t_i) X_q(t_i)}, \quad (37)$$

contains strong nonlinearity of parameters f_j, g_{jpq} . In analogy to Eqs. 2 and 50 in Appendix II in the Supplementary Material, the least squares criterion:

$$\Phi(a_{jpq}, g_{jpq}; f_j) = \sum_{i=1}^{nt} \sum_{j=0}^{ns} \left[\left(f_j + \sum_{p=0}^{ns} \sum_{q=0}^p g_{jpq} X_p(t_i) X_q(t_i) \right) \times \left(\frac{dX_j(t_i)}{dt} \right)_{\text{obs}} - \sum_{p=0}^{ns} \sum_{q=0}^p a_{jpq} X_p(t_i) X_q(t_i) \right]^2, \quad (38)$$

can be used to solve the reaction identification problem for the nonlinear parameters using a tractable system of linear equations for all of the parameters in Eq. 37, including the nonlinear parameters f_j, g_{jpq} .

Multiple initial states

Application of the NMM is complicated by linear dependencies among the species in the time-dependent data. In the absence of knowledge about the reaction network structure or the stoichiometry matrix, the linear system of equations for the rate constants becomes singular. However, recording data sets with multiple initial states where the concentrations of the various species are changed in combination results in a numerically tractable problem. The choice of the number and the nature of the initial states will depend on the nature of the kinetics present, which cannot be known a priori for an unknown reaction structure. This consideration offers an important guiding principle for designing experiments to determine rate constants in complex kinetic systems, and it will be generally necessary to collect data using a range of concentrations of all of the initial species present.

CONCLUSION

The matrix methods described here are based on a very general kinetic framework that is applicable to a wide variety of dynamic systems, particularly chemical and biochemical reaction networks. An important and simplifying principle is that most chemical reactions can be represented by a linearly quadratic kinetic equation, making it possible to enumerate all possible uni- and bimolecular reactions. For all of the above methods, the description of the dynamics is based on the matrix of complexes \tilde{Y} that correspond to these reactions. An important feature of the approach is that rate constants for these reactions can be obtained by solving a linear system of equations with a least squares solution to the observed data in closed form. The various methods can be readily combined with each other and can be applied in succession as the understanding of the dynamics increases. The individual matrix methods constitute important parts of a universal approach to the task of the system identification that we term the Numerical Matrices Method.

The NMM provides an opportunity to use a continuously changing amount of prior information for system identification and rate constant determination. This in turn provides the possibility of developing iterative methods for identification of large chemical and biochemical reaction networks. Information that is obtained in early steps can be subsequently included to iteratively improve the description of the dynamics. A complete implementation of such an iterative procedure will include analysis of errors at each stage to guide the reconstruction of the reaction network. Here we have described the essential elements of the NMM that will provide the basis for further development of such powerful tools for analysis of kinetic data and for system identification.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>.

This work was supported by grants from the National Institutes of Health (TW-007478 to A.V.K. and J.R.W and GM-53757 to J.R.W.).

REFERENCES

- Maria, G. 2004. A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chem. Biochem. Eng. Q.* 18:195–222.
- Crampin, E. J., S. Schnell, and P. E. McSharry. 2004. Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. *Prog. Biophys. Mol. Biol.* 86:77–112.
- Moles, C. G., P. Mendes, and J. R. Banga. 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13:2467–2474.
- Famili, I., and B. O. Palsson. 2003. Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices. *J. Theor. Biol.* 224:87–96.
- Schuster, S., D. A. Fell, and R. Dandekar. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18:326–332.
- Famili, I., and B. O. Palsson. 2003. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys. J.* 85:16–26.
- Forster, J., I. Famili, P. Fu, B. O. Palsson, and J. Nielsen. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13:244–253.
- Schuster, S., C. Hilgetag, J. H. Woods, and D. A. Fell. 2002. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* 45:153–181.
- Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature.* 420:190–193.
- Arkin, A., and J. Ross. 1995. Statistical construction of chemical reaction mechanisms from measured time-series. *J. Phys. Chem.* 99:970–979.
- Arkin, A., P. Shen, and J. Ross. 1997. A test case of correlation metric construction of a reaction pathway from measurements. *Science.* 277:1275–1279.
- Vance, W., A. Arkin, and J. Ross. 2002. Determination of causal connectivities of species in reaction networks. *Proc. Natl. Acad. Sci. USA.* 99:5816–5821.
- Bard, Y. 1974. *Nonlinear Parameter Estimation*. Academic Press, New York and London, UK.
- Hemker, P. W. 1972. Numerical methods for differential equations in system simulation and in parameter estimation. In *Analysis and Simulation of Biochemical Systems*. H. C. Hemker and B. Hess, editors. North Holland, Amsterdam. 59–80.
- Stortelder, W. J. H. 1998. *Parameter Estimation in Nonlinear Dynamic Systems*. Stichting Mathematische Centrum, Amsterdam.
- Stortelder, W. J. H., P. W. Hemker, and H. C. Hemker. 1997. *Mathematical Modelling in Blood Coagulation; Simulation and Parameter Estimation. Modelling, Analysis and Simulation*. Report CWI, Stichting Mathematische Centrum, Amsterdam.
- Bard, Y. 1970. Comparison of gradient methods for the solution of nonlinear parameter estimation problems. *SIAM J. Numer. Anal.* 7:157–186.
- Himmelblau, D. M., C. R. Jones, and K. B. Bischoff. 1967. Determination of rate constants for complex kinetic models. *Ind. Eng. Chem. Fundam.* 6:539–543.
- Hosten, L. H. 1979. A comparative study of shortcut procedures for parameter estimation in differential equations. *Comp. Chem. Eng.* 3:117–126.
- Vajda, S., P. Valko, and A. Yermakova. 1986. A direct-indirect procedure for estimation of kinetic parameters. *Comp. Chem. Eng.* 10:49–58.
- Yermakova, A., and P. Valco. 1985. A remark on parameter estimation in dynamic models. *AIChE J.* 31:1213.
- Yermakova, A., P. Valco, and S. Vajda. 1982. Direct integral method via spline-approximation for estimating rate constants. *Appl. Cat.* 2:139–154.
- Prony, R. 1975. Essai experimental et analytique. *J. de L'ecole Polytechnique.* 1:24–76. [in French].
- Bak, K. 1963. Numerical differentiation and integration in chemical kinetics. *Acta Chem. Scand.* 17:985–988.
- Fay, L., and A. Balogh. 1968. Determination of reaction order and rate constants on the basis of the parameter estimation of differential equations. *Acta Chim. Acad. Sci. Hung.* 57:391–401.
- Garfinkel, D., J. D. Rutledge, and J. J. Higgins. 1961. Simulation and analysis of biochemical systems. *Comm. Assoc. Computing Machinery.* 4:559–562.
- Levenspiel, O. 1962. *Chemical Reaction Engineering*. Wiley, New York.

28. Lindsay, K. L. 1962. Optimum time scheduling of kinetic experiments. *Ind. Eng. Chem. Fundamentals*. 1:241–245.
29. Rudakov, E. 1960. Differential methods of determination of rate constants of noncomplicated chemical reactions. *Kinetics and Catalysis*. 1:177–187.
30. Rudakov, E. 1970. Determination of rate constants. Method of support function. *Kinetics and Catalysis*. 11:228–234.
31. Steiner, R., and K. Schoenemann. 1965. Korrelationsrechnung beim Aufstellen von Geschwindigkeitsgleichungen für Komplexe Chemische. *Chem. Ing. Tech.* 37:101–107. [in German].
32. Feinberg, M., and F. Horn. 1977. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspace. *Arch. Rational Mech. Anal.* 66:83–97.
33. Horn, F., and R. Jackson. 1972. General mass action kinetics. *Arch. Rational Mech. Anal.* 47:81–116.
34. Karnaukhov, A. V., and E. V. Karnaukhova. 2003. Application of a new method of nonlinear dynamical system identification to biochemical problems. *Biochemistry (Mosc.)*. 68:253–259.
35. Karnaukhov, A. V., E. V. Karnaukhova, and V. I. Zarnitsina. 2005. Consideration of prior information and the type of experimental error in the identification method based on minimizing square residuals. *Biophysics (Moscow)*. 50:309–313.
36. Schluenzen, F., A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*. 102:615–623.
37. Wimberly, B. T., D. E. Brodersen, W. M. Clemons Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonnrhein, T. Hartsch, and V. Ramakrishnan. 2000. Structure of the 30S ribosomal subunit. *Nature*. 407:327–339.
38. Traub, P., and M. Nomura. 1968. Structure and function of E. coli ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc. Natl. Acad. Sci. USA*. 59: 777–784.
39. Trevathan, M. T., G. Siuzdak, and J. R. Williamson. 2005. Assembly landscape of the 30S ribosomal subunit. *Nature*. 438:628–632.