

# On the deduction of chemical reaction pathways from measurements of time series of concentrations

Michael Samoilov

*Department of Chemistry, Stanford University, Stanford, California 94305*

Adam Arkin

*Department of Bioengineering and Chemistry, University of California, Berkeley, California 94720*

John Ross

*Department of Chemistry, Stanford University, Stanford, California 94305*

(Received 12 July 2000; accepted for publication 30 October 2000)

We discuss the deduction of reaction pathways in complex chemical systems from measurements of time series of chemical concentrations of reacting species. First we review a technique called correlation metric construction (CMC) and show the construction of a reaction pathway from measurements on a part of glycolysis. Then we present two new improved methods for the analysis of time series of concentrations, entropy metric construction (EMC), and entropy reduction method (ERM), and illustrate (EMC) with calculations on a model reaction system. © 2001 American Institute of Physics. [DOI: 10.1063/1.1336499]

**Complex chemical kinetic systems have many reacting species and catalysts, and the reaction mechanism describes their chemical connections in a series of elementary reactions. Each step gives the reactants which, in a single collision, yield the stated products. We discuss the determination of reaction mechanisms deduced from measurements of the time series of the concentrations of the chemical species in the system. We review the method of using concentration correlation functions to construct a correlation metric for this purpose, and present two improved procedures, which require a larger number of measurements.**

## I. INTRODUCTION

A series of studies have been concerned with new approaches to the deduction of the connectivities of chemical species, the reaction pathway, and the reaction mechanism of complex reaction systems from measurements. These approaches try to go beyond the art practiced in chemical kinetics for at least 100 years, that is of guessing a reaction mechanism and comparing its calculated predictions with measurements. Stable and radioactive tracers have been used to follow the transfer of a given tracer from one molecule to another, or the transformation of a tagged molecule from one type to another.<sup>1</sup> In Refs. 2–9 we presented and tested an approach suitable for oscillatory reactions, which consisted of a categorization of such reactions, the identification of species essential and nonessential for oscillations, and the connectivities of these species, all deduced from several experiments. One class of such experiments is concerned with the determination of Jacobian matrix elements, either from time-series analysis<sup>10,11</sup> or from quenching studies.<sup>12–14</sup> In another approach, still to be developed much further, we introduced the technique of genetic algorithms for a systematic search for a reaction mechanism assigned a specific task.<sup>15</sup> In

this paper we focus on the deduction of reaction pathways and mechanisms from measurements of time series of chemical concentrations.<sup>16,17</sup> In Ref. 16 we presented a new technique, called correlation metric construction (CMC) of reaction pathways. A prediction of the reaction pathway is deduced from time-lagged correlation functions of two chemical species at a time, obtained from concentration measurements. These functions are converted into interspecies distances, which are then used in the construction of a multidimensional object; a specified two-dimensional projection yields the reaction pathway of the reacting system. This approach came from earlier theoretical and experimental work on the implementation of computations by macroscopic, kinetic systems,<sup>18–24</sup> from the theoretical and experimental demonstration that complex biochemical reaction networks implement logic functions,<sup>25,26</sup> and from prior work in electronic circuit theory, system theory,<sup>27,28</sup> and multivariate statistics.<sup>29–31</sup> In Ref. 17 we reported an experimental test of the theory on a part of an *in vitro* glycolysis system containing 8 enzymes and 14 metabolites. We review this work in Sec. II. We then turn in Secs. III and IV to the presentation of alternative and improved, procedures for the analysis of time series of concentrations, and illustrate one of these with calculations on a model reaction system.

## II. CORRELATION METRIC CONSTRUCTION

Consider a reacting system with two sets of species designated I and S; we suppose that we can perturb each of the I species externally. In the absence of such perturbations the system is in a stationary state. The perturbations of the I species are chosen such that the concentration state space of all I+S species is adequately sampled. Imposing the choice of an uncorrelated multivariable Gaussian distribution of the variation of input species concentrations guarantees that the entire state space of the input species is sampled and that there are no auto or cross correlations among the input spe-

cies. For each of the I+S species we have a deterministic kinetic equation for the concentration  $x_j$  of species  $j$ , which may be from within either the I or S set

$$\frac{dx_j}{dt} = F_j(\mathbf{x}), \quad (2.1)$$

where  $\mathbf{x}$  denotes a vector all the I+S species concentrations. The solutions of Eq. (2.1) are time-dependent deterministic trajectories, which lie on a hypersurface of I+S-1 dimensions. If we vary the concentrations of the input species randomly and thus move the system away from its stationary state, then the system will return towards that state along a deterministic trajectory. Repeated random variations of input concentrations and measurement of the responses of the I+S species will then sample the hypersurface effectively, and we obtain a time-series measurement of the stochastic distribution of inputs and responses. The perturbations are chosen to remove the system from its linear regime, and the time interval between concentration measurements are assumed to be less than the slowest relaxation time in the system.

In CMC we form from these measurements time-lagged correlation functions of a pair of species, one such function for each pair,

$$S_{ij}(\tau) = \langle (x_i(t) - \bar{x}_i)(x_j(t + \tau) - \bar{x}_j) \rangle, \quad (2.2)$$

where the bar on  $x_i$  denotes the time average of the concentration of the time series of species  $i$ , and the brackets denote the time average over all measurements. The time interval  $\tau$  may be positive, negative, or zero, and the time-lagged correlation matrix  $R(\tau)$  is defined with the reduced matrix elements

$$r_{ij}(\tau) = \frac{S_{ij}(\tau)}{\sqrt{S_{ii}(\tau)S_{jj}(\tau)}}. \quad (2.3)$$

The pair correlations depend on the elementary reactions in the system and their rate coefficients in a complicated way.

A simple agglomerative dependency algorithm<sup>16</sup> selects the most important, significant correlations and creates a singly linked graph in which every chemical species is connected to at least one other species.

Next the correlation matrix elements are converted into distance matrix elements with the definitions

$$d_{ij} = (c_{ii} - 2c_{ij} + c_{jj})^{1/2} = \sqrt{2}(1.0 - c_{ij})^{1/2}, \quad (2.4)$$

$$c_{ij} = \max|r_{ij}(\tau)|_{\tau},$$

where max specifies the absolute value of the maximum of a given correlation regardless of the value of  $\tau$ .

The distance matrix elements are analyzed with a multi-dimensional scaling (MDS) method to construct a multidimensional object. The inputs for the MDS method are the measurements of time series of concentrations and the outputs are a map of the connections due to correlations of the species. The object presents measures of relatedness of time series; the more related the more likely that two species are connected by a single reaction. The dimensions of the object give an indication of the complexity of the reaction network.

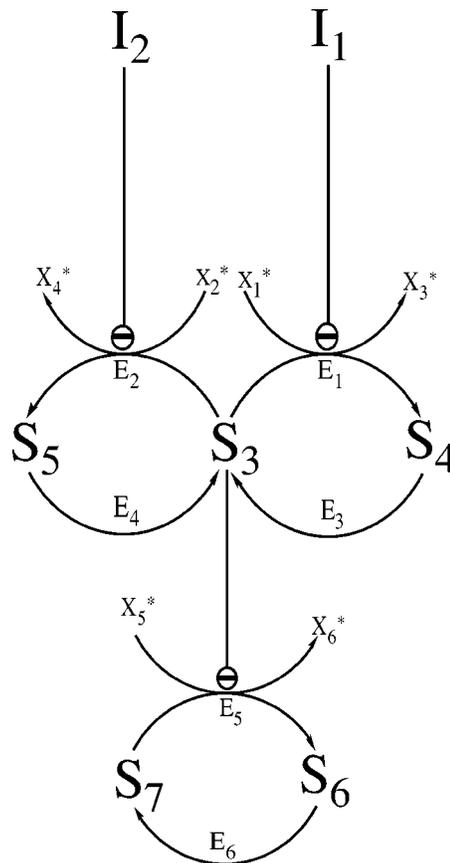


FIG. 1. A model reaction mechanism: the species varied randomly from outside are labeled by I, and other species by S. Species marked with an asterisk are held constant. The enzyme  $E_2$  is inhibited noncompetitively (minus sign in the circle) by  $I_2$ , and similarly for  $E_1$  and  $E_5$ . From Ref. 16.

A projection of that object onto two dimensions can be calculated which provides the most information about the multidimensional object.

Tests of this theory were made on several model reaction systems; one is shown in Fig. 1, which is an open system, a combination of three futile cycles, and functions as a NAND gate. All enzyme catalyzed reactions are given by Michaelis-Menton mechanisms. Calculations of the responses of the S species to random inputs of the I species gives the necessary information for the calculations of the pair correlation functions, the correlation distances, the formation of the correlation object, and its projection onto a plane. This projection is shown in Fig. 2(A), which gives a good representation of the reaction pathway of the model shown in Fig. 1. Figure 2(B) is obtained by allowing the distances between points to be greater, equal to, or less than the actual distances by use of an optimization procedure which minimizes an assumed stress function.

A cluster analysis can also be made which summarizes the groupings of chemical subsystems in the reaction system and gives a hierarchy of interactions among the subsystems. Several caveats about the CMC method are given in Ref. 16, particularly in regard to the simple connection algorithm used for choosing the most important correlations regardless of the time lag  $\tau$ .

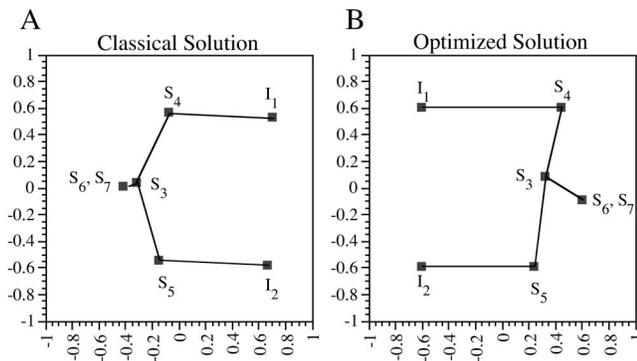


FIG. 2. Results of calculations of the multidimensional scaling analysis of the model reaction system shown in Fig. 1(A). Projection onto two dimensions; the distances between points are equal to or less than the calculated distances in the multidimensional object. (B) A projection obtained with an optimization method, see Ref. 16.

Other models were investigated in Ref. 16, some with more variables, and one with two distinct time scales. The CMC provided the reaction pathways as in the first example and further showed a separation of species according to the separation of time scales.

The theory of correlation metric construction was tested by application to *in vitro* experiments<sup>17</sup> on a part of glycolysis, see Fig. 3. The experiments were carried out in a continuous-flow stirred reactor vessel, with the enzymes being confined to the vessel by a membrane. Capillary electrophoresis was used to analyze quantitatively the outflow from the vessel, which was sampled periodically. The species citrate and AMP were selected for random variations of their concentrations, and consequent perturbation of the glycolysis system from its nonequilibrium stationary state. The concentrations of these two species and six other metabolites were

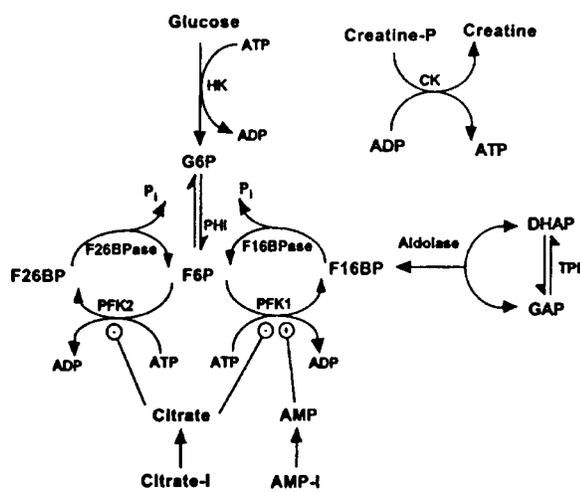


FIG. 3. The first few initial steps of glycolysis; minus (*plus signs*) indicate negative (*positive*) effectors on the respective enzymes. Regulatory interaction: (*minus signs*) a negative effector; (*plus signs*) a positive effector. Creatine-P (phosphate) and C (creatine kinase) keep the concentrations of ATP and ADP constant.  $P_i$ , inorganic phosphate; HK, hexokinase; G6P glucose-6-phosphate; PHI, phosphohexose isomerase; F6P, fructose-6-phosphate; F26BPase, fructose-26-biphosphatase; F26BP, fructose-26-biphosphate; PFK, phosphofructokinase; TPI, triose phosphate isomerase; GAP, glyceraldehyde phosphate. From Ref. 17.

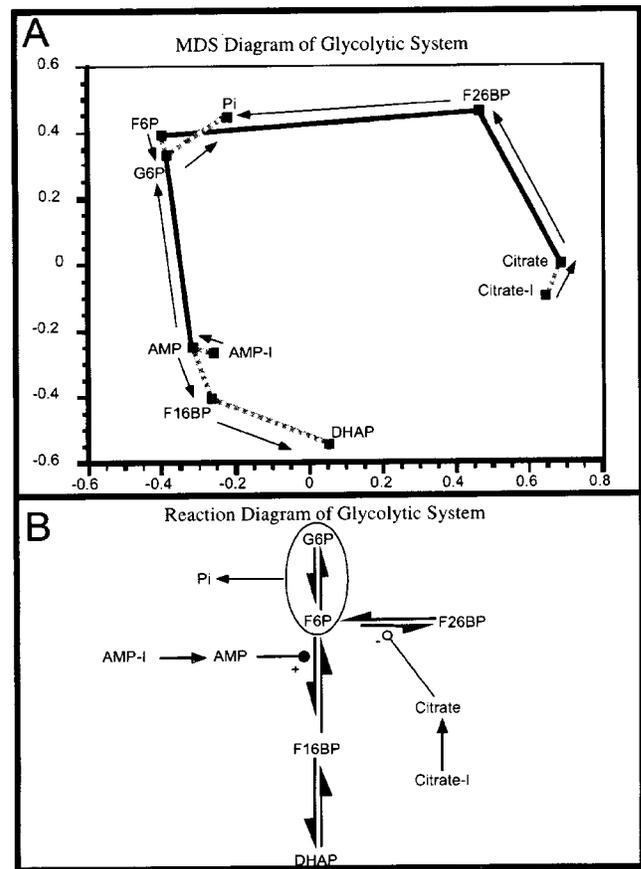


FIG. 4. (A) The two-dimensional projection of the multidimensional scaling diagram for the measured time series. The closer two points are the higher the correlation between the respective time series. Black (*gray*) lines represent negative (*positive*) correlations between the respective species. (*Arrows*) Temporal ordering among species based on the lagged correlations between their time series. (B) The predicted reaction pathway derived from the correlation metric construction diagram. Its correspondence to the known mechanism (Fig. 3) is high. See Fig. 3 for abbreviations. From Ref. 17.

measured at given time intervals. From these measurements two-species time lagged correlation functions were calculated. These results give information on the temporal sequence of events, for instance, whether a perturbation of one species follows or proceeds that of another species. These results combined with multidimensional scaling analysis led to the two-dimensional projection of the MDS object formed, as shown in Fig. 4(A). The reaction pathway deduced from that projection is shown in Fig. 4(B). Both of these deductions from the measurements resemble closely the reaction pathway obtained by traditional methods. The connectivities of the species, as well as the locations of the positive and negative effectors, are obtained correctly. No postulates of any kind about the reaction mechanism had to be made to obtain the results in Fig. 4(A).

The CMC method requires that the random Gaussian inputs elicit responses in the concentrations of the other chemical species which are adequately represented by Eq. (2.2); that expression is the second moment of the pair distribution function, and hence the response correlations should also be Gaussian. This condition seems to suffice for the example studied and for the interpretation of the experi-

ments cited, all of which have nonlinearities in the kinetics. However, as we show in the next section, for certain strong nonlinearities and correlations extending over several species, higher than the second moment of the pair probability distribution function may contribute and an improvement on CMC is required.

### III. ENTROPY METRIC CONSTRUCTION METHOD (EMC)

#### A. General

In this new method we determine from the measurements of time series of concentrations of species more information than in CMC. From the responses to the stochastic inputs we determine the pair distribution function  $p(x_i, x_{j\tau})$  of chemical concentrations in a method explained below. The pair correlation function defined in Eq. (2.1) is the second moment of the pair distribution function and is obtained from it by integration

$$S_{ij}(\tau) = \int (x_i(\tau) - \bar{x}_i)(x_j(t + \tau) - \bar{x}_j) \times p[x_i(t)x_j(t + \tau)] dx_i dx_j, \quad (3.1)$$

where an ensemble average replaces the average over a time series of experiments on a single system. More measurements are needed to determine the pair distribution function than the pair correlation function.

Further we choose a new measure of the correlation distance, one based on an information theoretical formulation.<sup>32,33</sup> A natural measure of the correlation distance between two variables is the number of states jointly available to them (the size of the support set) compared to the number of states available to them individually. We therefore require that the measure of the statistical closeness between variables  $X$  and  $Y$  be the fraction of the number of states jointly available to them versus the total possible number of states available to  $X$  and  $Y$  individually. Further, we demand that the measure of the support sets weighs the states according to their probabilities. Thus two variables are close and the support set is small, if the knowledge of one predicts the most likely state of the other, even if there exists simultaneously a substantial number of other states.<sup>34-36</sup>

The information entropy gives the distance we demand in these requirements. The effective size of the support set of a continuous variable is<sup>37</sup>

$$S(X) = e^{h(X)} \quad (3.2)$$

in which the entropy  $h(X)$  is defined by

$$h(X) = - \int_s p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}, \quad (3.3)$$

where  $S(X)$  is the support set of  $X$  and  $p(x)$  is the probability density of  $X$ . Similarly we denote the entropy of a pair of continuous variables  $X, Y$ , as  $h(X, Y)$ , which is related to the pair distribution function  $p(\mathbf{x}, \mathbf{y})$  by an equation analogous to (3.3). The mutual information  $I(X; Y)$  between two stochastic variables is

$$I(X; Y) = h(X) + h(Y) - h(X, Y) \quad (3.4)$$

or in terms of the effective sizes of the support sets

$$\frac{S(X, Y)}{S(X)S(Y)} = \frac{e^{h(X, Y)}}{e^{h(X)}e^{h(Y)}} = e^{h(X, Y) - h(X) - h(Y)} = e^{-I(X; Y)}. \quad (3.5)$$

We define an EMC correlation distance based on information entropy as the minimum of Eq. (3.5) regardless of the value of  $\tau$ ,

$$d_{xy} = \min_{\tau} \frac{S(X_{\tau}, Y)}{S(X_{\tau})S(Y)} = \min_{\tau} e^{-I(X_{\tau}; Y)}. \quad (3.6)$$

If the correlation of the species  $X$  and  $Y$  is Gaussian, then the expression (3.6) for the EMC distance usually leads to a result similar to that of the CMC distance, Eq. (2.4).

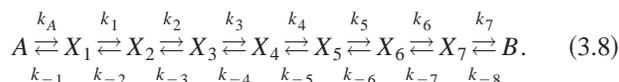
The CMC and EMC distances satisfy the first three requirements of a metric:

- (1)  $d_{xy} \geq 0$ ,
- (2)  $d_{xy} = d_{yx}$ ,
- (3)  $d_{xy} = 0$  iff  $x \triangleq y$ ,
- (4)  $d_{xy} + d_{zy} \geq d_{xz}$

but not the triangular inequality, condition (4). This situation can be remedied with the technique of stress minimization.<sup>38</sup> There is frequently little difference in the results of the prediction of a reaction pathway whether the stress minimization technique is applied or not.

#### B. An example

We choose an example of a reaction mechanism to illustrate the calculation of EMC distances and the prediction of a reaction pathway from simulated time series of concentrations. We compare these results with predictions of the CMC method. The example is chosen to show differences between these two approaches, and to indicate the origin of these differences. The example is the mechanism



All of the reactions are first order with constant rate coefficients, except for the  $A \rightarrow X_1$  step, which is enzyme catalyzed. The forward rates of reactions inside the chain are  $k_i = 0.7$  for all  $i$ . The backward rates for the reactions inside the chain are  $k_{-2} = 0.3$ ,  $k_{-5} = 0.2$ , and  $k_{-i} = 0.1$  otherwise. The species  $A + B$  are held at constant concentrations, with  $A = B = 1$ , and  $X_8$  is varied in a prescribed way (i.e.,  $X_8$  is chosen to be the input species  $\mathbf{I}$ ) with  $k_B = 1.0$ . Finally, the enzyme-catalyzed step has the effective rate coefficient  $k_A = 60.0X_8 / (40.0 + X_8)(60.0 + X_8)$ .

The difficulties in this example arise from the self-inhibition of the enzyme catalysis by  $X_8$ . The rate coefficient  $k_A$  first increases with increasing concentration  $X_8$  and then

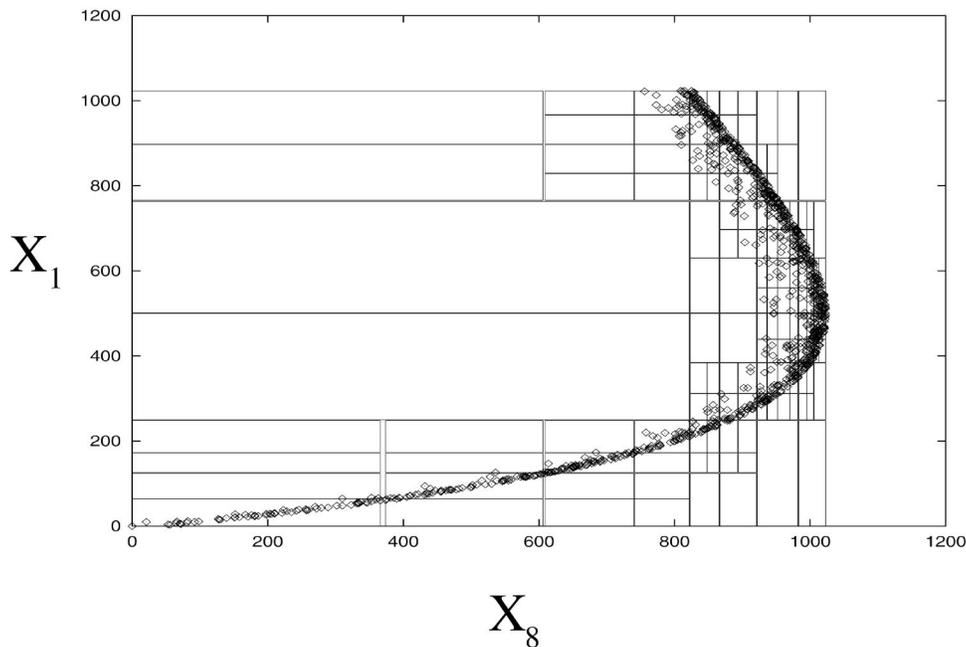


FIG. 5. The diamonds plot the values of  $X_1$  vs  $X_8$  obtained from the simulated time series. The rectangles are the result of a partitioning algorithm, see the text. From Ref. 32.

decreases. The hypersurface formed by eliminating the time dependence from the set of equations (2.1), by dividing the equation for each but one of the species by the equation for that one species, is folded over due to the quadratic dependence of  $k_A$  on  $X_8$ .

In the simulation the concentration of  $X_8$  is varied randomly and the responses of the other species are calculated to give time series of 2000 data points; these series are the starting point for both the EMC and the CMC analysis. The diamonds in Fig. 5 show the values of  $X_8$  vs  $X_1$  obtained from these time series. The folding of the hypersurface in the concentration space is shown in this projection onto the  $X_1$ - $X_8$  plane. The space in this plane is divided into rectangles of varying size so that the distribution of points is uniform in each rectangle to within a given accuracy.<sup>39</sup> The density in each rectangle ( $i, j$ ) where  $i$  and  $j$  are indices of the discretization of the continuous ranges of values of  $X_1$  and  $X_8$ , is the pair probability distribution

$$p_{ij}(x, y) = \frac{N_{ij}}{N_{tot}A_{ij}},$$

where  $N_{ij}$  is the number of points in the particular rectangle labeled  $\{i, j\}$ ,  $N_{tot}$  is the total number of points and  $A_{ij}$  is the

area of the rectangle. We use the pair distribution  $p(X_1, X_8)$  to calculate the singlet distributions  $p(X_1)$  and  $p(X_8)$  by integrations. Then, with Eq. (3.3), and its analog for  $h(X, Y)$ , we calculate EMC distances, see Eq. (3.6), for pairs of species. The primary connections among the species are thus obtained and their corresponding distances, as derived from a multidimensional scaling analysis, are shown in Fig. 6.

The calculations of the CMC distances by means of Eq. (2.4), and subsequent multidimensional scaling analysis, yields a 2D projection, see Fig. 7, that differs from the EMC analysis in two respects. The reaction pathway predicted by EMC correctly shows the close correlation of the enzymatic catalyst  $X_8$  and species  $X_1$ . It also shows correctly the chain of linear reactions from  $X_1$  to  $X_7$ . The CMC method also shows that  $X_8$  is correlated with  $X_1$ , but more weakly. Less importantly, the CMC method yields a wrap around in the placing of species 7. However, the three-dimensional representation of the multidimensional CMC object implemented by stressing the original CMC distances into three rather than two dimensional results in the correct sequencing of the species: the species points lie on a 3D spiral, which when projected onto two dimensional results in the wrap around effect.

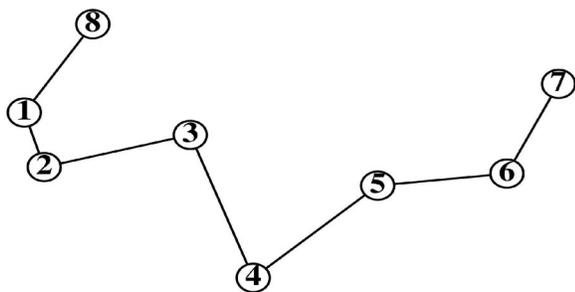


FIG. 6. EMC construction of the reaction mechanism of the example. From Ref. 32.

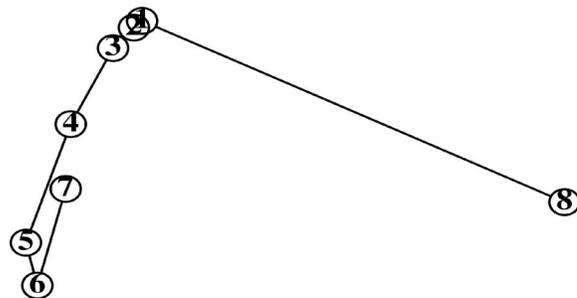


FIG. 7. CMC construction of the reaction mechanism given in Eq. (3.8). From Ref. 32.

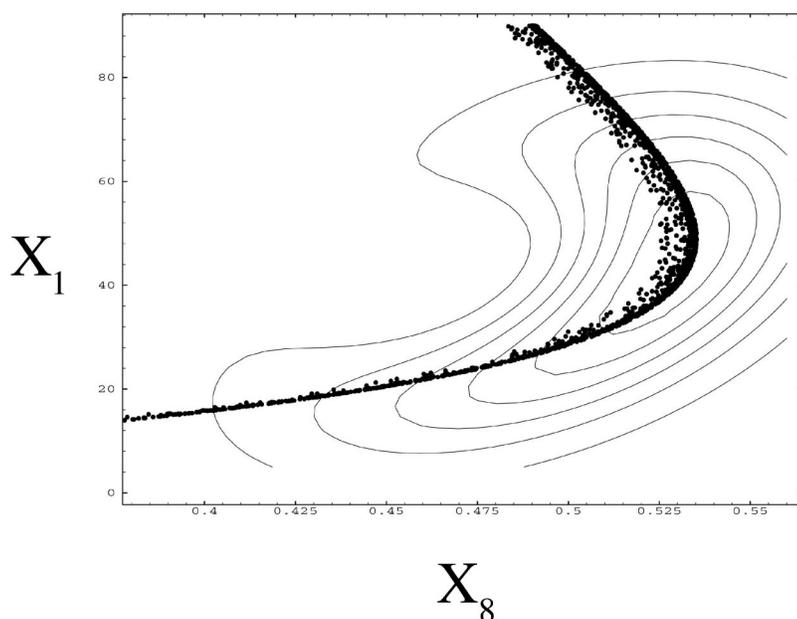


FIG. 8. Contour plot of the pair distribution function  $p(X_1, X_8)$ ; the values of the concentrations of  $X_1$  vs  $X_8$ , as obtained from the simulations, are also shown. From Ref. 33.

The differences in the CMC and EMC predictions can be traced to the different pair probability densities estimated by these methods from the given time series. In Fig. 8, we show a contour plot of the pair distribution function  $p(X_1, X_8)$  as calculated by a semi-non-parametric (SNP) method<sup>40</sup> with the time series simulations for these two species superimposed. In comparison, we show in Fig. 9 a Gaussian pair probability distribution, as is consistent with CMC, with the same means and variances as those of the EMC distribution. The deviations of the EMC from the Gaussian distribution show that higher than second moments contribute. Since the information entropy for any distribution is less than or equal to that of a Gaussian distribution, we see that the CMC method gives upper bounds to correlation distances between species.

The CMC method gives a reasonable approximation to the reaction pathways of many systems, including nonlinear systems; strong nonlinearities may demand the use of the EMC method, which requires more data points than the CMC method; we do not know how widespread the need is for the EMC method.

#### IV. ENTROPY REDUCTION METHOD

EMC provides a more accurate representation of the relationship between two variables because nonlinearities are better described if higher moments of the pair distribution are considered. There is, of course, a price to be paid for this higher accuracy and that is the requirement for large amounts

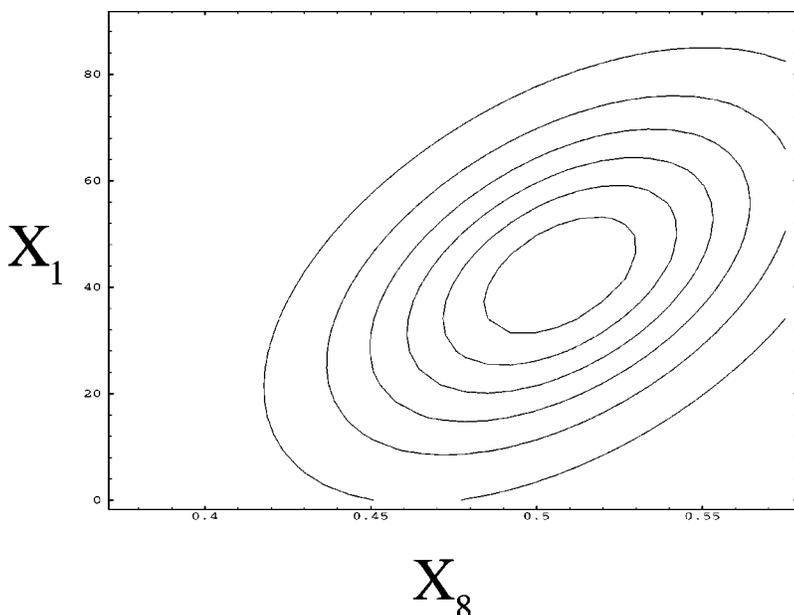


FIG. 9. Gaussian joint normal distribution of the variables  $X_1$  and  $X_8$  with the means and covariances of those of the distribution in Fig. 8. From Ref. 32.

of data. However, even with well-estimated pair distribution functions one can, theoretically, make better, less ambiguous predictions of network structure from these measurements. One possible extension of the EMC method is something we call the entropy reduction method. Here, we essentially require that the nonlinear variation in a variable  $Y$  be explainable by the variations in the concentration of a subset (possibly all) of the other variables in the system. In terms of entropy this condition is obtained by minimizing the differential entropy  $h(Y|\mathbf{X})$  over the set of variable  $\mathbf{X}$ . The term  $h(Y|\mathbf{X})$  is simply  $h(\mathbf{X}, Y) - h(\mathbf{X})$  where  $h(\mathbf{X}, Y) = -\text{Integral}[p(\mathbf{X}, Y)\log(p(\mathbf{X}, Y)/p(\mathbf{X}))d\mathbf{X} dY]$ . If  $Y$  is completely independent of  $\mathbf{X}$  then  $h(Y|\mathbf{X}) = h(Y)$  otherwise it is less than  $h(Y)$ . Formally,  $h(Y|\mathbf{X})$  goes to negative infinity when  $\mathbf{X}$  completely determines  $Y$ . This latter statement is equivalent to saying the size of the support set,  $\text{EXP}(h(Y|\mathbf{X}))$ , is zero. Thus, by iterating through cycles of adding a variable  $x_i$  to  $\mathbf{X}$  that minimizes  $\text{EXP}(h(Y|\mathbf{X}))$  until further additions do not decrease the support set, we can find an ordered set of variables that control the variation in  $Y$ . This technology uses the full joint probability distributions between  $Y$  and the set  $\mathbf{X}$  and so yet again requires a larger amount of data to estimate them. (For multivariate Gaussian distributions, for example, the amount of data needed goes up exponentially with the number of variables.<sup>41</sup>) The quality of data and the measurement error, as in the other methods, also plays a role in determining the effectiveness of this method. However, it may be shown that this method eliminates a great deal of the ambiguity that arises when only the second moment or a few moments of the distribution are used to define the ordering and causality among the chemical concentration variables. For an expanded discussion of this method see Ref. 32.

Recently, driven by advances in molecular profiling techniques such as DNA microarrays, other techniques have been proposed for deduction of interaction networks from molecular concentration time series and/or perturbation data. These include variants of linear regression modeling (with something like SEPATH)<sup>42</sup> and Boolean network reverse engineering.<sup>43</sup> There are also neural network like procedures that have been tried for the discussion developmental gene expression network in *Drosophila*. In addition, there has been much discussion about application of Bayesian Network Analysis to the problem. This is related to some of the information theoretical techniques described above but uses a slightly different set of distributional manipulations. Finally, there has been some progress on particular experimental designs optimized for deduction of causal networks.<sup>44,45</sup>

## ACKNOWLEDGMENTS

This work was supported in part by the Defense Advances Research Project Agency and the Department of Energy (Adam Arkin), and the National Science Foundation (John Ross).

- <sup>1</sup>T. J. Simpson, *Top. Curr. Chem.* **195**, 1 (1998).
- <sup>2</sup>M. Eiswirth, A. Freund, and J. Ross, *Adv. Chem. Phys.* **80**, 127 (1991).
- <sup>3</sup>P. Strasser, J. Stemwedel, and J. Ross, *J. Phys. Chem.* **97**, 2851 (1993).
- <sup>4</sup>T. Chevalier, I. Schreiber, and J. Ross, *J. Phys. Chem.* **97**, 6776 (1993).
- <sup>5</sup>J. Stemwedel, I. Schreiber, and J. Ross, *Adv. Chem. Phys.* **89**, 327 (1995).
- <sup>6</sup>Y. Hung and J. Ross, *J. Phys. Chem.* **99**, 1974 (1995).
- <sup>7</sup>Y. Hung, I. Schreiber, and J. Ross, *J. Phys. Chem.* **99**, 1980 (1995).
- <sup>8</sup>J. Stemwedel and J. Ross, *J. Phys. Chem.* **99**, 1988 (1995).
- <sup>9</sup>Y. Hung, I. Schreiber, and J. Ross, *J. Phys. Chem.* **100**, 8556 (1996).
- <sup>10</sup>J. J. Tyson, *J. Chem. Phys.* **62**, 1010 (1975).
- <sup>11</sup>E. Mihaliuk, H. Skódt, F. Hynne, P. G. Sørensen, and K. Showalter, *J. Phys. Chem.* **103**, 8246 (1999).
- <sup>12</sup>F. Hynne, P. G. Sørensen, and T. Möller, *J. Chem. Phys.* **98**, 211 (1993).
- <sup>13</sup>P. G. Sørensen and F. Hynne, *J. Phys. Chem.* **93**, 5467 (1989).
- <sup>14</sup>F. Hynne, P. G. Sørensen, and K. Nielsen, *J. Chem. Phys.* **92**, 1747 (1990).
- <sup>15</sup>A. Gilman and J. Ross, *Biophys. J.* **69**, 1321 (1995).
- <sup>16</sup>A. Arkin and J. Ross, *J. Phys. Chem.* **99**, 970 (1995).
- <sup>17</sup>A. Arkin, P. Shen, and J. Ross, *Science* **277**, 1275 (1997).
- <sup>18</sup>A. Hjelmfelt, E. Weinberger, and J. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10983 (1991).
- <sup>19</sup>A. Hjelmfelt, E. Weinberger, and J. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 383 (1992).
- <sup>20</sup>A. Hjelmfelt and J. Ross, *J. Chem. Phys.* **96**, 7019 (1992).
- <sup>21</sup>A. Hjelmfelt, F. Schneider, and J. Ross, *Science* **260**, 335 (1993).
- <sup>22</sup>A. Hjelmfelt and J. Ross, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 63 (1994).
- <sup>23</sup>A. Hjelmfelt and J. Ross, *Physica D* **84**, 180 (1995).
- <sup>24</sup>J. Laplante, M. Pemberton, A. Hjelmfelt, and J. Ross, *J. Phys. Chem.* **99**, 10063 (1995).
- <sup>25</sup>A. Arkin and J. Ross, *Biophys. J.* **67**, 560 (1994).
- <sup>26</sup>D. Hauri, P. Shen, A. Arkin, and J. Ross, *J. Phys. Chem.* **101**, 3872 (1997).
- <sup>27</sup>G. J. Klir, *Int. J. Gen. Syst.* **7**, 1 (1981).
- <sup>28</sup>R. C. Conant, *Int. J. Gen. Syst.* **14**, 125 (1988).
- <sup>29</sup>F. H. C. Marriott, *The Interpretation of Multiple Observations* (Academic, New York, 1974).
- <sup>30</sup>K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis* (Academic, San Francisco, 1979).
- <sup>31</sup>W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numeical Recipes in C* (Cambridge University Press, New York, 1988).
- <sup>32</sup>M. Samoilov, Ph.D. thesis, Stanford University, 1997.
- <sup>33</sup>G. S. Michaels, D. B. Carr, M. Askenazi, X. Wen, S. Fuhrman, and R. Somogyi, *Pacific Symposium on Biocomputing* **3**, 42 (1998).
- <sup>34</sup>S. Liang, S. Fuhrman, and R. Somogyi, *Pacific Symposium on Biocomputing* **3**, 18 (1998).
- <sup>35</sup>S. Fuhrman, M. J. Cunningham, X. Wen, G. Zweiger, J. J. Seilhamer, and R. Somogyi, *BioSystems* **55**, 5 (2000).
- <sup>36</sup>A. J. Butte and I. S. Kohane, *Pacific Symposium on Biocomputing* **5**, 401 (2000).
- <sup>37</sup>T. Cover and J. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- <sup>38</sup>R. N. Shephard, *Science* **210**, 390 (1980).
- <sup>39</sup>A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**, 1134 (1985).
- <sup>40</sup>A. Gallant and G. Tauchen, *New Directions in Time Series Analysis, Part II* (Springer-Verlag, New York, 1992), p. 71; SNP: A Program for Nonparametric Time Series Analysis, Version 8.4. User's Guide, Chapel Hill, 1995.
- <sup>41</sup>B. W. Silverman, *Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Applied Probability* **26** (Chapman and Hall, New York, 1986), p. 94.
- <sup>42</sup>P. D'Haeseleer, X. Wen, S. Fuhrman, and R. Somogyi, *Pacific Symposium on Biocomputing* **4**, 41 (1999).
- <sup>43</sup>S. Liang, S. Fuhrman, and R. Somogyi, *Pacific Symposium on Biocomputing* **3**, 18 (1998).
- <sup>44</sup>T. Akutsu, S. Miyano, and S. Kuhara, *Pacific Symposium on Biocomputing* **4**, 17 (1999).
- <sup>45</sup>S. Akutsu, T. Kuhara, O. Maruyama, and S. Minyano, *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms. ACM-SIAM*, 1998.