Acc. Chem. Res. **2003,** *36,* 839−847

New Approaches to the Deduction of Complex Reaction Mechanisms

JOHN ROSS*

Department of Chemistry, Stanford University, Stanford, California 94305-5080

Received March 27, 2003

ABSTRACT

The formulation of a macroscopic reaction mechanism, the sequence of elementary reaction steps by which reactants are turned into products, is difficult. We review several new methods of determining the causal connectivity of chemical species, the reaction pathway (the sequence of chemical species), and the reaction mechanisms of complex reaction systems from prescribed measurements and theories.

I. Introduction

Chemical and biochemical reaction systems may have many species: reactants, products, intermediates, catalysts, and positive and negative effectors on the catalysts. These species may be involved in many elementary reaction steps, each of which details the particular reactants and products involved in a single reactive collision. The sum total of these elementary steps constitutes the reaction mechanism of the given system by which the initial reactants are turned into final products.

The goal of establishing reaction mechanisms has long been sought in chemistry. For more than 100 years this goal was approached by (1) identifying individual chemical species, either by physical or chemical means; (2) isolating the species contributing to one elementary step in the mechanism of that system; (3) determining the stoichiometry of that step; and (4) determining the kinetics of that step. This has been an arduous task, in part due to the difficulties, until recent years, of measuring the concentrations of more than a few species as a function of time. The use of radioactive tracers has helped significantly. When these tasks were done, then began the guessing of the reaction mechanism, followed by writing the kinetic equations for the hypothesized mechanism and deducing kinetic predictions to be compared with available experiments. If the predictions fit the experiments, the guessed mechanism is possible but not necessary. There has not existed a prescribed method of deducing a reaction mechanism from measurements; there are a few clues for small systems (few species),[1] but no more.

This Account describes some new approaches to the deduction, not guessing, of reaction mechanisms, reaction pathways, which contain less information, and causal connectivites of the chemical species from specially designed experiments and necessary theories for their interpretation. (For several such theoretical studies of genetic networks, restricted mostly to Boolian networks, see refs 2−4.) The deduction of a mechanism from experiments still leads only to a sufficient mechanism, not a necessary or unique one, in that more measurements may lead to changes in the mechanism. The advantage of our approaches lies in prescribing definite procedures for obtaining reaction mechanisms and pathways and their predictions: if they check with experiments they are sufficient, but not necessary or unique.

Compare the determination of the long-practiced art of guessing reaction mechanisms as described above with the determination of the logic functions of an electronic device. An electronic engineer imposes electronic inputs (voltages, currents) and measures outputs of the entire system; this leads to the construction of a truth table from which the functions of the device, and at least some of its components, may be deduced. A chemist would take a sledgehammer and knock the device to pieces, look for circuit elements such as transistors, capacitors, etc., and from that information try to guess the functions of the device (overstated, but indicative).

We learned (slowly) to follow the example of the analysis of electronic devices and apply it to chemical kinetics.[5] First we showed the possibility of constructing logic devices by means of macroscopic kinetics. For example, if and only if the concentrations of species 1 and 2 are high, then and only then is the concentration of species 3 high; the mechanism acts as a logic AND gate. We also constructed on paper various logic gates and with these designed a sequential computer called a universal Turing machine, such as a pocket computer and many much larger machines. We then constructed a parallel computer, first on paper, and then we used chemical bistable systems to carry out the computations of a pattern recognition experiment, the first by implementation of a computation by macroscopic chemical kinetics.[6]

Next we came to the question if chemical, or more pertinently biochemical, systems can carry out computations. We answered that question in the affirmative by analyzing a large part of the glycolytic pathway, including the tricarboxylic acid cycle, to show that the bifurcation from fructose-6-phosphate to 1,6- and 2,6-fructose biphosphates acts as a logic gate that controls the important switch from glycolysis to gluconeogenesis. This gate is not Boolian, with a sharp transition from one path to its effective reverse, which would serve poorly, but a fuzzy logic gate that changes slowly from one path to the other

John Ross was born in 1926 in Vienna, Austria. He emigrated to the United States in 1940. He obtained his B.S. from Queens College, NY; served in the U.S. Army, 2nd Lt, 1944−46; and received his Ph.D. from MIT in 1951. He is currently Professor Emeritus at Stanford University. He has received the National Medal of Science, the Peter Debye Award, the Irving Langmuir Award, and the Austrian Cross of Honor for Science and Art. He is a member of the National Academy of Science and the American Academy of Arts and Science. His research interests are in thermodynamic and stochastic analysis and experiments of systems far from equilibrium.

* E-mail: john.ross@stanford.edu.

10.1021/ar020285f CCC: $25.00
Published on Web 09/05/2003

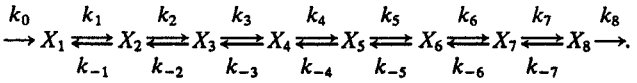

FIGURE 1. An unbranched chain of reversible first-order reactions.

as it responds to multiple inputs of many effectors that control the gate.[7]

If computational functions are indeed built into biochemical reaction pathways and mechanisms, then new approaches with suggestions from circuit theory, system analysis, multivariate statistics, and others may be available for the analysis of chemical kinetic systems as a whole, with retention of all interactions among the species. Keep the system together and measure the concentrations of as many species as possible. With advances in several fields of analytic instrumentation such as a variety of mass spectrometers, capillary electrophoresis, high-pressure liquid chromotography, and others, the measurement of the concentrations of many, if not all, species at a time and as a function of time has become possible. Hence it is timely to consider the design of suitable experiments and appropriate theories for the analysis of complex reaction systems in a way such that we may deduce from available experiments, rather than guess, reaction pathways and mechanisms. If species are not detected with the measurements available, for example, short-lived species, then in general little can be said. Sometimes chemical knowledge may indicate a missing species; sometimes inferences about species not measured may be made.

We begin with our latest approach, pulse change of concentrations, which is conceptually the simplest.

II. Determination of Causal Connectivities of Species in Reaction Networks (Mechanisms)

A. Introduction and Theory.[8] Consider the simple case of a sequence of first-order reactions as shown in Figure 1. Let this isothermal system be in a stationary state, not at equilibrium, where the concentrations are constant but there is a steady flux, say from left to right; or at equilibrium where there is zero flux. The stationary state not at equibrium is maintained by a balance of mass flux into the system (k_0) and out of the system (k_8). Now pulse (increase) the concentration of one of the species by an arbitrary amount. The theoretical analysis of the pulse methods is given in ref 8, and an experimental test of that method is presented in ref 9. (For studies of reaction mechanisms by pulse methods restricted to small pulses, see refs 10−12 for chemical systems. A theoretical study of functional, not molecular interactions, in signaling and gene networks is presented in ref 13.) As the concentration of one species is changed, other species respond with changes of concentrations. For a pulse in species X_1 in Figure 1, the responses of the other species are shown in Figure 2. Hence the order of the responses yields the causal (direct) connectivities of the species in the reaction mechanism.

If the system in Figure 1 is in a nonequilibrium stationary state, then there is a net flux, say from left to

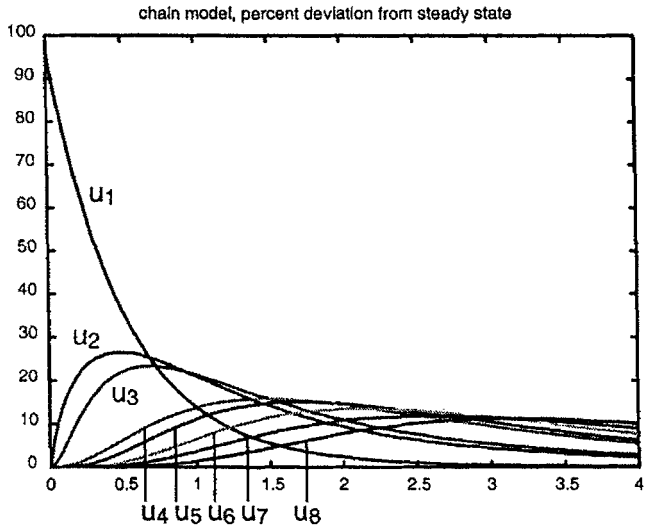

FIGURE 2. Plots of the relative deviation in concentration from the stationary state versus time for all the species of the mechanism in Figure 1. The maxima are ordered according to the number of reaction steps separating that species from the initially perturbed species.

right. If we pulse some other species, say X_5, then that pulse will propagate downhill in Gibbs free energy, to the right strongly, but uphill to the left weakly. If the reverse rate coefficients k_{-1} etc. are all set to zero, then the reaction is irreversible, and no pulse of any species propagates to the left.

We see in Figure 2 that a maximum occurs in the relative concentration of any species between the curves for the preceding and succeeding species, a result which is predicted by the solutions of the kinetic equations of system 1. Several rules often hold for systems of this type: (1) The time of (appearance of) an extremum increases (and its amplitude decreases) as the number of reaction steps separating that species from the initially perturbed species increases, unless some species act as effectors in distant reactions. (2) Conversely, the initial curve of the relative concentration changes of a species with time approaches the time axis (closer) as the number of reaction steps separating that species from the initially perturbed species increases. (3) Species that are directly connected through reactions to the initially perturbed species exhibit nonzero initial slopes. (4) Species that are not directly connected through reactions to the initially perturbed species exhibit zero initial slopes. (5) All responses are positive deviations from the stationary state unless there is a feedback, feedforward, or higher order (>1) kinetic step. (6) For short times, before the exit of material from the pulse to the surroundings of system, the concentration change of the pulse is conserved: the sum of deviations of concentrations (weighted by stoichiometric coefficients) is constant and equal to the change in concentration of the initial pulse. This property is useful in determining that all species produced from the pulse through reactions have been detected, and can help in determining correct stoichiometric coefficients.[8]

If tracers (radioactive, fluorescent,etc.) are used, such that the concentration of the sum of the labeled and

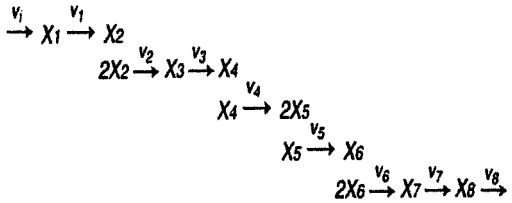

FIGURE 3. A linear chain of coupled first- and second-order reactions.

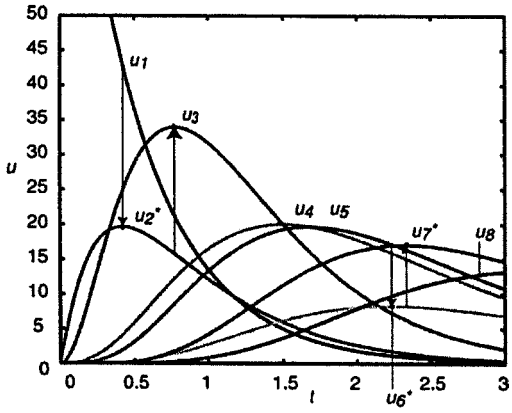

FIGURE 4. Plots of relative deviation in concentration versus time for species of the mechanism in Figure 3. A pulse perturbation of the concentration of species X_1 results in the responses shown.

unlabeled compound is constant, then the response to a tracer perturbation is always first order, even if the kinetics of the system is nonlinear.

An example of mixed first- and second-order irreversible reactions is shown in Figure 3. These are also respectively unimolecular and bimolecular reactions if each step is an elementary step, which need not be the case. For example, a first-order reaction may describe an enzymatic (Michaelis−Menten) mechanism at high substrate concentration. If X_1 is pulsed in this system, then the responses are as shown in Figure 4. The occurrence of the maxima of the responses is ordered in time according to the distances from the pulsed species in the mechanism. There are approximate relations among the maximum responses in the relative concentrations, labeled u_i^*,

$$u_2^* \approx (^1/_2)u_1; \quad u_3^* \approx 2u_2; \quad u_6^* \approx (^1/_2)u_5; \quad u_7^* \approx 2u_6 \quad (1)$$

which can be derived from the deterministic kinetic equations for this system for small pulses. The coefficients give the stoichiometric coefficient of 2 in front of X_2, X_5, and X_6.

Complex reactions may occur also in branching or coalescing chains (see Figure 5); in cycles; with feedforward and feedback reactions; and any combinations thereof. An illustration of feedback is given in Figure 6. If species X_7 is pulsed in that mechanism, and X_7 increases the rate v_3 as indicated, then the concentration X_3 will decrease initially from its value in the stationary state and then return to that value.

In ref 8 we work out an example of the application of the pulse method to a simple model of glycolysis by

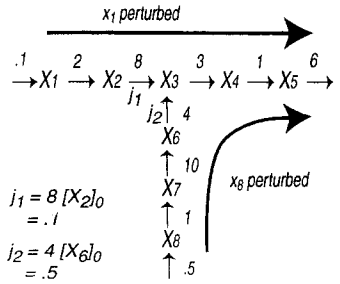

FIGURE 5. Chemical reaction mechanism for converging chains of irreversible first-order reactions. The rates of production of species X_1 and X_8 are held constant at 0.1 and 0.5, respectively. $[X_i]_0$ denotes the stationary-state concentration of species X_i.

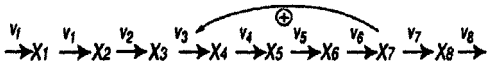

FIGURE 6. Illustration of positive feedback in a linear chain of first-order reactions.

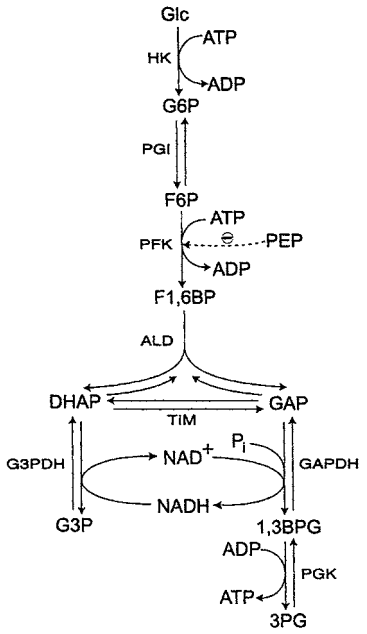

FIGURE 7. Experimental test system. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F1,6BP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone 3-phosphate; GAP, glyceraldehydes 3-phosphate; G3P, glycerol 3-phosphoglycerate; CSTR, continuous-flow stirred-tank reactor.

solving the kinetic equations numerically. We thus generate response curves which are simple to interpret for the deduction of the connectivity and reaction pathway. Rather than discuss this exercise, we turn to an experimental test of the pulse method.

B. Experimental Test of the Pulse Method. We tested[9] the pulse method for determining reaction systems with the experimental system shown in Figure 7, which is a part — the beginning — of glycolysis. This system is studied in a continuous stirred tank reactor (CSTR), a stirred vessel with an inflow and outflow. The enzymes in Figure 7 are introduced at an initial time and kept in the vessel at a determined concentration by a membrane which blocks their outflow, but not that of the remainder of the system. When a stationary state has been achieved, which

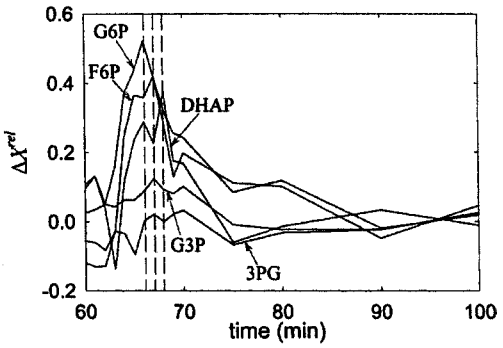

FIGURE 8. Responses in relative concentration to a pulse of G6P.

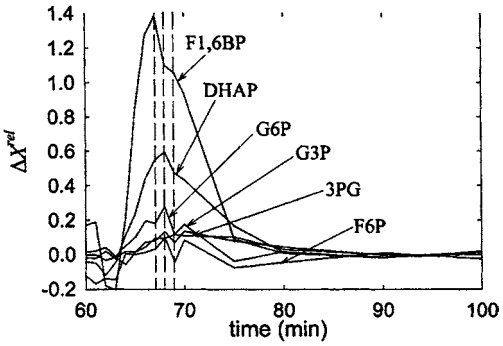

FIGURE 9. Responses in relative concentration to a pulse of F1,6BP.

takes about 30 min, one of the metabolites is pulsed by an injection of that species into the inflow. The responses of six metabolites are measured with capillary electrophoresis until the concentrations again return to their values at the stationary state. The measurements are of modest precision: typical relative errors were 4% for G6P, 11% for F6P, 15% for F1,6BP, 9% for DHAP, 6% for 3PG, and 3% for G3P. However, this precision is sufficient for obtaining substantial information about the reaction mechanism.

Measurements of the responses in relative concentration due to a pulse in G6P are shown in Figure 8. The temporal order of propagation of the pulse is from G6P to F6P, then DHAP, G3P, and 3PG. The time of appearance of the maximum response of the first three species is in that same order. F1,6BP was not measured adequately in this pulse and is not shown.

Figure 9 shows the responses due to a pulse of F1,6BP. DHAP follows, and then G3P and 3PG. The uphill species (in Gibbs free energy), as determined from Figure 7, respond with low amplitude (the order of G6P and F6P is the reverse from that expected from Figure 7, but the precision of the measurements is low at low amplitude).

The response to a pulse of DHAP is interesting (Figure 10). F1,6BP has a substantially higher response than expected, unless there is a stoichiometric coefficient other than unity (see Figure 4, the curve for u_3 and the corresponding reaction $2X_2 \rightarrow X_3$). G3P and 3PG follow DHAP.

The only response to a pulse of G3P is the one of that species and DHAP. There are no responses to a pulse of 3PG other than that species. 3PG is therefore at the end of the line, a branch of the reaction mechanism, with no

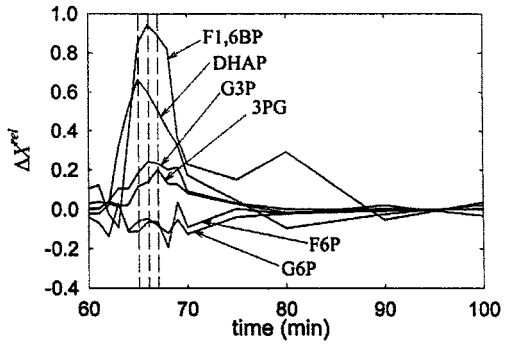

FIGURE 10. Responses in relative concentration to a pulse of DHAP.

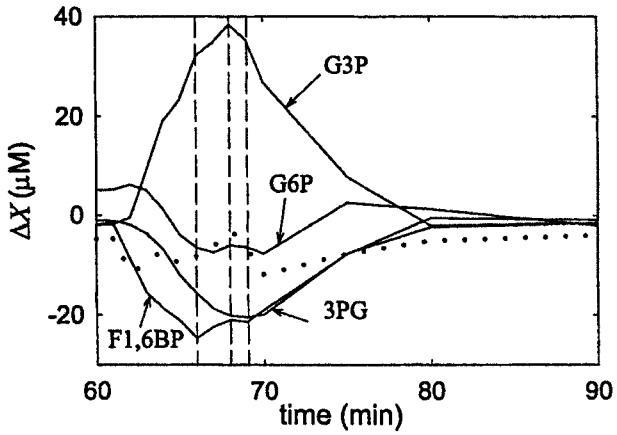

FIGURE 11. Responses in concentration changes to a pulse of NADH. The dotted line is the sum of the variations of F1,6BP, G3P, and 3PG.

significant responses of uphill species, but not the same branch that has G3P in it.

Although we did not measure the concentration of NADH, we applied a pulse of that species and found the responses shown in Figure 11. Note the very different responses of 3PG and G3P, which confirm that these two species are indeed in different branches. NADH is either an effector or a reactant which increases G3P, and at the same time either a reactant or effector that decreases 3PG. Both G3P and 3PG are connected to DHAP, which shows no change. The effect of NADH works through DHAP to F1,6BP which shows a change in concentration. The dotted line gives the sum of the changes of F1,6BP, G3P, and 3PG, which is essentially constant.

These experiments lead by deduction from them to the reaction pathway shown in Figure 12. The first attempt at that deduction was made by assigning random integers to the chemical species so as not to be prejudiced in our deductions from the measurements. We then repeated this process by using the names of the species, and obtained the same results. We did the analysis not using common knowledge such as necessary stoichiometric relations due to conservation of mass. We did assume that DHAP and GAP are in equilibrium (quickly attained), which is the case and which we tested. We were not able to measure GAP. That PEP is an inhibitor of PFK was found in a separate experiment.

The main features of the reaction pathway are well predicted by the pulse method, in particular the bifurca-

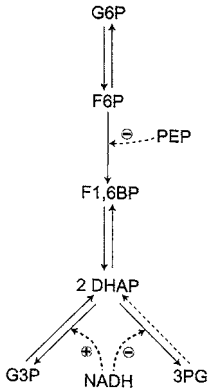

FIGURE 12. Proposed reaction scheme based on experiments. Dashed lines with circles indicate that activation (+) or inhibition (−) may be effected by a metabolite either as a substrate or product, or as an effector. PEP, phosphoenolpyruvate.

tion of the sequence G6P, F6P, F1,6BP into two branches of reaction sequences, one ending in G3P and the other in 3PG. The fast equilibrium of DHAP with GAP (not measured) places that bifurcation at DHAP rather than that shown in Figure 7. Much of the reaction mechanism can be deduced from the reaction pathway.

We believe the pulse method to be relatively simple, effective, and generally applicable.

III. Statistical Construction of Reaction Mechanism. Correlation Functions from Measurements of Time Series of Concentrations

A. Introduction and Theory.[14] In the last section we discussed the issue and measurements of the causal connectivity of species in a reaction mechanism. We now turn to related concepts, those of the correlation of time series of reacting species and correlation metric construction (CMC), and their relations to the reaction mechanism of the system. Causally connected species are generally highly correlated; however, highly correlated species may, but need not be causally directly connected, as for instance in branched networks or networks with feedback. The goal of CMC is the determination of reaction pathways and mechanisms, the regulatory structure of the mechanism, and the connectivity of the species from the measured responses of the species to imposed fluctuations of some chosen species.

Consider a simple hypothetical reaction network such as that shown in Figure 13, which is common in biochemical reactions. Let this open system be maintained in a nonequilibrium stationary state. Perturb the concentrations of the arbitrarily chosen species I_1 and I_2 randomly by arbitrary amounts, and let the system relax back toward the stationary state after each perturbation. Measure during this relaxation the concentrations of all seven species as a function of time (the enzymes E_i are at constant concentrations).

The reciprocal of the characteristic frequency of the random variations of the species I_1 and I_2 is of the order of the longest relaxation time in the system. For given rate

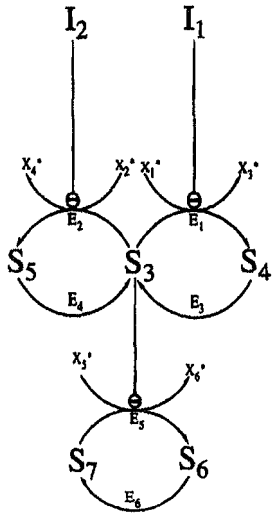

FIGURE 13. Chemical reaction mechanism representing a biochemical NAND gate. All species with asterisks are held constant by buffering. Lines ending in a circle-enclosed minus sign over an enzymatic reaction step indicate that the corresponding enzyme is inhibited (noncompetitively) by the relevant chemical species.

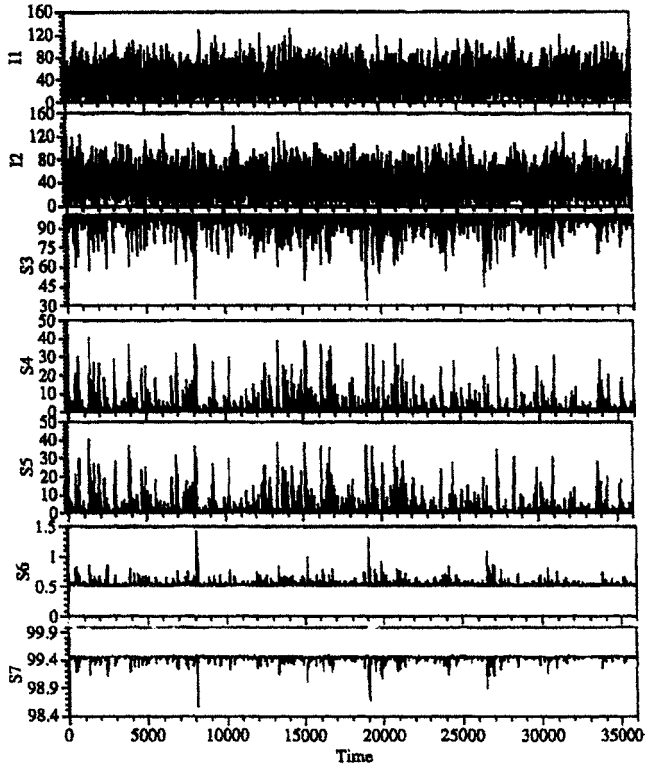

FIGURE 14. Plot of the calculated concentration time series for all the time-varying species composing the mechanism in Figure 13. Only the first two time courses (those for I_1 and I_2) are set by the experimenter. The concentration of I_1 and I_2 are chosen independently from a Gaussian distribution with a mean and standard deviation of 30.0 concentration units.

coefficients for the system in Figure 13, the responses of the species S_3 to S_7 to the imposed fluctuations (perturbations) in I_1 and I_2 are shown if Figure 14. Note that this, as well as most other chemical reaction mechanisms, may act as frequency filters of various types, and this property may have applications in biological systems (see ref 15).

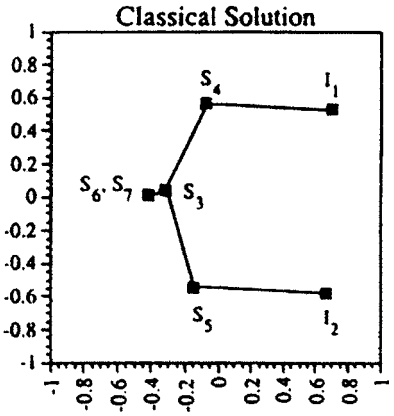

FIGURE 15. Multidimensional scaling analysis obtained from the correlation functions calculated from the time series of concentrations shown in Figure 14. The scales on the axes give correlation distances.

From these time series, correlation functions are formed, for example the correlation of species i and j,

$$S_{ij}(\tau) = \langle (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j) \rangle \qquad (2)$$

where $x_i(t)$ is the concentration of species i at time t, $\bar{x}_i$ is the average concentration of that species over time for a given time series, and τ is a chosen time interval. Some representative correlations are shown later. We normalize these correlations,

$$r_{ij}(\tau) = \frac{S_{ij}(\tau)}{\sqrt{S_{ii}(\tau)S_{jj}(\tau)}} \qquad (3)$$

define the maximum of that correlation for any τ, and define a distance,

$$d_{ij} = (c_{ii} - 2c_{ij} + c_{jj})^{1/2} = \sqrt{2}(1.0 - c_{ij})^{1/2} \qquad (4)$$

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

If the correlation r_{ij} is large, say the maximum of unity, then the distance d_{ij} is zero; if there is no correlation, $r_{ij} = 0$, then $d_{ij} = 1.41$ (an arbitrary number). With these distances we can carry out a mathematical procedure called multidimensional scaling analysis to build an object.[14] A simple description of this procedure is this: take a stick and write the number of one of the species on one end of the stick, and the number of another species on the other end. The stick is small for large correlations and larger for smaller correlations. Pick all the ends of sticks with the number one and place these ends at a point. Do the same with the number two, and so on for all the species. You will need a multidimensional space to accomplish the task of building this object.

Shine a light beam on the object and project its image on a screen. Rotate the object until its image on the screen gives you the maximum information about the object. If all this, or its mathematical equivalent, is done with the reaction mechanism in Figure 13, then we obtain the projection in Figure 15. The projection of the multidimensional object constructed from the correlation distances (CMC) gives quite closely the reaction pathway shown in Figure 13.

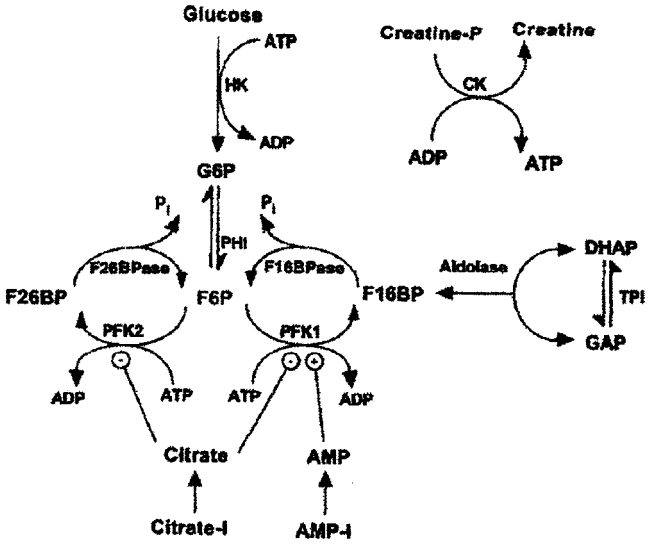

FIGURE 16. The first few reaction steps of glycolysis. Regulatory interaction: (−) a negative effector, (+) a positive effector. Creatine-P (phosphate) and CK keep the concentrations of ATP and ADP constant. Abbreviations: P_i, inorganic phosphate; HK, hexokinase; PHI, phosphoinositol; F26Bpase, fructose-2,6-biphosphatase; TPI, triphosphoinositol; GAP, glutamate phosphate.

With seven species there are $(7 \times 6)/2$ binary correlations. We retained only the ones shown in Figure 15 by a procedure which ensures that each species is connected to at least one other species, and only the largest correlations are kept. Species 6 and 7 are in a single point: they are completely correlated by conservation of mass. The closer connection of species 1 to species 4, rather than 3, depends on the rate coefficients in the S_3-to-S_4 interconversion. The closeness of species 3 to species 6 and 7 indicates a point of control of 3 on 6 and 7. Such information, available from correlation metric constructions, is not available from the usual listings of elementary reactions steps in a reaction mechanism.

For further testing of CMC, we chose another example with two groups each having several futile cycles; to one of these groups we assigned faster reactions than for the other group (so that we have a two-time-scale reaction mechanism). In this case the correlation diagram analogous to Figure 15 showed a clear separation of the two groups, and hence the existence of two time scales. It also represented the reaction pathway of each group.

We have presented only the simplest analysis. There are more sophisticated methods, such as multiple regression analysis, which can provide information about missing (or not measured) variables.[14]

B. Experimental Test of CMC. To test the correlation metric construction method,[16] we chose a part of the much studied glycolysis system shown in Figure 16, which differs in some details from that in Figure 7. The system is established in a nonequilibrium stationary state with a

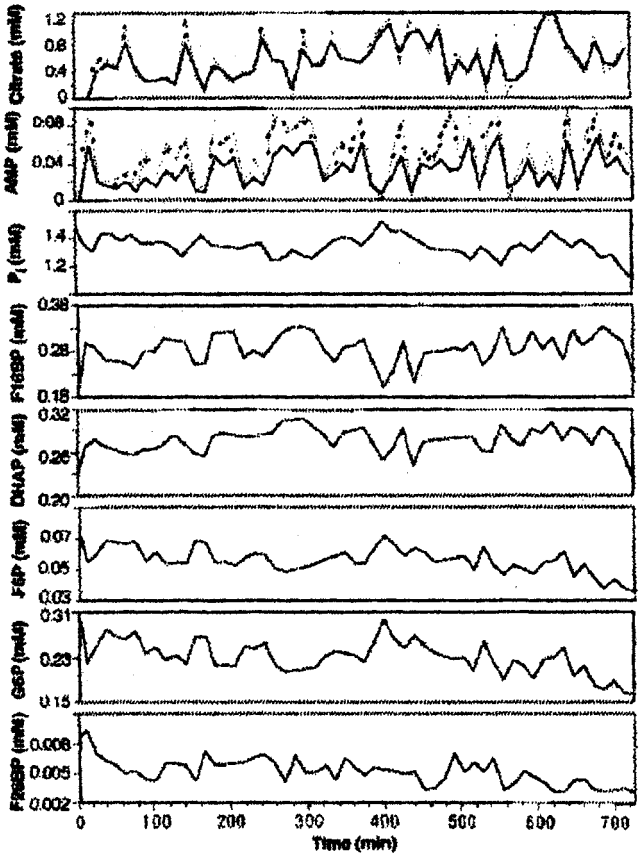

FIGURE 17. Time courses of measured concentration of the inputs AMP and citrate in the experiments with the responses of the concentrations of P_i and the species F1,6BP, DHAP, F6P, G6P, and F2,6BP. The dotted curves are the known input concentrations for AMP and citrate.

constant inflow of glucose and buffer. Metabolites were measured by capillary electrophoresis; the concentrations of the enzymes were kept constant (see section II.B); and the ATP/ADP ratio was held constant. The two effectors citrate-1 and AMP-1 were chosen for the species to be perturbed randomly by arbitrary amounts. All the metabolites listed were measured at regular intervals as, after each perturbation, the system returned to its nonequilibrium stationary state. Typical measurements are shown in Figure 17. A few of the correlations are given in Figure 18. The correlation of G6P with itself peaks at zero time lag τ and decays symmetrically with positive and negative τ, which shows that G6P is not in a stationary state during this perturbation. The correlation of G6P with AMP-1 is larger for positive than negative τ, which indicates that a variation in AMP-1 precedes a variation in G6P. From such information knowledge is obtained about the connectivity of the species.

From the measured correlations we constructed, as explained earlier, the multidimensional diagram (also called the correlation metric construction) shown in Figure 19A. Solid lines indicate negative correlations, shaded lines positive correlations; arrows indicate the time sequence of events: for example, an increase in AMP is followed by a decrease in G6P and an increase in F1,6BP. In Figure 19B the MDS diagram of Figure 19A is rearranged to show the usual reaction pathway determined over many years of effort. The agreement with prior work is excellent and shows the viability and utility of the CMC

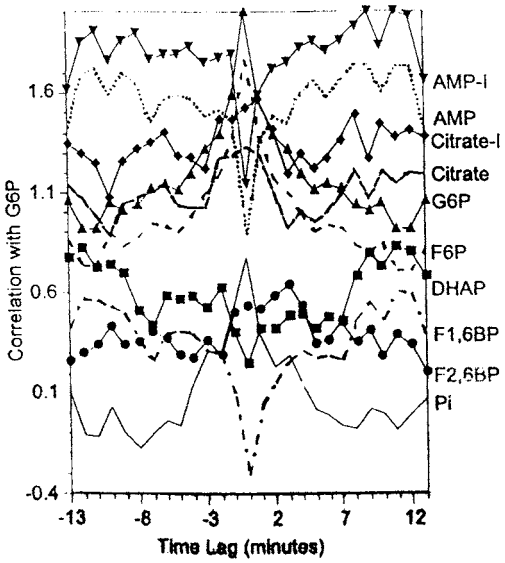

FIGURE 18. Graphs of the lagged correlation functions offset to distinguish individual correlation functions. Each successive correlation function, starting with F2,6BP, is offset by 0.2 unit on the correlation axis. For example, the maximum correlation ofG6P with AMP-AMP-11occurs at significantly positive lags, implying that variation in AMP-1 precedes variation in G6P.

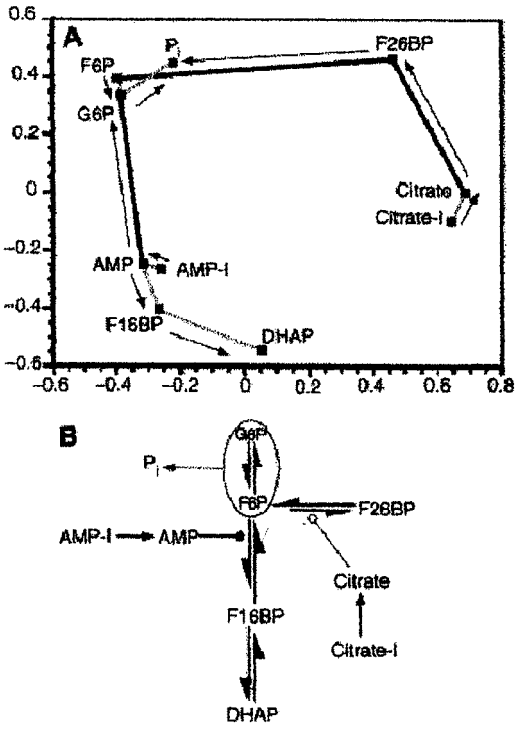

FIGURE 19. (A) 2D projection of the CMC diagram for the time series shown in Figure 17. Each point represents the time series of a given species. The closer two points are, the higher the correlation between the respective time series. The black (gray) lines indicate negative (positive) correlation between the respective species. Arrows indicate temporal ordering among species based on the lagged correlations between their time series. (B) Predicted reaction pathway derived from the CMC diagram. Its correspondence to the known mechanism is high.

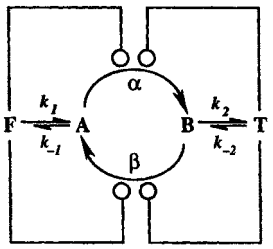

FIGURE 20. Diagram of the model. F and T are the reservoir species. A and B are the cycle intermediates, interconverted by enzymes α and β. Arrows indicate reactions, and knobs indicate regulation. The kinetic parameters are given in Figure 1 of ref 17.

approach. The applicability of the pulse method and the CMC approach is retained even if there are one or a few missing species. Many other details are given in ref 16.

IV. Applications of Genetic Algorithm Methods to Chemical Kinetics

A. Introduction. Genetic algorithms are one class of mathematical techniques for finding stated optimal goals or conditions in a given problem. The search for the chosen optimum conditions are started randomly but then directed toward the stated goals. The first tries, the first generation, are judged in fitness to the stated goals; the unfit are removed, the semi-fit are altered, and the fit are retained for the next tries, the next generation, and so on, until solutions are found that fulfill the stated goals adequately. Alterations are made by "mutations" in one or more parameters or variables, by crossover, and by other methods reminiscent of biological evolution due to genetic changes, hence the name genetic algorithms.

B. Selection of a Regulatory Structure for Flux Direction in a Simple Metabolic Model. Consider a schematic mechanism (Figure 20) for the study of the selection, by means of genetic algorithms, of a regulatory structure that directs flux in a simple metabolic model.[17] (For some references on prior studies of optimization of metabolic reaction networks, see refs 18 and 19.) F represents food (glucose), T an energetic molecule (ATP), the *k*'s are rate coefficients, A and B are intermediates, α and β are enzymes, and circles denote the effector action of F and T on the enzymes. We assign a task to this mechanism, that of controlling the proper direction of the flux from F to T (glycolysis), or the reverse (gluconeogenesis), as we externally vary the concentrations of F and T. If the concentration of F happens to be high and that of T low, then we want the system to direct mass flux quickly from F to T, and similarly for the reverse situation from T to F.

Let the rate equations for the temporal variation of A and B be

$$\dot{A} = k_1F + \nu_\beta - k_{-1}A - \nu_\alpha$$
$$\dot{B} = k_{-2}T + \nu_\alpha - k_2B - \nu_\beta \qquad (5)$$

For the effector control of F and T on the enzymes we take a model of noncompetitive allosteric binding[20] (Figure 21) with the rate (for enzyme α)

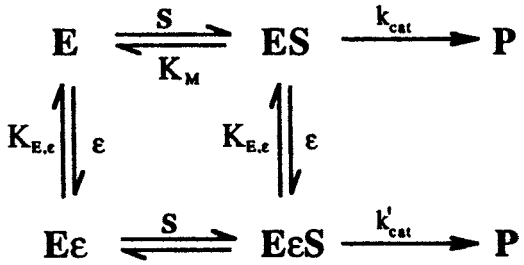

FIGURE 21. Mechanism of noncompetitive binding used to model the interaction between an enzyme (E) and an effector (ϵ). In this approximation, K_M is the dissociation constant of the enzyme–substrate complex. The binding interaction between substrate and enzyme is unperturbed by the presence of the effector. Formation of product from the enzyme–substrate complex with effector bound proceeds with the altered rate constant K'_{cat}.

$$\nu_\alpha = \frac{V_{\max,\alpha}A}{K_{M,\alpha} + A} R_{\alpha,F} R_{\alpha,T} \qquad (6)$$

where the factors modifying the intrinsic Michaelis–Menten rate expressions are

$$R_{\alpha,\epsilon} = \frac{K_{\alpha,\epsilon} + r_{\alpha,\epsilon}\epsilon}{K_{\alpha,\epsilon} + R_{\alpha,\epsilon}\epsilon} \qquad (7)$$

The parameter $K_{\alpha\epsilon}$ is the dissociation constant for the complex of enzyme α and the effector T or F, labeled ϵ, and $r_{\alpha\epsilon}$ is the ratio of the catalytic rate coefficient for the enzyme with and without effector bound, respectively. If *r* is greater than unity, then the effector is an activator, and if *r* is less than unity the effector is an inhibitor. We see in Figure 20 that there are four effector interactions, each with two parameters, one *K* and one *r*, for a total of eight parameters. Hence there are eight parameters that specify the regulatory response of the system for differing conditions of the reservoirs F and T. Thus it is possible for F or T to be selected a positive, or negative, or no effector on the enzymes α and β.

To specify the optimization, we define a need state for each reservoir which is positive if the concentration of that reservoir falls below a given value and negative if it exceeds another given value. A functional form for the need state is chosen such that there is an acceptable window of concentration around the target, within which the numerical value of the need is close to zero, but at the edges of which that value changes rapidly to a positive or negative one. With the defined need states we wish to optimize the function

$$f = \xi_T(k_2B - k_{-2}T) + \xi_F(k_{-1}A - k_1F) \qquad (8)$$

or its time average

$$J = \int_0^\tau f \, d\tau \qquad (9)$$

where ξ_F is the need state for F and ξ_T for T. The need states of F and T are multiplied respectively by the net flux of concentration into F and T. Hence if *f* is positive, then the mechanism is directing the net flux in accordance with the need states.

We have to begin with a chosen course of variation of the concentration of F and T. Next we start with a set of the eight parameters and integrate eqs 5−7 to evaluate J in eq 9. In the use of a genetic algorithm we wish to vary the eight parameters in a systematic way. Further details are given in the Appendix of ref 17.

Five different courses of variations of the concentrations of F and T were chosen to train 100 individuals (systems) for proper flux control. The results are interesting. In all but a few cases, the effect of F on the enzyme α is one of activation, and that of T on the same enzyme one of inhibition. For the enzyme β the reverse is true: T is an activator and F an inhibitor. There are exceptions, but then the networks perform not as well. This reciprocal effect on the opposing branches of the cycle is the regulatory pattern to be expected for efficient homeostasis: the mechanism seeks to control the fluxes so as to keep the concentrations of the reservoirs at the desired levels. This result arises solely from the optimization procedure of flux direction carried out by the genetic algorithm.

No single network was found that performs best on all the courses of changing environments. There is no single winner; the winners are survivors that perform adequately, but not necessarily outstandingly, from course to course. The absence of a single winner prevents global dominance and presents the opportunity for biological diversity.

C. Systematic Determination of a Reaction Mechanism and Rate Coefficients. For an application of genetic algorithms to this subject, see ref 20.

I thank all my co-workers cited in this review for their contributions and friendship.

References

(1) Berry, R. S.; Rice, S. A.; Ross. J. *Physical Chemistry*, 2nd ed.; Oxford University Press: Oxford, 2000; p 1064.
(2) Liang, S.; Fuhrmann, S.; Simogyi, R. Reveal, A General Reverse Engineering Algorithm For Inference of Genetic-Network Architectures. *Pac. Symp. Biocomput. '98* **1998**, *3*, 18−29.
(3) Akutsu, T.; Miyano, S.; Kuhara, S. Algorithms For Inferring Qualitative Models of Biological Networks. *Pac. Symp. Biocomput. '00* **2000**, *5*, 293−304.
(4) D'Haeseleer, P.; Liang, S.; Simogyi, R. Genetic Network Inference: From Co-Expression Clustering to Reverse Engineering. *Bioinformatics* **2000**, *16*, 707−726.
(5) Hjelmfelt, A.; Ross, J. Implementation of Logic Functions and Computations by Chemical Kinetics. *Phys. D* **1995**, *84*, 180−193 and references therein.
(6) Laplante, J.; Pemberton, M.; Hjelmfelt, A.; Ross, J. Experiments on Pattern Recognition by Chemical Kinetics. *J. Phys. Chem.* **1995**, *99*, 10063−10065.
(7) Arkin, A.; Ross, J. Computational Functions in Biochemical Reaction Networks. *Biophys. J.* **1994**, *67*, 560−578. For related experiments, see: Hauri, D. C.; Shen, P.; Arkin, A. P.; Ross, J. Steady State Measurements on the Fructose 6-Phosphate/Fructose 1,6-Bisphosphate Interconversion Cycle. *J. Phys. Chem. B* **1997**, *101*, 3872−3876.
(8) Vance, W.; Arkin, A.; Ross, J. Determination of Causal Connectivities of Species in Reaction Networks. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5816−5821.
(9) Torralba, A. S.; Yu, K.; Shen, P.; Oefner, P. J.; Ross, J. Experimental Test of a Method For Determining Causal Connectivities of Species in Reactions. *Proc. Natl. Acad. Sci.* **2003**, *100*, 1494−1498.
(10) Tyson, J. Classification of Instabilities in Chemical-Reaction Systems. *J. Chem. Phys.* **1975**, *62*, 1010−1015.
(11) Chevalier, T.; Schreiber, I.; Ross, J. Toward a Systematic Determination of Complex Reaction Mechanisms. *J. Phys. Chem.* **1993**, *97*, 6776−6787.
(12) Mihalik, E.; Skodt, H.; Hinne, F.; Sorenson, P. G.; Showalter, K. Normal Modes For Chemical Reactions From Time Series Analysis. *J. Phys. Chem. A* **1999**, *103*, 8246−8251.
(13) Kholodenko, B. N.; Klyatkin, A.; Bruggeman, F. J.; Sontag, E.; Westerhoff, H. V.; Hoek, J. B. Untangling The Wires: A Strategy to Trace Functional Interactions in Signaling and Gene Networks. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12841−12846.
(14) Arkin, A.; Ross, J. Statistical Construction of Chemical Reaction Mechanisms From Measured Time-Series. *J. Phys. Chem.* **1995**, *99*, 970−979.
(15) Samoilov, M.; Arkin, A.; Ross, J. Signal Processing by Simple Chemical Systems. *J. Phys. Chem.* **2002**, *106*, 10205−10221.
(16) Arkin, A.; Shen, P.; Ross, J. A Test Case of Correlation Metric Construction of a Reaction Pathway From Measurements. *Science* **1997**, *277*, 1275−1279.
(17) Gilman, A.; Ross, J. Genetic-Algorithm Selection of a Regulatory Structure That Directs Flux in a Simple Metabolic Model. *Biophys. J.* **1995**, *69*, 1321−1333.
(18) Bray, D.; Lay, S. Computer Simulated Evolution of a Network of Cell-Signaling Molecules. *Biophys. J.* **1994**, *66*, 972−977.
(19) Heinrich, R.; Hotzhutter, H. G. Efficiency and Design of Simple Metabolic Systems. *Biochim. Biophys. Acta* **1985**, *44*, 959−969.
(20) Fersht, A. *Enzyme Structure and Mechanism*, 2nd ed.; W. H. Freeman and Co.: New York, 1984; p 475.
(21) Tsuchiya, M.; Ross, J. Determination of Reaction Mechanism and Rate Coefficients for a Complex Reaction Network. *J. Phys. Chem. A* **2001**, *105*, 4052−4058.

AR020285F