CENTENNIAL FEATURE ARTICLE

# Determination of Complex Reaction Mechanisms. Analysis of Chemical, Biological and Genetic Networks†

**John Ross**

*Department of Chemistry, Stanford University, Palo Alto, California 94305-5080*

*Received: November 29, 2007; In Final Form: January 7, 2008*

We present several methods of determining, not guessing, complex chemical reaction mechanisms and their functions. One method is based on the theory of correlation functions of measured time series of concentrations of chemical species; another is on measurements of temporal responses of concentrations to various perturbations of arbitrary magnitude; a third deals with the analysis of oscillatory systems; a fourth is on the use of genetic algorithms to determine functions of chemical reaction networks. All methods are applicable to chemical, biochemical, and biological reaction systems and to genetic networks and systems biology. The methods depend on the design of appropriate experiments on the whole system and corresponding theories for interpretation that lead to information on the causal chemical connectivity of species, on reaction pathways, on reaction mechanisms, on control centers in the system, and on functions of the system. The first three methods require no assumption of a model or hypothesis, nor extensive calculations, unlike the interpretation of measurements made on a gene network at only one time.

## I. Introduction

The mechanism of a reaction system, at the macroscopic level of description, consists of the specification of the participating species, reactants, products, and catalysts and the connectivity of these species due to chemical reactions.[1] This connectivity can be expressed by a series of elementary reactions, the sum of which yields the reaction mechanism. The art and practice of guessing the mechanisms of reaction systems in the past, and still used to date, consists of (1) identifying individual chemical species, either by physical or by chemical means, (2) isolating the species contributing to one elementary reaction step in the 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 simultaneously the concentrations of more than a few chemical species as a function of time. The use of isotopic tracers, as for example, in metabolic flux analysis, has helped significantly. Following these steps began the guessing, hypothesizing, of a possible reaction mechanism, a sequence of elementary steps leading from reactants to products, with specifications of catalysts and their effectors. If the predictions of a guessed mechanism fit the available experiments, then that mechanism is possible, is sufficient; it is not necessary. (This holds for all hypotheses in science, from reaction mechanisms to the Schroedinger equation.)

Compare the determination of the long-practiced art of guessing reaction mechanisms as described above with the determination of the logic functions and components of an electronic device. An electronic engineer imposes electronic inputs (voltages, currents) on the device 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. The art of guessing a reaction mechanism is similar to taking a sledgehammer and knocking the device to pieces, looking for circuit elements such as transistors, capacitors, etc., and from that information attempting to guess the functions of the device.

The various methods we discuss here[2−5] for determining reaction mechanism have several common features that run parallel to the determination of logic functions: (1) keep the system intact; (2) perturb the system by changing the concentrations of one or more of the reactants (or other state variables) temporarily or permanently, or change the influx conditions into the (open) system; (3) determine the responses of the system following a perturbation by measuring the time dependence of as many concentrations as possible; (4) interpret these results with theories appropriate to the measurements to determine the chemical connectivity of the species, the reaction pathways, the reaction mechanisms, the network structure, and the functions of the system. These features also offer advantageous approaches to systems biology.[5]

## II. Computations in Chemical and Biochemical Reaction Systems

To follow the approach of the electronic engineer more closely, we needed to accomplish some prior goals. First we showed the possibility of constructing logic devices by means

† This year marks the Centennial of the American Chemical Society's Division of Physical Chemistry. To celebrate and to highlight the field of physical chemistry from both historical and future perspectives, *The Journal of Physical Chemistry* is publishing a special series of Centennial Feature Articles. These articles are invited contributions from current and former officers and members of the Physical Chemistry Division Executive Committee and from *J. Phys. Chem.* Senior Editors.

10.1021/jp711313e CCC: $40.75 
Published on Web 02/15/2008

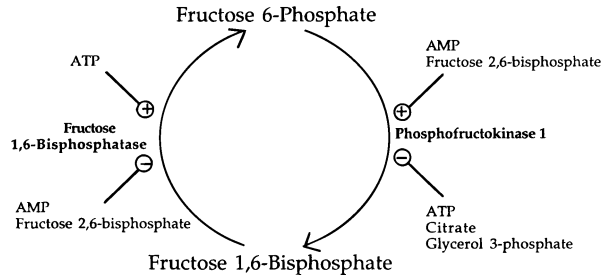



**Figure 1.** Schematic diagram of the fructose 6-phosphate/fructose 1,6-bisphosphate cycle. Some effectors of the enzymes are noted by lines ending in circles; + indicates activators, and − indicates inhibitors. Enzymes are noted in boldface type. Reproduced from ref 11.

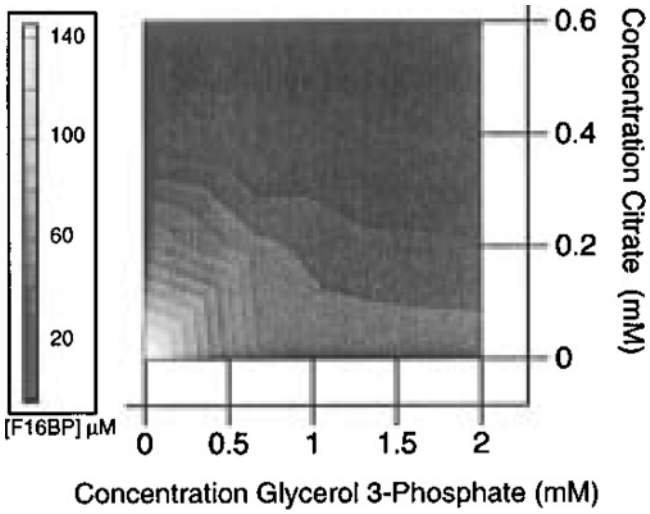



**Figure 2.** Experimentally determined plots of the stationary state concentration of F16BP vs the concentrations of effectors: citrate and G3P. Black is the lowest concentration of F16BP in each graph, and white is the highest (see the shaded quantitative scale next to the plot). Reproduced from 11.

of macroscopic kinetics.[6−9] For example, if and only if in a reaction system the concentrations of species 1 and 2 are high, then and only then is the concentration of species 3 high: in this case the mechanism acts as a logic AND gate. We also constructed on paper various other logic gates and with these designed a sequential computer called a universal Turing machine, such as a pocket computer and much larger machines. We then constructed a parallel computer, first on paper, and then we used a chemical bistable systems to carry out an experiment on pattern recognition, the first by implementation of a computation by means of macroscopic chemical kinetics.[10]

**John Ross** was born in Austria in 1926 and emigrated to the U.S. in 1940. He was educated at Queens College, New York, spent 2 years in the U.S. Army, and was granted a Ph.D. from MIT. He advanced through academic ranks at Brown University, was professor and chair in chemistry at MIT, and professor and chair at Stanford University. His research interests are in the subject of this article, in thermodynamics far from equilibrium, and in the kinetics of nonlinear systems.

Next we investigated the possibility that biochemical reactions can perform logic functions and calculations.[11] For example, the system in Figure 1 was studied experimentally as well as modeled theoretically, and one experimental result is shown in Figure 2.

The experimental results in Figure 2 correspond well to a calculation on this system. If the concentration of F16BP is small, then the gate is off. If, however, the concentration of both citrate and G3P (glycerol 3-phosphate) are small, then the gate, a NOR gate, is on and the concentration of F16BP is large. We can correlate the concentration of F16BP with the output of this system to be glycolysis (high concentration of F16BP) or gluconeogenesis (low concentration of F16BP). These results hold also for an extensive calculation on glycolysis that includes the system in Figure 1 and the tricarboxylic acid cycle.[3] Thus subsystems with various computational functions may be fitted together for carrying out more extensive calculations.

From these studies there emerges the possibility of a new approach to the problem of the construction and interpretation of reaction mechanisms by the development and application of analyses from electronic circuit theory, general systems theory, and multivariate statistic. We first discuss response methods, that is the responses of a reaction system to perturbations of several kinds.

## III. Response Methods

**A. Small Perturbations.**[12] Suppose we have a reaction mechanism with $n$ species; the deterministic kinetic evolution equations are

$$dX_i/dt = F_i(X_1,..,X_n) \qquad i = 1, ..., n \tag{1}$$

where $F_i$ is in general a nonlinear function of the concentrations $X_i$. For a stationary state we have

$$F_i(X^s) = 0 \qquad i = 1, ..., n \tag{2}$$

where the superscript "s" on $X$ denotes a stationary state. An experimental way of keeping the system in a stationary state far from equilibrium is to use a continuous flow stirred tank reactor (CSTR).

We may expand eq 1 in a Taylor series about a reference state, $X_\gamma$, which could be a stationary state, and obtain to first order

$$\frac{d\delta X_i}{dt} = \sum_{j=1}^{n} \left.\frac{\partial F_i}{\partial X_j}\right|_{X_\gamma} \delta X_j \qquad i = 1, ..., n \tag{3}$$

where the partial derivative is a Jacobian matrix element, $J_{ij}$. The sign of $J_{ij}$ determines whether the concentration of species $i$ increases, decreases, or is unaffected by adding a small amount of species $j$. The numerical values of the Jacobian matrix elements are determined by the reaction mechanism through its stoichiometry, the rate coefficients, and the concentrations. Thus measurements of these elements provide information on the reaction mechanism. Illustrations of this procedure are given in refs 12 and 13. The disadvantage of this approach lies with the need for small perturbations to be measured with sufficient accuracy.[14]

Relations among the matrix elements that need to be satisfied so that either positive or negative feedback loops in a mechanism produce a chemical instability are given in ref 15. Classification schemes for oscillating chemical systems, based on the interplay of positive and negative feedback loops have been suggested;[16] another method for such classifications is based on stoichiometric network analysis and is presented in refs 17−19.

The small perturbation approach has been extended from chemical species to biochemical modules[20] and applied to the mitogen-activated protein network.[21]

**B. Single Pulses of Arbitrary Magnitude.**[22] Consider the simple case of a sequence of first-order reactions

$$\xrightarrow{k_0} X_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} X_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} X_3 \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} X_4 \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} X_5 \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} X_6 \underset{k_{-6}}{\overset{k_6}{\rightleftharpoons}} X_7 \underset{k_{-7}}{\overset{k_7}{\rightleftharpoons}} X_8 \xrightarrow{k_8} \quad (4)$$

Let this isothermal system be in a stationary state, not at equilibrium, where the concentrations are constant but there is a steady flux, e.g., 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. As the concentration of one species is changed, other species respond with changes of concentrations. For a pulse in species $X_1$ the responses of the other species are shown in Figure 3.

A maximum occurs in the relative concentration of any species between the curves for the preceding and succeeding species, a result predicted by the solutions of the deterministic kinetic equations of system. Hence the order of the responses yields the causal (direct) connectivities of the species in the reaction mechanism. If the reaction in Figure 4 is perturbed at $X_4$, for example, then the pulse propagates down the chain, but more weakly up the chain against the free energy gradient down the chain.

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 it can help in determining correct stoichiometric coefficients. (7) If the reverse reaction rates are small (negligible) then the curve of the relative concentration of species $X_i$ intersects the curve of species $X_{i+1}$ at its maximum, making the identification of connectivities particularly simple.

An example of mixed first and second-order irreversible reactions is

$$\rightarrow X_1 \rightarrow X_2;\ 2X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow 2X_5;$$
$$X_5 \rightarrow X_6;\ 2X_6 \rightarrow X_7 \rightarrow X_8 \rightarrow \quad (5)$$

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 $u^*_i$,

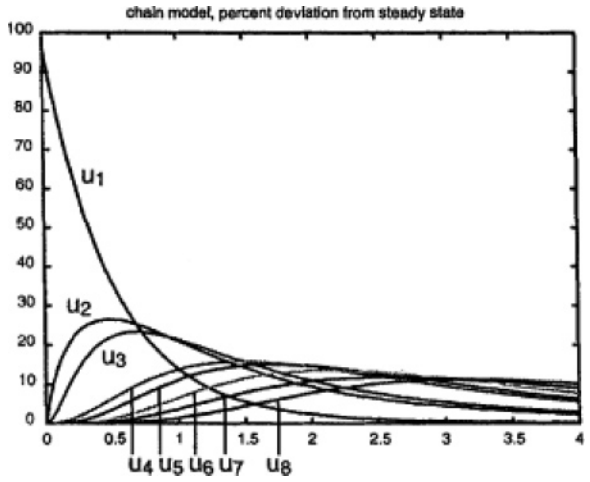



**Figure 3.** Plots of the relative deviation in concentration from the stationary state versus time for all the species of the mechanism in eq 4. The maxima are ordered according to the number of reaction steps separating that species from the initially perturbed species. Reproduced with permission from ref 22. Copyright 2002 American Institute of Physics.

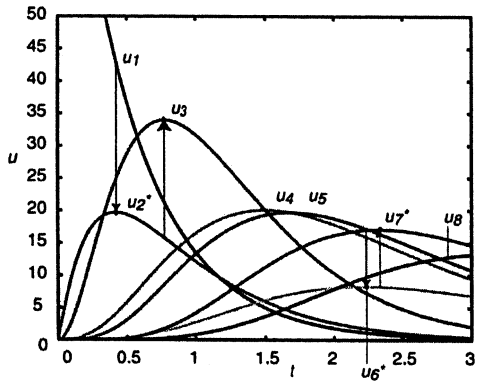
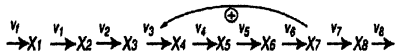



**Figure 4.** Plots of relative deviation in concentration versus time for species of the mechanism in eq 5. A pulse perturbation of the concentration of species $X_1$ results in the responses shown. The arrows indicate the approximate relations among concentrations shown in Figure 4. Reproduced with permission from ref 22. Copyright 2002 American Institute of Physics.

$$u_2^* \approx \frac{1}{2} u_1 \qquad u_3^* \approx 2u_2 \qquad u_6^* \approx \frac{1}{2} u_5 \qquad u_7^* \approx 2u_6 \quad (6)$$

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; in cycles; with feedforward and feedback reactions, and any combinations thereof. All can be analyzed with the pulse method. An illustration of feedback is given in the sequence




(7)

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.

The pulse method was tested[23] on the experimental system shown in the left part of Figure 5. The measured responses due

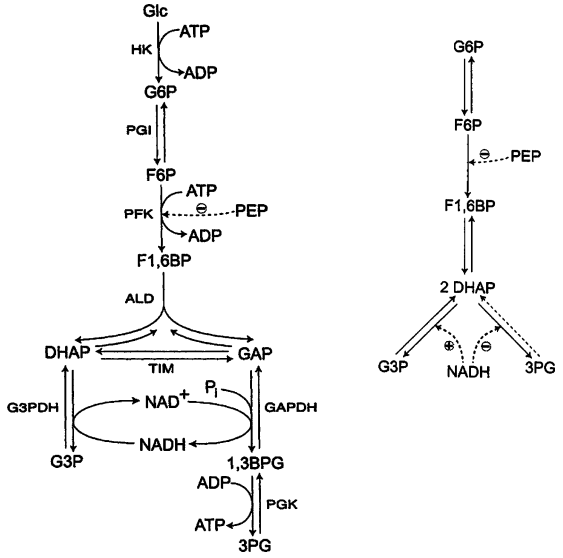



**Figure 5.** Left figure: experimental test system. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F1,6BP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone 3-phosphate; GAP, glyceraldehydes 3-phosphate; G3P, glycerol 3-phosphoglycerate. Reproduced with permission from ref 23. Copyright 2003 American Institute of Physics. Right figure: proposed reaction scheme based on pulse 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. Reproduced with permission from ref 23. Copyright 2003 American Institute of Physics.

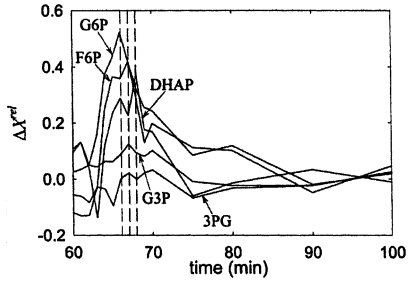



**Figure 6.** Responses in relative concentration to a pulse of G6P. Reproduced with permission from ref 23. Copyright 2003 American Institute of Physics.

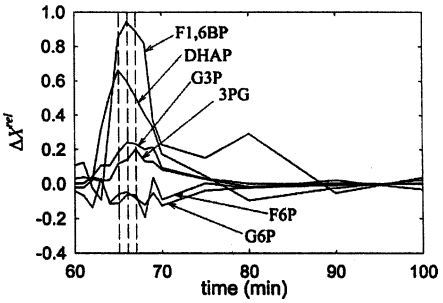



**Figure 7.** Responses in relative concentration to a pulse of DHAP. Reproduced with permission from ref 23. Copyright 2003 American Institute of Physics.

to a pulse in G6P give the temporal order of propagation of the pulse from G6P to F6P, then DHAP, G3P, and 3PG; see Figure 6.

The responses to a pulse of DHAP are shown in Figure 7. Note the much larger response of F16BP than that of DHAP due to the fact that essentially two molecules of DHAP are formed from one molecule of F16BP, because there is a rapidly established equilibrium between DHAP and GAP with the GAP concentration too small to be detected. Thus stoichiometric coefficients can be established from such experiments. From the responses to pulses of 6GP, F16BP, DHAP, and NADH the reaction sequence shown in the right part of Figure 5 was constructed and the main features of the experimental system are captured. The pulse method is relatively simple, effective, and generally applicable.

Other types of perturbations may be applied, such as sinusoidal variations of a concentration or other state variable, pulse sequences, etc.

**C. Neutral Response Methods.** If in a (radioactive) tracer experiment a pulse of tracer is introduced into the system such that the concentration of the sum of labeled and unlabeled species remains constant, and the kinetic isotope effects are small (negligible), then the response in the output of the system to such a pulse is linearly related to the input, whether the kinetics of the system are linear or not. The response and the input for a species u are related by a susceptibility, eq 8,

$$\beta_u(\mathbf{r},t) = \sum_{u'} \int_{-\infty}^{t} \int_{\mathbf{r}'} \chi_{uu'}(\mathbf{r}',t' \rightarrow \mathbf{r},t)\, \alpha_{u'}(\mathbf{r}',t')\, d\mathbf{r}'\, dt' \quad (8)$$

$$\alpha_{u'}(\mathbf{r}',t') = \text{input functions}$$

$$\beta u(\mathbf{r},t) = \text{output functions}$$

$$\chi_{uu'}(\mathbf{r}',t' \rightarrow \mathbf{r},t) = \text{susceptibility functions}$$

with the physical interpretation that this susceptibility equals the probability that a given species leaves the system at time $t$ and enters the system at $t'$, with a residence time of $\tau = t - t'$,

$$\varphi_{u'}(\tau,\Delta\mathbf{r}|\mathbf{r},u;t) = \chi_{uu'}(\mathbf{r}-\Delta\mathbf{r},t-\tau \rightarrow \mathbf{r},t) \quad (9)$$

The susceptibility matrix is proportional to a Green function matrix **G** which is the exponential of a connectivity matrix **K**

$$\chi = [\chi_{uu'}] = \chi(\tau) \sim \mathbf{G}(\tau) = \exp[\tau\mathbf{K}] \quad (10)$$

If an element ij of the connectivity matrix is zero then species i and j are not connected directly by a chemical reaction.

This simplification of neutral response experiments makes it possible to determine the connectivity of the species.[24,25]

## IV. Correlation Metric Construction

We turn next to correlation function methods for determining reaction mechanisms. We study the correlations of time series of reacting species and correlation metric construction (CMC), and their relations to the reaction mechanism of the system.[26] Causally connected species, by reactions, 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 (perturbations) of some chosen species.

**1. The CMC Method and a Test Calculation.** As an illustration consider a simple hypothetical reaction network, Figure 8, which is common in biochemical reactions. Let this thermodynamically 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 responses to these perturbations in the concentrations of all seven

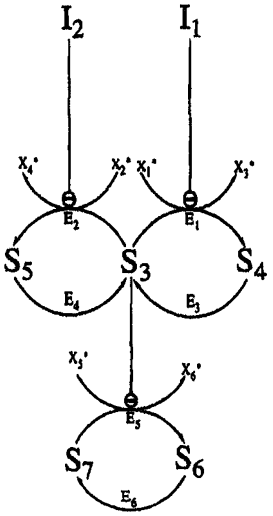



**Figure 8.** Chemical reaction mechanism representing a biochemical NAND gate. All species with asterisks are held constant by buffering. Lines ending in a circle-enclosed minus sign over an enzymatic reaction step indicate that the corresponding enzyme is inhibited (noncompetitively) by the relevant chemical species. Reprinted from ref 26.

species as a function of time (the enzymes $E_i$ are at constant concentrations). From these time series form correlations functions defined by, for species i and j,

$$S_{ij}(\tau) = \langle (x_i(t) - \bar{x}_i)(x_j(t) - \bar{x}_j) \rangle \tag{11}$$

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

$$r_{ij}(\tau) = \frac{S_{ij}(\tau)}{\sqrt{S_{ii}(\tau)\, S_{jj}(\tau)}} \tag{12}$$

define the maximum of that correlation for any $\tau$, and define a distance $d_{ij}$

$$d_{ij} = (c_{ii} - 2c_{ij} + c_{jj})^{1/2} = \sqrt{2}(1.0 - c_{ij})^{1/2}$$
$$c_{ij} = \max |r_{ij}(\tau)|_\tau \tag{13}$$

If the correlation $r_{ij}$ is large, e.g., the maximum of unity, then the distance $d_{ij}$ is zero; if there is no correlation, $r_{ij} = 0$, and 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.[26] 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. A multidimensional space is needed 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 on a computer, is done with the reaction mechanism in Figure 8, then we obtain the projection in Figure 9.

The projection of the multidimensional object constructed from the correlation distances (CMC) gives quite closely the reaction pathway shown in Figure 8. Note that no assumption

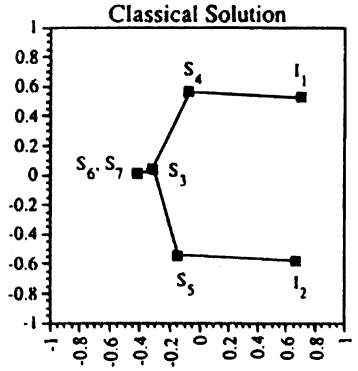



**Figure 9.** Projection of the multidimensional scaling analysis obtained from the correlation functions calculated from the time series of concentrations following perturbations of the system. The scales on the axes give correlation distances. Reprinted from ref 26.

of any model for this reaction system had to be made in this construction.

With seven species there are $(7 \times 6)/2$ binary correlations. We retained only the ones shown in Figure 9 by a procedure that ensures 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 S3-to-S4 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 to the other group (so that we have a two-time-scale reaction mechanism). In this case the correlation diagram analogous to Figure 4 shows a clear separation of the two groups, and hence the existence of two time scales. It also represents well the reaction pathway of each group.[26] Further, if there is a rate-limiting step in the reaction mechanism, then the correlation construction consists of only two points, one for all the reactions before that limiting step and one for all the reactions after that step.[26]

**2. An Experiment and Test of CMC.**[27] For the first experimental test of the correlation metric construction method, we chose a part of the much studied glycolysis system, shown in Figure 10. The system is established in a nonequilibrium stationary state with a constant inflow of glucose and buffer. Metabolites were measured with capillary electrophoresis; the concentrations of the enzymes were kept constant and so was the ATP/ADP ratio. 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 toward its nonequilibrium stationary state. The correlation functions themselves, eq 11, provide useful information. For example, the correlation of G6P with itself peaks at zero time lag $\tau$ and decays symmetrically with positive and negative $\tau$, 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 $\tau$, 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 and a sense of the temporal sequence in the reaction system.[27]

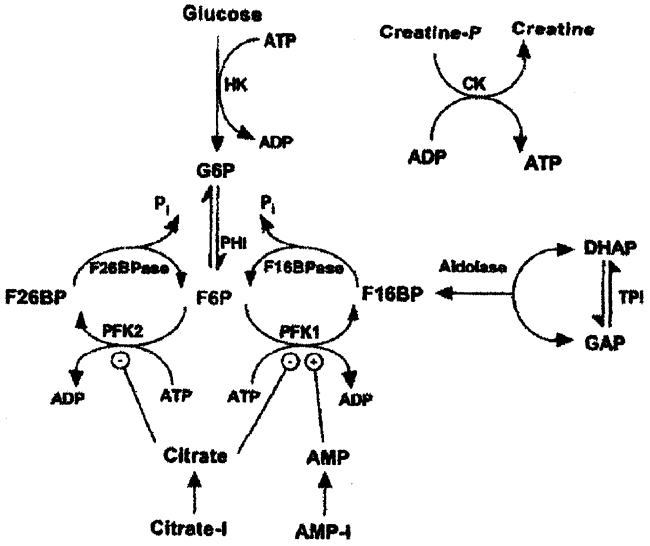



**Figure 10.** First few reaction steps of glycolysis. Regulatory interaction: (−) a negative effector, (+) a positive effector. Creatine-P (phosphate) and CK keep the concentrations of ATP and ADP constant. Abbreviations: Pi, inorganic phosphate; HK, hexokinase; PHI, phosphoinositol; F26Bpase, fructose 2,6-biphosphatase; TPI, triphosphoinositol; GAP, glutamate phosphate. Reprinted with permission from ref 27. Copyright 1997 AAAS.

From the experimentally determined correlations we constructed the multidimensional correlation metric diagram shown in Figure 11A.

Solid lines indicate negative correlations, dotted 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 11B the MDS diagram of Figure 11A is rearranged to show the usual reaction pathway determined over many years of effort. Much more detail is given in ref 27. The agreement with prior work is excellent and shows the viability and utility of the CMC. The applicability of the CMC approach is retained even if there are one or a few missing species.

For further theoretical development of the correlation function method based on an entropy metric construction see ref 28.

**3. Some Applications of Correlation Function Methods.** (a) In ref 29 a gene network of 259 genes from a photosynthetic cyanobacterium is perturbed with a flux of light and the responses of the genes, the gene transcription, is recorded on DNA microchips at 20 min intervals for a period of 8−16 h. The presence of time lags in the correlations indicates a cascade of biochemical reactions. The subdivision of the responses are obtained from the maximum of the correlations of the light pulse with the gene transcription: in group 1 that maximum occurred after 20 min, in group 2 after 40 min, etc. Within each time group, further subgroupings are obtained by an analysis of the profiles of the responses in the transcription. A listing of the genes is given for each group and subgroup. The authors state in the summary of this work: "Although substantial effort is required to plan and perform this type of experiment, an enormous amount of information is obtained. The directionality of the resulting networks provides more information than clustering alone and therefore allows the researcher to generate hypotheses based on the system structure. Additionally, it is important to consider similarly expressed genes as potential regulon members. Regulons are sets of coregulated genes with common promoter regions differing from operons in that they are not necessarily sequentially oriented in the genome. To this end, genes with the same time-lagged correlations may be considered as good regulon candidates .... This suggests that dynamic studies of transcriptional behavior with significant number of time points can play a key role in understanding cellular regulation."

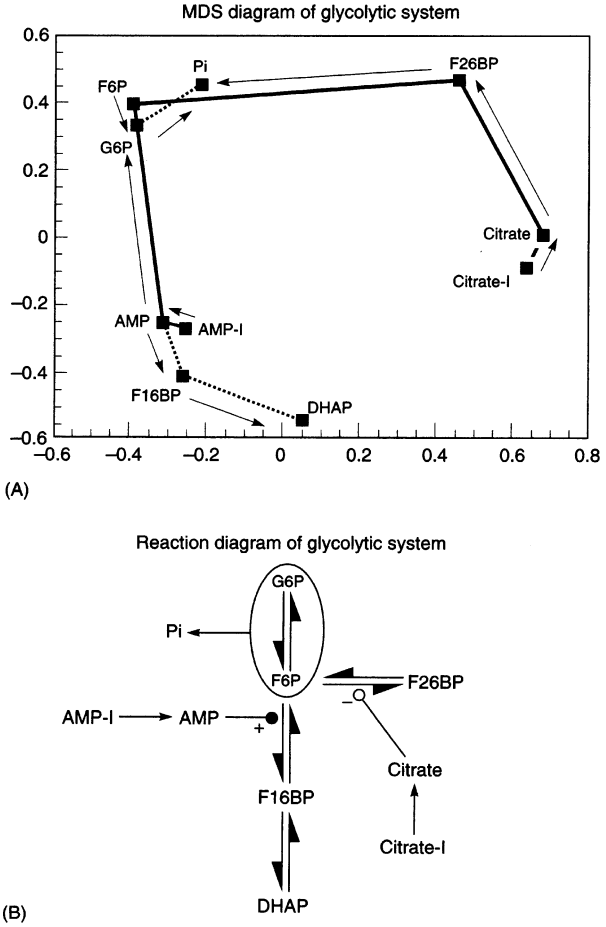



**Figure 11.** (A) 2D projection of the CMC diagram. Each point represents the time series of a given species. The closer two points are, the higher the correlation between the respective experimentally measured time series. The black (dotted) lines indicate negative (positive) correlation between the 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, Figure 10, is high for the species measured. Reprinted with permission from ref 27. Copyright 1997 AAAS.

(b) The problem of finding targets of the c-*myc* protooncogene is studied in ref 30. The data are a time series of gene expressions collected by microarray analysis. The correlations of gene expressions are used to construct a robust network that shows global dynamic properties following a cell state perturbation. The authors demonstrate that the correlation method establishes a clear relation between network structure and the cascade of c-*myc* activated genes.

(c) Applications of the correlation function method to parts of glycolysis are given in refs 31 and 32.

## V. Oscillatory Reactions

Many chemical, biochemical, biological, and some genetic network reactions are oscillatory when run far from equilibrium; that is, concentrations of some, not necessarily all chemical species, vary periodically in time. These variations may be, but need not be, sinusoidal. Biochemical examples include the horseradish peroxidase reaction and the beginning part of

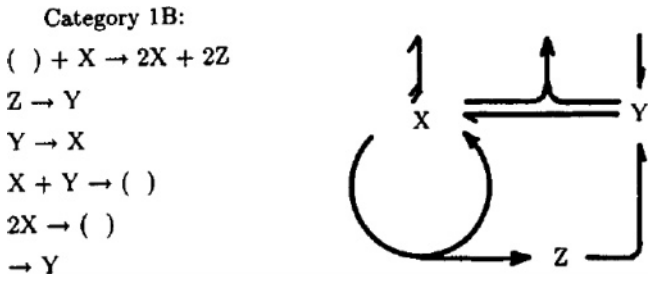



**Figure 12.** Basic reaction mechanism for category 1B in the classification of oscillatory reactions. The arrow notation, made clear by a comparison of the two parts of the figure, is defined in the caption of Figure 14. Reprinted with permission from ref 17. Copyright 1991 Wiley.

anaerobic glycolysis, in vivo and in vitro. In ref 33 there is an analysis of a periodic genetic network.

In 1991 we published studies of oscillatory reactions[17−19] that produced a mechanistic classification of such systems on the basis, in part, of stoichiometric net work analysis and a series of more than 10 experiments done on the whole system. (An extensive introduction to this subject is given in chapter XI of ref 3.) In oscillatory reactions there are essential species that upon holding one of these constant, all oscillations stop. The essential species fall into four categories, labeled X for autocatalytic species, Y for exit species, Z for negative feedback species, and W for recovery species. In terms of these essential species all known oscillatory reactions can classified into two categories, one with three subcategories, the other with two subcategories. For each subcategory there is a skeleton of elementary reaction steps, and we show one example of these in Figure 12 for category 1B.

The figure shows the essential species, their roles in the reaction mechanism, and their connectivity. There may be more than one species in the role of X, and so for the other essential species.

There are also nonessential species for which, if the concentration of one were held constant, the oscillations of the other species would not stop.

The experiments mentioned above allow the determination of essential and nonessential species. Just two examples must suffice. Comparison of relative amplitudes leads to identification of essential (large amplitudes of oscillations) and nonessential (small amplitudes) variables. The measurement of the phase shift of the oscillation of one essential species with respect to another leads to the roles of the essential species and the category of the oscillatory reaction, and hence the connectivity of the essential species. We use the qualitative designation (I) for in phase, (A) for antiphase, (+) for advanced, and (−) for delayed with respect to the oscillation of a reference species. For the category 1B the qualitative phase shifts are given in Figure 13 in which the phase shift of a species in the top line is given with respect to a species on the left. The more then 10 experiments discussed in ref 18 are sufficient to determine the reaction mechanism of a given oscillatory reaction. An analysis of the Belusov−Zhabotinsky reaction (ref 17 p 154 ff) identifies the essential and nonessential species, their connectivities, and the classification 1B; see Figure 14.

A study of the oscillatory peroxidase-oxidase reaction, including experiments and the formation of a reaction mechanism, by the methods described in this section is given in ref 18.

| | X | Y | Z |
|---|---|---|---|
| X | I | − | − |
| Y | + | I | + |
| Z | + | − | I |

**Figure 13.** Phase relations of the oscillations for the essential species in category 1B. We use the qualitative designations (I) for in phase, (A) for antiphase, (+) for advanced, and (−) for delayed with respect to the oscillation of a reference species. For example, the phase of the oscillation of X is advanced with respect to that of Y. Reprinted with permission from ref 18. Copyright 1995 Wiley.

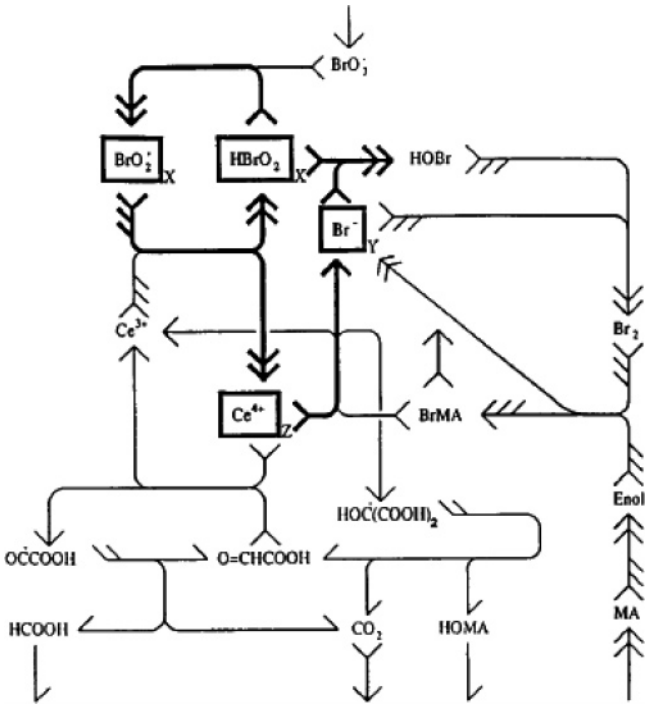



**Figure 14.** Belusov−Zhabotinsky reaction. The essential species and their reactions are shown in bold lines; the indices of the boxes denote the role of the essential species. The number of barbs (total number of feathers) on an arrow at a product (reactant) denotes the stoichiometric coefficient of this product (reactant); the number of left feathers denotes the kinetic exponent of the reactant. No feathers are shown when the stoichiometric and kinetic coefficient of a reactant are both unity. Vertical arrows indicate flows, not distinguished from reactions to (formation from) otherwise inert products (reactants). Reprinted with permission from ref 17. Copyright 1991 Wiley.

## VI. The Evolution of Functions of Reaction Mechanisms

Chemical and biochemical reactions can perform certain functions, such as computations, as discussed in section II, frequency filtering, signal transmission, etc. It is important to determine what functions a mechanism can perform, and how to evolve desirable properties in a given mechanism. We describe briefly two calculations on the evolution of a reaction mechanism to achieve a prescribed task. For this purpose we use genetic algorithms (GA), which is a mathematical optimization method.[34] Other optimization procedures may be used.

**A. Selection of Regulation of Flux in a Metabolic Model.[35]** We examine a simple biochemical reaction model in which certain functional parameters are deliberately left unspecified and are made the object of an optimization procedure. Our model cycle, shown on the left of Figure 15, consists of two "irreversible" enzymes that are regulated by signals from

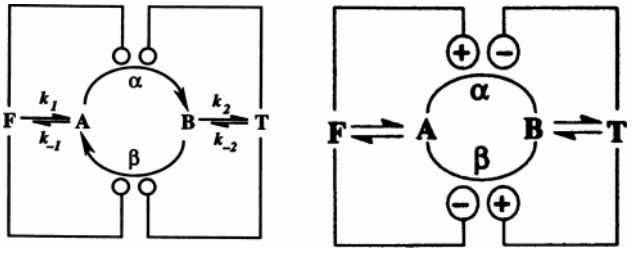



**Figure 15.** Left: diagram of the model. F and T are the reservoir species. A and B are the cycle intermediates, interconverted by enzymes α and *β*. Arrows indicate reactions, and knobs indicate regulation. Reprinted with permission from ref 35. Copyright 1995 Biophysical Society. Right figure: limiting network diagram, obtained from calculations to be described presently, that passed five courses (a course being a given set of changes in the reservoir concentrations). Note the reciprocal negative feedback asserted by F and T on the enzymes: ⊕ indicates increased catalytic activity, minus the reverse Reprinted with permission from ref 35. Copyright 1995 Biophysical Society.

external species (that is, species in the pathway but downstream or upstream from the cycle). We fix the reaction structure of the pathway, as indicated in the figure, and certain intrinsic kinetic parameters for the enzymes (the Michaelis−Menten parameters $K_m$ and $V_{max}$). We also specify that the regulation of the enzymes occurs only by modulation of their $V_{max}$ through noncompetitive binding of the external species. The parameters governing this modulation, however, are not specified but are left to be optimized. The criterion for optimization is that the regulated futile cycle be able to carry out a metabolic function that we specify: the net flux through the cycle is required to be sent in the direction given according to a simple model of biochemical "need". The cycle carrying out this task idealizes an animal cell that metabolizes blood glucose for energy (F in Figure 15) as long as the glucose concentration in the blood is adequate but synthesizes glucose for export if the glucose concentration in the blood drops too low.

There are four points of regulation with two parameters each, which need to be optimized to fulfill the function of flux control. We apply genetic algorithms to this problem. We choose a path of given external variations of F and T and vary these eight parameters, at first randomly. A given system must achieve a measure of controlling the flux from F to T, and its reverse, to maintain the stated requirements for F and T. If the system fails, then it is rejected from the pool of systems; if it passes, then the system goes on to the next generation with mutations, that is, changes of the eight parameters. The systems must pass five different paths (courses) of external variations of F and T. The results are interesting; one typical result is shown on the right in Figure 15.

We studied 25 systems on five courses, and all survivors developed over generations some negative reciprocal feedback. Most optimal individuals observed do respond rapidly to changes in both food supply and energy charge, fully reversing the direction of net flux in accordance to the need state. The regulatory pattern evident in these systems shows negative feedback and reciprocal effect on the opposing branches of the cycle. Although these regulatory motifs are fully consistent with intuitive expectations, the finding is significant in that it arises purely as a consequence of specifying the task to be performed. It thus serves as a numerical confirmation of the intuition.

The other findings are perhaps more surprising and appear in connection with the specific form of the time-varying external constraints. Optimization runs were performed on five different time courses, and strikingly, no global winner was found. Individuals found to be optimal on one or several of the courses proved not to be optimal on others. Indeed, we observe the appearance of "generalists", which perform well, if not optimally, on all of the course, and "specialists", which perform well on a single course but poorly on the others. The performance of generalists is generally accompanied by a higher expenditure of energy cost, although high cost does not appear to be sufficient for good performance. On some courses, the regulatory structures selected as optimal did not conform to intuitive expectations and in fact did not perform well relative to those selected on other courses. In this regard, some courses proved to be less stringent than others. This finding corresponds to the notion that under less stringent conditions even a suboptimal way of doing things is adequate. The most important point is survival, not energy expenditure or other criteria.

It is interesting to find results reminiscent of biodiversity and ecological evolutionary effects in a system as simple as this one. The analogy arises because of the fact that, although the GA is not intended to model the process of biological evolution, the two processes share certain features in common.

We believe this interesting example to be a demonstration of discovering by evolutionary methods the way reaction mechanisms can change to accomplish an assigned tasks.

**B. Development of Oscillations in Biochemical Systems for the Purpose of Increasing the Efficiency of the Usage of Nutrients.**[36] In unicellular and higher organisms, there exist oscillatory reactions, far from equilibrium (e.g., glycolysis under certain conditions including constant influx of glucose). External periodic perturbations of such reactions in concentration, temperature, pressure, light intensity, or imposed electric fields can phase shift the oscillatory rate compared to the oscillatory Gibbs free energy change, with consequent changes in the dissipation and conversely in the efficiency of the reaction, e.g., the conversion of glucose to ATP.[37,38] Similar phase shifts occur in alternating current (ac) networks, where the analog of the rate is the current, and that of the Gibbs free energy change is

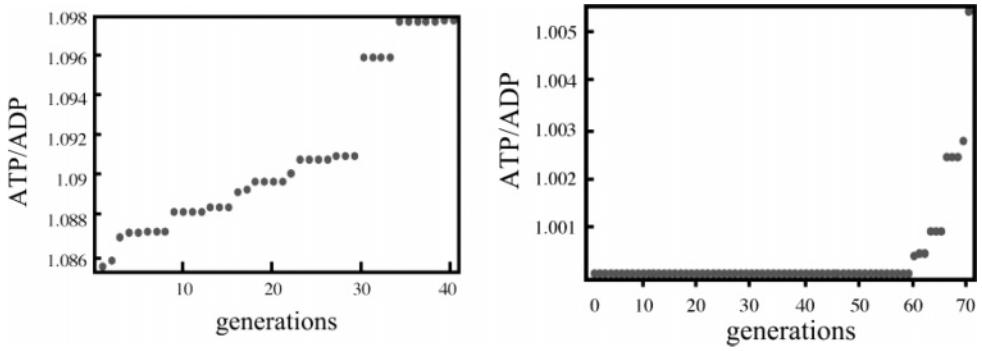
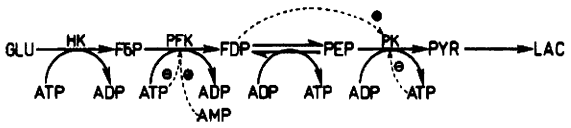



**Figure 16.** Left: plot of the ATP/ADP ratio in a given generation of the genetic algrithm calculation for the case of oscillatory input of glucose. Reproduced with permission from ref 36. Copyright 2003 American Institute of Physics. Right: same plot as on the left but here for a constant influx of glucose. Reproduced with permission from ref 36. Copyright 2003 American Institute of Physics.

proportional to the voltage. These effects of an "ac chemistry" have been shown in experiments on the oscillatory horseradish peroxidase reaction,[39,40] in combustion reactions,[41] in photosynthesis in a C3 plant,[42] and in theoretical studies in proton transfer.[43] In calculations on a model of glycolysis[37]



(14)

with a constant input of glucose, it was shown that the ratio of the concentrations ATP/ADP is substantially increased after a transition from steady-state kinetics to oscillatory kinetics, with that ratio averaged over a full oscillation.

A stimulating report[44] on the increased rate of growth of plants and animals exposed to ocean wave action, compared to that life in an estuary, led us to consider the possibility that oscillatory reactions in biological systems may have evolved in response to periodic perturbations of water waves impacting on a rocky shore. We investigated this conjecture[36] by considering a simple system, a part of the reaction mechanism of glycolysis shown in eq 8, and employed a genetic algorithm (GA) that is analogous to mutations in the gene pool.

Oscillation occur in the reaction mechanism in eq 14, for constant influx of glucose, due to the feedback that occurs in the PFK and the PK catalyzed reactions. If the binding constants of ATP and AMP to these enzymes are changed, then the oscillations cease. Consider now starting 2*n* systems in a nonoscillatory state with *n* systems having a constant influx of glucose (modeling the glucose input in a calm estuary) and *n* systems having an oscillatory influx of glucose (modeling the glucose input by wave action). Alter the binding constants in a genetic algorithm and see which mode of glucose input reaches oscillatory conditions of the autonomous system first and which therefore attains a higher ratio of ATP/ADP, and hence more efficient utilization of glucose. Indicative results of these calculations are shown in Figure 16.

Comparison of these two figures shows that for an oscillatory input of glucose the ATP/ADP ratio is increased nearly 10%, with the occurrence of oscillations in the autonomous system, after 34 generations of changes of the binding constants, whereas for constant input of glucose it takes 70 generations for a change of the ATP/ADP ratio of about 0.5%. The imposition of an oscillatory flux of glucose brings the system to oscillatory conditions much more rapidly and achieves more rapidly better utilization of the food supply.

The time scale of successive wave action in oceans, on the order of 20−40 s, may be of importance for the evolutionary development of oscillatory cell reactions in animals and plants.

## VII. Conclusions

The methods reviewed here of **determining** complex reaction mechanisms and their functions have been applied to chemical, biochemical, biological, and genetic systems, in some cases small systems and in some cases large systems. The methods emphasize the need to investigate the whole system, to keep intact all interactions. The suggested experiments require the measurements of concentrations over time and require the appropriate theories that allow the deduction of information on the reaction pathway, the reaction mechanism, the connectivity of the species, control features in the system, and functions of the system. No reaction mechanism has to be assumed and no prior hypotheses are necessary. Thus we believe that theses methods, going back to 1991, contribute to chemistry, biochemistry, genetic networks, biotechnology, and systems biology.

If experiments on a system are made at one time only, extensive as they may be (see, for example, the major contribution of Ishii et al.[45]), the information available directly appears limited, especially in view of the tremendous effort expended. Such experimental results require extensive calculational procedures and analysis, such as Baysian network analysis, cluster analysis, model formation, among others. These approaches have had successes, and problems, that require another review.

**Acknowledgment.** I thank my students and associates who contributed to the research reviewed here. Part of this work was supported by the National Science Foundation.

## References and Notes

(1) Berry, R. S.; Rice, S. A.; Ross, J. *Physical Chemistry*, 2nd ed.; Oxford University Press: Oxford, U.K., 2000.
(2) All the subjects presented here have been published in the literature, in a book, ref 3, and parts of it in two review articles, refs 4 and 5.
(3) Ross, J.; Schreiber, I.; Vlad, M. O. *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks*; Oxford University Press: New York, 2006.
(4) Ross, J. *Acc. Chem. Res.* **2003**, *36*, 839−847.
(5) Ross, J. To appear in *Annu. Rev. Biochem.*
(6) Ross, J.; Schreiber, I.; Vlad, M. O. *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks*; Oxford University Press: New York, 2006; pp 34−35.
(7) Hjelmfelt, A.; Weinberger, E. D.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 383−387.
(8) Hjelmfelt, A.; Schneider, F. W.; Ross, J. *Science* **1993**, *260*, 335−337.
(9) Hjelmfelt, A.; Ross, J. *Physica D* **1995**, *84*, 180−193.
(10) Laplante, J. P.; Pemberton, M.; Hjelmfelt, A.; Ross, J. *J. Phys. Chem.* **1995**, *99*, 10063−10065.
(11) Hauri, D. C.; Shen, P.; Arkin, A. P.; Ross, J. *J. Phys. Chem. B* **1997**, *101*, 3872−3876.
(12) Chevalier, T.; Schreiber, I.; Ross, J. *J. Phys. Chem.* **1993**, *97*, 6776−6787.
(13) Mihalik, E.; Skodt, H.; Hinne, F.; Sorenson, P. G.; Showalter, K. *J. Phys. Chem. A* **1999**, *103*, 8246−8251.
(14) Vlad, M. O.; Arkin, A.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7223−7228.
(15) Tyson, J. *J. Chem. Phys.* **1975**, *62*, 1010−1015.
(16) Luo, Y.; Epstein, I. R. *Adv. Chem. Phys.* **1990**, *79*, 269−299.
(17) Eiswirth, M.; Freund, A.; Ross, J. *Adv. Chem. Phys.* **1991**, *90*, 127−199.
(18) Stemwedel, J. D.; Schreiber, I.; Ross, J. *Adv. Chem. Phys.* **1995**, *89*, 327−388.
(19) Schreiber, I.; Ross, J. *J. Phys. Chem.* **2003**, *107*, 9846−9859.
(20) Kholodenko, B. N.; Kiyatkin, A.; Bruggeman, F. J.; Sontag, E.; Westerhoff, H. V.; Hoek, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12841−12846.
(21) Santos, S. D. M.; Verveer, P. J.; Bastiaens, P. I. H. *Nat. Cell Biol.* http://www.nature.com/ncb/journal/v9/n3/full/ncb1543.html (accessed 2007).
(22) Vance, W.; Arkin, A.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5816−5821.
(23) Sanchez-Torralba, A.; Yu, K.; Shen, P.; Oefner, P. J.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1494−1498.
(24) Ross, J.; Schreiber, I.; Vlad, M. O. *Determination of Complex Reaction Mechanisms: Analysis of Chemical, Biological, and Genetic Networks*; Oxford University Press: New York, 2006, pp 170−206.
(25) Vlad, M. O.; Moran, F.; Rodriguez, Y.; Ross, J. *Int. J. Bif. Chaos* **2002**, *12*, 2599−2618.
(26) Arkin, A.; Ross, J. *J. Phys. Chem.* **1995**, *99*, 970−979.
(27) Arkin, A.; Shen, P.; Ross, J. *Science* **1997**, *277*, 1275−1279.
(28) Samoilov, M.; Arkin, A.; Ross, J. *Chaos* **2001**, *11*, 108−114.
(29) Schmitt, W. M. Jr.; Raab, R. M.; Stephanopoulos, G. *Genome Res.* http://www.genome.org/cgi/doi/10.1101/gr.2439804 (accessed 2004).
(30) Remondini, D.; O'Connell, B.; Intrator, N.; Sedivy, J. M.; Neretti, N.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6902−6906.
(31) Steuer, R.; Kurths, J.; Fiehn, O.; Weckwerth, W. *Bioinformatics* **2003**, *19*, 1019−1026.
(32) Kaufmann, K. J.; Pajerowski, J. D.; Jamshidi, N.; Paulson, B. O.; Edwards, J. S. *Biophys. J.* **2002**, *83*, 646−662.
(33) Pigolotti, S.; Krishna, S.; Jensen, M. H. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6533−6537.
(34) Mitchell, M. *An Introduction to Genetic Algorithms*; MIT Press: Cambridge, U.K., 1996.

(35) Gilman, A.; Ross, J. *Biophys. J.* **1995**, *69*, 1321−1333.
(36) Tsuchiya, M.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9691−9695.
(37) Termonia, Y.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 2952−2956.
(38) Termonia, Y.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2878−2881.
(39) Lazar, J. G.; Ross, J. *Science* **1990**, *247*, 189−192.
(40) Lazar, J. G.; Ross, J. *J. Chem. Phys.* **1990**, *92*, 3579−3589.
(41) Schram, H.; Schell, M.; Ross, J. *J. Chem. Phys.* **1988**, *88*, 2730−2734.
(42) Kocks, P.; Ross, J.; Bjorkman, O. *J. Phys. Chem.* **1995**, *99*, 16483−16489.
(43) Schell, M.; Kundu, K.; Ross, J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 424−428.
(44) Leigh, E. G., Jr.; Paine, R. T.; Quinn, J. F.; Suchanek, T. H. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 1314−1318.
(45) Ishii, N.; Nakahigashi, K.; Baba, T.; Robert, M.; Soga, T.; et al. *Science* **2007**, *316*, 593−597.