Biophysical Chemistry, 35 (1990) 1–24
Elsevier

BIOCHE 01401

Review

# Applications of graph theory to enzyme kinetics and protein folding kinetics

## Steady and non-steady-state systems

Kuo-Chen Chou
*Computational Chemistry, Upjohn Research Laboratories, Kalamazoo, MI 49001, U.S.A.*

Received 27 June 1989
Revised manuscript received 11 September 1989
Accepted 12 September 1989

Directed graph; Spanning in-tree; Sink; Cycle; Loop; Path; Product-creating cycle; Product-annihilating cycle; Phase concentration; Transient concentration

Graphic methods have proved to be very useful in enzyme kinetics, as reflected in both raising the efficiency of performing calculations and aiding in the analysis of catalytic mechanisms. The kinetic relations among protein folding states are very similar to those between enzyme-catalyzed species. Therefore, it should be equally useful to provide a visually intuitive relation between kinetic calculations and folding mechanisms for protein folding kinetics, as manifested by the graphic rules in enzyme kinetics. It can actually be anticipated that, due to increasing interest in protein folding, the graphic method will become an important tool in folding kinetics as well. Based on the recent progress made in graphic methods of enzyme kinetics, in this review four graphic rules are summarized, which can be used to deal with protein folding systems as well as enzyme-catalyzed systems. Rules 1–3 are established for deriving the kinetic equations for steady-state processes and Rule 4 for those in the case of non-steady-state processes. In comparison with conventional graphic methods, which can only be applied to a steady-state system, the current rules have the following advantages: (1) Complicated and tedious calculations can be greatly simplified. (2) A lot of wasted labor can be turned away. (3) Final results can be double-checked by a formula provided in each of the graphic rules. (4) Transient kinetic systems can also be treated. The mathematical proof of Rules 1–4 is given in appendices A–D, respectively.

## 1. Introduction

33 years ago King and Altman [1] proposed a graphic method for deriving the steady-state rate equations in enzyme kinetics. Their method can make the calculation more convenient and intuitive, and hence has proved to be very useful in enzymology. However, the King-Altman method can only be used to deal with very simple, steady-state enzyme-catalyzed systems. For the case of an enzyme-catalyzed reaction system that is a little more complicated, the task of performing such calculations would become quite formidable without the aid of a computer. Unfortunately, by means of a computer, we can in general only obtain numerical results, and not an analytic solution which, however, is crucially important for the analysis of the mechanisms (cf., e.g., refs. 2–5). Therefore, many attempts [6–24] have been made since then in endeavoring to improve and to develop the original graphic method. To develop the graphic method more effectively and rigorously even for the steady-state system only, it is necessary to introduce some knowledge of graphic theory, which, however, will present a barrier to most readers, whose interest

Correspondence address: K.-C. Chou, Computational Chemistry, Upjohn Research Laboratories, Kalamazoo, MI 49001, U.S.A.

lies solely in the application of graphic methods. As a consequence, most biologists and chemists still prefer to adopt the original King-Altman method although the derivation would thus be much more time-consuming and error-prone. Moreover, the developments in the past 30 years cover a rather wide range of areas of interest, which presents additional difficulty to users. Furthermore, on going more deeply into the principles of enzyme action as well as protein folding, one is often confronted with transient processes, which cannot be treated in terms of the conventional graphic methods. Therefore, it is highly desirable to present a systematic and readable description of the contemporary graphic methods that can be used to deal with both steady-state and non-steady-state systems. To satisfy this kind of request from many users, in this review the recent progress of graphic methods has been systematized into four rules. Rules 1–3 are established for improving the conventional calculation methods for the kinetic system of steady-state processes, and Rule 4 is set up in order to extend the application of graphic methods to non-steady-state systems which are beyond the scope of conventional graphic methods. The norms of introducing each of these rules are: (1) The application should be as simple and intuitive as possible. (2) The description should be as readable as possible; i.e., the special and uncommon terminologies in graphic theory should be used as little as possible except for those whose meanings are self-evident. (3) Each rule should be followed by some illustrations demonstrating its advantages as well as implications. (4) The establishment of each rule should be supported by a solid mathematical proof. In order to meet all of the above four norms, the present review is divided into two parts: the text and appendices. In the text, novel and powerful rules are written and demonstrated in a way that is easy to follow for most chemists and biologists. In the appendices, however, each of the four rules will be rigorously proved in order to cater for readers who wish to peruse the details of the mathematical principles for these novel rules.

At the outset, let us consider an enzyme-catalyzed system in which there are $n$ different enzyme-containing species $E_i$ $(i = 1, 2, \ldots, n)$. Suppose that the rate constant for $E_i \to E_j$ is $k_{ij}$, and that for $E_j \to E_i$ is $k_{ji}$, etc., as schematized in fig. 1. Here, $k_{ij}$ represents the first-order rate constant or the pseudo first-order rate constant [1,3,16]. According to definition, we have $k_{ij} = 0$ when $i = j$. For a reaction system of this type, the following two questions are often asked:

$$e_m \equiv [E_m] = ?\ (m = 1, 2, \ldots, n) \tag{1}$$

$$v = \frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} = \frac{\mathrm{d}p}{\mathrm{d}t} = \sum_{\alpha \neq \beta} \left\{ k_{\alpha\beta}[\mathrm{E}_\alpha] - k_{\beta\alpha}[\mathrm{E}_\beta] \right\} = \sum_{\alpha \neq \beta} \left\{ k_{\alpha\beta} e_\alpha - k_{\beta\alpha} e_\beta \right\} = ? \tag{2}$$

where $e_m \equiv [\mathrm{E}_m]$ is the concentration of the $m$-th enzyme species, $\mathrm{E}_m$. In this article, the concentration of a reactant X is represented by either [X] or just the corresponding lower-case letter $x$, $v$ denotes the rate of formation of the product P, and $\mathrm{E}_\alpha$ and $\mathrm{E}_\beta$ are the enzyme-containing species that directly involve the releasing and binding of the product P, respectively.

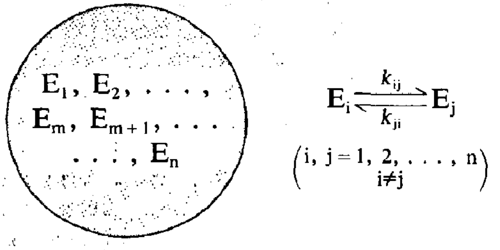


Fig. 1. An enzyme-catalyzed reaction system in which there are $n$ different enzyme-containing species $\mathrm{E}_i$ $(i = 1, 2, \ldots, n)$. The conversion rate constant from the $i$-th enzyme species $\mathrm{E}_i$ to the $j$-th enzyme species $\mathrm{E}_j$ is $k_{ij}$, and that from the $j$-th enzyme species $\mathrm{E}_j$ to the $i$-th enzyme species $\mathrm{E}_i$ is $k_{ji}$, where $k$ represents the first-order rate constant or pseudo first-order rate constant [1,3,16].

Obviously, for a protein folding system in which there are $n$ different folding states, all of the above descriptions remain valid as well. The following equations will therefore not explicitly refer to the protein folding system, since the current formulations will always be valid for either of these two kinetic systems.

## 2. Steady-state kinetic systems

For steady-state systems, in order to determine the solutions to the questions raised in eqs. 1–2, one usually has to solve the following equations:

$$0 = e_i \sum_{j=1}^{n} k_{ij} - \sum_{j=1}^{n} k_{ji} e_j \quad (i = 1, 2, \ldots, n) \tag{3}$$

$$\sum_{i=1}^{n} e_i = e_0 \tag{4}$$

where $e_0$ is the total concentration of all the enzyme species. Note that in eq. 3 only $n - 1$ equations are independent. To obtain the unique solution, we can use eq. 4 to replace any one of the $n$ equations in eq. 3. Thus, according to Cramer's rule, the solutions of eqs. 3–4 can be expressed as

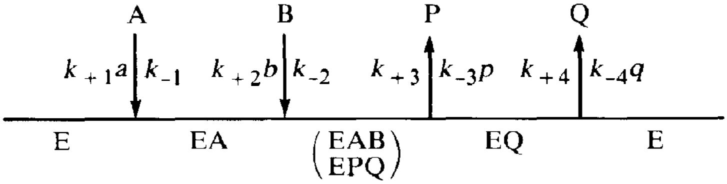
$$e_m = \frac{N_m}{\Sigma_{i=1}^{n} N_i} e_0 \tag{5}$$

where

$$\begin{array}{c|ccccccc|l}
 & & & & \begin{pmatrix}\text{the } m\text{-th}\\ \text{column}\end{pmatrix} & & & & \\
 & & & & \downarrow & & & & \\
 & \Sigma_{j=1}^{n} k_{1j} & -k_{21} & \cdots & 0 & \cdots & -k_{n1} & & \\
 & -k_{12} & \Sigma_{j=1}^{n} k_{2j} & \cdots & 0 & \cdots & -k_{n2} & & \\
 & \vdots & \vdots & & \vdots & & \vdots & & \\
N_m = & 1 & 1 & \cdots & 1 & \cdots & 1 & & \leftarrow \begin{pmatrix}\text{the } m\text{-th}\\ \text{row}\end{pmatrix} \\
 & \vdots & \vdots & & \vdots & & \vdots & & \\
 & -k_{1n} & -k_{2n} & \cdots & 0 & \cdots & \Sigma_{j=1}^{n} k_{nj} & &
\end{array} \tag{6}$$

Therefore, once the determinant in eq. 6 has been expanded, all the answers for the above questions can be easily obtained merely by substituting the result into eq. 5 as well as eqs. 1–2. Unfortunately, it is an extremely tedious and error-prone procedure to expand the determinant in eq. 6 except when the case considered is very simple.

To deal with this difficulty, King and Altman [1] proposed a graphic method, which can be illustrated by considering the following bi-bi ordered reaction mechanism [25]:



(7)

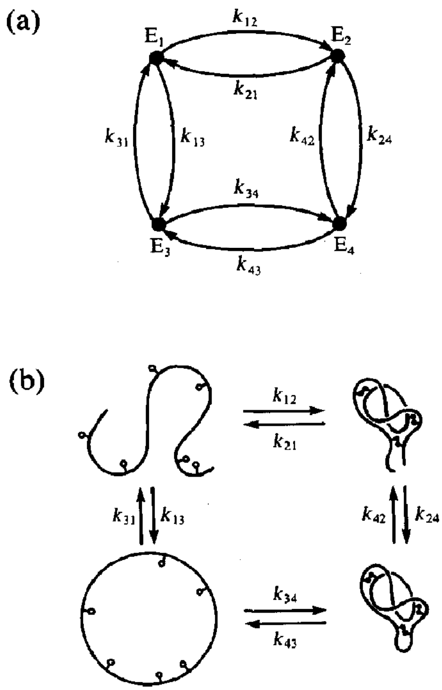
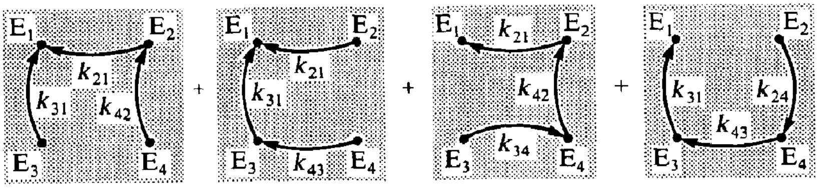


Fig. 2. (a) The directed graph $G$ used to express the *bi-bi ordered* reaction mechanism [25] as formulated in eq. 7. (b) A hypothetical thermodynamic cycle made up of the unfolded and native conformations of BPTI with and without the peptide bond between termini (adapted from ref. 26).

Assigning $E_1 = E$, $E_2 = EA$, $E_3 = EQ$, $E_4 = \binom{EAB}{EPQ}$, and $k_{12} = k_{+1}a$, $k_{21} = k_{-1}$, $k_{13} = k_{-4}q$, $k_{31} = k_{+4}$, $k_{24} = k_{+2}b$, $k_{42} = k_{-2}$, $k_{34} = k_{-3}p$, $k_{43} = k_{+3}$, then the bi-bi ordered mechanism can be described by a *directed graph*, or the abbreviation, *digraph G*, as given in fig. 2a. The directed graph $G$ consists of points and arcs. In such a graph, various enzyme species are represented by different points, and the interconversion between any two enzyme species by an arc with an arrow and weighted by a rate constant to indicate the conversion direction and rate, respectively. If $k_{ij} = 0$, then the arc from $E_i$ to $E_j$ is not depicted, implying no direct conversion from the $i$-th enzyme species to the $j$-th one. King and Altman [1] suggested that, instead of expanding the determinant of eq. 6, one can derive $N_m$ directly from the directed graph; e.g., for the enzyme-catalyzed reaction system as described by fig. 2a, it follows that according to the graphic method suggested by King and Altman [1]:

$N_1 =$



$$= k_{31}k_{42}k_{21} + k_{21}k_{43}k_{31} + k_{34}k_{42}k_{21} + k_{24}k_{43}k_{31} \tag{8}$$

According to the graphic theory, each term in eq. 8 actually corresponds to a *spanning in-tree* of fig. 2a with $E_1$ as a *sink*, and its value is the product of all the weights in each such subgraph. Accordingly,

$$\begin{aligned} N_2 &= \{\text{all spanning in-trees of fig. 2a with } E_2 \text{ as a sink}\} \\ &= k_{42}k_{31}k_{12} + k_{12}k_{34}k_{42} + k_{13}k_{34}k_{42} + k_{43}k_{31}k_{12} \end{aligned} \tag{9}$$

$$\begin{aligned} N_3 &= \{\text{all spanning in-trees of fig. 2a with } E_3 \text{ as a sink}\} \\ &= k_{13}k_{24}k_{43} + k_{43}k_{21}k_{13} + k_{42}k_{21}k_{13} + k_{12}k_{24}k_{43} \end{aligned} \tag{10}$$

$$\begin{aligned} N_4 &= \{\text{all spanning in-trees of fig. 2a with } E_4 \text{ as a sink}\} \\ &= k_{34}k_{12}k_{24} + k_{24}k_{13}k_{34} + k_{31}k_{12}k_{24} + k_{21}k_{13}k_{34} \end{aligned} \tag{11}$$

As we can see from the above, King-Altman's method provides a one-to-one relation between the calculated terms and the subgraphs, which undoubtedly is an advantage from the viewpoints of both calculation and analysis. It is also apparent that the directed graph $G$ can be used to represent a protein folding kinetic mechanism as well, in which case different points in the graph instead represent various folding states. For example, fig. 2a is also the directed graph of the folding mechanism as shown in fig. 2b. Therefore, after such a mathematical abstraction, the kinetic mechanism as shown in fig. 2b for a protein folding system is actually the same as that given by eq. 7 for an enzyme-catalyzed system, at least from the point of view of graph theory. However, when an enzyme-catalyzed reaction system or a protein folding mechanism becomes a little more complicated, the number of the spanning in-trees, i.e., the subgraphs to be counted, will increase rapidly. Thus, the following three problems emerge:

1. Some of the subgraphs are easily missed, leading to a loss of the corresponding terms. To avoid such an error, can we determine a reliable method by which to predict the number of subgraphs and hence the number of total terms in $N_m$ beforehand?
2. Even if the number of subgraphs to be counted is known, it is still very difficult to carry out the calculation because such a number might be very large even for a moderately complicated system. To solve this problem, can we find a new graphic method by which the required subgraphs can be greatly reduced?
3. When calculating the rate of product formation as formulated by eq. 2, many reciprocally cancelled terms between $k_{\beta\alpha}e_\beta$ and $k_{\alpha\beta}e_\alpha$ will appear explicitly and hence one must determine all of them pair by pair followed by cancelling them out against each other. To avoid this kind of cancellation operation that is both wasted labor and error-prone, can we derive a new approach by which all those useless cancelled terms will automatically disappear from the beginning?

Below, we shall present three rules, which can be used to deal with the difficulties as mentioned in the above three aspects, respectively. The rules given below are valid irrespective of whether the kinetic system concerned is with or without parallel reaction pathways [23].

*Rule 1: The method for calculating the number of terms in $N_m$*

(1) According to the directed graph $G$, construct a matrix $A = [a_{ij}]$, where

$$a_{ij} = \{\text{the number of arcs from } E_i \text{ to } E_j \text{ in } G\} \tag{12}$$

(2) Based on matrix $A$, the matrix $B = [b_{ij}]$ is built, in which

$$b_{ij} = \begin{cases} \sum_{j'=1}^{n} a_{ij'} & \text{if } i = j \\ -a_{ij} & \text{if } i \neq j \end{cases} \tag{13}$$

(3) The number of total terms in $N_m$ is then given by

$$P_m = \det B_{m,m} \tag{14}$$

where $B_{m,m}$ denotes the submatrix obtained by removing the $m$-th row and $m$-th column from the matrix $B$, and 'det' denotes taking the determinant value for the matrix next to it.

(4) If the directed graph $G$ is a reversible one, i.e., one in which the number of arcs from $E_i$ to $E_j$ is the same as that from $E_j$ to $E_i$, then it is sufficient merely to calculate any one of the $P_m$ $(m = 1, 2, \ldots, n)$ because all of them must be the same under such a circumstance.

The mathematical proof of the above rule is given in appendix A.

*Example 1*

For the kinetic system as shown in fig. 2a, according to Rule 1 we have

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix} \tag{15}$$

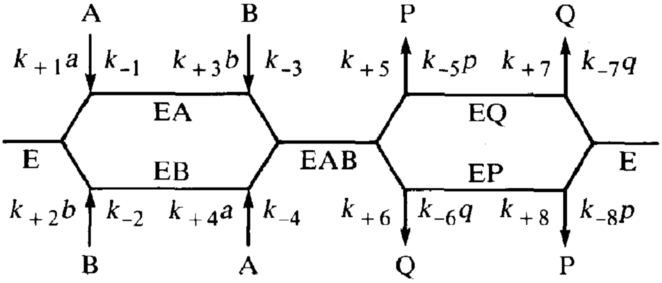
Since the directed graph in fig. 2a is a reversible one, it follows that

$$P_1 = P_2 = P_3 = P_4 = \det B_{4,4} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} = 4 \tag{16}$$

which means that there are four spanning in-trees in the directed graph $G$ for each of the $E_m$ $(m = 1, 2, 3, 4)$ as a sink, or alternatively, each of $N_m$ $(m = 1, 2, 3, 4)$ contains four terms (cf. eqs. 8–11).

*Example 2*

Consider the following bi-bi random mechanism [25]:



(17)

Let $E_1 = E$, $E_2 = EA$, $E_3 = EB$, $E_4 = EAB$, $E_5 = EQ$, $E_6 = EP$, and $k_{12} = k_{+1}a$, $k_{21} = k_{-1}$, $k_{13} = k_{+2}b$, $k_{31} = k_{-2}, \ldots$, etc., then such an enzyme-catalyzed reaction can be expressed by the directed graph $G$ as shown in fig. 3a. Thus according to eqs. 12–13, we have

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & -1 & -1 & 0 & -1 & -1 \\ -1 & 2 & 0 & -1 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 \\ 0 & -1 & -1 & 4 & -1 & -1 \\ -1 & 0 & 0 & -1 & 2 & 0 \\ -1 & 0 & 0 & -1 & 0 & 2 \end{bmatrix} \tag{18}$$

Since the directed graph in fig. 3a is a reversible one, it follows in terms of eq. 14 that

$$P_2 = P_3 = P_4 = P_5 = P_6 = P_1 = \det B_{1,1} = \begin{bmatrix} 2 & 0 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 & 0 \\ -1 & -1 & 4 & -1 & -1 \\ 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 0 & 2 \end{bmatrix} = 32 \tag{19}$$

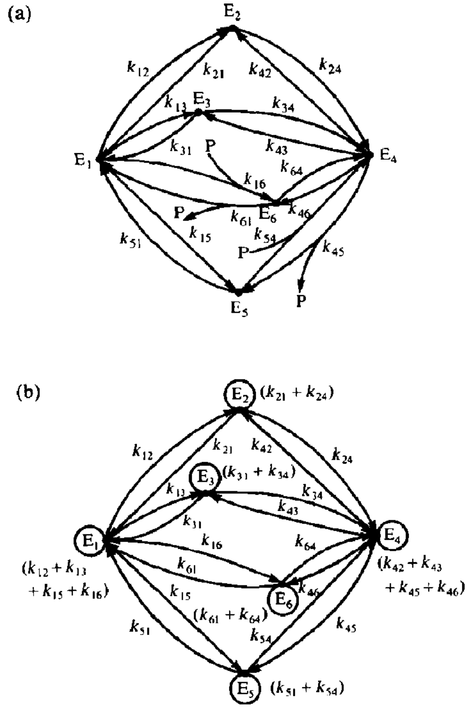


Fig. 3. (a) The directed graph $G$ used to express the *bi-bi random* reaction mechanism as formulated in eq. 17. P is included explicitly as an aid for calculating the rate of product formation (cf. Rule 3). (b) The transformed graph $D^\dagger$ obtained from panel a following the procedure described in Rule 2(1).

which also means that, if using the King-Altman method, one has to find 32 subgraphs (in-trees) in order to calculate each of $N_m$ $(m = 1, 2, \ldots, 6)$.

*Example 3*

For the Botts-Morales mechanism [27] as schematized in fig. 4a, assigning $E_1 = E$, $E_2 = ES$, $E_3 = EIS$ and $E_4 = EI$, as well as $k_{12} = k_{+1}[\mathrm{S}]$, $k_{21} = k_{-1}$, $k^*_{12} = k^*_{+1}[\mathrm{P}]$, $k^*_{21} = k^*_{-1}$, $k_{23} = k_{+2}[\mathrm{I}]$, $k_{32} = k_{-2}$, $k_{34} = k_{-3}$, $k_{43} = k_{+3}[\mathrm{S}]$, $k^*_{34} = k^*_{-3}$, $k^*_{43} = k^*_{+3}[\mathrm{P}]$, $k_{41} = k_{-4}$ and $k_{14} = k_{+4}[\mathrm{I}]$, then the directed graph for the Botts-Morales mechanism can be depicted as in fig. 4b. Note the difference in the reaction mechanism here with the mechanism in example 1: The current one is with parallel reaction pathways as reflected by fig. 4b, in which, e.g., there are two arcs from $E_1$ to $E_2$, rather than only one arc as in fig. 2a. Thus, according to Rule 1, we have

$$A = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -2 & 0 & -1 \\ -2 & 3 & -1 & 0 \\ 0 & -1 & 3 & -2 \\ -1 & 0 & -2 & 3 \end{bmatrix} \tag{20}$$

and

$$P_1 = P_2 = P_3 = P_4 = \det B_{4,4} = \begin{vmatrix} 3 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 3 \end{vmatrix} = 12 \tag{21}$$

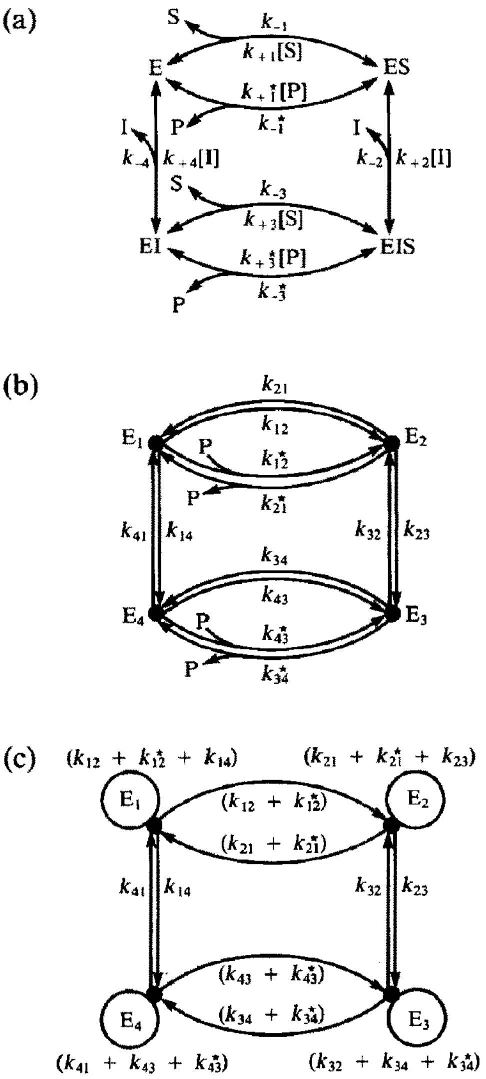


Fig. 4. (a) The Botts-Morales reaction mechanism [27]. (b) The directed graph $G$ depicted to express the Botts-Morales mechanism as shown in panel a. The special mark of P is helpful while calculating the rate of product formation (cf. Rule 3). (c) The transformed graph $D^{\dagger}$ obtained from panel b according to Rule 2(1).

It is interesting to compare eqs. 21 and 16: Because of the existence of the two parallel pathways between $E_1$ and $E_2$, and between $E_3$ and $E_4$, the number of terms for each of $N_m$ ($m = 1, 2, 3, 4$) will increase from 4 to 12, i.e., 3-fold.

*Rule 2: A new method for calculating $N_m$*

(1) Transform the directed graph $G$ to $G^{\dagger}$ according to the following procedure: To each point in $G$ add a loop with a weight equal to the sum of the weights of the arcs departing from that point. If there are

two or more arcs from one enzyme species to another, i.e., the so-called 'parallel pathways' (e.g., see fig. 4b), then condense them into one by adding their rate constants together (cf. fig. 4c).

(2) For any selected reference point, e.g., $E_s$, in $G^\dagger$, find all those subgraphs each of which must contain one, and only one, path from $E_s$ to $E_m$, as well as all cycles and loops that intersect with neither each other nor the path. Then for each of such subgraphs, take the product of all its weights, multiplied with a sign factor given by

$$(-1)^{C_y} \tag{22}$$

where $C_y$ is the number of the cycles (not including loops) in the subgraph concerned. Taking a sum of all these results will immediately give $N_m$ of eq. 5. Note that the result obtained for $N_m$ is always the same regardless of which point [23] is chosen as the reference point $E_s$. It is recommended, however, to choose the reference point $E_s$ as far from $E_m$ as possible because it will in most cases further reduce the number of the subgraphs.

(3) The number of the total subgraphs can be predicted as follows: According to the transformed graph $G^\dagger$, construct a matrix $C = [c_{ij}]$, where

$$c_{ij} = \begin{cases} 1, & \text{if there is an arc from } E_i \text{ to } E_j \text{ in } G^\dagger \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

Thus, when $E_s$ is selected as a reference point to calculate $N_m$, the number of the subgraphs will be given by

$$n^{s \to m} = \text{per } C_{m,s} \tag{24}$$

where $C_{m,s}$ is the submatrix obtained by removing the $m$-th row and $s$-th column from the matrix $C$, and per $C_{m,s}$ denotes the sum of all terms obtained by expanding the determinant of $C_{m,s}$ but taking all the signs of the expanded terms as plus, e.g.,

$$\text{per}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 2, \ \text{per}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \text{per}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \text{per}\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = 2 + 1 = 3$$

and so forth, Therefore, it is very easy to calculate the number of subgraphs in terms of eq. 24 in which the matrix elements are either 1 or 0. The figure obtained from eq. 24 can serve to check if any subgraphs have been missed during the operation of step (2).

The mathematical proof of Rule 2 is presented in appendix B.

*Example 4*

Consider the bi-bi random mechanism whose directed graph is depicted in fig. 3a. Following Rule 2(1), we obtain its transformed graph $G^\dagger$ as shown in fig. 3b, and according to Rule 2(3) we have

$$C = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{25}$$

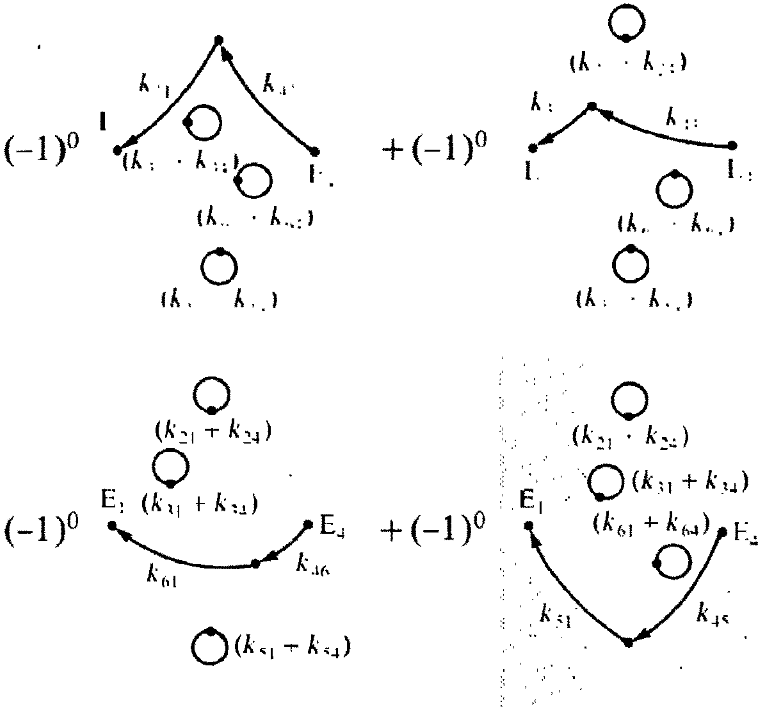
Thus, it follows by means of eq. 24 that

$$n^{4\to1} = \text{per}\,\boldsymbol{C}_{1,4} = \text{per}\begin{bmatrix}1&1&0&0&0\\1&0&1&0&0\\0&1&1&1&1\\1&0&0&1&0\\1&0&0&0&1\end{bmatrix} = \text{per}\begin{bmatrix}0&1&0&0\\1&1&1&1\\0&0&1&0\\0&0&0&1\end{bmatrix} + \text{per}\begin{bmatrix}1&1&0&0\\0&1&1&1\\1&0&1&0\\1&0&0&1\end{bmatrix}$$

$$= \text{per}\begin{bmatrix}1&1&1\\0&1&0\\0&0&1\end{bmatrix} + \text{per}\begin{bmatrix}1&1&1\\0&1&0\\0&0&1\end{bmatrix} + \text{per}\begin{bmatrix}0&1&1\\1&1&0\\1&0&1\end{bmatrix} = 1+1+2=4 \qquad (26)$$

which means only 4 subgraphs need to be counted in order to calculate $N_1$ if $E_4$ is chosen as a reference point. Using the King-Altman method for calculation, however, one has to find 32 in-trees (cf. eq. 19), which seems a formidable task. Since the number of subgraphs has been tremendously reduced, the result of $N_1$ can be readily obtained in terms of Rule 2(2) as follows:

$N_1 =$



$$= k_{42}k_{21}(k_{31}+k_{34})(k_{51}+k_{54})(k_{61}+k_{64}) + k_{43}k_{31}(k_{21}+k_{24})(k_{51}+k_{54})(k_{61}+k_{64})$$
$$+ k_{46}k_{61}(k_{21}+k_{24})(k_{31}+k_{34})(k_{51}+k_{54}) + k_{45}k_{51}(k_{21}+k_{24})(k_{31}+k_{34})(k_{61}+k_{64}) \qquad (27)$$

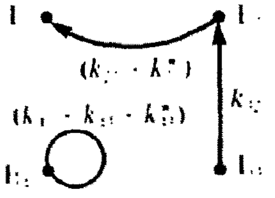
which really contains 32 terms as predicted by Rule 1 (cf. eq. 19). Likewise, $N_2, \ldots, N_6$ can all be easily obtained by means of Rule 2.

*Example 5*

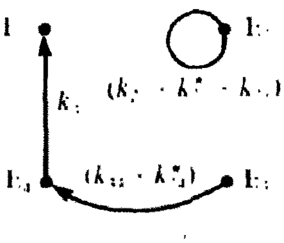
Let us now consider the Botts-Morales mechanism [27] whose directed graph is given in fig. 4b. Using the King-Altman method, 12 subgraphs (cf. eq. 21) need to be counted for calculating each of $N_m$ ($m = 1, 2, 3, 4$). Now let us use Rule 2. Note that the current directed graph contains parallel pathways. Following Rule 2(1), the graph $G$ in fig. 4b is transformed to $G^\dagger$ of fig. 4c. According to Rule 2(2), it follows that

$N_1 =$

$(-1)^0$



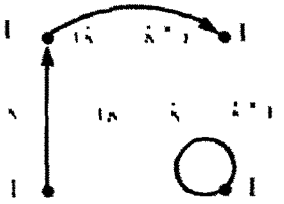
$+(-1)^0$



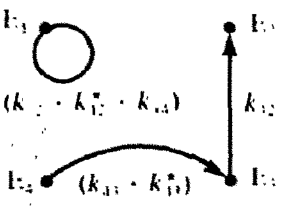
$$= k_{32}(k_{21}+k_{21}^*)(k_{41}+k_{43}+k_{43}^*) + k_{41}(k_{34}+k_{34}^*)(k_{21}+k_{21}^*+k_{23}) \tag{28}$$

$N_2 =$

$(-1)^0$



$+(-1)^0$



$$= k_{41}(k_{12}+k_{12}^*)(k_{32}+k_{34}+k_{34}^*) + k_{32}(k_{43}+k_{43}^*)(k_{12}+k_{12}^*+k_{14}) \tag{29}$$

$$N_3 = k_{23}(k_{12}+k_{12}^*)(k_{41}+k_{43}+k_{43}^*) + k_{14}(k_{43}+k_{43}^*)(k_{21}+k_{21}^*+k_{23}) \tag{30}$$

$$N_4 = k_{14}(k_{21}+k_{21}^*)(k_{32}+k_{34}+k_{34}^*) + k_{23}(k_{34}+k_{34}^*)(k_{12}+k_{12}^*+k_{14}) \tag{31}$$

Now let us use Rule 2(3) to check the above calculation. From eqs. 23–24, it follows that

$$C = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \tag{32}$$

and

$$n^{3\to 1} = \operatorname{per} C_{1,3} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = 2 \tag{33}$$

$$n^{4\to 2} = n^{1\to 3} = n^{2\to 4} = \operatorname{per} C_{4,2} = 2 \tag{34}$$

Eqs. 33–34 indicate that no subgraph was missed in the above calculation. Besides, this example further demonstrates the merit of Rule 2: the amount of work involved in searching the subgraphs can be reduced 6-fold in comparison with the King-Altman method.

*Rule 3: A new method for calculating the rate of product formation*

(1) The rate of product formation can be written as follows

$$v=\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t}=\frac{F^{+}-F^{-}}{\Sigma_{i=1}^{n}N_i}e_0 \tag{35}$$

where only the factors $F^{+}$ and $F^{-}$ are unknown which, however, are closely associated with the product-creating and product-annihilating cycles, whose definitions are given below, respectively.

(2) A product-creating cycle is a cycle in $G$ along which there is a net release of product P (cf. fig. 3a as well as the cycles in eq. 39), while a product-annihilating cycle is a cycle along which for a circuit there is a net consumption of product P (cf. fig. 3a as well as the cycles in eq. 40). Both types of cycle play a key role in calculation of the rate of product formation and hence may be termed master cycles. In calculations, however, all these master cycles should be found directly from the original directed graph $G$ rather than the transformed graph $G^{\dagger}$, especially for the case with parallel pathways.

(3) Find all those subgraphs each of which must have one, and only one, product-creating cycle (or product-annihilating cycle), and all other cycles and loops. The latter, however, are found from the transformed graph $G^{\dagger}$, but they must intersect with neither each other nor the master cycle even when the graphs $G$ and $G^{\dagger}$ overlap. Then, for each of these subgraphs, take the product of all its weights, multiplied with a factor given by

$$(-1)^{C_y^*}g \tag{36}$$

where $C_y^*$ is the number of cycles (not including the master cycle and loops), and $g$ the number of the product P released (or consumed) when a circuit along the product-creating cycle (or product-annihilating cycle) is completed. Taking a sum of these results will immediately give $F^{+}$ (or $F^{-}$).

The mathematical proof is presented in appendix C.

*Example 6*

Calculate the rate of product formation for the bi-bi random mechanism as shown in fig. 3a. According to eqs. 2 and 5, we have

$$v=\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t}=\left[(k_{61}N_6-k_{16}N_1)+(k_{45}N_4-k_{54}N_5)\right]\frac{e_0}{\Sigma_{i=1}^{6}N_i} \tag{37}$$

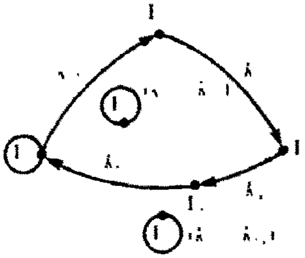
When using the King-Altman method or other conventional methods to calculate the numerator of the above equation, one has to deal with 4 × 32 = 128 terms (cf. eq. 19) among which 96 terms will eventually be cancelled out with each other (cf. eqs. 38–40). This kind of cancellation operation is not only time-consuming but also error-prone. By means of Rule 3, however, eq. 37 can be expressed as

$$v=\frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t}=\frac{F^{+}-F^{-}}{\Sigma_{i=1}^{6}N_i}e_0 \tag{38}$$

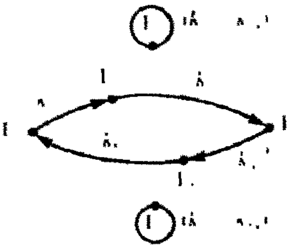
where

$F^{+} =$

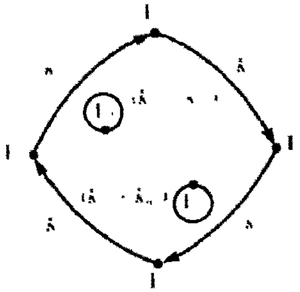
$(-1)^0$



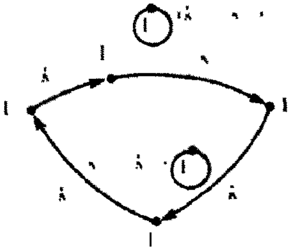
$+(-1)^0$



$+(-1)^0$



$+(-1)^0$

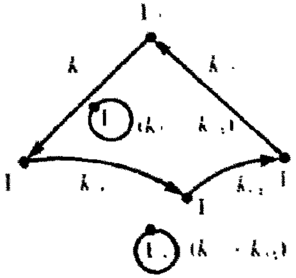


$$\begin{aligned}&= k_{12}k_{24}k_{46}k_{61}(k_{31}+k_{34})(k_{51}+k_{54}) + k_{13}k_{34}k_{46}k_{61}(k_{21}+k_{24})(k_{51}+k_{54})\\ &\quad + k_{12}k_{24}k_{45}k_{51}(k_{31}+k_{34})(k_{61}+k_{64}) + k_{13}k_{34}k_{45}k_{51}(k_{21}+k_{24})(k_{61}+k_{64})\end{aligned} \tag{39}$$

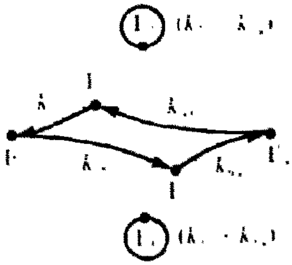
and

$F^{-} =$

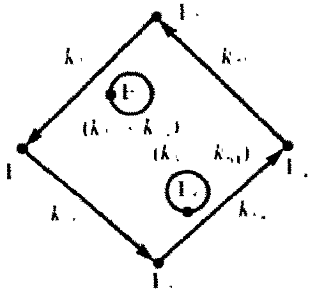
$(-1)^0$



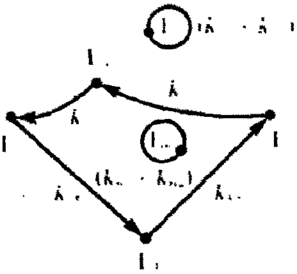
$+(-1)^0$



$+(-1)^0$



$+(-1)^0$



$$= k_{16}k_{64}k_{42}k_{21}(k_{31}+k_{34})(k_{51}+k_{54}) + k_{16}k_{64}k_{43}k_{31}(k_{21}+k_{24})(k_{51}+k_{54})$$
$$+ k_{15}k_{54}k_{42}k_{21}(k_{31}+k_{34})(k_{61}+k_{64}) + k_{15}k_{54}k_{43}k_{31}(k_{21}+k_{24})(k_{61}+k_{64}) \qquad (40)$$

Substituting eqs. 39–40 into eq. 38, we immediately obtain the desired result, the tedious job involved in finding and cancelling the 96 terms from the 128 terms being completely avoided.

*Example 7*

Calculate the rate of product formation for the Botts-Morales mechanism [27]. Note that this mechanism contains parallel pathways. According to its directed graph as shown in fig. 4b, there are 4 product-creating cycles and 4 product-annihilating cycles. Actually, after becoming familiar with these rules, we do not have to depict the subgraphs one by one as illustrated above. The desired results can be directly written out according to the directed graph $G$ and its transformed graph $G^\dagger$ as follows:

$$F^+ = k_{12}k_{21}^*(k_{32}k_{41}+k_{32}k_{43}+k_{43}^*k_{32}+k_{34}k_{41}+k_{34}^*k_{41}) + k_{14}k_{43}k_{32}k_{21}^*$$
$$+ k_{43}k_{34}^*(k_{21}k_{14}+k_{21}^*k_{14}+k_{23}k_{14}+k_{12}k_{23}+k_{12}^*k_{23}) + k_{12}k_{23}k_{34}^*k_{41} \qquad (41)$$
$$F^- = k_{21}k_{12}^*(k_{32}k_{41}+k_{32}k_{43}+k_{43}^*k_{32}+k_{34}k_{41}+k_{34}^*k_{41}) + k_{12}^*k_{23}k_{34}k_{41}$$
$$+ k_{34}k_{43}^*(k_{21}k_{14}+k_{21}^*k_{14}+k_{23}k_{14}+k_{12}k_{23}+k_{12}^*k_{23}) + k_{14}k_{43}^*k_{32}k_{21} \qquad (42)$$

Substituting eqs. 41–42 as well as eqs. 28–31 into

$$v = \frac{\mathrm{d}[\mathrm{P}]}{\mathrm{d}t} = \frac{F^+ - F^-}{\sum_{i=1}^4 N_i} e_0 \qquad (43)$$

we immediately obtain the rate of product formation for the Botts-Morales mechanism. In comparison with the King-Altman method according to which

$$v = (k_{21}^*N_2 + k_{34}^*N_3 - k_{12}^*N_1 - k_{43}^*N_4)\frac{e_0}{\sum_{i=1}^4 N_i} \qquad (44)$$

the tedious work required for finding 12 × (4 − 2) = 24 reciprocally cancelled terms (cf. eq. 21) as well as (12–2) × 4 = 40 subgraphs is completely avoided.

## 3. Non-steady-state kinetic systems

For non-steady-state kinetic systems, instead of eqs. 3–4, we should consider the following equations:

$$-\frac{\mathrm{d}e_i}{\mathrm{d}t} = e_i\sum_{j=1}^{n} k_{ij} - \sum_{j=1}^{n} k_{ji}e_j \quad (i = 1, 2, \ldots, n) \qquad (45)$$

$$\sum_{i=1}^{n} e_i = e_0 \qquad (46)$$

with the following initial conditions:

$$\left.\begin{array}{l} e_1(t) = e_0 \\ e_i(t) = 0 \ (i \neq 1) \end{array}\right\} \text{when } t = 0 \qquad (47)$$

where $e_1 = [\mathrm{E}_1]$ is the concentration of the free enzyme. As is well known, it is much more difficult and complicated [28,29] to determine the solution for a non-steady-state system in comparison with the corresponding steady-state one as described by eqs. 3–4. The following question is naturally raised: Can we also find a graphic method to deal with the non-steady-state system as we did for the steady-state one? The answer is yes. As an approach to realize that, let us first make the Laplace transformation for eq. 45–46 with the initial condition (eq. 47), which yields

$$\left(s+\sum_{j=1}^{n} k_{ij}\right)\tilde{e}_i - \sum_{j=1}^{n} k_{ji}\tilde{e}_j = \delta_{1i} e_0 \quad (i = 1, 2, \ldots, n) \tag{48}$$

$$s\sum_{i=1}^{n} \tilde{e}_i = e_0 \tag{49}$$

where $\delta_{1i}$ is the Kronecker delta, $s$ an interim parameter introduced by the following Laplace transform

$$\tilde{e}_i(s) = \int_0^{\infty} e_i(t)\exp(-ts)\,\mathrm{d}t \tag{50}$$

and $\tilde{e}_i$ is the *phase concentration* of $E_i$. The relation between the phase concentration $\tilde{e}_i$ and the usual concentration $e_i$ can be simply expressed as

$$\left.\begin{aligned} \tilde{e}_i &= \hat{\boldsymbol{L}} e_i \\ e_i &= \hat{\boldsymbol{L}}^{-1}\tilde{e}_i \end{aligned}\right\} \tag{51}$$

where $\hat{\boldsymbol{L}}$ and $\hat{\boldsymbol{L}}^{-1}$ are the Laplace transformation and inverse transformation operator, respectively.

Below, a parallel graphic method is presented by which one can directly write out the phase concentrations according to the directed graph without the need to solve eqs. 48–49 whose operation is even more tedious and error-prone than in the case of steady-state systems. Once the phase concentrations $\tilde{e}_i$ $(i = 1, 2, \ldots, n)$ are known, the corresponding normal concentrations $e_i(t)$ $(i = 1, 2, \ldots, n)$ can be immediately obtained according to the Laplace transform table available in any mathematical handbook.

*Rule 4: A graphic method for calculating the phase concentrations $\tilde{e}_m$ $(m = 1, 2, \ldots, n)$*

(1) Transform the directed graph $G$ to $\tilde{G}$ according to the following procedures: To each of the points $E_m$ $(m = 1, 2, \ldots, n)$ add a loop with the weight $s + \Sigma_{j=1}^{n} k_{mj}$, respectively. If there are two or more arcs from one point to the same other, then condense them into one by adding their rate constants together.

(2) The graph obtained through the above procedure is called the phase graph $\tilde{G}$. Thus the phase concentration $\tilde{e}_m$ for the $m$-th enzyme species $E_m$ is given by

$$\tilde{e}_m = \frac{\tilde{N}_m}{s\Sigma_{i=1}^{n}\tilde{N}_i} e_0 \tag{52}$$

where $\tilde{N}_m$ can be obtained as follows: From the phase graph $\tilde{G}$ find all the subgraphs each of which has one, and only one, path from $E_1$ to $E_m$, as well as all the cycles and loops that intersect with neither each other nor the path. For each such subgraph, multiply all its weights and a sign factor given by

$$(-1)^{\tilde{C}_y} \tag{53}$$

where $\tilde{C}_y$ is the number of the cycles (not including loops) in the respective subgraph. The sum of all these results will immediately give $\tilde{N}_m$. When $m = 1$, however, the path from $E_1$ to $E_m$ will reduce to a point, whose weight in this case should be assigned as 1 for calculations (e.g., see eq. 57).

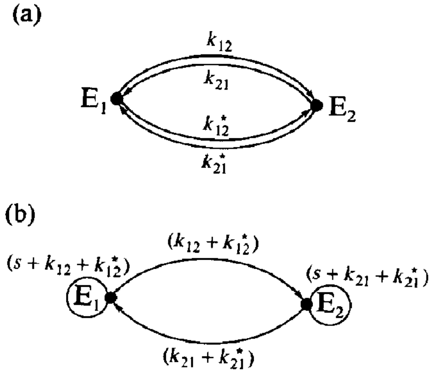


Fig. 5. (a) The directed graph $G$ for the Michaelis-Menten mechanism as formulated in eq. 56. (b) The phase graph $\tilde{G}$ obtained from panel a according to the procedure as stated in Rule 4(1).

(3) In order to facilitate checking and avoiding missing any subgraphs, we can also predict the number of subgraphs to be counted. The method is as follows. Construct a matrix $\tilde{C} = [\tilde{c}_{ij}]$, where

$$\tilde{c}_{ij} = \begin{cases} 1, & \text{if there is an arc from } E_i \text{ to } E_j \text{ in } \tilde{G} \\ 0, & \text{otherwise} \end{cases} \tag{54}$$

then the number of subgraphs involved in calculating $\tilde{N}_m$ must be

$$n^{1\to m} = \text{per}\ \tilde{C}_{m,1} \tag{55}$$

where $\tilde{C}_{m,1}$ is the submatrix obtained by removing the $m$-th row and the 1st column from $\tilde{C}$, and 'per' has the same definition as in eq. 24.

The mathematical proof of Rule 4 is given in appendix D.

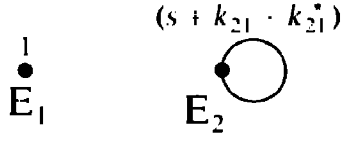
*Example 8*

Find the non-steady-state solution for the Michaelis-Menten mechanism [30]:

$$E + S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} E + P \tag{56}$$

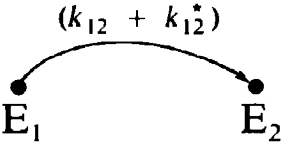
where the concentration of E is equal to $e_0$ and that of ES equal to 0 when $t = 0$. Let $E_1 = E$, $E_2 = ES$, $k_{12} = k_{+1}[S]$, $k_{21} = k_{-1}$, $k_{12}^* = k_{-2}[P]$, $k_{21}^* = k_{+2}$, then the Michaelis-Menten mechanism of eq. 56 can be expressed by the directed graph as shown in fig. 5a. Following the procedure as described in Rule 4(1), the directed graph $G$ in fig. 5a can be transformed to the phase graph $\tilde{G}$ as given in fig. 5b. Then according to Rule 4(2), it follows that

$\tilde{N}_1 = (-1)^0$



$$= s + k_{21} + k_{21}^* \tag{57}$$

$$\tilde{N}_2 = (-1)^0$$



$$= k_{12} + k_{12}^* \tag{58}$$

Substituting the above results into eq. 52, we obtain

$$\tilde{e}_1 = \frac{s + k_{21} + k_{21}^*}{s(s + k_{12} + k_{21} + k_{12}^* + k_{21}^*)} e_0 \tag{59}$$

$$\tilde{e}_2 = \frac{k_{12} + k_{12}^*}{s(s + k_{12} + k_{21} + k_{12}^* + k_{21}^*)} e_0 \tag{60}$$

Using the table of Laplace transforms (cf., e.g., ref. 31), we immediately obtain

$$e_1(t) = e_0 \left\{ \frac{k_{21} + k_{21}^*}{k_{12} + k_{21} + k_{12}^* + k_{21}^*} + \frac{k_{12} + k_{12}^*}{k_{12} + k_{21} + k_{12}^* + k_{21}^*} \exp[-(k_{12} + k_{21} + k_{12}^* + k_{21}^*)t] \right\} \tag{61}$$

$$e_2(t) = e_0 \left\{ \frac{k_{12} + k_{12}^*}{k_{12} + k_{21} + k_{12}^* + k_{21}^*} - \frac{k_{12} + k_{12}^*}{k_{12} + k_{21} + k_{12}^* + k_{21}^*} \exp[-(k_{12} + k_{21} + k_{12}^* + k_{21}^*)t] \right\} \tag{62}$$

Now let us see how to use the check formula of eq. 55. According to the phase graph $\tilde{G}$ of fig. 5b we have

$$\tilde{C} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{63}$$

thus it follows from eq. 55 that

$$n^{1 \to 1} = \text{per } C_{1,1} = [1] = 1, \quad n^{1 \to 2} = \text{per } C_{2,1} = [1] = 1 \tag{64}$$

which means that no subgraphs were missed in calculating the phase concentrations as illustrated in eqs. 57–58.

*Example 9*

Consider the non-steady-state kinetics of the following three-state model of protein folding

$$\mathrm{D} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{X} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathrm{N} \tag{65}$$

where D represents the denatured and unfolded protein species, N the native and folded protein species, and X the intermediate in the pathway between unfolded and folded states. Let $E_1 = D$, $E_2 = X$, $E_3 = N$, $k_{12} = k_1$, $k_{21} = k_{-1}$, $k_{23} = k_2$, $k_{32} = k_{-2}$, then the three-state folding model can be expressed by the directed graph $G$ as shown in fig. 6a. Assume that initially only denatured protein is present, viz., $e_1 = [D] = e_0$, $e_2 = [X] = 0$ and $e_3 = [N] = 0$ when $t = 0$. Here $e_0$ is the total concentration of protein. The protein is then subjected to a rapid temperature jump, a sudden change in solvent, or some other quick change that causes the protein to fold. For simplifying illustration, suppose $k_{12}$, $k_{23} \gg k_{21}$, $k_{32}$, which

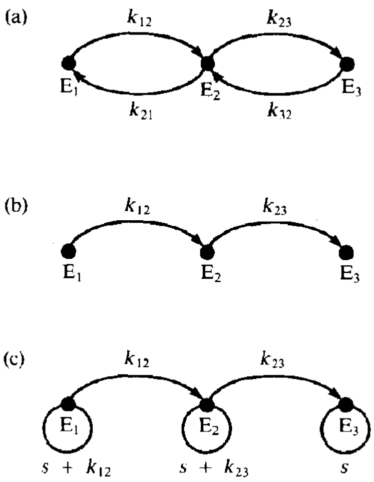
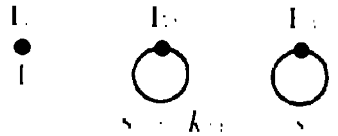


Fig. 6. (a) The directed graph for the three-state folding model as formulated in eq. 65. (b) The directed graph reduced from panel a when $k_{12}, k_{23} \gg k_{21}, k_{32}$. (c) The phase graph $\tilde{G}$ obtained from panel b according to Rule 4(1).

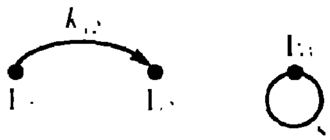
means the tendency of protein towards folding is dominant over its tendency towards unfolding after such a sudden change. In this case, the directed graph $G$ in fig. 6a can be reduced to fig. 6b. According to Rule 4(1), the directed graph $G$ in fig. 6b can be transformed to the phase graph $\tilde{G}$ of fig. 6c. It follows thus by Rule 4(2) that

$$\tilde{N}_1 = (-1)^0$$



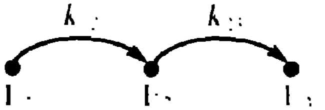
$$= (s + k_{23})s \qquad (66)$$

$$\tilde{N}_2 = (-1)^0$$



$$= k_{12}s \qquad (67)$$

$$\tilde{N}_3 = (-1)^0$$



$$= k_{12}k_{23} \qquad (68)$$

Substituting the above results into eq. 52 yields

$$\tilde{e}_1 = \frac{(s+k_{23})s}{s\left[s^2+(k_{23}+k_{12})s+k_{12}k_{23}\right]}e_0 = \frac{(s+k_{23})s}{s(s+k_{12})(s+k_{23})}e_0 = \frac{1}{s+k_{12}}e_0 \tag{69}$$

$$\tilde{e}_2 = \frac{k_{12}}{(s+k_{12})(s+k_{23})}e_0 \tag{70}$$

$$\tilde{e}_3 = \frac{k_{12}k_{23}}{s(s+k_{12})(s+k_{23})}e_0 \tag{71}$$

Using the table of Laplace transforms (cf., e.g., ref. 31), we immediately obtain the transient concentrations of the protein at three different folding states. The solutions may be classified into the following two cases:

$$e_1(t) = e^{-k_{12}t}e_0 \tag{72}$$

$$e_2(t) = \frac{k_{12}}{k_{23}-k_{12}}\left(e^{-k_{12}t} - e^{-k_{23}t}\right)e_0 \tag{73}$$

$$e_3(t) = \left[\frac{1}{k_{23}-k_{12}}\left(k_{12}e^{-k_{23}t} - k_{23}e^{-k_{12}t}\right) + 1\right]e_0 \tag{74}$$

(b) When $k_{12} = k_{23} = k$, we have

$$e_1(t) = e^{-kt}e_0 \tag{75}$$

$$e_2(t) = kte^{-kt}e_0 \tag{76}$$

$$e_3(t) = (1 - e^{-kt} - kte^{-kt})e_0 \tag{77}$$

From examples 8 and 9 we see that the operation neither of expanding determinants nor of solving differential equations is needed; we can directly write out the transient concentrations of enzyme species or protein folding states just by looking up the Laplace transform table as well as its directed graph. Many complicated and difficult mathematical derivations can thereby be avoided.

## Appendix A: The mathematical principle of Rule 1

First, let us point out that an $n \times n$ matrix $X = [x_{ij}]$ can be expressed by a weighted *digraph* (i.e., directed graph) $G(X)$ with $n$ vertices ($V_1$, $V_2, \ldots, V_n$). The method is as follows: If $x_{ij} \neq 0$ ($i \neq j$), draw an arc from vertex $V_i$ to vertex $V_j$, and weight it with $x_{ij}$; if $x_{ij} = 0$ ($i \neq j$), do not depict an arc from $V_i$ to $V_j$. If $x_{ii} \neq 0$, draw a loop at vertex $V_i$, i.e., an arc from vertex $V_i$ to itself, and weight it with $x_{ii}$; if $x_{ii} = 0$, however, no loop should be drawn at vertex $V_i$. Through such a procedure, a one-to-one correspondence between the matrix and the digraph is established. This is the essence of why some calculations relevant to a matrix can be approached through the graphic methods as well.

According to the graphic theory, the graph $G$ as illustrated in the text is actually a weighted digraph, in which the points representing the enzyme species (or protein folding states) are the vertices, and the reaction rate constants $k_{ij}$ are the matrix elements $x_{ij}$. Also, as mentioned in the text, each of the counted subgraphs in the King-Altman method corresponds to a spanning in-tree of $G$ with some special point as a sink. A *sink* is a vertex which all the other vertices in a digraph can reach via at least one path, and a *spanning in-tree* of graph $G$ is a subgraph of $G$ that contains no cycles or semicycles but that contains one (and only one) sink of $G$.

As indicated in eqs. 8–11, the number of terms in $N_m$ is actually equal to that of spanning in-trees with $E_m$ as a sink in the corresponding graph $G$. On the other hand, the matrix $\boldsymbol{B}$ obtained through steps (1)–(2) of Rule 1 may also be expressed as

$$\boldsymbol{B} = \boldsymbol{A}^* - \boldsymbol{A} \tag{A1}$$

where $\boldsymbol{A}^*$ is a diagonal matrix whose diagonal element $a_{ii}^*$ is equal to the sum of all the elements of the $i$-th row in matrix $\boldsymbol{A}$. Thus, according to the theorem given by Tutte [32], that the number of spanning in-tree with point $E_m$ as a sink in the digraph $G$ is equal to the cofactor of any element in the $m$-th row of the matrix $\boldsymbol{A}^* - \boldsymbol{A}$; i.e.

$$P_m = \det \boldsymbol{B}_{m,1} = \det \boldsymbol{B}_{m,2} = \cdots = \det \boldsymbol{B}_{m,m} = \cdots = \det \boldsymbol{B}_{m,n} \tag{A2}$$

In other words, eq. 14 is only one choice from the $n$ equivalent equations in eq. A2. Furthermore, when the directed graph $G$ is a reversible one (i.e., whenever there is an arc from $E_i$ to $E_j$, there must be an arc from $E_j$ to $E_i$), both $\boldsymbol{A}$ and $\boldsymbol{B}$ must be symmetric and we therefore have

$$\boldsymbol{B}_{i,j} = \boldsymbol{B}_{j,i} \quad (i, j = 1, 2, \ldots, n) \tag{A3}$$

From eqs. A2 and A3 it follows that, when $G$ is a reversible graph,

$$P_1 = P_2 = \cdots = P_m = \cdots = P_n \tag{A4}$$

This completes the mathematical proof of Rule 1.

## Appendix B: The mathematical principle of Rule 2

Let us first prove that for any $n \times n$ matrix $\boldsymbol{X} = [x_{ij}]$, we have

$$\det \boldsymbol{X} = \sum_u (-1)^{q_u} f(G_u) \tag{B1}$$

where $G_u$ is the $u$-th of those subgraphs each of which contains all disjoint cycles and loops of $G(\boldsymbol{X})$ (see the first paragraph of appendix A for the definition of $G(\boldsymbol{X})$), $q_u$ is the number of cycles with even number of arcs in $G_u$, and $f(G_u)$ denotes the product of the weights of all the arcs in $G_u$. The proof of eq. B1 is as follows. According to the definition of a determinant

$$\det \boldsymbol{X} = \sum_\sigma \operatorname{sgn} \sigma \, x_{1j_1} x_{2j_2} \cdots x_{nj_n} \tag{B2}$$

where $\sigma = \{j_1, j_2, \ldots, j_n\}$ is a permutation of $\{1, 2, \ldots, n\}$, and

$$\operatorname{sgn} \sigma = \begin{cases} 1, & \text{for even permutation} \\ -1, & \text{for odd permutation} \end{cases} \tag{B3}$$

If $x_{1j_1} x_{2j_2} \cdots x_{nj_n} \neq 0$, it must also equal the product of weights of the arcs $\overline{V_1V_{j_1}}, \overline{V_2V_{j_2}}, \ldots, \overline{V_nV_{j_n}}$ in $G(\boldsymbol{X})$, where each given subscript appears only twice: one is in the initial point of an arc, and the other is in the terminal point of an arc. Consequently, $\overline{V_1V_{j_1}}, \overline{V_2V_{j_2}}, \ldots, \overline{V_nV_{j_n}}$ must correspond to a $G_u$ of eq. B1. In other words, there is a one-to-one correspondence between the non-zero terms of det $X$ and $G_u$ ($u = 1, 2, \ldots$) of $G(\boldsymbol{X})$; i.e., we have

$$\det \boldsymbol{X} = \sum_u \operatorname{sgn} \sigma_u f(G_u) \tag{B4}$$

where sgn $\sigma_u$ is the sign which can be determined as follows. As mentioned above, each $G_u$ $(u = 1, 2\ldots)$ consists of disjoint cycles and loops. Suppose $V_{i_1}$, $\widehat{V_{i_1}V_{i_2}}$, $V_{i_2}$, $\widehat{V_{i_2}V_{i_3}}, \ldots, V_{i_l}$, $\widehat{V_{i_l}V_{i_1}}$, $V_{i_1}$ is a cycle formed by $l$ vertices and $l$ arcs in $G_u$. According to eq. B3 the term $x_{i_1i_2}x_{i_2i_3}\cdots x_{i_li_1}$, which is actually the product of the weights of all the arcs of such a cycle, will contribute a factor of $(-1)^{l-1}$ to sgn $\sigma_u$. Therefore, if taking all disjoint cycles of $G_u$ into account, we have

$$\operatorname{sgn}\sigma_u = \prod_{\{l\}}(-1)^{l-1} \tag{B5}$$

where $l$ is the number of arcs in each of the disjoint cycles in $G_u$. Since

$$(-1)^{l-1} = \begin{cases} -1, & \text{when } l \text{ is an even number} \\ 1, & \text{when } l \text{ is an odd number} \end{cases} \tag{B6}$$

it follows that

$$\operatorname{sgn}\sigma_u = (-1)^{\{\text{the number of cycles in } G_u \text{ that have even number of arcs}\}} \tag{B7}$$

Substitution of eq. B7 into eq. B4 completes the proof of eq. B1.

On the other hand, according to the definition of the *permanent* of a matrix, we have

$$\operatorname{per} \boldsymbol{X} = \sum_{\sigma} x_{1j_1}x_{2j_2}\cdots x_{nj_n} \tag{B8}$$

In comparison with det $\boldsymbol{X}$ of eq. B2, the only difference is in the sign factor sgn$\sigma$ which appears in det $\boldsymbol{X}$ but not in per $\boldsymbol{X}$. Therefore, if $\boldsymbol{C}$ is built from $G(\boldsymbol{X})$ according to Rule 2(3), it is obvious that

$$\begin{aligned} \operatorname{per} \boldsymbol{C} &= \{\text{the number of non-zero terms in det } \boldsymbol{X}\} \\ &= \{\text{the number of } G_u \text{ in } G(\boldsymbol{X})\} \end{aligned} \tag{B9}$$

Now we can use eqs. B1 and B9 to prove Rule 2. Note that the $n$ equations in eq. 3 are not independent because the sum of these $n$ equations is identically equal to zero. To obtain the unique solution, therefore, we can replace any one of them by eq. 4. The expression for $N_m$ as given in eq. 6 is obtained by replacing the $m$-th equation with eq. 4. However, if we use eq. 4 to replace any other equation in eq. 3, the solution should remain the same. Besides, a matrix and its transposed matrix should have the same determinant value. Based on the above two grounds, eq. 6 can be rewritten as

$$N_m = \begin{vmatrix} 1 & -k_{12} & \ldots & -k_{1m} & \ldots & -k_{1n} \\ 1 & \Sigma_{j=1}^{n}k_{2j} & \ldots & -k_{2m} & \ldots & -k_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & -k_{n2} & \ldots & -k_{nm} & \ldots & \Sigma_{j=1}^{n}k_{nj} \end{vmatrix} = \begin{vmatrix} \Sigma_{j=1}^{n}k_{1j} & 1 & \ldots & -k_{1m} & \ldots & -k_{1n} \\ -k_{21} & 1 & \ldots & -k_{2m} & \ldots & -k_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 1 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -k_{n1} & 1 & \ldots & -k_{nm} & \ldots & \Sigma_{j=1}^{n}k_{nj} \end{vmatrix}$$

$$= \cdots = \begin{vmatrix} \Sigma_{j=1}^{n}k_{1j} & -k_{12} & \ldots & 1 & \ldots & -k_{1n} \\ -k_{21} & \Sigma_{j=1}^{n}k_{2j} & \ldots & 1 & \ldots & -k_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -k_{n1} & -k_{n2} & \ldots & 1 & \ldots & \Sigma_{j=1}^{n}k_{nj} \end{vmatrix} =$$

$$= \cdots = \begin{vmatrix} \Sigma_{j=1}^{n} k_{1j} & -k_{12} & \cdots & -k_{1m} & \cdots & 1 \\ -k_{21} & \Sigma_{j=1}^{n} k_{2j} & \cdots & -k_{2m} & \cdots & 1 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots & & \vdots \\ -k_{n1} & -k_{n2} & \cdots & -k_{nm} & \cdots & 1 \end{vmatrix} \tag{B10}$$

On inspecting eq. B10 we discover the following:

(1) According to the relation between terms and graphs as given by eq. B1, in all the graphs corresponding to $N_m$, there is no arc starting from $E_m$. Therefore, any cycle containing point $E_m$ must be reduced to a path. On the other hand, if we choose eq. 4 to replace the $s$-th equation of eq. 3, in all the corresponding graphs, there is no arc ending at the point $E_s$. Therefore, the reduced cycle must be a path from point $E_s$ to $E_m$. This is actually reflected by selecting $E_s$ as a reference point to calculate $N_m$ as stated in Rule 2(2). In a particular case, if $s = m$, i.e., the reference point is $E_m$ itself, the path is further reduced to a point with weight of 1 as implied in eq. B10. (Note that according to eq. B1 if the path is reduced from a cycle with even number of arcs, then a factor of $(-1)$ will be added; if the path is reduced from a cycle with odd number of arcs, no such a factor should be added.)

(2) For brevity, all the minus signs before the elements of $N_m$ in eq. B10 can be removed through an appropriate adjustment in sign. Such an adjustment, combined with $(-1)^{q_u}$ in eq. B1 as well as the sign contributed from the path as described in step (1) of this section, will eventually lead to eq. 22.

(3) Again, from eqs. B9 and B10 it is obvious that if $E_s$ is selected as a reference point, the number of subgraphs needed to be counted for calculating $N_m$ is given by eq. 24.

This completes the proof of Rule 2.

## Appendix C: The mathematical principle of Rule 3

Rule 3 can be directly derived from Rule 2. Let us first consider $k_{\alpha\beta}e_\alpha$ in eq. 2. If we take $E_\beta$ as a reference point to calculate $e_\alpha$ as stated in Rule 2, then the path from $E_\beta$ to $E_\alpha$ can be augmented to become a master cycle due to the multiplicator $k_{\alpha\beta}$. This master cycle is none other than the product-creating cycle as defined in Rule 3. Similarly, for $k_{\beta\alpha}e_\beta$, if $E_\alpha$ is chosen as a reference point to calculate $e_\beta$, then the path from $E_\alpha$ to $E_\beta$ as mentioned in Rule 2 will become a product-annihilating cycle because of the multiplicator $k_{\beta\alpha}$.

Unlike the other cycles as defined in eq. 22, these two types of cycles are generated from a path and hence should have no contribution to the sign factor, as reflected by the definition of $C_y^*$ in eq. 36.

Furthermore, from eq. 2 we see that the master cycle must be those cycles which involve the releasing and (or) consuming of product P. Nevertheless, even though a cycle has such an involvement, but the net number of P created (or annihilated) is zero, such as the cycles $\widehat{E_1E_6E_4E_5E_1}$ and $\widehat{E_1E_5E_4E_6E_1}$ of fig. 3a, it is still not a master cycle because its actual contribution is reciprocally cancelled. This point is just reflected by the size factor $g$ in eq. 36.

This completes the proof of Rule 3.

## Appendix D: The mathematical principle of Rule 4

The phase concentrations are the solutions of eqs. 48 and 49. However, the $n+1$ equations in eqs. 48–49 are not independent. The first equation in eq. 48 is equal to eq. 49 minus all the other equations, i.e., $i = 2, 3, \ldots, n$, of eq. 48. Therefore, instead of solving eqs. 48–49, we can consider the following equations:

$$s \sum_{i=1}^{n} \tilde{e}_i = e_0 \tag{D1}$$

$$\left( s + \sum_{j=1}^{n} k_{ij} \right) \tilde{e}_i - \sum_{j=1}^{n} k_{ji} \tilde{e}_j = 0 \; (i = 2, 3, \ldots, n) \tag{D2}$$

Making a comparison between eqs. D1–D2 and eqs. 3–4 in which only $n$ equations are independent, we find that, if using eq. 4 to replace the first equation in eq. 3, the $n$ independent equations thus obtained are completely in the same form as the $n$ equations in eqs. D1 and D2. In other words, eqs. D1–D2 are only a special case of eqs. 3–4. Therefore, the mathematical proof as given in appendix B for Rule 2 has already covered the proof for Rule 4. The difference in the form of coefficients would not influence the established method itself but only change the weighting of arcs during the transformation of the directed graph as described by Rule 4(1).

Furthermore, the fact that the form of eqs. 3–4 is more general than that of eqs. D1–D2 is also reflected by such a difference between Rule 2 and Rule 4: In Rule 2, any point can be selected as a reference point to calculate $N_m$; in Rule 4, however, the reference point must be $E_1$ since all paths start from $E_1$.

This completes the proof of Rule 4.

## Acknowledgments

Valuable discussions with Dr. Gerald M. Maggiora, Dr. Boryeu Mao, and Dr. Mark A. Johnson are gratefully acknowledged. The author is also indebted to Dr. W. Jeffrey Howe, whose comments led to the incorporation of the mathematical principles into this paper.

## References

1 E.L. King and C. Altman, J. Phys. Chem. 60 (1956) 1375.
2 C. Frieden, J. Biol. Chem. 245 (1970) 5788.
3 G.R. Ainslie, J.P. Shill and K.E. Neet, J. Biol. Chem. 247 (1972) 7088.
4 M. Eigen and P. Schuster, The hypercycle: a principle of natural self-organization (Springer, Berlin, 1979).
5 R.H. Jackson, J.A. Cole and A. Cornish-Bowden, Biochem. J. 199 (1981) 171.
6 J.T.F. Wong and C.S. Hames, Can. J. Biochem. Physiol. 40 (1962) 763.
7 M.V. Volkenstein and B.N. Goldstein, Biochim. Biophys. Acta 115 (1966) 471.
8 T.L. Hill, J. Theor. Biol. 10 (1966) 442.
9 M.V. Volkenstein, B.N. Goldstein and V.E. Stefanov, Mol. Biol. (Moscow) 1 (1967) 9.
10 S. Cha, J. Biol. Chem. 243 (1968) 820.
11 B.A. Orsi, Biochim. Biophys. Acta 258 (1972) 4.
12 N. Seshagiri, J. Theor. Biol. 34 (1972) 469.
13 S. Ainsworth, J. Theor. Biol. 43 (1974) 161.
14 K.J. Indge and R.E. Childs, Biochem. J. 155 (1976) 567.
15 A. Cornish-Bowden, Biochem. J. 165 (1977) 55.
16 K.C. Chou, S.P. Jiang, W.M. Liu and C.H. Fee, Sci. Sin. (Engl. Ed.) 22 (1979) 341.
17 E.P. Whitehead, J. Theor. Biol. 80 (1979) 355.
18 K.C. Chou and S. Forsén, Biochem. J. 187 (1980) 829.
19 K.C. Chou, Eur. J. Biochem. 113 (1980) 195.
20 K.C. Chou, J. Theor. Biol. 89 (1981) 581.
21 K.C. Chou and W.M. Liu, J. Theor. Biol. 91 (1981) 637.

22 B.N. Goldstein, J. Theor. Biol. 103 (1983) 247.
23 G.P. Zhou and M.H. Deng, Biochem. J. 222 (1984) 169.
24 K.C. Chou, J. Biol. Chem. 264 (1989) 12074.
25 W.W. Cleland, Biochim. Biophys. Acta 67 (1963) 104.
26 D.P. Goldenberg, J. Cell. Biochem. 29 (1985) 321.
27 J. Botts and M. Morales, Trans. Faraday Soc. 49 (1953) 696.
28 I.G. Darvey, J. Theor. Biol. 19 (1968) 215.
29 J. Galvez and R. Varon, J. Theor. Biol. 89 (1981) 1.
30 L. Michaelis and M.L. Menten, Biochem. Z. 49 (1913) 49.
31 W.H. Beyer, CRC handbook of mathematical sciences (CRC Press, Boca Raton, FL, 1988) 6th edn., p. 544.
32 F. Harary, Graph theory (Addison-Wesley, New York, 1969) p. 203.