J. theor. Biol. (1978) **70,** 199-212

A Stochastic Model for a Closed Biochemical System at Equilibrium†

SANDRA HASSTEDT‡

Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, *U.S.A.*

(*Received* 19 *July* 1976, *and in revised form* 27 *June* 1977)

This paper presents a stochastic model, a theoretical multi-variate probability function describing concentrations of reactants in a closed biochemical system at equilibrium. The theory applies to the complete range of biochemical systems from single enzyme reactions to combinations of reactions to complete pathways. Prior to examining the general system, probability functions are derived for the following systems as examples: a reaction with a competitive inhibitor, a bisubstrate reaction using the ping-pong mechanism and a series of two mono-substrate reactions. The theory of Markov processes is used to derive the probability functions for each of the example systems and then for the general system which includes the example systems as special cases. The probability function for any appropriate biochemical system proves to be the product of independent Poisson probabilities conditioned on the conservation equations. Finally, the implications of the theory are briefly discussed and possible extensions proposed.

1. Introduction

Previous stochastic treatments of closed biochemical systems consider only one type of enzyme reaction, the reversible Michaelis–Menten mechanism defined by the model:

$$S+E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P+E.$$

S, E, C and P represent free substrate, free enzyme, enzyme-substrate (or enzyme-product) transition complex and free product, respectively. Each of the earlier considerations of this reaction concerns deriving a probability generating function (pgf) in which the random variables represent the

† Supported by NIH training grant 5T01-GM-711.

‡ Present address: Department of Medical Biophysics and Computing, University of Utah Medical Center, Salt Lake City, Utah 84143, U.S.A.

concentrations of these reactants and k_1, k_2, k_3 and k_4, the rate constants for the individual steps of the reaction, serve as the parameters. The derivations employ the theory of Markov processes (Feller, 1968).

The initial (unsuccessful) attempts (Bartholomay, 1962*a*,*b*, 1964; Jachimowski, McQuarrie & Russell, 1964) to derive a pgf consider the complete time course of the reaction when $k_4 = 0$. Later investigators are more successful as a result of restricting the phase of the reaction. Specifically, the cases considered are the initial phase when $k_4 = 0$ (Heyde & Heyde, 1969), the equilibrium phase when $k_4 = k_1$ (Darvey & Staff, 1967) and the equilibrium phase without any restrictions on k_4 (Staff, 1970). A similar model, but for chemical reactions, is successfully derived to describe the number of conversions in a reaction (Orriss, 1969).

This consideration differs from previous ones in not being limited to the Michaelis–Menten mechanism. It treats a wide range of additional biochemical systems in the equilibrium phase, to extend Staff's results. Staff, however, solves a differential equation for the pgf, whereas here a system of linear equations is solved to find the probability function. The derivation of the linear equations and their solution for each of three example systems, chosen to demonstrate various features found in biochemical systems, is presented in the next three sections. These example systems are:

(i) A reaction with a competitive inhibitor.

(ii) A bisubstrate reaction using the ping-pong mechanism.

(iii) A series of two mono-substrate reactions. (Mahler & Cordes (1971) list and describe the mechanisms of these and other systems.) With the insight gained from these example systems' derivations, the general case equations are solved.

2. A Reaction with a Competitive Inhibitor

Consider a reversible Michaelis–Menten mechanism with a competitive inhibitor defined by the model:

$$S+E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P+E \qquad I+E \underset{k_6}{\overset{k_5}{\rightleftharpoons}} D. \tag{1}$$

S, E, C, P, I and D represent free substrate, free enzyme, enzyme-substrate (or enzyme-product) transition complex, free product, inhibitor and enzyme-inhibitor complex, respectively. The rate constants for the individual steps of the reaction are given by $k_j : j = 1,2,\ldots,6$. The conservation equations,

$$\begin{aligned} s+p+c &= S_0, \\ e+c+d &= E_0, \\ i+d &= I_0, \end{aligned} \tag{2}$$

complete the definition of the system. s, e, c, p, i and d symbolize concentrations of reactants designated by the corresponding large letters: S_0, E_0 and I_0 are initial concentrations of substrate, enzyme and inhibitor (assuming the initial concentrations of product and the two complexes are zero).

In order to derive the probability function for the reactants in system (1) the following assumptions are made.

(i) s, e, c, p, i and d take on integer values only (reactant concentrations are expressed in molecules).

(ii) The future state of the system depends only on the present state and not on the history of the system (the Markov property).

(iii) The probability of a conversion from one set of reactants to another is proportional to the concentrations of the reactants making the conversion.

(iv) The probability of more than one conversion in a time unit Δt is ${}_0(\Delta t)$ where:

$$ {}_0(\Delta t)/\Delta t \to 0, $$

as

$$ \Delta t \to 0. $$

Each arrow in system (1) represents a conversion from one set of reactants to another and therefore from one state (s, e, c, p, i, d) to another $(s', e', c' p', i', d')$. Using the assumptions made above, all possible conversions between states and their respective probabilities can be expressed as follows:

Conversions	Probabilities
$(s,e,c,p,i,d) \to (s-1,e-1,c+1,p,i,d)$	$k_1 se\Delta t + {}_0(\Delta t)$
$(s,e,c,p,i,d) \to (s+1,e+1,c-1,p,i,d)$	$k_2 c\Delta t + {}_0(\Delta t)$
$(s,e,c,p,i,d) \to (s,e+1,c-1,p+1,i,d)$	$k_3 c\Delta t + {}_0(\Delta t)$
$(s,e,c,p,i,d) \to (s,e-1,c+1,p-1,i,d)$	$k_4 pe\Delta t + {}_0(\Delta t)$ (3)
$(s,e,c,p,i,d) \to (s,e-1,c,p,i-1,d+1)$	$k_5 ie\Delta t + {}_0(\Delta t)$
$(s,e,c,p,i,d) \to (s,e+1,c,p,i+1,d-1)$	$k_6 d\Delta t + {}_0(\Delta t)$
$(s,e,c,p,i,d) \to (s,e,c,p,i,d)$	$1-(k_1 se+k_2 c+k_3 c+$ $k_4 pe+k_5 ie+k_6 d)\Delta t+{}_0(\Delta t)$

From the conservation equations (2),

$$ \begin{aligned} c &= S_0 - s - p, \\ d &= I_0 - i, \\ e &= E_0 - S_0 - I_0 + s + p + i. \end{aligned} \tag{4} $$

Therefore only s, p and i are needed to describe the state of the system. Rewriting expression (3) by making each conversion to the state (s, p, i) and replacing c, d and e from equations (4) yields:

Conversions	Probabilities
$(s+1,p,i) \to (s,p,i)$	$k_1(s+1)(E_0-S_0-I_0+s+p+i+1)\Delta t+{}_0(\Delta t)$
$(s-1,p,i) \to (s,p,i)$	$k_2(S_0-s-p+1)\Delta t+{}_0(\Delta t)$
$(s,p-1,i) \to (s,p,i)$	$k_3(S_0-s-p+1)\Delta t+{}_0(\Delta t)$
$(s,p+1,i) \to (s,p,i)$	$k_4(p+1)(E_0-S_0-I_0+s+p+i+1)\Delta t+{}_0(\Delta t)$
$(s,p,i+1) \to (s,p,i)$	$k_5(i+1)(E_0-S_0-I_0+s+p+i+1)\Delta t+{}_0(\Delta t)$
$(s,p,i-1) \to (s,p,i)$	$k_6(I_0-i+1)\Delta t+{}_0(\Delta t)$ (5)
$(s,p,i) \to (s,p,i)$	$1-[(k_1s+k_4p+k_5i)(E_0-S_0-I_0+s+p+i)$
	$+(k_2+k_3)(S_0-s-p)+k_6(I_0-i)]\Delta t+{}_0(\Delta t)$

This list includes all possible ways of attaining the state (s, p, i) in a time unit Δt. Therefore it is possible to write the probability of a state (s, p, i) at time $t+\Delta t$ in terms of the probabilities of all the origin states at time t as:

$$
\begin{aligned}
&P(s,p,i;\, t+\Delta t)\\
&\quad = \{1-[(k_1s+k_4p+k_5i)(E_0-S_0-I_0+s+p+i)\\
&\qquad +(k_2+k_3)(S_0-s-p)+k_6(I_0-i)]\Delta t\}P(s,p,i;\, t)\\
&\qquad +(1-\delta_{s,S_0})k_1(s+1)(E_0-S_0-I_0+s+p+i+1)\Delta t\ P(s+1,p,i;\, t)\\
&\qquad +(1-\delta_{s,0})k_2(S_0-s-p+1)\Delta t\ P(s-1,p,i;\, t)\\
&\qquad +(1-\delta_{p,0})k_3(S_0-s-p+1)\Delta t\ P(s,p-1,i;\, t) \qquad (6)\\
&\qquad +(1-\delta_{p,S_0})k_4(p+1)(E_0-S_0-I_0+s+p+i+1)\Delta t\ P(s,p+1,i;\, t)\\
&\qquad +(1-\delta_{i,I_0})k_5(i+1)(E_0-S_0-I_0+s+p+i+1)\Delta t\ P(s,p,i+1;\, t)\\
&\qquad +(1-\delta_{i,0})k_6(I_0-i+1)\Delta t\ P(s,p,i-1;\, t)+{}_0(\Delta t), \qquad (s,p,i)\in\Omega_1
\end{aligned}
$$

where:

$$
\begin{aligned}
\Omega_1 = \{&(s,p,i):\ s=0,1,\ldots,S_0,\ p=0,1,\ldots,S_0,\\
&i=0,1,\ldots,I_0,\ S_0-E_0\le s+p\le S_0,\\
&I_0-E_0\le i,\ S_0+I_0-E_0\le s+p+i\}
\end{aligned} \qquad (7)
$$

and $\delta_{m,n}$ is the Kronecker delta,

$$
\delta_{m,n}=\begin{cases}1 & \text{if } m=n,\\ 0 & \text{if } m\neq n.\end{cases} \qquad (8)
$$

Subtracting $P(s, p, i; t)$ from both sides, dividing by Δt and taking the limit

as $\Delta t \to 0$ yields the forward Kolmogorov equations:

$$\lim_{\Delta t \to 0} \frac{P(s, p, i; t+\Delta t) - P(s, p, i; t)}{\Delta t} = \frac{\partial P(s, p, i; t)}{\partial t}$$

$$\begin{aligned}
&= -[(k_1 s + k_4 p + k_5 i)(E_0 - S_0 - I_0 + s + p + i) + (k_2 + k_3)(S_0 - s - p) \\
&\quad + k_6(I_0 - i)]\, P(s, p, i; t) \\
&\quad + (1 - \delta_{s,S_0}) k_1 (s+1)(E_0 - S_0 - I_0 + s + p + i + 1)\, P(s+1, p, i; t) \\
&\quad + (1 - \delta_{s,0}) k_2 (S_0 - s - p + 1)\, P(s-1, p, i; t) \\
&\quad + (1 - \delta_{p,0}) k_3 (S_0 - s - p + 1)\, P(s, p-1, i; t) \\
&\quad + (1 - \delta_{p,S_0}) k_4 (p+1)(E_0 - S_0 - I_0 + s + p + i + 1)\, P(s, p+1, i; t) \\
&\quad + (1 - \delta_{i,I_0}) k_5 (i+1)(E_0 - S_0 - I_0 + s + p + i + 1)\, P(s, p, i+1; t) \\
&\quad + (1 - \delta_{i,0}) k_6 (I_0 - i + 1)\, P(s, p, i-1; t), \qquad (s, p, i) \in \Omega_1.
\end{aligned} \tag{9}$$

To this point the standard Markov procedure used by other investigators has been followed. Their next step would be to transform the system of forward Kolmogorov equations into a differential equation in the probability generating function. In making this transformation, probabilities containing a product of two concentrations, which always occur in biochemical systems, cause this partial differential equation to be second order. Therefore, solution is difficult and an alternative method is used here.

The equilibrium phase of the reaction is being considered ($t = \infty$), so the derivative of the probability with respect to t is zero. Letting $\partial P(s,p,i;t)/\partial t = 0$ in expression (9) leaves a system of linear equations in the unknown probabilities:

$$\begin{aligned}
&[(k_1 s + k_4 p + k_5 i)(E_0 - S_0 - I_0 + s + p + i) + (k_2 + k_3)(S_0 - s - p) \\
&\quad + k_6(I_0 - i)]\, P(s, p, i) \\
&= (1 - \delta_{s,S_0}) k_1 (s+1)(E_0 - S_0 - I_0 + s + p + i + 1)\, P(s+1, p, i) \\
&\quad + (1 - \delta_{s,0}) k_2 (S_0 - s - p + 1)\, P(s-1, p, i) \\
&\quad + (1 - \delta_{p,0}) k_3 (S_0 - s - p + 1)\, P(s, p-1, i) \\
&\quad + (1 - \delta_{p,S_0}) k_4 (p+1)(E_0 - S_0 - I_0 + s + p + i + 1)\, P(s, p+1, i) \\
&\quad + (1 - \delta_{i,I_0}) k_5 (i+1)(E_0 - S_0 - I_0 + s + p + i + 1)\, P(s, p, i+1) \\
&\quad + (1 - \delta_{i,0}) k_6 (I_0 - i + 1)\, P(s, p, i-1), \qquad (s, p, i) \in \Omega_1,
\end{aligned} \tag{10}$$

where Ω_1 is given in expression (7) and t no longer enters into the probabilities.

The solution to the system of linear equations in equation (10) is a probability function for which the random variables are the concentrations of the reactants in system (1). Since there is an equation for each probability, there are the same number of equations as unknowns; there is one dependency in the system of equations but the requirement that the probabilities sum to

one compensates for it to yield a unique solution given by:

$$P(s, p, i) = \frac{1}{\alpha}\frac{(k_2/k_1)^s(k_3/k_4)^p(k_6/k_5)^i}{s!p!i!(I_0-i)!(S_0-s-p)!(E_0-S_0-I_0+s+p+i)!}, \quad (s, p, i) \in \Omega_1 \tag{11}$$

where Ω_1 is given in expression (7) and α is a normalizing constant.

3. A Bisubstrate Reaction using the Ping-pong Mechanism

Now consider a bisubstrate reaction operating by the ping-pong mechanism defined by the model:

$$S+E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C_1 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P+E' \qquad T+E' \underset{k_6}{\overset{k_5}{\rightleftharpoons}} C_2 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} Q+E. \tag{12}$$

Two products, P and Q, are formed from two substrates, S and T. The obligatory first substrate, S, combines with the enzyme, E, to form a complex, C_1; then product, P, and an altered enzyme, E', are released. The second substrate, T, combines with E' to form complex C_2; then product, Q, and the original enzyme, E, are released. Representing the concentrations of the reactants by the corresponding small letters, the conservation equations are:

$$\begin{aligned} s+c_1+p &= S_0, \qquad t+c_2+q = T_0, \\ e+e'+c_1+c_2 &= E_0, \qquad s+q-e = S_0-E_0, \end{aligned} \tag{13}$$

where S_0, T_0 and E_0 are initial concentrations of S, T and E, respectively (assuming the initial concentrations of P, Q, E', C_1 and C_2 equal zero). With the addition of these four equations to the definition of system (12), only four random variables are needed to describe the state of the system; s, p, t and q are used.

Making the same assumptions as for the inhibited reaction already discussed, the conversions between states and the probabilities of each occurring are:

Conversion	Probability
$(s+1, p, t, q) \to (s, p, t, q)$	$k_1(s+1)(e+1)\Delta t+{}_0(\Delta t)$
$(s-1, p, t, q) \to (s, p, t, q)$	$k_2(c_1+1)\Delta t+{}_0(\Delta t)$
$(s, p-1, t, q) \to (s, p, t, q)$	$k_3(c_1+1)\Delta t+{}_0(\Delta t)$
$(s, p+1, t, q) \to (s, p, t, q)$	$k_4(p+1)(e'+1)\Delta t+{}_0(\Delta t)$
$(s, p, t+1, q) \to (s, p, t, q)$	$k_5(t+1)(e'+1)\Delta t+{}_0(\Delta t)$
$(s, p, t-1, q) \to (s, p, t, q)$	$k_6(c_2+1)\Delta t+{}_0(\Delta t)$
$(s, p, t, q-1) \to (s, p, t, q)$	$k_7(c_2+1)\Delta t+{}_0(\Delta t)$
$(s, p, t, q+1) \to (s, p, t, q)$	$k_8(q+1)(e+1)\Delta t+{}_0(\Delta t)$
$(s, p, t, q) \to (s, p, t, q)$	$1-(k_1se+k_2c_1+k_3c_1+k_4pe'+k_5te'+k_6c_2+k_7c_2+k_8qe)\Delta t+{}_0(\Delta t)$

(14)

where e, e', c_1 and c_2 are expressed in terms of the random variables, s, p, t and q, from equations (13) as:

$$\begin{aligned} e &= E_0 - S_0 + s + q, \qquad e' = p + t - T_0, \\ c_1 &= S_0 - s - p, \qquad c_2 = T_0 - t - q. \end{aligned} \tag{15}$$

Following the same procedure as before, the system of linear equations arrived at is:

$$\begin{aligned} &(k_1 se + k_2 c_1 + k_3 c_1 + k_4 pe' + k_5 te' + k_6 c_2 + k_7 c_2 + k_8 qe)\, P(s, p, t, q) \\ &\quad = (1 - \delta_{s,S_0}) k_1 (s+1)(e+1) P(s+1, p, t, q) \\ &\qquad + (1 - \delta_{s,0}) k_2 (c_2 + 1)\, P(s-1, p, t, q) \\ &\qquad + (1 - \delta_{p,0}) k_3 (c_1 + 1)\, P(s, p-1, t, q) \\ &\qquad + (1 - \delta_{p,S_0}) k_4 (p+1)(e'+1)\, P(s, p+1, t, q) \\ &\qquad + (1 - \delta_{t,T_0}) k_5 (t+1)(e'+1)\, P(s, p, t+1, q) \\ &\qquad + (1 - \delta_{t,0}) k_6 (c_2 + 1)\, P(s, p, t-1, q) \\ &\qquad + (1 - \delta_{q,0}) k_7 (c_2 + 1)\, P(s, p, t, q-1) \\ &\qquad + (1 - \delta_{q,T_0}) k_8 (q+1)(e+1)\, P(s, p, t, q+1), \\ &\qquad\qquad (s, p, q, t) \in \Omega_2, \end{aligned} \tag{16}$$

where:

$$\begin{aligned} \Omega_2 = \{&(s, p, t, q): s = 0, 1, \ldots, S_0,\ p = 0, 1, \ldots, S_0, \\ &t = 0, 1, \ldots, T_0,\ q = 0, 1, \ldots, T_0,\ T_0 \le p + t, \\ &S_0 - E_0 \le s + p \le S_0,\ T_0 - E_0 \le t + q \le T_0,\ S_0 - E_0 \le s + q\} \end{aligned} \tag{17}$$

and $\delta_{m,n}$ is given by expression (8).

The solution to this system of equations can be shown to be:

$$P(s, p, t, q) = \frac{1}{\alpha} \frac{(k_2/k_1)^s (k_3/k_4)^p (k_6/k_5)^t (k_7/k_8)^q}{s!\,p!\,t!\,q!\,e!\,e'!\,c_1!\,c_2!}, \qquad (s, p, t, q) \in \Omega_2 \tag{18}$$

where Ω_2 is given in expression (17), α is a normalizing constant and e, e', c_1 and c_2 are given in equations (15).

4. A Series of Two Enzyme Reactions

The third example biochemical system considered here is a series of two reversible Michaelis–Menten mechanisms defined by the model:

$$S + E_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C_1 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P + E_1, \qquad P + E_2 \underset{k_6}{\overset{k_5}{\rightleftharpoons}} C_2 \underset{k_8}{\overset{k_7}{\rightleftharpoons}} Q + E_2. \tag{19}$$

The product of the first reaction serves as the substrate of the second reaction to form a short pathway. S, P and Q represent the intermediates in the

pathway; E_1 and E_2 symbolize enzymes for the two reactions and C_1 and C_2 are complexes containing E_1 and E_2, respectively. Letting the corresponding small letters represent concentrations of the reactants in equations (19), the conservation equations are:

$$s+p+q+c_1+c_2 = S_0, \qquad e_1+c_1 = E_{01}, \qquad e_2+c_2 = E_{02}, \tag{20}$$

where S_0, E_{01} and E_{02} are initial concentrations of S, E_1 and E_2 (assuming the initial concentrations of P, Q, C_1 and C_2 are zero). These equations reduce the number of variables needed to describe the state of the system to four, s, p, q and e_1 are used.

Making the same assumptions and following the same procedure as before, the system of linear equations becomes:

$$\begin{aligned}
&(k_1 s e_1 + k_2 c_1 + k_3 c_1 + k_4 p e_1 + k_5 p e_2 + k_6 c_2 + k_7 c_2 + k_8 q e_2)\, P(s, p, q, e_1) \\
&\quad = (1-\delta_{s,S_0})(1-\delta_{e_1,E_{01}}) k_1 (s+1)(e_1+1)\, P(s+1, p, q, e_1+1) \\
&\qquad +(1-\delta_{s,0})(1-\delta_{e_1,0}) k_2 (c_1+1)\, P(s-1, p, q, e_1-1) \\
&\qquad +(1-\delta_{p,0})(1-\delta_{e_1,0}) k_3 (c_1+1)\, P(s, p-1, q, e_1-1) \\
&\qquad +(1-\delta_{p,S_0})(1-\delta_{e_1,E_{01}}) k_4 (p+1)(e_1+1)\, P(s, p+1, q, e_1+1) \\
&\qquad +(1-\delta_{p,S_0}) k_5 (p+1)(e_2+1)\, P(s, p+1, q, e_1) \\
&\qquad +(1-\delta_{p,0}) k_6 (c_2+1)\, P(s, p-1, q, e_1) \\
&\qquad +(1-\delta_{q,0}) k_7 (c_2+1)\, P(s, p, q-1, e_1) \\
&\qquad +(1-\delta_{q,S_0}) k_8 (q+1)(e_2+1)\, P(s, p, q+1, e_1), \qquad (s, p, q, e_1) \in \Omega_3.
\end{aligned} \tag{21}$$

where:

$$\begin{aligned}
\Omega_3 = \{&(s, p, q, e_1): s = 0, 1, \ldots, S_0, p = 0, 1, \ldots, S_0, \\
&q = 0, 1, \ldots, S_0, e_1 = 0, 1, \ldots, E_{01}, \\
&S_0 - E_{01} - E_{02} \le s+p+q \le S_0, E_{01} - S_0 \le e_1\},
\end{aligned} \tag{22}$$

$\delta_{m,n}$ is the Kronecker delta given in expression (8) and c_1, c_2 and e_2 are given by the random variables from equations (20) as:

$$\begin{aligned}
c_1 = E_{01} - e_1,\ c_2 = S_0 - E_{01} - s - p - q + e_1, \\
e_2 = E_{01} + E_{02} - S_0 + s + p + q - e_1.
\end{aligned} \tag{23}$$

The solution to this system of equations is:

$$P(s, p, q, e_1) = \frac{1}{\alpha} \frac{(k_2/k_1)^s (k_4/k_3)^{s+e_1} (k_6/k_5)^{s+e_1+p} (k_7/k_8)^q}{s!\,p!\,q!\,e_1!\,e_2!\,c_1!\,c_2!}, \qquad (s, p, q, e_1) \in \Omega_3, \tag{24}$$

where Ω_3 is given in expression (22), c_1, c_2 and e_2 are given in equations (23) and α is a normalizing constant.

It is convenient at this point to rewrite the equations in another form. Let:

$$\mathbf{x}' = (s, e_1, p, q, c_1, e_2, c_2) = (x_1, \ldots, x_7), \tag{25}$$

the reactants,

$$\mathbf{x}_1' = (s, e_1, p, q), \tag{26}$$

the random variables,

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \cdots \\ \mathbf{A}_2 \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \cdots \\ \mathbf{B}_2 \end{bmatrix}. \tag{27}$$

A and **B** are formed by placing a 1 in row i and column j if reactant x_i is converted to another reactant in step j; **A** specifies conversions in the forward direction and **B** in the reverse direction. Using this notation, equation (21) can now be rewritten as:

$$\begin{aligned} &\sum_{i=1}^{4} \left(k_{2j-1} \prod_{i=1}^{7} (x_i^{a_{ij}}) + k_{2j} \prod_{i=1}^{7} (x_i^{b_{ij}}) \right) P(\mathbf{x}_1) \\ &= \sum_{j=1}^{4} (1 - \delta_{\mathbf{x}_1 + \mathbf{A}_{1j} - \mathbf{B}_{1j}, \Omega_3}) k_{2j-1} \prod_{i=1}^{7} [(\mathbf{x}_i + 1)^{a_{ij}}] \, P(\mathbf{x}_1 + \mathbf{A}_{1j} - \mathbf{B}_{1j}) \\ &\quad + \sum_{j=1}^{4} (1 - \delta_{\mathbf{x}_1 - \mathbf{A}_{1j} + \mathbf{B}_{1j}, \Omega_3}) k_{2j} \prod_{i=1}^{7} [(x_i + 1)^{b_{ij}}] \, P(\mathbf{x}_1 - \mathbf{A}_{1j} + \mathbf{B}_{1j}), \\ &\qquad\qquad \mathbf{x}_1 \in \Omega_3, \end{aligned} \tag{28}$$

where Ω_3 is given in expression (22), $\mathbf{A}_{1j}$ and $\mathbf{B}_{1j}$ are columns of $\mathbf{A}_1$ and $\mathbf{B}_1$, a_{ij} and b_{ij} are elements of **A** and **B** and $\delta_{m,n}$ is now:

$$\delta_{m,n} = \begin{cases} 0 & m \in n, \\ 1 & m \notin n. \end{cases} \tag{29}$$

Also using the new notation, the probability function (24) is:

$$P(\mathbf{x}_1) = \frac{1}{\alpha} \frac{\prod_{j=1}^{4} \left(\frac{k_{2j-1}}{k_{2j}} \right)^{g_j(\mathbf{x}_1)}}{\prod_{i=1}^{7} x_i!}, \qquad \mathbf{x}_1 \in \Omega_3, \tag{30}$$

where:

$$\mathbf{g}(\mathbf{x}_1) = (\mathbf{B}_1 - \mathbf{A}_1)^{-1} \mathbf{x}_1, \tag{31}$$

and α is a normalizing constant.

5. A General Biochemical System

The probability function for a general biochemical system results from following the same procedures used for the example systems. Instead of a model of mechanism(s), an $m \times 1$ vector, $\mathbf{x}$, of reactants and corresponding $m \times n$ matrices, $\mathbf{A}$ and $\mathbf{B}$, of conversions specify this system of n steps with m reactants. Each of the n steps proceeds at forward and reverse rates, k_{+j} and k_{-j}, $j = 1,2,\ldots, n$. A set of $m-n$ conservation equations restrict the concentration ranges of the reactants.

The elements of $\mathbf{x}$ represent the reactants in the system; the first n elements are the random variables. $\mathbf{A}$ and $\mathbf{B}$ are formed by placing a 1 in row i and column j if x_i undergoes a conversion to another reactant in step j and a 0 otherwise; $\mathbf{A}$ specifies conversions in the forward direction and $\mathbf{B}$ in the reverse direction. The $m-n$ conservation equations are linear combinations of reactants for which the same linear combinations of rows of $\mathbf{B}-\mathbf{A}$ produce rows of zeros. Each of $\mathbf{x}$, $\mathbf{A}$ and $\mathbf{B}$ may be partitioned into separate matrices:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \cdots \\ \mathbf{x}_2 \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \cdots \\ \mathbf{A}_2 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \cdots \\ \mathbf{B}_2 \end{bmatrix}$$

where $\mathbf{x}_1$ is $n \times 1$ and $\mathbf{A}_1$ and $\mathbf{B}_1$ are each $n \times n$. For the theory to hold, the system must be such that $m \geqslant n$ (at least as many reactants as steps), $\mathbf{B}_1 - \mathbf{A}_1$ is non-singular and k_{+j} and k_{-j} are non-zero for $j = 1,2,\ldots, n$.

Making the same assumptions as before, the conversions between states and the probabilities of each occurring for a general system are:

Conversions	Probabilities	
$(\mathbf{x}_1 + \mathbf{A}_{1j} - \mathbf{B}_{1j})' \to \mathbf{x}_1'$	$k_{+j} \prod_{i=1}^{m} (x_i+1)^{a_{ij}} \Delta t + {}_0(\Delta t)$,	$j = 1, 2, \ldots, n$
$(\mathbf{x}_1 - \mathbf{A}_{1j} + \mathbf{B}_{1j})' \to \mathbf{x}_1'$	$k_{-j} \prod_{i=1}^{m} (x_i+1)^{b_{ij}} \Delta t + {}_0(\Delta t)$,	$j = 1, 2, \ldots, n$
$\mathbf{x}_1' \to \mathbf{x}_1'$	$1 - \sum_{j=1}^{n} \left(k_{+j} \prod_{i=1}^{m} x_i^{a_{ij}} + k_{-j} \prod_{i=1}^{m} x_i^{b_{ij}} \right) \Delta t + {}_0(\Delta t)$	(32)

where $\mathbf{A}_{1j}$ and $\mathbf{B}_{1j}$ are columns of $\mathbf{A}_1$ and $\mathbf{B}_1$ and a_{ij} and b_{ij} are elements of $\mathbf{A}$ and $\mathbf{B}$.

Using the forward Kolmogorov equations and setting the time derivative to zero (as for the example systems), the system of linear equations reduces

to:

$$\sum_{j=1}^{n}\left(k_{+j}\prod_{i=1}^{m}x_i^{a_{ij}}+k_{-j}\prod_{i=1}^{m}x_i^{b_{ij}}\right)P(\mathbf{x}_1)$$
$$=\sum_{j=1}^{n}(1-\delta_{\mathbf{x}_1+\mathbf{A}_{1j}-\mathbf{B}_{1j},\Omega})k_{+j}\prod_{i=1}^{m}(x_i+1)^{a_{ij}}P(\mathbf{x}_1+\mathbf{A}_{1j}-\mathbf{B}_{1j}) \tag{33}$$
$$+\sum_{j=1}^{n}(1-\delta_{\mathbf{x}_1-\mathbf{A}_{1j}+\mathbf{B}_{1j},\Omega})k_{-j}\prod_{i=1}^{m}(x_i+1)^{b_{ij}}P(\mathbf{x}_1-\mathbf{A}_{1j}+\mathbf{B}_{1j}),\quad \mathbf{x}_1\in\Omega,$$

where Ω is determined by the conservation equations and initial conditions, $\mathbf{A}_{1i}$, $\mathbf{B}_{1j}$, a_{ij} and b_{ij} are defined above and $\delta_{m,n}$ is given in expression (29). The first example system derivation considers the uniqueness of the function.

Equation (33) is satisfied if each of the $2n$ terms on the left side equals a different term on the right side; i.e. if:

$$k_{+j}\prod_{i=1}^{m}x_i^{a_{ij}}P(\mathbf{x}_1)=k_{-j}\prod_{i=1}^{m}(x_i+1)^{b_{ij}}P(\mathbf{x}_1-\mathbf{A}_{1j}-\mathbf{B}_{1j}), \tag{34}$$
$$j=1,2,\ldots,n,\qquad \mathbf{x}_1,\ \mathbf{x}_1-\mathbf{A}_{1j}+\mathbf{B}_{1j}\in\Omega.$$

(The other n terms yield identical equations.) Assuming $P(\mathbf{x}_1)\neq 0$ and $k_{-j}\neq 0$,

$$\frac{P(\mathbf{x}_1-\mathbf{A}_{1j}+\mathbf{B}_{1j})}{P(\mathbf{x}_1)}=\prod_{i=1}^{m}\frac{x_i^{a_{ij}}}{(x_i+1)^{b_{ij}}}\frac{k_{+j}}{k_{-j}},\qquad j=1,2,\ldots,n. \tag{35}$$

For some j, when $b_{ij}=1$, x_i+1 occurs in the denominator on the right in this ratio and in the argument of the numerator probability function on the left; when $a_{ij}=1$, x_i occurs in the numerator of the ratio on the right and x_i-1 in the argument of the numerator probability function on the left. This occurs if the probability function contains

$$\left[\prod_{i=1}^{m}x_i!\right]^{-1},$$

so that:

$$P(\mathbf{x}_1)=\frac{1}{\alpha}\frac{\prod_{j=1}^{n}\left(\frac{k_{+j}}{k_{-j}}\right)^{g_j(\mathbf{x}_1)}}{\prod_{i=1}^{m}x_i!},\qquad \mathbf{x}_1\in\Omega, \tag{36}$$

where:

$$\begin{aligned}&g_j(\mathbf{x}_1-\mathbf{A}_{1j}+\mathbf{B}_{1j})-g_j(\mathbf{x}_1)=1\\&g_j(\mathbf{x}_1-\mathbf{A}_{1j'}+\mathbf{B}_{1j'})-g_j(\mathbf{x}_1)=0,\quad j'=1,2,\ldots,n,\quad j'\neq j.\end{aligned} \tag{37}$$

The vector $\mathbf{g}$ has this property if:

$$\mathbf{g}(\mathbf{x}_1)=(\mathbf{B}_1-\mathbf{A}_1)^{-1}\,\mathbf{x}_1. \tag{38}$$

The probability function given in equation (36) is the product of independent

Poisson functions conditioned on the conservation equations. The moments remain to be determined.

Considering the reversible Michaelis–Menten mechanism,

$$S+E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P+E, \tag{39}$$

as did Staff (1970),

$$\mathbf{x}=\begin{bmatrix} s \\ p \\ \dots \\ c \\ e \end{bmatrix}=\begin{bmatrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_2 \end{bmatrix}, \quad \mathbf{A}=\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \dots & \\ 0 & 1 \\ 1 & 0 \end{bmatrix}=\begin{bmatrix} \mathbf{A}_1 \\ \dots \\ \mathbf{A}_2 \end{bmatrix}, \quad \mathbf{B}=\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ \dots & \\ 1 & 0 \\ 0 & 1 \end{bmatrix}=\begin{bmatrix} \mathbf{B}_1 \\ \dots \\ \mathbf{B}_2 \end{bmatrix} \tag{40}$$

The conservation equations,

$$e+c=E_0+C_0, \qquad s+p+c=S_0+P_0+C_0, \tag{41}$$

result from consideration of $\mathbf{B}-\mathbf{A}$,

$$\mathbf{B}-\mathbf{A}=\begin{bmatrix} -1 & 0 \\ 0 & 1 \\ \dots & \\ 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{42}$$

since the sums of rows three and four and of rows one, two and three are rows of zeros. S_0, P_0, C_0 and E_0 are initial concentrations of s, p, c and e, respectively. Taking the inverse,

$$(\mathbf{B}_1-\mathbf{A}_1)^{-1}=\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{43}$$

Therefore,

$$\mathbf{g}(\mathbf{x}_1)=\begin{bmatrix} -s \\ p \end{bmatrix}. \tag{44}$$

Letting $P_0=C_0=0$ as did Staff, from equation (36),

$$P(s,p)=\frac{1}{\alpha}\frac{(k_1/k_2)^{-s}(k_3/k_4)^{p}}{s!p!c!e!}, \quad (s,p)\in\Omega, \tag{45}$$

where:

$$c=S_0-s-p, \qquad e=E_0-S_0+s+p, \tag{46}$$

and

$$\Omega=\{(s,p): s=0,1,\dots,S_0, p=0,1,\dots,S_0, S_0-E_0\le s+p\le S_0\} \tag{47}$$

in agreement with Staff. Since the means of s and p, derived from the probability function, are:

$$\mu_s = \frac{k_2}{k_1}\beta, \qquad \mu_p = \frac{k_3}{k_4}\beta, \tag{48}$$

where β is a constant, the ratio:

$$\frac{\mu_p}{\mu_s} = \frac{k_1 k_3}{k_2 k_4} = K_{eq}, \tag{49}$$

where K is the equilibrium constant for the reaction.

6. Discussion

The state of the art of analyzing biochemical systems would be greatly improved by the development of a comprehensive probability theory for both open and closed systems. Functions of rate constants could be estimated from *in vivo* measurements of concentrations of intermediate reactants, eliminating the need for the assumption that enzymes operate identically *in vitro* as *in vivo* (Walter, 1966, preface). A larger proportion of the total number of mutations that occur could be detected by statistical testing procedures applied to intermediate data than are detectable by electrophoresis. Nei (1975) estimates that 70% of all mutations result in amino acid changes while only 30% are detectable by electrophoresis. The source of altered enzyme activity resulting from a mutation could possibly be determined by testing for a change in the amount of enzyme, thereby simplifying an otherwise difficult problem (Harris, 1975). Regulatory enzymes could be located with statistical rigor, not obtainable using crossover plots (Rolleston, 1972) Finally, models of enzyme mechanisms could be tested against alternatives using *in vivo* data.

Analyses, which employ the probability function derived here, are limited to those that can use data from a closed (*in vitro*) system at equilibrium. Functions of rate constants, or equilibrium constants in this case, can be estimated. Both detecting mutant enzymes and testing models for enzyme mechanisms require that enzyme properties be reflected in the data. One would not expect this to be true at equilibrium since the equilibrium constant does not depend on the enzyme. The location of regulatory enzymes and the determination of enzyme level necessitate *in vivo* data. The probability function finds its main importance as the first step in the development of a more complete theory, one which includes open biochemical systems.

On the other hand, while the applications of this theory are limited, the systems to which it applies are numerous and varied. Three restrictions are required to develop the theory. Two of them qualify easily since most

reactions can be considered reversible and closed systems generally have as many reactants as steps. The third restriction, that $\mathbf{B}_1 - \mathbf{A}_1$ be non-singular, eliminates certain allosteric mechanisms, yet, in the cases tried, a solution proves possible if particular rate constants are equated. There is no limit to the length of the pathway, to the number of reactants or to the mechanisms assumed for the reactions as long as these three restrictions are met.

Consideration of the plausible uses and shortcomings of this theory leads to proposals for its extension. Measurement error will have to be incorporated for the function to be practical since the variation included in the present model may not nearly represent all that occurs in the data (Heyde & Heyde, 1971). The applications will increase with extension to open systems. Logically, the reactants in the model should have a continuous density rather than discrete probabilities, for that is the nature of their measurements. Including compartmentation (as in mitochondria or in different cells), variable enzyme levels (as with degradation), or more than one set of rate constants (as with heterozygotes), may add useful, though unessential, features.

I would like to thank Dr Peter Smouse, Dr Michael Savageau, Dr George Brewer, Dr Michael Woodroofe and especially Dr Charles Sing for reviewing and offering valuable, timely suggestions which improved this manuscript. I also wish to thank Dr Jeffrey Weisberg for proposing needed revisions to increase the clarity of the language.

REFERENCES

BARTHOLOMAY, A. F. (1962*a*). *Ann. N. Y. Acad. Sci.* **96,** 897.

BARTHOLOMAY, A. F. (1962*b*). *Biochem.* **1,** 223.

BARTHOLOMAY, A. F. (1964). In *Stochastic Models in Medicine and Biology* (J. Gurland, ed.), p. 101. Madison: University of Wisconsin Press.

DARVEY, I. G. & STAFF, P. J. (1967). *J. theor. Biol.* **14,** 157.

FELLER, W. (1968). *An Introduction to Probability Theory and its Application*, Vol. 1. New York: John Wiley & Sons.

HARRIS, H. (1975). *The Principles of Human Biochemical Genetics*. New York: American Elsevier Publishing Company, Inc.

HEYDE, C. C. & HEYDE, E. (1969). *J. theor. Biol.* **25,** 159.

HEYDE, C. C. & HEYDE, E. (1971). *J. theor. Biol.* **30,** 395.

JACHIMOWSKI, C. J., MCQUARRIE, D. A. & RUSSELL, M. E. (1964). *Biochem.* **3,** 1732.

MAHLER, H. R. & CORDES, E. H. (1971). *Biological Chemistry*. New York: Harper & Row.

NEI, M. (1975). *Molecular Population Genetics and Evolution*. New York: American Elsevier Publishing Company, Inc.

ORRISS, J. (1969). *J. appl. Prob.* **6,** 505.

ROLLESTON, F. S. (1972). In *Current Topics in Cellular Regulation* (B. L. Horecker & E. R. Stadtman, eds), Vol. 5, p. 47. New York: Academic Press.

STAFF, P. J. (1970). *J. theor. Biol.* **27,** 221.

WALTER, C. (1966). *Enzyme Kinetics: Open and Closed Systems*. New York: Ronald Press Company.