Reprinted from
Proc. Natl. Acad. Sci. USA
Vol. 75, No. 1, pp. 26–29, January 1978
Chemistry

Stochastic theory of ligand migration in biomolecules

(flash photolysis/myoglobin kinetics)

N. ALBERDING, H. FRAUENFELDER*, AND P. HÄNGGI

Department of Physics, University of Illinois at Urbana–Champaign, Urbana, Illinois 61801

Contributed by Hans Frauenfelder, November 3, 1977

ABSTRACT **When ligand binding to proteins involves the presence of more than one ligand inside a given biomolecule, linear deterministic rate equations become useless. A stochastic approach, however, permits a treatment of the migration and binding of small molecules to proteins even at high ligand concentrations. An appropriate linear master equation and its analytic solution are given. As an example, the binding of carbon monoxide to myoglobin at partial pressures from 1 to 10^3 bars (0.1 to 100 MPa) is treated.**

1. Deterministic and stochastic approach

The discovery that access to the binding site in heme proteins is governed by multiple barriers (1) leads to interesting problems in reaction kinetics. Consider, as an example, the binding of carbon monoxide to myoglobin (Fig. 1). The ligand CO may encounter, on its way from the solvent S to the binding site 1 at the heme iron, four potential barriers in succession.† Assume that the ligand concentration [CO] in the solvent S is so small that at any given time there is at most one ligand within any biomolecule. The binding kinetics can then be described by *deterministic linear rate equations*, of the form

$$\frac{dN_2(t)}{dt} = k_{12}N_1(t) - k_{21}N_2(t) - k_{23}N_2(t) + k_{32}N_3(t). \quad [1]$$

Here, $N_2(t)$, for instance, denotes the fraction of biomolecules with a CO molecule in well 2 at time t and k_{21}, the rate parameter for the step $2 \rightarrow 1$. The initial conditions depend on the experimental arrangement. In flash photolysis, all binding sites are initially occupied, all other wells are empty. The photolyzing light pulse breaks the bond between the heme iron and CO, and the ligand is promoted to well 2. The initial conditions for photodissociation thus are $N_2(0) = 1$, $N_1(0) = N_3(0) = N_4(0) = 0$. In a stopped-flow experiment, the ligand is initially in S and all other wells are empty.

The sequential model of Fig. 1 and deterministic linear rate equations adequately describe many experiments. However, a more powerful approach is needed for more complex situations. The number of wells can be larger than five, transitions may occur between any two wells, and many ligands may simultaneously occupy a given biomolecule. In myoglobin, for instance, binding at the iron is covalent; the first ligand that occupies well 1 blocks further transitions. The other wells, however, can very likely accept more than one ligand. These features call for a generalization of the treatment of migration and ligand binding, and we present here a stochastic approach.

In the *stochastic approach* (2–5) the system at time t is described by a set of stochastic variables $[x_1(t), x_2(t), \ldots, x_L(t), \ldots] \equiv \mathbf{x}(t)$. Here $x_L(t)$ denotes the number of ligands in well L at time t in a given biomolecule with $L = 1,2, \ldots L_{\max}$ ($L_{\max} = 4$ in Fig. 1). Each well in a given biomolecule can be occupied by at most $x_L^{\max}$ ligands. The total number of ligands in all wells is not fixed, and fluctuations can thus be handled. Blocking of well 1 by the first bound ligand is expressed by $x_1^{\max} = 1$. Blocking of other wells by a given number of ligands can also be stated by corresponding restrictions on $x_L^{\max}$. A probability $p(\mathbf{x};t,\lambda)$ is introduced for finding the system at time t in configuration $\mathbf{x}$; λ describes parameters such as temperature, pH, ligand concentration, and hydrostatic pressure. The time dependence of p is determined by a master equation. In the present paper we only sketch the essential steps; the full theory will be published elsewhere.‡

2. Experimental approach

In a typical experiment, ligand binding is monitored optically. The optical absorption spectra of a free heme protein and one with bound ligand differ. From the absorbance measured at a suitable wavelength, the fraction $N_{\exp}(t,\lambda)$ of biomolecules without bound ligand can therefore be determined. To compare $N_{\exp}(t,\lambda)$ with the result of a calculation, two features must be noted. First, migration in each individual biomolecule is a stochastic process, with large fluctuations. An experimental observation measures an average over a very large number of independent biomolecules; $N_{\exp}(t,\lambda)$ thus is an ensemble average and fluctuations in $N_{\exp}(t,\lambda)$ can be neglected (central limit theorem) (6). Second, the optical spectrum changes when the ligand binds covalently to the heme iron. In terms of Fig. 1, the bound state corresponds to one, the unbound, to zero ligands in well 1. Ligands in wells $2 \ldots L_{\max}$ have a negligible effect on the optical spectrum. We can thus make the identification

$$N_{\exp}(t,\lambda) = \sum_{x_2, \ldots x_{L_{\max}}} p(x_1 = 0, x_2, \ldots, x_{L_{\max}};t,\lambda). \quad [2]$$

The experimental binding data can be evaluated if the time development of the probability function $p(\mathbf{x};t,\lambda)$ is known.

3. Microscopic description

To arrive at a microscopic picture of ligand migration, we postulate sequential, locally stable, potential wells as in Fig. 1. The random variable $x_L(t)$ describes the occupation of well L at time t in a given biomolecule and also the actual value in the state-space $x_L = 0,1,2 \ldots$. The meaning of $x_L(t)$ will always be clear from the context. The stationary transition rate for a transition in which one ligand jumps from well L to well K is denoted by $\Gamma(x_K|x_L;\lambda)$,

$$(\ldots, x_L, \ldots, x_K - 1, \ldots) \xrightarrow{\Gamma(x_K|x_L;\lambda)} (\ldots, x_L - 1, \ldots, x_K, \ldots).$$

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* To whom reprint requests should be addressed.

† In ref. 1, we denoted wells 1 to 4 by A to D and the solvent by E. The notation used here makes the equations simpler.

‡ P. Hänggi, unpublished.

For the evaluation, we make a number of assumptions:

(*i*) We assume that the biomolecules are independent and do not influence each other.

(*ii*) We neglect transitions in which two or more ligands within the same biomolecule jump simultaneously.

(*iii*) We assume the transition rate $L \rightarrow K$ to be proportional to the number x_L initially present in well L and to depend on all other variables in the form

$$\Gamma(x_K|x_L;\lambda) = x_L \gamma_{KL}(x_K;\lambda). \quad [3]$$

The rate parameters γ_{KL} are in general still functions of the occupation number x_K of the final well K. In the innermost well ($K = 1$), the ligand binds covalently to the heme iron. This fact leads to two conditions:

(*iv*) The binding is so tight that well 1 acts as a trap; transitions from well 1 to all others can be neglected,

$$\gamma_{K1}(x_K;\lambda) \cong 0, \quad K = 2, \ldots L_{max}. \quad [4]$$

(*v*) The first ligand to occupy well 1 blocks further transitions so that

$$\Gamma(x_1|x_L;\lambda) = x_L \gamma_{1L}(\lambda)\delta_{x_1,1}. \quad [5]$$

(*vi*) We assume that the solvent acts as a bath with constant ligand concentration, $[S]$. For transitions involving S, Eq. 3 consequently becomes

$$\Gamma([S]|x_L;\lambda) = x_L \gamma_{SL}(\lambda), \quad [6]$$

$$\Gamma(x_L|[S];\lambda) = [S]\gamma'_{LS}(x_L;\lambda). \quad [7]$$

The prime on γ'_{LS} indicates that it is a second-order rate parameter. The binding site, $L = 1$, is assumed not to couple directly to the solvent:

$$\gamma_{S1} = \gamma'_{1S} = 0. \quad [8]$$

(*vii*) We further assume that all wells except 1 can accept an arbitrary number of ligands and that the relevant rate parameters are independent of the occupation number of the well. Eqs. 3 and 7 then become for $K,L \neq 1$

$$\Gamma(x_K|x_L;\lambda) = x_L \gamma_{KL}(\lambda) \quad [9]$$

$$\Gamma(x_L|[S];\lambda) = [S]\gamma'_{LS}(\lambda) \equiv \beta_L(\lambda). \quad [10]$$

In Eq. **10**, $[S]$ can be considered one of the parameters of the set λ.

(*viii*) The time development of the probability $p(x;t,\lambda)$ shall be Markovian. The transition rates then depend only on internal physical parameters and variables, but not on the initial preparation, $p(t_0)$, of the system (7).

(*ix*) All biomolecules are assumed to be identical and the potential, given for a special case in Fig. 1, shall be time independent.

In the specific example treated in *Section 4*, we add two more assumptions, namely

(*x*) Transitions occur only between neighboring wells.

(*xi*) In photodissociation, the flash breaks the bond between the heme iron and the ligand. At time $t = 0^+$, the ligand moves to well 2 without perturbing the other initial probabilities (sudden approximation), and is immediately thermalized.

Not all of these assumptions are based on equally firm grounds. *i*, *ii*, *iv*, and *v* can be well justified. Assumption *ix* is certainly too restrictive, as is known from experiment (1), and we will lift this restriction in another publication. Assumptions *vii*, *x*, and *xi* are most likely inadequate, but the restrictions on the occupation number of the various wells, alternate pathways, and the processes during photodissociation are not yet known

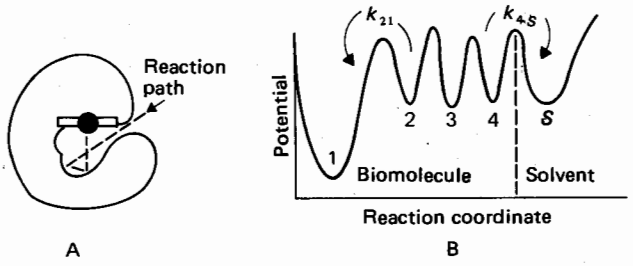

FIG. 1. Access to the binding site in a heme protein is governed by multiple barriers. (*A*) Reaction path; (*B*) potential along the reaction path.

and probably differ from biomolecule to biomolecule. The goal is to find an analytic expression for $N_{exp}(t,\lambda)$, with all assumptions valid. Comparison with experiment tests the assumptions and the theoretical treatment can then be generalized.

The nonlinear deterministic rate equations for the time evolution of $x(t,\lambda)$ follow from Eq. 3 as

$$\frac{d\langle x(t,\lambda)\rangle}{dt} = M\langle x(t,\lambda)\rangle + \beta \quad [11]$$

in which

$$\beta = (\beta_1, \ldots, \beta_{L_{max}}) \quad [12]$$

and in which the matrix M is defined by

$$M_{KL} = \gamma_{KL}(\langle x_K(t)\rangle), \quad K \neq L \quad [13]$$

$$M_{KK} = -\gamma_{SK} - \sum_{\substack{L=1 \\ L \neq K}}^{L_{max}} \gamma_{LK}(\langle x_L(t)\rangle). \quad [14]$$

The bracket, $\langle x_K(t)\rangle$, denotes the average occupation number in well K at time t. Eq. 11 is nonlinear. In the low concentration limit, $[S] \rightarrow 0$, at which each biomolecule is occupied by at most one ligand, the blocking of well 1 can be neglected. Eq. 11 then coincides with the linear rate equations of the type of Eq. 1 with $\gamma_{LK} \rightarrow k_{KL}$ and $\langle x_K(t)\rangle \rightarrow N_K(t)$.

Because it is difficult to describe the blocking of well 1 with a deterministic approach, we now turn to the stochastic treatment. If, as assumed in *viii*, migration of the ligands is Markovian, the rate of change of the probability $p(x;t,\lambda)$ obeys the linear master equation

$$\frac{\partial p(x)}{\partial t} = \sum_{L=2}^{L_{max}} \sum_{\substack{K=2 \\ L \neq K}}^{L_{max}} (x_L + 1)\gamma_{KL} p(x_L + 1, x_K - 1, x')$$

$$\sum_{L=2}^{L_{max}} (x_L + 1)\gamma_{SL} p(x_L + 1, x') + \sum_{L=2}^{L_{max}} \beta_L p(x_L - 1, x')$$

$$- \sum_{L=2}^{L_{max}} \sum_{\substack{K=2 \\ L \neq K}}^{L_{max}} x_L \gamma_{KL} p(x) - \sum_{L=2}^{L_{max}} x_L \gamma_{SL} p(x)$$

$$- \sum_{L=2}^{L_{max}} \beta_L p(x) + \sum_{L=2}^{L_{max}} (x_L + 1)\gamma_{1L}\delta_{x_1,1}\, p(x_L + 1, x_1 - 1, x')$$

$$- \sum_{L=2}^{L_{max}} x_L \gamma_{1L}\delta_{x_1,0}\, p(x). \quad [15]$$

Here, x' is the reduced vector derived from x by removal of the explicitly written components. For brevity, we have also omitted the parameters t and λ in γ and p. In order to solve for the set of probabilities obeying this master equation, a gener

ating function $G(y;t,\lambda)$ is introduced through the definition

$$G(y_1,\ldots,y_{L_{max}};t,\lambda) = \sum_{x_1=0}^{1}\sum_{x_2=0}^{\infty}\cdots\sum_{x_{L_{max}}=0}^{\infty} p(x;t,\lambda)\prod_{L=1}^{L_{max}} y_L{}^{x_L}. \quad [16]$$

The evolution equation for G then follows from Eq. **15** as

$$\frac{\partial G}{\partial t} = \sum_{L=2}^{L_{max}}\beta_L(y_L-1)G + \sum_{L=2}^{L_{max}}\gamma_{SL}(1-y_L)\frac{\partial G}{\partial y_L} + \sum_{L=2}^{L_{max}}\sum_{K=2}^{L_{max}}\gamma_{KL}(y_K-y_L)\frac{\partial G}{\partial y_L} + \sum_{x_1}\cdots\sum_{x_{L_{max}}}\sum_{L=2}^{L_{max}}\gamma_{1L}[(x_L+1)\delta_{x_1,1}\,p(x_L+1,x_1-1,x') - x_L\delta_{x_1,0}\,p(x)]\prod_{I=1}^{L_{max}} y_I{}^{x_I}. \quad [17]$$

Comparison of Eqs. **2** and **16** shows that

$$N_{exp}(t,\lambda) = G(0,1,\ldots 1;t,\lambda). \quad [18]$$

Eq. **17** can be solved analytically for $G(0,1,\ldots,1;t,\lambda)$; with the initial probability $p(x;t=0^+,\lambda)$, the desired expression becomes

$$N_{exp}(t,\lambda) = \exp\left\{-\|b\|^{-1}\sum_{L=2}^{L_{max}}\beta_L(\lambda) \times\left[\sum_{K=2}^{L_{max}}\mu_K{}^{-1}B^{KL}(\exp\mu_K t-1)b_{K1}+B^{1L}b_{11}t\right]\right\} \times\sum_{x_2=1}^{\infty}\cdots\sum_{x_{L_{max}}=1}^{\infty}p(0,x_2\ldots x_{L_{max}};t=0,\lambda) \times\prod_{L=2}^{L_{max}}\left\{1-\|b\|^{-1}\sum_{K=1}^{L_{max}}B^{KL}b_{K1}\exp\mu_K t\right\}^{x_L}. \quad [19]$$

Here, $b_K = (b_{K1},\ldots b_{KL_{max}})$ are the eigenvectors of the matrix M, with corresponding eigenvalues μ_K. The matrix M, Eqs. **13** and **14**, is defined with constant elements $\gamma_{KL}(\lambda)$ (assumption *vii*). B is the cofactor matrix of the matrix of eigenvectors, b; $\|b\|$ denotes the determinant of b.

4. An example—myoglobin

As application of Eq. **19**, we consider binding after photodissociation in Mb. Flash photolysis experiments have shown that at least four wells exist inside Mb, so that the potential encountered by CO is as given in Fig. 1 (1). At CO partial pressures of less than one bar (1 bar = 10^5 Pa), the probability of having more than one ligand inside a given protein is small and the use of the linear deterministic equations (1) is justified. At higher CO concentrations, however, these equations are no longer adequate, while the stochastic approach is still valid.

With the assumptions *i–xi*, and $L_{max} = 4$, M and β are given by

$$M = \begin{pmatrix} 0 & \gamma_{12} & 0 & 0 \\ 0 & -(\gamma_{12}+\gamma_{32}) & \gamma_{23} & 0 \\ 0 & \gamma_{32} & -(\gamma_{23}+\gamma_{43}) & \gamma_{34} \\ 0 & 0 & \gamma_{43} & -(\gamma_{34}+\gamma_{S4}) \end{pmatrix}, \qquad \beta = (0,0,0,\beta_4). \quad [20]$$

Before the photoflash, all wells are in thermal equilibrium with the solvent; because of the covalent binding, well 1 is fully occupied. We assume with *xi* that the flash moves the ligand from well 1 to 2, but leaves the other initial probabilities unperturbed (sudden approximation). At time $t = 0^+$, the initial probability

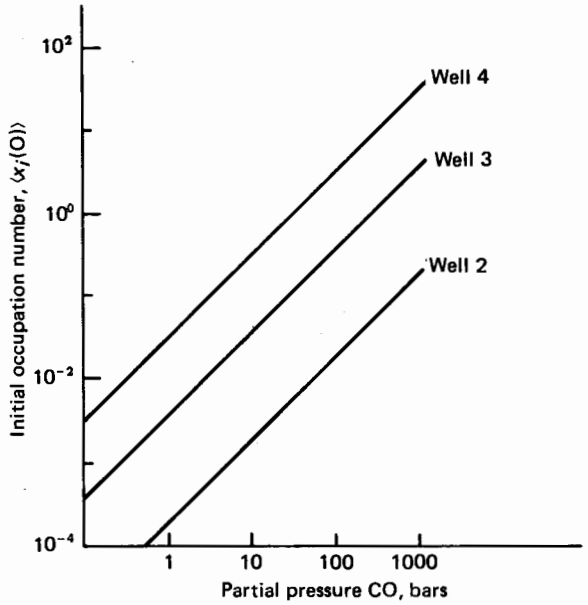

FIG. 2. Mean initial occupation of wells 2–4 at 230 K in Mb before photodissociation as function of the CO partial pressure.

therefore satisfies

$$p(1,x_2,x_3,x_4;0^+,\lambda) = 0,\ p(0,0,x_3,x_4;0^+,\lambda) = 0, \quad [21]$$

and the grand canonical initial probability becomes

$$p(x;0^+,\lambda) = \frac{\langle x_2(0,\lambda)\rangle^{x_2-1}}{(x_2-1)!}\exp(-\langle x_2(0,\lambda)\rangle) \times(1-\delta_{0,x_2})(1-\delta_{1,x_1})\frac{\langle x_3(0,\lambda)\rangle^{x_3}}{x_3!}\exp(-\langle x_3(0,\lambda)\rangle) \times\frac{\langle x_4(0,\lambda)\rangle^{x_4}}{x_4!}\exp(-\langle x_4(0,\lambda)\rangle). \quad [22]$$

The initial mean values x_2 to x_4 are determined by

$$\begin{pmatrix}\langle x_2(0,\lambda)\rangle\\ \langle x_3(0,\lambda)\rangle\\ \langle x_4(0,\lambda)\rangle\end{pmatrix} = -M_r{}^{-1}\begin{pmatrix}0\\0\\\beta_4\end{pmatrix}, \quad [23]$$

in which M_r denotes the matrix M without the first row and column and with $\gamma_{12} = 0$. With Eq. **19**, $N_{exp}(t,\lambda)$ now becomes

$$N_{exp}(t,\lambda) = \exp(-\langle x_1^*(t,\lambda)\rangle) \times\left\{1-\|b\|^{-1}\sum_{K=1}^{4}B^{K2}b_{K1}\exp\mu_K t\right\} \quad [24]$$

in which

$$\langle x_1^*(t,\lambda)\rangle = \|b\|^{-1}\beta_4\left[B^{14}b_{11}t+\sum_{K=2}^{4}\frac{B^{K4}}{\mu_K}(\exp\mu_K t-1)b_{K1}\right] + \sum_{K=2}^{4}\langle x_K(0,\lambda)\rangle\|b\|^{-1}\sum_{L=1}^{4}B^{LK}b_{L1}\exp\mu_K t. \quad [25]$$

The mean value $\langle x_1^*(t,\lambda)\rangle$ has a simple interpretation: If the recombination were to start with an initial occupation of zero at the binding site and the equilibrium Poisson distributions in wells 2–4 with mean values given by Eq. **23**, $\langle x_1^*(t,\lambda)\rangle$ would be the mean occupation number of site 1 if no blocking were to occur. In reality, the actual mean $\langle x_1(t,\lambda)\rangle$ has little in common with $\langle x_1^*(t,\lambda)\rangle$.

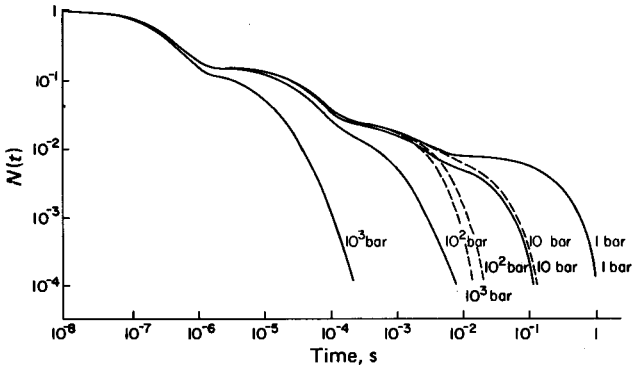

FIG. 3. Computer-simulated rebinding curves for MbCO. $N(t)$ after photodissociation at 230 K. The CO partial pressure increases from 1 to 10^3 bars. The solid lines are calculated with the stochastic model, the broken ones with deterministic linear rate equations. The top lines of labels in the figure refer to the stochastic calculation.

$N_{exp}(t,\lambda)$ depends, for a fixed value of λ, on the rate parameters γ and β. The rate parameters are determined by fitting Eq. 24 to the experimental data. Here we demonstrate how the deterministic and stochastic approaches give different results when the probability of finding more than one ligand inside a given biomolecule can no longer be neglected. As an example, we consider the rebinding of CO to Mb after photodissociation at 230 K, with rate parameters $\gamma_{12} = 2.8 \times 10^6\ s^{-1}$, $\gamma_{32} = 4.7 \times 10^5\ s^{-1}$, $\gamma_{23} = 2.3 \times 10^4\ s^{-1}$, $\gamma_{43} = 3.7 \times 10^3\ s^{-1}$, $\gamma_{34} = 4.9 \times 10^2\ s^{-1}$, $\gamma_{S4} = 1.9 \times 10^2\ s^{-1}$, $\beta_4 = [CO]\gamma'_{4S} = 5.5 \times 10^5\ p_{CO}\ s^{-1}$, in which p_{CO} is the CO partial pressure in bars (1). For clarity, we assume *ix* to be valid and neglect distributed barriers. The initial mean values $\langle x_i \rangle$, determined from Eq. 23, are given in Fig. 2 as a function of p_{CO}. The outermost well, 4, is appreciably populated above about 10 bars, the next one above 100 bars. The solubility of CO inside Mb therefore is considerably larger than in the solvent, and deviations from the linear deterministic rate equations should be expected already around 10 bars. The curves for $N_{exp}(t, 230\ K, p_{CO})$, calculated for both the stochastic (Eqs. 24 and 25) and linear deterministic rate equations (ref. 1) and shown in Fig. 3, bear out this expectation. At 10 bars, the stochastic theory predicts a slightly faster rebinding. At 100 bars and above, the discrepancy becomes large, and the linear deterministic rate approach is no longer useful. The results in our preliminary experiments with CO partial pressures up to 128 bars obey the qualitative features of the stochastic description and disagree with the deterministic one. We cannot, however, expect that the actual data will follow the stochastic prediction exactly, because some of the assumptions *i–xi* are too restrictive. It is, for instance, unlikely that an arbitrary number of ligands can occupy wells 2–4. It will be the goal of future work to remove restrictions.

Experiments with high ligand concentrations, to which the present paper is mainly addressed, appear at first sight to have little direct bearing on biological processes. Such experiments may, however, help elucidate ligand migration within biomolecules and explore the limitations on the capacity of accommodating ligands. In some systems more complex than Mb, nonlinear situations may occur even under biological conditions. In cytochrome oxidase, for instance, Sharrock and Yonetani (8) have found evidence for an oxygen reservoir that connects to a number of heme groups and is occupied by many CO_2 molecules. The processes that take place under such circumstances can be treated by a straightforward adaptation of the ideas presented here.

The motivation for the present investigation came from R. H. Austin, who performed our first experiments with high CO concentrations. We thank L. Eisenstein, L. B. Sorensen, and H. Thomas for comments and discussions and K. T. Yue for assistance with computer programming. The work was supported by the Swiss National Science Foundation (P.H.); by the U.S. Department of Health, Education, and Welfare under Grant GM 18051, and by the U.S. National Science Foundation under Grant PCM 74-01366.

1. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. (1975) *Biochemistry* **14**, 5355–5373.
2. Haken, H. (1975) *Rev. Mod. Phys.* **47**, 67–121.
3. Iosifescu, A. & Tăutu, P. (1973) in *Stochastic Processes and Applications in Biology and Medicine*, eds. Krickeberg, K., Lewontin, R. C., Neyman, J. & Schreiber, M. (Springer, New York), Biomathematics, Vol. 3, pp. 163–303.
4. McQuarrie, D. A. (1967), *J. Appl. Prob.* **4**, 413–478.
5. Goel, N. S. & Richter-Dyn, N. (1974) *Stochastic Models in Biology*, (Academic Press, New York), Chap. 9, pp. 206–222.
6. Feller, W. (1966) *An Introduction to Probability Theory and Its Application* (John Wiley, New York), Vol. 2, pp. 252–264.
7. Hänggi, P. & Thomas, H. (1977) *Z. Phys.* **B26**, 98–105.
8. Sharrock, M. & Yonetani, T. (1977) *Biochim. Biophys. Acta* **462**, 718–730.