Stochastic theory of ligand migration in biomolecules

N. ALZEBIDI, 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 or 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 100 bars (0.1 to 100 MPa) is treated.

1. Deterministic and stochastic approach

The discovery that access to the binding site in human 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 I at the heme iron, four potential barriers in succession (2) 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

\[ \frac{dN(t)}{dt} = k_2 N(t) - k_1 N(t) - k_3 N(t) + k_4 N(t). \]

Here, \( N(t) \), for instance, denotes the fraction of biomolecules with a CO molecule in well 2 at time \( t \) and \( k_3 \), 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 photoinduced 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 \( N(0) = 1 \), \( N(0) = N(0) = 0 \). In a stopped-flow experiment, the ligand is initially in S and of all wells, \( 1 \)-th is occupied.

The sequential model of Fig. 1 and deterministic linear rate equations adequately describe many experiments. However, more powerful approaches are needed in 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 covariant; 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 (3) the system at time \( t \) is described by a set of stochastic variables \( x(t), x_1(t), \ldots, x_n(t), \ldots, x_1(t), \ldots, x_n(t), \ldots \). Here \( x(t) \) denotes the number of ligands in well 1 at time \( t \) in a given biomolecule with \( 1 \leq N \leq 100 \).

\[ \frac{dN(t)}{dt} = \sum_{x(t)} \frac{dN(t)}{dx(t)} \]

\[ \frac{dN(t)}{dt} = \sum_{x(t)} \frac{dN(t)}{dx(t)} \]

The costs of publishing 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.

2. Experimental approach

In a typical experiment, ligand binding is monitored optically. The optical absorption spectra of a free protein and one with bound ligand differ. From the absorbance measured at a suitable wavelength, the fraction \( N_{\text{rel}}(t) \) of biomolecules without bound ligand can therefore be determined. To compare \( N_{\text{rel}}(t) \) 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, \( N_{\text{rel}}(t) \) thus is an ensemble average and fluctuations in \( N_{\text{rel}}(t) \) can be neglected (central limit theorem) (4). 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 \( \ldots, N_{\text{max}} \) have a negligible effect on the optical spectrum. We can thus make the identification

\[ N_{\text{rel}}(t) \approx \frac{1}{N_{\text{max}}} \sum_{x(t)} p(x(t) = 0, x_2(t), \ldots, x_{N_{\text{max}}}(t)). \]

The experimental binding data can be evaluated if the time development of the probability function \( p(x(t), t) \) 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 \( x(t) \) describes the occupation of well \( L \) at time \( t \) in a given biomolecule and also the actual value in the state-space \( x \approx 0, 1, \ldots, L_{\text{max}} \). The meaning of \( x(t) \) (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 \( F_{L_{\text{max}}} \).

\[ \left( \ldots, x_{L_{\text{max}}}, \ldots, x_{L_{\text{max}} - 1} \right) \rightarrow \left( \ldots, x_{L_{\text{max}}}, \ldots, x_{L_{\text{max}} - 1} \right) \]

\[ \left( \ldots, x_{L_{\text{max}}}, \ldots, x_{L_{\text{max}} - 1} \right) \rightarrow \left( \ldots, x_{L_{\text{max}}}, \ldots, x_{L_{\text{max}} - 1} \right) \]

4 To whom reprint requests should be addressed.

5 In ref. 1, we denote wells 1 to \( 5 \) by A to D and the solvent by E. The notation used here makes the calculations simpler.

6 F. H"{a}nggi, unmarked.
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 of ligands initially present in well \( L \) and to depend on all other variables in the form

\[
\Gamma_{iKL}(x_L) = \gamma_{x_L}(x_L) \psi_i(x_L).
\]  

[3]

The rate parameters \( \gamma_{x_L} \) are in general still functions of the occupation number \( x_L \) of the final well \( K \) in the inequivalent well \( (K = 1) \), the ligand bonds covalently to the hinge iron. This fact leads to two conditions:

(i) The binding is so tight that well 1 acts as a trap, transitions \( t_{1K} \) are irrelevant to 1 and others can be neglected,

\[
\gamma_{x_L}(x_L) \approx 0, \quad K = 2 \ldots L_{\text{max}}
\]  

[4]

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

\[
\Gamma_{21L}(x_L) = \gamma_{x_L}(x_L) \psi_i(x_L)
\]  

[5]

We assume that the selection acts as a bath with constant ligand concentration, \( n_L \). For transitions involving 5, Eq. 3 consequently becomes

\[
\Gamma_{iKL}(x_L) = \gamma_{x_L}(x_L) \psi_i(x_L).
\]  

[6]

The prime on \( \gamma_{x_L} \) indicates that it is a second-order rate parameter. The binding site, \( L = 1 \), is assumed not to couple directly to the solvent:

\[
\gamma_{x_L} = 0.
\]  

[7]

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. Eqns. 3 and 7 then become for \( K, L \neq 1 \)

\[
\Gamma_{iKL}(x_L) = \gamma_{x_L}(x_L) \psi_i(x_L)
\]  

[9]

\[
\Gamma_{1KL}(S_L) = \gamma_{x_L}(x_L) \psi_i(x_L)
\]  

[10]

In Eq. 10, \( [S_L] \) can be considered one of the parameters of the system.

(iii) The time development of the probability \( p_i(x_L) \) shall be Markovian. The transition rate then depends only on internal physical parameters and variables, but not on the initial preparation, \( p_i(0) \), of the system.

\[
\Gamma_{iKL}(x_L) = \gamma_{x_L}(x_L) \psi_i(x_L)
\]  

[11]

The prime on \( \gamma_{x_L} \) indicates that it is a second-order rate parameter. The binding site, \( L = 1 \), is assumed not to couple directly to the solvent:

\[
\gamma_{x_L} = 0.
\]  

[12]

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. Eqns. 3 and 7 then become for \( K, L \neq 1 \)

\[
\Gamma_{iKL}(x_L) = \gamma_{x_L}(x_L) \psi_i(x_L)
\]  

[13]

The prime on \( \gamma_{x_L} \) indicates that it is a second-order rate parameter. The binding site, \( L = 1 \), is assumed not to couple directly to the solvent:

\[
\gamma_{x_L} = 0.
\]  

[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, \( [S_L] \rightarrow 0 \), 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. 11 with \( \gamma_{x_L} \rightarrow \Delta x_L \) 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 rel. migration of the ligands is Markovian, the rate of change of the probability \( p(x_i) \) obeys the linear master equation

\[
\frac{\partial p(x_i)}{\partial t} = \sum_{x \neq x_i} \left( \gamma_{x_L}(x_L + 1)p(x_i + x_L - 1) - \gamma_{x_L}(x_L)p(x_i) \right)
\]  

[15]
\[ G(x, \tau) = \frac{1}{2} x_1 x_2 x_3 \exp(-\tau) \]

From this, we can calculate the probability of finding the system in state \( x \) at time \( t \).

\[ P(x,t) = \frac{1}{2} \left[ \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \right] \]

where \( \sigma \) is the standard deviation and \( x \) is the position of the particle.

This result is analogous to the Gaussian distribution in statistics, which is widely used to model random variables.

\[ \text{Normal}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \]

where \( \mu \) is the mean and \( \sigma \) is the standard deviation.

The above expression is the probability density function of a normal distribution, which is a common distribution in physics, particularly in statistical mechanics and quantum mechanics.

\[ \text{Normal}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \]

This expression is used to describe the distribution of a random variable that is the sum of many independent and identically distributed random variables.
FIG. 3. Computer-simulated rebinding curves for MHC. N(t) after photodissociation at 230 K. The CO partial pressure increases from 1 to 10 bar. The solid lines are calculated with the stochastic model, the broken ones with deterministic linearized equations. The top lines of labeled in the figure refer to the stochastic calculation.

\[ N_{\text{eq}}(t, \lambda) \] for a fixed value of \( \lambda \), on the rate parameters \( \gamma \) and \( \beta \). 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 in a given biomolecule can no longer be neglected. As an example, we consider the rebinding of CO to MHC after photodissociation at 230 K, with the rate parameters \( \gamma_{12} = 2.8 \times 10^4 \text{ s}^{-1} \), \( \gamma_{23} = 4.7 \times 10^4 \text{ s}^{-1} \), \( \gamma_{34} = 2.5 \times 10^3 \text{ s}^{-1} \), \( \gamma_{41} = 3.7 \times 10^4 \text{ s}^{-1} \), \( \gamma_{52} = 4.9 \times 10^4 \text{ s}^{-1} \), \( \gamma_{63} = 1.9 \times 10^5 \text{ s}^{-1} \), \( \beta_1 = 1.7 \times 10^3 \text{ s}^{-1} \), \( \beta_2 = 5.5 \times 10^4 \text{ s}^{-1} \), in which \( \rho_{\text{CO}} \) is the CO partial pressure in bars (1). For clarity, we assume \( \beta \) to be valid and neglect distributed barriers. The initial mean values \( (\bar{a}) \), determined from Eq. 23, are given in Fig. 2 as a function of \( \rho_{\text{CO}} \). The outermost well, 4, is appreciably populated above about 10 bars, the next one above 100 bars. The solubility of CO inside MHC 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 \( N_{\text{eq}}(t, \lambda, \rho_{\text{CO}}) \), 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 bar 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 \( \gamma \) 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 ofaccommodating ligands. In some systems more complex than MHC, nonlinear situations may occur even under biological conditions. In cytochrome oxidase, for instance, Sharrack and Yonetani (8) have found evidence for an oxygen reservoir that connects to a number of heme groups and is occupied by many CO2 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 B. H. Austin, who performed our first experiments with high CO concentrations. We thank L. Eisenstein, L. B. Sormen, and H. Thomas for comments and discussions and K. T. Yee 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 1805, and by the U.S. National Science Foundation under Grant PCM 74-01396.