Note: Network random walk model of two-state protein folding: Test of the theory

Alexander M. Berezhkovskii, Ronan D. Murphy, and Nicolae-Viorel Buchete

Citation: J. Chem. Phys. 138, 036101 (2013); doi: 10.1063/1.4776215
View online: http://dx.doi.org/10.1063/1.4776215
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v138/i3
Published by the American Institute of Physics.

Additional information on J. Chem. Phys.
Journal Homepage: http://jcp.aip.org/
Journal Information: http://jcp.aip.org/about/about_the_journal
Top downloads: http://jcp.aip.org/features/most_downloaded
Information for Authors: http://jcp.aip.org/authors
Single-exponential kinetics of an isomerization reaction $A \iff B$ implies that the equilibrium distribution of the molecule has two peaks localized in different parts of the phase space (two basins). Transitions between the two basins occur via a sparsely populated area, the so-called “bottleneck” region. Although the region does not manifest itself in the thermodynamics, dynamics of the system in this region determines the rate constants of the forward and reverse reactions. These rate constants are given by the Kramers ratio of the reactive flux at equilibrium to the equilibrium populations of $A$ and $B$ states of the molecule. A distinctive feature of the single-exponential kinetics is that there is some freedom in the choice of the boundaries separating the basins and the bottleneck region; both the reactive flux and the equilibrium populations are weakly sensitive to the exact location of the boundaries in some range of their positions.

Two-state protein folding described by the kinetic scheme $U \xrightleftharpoons[k_{U}]^{k_{F}} F$ is an example of the isomerization reaction. Here $U$ and $F$ denote unfolded and folded states of the protein, while $k_{F}$ and $k_{U}$ are the folding and unfolding rate constants given by $k_{U,F} = J / P_{F,U}^{eq}$, where $J$ is the reactive flux at equilibrium and $P_{F,U}^{eq}$ are equilibrium populations. An expression for $J$ has been recently derived assuming that the underlying protein dynamics is described in terms of Markovian transitions on a network of complex connectivity, which is formed by coarse-grained microstates of the protein. To test this expression, we apply it to a toy model of protein dynamics, which has an important advantage: it allows for an approximate analytical solution for the sum of the rate constants. Here, we show that the expression for the reactive flux, which follows from this solution, is identical to the expression obtained using the formula derived in Ref. 1.

Let $p_i(t)$ be the probability of finding the protein in microstate $i$, $i = 1, 2, \ldots, N$, where $N$ is the number of microstates. This probability satisfies

$$\frac{dp_i(t)}{dt} = K p_i(t),$$

where the evolution operator $K$ is the $N \times N$ rate matrix, in which the non-diagonal matrix element $K_{ij}$ is the rate constant for transition $j \to i$ ($i \neq j$) and the diagonal matrix element $K_{ii}$ is the rate constant for escape from microstate $i$, $K_{ii} = -\sum_{i \neq j} K_{ij}$. The elements of the rate matrix are $K_{ij}$ and their magnitudes, $\lambda_i = \lambda_{eq}$ and their corresponding eigenvectors $\phi_{eq}^{(n)}$ are solutions of the eigenvalue problem $K_{ij} \phi_{eq}^{(n)} = \lambda_{eq} \phi_{eq}^{(n)}$, $n = 1, 2, \ldots, N$. The smallest in magnitude eigenvalue is zero, while all other eigenvalues are negative and can be ordered in increasing values of their magnitudes, $\lambda_1 = 0 < \lambda_2 < \lambda_3 \leq \ldots \leq \lambda_N$. When the master equation, Eq. (1), describes stochastic dynamics that underlies two-state protein folding, $\lambda_2$ is the sum of the folding and unfolding rate constants, $\lambda_2 = k_F + k_U = J / P_{F,U}^{eq} P_{U,F}^{eq}$.

The main result of Ref. 1 is a formula giving the reactive flux in terms of the splitting (commitment) probabilities in the bottleneck region (see also Ref. 3). In deriving the formula, all microstates are divided into three groups: definitely unfolded ($UU$), definitely folded ($FF$), and intermediate ($I$) or bottleneck microstates. The latter separate the $UU$- and $FF$-microstates in the sense that all transitions between them occur only through the intermediate microstates. The splitting probability $p_{fold}(i)$ of microstate $i$ is defined as the probability of reaching the $FF$-basin before reaching the $UU$-basin starting from microstate $i$, so that $p_{fold}(i) = 1$ for $i \in FF$, $p_{fold}(i) = 0$ for $i \in UU$, and $0 < p_{fold}(i) < 1$ for $i \in I$. As shown in Ref. 1, the reactive flux through an arbitrary dividing surface $\Sigma$ in the bottleneck region, which is necessarily crossed by any trajectory connecting the $FF$- and $UU$-basins, is given by

$$J = \sum_{i \in F^*} \sum_{j \in U^*} K_{ij} p_{eq}^{i} [p_{fold}(i) - p_{fold}(j)],$$

where $F^*$ and $U^*$ are the sets of microstates located on the $F$- and $U$-sides of $\Sigma$.

Here, we use the formula in Eq. (2) to derive an explicit expression for the reactive flux assuming that (i) the protein is at its melting temperature, i.e., $P_{U,F}^{eq} = P_{F,U}^{eq} = 1/2$ and (ii) its dynamics is described by the simple model proposed in Ref. 2. This model is an extension of the model suggested by Zwanzig, Szabo, and Bagchi to study Levintal’s paradox.

The model assumes that the protein contains $M$ identical
residues, each of which can be in either a folded (f) or an unfolded (u) state. A microstate of the protein is completely characterized by indicating the state of each residue. Therefore, the total number of microstates is \(2^M\). In addition, it is assumed that transitions among microstates occur as a result of independent transitions between f- and u-states of individual residues described by the kinetic scheme

\[ u \xrightarrow{k} f, \text{ where the rate constant } k \text{ is assumed to be the same for transitions in both directions.} \]

The model introduces a collective behavior of the protein responsible for its folding by postulating that the first and second residue unfolding events occur with modified rate constants \(\varepsilon_0 k\) and \(\varepsilon_1 k\), respectively. \(\varepsilon_0, \varepsilon_1 \approx 1\). The model also postulates that (i) the FF-basin contains the only microstate in which all residues are in their f-states, while the UU-basin contains microstates with two and more unfolded residues, and (ii) microstates containing the only unfolded residue form a transition state ensemble (TSE), in the sense that transitions from these microstates to microstates with two and zero unfolded residues occur with probability \(1/2\). This leads to

\[ \varepsilon_0 = \frac{M - 1}{2^M + (M - 1)^2 - 2}, \quad \varepsilon_1 = \frac{1}{M - 1}. \]  

(3)

The equilibrium probability of finding the protein in a microstate with \(m\) unfolded residues \(P_m^{eq}\) is given by

\[ P_m^{eq} = \frac{\varepsilon_0 (1 - \varepsilon_0) \delta_{0m}}{2[1 - (1 - \varepsilon_1) \delta_{0m}]} \left( \frac{M}{m} \right). \]  

(4)

where \(\delta_{mn}\) is the Kronecker delta.

For this model of protein dynamics, an approximate solution for \(\lambda_2, \lambda_3 = M(M - 1)k/[2^M + (M - 1)^2 - 2]\), was found assuming that the protein was at its melting temperature.\(^2\) Comparison with \(\lambda_3\) obtained by numerical diagonalization of the rate matrix shows very good agreement between the analytical and numerical results for \(M \geq 20\).\(^2\) Since at the melting temperature \(J = \lambda_2/4\), we have

\[ J = \frac{M(M - 1)k}{4[2^M + (M - 1)^2 - 2]}. \]  

(5)

Below, we obtain this result from Eq. (2). This is done in two steps. First, by taking advantage of the fact that microstates with the same number of folded/unfolded residues are indistinguishable, we reformulate the problem in terms of effective microstates, which are unions of the indistinguishable initial microstates. Then, we find \(J\) for the reformulated problem.

Let \(P_m(t)\) be the probability of finding the protein in one of the microstates \(m\) of folded residues and \(M - m\) folded residues, \(m = 0, 1, 2, \ldots, M\). It can be shown that probabilities \(P_m(t)\) satisfy the master equation \(k^{-1}dP(t)/dt = K^{eff}P(t)\), where \(K^{eff}\) is the \((M + 1) \times (M + 1)\) three-diagonal rate matrix, with non-diagonal matrix elements given by \(K^{eff}_{j-1,j} = j\), \(j = 1, 2, \ldots, M\), \(K^{eff}_{10} = \varepsilon_0 M\), \(K^{eff}_{21} = 1\), \(K^{eff}_{j+1,j} = M - j\), \(j = 2, 3, \ldots, M - 1\), and the diagonal matrix elements are \(K^{eff}_{jj} = -(K^{eff}_{j-1,j} + K^{eff}_{j+1,j})\). Assuming that the dividing surface \(\Sigma\) separates microstates with the same number of folded/unfolded residues, we can write an expression for the reactive flux in terms of effective microstates using the equilibrium distribution \(P_m^{eq}\) and the corresponding splitting probabilities \(P_{fold}(m)\),

\[ J = k \sum_{m \in U^*} K^{eff}_{lm} P^{eq}_{m}(P_{fold}(l) - P_{fold}(m)). \]  

(6)

This expression is an equivalent of Eq. (2) formulated in terms of the initial microstates. In our model, the FF-basin contains the only effective microstate, in which all residues are in their folded states. The \(UU\)-basin contains effective microstates with at least two unfolded residues. The TSE is formed by the only effective microstate, which is a union of all initial microstates with the only unfolded residue.

There are two possibilities to draw the dividing surface \(\Sigma\) in such a system, between the TSE and (i) the FF-basin or (ii) the \(UU\)-basin. In the first case, we have

\[ J = kK^{eff}_{01} P_{fold}^{eq}[P_{fold}(0) - P_{fold}(1)]. \]  

(7)

Here, \(K^{eff}_{01} P^{eq}_{0} = K^{eff}_{10} P^{eq}_{1} = M \varepsilon_0/2\), \(P_{fold}(0) = 1\), and \(P_{fold}(1) = 1/2\), so that the flux is \(J = kM \varepsilon_0/4\). Substituting \(\varepsilon_0\) here, given in Eq. (3), we recover the expression for the flux in Eq. (5). In the second case, Eq. (6) leads to

\[ J = kK^{eff}_{12} P_{fold}^{eq}[P_{fold}(1) - P_{fold}(2)], \]  

(8)

where \(K^{eff}_{12} P_{2} = K^{eff}_{21} P_{1} = M \varepsilon_0/2\), \(P_{fold}(1) = 1/2\), and \(P_{fold}(2) = 0\). Substituting these into Eq. (8) and using \(\varepsilon_0\) in Eq. (3), we again recover the result in Eq. (5).

To summarize, we have demonstrated that the expression for the reactive flux given in Eq. (5) for the toy model of the protein dynamics can be obtained from the Kramers-type formula for the reactive flux, Eq. (2), derived under the assumption that the protein dynamics is described by a Markov random walk on a network of complex connectivity.

We are grateful to Gerhard Hummer and Attila Szabo for helpful discussions. This study was supported by the Intramural Research Program of the National Institutes of Health, Center for Information Technology, and by the Irish Research Council.

---