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

Alexander M. Berezhkovskii,¹ Ronan D. Murphy,^{2,3} and Nicolae-Viorel Buchete^{2,3,a)}

¹Mathematical and Statistical Computing Laboratory, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, USA

²School of Physics, University College Dublin, Belfield, Dublin 4, Ireland

³Complex and Adaptive Systems Laboratory, University College Dublin, Belfield, Dublin 4, Ireland

(Received 3 October 2012; accepted 2 January 2013; published online 18 January 2013)

[<http://dx.doi.org/10.1063/1.4776215>]

Single-exponential kinetics of an isomerization reaction $A \rightleftharpoons B$ 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 derived¹ 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.² 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 , $i = 1, 2, \dots, N$, where N is the number of microstates. This probability satisfies

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{K}\mathbf{p}(t), \quad (1)$$

where the evolution operator $\mathbf{K}$ is the $N \times N$ rate matrix, in which the non-diagonal matrix element K_{ij} is the rate con-

stant for transition $j \rightarrow i$ ($i \neq j$) and the diagonal matrix element K_{jj} is the rate constant for escape from microstate j , $K_{jj} = -\sum_{i \neq j} K_{ij}$. The elements of the rate matrix satisfy the condition of detailed balance $K_{ij}p_j^{eq} = K_{ji}p_i^{eq}$, where p_i^{eq} is the equilibrium probability of finding the protein in microstate i , which satisfies $\mathbf{K}\mathbf{p}^{eq} = 0$. Eigenvalues $-\lambda_n$ and their corresponding eigenvectors $\varphi^{(n)}$ are solutions of the eigenvalue problem $\mathbf{K}\varphi^{(n)} = -\lambda_n\varphi^{(n)}$, $n = 1, 2, \dots, N$. The smallest in magnitude eigenvalue is zero, while all other eigenvalues are negative and can be ordered in increasing values of their magnitudes, $\lambda_1 = 0 < \lambda_2 < \lambda_3 \leq \dots \leq \lambda_N$. When the master equation, Eq. (1), describes stochastic dynamics that underlies two-state protein folding, λ_2 is the sum of the folding and unfolding rate constants, $\lambda_2 = k_F + k_U = J/(P_U^{eq}P_F^{eq})$.

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 Σ in the bottleneck region, which is necessarily crossed by any trajectory connecting the FF - and UU -basins, is given by

$$J = \sum_{\substack{i \in F^* \\ j \in U^*}} K_{ij}p_j^{eq}[p_{fold}(i) - p_{fold}(j)], \quad (2)$$

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

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., $P_U^{eq} = P_F^{eq} = 1/2$ 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 Levinthal's paradox.⁴ The model assumes that the protein contains M identical

^{a)}E-mail: buchete@ucd.ie.

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 N of the 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 $u \xrightleftharpoons[k]{k} f$, where the rate constant k 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, $\varepsilon_0, \varepsilon_1 \ll 1$. 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

$$\varepsilon_0 = \frac{M-1}{2^M + (M-1)^2 - 2}, \quad \varepsilon_1 = \frac{1}{M-1}. \quad (3)$$

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

$$P_m^{eq} = \frac{[\varepsilon_0 + (1 - \varepsilon_0)\delta_{m0}][\varepsilon_1 + (1 - \varepsilon_1)\delta_{m1}]}{2[1 - (1 - \varepsilon_1)\delta_{m0}]} \binom{M}{m}, \quad (4)$$

where δ_{mm} is the Kronecker delta.

For this model of protein dynamics, an approximate solution for λ_2 , $\lambda_2 = M(M-1)k/[2^M + (M-1)^2 - 2]$, was found assuming that the protein was at its melting temperature.² Comparison with λ_2 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

$$J = \frac{M(M-1)k}{4[2^M + (M-1)^2 - 2]}. \quad (5)$$

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 unifications 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 with m unfolded residues and $M-m$ folded residues, $m = 0, 1, 2, \dots, M$. It can be shown that probabilities $P_m(t)$ satisfy the master equation $k^{-1}d\mathbf{P}(t)/dt = \mathbf{K}^{eff}\mathbf{P}(t)$, where $\mathbf{K}^{eff}$ is the $(M+1) \times (M+1)$ three-diagonal rate matrix, with non-diagonal matrix elements given by $K_{j-1,j}^{eff} = j$, $j = 1, 2, \dots, M$, $K_{10}^{eff} = \varepsilon_0 M$, $K_{21}^{eff} = 1$, $K_{j+1,j}^{eff} = M-j$, $j = 2, 3, \dots, M-1$, and the diagonal matrix elements

are $K_{jj}^{eff} = -(K_{j-1,j}^{eff} + K_{j+1,j}^{eff})$. Assuming that the dividing surface Σ 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 $P_{fold}(m)$,

$$J = k \sum_{\substack{l \in F^* \\ m \in U^*}} K_{lm}^{eff} P_m^{eq} [P_{fold}(l) - P_{fold}(m)]. \quad (6)$$

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 unification of all initial microstates with the only unfolded residue.

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

$$J = k K_{01}^{eff} P_1^{eq} [P_{fold}(0) - P_{fold}(1)]. \quad (7)$$

Here, $K_{01}^{eff} P_1^{eq} = K_{10}^{eff} P_0^{eq} = M\varepsilon_0/2$, $P_{fold}(0) = 1$, and $P_{fold}(1) = 1/2$, so that the flux is $J = kM\varepsilon_0/4$. Substituting ε_0 here, given in Eq. (3), we recover the expression for the flux in Eq. (5). In the second case, Eq. (6) leads to

$$J = k K_{12}^{eff} P_2^{eq} [P_{fold}(1) - P_{fold}(2)], \quad (8)$$

where $K_{12}^{eff} P_2^{eq} = K_{21}^{eff} P_1^{eq} = M\varepsilon_0/2$, $P_{fold}(1) = 1/2$, and $P_{fold}(2) = 0$. Substituting these into Eq. (8) and using ε_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.

¹A. Berezhkovskii, G. Hummer, and A. Szabo, *J. Chem. Phys.* **130**, 205102 (2009).

²A. M. Berezhkovskii, F. Tofoleanu, and N.-V. Buchete, *J. Chem. Theory Comput.* **7**, 2370 (2011).

³Ch. Schutte, F. Noe, E. Meerbach, P. Metzner, and C. Hartmann, in *Proceedings of the 6th International Congress on Industrial and Applied Mathematics, Zurich, Switzerland, 15–20 July, 2007*, edited by I. Jeltsch and G. Wanner (European Mathematical Society Publishing House, 2009), pp. 297–335; P. Metzner, Ch. Schutte, and E. Vanden-Eijnden, *Multiscale Model. Simul.* **7**, 1192 (2009).

⁴R. Zwanzig, A. Szabo, and B. Bagchi, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 20 (1992); R. Zwanzig, *ibid.* **92**, 9801 (1995); **94**, 148 (1997); N. V. Buchete and J. E. Straub, *J. Phys. Chem. B* **105**, 6684 (2001).