

# Reactive flux and folding pathways in network models of coarse-grained protein dynamics

Alexander Berezhkovskii,<sup>1,a)</sup> Gerhard Hummer,<sup>2</sup> and Attila Szabo<sup>2</sup>

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

<sup>2</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

(Received 9 March 2009; accepted 27 April 2009; published online 29 May 2009)

The reactive flux between folded and unfolded states of a two-state protein, whose coarse-grained dynamics is described by a master equation, is expressed in terms of the commitment or splitting probabilities of the microstates in the bottleneck region. This allows one to determine how much each transition through a dividing surface contributes to the reactive flux. By repeating the analysis for a series of dividing surfaces or, alternatively, by partitioning the reactive flux into contributions of unidirectional pathways that connect reactants and products, insight can be gained into the mechanism of protein folding. Our results for the flux in a network with complex connectivity, obtained using the discrete counterpart of Kramers' theory of activated rate processes, show that the number of reactive transitions is typically much smaller than the total number of transitions that cross a dividing surface at equilibrium. [DOI: [10.1063/1.3139063](https://doi.org/10.1063/1.3139063)]

## I. INTRODUCTION

Discrete coarse-graining of the dynamics is a powerful approach to studying protein folding and other molecular processes that occur on time scales inaccessible to current atomistic simulations.<sup>1–17</sup> Configuration space is first divided into appropriately defined discrete microstates, and then the dynamics is modeled using a master equation that describes stochastic transitions among these microstates. The required rate coefficients can be extracted using a variety of strategies, for instance by monitoring transitions between coarse-grained states in molecular simulations.<sup>7–17</sup>

Here we establish the relation between the macroscopic kinetics and the underlying microscopic dynamics on a network with complex connectivity (Fig. 1). The focus is on providing a theoretical framework to unravel mechanisms of protein folding when the dynamics is described by a master equation. However, our results are applicable to any discrete kinetic system, such as those that arise in systems biology or the modeling of enzyme catalysis.

One of the main results of this paper is a remarkably simple expression for the reactive flux in terms of the commitment or splitting probabilities,  $p_{\text{fold}}$ , of microstates in the bottleneck region that separates reactants from products. The  $p_{\text{fold}}$  of a microstate is defined as the probability that a protein starting from this microstate folds before it unfolds. While this result simplifies the calculation of the rate, its primary importance is that it gives the contribution to the reactive flux of each transition through a dividing surface located in the bottleneck region. By repeating such an analysis for a series of dividing surfaces or from the decomposi-

tion of the reactive flux into a sum of contributions of unidirectional pathways, one can gain insight into the mechanism of protein folding.

Our expression for the reactive flux is applicable to discrete dynamics with complex connectivity of the microstates (Fig. 1) and is not based on even an implicit assumption that low-dimensional reaction coordinates exist. We shall show that in general the reactive flux is much smaller than the upper bound given by transition state theory (TST). In TST, which ignores recrossings, the reactive flux is approximated by the number of transitions per unit time that cross a dividing surface in one direction. Our expression for the reactive flux shows that the TST contribution of every transition  $i \rightarrow j$  that crosses the dividing surface is reduced by the difference in  $p_{\text{fold}}$ 's of microstates  $j$  and  $i$ . When transitions occur between microstates with similar values of  $p_{\text{fold}}$ , this difference is much smaller than unity.

The commitment or splitting probability,  $p_{\text{fold}}$ , plays a key role in our analysis.  $p_{\text{fold}}(i)$  can be found from trajectories obtained in simulations or in state-resolved single-molecule experiments by counting the fraction of trajectory fragments starting in microstate  $i$  that reach the fully folded state before the unfolded state. The importance of this quantity appears to have been first recognized by Ryter<sup>18</sup> who suggested using the stochastic separatrix (the locus of the phase space points for which the splitting probability is equal to  $\frac{1}{2}$ ) as the boundary separating reactants from products. This idea has been exploited in the theory of chemical reactions<sup>19</sup> and was introduced into the protein-folding literature by Du *et al.*<sup>20</sup> who defined the transition state ensemble (TSE) as the set of microstates with  $p_{\text{fold}}$  close to  $\frac{1}{2}$ . In this paper we shall derive the expression for the reactive flux in terms of  $p_{\text{fold}}$  using the discrete analogs of the Kramers<sup>21</sup> and Chandler<sup>22</sup> approaches.

<sup>a)</sup>Electronic mail: berezh@helix.nih.gov.

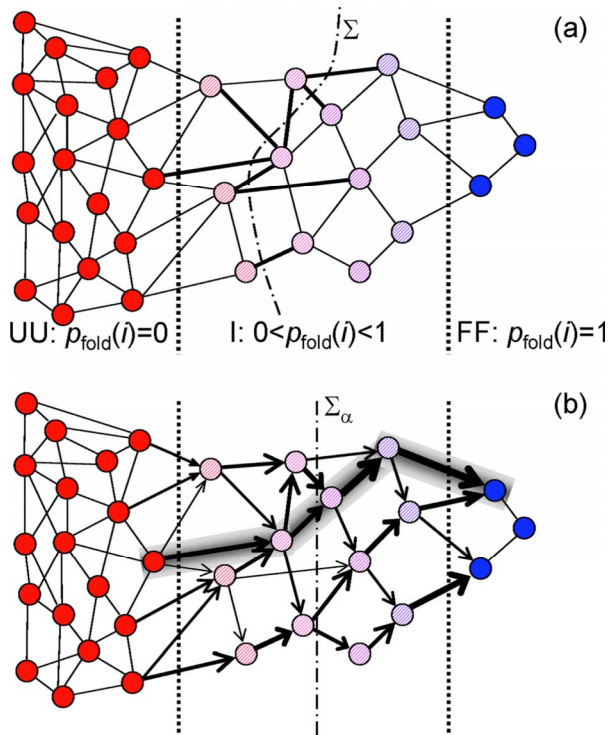


FIG. 1. (Color online) Schematic representation of a network of coarse-grained microstates of a two-state protein (circles). (a) Connectivity network. Reversible transitions among the microstates are denoted by solid lines. Microstates are divided into three groups: Definitely unfolded [ $UU$  with  $p_{\text{fold}}(i)=0$ ; red], intermediate [ $I$  with  $0 < p_{\text{fold}}(i) < 1$ ; shaded] and definitely folded [ $FF$  with  $p_{\text{fold}}(i)=1$ ; blue]. The boundaries separating these groups are shown by dashed lines. All transitions between the  $UU$ - and  $FF$ -microstates occur through the  $I$ -microstates, which form the bottleneck. Bold edges indicate the transitions across a dividing surface  $\Sigma$  (shown by the dashed-dotted curve) contributing to the reactive flux. (b) Unidirectional reactive flux diagram showing protein folding currents through the bottleneck region. The direction and thickness of the arrows are proportional to  $J_{i-j}$  defined in Eq. (14). The dominant unidirectional path is indicated by shading. The dashed-dotted line is a dividing surface  $\Sigma_\alpha$  at  $p_{\text{fold}}=\alpha$ .

## II. THEORY

Our analysis is based on the master equation that describes the evolution of the probability  $p_i(t)$  of finding the protein in microstate  $i$  at time  $t$ ,  $i=1, 2, \dots, N$ ,

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

The  $ij$ -th element of the  $N \times N$  rate matrix  $\mathbf{K}$ ,  $K_{ij}$ , is the rate constant for transition  $j \rightarrow i$  ( $i \neq j$ ) and  $K_{jj} = -\sum_{i \neq j} K_{ij}$ . The equilibrium probability of finding the protein in microstate  $i$ ,  $p_{\text{eq}}(i)$ , is the solution to  $\mathbf{K}\mathbf{p}_{\text{eq}}=0$  with the normalization  $\sum_{i=1}^N p_{\text{eq}}(i)=1$ . The elements of the rate matrix satisfy the condition of detailed balance,  $K_{ij}p_{\text{eq}}(j)=K_{ji}p_{\text{eq}}(i)$ , which follows from the requirement that there is no net flux between microstates at equilibrium.

We shall assume that the structure of the rate matrix,  $\mathbf{K}$ , is such that folding can be accurately described by a phenomenological kinetic scheme



where  $k_F$  and  $k_U$  are the folding and unfolding rate constants for transitions between unfolded ( $U$ ) and folded ( $F$ ) states of the protein. According to this scheme, the reactive flux in the  $U \rightarrow F$  direction at time  $t$  is  $k_F f_U(t)$ , where  $f_U(t)$  is the population of the unfolded state at time  $t$ . At equilibrium this flux, which is the number of  $U \rightarrow F$  transitions per unit time, is equal to the reactive flux in the opposite ( $F \rightarrow U$ ) direction. Both fluxes, denoted by  $J$ , are given by

$$J = k_F f_U^{\text{eq}} = k_U f_F^{\text{eq}}, \quad (3)$$

where  $f_U^{\text{eq}}$  and  $f_F^{\text{eq}}=1-f_U^{\text{eq}}$  are fractional populations of the folded and unfolded states of the protein at equilibrium.

For the kinetics to be described by the scheme in Eq. (2), the majority of the protein population must be localized in two basins separated by a narrow bottleneck, so that the intrabasin relaxation occurs much faster than the interbasin exchange, and thus kinetics of equilibration is essentially single exponential. In such a system all microstates can be separated into three groups: definitely unfolded ( $UU$ ), definitely folded ( $FF$ ), and intermediate ( $I$ ) or bottleneck microstates (Fig. 1). There is some freedom in the choice of the bottleneck boundaries (two dashed lines in Fig. 1) since the reactive flux is rather insensitive to their precise location. Although the equilibrium population of the  $I$ -microstates is small,  $\sum_{i \in I} p_{\text{eq}}(i) \ll \sum_{i \in L} p_{\text{eq}}(i)$ ,  $L=FF, UU$ , these microstates play an important role since all transitions between the  $UU$ - and  $FF$ -basins occur only through the bottleneck,  $UU \leftrightarrow I \leftrightarrow FF$  (i.e., we assume that there are no direct transitions between  $FF$ - and  $UU$ -basins,  $K_{ij}=K_{ji}=0$ ,  $i \in UU$ ,  $j \in FF$ ).

To establish the relationship between the reactive flux at equilibrium and the discrete protein dynamics, we follow Kramers and consider a steady state in which all trajectories entering the  $FF$ -basin from the bottleneck are instantly returned to the  $UU$ -basin to maintain local equilibrium in this basin. The resulting steady-state distribution,  $p_{ss}(i)$ , satisfies

$$\sum_j K_{ij} p_{ss}(j) = 0 \quad (4)$$

subject to the boundary conditions  $p_{ss}(i)=p_{\text{eq}}(i)$  if  $i \in UU$  and  $p_{ss}(i)=0$  if  $i \in FF$ . The reactive flux  $J$  is approximated by the steady-state flux through an arbitrary dividing surface  $\Sigma$  in the bottleneck region, which is necessarily crossed by any trajectory connecting the  $UU$ - and  $FF$ -basins,

$$J = \sum_{i \in F^*, j \in U^*} [K_{ij} p_{ss}(j) - K_{ji} p_{ss}(i)], \quad (5)$$

where  $F^*$  and  $U^*$  denote the sets of the microstates in the  $I$  region located on the  $F$ - and  $U$ -sides of  $\Sigma$  that are connected by a single transition.

The solution to Eq. (4) can be written in terms of  $p_{\text{eq}}(i)$  and the splitting probability,  $p_{\text{fold}}(i)$ , defined as the probability of reaching the  $FF$ -basin before reaching the  $UU$ -basin starting from microstate  $i$ . For  $i \in FF$   $p_{\text{fold}}(i)=1$ , while for

$i \in UU$   $p_{\text{fold}}(i)=0$ . The remaining  $p_{\text{fold}}$ 's, as indicated in Ref. 23, can be found by solving the discrete version of Onsager's equation<sup>24</sup> for the splitting probability,

$$\sum_i p_{\text{fold}}(i)K_{ij} = \sum_{i \in I} p_{\text{fold}}(i)K_{ij} + \sum_{i \in FF} K_{ij} = 0, \quad j \in I. \quad (6)$$

This equation can be derived by considering the relation between the  $p_{\text{fold}}$  of microstate  $j$  and those of microstates that are connected to it by a single transition (i.e., microstates  $i$  for which  $K_{ij} \neq 0$ ). After one transition from microstate  $j$ , the protein ends up in microstate  $i$  with probability  $K_{ij}/\sum_{l \neq j} K_{lj}$ . From the probability conservation it then follows that

$$p_{\text{fold}}(j) = \sum_{i \neq j} p_{\text{fold}}(i) \frac{K_{ij}}{\sum_{l \neq j} K_{lj}}. \quad (7)$$

Since  $\sum_{l \neq j} K_{lj} = -K_{jj}$ , Eq. (7) is equivalent to Eq. (6).

Comparing Eqs. (4) and (6) and using the condition of detailed balance,  $K_{ij}p_{\text{eq}}(j) = K_{ji}p_{\text{eq}}(i)$ , it can be shown that

$$p_{ss}(i) = [1 - p_{\text{fold}}(i)]p_{\text{eq}}(i). \quad (8)$$

Substituting this solution for  $p_{ss}(i)$  into Eq. (5) and again using the condition of detailed balance, we obtain one of the main results of this paper,

$$J = \sum_{i \in F^*, j \in U^*} K_{ij}p_{\text{eq}}(j)[p_{\text{fold}}(i) - p_{\text{fold}}(j)], \quad (9)$$

which is the expression for the reactive flux when the underlying dynamics involves stochastic transitions on a network of microstates with complex connectivity (Fig. 1). One can use this expression to find the folding and unfolding rate constants by means of the flux-over-population formula,  $k_{F,U} = J/f_{U,F}^{\text{eq}}$ , [see Eq. (3)]. In the special case when the microstates correspond to lattice points in a multidimensional space and the transitions occur only between the nearest neighbors, the continuum limit of Eq. (9) has been obtained by E and Vanden-Eijnden.<sup>25</sup> We now rederive the result in Eq. (9) by another method.

To obtain this expression using Chandler's approach,<sup>22</sup> consider the "number" correlation function  $C_F(t) = \langle I_F(t)I_F(0) \rangle$ , where  $I_F$  is unity when the protein is in one of the microstates on the  $F$  side of  $\Sigma$  and zero otherwise. In the two-state phenomenological description, Eq. (2), this correlation function is  $C_F(t) = f_F^{\text{eq}}[f_F^{\text{eq}} + f_U^{\text{eq}} \exp(-(k_F + k_U)t)]$ , so that its time derivative at  $t=0$  is equal to the reactive flux,  $dC_F(t)/dt|_{t=0} = f_F^{\text{eq}}f_U^{\text{eq}}(k_F + k_U) = J$ . In the framework of the more detailed coarse-grained description, Eq. (1), this correlation function is given by  $C_F(t) = \sum_{i,j \in F} G_{ij}(t)p_{\text{eq}}(j)$ , where  $G_{ij}(t)$  is the probability of finding the protein in microstate  $i$  at time  $t$ , given that it starts from microstate  $j$  at  $t=0$ . Using the evolution equation,  $dG_{ij}(t)/dt = \sum_n K_{in}G_{nj}(t)$  with the initial condition  $G_{ij}(0) = \delta_{ij}$ , the time derivative of the correlation function can be written as,

$$\begin{aligned} \frac{dC_F(t)}{dt} &= \sum_{i,j \in F, n \in U} K_{in}G_{nj}(t)p_{\text{eq}}(j) \\ &+ \sum_{i,j,n \in F} K_{in}G_{nj}(t)p_{\text{eq}}(j). \end{aligned} \quad (10)$$

Using the relations  $G_{ij}(t)p_{\text{eq}}(j) = G_{ji}(t)p_{\text{eq}}(i)$  and  $K_{ij}p_{\text{eq}}(j)$

$= K_{ji}p_{\text{eq}}(i)$ , which follow from the condition of detailed balance, and the fact that  $\sum_{i \in F} K_{ij} = -\sum_{i \in U} K_{ij}$ , one can recast Eq. (10) as

$$\frac{dC_F(t)}{dt} = \sum_{i \in F^*, n \in U^*} K_{in}p_{\text{eq}}(n) \sum_{j \in F} [G_{jn}(t) - G_{ji}(t)], \quad (11)$$

where  $i$  and  $n$  are "connected" microstates located on opposite sides of the dividing surface. As  $t \rightarrow 0$ ,  $-dC_F(t)/dt$  reduces to the TST estimate for the reactive flux through the dividing surface  $\Sigma$ ,

$$-\left. \frac{dC_F(t)}{dt} \right|_{t=0} = J_{\text{TST}}(\Sigma) = \sum_{i \in F^*, j \in U^*} K_{ij}p_{\text{eq}}(j). \quad (12)$$

To get a better estimate of the reactive flux, consider  $-dC_F(t)/dt$  on a timescale that is much smaller than the mean interbasin equilibration time,  $(k_F + k_U)^{-1}$ , but much larger than the time spent in the bottleneck region by a protein starting in a microstate near the dividing surface. On such a time scale, once the protein reaches the  $FF$ - or  $UU$ -basin, it does not return to the bottleneck region, so these basins are effectively absorbing. This allows us to approximate the sums in Eq. (11) as:  $\sum_{j \in F} G_{jn}(t) \approx \sum_{j \in FF} G_{jn}(t) \approx p_{\text{fold}}(i)$ ,  $i \in U^*$  or  $F^*$ . As a result,  $-dC_F(t)/dt$  in Eq. (11) reduces to the reactive flux in Eq. (9).

One can also obtain the expression for the reactive flux, Eq. (9), using the following simple argument. Consider an equilibrium ensemble of identical proteins. Each protein in the ensemble can be in one of the two basins or in the bottleneck. We denote the fraction of proteins in microstate  $i$ ,  $i \in I$ , that entered the bottleneck from the  $UU$ -basin by  $\nu_U(i)$ . The reactive flux is the average number of trajectories that go from the  $UU$ -basin to the  $FF$ -basin per unit time. The forward (i.e., in the  $U \rightarrow F$  direction) flux through  $\Sigma$ ,  $J_f(\Sigma)$ , due to such trajectories is given by  $J_f(\Sigma) = \sum_{i \in F^*, j \in U^*} p_{\text{fold}}(i)K_{ij}p_{\text{eq}}(j)\nu_U(j)$ . Since a reactive trajectory can cross  $\Sigma$  several times, the reactive flux is  $J_f(\Sigma)$  minus the flux due to backward transitions of these trajectories through  $\Sigma$ ,  $J_b(\Sigma) = \sum_{i \in F^*, j \in U^*} p_{\text{fold}}(j)K_{ji}p_{\text{eq}}(i)\nu_U(i)$ . Thus, the reactive flux is

$$\begin{aligned} J &= J_f(\Sigma) - J_b(\Sigma) = \sum_{i \in F^*, j \in U^*} [p_{\text{fold}}(i)K_{ij}p_{\text{eq}}(j)\nu_U(j) \\ &- p_{\text{fold}}(j)K_{ji}p_{\text{eq}}(i)\nu_U(i)]. \end{aligned} \quad (13)$$

From the time-reversal symmetry of the underlying microscopic dynamics it follows that, at equilibrium, for any trajectory fragment that goes from microstate  $j$  to microstate  $i$  in time  $t$ , there is a counterpart going in the opposite direction. As a consequence, the fraction of proteins in microstate  $i$ ,  $i \in I$ , that entered the bottleneck from the  $UU$ -basin,  $\nu_U(i)$ , is the same as the fraction of proteins that start from this microstate and reach the  $UU$ -basin before reaching the  $FF$ -basin, so that

$$\nu_U(i) = 1 - p_{\text{fold}}(i). \quad (14)$$

Using this relation and the condition of detailed balance,  $K_{ij}p_{\text{eq}}(j) = K_{ji}p_{\text{eq}}(i)$ , in Eq. (13), we recover our result for the reactive flux in Eq. (9).



### III. DISCUSSION

As the simplest example, consider the three-state scheme  $U \leftrightarrow I \leftrightarrow F$ , where  $k_{XY}$  is the rate constant for the transition  $Y \rightarrow X$  with  $X, Y = U, I, F$ . When the population of  $I$  is small, the rates  $k_F$  and  $k_U$  can be obtained from the three rate equations by setting  $dI/dt=0$ . For the folding rate this procedure yields  $k_F = k_{FI}k_{IU}/(k_{UI} + k_{FI})$ . If one chooses the dividing surface to be between the  $U$  and  $I$  states, our formalism [Eq. (9)] predicts that the reactive flux is  $J = [p_{\text{fold}}(I) - p_{\text{fold}}(U)]k_{IU}p_{\text{eq}}(U)$ . Since  $p_{\text{fold}}(U)=0$  and  $p_{\text{fold}}(I) = k_{FI}/(k_{UI} + k_{FI})$ ,  $k_F = J/p_{\text{eq}}(U)$  agrees with the above result obtained by making the steady-state assumption for the population of  $I$ . It can be readily verified that if one chooses the dividing surface to be between  $I$  and  $F$ , our formalism yields the identical expression for  $k_F$ . In fact, it can be shown in general that the reactive flux in Eq. (9) is the same for any dividing surface chosen so that any reactive trajectory unavoidably crosses this surface.

Whereas  $J$  is independent of the choice of  $\Sigma$ , individual contributions to the reactive flux may be both positive and negative. However, if we specify  $\Sigma$  using  $p_{\text{fold}}(i)$ , then all contributions have the same sign. Let us choose the dividing surface  $\Sigma_\alpha$  so that all microstates with  $p_{\text{fold}}(i) < \alpha$  are on the  $U$ -side of  $\Sigma_\alpha$ , while all microstates with  $p_{\text{fold}}(i) > \alpha$  are on the  $F$ -side of this dividing surface,  $0 < \alpha < 1$ . We denote the local reactive flux flowing in the  $U \rightarrow F$  direction from microstate  $i$  to microstates  $j$  that has a greater  $p_{\text{fold}}$  by  $J_{i \rightarrow j}$ ,

$$J_{i \rightarrow j} = K_{ji} p_{\text{eq}}(i) [p_{\text{fold}}(j) - p_{\text{fold}}(i)], \quad p_{\text{fold}}(j) > p_{\text{fold}}(i). \quad (15)$$

Since all fluxes  $J_{i \rightarrow j}$  crossing  $\Sigma_\alpha$  are positive, we can determine the importance of each transition based on the magnitude of its contribution to the reactive flux [Fig. 1(b)]. By repeating this procedure for a series of dividing surfaces  $\Sigma_\alpha$ , one can evaluate the importance of different folding/unfolding pathways. Of particular interest are transitions that cross the dividing surface with  $\alpha=1/2$ , since microstates involved in such transitions form the TSE. Using Eq. (9) one can ascertain the dynamically most relevant members of the TSE, i.e., microstates  $i$  and  $j$ , for which the local reactive flux is the largest. Interestingly, these microstates need not be those with the highest equilibrium populations. Thus, one can determine whether the reactive flux is localized (i.e., only a few pairs of microstates contribute) or delocalized (i.e., there are similar contributions from many transitions). Finally we note that when transitions occur between microstates with similar values of  $p_{\text{fold}}(i)$ , the difference of the splitting probabilities is much smaller than unity,  $[p_{\text{fold}}(j) - p_{\text{fold}}(i)] \ll 1$ . The reactive flux in Eq. (9) is then well below its TST upper bound given in Eq. (11), in which recrossings are ignored.

Instead of dividing the total reactive flux into contributions due to transitions crossing a dividing surface, we can decompose it into a sum of contributions due to unidirectional paths connecting the  $UU$ - and  $FF$ -basins. The local reactive flux between microstate  $i$  and  $j$  ( $p_{\text{fold}}(j) > p_{\text{fold}}(i)$ ) is positive,  $J_{i \rightarrow j} > 0$ , and directed in the  $U \rightarrow F$  direction. Because of the probability conservation, the sums of local re-

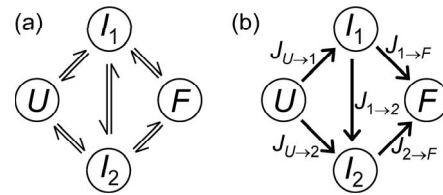


FIG. 2. Four-state model of coarse-grained protein dynamics. (a) Kinetic scheme with two intermediate (or bottleneck) microstates  $I_1$  and  $I_2$ . (b) Local unidirectional reactive fluxes for  $p_{\text{fold}}(I_2) \geq p_{\text{fold}}(I_1)$ .

active fluxes entering and exiting a microstate  $i$  are equal,  $\sum_k J_{k \rightarrow i} = \sum_j J_{i \rightarrow j} = J_i$ , where  $p_{\text{fold}}(k) < p_{\text{fold}}(i) < p_{\text{fold}}(j)$ . One can construct a directed graph that connects the  $UU$ - and  $FF$ -basins by following the local reactive fluxes [Fig. 1(b)]. The contribution of each unidirectional path on this graph to the total reactive flux can be determined as follows. A path that starts from  $i_0 \in UU$  and ends up at  $i_L \in FF$  passing through microstates  $i_1, i_2, \dots, i_{L-1} \in I$  contributes  $J(i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_{L-1} \rightarrow i_L) = J_{i_0 \rightarrow i_1} \prod_{k=1}^{L-1} (J_{i_k \rightarrow i_{k+1}} / J_{i_k})$  to the reactive flux. The sum of all “path” fluxes is equal to the total reactive flux  $J$ . The unidirectional path with the largest contribution can be found, for example, by using Dijkstra’s algorithm for the shortest path, identifying  $\ln(J_j / J_{i \rightarrow j})$  as the “length” associated with the transition from  $i$  to  $j$ , and  $\ln(J / J_{i_0 \rightarrow i_1})$  as the length associated with the entrance into the bottleneck from the  $UU$ -basin.

As an illustration consider the four-state model shown in Fig. 2(a). In this model the unfolded ( $U$ ) and folded ( $F$ ) states of the protein are separated by two bottleneck states  $I_1$  and  $I_2$ , through which all  $U$ -to- $F$  transitions occur. Given rate constants for which  $p_{\text{fold}}(I_2) \geq p_{\text{fold}}(I_1)$ , the local reactive fluxes obtained from Eq. (15) are shown in Fig. 2(b). There are four distinct dividing surfaces in the bottleneck region. Since the total reactive flux  $J$  is independent of the choice of the dividing surface we have

$$\begin{aligned} J &= J_{U \rightarrow I_1} + J_{U \rightarrow I_2} = J_{I_1 \rightarrow F} + J_{I_2 \rightarrow F} \\ &= J_{U \rightarrow I_2} + J_{I_1 \rightarrow F} + J_{I_1 \rightarrow I_2} \\ &= J_{U \rightarrow I_1} + J_{I_2 \rightarrow F} - J_{I_1 \rightarrow I_2}. \end{aligned} \quad (16)$$

From this it follows that  $J_{U \rightarrow I_1} = J_{I_1 \rightarrow I_2} + J_{I_1 \rightarrow F}$  and  $J_{I_2 \rightarrow F} = J_{U \rightarrow I_2} + J_{I_1 \rightarrow I_2}$ . These relations represent flux conservation at the nodes. Thus, there are only three independent local reactive fluxes. The total reactive flux can be partitioned among the three unidirectional “flux” paths as follows:  $J(U \rightarrow I_2 \rightarrow F) = J_{U \rightarrow I_2}$ ,  $J(U \rightarrow I_1 \rightarrow F) = J_{U \rightarrow I_1} [J_{I_1 \rightarrow F} / (J_{I_1 \rightarrow I_2} + J_{I_1 \rightarrow F})] = J_{I_1 \rightarrow F}$ , and  $J(U \rightarrow I_1 \rightarrow I_2 \rightarrow F) = J_{U \rightarrow I_1} [J_{I_1 \rightarrow I_2} / (J_{I_1 \rightarrow I_2} + J_{I_1 \rightarrow F})] = J_{I_1 \rightarrow I_2}$ . The sum of these fluxes is equal to  $J$ .

### IV. CONCLUDING REMARKS

To summarize, the phenomenological kinetic scheme in Eq. (2) describes the equilibration of the populations of two basins where the majority of the protein population is localized. The equilibration occurs via a sparsely populated bottleneck region, which connects these basins. Although the equilibration rate is determined by dynamics in the bottleneck, this region does not appear in the phenomenological description since the protein spends a small fraction of time

there. One of the main results of our analysis is the expression for the reactive flux in Eq. (9), which relates the phenomenological description of two-state protein folding and a more detailed description of the protein dynamics that involves transitions among discrete coarse-grained microstates. This expression allows one to determine how much each transition through a dividing surface contributes to the observed rate. Alternatively, one can divide the total reactive flux into a sum of contributions due to all directed paths connecting the definitely unfolded and definitely folded microstates. Although our primary focus has been on problems where the reactive flux is related to the phenomenological rate constants describing two-state kinetics, the expression in Eq. (9) is actually valid for an arbitrary partitioning of the microstates into three subsets. Specifically, it gives the number of trajectory fragments per unit time that leave the *UU* region and enter the *FF* region, irrespective of whether the intermediate microstates are sparsely populated or even exist.

## ACKNOWLEDGMENTS

This study was supported by the Intramural Research Program of the NIH, Center for Information Technology and National Institute of Diabetes and Digestive and Kidney Diseases.

<sup>1</sup>R. Zwanzig, A. Szabo, and B. Bagchi, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 20 (1992); R. Zwanzig, *ibid.* **92**, 9801 (1995).

<sup>2</sup>V. Munoz and W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11311 (1999); E. R. Henry and W. A. Eaton, *Chem. Phys.* **307**, 163 (2004).

<sup>3</sup>S. B. Ozkan, I. Bahar, and K. A. Dill, *Nat. Struct. Biol.* **8**, 765 (2001); J. Schonbrun and K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12678 (2003).

<sup>4</sup>W. Zhang and S. J. Chen, *J. Chem. Phys.* **119**, 8716 (2003).

<sup>5</sup>D. A. Evans and D. J. Wales, *J. Chem. Phys.* **121**, 1080 (2004).

<sup>6</sup>E. Pollak, A. Auerbach, and P. Talkner, *Biophys. J.* **95**, 4258 (2008).

<sup>7</sup>O. M. Becker and M. Karplus, *J. Chem. Phys.* **106**, 1495 (1997).

<sup>8</sup>M. Cieplak, M. Henkel, J. Karbowski, and J. R. Banavar, *Phys. Rev. Lett.* **80**, 3654 (1998).

<sup>9</sup>B. L. de Groot, X. Daura, A. E. Mark, and H. Grubmüller, *J. Mol. Biol.* **309**, 299 (2001).

<sup>10</sup>W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).

<sup>11</sup>N. Singhal, C. D. Snow, and V. S. Pande, *J. Chem. Phys.* **121**, 415 (2004); S. Park and V. S. Pande, *ibid.* **124**, 054118 (2006); J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *ibid.* **126**, 155101 (2007); N. Singhal Hinrichs and V. S. Pande, *ibid.* **126**, 244101 (2007).

<sup>12</sup>F. Rao and A. Caflisch, *J. Mol. Biol.* **342**, 299 (2004).

<sup>13</sup>D. S. Chekmarev, T. Ishida, and R. M. Levy, *J. Phys. Chem. B* **108**, 19487 (2004); M. Andrec, A. K. Felts, E. Gallicchio, and R. M. Levy, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6801 (2005).

<sup>14</sup>S. Sriraman, I. G. Kevrekidis, and G. Hummer, *J. Phys. Chem. B* **109**, 6479 (2005).

<sup>15</sup>N. V. Buchete and G. Hummer, *J. Phys. Chem. B* **112**, 6057 (2008).

<sup>16</sup>F. Noé, I. Horenko, C. Schütte, and J. C. Smith, *J. Chem. Phys.* **126**, 155102 (2007).

<sup>17</sup>A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).

<sup>18</sup>D. Ryter, *Physica A* **142**, 103 (1987); *J. Stat. Phys.* **49**, 751 (1987).

<sup>19</sup>M. M. Klosek, B. J. Matkowsky, and Z. Schuss, *Ber. Bunsenges. Phys. Chem.* **95**, 331 (1991); E. Pollak, A. M. Berezhkovskii, and Z. Schuss, *J. Chem. Phys.* **100**, 334 (1994); P. Talkner, *Chem. Phys.* **180**, 199 (1994); A. N. Drozdov and P. Talkner, *Phys. Rev. E* **54**, 1660 (1996); P. L. Geissler, C. Dellago, and D. Chandler, *J. Phys. Chem. B* **103**, 3706 (1999); P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002); G. Hummer, *J. Chem. Phys.* **120**, 516 (2004).

<sup>20</sup>R. Du, V. S. Pande, A. Yu. Grosberg, T. Tanaka, and E. Shakhnovich, *J. Chem. Phys.* **108**, 334 (1998).

<sup>21</sup>H. A. Kramers, *Physica (Amsterdam)* **7**, 284 (1940); P. Hänggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990); A. Nitzan, *Chemical Dynamics in Condensed Phases* (Oxford University Press, Oxford, 2006).

<sup>22</sup>D. Chandler, *J. Chem. Phys.* **68**, 2959 (1978).

<sup>23</sup>A. Berezhkovskii and A. Szabo, *J. Chem. Phys.* **121**, 9186 (2004); **122**, 079902 (2004).

<sup>24</sup>L. Onsager, *Phys. Rev.* **54**, 554 (1938).

<sup>25</sup>W. E and E. Vanden-Eijnden, *J. Stat. Phys.* **123**, 503 (2006).