

Perturbation theory of Φ -value analysis of two-state protein folding: Relation between p_{fold} and Φ values

Alexander Berezhkovskii^{a)}

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

Attila Szabo

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

(Received 31 May 2006; accepted 8 August 2006; published online 8 September 2006)

In protein folding, the transition state ensemble is defined as the set of conformations with $p_{\text{fold}} = 1/2$, where the p_{fold} of a conformation is the probability that starting from this conformation the protein folds before it unfolds. Experimentally, this ensemble is probed by the Φ -value analysis, where Φ is the ratio of the changes in the logarithms of the folding rate and the equilibrium constant when the system is perturbed by a mutation. We show that for a two-state protein the Φ value can be expressed in terms of the perturbation and only the first two eigenfunctions of the evolution operator (e.g., a rate matrix) of the wild-type protein. The first eigenfunction is the equilibrium probability distribution while the second is proportional to p_{fold} , thus establishing a formal relation between p_{fold} and Φ values. In addition to providing insight into the theoretical foundation of the Φ -value analysis, our results may prove practically useful in performing such analyses within the framework of models containing a large number of states. © 2006 American Institute of Physics. [DOI: 10.1063/1.2347708]

I. INTRODUCTION

The kinetics of two-state protein folding can be described by the scheme



where U and F denote unfolded and folded states while k_F^0 and k_U^0 are folding and unfolding rates (the superscript zero is used to refer to the wild-type protein). When such a protein is perturbed the Φ value is the ratio of the changes in the logarithms of the folding rate and the equilibrium constant,

$$\Phi = \frac{\ln(k_F/k_F^0)}{\ln(K_{\text{eq}}/K_{\text{eq}}^0)} = -\frac{\Delta \ln k_F}{\beta \Delta \Delta G_F}. \quad (1.2)$$

Here $K_{\text{eq}}^0 = k_F^0/k_U^0 = f_F^0/f_U^0$ is the equilibrium constant of the wild-type protein; f_I^0 , $I=F, U$, is the fractional occupancy of the state I at equilibrium, $f_F^0 + f_U^0 = 1$. The change in the free energy of folding is $\Delta \Delta G_F$; $\beta = 1/k_B T$, where k_B and T are the Boltzmann constant and the absolute temperature. The folding rate and equilibrium constant of the perturbed system are denoted as k_F and K_{eq} . In practice, the perturbations are a series of mutations and the resulting Φ values are used to infer the nature of the transition state and the mechanism of protein folding.¹⁻⁵

In this paper we show that when a *two-state* protein is perturbed the corresponding Φ value can be expressed in terms of the perturbation and the first two eigenfunctions of the evolution operator of the wild-type protein. A two-state

protein is special because there is a gap in the eigenvalue spectrum of its evolution operator. The first eigenvalue is zero, and the corresponding eigenfunction is the equilibrium distribution of the microstates. The second eigenvalue is much closer in magnitude to the first than to the third one. It determines the rate of equilibration of the populations of folded and unfolded states and is essentially equal to the sum of the folding and unfolding rate constants. To calculate a Φ value one must find the effect of the perturbation on the second eigenvalue (i.e., the reciprocal of the relaxation time) and on the first eigenfunction (i.e., the equilibrium distribution of the microstates). Because the eigenvalue spectrum has a gap, this can be done to an excellent approximation using only the first two eigenfunctions of the unperturbed system.

The transition state ensemble is defined as a set of conformations with $p_{\text{fold}} = 1/2$, where the p_{fold} of a conformation is the probability that starting from this conformation the protein folds before it unfolds.⁶ The idea of using the splitting or commitment probabilities to identify the transition state (i.e., the dividing surface separating reactants from products) was first suggested in the theory of activated rate processes governed by stochastic dynamics by Ryter⁷ and has been exploited in the study of chemical reactions in solutions.⁸⁻¹³ The Φ value contains information about the transition state ensemble because for a two-state protein the second eigenvector of the wild-type evolution operator is related to p_{fold} .¹⁴

The outline of the paper is as follows. We derive our key

^{a)}Electronic mail: berezh@mail.nih.gov

results in the next section. Two simple illustrative examples are discussed in Sec. III. Some concluding remarks are given in the last section.

II. THEORY

Let $p_0(\mathbf{r}, t)$ be the probability density of the multidimensional conformational coordinate \mathbf{r} of the wild-type protein at time t . The dynamics is described by an evolution equation of the form

$$\frac{\partial p_0}{\partial t} = \mathbf{L}_0 p_0, \quad (2.1)$$

where \mathbf{L}_0 is the evolution operator. The equilibrium distribution $p_{\text{eq}}^0(\mathbf{r})$ satisfies $\mathbf{L}_0 p_{\text{eq}}^0(\mathbf{r}) = 0$ and is normalized to unity.

Let us consider two specific examples of the operator \mathbf{L}_0 . If the dynamics can be described as diffusion under the influence of a potential of mean force $V_0(\mathbf{r})$ then

$$\mathbf{L}_0 = \nabla \cdot \mathbf{D} \exp[-\beta V_0(\mathbf{r})] \cdot \nabla \exp[\beta V_0(\mathbf{r})], \quad (2.2)$$

where \mathbf{D} is the diffusion tensor. If one adopts a more coarse-grained picture where the conformational coordinate is discrete and the dynamics is described by the rate equations of chemical kinetics then

$$\mathbf{L}_0 = \mathbf{K}^0, \quad (2.3)$$

where \mathbf{K}^0 is the rate matrix. The off-diagonal element K_{ij}^0 is the rate constant for the transition from conformation j to conformation i . The diagonal element K_{jj}^0 is given by $K_{jj}^0 = -\sum_{i \neq j} K_{ij}^0$. The matrix elements satisfy the condition of detailed balance $K_{ij}^0 p_{\text{eq}}^0(j) = K_{ji}^0 p_{\text{eq}}^0(i)$.

The eigenvalue problem for the operator \mathbf{L}_0 is

$$\mathbf{L}_0 \chi_n^0 = -\lambda_n^0 \chi_n^0, \quad n = 1, 2, \dots, \quad (2.4)$$

where χ_n^0 is the eigenfunction corresponding to the eigenvalue $-\lambda_n^0$. We restrict ourselves to the class of evolution operators for which all eigenvalues are real and nonpositive. The eigenstates are labeled in increasing magnitude of the eigenvalues so that $\lambda_1^0 = 0 < \lambda_2^0 < \lambda_3^0 < \dots$. The first eigenvalue is zero and the corresponding eigenfunction is the equilibrium distribution, $\chi_1^0 = p_{\text{eq}}^0(\mathbf{r})$, $\int p_{\text{eq}}^0(\mathbf{r}) d\mathbf{r} = 1$. The eigenfunctions are normalized in the sense

$$\int p_{\text{eq}}^0(\mathbf{r})^{-1} \chi_m^0 \chi_n^0 d\mathbf{r} = \delta_{mn}. \quad (2.5)$$

Multiplying both sides of Eq. (2.4) by $(p_{\text{eq}}^0)^{-1} \chi_n^0$ and integrating over \mathbf{r} we obtain

$$\lambda_n^0 = - \int (p_{\text{eq}}^0)^{-1} \chi_n^0 \mathbf{L}_0 \chi_n^0 d\mathbf{r}. \quad (2.6)$$

Two-state proteins are special because the equilibration of folded and unfolded populations occurs much more slowly than local equilibration within each of these states. This hierarchy of time scales is manifested by the existence of a gap in the eigenvalue spectrum between the first two nonzero eigenvalues, namely, $\lambda_2^0 - \lambda_1^0 = \lambda_2^0 \ll \lambda_3^0 - \lambda_2^0$. The folding relaxation rate is determined by the second eigenvalue,

$$\lambda_2^0 \approx k_F^0 + k_U^0 \ll \lambda_3^0. \quad (2.7)$$

When a protein is perturbed the dynamics changes and the evolution operator is modified,

$$\mathbf{L} = \mathbf{L}_0 + \delta\mathbf{L}. \quad (2.8)$$

The operator $\delta\mathbf{L}$ will be assumed to be a small perturbation. To find a Φ value one needs to know the second eigenvalue λ_2 and the first eigenfunction $\chi_1 = p_{\text{eq}}(\mathbf{r})$ of the operator \mathbf{L} . Because of the gap in the eigenvalue spectrum we can solve the eigenvalue problem,

$$\mathbf{L} \chi = -\lambda \chi, \quad (2.9)$$

to a good approximation in the basis of the first two eigenfunctions of the unperturbed operator \mathbf{L}_0 . Substituting

$$\chi = c_1 \chi_1^0 + c_2 \chi_2^0 \quad (2.10)$$

into Eq. (2.9), multiplying both sides by $(p_{\text{eq}}^0)^{-1} \chi_1^0$ and $(p_{\text{eq}}^0)^{-1} \chi_2^0$, and integrating over \mathbf{r} we obtain

$$\begin{pmatrix} 0 & 0 \\ \langle 2 | \delta\mathbf{L} | 1 \rangle & -\lambda_2^0 + \langle 2 | \delta\mathbf{L} | 2 \rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = -\lambda \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (2.11)$$

where we have defined the matrix elements by

$$\langle m | \delta\mathbf{L} | n \rangle = \int (p_{\text{eq}}^0)^{-1} \chi_m^0 \delta\mathbf{L} \chi_n^0 d\mathbf{r}. \quad (2.12)$$

Solving the set of equations in Eq. (2.11) one finds that $\lambda_1 = 0$, as it must be, and

$$\lambda_2 = \lambda_2^0 - \langle 2 | \delta\mathbf{L} | 2 \rangle. \quad (2.13)$$

This is identical to the expression for λ_2 given by the first-order perturbation theory. The first eigenfunction is

$$\chi_1(\mathbf{r}) = p_{\text{eq}}(\mathbf{r}) = p_{\text{eq}}^0(\mathbf{r}) + \frac{\chi_2^0(\mathbf{r})}{\lambda_2} \langle 2 | \delta\mathbf{L} | 1 \rangle. \quad (2.14)$$

When λ_2 is replaced by λ_2^0 , the second term is identical to the dominant term in the first-order eigenfunction given by perturbation theory.

The equilibrium fraction of folded proteins is insensitive to the precise location of the transition state, and so we can assume here that the dividing surface between the folded and unfolded states is unaltered by the perturbation. Integrating both sides of Eq. (2.14) over the ensemble of folded conformations we have

$$f_F = \int_F p_{\text{eq}} d\mathbf{r} = f_F^0 + \frac{1}{\lambda_2} \langle 2 | \delta\mathbf{L} | 1 \rangle \int_F \chi_2^0 d\mathbf{r}. \quad (2.15)$$

The expression for the Φ value in Eq. (1.2) can be written in terms of λ_2 , λ_2^0 , f_F , and f_F^0 as

$$\Phi = \frac{\ln[\lambda_2 f_F / (\lambda_2^0 f_F^0)]}{\ln[f_F(1 - f_F^0) / f_F^0(1 - f_F)]}. \quad (2.16)$$

Substituting λ_2 and f_F given in Eqs. (2.13) and (2.15) into Eq. (2.16) and linearizing the result with respect to the matrix elements of $\delta\mathbf{L}$, we obtain

$$\Phi = 1 - f_F^0 - \frac{\langle 2|\delta\mathbf{L}|2\rangle f_F^0(1-f_F^0)}{\langle 2|\delta\mathbf{L}|1\rangle \int_F \chi_2^0 d\mathbf{r}}. \quad (2.17)$$

For a two-state protein,

$$\int_F \chi_2^0 d\mathbf{r} = \sqrt{f_F^0(1-f_F^0)}, \quad (2.18)$$

to a good approximation.¹⁴ Using this and the relation $f_F^0 = K_{\text{eq}}^0/(1+K_{\text{eq}}^0)$ we finally have

$$\Phi = \frac{1}{1+K_{\text{eq}}^0} \left[1 - \sqrt{K_{\text{eq}}^0} \frac{\int (\rho_{\text{eq}}^0)^{-1} \chi_2^0 \delta\mathbf{L} \chi_2^0 d\mathbf{r}}{\int (\rho_{\text{eq}}^0)^{-1} \chi_2^0 \delta\mathbf{L} \rho_{\text{eq}}^0 d\mathbf{r}} \right]. \quad (2.19)$$

For a two-state protein the second eigenfunction χ_2^0 is approximately proportional to the wild-type p_{fold} . Specifically¹⁴

$$\chi_2^0(\mathbf{r}) \approx \frac{p_{\text{fold}}^0(\mathbf{r}) - f_F^0}{\sqrt{f_F^0(1-f_F^0)}} p_{\text{eq}}^0(\mathbf{r}). \quad (2.20)$$

This is the formal reason why a Φ value contains information about the transition state ensemble, defined as the set of conformations with $p_{\text{fold}} = 1/2$.

Finally we note that within the framework of the above formalism it is possible to determine how the location of the transition state changes as a result of a mutation (i.e., how p_{fold} changes). Starting with the second eigenfunction χ_2 of the matrix in Eq. (2.11), which is related to p_{fold} via the analog of Eq. (2.20), it can be shown that

$$p_{\text{fold}}(\mathbf{r}) = \frac{p_{\text{fold}}^0(\mathbf{r})}{p_{\text{fold}}^0(\mathbf{r}) + \exp(\beta\Delta\Delta G_F)(1-p_{\text{fold}}^0(\mathbf{r}))}. \quad (2.21)$$

This expression shows that a small perturbation does not alter the p_{fold} values of the folded and unfolded states and affects only intermediate states in the vicinity of the wild-type transition state. It predicts that the p_{fold} of such an intermediate state rapidly goes to zero as $\Delta\Delta G_F$ increases and quickly approaches unity as $\Delta\Delta G_F$ decreases. Even though this limiting behavior is just what one would expect, Eq. (2.21) is useful only when $\Delta\Delta G_F$ is small ($\beta\Delta\Delta G_F \ll 1$).

III. ILLUSTRATIVE EXAMPLES

Suppose we model protein dynamics as one-dimensional diffusion along a reaction coordinate x in the presence of a double-well potential of mean force $V_0(x)$. A mutation changes this potential to $V(x)$. From the one-dimensional version of Eq. (2.2) one can see that

$$\delta\mathbf{L} = \beta D \frac{\partial}{\partial x} \left[\frac{d\delta V(x)}{dx} \right], \quad (3.1)$$

where $\delta V(x) = V(x) - V_0(x)$. In this case the expression for Φ in Eq. (2.19) becomes

$$\Phi = \frac{1}{1+K_{\text{eq}}^0} - \frac{\int p_{\text{eq}}^0 \delta V [(\chi_2^0/p_{\text{eq}}^0)^2 - D(d(\chi_2^0/p_{\text{eq}}^0)/dx)^2/\lambda_2^0] dx}{(1+K_{\text{eq}}^0) \int \delta V \chi_2^0 dx / \sqrt{K_{\text{eq}}^0}}, \quad (3.2)$$

where we have integrated by parts and used Eq. (2.4). To write this expression in terms of $p_{\text{fold}}^0(x)$ we use the relation between χ_2^0 and p_{fold}^0 in Eq. (2.20). The result is

$$\Phi = \frac{\int p_{\text{eq}}^0 \delta V [(p_{\text{fold}}^0 - f_F^0)(1-p_{\text{fold}}^0) + D(dp_{\text{fold}}^0/dx)^2/\lambda_2^0] dx}{\int p_{\text{eq}}^0 \delta V (p_{\text{fold}}^0 - f_F^0) dx}, \quad (3.3)$$

Using the same relation in Eq. (2.6) and integrating by parts we obtain

$$\lambda_2^0 = \frac{D}{f_F^0(1-f_F^0)} \int p_{\text{eq}}^0 \left(\frac{dp_{\text{fold}}^0}{dx} \right)^2 dx. \quad (3.4)$$

These two expressions give the Φ value in terms of $p_{\text{fold}}^0(x)$ and $p_{\text{eq}}^0(x)$ that characterize the wild-type protein and the perturbation $\delta V(x)$.

We now use the approximations,

$$\int p_{\text{eq}}^0 \delta V p_{\text{fold}}^0 dx = f_F^0 \langle \delta V \rangle_F, \quad (3.5a)$$

$$\int p_{\text{eq}}^0 \delta V p_{\text{fold}}^0 (1-p_{\text{fold}}^0) dx = 0, \quad (3.5b)$$

and find that

$$\Phi = \frac{\int p_{\text{eq}}^0 (\delta V - \langle \delta V \rangle_U) (dp_{\text{fold}}^0/dx)^2 dx}{(\langle \delta V \rangle_F - \langle \delta V \rangle_U) \int p_{\text{eq}}^0 (dp_{\text{fold}}^0/dx)^2 dx}, \quad (3.6)$$

where we have defined

$$\langle \delta V \rangle_I = \frac{\int_I \delta V(x) \exp[-\beta V_0(x)] dx}{\int_I \exp[-\beta V_0(x)] dx}, \quad I = F, U. \quad (3.7)$$

The approximations in Eq. (3.5) are accurate because in a two-state system $p_{\text{fold}}^0(x)$ differs from a step function only in the barrier region where $p_{\text{eq}}^0(x)$ is very small.

Finally we note that $dp_{\text{fold}}^0(x)/dx$ is zero except in the barrier region. By solving $L_0 \exp(-\beta V_0) p_{\text{fold}}^0 = 0$ in this region we find

$$\frac{dp_{\text{fold}}^0(x)}{dx} = \frac{\exp[\beta V_0(x)]}{\int_{\neq} \exp[\beta V_0(x)] dx}, \quad (3.8)$$

where “ \neq ” indicates that the integration is over only the barrier region. Using this in Eq. (3.6) we obtain

$$\Phi = \frac{\langle \delta V \rangle_{\neq} - \langle \delta V \rangle_U}{\langle \delta V \rangle_F - \langle \delta V \rangle_U}, \quad (3.9)$$

where

$$\langle \delta V \rangle_{\neq} = \frac{\int_{\neq} \delta V(x) \exp[\beta V_0(x)] dx}{\int_{\neq} \exp[\beta V_0(x)] dx}. \quad (3.10)$$

Note that $\langle \delta V \rangle_{\neq}$ involves the reciprocal of the Boltzmann factor in the barrier region.

The above expression for Φ provides an important consistency check of our formalism because it can be readily derived in the framework of the Kramers theory of the reaction rates.¹⁵ It can be shown that using the following expression for the rate constant:

$$k_I = \frac{D}{\left\{ \int_J \exp[-\beta V(x)] dx \right\} \left\{ \int_{\neq} \exp[\beta V(x)] dx \right\}}, \quad (3.11)$$

$$J = U \quad \text{for } I = F,$$

$$J = F \quad \text{for } I = U,$$

with $V(x) = V_0(x) + \delta V(x)$, in the definition of Φ and linearizing the result with respect to δV we recover Eq. (3.9).

When the dynamics of the protein is described in terms of transitions among discrete microstates (conformations) $\delta \mathbf{L}$ is given by the difference of the rate matrices of the mutant (\mathbf{K}) and wild-type (\mathbf{K}^0) proteins $\delta \mathbf{K} = \mathbf{K} - \mathbf{K}^0$. The expression for the Φ value in Eq. (2.19) then becomes

$$\Phi = \frac{1}{1 + K_{\text{eq}}^0} \left[1 - \frac{\sum_{ij} (p_{\text{eq}}^0(i))^{-1} \chi_2^0(i) \delta K_{ij} \chi_2^0(j)}{\sum_{ij} (p_{\text{eq}}^0(i))^{-1} \chi_2^0(i) \delta K_{ij} p_{\text{eq}}^0(j)} \right]. \quad (3.12)$$

In general, the mutation of a residue changes many rate constants. For illustrative purposes let us assume that somehow one could change only the rate constant for the transition from conformation l to conformation m , $m \neq l$. Then $\delta K_{ij} = (K_{ml} - K_{ml}^0)(\delta_{im} - \delta_{il})\delta_{ij}$ and Eq. (3.12) becomes

$$\Phi = 1 - p_{\text{fold}}^0(l). \quad (3.13)$$

Note that the Φ value in this case depends only on the starting conformation. If this conformation belongs to the $U(F)$ state of the protein then $\Phi = 1(0)$. If it is a member of the transition state ensemble then $\Phi = 1/2$.

As another example let us consider a system with an intermediate state I whose equilibrium population is negligible,



The rate constants are assumed to be given by $k_{ij}^0 = \nu_j^0 \exp[-\beta(V_{ij}^{\neq 0} - V_j^0)]$, $i = U, I, F$, where ν_j^0 is the preexponential factor and $V_{ij}^{\neq 0} - V_j^0$ is the height of the barrier that the system has to overcome when jumping from state j to state i . Note that $V_{ij}^{\neq 0} = V_{ji}^{\neq 0}$ since transitions in both directions involve the same barrier. We assume that $k_{IF}^0, k_{IU}^0 \ll k_{FI}^0, k_{UI}^0$ which guarantees that the equilibrium population of the intermediate state is low, $p_{\text{eq}}^0(I) \ll p_{\text{eq}}^0(F), p_{\text{eq}}^0(U)$, and that the kinetics is essentially single exponential.

For this system,

$$p_{\text{fold}}^0 = \begin{pmatrix} 0 \\ p_{\text{fold}}^0(I) \\ 1 \end{pmatrix}, \quad p_{\text{fold}}^0(I) = \frac{k_{FI}^0}{k_{FI}^0 + k_{UI}^0}, \quad (3.15)$$

and one can use this to find χ_2^0 by Eq. (2.20) and then the Φ value via Eq. (3.12). Assuming that the perturbation changes only the barrier heights and does not change the preexponential factors, $\delta K_{ij} \approx k_{ij}^0 \beta (\Delta V_j - \Delta V_{ij}^{\neq})$, $i \neq j$, where ΔV_j and ΔV_{ij}^{\neq} are variations of the energy of state j and the energy barrier between states i and j , respectively. Eventually we obtain

$$\Phi = \frac{p_{\text{fold}}^0(I) \Delta V_{UI}^{\neq} + [1 - p_{\text{fold}}^0(I)] \Delta V_{FI}^{\neq} - \Delta V_U}{\Delta V_F - \Delta V_U}, \quad (3.16)$$

where $p_{\text{fold}}^0(I)$ is given in Eq. (3.15). In the special case that $\Delta V_U = \Delta V_{UI}^{\neq}$ and $\Delta V_F = \Delta V_{FI}^{\neq}$, we are changing only the rates from state I , and this result reduces to Eq. (3.13) with $l = I$.

If we further invoke the commonly used linear free energy relationships $\Delta V_{FI}^{\neq} = (\Delta V_F + \Delta V_I)/2$ and $\Delta V_{UI}^{\neq} = (\Delta V_U + \Delta V_I)/2$ we arrive at

$$\Phi = \frac{1}{2} \left[1 - p_{\text{fold}}^0(I) + \frac{\Delta V_I - \Delta V_U}{\Delta V_F - \Delta V_U} \right]. \quad (3.17)$$

For $\Delta V_I = \Delta V_U$ this expression predicts that $\Phi = [1 - p_{\text{fold}}^0(I)]/2$ rather than $\Phi = 0$ as would be expected at first sight. The reason for this apparent discrepancy can be traced back to our use of the linear free energy relation for the change of the energy barrier between the I and F states. Variation in the energy of the I state changes the energy of this barrier and, hence, the folding rate.

IV. CONCLUDING REMARKS

In this paper we have examined the formal theoretical foundation of Φ -value analysis using perturbation theory. The main results of our analysis are the expressions for the Φ value in Eqs. (2.19) and (3.12) for continuous and discrete conformational dynamics, respectively. The key step, which allowed us to derive these results, is the solution of the eigenvalue problem in Eq. (2.9) in the basis of the first two eigenfunctions of the wild-type protein. This simplification is justified for a two-state protein because of the gap in the

eigenvalue spectrum of its evolution operator that is a consequence of the slow equilibration of the populations of the folded and unfolded states relative to local equilibration within each of these states. The existence of this gap also implies¹⁴ that the second eigenfunction and hence the Φ value is related to p_{fold} , the quantity which can be used to define the transition state ensemble. The implication of our work on various qualitative interpretations of Φ values found in the literature remains to be investigated.

Although this paper is rather formal, our results should prove practically useful in performing Φ -value analysis within the framework of a microscopic model involving a large number of discrete states. Suppose one determines $\delta\mathbf{K}=\mathbf{K}-\mathbf{K}^0$ for several different perturbations either from a statistical mechanical model or from large-scale computer simulations.¹⁶⁻¹⁹ Then we can determine all the resulting Φ values using only the first two eigenfunctions of \mathbf{K}^0 . Computationally it is more convenient to work with symmetrized rate matrices $H_{ij}^0=K_{ij}^0\sqrt{K_{ji}^0/K_{ij}^0}$ (Ref. 20) whose first two eigenfunctions satisfy

$$\sum_j H_{ij}^0 \Psi_n^0(j) = -\lambda_n^0 \Psi_n^0(i), \quad n=1,2, \quad (4.1)$$

$$\sum_i \Psi_m^0(i) \Psi_n^0(i) = \delta_{mn}. \quad (4.2)$$

The equilibrium distribution is $p_{\text{eq}}^0(l)=\Psi_1^0(l)^2$. When the system is perturbed \mathbf{H}^0 changes. For example, in the special case when $K_{ij}^0=\nu \exp[-\beta(V_i^0-V_j^0)/2]$ and the perturbation changes only the energies, the difference $\delta\mathbf{H}=\mathbf{H}-\mathbf{H}^0$ turns out to be a diagonal matrix. In general, the Φ values can be found using Eq. (2.16) in conjunction with discrete analogs of Eqs. (2.13) and (2.14),

$$\lambda_2 = \lambda_2^0 - \sum_{ij} \Psi_2^0(i) \delta H_{ij} \Psi_2^0(j), \quad (4.3)$$

$$p_{\text{eq}}(l) = p_{\text{eq}}^0(l) + \frac{\Psi_1^0(l) \Psi_2^0(l)}{\lambda_2} \sum_{ij} \Psi_2^0(i) \delta H_{ij} \Psi_1^0(j). \quad (4.4)$$

This is clearly computationally less demanding than having to re-diagonalize the rate matrix corresponding to each muta-

tion. In addition, the validity of this procedure can be checked by generalizing our analysis to include additional wild-type eigenfunctions.

ACKNOWLEDGMENT

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.

- ¹A. Matouschek, J. T. Kellies, Jr., L. Serrano, and A. R. Fersht, *Nature* (London) **340**, 122 (1989).
- ²A. R. Fersht, A. Matouschek, and L. Serrano, *J. Mol. Biol.* **224**, 771 (1992).
- ³A. Matouschek, D. E. Otzen, L. S. Itzhaki, S. E. Jackson, and A. R. Fersht, *Biochemistry* **34**, 13656 (1995).
- ⁴A. R. Fersht, *Curr. Opin. Struct. Biol.* **5**, 79 (1995).
- ⁵A. R. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, San Francisco, 1999).
- ⁶R. Du, V. S. Pande, A. Yu. Grosberg, T. Tanaka, and E. Shakhnovich, *J. Chem. Phys.* **108**, 334 (1998).
- ⁷D. Ryter, *Physica A* **142**, 103 (1987); *J. Stat. Phys.* **49**, 751 (1987).
- ⁸M. M. Klosek, B. J. Matkowsky, and Z. Schuss, *Ber. Bunsenges. Phys. Chem.* **95**, 331 (1991).
- ⁹E. Pollak, A. M. Berezhkovskii, and Z. Schuss, *J. Chem. Phys.* **100**, 334 (1994).
- ¹⁰P. Talkner, *Chem. Phys.* **180**, 199 (1994); A. N. Drozdov and P. Talkner, *Phys. Rev. E* **54**, 1660 (1996).
- ¹¹P. L. Geissler, C. Delago, and D. Chandler, *J. Phys. Chem. B* **103**, 3706 (1999).
- ¹²P. G. Bolhuis, D. Chandler, C. Delago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- ¹³G. Hummer, *J. Chem. Phys.* **120**, 516 (2004).
- ¹⁴A. Berezhkovskii and A. Szabo, *J. Chem. Phys.* **121**, 9186 (2005); **122**, 079902 (2006).
- ¹⁵H. A. Kramers, *Physica* (Amsterdam) **7**, 284 (1940).
- ¹⁶W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).
- ¹⁷W. C. Swope, J. W. Pitera, F. Suits *et al.*, *J. Phys. Chem. B* **108**, 6582 (2004).
- ¹⁸S. Sriraman, I. G. Kevrekidis, and G. Hummer, *J. Phys. Chem. B* **109**, 6479 (2005).
- ¹⁹S. P. Elmer, S. Park, and V. S. Pande, *J. Chem. Phys.* **123**, 114903 (2005).
- ²⁰R. J. Wittebort and A. Szabo, *J. Chem. Phys.* **69**, 1722 (1978).