Quantifying Kinetic Paths of Protein Folding

Jin Wang,*† Kun Zhang,* Hongyang Lu,* and Erkang Wang*
*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin 130021, People's Republic of China; and †Department of Chemistry and Physics, State University of New York, Stony Brook, New York 11794

ABSTRACT We propose a new approach to activated protein folding dynamics via a diffusive path integral framework. The important issues of kinetic paths in this situation can be directly addressed. This leads to the identification of the kinetic paths of the activated folding process, and provides a direct tool and language for the theoretical and experimental community to understand the problem better. The kinetic paths giving the dominant contributions to the long-time folding activation dynamics can be quantitatively determined. These are shown to be the instanton paths. The contributions of these instanton paths to the kinetics lead to the "bell-like" shape folding rate dependence on temperature, which is in good agreement with folding kinetic experiments and simulations. The connections to other approaches as well as the experiments of the protein folding kinetics are discussed.

INTRODUCTION

Protein folding is one of the most important issues in modern molecular biology. Studying the dynamics is essential in understanding how the protein folds. The question is how the many conformational degrees of freedom can converge to the native state in a finite biological time scale (millisecond to second) instead of cosmological timescale (1). The energy landscape theory of protein folding resolves this issue naturally by assuming there is a bias or funnel toward the native state (2–4). This bias is believed to be from the natural evolution. Superimposed on the funneled landscape are the local traps. The slope of the funnel must be steep enough to overcome the traps to reach the folded state. The energy landscape theory is successful in explaining many experiments (5) at both the qualitative and quantitative levels.

According to the energy landscape theory, in general at the initial stage of folding, there are multiple paths toward native state. The discrete paths emerge only when the landscape becomes rough and local traps are important at late stage of folding. Searching for kinetic paths has been a central issue for the folding experimental community for many years (6–18). Unfortunately, most of the current kinetic folding studies are formulated in terms of the rate dynamics giving only the end results, rather than the paths that represent the full intermediate histories connecting the initial and final ends. It is, therefore, important and natural to formulate the theory in terms of path language. Such a formulation would help to resolve the challenging kinetic path issue of the folding problem and provide a direct tool and language for the theoretical and experimental community to understand each other better. Another advantage of using paths is that the direct integration over paths is normally easier computationally than solving differential equations locally in microscopic details.

Path integral methods since first appearing (19,20) have been successfully applied to many areas in physics (20–22) and chemistry (23–25). There are so far very limited studies on folding paths. Wang et al. (4) have studied a downhill folding process (very steep funnel) without activation barrier. It is shown that there exists a multiple path to discrete path transition at a temperature higher than the thermodynamic glassy trapping temperature. The relevance to single molecule dynamics is studied (26,27). Olender and Elber and Elber et al. (28,29) studied peptide folding with atomic level simulations and identify some key paths. The purpose of this study is to formulate a diffusive path integral framework for the general case where there exists activation free-energy barriers on the folding landscape, and to identify and quantify the dominant path contributions to the kinetics.

METHOD AND MATHEMATICAL DETAILS

For mathematical simplicity, we study the protein folding problem not at atomic level but at the coarse-grained level—the residue-residue level. This will reduce significantly the computational unforeseeable tasks without the loss of too many important universal features and serve as a guiding force for the more detailed atomic level investigations.

Let us turn to a model Hamiltonian that describes protein folding. To first order approximation, we assume that the energetics that favors bringing two or multiple residues close together from the protein is due mainly to the short-range (in space) hydrophobic driving force. The form of the interactions is $-\epsilon_{ijk\ldots p}(\alpha_i, \alpha_j, \alpha_k, \ldots \alpha_p, r_i, r_j, r_k, \ldots, r_p)$, where $\epsilon_{ijk\ldots p}$ is the multibody coupling strength, r_i is the position of the ith residue, and α_i represents the physical properties of the residue i, for example, hydrophobic charges, etc. Here,

Submitted October 28, 2004, and accepted for publication June 17, 2005.
Address reprint requests to Jin Wang, E-mail: jin.wang.1@stonybrook.edu.

0006-3495/05/09/1612/09 $2.00
doi: 10.1529/biophysj.104.055186

we also assume that the environmental solvent effects are already averaged out, resulting in the multibody cooperative hydrophobic interactions among residues upon folding.

We may write down the Hamiltonian energy function of a polypeptide sequence as:

$$H = -\sum_{\mathrm{ijk\ldots p}} \epsilon_{\mathrm{ijk\ldots p}} \sigma_{\mathrm{ijk\ldots p}}, \tag{1}$$

where $\sigma_{\mathrm{ijk\ldots p}} = 1$ when there is a multibody contact adjacent in space made among monomers ijk…p and $\sigma_{\mathrm{ijk\ldots p}} = 0$ otherwise. $\sum$ is up to N, N is the length of the polypeptide sequence, and $E_{\mathrm{ijk\ldots p}}$ is quite random due to the sequence and interaction heterogeneity. Notice that this is mathematically closely related to the random energy model (30,31).

Suppose there exists a native configurational state n of energy E_{n}. We can find the probability that configuration a has energy E_{a}, given that a has an overlap Q with n, where Q is the fraction of native contacts of state a: $Q = (1/N)\sum_{\mathrm{ij}} \sigma^{\mathrm{a}}_{\mathrm{ij}} \sigma^{\mathrm{n}}_{\mathrm{ij}}$ and N is the total number of native contacts. Q can be used as an order parameter or a reaction coordinate for the physical folding process that measures how close the states are toward native state. Note that for $Q = 1$, the state is in the native folded state and for $Q = 0$, the configurations are in totally nonnative unfolded states.

The conditional probability is obtained directly by averaging over the Gaussian distribution of contact energy $\epsilon_{ijk\ldots p}$ $(\langle \delta[E_{\mathrm{a}} - H(\{\sigma^{\mathrm{a}}_{\mathrm{ijk\ldots p}}\})]\delta[E_{\mathrm{n}} - H(\{\sigma^{\mathrm{n}}_{\mathrm{ijk\ldots p}}\})]\rangle / \langle \delta[E_{\mathrm{n}} - H(\{\sigma^{\mathrm{n}}_{\mathrm{ijk\ldots p}}\})]\rangle)$. By approximating the cooperative multibody interactions $\sigma_{\mathrm{ijk\ldots p}}$ in the Hamiltonian into the factorization of pair interaction terms $\sigma_{\mathrm{ij}}\sigma_{\mathrm{jk}}\ldots$ through a suitable decomposition law such as in the superposition approximation in the theory of fluid, the expression can be simplified as: $(P_{\mathrm{an}}(E_{\mathrm{a}}, Q, E_{\mathrm{n}})/P_{\mathrm{n}}(E_{\mathrm{n}})) \sim exp[-([(E_{\mathrm{a}} - \bar{E}) - Q^{\mathrm{m}-1}(E_{\mathrm{n}} - \bar{E})]^2 / 2N\Delta\epsilon^2(1 - Q^{2(\mathrm{m}-1)}))]$, where m is the order of the interactions ($m = 2$ for two-body interactions, $m = 3$ for three-body interactions, and $m = p$ for p-body interactions), $\bar{E}$ is the average mean energy, and $\Delta\epsilon$ is the effective width of the energy distribution per contact.

The configurational entropy S_{tot} as a function of similarity Q with a given state is treated in details by the previous studies (32,33).

Given the $S_{\mathrm{tot}}(Q)$ and conditional probability distribution obtained earlier, the average numbers of states of energy E and overlap Q with native state n is: $\langle n(E, Q, E_{\mathrm{n}})\rangle = exp[S_{\mathrm{tot}}(Q)](P(E, Q, E_{\mathrm{n}})/P(E_{\mathrm{n}}))$. This is effectively the microcanonical ensemble description of the thermodynamics. At each stratum of the order parameter or reaction coordinate Q, the set of states is modeled by a random energy model. By the thermodynamic relation of $(\partial log\langle n(E, Q, E_{\mathrm{n}})\rangle/\partial E) = 1/T$, we can obtain the energy and entropy of the biomolecular folding as: $E(T, Q, E_{\mathrm{n}}) = \bar{E} + Q^{\mathrm{m}-1}(E_{\mathrm{n}} - \bar{E}) - (N\Delta\epsilon^2(1 - Q^{2(\mathrm{m}-1)})/T)$ and $S(T, Q, E_{\mathrm{n}}) = Ns_{\mathrm{tot}}(Q) - (N\Delta\epsilon^2(1 - Q^{2(\mathrm{m}-1)})/2T^2)$, where $s_{\mathrm{tot}}(Q) = S_{\mathrm{tot}}(Q)/N$. The entropy vanishes at a characteristic temperature: $T_{\mathrm{g}} = \Delta\epsilon\sqrt{((1 - Q^{2(\mathrm{m}-1)})/2s_{\mathrm{tot}}(Q))}$, which signals the trapping of the polypeptide chain into a low-energy conformational state within the stratum characterized by Q. Notice that when $Q = 0$ (nonnative unfolded states), $T_{\mathrm{g}} = \Delta\epsilon\sqrt{(1/2s_{\mathrm{tot}}(Q = 0))}$.

From the thermodynamic expression of the energy and the entropy given above, we can easily obtain the expression for the free energy per contact as (33): $(F/N)(T, Q, E_{\mathrm{N}}) = -Ts_{\mathrm{tot}}(Q) - Q^{\mathrm{m}-1}\delta\epsilon_{\mathrm{n}} - (\Delta\epsilon^2/2T)(1 - Q^{2(\mathrm{m}-1)})$, where $\delta\epsilon_{\mathrm{n}} = |(E_{\mathrm{n}} - \bar{E})/N|$. The free energy is composed of three terms, the entropy, the native driving force, and roughness contribution of the energy landscape. In the parameter space in ($\delta\epsilon_{\mathrm{n}}$, $\Delta\epsilon$, T), the expression above can have a double minimum structure in the reaction coordinate Q with one minimum at low Q corresponding to the nonnative states separated by a barrier from another minimum at high Q corresponding to the native folded state. As the cooperativity measured by multibody interaction order m increases, the free-energy minimum of nonnative states and native folded state shift toward $Q \sim 0$ and $Q \sim 1$, respectively. To the extent that this approximation is good ($m \to \infty$), we can equate the free energies of the nonnative states and native folding state to obtain the folding transition temperature ($F(Q = 0) = F(Q = 1)$):

$$T_{\mathrm{f}} = \frac{\delta\epsilon_{\mathrm{n}}}{2s_{\mathrm{tot}}(Q = 0)}\left(1 + \sqrt{1 - \frac{2s_{\mathrm{tot}}(Q = 0)\Delta\epsilon^2}{\delta\epsilon_{\mathrm{n}}^2}}\right).$$

Take the ratio of folding temperature and trapping temperature, we obtain:

$$T_{\mathrm{f}}/T_{\mathrm{g}}(Q = 0) = \Lambda + \sqrt{\Lambda^2 - 1}, \tag{2}$$

where $\Lambda = (\delta\epsilon_{\mathrm{n}}/\Delta\epsilon\sqrt{2s_{\mathrm{tot}}(Q = 0)})$ is the ratio of the energy gap between native state and the average of the energy landscape spectrum to the ruggedness or the width (spread) of the distribution of the energy landscape spectrum weighted by entropy per contact $\sqrt{2s_{\mathrm{tot}}(Q = 0)}$, which is on the order of 1 (34). To guarantee the folding without getting into the local traps, the ratio of $(T_{\mathrm{f}}/T_{\mathrm{g}})$ should be maximized; this, in turn, leads to the maximization of Λ.

Therefore, maximizing the ratio of the energy gap (or the slope) versus the roughness of the underlined energy landscape becomes the criterion for the thermodynamic stability of folding, implying a funneled energy landscape.

Under the free-energy profiles, the equation of motion for native contact Q formation can be formulated as:

$$\frac{dQ}{dt} = -\frac{\partial F(Q)}{\zeta\partial Q} + \eta, \tag{3}$$

where ζ is the friction coefficient; $-\partial F(Q)/\partial Q$ is the gradient force the motion of Q would follow. Due to the long timescale of folding compared with the short timescale fluctuations, the folding can be seen as overdamped. Therefore, the second derivatives of Q with respect to time t is ignored; η is assumed to be a Gaussian noise force term

where its correlation becomes $\langle \eta(Q,t)\eta(Q,0)\rangle = 2D(Q)\delta(t)$. $D(Q)$ is the Q-dependent diffusion coefficient. The noise term is related to the environmental fluctuations (temperature) through the Einstein relationship (fluctuation dissipation theorem) $D(Q)\zeta = k_B T$. The protein folding has many degrees of freedom, therefore, when looking at the motion along the reduced one-dimensional order parameter or reaction coordinate Q, the noise is effectively from the rest of the other multidimensions of folding and the environments surrounding it.

When taking into account the combination of multibody interactions (up to the six-body interactions because the order of the hydrophobic multibody interactions beyond two-body interactions is typically ranging from three or four up to six), the free energy becomes:

$$\begin{aligned} F(Q) = &-N\delta\epsilon(\alpha Q + c_1Q^2 + c_2Q^3 + c_3Q^4 + (1-\alpha-c_1 \\ &- c_2 - c_3)Q^5 - N\frac{\delta\epsilon^2}{2T}(1 - (\alpha Q + c_1Q^2 + c_2Q^3 \\ &+ c_3Q^4 + (1-\alpha-c_1-c_2-c_3)Q^5)^2) - NTS_{\text{tot}}(Q), \end{aligned} \tag{4}$$

where α, c_1, c_2, c_3, $1-\alpha-c_1-c_2-c_3$ are the coefficient mimicking the relative importance of the order of multibody (2–6) interactions. N is the length of the polypeptide chain. $S_{\text{tot}}(Q) \approx S_0(1-Q) - QLog(Q) - (1-Q)Log(1-Q)$, where $S_0 \approx ln(10/2.718)$; 10 in the ln is the degrees of freedom per residue whereas factor 2.718 in the ln takes into account the constraints of the phase space upon collapse. The first term of the entropy is the entropy loss forming a contact whereas the rest of the two terms is responsible for the entropy associated with the possible ways of forming a contact.

We can now formulate the dynamics with the Onsager-Machlup (21) functional path integral as:

$$\begin{aligned} P(Q_f,t,Q_i,0) &= \int DQ exp\left[-\int L(Q(t))dt\right] \\ &= \int DQ \exp\left[-\int\left(\frac{1}{4}\frac{\left(\frac{dQ}{dt} + \frac{D(Q)\partial\beta F(Q)}{\partial Q}\right)^2}{D(Q)}\right.\right. \\ &\quad \left.\left. -\frac{1}{2}\frac{\partial\left(D(Q)\frac{\partial\beta F(Q)}{\partial Q}\right)}{\partial Q}\right)dt\right], \end{aligned} \tag{5}$$

where $\beta = (1/k_B T)$. The DQ is summing over all possible paths connecting Q_i at time $t = 0$ to Q_f at time t. The exponential factor gives the weight of each path, so the probability of folding dynamics from nonnative configuration Q_i to native configuration Q_f is equal to the sum of the weights from the contributions of all the possible paths. $L(Q(t))$ is the Lagrangian of the system.

Each path in the path integral contributes a weight, but not every path gives the same contribution. In fact, the contribution from the paths to the weight is on the exponential, so the dominant paths with the largest weight contribute significantly larger than the ones with the subdominant or even smaller weights. We can then approximate the path integrals with a set of dominant paths and ignore the subleading terms. One can easily see to find the paths with the optimal weights, the dominant paths should satisfy the Euler-Lagrangian equation (see Fig. 1):

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{Q}} - \frac{\partial L}{\partial Q} = 0, \tag{6}$$

and the resulting equation becomes:

$$\frac{d^2Q}{dt^2} - \frac{1}{2}\frac{\frac{\partial D(Q)}{\partial Q}}{D(Q)}\dot{Q}^2 - 2D(Q)\frac{\partial V}{\partial Q} = 0, \tag{7}$$

where

$$\begin{aligned} V(Q) = &\frac{D(Q)}{4}\left(\frac{\partial\beta F(Q)}{\partial Q}\right)^2 - \frac{D(Q)}{2}\frac{\partial^2\beta F(Q)}{\partial Q^2} \\ &- \frac{1}{2}\frac{\partial D(Q)}{\partial Q}\frac{\partial\beta F(Q)}{\partial Q}. \end{aligned} \tag{8}$$

The equation of motion can be integrated out to obtain:

$$\frac{\left(\frac{dQ}{dt}\right)^2}{4D(Q)} - V(Q) = E, \tag{9}$$

where E is a constant. This is an energy conservation equation with $(((dQ/dt))^2/4D(Q))$ as kinetic energy term with position Q dependent mass and $U = -V$ as the effective potential, and E as the total energy. The problem becomes a one-dimensional particle moving in a potential well U.

The free energy as a function of Q at various temperatures T is plotted in Fig. 2, *A–C*, (for mixed but mainly four-body, five-body, and six-body interactions). The potential V as

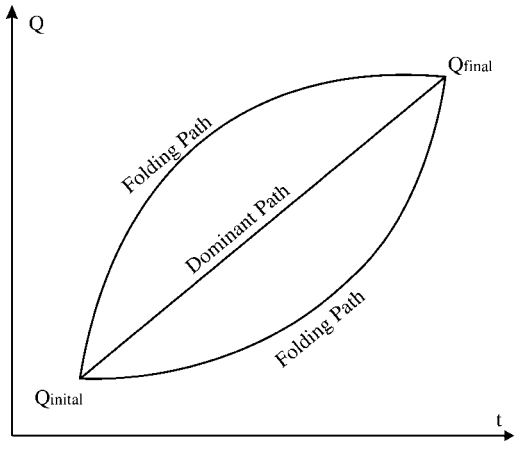

FIGURE 1 Folding paths from initial to final configuration.

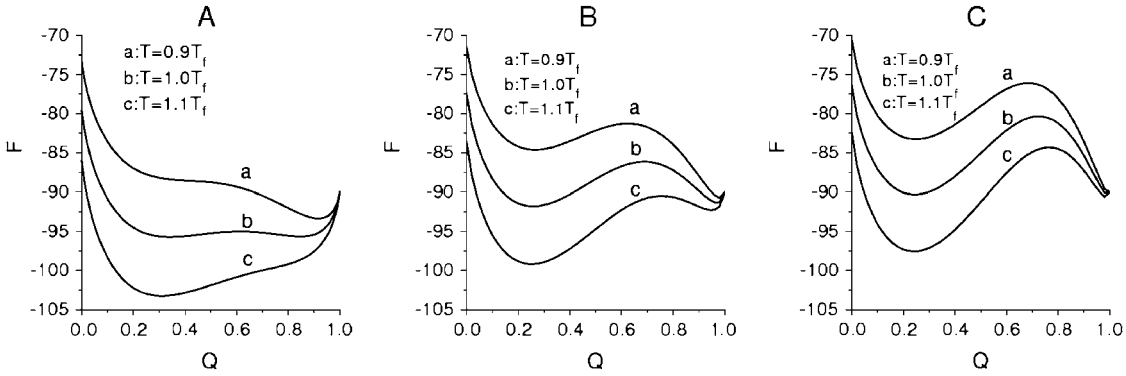

FIGURE 2 Free energy F as a function of Q at various temperatures near folding temperature. T_f for dominant four-body (A), dominant five-body (B), and dominant six-body interactions (C).

a function of Q is plotted in Fig. 3, *A–C* (for mixed but mainly four-body, mixed but mainly five-body, and mixed but mainly six-body interactions) as a function of Q at folding temperature T_f.

Because around folding temperature, the free energy F as a function of Q often has a double-well shape (Fig. 2, with given parameters specified later in the article) with one well corresponding to the nonnative unfolded states and the other one corresponding to the native folded state. The free-energy barrier is closely linked with the cooperative nature of multibody hydrophobic interactions for protein folding. We have done a careful analysis with different degrees of cooperativity in the inherent interactions in the Hamiltonian with mixed but mainly four-body interactions, mixed but mainly five-body interactions, and mixed but mainly six-body interactions. We see from Fig. 2, *A–C*, that as the degree of cooperativity increases, the barrier height increases too. In other words, for low cooperativity the barrier is small, but for high cooperativity the barrier is large. One can substitute the shape of $F(Q)$ into the expression of V and obtain the shape of the potential V as a function of Q (Fig. 3). Again, we see that the V has a minimum. The position of the minimum is close to the original minimum in Q in the free-energy profile F. The dominant contribution for the paths are from solving the equation of motion for Q. The effective potential $U = -V$. In the long time limit, there exists possibilities that the paths go back and forth many times from the hill (maximum) in the effective potential U (the minimum or the valley in V since $U = -V$) corresponding to the nonnative states to the other bounce-back point near the native state, where the value of U at the bounce is equal to that of the hill. This corresponds to the traversal of multiple times passing through the barrier to reach the native folded state. These oscillating back and forth solutions are called the instanton solutions (35–39). In Fig. 4, *A–C*, the instanton solutions are shown for dominant four-, five-, and six-body interactions. Each instanton (antiinstanton) corresponds to one transition from nonnative (native) to native (nonnative) states. The dominant path is composed of multiple instantons. The contribution can be summed in the dilute gas approximation by assuming no instanton-instanton interactions to obtain the final contribution to the probability of folding. The one instanton contribution to the weight is given by:

$$W = W_0 exp[-\gamma] = W_0 exp\left[-\int (L(Q_s(t) - L(Q_{min}))dt\right], \tag{10}$$

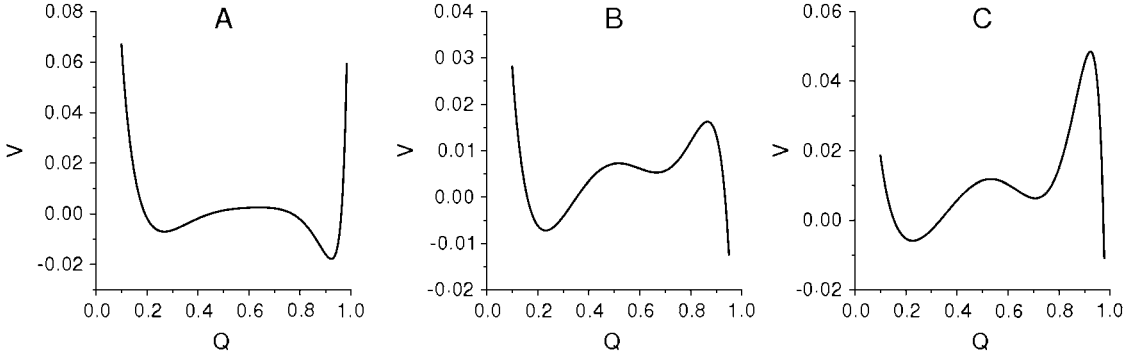

FIGURE 3 Potential V as a function of Q at folding temperature. T_f for dominant four-body (A), dominant five-body (B), and dominant six-body interactions (C) when $D = D(Q = 0)$.

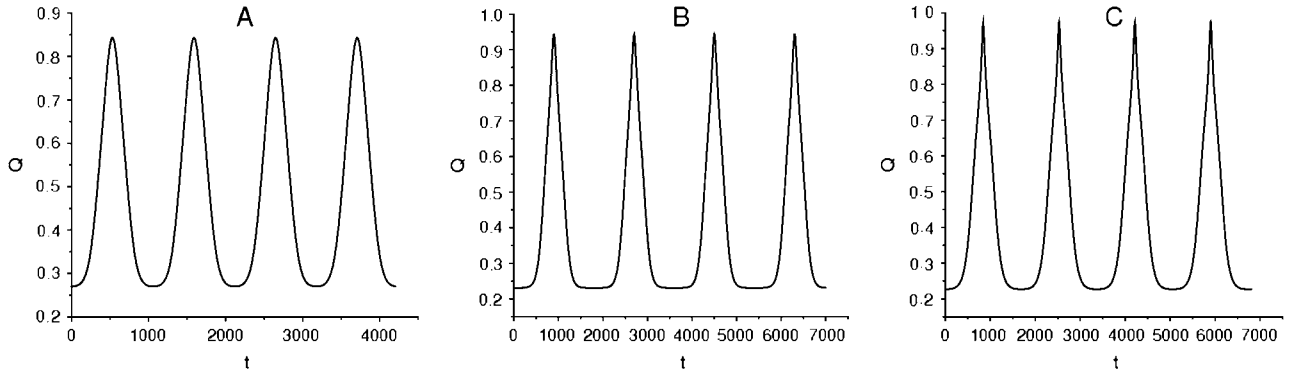

FIGURE 4 Instanton path Q as a function of time t of the protein folding dynamics at folding temperature. T_f for dominant four-body (A), dominant five-body (B), and dominant six-body interactions (C) when $D = D(Q = 0)$.

where W_0 is a constant and Q_{ins} is the instanton path ($Q_{ins}(t)$ is obtained through solving Eq. 11 with the boundary condition of $Q = Q_{min}$ and $\dot{Q} = 0$ at $t = 0$) and the integral is from the beginning of one instanton at $t = 0$ and at Q_{min} to the end of one instanton at the bounce-back point Q_{max} and at t_{max}. Q_{min} and Q_{max} correspond to approximately the minimum of V near the nonnative state and the bounce-back point of V near the native folded state, where the value of V at the bounce-back point is equal to that of the minimum of V.

The probability is determined from the optimal paths contributed by the sum of the multiinstanton contribution. It can be written as (Figs. 3 and 4):

$$P(t) \sim C \sum_{n=0}^{\infty} W^{2n} \int dt_1 \int dt_2 \ldots \int dt_{2n} \; exp[-V(Q_{min})((t_1 - t_0) + (t_3 - t_2) + \ldots + (t - t_{2n}))] \; exp[-V(Q_{max})((t_2 - t_1) + (t_4 - t_3) + \ldots + (t_{2n} - t_{2n-1}))], \tag{11}$$

where $t_1 - t_0, t_3 - t_2 \ldots t - t_{2n}$ are the time intervals staying at the position of the minimum of V near the nonnative states; $t_2 - t_1, t_4 - t_3 \ldots t_{2n} - t_{2n-1}$ are the time intervals staying at the position of the minimum of V near native folded state. The W is the one-instanton contribution to the probability $P(t)$. W^{2n} in the sum takes into consideration that the contribution to the probability for one traversal of the trajectory is from one minimum of V to the bounce-back point of V and back. This leads to one instanton (from one minimum Q_{min} of V near the nonnative states to the bounce-back point of V Q_{max} near the native state) and one antiinstanton (from the bounce-back point Q_{max} near native state to the minimum of V Q_{min} near the nonnative state); n in the sum is the number of times the trajectory is traversing from the minimum of V (Q_{min}) near nonnative state to the bounce-back point of V (Q_{max}) near native state and back.

The above expression can be easily evaluated in the Laplace representation s:

$$P(s) = C \sum_{n=0}^{\infty} W^{2n} (s + V(Q_{min}))^{-n-1} (s + V(Q_{max}))^{-n} = C(s + V(Q_{max})) \frac{1}{(s + V(Q_{min}))(s + V(Q_{max})) - W^2}. \tag{12}$$

By inverting the Laplace transform, we obtain:

$$P(t) = C_+ exp[\lambda_+ t] + C_- exp[\lambda_- t], \tag{13}$$

while λ_+ and λ_- are given by:

$$\lambda_\pm = -\frac{1}{2}(V(Q_{min}) + V(Q_{max})) \pm \frac{1}{2}\sqrt{(V(Q_{min}) - V(Q_{max}))^2 + 4W^2}. \tag{14}$$

When $V(Q_{min}) = V(Q_{max})$ as is the case in instantons, the expression is simplified as:

$$\lambda_\pm = -V(Q_{min}) \pm W, \tag{15}$$

because the $exp[-V(Q_{min})t]$ term accounts for the probability of staying at minimum of V near nonnative state or at bounce-back point of V. The real transition rate from one minimum to the bounce-back point is controlled by W (the $exp[-V(Q_{min})t]$ term is normalized out). Thus, the kinetic rate of folding k is determined by W ($k = W$).

DISCUSSIONS AND CONCLUSIONS

We take number of residues as $N = 30$, roughness or spread of the landscape $\Delta\epsilon = 1$, the bias or slope of the landscape toward folded state $\delta\epsilon = 3$ and $\alpha = 0.05$, $c_1 = 0.05$, $c_2 = 0$, $c_3 = 0$, and $1 - \alpha - c_1 = 0.90$ (for mixed two-, three-, and four-body interactions, but dominant four-body interactions); $\alpha = 0.05$, $c_1 = 0.05$, $c_2 = 0.05$, $c_3 = 0$, and $1 - \alpha - c_1 - c_2 = 0.85$ (for mixed two-, three-, four-, and five-body interactions, but dominant five-body interactions); $\alpha = 0.05$, $c_1 = 0.05$, $c_2 = 0.05$, $c_3 = 0.05$, and $1 - \alpha - c_1 - c_2 - c_3 = 0.80$ (for mixed two-, three-, four-, five-, and six-body

interactions, but dominant six-body interactions). The diffusion coefficients are given as (40): $D(Q) = D_0 exp[-S_0(Q)]$ for $T < T_g$; and $D(Q) = D_0 exp[-\beta^2 \Delta E(Q)^2]$ for $2T_g < T$; and $D(Q) = D_0 exp[-S_0(Q) + (\beta_g(Q) - \beta)^2]\Delta E(Q)^2]$ for $T_g < T < 2T_g$. Here, $T_g = (1/\beta_g) = \sqrt{(\Delta E(Q)^2/2S(Q))}$.

Fig. 4, *A–C*, shows the multiinstanton solutions at $T = T_f$ for dominant four-, five-, and six-body interactions, respectively. Fig. 5, *A* and *B*, show the temperature dependence of the logarithm (*ln*) of the folding rate (K) $-$ *lnK* for dominant four-, five-, and six-body interactions when diffusion coefficient is constant with $D = D(Q = 0)$ and when diffusion coefficient is reaction coordinate dependent with $D = D(Q)$.

As we can see folding kinetic rate has a ''bell''-like shape dependence with respect to the temperature. At high temperatures, the folding kinetic rate is small. This is due to the instability of proteins at high temperature. On the other hand, at low temperatures, the folding rate drops again. This is due to the possible trapping into the local valleys. Thus, the temperature varying kinetics provides a way of exploring the structure of the underlined folding energy landscape. The maximal rate for folding happens at certain optimal temperature. This is in good agreement with kinetic folding experiments (Chevron rollover) and theory/simulation studies (41–46). We also observed that as the cooperativity of the inherent hydrophobic interactions increases (from dominant four-body interactions to dominant six-body interactions), the free-energy barrier increases (as shown in Fig. 2), and the associated kinetic rate decreases. Furthermore, we can see that when the diffusion coefficient depends on the reaction coordinate, the kinetic rate for folding is significantly changed, especially at the low temperature regimes. In the low temperature regimes, the thermodynamic free-energy barrier for folding is less and less compared with the corresponding higher temperature case (see Fig. 2), and the effect of the diffusion on kinetics becomes more and more important.

This indicates that the kinetics is not only controlled by the inherent thermodynamic free energy but also by the diffusion. This is particularly important because the fast folding experiments are now approaching the speed limit where the kinetics of pure diffusion can be measured (47,15,48).

We can simplify the expression of the kinetic rate by assuming that diffusion coefficient is relatively small. In this case, we can substitute the instanton solution to the action of the probability expression of the path integral (49) and obtain analytic form of equilibrium probability as:

$$P(Q) \sim \exp[-\beta F(Q)/D], \tag{16}$$

with constant diffusion coefficient. With nonuniform diffusion coefficients, the result is

$$P(Q) \sim \exp\left[-\int_{Q_{\min}}^{Q} dQ' \frac{\partial \beta F(Q')}{\partial Q'} \Big/ D(Q')\right]. \tag{17}$$

The effective activation energy for transitions from nonnative unfolded state at $Q = Q_{\min}$ to the transition state $Q^{\#}$ is given by:

$$\beta \Delta F^{\#} = \int_{Q_{\min}}^{Q^{\#}} \frac{\partial \beta F(Q')}{\partial Q'} \Big/ D(Q') dQ'. \tag{18}$$

It is very important to realize that the current formalism implies both the diffusion and thermodynamic free-energy barrier control the kinetics of protein folding as mentioned above. When the underlying process is barrier limited, both the thermodynamic barrier and diffusion contribute to the kinetics although free energy contribution might be larger. The role of diffusion is to modify the effective free-energy profile and the corresponding barrier. In the case where there is no inherent free energy barrier, the kinetics is controlled by diffusion. Thus, the formalism in this article provides a route to look for the switching roles from thermodynamic-barrier-driven kinetics to downhill diffusion-driven kinetics, which is quite relevant for the experimental study of fast folding proteins where the speed of folding is determined from the thermodynamic-driven to the essentially diffusion-controlled process (47,15,48).

The current formalism can also be used to discuss the transition state property of protein folding. In the case of

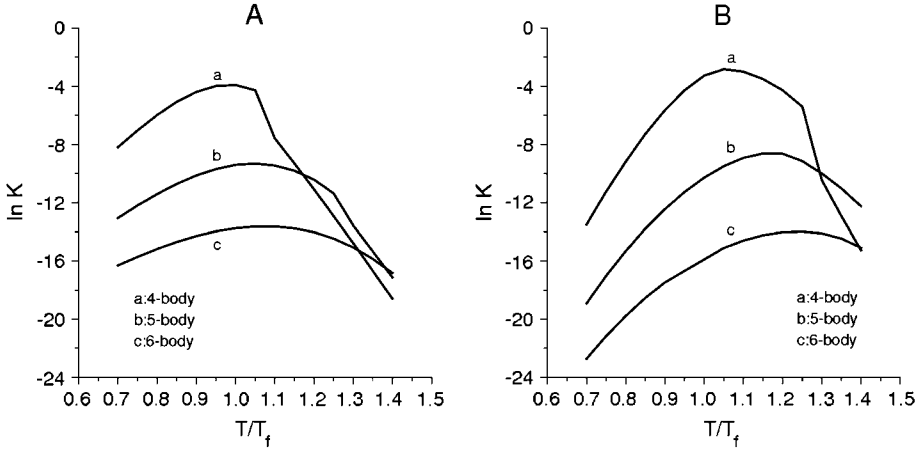

FIGURE 5 The logarithm of the kinetic protein folding rate *ln K* as a function of temperature T (in the units of folding temperature T_f) for dominant four-body, dominant five-body, and dominant six-body interactions when diffusion coefficient is constant, $D = D(Q = 0)$ (*A*) and when diffusion coefficient is Q dependent, $D = D(Q)$ (*B*).

constant diffusion, the current formalism reduces to the normal transition state theory and kinetics is controlled by the free energy barrier. As mentioned above, when the diffusion coefficient is not a constant, the kinetics is controlled by both the free energy barrier and diffusion. In the case when the thermodynamic barrier is large, the kinetics is dominated by the free energy profile. On the other hand, when the thermodynamic barrier is moderate, the effect of diffusion will come into play by modifying the original free energy profile. Both the position and value of the resulting effective transition state free energy will be shifted. So the kinetics will be modified by this shift of the original transition state. Details of the study will be given in a future publication.

Let us discuss the possible connections of our approach with another set of experimental observations (50). When the diffusion coefficient is a constant, the kinetics is controlled by the free energy profile as we have derived above. In the barrier limited case, if the barrier is caused mainly by topology instead of heterogeneity of the interactions, then the free energy barrier is mainly from entropy contribution of loop contacts (32,33,51). Thus, the free energy change with respect to the mean sequence length of making contacts $\bar{l}$ can be shown as:

$$\partial F/\partial \bar{l} \sim -T\partial S_{\text{contact}}/\partial \bar{l} \sim 1/\bar{l}^2. \quad (19)$$

Thus, the free energy barrier is linked to the average sequence length of making the contacts $\bar{l}$. The effect of increasing the mean loop length is to increase the barrier height. So the kinetics is faster (slower) when the mean contact distance is small (large). When the diffusion coefficient is not constant and interaction heterogeneity are taken into account, the free energy dependence on the mean contact distance might not be as strong. Further detailed investigations on this are needed and will be carried out in a future publication.

We discussed in this article the long-time dynamics of folding. In principle, the short-time dynamics can be revealed by solving the Euler-Lagrangian equation for the optimal paths. Because the time is short, the solutions typically don't have enough time forming multiple instantons. Finding dominant paths becomes solving ordinary differential equation for fixing two end points. One can expand around the dominant solution up to quadratic order and obtain the contribution to the probability of folding. In general the results are good for short times and the kinetics is usually nonexponential. This is in contrast with the long-time case where the dynamics is usually controlled by the longest timescale as we discussed here.

We obtain in this study the optimal instanton paths that determines the folding rate dynamics in the long time limit. We should mention that the optimal paths are actually a set of paths in the multidimensional configurational space. They represent the dominant flow of paths directed toward the native state. At low temperatures, the folding might be trapped into the local valleys, while the current continuous path approach can give some qualitative features as to approximately when the continuous flow of paths might break down; instead, the more appropriate approach seems to be the discrete version of the path integral we presented here. The formulation is currently under development. With this formulation, one can study and understand the transition from the multiple paths to the discrete path transition in the case of activated folding transition.

The kinetic rate dynamics is often studied by the Fokker-Planck type rate equations (or Brownian dynamics). This approach to the kinetics is mathematical, related to the path integral formulations presented here but emphasizing different aspects. Although the path method concentrates on intermediate processes and the corresponding contributions to the final kinetics, the Fokker-Planck type rate equation approach concentrates more on the end results. Therefore, it is convenient and advantageous to address the kinetic path issues for protein folding in the path integral formulation.

It is worth mentioning that biomolecular recognition (binding) often involves large fluctuations and conformational changes (52–59); sometimes local unfolding (60,61) for induced fit (62) is necessary, so in general folding and binding are dynamically coupled. It is tempting to study the kinetics of the folding-binding process using the current developed path integral methodology (J. Wang, K. Zhang, H. Y. Lu, and E. K. Wang, unpublished data). The crucial question would be what are the dominant kinetic paths for the folding-binding process in nature.

J.W. thanks Prof. Peter G. Wolynes, Prof. Jose N. Onuchic, and Prof. Andrew J. McCammon for helpful discussions.

The work of J.W. is supported by National Science Foundation Career Award, Petroleum Research Fund, K. C. Wong Foundation Research Award, and Stony Brook Faculty Funding. The work of K.Z., H.L., and E.K.W. is supported by the Chinese National Science Foundation.

REFERENCES

1. Levinthal, C. 1969. Mossbauer Spectroscopy in Biological Systems. University of Illinois Press, Urbana, IL.
2. Bryngelson, J. D., J. O. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein-folding: a synthesis. *Proteins.* 21:167–195.
3. Chan, H. S., and K. A. Dill. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J. Chem. Phys.* 100:9238–9257.
4. Wang, J., J. Onuchic, and P. Wolynes. 1996. Statistics of kinetic pathways on biased rough energy landscapes with applications to protein folding. *Phys. Rev. Lett.* 76:4861–4864.
5. Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes. 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.
6. Kim, P. S., and R. L. Baldwin. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu. Rev. Biochem.* 51:459–489.
7. Kim, P. S., and R. L. Baldwin. 1990. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59:631–660.

8. Baldwin, R. L. 1995. The nature of protein-folding pathways: the classical versus the new view. *J. Biomol. NMR.* 5:103–109.
9. Bai, Y. W., T. R. Sosnick, L. Mayne, and S. W. Englander. 1995. Protein-folding intermediates: native-state hydrogen-exchange. *Science.* 269:192–197.
10. Phillips, C. M., Y. Mizutani, and R. M. Hochstrasser. 1995. Ultrafast thermally induced unfolding of RNase A. *Proc. Natl. Acad. Sci. USA.* 92:7292–7296.
11. Williams, S., T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, and R. B. Dyer. 1996. Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide. *Biochemistry.* 35:691–697.
12. Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–288.
13. Jones, C. M., E. R. Henry, Y. Hu, C.-K. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, and W. A. Eaton. 1993. Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. USA.* 90:11860–11864.
14. Chan, C.-K., Y. Hu, S. Takahashi, D. L. Rousseau, W. A. Eaton, and J. Hofrichter. 1997. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. USA.* 94:1779–1784.
15. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
16. Lipman, E. A., B. Schuler, O. Bakajin, and W. A. Eaton. 2003. Single-molecule measurement of protein folding kinetics. *Science.* 301:1233–1235.
17. Sabelko, J., J. Ervin, and M. Gruebele. 1999. Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. USA.* 96:6031–6036.
18. Nguyen, H., M. Jäger, A. Moretto, M. Gruebele, and J. W. Kelly. 2003. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. USA.* 100: 3948–3953.
19. Wiener, N. 1964. Generalized Harmonic Analysis and Tauberian Theorems. MIT Press, Boston, MA.
20. Feynman, R. P., and A. R. Hibbs. 1965. Quantum Mechanics and Path Integrals. McGraw-Hill, New York, NY.
21. Onsager, L., and S. Machlup. 1953. Fluctuations and irreversible processes. *Phys. Rev.* 91:1505–1512.
22. Hanggi, P. 1989. Path integral solutions for non-Markovian processes. *Z. Phys. B.* 75:275–281.
23. Hunt, K. L. C., and J. Ross. 1981. Path integral solutions of stochastic equations for nonlinear irreversible processes: the uniqueness of the thermodynamic Lagrangian. *J. Chem. Phys.* 75:976–984.
24. Wang, J., and P. G. Wolynes. 1993. Passage through fluctuating geometrical bottlenecks. The general Gaussian fluctuating case. *Chem. Phys. Lett.* 212:427–433.
25. Wang, J., and P. G. Wolynes. 1994. Survival paths for reaction dynamics in fluctuating environments. *Chem. Phys.* 180:141–156.
26. Onuchic, J. N., J. Wang, and P. G. Wolynes. 1999. Analyzing single molecule trajectories on complex energy landscapes using replica correlation functions. *Chem. Phys.* 247:175–184.
27. Wang, J. 2003. Statistics, pathways and dynamics of single molecule protein folding. *J. Chem. Phys.* 118:952–958.
28. Olender, R., and R. Elber. 1996. Calculation of classical trajectories with a very large time step: formalism and numerical examples. *J. Chem. Phys.* 105:9299–9315.
29. Elber, R., J. Meller, and R. Olender. 1999. Stochastic path approach to compute atomically detailed trajectories: application to the folding of c peptide. *J. Phys. Chem. B.* 103:899–911.
30. Derrida, B. 1980. Random-energy model: limit of a family of disordered models. *Phys. Rev. Lett.* 45:79–82.
31. Derrida, B. 1981. Random-energy model: an exactly solvable model of disordered systems. *Phys. Rev. B.* 24:2613–2626.
32. Plotkin, S. S., J. Wang, and P. G. Wolynes. 1996. Correlated energy landscape model for finite, random heteropolymers. *Phys. Rev. E.* 53:6271–6296.
33. Plotkin, S. S., J. Wang, and P. G. Wolynes. 1997. Statistical mechanics of a correlated energy landscape model for protein folding funnels. *J. Chem. Phys.* 106:2932–2948.
34. Goldstein, R. A., Z. A. Luthey-Schulten, and P. G. Wolynes. 1992. Optimal protein-folding codes from spin-glass theory. *Proc. Natl. Acad. Sci. USA.* 89:4918–4922.
35. Langer, J. S. 1967. Theory of the condensation point. *Ann. Phys.* 41: 108–157 (NY).
36. Coleman, S. 1977. Fate of the false vacuum: semiclassical theory. *Phys. Rev. D.* 15:2929–2936.
37. Callen, C. G., and S. Coleman. 1977. Fate of the false vacuum. II. First quantum corrections. *Phys. Rev. D.* 16:1762–1768.
38. Jalickee, J. B., F. W. Wiegel, and D. J. Vezzetti. 1971. Role of droplets and bubbles in the 'condensation' of one-dimensional gas of Kac, Uhlenbeck, and Hemmer. *Phys. Fluids.* 14:1041–1048.
39. Wang, J., and P. G. Wolynes. 1996. Instantons and the fluctuating path description of reactions in complex environments. *J. Phys. Chem.* 100: 1129–1136.
40. Bryngelson, J. D., and P. G. Wolynes. 1989. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* 93:6902–6915.
41. Kaya, H., and H. S. Chan. 2000. Energetic components of cooperative protein folding. *Phys. Rev. Lett.* 85:4823–4826.
42. Kaya, H., and H. S. Chan. 2002. Towards a consistent modeling of protein thermodynamic and kinetic cooperativity: how applicable is the transition state picture to folding and unfolding? *J. Mol. Biol.* 315: 899–909.
43. Lee, C. L., G. Stell, and J. Wang. 2003. First-passage time distribution and non-Markovian diffusion dynamics of protein folding. *J. Chem. Phys.* 118:959–968.
44. Lee, C. L., C. T. Lin, G. Stell, and J. Wang. 2003. Diffusion dynamics, moments, and distribution of first-passage time on the protein-folding energy landscape, with applications to single molecules. *Phys. Rev. E.* 67:041905.
45. Zhou, Y., C. Zhang, G. Stell, and J. Wang. 2003. Temperature dependence of the distribution of the first passage time: results from discontinuous molecular dynamics simulations of an all-atom model of the second -hairpin fragment of protein G. *J. Am. Chem. Soc.* 125: 6300–6305.
46. Kuhlman, B., D. L. Luisi, P. A. Evans, and D. P. Raleigh. 1998. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9. *J. Mol. Biol.* 284:1661–1670.
47. Yang, W. Y., and M. Gruebele. 2003. Folding at the speed limit. *Nature.* 423:193–197.
48. Garcia-Mira, M. M., M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz, and V. Muñoz. 2002. Experimental identification of downhill protein folding. *Science.* 298:2191–2195.
49. Bialek, W. 2001. Stability and noise in biochemical switches. *Advances in Neural Information Processing.* 13:103–109.
50. Plaxco, K. W., K. T. Simons, and D. Baker. 1998. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* 277:985–994.
51. Plotkin, S. S., and J. N. Onuchic. 2002. Structural and energetic heterogeneity in protein folding I. Theory. *J. Chem. Phys.* 116:5263–5283.
52. McCammon, J. A. 1998. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.* 8:245–249.
53. Berkowitz, M., and J. A. McCammon. 1981. Brownian motion of a system of coupled harmonic oscillators. *J. Chem. Phys.* 75:957–961.

54. Berkowitz, M., J. D. Morgan, D. J. Kouri, and J. A. McCammon. 1981. Memory kernels from molecular dynamics. *J. Chem. Phys.* 75:2462–2463.

55. Berkowitz, M., and J. A. McCammon. 1982. Molecular dynamics with stochastic boundary conditions. *Chem. Phys. Lett.* 90:215–217.

56. Berkowitz, M., J. D. Morgan, and J. A. McCammon. 1983. Generalized Langevin dynamics simulations with arbitrary time-dependent memory kernels. *J. Chem. Phys.* 78:3256–3261.

57. Berkowitz, M., J. D. Morgan, J. A. McCammon, and S. H. Northrup. 1983. Diffusion-controlled reactions: a variational formula for the optimum reaction coordinate. *J. Chem. Phys.* 79:5563–5565.

58. Northrup, S. H., and J. A. McCammon. 1983. Saddle-point avoidance in diffusional reactions. *J. Chem. Phys.* 78:987–989.

59. Wang, J., and G. M. Verkhivker. 2003. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys. Rev. Lett.* 90:188101.

60. Shoemaker, B. A., J. J. Portman, and P. G. Wolynes. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. USA.* 97:8868–8873.

61. Papoian, G. A., and P. G. Wolynes. 2003. The physics and bio-informatics of binding and folding: an energy landscape perspective. *Biopolymers.* 68:333–349.

62. Koshland, D. E., Jr. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA.* 44: 98–104.