Protein Science (2000), *9*:452–465. Cambridge University Press. Printed in the USA.

Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model

D.J. BICOUT AND ATTILA SZABO

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bldg. 5, Bethesda, Maryland 20892

(RECEIVED May 20, 1999; FINAL REVISION December 3, 1999; ACCEPTED December 7, 1999)

Abstract

This paper presents an analytically tractable model that captures the most elementary aspect of the protein folding problem, namely that both the energy and the entropy decrease as a protein folds. In this model, the system diffuses within a sphere in the presence of an attractive spherically symmetric potential. The native state is represented by a small sphere in the center, and the remaining space is identified with unfolded states. The folding temperature, the time-dependence of the populations, and the relaxation rate are calculated, and the folding dynamics is analyzed for both golf-course and funnel-like energy landscapes. This simple model allows us to illustrate a surprising number of concepts including entropic barriers, transition states, funnels, and the origin of single exponential relaxation kinetics.

Keywords: diffusion; exponential kinetics; free energy barriers; protein folding

Protein folding is a complex chemical reaction in which a polymer interconverts between an ensemble of unfolded states and a compact native configuration. Over the last decade, considerable theoretical and computational effort (Bryngelson et al., 1995; Dill et al., 1995; Fersht, 1995; Karplus & Sali, 1995; Zwanzig, 1995; Orland et al., 1996; Thirumalai & Woodson, 1996; Onuchic et al., 1997; Shakhnovich, 1997; Dobson et al., 1998; Muñoz et al., 1998; Pande et al., 1998) has been devoted to understanding this phenomena on a fundamental level. Much of the conceptual framework that has evolved is based on simple ideas. The purpose of this paper is to illustrate some of these ideas in the context of an exceedingly simple model that can be analytically solved. This model is designed to capture only the most fundamental aspect of the folding problem, namely that both the entropy (i.e., the number of configurations) and the energy decrease as the protein folds.

In this model, protein folding is described as diffusion within a closed sphere in the presence of a spherically symmetric potential. The native state is represented by a small spherical region in the center, and the remaining space represents unfolded states. This model is simple enough to yield analytical results, yet it is rich enough to describe many of the features exhibited by more realistic models and even experiments.

In the most elementary version of this model, the energy of all unfolded states is zero, but the energy of the native region is so low that it acts as an irreversible trap or "black hole." Mathematically, this means that the surface of the inner (i.e., native) sphere is an absorbing boundary. In the biophysical context, this model has been previously used to analyze how the reduction of spatial dimensionality influences the rate of binding to receptors (Adam & Delbrück, 1968) and to estimate the rate of coalescence of two microdomains during folding (Karplus & Weaver, 1976). In this paper, this model will be used to illustrate the Levinthal "paradox" (Levinthal, 1969) on a golf-course landscape (Bryngelson & Wolynes, 1989). In three dimensions, when the native region is sufficiently small, the folding kinetics predicted by this model is essentially single exponential (Szabo et al., 1980). By rigorously mapping the three-dimensional problem onto a one-dimensional one involving the potential of mean force (i.e., the free energy), we will show that this behavior is due to the presence of an entropic barrier in the free energy.

This model can be generalized to describe reversible folding by assuming that the potential energy inside the inner sphere is $-\varepsilon$ (not $-\infty$, as above) and is zero outside. After solving this model analytically, we show how it can be mapped onto a two-state kinetic scheme, and examine the temperature dependence of the folding and unfolding rate constants. In addition, we discuss the optimum choice of the dividing surface separating reactants and products, i.e., the transition state.

This model can be further embellished by incorporating an additional attractive potential that is proportional to the distance to the origin (i.e., a constant force that drives the system toward the native state). A two-dimensional slice of the resulting potential energy surface has the shape of a funnel. This attractive potential increases the folding rate by reducing the magnitude of the free-

Reprint requests to: D.J. Bicout, National Institute of Diabetes and Kidney Disease, National Institute of Health, Bldg. 5, Rm. 136, Bethesda, Maryland 20892; e-mail: bicout@speck.niddk.nih.gov.

energy barrier. When the attractive potential becomes strong enough, the free-energy barrier can completely disappear. This is the so-called "downhill scenario" (Bryngelson et al., 1995), where the folding kinetics becomes nonexponential and the system cannot be described by a two-state kinetic scheme.

In this paper, for the sake of simplicity, the attractive potential energy is temperature independent, and thus hydrophobic interactions are ignored. However, one could incorporate the effect of such interactions by simply making the potential energy depend on temperature (and thus making it a free energy).

The model considered in this paper is closely related to and, in fact, leads to qualitatively the same results as the "correct–incorrect" model of protein folding kinetics (Zwanzig et al., 1992; Zwanzig, 1995). In this model, as applied to an N residue protein, states of the system correspond to corners of an N dimensional hypercube. The dynamics is described as hopping between adjacent vertices, and the native state is identified with a corner of this cube. This multidimensional model can be rigorously reduced to a one-dimensional one by using the appropriate reaction coordinate. When the energy difference ε between the folded state and unfolded states is infinite, the folded state is an absorbing point, and folding is irreversible (Zwanzig et al., 1992). If ε is finite (Zwanzig, 1995), the folding is reversible. When the correct $\rightarrow$ incorrect rate is smaller than the incorrect $\rightarrow$ correct one, the system is driven toward the native state by a constant force. The dynamics (hopping vs. diffusion), the topology (hypercube vs. hypersphere), the reaction coordinate ("correctness" vs. distance to the origin), and the free energy (harmonic vs. logarithmic) are, of course, different in the two models, but the spirit is the same.

Formulation of the model

We represent each configuration of the polypeptide chain by a point $\mathbf{r}$ inside a d-dimensional sphere of radius R. The energy landscape is specified by a spherically symmetric potential $U(r)$ that depends only on the distance $|\mathbf{r}| = r$ from the origin. Because of spherical symmetry, all configurations with the same radial coordinate have the same behavior. Thus, r specifies an ensemble of configurations with different angular coordinates. The number of these configurations is proportional to the area of a sphere of radius r.

The dynamics of the system is assumed to be diffusive in nature. The probability density $P(r,t)$ of finding the system at r at time t satisfies

$$\frac{\partial P(r,t)}{\partial t} = \frac{1}{r^{d-1}} \frac{\partial}{\partial r} D(r) r^{d-1} e^{-\beta U(r)} \frac{\partial}{\partial r} e^{\beta U(r)} P(r,t) \qquad (1)$$

where $\beta = 1/k_B T$. Illustrative numerical results will be presented only for the case where the diffusion coefficient is independent of r. While it can be argued that the dimension d should be proportional to the number of residues in the protein, for the sake of simplicity we will focus primarily on three dimensions. Our purpose here is not to present a model for the way a protein actually folds, but rather to show that an extremely simple model can capture so many of the features exhibited by more realistic models and experiment.

We will consider two forms of interaction potential $U(r)$. The simplest is the radial step function,

$$U(r) = \begin{cases} -\varepsilon, & 0 \le r \le R_N, \\ 0, & R_N < r \le R, \end{cases} \qquad (2)$$

where $\varepsilon \ge 0$. Any configuration inside the sphere with radius R_N is identified with the native state. The rest of the space represents unfolded configurations. Thus, the folded state is energetically more stable than the unfolded state by ε. To keep the notation as simple as possible, we rescale the variable r as $r \leftrightarrow r/R$ so that the radius of outer sphere is 1 and the radius of the sphere enclosing the native states is $a = R_N/R$.

Figure 1 shows the total configuration space representing all possible conformations of the polypeptide chain. To visualize the energy profile that corresponds to Equation 2, we plot the energy as a function of position on a plane through the center of the sphere in Figure 2A. It can be seen that the step function potential leads to a golf-course energy landscape (Bryngelson & Wolynes, 1989) with a "cup" depth equal to ε. When ε is finite, this model describes reversible folding, because any trajectory that enters the native region can subsequently leave it. When $\varepsilon \rightarrow \infty$, folding is irreversible. We shall call this limiting case, the black-hole golf-course landscape.

In the second form of the potential that will be considered in this paper, an attractive potential proportional to r is added to the radial step-function potential. Specifically,

$$U(r) = \begin{cases} -\varepsilon + F(r-1), & 0 \le r \le a, \\ F(r-1), & a < r \le 1, \end{cases} \qquad (3)$$

where $F \ge 0$. In this model, a constant force F drives the system toward the native state. A two-dimensional representation of this

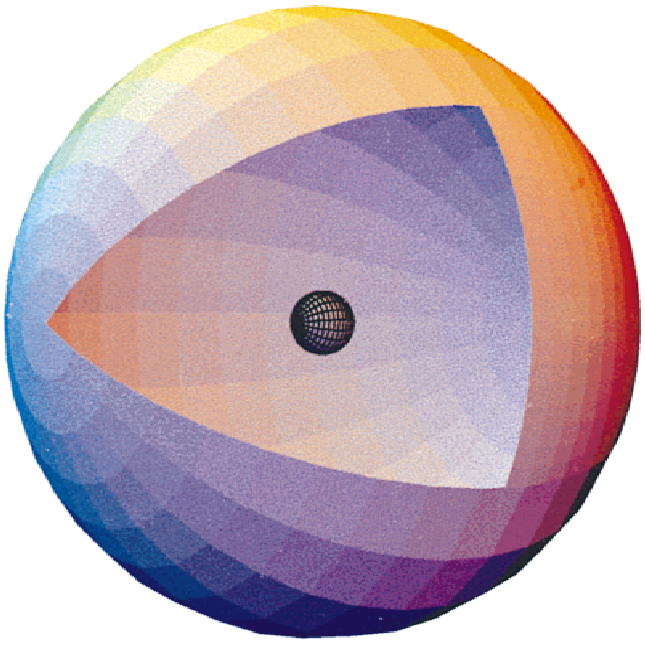

Fig. 1. Sketch of the three-dimensional version of our model representing the protein configurational space. Each point within the inner sphere of radius R_N is a native configuration and the system is bounded by an outer sphere of radius R ($R \gg R_N$). The two spheres are concentric and the volume between them contains the unfolded configurations.

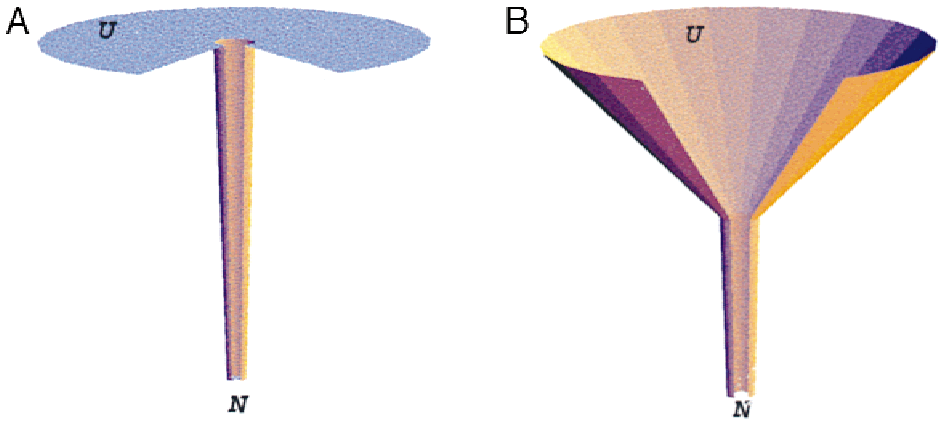

Fig. 2. The potential energy (vertical direction) on a plane bisecting the sphere, for $a = 0.1$ and $\varepsilon = 6.91k_BT$; N and U stand for native and unfolded states, respectively. **A:** Golf-course energy landscape that corresponds to the step function potential given in Equation 2. The equilibrium population of N and U are both equal, i.e., $K_{eq} = 1$. **B:** Funnel-energy landscape for $F = \varepsilon/2$ that corresponds to the attractive potential given in Equation 3. About 89% of the population at equilibrium is in N, i.e., $K_{eq} = 8.09$.

potential is shown in Figure 2B. It can be seen that this potential leads to a funnel landscape (Leopold et al., 1992; Wolynes et al., 1995; Onuchic et al., 1995) with a stem of length ε. The potentials given in Equations 2 and 3 were chosen because of their simplicity. Continuous version of them (i.e., replacing the step function by, say, a Lennard–Jones potential) would lead to similar results.

Black-hole golf course

We first consider the situation where the native state is described as an infinitely deep well ($\varepsilon \to \infty$ in Equation 2). This implies that any trajectory that reaches the surface of the sphere of radius a folds irreversibly. This mimics the situation when the system is under strongly folding conditions. We assume that at $t = 0$, the system is unfolded and uniformly distributed in the space bounded by the concentric spheres with radii a and 1. The diffusion equation, with this initial condition, and with absorbing (at $r = a$) and reflecting (at $r = 1$) boundary conditions can be solved analytically (see, e.g., Adam & Delbrück, 1968). The fraction of unfolded configurations at time t, denoted by $S(t)$, is then given by

$$S(t) = \sum_{n=1}^{\infty} \left[\frac{6a^2}{q_n^2(1-a^3)\{\sin^2[q_n(1-a)] - a\}} \right] e^{-q_n^2 Dt/R^2}, \quad (4)$$

where the eigenvalues q_n are determined from $\tan[q_n(1-a)] = q_n$. Using the theory of mean first passage times (Szabo et al., 1980), the folding rate constant k_f can be expressed as

$$\frac{1}{k_f} = \int_0^{\infty} S(t)\, dt = \frac{(1-a)^2(5+6a+3a^2+a^3)}{15a(1+a+a^2)} \frac{R^2}{D}. \quad (5)$$

For $a = R_N/R \ll 1$ (where R_N is the radius of the sphere enclosing the native configurations), this expression for the folding rate simplifies to $k_f \simeq 3aD/R^2 = 3DR_N/R^3$. In d ($d > 2$) dimensions, this generalizes to $k_f \simeq d(d-2)DR_N^{d-2}/R^d$. These results, which show that $k_f \to 0$ as $R_N \to 0$, are the quantitative statement of the Levinthal "paradox" (Levinthal, 1969). Finding the native state of a protein by an unbiased random search can take an astronomically long time when the space of unfolded configurations is sufficiently large.

Folding kinetics: Exponential relaxation

To examine the nature of the folding kinetics on a flat energy landscape, we consider the case where the radius of the native sphere is $a = 0.1$. The first two terms of $S(t)$ are

$$S(t) = 0.995e^{-Dt/2.755R^2} + 0.0028e^{-Dt/0.0397R^2} + \cdots. \quad (6)$$

This shows that the time evolution of the fraction of unfolded configurations is almost perfectly described by a single exponential. Consequently, the approximation $S_{\text{approx}}(t) = e^{-k_f t}$, where k_f is given in Equation 5 ($k_f = 0.365D/R^2$) is excellent.

Why is $S(t)$ so close to being a perfect exponential? In Figure 3 we present a typical folding trajectory starting from the point (0,0,0.8) and terminating upon reaching the inner sphere with radius $a = 0.1$. It can be seen that the system has a "hard time" finding the folded state and must first explore a large part of the unfolded configuration space. However, it is by no means clear from this single trajectory why $S(t)$ is almost exactly a single exponential. Let us now consider the behavior of an ensemble of such trajectories. Imagine that at $t = 0$ one initiates a large number of trajectories from positions that are uniformly distributed in the region between the two spheres. After a very short time, those configurations that happen to have started near to the inner sphere are absorbed with a high probability, and thus fold. After this initial transient, the shape of the distribution of the unfolded proteins remains virtually the same, because each unfolded conformation must explore a significant part of the total unfolded configuration space before arriving at the inner sphere. As the reaction progresses, the fraction of unfolded configurations, of course, decreases. If the shape of the distribution of the unfolded configurations does not change with time, then the kinetics must be single exponential (i.e., for $0 \le t_1 \le t_2$, the solution of the functional equation $S(t_2) = S(t_1)S(t_2 - t_1)$ is $S(t) = e^{-\alpha t}$).

It is well known from the chemical kinetics that if a system has to surmount a high barrier to react, the kinetics is single exponential. In the next subsection, we show how the single exponential

Fig. 3. A typical three-dimensional folding trajectory for free diffusion starting from the coordinate (0, 0, 0.8) and terminating on the surface of the native sphere, when $a = 0.1$ and the outer radius is equal to 1.

kinetics of the black-hole golf course can be rationalized using the concept of an entropic barrier.

Potential of mean force: Entropic barriers

Can our d-dimensional diffusion problem be rigorously mapped onto an effective one-dimensional one? Finding an exact one-dimensional reaction coordinate is impossible in general, but it is very easy for our spherically symmetric model. Let x be the reaction coordinate and denote its probability distribution by $p(x,t)$. To conserve the probability we require that $\int_a^1 p(x,t)\,dx = \sigma(d)\int_a^1 r^{d-1}P(r,t)\,dr$, where $P(r,t)$ is the solution of Equation 1, and $\sigma(d) = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of a unit d-dimensional sphere. This suggests the transformations $r \to x$ and $\sigma(d)r^{d-1}P(r,t) \to p(x,t)$ in Equation 1. We immediately find that

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x} D(x) e^{-\beta V_{pmf}(x)} \frac{\partial}{\partial x} e^{\beta V_{pmf}(x)} p(x,t), \qquad (7)$$

where the potential of mean force $V_{pmf}(x)$ is given by $\exp\{-\beta V_{pmf}(x)\} = \text{constant} \times x^{d-1}\exp\{-\beta U(x)\}$. Thus, $\exp\{-\beta V_{pmf}(x)\}$ is proportional to the equilibrium probability of finding the system at x. If we choose the constant so that $V_{pmf}(1) = U(1)$, then

$$V_{pmf}(x) = U(x) - (d-1)k_B T \ln(x). \qquad (8)$$

The potential of mean force is a free energy, and hence, $V_{pmf}(x) = U(x) - TS(x)$. Comparing this with Equation 8, we find that the configurational entropy is given by $S(x) = (d-1)k_B \ln(x)$.

In general, a multidimensional dynamical problem cannot be exactly reduced to a one-dimensional one. However, given any reaction coordinate x, it is always possible to obtain the free energy of the system as a function of this coordinate (i.e., the potential of mean force $V_{pmf}(x)$) using equilibrium statistical mechanics. Given such a potential, the simplest approximate treatment of stochastic dynamics that is consistent with the equilibrium properties is to assume that the system diffuses along the reaction coordinate in the presence of this potential. When the reaction coordinate is cleverly chosen, this description can be not only conceptually but also quantitatively useful (Socci et al., 1996).

For the black-hole golf course [$U(x) = -\infty$ for $x \le a$ and $U(x) = 0$ for $a < x < 1$], the potential of mean force is plotted as a function of x in Figure 4a. It can be seen that this potential has

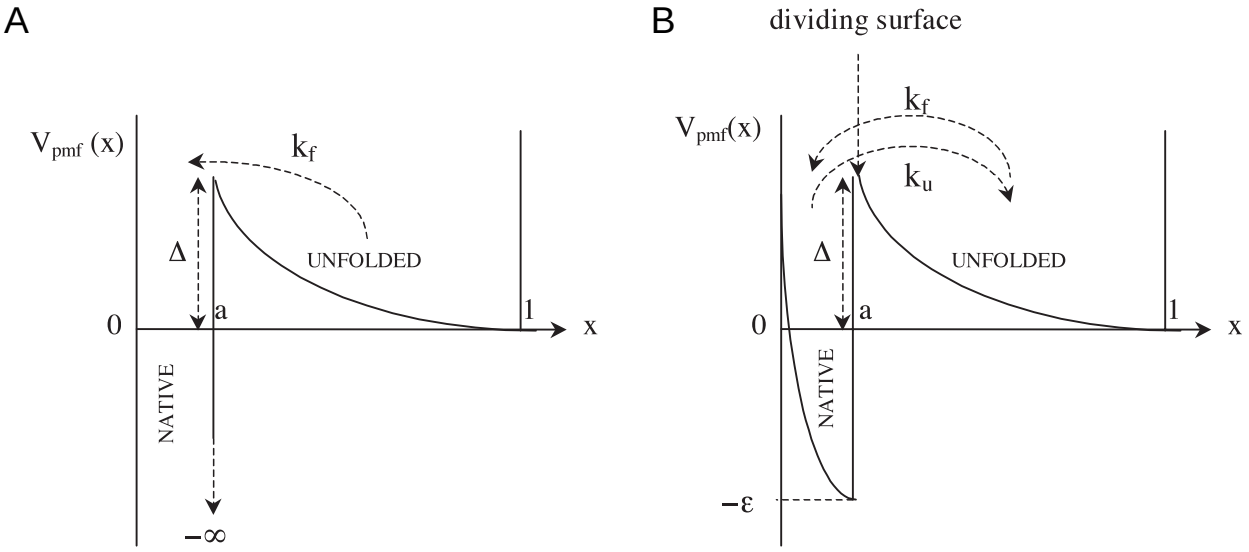

Fig. 4. Potential of mean force (or free energy) for (**A**) the black-hole golf course and (**B**) the reversible golf-course landscapes as a function of the reaction coordinate x, which is the radial distance to the origin. The barrier is due to the fact that the number of configurations (and, hence, the entropy) with a given radial coordinate r decreases as r decreases.

a cusp-like barrier at $x = a$. Physically, this is simply the reflection of the fact that the number of configurations with the same r decreases as r decreases, and so states with small r are less probable. Thus, the barrier in the one dimensional free-energy potential is entropic in nature. The barrier height Δ is given by

$$\Delta = V_{pmf}(a) - V_{pmf}(1) = T[S(1) - S(a)]$$

$$= (d-1)k_B T \ln[1/a]. \tag{9}$$

Since $\Delta \propto k_B T$ and $k_f \propto e^{-\Delta/k_B T}$, the rate of barrier crossing is weakly dependent on temperature. The rate constant given previously (i.e., $k_f \simeq d(d-2)DR_N^{d-2}/R^d$) can be expressed in terms of the entropic barrier height as $k_f \simeq [d(d-2)D/RR_N]e^{-\Delta/k_B T}$ and depends on temperature only through D. Thus, the folding rate goes exponentially to zero as the entropic barrier height increases.

When the barrier separating reactants and products is sufficiently high compared to the thermal energy, the populations in each subsystem rapidly equilibrate and the kinetics approaches purely single exponential behavior. Thus, single exponential behavior is expected for folding on the golf-course landscape when the entropic barrier Δ is much greater than $k_B T$. Indeed, for $a = 0.1$ and $d = 3$, we find that $\Delta = 4.61 k_B T$ and the kinetics is well approximated by a single exponential (see Equation 6).

Reversible golf course

We now consider the case when the depth of the native well is finite (see Fig. 2), beginning with the thermodynamics. The potential energy is given by Equation 2 and, thus, the normalized thermal equilibrium distribution $p_{eq}(r)$ of any configuration r is

$$p_{eq}(r) = \frac{e^{-\beta U(r)}}{\int_0^1 4\pi r^2 e^{-\beta U(r)}\, dr}$$

$$= \frac{3}{4\pi(a^3 e^{\beta\varepsilon} + 1 - a^3)} \begin{cases} e^{\beta\varepsilon}, & 0 \le r \le a, \\ 1, & a < r \le 1. \end{cases} \tag{10}$$

The fraction $f = \int_0^a 4\pi r^2 p_{eq}(r)\, dr$ is the probability of finding the system at equilibrium in the native state, and the equilibrium constant between the native and unfolded states is

$$K_{eq} = \frac{f}{1-f} = e^{-\beta \Delta G_{\text{fold}}}; \quad \Delta G_{\text{fold}} = -\varepsilon + k_B T \ln\left[\frac{1-a^3}{a^3}\right]. \tag{11}$$

Because the free energy of folding is $\Delta G_{\text{fold}} = \Delta U_{\text{fold}} - T\Delta S_{\text{fold}}$, where ΔU_{fold} and ΔS_{fold} are, respectively, the folding energy and entropy, $\Delta U_{\text{fold}} = -\varepsilon$ ($\varepsilon > 0$) and $\Delta S_{\text{fold}} = k_B \ln[a^3/(1-a^3)] = k_B \ln(\text{volume of native sphere}/\text{volume of unfolded space})$. $\Delta G_{\text{fold}} < 0$ for $T < T_f$ and $\Delta G_{\text{fold}} > 0$ for $T > T_f$, where T_f is the folding or transition temperature at which $\Delta G_{\text{fold}} = 0$ (or equivalently, $f(T_f) = \frac{1}{2}$ or $K_{eq}(T_f) = 1$). The folding temperature decreases like $k_B T_f \simeq \varepsilon/\ln[1/a^3]$ as the native sphere radius gets smaller. For $a = 0.1$ and $a = 0.2$ we have $\varepsilon = 6.91 k_B T_f$ and $\varepsilon = 4.82 k_B T_f$, respectively. Figure 5 shows that the temperature dependence of $f(T)$ has a sigmoidal shape. The system is under folding conditions for $T < T_f$ (i.e., low T where $\Delta G_{\text{fold}} < 0$) and under unfolding conditions for $T > T_f$ (i.e., high T where $\Delta G_{\text{fold}} > 0$).

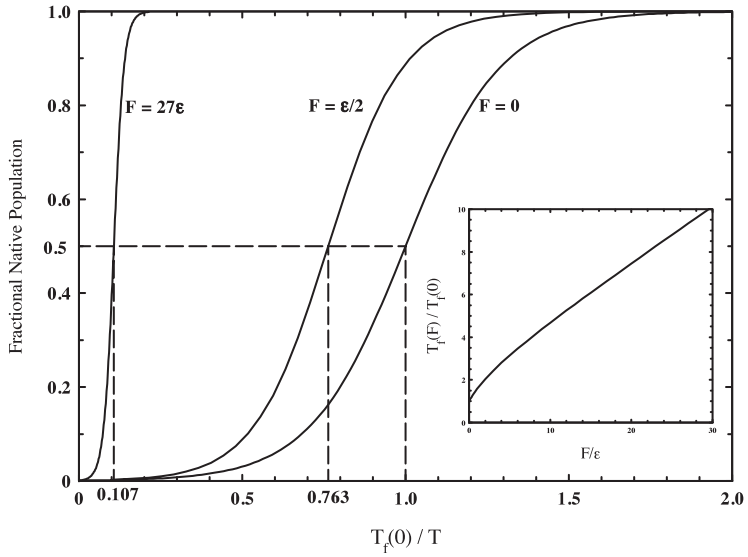

Fig. 5. Fractional native population for the golf-course landscape ($F = 0$), funnel landscape ($F = \varepsilon/2$), and downhill scenario ($F = 27\varepsilon$) vs. $T_f(0)/T$ with $a = 0.1$ and $\varepsilon = 6.91 k_B T_f(0)$, where $T_f(0)$ given by $k_B T_f(0) = \varepsilon/\ln[(1-a^3)/a^3]$ is the folding temperature for the reversible golf course. The folding temperature is $1.31 T_f(0)$ for the funnel landscape and $9.35 T_f(0)$ for the downhill scenario. **Inset:** Reduced folding temperature $T_f(F)/T_f(0)$ as a function of the reduced funnel slope F/ε.

Two-state kinetics

The exact description of the dynamics on a reversible golf-course landscape can be found by solving the diffusion equation subject to the appropriate boundary conditions (see Appendix A). We will soon see that, in certain circumstances, this dynamics can be well approximated by two-state chemical kinetics. To set the stage, consider the two-state kinetic scheme:

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} N$$

where $N(t)$ and $U(t)$ are, respectively, the fractional populations in the native and unfolded configurations, and k_f and k_u are the folding and unfolding rate, respectively. The rate equations corresponding to this kinetic scheme are

$$\frac{dN}{dt} = -k_u N + k_f U$$

$$\frac{dU}{dt} = +k_u N - k_f U \tag{12}$$

with $N(t) + U(t) = 1$. At equilibrium, $dN/dt = dU/dt = 0$, and so, $N_{eq} = 1 - U_{eq} = f$ and $K_{eq} = N_{eq}/U_{eq} = k_f/k_u$. If initially the protein is unfolded ($U(0) = 1$), $N(t)$ is given by

$$N(t) = \frac{K_{eq}}{1 + K_{eq}}(1 - e^{-(k_f + k_u)t}). \tag{13}$$

Defining the deviation from equilibrium by $\delta N(t) = N(t) - N_{eq}$, the combination of the above rate equations gives

$$\frac{d[\delta N(t)]}{dt} = -(k_f + k_u)\delta N(t) \Rightarrow \delta N(t) = \delta N(0) e^{-(k_f + k_u)t}, \tag{14}$$

and, hence, any deviation from equilibrium relaxes to zero exponentially. Multiplying the first equation of 14 by $\delta N(0)$, and taking the equilibrium average, one finds (Chandler, 1978; Skinner & Wolynes, 1979)

$$C(t) = \frac{\langle \delta N(t)\, \delta N(0) \rangle}{\langle (\delta N)^2 \rangle} = e^{-(k_f + k_u)t}. \tag{15}$$

The correlation function $C(t)$, which describes fluctuations in the populations at equilibrium, also decays as a single exponential. The relaxation rate k_{rx} is given by $k_{rx} = 1/[\int_0^\infty C(t)\, dt] = k_f + k_u$. In kinetic experiments, the observed relaxation time is k_{rx}^{-1} (compare Equations 15 and 13).

Folding kinetics

Two-state kinetics can adequately describe diffusive dynamics on the reversible golf-course landscape when the exact expression for the correlation function $C(t)$ is close to being single exponential. For the reversible golf-course landscape, $C(t)$ can be obtained analytically (see Appendix A). For now, it is sufficient to note only that $C(t)$ is multiexponential in which the structure is similar to the expression given in Equation 4, which describes the time dependence of the fraction of unfolded configurations in the irreversible case.

As discussed in the preceding subsection, potential of mean force, diffusion in a sphere in the presence of a spherically symmetric potential, is completely equivalent to one-dimensional diffusion in the presence of a potential of mean force given by Equation 8. For the reversible golf course in three dimensions, this equation gives

$$V_{pmf}(x) = U(x) - 2k_B T \ln(x) = \begin{cases} -\varepsilon - 2k_B T \ln(x), & 0 \le x \le a, \\ -2k_B T \ln(x), & a < x \le 1. \end{cases} \tag{16}$$

This potential is shown in Figure 4b. The barrier that must be crossed in order to fold is entropic in nature, and its height Δ is again given by $2k_B T \ln(1/a)$.

For an arbitrary one-dimensional potential, the relaxation rate can be expressed analytically in terms of quadratures (Schulten et al., 1981; Bicout & Szabo, 1997). For a potential that is infinite when $x = 0$ and $x = 1$, one has

$$\begin{aligned} \frac{1}{k_{rx}} &= \int_0^\infty C(t)\, dt \\ &= \frac{R^2 Q_+(a)}{[Q_+(a) + Q_-(a)]Q_-(a)} \int_0^a \frac{[Q_-(x)]^2}{D(x)}\, e^{\beta V_{pmf}(x)}\, dx \\ &\quad + \frac{R^2 Q_-(a)}{[Q_+(a) + Q_-(a)]Q_+(a)} \int_a^1 \frac{[Q_+(x)]^2}{D(x)}\, e^{\beta V_{pmf}(x)}\, dx, \end{aligned} \tag{17}$$

where the products are separated from the reactants by $x = a$, and where $Q_\pm(x)$ are defined as $Q_-(x) = \int_0^x \exp\{-\beta V_{pmf}(y)\}\, dy$ and $Q_+(x) = \int_x^1 \exp\{-\beta V_{pmf}(y)\}\, dy$. Using Equation 16 for $V_{pmf}(x)$, assuming $D(x) = D$, and evaluating the integrals, we find

$$\begin{aligned} \frac{D}{R^2 k_{rx}} &= \frac{a^2}{15}\left(\frac{1}{1 + K_{eq}}\right) \\ &\quad + \frac{(1-a)^2(5 + 6a + 3a^2 + a^3)}{15a(1 + a + a^2)}\left(\frac{K_{eq}}{1 + K_{eq}}\right), \end{aligned} \tag{18}$$

where K_{eq} is given in Equation 11.

We now return to the question of whether two-state chemical kinetics provides a satisfactory description of the dynamics of this system. When the barrier separating the unfolded configurations from the native ones (i.e., Δ in Fig. 4b) is high compared to $k_B T$, one expects that the folding dynamics can be described by two-state kinetics. This should be the case when $a = 0.1$, because $\Delta = 4.61 k_B T$. To verify this, we examine the correlation function $C(t)$ for $a = 0.1$, $\varepsilon = 6.91 k_B T$ ($K_{eq} = 1$):

$$C(t) = 0.989 e^{-Dt/1.3852R^2} + 0.0058 e^{-Dt/0.0396R^2} + \cdots. \tag{19}$$

Thus, the exact $C(t)$ is almost perfectly a single exponential and is virtually indistinguishable from $C_{\text{approx}}(t) = e^{-k_{rx}t}$, where k_{rx} is given by Equation 18.

Another somewhat more complex example of a multidimensional problem that can be mapped onto a two-state kinetic scheme is the diffusion of probes in the membranes of two fused cells (Chen et al., 1993). Specifically, the problem of diffusion on the surfaces of the two fused spheres can be rigorously reduced to one-dimensional diffusion in the presence of an entropic potential of mean force. When the radius of the pore at the fusion junction is small, the kinetics is exponential, and the relaxation rate can be found using Equation 17.

Temperature dependence of the folding and unfolding rate constants

The temperature dependence of the relaxation rate constant $k_{rx} = k_f + k_u$ can be found from the analytic expression given in Equation 18. Because $K_{eq} = k_f/k_u$, the individual rates are given by $k_f = k_{rx}K_{eq}/(1 + K_{eq})$ and $k_u = k_{rx}/(1 + K_{eq})$. When a is small, these rates are approximately given by $k_f \approx 3Da/R^2$ and $k_u \approx (3D/a^2R^2)e^{-\beta\varepsilon}$. Let us assume that the diffusion coefficient is essentially independent of temperature over the range of interest (this will be relaxed in the section devoted to the effect of ruggedness of the energy landscape). Then, the folding rate is independent of temperature because in this model the barrier between the unfolded and folded configurations is entropic. The unfolding rate constant, on the other hand, exhibits the familiar Arrhenius behavior, because unfolding is a thermally activated process because of low energy of the native state.

In Figure 6, k_{rx}, k_f, and k_u are plotted as a function of T_f/T for the reversible golf course. This is a kind of chevron plot with the reciprocal of the temperature playing the role of the denaturant concentration (Jackson & Fersht, 1991). The relaxation rate k_{rx} is equal to the folding rate k_f under folding conditions $T < T_f$ and to the unfolding rate k_u under unfolding conditions $T > T_f$. As explained above, k_u has an Arrhenius behavior, while k_f is almost independent of temperature. When a force that drives the system to the native state is incorporated into the model, k_f increases with decreasing the temperature (see below).

Transition states

The transition state is usually obtained by choosing the dividing surface between reactants (unfolded conformations) and products (folded conformations) at the top of the barrier in the potential of mean force along the reaction coordinate. This is why in the previous section we have chosen $r = a$ as the dividing surface. Incidentally, the transition state in our three-dimensional model is an ensemble of conformations all having $r = a$, but with different polar angles. Is there a better choice for the location of the dividing surface? In general, the correlation function $C(t)$, which describes population fluctuations, is multiexponential. One can choose the position $a^{\ddagger}$ of the dividing surface so as to maximize the amplitude of the exponential with the longest relaxation time. This makes $C(t)$ as single exponential as possible. It can be shown (see Appendix B) that $a^{\ddagger}$ turns out to be the root (i.e., the location of the node) of the eigenfunction of the diffusion operator with the lowest (in magnitude) nonzero eigenvalue. For high barriers, this, in turn, can be identified with the stochastic separatrix (Ryter, 1987). The stochastic separatrix is the dividing surface between reactants and products from which either is reached with equal probability. This property has been recently exploited (Du et al., 1998) as a practical way of finding the transition state for lattice models of protein folding.

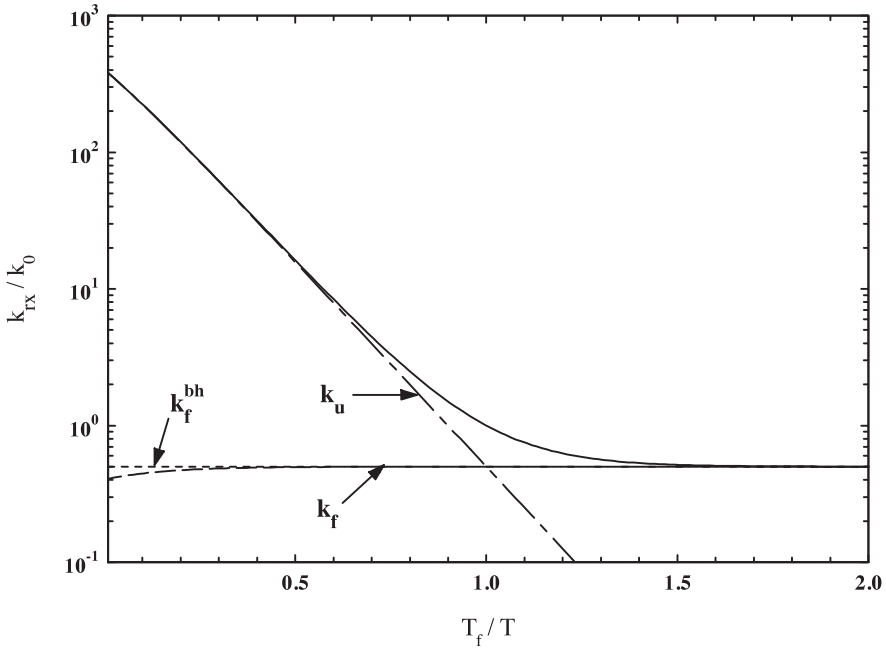

Fig. 6. Chevron plot for the reversible golf course. Reduced relaxation rate k_{rx}/k_0 as a function of T_f/T, for $a = 0.1$ and $\varepsilon = 6.91k_BT_f$. The value of $k_0 = 0.73D/R^2$ corresponds to the relaxation rate at $T = T_f$. The curves k_f and k_u represent the reduced folding and unfolding rates, respectively. For comparison, the reduced rate $k_f^{bh} = k_f/k_0$ (with k_f given in Equation 5) for the black-hole golf course is also displayed.

For a high parabolic barrier in one dimension, the location of the stochastic separatrix and the maximum of the potential are the same. In our problem, however, they are different. For example, when $a = 0.1$, the dividing surface obtained by solving Equation B3 with $q_1 = 0.8497$ is $a^{\ddagger} = 0.175$. If this choice of the dividing surface is used to calculate the correlation function $C(t)$, we find that the amplitudes of the two first exponential terms are $C_1 = 0.9913$ and $C_2 = 0.00466$, which should be compared with the coefficients in Equation 19 found using $a^{\ddagger} = 0.1$. It can be seen that $C(t)$ becomes even closer to being perfectly single exponential.

The entropic barrier $\Delta^{\ddagger} = 2k_BT \ln[1/a^{\ddagger}]$ is $3.49k_BT$ and $4.61k_BT$ for $a^{\ddagger} = 0.175$ and $a^{\ddagger} = 0.1$, respectively. This seems to suggest that the folding rate should be sensitive to the choice of the dividing surface or the transition state. However, this is not the case. In one dimension, for high barriers, both K_{eq} and k_{rx} (as obtained from the exact expression given in Equation 17) are insensitive to the choice of $a^{\ddagger}$ as long as it is located in a high free-energy region. For example, at $T = T_f$ we have $K_{eq}(a^{\ddagger} = 0.175)/K_{eq}(a^{\ddagger} = 0.1) = 1.008$ and $k_{rx}(a^{\ddagger} = 0.175)/k_{rx}(a^{\ddagger} = 0.1) = 0.997$.

Funnel landscape

We have seen that the Levinthal "paradox" can be viewed as the result of a high entropic barrier to folding. It can be resolved (Bryngelson & Wolynes, 1989; Zwanzig et al., 1992; Karplus, 1997) by adding an energetic bias in the direction of the folded states, which reduces the height of the free energy barrier. In this section, we consider the influence of a constant force F that drives the system toward the folded states. Specifically, we repeat the analysis of the previous sections using the potential given in Equation 3, which looks like a funnel in two dimensions (see Fig. 2b).

Choosing $r = a$ as the dividing surface, the equilibrium constant is

$$K_{eq} = \frac{k_f}{k_u} = \frac{e^{\beta\varepsilon}\int_0^a x^2\, e^{-\beta F x}\, dx}{\int_a^1 x^2\, e^{-\beta F x}\, dx}$$

$$= \frac{2e^{\beta\varepsilon} - [1 + (1 + \beta F a)^2]e^{\beta(\varepsilon - Fa)}}{[1 + (1 + \beta F a)^2]e^{-\beta F a} - [1 + (1 + \beta F)^2]e^{-\beta F}}. \quad (20)$$

The folding temperature T_f, which now depends on ε and F, can be obtained by numerically solving $K_{eq}(T_f) = 1$. For $a = 0.1$ and $F = \varepsilon/2$, we find that $\varepsilon = 5.27k_BT_f$, which is about 1.3 times the folding temperature of the reversible golf course. In Figure 5, the fraction of population in the native state $[f = K_{eq}/(1 + K_{eq})]$ is plotted for several values of F. As F increases, the curves remain sigmoidal but are shifted to higher temperatures. The inset of this figure shows that the folding temperature increases almost linearly with F.

The one-dimensional potential of mean force for the three-dimensional spherically symmetric potential given in Equation 3 is

$$V_{pmf}(x) = \begin{cases} -\varepsilon - 2k_BT \ln(x) + F(x-1), & 0 \le x \le a, \\ -2k_BT \ln(x) + F(x-1), & a < x \le 1. \end{cases} \quad (21)$$

The relaxation rate $k_{rx} = k_f + k_u$ in this case can be obtained by using this expression in Equation 17. This potential of mean force is shown in Figure 7. A new feature of this potential, not seen in Figure 5, is the shift of the position of the minimum in the unfolded space from $x = 1$. The location of this minimum, denoted by x_m, is given by $x_m = 1$ for $T \ge T_m$ and $x_m = T/T_m$ for $a < T/T_m < 1$, where $T_m = F/(2k_B)$. Thus, one effect of the constant force F is to make the ensemble of unfolded configurations more compact. The other effect is that it reduces the height of the folding barrier:

$$\Delta = V_{pmf}(a) - V_{pmf}(x_m) = 2k_BT \ln\left(\frac{x_m}{a}\right) - F(x_m - a), \quad (22)$$

and thus increasing the rate of folding. For $a = 0.1$, $\varepsilon = 5.27k_BT_f$, $F = \varepsilon/2$, and $T/T_f = 0.5$, 1, 2 we have $x_m = 0.38$, 0.76, 1, respectively. The corresponding barrier heights are $\Delta/k_BT \simeq 1.2$, 2.3, 3.42, which should be compared to the barrier height of $4.61k_BT$ found for the golf-course landscape.

Figure 8 displays two three-dimensional trajectories on a funnel landscape. As in Figure 3, both trajectories start from the coordinate (0, 0, 0.8) and terminate at the surface $r = a$. Clearly, these two trajectories explore different regions of the conformational space. Compared with Figure 3, it can be seen that these trajectories are now less dense because the rate of folding has increased. In addition, these trajectories explore less of the configuration space than before, and appear to linger around the surface with radius specified by the minimum of the potential of mean force ($r_m = x_m \simeq 0.76$, in this case).

Figure 9 shows the temperature dependence of the relaxation rate k_{rx}, the folding rate k_f, and the unfolding rate k_u. These were obtained using $V_{pmf}(x)$ given by Equation 21 in Equation 17 for

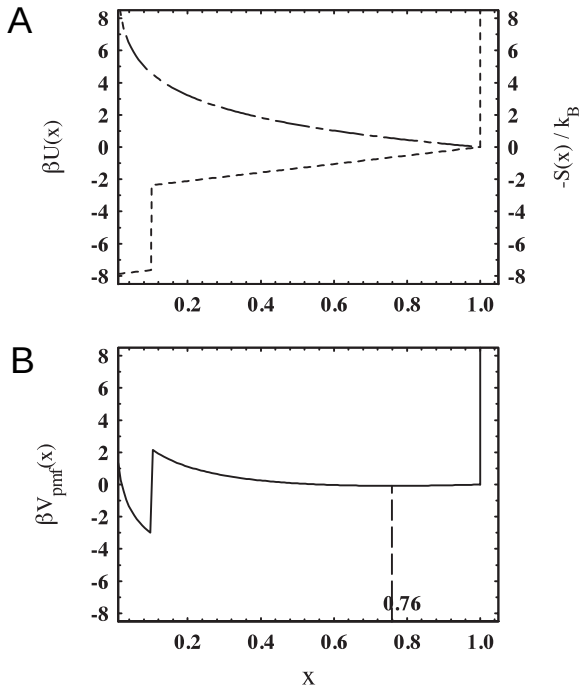

Fig. 7. Potential of mean force (or free energy) for the funnel landscape at $T = T_f$ for $a = 0.1$, $\varepsilon = 5.27k_BT_f$, and $F = \varepsilon/2$. **A:** Dashed line represents the potential energy $\beta U(x)$ and the dot-dashed line, the entropy $-S(x)/k_B$. **B:** The potential of mean force, $\beta V_{pmf}(x) = \beta U(x) - S(x)/k_B$.

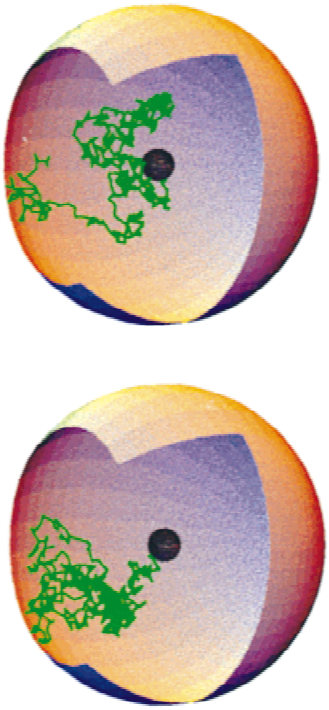

Fig. 8. Two three-dimensional folding trajectories on a funneled landscape. Both start from the coordinate (0, 0, 0.8) and terminate upon reaching the surface $r = a$ enclosing the native configurations. $a = 0.1$ as in Figure 3 and the slope of the funnel (i.e., the magnitude of the force that drives the system toward the native state) is $F = \varepsilon/2 = 2.635 k_B T$.

k_{rx}, and relations $k_f = k_{rx} K_{eq}/(1 + K_{eq})$ and $k_u = k_{rx}/(1 + K_{eq})$. Comparing this to Figure 6 where $F = 0$, we see that in both cases the unfolding rate increases with increasing the temperature. This is the usual Arrhenius behavior with activation energy ε.

The folding rate, on the other hand, decreases with increasing temperature so that the apparent activation energy is negative (Oliveberg et al., 1995; Muñoz et al., 1997). This is a simple consequence of the fact that, in our model, the energy of the unfolded states decreases the closer they are to the native state. To make this a bit more quantitative, consider the expression for the folding barrier height given by Equation 22. This barrier height is a sum of a positive entropic term and a negative energetic term. Because the folding rate is roughly proportional to $e^{-\Delta/k_B T}$, one finds that $k_f \propto e^{F(x_m - a)/k_B T}$.

The correlation function $C(t)$, which describes equilibrium population fluctuations, cannot be found analytically when $F \neq 0$. We calculated $C(t)$ by solving the one-dimensional diffusion equation involving $V_{pmf}(x)$ numerically using finite difference techniques (see, e.g., Bicout & Szabo, 1998). The time dependence of $C(t)$ is shown in Figure 10 for two values of F. The potential of mean force for $F = \varepsilon/2$ is shown in Figure 7. Because the barrier for folding is larger than $k_B T$, $C(t)$ is expected to be and, in fact, is nearly single exponential.

When F increases at fixed temperature (or the temperature is decreased at fixed F), the barrier for folding can disappear completely (i.e., the energetic contribution exactly counterbalances the entropic term). From Equation 21, it follows that this occurs when $Fa/k_B T \geq 1$. This has been called the "downhill scenario" (Bryngelson et al., 1995). As an example, consider the situation where

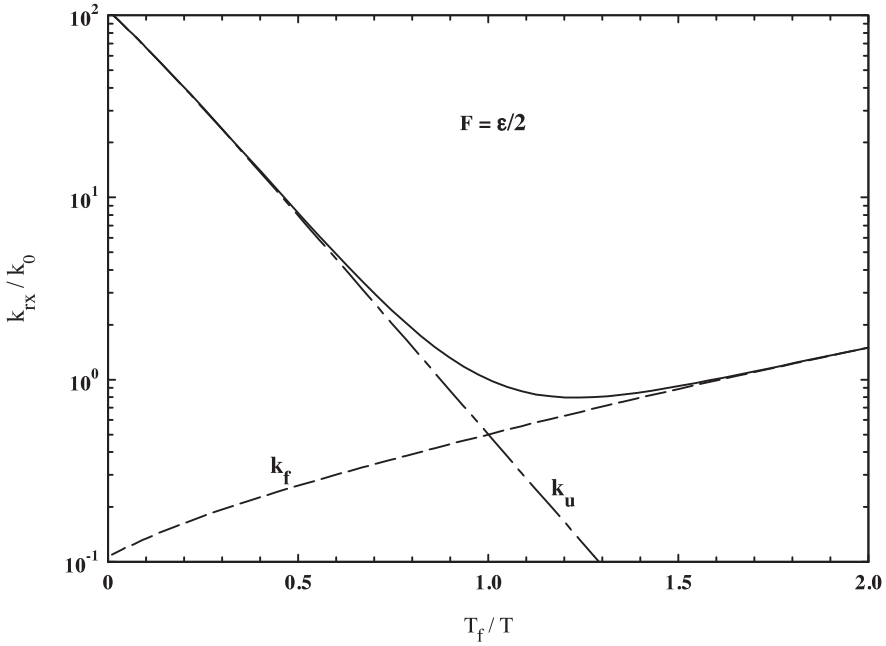

Fig. 9. Chevron plot for the funnel landscape. Reduced relaxation rate k_{rx}/k_0 (solid line), as a function of T_f/T, for $a = 0.1$ and $\varepsilon = 5.27 k_B T_f$. The value of $k_0 = 2.76 D/R^2$ corresponds to the relaxation rate at $T = T_f$. The curves k_f and k_u (dashed lines) represent the reduced folding and unfolding rates, respectively.

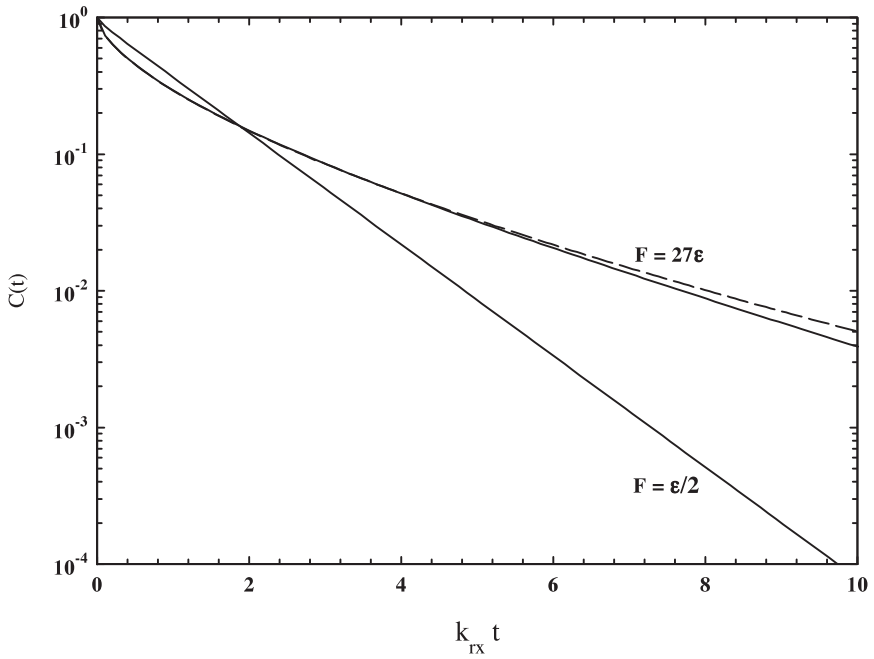

Fig. 10. The correlation function $C(t)$ that describes equilibrium population fluctuations vs. the reduced time, $k_{rx}t$ at $T = T_f$ for $a = 0.1$ for funnel landscapes. The curve with $F = \varepsilon/2$ corresponds to $\varepsilon = 5.27k_BT_f$ and relaxation rate $k_{rx} = 2.76D/R^2$. The large funnel slope $F = 27\varepsilon$ corresponds to the downhill scenario (where there is no free-energy barrier) with $\varepsilon = 0.74k_BT_f$ and relaxation rate $k_{rx} = 206D/R^2$. The dashed line represents the stretched exponential fit, $\exp\{-(1.4k_{rx}t)^{0.633}\}$, to the data.

$a = 0.1$, $\varepsilon = 0.74k_BT_f$ and $F = 27\varepsilon$. For these values of the parameters $K_{eq} = 1$. Because there is no barrier in the free energy, the kinetics is expected to be multiexponential. This is confirmed by the correlation function $C(t)$ labeled $F = 27\varepsilon$ in Figure 10. Amusingly, a stretched exponential, $\exp\{-(1.4k_{rx}t)^{0.633}\}$ (dashed line), turns out to be a very good fit to $C(t)$ for all but very long times where $C(t)$ is less than 0.01.

Figure 11 shows a typical trajectory for a downhill scenario. As in Figures 3 and 8, the trajectory starts at (0, 0, 0.8) and terminates at the surface $r = a$. Clearly, these trajectories are quite different. The Chan and Dill (1998) analogy involving an ensemble of skiers proceeding downhill to the native state is applicable only for the case where there is no barrier in the free energy and where the kinetics is nonexponential.

Rugged energy landscape

The influence of local minima and maxima of the potential energy surface of protein folding can be roughly treated in the framework of our model by making the diffusion coefficient depend on the reaction coordinate and/or temperature. Roughness of the potential surface slows down diffusion because local barriers must be surmounted for the system to jump from one configuration to another. Because some regions of the conformational space can be rougher than others, the diffusion coefficient can depend on x. This refinement will not be considered here.

Because local barriers must be overcome, diffusion is an activated process and, hence, $D(\mathrm{T}) = De^{-\Delta V/k_BT}$, where ΔV is the activation energy. When the amplitude of roughness is random,

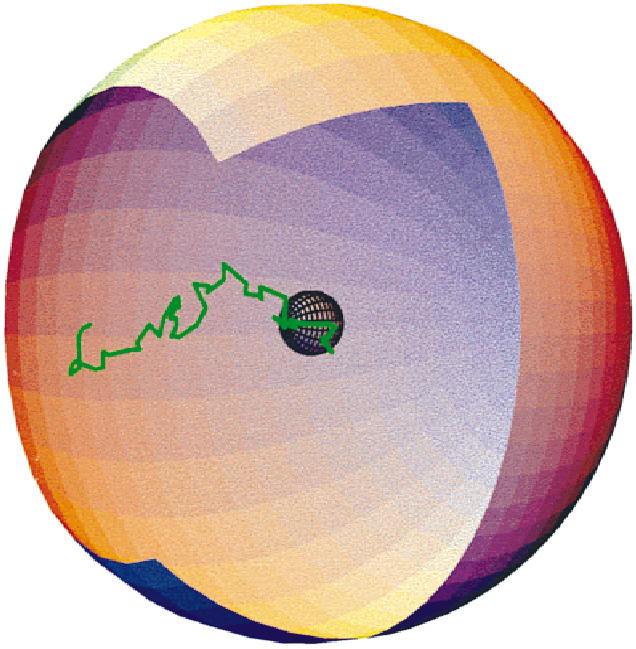

Fig. 11. Typical three-dimensional folding trajectory when the force F that drives the system toward the native state is so strong that the free-energy barrier disappears and the kinetics becomes multiexponential (i.e., downhill scenario). The diffusion process starts from the coordinate (0, 0, 0.8) and terminates at the surface $r = a = 0.1$ enclosing the native configurations. The funnel slope is $F = 27\varepsilon = 20k_BT$. Compare this with Figures 3 and 8, where the kinetics is exponential.

then (Ferry et al., 1953; Zwanzig, 1988; Bryngelson & Wolynes, 1989), $D(T) = De^{-\sigma^2/(k_BT)^2}$, which has a stronger temperature dependence.

We have seen in the previous section that in the presence of a constant force that drives the system toward the native state, the folding rate increases as the temperature decreases. This result was obtained under the assumption that the diffusion coefficient was temperature independent. Because the rate is proportional to D, it is clear that, if D decreases with decreasing the temperature sufficiently rapidly, the folding rate must eventually decrease as the temperature decreases. This is confirmed by illustrative calculations presented in Figure 12. Such turnover behavior of the folding rate as a function of temperature has been found in lattice simulations.

Summary

In this paper we have studied a simple exactly solvable model that allows us to illustrate some of main characteristic features of protein folding. In this model, each configuration of the polypeptide is represented by a point **r** inside a d-dimensional sphere of radius R. The dynamics of the system is described by diffusion within a closed sphere in the presence of a spherically symmetric potential $U(r)$ that depends only on the distance, $|\mathbf{r}| = r$, from the origin. Two types of potential surfaces have been considered.

In the simplest version of this model, we considered the radial step-function potential that divides the configurational space into two regions: $U(r) = -\varepsilon$ (with $\varepsilon > 0$) for configurations enclosed in the sphere of radius R_N, and $U(r) = 0$ otherwise. Any configuration with $0 \le r \le R_N$ is identified with the native state, and the rest of the space represents unfolded configurations. ε is the energy difference between folded and unfolded configurations. As shown in Figure 2a, this potential surface leads to the golf-course landscape.

When $\varepsilon \to \infty$, the model is even simpler and leads to the "black-hole" golf-course energy landscape describing irreversible folding. The fraction of unfolded configurations on this energy landscape decays exponentially for sufficiently small radius of the native sphere. Why the folding kinetics is single exponential can be understood in terms of the one-dimensional potential of mean force (or free energy) $V_{pmf}(x)$ (see Equation 8) along the reaction coordinate $x = |\mathbf{r}|$. Examination of $V_{pmf}(x)$ shows that the origin of single exponential folding kinetics stems from the presence of a high (compared to k_BT) configurational entropic barrier (given in Equation 9) between unfolded (reactants) and folded (products) states. In three dimensions, for $a = R_N/R = 0.1$, the entropic barrier height is $\Delta = 4.61k_BT$.

When ε is finite, the model describes reversible folding. Because the unfolded configuration space is not affected by making ε finite, the height Δ of the entropic barrier is the same as before. When $\Delta \gg k_BT$, the correlation function $C(t)$, which describes the relaxation of equilibrium population fluctuations, is virtually identical to $e^{-k_{rx}t}$, where k_{rx} is the relaxation rate given in Equation 18. It follows that the folding dynamics can be well described by a two-state chemical kinetics. Assuming that the diffusion coeffi-

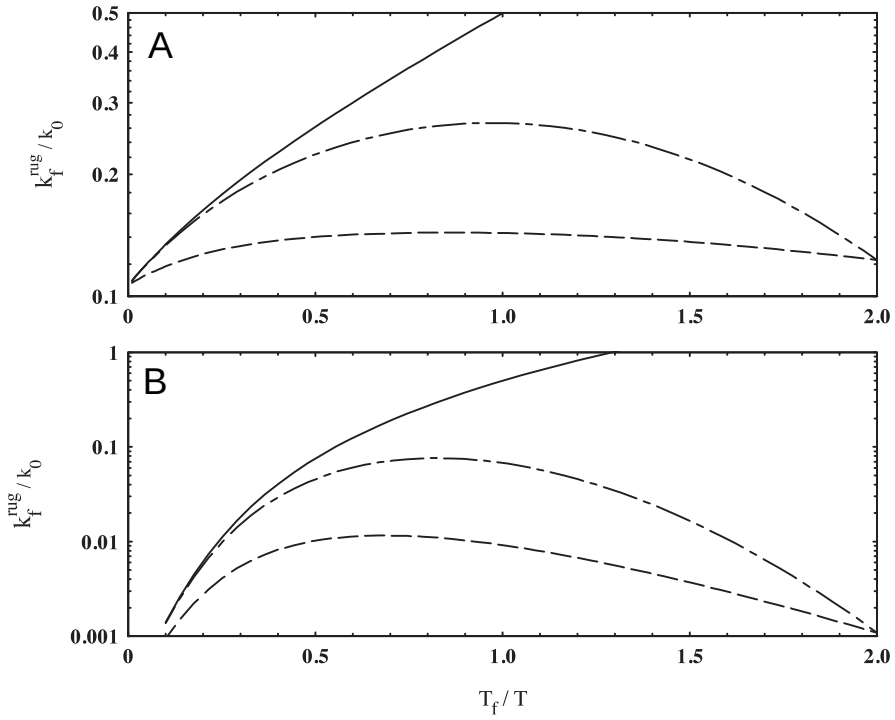

Fig. 12. Effect of ruggedness on the folding rate for the funnel energy landscape. Reduced folding rate as a function of T_f/T. **A:** The radius of native sphere is $a = 0.1$, $\varepsilon = 5.27k_BT_f$ and slope $F = 0.5\varepsilon$. The value of $k_0 = 2.76D/R^2$ corresponds to twice the folding rate on the smooth energy landscape at $T = T_f$. Solid line represents k_f/k_0 on the smooth funnel energy landscape, the dashed and dotted-dashed lines to k_f^{rug}/k_0 for uniform and random activated diffusion with $\Delta V = 1.25k_BT_f$ and $\sigma^2 = 0.625(k_BT_f)^2$, respectively (see the text). For comparison, $k_f/k_0 \simeq 1.5$ at $T = T_f/2$. **B:** Same thing as in **A** for $a = 0.01$, $\varepsilon = 5.74k_BT_f$, $F = 5\varepsilon$ and $k_0 = 126.17D/R^2$. The dashed and dotted-dashed lines correspond to k_f^{rug}/k_0 for $\Delta V = 4k_BT_f$ and $\sigma^2 = 2(k_BT_f)^2$, respectively. $k_f/k_0 = 3.24$ at $T = T_f/2$.

cient is independent of temperature, we examined the behavior of the relaxation rate k_{rx} as a function of T_f/T, where T_f is the folding temperature. Both k_{rx} and the unfolding rate k_u have an Arrhenius behavior for $T > T_f$, because the unfolding is a thermally activated process because of the energy difference between native and unfolded configurations. For $T < T_f$, on the other hand, both k_{rx} and the folding rate k_f are almost independent of temperature because the barrier in the folding direction is purely entropic.

Next, we have addressed the issue of finding the transition state, i.e., the optimal dividing surface between unfolded (reactants) and folded (products) conformations. Usually, the dividing surface is chosen at the top of the barrier in the free-energy surface along the reaction coordinate. Thus, the transitions state in our three-dimensional model is an ensemble of configurations specified by the surface $r = a$. As an alternative, the position $a^\ddagger$ of the dividing surface is chosen so as to make the correlation function $C(t)$ as single an exponential as possible. With this definition, we find that the transition state corresponds to the node of the eigenfunction of the diffusion operator with the lowest (in magnitude) nonzero eigenvalue. For high barriers, this is the same as the stochastic separatrix, which is the the dividing surface between reactants and products from which either is reached with equal probability. For $a = 0.1$ in three dimensions and at $T = T_f$, we find that $a^\ddagger = 0.175$, and the corresponding entropic barrier height is $\Delta^\ddagger = 3.49 k_B T$. However, when the relaxation rate is calculated using Equation 17, which is exact for any one-dimensional potential, the result is insensitive to the location $a^\ddagger$ as long as the free energy at $a^\ddagger$ is high.

The second form of the potential surface that we have considered is the radial step-function potential plus a linear (in r) attractive potential of constant slope F (i.e., there is an additional constant force F that drives the system toward the native state). This leads to the funnel energy landscape represented in Figure 2b. The folding temperature increases with F and the height Δ of the folding barrier decreases with F. When Δ is sufficiently high compared to $k_B T$, the correlation function $C(t)$ is again well approximated by a single exponential with the relaxation rate k_{rx}, which can be obtained using Equation 17. Thus, the folding dynamics on the funnel landscape can be described by a two-state chemical kinetics as long as F is not too large (see below). Assuming that the diffusion coefficient is temperature independent, the relaxation rate k_{rx} shows a chevron-like behavior as a function of T_f/T (see Fig. 9). As before, the unfolding rate k_u has the usual Arrhenius behavior expected for an activated process. The folding rate k_f, on the other hand, now increases with decreasing the temperature so that the apparent activation energy is negative. This results from the fact that the folding barrier Δ is a sum of a positive entropic term and a negative energetic term (proportional to F). When $Fa \geq k_B T$, the folding barrier Δ vanishes completely and the free-energy decreases monotonically. For the "downhill scenario," the folding kinetics is nonexponential, as illustrated in Figure 10.

Finally, we have briefly discussed how the effect of local minima and maxima (i.e., ruggedness) of the potential energy surface can be taken into account by simply making the diffusion coefficient decrease as the temperature is lowered. The folding rate now exhibits a non-Arrhenius concave turnover behavior as a function of T_f/T. This behavior can also arise if the parameters in the attractive potential energy are temperature dependent because of hydrophobic interactions.

The model analyzed in this paper is arguably the simplest example of a reaction that is energetically favorable but entropically unfavorable. It is not intended to be a realistic model of the molecular mechanism of protein folding or to be used to analyze experimental data. However, it does show that a surprising number of features of more complex models and even some experiments are the consequence of very simple assumptions, and it helps clarify a number of issues and concepts that arise in thinking about protein folding.

Acknowledgments

Parts of this paper were given as a lecture by D.J.B. in the spring school on "Self-Organization in Biological Systems" at the École de Physique, Les Houches, France, May 24–29, 1998.

We thank Sasha Berezhkovskii, Bill Eaton, Gerhard Hummer, Goury Jas, Victor Muñoz, and Bob Zwanzig for helpful comments on this paper.

References

Adam G, Delbrück M. 1968. Reduction of dimensionality in biophysical diffusion processes. In: Rich A, Davidson N, eds. *Structural chemistry and molecular biology*. Freeman: San Francisco. p 198.

Bicout DJ, Szabo A. 1997. First passage times, correlation functions, and reaction rates. *J Chem Phys 106*:10292–10298.

Bicout DJ, Szabo A. 1998. Electron transfer reaction dynamics in non-Debye solvents. *J Chem Phys 109*:2325–2338.

Bryngelson JD, Onuchic JN, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Genet 21*:167–195.

Bryngelson JD, Wolynes PG. 1989. Intermediates and barrier crossing in random energy model (with applications to protein folding). *J Phys Chem 93*:6902–6915.

Chan HS, Dill KA. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins Struct Funct Genet 30*:2–33.

Chandler D. 1978. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J Chem Phys 68*:2959–2970.

Chen Y-D, Rubin RJ, Szabo A. 1993. Fluorescence dequenching kinetics of single cell–cell fusion complexes. *Biophys J 65*:325–333.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles protein folding—A perspective from simple exact models. *Protein Sci 4*:561–602.

Dobson CM, Sali A, Karplus M. 1998. Protein folding: A perspective from theory & experiments. *Angew Chem Int Ed 37*:869–893.

Du R, Pande VS, Groeberg AYu, Tanaka T, Shakhnovich ES. 1998. On the transition coordinate for the protein folding. *J Chem Phys 108*:334–350.

Ferry JD, Grandine ER, Fitzgerald ER. 1953. Viscoelastic relaxation of polyisobutylene. *J Appl Phys 24*:911–921.

Fersht AR. 1995. Optimization of rates of protein-folding—The nucleation condensation mechanism and its implications. *Proc Natl Acad Sci USA 92*: 10869–10873.

Jackson SE, Fersht AR. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry 30*:10428–10435.

Karplus M. 1997. The Levinthal paradox: Yesterday and today. *Fold Des 2*:S69–S75.

Karplus M, Sali A. 1995. Theoretical studies of protein folding and unfolding. *Curr Opin Struct Biol 5*:58–73.

Karplus M, Weaver DL. 1976. Protein-folding dynamics. *Nature 260*:404–406.

Leopold PE, Montal M, Onuchic JN. 1992. Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci USA 89*:8721–8725.

Levinthal C. 1969. Mossbauer spectroscopy in biological systems. In: Debrunner P, Tsibiris JCM, Münck E, eds. *Proceedings of a meeting held at Allerton House, Monticello, Illinois*. Urbana: University of Illinois Press. pp 22–24.

Muñoz V, Henry ER, Hofritchter J, Eaton WA. 1998. A statistical mechanical model for β-hairpin kinetics. *Proc Natl Acad Sci USA 95*:5872–5879.

Muñoz V, Thompson AP, Hofritchter J, Eaton WA. 1997. Folding dynamics and mechanism of β-hairpin formation. *Nature 390*:196–199.

Oliveberg M, Tan Y-J, Fersht AR. 1995. Negative activation enthalpies in the kinetics of protein folding. *Proc Natl Acad Sci USA 92*:8926–8929.

Onuchic JN, Luthey-Schulten Z, Wolynes PG. 1997. Theory of protein folding: The energy landscape perspective. *Ann Rev Phys Chem 48*:545–600.

Onuchic JN, Wolynes PG, Luthey-Schulten Z, Socci ND. 1995. Toward an

outline of the topography of a realistic protein-folding funnel. *Proc Natl Acad Sci USA 92*:3626–3630.
Orland H, Garel T, Thirumalai DT. 1996. Analytical theories of protein folding. In: Elber R, ed. *Recent developments in theoretical studies of proteins*. Singapore: World Scientific. pp 197–268.
Pande VS, Groeberg AYu, Tanaka T, Rokhsar DS. 1998. Pathways for protein folding: Is a new view needed? *Curr Opin Struct Biol 8*:68–79.
Ryter D. 1987. Noise-induced transitions in a double-well potential at low friction. *J Stat Phys 49*:751–765.
Schulten K, Schulten Z, Szabo A. 1981. Dynamics of reactions involving diffusive barrier crossing. *J Chem Phys 74*:4426–4432.
Shakhnovich EI. 1997. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr Opin Struct Biol 7*:29–40.
Skinner JL, Wolynes PG. 1979. Relaxation processes and chemical kinetics. *J Chem Phys 69*:2143–2150.
Socci ND, Onuchic JN, Wolynes PG. 1996. Diffusive dynamics of the reaction coordinate for the protein folding funnels. *J Chem Phys 104*:5860–5868.
Szabo A, Schulten K, Schulten Z. 1980. First passage time approach to diffusion controlled reactions. *J Chem Phys 72*:4350–4357.
Thirumalai D, Woodson SA. 1996. Kinetics of folding of proteins and RNA. *Acc Chem Res 9*:433–439.
Wolynes PG, Onuchic JN, Thirumalai D. 1995. Navigating the folding routes. *Science 267*:1619–1620.
Zwanzig R. 1988. Diffusion in a rough potential. *Proc Natl Acad Sci USA 85*:2029–2030.
Zwanzig R. 1995. Simple model of protein folding kinetics. *Proc Natl Acad Sci USA 92*:9801–9804.
Zwanzig R, Szabo A, Bagchi B. 1992. Levinthal's paradox. *Proc Natl Acad Sci USA 89*:20–22.

Appendix A. Relaxation dynamics on the reversible golf-course landscape

The dynamics is described by the probability density $P(r,t)$ of finding the system with configuration r at time t. This satisfies

$$\frac{\partial P(r,t)}{\partial t} = \left(\frac{D}{R^2}\right)\frac{1}{r^2}\frac{\partial}{\partial r}\,r^2 e^{-\beta U(r)}\,\frac{\partial}{\partial r}\,e^{\beta U(r)}P(r,t), \tag{A1}$$

with the reflecting boundary conditions at edges:

$$\frac{\partial}{\partial r}\,[e^{\beta U(r)}P(r,t)]|_{r=0,1} = 0. \tag{A2}$$

Because the potential $U(r)$, as defined in Equation 2, is discontinuous at $r = a$, Equation A1 has to be solved separately in the region $r \le a_-$ (folded space) and in the region $r \ge a_+$ (unfolded space). These solutions are then matched using the conditions:

$$e^{\beta U(r)}P(r,t)|_{r=a_-} = e^{\beta U(r)}P(r,t)|_{r=a_+}, \tag{A3a}$$

$$e^{-\beta U(r)}\,\frac{\partial}{\partial r}[e^{\beta U(r)}P(r,t)]|_{r=a_-} = e^{-\beta U(r)}\,\frac{\partial}{\partial r}[e^{\beta U(r)}P(r,t)]|_{r=a_+}. \tag{A3b}$$

Using the transformation, $P(r,t) = e^{-q^2Dt/R^2}\,e^{-\beta U(r)/2}\psi(r)/r$, Equation A1 becomes

$$\frac{e^{\beta U(r)/2}}{r^2}\,\frac{d}{dr}\,r^2 e^{-\beta U(r)}\,\frac{d}{dr}\left[e^{\beta U(r)/2}\,\frac{\psi(r)}{r}\right] = -q^2\,\frac{\psi(r)}{r}. \tag{A4}$$

This equation is first solved in both folded and unfolded regions, and the two solutions are matched as indicated above. In this way, we find that the Green's function (i.e., the probability density of being at r at time t given that the system was at r_0 initially) of the diffusion equation (A1) is given by

$$G(r,t|r_0) = \frac{e^{-\beta[U(r)-U(r_0)]/2}}{4\pi r r_0}\sum_{n=0}^{\infty}\psi_n(r_0)\psi_n(r)e^{-q_n^2Dt/R^2}, \tag{A5}$$

where the eigenvalues q_n are solutions of

$$\frac{q_n\tan[q_n(1-a)]+1}{q_n-\tan[q_n(1-a)]} = \frac{q_n a e^{\beta\varepsilon}+(1-e^{\beta\varepsilon})\tan[q_n a]}{q_n a\tan[q_n a]}, \tag{A6}$$

and the corresponding eigenfunctions $\psi_n(r)$ are given by

$$\psi_n(r) = A_n\sin[q_n r]; \quad 0 \le r \le a, \tag{A7a}$$

$$= A_n e^{-\beta\varepsilon/2}\sin[q_n a]\left(\frac{\sin[q_n(1-r)]-q_n\cos[q_n(1-r)]}{\sin[q_n(1-a)]-q_n\cos[q_n(1-a)]}\right);$$

$$a \le r \le 1, \tag{A7b}$$

with the constants A_n found from the orthonormality condition, $\int_0^1\psi_m(r)\psi_n(r)\,dr = \delta_{mn}$. Note that $q_0 = 0$, and $[\psi_0(r)]^2 = 4\pi r^2 p_{eq}(r)$, where $p_{eq}(r) = e^{-\beta U(r)}/[\int_0^1 4\pi r^2 e^{-\beta U(r)}\,dr]$ is the equilibrium distribution.

Knowing now the Green's function, we next turn to the calculation of the correlation function $C(t)$ to study the relaxation dynamics, and thus folding kinetics.

The native state in this model is defined by configurations with $r \le a$, i.e., the fractional number configurations $N(r)$ in the native state is $N(r) = 1$ for $r \le a$ and zero for $a < r \le 1$. Hence, the deviation from equilibrium is $\delta N(r) = N(r) - N_{eq}$, where $N_{eq} = f$ is the fraction of native population at equilibrium given by $f = \int_0^1 4\pi p_{eq}(r)N(r)\,dr$. The correlation function $C(t)$ that describes fluctuations in the population at equilibrium is then defined as

$$C(t) = \frac{\langle\delta N(t)\,\delta N(0)\rangle}{\langle(\delta N)^2\rangle}$$

$$= \frac{\displaystyle\int_0^1 4\pi r_0^2\,dr_0\int_0^1 4\pi r^2\,dr\,\delta N(r)G(r,t|r_0)\,\delta N(r_0)p_{eq}(r_0)}{\displaystyle\int_0^1 4\pi r^2[\delta N(r)]^2 p_{eq}(r)\,dr}, \tag{A8}$$

where the Green's function $G(r,t|r_0)$ is given in Equation A5. Evaluating the integrals, we find that

$$C(t) = \frac{1}{f(1-f)}\sum_{n=1}^{\infty}\left[\int_0^a\psi_0(r)\psi_n(r)\,dr\right]^2 e^{-q_n^2Dt/R^2},$$

$$= \frac{3a}{1-f}\sum_{n=1}^{\infty}\{A_n j_1[q_n a]\}^2 e^{-q_n^2Dt/R^2}, \tag{A9}$$

where $j_1[x] = \sin[x]/x^2 - \cos[x]/x$ is the spherical Bessel function of the order of 1. The relaxation rate k_{rx} is given by

$$\frac{1}{k_{rx}} = \int_0^\infty C(t)\,dt = \frac{3aR^2}{(1-f)D}\sum_{n=1}^{\infty}\left\{\frac{A_n j_1[q_n a]}{q_n}\right\}^2. \tag{A10}$$

It worthwhile to note that this infinite series for the expression of k_{rx} is equivalent to the closed form given in Equation 18 obtained using the mean first passage theory.

Appendix B. Location of the transition state

Let $a^{\ddagger}$ be the position of the dividing surface between reactants and products. It follows from Appendix A that the correlation function $C(t)$, describing fluctuations in population at equilibrium is given by

$$C(t) = \frac{1}{f(1-f)} \sum_{n=1}^{\infty} \left[\int_0^{a^{\ddagger}} \psi_0(r)\psi_n(r)\, dr \right]^2 e^{-q_n^2 Dt/R^2} = \sum_{n=1}^{\infty} C_n e^{-t/\tau_n}, \tag{B1}$$

where the eigenvalues q_n and eigenfunctions $\psi_n(r)$ are given in Equations A6, A7a, and A7b, respectively, and the relaxation times τ_n are such that $\tau_1 > \tau_2 > \cdots$. One can choose the position $a^{\ddagger}$ of the dividing surface so as to maximize the first amplitude

$$C_1 = \frac{1}{f(1-f)} \left[\int_0^{a^{\ddagger}} \psi_0(r)\psi_1(r)\, dr \right]^2, \tag{B2}$$

i.e., make $C(t)$ as single exponential as possible. As in the quantum mechanics, the lowest eigenfunction, $\psi_0(r) = [4\pi r^2 p_{eq}(r)]^{1/2}$, has no nodes, and the next eigenfunction has one, given by $\psi_1(r^{\ddagger}) = 0$. Thus, $\psi_1(r)$ is either negative for $r < r^{\ddagger}$ and positive for $r > r^{\ddagger}$, or vice versa. It follows that C_1 will be maximized if the upper limit of the integral in Equation B2 coincides with the node of ψ_1, i.e., when the dividing surface $a^{\ddagger}$ for transition states is chosen such that

$$\psi_1(a^{\ddagger}) = 0 \Rightarrow \tan[q_1(1 - a^{\ddagger})] = q_1, \tag{B3}$$

where $\psi_1(r)$ is given in Equation A7b (because $a^{\ddagger} \geq a$) and q_1 is the first root of the eigenvalue equation A6. Because q_1 depends on temperature, it follows from Equation B3 that $a^{\ddagger}$ is also temperature dependent.