A simple formalism on dynamics of proteins on potential energy landscapes

RAJAMANICKAM MURUGAN AND SHYMALAVA MAZUMDAR

Department of Chemical Sciences, Tata Institute of Fundamental Research (TIFR), Colaba, Mumbai, India, 400005

(RECEIVED August 1, 2003; FINAL REVISION October 10, 2003; ACCEPTED October 10, 2003)

Abstract

We present a simple formalism for the dynamics of proteins on a potential energy landscape, using connectedness of configurational domains as an order parameter. This formalism clearly shows that the energy bias required to form a unit correct contact toward the native configuration of a two-state folder, to overcome Levinthal's paradox, is $E_{bias} \cong RT \ln 2$. This result agrees well with earlier studies and indicates that the bias is mainly due to hydrophobic interaction. Further investigations have shown that the landscape funnel could be experimentally mapped onto a two-dimensional space formed by denaturant concentration and the connectedness of configurational domains. The theoretical value of the depth-of-folding funnel in terms of denaturant concentration has been calculated for a model protein (P450cam), which agrees well with the experimental value. Using our model, it is also possible to explain the turnover nature of heat-capacity change upon unfolding of proteins and the existence of enthalpy and entropy convergence temperatures during unfolding without any strict assumptions as proposed in earlier studies.

Keywords: Protein dynamics; energy landscape; folding funnel; convergence temperatures

Understanding the dynamics of macromolecules such as proteins and nucleic acids is important to elucidate their functional role in biological systems. Macromolecular dynamics differs from that of simple molecules in such a way that we cannot describe the processes here just using simple two-dimensional diagrams with free energy as the ordinate and reaction coordinate as the abscissa; instead, the potential energy landscapes are used to describe such dynamics. The theory of energy landscapes has already been well developed and applied to problems such as protein folding and protein–nucleic acid interactions (Leopold et al. 1992; Dill et al. 1993; Boczko and Brookes 1995; Bryngelson et al. 1995; Onuchic et al. 1995). Because the free-energy landscape of proteins is a complicated hypersurface, to simplify the problem, earlier theoretical works have suggested a number of reaction coordinates to describe protein dynamics such as:

1. Number of correct contacts. Here the native protein is said to have N number of correct contacts and the random coil a zero number of correct contacts.
2. Connectedness of configurational domains. Here the native configuration of the protein is probabilistically unconnected with the nonnative configurations under native conditions, whereas it is connected under denaturing conditions, and the degree of connectedness decides the position of the protein molecule on the energy landscape. The connectedness between native and nonnative configurations can be understood as the probability of finding the protein molecule simultaneously in both configurations.
3. Similarity index. The percentage of structural similarity when the unfolded protein configuration is projected over its native configuration.

Because the configurational space of a macromolecule is astronomically large, finding the global minimum just by a random search will not be possible unless there is a kind of energy bias toward the minimum, which forces us to assume a funnel-shaped landscape to explain the experimental ob-

Reprint requests to: Rajamanickam Murugan, Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai, India, 400005; e-mail: muruga@tifr.res.in; fax: 91-22-2280-4610.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.03347504.

servations. For example, the observed folding time and the folding rate of a protein cannot be explained unless we introduce an energy bias toward the native configuration. Moreover, extensive protein-folding simulations have confirmed this fact. But the main drawback of such simulations is that the energy landscape of a real protein is correlated and rugged and therefore far away from these theoretical models. Unfortunately, the usual thermodynamic and kinetics studies on proteins will tell us only about the overall free-energy change of folding/unfolding transitions and the average height of the transition-state ensemble (TSE), but we cannot get any information about the folding funnel and its nature owing to the fact that the TSE lies far above the funnel. The depth of the funnel is an important parameter, which decides the stability and foldability of a particular protein. Suppose that the folding funnel is shallow, then the protein's native conformation would be unstable, heterogeneous, and prone to conformational fluctuations even though its overall folding free energy is very high. From the depth of the folding funnel it is also possible to discriminate a well-folding protein from other proteins. In this article, we present a theory and an experimental methodology to map the protein free-energy landscape onto a two-dimensional space from which we can easily estimate the depth of the folding funnel, which is not possible in standard folding/unfolding studies. First, we present the formal theoretical principles, and then we present the experimental method with one example.

Theoretical concepts

The native configuration of a protein, which is denoted as n, lies at the bottom of the landscape funnel (Fig. 1), which is, of course, very narrow and deep enough to stabilize the native configuration. Here we say that the connectedness of the native form with rest of the configurational space, which is denoted as u, is almost zero. When we apply free-energy perturbations via temperature or denaturants, the native configuration will get probabilistically dispersed into the nonnative domain according to Boltzmann's distribution. Therefore, here we say that the connectedness of the native state (n) configuration with the rest of the configurational space (u) is increasing. This increasing trend will hold until the energy perturbation reaches the bottleneck of the folding funnel, after which the connectedness will start to decrease, which is caused by the fact that inside the funnel, the molecules can explore only a small configurational area and therefore get strongly connected, whereas once the molecule comes out of the funnel, it can explore a large area of landscape and therefore be weakly connected. Here we have complete information only about the native state (n), whereas the nonnative states (u) are not defined or heterogeneous. Therefore, the connectedness of n with u is meaningful but not vice versa. Now let us define this concept slightly more rigorously.

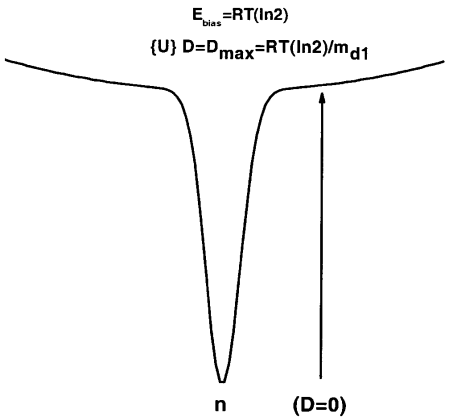

Figure 1. A simplified two-state landscape funnel, where n denotes the native state, which lies at the *bottom* of the funnel, and $\{U\}$ denotes the rest of the configurational space. The upward arrow is the denaturant scale (D), where $D = 0$ corresponds to the native state (n) and D_{max} (in moles per liter) corresponds to the bottleneck of the folding funnel. Here m_{d1} is the denaturant m value (in calories per mole per molar), which is obtained from the linear energy model.

Mathematical derivations

The dynamics of a protein on a potential energy landscape with respect to a symbolic reaction coordinate x can be well described by the Langevin equation, where the discrete variable x denotes the position of the protein molecule on the configurational space. Let us assume that $x = n$ denotes the native configuration, $x = \{u\}$ denotes the rest of the configurational space, and $n \notin u$, $n \cup u = \Omega$, $x \subset \{n, u\}$. Suppose that if the native form n contains C correct contacts (in other words, native contacts), then the nonnative form u can have a maximum of $C - 1$ correct contacts. Because the configurational space is extraordinarily large, for the purpose of analysis, we can assume that x is continuous and the corresponding Langevin equation can be written as:

$$d_t x = \frac{F(x)}{\gamma m} + \frac{\Gamma(t)}{\gamma} \tag{1}$$

Here $F(x)$ denotes the potential of mean force (pmf) acting on the protein molecule [i.e., $F(x) = -d_x\tilde{f}(x)$, where $\tilde{f}(x) = \Delta G(x)$ is the folding free energy], m is its effective mass, γ denotes the internal friction coefficient, and $\Gamma(t)$ is delta-correlated Gaussian noise, which satisfies the fluctuation dissipation theorem as:

$$\langle \Gamma(t)\Gamma(t') \rangle = (2\gamma kT/m) \times \delta(t - t').$$

The corresponding Fokker-Plank Equation (FPE) for the probability of finding the protein molecule at position x in time t can be given as (i.e., the Smoluchowski equation):

$$\partial_t P(x,t) = -(1/m\gamma) \times [\partial_x F(x) - kT \times \partial_x^2] P(x,t) \quad (2)$$

Here T is the absolute temperature (in kelvins) and k is the Boltzmann constant. Because we are interested in the stationary solution [i.e., at $\partial_t P(x, t) = 0$], that can be given as follows:

$$P_{st}(x) = N \times e^{-\tilde{f}(x)/kT} \quad (3)$$

Here

$$N = \left(\int_l^u e^{-\tilde{f}(x)/kT} dx \right)^{-1}$$

is the normalization constant. Now the stationary probability of finding the native form n is

$$P_{\mathrm{st}}(n) = N \times e^{-\tilde{f}(n)/kT}$$

and the connectedness of the native form n with rest of the configurational space u can be easily given as follows:

$$P_{st}(n \cap u) = N \times e^{-\tilde{f}(n)/kT} \times (1 - N \times e^{-\tilde{f}(n)/kT}) \quad (4)$$

Proof: Given that $n \notin u$ and $\{n, u\} = \Omega$, the following equality is true, which proves equation 4:

$$P_{\mathrm{st}}(n \cap u) = P_{\mathrm{st}}(n) \times P_{\mathrm{st}}(u) = P_{\mathrm{st}}(n) \times (1 - P_{\mathrm{st}}(n))$$

When we unfold the protein by denaturants such as urea and guanidine hydrochloride, the Linear Energy Model (LEM) predicts the following relation:

$$\tilde{f}(n) = -\Delta G_0^{H_2O} + m_d D \quad (5)$$

Here, $\tilde{f}(n) = -G_0{}^{\mathrm{H_2O}}$ is the folding free energy (the potential of mean force) acting on the native protein, m_{d} (in kilocalories per mole per molar) denotes the denaturant m value, and D (in moles per liter) denotes the denaturant's activity. From equations 3 and 5, the probability $P_{\mathrm{st}}(n, D)$ of observing the native configuration n at the denaturant concentration of D is given by

$$P_{\mathrm{st}}(n, D) = N \times e{-}m_d D/RT,$$

where N is the normalization constant. Using the boundary conditions $P_{\mathrm{st}}(n, 0)$ (therefore, $N = e^{-\Delta G_0{}^{\mathrm{H_2O}}}$/RT and $P_{\mathrm{st}}(n, \infty)$, it can be shown that:

$$P_{st}(n,D) = e^{-m_d D/RT} \quad (6)$$

From equations 4 and 6, the connectedness (denoted as λ') of the native configuration with the rest of the configurational space in the presence of denaturant can be given as:

$$\begin{aligned} \lambda' = P_{st}(n \cap u, D) &\cong P_{st}(n, D) \times (1 - P_{st}(n,D)) \\ &= e^{-m_d D/RT} \times (1 - e^{-m_d D/RT}) \end{aligned} \quad (7)$$

From the relation $\partial_D P(n \cap u, D) = 0$, we can easily show that the denaturant concentration at which the maximum of connectedness occurs is equal to $D_{\max} = RT \ln 2/m_d$. One also should note that at $D_{\max}$, $P_{\mathrm{st}}(n) = P_{\mathrm{st}}(u) = 0.5$, which is known as a stochastic separatrix; that is, the probabilities of existence of the protein molecule in the folded form and the unfolded form are equal. The stochastic separatrix is an abstract point that decides whether a protein molecule is in folded form [$P_{\mathrm{st}}(n) > 0.5$] or in unfolded form [$P_{\mathrm{st}}(u) > 0.5$], which otherwise can be viewed as the bottleneck of the folding funnel. There exists one more point that is different from $D_{\max}$ called the midpoint denaturant concentration, at which the population of the folded form is equal to the population of the unfolded form, that is, $D_{\mathrm{mid}} = \Delta G_0{}^{\mathrm{H_2O}}/m_d$, and therefore $\tilde{f}(n) = 0$ in equation 5. Here one should note that from the $D_{\max}$, which is obtained from the usual equilibrium unfolding experiments, it is not possible to derive any information about the folding funnel and the energy landscape. But using the aforementioned formalism, we show in the following sections that it is possible to estimate the energy bias toward the native configuration as well as the depth of the folding funnel.

Estimation of energy bias toward native configuration

It is known from earlier studies on denaturant-mediated protein unfolding that (Murugan 2003):

$$m_d \cong RT\, \alpha\beta/h_0, \quad (8)$$

where h_0 denotes the activity of water (55.5 M), α denotes the number of bound water molecules on protein (i.e., interaction potential of water with protein), and β is the number of water molecules interacting with the denaturant. Therefore, by putting the value of m_d in $D_{\max}$, we obtain $D_{\max} = h_0 \ln 2/\alpha\beta$. Because $D_{\max}$ is inversely proportional to water–protein interactions (i.e., α), it is an indirect measure of hydrophobic interactions within protein, that is, $D_{\max} \propto$ hydrophobic interactions. Comparing this with equation 5, we obtain the approximate energy bias per correct contact as $E_{\mathrm{bias}} = m_d \times D_{\max} \cong RT \ln 2$, which is very close to the earlier suggested value (Zwanzig et al. 1992) of few RTs per native contact and also indicates that it is mainly contributed by hydrophobic interactions! Here we

should note that n and u could be differentiated by a unit number of correct contacts.

Estimation of depth of folding funnel

Now, finding the depth of the protein's landscape funnel is a simple procedure. We just give incremental free-energy perturbations to the native state (n) by applying denaturants or temperature and measure the connectedness of n with rest of the configurational space u as a function of this perturbation by some means. Here we have used the sample-size autocorrelation method as described in Materials and Methods. From our prediction, the depth of the folding funnel is simply the amount of free-energy perturbation at which connectedness attains maximum. Because the free-energy perturbation can be measured in terms of the amount of perturbing agents such as urea, we express the connectedness in terms of urea concentration, where the maximum connectedness occurs at the concentration of $D_{\max}$ (moles per liter). In the next section, we generalize the concept to thermal free-energy perturbations.

Existence of enthalpy and entropy convergence temperatures

The aforementioned model can be easily extended to temperature-mediated unfolding too. The potential of mean force of the native state can be expressed as a function of temperature using the following expression:

$$\begin{aligned}\tilde{f}_T(n, T) = \Delta G^{H_2O}_{0,T} = \Delta H_n + \Delta C_p^{NU} \times (T - T_n) \\ - T \times \Delta S_n - T \times \Delta S_n \\ - T \times \Delta C_p^{NU} \times \ln\left(\frac{T}{T_n}\right)\end{aligned} \quad (9)$$

Here the subscript n denotes the corresponding parameter values at $T = T_{\rm n}$, which is the temperature ($T = T_{\rm n}$) at which the protein exists completely in native form, ΔS, ΔH denotes the entropic and enthalpic changes caused by the change in temperature, and ΔC is the corresponding change in heat capacity of the protein during unfolding. We should note that $\tilde{f}_T(n, T_{\rm n}) = \Delta H_{\rm n} - T_{\rm n}\,\Delta S_{\rm n}$ and the corresponding probability of observing the native conformation (n) at temperature T is $P(n, T) = N \times e^{-\tilde{f}_T}(n, T)/RT$ (from equation 3), where N is the usual normalization constant. Using the initial condition, N can be shown to be $N = e^{\tilde{f}_T}(n, T_{\rm n})/RT$. And thus the connectedness of the native configuration n with the rest of the configurational space u at a temperature T can be given as:

$$P(n \cap u,T) = e^{-\tilde{g}_T(n,T)/RT} \times (1 - e^{-\tilde{g}_T(n,T)/RT}) \quad (10)$$

where,

$$\begin{aligned}\tilde{g}(n,T) = \tilde{f}_T(n,T) - \tilde{f}_T(n,T_n) = (\Delta C_p^{NU} - \Delta S_a) \times (T - T_n) \\ - T \times \Delta C_p^{NU} \times \ln(T/T_n).\end{aligned}$$

Now it is very easy to show [by solving the equation $\partial_T P(n \cap u, T) = 0$ using MAPLE 7] that the function has two maxima at:

$$T_{m,1} = (1 - \Delta S_n/\Delta C_p^{NU}) \times T_n \le T_n \quad (11)$$

$$T_{m,2} = T_n \times e^{\delta + (1 - 1/T_n) - (R/\Delta C_p^{NU}) \times \ln 2} \ge T_n \quad (12)$$

Here

$$\delta = \text{LambertW}\{1/T_n - 1) \times e^{(1/T_{n-1}) + (R/\Delta C_p^{NU}) \times \ln 2}\},$$

where $y = \text{LambertW}(x)$ is the solution of the equation $y \exp(y) = x$. We are generally interested in the maximum $T_{\rm n} > T$ (here it is $T_{\rm m,\ 2}$), whose approximate value can be calculated as follows: Given that $(1/T_{\rm n}) \gg 1$ ($T_{\rm n}$ is generally close to room temperature, i.e., ~298 K) and $R/\Delta C_p^{NU} \ll 1$, by neglecting those terms in the expression of $T_{\rm m,\ 2}$, we get:

$$T_{m,2} \cong T_n \times e^{[LambertW\{-e^{-1}\}+1]} \cong 1.11 \times T_n \quad (13)$$

The numerical value of LambertW($-1/e$) can be obtained from MAPLE 7. One also should note that $P(n \cap u, T)$ is a turnover function only when $\Delta C_p^{NU} < 0$. Although the expressions for change in enthalpy and entropy due to temperature have the forms as:

$$\Delta H_T^{NU} = \int_0^T \Delta C_p^{NU}\, dT \text{ and } \Delta S_T^{NU} = \int_0^T \Delta C_p^{NU} d \ln T,$$

where

$$\Delta C_p^{NU} = C_p^U - C_p^N \text{ is the}$$

is the heat capacity change due to unfolding, detailed calorimetric studies showed that it was not valid in the case of protein unfolding because of the fact that ΔC_p^{NU} itself was a function of temperature. Moreover, upon unfolding, the protein generally would not be only in native and unfolded forms but exist in a continuum of states. Because n and u are two different sets of configurations of the same molecule, earlier studies (Zhou et al. 1999) showed that it was necessary to introduce a measure called the weighting factor (i.e., $\Delta C_p^{NU} = f_{\rm U} C_p^U - {\rm f_N} C_p^N < 0$, where $f_{\rm N}$ and $f_{\rm U}$ are the corresponding fractions of native and unfolded forms) as a correction. In our model, we propose that the correct weighting factor should be the connectedness of configurational domains [i.e., $P(n \cap u, T)$], because the change in

heat capacity upon unfolding is directly proportional to the connectedness of configurational domains. Moreover, weighting by fraction of folded/unfolded population does not carry any meaning because there is a possibility of occurrence of protein molecules with partially folded configuration, which cannot be accounted for either in the folded fraction or in the unfolded fraction. In other words, it is not valid in the vicinity of the stochastic-separatrix. Because the connectedness has a turnover behavior, it is easy to conclude that the function $\Delta C_p^{NU} \times P(n \cap u, T)$ also should be a turnover function, which is the usual observation in calorimetric studies on protein unfolding.

Explaining the existence of the enthalpy and entropy convergence temperatures (i.e., the temperature, $T^* = 110°C$, at which common enthalpy, ΔH^*, and entropy values, ΔS^*, occur) is still under debate, although many possible explanations based on balance between hydrophobic and hydrophilic interaction of water with proteins have been proposed (Privalov 1979, 1996, 1997; Baldwin 1986; Lee 1991; Baldwin and Muller 1992; Fu and Freire 1992; Ragone and Colonna 1994) so far. Using our model, we were able to predict such convergence temperatures under certain conditions as follows: Using our weighted expression for the change in heat capacity, the possible correct equation for enthalpy and entropy change with temperature can be given as:

$$\Delta H_T^{NU} = \int_{T_N}^{T} \Delta C_p^{NU} \times P_{st}(n \cap u, T)_T dT \qquad (14)$$

$$\Delta S_T^{NU} = \int_{T_N}^{T} \Delta C_P^{NU} \times P_{st}(n \cap u, T)_T d \ln T \qquad (15)$$

Because our studies showed that $T_{m,\,2}$ [this is the point at which $P(n \cap u, T)$ becomes maximum] depends only on T_n, for a set of proteins with similar ΔC_p^{NU}, it is easy to verify that although there is a small difference in the T_n values of the proteins, there exists a convergence temperature T^* at which common ΔH^* and ΔS^* values can occur beyond the temperature $T_{m,\,2}$, which is the usual observation. This is because the tail region ($T > T_{m,\,2}$) of the function $P(n \cap u, T)$ contributes much less to the integrals given by equations 11 and 12. In the following section, we see how can one measure the depth of the folding funnel of a model protein, in terms of denaturant concentration experimentally.

Materials and methods

A simple method to estimate the connectedness parameter in the case of macromolecular dynamics has already been reported (Murugan 2002; Appendix). To apply our model, we chose a protein called Cytochrome P450cam, which is a 46-kD protein, obtained from a soil bacterium *Pseudomonas putida*, where "cam" stands for its substrate 1R-camphor. DEAE Sepharose, Q Sepharose and Sephadex G-10 column, and urea were purchased from Roche chemicals. Camphor was purchased from Sigma, and all other mentioned chemicals were of analytical grade. *P. putida* cytochrome P450cam (P450cam) was overexpressed in *Escherichia coli* and purified using a reported protocol (Unger et al. 1986). The concentration of the enzyme was determined using heme absorbance at 392 nm for camphor-bound P450cam ($\varepsilon_{392} = 102\ mM^{-1}\ cm^{-1}$; Gunsalus and Wagner 1978). A protein concentration of ~3 μM was used. Experiments were conducted at room temperature (298 K). The unfolding kinetic experiments were done using Hi-Tech SF61MX stopped flow spectrometer. Here a fixed N of 2000 data points (i.e., absorbance at the Soret peak of 392 nm) and a total time ($T = N\,\Delta t$) of data collection of 200 msec (i.e., $\Delta t = 100$ μsec) were used. From these collected data, the corresponding sample-size autocorrelation functions were constructed using equation A6 of the Appendix, and the autocorrelation functions obtained accordingly were fitted to equation A5 to obtain the corresponding λ' values.

Results and Discussion

A typical stopped-flow trace of unfolding of P450cam by 3 M urea is shown in Figure 2A, which clearly indicates that there was a fast process with a lifetime of ~50 msec; the corresponding sample-size autocorrelation function constructed from equation A6 is shown in Figure 2B. Figure 3 shows the variation of λ' obtained by fitting the sample-size autocorrelation function constructed accordingly to equation A5 with different concentrations of urea, which clearly indicated the turnover behavior as given in equation 7, with a maximum at ~2 M urea. Above 2 M urea, the connectedness parameter started to decline, indicating that protein molecules have come out of the folding funnel (i.e., completely to the u state) and thus can explore the rest of the nonnative configuration space. In other words, below 2 M urea concentration [$P_{st}(n) > 0.5$], the native character dominates, whereas above 2 M urea [$P_{st}(n) < 0.5$], nonnative character dominates. Therefore, we can conclude that the free-energy value corresponding to 2 M urea is the depth of the folding funnel of P450cam. Here one should note that the usual kinetic or thermodynamic analysis of unfolding/refolding in presence of denaturants will only give information about the midpoint denaturant concentration (D_{mid}), whereas from our analysis it is possible to obtain the depth of the folding funnel itself.

Acknowledgments

This work was supported by TIFR, Mumbai. We thank the Referees for their constructive comments and suggestions, which improved this manuscript a lot. R.M. acknowledges the support from the TIFR endowment fund for career development.

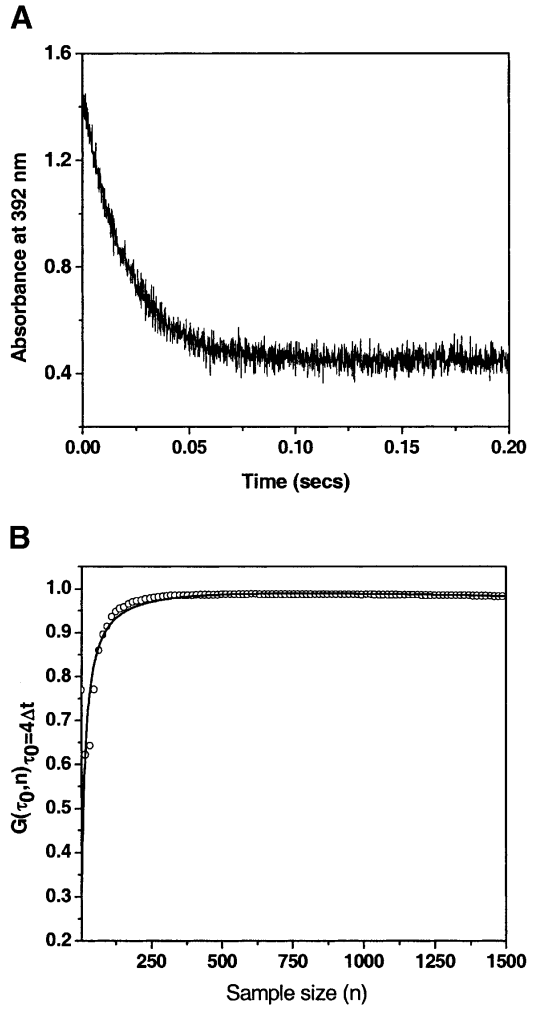

Figure 2. (*A*) Stopped-flow kinetic trace of unfolding of cytochrome P450cam (3 μM of protein) by 3 M urea monitored at 392 nm (Soret peak) for 200 msec with $\Delta t = 100$ μsec (i.e., the total number of data points, $N = 2000$). (*B*) The corresponding sample-size autocorrelation function constructed (open circles) from equation A6 with a delay of $4\Delta t = 400$ μsec. The solid line shows the fitted curve ($\chi^2 = 0.98$) of equation A5 with $\lambda' = 0.03$.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Appendix

An experimental method to estimate the connectedness using sample-size autocorrelation analysis of kinetic data has been reported (Murugan 2002). The principle is as follows: The local dynamics of the protein molecule on configurational space can be well modeled by a birth–death master equation. Let there be an M number of molecular conformations initially at a given energy level and the transition probability from the $(x-1)$-th conformation to the x-th conformation in an infinitesimal time ΔT is λ, which is the connectedness parameter. Assuming equal initial probability, the birth–death master equation becomes (because the configurational space is extraordinarily large, we can neglect the reverse probability terms):

$$\partial_t P(x,t) = \lambda P(x-1,t) - \lambda P(x,t) \tag{A1}$$

Equation A1 can be simply solved by using the generating function

$$G(s,t) = \sum_{x=0}^{M} s^x P(x,t)$$

with initial condition

$$G(s,0) = (1/M) \times \sum_{x=0}^{M} s^x$$

which is due to the fact that $P(x,0) = 1/M$ to give:

$$P(x,t) = \frac{e^{-\lambda t}}{M}\left(\sum_{i=0}^{x}\frac{[\lambda t]^i}{i!}\right) \tag{A2}$$

Now the variance of x can be given as:

$$\begin{aligned} Var\{x(t)\} &= \langle x(t),x(t)\rangle = \lim_{s\to 1}\partial_s{}^2 G(s,t) + \langle x(t)\rangle - \langle x(t)\rangle^2 \\ &= \left(1\frac{1}{M}\sum_{j=0}^{M} j^2 - \frac{(M+1)^2}{4}\right) + \lambda t = \delta + \lambda t \qquad \text{(A3)} \\ \text{Where } \delta &= \left(\frac{1}{M}\sum_{j=0}^{M} j^2 - \frac{(M+1)^2}{4}\right) = \frac{M^2-1}{12} \end{aligned}$$

Any kinetic data involving biomacromolecules like protein and DNA are the sum [$S(t)$, where t is time] of reactive [e.g., folding–unfolding transitions that involve changes in energy state, $f(t)$], nonreactive [local dynamics, $h(t)$], and instrumental noise components [$e(t)$], that is, $S(t) = f(t) + h(t) + e(t)$, whose sample-size autocorrelation with a delay of τ sec can be easily obtained by assuming $e(t)$ as an additive Gaussian noise with $\langle e\rangle = 0$, $\sigma_e{}^2 \approx 1/n$ for $n \geq 30$, where n is the sample size, $\langle e_i, e_{i+\tau}\rangle = 0$, $\langle e, f\rangle = 0$, and $\langle e, h\rangle = \langle h, f\rangle = 0$, as follows:

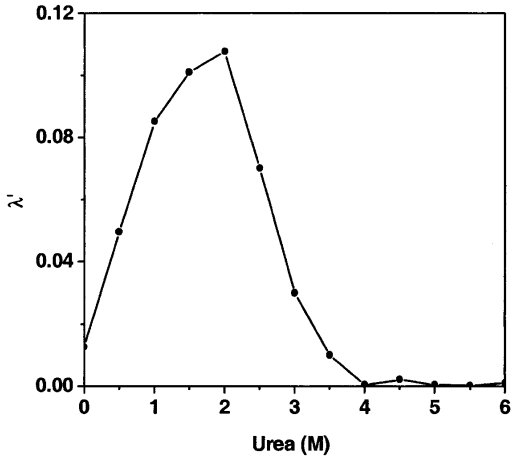

Figure 3. Variation of $\lambda' = P_{st}(n \cap u, D)$ with concentration of urea (D moles/L), showing a maximum at 2 M urea (D_{max}) and almost zero after 3.5 M urea, indicating that protein molecules have come out of the landscape funnel and thus can explore a large area of landscape.

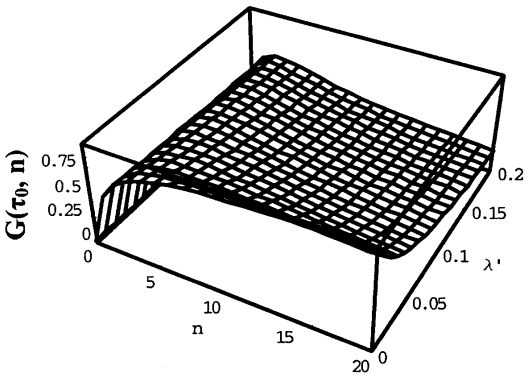

Figure 4. Variation of sample-size autocorrelation function $G(\tau_0, n)$ with λ', which is the measure of connectedness of configurational domains. The model function shown here is $G(\tau_0, n) = n/(1 + n + \lambda' n^2)$, where n is the sample size and τ_0 ($= 1$ here) is the autocorrelation delay (in seconds).

$$G(\tau,n) = \frac{\langle f_i f_{i+\tau}\rangle}{\left[\left(\langle f_i,f_i\rangle + \langle h_i,h_i\rangle + \frac{1}{n}\right) + \left(\langle f_{i+\tau},f_{i+\tau}\rangle + \langle h_{i+\tau},h_{i+\tau}\rangle + \frac{1}{n}\right)\right]^{0.5}} \tag{A4}$$

Given that $\langle h_i, h_{i+\tau}\rangle_{\tau \neq 0} = 0$, $\langle h_i, h_i\rangle = (\varepsilon/A) \times \mathrm{Var}\{x(t)\} = (\varepsilon/A) \times (\delta + \lambda t) = \delta' + \lambda' t$, where ε is the corresponding spectroscopic conversion factor (e.g., molecular extinction coefficient) and A is Avagadro's number,

$$G(\tau_0,n) = \frac{\langle f,f\rangle_{r\to 0}}{\langle f,f\rangle_{\tau\to 0} + \delta' + (\lambda'\Delta t)n + \frac{1}{n}} = \frac{\langle f,f\rangle_{\tau\to 0}}{\langle f,f\rangle_{\tau\to 0} + \delta' + \varsigma n + \frac{1}{n}}. \tag{A5}$$

where Δt is the time difference between two consecutive data points, which is constant, $t = (\Delta t)n$ and $\zeta = \lambda' \Delta t$. From equation A5, we can easily conclude that as λ' increases from zero, $G(\tau_0, n)$ exhibits the turnover behavior (Fig. 4). The experimental $G(\tau_0, n)$ can be constructed from the kinetic data (in our case, unfolding of cytochrome P450cam by urea) using the following relation:

$$G(\tau_o,n) = \frac{\sum_{i=0}^{n} g_i g_{i+\tau_0} - \left(\frac{\sum_{i=0}^{n} g_i \sum_{i=0}^{n} g_{i+\tau_0}}{n-\tau_0}\right)}{\sigma_{gi}\sigma_{g_{i+\tau_0}}} \tag{A6}$$

Where,

$$\sigma_{g_{i+t_0}}{}^2 = \sum_{i=0}^{n} g_{i+\tau_0}{}^2 - \frac{\left(\sum_{i=0}^{n} g_{i+\tau_0}\right)^2}{n-\tau_0} \quad \sigma_{gi}{}^2 = \sum_{i=0}^{n} g_i{}^2 - \frac{\left(\sum_{i=0}^{n} g_i\right)^2}{n-\tau_0}$$

Here g_i is the i-th data point from the experiment and $\tau_0 < n < (N - \tau_0)$, where N is the total number of data points collected. Now, by fitting equation A5 to the function constructed from equation A6 by a standard nonlinear method, we can easily obtain the parameter λ'.

References

Baldwin, R.L. 1986. Temperature dependence of the hydrophobic interactions in protein folding. *Proc. Natl. Acad. Sci.* **83:** 8069–8072.

Baldwin, R.L. and Muller, N. 1992. Relation between the convergence temperatures T*h and T*s in protein unfolding. *Proc. Natl. Acad. Sci.* **89:** 7110–7113.

Boczko, E.M. and Brookes, C.L. 1995. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269:** 393.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **21:** 167.

Dill, K.A., Fiebig, K.M., and Chan, H.S. 1993. Cooperativity in protein-folding kinetics. *Proc. Natl. Acad. Sci.* **90:** 1942–1946.

Fu, L. and Freire, E. 1992. On the origin of the enthalpy and entropy convergence temperatures in protein folding. *Proc. Natl. Acad. Sci.* **89:** 9335–9338.

Gunsalus, I.C. and Wagner, G.C. 1978. Bacterial P-450cam methylene monooxygenase components: Cytochrome m, putidaredoxin, and putidaredoxin reductase. *Methods Enzymol.* **52:** 166–188.

Lee, B. 1991. Isoenthalpic and isoentropic temperatures and the thermodynamics of protein denaturation. *Proc. Natl. Acad. Sci.* **88:** 5154–5158.

Leopold, P.E., Montalk, M., and Onuchic, J.N. 1992. Protein folding funnels: A kinetic approach to the sequence–structure relationship. *Proc. Natl. Acad. Sci.* **89:** 8721.

Murugan, R. 2002. Sample size autocorrelation analysis of kinetic data: Resolving reaction path heterogeneity. *J. Chem. Phys.* **117:** 6372–6376.

———. 2003. Competitive model on denaturant mediated protein unfolding. *Biophys. J.* **84:** 770–774.

Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z., and Socci, N.D. 1995. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci.* **92:** 3626.

Privalov, P.L. 1979. Stability of proteins: Small globular proteins. *Adv. Protein Chem.* **33:** 167–241.

———. 1996. Intermediate states in protein folding. *J. Mol. Biol.* **258:** 707–725.

———. 1997. Thermodynamics of protein folding. *J. Chem. Thermodynamics* **29:** 447–474.

Ragone, R. and Colonna, G. 1994. Enthalpy–entropy balance and convergence. *J. Biol. Chem.* **269:** 4047–4049.

Unger, B.P., Gunsalus, I.C., and Sligar, S.G. 1986. Nucleotide sequence of the *Pseudomonas putida* cytochrome P-450cam gene and its expression in *Escherichia coli*. *J. Biol. Chem.* **261:** 1158–1163.

Zhou, Y., Hall, C.K., and Karplus, M. 1999. The calorimetric criterion for a two-state process revisited. *Protein Sci.* **8:** 1064–1074.

Zwanzig, R., Szabo, A., and Bagchi, B. 1992. Levinthal's Paradox. *Proc. Natl. Acad. Sci.* **89:** 20–22.