PROTEINS: Structure, Function, and Genetics 29:282–291 (1997)

RESEARCH ARTICLES

Local Interactions and the Optimization of Protein Folding

Ross Doyle,[1] **Kim Simons,**[2] **Hong Qian,**[3] **and David Baker**[2]*
[1]*Department of Physics, University of Washington, Seattle, Washington*
[2]*Department of Biochemistry, University of Washington, Seattle, Washington*
[3]*Department of Biomathematics, UCLA School of Medicine, Los Angeles, California*

ABSTRACT **The role of local interactions in protein folding has recently been the subject of some controversy. Here we investigate an extension of Zwanzig's simple and general model of folding in which local and nonlocal interactions are represented by functions of single and multiple conformational degrees of freedom, respectively. The kinetics and thermodynamics of folding are studied for a series of energy functions in which the energy of the native structure is fixed, but the relative contributions of local and nonlocal interactions to this energy are varied over a broad range. For funnel shaped energy landscapes, we find that 1) the rate of folding increases, but the stability of the folded state decreases, as the contribution of local interactions to the energy of the native structure increases, and 2) the amount of native structure in the unfolded state and the transition state vary considerably with the local interaction strength. Simple exponential kinetics and a well-defined free energy barrier separating folded and unfolded states are observed when nonlocal interactions make an appreciable contribution to the energy of the native structure; in such cases a transition state theory type approximation yields reasonably accurate estimates of the folding rate. Bumps in the folding funnel near the native state, which could result from desolvation effects, side chain freezing, or the breaking of nonnative contacts, significantly alter the dependence of the folding rate on the local interaction strength: the rate of folding decreases when the local interaction strength is increased beyond a certain point. A survey of the distribution of strong contacts in the protein structure database suggests that evolutionary optimization has involved both kinetics and thermodynamics: strong contacts are enriched at both very short and very long sequence separations. Proteins 29:282–291, 1997.**

Key words: protein folding kinetics; folding energy landscapes; transition state approximation

INTRODUCTION

The role of local interactions in protein folding has been the subject of considerable debate over the last year. A number of recent theoretical studies have argued that folding is optimized when local interactions make a relatively small contribution to the energy of the native state. Fersht[1] argued from simple transition state theory that folding rates are maximized when few interactions are made before the rate-limiting transition state. The argument is that even native local interactions lower the free energy of the unfolded state and thus increase the size of the barrier that must be climbed to reach the transition state relative to what it would be if all interactions were formed at the transition state. Govindarajan and Goldstein[2] argued that "optimal conditions for folding" are obtained when local interactions are weak by using a simple lattice model in conjunction with techniques drawn from spin glass theory: increasing the strength of local interactions was found to increase the roughness of the energy landscape. Abkevich et al.[3] found that structures with substantial numbers of nonlocal interactions folded significantly faster than structures with primarily local interactions at temperatures in which the native states were stable. In contrast to these conclusions, Unger and Moult[4] found that the greater the strength of local interactions, the more foldable a sequence in a simple lattice model.

To investigate the rather discordant conclusions reached in these studies, we studied an extension of

---

Contract grant sponsor: NIH; Contract grant number: GM07270; Contract grant sponsors: NSF and Packard Foundation.

*Correspondence to: David Baker, Department of Biochemistry, University of Washington, Seattle, WA 98195.
E-mail: baker@ben.bchem.washington.edu

Received 6 January 1997; Accepted 21 April 1997

Zwanzig's simple model of protein folding.[5,6] The model captures the general features of protein folding landscapes, notably the trade-off between loss of configurational entropy and the gain of favorable interactions during folding, and importantly allows exact calculation of folding rates. Keeping the total energy of interactions in the native structure fixed, we examine the dependence of folding kinetics and thermodynamics on the balance between local and nonlocal interactions.

Zwanzig's Model

We first briefly summarize the simple model of folding studied by Zwanzig.[5] A generalized protein is considered to have a large number N of independent degrees of freedom, which are perhaps most simply viewed as backbone torsion angles. Each angle has $v + 1$ allowed states, so there are a total of $(v + 1)^N$ conformations. For each angle, the state that is found in the native conformation is called "correct." Folding kinetics are modeled by allowing any individual angle to change from correct to incorrect or from incorrect to correct. In a "golf course" energy landscape, in which the energy of the native conformation is large and negative and all nonnative conformations have the same (zero) energy, all transitions between nonnative conformations occur at the same rate, and folding is extremely slow because a near exhaustive search is required to locate the native conformation (Levinthal's paradox).

Zwanzig showed that folding can proceed rapidly if the correct state for each angle has slightly lower energy than the incorrect states. A simple reaction coordinate (S), the number of incorrectly oriented angles, emerges if this energy bias is the same for all angles because all protein conformations with the same number of angles correctly oriented have the same energy and can thus be grouped together. To make the native state stable despite the large entropic cost of folding (there are $\sim[(v + 1)^N - 1]$ nonnative conformations), Zwanzig added an extra energy bonus when all angles are in the correct configuration ($S = 0$). Zwanzig obtained analytic solutions for the kinetics of folding and showed that the system undergoes highly cooperative folding-unfolding transitions.[5]

Extension of Model

We consider the contributions of local and nonlocal interactions to the kinetics and thermodynamics of folding by using an extension of Zwanzig's model. To illustrate our use of the terms "local" and "nonlocal" in the context of the model, we begin with a simple example. Let $E(x_1,x_2,x_3, \ldots, x_N)$ be the energy function describing a physical system with N independent degrees of freedom where the contribution of each of the x_i to the total energy is independent of the state of the other variables: $E(x_1,x_2,x_3, \ldots, x_N) = e1(x_1) + e2(x_2) + e3(x_3) + \ldots + eN(x_N)$. In such a system, a change in one of the variables, say x_1, has no effect on the values of any of the other x_i. Because changes in one part of the system have no effect on any of the other parts of the system, we call such interactions "local." If, on the other hand, the contributions from the different variables to the energy are highly coupled, changes in one part of the system will generally affect the other parts considerably, and we call such interactions "nonlocal." *Local* and *nonlocal* thus refer to interactions that are functions of single and large numbers of degrees of freedom, respectively. This corresponds to the sense in which the terms are usually used in the context of protein structures: interactions between residues close in the linear sequence of a protein depend only on the values of the small number of torsion angles between them, whereas bringing together residues distant along the sequence requires the proper orienting of many torsion angles.

For simplicity, we continue with the assumption that the energy of a configuration depends only on the number of incorrectly oriented angles. The energy can thus be written as a function of S:

$$E(S) = \epsilon_L f_L(S) + \epsilon_{NL} f_{NL}(S) \tag{1}$$

where ϵ_L and ϵ_{NL} determine the overall contributions of local and nonlocal interactions to the total energy, and $f_L(S)$ and $f_{NL}(S)$ describe the dependencies of the local and nonlocal interactions on the number of incorrectly oriented angles. We identify the energy biases toward the correct orientation of each angle in Zwanzig's model as local interactions; they clearly have the important features of being independent and additive. With this choice, $f_L(S)$ is proportional to the number of incorrectly oriented angles and hence is a linear function of S. To explore the effects of cooperative interactions that occur prior to the formation of the native structure, we take $f_{NL}(S)$ to be a smoothly decreasing function of the amount of native structure (the number of correctly oriented angles) rather than the abrupt drop in the energy of the native state considered by Zwanzig. Because of the coupling between the different degrees of freedom, $f_{NL}(S)$ is a strongly nonlinear function of S; physical intuition suggests that nonlocal interactions should play an increasingly important role as the amount of native structure increases and more and more tertiary interactions and contacts are formed and thus that $d^2f_{NL}(S)/dS^2 < 0$. The specific forms for $f_L(S)$ and $f_{NL}(S)$ used in this paper are

$$f_L(S) = S$$

$$f_{NL}(S) = \frac{\{1 - \exp(-\alpha S)\}}{\{1 - \exp(-\alpha N)\}} N. \tag{2}$$

Both functions decrease monotonically from N (at $S = N$, the fully unfolded structure) to zero (the native structure); the important features are the linearity of $f_L(S)$ and the rapid and nonlinear decrease in $f_{NL}(S)$ as S approaches 0. With this choice for $f_L(S)$, the local interaction strength ϵ_L is just the local energy bias toward the correct orientation of a single torsion angle. The parameter α determines the cooperativity of the nonlocal interactions; the problem studied by Zwanzig corresponds to $\alpha \gg 1$.[5]

To explore the contributions of local and nonlocal interactions to folding kinetics, we consider the problem of folding on an "energy budget": the total energy of the native structure is held constant, but the relative contributions of local versus nonlocal interactions to this energy are varied over a broad range. The justification for the "energy budget" concept is the finite number of energetically favorable interactions possible in the native state of a protein. For example, because there are only a finite number of contacts possible in a compact polymer, an increase in the number of local hydrophobic contacts (between residues close in the sequence) necessarily requires a decrease in the number of nonlocal hydrophobic contacts. With the imposition of the energy budget, equation (1) becomes

$$E(S) = \epsilon_L f_L(S) + (\epsilon_T - \epsilon_L) f_{NL}(S) \qquad (3)$$

where ϵ_T is a constant and ϵ_L ranges from zero to ϵ_T.

MATERIALS AND METHODS

Free energy function

The entropy for a given S (the number of incorrectly oriented angles) is

$$\sigma(S) = k_B \ln \{N^S \cdot N!/(S!(N - S)!)\} \qquad (4)$$

where k_B is Boltzmann's constant because there are N^s different ways of orienting the S incorrect angles, and $N!/(N - S)!S!$ ways of choosing the S incorrect angles from the N total angles.[5] The free energy $F(S)$ is [combining Eqs. (2) and (3)]

$$F(S) = \epsilon_L S + (\epsilon_T - \epsilon_L) \frac{\{1 - \exp(-\alpha S)\}}{\{1 - \exp(-\alpha N)\}} N - T\sigma(S). \qquad (5)$$

$F(N) - F(0)$ is independent of ϵ_L in keeping with the energy budget concept. For computational convenience v and N were chosen to be 5 and 20, respectively; results were qualitatively the same for larger N and energy functions with similar curvatures. The maximal energy bias toward the correct orientation of a single angle, ϵ_T, was set to $2.36 k_B T$ so that $F(20) - F(0) = 15 k_B T$. All energies and free energies are in units of k_B T.

Calculation of folding rates

Kinetic analysis requires assumptions about how the different states are connected. In Zwanzig's model, each individual angle can independently change from correct to incorrect or from incorrect to correct. The rate of transitions from incorrect to correct is taken to be a constant (k_1). The net rate of transitions from S to S-1 is $k_1 \cdot S$ because any of the S nonnative angles can become properly oriented. To satisfy detailed balance, the rate of transitions from S to S +1 is then $(N - S)\, k_1 \exp -\beta(U(S + 1) - U(S))$.

The energy function that Zwanzig used was simple enough to permit analytical calculation of folding rates. We have not been able to obtain analytical solutions for the class of functions considered in this paper. Instead, the rate of folding was calculated for each of the curves by using the transition matrix method (see for example Ref. 7). The transition matrix element $T_{i,j}$ describes the probability of transitions from state i to j in unit time; thus the probability of transitions from i to j in t time units is $(T^t)_{i,j}$. With transitions from a given state S to states $S - 1$ and $S + 1$ chosen according to the standard Metropilis scheme, the nonzero elements of the transition matrix are:

$$\begin{aligned} T_{S,S+1} &= (N - S)k_1 \min \cdot \{1, \exp((E(S) - E(S + 1))/k_B T)\} \\ T_{S,S-1} &= Sk_1 \min \{1, \exp((E(S) - E(S - 1))/k_B T)\} \\ T_{S,S} &= 1 - T_{S,S+1} - T_{S,S-1}. \end{aligned} \qquad (6)$$

The probability of arriving at the completely folded state ($S = 0$) starting from the completely unfolded state ($S = 20$) in time t is given by the matrix element $(T^t)_{20,0}$. In all calculations $k_1 = 0.01$ and $k_B T = 1$.

Folding rates were obtained from simple exponential fits of kinetic curves calculated by using the transition matrix method. For folding conditions in which the kinetics deviated significantly from simple exponential, the reciprocal of the time required for the probability of nonnative configurations to decay to $1/e$ of the equilibrium value is given in the figures and table in lieu of a rate.

RESULTS

The kinetics and thermodynamics of folding were characterized for a series of energy functions (Fig. 1A) in which the local interaction strength ϵ_L was varied while keeping the total energy difference between the fully folded ($S = 0$) and fully unfolded ($S = 20$) structures constant through compensating changes in the nonlocal interaction strength [Eq (3)]. As ϵ_L increases, the contribution of the nonlocal interaction term decreases, and the energy becomes an increasingly linear function of S. The number of states and hence the entropy $\sigma(S)$ decreases rapidly as S decreases [Eq. (4)]. The corresponding free energies $F(S) = U(S) - T\sigma(S)$ [Eq. (5)] are shown in

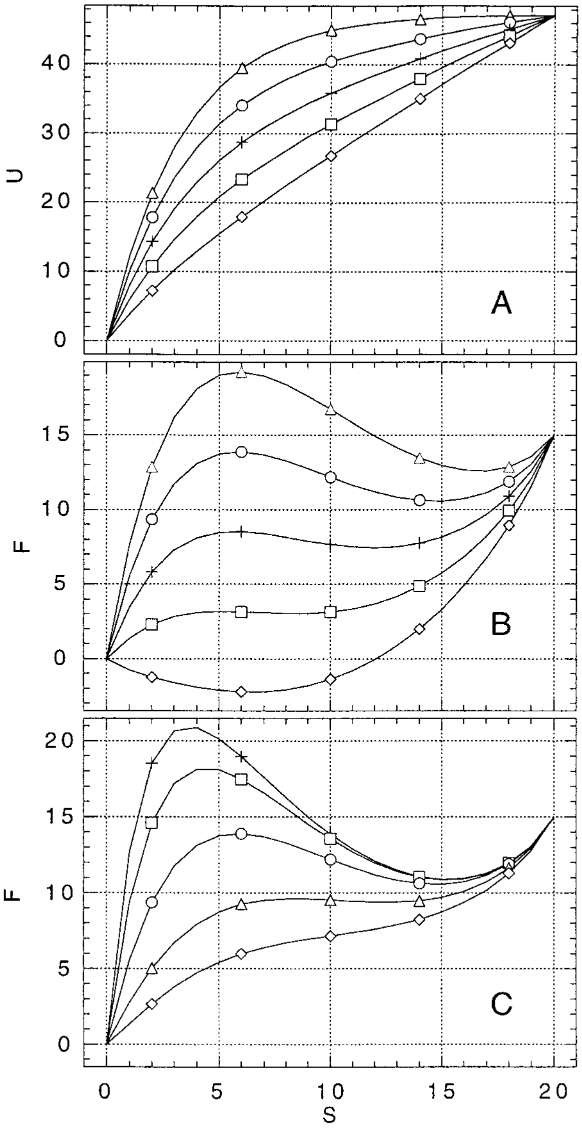

Fig. 1. Energy and free energy for different ϵ_L and α. The reaction coordinate S is the number of angles not in their native configuration. **A:** Energy and **B:** Free energy for $\alpha = 0.3$ and $\epsilon_L = 0$ (triangles), 0.5 (circles), 1 (plusses), 1.5 (squares), and 2 (diamonds). As indicated in Eq. (3), increasing ϵ_L increases the strength of local interactions and decreases the strength of nonlocal interactions. **C:** Free energy for $\epsilon_L = 0.5$ and $\alpha = 0.15$ (diamonds), 0.2 (triangles), 0.3 (circles), 0.45 (squares), and 0.6 (plusses).

Figure 1B. All energies and free energies are in units of k_BT, where $k_BT \equiv 1$ (the temperature is not varied in any of the calculations). For small ϵ_L (upper curves) there is a pronounced free energy barrier to folding at S $\sim$ 5, which decreases as ϵ_L increases (lower curves). The free energy barrier arises because the decrease in energy with decreasing S is initially not sufficient to compensate for the large decrease in entropy. The free energy minimum corresponding to the unfolded state becomes lower and moves toward the folded state (lower S) as ϵ_L increases. The effect of changing the cooperativity of acquisition of energetically favorable interactions [α in Eq. (2)] is illustrated in Figure 1C for a fixed value of ϵ_L. As α decreases, the free energy barrier to folding decreases and moves away from the folded state, whereas the free energy of the unfolded state is not substantially changed.

The rate of folding was calculated for the different values of ϵ_L and α by using the transition matrix method (see Methods). Although calculated numerically, the results are exact and do not depend on transition state theory type approximations. A pronounced lag phase and distinctly nonexponential kinetics (Fig. 2C) were obtained for energy functions with free energy barriers of less than $2k_BT$ (bottom curve in Fig. 1C). As the free energy barrier increased, the kinetic curves approached simple exponentials (Fig. 2A,B). Very low residuals from simple exponential fits were obtained when the free energy barrier was greater than $6k_BT$.

Figure 3 shows the distribution of probability density along the reaction coordinate at different times for high and low free energy barriers. For barriers of greater than $\sim 3k_BT$ (Fig. 3A), there is a rapid relaxation from the completely unfolded state (S = 20) to the equilibrium unfolded distribution (centered at $S \approx 15$), which may resemble the rapid changes in spectroscopic signals following changes in solvent conditions in experimental protein folding studies. After this initial relaxation, the kinetics closely resemble a two-state process: proteins are transferred from the unfolded population to the folded population with little change in the shape of either population. In contrast, with barriers of less than $2k_BT$, the population steadily migrates in the direction of decreasing S (Fig. 3B), producing the characteristic lag phase evident in Figure 2C.

The rate and free energy of folding for the energy functions in Figure 1A are plotted versus ϵ_L in Figure 4. The rate of folding increases with increasing ϵ_L (closed circles), whereas the stability of the native state decreases (open circles; the native state is not stable for $\epsilon_L > 1.5$). Thus, for a smooth funnel-shaped energy landscape, optimization of folding would involve a trade-off between kinetics and thermodynamics: increasing the strength of local interactions increases the rate of folding but decreases the overall stability. A similar trade-off is observed when α is varied (Table I); as the cooperativity of forming interactions increases, the stability of the native state increases and the folding rate decreases.

The preceding analysis assumes that the energy landscape is a smooth downhill funnel: the energy decreases monotonically with decreasing S in Figure 1A. However, there may well be bumps in protein folding funnels due to desolvation effects[8,9] side chain freezing,[10] and/or the breaking of nonnative interactions formed earlier in folding.[11] Such barri-

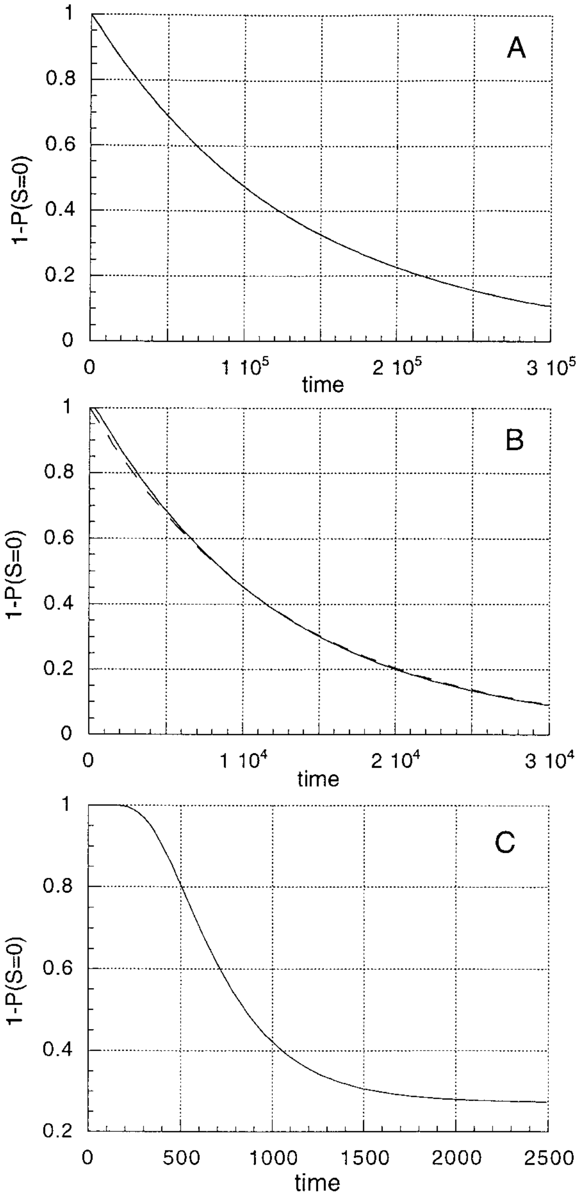

Fig. 2. Representative kinetic traces. The *y* axis is the total probability of configurations with S $>$ 0. **A:** Large free energy barriers ($\epsilon_L = 0$, $\alpha = 0.3$, Fig. 1B, triangles) give rise to simple exponential kinetics. **B:** Modest free energy barriers ($\epsilon_L = 0.5$, $\alpha = 0.3$; Fig. 1B, circles) give rise to near exponential kinetics with a small lag time. **C:** In the absence of a barrier ($\epsilon_L = 0.5$, $\alpha = 0.15$, Fig. 1C, diamonds) the kinetics are clearly nonexponential. Dashed lines in A and B are simple exponential fits.

ers may separate "molten globule" states of larger proteins from their native conformations. To investigate the effect of such bumps on the relationship between local interaction strength and folding rate, the energy for S $=$ 2 was increased by $6k_BT$ for all the curves in Figure 1A. The most dramatic effect is the reduction in the increase in folding rate with increasing local interactions (Fig. 4, closed squares); the folding rate actually decreases when ϵ_L is increased from 1.5 to 2. The presence of the barrier considerably reduces the changes in the positions of the unfolded and transition states upon changes in local interaction strength.

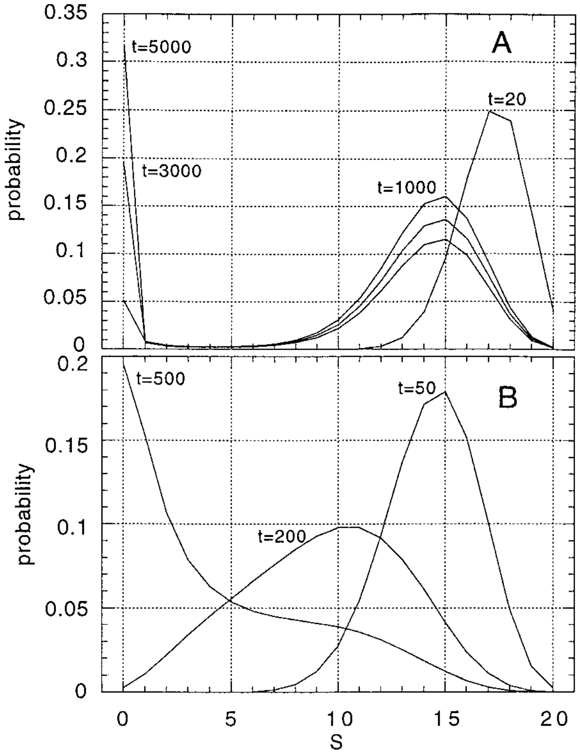

Fig. 3. Snapshots of folding. Probability distributions at early times in folding (indicated at the highest point of each distribution) for the traces shown in (**A**) Figure 2B ($\sim 3k_BT$ barrier) and (**B**) Figure 2C (no barrier).

Φ Value Distribution

To simulate the effect of point mutations on the kinetics and thermodynamics of folding, a series of "mutant" energy functions were generated from the function shown in Figure 1B, circles. Particular angles cannot be singled out in the model; instead, the "mutations" correspond to disruptions at different stages in the aquisition of native structure. A "mutation" at S′ consisted of an addition of $1.5k_BT$ to the energy for all values of S less than or equal to S′; the consequences of the mutation are spread out among all of the properly oriented degrees of freedom. The rate of folding decreased in all of the "mutants" (Fig. 5, circles), the decrease was maximal at S $\sim$10, which corresponds to the region between the unfolded minimum and the peak of the barrier for the parent energy function. Φ values were calculated from the changes in stability and folding rate for each of the mutations as described by Itzhaki and co-workers.[12] The Φ values increased from 0 to 1 as S increased from 0 to 9 (Fig. 5, triangles) and then remained close to 1 up to S $=$ 17. The observation of

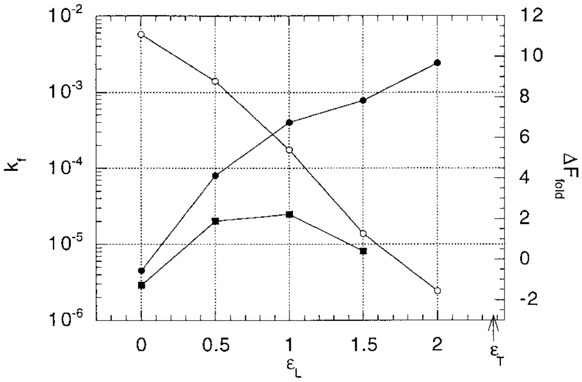

Fig. 4. Optimization of folding is a trade-off between kinetics and thermodynamics. The folding rate (closed circles) and free energy of folding (open circles) are plotted as a function of ϵ_L, with $\alpha = 0.3$ (the corresponding energy and free energy functions are shown in Fig. 1A,B). Because of the fixed energy budget, increasing ϵ_L reduces the nonlocal interaction strength [Eq. (3)]. The dependence of the folding rate on the local interaction strength changes significantly when the energy for $S = 2$ is increased by $6k_BT$ (closed squares). The free energy of folding is

$$\Delta F_{\text{fold}} = -k_BT \ln [P_{eq}(S < 4)/P_{eq}(S \geq 4)]$$

where P_{eq} is the probability of the indicated values of S at equilibrium.

TABLE I. Exact and Approximate Folding Rates for Different Local Interaction Strengths*

ϵ_L	α	Folding rate	Approximate folding rate
0	0.3	4.51E-06	2.78E-06
0.5	0.3	7.96E-05	6.22E-05
1	0.3	4.00E-04	4.76E-04
1.5	0.3	7.79E-04	1.17E-03
2	0.3	2.40E-03	—
0.5	0.6	1.33E-07	9.16E-08
0.5	0.45	1.94E-06	1.38E-06
0.5	0.3	7.96E-05	6.22E-05
0.5	0.2	7.46E-04	1.00E-03
0.5	0.15	1.27E-03	—

*The rates in column three were obtained by using the exact transition matrix method; those in column four were estimated by using the approximation:

$$k_{\text{APPROX}} = k_1 \frac{e^{-F(S_{TS})/k_BT}}{\sum_{S \geq S_{TS}} e^{-F(S)/k_BT}}$$

where S_{TS} is the value of S for which F is maximum.

Φ values intermediate between 0 and 1 for mutations in small proteins[12] is consistent with the expectation that mutations primarily affect the energies of states in which large numbers of interactions are starting to be formed (Fig. 1A; 75% of the energy of the native state is gained in going from $S = 6$ to $S = 0$).

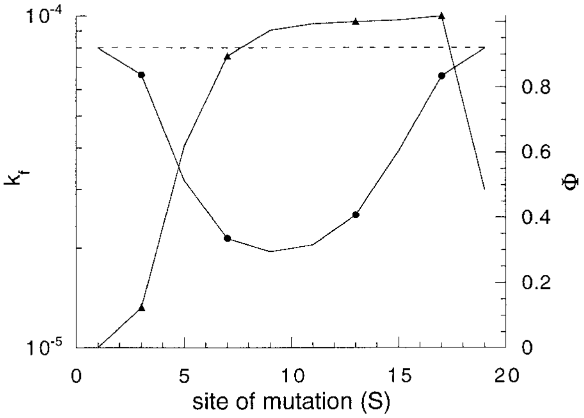

Fig. 5. Effect of mutations on Φ values and folding rates. $1.5k_BT$ was added to the free energy for $\epsilon_L = 0.5$, $\alpha = 0.3$ (Fig. 1B, circles) for all values of S less than or equal to that indicated on the horizontal axis. The Φ value is defined as

$$-k_BT \ln [k_f(\text{mutant})/k_f(\text{wt})]/\Delta F_{\text{fold}}(\text{mutant}) - \Delta F_{\text{fold}}(\text{wt})$$

according to Ref. 12. Circles, folding rate; trianges, Φ values.

Transition State Theory Approximation

Approximations based on simple transition state theory are frequently used in treatments of protein folding kinetics. Because folding rates are determined independent of such approximations in this study, there is a rare opportunity to test the validity of such approximations (a related approximation was considered by Zwanzig). The rate was estimated by using $k_f = k_1 \cdot P(\text{transition state})$, where P(transition state) is calculated assuming thermal equilibrium between the transition state (the highest free energy state) and states with larger S. As shown in Table I, the agreement is quite good for all but the smallest free energy barriers, for which the assumption of a rapidly generated Boltzmann distribution among nonnative configurations clearly breaks down.

DISCUSSION

Protein Folding Transition States

One of the striking features of experimental data on the folding of small proteins is the prevalence of simple exponential relaxation kinetics. As shown in Figure 2, simple exponential kinetics arise in the simple model when the free energy barrier to folding is many k_BT. Under such conditions, the rate of equilibration within the two free energy minima is fast compared with the rate of folding and unfolding, and thus the probability of barrier crossing is independent of time. The experimental observation of simple exponential kinetics may be related to the correlation evident in Figures 1–3 between deviations from simple exponential kinetics and energy functions which do not produce highly stabilized native states. Stabilization of the native state requires that a substantial number of energetically favorable interactions be made late in folding, and this produces an overall free energy barrier because the entropic cost of fixing degrees of freedom is initially not compensated.

The concept of a "transition state" appears in much discussion of protein folding kinetics, often borrowed directly from the theory of simple bond breaking or making reactions. Given the dramatic differences between protein folding and such reactions, it is not necessarily obvious how transition state-based pictures apply to protein folding. The success with the simple transition state approximation (Table I) shows that from the point of view of estimating the rate of folding, a "transition state" is a useful construct, but in the examples discussed here, the transition state consists of a vast ensemble of structurally unrelated states that have some particular value of S. In the presence of free energy barriers of at least several $k_B T$, the kinetics begin to approach a simple exponential, and the concept of a transition state takes on some meaning; in the absence of such barriers the concept breaks down entirely. The position of transition state and unfolded state ensembles along the reaction coordinate is strongly dependent on the shape of the energy function (Fig. 1B).

Fersht[1] argued that "A simple analysis of the optimization of the rate of protein folding predicts that rates are highest when the denatured state has little residual structure under physiological conditions." The opposite is seen in Figure 1B and Figure 4: as the strength of local interactions increases, the amount of "native-like structure" increases (the average number of the 20 degrees of freedom not correctly oriented decreases from 17 to 9; Fig. 1B) and the rate of folding increases (Fig. 4, closed circles). The explanation for the discrepancy is simple: Fersht assumed that the free energy of the transition state remained unchanged, whereas in Figure 1B it is clear that the free energy of the transition state decreases dramatically as the local interaction strength increases. The presence of a bump in the funnel near the native state greatly reduces the movement of the transition state, and the folding rate decreases with increases in the local interaction strength (Fig. 4, closed squares) as argued by Fersht.

One-dimensional free energy profiles, such as those in Figure 1B,C appear frequently in discussions of protein folding. It is important to note that for the model described here, the one-dimensional profile captures the kinetic and thermodynamic properties of the system exactly despite the large decrease in dimensionality. Folding for real chains and in the simple model involves a search for the native state in a very high dimensional space; folding dynamics in the latter case can be accurately described in one dimension because 1) all angles are treated equivalently and hence the energy depends only on the total number of correctly oriented angles and 2) only one angle is allowed to change at a time during the search, so a configuration with S angles incorrectly oriented is only kinetically connected to configurations with $S + 1$, S, or $S - 1$ angles incorrectly oriented. Both the energy function and the connectivity between states (determined by the move set in Monte Carlo studies of lattice chains) are considerably more complex for real chains, and thus one-dimensional free energy profiles can only provide an approximate description of their dynamics. A challenge for future work is to extend the simple model to energy landscapes more complex than the smooth N dimensional funnels studied here.

Folding Free Energy Landscapes

General properties of the folding free energy landscape (see ref. 11, 13) for excellent recent reviews) can be deduced from fairly elementary considerations. First, to avoid the impossible exhaustive search through conformational space, there must be some bias toward the native state in the folding energy landscape; the landscape must be to some extent funnel shaped. Second, thermodynamic stability requires that the native state be considerably lower in energy than typical unfolded configurations. Third, the energy surface is not likely to be smooth; the multitude of weak interactions possible between the large number of atoms in a polypeptide chain necessarily leads to a very large number of local minima, a subset of which will be entered and exited (with some increase in energy) during folding. This "roughness" leads to a reduction in the rate at low temperatures. The debate over the relative importance of local versus nonlocal interactions is connected to the relative importance of the three defining features of the energy landscape: increasing the local interaction strength has the rather opposite effects of 1) increasing the bias toward the native state in the folding funnel, 2) decreasing the energy of nonnative states (which can contain substantial local native interactions) and thus reducing the stability of the native state, and 3) increasing the variation in the energies of the nonnative states and thus increasing the roughness of the landscape.[11] Claims that local interactions favor folding focus on 1), whereas arguments to the contrary focus on 2) and 3).

For example, Govindarajan and Goldstein[2] focused on 2) and 3) in their argument that "optimal conditions for folding are achieved when the contributions from local interactions to the stability of the native state is small." Folding dynamics were assumed to be related to the ratio of the folding transition temperature to the glass transition temperature, which is maximized in a random energy model by maximizing $R = \Delta/\Gamma$, where Γ is the width of the distribution of energies of unfolded states, and Δ is the average energy difference between these states and the native state. Optimization of the ratio was found to involve reducing the strength of local interactions. However, the argument depends on a correspondence between thermodynamic param-

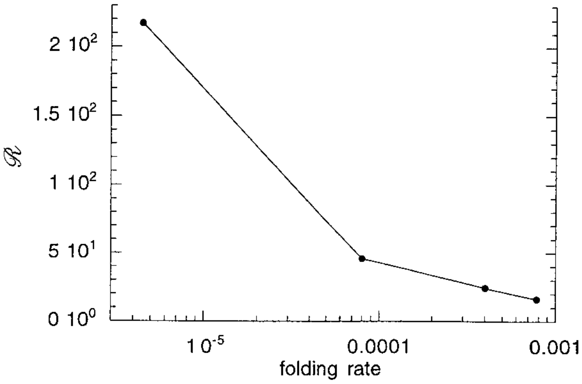

Fig. 6. The folding rate decreases with increasing R. Shown are the results for the top four energy functions in Figure 1A; R decreases as ϵ_L increases. As defined by Govindarajan and Goldstein,[2]

$$R = \frac{\Delta}{\Gamma} = \frac{\overline{E}}{\sqrt{\dfrac{\sum_{S=4}^{20} N(S)(E(S) - \overline{E})^2}{\sum_{S=4}^{20} N(S)}}}\,; \quad \overline{E} = \frac{\sum_{S=4}^{20} N(S)E(S)}{\sum_{S=4}^{20} N(S)}$$

where N(S) is the degeneracy of the state S.

eters, such as R, and dynamic properties, such as foldability, which may not always hold. Our simple model is an extreme example: R indeed increases as the strength of local interactions decreases, as in the lattice model used by Govindarajan and Goldstein,[2] but the rate of folding also decreases, and becomes extremely slow for large values of R (Fig. 6). Maximizing R, while increasing the stability of the folded state, clearly does not “optimize” the foldability in this model. As mentioned above, increasing the strength of local interactions increases both the bias toward the native state in the folding funnel and the roughness of the energy landscape; the earlier study only considered the latter effect, but an obvious limitation of the current study is that only the former is considered (the energy landscapes are smooth). A better treatment of local interactions would incorporate both the likely increase in roughness found by Govindarajan and Goldstein[2] and the effects noted in this study. Roughness can be readily incorporated into Zwanzig’s model through an activation barrier in the elementary transition rate k_1, but it is not obvious how to relate the activation barrier to the local interaction strength without introducing additional parameters (Wolynes and coworkers[13] have argued that the effective diffusion constant changes along the reaction coordinate, a further complication).

A related limitation of the current model is that there is no provision for attractive nonlocal interactions that are not present in the native structure (the “principle of minimal frustration”[13] taken to the extreme). This limitation may be considerably less severe for proteins than for heteropolymers in general; the simple two-state folding behavior and monotonic decrease in the folding rate with increasing denaturant observed for most small (<100 amino acids) proteins suggest that misfolded kinetically trapped states do not contribute appreciably to their folding dynamics.

Because of the shortcomings of the simple model, it is important that studies of more concrete lattice models support the general conclusions about the role of local interactions in folding. In one of the first lattice simulation studies of protein folding, Go and Taketami[14] found that, for a 49 residue chain on a two-dimensional square lattice, increasing the strength of local interactions relative to nonlocal interactions increased the rate of folding but decreased overall stability. In studies of a 27 residue chain on a cubic lattice, Akbevich et al.[3] found that structures with large numbers of local interactions were less stable than structures with primarily nonlocal interactions. In studies of a similar model, Unger and Moult[4] found that progressively increasing the local interaction strength increased the frequency of folding to the native conformation in long Monte Carlo simulations. Akbevich et al.[3] noted that, if the temperatures of folding simulations for different lattice proteins are chosen so that all of the proteins have the same stabilities, faster folding is observed for proteins with larger numbers of nonlocal interactions. Depending on the choice for the temperature dependence of the intrinsic rate constant k_1, this could well occur in our simple model because stability decreases dramatically with increasing local interaction strength (Fig. 4).

Local Interactions and Evolutionary Optimization

In contrast to the predictions of diffusion-collision type models of folding, increasing the helical propensity of segments of a protein that are helical in the folded state has been shown not to increase the rate of folding in several different proteins.[12,15] In a recent study of CheY, increases in helical propensities were found to actually decrease the rate of folding (Victor Munoz, personal communication). Premature formation of rigid helical structure could perhaps slow the rate of folding by interfering with desolvation, side chain interdigitation, or other slow steps associated with the docking of different portions of the chain. Increases in the local stiffness of the chain could also reduce the frequency of the chain reversals required for folding. In contrast, we found that mutations that disrupt local interactions in turns can significantly slow the rate of folding of small proteins[16]. It is possible that local interactions associated with chain reversals play a particularly important role in folding; more experiments are clearly required in this area.

The protein database may be viewed as a large evolutionary optimization experiment. If the tradeoff between kinetics and thermodynamics is resolved in biological proteins in favor of kinetics, then strong contacts should be enriched at short sequence separations, whereas the converse is expected if thermodynamics rather than kinetics has been optimized. Abkevich et al.[3] observed that a relatively large fraction of contacts in protein structures involve pairs separated by more four residues along the sequence and argued that this reflected evolutionary optimization. However, a sizable fraction of contacts are necessarily nonlocal in compact structures. To correct for the strong effects of chain connectivity and compactness on the distribution of contacts, we investigated not the distribution of total contacts, but the frequency of strong contacts (2 side chains with more than 15 pairs of adjacent carbon atoms) relative to the frequency of total contacts (2 side chains with at least 1 pair of adjacent carbon atoms). As expected, the frequency of both strong contacts and total contacts decreases monotonically with increasing sequence separation (data not shown). The ratio of the strong contact frequency to the total contact frequency is more interesting; as shown in Figure 7, there is a significant enrichment in strong contacts both at very short sequence separations (the major contributions are from i,i + 3 and i,i + 4 contacts) and at very long sequence separations. In light of the results with the simple model, it is tempting to speculate that the former reflects optimization of kinetics, and the latter, optimization of thermodynamic stability.

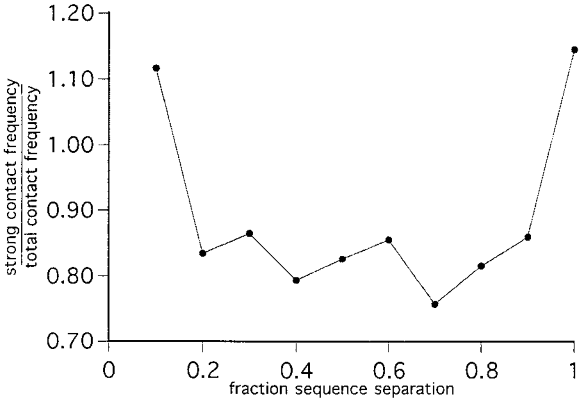

Fig. 7. Strong contacts are enriched at both very short and very long sequence separations. Two residues with at least one pair of side chain carbon atoms separated by <5 Å were considered to be in contact; two residues with >15 pairs of side chain carbons separated by <5 Å were considered to form a strong contact. The relative sequence separation of the two residues is $(n1 - n2)/N$ where $n1$ and $n2$ are the positions of the two residues in the sequence and N is the number of amino acids in the protein. Contacts in the protein structure database (August 1996 PDB-Select 25 set of proteins[17]) were collected into 10 bins; the $n/10$ bin included all contacts with relative sequence separations between $(n - 1)/10$ and $n/10$. The frequency of strong contacts and total contacts was determined for each bin: contact frequency = (number of contacts in bin)/(number of contacts in all bins). The ratio of strong contact frequency to the total contact frequency is plotted as a function of the relative sequence separation in the figure; the denominator corrects for the decrease in contact frequency with increasing sequence separation expected in a compact polymer.

Comparison with Other Models

The simple model introduced by Zwanzig and developed in this paper captures effects that are largely lost or blurred in the simple two-state (U ⇔ N) models often used to analyze experimental data: for example, the distinction between configurational entropy and the other energetic and entropic contributions to the free energy, the possibility of movement of the position of the denatured state and transition state along the reaction coordinate with changes in the energy function, and the ensemble nature of the transition and denatured states. The limitations of the two-state model greatly complicate the analysis of the effect of changes in local interaction strength on the kinetics of folding due to the neglect of the first two of these features. The current model provides a simple conceptual framework that captures many of the qualitative features of folding extracted from lattice models, such as the largely entropic nature of the barrier and the rate limiting search for a transition state with substantial native-like features, and in fact lattice simulation results are often interpreted by constructing free energy profiles similar to those in Figure 1. Although certainly more realistic than that of the simple model used here, the free energy landscapes of lattice chains are somewhat parameter dependent, and it is not always clear how closely they resemble those of proteins; in particular, the roughness of the landscape may be considerably greater than that for real proteins given the on-off character of the interactions. However, the great advantages of lattice models are obvious: they incorporate excluded volume, connectivity, and specific intrachain interactions, and the lack of treatment of these effects are limitations of the current model.

ACKNOWLEDGMENTS

We thank D. Shortle, R. Zwanzig, V. Munoz, D. Teller, H.S. Chan, and R. Baldwin for valuable comments on the manuscript and V. Munoz for communicating results prior to publication. K.S. was supported by an NIH predoctoral training grant (GM07270). This work was supported by young investigator awards to D.B. from the NSF and the Packard Foundation.

REFERENCES

1. Fersht, A.R. Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications. Proc. Natl. Acad. Sci. USA 92:10869–10873, 1995.
2. Govindarajan, S., Goldstein, R.A. Optimal local propensities for model proteins. Proteins Struct. Funct. Genet. 22:413–418, 1995.
3. Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. J. Mol. Biol. 252:460–471, 1995.
4. Unger, R., Moult, J. Local interactions dominate folding in a simple protein model. J. Mol. Biol. 259:988–994, 1996.
5. Zwanzig, R. Simple model of protein folding kinetics. Proc. Natl. Acad. Sci. USA 92:9801–9804, 1995.
6. Zwanzig, R., Szabo, A., Bagchi, B. Levinthal's paradox. Proc. Natl. Acad. Sci. USA 89:20–22, 1992.
7. Chan, H.S., Dill, K.A. Transition states and folding dynamics of proteins and heteropolymers. J. Chem. Phys. 100: 9238–9257, 1994.
8. Rank, J., Baker, D. A desolvation barrier to the association of nonpolar solutes in water may contribute to the rate limiting step in protein folding. Protein Sci. 6:347–354, 1997.
9. Waldburger, D.D., Jonsson, T., Sauer, R.T. Barriers to protein folding: Formation of buried polar interactions is a slow step in acquisition of structure. Proc. Natl. Acad. Scie. 93:2629–2634, 1996.
10. Shakhnovich, E., Finkelstein, A.V. Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. Biopolymers 28:1667–1681, 1989.
11. Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D., Chan, H.S. Principles of protein folding—A perspective from simple exact models. Protein Sci. 4:561–602, 1995.
12. Itzhaki, L.S., Otzen, D.E., Fersht, A.R. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. J. Mol. Biol. 254:260–288, 1995.
13. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins Struct. Funct. Genet. 21:167–195, 1995.
14. Go, N., Taketomi, H. Respective roles of short- and long-range interactions in protein folding. Proc. Natl. Acad. Sci. USA 75:559–563, 1978.
15. Sosnick, T.R., Jackson, S., Wilk, R.R., Englander, S.W., DeGrado, W.F. The role of helix formation in the folding of a fully α-helical coiled-coil. Proteins Struct. Funct. Genet. 24:427–432, 1996.
16. Gu, H., Kim, D.E., Baker, D. Contrasting roles for symetrically disposed β-turns in the folding of a small protein. J. Mol. Biol. (in press).
17. Hobohm, U., Scharf, M., Schneider, R., Sander, C. Selection of a representative set of structures from the Brookhaven Protein Data Bank. Protein Sci. 1:409–417, 1992.