Protein Science (1996), 5:1826–1832. Cambridge University Press. Printed in the USA.

Denaturants can accelerate folding rates in a class of globular proteins

CARLOS J. CAMACHO[1] AND D. THIRUMALAI[2]

[1] Facultad de Fisica, Pontifica Universidad Catolica de Chile, Casilla 306, Santiago 22 Chile
[2] Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742

(RECEIVED March 6, 1996; ACCEPTED June 12, 1996)

Abstract

We present a lattice Monte Carlo study to examine the effect of denaturants on the folding rates of simplified models of proteins. The two-dimensional model is made from a three-letter code mimicking the presence of hydrophobic, hydrophilic, and cysteine residues. We show that the rate of folding is maximum when the effective hydrophobic interaction ϵ_H is approximately equal to the free energy gain ϵ_S upon forming disulfide bonds. In the range $1 \leq \epsilon_H/\epsilon_S \leq 3$, multiple paths that connect several intermediates to the native state lead to fast folding. It is shown that at a fixed temperature and ϵ_S the folding rate increases as ϵ_H decreases. An approximate model is used to show that ϵ_H should decrease as a function of the concentration of denaturants such as urea or guanidine hydrochloride. Our simulation results, in conjunction with this model, are used to show that increasing the concentration of denaturants can lead to an increase in folding rates. This occurs because denaturants can destabilize the intermediates without significantly altering the energy of the native conformation. Our findings are compared with experiments on the effects of denaturants on the refolding of bovine pancreatic trypsin inhibitor and ribonuclease T1. We also argue that the phenomenon of denaturant-enhanced folding of proteins should be general.

Keywords: denaturants; energy landscape; folding rates; lattice model

It is believed that, in most instances, denaturants such as urea or guanidine hydrochloride (Gdn-Hcl) strongly decrease the rate of folding as their concentration is increased. However, there are examples in which the rate of folding is increased as the concentration of denaturants increases. One of the earliest experiments that demonstrated this effect was in the refolding of carbonic anhydrase (McCoy et al., 1980). Two more recent examples in which this has been illustrated nicely is the refolding of ribonuclease T1 (Kiefhaber et al., 1992) and the rearrangement of a kinetically trapped native-like intermediate in bovine pancreatic trypsin inhibitor (BPTI) (Weissman & Kim, 1991). We now discuss the features of these experiments as they pertain to this article.

It has been suggested that the folding kinetics of RNase T1 (Kiefhaber et al., 1992) is complex, involving parallel pathways. In the predominant pathway, the rapidly formed partially ordered structure reaches the native conformation through a series of two slow kinetic steps. Kiefhaber et al. have argued that the slowest step in this pathway, with time constant of 7,000 s at pH 5 at 10 °C, involves proline isomerization. The very small rate constant for this first-order kinetics is apparently unusually slow compared with other proteins whose rate-determining step also involves proline isomerization. These experiments seem to indicate that the incorrect isomer (the C conformation in RNase T1) is trapped in a stable native-like conformation. Consequently, the addition of urea or Gdn-Hcl in concentrations that does not denature the protein could destabilize the native-like intermediate, thus accelerating the folding rate.

Weissman and Kim (1991) have reported similar findings in the rearrangement kinetics of the rate-determining step in the refolding of BPTI. In particular, their experiments suggest that the rate-limiting step in this protein involves the rearrangement of a native-like intermediate to the native conformation. Weissman and Kim observed that, in a limited concentration regime, the rearrangement kinetics in the rate determining step is accelerated with addition of urea. Although they did not provide an explanation, they did suggest that the increase in the folding rate is due to the destabilization of the stable native-like intermediate.

These experiments strongly suggest that whenever an extremely stable native-like intermediate (containing many features in common with the native conformation) dominates the productive pathway in the folding of a protein, the rate of folding can be very slow. In fact, these native-like intermediates could effectively act as kinetic traps, preventing fast folding of proteins. The slow kinetics result because the rearrangement kinetics could be sterically hindered, and thus subsequent steps involved in the transition to the native conformation could involve overcoming large barriers. These

Reprint requests to: D. Thirumalai, Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742; e-mail: thirum@glue.umd.edu.

effects have been observed already in simulations (Camacho & Thirumalai, 1993; Abkevich et al., 1994; Socci & Onuchic, 1994, 1995; Dill et al., 1995). In this article, we propose simple theoretical arguments that confirm this picture and we relate this to the effect of denaturants on the folding kinetics. In addition, our results provide certain guidelines for obtaining estimates for the barriers involved in the slow processes. Our calculations are based on Monte Carlo (MC) calculations on minimal lattice models for proteins. For concreteness, we use the recently introduced class of models for disulfide-bonded proteins (Camacho & Thirumalai, 1995a). Hence, the results of our computations are perhaps most directly relevant for the experiments of Weissman and Kim (1991). By recognizing the generality of this phenomenon, we have provided a possible explanation of similar findings in RNase T1 (Kiefhaber et al., 1992).

Minimal lattice models and computational details

Following our recent work (Camacho & Thirumalai, 1995a), we consider minimal models of disulfide-bonded proteins so that the kinetics of approach to the native state starting from an ensemble of heterogeneous unfolded conformations can be monitored. Such models of proteins must include at least three features: (1) a flexible connectivity of the basic monomers; (2) self-avoidance between different parts of the protein chain; and (3) a well-defined native conformation. The formation of disulfide bonds are mimicked by assigning short-range pairwise interactions between certain monomers. Because we are particularly interested in understanding the folding kinetics in detail, a natural choice appears to be lattice models of proteins that seem to mimic many aspects expected in the folding of proteins (Bryngelson et al., 1995; Dill et al., 1995). These lattice models can be designed to include all three features mentioned above. Our models do not consider side chains and, consequently, the extent to which their neglect affects the overall kinetics cannot be ascertained. Despite this, based on similar studies reported elsewhere (Dill et al., 1995), we expect that qualitative aspects of the experiments can be understood using very simple models.

We model proteins as self-avoiding walks (SAW) or chains of M monomers on the square lattice. The sites in the chain can be of three types: hydrophobic (H), hydrophilic (P), or covalent (S), the latter mimicking the cysteine residues. All interactions are short range and depend exclusively on the nature of the nearest neighbors' sites surrounding each monomer. Nearest neighbors' sites satisfy the following criteria: $|i - j| \leq 3$ and $r_{ij} \equiv |r_i - \mathbf{r_j}| = a$, where $\mathbf{r_i}$, $i = 1, 2, \ldots, M$ are the coordinates of the SAW sites, and a is the lattice spacing. An attractive energy $-\epsilon_H$, with $\epsilon_H \leq 0$, is assumed whenever two H sites are nearest neighbors. Two cysteine (or S) residues that are nearest neighbors form a disulfide bond (S-S) with an associated energy gain of $-\epsilon_S$, with $\epsilon_S \leq 0$; ϵ_S can be thought of as the free energy gain in the formation of disulfide bonds under renaturing conditions. Interactions involving P sites are taken to be zero. This class of models is a generalization of the ones discussed elsewhere (Dill et al., 1995).

We study the dynamics of this model using the time dependence of an overlap function. The overlap function $\chi(\leq 1)$ measures the degree of disorder of the protein structure with respect to the ground state. The time-dependent statistical average of χ is:

$$\langle \chi(t) \rangle = 1 - \frac{1}{M^2 - 3M + 3} \left\langle \sum_{i \neq j,\, j \pm 1} \delta\ (r_{ij}(t) - r_{ij}^N) \right\rangle, \qquad (1)$$

where $\delta(0) = 1$ whenever $r_{ij}(t) = r_{ij}^N$ and 0 otherwise, and the superscript N refers to the ground state or the native state. The overlap function was introduced by us as a very general probe of the folding kinetics. In our earlier studies (Camacho & Thirumalai, 1993; Guo & Thirumalai, 1995), we showed that the analysis of the time dependence of $\langle \chi(t) \rangle$ revealed a three-stage folding kinetics for generic protein models. Furthermore, we found that the temperature dependence of the equilibrium fluctuations $\Delta\chi = \langle \chi^2 \rangle - \langle \chi \rangle^2$ could be used to determine the folding temperature T_f at which the protein undergoes a first-order transition to the native conformation. Thus, χ proves to be a useful order parameter for describing the kinetics and thermodynamics of protein folding. Physically, the function χ can also be thought of as a measure of the amount of nonnative-like structure present in the protein. Notice that $\chi \approx 0$ only when the protein is folded in its ground state conformation.

The thermodynamic properties of our model, such as T_f, $\langle \chi \rangle$, and so forth, have been obtained exactly by enumeration of the whole space of conformations. The kinetics of the model have been calculated using MC simulations: single-site Verdier's lattice dynamics (Verdier & Stockmeyer, 1967), which satisfy: (1) the connectivity constraints; (2) self-avoidance, which prevents two beads from occupying the same lattice site; and (3) Metropolis algorithm (Metropolis et al., 1953). The breakage of a disulfide bond is energetically unfavorable. Thus, MC steps (MCS) that break an S-S bond are accepted with probability $\exp(-\epsilon_S/k_B T)$, where k_B corresponds to the Boltzmann constant, which has been set to unity. Similar considerations apply to the formation and breakage of the hydrophobic contacts. In order to identify the physical events involved in the folding process, we have performed exhaustive MC simulations in the neighborhood of all low-lying minimum energy structures so that the smallest energy barriers connecting these structures to other low-lying minima can be determined. These energy barriers are used to characterize the energy landscape and hence the time scales for folding.

Rates of folding in the minimal model

We have studied quantitatively the folding kinetics for one generic sequence for the model for disulfide-bonded proteins. In our earlier study, we had argued that qualitatively similar results are obtained for other sequences as well (Camacho & Thirumalai, 1995a). We emphasize that the computational studies of the sort attempted here are at best caricatures of real proteins. Therefore, it does not seem necessary to obtain quantitative results for all possible sequences. Rather, using the exhaustive studies on a few sequences, we have attempted to provide a coherent microscopic interpretation of the effect of denaturants on refolding rates of proteins.

The native conformation for the specific sequence considered here is displayed in Figure 1 of an earlier paper (Camacho & Thirumalai, 1995a). The model protein consists of $M = 15$ sites and has two S sites at positions 4 and 15. In addition, there are nine hydrophobic sites and four hydrophilic (or neutral) ones. This model is, therefore, a minimal representation of a three-letter code, with one of them being a model for a Cys residue.

The independent energy scales for our model are $k_B T$ and ϵ_H, whereas the variables that are used in the experiments are the concentration of the denaturant and temperature. A plausible relationship between the concentration of denaturants and ϵ_H is suggested in the next section. We have calculated the folding rates as a function of ϵ_H and T. The result of the computations, which is the summary of the work for our model, is presented in Figure 1. There

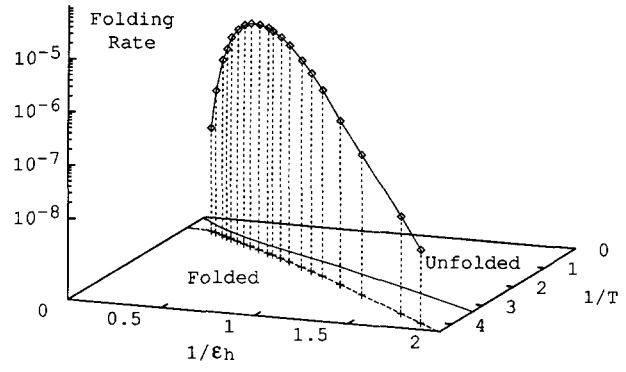

Fig. 1. Folding rate to a constant amount of native structure $\langle\chi(T, \epsilon_H)\rangle = 0.1$. In this study, the energy is measured in units of k_BT. Comparison with real proteins can be made by taking $\epsilon_H \sim$ 1–2 kcal/mol. The diamond symbols, joined by a spline fit to guide the eye, correspond to the rate measured in [1/MCS]. The binding energy of covalent bonds is always fixed to $\epsilon_S = 3$. The + symbols and dashed line represent the exact temperature T and hydrophobic binding energy, ϵ_H, for which $\langle\chi\rangle = 0.1$. The solid line on the $(1/\epsilon_H, 1/T)$ plane corresponds to the exact loci of the folded–unfolded transition, which is defined by the maximum on the fluctuations of the overlap correlation function $\Delta\chi = \langle\chi^2\rangle - \langle\chi\rangle^2$.

are several aspects of this figure that are relevant for subsequent discussions.

Equilibrium folding-unfolding transition

By examining the behavior of the chain in the plane formed by ϵ_H and T, we discover the condition for transition from folded conformation to the unfolded state. This is displayed as a two-dimensional plot in Figure 2. The solid line in this figure [also shown in Fig. 1 in the (ϵ_H, T) plane] separates the folded and unfolded states. This line was obtained as follows. For a given value of ϵ_H ($\epsilon_S = 3$ in all of our computations), the folding transition temperature is determined from the peak of the fluctuations

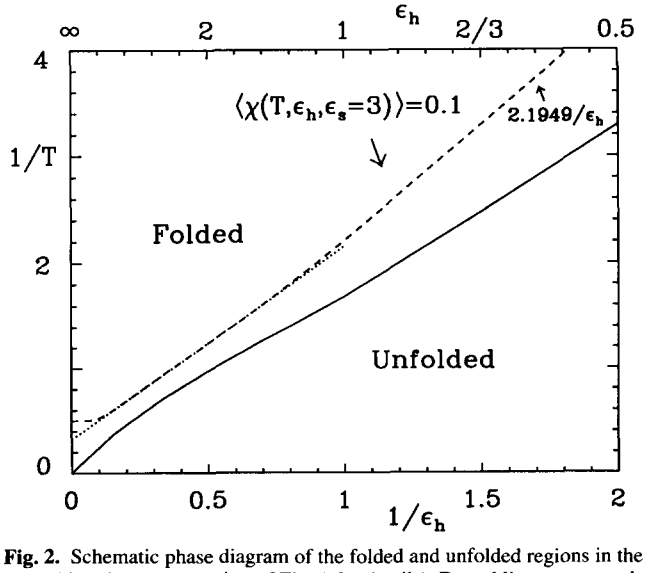

Fig. 2. Schematic phase diagram of the folded and unfolded regions in the $(1/\epsilon_H, 1/T)$ plane (see caption of Fig. 1 for details). Dotted line corresponds to the fit given in Equation 3B. As indicated in the figure, for $\epsilon_H \sim 1$, the $\langle\chi\rangle = 0.1$ curve (dashed line) is very close to a straight line.

in the overlap function, namely from the temperature dependence of $\Delta\chi$. Our previous studies (Camacho & Thirumalai, 1993) had shown that this transition is a finite-sized first-order-like phase transition.

Because the main purpose of our study is to compare folding rates as a function of ϵ_H, we have adjusted the temperature of our simulations for all values of ϵ_H so that the statistical average of the overlap function is exactly 0.1, i.e.,

$$\langle\chi(T, \epsilon_H, \epsilon_S = 3)\rangle = 0.1. \tag{2}$$

Recall that $\langle\chi\rangle$ measures the amount of non-native structure in a given conformation (see Equation 1). Thus, by fixing $\langle\chi\rangle$ for different conditions, the folding rate to achieve the same level of the native structure can be obtained and compared. When Equation 2 is used to divide the (ϵ_H, T) plane into folded and unfolded states, the dotted line in Figure 1 results. This line is also displayed explicitly in Figure 2. We note that the dotted line closely follows the folding–unfolding transition line and for our purposes it will be used as the "phase boundary." For this system, the dotted line for $\epsilon_H \lesssim 1$ is well characterized by:

$$\frac{1}{T_-} = C_1/\epsilon_H, \tag{3A}$$

where $C_1 = 2.1949$ is computed exactly using Equation 2 with $\epsilon_H = 0$. For $\epsilon_H \gtrsim 1$, $1/T$ is well approximated as:

$$\frac{1}{T_+} \approx C_2/\epsilon_H + C_3, \tag{3B}$$

with $C_2 = 1.84$ and $C_3 = 0.32$. When $\epsilon_H \gg 1$, $1/T$ approaches a constant value $1/T_0 = 0.5001$.

Energy landscape

In order to establish a direct relationship between the kinetics of folding and the underlying microscopic aspects of the folding dynamics, we have mapped the minimum energy barriers connecting all the relevant low-lying minimum energy structures. This method, first used in our earlier paper (Camacho & Thirumalai, 1993) on protein folding kinetics of lattice models, has proven useful in revealing the key features of the folding dynamics like the intermediate regime in the three-stage folding process and the role of conformational entropy. It is particularly suitable for lattice simulations of small enough systems where the whole spectrum of conformations can be determined explicitly by direct enumeration. Some generalized version of the algorithm, however, can also be used in continuous models.

The algorithm is as follows. The total energy is discretized in the smallest energy scale ΔE. Then, one starts an MC simulation with the native state structure, with energy E_N, as the initial condition, and arbitrary positive temperature. For optimum sampling, it proves convenient if $T \sim T_\theta$, the collapse transition temperature. The energy is initially restricted to vary between E_N and $E_N + \Delta E$. After all new states have been sampled and stored, the maximum possible energy is increased to $E_N + 2\Delta E$, and so on. Every time there is an attempt to decrease the energy, one checks whether or not the new structure has been stored already. If the simulation samples a lower energy conformation for the first time, it means that it belongs to a new minimum whose energy barrier to the

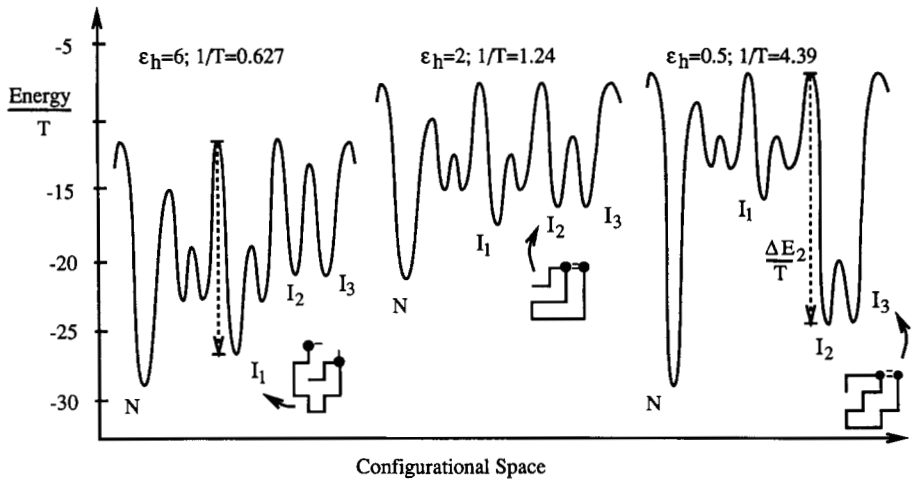

Fig. 3. Schematic energy landscapes for ϵ_H = 6, 2, and 0.5, and $\langle\chi\rangle$ = 0.1. The smallest energy barriers in the whole configurational space connecting the minimum energy conformations N, I_1, I_2, I_3 (and three others) have been plotted. The horizontal axis is rather artificial; however, it should correspond to some measure of distance in conformational space between the different minima. The dashed arrows highlight the largest energy barriers for large and small ϵ_H. The structure of the minima I_1, I_2, and I_3 are also shown.

native state is precisely the allowed energy gap at that moment. Then, one identifies the new local minimum and follows the same recipe. With this strategy, all trivially connected states, i.e., states that have at least one downhill pathway to the local minimum, can be mapped easily. The method, however, becomes inefficient if two low-lying minima are far apart in configurational space. In this case, one does an exhaustive MC simulation to determine the minimum energy gap for which a conformation in one minimum reaches a conformation of the other one, by once again increasing the energy in units of ΔE. The algorithm becomes computationally intensive for large values of E.

In Figure 3, the energy landscape showing the connectivity between the native state and certain low-energy structures like I_1, I_2, I_3, and others as a function of conformation space for three values of ϵ_H is displayed. The method gives the energy barrier between a pair of states. For example, the energy barriers correspond to the minimum activation energy separating certain low-lying metastable states (I_1, I_2, I_3, etc.) and the native conformation. Notice that the energy in Figure 3 has been divided by temperature. A plot of some low-energy structures is also shown in Figure 3. There are several aspects of this figure that are worth noting: First, for $\epsilon_H/\epsilon_S \gg 1$, the predominant barrier is between the single metastable intermediate I_1 and N with $\Delta E_+ = 4\epsilon_H$. Second, in the opposite limit, with $\epsilon_H \ll 1$, the predominant barriers are between several metastable states such as I_2 and I_3 (and five other similar states that are not explicitly displayed) and the native conformation, with $\Delta E = 2\epsilon_H + \epsilon_S$. This situation is exactly the same as the one analyzed in our earlier study, in which we showed that, under these circumstances, entropic barriers are more relevant (Camacho & Thirumalai, 1995a). This implies that the presence of a large number of similar native-like states (such as I_2, I_3) slows down the folding process (see Equation 5 below) because the chain can get trapped easily in one of these conformations. Then the loss in entropy has to be compensated to reach the native conformation. Hence, the origin of entropic barriers. A quantitative analysis of the origin of entropic barriers is given elsewhere (Camacho & Thirumalai, 1995a). From the arguments given above, it follows that optimal rates of folding are obtained only for certain intermediate values of ϵ_H. This, in fact, is confirmed in Figure 4.

Folding rates

The folding rate k_f is computed by fitting an exponential of the form $\exp(-k_f t)$, where t is measured in terms of MCS, to the long time decay of the correlation function $\langle\chi(t)\rangle$. In our computation, folding is assumed to be complete when $\langle\chi\rangle$ reaches the preset value of 0.1.

The dependence of the rate on ϵ_H and T is displayed in Figure 1 and is analyzed in more detail in Figure 4. It is clear from Figure 4 that folding rate is vanishingly small in the extreme limit of small or large values of ϵ_H (Camacho & Thirumalai, 1995a). The optimal

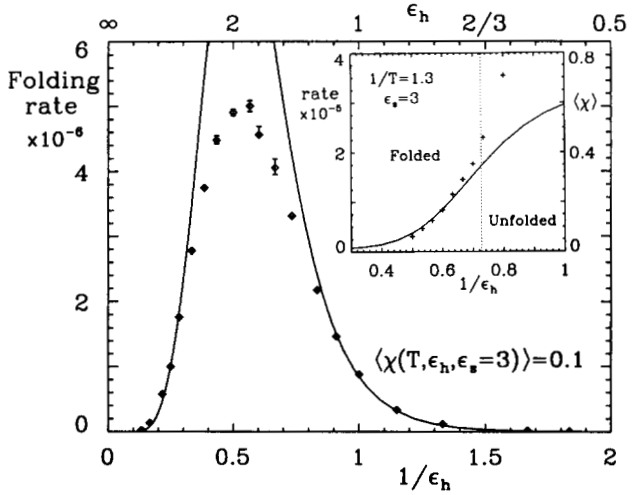

Fig. 4. Folding rate (diamond symbols), measured in units of [1/MCS], as a function of ϵ_H^{-1}. The solid line is a fit to the data using Equation 5. The inset shows the dependence of the folding rate on ϵ_H^{-1} for a fixed value of temperature ($1/T$ = 1.3). The dotted line in the inset separates the folded and unfolded states, i.e., it defines the locus of folding–unfolding transition. The + signs is the rate for reaching equilibrium with T^{-1} = 1.3. It is clear that, when ϵ_H is above a certain value, the chain does not fold. $1/\epsilon_H$ at constant $\langle\chi\rangle$ = 0.1. Error bars measure the uncertainty in the numerical determination of the rate. At constant temperature, the amount of native structure (solid line) changes as a function of ϵ_H.

folding rate, i.e., minimum in the folding time, is achieved for $\epsilon_H/\epsilon_S \sim 1$. It has been argued elsewhere (Camacho & Thirumalai, 1995a) that, for models of the kind discussed here, the folding rates can asymptotically be written as:

$$k \approx \begin{cases} \exp(-D_1\epsilon_H) & \epsilon_H \gg 1 \\ \exp(-D_2/\epsilon_H) & \epsilon_H \ll 1 \end{cases}, \quad (4)$$

where D_1 and D_2 are model dependent constants. The form for k given in Equation 4 can be used as the basis for fitting the rate. Not surprisingly, we find that the following expression:

$$k = 0.27[\exp(\Delta E_+/T_+) + C_4\exp(\Delta E_-/T_-)]^{-1}, \quad (5)$$

with $C_4 = 5.4$, T_+ and T_- given in Equation 3 are consistent with the computational results given in Figure 3 (see the solid line in Fig. 4). The barriers in Equation 5 are $\Delta E_+ = 4\epsilon_H$ and $\Delta E_- = 2\epsilon_H + \epsilon_S$. The scaling form used in Equation 5 embodies the key aspects of the energy landscape. Note that ΔE_+ and ΔE_- are the largest barriers separating the relevant metastable states and the native conformation. The constant C_4 measures the higher entropic weight associated with the seven metastable states compared with I_1. Thus, in the extreme cases, namely large and small values of ϵ_H, the folding rates are essentially determined by the largest energy barriers separating the metastable states and the native conformation.

The rate of folding is largest when $\epsilon_H/\epsilon_S \sim 1$ and, in this region ($1 < \epsilon_H < 3$), the fit given by Equation 5 is not adequate. In this interesting region, the rate of folding is determined by an effective free energy barrier involving an interplay between energetic and, perhaps more importantly, entropy considerations. Thus, it appears that, if the transition from metastable states to the native conformation is dominated by energetic barriers, then these states lead to slow kinetics. In this case, there are perhaps very few paths (or transition states) connecting the stable intermediates to the native conformation. In the fast folding regime (see Fig. 3 for $\epsilon_H = 2$), there are many paths connecting several metastable intermediates to the native conformations and this leads to an effective lowering of the free-energy barriers. It appears that, for fast folding proteins, there are several transition states associated with multipathway mechanism of protein folding (Sali et al., 1994; Bryngelson et al., 1995; Dill et al., 1995).

The maximum in the rate of folding in the region $1 < \epsilon_H < 3$ is reminiscent of similar behavior observed in other simulations (Miller et al., 1992; Socci & Onuchic, 1994). However, there are certain differences between the model considered here and the previous studies. In the two earlier studies (Miller et al., 1992; Socci & Onuchic, 1994), there is essentially only one independent energy scale, namely ϵ_H and T. For this class of models, it was shown that folding rate has a maximum as a function of T. In our model, there are two energy scales and the maximum in the rate occurs due to the changes in the energy landscape induced by the interplay of the effective hydrophobic interaction ϵ_H and the free energy gain due to disulfide bond formation.

The inset in Figure 4 shows the dependence of the folding rate on ϵ_H at a fixed value of the temperature ($1/T = 1.3$). The dotted line in the inset is the phase boundary between the folded and unfolded states. The plus signs (+) in the inset gives the rate for reaching the equilibrium conformation at constant temperature ($T^{-1} = 1.3$) and for different values of ϵ_H. It is clear that when ϵ_h^{-1} is below a certain value, the chain does not fold. When ϵ_H is such that the system is classified thermodynamically as unfolded, the rate of approaching the equilibrium conformation is rapid. However, the value of $\langle\chi\rangle$ (shown as a solid line in Fig. 4) is large, which implies that the amount of native conformation is a minimum.

Analysis of experiments

We have already mentioned that the experimental measurements usually maintain temperature at a constant value and vary the concentration of the denaturants (usually urea or guanidinium hydrochloride). In our computational studies, we have varied T and ϵ_H. Thus, in order to make a direct comparison with the experiments to assess the validity of the models we have to know the dependence of ϵ_H on concentration (c_D) of the denaturant. This is an extremely difficult problem in the physical chemistry of ternary systems, and, in general, it would be difficult to obtain an explicit relationship between the microscopic variable ϵ_H and the macroscopic concentration variable c_D. Here we rely on physical arguments to propose the simplest possible dependence of ϵ_H on c_D. As a consequence, our analysis of the experiments should be viewed as tentative.

The problem is to assess the variation of the average hydrophobic interaction ϵ_H with c_D. It is worth emphasizing here that $\epsilon_H(c_D)$ and $\epsilon_H(0)$ measure the gain in free energy upon transferring hydrophobic species from an organic solvent to water. It is known that the potential of mean force between spherical hydrophobic species in water has a primary minimum whose strength can be identified with ϵ_H. At sufficiently high concentration of the denaturants, proteins denature. If the driving force for protein folding is hydrophobic, then we can assume that

$$\epsilon_H(c_D)/k_BT \ll 1 \quad (6)$$

for large enough c_D. For $c_D = 0$, ϵ_H retains the pure value (in the absence of the denaturant) that is expected to be significantly greater than the high concentration limit. Thus, at least within a limited concentration range, one can suggest that

$$\epsilon_H(c_D) \simeq \epsilon_H(0) - k_Cc_D, \quad (7)$$

where k_C is constant. There are, of course, higher-order corrections that we neglect here. With this ansatz, one can map the results for the model presented above to analyze some experiments.

The dependence of ϵ_H on c_D given in Equation 7 suggests that *the primary role of the concentration of the denaturant is to decrease the strength of the hydrophobic interaction.* One possible mechanism for how this can happen is the following. It is likely that the denaturant may associate preferentially with the hydrophobic species. If this is the case, then the denaturant molecules effectively act as bumpers, causing a repulsive interaction between two hydrophobic molecules with which the denaturant has formed a complex. The presence of urea also disrupts the hydrogen bond network of the water. These effects, which combine entropy loss due to local ordering of water molecules around the hydrophobic species–denaturant complex and an increase in enthalpy due to repulsive interaction in the complex, would lead to a lowering of ϵ_H. At present, the precise mechanism of the action of denaturants is not known. The relationship in Equation 7 is perhaps the simplest conjecture of how ϵ_H depends on c_D.

This mechanism also suggests why denaturants do not affect the native conformation significantly. In the native state, the majority of the hydrophobic species are buried in the interior. Consequently, they are shielded from the denaturant molecules, which, in our model, are assumed to interact predominantly with the hydrophobic residues. Thus, over a limited concentration regime, the native state is relatively unaffected by denaturants. At sufficiently high concentration, our model breaks down.

The aforementioned arguments leading to the simple representation of the dependence of ϵ_H on c in Equation 7 are not entirely new. The effects of urea and Gdn-Hcl on the denaturation process have been explained based on localized free energy effects at the hydrophobic sites. In fact, the dependence of the equilibrium constant between denatured states and the native conformation on the activity of denaturants have been explained by assuming that the hydrophobic sites act as binding sites for denaturants (Tanford, 1970). If one assumes that all sites are roughly equivalent, then one can show that the equilibrium constant K is given by:

$$K = K_0(1 + \alpha c_D). \tag{8}$$

By taking logarithm on both sides of Equation 8 and by identifying $-k_B T \log K$ as the average hydrophobic interaction between the residues, one recovers Equation 7 in the limit of small c_D. Thus, our model is closely related to the ones proposed earlier. More recently, a statistical thermodynamic theory has been developed to examine the effect of solvents on the stability of globular proteins (Alonso & Dill, 1991; Shortle et al., 1992). Their theory also leads to a linear correction to the effective hydrophobic interaction. Thus, it appears that the effect of the denaturant is qualitatively captured by the model presented here. It should be stated that the model leading to Equation 8 has been quite successful in the analysis of denaturation induced by urea and Gdn-Hcl (Greene & Pace, 1974; Barrick & Baldwin, 1993).

We begin with the experiments of Weissman and Kim (1991). These authors performed refolding experiments on BPTI using acid quench techniques to trap the relevant intermediates. It is known that, in this problem, the folding intermediates can be characterized by the formation of disulfide bonds between the various cysteine residues (Creighton, 1978). The process of folding is well correlated with the formation of S-S bonds. As a result, the "reaction coordinates" can be associated with the S-S bonds. The rate-limiting step in the transition to the native conformation involves the rearrangement of the native-like intermediate [30–51; 14–38] to [30–51; 5–55]. We note parenthetically that, even in this well-studied case, there are still controversies concerning the folding pathways (Creighton, 1992; Weissman & Kim, 1992). This nomenclature of the intermediate means that a disulfide bond between Cys residues at between 30 and 51 and one at between 14 and 38 have been formed. In this notation, the native conformation is denoted as [30–51; 14–38; 5–55]. Weissman and Kim showed that, as the concentration of urea is increased from 0 to 6 M, the rate for rearrangement of [30–51; 14–38] to [30–51; 5–55] and [14–38; 5–55] increases by a factor of eight. Because urea does not effect the primary mechanism of formation of the disulfide bonds via a thiol-disulfide exchange mechanism, it is reasonable to suggest that the enhancement in the rate is due to the destabilization of the native-like species [30–51; 14–31] (Weissman & Kim, 1991).

The inset in Figure 4 that shows the rate of approach to equilibrium at a fixed T and varying ϵ_H^{-1} bears a resemblance to Figure 6B in the refolding of BPTI (Weissman & Kim, 1991). It should be emphasized that, even when we constrain the system to have the same extent of native structure (as measured by $\langle\chi\rangle$), the folding rate increases between $0.1 < \epsilon_H^{-1} < 0.5$. If we assume that some general trends seen in proteins can be understood in terms of minimal models, then Figure 4, together with Equation 7, offers a tentative rationale for the findings reported in Weissman and Kim. Using Equation 7, it follows that increasing $1/\epsilon_H$ is equivalent to increasing the concentration of the denaturant. Thus, our Figure 4 shows that, as ϵ_H^{-1} increases the rate increases. It is only meaningful to use the inset in Figure 4 in the region where the chain is essentially folded. Notice that the increase in the rate is more drastic precisely in the range of the ratio ϵ_H/ϵ_S, in which the overall folding rate is maximized (see Fig. 4). The reason for the increase in the rate is most evident from the energy landscape profile in Figure 3, which shows that, for $\epsilon_H = 2$, there is an effective destabilization of all intermediates that in turn leads to a decrease in the barriers between the kinetic traps and the native conformation. This explanation, which is made quantitative within the context of our model, confirms the expectations of the experimentalists. Our results, however, show that at higher concentrations the folding rate should decrease provided folding to a state with very similar native character is monitored. Such a turnover has not been observed experimentally.

It should be emphasized that both this study and that of Weissman and Kim (1991) suggest the rate increase with the addition of urea is only modest—the enhancement is about a factor of six. This observation allows us to obtain the extent to which urea destabilizes the native-like intermediate [30–51;14–38] in BPTI. The experiments suggest that the rate-limiting step is the conversion of this intermediate to [30–51; 5–55]. Assuming that this involves overcoming an average free energy barrier and using the fact that the rate for this process increases by roughly a factor of six when urea concentration increases from 0 to 6 M, it follows that:

$$\Delta F\,[30\text{–}51;\ 14\text{–}38] \sim k_B T\, \ell n(6) \sim 1.6 \text{ kcal/mol}. \tag{9}$$

Here ΔF is the extent of destabilization of [30–51; 14–38] with the addition of urea. We have assumed that $T = 37$ °C.

Another consequence of our analysis is that the folding rates can be altered significantly by a combination of mutations and denaturants that would change $\epsilon_H(0)$ and $\epsilon_H(c_D)$ (see Equation 7). This is particularly relevant to the experiments on RNase T1, which can be interpreted using Equation 7 and the general notion of the denaturant-enhanced folding rates. These experiments showed that, when the concentration of the denaturant was increased, the rate of the folding of the slow step increased, which fits in with the general scenario discussed thus far. In addition, it was observed that when tryptophan (Trp) 59 of RNase T1 is replaced by tyrosine (Tyr) the folding rate was accelerated. The folding rate of the resultant mutant W59Y, however, decreased with addition of the denaturant that is normally observed in other proteins (Kelley et al., 1986; Schmid et al., 1986). Both of these salient observations can be rationalized in terms of the model suggested here. When Trp is substituted at 59 by Tyr, there is an effective decrease in ϵ_H, the average hydrophobic attraction. This alone is sufficient to speed up the folding rate, as is evident from Equation 7 and Figure 4. Note that, as implied by Equation 9, only a relatively small amount of decrease in ϵ_H can lead to moderate increase in the rate of folding. When the denaturant is added to the mutant W59Y, the

effective hydrophobic interaction is decreased further. This could cause destabilization of the rapidly formed intermediate to such an extent that the rate actually would start to decrease (see Fig. 4). This is again in accord with previous experiments (Kiefhaber et al., 1992), and seems to suggest that the folding rate is optimum only in a limited range of ϵ_H. The averaged hydrophobic interaction can be changed by mutation, by adding denaturants, or both.

Conclusions

In this article, using simple models of disulfide-bonded proteins, we have shown that the strength of the average hydrophobic interaction plays an essential role in determining the rate of folding. If the temperature is fixed, as is usually the case in most experiments, then it has been established that the maximal folding rates are achieved only for a limited range of the effective hydrophobic interaction. A detailed analysis of the underlying energy landscape reveals that, in this range, there are many (10–20) low-energy conformations; their combined entropy makes up for the energy difference between the native state and these structures. In the late stages of folding, this entropic contribution has to be overcome. It is interesting that, in refolding of BPTI, the rate-determining step is, in fact, the transition from the native-like intermediate. This appears to be in conformity with our conclusions here and elsewhere (Weissman & Kim, 1991; Camacho & Thirumalai, 1995b). Our calculations also show that, in general, fast folding is achieved in situations in which there are several low-energy states that connect to the native states, thus involving multiple transition states.

We have also argued that the studies here provide support for the experimental findings that denaturants can speed up the folding rate of certain classes of proteins. By using a model that correlates the concentration of urea with the strength of the hydrophobic interaction (see Equation 7), we have suggested that the folding rate can increase the rate of folding in a limited concentration regime. This rather unexpected result can be found whenever there are stable native-like intermediates; then, denaturants can effectively destabilize them, leading to enhanced folding rates. Both previous work (Weissman & Kim, 1991) and our simulations suggest that the increase is only modest, i.e., the rate increases by a factor of about six.

Although we have formulated the problem in terms of simplified models of disulfide-bonded proteins with the experiments of Weissman and Kim in mind, our conclusions may be fairly general. In using the simulation results to interpret experiments, we suggested that the mechanism by which urea increases the rate of folding is effectively decreasing the strength of the hydrophobic interaction. This, in turn, leads to the destabilization of the metastable intermediates without significantly altering the energy of the native conformations. These effects lead to lowering of the free energy barriers between the metastable states and the native conformation. Thus, our prediction is that the enhancement of the rate of folding by the addition of denaturants should be a general phenomenon that should be observable wherever long-lived partially folded proteins serve as dominant intermediates. It is likely that some of the slow phases in the folding kinetics may be almost eliminated by adding moderate amounts of denaturants.

Acknowledgments

D.T. thanks Prof. S. Doniach for useful discussions. This work was supported in part by a grant from the National Science Foundation (through grant number NSF 93-07884) and the Air Force Office of Scientific Research. C.J.C. acknowledges support from FONDECYT no. 3940016.

References

Abkevich VI, Gutin AM, Shakhnovich EI. 1994. Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations. *J Chem Phys 101*:6052–6062.

Alonso DOV, Dill KA. 1991. Solvent denaturation and stabilization of globular proteins. *Biochemistry 30*:5974–5985.

Barrick D, Baldwin RL. 1993. Three state analysis of sperm whale apomyoglobin folding. *Biochemistry 32*:3790–3796.

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

Camacho CJ, Thirumalai D. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA 90*:6369–6372.

Camacho CJ, Thirumalai D. 1995a. Modeling the role of disulfide bonds in protein folding: Entropic barriers and pathways. *Proteins Struct Funct Genet 22*:27–40.

Camacho CJ, Thirumalai D. 1995b. Theoretical predictions of folding pathways using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci USA 92*:1277–1281.

Creighton TE. 1978. Experimental studies of protein folding and unfolding. *Prog Biophys Mol Biol 33*:231–297.

Creighton TE. 1992. The disulfide folding pathway of BPTI. *Science 256*:111–112.

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

Greene RF, Pace CN. 1974. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *J Biol Chem 249*:5388–5393.

Guo Z, Thirumalai D. 1995. Kinetics of protein folding: Nucleation mechanism, time scales and pathways. *Biopolymers 36*:83–102.

Kelley RF, Wilson J, Bryant C, Stellwagen E. 1986. Effects of guanidine hydrochloride on the refolding kinetics of denatured thioridazine. *Biochemistry 25*:728–732.

Kiefhaber T, Grunert HP, Hahn U, Schmid FX. 1992. Folding of RNase T1 is decelerated by a specific tertiary contact in a folding intermediate. *Proteins Struct Funct Genet 12*:171–179.

McCoy LF, Rowe ES, Wong KP. 1980. Multiparameter kinetic study on the unfolding and refolding of bovine carbonic anhydrase B. *Biochemistry 19*:4738–4743.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AN, Teller E. 1953. Equation of state calculations by fast computing machines. *J Chem Phys 21*:1087–1092.

Miller R, Danko CA, Fasolka MJ, Balazs AC, Chan HS, Dill KA. 1992. Folding kinetics of proteins and copolymers. *J Chem Phys 96*:768–780.

Sali A, Shakhnovich E, Karplus M. 1994. How does a protein fold? *Nature 369*:248–251.

Schmid FX, Grafl R, Wrba A, Beintama JJ. 1986. Role of proline peptide band isomerization in unfolding and refolding ribonuclease. *Proc Natl Acad Sci USA 83*:872–876.

Shortle D, Chan HS, Dill KA. 1992. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci 1*:201–215.

Socci ND, Onuchic JN. 1994. Folding kinetics of protein-like heteropolymers. *J Chem Phys 101*:1519–1528.

Socci ND, Onuchic JN. 1995. Kinetic and thermodynamic analysis of protein like heteropolymers: Monte Carlo histogram technique. *J Chem Phys 103*:4732–4744.

Tanford C. 1970. Protein denaturation: Theoretical models for the mechanism of denaturation. *Adv Protein Chem 24*:1–95.

Verdier PH, Stockmeyer WH. 1967. Monte Carlo calculations on the dynamics of polymers in dilute solution. *J Chem Phys 36*:277–285.

Weissman JS, Kim PS. 1991. Reexamination of the folding of BPTI: Predominance of native intermediates. *Science 253*:1386–1393.

Weissman JS, Kim PS. 1992. The disulfide folding pathway of BPTI: Response. *Science 256*:112–114.