ELSEVIER

Thermal denaturation and folding rates of single domain proteins: size matters

Mai Suan Li[a], D.K. Klimov[b], D. Thirumalai[b],*

[a]*Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland*
[b]*Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

Abstract

We analyze the dependence of thermal denaturation transition and folding rates of globular proteins on the number of amino acid residues, N. Using lattice Go models we show that $\Delta T/T_F \sim N^{-1}$, where T_F is the folding transition temperature and ΔT is the transition width computed using the temperature dependence of the order parameter that distinguishes between the unfolded state and the native basin of attraction. This finding is consistent with finite size effects expected for the systems undergoing a phase transition from a disordered to an ordered phase. The dependence of the folding rates on N for lattice models and the dataset of 57 proteins and peptides shows that $k_F \simeq k_F^0 \exp(-CN^\beta)$ with $0 < \beta \leq 2/3$ provides a good fit, where C is a β-dependent constant. We find that $k_F \simeq k_F^0 \exp(-1.1N^{1/2})$ with an average (over the dataset of proteins) $k_F^0 \approx (0.4 \ \mu s)^{-1}$, can estimate optimal protein folding rates, to within an order of magnitude in most cases. By using this fit for a set of proteins with β-sheet topology we find that $k_F^0 \approx k_U^0$, the prefactor for unfolding. The maximum ratio of $k_U^0/k_F^0 \approx 10$ for this class of proteins.
© 2003 Published by Elsevier Ltd.

Keywords: Folding rates; Folding barriers; length scaling

1. Introduction

Deciphering the factors that determine the foldability of protein sequences [1–5] is an important problem from the perspective of protein design, protein structure prediction, and in vitro and in vivo protein folding. Foldability refers both to the folding rate, k_F, and thermodynamics of the transition from the ensemble of unfolded states (U) to the native state or, more precisely, to the native basin of attraction (NBA). Folding rates and the associated equilibrium characteristics depend on intrinsic factors (sequence and topology) as well as on external conditions (pH, temperature, salt concentration, and viscosity). Variation in external conditions cannot only alter the rates, but also the mechanism of folding. Despite this obvious fact most of the studies have been focused on the dependence of k_F solely on the characteristics of the native states as described by crystal (or NMR) structures.

The role of finite size effects on the thermodynamics of protein folding has received very little attention. The emphasis on the cooperativity of the transition from U state to NBA seems to have precluded consideration of the

role of N, the number of residues in a sequence. The transition to the folded state, for apparent two-state folders, has all the hallmarks of (weak) first-order phase transition. The highly cooperative U $\leftrightarrow$ NBA transition has lead some authors to suggest that there is no evidence that partially structured states contribute to the thermodynamic properties of proteins. Computational studies have shown [6] that in β-hairpin forming sequence from the C terminus of GB1 protein structure is acquired over a finite range of temperatures, even though the overall folding can be described as a broad 'two-state' transition [7]. Experiments on refolding of barnase have also suggested that structure is lost incrementally upon temperature induced unfolding [8]. Direct temperature dependence of structure formation in leucine zipper using one-dimensional NMR experiments has established that melting temperature varies across the structure [9]. Although the variations occur over a relatively narrow range of temperatures, it is clear from these experiments that because of the finite size of proteins partially folded structures contribute to folding thermodynamics. These observations warrant an examination of finite size effect on the U $\leftrightarrow$ NBA transition. Building on our previous study [10], we further investigate the role of N in thermal denaturation using lattice models of proteins.

It has been noted [11] that k_F correlates well with the relative contact order (RCO), which measures the proximity of side-chain contacts in the folded state. The notion that protein folding is initiated with residues forming local structures and, thus, is determined by their proximity along the sequence is the basis of the hierarchical folding mechanism [12]. Thus, in retrospect, the correlation between the RCO and k_F is not entirely unanticipated, especially in α-helical proteins. Although RCO is an important indicator of the folding rates, it should be pointed out that there is little correlation between RCO and k_F for proteins with β-sheet topology. Clarke et al. [13] showed that neither k_F nor the unfolding rates k_U correlate with RCO for a class of β-sheet proteins belonging to the immunoglobulin (Ig) fold. The RCO for the six proteins examined [13] is in the very narrow range (0.17 ≤ RCO ≤ 0.20). Nevertheless, the refolding rates for these proteins vary by a factor of 800. More recently, Clarke and co-workers have shown that for a number of Ig domains from the muscle protein titin k_F can vary by over four orders of magnitude [14], although their RCO values are nearly the same. These studies show that factors besides RCO play an important role in the determination of k_F.

Surprisingly, it was initially suggested [11] that neither stability nor the size of proteins plays a role in determining k_F. These counterintuitive observations contradict several theoretical [15–18] and a few experimental studies [13]. More careful examination of the database of well-characterized proteins has shown that, although there are exceptions [19], stability is an important factor that determines k_F [13,20]. Recently, several studies [21,22] have concluded that N, the number of amino acids, must also play an important role in determining k_F. In this paper we examine the dependence of rates as well as thermal denaturation of single domain proteins on N.

Beginning with the paper by one of us [15] a number of theoretical studies [16–18] have predicted that N should play a significant role in controlling k_F. Given that polypeptide chains are heteropolymers we expect that their relaxation times in both the folded and unfolded states must depend on N. Theoretical studies [15,16] suggest that the dependence of k_F on N is dictated by the interplay of three characteristic temperatures of the polypeptide chain, namely, T_F (the folding transition temperature), T_θ (the collapse transition temperature), and T_g (the glass transition temperature). It appears that in most experiments the external conditions are such that fastest folding is observed near the 'tricritical' point, where $T_F \approx T_\theta$ in accord with the prediction by Camacho and Thirumalai [23]. For near optimal folding, as may be the case for minimally frustrated sequences, it has been argued that

$$\ln(k_F/k_F^0) \sim \alpha \ln N, \tag{1}$$

where $\alpha \approx 4$ [15] and k_F^0 is an undetermined prefactor. For artificial Go models $\alpha \approx 3$ [18,24]. On the other hand, due

to topological frustration, even the sequences following two-state kinetics have a rough energy landscape. In this case

$$\ln(k_F/k_F^0) \sim N^\beta. \tag{2}$$

The value of β has been suggested to be less than unity and is probably in the range $0.5 \le \beta \le 2/3$ [15–17]. Given the limited range of N for single domain proteins it is difficult (see below) to determine, β precisely.

To probe finite size effects on thermally induced folding we have performed Monte Carlo simulations using Go lattice models. These results are used to quantitatively establish the effect of finite N on rounding the U ↔ NBA transition. A dataset of proteins, for which k_F is available, is used to draw lessons on the dependence of k_F on N. Using these results we show that unambiguous determination of β is not possible. However, we argue that the N dependence given in Eq. (2) is useful in analyzing the experimental data. As a byproduct of this work we also provide estimates of the folding and unfolding prefactors, k_F^0 and k_U^0.

2. Models and methods

For the numerical simulations we represent a polypeptide chain using lattice Go model without side chains. The energy of a conformation

$$E = \sum_{i<j} \epsilon_{ij} \delta_{r_{ij},a}, \tag{3}$$

where a is a lattice spacing, r_{ij} is the distance between non-bonded beads i and j, and the contact energies ϵ_{ij} are chosen to be -1 for native contacts and 0 for non-native ones. Go models are useful in exploring general physical principles that govern protein folding under the condition of marginal stability of the native state [25,26]. The sequences were selected by a standard sequence space Monte Carlo algorithm, which maximizes the Z-score for a given target structure. The target structures for each N were chosen to be maximally compact. For example, for $N = 18$ and $N = 80$ the native structures occupy the vertices of $3 \times 3 \times 2$ and $4 \times 4 \times 5$ cubes, respectively.

The thermodynamics of folding were determined using Monte Carlo simulations based on MS3 move set [27–30], which involves single, double and triple bead moves. Because this move set involves multiparticle updates, it is much more efficient compared to the standard move set [29–31]. The thermodynamic properties of the sequences are calculated using the multiple histogram method [32]. Typical number of Monte Carlo trajectories used to collect histograms is 50–100 depending on N. The free energy is calculated as a function of the number of native contacts Q, which is treated as an approximate reaction coordinate for Go models. This allows us to estimate the dependence of folding and unfolding free energy barriers on N.

For lattice models the structural similarity with the native

conformation is measured by the overlap function [23]

$$\chi = 1 - \frac{1}{N^2 - 3N + 2} \sum_{i<j+1}^{N} \delta(r_{ij} - r_{ij}^0), \qquad (4)$$

where the superscript 0 refers to the native state. The folding temperature T_F is defined as a temperature at which $d\langle\chi\rangle/dT$ is maximum and the transition width ΔT is defined as the full width at half maximum of $d\langle\chi\rangle/dT$ at $T = T_F$.

3. Results

3.1. Finite size effects in thermal denaturation

The transition width ΔT is obtained from the temperature dependence of $d\langle\chi\rangle/dT$ (for e.g. see Fig. 1(a)). For all the sequences considered here $T_F \approx T_\theta$. For finite size systems the $U \leftrightarrow NBA$ transition is expected to be rounded. The rounded nature of the transition which has been seen in simulations, is reflected in the temperature dependence of $d\langle\chi\rangle/dT$ (Fig. 1(a)). More importantly, we expect $\Delta T/T_F$, to scale as

$$\frac{\Delta T}{T_F} \sim N^{-1}. \qquad (5)$$

The data for lattice Go models show that $\Delta T/T_F \sim N^{-\lambda}$ with $\lambda = 1.2 \pm 0.1$ (Fig. 1(b)). The small deviation from the expected theoretical result (Eq. (5)) may be a consequence of the relatively small $N \leq 80$ in the sample. For small values of N the native state does not have a well-defined core. As a result fluctuations are relatively large, which may explain the observed deviation. Analysis of the experimental data indeed shows that (Eq. (5)) is obeyed with great precision [10,33].

3.2. N dependence of folding and unfolding barrier heights at T_F for Go models

To compute the free-energy folding barriers, $\Delta F_F^{\ddagger}$ ($\simeq \Delta F_U^{\ddagger}$, the unfolding barrier, at T_F) it is necessary to define a reaction coordinate. The precise reaction coordinate for a multi-dimensional process such as protein folding is difficult to ascertain. However, Onuchic and co-workers [34] have argued that, for minimally frustrated systems such as the Go models, the fraction of native contact Q may be used as approximate reaction coordinate. Accordingly, we have computed $F(Q)$ for about 80 sequences with N ranging from 18 to 80. This is the largest number of sequences used so far to test the expected scaling of $\Delta F_F^{\ddagger}$, and $\Delta F_U^{\ddagger}$. At T_F, $\tau_F^0 \exp(\Delta F_F^{\ddagger}/k_B T_F) = \tau_U^0 \exp(\Delta F_U^{\ddagger}/k_B T_F)$. Because it is not obvious that $\tau_F^0 \approx \tau_U^0$, $\Delta F_F^{\ddagger}$, and $\Delta F_U^{\ddagger}$ may, in principle, exhibit different scaling behavior with N.

From the typical free-energy profile $F(Q)$ (Fig. 2(a)) we computed $\Delta F_F^{\ddagger}$ and $\Delta F_U^{\ddagger}$. The variation of $\Delta F_F^{\ddagger}/k_B T$ as a function of $\ln N$, $N^{1/2}$ and $N^{2/3}$ for the Go sequences plotted
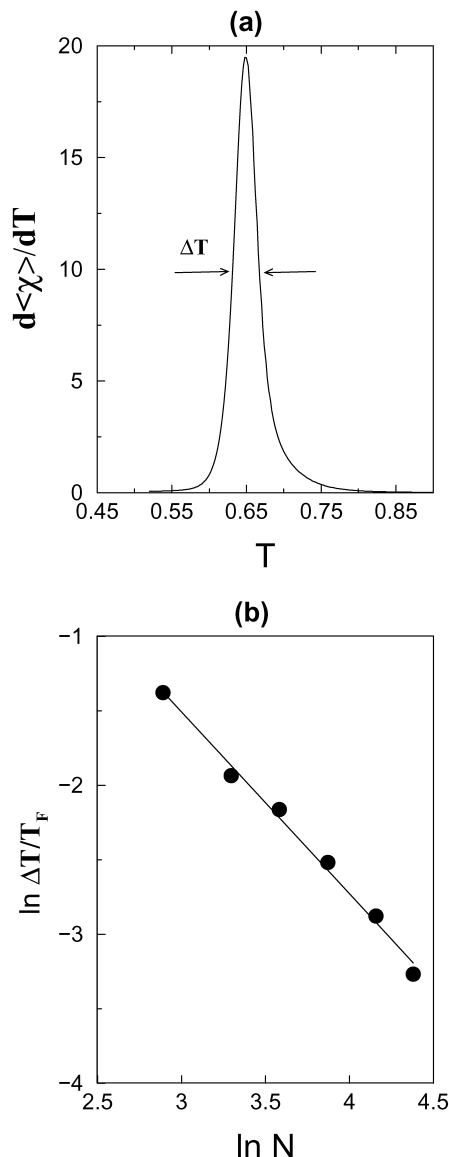

Fig. 1. (a) Temperature dependence of $d\langle\chi\rangle/dT$ for the lattice sequence with $N = 64$. The folding transition temperature is identified with the peak in $d\langle\chi\rangle/dT$. The full width at half-maximum is indicated by ΔT. (b) The dependence of $\Delta T/T_F$ as a function of N. The straight line gives the fit $\Delta T/T_F \sim N^\lambda$ with $\lambda = 1.2 \pm 0.1$.

in Fig. 2(b), (c) and (d), respectively, shows that all three fits quantitatively reproduce the simulation results. However, we argue below using the analysis of experimental data that $\Delta F_F^{\ddagger} \sim \ln N$ is not viable. Based on experimental estimates of τ_F^0 and τ_U^0 we find that $\Delta F_F^{\ddagger} \sim N^{1/2}$ provides the best physically acceptable representation of the data. From the lattice model computations we find $\Delta F_F^{\ddagger}$ and $\Delta F_U^{\ddagger}$ have the same dependence on N, which implies that $\tau_F^0 \approx \tau_U^0$.

3.3. Chain length dependence of folding rates

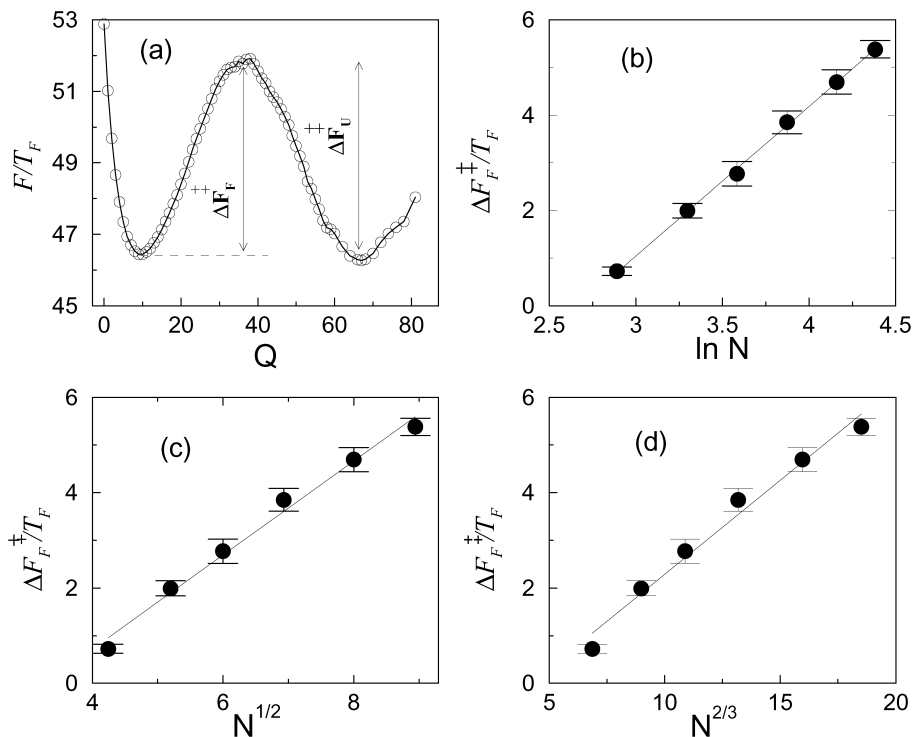The RCO, which is a characteristic of the native topology

Fig. 2. (a) Dependence of $F/k_B T$ (F is the free energy of a sequence) as a function of the presumed reaction coordinate Q, the number of native contacts, for one of the $N = 64$ Go sequences. The unfolding and refolding barriers are extracted from the free energy profile as indicated. Panel (b) shows the fit $\Delta F_F^{\ddagger}/k_B T_F \sim$ ln N. Panels (c) and (d) correspond to the fits $\Delta F_F^{\ddagger}/k_B T_F \sim N^{\beta}$ with $\beta = 0.5$ and 2/3, respectively. The results were computed for $N = 18(20)$, $27(17)$, $36(18)$, $48(18)$, $64(15)$, and $80(12)$, where the number in parenthesis refers to the number of sequences used for averaging $\Delta F_F^{\ddagger}/k_B T_F$. Similar scaling with N is obtained for $\Delta F_U^{\ddagger}/k_B T_F$.

of proteins, is [11]

$$\text{RCO} = \frac{\Sigma_{i,j}\Delta_{ij}|i-j|}{N\Sigma_{i,j}\Delta_{ij}}, \tag{6}$$

where $|i-j|$ is the sequence separation between the residues i and j and Δ_{ij}, is unity, if i and j form a native contact, or zero, otherwise. The observed correlation between RCO and ln k_F suggests that folding is most rapid, if the native state has a large fraction of local contacts. The importance of RCO is based on the sound physical idea that residues local in sequence space tend to form interactions early in the folding process and, if these substructures can 'coherently' add to produce the folded structure, efficient folding may be realized. However, almost all proteins are stabilized by a sizable fraction of long-ranged (non-local) contacts. This suggests that ln k_F may also depend on other factors (for example, stability [20] and N) besides RCO. Lattice model simulations [35] and experiments [13] have shown that native state stability is also a contributing factor to refolding rates.

Depending on the extent of energy frustration one of us suggested [15] that $k_F \sim N^{\alpha}$ (for optimized sequences) or $k_F \sim \exp(-C_1 N^{1/2})$, where C_1 is a constant (Eqs. (1) and (2)). By balancing the 'bulk' free-energy gain due to the formation of a stable hydrophobic core and the surface tension cost due to interface formation it has been proposed [16,17] that for optimal folding $k_F \sim \exp(-C_2 N^{2/3})$, where

C_2 is a constant. Although, the limited range of N values accessible in proteins makes it difficult to unambiguously determine the precise way k_F decreases upon increasing N, it is generally agreed that free-energy barriers in proteins shall be relatively small. Moreover, the transition region could be broad with roughness superimposed on it. As a result $\Delta F_F^{\ddagger}/k_B T_F$ is expected to grow only as N^{β} with $\beta < 1$. The sublinear growth of $\Delta F_F^{\ddagger}/k_B T$ with respect to N naturally explains both the rapid folding (kinetics) and marginal stability (thermodynamics) of folded states of proteins.

Recently, Koga and Takada [36] have computed folding rates for 18 proteins using C_{α}-Go models. They fit the data using $k_F \sim \exp(-C_3 \text{RCO} \times N^{\beta})$ with $\beta = 0.607 \pm 0.179$ and C_3 is a constant. Within the error bar of their fit it is impossible to distinguish between $\beta = 0.5$ or 2/3. Their results showed, as argued on theoretical grounds, that $\beta < 1$. In addition, due to the possibility that RCO decreases with N [22] it is likely that the actual value of β in Ref. [36] is considerably smaller. By focusing on the proteins that fold by three-state kinetics Galzitskaya et al. [21] have argued that chain length N is the major determinant of folding rates. However, they were unable to determine the precise dependence of k_F on N.

Ivankov et al. [22] have reconsidered chain length dependence of k_F by analyzing experimental data for 57 proteins (both two and three state folders) and peptides. They suggested that ln $k_F \sim -0.44 \text{RCO} \times N + 11.15$ for

the set of 57 proteins with the correlation co-efficient $p = 0.74$. For this dataset of proteins it is argued that $RCO \sim N^{-0.3}$, so that $k_F \sim \exp(-C_4 \times N^{0.7})$, where C_4 is a constant. Because there are errors in fitting RCO to a power law decay with N, the indirect inference that $\beta \approx 0.7$ is not transparent. To circumvent this problem we have directly examined the dependence of $\ln k_F$ on N. The fit of $\ln k_F$ using the theoretically proposed models are shown in Fig. 3(a), (b) and (c). The correlation coefficient for the fits $\ln k_F \sim N^\beta$ is nearly constant for $0 \le \beta \le 2/3$ and begins to decrease modestly for $\beta > 2/3$ (Fig. 3(d)). We have also established that the folding rates in lattice Go models can be adequately fit with $\beta = 0$, 0.5, or 2/3 [30]. From this perspective alone it is difficult to distinguish between the three theoretical values of $\beta(0, 0.5, \text{and } 2/3)$. However, we rule out $\ln k_F \sim \ln N$ (Fig. 3(b)) based on the following arguments: (1) The power law fit yields $\ln k_F = -5.5 \ln N + 28.5$ which implies $k_F^0 \approx e^{28.5} \text{ s}^{-1} = (0.4 \text{ ps})^{-1}$. This value for the prefactor k_F^0 is nearly the same as $k_B T/h \approx (0.2 \text{ ps})^{-1}$, which is reasonable for small molecules, but is not appropriate to describe folding reactions. (2) The value of the exponent $\alpha = 5.5$ is too large to be justified theoretically. Such a large value of α is usually indicative of an underlying activated process with a relatively small barrier [30].

The fits to the data in Fig. (3) cannot distinguish the scalings of $\ln k_F$ with $N^{1/2}$ or $N^{2/3}$. This is consistent with

our results presented in Fig. 2. In an attempt to further discriminate between $\beta = 0.5$ and $\beta = 2/3$ we focus on the numerical values of the prefactor k_F^0. The inverse of the prefactor $1/k_F^0$ for the $N^{1/2}$ scaling of the barrier height is 0.4 μs, whereas $1/k_F^0 \approx 8$ μs for $N^{2/3}$ scaling (see caption to Fig. 3). By applying Kramer's theory to describe the $U \leftrightarrow NBA$ transition it has been argued that $\tau_0 = 1/k_F^0$ should be considerably greater than $h/k_B T$ [37–39]. The range $0.4 \text{ μs} \le \tau_0 \le 8 \text{ μs}$ obtained from the two fits is consistent with this expectation. Therefore, it follows that, unless a direct experimental measurement of k_F^0 is made, it would be difficult to determine the precise value of β. The goodness of fits with $\beta = 1/2$ or $\beta = 2/3$ shows clearly that barriers to folding scale sublinearly with N.

3.4. Prefactors for folding and unfolding

There is considerable interest in obtaining a fairly accurate estimate of $\tau_F^0(\sim (k_F^0)^{-1})$ at near neutral pH and $T = 25\,^{\circ}\text{C}$ so that the measurements of average barrier heights can be made directly. Estimates of τ_F^0 have been made using few physically motivated arguments:

(1) Assuming that the most elementary step in the folding process is the formation of a single tertiary contact (a loop between two residues separated by l intervening residues) it was argued that the speed limit for folding is about 1 μs [40]. Because most probable loops are predicted to form in
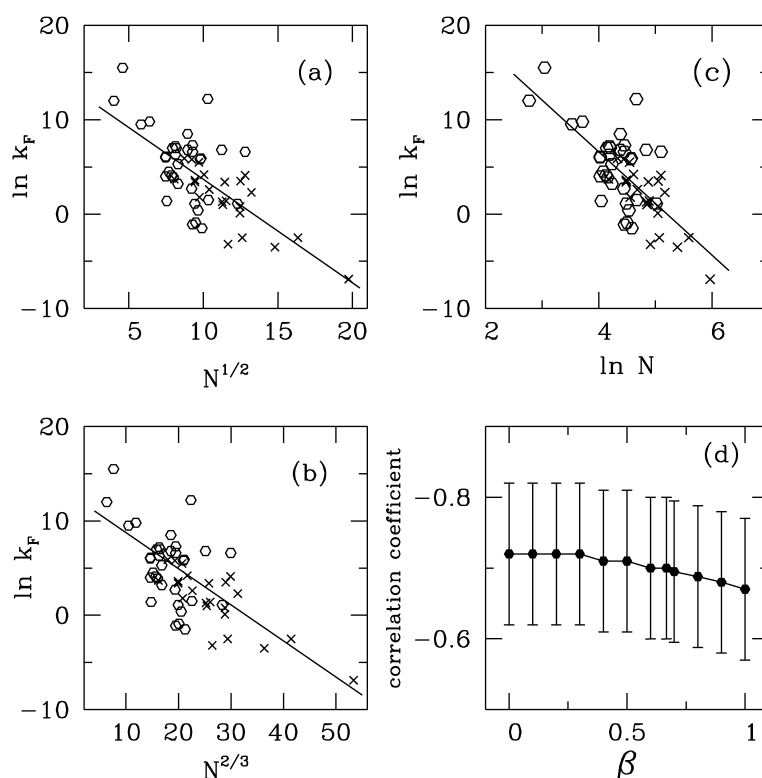

Fig. 3. Fits of $\ln k_F$ as a function of N for the dataset of 57 proteins and peptides taken from Ref. [22]. Cross and hexagon symbols correspond to three and two state folders, respectively. (a) The fit based on $\ln k_F \sim N^{1/2}$. The straight line is $y = -1.1x + 14.7$ and the correlation coefficient is 0.71. (b) The fit based on $\ln k_F \sim N^{2/3}$. The straight line is $y = -0.36x + 11.7$ and the correlation coefficient is 0.70. (c) Fits of $\ln k_F \sim \ln N$ gives $y = -5.5x + 28.5$ with the correlation coefficient of 0.72. (d) Variation of the correlation coefficient with β. The correlation becomes weaker at $\beta > 2/3$.

about 1 μs [42], it follows that $\tau_F^0 \approx 1$ μs. Eaton and co-workers [40,41] have provided additional arguments that proteins are unlikely to fold faster than $\tau_F^0 \approx 1$ μs.

(2) Yang and Gruebele [43] argue using refolding data of mutants of a helical protein λ_{6-85} that $\tau_F^0 \approx 2$ μs. We believe that, in general, for the majority of proteins $\tau_F^0 \approx 2$ μs should be near the upper limit for the following reasons. Based on theories of collapse dynamics we expect that the 80 residue protein λ_{6-85} becomes compact in about $\tau_c \approx (\eta a/\gamma)N^\theta \approx 1.5$ μs, where η is the solvent viscosity, a is the Flory characteristic ratio, γ is the surface tension (50 cal/mol Å^2), $\theta \approx 2.2$ and $N = 80$. This estimate is close to the folding time for λ_{6-85}, which suggests that collapse and folding are nearly simultaneous for this protein. Because these two processes cannot be separated for proteins with $N = 80$ that fold in about ≈ 1 μs, it appears that one can assume that $\tau_F^0 \approx 2$ μs may be an upper bound. We believe that $\tau_F^0 \approx 1$ μs could serve as a practical estimate for the prefactor, because on time scales greater than 1 μs multiple loops can form and collapse of the entire polypeptide chain can occur, which could obfuscate direct determination of τ_F^0. In arriving at these estimates for λ_{6-85} we have assumed that internal viscosity does not alter folding rates appreciably. Although, a similar observation has been made for refolding of protein L [44] and CspB [45] it is unclear how important internal viscosity of proteins is in the determination of τ_F^0 [39]. In addition, external conditions can alter τ_F^0. Thus, $\tau_F^0 \approx 1$ μs should be taken merely as a useful estimate for the prefactor.

The dependence of $\ln k_F$ on N (Fig. 3) allows us to estimate τ_F^0 and τ_U^0 using the experimental data for proteins that are not fully represented in Fig. (3). We use refolding rates for several β-sheet proteins to estimate τ_F^0 (Table 1). Assuming $N^{1/2}$ scaling we find that τ_F^0 is in the range (0.1–18) μs. Except for $\tau_F^0 \simeq 18$ μs obtained for twitchin (TWIg18$'$) with low native state stability the average

value of the prefactor is $\overline{\tau_F^0} \approx 3.5$ μs. If we use $\tau_F^0 \simeq \tau_F \times \exp(-0.36N^{2/3})$ (Fig. 3), we find 2 μs $\lesssim \tau_F^0 \lesssim 400$ μs (Table 1). For four immunoglobulin proteins with the exception of FNfn10 (Table 1) the estimated values of τ_F^0 using the $N^{2/3}$ scaling for the barrier height seem too large. Thus, assuming $N^{(1/2)}$ scaling τ_F^0 appears to be in the neighborhood of few μs for the β-sheet proteins and for the α-helical protein λ_{6-85}.

Another question of interest is whether $\tau_F^0 \approx \tau_U^0$? Using lattice model simulations we have previously argued that the unfolding and folding prefactors are similar [35]. This conclusion was reached using the number of native contacts Q as a reaction coordinate. It is unclear whether this result is a consequence of our choice of the reaction coordinate. The results in Fig. (3) and the measured unfolding rates in Table 1 allow us to directly estimate

$$\tau_U^0 \simeq \tau_U \exp[-(1.1N^{1/2} + \beta\Delta G)], \tag{7}$$

where τ_U is the unfolding time, ΔG is the free energy of stability of the native state, and $\beta = (k_BT)^{-1}$. With the exception of TWIg18$'$ the ratio $\tau_U^0/\tau_F^0 < 1$ and is in the range $0.1 \lesssim \tau_U^0/\tau_F^0 \lesssim 1.0$. For this class of proteins the maximum value of $\tau_F^0/\tau_U^0 \lesssim 10$ (Table 1). Similar conclusions have been drawn for α-helical proteins as well. Thus, it appears that $\tau_U^0 \approx \tau_F^0$.

4. Conclusions

In this article we have considered finite size effects in thermal denaturation and folding kinetics. We have established using lattice models that the rounded transition as quantified by $\Delta T/T_F$, obeys the expected scaling (Eq. (5)). This is in accord with the earlier analysis of the experimental data [10], which further suggests that qualitative features of folding transition can be gleaned

Table 1
Estimates of the folding and unfolding prefactors

Protein[a]	$\beta\Delta G$[b]	τ_F[c]	τ_U[d]	τ_F^{0}[e]	τ_F^{0}[f]	τ_U^{0}[g]	τ_U^{0}[h]
TI 127 (89)	12.7	0.0313	2041	0.974	23.9	0.194	4.76
TWIg18$'$ (93)	6.9	0.667	3571	16.5	412	89.0	2220
CD2d1 (98)	11.5	0.0556	588	1.04	26.4	0.111	2.83
TNfn3 (92)	9.1	0.344	2174	9.00	224	6.35	158
FNfnlO (96)	15.9	0.00417	4348	0.0870	2.20	0.0113	0.285
CspB (*B. subtilis*) (67)	4.6	0.00145	0.101	0.178	3.82	0.125	2.68
CspB (*B. caldolyticus*) (66)	8.1	0.000730	1.56	0.0960	2.04	0.0623	1.32
CspB (*T. maritima*) (68)	10.6	0.00177	55.6	0.203	4.40	0.159	3.44

Data for the first five proteins are from Ref. [13] and the data for CspB proteins are from Ref. [19].
[a] Numbers in parenthesis are the values of N.
[b] Free energy of stability extrapolated to zero denaturant concentration.
[c] Folding times in seconds.
[d] Unfolding times in seconds.
[e] Folding prefactor (in units of μs) calculated using $\tau_F^0 = \tau_F \exp(-1.1N^{1/2})$.
[f] Folding prefactor (in units of μs) calculated using $\tau_F^0 = \tau_F \exp(-0.36N^{2/3})$.
[g] Unfolding prefactor (in μs) calculated using $\tau_U^0 = \tau_U \exp(-1.1N^{1/2} - \Delta G/(k_BT))$.
[h] Unfolding prefactor (in μs) calculated using $\tau_U^0 = \tau_U \exp(-0.36N^{2/3} - \Delta G/(k_BT))$.

using lattice models. Unlike the case of thermal denaturation the situation is far more ambiguous when the scaling of k_F with N is considered. The dependence of $\ln k_F$ on N does not match the quality of correlation noted for thermodynamics. If we delete the fastest folding proteins and peptides and the slowest folding proteins from the dataset in Fig. 3, the correlation coefficient becomes considerably worse (≈ 0.56) regardless of the scaling ($\beta = 0.5$ or 2/3) used. Nevertheless, the inclusion of the N dependence does improve the correlation between $\ln k_F$ on RCO [22]. Using the expected values (from a number of unrelated studies) for the prefactor, we suggest that the $N^{1/2}$ scaling for barrier height $\Delta F_f^{\ddagger} B T$ may be useful in making order of magnitude estimates of refolding rates. This scaling also implies that the energy landscape of two-state proteins is rugged. The energy scale for roughness may be of order of a few $k_B T$.

Acknowledgements

References

[1] Onuchic JN, Luthey-Schulten Z, Wolynes PG. Ann Rev Phys Chem 1997;48:545–600.
[2] Thirumalai D, Klimov DK. Curr Opin Struct Biol 1999;9:197–207.
[3] Dill KA, Chan HS. Nat Strut Biol 1997;4:10–19.
[4] Mirny L, Shakhnovich E. Ann Rev Biophys Biomol 2001;30:361–96.
[5] Bonneau R, Baker D. Ann Rev Biophys Biomol Struct 2001;30:173–90.
[6] Klimov DK, Thirumalai D. Proc Natl Acad Sci USA 2000;97:2544–9.
[7] Munoz V, Thompson PA, Hofrichter J, Eaton WA. Nature 1997;390:196–9.
[8] Lakshmikanth GS, Sridevi K, Krishnamoorthy G, Udgaonkar JB. Nat Struct Biol 2001;8:799–804.
[9] Holtzer ME, Lovett EG, d'Avignon DA, Holtzer A. Biophys J 1997;73:1031–41.
[10] Klimov DK, Thirumalai D. J Comput Chem 2002;23:161–5.
[11] Plaxco KW, Simons KT, Baker D. J Mol Biol 1998;277:985–94.
[12] Baldwin RL, Rose G. Trends Biochem Sci 1999;24:77–83.
[13] Clarke JR, Cota E, Fowler SB, Hamill SJ. Structure 1999;7:1145–53.
[14] Scott KA, Steward A, Fowker SB, Clarke. J Mol Biol 2002;315:819–29.
[15] Thirumalai D. J Phys I (France) 1995;5:1457–67.
[16] Wolynes PG. Proc Natl Acad Sci USA 1997;94:6170–5.
[17] Finkelstein AV, Badretdinov AYa. Fold Des 1997;2:115–21.
[18] Gutin AM, Abkevich VI, Shakhnovich EI. Phys Rev Lett 1996;77:5433–6.
[19] Perl D, Welker C, Schindler T, Schroder K, Marahiel MA, Jaenicke R, Schmid FX. Nat Struct Biol 1998;5:229–35.
[20] Dinner AR, Karplus M. Nat Struct Biol 2001;8:21–2.
[21] Galzitskaya OV, Garbuzynskiy SO, Ivankov DN, Finkelstein AV. Proteins Struct Funct Cenet 2003;51:162–6.
[22] Ivankov DN, Garbuzynskiy SO, Alm E, Plaxco KW, Baker D, Finkelstein AV. Protein 2003;12:2057–62.
[23] Camacho CJ, ThirumaIai D. Proc Natl Acad Sci USA 1993;90:6369–72.
[24] Cieplak M, Hoang TX, Li MS. Phys Rev Lett 1999;83:1684–7.
[25] Kaya H, Chan HS. J Mol Biol 2002;315:899–909.
[26] Kaya H, Chan HS. J Mol Biol 2003;326:911–31.
[27] Betancourt MR. J Chem Phys 1998;109:1545–54.
[28] Betancourt MR, Thirumalai D. J Phys Chem B 2002;106:599–609.
[29] Li MS, Klimov DK, Thirumalai D. Comput Phys Commun 2002;147:625–8.
[30] Li MS, Klimov DK, Thirumalai D. J Phys Chem B 2002;106:8302–5.
[31] Hilhorst HJ, Deutch JM. J Chem Phys 1975;63:5153–61.
[32] Ferrenberg AM, Swendsen RH. Phys Rev Lett 1989;63:1195–8.
[33] Li M S, Klimov DK, Thirumalai D. Unpublished results.
[34] Nymeyer H, Garcia AE, Onuchic JN. Proc Natl Acad Sci USA 1998;95:5921–6.
[35] Klimov DK, Thirumalai D. J Chem Phys 1998;109:4119–25.
[36] Koga N, Takada S. J Mol Biol 2001;313:171–80.
[37] Klimov DK, Thirumalai D. Phys Rev Lett 1997;79:317–20.
[38] Socci ND, Onuchic JN, Wolynes PG. J Chem Phys 1996;104:5860–8.
[39] Portman JJ, Takada S, Wolynes PG. J Chem Phys 2001;114:5069–81.
[40] Hagen SJ, Hofrichter J, Szabo A, Eaton AW. Proc Natl Acad Sci USA 1996;93:11615–9.
[41] Munoz V, Eaton WA. Proc Natl Acad Sci USA 1999;96:11311–6.
[42] Guo Z, Thirumalai D. Biopolymers 1995;36:83–102.
[43] Yang WY, Gruebele M. Nature 2003;423:193–7.
[44] Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D. Nat Strut Biol 1999;6:554–6.
[45] Jacob M, Schindler T, Balbach J, Schmid FX. Proc Natl Acad Sci USA 1997;94:5622–7.