Specific and Nonspecific Collapse in Protein Folding Funnels

Jorge Chahine,[1] Hugh Nymeyer,[2] Vitor B. P. Leite,[1] Nicholas D. Socci,[3] and José Nelson Onuchic[2]

[1]*Departmento de Física, IBILCE, Universidade Estadual Paulista, S. J. Rio Preto, Brasil 15054-000*
[2]*Department of Physics, University of California, San Diego, California 92093*
[3]*Center for Studies in Physics and Biology, The Rockefeller University, New York, New York 10021*
and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

Experiments with fast folding proteins are beginning to address the relationship between collapse and folding. We investigate how different scenarios for folding can arise depending on whether the folding and collapse transitions are concurrent or whether a nonspecific collapse precedes folding. Many earlier studies have focused on the limit in which collapse is fast compared to the folding time; in this work we focus on the opposite limit where, at the folding temperature, collapse and folding occur simultaneously. Real proteins exist in both of these limits. The folding mechanism varies substantially in these two regimes. In the regime of concurrent folding and collapse, *nonspecific* collapse now occurs at a temperature below the folding temperature (but slightly above the glass transition temperature).

One intriguing property of proteins is their ability to self-assemble into specific, functional native structures. The energy landscape theory and the folding funnel concept have provided a framework for distinguishing heteropolymers which easily fold without help (weakly frustrated) and those that have difficulty folding alone (strongly frustrated) [1–9].

Bryngelson and Wolynes proposed that one way to measure the frustration (and hence foldability) of different sequences is through the ratio of two temperatures: the folding (denaturation) temperature T_f and the glass temperature T_g [1,10,11].

There has been much debate over the role of collapse and its relationship to frustration and foldability. In this Letter, we focus on folding mechanisms that apply to small two-state fast folding proteins, although the main ideas may be generalized to proteins with more complex folding mechanisms. While some proteins appear to collapse rapidly to a partially compact state which has some residual native structure but is mostly disordered, other proteins collapse and form native structures concurrently. This diversity has been observed in all atom simulations and experimentally. Simulations of the protein GB1 [12] show rapid collapse in the early stages of folding to structures with nearly nativelike volumes but with only 35% of the native interactions. In contrast, simulations of fragment *B* of protein *A* [13] show a reduction of the radius of gyration concurrent with the formation of native contacts. Similar mechanisms can now be seen experimentally. Single molecule fluorescence resonance energy transfer (FRET) studies on *cold-shock protein* show no observable equilibrium collapse transition in the non-native population as the denaturant is titrated out until well below the folding-unfolding transition [14]. Time-resolved small-angle x-ray scattering studies of *protein L* suggest that it can also fold directly from a noncompact unfolded

state to the native without passing through a compact phase [15]. On the other hand, proteins which fold by first passing through a partially hydrophobically driven compact phase are numerous and well known [16].

In this Letter, we study a simple model of proteins in which the variation from concurrent collapse and folding to collapse preceding folding can be controlled. We find that when collapse and folding are concurrent, a richer behavior results through the competition of transiently populated collapsed states and the coil-like unfolded state. Because of this more complex behavior, measures of foldability based on collapse temperatures [17] have only limited validity for proteins with concurrent folding and collapse. T_f/T_g remains an effective measure of foldability in all limits.

The term "nonspecific collapse" is used to describe collapsed states which are non-native-like. When collapse and folding are concurrent, the temperature at which nonspecific collapse occurs T_{nsz} lies below the folding temperature T_f instead of above it. Recall that even this nonspecific collapse ensemble still has some remnants of the native structure. In this case the nonspecific collapsed state is metastable and only weakly populated, however, collapsed conformations can still be transiently populated and act as off-pathway folding traps.

Evidence for nonspecific collapse.—Previous studies [18,19] have examined the thermodynamic properties of various sequences modeled by a 27 length polymer chain (27-mer) on a three-dimensional cubic lattice. In this manuscript we extend these studies, using the same model, to sequences with a variable nonspecific collapse parameter. The energy for the heteropolymer is given by

$$E = N_l E_l + N_u E_u, \qquad (1)$$

where E_l is the nonbonded contact energy between monomers of the same type and E_u is the energy between

monomers of different types. N_l is the number of contacts between monomers of similar type, and N_u is the number of contacts between monomers of dissimilar types. For a given chain in a random collapsed state with m different types of monomers, the average (nonspecific) contact energy between nonbonded monomers is approximately

$$\overline{E}_{\mathrm{ns}} \cong \sum_{i=1}^{m} [f_i^2 E_l + f_i(1 - f_i)E_u], \qquad (2)$$

where f_i is the occurrence of monomer type i in the heteropolymer chain, and m is the number of monomer types. The dispersion in contact energies, Δ, is approximately

$$\Delta \cong \left\{ \sum_{i=1}^{m} [f_i^2(E_l - \overline{E}_{\mathrm{ns}})^2 \right.$$
$$\left. + f_i(1 - f_i)(E_u - \overline{E}_{\mathrm{ns}})^2] \right\}^{1/2}, \qquad (3)$$

which is associated with the roughness of the energy landscape. The ratio

$$\kappa = -\frac{\overline{E}_{\mathrm{ns}}}{\Delta} \qquad (4)$$

is the collapse parameter. There are only two independent scaled parameters: T/Δ and κ, namely, a scaled temperature and a parameter measuring the average strength of the attractive interaction between monomers (both scaled by the energetic variance Δ). Collapse precedes folding when $\kappa \gtrsim 0.5$; collapse and folding are concurrent when $\kappa \approx 0$. Temperature is in units of the monomer interaction energy, which is set such that relevant transitions occur at about a temperature ≈ 1.

Monte Carlo sampling with a local move set [18] is used to determine the density of states. The density of states Ω is a function of energy (E), number of native nonbonded contacts (Q), and the total number of nonbonded contacts Z. For this 27-mer model, the largest possible value for Z and Q is 28. The probability of finding the system in a state with any given value of E, Q, and Z is then

$$P(E, Q, Z) = \frac{\Omega(E, Q, Z)e^{-E/T}}{\sum \Omega(E, Q, Z)e^{-E/T}}. \qquad (5)$$

We used the designed sequence ABABBBCBAC-BABABACACBACAACAB which has a unique (aside from symmetries), maximally compact native state. Since we are interested in the concurrent folding and collapse regime, studies have been performed for $\kappa \approx 0$. (All $\kappa \approx 0$ simulations reported were made with $E_l = -3$ and $E_u = +1.6$ in arbitrary energy units. This choice of units provides typical transition temperatures about 1.) We consider folding and nonspecific collapse based on Z and Q criteria. The specific and nonspecific collapsed states can be distinguished by filtering out all the conformations of the designed sequence with large values of Q. For example, filtering out the states with $Q \geq 4$ (more than 14% of the native contacts) leaves states with the same distribution of collapse as other typical random (unde-

signed) sequences which have identical compositions. These results depend only weakly on the cutoff in Q used for filtering states (for cutoffs below $Q = 10$).

The effects of filtering on the specific heat can be seen in Fig. 1. Filtering shows the existence of a broad specific heat peak centered below that of the folding transition. This peak is centered at a temperature of $T/\Delta \approx 0.64$, below the folding transition peak at $T/\Delta \approx 0.89$. Without filtering, the nonspecific collapse peak is obscured by the "latent heat" of folding. The increase in the collapse order parameter around this temperature indicates the origin of this peak. To confirm this, we computed the specific heat for a random sequence with the same composition as the native sequence (ABCAACCBAACBAAAACBBA-CBBBBBA). This random sequence does not have the folding specific heat peak, but it has a peak in the same location as the nonspecific collapse peak. The width of these peaks suggests this transition is less cooperative and more second-order-like than the folding transition. Because the nonspecific collapse temperature T_{nsz} lies below T_f, this phase has a higher free energy than the native state and is only metastable. However, when folding is occurring at temperatures below T_{nsz}, this phase is more stable than the random-coil phase and more accessible than the native state, so it can still affect the folding behavior. The existence of this transition can be understood as the continuation of the collapse/random-coil
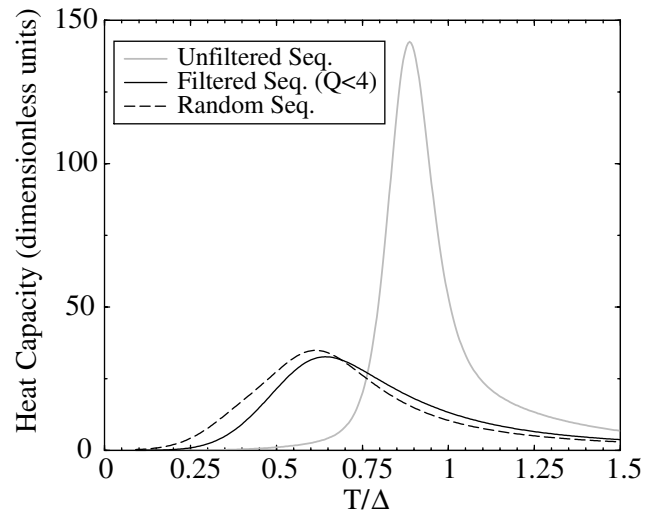

FIG. 1. The heat capacity as a function of temperature in the limit of concurrent collapse and folding ($\kappa \approx 0$). For the designed sequence without the filtering procedure, it shows the peak due to the latent heat folding transition. For the filtered states with $Q < 4$, the dominant effect of the latent heat of folding is removed, revealing a smaller and broader specific heat peak at $T = 1.30$ due to nonspecific collapse. The specific heat of a randomly shuffled sequence, which has the same composition as our native sequence, shows a broad peak at about the same location as the filtered native sequence, suggesting the similarity of the collapse transition in the randomized sequence and nonspecific collapse in the native sequence.

phase boundary below the folding transition boundary as the average monomer-monomer interaction becomes less attractive. Around the point where κ falls below ≈ 0.7, the T_{nsz} boundary will fall below the T_f boundary (see Fig. 2). The phase diagram is sequence dependent, i.e., the area of the folded region will depend on how well designed the sequence is. For random sequences, where $T_f < T_g$, the folded region will not exist, and the folded region in Fig. 2 corresponds to a collapsed frozen phase.

Kinetic effects of nonspecific collapse.—Below T_f but above T_{nsz}, few unfolded conformations undergo nonspecific collapse, so the relaxation behavior is determined by transitions between native and noncompact unfolded states, i.e., two phases. Below T_{nsz} three phases are involved: native, noncompact unfolded, and compact unfolded. Around these temperatures the energy landscape is sufficiently rough so that folding occurs primarily from the noncompact states, thus the compact unfolded states behave as traps to slow folding and introduce multiexponential kinetics in the folding process. There has been much debate about glassy behavior in proteins, and several papers have addressed the relationship between kinetics and glassy behavior [20–22]; however, it is clear that glasslike behavior cannot occur without significant collapse [23], hence, any T_g must occur below T_{nsz}. T_{nsz} appears to be a good estimator of the onset of nonexponential kinetics for small values of κ.

We measured first passage times to the native state for the designed sequence in the $\kappa \approx 0$ limit starting from random-coil configurations. Above $T/\Delta \approx 0.68$, the fraction of the initial population which has not yet visited the native state decays exponentially with time. The

degree of deviation from exponentiality is quantified by fitting the uncollapsed population to a stretched exponential, $\exp[-(t/\tau)^\beta]$. Figure 3 shows the onset of nonexponential kinetics at $T/\Delta \approx 0.68$, which agrees with T_{nsz} determined from the nonspecific collapse peak of Fig. 1. Transition from single exponential to stretched exponential relaxation in time with reduction of temperature has been observed in yeast *phosphoglycerate kinase* and a *ubiquitin* mutant [24]. We should, however, be cautious when interpreting these results, since they may be due to other effects such as diffusion on a barrier-free landscape.

Figure 4 shows the dependence of the collapse time τ versus $1/T$ for both the designed and randomized sequences and for a variety of values of κ. When collapse precedes folding ($\kappa = 2$), collapse is sequence independent, i.e., the curves for both the designed and randomized sequences are identical. The collapse times are sequence independent because the nonspecific collapse temperature T_{nsz} is larger than T_f for large κ, thus quenching from high temperature almost always results in nonspecific collapse occurring before folding. However, when collapse and folding are concurrent ($\kappa \approx 0$), collapse is sequence dependent for temperatures above T_{nsz}. This is expected since collapse is dominated by the native basin in this regime. Below T_{nsz} both the designed and randomized sequences collapse to non-native-like structures before folding, so their collapse times are similar. For negative values of κ, most of the random sequence runs do not collapse during the allowed time frame (10^9 Monte Carlo steps), as expected from Fig. 2.
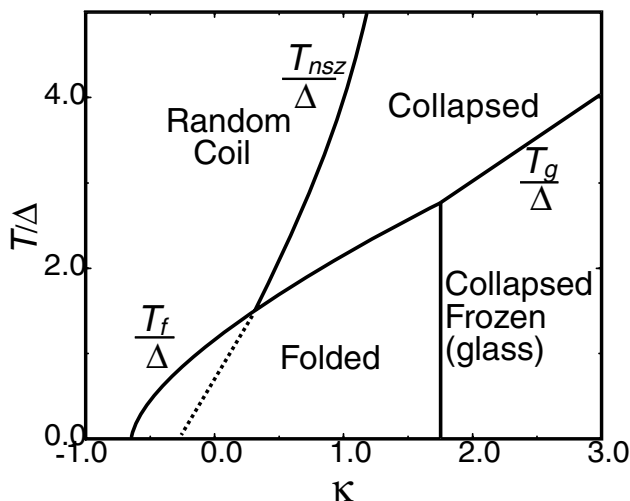

FIG. 2. Illustration of how, for a well-designed sequence, the nonspecific collapse and folding phase boundaries vary with our two system parameters: the effective temperature and the parameter measuring the amount of nonspecific attraction. The exact crossing point of these curves is not known, but it lies somewhere between 0 and 1.
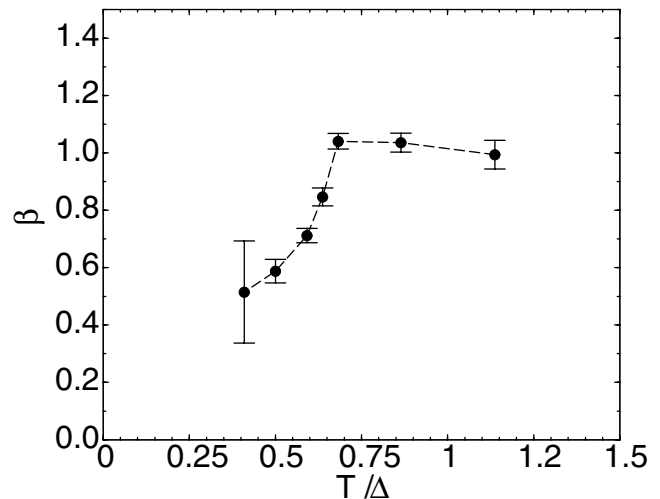

FIG. 3. The exponent β in a stretched exponential fit, $\exp[-(t/\tau)^\beta]$, to the population which has not visited the native state versus temperature. This is shown for the designed sequence with $\kappa \approx 0$. The transition from exponential relaxation to nonexponential relaxation occurs at the same temperature at which collapse times become sequence specific, as shown in Fig. 4. Error bars show 68% confidence limits estimated from ten bootstrapping runs.
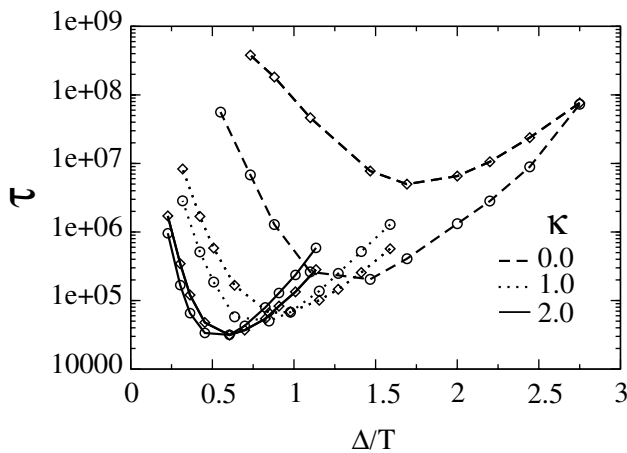
FIG. 4. Characteristic collapse times τ as a function of inverse temperature for different parameters κ. Values of (E_l, E_u) used are $(-3.0, 1.6)$, $(-3.0, -0.34)$, and $(-3.0, -1.1)$. $\circ$ is the designed sequence result and $\diamond$ is the randomized one.

Conclusion.—The ratio of the folding to glass transition temperatures T_f/T_g is believed to correlate well with folding rates. The parameter $\sigma = (T_z - T_f)/T_z$, where T_z and T_f are the collapse and folding temperatures, respectively, has been shown to work as an alternate measure of foldability [17]. This makes sense since σ is an approximate measure of the gap [25], which is the stability in units of the energy width of the molten globule, itself an approximate measure of T_f/T_g. It is usually maintained [18] that T_z and T_g depend strongly on sequence composition but only weakly on the specific sequence. We show in this paper that T_z can be itself sequence specific when collapse and folding are concurrent. These two different measures for foldability strongly correlate as long as T_z is larger than T_f, but, when κ is small, $T_z \approx T_f$ and therefore σ does not constitute a sensitive measure of foldability. In this regime, information about the nonspecific collapse temperature T_{nsz} or the glass temperature T_g is needed to differentiate sequences.

Landscape theory has provided a general framework for understanding folding, but new experiments are pointing out the diversity of folding behavior described within this framework. For example, experiments have demonstrated that folding can occur directly from random-coil states without a preceding collapse. The results in this paper suggest the existence of a temperature below the folding temperature at which the dynamics of such proteins can change dramatically. The details of the phase diagram in Fig. 2 can be used to guide experiments to observe such behaviors in real proteins.

[1] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[2] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **89**, 8721 (1992).
[3] C. J. Camacho and D. Thirumalai, Phys. Rev. Lett. **71**, 2505 (1993).
[4] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich, J. Chem. Phys. **101**, 6052 (1994).
[5] P. G. Wolynes, Z. L. Schulten, and J. Onuchic, Chem. Biol. **3**, 425 (1996).
[6] L. A. Mirny, V. Abkevich, and E. I. Shakhnovich, Folding Des. **1**, 103 (1996).
[7] K. A. Dill and H. S. Chan, Nat. Struct. Biol. **4**, 10 (1997).
[8] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[9] J. N. Onuchic *et al.,* Adv. Protein Chem. **53**, 87 (2000).
[10] J. D. Bryngelson and P. G. Wolynes, J. Phys. C **93**, 6902 (1989).
[11] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Protein Struct. Funct. Gen. **21**, 167 (1995).
[12] F. B. Sheinerman and C. L. Brooks, III, Proc. Natl. Acad. Sci. U.S.A. **95**, 1562 (1998).
[13] E. M. Boczko and C. L. Brooks, Science **269**, 393 (1995).
[14] B. Schuler, E. A. Lipman, O. Bakajin, and W. A. Eaton, report, 2002 (to be published).
[15] K. W. Plaxco *et al.,* Nat. Struct. Biol. **6**, 554 (1999).
[16] O. B. Ptitsyn, Adv. Protein Chem. **47**, 83 (1995).
[17] D. K. Klimov and D. Thirumalai, Protein Struct. Funct. Gen. **26**, 411 (1996).
[18] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[19] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).
[20] H. Nymeyer, A. E. Garcia, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **95**, 5921 (1998).
[21] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Protein Struct. Funct. Gen. **32**, 136 (1998).
[22] S. S. Plotkin, J. Wang, and P. G. Wolynes, J. Chem. Phys. **106**, 2932 (1997).
[23] J. D. Bryngelson and P. G. Wolynes, Biopolymers **30**, 177 (1990).
[24] J. Sabelko, J. Ervin, and M. Gruebele, Proc. Natl. Acad. Sci. U.S.A. **96**, 6031 (1999).
[25] A. R. Dinner, V. Abkevich, E. Shakhnovich, and M. Karplus, Protein Struct. Funct Gen. **35**, 34 (1999).