Structural Basis of Folding Cooperativity in Model Proteins: Insights from a Microcanonical Perspective

Tristan Bereau,[†] Markus Deserno,[†] and Michael Bachmann[‡*]
[†]Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania; and [‡]Center for Simulational Physics, Department of Physics and Astronomy, University of Georgia, Athens, Georgia

ABSTRACT Two-state cooperativity is an important characteristic in protein folding. It is defined by a depletion of states that lie energetically between folded and unfolded conformations. There are different ways to test for two-state cooperativity; however, most of these approaches probe indirect proxies of this depletion. Generalized-ensemble computer simulations allow us to unambiguously identify this transition by a microcanonical analysis on the basis of the density of states. Here, we present a detailed characterization of several helical peptides obtained by coarse-grained simulations. The level of resolution of the coarse-grained model allowed to study realistic structures ranging from small α-helices to a de novo three-helix bundle without biasing the force field toward the native state of the protein. By linking thermodynamic and structural features, we are able to show that whereas short α-helices exhibit two-state cooperativity, the type of transition changes for longer chain lengths because the chain forms multiple helix nucleation sites, stabilizing a significant population of intermediate states. The helix bundle exhibits signs of two-state cooperativity owing to favorable helix-helix interactions, as predicted from theoretical models. A detailed analysis of secondary and tertiary structure formation fits well into the framework of several folding mechanisms and confirms features that up to now have been observed only in lattice models.

INTRODUCTION

Two-state protein folding is characterized by a single free-energy barrier between folded and unfolded conformations at the transition temperature T_c, whereas downhill folders do not exhibit folding barriers (1,2). The analysis of this property conveys important information on both the thermodynamics and the kinetic pathways of proteins (2,3). A widely used test for a two-state transition is the calorimetric criterion, which probes features in the canonical specific heat curve (4). However, this criterion does not provide a sufficient condition to identify two-state transitions (5), and makes no clear distinction between weakly two-state and downhill folders. Other experimentally observable aspects of two-state cooperativity include sharp transitions in certain order parameters, or features in chevron plots (3,6). All these methods focus on the thermodynamic consequences of a depletion of intermediate states, rather than studying it directly.

However, it is possible to determine the density of states in a standard canonical computer simulation at temperature T^* of interest by sampling the probability density $p(E)$ of finding an energy E. The density of states $\Omega(E)$ is then proportional to $p(E)\mathrm{e}^{E/k_\mathrm{B}T^*}$, and hence the entropy is (up to a constant) given by $S(E) = k_\mathrm{B} \ln \Omega(E) = \text{const.} + k_\mathrm{B} \ln p(E) + E/T^*$. We can then proceed to analyze the system microcanonically, i.e., by studying the thermodynamics of $S(E)$ in the neighborhood of $\langle E \rangle_{T^*}$. This is advantageous because essentially we can directly analyze the probability density $p(E)$ instead of merely looking at its lowest moments, such as the specific heat. Such microcanonical analyses have been applied to a wide variety of problems, including spin systems (7–13), nuclear fragmentation (14,15), colloids (16), gravitating systems (17,18), off-lattice homo- and heteropolymer models (19,20), and protein folding (21–27). Two remarks are worthwhile:

1. If the transition is characterized by a substantial barrier, standard canonical sampling suffers from the usual getting-stuck problem. During a simulation, the system may not cross the barrier often enough to equilibrate the two coexisting ensembles. This must be avoided regardless of whether one is performing a canonical or microcanonical analysis. Investigators have proposed many ways to get around this problem, including multicanonical (28) or Wang-Landau (29) sampling. In our study, we employed replica-exchange molecular dynamics (MD) to sample coupled canonical ensembles (33) and combined the overlapping energy histograms by means of the weighted histogram analysis method (WHAM) (30–32), a minimum variance estimator for $\Omega(E)$.
2. Accurately sampling the whole distribution $p(E)$ over some range of interest requires better statistics than is needed for merely sampling its lowest moments; there is a price for higher-quality data. A microcanonical analysis can tap into this quality, whereas a canonical analysis of the much longer simulation run will not significantly improve the observables. Recall that the canonical partition function $Z(T) = \int \mathrm{d}E\, \Omega(E)\mathrm{e}^{-E/k_\mathrm{B}T}$ is the Laplace transform of the density of states $\Omega(E)$, an operation that is well known to be strongly smoothing and thus difficult to invert.

Submitted December 9, 2010, and accepted for publication March 31, 2011.

*Correspondence: bachmann@smsyslab.org

Editor: Nathan Andrew Baker.

0006-3495/11/06/2764/9 $2.00

doi: 10.1016/j.bpj.2011.03.056

From a thermodynamic standpoint, a two-state transition is characterized by two coexisting ensembles of conformations (6). Although this does not qualify as a genuine (first-order) phase transition (because the free energy of finite systems is always analytical), we can unambiguously characterize its finite-size equivalent by monitoring the entropy $S(E)$. In the phase-coexistence region, it will exhibit a convex intruder due to the suppression of states of intermediate energy. This can best be observed by defining the quantity $\Delta S(E) = \mathcal{H}(E) - S(E)$, where $\mathcal{H}(E)$ corresponds to the (double-)tangent to $S(E)$ in the transition region (8–10,23,24). In a finite system, the existence of a barrier in $\Delta S(E)$ will imply a nonzero microcanonical latent heat ΔQ, defined by the interval over which $S(E)$ departs from its convex hull, and in turn leads to a backbending effect (akin to a van der Waals loop) in the inverse microcanonical temperature $T_{\mu c}^{-1}(E) = \partial S/\partial E$ (8–10,12,13,23). A nonzero ΔQ demarcates a transition region, whereas a downhill folder (continuous transition) will only exhibit a transition point at which the concavity of $S(E)$ is minimal.

Extending a recent study (27), we focus here on the link between 1), the nature of the transition (i.e., two-state versus downhill); 2), secondary structure; and 3), tertiary structure formation for several helical peptides using a high-resolution, implicit-solvent, coarse-grained (CG) model. The results are interpreted in terms of different frameworks of folding mechanisms, such as the molten-globule model and simple polymer-collapse models (34,35). Although all of the helical peptides presented in this work were artificially constructed (de novo), and thus have not naturally evolved, they exhibit the relevant physics in a particularly clean way and thus provide useful model systems. (See Supporting Material for a further discussion of this point.)

MATERIALS AND METHODS

We performed CG-MD simulations based on an intermediate resolution, implicit-solvent peptide model (36). This model accounts for amino acid specificity and is capable of representing genuine secondary structure without explicitly biasing the force field toward any particular conformation (native or not). Table 1 lists the sequences of all of the studied peptides. More details can be found in the Supporting Material.

We performed replica-exchange MD simulations using the ESPResSo package (37), and ran all simulations in the canonical (NVT) ensemble using a Langevin thermostat with friction constant $\Gamma = \tau^{-1}$, where τ is the intrinsic unit of time of the CG model. The CG unit of energy, $\mathcal{E}$, relates to thermal energy at room temperature via $\mathcal{E} = k_B T_{\text{room}} = 1.38 \times 10^{-23}\text{JK}^{-1} \times 300\text{K} \approx 0.6\ \text{kcalmol}^{-1}$. The temperature T was expressed in terms of the intrinsic unit of energy $[T] = \mathcal{E}/k_B$. The equations of motion were integrated with a time step $\delta t = 0.01\tau$.

Entropy, order parameters, and canonical averages were obtained from the density of states, $\Omega(E)$, which itself was calculated from WHAM (30–32). Details can again be found in the Supporting Material.

Finally, the reader should observe that CG force fields, including the one used here, are usually constructed to reproduce the canonical ensemble, and hence they strive to reproduce the free energy. However, individual enthalpic and entropic contributions will generally be off, because the reduced number of degrees of freedom lowers the entropy of CG conformations, and thus the energies must be adjusted to leave the free energy correct. For instance, in the absence of solvent, both solvent energy and entropy must be parametrized into effective solute interaction energies. Therefore, the entropies we calculate in this work are not to be confused with the entropies of the actual system. On the other hand, this does not preclude them from being exquisitely sensitive observables for the thermodynamics of the CG model.

TABLE 1 Amino acid sequences of the peptides studied in this work

Peptide	Sequence
helix $n = 3$	$(\text{AAQAA})_3$
helix $n = 7$	$(\text{AAQAA})_7$
helix $n = 10$	$(\text{AAQAA})_{10}$
helix $n = 15$	$(\text{AAQAA})_{15}$
bundle α3D	MGSWA EFKQR LAAIK TRLQA LGGSE …
	AELAA FEKEI AAFES ELQAY KGKGN …
	PEVEA LRKEA AAIRD ELQAY RHN

The three helical regions of the native state (from NMR structure, PDB 2A3D) of the helix bundle α3D (38) are underlined (as predicted by STRIDE (39)).

RESULTS

Secondary structure

We first examine the structural and energetic properties of the sequence $(\text{AAQAA})_n$ with various chain lengths ($n = 3, 7, 10, 15$). The $n = 3$ variant is known as a stable α-helix folder and has been studied both experimentally and computationally (40–44). The $n = 7$ peptide has also been shown to fold into a helix (42). We find that all four peptides form a stable, long helix in the lowest-energy sector (see below), but we are not aware of any structural study that would confirm this for the longer peptides with $n = 10,\ 15$. Since we will soon show that the latter two fold differently from the shorter ones, an experimental confirmation of their ground state structure would be very useful.

For $(\text{AAQAA})_3$, Fig. 1 *a* shows a barrier in $\Delta S(E)$ as well as a backbending in the inverse microcanonical temperature $T_{\mu c}^{-1}(E)$, indicative of a first-order like transition. The two vertical lines mark the transition region with the corresponding microcanonical latent heat ΔQ. In the region between $E = (40 - 80)\mathcal{E}$, mostly helical and mostly coil conformations coexist, in agreement with the sharp transitions in the helicity $\theta(E)$ (as determined by the STRIDE algorithm (39)) and the number of helices in the chain, $H(E)$. These results point to a clear two-state folder.

Increasing the chain length from $n = 3$ to $n = 15$ (Fig. 1, *b–d*) changes the nature of the transition significantly. Whereas $n = 7$ still shows a (lower) barrier in $\Delta S(E)$ and a nonzero microcanonical latent heat ΔQ, the cases $n = 10$ and $n = 15$ are downhill folders (no barrier in $\Delta S(E)$ and monotonic $T_{\mu c}^{-1}(E)$ curves). The transition region is replaced by a transition point for which the concavity of $S(E)$ is minimal and $\Delta Q = 0$. This process is associated with important structural changes around the transition region/point, as seen in the number of helices

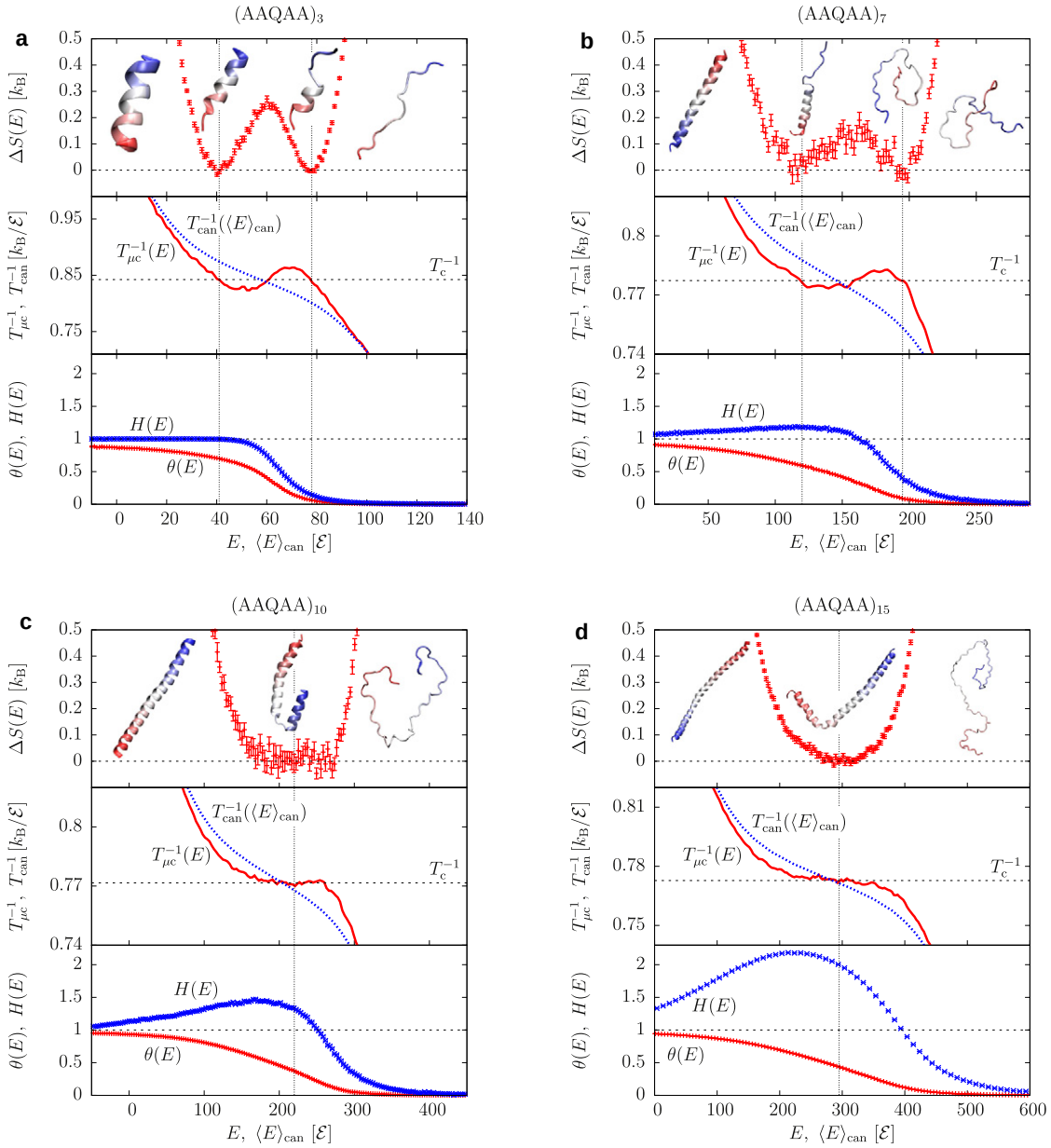

FIGURE 1 Various observables as a function of energy for $(AAQAA)_n$: (*a*) $n = 3$, (*b*) $n = 7$, (*c*) $n = 10$, and (*d*) $n = 15$. From top to bottom for each inset: $\Delta S(E)$, error bars reflect the variance of the data points (1σ interval); inverse temperatures from a canonical ($T_{can}^{-1}(\langle E_{can}\rangle$, *blue*) and a microcanonical ($T_{\mu c}^{-1} = \partial S/\partial E$, *red*) analysis, where E_{can} is the canonical average energy; helicity $\theta(E)$ (*red*) and number of helices $H(E)$ (*blue*), both with the error of the mean. Vertical lines mark either the transition region ($n = 3, 7$) or the transition point ($n = 10, 15$). Representative conformations at different energies, visualized with the use of VMD (45), are shown.

$H(E)$: although the curve is monotonic for $n = 3$, it exhibits a peak with $H(E)$>1 for larger n, showing that during the transition most conformations form more than one helix. This suggests the existence of multiple helix nucleation sites upon folding (see representative conformations at the transition point for $n = 10,\ 15$ in Fig. 1).

To further elucidate the structural features of these chains around the transition region/point, we analyzed the fraction of secondary structure (i.e., helicity) in dependence of both energy and the residue index for helices $n = 3,\ 15$. For $n = 3$, helix nucleation appears mostly around the center of the peptide and propagates symmetrically to the termini (Fig. 2 *a*); however, $n = 15$ shows two distinct peaks at an energy E slightly below the transition point (Fig. 2 *b*). The results suggest the formation of two individual helices placed symmetrically from the midpoint of the chain, around residue 35, which only join into one long helix significantly below the transition point. As discussed further below, these two helices divide the system into two distinct melting domains that fold noncooperatively (i.e., folding

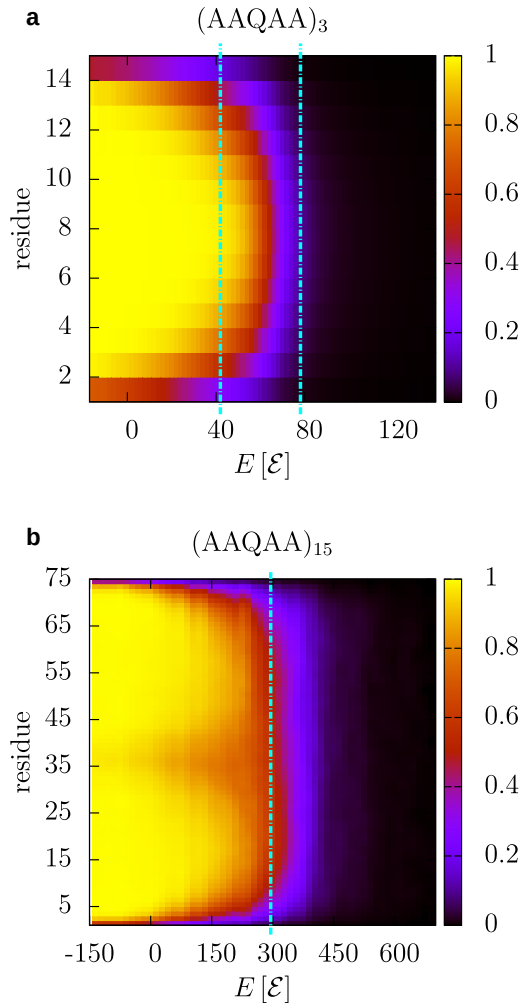

FIGURE 2 Fraction of secondary structure as a function of energy and residue for (*a*) $(AAQAA)_3$ and (*b*) $(AAQAA)_{15}$. Vertical lines mark the transition region (*a*) and point (*b*).

one helix does not help fold the other) (46,47). The same conclusion can be drawn from the probability distributions of forming an *m*-helix (see Supporting Material).

To probe the behavior of simultaneous folding motifs within a chain, we performed a microcanonical analysis of the 73 residue de novo three-helix bundle α3D (38) (for the amino acid sequence, see Table 1). The CG model used here has been shown to fold α3D with the correct native structure, up to a root mean-square deviation of 4Å from the NMR structure (36). Although its length is similar to that of $(AAQAA)_{15}$, it shows a discontinuous transition (see Fig. 3) and thus a nonzero microcanonical latent heat during folding. In the transition region, the helicity increases sharply from 20% to ~65%, and the average number of helices also increases sharply but monotonically from 1.5 to 3. In contrast to the simple $n = 7, 10, 15$ helices, the transition region never samples more helix nucleation sites than the number of helices at low energies. As can be seen from the representative conformations shown in Fig. 3, the

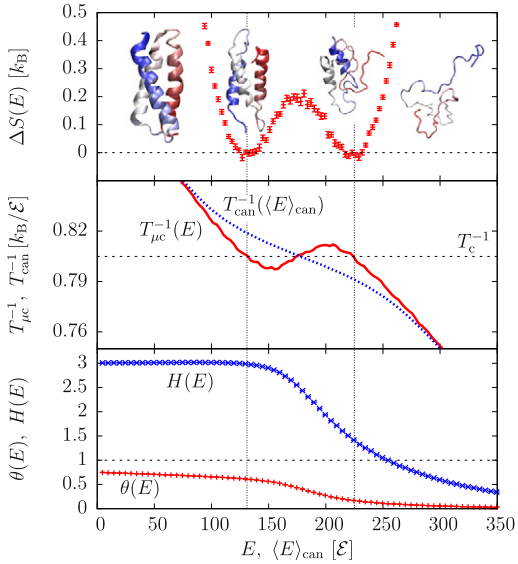

FIGURE 3 Various observables as a function of energy for α3D. Plots and definitions agree with the conventions in Fig. 1.

ensemble of folded states ($E \approx 130\,\mathcal{E}$) consists of three partially formed helices in largely native chain topology, and the coexisting unfolded ensemble ($E \approx 225\,\mathcal{E}$) consists of a compact structure containing transient helices. All of these findings identify α3D as a two-state folder.

To better monitor the formation of individual helices, we measured the fraction of helicity as a function of energy and residue (see Fig. 4). Unlike $(AAQAA)_n$ (Fig. 2), α3D shows strong features due to its more interesting primary sequence. The turn regions (*dark color*) delimiting the three helices (*light color*) are clearly visible at low energies and correspond well to the STRIDE prediction of the NMR structure, as shown in Table 1. Moreover, it is clear from this figure that secondary structure formation occurs simultaneously (i.e., at the same energy) for all three helices, and that most of the folding takes place within the coexistence region

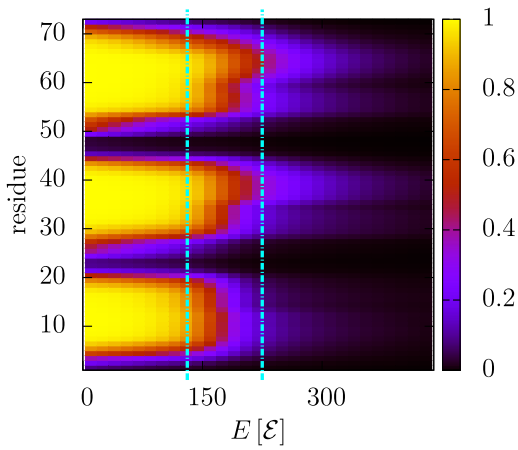

FIGURE 4 Fraction of secondary structure as a function of energy and residue for α3D. Vertical lines mark the transition region.

(marked by the two vertical lines). The residues that form the native turn regions do not show any statistically significant signal of helix formation at any energy. Secondary structure is formed almost entirely close to the folded ensemble in the transition region (*leftmost vertical line*), in line with the representative conformations shown in Fig. 3.

Tertiary structure

A secondary structure analysis alone can only provide information on the local aspects of folding. Several studies have highlighted the role of an interplay between local and nonlocal interactions in protein folding cooperativity (27,48–50). Here, we first analyze the size and shape of the overall molecule by monitoring the radius of gyration $R_g = \sqrt{\lambda_x^2 + \lambda_y^2 + \lambda_z^2}$ and the normalized acylindricity $c = (\lambda_x^2 + \lambda_y^2)/2\lambda_z^2$ as a function of E, expressed in terms of the three eigenvalues of the gyration tensor $\lambda_x^2 < \lambda_y^2 < \lambda_z^2$, respectively. The results for the single helices $n = 3$ and $n = 15$ and the three-helix bundle α3D are shown in Fig. 5. $(AAQAA)_3$ shows sharp features in both order parameters within the transition region, indicating an overall structural compaction (in shape and size) of the chain as energy is lowered. Observe that c approaches 0.13 at high energy, which is close to the random walk or self-avoiding walk values, both of which are close to $c \approx 0.15$ (51,52). The longer helix $n = 15$ shows a nonmonotonic behavior in both $R_g(E)$ and $c(E)$: whereas the radius of gyration exhibits a minimum around $E = 400\,\mathcal{E}$, the normalized acylindricity displays a maximum. This indicates a structure that is most compact and spherical 100 $\mathcal{E}$ above the transition point. This dip in $R_g(E)$ corresponds to a chain collapsing into maximally compact nonnative states (34) due to a nonspecific compaction of the chain gradually restricted by steric clashes, at which point secondary structure becomes favorable. When the energy is lowered, R_g increases and the acylindricity decreases, because the peptide elongates while folding from a compact globule into an α-helix. Results for the three-helix bundle are similar: $R_g(E)$ and $c(E)$ also show a minimum and a maximum, respectively, slightly above the transition region. This indicates a similar type of chain collapse mechanism. However, nonmonotonic features appear also at the other end of the transition region ($E \approx 130\,\mathcal{E}$) where R_g shows a maximum and the acylindricity plateaus. The evolution of the two order parameters below the transition region is rather limited, suggesting that only minor conformational changes take place (i.e., the shape of the molecule stays steady while its size decreases slightly). In contrast, at high energy, both $(AAQAA)_{15}$ and α3D are still far away from a random walk limit, as evidenced by the acylindricity being far away from 0.15.

We can readily observe chain collapse in longer chains (such as $(AAQAA)_{15}$ and α3D) by monitoring tertiary contacts as a function of energy. Fig. 6 shows the total number of nonlocal contacts (*red curve*) as well as the number of native contacts alone (*blue curve*). Tertiary contacts are defined here as pairs of residues that are more than five amino acids apart (which prevents chain connectivity artifacts) and within a 10Å distance (these numbers are somewhat arbitrary, but their value does not affect the qualitative behavior of Fig. 6). Native contacts correspond here to the set of abovementioned tertiary contacts sampled with a frequency higher than 1% from a set of 10,000

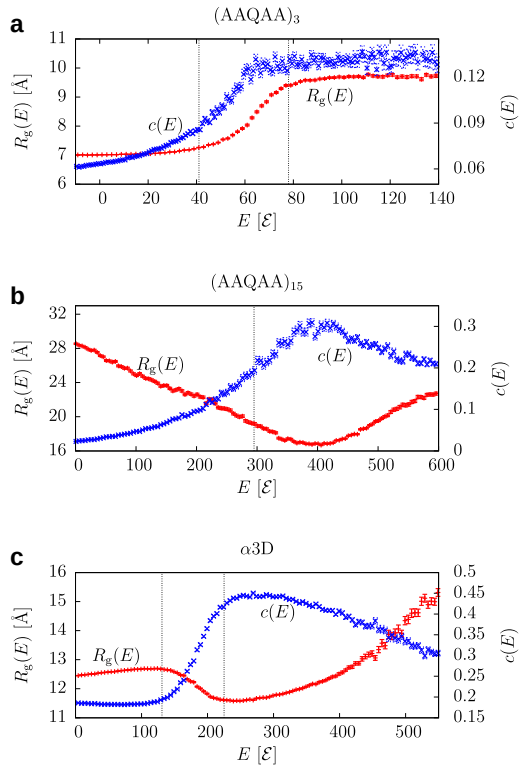

FIGURE 5 Radius of gyration $R_g(E)$ (*red*) and normalized acylindricity parameter $c(E)$ (*blue*), both with the error of the mean, for (*a*) $(AAQAA)_3$, (*b*) $(AAQAA)_{15}$, and (*c*) α3D. Vertical lines mark either the transition region ($n = 3$, α3D) or the transition point ($n = 15$).

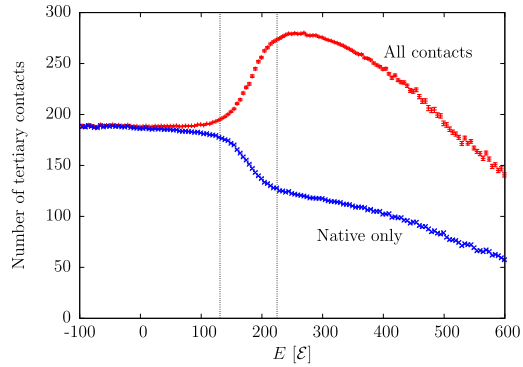

FIGURE 6 Number of tertiary contacts for α3D as a function of energy. The All-contacts curve (*red*) averages over all nonlocal pairs, whereas the Native-only curve (*blue*) counts only native pairs (see text for details). Vertical lines mark the transition region.

low-energy conformations ($E \leq 50\,\mathcal{E}$). Although the two curves are virtually identical below the transition region (i.e., all contacts are native) and show a similar trend above it, they behave very differently inside that interval. The number of native contacts monotonically increases as the energy is lowered (i.e., in the transition from a globule to a native-like structure), while the total number of contacts shows a peak above the transition region and sharply decreases inside it. To approach the native state, the peptide needs to break more contacts of the nonnative type than it gains native contacts.

The nonmonotonicity of this curve, as well as the R_g data, invite a comparison with the thermodynamics of water. Upon cooling, liquid water expands below 4°C. Weak but isotropic van der Waals interactions are given up for strong but directional hydrogen bonds. This energy/entropy balance seems to occur in a very similar manner here, and essentially for the same reason. Weak van der Waals side-chain interactions (i.e., tertiary contacts) are replaced by hydrogen-bond interactions (i.e., secondary structure) at lower energies. This further confirms the concept of a chain collapsing into maximally compact nonnative states: after a decrease in energy (above the transition region), the system accumulates a large number of nonnative contacts due to a simple hydrophobicity-driven compaction mechanism. This idea was proposed early on as the hydrophobic-collapse model or molten-globule model (34,35). Hills and Brooks (53) observed a similar effect by using a Gō model in which out-of-register contacts had to unfold to reach the native state.

Although a transient chain collapse upon cooling is present in both $(AAQAA)_{15}$ and α3D ($R_g(E)$ is nonmonotonic; see Fig. 5, *b* and *c*), its effect on tertiary structure formation will greatly depend on the amino acid sequence. Fig. 7 shows the number of tertiary contacts of the two peptides as a function of energy and residue. The single helix $n = 15$ shows a uniformly small number of tertiary contacts in the low-energy region (due to the linearity of the helix) and peaks above the transition point (which corresponds to the energy where $R_g(E)$ is smallest). The tertiary contact distribution in the maximally compact nonnative states is homogeneous along the chain (i.e., all residues have the same number of contacts). On the other hand, the number of tertiary contacts along the three-helix bundle (Fig. 7 *b*) is highly structured, forming stripes as a function of residue that extend below the transition region. This follows directly from the amphipathic nature of the subhelices that constitute α3D: residues that form the native hydrophobic core of the bundle have a higher number of contacts. The presence of these stripes in the energetic region of collapsed structures ($E \approx 300\,\mathcal{E}$) is due to a strong selection between hydrophobic and polar amino acids during the hydrophobic collapse, burying hydrophobic groups inside the globule. The low number of tertiary contacts in the turn regions indicates that they remain on the surface of the maximally compact globule during chain collapse.

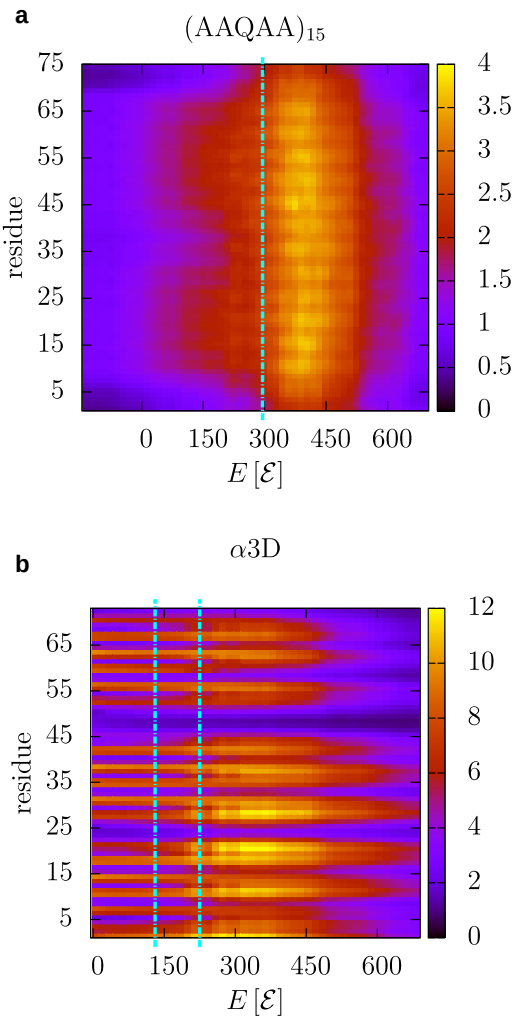

FIGURE 7 Number of tertiary contacts as a function of energy and residue for (*a*) $(AAQAA)_{15}$ and (*b*) α3D. Observe that the dynamic range of *b* is four times as wide as that for *a*. Vertical lines mark the transition point (*a*) and region (*b*).

DISCUSSION

Two-state cooperativity has been characterized as a common signature of small proteins for which the transition of the cooperative domain corresponds to the whole molecule (i.e., the protein undergoes a transition as a whole) (54). Although this framework applies well to the small helix $(AAQAA)_3$, it is difficult to predict its thermodynamic signature from other grounds. A description of the conventional helix-coil transition is not appropriate given the small size of the system and the correspondingly important finite-size effects.

The thermodynamic signature of proteins can better be described for longer chains. Several arguments can be brought forward to explain the transition we observe for the longer helices $(AAQAA)_n$ for $n = 10$, 15:

1. Most theoretical models of the helix-coil transition, such as the Zimm-Bragg model (55), are based on the

one-dimensional Ising model, which, being one-dimensional, shows no genuine phase transition but only a finite peak in the specific heat. The entropic gain of breaking a hydrogen bond (i.e., forming two unaligned spins) outweighs the associated energetic cost for a sufficiently long chain.
2. The structure of the maximally compact state right above the transition (see Fig. 7) indicates that there is no statistically significant competition between amino acids (i.e., all residues have the same number of tertiary contacts), and therefore it is associated with a homopolymer-type of collapse, which is indeed barrierless (34,56).
3. The denaturation of large proteins composed of several melting domains is not a two-state transition (46,47). The presence of two helix nucleation sites around the transition point (Fig. 2) indicates the existence of two such melting domains that fold noncooperatively (i.e., the folding of one helix is not correlated with the formation of the other). We ascertained that there were no statistically significant helix-helix interactions between the two domains by calculating contact maps. These were averaged over the ensemble of conformations for which $50 \leq E \leq 150\ \mathcal{E}$ (data not shown).

A common expectation is that larger systems will show sharper transition signals, and thus it may appear surprising that the transition of the $(AAQAA)_n$ sequence weakens for increasing n. However, one must bear two things in mind: First, size alone is not sufficient—dimensionality counts as well. In the Supporting Material we show examples of quasi-one-dimensional systems for which the transitions become weaker for larger systems because in the process of growing, they become more one-dimensional. When size is associated with cooperativity, one tends to think of globular (three-dimensional) systems, for which the size-cooperativity connection is true, but this is not the most general case. Second, the sharpness may depend on which observable is being studied. The helicity θ as a function of temperature indeed varies more sharply for larger n, making the response function $(\partial\theta/\partial T)_n$ peak more strongly for bigger n. This steepening would suggest a stronger two-state nature; however, this goes against every other observable, which suggests a downhill folder, including the calorimetric criterion (see below). Therefore, observing response functions alone can be misleading. More details on this can be found in the Supporting Material.

The two-state signature of the helix bundle α3D can be understood from two different perspectives:

1. Although there are clearly three distinct folding motifs (i.e., three helices), the selective hydrophobicity (i.e., amphipathic sequence) between residues provides cooperativity (i.e., the folding of one helix helps the formation of the others).
2. The barrier associated with a two-state transition is interpreted in the hydrophobic-collapse model as the result of the cost of breaking hydrophobic contacts from a maximally compact state into the folded ensemble (34). A further discussion on the order of appearance of secondary versus tertiary structure formation can be found in the Supporting Material.

Previous experimental studies of α3D showed a fast folding rate of $(1-5)\mu$s and single-exponential kinetics (57), compatible with a two-state cooperative transition. As presented here, this highlights the interplay between secondary structure formation (see Fig. 4) and the loss of nonnative tertiary contacts (see Fig. 6), both occurring exactly within the coexistence region, as a possible mechanism for folding cooperativity (27).

Compaction of the unfolded state upon an increase in temperature was observed experimentally by Nettels et al. (58) using single-molecule fluorescence resonance energy transfer. The decrease in R_g in our simulations can be explained by a combination of the hydrophobic effect and the loss of helical structure; however, Nettels et al. also showed a similar behavior for an intrinsically disordered hydrophilic protein, in which other mechanisms likely play a role.

So far, we have avoided any reference to free-energy barriers. Although the nature of the finite-size transition can unambiguously be characterized from the presence of a convex intruder in the entropy $S(E)$ (8), which implies a nonzero latent heat ΔQ, the mere existence of a free-energy barrier is not a strong criterion, for two reasons: 1), the definition of a free-energy barrier is not unique in a finite-size system (7,11); and 2), the height of the barrier depends on the reaction coordinate used. Chan (59) therefore argued that the calorimetric criterion, which relates the van't Hoff and calorimetric energies, is often more restrictive for protein models than the existence of such a free-energy barrier. Still, the density-of-states calculations performed here correlate well with calorimetric ratios for $(AAQAA)_n$, $n = \{3, 7, 10, 15\}$ and α3D: $\delta = 0.78$, 0.76, 0.51, 0.52 and 0.78, respectively. We determined these values by analyzing the canonical specific heat curve $C_V(T)$ as previously described by Kaya and Chan (4) (κ_2 without baseline subtraction). The value $\delta = 0.78$ for α3D also agrees with an earlier theoretical calculation of the similar bundle α3C from Ghosh and Dill (49), who found $\delta = 0.72$.

CONCLUSIONS

Replica-exchange MD simulations of an intermediate resolution CG implicit-solvent peptide model allowed us to accurately determine the thermodynamics of folding for several helical peptides without biasing the force field toward a particular native structure. We argued that a microcanonical analysis would be extremely valuable for characterizing the energetics and structure of peptides, for two reasons: First, an accurate density-of-states calculation would allow an unambiguous characterization of the nature

of the folding transition. Second, different order parameters, analyzed as a function of E, exhibit highly nonmonotonic behavior inside the (first-order-like) transition region. A corresponding canonical analysis (i.e., as a function of T) would not allow us to observe in such detail many of the abovementioned features around transition regions.

The results showed that simply elongating the $(AAQAA)_n$ sequence induced a change in the nature of the transition, from two-state ($n = 3, 7$) to downhill ($n = 10, 15$). This correlated with the number of helices sampled around the transition region/point, which is indicative of the average number of helix nucleation sites, and thus characterized the number of distinct melting domains and the structural diversity of intermediates. Remarkably, the loss of a first-order signature still goes along with a potentially misleading steepening of the helicity as a function of temperature for longer chains (see Supporting Material). The bundle α3D was found to be two-state cooperative, in agreement with theoretical models (49,50). The analysis of tertiary structure formation highlighted the influence of the amino acid sequence on the folding mechanism, with the hydrophobic-collapse model used as a starting point.

Although previous studies have brought forward the coupling between secondary and tertiary structure formation for two-state cooperativity (27,48–50), in this study we have illustrated several links between the nature of the transition and secondary/tertiary structure signatures of folding for realistic representations of peptide chains. Gaining a thorough understanding of structure formation in two-state cooperative proteins will provide insight into the stability of their folded conformation. Cooperativity improves the stability of the folded state by suppressing the population of intermediates. Mutations that lower cooperativity not only decrease stability, they can also promote misfolding in certain cases (60). The resolution of the CG model provides a useful compromise between computational efficiency and resolution, and allows us to access features that thus far have only been observed in less-realistic lattice models.

SUPPORTING MATERIAL

Additional text, references, and five figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)00414-0.

We thank K. Binder, W. Paul, B. Schuler, R.H. Swendsen, M. Taylor, and T.R. Weikl for stimulating discussions.

This work was partially supported by the National Institutes of Health (grant P01AG032131). M.B. received grants jiff39 and jiff43 from the Forschungszentrum Jülich for supercomputer time. T.B. received an Astrid and Bruce McWilliams Fellowship.

REFERENCES

1. Dobson, C. M., A. Šali, and M. Karplus. 1998. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed.* 37:868–893.
2. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.
3. Jackson, S. E. 1998. How do small single-domain proteins fold? *Fold. Des.* 3:R81–R91.
4. Kaya, H., and H. S. Chan. 2000. Polymer principles of protein calorimetric two-state cooperativity. *Proteins.* 40:637–661.
5. Zhou, Y. Q., C. K. Hall, and M. Karplus. 1999. The calorimetric criterion for a two-state process revisited. *Protein Sci.* 8:1064–1074.
6. Chan, H. S., S. Bromberg, and K. A. Dill. 1995. Models of cooperativity in protein folding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 348:61–70.
7. Borgs, C., and S. Kappler. 1992. Equal weight versus equal height: a numerical study of an asymmetric first-order transition. *Phys. Lett. A.* 171:37–42.
8. Gross, D. H. E. 2001. Microcanonical Thermodynamics: Phase Transitions in 'Small' Systems. World Scientific Publishing, Hackensack, NJ.
9. Hüller, A. 1994. First order phase transitions in the canonical and the microcanonical ensemble. *Zeit. Phys. B.* 93:401–405.
10. Deserno, M. 1997. Tricriticality and the Blume-Capel model: a Monte Carlo study within the microcanonical ensemble. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* 56:5204–5210.
11. Janke, W. 1998. Canonical versus microcanonical analysis of first-order phase transitions. *Nucl. Phys. B Proc. Suppl.* 63:631–633.
12. Hüller, A., and M. Pleimling. 2002. Microcanonical determination of the order parameter critical exponent. *Int. J. Mod. Phys. C.* 13:947–956.
13. Pleimling, M., and H. Behringer. 2005. Microcanonical analysis of small systems. *Phase Transit.* 78:787–797.
14. Gross, D. H. E. 1993. Multifragmentation, link between fission and the liquid-gas phase-transition. *Prog. Part. Nucl. Phys.* 30:155–164.
15. Koonin, S. E., and J. Randrup. 1987. Microcanonical simulation of nuclear disassembly. *Nucl. Phys. A.* 474:173–192.
16. Fernández, L. A., V. Martín-Mayor, …, P. Verrocchio. 2010. Separation and fractionation of order and disorder in highly polydisperse systems. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82:021501–021507.
17. Komatsu, N., S. Kimura, and T. Kiwata. 2009. Negative specific heat in self-gravitating N-body systems enclosed in a spherical container with reflecting walls. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 80:041107–041115.
18. Posch, H. A., and W. Thirring. 2006. Thermodynamic instability of a confined gas. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 74:051103–051108.
19. Chen, T., X. Lin, …, H. Liang. 2007. Microcanonical analysis of association of hydrophobic segments in a heteropolymer. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 76:046110–046113.
20. Taylor, M. P., W. Paul, and K. Binder. 2009. All-or-none proteinlike folding transition of a flexible homopolymer chain. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 79:050801–050804.
21. Hao, M. H., and H. A. Scheraga. 1994. Monte Carlo simulation of a first-order transition for protein folding. *J. Phys. Chem.* 98:4940–4948.
22. Sikorski, A., and P. Romiszowski. 2003. Thermodynamical properties of simple models of protein-like heteropolymers. *Biopolymers.* 69:391–398.
23. Junghans, C., M. Bachmann, and W. Janke. 2006. Microcanonical analyses of peptide aggregation processes. *Phys. Rev. Lett.* 97:218103–218106.
24. Junghans, C., M. Bachmann, and W. Janke. 2008. Thermodynamics of peptide aggregation processes: an analysis from perspectives of three statistical ensembles. *J. Chem. Phys.* 128:085103–085111.
25. Hernández-Rojas, J., and J. M. Gomez Llorente. 2008. Microcanonical versus canonical analysis of protein folding. *Phys. Rev. Lett.* 100:258104–258107.

26. Kim, J., T. Keyes, and J. E. Straub. 2009. Relationship between protein folding thermodynamics and the energy landscape. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 79:030902 (R).
27. Bereau, T., M. Bachmann, and M. Deserno. 2010. Interplay between secondary and tertiary structure formation in protein folding cooperativity. *J. Am. Chem. Soc.* 132:13129–13131.
28. Berg, B. A., and T. Neuhaus. 1991. Multicanonical algorithms for first order phase transitions. *Phys. Lett. B.* 267:249–253.
29. Wang, F., and D. P. Landau. 2001. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* 86:2050–2053.
30. Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.
31. Kumar, S., D. Bouzida, …, J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* 13:1011–1021.
32. Bereau, T., and R. H. Swendsen. 2009. Optimized convergence for multiple histogram analysis. *J. Comput. Phys.* 228:6119–6129.
33. Swendsen, R. H., and J. S. Wang. 1986. Replica Monte Carlo simulation of spin glasses. *Phys. Rev. Lett.* 57:2607–2609.
34. Dill, K. A., and D. Stigter. 1995. Modeling protein stability as heteropolymer collapse. *Adv. Protein Chem.* 46:59–104.
35. Baldwin, R. L. 1989. How does protein folding get started? *Trends Biochem. Sci.* 14:291–294.
36. Bereau, T., and M. Deserno. 2009. Generic coarse-grained model for protein folding and aggregation. *J. Chem. Phys.* 130:235106–235120.
37. Limbach, H. J., A. Arnold, …, C. Holm. 2006. ESPResSo—an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.* 174:704–727.
38. Walsh, S. T. R., H. Cheng, …, W. F. DeGrado. 1999. Solution structure and dynamics of a de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. USA.* 96:5486–5491.
39. Frishman, D., and P. Argos. 1995. Knowledge-based protein secondary structure assignment. *Proteins.* 23:566–579.
40. Scholtz, J. M., E. J. York, …, R. L. Baldwin. 1991. A neutral, water-soluble, α-helical peptide: the effect of ionic-strength on the helix coil equilibrium. *J. Am. Chem. Soc.* 113:5102–5104.
41. Shalongo, W., D. Laxmichand, and E. Stellwagen. 1994. Distribution of helicity within the model peptide acetyl(AAQAA)3amide. *J. Am. Chem. Soc.* 116:8288–8293.
42. Zagrovic, B., G. Jayachandran, …, V. S. Pande. 2005. How large is an α-helix? Studies of the radii of gyration of helical peptides by small-angle X-ray scattering and molecular dynamics. *J. Mol. Biol.* 353:232–241.
43. Ferrara, P., J. Apostolakis, and A. Caflisch. 2000. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. *J. Phys. Chem. B.* 104:5000–5010.
44. Chebaro, Y., X. Dong, …, N. Mousseau. 2009. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J. Phys. Chem. B.* 113:267–274.
45. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
46. Privalov, P. L. 1989. Thermodynamic problems of protein structure. *Annu. Rev. Biophys. Biophys. Chem.* 18:47–69.
47. Privalov, P. L. 1982. Stability of proteins. Proteins which do not present a single cooperative system. *Adv. Protein Chem.* 35:1–104.
48. Kaya, H., and H. S. Chan. 2000. Energetic components of cooperative protein folding. *Phys. Rev. Lett.* 85:4823–4826.
49. Ghosh, K., and K. A. Dill. 2009. Theory for protein folding cooperativity: helix bundles. *J. Am. Chem. Soc.* 131:2306–2312.
50. Badasyan, A. V., G. N. Hayrapetyan, …, V. F. Morozov. 2009. Intersegment interactions and helix-coil transition within the generalized model of polypeptide chains approach. *J. Chem. Phys.* 131:115104–115111.
51. Solc, K. 1971. Shape of a random-flight chain. *J. Chem. Phys.* 55:335–344.
52. Sciutto, S. J. 1996. The shape of self-avoiding walks. *J. Phys. Math. Gen.* 29:5455–5473.
53. Hills, Jr., R. D., and C. L. Brooks, 3rd. 2008. Subdomain competition, cooperativity, and topological frustration in the folding of CheY. *J. Mol. Biol.* 382:485–495.
54. Privalov, P. L. 1979. Stability of proteins: small globular proteins. *Adv. Protein Chem.* 33:167–241.
55. Zimm, B. H., and J. K. Bragg. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31:526–535.
56. Tiktopulo, E. I., V. E. Bychkova, …, O. B. Ptitsyn. 1994. Cooperativity of the coil-globule transition in a homopolymer: microcalorimetric study of poly(n-isopropylacrylamide). *Macromolecules.* 27:2879–2882.
57. Zhu, Y., D. O. V. Alonso, …, F. Gai. 2003. Ultrafast folding of α3D: a de novo designed three-helix bundle protein. *Proc. Natl. Acad. Sci. USA.* 100:15486–15491.
58. Nettels, D., S. Müller-Späth, …, B. Schuler. 2009. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 106:20740–20745.
59. Chan, H. S. 2000. Modeling protein density of states: additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins.* 40:543–571.
60. Booth, D. R., M. Sunde, …, M. B. Pepys. 1997. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature.* 385:787–793.