Evidence, from Simulations, of a Single State with Residual Native Structure at the Thermal Denaturation Midpoint of a Small Globular Protein

Gia G. Maisuradze,[†] Adam Liwo,[†,‡] Stanisław Ołdziej,[†,§] and Harold A. Scheraga*,[†]

Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland, and Intercollegiate Faculty of Biotechnology, University of Gdańsk, Medical Univeristy of Gdańsk, Kładki 24, 80-222 Gdańsk, Poland

Received April 14, 2010; E-mail: has5@cornell.edu

Abstract: The folding of the B-domain of staphylococcal protein A has been studied by coarse-grained canonical and multiplexed replica-exchange molecular dynamics simulations with the UNRES force field in a broad range of temperatures (270 K $\leq T \leq$ 350 K). In canonical simulations, the folding was found to occur either directly to the native state or through kinetic traps, mainly the topological mirror image of the native three-helix bundle. The latter folding scenario was observed more frequently at low temperatures. With increase of temperature, the frequency of the transitions between the folded and misfolded/unfolded states increased and the folded state became more diffuse with conformations exhibiting increased root-mean-square deviations from the experimental structure (from about 4 Å at $T = 300$ K to 8.7 Å at $T = 325$ K). An analysis of the equilibrium conformational ensemble determined from multiplexed replica exchange simulations at the folding-transition temperature ($T_f = 325$ K) showed that the conformational ensemble at this temperature is a collection of conformations with residual secondary structures, which possess native or near-native clusters of nonpolar residues in place, and not a 50−50% mixture of fully folded and fully unfolded conformations. These findings contradict the quasi-chemical picture of two- or multistate protein folding, which assumes an equilibrium between the folded, unfolded, and intermediate states, with equilibrium shifting with temperature but with the native conformations remaining essentially unchanged. Our results also suggest that long-range hydrophobic contacts are the essential factor to keep the structure of a protein thermally stable.

Introduction

To perform their functions in living organisms, most proteins must fold into a unique three-dimensional structure. Starting from the famous experiments of Anfinsen et al.,[1] the question of how proteins reach their biologically active ensembles of conformations by interconversion among ensembles of conformations in a back-and-forth progression from the non-native to the native state still remains to be answered.

One of the widely used models for the description of folding of single-domain proteins is the two-state model, deduced from calorimetric experiments,[2] consisting of the non-native and the native state separated by the energetically unfavorable transition state. In this scenario, at the folding-transition temperature (T_f), the non-native (fully unfolded) and native (fully folded) conformations are equally populated, that is, the equilibrium ensemble contains their 50−50% mixture. This means that, for a folding scenario characterized by single-exponential kinetics, a protein goes directly from the unfolded to the native state

without detectable intermediates on the way. However, many recent experimental studies[3−7] have shown that proteins can fold through intermediate states or undergo one-state downhill folding. This scenario of folding was pointed out by Poland and Scheraga 45 years ago.[8]

On the basis of the classical Landau theory for critical transitions, Munoz and Sanchez-Ruiz[5] showed that the appearance of a peak in heat-capacity curves can very well be explained in terms of a continuous order parameter (enthalpy) at various temperatures and not the competition between the well-defined folded and unfolded state. An analysis of experimental heat-capacity curves showed that the density of states (as a function of enthalpy) has either one maximum or two maxima, indicating either a barrierless folding (e.g., for 1BBL)

[†] Cornell University.

[‡] Faculty of Chemistry, University of Gdańsk.

[§] Intercollegiate Faculty of Biotechnology, University of Gdańsk.

(1) Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H. *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1309–1314.

(2) Privalov, P. L.; Khechinashvili, N. N. *J. Mol. Biol.* **1974**, *86*, 665–684.

(3) Garcia-Mira, M. M.; Sadqi, M.; Fisher, N.; Sanchez-Ruiz, J. M.; Munoz, V. *Science* **2002**, *298*, 2191–2195.

(4) Nguyen, H.; Jäger, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3948–3953.

(5) Munoz, V.; Sanchez-Ruiz, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17646–17651.

(6) Li, P.; Oliva, F. Y.; Naganathan, A.; Munoz, V. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 103–108.

(7) Jha, S. K.; Dhar, D.; Krishnamoorthy, G.; Udgaonkar, J. B. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11113–11118.

(8) Poland, D. C.; Scheraga, H. A. *Biopolymers* **1965**, *3*, 401–419.

or the existence of a barrier between the folded and unfolded states (e.g., for thioredoxin).[5]

By using simplified atomistic and coarse-grained models, Knott and Chan[9] demonstrated that sharp heat-capacity peaks and steeply increasing sigmoidal curves of the temperature dependence of the radius of gyration generally indicate the presence of a barrier (cooperative folding), while low and broad peaks and less steeply increasing curves of the radius of gyration (such as obtained for 1BBL) indicate barrierless (noncooperative) folding. However, the sigmoidal shape of the radius of gyration profile is not a reliable criterion for two-state behavior, and the heat-capacity baselines can introduce ambiguity.[9]

In this article, by studying the molecular dynamics trajectories for folding the B domain of staphylococcal protein A [1BDD (a three-α-helix bundle; 46 residues)],[10] generated with the coarse-grained united-residue (UNRES) force field[11,12] in a broad range of temperatures (270 K ≤ T ≤ 350 K), we demonstrate that the folding of protein A does not seem to be reducible to a simple "quasi-chemical" picture. Protein A was selected for this study because its folding dynamics and thermodynamics have been investigated extensively both experimentally[13−17] and theoretically.[18−22] On the other hand, using the coarse-grained UNRES force field enables us to extend the time scale of simulations by about 3 orders of magnitude compared to all-atom simulations[11] which is a tremendous advantage, even though accuracy is sacrificed because a simplified treatment is used. Regarding the accuracy of the representation of polypeptide chains in solution at various temperature, UNRES seems to be more appropriate than other coarse-grained force fields because it has been derived consistently as a temperature-dependent potential of mean force of a polypeptide chain in solution and parametrized to reproduce not only the folded structures of proteins but also partially unfolded and unfolded states at higher temperatures, as well as the thermodynamics of the folding transition.[12]

Results and Discussion

To determine the temperatures at which to carry out canonical MD runs, we first performed a multiplexed replica exchange (MREMD) simulation[23] of the system (see Materials and Methods). This calculation enabled us to determine the heat-

(9) Knott, M.; Chan, H. S. *Proteins* **2006**, *65*, 373–391.
(10) Gouda, H.; Torigoe, H.; Saito, A.; Sato, M.; Arata, Y.; Shimada, I. *Biochemistry* **1992**, *31*, 9665–9672.
(11) Liwo, A.; Khalili, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2362–2367.
(12) Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. A. *J. Phys. Chem. B* **2007**, *111*, 260–285.
(13) Bai, Y. W.; Karimi, A.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **1997**, *6*, 1449–1457.
(14) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552–558.
(15) Sato, S.; Religa, T. L.; Daggett, V.; Fersht, A. R. *Proc. Natl Acad. Sci. U.S.A.* **2004**, *101*, 6952–6956.
(16) Huang, F.; Sato, S.; Shape, T. D.; Ying, L.; Fersht, A. R. *Proc. Natl Acad. Sci. U.S.A.* **2007**, *104*, 123–127.
(17) Huang, F.; Lerner, E.; Sato, S.; Amir, D.; Haas, E.; Fersht, A. R. *Biochemistry* **2009**, *48*, 3468–3476.
(18) Boczko, E. M.; Brooks, C. L., III. *Science* **1995**, *269*, 393–396.
(19) Zhou, Y.; Karplus, M. *Nature* **1999**, *401*, 400–403.
(20) Vila, J. A.; Ripoll, D. R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14812–14816.
(21) Khalili, M.; Liwo, A.; Scheraga, H. A. *J. Mol. Biol.* **2006**, *355*, 536–547.
(22) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. *Phys. Rev. Lett.* **2009**, *102*, 238102.
(23) Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H. A. *J. Chem. Theory. Comput.* **2009**, *5*, 627–640.
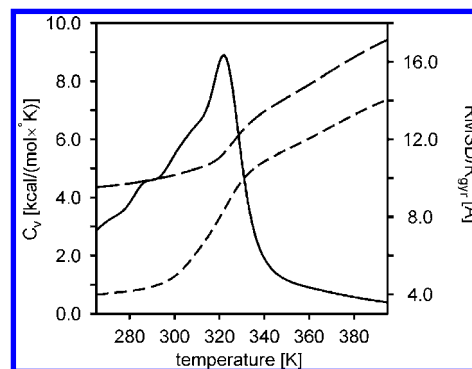
Figure 1. Heat-capacity (solid line), ensemble-averaged radius of gyration (long-dashed line), and ensemble-averaged rmsd (short-dashed line) of 1BDD, calculated in MREMD simulations.
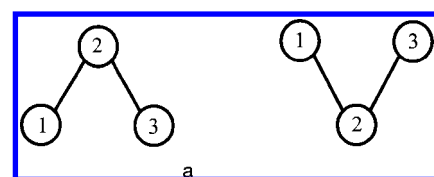

Figure 2. Scheme of the top view of (a) native and (b) mirror-image topology of the three-helix bundle of protein A. Each helix is represented by a circle with the number identifying its position in the chain (1, N-terminal; 2, middle; 3, C-terminal) and lines showing connections between the helices. The C-terminal α-helix runs from bottom to top.

capacity curve and ensemble-averaged quantities as functions of temperature. In comparison with multiple-trajectory canonical simulations, MREMD provides much faster convergence to the equilibrium ensemble. The temperature profiles of the heat capacity, the ensemble-averaged radius of gyration (R_{gyr}), and the C^{α} root-mean-square deviation (rmsd) from the experimental structure, obtained by processing the results of simulations with the weighted-histogram analysis method (WHAM),[24] are shown in Figure 1.

As shown in Figure 1, the heat capacity has a maximum at $T_f \approx 325$ K; this temperature corresponds to the inflection points in the ensemble-averaged rmsd and R_{gyr} as functions of temperature which identifies T_f as the folding-transition temperature.[25] It should also be noted that the rmsd keeps decreasing with temperature below the folding-transition temperature, which means that the native-like conformations have increasing thermodynamic stability with decreasing temperature.

To examine the folding pathways of 1BDD, we ran 16 0.5-μs long trajectories at $T = 270$ and 280 K (far below T_f), at $T = 310$ K (near T_f), and at $T = 325$ K (T_f), starting from the fully extended unfolded conformation. Our previous studies[21,22,26] showed that the simulated folding of 1BDD occurs either almost instantly from a fully unfolded conformation or, before jumping to the native state, the protein becomes trapped in a metastable state (the so-called mirror image of the native structure; Figure 2). In this study, we found that folding through the kinetic trap, forming quite a deep metastable state, occurs more frequently at low temperatures ($T = 270$ and 280 K) than near ($T = 310$ K) the folding-transition temperature. However, as demonstrated later in this section, this metastable state constitutes only a very

(24) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(25) Klimov, D. K.; Thirumalai, D. *J. Chem. Phys.* **1998**, *109*, 4119–4125.
(26) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. *J. Chem. Theory. Comput.* **2010**, *6*, 583–595.
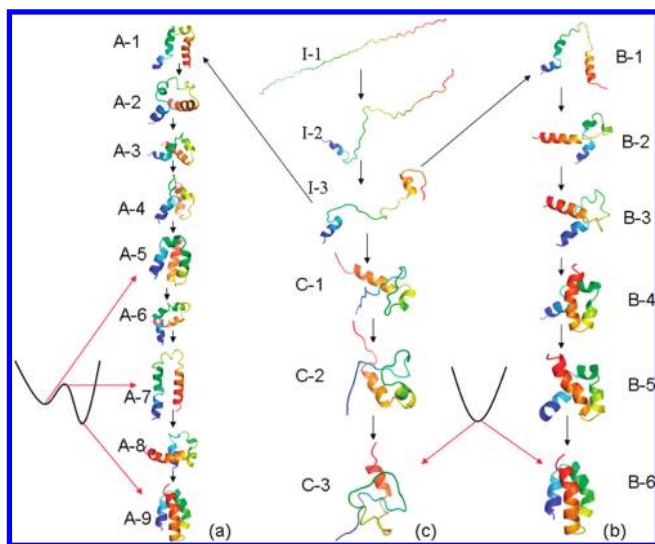
Figure 3. Structures and FEPs of pathways corresponding to (a) folding through a kinetic trap, with such pathways occurring most frequently at low temperatures (T = 270 and 280 K); (b) downhill folding, with such pathways occurring near (310 K) but below the folding-transition temperature (T = 325 K); and (c) residual folding, with such pathways occurring at the folding-transition temperature (T = 325 K). The structures are colored from blue to red from the N- to the C-terminus. A-5, A-7, and A-9 are representative structures of a kinetic trap (mirror image), the transition state, and the native state, respectively. B-6 and C-3 correspond to native and residually folded structures, respectively.

tiny fraction of the conformational ensemble at T = 300 K and below, and its appearance and temporary persistence on the folding pathways results from the fact that a free-energy barrier has to be overcome to unpack the helices once they are incorrectly packed. Ultimately, when the simulations at T = 300 K and lower are continued, nearly all trajectories end up in the native topology. The dependence of the folding pathway on temperature and friction was also demonstrated in our earlier all-atom study of protein A.[27] Structures along representative folding pathways, together with the corresponding nonequilibrium free-energy profiles (FEPs) (i.e., corresponding to the canonical-simulation trajectories of finite length) mapped onto the conformations encountered on the folding pathways are shown in Figure 3 as pathways a and b, respectively. The behaviors of the protein in the initial stages for both pathways resemble each other (structures I-1, I-2, I-3); however, folding starts to deviate after that. In particular, pathway (a) (from A-1 to A-9) describes the folding scenario occurring in most of the trajectories at T = 270 and 280 K and in a few trajectories at T = 310 K; pathway (b) (from B-1 to B-6) shows how 1BDD folds most of the time at T = 310 K and a few times at T = 270 and 280 K. It should also be noted that the protein does not remain in the folded state permanently but, sometimes, jumps back to the unfolded/misfolded state. This feature indicates that the system is ergodic. As expected, interconversions between the folded and misfolded/unfolded states are more frequent at T = 310 K than at T = 280 K.

Pathway (c) (from C-1 to C-3) represents trajectories at the folding-transition temperature (325 K). From the simple two-state picture (in which the native state is in equilibrium with the unfolded state and, thus, only the proportion of these two states, but not the shape of the conformations constituting the

native ensemble, depends on temperature), we would expect a behavior qualitatively similar to that at T = 310K with only more frequent transition between the folded and unfolded/misfolded conformations. Instead, the conformations of 1BDD at the minimum of the FEP are only partially unfolded (or "residually folded"), with high rmsd ∼8.7 Å, which contradicts the established idea that only fully folded and fully unfolded species coexist at the transition temperature.[2] In particular, the α-helical structure is present only in the C-terminal part in these conformations. On the other hand, most of these conformations at T = 325K do not have a very much greater radius of gyration with respect to that in the experimental structure (Figure 1) with the N-terminal and middle (unfolded) parts partially packed to the C-terminal α-helix which persists in this partially unfolded state.

For deeper insight into the properties of the conformations at the folding-transition temperature, we used MREMD simulations and WHAM to determine the equilibrium canonical conformational ensemble at T = 325 K, as described in our earlier work.[12] Such a procedure produces more reliable ensembles than those obtained by canonical simulations at a given temperature, the latter being prone to nonequilibration. In Figure 4, representative conformations of the four clusters with the highest probability of this ensemble at T = 325 K are shown. It can be seen that their common characteristics are (i) chain reversals close to sequence regions that form loops in the native structure of 1BDD and (ii) a cluster that includes the following nonpolar residues which make contacts present in the experimental structure (Figure 4a): Ala13, Phe14, Ile17, and Leu18 (at the N-terminal α-helix in the experimental structure), Phe31, Ile32, and Leu35 (at the middle α-helix), and Leu45, Leu46, and Ala49 (at the C-terminal α-helix), even though the detailed contact pattern is not exactly the same as in the experimental structure. These residues make contacts to bind the three segments of the chain together. It should be noted that the segments, which form α-helices in the experimental structure, have only some secondary (α-helical) structure in the conformations from the ensemble at T = 325 K. The contacts between the residues listed above in the experimental and selected calculated conformations are shown in Figure 4 and are also listed in Table 1.

Additionally, for conformations shown in Figure 4b,d, the neighboring nonpolar residues, namely, Ala43 and Tyr15, respectively, also take part in binding the N-terminal and the C-terminal, or the N-terminal and the middle α-helix, respectively. For all conformations at the folding-transition temperature, non-native contacts are formed (Table 1). Consequently, it can be concluded that the three patches of nonpolar residues in the N-, middle, and C-terminal parts provide the necessary redundancy to overcome the unfavorable loop-closing entropy and to form long-range contacts between the N- and the C-terminal part, which seems to be a necessary condition for this protein to fold.

Usually, the C-terminal helix is the most mature (Figure 4b,d), as also found earlier in other experimental[13] and theoretical[28] studies, but conformations without any C-terminal α-helix (but still preserving hydrophobic pairing) are also observed (Figure 4c,e). The prototypes of both the native structure (Figure 4b,c) and the topological mirror image (Figure 4d,e) can be found.

The gradual melting of protein A and appearance of topological mirror-image conformations with increasing temperature are

(27) Jagielska, A.; Scheraga, H. A. *J. Comput. Chem.* **2007**, *28*, 1068–1082.

(28) Ghosh, A.; Elber, R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10394–10398.
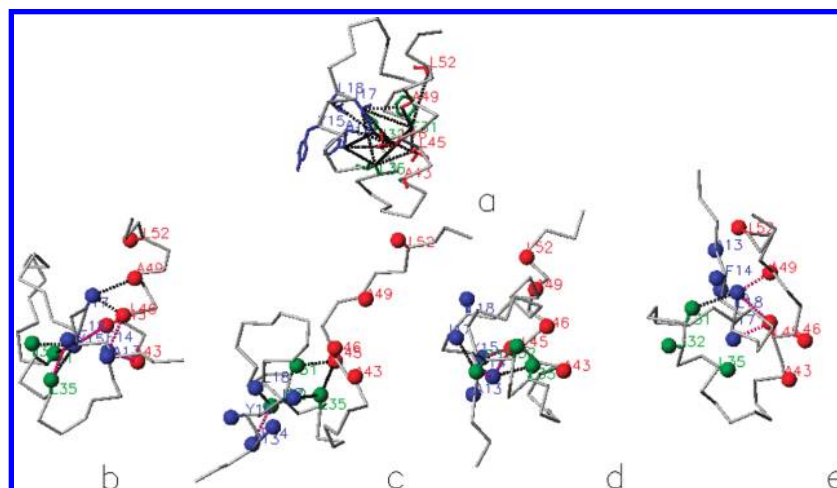
Figure 4. Experimental structure of 1BDD (a) and four representatives from the canonical ensemble determined by MREMD simulations at $T = 325$ K (b−e) with nonpolar residues (Ala13, Phe14, Ile17, Leu18, Phe31, Ile32, Leu35, Leu45, Leu46, Ala49, and Leu52) that make contact between helices in the N-terminal and the middle or the N-terminal and the C-terminal α-helix in the experimental structure and in the conformations of the ensemble at $T = 325$ K, shown as bonds between heavy atoms for the experimental structure or as spheres at the united side chain centers for the UNRES structures. The residues of the N-terminal, middle, and C-terminal helices are colored blue, green, and red, respectively. Native contacts between the above-mentioned nonpolar residues are shown as black dashed lines, and non-native contacts (in panels b−e) are shown as pink dashed lines. It is shown that a cluster composed of all or some of the above-mentioned nonpolar residues is preserved regardless of the formation of secondary structure or topology which is similar to that of the native structure in conformations (b) and (c) or mirror image (d) and (e). Additionally, in the structures shown in panels (b) and (d), the neighboring nonpolar residues that are not involved in the formation of long-range contacts in the experimental structure, Ala43 and Tyr15, respectively, can assist in the formation of long-range contacts between the N-terminal, middle, and C-terminal segments of the chain.

Table 1. Contacts between Core Hydrophobic Residues in the Experimental Structure of Protein A and in the Sample Conformations of the Ensemble at the Folding-Transition Temperature Shown in Figure 4b−e[a]

	native topology		mirror image	
experimental structure	Figure 4b	Figure 4c	Figure 4d	Figure 4e
	Helix 1− Helix 2			
		Ala13−Ile32	Phe14−Phe31	
Phe14−Ile32	-	+	-	-
Phe14−Leu35	-	-	+	-
	Tyr15−Leu35			-
Ile17−Phe31	-	-	+	-
Ile17−Ile32	-	-	-	-
Ile17−Leu35	-	+	-	-
Leu18−Phe31	+	-	-	+
Leu18−Ile32	+	+	-	-
Leu18−Leu35	+	-	-	-
	Helix 1−Helix 3			
Ala13−Leu46	-	-	-	-
	Phe14−Ala43		Phe14−Leu45	Ile17−Leu45
			Tyr15−Leu45	
Phe14−Leu46	+	-	-	-
Ile17−Leu46	+	-	-	-
Ile17−Ala49	+	-	-	-
	Leu18−Leu45		-	Leu18−Leu45
				Leu18−Ala49
	Helix 2−Helix 3			
Phe31−Leu45	-	-	-	-
Phe31−Leu46	-	+	-	-
Phe31−Ala49	-	-	-	-
Phe31−Leu52	-	-	-	-
Ile32−Leu45	-	-	-	-
Ile32−Leu46	-	-	-	-
Ile35−Leu45	-	-	-	-
Ile35−Leu46	-	+	-	-

[a] The native contacts are specified in the first column, while their presence or absence in the other conformations is marked by a "+" or a "−", respectively. Non-native contacts are specified explicitly in columns 2−5.

illustrated in Figure 5, where the distributions of the C^{α} rmsd from the experimental 1BDD structures of the conformations with the native- and mirror-image topology, respectively, are shown. It can be seen that the mirror-image conformations are very rare at $T = 300$ K and their population increases with

temperature, reaching nearly 50% at the folding-transition temperature, after which it remains effectively unchanged.

At first glance, such behavior appears qualitatively indicative of a two-state model. However, it can be seen that the center of the rmsd distribution corresponding to the conformations with
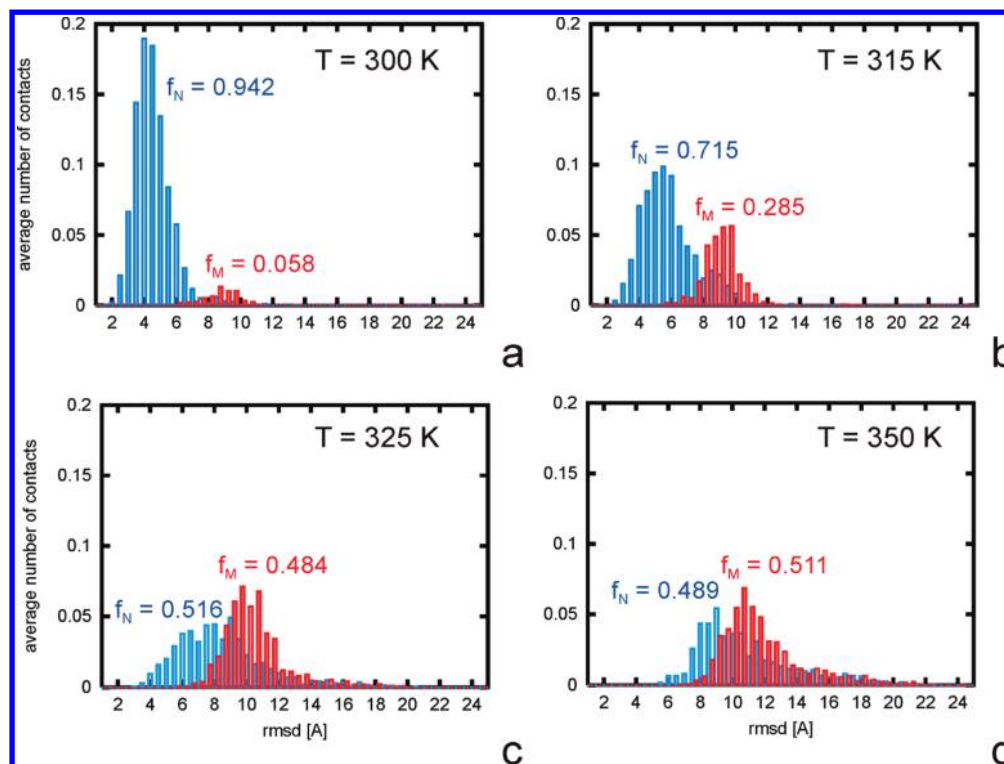
Figure 5. Distributions of C^α rmsd from the NMR structure of 1BDD, at (a) $T = 300$ K, (b) $T = 315$ K, (c) $T = 325$ K (T_f), and (d) $T = 350$ K, of the conformations with native (blue) and mirror-image (red) topology of the three-helix bundle. The fractions of the conformations with native (f_N) and mirror-image bundle topology (f_M) are shown in the respective panels. The bin size is $\Delta d = 0.5$ Å.

the native topology moves steadily to higher rmsd values as the temperature increases, reaching about 8 Å at the folding-transition temperature (Figure 5c). This observation conforms to the qualitative picture of the conformational ensemble at the folding-transition temperature (Figure 4b–e) as an ensemble of residually folded conformations and not that consisting of perfectly folded and unfolded conformations. Furthermore, if the mirror-image state could be identified with the "unfolded" state in the two-state model, its population should tend to 1 and not only to 1/2 with the increase of temperature, while the population of the "folded" state should tend to zero.

From Figure 5, it can also be concluded that the appearance of a mirror-image does not seem to be the artifact of the force field used. At low temperatures, the mirror-image conformations are scarce, which means that the force field correctly favors the native topology. At the folding-transition and higher temperatures, the populations of the conformations with the native and mirror-image topology are nearly equal. This feature can be expected because, as the temperature increases, the entropic factor, which favors equal population of conformations with each bundle chirality, takes over.

In Figure 6, the probability of states is plotted as a function of the order parameter q defined in ref 12 (q is a measure of the deviation of the distances in a given structure from those in the experimental structure; see eqs 8–11 of ref 12) and energy U at three selected temperatures. It can be seen that both series of plots indicate single-state behavior.

The preservation of the side chain contact pattern and the persistence of the C-terminal α-helix relative to the N-terminal and middle helices in the partially thermally unfolded simulated structure of protein A is a general feature of the conformational ensemble at the folding-transition temperature. Figure 7b–d shows the side chain/side chain and hydrogen-bonding contact probability maps of the conformational ensembles at $T = 300$
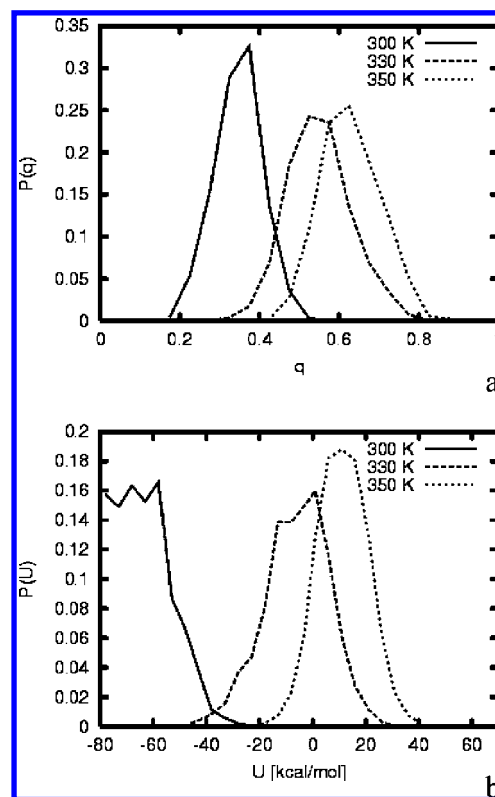

Figure 6. Plots of the probability of states at $T = 300$ K (solid lines), $T = 330$ K (short-dashed lines), and $T = 350$ K (dotted lines) as a function of (a) the order parameter q; (bin size $\Delta q = 0.1$) and (b) the UNRES energy (U; bin size $\Delta U = 5$ kcal/mol).

K, $T = 325$ K (the folding-transition temperature), and $T = 350$ K (at which temperature the protein is denatured). For
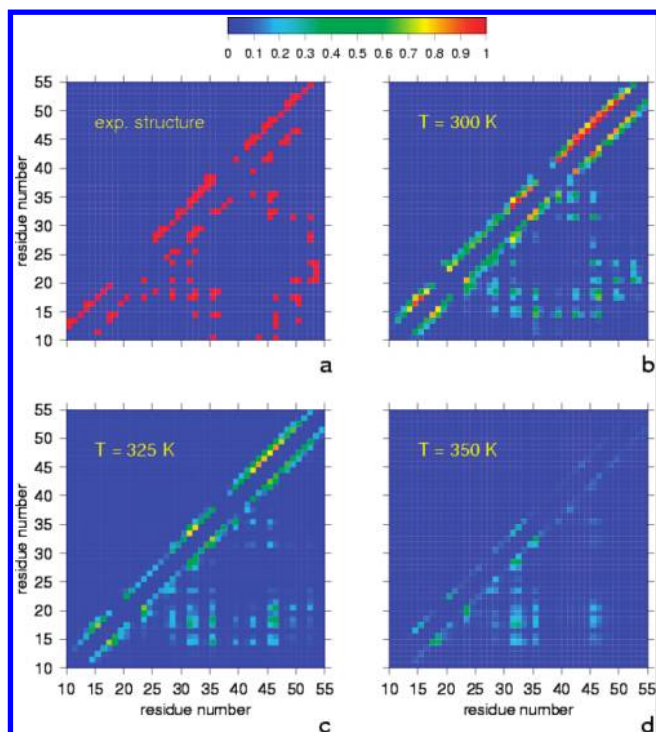
Figure 7. (a) Map of the side chain/side chain (right triangle) and side backbone peptide-group−hydrogen-bonding contacts (left triangle) of the experimental structure of 1BDD. (b− d) Probability maps of these contacts corresponding to the conformational ensembles at (b) $T = 300$ K, (c) $T = 325$ K, and (d) $T = 350$ K. The probability color scale is shown in the upper panel.

comparison, the contact map of the experimental structure of 1BDD is shown in Figure 7a.

The side chain contact pattern of the experimental structure clearly shows the contacts between the N-terminal and the middle, the N-terminal and the C-terminal, and somewhat fewer contacts between the middle and the C-terminal α-helices involving the residues listed above. The backbone−hydrogen-bonding-contact pattern clearly shows the three α-helices (Figure 7a). At $T = 300$ K (Figure 7b), the ensemble-averaged side chain/side chain and backbone−hydrogen-bonding-contact patterns are somewhat more diffuse than that of the experimental structure; this is a feature of the medium-resolution UNRES force field. However, the contacts are essentially the same as in the experimental structure.

At the folding-transition temperature (Figure 7c), the side chain contacts have a lower probability but are still firmly established, while the hydrogen-bonding contacts start to die out except those of the C-terminal α-helix. Finally, at $T = 350$ K, only some isolated backbone−hydrogen-bonding contacts remain, while the long-range side chain/side chain contacts between the N-terminal and the middle and the N-terminal and the C-terminal α-helices have considerable probability. This observation suggests that these hydrophobic contacts form a folding nucleus, as proposed earlier by Matheson and Scheraga[29] (however, we must keep in mind that we analyze thermal equilibrium unfolding here, i.e., the order of stability of the elements of the structure and the order of the formation of structural elements at a subfolding transition temperature need not be the same).
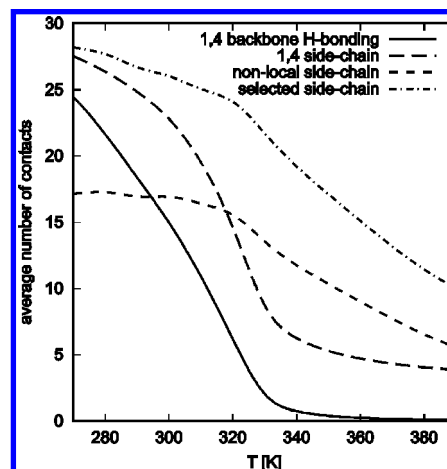
(29) Matheson, R. R., Jr.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 819–829.

Figure 8. Plots of the ensemble-averaged number of 1,4-backbone-hydrogen-bonding contacts (characteristic of α-helix; solid line; inflection point at $T_i = 315$ K, 32% of maximum number of contacts at T_i), 1,4-side-chain contacts (long-dashed line; $T_i = 320$ K, 50% of the maximum number of contacts at T_i), all long-range side-chain contacts (short-dashed line; $T_i = 330$ K, 73% of the maximum number of contacts at T_i), and the contacts between the selected hydrophobic side chains that form the hydrophobic core and are labeled in Figure 4 (dot-dashed lines; $T_i = 330$ K, 78% of the maximum number of contacts at T_i).

The change of contact probability with temperature is also illustrated in Figure 8, where the ensemble-averaged number of 1,4-contacts between the peptide groups (characteristic of an α-helix), that of the 1,4- and long-range (i.e., farther than 1,4) contacts between the side chains, as well as that of the long-range contacts between the nonpolar side chains indicated in Figure 4, are plotted as functions of temperature. It can be seen that the number of peptide-group contacts decays with temperature first and the last to go are the long-range contacts between the nonpolar residues that form the hydrophobic cluster. Another important observation is that, if the "folding-transition" temperature were determined from the position of the inflection point in the temperature dependence of the number of contacts of a particular category, each of these temperatures would be different and these temperatures would differ from that of the position of the heat-capacity peak. Because the average number of contacts of a particular category can, to some extent, be identified with such experimental quantities as ellipticity (the number of 1,4-contacts between the backbone peptide groups) or fluorescence-quenching intensity (the number of nonlocal contacts between side chains), the above conclusion is the same as that drawn in our recent experimental study of the folding of the C-terminal β-hairpin of the B1 domain of protein G.[30]

On the basis of the graphs of Figure 8, it can, additionally, be concluded that in no property does the position of the inflection point in the temperature dependence of the ensemble-averaged number of contacts correspond to the decrease of the number of contacts to 50%; it is either smaller (for the number of backbone peptide group contacts) or higher (for the number of long-range side-chain contacts); only the number of 1,4 side-chain contacts is 50% of the maximum value at the inflection point. Thus, not only do the folded conformations not seem to constitute 50% of the equilibrium ensemble at the folding-transition temperature, but it does not appear easy to select any conformation-dependent observable which would exhibit such behavior.

(30) Lewandowska, A.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Biopolymers* **2010**, *93*, 469–480.

The probe dependence of the position of the inflection point of the folding/unfolding curve (and, thereby, that of the "folding-transition temperature" determined from the inflection point) illustrated in Figure 8 conforms to findings in earlier experimental and theoretical downhill folding studies of BBL,[3,31−34] gpW,[35] and some histidine-containing mutants of λ_{6-85},[36] and in our recent study of the folding of peptides derived from the C-terminal fragment of the B3 domain of immunoglobulin-binding protein G.[37,38]

Fersht and co-workers[16,17] have recently carried out FRET studies of fluorophore-labeled protein A denatured to various extents by guanidinium chloride at concentrations up to 7 M. In their first study,[16] they found that the distribution of quenching efficiency was bimodal with the relative intensity of the band corresponding to greater efficiency decreasing and that of the smaller efficiency decreasing. In their second study, residues 22 and 55, which are located at the ends of helices 2 and 3 and are close to each other in the experimental structure, were labeled and their distance distribution determined from the FRET experiments. For up to 5 M GdmCl concentrations, the distance distribution turned out to be bimodal, and the authors attributed the band at shorter distances (about 15 Å) to the folded state and the second band (at about 25 Å) to the denatured state. In Figure 9, we present the $C^{\alpha}_{10} \cdots C^{\alpha}_{55}$ and $C^{\alpha}_{22} \cdots C^{\alpha}_{55}$ distance distribution, respectively, at $T = 300, 315, 325$, and 350 K obtained from our simulations. The end-to-end distance distributions (Figure 9a) seem unimodal, which does not seem to agree with the FRET results of Fersht and co-workers.[16] Closer examination of the distance distributions in Figure 9a shows that they are not symmetric and can consist of two closely overlapping peaks, one at about 25 Å and one at about 35 Å, respectively; this bimodality is most pronounced at $T = 325$ K (the folding-transition temperature). However, the lobes are broad and, therefore, overlap quite closely. It should be noted that the Förster distance (R_0) for the donor−acceptor pair (Alexa Fluor 488 and Alexa Fluor 647) used by Fersht and co-workers[16] is about 56 Å, while only the largest distances obtained in our simulations, even at $T = 350$ K, reach this value; therefore, the range of the distances obtained in our simulations is apparently narrower than that in the chemical-denaturation study of ref 16. It is likely that chemical denaturant forces increase the dimension of the chain to a greater extent (because of interactions with the peptide groups) than does the increase of temperature (we note that the strength of hydrophobic interactions, which promote less-extended structures, increases with increasing temperature).

The distributions of the $C^{\alpha}_{22} \cdots C^{\alpha}_{55}$ distance shown in Figure 9b are in closer agreement with those between photolabeled residues 22 and 55 determined by Fersht and co-workers by FRET of chemically denatured protein A.[17] The distributions are apparently bimodal with the first band centered at about 11−15 Å and the second one at about 25 Å, even though the
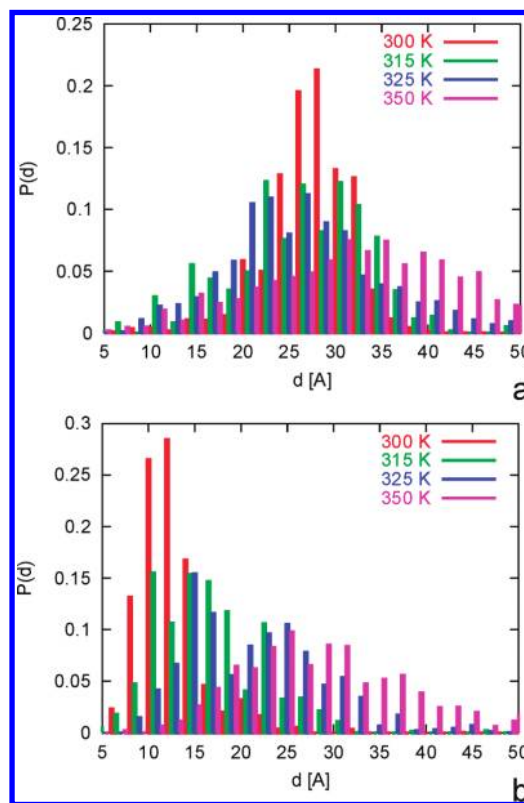
(31) Sadqi, M.; Fushman, D.; Munoz, V. *Nature* **2006**, *442*, 317–321.
(32) Zuo, G.; Wang, J.; Wang, W. *Proteins* **2006**, *63*, 165–173.
(33) Zhang, J.; Li, W.; Wang, J.; Qin, M.; Wang, W. *Proteins* **2008**, *72*, 1038–1047.
(34) Pitera, J. W.; Swope, W. C.; Abraham, F. F. *Biophys. J.* **2008**, *94*, 4837–4846.
(35) Fung, A.; Li, P.; Godoy-Ruiz, R.; Sanzhez-Ruiz, J. M.; Munoz, V. *J. Am. Chem. Soc.* **2008**, *130*, 7489–7495.
(36) Lui, F.; Gruebele, M. *J. Mol. Biol.* **2007**, *370*, 574–584.
(37) Skwierawska, A.; Makowska, J.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Proteins: Struct., Funct., Bioinf.* **2009**, *75*, 931–953.
(38) Skwierawska, A.; Zmudzińska, W.; Ołdziej, S.; Liwo, A.; Scheraga, H. A. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 637–654.

Figure 9. Histograms of the end-to-end distance distribution (a) and the distributions of the distances between the C^{α} atoms of Gln10 and Ala55 (b) at $T = 300, 315, 325$, and 350 K. Colors correspond to temperatures, as defined in the insets. The bin size is $\Delta d = 2$ Å.

position of the first band shifts to the right with temperature. The bimodality is especially evident at the folding-transition temperature. Thus, gradual unfolding of all of the protein A molecules (which is apparent because of the presence of only single maxima in the graphs of the probability of states in Figure 6) rather than an equilibrium between folded and unfolded conformations at every temperature, which we found in this study, can also generate a bimodal distribution of selected interresidue distances. The explanation of this fact can be deduced from Figures 4b−e and 8. Even though the secondary structure at the folding-transition temperature is largely destroyed in all of the conformations, contacts between the segments that form helices 2 and 3 in the native structure are still preserved in many of them (from Figure 8 we find that the fraction of conformations with backbone α-helical contacts is much more depleted at the folding-transition temperature than that with long-range side chain contacts) and such conformations contribute to the first lobe of the distance distribution. The presence of conformations with the long-range side chain contacts broken at the folding-transition temperature (and also at temperatures lower but close to it) contributes to the band centered at 25 Å.

Conclusions

The "quasi-chemical" two-state picture of protein folding at equilibrium, which is used to interpret experimental data and calculate thermodynamic parameters of folding, effectively identifies folding at different temperatures with the establishment of a chemical equilibrium between the folded and unfolded state. According to this picture, the change of temperature just changes the ratio between the folded and unfolded state. The confor-

mational ensemble at the folding-transition temperature consists of a 50−50% mixture of the fully folded and fully unfolded state. This simple model perfectly explains the appearance of the heat-capacity peak at the folding-transition temperature, a feature that constitutes the less-common thermodynamic definition of the folding-transition temperature.[25,39,40]

Instead of the shift of equilibrium between the folded and unfolded conformations with temperature, in the present study, we found that the bulk of the conformations of protein A becomes gradually unfolded with increasing temperature and, at the folding-transition temperature, the conformational ensemble of protein A is a collection of residually folded structures and not a 50−50% mixture of native and non-native conformations. This result is in agreement with experimental[5,6,31] and recent simulation studies of BBL.[32−34] A good illustration of this agreement is the folding dynamics, observed in BBL by a FRET experiment, sketched in Figure 4 of ref 6. We have also demonstrated that the bimodal distribution of distances between residues 22 and 55, found recently by Fersht and co-workers by FRET monitoring of the chemical denaturation of protein A,[17] is also obtained from the results of our simulations; the bimodal end-to-end distance found by these authors[16] also does not seem to be incompatible with our picture in which all of the molecules gradually unfold.

It is remarkable that the UNRES force field was parametrized using our hierarchical method which assumes a modified quasi-chemical approximation of the temperature-dependent equilibrium between species with various native-like segments (although the force field was optimized with a protein unrelated to 1BDD)[12] and, nevertheless, the results obtained in our study strongly suggest gradual formation/breaking of native-structure elements in all of the molecules rather than an equilibrium between molecules with those elements (e.g., α-helices or completely unfolded structures) intact.

The only kind of "equilibrium" that we found was that between the conformations with the native three-helix bundle of protein A and its topological mirror image. However, this equilibrium does not fit into the two-state model because, first, the topological mirror image appears to be a kinetic trap in the folding which lives for a longer time as the temperature decreases; second, the populations of the conformations with the native and mirror-image topology become equal (as dictated by the entropic factor) at higher temperatures and; third, near the folding-transition temperature, the helices are virtually unfolded except for the C-terminal one.

In conclusion, the folding transition of protein A is not of the all-or-none type but involves gradual melting of protein conformation, and the conformational ensemble at the thermodynamic folding-transition temperature might not contain any amount of native conformations. To generalize this finding to other proteins, more studies of the folding of the proteins of different structural classes (β and α + β) are required. These studies are currently being carried out in our laboratory. Nevertheless, it should be noted that our results appear reasonable in terms of functionality, that is, if a protein contained 50% of perfectly folded structure at the folding-transition temperature, it would retain some function. However, many enzymes lose most of their activity at lower temperature than the folding-

transition temperature.[41,42] Using the two-state model or, in general, any model involving a given number of well-defined states to analyze protein folding might, therefore, lead to a wrong description of this phenomenon and the thermodynamic quantities derived from such an analysis might, therefore, have no completely clear meaning.

It can also be concluded from our study that nonlocal interactions between the nonpolar side chains are the factor that preserves a rough protein shape even at quite high temperatures, at which the hydrogen-bonding interactions do not hold and the secondary structure is, therefore, melted. On the other hand, even the number of contacts between the nonpolar residues or any other conformation-dependent observable does not seem to follow a "quasi-chemical" picture with 50% content at the folding-transition temperature.

Materials and Methods

We used the coarse-grained UNRES model of polypeptide chains developed in our laboratory,[11,12,43−45] which we recently extended to perform molecular dynamics simulations;[46,47] we also implemented[23] multiplexed-replica exchange (MREMD) in UNRES. In the UNRES force field, a polypeptide chain is represented by a sequence of α-carbon (C^α) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p). Each united peptide group is located in the middle between two consecutive α-carbons.[45] Only these united peptide groups and the united side chains serve as interaction sites, the α-carbons serving only to define the chain geometry. The UNRES force field has been derived as a restricted free energy (RFE) function[44] of an all-atom polypeptide chain plus the surrounding solvent, where the all-atom energy function is averaged over the degrees of freedom that are lost when passing from the all-atom to the simplified system. Consequently, the UNRES effective energy function depends on temperature and this dependence has been incorporated in the version of UNRES used in our present study.[12] The complete effective energy function was optimized by using our hierarchical-optimization procedure[48] to reproduce the structure and folding thermodynamics of selected training proteins. The hierarchical procedure is based on the division of the conformational ensemble into levels, each of which contains conformations with a different number of native-like elements; level 0 contains completely unfolded structures and the highest level contains native-like structures. Intermediate levels contain partially folded conformations, the number of the level increasing with decreased thermal stability of a structure (e.g., if conformations with a single α-helix survive until the highest temperatures, they constitute level 1). Optimization is aimed at ordering the free energies of the levels such that they decrease with level number below the folding-transition temperature, become equal at the folding-transition

(39) Makhatadze, G. I.; Privalov, P. L. *Adv. Protein Chem*. **1995**, *47*, 307–425.

(40) Prabhu, N. V.; Sharp, K. A. *Annu. Rev. Phys. Chem*. **2005**, *56*, 521–548.

(41) Bayol, A.; Dupin, P.; Boe, J. F.; Claudy, P.; Letoffe, J. M. *Biophys. Chem*. **1995**, *54*, 229–235.

(42) Zavodszky, P.; Kardos, J.; Svingor, A.; Petsko, G. A. *Proc. Natl. Acad. Sci. U.S.A*. **1998**, *95*, 7406–7411.

(43) Liwo, A.; Ołdziej, S.; Pincus, M. R.; Wawak, R. J.; Rackowsky, S.; Scheraga, H. A. *J. Comput. Chem*. **1997**, *18*, 849–873.

(44) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Chem. Phys*. **2001**, *115*, 2323–2347.

(45) Liwo, A.; Czaplewski, C.; Ołdziej, S.; Rojas, A. V.; Kaźmierkiewicz, R.; Makowski, M.; Murarka, R. K.; Scheraga, H. A. "Simulation of protein structure and dynamics with the coarse-grained UNRES force field". In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth; G., Ed., CRC Press: Boca Raton, FL, 2008; Chapter 8, pp 107–122.

(46) Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2005**, *109*, 13798–13810.

(47) Rakowski, F.; Grochowski, P.; Lesyng, B.; Liwo, A.; Scheraga, H. A. *J. Chem. Phys*. **2006**, *125*, 204107.

(48) Liwo, A.; Arłukowicz, P.; Czaplewski, C.; Ołdziej, S.; Pillardy, J.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A*. **2002**, *99*, 1937–1942.

temperature, and the free-energy relations become inverted after the folding-transition temperature. In a sense, the procedure implies adapting a "quasi-chemical approximation" to represent the thermal folding/unfolding process. With the optimized force field, the heat-capacity curves of the training proteins have peaks at the respective folding-transition temperatures. In this study, we used the version of the force field parametrized on the 1GAB protein.[12] The theoretical basis of the force field and its molecular-dynamics implementation are described in detail in our earlier papers[43,44,46,47] and in a recent book chapter.[45]

Canonical and MREMD simulations were started from an extended conformation. Because the solvent is implicit in UNRES, the Berendsen thermostat[49] was used with the coupling parameter $\tau = 48.9$ fs. To integrate the equations of motion, we developed an adaptive multiple time step (A-MTS) algorithm[47] which is based on the velocity Verlet algorithm.[50] The integration time step was 5 fs. MREMD[23] simulations were performed with 80 trajectories in the range of temperatures from 250 to 500 K, and 28 000 000 MD steps for each trajectory. To determine the probabilities of the conformations at various temperatures and, consequently, to calculate ensemble-averaged quantities, we used the WHAM[24] algorithm implemented in UNRES as described in our earlier work.[12]

To determine the contact maps in the experimental 1BDD structure, we proceeded as follows. Two peptide groups p_1 and p_2 were considered to be at contact if $d(H_{p1}\cdots O_{p2}) < 2.5$ Å and $\alpha(N_{p1}-H_{p1}\cdots O_{p2}) > 90°$ or $d(H_{p2}\cdots O_{p1}) < 2.5$ Å and $\alpha(N_{p2}-H_{p2}\cdots O_{p1}) > 90°$, where d denotes the distance and α the planar angle. Two side chains in the experimental structure were considered at contact if the distance between any non-hydrogen atom of the first and any non-hydrogen atom of the second one was less than 5 Å. To determine if two peptide groups are in contact in a calculated structure, we computed their average energy of interaction from eq 5 of ref 51. If this energy was lower than -0.3 kcal/mol, the two groups were considered to be at contact. To

determine if two side chains of a calculated conformation were in contact, we used the contact function defined by eq 8 of ref 52. If the value of the contact function was lower than the critical value corresponding to given types of side residues, the side chains were considered to be at contact. The ensemble-average probability of a given contact was calculated from eq 1.

$$P_{ij}(T) = \sum_k h_{ijk} P_k(T) \qquad (1)$$

where $P_{ij}(T)$ is the probability of the contact between peptide groups/side chains i and j at temperature T, $P_k(T)$ is the probability of conformation k at temperature T (calculated by using WHAM as described in ref 12), and h_{ijk} is defined by eq 2.

$$h_{ijk} =$$
$$\begin{cases} 1 & \text{if peptide groups/side chains } i \text{ and } j \text{ are at contact in conformation } i \\ 0 & \text{otherwise} \end{cases}$$

$$(2)$$

(49) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(50) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637–649.

(51) Liwo, A.; Pincus, M. R.; Wawak, R. J.; Rackovsky, S.; Scheraga, H. A. *Protein Sci.* **1993**, *2*, 1715–1731.

(52) Ołdziej, S.; Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. A. *J. Phys. Chem. B* **2004**, *108*, 16934–16949.